Systematic analysis of the frequently amplified 2p15-p16.1 locus reveals PAPOLG as a potential proto-oncogene in follicular and transformed follicular lymphoma

Deniz KURŞUN1,2, Can KÜÇÜK1,2,3,*

1Izmir International Biomedicine and Genome Institute (iBG-İzmir), Dokuz Eylül University, İzmir, Turkey
2Izmir Biomedicine and Genome Center (IBG), İzmir, Turkey
3Department of Medical Biology, Faculty of Medicine, Dokuz Eylül University, İzmir, Turkey

Abstract: Transformed follicular lymphoma (tFL) originates from histological transformation of follicular lymphoma (FL), which is the most common indolent non-Hodgkin lymphoma. High-resolution genomic copy-number analysis previously identified frequent amplification of the 2p15-p16.1 locus in FL and tFL cases. The genes (i.e. BCL11A, PAPOLG, PUS10, and USP34) in this amplified locus have not been systematically investigated to date in terms of their role in FL pathogenesis or transformation to tFL. Here we investigated the relationship between amplification and expression of genes in 2p15-p16.1 as well as their expression after histological transformation. NCBI GEO SNP array and gene expression profile (GEP) data of tFL cases were analyzed to evaluate the relationship between amplification and mRNA expression. Moreover, transcript levels of these four genes in FL cases were compared with those of patient-matched tFL cases and normal B-cells. Amplification of the 2p15-p16.1 locus is associated with increased transcription of BCL11A and PAPOLG in tFL cases, of which the latter showed increased expression after histological transformation. Compared with the level in normal B-cells, PAPOLG was significantly overexpressed in FL cases, but expression levels of the other three genes did not show any significant difference. Altogether these results suggest that PAPOLG may be the most critical gene in terms of transformation to tFL.

Key words: Amplification, 2p15-16.1, BCL11A, PAPOLG, PUS10, USP34, proto-oncogene, tFL

1. Introduction

Follicular lymphoma (FL) is the second most common type of lymphoma in the world (Diumenjo et al., 2016). FL occasionally undergoes histological transformation to higher grade lymphomas such as diffuse large B-cell lymphoma (DLBCL) or Burkitt lymphoma (Lossos and Gascoyne, 2011). The characteristic genetic abnormality of FL and transformed follicular lymphoma (tFL) is the t(14;18) chromosomal translocation that leads to constitutive overexpression of BCL2, which promotes survival of germinal center B-cells (Tsujimoto et al., 1985). Recurrent mutations in chromatin regulator genes such as MLL2, EZH2, or CREBBP have been identified to drive FL tumorigenesis (Okosun et al., 2014). Recent studies showed that protection against apoptosis allows early-stage FL cells to proliferate rapidly and acquire additional genetic, epigenetic, and metabolic abnormalities associated with FL pathogenesis (Küppers and Stevenson, 2018; Link, 2018; Minoia et al., 2018). However, the full list of molecular abnormalities related to the development of FL has not been identified or characterized yet.

Histological transformation of FL to higher grade malignancies such as DLBCL has been estimated to occur in ~50% of cases during the course of the disease (Kridel et al., 2012). These tFL cases have a very aggressive clinical course with a median survival time of 1.7 years posttransformation (Al-Tourah et al., 2008). P53 mutations (Coco et al., 1993) and a pluripotency signature related to MYC overexpression (Gentles et al., 2009) have been identified to be associated with transformation of FL to tFL; however, the full list of molecular aberrations associated with tFL has not been addressed yet.

Genome-wide copy number analysis of FL and tFL cases identified recurrent gain/amplification of the 2p15-p16.1 locus that includes BCL11A, PAPOLG, PUS10, REL, and USP34, which are candidate proto-oncogenes (Eide et al., 2010; Bouska et al., 2014). A recent report (Hu et al., 2017) showed a weak correlation between REL amplification and its mRNA expression. The same study also showed that ectopic expression of REL marginally promoted cell growth in an FL cell line in limiting serum concentrations. Altogether these observations suggest that other gene(s)
can be recognized by CD4+ T cells in melanoma patients. PAPOLG was described as a tumor-associated antigen that facilitates correct folding and assembly of tRNAs, which has an important role in several human cancers. Pseudouridine synthases act as RNA chaperones; they facilitate correct folding and assembly of tRNAs.

