

# GWAS Analysis of Young Adult Nodular Sclerosis Hodgkin Lymphoma

## A CCSS Concept Proposal

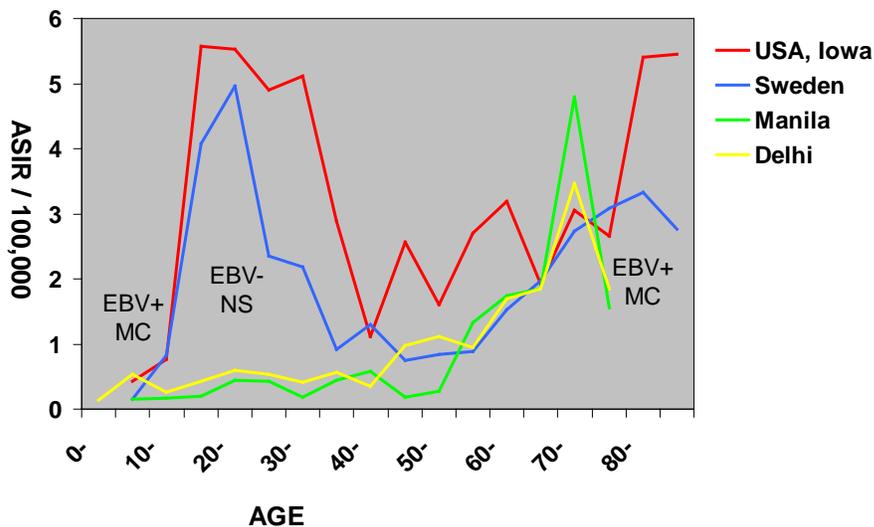
Submitted by Kenan Onel, David Conti, Wendy Cozen (Co-PI's)  
and Co-Investigators Andrew Skol, James R. Cerhan, Timothy Best, Dalin Li, Sally Glaser and Tom Mack

### Rationale

#### *General Background*

NSHL is the most common subtype of Hodgkin lymphoma in young adults, comprising between 60-80% of the disease in this age group; it occurs equally in males and females, peaks in incidence in the early 20's, and is associated with higher socioeconomic status and low prevalence of Epstein-Barr virus in the tumors (1,2). Although commonly classified together with the mixed cellularity subtype as "classical" Hodgkin lymphoma (HL), NSHL is now considered to be a specific etiologic entity separate from other types of HL because of its distinct pathology and epidemiology (3). Since age (young adult, roughly 10-49 years at diagnosis), histology (nodular sclerosis) and EBV tumor status (EBV negative) are correlated, some studies have used only one of these variables to represent the general young adult nodular sclerosis EBV negative subtype, which is believed to be a distinct entity (4).

### Age-specific incidence rates (ASIR) of Hodgkin lymphoma in females, 1993-1997



Cancer in Five Continents, Vol VIII, IARC

HL, especially in young adults, is an immunologically active tumor whose relatively rare neoplastic giant cells (Hodgkin-Reed Sternberg [HRS] cells), produce copious amounts of T-helper-2 (Th2) cytokines (ref). A viral etiology has long been suspected but despite numerous studies, a plausible causative agent has not been identified for EBV-negative (mostly NSHL) disease (5-10). (HL positive for EBV DNA has a different risk pattern and has been linked to prior EBV infection, especially infectious mononucleosis) (11).

Unaffected MZ twins of patients have a ~14-fold greater risk of acquiring young adult HL than unaffected DZ twins of patients, suggesting multigenic susceptibility (12). Concordant twin pairs had primarily the EBV-negative NSHL subtype.

