Single nucleotide variation in the $TP53$ 3′ untranslated region in diffuse large B-cell lymphoma treated with rituximab-CHOP: a report from the International DLBCL Rituximab-CHOP Consortium Program

Yong Li,1, 2¶ Michael W. Gordon,1 Zijun Y. Xu-Monette,2 Carlo Visco,3 Alexander Tzankov,4 Dehui Zou,5 Lugui Qiu,5 Santiago Montes-Moreno,6 Karen Dybkaer,7 Attilio Orazi,8 Youli Zu,9 Govind Bhagat,10 Kristy L. Richards,11 Eric D. Hsi,12 William WL. Choi,13 J. Han van Krieken,14 Qin Huang,15 Weiyun Ai,16 Maurilio Ponzoni,17 Andrés JM. Ferreri,17 Jane N. Winter,18 Ronald S. Go,19 Miguel A. Piris,6 Michael B. Møller,20 Lin Wu,21 Michael Wang,22 Kenneth S. Ramos,1 L. Jeffrey Medeiros,2 and Ken H. Young2¶

1Department of Biochemistry and Molecular Biology, University of Louisville School of Medicine, Brown Cancer Center, Louisville, KY, USA; 2Department of Hematopathology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA; 3San Bortolo Hospital, Vicenza, Italy; 4University Hospital, Basel, Switzerland; 5Institute of Hematology, CAMS & PUMC, 288 Nanjing Road, Tianjin, China; 6Hospital Universitario Marques de Valdecilla, Santander, Spain; 7Aalborg Hospital, Aarhus University Hospital, Aalborg, Denmark; 8Weill Medical College of Cornell University, New York, NY, USA; 9The Methodist Hospital, Houston, TX, USA; 10Columbia University School of Medicine, New York, NY, USA; 11University of North Carolina School of Medicine, Chapel Hill, NC, USA; 12Cleveland Clinic, Cleveland, OH, USA; 13University of Hong Kong Li Ka Shing Faculty of Medicine, Hong Kong, China; 14Radboud University Nijmegen Medical Centre, Nijmegen, Netherlands; 15City of Hope National Medical Center, Duart, CA, USA; 16University of California San Francisco School of Medicine, San Francisco, CA, USA; 17San Raffaele H. Scientific Institute, Milan, Italy; 18Feinberg School of Medicine, Northwestern University, Chicago, IL, USA; 19Gundersen Lutheran Health System, La Crosse, WI, USA; 20Odense University Hospital, Odense, Denmark; 21Roche Molecular Systems, Inc., Pleasanton, CA, USA; 22Department of Lymphoma and Myeloma, The University of Texas MD Anderson Cancer Center, Houston, TX, USA

Key Words: diffuse large B-cell lymphoma, p53, 3′untranslated region, noncoding polymorphisms

Running Title: $TP53$ 3′UTR Variation in DLBCL

¶ Correspondence:

Yong Li, PhD
Department of Biochemistry and Molecular Biology
School of Medicine, University of Louisville
319 Abraham Flexner Way, Louisville, KY 40202
Tel: 502-852-7551
Fax: 502-852-6222
Email: yong.li@louisville.edu
Ken H. Young, MD, PhD
Department of Hematopathology, Unit 72
The University of Texas MD Anderson Cancer Center
1515 Holcombe Boulevard
Houston, Texas 77230-1439
Phone: 713-745-2598
Pager: 713-404-8460
Fax: 713-792-7273
Email: khyoung@mdanderson.org
Key Point: TP53 3'UTR variations demonstrate prognostic value in DLBCL

Abstract
We identified multiple single nucleotide variants (SNVs) in the TP53 3' untranslated region (3'UTR) in tumor specimens from 244 patients with diffuse large B-cell lymphoma (DLBCL). Patients carrying a wild-type TP53 coding sequence (CDS) and one or more 3'UTR SNV had better 5-year survival rate than patients carrying a wild-type CDS and the reference 3'UTR, yet there is no statistically significance difference in overall survival (OS). In contrast, 3'UTR variation predicted poorer OS for patients with a mutant TP53 CDS. We then sequenced the TP53 3'UTR in 247 additional DLBCL patients as a validation set. Altogether, we identified 187 novel SNVs; 36 occurred at least twice. Most of the newly identified 3'UTR SNVs were located at sites that are complementary to seed sequences of microRNAs that are predicted or experimentally known to target TP53. Three SNVs disrupt the seed match between miR-125b and the TP53 3'UTR thereby impeding suppression of p53 by this microRNA. In addition, a germline SNV (rs78378222) located in the TP53 polyadenylation signal resulted in down-regulation of both p53 mRNA and protein levels and reduction of cellular apoptosis. This study is the first to demonstrate the prognostic value of the TP53 3'UTR in cancer.
Introduction

The tumor suppressor p53, encoded by TP53, is recognized as the guardian of the human genome. p53 responds to diverse stressors by regulating many downstream target genes that promote cellular DNA repair, apoptosis, cell cycle arrest, and senescence. Restoring p53 function alone is sufficient to cause regression of several types of cancer in mice.\textsuperscript{1,2} This characteristic suggests that p53 itself or the p53 pathway must be derailed for tumor development, genetically or otherwise. Three distinct sets of nucleotide alterations can occur in the human TP53 gene: tumor-associated (somatic) mutations, germline Li-Fraumeni mutations, and germline p53 polymorphisms.\textsuperscript{3,4} Both tumor-associated mutations and Li-Fraumeni mutations occur in the p53 coding sequence (CDS) and produce mutant p53 proteins that lack most or all of the normal p53 functions and confer oncogenic properties \textsuperscript{5}. For instance, a missense base substitution at codon 273 of human TP53 (R273H) is frequently observed in somatic tumors and in tumors from Li-Fraumeni syndrome patients. Mouse models demonstrate that this mutant p53 allele has enhanced oncogenic potential that cannot be explained by simple loss of p53 function.\textsuperscript{6,7}

