



Published in final edited form as:

Leukemia. 2009 March ; 23(3): 596–602. doi:10.1038/leu.2008.240.

Polymorphisms in DNA Repair and One-Carbon Metabolism Genes and Overall Survival in Diffuse Large B-Cell Lymphoma (DLBCL) and Follicular Lymphoma

Sophia S Wang, PhD^{1,*}, Matthew J Maurer, MS², Lindsay M Morton, PhD¹, Thomas M Habermann, MD², Scott Davis, PhD³, Wendy Cozen, MD⁴, Charles F Lynch, MD, PhD⁵, Richard K Severson, PhD⁶, Nathaniel Rothman, MD¹, Stephen J Chanock, MD¹, Patricia Hartge, ScD¹, and James R Cerhan, MD, PhD²

¹Division of Cancer Epidemiology and Genetics, National Cancer Institute, Rockville, MD, United States

²Department of Health Sciences Research, Mayo Clinic College of Medicine, Rochester, MN, United States

³Fred Hutchinson Cancer Research Center, Seattle, WA, United States

⁴Department of Preventive Medicine, University of Southern California, Los Angeles, CA, United States

⁵Department of Epidemiology, University of Iowa, Iowa City, IA, United States

⁶Department of Family Medicine and Public Health Sciences and Karmanos Cancer Institute, Detroit, MI, United States

LETTER TO THE EDITOR

Diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL) are the two most commonly diagnosed subtypes of non-Hodgkin lymphoma (NHL) in the United States (1). Both diseases show variable progression and identifying prognostic markers applicable in a clinical setting could greatly improve patient management. Established adverse prognostic factors in DLBCL as delineated in the International Prognostic Factor Index (IPI) include older age, higher stage, poor performance score, extranodal involvement, and above normal lactate dehydrogenase (LDH). For FL, adverse prognostic factors are combined in the Follicular Lymphoma International Prognostic Index (FLIPI) and include older age, higher stage, lower hemoglobin level, more nodal areas, and above normal LDH. Because both IPI and FLIPI predict risk incompletely, it is thought that molecular characteristics of the tumor and its microenvironment may improve predicting DLBCL and FL survival. Notably, recent gene expression studies suggest a molecular profile of DLBCL with poorer survival (2).

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

Correspondence: Sophia S. Wang, PhD, Division of Cancer Epidemiology and Genetics, National Cancer Institute, 6120 Executive Blvd, Room 5104, Rockville, MD 20852-7234, U.S.A., Phone: (301) 402-5374, FAX: (301) 402-0916, wangso@mail.nih.gov.

Supplementary Information accompanies the paper on the Leukemia website: <http://www.nature.com/leu>

New prognostic markers for FL have also been suggested (3), but neither have yet been translated to the clinical setting. Evaluating host germline genetic variability offers another promising approach for refining prognosis of DLBCL and FL.

DNA repair and one-carbon metabolism pathways play important roles in maintaining DNA and chromosomal integrity. Briefly, chromosomal translocations, a hallmark of non-Hodgkin lymphoma (NHL), results from V(D)J recombination or persistent, unrepaired DNA double strand breakage. A host's ability to repair DNA damage reflects mutations and variations in a large suite of genes that affect non-homologous end joining, homologous recombinational repair, nucleotide excision repair, base excision repair, and direct damage reversal. DNA damage can also result indirectly when one-carbon metabolism is disrupted from nutrient deficiency (e.g., folate, methionine) or from genetic variation in one-carbon metabolism genes. Disruptions within one-carbon metabolism have been linked to insufficient DNA synthesis/repair and subsequent aberrant gene expression.

We hypothesized that DNA repair and one-carbon metabolism pathways may be related to NHL survival by affecting the host's sensitivity to DNA damage and ability to repair DNA damage that results either in response to treatment (e.g., radiation) or interaction with treatment (e.g., chemotherapy). Here, we evaluated the association between 39 genetic variations among 19 DNA repair genes and 27 SNPs in 17 one-carbon metabolism genes (Table 1) as related to overall survival (e.g., risk of dying) from DLBCL and FL, according to those SNPs. Our evaluation was based on population-based cases recruited as part of a large epidemiologic population-based case-control study of NHL in the United States (NCI-SEER NHL case-control study).