### Risk of HL in Twins of Patients: Heritability

MALIGNANT NEOPLASM		ZYGOSITY	NO. AT RISK	CASES EXPECTED†	CASES OBSERVED	SIR (95% CI)	SIR FOR MZ TWINS/ SIR FOR DZ TWINS
PROBAND	TWIN						
HD	HD	MZ	179	0.101	10	99 (48-182)	∞
HD	HD	DZ	187	0.100	0	—	∞
HD	HD‡	MZ	172	0.039	5	128 (42-299)	∞
HD	HD‡	DZ	181	0.035	0	—	∞
HD	Any other cancer	MZ	179	5.48	9	1.6 (0.8-3.1)	0.6
HD	Any other cancer	DZ	186	6.05	17	2.8 (1.6-4.5)	0.6
HD	Any other cancer‡	MZ	171	2.64	2	0.8 (0.1-2.7)	0.3
HD	Any other cancer‡	DZ	172	2.46	6	2.4 (0.9-5.3)	0.3
NHL	NHL	MZ	110	0.131	3	23 (4.7-67)	1.7
NHL	NHL	DZ	164	0.293	4	14 (3.8-35)	1.7

*T. Mack, et al., NEJM, 1995*

Up to now, most genetic studies of HL have concentrated on examination of HLA phenotypes as the putative genetic factors. HL is postulated to be an immune response to a virus or some other antigen that somehow becomes neoplastic. Because HLA class I and II proteins bind to and present antigen which then initiates the immune response, it makes sense that individuals with certain HLA phenotypes may be more susceptible. One might predict that HLA Class I variants would show a stronger association, since these proteins present antigen that has been processed from intracellular pathogens such as viruses. Instead, Klitz and colleagues found that the HLA class II region variants showed significant association with the

young adult nodular sclerosis HL (in the major focus of this study) (13). Other investigators identified other HLA class II haplotypes associated with increased risk (e.g. DRB1\*1501-DQA1\*0102-DQB1\*0602), along with a polymorphism of TAP1, an antigen processing gene located within the coding region of the HLA complex (14). One study suggested possibly modification by gender with a DPB1 (DPB1\*0301) allele associated with NSHL in females but not males (15). Diepstra and colleagues demonstrated a link between loci in the same HLA region and risk of EBV-negative (majority nodular sclerosis) Hodgkin lymphoma (16). The linkage disequilibrium relationships between these loci and specific etiologic inferences remain to be elucidated.

Candidate SNPs were also examined but have not been consistently replicated except those in interleukin 6 (IL6). We (Cozen et. al) were the first to show a link between higher IL6 levels and a functional single nucleotide polymorphism (SNP) and young adult Hodgkin lymphoma (17). The association was observed by Cordano et al (18), though not statistically significant, in young adult nodular sclerosis patients, and was recently demonstrated again in a case-control study in which all Hodgkin lymphoma was considered together (19, plus personal communication Goldin LR and Landgren O).

### **Goals / Specific Aims with planned publications**

1. To identify genetic risk factors for young adult NSHL by combining results from two separate but similar GWAS studies of young adult NSHL in a meta-analysis. (Publication # 1- for Nature Genetics)
  - a. To replicate findings from any other GWAS publications using our combined data (Publication # 2)
2. To further explore the results of the USC/UC GWAS meta-analysis using genotyping data from additional young adult NSHL cases and controls accrued in an ongoing study by PI James Cerhan, MD, PhD, from Mayo Clinic and the Iowa SEER Cancer Registry. (Publication # 3) No new genetic testing of the CCSS samples is necessary.
3. Examine associations in other subsets of HL including mixed cellularity (MC) and all Hodgkin lymphoma to evaluate heterogeneity of the results, power permitting. (Publication # 4 if possible)

Note that this is a purely genetic study; there is no environmental component.

### **Analytic Approach**

#### *Aim 1:*

- (1) Our proposal is to combined two case-control GWAS with a meta-analysis.

Two independent groups of adolescents and young adults with NSHL will be genotyped (Shell Table 1). One was analyzed at the University of Southern California (USC), and consisted of 251 NSHL patients of European descent (median age at diagnosis =28 years old, range 13-49 years old, except for one 7-year old) and 2,299 controls from the Cancer Genetic Markers and Susceptibility Project (CGEMS); all were genotyped using the Human610-Quad Illumina

platform with a total of 599,010 SNPs successfully genotyped in 251 NSHL patients (mean SNP call rate of > 99%; mean sample call rate among patients = 99.87%).

