Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis

René Scholtysik,[1] Markus Kreuz,[2] Wolfram Klapper,[3] Birgit Burkhardt,[4] Alfred C. Feller,[5] Michael Hummel,[6] Markus Loeffler,[2] Maciej Rosolowski,[2] Carsten Schwaenen,[7] Rainer Spang,[8] Harald Stein,[6] Christoph Thorns,[5] Lorenz Trümper,[9] Inga Vater,[10] Swen Wessendorf,[7] Thorsten Zenz,[7] Reiner Siebert,[10] and Ralf Küppers,[1] for the "Molecular Mechanisms in Malignant Lymphomas" Network Project of the Deutsche Krebshilfe

[1]Institute of Cell Biology (Cancer Research), Medical School, University of Duisburg-Essen, Essen; [2]Institute of Medical Informatics, Statistics and Epidemiology (IMISE), University of Leipzig, Leipzig; [3]Department of Pathology, Hematopathology Section and Lymph Node Registry, University Hospital Schleswig-Holstein, Campus Kiel/Christian-Albrechts-University, Kiel; [4]NHL-BFM Study Center, Department of Pediatric Hematology and Oncology, Justus-Liebig-University, Giessen; [5]Institute of Pathology, University of Lübeck, Lübeck; [6]Institute of Pathology, Charité, Campus Benjamin Franklin, Berlin; [7]Internal Medicine III, University Hospital of Ulm, Ulm; [8]Institute of Functional Genomics, University of Regensburg, Regensburg; [9]Department of Hematology/Oncology, University Hospital Göttingen, Göttingen and [10]Institute of Human Genetics, Christian-Albrechts University Kiel & University Hospital Schleswig-Holstein, Campus Kiel, Kiel, Germany

Acknowledgments: we thank Ludger Klein-Hitpaß for generating the SNP-chip profiles and Claudia Becher for support with the FISH analyses.

Funding: this work was supported by the Deutsche Krebshilfe through grant 107532.

Manuscript received on April 26, 2010. Revised version arrived on August 26, 2010. Manuscript accepted on August 26, 2010.

Correspondence: Ralf Küppers, Institute of Cell Biology (Cancer Research), University of Duisburg-Essen, Medical School, Virchowstr. 173, 45122 Essen, Germany. Phone: international +49.201.7233384. Fax: international +49.201.7233386. E-mail: ralf.kueppers@uk-essen.de

The online version of this article has a Supplementary Appendix.

ABSTRACT

Background
Knowledge about the genetic lesions that occur in Burkitt's lymphoma, besides the pathognomonic *IG-MYC* translocations, is limited.

Design and Methods
Thirty-nine molecularly-defined Burkitt's lymphomas were analyzed with high-resolution single-nucleotide polymorphism chips for genomic imbalances and uniparental disomy. Imbalances were correlated to expression profiles and selected micro-RNA analysis. Translocations affecting the *MYC* locus were studied by fluoresence *in situ* hybridization.

Results
We detected 528 copy number changes, defining 29 recurrently imbalanced regions. Five hundred and eighteen regions of uniparental disomy were found, but these were rarely recurrent. Combined imbalance mapping and expression profiling revealed a strong correlation between copy number and expression. Several recurrent imbalances affected the MYC pathway: the micro-RNA-supercluster 17-92 was frequently gained and the transcription factor E2F2 was recurrently deleted. Molecular Burkitt's lymphoma lacking *MYC* translocations showed *MYC* gains. Amplifications of the polymerase iota gene were associated with increased frequency of positions scored as aberrant.

Conclusions
The present findings suggest that uniparental disomies do not play a major role in the pathogenesis of Burkitt's lymphoma, whereas some genes may contribute to the development of this lymphoma through gene dosage effects. Amplifications of the polymerase iota gene may be functionally linked with increased genomic alterations in Burkitt's lymphoma. The pattern and rarity of chromosomal changes detectable, even at the high resolution employed here, together with aberrations of genes regulating MYC activity, support the hypothesis that deregulation of the MYC pathway is the major force driving the pathogenesis of Burkitt's lymphoma, but show that this deregulation is more complex than previously known.

Key words: Burkitt's lymphoma, genomic imbalance, MYC, uniparental disomy.

Citation: Scholtysik R, Kreuz M, Klapper W, Burkhardt B, Feller AC, Hummel M, Loeffler M, Rosolowski M, Schwaenen C, Spang R, Stein H, Thorns C, Trümper L, Vater I, Wessendorf S, Zenz T, Siebert R, and Küppers R, for the "Molecular Mechanisms in Malignant Lymphomas" Network Project of the Deutsche Krebshilfe. Detection of genomic aberrations in molecularly defined Burkitt's lymphoma by array-based, high resolution, single nucleotide polymorphism analysis. Haematologica 2010;95(12):2047-2055. doi:10.3324/haematol.2010.026831

Introduction

Burkitt's lymphoma (BL) is an aggressive mature B-cell lymphoma. Left untreated, it is fatal within months. There are three subtypes of BL: sporadic BL, mainly found in the Western world, endemic BL, mostly found in tropical Africa, and BL associated with immunosuppression.[1] About 90% of patients with endemic BL and 30% of those with sporadic BL are infected by Epstein-Barr virus (EBV), suggesting a pathogenic role of the virus, although only few viral genes are expressed in EBV$^+$ BL.[2]

The hallmark genetic lesion of BL is the so-called "Burkitt translocation" t(8;14)(q24;q32) and its variants t(8;22) and t(2;8), which juxtapose the *MYC* oncogene to one of the three immunoglobulin (*IG*)-loci.[2] One of these Burkitt translocations is present in almost all cases of BL investigated so far,[2] but is not specific because it is also found in other lymphoma types.[3] *IG* locus-driven *MYC* expression leads to strong proliferation signals, allowing the cells to grow rapidly.[2] However, strong expression of *MYC* also stimulates apoptotic pathways.[2] The BL clone must, therefore, acquire additional lesions to disrupt this signaling to benefit from the growth-enhancing effects of MYC. Some of the additional mutations described so far disrupt the p53-mediated pathway of apoptosis, by targeting *TP53* itself or components of its signaling cascade, allowing BL cells to evade external and internal death signals.[2] Compared to most other mature B-cell lymphomas, BL is characterized by the rarity of chromosomal aberrations secondary to the *IG-MYC* fusion. The most frequent secondary changes in BL detected by conventional cytogenetics are gains in 1q and in chromosomes 7 and 12.[3]

