Using advanced CRISPR-Cas9 functional genetics to model and probe T cell acute lymphoblastic leukemia (T-ALL)

A thesis submitted for the degree of

Doctor of Natural Sciences (Dr. rer. nat.)

in the subjects of *Hematology* and *Oncology*

by

Maximilian Schinke, M.Sc. Biomedicine

April 2023

Hannover Medical School

International MD/PhD Program "Molecular Medicine"

at Hannover Biomedical Research School (HBRS)

Institute of Experimental Hematology

Acknowledged by the MD / PhD committee and head of Hannover Medical School

President:	**Prof. Dr. med. Michael P. Manns** President of Hannover Medical School, Hannover, Germany
Supervisor:	**Dr. med. Adrian Schwarzer, PhD** Institute of Experimental Hematology, Hannover Medical School, Hannover, Germany
	Prof. Dr. med. Axel Schambach, PhD Institute of Experimental Hematology, Hannover Medical School, Hannover, Germany
Co-supervisors:	**Prof. Dr. med. Michael Heuser** Department of Hematology, Hemostasis, Oncology and Stem Cell Transplantation, Hannover Medical School, Hannover, Germany
	Prof. Dr. rer. nat. Dirk Heckl Pediatric Hematology and Oncology, Martin-Luther-University Halle-Wittenberg, Halle, Germany
External expert:	**Prof. Dr. rer. nat. Tony Gutschner** Institute of Molecular Medicine, Martin-Luther-University Halle-Wittenberg, Halle, Germany
Internal expert:	**PD Dr. med. Anna Saborowski,** Department of Gastroenterology, Hepatology, Infectious Diseases and Endocrinology, Hannover Medical School, Germany

Day of final exam / public defense: **16th of June 2023**

List of content

1 Abstract

T cell acute lymphoblastic leukemia (T-ALL) is an aggressive malignancy of the hematopoietic system characterized by the invasion of immature lymphatic blasts into the bone marrow (BM) and other peripheral organs. It accounts for ~15% of acute lymphoblastic leukemia (ALL) cases in children and ~25% in adults, with a five-year survival rate of ~80% in children and ~40-50% in adults. T-ALL is an extremely genetically diverse disease, reflecting different stages and genetic dependencies of T cell maturation in the thymus. This genetic complexity of T-ALL is a major cause of relapse and the resulting low patient survival rates.

For a long time, T-ALL has been modeled by the overexpression of certain disease-associated genes, such as *Jak3*, *Tal1*, *Notch1*, *Tlx1*, and *Il7r*. However, these models do not reflect the genetic and clinical heterogeneity and the variable therapeutic responses of the disease, which limits the conclusions that can be drawn from them. Therefore, there is a great need for new pre-clinical models that replicate the wide range of observed genetic alterations known from high-throughput sequencing studies of T-ALL patients in a more realistic and explorable murine model system.

To address this, in this work the *LSL.Cas9 x Lck-cre* mouse strain was developed that specifically expresses Cre and Cas9 in thymocytes, providing the ability to directly target the onset of multiple oncogenic insults to developing T cells. Comprehensive characterization of Cre and Cas9 expression and intact T cell differentiation potential demonstrated high T cell specificity, reducing the risk of lineage-nonspecific activation. Transduction of the mouse strains' hematopoietic stem and progenitor cells with a floxed inverse vector design (*lox66/lox71* sites), in which murine transcription factors and human mutant signaling proteins require activation by Cre (type A aberration), and sgRNA-encoding vectors introduce defined genetic alterations (type B aberration), led to efficient development of T-ALL in transplanted mice. In particular, *Tlx1.NRAS*^{G12D}, *Tlx1.PIK3CD*^{E1021K}, *Tlx1.IL7R*$^{c.731\_732insTTGTCCCAC}$, and *Tal1.Lmo2* were combined with 6 – 8 sgRNAs, targeting subtype-related vulnerable genes. The transforming potential of the type A vectors was previously verified by transforming IL3-dependent BAF3 and 32D cell lines. High on-target activity of used sgRNAs was achieved by testing the candidates in a reporter-based assay. Using this Cre/loxP and CRISPR-Cas9-based animal platform, several hundred genetic permutations could be screened for oncogenic cooperativity *in vivo*.

The effort of this work was focused on three different T-ALL subtypes. The phenotypes comprised immature/ early cortical CD4$^-$CD8$^-$ T-ALL (*Tlx1.NRas*^{G12D} + sgRNAs), the cortical

$CD3^{low}CD4^{+}CD8^{+}$ stage (*Tlx1.PIK3CD*^{E1021K} + sgRNAs) followed by the more mature $CD3^{+}CD8^{+}$ T-ALL subgroup (*Tal1.Lmo2* + sgRNAs). The resulting T-ALL contained up to 8 different disease-relevant genetic alterations, recapitulating the known complexity from the clinics and overcoming the long latency for spontaneous mutations in less complex mouse models. Among these, co-evolution of insertions and deletions in *Notch1*, *Cdkn2a*, *Bcl11b*, and *Pten* were among the most frequently observed mutations in definitive T-ALL, whereas others, e.g., in *Dnm2*, *Phf6*, *Etv6*, and *Lef1* occurred more randomly. These type B alterations were identified as the main drivers for proliferation, cell migration and disease onset, whereas the corresponding type A vectors determined the T cell phenotypical stage. The resulting T-ALL were marked by the upregulation off cell cycle genes and the Notch pathway and recapitulated disease formation after secondary transplantation. Co-cultivation with OP9 stromal cells facilitated *in vitro* cultivation of T-ALL cells, allowing the generation of various T-ALL cell lines. The culturing system only slightly altered the immune phenotypic characteristics and the underlying genetic aberrations, whereas analysis of the transcriptome revealed more changes, including increased mTOR signaling pathway and decreased differentiation markers.

In summary, we developed a functional genetics pipeline to model T-ALL *in vivo* by selectively multiplexing CRISPR-Cas9-mediated gene editing and Cre/loxP-mediated overexpression of transcription factors and signaling proteins. This novel T-ALL resource will help to understand disease biology and provides a clinically relevant screening platform for novel therapeutic strategies.

2 Introduction

2.1 The hematopoietic system

The hematopoietic system is a hierarchical and complex network that facilitates the lifelong production of all blood cells. For a long time, it was considered as a strict hierarchical hematopoietic tree, where long-term hematopoietic stem cells (LT-HSC) stand at the apex of the hierarchy (Figure 1)[1]. These quiescent cells have the capability to regenerate themselves, a potential that disappears throughout the branch of multipotent and oligopotent progenitor cells, following a progressing lineage restriction and increasing specialization in their functionality[2]. HSCs reside in a special bone marrow (BM) niche together with mesenchymal stromal cells (MSC), osteoblasts and endothelial and perivascular cells, which support HSC quiescence and maintenance[3,4]. These cells secrete different growth factors and cytokines and orchestrate the steady-state hematopoiesis[5], including C-X-C motif chemokine ligand 12 (CXCL12) and stem cell factor (SCF) that promote HSC retention and maintenance in the bone marrow (BM) through interaction with their receptors C-X-C motif chemokine receptor 4 (CXCR4) and c-KIT (CD117)[6]. Among others, Fms-like tyrosine kinase-3 ligand (FLT3-L)[7] and Thrombopoietin (TPO)[8] support self-renewal and survival of HSCs, while they as well support further differentiation pathways in more lineage-restricted progenitors. Only 1% of total BM cells include hematopoietic stem and progenitor cells (HSPC). In mice, they can be identified by the lack of lineage surface markers (Lin^-) and the expression of c-KIT and SCA-1 proteins on the cell surface. Within these cells, only those containing the signaling lymphocytic activation molecule family markers $CD150^+$ (SLAMF1) and lacking cluster of differentiation 48 (CD48) expression are identified as LT-HSCs[9]. Besides the microenvironment, the expression of stem-cell genes is another key factor in controlling HSC preservation. Expression of the transcription factors (TF) stem cell leukemia (SCL/TAL1)[10], hepatic leukemia factor (HLF)[11], LIM-domain-only 2 (LMO2)[12], GATA-binding protein 2 (GATA2)[13], the homeobox family[14] and the Notch complex[15] regulate, among many others, the maintenance of HSC. Dysregulation of their expression often correlates with malignant transformation into leukemia and other hematological malignancies.

In the classical tree-like structure, most LT-HSCs persist in a quiescent, reconstituting state[16]. In response to external stimuli like infections, bleeding or stress they undergo symmetric and asymmetric cell division, where they differentiate into short-term HSCs (ST-HSC)[1,17]. ST-HSCs have a shorter lifetime and further differentiate into so-called multipotent progenitors (MPP), which branch to myeloid and lymphoid oligopotent progenitor cells determining the two main lineages of hematopoiesis. Here, common lymphoid progenitors (CLP) give rise to lymphoid populations, including T cells, B cells, and natural killer (NK) cells. In contrast, common myeloid

progenitors (CMP) form either megakaryocyte-erythrocyte progenitors (MEP), from which erythrocytes and platelets arise, or granulocyte-monocyte progenitors (GMP) differentiating into granulocytes, mast cells and monocytes[18,19]. However, this strict and hierarchical concept oversimplifies the complexity of hematopoiesis. Recent advances in single-cell ribonucleic acid (RNA) sequencing and limiting-dilution assays identified a bias of HSCs towards specific lineages[20]. It seems that within the population of HSCs, some subtypes are more prone to differentiate towards megakaryocytes, granulocytes or lymphoid progenitor cells. In addition, various intermediate stages between the two major lineages indicate a less strict but flexible differentiation pathway[20–22].

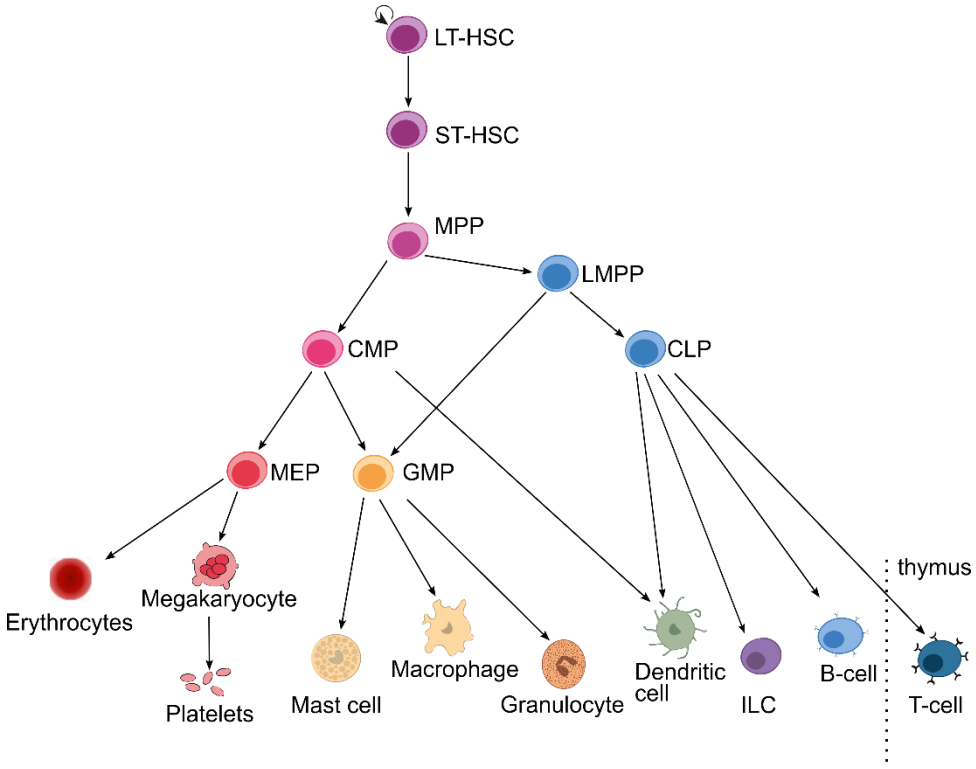


Figure 1: Overview of the hematopoietic system. Hematopoiesis is a hierarchical system with long-term hematopoietic stem cells at its apex (LT-HSCs). In classical perception, it is considered as a tree-like structure along whose branches (black arrows) the self-renewal capacity decreases and cell functions become more specialized. Upon stimulation, LT-HSCs differentiate into short-term HSCs (ST-HSC), which give rise to multipotent progenitors (MPP). These MPPs determine the myeloid lineage, by generating oligopotent common myeloid progenitors (CMP), megakaryocyte/erythrocyte progenitors (MEP), and granulocyte/macrophage progenitors (GMP) or the lymphoid lineage through lymphoid-primed multipotent progenitors (LMPP) and common lymphoid progenitors (CLP). The oligopotent progenitor cells give rise to all the mature blood cells, which subsequently pass to the blood-stream. ILC, innate lymphoid cells. The illustration was adapted and modified from Wilson *et al.*[23] (available under Public License CC BY 4.0; Copyright © 2015 by the Authors. Published by Elsevier Inc.)

2.2 Development of healthy T cells inside the thymus

T cell development starts with the migration of CLPs or lymphoid-primed multipotent progenitors (LMPPs) from the BM niche into the thymus. At this early stage, early T cell progenitors (ETP) still maintain the potential to differentiate towards the myeloid, B-, NK-, and dendritic cell lineage and are therefore not entirely restricted to the T cell fate yet. The murine thymic environment provides all signals to drive T cell development, including SCF, FLT3-L, interleukin 7 (IL7), and Notch ligands[24], thus committing T cell identity.

In mice, the early stages of differentiation are named double-negative (DN) as they lack expression of CD4 and CD8 and can be further discriminated based on the up- and downregulation of CD44 and CD25 (Figure 2). The least mature stage is referred as DN1 stage ($DN44^+$ $CD25^-$) and can further be characterized by the high expression and c-KIT and low Il7 receptor (IL7R) expression. Within this early stage, ETPs can be identified by the additional expression of FLT3[25]. Various stem cell-associated genes are still expressed at this stage, including *Erg*, *Hoxa9*, *CD34*, *Mef2c*, *Bcl11a*, *Lyl1*, *Spi1* (encoding PU.1) and *Lmo2* that help to maintain the early T cell phenotype[24,26]. The following DN2 stage ($CD44^+$ $CD25^+$) is marked by the upregulation of interleukin 2 receptor alpha chain (IL2Ra or CD25) and IL7R. Notch and IL7R-signaling play an essential role in these two phases (see more details in section 2.3), managing an excessive proliferation rate of the thymocytes[27]. In the early phase, the DN2a stage, *Tcf7* (encoding the TCF1 transcription factor) responds to Notch signaling, initiating *Runx1*, *Bcl11b* and *Gata3* expression. As a result of this cascade, T cell fate is enforced, while alternative lineage fates such as differentiation towards dendritic cells, granulocytes, macrophages, NK cells or mast cells are excluded[26,27]. The down-regulation of c-KIT marks the beginning of DN2b phase. Committed T cells become more independent of IL7R signaling while relying more on Notch signaling. Nevertheless, IL7R activity still contributes to cell survival[27]. In the following DN3 stage, CD3 chains are expressed intracellularly together with recombination activating 1 (RAG1) protein that mediates the V(D)J-recombination of the T cell receptor (TCR) β chain locus. T cells that failed to generate a functional pre-TCR undergo apoptosis in a process called “β-selection”[28]. From this point, pre-TCR signaling can bypass Notch signaling, leading to the downregulation of Notch target genes. When entering the DN4 stage ($CD44^-$ $CD25^-$), the *Il2ra* gene is downregulated and cells undergo an extensive proliferation mediated by CXCL12, TCF1, and LEF1[27].

Before reaching the CD4 and CD8 double-positive (DP) stage, cells have to pass through the immature single-positive (ISP) stage, where thymocytes only express $CD8^{low}$ in mice or $CD4^{low}$ in humans[29,30]. Merely upon the activation of the transcription factor bromodomain protein 4 (BRD4), T cells perform a single cell division, transit to the DP stage and undergo recombination of the TCR α chain (TCRα) locus[29]. During this transition, T cells completely

lose their Notch dependency for proliferation and survival[31]. The early DP phase at which TCRα recombines and assembles to functional TCRαβ is marked by the lack of extracellular CD3 [32]. At a later DP phase, fully mature TCRαβ is expressed together with low CD3, initiating negative and positive selection on a correct and functional TCRαβ complex[33,34]. During these selection processes, up to 90% of all developed T cells undergo apoptosis[35]. Finally, thymocytes choose the CD4 single-positive (SP4) or CD8 single-positive (SP8) lineage and migrate as fully mature and functional T cells from the thymus into the periphery.

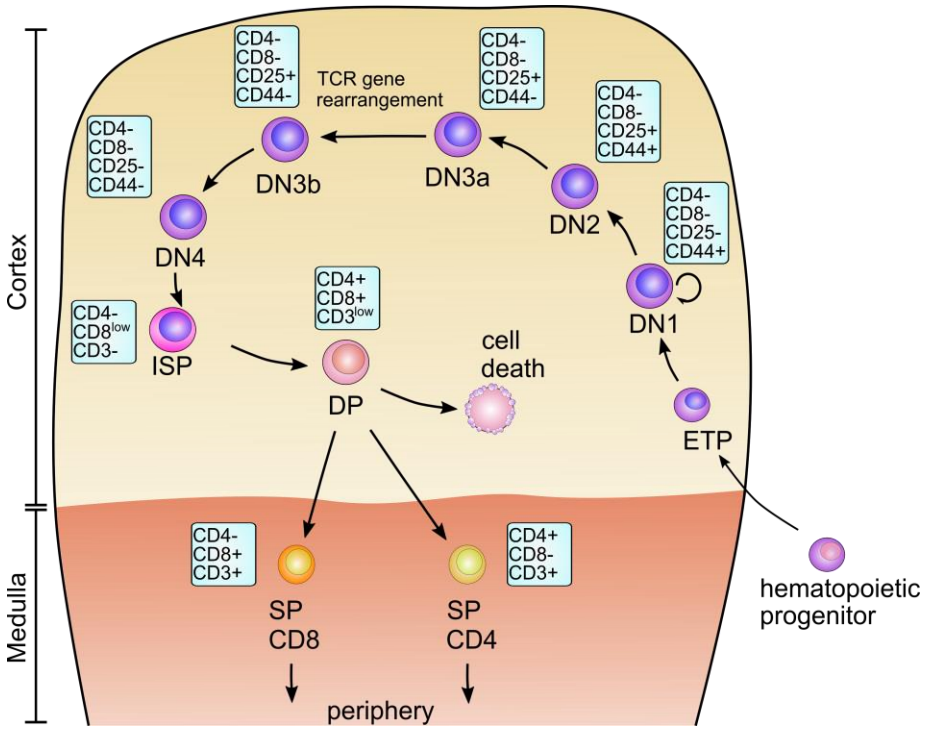


Figure 2: Murine T cell development within the thymus. Early T cell progenitors (ETP) migrate into the cortical region of the thymus and undergo T cell differentiation. Within the early immature double-negative (DN) stages, cells migrate through the outer and inner parts of the thymus, marked by up- and down-regulation of CD25 and CD44 and rearrangement of the T cell receptor (TCR). After the DN4 stage, T cells earn the immature single positive (ISP) stage before reaching the double-positive (DP) stage, where they undergo positive and negative selection for functional TCR. Mature T cells leave the thymus as either CD4 or CD8 single-positive (SP) T cells. The illustration was inspired by Germain *et al*.[36] and Rothenberg *et al*.[37]

To mimic the differentiation stages of developing thymocytes *in vitro*, a co-culture system described by Zúñiga-Pflücker will be used for this study[38]. The OP9 bone marrow stromal cell line was optimized to express delta-like 1 ligands (OP9-DL1), promoting necessary Notch signaling for driving T cell specification and differentiation[39]. In addition, the cytokines IL7 for differentiation and SCF and FLT3-L for proliferation and survival of the early progenitors will be added to this culture system. Thus, this creates a suitable environment in which lineage-negative BM cells undergo T cell lineage-specific differentiation, mirroring the different phases and stages of intra-thymic development[40].

2.3 Relevant cellular signaling pathways driving T cell differentiation and activation

Different cellular signaling pathways play an important role in the activation, proliferation and survival of functional T cells. These include the already mentioned Notch signaling[41], the IL7R mediated JAK-STAT signaling pathway[42], as well as the growth-factor induced PI3K-AKT-mTOR pathway[43] and the TCR-induced Ras-MAPK-ERK signaling[44,45].

The Notch pathway starts with the external engagement of Notch ligand (mainly Delta-like ligand 4, DLL4) with the extracellular receptor part of NOTCH1[46]. The intracellular part of NOTCH1 (ICN1) is then cleaved by a disintegrin and metalloproteinase domain-containing protein 10 (ADAM-10) and a γ-secretase to release the NOTCH1 intracellular domain (NICD protein). It translocates to the nucleus to interact with recombining binding protein suppressor of hairless (RBPJ) and activates, among others, expression of proto-oncogene *c-Myc* and other genes important for the regulation of normal T cell development[41,47,48]. Signaling is terminated through the ubiquitin ligase F-Box and WD repeat domain containing 7 (FBXW7) protein, which modulates the degradation of ICN1 or its effector protein MYC (encoded by *c-Myc* gene) (Figure 3). Therefore, inactivating mutations in *FBXW7* or activating mutations in *NOTCH1* are favorable events for human leukemia initiation[49,50].

During IL7R signaling, IL7 interacts with the heterodimer IL7R (consisting of IL7Rα subunit and common gamma chain γc), which subsequently induces the phosphorylation of Janus kinase (JAK) 1 and JAK3 and thus activates signal transducer and activator of transcription 5 (STAT5). Phosphorylated STAT5 translocates as a dimer into the nucleus to regulate for example antiapoptotic BCL-2 and genes involved in T cell homeostasis and differentiation[51]. Signaling gets terminated by the activation of, e.g., protein tyrosine phosphatases non-receptor type 2 (PTPN2) and dynamin 2 (DNM2)[52]. Activating mutations within signaling compartments of JAK-STAT-signaling (IL7R, JAK1, JAK3) and its effector proteins (BCL-2, MYC) or inactivation of negative regulators (PTPN2, DNM2) are frequently found in T-ALL (Figure 3)[53,54].

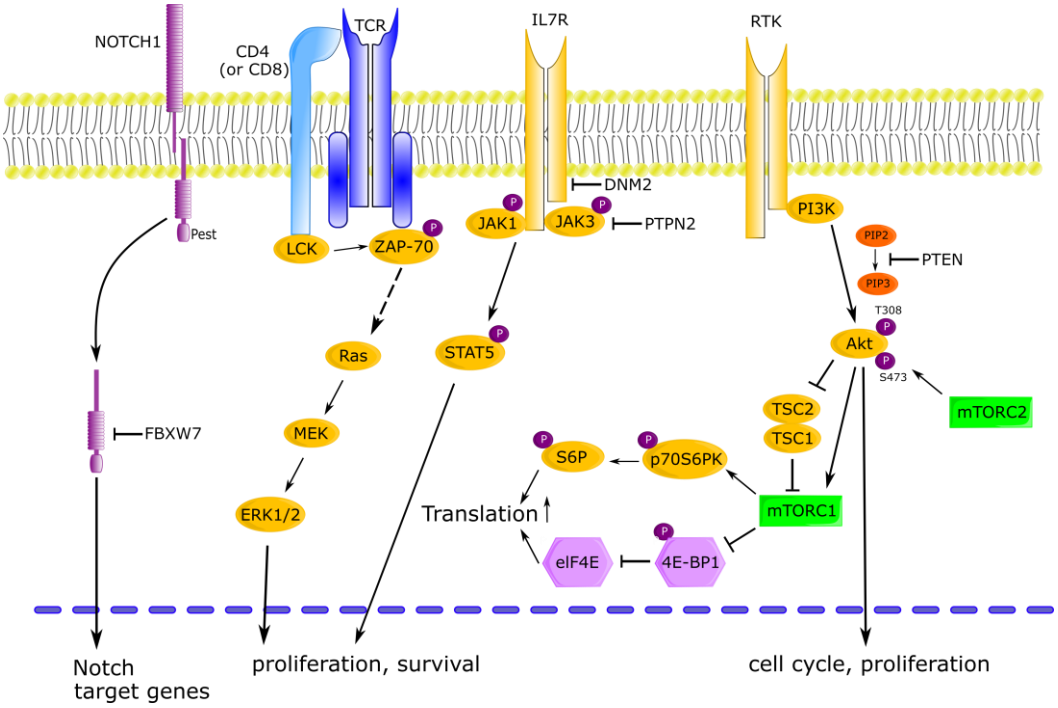


Figure 3: Important cellular signaling pathways in T cells. Notch signaling, T cell receptor (TCR) mediated RAS-MEK-ERK signaling, interleukin 7 receptor (IL7R) mediated JAK-STAT signaling and receptor tyrosine kinase induced PI3K-AKT-mTOR signaling are the most important signaling pathways in T cells regulating activation, proliferation, and survival. NOTCH1, notch homolog 1; FBXW7, F-box and WD repeat domain containing 7; CD4/CD8, cluster of differentiation 4/8; LCK, lymphocyte-specific protein tyrosine kinase; ZAP-70, zeta-chain-associated protein kinase 70; Ras, rat sarcoma virus protein; MEK, mitogen-activated protein kinase kinase; ERK1/2, extracellular signal-regulated kinase 1/2; JAK1/3, Janus Kinase 1/3; DNM2, dynamin 2; PTPN2, protein tyrosine phosphatases non-receptor type 2; STAT5, signal transducer and activator of transcription 5; PI3K, phosphoinositide 3-kinase; PIP2, Phosphatidylinositol 4,5-bisphosphate; PIP3, phosphatidylinositol-3, 4, 5-triphosphate; PTEN, phosphatase and tensin homolog; AKT, protein kinase B; mTORC1/2, mammalian target of rapamycin complex 1/2; TSC1/2, tuberous sclerosis 1/2 protein; p70S6K, ribosomal protein S6 kinase beta-1; S6P, S6 ribosomal protein; 4E-BP1, eukaryotic translation initiation factor 4E-binding protein 1; eIF4E, eukaryotic translation initiation factor 4E.

An important function of fully mature T cells is the recognition of presented peptides on other cell surfaces. When TCR recognizes an antigen, lymphocyte-specific protein tyrosine kinase (Lck) gets activated through the interaction of CD4 or CD8 with the major histocompatibility complex (MHC) I/II on the presenting cell[55]. Once zeta-chain-associated protein kinase 70 (ZAP-70) binds to the TCR, activated Lck facilitates its phosphorylation, which subsequently induces the rat sarcoma virus (Ras) signaling[44,56]. The Ras family consists of three isoforms, having N-Ras as the most abundant form[57]. Actively bound to guanosine-5'-triphosphate (GTP), Ras-GTP triggers the mitogen-activated protein kinase (MAPK) cascade and downstream extracellular-signal-regulated kinases 1 and 2 (ERK1/2) initiate transcription[56]. Hence, the transcription factor activator protein-1 (AP-1) is formed, which drives the transcription of interleukin 2 (IL2). By this, peripheral T cell is activated and its proliferation is induced[58].

Upon binding of a growth factor to a receptor tyrosine kinase (RTK) (including IL7 to IL7R), phosphatidylinositol-3 kinase (PI3K) is activated and phosphorylates phosphatidylinositol-3,4-

bisphosphate (PIP2) into phosphatidylinositol-3,4,5-trisphosphate (PIP3). This phosphorylation can be reversed by phosphatase and tensin homolog (PTEN). PIP3 next recruits the serine/threonine kinase (AKT) to the membrane, which results in its phosphorylation and activation at the threonine 308 (Thr308) site, mediated by phosphoinositide-dependent kinase-1 (PDK1), or at the serine 473 (Ser473) site, mediated by mammalian target of rapamycin complex 2 (mTORC2)[59,60]. Downstream of the activation cassette lays the mammalian target of rapamycin complex 1 (mTORC1), which in turn phosphorylates p70 ribosomal protein S6 kinase (p70S6K) and the effector protein S6 kinase-1 (S6K-1) and inactivates eukaryotic translation initiation factor 4E binding protein-1 (4EBP1) by phosphorylation, all leading to cell cycle progression and increased translation[61,62]. Dysregulation of the PI3K-AKT-mTOR pathway is frequently observed in various cancers[63,64].

2.4 T cell acute lymphoblastic leukemia (T-ALL)

2.4.1 T-ALL – classification and treatment

Aberrant activation of the described signaling pathways is a common event driving the development of T cell acute lymphoblastic leukemia (T-ALL)[50]. The complex orchestration of T cell development with its intended genetic rearrangements is a highly effective but also error-prone procedure, which could promote the acquisition of malignant alterations. T-ALL is an aggressive hematological malignancy, marked by the infiltration of the bone marrow, central nervous system, and other peripheral organs by immature blasts of the T cell lineage. It accounts for ~15% of pediatric and ~25% of adult acute lymphoblastic leukemia (ALL) cases, with a five-year survival rate of ~80% for children and ~40-50% for adults[65–67]. Diagnosis compromises morpho-cytochemistry, flow cytometry, and gene analysis, to classify the exact subtype and to distinguish from acute myeloid leukemia (AML) or B cell acute lymphoblastic leukemia (B-ALL), which can just hardly be discriminated morphologically. Indeed, the search for T cell antigens, including CD7, CD2, CD5, CD1a, CD3, CD4, CD8 and terminal deoxynucleotidyl transferase (TdT) allows discrimination not only from other hematologic lineages but also between early and late T cell developmental stages, indicating T-ALLs biological subtype[68].

In line with the different stages of T cell development in the thymus, T-ALL can be divided into early T cell progenitor ALL (ETP-ALL), early cortical T-ALL, and late cortical T-ALL. ETP-ALL comprises ~15-35% of T-ALL cases, and encompasses the most immature type of T cell identity with prolonged myeloid or B cell differentiation potential. Here, the lack of CD1a, CD4 and CD8 and the presence of at least one myeloid or stem cell antigen (CD34, CD33, CD11b, CD65, human leukocyte antigen – DR isotype (HLA-DR)) classifies this subtype[68,69]. Generally, patients from this subtype have a poor prognosis[70].

The early cortical stage is typically classified by the presence of CD1a, CD4 and CD8. It results from chromosomal translocations leading to aberrant expression of the related transcription factors T cell leukemia homeobox 1/3 (*TLX1/3*), homeobox *A* (*HOXA*) or NK2 homeobox 1 (*NKX2-1*)[65,71,72]. The late cortical group summarizes 40-60% of patients, is classically identified by the expression of CD3, CD4 and CD8, and is most often linked to overexpression of TAL basic helix-loop-helix transcription factor 1 (*TAL1/SCL*) and *LMO1/2.*[73]

Although the immune phenotype helps classify T-ALL biological subtypes, it just partly correlates to the clinical outcome. The complexity of the underlying genetic abnormalities within the leukemia has a much greater impact on patients' survival[74,75]. Therefore, the accumulation of more, and also more aggressive mutations in older patients may explain, why adults have a less favorable prognosis compared to children[76]. While in 2015, about twenty driving gene mutations were identified to distinguish biological T-ALL classes[77], two years later, Liu *et al.* described 106 presumed driver genes[50]. These include genes involved in transcriptional regulation, cell cycle, Notch pathway, epigenetics and cell signaling[50,78]. A summary of relevant genes and their assignment can be found in Table 1. For more details, see section 2.4.2.

Table 1: Classification of the most relevant genetic alterations in T-ALL.

Classification	Gene annotation	Gene name
Transcription factor	LYL	LYL1 basic helix-loop-helix (bHLH) family member
	TAL1/SCL	TAL bHLH transcription factor 1/Stem cell leukemia
	LMO1/2	Lim domain only 1/2
	TLX1/3	T cell leukemia homeobox 1/3
	HOXA	Homeobox A
	NKX2-1	NK2 homeobox 1
	ETV6	ETS variant transcription factor 6
	GATA3	GATA binding protein 3
	RUNX1	Runt-related transcription factor 1
Notch pathway	NOTCH1	Notch Receptor 1
	FBXW7	F-Box and WD repeat domain containing 7
	MYC	MYC proto-oncogene
Cell cycle	CDKN2A	Cyclin dependent kinase inhibitor 2A
Cell signaling	IL7R	Interleukin 7 receptor
	JAK1/3	Janus kinase 1/3
	PIK3CD	Phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta
	NRAS	Neuroblastoma RAS viral oncogene homolog
	PTEN	Phosphatase and tensin homolog
	PTPN2	Protein tyrosine phosphatase non-receptor type 2

Table 1: Classification of the most relevant genetic alterations in T-ALL (continued).

Classification	Gene annotation	Gene name
T cell development	BCL11B	B cell lymphoma/leukemia 11B
	LEF1	Lymphoid enhancer binding factor 1
Epigenetic	PHF6	PHD finger protein 6
	SUZ12	Polycomb protein SUZ12
	USP7	Ubiquitin specific peptidase 7
	EZH2	Enhancer of zeste homolog 2
	KMT2A	Histone-lysine N-methyltransferase 2A
	MLLT10	Histone-lysine methyltransferase DOT1L cofactor
Other	DNM2	Dynamin 2

Although recent applications of pediatric-oriented protocols in adults improved remission rates, the relatively high relapse rate originating from persistent leukemia-initiating cells (LIC) is still the biggest challenge in adults[65]. Indeed, 40% of adults relapse, which is very difficult to treat, even with prompt allogeneic stem cell transplantation (HSCT) as salvage therapy[79,80]. In general, many different cellular pathways or mutations can be held responsible for these resistance mechanisms. Degradation of the tumor suppressor protein P53 through murine double minute 2 (*MDM2*) imbalance or indirectly through disrupted cyclin-dependent kinase Inhibitor 2A *(CDKN2A*) mutations, which often occur in T-ALL, can be linked to resistance against standard chemotherapy like vincristine, anthracycline or glucocorticoids[81,82]. However, many resistance pathways do not result from single mutated genes but rather from combined mutations. Especially in the cellular signaling pathways JAK-STAT, PI3K-AKT-mTOR, or RAS-MAP-ERK, crosstalk between these pathways exists and co-occurring mutations can compensate for the drug-mediated intervention of single pathways. For example, the blocking of the NOTCH1 signaling pathway, which is mutated in ~60-70% of T-ALL[50,83], often only leads to a cell cycle arrest of the leukemic cell, but not to apoptosis. The interference can be counterbalanced by the abnormal activation of PI3K-AKT-mTOR signaling through deletions in *PTEN*[47] or activating mutations in *PIK3CA*, *PIK3CB*, and *PIK3CD*[50,76]. When thus the PI3K-AKT-mTOR pathway is targeted with novel and highly potent drugs like Buparlisib[84] or the PI3K-mTOR dual inhibitors NVP-BEZ235[85], promising results were observed in pre-/early clinical studies. However, breakthroughs are lacking, as resistance still occurs, highlighting the need to understand the genetic complexity of T-ALL, driving these resistance mechanisms, and to recognize novel genetic and cellular liabilities, by which refractory or relapsed leukemia is initiated.

2.4.2 T-ALL – genetic landscape shapes different biological subtypes

As already described, T-ALL can roughly be divided into ETP-ALL, early cortical T-ALL, and late cortical T-ALL by a variety of different genetic alterations and their respective cells of origin. Due to genetic rearrangements, promoters of different transcription factors are placed under the control of strong T cell-specific enhancers[28]. Consequently, the overexpressed driving oncogenic transcription factors, including TAL1, LMO1, LMO2, LYL, TLX3, TLX1, HOXA, NKX2 and KMT2A-rearrangements define the biological subtype of the disease. Chromosomal alterations affecting *HOXA, LMO2* or *LYL* genes, as well as mutations in early hematopoiesis genes such as *GATA3*, *ETV6*, *RUNX1*, *EZH2*, or *SUZ12*, are usually correlated with the ETP-ALL subtype. Leukemic thymocytes harboring activating aberrations in *TLX3*, *TLX1*, or *NKX2-1* are defined as early cortical T-ALL, while malignant activation of *TAL1* and/or *LMO1/2* are assigned to the more mature late cortical subtype[70,76]. Within this thesis, the main focus will be set on the early cortical and late cortical subtypes, including the deregulation of *TAL1*, *LMO2* and *TLX1*, comprising roughly 40% of T-ALL cases[86].

Aberrant expression of oncogenic transcription factors results in most cases from chromosomal translocations or rearrangements linking promoter and enhancer regions of the TCRβ locus to those of the foreign transcription factors[71,87]. The most prominent ones include for example t(7;10)(q35;q24) for aberrant *TLX1* expression, t(11;14)(p13;q11) for *LMO2*, t(1;14)(p32;q11) for *TAL1* genes but also others like *PICALM-MLLT* (t(10;11)(p13;q14)) and *NUP214-ABL1* (9q34) translocations[88]. It is thought that RAG1, which facilitates rearrangement during V(D)J-recombination in healthy thymocytes, could bind to cryptic recognition sites and thus recombines inappropriate chromosomal parts. Nevertheless, besides RAG1, there must also be other unknown mechanisms promoting translocations, as for many break points it could be shown that they are not mediated by RAG1[87,89]. Interestingly, these potential recognition sites for RAG1 are also found within the *PTEN* locus, hypothesizing that deletion of *PTEN* can also result from unwarranted RAG1-recognition[90].

In addition to the aforementioned deregulated transcription factors, activating mutations in the Notch signaling pathway, PI3K/AKT/mTOR pathway, and/or JAK-STAT pathway together with loss of *CDKN2A* and/or *PTEN* are mostly associated with the different T-ALL groups. Indeed, loss of *CDKN2A*, which stabilizes the tumor suppressor protein p53 and regulates cell cycle progression, is deleted in ~70% of all T-ALLs – most likely in the early and late cortical subtypes[70,74]. Activating mutations within *NOTCH1* are found in 60-70% of patients, although they appear less frequently in the ETP-ALL subset[50,83]. Mutations either target the NOTCH1-Pest domain, a negative regulatory domain from NOTCH1, or the *Fbxw7* gene, involved in the degradation of NICD and MYC, increasing their stability and abundance[49]. Moreover,

additional driver mutations often occur in hematopoietic transcription factors that play an active role during T cell development. These include *BCL11B*, *ETV6*, *GATA3*, *LEF1* and *RUNX1*[50,74].

In a sequencing study of 264 T-ALL cases by Liu *et al.* 2017, they could not only detect all of the described genetic alterations, but also designated a prevalence of certain loss-of-function mutations with respective subtypes. Indeed, they showed that *PTEN* and *USP7* gene mutations were mainly enriched in the TAL1 group, whereas *DNM2*, *PHF6* mutations and JAK/STAT activation correlated mostly with TLX1/TLX3 subtypes[50]. Among these, *PHF6* disruption has the highest prevalence[50,76]. *PHF6* regulates the self-renewal and cycling of HSCs, and it's inactivation contributes as one of the earliest factors in triggering leukemic transformation[91,92]. Mutations in other TFs like in *RUNX1*, *GATA3*, and *EZH2* or the genes coding for the signaling proteins NRAS, IL7R, and JAK3 are most prominently found in the immature ETP-ALL subtype[65,69]. LMO2 upregulation can lead to an early maturational arrest, when deregulated with LYL1, or to a late cortical T-ALL when associated with overexpressed TAL1[93].

Although the genetic complexity and growing number of known mutations have increased dramatically in recent years, our understanding of how different oncogenes and tumor suppressors interact is still limited. Especially in the context of chemotherapeutic resistance, understanding the aberrant genetic and cellular networks seems to be essential. In the context of disease progression, little is known about the acquisition steps of these mutations and the heterogeneity of the observed phenotypes and genotypes. Advances in single-cell RNA sequencing technologies gave the first insights into this topic. De Bie *et al.* showed that the very abundant mutation in *NOTCH1* is a relatively late event during oncogenesis, thereby questioning how promising its therapeutic targeting would be[94]. Furthermore, they could demonstrate that while mutations were acquired at a committed lymphoid developmental stage in some patients, in others they already occur at an early hematopoietic progenitor stage, leading to their presence in mature lymphoid but also myeloid cells[94]. This observation raises different concerns. On one side, it supports the need to target early hematopoietic progenitor cells. This aspect becomes important for autologous HSCT, in case some highly mutated progenitor cells were not eradicated properly before transplantation[79]. On the other side, acquisition at such an early stage of hematopoiesis favors the development of mixed lineage leukemia, further complicating the eradication of leukemia. Indeed, various gene mutations, which were just described, are also observed in acute myeloid leukemia (AML). This includes, for example, activating *NRAS* mutations ($NRas^{G12D}$) but also inactivating mutations in *CDKN2A*, *PHF6*, *ETV6*, *EZH2*, or *WT1*[95].

2.4.3 T-ALL – challenges in modeling a genetically complex disease

In parallel to the better knowledge of the underlying genetics of T-ALL, various murine model systems have been established that attempt to mimic the biology of this disease. Transgenic mice overexpressing *Notch3* or *Tal1* were one of the first T-ALL models established[96,97]. However, their long latency in disease progression led to the notion that additional co-occurring events are needed for frank leukemia. Therefore, double-mutant mice overexpressing, for example, *NUP214-ABL1* and *TLX1* or deleting *Runx1* and *Ezh2* in thymocytes were developed, representing disease phenotypes more realistically[98,99]. However, germline engineering is a time-intensive approach and transgenic animals require complex breeding schemes and intensive animal handling as all the animals are afflicted. Still, they only cover a few mutations, which normally occur in clinics.

The group of Gao *et al.* were able to improve the analysis of genetic heterogeneity by transplanting wild-type thymus into a *Rag2*$^{-/-}$ *Il2rg*$^{-/-}$ (interleukin-2 receptor subunit gamma) double knockout mouse model. This indeed spontaneously induced different complex T-ALLs with similar mutation patterns as observed in patients[100]. However, long latencies in disease progression and spontaneous T-ALL subtypes diminish their large-scale use for therapeutic screening purposes for reasons of time and consistency.

Another approach is patient-derived xenografts (PDX). There, T-ALL patient samples are transplanted into immune-compromised mice and thus represent a clinically relevant mutation pattern, allowing pre-clinical testing of potential drugs[101,102]. Nevertheless, transplanted animals consist of an atypical immunocompromised microenvironment, as they lack lymphoid immune cells and NK cells. The interaction of therapeutic approaches with the hosts' immune system remains unanswered. Optimizations have been implemented by co-transplanting human BM-derived mesenchymal stroma cells, helping to increase engraftment rates and reconstitute human hematopoiesis from functional progenitor cells. However, this system was only applied to myeloid leukemia samples so far, and accessibility and donor variety still challenge this model system[103].

A more reliable and reproducible approach is the transplantation of genetically altered HSCs. By manipulating murine BM cells, leukemia can be generated in a more controlled and modular context. With this approach, different T-ALL were generated through overexpressing *receptor tyrosin kinases* (*TRK*), *IL7R*α mutants or *Notch1* mutants[104–106]. Interestingly, while activated *Notch1* was intentionally introduced in the context of *IL7Rα*[106], TRK-driven leukemia acquired additional mutations in *Notch1* and *Pten* during their clonal development and hence accelerated disease progression and T-ALL formation[107]. However, these animal models describe only a fraction of potential mutations that can interact to generate complex T-ALL.

The additional manipulation of BM with CRISPR-Cas9 genome editing approaches can help to further increase complexity. In a co-transduction approach with lentiviral vectors Heckl *et al.* were able to introduce up to five different mutations into one single BM cell, leading to its clonal expansion and subsequent development of AML[108]. This approach illustrated that the interaction and cooperation of different oncogenic events can be mimicked in a mouse model.

All of the above-mentioned model systems have different advantages and disadvantages in time consumption, costs, translatability, availability and applicability. In times of high-throughput sequencing[50,76,77], we are gaining deeper insights into the complex interplay of different genetic events in the malignant transformation of healthy thymocytes. Appropriate model systems are needed to use the gained knowledge to increase patients' survival. In this work, we will therefore focus on a modular BM transplantation-based system that can be genetically manipulated by lentiviral transduction (described in section 2.5) and using CRISPR-Cas9 (described in section 2.6.2).

2.5 Using retroviral vectors for safe gene delivery

The genetic modification of somatic cells with retroviral vectors is a common tool in biomedical research. Thereby, the retrovirus-based viral vectors serve as a vehicle to deliver genetic information by stably integrating it into the hosts' genome. Retroviruses belong to single-stranded RNA-enveloped viruses. The retrovirus family includes the alpha-, gamma- and lentiviruses, among others, which share similar genomic structures consisting of structural elements (group associated antigen - gag), retroviral enzymes for replication (pol), and envelope proteins (env)[109]. They mainly differ in their integration pattern and restriction to cell division. While alpha- and gamma-retroviruses primarily, but not exclusively[110], enter the nucleus during cell division, lentiviruses hold the potential to actively enter the nucleus and integrate into non-dividing cells[111]. Moreover, lentiviruses preferably integrate within actively transcribed genes, while gamma-retroviruses are more prone to integrate into the proximity of transcriptional start sites (TTS) and promoter/enhancer regions. Only the alpha-retroviruses tend to integrate within a random pattern[112–114].

For the application in biomedical research or therapy as a gene delivery tool, the genus of retroviral vectors was changed by eliminating non-essential genes, including *gag*, *pol*, *env*, and the regulator of virion (*rev* – involved in nuclear export and packaging for lentiviruses only), leaving space for the gene-of-interest (GOI). To assemble the whole viral particle, the removed genes are provided as separate helper plasmids, lacking the packaging signal (ψ*,* split-packaging system). By this approach, the formation of replication-competent retroviruses, which could start independent infection cycles, is prevented[109]. For vector production, the transfer vector encoding the GOI is co-transfected with the helper plasmids (gag/pol, env, rev)

into a producer cell line. The natural tropism of the retrovirus can be modified by exchanging the envelope protein (pseudotyping). The most commonly used one is derived from the vesicular stomatitis virus (VSVg) and possesses a very broad tropism[115].

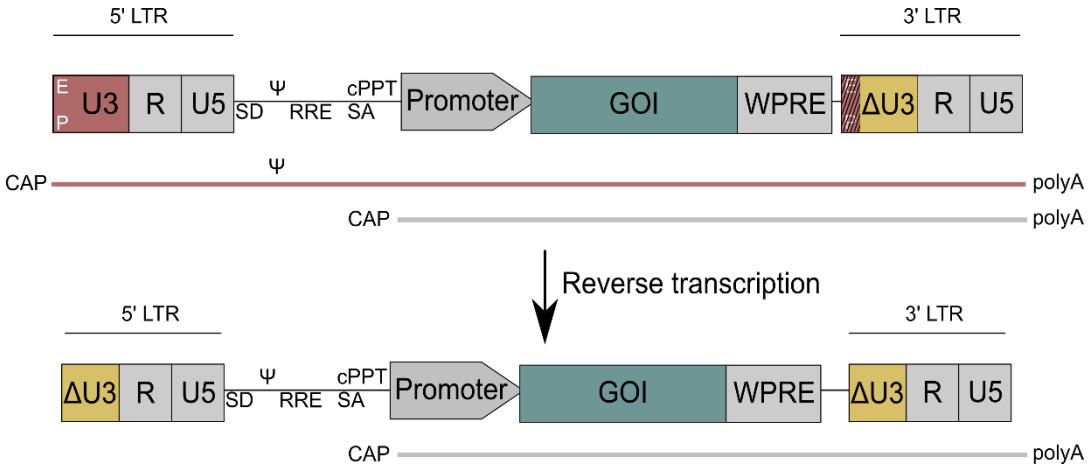


Figure 4: Configuration of a self-inactivating (SIN) lentiviral vector. The upper part represents the lentiviral RNA, which is reverse transcribed into DNA for integration into the host genome. During this process, the ΔU3 region of the 3' long terminal repeat (LTR) is copied to the 5' LTR and replaces its U3 region, containing strong enhancer (E) and promoter (P) regions. After reverse transcription and integration, the gene of interest (GOI) is only transcribed by the internal promoter. PBS, primer binding site; SD, splice donor; SA, splice acceptor; ψ, packaging signal; RRE, rev-responsive element; PPT, polypurine tract; WPRE, woodchuck hepatitis virus post-transcriptional regulatory element. The illustration was inspired by Maetzig *et al.* 2011.[109] (open access article under Public License CC BY 3.0; Copyright © 2011 by the authors)

The retroviral genome consists of different elements and is flanked by two long terminal repeats (LTR) (Figure 4). These contain unique 3' and 5' regions (U3 and U5) and a repeated sequence (R) in between. In the U3 region, enhancer and promoter sequences are located, which drive transcription of the following cassette, including the GOI. During reverse transcription, the 3' U3 region is copied to the 5' LTR and the 5' R and U5 region to their counterparts in the 3' LTR. The R region includes the transcriptional start site and a weak polyadenylation signal[109].