In general, p53 CDS mutations contribute to oncogenesis by two distinct mechanisms: (1) mutant alleles can have dominant negative effects on wild-type (WT) p53 function; and (2) mutant p53 protein can actively promote tumorigenesis independently of WT p53 function. It is accepted that the WT TP53 CDS produces native p53 tumor suppressor, whereas most TP53 CDS mutations produce oncogenic mutant p53. By contrast, only a few of the >200 naturally occurring germline variants of TP53 in human populations cause measurable perturbation of p53 function.\textsuperscript{3} More than 90\% of these p53 variants occur in 10 introns alleged to have no cancer-related biological consequences. None of the 20 variants in the TP53 CDS have been linked reproducibly to cancer predisposition \textsuperscript{3,4}. The best studied TP53 CDS polymorphism is codon 72 (R72P), owing to an earlier finding that human papillomavirus causes more efficient degradation of p53\textsuperscript{R72} than p53\textsuperscript{P72}.\textsuperscript{8} However, few associations between this polymorphism and various types of cancer have been found.\textsuperscript{3} Sequence variation in the untranslated regions (UTRs) of the TP53 gene has not been extensively studied.

Diffuse large B-cell lymphoma (DLBCL) is the most common type of lymphoma and is both clinically heterogeneous with subsets of patients with better or worse prognosis.\textsuperscript{9,10} Biologically, DLBCL has been known to be heterogeneous with germinal center B-cell (GCB) and activated B-cell (ABC) subtypes. Recent advancements in next-generation sequencing have further broadened our understanding of genetic lesions in DLBCL. Three DLBCL sequencing projects have provided a
wealth of information on whole genomes, exomes, and/or transcriptomes for over 100 DLBCL patient specimens, some with matched healthy tissues as controls.\textsuperscript{11-13} These studies have identified mutations in more than 100 genes not implicated previously in DLBCL pathogenesis, including genes involved in chromatin modification and immune regulation. The \textit{TP53} gene (the CDS) was found to be mutated in \textasciitilde20\% of DLBCL cases all three deep-sequencing projects. However, \textit{MLL2}\textsuperscript{11,13,11,13}, \textit{CREBBP},\textsuperscript{11} \textit{EZH2},\textsuperscript{12} \textit{BCL2},\textsuperscript{12} \textit{GNA13},\textsuperscript{13} \textit{BTG1},\textsuperscript{13} \textit{PIM1},\textsuperscript{13} and \textit{PCLO}\textsuperscript{13} were found to be mutated more frequently than the \textit{TP53} CDS in some of these studies. These findings challenge the notion that \textit{TP53} is the most commonly mutated gene in cancer\textsuperscript{14,15,14,15} at least within the context of DLBCL. In addition, noncoding genomic sequences have been largely overlooked in these sequencing projects\textsuperscript{11,12}.

In a previous study, we found that \textasciitilde20\% of DLBCL patients who received cyclophosphamide, doxorubicin, vincristine, and prednisone (CHOP) chemotherapy had \textit{TP53} CDS mutations and that these mutations adversely affected prognosis.\textsuperscript{16} To provide an in-depth analysis of DLBCL patients treated with CHOP plus the monoclonal antibody Rituximab (R-CHOP), we assembled a multi-center study designated as the DLBCL R-CHOP Collaborative Consortium. We reported that \textit{TP53} CDS mutations predicted poorer survival in 506 (largest ever cohort) R-CHOP-treated DLBCL patients with either the GCB or the ABC subtype, and that \textit{TP53} monoallelic deletion and loss of heterozygosity did not confer worse survival.\textsuperscript{17} These findings suggested that assessment of \textit{TP53} CDS mutation status can be useful for stratifying patients receiving R-CHOP treatment into distinct prognostic subsets.

In this study, we retrospectively sequenced the \textit{TP53} 5'UTR and 3'UTR in DLBCL patients to determine if \textit{TP53} 3'UTR sequence alterations occur in cancer patients and to assess their potential prognostic importance. We also evaluated the potential impact of \textit{TP53} 3'UTR sequence variation in combination with CDS mutations and we determined the functional significance of 3' UTR variants in p53 regulation.
Materials and Methods

Patients
For the initial study group, we randomly selected 244 patients with DLBCL from a total of 506 de novo DLBCL patients treated with rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisone (R-CHOP), whose TP53 CDSs had been sequenced in a previous study by our consortium.17 For the validation set, tumor specimens were also obtained from 247 DLBCL patients receiving R-CHOP treatment in the consortium program, none of whom were included in the 506 DLBCL patients in the previous p53 CDS study.17

All cases were reviewed by a group of hematopathologists (S.M.M., M.A.P., M.B.M., A.T., and K.H.Y.) and the diagnoses confirmed by World Health Organization classification criteria. The current study and material transfer agreement were approved by the institutional review board of each participating center. The overall collaboration was approved by the Institutional Review Board of The University of Texas MD Anderson Cancer Center, Houston, Texas. This study was conducted in accordance with the Declaration of Helsinki.

All of these patients have undergone molecular and phenotypic classification with available clinical data. There are no statistical differences in regard to clinical parameters (age, gender, stage, lactic dehydrogenase level, tumor size, performance status, chemotherapy regimen, survival time, and recurrence rate), GCB/ABC classification and distribution, and survival among three groups (244, 262 unselected and 247).

