PhD degree in Molecular Medicine

European School of Molecular Medicine (SEMM),

University of Milan and University of Naples "Federico II"

Faculty of Medicine

Settore disciplinare: BIO/11

# Sleeping Beauty finds cooperating partners of oncogenic Myc

## A transposon-based insertional mutagenesis screen in the Eμ-*myc* mouse.

*Heide-Marie Binder*

IFOM-IEO Campus, Milan

Matricola n. R08913

*Supervisor:* Dr. / Prof. Bruno Amati

IFOM-IEO Campus, Milan

*Added Supervisor:* Dr. / Prof. Stefano Casola

IFOM-IEO Campus, Milan

Anno accademico 2012-2013

**ABSTRACT** ..... **4**

**INTRODUCTION** ..... **5**
Myc ..... 5
*Myc in general* ..... *5*
*Myc and Transcription* ..... *5*
*Transforming Myc and the paradigm between proliferation and apoptosis* ..... *7*
Eμ-*myc* mouse model ..... 9
B-cell lymphoma ..... 12
*Hematopoiesis* ..... *12*
*Normal B-cell development* ..... *13*
*Myc and hematopoiesis* ..... *14*
*Myc involvement in hematologic disease* ..... *15*
The Sleeping Beauty System ..... 16
*Insertional mutagenesis, drivers and passengers: Cancer gene discovery* ..... *16*
*History and basics of transposons* ..... *18*
*The SB transposon system: structure and function* ..... *19*
*Transposons* ..... *20*
*Transposase* ..... *22*
*The Cre-LoxP system* ..... *23*
*Studies of conditional SB screens* ..... *23*

**MATERIALS AND METHODS** ..... **26**
Mice. ..... 26
Adoptive transplant model. ..... 26
Tumor analysis ..... 27
Flow Cytometry. ..... 27
Linker-Mediated PCR. ..... 27
Sequence analysis. ..... 28
Statistical evaluation of clonal insertion sites ..... 30
Statistical evaluation of common insertion sites ..... 31
Orientation of the gene ..... 33

**AIM** ..... **37**

**RESULTS** ..... **38**
Establishment of the mouse model and basic concepts of the SB screen ..... 38
*Breeding strategy and mice* ..... *38*
*Basic concepts of Sleeping Beauty screens* ..... *41*
*Monitoring the SB system* ..... *42*
*Adoptive transplantation serves as a tool for large-scale screening* ..... *43*
*Validation of the adoptive transplant strategy* ..... *44*
Strong genetic cooperation of SB-mutagenesis in the Eμ-myc mouse ..... 48
SB-induced tumors show heterogeneous lymphomagenesis ..... 52
Identification of SB-induced modifiers of Myc-induced lymphoma ..... 60
*Sequencing* ..... *60*
*Read analysis* ..... *64*
*Clonal Insertion sites* ..... *65*
*Common Insertion sites* ..... *71*
Characterization of CIS ..... 78
*Cross-Species Comparison* ..... *78*
*Common pathways and functions of genes cooperating with Myc* ..... *81*

**DISCUSSION** ..... **91**
Experimental system ..... 93
Insertion sites and drivers found ..... 96
Connection of CIS to Myc ..... 98
*Blocking senescence* ..... *102*
*Changing transcription* ..... *103*

**SUPPLEMENTS** ..... **105**

**REFERENCES** ..... **106**

## ABSTRACT

Myc triggers a transcriptional program inducing hyper-replication and proliferation but also tumor suppressive mechanisms like apoptosis. Therefore, Myc dependent tumors display high selective pressure to accumulate secondary mutations blocking these tumor-suppressive pathways. In the Eμ-*myc* mouse model, Myc is constitutively expressed in the B-cell linage under the control of the immunoglobulin heavy chain enhancer. The most prominent failsafe program known to be disrupted in Eμ-*myc* lymphomas is the p53-ARF pathway. In order to find novel cooperating partners of Myc leading to transformation we applied a conditional Sleeping Beauty transposon-based mutagenesis screen *in vivo*. By adoptive transplantation of Eμ-*myc* hematopoietic progenitors we generated 312 experimental animals, prone for lymphoma onset. Data show a strong genetic cooperation between the Eμ-*myc* transgene and SB transposon mobilization with accelerated tumor onset. Arising lymphomas were of the pre/pro- and immature B-cell stage and infiltration included extra-lymphatic tissue like the liver. The genomic sequences immediately adjacent to integrated transposons of 184 lymphomas were enriched by ligation-mediated PCR and were sequenced with a multiplexed approach. Based on published (Brett et al., 2011) and new bioinformatic methods, we identify 338 common integration sites (CIS) of which 188 were found mutated in human B-cell lymphoma. Pathway and GO term analysis reveal modulation of the Ras-MAP-kinase signaling pathway. Next to well-known modulators of Myc-induced lymphomagenesis including Bcl-XL, p53, ARF and Mdm2, we find CIS that were not yet reported in the Eμ-*myc* lymphoma like Map3K5.

# INTRODUCTION

## Myc

### Myc in general

More than 30 years ago, v- *myc* was discovered as the transforming part of the avian myelocytomatosis retrovirus MC29. Its homolog *c-Myc* (hereafter *myc*) and also the family members *N*- and *L-Myc* were further shown to be involved in transformation. In normal cells it was shown to be essential, as homozygotes for targeted *N-Myc* null mutations die between 10.5 and 12.5 days of gestation (Charron et al., 1992) and *myc* knockout embryos die before 10.5 days of gestation (Davis et al., 1993). Myc is highly conserved among species and has a short mRNA and protein half-life (reviewed in Dang, 2012). For being one of the first oncogenes studied, it is amazing that still many of its functions and mechanisms are not fully understood.

### Myc and Transcription

Myc is in the basic sense of its function a transcription factor that can only execute its function when binding to the protein Max (Figure 1). The basic helix loop helix and ZIP motif harbor Myc's recognition sequence for a manifold of E-boxes with the consensus sequence CACGTG dispersed in the genome (Lüscher and Larsson, 1999). Due to competition for its binding partner, by Max itself and also by other bHLH-domain containing proteins of the Mad/Mdx protein family, Myc's transcriptional activation is repressed (Amati et al., 1993; Grandori et al., 2000).

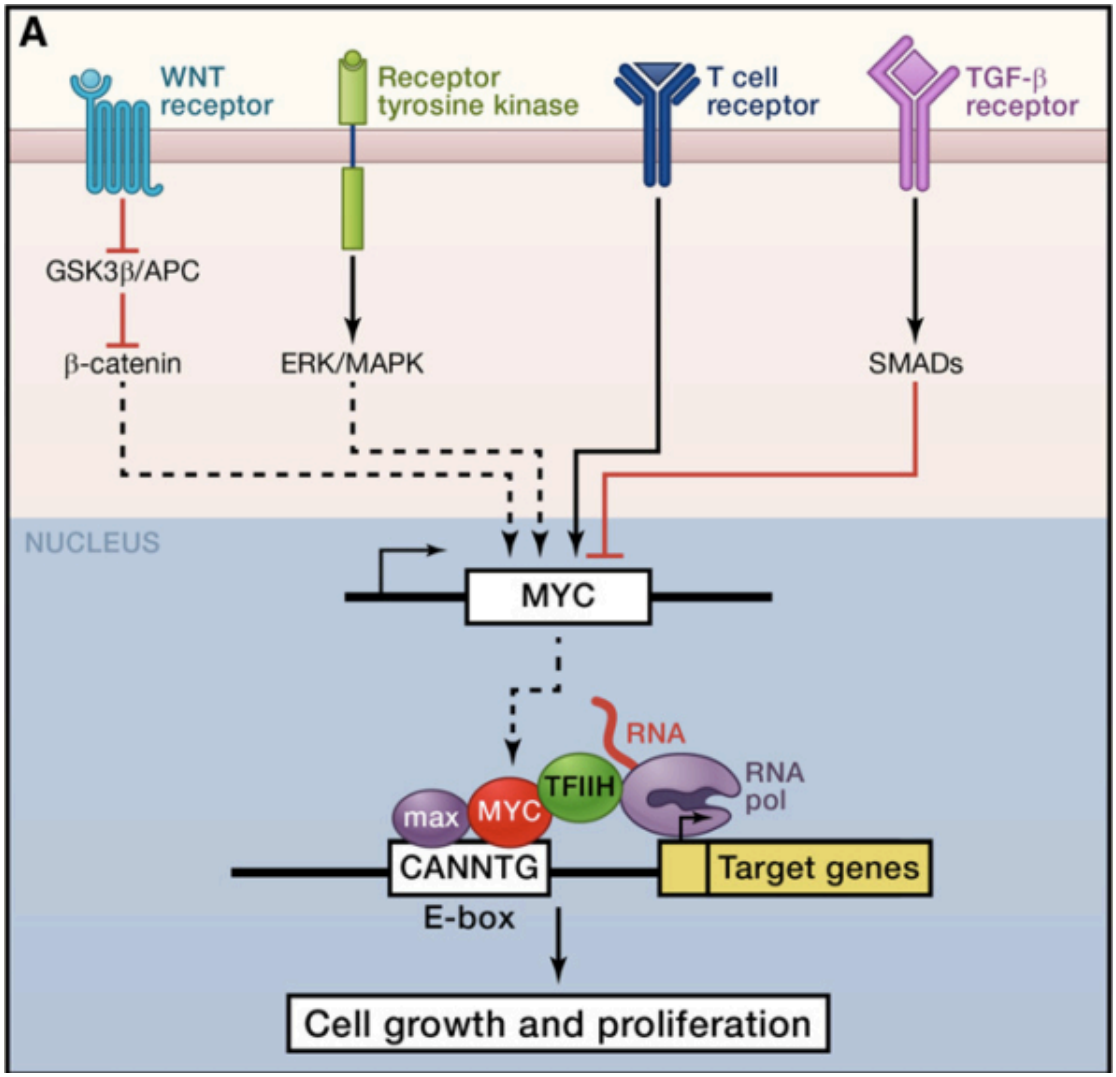


**Figure 1 Myc's function in transcription.** Mitogenic signals activate transcription of Myc, which in turn, together with its binding partner Max induces genes that lead to cell growth and proliferation. (Dang, 2012)

Moreover, Myc regulates transcription by recruitment of chromatin remodeling complexes that enhance the accessibility of DNA. Myc recruits acetyltransferase and ATPase containing complexes including the STAGA, Tip60 and SWI/SNF complex, that lead to an open chromatin conformation via TRRAP and SNF5 (Amati et al., 2001). Heterodimers composed of Max and a Mad/Mxi family member, on the contrary recruit chromatin modifying complexes that are involved in compaction of DNA, like the Sin3-complex which contains histone deacetylases (HDACs) and thus antagonize Myc's effect on chromatin. Moreover, transcriptional elongation is achieved by recruitment of the pTFEB complex, that phosphorylates the C-terminus of RNA Polymerase II abrogating transcriptional pausing (Rahl et al., 2010). Yet another layer of regulation is the displacement of core-transcription

activators like Miz-1 (Seoane et al., 2002) and Sp1 (Gartel et al., 2001) repressing transcribed genes and recruitment of methyltransferases like DNMT3a that render the chromatin inaccessible (Brenner et al., 2005). Myc's transcriptional regulation, depicted schematically in Figure 1, controls an estimate of 15% of all genes, which makes it challenging to find common determinants for the general function of this gene and lead to suggestions that Myc instead of being selective for its targets, amplifies transcription of already active genes (Littlewood et al., 2012). However, the main molecular functions associated with Myc are cell cycle, protein biosynthesis, regulation of metabolism and replication (Dang et al., 2006). These functions are in normal cells induced by mitogenic signaling downstream of many receptors including IL2-, MCSF-, Epo-, EGF- and PDGF-receptor with involvement of the Wnt, Ras, PI3K and MAPK signaling pathways leading to activation of Myc transcription (Grandori et al., 2000; Oster et al., 2002). In summary, Myc was defined by Oster et al. as "the intracellular sentinel of the extracellular milieu" as it senses external stimuli that allow growth and proliferation (Oster et al., 2002).

**Transforming Myc and the paradigm between proliferation and apoptosis**

Myc was found to be a major target driving tumorigenesis in retroviral mutagenesis screens (Kool and Berns, 2009; Selten et al., 1984), was inducing tumors of the hematopoietic system, the skin, bones, mammary gland, brain, pancreas and in the liver in mouse models (Morton and Sansom, 2013) and was found to transform mouse embryonic fibroblasts (MEFs) in collaboration with other oncogenes (Alexander et al., 1989a; Jacobs et al., 1999; Land et al., 1983).

Strong induction of proliferation and cell growth contribute to Myc's oncogenic potential but is strictly dependent on mitogenic signals. In detail, Myc activates transcription of cyclins (D1, D2, E1, A2) and abrogates the expression of checkpoint genes (*E2f1, E2f2, Cdk4, Cdc25, Gadd45* and *Gadd153*) and cell cycle

inhibitors (*Cdkn2a* and *Cdkn2b*) boosting the cell cycle and pushing cells into a metabolically active state by increasing energy metabolism and biosynthesis. Deregulation of this progressive program however, does not lead to transformation as tight feedback regulations activate tumor suppressive pathways. Thus, like other oncogenes, Myc only predisposes cells to transformation and only by acquisition of secondary mutations the multistep process of tumorigenesis can occur (Hanahan, 1988).

Mechanistically, deregulated Myc induces strong transcriptional activation of *Cdkn2a* encoded protein ARF (Zindy et al., 1998) most probably by involvement of FoxO proteins (Bouchard et al., 2007). ARF, in turn binds and sequesters Mdm2, an ubiquitin ligase and negative regulator of p53 transactivation and stability (Eischen et al., 1999; Zhang et al., 1998). p53, a transcription factor itself then is able to activate programmed cell death including apoptosis and autophagy, cell cycle arrest and senescence. By a feedback loop, p53 transcriptionally induces Mdm2, which leads to a cessation of the response. Another fail-safe mechanism is Myc's repression of transcription of Bcl-2 and Bcl-XL, anti-apoptotic proteins of the BH-3 only family (Nilsson and Cleveland, 2003). Deregulation of the balance between the pro- (Bax, Bak, Bim, Bad, Noxa, Puma) and anti-apoptotic proteins (Bcl-2, Bcl-XL, Mcl-1) induces apoptosis via mitochondrial depolarization, cytochrome C release and caspase activation (Lowe et al., 2004).

Although Myc overexpression induces prevalently an apoptotic responses as result of Myc induced transcriptional suppression of cell cycle inhibitors (Seoane et al., 2002), also senescence is suggested as a tumor suppressive mechanism (Campaner and Amati, 2012; Gorrini et al., 2007).

## Eμ-*myc* mouse model

In the early 80ies a mouse model was established, that allowed the *in vivo* study of Myc induced apoptotic and proliferative responses, the Eμ-*myc* mouse (Adams et al., 1985). Myc is constitutively expressed in the B-cell linage by mimicking the Burkitt lymphoma translocation of the *myc* gene to the Immunoglobulin heavy chain (*IgH*) Eμ enhancer. The mouse develops B-cell lymphomas with disseminated cells that carry the B220 antigen, a marker for the B-cell lineage and are staged early in development. The major pathological characteristics were enlargement and infiltration of the lymph nodes, spleen, liver and occasionally also the thymus. The initial mouse model developed lymphoma around 80 days with full penetrance on a C57Bl6 and SLJ mixed background (Adams et al., 1985; Harris et al., 1988). Backcrossing for several generations with C57Bl6 mice showed reduced tumor latency ranging from 100 to 150 days (Bouchard et al., 2007; Eischen et al., 1999, 2001a; Garrison et al., 2008; Schmitt et al., 1999), which Harris suggested to be due to C57Bl6 specific genes that confer a relative resistance to lymphomagenesis (Harris et al., 1988). Interestingly, Myc overexpression starts in fetal liver, where the first B-cells develop. However, tumors only establish after weeks. This latency phase reflects the time needed for spontaneous secondary mutations to by-pass Myc-induced tumor suppressive pathways. In the last 30 years many genetic screens of such potentially involved genes have been performed, increasing our knowledge about basic mechanisms in tumor development. The most prominent failsafe program to be disrupted in the Eμ-*myc* mouse is the p53-Mdm2-ARF pathway (Eischen et al., 1999; Schmitt et al., 1999). Lymphomas are mutated either in p53 (28% of cases) or show biallelic ARF deletion (24% of cases) and Mdm2 haploinsufficiency delays lymphomagenesis. In summary this suggests an important role of this tumor suppressive pathway for Myc-induced lymphomagenesis (Alt et al., 2003; Eischen et al., 1999). The p53

mediated response includes mitochondrial-mediated apoptosis, which in more than 50% of spontaneous lymphomas of the Eμ-*myc* mouse is abrogated by overexpression of anti-apoptotic proteins Bcl-2 (Strasser et al., 1990) together with Bcl-XL (Eischen et al., 2001). Consistent with this, the inactivation of the pro-apoptotic proteins Bim (Egle et al., 2004), Bmf (Frenzel et al., 2010), Bax (Eischen et al., 2001) and Puma (Garrison et al., 2008) accelerates lymphomagenesis in Eμ-*myc* transgenic mice to different extents but is not always independent of p53.

The concept of Myc-induced senescence led to several studies in Eμ-*myc* mice that suggest that enhanced DNA damage response by blocking the damage sensing kinase ATM (Maclean et al., 2007; Reimann et al., 2008), blocking Tip60, an acetyltransferase (Gorrini et al., 2007) leads to acceleration of lymphoma onset, whereas activation of ATM by blocking the negative regulating phosphatase Wip-1 lead to increased latency (Shreeram et al., 2006) and is suggested to be a p53 dependent, ARF independent branch of Myc's tumor suppressive function (Sluss et al., 2010) as shown in Figure 2.

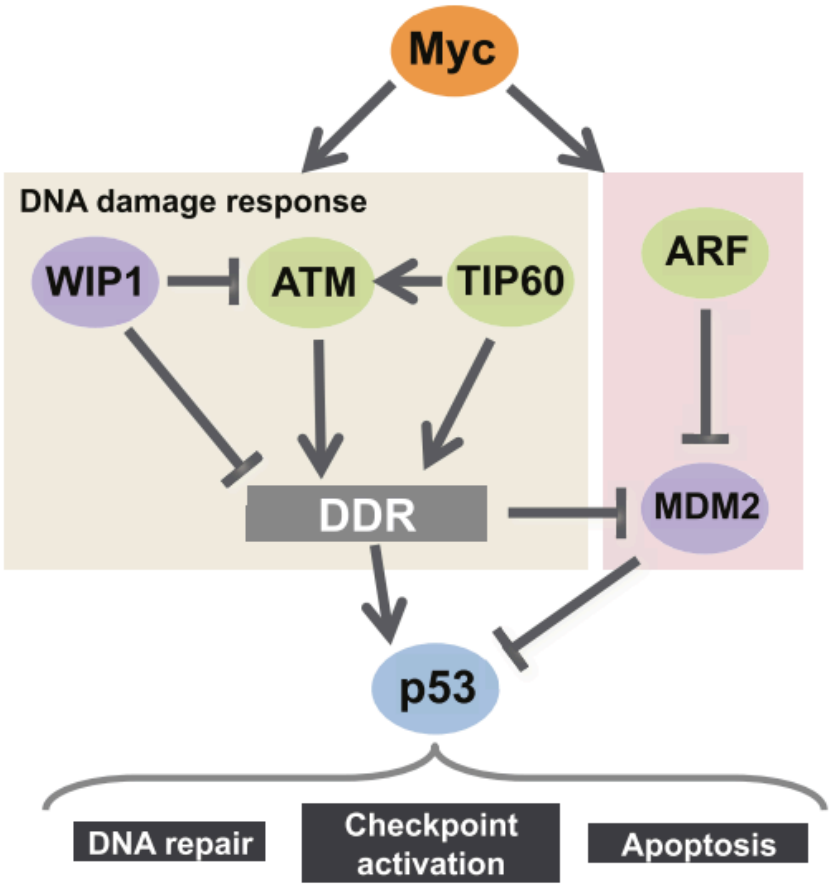


**Figure** 2 **Myc-induced tumor suppressive activation of p53.** (Campaner and Amati, 2012)

Moreover, delayed tumor onset was observed for *Cdk2* knock out mice due to induction of senescence (Campaner et al., 2009). Additionally, Myc's function to repress Miz-1 transcriptional activity directly connects cell cycle inhibitor expression with senescence. This is shown in a mouse model that harbors a mutation in the Miz-1 binding domain, leaving cells inactive to Myc-repressive functions which leads to senescence upon TGF-β stimulation (van Riggelen et al., 2010). Moreover, inhibition of a TGF-β induced senescence-related histone methyltransferase Suv39h1 resulted in acceleration of tumors (Reimann et al., 2010). Inhibition of the mTOR pathway restores senescence in Eμ-*myc* lymphoma leading to longer survival period (Wall et al., 2013). All together, this suggests senescence to be an important complementary tumor suppressive pathway in the Eμ-*myc* mouse model.

The imbalance of the Myc-induced response towards proliferation, impairing apoptosis, senescence and cell cycle arrest leads to the development of

lymphomas. Direct abrogation of genes involved in these mechanisms leads to tumor formation, but also activation of an additional oncogene can override failsafe mechanisms. 30 years ago, experiments in fibroblasts showed the transforming capabilities of Myc and Ras together, whereas either of them alone could not induce tumorigenesis (Land et al., 1983). Interestingly, Ras cooperation with Myc-induced large tumors that however were blocked by size restriction *in vivo*, whereas cooperation with large T antigen, a viral oncogene was shown to introduce limitless growth (Land et al., 1983). The collaborating effect is achieved as Myc is blocking the strong induction of senescence by Ras (Serrano et al., 1997), whereas Ras-signaling, which is strongly induced upon mitogenic signaling provides a survival signal that blocks the increased sensitivity of Myc expressing cells to apoptosis (Evan et al., 1992). The Eμ-*myc* mouse model was shown to harbor mutations in Ras, collaborate in tumorigenesis with Ras signaling and synergizes with Myc to abrogate tumor suppressive mechanism of apoptosis and senescence (Alexander et al., 1989a, 1989b; Gramling and Eischen, 2012).

Concluding, mechanisms abrogating apoptosis, senescence and cell cycle arrest are highly selected in the Eμ-*myc* mouse models and suggest the p53-ARF-Mdm2 axis of apoptosis, the p53-ATM-p21 axis of senescence and the partly p53 independent Bcl2 and BclX axis of apoptosis to be the key mechanism. Thus, parallel pathways acting in this failsafe program serve as a good therapy strategy to activate tumor regression.

## B-cell lymphoma

### Hematopoiesis

Hematopoiesis is a process in which many thousand cells of distinct lineages are produced every minute from hematopoietic stem cells to make up the blood. In

general hematopoietic stem cells (HSCs) divide and a daughter cell gives rise to multipotent progenitors (MPPs), while the other keeps the HSC state. MPPs further separate in the myeloid and the lymphoid lineage. The myeloid lineage further differentiates into megakaryocytes producing platelets, erythrocytes, granulocytes, monocytes, macrophages and mast cells, contributing to the innate immunity of an organism. The lymphocytic branch gives rise to B-cells, T-cells and natural killer cells. This lineage develops the adaptive immune system including CD4+ T helper cells and CD8+ cytotoxic T-cells and B-cells that also give rise to antibody producing plasma cells. Additionally, the NK cells and the γδ regulatory T-cells are part of the innate immune response. Dendritic cells, part of the innate immune system, evolve of both the myeloid and the lymphoid lineage.

The first step of distinction between the different lineages is primed in the bone marrow, where different niches and surrounding cells induce a change in the transcriptional program of hematopoietic stem and progenitor cells (HSPCs) via cytokines, growth factors and/or cell-to-cell contact. Focusing on the CLPs, the progenitors of the T-cells evade the bone marrow and differentiate further in the thymus. The B-cell lineage progenitors develop into differentiated sub-populations residing in the bone marrow. (Janeway et al., 2001)

**Normal B-cell development**

The final function of the B-cells is recognizing different antigens, which is achieved by the B-cell receptors (BCR). BCRs are heterodimers of pairs of immunoglobulin heavy and light chains that are expressed on the surface of the cell. The *Ig* genes are comprised of multiple segments of the variable (V), diversity (D) and joining (J) segments in the heavy chain, and only V and J segments in the light chain gene. High diversity of the BCRs is in a first step achieved by V(D)J recombination, a process in which RAG enzymes (recombination activating genes) join together

different V, D, and J segments encoded in the Immunoglobulin gene. In the progenitor B-cells the VDJ segments of the heavy chain are joined which leads to the expression of a pro-BCR in the pro B-cell stage. Further the light chain V-J recombination takes place and the assembly and expression of a complete BCR (IgM) marks the immature B-cell stage. These cells then leave the bone marrow to home the secondary lymphatic organs, the lymph nodes residing throughout the body and the spleen. In these secondary organs further diversity of the BCR is achieved by class switch recombination and somatic hyper mutation during the germinal center reaction. (Janeway et al., 2001)

Binding of antigens to their BCR activates B-cells leading to internalization and further processing of the antigen. MHC class II molecules are loaded with the antigens and are presented on the surface of the B-cell. Helper T-cells in turn bind to these antigen-receptor complexes, resulting in upregulation of costimulatory signals like cytokines and co-receptors like Cd40. These stimuli induce on the one hand proliferation to expand the B-cell pool recognizing this antigen and on the other hand induce differentiation into antibody producing plasma cells. (Janeway et al., 2001)

**Myc and hematopoiesis**

Myc is highly involved in hematopoiesis as knock out mice die by day 10.5 because of lacking fetal liver hematopoiesis (Davis et al., 1993). During development, Myc expression changes along with differentiation: long-term hematopoietic stem cells show lower Myc expression than MPP, which correlates with proliferation of these cells (Delgado and León, 2010). Also in B-cells Myc expression correlates with developmental stages in which the B-cell pool is strongly expanded and cells are proliferating. This occurs once during the pro-pre B-cell stage when the pre-B-cell

receptor (BCR) is interacting with stimulated IL7R and later in the mature B-cells when the BCR is activated by antigens to undergo a germinal center reaction (Delgado and León, 2010). The BCR signals through a Ras-MAP-kinase cascade leading to expression of Myc and Bcl-2, which induce proliferation and survival of this population (Gold, 2008). B-cell specific conditional *myc* knockout mice show reduced numbers of early and immature B-cells, which reflects a higher rate of apoptosis. In these mice Myc is essential for early B-cell development inducing a transcriptional program of B-cell identity by the transcription factors EBF-1 and Pax-5 (Vallespinós et al., 2011). Moreover, Myc is important for upregulating Bcl-6, which induces the germinal center reaction and is also a negative regulator of *myc* (Ott et al., 2013).

**Myc involvement in hematologic disease**

Shortly after finding that a human homolog of the v-*myc* exists, *myc* was found to be the driving oncogene in Burkitt lymphoma (BL)(Dalla-Favera et al., 1982). Most of these patients were bearing the translocation t(8,14)(q24;q32) of the *myc* gene to the Eμ enhancer of the *IgH* gene, although translocations could also occur with the λ and κ light chains. Such translocations, also occur in diffuse and large B-cell lymphoma (DLBCL), unclassifiable B-cell lymphoma (BCLU), a tumor in an intermediate stage between DLBCL and BL, the leukemic variant of Burkitt lymphoma, plasma cell myeloma and acute lymphatic leukemia (ALL), in T-ALL *myc* gets translocated to the *TCR* $\alpha$ and $\beta$ locus (Klein G, BioEssays 2000). Moreover, Myc is deregulated in follicular lymphoma, chronic lymphoid leukemia (CLL), acute and chronic myeloblastic by amplification, copy number variations and overexpression due to deregulated upstream signaling (Delgado and León, 2010).

## The Sleeping Beauty System

### Insertional mutagenesis, drivers and passengers: Cancer gene discovery

The background of insertional mutagenesis screens began in the 1930s when spontaneous tumor formation in mouse models was related to retroviral insertions. The underlying key was, that not the genetic material of the provirus per se, but the deregulation of the endogenous genes in which the provirus inserted, led to tumor formation. Only when a provirus was integrating upstream of the coding sequence of a gene that was important for tumor establishment, the cell obtained a proliferative or growth advantage, would transform and establish a tumor. Cancer is a genetic disease that implies a multi-step process in which oncogenes are activated and tumor suppressors are inactivated leading to transformation. Several cooperating somatic mutations in multiple independent genes arise, that cooperate to induce a malignant state of the cell. These essential mutations are also called driver mutations and are distinguished from a multitude of mutation that are present in a tumor that is resected from patients or animal models, so called passenger mutations. Thus the use of insertional mutagenesis empowers the cell to acquire a set of mutations that, if they induce an advantage, lead to tumor formation. A disadvantage of this retroviral system is the cellular tropism of such slow-transforming retroviruses, leading to a limited discovery of cancer causing genes in mouse models. This made different systems of insertional mutagenesis a strong focus to discover new cancer causing mutations in mouse models of human disease (Copeland and Jenkins, 2010; Koudijs et al., 2011; Mattison et al., 2009; Uren and Berns, 2009; Uren et al., 2005; Vassiliou et al., 2010).

The advances in sequencing technology that occurred in the last decade made it feasible to screen vast amounts of tumors from humans or mouse models (encode, Cosmic database, TGCA database). This led to the discovery of the high number and high complexity of mutations present in the tumor genomes. In order to deal with this load of information, and integrate it in biological terms for the use to discover unknown mechanisms in tumor development, computational biology strongly developed. The tools and algorithms that have been developed to distinguish the true oncogenes and tumor suppressors from the vast mass of mutated genes are therefore an essential part in the search of new cancer genes and are continuously evolving (Koudijs et al., 2011; de Ridder et al., 2006; Uren et al., 2005).

The main aim of constructing the SB mouse models was to develop an alternative system of mutational insertion screens in mouse models of human cancers. While virus based strategies, which were previously used for such purposes, could only be used for some cancer types, the transposon system was adapted and showed to cause mutations in virtually all tissues of the mouse (Collier et al., 2005; Dupuy et al., 2005). Moreover, while the insertions of the retrovirus were shown to have a preference for transcribed regions, the SB chooses relatively randomly from a range of TA dinucleotides. This leads to $20x10^6$ possible insertion sites in the mouse genome (Liu et al., 2005).

The main feature of the SB system is the two-component system, where the transposase is separated physically from the transposable element and acts in trans. In this way, transposition can be regulated, inducing mobilization of the transposon in a temporal or conditional manner. The detection of the targeted genes is relatively easy, as the transposons themselves mark these sites. Moreover the PCR methods to identify the genomic structures flanking the transposon, ligation-mediated PCR (LM-PCR) could already be adapted from retroviral screens, as both methods “tag” the insertion site. A disadvantage of the SB system is that the

reintegration of excised transposons highly favor nearby site, referred to as local hopping. Thus the mutagenized structures surrounding the donor site of the transposon, has to be excluded from integration site analysis.

In case of highly efficient transposition though, the transposase could mobilize a transposon twice, leaving only the 3bp transposon footprint behind, which can lead to non-identified integration sites.

### History and basics of transposons

First discovered and described by Barbara McClintock in zea maize as jumping genes in the 1930s and honored with a Nobel Prize for Physiology or Medicine in 1983, transposable elements moved their way into modern genetics.

Transposable elements, or transposons are patches of DNA that are flanked by a recognition sequence that can be placed from one genetic location in the genome to another. In essentially all organisms transposons were identified and form a large part of the genome, ranging from around 10 percent in C. elegans to around 40 -50 in mice and men and in some plants up to 80 percent (Muñoz-López and García-Pérez, 2010).

There are two classes of transposons that use diverse mechanisms to reintegrate into the genome, class I RNA-based retrotransposons and class II DNA transposons. RNA-retrotransposons are transcribed from the genomic location through a RNA intermediate. Then, they are reinserted by the transposon-encoded retrotranscriptase. This class of transposons adds an additional copy of the transposable element to the genome in a diverse location and leads, from an evolutionary point of view, to the formation of gene duplications. In contrast, DNA transposons use a cut and paste mechanisms and thus do not duplicate, but just change the insertion site.

DNA transposons in invertebrates serve already for a long time as tools of genetic engineering, including transgenesis and insertional mutagenesis (Dupuy, 2010). Good examples are the Drosophila P-elements, which have been used in various forward genetic screens as a tool to find genetic associations. In vertebrates, the endogenous DNA-transposable elements LINE, SINE and Alu repeats are inactive, thus excluding these elements as a tool for mammalian genetics. Non-vertebrate transposable elements showed to be able to transpose in vertebrates, but the efficiency was low. Therefore, a Tc1/mariner-type transposon was synthesized by reconstructing inactivating mutated sites of salmoid fish transposase gene. The modified transposon was called Sleeping Beauty (SB), as it was „kissed" awake by scientists after a long time being „asleep". It was shown to be valid as a genetic tool in vertebrates as its specific activity was shown in mouse, human and fish cells (Ivics et al., 1997, 2004).

**The SB transposon system: structure and function**

The SB transposase contains a N-terminal DNA-binding motif, with a nuclear localization signal, a central glycin rich motif and the DDE domain consisting of two aspartic acids (D) followed by a glutamic acid (E) in the catalytic pocket of the C-terminus, which performs DNA cleavage and joining reactions (Ivics et al., 1997). The most important structures of the transposon are the 200-250bp inverted repeats (IR) which contain two 15-20bp direct repeats (DR) (Ivics et al., 2004). The IR/DR structures are the recognition site for the transposase and define highly specific binding. Moreover, the sequence of the IR/DR left and right of the exogenous transposable DNA differs in its base pair composition leading the specific recognition of SBase to only the SB transposon and not other transposon terminal repeats (Ivics et al., 1997). Moreover, this feature enables the specific amplification of the genomic region upstream and downstream of the insertion

site. In mammalian cells, for binding of two heterodimeric SB transposase monomers to each IR/DR, the cellular host cofactor HMGB1 is recruited to bend DNA (Zayed et al., 2003). A synaptic complex is formed in which the inverted repeats are paired and held together by the SBase. The catalytic domain then introduces a DNA double strand break and 3' overhangs at the donor sites. These leave behind a 3 basepair footprint, if repaired by the host non-homologous end joining repair mechanism or, lead to transposon duplication if homologous recombination is used for repair. Recently it was shown that the NHEJ pathway is the preferential pathway to repair the double-strand breaks introduced (Izsvák et al., 2004; Yant and Kay, 2003).

Reintegration of the excised transposon happens at TA dinucleotides, which serve as the recognition site for the Tc-1/mariner like transposons like SBase. In contrast to virus-based integration vectors, the SB does not favor transcriptional active sites, but rather an angled conformation of DNA as preferential target site for integration (Liu et al., 2005; Yant et al., 2005). At the insertion site, the missing nucleotides are filled up by host factors of the NHEJ and HR DNA repair pathway similar to VDJ recombination mechanisms and generate target site duplication (Muñoz-López and García-Pérez, 2010).

## Transposons

The initial mouse models of the transposable element, called T2Onc were established by the work group of Largaespada and is depicted in Figure 2 (Collier et al., 2005).


**Figure 2. T2Onc transposon. IR/DR are depicted by arrows, splice acceptors (SA) followed by polyadenylation signals (pA) result in premature termination of transcription, whereas the mouse stem cell virus 5' long terminal repeat (MSCV 5' LTR) followed by a splice donor sites (SD) leads to enhanced expression of downstream exons (Collier et al., 2005).**

The T2Onc harbors genetic elements between the IR/DR that deregulate genes depending on where and how they are inserting. Bidirectional splice acceptors (one from exon 2 of the mouse *engrailed-2* gene and the other from the carp *β-actin* gene) followed by polyadenylation signals (bi-directional SV40 poly(A)) lead to the disruption of the full-length transcript when inserted in either orientation (5' -3' or 3'-5') into a gene. A 5' long terminal repeat from the murine stem cell virus (MSCV) followed by a splice donor (splice donor is from exon 1 of the mouse Foxf2) leads to overexpression of the downstream exons if the transposon integrates in transcriptional direction. Two mouse lines carrying a transgene containing around 25 copies of the T2Onc in a concatameric structure on either chromosome 15 or chromosome 1 were established.

This system was ameliorated by Adam Dupuy in 2005 (Dupuy et al., 2005) who produced three other mouse lines Tg6057, Tg6070 and Tg6113. They carried a high copy transgene containing up to 148, 214 or 358 copies of the transposon, respectively. As methylation related silencing at the MSCV 5'LTR reduced the number of activating mutations that can be obtained during a screen, the above mentioned mouse lines were estimated to contain 30, 36 and 77 non-methylated transposons, respectively. T2Onc2, as the new transposon was called, was optimized for the transposase binding sites, reduced in size to around 2kb and the engrailed-2 splice acceptor was enlarged. By introducing more transposable elements within the concatameric transgene, a higher mutagenesis rate was obtained.

