Published in final edited form as:

Leuk Lymphoma. 2012 June ; 53(6): 1105–1112. doi:10.3109/10428194.2011.638717.

LMO2 protein expression, *LMO2* germline genetic variation, and overall survival in diffuse large B-cell lymphoma in the pre-rituximab era

James R. Cerhan¹, Yasodha Natkunam², Lindsay M. Morton³, Matthew J. Maurer⁴, Yan Asmann⁴, Thomas M. Habermann⁵, Mohammad A. Vasef⁶, Wendy Cozen⁷, Charles F. Lynch⁸, Cristine Allmer⁴, Susan L. Slager⁴, Izidore S. Lossos⁹, Stephen J. Chanock³, Nathaniel Rothman³, Patricia Hartge³, Ahmet Dogan¹⁰, and Sophia S. Wang¹¹

¹Division of Epidemiology, College of Medicine, Mayo Clinic, Rochester, MN ²Department of Pathology, Stanford University School of Medicine, Stanford, CA ³Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, Maryland ⁴Division of Biomedical Statistics and Informatics, College of Medicine, Mayo Clinic, Rochester, MN ⁵Division of Hematology, College of Medicine, Mayo Clinic, Rochester, MN ⁶Department of Pathology, University of New Mexico, Albuquerque, NM ⁷Department of Preventive Medicine, University of Southern California, Los Angeles, CA ⁸Department of Epidemiology, University of Iowa, Iowa City, IA ⁹Division of Oncology-Hematology, Department of Medicine, University of Miami, Miami, FL ¹⁰Department of Laboratory Medicine and Pathology, College of Medicine, Mayo Clinic, Rochester, MN ¹¹Department of Population Sciences, City of Hope, Duarte, CA

Abstract

Both *LMO2* mRNA and protein expression in diffuse large B-cell lymphoma (DLBCL) have been associated with superior survival; however, a role for germline genetic variation in *LMO2* has not been previously reported. Immunohistochemistry (IHC) for *LMO2* was conducted on tumor tissue from diagnostic biopsies, and 20 tag single nucleotide polymorphisms (SNPs) from *LMO2* were genotyped from germline DNA. *LMO2* IHC positivity was associated with superior survival (HR=0.55; 95% CI 0.31–0.97). Four *LMO2* SNPs (rs10836127, rs941940, rs750781, rs1885524) were associated with survival after adjusting for *LMO2* IHC and clinical factors ($p < 0.05$), and one of these SNPs (rs941940) was also associated with IHC positivity ($p = 0.02$). Compared to a model with clinical factors only (c-statistic=0.676), adding the 4 SNPs (c-statistic=0.751) or *LMO2* IHC (c-statistic=0.691) increased the predictive ability of the model, while inclusion of all 3 factors (c-statistic=0.754) did not meaningfully add predictive ability above a model with clinical factors and the 4 SNPs. In conclusion, germline genetic variation in *LMO2* was associated with DLBCL prognosis and provided slightly stronger predictive ability relative to *LMO2* IHC status.

Keywords

Diffuse large B-cell lymphoma; *LMO2*; prognosis; single nucleotide polymorphisms

Correspondence: James R. Cerhan, MD, PhD, Division of Epidemiology, College of Medicine, Mayo Clinic, 200 First Street SW, Rochester, MN 55905, Phone: 507-538-0499; Fax: 507-266-2478; cerhan.james@mayo.edu.

Potential conflict of interest: There are no conflicts of interest.

Introduction

LMO2 (LIM domain only 2) is located on 11p13 and belongs to a family of four genes encoding LIM-only proteins, which are transcription regulators that control cell fate in normal hematopoiesis [1] and endothelial cell remodeling [2]. *LMO2* encodes a 156 amino acid protein comprised of two zinc-binding LIM domains, which function in protein-protein interactions in the transcription factor complex that includes E2A, TAL2, GATA1, and LDB1 in erythroid cells [3,4].

While *LMO2* is perhaps most well-known for its role as a T-cell oncogenic protein [5,6], gene expression studies have found that *LMO2* mRNA expression in diffuse large B-cell lymphoma (DLBCL) was part of the “germinal center” expression profile [7], and it emerged as the strongest predictor of overall survival in DLBCL both in the univariate and multivariate setting of a six-gene model [8]. A monoclonal anti-*LMO2* antibody was subsequently developed, and immunohistochemical (IHC) analysis showed that *LMO2* protein was expressed as a nuclear marker in normal germinal center B-cells and hematopoietic lineages, as well as leukemias and a subset of germinal center derived B-cell lymphomas [9].

Approximately 50% of DLBCL patients express *LMO2* protein by IHC analysis, and *LMO2* protein expression has been associated with better progression-free and overall survival among DLBCL patients treated with anthracycline-based chemotherapy or rituximab plus anthracycline-based chemotherapy [10]. Unlike its role in leukemias, *LMO2* protein expression in DLBCL has not been associated with any somatic genetic alterations [9,10]. This raises the hypotheses that germline genetic variation could play a role in expression of *LMO2*, and in the setting of DLBCL, could also be associated with prognosis. We tested these hypotheses in a prognostic cohort of DLBCL patients who participated in a population-based study conducted from 1998–2000 and had data on both germline genotyping for *LMO2* [11], and *LMO2* expression as assessed by immunohistochemistry in formalin-fixed, paraffin embedded tumor tissue [12].