DNA sequencing and cell biology
Genomic DNA was extracted from formalin-fixed, paraffin-embedded tissue samples using the HighPure paraffin DNA extraction kit (Qiagen). The TP53 UTRs were sequenced bi-directionally using the Sanger method at Polymorphic DNA, Inc. (Alameda, CA). All sequences were compared with the TP53 reference sequence (NM_000546.5). Single nucleotide variants (SNVs) in the 3’UTR were grouped into single nucleotide polymorphisms (SNPs) and novel SNVs (nSNVs). Any nucleotide change listed in the dbSNP database was considered an SNP; the rest (none occurring in ≥1% of patients) were considered nSNVs.
The p53-null human lung adenocarcinoma cancer cell line H1299 was obtained from the American Type Culture Collection (Manassas, VA). For mutational analysis, the p53 expression vector pCMV-p53 (Clontech, Mountain View, CA) was modified to contain TP53 5' and 3'UTRs as well as the 2.5 kb genomic DNA downstream of the 3'UTR. The resulting constructs were designated as TP53Ext, TP53ExtC1170T, and TP53ExtA1175C, respectively. MicroRNA (miRNA) expression vectors and Western blotting analysis were described previously and are included in the Supplemental Materials and Methods.

**Statistical analysis**

All statistical calculations were performed using GraphPad Prism 5. Overall survival (OS) was calculated from time of diagnosis to time of death from any cause or last follow-up. Progression-free survival (PFS) was calculated from time of diagnosis to time of progression or death from any cause. The actuarial probabilities of OS and PFS were determined using the Kaplan–Meier method, and differences were compared using the log-rank test. The 5-year survival and 5-year progression-free survival rates describe the percentage of patients that are alive five years after their cancer is diagnosed; they were compared using Fisher’s exact test. All differences with \( P \leq 0.05 \) were considered statistically significant.
Results

Distribution and frequency of single nucleotide variants in TP53 UTRs in 244 de novo DLBCL

In the initial group of 244 DLBCL samples we identified 128 single nucleotide variants (SNVs) in the UTRs of 107 patients, with 14 in the 5'UTR and 114 in the 3'UTR (Fig. 1A; Table S1). Six 3'UTR SNVs have been reported as single nucleotide polymorphisms in dbSNP and occurred 37 times (Table 1). We designated the newly-found noncoding SNVs in the TP53 gene as new (n) SNVs. A total of 108 nSNVs were identified with 121 occurrences in the 3'UTR in 76 patients. More patients had at least one nSNV in the 3'UTR (n=76, 31%) than they had non-synonymous mutations in the CDS (n=49, 20%). In the 76 patients with 3'UTR nSNVs, 32 (42%) had multiple lesions, compared with 2 of 49 (4%) patients with 2 or more CDS mutations. Ten of the TP53 3'UTR nSNVs occurred twice and one occurred three times (Fig. 1A). 5'UTR SNVs occurred at a much lower frequency (n=15 patients, 6%; Table S1; none of them reported in dbSNP) and was not further analyzed in the current study.

Most TP53 CDS somatic mutations are clustered within exons 5-8.19 While the nucleotide length (544 bp) of exons 5-8 accounts for only ~46% of the coding exons (1182 bp), mutations in these exons account for 94% of all CDS somatic mutations.19 Our analysis of TP53 CDS in the 506 DLBCL patients confirmed the uneven codon distribution of single nucleotide missense substitutions: mutations in exons 5-8 account for 90% of all mutations in coding exons (exons 3-10 plus parts of exon 2 and exon 11).17 In the initial study group of 244 patients, 92% of all CDS mutations were found in exons 5-8 (45 of 49 cases, Fig. 1B). In contrast, 3'UTR nSNVs were evenly dispersed across the 1207-bp 3'UTR in the 76 patients with nSNVs (Fig. 1A). The distribution of SNVs in the 5'UTR in the 244 cases was similar to that in the 3'UTR (Fig. 1C). These data suggest that UTR variants in the TP53 gene are distributed more evenly than are CDS mutations.

Survival in relation to clinical characteristics and TP53 3'UTR variation

The clinical and biological characteristics of the 244 DLBCL patients and their impact on OS and PFS are summarized in Table S2. We first verified that no adverse clinical variables were associated with variation in the 3'UTR. When patients were divided based on the presence of TP53 3'UTR lesions (SNPs only, nSNVs only, or any SNVs [either SNP or nSNV]), no significant difference in OS or PFS was found (Fig. 2). TP53 CDS mutations (MUT-CDS) predicted poorer OS ($P = 0.0028$) and PFS ($P = 0.0086$) in this 244-patient cohort (Fig. 3A-B). We then divided patients into two groups according
to the p53 CDS status and analyzed survival in relation to SNVs. We found that SNPs in the TP53 3'UTR had no impact on patient OS or PFS (Fig. S1). Of the patients with WT-CDS only, those without nSNVs (n = 136) seemed to have poorer OS and PFS than those with nSNVs (n = 59), but the difference was not statistically significant ($P = 0.15$ for OS and $P = 0.17$ for PFS, respectively; Fig. 3C-D). When 5-year survival rates were calculated, nSNVs confers better survival in patients with WT-CDS (78% vs 59%, $P = 0.025$) and 5-year progression-free survival also showed a trend (76% vs 58%, $P = 0.079$). In contrast, patients with MUT-CDS and nSNVs (n = 17) had poorer OS and PFS than did patients with MUT-CDS and no nSNVs (n = 32; $P = 0.019$ for OS and $P = 0.018$ for PFS; Fig. 3E-F). Indeed, all patients with MUT-CDS and nSNVs died within 76 months after diagnosis (Fig. 3E). We also combined SNPs and nSNVs as total SNVs; there were similar survival trends, but not statistical significant (Fig. S2). After the cohort was divided into ABC- and GCB-DLBCL subsets, the prognostic impact of nSNVs in OS seemed to favor patients with WT-CDS and against patients with MUT-CDS in the GCB subset. Yet none were statistically significantly, most likely due to a smaller cohort (Fig. S3). Again, nSNVs confers better 5-year survival in GCB-patients with WT-CDS ($P = 0.0006$). We did not find statistical differences in frequencies of subtypes, CDS mutation, and TP53 monoallelic deletion between patients with nSNVs and those without.