Viral promoters have a tissue tropism, like the MSCV 5'LTR of the T2Onc and T2Onc2 transposons shows a preference for the hematopoetic lineage (Howell, 2012; Uren et al., 2005). Thus a composite promoter called CAG which contains the β-actin promoter fused to the CMV immediate early enhancer was introduced in the T2Onc3 version. Two lines were established carrying 11 and 28 copies of

T2Onc3 and they were found to reduce the preference for hematopoietic tumor formation (Dupuy et al., 2009).

**Transposase**

The second component of the SB mutagenesis system also evolved over the past decade. The initial mouse line was carrying as transgene the SB10 transposase driven by the CAG promoter (CAG-GS-SB). In combination with the T2Onc transposition occurred in somatic cells in mainly all tissue types. The amount of integration was however not enough to induce tumor formation in up to one year aged mice (Collier et al., 2005). The same was true for the breeding of T2Onc2 high copy transposon line with the CAG-GS-SB lines (Collier et al., 2009). Low transposase expression probably related to epigenetic silencing was suggested to be one of the reasons for the failure of tumor formation. Therefore, the promoter of the SB was changed to the Rosa26 promoter, which is ubiquitously expressed in the mouse. The established mouse line, R26-SB11 carried a knock-in on chromosome 6 of an optimized transposase, in which four amino acids were changed to increase its activity. Breeding with the high copy transgene T2Onc2 mouse induced up to 20% embryonal lethality and the remaining mice all succumbed to tumors by the age of 17 weeks. The main cancer types were lymphoma and leukemia, but also astrocytomas were observed (Dupuy et al., 2005). When a low copy T2Onc was bred with the R26-SB11, however, no embryonic lethality, but lymphomas, leukemias and low penetrance gliomas were observed at an average onset of 6 month. This underlines the flexibility of the system and also the high mutagenic rate of the high copy system with the R26-SB11 transposase. To overcome the developmental related lethality, a conditional version of the R26-SB11 was established, taking advantage of the Lox-Cre system.

### The Cre-LoxP system

The Lox-Cre system is in ways similar to the transposase system. Like in the transposon system two components are being used: An enzyme, Cre recombinase, acting in trans and a pair of repeats called LoxP sites, that flank a genetic region of interest. The Cre recombinase excises the LoxP sites and thus deletes the included DNA sequence from the locus. As Cre-recombinase can be knocked into any tissue specific promoter, this system serves as optimal tool for tissue specific exclusion of genetic elements flanked by the LoxP sites. Moreover, due to interest in turning on genes in a conditional manner, the lox-stop-lox Cre construct was developed. A reporter gene followed by several polyadenylation sites that lead to transcriptional termination, is placed in between the two loxP sites downstream of the promoter and upstream of the gene that should be turned on (Srinivas et al., 2001). In the tissue that expresses the Cre under the specific promoter, the reporter gene is floxed out together with the stop signal, and instead the gene of interest is expressed. In the case of the R26-SB11 system, a loxP-EGFP-stop-LoxP cassette was introduced between the R26 promoter and the SB11 trasposase gene (Dupuy et al., 2009).

### Studies of conditional SB screens

An initial experiment showed -contrary to expectations- no embryonic lethality in R26-lsl-SB11/T2Onc2 mice crossed to β-actin driven Cre mice, which induce expression of SBase in virtually all tissue types. A slight reduction in the SBase expression levels was observed which was suggested to be the reason for the reduced lethality, which would fit with the study of the CAG-GS-SB/T2Onc2 mice. Also there expression levels are lower resulting in no embryonic lethality. Although the embryonic lethality dropped down, the tumor development was comparable to the one seen in R26-SB11/T2Onc2 mice, and hematopoietic tumors,

mainly T-cell lymphomas developed by 8 weeks (Dupuy et al., 2009). When crossed to AID Cre, which is expressed in germinal center B-cells, R26-lsl-SB11/T2Onc2 mice developed tumors, mainly of B-cell origin, with an average latency of 46 weeks and a penetrance of 90% (Dupuy et al., 2009).

A big step in cancer gene discovery was taken, when the first two papers showing specific human solid cancer models in mouse were shown to be screenable: Albumin-Cre/T2Onc/R26-lsl-SB11 mice were shown to develop hepatocellular carcinoma by 15 month with lung metastasis and a male susceptibility that is also observed in human HCC (Keng et al., 2009). When combined with a Villin-Cre, many intestine specific types of tumors were found between 15 and 18 month (Starr et al., 2009). Since then the CRE-system combined with the SB system successfully discovered novel genes involved in the development of prostate cancer (Pdx1-Cre) (Rahrmann et al., 2009), neuronal (Nestin-Cre) (Koso et al., 2012; Mumert et al., 2012) and hematopoietic tumors (Vav-, Lsk-, Cd4-Cre) (Berquam-Vrieze et al., 2011), to mention some.

Very often such screens also take advantage of known mutations that occur frequently in certain types of tumor. For example, mice that are null for the tumor suppressor ARF were crossed to mice carrying the Sleeping Beauty system. This resulted in very fast onset tumorigenesis with mainly sarcomas arising that harbored truncated versions of the catalytic domain of BRAF in 80% of the cases (Collier et al., 2005).

Further the combination of mouse models with known predisposing mutations or translocations were combined in a manifold way to elucidate the underlying signaling pathways and cooperation of mutations that lead to tumor formation. Examples of studies include mutated Kras for pancreatic adenocarcinomas (Mann et al., 2012), mutated BRAF for melanoma (Karreth et al., 2011), APCmin for colorectal tumors (March et al., 2011; Starr et al., 2011), overexpressed Myc for

hepatocellular carcinoma (O'Donnell et al., 2012), mutated Janus kinase in myeloproliferative disease (Tang et al., 2013), genes that collaborate with loss of Rassf1a (van der Weyden et al., 2012a) and with loss of Cadm1 (van der Weyden et al., 2012b) in tumorigenesis. The aim of all of these studies was to find candidate cancer genes and distinguish them from passenger mutations that are non transforming events, co-evolving with the driver mutation. Statistical methods including Kernel deconvolution and Monte Carlo simulation define common integration sites (CIS) that are integrations that are represented in a higher than randomly expected manner in the Sleeping Beauty evolved tumors. The genes found in the previous mentioned publications were also relevant in human disease and could unravel new findings, making the Sleeping Beauty system an attractive means of deciphering molecular mechanisms underlying different types of cancer.

# MATERIALS AND METHODS

## Mice.

E*μ-myc* transgenic mice (Myc) (Adams et al., 1985) were crossed with mice bearing the *CD19-Cre* knock-in (Cre) (Rickert et al., 1997) to maintain one transgene and Cre homozygous in the F1 animals. In parallel, *R26-lsl-SB11* (SB)*; Tg6113* (T) (Dupuy et al., 2009) obtained from Stefano Casola were intercrossed to keep SB in homozygosity and retain one transgene in the F1 animals. F1 animals were intercrossed to obtain F2 generations carrying all 4 genetic elements of the experimental mouse. All mice were on an isogenic C57BL/6J background. Mice were monitored weekly for lymphoma onset by lymph node palpation and sacrificed when moribund. Donor mice for adoptive transfer were sacrificed at maximum 56 days after birth, to reduce potential selection and outgrowth of tumor clones.

Recipient mice for adoptive hematopoietic transfer were purchased at Harlan Laboratories and transplanted at an age between 6 and 10 weeks. Transgene and knock-in animals were genotyped by PCR, as described previously for E*μ-myc* (Gorrini et al., 2007), *CD19-Cre* (Rickert et al., 1997), *R26-lsl-SB11* and *T2Onc2* (Dupuy et al., 2009).

## Adoptive transplant model.

Bone marrow was extracted by crushing the femurs, tibias and hips from the aforementioned donor mice. By magnetic bead sorting lineage committed cells and Cd19 positive cells sequentially were depleted from whole bone marrow following the user instructions of the MACS bead kits Lineage Cell Depletion Kit and Cd19 MicroBeads (Miltenyi Biotec). Between 10000 and 200000 Lin- Cd19- hematopoietic stem and progenitor cells (HSPC) were used to reconstitute sub-lethally irradiated (7Gy) syngeneic C57BL/6 females (6-10 weeks) with addition of 300000 C57BL/6 splenic cells. Neomycin was added to the drinking water of irradiated, transplanted animals for 1 week. Reconstitution was monitored by analyzing GFP (present in the R26-lsl-SB11 construct) positivity by flow cytometry in peripheral blood of a sample taken from mice at different stages after transplantation, as indicated. Mice were monitored weekly for lymphoma onset by lymph node palpation and sacrificed when moribund.

## Tumor analysis.

Genomic DNA was extracted from lymphoma cells using standard procedures, quantified with picogreen (Quant-It, Invitrogen) to be used for ligation-mediated PCR (LM-PCR), in detail below), VDJ-PCR, excision PCR and re-genotyping. For LM-PCR, primers were used as previously described (Berquam-Vrieze et al., 2011; Brett et al., 2011), VDJ-PCR primers annealing in the CDR3 and in the JH4 of the *IgH* gene were used as previously described (Yu and Thomas-Tikhonenko, 2002). This gives rise to 4 main amplicons depending on which *JH* segment was used for *IgH* rearrangement. Excision PCR was performed as described previously (Collier et al., 2005) with the forward primer of the *R26* gene and the reverse primer annealing to the *SB11* gene (SBR2) and gives rise to a 800 basepair product after Cre-mediated excision of the Lox-stop-Lox cassette. Hematoxylin and Eosin (HE) stainings were performed on formalin fixed tissue and paraffin embedded tissue sections according to standard procedures. 50$\mu$l of peripheral blood samples were analyzed with the automatic Coulter Ac hematological analyzer (Beckman).

## Flow Cytometry.

Single cell suspension was prepared from tissues and infiltrated lymphnodes obtained from mice. Alternatively peripheral blood mononuclear cells (PBMCs) were derived from the bleeding of animals. Flow cytometry was performed either immediately or in a retrospective manner by freezing samples in 10% DMSO supplemented FBS. In short, single cell suspensions were washed with PBS, blocked in 1% BSA solution and antibody-incubation was performed for at least 30 minutes at 4°C. Antibodies used include anti-mouse IgM APC (eBioscience) and PE-Cy7 Rat anti-mouse CD45R/B220 (BD), anti-mouse CD25 PerCP-Cyanine5.5 (eBioscience), anti-mouse CD117 (c-Kit) PE (eBioscience), anti-mouse Cd3ε PerCP-Cyanine5.5 (eBioscience) and anti-mouse Ly-6G (Gr-1) PE (eBioscience) and were used dilutions of 1:200-1:400. Dead cell exclusion was performed with addition of DAPI or PE to stained single cell suspensions. Acquisition was performed on either FACS Calibur (BD) or FACS Canto II (BD). Analysis of data was performed with the BD FACSDiva (BD) and FlowJo (Treestar) software.

## Linker-Mediated PCR.

Protocols were adapted from (Berquam-Vrieze et al., 2011; Brett et al., 2011) as follows: 1 $\mu$g of genomic DNA (gDNA) was digested in a volume of 40$\mu$l followed by adaptor ligation over night using high concentration T4 ligase. To deplete the ligated fragments from excess unligated adaptors that might interfere with later PCR steps,

purification with AMPure XP (Beckman Coulter) was performed following the user instructions. A BamHI digest was performed to destroy concatamer-generated fragments. Primary PCR was performed using primers specific for AluI or NlaII adaptors together with reverse primers annealing either in the left (IRL) or the right (IRR) inverted repeats of the T2Onc2 sequence. As the adaptor is composed of a 3'-C3spacer that blocks annealing of adaptor primers to the template, amplification is only performed from fragments that can be amplified with the transposon specific primer. The nested, second PCR is performed with primers including tags for direct amplicon sequencing on the Illumina platforms and a 8 basepair barcode for sample recognition. PCR amplicons were multiplexed in equimolar concentrations and sequenced using the Illumina Genome Analyzer 2 or the HighSeq2000.

**Sequence analysis.**

**Raw data processing, mapping and defining clonal insertion sites**

The sequencing data from the Illumina HighSeq2000 were analysed as follows:
A typical sequence read is composed as follows:

GGAGAATA-TGTATGTAAACTTCCGACTTCAACTG-TATTACTATTATTTTTAAGAG-GTCCCTTAAGCGGAGCCCTAGA

Meaning
Barcode-Transposon sequence-mappable sequence of varying size-Adaptor sequence

Sequences not containing barcode and transposon sequence were discarded in a first filtering step. Remaining sequences were demultiplexed according to their barcodes. Adaptor and transposon sequences were cropped. Depending on the length of the mappable sequence, the adaptor sequence may be represented by only a few nucleotides or may be missing entirely from the sequence read. Thus, cropping the adaptor sequence involves a step of pattern recognition. We searched for the first eight nucleotides of the adaptor sequence (GTCCCTTA). Whenever this sequence was found, it was cropped off. When the adaptor sequence was represented by less than 8 nucleotides, it was ignored, leading to increased mismatches during the mapping step. Sequence reads shorter than 25 nucleotides after the two cropping steps were discarded. Alignment to the mouse genome reference assembly (NCBI37/mm9) was performed using Bowtie. Additional optional filters were applied to mapped reads: Reads were required to map

uniquely to the genome, to start with a TA dinucleotide (the transposase target recognition site), and to map with either 0, 1, or 2 mismatches.

The mapping data were represented as bedgraphs, where for each genome segment the corresponding number of mapped reads was recorded. Mapping positions for each library were extended by 50 bases upstream and downstream leading to the final IS.

To define clonally expanded insertions, two different filtering strategies were performed in house. The first one was developed by our collaborator Heiko Müller based on the publication from Brett et al. and is named here HM101 (Brett et al., 2011). It is based on the rationale that clonally expanded IS have a higher read count than background insertions. Thus, clonally expanded IS had to be represented by a read count per library of at least 0.1% of the total reads per library and of at least 1% of the maximal number of reads in a library.

> Example: Ct73947_B library IRR AluI has a sum of reads of 245986 and a maximum read count of 97354. Thus the requirement of clonally expanded IS in this library is a minimum of 973.54 reads. All IS that have a readcount higher are by definition clonally expanded, all the others were set to have a value of 0.

The second definition of clonally expanded IS was also developed by Heiko Muller and is named here HM34. This does not take into account the read count of the libraries, but instead, consistent amplification in the 4 libraries of the LM PCR prepared per sample. Clonally expanded IS had to be represented by insertions with a read count of at least 1 in at least 3 out of 4 libraries per sample prepared.

> Example: The IS at the position chr8:68211625-68211808 of sample Ct81496_D has 8 reads in the library IRL AluI, 0 in IRL NlaIII, 30 in IRR AluI and 208 in IRR NlaIII and thus fulfills the criteria of having more than 1 read in at least 3 out of 4 libraries.

The result of these analyses is an annotated excel file Supplementary table 1 in two spread sheets with the annotated IS found by each of the filtering methods.

Additionally to the above analysis, the raw sequencing data were sent to the IAS pipeline that was published as an online tool (Brett et al., 2011). Per sequencing lane, the fastQ files were compiled and uploaded on a FTP server, together with a file containing information of sample ID and the barcode referring to the different libraries prepared. In total 5 fastQ files and 5 sample description files were uploaded. The output files were a .uniq file and .gff3 file for each sample. The .uniq files are single files for each sample containing all insertions identified in either the

IRR or IRL barcodes. The .gff3 files are single files that contain the insertion sites that were considered clonally expanded. Clonally expanded insertion sites have a read count >10% of most frequently seen site in that barcode (IRL/IRR).

> Example: IS Chr8:9921367-9921467 has 955 reads from the libraries amplified with the IRL primer and 263 reads amplified with the IRR primer. The maximal amount of reads in the IRL libraries in this tumor was 3406 and 1499 for the IRL libraries. Thus the %IRL reads at insertion site are 87.215 and 21.808% for IRR. The average percentage of reads, which is in this case 54.511, was calculated by taking the average of the % of reads in IRR and IRL.

The .gff3 files were complied to an excel file and are listed in the third sheet of Supplementary table 1.

Summarizing, the different filtering pipelines were performed as follows:

HM101) Clonally expanded insertion sites have a read count >0.1% of total reads of a library and >1% of reads of the most frequently seen site of a library. Both criteria have to apply.

HM34) Clonally expanded insertions have a read count >1 in at least 3 out of 4 libraries.

BB) Clonally expanded insertion sites have a read count >10% of the most frequently sequenced site in the corresponding IRL or IRR library.

**Overlap of different tumor cohorts**

In order to assess and validate the clonal filtering and the clonal integration sites between different cohorts of tumors, Venn diagrams were produced with the genes that were found hit by the clonally expanded IS (Oliveros, 2007). The percentage of overlap with the donor was calculated for each of the different filtering methods using Fisher exact test comparing the overlap of the clonal samples to the overlap of control samples.

**Statistical evaluation of clonal insertion sites**

Per mouse cohort (control group n=15, lym group n=184, clonal group 64813 n=3, clonal group 73948 n=4, clonal group 82091 n= 3) all the clonal IS that were annotated with a Ref-Seq ID were counted as intragenic, and non-annotated IS were assigned to the inergenic clonal IS. These numbers were normalized to the

total number of clonal IS per cohort and gives the proportion of intergenic to intragenic clonal IS in %. The distribution of clonal IS to intragenic and intergenic regions were compared to the distribution of the proportion of the mouse genome occupied by genes to intergenic regions. As the Sleeping Beauty transposon integrations distribute evenly in the intragenic to intergenic regions of a mouse genome under non-selective conditions, we can monitor the clonal selection as indication of non-random distribution of SB IS in tumors with a probability calculated by Fisher's exact test (Supplementary table 1).

**Statistical evaluation of common insertion sites**

Common Insertion sites are listed in the Supplementary table 3 with a spread sheet for each method, overlap of CIS from lymphoma cohort with CIS found in control cohort are shown in Supplementary table 2.

**Enrichment analysis**

For CIS enrichment analysis (eCIS), the number of clonal IS was counted per gene, including a 1 kb or 10 kb promoter region upstream of the transcription start site. This IS count was normalized to the total number of independent clonal IS per cohort (72338 in lymphomas and 15786 in controls). The enrichment of clonal IS in the lymphoma cohort compared to the control cohort was calculated for all genes with either of the two different promoter regions by dividing the normalized IS in lymphomas (LYM) by the normalized IS in controls. If no IS were found in the control, the IS count was divided by the factor 4.58, which reflects the number of insertion sites found in tumors versus control samples. eCIS were defined to have an enrichment of more than 2.5 and be hit in more than 3 lymphomas. Input files as well as results for genes hit by clonal IS are in Supplementary table 2.

**Monte Carlo and gCIS analysis**

Clonal insertions of the lymphoma cohort and of the control cohort were extracted from the filtering obtained by Benjamin Brett (BB). These files were sent to the IAS online analysis platform (http://ias.eng.uiowa.edu/uploader/). A promoter size of 1 and 10 kb was chosen. The program uses a Chi-squared test statistic to calculate gCIS and Monte Carlo Simulation to define CIS as published (Brett et al., 2011). In short, the Chi-squared test statistic compares the distribution of IS in the TA dinucleotides in a gene-associated region (gene size with a 1kb / 10kb promoter

region) to the distribution of all TA dinucleotides hit by a transposon. For the Monte Carlo simulation, the set of integrations in TA dinucleotides per tumor are randomly simulated to hit all possible TA dinucleotides in the mouse NCBI37/mm9 reference genome. The probability of random simulated integrations to real integrations is calculated to define the CIS. gCIS and CIS were considered as valid if they were not found on chromosome 1 and if they were not present in the control cohort as CIS/gCIS.

**Grouping of clonal IS**

For further characterization of the CIS, the clonal insertion sites were compiled for the control cohort aswell as for the tumor cohort and are listed in Supplementary table 2. The file containing the clonal IS sequenced from lymphomas was uploaded to galaxy, a bioinformatical program available on the server of the institute (Blankenberg et al., 2010; Giardine et al., 2005; Goecks et al., 2010).

The uploaded file was grouped with the following tool:

Join, Subtract and Group/ Group (version 2.0.0) was performed on the gene ID (column 6), performing the operations that included "count distinct" and "concatenate distinct" on the following:

lymphomas harbouring clonal IS in CIS, to know which and how many lymphoma were hit by a gene, defining the frequency of CIS per lymphoma

filtering methods, to know by which method the IS in the gene was obtained

sample ID, to get the number and tumors that harbored a clonal IS in the gene associated with the CIS

identifier, harbouring the information about the genomic location (chromosome:start-end) of the clonal insertion site assigned to a gene associated CIS

**Gene annotation**

For annotation of the CIS, the gene list was uploaded to the Panther Classification System (Mi et al., 2013) and only 9 of the genes could not be annotated. The annotated gene list was uploaded to galaxy.

### GO Terms and Pathways

CIS were analyzed for their overrepresentation with the online tools DAVID (Huang et al., 2009), STRING (Franceschini et al. 2013), PathJam (Glez-Pena et al., 2013) and PANTHER (Mi et al., 2013). Moreover, Ingenuity Pathway analysis tool was used for pathway analysis (Ingenuity Systems Inc.).

### RNA Seq and Chip Seq data

The RNA Seq and Myc ChIP-Seq data used as tracks in the UCSC genome browser and added to the CIS found in this screen were performed by Arianna Sabò in our group (Sabò and Amati, 2014)(Sabò et al. submitted). These data were generated from splenic B-cells of 4 control C57/Bl6 mice (RPKM C_1, RPKM C_3, RPKM C_4, RPKM C_6), splenic B-cells of 4 pretumoral Eμ-*myc* mice (< 8 weeks of age, not showing signs of lymphoma) (RPKM P_2, RPKM P_3, RPKM P_4, RPKM P_5) and on the infiltrated lymphnodes of 3 individual diseased Eμ-*myc* mice (RPKM T_1, RPKM T_2, RPKM T_3). Moreover, the binding of Myc to the promoters of the genes (TSS+1kb/-3kb) was calculated and represented as + when a significant peak was observed in the above cohorts. Control B-cells are labeled with mycCbound, pretumoral samples with mycPbound, and lymphoma samples with mycT1bound, mycT2bound and mycT3bound. This file was also uploaded to galaxy.

## Orientation of the gene

We obtained a) the strand information regarding the orientation of the T2Onc2 of every IS regarding the primer used to amplify the fragment. The IRL primer is complementary to the left inverted repeat in the – strand, whereas the IRR primer is complementary to the right inverted repeat in the +strand of the transposon and the MSCV is acting as a promoter activating locations downstream of the IRR. Thus the direction of the MSCV LTR is in sense when the product of the IRL maps to the – strand and the product of the IRR maps to the + strand. Due to PCR bias and as a result of mismapping, not all of the insertions give a clear indication about strand location and were assigned as N/D (not defined). In a second step the strand information of the genes that were hit by the IS were added. Combining the data, genes were defined to be hit in SENSE when the MSCV and gene are in the same orientation, ANTISENSE when orientation is opposite and N/D for regions that are not gene-centric or when a N/D (not defined) was assigned for the transposon. Genes that have less than 26% of IS in sense orientation are suggested to have putative downregulation as the probability of enhanced transcription due to the

sense oriented MSCV is low. These 34 genes serve as putative targets for shRNA mediated *in vivo* knock-down experiments. The remaining 299 genes have to be asessed for putative knockdown or overexpression depending on the location of the ISs: Putative upregulation of a full-length transcript occurs when the IS integrate in sense upstream of the coding sequence whereas sense insertions into the coding sequence can lead to overexpression of a truncated version of the protein or to premature termination of the transcript due to the polyadenylation signals present in the T2On2, which act in both sense and antisense orientation. Therefore, genes that have more than 25% of the IS in sense orientation have to be analyzed one by one with regard to location of IS, regarding exons, introns and coding sequence, the protein domains that would putatively be upregulated by the IS, biological functions of the protein and if already known, their function in tumor context.

For these informations the clonal IS were color coded and uploaded to the UCSC genome browser. Green colored IS are antisense oriented, red IS are sense oriented and blue IS are N/D. In the UCSC genome browser, I obtained the information about which exons are disrupted, in which introns the IS fall and also, if the IS were pointing towards putative upregulation, whether in the following exons putative alternative start codons are located. A good indication about disruption of the gene was given, when the IS were spread across the gene body. On the other hand, IS that cluster around the same site seem to delete a certain exon, prematurely terminate transcription leading to a short isoform, and/or overexpress the downstream transcript. To see which are the domains that are putatively up or downregulated, the NCBI nucleotide graphic information was obtained, that shows together with exons present in a gene, also the functional annotated domains and also the protein coding sequence. Moreover, on this page, the functional information about the genes were observed. Moreover, from the MGI webpage I obtained information about the knock-down function of these genes if knock-out mice were developed already. Moreover, I obtained further literature form this site. The obtained information were collected in the form of a PowerPoint presentation, indicating for each CIS a screenshot of the whole gene, if necessary a zoom in of the region of the gene that was hit by the IS; a screenshot of the protein domain information from NCBI indicating the site of IS (if it is in intronic sequences a line was drawn in between exons, otherwise the whole region that was hit is marked by a colored square), informations regarding mouse models

that were established for the gene of interest if available and further publications in context with tumorigenesis or Myc.

**Comparison to databases**

For the cross comparison of the mouse CIS to human databases and gene lists, a gene ID conversion was performed for every gene set with the Ingenuity Pathways Analysis system (thereafter IPA), to assure consistency of the annotations used. Genes found in the different databases were overlapped and Venn diagrams were produced with an online tool (Oliveros, 2007). The overlap of the CIS gene set with a database compared to the human protein coding genes (Harrow et al., 2012) was used to calculate p-values by using a binomial distribution test. CIS from the Retrovirus and Transposon tagged Cancer Gene Database (RTCGD) were obtained from (Akagi et al., 2004). The tumor types included are solid tumors and 12 hematological tumor types. For the Myc_SB enriched dataset CIS from the mouse models RosaSB;T2Onc2 (Dupuy et al., 2005), p19(Arf-/-);CAGGSB10,T2/Onc (Collier and Largaespada, 2007), Eu-*myc* (van Lohuizen et al., 1991), Eu-*myc*;Pim1(-);Pim2(-) (Mikkers et al., 2002) and CD2-MYC/Runx2 (Blyth et al., 2001) were obtained.

Moreover, curated mutations of the COSMIC cancer genes (Futreal et al., 2004), published mutation by sequencing human Burkitt lymphoma samples(Giulino-Roth et al., 2012; Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012), mutations defined to be common in humans to drive tumorigenesis (Vogelstein et al., 2013) and genes that defining known oncogenes (OG) and tumor suppressor genes (TSG) (Davoli et al., 2013) were obtained to define different genesets.

Mutations occurring in the following human tumor types were obtained from the TGCA database at http://cancergenome.nih.gov/ and the number of somatic mutations per gene were listed per tumor:

Burkitt Lymphoma
Diffuse large B-cell lymphoma
Follicular Lymphoma
Bladder urothelial carcinoma
Breast invasive carcinoma
Cervical squamous cell carcinoma and endocervical adenocarcinoma
Colon adenocarcinoma
Glioblastoma multiforme
Head and Neck squamous cell carcinoma
Kidney renal clear cell carcinoma
Kidney renal papillary cell carcinoma
Acute myeloid leukemia
Lower grade glioma
Lung adenocarcinoma
Lung squamous cell carcinoma
Serous cystadenocarcinoma
Pancreatic adenocarcinoma
Prostate adenocarcinoma
Rectum adenocarcinoma
Skin cutaneous melanoma
Thyroid carcinoma
Uterine corpus endometrioid carcinoma

All the above generated information concerning the independent CIS were joined to provide a final table of genes hit in the Eμ-*myc* mouse:

Join, Subtract and Group/ Join two Datasets (version 2.0.2), by using the mouse gene ID as common field to join the datasets. The result is the excel file Supplementary table 3 which allows easy and fast accession of information collected on the CIS in order to chose targets for validation based on communalities assigned to them, for example CIS that belong to the same GO ontology.

# AIM

Myc triggers a transcriptional program inducing hyper replication and proliferation but also tumor suppressive mechanisms like apoptosis when deregulated. Therefore, Myc dependent tumors display high selective pressure to accumulate secondary mutations blocking these tumor-suppressive pathways (Hoffman and Liebermann, 2008). The Eμ-*myc* mouse is very well characterized and therefore serves as a good model system in which deregulation of Myc-cooperating tumor suppressive pathways can be studied (Eischen et al., 1999; Hemann et al., 2005; Schmitt et al., 1999, 2002; Sherr, 2001; Zindy et al., 1998). The Sleeping Beauty insertional mutagenesis system (SB) is an attractive means to decipher molecular mechanisms underlying different types of cancers also relevant in human disease (Howell, 2012). Therefore, we apply a SB screen in Eμ-*myc* mice to find rare events abrogating tumor suppressive functions needed for the development of cancer in the Eμ-*myc* model.

# RESULTS

## Establishment of the mouse model and basic concepts of the SB screen

### Breeding strategy and mice

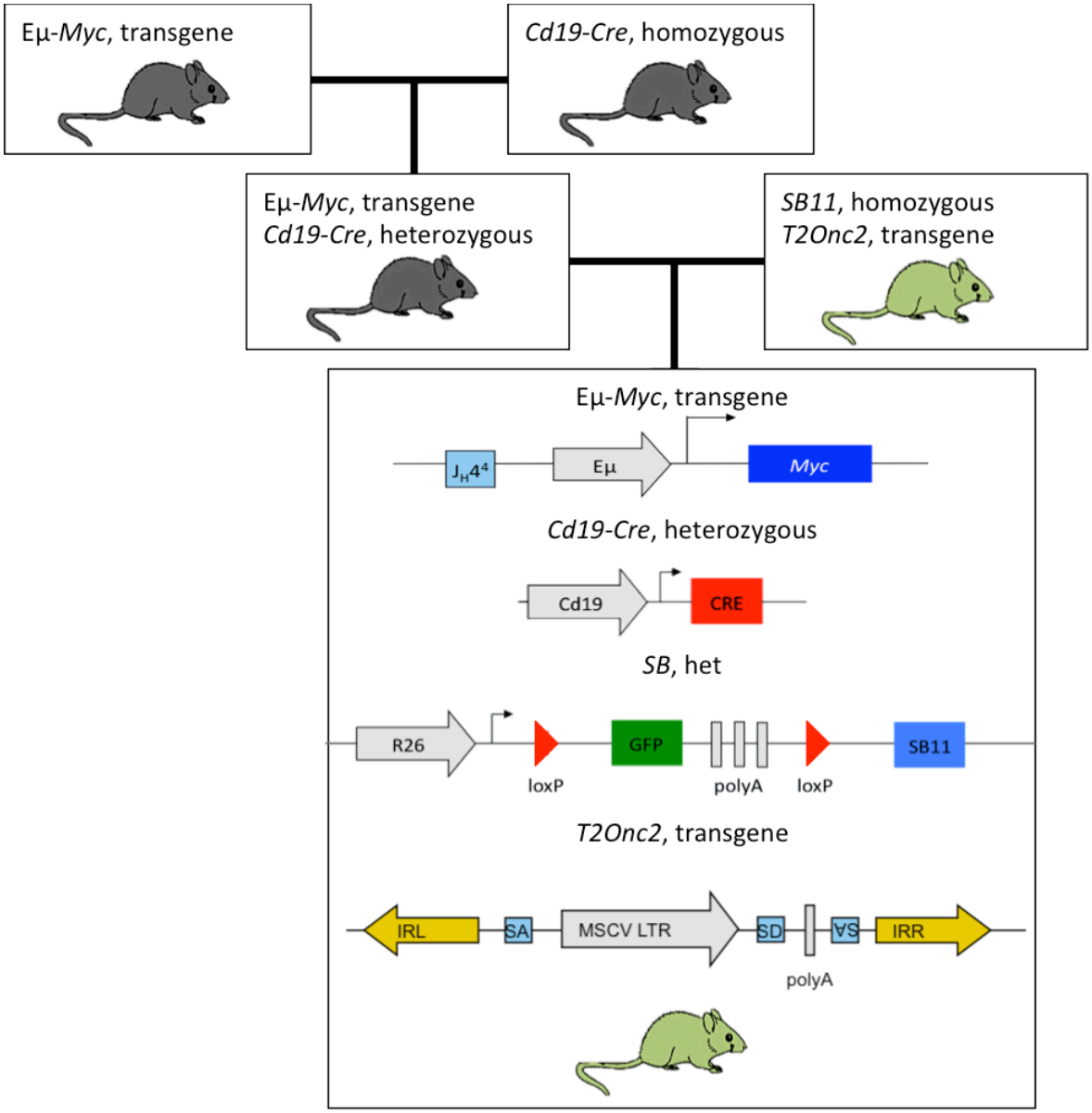


**Figure 3 Breeding strategy.** Eμ-*myc*/Cd19Cre mice were crossed with SB11/T2Onc2 mice to obtain animals harboring all 4 genetic elements (SB11, T2Onc2, Cd19-cre and Eμ-*myc*) needed for this insertional mutagenesis screen. Abbreviations: Jh44: joining segment 4 of the Immunoglobulin heavy chain. Eμ: immunoglobulin heavy chain enhancer μ, CRE: Cre recombinase, R26: Rosa 26, SB11: Sleeping Beauty Trasposase 11, IRL: IR/DR left, IRR: IR/DR right, SA: splice acceptors, MSCV LTR: murine stem cell virus 5' long terminal repeat, SD: splice donor.

To perform a forward genetic screen to find cooperating partners of Myc we chose the Eμ-*myc* mouse model. It carries a construct of the *myc* oncogene under the control of the Eμ heavy chain enhancer (*IgH-myc*) which is the translocation found to be responsible for some B-cell lymphomas in humans (Delgado and León, 2010).

The transgene leads to strong overexpression of Myc in the B-cell lineage (Langdon et al., 1986).

In our screen we aimed to target the insertional mutagenesis selectively to the B-cell compartment, where Myc is overexpressed from the Eμ-*myc* transgene, for which we took advantage of the *Cd19-Cre* mouse model. This mouse harbors a targeted knock-in of the Cre-recombinase in the *Cd19* gene, leading to Cre expression from the pro-B-cell stage, and can thus be used to delete floxed targets in the whole B-cell lineage (Rickert et al., 1997). Being null for the *CD19* gene, homozygous *Cd19-Cre* mice are normal except for a deficiency of the B-1 subset. As the Eμ-*myc* mouse was bred with the homozygous *Cd19-Cre* mouse (**Figure 3**), offspring carry one functional allele of the *Cd19* gene, which was shown to be enough to provide normal B-cell development (Rickert et al., 1997).

The strain carrying both components of the conditional Sleeping Beauty system (Dupuy et al., 2005) was obtained from the group of Stefano Casola. It harbors the T2Onc2 transgene on chromosome 1 and this transgene contains a cassette with 358 copies of the T2Onc2 transposable element, which is mobilized upon expression of the Sleeping Beauty transposase. Additionally, this mouse carries a conditional SB-coding gene (*SB11*) that is expressed under the control of the ubiquitous *Rosa26* promoter. Insertion of a loxP-flanked stop signal and an *EGFP* sequence between the promoter and *SB11* renders SB expression strictly dependent upon expression of the Cre recombinase, expressed from an independent transgene. Hence, animals are exposed to insertional mutagenesis exclusively in tissues expressing a Cre transgene.

By breeding Eμ-*myc* (Myc)/*Cd19*-Cre (Cre) mice to the strain carrying the conditional transposon system (R26lslSB11 (SB)/Tg6113 (T) we obtained animals harboring all 4 genetic elements further referred to as experimental mice (**Figure 3**). Additionally this breeding gave rise to control mice lacking one of these genetic

elements: Myc-Cre-SB, Myc-SB-T and Cre-SB-T. 24 of the 181 mice from the final breeding contributed to the experimental cohort, thus fulfilling the expected Mendelian frequency of 1:8. Of the control cohorts, 15 Cre-SB-T, 27 Myc-Cre-SB mice and 18 Myc-SB-T mice were born.

