

# Host immune gene polymorphisms in combination with clinical and demographic factors predict late survival in diffuse large B-cell lymphoma patients in the pre-rituximab era

Thomas M. Habermann,<sup>1</sup> Sophia S. Wang,<sup>2</sup> Matthew J. Maurer,<sup>1</sup> Lindsay M. Morton,<sup>2</sup> Charles F. Lynch,<sup>3</sup> Stephen M. Ansell,<sup>1</sup> Patricia Hartge,<sup>2</sup> Richard K. Severson,<sup>4</sup> Nathaniel Rothman,<sup>2</sup> Scott Davis,<sup>5</sup> Susan M. Geyer,<sup>1</sup> Wendy Cozen,<sup>6</sup> Stephen J. Chanock,<sup>2</sup> and James R. Cerhan<sup>1</sup>

<sup>1</sup>Mayo Clinic College of Medicine, Rochester, MN; <sup>2</sup>National Cancer Institute, Bethesda, MD; <sup>3</sup>University of Iowa, Iowa City; <sup>4</sup>Wayne State University, Detroit, MI; <sup>5</sup>Fred Hutchinson Cancer Research Center, Seattle, WA; and <sup>6</sup>University of Southern California, Los Angeles

To evaluate the hypothesis that host germ line variation in immune genes is associated with overall survival in diffuse large B-cell lymphoma (DLBCL), we genotyped 73 single nucleotide polymorphisms (SNPs) from 44 candidate genes in 365 DLBCL patients diagnosed from 1998 to 2000. We estimated hazard ratios (HRs) and 95% confidence intervals (CIs) for the association of SNPs with survival after adjusting for clinical factors. During follow-up, 96 (26%) patients died, and the

median follow-up was 57 months for surviving patients. The observed survival of this cohort was consistent with population-based estimates conditioned on surviving 12 months. An *IL10* haplotype (global  $P = .03$ ) and SNPs in *IL8RB* (rs1126580;  $HR_{AG/GG} = 2.11$ ; CI, 1.28-3.50), *IL1A* (rs1800587;  $HR_{CT/TT} = 1.90$ ; CI, 1.26-2.87), *TNF* (rs1800629;  $HR_{AG/GG} = 1.44$ ; CI, 0.95-2.18), and *IL4R* (rs2107356;  $HR_{CC/CT} = 1.97$ ; CI, 1.01-3.83) were the strongest predictors of overall survival. A risk score that com-

bined the latter 4 SNPs with clinical factors was strongly associated with survival in a Cox model ( $P = 6.0 \times 10^{-11}$ ). Kaplan-Meier 5-year survival estimates for low, intermediate-low, intermediate-high, and high-risk patients were 94%, 79%, 60%, and 48%, respectively. These data support a role for germ line variation in immune genes, particularly genes associated with a proinflammatory state, as predictors of late survival in DLBCL. (Blood. 2008;112:2694-2702)

## Introduction

Diffuse large B-cell lymphoma (DLBCL) is the most commonly diagnosed subtype of non-Hodgkin lymphoma (NHL) in Western countries.<sup>1,2</sup> DLBCL is potentially curable, but the course of the disease is variable.<sup>3-5</sup> Established adverse prognostic factors in DLBCL include older age, higher stage, poor performance score, and above normal lactic dehydrogenase, and these factors have been validated as part of the International Prognostic Factor Index (IPI). However, the IPI predicts outcome incompletely. Molecular features of the tumor offer significant promise in providing additional information on prognosis.<sup>6,7</sup> For example, gene expression profiling of DLBCL has defined 2 major subgroups, one with a gene expression profile similar to normal germinal center B cells (60% 5-year survival) and the other mimicking activated peripheral B cells (30% 5-year survival).<sup>8-10</sup> Besides clinical and tumor molecular characteristics, there is a growing appreciation that the tumor microenvironment,<sup>7</sup> and more broadly the host genetic background,<sup>11</sup> may be an additional critical factor in cancer progression and outcome and therefore may be useful as a prognostic marker.<sup>12</sup>

Cytokines and related immune factors have been hypothesized to play an important role in lymphomagenesis,<sup>13</sup> as they appear to influence proliferation, differentiation, and movement of both tumor and stromal cells, regulate communication between tumor and stroma, and regulate tumor interactions with the extracellular matrix.<sup>14</sup> Immunologic function is in part influenced by host genetics, and germ line genetic variation in cytokine and related

immune genes have been associated with risk of developing DLBCL<sup>15-17</sup> and disease-free and overall survival after DLBCL diagnosis.<sup>18-20</sup> To test the hypothesis that inherited variation in cytokine and related immune genes impact DLBCL survival more comprehensively, we evaluated the role of 73 single nucleotide polymorphisms (SNPs) from 44 candidate immune genes (Table 1) and overall survival in DLBCL, using cases that were recruited as part of a population-based case-control study. We have previously reported the association of immune SNPs with risk of developing DLBCL in this study population.<sup>16</sup> Here we present the risks of dying from DLBCL according to those SNPs. By measuring the effects of the same markers on both etiology and survival, we examine a larger area of influence of genes on this immune system malignancy.

## Methods

### Study population

This study was reviewed and approved by human subjects review boards at the National Cancer Institute and each of the participating study centers, and written, informed consent was obtained from all participants in accordance with the Declaration of Helsinki. Methods for this molecular epidemiology study have been described previously.<sup>16</sup> Briefly, we enrolled 1321 patients (20-74 years of age) with newly diagnosed, histologically confirmed NHL in 4 Surveillance, Epidemiology, and End Results (SEER) cancer registries (Detroit, MI, metropolitan area; northwestern Washington

Submitted September 8, 2007; accepted June 8, 2008. Prepublished online as *Blood* First Edition paper, July 16, 2008; DOI 10.1182/blood-2007-09-111658.

The publication costs of this article were defrayed in part by page charge payment. Therefore, and solely to indicate this fact, this article is hereby marked "advertisement" in accordance with 18 USC section 1734.

The online version of this article contains a data supplement.