**Validation of TP53 3'UTR variants**

To determine whether nSNVs are common in DLBCL, we sequenced the TP53 3'UTR in tumor samples from an additional 247 DLBCL patients. We found 106 nSNVs in 70 (28%) patients with 5 occurring twice. All 6 SNPs identified in the initial set were found in this 247 patient group (Table 1 and Table S3). For the combined 491 patients, 187 nSNVs were identified in the TP53 3'UTR: 151 occurred once, 31 occurred twice, and 5 occurred three times (Table S4). These data indicate that nSNVs in the TP53 3'UTR occur frequently in DLBCL. In addition, nSNVs were dispersed across the whole TP53 3'UTR (Fig. 4A), providing added evidence that 3'UTR nSNVs are distributed more evenly than are CDS mutations.

**3'UTR variants in miRNA binding sites**

Eleven miRNAs are reported to target the human TP53 gene: miR-25, miR-30d, miR-125b, miR-504, miR-380-5p, miR-92a, miR-141, miR-200a, miR-1285, miR-15a and miR-16. Among these miRNAs, three pairs have the same seed sequence: miR-25 and miR-92a, miR-141 and miR-200a, and miR-15a and miR-16. In addition, miR-504 and miR-1285 target two sites each. Therefore, the TP53 3'UTR has a total of 10 experimentally validated miRNA targeting sites. Except for the
miR-141 and miR-380-5p sites, 8 of 10 miRNA sites had one or more nSNVs occurring at least once in 491 patients (Table 2). We analyzed the potential of nSNVs to affect putative miRNA binding as predicted by TargetScan,\textsuperscript{28,29} in which 8 nucleotides in the target 3'UTR are critical to miRNA suppression: a sequence complementary to the 6-nucleotide seed sequence of the miRNAs and two flanking nucleotides (A1 and m8). Of the 146 patients carrying one or more nSNVs in the 3'UTR, 95 (65\%) had nSNVs that disrupt and/or generate miRNA binding sites as predicted by TargetScan; 78 of 95 patients carried nSNVs disrupting putative miRNA binding (Table 3 and Fig. 4B).

\textit{p53 expression analysis of 3'UTR nSNVs}

To test whether 3'UTR variation impacts p53 expression, we introduced gene expression cassettes carrying human TP53 cDNA with 3'UTR variants in the miR-125b binding site (Fig. 5A) into p53-null H1299 cells. As shown in Fig. 5B, TP53 3'UTR nSNVs did not alter p53 expression compared with WT. However, when cells were co-transfected with miR-125b, three nSNVs (733U, 737A and 739A) disrupting the seed match between miR-125b and the TP53 3'UTR released the negative regulation of p53 by miR-125b (Fig. 5C). One variant (740A) that converts the seed match from 7mer to 8mer, which is predicted to have a greater effect on p53 down-regulation, did not enhance p53 suppression by miR-125b. These results demonstrate that miRNAs and nSNVs in the TP53 3'UTR significantly alter p53 protein levels \textit{in vitro}. We next analyzed the relationship between p53 protein expression with a 50\% cut-off\textsuperscript{17} and 3'UTR nSNVs in the initiation set. We found that there was no correlation between nSNVs and p53 expression in the entire cohort, the WT-CDS group, or the MUT-CDS group (Table S5). There are at least two possibilities: (1) nSNVs may have either inhibitory or stimulatory effects on p53 regulation; and (2) the standard assay of p53 expression in tumors (percentage of cells with positive staining rather the signal strength)\textsuperscript{17} cannot detect the subtle modulation of p53 protein levels by nSNVs.

\textit{Functional significance of a germline polymorphism in the TP53 3' UTR}

SNP rs78378222 (A\textsuperscript{1175}C) is located in the polyadenylation signal (PAS) of TP53 and is strongly associated with basal cell carcinoma (BCC).\textsuperscript{30} This germline variant was found in 7 patients (6 A/C and 1 C/C). We predicted a decrease in p53 expression when the TP53 gene contains the PAS variant (AATAAA > AATACA). We introduced gene expression cassettes carrying human TP53 cDNA with this PAS variant (TP53ExtA\textsuperscript{1175}C), or a control SNP rs114831472 (TP53ExtC\textsuperscript{1170}T, not associated with BCC and other cancers) located one nucleotide upstream of the PAS, and the 2.5kb genomic sequence downstream of the 3'UTR (Fig. 6A) into p53-null H1299 cells. We found significantly
lower p53 protein levels in cells with TP53ExtA^{1175}C than in cells with WT control or TP53ExtC^{1170}T (Fig. 6B). Northern blot analysis confirmed that p53 mRNA levels were down-regulated in cells with the p53 PAS variant (Fig. 6C). We assessed the effect of the TP53 PAS variant on apoptosis by using the terminal deoxynucleotidyl transferase dUTP nick end labeling (TUNEL) assay. In cells harboring the p53 PAS minor allele, cellular apoptosis was significantly lower than that in cells expressing the WT TP53 gene (Fig. 6D). These data indicate that the TP53 variant rs78378222 hinders p53 expression and its downstream cell functions in apoptosis.
Discussion