Design and methods

Study population

Institutional Review Boards at the National Cancer Institute and each Surveillance, Epidemiology and End Results (SEER) center approved the study protocol. Participants provided written, informed consent prior to completing an in-person interview. Details on survival of the DLBCL patients has been previously described [12,13]. Briefly, 1,321 subjects aged 20–74 years with incident, histologically-confirmed NHL first diagnosed from July 1998 through June 2000 were enrolled in a population-based case-control study. All cases were rapidly reported from four SEER cancer registries in the Detroit metropolitan area, the state of Iowa, Los Angeles County, and Seattle. Any patients known to be HIV-positive were excluded. A total of 1172 cases (89%) provided either a peripheral blood (N=773) or mouthwash buccal sample (N=399) as a source for DNA.

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 registry databases at each study site. Data on first course of therapy included use of single or multi-agent chemotherapy, radiation, other therapies exclusive of chemotherapy and/or radiation, and no therapy (presumed to be observation); information on individual agents and doses was not available. The SEER registries collect date and cause of death, but do not collect data on treatment response, disease recurrence or progression. In 2008, we conducted a second linkage to each SEER registry to update survival information through the end of 2007.

Pathology classification

All cases in the study were initially histologically confirmed as NHL and coded according to the International Classification of Diseases for Oncology, 2nd Edition (ICD-O-2) [14] by the local diagnosing pathologist. Final ICD-O-3 codes based on local pathology review were received during the SEER record linkage. Pathology reports were obtained for 93% (1228 of 1321) of patients for review by an expert hematopathologist (MAV), who classified cases according to the World Health Organization classification for lymphoid neoplasms [15] and assigned a confidence score to the subtype diagnosis (90% versus <90%). For all cases with low confidence in the NHL subtype classification as well as a 20% random sample of cases with high confidence in the NHL subtype classification, additional immunostaining was conducted to establish the NHL subtype for those patients with available pathology material (N=472). All cases were then assigned a final diagnosis based on the pathology review, updated SEER record linkage if pathology review was not available, or original SEER data if updated data were not available. A total of 417 cases had a final diagnosis of DLBCL (ICD-O-2: 9680–84, 9688, 9712; ICD-O-3: 9678–80, 9684), and all but 36 of these were confirmed by pathology report or slide review.

Immunohistochemistry

Sufficient archived, unstained 5-micron slides from formalin-fixed, paraffin-embedded tumor biopsies were available for 239 of the 417 DLBCL cases (57%) for LMO2 IHC staining, which was performed at Stanford University under the direction of a hematopathologist (Y.N.) using an established protocol [9,10]. Two hematopathologists (Y.N. and A.D.) then independently scored the slides as no expression (0%), expression 30%, and expression >30%; staining greater than 30% was classified as LMO2 positive, while 0% and <30% staining was classified as LMO2 negative using the previously established threshold [10]. LMO2 staining of endothelial cells served as an internal control. The concordance between the two hematopathologists was 89%; discordant cases were reviewed by the two hematopathologists together to assign a final consensus score. Of the 239 cases stained, 5 cases were excluded due to insufficient tissue quality to make a final LMO2 determination, leaving 234 cases for analysis.

Germline genotyping

Details on genotyping for these samples have been previously described [11]. Briefly, genotyping of *LMO2* SNPs was conducted at the NCI Core Genotyping Facility (Advanced Technology Center, Gaithersburg, MD; <http://snp500cancer.nci.nih.gov>) [16] as part of a larger, custom-designed GoldenGate assay (Illumina, www.illumina.com), supplemented by TaqMan genotyping. *LMO2* tagging SNPs were chosen from the designable set of common SNPs at a minor allele frequency of >5% and a binning threshold of $r^2 > 0.8$; [17] a total of 22 SNPs were selected, which represents 88% coverage based on the number of SNPs genotyped divided by the number of bins from the designable set of SNPs.

Quality control was implemented at the level of the entire 1536 SNP OPA for all 1001 (of 1321) NHL cases with a DNA sample [11]. We excluded SNPs that failed to properly cluster in the genotyping calling algorithm (N=3) and SNPs with low completion rate (<90% of samples; N=5). Quality control duplicates with concordance <95% were excluded (N=1). We also excluded samples with a low completion rate (<90% of the full panel of 1536 SNPs, N=17). Hardy–Weinberg equilibrium was evaluated among non-Hispanic Caucasian controls of the parent study, and 3 SNPs showed evidence ($p < 0.001$) for deviation from Hardy–Weinberg proportions, including one *LMO2* SNP (rs7941248), but this SNP was retained in the analysis as there was no obvious genotyping error. Thus, all 22 *LMO2* SNPs were successfully genotyped. However, we excluded rs2038602 due to lack of any variation

(all wild type homozygotes) and rs941941 as it was only available on 50 cases (rest of cases not genotyped on that SNP due to availability of only a buccal DNA sample).

Of the 234 cases with clinical, outcome and IHC data, 162 also had genotyping data, and compromised the final analysis sample.

Data analysis

Overall survival was defined as the time from diagnosis to the date of death or last follow-up; patients alive at last follow-up were censored at that time point. We used Cox proportional hazards regression [18] to estimate hazard ratios (HRs) and 95% confidence intervals (CIs). To efficiently adjust for clinical and demographic factors, we used a previously developed risk score ("clinical risk score"), analogous to a propensity score in logistic regression [19]. The clinical risk score is a linear combination of age (<60 versus 60+ years), stage (local, regional, distant, or missing), B-symptoms (no, yes, missing), initial therapy (chemotherapy + radiation, chemotherapy + other therapy, radiation only, all other therapy, or missing), sex, race (white, other), years of education (<12, 12–15, 16+), and study center (Detroit, Iowa, Los Angeles, Seattle). To address multiple testing, we calculated q-values [20] for the primary test of each SNP with overall survival; a q-value < 0.05 was considered to be noteworthy.