USC patients were ascertained twin registries maintained at USC in a case-sibling study design (NCI/ 1R03CA110836-01A2, PI: Cozen) and from California SEER registries in a case-parent trio design (USAMRMC (DOD) PR054600; Leukemia Lymphoma Society, 6137-07, PI of both studies: Cozen). We collected blood and/or saliva specimens from 584 young adult Hodgkin lymphoma patients, 828 siblings and parents of patients, and 107 spouses (of twin cases only). Both studies were originally designed to examine candidate SNPs but by the time the DNA samples were collected, the technology had advanced and GWAS was used instead. A GWAS was conducted on 380 patients of European descent from the combined sources. Of the total combined USC samples assayed, 37 (10%) were diagnosed as mixed cellularity HL, 12 (3%) as "classical HL, 65 (17%) as Hodgkin lymphoma, not otherwise specified, 42 (11%), and the remaining 255 were diagnosed with nodular sclerosis (NSHL) (67%). We limited the analysis to the NSHL subset since that was the most heritable, epidemiologically distinct, and numerous subset; and, based on previously studies, most likely to detectable genetic associations.

For the USC analysis, we used the PLINK software package to calculate missingness, allele frequency, and deviation from the Hardy-Weinberg Equilibrium (<http://pngu.mgh.harvard.edu/~purcell/plink/>) (20). Association of single markers with disease was assessed using unconditional logistic regression after adjusting for gender and the top 10 eigenvalues in the population stratification analysis (21). We assessed association of imputed genotypes with disease by an unconditional logistic regression after adjusting for gender and the top 10 principal components. Genotyped SNPs with minor allele frequencies (MAFs) of > 0.05 were compared in patients and controls and *P* values, odds ratios (ORs) and 95% confidence intervals (CIs) were calculated using unconditional logistic regression.

The second group, analyzed at the University of Chicago (UC, PI: K Onel), consisted of an independent set of 144 NSHL patients of European descent (median age at diagnosis =16) from the Children's Cancer Survival Study (CCSS) and 1,016 controls from the Genetic Association Information Network (GAIN) genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0 After exclusion of 339,905 SNPs with minor allele frequency <0.10, 17,035 SNPs with genotyping call rate < 0.95 and 31 SNPs that strongly deviated from Hardy-Weinberg equilibrium (HWE;  $P < 1.0 \times 10^{-5}$ ), 552,651 markers were investigated for association. The GWAS was performed as a case-case study within CCSS HL patients to investigate the contribution of genetic factors to the risk of solid tumor second malignant neoplasms (SMNs) in patients previously treated for HL (See attached draft, concept previously approved by CCSS). In the UC sample, both genotyped SNPs and SNPs imputed using the Mach software package with MAFs of > 0.05 were compared between patients and controls using unconditional logistic regression.

A meta-analysis will be conducted to obtain combined estimates using an inverse variance weighting of study-specific estimates. SNPs will be determined noteworthy if their *p*-value from the corresponding test of the combined estimate is genome-wide significant ( $p < 5 \times 10^{-8}$ ). Quantile-quantile (Q-Q) plots will be produced to evaluate whether the observed distribution of *P* values for SNPs passing quality control criteria for each group fit the null distribution. Heterogeneity in effect estimates between studies will be investigated using several statistics (H and I<sup>2</sup>) (22). In addition to study-specific adjustment for potential confounding by population substructure, principal components will be investigated on the combined sample to test whether the two sets of patients were sampled from genetically similar populations.

All cases from both study populations will be restricted to NSHL. Over 85% were in the adolescent/young adult age range, so we assume that the patients in both samples have the

same etiologic subtype (since this subtype is considered a single etiologic entity regardless of age of onset, within the “adolescent/young adult age range”). Nevertheless since there were some differences in age distribution between the two samples we will perform a sensitivity analysis limiting the cases from each sample to the same age range (10-20 years of age).