The current knowledge on secondary chromosomal changes in BL predominantly relies on analyses performed by conventional cytogenetics and comparative genomic hybridization (CGH) to chromosomes or low-resolution arrays.[4-8] These techniques are not suitable for detecting chromosomal imbalances at high-resolution and might, therefore, have failed to detect pathogenically important changes. Moreover, they are also not suitable for detecting chromosomal changes not causing structural alterations, such as uniparental disomies (UPD). UPD were recently detected in various cancers[9] and have been proposed to be an alternative mechanism to down-regulate tumor suppressor genes by duplication of inactivating mutations or deletions on one allele.[10,11] Finally, previous studies have focused on cases not classified by gene expression. As we recently showed, a considerable number of mature aggressive B-cell lymphomas would have to be re-classified if molecular signatures are applied.[8] We, therefore, performed high-resolution single nucleotide polymorphism (SNP) chip analysis on 39 sporadic BL cases collected by the "Molecular Mechanisms of Malignant Lymphomas" (MMML) consortium and defined as "molecular Burkitt's lymphoma (mBL)".

Design and Methods

Collection and extraction of DNA

Biopsies were diagnosed by experienced pathologists from the MMML consortium according to WHO criteria.[1] Five cases of BL were used as "core" BL in the classification based on gene expression,[8] 29 were in the "atypical" group. A tumor cell content of at least 70% was an inclusion criterion. Whole tissue DNA of tumor sections was extracted with the QiaAmp DNA Blood Kit according to the manufacturer's manual (Qiagen, Hilden, Germany). Central approval for the MMML network was obtained through the institutional review board of the University of Göttingen (D403/05).

Single nucleotide polymorphism chips

All 39 samples were analyzed on 250k Sty GeneChips (Affymetrix, Santa Clara, CA, USA), and 30 of these with enough material were additionally hybridized to Nsp-chips for a combined resolution of approximately 500k SNP. Each 250k GeneChip was prepared and hybridized according to the Affymetrix manual. The GeneChips were scanned by a GeneChip scanner 3000 with G7 update (Affymetrix). The sample files were genotyped with the BRLMM-algorithm (*see below*). The SNP-chip files have been submitted to the GEO database under accession number GSE21597.

Genotyping and copy number analysis

The BRLMM algorithm[12] was applied with default parameters (*score threshold*=0.5, *prior size*=10000 and *DM threshold*=0.17) to genotype mBL and 63 non-mBL tumor samples (*unpublished data*) using 39 Hapmap samples provided by Affymetrix (*http://www.affymetrix.com/support/technical/sample_data/500k_data.affx*) as a reference. The reference set was complemented by 20 laboratory-specific reference samples (11 female, 9 male) for the Sty array and 10 (6 female, 4 male) for the Nsp array. We supplemented the Hapmap samples with laboratory-specific samples because the Copy Number Analyzer for GeneChips (CNAG) software selects the reference samples that minimize the signal variance, so that a larger collection of references results in a lower "noise", and the laboratory-specific references should better eliminate a laboratory-specific signature. Median call rates of the mBL tumor samples were 96.23% and 98.96% for Sty- and Nsp-arrays, respectively (range, 90.44-98.63% and 90.26-99.65%).

Copy number analysis was performed using the CNAG program v2.0,[13] employing the same reference samples as for genotyping. CNAG was configured to select an optimal gender-specific reference set individually for each array.[13] This resulted in selection of 97.6% of the laboratory-specific references for Sty-arrays and 81.6% for Nsp-arrays. The lower proportion for Nsp possibly reflects the lower number of references for the Nsp-samples, giving the software less variance in references to choose from. For samples with Sty and Nsp arrays available, data from both chips were combined. Segmentation of raw copy number data was performed using the hidden Markov model (HMM) approach provided by CNAG.

HMM parameters were adjusted individually for each array to adapt the segmentation to differences in hybridization quality and tumor cell content of the analyzed samples. Starting with default parameters, the mean levels of HMM states were adjusted to optimize the segmentation results, i.e. to avoid missing clearly aberrant regions, as well as to prevent frequent successive alternation between neighboring HMM states.

With regard to outliers and technical artifacts, HMM segments were considered as copy number aberration only if they consisted of at least five consecutive imbalanced SNP. A preliminary analysis with a more stringent limit of ten consecutive SNP resulted in the identification of the same 33 recurrent regions, and only in small shifts in the borders of six of these regions (see Genomic Identification of Significant Targets in Cancer [GISTIC] analysis), due to exclusion of one imbalanced case for each of these six regions. This shows the robustness of the approach. High level amplifications were defined as aberrations with an HMM copy number of at least five, homozygous deletions as aberrations with a copy number of zero.

For the chromosome X in males (except for pseudo-autosomal region 1), interpretation of $\log_2$ ratios had to be adjusted with respect to the gender-specific (single copy) reference. Therefore, segments with an estimated HMM copy number of one to three copies were assumed to be normal (copy number n=1). Segments with an estimated copy number of at least four were selected as copy number gains and segments with an estimated copy number of zero as losses.

Loss of heterozygosity and uniparental disomy analysis

An HMM-based method[14] implemented in the dChip program[15,16] (Build date: Apr 11, 2007) was used to infer regions with loss of heterozygosity (LOH). The "HMM considering haplotype" (LD-HMM)[14] method was selected for the LOH calculations to account for linkage disequilibrium-induced SNP dependencies. The LOH call threshold was set at 0.99, applying an empirical haplotype correction.[14] Thus, putative LOH regions were excluded if there was 95% concordance of the homozygous genotypes of the candidate LOH region with respective regions of more than 5% of the reference samples. For samples for which both Sty and Nsp arrays were available, the combined Sty/Nsp set of 39 Affymetrix Hapmap samples was used as the normal reference set. The complete set of Sty reference arrays was selected as the reference for tumor samples without Nsp array data.

LOH regions were called UPD if no copy number aberrations were present in the region. In LOH regions partially affected by copy number aberrations, subregions without copy number aberrations were classified as UPD if they comprised 50 or more neighboring SNP.

Test for recurrence using Genomic Identification of Significant Targets in Cancer

For each imbalance detected by CNAG, the mean $\log_2$ ratio was determined. For gains and losses, the segment level was set to 0.1 and -0.1 if the mean $\log_2$ ratio was less than 0.1 or greater than -0.1, respectively. The segment level of balanced regions was set to 0. Using these segment values we selected recurrent gains and losses using GISTIC with standard parameters and thresholds.[17] Regions with known copy number polymorphisms were filtered by GISTIC using the Database of Genomic Variants (Version: June 2008).[18] Each recurrent aberration identified by GISTIC was represented by a region with the highest G-score (peak region) and a robustified wide peak region.[17] If the majority of the markers showed a concordant copy number status for the peak or the wide peak region, the respective region in each of the affected cases was classified as present.