In this work, we found many previously unreported SNVs in the TP53 3'UTR in DLBCL tumor DNA samples. 3'UTR shortening and dysregulation of 3'UTR-binding proteins or miRNAs are reported. This study, however, is the first to demonstrate that 3'UTR variation is widespread in the TP53 gene in cancer patients. Most DLBCLs originate from B-cells in the GC, which have a high proliferation rate and are prone to genomic alterations. GC B-cells expand clonally and undergo physiological somatic hypermutation of their immunoglobulin (Ig) variable-region genes and class-switch recombination to alter the Ig subtype mediated by activation-induced cytidine deaminase (AID) as part of the immune response. BCL2 point mutations in DLBCL are suspected to be caused by this somatic hypermutation mechanism. TP53 3'UTR nSNVs found in this study are predominantly G→A and C→T substitutions and the frequency ratio of transition over transversion is 16.7, significantly higher than 2.1 (for all human variants in dbSNP), 2.0 (TP53 CDS mutation in DLBCL), or 7.3 (known SNPs of TP53 3'UTR). However, only a small number of nSNVs are located in WRCY/RGYW sites (Table S6), hotspots for AID-mediated physiological hypermutations. In addition, these nSNVs are not found in many projects that clearly sequence the full TP53 3'UTR from hundreds of individuals. Therefore, most nSNVs found in this study are likely cancer-specific mutations rather than germline polymorphisms or a result of a normally operative somatic hypermutation process. We acknowledge that the absence of patient-matched noncancerous tissue samples prevents us to completely rule out the possibility of rare private germline variants in these patients. Counting CDS mutations and 3'UTR nSNVs, approximately 45% of DLBCL patients (109 of 244) had alterations in TP53, making TP53 the most frequently mutated gene identified in DLBCL. A recent study reported an R-CHOP treatment-associated pattern of p53 and cell cycle dysregulation in DLBCL, further emphasizing the importance of p53 in DLBCL pathogenesis and prognosis.

Most of the TP53 3'UTR nSNVs are located in validated or putative miRNA binding sites. Each specific mammalian cell expresses about 70 miRNAs at an appreciable level, yet normal B-cell development and the initiation of malignancy and tumor progression are ordered and dynamic processes in which hundreds of miRNAs are differentially expressed in over a dozen distinct cell populations. Thus, 3'UTR variation may have a direct biological impact on potential interaction between miRNAs and p53 in B-cell maturation or DLBCL pathogenesis. Since miRNAs repress gene expression, 3'UTR alterations that are predicted to disrupt miRNA binding sites, as occurred in 78 of 146 patients with one or more nSNVs in this study (Table 3), could reduce the efficiency of miRNA
suppression and subsequently increase p53 expression. This has been experimentally validated for a few nSNVs located at the miR-125b site. We reasoned that when the TP53 CDS is WT, the increased levels of tumor-suppressive p53 protein confer better 5-year survival in patients with 3'UTR nSNVs, as WT p53 is advantageous for patients undergoing R-CHOP chemotherapy; in contrast, when the TP53 CDS is mutated concomitantly, the elevated levels of oncogenic p53 will worsen survival in patients with 3'UTR nSNVs. Our findings do show that variation in the TP53 3'UTR as a prognostic indicator of survival in DLBCL patients is dependent on the status of the CDS, yet there is no correlation between nSNV and p53 protein expression in the entire cohort, the WT-CDS group, or the MUT-CDS group. It should be noted that 3'UTR variation may also disrupt interactions between p53 mRNA and RNA-binding proteins, although the exact nucleotides have not been pinpointed. Additionally, 3'UTR variation could affect mRNA stability or translation efficiency due to other mechanisms. Further investigation is needed to ascertain the role of the nSNVs in p53 regulation and DLBCL prognosis.

The disappearance of the survival advantage at about 10 years for DLBCL patients with a WT-CDS and nSNVs is noteworthy. The frequency of TP53 CDS mutations may be higher in relapsed DLBCL patients, and this may partially override the impact of 3'UTR variation on survival beyond 5 years. In patients with multiple myeloma, the frequency of TP53 CDS mutations in de novo disease is only ~3%, yet this number rises up to 30% in patients with relapsed disease. Relapsed DLBCL is generally diagnosed by imaging and limited sampling such as fine-needle aspiration biopsy rather than by excisional biopsy; the lack of adequate specimens for genetic analysis precluded our testing for TP53 CDS mutations in DLBCL patients at relapse and is a limitation of this study.

TP53 germline variants have been investigated for cancer susceptibility using genome-wide association studies (GWAS), in which a type I error rate (P value threshold for significance) of $10^{-7}$ is generally regarded as indicating genome-wide significance. Until 2011, no GWAS study reported a significant association between any germline TP53 polymorphism (other than LFS variants) and any cancer with a $P$ value $\leq 10^{-7}$. In 2011, however, rs78378222 was found to be strongly associated with BCC, with an odds ratio (OR) = 2.16; $P = 2.2 \times 10^{-20}$. rs78378222 is located in the 5th nucleotide of the TP53 PAS that is required for polyadenylation machinery recognition and subsequent cleavage, polyadenylation and export of mature mRNAs to the cytoplasm. The risk-associated allele is a “C”, resulting in an alternative PAS (AATACA) replacing the canonical PAS (AATAAA). Among millions of SNPs in the human genome, rs78378222 has the strongest association with BCC. This
SNP is also associated with an increased risk of prostate cancer (OR = 1.44, \( P = 2.4 \times 10^{-6} \)), glioma (OR = 2.35, \( P = 1.0 \times 10^{-5} \)), and colorectal adenoma (OR = 1.39, \( P = 1.6 \times 10^{-4} \)),\(^3\) as well as esophageal squamous cell carcinoma (OR=3.22, \( P=1.34\times10^{-4} \)).\(^4\) In an independent confirmation study, rs78378222 was found to be associated with a 3.5 times higher risk of glioma (\( P=1.34\times10^{-4} \)).\(^5\) Stefansson and colleagues showed that rs78378222[A/C] heterozygotes (n=7) expressed “somewhat less” p53 transcript levels than WT homozygotes in human blood specimens (\( P=0.041 \)).\(^3\) In the present study, we tested the role of the alternative PAS in p53 regulation using a p53-null cell line with exogenous p53 constructs. We found that the C allele, as compared with the A allele, dramatically lowers p53 mRNA levels and results in reduced p53 protein expression and cellular apoptosis. Our data are in agreement with a previous report that simian virus 40 mRNA with AATACA as PAS exhibits an 8-fold decrease in cleavage efficiency compared with the canonical site.\(^6\) These findings provide direct evidence that rs78378222 negatively impacts p53 expression and function. p53 stabilization and function rely on various post-translational modifications. As rs78378222 occurs in the PAS signal of \( TP53 \), none of the downstream p53 protein modifications are affected. Thus, this variant presents a unique instance of \( TP53 \) alleles with unambiguous reduced-function, in contrast with most CDS mutants that are associated with a “loss-of-function” and/or a “gain-of-function”.\(^7\)