#### *SubAim 1a*

There are two other groups currently working on an HL GWAS. If one of these groups publishes prior to the completion of these analyses, we will compare the published SNPs to our combined top hits and the results will be presented as a replication study.

#### *Aim 2*

We plan to add a third set of young adult Hodgkin lymphoma cases and controls of European descent. Dr. James Cerhan, Chief of Epidemiology at Mayo Clinic and an international expert in the etiology of lymphoma, has been prospectively recruiting consecutive, newly diagnosed HL patients age 20 and older seen at Mayo Rochester and the University of Iowa as part of the Iowa/Mayo Lymphoma SPORE Molecular Epidemiology Resource. All pathology is centrally reviewed, and DNA and serum are banked for future studies. Controls consist of regional patients attending Mayo Clinic for a routine physical exam. This control group has been extensively used for genetic and risk factor studies and the MAF of the controls for over 6,000 SNPs are in line with other Caucasian, population-based controls (unpublished data). There are a total of 260 HL, 125 of these are NSHL diagnosed from 10-49 years of age, and 343 controls with DNA available for study. From Aim 1, we will take our top SNPs (the number to be decided by budget and significance level) and genotype these in the Mayo/Iowa cases and controls. Final significance will be determined by combining stage 1 (USC/UC-CCSS) with stage 2 (Mayo/Iowa cases and controls).

#### *Aim 3*

Although numbers are too small in the USC/UC-CCSS meta-analysis to examine risk of other distinct HL subtypes such as mixed cellularity or lymphocyte predominant, it may be possible to do so in combination with the Mayo/Iowa cases, or cases from the other international HL groups. The details of any such comparisons would have to be worked out with all collaborators.

### **Possible Limitations/Concerns**

#### **1) Would survival bias affect the results?**

The 5-year survival rate of nodular sclerosis Hodgkin lymphoma diagnosed between 10-49 years of age based on data from the 17 SEER registries from 1996-2007 is 92.2%. This very high survival rate (and probably cure rate) suggests that differential survival would probably not appreciably affect the results.

#### **2) Why did we limit Aim 1 to nodular sclerosis HL?**

Nodular sclerosis is considered to be a distinct etiological and pathological entity from other HL, and is the most heritable subtype. Because we are looking for a genetic risk pattern, we wanted

to have a definitive phenotype; otherwise the associations could be muted due to misclassification.

### **3) Would response rates affect the results?**

The response rates for the USC subjects varied by study. The response rate was high for the twins (~75%) but lower for the recently diagnosed population-based cases (68%). It is unlikely that response rate among persons of European descent would be correlated with genetic risk factors in a purely genetic study. If environmental risk factors were to be evaluated, then responses rates and response bias could affect the results, but we are not examining environmental factors in this study.

### **4) The subjects derive from three population sources: Twin Registries, SEER population-based cancer registries, and the CCSS, a cohort of long term pediatric cancer survivors. Will use of these disparate populations affect study results?**

The subjects are all nodular sclerosis HL patients of European descent but were ascertained in different ways. Some are long term survivors (CCSS and twins) and some more recently diagnosed (SEER-ascertained patients). For the type of genetic study proposed, replication of the results is important to validate the findings and ensure that they aren't false positives. In this context, replication in samples from somewhat disparate populations is helpful and will actually enhance the credibility of the findings.

### **5) Most GWAS's have been conducted with very large numbers (e.g. breast GWAS with 20,000+ cases). Will there be enough power in these small studies to be able to reliably detect an affect?**

We have a better chance of detecting an effect for this phenotype of young adult nodular sclerosis HL with small numbers because it is one of the most heritable cancers, so the effect size will be larger and easier to detect than for other common cancers. We (Mack, Cozen et al.) previously showed that the risk to an identical twin of a case was ~100 times that expected, and ~14 times more than that to the fraternal twin of a case, suggesting a very strong genetic effect. When heritability is strong, genetic risk factors (at least, the "low hanging fruit") can be identified using smaller numbers, as demonstrated in several studies of testis cancer. Ultimately, confirmation of results by combining study-specific estimates and using conservative genome-wide levels of significance are the key to valid conclusions in this area of research.