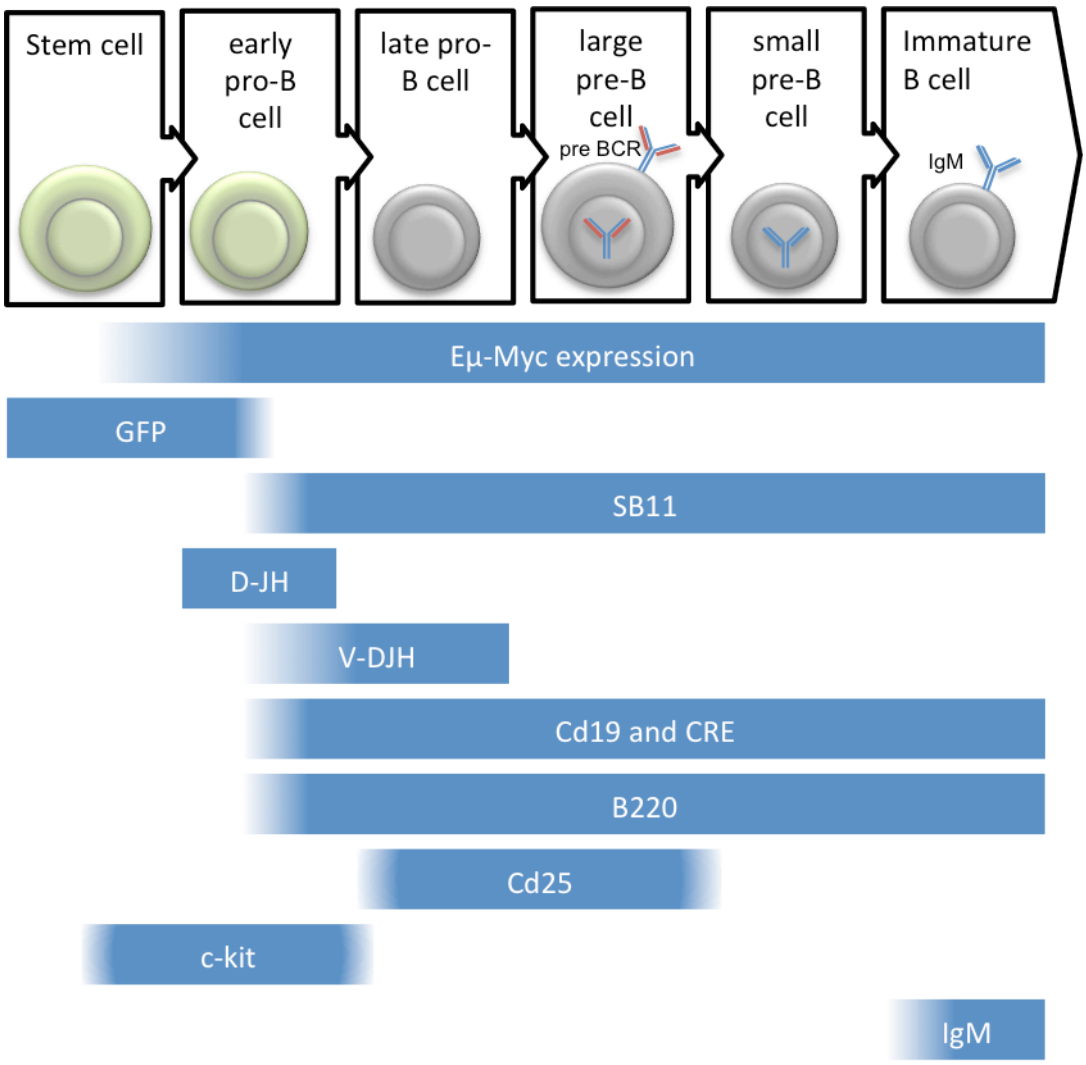


**Figure 4 B-cell development and associated changes in the experimental mouse.** A schematic presentation of the developmental stages in the Eμ-*myc*-Cre-SB-T mouse, showing the overlap of stage specific markers and induction of the Sleeping Beauty mediated insertional mutagenesis, Adapted from (Janeway et al., 2001)

Figure 4 shows the different stages in B-cell development and the associated changes in expression of the different components used to monitor our mouse model: Overexpression of Myc driven by the IgH Eμ enhancer occurs only in the B-cell lineage from an early progenitor stage onwards (Adams et al., 1985; Harris et al., 1988), as also indicated by immunophenotyping experiments of a mouse that

harbors GFP as a reporter gene cloned downstream of the Eμ enhancer (Guglielmi et al., 2003). Myc induces increased proliferation, coupled to induction of programmed cell death (Adams et al., 1985; Eischen et al., 1999), leading to an expansion of the B-cell progenitor compartment, which can be monitored as a B220 low, IgM negative population in flow cytometry. The Cre-recombinase deletes the lox-stop-lox cassette with the GFP reporter from an early B-cell progenitor stage on leading to Sleeping Beauty transposase expression. This activates the mobilization of transposons and thus insertional mutagenesis coinciding with the expansion of the B-cell compartment.

**Basic concepts of Sleeping Beauty screens**

The basic concepts of the SB screen are summarized in Figure 5. SB transposase activates transposon mobilization in B-cells overexpressing Myc. Cells that obtain a selective advantage by transposon insertion perform clonal expansion, which leads to tumor formation. As mutagenesis is an ongoing process, initial transforming mutations along with passenger mutations and newly acquired mutations, that do not contribute to tumorigenesis are found in the obtained tumors. Thus it is essential to apply a filtering system in which clonal derived insertions can be isolated which are used to further discriminate between passenger mutations and cooperating driver mutations. Clonal integration sites define abundant mutations in a heterogeneous tumor that are used to define common integration sites (CIS). CIS are derived by comparison of clonal insertion sites found in all tumors and are defined as regions of the genome that are mutated at a rate higher than expected by chance in more than two tumors (Brett et al., 2011). By cross comparison of found CIS with databases of mutations in cancers in mouse and humans new driving mutations are found, which are further validated *in vitro* and *in vivo*.

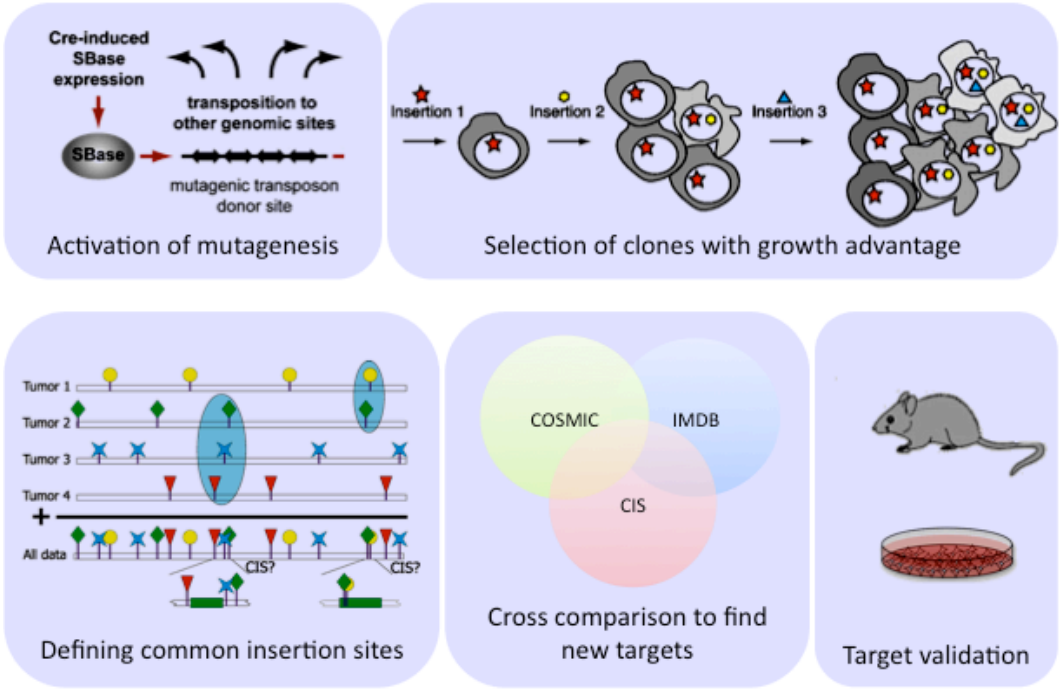


**Figure 5 Basic Concepts of the SB transposon screen.** In a first step CIS are defined statistically from the sequencing data obtained from the independent tumors. These CIS are cross-compared to databases listing human cancer genes like COSMIC and previous insertional mutagenesis databases like IMDB (Insertional Mutagenesis Database). Like this unknown genes involved in cancer can be discovered, which can be further validated by *in vivo* and *in vitro* knockdown or overexpression experiments. Adopted from (Berquam-Vrieze et al., 2011; Koudijs et al., 2011; de Ridder et al., 2006).

## Monitoring the SB system

To test the efficiency of Cre recombination in different B-cells containing compartments, flow cytometric analysis was performed on experimental mice harboring or lacking Cre-recombinase (Figure 6). Concordant with the literature (Zambrowicz et al., 1997), genes under the control of the *Rosa26* locus are ubiquitously expressed, as shown by the GFP positivity of practically all cells in the tissues of mice not carrying *Cd19-cre* analyzed by flow cytometry (Figure 6). Mice that harbor the *Cd19-Cre* knock-in show no decrease in the percentage of GFP expressing cells in the thymus. This compartment is known to harbor mainly T-cells but also contains around 3% B-cells (Miyama-Inaba et al., 1988), which however are not detected to lose GFP (Figure 6). Nonetheless, in other tissues described to undergo Cd19-Cre mediated recombination (Rickert et al., 1997), the highest amount of recombination was observed in the peripheral blood (PBMCs, 55%), followed by the spleen with 38%, the bone marrow with 35% and the lymph

nodes with 25% of deletion efficiency, indicating productive expression of the SB11 transgene.

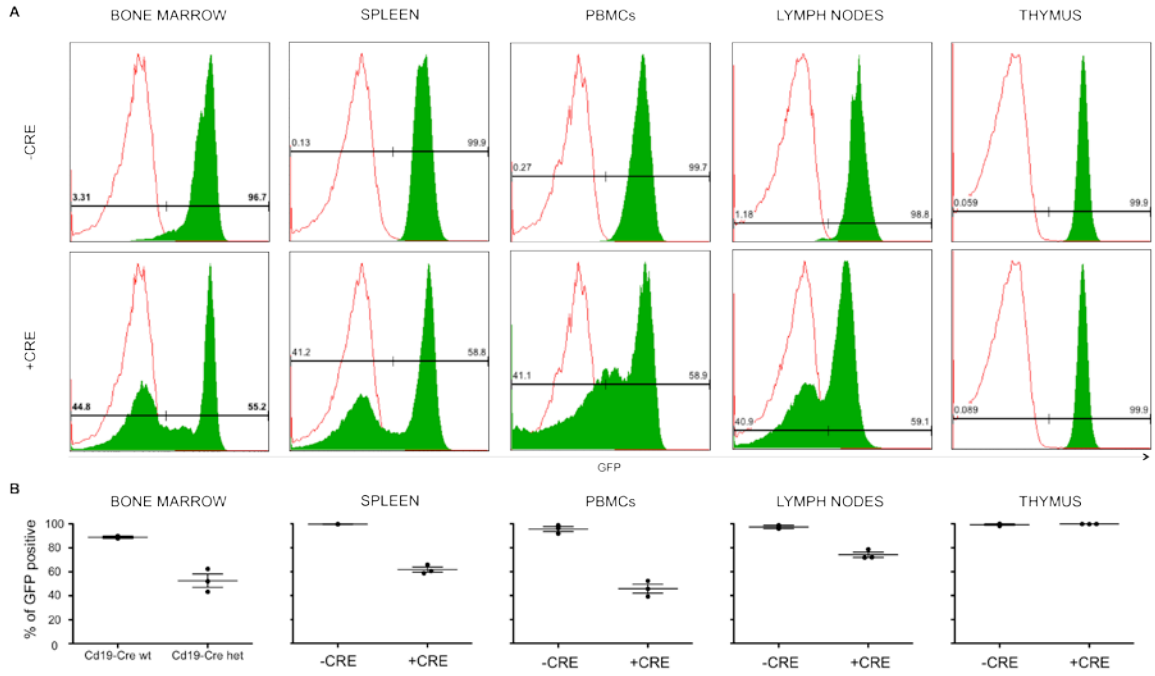


**Figure 6 Cre-mediated activation of insertional mutagenesis** FACS analysis was performed with single cell suspensions of tissue obtained of mice with the genotype C57 trans Myc-Cre-SB-T and C57 trans Myc-SB-T (in green). A C57 mouse served as GFP negative control (in red). Histograms present cell count on the y-axis and GFP positivity on the x-axis. GFP is expressed ubiquitously from the R26 promoter. Upon Cre mediated excision of the stop-lox-cassette, SB11, the transposase is expressed and leads to mobilization of transposons from the donor chromosome. This is observed as a loss in the GFP positive population, as the GFP reporter is also flanked by the lox-P sites. Ubiquitous GFP expression was shown in bone marrow (BM), spleen, peripheral blood mononuclear cells (PBMCs), lymph nodes and thymus in mice carrying the conditional transposase construct (upper panel). Mice that carry additionally the Cd19-Cre knock-in show a second population that is GFP negative and has undergone Cre-mediated excision. Quantification of the % of GFP positive cells of 3 mice per genotype was performed and is shown in the lower panel.

Thus, although the Cre-mediated deletion is not complete, from a quarter to one half of the cells in the compartments affected by lymphoma onset have active Sleeping Beauty, and are thus likely to undergo insertional mutagenesis.

### Adoptive transplantation serves as a tool for large-scale screening

A key requirement of this study was to generate a high number of tumors to define common integration sites (CIS) with a good level of certainty. Therefore an additional means to amplify the throughput of the screen was applied (Figure 7): instead of aging bred animals to develop lymphoma, we transplanted the isolated hematopoietic stem and progenitor cells (HSPCs) of 4-6 week old experimental mice to irradiated recipient mice.

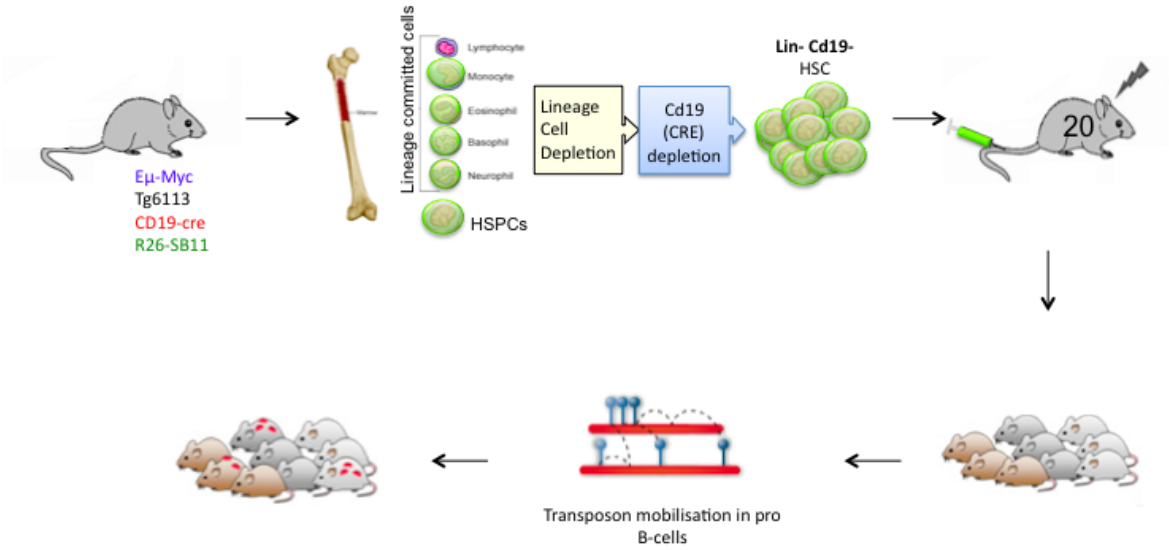


**Figure 7 Adoptive transfer strategy.** Hematopoietic stem and progenitor cells (HSPCs) are obtained by depleting sequentially lineage committed and Cd19 positive cells from the bone marrow fraction of Myc-Cre-SB-T mice. 10-300 x $10^3$ HSPCs are injected each into 20 irradiated C57JHsd mice. Reconstituted mice activate SB-induced mutagenesis in B-cells that also overexpress Myc. Mice were monitored for lymphoma onset by weekly palpation and checking for overall health of the mice. Moribund mice were sampled by taking blood, spleen, lymph nodes, and additional infiltrated organs to prepare single cell suspension for further flow cytometric analysis and to prepare paraffin sections for morphological analysis. Moreover, cell pellets were prepared for extraction of genomic DNA.

The rationale is that HSPCs do not yet express SB and have not yet experienced transposition events. Every HSPC gives rise to a multitude of B-cells that can all gain individual insertions by the transposon system. Thus, the recipients derived from a single donor will develop independent lymphomas with different genes hit by the transposon system. In this way we can amplify up to 20 fold the number of animals that develop independent tumors by insertional mutagenesis, hence significantly decreasing breeding time and increasing data collection from every quadruple transgenic mouse.

### Validation of the adoptive transplant strategy

As described above, it is essential that insertional mutation is not yet active in HSPCs: otherwise clonal selection would take place already before transplantation and give rise to lymphomas with the same driving event/s. In a first attempt we isolated HSPCs with magnetic beads that are coupled to antibodies recognizing lineage specific proteins (Lin) expressed on committed cells of the bone marrow. Excision PCR (Figure 8) on genomic DNA of these Lin- cells indicated that deletion

of the stop cassette had already occurred. As Eμ-*myc* mice show expansion of B-cell progenitors in the bone marrow (Langdon et al., 1986), we sequentially depleted for lineage positive cells and additionally for the excess of B-cell progenitors that most probably contaminated the pool of Lin- cells. Excision PCR of these double depleted cells showed that none of the cells had undergone Cre-mediated recombination: we thus decided to use these cells for adoptive transfer.

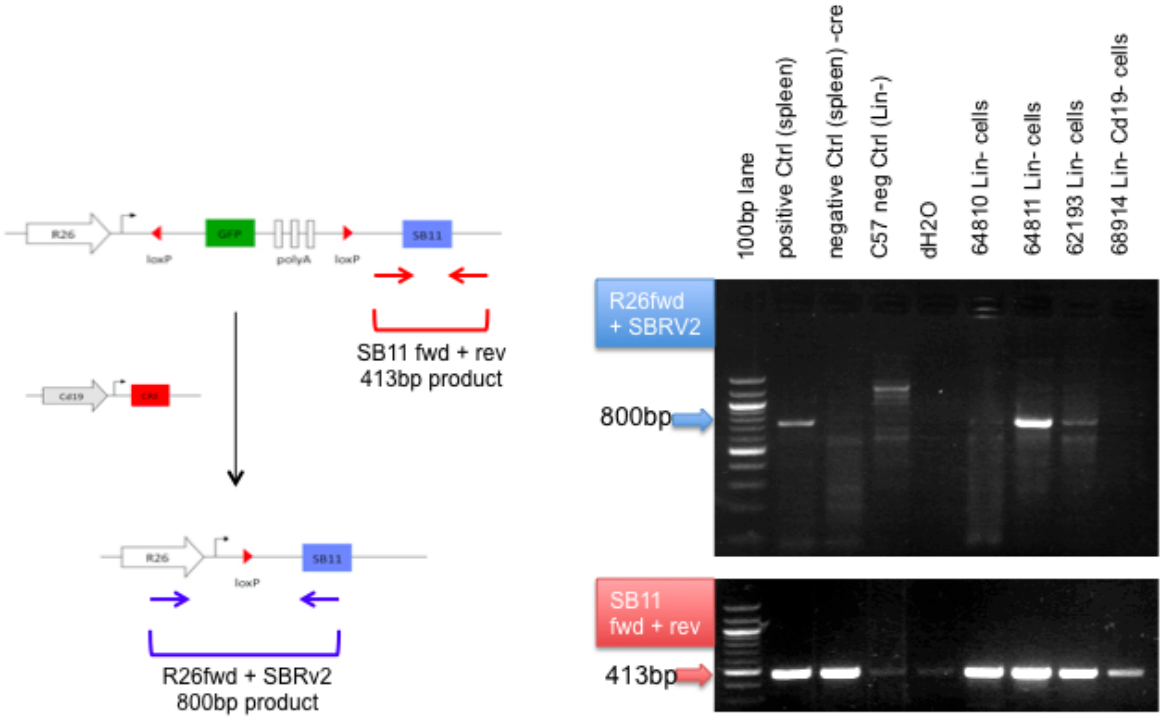


**Figure 8** Excision PCR amplifies a product only if Cre-mediated excision has occurred like shown in the scheme. In the upper panel as a positive control splenic cells from mice of the genotype Myc-Cre-SB-T and as a negative control Myc-SB-T splenic cells were analyzed. Cells isolated by magnetic bead sorting of the bone marrow of MycCreSBT mice to deplete lineage expressing (Lin-) and Cd19 expressing (Lin- Cd19-) cells show little to no Cre-mediated excision, whereas the fraction of cells depleted only for lineage committed cells shows Cre-mediated deletion.

Spontaneous secondary mutations in the Eμ-*myc* mouse lead to lymphoma from 3 to 5 month of age (Adams et al., 1985; Harris et al., 1988). Therefore it was essential to assure that the reconstitution persists for the time that is needed for lymphomas to develop, and to prove functionality of the adoptive transfer strategy by titration of the number of cells needed for stable long-term bone marrow reconstitutions. An initial experiment shown in **Figure 9** was performed taking

advantage of the Ly5.1/Ly5.2 allelic variants of CD45, a glycoprotein expressed on hematopoietic cells. The scope was to transplant different amounts of HSPCs (50.000, 100.000 and 300.000) from a Ly5.2 mouse to lethally irradiated Ly5.1 recipients. By flow cytometric analysis of the peripheral blood stained against the allelic variants we observed that reconstitution after 8 weeks was up to 70%. Stable long-term reconstitution was established, with up to 90% of the peripheral blood population was deriving from donor progenitors (Ly5.2) after 19 weeks. Moreover, the lowest number of cells used in this titration - 50000 cells - was enough to stably reconstitute the hematopoietic system of lethally irradiated recipients.

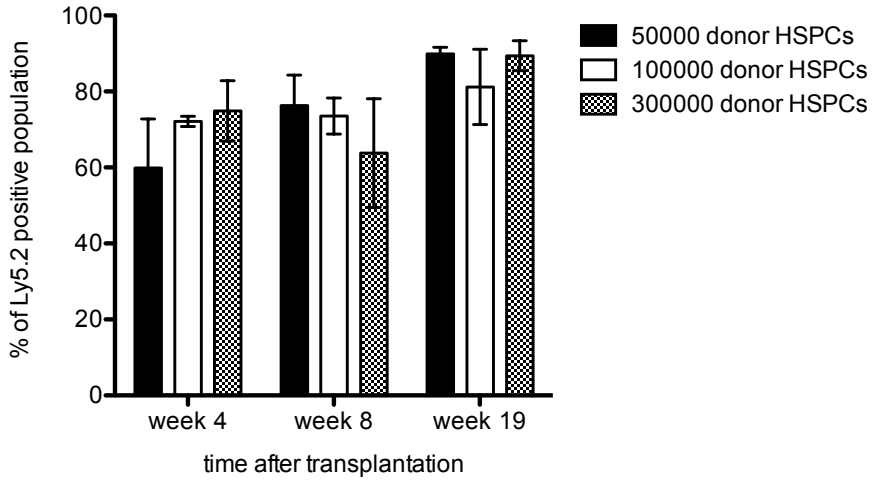


**Figure 9 Amount of HSPCs needed for stable reconstitution of the recipient hematopoietic system.** By hydrodynamic tail vein injection 50000, 100000 or 300000 lineage negative cells carrying the Ly5.2 allele were transplanted into cohorts of each 3 recipient Ly5.1 mice. After 4, 8 and 19 weeks peripheral blood was taken from these animals and PBMCs stained for Ly5.1 and Ly5.2 were analyzed by FACS. After 19 weeks the amount of hematopoietic cells deriving from the donors cells (Ly5.2) was between 80 and 90% when injecting 50000, 100000 or 300000 cells.

The strategy was then further validated in the experimental cohort: PBMCs from transplanted animals were stained for B220 and analyzed by flow cytometry (Figure 10A). In contrast to the experiment mentioned above using the Ly5.1/Ly5.2 allelic variants, we exploit the GFP reporter gene as marker of transplantation efficiency. All cells of the hematopoietic system ubiquitously express GFP from the *R26 locus*, apart from the B-cell lineage that deletes the Lox-

GFP-stop-lox cassette. Therefore the % of GFP positive, B220 negative cells were monitored to obtain a representative measure of the population derived from donor HSPCs.

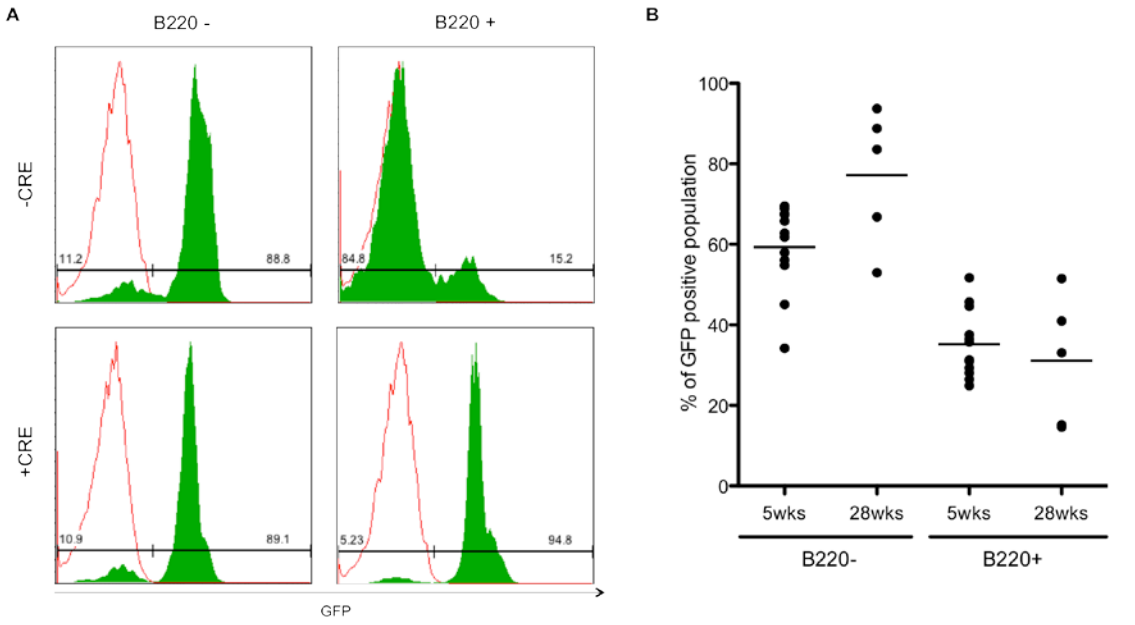


**Figure 10 Stable reconstitution and active mutagenesis in B-cells of transplanted mice.** Deletion of the GFP-reporter is dependent on specific expression of Cre-recombinase in B-cells. Therefore GFP positivity of non B-cells (B220-) and B-cells (B220+) of mice carrying Cre-recombinase (Cre) or devoid of it (-Cre) was assessed 5 and 28 weeks after transplantation by flow cytometry. **A** Examples of a histograms depicting GFP positivity of PBMCs of mice 28 weeks after transplantation show high reconstitution efficiency (88.8% in Cre-, 89.1% in Cre+), whereas recombination is only occurring in Cre+ B220+ mice with a remaining GFP positive population of 15.2% compared to 94.8% GFP positive cells in mice devoid of Cre-recombinase (-Cre B220+). As a GFP control a non-transplanted recipient was analyzed in the same way and is depicted as the red curve in the histograms. **B** Quantification of stable reconstitution and efficient recombination in mice 5 weeks (n=12) and 28 weeks (n=5) after transplantation. In transplanted mice, reconstitution is monitored by the GFP positivity of B220- cells and Cre-mediated recombination efficiency is assessed by the loss of GFP positivity of B220+ population in comparison to the B220+ population. Stable transplantation and Cre-mediated recombination is shown by the quantification of GFP positivity in B220- and B220+ PBMCs, respectively.

A mean reconstitution of 60% (minimum 34%, maximum 70%) was observed 5 weeks after transplantation, which was increased to 77% of reconstitution (minimum 53%, maximum 94%) after 28 weeks (Figure 10B). At the same time, this assay can be used to indirectly monitor the deletion frequency in B-cells, by monitoring the reduction of GFP expression relative to other blood cell populations (B220-). After 5 weeks there was a reduction of 40% of GFP expression (minimum 14%, maximum 63%) and of 57% (minimum 23%, maximum 83%) 28 weeks after transplantation, suggesting active mutagenesis in the B-cell lineage of transplanted mice.

Summarizing the above points, the adoptive transfer strategy was shown to be a valid and feasible method to enlarge the cohort of experimental mice that undergo SB activation, and hence insertional mutagenesis. This allowed us to derive up to 20 mice prone for lymphomagenesis out of one mouse harboring the right genotype and was an essential strategy for obtaining a high number of experimental animals in order to identify even low-frequency cooperating partners in Myc-induced lymphomagenesis.

**Table 1 Donors for adoptive transfer**

| Donor ID# | Age at HSPC collection (days) | Eμ-*Myc* (TG) | Tg6113 (TG) | R26-lsl-SB11 | Cd19-CRE | # of recipients | cohort |
|---|---|---|---|---|---|---|---|
| 62193 | 56 | + | + | het | het | 19 | experimental |
| 64810 | 40 | + | + | het | het | 20 | experimental |
| 64811 | 36 | + | + | het | het | 20 | experimental |
| 64813 | 33 | + | + | het | het | 20 | experimental |
| 68729 | 37 | + | + | het | het | 20 | experimental |
| 68732 | 37 | + | + | het | het | 20 | experimental |
| 69814 | 46 | + | + | het | het | 20 | experimental |
| 73947 | 39 | + | + | het | het | 15 | experimental |
| 73948 | 39 | + | + | het | het | 10 | experimental |
| 81496 | 36 | + | + | het | het | 20 | experimental |
| 82089 | 42 | + | + | het | het | 20 | experimental |
| 82091 | 43 | + | + | het | het | 20 | experimental |
| 85620 | 41 | + | + | het | het | 5 | experimental |
| 85630 | 41 | + | + | het | het | 20 | experimental |
| 85631 | 43 | + | + | het | het | 20 | experimental |
| 88599 | 33 | + | + | het | het | 20 | experimental |
| 90489 | 45 | + | + | het | het | 23 | experimental |
| 62200 | 58 | - | + | het | het | 20 | control |
| 73949 | 41 | + | - | het | het | 5 | control |
| 73951 | 41 | + | - | het | het | 5 | control |
| 75173 | 41 | + | - | het | het | 5 | control |
| 88600 | 33 | + | - | het | het | 5 | control |
| 81490 | 38 | + | + | het | wt | 5 | control |
| 82086 | 42 | + | + | het | wt | 5 | control |
| 82098 | 43 | + | + | het | wt | 5 | control |
| 88594 | 33 | + | + | Het | wt | 5 | control |
| 88601 | 33 | + | + | Het | wt | 5 | control |

## Strong genetic cooperation of SB-mutagenesis in the Eμ-myc mouse

Hematopoietic stem and progenitor cells (HSPCs) of 17 donors with the quadruple genotype were collected and transplanted to 312 lethally irradiated C57JHsd mice.

As controls, HSPCs from one Cre-SB-T, five Myc-SB-T and four Myc-Cre-SB donors were transplanted to a total of 20, 25 and twenty 20 recipients, respectively as shown in Table 1. These mice, together with non-transplanted, bred animals (12 with the genotype Myc-Cre-SB-T, 13 Cre-SB-T, 25 Myc-Cre-SB and 15 Myc-SB-T) were monitored for tumor development.

Tumor latencies of the transplanted mice of different genotypes were blotted in a frequency plot shown in Figure 11.

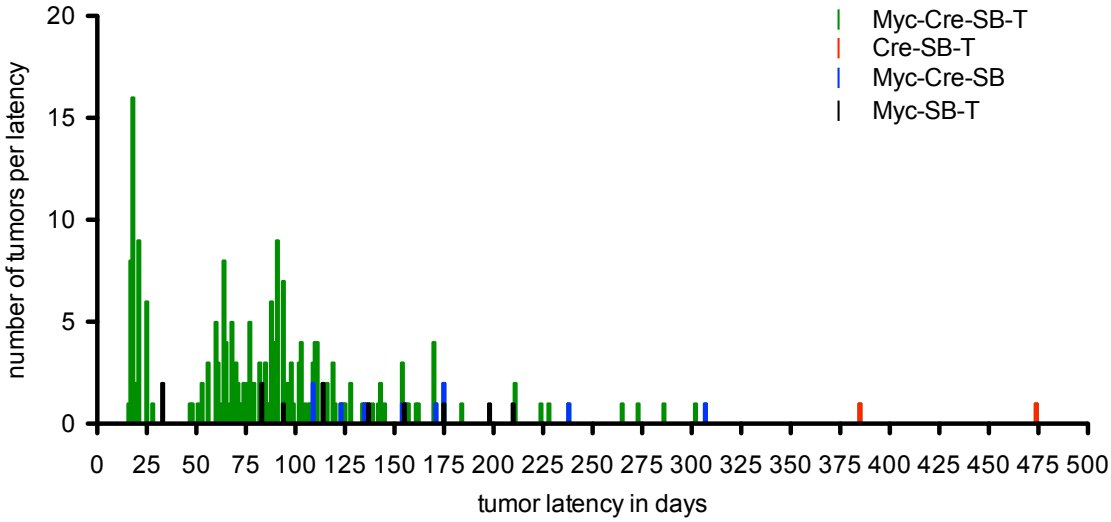


**Figure 11 Distribution of tumor onset in recipient mice transplanted with HSPCs from donors with different genotypes.** In this frequency plot, on the x-axis tumor latency in days after transplantation is plotted. In the y-axis the number of animals diseased at a certain tumor latency are depicted.

Tumor latency was decreased in experimental mice (average 91 days) compared to Cre-SB-T (429 days), Myc-Cre-SB (168 days) and Myc-SB-T (119 days) mice. Interestingly, tumor onset showed three peaks at days 18, 64 and 91. The first group comprised 45 animals that showed a high tumor burden which were transplanted with HSPCs obtained from the same Myc-Cre-SB-T donors. HSPCs from donor 64813 in particular showed signs of contamination of B-cells when monitored for excision PCR, in contrast to donor 69814 (Figure 8). This suggests that fast aggressive tumors stemmed from donor-derived tumor clones already present in some HSPC preparations. By monitoring the Kaplan-Meier curves of

recipients transplanted with individual donors (Figure 12), three donors (64813, 82091 and 73948) were found to give rise to such fast tumors in recipients with a median survival of 21 days (64813), 18 days (82091) and 25 days (73948) compared to the usual median tumor onset of 94 days (90489) and 128 days (69814).

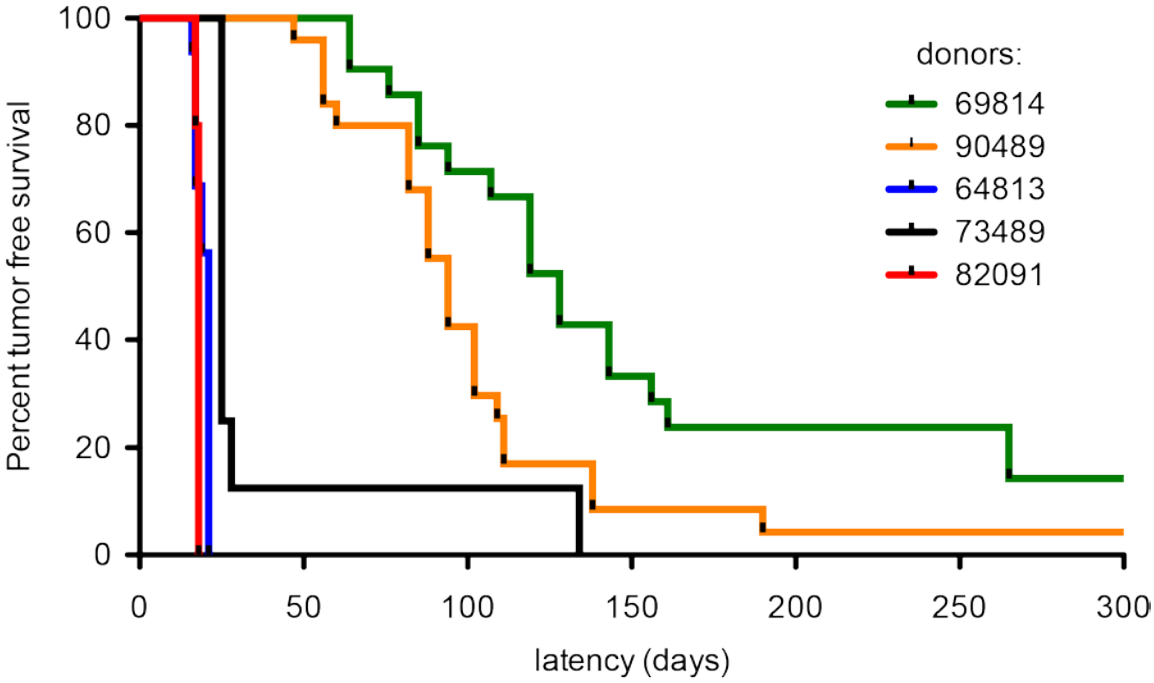


**Figure 12 Tumor free survival of groups of recipients transplanted with the same donor.** Examples of the survival of 5 groups of mice transplanted with each 50000 HSPCs of MycCreSBT donors 69814 (n=20, light green), 90489 (n=25, dark green), 64813 (n=20, purple), 73489 (n=10, violet) and 82091 (n=20, pink). Transplanted donors in reddish color are suggested to evolve from the same tumor clone present in the donor, whereas tumorigenesis is random and independent in the greenish cohorts.