Full details on this prognosis study have been described previously (4). Briefly, histologically-confirmed NHL patients aged 20–74 years in four Surveillance, Epidemiology, and End Results (SEER) cancer registries (Detroit; Washington; Iowa; Los Angeles) were enrolled from July 1998 through June 2000. Histology was coded initially according to the International Classification of Diseases-Oncology (ICD-O), 2nd Edition (5), and later updated to the 3rd Edition by each registry. For the present analysis, we included the 215 cases of DLBCL and 192 cases of FL with DNA extracted from blood. Follow-up of cases for overall survival was conducted. Date of diagnosis, histology, stage, presence of B-symptoms, first course of therapy, date of last follow-up, and vital status were derived from registry databases at each study site in the spring of 2005. 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 (observation); information on individual agents and doses was not available. Age, sex, race, Hispanic ethnicity, and education level were derived from patient interviews as part of the case-control study. Written, informed consent was provided by all participants.

The median age at diagnosis for both DLBCL and FL patients was 57 years (range, 20–74). Males comprised 54% and 51% of DLBCL and FL, respectively. 91% of DLBCL and FL patients identified themselves as non-Hispanic Caucasians. 59% of DLBCL and 40% of FL patients had advanced stage disease. 40% of DLBCL and 20% of FL patients had B-symptoms. Chemotherapy-based regimen was the most common initial therapy, followed by

observation and radiation only. With a median follow-up of 57 months (range, 27–78 months), 50 (23%) DLBCL and 40 (21%) FL patients died.

Details on genotyping have been published previously; SNPs were previously selected based on prior functional data or expected functional consequences (6, 7). All genotyping was conducted at the National Cancer Institute Core Genotyping Facility using the Taqman or EPOCH platforms (<http://cgf.nci.nih.gov>). Sequence data and assay conditions are provided at <http://snp500cancer.nci.nih.gov> (8). We used Cox proportional hazards regression (Cox) to estimate hazard ratios (HR) and 95% confidence intervals (95% CI) for the association of each individual genotype with overall survival, adjusting for age, clinical and demographic factors. Age was modeled according to the standard IPI score as <60 versus 60+ years. A clinical and demographic risk score was created using a linear combination of the clinical and demographic variables and their multivariable Cox model parameter estimates. The clinical variables used in the risk score included 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); demographic variables included sex, race (white, other), study center (Detroit, Iowa, Los Angeles, Seattle), and education (<12, 12–15, 16+ years).

We identified polymorphisms in four DNA repair genes associated with DLBCL survival. Specifically, after adjusting for demographic and clinical variables, SNPs in *BRCA1* (rs16942; HR_{AA}=2.53, 95% CI=1.41–4.55), *XRCC4* (rs1056503; HR_{TT}=2.70, 95% CI=1.26–5.75), *XRCC2* (rs3218536; HR_{AG/AA}=2.07, 95% CI=1.05–4.05), and *ERCC2* (rs13181; HR_{CC}=2.02, 95% CI=1.03–3.96) were associated with poorer overall survival for DLBCL. No individual DNA repair gene polymorphisms were associated with overall survival for FL, after adjusting for clinical and demographic factors.

We identified polymorphisms in three one-carbon metabolism genes associated with DLBCL overall survival. After adjusting for demographic and clinical variables, SNPs in *SHMT1* (rs1979276; HR_{CT/TT}=2.47, 1.31–4.67), *BHMT* (rs585800; HR_{AT/TT}=2.02, 1.16–3.54), and *TCN1* (rs526934; HR_{TT}=1.86, 1.04–3.33) were associated with DLBCL survival. In FL, *FTHFD* (rs1127717) AG/GG, *MTHFR* (rs1801131) AC/CC, and *GGH* (rs719235) GT/TT genotypes were associated with overall survival with hazard ratios of 1.99 (95% CI=1.07–3.7), 2.00 (95% CI=1.04–3.84), and 2.49 (95% CI=1.21–5.14), respectively.

We evaluated all possible SNP combinations, summing the number of deleterious genotypes in each multi-SNP combination, and fitting a Cox model adjusting for age and the clinical and demographic risk scores (demographic and clinic risk score was categorized into tertiles, with values of 0–2 (low to high risk)). The models were stratified by number of SNPs included in the SNP score variable (i.e., 1 SNP, 2 SNPs,...) and ranked by likelihood. We ran 1,000 bootstrap stepwise selection Cox models (adjusting for age, clinical and demographic risk scores) using the selected SNPs, and calculated the percentage of models that each SNP was included. We plotted the likelihood to determine the number of SNPs where the reduction in the likelihood became marginal. In parallel, we used results from the bootstrap analysis and from the best-fitting multi-SNP risk score model to develop a final multi-SNP risk score.

