Published in final edited form as:

Nat Immunol. 2015 June ; 16(6): 653–662. doi:10.1038/ni.3148.

Tet1 is a tumor suppressor of hematopoietic malignancy

Luisa Cimmino[1,2,9], **Meelad M. Dawlaty**[3,4,9], **Delphine Ndiaye-Lobry**[1,2], **Yoon Sing Yap**[1,2], **Sofia Bakogianni**[1,2], **Yiting Yu**[5], **Sanchari Bhattacharyya**[5], **Rita Shaknovich**[6], **Huimin Geng**[7], **Camille Lobry**[1,2], **Jasper Mullenders**[1,2], **Bryan King**[1,2], **Thomas Trimarchi**[1,2], **Beatriz Aranda-Orgilles**[1,2], **Cynthia Liu**[1], **Steven Shen**[8], **Amit K. Verma**[5], **Rudolf Jaenisch**[3,4], and **Iannis Aifantis**[1,2]

[1]Howard Hughes Medical Institute and Department of Pathology, NYU School of Medicine, New York, NY, 10016, USA

[2]NYU Cancer Institute and Helen L. and Martin S. Kimmel Center for Stem Cell Biology, NYU School of Medicine, New York, NY, 10016, USA

[3]Whitehead Institute for Biomedical Research, Cambridge, MA, 02142, USA

[4]Department of Biology, Massachusetts Institute of Technology, Cambridge, MA, 02139, USA

[5]Department of Hemato-Oncology, Albert Einstein College of Medicine, Montefiore Medical Center, Bronx, NY, 10461, USA

[6]Department of Medicine, Weill Cornell Medical College, Cornell University, New York, NY, 10065, USA

[7]Department of Laboratory Medicine, University of California San Francisco, San Francisco, CA, 94143, USA

[8]Center for Health Informatics and Bioinformatics, NYU School of Medicine, New York, NY, 10016, USA

Abstract

The TET methylcytosine dioxygenase 1 (TET1) enzyme is an important regulator of 5-hydroxymethylcytosine (5hmC) in embryonic stem cells. Decreased expression of TET proteins and loss of 5hmC in many tumors suggests a critical role for the maintenance of this epigenetic modification. Here we show that deletion of *Tet1* promoted the development of B cell lymphoma

Correspondence should be addressed to I.A. (Iannis.aifantis@nyumc.org) or R.J. (jaenisch@wi.mit.edu).

[9]These authors have contributed equally to this work

ACCESSION CODES

The NCBI Gene Expression Omnibus accession numbers for the Microarray, RNA-seq, RRBS assays and HELP-GT assays in this paper are GSE65957 and GSE65895.

AUTHOR CONTRIBUTIONS

L.C., M.M.D., R.J. and I.A., designed and performed experiments. D.N-L. and S.Bak. and J.M. provided technical assistance. Y.S.Y. performed bisulfite-sequencing experiments. Y.Y., S.Bhat. and A.K.V., performed, analyzed and provided bioinformatics support for HELP-GT assays. R.S. and H.G., provided bioinformatics support of patient HELP-Assay, RNA-seq and R.S. performed MassArray experiments. T.T., B.A-O., B.K. and S.S. participated in bioinformatics analysis. C.L. provided histopathology advice. L.C. and I.A. wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

in mice. *Tet1* was required for maintaining normal content of 5hmC, preventing DNA hypermethylation and in the regulation of B cell lineage, chromosome maintenance and DNA repair genes. Whole-exome sequencing of Tet1-deficient tumors revealed mutations frequently found in Non-Hodgkin B cell lymphoma, where *TET1* was hypermethylated and transcriptionally silenced. These findings provide *in vivo* evidence of TET1 function as a tumor suppressor of hematopoietic malignancy.

Epigenetic pathways regulating DNA methylation and chromatin modifications are frequently found to be dysregulated in human cancers[1]. The Ten-eleven-translocation (TET1-3) proteins are a family of 2-oxoglutatrate (2OG) and Fe-(II) dependent dioxygenases that catalyze the oxidation of 5-methylcytosine (5mC) to sequentially generate 5-hydroxymethylcytosine (5hmC), 5-formylcytosine (5fC) and 5-carboxycytosine (5caC), modifications that have been shown to be intermediates in the regulation of DNA demethylation[2–4]. Decreased expression of TET proteins and loss of 5hmC has been reported in breast, colorectal, skin, stomach and lung cancer, suggesting a critical role for the maintenance of this epigenetic modification in normal cellular function[5,6].

Genome wide studies have shown 5hmC to be enriched at enhancers, promoters and gene bodies of actively expressed genes[7,8]. The presence of 5hmC may contribute to both passive and active DNA demethylation in the mammalian genome. Maintenance methylation mediated by DNA methyltransferase 1 normally follows replication, and may be unable to recognize 5hmC[9], causing 5mC to be lost passively during cell division. The TET proteins may also actively prevent DNA hypermethylation and promote demethylation by a sequential process involving AID- or APOBEC-mediated deamination of 5hmC to 5hmU followed by base excision repair (BER)[4]. TET-mediated oxidation of 5hmC to 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) may also trigger DNA demethylation by BER independently of deamination[10].

The *TET* gene family was first identified because of the involvement of *TET1* as a fusion partner of MLL in acute myeloid leukemia (AML)[11], a translocation event that has also been detected in patients with T cell lymphoma[12] and B-ALL[13]. *TET1* mutations are also found at a low frequency in AML (~1%)[14] compared to T cell acute lymphoblastic leukemia (T-ALL) (~14%)[15] although mutations in B cell malignancy have not been reported. Mutations in *TET2* occur in ~30% of patients with myeloproliferative neoplasms and acute myeloid leukemia (AML)[16,17] and loss of TET2 function is associated with aberrant DNA methylation in the hematopoietic system[18,19]. TET1 is known to be an important regulator of 5hmC in embryonic stem cells[7,20,21], adult[4] and reprogrammed cells[22,23]. However, loss of TET1 function in the etiology of cancer has not been directly investigated. Here we show that deletion of the *Tet1* gene in mice promoted the development of B cell lymphoma after an extended period of latency. Whole-exome sequencing of Tet1-deficient tumors revealed mutations frequently found in Non-Hodgkin B cell lymphoma (B-NHL) patients[24–27] where *TET1* was also shown to be hypermethylated and transcriptionally silenced. Hematopoietic stem and progenitors cells deficient in *Tet1* displayed decreased 5hmC, increased 5mC and altered expression of transcriptional programs involved in B cell-lineage specification, chromosome maintenance, and DNA repair. Loss of Tet1 promoted increased progenitor B

cell self-renewal *in vitro*, enhanced stem cell self-renewal *in vivo* and cooperates with Bcl2 overexpression to drive B lymphocytosis in mice. Tet1-deficient pro-B cells also displayed increased DNA damage. These findings provide the first *in vivo* evidence of TET1 function as a tumor suppressor of hematopoietic malignancy and more specifically B cell lymphoma.

Results

Tet1-deficiency drives B cell malignancy upon advanced age

We aged a cohort of Tet1-deficient animals[28] and monitored their health for over two years. The majority of heterozygous ($Tet1^{+/-}$) and homozygous ($Tet1^{-/-}$) Tet1-deficient mice remained healthy for up to 1 year of age after which a loss in survival was observed, with less than ~20% of $Tet1^{-/-}$ mice remaining viable at 2 years of age compared to ~70% of wild-type animals (Fig. 1a). Peripheral blood analysis of mice aged 18–24 months showed mild but significant lymphocytosis (Fig. 1b and Supplementary Fig. 1a–d). Upon necropsy, moribund Tet1-deficient mice often displayed severely enlarged peripheral lymph nodes (Fig. 1c) and hepato-splenomegaly. By histological analysis, spleens of Tet1-deficient animals revealed disrupted architecture with loss of delineation between the red and white pulp and extensive infiltration by atypical lymphocytes with high Ki67 positivity (Supplementary Fig. 1e). Immunostaining of cells in the peripheral blood, lymph nodes or spleen of moribund Tet1-deficient mice revealed that the majority displayed a mature B lymphoma phenotype of $SSC^{hi}B220^{+}CD19^{+}CD43^{-}IgM^{+}IgD^{hi/lo}$ cells (Fig. 1d). Co-expression of the myeloid lineage marker CD11b, and in some cases erythroid lineage markers was also observed (Supplementary Fig. 2a,b, Supplementary Table 1). Histological analysis of the liver, lung and kidney in Tet1-deficient mice revealed destructive tissue infiltration by atypical lymphocytes (Fig. 1e and Supplementary Fig. 1f,g). The lymph nodes of Tet1-deficient mice showed a diffuse staining pattern of lymphocytes by hematoxylin and eosin (Fig. 1f). In addition, these cells stained positive by immunohistochemistry for Bcl-6 and IRF4 that are typically expressed in mature germinal center B cells, and negative for CD138 (Syndecan-1), a marker of plasma cells (Fig. 1g). In approximately 20% of cases, multinucleated cells and histiocytic sarcomas were also observed in lymph nodes, spleens and infiltrated solid organs (Supplementary Fig. 2c,d). The phenotypic and malignant characteristics of Tet1-deficient spleen and lymph node cells were transplantable, causing rapid disease 12 weeks post-transplant into congenic mice (Supplementary Fig. 2e–g). These data suggested that Tet1 functions as a tumor suppressor of hematopoietic malignancy, specifically in the B cell lineage.

Tet1-deficient tumors harbor mutations of B-lymphoma

To understand the etiology of Tet1-deficient hematopoietic malignancy in mice we performed whole-exome sequencing on genomic DNA isolated from thirteen Tet1-deficient tumors. Tumor samples were derived from the spleen, lymph node and peripheral blood of $Tet1^{+/-}$ and $Tet1^{-/-}$ mice, including a three tumors that stained negative for IgM by flow cytometry (Supplementary Table 2). Matched germline tail DNA was used as the reference genome for each mouse. In the coding genome, we identified >2000 somatic mutations, involving 1785 unique genes, of which 157 genes were recurrently mutated in 2 or more tumors (Supplementary Data Set 1–3). While the proportion of indels was equivalent across

all tumors, the frequency of non-synonymous single nucleotide variants (nsSNVs) differed by more than 10-fold for several samples (Fig. 2a, Supplementary Fig. 3). Increased numbers of nsSNV in the highly mutated tumors corresponded to an increased frequency of A/T base substitutions (Fig. 2b), which has been previously reported in patients with diffuse large B cell lymphoma (DLBCL) and associated with increased frequency of mutation in DNA repair genes[29]. In the combined tumors, nsSNVs were predominantly missense mutations, making up 83% of all variations observed (Fig. 2c). Independent of mutation load we observed more transitions than transversions (Fig. 2d,e) and a prevalence of C>T transitions in the context of GCG trinucleotides, suggesting a bias toward mutation of CpG cytosines (Fig. 2f).

Several whole genome and exome sequencing studies of B-NHL patients incorporating both DLBCL and follicular lymphoma (FL) have revealed a core set of recurrently mutated genes involved in RhoA signaling (*GNA13*), histone modification (*KMT2D/MLL2*), and B cell function (*MYD88, CD83*)[24–26]. Several of these genes are also known targets of aberrant somatic hypermutation (*PIM1, CD79B, FAS*)[30]. We found that many of the genes mutated in ≥3 Tet1-deficient tumors (both IgM^- or IgM^+) corresponded with those recurrently identified in B-NHL patients (Fig. 2g, Supplementary Fig. 4a–c and Supplementary Table 3). These included *Kmt2d*, which encodes for the H3K4-methyltransferase protein, *Mll2*, mutated in 30–50% of DLCBL and up to 89% of FL patients[24,25]. In addition, we identified mutations in several linker histone genes (*Hist1h1c, Histh1d and Hist1h1e*) along with other histone variants and histone modifying enzymes (Fig. 2h,i). *HIST1H1 B–E* mutations were recently reported in up to 27% of FL patients[26,27]. Together, the immunophenotype and mutation spectra of Tet1-deficient tumors suggested that loss of *Tet1* predisposed mice to the formation of B cell lymphoma.