The group of transplanted mice that was found diseased during the first 35 days post transplantation was thus excluded from the Kaplan-Meier curves monitoring the whole cohort of transplanted mice. The second group of Sleeping Beauty mice with a peak of disease onset at 64 days (considered +/- 3 days) showed an enrichment in tumors from donor 64810 (n=8) but also included tumors from 3 other donors (n=9), suggesting that results obtained from tumors transplanted with HSPCs of donor 64810 have to be interpreted with caution. The third group peaking around day 94 (+/- 3 days) were from 10 different donors comprising 27 tumors, with no clear communalities, suggesting that genetic factor changes might be the reason for this group.

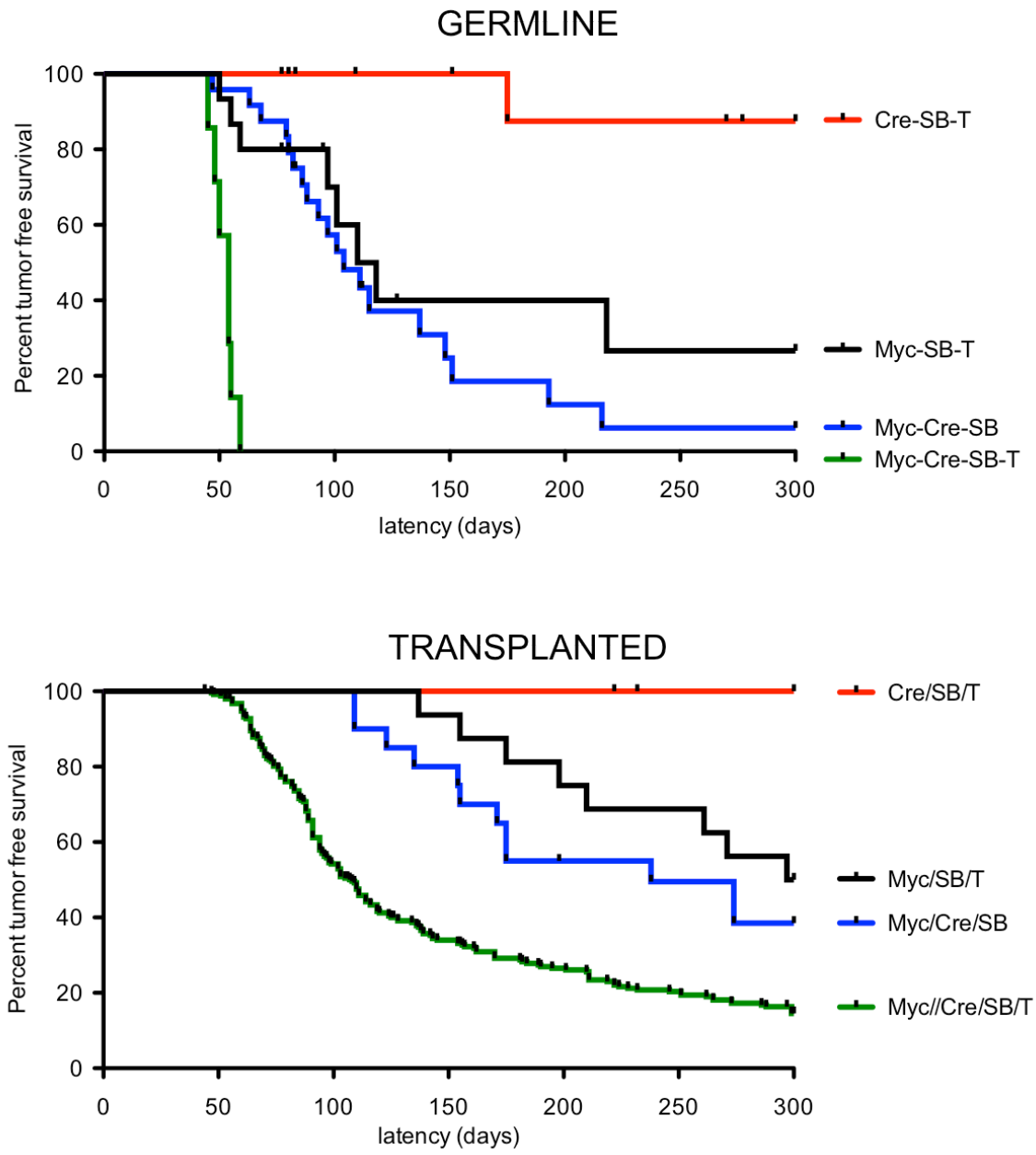


**Figure 13 Genetic cooperation of Eμ-*myc* and SB induced mutagenesis.** Tumor free survival was monitored in germ line (upper panel) MycCreSBT mice and controls and in transplanted animals (lower panel). The SB system alone (red) is not strong enough to induce tumorigenesis, whereas control mice that lack one element of the SB mutagenesis system (black, no active transposase and blue, no transposable elements) show lymphoma latencies comparable to spontaneous Eμ-*myc* tumors (104 days and 114 days in germ line, 238 days and 284 days in transplanted animals, respectively). Cooperation of the Eμ-*myc* transgene with insertions from the SB accelerate tumor onset dramatically with a median latency of 55 day in germ line and 94 days in transplanted animals (green). The day of necropsy of mice found moribund after birth or after transplantation is depicted on the x-axis.

Figure 13 shows the Kaplan-Meier survival curves for germ line (top) and transplanted animals (bottom). Median tumor free survival for germ line animals/transplanted animals is 54/108 days compared to more than 300/360 days in Cre-SB-T mice, 104/238 days in Myc-Cre-SB and 114/298.5 days in Myc-SB-T mice.

These data show that SB-induced insertional mutagenesis significantly accelerates lymphomagenesis in Eμ-*myc* mice. The selection of tumor clones in transplanted mice seems random as latencies within - and between - the different cohorts are variable. Of note, the resulting lymphomas are also transplantable, as shown by rapid and homogenous disease onset in the recipients of donors 64813, 73489, and 82091. These results confirm the genetic cooperation of Sleeping Beauty induced mutations with Myc to induce lymphomagenesis, and provides a valid basis for the pursuit of the proposed screen.

## SB-induced tumors show heterogeneous lymphomagenesis

Characterization of the tumors was performed in a first step during lymph node resection, which showed a high amount of diversity in lymphoma size, color, dissemination and metastasis. An example is shown in Figure 14A together with a control mouse.

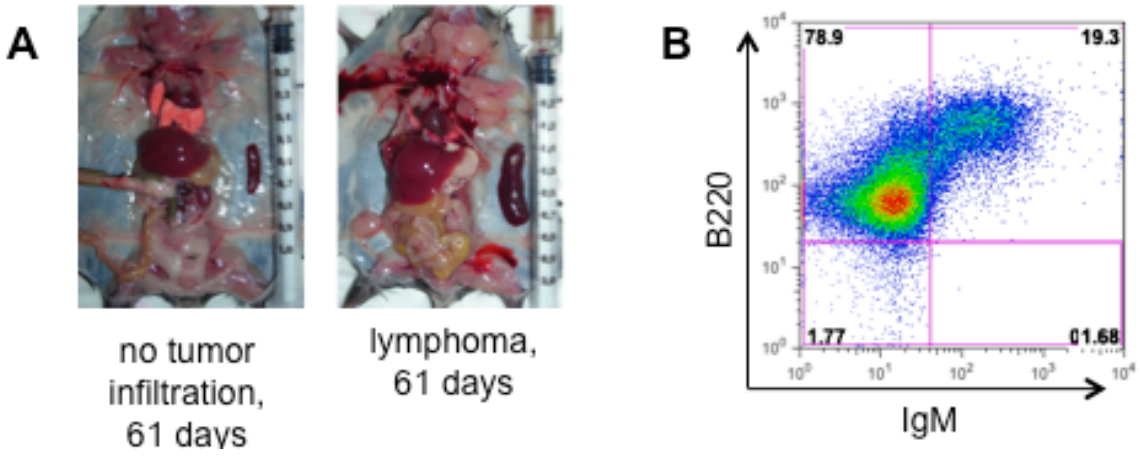


**Figure 14 Characterization of tumors. A** C57 trans MycCreSBT mice develop lymphoma with splenomegaly and infiltration of lymph nodes. An age-matched control transplanted mouse is shown on the left. **B** An example of an Eμ-*myc* lymphoma stained for B220, a marker of the B-cell lineage and IgM, a marker of immature B-cells that shows enrichment in the pre-pro B-cell population (B220 low/IgM negative).

The infiltration was usually spread to all the lymph nodes and mice also showed paralysis of the hind legs, splenomegaly without infiltration of lymph nodes, enlarged liver and infiltration of kidneys or thymus, as described previously for the Eμ-*myc* mouse (Adams et al., 1985; Harris et al., 1988; Strasser et al., 1990).

Moreover, using flow cytometry the B-cell stage of the lymphoma can be assessed as shown in a typical example of a pre/pro-B-cell lymphoma stained with B220 and IgM arising spontaneously in an Eμ-*myc* mouse (Figure 14B).

Histological analysis was performed with fixed and paraffin embedded liver sections and is shown in Figure 15 and show infiltration with typically observed small round cells, suggested to be lymphocytes of the B-cell lineage.

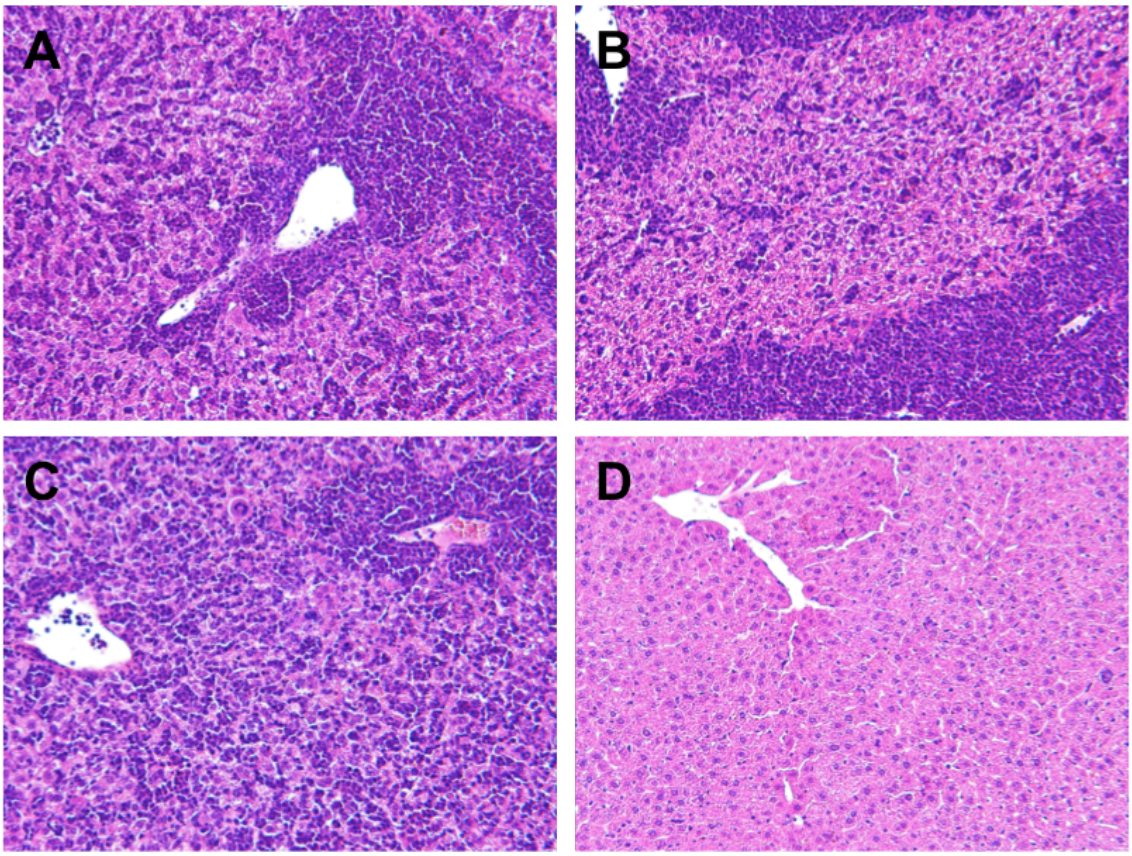


**Figure 15 Histological analysis.** Representative histology of HE stained liver sections from Myc-Cre-SB-T transplanted mice developing lymphoma 85 days (Ct69814E, A), 64 days (Ct64810H, B) and 25 days (Ct73948F, C) after transplantation. The liver of a control mouse is shown in D. Original magnification was 10x.

Further, we analyzed hematopoiesis by monitoring the abundance of white blood cells (WBCs) as well as lymphocytes in peripheral blood (Figure 16). Of the mice that developed lymphomas 125 were analyzed and showed a significant increase in white blood cells with an average count of $38.3 \times 10^6$ cells compared to WBC of 9 healthy C57JHsd ($11.3 \times 10^6$) or 13 Myc-Cre-SB-T donors ($16.15 \times 10^6$).

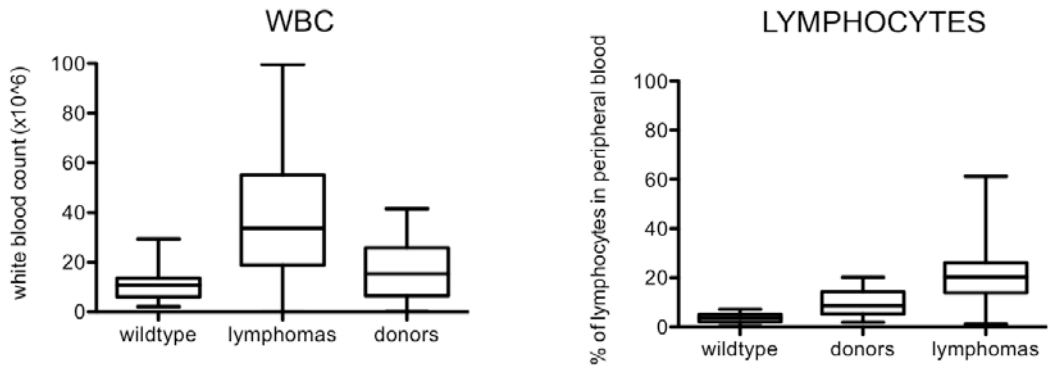


**Figure 16 Analysis of PBMCs of mice developing tumors.** On the left the white blood count (WBC) and on the right the % of lymphocytes present in the peripheral blood of the cohort of healthy C57JHsd wild type mice (n=9), donors at the age of HSPC recovery (n=13) and lymphomas at the day of necropsy (n=125) are depicted. A difference between wild type or donors and lymphomas was significant in WBC ($p<0.0017$). Difference between wild type and donor or lymphoma lymphocyte count in the peripheral blood sample was also significant ($p<0.006$).

In the same samples, a significant increase in the mean percentage of lymphocytes present in the peripheral blood was observed when comparing wild type (3.5%), Myc-Cre-SB-T donors (9.9%) accounting for the increased pre-B-cell population due to overexpression of Myc as shown for the spleens of healthy 4-6 week old Eμ-*myc* mice (Gorrini et al., 2007). Moreover, increased lymphocyte count in the lymphoma samples (20%) account for circulating tumor cells in the moribund animals as reported before (Adams et al., 1985).

To verify the differentiation of the lymphomas, single cell suspensions of the resected infiltrated lymph nodes were stained for their immune phenotype. The different markers are summarized in **Figure 4** of this chapter. Surprisingly the lymph nodes also included a B220- population, suggesting other cells than the lymphoma population being present in the resected lymph nodes (**Figure 17**).

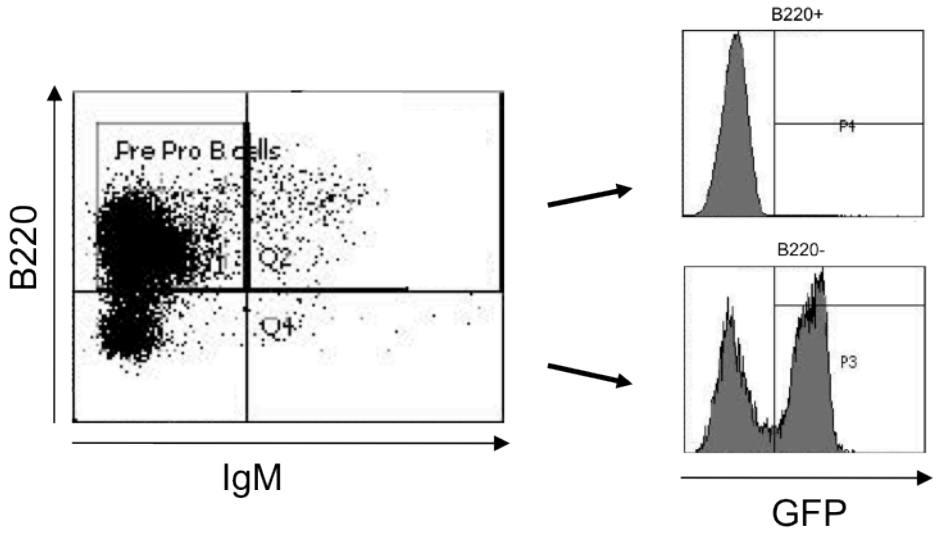


**Figure 17 Immune phenotyping on resected lymph nodes from a transplanted mouse with lymphoma.** Single cell suspension was stained for B220 and IgM and analyzed by flow cytometry. In the dot plot, IgM staining is depicted on the x-axis and B220 staining on the y-axis. B220$^{low}$IgM negative cells are defining the pre-/pro-B cells, and cells in quadrant 2 (Q2) are defined as immature B-cells expressing B220 and IgM. Non B-cells were defined as negative for B220 and IgM. The dot blot shows a prevalent pre/pro B-cell population (81.4%) and includes residual B220 negative cells (16%). Histograms of B220+ cells and B220- cells show that B220+ lymphoma cells have undergone deletion of the GFP-containing Stop-cassette, whereas the non-B-cell population in the resected lymph nodes shows the GPF positive and negative populations, reflecting the % of reconstitution.

Thus, staining on selected samples having a low amount of B220+ cells were performed using the T-cell specific marker Cd3ε and the myeloid marker Gr-1, together with B220 and IgM. The population not staining for B220 was to a major extent Cd3ε positive and to a small extent Gr-1 positive, suggesting T-cells to be the second main population in lymph nodes of the diseased animals that were analyzed. In a study that performed a shRNA knockdown screen using transplantation of fetal liver HSPCs from Eμ-*myc* mice, immune phenotyping of lymph nodes was performed as here (Bric et al., 2009, Figure S3). The single cell suspensions of lymph nodes of moribund mice that were used for immune phenotyping showed a predominant B220 positive population in three out of five experiments shown and additional non-B220 populations, like we find here, in two analysis (Rad17, Srfp). Moreover, four out of five lymphomas analyzed by Bric et al. stained also positive for the T-cell marker Thy1.2+ (Mek1, Angpt2, Rad17 and Numb) and for the myeloid marker Cd11b+ (Mek1, Rad17). Nevertheless, the

disease assigned from the flow cytometric analysis was pre B-cell lymphomas (B220+/IgM-). This suggests that in transplantation experiments with Eμ-*myc* HSPCs, cells from other lineages can be present in infiltrated lymph nodes next to the lymphoma population. Previous analyses show no expression of the Eμ-*myc* construct in enriched T-cell, macrophage and myeloid populations (Alexander et al., 1987). Moreover, coherent to our findings of inactive SB in the thymus (see Figure 8), splenic Thy2.1 cells in the Cd19-Cre mouse model did not delete a floxed *polβ* gene (Rickert et al., 1997). Concluding, this suggests strongly that SB induced tumors are lymphomas of the B-cell lineage.

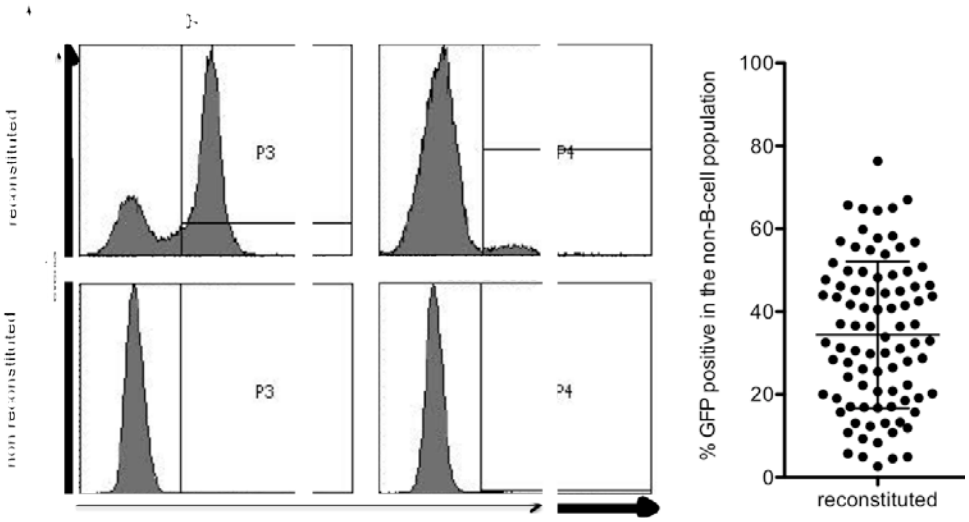


**Figure 18 Efficient transplantation** was observed in the mice that were analyzed (n=95). Single cell suspensions of lymph nodes were monitored by flow cytometry staining the B-cell lineage with B220. The histograms in **A** depict examples of mice that have undergone stable adoptive transfer (reconstituted) or where donor HSPCs failed to reconstitute the hematopoietic system of recipients (not reconstituted) as visible as the amount of GFP positive cells in the non B-cell compartment (B220-, P3). Quantification of 95 mice subjected to this analysis **B** show the amount of GFP+ cells in 96.2% of the mice reconstituted, with an average 34.5% of the population deriving from GFP-positive donor cells (SD 17.7%).

Taking advantage of the residual non-B-cell population in the lymph node, several issues of the experimental system can be analyzed by flow cytometry. First of all, as stated above, to evaluate engraftment of donor progenitors was assessed by GFP expression in the non-B-cell compartment (B220 negative) of the lymph nodes, where Cre-mediated recombination of the lox-GFP-stop-lox cassette has not taken place. Engraftment of the donor HSPCs was shown in 95.8 % of mice analyzed (n=95), with an average reconstitution of recipients with donor HSPCs of 34.5%,

suggesting that long-term transplantation was efficient in this experimental setting (Figure 18).

Furthermore, active Sleeping Beauty insertional mutagenesis could be assessed in the lymph nodes by comparing GFP expression of non-B cells to the B220+ population (Figure 19) as CD19-Cre expression is restricted to the B-cell lineage (Rickert et al., 1997). Of the lymph nodes sampled, only one showed no reduction of GFP and thus did not undergo Cre-mediated recombination of the lox-GFP-stop-lox cassette. The other samples (Figure 19) showed an average reduction of 84.8% with residual 5% of the B-cell population expressing GFP. Not complete loss of GFP suggests for the presence of B-cells that have not undergone Cre-mediated deletion as described for the Cd19-Cre mice that do not delete 20-25% of the pre-B-cell compartment and 5-10% of mature splenic B-cells (Rickert et al., 1997).

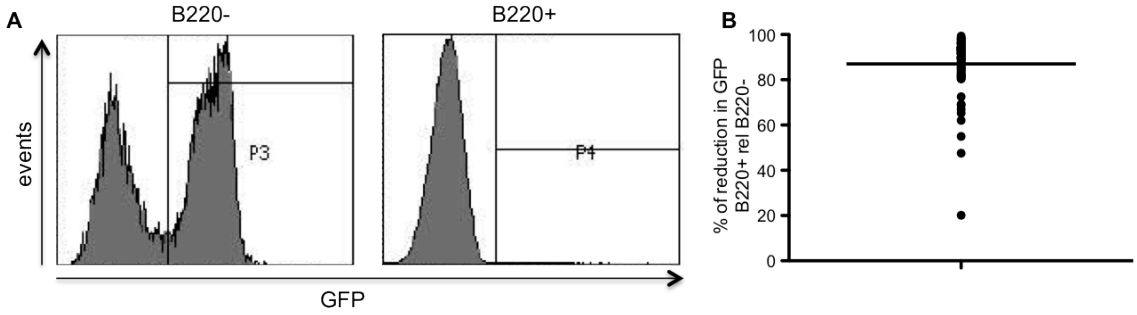


**Figure 19 Selection for tumor clones that have undergone Cre-mediated excision and have an active insertional mutagenesis.** Single cell suspensions from infiltrated lymph nodes were analyzed by flow cytometry. The reduction in GFP comparing the non-B-cell compartment (BB20-) to the B-cell lineage (B220+) is indicative for SB-mediated insertional mutagenesis. An example of histograms showing GFP expression on the X-axis and the cell count on the Y-axis in shown in A. 57% of the non-B-cells (P3) and 0.4% of the B-cells (P4) are GFP positive. Quantification of all lymph nodes analyzed is shown in B and shows an average of 87% loss of GFP expression comparing the B220+ to B220- population.

To further define the developmental stage of the lymphomas, we defined progenitor B-cells as B220+c-kit+ (pro B), early pre B cells as B220+c-kit+Cd25+ (pre/pro B), pre-B-cells that have completed VDJ recombination as B220+Cd25+ (pre B) and immature B-cells, that express a functional BCR as B220+IgM+. 46% of the lymphomas show one B-cell population prevailing, whereas 38% and 16% show two populations or three individual populations, respectively as shown in examples in Figure 20. The tumor populations found were spread among pro B

(32%), immature B (30%), pre B (23%), pre/pro B (12%) and populations that do only stain positive for B220 (c-kit-Cd25-IgM-, 3%) as shown in Figure 21. This reflects the range of developmental stages described in literature for the Eμ-*myc* mouse model (Harris et al., 1988). However this also indicates that 54% of the mice immunophenotypically analyzed develop more than one independent tumor, in contrast to Eμ-*myc* mice, which develop mainly monoclonal tumors and only in few cases mice harbor more than one independent tumor (Adams et al., 1985; Alexander et al., 1987). Thus spontaneous onset of lymphoma favors the outgrowth of one clone, whereas the involvement of a highly active insertional mutagenesis seems to favor the selection of multiple tumor clones in the same mouse.

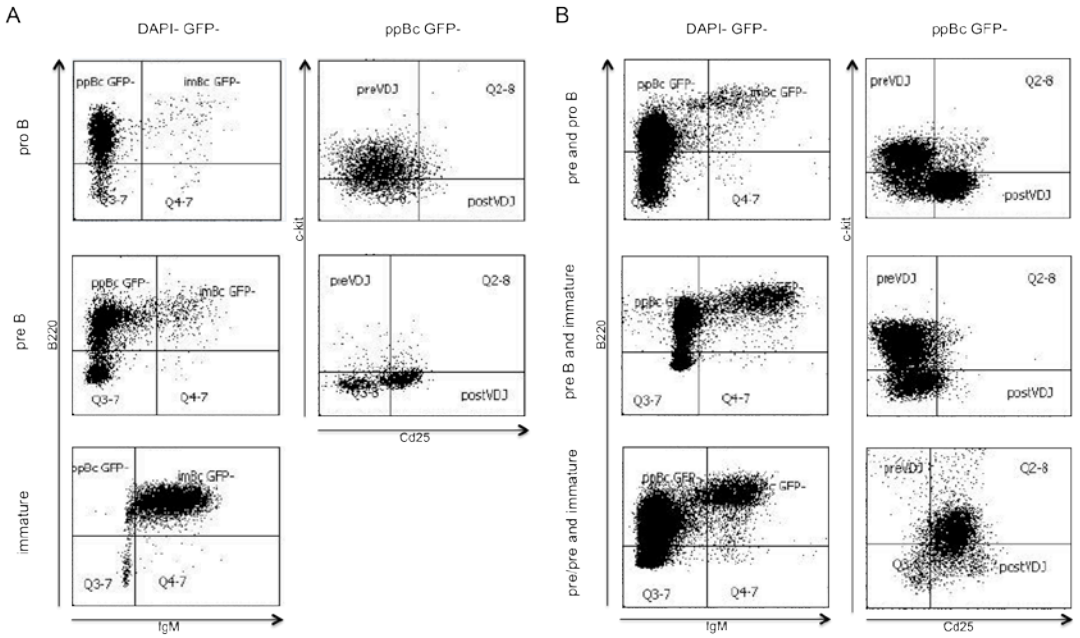


**Figure 20 Examples of immunophenotyping**. Single cell suspensions of lymph nodes were stained for DAPI, B220 IgM Cd25 and c-kit to define developmental stages of B-cells. The gating was performed as follows: First the cell population was gated according to granularity and cell size using FSC and SSC. In a next step DAPI negative cells were gated to exclude dead cell. The DAPI negative gate was divided into GFP positive and negative cells. In this image the live cells negative for GFP stained for B220 and IgM are depicted in the left panel (DAPI-GFP-). Pre and pro B-cells express low B220 and are IgM- (ppBc GFP-) and immature B-cells are B220+IgM+. Further gating on ppBc GFP- is shown in the right panel with B220+c-kit+ cells defining the pro-B-cell population and labeled as preVDJ. Pre-B-cells are positive for Cd25 and are labeled postVDJ and pre/pre B-cells show staining for both Cd25 and c-kit. In A single populations are depicted, whereas in B multiple populations present in the same resected lymph node of a mouse developing lymphoma is depicted.

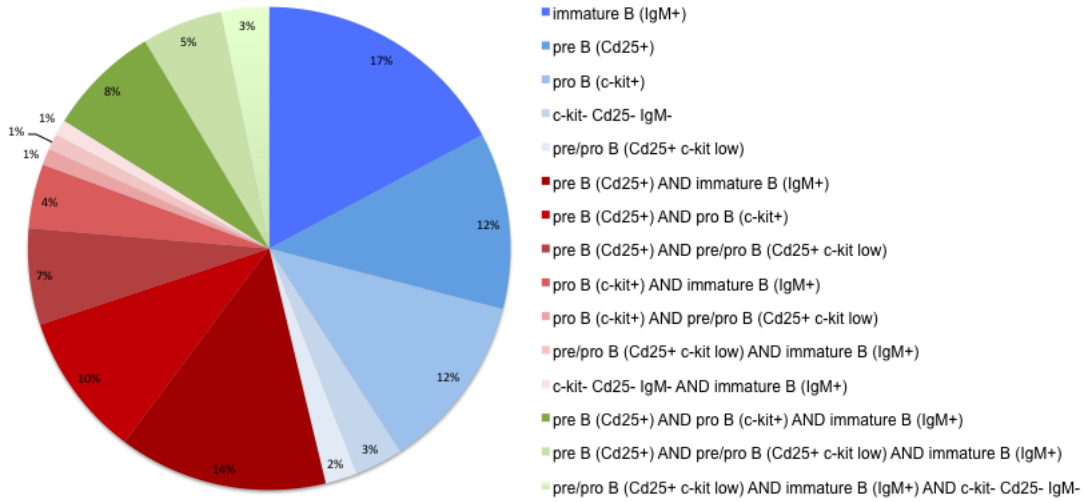


**Figure 21 Immunophenotyping.** 95 mice of the experimental cohort that developed lymphoma used for this analysis. Single cell suspensions of the lymph nodes were stained with the indicated markers and analyzed by flow cytometry. 46% of the tumors analyzed showed one prominent population (blue), whereas 38% of tumors show two populations (red) and 16% of the tumors are comprised of three populations (green). The predominant immune phenotype is the immature B-cell lymphoma (17%), followed by lymphomas that comprised both pre B and immature B-cells (14%) and pre or pro-B-cell lymphomas (both 12%), reflecting the typical early B-cell phenotype observed in lymphomas of the Eμ-*myc* mouse.

To confirm the oligoclonality of most tumors, a PCR was performed with primers annealing to the *CDR3* locus and to the *JH4* locus of *IgH* (Figure 22). This method allows looking at clonality status as during B-cell differentiation the V(D)J rearrangement gives rise to different subsets of clones (Rajewsky, 1996; Tiller et al., 2009; Yu and Thomas-Tikhonenko, 2002).

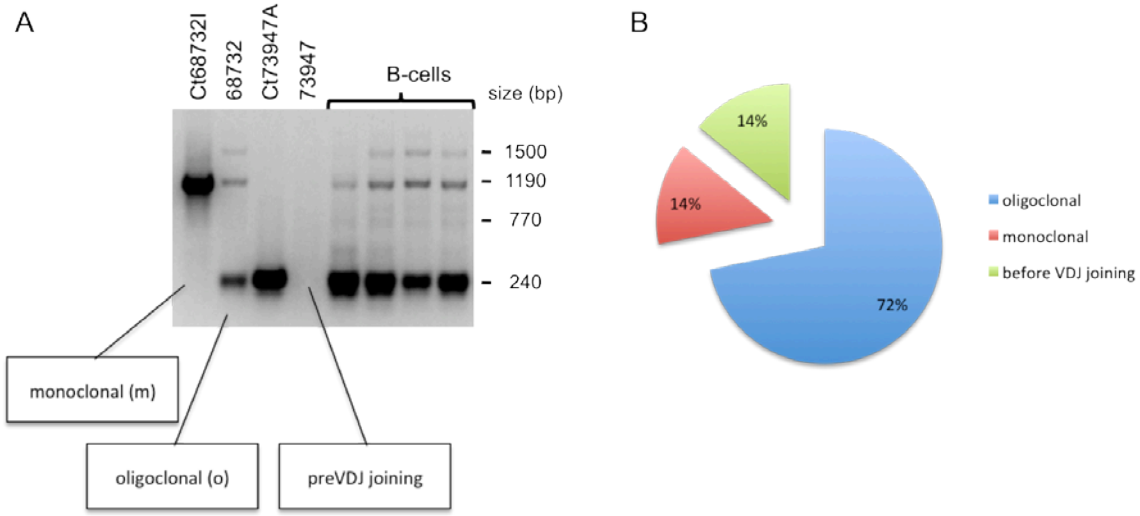


**Figure 22 Oligoclonality of tumors. A** VDJ PCR was performed to assess B-cell clonality. An example of an oligoclonal (68732), monoclonal (Ct68732I and Ct73947A) and HSPCs of a donor (73947), that has not yet undergone gene rearrangement of the IgH variable (V)-diversity (D)-joining (J) region is shown. As positive control of the reaction Cd19 positive isolated B-cells cells isolated from 4 healthy donors are shown. Note that there is an PCR bias towards the smallest fragment (240 basepairs(bp)) if it is present. **B** Of the 87 tumors analyzed more than 70% were found to be oligocolonal whereas less than 30% showed either monoclonality or did not yet undergo V(D)J joining.

It is limited by the fact that only 4 different *JH* segments are present in the *IgH*, but serves as fast and inexpensive method to define oligoclonality. A monoclonal tumor thus shows the same V(D)J rearrangement, whereas oligoclonal tumors may show more than one possible rearrangement that can be amplified by PCR. In line with our data from immunophenotyping to find multiple populations arising in the pooled infiltrated lymph nodes of one mouse, we find that more than 70% of tumors contain more than one V(D)J recombination status. This suggests that the SB system is a potent inducer of cooperating mutations in the Eμ-*myc* mouse model leading to the expansion and selection of more than one tumor clone per mouse. The same was observed for a SB screen performed with constitutively active SB transposase and a *T2Onc3* transposon elements, in which each mouse developed on average 3 independent solid tumors of different etiologies (Dupuy et al., 2009) but not in another SB screen using the conditional targeting of SB-mutagenesis to hematopoietic progenitors using *LcK-Cre* and *CD4-Cre* together with the *T2Onc2* transposon elements that induces T-cell lymphoma (Berquam-Vrieze et al., 2011).