**Table 1. Candidate genes and SNPs, NCI-SEER NHL Survival Study**

Gene	Name	Location	SNP rs no.
<i>CARD15</i>	Caspase recruitment domain family, member 15	16q12	rs2066842,* rs2066844,* rs2066847
<i>CCR2</i>	Chemokine, CC motif, receptor 2	3p21	rs1799864
<i>CCR5</i>	Chemokine, CC motif, receptor 5	3p21	rs333
<i>CTLA4</i>	Cytotoxic T lymphocyte-associated 4	2q33	rs231775
<i>CSF3</i>	Colony stimulating factor 3 (granulocyte)	17q11.2-q12	rs25645*
<i>CX3CR1</i>	Chemokine, CXC motif	3p21	rs3732379*
<i>CXCL12</i>	Chemokine, CXC motif, ligand 12	10q11.1	rs1801157
<i>FCGR2A</i>	Receptor for Fc fragment of IgG, low affinity IIa (CD32)	1q21-q23	rs1801274
<i>ICAM1</i>	Intercellular adhesion molecule 1 (CD54)	19p13.3-p13.2	rs5491
<i>INFA1</i>	Interferon $\alpha$ 1	9p22	rs1758566*
<i>IFNG</i>	Interferon $\gamma$	21q14	rs1861494, rs2069705
<i>IFNGR1</i>	Interferon $\gamma$ receptor 1	6q23-q24	rs3799488*
<i>IFNGR2</i>	Interferon $\gamma$ receptor 2	21q22.11	rs9808753
<i>IL1A</i>	Interleukin 1 $\alpha$	2q13	rs17561, rs1800587
<i>IL1B</i>	Interleukin 1 $\beta$	2q14	rs16944, rs1143627, rs1143634
<i>IL1RN</i>	Interleukin 1 receptor antagonist	2q14.2	rs454078
<i>IL2</i>	Interleukin 2	4q26-q27	rs2069762
<i>IL3</i>	Interleukin 3	5q31.1	rs40401*
<i>IL4</i>	Interleukin 4	5q31.1	rs2070874, rs2243248, rs2243250
<i>IL4R</i>	Interleukin 4 receptor	16p12.1-p11.2	rs2107356
<i>IL5</i>	Interleukin 5	5q31.1	rs2069807, rs2069812, rs2069818*
<i>IL6</i>	Interleukin 6	7p15.3	rs1800795, rs1800797
<i>IL7R</i>	Interleukin 7 receptor (CD127)	5p13	rs1494555*
<i>IL8</i>	Interleukin 8	4q12-q13	rs4073, rs2227307, rs2227306
<i>IL8RB</i>	Interleukin 8 receptor $\beta$	2q35	rs1126579, rs1126580, rs2230054*
<i>IL9</i>	Interleukin 9	5q31.1	rs1799962*
<i>IL9R</i>	Interleukin 9 receptor	Xq28 or Yq12	rs6522*
<i>IL10</i>	Interleukin 10	1q31-q32	rs1800871, rs1800872, rs1800896, rs1800890
<i>IL10RA</i>	Interleukin 10 receptor $\alpha$	11q23.3	rs9610
<i>IL12A</i>	Interleukin 12 $\alpha$	3q25.33	rs568408
<i>IL12B</i>	Interleukin 12B (natural killer cell stimulatory factor 2, cytotoxic lymphocyte maturation factor 2, p40)	5q33.3	rs3212227
<i>IL13</i>	Interleukin 13	5q23.3	rs20541, rs1800925
<i>IL15</i>	Interleukin 15	4q31.21	rs10833
<i>IL15RA</i>	Interleukin 15 receptor $\alpha$	10p15.1	rs2296135
<i>IL16</i>	Interleukin 16	15q25.1	rs859, rs11325
<i>JAK3</i>	Janus kinase 3	19p13.1	rs3008,* rs3212713*
<i>LTA</i>	Lymphotoxin- $\alpha$	6p21.3	rs909253, rs2239704
<i>MBL2</i>	Mannose-binding lectin (protein C) 2, soluble	10q11.2-q21	rs7095891,* rs11595876,* rs10824792,* rs2083771*
<i>SELE</i>	Selectin E	1q22-q25	rs5361*
<i>STAT1</i>	Signal transducer and activator of transcription 1	2q32.2-q32.3	rs2066804*
<i>TLR4</i>	Toll-like receptor 4	9q32-q33	rs4986790
<i>TNF</i>	Tumor necrosis factor	6p21.3	rs1800629, rs361525, rs1799724, rs1800630
<i>TNFRSF10A</i>	Tumor necrosis factor receptor superfamily, 10a	8p21	rs20577*
<i>VCAM1</i>	Vascular cell adhesion molecule 1	1p32-p31	rs1041163, rs3176879

\*Genotyped in blood-based samples only.

state; the state of Iowa; and Los Angeles County, CA) from July 1998 through June 2000. Known HIV-positive cases were excluded. All participants completed an in-person interview, and 1172 (89%) provided either a venous blood sample ( $n = 773$ ) or mouthwash buccal cell sample ( $n = 399$ ). This analysis is restricted to the 365 cases of DLBCL in the case group, as defined by SEER coding and using the InterLymph classification system,<sup>21</sup> who had a DNA sample available for genotyping. Of the eligible DLBCL cases for this study, 21% died before we could contact them, 9% had a physician refusal or could not be located, 17% were approached but refused participation (many were too ill), and 53% participated.

### Genotyping

The strategy for candidate gene selection focused on genes involved in key immune pathways, particularly those related to cytokine regulation and function. Priority in gene selection was given to those genes with data suggesting functional and biologic significance, an association with NHL etiology or prognosis, an association with other immune diseases, and a

minor allele frequency (MAF) of more than 5% in the white population. Full details on candidate gene selection have been previously published.<sup>16</sup>

Details on DNA extraction and genotyping have been published previously.<sup>16</sup> All genotyping was conducted at the National Cancer Institute Core Genotyping Facility using the Taqman or EPOCH platforms (<http://cgf.nci.nih.gov>), and sequence data and assay conditions are provided at [http://snp500cancer.nci.nih.gov/home\\_1.cfm?CFID=2106952&cftoken=53492492](http://snp500cancer.nci.nih.gov/home_1.cfm?CFID=2106952&cftoken=53492492).<sup>22</sup> Genotyping was conducted first on blood-based DNA samples ( $n = 215$  DLBCL patients) and then was expanded to patients with only buccal samples ( $n = 150$  DLBCL patients) for 52 of the 73 candidate SNPs. The decision to genotype buccal samples was based on a risk association in the subset of participants with blood-based DNA samples in the parent case-control study.<sup>16</sup> For the 21 SNPs that we genotyped on DLBCL patients with blood samples ( $N = 215$ ), the call rates ranged from 91.2% to 100% (median call rate, 93.0%); for the 52 SNPs that we genotyped on DLBCL patients with blood or buccal ( $n = 365$ ), the call rates ranged from 89.6% to 99.5% (median, 95.9%). Although we evaluated all SNPs

**Table 2. SNPs most strongly associated with DLBCL survival, ordered by *P* value from the codominant model**

SNP	Gene	<i>P</i> *	Deleterious genotype(s)†	Patients with deleterious genotype, %	HR‡	95% CI
rs2069807	<i>IL5</i> <sup>d</sup>	<.001	CT/TT	2.9	4.56	1.98-10.5
rs1126580	<i>IL8RB</i>	.002	AG/GG	68.6	2.11	1.28-3.50
rs1800587	<i>IL1A</i>	.004	CT/TT	49.2	1.90	1.26-2.87
rs17561	<i>IL1A</i>	.03	GT/TT	46.6	1.56	1.04-2.35
rs3176879	<i>VCAM1</i> §	.03	AG/GG	6.9	2.10	1.04-4.22
rs1800795	<i>IL6</i>	.03	GG/CG	86.1	1.57	0.84-2.94
rs1800797	<i>IL6</i>	.04	GG	40.6	1.47	0.99-2.21
rs5491	<i>ICAM1</i> §	.04	AT/TT	2.3	2.80	1.01-7.72
rs25645	<i>CSF3</i>	.04	AA	14.7	2.37	1.34-4.21
rs1800896	<i>IL10</i>	.08	AG/GG	71.0	1.48	0.91-2.38
rs1143634	<i>IL1B</i>	.09	CT/TT	33.8	1.31	0.86-1.98
rs1800629	<i>TNF</i>	.09	AG/AA	33.1	1.44	0.95-2.18
rs1801157	<i>CXCL12</i>	.09	GG	64.8	1.62	1.02-2.58
rs1041163	<i>VCAM1</i>	.10	TT	69.4	1.35	0.88-2.08
rs5361	<i>SELE</i>	.12	AC/CC	17.0	1.41	0.74-2.66
rs2107356	<i>IL4R</i>	.14	CC/CT	82.4	1.97	1.01-3.83
rs1799724	<i>TNF</i>	.15	CC	83.4	1.35	0.78-2.34

\*Observed *P* value from the trend test (codominant model) from Table S1.