TET1 is hypermethylated and silenced in human B-NHL

Despite the overlapping spectrum of mutations that correspond to human B-NHL, *TET1* somatic mutations have not been reported in this disease. We investigated alternative mechanisms of *TET1* inactivation in human B-NHL and found that the *TET1* gene is epigenetically silenced in FL (Fig. 3a) and Multiple Myeloma (MM) (Supplementary Fig. 4d) and transcriptionally downregulated in both DLBCL and FL patients (Fig. 3b). MassARRAY analysis of *TET1* CpG methylation was performed in an independent cohort of 26 FL patients compared to normal germinal center B (GCB) cells (Fig. 3c, Supplementary Fig. 4e). Corresponding RNA-seq analysis of *TET1* expression in the same patients showed a direct correlation between hypermethylation and decreased expression of *TET1* (Fig. 3d). These studies are consistent with our animal findings that demonstrate that $Tet1^{-/-}$ animals develop exclusively B cell lymphoma and with the fact that the $Tet1^{-/-}$ lymphomas carry mutations in genes consistent with those found in FL and DLBCL patients. Interestingly, we have noticed no hypermethylation of *TET2* in human B-NHL samples. This finding is in agreement to the spectra of hematopoietic malignancies reported in *Tet2*-deficient mice, which developed primarily myeloid disease (Supplementary Fig. 4f–h). Compared to other hematopoietic malignancies, the significant hypermethylation of the *TET1* promoter in FL patients, which correlated directly with decreased *TET1* expression,

would suggest that consistent with the phenotype observed in Tet1-deficient mice, *TET1* plays a role as a tumor suppressor specifically in B cell lymphoma.

Loss of *Tet1* co-operates with *Bcl2* overexpression

To determine whether loss of *Tet1* alters stem and progenitor differentiation in a cell intrinsic manner we performed primary transplants of total bone marrow and secondary transplants of purified long-term HSCs (LT-HSCs) (Fig. 4). In both the primary and secondary transplants, $CD45.2^{+}$ $Tet1^{-/-}$ donor cells exhibited a greater contribution to the maintenance of peripheral blood reconstitution (Fig. 4a) and a significant increase in B cell frequency 20 weeks post-transplant compared to $CD45.2^{+}$ $Tet1^{+/+}$ donor cells (Fig. 4b). Together these data suggest that the primary defect in Tet1-deficient mice is enhanced HSC activity with a bias toward B lymphopoiesis.

Given the prevalence of *BCL2* translocations in FL patients (~90%)[31] that drive overexpression of this pro-survival gene, we tested whether retroviral introduction of *Bcl2* overexpression in $Tet1^{-/-}$ LSKs could cooperate to promote B lymphomagenesis. Up to 10 weeks post-transplant we found that neither *Tet1* loss nor *Bcl2* overexpression alone had an effect on the total leukocyte counts. However, the combination of loss of *Tet1* and overexpression of *Bcl2* drove a rapid and sustained elevation in circulating total leukocytes, specifically B lymphocytes, with little effect on the number of circulating myeloid cells (Fig. 4c,d).

To gain further insight into the development of hematopoietic malignancies in Tet1-deficient mice we quantified *Tet1* mRNA abundance at various stages of hematopoiesis. We found that *Tet1* was most highly expressed in LT-HSCs, multipotent progenitors (MPPs) and common lymphoid and myeloid progenitors (CLPs and CMPs, respectively), decreased during B lineage commitment and was undetectable in immature and mature myeloid cells (Fig. 5a). Flow cytometric analysis of Tet1-deficient bone marrow and spleen cells from young mice (1–3 months) showed no perturbation in hematopoietic development (Supplementary Fig. 5a,c). However in sick mice, the total $B220^{+}$ cell frequency was decreased in both the spleen and bone marrow, with progenitor and precursor B cells ($B220^{+}IgM^{-}IgD^{-}$) making up the majority of the B cell compartment (Supplementary Fig. 5b,d–g). In young animals, loss of *Tet1* alone was sufficient to cause a decrease in 5hmC content in total bone marrow and B cells (Fig. 5b,c). Flow cytometry of the bone marrow stem cell compartment from sick Tet1-deficient mice showed no significant difference in the frequency of lineage negative c-Kit^{+} progenitors or LSKs ($Lin^{-}Sca\text{-}1^{+}cKit^{+}$) compared to wild-type animals (Fig. 5d). However, within the LSK compartment, staining with CD150 and CD48 showed that LT-HSCs ($CD150^{+}CD48^{-}$) were significantly decreased, whereas MPP2 cells ($CD150^{-}CD48^{+}$) and LMPP cells ($CD34^{+}Flt3^{+}$) were increased in frequency (Fig. 5e–g). These data suggested that loss of *Tet1* caused a bias towards lymphoid differentiation in hematopoietic stem and progenitor cells despite a later block in B cell maturation.

Tet1-loss promotes B-lineage differentiation

In the Tet1-deficient LSK compartment, the decreased frequency of LT-HSCs and increased frequency of lymphoid-primed MPPs prompted us to identify molecular signatures that define these subpopulations of LSKs in the absence of Tet1. We performed microarray analysis of RNA isolated from total LSKs and RNA-seq analysis on RNA from purified *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ LT-HSCs and MPPs (Fig. 5h–j and Supplementary Fig. 6). Genes with greater than 2-fold change in expression were equally up- and down-regulated in *Tet1*$^{-/-}$ compared to *Tet1*$^{+/+}$ LSKs (Fig. 5h). By gene set enrichment analysis (GSEA), we observed that in total LSKs and MPPs there was a significant up-regulation in mature hematopoietic cell genes and specifically B lymphoid and lymphoma expression profiles[32,33]. This included the transcription factors *Ebf1*, *Pax5*, *Irf4*, and their downstream targets, *VpreB*, *Cd79a*, *Igh*, and *Blnk*, suggesting an increased differentiation bias toward the B lymphoid lineage (Fig. 5h,i and Supplementary Table 4). Down-regulated and negatively enriched gene sets by GSEA included pathways involved in chromosome maintenance, transcription, and DNA repair (Fig. 5h,j and Supplementary Table 5). Histone cluster 1 genes were found to be significantly down-regulated, including the linker histone genes (*Hist1h1a,b,d,e*) and other histone 2, 3 and 4 variants, several of which (including *Hist1h1d,e*, *Hist1h2bc,bb* and *Hist1h3a,b,e*) were also found mutated in Tet1-deficient lymphoma cells. The increased expression of histone variants in ST-HSCs, LMPPs and pre-B cells (Supplementary Fig. 6d) and their frequent mutation in B cell lymphoma implicates these proteins as potentially important regulators of B cell development and function.

Aberrant DNA hydroxymethylation of Tet1-deficient cells

The role of TET proteins in DNA-demethylation suggests that in the absence of these enzymes, the genome may be vulnerable to the acquisition of increased DNA methylation. We performed a genome-wide analysis of 5hmC and 5mC at CpG sites in *Tet1*$^{+/+}$ and *Tet*1$^{-/-}$ LSKs using HELP-GT[34] and reduced representative bisulfite sequencing (RRBS)[35] assays, respectively. The frequency of CpG sites with differential 5hmC or 5mC was calculated for *Tet*1$^{-/-}$ compared to *Tet1*$^{+/+}$ LSKs (Supplementary Fig. 7). A global loss of 5hmC and gain in 5mC was observed across all chromosomes (Fig. 6a–d), with greater losses in 5hmC occurring in gene bodies (exons and introns) and intergenic regions (Supplementary Fig. 7a) that overlapped significantly with histone modifications (H3K4me1, H3K4me2, and H3K27Ac) characteristic of both poised and active enhancers in mouse MPPs[36] (Supplementary Fig. 7b). Interestingly, only 10% of promoters exhibited losses in 5hmC whereas the frequency of 5mC gain at promoters contributed to 27% of all differentially methylated CpGs (Fig. 6e, Supplementary Fig. 7c–e). Using Ingenuity Pathway Analysis (IPA) of loci that lose 5hmC in Tet1-deficient LSKs we found that amongst the top 20 networks and disease categories were multiple regulators of DNA replication, recombination and repair (Supplementary Tables 6, 7). Several genes involved in RhoA signaling (*RhoA*, *Gna12*, *Arghap12*) and specifically base excision repair (*Apex1*, *Lig1*, *Exo1*) lost 5hmC, whereas genes that displayed both 5hmC loss and 5mC gain were enriched in G-protein coupled receptor (*Gna14*), TGFβ (*Smad 2–4*) and WNT (*Ctnnb1*) signaling pathways known to play important roles in stem cell homeostasis and transformation[37,38] (Supplementary Fig. 7f, Supplementary Table 8). Many down-regulated

genes in Tet1-deficient LSKs or MPPs also exhibited loss of 5hmC in their gene bodies or promoters, including B cell function (*Cd74*), apoptotic (*Casp3*) and chromosome maintenance (*Mcm4*, *Mcm10*) genes, in agreement with our previous data demonstrating dysregulation of these gene pathways by gene set enrichment analysis in stem and progenitor cells. These data show for the first time in hematopoietic stem and progenitor cells that *Tet1* expression is required for the maintenance of DNA hydroxymethylation and to protect the genome from aberrant DNA hypermethylation.

Increased self-renewal and DNA damage in Tet1-deficient cells

DLBCL and FL are characterized as diseases of mature lymphocytes, originating from germinal center or post-germinal center lymphocytes[31]. It has been shown however that HSCs from patients with chronic lymphocytic leukemia (CLL) already display lymphoid lineage gene priming that in xenograft transplants develop into mature CLL-like B cells[39]. To assess the self-renewal potential of Tet1-deficient stem and progenitor cells *in vitro* we performed pro-B, pre-B and myeloid cell colony formation assays. Loss of *Tet1* promoted a significant increase in the number of self-renewing pro-B cell colonies (Fig. 7a) with only a moderate increase in colony formation for pre-B cells (Fig. 7b) and no effect on myeloid CFU replating capacity (data not shown). In both pro-B and pre-B CFU cells from Tet1-deficient mice, V(D)J recombination appeared normal (data not shown). However, the expression of several Histone Cluster 1 (with the exception of *Hist1h1c*) and DNA repair genes showed on average 2-fold decrease in mRNA expression (Fig. 7c,d). Consistent with deficient DNA repair pathway gene expression, ProB cells from *Tet1*$^{-/-}$ mice were found to display increased DNA damage with a greater percentage of γ-H2AX positive cells and significantly more foci per cell (Fig. 7e–g). These data show that Tet1-deficiency has the potential to generate pre-malignant B cells at the pro-B cell stage of differentiation.

Discussion

NHL constitutes a diverse array of malignancies arising from B or T lymphocytes of various developmental stages, immunophenotype, size, clinical presentation and disease progression[31,40] DLBCL, FL and Mantle cell lymphoma (MCL) are the most common B cell NHLs and together comprise over 60% of all NHL. Epigenetic factors have been shown to contribute to the heterogeneity of NHL, with several recent studies implicating aberrant DNA methylation in disease progression. Methylation heterogeneity in FL and DLBCL increases with disease severity and aberrant DNA methylation can predict patient survival[41].

Interestingly, B and T cell lymphoma with different histological subtypes are associated with distinct global DNA methylation states[42,43]. Key genes involved in cell cycle regulation, survival signaling and DNA repair have been shown to be hypermethylated in FL (germinal center tumors) compared to MCL (pre-germinal center) and chronic lymphocytic leukemia (CLL)/small lymphocytic lymphoma (pre- or post-germinal center tumors)[44] and may reflect either the stage of normal lymphoid maturation from which these lymphomas arise or etiology of the disease.