## Identification of SB-induced modifiers of Myc-induced lymphoma

### Sequencing

To retrieve the genomic loci mutated by integration of the *T2Onc2* leading to tumor development, ligation-mediated PCR (LM-PCR) is used as depicted in Figure 23. This method allows specific enrichment of genomic regions flanking the inserted transposon from genomic DNA isolated from the tumor or control samples. In short, genomic DNA from infiltrated lymph nodes was extracted and digested with AluI or NlaIII, two frequently cutting restriction endonucleases. The 300-900 base pair fragments were ligated to the corresponding adaptor linker

oligos for each restriction enzyme and a clean up step was performed to deplete for unligated adaptors. Following this step, a PCR was performed using primers annealing in the adaptor sequence and in the transposon inverted repeats (IR). Specific amplification of transposon containing elements is achieved by using modified adaptors that block primer annealing. From transposon specific primers however a primary transcript is obtained, which serves as a template for further amplification rounds with both primer pairs. The product obtained was then used as template in a nested PCR that included an 8bp barcode and the adaptor sequences used for Illumina sequencing.

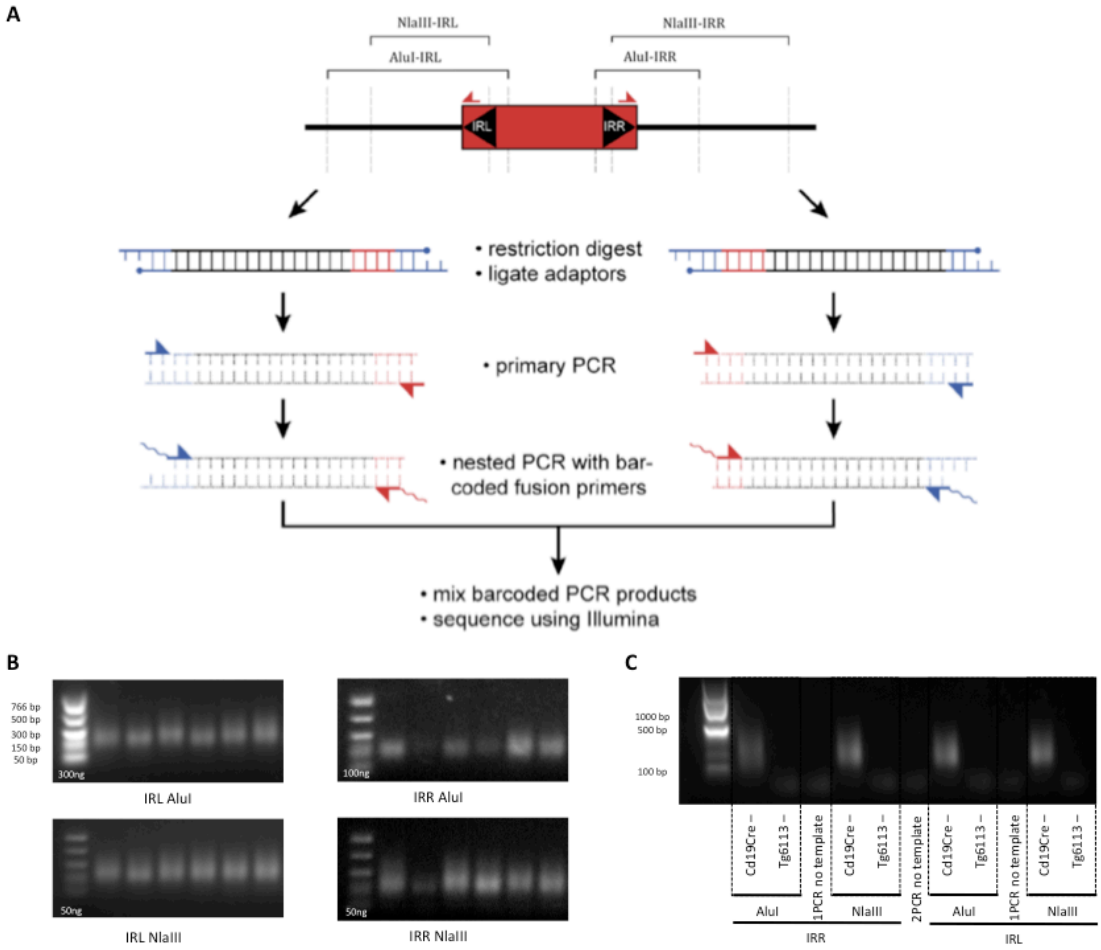


**Figure 23 Library preparation by LM-PCR. A** By LM-PCR 4 libraries are prepared for each tumor by cutting each tumor with 2 AluI and NlaIII, followed by adaptor ligation. In a next step, BamHI is used to cut *T2Onc2* concatameres on chromosome1, followed by a transposon enriching primary PCR. As the adaptor is composed of a 3'-C3spacer that blocks annealing of adaptor primers these can only anneal to primary transcripts amplified with the transposon specific primer. In a second, nested PCR primers introduce the sequence needed for sequencing and a barcode of 8 nucleotides. 192 barcoded libraries from 48 samples are pooled in equimolar concentrations and are deep sequenced on one flow cell lane of the HiSeq2000. **B** Examples of the 4 libraries prepared for 6 different tumors using AluI digestion and adaptors, NlaIII digestion and adaptors with IRL or IRR primers. **C** LM-PCR was performed on lymphoma gDNA of mice not carrying the *CD19-Cre* allele or not carrying the *Tg6113* transgene to test specificity of this method. As expected, without transposon, not product can be amplified. Surprisingly, without transposon mobilization (-*CD19-Cre*) libraries are produced that are similar to the ones produced in mice having the complete SB system active.

In summary, 5 sequencing libraries, each containing 48 samples prepared in 4 different ways (IRR-AluI, IRR-NlaIII, IRL-AluI and IRL-NlaIII) were loaded per sequencing lane of the Illumina platform. In total, the sequenced samples of the Myc-Cre-SB-T genotype include 175 lymphomas derived from transplanted animals, 7 lymphomas of bred animals, 7 samples of transplanted mice that are supposed to have developed from a tumor clone present in the donor, 5 samples of different infiltrated hematological tissues of the same mouse that developed lymphoma, 14 lymph nodes of donor mice. Additional, 15 samples of isolated B-cells from spleen of mice transplanted with a donor of the genotype Cre-SB-T, 4 lymphomas of transplanted mice with the genotype Cre-SB-T and 4 lymphomas of mice transplanted with a Myc-Cre-T donor were included as control samples.

A pre-requisite for this method to distinguish background mutations from common integration sites (CIS) is the right amount of sequencing depth, with at least 100000 sequence tags per tumor for statistically relevant analysis (Brett, 2011). In Figure 25 the summary of all the steps used to optimize and improve the sequencing is shown. On the left Y-axis the amount of available reads per sequencing library (blue bars) and the aligned reads (red bars) are depicted. On the right y-axis, the amount of demultiplexed aligneable reads per library of a sample is shown. In a first attempt SEQ1 (which comprises 48 samples and thus 196 libraries)(Seq1 GA2) and SEQ2 (Seq2 GA2) were sequenced each on a lane of the GA2, reading 75bp in single end. This method however did not give us the right amount of sequencing depth of 100000 reads per library needed. Moreover, as an internal control for the PCR method we used a Chi-squared test to see the overlap between the 4 libraries that were produced per sample and got less than 5% overlap. This suggested that both the PCR method and the sequencing depth needed to be optimized. Therefore, to increase the amount of reads available on each sequencing lane, the HiSeq 2000 sequencing platform was used, that allows

up to 8-fold more reads per lane than the GA2. The resequencing of SEQ2 (SEQ2 HiSeq) increased the amount of reads per sequencing lane as expected and, moreover, as 100bp single end reads were performed, also increased the amount of sequences that could be mapped to the genome (70%). LM PCR was modulated by changing the concentration and incubation time of the restriction enzymes, introducing clean up steps after every step to remove probable contaminants for the follow up step, changing polymerase, template concentration and PCR program. However, these steps did not change the appearance of the libraries, nor did it increase the overlap between the individual libraries per sample.

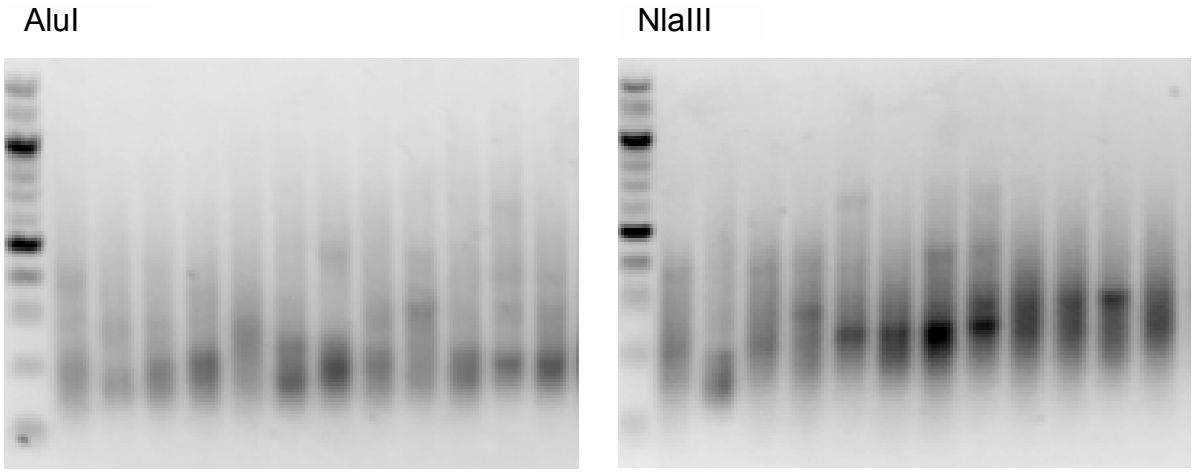


**Figure 24 LM-PCR fragment pools.** An aliquot of the LM-PCR product is analyzed by agarose gel electrophoresis to verify the quality of the sample. The above example is a typical result for the LM-PCR process. Products appear as published before as a low molecular weight smear (Berquam-Vrieze et al., 2011). DNA marker is the 100bp ladder (NEB).

The sequencing of these newly prepared libraries (seq1, seq2, seq3, seq4 and seq5) gave sufficient sequencing depth in 2 out of 5 sequencing libraries. The remaining 3 were suggested to have had a poor sequencing quality due to the low complexity of PCR products (all have a similar starting sequence due to the barcode and transposon primer) that resulted in low cluster formation. Augmenting the complexity of the DNA samples per sequencing lane by mixing the samples with unrelated samples of higher complexity resulted in optimized sequencing depth also for these libraries (Seq1 2013, Seq4 2013 and Seq5 2013).

In general, 30.6-70% of reads that were assigned to a specific barcode could be mapped to the mouse genome. As shown in **Figure 25**, each sequencing library could be optimized to fulfill the needed requirements for finding CIS.

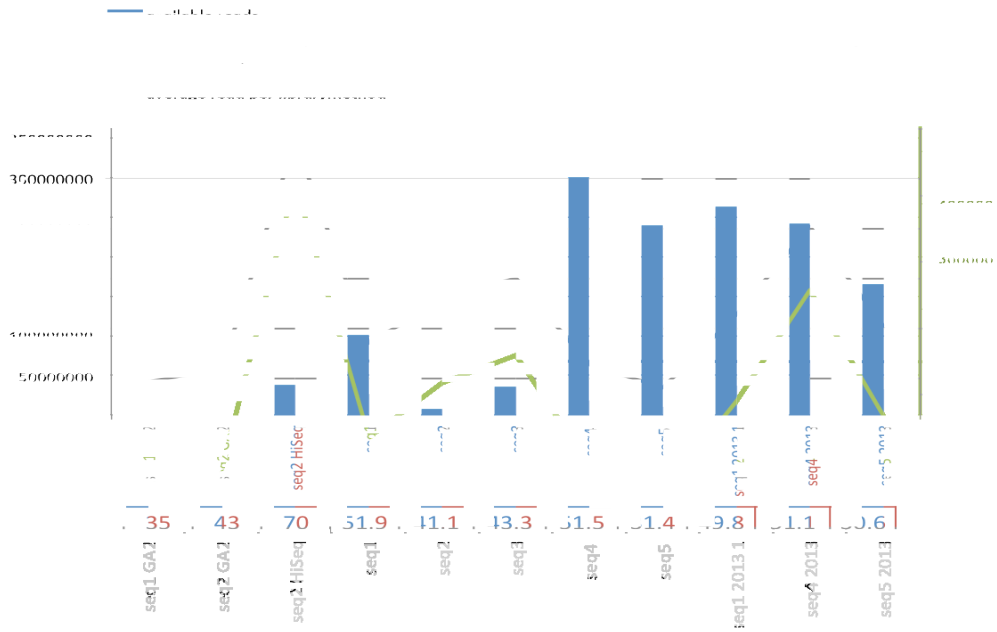


**Figure 25 Sequencing results.** During the process of optimizing the sequencing strategy we obtained different amounts of reads before (bars in blue) and after demultiplexing (bars in red) (left x-axis of the graph) from different samples sequenced on the GA2 or on the HiSeq (depicted in the y-axis of the graph). Libraries seq1 and seq2 were sequenced at first on the GA2, obtaining an insufficient coverage with less than 100000 reads per tumor (right x-axis). Resequencing of the same library on the HiSeq (seq2 HiSeq) resulted in optimal coverage with more than 500000 mapped reads per library method.

### Read analysis

To compare the data obtained from the sequencing applied the common criteria used in Sleeping Beauty screens with Vav-Cre, CD4-Cre and LcK-Cre leading to T-cell lymphoma (Berquam-Vrieze et al., 2011). As reported by these authors, we observed variations in the number of sequences obtained for each sample, probably due to different efficiency of LM-PCR amplification and/or complexity of tumors. The amount of sequences we obtained for each sample per LM-PCR method did not reflect a bias towards one primer set or one restriction enzyme performing better than the other. Another typical feature of SB screens is local hopping; indeed, we find 30-50% of the IS mapping to chromosome 1, which harbors the transposon-containing transgene. Moreover, it was described in the above study that a high amount of background insertion sites was present in the

sequencing data, which are defined by a low read count. Indeed, we find more than half of the insertion sites represented by a single sequencing read, which raises the need for a thorough filtering. In conclusion, our sequencing results are comparable to the study performed by (Berquam-Vrieze et al., 2011). Moreover, differences between the systems could arise because of different complexities of tumors in relation to the number of insertions they harbor. This was shown for the different model systems used in the previous study, in which SB-Cd4-Cre induced T-cell lymphomas had a much higher complexity than SB-Vav-Cre induced tumors (Berquam-Vrieze et al., 2011). Thus, the deep sequencing of more than 4 million sequences for 240 samples represented by 980 libraries overall seems to meet the prerequisites for further analysis.

**Clonal Insertion sites**

In our screen Sleeping Beauty mediated transposition induces insertions in the genome of Eμ-*myc* expressing B-cells. Most of the insertions should not cooperate with Myc in cellular transformation and should therefore not undergo positive selection. Insertions that provide a selective advantage, on the other hand, will most of the times persist in tumor cells that descend from this clone and will favor expansion of the clone. As SB mutagenesis is occurring continuously in B-cells, besides initiating insertions giving rise to tumors, background and sub-clonal events providing advantages in tumor maintenance are also present in the final tumor population. The selective advantage of the insertion in comparison to background mutations is believed to be proportional to the amount of tumor cells that carry an insertion site. In this sense we defined clonal insertion sites as those integrations that are present in the bulk of the sequenced fragments in a higher abundance compared to randomly occurring background mutations.

Three independent filtering systems for deriving clonal IS were applied (Table 2). The first, HM101, is based on the rationale that clonally expanded IS have a higher read count than background insertions; we set a threshold for IS to be represented by a read count of at least 0.1% of the total reads per library and of at least 1% of the maximal number of reads in a library as shown before (Brett et al., 2011). The second filtering method HM34, defines clonally expanded IS based on the consistent amplification in the 4 libraries of the LM-PCR prepared per sample and does not take into account the read count of the libraries. The third method, hereby termed BB, is the filtering for clonal IS performed by Benjamin Brett from the Dupuy work group. It defines clonally expanded insertion sites that have a read count >10% of the most frequently sequenced site in the corresponding IRL or IRR library. Details regarding the different filtering systems applied can be found in the Materials and Methods section.

HM101 is the most relaxed filtering method which finds most clonal IS, whereas the method HM34 and BB show similar numbers of IS but different numbers of genes hit by the IS as shown in Table 2.

**Table 2 Clonal Insertion sites.**

| | HM101 | HM34 | BB |
|---|---|---|---|
| **Clonal insertion sites** | 62698 | 22574 | 20404 |
| **Intragenic IS (RefSeq)** | 26638 | 16213 | 8782 |
| **Individual genes hit** | 8028 | 5379 | 3714 |

When considering the targeted loci, we find a higher fraction of gene centric IS with BB (23%) compared to HM101(12.8%) or HM34 (18.2%), although the mean number of IS per sample is lower (HM101: 261, HM34: 94, BB: 85) and correlates with the total number of clonal insertion sites found per filtering methods. This suggests that the IS are not distributed evenly between intragenic and intergenic regions in the different filtering methods and led us look at this in detail.

Sleeping Beauty based insertions are distributed evenly in the genome in a non selective population in contrast to tumor cells that acquire a selective advantage by an insertion site (Berquam-Vrieze et al., 2011; Brett et al., 2011). Thus, we expect to find integrations in protein coding genes to a higher extent than in intergenic regions with the different clonal filtering systems applied. To confirm this, we compare the clonal IS filtered for control and tumor samples to the distribution of intragenic and intergenic regions of the mouse genome (Figure 26).

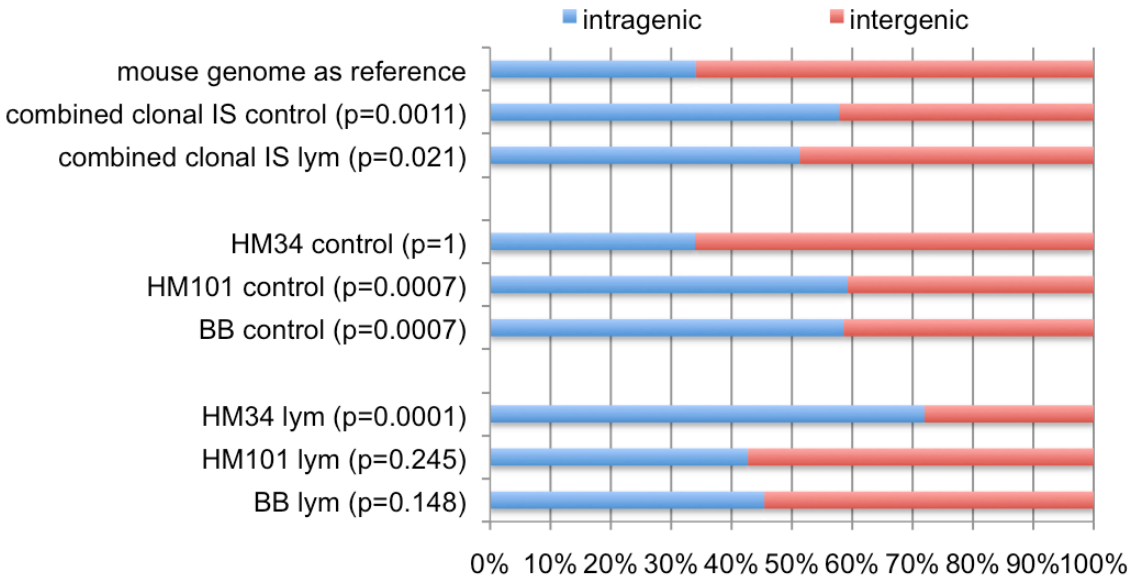


**Figure 26 Distribution of clonal IS** to intragenic and intergenic regions of the mouse genome. The clonal IS per filtering method was assigned to the intragenic or intergenic regions and its fraction to all insertion sites found in this method was calculated and is blotted for intragenic regions in blue and for intergenic regions in red. To see independence of the distribution of IS we compared the observed percentage to the one expected from intragenic to intergenic regions of the mouse genome. By Fisher's exact test these distributions are compared and p-values are assigned (in brackets next to the clonal insertions obtained from different filtering methods).

Surprisingly, looking at clonal insertion sites compiled from all methods we find a slight enrichment in genes for control and tumor cohort, but only for the control cohort this difference is significant. When we further look at the different filtering methods, only the HM34 filter, which takes into account PCR method accuracy rather than applying a threshold of reads to be reached, reflects the expected distribution of non-randomness in tumor and randomness of IS in control samples. This suggests that the control cohort is enriched for clonal insertion sites and/or that the clonal insertion sites defined by different methods include a high number

of background insertions.Comparison between the 9883 independent genes found with all filtering method suggests that 20% are consistently found by all three methods (Figure 27).

Moreover, 44.7% of genes found were selected by the filters HM101 and HM34, suggesting that many of the defined clonal IS (0.1% of the reads) were amplified by 3 out of 4 LM-PCR reactions, excluding possible PCR bias from samples that are unspecifically amplified and thus show low overlap between the different LM-PCR-libraries prepared. The BB and HM101 filters both define background insertions as a fraction of IS that was below a threshold of reads taking into account the size of the libraries (HM101 0.1%, BB 10%). Thus we expected the more stringent filtering of BB to overlap completely with the less stringent filtering of HM101 but only 28% of the genes overlap. This might relate to the fact that the BB filter was applied after the reads of libraries -prepared with the two different restriction enzymes- were pooled according to usage of the IRL and IRR primers. Instead, the HM101 filter considers every library independently. Summarizing, we can say that different filter methods exclude different background insertions. However, more than 40% of the clonal insertion sites were found by all three methods, suggesting that these clonal insertion sites include less false positive background insertions.

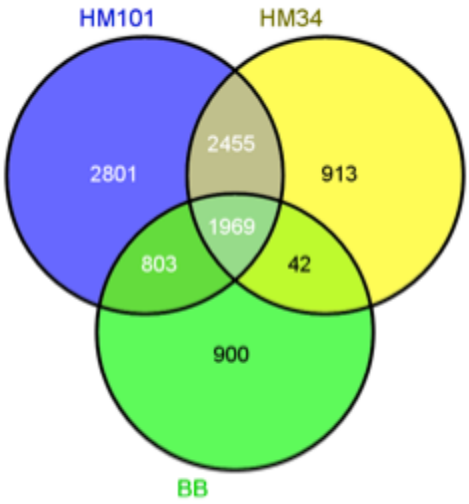


**Figure 27** Overlap of genes found clonally expanded in the 3 different filtering methods. 1969 genes were found in total by the three calling methods (HM101, HM34 and BB), with the overlaps indicated in the Venn diagram.

Based on the above, further evaluation on the quality of the filtering is needed as control and tumor cohorts do not reflect the expected distributions in the genome. We take advantage of the transplantation of HSPCs of donor 64813, 82091 and 73948 that gave rise to tumors in recipients with less than 35 days, all showing the same latency period (Figure 12). Therefore we assumed that clonal events were already present in the donor mouse that has been transplanted, thus serving as positive control for clonally expanded sites. Additionally, infiltrated tissue from different sites (spleen, axillary, mesenteric/inguinal and cervical lymph nodes) of mouse 112979 were collected and processed independently. Also in this case we suggest that the same driving genes are found in different infiltrated tissue of the same mouse, thus defining clonal events and not random distribution in the mouse genome. Overlaps were assessed for all of the filtering methods and are summarized in Supplement 1 and depicted in Figure 28.

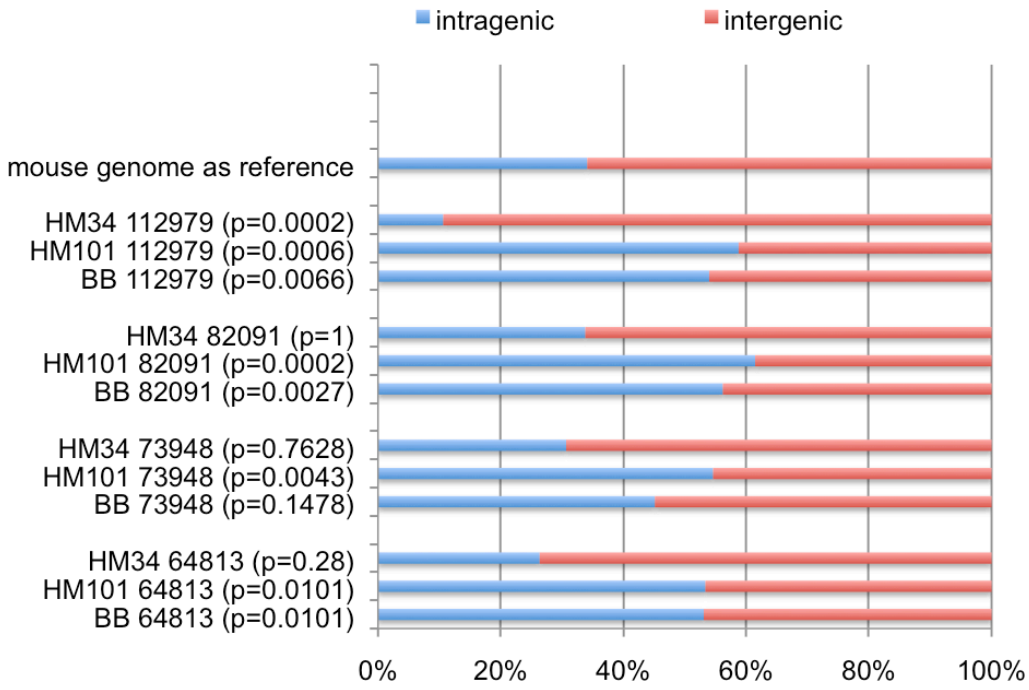


**Figure 28 Overlap between clonal cohorts.** The distribution of intragenic and intergenic regions of the mouse genome were compared to the distribution of clonal IS in different transplanted cohorts that were defined by 3 filtering methods. Cohort 112979 defines different infiltrated tissue of the same mouse that developed lymphoma. Cohorts 82091, 73948 and 64813 were transplanted with donor HSPCs that had already undergone clonal selection and reflects expansion of this clone in different recipients. The clonal IS per filtering method was assigned to the intragenic or intergenic regions and its fraction to all insertion sites found in this method was calculated and is blotted for intragenic regions in blue and for intergenic regions in red. To see independence of the distribution of IS we compared the observed percentage to the one expected from intragenic to intergenic regions of the mouse genome. By Fisher's exact test these distributions are compared and p-values are assigned (in brackets next to the clonal insertions obtained from different filtering methods).

The methods HM101 and HM34 find significant differences in the distribution of insertion sites compared to the mouse genome, whereas method HM34 fails to do so in all of the clonal cohorts, suggesting that an overlap between the four different LM-PCR amplification methods does not discriminate background insertions from clonally amplified insertions. However, also with the filters HM101 and BB we find not complete overlap. This can account either for different subclonal populations selected in the cohorts, as suggested by immunophenotying and/or by a high amount of background insertion sites present after filtering. However, when a ranking is performed according to the number of lymphoma samples that carry a certain gene-centric clonal IS, the top 158 genes are consistently found by either

filtering method, suggesting good validity of all filtering methods for highly abundant IS (Figure 29).

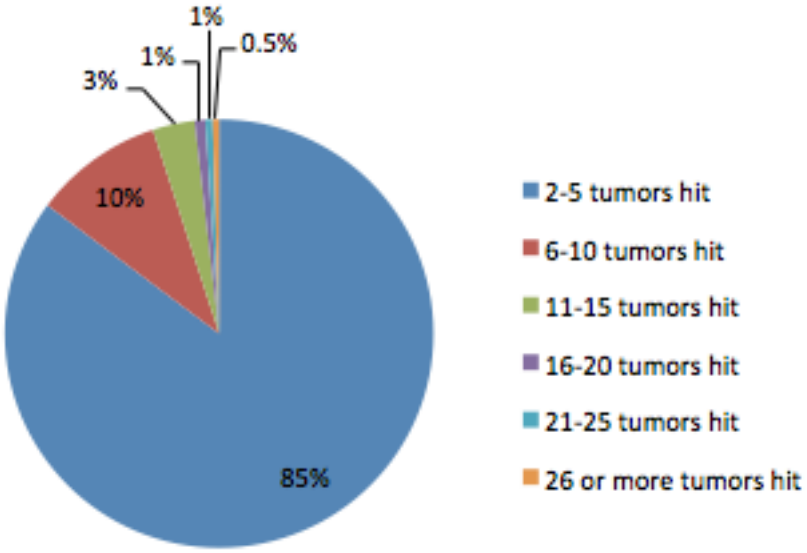


**Figure 29 Genes defined to be clonally expanded in the lymphoma cohort.** Ranking of genes was performed according to the frequency they were found clonally amplified in lymphomas (%). The clonally expanded gene centric IS that were hit in more than 10 tumors made up only 5% of all clonally amplified genes. These are consistently selected with either filtering method applied to reduce background insertions. Moreover, 85% of all clonal IS were hit in a less than 5 lymphomas.

Taken together, this suggests that we cannot rule out the presence of false positive clonal IS in any of the methods, however, we cannot rule out that these are real subclonal events. Thus in expense of probable false positives we include these and combine all clonal insertion sites to be able to find also rare insertions that are inducing a selective advantage in Myc-induced HSPCs. Moreover, the next layer of statistical analysis to define common insertion sites will further deplete background insertions. Thus, 72338 clonal IS for the lymphoma and 15786 for the control cohort which refer to 23.839 unique loci are used for further analysis.

### Common Insertion sites

Common insertion sites (CISs) are regions in the genome hit by clonal insertions sites in different tumors at higher than random frequency. To identify CISs involving more than two independent tumors, we used three different strategies:

I. Monte Carlo simulation defines CIS by comparing a set of simulated random integrations in TA dinucleotides to real integrations (hereafter mcCIS). This statistical analysis was performed with the Integrated Analysis System (Brett et al, 2011) and was a follow up on the filtering of clonal insertion sites (BB filter) performed by Benjamin Brett. mcCIS found also in the control cohort and in the lymphoma cohort (3) were subtracted, leaving 69 mcCIS assigned to the lymphoma cohort. Of these 75% are found in gene-centric regions (hitting 64 genes), whereas 17 mcCIS are mapping to non-annotated regions of the genome.

II. The gene centric analysis uses a Chi-squared test to compare the distribution of IS in the TA dinucleotides in a gene-associated region (gene size with an additional 1kb or 10kb upstream region) to the distribution of all TA dinucleotides hit by the transposon (forth on gCIS). The addition of a region upstream of the gene is used to include possible up-regulating events that occur in the promoter region of a gene. Including a promoter size of 1kb, 151 gCIS were found in the lymphoma cohort and 15 of them were also found as part of the 30 gCIS found in the control cohort. It is important to state that 80% of the genes found in the control cohort map to chromosome 1, where the Tg6113 transgene is located, and reflect local hopping. When increasing the promoter size to 10kb 165 gCIS are defined in the lymphoma cohort and 29 in the control group, again 80% mapping to chromosome 1. Subtracting the gCIS found in control samples from tumor samples leaves in total 155 gCIS, 132 of which were found using 1kb or 10kb promoters that are listed in Supplement 2. Thus 92 % of the gCIS including a 10 kb promoter are found already when using the 1kb promoter window, suggesting that SB has a preference of placing insertions inside or in close proximity (1kb) of genes and can be explained by the preference of SB to target the intronic sequence in 96% of all genes hit (Yant et al., 2005).

III. The third way of defining CIS involved enrichment analysis (hereafter eCIS). The combined clonal IS (BB, HM101 and HM34) of Myc-Cre-SB-T lymphomas were compared to control B-cells with the genotype Cre-SB-T. By doing this, we could exclude IS that occur due to a bias in the model system used, as we compare in both cases samples of the B-cell lineage. Moreover, as the control cells do not over express Myc, we enrich for clonal IS that need *myc* as driving oncogene. The analysis included, like for the previously described gCIS method, two different sized promoter regions of 1kb and 10kb. Using a cutoff for enrichment of 2.5, 273 genes were found within a 1kb region, whereas 318 genes are found including a 10kb promoter region.

CIS were listed in Supplementary table 2 in different spreadsheets for I, II and III.

In the latter, two statistical methods two different promoter sizes are used. By 10 fold increase of the promoter size, only an 1.2 fold increase in the number of CIS was observed suggesting that SB integration favor nearby regions to deregulate genes.

Control CIS and CIS found in the lymphoma cohort, control and lymphoma gCIS as well as eCIS with an enrichment score of a) at least 5, b) at least 2.5 and c) less than 1 that define background insertions were depicted in a circster plot that positions all the integration sites in concentric circles representing the mouse genome in Supplement 2.

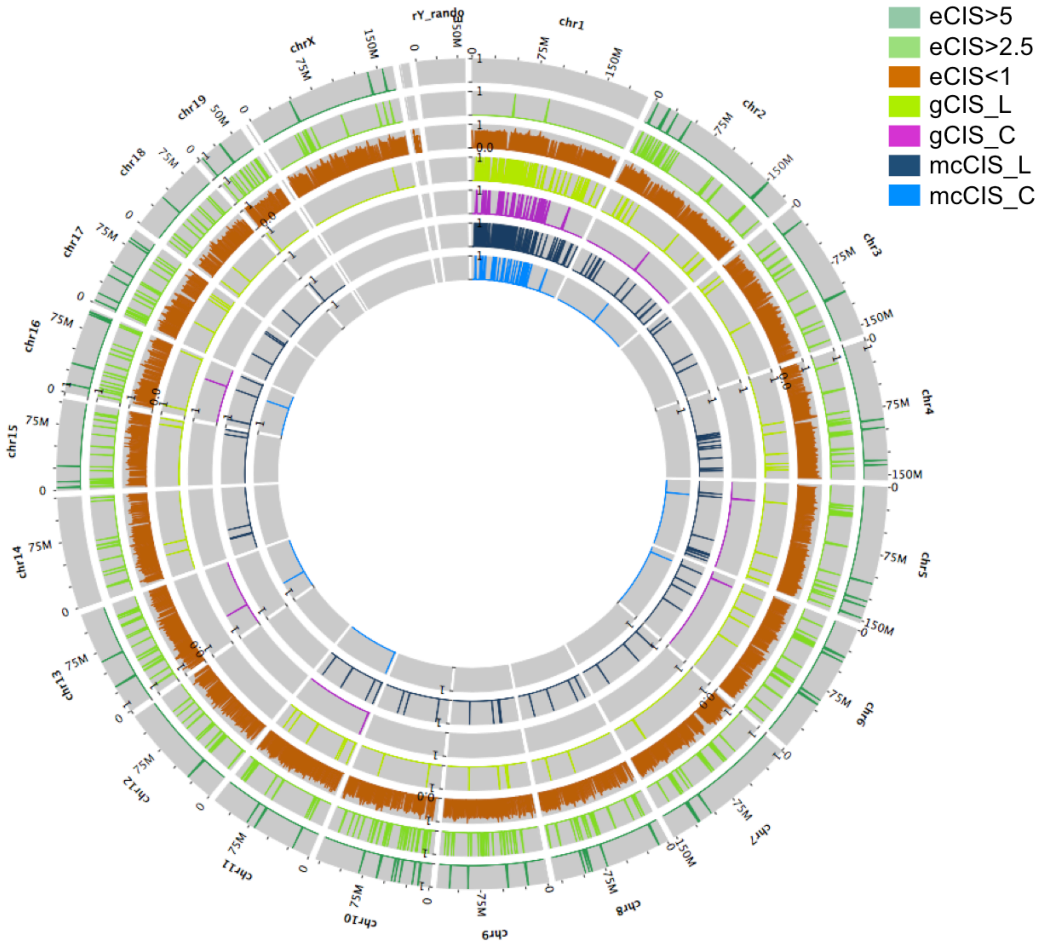


**Figure 30 Visualization of common insertion sites.** The information regarding the location of common insertion sites obtained by different methods in control and lymphoma cohort was uploaded to galaxy and visualization was performed using the circster tool. The mm9 genome is depicted as a circle, dived into the different chromosomes. Each circle depicts one analysis: mcCIS_C and mcCIS_L define the genesets found in the control cohort and lymphoma cohort, respectively by Monte Carlo simulation. gCIS_C and gCIS_L depict the regions found by the gene centric statistical analysis in the control and lymphoma cohort, respectively. The eCIS<1 defines the regions not enriched and suggest background insertions, whereas eCIS>2.5 and eCIS>5 depict the regions found by using enrichment analysis with a score at least 2.5 or 5, respectively.

gCIS mapped in 63.9% of the cases to chromosome 1, suggesting this method to be not selective enough against background insertion from local hopping. With the Monte Carlo or enrichment method on contrary, less than 1% of the mcCIS were mapped to chromosome 1. This reduction is observable in **Figure 30** where the non-enriched IS (eCIS<1) are covering the whole genome, whereas in the eCIS>2.5 a lot of IS overlap with the mcCIS and gCIS genes found in the lymphoma set. Apart from the high frequency of IS on chromosome 1 for all but the eCIS>2.5 and eCIS>5 datasets, the CIS confirm even distribution over the mouse genome shown in mouse liver tissue (Yant et al., 2005). Moreover, except for methods gCIS and mcCIS, which do not find CIS on chromosome 15, all autosomes harbored hits,

suggesting no biased insertion of the SB system in our settings. Interestingly, in the control set of mCIS (referred to a CIS_C in Figure 30) and gCIS, we can find CIS that are not found in the lymphoma sets of these groups (CIS_L and gCIS_L, respectively) nor in the eCIS>2.5 and eCIS>5 cohort as found by overlapping the latter with the CIS found in the control cohorts. This suggests, that these B-cells might be subjected to a clonal selection already. However, no signs of disease were observed in the mice used for this cohort at the time of sampling, although it is worthwhile mentioning that 2 mice of this cohort did develop lymphoma after a median latency of 429 days after transplantation.