In the past, the strong U3 enhancer/promoter region and the integration pattern of gamma-retroviruses into TSS led to insertional mutagenesis, in which the gene expression of neighboring genes is strongly transactivated by the retroviral enhancer elements of the inserted provirus[116]. Due to the uncontrolled activation of proto-oncogenes by the inserted viral genome, this approach could lead to the transformation of transduced cells[116,117]. To decrease this risk, the self-inactivating (SIN) lentiviral (LV) vector system was developed. There, the viral enhancer/promoter elements were depleted from the 3' U3 region (ΔU3). The enhancer/promoter sequence in the 5' LTR still drives transcription during packaging but will be replaced by ΔU3 during reverse transcription in the target cell. Thus, unwanted expression of proto-oncogenes is decreased without titer loss during vector production[118]. However, the GOI needs to contain an internal promoter then to drive its expression upon integration. Here,

different promoter systems can be chosen like the ubiquitous spleen-focus forming virus (SFFV) promoter. By using the HIV-1 derived SIN lentiviral (SIN-LV) vector system in conjunction with less strong internal promoters, insertional mutagenesis, and transformation was strongly reduced in HSCs and is since then used more widely[119].

2.6 Modulation of gene expression via Cre recombinase and CRISPR-Cas9-based knockout

2.6.1 Using Cre-recombinase for tissue-specific gene expression

Use of retroviral vectors with broad tropism (VSVg enveloped) and ubiquitous promoters (SFFV) allow an effective delivery and expression of genetic information but neglects the navigation into specific cell-types. In particular, studying the genetic and pathophysiologic role of certain genes in mouse models often requires a time and tissue-specific expression or deletion of the target gene. Although cell-type specific pseudotyping or tissue-specific promoter choice can be used to target gene delivery, its application is more time-consuming and elaborate (reviewed in [120,121]). In this regard, the Cre-loxP system has evolved as an efficient tool of choice facilitating tissue-specific gene expression. Here, mouse strains are modified to express Cre recombinase (Cre) under the control of tissue-specific promoters. When breeding with another mouse strain or providing retroviral vectors both containing a floxed gene-of-interest, its tissue-specificity can easily be achieved by relying on the expression of Cre[122].

Cre consists of a flippase and a D6-specific recombinase[123], which is normally used for the life cycle of the bacteriophage P1[124]. Cre recognizes a 34-base pair (bp) sequence, the so-called *loxP* (locus of x-over P1) site, consisting of two 13 bp inverted and palindromic repeats and 8 bp serving as core sequence[125]. Two *loxP* sites are needed for Cre to bind and excise the deoxyribonucleic acid (DNA), flanked by both *loxP* sites (floxed)[124]. Depending on the orientation of the *loxP* sites, excision or inversion can be achieved. These systems can be used for different purposes. Flanking a target gene with two sense *loxP* sites removes it upon activation of Cre. By this, conditionally knockouts can be generated[122]. Inserting a floxed stop cassette in front of a GOI promotes gene expression only after the excision of the stop cassette[126]. Another system is the usage of mutated *loxP* sites. When the single mutated *loxP* sites *lox66* and *lox71* flank a target gene in head-to-head orientation, the cassette gets inverted once, leading to a wild-type *loxP* site and a double mutant *lox72* site (product of *lox66* and *lox71*), with much lower affinity for the re-binding to Cre[127,128]. Enabled by this strategy, gene expression can accurately be coordinated for this work, allowing the targeted *in vivo* manipulation of the genome of developing T cells.

2.6.2 Gene editing with CRISPR-Cas9

Precise genome editing is one of the key revolutionary discoveries of this century. Honored by the Nobel Prize in 2020, gene editing by the clustered regularly interspaced short palindromic repeats (CRISPR) system became one of the hallmark tools in molecular biology. Originally, CRISPR-Cas9 (CRISPR-associated protein 9) is derived from bacteria and archaea as part of their immunity against phages. By storing parts of the viral DNA after an infection, prokaryotes can build up resistance against future viral attacks[129]. Nowadays, its application ranges from plant biotechnology[130,131] to treating genetic diseases[132,133], infectious diseases[134], and cancer research[135,136].

The CRISPR complex consists of non-coding, repetitive RNAs with short variable sequences. These elements form the CRISPR RNA (crRNA), which binds for further processing to the partly complementary trans-activating crRNA (tracrRNA) and builds a complex together with the Cas9 protein. Besides the protospacer adjacent motif (PAM), which varies between different species, the crRNA contains a 20 nucleotide (nt) guide-sequence, which allows the CRISPR-Cas9 complex to bind complementary to its target DNA and induces a double-strand break (DSB)[137]. For mammalian cells, the type-II CRISPR-Cas9 system derived from *Streptococcus pyogenes* is the best described[129,138]. To further simplify the system, the crRNA and tracRNA were fused with an artificial RNA linker into one single-guided RNA (sgRNA)[139]. Thus, by exchanging the 20 nt sequence, almost every genetic location can be targeted, restricted only by the PAM motif, which is NGG for the *Streptococcus pyogenes* Cas9 (spCas9)[139].

The binding of the CRISPR-Cas9 complex to its target region introduces a DSB, which is repaired by two different pathways (Figure 5). The more favorable and likely, but also more error-prone, is non-homologous end-joining (NHEJ)[140,141]. Thereby, random insertions and deletions (indels) occur, which can disrupt the reading frame and hence the protein expression. By designing experimental approaches to target important structural protein domains or conserved amino acids, even in-frame deletions strongly interfere with protein function or expression, thereby increasing the protein knockout potential[142]. In the presence of a DNA donor template and only during the S and G2-phase of the cell cycle, another DNA repair pathway can occur[140]. Through homology-derived repair (HDR), the donor template is used to correct the disrupted gene through recombination. In the presence of an artificial DNA template with homologous sequences, insertional replacement or gene correction can occur. This approach can theoretically be used to treat monogenetic disorders like cystic fibrosis[143], beta-thalassemia, and sickle cell disease[144].

During the last few years, different algorithms were created to predict efficient sgRNAs with high on-target and low off-target activity[145–147]. It was already known from short hairpin RNA (shRNA) approaches that sequence composition greatly impacts its functionality. Therefore, many different algorithms were created to further optimize their on-target knockdown capacity[145–147]. These algorithms provide theoretically potent sgRNAs for every genomic region of interest and offer a scoring system to evaluate its on-targeting activity, provide potential off-targets, depending on a given number of mismatches, and even calculate the potential indel formation upon cleavage[148]. However, translation of these *in silico* predictions could still lead to moderate genome editing, highlighting the requirement for independent validation by e.g. fluorescence reporter systems, before its application[149].

In the following work, the Cre-lox system will be used to coordinate T cell specific expression of various oncogenes, while the CRISPR-Cas9 system facilitates the insertion of mutations to disrupt gene function of crucial tumor suppressors. By merging the different gene editing tools, the occurrence of unwanted aberrations in other tissues or cell types could be reduced.

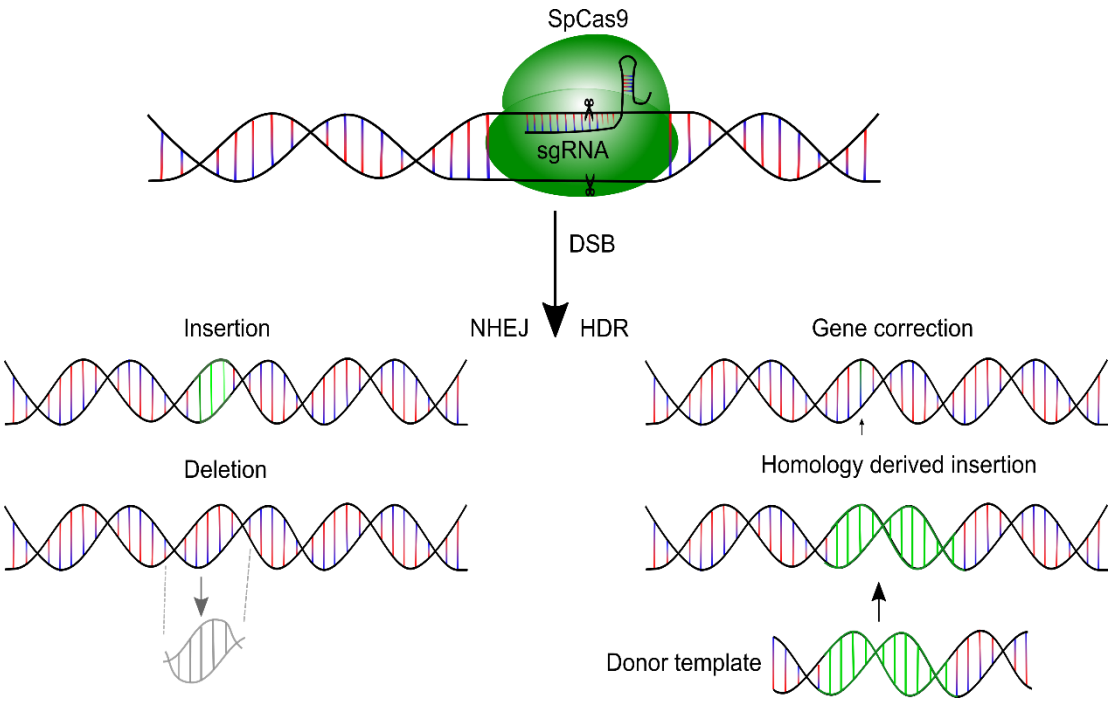


Figure 5: DNA repair events after the introduction of double-strand breaks (DSB) by CRISPR-Cas9. After DSB, DNA is more likely to be repaired by non-homologous end-joining (NHEJ), which results in the insertion of random nucleotides or deletion of the nearby sequence. Less likely and only in presence of a homologous DNA donor template, homology-directed repair (HDR) can occur, resulting, for example, in the correction of the sequence from the homologous chromosome. In presence of a partly homologous donor template, larger regions can be introduced, which is used for gene correction or insertion.

3 Aims of the PhD thesis

For decades, T-ALL models were based on the retroviral overexpression of *Notch1* in mouse hematopoietic stem cells. This induces aggressive T-ALL, but does not reflect the clinical heterogeneity and behavior of the disease, making it questionable to what extent the insights gained from these models are generalizable. The genetic complexity of T-ALL in particular is a major cause of relapse and subsequent low survival rates in patients. Although more advanced animal models that focus on the combination of two to three different observed genetic mutations helped to understand the mechanisms behind the acquisition of resistance to standard therapy, they hold many disadvantages. Transgenic mice harboring these oncogenic alterations require complex breeding strategies, yield long disease latencies and capture only a fraction of T-ALLs' genetic complexity. PDX models might bridge the gap from mice to patients, however, their low availability and the artificial setup limits larger screening approaches. Therefore, there is a high need for novel approaches mimicking the observed genetic alterations known from patient-related high-throughput sequencing studies in a more realistic and screenable mouse system.

In this context, the aims of this thesis will be

1) **Generation of genetically complex T-ALL in mouse by CRISPR-Cas9 genome editing matching the major subgroups of T-ALL**
2) **Characterization of novel T-ALL models and evaluation of their potential as resource to screen for future therapies**

The first part includes the generation of a BM transplantation-based mouse model that can easily be scaled and modified. Recent understandings in CRISPR-Cas9 allows to introduce T-ALL-specific gene mutations into BM-transplanted animals. To reach the genetic complexity of clinically relevant T-ALL, a random mutation pattern and polyclonal leukemia development might provide the most realistic outcome. Therefore, a set of 7 – 10 different oncogenic events was chosen, which are often observed in the clinics, and which will be set in defined combinations. These include genes coding for murine transcription factors (*Tlx1*, *Tal1*, *Lmo2*, and *Etv6*), genes involved in cellular signaling (human *NRAS*, *IL7R*, PIK3CD and murine *Pten*, *Notch1*, *Ptpn2*, and *Dnm2*), cell cycle (*Cdkn2a*), T cell development (*Bcl11b* and *Lef1*) and epigenetics (*Phf6* and *Usp7*). Different arrangements, as seen in patients, will help to investigate the cooperation between dysregulated genetic abnormalities and reveal vulnerabilities for therapeutic intervention. However, introducing oncogenic events into BM cells might favor the formation of leukemia besides the T cell lineage. Therefore, tight T cell specific expression will be the first striking challenge.

Based on the aim to generate a useful mouse model, the generation of various T-ALL subsets was aimed for, differing in their phenotype and reflecting different stages of T cell development at where transformation can take place. Thereby, the main focus was to model the Tlx1 and Tal1 T-ALL groups in the context of different aberrant cellular signaling pathways. To support their acute leukemia phenotype and prove their applicability as a reproducible leukemic cell source, developed leukemias will be transplanted into secondary recipient mice to observe the reproduction of the leukemia formation. Next-generation sequencing on the CRISPR-Cas9-targeted gene regions shall verify the genetic complexity and reveal a mono- or poly-clonal leukemia development.

In the second part of the thesis, it was intended to generate cell lines from obtained T-ALL. For this purpose, optimal culturing conditions have to be established for the leukemia cells, by relying on previous publications about murine T-ALL and known T cell differentiation protocols[38,104]. By sequencing the cell lines for the inserted genetic variations, we will gain an understanding of favorable cooperating events in the *in vitro* leukemia progression. In an all-combining RNA sequencing approach, transcriptomic differences between the different subtypes and alterations through the *in vitro* cultivation will help to evaluate the novel mouse model as suitable platform for future drug screenings.

Overall, this thesis will explore how different oncogenic events cooperate in the formation of T-ALL in mouse and will provide a clinically relevant model system, which will lead to a better understanding of T-ALL biology. This should help unravel new cellular dependencies and will serve as a reliable platform to test new treatment strategies in a clinically relevant disease context.

4 Materials and Methods

4.1 Molecular Biology

4.1.1 Buffer Composition

All self-made buffers and solutions used for molecular biology methods are listed in Table 2.

Table 2: Composition of buffers and solutions for molecular biology methods

Buffer	Composition
50x Tris-Acetate-EDTA (TAE) buffer	484 g Tris-HCl (Ambion, Dresden, Germany) 14.2 mL 100% Glacial acetic acid (Carl Roth, Karlsruhe, Germany) 200 mL 0.5M EDTA (pH 8.0) (Sigma Aldrich, Steinheim, Germany) Filled up to 2 L
1x P1 buffer	50 mM Tris-HCl 10 mM EDTA Adjust to pH 8 Add 100 µg/mL RNase A (Carl Roth)
1x P2 buffer	200 mM sodium hydroxide (Merck, Darmstadt, Germany) 1% Sodium dodecyl sulfate (SDS, BioRad, Hercules, CA, USA)
1x N3 buffer	4.2 M Guanidine hydrochloride (Carl Roth) 0.9 M Potassium acetate (Carl Roth) Adjust to pH 4.8
1x PE buffer	10 mM Tris-HCl (pH 7.5) 80% Ethanol (J.T. Baker, Thermo Fisher Scientific, Waltham, MA, USA)
1x EB buffer	10 mM Tris-HCl (pH 8.5)
1x TE buffer	10 mM Tris-HCl (pH 8.0) 1 mM EDTA (pH 8.0)
Lysogeny Broth (LB) medium	10 g/L Tryptone 5 g/L yeast-extract (Carl Roth, Karlsruhe, Germany) 5 g/L sodium chloride (Sigma-Aldrich, Steinheim, Germany) (100 µg/mL ampicillin (Sigma-Aldrich, Steinheim, Germany)

4.1.2 Cloning of viral vectors

4.1.2.1 Restriction enzyme digestion

Restriction digestion of plasmids and gene fragments was mainly performed with restriction enzymes from New England Biolabs GmbH (NEV; Frankfurt am Main, Germany) in NEBuffer™ 3.1 for BsmBI restriction enzyme or CutSmart® Buffer for all other enzymes. For digestions, DNA and enzymes were mixed in a total volume of 20 µL and incubated, if not otherwise recommended by the manufacturer, at 37°C for 1 hour. For further cloning, the plasmid

backbone could be dephosphorylated at their open 5'-ends by adding 1 unit (U) thermosensitive alkaline phosphatase (Thermo Scientific) to the digestion approach and incubation for 20 min at 37°C. Subsequently, enzymes were heat-inactivated for 10 min at 80°C.

4.1.2.2 Gel electrophoresis and gel purification

Depending on the size of generated DNA fragments, 1 – 2% 1x TAE agarose gel was prepared and used for gel electrophoresis. Therefore, UltraPure agarose (Invitrogen, Carlsbad, California, USA) was dissolved in 1x TAE buffer using microwave heating for 2 minutes. After short cooling, 3 – 4 µL MidoriGreen Advanced (Nippon Genetics, Düren, Germany) was added as DNA intercalator. As reference marker for fragment size determination, GeneRuler DNA Ladder Mix (Thermo Fisher) was used. If not already involved in the reaction mix, 4µL of 6x loading dye was added to each sample prior to loading onto the gel. For appropriate separation of DNA fragments, 90-140 V were applied for 30-50 min. In order to use the separated DNA fragments for further cloning or sequencing, the gel was placed onto an UV table and respective fragments were excised. Further purification was performed with the QIAquick Gel Extraction Kit (Qiagen, Hilden, Germany) according to manufacturer's protocol with elution in maximum 20 µL EB buffer.

4.1.2.3 Ligation and Gibson Assembly

In order to assemble few digested fragments, ligation was performed with purified DNA fragments in a 20 µL ligation approach, containing 2 µL 10x T4 ligase buffer (NEB) and 1 µL T4 ligase (400 U/µL; NEB). Depending on the size of deployed fragments, 2 µL backbone DNA was mixed with 3 – 6 µL insert DNA fragments. The reaction mix was incubated for 2 h at room temperature (RT) or overnight at 16°C.

In case of difficulties in the normal ligation approach or assembling more than three fragments together, Gibson Assembly was performed with NEBuilder® HiFi DNA Assembly (NEB). For this, used fragments were modified by polymerase chain reaction to contain 3' and 5' overhangs overlapping 20 base pairs (bp) from the neighboring fragment's sequence. Inserted fragments were used in a ratio of 2:1 on their number of pmol compared to the backbone. Respective amounts were mixed in a 20 µL approach with NEBuilder HIFI DNA Assembly Master Mix (NEB) and processed according to manufacturer's protocol.

4.1.2.4 Transformation and plasmid purification

Half of the assembled DNA mixture was used to transform 50 µL freshly thawed chemically competent *Escherichia coli* XL1-blue (Agilent Technologies, Waldbronn, Germany). After

20 min incubation of bacteria with DNA reaction on ice, heat-shock was performed at 42°C for 45 seconds, followed by 2 min cooling on ice. In case of retransforming complete plasmids, ~ 500 µg plasmid DNA were used in the incubation step. After heat-shock, 600 µL LB-medium (Luria / Miller, Carl Roth, Karlsruhe, Germany) without any antibiotics were added to the suspension and incubated for 45 – 60 min at 37°C on an orbital shaker. Pre-incubated bacteria culture was centrifuged at 3000 rpm for 3 min and plated onto a LB-agar plate, supplemented with 75 µg/mL ampicillin (Ratiopharm, Ulm, Germany). Following incubation at 37°C for 18 h led to the growth of bacteria colonies, which were picked and transferred into 2 – 3 mL LB-medium supplemented with ampicillin for overnight incubation at 37°C on an orbital shaker.

Obtained bacteria suspension contained the plasmid of interest. To isolate it, the suspension was transferred into a 1.5 mL Eppendorf tube (Eppendorf GmbH, Wesseling-Berzdorf, Germany) and centrifuged at 17,000 x g for 15 seconds. LB medium supernatant was discarded and 250 µL P1 buffer was added. Bacteria pellet was dissolved by strong vortexing prior to addition of 250 µL P2 buffer. After short inversion for mixture, 350 µL N3 buffer was added and inverted again. Lysates were centrifuged for 5 min at 17,000 x g, before supernatants could be transferred onto an EconoSpin® column for DNA (Epoch Life Science Inc., Missouri City, Texas, USA). Column was centrifuged for 30 seconds at 17,000 x g and flow-through was discarded. Membrane was washed with 700 µL PE buffer and column was centrifuged again for 30 seconds at 17,000 x g. The flow-through was discarded and the membrane was dried by centrifugation for 1 min at 17,000 x g. To elute the plasmid from the membrane, the column was placed into a new 1.5 mL collection tube. Up to 40 µL EB buffer were added directly onto the membrane, incubated for 2 min before a final centrifugation step at 17,000 x g for 1 min.

Larger amounts of plasmid DNA were isolated and purified utilizing the QIAGEN Plasmid Kits. In brief, up to 200 mL LB medium were inoculated with remaining bacteria suspension or grown colonies and incubated at 37 °C at 200 rpm on an orbital shaker overnight. Next, bacteria suspension was centrifuged at 6000 x g for 10 min. Bacteria pellet was resuspended in P1 buffer, supplemented with RNAse A and mixed with P2 buffer. After 5 min incubation at room temperature, pre-chilled P3 buffer was added for neutralization and incubated for another 20 min on ice. Following centrifugation for 10 min at 8500 x g allowed removal of cell debris. The supernatant was applied onto a DNA-binding column, previously equilibrated with QBT buffer. The loaded column was washed twice with QC buffer, before plasmid DNA was eluted with QF buffer. Following the addition of 0.7 amounts of isopropanol led to precipitation, which could be enriched by centrifugation at 8500 x g at 4°C for at least 30 min. The supernatant was discarded and a white pellet, containing the plasmid precipitate was transferred to a 1.5 mL Eppendorf tube and washed twice with 70% ethanol, to remove remaining isopropanol. The

pelleted plasmid DNA was air-dried and solved in 100 – 200 µL TE buffer, depending on the pellet size. Plasmid concentration was determined using the NanoDrop spectrophotometer (Peqlab, Erlangen, Germany) and concentration was adjusted to 1 µg/µL with TE buffer.

4.1.2.5 Sequence verification

Sanger sequencing was performed for two purposes. Either to confirm correct cloning of gained plasmids or to analyze insertions and deletions (indels) in particular PCR-amplified genes (see 4.4.1). Prior to sequencing of plasmids, control digestion was performed with 0.3 µL restriction enzyme. Sanger sequencing was performed at Microsynth Seqlab (Göttingen, Germany). Therefore, 500 – 1000 ng DNA were mixed with 3 µL sequencing primer in a total volume of 15 µL. Gained sequences were aligned against the computational reference plasmid using clone manager 9 software (Sci-Ed Software, Denver, CO, USA).

4.1.2.6 Cloning strategy for various used vectors

4.1.2.6.1 Cloning strategy of lox-site containing vectors

The lentiviral pL40C vector served as vector backbone for all used vectors and was kindly provided by Dr. Adrian Schwarzer, PhD[150]. It contains a modified wPRE element for lentiviruses (LPRE[151]), but will be referred as WPRE.

A codon-optimized *dTomato* flanked by wild-type loxP sites or mutated loxP sites (lox66/71) was cloned from a gBlock™ generated by the company Integrated DNA Technology (IDT) into the pL40C backbone (exemplary design in Figure 9). Downstream of *dTomato* followed the inverted sequence of the gene-of-interest. In the final vector design, two AgeI restriction sites flank the gene-of-interest (oncogenes or reporter) to allow an easy exchange of this cassette. The whole floxed *dtomato*-gene-of-interest cassette is flanked by BamHI and MluI sites to allow the exchange of vector backbones. For initial comparison of different loxP vector systems, an *mTagBFP2* was cloned from gBlock™ either in sense (for wild-type loxP) behind *loxP*.*dTomato*.*loxP* or antisense (for *lox66/71*) between *lox66.dTomato* and the *lox71* site.

Murine *Tlx1*, human *IL7R*, containing a TTGTCCCAC insertion between base pairs 731 and 732 (*IL7R*$^{c.731\_732insTTGTCCCAC}$, named *IL7R*$^{c.731ins732}$), murine *Tal1* and murine *Lmo2* were ordered as gBlock™ at IDT, containing 5' tags (Table 4) for later detection and respective restriction sites for cloning. *NRAS*^{G12D} was kindly provided by Daniel Brand and Michael Morgan. Restriction sites and FLAG-tag were added to *NRAS*^{G12D} with the primers from Table 3. The *IL7R*$^{c.731ins732}$ sequence was provided by Dr. Teng Cheong Ha. A *PIK3CD*^{E1021K} encoding plasmid was bought at Addgene (Watertown, Massachusetts, USA) (plasmid #116566)[152].

After cloning, correct orientation of the oncogene/reporter cassette was verified by control digestion and sequencing from the codon-optimized *dTomato* and WPRE elements, using the sequencing primers from Table 8.

Table 3: Primer pair for addition of tag and restriction sites for *NRAS^{G12D}* and *mTagBFP2*

Primer	Sequence 5' – 3'
NRAS^{G12D}_BamHI_FLAG_fw	CAGACTGAGTCGGATCCTAACCGGTGAGCCACCATGGACTACAAGGACGACGATGACAAGATGACTGAGTACAAACTGGTGGTGGTTGGAG
NRAS^{G12D}_MluI_rev	GATTGTCGACTTAACGCGTGACCGGTCTTACATCAGCACACAGGGCAGCCC
mTagBFP2_AgeI_FW	CGGCAATTGAACCGGTCGCCACCATGAGCGAGCTGATTAAGGAGAACATG
mTagBFP2_AgeI_rev	GCAATAACTTCACCGGTCTCACTTGTACAGATTAAGCTTGTGCCCCAG

Table 4: Overview of protein tags on used transcription factors and signaling proteins.

Gene	Protein-tag	Sequence 5' – 3'
Tlx1 *Tal1*	HA	ATGTACCCCTACGACGTGCCCGACTATGCC
IL7R$^{c.731ins732}$ *PIK3CD^{E1021K}* *NRAS^{G12D}*	Flag	GACTACAAGGACGACGATGACAAG
Lmo2	Myc	GAACAAAAACTCATCTCAGAAGAGGATCTG

4.1.2.6.2 Cloning of reporter vector for sgRNA efficiency analysis

The reporter vector for validation of sgRNA knockout capacity (see 4.2.7) was cloned by inserting the target site cassette via AgeI and MluI restriction sites into the pRRL.SFFV.sfEGFP.IRES.Puro.WPRE (kindly provided by Dirk Heckl[149]) backbone. Up to 20 target sites and their original PAM sequence were lined up and put in-frame 5' of the superfolder GFP (sfGFP). The cassette was ordered at IDT as gBlock.

4.1.2.6.3 Cloning of sgRNAs containing plasmids

Plasmids, containing sgRNAs were cloned by inserting freshly annealed sgRNAs via BsmBI sites into the pL40C.U6.sgRNA.EFS.dTomato.WPRE backbone for screening purposes and the most suitable for final application into the pL40C.U6.Stuffer.sgRNA.EFS.Thy1.1.WPRE (both kindly provided by Dirk Heckl). The fluorescence reporter dTomato or the thymus cell antigen 1.1 (Thy1.1 or CD90.1) under the control of EF-1 alpha short (EFS) promotor served to determine transduction efficiency. All sgRNAs were designed with the online tool CCTop

(https://cctop.cos.uni-heidelberg.de). For each target gene, three highly scored guides were ordered as oligonucleotides at IDT, containing a CACC overhang at the sense strand and an AAAC overhang at the antisense site (Table 5). All sgRNAs started with a G at the 5' site of the sense strand, even if there was none in the original target sequence (see Table 5). Length varied between 19 and 20 nucleotides. Oligonucleotides were prepared as 100 pmol stock and used in an oligo reaction mix (see Table 6).

Table 5: List of suitable guide sequences, their PAM sequences for various genes and later usage for *in vivo* studies

Target gene	Guide sequence 5' – 3'	PAM	length	Used in mice
m*Bcl11b*_CR1	GTGGAGGCTACCATCCTCG	AGG	19	No
mBcl11b_CR2	GAGGATGGTAGCCTCCACA	TGG	19	No
mBcl11b_CR3	GACCCTGATCTACTCACCTG	TGG	20	Yes
m*Phf6*_CR1	GTGGTGTGCTGCCACCTTC	TGG	19	No
m*Phf6*_CR2	GCACAGCCAATGGTTGCTCC	AGG	20	Yes
m*Phf6*_CR3	GTGGTAGTGGTATGTCCTG	TGG	19	No
m*Dnm2*_CR1	GAGGTGGCAGCTCTGACCGA	TGG	20	Yes
m*Dnm2*_CR2	GATCGGTCAGAGCTGCCACC	TGG	20	No
m*Dnm2*_CR3	GATCGAAGCAGAGACTGACC	GGG	20	No
m*Etv6*_Cr1	GCGGTAGCGGAAATCCTCTT	TGG	20	No
m*Etv6*_Cr2	GCAGAGATGACGTAGCCCAG	TGG	20	No
m*Etv6*_Cr3	GCTCCTGCTGCTGACCAAAG	AGG	20	Yes
m*Pten*_CR1	GCTCAGCCATTGCCTGTGTG	TGG	20	Yes
m*Pten*_CR2	GTCAGGACCCACGCGGCGGG	AGG	20	No
m*Ptpn2*_CR1	GTGCTGCCATTTCTGGCTCA	TGG	20	Yes
m*Ptpn2*_CR2	GATCTGGCCAGGTGGTATAA	TGG	20	No
m*Ptpn2*_CR3	GCCCGATGCCCGCACTGCAA	TGG	20	No
m*Cdkn2a*_CR1	GCCCGGTGCACGACGCAGCG	CGG	20	Yes
m*Cdkn2a*_CR2	GCAGCTCTTCTGCTCAACTA	CGG	20	No
m*Cdkn2a*_CR3	GTGGTGGTGCTGCACGGGTC	AGG	20	No
m*Usp7*_CR1	GTCTGTCTGGATAAAAGCG	TGG	19	No
m*Usp7*_CR2	GTTTTGTGCGAAATCTGCCA	TGG	20	Yes
m*Lef1*_CR1	GACGGGTCGCTGTTCATATT	GGG	20	Yes
m*Lef1*_CR2	GACGGAGGCCTGTACAACAA	GGG	20	No
m*Lef1*_CR3	GTGGTGCAGCCCTCTCACG	CGG	19	No

Table 6: Annealing and phosphorylation mix for oligonucleotides

Component	Volume
FW oligo (100 µM)	1 µL
REV oligo (100 µM)	1 µL
T4 DNA ligase buffer	1 µL
T4 PNK	0.5 µL
H_2O	6.5 µL

Annealing and phosphorylation of 100 µM sgRNA single strands were performed as reaction mix (Table 6) in a PCR cycler with 37°C for 45 min, followed by 95°C for 2 min and a cool down to 4°C at 0.1°C/s. The sgRNA duplex was next freshly diluted 1:500 in H_2O for cloning and used in a ligation mix (Table 7). Ligation was performed as described before and transformed into competent E.coli (see 4.1.2.4) afterwards. Correct insertion of sgRNA sequence into the vector backbone was confirmed by sequencing from the hU6 promoter (Table 8).

Table 7: Ligation mix for cloning sgRNAs into lentiviral backbones

Component	Volume
BsmBI digested backbone	30-50 ng
Annealed and phosphorylated oligos (1:500)	1 µL
T4 DNA ligase buffer	1 µL
T4 DNA ligase	1 µL
H_2O	Up to 10 µL

4.1.2.6.4 Cloning of double sgRNA CRISPR vector

The most suitable sgRNAs for *Cdkn2a* and *Notch1* (Pest domain) were cloned into a two-in-one vector, containing a human and a murine U6 promoter. First, the most suitable sgRNA for *Notch1* was ordered at IDT, containing a TGTT overhang and an additional T at the 5' site of the sense strand and an AAAC overhang and complementary A at the antisense strand. Oligonucleotides were annealed and phosphorylated as described before (see 4.1.2.6.3) and cloned into pSGL40C-NL-SL.mU6.EFS.tRFP657.PRE (kindly provided by Dirk Heckl) via *BsmbI* restriction sites. The whole mU6.sgRNA cassette was subsequently cloned into pL40C.U6.Cdkn2a.EFS.Thy1.1.WPRE via *XhoI* and *EcoRI* restriction sites. Correct insertion of sgRNA sequences into the vector backbone was confirmed by sequencing from the hU6 and mU6 promoter (Table 8).

Table 8: Sequencing primers for lentiviral vectors

Primer binding site	Sequence 5' – 3'
SFFV promoter	GCGCTTCTGCTTCCCGAGCTC
hU6 promoter	GACTGTAAACACAAAGATATTAG
mU6 promoter	AAAGGAAACTCACCCTAACTG
dTomato (codon optimized)	GCCCGGCTACTACTATGTTG
WPRE	AGCGTAAAAGGAGCAACATAG

4.1.3 Polymerase Chain Reaction (PCR)

PCR was used to insert restriction site(s), amplify specific genes from other plasmids for cloning purposes, to conduct genotyping or investigate indel formation and mutations in certain genes in primary cell material. Primers, which bind to plasmids or genomic DNA (gDNA), were designed at https://www.ncbi.nlm.nih.gov/tools/primer-blast/ and synthesized by Eurofins Genomics (Ebersberg, Germany). Reaction was performed using the Phusion Hot Start II High-Fidelity PCR Master Mix (Thermo Fisher Scientific, Schwerte, Germany) or Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific) according to the manufacturer's instructions (setup in Table 9). PCR reaction setup and elongation programs used for the different PCR approaches are summarized in Tables 9 – 11.

Table 9: Setup for PCR reaction with Phusion or Phire 2x Master Mix

Component	Volume
2x Phusion Green HS II HF Master Mix / 2x Phire Tissue Direct PCR Master Mix	10 µL
Forward Primer (10 mM)	1 µL
Reverse Primer (10 mM)	1 µL
Template DNA	50 – 500 ng
H_2O	Up to 20 µL

Table 10: PCR program for Phusion Green HS II HF Master Mix

Temperature	Time min:second	Cycles
98°C	2:00	1
98°C	0:10	1
60-65°C	0:15	35
72°C	0:30	35
72°C	10:00	35

Table 11: PCR program Phire Tissue Direct PCR Master Mix

Temperature	Time min:second	Cycles
98°C	5:00	1
98°C	0:05	1
60-65°C	0:05	35
72°C	0:20	35
72°C	5:00	35

4.1.3.1 PCR approach for proving Cre-mediated vector inversion

Proving successful inversion of the primary oncogenes from 4.1.2.6.1 in presence of Cre recombinase was performed in an all-in-one PCR approach, using Phusion Green HS II HF Master Mix. Three primers were mixed together and orientation of the floxed inverse construct

could be analyzed by gel-electrophoresis. The forward primer binds upstream of the SFFV promoter, while the reverse primers bind within the *dTomato* or the *Tlx1*/*Tal1* sequence. For the PIK3CD^{E1021K} vector, one primer binds in the WPRE element, instead of *dTomato* (see Figure 19). Depending on the orientation, only one of the reverse primers mediates the PCR reaction. Table 12 summarizes primer sequences and expected fragment size.

Table 12: Primer pairs for inverse vector orientation in leukemic samples

Primer name	Sequence 5' – 3'	Fragment size
backbone_FW	CAGCAGAGATCCACTTTGGC	
dTomato_Rev	CGGTTACAAGTCCACCGTCT	758 bp
Tlx1_Rev	CAAAGGAGATGGGTTCTGCG	587 bp
backbone_FW	CAGCAGAGATCCACTTTGGC	
backbone_Rev	GCAACCAGGATTTATACAAGGAGG	309 bp
PIK3CD_Rev	TCCAGAATTCCATGGGGCAG	581 bp
backbone_FW	CAGCAGAGATCCACTTTGGC	
dTomato_Rev	CGGTTACAAGTCCACCGTCT	758 bp
Tal1_Rev	CTGACCTAGCAGCCTCAGAG	567 bp

4.1.3.2 PCR amplification of genomic regions for targeted sequencing

For targeted amplification of genomic loci that were potentially targeted by CRISPR-Cas9, PCR primers were designed to bind 60 – 80 bp upstream and 80 – 100 bp downstream from the PAM sequence of the sgRNA target binding site of the respective Cas9-targeted locus. In addition, overlapping sequences for a following nested PCR were added 5' of each sequence. A summary of all primer pairs can be found in Table 16. Each locus was amplified with Q5® High-Fidelity DNA Polymerase (NEB) using the reaction setup and elongation program from Tables 13 and 14. Afterwards, 5 µL of each reaction approach were put onto a 1.5% 1xTAE agarose gel to verify successful PCR reaction, while the other remaining 15 µL were mixed with 75 µL PB buffer (Qiagen). The mixture was directly transferred to an EconoSpin® column (Epoch Life Science Inc.). Column was centrifuged as described before for 30 seconds at 17,000 x g and flow-through was discarded. Membrane was washed with 700 µL PE buffer, centrifuged again for 30 seconds at 17,000 x g and finally dried by centrifugation for 1 min at 17,000 x g. For elution, 20 µL EB buffer were added directly onto the membrane. Final centrifugation occurred into a new 1.5 mL collection tube at 17,000 x g for 1 min. A nested PCR on the first product added P5/P7 adapters, barcodes and sequencing primer binding regions to each amplified locus (PCR setup Table 14 and 15, exemplary primer design in Table 17). All amplified genomic loci from the same animal received the same genetic barcode.

Purification of the final PCR approach occurred as described above. Sample concentration was determined using the Nanodrop spectrophotometer. Equal amounts of all PCR products were mixed together and used for next-generation sequencing (NGS, see 4.4.2).

Table 13: PCR program Q5® High-Fidelity DNA Polymerase – first PCR

Temperature	Time min:second	Cycles
98°C	0:30	1
98°C	0:10	30
64°C	0:30	30
72°C	2:00	1

Table 14: Setup for primary and secondary PCR reaction with Q5® High-Fidelity DNA Polymerase

Component	Volume
Q5® High-Fidelity DNA Polymerase	10 µL
Forward Primer (10 mM)	1 µL
Reverse Primer (10 mM)	1 µL
Template DNA/ Purified first PCR product	20 – 50 ng 2 µL
H_2O	Up to 20 µL

Table 15: PCR program Q5® High-Fidelity DNA Polymerase – second PCR

Temperature	Time minute:second	Cycles
98°C	0:30	1
98°C	0:10	12
66°C	0:10	12
72°C	0:30	12
72°C	2:00	1

Table 16: Primer pairs for targeted sequencing

Genomic locus	Sequence 5’ – 3’	Fragment size
Bcl11b fw	TTGTGGAAAGGACGAAACACCTCGAGGAAGACGAGGGTCTG	212 bp
Bcl11b rev	TTGTGGATGAATACTGCCATTTGTCTCCAGGCCTCCACACTGTTTCT	
Cdkn2a fw	TTGTGGAAAGGACGAAACACCAGCTCTTCTGCTCAACTACGG	248 bp
Cdkn2a rev	TTGTGGATGAATACTGCCATTTGTCTCCGATGTCTTGATGTCCCCGC	

Table 16: Primer pairs for targeted sequencing (continued)

Genomic locus	Sequence 5' – 3'	Fragment size
Notch1-Pest fw	TTGTGGAAAGGACGAAACACCTGGGCCCCAGCAGTCT	198 bp
Notch1-Pest rev	TTGTGGATGAATACTGCCATTTGTCTCCCACAGGGGAGGAGGAGTAA	
Dnm2 fw	TTGTGGAAAGGACGAAACACCGCGGAGGCGGACACCAT	230 bp
Dnm2 rev	TTGTGGATGAATACTGCCATTTGTCTCGCTCACCGGCCCACG	
Etv6 fw	TTGTGGAAAGGACGAAACACCGAAAATGAGTTTTCTTTAAGGCCCA	235 bp
Etv6 rev	TTGTGGATGAATACTGCCATTTGTCTCTCCCATGGGTTATGATGAGTCTG	
Lef1 fw	TTGTGGAAAGGACGAAACACCTGGATTCTCCCAGGAAAGCAT	208 bp
Lef1 rev	TTGTGGATGAATACTGCCATTTGTCTCCATCAGCATGTTACTCACTGTCC	
Phf6 fw	TTGTGGAAAGGACGAAACACCCCTAGCCTGTTTGGGACAGT	240 bp
Phf6 rev	TTGTGGATGAATACTGCCATTTGTCTCTCCCTTGCGAAGGTTTCTCTC	
Pten fw	TTGTGGAAAGGACGAAACACCTCCTCCAATTCAGGACCCAC	216 bp
Pten rev	TTGTGGATGAATACTGCCATTTGTCTCTGAGGAATGGCTATTTCTACACA	
Ptpn2 fw	TTGTGGAAAGGACGAAACACCTCCAGAGTCACCAGCTTCATTT	225 bp
Ptpn2 rev	TTGTGGATGAATACTGCCATTTGTCTCCACCAATGAGATCGCATGACA	
Usp7 fw	TTGTGGAAAGGACGAAACACCCTTTCAGTTCACTGTCGAGCG	241 bp
Usp7 rev	TTGTGGATGAATACTGCCATTTGTCTCGCATCAGTTAAACTCTTACGTGG	

Table 17: NGS sequencing barcode primers

	P5 adapter - primer binding site - barcode - overlap
forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTTCTCAAAGTTGTGGAAAGGACGAAACACC
	P7 adapter - overlap
reverse	CAAGCAGAAGACGGCATACGAGATTTGTGGATGAATACTGCCATTTGTCTC

4.1.4 Genomic DNA (gDNA) isolation

Up to 5 x 10^6 cells were harvested, washed with phosphate-buffered saline (PBS), centrifuged for 5 min at 300 x g and either stored as cell pellet at -20°C or directly resuspended in 200 µL PBS for further preparation. The isolation of gDNA was performed using the QIAamp DNA Blood Mini Kit (Qiagen) according to manufacturer's instruction. Depending on the number of harvested cells, elution was done in 40 – 100 µL AE buffer from the kit. DNA concentration was measured using 2 µL of sample with the Nanodrop spectrophotometer.

4.1.5 RNA extraction

Cells were harvested the same way as for gDNA isolation, by washing them with PBS and centrifuging at 300 x g for 5 min. For isolation, the RNeasy Mini Kit (Qiagen) was used according to the manufacturer's protocol. If not directly performed upon harvest, cell lysates were stored after disruption in RLT buffer at -80°C. For all samples, removal of DNA was performed with DNase accordingly. After isolation, RNA was eluted in up to 40 µL RNAse-free 1 mM Tris and stored at -80°C. RNA concentration was measured with the Nanodrop spectrophotometer.

RNeasy Mini Kit was used for RNA extraction from all leukemia samples and cell lines. However, for the isolation of RNA from freshly sorted thymocytes, the Direct-zol RNA Miniprep Kit from Zymo research (Irvine, California, USA) was used. A maximum of 400,000 thymocytes (see 4.2.12) were sorted into 1.2 mL Trizol LS (Thermo Fisher Scientific) to preserve the 3:1 ratio of water to Trizol LS. After sorting, the volume was measured and filled to 1.6 mL using nuclease-free distilled water (UltraPure™, Invitrogen, Thermo Fisher Scientific, Schwerte, Germany). The following purification was carried out according to the manufacturer's instructions. Final elution occurred in 15 µL nuclease-free water (Zymo research). RNA concentration was measured with the Nanodrop spectrophotometer.

4.1.6 VCN by real-time quantitative polymerase chain reaction (RT-qPCR)

In order to determine the number of vector copies in transduced murine lineage-negative bone marrow cells, a quantitative real-time TaqMan PCR was accomplished. Two primer sets and probes (Table 18) detect the vector intrinsic WPRE element and the endogenous and conserved DNA locus for polypyrimidine tract binding protein 2 (*PTBP2*), which allows the quantitative calculation of vector insertions[109]. Each probe was labeled with either FAM or JOE as fluorescent dye and a quencher, which suppresses the fluorescence signal until its release by the Taq polymerase during extension. In a 15 µL RT-qPCR reaction approach, containing 13 µL master mix (see Table 19) and 2 µL gDNA (30-50 ng) were mixed and loaded into a MicroAmp 96-well plate. The plate was sealed with a MicroAmp™ Optical Adhesive Film

(Applied Biosystems, ThermoFisher Scientific). Amplification and detection of the fluorescence were done on a StepOnePlus Real-Time PCR System (Thermo Fisher Scientific). The used temperature protocol is summarized in Table 20.

Table 18: Probes and primers used for VCN determination by RT-qPCR

Target	Sequence 5' – 3'	Labeling
WPRE forwards	GAGGAGTTGTGGCCCGTTGT	
WPRE reverse	TGACAGGTGGTGGCAATGCC	
WPRE probe	CTGTGTTTGCTGACGCAAC	5'-FAM, 3'-BHQ1
PTBP2 forwards	TCTCCATTCCCTATGTTCATGC	
PTBP2 reverse	GTTCCCGCAGAATGGTGAGGTG	
PTBP2 probe	ATGTTCCTCGGACCAACTTG	5'-JOE, 3'-BHQ1

Table 19: Setup master mix for VCN determination using a RT-qPCR approach

Reagent	Concentration	Volume [µL]
ABI Taqman Fast Advanced Master Mix	2x	7.5
Primermix WPRE	20 µM	0.5
Probe WPRE	20 µM	0.11
Primermix PTBP2	20 µM	0.5
Probe PTBP2	20 µM	0.11
H_2O		4.29

Table 20: PCR program for RT-qPCR using TaqMan

Cycle step	Cycles	Temperature [°C]	Time [min:s]
	1	50	02:00
Heat inactivation	1	95	00:20
Denaturation	40	95	00:05
Annealing	40	56	00:20
Extension	40	65	00:20

Evaluation of the results and extraction for further analysis was done with the Step-One Software (Applied Biosystems, ThermoFisher Scientific). To verify an optimal amplification during each cycle, a dilution series with standard plasmids was run, too. The regression curve based on these values was used to calculate the efficiency and applied to determine the absolute copy number of PRE and PTBP2[109].

4.1.7 Western Blot

Cells were washed with ice-cold PBS, pelleted by centrifugation and incubated in 100 – 200 µL RIPA lysis buffer (Cell Signaling, Danvers, Massachusetts, United States) for 10 min on a shaking Table at 4 °C. Following shock freezing with liquid nitrogen disrupted the remaining

cell fragments. Lysates were centrifuged for 10 min at 17,000 x g and the supernatant was stored at -80°C. The protein concentration was determined with Thermo Scientific™ Pierce™ BCA Protein Assay Kit (Thermo Fisher Scientific) according to manufacturer's instructions. Equal amounts of protein (15 µg or 20 µg) were mixed from each sample with 3x Blue Loading Buffer (Cell Signaling) + 1/30 dithiothreitol (DTT; Cell Signaling) and denatured for 5 min at 95°C. Samples were loaded onto a 12% polyacrylamide gel in SDS running buffer and separated by applying 30 mA per gel. Proteins were blotted onto a nitrocellulose membrane using the semi-dry Trans-Blot® SD Semi-Dry Transfer Cell (BioRad) and manufacturer's transfer buffer for 30 min at 25 V and up to 1.0 A. Nitrocellulose membrane was washed with TBS-T and blocked for one hour in 5% non-fat milk at room temperature. Primary antibody staining was performed with 5% bovine serum albumin (BSA; PanBiotech, Aidenach, Germany) in TBS-T overnight at 4 °C (Table 21). Membrane was washed three times with TBS-T and stained with the horseradish peroxidase-conjugated secondary antibody in 5% non-fat milk for two hours at room temperature. After washing with TBS-T and finally TBS only, signal was detected using the SuperSignal West Pico Chemiluminescent Substrate (Thermo Fisher Scientific) or SuperSignal West Femto Chemiluminescent Substrate (Thermo Fisher Scientific) and the FusionX detection system (Peqlab). Stripping of stained membranes was performed with stripping buffer for 30 min in a 70°C water bath. After four washings, membrane was blocked and stained again as described above. A list of used buffers and their compositions are listed below (Table 21).