In the present study, we evaluated the relationship between gene copy number and mRNA expression of BCL11A, PAPOLG, PUS10, and USP34 genes located in the frequently amplified 2p15-p16.1 locus, were chosen based on a previous study that comprehensively analyzed copy number alterations in FL and tFL cases (Bouska et al., 2014). The amplified 2p15-p16.1 locus corresponded to recurrent copy number aberration (rCNA) ID: 693, which is a minimal region of abnormality that includes REL, BCL11A, PAPOLG, PUS10, and USP34 genes. As the role of REL in FL/tFL pathobiology was characterized in a recent report (Hu et al., 2017), the other four genes were evaluated in the current study.

2. Materials and methods

2.1. Patient samples

The characteristics of the 42 tFL cases used to evaluate the relationship between amplification and mRNA expression of genes located in 2p15-p16.1 were defined previously (Bouska et al., 2014; Hu et al., 2017). Similarly, the characteristics of the 12 tFL cases whose patient-matched diagnostic pretransformation FL biopsy samples were available were also described previously (Lossos et al., 2002). The ethics committee approval information for the patient samples used in this study was provided in previous studies of the publicly available data. The available ethical committee numbers are as follows: the regional committee of Oslo, Norway, for research ethics (protocol no. S-05 209) and the Institutional Review Board (IRB) of University of Nebraska Medical Center (no: IRB# 513-08-EP). The descriptions for DNA and/or RNA isolation and all subsequent experimental procedures of SNP array or DNA microarray experiments are available in the NCBI Gene Expression Omnibus (GEO) database (accession numbers: GSE67385, GSE81183, GSE3458, and GSE55267). The characteristics of the tFL cases used in this study to evaluate the relationship between gene copy number and expression are shown in Table S1.

2.2. Selection of genes in the 2p15-p16.1 locus for copy number and expression analyses

The genes evaluated in the present study, which are located in the frequently amplified 2p15-p16.1 locus, were chosen based on a previous study that comprehensively analyzed copy number alterations in FL and tFL cases (Bouska et al., 2014). The amplified 2p15-p16.1 locus corresponded to recurrent copy number aberration (rCNA) ID: 693, which is a minimal region of abnormality that includes REL, BCL11A, PAPOLG, PUS10, and USP34 genes. As the role of REL in FL/tFL pathobiology was characterized in a recent report (Hu et al., 2017), the other four genes were evaluated in the current study.

2.3. Selection of DNA microarray probe sets for gene expression analysis

The sensitivity and specificity values of HG U133 Plus 2.0 DNA microarray probe sets available for the analyzed genes were determined using the GeneAnnot database (Chalifa-Caspi et al., 2003). If more than one probe set is available for a gene, then the one with the highest sensitivity and specificity value was chosen for subsequent analyses. If more than one probe set for a gene (e.g., PAPOLG and USP34) has specificity and sensitivity values equal to one, then both probe sets were evaluated in transcript expression analyses.

2.4. Gene copy number and transcript expression analysis

Normalized SNP array (Affymetrix Mapping 250 K Nsp SNP Array) probe values of the BCL11A, PAPOLG, PUS10, and USP34 genes were obtained from the NCBI GEO database (accession number: GSE67385). These probe values were determined using the Genotyping Console software (Affymetrix Inc., Santa Clara, CA, USA). Gene copy numbers were estimated by calculating the average value of all corresponding SNP array probes for each gene in 42 tFL samples with available DNA microarray
data (Figures 1A–1D). These 42 tFL cases were then divided into two 10-sample groups based on the highest or lowest SNP array probe values for each gene evaluated (i.e. BCL11A, PAPOLG, PUS10, or USP34). After that, the NCBI GEO2R gene expression analysis tool (Barrett et al., 2013) was applied to log2-normalized, median centered Affymetrix Human Genome U133 Plus 2.0 data available as a part of the GSE81183 data set to address whether mRNA expression of each of these four genes differed in tFL cases with low or high gene copy numbers (Figure 2A).

The SNP array probe sets evaluated for these genes were as follows: BCL11A: SNP_A-1794265, SNP_A-1813498, SNP_A-2071773, SNP_A-2256538, and SNP_A-4193303; PAPOLG: SNP_A-1892864, SNP_A-2054939, and SNP_A-2094019; PUS10: SNP_A-1808220, SNP_A-2198855, and SNP_A-4208127; USP34: SNP_A-1788538, SNP_A-1845416, SNP_A-1943045, SNP_A-1956271, and SNP_A-1963530.