We assessed the association of each multi-SNP risk score with overall survival using Kaplan-Meier curves, Cox proportional hazards models, and time-dependent receiver-operator characteristics (ROC) methodology for censored data. We also developed a single risk score that combined the number of deleterious genotypes from the multi-SNP models (0-n) with a combined clinical and demographic risk score (0–2) to create a single SNP and clinical/demographic risk score (0-n).

We selected a three-SNP model for DLBCL that included *BRCA1* (rs16942), *XRCC4* (rs1056503), and *ERCC2* (rs13181). The reduction in the likelihood became marginal after three SNPs and the three selected comprised the best-fitting multi-SNP score model. The number of deleterious genotypes (as previously defined by the single-SNP results) was summed from these 3 SNPs (0–3), and this score was strongly associated with survival in both univariate ($p=0.0002$) and multivariate ($p=0.0001$) analyses. To demonstrate the independent effect of genotype on NHL OS, Figure 1a demonstrates that there was a clear gradient in risk (Figure 1a) (HR=2.27, 95% CI=1.57–3.27) for each additional deleterious genotype in the multivariate model, adjusted for demographic and clinical variables. Our combined score of deleterious genotypes (0–3) with the clinical and demographic risk score (0–2) into a single summary SNP and clinical risk score (0–5) was further strongly associated with survival ($p=1.3 \times 10^{-7}$). Again, a clear gradient in DLBCL mortality risk was evident (Figure 1b) (HR=2.23, 95% CI=1.69–2.95). We note that p-values for Kaplan Meier curves as shown in Figure 1 are also statistically.

Among one-carbon metabolism genes, we also selected a three SNP model for DLBCL that included *BHMT* (rs585800), *SHMT1* (rs1979276), and *TCN1* (rs526934)). The summary score (0–3) was strongly associated with survival in both univariate ($p = 0.0002$) and multivariate analyses ($p = 1.7 \times 10^{-5}$). There was a gradient in risk (Figure 2a) with each additional deleterious genotype (HR=2.55, 95% CI=1.74–3.76) in the multivariate model. Our score of deleterious genotypes (0–3) and clinical and demographic risk score (0–2) combined into a single summary SNP and clinical risk score (0–5) was further strongly associated with survival ($p = 2.1 \times 10^{-9}$). Risk of dying from DLBCL also increased with severity of score (HR=2.44, (95% CI = 1.82–3.29), $p=2.1 \times 10^{-9}$) (Figure 2b).

We also selected three one-carbon metabolism SNPs for FL: *MTHFR* (rs1801131), *FTHFD* (rs1127717) and *GGH* (rs719235). There was a gradient in mortality risk (Figure 2c) with a 2.14-fold increased risk of dying from FL with each additional deleterious genotype (95% CI=1.44–3.16) in the multivariate model ($p=0.0002$). Our summary SNP and clinical risk score (0–5) was further strongly associated with survival ($p=1.9 \times 10^{-6}$). For each severity in score (Figure 2d), there was a 2.06-fold increase in risk of dying from FL (95% CI=1.55–2.73).

We conducted a time-dependent ROC analysis for censored data to measure the prognostic capacity of our survival model based on the clinical and demographic risk score with and without genotype information, as measured by the area under the curve (AUC). For DLBCL survival, the DNA repair and one-carbon metabolism gene models' combined risk scores had a concordance index (AUC) at 5 years of follow-up of 0.73 (Figure 3a). For FL survival, the one-carbon metabolism gene model combined risk score had an AUC at 5 years of

follow-up of 0.72 (Figure 3b). Combined DNA repair and one-carbon metabolism gene AUC for DLBCL was 0.78 (Figure 3a). These AUCs approach the predictive range needed for clinically useful tests.

Finally, to ensure the validity of our multi-SNP models results, we also conducted a permutation analysis to evaluate the statistical significance of our results compared to chance. Specifically, we applied our model building approach to each of the 200 permuted datasets; the significance of the best model for each dataset was recorded. For DLBCL, both our three-SNP models for DNA repair genes and one-carbon metabolism genes from the true data outperformed the best model generated from >95% of the permuted random datasets, suggesting that our multi-SNP model has a high level of significance. Our 3-SNP model for FL and one-carbon metabolism genes from the true data performed better than the best model generated from >90% of the permuted random datasets.