Methylation profiling of DNA from AML patients led to the first observation that *TET2* mutant AMLs are characterized by a hypermethylation phenotype[18]. *TET2* is also mutated

or deleted in 6–12% of DLBCL[45,46]. Genome-wide profiling studies have found that DLBCL patients also display a *TET2* mutant hypermethylation signature compared to *TET2* wild-type DLBCL[46]. Together these findings are consistent with a role for TET proteins in generating 5hmC as an intermediate step in DNA demethylation at sites that would otherwise remain marked by 5mC. Gene silencing by DNA methylation in somatic cells is relatively stable and compounded by repressive histone modifications and the recruitment of proteins involved in histone deacetylation. Studies have shown that DNA methylation inhibits H3K4 methylation[47], a marker of gene activation, and is associated with H3K27 methylation[35] a marker of repressive chromatin. Reversal of this gene silencing therefore provides an attractive strategy for pharmacologic intervention. Epigenetic treatments that promote gene activation at silenced chromatin include histone deacetylase inhibitors (HDACis), such as Vorinostat[48]. Many phase I/II clinical trials are currently being conducted to expand the clinical use of HDACis in MM, DLBCL, MDS and AML patients. DNA demethylating agents such as 5-aza-2′-deoxycytidine are approved by the FDA for use in the treatment of MDS however, very little is known of the molecular effect of using DNA hypomethylating agents alone or in combination with HDACis in NHL patients. Additional DNA methylation profiling in various lymphoma subtypes may help facilitate both diagnosis and clinical treatment.

Leukemia is appreciated as a disease of stem cell origin, however lymphoma is more commonly known as a disease of mature lymphocytes and in the case of DLBCL and FL, a manifestation of germinal center or post-germinal center reactions[31]. However, HSCs from patients with CLL were shown to already display lymphoid lineage gene priming that develop into mature CLL-like B cells in xenograft transplants[39]. Clonal evolution of the same biallelic mutations in *TET2* within $CD34^+$ cells of a patient has been shown to contribute to both B cell lymphoma and AML, suggested that the patient was predisposed to malignancy from cells that retained multi-lineage differentiation potential. In a donor-recipient pair that developed FL 7 years after allogeneic bone marrow transplant, the donor-lymphocyte infusions were found to harbor the same *BCL2-IGH* and *IGH V-DJ* rearrangements[49]. While the *BCL2-IGH* rearrangement was shown to be restricted to $CD19^+$ cells, ultra deep-sequencing of $CD34^+CD10^-CD19^-$ cells within the donor-lymphocyte infusions, comprising both MPP and HSC cells, resulted in detection of 3/14 shared mutations between donor and recipient lymphomas, including mutations in *EP300* and *KLHL6*[49], known to be recurrently mutated in FL patients[24]. How epigenetic dysregulation within stem and progenitor cells may create a premalignant state or increased susceptibility to the acquisition of transforming alterations is not clear. Further deep-sequencing studies to determine the mutational status and epigenetic state of stem and progenitor cells from patients with DLBCL and FL may provide additional insight into the molecular ontogeny of B cell lymphoma.

Despite the evidence that loss of either the TET proteins or 5hmC correlates with cancer progression, a role for TET1 as a tumor suppressor in hematopoietic malignancy had not been previously established. Our findings suggest that loss of Tet1 function predisposes hematopoietic stem cells to malignancy and more specifically, B cell lymphoma. Deletion of either *Tet1* (our study) or *Tet2*[50] in mice causes phenotypically distinct tumors affecting

predominantly B or myeloid lineages, respectively, suggesting non-redundant or lineage-specific roles for these enzymes during hematopoiesis or transformation. *TET1* mutations in patients with hematopoietic malignancy are reported at low frequency compared to *TET2*, however loss of TET1 function due to altered transcription or post-translational regulation in hematopoietic stem and progenitor cells may be an overlooked mechanism by which these cells are able to transform into lymphoma. These data suggest that hypomethylating agents could be an efficient therapy for B-NHL patients characterized by epigenetic silencing of *TET1*.

Methods

Generation of Tet1-deficient mouse cohort

The *Tet1* knockout mouse strain used in this study has been previously reported[28]. $Tet1^{+/-}$ animals were intercrossed to generate cohorts of $Tet1^{+/+}$, $Tet1^{+/-}$ and $Tet1^{-/-}$ mice. All mice in the study were maintained on a mixed 129 and C57BL/6 background and monitored weekly for signs of poor health. Moribund animals were sacrificed and subjected to necropsy. Spleen, liver, lymph nodes, kidneys, lungs, bone marrow and blood were harvested for further analysis. Animal care was in accordance with institutional guidelines and approved by the Institutional Animal Care and Use Committee of the NYU School of Medicine and the Committee on Animal Care, Department of Comparative Medicine, Massachusetts Institute of Technology. Log Rank test was used for statistical analyses. Plotting of Kaplan-Meier survival curves and statistical analyses (*t*-test) were performed using Prism Graphpad software.

Histology and Immunohistochemistry

Peripheral blood smears were briefly fixed in methanol and stained with Wright-Giemsa solution (Fisher). Slides were rinsed with water, dried, mounted with Cytoseal 60 and cover slipped. Tissues were dissected from mice for fixation overnight in 10% formalin (Fisher). Fixed tissues were dehydrated and embedded in paraffin for sectioning. 5 μm paraffin sections were prepared and stained with hematoxylin and eosin (H&E) (Leica Autostainer XL). For immunohistochemistry, Bcl6 (N3, Santa Cruz Bio, SC-858), IRF4 (Santa Cruz Bio, SC-6059), CD138 (BD Pharmingen, 553712) and Ki67 (Clone SP6, Epitomics) antibodies were used and counterstained with hematoxylin. Light microscopy was performed using a Zeiss Axio Observer microscope.

Array-based methylation analysis using HELP

After informed consent from all participants, the genomic DNA from the primary leukemia and lymphoma cases were obtained for the HELP assay with the approval of the Institutional Review Boards of Weill Cornell Medical College. DNA methylation of patient samples was performed and analyzed using the HELP (HpaII tiny fragment enrichment by ligation mediated PCR) as previously published[52,53]. The HELP array data was available from the GEO database accession numbers GSE31671, GSE23967, GSE34937, GSE44860, and GSE18700. Basically, raw data (.pair) files were generated using NimbleScan software. Signal intensities at each HpaII amplifiable fragment were calculated as a robust (25% trimmed) mean of their component probe-level signal intensities. Any fragments found

within the level of background MspI signal intensity, measured as 2.5 mean-absolute-deviation (MAD) above the median of random probe signals, were considered as [3]failed[2] probes and removed. A median normalization was performed on each array by subtracting the median log-ratio (HpaII/MspI) of that array (resulting in median log-ratio of 0 for each array).

Gene expression microarray data

Gene expression microarray data of patient samples was obtained from the GEO database accession number GSE12453[54]. FL patient samples were used after informed consent from all participants with the approval of the Institutional Review Boards of Weill Cornell Medical College. The microarray raw data were normalized using the MAS5 method with Expression Console software (Affymetrix).

Single locus quantitative DNA methylation assays

EpiTYPER assays (Sequenom) were performed on bisulfite-converted DNA from FL patients as previously described[55] after informed consent from all participants with the approval of the Institutional Review Boards of Weill Cornell Medical College. The primers were designed using Sequenom EpiDesigner beta software (http://www.epidesigner.com/). The primer sequences are available in (Supplementary Table 9).

Fluorescence activated cell sorting and flow cytometry

Single cell suspensions were prepared from bone marrow (femur and tibia), spleen and lymph nodes. Red blood cells were lysed with ammonium-chloride-potassium (ACK) buffer and remaining cells were resuspended in PBS with 3% FBS. Nonspecific antibody binding was blocked by incubation with 20 μg/ml Rat IgG (Sigma-Aldrich) for 15 min. Cells were incubated with antibodies for 30 min on ice. All antibodies were purchased from BD-PharMingen, eBioscience or BioLegend and consisted of fluorochrome or biotin-conjugated anti-mouse antibodies: CD11b (M1/70), Gr-1 (RB6-8C5), CD71 (RI7217), Ter-119, CD3 (145-2C11), B220 (RA3-6B2), IgD (IA6-2), IgM (II/41), cKit (CD117, 2B8), Sca-1 (D7), CD150 (9D1), CD48 (HM481), CD45.1 (A20), CD45.2 (104). Bone marrow mature lineage positive markers were defined as CD11b, Gr-1, Ter-119, CD4, CD8, CD3, and B220. For 5hmC detection, 10^6 cells were fixed at 25 °C with 3% paraformaldehyde, permeabilized with 0.2% Triton X-100, denatured with 2N HCl followed by neutralization with 100 mM Tris-HCl, pH 8.0. Cells were then incubated with anti-5hmC (1:400) (Active Motif) and isotype control antibodies for 30 min, washed with PBS and incubated with secondary antibodies conjugated with Alexa Fluor 488 (1:500) (Invitrogen) for 30 min followed by wash and resuspension in PBS for visualization. Stained cells were quantified using a BD Fortessa analyzer or isolated with a MoFlo cell sorter (Beckman Coulter) or BD ARIA II. FlowJo software (TreeStar) was used to generate flow cytometry plots.

Whole-exome sequencing and identification of tumor-specific variants

Genomic DNA (gDNA) was isolated from the spleen, peripheral blood and lymph node tumor cells of Tet1-deficient mice and paired with gDNA isolated from the tail of each respective mouse as a reference genome. The mouse exome was enriched using SeqCap EZ

design mouse exome capture kit (Roche NimbleGen, Inc.) The coding sequence selected for the mouse exome probe pool design includes 203,225 exonic regions, including microRNAs, and collectively comprises over 54.3 Mb of target sequence (C57BL/6J, NCBI37/mm9). Sequencing libraries were generated from captured exomes using standard Illumina protocols and sequenced on the Illumina HiSeq 2500 using 100 bp paired-end reads. Approximately 100 million reads were sequenced per tumor at 2X the coverage of matched tail DNA to allow for more sensitive somatic variant calling. All tumor samples achieved average coverage above 85X with over 90% of bases above 20X coverage. Reads were aligned to the mouse genome (UCSC build mm10) using the Burrows-Wheeler Aligner (BWA). Additional refinement included duplicate-read removal with Picard as well as realignment and base quality score recalibration using the Genome Analysis Toolkit (GATK). Somatic variant callers MuTect and GATK Somatic Indel Detector were used to generate single nucleotide substitution and indel calls, respectively, in matched tumor-normal pairs. Mutations were annotated using Annovar based on all annotated transcripts in RefSeq Gene with likely germline variants marked using data from dbSNP (build 137). Recurrent mutations were visualized using Circos online software[56]. Trinucleotide mutational signatures were calculated based on equal trinucleotide frequency.

Bone marrow-competitive transplantation

Freshly dissected femurs and tibias were isolated from donor *Tet1*$^{+/+}$, *Tet1*$^{-/-}$ (CD45.2^{+}) and support *Tet1*$^{+/+}$ (CD45.1^{+}) mice. Bones were flushed with PBS using a 22.5 gauge needle, and the isolate was centrifuged (5 min, 0.5*g*, 4 °C) and treated with red cell lysis buffer as described above. Total nucleated bone marrow cells were re-suspended in PBS, passed through a 40 μM cell strainer and counted. For primary transplants, donor cells (0.5×10^6 per genotype, per mouse) were mixed 50:50 with support bone marrow cells and transplanted via retro-orbital injection into lethally irradiated (2 × 550 Rad) CD45.1^{+} recipient mice. Chimerism was monitored by flow cytometry (anti-CD45.1 and anti-CD45.2, BD Bioscience) of peripheral blood at 4-week internals post-transplant for 20 weeks at which time mice were sacrificed to assess chimerism in total cell isolates of other hematopoietic compartments (bone marrow and spleen). For secondary transplants, CD45.2^{+} LT-HSCs were purified by flow cytometry from the bone marrow of primary transplanted mice 20 weeks post transplant and re-injected into lethally irradiated CD45.1^{+} recipient mice (500 LT-HSCs per mouse) with support cells (0.5×10^6 total bone marrow cells per mouse) and monitored for chimerism as described above.