†SNPs recoded as 0 for low-risk genotype(s) and 1 for deleterious genotype(s) based on results from Table S1. Missing genotype data were included with the low-risk category.

‡Hazard ratio (HR) adjusted for age and clinical and demographic factors. The HR is based on assigning missing SNP data to the reference group (low-risk genotype).

§SNPs with a MAF less than 0.05 (not eligible for the multi-SNP risk score).

||SNPs that were only genotyped in patients with a blood-based DNA sample (*N* = 215) were not eligible for the multi-SNP risk score.

with survival (Table S1, available on the *Blood* website; see the Supplemental Materials link at the top of the online article), only those 52 SNPs that were also genotyped in patients with buccal samples were eligible for multi-SNP models.

Positive and negative controls and 140 replicate samples were interspersed for all genotyping assays and blinded from the laboratory. Agreement for quality control replicates and duplicates was more than 99% for all assays. Only 1 SNP (rs1801157) in black controls was not in Hardy-Weinberg equilibrium (*P* < .01); reviewing of all genotyping data (including quality control samples) confirmed the accuracy of this assay.

### Prognosis study

Full details on the prognosis study using cases from the case-control study have been previously published.<sup>12</sup> Briefly, age, sex, race, Hispanic ethnicity, and education level were derived from patient interviews as part of the case-control study. Date of diagnosis, histology, stage, presence of B symptoms, first course of therapy, date of last follow-up, and vital status were derived from linkage to individual SEER registry databases in early 2005. Data on first course of therapy include use of single-agent or multiagent chemotherapy, radiation, and other therapies exclusive of chemotherapy and/or radiation. Individual agents and doses, as well as indications for use of nonstandard therapy, were not available. The SEER registries collect date and cause of death but do not collect data on treatment response or disease recurrence or progression.

### Data analysis

**Evaluation of single SNPs and haplotypes.** Our overall data analysis approach has been previously published.<sup>12</sup> We first used Cox proportional hazards regression<sup>23</sup> to estimate hazard ratios (HRs) and 95% confidence intervals (95% CIs) for the association of each individual genotype with overall survival, adjusting for age and clinical and demographic factors (Table S1). The primary test of association for each SNP with survival used a codominant coding of the alleles (ie, 0, 1, and 2 variant alleles). This statistic was chosen because it has good power to detect genotype associations under a range of genetic models.<sup>24</sup> Age was modeled according to the standard IPI score as less than 60 versus 60 or more years.<sup>25</sup> Clinical and demographic factors were each modeled as 2 separate risk scores, analogous to a propensity score for logistic regression.<sup>26</sup> The clinical risk score was a linear combination of stage (local, regional, distant, missing),

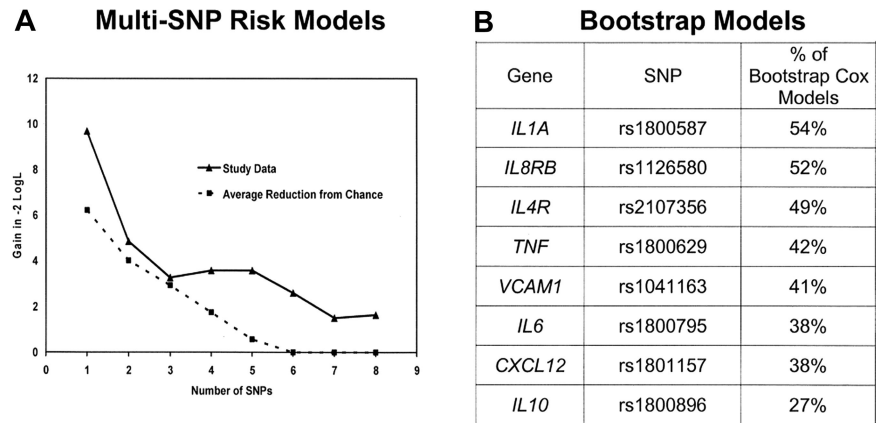
presence of B symptoms (no, yes, missing), and type of initial therapy (chemotherapy + radiation, chemotherapy + other therapy, radiation only, all other, or missing therapy). The demographic risk score was a linear combination of sex, race (white, all other), study center (Detroit, Iowa, Los Angeles, Seattle), and years of education (< 12, 12-15, 16+ years). For multi-SNP models, the demographic and clinical risk scores were combined into a single score, with values of 0 to 2 (low to high risk) as previously described.<sup>12</sup>

We were able to construct haplotypes for 4 genes: *TNF/LTA*, *IL8*, *IL8RB*, and *IL10*. Haplotype frequencies for selected genes were estimated by an expectation-maximization algorithm,<sup>27</sup> and the posterior probabilities for each haplotype were included in a Cox model to assess the association with survival.

To address concerns about multiple testing, we computed the tail strength of all 73 SNPs initially evaluated for their role in survival (Table S1). This measure<sup>28</sup> is closely related to the false discovery rate<sup>29</sup> and assesses the relative strength of the collection of observed *P* values from an analysis of a large number of markers.

**Selection of the best multi-SNP risk score.** Because many of these genes have overlapping functions and are part of complex networks, it is useful to identify a parsimonious multivariable prediction model. To achieve this, we first brought forward 17 SNPs (from 13 genes) with a *P* less than or equal to .15 based on the recoded results in Table 2. To preserve power, we eliminated from further consideration for multivariable modeling the 5 SNPs with an MAF less than 0.05 or more than 10% missing data (missing data resulted mainly from buccal samples not being genotyped in the parent study). After removing these SNPs, remaining SNPs with missing genotype were then assigned to the low-risk genotype. For the remaining SNPs in high linkage disequilibrium, we selected a single SNP based on the highest MAF and the strongest HR. After this data reduction process, there were 8 SNPs from 8 genes. To evaluate each of the 255 possible combinations of these 8 SNPs, we created an SNP score variable by summing the number of deleterious genotypes (as categorized in Table 2) in each particular combination.<sup>12,30</sup> Each multi-SNP score variable was fit in a Cox model adjusting for age and clinical and demographic risk scores. The models were grouped by number of SNPs included in the SNP score variable (ie, 1 SNP, 2 SNPs, . . . 8 SNPs) and were ranked by likelihood. A comparison of the likelihood for the best 1 SNP, 2 SNP, . . . 8 SNP models according to number of SNPs included in the model is shown in Figure 1A. In parallel, we ran 1000 bootstrap stepwise

**Figure 1. Selection of the multi-SNP model.** (A) Reduction in the  $-2 \log$  likelihood comparing the best 1 SNP, 2 SNP, . . . 8 SNP model from the study data (—) as well as expected reduction resulting from chance (---). (B) Ranking of the percentage of 1000 stepwise bootstrap Cox models that included each SNP. The top 4 SNPs in the bootstrap model are the same 4 SNPs in the best 4 SNP risk model.



selection Cox models (adjusting for age and clinical and demographic risk scores) using the 8 SNPs and calculated the percentage of models that included each SNP (Figure 1B). The plot of the likelihood suggested a model with 5 or fewer SNPs, as the reduction in likelihood was marginal after 5 SNPs (Figure 1A). The 4 highest ranked SNPs from the bootstrap analyses coincided with the most prognostic 4-SNP risk model; we therefore proceeded with a 4-SNP risk score as the final model.