Table 21: Buffer and gel compositions for Western Blot

Buffer	Composition
separation buffer (pH = 8.8)	1.5 M Tris-HCl in H_2O
collection buffer (pH = 6.8)	0.5 M Tris-HCl in H_2O
10% SDS	100 g SDS (Biorad) 1 L H_2O (68 °C until completely solved)
10% separation gel	2.9 mL H_2O 3.75 mL separation buffer (pH = 8.8) 3.35 mL 30% Acrylamide (BioRad) 100 µL 10% SDS 6 µL Tetramethylethylenediamine (TEMED) (Sigma Aldrich) 100 µL Ammonium persulfate (APS) (Sigma Aldrich)
5% collection gel	3.35 mL H_2O 0.625 mL collection buffer (pH = 6.8) 0.85 mL 30% Acrylamide (BioRad) 50 µL 10% SDS 5 µL (TEMED)

Table 21: Buffer and gel compositions for Western Blot (continued)

Buffer	Composition
	50 µL APS
10x TBS	24.2 g Tris-HCl 80 g NaCl in 1 L H_2O (adjusted to pH = 7.6)
1x TBST	900 mL H_2O 100 mL 10x TBS 1 mL Tween 20 (Carl Roth)
stripping buffer	36.52 mL H_2O 10 mL 10% SDS 3.12 mL 1 M Tris-HCl (pH 6.8) 0.36 mL β-Mercaptoethanol

Table 22: Antibodies used for Western Blot

Antibody	Clone	Manufacturer	Dilution
mTOR	7C10	Cell Signaling	1:1000
P-mTOR	D9C2	Cell Signaling	1:1000
STAT5	A-9	Santa Cruz Biotech	1:1000
P-STAT5	D47E7	Cell Signaling	1:1000
AKT	C67E7	Cell Signaling	1:1000
P-AKT-S473	D9E	Cell Signaling	1:2000
ERK	9102	Cell Signaling	1:1000
P-ERK	D13.14.4E	Cell Signaling	1:1000
S6RP	5G10	Cell Signaling	1:1000
P-S6RP	D68F8	Cell Signaling	1:2000
p70S6K	49D7	Cell Signaling	1:1000
P-p70S6K	108D2	Cell Signaling	1:1000
4-EBP1	53H11	Cell Signaling	1:1000
P-4-EBP1	236B4	Cell Signaling	1:1000
Cre	69050	Novagene	1:5000
Cas9	7A9	Biolegend	1:1500
b-Tubulin	9F3	Cell Signaling	1:1000
GAPDH-HRP	GT239	Genetex	1:8000
Goat anti rabbit-HRP	10004301	Cayman Chemicals	1:4000
Goat anti mouse-HRP	10004302	Cayman Chemicals	1:4000

4.2 Cell Culture

4.2.1 Buffers and cell culture media

Table 23: Composition of buffers and media used in cell culture

Media/Buffer	Composition
DMEM complete medium (for HEK293T, HT1080 cells)	Dulbecco's modified Eagle's medium (DMEM, 4.5 g/l glucose; Biochrom, Berlin, Germany) 10% heat-inactivated fetal bovine serum (FBS) 100 U/ml penicillin and 100 µg/ml streptomycin (Pen/Strep) 10 mM HEPES buffer (Gibco/Thermo Fisher Scientific) 1 mM sodium pyruvate (all PanBiotech, Aidenach, Germany)
RPMI complete medium (for 32D, BAF3 cells, T-ALL cell lines)	Roswell Park Memorial Institute (RPMI) 1640 medium (PanBiotech) 10 % heat-inactivated FBS Pen/Strep
αMEM reconstituted	10 g α-MEM, no nucleosides, powder (Gibco/Thermo FisherScientific) 2.2 g/L of $NaHCO_3$ (Pan Biotech) Adjusted to pH 7.20 and brought to final volume of 1 L with distilled water Sterile filtered through 0.22 µm filter - stored at 4°C
αMEM complete medium (for OP9, OP9DL1 cells)	αMEM reconstituted 20% heat-inactivated FBS (Standard or Brazil One) (PanBiotech) Pen/Strep
αMEM for T cell differentiation	αMEM reconstituted 20% heat-inactivated FBS Brazil One Pen/Strep 20 ng/mL mSCF (Peptrotech, Hamburg, Germany) 5 ng/mL mFLT3L (Peptrotech) 1 ng/mL mIL7 (Peprotech, Hamburg, Germany)
StemSpan medium	StemSpan™ SFEM (Serum-free medium for culture and expansion of hematopoietic cells; StemCell Technologies, Vancouver, Canada) Pen/Strep 2 mM L-glutamine (Biochrom)
Murine HSPC cultivation medium "StemSpan + STF"	StemSpan medium 50 ng/mL mSCF (Peprotech) 50 ng/mL mTPO (Peprotech) 50 ng/mL hFLT3L (Peprotech) 20 mM Meropenem (Inresa, Freiburg)
Freezing medium	90% FBS 10% dimethyl sulfoxide (DMSO, Merck)

Table 23: Composition of buffers and media used in cell culture (continued)

Media/Buffer	Composition
FACS buffer	PBS 4% FBS 2 mM EDTA (Invitrogen / Thermo Fisher Scientific)
MACS buffer	PBS 0.5% BSA (PanBiotech) 2 mM EDTA
Red blood cell (RBC) lysis buffer	8.26 g Ammonium chloride 1 g Potassium bicarbonate 0.037 g EDTA Dissolved and diluted to a final volume of 1 L dH2O Adjusted pH to 7.3
HeBs buffer	50 mM HEPES buffer (Gibco/Thermo Fisher Scientific) 280 mM NaCl (Sigma-Aldrich) 1.5 mM Na_2HPO_4 (Sigma-Aldrich) Adjusted to pH 7.0

4.2.2 Lentiviral vector production

Lentiviral vectors were produced by calcium-phosphate transfection of human embryonic kidney (HEK) 293T cells (German Collection of Microorganisms and Cell Cultures GmbH – DSMZ number ACC 635) as described before[109,153].

For maintenance, HEK293T cells were cultivated in DMEM complete medium at 37 °C and 5% CO_2 and splitted twice per week using Trypsin/EDTA (PanBiotech). One day prior to transfection, 12 x 10^6 HEK293T cells were seeded onto 15 cm dishes (TPP, Trasadingen, Switzerland). When cells reached 80% confluency the following day, cultivation medium was replaced by DMEM complete, supplemented with 20 mM HEPES buffer and 25 µM chloroquine (Sigma Aldrich, Steinheim, Germany). In the meantime, 20 µg vector of interest were mixed together with 30 µg pcDNA3.HIV-1.GP 4xCTE (LV gag/pol) [148], 12.5 µg of pRSV-Rev (only for LV, provided by T. Hope, Northwestern University, Chicago, IL, USA), 1.75 µg of pMD.G (VSVg pseudotype[154]). Plasmid mixture was transferred to 1250 µL UltraPure water (Invitrogen), containing 0.25 M $CaCl_2$ (Merck), mixed slowly with 1250 µL 2x Hepes-buffered saline (HeBs) solution with continuous bubbling and incubated at room temperature for 15 min to let the DNA and calcium form precipitates. Afterwards, 2.5 mL DNA mixture was added dropwise onto the HEK293T cells and incubated for at least eight hours at 37 °C or overnight. Next, the medium was completely removed and replaced by 36 mL DMEM complete medium with 20 mM HEPES. In case the vectorized construct contained a fluorescence reporter, cells were checked under a fluorescent microscope for the successful transfection of HEK293T cells.

After a maximum of 48 h, supernatants were harvested and filtered through a 0.22 µm filter (Merck). Concentration of viral particles was performed by ultra-centrifugation for 2 h at 25,000 rpm (106,800 x g) at 4 °C or >14 h at 10,000 rpm (17,100 x g) at 4 °C (Rotor: SW32Ti, Beckman Coulter, Krefeld, Germany). Pelleted virus particles were resuspended in 200 µL StemSpan medium (StemCell Technologies), aliquoted in 1.5 mL tubes and stored at -80°C.

4.2.3 Viral vector titration

Lentiviral particles were titrated either on HT1080 or directly on lineage-negative bone marrow cells (see 4.3.3).

For maintenance, HT1080 cells were cultivated in DMEM complete medium at 37 °C and 5% CO_2 and were passaged twice per week using Trypsin/EDTA (PanBiotech). In contrast, lineage-negative BM cells were freshly thawed and cultivated in StemSpan + STF (see Table 23) supplemented with 20 mM Meropenem (Eberth, Ursensollen, Germany) and used two days after thawing for titration. Therefore, 100,000 HT1080 cells or 50,000 lineage-negative BM cells were seeded to 12-well or 96-well plates, respectively. While HT1080 cells needed to attach for at least four hours, lineage-negative BM cells were transduced instantly. Medium was changed to either DMEM complete with 4 µg/mL protamine sulfate or StemSpan + SFT with 4 µg/mL protamine sulfate and 1 mM Poloxamer Synperonic® F 108 (Sigma Aldrich). Various dilutions and volumes of concentrated viral supernatant was added and incubated overnight at 37°C. HT1080 were additionally centrifuged for one hour at 2000 rpm (863 x g) and 37 °C to increase gene transfer. The next day, media were replaced and cells were analyzed by flow cytometry after 5 days post-transduction. The viral titer was defined from the mean of each tested volume (only conditions with transgene positive cells < 30% were used) calculated with the following formula:

$$Titer\ \left(\frac{TU}{mL}\right) = \frac{\frac{transgene\ positive\ cells\ (\%)}{100}\ x\ total\ seeded\ cell\ number\ x\ dilution\ factor}{supernatant\ volume\ (mL)}$$

4.2.4 Cell culture of suspension cells (32D and BAF3)

The murine bone marrow cell line 32D (DSMZ number ACC 411) and the murine pro-B cell line BAF3 (DSMZ number ACC 300) were cultivated in RPMI complete medium with 2 ng/mL murine interleukin 3 (IL3; Peprotech). All cells were cultivated at 37 °C and 5% CO_2 in a humidified incubator (Binder, Tuttlingen, Germany)

4.2.5 Transduction of 32D, BAF3 cells

Suspension cells were transduced by transferring 100,000 cells in 400 µL RPMI complete medium, supplemented with 4 µg/mL protamine sulfate and 1 mM Poloxamer into 12-well suspension plate. Viral supernatant was added at appropriate multiplicity of infection (MOI) and incubated at 37 °C overnight. No centrifugation was necessary. The next morning, cells received 1 mL fresh cultivation medium.

4.2.6 Reporter vector leakiness and Cre-mediated switch in 32D cells

In order to investigate leakiness and activation of Cre-dependent reporter vectors, cells were transduced in two rounds. First, the reporter vector was introduced, expressing a dTomato fluorescence reporter in the unmodified state and an mTagBFP2 fluorescence reporter after activation. If transduction rates reached >70%, no further selection or sorting was necessary. In a second step, a vectorized Cre-recombinase (pRRL.PPT.CbxSF.nlsCre_co.i2.GFP.pre.loxP, provided by Tobias Mätzig) was introduced into the cells. Cells were kept for five more days in maintenance and were then analyzed by flow cytometry on their ratio of dTomato and mTagBFP2 expression. Percentage of leakiness was determined by normalizing to the total transduction efficiency:

$$leakiness\ (\%) = \frac{\%(dTomato\ single\ positive) + \%(dTomato/mTagBFP2\ \ double\ positive)}{\%(dTomato\ single\ positive)}$$

Successful activation upon introduction of Cre recombinase was calculated by

$$activation\ (\%) = \frac{\%(mTagBFP2\ single\ positive)^{+Cre}\ +\ \%(dTomato/mTagBFP2\ double\ positive)^{+Cre}}{\%(dTomato)^{no\ Cre}}$$

4.2.7 Fluorescence reporter assay for CRISPR-Cas9 guide RNA efficacy

Based on the work of Labuhn *et al*., the cleavage activity and efficiency of sgRNAs were assessed in a fluorescent reporter assay[149]. The target sequence of each sgRNA (with PAM sequence) was cloned in-frame of a sfGFP, as described in section 4.1.2.6.2. 32D cells were first lentivirally transduced at different MOIs with pRRL.PPT.SFFV.target-region.sfGFP.IRES.Puro.pre, containing the reporter construct and the respective target sites for each sgRNA. Following selection was performed with 1 µg/mL Puromycin (Invivogen, San

Diego, CA, USA) on one approach with a transduction efficiency < 20%, avoiding multiple integrations. Next, lentiviral particles were generated from each sgRNA (pL40C.U6.sgRNA.EFS.Thy1.1.WPRE) and used to transduce selected 32D reporter cells. Double-transduced cells were kept in culture for 5 days, until they were harvested, stained against Thy1.1 (see section 4.2.11) and measured by flow cytometry at the CytoFlex S (Beckman Coulter). As positive control, one known sgRNA targeting the Tet sequence, as pre-integrated element of the target site cassette up-stream of the sfGFP, and a non-targeting sgRNA as negative control were used[155].

The sgRNAs guides the Cas9 protein to its target site to facilitate a DSB of the DNA. Subsequently, DNA repair occurs primarily via NHEJ, which leads to the introduction of insertion and deletions. This could lead to a frame-shift, which prevents the translation of down-stream sfGFP. Loss of sfGFP expression could consequently be used to determine the cleavage capability of particular sgRNAs, by using the mean fluorescence intensity (MFI) of the sfGFP in sgRNA-Thy1.1 positive cells (DP), sfGFP single positive cells (SP) and non-transduced cells (NT) in the following formulation:

$$KD\ (\%) = 1 - \frac{MFI^{sfGFP}(DP) - MFI^{sfGFP}(NT)}{MFI^{sfGFP}(SP) - MFI^{sfGFP}(NT)}$$

4.2.8 *In vitro* T cell differentiation on OP9-DL1 co-culture

The differentiation of hematopoietic stem cells towards T cells was achieved in a co-culture system with OP9 mouse stromal cells expressing the Notch ligand delta like 1 (DL1) (kindly provided by J. C. Zuñiga−Pflücker). The OP9-DL1 feeder cells were cultivated in αMEM complete medium (with FBS Brazil One) and passaged twice per week with 1x Trypsin/EDTA. Confluency of cells was checked regularly and kept below 70-80%. The day before replating T cell differentiation, OP9-DL1 cells were seeded to 12-well (5 x 10^4 cells) or 6-well (1 x 10^5 cells) in 1 mL or 2 mL αMEM complete medium, respectively[39,40].

Murine lineage-negative bone marrow (BM) cells (see section 4.3.3) where pre-conditioned 24h before transduction with StemSpan medium supplemented with the same cytokine conditions as for the differentiation – namely 20 ng/mL mSCF, 5 ng/mL hFLT3L, 1 ng/mL mIL7 (all Peprotech) and additionally 20 nM Meropenem (Inresa, Freiburg). On the day of transduction, lineage-negative BM cells were seeded in the same cultivation medium, supplemented with 1 mM Poloxamer and 4 µg/mL protamine sulfate and were transduced with LV supernatant containing a floxed *dTomato-mTagBFP2* fluorescence reporter cassette (SFFV.lox66.dTomato.reverse-mTagBFP2.lox71.PRE).

The day after transduction, lineage-negative BM cells were transferred to OP9-DL1 co-culture in αMEM for T cell differentiation (FBS Brazil One, 20 ng/mL mSCF, 5 ng/mL hFLT3-L, 1 ng/mL mIL7). On Tuesdays and Fridays, replating was performed. Cells were collected and filtered through a 40 µm cell strainer (Sarstedt, Nümbrecht, Germany) to remove the larger feeder cells. After washing the filter with 10 mL PBS, cells were centrifuged at 400 x g for 5 min and counted using the CASY Cell Counter (Schärfe System GmbH, Germany). For differentiation, ~ 4 x 10^5 differentiating cells were seeded in 2 mL αMEM for T cell differentiation to 12-well for the first replating time point and ~ 1.2 – 1.5 x 10^6 cells in 3.5 mL differentiation medium for the following time points onto the layer of OP9-DL1 cells. Remaining cells were used for flow cytometry to assess the stage of T cell development (see Table 19).

4.2.9 Immortalization assay

The immortalized bone marrow cell line 32D and the leukemic pro-B cell line BAF3 are both murine IL-3 dependent cell lines, which were transduced with various oncogenic signaling proteins in order to evaluate their transforming capability. Successful immortalization was assigned as cytokine-independent proliferation. For this purpose, both cell lines were transduced with either:

pL40C.SFFV.lox66.dTomato.reverse-Tlx1.PIK3CD^{E1021K}.WPRE,
pL40C.SFFV.lox66.dTomato.reverse-Tlx1.NRAS^{G12D}.lox71.WPRE,
pL40C.SFFV.lox66.dTomato.reverse-Tlx1.IL7R$^{c.731_732insTTGTCCCAC}$.lox71.WPRE
pL40C.SFFV.lox66.dTomato.reverse-Tal1.Lmo2.lox71.WPRE
pL40C.SFFV.lox66.dTomato.lox71.WPRE (as mock)

In a second step, cells were additionally transduced with a vectorized Cre-recombinase (pRRL.PPT.CbxSF.nlsCre_co.i2.GFP.PRE.loxP, vector provided by Tobias Mätzig). After 4 days of cultivation, transduced cells were washed three times with PBS to remove the murine IL3 (mIL3) from the cells. Rigorous vortexing was applied between each washing step to separate mIL3, which was still attached to the surface of the cells. Next, 500,000 cells were seeded in triplicate into 24-well plate with 500 µL RPMI complete. Every day, cells were resuspended and 50 µL were used in order to determine the cell number and viability. Afterwards, RPMI complete medium was filled up to 500 µL. In case of proliferation, 250 µL cell suspension were seeded into a new 24-well and filled with 250 µL RPMI complete. Dilutions were incorporated into later calculations.

4.2.10 Cultivation of primary and secondary murine leukemia cells

Leukemia cells, which infiltrated the spleen of diseased animals (see section 4.3.4), were used to generate murine T-ALL cell lines. Cells were isolated from the BM or spleen of primary or

secondary transplanted animals as described in chapter 4.3.4. Cells were either purified for Thy1.1-positive cells via magnetic-activated cell sorting (MACS) using the mouse and rat CD90.1 MicroBeads (Miltenyi Biotec, Bergisch Gladbach, Germany) beforehand or seeded directly as bulk culture onto pre-seeded OP9/OP9-DL1 cells. Except for the "Tlx1.NRAS^{G12D} + sgRNAs" group (see 4.3.4) that were grown on OP9-DL1 cells, OP9 cells were used as the main feeder layer. In brief, OP9/OP9-DL1 cells were seeded to 12-well adherent plate (Sarstedt) at least four hours before co-cultivation. Freshly isolated leukemia cells were seeded in different densities (0.5 x 10^6, 1 x 10^6 or 2 x 10^6) onto the pre-seeded OP9/OP9-DL1 cells. For cultivation, RPMI complete medium was added supplemented with either 5 ng/mL or 10 ng/mL mIL2 and mIL7 or without cytokines. For the first weeks, leukemia cells were splitted twice per week onto new pre-seeded OP9/OP9-DL1 cells in RPMI complete with either 5 ng/mL or 10 ng/mL mIL2 and mIL7, with filtering cells through a 70 µm cell strainer mesh (Sarstedt) before transfer. When leukemia cells started to proliferate more stable, cytokines were slowly diminished to zero for one fraction of cells, to test cytokine-independent proliferation. In addition, cells were filtered through 70 µm mesh, to remove OP9/OP9-DL1 cells and were seeded into suspension plates. For maintenance, one fraction of cells always remained in co-culture. Flow cytometry analysis was performed every few weeks to check the stability of the cell lines phenotype. Thereby, different marker panels were applied to define the T cell development stage and exclude expression of other blood lineage markers.

4.2.11 Surface staining and flow cytometry

All collected and washed cells were treated with a murine Fc-receptor blocking reagent (CD16/32 TruStain FcX, Biolegend) for 5 – 10 min in PBS at 4°C in order to reduce unspecific antibody binding. Surface staining was performed in FACS buffer for 20 – 25 min at 4°C in absence of light. Live/dead staining with 0.2 mg/mL 4′,6-Diamidin-2-phenylindol (DAPI; Sigma-Aldrich), propidium iodide (PI; Sigma Aldrich) or Zombie Aqua (Biolegend) excluded dead cells. Information on all used antibodies and staining reagents are listed below (Table 24), including dilution factors and fluorophores. Data were acquired at the CytoFlex S (Beckman Coulter) or FACS Calibur (BD) and analyzed using the software FlowJo 10 (Tree Star Incorporation, Ashland, OR, USA).

Cell phenotypes were partly summarized and presented as heatmap using R studio and the *ggplot2* package.

Table 24: Antibodies used for flow cytometry

Experiment	Antibody	Fluorophore	clone	Dilution	Company
sgRNA titration	Thy1.1 (CD90.1)	AF647	OX-7	1:600	Biolegend
Lineage negative BM isolation	Lineage cocktail	Biotin-PerCP	n.a.	1:1000	Caltag (Buckingham, UK)
	cKit	PE	2B8	1:300	Biolegend
	Sca1	AF700	D7	1:400	eBioscience
	CD150	APC	TC15-12F12.2	1:300	Biolegend
	CD48	PB	HM48-1	1:300	Biolegend
T cell differentiation on OP9-DL1 cells	CD4	BV605	RM4-5	1:300	Biolegend
	CD8	APC-Cy7	53-6.7	1:333	Biolegend
	CD25	APC	PC61.5	1:333	eBioscience
	CD44	PE-Cy7	IM7	1:2000	Biolegend
	CD11b	AF700	M1/70	1:1000	Biolegend
	Gr1	AF700	RB6-8C5	1:2000	Biolegend
Peripheral blood control: Final mouse analysis – panel 1	CD45.1	APC-eF780	A20	1:333	eBioscience
	CD45.2	AF700	104	1:333	Biolegend
	CD3	APC	145-2C11	1:333	eBioscience
	CD11b	eF450	M1/70	1:1000	eBioscience
	B220	Biotin	RA3.6B2	1:400	Biolegend
	Streptavidin	BV650	n.a.	1:1000	Biolegend
Final mouse analysis –panel 2	CD45.2	AF700	104	1:333	Biolegend
	Gr1	PE-Cy7	RB6-8C5	1:2000	Biolegend
	CD11b	eF450	M1/70	1:1000	eBioscience
	Thy1.1	AF647	OX-7	1:600	Biolegend
Final mouse analysis –panel 3	CD45.2	AF700	104	1:333	Biolegend
	CD4	BV605	RM4-5	1:300	Biolegend
	CD8	APC-Cy7	53-6.7	1:333	Biolegend
	CD3	APC	145-2C11	1:333	eBioscience
Final mouse analysis –panel 4 (thymus)	CD45.2	AF700	104	1:333	Biolegend
	CD4	BV605	RM4-5	1:300	Biolegend
	CD8	APC-Cy7	53-6.7	1:333	Biolegend
	CD25	APC	PC61.5	1:333	eBioscience
	CD44	PE-Cy7	IM7	1:2000	Biolegend
T-ALL mouse cell lines from primary cells – panel 1	CD3	APC	145-2C11	1:333	eBioscience
	CD11b	eF450	M1/70	1:1000	eBioscience
	B220	PE	RA3.6B2	1:400	eBioscience
T-ALL mouse cell lines from primary cells – panel 2	CD4	BV605	RM4-5	1:300	Biolegend
	CD8	APC-Cy7	53-6.7	1:333	Biolegend
	CD25	APC	PC61.5	1:333	eBioscience
	CD44	PE-Cy7	IM7	1:2000	Biolegend

4.2.12 Fluorescence-activated cell sorting (FACS) of various thymocytes populations

Obtained primary murine T-ALL samples arrested at different stages of T cell development, which is a hallmark of different T-ALL subsets[76,88]. To compare transcriptional changes within the developmental arrested leukemia, the different stages of T cell differentiation from healthy animals were used as reference. Therefore, thymus was removed from CD45.1$^+$ B6.SJL-Ptprca Pepcb/BoyJ mice (JAX stock #002014; referred as CD45.1 B6J strain) and scrunched on a 70 µM mesh using the back side of a syringe plunger. The mesh was washed with 5 mL MACS buffer and cells were centrifuged for 5 min at 300 x g, followed by an erythrocyte removal for 10 min in RBC lysis buffer. Lysis was stopped with PBS and centrifugation as before. Mature thymocytes were sorted by staining 30 x 10^6 cells in 750 µL FACS buffer with a T cell panel summarized in Table 25. Cell sorting was performed at the cell sorting core facility on the MHH campus by Matthias Ballmaier into 1.2 mL Trizol LS (Thermo Fisher Scientific) or 0.5 mL RLT lysis buffer (Qiagen) in 5 mL FACS tubes (Sarstedt). Not more than 0.4 – 0.5 x 10^6 cells were sorted, to minimize dilution of lysis buffer.

In order to sort immature double-negative thymocytes, a pre-enrichment was performed by depleting mature CD4 – and/or CD8 – positive thymocytes with MACS. Therefore, 1 x 10^8 cells were stained with biotinylated CD4 and CD8 antibodies (see Table 25) for 20 min on ice. Following washing with 1 mL MACS buffer per 10^7 cells, cells were secondary stained with 0.5 µL Streptavidin Nanobeads (BioLegend) in 100 µL FACS buffer per 10^7 cells. Purification was performed on equilibrated LS columns (Miltenyi) in a magnetic field. The flow through contained the double-negative (DN) thymocytes. Isolated cells were subsequently stained with the T cell panel from Table 25 and sorted at the sorting core facility on the MHH campus by Matthias Ballmaier with the FACSAria III Fusion (BD Biosciences, Franklin Lakes, New Jersey, USA) device into 1.2 mL Trizol LS (Thermo Fisher Scientific) or 0.5 mL RLT lysis buffer (Qiagen) in 5 mL FACS tubes (Sarstedt). Again, not more than 0.4 – 0.5 x 10^6 cells were sorted, to minimize dilution of lysis buffer.

Table 25: Antibodies used for cell sorting via FACS

Experiment	Antibody	Fluorophore	clone	Dilution	Company
T cell panel	CD4	BV605	RM4-5	1:300	Biolegend
	CD8	APC-Cy7	53-6.7	1:333	Biolegend
	CD25	APC	PC61.5	1:333	eBioscience
	CD44	PE-Cy7	IM7	1:1500	Biolegend
DN enrichment	CD4	Biotin	RM4-5	1:300	Biolegend
	CD8	Biotin	53-6.7	1:300	Biolegend

4.3 Mouse Work

All animal experiments were performed according to federal law and institutional guidelines under the control of the Lower Saxony State Office for Consumer Protection and Foot Safety (LAVES); approval number 33.19-42502-04-18/2906. Animals used in the following experiments were bred and maintained in a pathogen-free environment inside of ventilated cages (IVC) at the central animal facility of Hannover Medical School. All animals had unlimited access to food and water. Environmental settings were centrally monitored and set to 22°C ± 2°C, with a humidity of 55 % ± 5 % and a light cycle of 14h:10h (light:dark).

4.3.1 Generation of *LSL.Cas9 x Lck-cre* mouse strain

We generated the novel mouse strain Tg(Lck-cre)548Jxm; Gt(ROSA)26Sortm1(CAG-cas9+,-EGFP)FEZH/J (short *LSL.Cas9 x Lck-cre* mice), by breeding B6;129-Gt(ROSA)26Sortm1(CAG-cas9*,-EGFP)Fezh/J (short *LSL.Cas9* mice, JAX stock #024857) together with B6.Cg-Tg(Lck-cre)548Jxm/J (short *Lck-cre* mice, JAX stock #003802)[156]. At this point, generated animals were heterozygous for both loci (*LSL.Cas9* and *Lck-cre*). Descent mice were next bred with *LSL.Cas9* mice, until this locus became homozygous in the offspring. From this point, animals were bred within the same strain, to get a homozygous locus for the *Lck-cre* allele. Genotyping confirmed the presence of both alleles. However, as the official genotyping settings for the *Lck-cre* allele, provided by JacksonLab, only prove its presence but does not allow to distinguish between homozygous and heterozygous alleles, constant genotyping was performed on each animal, to confirm a correct genotype.

4.3.2 Genotyping

Tissue samples were collected from respective animals during ear marking. DNA was prepared with the Phire Tissue Direct PCR Master Mix (ThermoFisher) according to manufacturer's instruction. For PCR, the same cycle settings were used as described in section 4.1.3. The following primers were used for genotyping of mice:

Table 26: Genotyping primers for *LSL.Cas9 x Lck-cre* breeding

Gene locus	Primers 5' – 3'	PCR fragment size
Lck-cre	CTAGGCCACAGAATTGAAAGATCT GTAGGTGGAAATTCTAGCATCATCC	reference: 324 bp
	TGTGAACTTGGTGCTTGAGG CAGGTTCTTGCGAACCTCAT	insertion: 250 bp
LSL.Cas9	AGAATCCCTTCCCCCCTCTTC TTATGTAACGCGGAACTCCA	wild-type: 280 bp
	GGCGGATCACAAGCAATAAT	insertion: 180 bp

4.3.3 Isolation and cultivation of murine hematopoietic stem cells

Murine bone marrow cells were isolated by flushing ilium, femur and tibia from sacrificed mice with MACS buffer (Table 21) using a syringe. Next, cell suspension was filtered through a 70 µm mesh (Thermo Fisher Scientific) and centrifuged for 5 min at 300 x g. To remove erythrocytes, cells were incubated for 10 min in RBC lysis buffer (Table 22) and finally washed with PBS. Lineage-negative cells were purified using the MojoSort Mouse Hematopoietic Progenitor Isolation Kit (BioLegend) according to the manufacturer's protocol. Purification was performed on equilibrated LS columns (Miltenyi) with < 10^8 cells per column in a magnetic field. The flow-through contained the lineage-negative bone marrow cells, which were analyzed by flow cytometry for their purity (Table 24). Between 70 – 85% purity was reached. Depleted cells were either cultivated directly or frozen for later application in 90% FBS and 10% dimethyl sulfoxide (DMSO; Merck). Cultivation of lineage-negative BM cells was done in FBS-free StemSpan, supplemented with 50 ng/mL mSCF, 50 ng/mL mTPO, 50 ng/mL hFLT3-L, 1% Penicillin/Streptomycin, 1% L-glutamine and 0.1% Meropenem.

4.3.4 Co-transduction and transplantation of murine hematopoietic stem cells

For transduction, 700,000 lineage-negative BM cells were transferred to 12-well suspension plate with 700 µL cultivation medium, additionally supplemented with 1 mM Poloxamer (Sigma Aldrich) and 4 µg/mL protamine sulfate. To increase transduction efficiencies for each vector, two to three transduction rounds were performed with different fractions of the final co-transduction cocktail and a minimum of 4 h in between. For each new transduction round, all cells were washed with PBS, centrifuged for 5 min at 400 x g and reseeded in fresh transduction medium. The final transduction round was performed overnight and terminated by washing the cells twice and reseeding in 1 mL fresh cultivation medium. An overview of the vector pool composition for the different experimental main and control groups can be found in Table 27.

Table 27: List of all transplanted groups containing each target gene and the vector type

Group	Target gene	Vector type
Pool 1 - Tlx1.NRAS^{G12D} + sgRNAs	*Tlx1*	Inverted floxed oncogene
	NRAS^{G12D}	Inverted floxed oncogene
	Notch1	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Bcl11b	sgRNA
	Lef1	sgRNA
	Etv6	sgRNA
	Phf6	sgRNA

Table 27: List of all transplanted groups containing each target gene and the vector type (continued)

Group	Target gene	Vector type
Pool 2 - Tlx1.PIK3CD^{E1021K} + sgRNAs	*Tlx1*	Inverted floxed oncogene
	PIK3CD^{E1021K}	Inverted floxed oncogene
	Notch1	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Bcl11b	sgRNA
	Lef1	sgRNA
	Etv6	sgRNA
	Phf6	sgRNA
	Pten	sgRNA
	Dnm2	sgRNA
Pool 3 - Tlx1.IL7R$^{c.731ins732}$ + sgRNAs	*Tlx1*	Inverted floxed oncogene
	IL7R$^{c.731_732insTTGTCCCAC}$	Inverted floxed oncogene
	Notch1	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Bcl11b	sgRNA
	Lef1	sgRNA
	Etv6	sgRNA
	Phf6	sgRNA
	Ptpn2	sgRNA
	Dnm2	sgRNA
Pool 4 - Tal1.Lmo2 + PIK3CD^{E1021K}	*Tal1*	Inverted floxed oncogene
	Lmo2	Inverted floxed oncogene
	PIK3CD^{E1021K}	Inverted floxed oncogene
	Notch1	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Lef1	sgRNA
	Usp7	sgRNA
	Pten	sgRNA
Tlx1.NRAS (Pool 1) control group A	*Tlx1*	Inverted floxed oncogene
	NRAS^{G12D}	Inverted floxed oncogene
Tlx1.NRAS (Pool 1) control group B	*Notch1*	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Bcl11b	sgRNA
	Lef1	sgRNA
	Etv6	sgRNA
	Phf6	sgRNA
Tlx1.PIK3CD (Pool 2) control group A	*Tlx1*	Inverted floxed oncogene
	PIK3CD^{E1021K}	Inverted floxed oncogene

Table 27: List of all transplanted groups containing each target gene and the vector type (continued)

Group	Target gene	Vector type
Tlx1.Pik3cd (Pool 2) control group B	*Notch1*	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Bcl11b	sgRNA
	Lef1	sgRNA
	Etv6	sgRNA
	Phf6	sgRNA
	Pten	sgRNA
	Dnm2	sgRNA
Tal1.Lmo2 (Pool 4) control group A	*Tal1*	Inverted floxed oncogene
	Lmo2	Inverted floxed oncogene
Tal1.Lmo2 (Pool 4) control group B	*Notch1*	Dual sgRNA
	Cdkn2a	Dual sgRNA
	Lef1	sgRNA
	Usp7	sgRNA
	Pten	sgRNA
reporter	*mTagBFP2*	Inverted floxed oncogene

Two days after transduction, cells were harvested for transplantation. At least 4 washing steps with PBS and centrifugation at 400 x g for 5 min were performed before cell number was determined and concentration was set up to 500,000 cells/100 µL PBS. Remaining cells were used for transduction efficiency control via flow cytometry, by staining cells with anti-Thy1.1-AF647.

One day before transplantation, CD45.1$^+$ B6.SJL-Ptprca Pepcb/BoyJ recipient mice (JAX stock #002014; referred as CD45.1 B6J strain) were sub-lethally irradiated with a single dose of 7.5 Gy using the BioBeam200 irradiator (Eckert & Ziegler, Berlin, Germany). From the same day on, mice received the broad-spectrum antibiotic Ciprofloxacin (Fresenius Kabi, Bad Homburg von der Höhe, Germany) via their drinking water for three weeks. The co-transduced lineage-negative BM cells were injected intravenously (100 µL per mouse). Donor cell chimerism and blood type composition was analyzed via flow cytometry by retroorbital sampling every 4 weeks until reaching an end-point (first measurement after 5 week). In addition, blood parameters were monitored with the scil VET animal blood counter (scil animal care company, Viernheim, Germany). Animals were visited and scored daily after transduction and later at least three times per week. When showing signs of sickness, frequency of visits was increased and as soon as mice reached the approved end-points, animals were removed from the experiment and euthanized. Spleen, BM, thymus and blood were removed and single cell suspension was generated by scrunching spleen and thymus on a 70 µM mesh or flushing ilium, femur and tibia for BM. Following erythrocyte removal by 10 min incubation in RBC lysis

buffer and washing, cells were cryopreserved in freezing medium. At the day of sacrifice, blood parameters were acquired with the scil VET animal blood counter and a blood smear was prepared on a glass slide (Thermo Fisher Scientific). For further analysis, frozen tissue samples were thawed and analyzed by flow cytometry for blood lineage distribution and donor cell engraftment with four different panels, summarized in Table 24. For a detailed analysis and cell cultivation, MACS enrichment of Thy1.1-positive leukemia cells was performed mainly from spleen or BM. Cells were used for preparation of DNA, RNA and protein lysates and were partly seeded to co-culture with OP9 or OP9-DL1 cells, as described in chapter 4.2.10. Visualization of the animals' residence time in the experiment was performed with R Studio (R version 4.1.0) using the *survival* and *survminer* packages.

4.3.5 Secondary transplantation of primary leukemia cells

Recipient CD45.1 B6J mice were pre-conditioned the same way as described in 4.3.4, including irradiation with 7.5 Gy and treatment with Ciprofloxacin via their drinking water starting one day before transplantation and continued for three weeks. In total, leukemia cells from one primary transplanted mouse were transplanted into three secondary recipients. Therefore, leukemia cells were freshly isolated from frozen BM or spleen cells from the primary animal via MACS for Thy1.1$^+$ cells. In total, 500,000 leukemia cells and 500,000 lineage-negative BM cells as helper cells from *LSL.Cas9 x Lck-cre* mice were pooled together in 100 µL PBS for each mouse to be transplanted. Cell mix was injected intravenously into pre-conditioned mice. Animals were visited and scored daily after transduction and later at least three times per week. When showing signs of sickness, frequency of visits was increased and as soon as mice reached the approved end-points, animals were removed from the experiment and euthanized. Final analysis and organ preparation of sickened animal, cryopreservation and later sample processing were performed the same way as described in 4.3.4. However, flow cytometry analysis was simplified to Table 24, panel 1 and 4.

4.3.6 Pappenheim staining of blood samples and cytospins

Blood smears of sacrificed mice and cytospins of obtained murine T-ALL tissues and cell lines were stained with a standardized Pappenheim staining protocol. Cytospins were generated either from cell lines or from isolated BM and spleen suspension cells by spinning 7.5 x 10^4 cells in 200 µL PBS for 10 min at 800 x g onto glass slides using Cytospin 4 centrifuge (Thermo Fisher Scientific). For staining, glass slides were transferred to a staining chamber and incubated for 5 min in May-Grünwald solution (Carl Roth GmbH). Next, slides were thoroughly washed twice in dH_2O and subsequently stained for another 20 min in diluted Giemsa solution (1:20 in dH_2O; Sigma Aldrich). After two more washings in dH_2O, glass slides were dried overnight and mounted with a coverslip utilizing one drop of Roti® Histokitt (Carl Roth GmbH)

the next day. The prepared slides were scanned by Dirk Schlaudien (Fraunhofer Institute of Toxicology and Experimental Medicine, Hannover, Germany). Digital pictures were analyzed and prepared with NDP.view2 viewing software (Hamamatsu Photonics, Japan).

4.4 Bioinformatics

4.4.1 Tracking of Indels by Decomposition (TIDE)

In order to spot insertions and deletions, mediated by CRISPR-Cas9 induced double-strand breaks, the genomic target region was amplified by PCR (see 4.1.3) and SANGER-sequenced (see 4.1.2.5). Sequencing results including the chromatograms (.ab1 data format) of treated and untreated (= wild-type locus) samples were uploaded at http://shinyapps.datacurators.nl/tide/ together with the used guide sequence[157]. As results, frequencies of small insertions and deletions were generated. Table 28 summarizes the used primers to amplify the gene locus, where the sgRNAs bound.

Table 28: Primer pairs for amplification of CRISPR-Cas9 modified genetic regions

Primer name	Sequence 5' – 3'	PCR size in bp
Cdkn2a_locus_FW	CACACAATCCCAGTTCGGCT	692
Cdkn2a_locus_Rev	TGTTCTCAGGCTCATTTGGGT	
Bcl11b_locus_FW	CTCGGGGCTATGATGGTGAAC	524
Bcl11b_locus_Rev	CTGCTTGGGACAGATGCCTT	
Phf6_locus_FW	CTCCTGACCACAGTAGACCA	435
Phf6_locus_Rev	TGGAAAAGCCTGAAGGACAG	
Dnm2_locus_FW	ATAGCAGTAGGGTTGCGGAC	616
Dnm2_locus_Rev	CCAAACACCGCCCACTATCA	
Etv6_locus_FW	AGCCTCAGACCTATCGTCCC	902
Etv6_locus_Rev	CCTGAATGCTATGCGTGCAA	
Pten_locus_FW	AAGTTGCCTCTATGCAGTTCAC	784
Pten_locus_Rev	AGCCAACAAACTAAGGGTCGG	
Usp7_locus_FW	CTCTTCCCTCCTCAGTCCCA	684
Usp7_locus_Rev	TCCATCCCCAGAGCCAAAAC	
Lef1_locus_FW	GACGGAGGCCTGTACAACAA	388
Lef1_locus_Rev	CCTCCGACTTGGCTCCATTT	
Notch1_locus_FW	CATAGCATGATGGGGCCACT	539
Notch1_locus_Rev	GAGATGCCCTCGGACCAATC	

4.4.2 Next Generation Sequencing (NGS)

Barcoded PCR products, comprising each CRISPR-Cas9 targeted gene locus from each CD45.2^{+} T-ALL within the experimental main groups, were pooled at equal ratios. Targeted sequencing was conducted at the Research Core Unit Genomics (RCUG) of Hannover Medical School on a MiSeq Reagent Micro Kit v2 (300 cycles) from Illumina. The sequences were

bioinformatically exploited using Galaxy (https://usegalaxy.org/ , version 23.0.rc1). First, reads were demultiplexed using the "Barcode Splitter"[158] tool and trimmed from PCR primer sequences using the "Trim Sequences" tool[158]. Next, reads were mapped with "Map with BWA-MEM"[159] against the *mus musculus* genome assembly GRCm38 (mm10) reference genome. Using the "FreeBayes" tool[160], genetic variations were detected and annotated with the "SnpEff eff" tool[161]. Genetic aberrations and mapped sequences were visualized and verified with "IGV" software (https://software.broadinstitute.org/software/igv/home, Broad Institute, University of California).

4.4.3 RNA Sequencing

Purified RNA from different thymocytes subpopulations, primary and secondary transplanted leukemia and generated T-ALL cell lines was collected and processed as described in chapter 4.1.5. Library construction, RNA sequencing, quality control and mapping to *mus musculus* (GRCm39/mm39) genome was performed at Novogene Co., Ltd (UK). The extracted and mapped raw counts were analyzed using R studio (R version 4.1.0), following an adapted RNA-analysis script to Charity W. Law *et al.*[162] In brief, the raw counts with > 1 "counts per million" (CPM) were $\log_2$-transformed using the *edgeR* package. Next, counts were normalized on the library size by applying the method of "trimmed mean of M-values" (TMM) within *edgeR*. Principal component analysis (PCA) and heatmap visualization was conducted on processed gene counts filtered for an interquartile range (IQR) > 2. Results were visualized as two-dimensional PCA, using the "prcomp" function within the *stats* package and "ggplot" or "ggplot2" from the *ggplot2* package.

Differentially expressed genes were calculated using the *limma* package and its "voom" method on DGEList objects. For single-sample gene set enrichment analysis (ssGSEA), gene sets were extracted from the different thymocytes populations. Therefore, dysregulated genes were calculated for each thymocytes subset and selected for features that were significant in all comparisons using the "reduce" function. With the obtained gene lists, ssGSEA was conducted using the R package *GSVA* and the "ssgsea" method. As another analysis tool, gene set enrichment analysis (GSEA) was applied for the experimental groups, including, for example, the thymus subsets, leukemia from each transplanted condition, and cell types (primary cells versus cell lines). Deregulated genes (log fold change > 1) were assigned to functional KEGG pathways (https://www.genome.jp/kegg/pathway.html) or geneontology on biological processes using the web-based enrichment analysis platform WEBGestalt (https://www.webgestalt.org/)[163].

For the final application of the R script and support in data processing, I received help by Dr. Adrian Schwarzer, PhD and Dr. Anton Selich. The gene count table as provided by Novogene can be found on an attached hard drive, together with a target file and defined gene lists.

4.4.4 Statistics

All graphs represent the mean ± standard deviation (SD). Statistical significance was determined using non-parametric t-test (Mann-Whitney U test), one-way or two-way ANOVA with GraphPad Prism 5 (GraphPad Prism Incorporation, San Diego, CA, USA). *P≤0.05, **P≤0.01, ***P≤0.001

4.4.5 Preparation of Figures

All figures were designed and prepared using the software Inkscape (https://inkscape.org) and other sources used indicated as mentioned.

4.4.6 Grammatical inspection

The written text from this thesis was checked with the ‘Grammarly’ software (San Francisco, USA) for grammar mistakes.

5 Results

In this work, to target the onset of oncogenic events required for modeling the genetic complexity of T-ALL to developing T cells in the thymus, we developed the *LSL.Cas9 x Lck-cre* (LSL – loxP-stop-loxP) mouse strain (see chapter 4.3.1). In these animals, *Cre* is expressed under the control of the lymphocyte protein tyrosine kinase (*Lck*) promoter and thus specifically during T cell development[164]. Upon its activation, the Cre recombinase removes a stop-cassette, which is flanked by two *loxP* sites, and induces the transcription of *Cas9* and *EGFP* (Figure 6 A). Based on Liu *et al.*, who presented extensive high-throughput sequencing data from T-ALL patients, specific subgroups were formed and modeled as distinct subtypes[50]. The overall concept of this thesis was to combine subtype-specific overexpression of oncogenic transcription factors and mutant gain-of-function signaling proteins along with various sgRNAs targeting subtype-related genes, into the HSPCs of the novel mouse strain, to model clinically relevant T-ALL *in vivo*. For T cell-specific expression of the introduced oncogenes, their expression was linked to the *Lck-cre* expression by adding loxP sites that prevented lineage-nonspecific expression. After thoroughly characterizing the oncogenic effect of these vectors, genetically altered cells were transplanted into sub-lethally irradiated recipient mice. With the reconstitution of the hematopoietic system, a subset of modified HSPCs underwent thymus-driven T cell development, which led to the activation of the introduced pool of genetic aberrations und thus provoked transformation into T-ALL. In following experiments, gained leukemias from different genetic approaches were characterized by flow cytometry for their immune phenotype and transplanted into secondary recipient mice to prove a stem-cell like character. Through sequencing of the DNA and RNA, the subtypes were further described and the heterogeneity linked to fundamental genetic and transcriptomic changes. In a last step, cell lines were generated from the leukemia and compared to their leukemia of origin.

5.1 Generation and characterization of *LSL.Cas9 x Lck-cre* mice

Many different cancer types share sets of commonly mutated and dysregulated genes (reviewed in [165,166]). The uncontrolled overexpression or knockout of these genes can be lethal in a systemic setting[167] or, in the context of leukemia, provoke the development of different mixed cancer types[53,95,168]. Therefore, T cell-specific expression of oncogenes and knockout of tumor suppressor genes was the first challenge to model different subsets of T-ALL. *Lck-cre* mice, expressing a Cre recombinase in thymocytes, were bred with *LSL.Cas9.EGFP* mice, harboring a loxP-stop-loxP cassette in front of a Cas9 and an EGFP fluorescence reporter (Figure 6 A). The newly generated *LSL.Cas9 x Lck-cre* mouse strain should therefore express

the Cre recombinase and Cas9 protein only in thymocytes and not in any other organ or blood type. By this, the tissue-specific introduction of mutations and oncogenes would prevent the generation of other leukemia types besides T-ALL[53]. Of note, the *LSL.Cas9 x Lck-cre* mice were bred to contain homozygous *LSL.Cas9* alleles, which was checked by genotyping. Since the "Jackson Laboratory" genotyping protocol for the *Lck-cre* locus could not distinguish between homozygous and heterozygous alleles, no different genotypes were pointed out in this work (see chapters 4.3.1 and 4.3.2).