Retroviral transduction

Hematopoietic stem and progenitor cells were isolated by sorting Lin^{-} c-Kit^{+}Sca1^{+} cells, cultured in the presence of 50 ng/ml SCF, 50 ng/ml Flt3 ligand, 10 ng/ml IL-3, and 10 ng/ml IL-6 and infected with concentrated retroviral supernatants after 24 and 48 h. Transduction efficiency was determined by reporter fluorescence at 96 h, and total or sorted populations were transferred via retro-orbital injection into irradiated (2 × 550 Rad) congenic recipients along with 2×10^5 unfractionated bone marrow cells (BMMCs) for hemogenic support.

In vitro colony formation assays

Total bone marrow from *Tet1*$^{+/+}$, *Tet1*$^{+/-}$ and *Tet1*$^{-/-}$ mice (24 weeks of age) was plated in methylcellulose medium (Stem Cell Technologies); M3630 for pre-B, M3434 for myeloid/erythroid and M3630 supplemented with Flt3L (50ng/ml) for pro-B colony formation. Cells were seeded at a density of 20,000 cells/ml for pro-B or pre-B and 2,000 cells/ml for myeloid progenitor colony formation assay. Colony forming units were counted and re-plated (2000 cells/replicate) every 7–10 days.

Quantitative RNA expression assays

Total RNA was extracted from flow cell sorted LSK cells using the RNeasy Plus Mini Kit (Qiagen) from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice (24 weeks of age, $n = 3$ per genotype). RNA quantification and quality was determined using an Agilent 2100 Bioanalyzer. The Ovation RNA Amplification System V2 (NuGEN) kits were used for amplification. Amplified RNA was labeled and hybridized to the Mouse 430.2 microarrays (Affymetrix). Mouse hematopoietic microarray samples used in Extended Data Fig. 8f have been described previously[51] and are available at the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under the reference GSE14833. The Affymetrix gene expression profiling data was normalized using the previously published Robust Multi-array Average (RMA) algorithm using the GeneSpring GX software (Agilent). RNA-seq libraries were prepared from total RNA of flow cell sorted CD45.2^{+}LT-HSCs and MPPs from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice (20 weeks post-primary bone marrow transplant, combined from 3 mice per genotype per library) by PolyA selection using oligo-dT beads (Life Technologies) according to the manufacturer's instructions. The resulting RNA samples were then used as input for library construction using standard Illumina protocols and sequenced on the Illumina HiSeq 2000 using 50 bp paired-end reads. All RNA-Seq data was aligned to mm9 using TopHat v1.4 with default parameters and Cuffdiff v1.3 for all differential expression (DE) analyses. In all DE tests, a gene was considered significant if the q-value was less than 0.05 (Cuffdiff default). All gene expression data were gene/row normalized (*z*-score transformation) using Multiple Experimental Viewer (MeV) software (v4.7.4) for gene expression intensity visualization. For mRNA quantification, total RNA was used for cDNA synthesis using Superscript III first strand synthesis kit (Invitrogen). Real-time PCR reactions were carried out using SYBR Green Master Mix (Roche) and run with a Lightcycler 480 II (Roche). The following primer sequences were used for cDNA quantification: *Hprt* (mouse) 5′-CAGTACAGCCCCAAAATGGT-3′ and 5′-CAAGGGCATATCCAACAACA-3′, *Tet1* 5′-AGCTGGATTGAAGGAACAGG-3′ and 5′-GTCTCCATGAGCTCCCTGAC-3′. The NanoString nCounter gene expression system (NanoString Technologies Inc.) was utilized to directly measure mRNA abundance in pro-B colony forming cells. A custom nCounter Elements probeset (IDT) composed of probes to Histone Cluster 1, DNA repair and control genes (Supplementary Table 10), together with genes unrelated to this study was hybridized with 100 ng samples of total RNA and processed according to manufacturer's instructions.

Gene set enrichment analysis

Gene set enrichment analysis was performed using gene set as permutation type, 1,000 permutations and log2 ratio of classes or with gene_set and Signal2Noise as metrics for ranking genes. Gene sets used in this study were identified from the Molecular Signatures Database (MSigDB Curated v4.0). Previously published LSK ChIP-seq data[57], available at the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/) under the reference (GSE22075), was downloaded and aligned to mm9 using bowtie. Following alignment regions that showed significant enrichment for H3K4me3 or H3K27me3 were determined using MACSv1.4 (for H3K4me3 a $P < 10^{-13}$, for H3K27me3 $P < 10^{-5}$ and fold enrichment > 7). Enriched regions were extended by 500 bp in both directions and overlapping intervals were merged. Bivalent regions were identified as regions of H3K27me3 and H3K4me3 enrichment according to the cutoffs above. Genes (RefSeq mm9) were identified as bivalent if a bivalent region had any overlap in the interval spanning 5 kb upstream of the TSS to the TTS and intersects with Refseq genes containing CpG islands in promoters (2 kb upstream of the TSS) were generated using BED Tools software.

HELP-GT assay

Genomic DNA (gDNA) was isolated from sort-purified LSK ($Lin^{-}c\text{-}Kit^{+}Sca1^{+}$) cells from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice (24 weeks of age, $n = 2$ per genotype). The HELP-GT assay was used to map cytosine hydroxymethylation as previously described[34]. Briefly, gDNA with or without β-glucosyltransferase (β-GT) pre-treatment was digested by HpaII or MspI. HpaII only cuts at CCGG sequences where the central CCGG dinucleotide is unmethylated. The first Illumina adapter (AE) is ligated to the compatible cohesive end created, juxtaposing an EcoP15I site beside the HpaII/MspI digestion site and allowing EcoP15I to digest within the flanking DNA sequence. An A overhang is created, allowing the ligation of the second Illumina adapter (AS). This creates AE-insert-AS products and AS-insert-AS molecules. By performing a T7 polymerase-mediated *in vitro* transcription from a promoter sequence located on the AE adapter, the AE-insert – AS product is selectively enriched, following which, limited PCR amplification is performed to generate a single sized product for Illumina sequencing. The final library is sequenced by multiplexing in-house adapter primers (in multiples of 4) using an Illumina HiSeq2000 (50 bp single end reads). An average coverage of 10–15× per base was achieved and comparison of β-GT+MspI and MspI only digested sample was used to determine hydroxymethylated sites.

Bisulfite sequencing

Genomic DNA (gDNA) was isolated from sort-purified LSK ($Lin^{-}c\text{-}Kit^{+}Sca1^{+}$) cells from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice (24 weeks of age, $n = 3$ per genotype). DNA methylation sequencing was performed and analyzed using an enhanced reduced representation bisulphite sequencing (ERRBS) methodology as previously described[58]. Bisulfite treatment was performed using the EZ DNA Methylation Kit (Zymo Research) and upon addition of the CT conversion reagent, the incubation was conducted in a thermocycler (Eppendorf MasterCycler). Purified products were subjected to PCR amplification using the FastStart High Fidelity PCR System (Roche). PCR products were isolated using AMPure XP beads (Life Sciences). Library amplification and quality was determined using the Agilent 2100

Bioanalyzer. The amplified libraries were sequenced on an Illumina HiSeq2500 with 50 bp single-end reads following standard Illumina sequencing protocols. Primary and downstream data analysis including calculation of differential methylation was performed as previously described[58].

Immunofluorescence staining

Cytospins of sort-purified pro-B cells from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice (24 weeks old) on poly-lysine coated slides (50,000 per slide) were fixed with 4% paraformaldehyde for 10 min at 25 °C, permeabilized with 0.15% triton X-100 in PBS with for 2 min and blocked with 3% FBS/PBS for 30 min then incubated with anti-phospho-H2A.X (Ser139) (20E3) rabbit mAb (Cell Signaling Technology) overnight at 4 °C. Slides were washed three times in PBS and incubated for 1 hour at room temperature in 3% FBS/PBS with Alexa 488-conjugated goat anti-rabbit antibody (Life Technologies). Slides were washed three times in PBS and mounted using VectaShield (Vector Laboratories) containing 1 μg/ml DAPI.

V(D)J recombination

Igh, *Igk*, and *Igl* rearrangements were amplified by PCR using *Ig* gene-specific primers (Supplementary Table 11). DNA was isolated from cells by proteinase K digestion, phenol extraction and ethanol precipitation. PCR cycle numbers were adjusted to be in the linear range, based on the analysis of serially diluted DNA and PCR products were separated on agarose gels.

Statistics

All data were expressed as Mean ± SEM as indicated and *P*-values calculated using a Student's *t*-test with Graphpad Prism software unless otherwise described in the methods or figure legend. No specific randomization or blinding protocol was used for these analyses.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We would like to thank the members of the Aifantis laboratory for helpful discussions, M. Xu and F. Yang at Indiana University for valuable help and advice and L. Pasqualucci at Columbia University for providing reagents and expertise. We thank K. Ganz., R. Flannery., D. Fu. and T. Adil. from the Jaenisch lab for help with animal husbandry and tissue collection. Also, we thank A. Heguy and I. Dolgalev from the NYU Genome Technology Center, for expert assistance with microarray experiments, RNA- and exome-sequencing, the NYU Flow Cytometry facility for expert cell sorting and the NYU Histology Core. We also thank E. Oricchio from the Swiss Institute for Experimental Cancer Research, Ecole Polytechnique Federale de Lausanne for assisting with analysis of FL patient data. M.M.D. is a Damon Runyon Postdoctoral Fellow. Work in R.J.'s lab is funded by NIH grants 5RO1HD045022 and 5R37CA084198 and the Simons Foundation. The Aifantis Lab is supported by the National Institutes of Health (1R01CA169784, 1R01CA133379, 1R01CA105129, 1R01CA149655, 5R01CA173636), the William Lawrence and Blanche Hughes Foundation, The Leukemia & Lymphoma Society (TRP#6340-11, LLS#6373-13), The Chemotherapy Foundation, The V Foundation for Cancer Research, the Alex's Lemonade Stand Foundation for Childhood Cancer and the St. Baldrick's Cancer Research Foundation. L.C. is supported by an NHMRC overseas biomedical research fellowship. I.A. is a Howard Hughes Medical Institute Early Career Scientist.