Correlation of copy number and gene expression

We compared the mRNA expression level of genes within a recurrent region detected by GISTIC between cases with aberrant and balanced genomic status using Student's t-test, employing Hg17 (NCBI Build 35) for annotation, considering only regions with more than two contributing samples. Gene expression profiles of BL have been published.[8] Due to the small sample size we combined gains and high level amplifications. Adjustment for multiple testing within each recurrent region was performed with the step-down minP method implemented in the R-package multtest[19] using a family wise error rate of 5%.[20]

RNA extraction, reverse transcription and real-time polymerase chain reaction quantification of micro-RNA

RNA extraction, reverse transcription and real-time polymerase chain reaction (PCR) quantification were performed as described previously.[21] Briefly, the RecoverAll kit (Ambion, Austin, Texas, USA) was used for RNA-extraction from four 20 µm sections of formalin-fixed, paraffin-embedded tissues. The TaqMan® universal PCR master mix, No AmpErase® UNG-kit and the TaqMan® microRNA (miRNA) reverse transcription kit from Applied Biosystems (Foster City, California, USA) were used for cDNA synthesis and quantitative PCR. The reactions were performed in a 384-well format for 377 different miRNA (U6 in quadruplicate and miR16 in quintuplicate). The profiles were measured on the LightCycler® 480 instrument (Roche, Basel, Switzerland). To normalize the obtained raw CT values, we shifted the values of each sample such that the means of the probes of U6 and miR-16 were constant across all samples. That constant was chosen to be the mean of the means of U6 and miR-16.

Results

We generated 250k Sty SNP chips from 39 mature, aggressive B-cell lymphomas that had an mBL index ≥0.95 according to gene expression. For 30 of the tumors additional 250k Nsp chips were generated for a combined interrogation of 500k SNP, employing a resolution not previously attained for BL. Thirty-five of 38 evaluated tumors had an *IG-MYC* translocation detectable by fluorescence *in situ* hybridization (FISH).[8,22] Histological diagnoses by experienced pathologists did not always agree with the gene expression signature of mBL, with the series encompassing 27 BL, 6 centroblastic diffuse large B-cell lymphomas (DLBCL), one follicular lymphoma grade 3b, and 5 unclassifiable lymphomas. Twenty-four patients were 16 years or younger at the time of diagnosis.

As signals from single SNP probes are rather noisy, we only considered aberrations detected by at least five consecutive SNP for copy number analysis, resulting in a median detection window size of 33.8 kb and 16.6 kb for 250k and 500k arrays, respectively. Overall, the analysis detected 484 gains, 388 losses and 518 UPD before filtering for known copy number variations, combining the datasets from both arrays. Filtering left 308 gains and 220 losses that had less than 50% marker overlap with anno-

Table 1. Overview of the aberrations found in 39 cases of BL, after filtering for copy number polymorphisms.

Sample set	9 BL (Sty chips, 250k)			30 BL (Sty+Nsp chips, 500k)			All 39 BL
	Total	Mean	Median	Total	Mean	Median	Median size (bp)
Copy number changes	93	10.3	6	435	14.5	8.5	523,233
- Gains	63	7	3	245	8.2	4	362,547
- Losses	30	3.3	3	190	6.3	2.5	1 041,924
Uniparental disomies	58	6.4	4	460	15.3	14.5	770,662

A region was defined as aberrant if the HMM assigned a concordant copy number other than two to at least five consecutive SNP. Filtering for copy number polymorphisms was done using the Database of Genomic Variants (Version: June 2008).[40] As no corresponding germline DNA was available, UPD were detected by a statistical approach. The smallest detected aberration was a gain of 886 bases.

tated copy number polymorphisms (Table 1, Figure 1). Of these, 32 were high level amplifications and 6 were homozygous deletions (*Online Supplementary Tables S1 and S2*). Concerning the size of the copy number changes, the smallest aberrant region was 886 bases long and the median size was 522,475 bases (Table 1). The fraction of SNP detected in the normal copy number state of two was calculated to be more than 99% for 15 of the 39 tumors, and more than 95% for an additional 14 samples.

After filtering, data processing and testing for recurrent events, we detected 13 gains and 16 losses passing the significance threshold given by GISTIC (Tables 2 and 3; *Online Supplementary Figure S1*; complete annotation of genes in *Online Supplementary Tables S3 and S4*). It was not possible to calculate recurrence for UPD, as the tumors had too few overlapping UPD to reach a meaningful threshold. Regions with at least a few overlapping broad UPD encompass major parts of both arms of chromosomes 1 and 17, and the chromosomal arms 6p and 13q (Figure 1). The highest recurrence in gains (peak limit region aberrant in ≥7 cases) was detected for 1q25.1, 1q31.3, 3q27.2, 6q15, 11q24.3, and 13q31.3 (Table 2). The most frequent regions for losses of genetic material (peak limit region aberrant in ≥6 cases) were 3q13.13, 17p13.1, 19q13.42 and Xp22.33 (Table 3). A recent conventional CGH study which also employed a molecular definition of BL,[6] also described gains on 1q, 8q24-ter and 13q31-q32, and a recurrent loss on 17p (Tables 2 and 3).

Some regions contained genes with already known relevance for BL or cancer in general. For example, the gains at 3q27.3 (7 cases) always involved *BCL6*. Although a deregulation by translocation for this oncogene was described for large B-cell lymphomas,[23] none of the affected BL samples showed a translocation detectable by FISH.[8] Increased expression of *BCL6* was seen in affected cases (Figure 2B). Thus, BCL6 might be up-regulated solely by genomic gains in BL. Deleted regions involving or near previously known tumor suppressors were 9p21.3 (in only 2 cases) and 17p13.1, as these regions harbor *CDKN2B* and *TP53*, respectively, two well described tumor suppressor genes in BL.[24,25] While the recurrent loss on 9p21.3 narrowed exactly down to the *CDKN2B* gene, the target in the region 17p13.1 is unclear: four cases had lost one copy of the entire arm of chromosome 17, thereby deleting a copy of *TP53*. We also detected a partial UPD overlapping with the entire chromosomal region in three further cases. However, three additional cases had smaller heterozygous deletions, approximately 2 Mb away from the *TP53* gene,