To assess the correlation of LMO2 IHC staining and host genotypes, we use Chi-squared tests and logistic regression. To address the predictive ability of these results, we used time dependent ROC curves and corresponding c-statistics at 8 years (the approximate median follow-up) were used to assess and compare the prognostic ability of survival models [21].

Several bioinformatics tools were utilized to identify biological significance of the genetic variants. First, to identify evolutionarily conserved regions, we used phastCons, which is part of the PAST (PHYlogenetic Analysis with Space/Time models) package. The phastCons is a hidden Markov model-based method that estimates the probability that each nucleotide belongs to a conserved element based on the multiple alignments. It considers not just each individual alignment column, but also its flanking columns. We also used ESPERR (Evolutionary and Sequence Pattern Extraction through Reduced Representations) to calculate a Regulatory Potential score, which can discriminate regulatory regions from neutral sites with excellent accuracy (approximately 94%) [22,23]. To identify SNPs that might be in regions that bind transcription factors, we used Transfac Matrix Database (v.7.0) Public 2005 (<http://www.gene-regulation.com/pub/databases.html>) which contains position-weight matrices for 398 transcription factor binding sites, as characterized through experimental results in the scientific literature. Finally, we used ENCODE (Encyclopedia of DNA Elements, <http://www.genome.gov/10005107>) Integrated Regulation tracks, which can be accessed through the UCSC genome browser, to identify functional elements of the human genome for transcription regulation.

Results

Cohort characteristics

The mean age at diagnosis of the 162 DLBCL patients in this analysis was 60 years (range 24–74), and 58% were male. A majority of patients were white (90%). Clinically, 28% of the patients had advanced stage disease, and 27% had B-symptoms. Based on cancer registry data, the most common treatment was a chemotherapy-based regimen (84%). During a median follow-up of 92 months (range, 27 to 110 months), there were 52 deaths (32%). Of the 52 deaths, 36 (69%) were due to lymphoma. Our clinical risk score (a linear combination of age, stage, B-symptoms, initial therapy, sex, race, education, and study center as described in the Methods section) ranged from –1.34 to 1.43, and as expected, it was

strongly predictive of overall survival (HR=3.12 per unit increase in the score; 95% CI 1.86–5.25; $p<0.0001$). Comparing the 162 patients in this analysis to the 249 that did not have tissue or genotyping data, there were no statistically significant differences based on age, sex, education, stage or treatment status (data not shown).

LMO2 IHC staining and survival

Defining positivity based on the 30% cutpoint [10], 79 of the 162 patients (48%) were classified as LMO2 positive. Compared to LMO2 negative patients, LMO2 positive patients had superior overall survival (HR=0.55; 95% CI 0.31–0.97; $p=0.04$). Further adjustment for the clinical risk score did not alter this association (HR=0.56; 95% CI 0.32–0.98; $p=0.04$).

SNP-level associations with survival

Table I summarizes the SNP-level results with overall survival after adjustment for the clinical risk score. Two SNPs (rs10128650 and rs10836127) were statistically significantly associated with overall survival at $p<0.05$. The SNP rs10128650 (MAF of 0.06) is located in a highly conserved domain in the promoter region (by PhastCons analysis), and carrying a variant allele was associated with inferior survival (HR=2.89 for each variant allele, 95% CI 1.72–4.87, p -trend=0.00007; q -value 0.0014). The conserved domain in which rs10128650 resides has a very high regulatory potential score of >0.4 [22]. The intronic SNP rs10836127 had a MAF of 0.18, and carrying a variant allele was also associated with inferior survival (HR=2.18 per variant allele, 95% CI 1.39–3.43; p -trend=0.001; q -value=0.010). The SNP rs10836127 was not located in any known functional genomic domains, but was in linkage disequilibrium (LD) with rs10128650 ($D'=0.79$; $r^2=0.19$).

Four other SNPs approached statistical significance (0.05 p -trend 0.15): rs4756077 (intronic SNP, p -trend=0.061), rs750781 (promoter SNP, p -trend=0.087), rs941940 (promoter SNP, p -trend=0.13), and rs1885524 (intronic SNP, p -trend=0.15). With the exception of rs1885524, all of the remaining SNPs were in weak LD with the top two SNPs (Figure 1).

Using a manual backwards selection strategy to evaluate these six SNPs in a multivariate model adjusted for the clinical risk score, four SNPs remained statistically significant at $p<0.05$: rs1885524 (HR=1.78, 95% CI 1.16–2.72; p -trend=0.008), rs750781 (HR=2.23, 95% CI 1.33–3.73; p -trend=0.0024), rs941940 (HR=0.54, 95% CI 0.32–0.90; p -trend=0.019), and rs10836127 (HR=1.85, 95% CI 1.14–3.01; p -trend=0.013). In a model that further included LMO2 IHC expression, the HRs for the four *LMO2* SNPs were essentially unchanged (data not shown), while the HR for LMO2 IHC expression attenuated slightly (HR=0.63, 95% CI 0.36–1.13; $p=0.12$).