2.5. Evaluation of transcript levels of the amplified genes in FL cases and normal B cell subsets
The transcript expression level of the genes located in the 2p15-p16.1 amplicon was determined in FL cases (n = 63) and normal B-cell subsets (n = 6) by applying the GEO2R bioinformatics tool to the normalized GEP data set available as a part of the NCBI GEO database (accession number: GSE55267) by comparing the expression levels in FL cases to those in normal B cell subsets (Figure 2B). Similarly, transcript expression levels of these genes in FL cases were compared using GEO2R by dividing the FL cases into low- and high-grade groups (i.e. stage 1–2 vs. stage 3–3a) using the metadata available in GSE55267.

2.6. Comparison of mRNA expression levels of genes in the 2p15-p16.1 locus in tFL cases before and after histological transformation
Previously reported gene expression profiles of 12 tFL cases, which were derived from the Lymphochip cDNA microarray platform, were obtained from the NCBI GEO database (accession number: GSE3458). The Lymphochip microarray platform consists of 37,632 hotspots that represent 32,876 unique cDNA clones. In this microarray platform, the posttransformation (i.e. tFL) and the diagnostic FL samples that were obtained from the same patients were labeled with Cy5 or Cy3 fluorescent dye, respectively. For each transcript, the Cy5 to Cy3 ratio was used to calculate the fold-change in expression after histological transformation of FL cases (Figure 2C).

Relative transcript expression values from BCL11A cDNA clones (clone IDs: 7940 and 18549), PAPOLG (clone IDs: 845 and 6572), PUS10 (clone IDs: 32285 and 12710), and USP34 (clone IDs: 21935) were determined using the

Figure 1. Copy number estimates of four genes located in the 2p15-p16.1 locus in transformed follicular lymphoma cases. Gene copy numbers were estimated by calculating the average SNP array (Affymetrix mapping 250K Nsp SNP array) probe values for BCL11A (A), PAPOLG (B), PUS10 (C), and USP34 (D) for each of 42 tFL cases for which corresponding DNA microarray data are also available. Means ± SD for values of SNP array probes are shown for each gene. SNP probes used to estimate the copy number of each gene are described in the Materials and methods section. SD: Standard deviation.
NCBI GEO2R gene expression analysis tool (Barrett et al., 2013) in these tFL cases.

2.7. Statistical analysis
A two-sample Student's t-test was applied to evaluate the statistical significance for the difference observed in mRNA expression of the evaluated genes between tFL cases with amplification and tFL cases without amplification using Microsoft Excel 2016. Any difference with \( P < 0.05 \) was considered statistically significant.

3. Results
3.1. BCL11A and PAPOLG are significantly overexpressed in tFL cases with amplification
Forty-two tFL cases (Bouska et al., 2014; Hu et al., 2017) were divided into two subgroups based on the average probe intensity values of each of BCL11A, PAPOLG, PUS10, and USP34, which is determined using the values of normalized SNP array data. mRNA expression values were then analyzed with the GEO2R bioinformatics tool using the available DNA microarray probe sets (GEO accession number: GSE81184); B) Transcript expression of four amplified genes was evaluated with the GEO2R tool by comparing the expression levels in FL cases vs. normal B cell subsets (GEO accession number: GSE55267); C) Transcript expression levels of the four amplified genes in 12 tFL cases (i.e. DLBCL) were compared one by one with those of pretransformation levels for each tFL case by using the GEO2R tool (GEO accession number: GSE3458).

3.2. PAPOLG and PUS10 mRNA expression increases after histological transformation to tFL
To address whether expression of genes located in 2p15-p16.1 increases in FL samples after histological transformation to aggressive B-cell lymphoma (e.g., DLBCL), we evaluated whether mRNA expression of genes in 2p15-p16.1 (i.e. BCL11A, PAPOLG, PUS10, and USP34) increased in 12 FL cases after histological transformation to tFL using the NCBI GEO2R bioinformatics tool (Figure 4). We did not observe upregulation of BCL11A mRNA expression in most
Figure 3. Amplification of BCL11A or PAPOLG is associated with increased transcript expression levels in transformed follicular lymphoma cases. mRNA expression levels of tFL cases with BCL11A (A), PAPOLG (B), PUS10 (C), and USP34 (D) amplifications were compared to those with no amplification of the corresponding gene using probe set values (NCBI GEO dataset: GSE81183) individually and displayed using box-and-whisker plots (t-test, *: P < 0.01). n.s: nonsignificant.