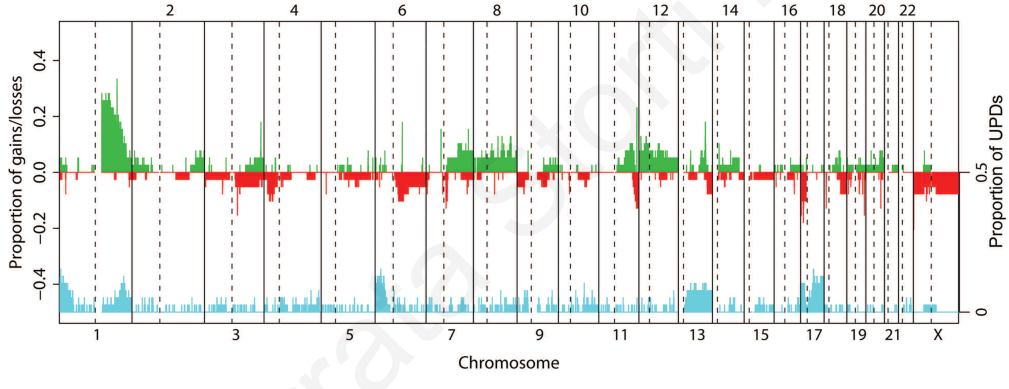

Figure 1. Proportion of gains, losses and UPD along chromosomes. y-axis: proportion of cases with the respective aberration, to the left copy number, to the right partial UPD; green: gains; red: losses; blue: UPD; dotted vertical lines depict centromeres. SNP that mapped to regions of known copy number variations and partial UPD that were also present in normal reference samples were excluded (see *Design and Methods* section).

Table 2. Recurrent gains in 39 BL as defined by GISTIC.

Cytoband	*q*-value[1]	Wide peak[2]	Peak limit[2]	N. wide peak	N. peak limit	Selected genes[3]
1q25.1	6.38E-06	chr1:166 155 454-171 863 618	chr1:171 851 737-171 863 618	10	10	*FASLG*
1q31.3	7.93E-06	chr1:195 253 751-195 672 280	chr1:195 369 624-195 638 957	12	13	*PTPRC*
3q27.3	0.058	chr3:188 956 472-189 175 357	chr3:189 056 204-189 167 011	6	7	*BCL6*
6q15	0.095	chr6:91 249 512-91 370 379	chr6:91 269 731-91 295 723	6	7	*MAP3K7*
7p12.2	0.12	chr7:49 995 112-50 127 527	chr7:49 995 112-50 127 527	6	6	*IKZF1*
7q34	0.11	chr7:141 162 879-142 244 396	chr7:141 162 879-142 244 396	6	6	
8q24.13	0.16	chr8:121 784 261-129 821 355	chr8:126 106 216-126 249 512	4	5	*MYC*
11q23.3	0.11	chr11:115 828 212-118 580 638	chr11:117 008 605-118 321 365	3	3	
11q24.3	0.012	chr11:127 825 348-127 931 093	chr11:127 825 348-127 919 072	9	9	*ETS1*
12q15	0.0051	chr12:61 724 591-82 263 925	chr12:67 278 531-68 393 145	3	4	
12q24.33	0.19	chr12:128 465 362-132 449 811	chr12:131 860 026-132 008 369	2	3	
13q31.3	1.16E-05	chr13:90 362 259-90 843 579	chr13:90 766 137-90 811 413	7	7	*hsa-mir-17-92*
18q21.2	7.75E-05	chr18:49 301 605-55 260 237	chr18:49 999 396-53 164 071	2	3	*POLI*

[1]significance as reported by GISTIC; [2]for the meaning of different regions see the Design and Methods section; [3]for a complete list of all annotated genes see Online Supplementary Table S3.

each including *STX8*, *USP43* and *WDR16* as the only annotated genes. In our gene expression data, only *STX8* is represented on the chip and is significantly down-regulated in cases with deletion (Figure 2A). Deletions involving 13q14.3 and/or 13q34 were recently observed by FISH in over 40% of childhood BL.[26] We, however, detected a recurrent loss centered at 13q32 only in 3/39 (7.7%) cases.

Concerning *MYC*, we detected a recurrent gain in five cases at 8q24.13, the location of this gene. Notably, the expression of *MYC* is not significantly different between cases with the respective gain and cases with an *IG-MYC* translocation, which is known to cause strong MYC over-expression (Figure 2A). It is interesting that three out of 39 BL cases had no detectable *MYC* translocation in FISH,[8,22] although this does not exclude cryptic insertions of *MYC* into an *IG* locus or vice versa. Nevertheless, they were scored as mBL by gene expression.[8] Two of these are among the cases with *MYC* gain.

The gain in 1q31.3, found in 13 cases, contained the gene *PTPRC*, a regulator of B-cell receptor and cytokine signaling,[27] and two annotated miRNA genes (hsa-mir-181b-1 and -213). The high recurrence of this gain hints at its importance for the tumor. Because the gene expression data showed no significant up-regulation of *PTPRC* in affected cases, and the gains often involved not the complete coding sequence of this gene (*data not shown*), the miRNA genes are strong target candidates.

The region gained on 13q31.3 in seven BL harbored the miRNA-17-92 supercluster. Only one of these cases showed a high level amplification at this location. For some of the BL analyzed (but not for the single case showing the amplification), miRNA expression profiles for their mature forms were recorded. The miRNA contained in the supercluster were consistently expressed at higher levels in cases with the respective gain (*Online Supplementary Table S5*), though this failed to be statistically significant after correcting for multiple testing. Three cases harbored losses at 1p36.12 (and 4 further cases had large UPD involving the same region) which contained the gene *E2F2*, a transcription factor that binds to the promoter of the miRNA-17-92 cluster.

Three BL showed a high level amplification of 18q21.2. Remarkably, these three are first, third and seventh in a ranked list of all 39 BL ordered according to the proportion of aberrant SNP (Figure 3). The gene expression data for this region (*Online Supplementary Figure S2*) show that the *POLI* gene was significantly up-regulated, as were three other genes with unclear relevance.

Using expression data available for all BL,[8] we compared the expression of genes in a recurrent region in cases with an aberration detected by SNP-chip to cases without the respective aberration, to pinpoint gene dosage effects of possible oncogenes or tumor suppressors, if present. Some exemplary regions are shown in Figure 2. We detected a clear concordant effect of copy number on expression strength: a gain led preferentially to a stronger expression of a large fraction of gained genes, with 12 of 108 probe-sets even reaching statistical significance after correction for multiple testing (and not a single significantly down-regulated gene). Likewise, the majority of genes in a heterozygously lost region displayed reduced expression (52 of 258 probe-sets significant after correction, 2 inversely correlated). A statistically significant positive correlation was, therefore, detected in 17.5% of all probe-sets in aberrant regions.