In summary, in this study we have shown that sequence variation in the \( TP53 \) 3′UTR is frequent in DLBCL and that the role of this sequence variation in patient survival depends on the status of the \( TP53 \) CDS. To our knowledge, this study is the first to demonstrate the prognostic importance of the \( TP53 \) 3′UTR in cancer or any other disease. The full extent of the pathobiology of 3′UTR variants remains to be determined.
Acknowledgments

The authors thank the consortium program team of pathologists, hematologists, and clinicians; each of the contributing center principal physicians; their patients; and former and current hematopathology and hematology/oncology fellows for their support. The Consortium is in memory of Michael W. Gordon, a gifted PhD graduate who contributed significantly to this work, conducted the research, analyzed the data, performed the statistical analysis, and wrote the manuscript.

This study was supported by the National Cancer Institute/National Institutes of Health (R01CA138688, 1RC1CA146299, P50CA136411, and P50CA142509). Z.Y.X.M. is the recipient of the Shannon Timmins Leukemia Fellowship Award at The University of Texas MD Anderson Cancer Center. K.H.Y is supported by The University of Texas MD Anderson Cancer Center Institutional Research and Development Fund, Institutional Research Grant Award, An MD Anderson Cancer Center Lymphoma SPORE Research Development Program Award, An MD Anderson Cancer Center Myeloma SPORE Research Development Program Award, Gundersen Lutheran Medical Foundation Award, MD Anderson Cancer Center Collaborative Funds with Roche Molecular System, HTG Molecular Diagnostics and Daiichi Sankyo Pharmaceutical.

Authorship


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


Table 1. Frequency of known TP53 3’UTR SNPs in DLBCL cohorts.

<table>
<thead>
<tr>
<th>SNP ID</th>
<th>Position¹</th>
<th>MAF²</th>
<th>Heterozygous frequency</th>
<th>Validation³</th>
<th>SNP occurrence⁴</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs8065799</td>
<td>9 A&gt;C</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs1042526</td>
<td>37 A&gt;T</td>
<td>N/A</td>
<td>N/A</td>
<td>Precious</td>
<td>0</td>
</tr>
<tr>
<td>rs35919705</td>
<td>105 C&gt;T</td>
<td>0.010</td>
<td>0.015</td>
<td>Frequency</td>
<td>0</td>
</tr>
<tr>
<td>rs16956880</td>
<td>205 G&gt;A</td>
<td>0.032</td>
<td>0.024</td>
<td>Cluster</td>
<td>3</td>
</tr>
<tr>
<td>rs79516338</td>
<td>309 G&gt;A</td>
<td>N/A</td>
<td>0.500</td>
<td>Frequency</td>
<td>0</td>
</tr>
<tr>
<td>rs34486624</td>
<td>314 G&gt;A</td>
<td>N/A</td>
<td>0.020</td>
<td>Frequency</td>
<td>0</td>
</tr>
<tr>
<td>rs17881366</td>
<td>328 G&gt;A</td>
<td>0.026</td>
<td>0.010</td>
<td>Cluster</td>
<td>0</td>
</tr>
<tr>
<td>rs4968187</td>
<td>485 G&gt;A</td>
<td>0.035</td>
<td>N/A</td>
<td>Cluster</td>
<td>11</td>
</tr>
<tr>
<td>rs916131</td>
<td>491 G&gt;A</td>
<td>N/A</td>
<td>N/A</td>
<td>Precious</td>
<td>0</td>
</tr>
<tr>
<td>rs916132</td>
<td>514 T&gt;C</td>
<td>N/A</td>
<td>N/A</td>
<td>Precious</td>
<td>0</td>
</tr>
<tr>
<td>rs1811095</td>
<td>534 C&gt;T</td>
<td>N/A</td>
<td>N/A</td>
<td>Cluster</td>
<td>0</td>
</tr>
<tr>
<td>rs35944977</td>
<td>534-535 TC&gt;CT</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs1811096</td>
<td>535 C&gt;T</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs17886358</td>
<td>569-570 GT&gt;del</td>
<td>N/A</td>
<td>0.435</td>
<td>Cluster</td>
<td>0</td>
</tr>
<tr>
<td>rs34607064</td>
<td>595 C&gt;del</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs66759322</td>
<td>611 A&gt;del</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs17879353</td>
<td>613 C&gt;A</td>
<td>0.007</td>
<td>N/A</td>
<td>Cluster</td>
<td>1</td>
</tr>
<tr>
<td>rs1042535</td>
<td>634 C&gt;CG&gt;GA</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs1794291</td>
<td>635 A&gt;G</td>
<td>N/A</td>
<td>N/A</td>
<td>Cluster</td>
<td>0</td>
</tr>
<tr>
<td>rs34182553</td>
<td>674 C&gt;del</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs77011241</td>
<td>773 C&gt;T</td>
<td>N/A</td>
<td>0.389</td>
<td>Frequency</td>
<td>0</td>
</tr>
<tr>
<td>rs35940853</td>
<td>798 T&gt;del</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs17884306</td>
<td>826 G&gt;A</td>
<td>0.044</td>
<td>N/A</td>
<td>Cluster</td>
<td>17</td>
</tr>
<tr>
<td>rs17884586</td>
<td>855 C&gt;T</td>
<td>N/A</td>
<td>0.012</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs35659787</td>
<td>886 G&gt;A</td>
<td>N/A</td>
<td>0.10</td>
<td>Frequency</td>
<td>0</td>
</tr>
<tr>
<td>rs55817367</td>
<td>925 C&gt;T</td>
<td>0.007</td>
<td>N/A</td>
<td>1000Genome</td>
<td>0</td>
</tr>
<tr>
<td>rs1794292</td>
<td>995 G&gt;C</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs1794293</td>
<td>1057 C&gt;G</td>
<td>N/A</td>
<td>N/A</td>
<td>-</td>
<td>0</td>
</tr>
<tr>
<td>rs114831472</td>
<td>1070 C&gt;T</td>
<td>0.032</td>
<td>0.065</td>
<td>1000Genome</td>
<td>2</td>
</tr>
<tr>
<td>rs78378222</td>
<td>1175 A&gt;C</td>
<td>0.010</td>
<td>N/A</td>
<td>Cluster</td>
<td>3</td>
</tr>
</tbody>
</table>