References

1. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012; 150:12–27.10.1016/j.cell.2012.06.013 [PubMed: 22770212]
2. Tahiliani M, et al. Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. Science. 2009; 324:930–935.10.1126/science.1170116 [PubMed: 19372391]
3. Ito S, et al. Tet Proteins Can Convert 5-Methylcytosine to 5-Formylcytosine and 5-Carboxylcytosine. Science. 201110.1126/science.1210597
4. Guo JU, Su Y, Zhong C, Ming GL, Song H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. Cell. 2011; 145:423–434.10.1016/j.cell.2011.03.022 [PubMed: 21496894]
5. Yang H, et al. Tumor development is associated with decrease of TET gene expression and 5-methylcytosine hydroxylation. Oncogene. 2013; 32:663–669.10.1038/onc.2012.67 [PubMed: 22391558]
6. Kudo Y, et al. Loss of 5-hydroxymethylcytosine is accompanied with malignant cellular transformation. Cancer Sci. 2012; 103:670–676.10.1111/j.1349-7006.2012.02213.x [PubMed: 22320381]
7. Xu Y, et al. Genome-wide regulation of 5hmC, 5mC, and gene expression by Tet1 hydroxylase in mouse embryonic stem cells. Mol Cell. 2011; 42:451–464.10.1016/j.molcel.2011.04.005 [PubMed: 21514197]
8. Stroud H, Feng S, Morey Kinney S, Pradhan S, Jacobsen SE. 5-Hydroxymethylcytosine is associated with enhancers and gene bodies in human embryonic stem cells. Genome Biol. 2011; 12:R54.10.1186/gb-2011-12-6-r54 [PubMed: 21689397]
9. Valinluck V, Sowers LC. Endogenous cytosine damage products alter the site selectivity of human DNA maintenance methyltransferase DNMT1. Cancer Res. 2007; 67:946–950.10.1158/0008-5472.CAN-06-3123 [PubMed: 17283125]
10. He YF, et al. Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. Science. 2011; 333:1303–1307.10.1126/science.1210944 [PubMed: 21817016]
11. Lorsbach RB, et al. TET1, a member of a novel protein family, is fused to MLL in acute myeloid leukemia containing the t(10;11)(q22;q23). Leukemia. 2003; 17:637–641.10.1038/sj.leu.2402834 [PubMed: 12646957]
12. Ittel A, et al. First description of the t(10;11)(q22;q23)/MLL-TET1 translocation in a T-cell lymphoblastic lymphoma, with subsequent lineage switch to acute myelomonocytic myeloid leukemia. Haematologica. 2013; 98:e166–168.10.3324/haematol.2013.096750 [PubMed: 24323992]
13. Burmeister T, et al. The MLL recombinome of adult CD10-negative B-cell precursor acute lymphoblastic leukemia: results from the GMALL study group. Blood. 2009; 113:4011–4015.10.1182/blood-2008-10-183483 [PubMed: 19144982]
14. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013; 368:2059–2074.10.1056/NEJMoa1301689 [PubMed: 23634996]
15. De Keersmaecker K, et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. Nat Genet. 2013; 45:186–190.10.1038/ng.2508 [PubMed: 23263491]
16. Abdel-Wahab O, et al. Genetic characterization of TET1, TET2, and TET3 alterations in myeloid malignancies. Blood. 2009; 114:144–147.10.1182/blood-2009-03-210039 [PubMed: 19420352]
17. Delhommeau F, et al. Mutation in TET2 in myeloid cancers. N Engl J Med. 2009; 360:2289–2301.10.1056/NEJMoa0810069 [PubMed: 19474426]
18. Figueroa ME, et al. Leukemic IDH1 and IDH2 mutations result in a hypermethylation phenotype, disrupt TET2 function, and impair hematopoietic differentiation. Cancer Cell. 2010; 18:553–567.10.1016/j.ccr.2010.11.015 [PubMed: 21130701]
19. Ko M, et al. Impaired hydroxylation of 5-methylcytosine in myeloid cancers with mutant TET2. Nature. 2010; 468:839–843.10.1038/nature09586 [PubMed: 21057493]

20. Ficz G, et al. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. Nature. 2011; 473:398–402.10.1038/nature10008 [PubMed: 21460836]
21. Williams K, et al. TET1 and hydroxymethylcytosine in transcription and DNA methylation fidelity. Nature. 2011; 473:343–348.10.1038/nature10066 [PubMed: 21490601]
22. Chen J, et al. Vitamin C modulates TET1 function during somatic cell reprogramming. Nat Genet. 2013; 45:1504–1509.10.1038/ng.2807 [PubMed: 24162740]
23. Gao Y, et al. Replacement of Oct4 by Tet1 during iPSC induction reveals an important role of DNA methylation and hydroxymethylation in reprogramming. Cell Stem Cell. 2013; 12:453–469.10.1016/j.stem.2013.02.005 [PubMed: 23499384]
24. Morin RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011; 476:298–303.10.1038/nature10351 [PubMed: 21796119]
25. Pasqualucci L, et al. Analysis of the coding genome of diffuse large B-cell lymphoma. Nat Genet. 2011; 43:830–837.10.1038/ng.892 [PubMed: 21804550]
26. Okosun J, et al. Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. Nat Genet. 2014; 46:176–181.10.1038/ng.2856 [PubMed: 24362818]
27. Li H, et al. Mutations in linker histone genes HIST1H1 B, C, D, and E; OCT2 (POU2F2); IRF8; and ARID1A underlying the pathogenesis of follicular lymphoma. Blood. 2014; 123:1487–1498.10.1182/blood-2013-05-500264 [PubMed: 24435047]
28. Dawlaty MM, et al. Tet1 is dispensable for maintaining pluripotency and its loss is compatible with embryonic and postnatal development. Cell Stem Cell. 2011; 9:166–175.10.1016/j.stem.2011.07.010 [PubMed: 21816367]
29. Morin RD, et al. Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood. 2013; 122:1256–1265.10.1182/blood-2013-02-483727 [PubMed: 23699601]
30. Liu M, et al. Two levels of protection for the B cell genome during somatic hypermutation. Nature. 2008; 451:841–845.10.1038/nature06547 [PubMed: 18273020]
31. Shaffer AL 3rd, Young RM, Staudt LM. Pathogenesis of human B cell lymphomas. Annu Rev Immunol. 2012; 30:565–610.10.1146/annurev-immunol-020711-075027 [PubMed: 22224767]
32. Pasqualucci L, et al. AID is required for germinal center-derived lymphomagenesis. Nat Genet. 2008; 40:108–112.10.1038/ng.2007.35 [PubMed: 18066064]
33. Ivanova NB, et al. A stem cell molecular signature. Science. 2002; 298:601–604.10.1126/science.1073823 [PubMed: 12228721]
34. Bhattacharyya S, et al. Genome-wide hydroxymethylation tested using the HELP-GT assay shows redistribution in cancer. Nucleic Acids Res. 2013; 41:e157.10.1093/nar/gkt601 [PubMed: 23861445]
35. Meissner A, et al. Genome-scale DNA methylation maps of pluripotent and differentiated cells. Nature. 2008; 454:766–770.10.1038/nature07107 [PubMed: 18600261]
36. Lara-Astiaso D, et al. Immunogenetics. Chromatin state dynamics during blood formation. Science. 2014; 345:943–949.10.1126/science.1256271 [PubMed: 25103404]
37. Campbell C, Risueno RM, Salati S, Guezguez B, Bhatia M. Signal control of hematopoietic stem cell fate: Wnt, Notch, and Hedgehog as the usual suspects. Curr Opin Hematol. 2008; 15:319–325.10.1097/MOH.0b013e328303b9df [PubMed: 18536569]
38. Lin HK, Bergmann S, Pandolfi PP. Deregulated TGF-beta signaling in leukemogenesis. Oncogene. 2005; 24:5693–5700.10.1038/sj.onc.1208923 [PubMed: 16123802]
39. Kikushige Y, et al. Self-renewing hematopoietic stem cell is the primary target in pathogenesis of human chronic lymphocytic leukemia. Cancer Cell. 2011; 20:246–259.10.1016/j.ccr.2011.06.029 [PubMed: 21840488]
40. A clinical evaluation of the International Lymphoma Study Group classification of non-Hodgkin's lymphoma. The Non-Hodgkin's Lymphoma Classification Project. Blood. 1997; 89:3909–3918. [PubMed: 9166827]
41. De S, et al. Aberration in DNA methylation in B-cell lymphomas has a complex origin and increases with disease severity. PLoS Genet. 2013; 9:e1003137.10.1371/journal.pgen.1003137 [PubMed: 23326238]

42. Martin-Subero JI, et al. A comprehensive microarray-based DNA methylation study of 367 hematological neoplasms. PLoS One. 2009; 4:e6986.10.1371/journal.pone.0006986 [PubMed: 19750229]
43. Matsuda I, Imai Y, Hirota S. Distinct global DNA methylation status in B-cell lymphomas: immunohistochemical study of 5-methylcytosine and 5-hydroxymethylcytosine. J Clin Exp Hematop. 2014; 54:67–73. [PubMed: 24942948]
44. Guo J, et al. Differential DNA methylation of gene promoters in small B-cell lymphomas. Am J Clin Pathol. 2005; 124:430–439.10.1309/LCGN-V77J-464L-NFD6 [PubMed: 16191512]
45. Quivoron C, et al. TET2 Inactivation Results in Pleiotropic Hematopoietic Abnormalities in Mouse and Is a Recurrent Event during Human Lymphomagenesis. Cancer Cell. 2011; 20:25–38.10.1016/j.ccr.2011.06.003 [PubMed: 21723201]
46. Asmar F, et al. Genome-wide profiling identifies a DNA methylation signature that associates with TET2 mutations in diffuse large B-cell lymphoma. Haematologica. 2013; 98:1912–1920.10.3324/haematol.2013.088740 [PubMed: 23831920]
47. Hashimshony T, Zhang J, Keshet I, Bustin M, Cedar H. The role of DNA methylation in setting up chromatin structure during development. Nat Genet. 2003; 34:187–192.10.1038/ng1158 [PubMed: 12740577]
48. Grant S, Easley C, Kirkpatrick P. Vorinostat. Nat Rev Drug Discov. 2007; 6:21–22.10.1038/nrd2227 [PubMed: 17269160]
49. Weigert O, et al. Molecular ontogeny of donor-derived follicular lymphomas occurring after hematopoietic cell transplantation. Cancer discovery. 2012; 2:47–55.10.1158/2159-8290.CD-11-0208 [PubMed: 22585168]
50. Moran-Crusio K, et al. Tet2 loss leads to increased hematopoietic stem cell self-renewal and myeloid transformation. Cancer Cell. 2011; 20:11–24.10.1016/j.ccr.2011.06.001 [PubMed: 21723200]
51. Di Tullio A, et al. CCAAT/enhancer binding protein alpha (C/EBP(alpha))-induced transdifferentiation of pre-B cells into macrophages involves no overt retrodifferentiation. Proc Natl Acad Sci USA. 2011; 108:17016–17021.10.1073/pnas.1112169108 [PubMed: 21969581]
52. Shaknovich R. DNA methylation signatures define molecular subtypes of diffuse large B-cell lymphoma. Blood. 2010; 116:e81–89.10.1182/blood-2010-05-285320 [PubMed: 20610814]
53. Geng H, et al. Integrative epigenomic analysis identifies biomarkers and therapeutic targets in adult B-acute lymphoblastic leukemia. Cancer Discov. 2012; 2:1004–1023.10.1158/2159-8290.CD-12-0208 [PubMed: 23107779]
54. Brune V, et al. Origin and pathogenesis of nodular lymphocyte-predominant Hodgkin lymphoma as revealed by global gene expression analysis. J Exp Med. 2008; 205:2251–2268.10.1084/jem.20080809 [PubMed: 18794340]
55. Ehrich M, et al. Quantitative high-throughput analysis of DNA methylation patterns by base-specific cleavage and mass spectrometry. Proc Natl Acad Sci USA. 2005; 102:15785–15790.10.1073/pnas.0507816102 [PubMed: 16243968]
56. Krzywinski M, et al. Circos: an information aesthetic for comparative genomics. Genome Res. 2009; 19:1639–1645.10.1101/gr.092759.109 [PubMed: 19541911]
57. Adli M, Zhu J, Bernstein BE. Genome-wide chromatin maps derived from limited numbers of hematopoietic progenitors. Nat Methods. 2010; 7:615–618.10.1038/nmeth.1478 [PubMed: 20622861]
58. Akalin A, et al. Base-pair resolution DNA methylation sequencing reveals profoundly divergent epigenetic landscapes in acute myeloid leukemia. PLoS Genet. 2012; 8:e1002781.10.1371/journal.pgen.1002781 [PubMed: 22737091]