Discussion

Cytogenetic and low-resolution CGH studies previously performed on BL revealed few consistent genomic imbalances.[4,5] A caveat regarding these studies was the definition of BL on non-molecular grounds. These earlier studies probably included unrecognized molecular DLBCL in their

Table 3. Recurrent losses in 39 cases of BL as defined by GISTIC.

Cytoband	*q*-value[1]	Wide peak[2]	Peak limit[2]	N. wide peak	N. peak limit	Selected genes[3]
1p36.12	0.014	chr1:23,457,835-23,714,048	chr1:23,535,005-23,673,234	2	3	*E2F2*
3p14.2	0.15	chr3:60,396,160-60,637,030	chr3:60,452,472-60,580,245	3	3	*FHIT*
3q13.13	0.085	chr3:110,456,383-110,657,226	chr3:110,482,177-110,611,242	6	6	*DPPA4, DPPA2*
4p15.32	0.15	chr4:17,969,802-29,966,659	chr4:17,991,350-18,456,461	3	4	
6q14.3	0.094	chr6:76,830,186-107,898,353	chr6:78,151,877-96,248,104	4	4	
7q11.22	0.15	chr7:66,573,850-67,281,794	chr7:66,810,706-67,240,659	5	5	
10p11.21	0.22	chr10:36,983,732-37,133,784	chr10:37,007,511-37,106,346	3	3	
11q24.3	0.085	chr11:122,263,650-134,452,384	chr11:126,328,752-132,876,984	4	5	
13q33.2	0.20	chr13:96,008,275-114,142,980	chr13:96,913,194-112,193,013	3	3	
17p13.1	0.014	chr17:9,125,548-9,654,733	chr17:9,125,548-9,598,824	7	7	*STX8, TP53*
18p11.23	0.094	chr18:7,079,886-7,359,571	chr18:7,125,227-7,359,571	5	5	
18q22.3	0.15	chr18:55,355,976-76,117,153	chr18:55,355,976-76,117,153	3	3	
19q13.42	0.085	chr19:58,746,599-59,029,391	chr19:58,758,793-58,990,786	6	6	
20q13.2	0.094	chr20:49,856,082-50,168,587	chr20:49,958,057-50,144,808	5	5	
Xp22.33	0.0042	chrX:1-1,499,465	chrX:1-1,499,465	8	8	
Xp11.3	0.085	chrX:1-154,824,264	chrX:46,975,370-47,150,005	3	3	

[1]significance as reported by GISTIC; [2]for the meaning of different regions, see the Design and Methods section; [3]for a complete list of all annotated genes, see Online Supplementary Table S4.

BL set, which contributed changes that newer studies do not recognize in mBL. The single high-resolution oligonucleotide array-CGH analysis in BL included, besides cell lines, only 13 primary tumors which were analyzed by a 44k array,[7] limiting resolution and sensitivity. Furthermore, lymphomas were selected as BL based only on histological diagnosis.[7] In our approach using molecularly-defined BL, we detected several of the aberrations that this paper found in only one to three histological BL. In most case we were able (due to the larger number of samples and the higher resolution) to set narrower limits to these regions and confirm them as recurrent events in primary tumors.

Concerning our published CGH analysis,[8,22] at a resolution of 3000 probes of a subset of cases also analyzed here, we note a good concordance between the two methods. As expected, the SNP analysis detected multiple additional aberrations that the CGH could not detect because of res-

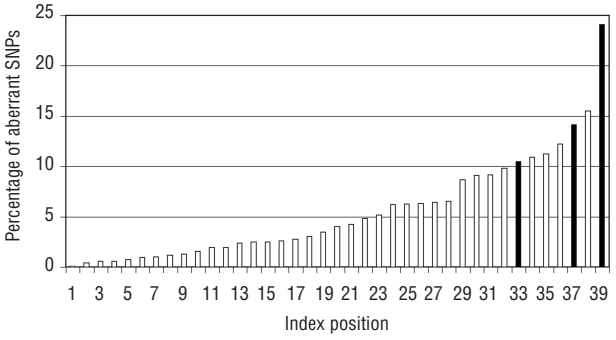

Figure 3. Correlation of the recurrent gain at chr18 involving *POLI* and the proportion of aberrant SNP. In a list ordered according to the proportion of aberrant SNP detected (y-axis), the three cases with an amplification involving *POLI* rank first, third and seventh (columns marked in black; *P*=0.0035 in a Mann-Whitney U test).