**Evaluation of the multi-SNP risk score.** We assessed the association of the 4-SNP risk score with overall survival using Kaplan-Meier curves, Cox proportional hazards models, and time-dependent receiver-operator characteristic (ROC) curves for censored data.<sup>31</sup> We also developed a single risk score that combined the number of deleterious genotypes from the 4 SNPs (0-4) plus a combined clinical and demographic risk score (0-2) to create a single SNP and clinical risk score (0-6). To compare the significance of the SNP and clinical and demographic risk score observed in our data to what we would expect from chance, we performed a permutation analysis. The test statistic observed in our study data was then compared with the distribution of test statistics from 200 permutation iterations.

## Results

### Descriptive results

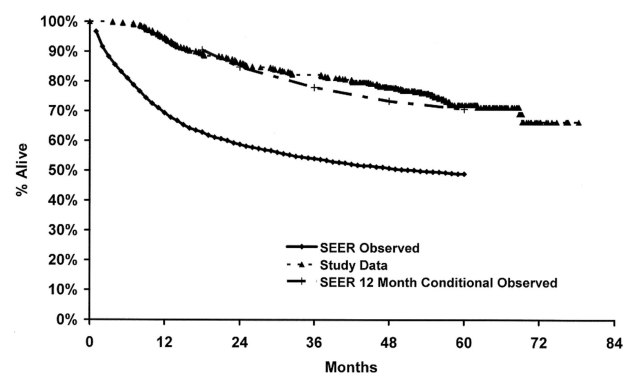
The median age at diagnosis of the 365 cases was 57 years (range, 20-74 years; 41% were age 60 or older). A majority of the patients (88%) were white, and 56% were male. Clinically, 41% had advanced stage disease and 27% had B symptoms. Based on cancer registry data, the most common initial therapy was a chemotherapy-based regimen (88%). During follow-up, 96 (26%) of the patients died, and 68% of the underlying causes of death were coded on the death certificate as lymphoma. The median follow-up of living patients was 57 months (range, 27-78 months). The age (age 60 years or older; HR = 1.80; 95% CI, 1.22-2.76), demographic (combination of sex, race, study center, and education; HR = 2.26; 95% CI, 1.36-3.76), and clinical (combination of stage, B symptoms, and type of treatment; HR = 2.66; 95% CI, 1.63-4.33) risk scores were associated with overall survival when included in the same Cox model.

In the parent case-control study, only 53% of the eligible cases were enrolled in the study, and we did not have genotype and survival data on the nonparticipants. Therefore, to address the potential impact of nonresponse on our results, we compared our observed survival to survival reported in the SEER program from the same registries as our cases (ie, Detroit, Iowa, Los Angeles, and Seattle) for white DLBCL patients 20 to 74 years of age and diagnosed from 1995 to 2000.<sup>32</sup> As shown in Figure 2, our observed survival was higher than that observed in the SEER data for

DLBCL from the same time frame of this study but was very similar to SEER for 12-month conditional survival (ie, survival given that a patient survives 12 months). This is consistent with the enrollment pattern of these patients into our case-control study, whereby patients with early mortality were less likely to be enrolled into the study.

### Single SNP results

We identified 17 SNPs from 14 genes of potential interest based on our statistical criteria ( $P$  trend  $\leq .15$ , Table S1). These SNPs were rescored so that all HRs were more than 1, and these are reported in Table 2. Most of the HRs were modest and in the range of 1.4 to 2.5. The smallest observed  $P$  value ( $< .001$ ) was for an *IL5* SNP (rs2069807, C-1551T, HR<sub>CT/TT</sub> = 4.56; 95% CI, 1.98-10.5), which was relatively rare (only 2.9% of patients carried a variant allele). The next smallest observed  $P$  value (.002) was for an *IL8RB* SNP (rs1126580; HR<sub>AG/GG</sub> = 2.11; 95% CI, 1.28-3.50). The only other  $P$  less than or equal to .01 was for an *IL1A* SNP (rs1800587; HR<sub>CT/TT</sub> = 1.90; 95% CI, 1.26-2.87). The tail strength of our set of 73 immune SNPs was 0.20 (95% CI,  $-0.03$ -0.43). A positive tail strength indicated that the observed  $P$  values were more significant than what would be expected resulting from chance; the tail strength of 0.20 in our study suggests that this set of SNPs displayed approximately 20% more signal than expected if all markers were null.



**Figure 2. Comparison of study data to SEER DLBCL survival.** SEER data (observed and conditioned on surviving 12 months) are based on whites from Detroit, Iowa, Los Angeles, and Seattle sites, 20 to 74 years of age at diagnosis, and diagnosis dates 1995 to 2000.<sup>32</sup>

**Table 3. Haplotype analyses**

Genes	SNPs in haplotype				Haplotype, %	HR*	95% CI	P†	
<b>TNF/LTA</b>	rs1800629 TNF G-308A	rs909253 LTA A252G							
	1	G	A		62	1.00	reference		
	2	A	A		< 1				
	3	G	G		19	1.05	(0.74, 1.50)	.8	
4	A	G		19	1.36	(0.96, 1.94)	.09		
<b>IL10</b>	rs1800871 C-819T	rs1800872 C-592A	rs1800896 A-1082G	rs1800890 T-3575A					
	1		A	T	55	1.00	reference		
	2		G	A	38	1.21	(0.89, 1.65)	.22	
	3		G	T	7	1.82	(1.08, 3.07)	.02	
4		A	A	< 1					
<b>IL10</b>	rs1800871 C-819T	rs1800872 C-592A	rs1800896 A-1082G	rs1800890 T-3575A					
	1	C	C	A	T	28	1.00	reference	
	2	C	C	G	A	39	1.42	(0.94, 2.14)	.10
	3	T	A	A	T	26	1.59	(1.01, 2.50)	.04
4	C	C	G	T	7	2.44	(1.34, 4.44)	.003	

\*Hazard ratio (HR) adjusted for age and clinical and demographic factors.

†Global *P* values are *P* = .23 for the *TNF/LTA* haplotype, *P* = .07 for the *IL10* 2-SNP haplotype, and *P* = .03 for the *IL10* 4-SNP haplotype. Global *P* value is testing the association of all observed haplotypes on a gene with survival.

### Haplotype results

There was no significant association with the risk of DLBCL (global *P* = .4) for the *TNF/LTA* haplotype constructed from 3 *TNF* SNPs (rs1799724, rs1800629, and rs361525) and 2 *LTA* SNPs, rs2239704 and rs909253 (Table S2), which has been previously reported to be associated with risk of developing DLBCL in this study population.<sup>16</sup> However, using a haplotype limited to *TNF* G-308A (rs1800629) and *LTA* A252G (rs909253) reported to be associated with risk of DLBCL by the InterLymph Consortium,<sup>15</sup> the AG versus GA haplotype was associated with a marginally significant higher risk of death (HR = 1.36; 95% CI, 0.96-1.94; Table 3). When the number of adverse alleles from *TNF* G-308A (A allele) and *LTA* A252G (G allele) were summed according to the approach of Warzocha et al,<sup>18</sup> patients with 2 to 4 adverse alleles (38% of patients) had poorer survival compared with patients with 0 or 1 allele (HR = 1.27; 95% CI, 0.84-1.90), and this association was stronger for patients with 3 or 4 adverse alleles (10% of the patients) compared with patients with 0 to 2 alleles (HR = 1.72; 95% CI, 0.95-3.09).

An *IL10* haplotype based on *IL10* A-1082G (rs1800896) and *IL10* T-3575A (rs1800890) showed a suggestive association with survival (global *P* = .07), and the GT haplotype was associated with poorer survival compared with the most common (AT) haplotype (HR = 1.82; 95% CI, 1.08-3.07; Table 3). Inclusion of 2 additional SNPs to the haplotype was even more strongly associated with survival (global *P* = .03), and 3 of the most common haplotypes were associated with poorer survival (Table 3). There were no associations of haplotypes in *IL8* or *IL8RB* with survival (Table S2).