The thymus, spleen, liver, and bone marrow (BM) from mice were first analyzed by Western Blot analysis on the expression of Cre recombinase and Cas9. For comparison and to prove the induction of Cas9 only in the presence of Cre, the parental mouse strains *LSL.Cas9* mice and *Lck-cre* mice were checked comparably. In *LSL.Cas9 x Lck-cre* mice, both proteins were strongly expressed within the thymus and spleen (Figure 6 B). The expression in the spleen was expected, as mature T cells, migrate into the spleen[169]. Neither liver nor BM cells showed Cre or Cas9 expression, indicating the tight and specific regulation by the *Lck* promoter. The same expression pattern for Cre was observed from *Lck-cre* mice, which expressed Cre inside the thymus and spleen at similar levels. For *LSL.Cas9* mice, no expression of Cas9 protein was observed in any organ.

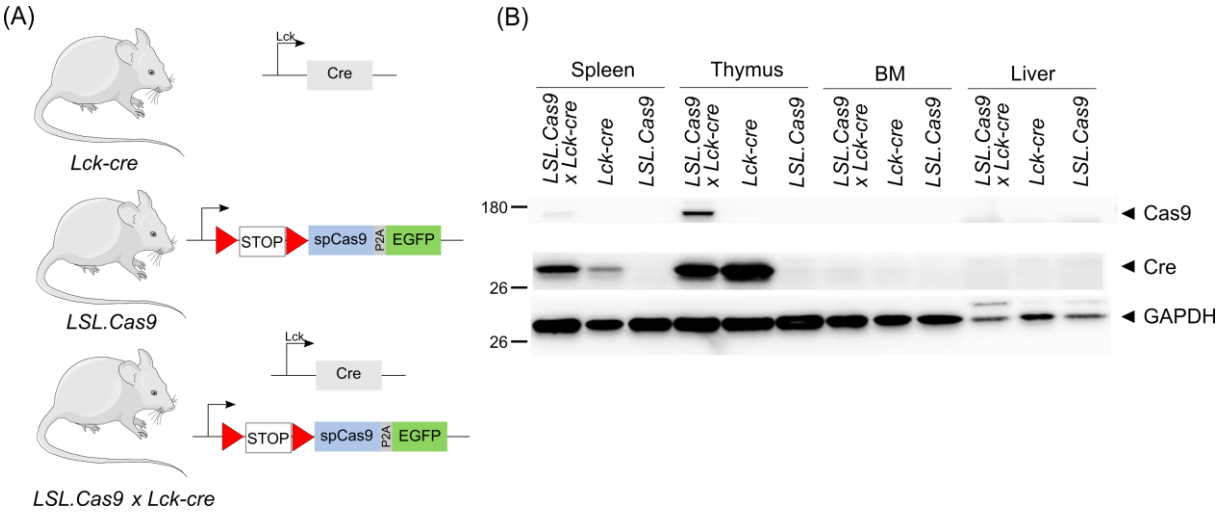


Figure 6: Organ-specific expression of Cre and Cas9 in *LSL.Cas9 x Lck-cre*. (A) Mice expressing Cre recombinase under the control of the *Lck* promoter were bred with mice harboring *Streptococcus pyogenes* Cas9 (spCas9) with an upstream floxed (red triangles) stop-cassette to generate the *LSL.Cas9 x Lck-cre* mouse strain. (B) Spleen, thymus, bone marrow (BM) and liver were analyzed by Western Blot analysis on the presence of Cas9 and Cre protein in the novel *LSL.Cas9 x Lck-cre* mouse strain and its parental strains.

In the next approach, different blood cell types were analyzed for the presence of Cas9 protein. After removal of the floxed stop-cassette, the enhanced green fluorescent protein (EGFP) fluorescence reporter linked to the *spCas9* gene via a P2A site could be measured by flow cytometry. Again, results were compared to the parental *LSL.Cas9* mice to underline the strict

requirement of Cre presence for expression of Cas9. Blood cells were stained for the B cell marker B220, the myeloid cell surface marker CD11b, the granulocyte cell surface marker Gr1, and the T cell marker CD3 (Figure 7 A). Indeed, no expression of EGFP was measured in myeloid or B cells from *LSL.Cas9 x Lck-cre* mice (Figure 7 B). Only in $CD3^+$ cells high expression of EGFP was detected in 80.9% ± 4.7% cells. Interestingly, a minor population of $Gr1^+$ granulocytes also expressed EGFP. However, only 14.5% ± 5.2% of the cells were affected. In *LSL.Cas9* mice, no such expression was seen in any blood cell type (Figure 7 C).

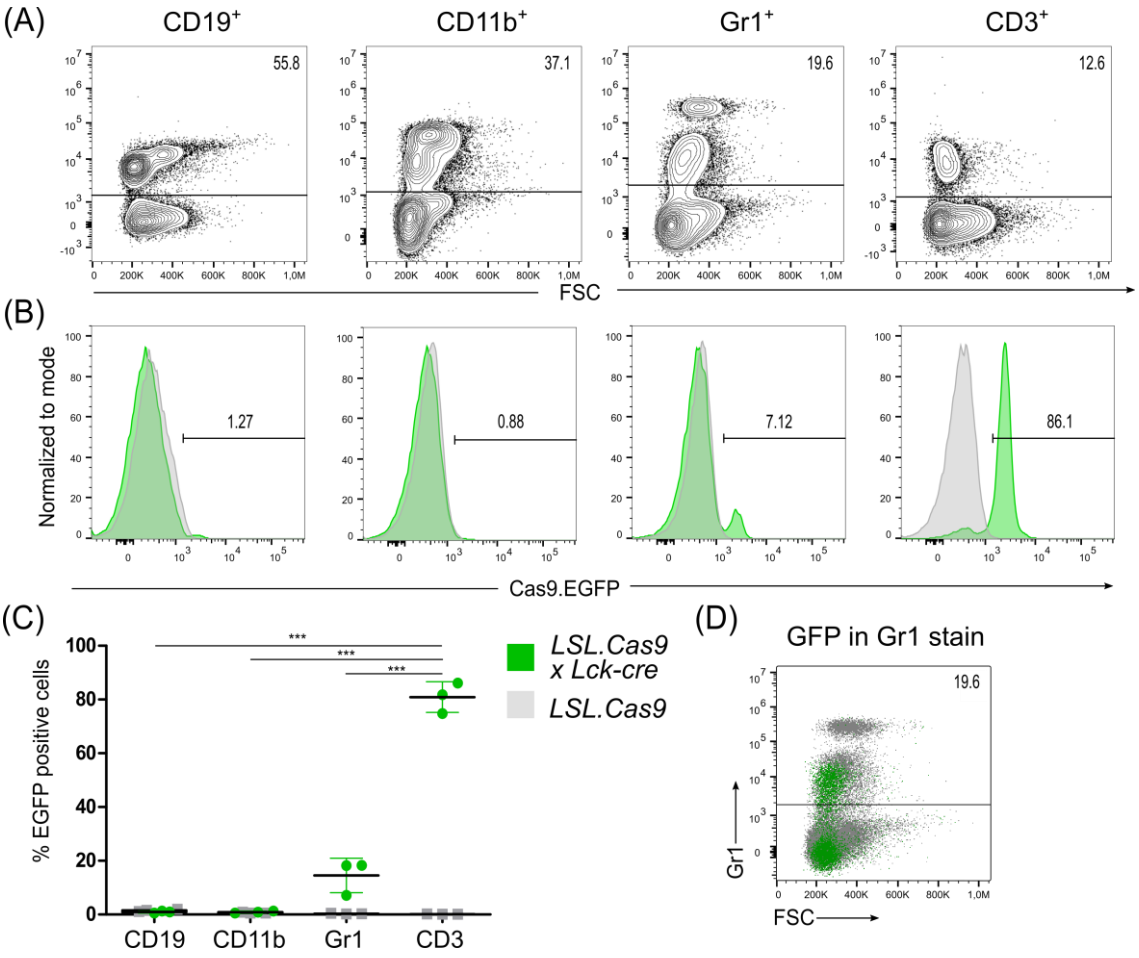


Figure 7: Cas9.EGFP expression within different blood cell types. Blood cells from *LSL.Cas9 x Lck-cre* and *LSL.Cas9* mice were analyzed for the presence of Cas9.EGFP in B cells (CD19), myeloid cells (CD11b), granulocytes (Gr1) and T cells (CD3). (A) Representative gating strategy on the different hematopoietic populations. (B) EGFP expression in the described blood compartments as representative histograms and (C) summarized for n=3 animals. Mean ± SD is shown with ***p≤0.001 conducting one-way ANOVA on *LSL.Cas9xLck-cre* samples together with Tukey's multiple comparison post hoc test. (D) Fraction of $EGFP^+$ population within Gr1-expressing cells.

Having confirmed expression of Cas9 in thymocytes, we next investigated at which stage of T cell development its expression started. As various T-ALL subtypes, including ETPs or early cortical immature types, stop their development at an early stage of T cell development[68], early activation of the Cre/Cas9 system would be necessary to generate these premature types of T-ALL. Therefore, the thymus of *LSL.Cas9 x Lck-cre* was extracted, and thymocytes were

isolated. Staining the immature T cell markers CD25 and CD44 and the mature T cell markers CD4 and CD8 allowed distinguishing different developmental stages of T cell maturation (Figure 8 A). The different populations existed in comparable proportions in the thymus of *LSL.Cas9 x Lck-cre* and *LSL.Cas9* mice (Figure 8 B). The expression of EGFP implicated the presence of Cas9 and, thereby, the previous activation of Cre. All results were compared to isolated thymocytes from parental *LSL.Cas9* mice to prove tightness and the necessity of Cre expression to activate Cas9. At the earliest stage of T cell development, CD4 and CD8 double-negative $CD25^{-}$ $CD44^{+}$ (DN1) cells already expressed EGFP, albeit at a low levels (Figure 8 C + D). As the T cell development progressed, EGFP expression decreased in $CD25^{+}$ $CD44^{+}$ (DN2), then increased again in $CD25^{+}$ $CD44^{-}$ (DN3) cells and reached its highest peak in $CD25^{-}$ $CD44^{-}$ (DN4) cells. Although the total median fluorescent intensity (MFI) of EGFP expression decreased in the mature stages of T cell development compared to the DN4 stage (Figure 8 D), nearly all cells were positive for EGFP with no discernible differences between $CD4^{+}$ and $CD8^{+}$ DP or $CD4^{+}$ or $CD8^{+}$ SP cells (Figure 8 C). In *LSL.Cas9* mice, no EGFP expression was observed in any developmental stage (Figure 8 D). These data demonstrate

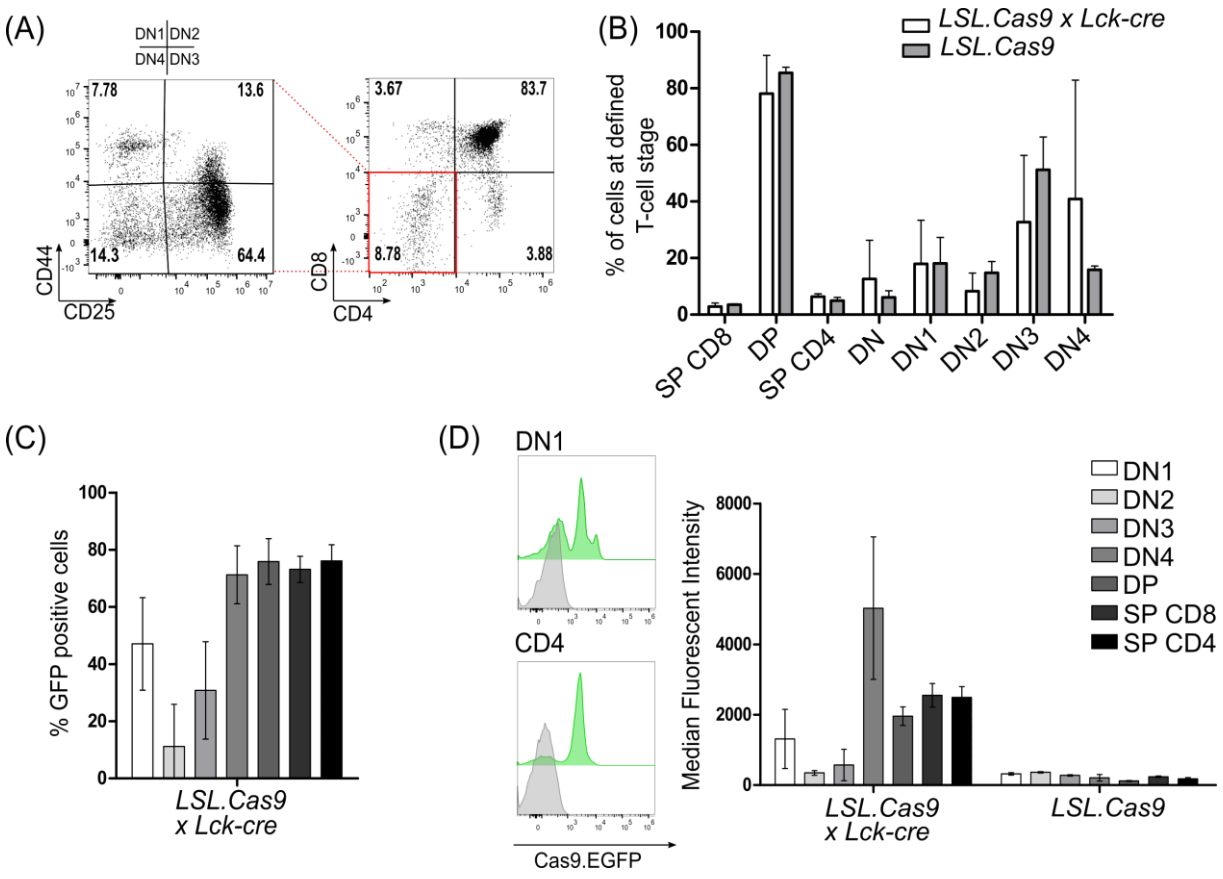


Figure 8: Cas9.EGFP expression inside the thymus of *LSL.Cas9 x Lck-cre* mice. The different stages of T cell development were analyzed for their Cas9.EGFP expression. (A) The immature CD4 and CD8 double-negative (DN) stages 1-4 can be distinguished by their up- and downregulation of CD44 and CD25. The mature thymocytes first become double-positive (DP) and then single-positive (SP) for

CD4 and CD8. (B) Percentages of the different T cell populations within the thymus between the *LSL.Cas9 x Lck*-cre and the *LSL.Cas9* mouse strain; n = 3. (C) Percentages of cells expressing Cas9.EGFP during the different stages of T cell development presented as mean ± SD. (D) Median fluorescence intensity of EGFP changes during differentiation exemplary shown for DN1 and CD4 stages. Total MFI is presented for each thymic stage as mean ± SD in *LSL.Cas9 x Lck-cre* mice and the parental strain *LSL.Cas9*.

that Cre and Cas9 are expressed early in T cell development with increasing activity in more mature stages. Thus, the system should enable to introduce mutations and oncogenes at all stages of T cell development, thereby allowing to mimic a broad spectrum of T-ALL subtypes, including early thymocyte progenitor (ETP)-ALL. In summary, we created a novel *LSL.Cas9 x Lck-cre* mouse strain that specifically expresses Cre and Cas9 only in developing thymocytes and mature T-cells allowing *in vivo* gene editing and conditional gene expression exclusively in these populations.

5.2 Generation of a non-leaky Cre-dependent gene expression system for multiplexing gain-of-function lesions

Next, it was aimed to develop a modular lentiviral vector system to genetically modify thymocytes *in vivo* in order to alter their developmental path towards malignant progression and, finally, T-ALL. Although we proved that Cre and Cas9 are only expressed during T cell development, the genetic intervention will already occur at a stem-cell stage when transducing BM cells if conventional gene expression vector systems are used. Hence, Cre or Cas9-independent expression of any introduced gene or sgRNA must be eliminated to prevent the generation of other leukemia types. Fortunately, sgRNAs possesses no intrinsic activity[170], and the tight T cell-specific expression of Cas9 was already proven in chapter 5.1. Therefore, in the absence of Cas9, no activity is expected from these vectors. For the overexpression of oncogenes a Cre-based approach was used, which is only expressed in the thymus of our mouse strain. A lentiviral approach was chosen because it allowed transducing low-proliferating HSCs[111]. To prevent expression in the absence of Cre, two different vector designs were tested. In the first design, a stop-cassette containing stop-codons in all three reading frames was cloned behind a *dTomato* fluorescence reporter gene, flanked by two *loxP* sites, and driven by the SFFV promotor. Right behind this cassette, an *mTagBFP2* fluorescence reporter was placed, which will only be expressed if *dTomato*, including the stop-cassette, is removed by Cre recombinase (named “stop construct”). Although long-term silencing effects are known for the SFFV promoter[171], it was chosen for its strong expression and high titers during virus production[172]. In a second design, *mTagBFP2* was cloned in antisense as inverted sequence 3’ of the *dTomato* gene. Both genes were flanked by two mutant *loxP* sites (*lox66* and *lox71*), which will lead to a Cre-mediated inversion of the whole construct (in the following named as “inverse construct”) (Figure 9 A). However, if the cassette got inverted, the mutated *loxP* sites were altered and prevented further inversion[128]. Both constructs were introduced

into the myeloid 32D suspension cell line with a Cre recombinase vector. In the absence of Cre, the stop construct showed a substantially more of 1.31% ± 0.09%, compared to 0.025 ± 0.004 % for the inverse construct (Figure 9 B). The leakiness remained low for the inverse construct even at high transduction rates of > 90 %, implicating multiple insertion sites within the transduced cells[173]. Upon introduction of Cre, both constructs were equally activated, indicated by the loss of *dTomato* expression and induction of *mTagBFP2* expression (Figure 9 C). In summary, the inverse construct was superior to the stop construct due to the substantially lower leakiness and was therefore used as vector design for the Cre-inducible expression of oncogenes.

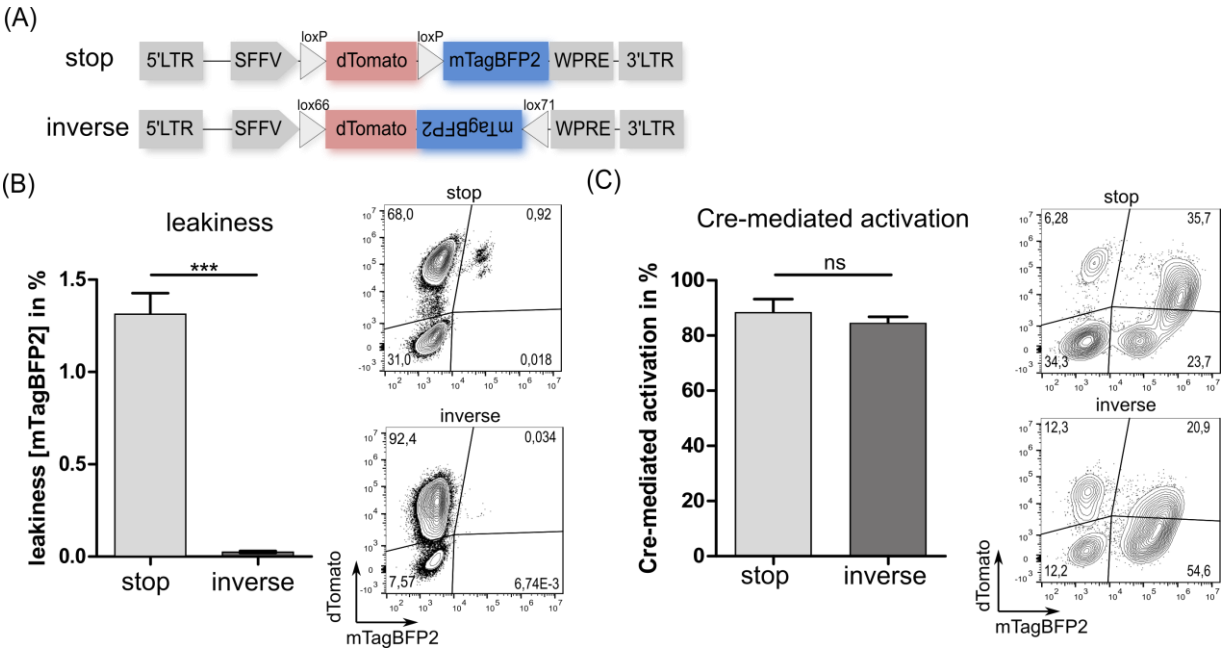


Figure 9: Leakiness and tightness of two different vector systems. (A) Lentiviral vector design for the "stop construct" vector system, containing a strong stop signal to prevent expression of the fluorescence reporter mTagBFP2, and the "inverse construct" vector system, containing the mTagBFP2 sequence in antisense orientation flanked by two mutated *loxP* sites (*lox66/71*). (B) Leakiness of both vector systems measured by unwanted expression of mTagBFP2$^+$ cells in transduced 32D cells. Presented are mean ± SD, calculated from the ratio of mTagBFP2$^+$ cells to the overall transduction rate. (C) Successful induction of mTagBFP2 expression in the presence of Cre, presented as mean ± SD from n=3 technical replicates. For calculation see chapter 4.2.6. Statistical analysis was performed by two-tailed t-test. ***p≤0.001, ns = not significant.

Prior to the final application of the inverse vector system, it was intended to additionally check its switching potential in developing thymocytes. In this attempt, lineage-negative BM cells from the novel *LSL.Cas9 x Lck-cre* mouse strain were differentiated into fully mature T cells. By this, the efficiency of the vector design and intact differentiation potential of Cre- and Cas9-expressing HSPCs was proven *ex vivo*. For differentiation, an established co-culture differentiation protocol on OP9-DL1 cells was used[39,40]. Thawed and pre-conditioned BM cells were initially transduced with the inverse reporter system (Figure 10 A). In the presence of the cytokines mSCF, hFLT3-L and mIL-7 (at 20 ng/mL, 5 ng/mL, and 1 ng/mL, respectively), BM

cells differentiated through the different stages of thymocyte development. Developing cells were passaged every three to four days onto freshly seeded OP9-DL1 cells[40] (Koniaeva 2020).

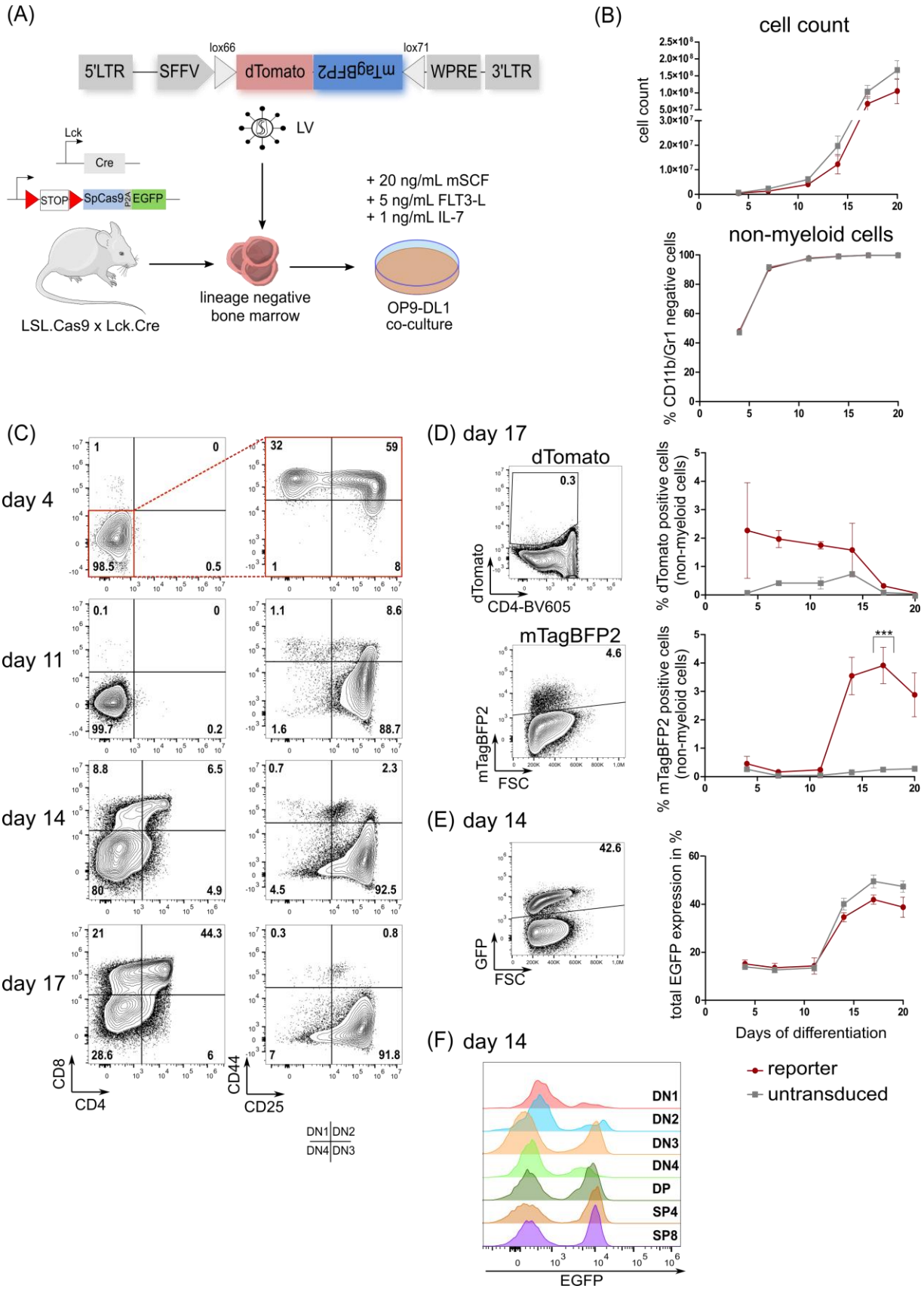


Figure 10: *In vitro* T cell differentiation proves inversion and activation of the inverse vector system. (A) Lineage-negative BM cells from *LSL.Cas9 x Lck-cre* mice were lentivirally transduced with the inverse reporter construct, harboring the fluorescence reporter proteins dTomato and mTagBFP2. Transduced cells were seeded onto OP9-DL1 cells for co-cultivation and differentiation in the presence of mSCF, FLT3-L and mIL7 towards T cells. (B) Total cell count and percentage of non-myeloid ($CD11b^-$ and $Gr1^-$) cells after start of differentiation for transduced and untransduced conditions. (C) Flow cytometry classification of the different stages of T cell development at different time points of differentiation. CD25 and CD44 were gated only on the CD4 and CD8 double-negative population. (D) Reporter switch, marked by the downregulation of the dTomato fluorescence protein and induction of mTagBFP2 expression. (E) Cas9.EGFP expression at different time points of differentiation and (F) as representative histogram in the different stages of T cell development on day 14 of differentiation. Presented are mean ± SD from n=3. Mann-Whitney U test (*** $p< 0.001$). SFFV – spleen focus forming virus promoter, LTR – long terminal repeat, Lck - Leukocyte C-Terminal Src Kinase promoter, WPRE - woodchuck hepatitis virus posttranscriptional regulatory element, spCas9, *Streptococcus pyogenes*-derived Cas9, DN – double-negative, SP – single-positive, DP – double-positive.

First, it was checked if the transduction affected the differentiation. Therefore, the cell count and the absence of myeloid cells throughout the differentiation were compared between transduced and untransduced differentiating cells. Besides a minor decrease in the total number of cells, no differences were observed between both conditions (Figure 10 B). Next, the stage of T cell differentiation was assessed at which the inversion of the *dTomato-mTagBFP2* cassettes was mediated. On day 4, developing T cells reached the DN2 stages ($CD25^+$ $CD44^+$), progressed to DN3 stage on day 11, and most cells reached mature DP and single-positive CD4 (SP4) and single-positive CD8 (SP8) stages on day 17 (Figure 10 C). At this stage, a bias towards SP8 and fewer SP4 cells were observed for transduced and untransduced cells (Figure 10 C – day 17). The transduction rate with the dual reporter vector was relatively low, ranging from 1.1% to 3.5% on day 4. Still, the switch in the expression of both fluorescence reporters was observed at day 14 of differentiation, indicated by the decline of $dTomato^+$ cells and the increase of $mTagBFP2^+$ cells (Figure 10 D). This switch aligned with the increase in EGFP expression (Figure 10 E) in developing T cells from day 14 of differentiation. Until day 20, the remaining expression of dTomato went to undetectable levels, while mTagBFP2 peaked at 3.55% ± 0.54% on day 17 before it decreased slightly again. EGFP expression also peaked on day 17 and slightly diminished on day 20 again (Figure 10 D). Surprisingly, only 41.9% of transduced and 49.5% of untransduced cells started to express EGFP, which is lower compared to the *in vivo* situation observed (Figure 8 B). The EGFP induction peaked in the DN3 stage (Figure 10 F), which was earlier than in *in vivo*. Nevertheless, the complete inversion of the expression cassette during T cell development supported an efficient vector system suitable for *in vivo* use. In summary, cloning the oncogenes of interest in antisense flanked by *lox66* and *lox77* sites combines non-leakiness and effective onset of gene expression by inversion upon Cre exposure.

5.3 Screening of single guide RNAs for their cleavage efficacy

Besides introducing specific overexpressed genes, T-ALL subtype-related mutations will be set to disrupt gene function of, e.g. tumor suppressors or epigenetic regulators. While CRISPR

screens use multiple sgRNAs simultaneously to target the same gene and ensure efficient cleavage[142], only one highly efficient sgRNA for each candidate gene was used in this work to avoid further complicating the co-transduction protocol. To further increase likelihood of a loss of gene function, gene regions were targeted that encode critical functional protein domains or were frequently mutated within T-ALL patients[50].

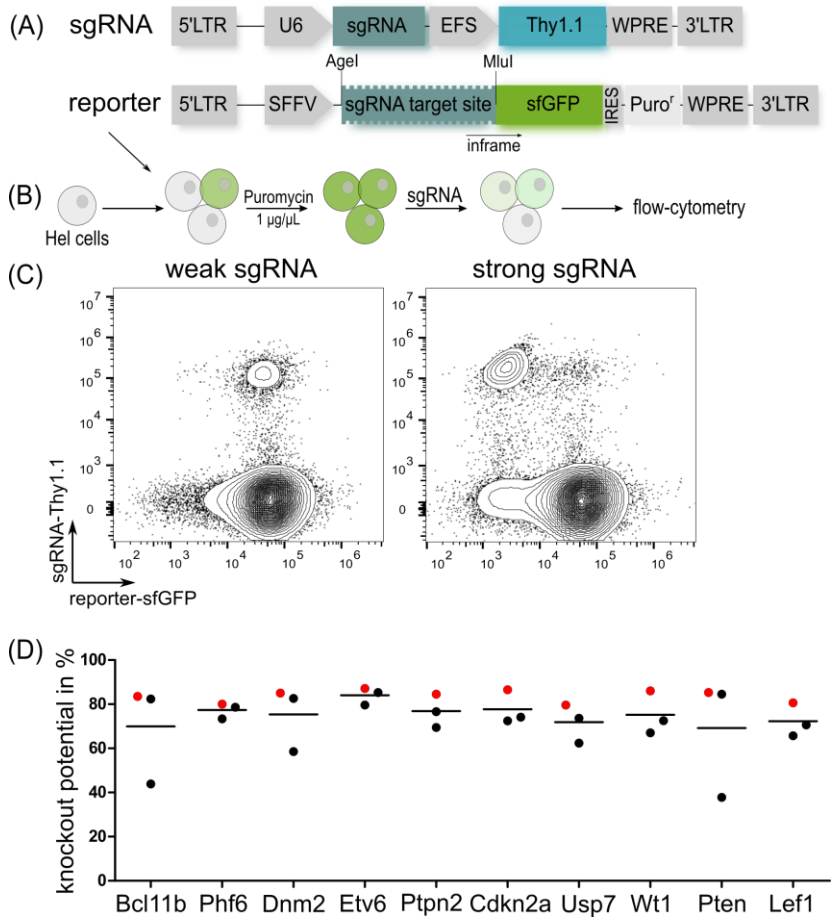


Figure 11: Evaluation of sgRNA cleavage potential for selected genes. Vector design for sgRNA containing vectors, and the reporter vector containing the sgRNA targets sites and their PAM sequences in-frame of a superfolder green fluorescence protein (sfGFP). (B) HEL cells were transduced with the reporter, selected with Puromycin and transduced again with a vector containing the sgRNA and a Thy1.1 for transduction control. (C) Representative flow cytometry blot after sgRNA cleavage, with a representative shift in sfGFP expression of sgRNA-Thy1.1 positive cells after cleavage. (D) Summary of the cleavage potential of three considered sgRNAs for each gene. Results are presented as individual knockout potential of each sgRNA and their mean for each gene. The best candidate, highlighted in red, was used for further application. SFFV – spleen focus forming virus promoter, EFS – elongation factor 1α short, LTR – long terminal repeat, WPRE- woodchuck hepatitis virus posttranscriptional regulatory element

For each gene to be targeted, three different high-ranked sgRNAs were designed using the online tool CCTop (https://cctop.cos.uni-heidelberg.de:8043/) and screened for their cleavage efficacy[146,149]. Therefore, a fluorescence reporter was cloned, harboring all target sequences for each sgRNA in-frame of a superfolder green fluorescent protein (sfGFP) and a puromycin-N-acetyltransferase[149]. In the case of Cas9-mediated cleavage, DNA repair could lead to a

frameshift and thus prevents translation of sfGFP (Figure 11 A). Since murine gene sequences will be targeted within the reporter assay, the human HEL cell line (AML cell line) was used for the screenings to avoid unwanted cleavage of the endogenous genetic regions. Cells were transduced with the reporter at low transduction efficiency of 10 – 20%, preventing multiple insertions[173], and selected with Puromycin subsequently. The transduced and selected HEL cells were then transduced with a vector expressing the sgRNA under the control of the RNA polymerase III promotor U6 and the thymus cell antigen 1.1 (Thy1.1 or CD90.1) (Figure 11 A). The surface protein Thy1.1 allowed staining for transduced cells and later cell sorting by magnetic-activated cell sorting (MACS). After five days, expression of sfGFP was measured by flow cytometry. The shift in its expression was used to calculate the cleavage efficacy (Figure 11 B + C). With only a few exceptions, most predicted sgRNAs possessed high cleavage capacities of > 80% for most target genes (Figure 11 D). The most promising sgRNAs candidates (marked in red in Figure 11 D) were chosen for the later application in chapter 5.6.

5.4 Immortalization potential of inverted oncogenes

Many activating or inactivating mutations can be linked to the development of T-ALL. In the previous approach, the inactivation of specific genes that are related to T-ALL formation by Cas9 was established. In addition, it was aimed to overexpress transcription factors frequently found to be overexpressed in T-ALL and alter cell signaling, particularly the JAK-STAT, RAS-MAPK-ERK and PI3K-AKT-mTOR pathways[65] by using the inverse vector construct. In particular, human IL7R$^{c.731_732insTTGTCCCAC}$, NRAS^{G12D}, and PIK3CD^{E1021K} were introduced to hyperactivate these signaling pathways. The NRAS^{G12D} mutation is well described and can often be found in various leukemia types[95,174], while activating mutations inside the human *IL7R* have a broader spectrum of genetic variations. The used *IL7R* mutation contained a TTGTCCCAC insertion between base pairs 731 and 732 of *IL7R* (named IL7R$^{c.731ins732}$), introducing disulfide bonds to promote ligand-independent homodimer formation and constitutive signaling transduction[175,176]. For the hyperactivation of the PI3K-AKT-mTOR pathway, we decided to use a constitutively active form of the catalytic subunit PIK3CD, which predominantly regulates this signaling in leukocytes[177]. Based on a somatic mutation screen of Kwok-Shing Ng *et al.*, we decided on the E1021K mutation, which was best described for its activating function and prevalence among PIK3CD mutations[152,178]. Therefore, these mutant human signaling proteins were included in the oncogenic gene pool. As potent oncogenic transcription factors, murine TAL1, LMO2 and TLX1 were selected, which are linked to the early and late cortical stages of T-ALL and, by this, relate to distinct biological subtypes of the disease[65,67]. Before introducing these genes into animals, the sufficient activation and their transforming effect was verified in murine cell lines.


(A)
Tlx1.PIK3CD^{E1021K}
Tlx1.NRAS^{G12D}
Tal1.Lmo2
Tlx1.IL7R$^{c.731ins732}$
mock
5'LTR
SFFV
lox66
dTomato
P2A
lox71
WPRE
3'LTR
Tlx1
Tal1
Lmo2
(B)
oncogene casette
Cre-recombinase
washing and reseeding
+ IL-3
cultivation
- IL-3
flow cytometry
BAF3 or 32D

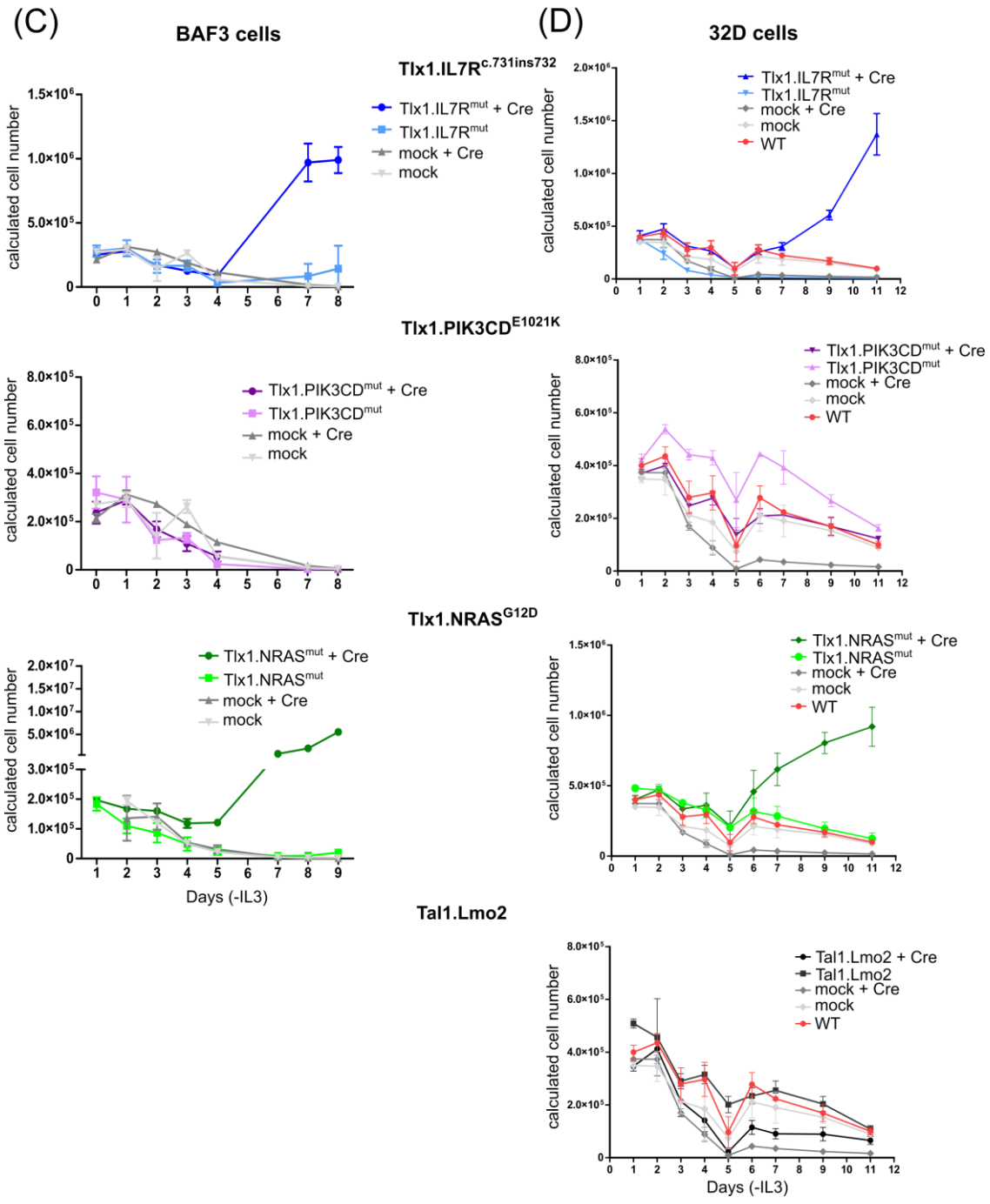

(C)
BAF3 cells
(D)
32D cells
Tlx1.IL7R$^{c.731ins732}$
Tlx1.PIK3CD^{E1021K}
Tlx1.NRAS^{G12D}
Tal1.Lmo2
calculated cell number
Days (-IL3)
Tlx1.IL7R^{mut} + Cre
Tlx1.IL7R^{mut}
mock + Cre
mock
Tlx1.PIK3CD^{mut} + Cre
Tlx1.PIK3CD^{mut}
Tlx1.NRAS^{mut} + Cre
Tlx1.NRAS^{mut}
Tal1.Lmo2 + Cre
Tal1.Lmo2
WT

Figure 12: Immortalization potential of different transcription factors and mutant cell signaling proteins (type A vectors) in BAF3 and 32D cells. (A) Vector design for the expression of Tlx1.PIK3CD^{E1021K}, Tlx1.NRAS^{G12D}, Tal1.Lmo2 and Tlx1.IL7R$^{c.731ins732}$ as inverse vector construct. (B) 32D or BAF3 cells were transduced in RPMI + IL3 medium with the different oncogenes and subsequently with a Cre recombinase. After stable transduction, cells were washed and cultivated in the absence of IL3. Cell viability and numbers were checked daily by flow cytometry. (C) Recalculated cell count of transduced BAF3 cells without IL3 supplementation in absence and presence of Cre recombinase. Mock transduced cells served as control. (D) Recalculated cell count of transduced 32D without IL3 supplementation in the absence and presence of Cre recombinase. Wild-type (WT) and mock transduced cells served as control. Presented are mean ± SD of three technical replicates. SFFV – spleen focus forming virus promoter, LTR – long terminal repeat, WPRE- woodchuck hepatitis virus posttranscriptional regulatory element

The transforming ability of the mutant signaling proteins in the context of Tlx1 was investigated in the two IL3-dependent cell lines, BAF3 and 32D. Both cell lines resemble immortalized blast-like cell types from early hematopoietic development. While 32D cells are myeloblast-like cells derived from long-term cultivated murine BM cells[179], BAF3 cells are pro-B cells[180]. Both cell lines can be used to test intracellular signaling and immortalization properties of genes of interests after withdrawal of IL3[179,181]. Next, *Tlx1.P2A.NRAS^{G12D}* (named Tlx1.NRAS^{G12D}), *Tlx1.P2A.PIK3CD^{E1021K}* (named Tlx1.PIK3CD^{E1021K}), *Tlx1.P2A.IL7R$^{c.731ins732}$* (named Tlx1.IL7R$^{c.731ins732}$) and *Tal1.P2A.Lmo2* (named Tal1.Lmo2) were cloned into the inverse construct, as described before (Figure 12 A). In the presence of a Cre recombinase, the whole cassette between both mutated *loxP* sites (*lox66* and *lox71*) gets inverted once, leading to the loss of dTomato protein expression and transcription of the aforementioned gene cassettes. After the introduction of Cre, cells were washed to remove the remaining IL3 and cultivated in the absence of cytokines (Figure 12 B). Cell survival and proliferation were measured by DAPI staining and cell counting using flow cytometry. In all cell lines, most cells died after IL3 removal. After the introduction of Cre, *Tlx1.IL7R$^{c.731ins732}$* and *Tlx1.NRAS^{G12D}* transduced cells started to grow cytokine-independently within 5 days (BAF3 cells) or 6 days (32D cells) (Figure 12 C+D). For the *Tal1.Lmo2* transduced group in 32D cells and *Tlx1.PIK3CD^{E1021K}* group in both cell lines, no growth and survival was observed, which was expected, since AKT-signaling alone is not sufficient to immortalize these cell lines (A. Schwarzer, unpublished observation). In BAF3 cells, a delayed cell rescue was observed after 7 days in non-Cre transduced cells from the *Tlx1.IL7R$^{c.731ins732}$* and to a minor tendency the *Tlx1.NRAS^{G12D}* groups.

Next, it was examined in which cellular signaling pathways were hijacked by the transforming Tlx1.IL7R$^{c.731ins732}$ and Tlx1.NRAS^{G12D} expression in immortalized BAF3 cells, facilitating cytokine-independent growth. Therefore, Western Blot analysis was performed on various compartments of MAPK/ERK, PI3K/AKT/mTOR, and JAK/STAT signaling. Although Tlx1.IL7R$^{c.731ins732}$ + Cre immortalized BAF3 cells proliferated much faster than Tlx1.NRAS^{G12D} + Cre immortalized cells, a more robust activation of signaling pathways was observed for Tlx1.NRAS^{G12D} + Cre (Figure 13). Indeed, a strong upregulation of PI3K/AKT/mTOR pathway was detected, including stronger phosphorylation of mTOR, S6RP, P-p70S6K, and slightly

increased phosphorylation of 4EBP1 and AKT at site serine 473 (P-Akt-S473) compared to Tlx1.IL7R$^{c.731ins732}$ + Cre immortalized BAF3 cells. In addition, MAPK/ERK signaling was slightly more enhanced in Tlx1.NRAS^{G12D} + Cre immortalized cells, indicated by phosphorylation of ERK. However, Tlx1.IL7R$^{c.731ins732}$ + Cre immortalized BAF3 showed increased JAK/STAT signaling, marked by phosphorylation of STAT5, which is missing in Tlx1.NRAS^{G12D} + Cre cells. Nevertheless, none of the observed activations reached the amount of activation and phosphorylation measured in WT cells in the presence of IL-3. Untreated BAF3-WT cells exhibited reduced but not absent phosphorylation, possibly due to the shorter starvation period compared to the immortalized cells. Otherwise, cell death, would have occurred too rapidly. In summary, the constructs were efficiently inverted upon Cre exposure, revealing that IL7R$^{c.731ins732}$ and NRas^{G12D} led to cytokine independency in 32D and BAF3 cell lines.

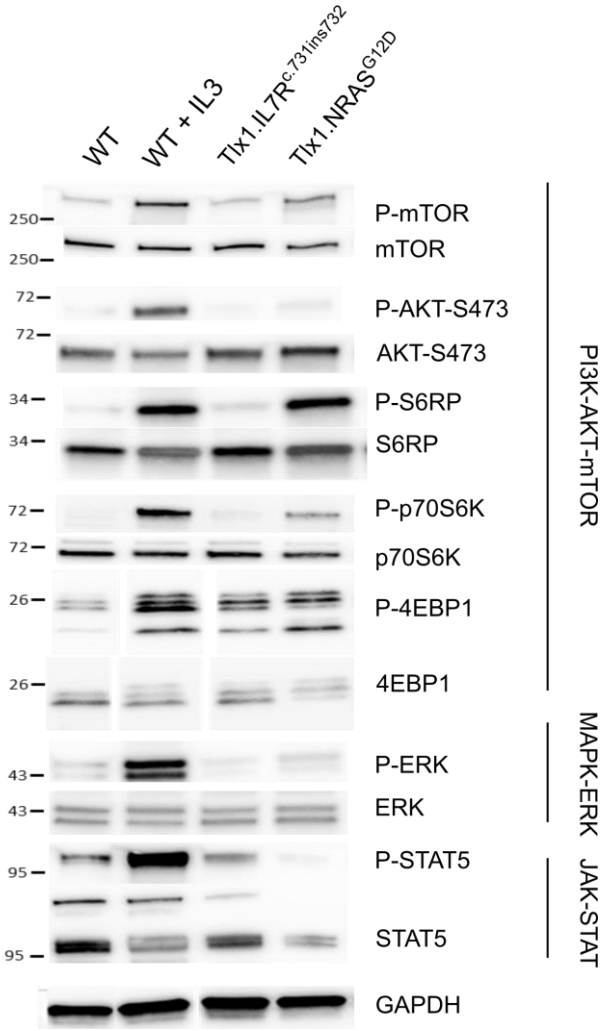


Figure 13: Tlx1.IL7R$^{c.731ins732}$ and Tlx1.NRAS^{G12D} enhanced different cellular signaling to promote cytokine-independent growth. Immunoblot of immortalized BAF3 cells, from the Tlx1.IL7R$^{c.731ins732}$ and Tlx1.NRAS^{G12D} groups on different compartments of the PI3K-AKT-mTOR, MAPK-ERK and JAK-STAT pathways. Untransduced BAF3 cells served as control, either after overnight starvation from IL3 (wild-type – WT) or in presence of IL3 (WT + IL3).