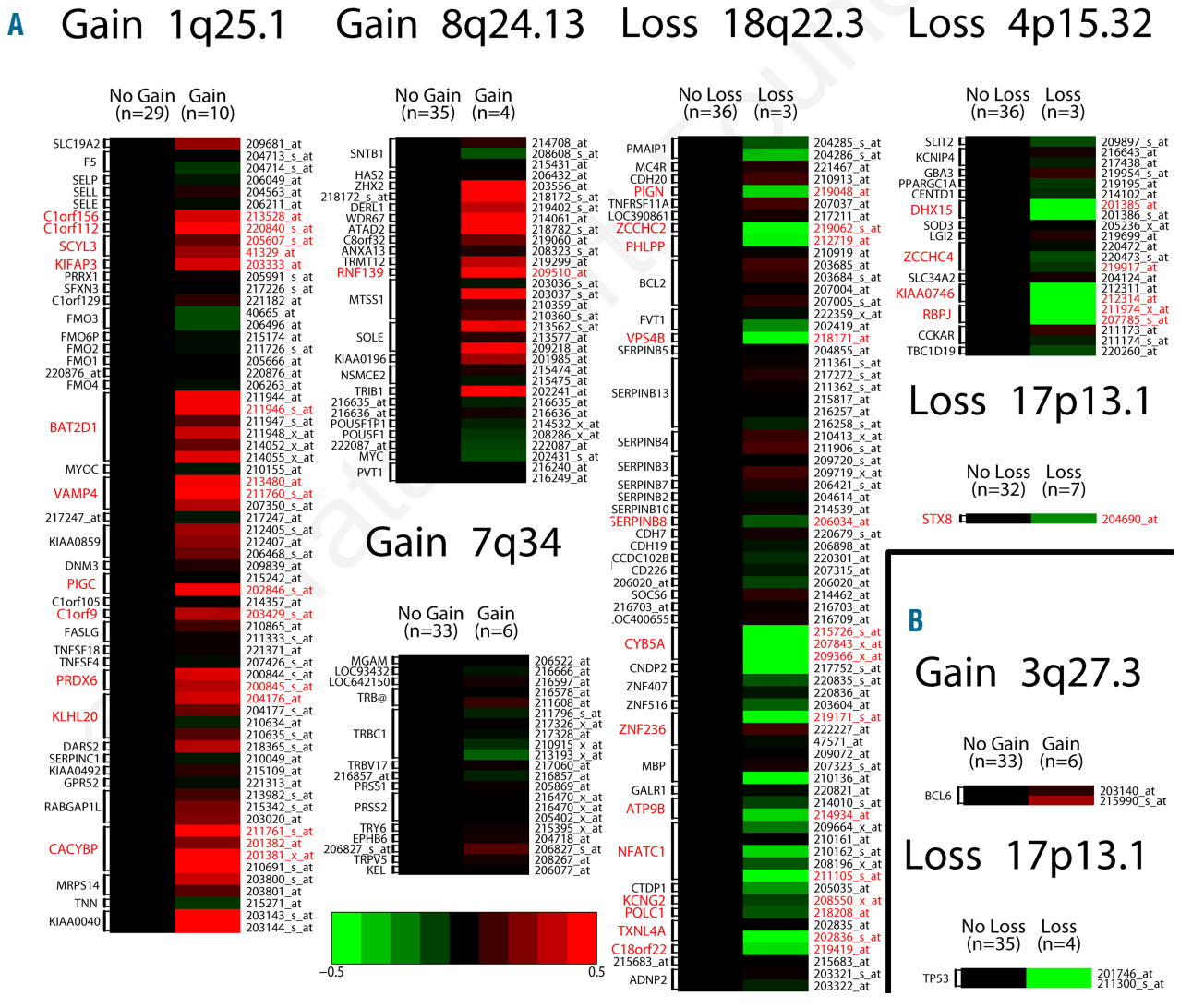

Figure 2. Gene dosage effects for genes in recurrently imbalanced regions. Green: reduced expression; red: stronger expression. The difference in expression between aberrant and non-aberrant cases is displayed on a log-scale. For all regions analyzed except 7q34, a clear gene dosage effect is visible. Gene names marked in red indicate significant differences in expression after correcting for multiple testing. (A) whole recurrently aberrant regions depicted; (B) single genes of interest out of a larger region.

olution constraints, but the raw signal values for the CGH clones gave the correct tendency for each aberration detected with SNP chips (*unpublished data*). Regarding high level amplifications and homozygous losses before filtering for copy number polymorphisms, 22 concordant regions were identified with both platforms. Importantly, however, 35 additional imbalances were only seen with the SNP-chips as the respective regions were not covered on the CGH-array, and in ten further instances of imbalances detected by SNP-chip analysis, CGH signals classified the regions as balanced.

Regarding the extent of genomic lesions in BL as compared to in other aggressive lymphomas, e.g. DLBCL, a direct comparison is currently not possible due to lack of data from identical high-resolution platforms. Nevertheless, an array CGH study with 3000 BAC clones detected a median of 8.5 copy number changes for DLBCL,[8] while, with our higher resolutions, the median copy number changes we found in BL was 6.0 with the 250k array and 8.5 with the 500k array (Table 1). Recent publications on comparable high-resolution SNP-chip approaches to mostly solid tumors reported more copy number changes per case (e.g. a median of 12 losses and 12 gains; although it should be considered that a change of >0.1 copies was already considered aberrant in one of these studies).[28,29] We, therefore, confirm that BL has a relatively stable genome.

An important advantage of the SNP chip method compared to CGH is the former's ability to detect UPD, with the caveat that the exact detection of these disomies was hampered in this study as no normal tissue DNA was available for germline comparison. Nevertheless, using a statistical approach, robust pinpointing of larger stretches of LOH is possible. In several tumors, UPD are an additional way to inactivate tumor suppressor genes,[9,10] beside deletions and/or mutations. Thus, the integration of losses and UPD may allow better detection of new candidate tumor suppressors. Instead, we show that in BL there is little congruency between UPD and regions recurrently lost and no strong accumulation of UPD in specific regions (Figure 1), suggesting that UPD do not play a major role in BL. Nevertheless, the few UPD detected here may still be of functional relevance in the affected cases, as exemplified for *TP53*, for which cases with UPD were found besides cases with deletions.

Regarding the differential gene expression analysis, we detected a clear gene dosage effect for the recurrently aberrant regions (Figure 2). The only exception to this is the recurrent gain at 7q34, which includes nine genes, but not a single one with a strong dosage effect. This underlines that the gain and loss of genetic material can be a potent mechanism to deregulate certain genes in BL, in line with the results of a recent study comparing array-CGH data with gene expression levels.[6]

Four cases that had a large deletion on chromosome 17 retained only a single copy of *TP53*, encoding the p53 tumor suppressor. These four cases showed lower *TP53* gene expression compared with the 35 unaffected cases (Figure 2B). Notably, three further cases showed UPD encompassing *TP53*, suggesting that these events may also have pathogenic relevance. Indeed, two of two cases with deletion and two of two cases with UPD that were sequenced had inactivating *TP53* mutations (*data not shown*). The exact target of smaller deletions on chromosome 17 in three additional cases was more ambiguous: the deletion was approximately 2 Mb away from *TP53* and included three annotated genes (*STX8*, *USP43* and *WDR16*) that were heterozygously lost in each affected case. However, as none of these genes shows differential expression between B-cell subsets (*data not shown*) and there were no previous reports about their functional involvement in cancer, the relevance of these deletions is unclear.

A significantly up-regulated gene in the amplified region 18q21.2, *POLI*, codes for DNA polymerase iota which plays a role as an error-prone polymerase in somatic hypermutation[30] and some forms of DNA-damage response.[31] We speculate that over-expression of POLI shifted the balance of DNA repair to more error-prone repairs, which would explain the high percentage of aberrant SNP detected in these cases (Figure 3). The expression of *POLI* was also positively correlated with an increased percentage of aberrant SNP in our set of tumors (Spearman's correlation=0.3; P=0.06). In breast cancer cells, a higher amount of POLI leads to a higher mutation burden,[32] suggesting a similar role in BL. A recent study in mice actually implicated *POLI* as a new oncogene in a colorectal cancer model.[33]

We found that the *MYC* gene was not only targeted by translocation events, but additionally by gains of genetic material in five of the 39 BL. The gene expression was not different in mBL with the gain, but this is explainable by the fact that, to become a mBL, the up-regulation of *MYC* is obligatory.[8] Indeed, we identified three mBL that had no detectable *MYC* translocation by FISH ([8] and *data not shown*), but two of these showed a low-level gain including *MYC*. These cases were nevertheless classified as mBL.[8] The most frequent deregulation of *MYC* is, therefore, by translocation to another enhancer (such as *IG-MYC*), but we suppose that the gain of genomic material is an additional mechanism by which the expression of this important driver gene can be increased. The three BL with a gain in addition to the translocation probably represent unbalanced translocation events. For the remaining single case without translocation and without gain, the mechanism of deregulation is unknown. *TP53* deletions, which may hamper the pro-apoptotic activity of MYC, were not specifically associated with BL harboring *MYC* gains.