There were no changes in these associations when analyses were restricted to lymphoma deaths as the outcome (non-lymphoma deaths censored; data not shown). When we excluded patients who did not receive any chemotherapy or restricted to white patients only, all significant SNP associations strengthened (data not shown).

Correlation of SNP genotype with LMO2 IHC expression

Of the four SNPs associated with survival from the multivariate model, only the *LMO2* promoter SNP rs941940 was significantly associated with LMO2 IHC expression ($p=0.02$). Inspection of the data in Table II supported a recessive model, and patients who were variant homozygotes were 2.89 times more likely to be LMO2 IHC positive compared to wild type homozygotes and heterozygotes combined (95% CI 1.26–6.54). This finding parallels the survival results, which showed that carrying a variant allele was associated with better survival (see previous section). For the other 3 SNPs associated with survival, two of them

showed no variability in LMO2 IHC expression (rs1885524 and rs750781), while rs10836127 variant homozygotes showed lower LMO2 IHC expression (20%) compared to heterozygotes (52%) and wild type homozygotes (48%), although the global test was not statistically significant ($p=0.29$).

The only other SNP to be significantly correlated with LMO2 IHC expression was the promoter SNP rs7941248 ($p=0.009$). Inspection of Table II supported a dominant model, and patients carrying 1 or 2 variant alleles were 2.50 times more likely to be LMO2 positive compared to wild type homozygotes (95% CI 1.33–4.71). However, this SNP itself was not significantly associated with survival ($p\text{-trend}=0.33$), although the per allele HR was <1.0 (HR=0.81; 95% CI 0.53–1.24; Table I), which (weakly) parallels the IHC results. In the HapMap Phase II Caucasian population, this SNP is in high LD with multiple conserved transcription factor binding sites (TRANSFAC 7.0 Public 2005), including the binding site of a transcription factor, GATA1, which has been reported to regulate the *LMO2* expression [24], suggesting a potential function underlying this association.

Predictive ability of a combined LMO2 expression and LMO2 SNP model

To assess the prognostic ability of LMO2 expression and *LMO2* SNPs, we used a time dependent area under the curve (AUC) model, and calculated the concordance index (c-statistic) at 8 years follow-up, which provides values from 0.500 (prediction no better than chance) to 1.000 (perfect prediction). The model with the clinical risk score alone had a c-statistic of 0.676, which is consistent with the prediction ability of the IPI for DLBCL [13]. The c-statistic for the model with LMO2 IHC expression alone was 0.573. A model with both the clinical risk score and LMO2 IHC was 0.691. The c-statistic for each of the four SNPs that remained statistically significant in a multivariate model ranged from 0.525–0.579 (Table III). When all four SNPs were included in the same model with no other factors, the c-statistic was 0.672. Adding LMO2 IHC to individual SNPs increased the c-statistic, although adding LMO2 IHC to the four SNP model had minimal impact on the model (c-statistic=0.683). Adding the clinical risk score to either individual SNPs or to the four SNP model increased model prediction. A full model of four SNPs, LMO2 IHC, and clinical risk score was essentially identical (c-statistic=0.754) to the four SNP and clinical risk score model (c-statistic=0.751).

Additional bioinformatics considerations

For the seven *LMO2* SNPs showing some association with either survival ($N=6$) or LMO2 IHC expression ($N=2$) (Figure 1), we further looked for additional SNPs in LD ($r^2 \geq 0.8$, HapMap CEU population) within 250kb of these seven SNPs to identify potential causal SNPs. There were two SNPs in LD with the seven candidate *LMO2* SNPs. The SNP rs7119405 is in LD with rs10128650 ($D'=1$, and $r^2=0.94$). However, SNP rs7119405 is not in a conserved region and does not overlap with any known functional genomic regions, suggesting that rs10128650 is more likely to be the biologically more relevant SNP. In addition, SNP rs1885523 is in LD with rs1885524 ($D'=1$, and $r^2=0.98$). SNP rs1885523 is located in the intron of *LMO2* and has been reported to overlap with a polymerase II binding site (ENCODE transcription binding project).

Discussion

Using a population-based cohort of 162 DLBCL patients diagnosed from 1998–2000 (pre-rituximab era) and followed through 2008, we show for the first time that germline genetic variation in four *LMO2* SNPs (rs1885524, rs10836127, rs750781, rs941940) was associated with overall survival after accounting for clinical factors. The correlation of *LMO2* SNP genotype and LMO2 IHC status was more variable, although the SNP rs941940 was

significantly correlated with both expression and survival, and several other SNPs trended towards a correlation with both expression and survival. Individually, the SNPs were weak predictors of outcome (c-statistics 0.525–0.579), similar to that of LMO2 IHC (c-statistic 0.573), and weaker than the clinical risk score (c-statistic=0.676). However, when the individual SNPs were combined with the clinical risk score, the c-statistics improved. Further addition of LMO2 IHC to models with SNPs and the clinical risk score only trivially improved the c-statistics. In total, these results suggest that germline genetic variation in *LMO2* provides as good or slightly better prognostic information than tumor LMO2 IHC status.

Strengths of this study include the population-based ascertainment of newly diagnosed cases, which enhances generalizability. Our genotyping included extensive quality controls. The LMO2 IHC uses a relatively robust antibody, and scoring was highly reliable as assessed by two independent pathologists. Further, the prevalence of LMO2 expression and its association with survival was in the range of previous studies using this antibody [10,25].