The potential relevance of DNA repair genes in cancer survival has been reported in other tumors. The *ERRC2* K751Q and *XRCC2* R188H polymorphisms have both been reported to be associated with autologous stem cell transplantation failure in multiple myeloma patients. The *XRCC2* R188H polymorphism was also associated with breast cancer and pancreatic cancer survival. Of note, genes we previously identified as relevant for etiology (6) were not predictive for survival, potentially suggesting distinct roles for these gene variations in etiology compared to survival. While we infer based on our results that homologous and nucleotide-excision repair may be relevant for DLBCL survival, we cannot exclude the potential relevance of other DNA repair pathways as we did not comprehensively evaluate all DNA repair genes and their variations. Of the one-carbon metabolism genes identified as relevant for DLBCL or follicular overall survival, no other reports have implicated *BHMT*, *TCN1*, *FTHFD* or *SHMT* with any cancer but numerous reports associate *MTHFR* polymorphisms with survival in NHL and cancers of the breast, colon, bladder, and stomach.

An important strength of our study was the population-based ascertainment of newly diagnosed cases making our results relevant for community-based treatment programs. Study limitations include the lack of detailed data on prognostic factors or treatment. Although our clinical and demographic risk score predicted survival with AUC's for DLBCL and FL comparable to their respective IPI's (4), future studies that incorporate Ann Arbor and IPI classifications will be important for translation of confirmed gene associations to clinical applications or any risk prediction model.

In summary, we show that host genetic variations in DNA repair and one-carbon metabolism gene polymorphisms can stratify risk for overall survival in DLBCL and FL after accounting for demographic and clinical factors. A detailed evaluation of genes in these two pathways and their interaction with dietary nutrients and specific therapeutic agents may be informative. Our results require replication in an independent population and further evaluation among cases treated with Rituximab. Understanding the functional relevance of confirmed gene variations and their influence on therapy response should also be pursued.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Peter Hui and Cristine Allmer for data management, and Laura Huennekens for secretarial support. We also thank Drs. Qing Lan and Deirdre Hill for their scientific input for selection of one-carbon metabolism and DNA repair genes, respectively, as hypothesized for NHL etiology. This work was supported by National Cancer Institute grants R01-CA96704 and P50-CA97274; the NCI Intramural Program; SEER contracts N01-PC-67010, N01-PC-67008, N01-PC-67009, N01-PC-65064, and N02-PC-71105.

References

1. 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 Jan 1; 107(1):265–276. [PubMed: 16150940]
2. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, Fisher RI, et al. The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. *N Engl J Med*. 2002 Jun 20; 346(25):1937–1947. [PubMed: 12075054]
3. Dave SS, Wright G, Tan B, Rosenwald A, Gascoyne RD, Chan WC, et al. Prediction of survival in follicular lymphoma based on molecular features of tumor-infiltrating immune cells. *N Engl J Med*. 2004 Nov 18; 351(21):2159–2169. [PubMed: 15548776]
4. Cerhan JR, Wang S, Maurer MJ, Ansell SM, Geyer SM, Cozen W, et al. Prognostic significance of host immune gene polymorphisms in follicular lymphoma survival. *Blood*. 2007 Jun 15; 109(12):5439–5446. [PubMed: 17327408]
5. Percy, C.; Van Holten, V.; Mui, C., editors. *International Classification of Diseases for Oncology*. 2nd ed.. Geneva: World Health Organization; 1990.
6. Hill DA, Wang SS, Cerhan JR, Davis S, Cozen W, Severson RK, et al. Risk of non-Hodgkin lymphoma (NHL) in relation to germline variation in DNA repair and related genes. *Blood*. 2006 Nov 1; 108(9):3161–3167. [PubMed: 16857995]
7. Lim U, Wang SS, Hartge P, Cozen W, Kelemen LE, Chanock S, et al. Gene-nutrient interactions among determinants of folate and one-carbon metabolism on the risk of non-Hodgkin lymphoma: NCI-SEER case-control study. *Blood*. 2007 Apr 1; 109(7):3050–3059. [PubMed: 17119116]
8. Packer BR, Yeager M, Burdett L, Welch R, Beerman M, Qi L, et al. SNP500Cancer: a public resource for sequence validation, assay development, and frequency analysis for genetic variation in candidate genes. *Nucleic Acids Res*. 2006 Jan 1; 34(Database issue):D617–D621. [PubMed: 16381944]