Regarding MYC deregulation, gain of the miRNA-17-92 supercluster in seven cases is of note, because it is directly transactivated by MYC.[34] This cluster has been shown to be frequently over-expressed in several types of solid tumors[35] and also B-cell malignancies,[36] in which it is also often amplified.[37] Over-expression of this cluster in mouse lymphocytes reduces apoptosis and accelerates lymphoma development,[38] defining it as an oncogene.[37,39,40] In addition, it was recently shown that these miRNA act synergistically with MYC in the development of aggressive cancer.[36] They provide the tumor with a means to counteract activation-induced apoptosis caused by the high *MYC* expression in BL, by reducing the expression of tumor suppressor genes, for example *TGFBR2*,[36] *BIM* and *PTEN*.[38,40]

The miRNA-17-92 supercluster is controlled by an autoregulatory loop with E2F-family proteins.[41] In our series, three cases had heterozygous deletion of the *E2F2* gene on 1p36.12, which directly binds the promoter of the supercluster and increases its expression. The three cases with *E2F2* deletion had lower mRNA levels of the gene than the other cases, although the difference did not reach

statistical significance (*data not shown*). The deletion of a positive regulator of the miRNA cluster at first seems counterintuitive, as other cases in our series had specifically amplified miRNA polycistron, hence over-expression of these miRNA seems to be advantageous for the clone. Besides regulating the miRNA cluster 17-92, *E2F2* thus presumably has other effects on the tumor which promoted the deletion. It was indeed shown that *E2F2* functions as a tumor suppressor in B cells of mice over-expressing *MYC*.[42] The mice with heterozygous *E2F2* deletions show strong haploinsufficiency, resulting in accelerated development of *MYC*-driven B-cell lymphomas, suggesting that also in human BL, the heterozygous losses could have a strong effect on the tumor clone. We thus propose that the heterozygous deletion of *E2F2* would indeed tend to decrease the expression of the miRNA-17-92 supercluster, but this potential down-regulation is overcompensated in BL by the high amount of MYC, a direct transactivator of this supercluster. In this way, the tumor can benefit from the remaining effects of reduced *E2F2* discussed above. In our dataset, gain of the miRNA-cluster 17-92 and deletion of *E2F2* were mutually exclusive, hinting at the possibility that either event is sufficient to deregulate the feedback loop.

In summary, we refined the make-up of known genetic events in BL with high resolution, and found many new recurrent aberrations, whose functional significance is mostly unclear so far. Some of these newly identified lesions affect the MYC pathway, suggesting that this pathway is deregulated in a more complex fashion than previously thought.

Authorship and Disclosures

The information provided by the authors about contributions from persons listed as authors and in acknowledgments is available with the full text of this paper at www.haematologica.org.

Financial and other disclosures provided by the authors using the ICMJE (www.icmje.org) Uniform Format for Disclosure of Competing Interests are also available at www.haematologica.org.