There are also several important limitations. We used a retrospective study design and patients were not uniformly treated as in a clinical trial. While this is an important limitation, it needs to be balanced against the limitations of a clinical trial, which are generally based upon a very select subset of generally otherwise healthy patients. We do note that the SNP associations were strongest for patients receiving chemotherapy, which in this setting would be presumably for curative intent. While we did not have data to calculate the IPI, we were able to adjust for clinical and demographic variables, and our clinical risk score predicted with the same robustness as the IPI in other datasets [13]. We also only had IHC and genotyping data on 38% of the original cohort of patients, although we did not see systematic differences based on inclusion in this analysis. We did not include an analysis of germinal center B-cell like (GCB) phenotype in this report. However, in a prior report, we found that 61% of GCB DLBCLs (as assessed by IHC) were LMO2 IHC positive but that 26% of non-GCB cases were also LMO2 positive. Furthermore, there was no significant association of GCB-phenotype with overall survival [12], supporting a role for LMO2 independent of GCB phenotype that has been previously reported [10].

Our study was unable to enroll patients with rapidly fatal disease, consistent with enrollment of only living cases into the parent case-control study. The impact of this bias is that our observed survival is consistent with SEER estimates conditioned on surviving 12 months after diagnosis [13]. Thus, the inferences from this study would not apply to early deaths. Nevertheless, LMO2 IHC positivity and the association with overall survival was similar to other published data [10], suggesting that the association of LMO2 IHC expression as a prognostic marker for overall survival is not particularly impacted by this bias. The number of events was modest (N=52 deaths), and so the models need to be considered in this context, and in the context that we did not have an independent validation sample. Finally, this study was based on patients initially treated in the pre-rituximab era, and thus our findings will need to be evaluated in rituximab era patients.

Our findings that LMO2 expression as measured by IHC in DLBCL is a strong prognostic factor after considering clinical factors replicates in a population-based setting the results of Natkunam and colleagues [10]. They found that LMO2 IHC was positive in 53% of 263 patients treated with anthracycline-based chemotherapy (pre-rituximab) and that LMO2 positive patients had significantly better progression-free (median 12 versus 49 months, $p=0.010$) and overall (median 21 versus 80 months, $p=0.018$) survival. Our results are also consistent with studies that have found higher *LMO2* expression as measured by cDNA microarray [7], and RT-PCR [8], but not qNPA [26], is associated with better survival in the pre-rituximab setting. While we did not have any data on R-CHOP treated patients, LMO2

IHC expression [10] and mRNA expression as assessed by RT-PCR [27] or qNPA [26] have all been found to be associated with better survival in R-CHOP treated patients, although LMO2 IHC expression was not statistically significantly associated with survival (63% 5-year survival for LMO2+ and 61% 5-year survival for LMO2- patients, $p=0.78$) in a study of DLBCL patients age 60 and older [25]. However, the proportion of LMO2 positive cases in this study was somewhat lower than in previous studies and only 82 cases were analyzed for LMO2 expression.

Our data also strongly suggest that common germline genetic variation in the *LMO2* gene is associated with DLBCL prognosis. There were four SNPs in *LMO2* that were associated with overall survival in a multivariate model, two were in the promoter (rs750781 and rs941940), one was intronic (rs10836127) but in weak LD with the promoter SNPs, and the fourth (rs1885524) was intronic and not in LD with the promoter SNPs. The latter intronic SNP was in high LD with rs1885523, which is located in a polymerase II binding site. Of these same four SNPs, only one SNP (rs941940) was significantly associated with LMO2 IHC expression, and the allele associated with expression was also associated with better survival. The only other SNP associated with LMO2 IHC expression, rs7941248, was in high LD with multiple conserved transcription factor binding sites, including GATA1, which has been shown to regulate LMO2 expression [24]. However, rs7941248 was not significantly associated with survival (p -trend=0.33), although patients carrying one or more variant alleles (which was associated with LMO2 IHC expression) did have better survival. The lack of a strong correlation between SNPs predicting survival and SNPs predicting LMO2 IHC expression may be due to variability in LMO2 protein expression, technical issues in LMO2 IHC staining and scoring, or other mechanisms that regulate LMO2 expression (e.g., epigenetics, miRNAs). Nevertheless, taken together, our data support the hypothesis that genetic variation in *LMO2*, particularly in the promoter, may play a role in both LMO2 expression and survival in DLBCL patients.

The biologic relevance of LMO2 in DLBCL is not known, but it appears to be involved in several important physiologic and pathologic processes relevant to lymphomagenesis. LMO2 is an important transcription factor and its expression is required for hematopoiesis, [1,28] angiogenesis early in development [2], and vascular endothelial remodeling in adults [2]. Nuclear LMO2 is also widely expressed in the vasculature of native tissues, including lymphatic vasculature, and is detected in the secretory but not proliferative phase of the endometrial gland, suggesting tissue-specific regulation [29]. From a pathologic perspective, chromosomal translocations with the T-cell receptor locus or insertional mutations have been associated with T-cell leukemias [5,30], although microarray analysis has found ectopic LMO2 expression in many cases without chromosomal changes [31]. In a transgenic mouse model of LMO2, the mice developed T-cell lymphoblastic lymphomas and associated leukemias [32,33]. LMO2 is also uniformly expressed in benign vascular and lymphatic neoplasms and in most malignant vascular neoplasms [29]. Of note, to date LMO2 expression in DLBCL has not been associated with any somatic genetic alterations [9,10]. Physiologic (and aberrant) control of LMO2 expression is only now being unraveled, and early reports support a role for tissue-specific regulatory elements [34] as well as microRNAs (specifically miR-223) during erythroid differentiation [35].