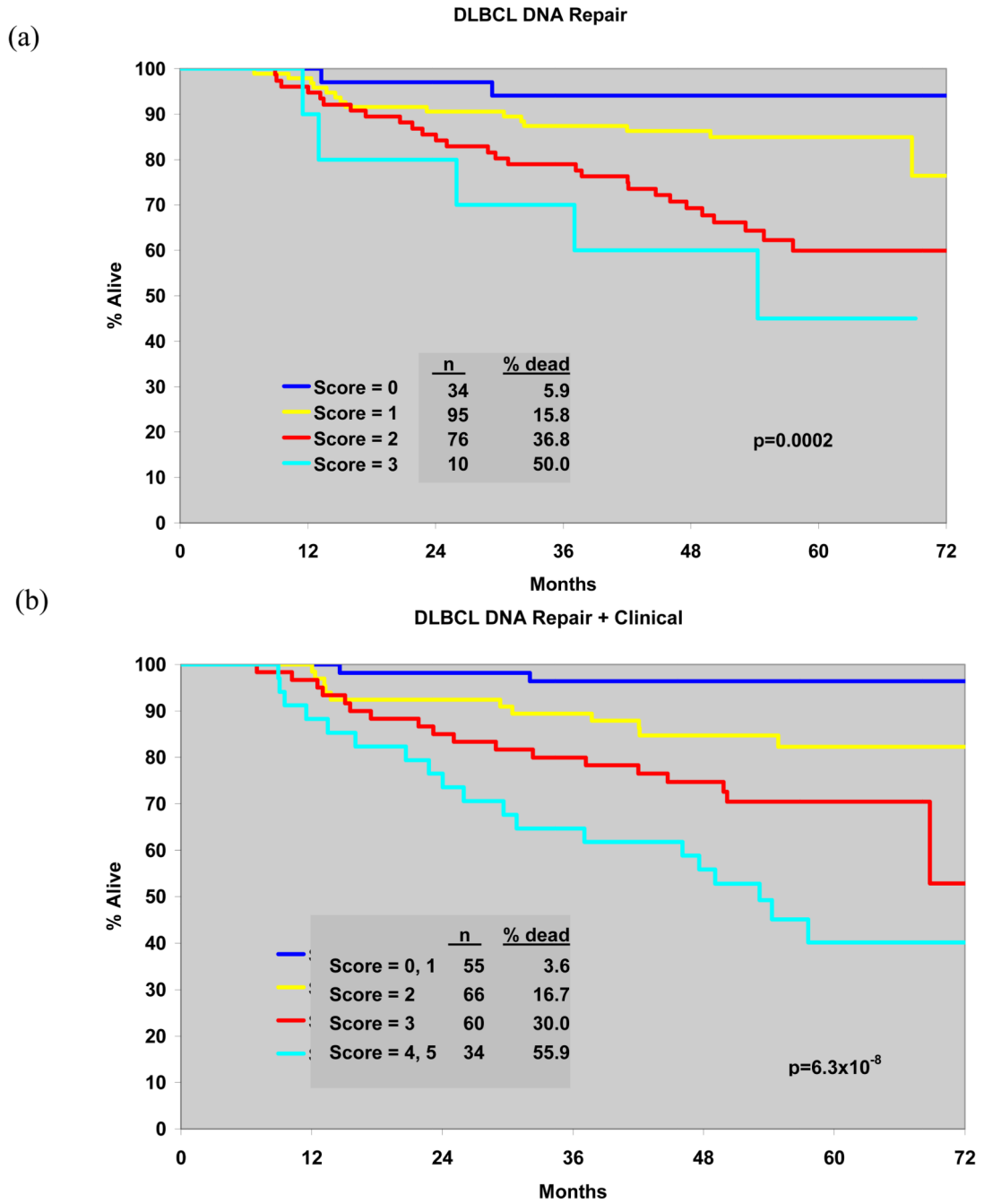


Figure 1. Kaplan-Meier curves for DNA repair genes and overall survival for diffuse large B-cell lymphoma (DLBCL) by a summary risk score of (a) the number of deleterious DNA repair genotypes from the final three-gene SNP risk score (0–3) and (b) from the final score for summary SNP and clinical and demographic factors (0–5). *Note:* p-values are provided for Kaplan Meier curves; Cox regression p-values are provided in manuscript text.