1. Position
2. Minor Allele Frequency
3. Validation
4. SNP occurrence in groups

For personal use only. on November 16, 2017. By guest www.bloodjournal.org From
1 Counted from first nucleotide of the TP53 3’UTR; “del”, deletion.
2 MAF, minor allele frequency; N/A, Not available.
3 Validation: Cluster, multiple submissions mapped to the same location and clustered into one refSNP cluster and assigned a reference SNP ID number; Frequency, validated by both population frequency data and genotype data; 1000Genome, validated by the 1000 Genomes Project; Precious, cited as clinically relevant in the literature and belonging to a reserved set of clinically associated SNPs in dbSNP.
4 Minor allele found in the initiation set (244 cases) or the validation set (247 cases). Some patients possess more than one SNPs.
Table 2. **SNVs in the 3′UTR disrupting the validated seed match of TP53: miRNA**

<table>
<thead>
<tr>
<th>TP53 3′UTR and miRNA</th>
<th>Consequential pairing ¹</th>
<th>Seed Match ²</th>
<th>No. of Patients ³</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Seed+1A</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>NO</td>
<td></td>
</tr>
</tbody>
</table>

1 Nucleotides in the TP53 3′UTR are numbered starting from the one after the stop codon. Nucleotides complementary to the seed sequence (positions 2-7) of miRNAs are indicated. Position (pos.) numbers correspond to nucleotides that are complementary to the miRNA seed sequence (in...
bold). rs114831472, affecting miR-504 binding, is “SNP 1070C>T”. One miR-504 site overlaps with the miR-125b site. “x” indicates that base pairing is disrupted.

2 The 7mer-m8 site comprises the seed match supplemented by a Watson–Crick match to miRNA nucleotide 8. The 7mer-A1 site comprises the seed match supplemented by an A across from miRNA nucleotide 1. The 8mer site comprises the seed match supplemented by both m8 and A1. “NO” indicates that the seed match was abolished. “?” indicates unconventional interaction.

3 Number of patients carrying SNVs affecting putative miRNA binding. Unique SNVs are indicated in parentheses. SNVs affecting TP53:miRNA complementarity other than the seed match are not included.
Table 3. Occurrence of nSNVs in the TP53 3′ UTR potentially affecting miRNA binding in DLBCL

<table>
<thead>
<tr>
<th>Subgroup</th>
<th>76 of 244</th>
<th>70 of 247</th>
<th>146 of 491</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>TP53 CDS status</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CDS Mut</td>
<td>16 (21)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>CDS WT</td>
<td>60 (79)</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td><strong>3′UTR nSNV affecting putative miRNA binding</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Generating only</td>
<td>12</td>
<td>5</td>
<td>17</td>
</tr>
<tr>
<td>Disrupting only</td>
<td>31</td>
<td>23</td>
<td>54</td>
</tr>
<tr>
<td>Generating &amp; disrupting</td>
<td>15</td>
<td>9</td>
<td>24</td>
</tr>
<tr>
<td>Others</td>
<td>18</td>
<td>33</td>
<td>51</td>
</tr>
<tr>
<td><strong>No. of 3′UTR nucleotides altered</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>42 (59.5)</td>
<td>42 (60)</td>
<td>84 (57.5)</td>
</tr>
<tr>
<td>2</td>
<td>19 (23.8)</td>
<td>18 (25.7)</td>
<td>37 (25.3)</td>
</tr>
<tr>
<td>3</td>
<td>8 (9.5)</td>
<td>7 (10)</td>
<td>15 (10.3)</td>
</tr>
<tr>
<td>4</td>
<td>5 (4.8)</td>
<td>3 (4.3)</td>
<td>8 (5.5)</td>
</tr>
<tr>
<td>5</td>
<td>1 (1.2)</td>
<td>0</td>
<td>1 (0.7)</td>
</tr>
<tr>
<td>6</td>
<td>1 (1.2)</td>
<td>0</td>
<td>1 (0.7)</td>
</tr>
</tbody>
</table>

1 nSNVs predicted by TargetScan 5.0 to be targeted by miRNAs (based on the seed sequence of the miRNA plus A1 and m8). Some patients had both mutations generating miRNA binding and mutations disrupting it. Three SNPs listed in dbSNP that putatively affect miRNA binding (rs16956880 [205 G>A], rs114831472 [1070 G>A], and rs78378222 [1175 C>A]) were excluded. Others: outside seed match or not in miRNA sites. NA, not applicable.
Legends for Figures

Figure 1. Distribution of variation in the TP53 gene in 244 de novo DLBCL patients. (A) Distribution of SNVs in the 3′UTR (part of exon 11; nucleotide 1385-2591 of NM_000546.5 of the TP53 reference mRNA). (B) Distribution of CDS mutations in exons 2-10 and part of exons 2 and 11 (top bar; nucleotide 203-1384 of NM_000546.5). (C) Distribution of 5′UTR SNVs in exon 1 and part of exon 2, nucleotide 1-202 of NM_000546.5. Numbers on the top indicate exons.