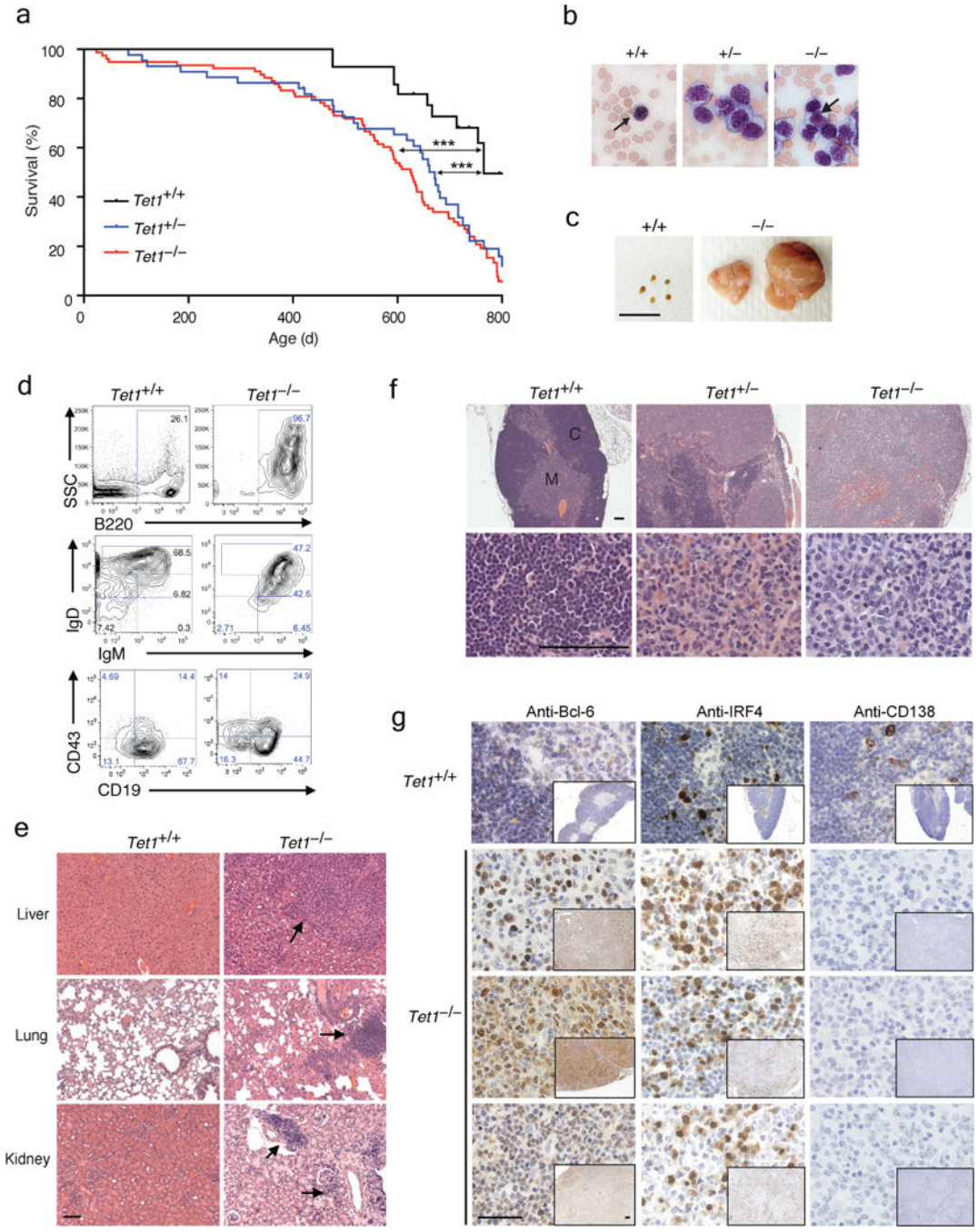

Figure 1. Tet1-deficiency drives B cell malignancy upon advanced age

a) Kaplan-Meier survival curve of Tet1-deficient mice with heterozygous (*Tet1*$^{+/-}$, n = 44) and homozygous (*Tet1*$^{-/-}$, n = 78) deletion compared to wild-type mice (*Tet1*$^{+/+}$ n = 28). * *P* = <0.0005. **b)** Peripheral blood smears stained with Wright-Giemsa; arrows indicate normal lymphocyte (left panel) and aberrant lymphocyte (right panel). Representative of n = 8–10 mice per genotype. **c)** Lymph nodes from *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice; representative of n = 8–10 mice per genotype. Scale bar = 1cm. **d)**. Flow cytometric analysis of malignant cells in the lymph nodes of moribund *Tet1*$^{-/-}$ compared to age-matched *Tet1*$^{+/+}$ mice. Upper panels;

side-scatter (SSC) vs. B cells ($B220^+$). Middle Panels; $B220^+$ gated cells stained for progenitor/precursor B cells (IgM^-IgD^-), immature B cells (IgM^+IgD^-), transitional B cells (IgM^+IgD^{low}) and mature B cells (IgM^+IgD^{high}). Lower panels; $B220^+$ gated cells stained for progenitor B cell marker CD43 vs. CD19 staining. Data are representative of 3 independent experiments, n = 8–10 mice per genotype. Histological analysis by H&E staining of **e)** liver, lung, kidney and **f)** lymph node sections from sick *Tet1*$^{+/-}$ and *Tet1*$^{-/-}$ mice compared to *Tet1*$^{+/+}$ controls. Arrows indicate infiltration of lymphocytes. C = cortex, M = Medulla; Scale bar = 100 μm in all panels **g)** Immunohistochemistry of lymph nodes from *Tet1*$^{+/+}$ and sick *Tet1*$^{-/-}$ mice stained with anti-Bcl-6, anti-IRF4, anti-CD138 antibodies (brown) and hematoxylin. Scale bar = 100 μm in all panels. Data are representative of n = 4 mice per genotype.

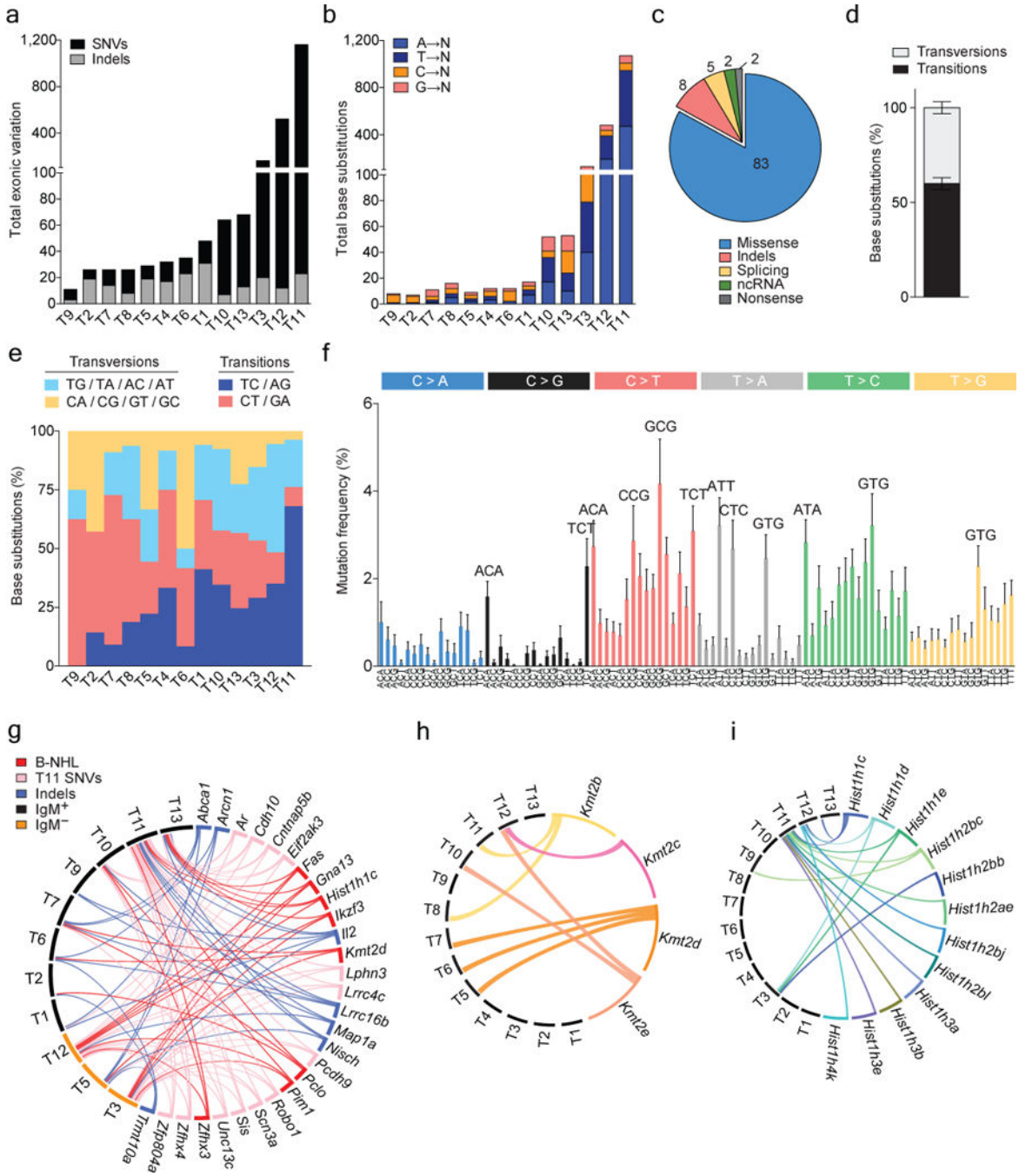

Figure 2. Whole-exome sequencing of Tet1-deficient tumors reveals mutations of B-non Hodgkin's lymphoma

a) Total number of non-synonymous single nucleotide variants (nsSNVs) and insertions and deletions (Indels) detected by whole exome sequencing in thirteen Tet1-deficient tumor cell populations (T1-13) in order of increasing mutation load. **b)** Total number of nsSNVs divided into A, T, C or G base substitutions, ordered from left to right in tumors according to increasing total exonic variations. **c)** The frequency of missense, indel, splicing, ncRNA, and nonsense mutations found in Tet1-deficient tumors and **d)** the average frequency of

nsSNVs that are transition and transversion mutations (mean ± SEM, n = 13) and **e)** according to base substitution frequency per tumor. **f)** Average mutation frequency of base substitutions in Tet1-deficient tumors according to trinucleotide context (mean ± SEM, n = 13). Circos plots of **g)** the most frequently mutated genes and their co-occurrence in Tet1-deficient tumors; red = B-NHL recurrently mutated genes, pink = T11 nsSNVs, blue = indels, black = IgM^+ tumors, orange = IgM^- tumors. Circos plots of mutations and their co-occurrence in **h)** histone-modifying enzymes and **i)** histone cluster 1 genes.