### Multi-SNP risk score

As outlined in "Evaluation of the multi-SNP risk score," we selected a 4-SNP risk score for further evaluation, which included polymorphisms in *IL1A* (rs1800587), *IL8RB* (rs1126580), *IL4R* (rs2107356), and *TNF* (rs1800629). The number of deleterious genotypes was summed from these 4 SNPs (0-4), and this score was strongly associated with survival in both univariate (*P* =  $2.8 \times 10^{-5}$ ) and multivariate (*P* =  $3.7 \times 10^{-6}$ ) analyses (Figure 3, Table 4). Patients with 4 deleterious genotypes were more than 6 times more likely to die compare with patients with zero deleterious genotypes (95% CI, 3.05-15.0), and there was a gradient in risk with the number of deleterious SNP genotypes. Both the 4 SNP risk score and *IL10* haplotype remained statistically significant when they were included in the same model along with the clinical and demographic variables (data not shown).

We next combined the number of deleterious genotypes (0-4) with the clinical and demographic risk score (0-2). This combined score was strongly associated with survival (*P* =  $6 \times 10^{-11}$ ), and patients with a score of 5 or 6 were more than 9 times more likely to die compared with those with a low (0-2) risk score (95% CI, 4.22-21.4; Figure 4, Table 5).

To further evaluate the predictive ability of our model, we conducted a time-dependent ROC analysis for censored data.<sup>31</sup> This analysis uses sensitivity and specificity, both of which are time-dependent, to measure the prognostic capacity of the survival model as measured by the area under the curve (AUC). As shown in Figure 5, our clinical and demographic risk score (0-2) compared favorably in a time-dependent ROC results to the IPI from a previously published series.<sup>8</sup> The time-dependent ROC analysis for

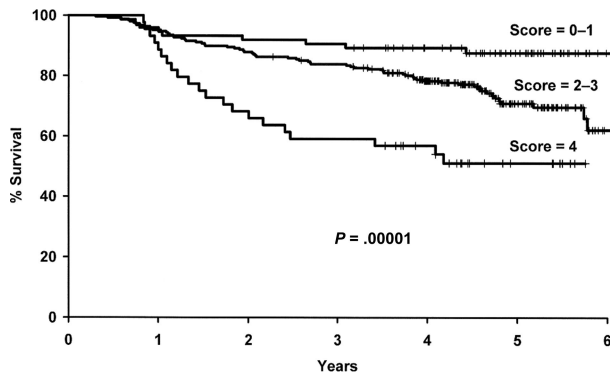
**Table 4. Univariate and multivariate Cox models for the 4 SNP risk score**

No. of deleterious genotypes*	n	Percent dead	Univariate model		Multivariate model†	
			HR	95% CI	HR	95% CI
0,1	74	12.2	1.00	(reference)	1.00	(reference)
2,3	247	26.7	2.47	1.23-4.96	2.76	1.37-5.56
4	44	47.7	5.63	2.57-12.3	6.77	3.05-15.0
Continuous			1.56	1.26-1.94	1.65	1.32-2.05

*P* values for trend test are *P* =  $2.8 \times 10^{-5}$  and *P* =  $3.7 \times 10^{-6}$  for univariate and multivariate models, respectively.

\*Deleterious genotypes from *IL1A* (rs1800587), *IL8RB* (rs1126580), *IL4R* (rs2107356), and *TNF* (rs1800629), as coded in Table 2.

†Adjusted for age and clinical and demographic factors.



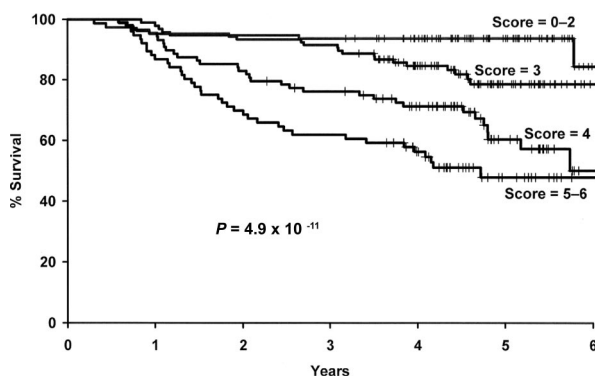
**Figure 3. Results for the 4 SNP risk score.** Kaplan-Meier curves by the number of deleterious genotypes from the 4 SNP risk score based on *IL1A* (rs1800587), *IL8RB* (rs1126580), *IL4R* (rs2107356), and *TNF* (rs1800629).

the 4 SNP risk score showed a lower ability to predict outcome; but when it was combined with the clinical and demographic risk score, the AUC at 2 years was more than 0.70, and at 5 years the AUC was 0.75 (95% CI, 0.68-0.80). Figure 5 also includes the predictive ability of the germinal center phenotype versus all others from a previously published DLBCL survival dataset,<sup>8</sup> and this characteristic predicted outcome at the same level as our 4 SNP risk score. Furthermore, when the germinal center phenotype was combined with the IPI, it predicted at the same level as our SNP plus clinical and demographic risk score.

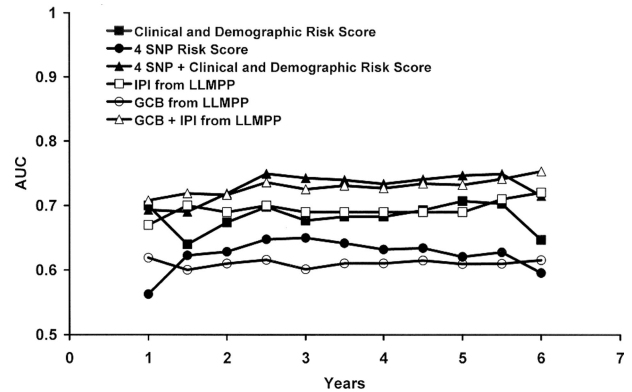
To assess the robustness of our multi-SNP risk score, we repeated our model-building strategy (starting with selection of SNPs with a  $P \leq .15$  through the final multi-SNP risk model) with the datasets generated in a permutation analysis. Our observed results were more significant than 82% of the results from randomly generated datasets, which suggests that our multi-SNP risk score has some degree of significance given the intense model building approach performed. In addition, the likelihood plot (Figure 1A) of the best 1 SNP, 2 SNP, . . . 8 SNP models suggests that as many as 5 SNPs may add information to predicting survival beyond what would be expected resulting from chance, although we opted for 4 SNPs based on the combined results that included the bootstrap modeling in selecting the most robust and parsimonious model.

**Sensitivity analyses**

All results were also similar when we excluded all nonwhites from the analysis (data not shown).



**Figure 4. Results for the combined SNP and clinical and demographic risk score.** Kaplan-Meier curves by level of the combined 4 SNP and clinical and demographic risk score.



**Figure 5. Time-dependent ROC analysis.** Time-dependent receiver-operator (ROC) analysis using the NCI-SEER dataset (clinical and demographic risk score, 4 SNP risk score, combined 4 SNP and clinical and demographic risk score) and using the Lymphoma/Leukemia Molecular Profiling Project (LLMPP) dataset<sup>8</sup> (International Prognostic Index [IPI], Germinal-center B cell-like [GCB], and GCB + IPI).

We fit the final 4-SNP plus clinical and demographic risk score model based on deaths resulting from lymphoma coded on the death certificate ( $n = 65$  of the 96 deaths; other deaths censored) and found the HR for the continuous score increased slightly from 1.9 (Table 5) to 2.0 (95% CI, 1.6-2.5).