Irrespective of *LMO2* biologic functions, our study shows that genetic variation in *LMO2* in combination with clinical factors is a robust prognostic factor for DLBCL, and that LMO2 IHC did not add additional predictive ability once these factors were considered. If replicated, future prognostic indices should consider germline genetic risk markers in *LMO2* and perhaps other genes known to be prognostic in DLBCL.

Acknowledgments

This work was supported by the National Cancer Institute (grants R01 CA96704, R01 CA129539, P50 CA97274, P01 CA17054, and P30 CA014089; NCI Intramural Program; SEER contracts N01-PC35139, N01-PC67008, N01-PC67009, N01-PC65064, N01-PC71105).

We thank Drs. Scott Davis (Fred Hutchinson Cancer Research Center) and Richard K. Severson (Wayne State University) for contributing data. The authors thank Sondra Buehler for editorial assistance.

References

- Warren AJ, Colledge WH, Carlton MB, Evans MJ, Smith AJ, Rabbitts TH. The oncogenic cysteine-rich LIM domain protein *rbtn2* is essential for erythroid development. *Cell*. 1994; 78:45–57. [PubMed: 8033210]
- Yamada Y, Pannell R, Forster A, Rabbitts TH. The oncogenic LIM-only transcription factor *Lmo2* regulates angiogenesis but not vasculogenesis in mice. *Proc Natl Acad Sci U S A*. 2000; 97:320–4. [PubMed: 10618416]
- Osada H, Grutz G, Axelson H, Forster A, Rabbitts TH. Association of erythroid transcription factors: complexes involving the LIM protein *RBTN2* and the zinc-finger protein *GATA1*. *Proc Natl Acad Sci U S A*. 1995; 92:9585–9. [PubMed: 7568177]
- Wadman IA, Osada H, Grutz GG, et al. The LIM-only protein *Lmo2* is a bridging molecule assembling an erythroid, DNA-binding complex which includes the *TAL1*, *E47*, *GATA-1* and *Ldb1/NLI* proteins. *Embo J*. 1997; 16:3145–57. [PubMed: 9214632]
- Boehm T, Foroni L, Kaneko Y, Perutz MF, Rabbitts TH. The rhombotin family of cysteine-rich LIM-domain oncogenes: distinct members are involved in T-cell translocations to human chromosomes 11p15 and 11p13. *Proc Natl Acad Sci U S A*. 1991; 88:4367–71. [PubMed: 2034676]
- Royer-Pokora B, Loos U, Ludwig WD. *TTG-2*, a new gene encoding a cysteine-rich protein with the LIM motif, is overexpressed in acute T-cell leukaemia with the *t(11;14)(p13;q11)*. *Oncogene*. 1991; 6:1887–93. [PubMed: 1923511]
- Alizadeh AA, Eisen MB, Davis RE, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature*. 2000; 403:503–11. [PubMed: 10676951]
- Lossos IS, Czerwinski DK, Alizadeh AA, et al. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *The New England journal of medicine*. 2004; 350:1828–37. [PubMed: 15115829]
- Natkunam Y, Zhao S, Mason DY, et al. The oncoprotein *LMO2* is expressed in normal germinal-center B cells and in human B-cell lymphomas. *Blood*. 2007; 109:1636–42. [PubMed: 17038524]
- Natkunam Y, Farinha P, Hsi ED, et al. *LMO2* protein expression predicts survival in patients with diffuse large B-cell lymphoma treated with anthracycline-based chemotherapy with and without rituximab. *J Clin Oncol*. 2008; 26:447–54. [PubMed: 18086797]
- Morton LM, Purdue MP, Zheng T, et al. Risk of non-Hodgkin lymphoma associated with germline variation in genes that regulate the cell cycle, apoptosis, and lymphocyte development. *Cancer Epidemiol Biomarkers Prev*. 2009; 18:1259–70. [PubMed: 19336552]
- Morton LM, Cerhan JR, Hartge P, et al. Immunostaining to identify molecular subtypes of diffuse large B-cell lymphoma in a population-based epidemiologic study. *IJMEG*. 2011 in press.
- Habermann TM, Wang SS, Maurer MJ, et al. 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. *Blood*. 2008; 112:2694–702. [PubMed: 18633131]
- Percy, C.; Van Holten, V.; Muir, C. International classification of diseases for oncology. 2. Geneva: World Health Organization; 1990.
- Jaffe, ES.; Harris, NL.; Stein, H.; Vardiman, JW. World Health Organization Classification of Tumours: Pathology and Genetics, Tumours of Hematopoietic and Lymphoid Tissues. Lyon: IARC Press; 2001.
- Packer BR, Yeager M, Burdett L, et al. *SNP500Cancer*: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic acids research*. 2006; 34:D617–21. [PubMed: 16381944]

17. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA. Selecting a maximally informative set of single-nucleotide polymorphisms for association analyses using linkage disequilibrium. *American journal of human genetics*. 2004; 74:106–20. [PubMed: 14681826]
18. Cox DR. Regression models and life tables (with discussion). *Journal of the Royal Statistical Society B*. 1972; 34:187–220.
19. Rosenbaum P, Rubin D. The central role of the propensity score in observational studies for causal effects. *Biometrika*. 1983; 70:41–55.
20. Storey JD. A direct approach to false discovery rates. *J R Statist Soc B*. 2002; 64:479–98.
21. Heagerty PJ, Lumley T, Pepe MS. Time-dependent ROC curves for censored survival data and a diagnostic marker. *Biometrics*. 2000; 56:337–44. [PubMed: 10877287]
22. Kolbe D, Taylor J, Elnitski L, et al. Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. *Genome research*. 2004; 14:700–7. [PubMed: 15060013]
23. Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, Chiaromonte F. ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. *Genome research*. 2006; 16:1596–604. [PubMed: 17053093]
24. Pruess MM, Drechsler M, Royer-Pokora B. Promoter 1 of *LMO2*, a master gene for hematopoiesis, is regulated by the erythroid specific transcription factor GATA1. *Gene Funct Dis*. 2000; 2:87–94.
25. Copie-Bergman C, Gaulard P, Leroy K, et al. Immuno-fluorescence in situ hybridization index predicts survival in patients with diffuse large B-cell lymphoma treated with R-CHOP: a GELA study. *J Clin Oncol*. 2009; 27:5573–9. [PubMed: 19786664]
26. Rimsza LM, Leblanc ML, Unger JM, et al. Gene expression predicts overall survival in paraffin-embedded tissues of diffuse large B-cell lymphoma treated with R-CHOP. *Blood*. 2008; 112:3425–33. [PubMed: 18544678]
27. Malumbres R, Chen J, Tibshirani R, et al. Paraffin-based 6-gene model predicts outcome in diffuse large B-cell lymphoma patients treated with R-CHOP. *Blood*. 2008; 111:5509–14. [PubMed: 18445689]
28. Yamada Y, Warren AJ, Dobson C, Forster A, Pannell R, Rabbitts TH. The T cell leukemia LIM protein Lmo2 is necessary for adult mouse hematopoiesis. *Proc Natl Acad Sci U S A*. 1998; 95:3890–5. [PubMed: 9520463]
29. Gratzinger D, Zhao S, West R, et al. The transcription factor LMO2 is a robust marker of vascular endothelium and vascular neoplasms and selected other entities. *Am J Clin Pathol*. 2009; 131:264–78. [PubMed: 19141387]
30. Hacein-Bey-Abina S, Von Kalle C, Schmidt M, et al. LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science (New York, NY)*. 2003; 302:415–9.
31. Ferrando AA, Neuberg DS, Staunton J, et al. Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer cell*. 2002; 1:75–87. [PubMed: 12086890]
32. Larson RC, Fisch P, Larson TA, et al. T cell tumours of disparate phenotype in mice transgenic for Rbtl-2. *Oncogene*. 1994; 9:3675–81. [PubMed: 7970726]
33. Larson RC, Osada H, Larson TA, Lavenir I, Rabbitts TH. The oncogenic LIM protein Rbtl2 causes thymic developmental aberrations that precede malignancy in transgenic mice. *Oncogene*. 1995; 11:853–62. [PubMed: 7545805]
34. Landry JR, Bonadies N, Kinston S, et al. Expression of the leukemia oncogene Lmo2 is controlled by an array of tissue-specific elements dispersed over 100 kb and bound by Tal1/Lmo2, Ets, and Gata factors. *Blood*. 2009; 113:5783–92. [PubMed: 19171877]
35. Felli N, Pedini F, Romania P, et al. MicroRNA 223-dependent expression of LMO2 regulates normal erythropoiesis. *Haematologica*. 2009; 94:479–86. [PubMed: 19278969]

Figure 1.
LMO2 gene structure and tagSNP mapping. Top: gene position and structure. Bottom: linkage disequilibrium plot of genotyped SNPs (numbers, $|D'|$ values; darker shading indicates higher r^2 values of correlation between SNPs). A box around the SNP rs number indicates an association with overall survival (p-trend 0.15), while an oval around the SNP rs number indicates a significant association with *LMO2* IHC positivity (p 0.15).

Table 1
Results for single *LMO2* SNPs with overall survival in DLBCL, NCI-SEER DLBCL Survival Study.