Figure 4. Transcript expression levels of PAPOLG or PUS10 increase in transformed follicular lymphoma tumors compared with those of patient-matched pretransformation tumors. The change in BCL11A (A), PAPOLG (B), PUS10 (C), and USP34 (D) mRNA expression in 12 tFL cases compared to that of pretransformation levels was shown using cDNA microarray clones available in the Lymphochip platform. Clone IDs for each gene are shown near each plot (NCBI GEO accession number: GSE3458).
of the tFL cases analyzed using two different BCL11A cDNA clones (Figure 4A). PAPOLG and PUS10 mRNA expression increased in 75% of FL cases (9 of 12 cases) (Figures 4B and 4C). Most cases did not show upregulation of USP34 expression after histological transformation to tFL (Figure 4D).

3.3. PAPOLG is overexpressed in follicular lymphoma cases compared to normal B cells
Next, transcript expression levels of the four amplified genes were evaluated in FL cases (n = 63) and normal B-cell subsets (n = 6) to observe whether their expression is upregulated in FL cases. Of all four evaluated genes, only PAPOLG showed significant upregulation of mRNA expression in FL cases, whereas no significant change was observed for the transcript levels of BCL11A, PUS10, and UPS34 genes (Figures 5A–5D).

3.4. Comparison of the expression of amplified genes in low- and high-grade follicular lymphoma cases
To address whether any of the amplified genes show higher expression in high-stage FL cases, transcript expression of genes in the 2p15-16.1 locus were determined in 41 high (i.e. stage 3–3a) and 18 low (i.e. stage 1–2) stage FL cases using the GEO2R bioinformatics tool by reanalyzing the GEP data available in GSE55267. No significant difference in gene expression was observed for any of the four evaluated genes in these FL tumor samples (Figures 6A–6D).

4. Discussion
Recurrent genomic gains/amplifications may lead to the activation of proto-oncogenes through elevated transcript expression. In most cases, elevated transcript levels lead to increased protein expression of these proto-oncogenes (Bhargava et al., 2005; Borah et al., 2015). This amplification-related elevated expression has been shown to promote tumorigenesis in breast cancer cells through constitutive activation of the Akt signaling pathway (She et al., 2008) or chemotherapy resistance in solid tumors such as lung cancer (Engelman et al., 2007) and breast cancer (Li et al., 2010). Interestingly, amplification of oncogenes such as n-MYC in neuroblastoma (Brodur et al., 1984) or HER2/NEU amplification in breast carcinoma (Press et al., 1997) was shown to predict poor prognosis in cancer patients. Recurrent gains/amplifications have
also been commonly observed in lymphoid malignancies. For instance, amplification of BMI-1, a proto-oncogene involved in regulation of proliferation and senescence (Jacobs et al., 1999), was reported to be associated with high BMI-1 expression in mantle cell lymphoma cases (Bea et al., 2001). Similarly, another study reported upregulation of the FOXP1 proto-oncogene due to trisomy 3 or focal amplifications in activated B-cell (ABC)-type DLBCL (Lenz et al., 2008). In particular, recurrent gain/amplification of the 2p15-p16.1 locus, where REL, BCL11A, PAPOLG, PUS10, and USP34 genes reside, is a common observation in different lymphoma types including classical Hodgkin lymphoma (Martin-Subero et al., 2002) and transformed follicular lymphoma (Bouska et al., 2014).

Our results suggest PAPOLG as a proto-oncogene candidate potentially critical in FL tumorigenesis or histological transformation of FL to tFL due to the following observations. First, gain/amplification of PAPOLG leads to its increased transcript expression in amplified tFL cases. Second, PAPOLG expression increases in most FL cases after transformation to tFL. Third, it is overexpressed in FL cases compared to normal B cell subsets. These observations together with the previous studies that reported overexpression of PAPOLG (neo-PAP) in many cancer types (Topalian et al., 2001) suggest that its overexpression may have critical consequences during the development of FL or tFL through elevated transcript polyadenylation that may result in a more aggressive tumor phenotype as observed for PAP, another poly A polymerase gene, overexpressed in breast cancer ( Scorilas et al., 2000).

In conclusion, systematic analyses of expression of genes located in the recurrently amplified 2p15-p16 locus

Figure 6. The relationship between expression of amplified genes in low- and high-grade follicular lymphoma cases. Transcript expression levels of BCL11A (A), PAPOLG (B), PUS10 (C), and USP34 (D) in low grade (stage 1–2) and high grade (stage 3–3a) FL cases are shown with box-and-whisker plots.
revealed PAPOLG as the most likely candidate proto-oncogene in development of FL and transformation of FL to tFL. These analyses provide the basis for functional assays in the future that will be needed to elucidate the role of PAPOLG in these malignancies.

Acknowledgment
We thank the Young Scientists Award Program of the Turkish Academy of Sciences (TÜBA GEBİP 2017) (C.K.) for supporting this study.