5.5 Primary transplantation of transduced BM from *LSL.Cas9 x Lck-cre* mice

After having determined the optimal vector design, the oncogenes and tumor suppressors of interest, and sgRNA efficacy, all approaches were combined to generate complex T-ALL in mice. The four groups from chapter 5.5 were tested in the following transplantation experiments. Lineage-negative BM cells were isolated from the *LSL.Cas9 x Lck-cre* mouse strain and lentivirally co-transduced in a step-wise process with the respective driver genes determining different biological subtypes in the inverse vector construct and sgRNAs targeting subtype-related tumor suppressor genes (details in chapter 4.3.4). Transduced HSPCs ($CD45.2^+$) were then transplanted into sublethally irradiated $CD45.1^+$ C57BL/6J mice (Figure 15 A). For the group design, between 5 to 8 frequently occurring mutations within the subtype were chosen from the large patient sequencing screen of Liu *et al.*[50] It was conceptualized that the overexpressed genes for murine transcription factors *Tlx1*, *Tal1*, and *Lmo2*, and for mutant human signaling proteins $NRAS^{G12D}$, $PIK3CD^{E1021K}$, $IL7R^{c.731ins732}$, which were cloned in antisense flanked by *lox66* and *lox71* sites, would have been activated by Cre recombinase and the subtype-related mutations set by Cas9 through interaction (Figure 15 B). To allow a high combination frequency of many introduced alterations, sufficient and comparable transduction rate efficiencies were aimed for. Therefore, all lentiviral particles were titrated on lineage-negative BM cells to prevent calculation inaccurateness when comparing transduction of different cell types. While sgRNA vectors generated high lentiviral titers (data not shown), transduction efficiencies of the primary oncogenes in the inverse construct remained surprisingly low. Except for the Tlx1.$NRas^{G12D}$ construct, transduction efficiencies remained at < 10%, even with an increasing volume of concentrated lentiviral supernatant (Figure 14). We hypothesized that because of the "inverse" vector design of the oncogenes, the stability of the primary transcript from which the *dTomato* reporter is transcribed is compromised due to its long and unstructured 3'-untranslated region encoding the oncogenes in antisense. If this is the case, one would expect to see an increased vector copy number (VCN) within transduced cells linear to the used volumes of lentiviral particles. Indeed, this pattern was observed for the reporter vector, containing the fluorescence protein genes *dTomato* in sense and *mTagBFP2* in antisense orientation (see Figure 10 A), serving as mock control. During flow cytometry, only 5.1% – 6.6% $dTomato^+$ cells were detected, independent of the used volume, and would have resulted in a VCN < 0.1. However, a linear increase of the VCN was observed with increasing volumes of viral supernatant up to a VCN > 3. Of note, untransduced cells already contained a VCN of 2 due to the *LSL.Cas9.EGFP.WPRE* locus, which was introduced into the ROSA26 locus for the elaboration of the *LSL.Cas9* mouse strain. In the presented data, each VCN was therefore adjusted by 2. A similarly linear increase was observed for the Tlx1.$NRAS^{G12D}$ construct. While the total VCN reached up to 1.96, only 28% $dTomato^+$ cells were observed by

flow cytometry. For the Tlx1.PIK3CD^{E1021K} construct, titration results from flow cytometry and VCN behaved comparably (Figure 15). All results were obtained seven days post-transduction to decrease the presence of episomal vector particles. At this stage, it is important to mention that EGFP expression was detected within the lineage-negative cells and in all cellular fractions (see Supplementary Figure 1), indicating the presence of Cre recombinase and impeding the reliability of flow cytometry results.

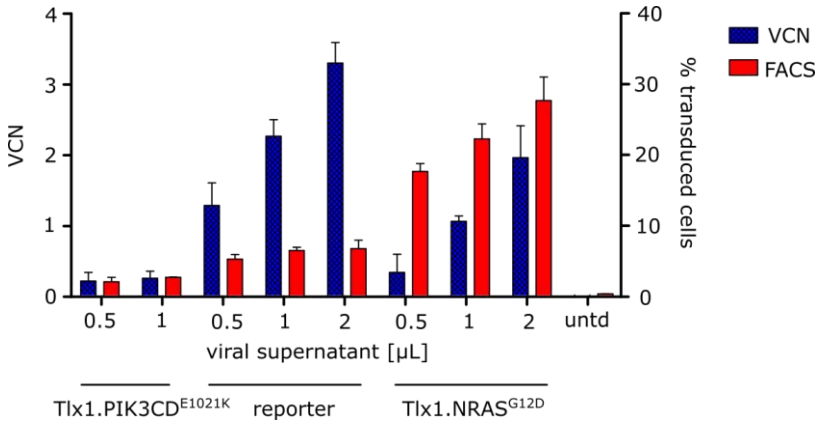


Figure 14: Titration of inverse vector constructs on lineage-negative BM cells. Measuring the transduction efficiency on lineage-negative bone marrow (BM) cells from *LSL.Cas9 x Lck-cre* mice via flow cytometry (FACS) or total vector-copy number (VCN) determination. Cells were transduced by applying different volumes of lentiviral supernatant and kept in culture for 5 days before measurement. The reporter vector encoded for a dTomato and an inverted mTagBFP2 fluorescence protein. VCN values were normalized to the untransduced (untd) control cells, which already possessed a value of 2 by the *LSL.Cas9* transgene. Presented are mean ± SD of three technical replicates.

For the final transduction of lineage-negative BM cells before transplantation, an MOI of 0.2 was aimed for the driving oncogenes and an MOI of 0.5 for the sgRNAs. The high MOI for the sgRNA vectors was chosen to increase the percentage of multiple co-transduction of different sgRNAs within the same cell. With an average of 6 sgRNAs vectors for the different experimental groups, only 1.56% of all transduced cells could harbor all individual sgRNAs (MOI 0.5^6 = 0.015625; represents the percentage for independent co-occurrence). This led to a very high number of possible random combinations of all constructs within different cells to ensure an effective screening for oncogenic cooperativity. Transplantation was performed in one or two rounds. A summary of each transplanted group, and its group name is summarized in Table 26 in chapter 4.3.4. For simplicity of this work, the different experimental main groups will be named pool 1 – 4. Thereby, pool 1 consists of Tlx1.NRAS^{G12D} with dedicated sgRNAs in the primary co-transduction mix. Moreover, pool 2 resembles Tlx1.PIK3CD^{E1021K}, pool 3 Tlx1.IL7R$^{c.731ins732}$ and pool 4 Tal1.Lmo2 and PIK3CD^{E1021K}, each with their respective set of sgRNAs. Figure 15 B summarizes all targeted genes. Final transduction efficiencies for each

transplanted group are summarized in Supplementary Figure 3. Transduction rates differed between the groups, ranging from 2.5% in pool 2 to 17.02% in pool 1 for the inverse vector and for the sgRNA vectors from 48.7 ± 15.5% in pool 4 to 98.8% in pool 2.

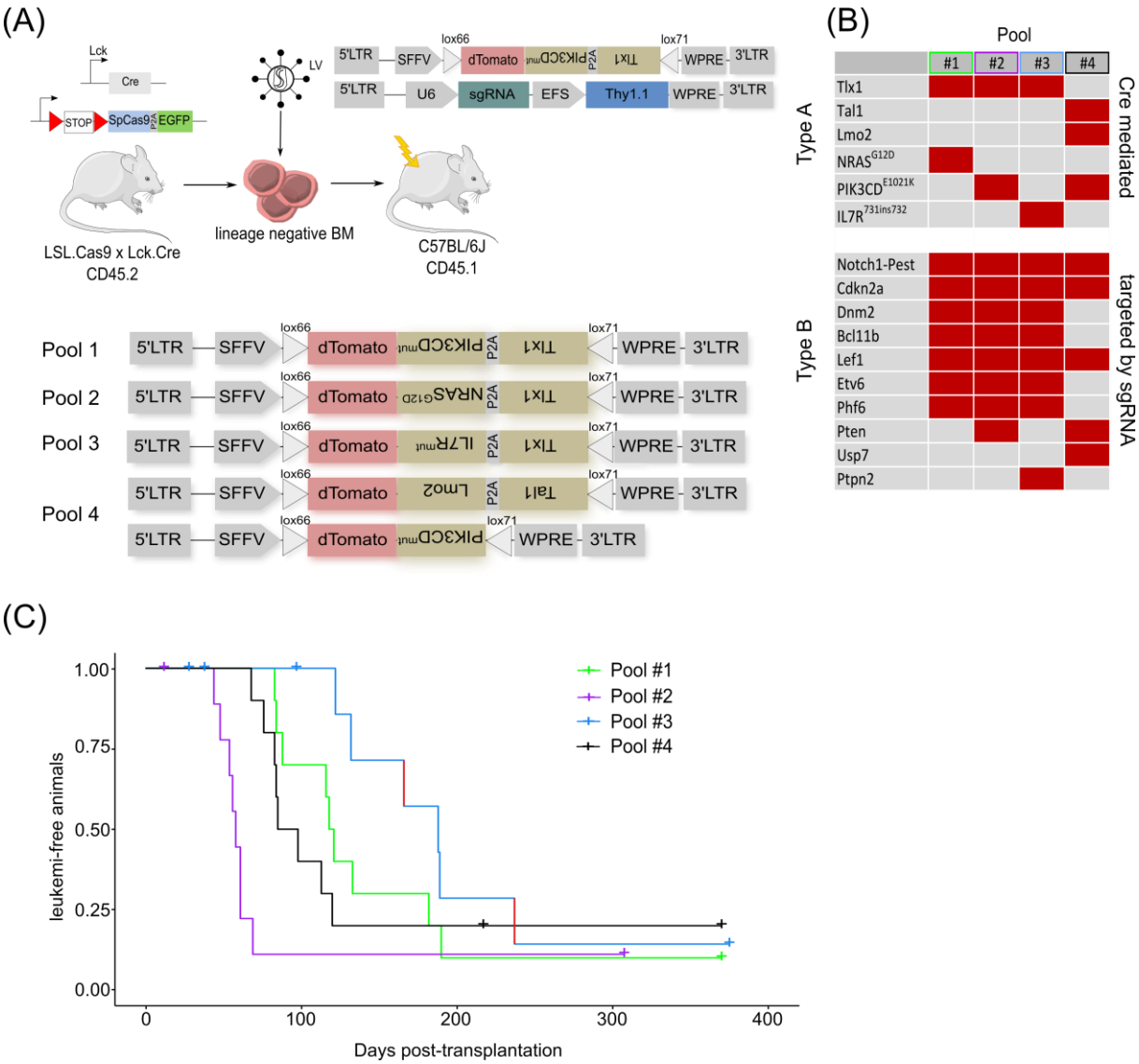


Figure 15: Leukemia development and its relation to animals' chimerism after transplantation. (A) Schematic overview of the co-transduction of lineage-negative BM cells from *LSL.Cas9 x Lck-cre* mice with an inverse vector system, containing an oncogenic transcription factor (Tlx1, Tal1, Lmo2) and/or a mutated signaling protein ($NRAS^{G12D}$, $PIK3CD^{E1021K}$, $IL7R^{c.731ins732}$) as type A aberration. In addition, sgRNAs were transduced, targeting different genetic regions. (B) Summary of targeted genetic aberration within each transplanted group, including overexpressing type A genes and type B genes targeted by sgRNAs. (C) Kaplan-Meier curve of transplanted animals for all four groups. Leukemia development that could not be assigned to the transplanted cells was marked as red line. Animals that sickened for other reasons were excluded from the curve (marked as crosses). n = 10 animals for each transplanted group.

The four main groups were transplanted at two time points, with 5 animals each time and an independent transduction. Five weeks after transplantation, animals were checked for the first time for the chimerism of donor cells in the blood. Peripheral blood cells were stained for the hematopoietic surface markers CD45.1 and CD45.2 to distinguish between donor cells (CD45.2) and recipient cells (CD45.1). The four different groups achieved a median chimerism

of 1.6%, 11.4%, 1.7% and 10.8%, respectively. Interestingly, in each group one to two animals possessed much higher engraftment, while all other animals only had chimerism of < 5 – 10%, partly even below 1% (Supplementary Figure 4 A).

Leukemia formation was observed in each of the transplanted main groups (pool 1 – 4). The fastest leukemia development was reached for pool 2 with a mean of 56 days until the pre-defined endpoint criteria were reached. Animals from pool 1 and pool 4 developed leukemia at an average of 119 and 90 days, respectively. Although Tlx1.IL7R$^{c.731ins732}$ exhibited the strongest transforming effect in the *in vitro* assays, *in vivo* leukemia formation for pool 3 showed the longest latency, with 197 days on average (Figure 15 C). Within this group, three animals showed early signs of BM failure and had therefore been removed from the experiment before leukemia was detected. Only 6 out of 10 transplanted animals developed leukemia. However, two of the obtained leukemias contained a CD45.1^{+} recipient phenotype (Supplementary Figure 5), attributed to transforming effects from the irradiation or other procedural implications on recipients' cells. A summary of the fate of each animal is summarized in Supplementary Table 1. Because of the smaller evaluable group size of pool 3 (4 versus 8-9 animals), the main focus of this thesis was put on pool 1 (*Tlx1.NRAS*^{G12D} + sgRNAs transduction mix), pool 2 (*Tlx1.PIK3CD*^{E1021K} + sgRNAs transduction mix) and pool 4 (*Tal1.Lmo2* + *PIK3CD*^{E1021K} + sgRNAs transduction mix).

Animals from each primary transplanted group showed typical signs of T cell leukemia, including infiltration of the thymus, spleen, liver, and BM with lymphoid blasts and a rapid increase in the white blood count (WBC) due to circulating blasts. Within the different groups, minor differences could already be observed on a morphological level. For pool 2 and pool 4 transplanted animals, most cells infiltrated into the liver and spleen leading to an average of 7.7 and 5.7-fold increase in size for the spleen and 2.4 and 2.5-fold for the liver, compared to size-matched untransplanted animals (Figure 16 A). A different pattern was observed for the first group. Here, the liver and spleen were less affected, while most tumor cells populated in the thymus, leading to a 7.8-fold increase in its size. Within the pool 3 transplanted animals, liver size did not change and spleen size increased partly. Nevertheless, all four animals exhibited a very strong thymus growth (6.1 ± 2-fold). This phenotype was profound for the first and third groups. In addition, enlarged lymph nodes at the legs and arm-pits were only observed within single animals of these two groups. The WBC, as already indicated, rapidly increased at the time of sickening, reaching its highest values in group 4 with 238 ± 145 x 10^3 cells/mm³, compared to 108 ± 99 x 10^3 cells/mm³ for group 1, 116 ± 75 x 10^3 cells/mm³ for group 2 and 59 ± 19 x 10^3 cells/mm³ for group 3 (Figure 16 B). Pappenheim staining on blood smears of diseased animals at the time of sacrifice confirmed a high density of immature lymphatic blasts in the peripheral blood (Figure 16 C, white arrows). These blast-like cells were

also identified in the cytospins from the spleen and BM of diseased animals (Figure 16 D). In summary, we observed development of acute leukemias with high penetrance and variable latencies, which was dependent on the combination of sgRNA/oncogene pools used.

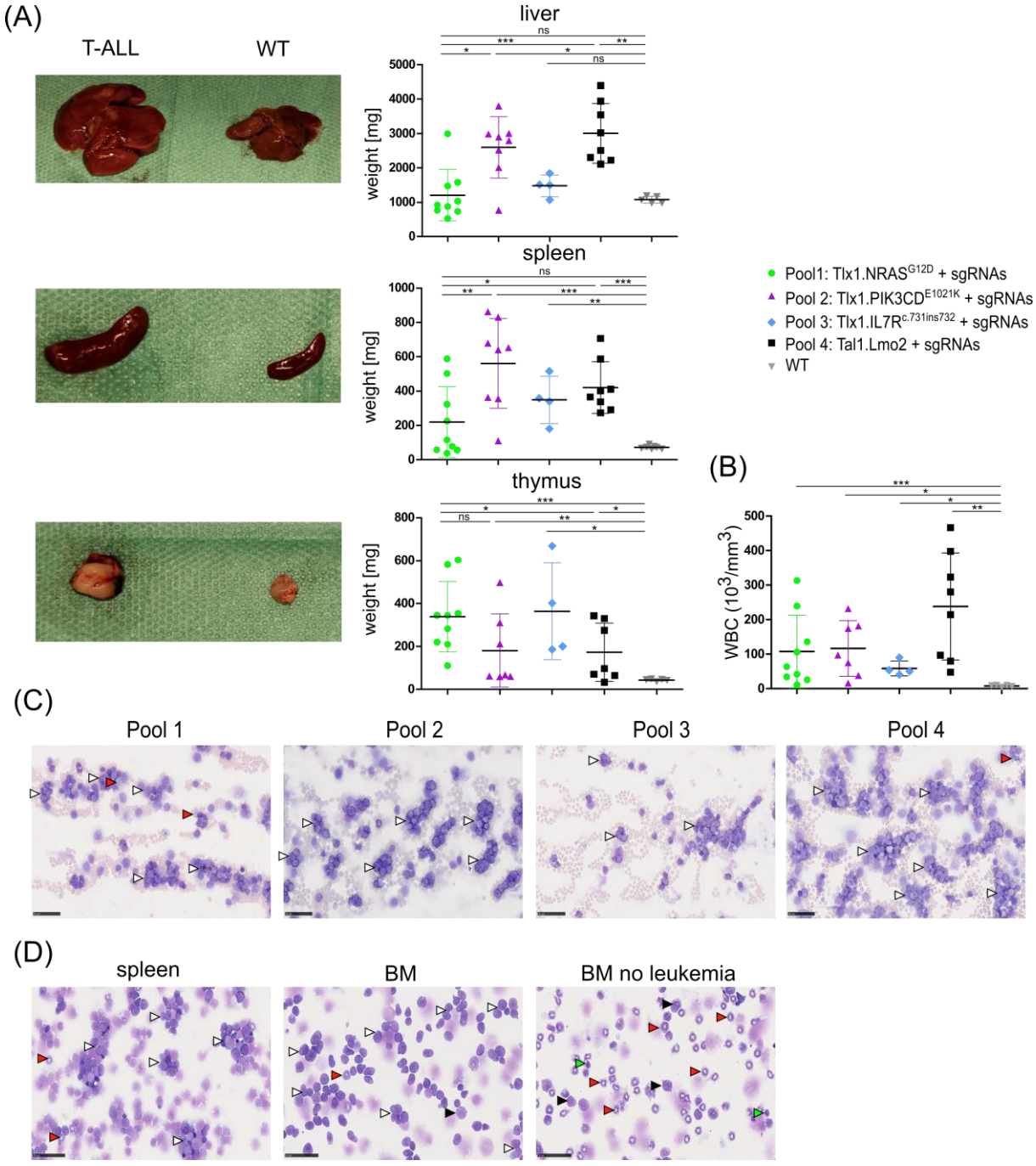


Figure 16: Morphological analysis of obtained leukemia from transplanted animals. (A) Representative picture of the organ sizes for liver, spleen, and thymus from pool 4 compared to a wild-type mouse. Summary of all organ sizes from each transplanted group (T-ALL). Size-matched CD45.1 mice (wild-type, WT) were used as reference. (B) White blood count (WBC) from heart blood on the day of sacrifice. Presented are mean ± SD and individual values. Depicted are only those animals, where T-ALL development could be assigned to an origin from CD45.2$^+$ transplanted cells. Mann-Whitney U test (* p< 0.05, ** p< 0.01, *** p< 0.001, ns – not significant). (C) Exemplary blood smears from each transplanted main group. (D) Exemplary cytospins from spleen and bone marrow (BM) after Pappenheim staining. Arrows show leukemic blasts (white), myeloid cells (red), macrophages/mast cells (black) and mature lymphocytes (green).

5.6 Obtained primary T-ALL represent distinct developmental stages of T cell maturation

The stage of T cell development, at which the T cell is arrested in its differentiation, determines the biological subtype of the T-ALL and corresponds to its behavior and clinical prognosis[28,182]. Therefore, the immune phenotype of obtained leukemia was investigated to characterize their stage of development and allow comparison to known descriptions from human patients. In the clinics, T-ALL containing mutations in *TLX1* are assigned to the early cortical-stage ($CD4^+CD8^+$ double-positive), whereas aberrant *TAL1* expression correlates with a late cortical phenotype (CD3 and CD8 or CD4 single-positive)[65,183].

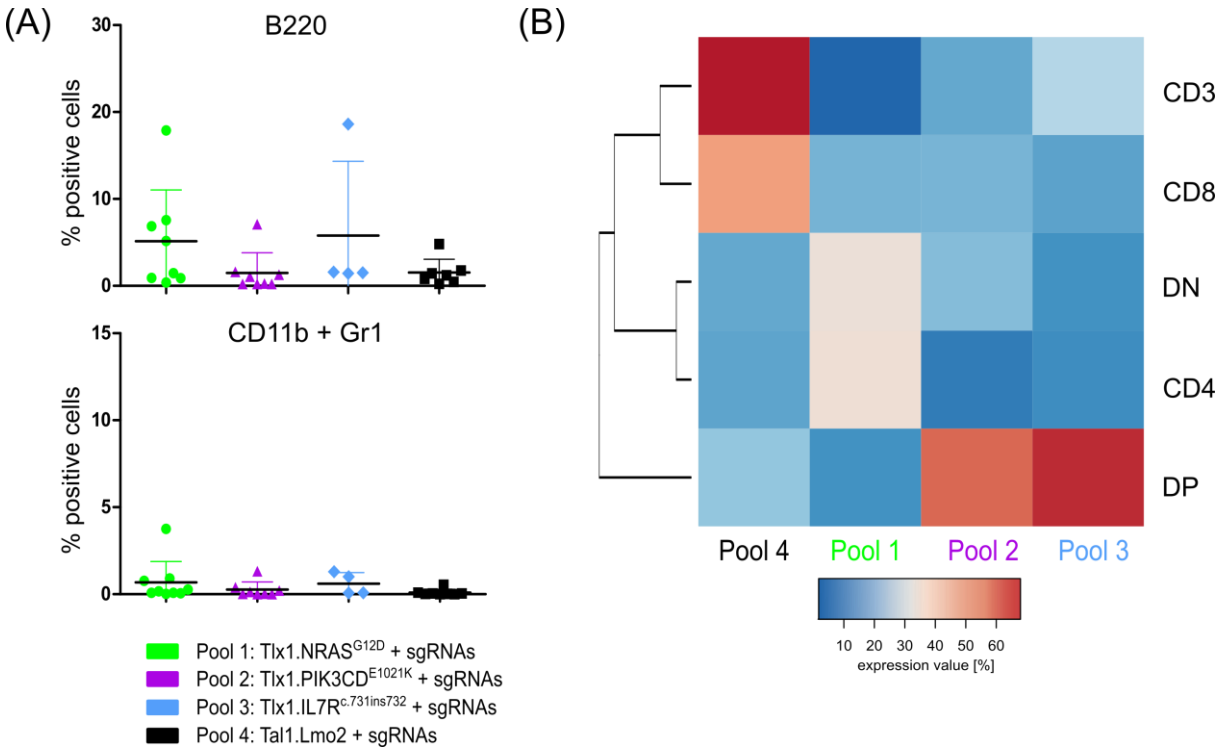


Figure 17: Summarized immune phenotypes of leukemia samples from transplanted animals. (A) Expression of the B cell marker B220 and the myeloid cell marker CD11b and Gr1 within each leukemia that derived from transplanted $CD45.2^+$ cells. Presented are mean ± SD and individual values. (B) Heatmap summarizing the mean proportion of cells positive for the mature T cell surface markers $CD3^+$, $CD4^+$, $CD8^+$, $CD4^+CD8^+$ (double-positive, DP) and $CD4^-CD8^-$ (double-negative, DN) from blood and spleen-infiltrating T-ALL cells.

Although the increased thymus size already indicated leukemia formation within the T cell lineage, phenotypic analysis was performed for blood and spleen-infiltrating cells to prove the T cell phenotype. First, samples were checked for the expression of the B cell marker CD45R/B220, the myeloid surface protein CD11b and the granulocyte marker Gr1, to exclude malignant cell growth from other hematologic lineages. Indeed, nearly no expression of CD11b and Gr1 was detected by flow cytometry in any groups (Figure 17 A). Except for one outlier

within group 1 (*Tlx1.NRAS*^{G12D} + sgRNAs transduction mix) and group 3 (*Tlx1.IL7R*$^{c.731ins732}$ + sgRNAs transduction mix), B220-expressing cells were similarly underrepresented, indicating the absence of B cells, monocytes, macrophages, and granulocytes. At this point, it couldn't be determined whether the B220-expressing population possessed a leukemia phenotype or was simply a remnant of untransformed donor cells.

In the following immune-phenotypic characterization, the focus was set on the mature T cell markers CD4, CD8, and CD3, which are frequently expressed in T-ALL and allow the discrimination of immature and more mature phenotypes[73,184–186]. Therefore, all T-ALL samples were analyzed by flow cytometry for the cell surface expression of these markers. On average, pool 4 leukemia cells highly expressed CD3 and CD8 on their cell surface. For pool 1 leukemia cells, the presence of a CD4$^+$ and DN (CD4$^-$CD8$^-$) population was characteristic. Leukemia cells from pool 2 and pool 3 were both expressing CD4 and CD8 (DP) and low CD3 (Figure 17 B). However, these data only present mean expression values from each leukemia sample within one group. When comparing surface expression levels of the leukemia cells from each individual animal, more variances could be detected within the different groups. Figure 18 summarizes the immune phenotypes of each donor-derived leukemia as a heatmap. In general, samples from pool 1, pool 2, and pool 4 clustered in their surface marker expression pattern. In pool 1, two outliers harbored a CD8$^+$ or DP phenotype, while the other samples were mainly positive for CD4 or did not express any of the measured surface markers (DN). Similar observations were done for pool 2, pool 3, and pool 4, with few outliers each. Samples from pool 2 and 3 mainly clustered together through a high DP phenotype and mostly lacking CD3 expression. One outlier from each of both groups had higher expression of CD3. Leukemia, obtained from pool 4, possessed higher CD3 expression compared to most other samples. The percentage of expressing cells ranged from 50 – 98 %. Here, most samples were single positive for CD8, with three exceptions that were instead positive for CD4, DP, or DN, respectively.

Next, we performed a deeper characterization of the immature T cells. Therefore, additional staining of CD25 and CD44 was performed to distinguish DN1 (CD44$^+$CD25$^-$), DN2 (CD44$^+$CD25$^+$), DN3 (CD44$^-$CD25$^+$) and DN4 (CD44$^-$CD25$^-$) developmental stages. This complete T cell staining was not applied on the periphery leukemia cells, as analyzed in Figure 18 A, but on thymus cells (Figure 18 B). These cells might not recapitulate the same phenotype as the infiltrating T-ALL cells. However, the increased thymus size in animals with a more immature T-ALL phenotype (mostly from pool 1 (P1)) still recapped the major developmental stages. For P1-1 (pool 1, animal number 1), P1-2, P1-7, P1-8 and P1-9 with a predominant DN phenotype in the periphery, similar was observed in the thymus. Here, cells were stuck at

the DN1 and DN2 stages. Surprisingly, P1-5, which possessed a $CD4^+$ and DN phenotype in the spleen, was mostly $CD4^+CD8^+$ (DP) within the thymus.

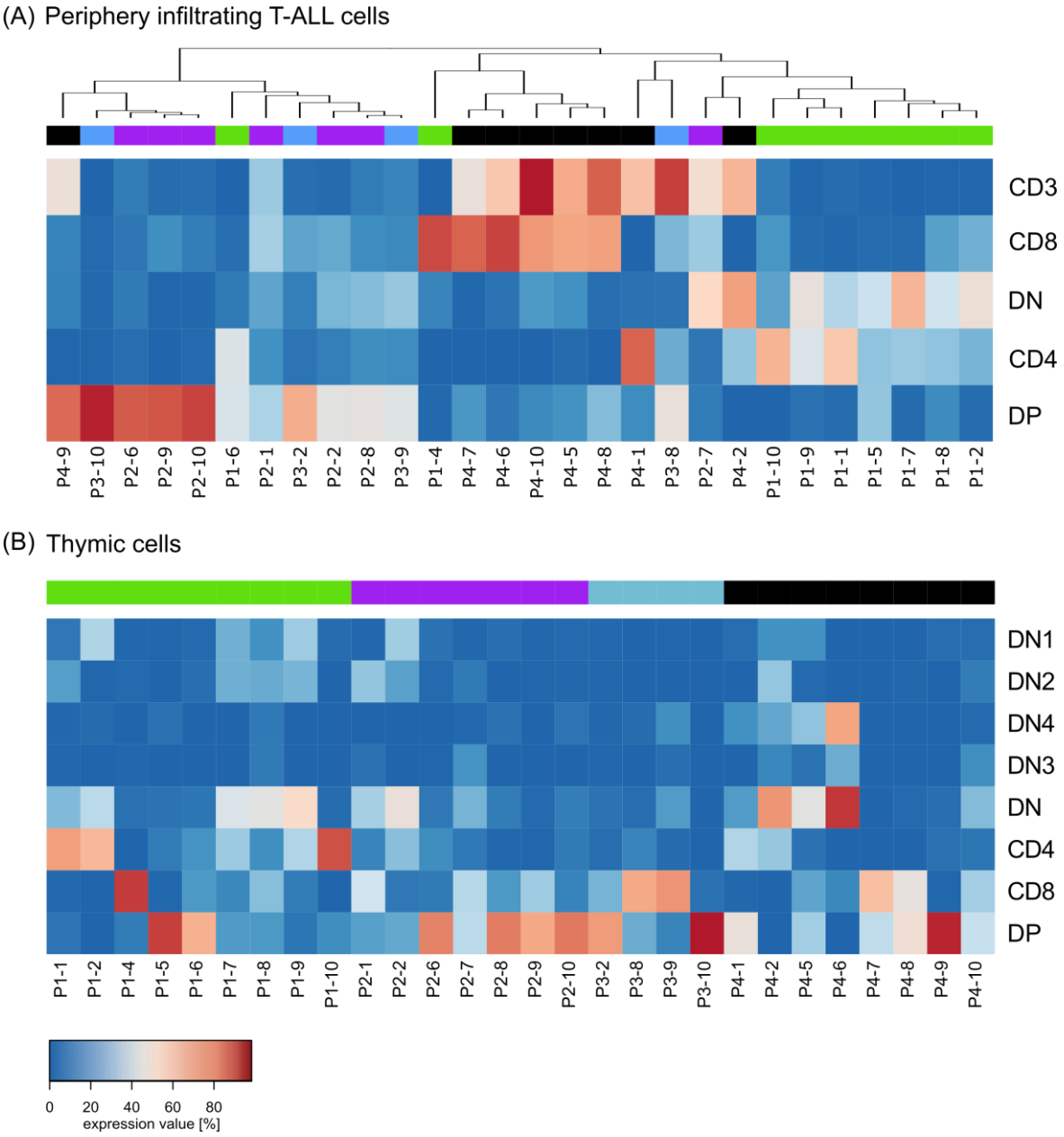


Figure 18: T-ALL immunophenotype from each individual animal. Heatmap summarizing the expression rate of (A) $CD3^+$, $CD4^+$, $CD8^+$, $CD4^+CD8^+$ (double-positive, DP) or absence of $CD4^+$ or $CD8^+$ (double-negative, DN) from spleen infiltrating T-ALL cells. Analysis of thymus cells (B) further distinguishes DN stages towards DN1 ($CD44^+CD25^-$), DN2 ($CD44^+CD25^+$), DN3 ($CD44^-CD25^+$) and DN4 ($CD44^-CD25^-$). Group identity of each animal is color-coded as green (pool 1), purple (pool 2), blue (pool 3) or black (pool 4).

Within pool 2, only P2-1 (Pool 2, animal number 1) and P2-2 showed a dominant DN phenotype within the thymus. Immature cells were mostly at DN2, or DN1 and DN2 stages, respectively. Samples from pool 2 and 3 that showed a DP phenotype in the periphery also showed a DP and CD8 phenotype in the thymus. While P4-2 was the only sample from pool 4 displaying a DN phenotype in the spleen, the thymuses of P4-1, P4-5 and P4-6 also showed a dominant

DN fraction. These displayed mainly a DN3 and DN4 phenotype, while P4-2 consisted mostly of DN2 cells, but contained proportions of all other stages, too. The remaining samples from this group were mostly $CD8^+$ or $CD4^+CD8^+$ within the thymus. One striking outlier was P4-6, which possessed a dominant immature DN4 phenotype within the thymus, while the periphery T-ALL cells expressed mostly $CD8^+$ and $CD3^+$. In summary, the obtained T-ALLs represented different immune phenotypes determined by the sgRNA/oncogene pools used. Most importantly, the different pools elicited leukemia phenotypes that corresponded to the clinical subgroups for which the pools were designed.

5.7 Contribution of type A and type B delivering vectors on T-ALL development

T-ALL subsets are mainly classified by chromosomally rearranged transcription factors, such as HOXA, TLX1, TLX3, TAL1, and LMO[65,76,187,188]. Following this assignment, obtained T-ALL samples were screened for the presence of the transduced inverse vector constructs, harboring either *Tlx1* or *Tal1* and *Lmo2*. In parallel, not only its presence was verified but also whether the vector was successfully inverted through Cre, which initiates the expression of the vectors' transgene. At this point it is important to mention that only the vector orientation was assessed in this work and not the expression of the encoded proteins. A PCR approach was designed to capture the orientation of the inverse vector construct by placing one forward primer within the promoter sequence (upstream of the *lox61* site) and two reverse primers aligned to the reading frame of the *dTomato* or *Tlx1*/*Tal1* sequences (Figure 19 A). For the *PIK3CD*^{E1021K} vector, one reverse primer was placed on the WPRE sequence instead of *dTomato*. Thus, a misleading signal of the co-transduced Tal1.Lmo2 vector in pool 4 should be prevented. PCR analysis revealed the presence of inverted Tlx1.NRas^{G12D} in each obtained T-ALL from pool 1 with three leukemia additionally carrying a non-inverted vector (P1-5, P1-7, and P1-10). Within animals transplanted with pool 2, only the animals P2-2, P2-8 and P2-10 harbored an inverted Tlx1.PIK3CD^{E1021K}. One very weak inverted vector signal was identified for P2-6. For the animals P2-7 and P2-9, a dim non-inverted band was detected. No Tlx1.IL7R$^{c.731ins732}$ containing vector was measured in any pool 3-derived leukemia. Within leukemia obtained from pool 4, the presence of Tal1.Lmo2 vector was confirmed in each individual. However, leukemia from animal numbers 7 and 9 from pool 4 showed no inversion of the vector. Conspicuously, both animals harboring only the non-inverted vector had the longest latency until leukemia development within this group (120 and 113 days, respectively). For the PIK3CD^{E1021K} vector in pool 4, another PCR approach was performed to prevent misleading signals from the Tal1.Lmo2 vector. Instead of the reverse primer within the *dTomato* sequence, one primer binding to the WPRE element was chosen (Figure 19 A).

Within P4-6 and P4-9, a PIK3CD^{E1021K} vector in its active orientation was detected, and a dim mixture of inverted and non-inverted vectors for P4-8. For unknown reasons, a positive signal appeared in non-Cre-transduced cells, which served as control, and that did not disappear in different tested cell lines, which were transduced with the type A vector. However, absence of the band in the T-ALL samples excluded general contamination or off-targeting of the PCR reaction.

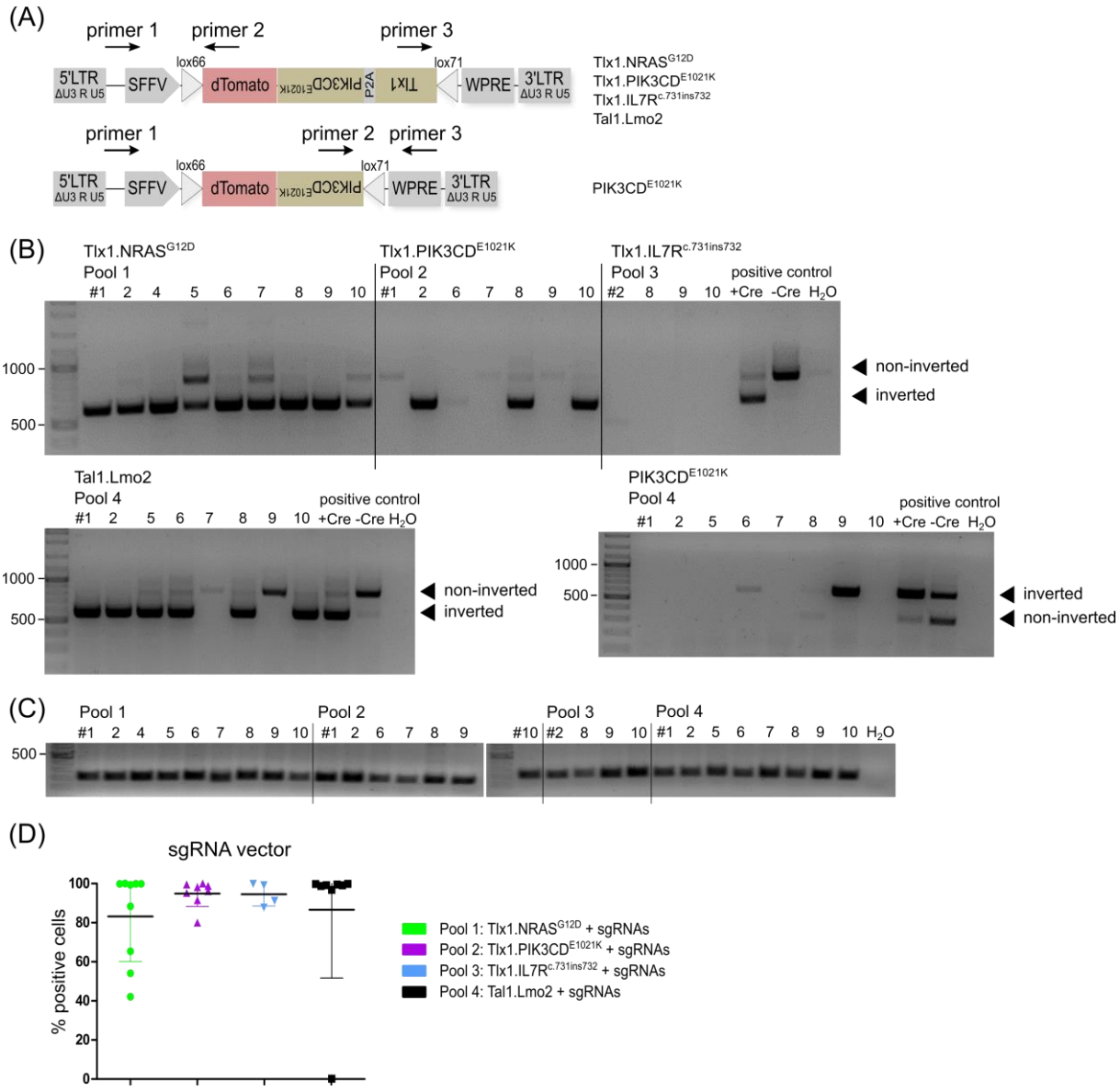


Figure 19: Presence and orientation of the transduced vector constructs within obtained T-ALL. (A) A PCR was designed to capture the orientation of the inverse vector construct. Depending on the Cre-mediated orientation of the *dTomato*-oncogene cassette, different PCR products were generated. Exemplary design for the *Tlx1.PIK3CD^{E1021K}* coding inverse vector and the *PIK3CD^{E1021K}* encoding vector were depicted. (B) Presence of differently orientated inverse vector constructs within each CD45.2$^+$ T-ALL. (C) Reference PCR on *Notch1* gene locus as control for intact DNA (Table 28). (D) Percentage of sgRNA expressing T-ALL cells measured by Thy1.1$^+$ phenotype within spleen or BM-infiltrating cells. Presented are mean ± SD and individual values.

Besides the primary oncogene cassette in the inverse vector construct, the presence of sgRNA vectors within the final leukemia cells was investigated. Each sgRNA vector that was part of the initial transduction mix encoded the thymus cell antigen 1.1 (Thy1.1 or CD90.1). Cells that received at least one of the sgRNA encoding vectors expressed the THY1.1 protein on their surface. This implementation allowed us to assign transformation to the introduced genetic modifications and to enrich leukemia cells from the spleen or BM. For most leukemia samples, high amounts of THY1.1 were detected by flow cytometry on CD45.2^{+} infiltrating cells from the spleen (Figure 19 C). For pool 2 and 3, more than 78% of the cells were positive for THY1.1. The same was observed for pool 4. However, no THY1.1 was detected on the cell surface of one CD45.2^{+} T-ALL in this group. This sample was collected from animal P4-7 and showed the same phenotype as the other leukemia samples from this group, where more genetic abnormalities were expected due to Cas9-mediated mutations. Leukemia samples from pool 1 were more heterogeneous. For six leukemia samples, > 88% of cells expressed THY1.1. In the other three samples, only 42% – 65% of the cells were positive for THY1.1.

As the presence of THY1.1 did not allow any conclusions on which altered genes contributed to the final leukemia progression, targeted NGS analysis was performed on the CD45.2^{+} T-ALL samples. Each gene locus that was potentially targeted by Cas9 was amplified by PCR and sequenced subsequently. Genetic variances were detected and quantified using the Galaxy platform (https://usegalaxy.org/). The *mus musculus* genome assembly GRCm38 (mm10) was used as a reference genome. On average, 4 mutated genes were found within the T-ALL samples from pool 1 (7 sgRNAs). In pool 2 (8 sgRNAs), an average of 5 – 6 mutant genes were found. Less mutated genes at the same time were found within pool 3 (8 sgRNAs) with 3 mutations and pool 4 (5 sgRNAs) with 2 – 3 hits (Figure 20). The most frequent aberrations for pool 1 were detected in *Bcl11b*, *Notch1*, *Lef1*, *Cdkn2a*, and *Dnm2*. However, not all mutations were found within each T-ALL sample of pool 1. Mutations in *Etv6* and *Phf6* occurred less frequently. A different pattern was observed for pool 2, in which each T-ALL possessed indels within *Bcl11b*, *Notch1*, *Cdkn2a*, and *Pten*. Sporadic hits were found within individual samples for *Lef1*, *Etv6* and *Dnm2*. No *Phf6* alteration was detected in any sample of pool 2, even though the transduction mix contained the functional sgRNA. For leukemia samples from pool 3, no common mutation pattern was observed. Each gene locus was mutated in at least one leukemia. Only *Notch1*, *Cdkn2a*, and *Dnm2* were affected in two of four leukemias. Absence of Tlx1.IL7R$^{c.731ins732}$ and the heterogeneous genotype did not allow further generalization of this group's observations. Within pool 4, *Pten*, *Notch1* and *Cdkn2a* indels were among the most prevalent aberrations. No mutation was detected for the targeted *Usp7* gene locus.

sample	Tlx1.Nras^{G12D}	Bcl11b	Notch1	Lef1	Cdkn2a	Etv6	Dnm2	Phf6
P1-1	Y							
P1-2	Y							*
P1-4	Y							
P1-5	Y							
P1-6	Y							
P1-7	Y							
P1-8	Y							
P1-9	Y							
P1-10	Y							

sample	Tlx1.PIK3CD^{E1021K}	Bcl11b	Pten	Notch1	Lef1	Cdkn2a	Etv6	Dnm2	Phf6
P2-1	(Y)								
P2-2	Y								
P2-6	(Y)								
P2-7	(Y)								
P2-8	Y								
P2-9	(Y)								
P2-10	Y								

sample	Tlx1.IL7R$^{g.731ins732}$	Bcl11b	Notch1	Lef1	Cdkn2a	Etv6	Dnm2	Phf6	Ptpn2
P3-2	N								
P3-8	N								
P3-9	N								
P3-10	N								

sample	Tal1.Lmo2	PIK3CD^{E1021K}	Pten	Notch1	Lef1	Cdkn2a	Usp7
P4-1	Y	N					
P4-2	Y	N					
P4-5	Y	N					
P4-6	Y	Y					
P4-7	Y	N					
P4-8	Y	(Y)					
P4-9	Y	Y					
P4-10	Y	N					

2x KO alleles (both alleles with frameshift)
1x KO, 1x inframe indel allele within regulatory/functional domain
2 inframe indels in both alleles within regulatory/functional domain
1x KO, 1x inframe indel allele outside regulatory/functional domain
1x KO, 1x WT allele
2 inframe indels in both alleles outside regulatory/functional domains
1x WT, 1x inframe indel outside regulatory/functional domain
WT

Figure 20: Genotyping primary T-ALL samples. The first column summarizes the presence (Y – yes, N – no) and orientation of the respective primary inverse oncogene construct (based on Figure 19). Brackets indicate detection but no inversion. Targeted Next-Generation-Sequencing on each genomic region that was targeted by a sgRNA from the respective lentiviral pool. Color-coded mutation profile indicating the functional consequences of the observed insertions/deletions (indel). Assignment of regulatory/functional regions is summarized in Supplementary Table 3 and additionally on an attached hard drive. * - Evaluation of the sequencing results (P1-2) revealed contamination with recipient DNA, impeding quantification.

Most detected indels within *Notch1*, *Cdkn2a, Etv6, Phf6, Lef1*, *Pten*, and *Ptpn2* led to a frameshift and, by this, disrupted gene function. The occurrence of in-frame indels for *Etv6*,

Phf6 and Lef1 was presumed to disrupt protein function potentially, as its location lay within functional domains or affected regions frequently mutated in T-ALL[50,189–191]. For *Pten*, *Notch1*, and *Cdkn2a*, the targeted region could not be directly linked to a potential loss of function. Nevertheless, at least one allele of them was knocked out. Interestingly, for the *Bcl11b* and *Dnm2* gene loci, only heterozygous knockouts occurred. While one allele was usually disrupted through frameshift insertion, the other allele was usually not affected, or the indel was assigned to no important functional protein domain. Hence, for these genes the loss of one allele promoted leukemogenesis, whereas the loss of both alleles seems to be deleterious and was never observed.

Surprisingly, in the P4-7 T-ALL sample, showing no sgRNA-dedicated THY1.1 expression (Figure 19 D), sgRNA-associated mutations were detected within the *Notch1* and *Cdkn2a* locus, suggesting that the THY1.1 reporter might have been damaged or silenced, or a site-specific knockout without an active sgRNA-containing vector occurred. The opposite was observed for P1-10. Even though THY1.1 expression was measured by flow cytometry, no mutation within one of the targeted loci was detected. Another conspicuous feature was found within P3-2, P3-9, P3-10, and P4-1 T-ALL samples. Here, either *Cdkn2a* or *Notch1* contained indel alleles, even though a dual sgRNA vector facilitated targeting of both genes. A detailed overview of each detected mutation within all T-ALL samples can be found in Supplementary Table 3. In summary, our strategy was able to generate genetically complex T-ALL with multiplexed loss-of-function and gain-of-function lesions. The frequency of co-occurrence of the lesions and the gene dose (homozygous vs heterozygous hits) gave first insights into the oncogenic potency and oncogenic cooperativity of individual events.

Complex genetic interactions facilitated T-ALL development with the multiplexed combination of overexpressed oncogenes (type A, Cre-mediated) and the disruption of relevant genes (type B, Cas9-targeted). In the following, the focus was set on which contribution type A and type B mutations had on leukemia formation. For this purpose, either the inverse vector system containing the type A aberrations or the pool of sgRNAs from each group was transplanted into six CD45.1^{+} recipient animals each. With the number of transplanted control animals, it was intended to capture enough phenotypic variance, which was already observed within the main groups. The primary transplantation of pool 3, including *Tlx1.IL7R*$^{c.731ins732}$ and dedicated sgRNAs, led to only four leukemia formations originating from transplanted modified BM cells. Together with the long latency until the leukemia outbreak (mean = 172 days) and the absence of *Tlx1.IL7R*$^{c.731ins732}$ containing vector, this group was excluded for the comparison of the vector influence on disease development. In the following, BM from *LSL.Cas9 x Lck-cre* mice was transduced either with the sgRNA pools from groups 1, 2 and 4, or only with *Tlx1.NRAS*^{G12D}, *Tlx1.PIK3CD*^{E1021K} or *Tal1.Lmo2* and transplanted subsequently.