### **6) What about the use of publically available GWAS data from controls?**

Because this study addresses only genetic risk factors with no environmental (e.g. non-genetic components), confounding by population stratification can be controlled in the analysis through the use of ancestry informative markers and principal component analysis. The cases and controls in both the USC and UC/CCSS samples are of European descent so only minor adjustments will need to be made. Publically available GWAS data increases the efficiency and lowers the expense new assays need only be performed on cases.

## References

1. Cozen W, Katz J, Mack TM: Hodgkin's disease varies by cell type in Los Angeles. *Cancer Epidemiol Biomarkers Preven*, 1: 261-268, 1992.
2. Glaser SL, Gulley ML, Clarke CA, Keegan TH, Chang ET, Shema SJ, Craig FE, Digiuseppe JA, Dorfman RF, Mann RB, Anton-Culver H, Ambinder RF. Racial/ethnic variation in EBV-positive classical Hodgkin lymphoma in California populations. *Int J Cancer*;123:1499-507, 2008.
3. Mani H, Jaffe ES. Hodgkin lymphoma: An update on its biology with newer insights into classification. *Clin Lymphoma Myeloma*; 9:206-216, 2009.
4. Mueller NE, Grufferman S. Hodgkin's lymphoma. in *Cancer Epidemiology and Prevention* in Schottenfeld D and Fraumeni JF, Jr., Third Edition, Oxford University Press, Third Edition, New York, New York, 2006.
5. Armstrong A, Shield L, Gallagher A, Jarrett R. Lack of involvement of known oncogenic DNA viruses in Epstein-Barr virus-negative Hodgkin's disease. *Br J Cancer*. 1998; 77:1045-7.
6. Jarrett R. Risk factors for Hodgkin's lymphoma by EBV status and significance of detection of EBV genomes in serum of patients with EBV-associated Hodgkin's lymphoma. *Leuk Lymph*. 2003; 44:S27-S32.
7. Gallagher A, Perry J, Shield L, Freeland J, Mackenzie J, Jarrett R. Viruses and Hodgkin Disease: No evidence of novel Herpesviruses in non-EBV-associated lesions. *Int J Cancer*. 2002; 101:259-64.
8. Gallagher A, Perry J, Freeland J, et al. Hodgkin lymphoma and Epstein-Barr virus (EBV): No evidence to support hit-and-run mechanism in cases classified as non-EBV-associated. *Int J Cancer*. 2003; 104:624-30.
9. Wilson K, Gallagher A, Freeland J, Shield L, Jarrett R. Viruses and Hodgkin lymphoma: No evidence of polyomavirus genomes in tumor biopsies. *Leuk Lymph*. 2006; 47:1315-21.
10. Wilson K, Freeland J, Gallagher A, et al. Measles virus and classical Hodgkin lymphoma: No evidence for a direct association. *Int J Cancer*. 2007; 121:442-7.
11. Jarrett R. Viruses and leukemia/lymphoma. *J Pathol*. 2006; 208:176-86.
12. Mack TM, Cozen W, Shibata DK, Weiss LM, Nathwani BN, Hernandez AM, Taylor CR, Hamilton AS, Deapen DM, Rappaport EB: Concordance for Hodgkin's disease in identical twins suggests genetic susceptibility to the young-adult form of the disease. *N Eng J Med*, 332: 413-8, 1995.
13. Klitz W, Aldrich C, Fildes N, Horning S, Begovich A. Localization of predisposition to Hodgkin's disease in the HLA class II region. *Am J Hum Genet* 1994;54:497-505.
14. Harty LC, Lin AY, Goldstein AM, Jaffe ES, Carrington M, Tucker MA, and Modi WS. HLA-DR, HLA-DQ, and TAP genes in familial Hodgkin disease. *Blood*, 99: 690-693, 2002.
15. Staratschek-Jox A, Shugart YY, Strom SS, Nagler A, Taylor GM. Genetic susceptibility to Hodgkin's lymphoma and to secondary cancer: workshop report. *Ann Oncol*. 2002;13 Suppl 1:30-3.
16. A Diepstra, M Niens, E Vellenga, G W van Imhoff, I M Nolte, M Schaapveld, G van der Steege, A van den Berg, R E Kibbelaar, G J te Meerman, S Poppema . Association with HLA class I in Epstein-Barr-virus-positive and with HLA class III in Epstein-Barr-virus-negative Hodgkin's lymphoma. *Lancet*, 365, 2216-2224, 2005.