Another potential concern is that not all patients received standard of care. Although we did not have sufficient data to fully evaluate treatment decisions and standard of care for these patients, we were able to exclude patients who did not receive multiagent chemotherapy, the presumed standard of care at the time of enrollment (rituximab was unlikely to have been used in the community setting from 1998 to 2000). After excluding these patients ( $n = 73$ ), results from Tables 4 and 5 were essentially unchanged (data not shown). Finally, there was no survival difference by enrollment year ( $P = .42$ ) and no impact on the final results in Table 5 after adjusting for enrollment year (data not shown).

**Discussion**

Using a population-based sample of 365 DLBCL cases diagnosed from 1998 to 2000 and followed through early 2005, we identified 17 SNPs from 14 cytokine and related immune regulation genes and a haplotype from the *IL10* gene that were all associated with overall survival from DLBCL independent of clinical and demographic factors. Furthermore, we observed a strong effect from the combination of 4 SNP markers: 3 SNP markers from 3 genes, namely, *IL1A1*, *IL8RB*, and *IL4R*, that had not been previously reported in NHL prognosis, and 1 SNP marker in *TNF*, which has

**Table 5. Cox model for the combined SNP and clinical and demographic risk score**

Risk score*	n	Percent dead	HR	95% CI
0-2	95	7.3	1.00	(reference)
3	106	18.9	2.76	1.17-6.53
4	88	36.4	5.91	2.61-13.4
5,6	76	48.7	9.50	4.22-21.4
Continuous			1.86	1.56-2.22

\*P value for trend test is  $6 \times 10^{-11}$ .

\*Number of hazardous genotypes (0-4) plus the clinical and demographic risk score (0-2).

been shown to have a deleterious effect on survival. The preliminary results are particularly encouraging when combined into a common carrier model that linearly summed the number of deleterious genotypes. Combining the SNP risk score with demographic and clinical factors increased the predictive ability of the model, with AUCs of more than 0.70 after 24 months of follow-up in the time-dependent ROC analysis, which is approaching the predictive range needed for clinically useful tests. Compared with gene expression profiling data from biopsy samples,<sup>8</sup> host genetics demonstrated a similar prognostic ability. Overall, these results support the importance of germ line variation in immune genes as predictors of DLBCL prognosis. Although steps were taken to guard against overfitting the data, including a permutation analysis of our entire model building strategy to assess the significance of our results compared with chance, these results clearly require replication in independent populations. In addition, it will be important to evaluate a population of DLBCL patients treated with rituximab in combination with CHOP chemotherapy and evaluate these findings in conjunction with molecular subtypes of DLBCL.

Even with rapid reporting by population-based cancer registries, we were able to enroll only approximately 50% of eligible DLBCL patients; therefore, we systematically missed those patients with the most aggressive disease leading to early mortality. Indeed, the observed survival of our patient cohort was much better than SEER population-based estimates for DLBCL patients overall but was quite consistent with survival estimates that were conditioned on DLBCL patients who survived 12 months after diagnosis (Figure 2). Therefore, our results will not apply to early mortality, and this will need to be addressed in future studies. A majority of deaths during the first year after diagnosis resulted from disease; and in the SEER data from Figure 2, approximately 58% of the deaths that occurred in the first 5 years after diagnosis occurred during the first year, and 75% of those deaths resulted from disease. Although our data are not informative for early mortality, our data will be robust to patients who survive to 12 months after their diagnosis. Furthermore, in sensitivity analyses, we found that our results held for patients who died of their disease (68%).

A proinflammatory state may both contribute to lymphomagenesis and lead to overall poorer survival. Of the genes in our final multi-SNP model, *TNF* has been most extensively studied as a prognostic factor in DLBCL. Higher levels of TNF have been associated with poorer outcome in NHL.<sup>33</sup> Warzocha et al<sup>18</sup> reported that an extended haplotype in *TNF* (G allele) and *LTA* (A allele) was associated with higher TNF production, and DLBCL patients (n = 126) with 2 to 4 high-risk alleles (33% of patients) had lower progression-free (HR = 2.33; 95% CI, 1.17-4.64) and overall (HR = 1.92; 95% CI, 0.63-5.80) survival. These results are consistent with our findings that patients with 2 to 4 risk alleles (38% of patients) had a lower overall survival compared with patients with 0 or 1 risk alleles (HR = 1.27; 95% CI, 0.63-1.48), although our HR was weaker and not statistically significant. However, our HR increased to 1.72 when we compared patients with 3 or 4 versus 0, 1, or 2 risk alleles, suggesting a gradient in risk with the number of adverse alleles. In a follow-up study, Warzocha et al reported that of SNPs in *LTA*<sub>+252</sub>, *TNF* (*TNF*<sub>-376</sub>, *TNF*<sub>-308</sub>, *TNF*<sub>-238</sub>, *TNF*<sub>-163</sub>), and *HLA DRB1\*02*, only the *TNF* -308A allele was associated with higher levels of TNF and its associated receptors p55 and p75. The *TNF* -308A allele was also an independent predictor of freedom from progression (relative risk (RR) = 1.63) and overall survival (RR = 1.51) in DLBCL,<sup>19</sup> and there was no evidence of a *TNF/LTA* haplotype effect in this analysis. Of note, the *TNF* G-308A promoter polymorphism has been associated

with the development of DLBCL in this study population<sup>16</sup> and in the InterLymph consortium pooled dataset,<sup>15</sup> suggesting a role for this SNP (or another in strong linkage disequilibrium with it) in both the etiology and prognosis of DLBCL.

The other SNPs in our multi-SNP model (*IL1A*, *IL8RB*, and *IL4R*) have not been previously evaluated as DLBCL prognostic factors. The *IL1A* (rs1800587) -889T allele has been associated with higher interleukin-1 (IL1) production<sup>34</sup> and an elevated erythrocyte sedimentation rate,<sup>35</sup> but does not appear to be associated with risk of developing DLBCL.<sup>15,36</sup> *IL8RB* encodes for the receptor of the chemokine IL8, a potent neutrophil chemoattractant whose expression is greatly enhanced by IL1 and TNF.<sup>37</sup> We observed an association with the *IL8RB* SNP rs1126580, which is located in the 3' untranslated region of the gene, but we did not observe any associations with the other *IL8RB* or *IL8* SNPs or with haplotypes in these genes. Although we observed lower survival with the *IL8RB* rs1126580 AG/GG genotype, this genotype was associated with lower risk of developing DLBCL in this study population,<sup>16</sup> although there was no association between this SNP and DLBCL risk in another study.<sup>36</sup> The common allele (C) for the *IL4R* SNP rs2107356 was associated with poorer survival in our study; in etiology studies, this allele has been associated with a lower risk of developing DLBCL in this study population<sup>16</sup> but not in 2 other studies.<sup>17,36</sup> IL4 is central to B cells switching to IgE antibody production and maturation of helper T cells to a Th2 phenotype, and the IL4 receptor is crucial for binding and signal transduction of both IL4 and IL13.<sup>38</sup> Several genes that correlate with DLBCL survival (eg, *BCL6*, *HGAL*)<sup>7</sup> are IL4-specific target genes,<sup>39,40</sup> although IL4 may use different signaling pathways in the germinal center B cell-like versus the activated B cell-like subtypes of DLBCL,<sup>41</sup> and future studies should consider these subtypes.