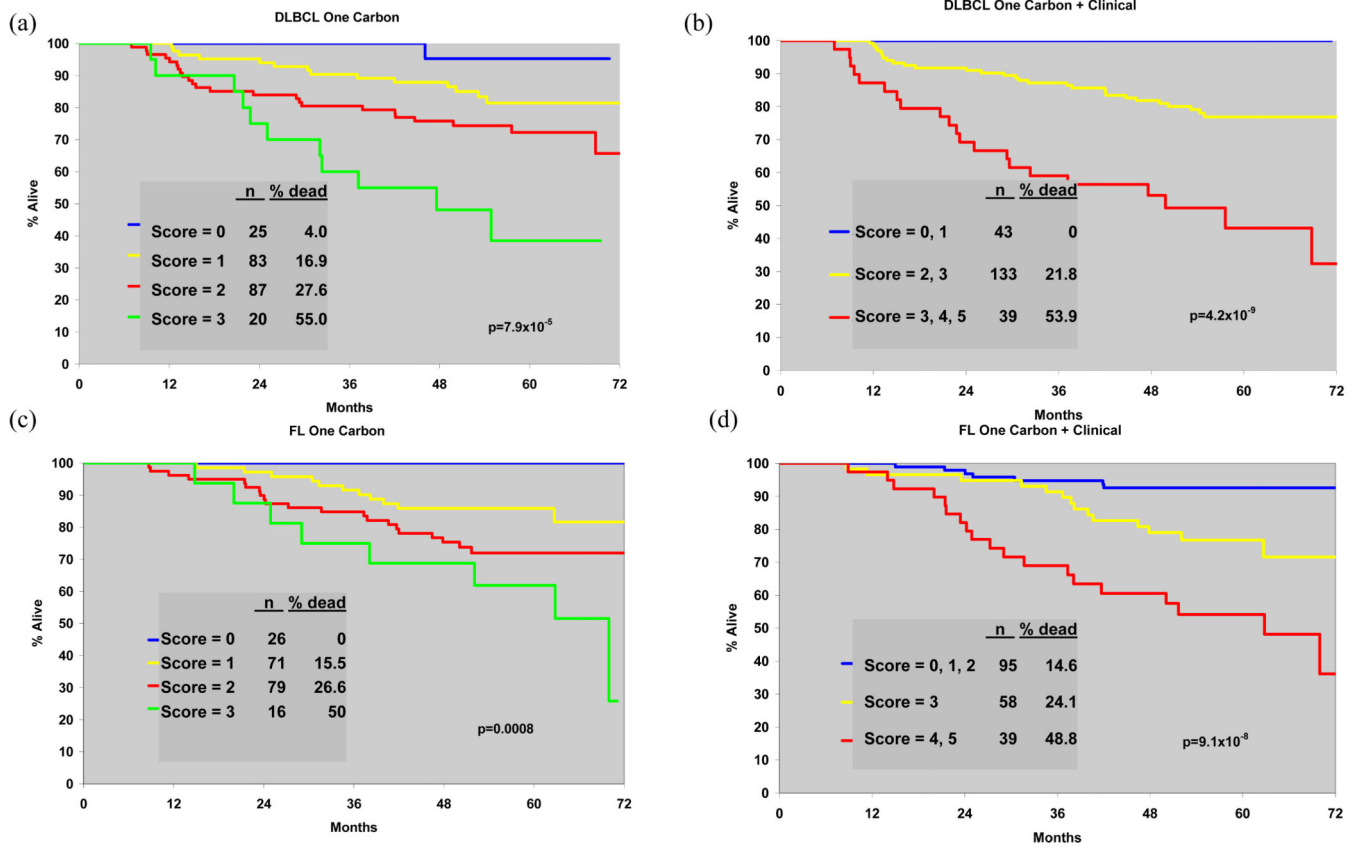


Figure 2. Kaplan-Meier curves for one-carbon metabolism genes and overall survival for diffuse large B-cell lymphoma (DLBCL) (a–b) and follicular lymphoma (FL) (c–d) by a summary risk score of deleterious DNA repair genotypes (0–3) and from the final score for summary SNP and clinical and demographic factors (0–5). *Note:* p-values are provided for Kaplan Meier curves; Cox regression p-values are provided in manuscript text.