To understand in detail which genes define putative cooperating partners of Myc in lymphomagenesis, the datasets of the three analysis methods - excluding genes mapped to chromosome 1 and genes found in the control cohorts - were overlapped and the distribution of the 340 individual CIS found are depicted in the Venn diagram in **Figure 31** and Supplementary table 2.

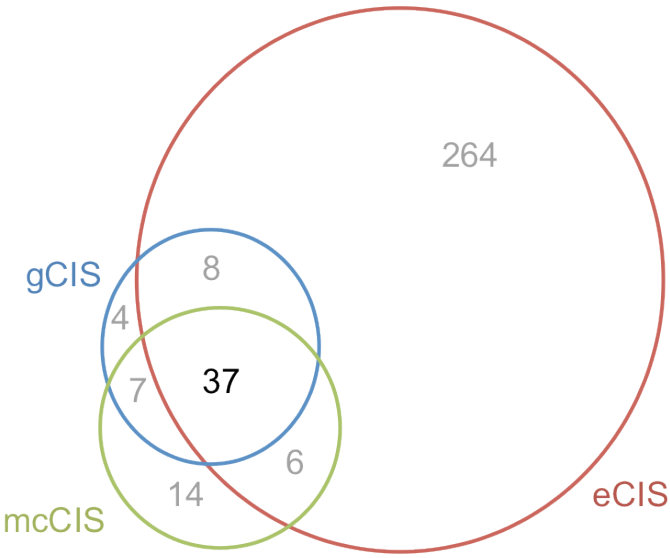


**Figure 31 Venn diagram overlapping genes found in the gene-centric analysis (gCIS), the enrichment method (eCIS) and the genes located in the regions defined statistically overrepresented by Monte Carlo simulation (mcCIS).** Genes found in control cohorts and/or mapping to chromosome 1 were excluded.

Of all the genes, 37 (10.85%) were found by all three statistical methods, including the following genes:

| | | |
|---|---|---|
| Armc1 | Gtf2ird1 | Snx19 |
| BC052040 | Hacl1 | Sos1 |
| BC061194 | Igsf21 | Tead1 |
| Bmi1 | Kdm2a | Tekt5 |
| Braf | Kdm4c | Thada |
| Cdk17 | Mll1 | Trp53 |
| Cdkn2a | Map3k5 | Tsc1 |
| Dyrk1a | Mef2b | Ttll11 |
| Eif2s2 | Pik3r5 | Zeb2 |
| Ets1 | Pou5f1 | |
| Flt3 | Ppp3r1 | |
| Gata2 | Rasgrf1 | |
| Gatad2b | Sap130 | |
| Gfi1b | Scmh1 | |

Known collaborators of Myc in lymphomagenesis in the Eμ-*myc* model were included, like Bmi-1 (Haupt et al., 1991; Jacobs et al., 1999) and Gfi1b, a paralog of Gfi1 (Scheijen et al., 1997), that were originally discovered in retroviral screens. Trp53 (Eischen et al., 1999) and Cdkn2a (Schmitt et al., 1999) are part of the main axis of tumor suppression in Eμ-*myc* mice and disruption is observed in up to 50% of lymphomas that develop spontaneously. Moreover, genes that have a functional implication with *myc* were found as shown by two examples: *Pou5f1*, encodes for Oct4, one of the core factors inducing pluripotency, and together with Myc induces a cancer-stem cell like signature in ES cells (Kim et al., 2010). Tsc1, a negative regulator of the mTOR pathway is a Myc-repressed target that has a function in Myc-induced transformation (Ravitz et al., 2007; Wall et al., 2013). This suggests that we could find known collaborators but also new interesting targets. However, this short list does not include genes involved in the p53 independent apoptotic response elicited by the BH3-only protein, which is deregulated in up to 50% of cases by upregulation of Bcl-2 (Eischen et al., 2001b; Strasser et al., 1990). This

might account for differences in the DNA structure, as SB transposase was shown to have preference for bent sequences (Liu et al., 2005).

25 genes were not overlapping with the eCIS>2.5 dataset, constituting 7.33% of all independent genes found, which are listed here:

| gCIS_L_10 | gCIS_L_10 and CIS_L_10 | CIS_L_10 |
|---|---|---|
| Stat5b | Gch1 | Fam18a |
| Abca5 | Sp1 | Nubp1 |
| Yaf2 | Sulf2 | Ctps |
| Terf2 | Znfx1 | Slfnl1 |
| | Pag1 | Lhx2 |
| | Aadacl3 | Gm13476 |
| | Fli1 | E130114P18Rik |
| | | Nfia |
| | | Fam188b |
| | | Frmd3 |
| | | Jak1 |
| | | Sdk1 |
| | | Cacna2d3 |
| | | Cux1 |

For further analysis, the genes found statistically represented were pooled and subsets were defined to analyze the consistency of the data obtained:

- CIS includes all the 338 non-synonymous CIS found by the Monte Carlo simulation, gene-centric analysis and enrichment analysis. From the previously found 340 CIS, two could not be annotated by IPA (details in Materials and Methods).
- coreCIS defines 37 non-synonymous CIS found by all three methods suggesting these 10.7% to be highly probable CIS.
- looseCIS, is defined by the eCIS2.5 set contributing 92.6% to the whole CIS but does not contain the coreCIS. It possibly contains a high rate of false positive CIS but allows us to find rare events that might be excluded with the gene-centric and the Monte Carlo-based methods. The rationale to include this gene-set was a) it was obtained using all clonal insertion sites selected by the different filtering systems, whereas the other methods to define CIS were based

only on clonal insertion sites selected by the stringent BB filter, thus possibly counter-selecting for rare events present in the dataset; and b) the enrichment analysis defines CIS as IS that are overrepresented in tumor samples compared to control samples. This allows for experimental bias as control and tumor samples were obtained and prepared in the same way. Moreover, this possibly excludes for IS that are hotspots for SB in the system.

## Characterization of CIS

### Cross-Species Comparison

Mouse models define an ideal experimental approach in which candidate loci observed in human diseases or in forward genetic screens (such as CIS loci in this SB screen) can be functionally analyzed by reverse genetics. A database that collects CIS of insertional mutagenesis screens has been created in the Retrovirus and Transposon tagged Cancer Gene Database (RTCGD). In the last few years increasingly better technology made it possible to perform whole exome sequencing of many human cancer samples. Somatic mutations found are collected in databases like The Cancer Genome Atlas (TGCA) and the Catalogue Of Somatic Mutations In Cancer (COSMIC). For the latter, a consensus gene set was defined, that describes mutations found with a strong probability and are suggested to contain drivers of tumorigenesis in human cancer. Furthermore, lately published data on somatic mutations in Burkitt lymphoma, the human cancer that develops due to translocation of *myc* and the *IgH* (Giulino-Roth et al., 2012; Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012) were compiled and overlapped to the CIS found in the Eμ-*myc* mouse.

Overlapping the CIS found in this study with the databases indicates the validity of the screen in several ways and is shown in Table 3. First, it indicates how many known cancer genes are found in the coreCIS compared to the looseCIS. Second, it

defines the relevance of the screen for human tumorigenesis, possibly allowing not known common denominators to be used for future therapeutical implications.

Regarding the RTCGD, we find 35 overlapping CIS (p=6.80E-14) of which half were found in screens using transposons and/or Myc-induced tumors including the RosaSB;*T2Onc2* mouse model (Dupuy et al., 2005) and a M-MuLV screen in the Eμ-*myc* model (Berns A), suggesting a high similarity of the CIS found in this screen with others performed with comparable systems. To understand whether the CIS found also have an implication in human cancer we overlapped the CIS with the TGCA, a database collecting somatic mutations found in human cancer types, and found 211 common genes. These 62.4% of the CIS were overlapping with the TGCA dataset, which is statistically significant (p>1E-15) and reflects the efficiency of the screen to identify functionally relevant mutations in human cancer. However, the genes defined in the TGCA are not curated genes and contains passenger and driver mutations. To further understand whether the CIS defined here also correspond to driver genes in humans, we cross-compared with the cosmic cancer census and also got highly significant overlap. Interestingly, the coreCIS and looseCIS subsets perform similarly well in overlapping with the genes in the different databases. This suggests that all methods to define CIS are robust and can distinguish background/passenger mutations from CIS that statistically significant overlap with known CIS in mice or tumor drivers in humans, validating the findings of this screen. Comparison to the Burkitt lymphoma mutational signature shows no statistically significant overlap with the coreCIS subset, whereas total CIS and looseCIS do show statistical significant overlap. This suggests that CIS found in the Eμ-*myc* mouse model reflects genetic lesions found in human Burkitt lymphoma only to a certain extent, as reported previously (Schmitt and Lowe, 2002). Moreover, as described in the introduction, the Eμ-*myc* mouse serves as a

model for finding tumor suppressor genes (TSG) and oncogenes (OG) deregulated in Myc-driven tumors in different tumor types.

**Table 3 Cross-comparison.**

| gene sets previously reported | gene sets obtained in this study | overlapping genes | p-value (binomial probability distribution) |
|---|---|---|---|
| RTCGD (480) | CIS (338 genes) | 35 | 6.80E-14 |
| RTCGD_MYC_SB (94) | | 17 | 5.80E-14 |
| TGCA (4820) | | 211 | <1e-15 |
| COSMIC cancer gene sensus (503) | | 30 | 4.80E-10 |
| BL_mutations (3617) | | 86 | 9.80E-05 |
| OG_Davoli (998) | | 35 | 9.90E-06 |
| TSG_Davoli (1000) | | 41 | 3.10E-08 |
| RTCGD (481) | coreCIS (36 genes) | 10 | 3.80E-10 |
| RTCGD_MYC_SB (94) | | 4 | 6.50E-07 |
| TGCA (4820) | | 29 | 4.60E-14 |
| COSMIC cancer gene sensus (503) | | 7 | 2.00E-06 |
| BL_mutations (3617) | | 7 | 0.29 |
| OG_Davoli (998) | | 9 | 5.50E-06 |
| TSG_Davoli (1000) | | 6 | 1.50E-03 |
| RTCGD (482) | looseCIS (277 genes) | 22 | 2.20E-07 |
| RTCGD_MYC_SB (94) | | 12 | 8.10E-10 |
| TGCA (4820) | | 174 | <1e-15 |
| COSMIC cancer gene sensus (503) | | 20 | <1e-15 |
| BL_mutations (3617) | | 75 | <1e-15 |
| OG_Davoli (998) | | 24 | 2.20E-03 |
| TSG_Davoli (1000) | | 32 | 2.20E-06 |

To further define this concept, we took advantage of a recent study that analyzed the patterns of mutational signatures in tumors and predicts the likelihood that any individual gene functions as a tumor suppressor (TSG) or oncogene (OG) (Davoli et al., 2013). Indeed, we find 71 genes overlapping significantly with suggested TSGs and OGs in similar frequencies (35 OG to 41 TSG), again validating the findings.

Summarizing this part, there is strong statistical overlap between the CIS defined in this study to previous insertional mutation derived CIS, especially in Myc-driven tumor models, suggesting robustness of the experimental and bioinformatic methods used. Moreover, part of the genes found in this SB-induced mutagenesis screen are mutated in human cancers and although we cannot reconcile mutations

occurring in Burkitt lymphoma completely, we still have strong indications for the importance of these CIS in Myc-driven human malignancies.

### Common pathways and functions of genes cooperating with Myc

Inactivation of TSGs and activation of OGs is a central dogma in tumorigenesis. Lately however, the view has changed from a gene-centric point of view to a more global perception, taking into account different signaling pathways that need to be deregulated in order for transformation to occur. In this sense, all genes involved in a pathway can when mutated deregulate the effect of a signaling cascade, leading to an selective growth advantage and to the emergence of cancer, as suggested by Vogelstein (Vogelstein et al., 2013). He designated 12 major pathways that, when deregulated, lead to growth advantage by increasing cell survival, changing cell fate or deregulating genome maintenance (Figure 32).

To understand the relationship of the CIS to each other and to find common canonical pathways, the CIS gene set was analyzed using different online analysis tools (DAVID, PathJam, String) that derive information from curated pathway databases including KEGG, Biocarta, Reactome and The Biological Biochemical Image Database (BBID).

The top pathways that were found with all of the analysis tools are listed in Table 4 and in Supplementary table 4, and show a ranking according to the number of CIS that were present in the pathway.


**Figure 32 The 12 major pathways deregulated in cancer according to Bert Vogelstein (from** Vogelstein et al., 2013).

The pathway harboring most CIS (22) was the MAP-kinase pathway. As depicted in Figure 33, the main signaling branch hit is the classical Ras-signaling. CIS that mapped there are Rras2, a Ras related member of GTPases, direct positive modulators of Ras-activity (Sos1, Rasgrf, Rasgrp), the negative modulator Nf1 and downstream Braf. 90% of lymphomas contain a CIS that is found in this central signaling branch, suggesting this to be the main pathway deregulated by SB mediated mutagenesis. The Ras-MAP kinase signaling pathway is highly intertwined with CIS involved in PI3K and STAT signaling (Supplementary table 4). Growth factor signaling pathways were also highly represented by CIS, like PDGF and EGF signaling pathways. This again represents Ras-MAP-kinase signaling as ligand transmits these survival signals through this classical Ras-Raf-MEK-ERK cascade that in turn induces transcription of genes enhancing proliferation and differentiation. Interestingly, cells that express Myc are highly depend on mitogenic signals and induce an apoptotic response when it is missing (Harrington et al., 1994). Thus, the Ras-MAP-kinase signaling pathway cooperates in

lymphomagenesis, by counteracting Myc-induced apoptosis as demonstrated in Eμ-*myc* mice that were null for *Ksr1*, a central scaffold protein important upstream of Ras (Gramling and Eischen, 2012).

| Category | Term | CIS count | % | PValue | CIS ID | Benjamini | also in STRING | also in PATH-JAM |
|---|---|---|---|---|---|---|---|---|
| KEGG_ | MAPK signaling pathway | 22 | 6.61 | 2.58E-09 | TRP53, PRKCA, BRAF, NF1, PPP3R1, CACNB2, NFKB1, CACNB4, GNG12, CACNA2D3, MAP4K3, MAP3K5, MAP3K4, RASGRP3, RPS6KA2, RASGRF1, RASGRP1, SOS1, RRAS2, MAP3K8, TRAF6, GADD45B | 2.68E-07 | 1 | 1 |
| PANTHER_ | PDGF signaling pathway | 21 | 6.31 | 4.33E-08 | PRKCA, ERG, VAV3, BRAF, STAT5B, ARHGAP26, ITPR2, PDPK1, MAP3K4, FLI1, TIAM2, GAB2, RPS6KA2, ETS1, SOS1, RRAS2, GAB1, JAK1, OPHN1, PIK3R5, ARHGAP10 | 3.29E-06 | 0 | 0 |
| KEGG_ | Pathways in cancer | 19 | 5.71 | 7.72E-06 | PRKCA, TRP53, COL4A1, BRAF, MSH3, FLT3, STAT5B, FOXO1, NFKB1, BCL2L1, IGF1R, CCDC6, CDKN2A, SOS1, JAK1, MDM2, PIK3R5, RUNX1, TRAF6 | 2.68E-04 | 1 | 1 |
| PANTHER_ | Integrin signaling pathway | 13 | 3.90 | 2.34E-02 | COL4A1, BRAF, TLN2, ARHGAP26, ITGA9, MAP3K5, MAP3K15, MAP3K4, NAV2, RRAS2, SOS1, COL8A1, ARHGAP10 | 3.03E-01 | 0 | 0 |
| KEGG_ | Focal adhesion | 12 | 3.60 | 5.52E-04 | PRKCA, ITGA9, IGF1R, PDPK1, COL4A1, VAV3, CCND3, BRAF, TLN2, RASGRF1, SOS1, PIK3R5 | 7.15E-03 | 1 | 1 |
| PANTHER_ | EGF receptor signaling | 12 | 3.60 | 1.64E-03 | PRKCA, MAP3K5, MAP3K4, MAP3K15, GAB2, | 4.06E-02 | 0 | 0 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | pathway | | | | BRAF, SOS1, RRAS2, STAT5B, GAB1, PIK3R5, RASA3 | | | |
| KEGG_ | Chronic myeloid leukemia | 11 | 3.30 | 5.49E-07 | TRP53, CDKN2A, GAB2, BRAF, SOS1, STAT5B, MDM2, NFKB1, PIK3R5, BCL2L1, RUNX1 | 2.86E-05 | 1 | 1 |
| PANTHER_ | Ras Pathway | 10 | 3.00 | 6.53E-04 | PDPK1, MAP3K4, CDKN2A, BRAF, TIAM2, ETS1, RPS6KA2, SOS1, RRAS2, EXOC2 | 2.45E-02 | 0 | 0 |
| KEGG_ | Jak-STAT signaling pathway | 10 | 3.00 | 1.15E-03 | CCND3, PRLR, OSMR, SOS1, STAT5B, JAK1, CSF2RB, PIK3R5, BCL2L1, GHR | 1.32E-02 | 1 | 1 |
| PANTHER_ | Interleukin signaling pathway | 10 | 3.00 | 3.56E-02 | PDPK1, MAP3K4, BRAF, RPS6KA2, SOS1, RRAS2, STAT5B, FOXO1, CSF2RB, FOXJ3 | 2.92E-01 | 0 | 0 |
| KEGG_ | Prostate cancer | 9 | 2.70 | 1.43E-04 | TRP53, IGF1R, PDPK1, BRAF, SOS1, FOXO1, MDM2, NFKB1, PIK3R5 | 2.97E-03 | 1 | 1 |
| KEGG_ | Neurotrophin signaling pathway | 9 | 2.70 | 1.70E-03 | TRP53, MAP3K5, BRAF, RPS6KA2, SOS1, GAB1, NFKB1, PIK3R5, TRAF6 | 1.60E-02 | 1 | 1 |
| KEGG_ | Chemokine signaling pathway | 9 | 2.70 | 1.29E-02 | PARD3, VAV3, BRAF, TIAM2, SOS1, STAT5B, NFKB1, PIK3R5, GNG12 | 8.07E-02 | 1 | 1 |
| KEGG_ | Regulation of actin cytoskeleton | 9 | 2.70 | 3.30E-02 | ITGA9, VAV3, CHRM3, BRAF, TIAM2, SOS1, RRAS2, PIK3R5, GNG12 | 1.26E-01 | 1 | 0 |
| KEGG_ | Glioma | 8 | 2.40 | 1.02E-04 | PRKCA, TRP53, IGF1R, CDKN2A, BRAF, SOS1, MDM2, PIK3R5 | 2.65E-03 | 1 | 1 |
| PANTHER_ | Insulin/IGF pathway-protein kinase B signaling cascade | 8 | 2.40 | 1.01E-02 | IGF1R, RALGAPA1, PDPK1, TSC1, FOXO1, MDM2, PIK3R5, FOXJ3 | 1.76E-01 | 0 | 0 |
| BIOCARTA | MAPKinase Signaling Pathway | 8 | 2.40 | 2.31E-02 | MAP4K3, MAP3K5, MAP3K4, SP1, BRAF, RPS6KA2, MAP3K8, NFKB1 | 9.39E-01 | 0 | 0 |
| PANTHER_ | PI3 kinase pathway | 8 | 2.40 | 4.22E-02 | PDPK1, CCND3, SOS1, RRAS2, FOXO1, JAK1, PIK3R5, FOXJ3 | 3.05E-01 | 0 | 0 |
| KEGG_ | Non-small | 7 | 2.10 | 3.00E-04 | PRKCA, TRP53, | 5.18E-03 | 1 | 1 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | cell lung cancer | | | | PDPK1, CDKN2A, BRAF, SOS1, PIK3R5 | | | |
| KEGG_ | Acute myeloid leukemia | 7 | 2.10 | 4.04E-04 | BRAF, FLT3, SOS1, STAT5B, NFKB1, PIK3R5, RUNX1 | 5.98E-03 | 1 | 1 |
| KEGG_ | Pancreatic cancer | 7 | 2.10 | 1.41E-03 | TRP53, CDKN2A, BRAF, JAK1, NFKB1, PIK3R5, BCL2L1 | 1.46E-02 | 1 | 1 |
| KEGG_ | Phosphatidylinositol signaling system | 7 | 2.10 | 1.75E-03 | PRKCA, DGKG, PI4KA, SYNJ2, PIK3R5, DGKI, ITPR2 | 1.50E-02 | 1 | 1 |

**Table 4 Canonical pathway enrichment.** The table lists the CIS enriched pathways (column 2: name of the pathway) as defined by the DAVID tool. It lists the pathway databases (column 1: Category), states the number of CIS that map to pathway (column 3: CIS count), lists the % of CIS compared to all genes of the pathway (column 4: %), lists the CIS mapped to the pathway (column 6: CIS), indicates the p-value (column 5: p-value), a corrected p-value that takes into account the false discovery rate (column 7: Benjamini) and shows whether the pathway was also found by another analysis method (1 meaning present, 0 meaning absent).

Another strong implication, that the CIS found are good candidates for tumor suppressive and oncogenic mechanisms deregulated in tumor is the significant enrichment in pathways known to be deregulated in cancer. Signaling pathways specific for different types of tumors like CML (Vaqué et al., 2005), prostate cancer (Gurel et al., 2008), glioma (Shindo et al., 1993), non-small-lung cancer (Iwakawa et al., 2013), AML (Delgado et al., 2013) and pancreatic cancer were mapped to the CIS found. Of notice is that the shared CIS between the deregulated pathways of different cancer types enriched in the screen are Myc, p53 and Braf deregulation, and all but pancreatic carcinoma include Sos1 as deregulated CIS (Table 4). This suggests that p53 and Ras-Braf signaling are of central role in Eμ-myc lymphoma and other cancers. The BRAF V600E mutation, for example is typical for melanoma, but was also found in hairy cell lymphoma patients (Mullighan, 2013). Sos1 germline mutations lead to Noonan syndrome, a developmental disorder that leads to higher risk of patients for myeloproliferative disease and leukemia (Fernández-Medarde and Santos, 2011).

Other pathways also strongly enriched are chemokine and interleukin signaling. Lymphomas need constant activation of survival signaling pathways in order to block apoptosis (CIT). An example is the reduced survival of Eμ-*myc* lymphomas

deprived for the cytokine receptor CCR7, that induces homing to the spleen and further crosstalk with stromal cells via lymphotoxin signaling (Rehm et al., 2011). Summarizing this part suggests that cell survival is strongly induced by upregulating growth, proliferation and survival signals by Ras, MAPK, PI3K and STAT. On the other side, also genomic maintenance is deregulated by abrogation of the p53, that is represented in the pathways found in different cancer types and confirms that the CIS found can override Myc-induced tumor suppressive functions.

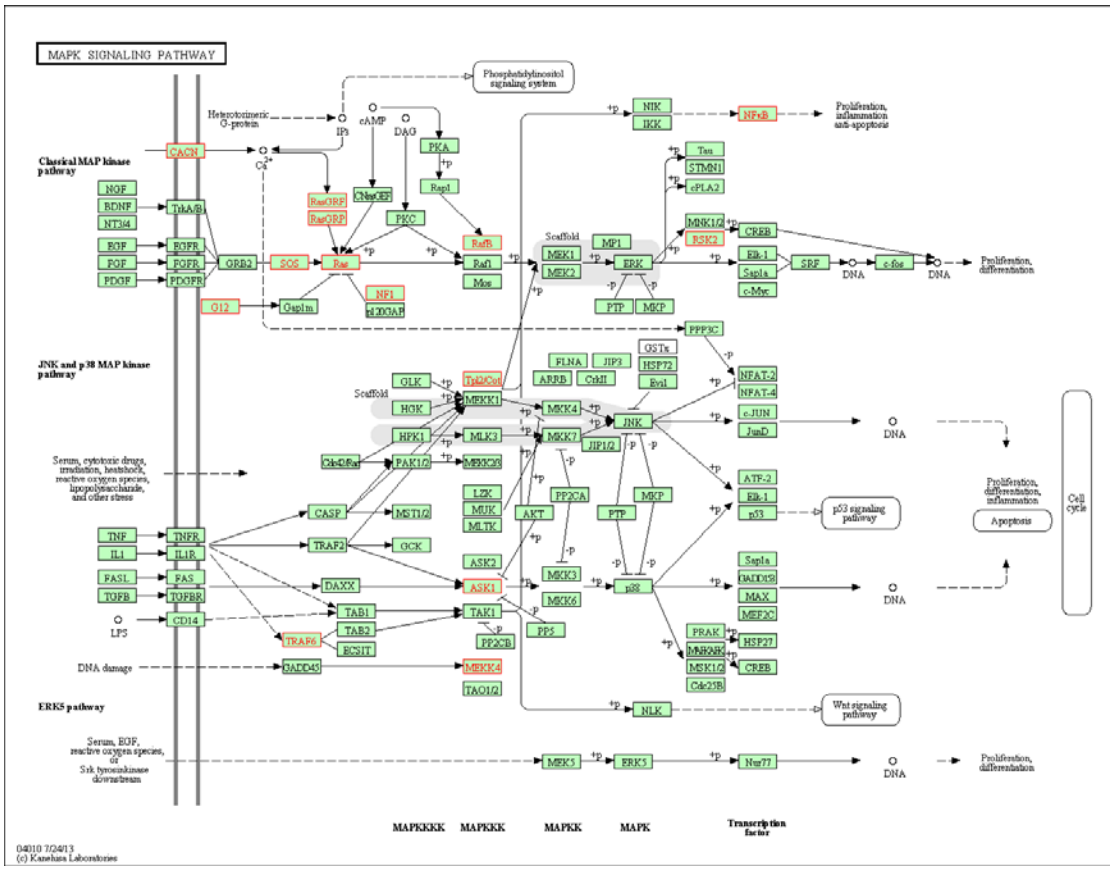


**Figure 33 MAPK signaling pathway. CIS that are involved in this pathway are shown in red.** From an activated receptor, signaling is passed on by receptor tyrosines by autophosphorylation. Receptor bound signaling molecules activate downstream proteins via direct exchange of GDP to GTP in Ras, Rho and Rac or by indirectly activating GEFs (Guanine nucleotide exchange factors). Further the downstream signal is transmitted via phosphorylation by kinases and dephosphorylation by phosphatases. After passing the signals from the MAP3 kinases to the MAP2 kinases, the MAP kinases activate effector proteins, the transcription factors, which carry out the cellular response from outside signals by changing the transcription profile in a cell.

To get further insight into mechanisms that are exploited by genes not belonging to canonical pathways, CIS were uploaded to the string database, which is an online database integrating known and predicted protein information. The aim is

to find interactions between the CIS in order to predict commonalities and differences. Of the 338 distinct CIS, 331 could be associated to an entry in the STRING database as depicted in Figure 34.A statistically significant enrichment in the 234 protein-protein interaction observed compared to the 120 interactions expected was found. This suggests that the reduction in background insertion sites could be achieved, as the CIS do reflect associative behavior. Moreover, the String analysis tool performs a clustering according to a binding distance that is calculated for every protein-protein interaction found, with highly validated binding interactions showing higher proximity of molecules. As expected, we find a strong clustering around the p53-Mdm2-Cdkn2a axis, a pathway that is highly deregulated in Eμ-*myc* lymphomas (Eischen et al., 1999; Schmitt et al., 2002; Zindy et al., 1998). Moreover, the CIS cluster around the transcription factors Foxo1, Sp1 and Ets1which showed to cooperate with Myc (Bonetti et al., 2013; Bouchard et al., 2007; Gartel et al., 2001), suggesting a high enrichment for transcription factors associated with Myc in our screen. Another cluster with many surrounding connections centers on Sos1-Rras2 and confirms the previous finding from the pathway analysis.

Moreover, enrichment analysis for GO terms belonging to biological processes, molecular functions, InterPro domains and PFAM entries are listed in Table 5 and suggest that lymphoma cells use mechanism to uncouple from negative extracellular signals like deprivation of growth factors that would induce an apoptotic response by deregulating receptor signaling pathways. Interesting, there is an enrichment in processes related to nerves, like neurotrophin signaling and NGF signaling, These are stimuli that are usually giving survival signals to nerve cells, which may be exploited by Myc-induced lymphoma cells to block apoptosis. Interestingly, activated B-cells show upregulation of NGF signaling molecules by induction of NFκB (Heese et al., 2006), which was found in the screen as target in

4.9% (Nfkb1) of lymphomas but showed to be dispensable for Eμ-*myc* lymphomagenesis, contrary to its paralog Nfkb2 (Keller et al., 2005, 2010).

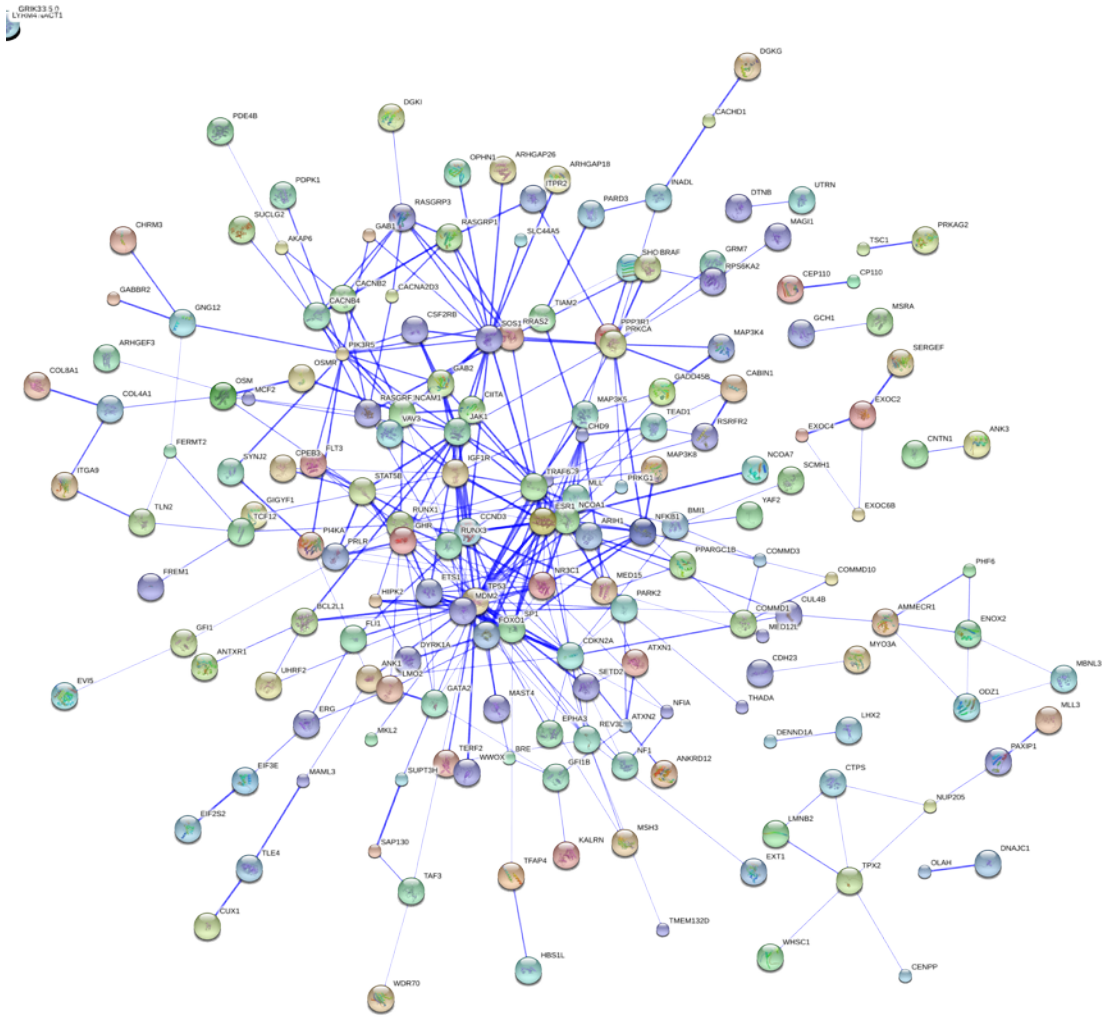

**Figure 34 Interaction between CIS.** The confidence of interaction between different genes depicted as nodes are shown by blue lines, with stronger associations referring to thicker lines. Nodes which show stronger association cluster together.

Enrichment analysis of GO biological processes and molecular functions suggests that the remaining CIS found also belong to receptor-mediated signaling, as already found by the enrichment in the pathway analysis. However, contrary to the mechanisms assumed to induce a survival signal, the term induction of apoptosis by extracellular signals is strongly enriched being represented by 11 CIS. This suggests that not only survival signaling is transmitted by deregulating Ras GTPase activator activity, but also that apoptotic signals that are mediated by the surrounding are abrogated, further suggesting that Myc cooperates with Ras signaling to counteract apoptosis during lymphoma onset. Searching the gene list

for enrichment in certain domains (Interpro) or motifs (PFAM), we can find enrichment of the RasGEF. Ras, in its active form is bound to GTP. Regulation of its activity is introduced by GEFs - guanosine nucleotide exchange factors (Rasgrp1, Rasgrp3, Ralgps1), that activate loading of GTP to Ras whereas GAPS are hydrolases, that convert GTP to GDP, leaving Ras inactive (Fernández-Medarde and Santos, 2011). Moreover, analysis for InterPro domains found enrichment in the p53/RUNX-like transcription factor binding domains, which are present in the transcription factors RUNX1, RUNX3 and p53 found to be hit in this screen. Interestingly, Runx family proteins, which can act as oncogenes or tumor suppressors (Blyth et al., 2005) were targeted in a retroviral insertion screen in a mouse model for Myc-induced T-cell lymphoma (Akagi et al., 2004). In line with that, crossing of the Eμ-*myc* mouse with a *Runx1* overexpressing mouse led to the fast onset of B- and T-cell lymphomas (Blyth et al.) supporting the found CIS as activating Runx1 to synergize in the onset of B-cell lymphoma in this screen. Summarizing this part suggests that receptor mediated signaling leads to upregulation of the Ras-MAPK signaling by modulation of regulators of RAS. This confers survival signals by inhibition of extracellular apoptotic stimuli that cooperate to induce transformation. The CIS found truly cluster around oncogenic Ras and suggest that upstream and downstream mechanisms that convey activity to Ras-MAPK signaling lead to abrogation of Myc induced apoptosis. Moreover, growth factor signaling was strongly enriched, which also uses the Ras-MAP kinase cascade to induce a survival signal in cells. Modulation MAP-kinase signaling is complex due to many signal proteins participating and many outcomes that depend on endurance of a signal and on the activity of collaborating pathways including the PI3K pathway. However, in this screen we could find some of the signalling proteins that are important for Myc-induced lymphoma, like Sos1, Rasgrf and Map3k5. Understanding their function in collaboration with Myc in

tumorigenesis could impose further insight and possibly new therapeutic strategies for tumors with dependence on Myc and Ras signaling.