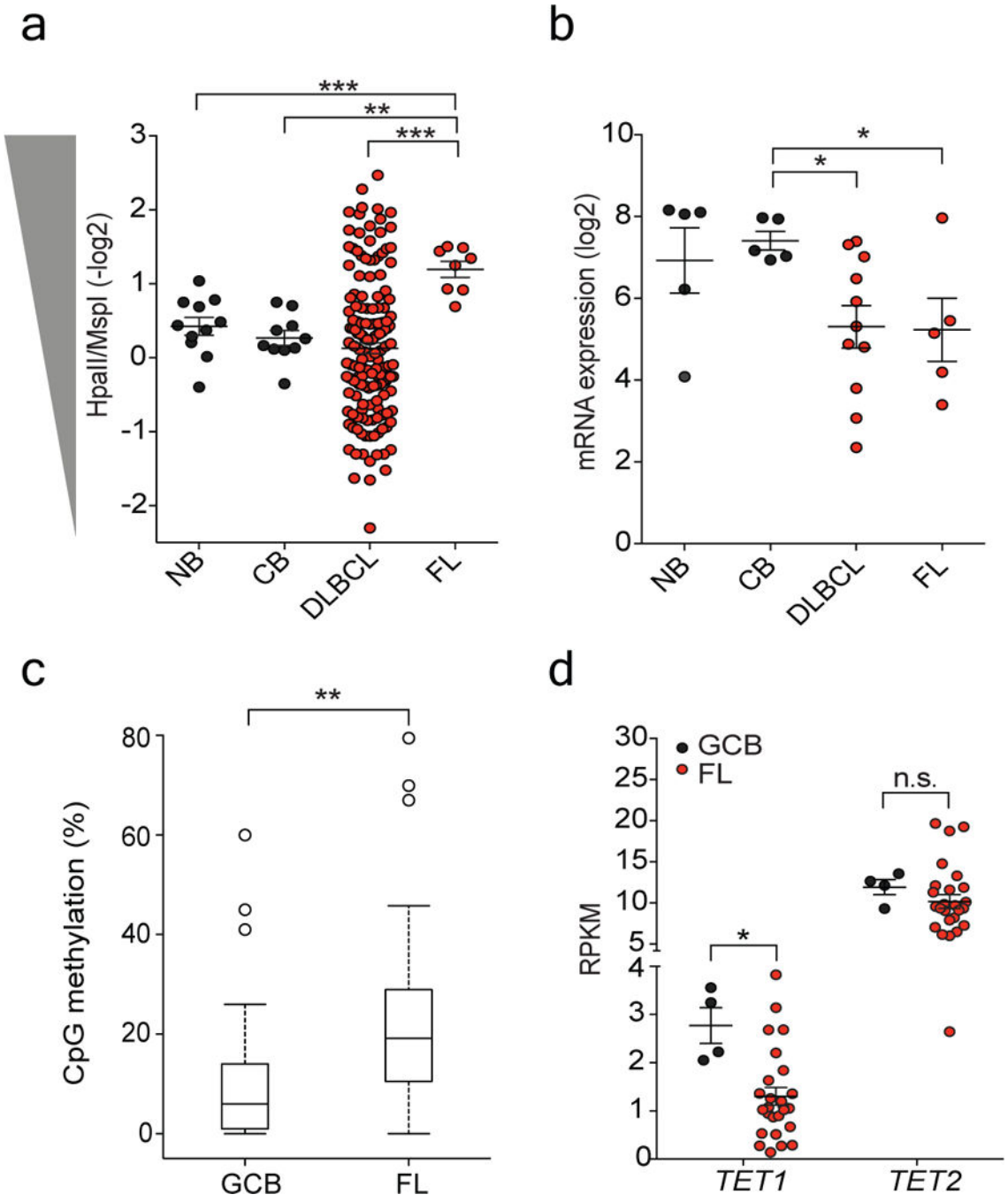

Figure 3. *TET1* is hypermethylated and transcriptionally down-regulated in B-NHL

a) Methylation profiling by HELP-assay of the *TET1* promoter in DLBCL and FL patients compared to normal naïve (NB) and centroblast (CB) B cells. **b)** mRNA expression analysis of *TET1* in B-NHL patients. **c)** MassARRAY analysis of *TET1* CpG methylation in 26 FL patients compared to normal Germinal Center B (GCB) cells and **d)** RNA-seq of *TET1* and *TET2* expression in the same 26 FL patients and normal GCB cells. Each circle indicates an individual patient in all panels; * P = <0.05, ** P = <0.005, *** P = <0.0005.

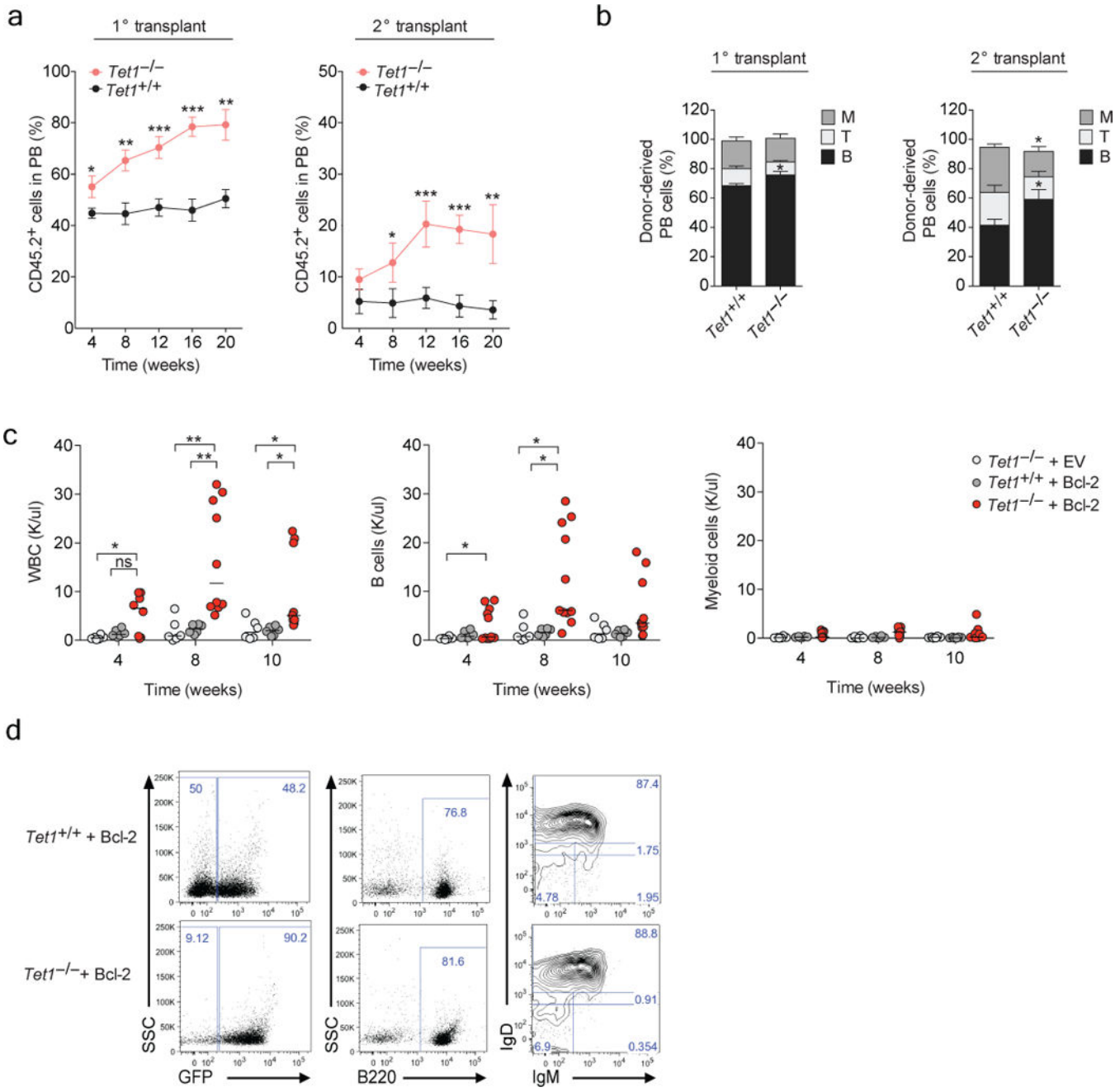

Figure 4. Tet1-deficient hematopoietic stem display increased self-renewal *in vivo* with a bias toward B cell differentiation

Competitive bone marrow (BM) reconstitution assays. Primary transplants were performed with CD45.2$^+$ *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ total BM cells (200,000 per mouse), mixed in equal ratio with CD45.1$^+$ support BM cells and transplanted into lethally irradiated recipient mice. 20 weeks post transplant, CD45.2$^+$ *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ purified LT-HSCs (500 per mouse) were serially transplanted into lethally irradiated recipient mice with CD45.1$^+$ support BM. **a)** Frequency of donor-derived CD45.2$^+$ cells in the peripheral blood (PB) of primary and secondary transplanted mice. Data are the average of two independent experiments (mean ±

SEM, n = 3 recipient mice per donor BM), n = 2 donor BM per genotype, per experiment. **b)** Average frequency of $CD45.2^{+}Lineage^{+}$ cells stained for CD11b/Gr1 (M), B220 (B) and CD3 (T) surface markers in peripheral blood 20-weeks post transplant in secondary recipient mice (mean ± SEM, n = 6 mice per genotype). Loss of *Tet1* cooperates with Bcl2 overexpression to drive B lymphocytosis in mice. Purified LSK cells from $Tet1^{+/+}$ and $Tet1^{-/-}$ mice were transduced with either pMIG-Bcl2 or an empty vector control retrovirus, and transplanted into lethally irradiated recipient mice (5000 LSKs were injected per recipient with 200,000 wild-type support bone marrow cells). **c)** Total numbers of GFP^{+} leukocytes, GFP^{+} B lymphocytes and GFP^{+} myeloid cells monitored 4, 8 and 10 weeks post-transplant in the peripheral blood of recipient mice. **d)** Representative flow cytometric analysis of peripheral blood for $CD45.2^{+}$ GFP^{+} cells, with B220, IgM and IgD staining as indicated (n = 6–12 mice per genotype). Small horizontal lines indicate the mean. * P = <0.01 ** P = <0.001, *** P = <0.0001 in all panels.

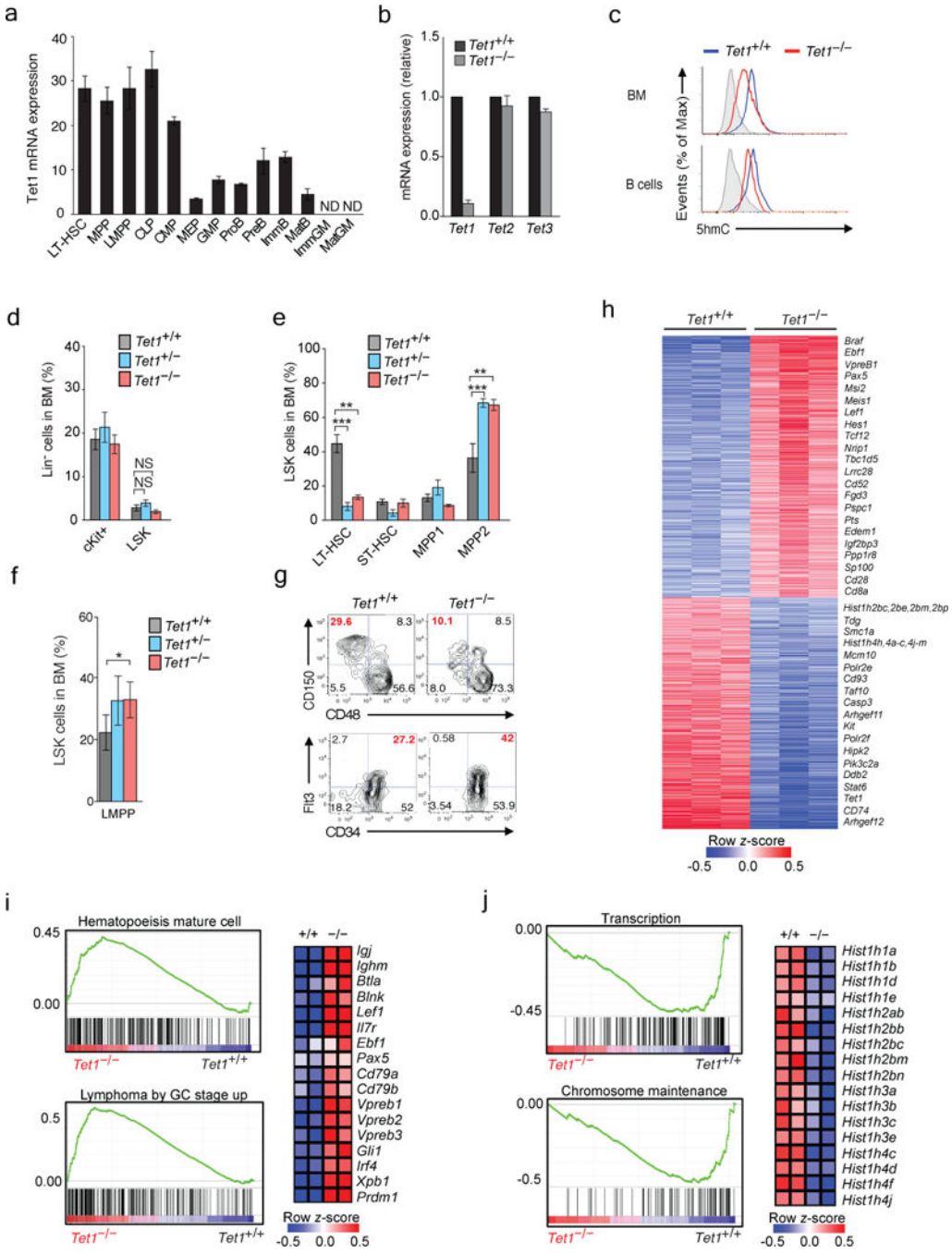

Figure 5. Loss of *Tet1* in hematopoietic stem cells promotes differentiation with a lymphoid bias