References

1. Swerdlow SH. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues (4th ed). Lyon: International Agency for Research on Cancer; 2008.
2. Hecht JL, Aster JC. Molecular biology of Burkitt's lymphoma. J Clin Oncol. 2000;18(21):3707-21.
3. Boerma EG, Siebert R, Kluin PM, Baudis M. Translocations involving 8q24 in Burkitt lymphoma and other malignant lymphomas: a historical review of cytogenetics in the light of todays knowledge. Leukemia. 2009;23(2):225-34.
4. Barth TF, Muller S, Pawlita M, Siebert R, Rother JU, Mechtersheimer G, et al. Homogeneous immunophenotype and paucity of secondary genomic aberrations are distinctive features of endemic but not of sporadic Burkitt's lymphoma and diffuse large B-cell lymphoma with MYC rearrangement. J Pathol. 2004;203(4):940-5.
5. Garcia JL, Hernandez JM, Gutierrez NC, Flores T, Gonzalez D, Calasanz MJ, et al. Abnormalities on 1q and 7q are associated with poor outcome in sporadic Burkitt's lymphoma. A cytogenetic and comparative genomic hybridization study. Leukemia. 2003;17(10):2016-24.
6. Salaverria I, Zettl A, Bea S, Hartmann EM, Dave SS, Wright GW, et al. Chromosomal alterations detected by comparative genomic hybridization in subgroups of gene expression-defined Burkitt's lymphoma. Haematologica. 2008;93(9):1327-34.
7. Toujani S, Dessen P, Ithzar N, Danglot G, Richon C, Vassetzky Y, et al. High resolution genome-wide analysis of chromosomal alterations in Burkitt's lymphoma. PLoS One. 2009;4(9):e7089.
8. Hummel M, Bentink S, Berger H, Klapper W, Wessendorf S, Barth TF, et al. A biologic definition of Burkitt's lymphoma from transcriptional and genomic profiling. N Engl J Med. 2006;354(23):2419-30.
9. Tuna M, Knuutila S, Mills GB. Uniparental disomy in cancer. Trends Mol Med. 2009;15(3):120-8.
10. Nielaender I, Martin-Subero JI, Wagner F, Martinez-Climent JA, Siebert R. Partial uniparental disomy: a recurrent genetic mechanism alternative to chromosomal deletion in malignant lymphoma. Leukemia. 2006;20(5):904-5.
11. Vater I, Wagner F, Kreuz M, Berger H, Martin-Subero JI, Pott C, et al. GeneChip analyses point to novel pathogenetic mechanisms in mantle cell lymphoma. Br J Haematol. 2009;144(3):317-31.
12. Affymetrix. BRLMM: an improved genotype calling method for the genechip human mapping 500k array set. Affymetrix, Inc White Paper; 2006.
13. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res. 2005;65(14):6071-9.
14. Beroukhim R, Lin M, Park Y, Hao K, Zhao X, Garraway LA, et al. Inferring loss-of-heterozygosity from unpaired tumors using high-density oligonucleotide SNP arrays. PLoS Comput Biol. 2006;2(5):e41.
15. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics. 2004;20(8):1233-40.
16. Zhao X, Li C, Paez JG, Chin K, Janne PA, Chen TH, et al. An integrated view of copy number and allelic alterations in the cancer genome using single nucleotide polymorphism arrays. Cancer Res. 2004;64(9):3060-71.
17. Beroukhim R, Getz G, Nghiemphu L, Barretina J, Hsueh T, Linhart D, et al. Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. Proc Natl Acad Sci USA. 2007;104(50):20007-12.
18. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, et al. Detection of large-scale variation in the human genome. Nat Genet. 2004;36(9):949-51.
19. Pollard KS, Ge Y, Dudoit S. Multtest: Resampling-based multiple hypothesis testing, R package version 1.12.0, http://CRAN.R-project.org/package=multtest; 2006.
20. Kreuz M, Rosolowski M, Berger H, Schwaenen C, Wessendorf S, Loeffler M, et al. Development and implementation of an analysis tool for array-based comparative genomic hybridization. Methods Inf Med. 2007;46(5):608-13.
21. Roehle A, Hoefig KP, Repsilber D, Thorns C, Ziepert M, Wesche KO, et al. MicroRNA signatures characterize diffuse large B-cell lymphomas and follicular lymphomas. Br J Haematol. 2008;142(5):732-44.
22. Klapper W, Szczepanowski M, Burkhardt B, Berger H, Rosolowski M, Bentink S, et al. Molecular profiling of pediatric mature B-cell lymphoma treated in population-based prospective clinical trials. Blood. 2008;112(4):1374-81.
23. Sanchez-Beato M, Sanchez-Aguilera A, Piris MA. Cell cycle deregulation in B-cell lymphomas. Blood. 2003;101(4):1220-35.
24. Gaidano G, Ballerini P, Gong JZ, Inghirami G, Neri A, Newcomb EW, et al. p53 mutations in human lymphoid malignancies: association with Burkitt lymphoma and chronic lymphocytic leukemia. Proc Natl Acad Sci USA. 1991;88(12):5413-7.
25. Klangby U, Okan I, Magnusson KP, Wendland M, Lind P, Wiman KG. p16/INK4a and p15/INK4b gene methylation and absence of p16/INK4a mRNA and protein expression in Burkitt's lymphoma. Blood. 1998;91(5):1680-7.
26. Nelson M, Perkins SL, Dave BJ, Coccia PF, Bridge JA, Lyden ER, et al. An increased frequency of 13q deletions detected by fluorescence in situ hybridization and its impact on survival in children and adolescents with Burkitt lymphoma: results from the Children's Oncology Group study CCG-5961. Br J Haematol. 2010;148(4):600-10.
27. Hermiston ML, Zikherman J, Zhu JW. CD45, CD148, and Lyp/Pep: critical phosphatases regulating Src family kinase signaling networks in immune cells. Immunol Rev. 2009;228(1):288-311.
28. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The

landscape of somatic copy-number alteration across human cancers. Nature. 2010;463(7283):899-905.
29. Leary RJ, Lin JC, Cummins J, Boca S, Wood LD, Parsons DW, et al. Integrated analysis of homozygous deletions, focal amplifications, and sequence alterations in breast and colorectal cancers. Proc Natl Acad Sci USA. 2008;105(42):16224-9.
30. Faili A, Aoufouchi S, Flatter E, Gueranger Q, Reynaud CA, Weill JC. Induction of somatic hypermutation in immunoglobulin genes is dependent on DNA polymerase iota. Nature. 2002;419(6910):944-7.
31. Petta TB, Nakajima S, Zlatanou A, Despras E, Couve-Privat S, Ishchenko A, et al. Human DNA polymerase iota protects cells against oxidative stress. EMBO J. 2008;27 (21):2883-95.
32. Yang J, Chen Z, Liu Y, Hickey RJ, Malkas LH. Altered DNA polymerase iota expression in breast cancer cells leads to a reduction in DNA replication fidelity and a higher rate of mutagenesis. Cancer Res. 2004;64(16):5597-607.
33. Starr TK, Allaei R, Silverstein KA, Staggs RA, Sarver AL, Bergemann TL, et al. A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science. 2009;323(5922):1747-50.
34. O'Donnell KA, Wentzel EA, Zeller KI, Dang CV, Mendell JT. c-Myc-regulated microRNAs modulate E2F1 expression. Nature. 2005;435(7043):839-43.
35. Volinia S, Calin GA, Liu CG, Ambs S, Cimmino A, Petrocca F, et al. A microRNA expression signature of human solid tumors defines cancer gene targets. Proc Natl Acad Sci USA. 2006;103(7):2257-61.
36. Tagawa H, Karube K, Tsuzuki S, Ohshima K, Seto M. Synergistic action of the microRNA-17 polycistron and Myc in aggressive cancer development. Cancer Sci. 2007;98(9):1482-90.
37. Ota A, Tagawa H, Karnan S, Tsuzuki S, Karpas A, Kira S, et al. Identification and characterization of a novel gene, C13orf25, as a target for 13q31-q32 amplification in malignant lymphoma. Cancer Res. 2004;64 (9):3087-95.
38. Xiao C, Srinivasan L, Calado DP, Patterson HC, Zhang B, Wang J, et al. Lymphoproliferative disease and autoimmunity in mice with increased miR-17-92 expression in lymphocytes. Nat Immunol. 2008;9(4):405-14.
39. He L, Thomson JM, Hemann MT, Hernando-Monge E, Mu D, Goodson S, et al. A microRNA polycistron as a potential human oncogene. Nature. 2005;435(7043): 828-33.
40. Mu P, Han YC, Betel D, Yao E, Squatrito M, Ogrodowski P, et al. Genetic dissection of the miR-17~92 cluster of microRNAs in Myc-induced B-cell lymphomas. Genes Dev. 2009;23(24):2806-11.
41. Sylvestre Y, De Guire V, Querido E, Mukhopadhyay UK, Bourdeau V, Major F, et al. An E2F/miR-20a autoregulatory feedback loop. J Biol Chem. 2007;282(4):2135-43.
42. Rempel RE, Mori S, Gasparetto M, Glozak MA, Andrechek ER, Adler SB, et al. A role for E2F activities in determining the fate of Myc-induced lymphomagenesis. PLoS Genet. 2009;5(9):e1000640.