We found a suggestive positive association of the *IL10* A-1082G allele with DLBCL survival (HR<sub>AG/GG</sub> = 1.48; 95% CI, 0.91-2.38); and although this SNP did not make it into our final multi-SNP risk score, it did enter 27% of the multigene bootstrap models (overall ranked 8th). In contrast, Lech-Maranda et al<sup>20</sup> found that this allele was inversely associated with overall survival in 199 DLBCL patients (RR = 0.78, *P* = .001), although 2 other studies reported no association.<sup>42,43</sup> We observed no association of *IL10* rs1800871 (C-819T), rs1800872 (C-592A), and rs1800890 (T-3575A) with DLBCL survival; the results of the former 2 SNPs are consistent with other studies,<sup>20,43</sup> whereas the latter SNP has not been previously evaluated for DLBCL survival. Two microsatellite loci<sup>44</sup> and 4 SNPs (-819C, -592C, -1082G, and -3575T)<sup>20,45,46</sup> in the *IL10* promoter have been associated with greater IL10 production, and higher IL10 levels have been associated with poorer prognosis in DLBCL in some,<sup>20,47</sup> but not all,<sup>48-50</sup> studies. We found that *IL10* haplotypes that included alleles with putative greater IL10 production were also the alleles most strongly associated with lower overall survival. With respect to etiology, the *IL10* -1082G and -3575T alleles have been associated with risk of developing DLBCL.<sup>15,17,36</sup>

A major strength of this study was the population-based ascertainment of incident cases of DLBCL. Whereas lack of standardization in treatment and clinical follow-up is a limitation of observational studies relative to clinical trials, clinical trials are often conducted in highly selected patient populations and therefore may not be representative of patients in the community. Furthermore, protection from confounding by the clinical trial design is less compelling in this context, where genotype is

unlikely to confound treatment choice. Our observations, if validated, could be considered for general application to community-based patients. This study is the largest study of immune candidate SNPs in relation to survival conducted to date in DLBCL. The genes and SNPs were selected based on either functional data or their association with cancer or other immune-related diseases, and extensive quality controls were used to ensure high-quality genotyping. Our statistical analyses were comprehensive, and we have been cautious to evaluate the robustness of our results to both false positives and false negatives. Nevertheless, this analysis must be acknowledged as a first step, as other SNPs or haplotypes for these genes, or other immune genes we did not assess, may be of greater prognostic relevance.

A limitation of this study was the lack of detailed data on prognostic factors or treatment. However, we did have age, stage, B symptoms, and treatment class, and these variables predicted survival with a level of predictive ability similar to the IPI for a large study of DLBCL patients.<sup>8</sup> Pathology classification was based on the cancer registry report without central review. Our sample was 20 to 74 years of age and may not generalize to patients 75 years of age and older. As discussed in "Descriptive results," the study design did not capture patients with aggressive disease who died shortly after diagnosis. Finally, all of these patients were initially treated before 2000, before the widespread use of rituximab in the treatment of DLBCL.

In conclusion, host genetic variation in the cytokine and chemokine genes *IL1A*, *IL8RB*, *IL4R*, *IL10*, and *TNF*, individually and particularly in combination, were associated with late survival (> 12 months) in DLBCL after accounting for clinical and demographic factors. Our results suggest that patients with a greater propensity to produce TNF- $\alpha$ , IL-10, and IL-1 (and thus a proinflammatory state) may promote lymphomagenesis, decrease the ability of the host to eradicate lymphoma, or perhaps impair therapeutic efficacy, leading to poorer overall survival. These same

*TNF* and *IL10* SNPs and haplotypes also appear to increase risk of developing DLBCL, supporting a shared mechanism in the etiology and prognosis of DLBCL. The association with *IL8RB* supports a role for the tumor microenvironment in the biologic and clinical behavior of DLBCL. Thus, immunogenetics represents a promising class of prognostic factors that warrants further evaluation in DLBCL.

## Acknowledgments

The authors thank Peter Hui and Cristine Allmer for programming assistance and Sondra Buehler for assistance in manuscript preparation.

This work was supported by National Institutes of Health (Bethesda, MD) grants R01 CA96704 and P50 CA97274; National Cancer Institute Intramural Program; and SEER contracts N01-PC-67 010, N01-PC-67 008, N01-PC-67 009, N01-PC-65 064, and N02-PC-71 105.

## Authorship

Contribution: J.R.C. and P.H. designed the study; J.R.C., P.H., N.R., and S.J.C. obtained funding; J.R.C., W.C., S.D., P.H., C.F.L., and R.K.S. obtained clinical data; S.J.C., S.S.W., and N.R. obtained genetic data; M.J.M. and S.M.G. performed statistical analysis; T.M.H. and J.R.C. drafted the manuscript; and all authors revised the manuscript and reviewed and approved the final manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Correspondence: James R. Cerhan, Division of Epidemiology, Mayo Clinic College of Medicine, 200 1st Street SW, Rochester, MN 55905; e-mail: cerhan.james@mayo.edu.

## References

- Jaffe ES, Harris N, Stein H, Vardiman J. World Health Organization Classification of Tumours. Pathology & Genetics: Tumours of Hematopoietic and Lymphoid Tissues. Lyon, France: IARC Press; 2001.
- Morton LM, Wang SS, Devesa SS, Hartge P, Weisenburger DD, Linet MS. Lymphoma incidence patterns by WHO subtype in the United States, 1992-2001. *Blood*. 2006;107:265-276.
- Coiffier B, Lepage E, Briere J, et al. CHOP chemotherapy plus rituximab compared with CHOP alone in elderly patients with diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346:235-242.
- Feugier P, van Hoof A, Sebban C, et al. Long-term results of the R-CHOP study in the treatment of elderly patients with diffuse large B-cell lymphoma: a study by the Groupe d'Etude des Lymphomes de l'Adulte. *J Clin Oncol*. 2005;23:4117-4126.
- Habermann TM, Weller EA, Morrison VA, et al. Rituximab-CHOP vs CHOP alone or with maintenance rituximab in older patients with diffuse large B-cell lymphoma. *J Clin Oncol*. 2006;24:3121-3127.
- Gascoyne RD. Emerging prognostic factors in diffuse large B cell lymphoma. *Curr Opin Oncol*. 2004;16:436-441.
- Lossos IS. Molecular pathogenesis of diffuse large B-cell lymphoma. *J Clin Oncol*. 2005;23:6351-6357.
- Rosenwald A, Wright G, Chan WC, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002;346:1937-1947.
- Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. *Proc Natl Acad Sci U S A*. 2003;100:9991-9996.
- Monti S, Savage KJ, Kutok JL, et al. Molecular profiling of diffuse large B-cell lymphoma identifies robust subtypes including one characterized by host inflammatory response. *Blood*. 2005;105:1851-1861.
- Hunter KW, Crawford NP. Germ line polymorphism in metastatic progression. *Cancer Res*. 2006;66:1251-1254.
- Cerhan JR, Wang SS, Maurer MJ, et al. Prognostic significance of host immune gene polymorphisms in follicular lymphoma survival. *Blood*. 2007;109:5439-5446.
- Kurzrock R. Cytokine deregulation in hematological malignancies: clinical and biological implications. *Clin Cancer Res*. 1997;3:2581-2584.
- Balkwill F, Mantovani A. Inflammation and cancer: back to Virchow? *Lancet*. 2001;357:539-545.
- Rothman N, Skibola CF, Wang SS, et al. Genetic variation in TNF and IL10 and risk of non-Hodgkin lymphoma: a report from the InterLymph Consortium. *Lancet Oncol*. 2006;7:27-38.
- Wang SS, Cerhan JR, Hartge P, et al. Common genetic variants in proinflammatory and other immunoregulatory genes and risk for non-Hodgkin lymphoma. *Cancer Res*. 2006;66:9771-9780.
- Lan Q, Zheng T, Rothman N, et al. Cytokine polymorphisms in the Th1/Th2 pathway and susceptibility to non-Hodgkin lymphoma. *Blood*. 2006;107:4101-4108.
- Warzocha K, Ribeiro P, Bienvu J, et al. Genetic polymorphisms in the tumor necrosis factor locus influence non-Hodgkin's lymphoma outcome. *Blood*. 1998;91:3574-3581.
- Juszczynski P, Kalinka E, Bienvu J, et al. Human leukocyte antigens class II and tumor necrosis factor genetic polymorphisms are independent predictors of non-Hodgkin lymphoma outcome. *Blood*. 2002;100:3037-3040.
- Lech-Maranda E, Baseggio L, Bienvu J, et al. Interleukin-10 gene promoter polymorphisms influence the clinical outcome of diffuse large B-cell lymphoma. *Blood*. 2004;103:3529-3534.
- 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). *Blood*. 2007;110:695-708.
- Packer BR, Yeager M, Staats B, et al. SNP500Cancer: a public resource for sequence validation and assay development for genetic variation in candidate genes. *Nucleic Acids Res*. 2004;32:D528-D532.
- Cox DR. Regression models and life tables (with discussion). *J R Statist Soc B*. 1972;34:187-220.
- Freidlin B, Zheng G, Li Z, Gastwirth JL. Trend tests for case-control studies of genetic markers: power, sample size and robustness. *Hum Hered*. 2002;53:146-152.
- Shipp MA, Harrington DP, Anderson JR, et al. A