17. Cozen W, Gill PS, Ingles SA, Masood R, Martinez-Maza O, Cockburn MG, Gauderman J, Pike MC, Bernstein L, Nathwani BN, Salam MT, Danley K, Wang W, Gage J, Gundell-Miller S, Mack TM: IL-6 Levels and Genotype are Associated with Risk of Young Adult Hodgkin Lymphoma. *Blood*, 103: 3216-21, 2004.
18. Cordano P, Lake A, Shield L, Taylor GM, Alexander FE, Taylor PRA, White J, Jarrett RF. Effect of IL-6 promoter polymorphism on incidence and outcome in Hodgkin's lymphoma. *BJH*, 128: 493-495, 2005.
19. Liang XS, Caporaso N, McMaster ML, Ng D, Landgren O, Yeager M, Chanock S, Goldin LR. Common genetic variants in candidate genes and risk of familial lymphoid malignancies. *Br J Haematol*. 146:418-23, 2009.
20. Purcell, S., B. Neale, et al. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**: 559-575, 2007.
21. Price, AL, Patterson NJ, et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**: 904-909, 2006.
22. Higgins, JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* **21**: 1539-1558, 2002.

Dr. Yasui's Review (Statistics):

Since the CCSS Biospecimen Committee approved the proposal, I will limit my comments to statistical issues only.

- (1) Two case-control studies are proposed to be combined through a meta analysis method. From epidemiological points of view, cases and controls must come from a well-defined study base. As with the case with many GWAS studies, this principle is violated in this proposal. Under this violation, is there any special reason to compare the CCSS samples with U of Chicago controls and not with USC controls in a combined analysis? At least, I would like to know more about the controls against whom the CCSS samples will be compared: who they are, what the inclusion/exclusion criteria of the studies were that collected these control samples. Also, the USC samples' control description is unclear to me: I am confused with "twins", "siblings" and "parents"(?) of the cases as controls.
- (2) Need more specific details of the method for combining the two case-control results.
- (3) Power needs to be presented.
- (4) The pre-analysis processing and SNP analysis methods are standard.

Our Response:

The first study uses independent and non-relative HL cases subsampled from ongoing HL studies at USC. While these parent USC studies contain twins and siblings, we only include independent cases in the GWAS analysis. These cases were genotyped on the Illumina platform. To avoid any issues of comparison or imputation across platforms, we compared these cases to publicly available controls from CGEMs (<http://cgems.cancer.gov/>) that were previously included in genome-wide scans of prostate and breast cancer. Most importantly, these controls were genotyped using the Illumina platform and our population substructure analysis demonstrates that our cases and controls come from the same genetic population (i.e. all individuals are of a European ancestry).

The second study uses CCSS HL cases genotyped using the Affymetrix platform (since this effort is lead by investigators at the University of Chicago we have referred to this comparison as the UC analysis, albeit there are in fact no samples coming from the U of Chicago). These cases are then compared to samples from another publicly available dataset, the Genetic Association Information Network (<http://www.genome.gov/19518664>), also genotyped on the Affymetrix platform. As before with the USC-CGEMS comparison, both cases and controls for the CCSS-GAIN comparison have been genotyped on the same platform and are confirmed to have the same genetic ancestry.