a) *Tet1* mRNA expression abundance in long-term hematopoietic stem cells (LT-HSC), multi-potent progenitors (MPP), lymphoid-primed multipotent progenitors (LMPP), common lymphoid (CLP) and myeloid progenitors (CMP), megakaryocyte and erythroid progenitor (MEP), granulocyte and macrophage progenitor (GMP), progenitor B (ProB), precursor B (PreB), Immature B (ImmB), Mature splenic B (MatB), immature myeloid cells in the bone marrow (ImmGM) and mature myeloid cells in the spleen (MatGM) of wild-type mice. *Tet1* mRNA expression was normalized to *Hprt* mRNA (mean ± SEM, n = 3

experiments). **b)** Loss of *Tet1* mRNA expression upon deletion in hematopoietic stem and progenitor cells ($Lin^{-}cKit^{+}$ cells) measured by qPCR and normalized to Hprt (mean and SEM, n = 3 experiments). **c)** Intracellular flow cytometric analysis of 5-hydroxymehtylcytosine (5hmC) in total bone marrow (BM) and B cells from $Tet1^{+/+}$ and $Tet1^{-/-}$ mice (data are representative of 3 experiments). **d–f)** Frequency of $Lin^{-}Sca1^{+}cKit^{+}$ (LSK) subsets in old Tet1-deficient mice compared to wild-type mice (mean ± SEM, n = 6–8 mice per genotype). * P = <0.01, ** P = <0.001, *** P = <0.0001. **g)** Representative flow cytometric analysis of $Tet1^{+/+}$ and $Tet1^{-/-}$ LSKs. Upper panels = LT-HSC ($CD150^{+}CD48^{-}$), ST-HSC ($CD150^{-}CD48^{-}$), MPP1 ($CD150^{+}CD48^{+}$) and MPP2 ($CD150^{-}CD48^{+}$); Lower panels = LMPP ($Flt3^{+}CD34^{+}$). Data are representative of 6–8 mice per genotype. **h)** Microarray analysis of LSK cells. Significantly differentially expressed genes (P = <0.05) with changes >2-fold are displayed. GSEA analysis of RNA-seq from purified $Tet1^{+/+}$ and $Tet1^{-/-}$ MPP cells with corresponding heatmaps of leading edge gene expression changes in i) B lineage and j) histone cluster 1 genes.

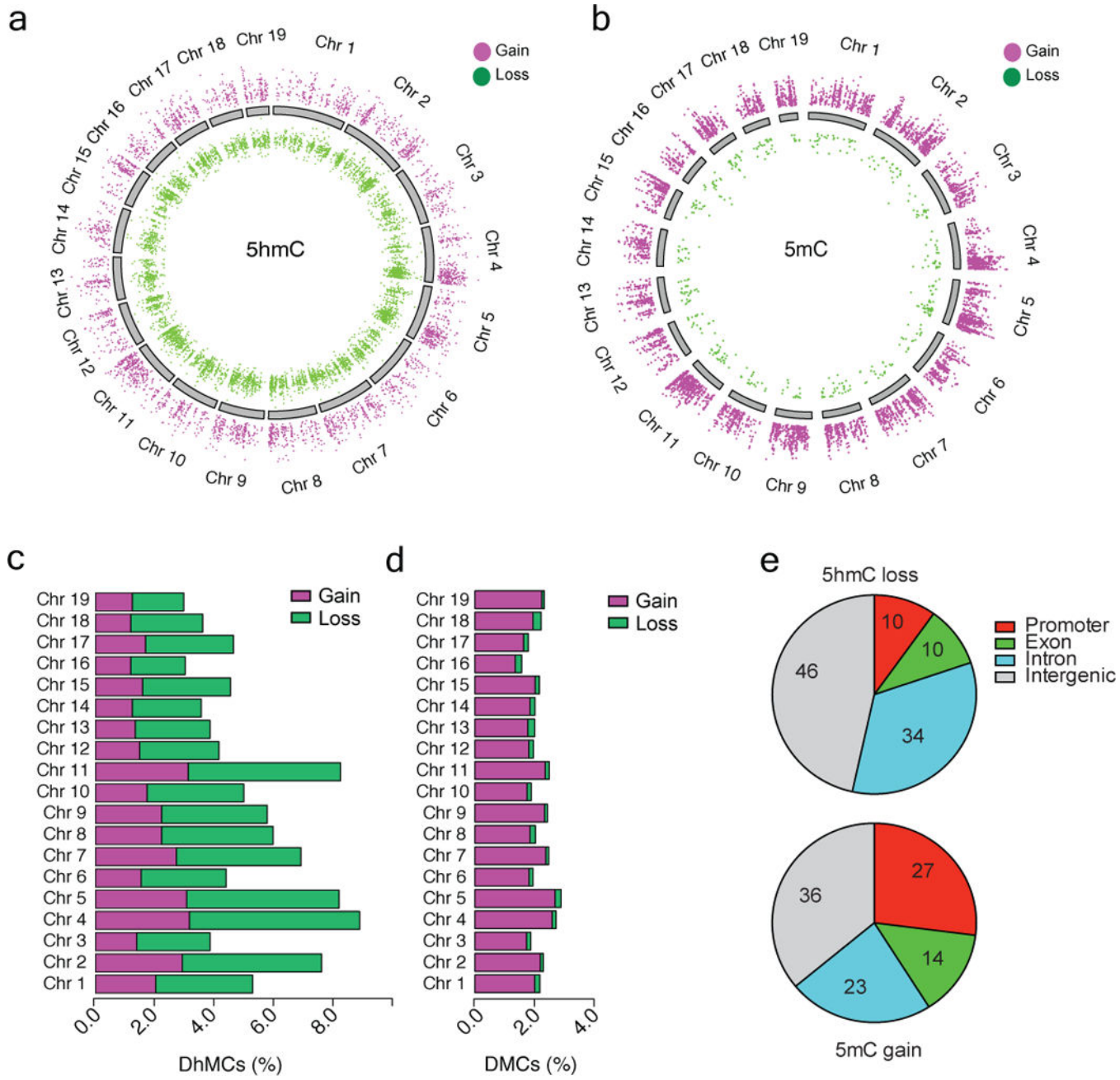

Figure 6. Aberrant DNA hydroxymethylation of Tet1-deficient stem and progenitor cells

DNA sequencing data was obtained from purified bone marrow LSK cells of 6 month-old *Tet1*$^{+/+}$ and *Tet1*$^{-/-}$ mice using HELP-GT (n = 2 per genotype) and RRBS (n = 3 per genotype) assays for quantitation of 5hmC and 5mC, respectively. Graphic representation of the differentially methylated genomic regions are displayed as circos plots for **a)** 5hmC and **b)** 5mC with respect to their chromosomal location. Bar graphs depict the overall frequency of **c)** differential hydroxymethylation (%DhMC) compared to **d)** differential methylation (%DMC) per chromosome and **e)** genomic location. Differential hyper- and hypo-

methylated sites were calculated for CpG sites with methylation difference cutoff >25%, q-value < 0.01.

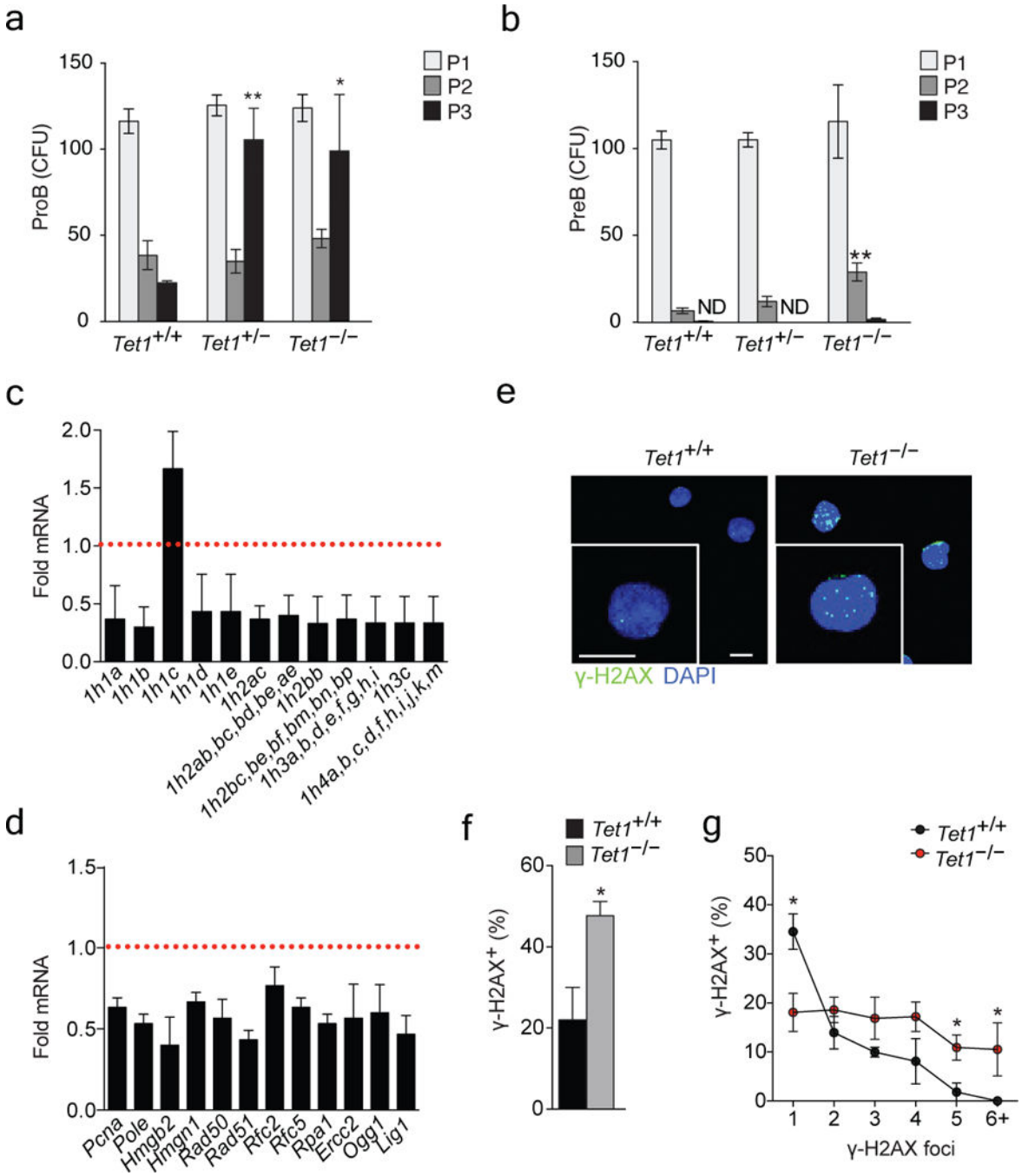

Figure 7. Enhanced colony formation and accumulation of DNA damage in *Tet1*- deficient progenitor B cells

Colony-formation assays in methylcellulose media were performed to measure **a)** pro-B and **b)** pre-B self-renewal capacity. Cells were passaged every 7 days for 3 successive weeks (P1-3), (mean ± SEM, n = 3 experiments). Quantitative mRNA expression analysis of pro-B cells from P1 for the relative fold-change in expression of **c)** histone cluster 1 gene variants and **d)** DNA repair genes in $Tet1^{-/-}$ compared to $Tet1^{+/+}$ (mean + SEM, n = 3 experiments, $P = <0.05$). **e)** Immunofluorescence staining of γH2AX foci in purified ProB cells from

Tet1$^{+/+}$ and *Tet1*$^{-/-}$ mice, scale bar = 10μm. Quantitation of f) the frequency of γH2AX positive cells and the number of foci per positive cell (mean ± SEM, n = 3 per genotype). In all panels: * *P* = <0.05, ** *P* = <0.005.