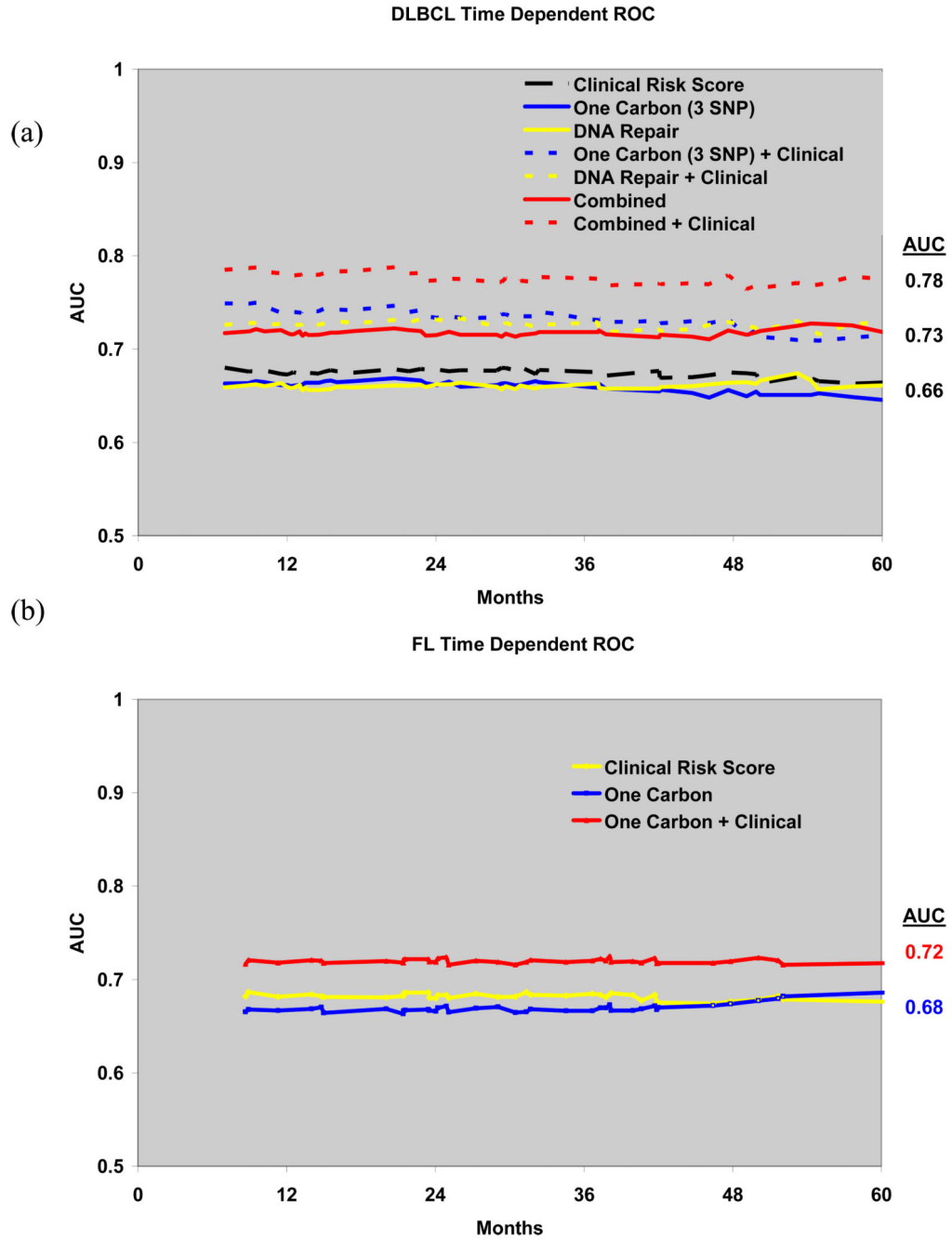


Figure 3. Time-dependent receiver-operator (ROC) analysis using the NCI-SEER dataset (clinical and demographic risk scores, final SNP risk scores, combined SNP and clinical and demographic risk scores) for (a) DNA repair and one-carbon metabolism SNP models in diffuse large B-cell lymphoma (DLBCL) and (b) one-carbon metabolism SNP model in follicular lymphoma (FL). “Combined” refers to scores derived from a single model that includes both one-carbon metabolism and DNA repair SNPs.

Table 1

DNA repair and one-carbon metabolism gene variants examined in relation to overall survival for diffuse large B-cell lymphoma (DLBCL) and follicular lymphoma (FL).