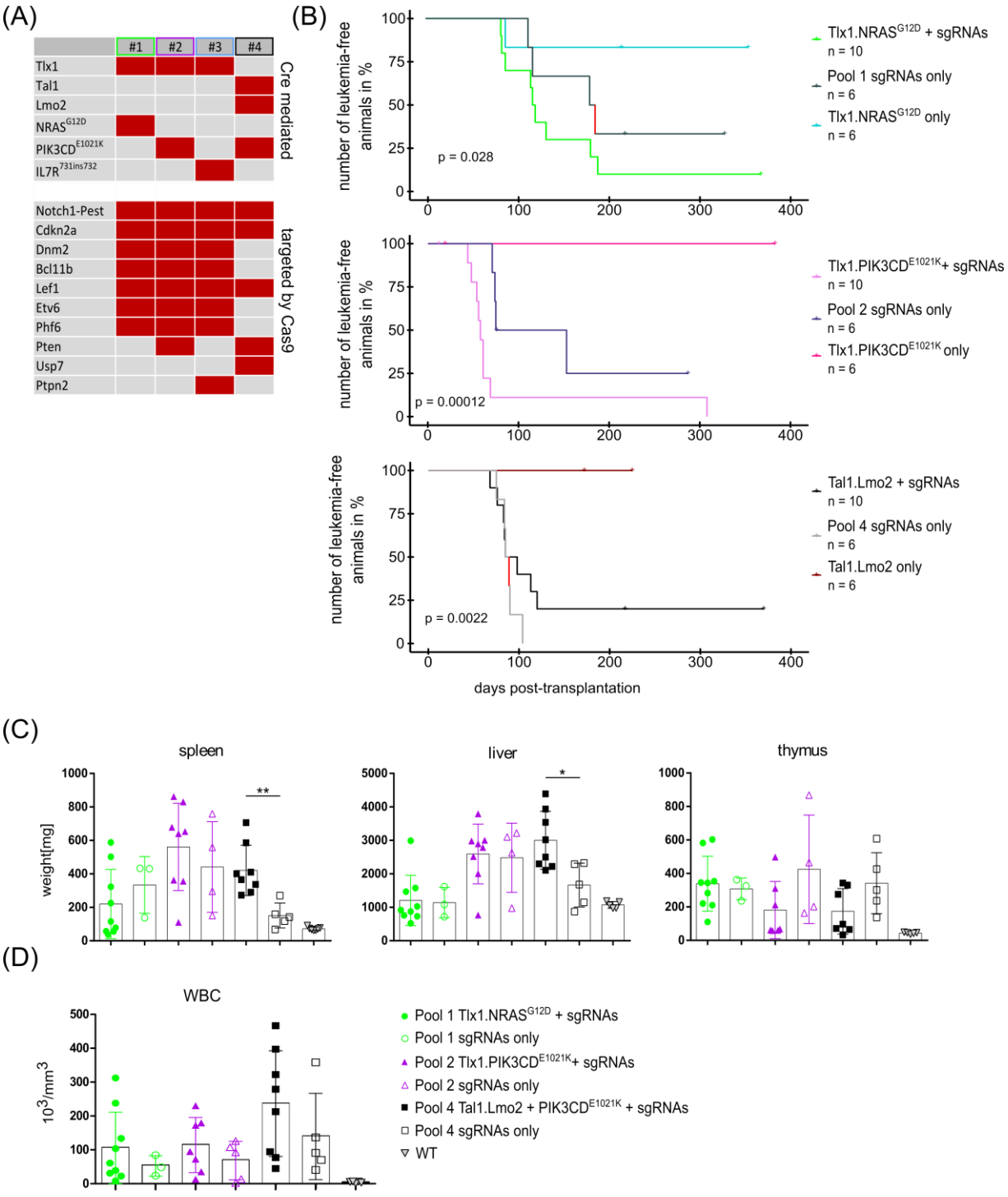


Figure 21: Overview of transplanted control groups and their influence on disease progression. (A) Summary of all transduced targeted genes among the transplanted groups, including overexpressing type A oncogenes and type B genes targeted by sgRNAs. Control groups included the target genes from one block for each group compromising only the primary oncogene or only the co-transduced sgRNAs. (B) Kaplan-Meier curve for leukemia-free animals from the main experimental groups and their corresponding control groups. Leukemia development that did not arise from CD45.2$^+$ transplanted cells was marked in red. (C) Organ sizes and (D) white-blood count (WBC) from leukemia-diseased animals with exclusive CD45.2$^+$ leukemia phenotype. Sizes-matched wild-type (WT) animals served as reference. Presented are mean ± SD and individual values. Statistical analysis was performed by two-tailed t-test. * p≤ 0.05, ** p≤ 0.01.

With two exceptions, animals that received either *Tlx1.NRAS*^{G12D}, *Tlx1.PIK3CD*^{E1021K}, or *Tal1.Lmo2* did not sicken during the one-year experimental period. Only one animal that was transplanted with *Tlx1.NRAS*^{G12D} transduced BM cells developed leukemia that could be assigned to the transplanted cells (CD45.2^{+}). In the *Tal1.Lmo2* transplanted control group, one animal developed a BM failure after 172 days, whereas the other animals developed no signs of disease. Of note, the remaining animals from the *Tal1.Lmo2* control group were at day 230 post-transplantation when these results were generated.

Transplantation of only the sgRNA pools promoted complete transformation and leukemia formation. Animals that were transplanted with the sgRNAs from pool 1 developed signs of leukemia within 137 days on average (n = 3). This was slightly later compared to 123 days on average for the main group, additionally containing *Tlx1.NRas*^{G12D} in the primary transduction mix. It should be noted that only the mice that developed leukemia with a CD45.2^{+} phenotype were used for the statistics. Mice that developed leukemia with a CD45.1^{+} phenotype are marked in red in the survival plot (Figure 21 B). Animals that became ill for other reasons or remained healthy throughout the duration of the experimental were not used for the comparison. Within the control group for pool 2, one mouse had to be removed from the experiment because of a late BM failure and another mouse did not develop any signs of disease within 330 days of the experiment. For animals transplanted with the pool 2 sgRNAs, mice sickened within 93 days (n = 4), compared to 56 days for pool 2 including *Tlx1.PIK3CD*^{E1021K}. Again, one animal was removed from this experiment due to a BM failure and another one for too low chimerism of < 0.1 % CD45.2^{+} cells after 287 days. The lowest difference between the control and the main groups was observed for sgRNA pool 4 with an average of 88 days (n = 5), compared to 91 days for pool 4, including the type A vector. One animal from this group had a massively enlarged thymus but was not analyzed further since an origin from transplanted BM cells could not be verified for this animal. Besides the time until leukemia development, the weight of the spleen, liver, thymus, and the WBC levels were compared between diseased animals from the main groups and the dedicated sgRNA-harboring control groups. For pool 1 and pool 2, no significant differences in organ weight and WBC were observed (Figure 21 C + D). For pool 4, the spleen and liver were significantly larger than in the corresponding sgRNA control group.

One experimental group contained the inverse mTagBFP2 fluorescence reporter. It was used as a control group to rule out the transforming effects of the inverse vector design itself and to gain insight into the switching kinetics of the inverse vector system. Unfortunately, all transplanted animals became ill during the experiment for different reasons. Four out of six mice developed anemia at different time points after transplantation, while the remaining two animals developed a CD45.1^{+} leukemia phenotype (Figure 22 A). One of the leukemia animals

showed an enlargement of the spleen, liver and thymus and a high WBC level, while the other one was identified by a minor increase of the WBC and in spleen and thymus size (Figure 22 B+C). Blood was collected every 4 weeks during the experiment to check for blood parameters. In all animals, except for #3 with the leukemia phenotype, the platelet count (PLT) and hemoglobin (Hb) concentration decreased during the experiment, leading to the termination of the experiment (Figure 22 D). Peripheral blood was collected from all animals and analyzed for the chimerism of transplanted cells. On average, only 4.1% ± 4.9% of the cells were derived from the donor, indicating malignancy of the recipients' hematopoietic system (Figure 22 E). Within the increased thymuses, no expression of dTomato or mTagBFP2 could be measured by flow cytometry (data not shown), preventing further analysis of the kinetics of the reporter vector switch *in vivo*.

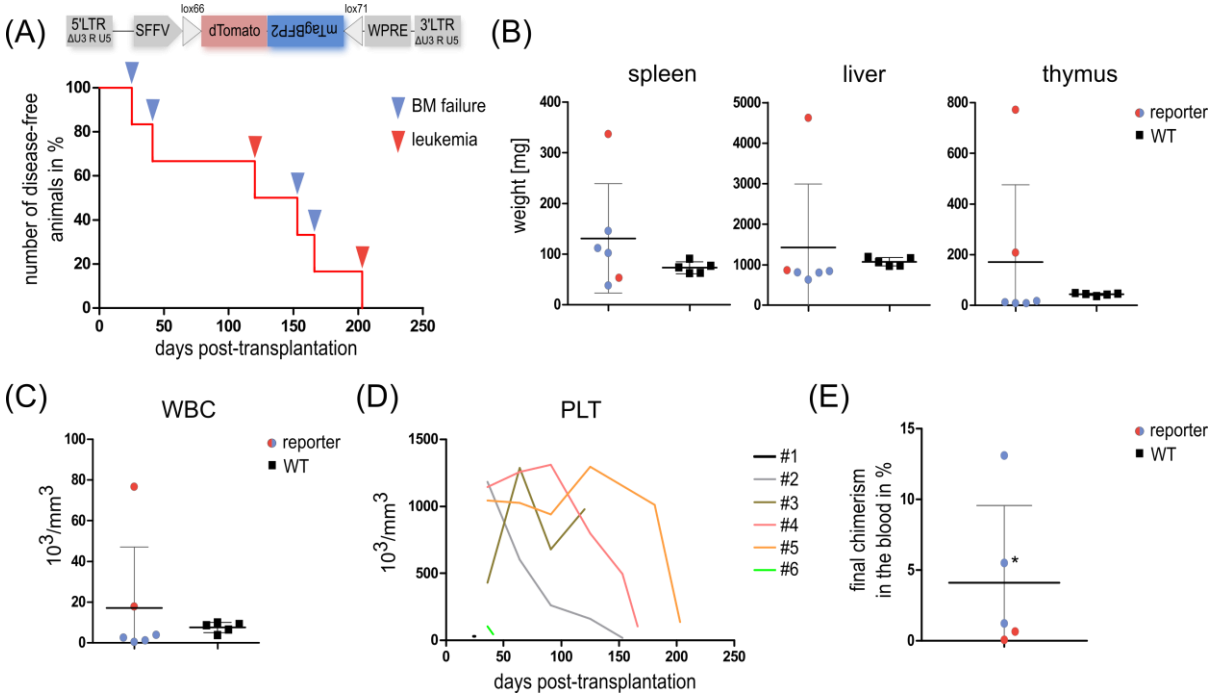


Figure 22: Morphological overview of diseased animals from the transplanted reporter control group. (A) Animals were transduced with the inverse construct containing the fluorescence proteins dTomato and mTagBFP2. Kaplan-Meier curve of disease-free animals. Animals sickened either from leukemia (red arrow) or bone marrow (BM) failure (blue arrow). (B) Summary of organ sizes for the spleen, liver and thymus, compared to wild-type (WT) animals for reference. (C) White blood count (WBC) from heart blood on the day of sacrifice. (D) Development of the platelet count (PLT) over the time of experimental duration for each individual mouse. (E) Chimerism of CD45.2$^+$ donor cells in the peripheral blood on the day of their sacrifice (* - chimerism of taken blood few days before sacrifice).

At this point, the obtained data did not exclude a potential oncogenic role of the inverse vector system for the disease development itself. No control transplantation with a mock vector without fluorescence genes was performed. Repetition of transplanting the reporter vector failed due to too low chimerism 50 days after transplantation. Still, a minor mTagBFP2$^+$ population could be detected in a very small fraction in CD45.2$^+$ thymocytes (Supplementary Figure 6). Nevertheless, the control groups comprising only *Tlx1.PIK3CD*^{E1021K}, *Tlx1.NRAS*^{G12D}

or *Tal1.Lmo2*, without associated sgRNA vectors, proved the safety of the vector design. The absence of the observed disease phenotypes from the reporter control group within the three control groups (except for one leukemia in *Tlx1.NRAS*^{G12D}, see Figure 21 B) diminished concerns regarding the general vector design.

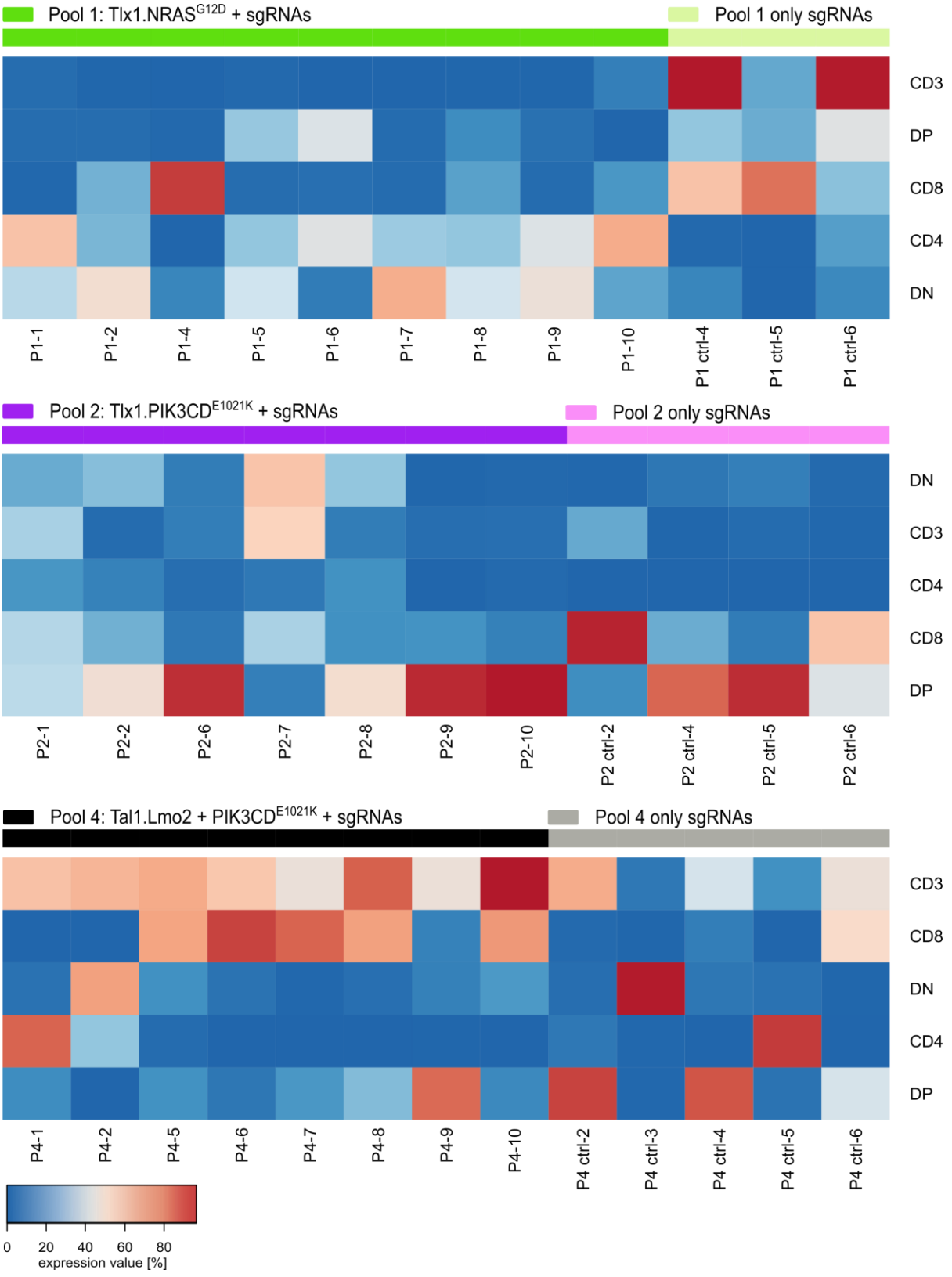


Figure 23: Immune phenotype for individual animals including control group. Leukemia phenotypes from transplanted sgRNA control groups were compared to their corresponding main groups, containing overexpressed oncogenes in addition. Heatmaps summarizing the percentages of cells expressing CD3, CD4, CD8, $CD4^+CD8^+$ (double-positive, DP) or do not express CD4 or CD8 at all (double-negative, DN) for each individual T-ALL with $CD45.2^+$ phenotype. Group identity of each animal is color-coded as depicted in the legend.

Besides the morphological comparison, we were interested in the specificity of the immune phenotypes, which were observed within the primary main groups. Therefore, the expression of the T cell surface markers CD3, CD4, and CD8 were analyzed within BM- or spleen-infiltrating leukemia cells.

Again, the absence of the myeloid, granulocytes, and B cell surface proteins CD11b, Gr1, and B220 indicated leukemia phenotype from mainly the T cell lineage (data not shown). Developmental differences were observed within the first pool between the main group and the sgRNA control group. While the main group consisted mainly of immature $CD3^-$DN or $CD3^-$ $CD4^+$ blasts, the control group had a more mature phenotype. Leukemia cells expressed high $CD3^+$ and were either DP or $CD8^+$. For the second transplanted pool, fewer variances were observed. Similar to the main group, transplantation of only the sgRNAs resulted in a DP or $CD8^+$ phenotype with low CD3 expression. In the case of P2-9 derived leukemia, in which no inverted and therefore active *Tlx1.PIK3CD*^{E1021K} vector was detected (Figure 19), a similar expression pattern was observed as seen within the sgRNA control group. However, for P2-7 we also could not detect inverted *Tlx1.PIK3CD*^{E1021K}, while its phenotype was mostly DN and $CD8^+$ with a slight expression of CD3, representing an outlier in this group. Leukemia from pool 4 sgRNAs had a very diverse phenotypic pattern. While the phenotypic variances within the main group with *Tal1.Lmo2* and *PIK3CD*^{E1021K} expressing vectors were more stable, a heterogeneous expression profile was observed for the control group. Instead of mostly expressing CD3 and CD8, one leukemia sample contained a primary immature DN phenotype, and another expressed only $CD4^+$. The other three samples were DP to varying degrees and had $CD3^+$ and single $CD8^+$ populations in different proportions. No NGS analysis was performed to identify and quantify genetic aberration within the control groups. In summary, whereas the sole expression of the sgRNA pools was able to induce T-ALL with comparable latency, the phenotypes observed were inconsistent and did not necessarily match the prespecified T-ALL subgroups they were designed for.

After exploring the potential of generating phenotypically and genetically complex T-ALL by relying on a combination of overexpressing oncogenes and introducing disease-related mutations, we next assessed the retransplantability of the obtained leukemia samples.

5.8 Primary T-ALL recapitulates leukemia formation after secondary transplantation

Secondary transplantation of primary leukemia samples is a hallmark of acute leukemia research[54,90,174,192]. The retransplantability supports the acute character of the obtained leukemia[54] and indicates the presence of leukemic stem cells, able to self-renew and re-establish the fully blown disease phenotype in secondary recipients[193]. To evaluate the acute leukemia nature of our T-ALL model and to investigate phenotype conservation, secondary transplantations were performed from one diseased animal into three recipients (Figure 24 A). Therefore, T-ALL cells were sorted by MACS targeting THY1.1 surface expression in transformed cells. This antigen contained a glycophosphatidylinositol (GPI)-linked surface membrane molecule, which was part of the sgRNA-containing vector and was therefore expressed in each cell harboring at least one inserted sgRNA. As already shown in Figure 19 D, flow cytometry analysis of primary leukemia samples revealed high THY1.1 expression in > 99% of CD45.2^{+} cells. Leukemia cells were purified from the spleen or BM for pool 2, 3, and 4, and transplanted together with lineage-negative BM cells from *LSL.Cas9 x Lck-cre* mice into three irradiated CD45.1 recipient mice. Since most leukemia cells resided in the thymus and not in the spleen or BM, isolation was adapted to the thymus. The co-transplanted healthy BM cells served as helper cells to transitionally reconstitute hematopoiesis until secondary leukemia developed (Figure 24 A). Within 14 – 15 days, animals transplanted with leukemia cells from pool 2 and pool 4 became ill and were sacrificed (Figure 24 B). Except for 2 animals, which had to be sacrificed because of an engraftment failure, all animals from these groups developed a leukemia. In the group of Tlx1.NRas^{G12D} + sgRNAs secondary recipients, only one group sickened and was sacrificed on day 41 post-transplantation (secondary pool 1 named SecP1). The three leukemia samples derived from primary P1-4 were appropriately designated SecP1-4 #1-3. Animals from another secondary transplantation, which received cells from P1-2 were withdrawn from the experiment after 30 days due to signs of transplantation failure. The absence of morphological changes suspended the onset of leukemia. Within the group of secondary transplantation from pool 3 T-ALL, only three out of 6 animals developed a secondary T-ALL (two transplanted groups). Two animals with anemia symptoms were removed after 38 and 89 days, indicating a late BM failure. The sixth animal did not sicken within 180 days post-transplantation. Supplementary Table 2 summarizes the experimental outcome of each secondary transplanted mouse.

Similar to primary transplantation, leukemia cells from pool 1 and 3 showed greater participation of the thymus during disease development (Figure 24 C). T-ALL cells from pool 1 mainly infiltrated the spleen and thymus in two of three animals, whereas the liver size and peripheral WBC levels were unaffected (Figure 24 D). In the secondary transplanted animals

from pool 2 and pool 4, T-ALL cells infiltrated mainly the spleen and slightly less into the liver, while they completely omitted infiltration of the thymus. Peripheral WBC levels varied widely between the different animals but were increased overall in pool 2 – 4.

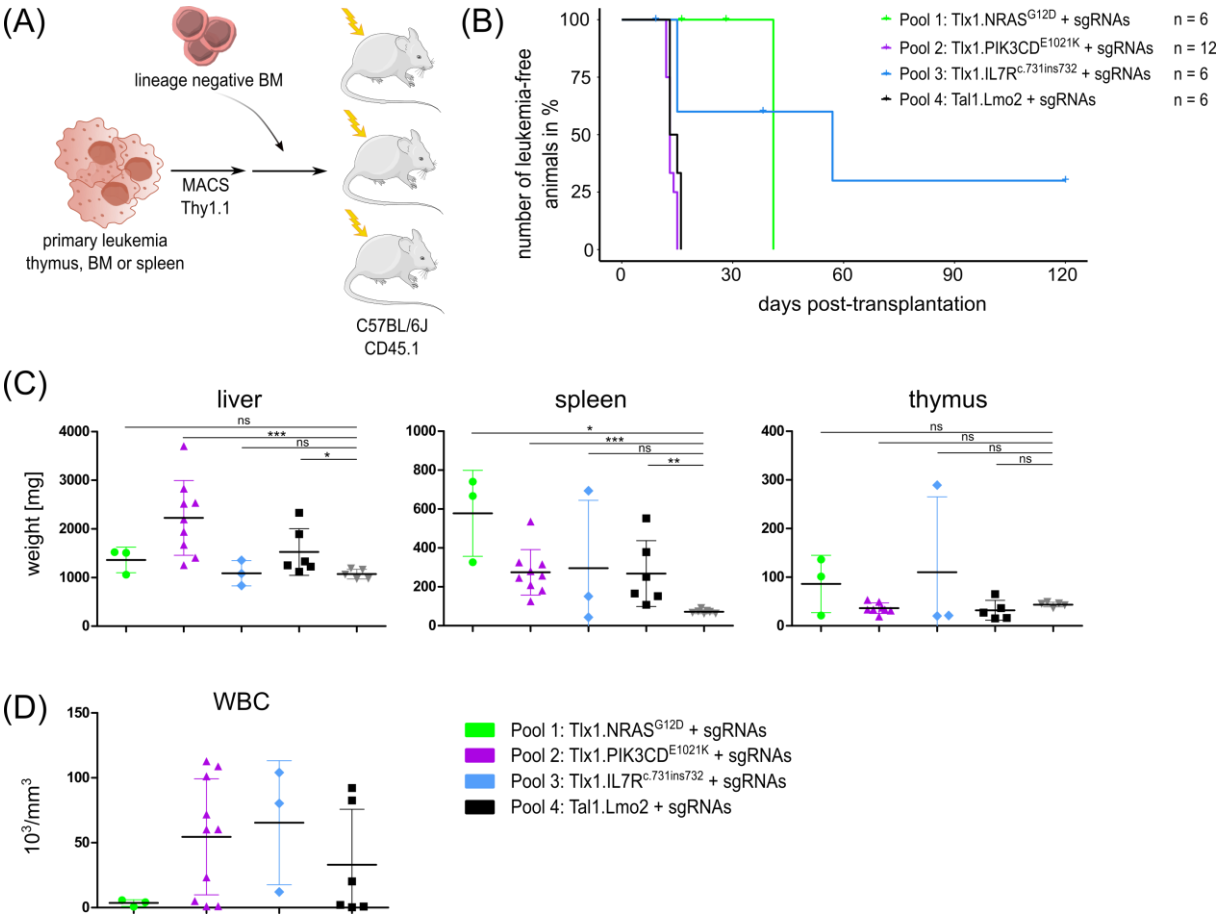


Figure 24: Transplantation of primary T-ALL cells replenished leukemia development in secondary recipients. (A) Primary leukemia cells were purified by MACS from the thymus, spleen, or bone marrow (BM) and transplanted together with helper cells (lineage negative BM cells) into three CD45.1 recipient mice. (B) Kaplan-Meyer curve for secondary transplanted animals with T-ALL from pool 1 to 4. (C) Organ size of liver, spleen, and thymus for secondary diseased animals. Size-matched non-transplanted animals were used as reference for organ size. (D) White-blood count (WBC) of secondary diseased animals assigned to each group. Presented are mean ± SD and individual values. Mann-Whitney U test (* p< 0.05, ** p< 0.01, *** p< 0.001, ns – not significant).

When analyzing the immune phenotype of CD45.2$^+$ secondary leukemia cells, the absence of the myeloid marker CD11b, and the B cell marker B220 for most samples supported the persistence of the T-ALL phenotype, with no previously underrepresented clones growing out from other hematopoietic lineages (Figure 25 A). However, in the three secondary P1-4 T-ALL (SecP1-4 #1-3), B220 expression was increased within the blood in all three recipients. However, this observation was only assigned to the blood cells and did not recapitulate within the BM cells (Supplementary Figure 7).

Phenotypic analysis of CD4, CD8, and CD3 expression highlighted mixed preservation of T-ALL phenotypes upon secondary transplantation. The three secondary transplanted animals

from P1-4 (SecP1-4 #1-3) kept their phenotype but varied within the three recipients in the percentage of $CD8^+$ and DN cells (Figure 25 B). For the secondary transplanted leukemia from pool 2, SecP2-1 #1+2 and SecP2-2 #1 secondary leukemia kept their phenotypic properties, compared to the primary T-ALL (P2-1 and P2-2). For the remaining secondary T-ALL from P2-9 and P2-10, the cell phenotype shifted from DP towards a primarily $CD8^+$ phenotype, with SecP2-10 additionally showing strong CD3 expression, which was absent in the primary T-ALL. Secondary T-ALLs from pool 3 kept their $CD3^{low}$ DP phenotype. A phenotypic conversion was observed for the two chosen primary T-ALL samples from pool 4. While P4-1 expressed CD4 and CD3, secondary T-ALL were characterized as $CD4^-CD8^-$ $CD3^+$ cells. The secondary T-ALL from P4-2 changed from a $CD4^-CD8^-$ $CD3^+$ towards a $CD4^+CD8^+/CD8^+$ and $CD3^+$ phenotype.

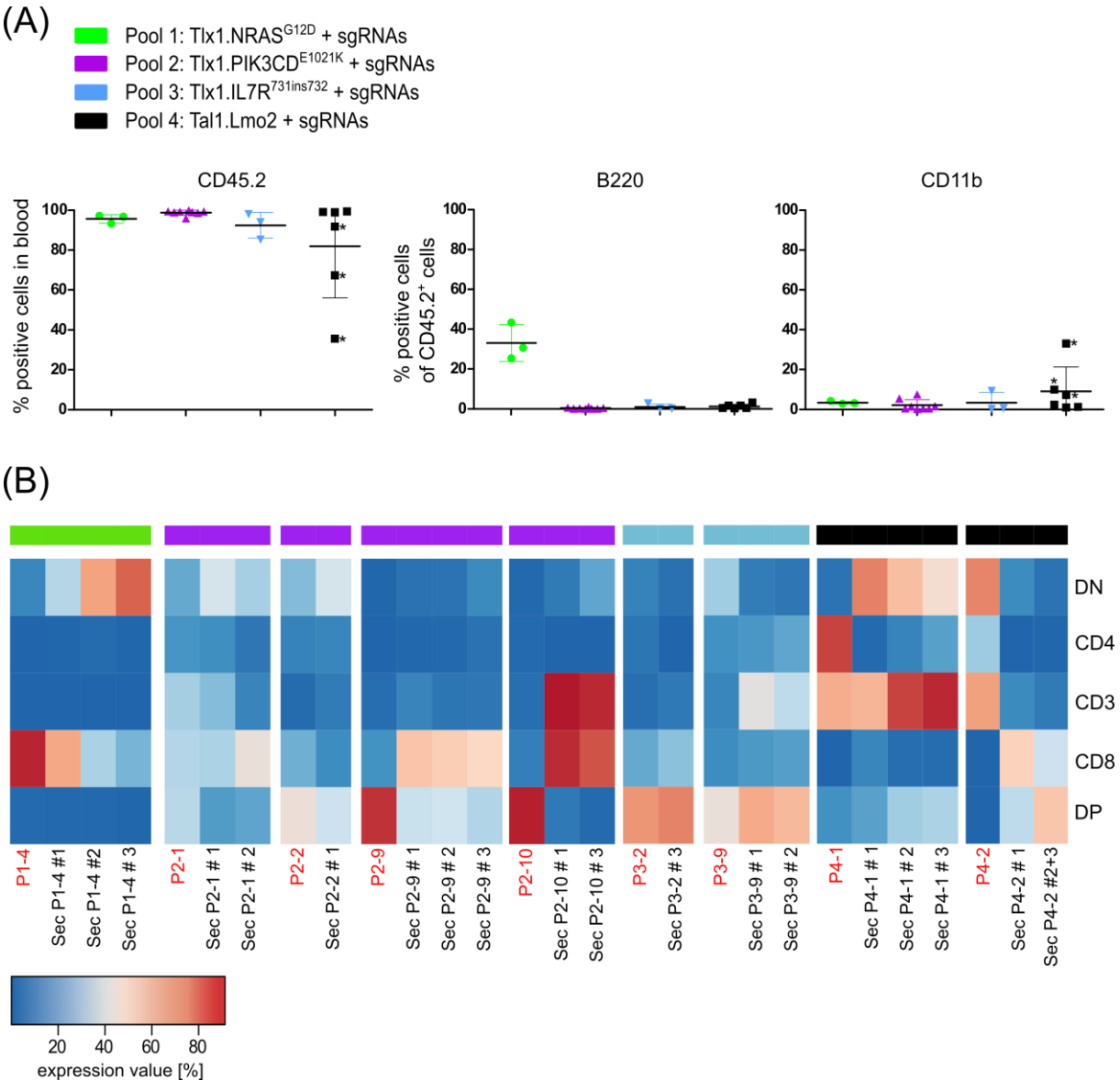


Figure 25: Immune phenotyping T-ALL from secondary transplanted animals. (A) Chimerism of $CD45.2^+$ cells in the blood and fraction of B220 (B cell) and CD11b (myeloid) expressing cells among them. * - values derived from BM cells (B) Individual phenotype of the primary T-ALL (red font) compared

to its secondary occurrence (black font). Expression of mature T cell surface markers $CD3^+$, $CD4^+$, $CD8^+$, $CD4^+CD8^+$ (double-positive, DP) and $CD4^-CD8^-$ (double-negative, DN). Group identity of each animal is color-coded as green (pool 1), purple (pool 2), blue (pool 3) or black (pool 4). Samples SecP4-2 #2+3 were accidentally mixed together but depicted here only for reference.

So far, the focus of the analysis has been on the phenotypic features and the genetic similarities and differences between the different groups to describe the impact of the different type A and type B mutations on the progression of T-ALL development. In another experiment, we were interested in the transcriptomic changes that characterize the different generated biological subgroups of T-ALL.

5.9 Transcriptomic analysis of primary and secondary T-ALL in comparison to healthy thymocytes

In this work, different subgroups of T-ALL were defined based on their morphological disease phenotype in mice, their surface phenotype and genetics. While the genetic characterization of T-ALL samples is the hallmark of classifying distinct biological subgroups[50,88], recent RNA screenings of patient samples highlighted complex transcriptome variations within different T-ALL groups[76]. In the following work, RNA samples from various primary and secondary transplanted animals were selected to investigate transcriptomic changes within the different transplanted groups and compare alterations to healthy and non-transformed thymocyte populations. At the time of this experiment, the underlying genetics of each sample was unknown. Therefore, sample sets were not chosen by similar genetic variations but by those, which incorporated secondary transplantations of primary leukemia and/or corresponding cell lines. From each transplanted group, three primary T-ALL samples were picked. In addition, pool 1 and pool 4 T-ALL were additionally represented by one corresponding secondary leukemia sample and two corresponding cell lines, each. Pool 2 was additionally expanded by three belonging secondary T-ALL and three cell lines each. For pool 3, only one cell line was chosen besides the three primary samples. As a control and reference, the different T cell developmental stages were FACS-sorted from the thymuses of three $CD45.1^+$ C56BL/6J animals. The analysis and evaluation of the cell line samples were summarized in chapter 5.10.

In the first part of the analysis, the different experimental groups were compared. Therefore, a principal component analysis (PCA) was performed to identify the main variance between the different samples and to spot outliers. With few exceptions, leukemia from pool 1, 2, and 4 transplanted animals clustered within their associated samples and separated from the other experimental groups (Figure 26 A). The T-ALL samples from pool 3 were mixed with pool 1 and pool 4 leukemia. In a first analysis, differentially expressed (DE) genes were calculated between the primary T-ALL of each experimental group. Almost no DE genes were examined

between the primary T-ALL samples based on the transplanted pools. Only 5 DE genes were found between pool 1 and pool 4, whereas 7 genes were deregulated between pool 2 and pool 4 (data not shown). Therefore, the primary and secondary T-ALL samples were analyzed together. Thus, three different clusters could be distinguished in the principal component analysis (Figure 26 A, red dotted line). Cluster 1 included P3-2 and all primary and secondary samples from pool 1. The second cluster included SecP4-2 and all samples from pool 2. The remaining pool 3 and pool 4 samples formed the third cluster. Strikingly, most primary T-ALL samples clustered close to their corresponding secondary T-ALL, as seen in PCA and hierarchical clustering analysis in Figure 26 B. However, for SecP2-1 and SecP4-2 samples, the transcriptome of the secondary leukemia differed compared to the primary T-ALL. In both secondary leukemia, most genes were less expressed compared to their primary T-ALL. However, transcripts involved in the Ras and MAPK pathway were upregulated in SecP2-1, and genes involved in Wnt and mTOR signaling pathway were upregulated in SecP4-2 (Supplementary Figure 8). However, these calculations were based solely on the arithmetic

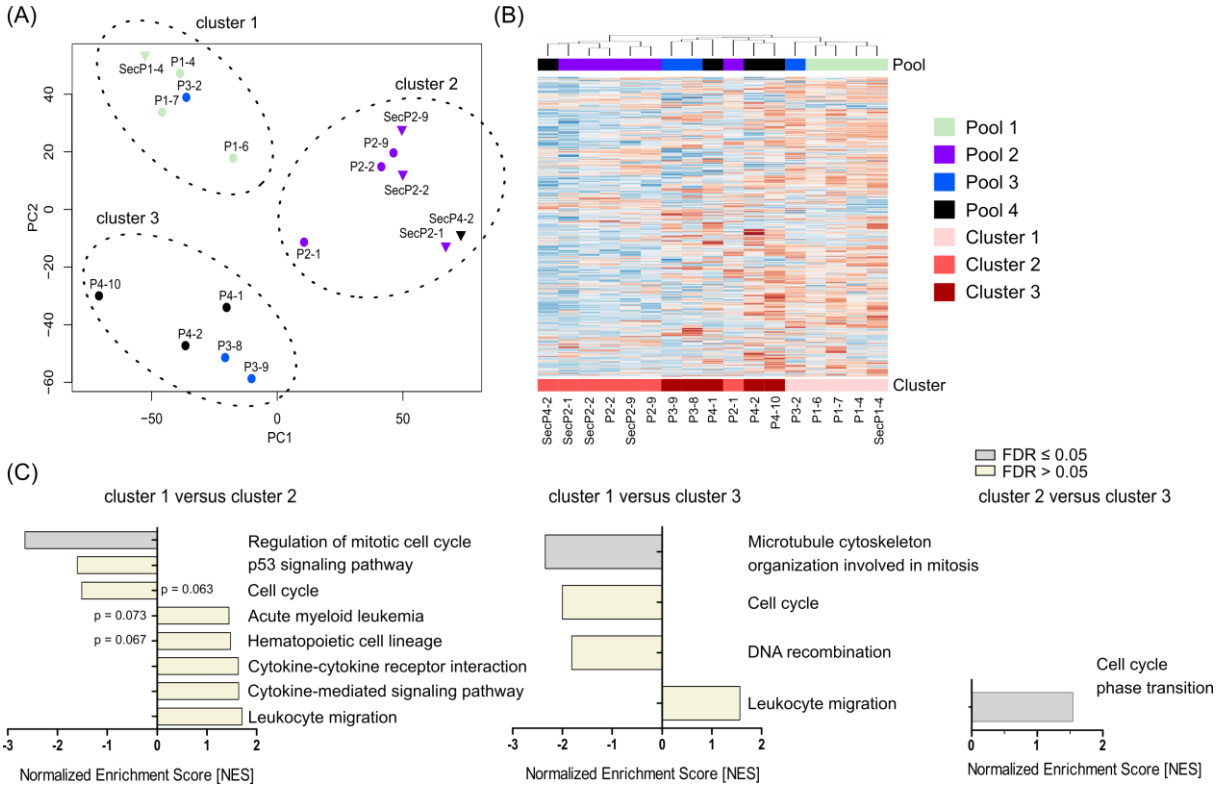


Figure 26: Murine T-ALL samples were classified into functional clusters. (A) Principal component (PC) analysis (PCA) and hierarchical clustering (B) were performed for all primary and secondary T-ALL. By this, three different clusters were identified, which were subsequently used for the calculation of differently expressed genes. Presented are z-score values with the color legends specified in (D). (C) Selection of significantly enriched KEGG pathways and gene ontology sets of biological processes that were identified by gene set enrichment analysis from https://www.webgestalt.org/ between the three defined clusters. Biologically matching pathways were selected for $p \leq 0.05$ (if not depicted differently) and marked based on their false discovery rate (FDR) for > 0.05 and ≤ 0.05. (D).

difference of the normalized and log_2 – logarithmized raw values. Due to the low sample size of these outliers, statistics and further quantification were not conducted.

Comparing the different clusters revealed 900 DE genes between clusters 1 and 2, 759 DE genes between clusters 1 and 3, and 460 DE genes between clusters 2 and 3. Deregulated gene lists were screened for Kyoto Encyclopedia of Genes and Genomes (KEGG) – pathways and gene ontology of biological processes, using the web-based gene set enrichment analysis (GSEA) tool https://www.webgestalt.org/. T-ALL samples from cluster 1 displayed decreased cell cycling and p53 pathway activity compared to the second cluster. Simultaneously, gene sets characterizing AML ($p = 0.073$) and hematopoietic cell lineage ($p = 0.067$) were enriched along with cytokine signaling and leukocyte migration (Figure 26 C). Alterations of cluster 1 T-ALL versus cluster 3 highlighted a comparatively decreased cell cycle activity and an enhanced leukocyte migration for cluster 1 samples. The highest cell cycle activity was assigned to cluster 2 T-ALL samples, which further stood out in comparison to cluster 3.

After comparing the different T-ALL samples, the next question was to which stage of T cell development the different leukemias could be assigned. Therefore, T-ALL samples were compared to the different T cell developmental stages within healthy $CD45.1^+$ C56BL/6J thymuses. During PCA, three clusters were observed for all measured thymus and T-ALL samples (Figure 27 A). One cluster contained only cells from the DN1 stage. All other developmental populations, including cells at the DN2 – 4, DP, CD4, and CD8 stages, could be separated into another cluster. Due to technical issues, the CD8 single-positive population contained only two replicates. The third cluster compromised all primary and secondary T-ALL samples. By performing hierarchical clustering and heatmap visualization of all transcribed genes, the separation of the three different clusters and their transcriptomic variance was underlined (Figure 27 B). The DN2 and DN3 stages clustered together within the thymus samples, similar as DN4 and DP, and the CD4 and CD8 stages, following the hierarchy of T cell development. In the following examination, gene sets were defined from the thymus samples, encompassing differentially expressed genes specific to the belonging T cell stage. These gene sets were used to perform single-sample gene set enrichment analysis (ssGSEA) on each T-ALL sample to arrange different leukemias to a particular T cell stage (Figure 26 C). The pool 1 samples, P1-6 and P1-7 were by trend enriched for DN1, DN2, and CD8 transcribed genes, whereas P1-4 and the belonging secondary transplanted SecP1-4 sample had no clear assignment to one particular stage. Downregulation of any T cell stage-specific gene set was observed in the second experimental group, besides a minor enrichment for DN3 and DP stages in P2-9 and SecP2-9. The same downregulation or absence of enriched transcripts was true for P3-9, P4-1, and SecP4-2 samples. In the third group, more enrichment was observed to a particular T cell stage. P3-2 sample was marked by upregulation of DP and CD8-specific

genes and minor enrichment for each of the other gene sets. The P3-8 sample mainly correlated with a DN4- and CD4-specific gene expression profile. Within the fourth group, enrichment for any gene set was negligible in the P4-1 and SecP4-2 samples. In contrast, P4-2 was affiliated with DP gene expression, while P4-10 could be allocated to a CD8 and DN1 expression profile. Based on these enrichment scores, a correlation to the measured surface phenotype was difficult because the depicted heatmap represented a row-Z score, only allowing the comparison within one gene set and not between ones. Nevertheless, for those samples with higher enrichment within one particular stage, like in P1-7 (DN), P3-2 (DP), P3-8 (CD4) and P4-10 (CD4, CD8, DN), a correlation with the results from flow cytometry analysis (Figure 18) was observed. However, this was not entirely true for all.

In the next part, deregulated genes were assigned to generic cellular pathways to understand which processes distinguished the leukemia cells from healthy tissues. Because of the enormous differences between DN1 cells and the other thymus samples, DE genes were seperately calculated to thymus samples without DN1 cells (Figure 27 D) and to DN1 cells alone (Figure 27 E) in comparison to T-ALL cells. Although 1630 DE genes with a $\log_2$-fold change (lfc) of > 1 were measured between the non-DN1 thymic samples and the entirety of primary T-ALL samples, only a few biologically matching and significantly deregulated KEGG-pathways and gene ontology sets could be assigned to them. The Notch signaling pathway, myeloid cell differentiation, and cell activation in immune responses were upregulated in the T-ALL samples. In contrast, DNA modification and the forkhead box O (FoxO) signaling pathway were downregulated (Figure 26 D). Within the DN1 cell cluster, 3003 genes were differently expressed than the T-ALL cluster (lfc > 1), indicating a greater distance of obtained T-ALL from this developmental stage. T-ALL cells displayed increased DNA repair, cell cycle, and Notch signaling pathway activity ($p = 0.0589$) compared to DN1 thymocytes. At the same time, endocytosis, leukocyte immunity, cytokine secretion and receptor interaction, and antigen processing were significantly downregulated, suggesting an immunogenic inactive state for the leukemia samples (Figure 27 E).

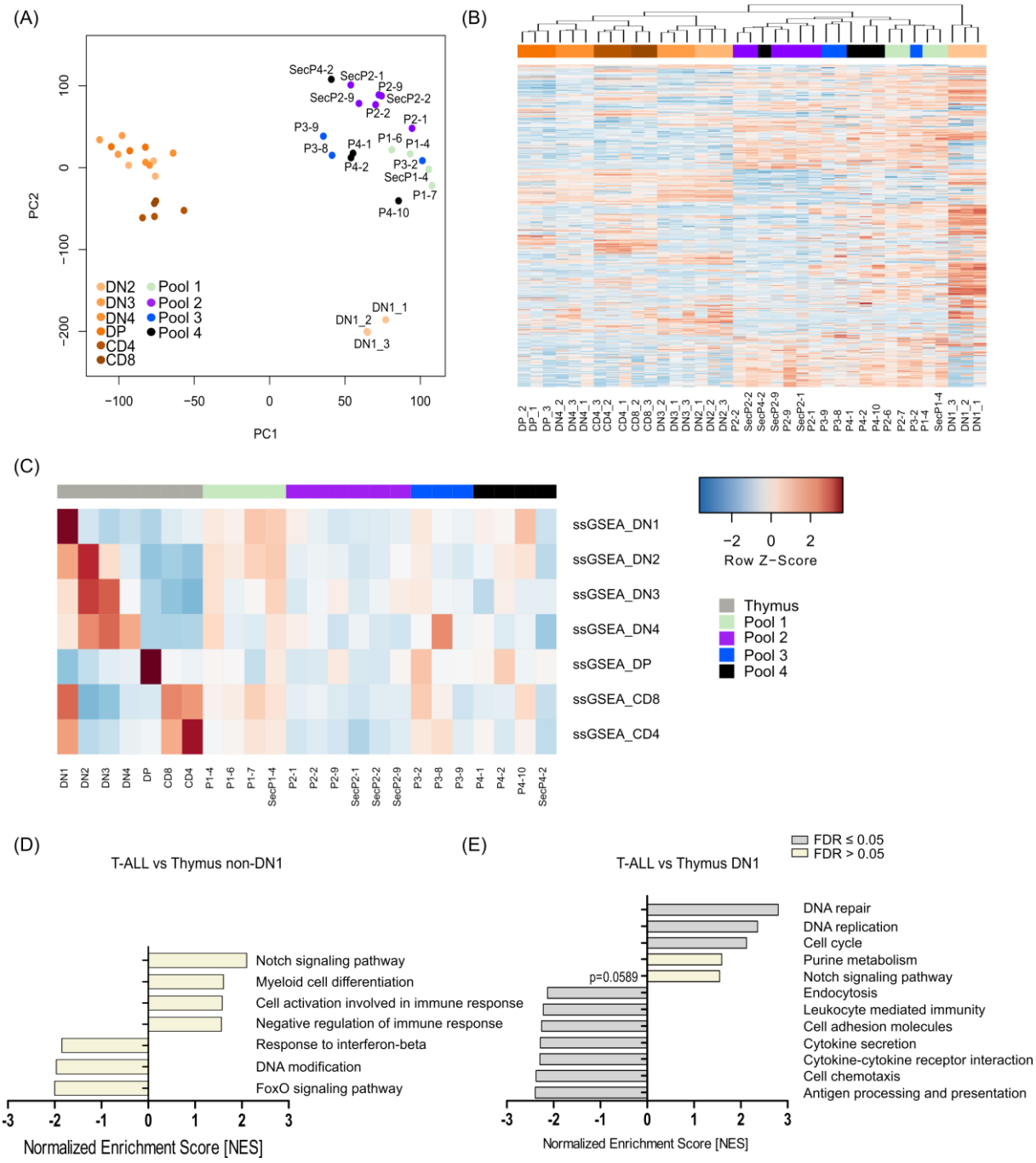


Figure 27: RNA sequencing of T-ALL samples revealed distinct transcriptional activity compared to healthy thymocytes. (A) Principal component (PC) analysis (PCA) was performed on CD45.1$^+$ C57BL/6J thymocytes at different stages of T cell development and chosen primary and secondary T-ALL. (B) Heatmap visualized hierarchical clustering of all expressed genes at an interquartile range of > 2 on normalized and logarithmized gene counts. Color legends were specified in (A). (C) Unique and top deregulated genes were summarized for DN1, DN2, DN3, DN4, DP, CD8 and CD4 thymocytes as gene sets. Single-sample gene set enrichment analysis (ssGSEA) was conducted with the R package "GSVA" to calculate enrichment scores for each sample set. Heatmap on one thymus set and the T-ALL samples visualized positive (reddish) or negative (bluish) gene expression enrichment for a specific thymus subtype. (D+E) Selection of significant KEGG pathways and gene ontology sets on biological processes between the T-ALL cells and the thymus group, without the DN1 set (D) and for only DN1 cells (E) using the web tool https://www.webgestalt.org/. Biologically matching pathways were selected for $p \leq 0.05$ (if not depicted differently) and marked based on their false discovery rate (FDR) for > 0.05 and ≤ 0.05. DN – double-negative ($CD4^-CD8^-$), DP – double-positive ($CD4^+CD8^+$), KEGG – Kyoto Encyclopedia of Genes and Genomes. The examination and calculation of deregulated genes between

the thymus clusters and the T-ALL samples were performed in cooperation with Dr. Anton Selich. The defined thymus gene lists can be find on an attached hard drive.