**Table 5 Enrichment analysis of CIS**

| **Biological Process** | **Term** | **NumberOf Genes** | **p-value** | **p-value_fdr** | **p-value_bonferroni** |
|---|---|---|---|---|---|
| GO:0007169 | transmembrane receptor protein tyrosine kinase signaling pathway | 26 | 1.24E-06 | 1.20E-02 | 1.20E-02 |
| GO:0048011 | nerve growth factor receptor signaling pathway | 17 | 3.09E-06 | 1.49E-02 | 2.98E-02 |
| GO:0007167 | enzyme linked receptor protein signaling pathway | 27 | 9.82E-06 | 2.80E-02 | 9.47E-02 |
| GO:0046677 | response to antibiotic | 5 | 1.16E-05 | 2.80E-02 | 1.12E-01 |
| GO:0008624 | induction of apoptosis by extracellular signals | 11 | 1.65E-05 | 3.18E-02 | 1.59E-01 |
| GO:0006917 | induction of apoptosis | 18 | 6.30E-05 | 9.07E-02 | 6.07E-01 |
| GO:0012502 | induction of programmed cell death | 81 | 6.58E-05 | 9.07E-02 | 6.35E-01 |
| GO:0032318 | regulation of Ras GTPase activity | 10 | 7.73E-05 | 9.32E-02 | 7.45E-01 |
| **Molecular Function** | **Term** | **NumberOf Genes** | **p-value** | **p-value_fdr** | **p-value_bonferroni** |
| GO:0005099 | Ras GTPase activator activity | 6 | 3.29E-04 | 4.03E-01 | 9.03E-01 |
| **GO_Cellular Component** | **Term** | **NumberOf Genes** | **p-value** | **p-value_fdr** | **p-value_bonferroni** |
| GO:0044451 | nucleoplasm part | 23 | 1.10E-04 | 7.31E-02 | 1.28E-01 |
| GO:0044428 | nuclear part | 56 | 2.10E-04 | 7.31E-02 | 2.44E-01 |
| GO:0005667 | transcription factor complex | 11 | 2.40E-04 | 7.31E-02 | 2.79E-01 |
| GO:0031981 | nuclear lumen | 50 | 2.51E-04 | 7.31E-02 | 2.92E-01 |
| GO:0016604 | nuclear body | 9 | 7.31E-04 | 1.62E-01 | 8.51E-01 |
| **InterPro** | **Term** | **NumberOf Genes** | **p-value** | **p-value_fdr** | **p-value_bonferroni** |
| IPR012346 | p53/RUNT-type transcription factor, DNA-binding domain | 3 | 3.07E-05 | 6.93E-02 | 1.40E-01 |
| IPR008937 | Ras guanine nucleotide exchange factor | 3 | 6.07E-05 | 6.93E-02 | 2.77E-01 |
| IPR001895 | Guanine-nucleotide dissociation stimulator CDC25 | 3 | 6.07E-05 | 6.93E-02 | 2.77E-01 |
| IPR023578 | Ras guanine nucleotide exchange factor, domain | 3 | 6.07E-05 | 6.93E-02 | 2.77E-01 |
| IPR016554 | Transcription factor, Runt-related, RUNX | 2 | 2.17E-04 | 1.68E-01 | 9.90E-01 |
| **PFAM** | **Term** | **NumberOf Genes** | **p-value** | **p-value_fdr** | **p-value_bonferroni** |
| PF00617 | RasGEF domain | 3 | 6.72E-05 | 1.56E-01 | 1.56E-01 |

# DISCUSSION

In this study we describe a Sleeping Beauty mediated insertional screen that aims to find new cooperation partners of Myc in lymphomagenesis. Myc is a transcription factor that is involved in tumorigenesis in more than 70% of human cancer (Nilsson and Cleveland, 2003), including Burkitt lymphoma which develops due to translocation of the myc gene to one of the enhancers of the immunoglobulin genes. This results in strong overexpression of *myc* in lymphocytes leading to an aggressive germinal-center B-cell tumor (Ott et al., 2013). Mouse models that tried to recapitulate this human tumor were the λ-*myc* mouse (Kovalchuk et al., 2000), the iMyc$^{Eμ}$ mouse (Park et al., 2005), the model used in this study, the Eμ-*myc* mouse (Adams et al., 1985) and a newly established model (Sander et al., 2012).

The Eμ-*myc* mouse model develops tumors of pre/pro and immature B-cells, although depending on the strain used and environmental factors, T-cell lymphomas were also observed (Yukawa et al., 1989). λ-*myc* mice carry a transgene that puts myc under the control of the λ enhancer of the immunoglobulin light chain and leads to longer latency periods and more mature B-cell lymphomas (Kovalchuk et al., 2000). The iMyc$^{Eμ}$ mouse model, carrying the translocation of *myc* to the IgH gene originally found in plasmacytomas recapitulates Burkitt lymphoma best as it induces germinal center B-cell lymphomas in the mouse (Park et al., 2005). Another mouse model for Burkitt lymphoma was established lately which does not rely on translocation of *myc* to the Ig gene. Instead it expresses a human *myc* gene under the control of a CAG promoter and a constitutively active form of PI3K, P110* under the control of a conditional *Rosa26* promoter only in germinal centers. This conditional expression was achieved by inducing *Cg1*-Cre expression in germinal center B-cells and

introducing a lox-stop-lox cassette upstream of the before mentioned constructs. Lymphomas reflect Burkitt lymphoma morphology, phenotye, expression profiles and in some cases could also reconcile some frequent somatic mutations like *Ccnd3* suggesting a major role for the PI3K pathway in the disease (Sander et al., 2012).

Although the Eμ-*myc* mouse model does not show stage specific overlap with the human disease, expression profile comparisons as well as cell morphology and tumor appearance suggest similarities to Burkitt lymphoma. Moreover, a correlation between the onset of tumorigenesis (and thus different cooperating mutations) and different types of non-Hodgkin lymphomas was found using expression profiles and morphological features (Mori et al., 2008). Burkitt lymphoma is the main tumor described to harbor the IgH-*myc* translocation, a fraction of plasmacytomas and DLBCLs also harbor such lesions. Other hematological diseases like acute lymphoblastic leukemia (ALL), acute myeloblastic leukemia (AML), plasmacytomas and bursal lymphomas show deregulation of Myc by overexpression or amplification (Delgado and León, 2010; Land et al., 1983). Altogether, this suggests that the Eμ-*myc* mouse serves as a relevant model deciphering the transforming potential of Myc per se as it was suggested to act in a non-tissue specific way (Land et al., 1983), which is supported by the manifold of tumor types that show deregulation of Myc, making up 70% of human cancers (Nilsson and Cleveland, 2003). Moreover, it serves a prototype model for general cooperating mechanisms applicable in Myc induced tumorigenesis, as p53 is found mutated in every human cancer type ranging from 5-50% of incidents (Levine, 1997) and the p53-mdm2-arf axis is deregulated in 60% of spontaneous Eμ-*myc* lymphomas (Eischen et al., 1999). Convenient is also that this model has been established in the 1980s and since then has been widely used to study Myc induced lymphomagenesis as a multistep disease.

## Experimental system

In this study we successfully could perform an insertional mutagenesis screen, circumventing complicated and long lasting breeding strategies by applying an adoptive transfer strategy. HSPCs obtained from 3 donors contained already existing tumor clones that gave rise to fast onset lymphomas when transplanted. These lymphomas had a very short latency in all recipient mice with similar disease dissemination and suggests the selection of strong candidate cooperating partners of Myc, as non-malignant Eμ-*myc* BM cells do not give rise to tumors when transplanted (Figure 7) (Alexander et al., 1989a). The other 14 donors gave rise to independent tumors in the recipient cohorts. The evolving lymphomas varied highly in tumor dissemination and showed more than 200 days of difference in tumor latency. Moreover immunophenotyping showed different developmental stages and populations, as well as diverse clonality status in a cohort of mice transplanted from the same donor indicating that these lymphomas had evolved independently. To confirm this statistically, comparisons between the insertional signatures of mice from a group of recipients deriving from the same donor cohort will be compared to mice from different donors by Chi-squared test (in progress).

Spontaneous lymphomas in the Eμ-*myc* mouse evolve from a single transformation at is highly selected and leads to clonal disease in most cases (Adams et al., 1985). Oligoclonality, however, was also reported in some cases of the Eμ-*myc* mouse (Adams et al., 1985) shown by PCR of the BCR VDJ recombination spectrum. In our Sleeping Beauty mediated screen we find oligoclonal BCR rearrangements in 72% of the cases, together with the co-existence of up to three lymphoma populations of different developmental stages as assayed by immunophenotyping. These data indicate that the augmented rate of Sleeping Beauty mediated mutations allow multiple tumor clones to co-evolve in the same mouse. This was previously

documented by several groups using the conditional SB system in solid tumors (Keng et al., 2009; March et al., 2011). In intestinal tumors March et al. suggested that tumors of the intestine that showed the same diversity of insertion site distribution in tumors of the same mouse as compared to the ones of different mice, concordant to our suggestion that multiple tumor clones are giving rise to independent tumors. However, the presence of multiple independent tumors in a mouse does not interfere with further analysis of CIS, as these are defined by the amount of individual IS in a region comparing different mice. As suggested by others (Dupuy et al., 2005)(March et al., 2011), subpopulations of tumor cells are represented by different insertions, which from an analytic point of view results in the same detection of CIS either in one mouse with multiple tumors, or in many mice each harboring one tumor. However, sequencing depth could be affected, leading to reduced calling of rare events.

During immunophenotyping we find more than one population present per mouse, which indicates oligoclonality of B-cell lymphomas as explained above. However, not all of the cells comprising the tumor were stained for B220, a marker of the B-cell lineage, although mice showed clear morphological features of lymphoma dissemination with prevalent splenomegaly, enlarged lymph nodes, and liver. Further staining of these populations defines the prevalent non B-cell population to be composed of Cd3ε positive T-cells. Moreover, the proportion of B-cell versus non-B-cell population was variable, non-B-cells ranging from 5 to more than 50% of the lymphoma suspension analyzed. Technical artifacts due to the freezing of lymphoma cells could account for that, although preexperiments showed no difference in staining efficiency of lymphomas. Also, incorrect tumor resection of thymic lymph nodes could lead to the high number of T-cells shown in some of the lympohoma samples. We also thought about the possibility of T-cell lymphomas arising along with B-cell lymphomas. Controversial information was found about

Eμ enhancer expression in different compartments of the hematopoietic system and disease type: expression of an Eμ-*myc* transgene induced mainly B-cell lymphomas in a B6 background whereas it showed mainly T-cell lymphomas in a C3H/HeJ background (Yukawa et al., 1989). However in most of the mentioned published Eμ-*myc* mice backcrossed to B6 background suggested solely B-cell lymphomas to arise (Adams et al., 1985; Schuster et al., 2011; Schmitt et al., 2002). Further Yukawa suggested that environmental factors like presence of viruses would change the etiology of cancers by showing transplantation experiments (Yukawa et al., 1989), which is shown by T-cell lymphomas arising in proviral insertional mutagenesis screens performed in the Eμ-*myc* model (van Lohuizen et al., 1991). Thus we thought that the transplantation strategy would favor the clonal selection of transformed B- and T-lymphocytes due to overexpression of Myc presumably already in the common lymphoid progenitor (CPL) stage. However, the Eμ enhancer induced expression of nearby genes strictly in the B-cell lineage as shown by the Eμ-*GFP* transgenic mouse (Guglielmi et al., 2003). Moreover, Cd19-Cre was shown exclusively expressed in the B-cell lineage (Rickert et al., 1997). Thus although T-cell lymphomas may arise in Eμ-*myc* mice, Sleeping Beauty insertional mutagenesis seems restricted to B-cells, as the thymus as organ comprised mainly by T-lymphocytes of Cre-carrying Eμ-*myc*-SB mice shows no activation of SB transposase. Thus SB lymphomagenesis is specific to B-cells and only these will undergo insertional mutations tagged by a transposon. All subsequent steps in our analysis rely on the presence of mobilized transposons in the genome, as LM-mediated PCR primers specifically amplify from these sites. Thus although another population could be present in the lymph nodes of lymphoma bearing mice, this should not impact on the analysis of cooperating hits in Eμ-*myc* B-cell lymphoma.

## Insertion sites and drivers found

In this screen we analyzed 184 lymphomas that developed in Eμ-*myc* mice carrying a conditional SB insertional mutagenesis system and we could identify 338 CIS.

The filtering methods HM101 and BB could exclude background insertions to a certain extent to define clonal insertion sites, whereas the method HM34 failed to do so. This is shown by a statistical significant selection of gene-centric regions compared to the expected random distribution of clonal IS introduced by SB (Dupuy 2009)(Figure 28). Unexpectedly, the same analysis performed on IS from splenic B-cells obtained from mice that did not carry the Eμ-*myc* transgene and had a fully functional SB insertional mutagenesis system indicates the presence of clonal selected IS (Figure 27). However the control mice did not show signs of disease at the time the samples were taken. This suggests that B-cells from younger mice would have served a better control for defining clonal insertion sites. In this way, however the clonal IS from control mice serve as good control for genes that are selected without the presence of myc as driving oncogene. Although we could not rule out the presence of false positives in our set of clonal IS, 65.2% of CIS passed the filtering criteria of all three filtering methods (BB, HM101 and HM34) and less than 1% of the CIS were found only by one filtering method, confirming the robustness of our strategy used to obtain CIS.

Regarding the methods to obtain CIS from clonal IS, 10.81% were found with all three approaches. This core data set will be highly relevant for future experiments as it includes previously reported modulators of lymphomagenesis in the Eμ-*myc* mouse model (Bmi-1 - 8.7%, Gfi1b - 36.4%, Cdkn2a - 12.5%, and p53 - 10.3%,) (Eischen et al., 1999; Jacobs et al., 1999; Schmidt et al., 1998; Zindy et al., 1998), as

well as not yet confirmed genes. These were shown to be mutated in B-cell lymphoma (Gata2, Braf, Flt3, Snx19, Thada, Mll1) (Giulino-Roth et al., 2012; Lohr et al., 2012; Love et al., 2012; Morin et al., 2013; Richter et al., 2012; Schmitz et al., 2012; Zhang et al., 2013). Also, CIS include putative tumor suppressors (Aadacl3, Tead1) and oncogenes (Sos1, Eif2s2) (Davoli et al., 2013), genes involved in modulating Myc's transcriptional response like the coactivator MEF2B in B-cells (Wang et al., 2009) and Sp1, a zinc-finger protein that is transcriptionally repressed by Myc binding (Gartel et al., 2001). Previously used methods to define CIS (mcCIS and gCIS) (Brett et al., 2011) showed stringent background exclusion and valid CIS, most of which were also found by the enrichment method (72.9% and 83.6%, respectively). Moreover, the new method used found a large number of additional CIS, that do not account for false positives as shown by statistical significant overlap with databases (Table 3). Moreover, the CIS found were strongly enriched in genes relevant for Myc-induced tumors, as suggested by significant overlap with an insertional mutagenesis screens performed in Myc-overexpressing mice (RTGCD). Rewardingly, we find more genes known to cooperate with Myc to induce lymphomagenesis in the Eμ-*myc* mouse model (Bmi-1 – 8.7%, Mdm2 – 1.1%) (Eischen et al., 1999; Jacobs et al., 1999). Additionally we find new cooperating partners of Myc-induced lymphomagenesis, previously not described, but being relevant for Burkitt lymphoma, as suggested by 91 genes overlapping with the somatic mutation found recently by sequencing Burkitt lymphoma patients (Giulino-Roth et al., 2012; Love et al., 2012; Richter et al., 2012; Schmitz et al., 2012). However, the coreCIS, which are insertion sites found with three independent statistically methods showed no statistical relevant overlap with this gene set. As Myc is also found deregulated due to translocation events including the immunoglobulin gene in DLBCL (Delgado and León, 2010), we took advantage of recently published mutations found in human DLBCL patients (Lohr

et al., 2012; Morin et al., 2013; Zhang et al., 2013). Indeed, 141 CIS overlapped with these mutations suggesting that the tumors enriched in this study resemble more human DLBCL than BL, and is coherent with previous data that Eμ-*myc* lymphomas show BL-line and DLBCL-like features (Mori et al., 2008). Interestingly, 44 CIS were found in both types of tumors suggesting general functions valid for transformation to be important in these tumors, including deregulation of cell cycle by cyclin D3 (13.6%), known to have stabilizing mutations which drives cell cycle progression and is found in 38% of Burkitt lymphoma patients (Schmitz et al., 2012).

## Connection of CIS to Myc

Myc induced tumor suppressive functions include at least three mechanistic entities: apoptosis induction, blockage in cell cycle progression and senescence. All of these mechanisms are an intrinsic failsafe program that was successfully abrogated by SB-mediated tumorigenesis.

Overexpression of Myc was shown to induce proliferation and also an apoptotic response (Evan et al., 1992). Myc-induced apoptosis is dependent upon the activation of p53 via induction of ARF (Eischen et al., 1999; Schmitt et al., 1999; Zindy et al., 1998), one of the products of the Cdkn2a locus, which was targeted by SB in this screen in 12.5% of lymphomas. ARF is suggested to be unregulated by induction FoxO proteins, which are induced upon Myc elevation (Bouchard et al., 2007), and in this screen hits in *FoxO1* were present in 7.6% of the tumors, which confirms the important role of ARF disruption seen in 24% of Eμ-*myc* mice (Eischen et al., 1999). ARF binds and thus inactivates Mdm2, a ubiquitin ligase regulating p53 proteolytic degradation. Interestingly, we find *Mdm2* targeted in a very small fraction of the lymphomas (1.1%): insertions occur upstream of the

coding sequence and in the sense orientation, suggesting Mdm2 upregulation. Consistent with a positive role of Mdm2 in Myc-induced lymphomagenesis, *Mdm2* heterozygous Eμ-*myc* mice showed delayed tumor onset (Alt et al., 2003). As observed in the Eμ-*myc* mouse (Schmitt et al., 2002) and in many human tumors, inactivation of p53 was also found in this screen, occurring in 10.3% of lymphomas. The role of p53 in Myc-induced apoptosis includes the induction of apoptotic mediators, such as the pro-apoptotic Bcl2-family members Bax, Noxa and Puma (Lowe, 2004). Members of this family were targeted in our screens, including *Bcl2l1* (6.5%), better known as Bcl-XL, a Bcl-2-like anti-apoptotic protein (Eischen et al., 2001). During B-cell development IL7/IL7R signaling together with the preBCR induces clonal expansion and differentiation of common lymphocytic progenitors to committed progenitor B-cells (Gold, 2008; Habib et al., 2007), and progenitors express Bcl-XL to circumvent apoptosis during development (Eischen et al., 2001). Together with Bcl-2, Bcl-XL was upregulated in more than 50% of Eμ-myc lymphomas (Eischen et al., 2001). In summary, we find an enrichment of CIS in genes known to modulate the apoptotic response induced by increased Myc levels.

On the other hand, most of the CIS were clustered in having a function in regulating Ras-signaling. Ras signaling and downstream cascades were highly enriched with almost 90% of the lymphomas carrying an insertion in at least one of the members of this pathway, confirming previously shown synergism of Myc and Ras signaling in the Eμ-*myc* mouse model (Alexander et al., 1989a, 1989b; Gramling and Eischen, 2012).

Most of the CIS were not found in the Ras or related family members of GTPases, although we did find *Rras2* in 9.2% of lymphomas. Instead, most of the CIS found were modulators of GTPases, signaling molecules that either activate the Ras and related members of the super family of G-proteins by exchanging GDP with GTP,

the so called GEF (guanine nucleotide exchange factors) or GAPs (GTPase-activating proteins), which control hydrolysis of GTP to GDP. Most of the hits were assigned to *Sos1* (75.5%) and *Rasgrf1* (41.8%), both GEFs activating Ras downstream signaling via the RAF-MEK-ERK cascade. Interestingly this signaling pathway is strongly induced upon antigen stimulation of the BCR-stimulation in normal B-cell development which as response leads to upregulation of Myc (Gold, 2008). Interestingly, some of the GEFs are known to be involved in B-cell development and activation, like Vav3 (Vigorito et al., 2005). Moreover, Ras activity is modulated by GTPase-activating proteins like the tumor suppressor Nf1, found in 12% of lymphomas. Homozygous *Nf1* knock-out mice show embryonic lethality and heterozygous animals show predisposition to a multitude of cancer types (Eppig et al., 2012). Moreover, in hematologic disease, reduced expression of Nf1 led to Ras-mediated hypersensitivity to granulocyte/macrophage-colony stimulating factor (GM-CSF) and cooperated in leukemogenesis in a SB-mediated screen (Largaespada et al., 1996).

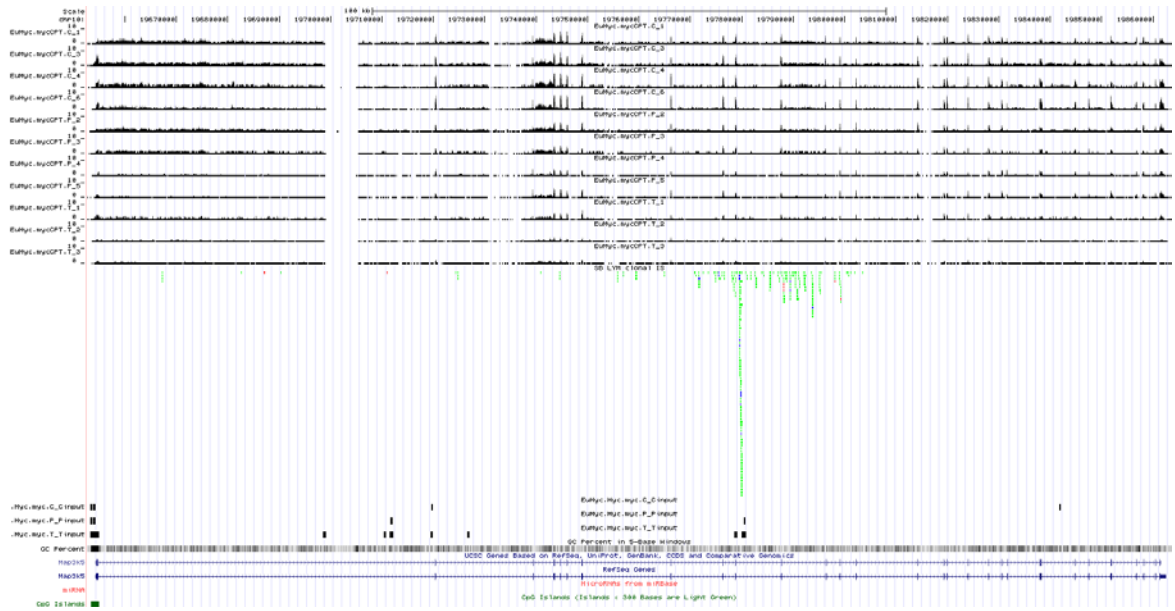

**Figure 35 Clonal insertions found for the CIS *Map3k5*.** In the upper 4 tracks results of RNA sequencing of isolated B-cells are performed. Followed by 3 lanes of RNA sequencing reads performed on pretumoral Eμ-*myc* mice, that did not develop yet lymphoma but show already expansion of the B-cell pool and the following 3 lanes are RNA sequenced Eμ-*myc* lymphomas. The following lane depicts clonally insertion sites, with green lines marking the insertions that are in sense orientation, whereas red ones mark the antisense insertions. Blue insertion sites depicted cannot be assesessed for their insertional orientation. The 3 lanes below represent Myc-binding sites in control B-cells (upper lane), B-cells isolated from pretumoral mice (middle lane) and lymphomas of diseased Eμ-*myc* mice (lower lane) assessed by ChIP-Seq analysis.

One of the most frequently hit genes in our screen was *Map3k5* (44.6%), a MAP3K family member encoding for Ask1. The insertions between intron 8 and intron 15 depicted in Figure 35 are in sense orientation in more than 90% of the cases, which suggest overexpression of a short isoform of this protein, containing the catalytic domain and the activating and regulating phosphosites on the C-terminus of the protein but missing the DUF4071 domain, of unknown function. Comparison with ChIP-seq data generated in our group (Sabò et al., submitted) showed that insertion sites center around Myc bound regions found in pretumoral and tumoral samples, but not in control B-cells (Figure 35), suggesting that this might be a regulatory mechanism in tumor initiation. Regarding transcription, it seems that the levels of Map3k5 RNA decrease, comparing the controls with tumor samples, suggesting that binding of Myc to this site blocks further transcription and abrogates Ask1 signaling, which might be a candidate for Myc-mediated repression of transcription. Ask1 sensors different stresses that occur in a cell, as summarized in **Figure 36.**

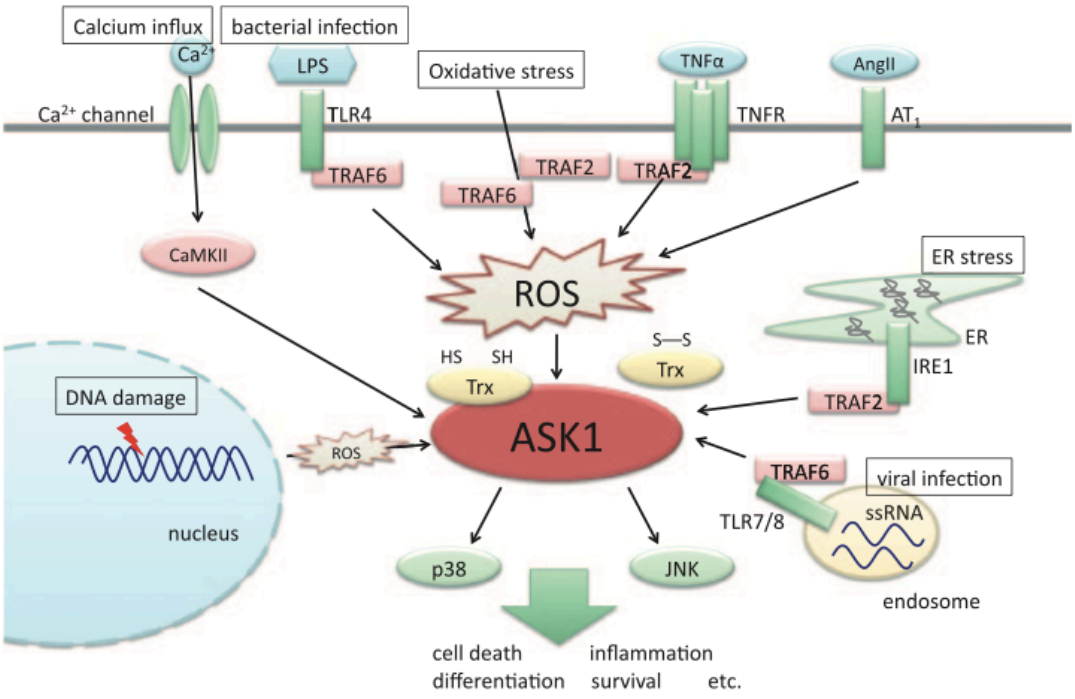


**Figure 36 ASK1 activation by stress.** (Hayakawa et al., 2012)

As such, it induces an apoptotic response in plasma cells, downstream of Blimp-1 signaling (Lin et al., 2012). A suggested mechanism for Ask1-mediated apoptosis is downstream JNK signaling, that leads to Myc phosphorylation at both Ser-62 and Ser-71. This leads to enhanced stability and accumulation of Myc protein to levels that in turn trigger an apoptotic response (Noguchi et al., 2001). Thus, Ask1 might act as a tumor suppressor in this setting. Thus, the CIS found in the Ras-MAP kinase signaling pathways might be involved in inducing a survival advantage by reducing apoptotic triggers that occur when growth factors are depleted or cellular stress is increased, as shown by Ask1.

**Blocking senescence**

Myc triggers not only a compensatory apoptotic response, but also senescence. The concept that senescence can act as a tumor suppressive mechanism in Myc overexpressing cells emerged from several studies. In particular, knock-out mice for Cdk2 showed increased senescence which delayed tumor onset (Campaner et al., 2009) Also, blocking senescence-related histone methyltransferase Suv39h1 (Reimann et al., 2010) and abrogated binding of Myc to Miz-1 (van Riggelen et al., 2010). Constant DNA damage, as well as increased levels of ROS and shortened telomeres all trigger cellular senescence, an irreversible growth arrest that is marked by upregulation of p16 and p21 (Campisi and d'Adda di Fagagna, 2007). Myc activation is known to induce DNA damage in vivo by upregulation of ROS (Reimann et al., 2010) and blocking ATM, p53 or Suv39h or Cdk2 prevented cells from undergoing Myc-induced senescence. Moreover, Miz-1 repression by Myc leads to abrogation of transcription of p15 and p21, having a role in irreversible cell cycle block that marks senescence. Targeting the mTOR pathway, that promotes nutrient and growth factor dependent signaling was shown to induce

senescence in Eμ-*myc* mice (Wall et al., 2013). Interestingly, we find Tsc1, a protein that regulates mTOR activity and ribosomal protein S6 kinase in 4.3% of lymphomas each, thus suggesting that mTOR mediated senescence was modulated in Myc overexpressing B-cells to induce lymphomagenesis.

**Changing transcription**

Myc changes transcription by recruiting histone deacetlyation complexes, HATs and mediators that change the accessibility of histones for transcription (Herkert and Eilers, 2010). The SAGA/STAGA complex is involved in recruiting mediator complexes to Myc (Liu et al., 2008). Interestingly, we find Supt3h, better known as Spt3 as part of the STAGA complex (Herkert and Eilers, 2010; Martinez et al., 1998) in 6% of the lymphomas and mediator complexes Med15 (3.8%) and Med12l (4.9%). Moreover, Sap130 (8.2%) is a subunit of the histone deacetylase-dependent Sin3a corepressor complex

Moreover, Myc associates with core transcription factors to activate transcription. We found the transcription factor Sp1 hit in 2.5% of the lymhpomas. Myc cooperates with Sp1 to activate transcription of the catalytic subunit of telomerase, that permits the tumor cells to cycle beyond replicative senescence (Kyo et al, 2000). Also, myc sequesters Sp1 from promoters, leading to transcriptional repression of cell cycle inhibitor p21 (Gartel et al., 2001). Moreover, also Ets trancription factor Ets-1 that has special roles in hematopoiesis was found as CIS in 8.5% of lymphomas and Fli-1 was found in 5.3% of lymphomas. Both transcription factors were shown to be involved in DLBCL to induce proliferation, survival, and germinal center differentiation by deregulating cytokine receptors (Bonetti et al., 2013).

Myc is mainly a transcription factor, however it also represses transcription by binding to Miz-1. Miz-1 is a transcriptional activator, that together with the co-factors p300 and NPM activates target genes involved in anti-mitotic signaling. It is important in mounting a senescence response upon DNA damage or TGFβ signaling by inducing cell cycle inhibitors p21 and p15, which are key players in cell cycle blockage (van Riggelen et al, 2010; (Seoane et al., 2002). Miz-1 was moreover shown to interact with Gfi1 and Myc to form a ternary complex, that represses Miz-1 target genes *Cdkn2a* and *Cdkn2b*, which are important for inducing cell cycle inhibition and senescence (Basu et al., 2009; Liu et al., 2010). Gfi1 and its paralog Gfi1b are hit in 62.5% of the lymphomas, suggesting this repression to be highly important for Myc-induced lymphomagenesis. This is also confirmed by 20% of retroviral CIS found in this gene in the in Eμ-*myc* mouse model (Scheijen et al., 1997). Coherent to the activation of transcription found in the retroviral screen, we find 82% of IS in the sense orientation, which suggests an activation of transcription of Gfi1 and Gfi1b also in this screen. Thus, we speculate that upregulation of Gfi1 circumvents senescence and cell cycle arrest by anitmitogenic stimuli, which together with Myc induces repression of Miz-1.

Together this suggests, that our screen found many cooperating partners of Myc in modulating a tumor suppressive response like Bcl-X, p53, ARF, Mdm2 and Gfi. The main cooperating pathway activated is the classic Ras-Raf-ERK pathways, that is suggested to be upregulated by CIS in Sos1 and Rasgfr1. Moreover, MAP3K signaling through JNK was hit in lymphomas, suggesting reduction of apoptosis by stabilizing Myc. Further interesting hits include genes that are involved in transcription, like Ets-1. In conclusion, the CIS found are ideal candidates for further experimental validation in order to understand mechanisms underlying Myc-induced lymphomagenesis. Further the CIS found can be directly assessed for their druggability in order to find new cures for human B-cell malignancies.

## SUPPLEMENTS

Supplements are provided as Excel tables enclosed on a DVD

Supplementary table 1: Clonal Insertion sites and analysis
Supplementary table 2: Common insertion sites and analysis
Supplementary table 3: Annotated CIS
Supplementary table 4: Pathway and enrichment analysis

# REFERENCES


Adams, J.M., Harris, A.W., Pinkert, C.A., Corcoran, L.M., Alexander, W.S., Cory, S., Palmiter, R.D., and Brinster, R.L. (1985). The c-myc oncogene driven by immunoglobulin enhancers induces lymphoid malignancy in transgenic mice. Nature *318*, 533–538.

Akagi, K., Suzuki, T., Stephens, R.M., Jenkins, N.A., and Copeland, N.G. (2004). RTCGD: retroviral tagged cancer gene database. Nucleic Acids Res. *32*, D523–7.

Alexander, W.S., Schrader, J.W., and Adams, J.M. (1987). Expression of the c-myc oncogene under control of an immunoglobulin enhancer in E mu-myc transgenic mice. Mol. Cell. Biol. *7*, 1436–1444.

Alexander, W.S., Adams, J.M., and Cory, S. (1989a). Oncogene cooperation in lymphocyte transformation: malignant conversion of E mu-myc transgenic pre-B cells in vitro is enhanced by v-H-ras or v-raf but not v-abl. Mol. Cell. Biol. *9*, 67–73.

Alexander, W.S., Bernard, O., Cory, S., and Adams, J.M. (1989b). Lymphomagenesis in E mu-myc transgenic mice can involve ras mutations. Oncogene *4*, 575–581.

Alt, J.R., Greiner, T.C., Cleveland, J.L., and Eischen, C.M. (2003). Mdm2 haplo-insufficiency profoundly inhibits Myc-induced lymphomagenesis. EMBO J. *22*, 1442–1450.

Amati, B., Littlewood, T.D., Evan, G.I., and Land, H. (1993). The c-Myc protein induces cell cycle progression and apoptosis through dimerization with Max. EMBO J. *12*, 5083–5087.

Amati, B., Frank, S.R., Donjerkovic, D., and Taubert, S. (2001). Function of the c-Myc oncoprotein in chromatin remodeling and transcription. Biochim. Biophys. Acta *1471*, M135–45.

Basu, S., Liu, Q., Qiu, Y., and Dong, F. (2009). Gfi-1 represses CDKN2B encoding p15INK4B through interaction with Miz-1. Proc. Natl. Acad. Sci. U. S. A. *106*, 1433–1438.

Berquam-Vrieze, K.E., Nannapaneni, K., Brett, B.T., Holmfeldt, L., Ma, J., Zagorodna, O., Jenkins, N.A., Copeland, N.G., Meyerholz, D.K., Knudson, C.M., et al. (2011). Cell of origin strongly influences genetic selection in a mouse model of T-ALL. Blood *118*, 4646–4656.

Blankenberg, D., Von Kuster, G., Coraor, N., Ananda, G., Lazarus, R., Mangan, M., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a web-based genome analysis tool for experimentalists. Curr. Protoc. Mol. Biol. *Chapter 19*, Unit 19.10.1–21.

Blyth, K., Slater, N., Hanlon, L., Bell, M., Mackay, N., Stewart, M., Neil, J.C., and Cameron, E.R. Runx1 promotes B-cell survival and lymphoma development. Blood Cells. Mol. Dis. *43*, 12–19.

Blyth, K., Terry, A., Mackay, N., Vaillant, F., Bell, M., Cameron, E.R., Neil, J.C., and Stewart, M. (2001). Runx2: a novel oncogenic effector revealed by in vivo complementation and retroviral tagging. Oncogene *20*, 295–302.

Blyth, K., Cameron, E.R., and Neil, J.C. (2005). The RUNX genes: gain or loss of function in cancer. Nat. Rev. Cancer *5*, 376–387.

Bonetti, P., Testoni, M., Scandurra, M., Ponzoni, M., Piva, R., Mensah, A.A., Rinaldi, A., Kwee, I., Tibiletti, M.G., Iqbal, J., et al. (2013). Deregulation of ETS1 and FLI1 contributes to the pathogenesis of diffuse large B-cell lymphoma. Blood *122*, 2233–2241.

Bouchard, C., Lee, S., Paulus-Hock, V., Loddenkemper, C., Eilers, M., and Schmitt, C.A. (2007). FoxO transcription factors suppress Myc-driven lymphomagenesis via direct activation of Arf. Genes Dev. *21*, 2775–2787.

Brenner, C., Deplus, R., Didelot, C., Loriot, A., Viré, E., De Smet, C., Gutierrez, A., Danovi, D., Bernard, D., Boon, T., et al. (2005). Myc represses transcription through recruitment of DNA methyltransferase corepressor. EMBO J. *24*, 336–346.

Brett, B.T., Berquam-Vrieze, K.E., Nannapaneni, K., Huang, J., Scheetz, T.E., and Dupuy, A.J. (2011). Novel molecular and computational methods improve the accuracy of insertion site analysis in Sleeping Beauty-induced tumors. PLoS One *6*, e24668.

Bric, A., Miething, C., Bialucha, C.U., Scuoppo, C., Zender, L., Krasnitz, A., Xuan, Z., Zuber, J., Wigler, M., Hicks, J., et al. (2009). Functional identification of tumor-suppressor genes through an in vivo RNA interference screen in a mouse lymphoma model. Cancer Cell *16*, 324–335.

Campaner, S., and Amati, B. (2012). Two sides of the Myc-induced DNA damage response: from tumor suppression to tumor maintenance. Cell Div. *7*, 6.