While it is technically feasible to combine all cases and all controls in a single combined analysis, our previous experience in combining samples across multiple platforms indicates that this is fraught with difficulties (i.e. strand issues, differential missingness for imputation, etc.). Thus, we believe that a more conservative and robust analyses uses the study-specific estimates in a meta-analysis.

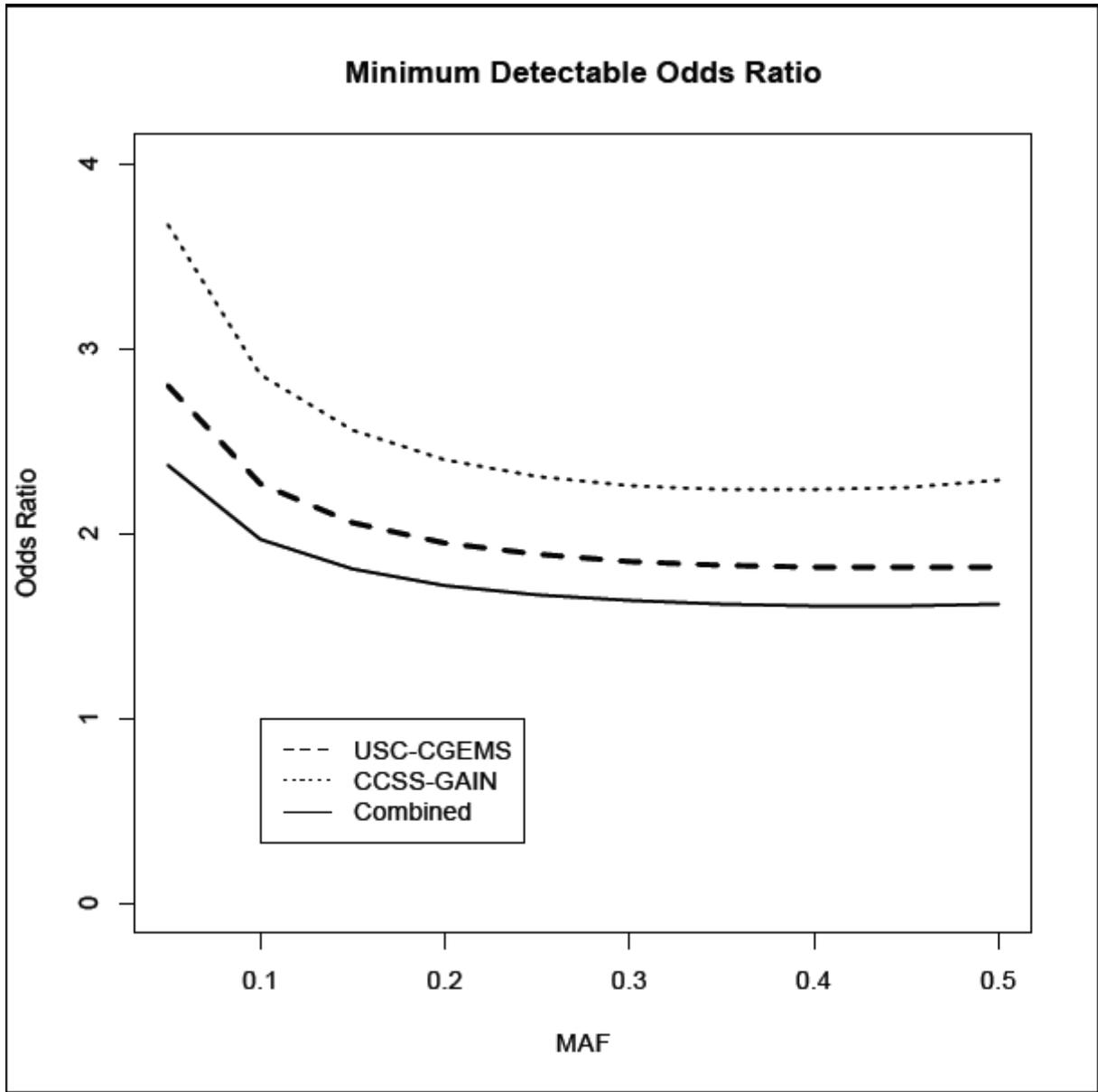
(2) For obtaining the combined results the most important aspect is to ensure SNP and allele consistency across platforms. To accomplish this we compare all results to HapMap genotypes and frequencies as well as making between study comparison of allele frequencies to identify any allele flips. Once this has been accomplished, the meta-analysis uses an inverse-variance weighting of the log odds ratio from the USC-CGEMS and CCSS-GAIN analyses. A single combined effect estimate is calculated as an average of the log odds ratio estimates, weighted by the inverse of the variance estimates from the two samples. This approach is computationally feasible and has been shown as an optimal weighting generating unbiased estimators with minimum variance (Hedges and Olkin 1985). In addition, we will investigate heterogeneity across study-specific estimates of effect using several statistics (H and I<sup>2</sup>) (see Higgins and Thompson 2002).