dbSNP ID	Position	Location	Alleles (wild type/variant)	MAF	Per copy of variant allele			One copy of variant allele			Two copies of variant allele			p-trend [†]	q-value [§]
					HR*	95% CI	HR*	95% CI	HR*	95% CI	HR*	95% CI	HR*		
rs3740617	33837592	exon	C/T	0.37	1.00	0.67–1.48	1.14	0.62–2.05	0.92	0.38–2.18			0.99	0.99	
rs10836123	33839045	intron	T/G	0.17	0.99	0.57–1.72	1.12	0.61–2.01					0.96	0.99	
rs3781575	33841895	intron	T/C	0.18	1.08	0.66–1.78	1.19	0.67–2.10	0.71	0.09–5.20			0.75	0.96	
rs3824848	33843123	intron	C/T	0.34	1.04	0.71–1.52	1.19	0.66–2.13	0.99	0.42–2.31			0.83	0.97	
rs4007	33844118	intron	A/T	0.20	0.87	0.53–1.42	0.87	0.48–1.56	0.76	0.17–3.25			0.58	0.96	
rs3758638	33856532	intron	C/A	0.25	0.89	0.56–1.43	0.93	0.52–1.63	0.71	0.16–2.97			0.64	0.96	
rs1885524	33860485	intron	C/T	0.51	1.33	0.90–1.95	2.29	1.00–5.21	2.06	0.82–5.10			0.15	0.50	
rs746481	33861947	intron	T/C	0.24	0.92	0.57–1.48	0.93	0.52–1.64	0.82	0.19–3.44			0.74	0.96	
rs10836126	33862777	intron	C/T	0.46	1.07	0.74–1.56	1.38	0.72–2.62	1.10	0.48–2.47			0.72	0.96	
rs3781578	33863402	intron	G/A	0.22	1.07	0.69–1.65	1.79	0.99–3.22	0.30	0.04–2.21			0.77	0.96	
rs10836127	33864374	intron	C/A	0.18	2.18	1.39–3.43	2.43	1.33–4.42	3.93	1.17–13.1			0.001	0.010	
rs4756077	33864531	intron	G/T	0.29	0.66	0.43–1.02	0.72	0.40–1.27	0.36	0.11–1.20			0.061	0.41	
rs911817	33865123	intron	T/C	0.29	0.74	0.47–1.16	0.98	0.56–1.71	0.18	0.02–1.32			0.19	0.54	
rs10836129	33869024	intron	T/G	0.41	1.10	0.75–1.62	1.07	0.57–1.99	1.21	0.55–2.64			0.64	0.96	
rs3758640	33871004	promoter	A/G	0.39	1.03	0.70–1.52	1.50	0.81–2.74	0.85	0.33–2.18			0.87	0.97	
rs3758642	33872467	promoter	A/G	0.45	1.08	0.75–1.56	1.54	0.82–2.89	1.07	0.46–2.43			0.68	0.96	
rs750781	33874042	promoter	T/A	0.26	1.41	0.95–2.09	1.82	1.02–3.22	1.54	0.57–4.10			0.087	0.44	
rs941940	33881363	promoter	T/C	0.46	0.73	0.48–1.10	0.83	0.43–1.57	0.51	0.21–1.21			0.13	0.50	
rs7941248	33882551	promoter	C/T	0.31	0.81	0.53–1.24	0.81	0.45–1.44	0.65	0.22–1.85			0.33	0.83	
rs10128650	33888627	promoter	A/C	0.06	2.89	1.72–4.87	4.35	2.17–8.70	3.22	0.43–24.0			0.00007	0.0014	

* Adjusted for the clinical risk score.

[†] P-value from the ordinal trend test in the Cox model, adjusted for clinical risk score.

[§] The q-value is the expected proportion of false positives among the significant results that had the same or smaller q-value for the single SNP analysis; a q<0.05 is considered noteworthy.

Table II

Association of *LMO2* SNPs with *LMO2* IHC status, NCI-SEER DLBCL Survival Study.

Association with LMO2 IHC Staining										
SNP	Position	Location	Alleles (wild type/variant)	Wild type/Wild type		Wild type/Variant		Variant/Variant		χ ² p-value [†]
				N	% Positive	N	% Positive	N	% Positive	
rs3740617	33837592	exon (K121K)	C/T	63	43.0%	76	48.7%	22	54.6%	0.79
rs10836123	33839045	intron	T/G	110	49.1%	49	46.9%	3	33.3%	0.85
rs3781575	33841895	intron	T/C	106	48.1%	51	29.0%	4	50.0%	0.99
rs3824848	33843123	intron	C/T	77	55.8%	61	41.0%	24	41.7%	0.18
rs4007	33844118	intron	A/T	101	43.6%	54	55.6%	5	60.0%	0.31
rs3758638	33856532	intron	C/A	91	45.1%	60	51.7%	10	60.0%	0.55
rs1885524	33860485	intron	C/T	37	56.8%	85	44.7%	40	47.5%	0.47
rs746481	33861947	intron	T/C	91	50.6%	62	45.2%	8	50.0%	0.80
rs10836126	33862777	intron	C/T	47	53.2%	79	43.0%	35	51.4%	0.48
rs3781578	33863402	intron	G/A	99	50.5%	53	39.6%	9	66.7%	0.22
rs10836127	33864374	intron	C/A	109	47.7%	48	52.1%	5	20.0%	0.39
rs4756077	33864531	intron	G/T	82	45.1%	66	50.0%	14	57.1%	0.66
rs911817	33865123	intron	T/C	77	42.9%	72	50.0%	11	72.7%	0.16
rs10836129	33869024	intron	T/G	57	50.9%	76	48.7%	29	41.4%	0.71
rs3758640	33871004	promoter	A/G	59	47.5%	77	45.5%	25	56.0%	0.66
rs3758642	33872467	promoter	A/G	51	54.9%	76	43.4%	34	47.1%	0.44
rs750781	33874042	promoter	T/A	90	46.7%	60	50.0%	12	50.0%	0.91
rs941940	33881363	promoter	T/C	45	48.9%	84	40.5%	32	68.8%	0.02
rs7941248	33882551	promoter	C/T	77	36.4%	68	61.8%	17	47.1%	0.009
rs10128650	33888627	promoter	A/C	142	50.0%	18	38.9%	1	0.0%	0.42

[†] p-value from χ^2 test for a global association of genotype with *LMO2* IHC expression.

Table III

Concordance index at 8 years for selected SNPs from a time-dependent receiver-operator curve (AUC) model.

SNP	C-statistic for Model			
	SNP(s) alone	SNP(s) + LMO2 IHC	SNPs + Clinical Risk Score	SNP(s) + LMO2 IHC + Clinical Risk Score
rs1885524	0.555	0.601	0.683	0.698
rs10836127	0.569	0.609	0.713	0.724
rs750781	0.579	0.617	0.686	0.699
rs941940	0.525	0.580	0.691	0.702
All 4 SNPs	0.672	0.683	0.751	0.754