References


### Table S1. Characteristics of tFL cases used in this study.*

<table>
<thead>
<tr>
<th>Case name</th>
<th>GEO accession # for U133+2 GEP</th>
<th>t(14;18) status</th>
<th>tFL biopsy tumor content</th>
<th>Overall survival (months)**</th>
<th>OS, posttransformation (months)</th>
</tr>
</thead>
<tbody>
<tr>
<td>tFL-1</td>
<td>GSM2144279</td>
<td>pos (MBR)</td>
<td>0.94</td>
<td>42.80</td>
<td>27.04</td>
</tr>
<tr>
<td>tFL-2</td>
<td>GSM2144287</td>
<td>pos (MBR)</td>
<td>0.32</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-3</td>
<td>GSM2144296</td>
<td>unknown</td>
<td>0.43</td>
<td>17.37</td>
<td>4.47</td>
</tr>
<tr>
<td>tFL-4</td>
<td>GSM2144305</td>
<td>unknown</td>
<td>0.57</td>
<td>82.89</td>
<td>5.20</td>
</tr>
<tr>
<td>tFL-5</td>
<td>GSM2144307</td>
<td>pos (MCR)</td>
<td>0.88</td>
<td>54.38</td>
<td>5.16</td>
</tr>
<tr>
<td>tFL-6</td>
<td>GSM2144315</td>
<td>pos (MCR)</td>
<td>0.61</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-8</td>
<td>GSM2144319</td>
<td>pos (MCR)</td>
<td>0.79</td>
<td>30.20</td>
<td>1.28</td>
</tr>
<tr>
<td>tFL-9</td>
<td>GSM2144320</td>
<td>unknown</td>
<td>0.84</td>
<td>118.06</td>
<td>28.26</td>
</tr>
<tr>
<td>tFL-10</td>
<td>GSM2144280</td>
<td>pos (MCR)</td>
<td>0.64</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-15</td>
<td>GSM2144282</td>
<td>pos (ICR)</td>
<td>0.68</td>
<td>58.42</td>
<td>0.56</td>
</tr>
<tr>
<td>tFL-16</td>
<td>GSM2144283</td>
<td>pos (ICR)</td>
<td>0.71</td>
<td>22.63</td>
<td>2.17</td>
</tr>
<tr>
<td>tFL-18</td>
<td>GSM2144285</td>
<td>unknown</td>
<td>0.49</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-19</td>
<td>GSM2144286</td>
<td>pos (clinical)</td>
<td>0.84</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-20</td>
<td>GSM2144288</td>
<td>pos (MBR)</td>
<td>0.95</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-21</td>
<td>GSM2144289</td>
<td>unknown</td>
<td>0.63</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-23</td>
<td>GSM2144295</td>
<td>pos (MBR)</td>
<td>0.26</td>
<td>234.21</td>
<td>171.81</td>
</tr>
<tr>
<td>tFL-32</td>
<td>GSM2144297</td>
<td>unknown</td>
<td>0.61</td>
<td>78.98</td>
<td>42.80</td>
</tr>
<tr>
<td>tFL-33</td>
<td>GSM2144298</td>
<td>pos (MBR)</td>
<td>0.4</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-34</td>
<td>GSM2144299</td>
<td>pos (MBR)</td>
<td>0.43</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-35</td>
<td>GSM2144300</td>
<td>pos (MBR)</td>
<td>0.91</td>
<td>53.82</td>
<td>1.18</td>
</tr>
<tr>
<td>tFL-36</td>
<td>GSM2144301</td>
<td>positive(der18 aCGH)</td>
<td>0.49</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-37</td>
<td>GSM2144302</td>
<td>unknown</td>
<td>0.37</td>
<td>168.75</td>
<td>144.18</td>
</tr>
<tr>
<td>tFL-38</td>
<td>GSM2144303</td>
<td>unknown</td>
<td>0.82</td>
<td>87.66</td>
<td>37.17</td>
</tr>
<tr>
<td>tFL-39</td>
<td>GSM2144304</td>
<td>positive (FISH)</td>
<td>0.77</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>tFL-40</td>
<td>GSM2144306</td>
<td>pos (MBR)</td>
<td>0.33</td>
<td>128.78</td>
<td>55.56</td>
</tr>
</tbody>
</table>

* This table is modified from Table S1 of Bouska et al. (2017), and it has also been reported recently by Hu et al. (2017).
** Overall survival values indicated are based on Bouska et al. (2014).

ICR: Internal Cluster Region
MBR: Major Breakpoint Region
MCR: Minor Cluster Region
NA: Not Available