Gene Abbreviations	Name	Chromosome Location	SNP database ID	Amino acid change
DNA Repair Genes				
<i>APEX1</i>	Apurinic/apyrimidinic Endonuclease	14q11.2	rs3136820	D148E
<i>BRCA1</i>	Breast Cancer 1, early onset	17q21	rs16940 rs799917 rs16941 rs16942 rs1799966	
<i>BRCA2</i>	Breast Cancer 2, early onset	13q12.3	rs144848 rs1801406 rs1799955 rs15869 rs766173 rs1799944 rs543304	N372H K1132 S2414 Ex27–336A>C N289H N991D
<i>ERCC1</i>	Excision repair cross-complementing 1	19q13.2	rs3212961	IVS5+33A>C
<i>ERCC2</i>	Excision repair cross-complementing 2	19q13.3	rs1799793 rs13181	D312N K751Q
<i>ERCC4</i>	Excision repair cross-complementing 4	16p13.3	rs1799802	P379S
<i>ERCC5</i>	Excision repair cross-complementing 5	13q22	rs17655	D1104H
<i>LIG4</i>	DNA Ligase IV	13q33	rs1805388	T91
<i>MGMT</i>	O-6-methylguanine DNA-methyltransferase	10q26	rs2308321 rs2308327 rs12917	I143V K178R L84F
<i>NBS1</i>	Nijmegen Breakage Syndrome	8q21	rs1805794	E185Q
<i>RAD23B</i>	Rad23, homolog of yeast, B	9	rs1805329	A249V
<i>RAG1*</i>	Recombination-Activating Gene 1	11p13	rs2227973	K820R
<i>TP53</i>	Tumor protein p53	17p13.1	rs1042522	P72R
<i>WRN</i>	Werner Syndrome	8p12	rs4987236 rs1800391 rs2725362 rs1346044	V114I M387I L1074F C1367R
<i>XPC</i>	Xeroderma Pigmentosum C	3p25	rs2228001	K939Q
<i>XRCC1</i>	X-Ray Repair Cross-Complementing 1	19q13.2	rs25487 rs25489 rs1799782	Q399R R280H R194W
<i>XRCC2</i>	X-Ray Repair Cross-Complementing 2	7q36.1	rs3218536	R188H
<i>XRCC3</i>	X-Ray Repair Cross-Complementing 3	14q32.3	rs861539	T241M
<i>XRCC4</i>	X-Ray Repair Cross-Complementing 4	5q13	rs1805377 rs1056503 rs3734091	N298S S307 A247S
One-carbon metabolism genes				
<i>BHMT</i>	Betaine-Homocysteine MethylTransferase	5q13.1-q15	rs585800	Ex8+453A>T
<i>CBS</i>	Cystathionine-Beta-Synthase	21q22.3	rs234706 rs12613 rs1801181	Y233Y Ex18–391G>A A360A
<i>FPGS</i>	FolylPolyGlutamate Synthase	9q34.1	rs10106	Ex15–263T>C

Gene Abbreviations	Name	Chromosome Location	SNP database ID	Amino acid change
DNA Repair Genes				
<i>FTHFD</i>	10-Formyl-TetraHydroFolate Dehydrogenase	3q21.3	rs2305230 rs1127717	L395L D793G
<i>GGH</i>	Gamma-Glutamyl Hydrolase	8q12.1	rs719236 rs1031552	-423G>T IVS7-3001C>T
<i>MBD2</i>	Methyl-CpG Binding Domain protein 2	18q21	rs7614 rs603097	Ex8+438A>G -2176T>C
<i>MLH1</i>	MutL Homolog 1	3p21.3	rs1799977 rs2286940	I219V -2176T>C
<i>MSH2</i>	MutS Homolog 2	2p22-p21	rs4987188	G322D
<i>MTHFD2</i>	Methylene-TetraHydroFolate Dehydrogenase	2p13.1	rs1667627	IVS1+3323T>C
<i>MTHFR</i>	(5,10-) Methylene-TetraHydroFolate Reductase	1p36.3	rs1801131 rs1801133	E429A A222V
<i>MTHFS</i>	(5,10-) Methenyl-TetraHydroFolate Synthetase	15q24.3	rs622506	IVS2-1411T>G
<i>MTR</i>	Methionine synthase	1q43	rs1805087	D919G
<i>MTRR</i>	Methionine synthase reductase	5p15.3-15.2	rs161870	L206L
<i>SHMT1</i>	Cytoplasmic Serine Hydroxy-MethylTransferase	17p11.2	rs1979277 rs1979276	L435F Ex12+236C>T
<i>SLC19A1</i>	Solute Carrier family 19 (folate transporter), member 1	21q22.3	rs12659 rs1051296	P232P Ex8-233T>G
<i>TCN1</i>	Transcobalamin I	11q11-q12	rs526934	IVS1+372T>C
<i>TYMS</i>	Thymidylate Synthase	18p11.32	rs699517 VNTR (Ex1+52-28base)	Ex8-233T>G 3R>2R