(3) We now include a power figure for each study and the combined analysis demonstrating the minimal detectable effect size at 80% power with an  $\alpha=5 \times 10^{-8}$  (see below). This is calculated using QUANTO (<http://hydra.usc.edu/gxe/>). For MAFs > 0.2, we have 80% power to detect an odds ratio of 1.70 or greater. While this minimum detectable odds ratio is slightly larger than many of those reported for complex diseases, we believe that we have a good opportunity for discovery since HL has larger sibling relative risks than other cancers (e.g. there is a 100 fold increase in risk for identical twins and a 7 fold increased risk for siblings). We are optimistic because young adult Hodgkin lymphoma is the most heritable cancer.

Higgins, JP, Thompson SG. Quantifying heterogeneity in a meta-analysis. *Stat Med* 21: 1539-1558, 2002.

Hedges L.V. and Olkin I. (1985). *Statistical methods for meta-analysis*. New York: Academic Press.

Power Calculations (see below)



## Proposed Tables and Graphs for Aim 1

**Figure 1a. Manhattan plot of the genome-wide association results for 395 cases and 3,245 controls.**

**Figure 1b.** Regional plot of regions reaching genome-wide significance for the combined genome-wide association results (directly genotyped SNPs in both patient groups and controls)

**Figure 1c.** Gene map of the regions reaching genome-wide significance in the combined analysis showing genes in LD with significant SNPs.

**Figure 1d.** LD structure of regions reaching genome-wide significance.

**Supplementary Figures 1a and 1b.** Quantile X Quantile plots for the USC (1a) and UC (1b) case samples.

**Supplementary Figures 2a and 2b.** Top 4 eigenvectors of the USC and UC sample in the population stratification analysis.

Table 1: SNPs associated with risk of nodular sclerosis Hodgkin lymphoma reaching genome-wide significance in a meta-analysis of two samples.

SNP	CHR	BP	USC				UC					
			Minor Allele	MAF (Cases)	MAF (Ctrls)	OR	P	MAF (Cases)	MAF (Ctrls)	OR	P	OR

Supplementary Table 1. Demographic and clinical information for nodular sclerosis Hodgkin lymphoma patients of European origin from two sites, University of Southern California (USC) and University of Chicago (UC).

USC	UC	Combined
n (%)	n (%)	n (%)

---

**Total patients**

**Male**

**Female**

**Median Age at Diagnosis**

**Age Range at Diagnosis**

**Age at Diagnosis in 10-year  
intervals**

**7-10**

**10-20**

**20-30**

**30-40**

**40-46**

**Mean Years of follow up**

---

Supplementary Table 2: Joint analysis of the two genome-wide significant genotyped SNPs in the USC and UC samples.

	USC		UC	
	Estimate	P-value	Estimate	P-value
<b>SNP 1</b>				
<b>SNP 2</b>				
<b>p for 2-df test</b>				