- predictive model for aggressive non-Hodgkin's lymphoma: the International Non-Hodgkin's Lymphoma Prognostic Factors Project. *N Engl J Med*. 1993;329:987-994.
26. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983;70:41-55.
  27. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA. Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet*. 2002;70:425-434.
  28. Taylor J, Tibshirani R. A tail strength measure for assessing the overall univariate significance in a dataset. *Biostatistics*. 2006;7:167-181.
  29. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J R Statist Soc B*. 1995;57:289-300.
  30. Wu X, Gu J, Grossman HB, et al. Bladder cancer predisposition: a multigenic approach to DNA-repair and cell-cycle-control genes. *Am J Hum Genet*. 2006;78:464-479.
  31. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000;56:337-344.
  32. Surveillance, Epidemiology, and End Results (SEER) Program. SEER\*Stat Database: Incidence-SEER 9 Regs Limited-Use, Nov 2007 Sub (1973-2005) <Katrina/Rita Population Adjustment>, Linked To County Attributes, Total U.S., 1969-2005 Counties, National Cancer Institute, DCCPS, Surveillance Research Program, Cancer Statistics Branch, released April 2008, based on the November 2007 submission. ([www.seer.cancer.gov](http://www.seer.cancer.gov)). Accessed on May 17, 2007.
  33. Warzocha K, Salles G, Bienvu J, et al. Tumor necrosis factor ligand-receptor system can predict treatment outcome in lymphoma patients. *J Clin Oncol*. 1997;15:499-508.
  34. Hulkkonen J, Laippala P, Hurme M. A rare allele combination of the interleukin-1 gene complex is associated with high interleukin-1 beta plasma levels in healthy individuals. *Eur Cytokine Netw*. 2000;11:251-255.
  35. McDowell TL, Symons JA, Ploski R, Forre O, Duff GW. A genetic association between juvenile rheumatoid arthritis and a novel interleukin-1 alpha polymorphism. *Arthritis Rheum*. 1995;38:221-228.
  36. Purdue MP, Lan Q, Kricker A, et al. Polymorphisms in immune function genes and risk of non-Hodgkin lymphoma: findings from the New South Wales non-Hodgkin Lymphoma Study. *Carcinogenesis*. 2007;28:704-712.
  37. De Larco JE, Wuertz BR, Furcht LT. The potential role of neutrophils in promoting the metastatic phenotype of tumors releasing interleukin-8. *Clin Cancer Res*. 2004;10:4895-4900.
  38. Shirakawa I, Deichmann KA, Izuhara I, Mao I, Adra CN, Hopkin JM. Atopy and asthma: genetic variants of IL-4 and IL-13 signalling. *Immunol Today*. 2000;21:60-64.
  39. Lossos IS, Alizadeh AA, Rajapaksa R, Tibshirani R, Levy R. HGAL is a novel interleukin-4-inducible gene that strongly predicts survival in diffuse large B-cell lymphoma. *Blood*. 2003;101:433-440.
  40. Schroder AJ, Pavlidis P, Arimura A, Capece D, Rothman PB. Cutting edge: STAT6 serves as a positive and negative regulator of gene expression in IL-4-stimulated B lymphocytes. *J Immunol*. 2002;168:996-1000.
  41. Lu X, Nechushtan H, Ding F, et al. Distinct IL-4-induced gene expression, proliferation, and intracellular signaling in germinal center B-cell-like and activated B-cell-like diffuse large-cell lymphomas. *Blood*. 2005;105:2924-2932.
  42. Berglund M, Thunberg U, Roos G, Rosenquist R, Enblad G. The interleukin-10 gene promoter polymorphism (-1082) does not correlate with clinical outcome in diffuse large B-cell lymphoma. *Blood*. 2005;105:4894-4895; author reply 4895.
  43. Kube D, Hua T-D, Kloss M, et al. The interleukin-10 gene promoter polymorphism-1087AG does not correlate with clinical outcome in non-Hodgkin's lymphoma. *Genes Immun*. 2007;8:164-167.
  44. Eskdale J, Gallagher G, Verweij CL, Keijsers V, Westendorp RG, Huizinga TW. Interleukin 10 secretion in relation to human IL-10 locus haplotypes. *Proc Natl Acad Sci U S A*. 1998;95:9465-9470.
  45. Turner DM, Williams DM, Sankaran D, Lazarus M, Sinnott PJ, Hutchinson IV. An investigation of polymorphism in the interleukin-10 gene promoter. *Eur J Immunogenet*. 1997;24:1-8.
  46. Gibson AW, Edberg JC, Wu J, Westendorp RG, Huizinga TW, Kimberly RP. Novel single nucleotide polymorphisms in the distal IL-10 promoter affect IL-10 production and enhance the risk of systemic lupus erythematosus. *J Immunol*. 2001;166:3915-3922.
  47. Blay JY, Burdin N, Rousset F, et al. Serum interleukin-10 in non-Hodgkin's lymphoma: a prognostic factor. *Blood*. 1993;82:2169-2174.
  48. Stasi R, Zinzani L, Galieni P, et al. Clinical implications of cytokine and soluble receptor measurements in patients with newly diagnosed aggressive non-Hodgkin's lymphoma. *Eur J Haematol*. 1995;54:9-17.
  49. Cortes JE, Talpaz M, Cabanillas F, Seymour JF, Kurzrock R. Serum levels of interleukin-10 in patients with diffuse large cell lymphoma: lack of correlation with prognosis. *Blood*. 1995;85:2516-2520.
  50. Fabre-Guillevin E, Tabrizi R, Coulon V, et al. Aggressive non-Hodgkin's lymphoma: concomitant evaluation of interleukin-2, soluble interleukin-2 receptor, interleukin-4, interleukin-6, interleukin-10 and correlation with outcome. *Leuk Lymphoma*. 2006;47:603-611.