Campaner, S., Doni, M., Hydbring, P., Verrecchia, A., Bianchi, L., Sardella, D., Schleker, T., Perna, D., Tronnersjö, S., Murga, M., et al. (2009). Cdk2 suppresses cellular senescence induced by the c-myc oncogene. Nat. Cell Biol. *12*, 54–59.

Campisi, J., and d'Adda di Fagagna, F. (2007). Cellular senescence: when bad things happen to good cells. Nat. Rev. Mol. Cell Biol. *8*, 729–740.

Charron, J., Malynn, B.A., Fisher, P., Stewart, V., Jeannotte, L., Goff, S.P., Robertson, E.J., and Alt, F.W. (1992). Embryonic lethality in mice homozygous for a targeted disruption of the N-myc gene. Genes Dev. *6*, 2248–2257.

Collier, L.S., and Largaespada, D.A. (2007). Transposons for cancer gene discovery: Sleeping Beauty and beyond. *8*, 1–8.

Collier, L.S., Carlson, C.M., Ravimohan, S., Dupuy, A.J., and Largaespada, D.A. (2005). Cancer gene discovery in solid tumours using transposon-based somatic mutagenesis in the mouse. Nature *436*, 272–276.

Collier, L.S., Adams, D.J., Hackett, C.S., Bendzick, L.E., Akagi, K., Davies, M.N., Diers, M.D., Rodriguez, F.J., Bender, A.M., Tieu, C., et al. (2009). Whole-body sleeping beauty mutagenesis can cause penetrant leukemia/lymphoma and rare high-grade glioma without associated embryonic lethality. Cancer Res. *69*, 8429–8437.

Copeland, N.G., and Jenkins, N. a (2010). Harnessing transposons for cancer gene discovery. Nat. Rev. Cancer *10*, 696–706.

Dalla-Favera, R., Bregni, M., Erikson, J., Patterson, D., Gallo, R.C., and Croce, C.M. (1982). Human c-myc onc gene is located on the region of chromosome 8 that is translocated in Burkitt lymphoma cells. Proc. Natl. Acad. Sci. U. S. A. *79*, 7824–7827.

Dang, C. V (2012). MYC on the path to cancer. Cell *149*, 22–35.

Dang, C. V, O'Donnell, K. a, Zeller, K.I., Nguyen, T., Osthus, R.C., and Li, F. (2006). The c-Myc target gene network. Semin. Cancer Biol. *16*, 253–264.

Davis, a C., Wims, M., Spotts, G.D., Hann, S.R., and Bradley, A. (1993). A null c-myc mutation causes lethality before 10.5 days of gestation in homozygotes and reduced fertility in heterozygous female mice. Genes Dev. *7*, 671–682.

Davoli, T., Xu, A.W., Mengwasser, K.E., Sack, L.M., Yoon, J.C., Park, P.J., and Elledge, S.J. (2013). Cumulative Haploinsufficiency and Triplosensitivity Drive Aneuploidy Patterns and Shape the Cancer Genome. Cell *155*, 948–962.

Delgado, M.D., and León, J. (2010). Myc roles in hematopoiesis and leukemia. Genes Cancer *1*, 605–616.

Delgado, M.D., Albajar, M., Gomez-Casares, M.T., Batlle, A., and León, J. (2013). MYC oncogene in myeloid neoplasias. Clin. Transl. Oncol. *15*, 87–94.

Dupuy, A.J. (2010). Transposon-based screens for cancer gene discovery in mouse models. Semin. Cancer Biol. *20*, 261–268.

Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature *436*, 221–226.

Dupuy, A.J., Rogers, L.M., Kim, J., Nannapaneni, K., Starr, T.K., Liu, P., Largaespada, D.A., Scheetz, T.E., Jenkins, N.A., and Copeland, N.G. (2009). A modified sleeping beauty transposon system that can be used to model a wide variety of human cancers in mice. Cancer Res. *69*, 8150–8156.

Eischen, C.M., Weber, J.D., Roussel, M.F., Sherr, C.J., and Cleveland, J.L. (1999). Disruption of the ARF-Mdm2-p53 tumor suppressor pathway in Myc-induced lymphomagenesis. Genes Dev. *13*, 2658–2669.

Eischen, C.M., Roussel, M.F., Korsmeyer, S.J., and Cleveland, J.L. (2001a). Bax loss impairs Myc-induced apoptosis and circumvents the selection of p53 mutations during Myc-mediated lymphomagenesis. Mol. Cell. Biol. *21*, 7653–7662.

Eischen, C.M., Woo, D., Roussel, M.F., and Cleveland, J.L. (2001b). Apoptosis triggered by Myc-induced suppression of Bcl-X(L) or Bcl-2 is bypassed during lymphomagenesis. Mol. Cell. Biol. *21*, 5063–5070.

Eppig, J.T., Blake, J.A., Bult, C.J., Kadin, J.A., and Richardson, J.E. (2012). The Mouse Genome Database (MGD): comprehensive resource for genetics and genomics of the laboratory mouse. Nucleic Acids Res. *40*, D881–6.

Evan, G.I., Wyllie, A.H., Gilbert, C.S., Littlewood, T.D., Land, H., Brooks, M., Waters, C.M., Penn, L.Z., and Hancock, D.C. (1992). Induction of apoptosis in fibroblasts by c-myc protein. Cell *69*, 119–128.

Fernández-Medarde, A., and Santos, E. (2011). Ras in cancer and developmental diseases. Genes Cancer *2*, 344–358.

Frenzel, A., Labi, V., Chmelewskij, W., Ploner, C., Geley, S., Fiegl, H., Tzankov, A., and Villunger, A. (2010). Suppression of B-cell lymphomagenesis by the BH3-only proteins Bmf and Bad. Blood *115*, 995–1005.

Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. Nat. Rev. Cancer *4*, 177–183.

Garrison, S.P., Jeffers, J.R., Yang, C., Nilsson, J. a, Hall, M. a, Rehg, J.E., Yue, W., Yu, J., Zhang, L., Onciu, M., et al. (2008). Selection against PUMA gene expression in Myc-driven B-cell lymphomagenesis. Mol. Cell. Biol. *28*, 5391–5402.

Gartel, A.L., Ye, X., Goufman, E., Shianov, P., Hay, N., Najmabadi, F., and Tyner, A.L. (2001). Myc represses the p21(WAF1/CIP1) promoter and interacts with Sp1/Sp3. Proc. Natl. Acad. Sci. U. S. A. *98*, 4510–4515.

Giardine, B., Riemer, C., Hardison, R.C., Burhans, R., Elnitski, L., Shah, P., Zhang, Y., Blankenberg, D., Albert, I., Taylor, J., et al. (2005). Galaxy: a platform for interactive large-scale genome analysis. Genome Res. *15*, 1451–1455.

Giulino-Roth, L., Wang, K., MacDonald, T.Y., Mathew, S., Tam, Y., Cronin, M.T., Palmer, G., Lucena-Silva, N., Pedrosa, F., Pedrosa, M., et al. (2012). Targeted genomic sequencing of pediatric Burkitt lymphoma identifies recurrent alterations in antiapoptotic and chromatin-remodeling genes. Blood *120*, 5181–5184.

Goecks, J., Nekrutenko, A., and Taylor, J. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. Genome Biol. *11*, R86.

Gold, M.R. (2008). B cell development: important work for ERK. Immunity *28*, 488–490.

Gorrini, C., Squatrito, M., Luise, C., Syed, N., Perna, D., Wark, L., Martinato, F., Sardella, D., Verrecchia, A., Bennett, S., et al. (2007). Tip60 is a haplo-insufficient tumour suppressor required for an oncogene-induced DNA damage response. Nature *448*, 1063–1067.

Gramling, M.W., and Eischen, C.M. (2012). Suppression of Ras/Mapk pathway signaling inhibits Myc-induced lymphomagenesis. Cell Death Differ. *19*, 1220–1227.

Grandori, C., Cowley, S.M., James, L.P., and Eisenman, R.N. (2000). The Myc/Max/Mad network and the transcriptional control of cell behavior. Annu. Rev. Cell Dev. Biol. *16*, 653–699.

Guglielmi, L., Le Bert, M., Comte, I., Dessain, M.L., Drouet, M., Ayer-Le Lievre, C., Cogné, M., and Denizot, Y. (2003). Combination of 3′ and 5′ IgH regulatory elements mimics the B-specific endogenous expression pattern of IgH genes from pro-B cells to mature B cells in a transgenic mouse model. Biochim. Biophys. Acta - Mol. Cell Res. *1642*, 181–190.

Gurel, B., Iwata, T., Koh, C.M., Jenkins, R.B., Lan, F., Van Dang, C., Hicks, J.L., Morgan, J., Cornish, T.C., Sutcliffe, S., et al. (2008). Nuclear MYC protein overexpression is an early alteration in human prostate carcinogenesis. Mod. Pathol. *21*, 1156–1167.

Habib, T., Park, H., Tsang, M., de Alborán, I.M., Nicks, A., Wilson, L., Knoepfler, P.S., Andrews, S., Rawlings, D.J., Eisenman, R.N., et al. (2007). Myc stimulates B lymphocyte differentiation and amplifies calcium signaling. J. Cell Biol. *179*, 717–731.

Hanahan, D. (1988). Dissecting multistep tumorigenesis in transgenic mice. Annu. Rev. Genet. *22*, 479–519.

Harrington, E.A., Bennett, M.R., Fanidi, A., and Evan, G.I. (1994). c-Myc-induced apoptosis in fibroblasts is inhibited by specific cytokines. EMBO J. *13*, 3286–3295.

Harris, A.W., Pinkert, C.A., Crawford, M., Langdon, W.Y., Brinster, R.L., and Adams, J.M. (1988). The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells. J. Exp. Med. *167*, 353–371.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Haupt, Y., Alexander, W.S., Barri, G., Klinken, S.P., and Adams, J.M. (1991). Novel zinc finger gene implicated as myc collaborator by retrovirally accelerated lymphomagenesis in E mu-myc transgenic mice. Cell *65*, 753–763.

Hayakawa, R., Hayakawa, T., Takeda, K., and Ichijo, H. (2012). Therapeutic targets in the ASK1-dependent stress signaling pathways. Proc. Jpn. Acad. Ser. B. Phys. Biol. Sci. *88*, 434–453.

Heese, K., Inoue, N., and Sawada, T. (2006). NF-kappaB regulates B-cell-derived nerve growth factor expression. Cell. Mol. Immunol. *3*, 63–66.

Hemann, M.T., Bric, A., Teruya-Feldstein, J., Herbst, A., Nilsson, J.A., Cordon-Cardo, C., Cleveland, J.L., Tansey, W.P., and Lowe, S.W. (2005). Evasion of the p53 tumour surveillance network by tumour-derived MYC mutants. Nature *436*, 807–811.

Herkert, B., and Eilers, M. (2010). Transcriptional repression: the dark side of myc. Genes Cancer *1*, 580–586.

Hoffman, B., and Liebermann, D. a (2008). Apoptotic signaling by c-MYC. Oncogene *27*, 6462–6472.

Howell, V.M. (2012). Sleeping beauty--a mouse model for all cancers? Cancer Lett. *317*, 1–8.

Ivics, Z., Hackett, P.B., Plasterk, R.H., and Izsvák, Z. (1997). Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. Cell *91*, 501–510.

Ivics, Z., Kaufman, C.D., Zayed, H., Miskey, C., Walisko, O., and Izsvák, Z. (2004). The Sleeping Beauty transposable element: evolution, regulation and genetic applications. Curr. Issues Mol. Biol. *6*, 43–55.

Iwakawa, R., Takenaka, M., Kohno, T., Shimada, Y., Totoki, Y., Shibata, T., Tsuta, K., Nishikawa, R., Noguchi, M., Sato-Otsubo, A., et al. (2013). Genome-wide

identification of genes with amplification and/or fusion in small cell lung cancer. Genes. Chromosomes Cancer *52*, 802–816.

Izsvák, Z., and Ivics, Z. (2004). Sleeping beauty transposition: biology and applications for molecular therapy. Mol. Ther. *9*, 147–156.

Izsvák, Z., Stüwe, E.E., Fiedler, D., Katzer, A., Jeggo, P.A., and Ivics, Z. (2004). Healing the wounds inflicted by sleeping beauty transposition by double-strand break repair in mammalian somatic cells. Mol. Cell *13*, 279–290.

Jacobs, J.J., Scheijen, B., Voncken, J.W., Kieboom, K., Berns, a, and van Lohuizen, M. (1999). Bmi-1 collaborates with c-Myc in tumorigenesis by inhibiting c-Myc-induced apoptosis via INK4a/ARF. Genes Dev. *13*, 2678–2690.

Janeway, C.A., Travers, P., Walport, M., and Shlomchik, M.J. (2001). Immunobiology: The Immune System in Health and Disease. (New York.: Garland Science).

Karreth, F.A., Tay, Y., Perna, D., Ala, U., Tan, S.M., Rust, A.G., DeNicola, G., Webster, K.A., Weiss, D., Perez-Mancera, P.A., et al. (2011). In vivo identification of tumor-suppressive PTEN ceRNAs in an oncogenic BRAF-induced mouse model of melanoma. Cell *147*, 382–395.

Keller, U., Nilsson, J. a, Maclean, K.H., Old, J.B., and Cleveland, J.L. (2005). Nfkb 1 is dispensable for Myc-induced lymphomagenesis. Oncogene *24*, 6231–6240.

Keller, U., Huber, J., Nilsson, J.A., Fallahi, M., Hall, M.A., Peschel, C., and Cleveland, J.L. (2010). Myc suppression of Nfkb2 accelerates lymphomagenesis. BMC Cancer *10*, 348.

Keng, V.W., Villanueva, A., Chiang, D.Y., Dupuy, A.J., Ryan, B.J., Matise, I., Silverstein, K.A.T., Sarver, A., Starr, T.K., Akagi, K., et al. (2009). A conditional transposon-based insertional mutagenesis screen for genes associated with mouse hepatocellular carcinoma. Nat. Biotechnol. *27*, 264–274.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010). A Myc network accounts for similarities between embryonic stem and cancer cell transcription programs. Cell *143*, 313–324.

Kool, J., and Berns, A. (2009). High-throughput insertional mutagenesis screens in mice to identify oncogenic networks. Nat. Rev. Cancer *9*, 389–399.

Koso, H., Takeda, H., Yew, C.C.K., Ward, J.M., Nariai, N., Ueno, K., Nagasaki, M., Watanabe, S., Rust, A.G., Adams, D.J., et al. (2012). Transposon mutagenesis identifies genes that transform neural stem cells into glioma-initiating cells. Proc. Natl. Acad. Sci. U. S. A. *109*, E2998–3007.

Koudijs, M.J., Klijn, C., van der Weyden, L., Kool, J., ten Hoeve, J., Sie, D., Prasetyanti, P.R., Schut, E., Kas, S., Whipp, T., et al. (2011). High-throughput semiquantitative analysis of insertional mutations in heterogeneous tumors. Genome Res. *21*, 2181–2189.

Kovalchuk, A.L., Qi, C.F., Torrey, T.A., Taddesse-Heath, L., Feigenbaum, L., Park, S.S., Gerbitz, A., Klobeck, G., Hoertnagel, K., Polack, A., et al. (2000). Burkitt lymphoma in the mouse. J. Exp. Med. *192*, 1183–1190.

Land, H., Parada, L., and Weinberg, R. (1983). Tumorigenic conversion of primary embryo fibroblasts requires at least two cooperating oncogenes. Nature *304*, 596–602.

Langdon, W.Y., Harris, A.W., Cory, S., and Adams, J.M. (1986). The c-myc oncogene perturbs B lymphocyte development in E-mu-myc transgenic mice. Cell *47*, 11–18.

Largaespada, D.A., Brannan, C.I., Jenkins, N.A., and Copeland, N.G. (1996). Nf1 deficiency causes Ras-mediated granulocyte/macrophage colony stimulating factor hypersensitivity and chronic myeloid leukaemia. Nat. Genet. *12*, 137–143.

Levine, A.J. (1997). p53, the cellular gatekeeper for growth and division. Cell *88*, 323–331.

Lin, F.-R., Huang, S.-Y., Hung, K.-H., Su, S.-T., Chung, C.-H., Matsuzawa, A., Hsiao, M., Ichijo, H., and Lin, K.-I. (2012). ASK1 promotes apoptosis of normal and malignant plasma cells. Blood *120*, 1039–1047.

Littlewood, T.D., Kreuzaler, P., and Evan, G.I. (2012). All things to all people. Cell *151*, 11–13.

Liu, G., Geurts, A.M., Yae, K., Srinivasan, a R., Fahrenkrug, S.C., Largaespada, D. a, Takeda, J., Horie, K., Olson, W.K., and Hackett, P.B. (2005). Target-site preferences of Sleeping Beauty transposons. J. Mol. Biol. *346*, 161–173.

Liu, Q., Basu, S., Qiu, Y., Tang, F., and Dong, F. (2010). A role of Miz-1 in Gfi-1-mediated transcriptional repression of CDKN1A. Oncogene *29*, 2843–2852.

Liu, X., Vorontchikhina, M., Wang, Y.-L., Faiola, F., and Martinez, E. (2008). STAGA recruits Mediator to the MYC oncoprotein to stimulate transcription and cell proliferation. Mol. Cell. Biol. *28*, 108–121.

Lohr, J.G., Stojanov, P., Lawrence, M.S., Auclair, D., Chapuy, B., Sougnez, C., Cruz-Gordillo, P., Knoechel, B., Asmann, Y.W., Slager, S.L., et al. (2012). Discovery and prioritization of somatic mutations in diffuse large B-cell lymphoma (DLBCL) by whole-exome sequencing. Proc. Natl. Acad. Sci. U. S. A. *109*, 3879–3884.

Van Lohuizen, M., Verbeek, S., Scheijen, B., Wientjens, E., van der Gulden, H., and Berns, a (1991). Identification of cooperating oncogenes in E mu-myc transgenic mice by provirus tagging. Cell *65*, 737–752.

Love, C., Sun, Z., Jima, D., Li, G., Zhang, J., Miles, R., Richards, K.L., Dunphy, C.H., Choi, W.W.L., Srivastava, G., et al. (2012). The genetic landscape of mutations in Burkitt lymphoma. Nat. Genet. *44*, 1321–1325.

Lowe, S.W., Cepero, E., and Evan, G. (2004). Intrinsic tumour suppression. Nature *432*, 307–315.

Lüscher, B., and Larsson, L.G. (1999). The basic region/helix-loop-helix/leucine zipper domain of Myc proto-oncoproteins: function and regulation. Oncogene *18*, 2955–2966.

Maclean, K.H., Kastan, M.B., and Cleveland, J.L. (2007). Atm deficiency affects both apoptosis and proliferation to augment Myc-induced lymphomagenesis. Mol. Cancer Res. *5*, 705–711.

Mann, K.M., Ward, J.M., Yew, C.C.K., Kovochich, A., Dawson, D.W., Black, M.A., Brett, B.T., Sheetz, T.E., Dupuy, A.J., Chang, D.K., et al. (2012). Sleeping Beauty mutagenesis reveals cooperating mutations and pathways in pancreatic adenocarcinoma. Proc. Natl. Acad. Sci. U. S. A. *109*, 5934–5941.

March, H.N., Rust, A.G., Wright, N. a, ten Hoeve, J., de Ridder, J., Eldridge, M., van der Weyden, L., Berns, A., Gadiot, J., Uren, A., et al. (2011). Insertional mutagenesis identifies multiple networks of cooperating genes driving intestinal tumorigenesis. Nat. Genet. *43*, 1202–1209.

Martinez, E., Kundu, T.K., Fu, J., and Roeder, R.G. (1998). A human SPT3-TAFII31-GCN5-L acetylase complex distinct from transcription factor IID. J. Biol. Chem. *273*, 23781–23785.

Mattison, J., van der Weyden, L., Hubbard, T., and Adams, D.J. (2009). Cancer gene discovery in mouse and man. Biochim. Biophys. Acta *1796*, 140–161.

Mi, H., Muruganujan, A., and Thomas, P.D. (2013). PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. Nucleic Acids Res. *41*, D377–86.

Mikkers, H., Allen, J., Knipscheer, P., Romeijn, L., Hart, A., Vink, E., Berns, A., and Romeyn, L. (2002). High-throughput retroviral tagging to identify components of specific signaling pathways in cancer. Nat. Genet. *32*, 153–159.

Miyama-Inaba, M., Kuma, S., Inaba, K., Ogata, H., Iwai, H., Yasumizu, R., Muramatsu, S., Steinman, R.M., and Ikehara, S. (1988). Unusual phenotype of B cells in the thymus of normal mice. J. Exp. Med. *168*, 811–816.

Mori, S., Rempel, R.E., Chang, J.T., Yao, G., Lagoo, A.S., Potti, A., Bild, A., and Nevins, J.R. (2008). Utilization of pathway signatures to reveal distinct types of B lymphoma in the Emicro-myc model and human diffuse large B-cell lymphoma. Cancer Res. *68*, 8525–8534.

Morin, R.D., Mungall, K., Pleasance, E., Mungall, A.J., Goya, R., Huff, R.D., Scott, D.W., Ding, J., Roth, A., Chiu, R., et al. (2013). Mutational and structural analysis of diffuse large B-cell lymphoma using whole-genome sequencing. Blood *122*, 1256–1265.

Morton, J.P., and Sansom, O.J. (2013). MYC-y mice: from tumour initiation to therapeutic targeting of endogenous MYC. Mol. Oncol. *7*, 248–258.

Mumert, M., Dubuc, A., Wu, X., Northcott, P. a, Chin, S.S., Pedone, C. a, Taylor, M.D., and Fults, D.W. (2012). Functional genomics identifies drivers of medulloblastoma dissemination. Cancer Res. *72*, 4944–4953.

Muñoz-López, M., and García-Pérez, J.L. (2010). DNA transposons: nature and applications in genomics. Curr. Genomics *11*, 115–128.

Nilsson, J.A., and Cleveland, J.L. (2003). Myc pathways provoking cell suicide and cancer. Oncogene *22*, 9007–9021.

Noguchi, K., Kokubu, A., Kitanaka, C., Ichijo, H., and Kuchino, Y. (2001). ASK1-Signaling Promotes c-Myc Protein Stability during Apoptosis. Biochem. Biophys. Res. Commun. *281*, 1313–1320.

O'Donnell, K. a, Keng, V.W., York, B., Reineke, E.L., Seo, D., Fan, D., Silverstein, K. a T., Schrum, C.T., Xie, W.R., Mularoni, L., et al. (2012). A Sleeping Beauty mutagenesis

screen reveals a tumor suppressor role for Ncoa2/Src-2 in liver cancer. Proc. Natl. Acad. Sci. U. S. A. *109*, E1377–86.

Oliveros, J. (2007). VENNY. An interactive tool for comparing lists with Venn diagrams. BioinfoGP, CNB-CSIC.

Oster, S.K., Ho, C.S.W., Soucie, E.L., and Penn, L.Z. (2002). The myc oncogene: MarvelouslY Complex. Adv. Cancer Res. *84*, 81–154.

Ott, G., Rosenwald, A., and Campo, E. (2013). Understanding MYC-driven aggressive B-cell lymphomas: pathogenesis and classification. Hematology Am. Soc. Hematol. Educ. Program *2013*, 575–583.

Park, S.S., Kim, J.S., Tessarollo, L., Owens, J.D., Peng, L., Han, S.S., Tae Chung, S., Torrey, T.A., Cheung, W.C., Polakiewicz, R.D., et al. (2005). Insertion of c-Myc into Igh induces B-cell and plasma-cell neoplasms in mice. Cancer Res. *65*, 1306–1315.

Rahl, P.B., Lin, C.Y., Seila, A.C., Flynn, R.A., McCuine, S., Burge, C.B., Sharp, P.A., and Young, R.A. (2010). c-Myc regulates transcriptional pause release. Cell *141*, 432–445.

Rahrmann, E.P., Collier, L.S., Knutson, T.P., Doyal, M.E., Kuslak, S.L., Green, L.E., Malinowski, R.L., Roethe, L., Akagi, K., Waknitz, M., et al. (2009). Identification of PDE4D as a proliferation promoting factor in prostate cancer using a Sleeping Beauty transposon-based somatic mutagenesis screen. Cancer Res. *69*, 4388–4397.

Rajewsky, K. (1996). Clonal selection and learning in the antibody system. Nature *381*, 751–758.

Ravitz, M.J., Chen, L., Lynch, M., and Schmidt, E. V (2007). c-myc Repression of TSC2 contributes to control of translation initiation and Myc-induced transformation. Cancer Res. *67*, 11209–11217.

Rehm, A., Mensen, A., Schradi, K., Gerlach, K., Wittstock, S., Winter, S., Büchner, G., Dörken, B., Lipp, M., and Höpken, U.E. (2011). Cooperative function of CCR7 and lymphotoxin in the formation of a lymphoma-permissive niche within murine secondary lymphoid organs. Blood *118*, 1020–1033.

Reimann, M., Loddenkemper, C., Rudolph, C., Schildhauer, I., Teichmann, B., Stein, H., Schlegelberger, B., Do, B., and Schmitt, C.A. (2008). The Myc-evoked DNA damage response accounts for treatment resistance in primary lymphomas in vivo. *110*.

Reimann, M., Lee, S., Loddenkemper, C., Dörr, J.R., Tabor, V., Aichele, P., Stein, H., Dörken, B., Jenuwein, T., and Schmitt, C. a (2010). Tumor stroma-derived TGF-beta limits myc-driven lymphomagenesis via Suv39h1-dependent senescence. Cancer Cell *17*, 262–272.

Richter, J., Schlesner, M., Hoffmann, S., Kreuz, M., Leich, E., Burkhardt, B., Rosolowski, M., Ammerpohl, O., Wagener, R., Bernhart, S.H., et al. (2012). Recurrent mutation of the ID3 gene in Burkitt lymphoma identified by integrated genome, exome and transcriptome sequencing. Nat. Genet. *44*, 1316–1320.

Rickert, R.C., Roes, J., and Rajewsky, K. (1997). B lymphocyte-specific, Cre-mediated mutagenesis in mice. Nucleic Acids Res. *25*, 1317–1318.

De Ridder, J., Uren, A., Kool, J., Reinders, M., and Wessels, L. (2006). Detecting statistically significant common insertion sites in retroviral insertional mutagenesis screens. PLoS Comput. Biol. *2*, e166.

Van Riggelen, J., Müller, J., Otto, T., Beuger, V., Yetil, A., Choi, P.S., Kosan, C., Möröy, T., Felsher, D.W., and Eilers, M. (2010). The interaction between Myc and Miz1 is required to antagonize TGFbeta-dependent autocrine signaling during lymphoma formation and maintenance. Genes Dev. *24*, 1281–1294.

Sabò, A., and Amati, B. (2014). Genome Recognition by MYC. Cold Spring Harb. Perspect. Med. *4*.

Sander, S., Calado, D.P., Srinivasan, L., Köchert, K., Zhang, B., Rosolowski, M., Rodig, S.J., Holzmann, K., Stilgenbauer, S., Siebert, R., et al. (2012). Synergy between PI3K signaling and MYC in Burkitt lymphomagenesis. Cancer Cell *22*, 167–179.

Scheijen, B., Jonkers, J., Acton, D., and Berns, A. (1997). Characterization of pal-1, a common proviral insertion site in murine leukemia virus-induced lymphomas of c-myc and Pim-1 transgenic mice. J. Virol. *71*, 9–16.

Schmidt, T., Karsunky, H., Gau, E., Zevnik, B., Elsässer, H.P., and Möröy, T. (1998). Zinc finger protein GFI-1 has low oncogenic potential but cooperates strongly with pim and myc genes in T-cell lymphomagenesis. Oncogene *17*, 2661–2667.

Schmitt, C.A., and Lowe, S.W. (2002). Apoptosis and chemoresistance in transgenic cancer models. J. Mol. Med. (Berl). *80*, 137–146.

Schmitt, C.A., McCurrach, M.E., de Stanchina, E., Wallace-Brodeur, R.R., and Lowe, S.W. (1999). INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. Genes Dev. *13*, 2670–2677.

Schmitt, C.A., Fridman, J.S., Yang, M., Baranov, E., Hoffman, R.M., and Lowe, S.W. (2002). Dissecting p53 tumor suppressor functions in vivo. Cancer Cell *1*, 289–298.

Schmitz, R., Young, R.M., Ceribelli, M., Jhavar, S., Xiao, W., Zhang, M., Wright, G., Shaffer, A.L., Hodson, D.J., Buras, E., et al. (2012). Burkitt lymphoma pathogenesis and therapeutic targets from structural and functional genomics. Nature *490*, 116–120.

Selten, G., Cuypers, H.T., Zijlstra, M., Melief, C., and Berns, A. (1984). Involvement of c-myc in MuLV-induced T cell lymphomas in mice: frequency and mechanisms of activation. EMBO J. *3*, 3215–3222.

Seoane, J., Le, H.-V., and Massagué, J. (2002). Myc suppression of the p21(Cip1) Cdk inhibitor influences the outcome of the p53 response to DNA damage. Nature *419*, 729–734.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. Cell *88*, 593–602.

Sherr, C.J. (2001). The INK4a/ARF network in tumour suppression. Nat. Rev. Mol. Cell Biol. *2*, 731–737.

Shindo, H., Tani, E., Matsumuto, T., Hashimoto, T., and Furuyama, J. (1993). Stabilization of c-myc protein in human glioma cells. Acta Neuropathol. *86*, 345–352.

Shreeram, S., Hee, W.K., Demidov, O.N., Kek, C., Yamaguchi, H., Fornace, A.J., Anderson, C.W., Appella, E., and Bulavin, D. V (2006). Regulation of ATM/p53-dependent suppression of myc-induced lymphomas by Wip1 phosphatase. J. Exp. Med. *203*, 2793–2799.

Sluss, H.K., Gannon, H., Coles, A.H., Shen, Q., Eischen, C.M., and Jones, S.N. (2010). Phosphorylation of p53 serine 18 upregulates apoptosis to suppress Myc-induced tumorigenesis. Mol. Cancer Res. *8*, 216–222.

Srinivas, S., Watanabe, T., Lin, C.S., William, C.M., Tanabe, Y., Jessell, T.M., and Costantini, F. (2001). Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev. Biol. *1*, 4.

Starr, T.K., Allaei, R., Silverstein, K.A.T., Staggs, R. a, Sarver, A.L., Bergemann, T.L., Gupta, M., O'Sullivan, M.G., Matise, I., Dupuy, A.J., et al. (2009). A transposon-based genetic screen in mice identifies genes altered in colorectal cancer. Science *323*, 1747–1750.

Starr, T.K., Scott, P.M., Marsh, B.M., Zhao, L., Than, B.L.N., O'Sullivan, M.G., Sarver, A.L., Dupuy, A.J., Largaespada, D.A., and Cormier, R.T. (2011). A Sleeping Beauty transposon-mediated screen identifies murine susceptibility genes for adenomatous polyposis coli (Apc)-dependent intestinal tumorigenesis. Proc. Natl. Acad. Sci. U. S. A. *108*, 5765–5770.

Strasser, A., Harris, A.W., Bath, M.L., and Cory, S. (1990). Novel primitive lymphoid tumours induced in transgenic mice by cooperation between myc and bcl-2. Nature *348*, 331–333.

Tang, J.Z., Carmichael, C.L., Shi, W., Metcalf, D., Ng, A.P., Hyland, C.D., Jenkins, N.A., Copeland, N.G., Howell, V.M., Zhao, Z.J., et al. (2013). Transposon mutagenesis reveals cooperation of ETS family transcription factors with signaling pathways in erythro-megakaryocytic leukemia. Proc. Natl. Acad. Sci. U. S. A. *110*, 6091–6096.

Tiller, T., Busse, C.E., and Wardemann, H. (2009). Cloning and expression of murine Ig genes from single B cells. J. Immunol. Methods *350*, 183–193.

Uren, A., and Berns, A. (2009). Jump-starting cancer gene discovery Combinatorial stem cell mobilization. *27*, 251–252.

Uren, a G., Kool, J., Berns, a, and van Lohuizen, M. (2005). Retroviral insertional mutagenesis: past, present and future. Oncogene *24*, 7656–7672.

Vallespinós, M., Fernández, D., Rodríguez, L., Alvaro-Blanco, J., Baena, E., Ortiz, M., Martínez, D., Rojas, A., Miguel, R., Dukovska, D., et al. (2011). B Lymphocyte commitment program is driven by the proto-oncogene c-Myc. J. Immunol. *186*, 6726–6736.

Vaqué, J.P., Navascues, J., Shiio, Y., Laiho, M., Ajenjo, N., Mauleon, I., Matallanas, D., Crespo, P., and León, J. (2005). Myc antagonizes Ras-mediated growth arrest in leukemia cells through the inhibition of the Ras-ERK-p21Cip1 pathway. J. Biol. Chem. *280*, 1112–1122.

Vassiliou, G., Rad, R., and Bradley, A. (2010). The use of DNA transposons for cancer gene discovery in mice. Methods Enzymol. *477*, 91–106.

Vigorito, E., Gambardella, L., Colucci, F., McAdam, S., and Turner, M. (2005). Vav proteins regulate peripheral B-cell survival. Blood *106*, 2391–2398.

Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L. a, and Kinzler, K.W. (2013). Cancer genome landscapes. Science *339*, 1546–1558.

Wall, M., Poortinga, G., Stanley, K.L., Lindemann, R.K., Bots, M., Chan, C.J., Bywater, M.J., Kinross, K.M., Astle, M. V, Waldeck, K., et al. (2013). The mTORC1 inhibitor everolimus prevents and treats Eμ-Myc lymphoma by restoring oncogene-induced senescence. Cancer Discov. *3*, 82–95.

Van der Weyden, L., Papaspyropoulos, A., Poulogiannis, G., Rust, A.G., Rashid, M., Adams, D.J., Arends, M.J., and O'Neill, E. (2012a). Loss of RASSF1A synergizes with deregulated RUNX2 signaling in tumorigenesis. Cancer Res. *72*, 3817–3827.

Van der Weyden, L., Arends, M.J., Rust, A.G., Poulogiannis, G., McIntyre, R.E., and Adams, D.J. (2012b). Increased tumorigenesis associated with loss of the tumor suppressor gene Cadm1. Mol. Cancer *11*, 29.

Yant, S.R., and Kay, M.A. (2003). Nonhomologous-end-joining factors regulate DNA repair fidelity during Sleeping Beauty element transposition in mammalian cells. Mol. Cell. Biol. *23*, 8505–8518.

Yant, S.R., Wu, X., Huang, Y., Garrison, B., Burgess, S.M., and Kay, M.A. (2005). High-resolution genome-wide mapping of transposon integration in mammals. Mol. Cell. Biol. *25*, 2085–2094.

Yu, D., and Thomas-Tikhonenko, A. (2002). A non-transgenic mouse model for B-cell lymphoma: in vivo infection of p53-null bone marrow progenitors by a Myc retrovirus is sufficient for tumorigenesis. Oncogene *21*, 1922–1927.

Yukawa, K., Kikutani, H., Inomoto, T., Uehira, M., Bin, S.H., Akagi, K., Yamamura, K., and Kishimoto, T. (1989). Strain dependency of B and T lymphoma development in immunoglobulin heavy chain enhancer (E mu)-myc transgenic mice. J. Exp. Med. *170*, 711–726.

Zambrowicz, B.P., Imamoto, A., Fiering, S., Herzenberg, L.A., Kerr, W.G., and Soriano, P. (1997). Disruption of overlapping transcripts in the ROSA  geo 26 gene trap strain leads to widespread expression of  -galactosidase in mouse embryos and hematopoietic cells. Proc. Natl. Acad. Sci. *94*, 3789–3794.

Zayed, H., Izsvák, Z., Khare, D., Heinemann, U., and Ivics, Z. (2003). The DNA-bending protein HMGB1 is a cellular cofactor of Sleeping Beauty transposition. Nucleic Acids Res. *31*, 2313–2322.

Zhang, J., Grubor, V., Love, C.L., Banerjee, A., Richards, K.L., Mieczkowski, P.A., Dunphy, C., Choi, W., Au, W.Y., Srivastava, G., et al. (2013). Genetic heterogeneity of diffuse large B-cell lymphoma. Proc. Natl. Acad. Sci. U. S. A. *110*, 1398–1403.

Zhang, Y., Xiong, Y., and Yarbrough, W.G. (1998). ARF promotes MDM2 degradation and stabilizes p53: ARF-INK4a locus deletion impairs both the Rb and p53 tumor suppression pathways. Cell *92*, 725–734.

Zindy, F., Eischen, C.M., Randle, D.H., Kamijo, T., Cleveland, J.L., Sherr, C.J., and Roussel, M.F. (1998). Myc signaling via the ARF tumor suppressor regulates p53-dependent apoptosis and immortalization. Genes Dev. *12*, 2424–2433.