Figure 2. Kaplan-Meier estimates of overall survival (OS) and progression-free survival (PFS) in DLBCL patients with SNVs in the TP53 3′UTR. (A) OS of patients with or without SNPs. (B) PFS of patients with or without SNPs. (C) OS of patients with or without nSNVs. (D) PFS of patients with or without nSNV. (E) OS of patients with or without SNVs (either SNP or nSNV). (F) PFS of patients with or without SNVs.

Figure 3. Kaplan-Meier estimates of overall survival (OS) and progression-free survival (PFS) in DLBCL patients classified according to TP53 CDS status. (A) OS of patients with or without a CDS mutation (MUT-CDS). (B) PFS of patients with or without a MUT-CDS. (C) OS of patients with a WT-CDS and nSNV. (D) PFS of patients with a WT-CDS and nSNVs. (E) OS of patients with a MUT-CDS and nSNV. (D) PFS of patients with MUT-CDS and nSNVs.

Figure 4. Variation in the TP53 3′UTR in 491 de novo DLBCL patients. (A) Distribution of SNVs in the 3′UTR. (B) Potential effects of SNVs on the interaction of p53 mRNA and miRNAs that are experimentally validated or predicted (by TargetScan) to target the TP53 3′. SNVs listed in dbSNP are marked with a “×”.

Figure 5. Expression analyses of TP53 3′UTR variation affecting miR-125b binding in p53-null cells. (A) Schematic representation of the p53 cDNA construct and 3′UTR variants. (B) Western blot of p53 expression in p53-null H1299 cells transfected with p53 constructs. (C) Western blot of p53 expression in p53-null H1299 cells transfected with p53 constructs with or without miR-125b. β-actin was used as a control.

Figure 6. p53 expression and function for the TP53 PAS variant in p53-null cells. (A) Schematic representation of the TP53 constructs. Each carries a WT 5′UTR, WT CDS, a WT or
mutant 3’UTR, and a 2.5kb native genomic sequence downstream of the $TP53$ PAS. (B) Western blot of p53 expression in p53-null H1299 cells transfected with p53 constructs. β-actin was used as a control. “OE”, over exposure. (C) Northern blot of p53 mRNAs in H1299 cells carrying exogenous p53 genes. A housekeeping RNA (H1RNA, the RNA subunit of human RNase P) was used as a control. (D) Apoptosis of cells with p53 constructs. Left, representative TUNEL images; right, percentage of TUNEL-positive cells. *, P≤0.05.
Figure-1

A

B

C
Figure-2
Figure-3

A

WT-CDS (n=195)

OS

P=0.0028

MUT-CDS (n=49)

Survival (%)

0 20 40 60 80 100

Months

B

WT-CDS (n=195)

PFS

P=0.0086

MUT-CDS (n=49)

Survival (%)

0 20 40 60 80 100

Months

C

5-Year OS (P=0.025)

With NSNV, WT-CDS (n=59)

OS

P=0.15

No NSNV, WT-CDS, (n=136)

Survival (%)

0 20 40 60 80 100

Months

D

5-Year PFS (P=0.079)

With NSNV, WT-CDS (n=59)

PFS

P=0.17

No NSNV, WT-CDS (n=136)

Survival (%)

0 20 40 60 80 100

Months

E

No NSNV, MUT-CDS (n=32)

OS

P=0.019

With NSNV, MUT-CDS (n=17)

Survival (%)

0 20 40 60 80 100

Months

F

No NSNV, MUT-CDS (n=31)

PFS

P=0.018

With NSNV, MUT-CDS (n=17)

Survival (%)

0 20 40 60 80 100

Months
Figure-4
Figure-5

A

![Diagram of TP53 and miR-125b binding sites.]

B

<table>
<thead>
<tr>
<th></th>
<th>WT</th>
<th>733U</th>
<th>737A</th>
<th>739A</th>
<th>740A</th>
</tr>
</thead>
<tbody>
<tr>
<td>p53</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

C

<table>
<thead>
<tr>
<th></th>
<th>Control</th>
<th>miR-125b</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>WT</td>
<td>WT</td>
</tr>
<tr>
<td>p53</td>
<td></td>
<td></td>
</tr>
<tr>
<td>β-actin</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 6

A. Schematic representation of the TP53 gene region, showing the CMV promoter, ATG translation initiation codons, and the CDS. The 5'UTR, CDS, and 3'UTR are indicated. The genomic region is shown with the nucleotide changes at positions 1170 and 1175.

B. Western blot analysis of control (Mock) and TP53Ext, TP53ExtC1170T, and TP53ExtA1175C constructs, showing expression levels of p53, p53 (OE), and β-Actin.

C. Northern blot analysis showing p53 mRNA levels in TP53Ext and TP53ExtC1170T constructs.

D. Microscopic images showing the effect of TP53Ext and TP53ExtA1175C on apoptosis, with quantification of % apoptosis shown in the bar graph.

* Significant difference.
Single nucleotide variation in the TP53 3' untranslated region in diffuse large B-cell lymphoma treated with rituximab-CHOP: a report from the International DLBCL Rituximab-CHOP Consortium Program

Yong Li, Michael W. Gordon, Zijun Y. Xu-Monette, Carlo Visco, Alexander Tzankov, Dehui Zou, Lugui Qiu, Santiago Montes-Moreno, Karen Dybkaer, Attilio Orazi, Youli Zu, Govind Bhagat, Kristy L. Richards, Eric D. Hsi, William W.L. Choi, J. Han van Krieken, Qin Huang, Weiyn Ai, Maurilio Ponzoni, Andrés J.M. Ferreri, Jane N. Winter, Ronald S. Go, Miguel A. Piris, Michael B. Möller, Lin Wu, Michael Wang, Kenneth S. Ramos, L. Jeffrey Medeiros and Ken H. Young