Next, individual gene expression was investigated in more details for certain candidates (Figure 28). These genes included *Notch1*, *Trp53*, *Il2ra* (*Cd25*), *Il7r* and *Cd44*, and have been associated with T cell development and their deregulation during transformation[73,185,192,194]. Except for P4-1, all T-ALL samples showed high *Notch1* and *Il2ra* expression and moderate levels of Il7r. *Cd44* was expressed in all T-ALL samples, except for P3-8 and P3-9. A similar expression pattern was dedicated to DN2 and DN3 T cells. Surprisingly, even though cell cycling was significantly increased in the T-ALL samples, no elevated *Trp53* expression was detected in comparison to the thymus samples.

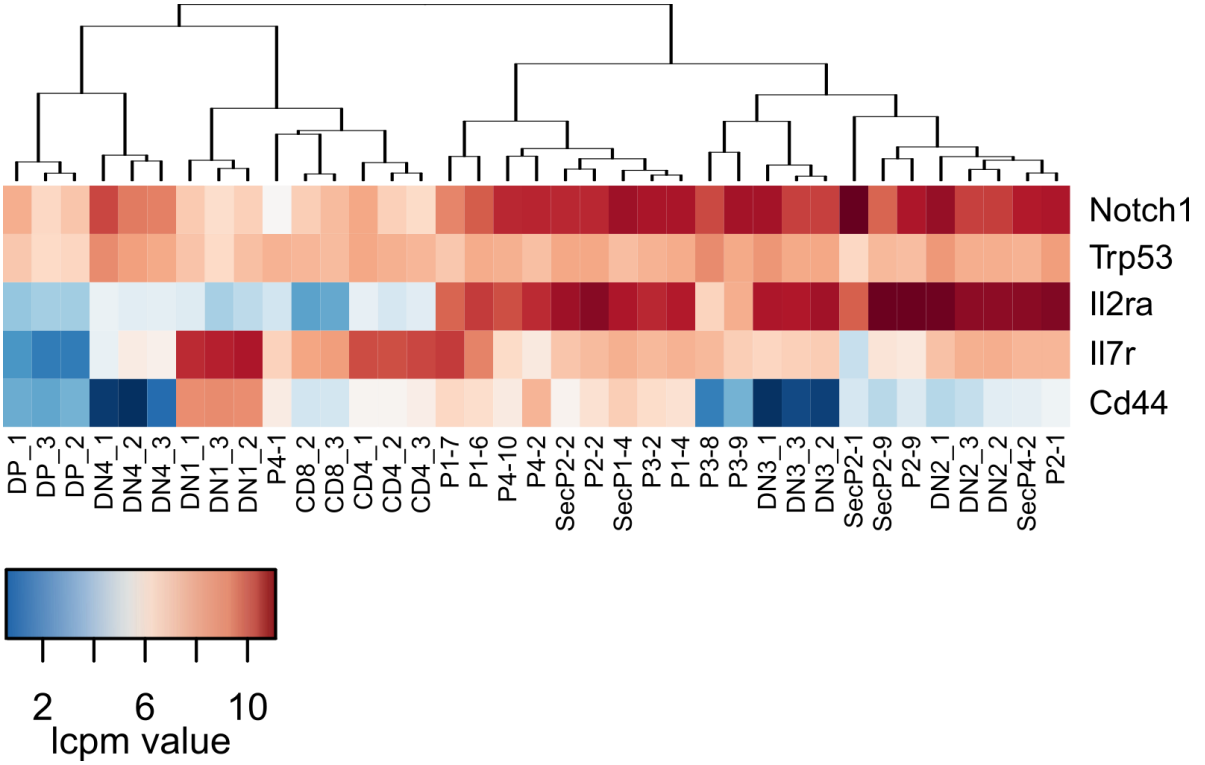


Figure 28: *Notch1* and *Il2ra* are highly upregulated in T-ALL samples. Hierarchical cluster analysis and heatmap visualization for the gene expression of *Notch1*, *Trp53*, *Il2ra*, *Il7r* and *Cd44* in primary and secondary T-ALL, and within the different developmental stages of T cell differentiation isolated from the thymus. Depicted are the normalized and logarithmized counts per million reads mapped (lcpm). DP – double positive ($CD4^+CD8^+$), DN – double negative ($CD4^-CD8^-$).

In summary, by using the *LSL.Cas9 x Lck-cre* mouse model, it was possible to generate different subtypes of T-ALL. The obtained subtypes reflected different stages of T cell development, as shown by the differential expression of CD3, CD4, and CD8. Genetically complex aberrations in Cas9-targeted genes were detected in all T-ALL samples. Depending

on the subtype, certain mutations, e.g., in *Notch1*, *Cdkn2a*, *Pten*, or *Etv6*, were conserved within a group, whereas others occurred more randomly. The major phenotypic and genotypic components were largely conserved within a transplanted group. Transcriptomic analyses of primary and secondary T-ALL samples revealed consistent upregulation of Notch signaling and cell cycle, with less overall variation within groups. Next, cell lines were generated from the obtained T-ALL samples to make the T-ALL model system more suitable as a potential platform for drug screenings and to further investigate the underlying genetics.

5.10 Cell lines derived from primary or secondary leukemia

After the characterization of primary and secondary T-ALL, thawed cells were transferred into *in vitro* culture to generate cell lines from the different subtypes. Leukemia cells were enriched by MACS for THY1.1$^+$ cells either from primary or secondary transplanted animals. Various groups have shown that co-cultivation with BM stromal cells increased the survival of T-ALL and B-ALL cells for the cultivation of primary ALL, respectively[195–197]. In line with the used T cell differentiation protocols (chapter 5.3) [38–40], OP9-DL1 cells were chosen as feeder cells for which adequate support for developing T cells was already shown (Figure 10). T-ALL cells were cultivated in RPMI medium supplemented with 5 – 10 ng/mL mIL2 and mIL7 and passaged twice weekly (Figure 29 A). Cytokine conditions were selected from Rhein *et al*., who described the generation of a murine T-ALL cell line derived from the transplantation of retrovirally modified BM cells[104]. Indeed, under these conditions, T-ALL cells from SecP2-1 and SecP2-2 animals manifested a stable proliferation and were maintained for several weeks up to months. However, the immune phenotype was lost for both cell lines over time. The primordial leukemia cells possessed a mixed phenotype of DP, CD8$^+$, CD4$^+$, and DN cells with partial expression of CD3. However, within a couple of weeks, the co-cultivated cSecP2-1 and cSecP2-2 (“c” for cell line from SecP2-1 and SecP2-2) cell lines lost most of their CD8 and CD4 expression, leading to a predominant DN phenotype. In parallel, CD3 expression drastically increased (shown for cSecP2-1 in Figure 29 B).

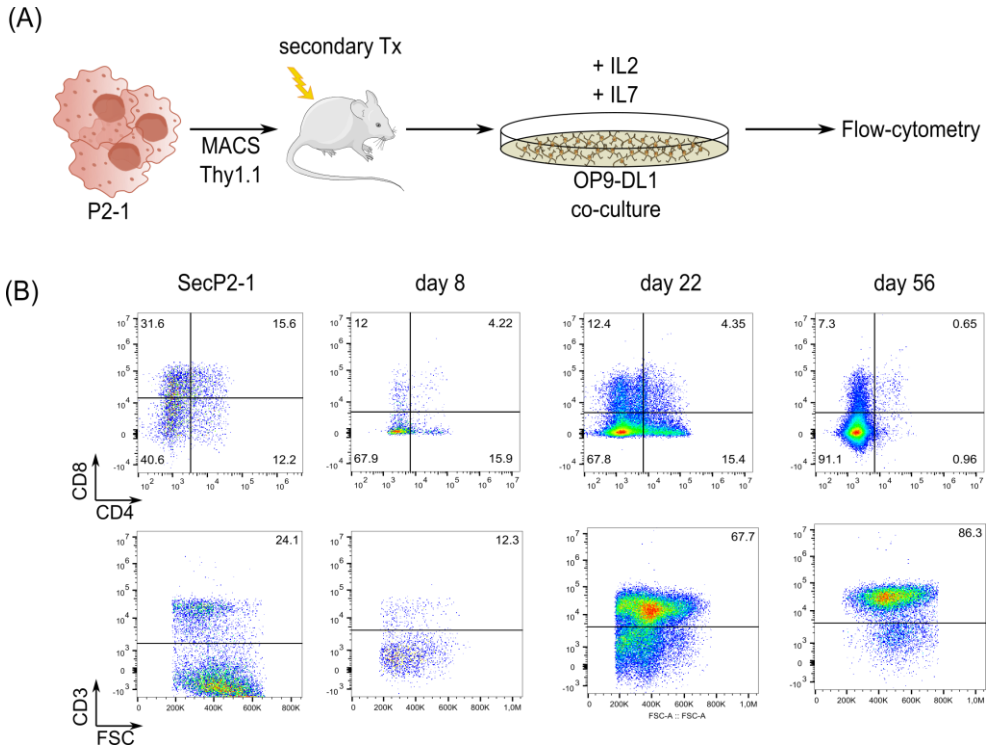


Figure 29: Immune phenotypes of developed T-ALL cell lines. (A) P2-1 T-ALL cells were enriched from the spleen and transplanted into secondary recipients. Gained secondary T-ALL cells were seeded onto pre-seeded OP9-DL1 cells. Cultivation was performed in RPMI medium, supplemented with 5 ng/mL mIL2 and mIL7. (B) Flow cytometry analysis of cell phenotype at different time points after initial seeding. Expression of the mature T cell surface markers CD8, CD4, and CD3.

It was assumed that Notch signaling from the co-cultured OP9-DL1 cells gave a stimulatory growth signal to the more immature DN cells, thereby outgrowing the other phenotypes or changing the physiology of the cells. Thus, co-cultivation conditions were changed to OP9 WT cells without an artificial Notch ligand for the mature pool 2 – 4 T-ALL samples, whereas immature pool 1 T-ALL were still cultivated with OP9-DL1 cells (Figure 30 A). With this approach, it was possible to establish several cell lines from all transplanted groups (Figure 30 B). Of note, mIL2 and mIL7 were the only tested cytokines within this work. The concentration compromised 5 ng/mL or 10 ng/mL with no striking differences (data not shown). For pool 1, the four different cell lines cP1-6, cP1-7, cP1-8 and cP1-9, respectively derived from primary P1-6, P1-7, P1-8 and P1-9 T-ALL samples, were established. Each relied on co-cultivation with OP9-DL1 cells and the constant supply of mIL2 and mIL7. The cell lines cP1-6, cP1-7 and cP1-9 were primarily negative for CD4 and CD8, while cP1-8 possessed a $CD8^+$ and $CD4^+CD8^+$ phenotype. All of these cell lines lacked CD3 expression. Besides the already described cSecP2-1 and cSecP2-2 cell lines from pool 2, three more cell lines were generated on OP9 WT cells from that group. The cell line cP2-9 was obtained from a primary T-ALL, while the other two cell lines cSecP2-9 #1 and cSecP2-9 #2 were cultivated from secondary transplanted animals. All of these three animals shared the same original T-ALL sample – P2-

9. When comparing their phenotype, the cP2-9 cell line could be distinguished by its mixed $CD4^{-}CD8^{-}$ and $CD8^{+}$ phenotype, while the cell lines from the secondary T-ALL had a predominantly $CD8^{+}$ phenotype. All of these cell lines lacked surface CD3 expression. Two cell lines were generated for the third transplanted group (pool 3). Both cell lines, cP3-8 and cP3-9, were mostly DN but partly expressed CD8 and CD3 on their cell membranes. All the described cell lines from pool 2 and pool 3 relied on co-cultivation with OP9 WT cells and the constant supply of mIL2 and mIL7. Only cP2-9 achieved stable proliferation in OP9-free suspension culture. The most cell lines were obtained from T-ALL samples dedicated to pool 4. The cell lines cP4-7 and cP4-10, derived from primary transplanted P4-7 and P4-10 T-ALL cells, were cultivated, after initial seeding with 5 ng/mL mIL2 and mIL7, in the absence of these cytokines in co-culture with OP9 cells. For comparison, an early passage of cP4-7 was constantly supplemented with 5 ng/mL mIL2 and mIL7 in parallel in co-culture. Interestingly, the cell lines cP4-7 (+ cytokines) and cP4-10 both lost CD8 expression, present in the primary T-ALL, and displayed a DN phenotype. However, cP4-7, in absence of cytokines, preserved the primary $CD8^{+}$ and low $CD3^{+}$ phenotype. Two other cell lines were obtained from the two secondary transplanted SecP4-1 #1 and #2 animals. After initial co-cultivation, both cell lines stably proliferated in suspension without OP9 co-culture. Thereby, the cSecP4-1 #1 became independent of supplied cytokines in the RPMI medium, whereas the cSecP4-1 #2 cell line kept its dependency. Both expressed high CD3 and showed a $CD8^{+}$ and $CD4^{+}CD8^{+}$ phenotype, indicating the reestablishment of the more mature phenotype of P4-1 that was lost during secondary transplantation. Another three cell lines were derived from three secondary recipients that were transplanted with P4-2 T-ALL cells. The three cell lines cSecP4-2 #1 – 3, depended on the co-cultivation with OP9 and the presence of 5 ng/mL mIL2 and mIL7. All three cell lines depicted a $CD8^{+}$ and low $CD3^{+}$ phenotype, preserving the phenotype of the respective secondary T-ALL.

For most cell lines, the cell surface phenotype was tracked for 1 – 5 months (Supplementary Figure 11). Over time, the phenotypes did not change substantially in their compositions. For the cell lines cP1-8, cSecP2-9 #2 and cSecP4-1 #2, no such tracking was performed. Besides CD3, CD4, and CD8 surface expression, the expression of CD25 and CD44 within $CD4^{-}CD8^{-}$ cells was also measured, describing phenotypic characteristics of early T cell stages. For the cell lines cP1-6, cP1-7, cP1-9, and cP2-9, high expression of CD25 and CD44 indicated a DN2 phenotype of these cells. No significant changes were observed for the cell lines cP4-2 #1 – 3, while in cSecP4-1 #1 the ratio of $CD4^{+}CD8^{+}$ cells decreased, and $CD8^{+}$ cells increased in percentage. More differences were observed for the cell lines cP4-7 and cP4-10. In the presence of cytokines, cP4-7 lost its CD4 expression and developed a DN3 phenotype, indicated by a $CD25^{+}CD44^{-}$ surface marker profile. In addition, CD3 expression increased over time. However, the $CD8^{+}CD3^{low}$ phenotype was more preserved in the absence of cytokines.

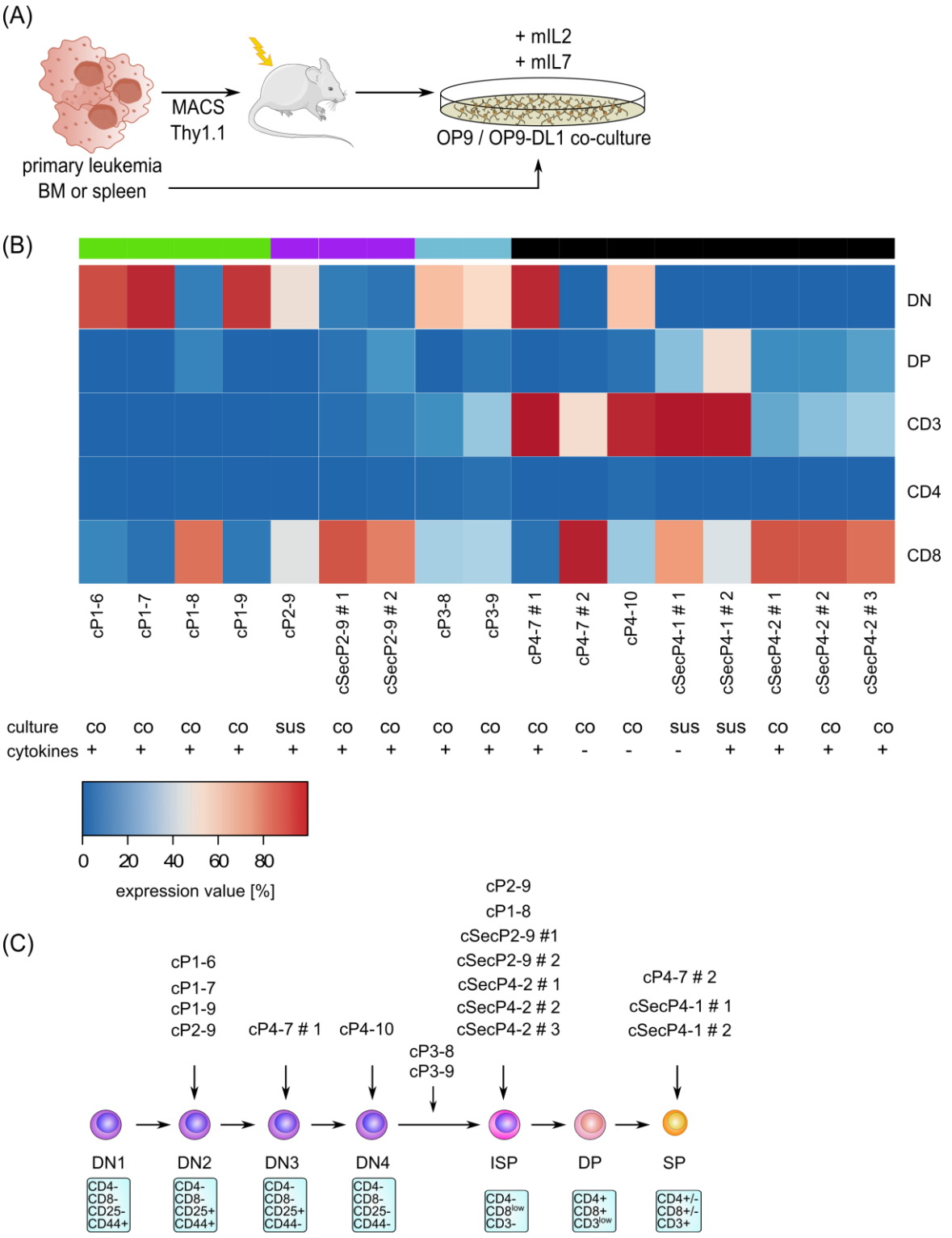


Figure 30: Immune phenotype of developed T-ALL cell lines. (A) T-ALL cells from bone marrow (BM) or the spleen of primary transplanted animals or after secondary transplantation were seeded onto pre-seeded OP9-DL1 (pool 1) or OP9 (pool 2 – 4) cells. Cultivation was performed in RPMI medium, supplemented with 5 ng/mL mIL2 and mIL7. (B) Individual phenotype of each obtained cell line. Expression of the mature T cell surface markers $CD3^+$, $CD4^+$, $CD8^+$, $CD4^+CD8^+$ (double-positive, DP) and $CD4^-CD8^-$ (double-negative, DN). Group identity of each animal is color-coded as green (pool 1), purple (pool 2), blue (pool 3) or black (pool 4). Co-culture (co) or suspension culture (sus) in the presence (+) or absence (-) of 5 ng/mL mIL2 and mIL7 indicate final cell culture conditions in which a stable proliferation was reached. (C) Schematic overview of T cell development and assigned cell lines based on their immune phenotype. ISP – intermediate single positive, SP – single positive.

For the cP4-10 cell line, a switch from $CD8^+CD3^+$ towards $CD4^-CD8^-CD25^-CD4^-CD3^+$ was striking. High CD25 (IL2Rα) expression was commonly observed among all cell lines that depended on supplied cytokines. Cytokine-independent proliferating cells, like cP4-7, and cP4-10, but not cSecP4-1 #1, downregulated CD25/IL2Rα expression over time (Supplementary Figure 11).

In the next experiment, we wanted to elucidate whether the obtained cell lines could genetically still be assigned to the primary leukemia samples or whether cultivation resulted in the outgrowth of a different clone. Therefore, the presence of the inverse vector construct and the existence and distribution of the type B mutations were confirmed and compared to the leukemia of origin. First, the already-used PCR approach was used to determine the presence and orientation of the inverse vector. Herein, one forward primer was bound within the promoter sequence (upstream of the lox61 site) and two reverse primers aligned to the reading frame of the *dTomato* or *Tlx1*/*Tal1* sequence (Figure 31 A). For the *PIK3CD*^{E1021K} vector in pool 4, another PCR approach was used to prevent misleading signals from the Tal1.Lmo2 vector. Instead of the reverse primer within the *dTomato* sequence, one primer was bound to the WPRE element (like in Figure 19 A). Inverted *Tlx1.NRas*^{G12D}, thereby, in its active form was detected in each cell line of pool 2 (Figure 31 B). No *Tlx1.PIK3CD*^{E1021K} sequence was found in any cell line of pool 2, besides a dim non-inverted signal in cSecP2-9 #1. Still, no *Tlx1.IL7R*$^{c.731ins732}$ sequence was detected in both cell lines from pool 3, even though a small band appeared for this sample that could not be allocated. Within pool 4, the cell lines cSecP4-1 #1+2 and cP4-10 contained the *Tal1.Lmo2* vector in its active orientation, while cSecP4-2 #2+3 contained the vector in the non-inverted orientation. This orientation was not detected in the primary P4-2 sample, although its underrepresentation in primary leukemia was definitely possible. No *Tal1.Lmo2* encoding vector was detected in cP4-7, and also not in cSecP4-2 #1. The same was true for the *PIK3CD*^{E1021K} vector, which was not detected in any cell line of pool 4. For unknown reasons, a positive signal appeared in non-Cre-transduced cells of the *PIK3CD*^{E1021K} vector control. However, its absence in any other T-ALL cell line excluded contamination or off-targeting of the PCR reaction.

In addition to the type A vector inversion, each gene locus that was primarily targeted by a sgRNA was PCR-amplified, sequenced and examined by using Tracking of Indels by Decomposition (TIDE) analysis (http://shinyapps.datacurators.nl/tide/)[157]. Subsequently, the Sanger sequencing results of the bulk PCR products were decomposed with TIDE to identify sequence variations or indels. As a reference, the same PCR amplicons were generated from the murine 32D and BAF3 cell lines and sequenced subsequently. For all obtained cell lines, indels were detected in some of the CRISPR-Cas9-targeted genes. However, only the size of

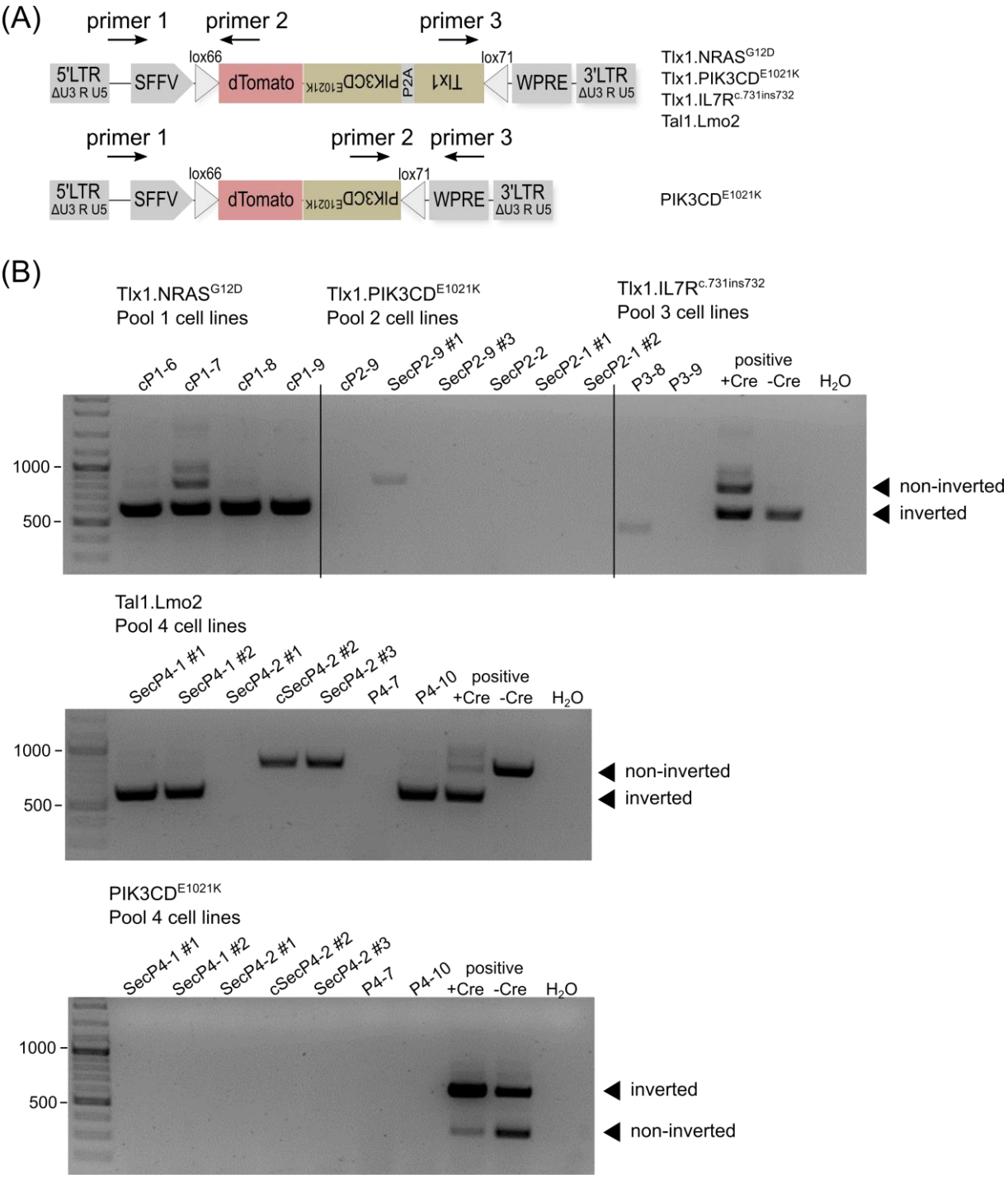


Figure 31: Presence and orientation of the Type-A vector constructs within obtained T-ALL cell lines. (A) A PCR was designed to capture the orientation of the inverse vector construct. Depending on the Cre-mediated orientation of the *dTomato*-oncogene cassette, different PCR products were generated. Exemplary design for the *Tlx1.PIK3CD^{E1021K}* coding inverse vector and the *PIK3CD^{E1021K}* encoding vector were depicted. (B) Presence of differently orientated inverse vector constructs within each cell line.

the indel was reported and not its specific sequence. Therefore, variations with the same indel size couldn't be distinguished. The absence of a WT peak (± 0 bp) was considered as a homozygous indel variation when only one indel was detected. Next, the results were compared to the NGS indel analysis of the primary T-ALL samples (Figure 20). For the four

cell lines cP1-6, cP1-7, cP1-8 and cP1-9 from the primary pool 1, the genetic variations were quite similar to those of the primary T-ALL (Figure 32 A). Minor differences were observed for cP1-8 with two out-of-frame indels for *Cdkn2a* and one in-frame indel for *Lef1* and *Etv6*, which were differently affected in the primary T-ALL. Surprisingly, for cP1-6, a Phf6 out-of-frame indel was detected, which was absent in the primary P1-6 sample. For the cell lines from pool 2, comparable observations were achieved (Figure 32 B). In general, most aberrations were pretty similar to their T-ALL of origin. Nevertheless, minor variances were observed, for example, in the cSecP2-1 #1 cell line with more than 2 different indels within the targeted *Etv6* locus. The same appeared for the *Notch1* locus from cP2-9 and cSecP2-9 #2, which was less heterogeneous within the original P2-9 sample. Concerning these cell lines, the inverse vector, encoding *Tlx1.PIK3CD*^{E1021K}, was no longer detected in P2-9-derived cell lines (Figure 30 B). Most changes occurred within the cSecP2-2 #1 cell line. Here, no genetic variations could be detected within *Cdkn2a* and *Lef1*, both mutated in the primary T-ALL. Instead, more different indels were detected in the *Bcl11b* locus. In the third group, minor changes in the detected indel sizes were observed for cP3-8, whereas the genetic alterations in cP3-9 and P3-9 were similar (Figure 32 C). In the fourth group (Figure 32 D), the cell line cSecP-1 #2 acquired a homozygous in-frame insertion of 7 bp in the *Cdkn2a* locus, which was absent in the primary P4-1 T-ALL. In the cP4-7 cell line the two indels in *Cdkn2a* from P4-7 were no longer detected. Besides these differences, all cell lines from pool 4 contained similar indels within *Notch1*, *Cdkn2a*, and *Pten* as in their primary T-ALL.

In addition to the phenotypic and genotypic characterization, the influence of the *in vitro* cultivation on the cell line's transcriptome was investigated. RNA was collected from chosen samples, from which the primary and/or secondary T-ALL were analyzed before (see chapter 5.9, Figure 26). In the principal component analysis of thymus samples, primary and secondary T-ALL, and T-ALL cell lines, the cell lines clustered as a distinct group apart from their primary cell source (Figure 33 A). Heatmap-visualization indicated major transcriptomic changes compared to the primary and secondary T-ALL samples (Figure 33 B). Therefore, transcriptomic differences were calculated between all cell lines together and the whole set of primary and secondary T-ALL. Due to the *in vitro* cultivation, cell lines increased mTOR and PI3K-AKT signaling pathways on trend (Figure 33 C). Cancer-associated pathways, as well as, cancer-associated proteoglycans and microRNAs that were defined by KEGG tended to be enriched within the cell lines. At the same time, cells became less mature, highlighted by downregulation of lymphocyte differentiation, immune response activation and antigen processing. Surprisingly, the hematopoietic cell lineage genes were downregulated by trend compared to the primary and secondary T-ALL. These transcriptomic differences were true for the whole set of cell lines, compared to all primary and secondary T-ALL. However, when comparing changes between the different cell lines themselves, less variance was observed.

Indeed, although two clusters became apparent within the heatmap visualization, (Figure 33 B) and PCA (Figure 33 D) no significant DE genes were calculated between these groups (cSecP4-1, cSecP4-2, cP1-7, and cSecP2-1 versus the others) (data not shown).

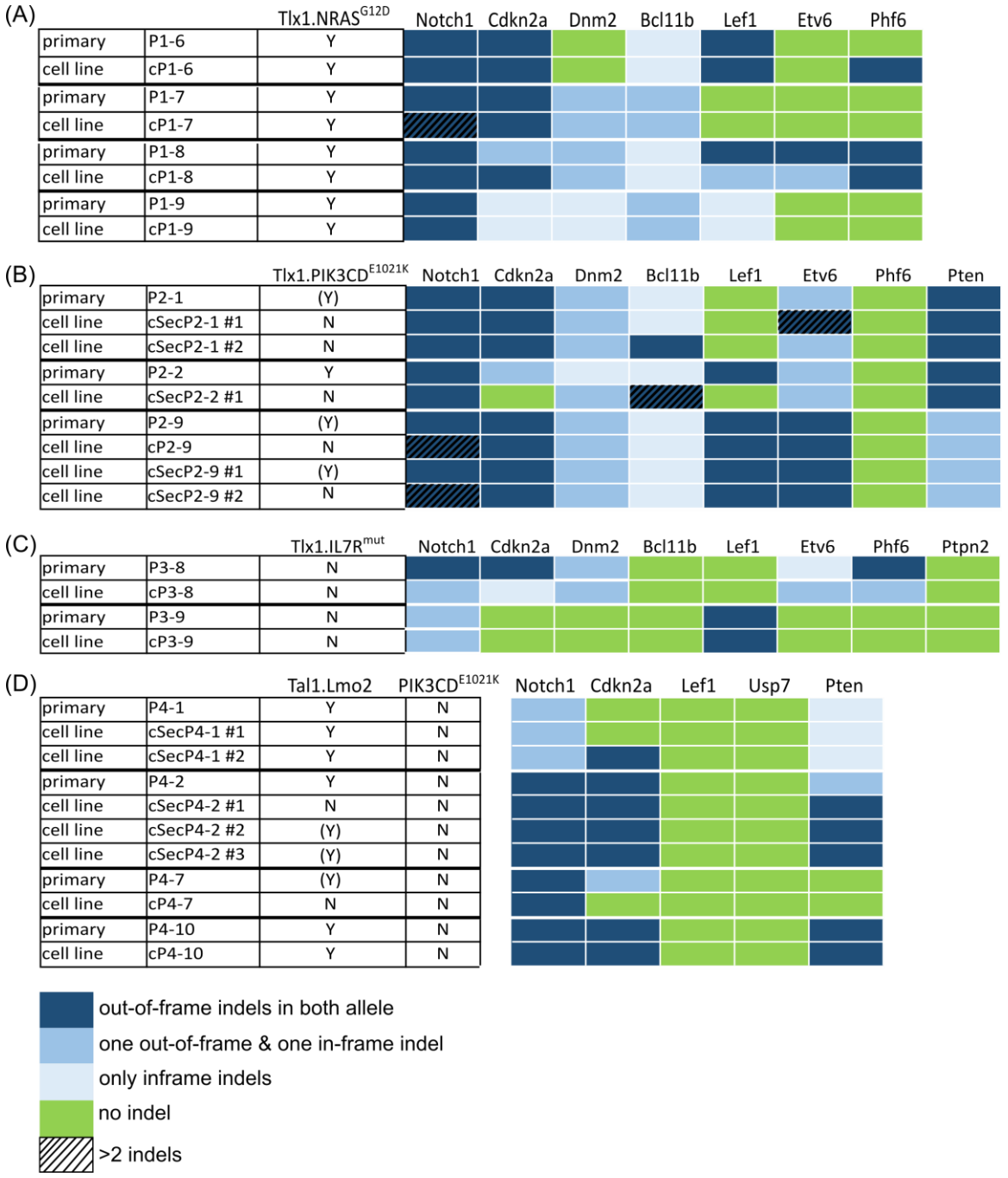


Figure 32: Genotyping primary T-ALL samples. Genotyping of pool 1 (A), pool 2 (B), pool 3 (C) and pool 4 (D) -derived cell lines, comparing the genetic aberrations of the primary T-ALL with obtained cell lines. The first column summarized the presence and orientation of the respective primary inverse oncogene construct (based on Figure 20). Brackets indicated their detection but no inversion. Targeted Next-Generation-Sequencing at each genomic region that was potentially targeted by a sgRNA from the respective lentivirus pool. Color-coded mutation profile indicated in-frame and out-of-frame insertions/deletions (indel).

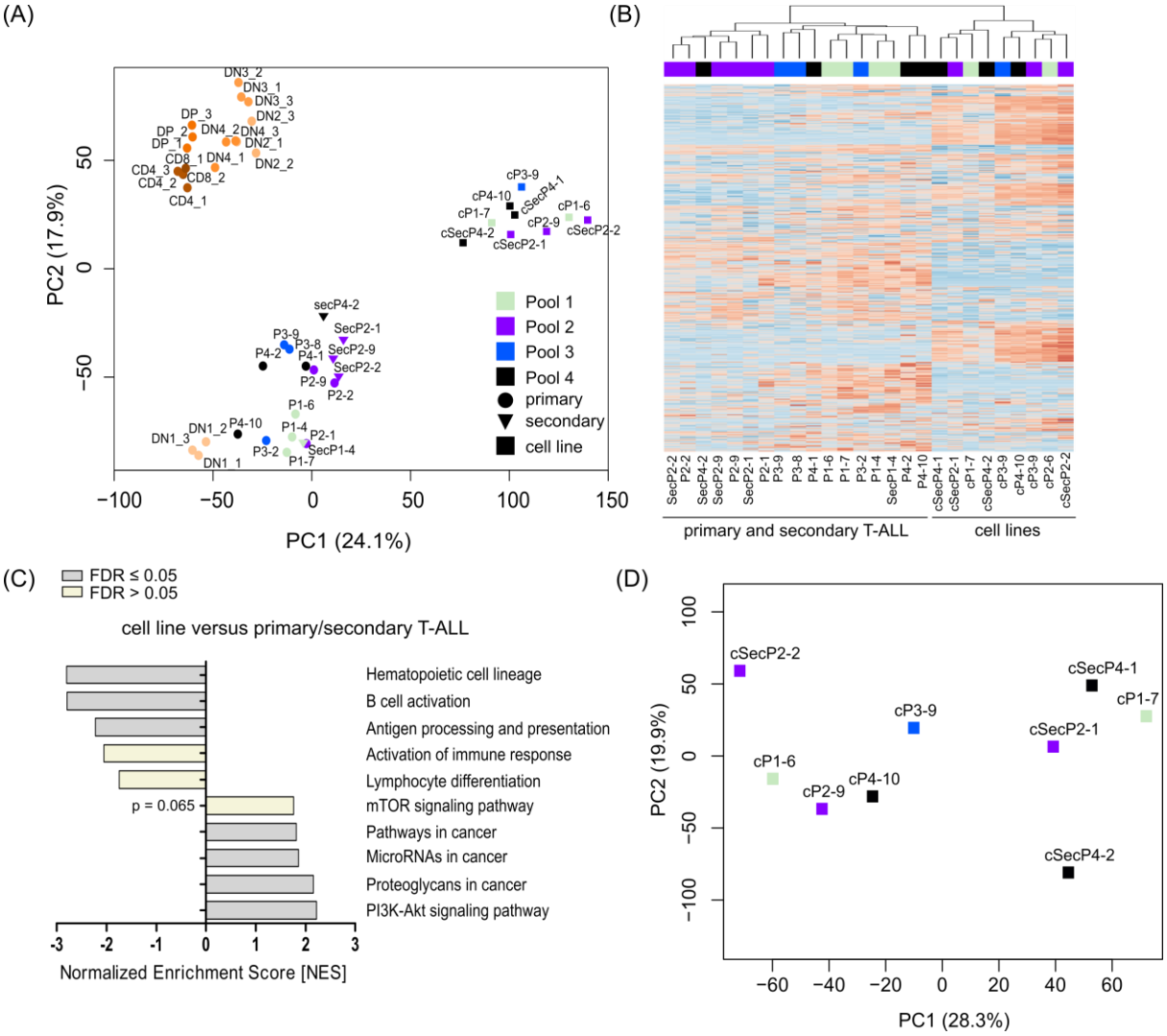


Figure 33: *In vitro* cultivation of T-ALL cells changed the transcriptome into a less mature stage. (A) Principal component (PC) analysis (PCA) was performed for all thymic, primary and secondary T-ALL and cell line samples, highlighting three distinct clusters. Hierarchical clustering (B) of primary and secondary leukemia and gained cell lines revealed a transcriptome alteration upon *in vitro* cultivation. Color legends were specified in (A). (C) Gene set enrichment analysis on https://www.webgestalt.org/ identified certain enriched KEGG pathways and gene ontology sets of biological processes. Selection of significant pathways between the cell lines versus the primary and secondary T-ALL cells. Biologically matching pathways were selected with $p \leq 0.05$ (if not depicted differently) and marked based on their false discovery rate (FDR) for > 0.05 and ≤ 0.05. (D) PCA within the different obtained cell lines didn't recapitulate the clustering behavior of the primary cells. Raw data can be find in an attached hard drive.

In this thesis, a mouse and a vector model were developed to modulate different biological groups of T-ALL. Complex disease phenotypes were generated *in vivo* and *in vitro* by multiplexing different genetic aberrations. This approach reinforced our knowledge of carcinogenesis, in which multiple genetic hits are required for malignant transformation, ranging from two to eight mutations depending on the cancer type[198]. Further studies will help elucidate new highly specific therapeutic targets on the different genetically modeled T-ALL entities.

6 Discussion and Outlook

Despite years of research on T-ALL, the 5-year survival rate for patients with T-ALL is with < 50% still relatively low[199]. Although most patients achieve complete remission, high relapse rates and resistance to chemotherapy remain the major obstacles to overall survival[68,182]. A promising treatment strategy is to inhibit NOTCH1, which is mutated in most patients. However, this intervention has failed in clinical trials[200–202]. Therefore, there is an urgent need to understand the underlying resistance mechanisms in T-ALL and to figure out how to circumvent them. Recently, PDX models have shown promising results as a tool for drug profiling[203,204]. Nevertheless, their availability is less applicable, complicating further studies to understand the driving mechanisms that lead to differential treatment susceptibility. Therefore, this thesis focused on the generation of a modular mouse model better reflecting the heterogeneous nature and mutations of T-ALL and its potential as a resource to study disease driving mechanisms and vulnerabilities.

6.1 Establishment and characterization of the *LSL.Cas9 x Lck-cre* mouse strain

The first part focusing on the generation and modeling of genetically complex T-ALL involved the development and characterization of a suitable mouse strain. In the style of Jan Cools group, overexpression of T-ALL related oncogenes was achieved by the T cell specific expression of Cre[53,205]. With this approach, Thielemans and colleagues were able to show that cooperation of aberrant TAL1 expression and mutated AKT^{E17K} accelerated leukemia burden. However, co-occurrence also increased their vulnerability to PI3K-AKT inhibitors that was less pronounced in leukemia with only one of the genetic variations[205]. Even though they showed the importance of understanding cooperating events in T-ALL, their approach focused on only one interaction. In a triple transgenic mouse model, Tremblay and colleagues were also interested in T-ALL expressing TAL1 and its cooperation with overexpressed LMO1 and activated NOTCH1[73]. They highlighted the insufficiency of overexpressed TAL1/LMO1, gained by chromosomal translocation, for leukemia development and the requirement of *Notch1* mutation for complete transformation.

Although the disadvantages of transgenic mice were already discussed (see chapter 2.4.3), their findings underlined the necessity for genetically more complex mouse models to understand disease development and evaluate severity more precisely. This is particularly important for adult T-ALL, which has been linked to an increase in genetic abnormalities

compared to childhood T-ALL[76]. Using the *LSL.Cas9 x Lck-cre* mouse model, the combinatorial introduction of up to 10 different genetic lesions was aimed for, allowing a more flexible and modular application with clonal outgrowth of the fittest mutation sets. Unlike Thielemans and de Bock, the *Lck-cre* mouse strain was used instead of *Cd2-cre* mice[53,205]. With this, it was expected to gain a more T cell specific expression, as *Cd2-cre* expression is also observed in early B cells[206]. However, the complexity of the multiplex approach and the overlap of certain oncogenes with other leukemia types[95,207] encouraged us to characterize the mouse strain and vector system more precisely. In particular, the publication of de Bock and colleagues, who investigated the cooperation of aberrant HOXA9 and $JAK3^{M511I}$ expression within T-ALL, supported this approach. In their first attempt, they introduced both genes without Cre-mediated guidance into BM and transplanted these cells into recipient mice. The resulting leukemia had a mixed-lineage phenotype of both myeloid and T cell lineages. Only after combining their transplantation approach with a *Cd2-cre* mouse strain, a T cell-specific transformation was achieved.

During the characterization of the *LSL.Cas9 x Lck-cre* model, expression of Cre and Cas9 proteins were detected by Western Blot analysis in T cell infiltrating organs, such as thymus and spleen (Figure 6 B). Although a minor Cre and Cas9 expression was expected within the BM from infiltrating T cells, it was not observed. Among the different leukocyte populations in the blood, expression of Cas9.EGFP was highly present in $CD3^+$ T cells, underlining the tight expression profile for both transgenes. Nevertheless, a minor population of $Gr\text{-}1^{low}$ cells also expressed Cas9.EGFP (Figure 7). At that point, it was unclear whether these were derived from a more immature group of granulocytes/monocytes progenitors or a specific differentiated subtype. On top, when analyzing purified lineage-negative BM cells of the *LSL.Cas9 x Lck-cre* strain, which were used for the transplantation, EGFP expression was already observed for 21 ± 19.8% of cells (Supplementary Figure 1). $EGFP^+$ cells were found among all BM subsets, but to a lesser extent in the LSK population. This was suspicious, as removal of the stop cassette upstream of *Cas9.EGFP* by *Lck-cre* would have led to its permanent expression. However, no such expression was found neither in mature $Gr1^{high}$ cells nor in other mature blood cells besides T cells. An explanation might lie within the expression of Cas9 and Cre itself. *Lck* can already be slightly expressed at the early LMPP stage (identified on www.bloodspot.eu using Mouse normal hematopoietic system[208]). Constitutive expression of Cas9 and Cre in the absence of *loxP* sites have both been linked to DNA damage[209,210]. This could have induced p53-mediated cell death in HSCs and early progenitors and explained, why there were no more progeny of EGFP-expressing cells in other lineages. The absence of Cre and Cas9 signals in Western Blot analysis on the whole BM of *LSL.Cas9 x Lck-cre* mice contradicted the flow cytometry-based detection of Cas9.EGFP in the BM and, thus, indicated

very minor and residual Cre and Cas9.EGFP expression within the BM cells (Figure 6 B and 7). Hence, its impact on hematopoiesis was neglected.

Besides the tissue-specific expression profile of Cre and Cas9.EGFP, a focus was set on the intact differentiation potential of T cells. Carow and colleagues described the impairment of T cell development and function in *Lck-cre* transgenic mice[164]. By comparing thymocytes of *Lck-cre*$^{+}$ mice with WT animals, they noticed a reduction of cellularity within the *Lck-cre*$^{+}$ thymus, leading to lower numbers of DP cells and elevated numbers in the DN population. *In vivo*, no striking alteration of the T cell differentiation compared to the parental *LSL.Cas9* mouse strain could be detected, indicated by comparable percentages of cells within the different stages of T cell development (Figure 8 B). The only distinctive feature was a higher variation in the abundance of the different DN populations within different thymuses of *LSL.Cas9 x Lck-cre* mice. By analyzing the EGFP expression, an increase from DN2 to DN4 stage was observed (Figure 8 C), which is in line with *Lck* expression within the different thymocytes populations[73,211]. However, a high EGFP expression was detected at the DN1/ETP stage. Cells at this early stage of development still hold the potential to differentiate towards B cells, NK cells and granulocytes[24]. Thus, the minor Cas9.EGFP expression within the Gr1^{low} population in the blood might be an artifact of ETPs lacking NOTCH1 signal and thus differentiated into granulocytes. Nevertheless, the transition from DN1 to DN2 was accompanied by a huge drop in EGFP expression, before it rose again (Figure 8 C). As discussed before, it could have been a consequence of p53-mediated apoptosis. However, the expression of Cre and Cas9 in more mature cells had no measurable effects, indicating intact T cell development. Strikingly, in the mature thymus, approximately 25% of mature thymocytes lacked EGFP expression, indicating silencing events or incomplete penetrance[212].

Taking advantage of the T cell-specific Cre and Cas9 expression, two vector systems were developed to introduce T-ALL-specific genetic aberrations – one for artificial gene overexpression and one for the disruption of autologous gene function (Figure 9 A). The inverse vector structure, with the GOI in antisense orientation, correlated with a very low leakiness in the absence of Cre (Figure 9 B). For the comparably tested stop-cassette vector system, it is important to mention that the used vector design contained only one stop-codon in each reading frame between the *dTomato* and *loxP* sequence. This design was much simpler, compared to, for example, the stop-cassette from the *LSL.Cas9* mouse strain, which consists of 3x polyA signals[213]. However, for the sake of lentiviral vector packaging, repetitive sequences and strong polyA signals were minimized. The inverse vector system allowed overexpression of transcription factors and mutated signaling proteins in the presence of Cre (Figure 9 C). Transgene expression under the control of an SFFV promoter was chosen to increase expression rates in hematopoietic cells[172].

In an *in vitro* T cell differentiation assay, the performance of the inverse vector system was demonstrated (Figure 10). Lineage-negative BM cells from *LSL.Cas9 x Lck-cre* mice were differentiated using an OP9-DL1 co-culture system, described by Zúñiga-Pflücker[38–40]. Although the transduction rate was relatively low, the inversion of a reporter vector with inverse vector design was demonstrated during the T cell development. When cells reached the DN4 stage on day 14, the inverted mTagBFP2 was expressed, and dTomato signal declined within the following days (Figure 10 B). This switch correlated with the expression of Cas9.EGFP, which also rose from day 14 of differentiation on (Figure 10 E). Interestingly, the increased Cre-mediated induction of Cas9.EGFP occurred later compared to the thymus measurements. Keeping a relatively high basal level of 14.2% ± 2.1% from DN1-DN3 stage, full EGFP expression increased as soon as the first cells reached DN4 and DP stages. This suggests that the stop-cassette in front of Cas9.EGFP was already removed in the $EGFP^+$ cells by the expression of Cre under the control of the *Lck* promoter. Along this line, evaluation of transduction efficiencies was difficult, since Cre could already invert the inverse cassette, preventing *dTomato* expression. While Cre or Cas9 expression could not be confirmed by Western Blot analysis in whole BM, it remained unclear, at which levels it was actually expressed and which progenitor cell types were affected. Therefore, transduction levels for the inverse vector construct might have been underestimated. Nevertheless, only 39% – 50% of cells showed complete penetrance at the mature DP and SP stages within this *in vitro* setting. *In vivo* verification was performed to determine the *in vivo* developmental stage of T cells at which the inverse vector is inverted into its active form. Unfortunately, a reporter-based assay failed due to insufficient chimerism of the transduced donor cells in transplanted recipient mice (Supplementary Figure 6).

The vector design and chosen genes were additionally evaluated in BAF3 and 32D cells (Figure 12). Only *Tlx1.IL7R*$^{c.731ins732}$ and *Tlx1.NRAS*^{G12D} transformed the cells and led to cytokine-independent growth. This was achieved by hijacking JAK-STAT signaling or the PI3K-AKT-mTOR and MAPK axis, respectively, and is in line with the previous description on IL7R$^{c.731ins732}$ and NRAS^{G12D} [175,214]. For the *IL7R*$^{c.731ins732}$ construct, it was experimentally confirmed beforehand that the mutant form, but not the WT version, can rescue cytokine deprivation-induced cell death in BAF3 cells (Supplementary Figure 2). While no such effect was expected for *Tal1.Lmo2*, lack of transformation capability for *Tlx1.PIK3CD*^{E1021K} was surprising. Kwok-Shing Ng and colleagues revealed the immortalization potential of *PIK3CD*^{E1021K} in a systematic screen on BAF3 cells, contradicting our experimental observations with no immortalization in BAF3 cells (Figure 12 C)[152]. Of important note, minor leakiness of the vector design with Cre-independent expression was observed for BAF3 cells transduced with *Tlx1.IL7R*$^{c.731ins732}$. Due to the strong selection pressure on these cells, a translation towards an *in vivo* application could not be guaranteed. Nevertheless, respective

concerns of off-target activity remained. Especially for the *NRAS*^{G12D} mutation and its implication in B cell leukemia[207], unwanted activation could have resulted in the malignant outgrowth of B cells. The comparison of the results from BAF3 and 32D cells indicated a similar immortalization pattern for both cell types but incomplete cell death for 32D cells. Here, the MOHITO cell line could have been a better alternative. It possesses an IL7 dependency for cell growth and resembles T-ALL physiology more precisely than the myeloid 32D cell line[215].

A similar mouse model has recently been described, in which B cell chronic lymphocyte leukemia (B-CLL) was modeled through the combinatorial disruption of B-CLL-related genes via CRISPR-Cas9 KO[216]. Through the combinatorial targeting of six different genes, genetically complex leukemias have been generated, capturing the genetic and transcriptomic alterations as observed in human CLL. These findings encouraged us to further use the developed mouse and vector systems to shape human T-ALL in mice.

Summarizing the first part of this thesis, a mouse strain was created, which expressed Cre and Cas9 specifically within the developing thymocytes. In combination with the efficient inverse vector system and potent sgRNAs, it was possible to modularly overexpress oncogenes and disrupt gene function by nonsense mutations in developing thymocytes to promote T-ALL development. However, due to the minor expression within Gr1^{low} or lineage-negative cells, the activation of introduced genes or sgRNAs within other hematopoietic lineages could not be excluded. Given the partial overlap of genetic aberrations in AML and T-ALL, careful examination of the leukemia phenotype was critical[95].

6.2 Modeling TLX1-related T-ALL

The disease modeling within this work was focused on the TLX1 and TAL/LMO T-ALL groups, which compromise roughly 30 % of adult T-ALL cases (Figure 34). Due to the high overlap of target genes of malignantly expressed TLX1 and TLX3, their control on leukemia development is highly related[217]. While TLX3 dominates children's disease, TLX1 is more prominent in adult T-ALL and was, therefore, in the main focus of this work[65,218]. In contrast to TAL1 and LMO2, TLX1 is not expressed during hematopoiesis but plays an essential role in spleen development[219]. Its inappropriate activation, induced by the translocation to the TCR locus, is a prevalent event in T-ALL[220]. In the work of De Keersmaecker *et al.*, transgenic mice overexpressing *Lck* promoter-driven *TLX1* developed T-ALL within 20 – 30 weeks[221]. Immature lymphoblasts infiltrated the thymus, BM, spleen, lymph nodes, and liver. The phenotypic analysis of the leukemia cells revealed CD3, CD4, and CD8 expression in most cases. Through the constant expression of TLX1 in developing thymocytes, T cell development was impaired, and thymocytes arrested at the DN2 stage, leading to increased apoptosis of the cells. By this, selection for secondary mutations was favored, and cooperating mutations within *Notch1*,

Bcl11b, and *Pten* were among the most frequently detected genetic alterations. This demand for secondary tumor-driving events is frequently observed in various T-ALL studies[67,73,192,222]. Thus, this thesis intentionally focused on providing these leukemic drivers.

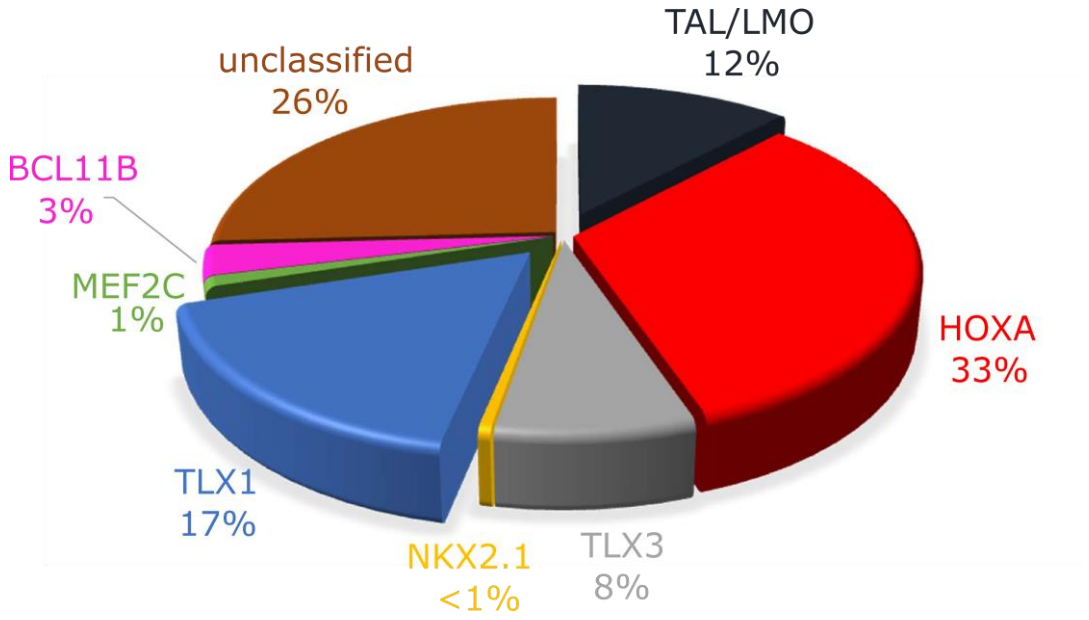


Figure 34: Distribution of T-ALL subgroups in adults. Distribution of adult T-ALL highlighting TAL/LMO and TLX1 as two major subgroups of T-ALL. Adapted and modified from Bardelli *et al.* 2021 ([65]; open access article distributed under Public License CC BY 4.0; Copyright © 2021 by the Authors).

One enormous benefit of this work's approach was its acceleration of T-ALL formation. Instead of 20 – 30 weeks, as seen in De Keersmaecker *et al.*, T-ALL developed within 8 – 17 weeks for pool 1 and pool 2[221]. Both groups contained murine Tlx1 in their primary transduction mix, together with secondary events promoting MAPK/ERK (pool 1) or PI3K-AKT signaling pathway activation (pool 2). The third group (pool 3) was intended to contain Tlx1 and a mutant IL7R to hyperactivate JAK-STAT signaling. These three branches of cellular signaling pathways are commonly dysregulated in TLX1 T-ALL patients[50].

Tlx1.IL7R$^{c.731ins732}$ had the highest immortalization capability in both BAF3 and 32D cell lines (Figure 12) merely by activating JAK-STAT signaling (Figure 13). However, the absence of *Tlx1.IL7R*$^{c.731ins732}$ vector within the final leukemia and low numbers of successfully developing T-ALL made evaluation and comparison of this group difficult. This was very unfortunate because Vanden-Bempt *et al.* modeled in transgenic mice the cooperation of TLX1 with the NUP214-ABL1 fusion protein, which is exclusively associated with TLX1 and TLX3 in T-ALL[98]. NUP214-ABL1 is known to directly phosphorylate parts of the JAK-STAT and MAPK signaling pathways[223]. In the transplantation study of Vanden-Bempt, they highlighted the role of STAT5, which shared genomic binding sites with TLX1 and thus corporately regulated gene expression, including *Myc*. The presence of the *Tlx1.ILR7*$^{c.731ins732}$ vector would have enabled

us to mimic similar cooperating events at high STAT5 activation, emphasizing the importance of malignant IL7R-signaling in T-ALL[175]. In the context of this work, most likely, the low titer for *Tlx1.IL7R*$^{c.731ins732}$ vector hindered sufficient transduction of BM cells (Supplementary Figure 4). Combined with the different sgRNA-containing vectors, transduced cells with the right combination of mutated genes have outcompeted *Tlx1.IL7R*$^{c.731ins732}$ containing cells in the final T-ALL.

Since transgenic mice expressing human *TLX1* are known to develop T-ALL within one year[98,221], lack of leukemic events in *Tlx1.NRAS*^{G12D} and *Tlx1.PIK3CD*^{E1021K} control groups were unexpected (Figure 21). One reason for the missing leukemia appearance might have laid within the co-expressed mutant cellular signaling proteins (NRAS^{G12D} or PIK3CD^{E1021K}). During the oncogene-induced senescence (OIS) process, abnormal cellular signaling activates p53 and promotes the senescence of the cells. Studies have indicated that this process can be induced through the p38 MAP kinase pathway in the presence of atypical Ras signaling[224,225]. A similar mechanism was described for high amounts of phosphorylated AKT, which leads to the accumulation of reactive oxygen species[226]. Both abnormal signaling pathways direct the activation of p53 activity and stop cell growth. In the same way, deregulated JAK-STAT signaling leads to the increase of suppressor of cytokine signaling 1 (SOCS1), which translocates at higher levels into the nucleus to phosphorylate and activate p53 [227]. In the absence of other genetic variations, which counteracted the p53 response, overexpression of *NRAS*^{G12D}, *PIK3CD*^{E1021K}, and *IL7R*$^{c.731ins732}$ might have induced cell growth arrest and thus prevented the transforming capability of TLX1. In this case, a loss of respectively transduced cells would have occurred in Cre-expressing cells after transplantation. Due to the inverse vector design, in which the *dTomato* fluorescence reporter was turned off in the presence of Cre, and the low chimerism of transplanted cells (Supplementary Figure 4 A), it was difficult to control the presence or absence of transduced cells after transplantation. Alternatively, VCN determination could be performed on the transplanted animals. Future work will have to examine if the transduced vectors were still detectable after one year in the experiment.

The immune-phenotypic characterization of generated T-ALL with *Tlx1* in the primary transduction mix was marked by a mixture of different immune phenotypes (Figure 18). This aligned with previous publications on T-ALL[221,228,229]. Within these three groups, T-ALL phenotypes were assigned to either a DN and CD4^{+}CD3^{-} (pool 1) or a DP and CD3^{low} (pool 2) phenotype. Abnormal expression of human *TLX1* can block the rearrangement of the *TCRα* locus and thus stops T cell differentiation at the DN3 stage[230]. However, T-ALL cells still reached an arrested DP phenotype, indicating an *in vivo* differentiation pathway towards DP, independent of functional TCR signaling[231]. This has frequently been described for TLX1^{+} T-

ALL and indicated a separation of the arrested T cell developmental stage and the measured immune phenotype[221,222,228] (more to this in chapter 6.3). The differential block has been associated as a critical feature in maintaining leukemia cell survival. Dadi *et al.* analyzed the effect of reinitiating the TCR-mediated differentiation pathway in the TLX1$^+$ TCR-deficient ALL-SIL cell line[230]. By transducing cells with a TCR alpha and beta chain, cells induced cell death in co-cultivation with OP9-DL1 cells. This apoptosis was detected only in the presence of the OP9-DL1 cells, highlighting the role of stromal cells for the differentiation of T cells and induction of cell death. A knockdown of TLX1 by shRNA resulted in apoptosis of the leukemia cells, too[230].

Such a TLX1 addiction of the leukemic cells might be challenged in some animals of pool 2 with an inserted but not inverted *Tlx1.PIK3CD^{E1021K}* vector. Transcription of non-inverted vectors could independently be regulated through the genomic insertion site and be expressed even without inversion. The lower number of T-ALL samples with integrated and/or inverted *Tlx1.PIK3CD^{E1021K}* vector, compared to *Tlx1.NRAS^{G12D}* and *Tal1.Lmo2* (Figure 19), indicated a minor contribution of *Tlx1.PIK3CD^{E1021K}* to leukemia. Western Blot analysis would be important to perform in order to confirm the expression of TLX1 in those samples. Consistent with these assumptions, mutation of *Pten*, *Bcl11b*, *Notch1*, and *Cdkn2a* majorly drove transformation. This hypothesis is supported by the presence of the same CD3$^-$DP phenotype in the sgRNA control group, lacking the *Tlx1.PIK3CD^{E1021K}* vector. Further generation of genomic data for the control group would unravel a contribution of which type B targeted genes were involved in final leukemia.

In contrast to the DP arrest in pool 2 (*Tlx1.PIK3CD^{E1021K}* + sgRNAs), pool 1 (*Tlx1.NRas^{G12D}* + sgRNAs) developed T-ALL with a very different phenotypical and morphological behavior. Here, transforming cells generated a lymphoma phenotype within the thymus with less involvement of peripheral organs. The thymus-persisting T-ALL cells gained a DN or CD4$^+$ CD3$^-$ phenotype, which was conserved after leaving the thymus (Figure 18 comparing thymus (B) and spleen (A)). The lymphoma phenotype arose most likely from the ectopic NRAS^{G12D} expression[185,232,233]. It has been described by Wang *et al.* that the transplantation of transgenic BM with homozygous *NRas^{G12D}* alleles into sublethally irradiated mice induced T-ALL within 6 – 7 months, mainly infiltrating the thymus[185]. These cells compromised a DP phenotype with elevated CD44 expression. In their final leukemia, random mutations occurred in *Notch1*, specifically targeting the PEST domain. The combination of both tumor-driving events made cells very sensitive to NOTCH1, MEK, and mTOR inhibition, suppressing cellular growth. Interestingly, when BM containing only one *NRas^{G12D}* allele was transplanted into recipients, animals developed a myeloproliferative disease with only a minor formation of T-ALL.

In this thesis, we combined a *Tlx1.NRAS*^{G12D} expressing vector with different secondary events to intentionally accelerate disease progression. A myeloproliferative outbreak could be prevented by applying the Cre-mediated T cell-specific expression of *Tlx1.NRAS*^{G12D}. In addition, *Bcl11b* mutations were most prominently associated with *Tlx1.NRAS*^{G12D} in pool 1 T-ALL, even more than *Notch1* mutations, most likely stabilizing the phenotype (Figure 20). Since *Bcl11b* haploinsufficiency accelerates lymphoma development in p53$^{+/-}$ mice[234], the cooperation of mutant murine *Bcl11b* and human *NRAS*^{G12D} gene expression further promoted the lymphoma phenotype. However, in contrast to all previous descriptions, this cooperation did not explain the arrest at the DN stage with a dominant CD4^{+} population, instead of a DP phenotype. DN arrest was also confirmed by similarity to DN1 and DN2 thymocytes in terms of their transcriptome data (Figure 27). In turn, it may be that Tlx1 has possessed a rather minor role during leukemogenesis. Besides the dominant DN phenotype, this hypothesis is further supported by the lower frequency of *Notch1* mutations, which are strongly associated with Tlx1[222]. Future work has to examine the expression of functional TCRαβ in the CD4^{+} cells and compare the retransplantation capability of CD4^{+} and DN T-ALL cells to decipher if both populations behave similarly or represent a heterogeneous leukemia cell assembly. In addition, a chromosomal analysis could reveal non-intended chromosomal translocation, which could explain the phenotypes' change.

Because Western Blot analysis was not performed in this work to confirm the expression of TLX1 or any of the mutant signaling proteins in the final T-ALL, a collaboration between ectopic TLX1 and the detected type B mutations can only be suspected. However, the presence of *Tlx1* in every single T-ALL from pool 1 and pool 2 made a random event less likely. The relatively low transduction rate with the type A vector made this high prevalence by chance less likely (Supplementary Figure 4). However, the initial transduction rates were difficult to evaluate, since the titration of the used vectors by flow cytometry and VCN showed enormous differences (Figure 14). This discrepancy could be due to a preexisting Cre protein in the lineage-negative BM cells (EGFP^{+} cells in Supplementary Figure 1) that switched vector orientation early, preventing the expression of the dTomato fluorescent reporter protein. Another explanation would be the process of nonsense-mediated mRNA decay (NMD). NMD can be induced by the long gene cassette in antisense orientation downstream of the *dTomato* gene. Accumulation of various termination codons upstream of the *polyA* signal recruits degradation factors and degrades mRNA (reviewed in [235]). Therefore, the percentage of dTomato^{+} cells measured by flow cytometry did not necessarily reflect the actual transduction rate. However, the omnipresence of *Tlx1.NRAS*^{G12D} and *Tlx1.PIK3CD*^{E1021K} in all T-ALL samples and, most importantly, the detection of known and described mutant genes cooperating with TLX1 indicated the involvement of *Tlx1* in leukemogenesis for pool 1- and pool 2-derived T-ALL. These mutant genes include *DNM2*, *BCL11B*, and *PHF6*, and are known

to cooperate mainly with aberrant TLX1 or TLX3 transcription factor expression in human T-ALL[50,86,91,221].

In contrast to the first two groups, leukemia from pool 3 completely lacked the *Tlx1*-containing vector but still represented the same dominant DP phenotype, as observed in T-ALLs from pool 2 (Figure 18). Despite the *Cdkn2a* mutation in 3 of 4 samples (with one being heterozygous), no typical mutation pattern was observed among pool 3 T-ALLs that could have explained this phenotype (Figure 20). Two T-ALLs contained mutant *Notch1*, while the other two showed no mutation in the sequenced region. The absence of *Notch1* mutations in P3-2 and P3-10, and the same for *Cdkn2a* in P3-9, was conspicuous (Figure 20). Since the two sgRNAs targeting both genes were expressed from the same vector and proved a high cleavage efficiency within the other leukemia samples, mutation of only one gene could indicate Cas9-independent alterations of the target region. However, even if other mutations could have independently accumulated within *Notch1*, the more extended T-ALL development of 188 and 189 days compared to 122 and 132 days clearly separated the leukemias, with no "detected" *Notch1* mutations from the others. Since no *Tlx1.IL7R*$^{c.731ins732}$ vector was detected in the final T-ALL of pool 3 transplanted animals, this group could be considered as a sort of sgRNA control group. In agreement with the sgRNA control group from pool 2, a Tlx1-independent predisposition to the DP phenotype can be inferred.

Based on the described contribution of the secondary mutations, the Cas9-targeted genes were hypothesized as the main drivers for the onset and rapidity of leukemia formation. This observation was in line with the control experiments, in which BM was transduced with the respective viral pools, lacking the primary oncogene cassette in the inverse vector (Figure 21). These control groups clearly showed that they mainly accelerated T-ALL formation. In contrast, the control group containing the transplantation of only *Tlx1.NRAS*^{G12D} transduced BM into six recipient mice generated only one T-ALL, while for *Tlx1.PIK3CD*^{E1021K} no transformation was detected even after one year post-transplantation. Thus, the Cas9 targeted mutations mainly contributed to the T-ALL outbreak. Nevertheless, in combination with *Tlx1.NRAS*^{G12D} and *Tlx1.PIK3CD*^{E1021K}, T-ALL formation was additionally enhanced and the immune phenotypes were better conserved. This finding is in line with previous publications in which *Tlx1* or *Nras*^{G12D} transgenic mice required secondary mutations for T-ALL development[185,221]. In contrast to these transgenic mouse models, the small number of genetically modified cells after transplantation in this work further reduced the likelihood of acquiring random mutations to accelerate disease development. In terms of future experiments, a limitation of this work was the simultaneous acquisition of these genetic lesions, which neglected a step-wise combinatorial interaction of different transforming events[94].

A correlation analysis was performed to determine whether there was an association between the chimerism of CD45.2$^+$ cells in the blood 5 weeks after transplantation and the overall survival. A lacking or marginal correlation was detected for pool 2 – 4 (Spearman r = 0.01, -0.2, -0.21, respectively). Only for pool 1, a slight connection between the chimerism and the overall survival was seen (Spearman r = -0.52), which could be due to an outlier having 49.8% CD45.2$^+$ cells when the remaining animals ranged from 0.075% – 3.76% (Supplementary Figure 4 B). Within the sgRNA control groups a more precise correlation was assigned to the pool 1 control group (Spearman r = -0.8), limiting its comparability to the main group (Supplementary Figure 4 C). For future experiments, using a higher irradiation dose of up to 9 Gy could help to increase overall engraftment[236]. Nevertheless, in all other groups the number of engrafted cells did not impact the disease formation, making the T-ALLs comparable to each other.

6.3 Modeling TAL/LMO-related T-ALL

Besides the TLX1 T-ALL group, TAL1/SCL and LMO2 were used to model a more mature and cortical stage T-ALL. Both transcription factors are regularly expressed during the early phase of hematopoiesis and regulate the differentiation and survival of HSCs[237,238]. Aberrant expression of either *Tal1* [97,239] or *Lmo1/Lmo2* [184,240,241] in developing thymocytes induces leukemia after a long progression period of up to one year. TAL1 binds to basic helix-loop-helix E-proteins, like E47 [242], forming a DNA-binding transcription-initiation complex with LMO2 and others to maintain hematopoiesis[243,244]. E47 alone acts as a homodimer transactivating the *TCRα* enhancer[245]. The loss of functional E47 has been linked to the development of T cell lymphomas with an immature phenotype[246]. During the ectopic expression of TAL1 and LMO2, E47 is sequestrated and bound tightly to TAL1. LMO2 further stabilizes this interaction and reduces the overall DNA-binding capacity. Thus, alternative genomic regions can be targeted[247]. The loss of functional E47 dimers had, thereby, a greater impact than the binding of TAL1 itself[248].

In combination, transgenic mice overexpressing *Tal1* and *Lmo1* or *Lmo2* generate leukemia within 3 – 6 months[73,184]. Of note, different publications refer to TAL1 interaction either with LMO1 or LMO2. However, no tremendous differences in the oncogenic function of TAL1 with either one of them have been observed[187,240,249]. In *Lck-Notch1*/*pSil-TSCL*/*Lck-LMO1* (*Notch1/TAL1/LMO1*) triple-transgenic mice, leukemia development was accelerated to 31 days[73]. Similar to TLX1, this observation highlights the need for additional genetic mutations to fully transform thymocytes[73,184,250,251]. This was in line with the lack of any leukemia phenotype within the *Tal1.Lmo2* transplanted control group, which received no sgRNA vectors within the primary transduction pool (Figure 21). However, animals from the *Tal1.Lmo2* control group were at day 225 post-transplantation when this thesis was summarized. Therefore, they

can still acquire secondary mutations. While the described *TAL1/LMO1*-transgenic mice expressed both transcription factors in each thymocyte, the low chimerism and transduction rate within our work further decreased the number of cells in which additional mutations could be accumulated (Supplementary Figure 3 + 4 A).

In theory, the aberrant expression of TAL1 and LMO2 arrests developing T cells at the DN3 – DN4 stage, inducing a stem cell-like phenotype and repressing the transition to the DP stage. In this highly proliferating phase, cells become more susceptible to acquiring additional mutations like in *Notch1*, driving disease development[73,251]. However, despite this early arrest, *Tal1/Lmo1* transgenic animals are described to represent a strong DP or intermediate single positive (ISP) $CD8^+$ phenotype[73,187,241,252]. Within this work, only P4-9 T-ALL depicted the DP phenotype, whereas all other T-ALL from pool 4 showed a $CD8^+$ phenotype.

A striking notice for pool 4 T-ALL was the high expression of CD3. Indeed, Tremblay *et al.*[73] indicated a 48% decrease in T-ALL outbreak in *Cd3*$^{-/-}$*TAL1/LMO1* transgenic mice, whereas Fasseu *et al.*[184] described a complete block of leukemia development in *Cd3*$^{-/-}$*TAL1/LMO1* transgenic mice. Since *TCR* β-chain rearrangement can still occur in human *TAL1/LMO1* transgenic mice, the pre-TCR signal in combination with CD3 provides a clonal advantage. This contributed to the prevalence of a $CD3^+CD8^+$ phenotype in pool 4 T-ALL within this work. Therefore, the classification as ISP or fully mature $CD8^+$ phenotype cannot be assigned. A clonal outgrow was further supported by the predominant presence of mainly two different allele variances within all mutated genes, verified by NGS (Supplementary Table 3). Interestingly, Tremblay *et al.* showed that *Cd3*$^{-/-}$*TAL1/LMO1*-derived T-ALL revealed either a DN or a DP phenotype. Since the preTCR cannot assemble in these animals, the transition toward the DP stage resolved a proscribed expression pattern. Thus, the observed immune phenotypes within this work did not represent the stage of T cell development at which the developmental arrest and transformation occurred[73,252]. This was further underlined by the strong upregulation of *Il2ra* (*Cd25*, Figure 28), which probably resulted from enhanced Notch signaling[253]. In addition, the elevated RNA expression of *Cd44* in $CD8^+$ T-ALL from pool 4 but also on T-ALLs from the other groups (Figure 28) supported the ineligibility of the two immature phenotype markers CD25 and CD44 to define the T cell developmental stage (Figure 28). Especially the upregulation of CD44 has been described in various T-ALL models[185,252] but is also linked to memory T cells and therefore expressed under physiological conditions in $CD4^+$ or $CD8^+$ T cells[254,255].

By performing transcriptome analysis and comparing the expression profile to a particular stage of T cell development, no clear assignment could be reached for most T-ALL, although some samples showed minor enrichment for a specific stage (Figure 27 C). This thesis provided precise insights into the transcriptional changes within healthy developing

thymocytes. Deeper analysis and comparisons have the potential to elucidate disease-driving mechanisms and further define which alterations hallmark the leukemia transcriptome. A striking discrepancy has been observed for the DN1 thymus subset. Huge transcriptomic differences were detected in comparison to the other thymocytes contradicting comparable published data (Figure 27)[256]. Further comparison of the obtained data to the published data sets has to be done in the future. Contamination likely resulted from $CD44^+$ macrophages and fibroblasts within the thymus (expression data extracted from Human Protein Atlas proteinatlas.org[257]). Additional staining and sorting for Kit^+ could have prevented this impurity[256]. Interestingly, only in some cases are there transcriptomic similarities between a given T cell stage and a T-ALL sample that overlap with the measured immune phenotype (for example for P1-7, P3-2, and P4-10). This discrepancy of the leukemia immune phenotype and the T cell transcriptome could explain why so few differentially expressed genes were identified within the different experimental pools or PCA-determined clusters. The physiology of the individual T-ALLs differ, resulting in high variability even within the same phenotype, making it difficult to cluster phenotypically similar T-ALL together. Even for the assessment of differentially expressed genes between all potential *Tlx1*- and *Tal1.Lmo2*-expressing T-ALL samples, only six significantly deregulated genes were calculated (adjusted p-value <0.05). Hence, sequencing more samples with similar genotypes could help elucidate crucial transcriptomic changes and identify potential target genes or pathways more vulnerable in a specific mutation set.

Surprisingly, the animals in the control group containing only the sgRNAs from pool 4 became ill at the same time as the animals that also had the *Tal1.Lmo2* and *PIK3CD*^{E1021K} vectors in the primary transduction mixture. The *PIK3CD*^{E1021K} vector was present only in two T-ALL, whereas the *Tal1.Lmo2* inverse vector was detected in its active orientation in every T-ALL from pool 4. While leukemia from the sgRNA control group showed a mixture of various immune phenotypes, with some expressing CD3 and others not, a remarkable conservation of CD3 expression was observed for each T-ALL in pool 4. Thus, from a functional consequence, the expression and contribution of *Tal1.Lmo2* to the leukemia development in pool 4 transplanted animals was very likely. However, future experiments must confirm their presence in final leukemia. Also, Western Blot analysis should be performed to confirm the presence of TAL1, LMO2 or PIK3CD^{E1021K}. Interestingly a $CD3^+$ phenotype was also detected within the sgRNA control group from pool 1 but not for pool 2. Further analysis of which transcription factors might drive the leukemias within the sgRNA control group would be interesting. This information could be given by genome translocation sequencing, karyotyping or RNA sequencing analysis[76,258].

Due to the co-transduction design of this work, 6 – 8 different vectors were introduced simultaneously into BM cells, facilitating up to 10 different possible genetic anomalies (overexpression and cleavage). The presence of that many sgRNAs within one cell promoted several DSB in the DNA, induced by Cas9-mediated cleavage and lentiviral insertion. This damage can activate the p53-induced apoptosis pathway[210,259]. However, this signaling pathway was most likely disrupted by the detected *Cdkn2a* mutations (Figure 20), indicated by the high expression levels of *Cdkn2a* without apoptosis induction (Supplementary Figure 13), or independent mutations in *Trp53* itself. Previous publications described models in which 5 – 6 other genes were targeted[108,216,260]. This work surpassed these numbers of intended variations. Some T-ALL samples had up to 7 targeted genetic aberrations and very likely high VCNs, although it was not measured. Therefore, insertional mutagenesis could have been a possible event that further contributed to leukemogenesis[119]. Indeed, Mitra Sharma *et al.* described the accumulation of activating retroviral insertions in the *Notch1* and *Myc* genes in their transduced *Tal1*-transgenic mouse strain, which significantly increased leukemia development compared to non-infected *Tal1*-transgenic mice[261]. A vector insertion site analysis could contribute to this question[262,263]. Although insertional mutagenesis may have contributed to the development of T-ALL in this work, the absence of other hematopoietic lineages and the marked differences in disease latency between the different groups suggest that the introduced alterations were the major contributor.

The preliminary genetic characterization of the different T-ALL sets revealed first interesting cooperating events, which will require further analysis in the future. Besides the already described common mutations in *Bcl11b*, *Notch1*, *Cdkn2a*, and *Pten*, some samples included mutations in *Lef1*, *Phf6*, *Dnm2* and *Etv6*. While most arose randomly, mutations in *Etv6* always co-occurred with *Dnm2*, indicating potential cooperation or dependency. Genetic translocation leading to the fusion of both genes has been described for precursor T-ALL[264]. No single T-ALL from pool 4 harbored a mutation in *Usp7* (Figure 20), which is known to cooperate with TAL1 in the sequestration of E-proteins and was therefore part of the vector pool[188]. Since the used sgRNA proved its efficiency in a reporter-based evaluation process, targeted cells were most likely clonally outcompeted by other combinations.

The peculiar differences in the duration of T-ALL development within the four transplanted groups were probably driven by *Pten* mutations. The experimental groups 1 and 3 were marked by the longest disease latency and harbored no sgRNA vector targeting *Pten* as part of the transduced lentivirus pool. PTEN is a prominent regulator of PI3K-AKT signaling and is often dysregulated in many different cancer types (reviewed in [166]). Its involvement in T-ALL pathophysiology has intensively been studied[47,75,168,265,266]. When conditionally deleting *Pten* in 40% of fetal liver HSCs (*VEC-cre* mouse), its high penetrance drove leukemia formation in

each mutant mouse[168]. Derived leukemia was identified as T-ALL in 70% of the cases, and as mixed AML and T-ALL in the remaining mice. Thus, the lack of *Pten* mutations could explain the delayed leukemia development in pool 1 and 3 transplanted animals. Furthermore, P4-7 was the only T-ALL within pool 4 lacking mutant *Pten* and exhibited the slowest time until leukemia. All others harbored at least one allele with a disruptive mutation. In patients, both *Pten* ablation and *Notch1* mutations are linked to increased resistance of T-ALL against NOTCH1 inhibition[47]. In this regard, future experiments should investigate the sensitivity of the different experimental groups toward NOTCH1 inhibition. In early studies, deleted *Pten* was also linked to the ability of leukemic cells to escape the thymus and infiltrate periphery organs like the liver and spleen[168,266]. Indeed, pool 1 and 3, lacking Cas9-mediated *Pten* mutation, had significantly less signs of splenohepatomegaly than pool 2- and pool 4-derived T-ALL. In addition, secondary transplantation of pool 1- and pool 3-derived T-ALL was only successful for three of six transplanted mice each and featured a decelerated progression compared to secondary transplantations from the other two pools. However, even though splenomegaly and hepatomegaly were less profound in these two groups, organs were still enlarged, and secondary transplanted T-ALL re-induced a new disease progression, indicating other aberrations besides the intended *Pten* mutation directed metastasis and leukemia re-initiation.

Zhu *et al.* investigated which genes assigned the leukemic stem cell property to *Pten*$^{-/-}$ T-ALL[267]. By enriching Lin^{-}CD3^{+}KIT^{mid} from a *Pten*$^{-/-}$ mouse strain (*Cdh5-Cre*$^{+}$*;Pten*$^{loxP/loxP}$*;Rosa26*^{floxedSTOP}*-LacZ*$^{+}$*)*, they identified *Spi-1 Proto-Oncogene* (*Spi1*) along with its transcriptional downstream targets Hepatitis A virus cellular receptor *2* (*Havcr2*) and Integrin subunit alpha x (*Itgax*). These genes were significantly enriched in leukemic stem cells (LSC) of *Pten*$^{-/-}$ T-ALL. In a limiting dilution assay, they linked high *Havcr2* expression to increased leukemia-initiating potential of *Pten*$^{-/-}$ T-ALL upon secondary transplantation. In double-transgenic mice with additionally deleted *Spi1*, no infiltrating cells could be detected in the thymus and spleen. Nevertheless, when treating *Pten*$^{-/-}$ mice with a SPI1-inhibitor, no significant beneficial effect on the survival of the mice was observed. This was also the case when animals were treated with a PI3K inhibitor, which prevented initial T-ALL formation when applied in a pre-leukemic stage[268], but only slightly slowed disease progression once T-ALL cells had formed. This indicated a resistance of the fully grown T-ALL against PI3K inhibition. However, combining both inhibitors significantly reduced leukemia outbreaks[267]. When analyzing the expression of *Spi1*, *Havcr2*, and *Itgax* in the developed T-ALL models of this thesis, most of the obtained T-ALL showed elevated expression of these LSC-marker genes (Supplementary Figure 9). In healthy thymocytes, *Spi1* expression was rapidly downregulated after DN1 and DN2 stages, which is in line with previous descriptions on thymocytes (Supplementary Figure 9 A)[256]. *Spi1* was just barely expressed in pool 2 T-ALLs, SecP4-2 and P3-8, while its expression was higher in the remaining T-ALL samples. Similarly, increased

expression of *Havcr2* was only valid for half of the leukemia samples. Interestingly, three secondary transplants (SecP4-2, SecP3-1, and SecP3-9) with low Havcr2 expression were found among the T-ALL samples. Since these samples had to rise from cells with high LSC potential, this observation contradicted the identified LSC markers of Zhu *et al.*[267] These published transgenic mouse strains do not recapitulate the whole picture of the *Pten*-depletion effect on tumor development since aberrant PI3K-AKT signaling has already started at a very early stage of hematopoiesis in their animal model[168]. In addition, the genetic background of different $Pten^{-/-}$ mouse strains has been described to significantly influence the impact of *Pten* KO[265]. Strikingly, four samples from pool 1 were among the samples with the highest expression of *Spi1*, *Itgax,* and *Havcr2*, even though *Pten* was not intentionally targeted within these leukemias. However, since they contained a hyperactive $NRAS^{G12D}$ signaling protein, PTEN-independent activation of PI3K-AKT signaling was still possible. The lack of *Havcr2* in half of the T-ALL samples could have resulted from the higher genetic complexity compared to the oversimplified T-ALL model of Zhu *et al*.[256] This highlights the importance of this thesis, in which the increased genetic complexity and the reproduction of different disease phenotypes provide a confidential tool to dissect T-ALL biology. Although, it is important to remember that LSCs compromise only a fraction of cells within heterogeneous leukemia. And even though NGS analysis indicated a profound monoclonal leukemia, non-assessed genomic alterations could still contribute to a heterogeneous character, in which leukemias are orchestrated as a hierarchical organization, and different (phenotypic) subsets can be restricted in reproducing the observed phenotype[269]. Thus, the overall expression level of the mentioned LSC-marker genes could have disappeared in a heterogeneous T-ALL mixture, since in this thesis RNA was prepared from the bulk cultures. Therefore, a RT-qPCR analysis on the different phenotypic populations within the obtained T-ALL could help to clarify this issue. In addition, transplantation of different populations of the phenotypically diverse T-ALL samples could help identify novel LSC marker genes in a more genetically complex context that could be used for new therapeutic approaches. Genetic alterations should be tracked after secondary transplantation to prove the clonal character of the preserved leukemia. In this context, genetic barcoding of the transplanted cells could help to elucidate heterogeneous characteristics and identify LSC populations. This would involve transduction of bulk cells with a number of different vectors encoding different fluorescence proteins and/or genetic barcodes. By acquiring the respective fluorophore combination over time, clonal variations in the final leukemia outgrowth can be identified and compared to the other populations before transplantation[269,270].

In summary, different phenotypical entities of T-ALL have been developed, reflecting known findings from published transgenic mouse experiments. However, due to the combinatorial approach, disease development was strongly accelerated, and phenotypic characteristics, like

in vivo behavior and immune phenotype, were more preserved. The obtained T-ALL models can be transplanted into secondary recipient mice to recapitulate disease development, highlighting stem cell-like features of the leukemia cells. Targeted sequencing revealed up to seven dedicated and introduced genetic aberrations, exposing strong cooperation of *Notch1*, *Cdkn2a*, *Bcl11b*, and *Pten* mutations. These type B mutations were identified as the main drivers of T-ALL induction and proliferation, whereas the type A vectors directed the immune phenotypical characteristics. Thus, the obtained T-ALL models' genetic complexity mimics the clinical reality more realistically. Future experiments will now have to elaborate on the effect of different treatment conditions on these models to examine resistance mechanisms that prevented the translation of previous treatment strategies into patients.

6.4 Murine T-ALL-derived cell lines partially recapitulate *in vivo* leukemia

When treating T-ALL, the tumor microenvironment plays an important factor in the drug response. In particular, mesenchymal stroma cells (MSC) and endothelial cells (EC) accumulate in specific niches within the BM and foster T-ALL survival through various pathways[271–273]. Through the co-cultivation of primary T-ALL cells with MSCs or ECs, an *in vitro* drugs-response can be modeled, which behaves similarly to the *in vivo* conditions[203,204]. In addition, *in vitro* co-cultivation has been described to correlate with T-ALL growth variability that can be observed *in vivo*[274]. Therefore, this work focused on a co-cultivation system to mimic a more physiological microenvironment for the T-ALL cells *in vitro*.

Different cell lines have been described for the co-cultivation of various ALL samples. These included mouse stromal MS5 cells[275], OP9-DL1 [196], immortalized MSCs[197,203], OP9-DL4[276] or human umbilical vein endothelial cells[204], covering a wide variety of supporting feeder cells. In the first attempt of this thesis, OP9-DL1 cells were used and a change in CD3, CD4 and CD8 cell surface marker expression towards a $CD3^+$DN phenotype was observed (Figure 29 B). Since T cells usually become independent from Notch signaling at the DN3 stage[27], it was hypothesized that the Notch-ligand provided a selective proliferation signal to the heterogeneous T-ALL population, while aberrant CD3 expression substituted missing TCR signaling and supported proliferation[231]. Thus, the DN population from the first pool 2 cell lines grew over and upregulated CD3 expression (Figure 29). Most cultivated mature T-ALL cell lines kept their phenotypic characteristics by changing the feeder cells to OP9 cells. Since pool 1-derived T-ALLs were already at the DN stage, their *in vitro* cultivation was continued on OP9-DL1 cells. However, a phenotypic change still occurred during the very first days of cultivation. Thereby, CD4 expression of pool 1 T-ALL disappeared, and a profound $CD3^-$DN2 phenotype persisted. Preassigned DP T-ALL from pool 2 and 3 changed to a $CD8^+$ or DN phenotype. Only pool 4-derived cell lines kept their characteristics. The preservation of a DP phenotype

was only achieved within one generated cell line from the sgRNA control of pool 2 harboring mutations in *Notch1*, *Cdkn2a*, *Bcl11b*, *Phf6* and *Pten* (Supplementary Figure 12). CD25 (IL2RA) was highly upregulated in all cell lines cultivated with mIL2 and mIL7, while it was downregulated over time in those conditions lacking mIL2 and mIL7 (Supplementary Figure 11). Herein, it was noticed for cP4-7 that in the presence of cytokines its phenotype changed towards $CD3^{-}$DN, whereas cultivation in the absence of any cytokines preserved the phenotypic characteristics. However, this was not true for all cell lines, especially since the withdrawal of cytokines was toxic for most cells. Different cytokine cocktails have been used in other publications, ranging from stem-cell supporting conditions (FLT3-L, IL7, hSCF and Insulin)[275] to individual receptor-expression-based conditions (IL2, IL6, IL7, IL12, and IL15)[204]. Therefore, phenotypic changes and variability in mIL2 and mIL7 dependency highlighted further optimization potential for the cytokine conditions.

The genotypic analysis of the obtained cell lines compromised the verification of type A vector insertion and inversion and the analysis of indels within Cas9-targeted genomic regions. The general overlap of type A vector insertion and type B indels between the cell lines and the primary T-ALL indicated high similarity and a clonal origin. Although TIDE analysis[157] limits the analysis of which specific indel sequence was altered in the cell lines, the overall resemblance to the primary T-ALL underlined the clonal character of the obtained T-ALL and its genetic preservation during cultivation. However, minor differences were still observed and indicated a not completely mono-clonal nature of the T-ALL. It would be helpful to obtain the genetic information for the T-ALL samples from secondary transplantation. For examples, the loss of the *Tal1.Lmo2* vector in the cell line ‘cSecP4-2 #1’ could have already appeared in the secondary T-ALL or was the result of a cultivation-induced selection from a polyclonal population. The same is true for ‘cSecP2-2 #1’, which lost *Cdkn2a* and *Lef1* mutations, while others such as ‘cSecP4-1 #2’, acquired mutations in *Cdkn2a* compared to the corresponding primary T-ALL. Throughout the secondary transplantation and the *in vitro* cultivation, certain cell subsets most probably outgrew the main population, potentially through additionally acquired mutations, which were not assessed in the genotyping approach. These could have occurred by chance or as a result of Cas9-off-targeting. The sgRNAs were designed to be highly specific and required ≥ 3 mismatches to other gene exons for off-target binding, effectively preventing unwanted cleavage[108]. Targeted sequencing of the top 5 off-target candidates has to be performed in future experiments to validate this assumption.

Significant differences were observed when comparing the primary and secondary T-ALL transcriptome with the derived cell lines (Figure 33). This phenomenon has also been described by Vincenzo *et al.,* who cultivated PDX T-ALL cells in co-culture with human umbilical vein endothelial cells[204]. The cultivated T-ALL cells expressed genes related to the

DN or DP stage of T cell development. Co-culturing resulted in profound changes in the transcriptome associated with the upregulation of specific cellular signaling pathways, including EGFR, JAK-STAT, MAPK, TGFB, and EGFR signaling. Single-cell RNA sequencing confirmed that mono-cultivation led to more drastic changes in the transcriptome compared to co-cultivated T-ALL cells. They claimed a high translatability of *in vitro* identified drugs and proved the increased survival for 5 predicted candidates in 11 PDX models. The observed transcriptomic changes in this thesis mostly correlated with an increased mTOR and PI3K-AKT signaling and a more immature immune cell signature. Increased expression of *Tgfb2* and *Egfr* was in line with the described alteration during cultivation from Vincenzo *et al.* (Supplementary Figure 10)[204]. These findings highlighted the plasticity of the generated T-ALL models. Future experiments have to involve testing different pre-clinical drugs to verify, if the genetic and transcriptomic complexity of these works' models can be used to predict potential candidates.

In contrast to the PDX models, the native interaction of murine stroma cells and murine T-ALLs can be targeted with this model *in vivo*. One such target could compromise the CXCR4-CXCL12 axis between BM stromal cells and leukemia cells. In cancer cells, CXCR4 signaling drives metastasis and promotes angiogenesis through the secretion of vascular endothelial growth factor (VEGF)[277,278]. Through the high expression of CXCR4, tumor cells migrate to the BM, in which CXCL12 is secreted[279]. In B-ALL, treatment with CXCR4 inhibitors disturbs the interaction in the BM niche and mobilizes the cells to the periphery. Thus, they become more vulnerable for conventional drugs[280]. For T-ALL cells, such susceptibility was described as even more profound. CXCR4 is found to be highly expressed in PDX T-ALL. And indeed, when PDX are transplanted into mice with CXCL12-knockout in vascular endothelial cells, engraftment is highly reduced. In addition, the conditional knockout of *Cxcr4* in a *Notch1*-overexpressed T-ALL model can reduce leukemia burden even after its outbreak[281]. Among all primary and secondary T-ALL, and all cell lines, high expression levels of *Cxcr4* and *Vegfa* supported the described observations. Interestingly, *Cxcl12* was highly upregulated in the cell lines (Supplementary Figure 10). The tumor microenvironment would usually secrete this cytokine. However, in this work's T-ALL model, tumor cell lines expressed high levels of *Cxcl12*, and other cytokine genes like *Il7* and *C-Kit ligand* (*Kitl*) themselves (Supplementary Figure 10), indicating that some of these cytokines, like KIT-L, were potentially missing in the cultivation conditions. At this point, it was not clear whether an RNA contamination by OP9 cells was the case or cells independently secreted these cytokines for extended survival. Hence, the transcriptome analysis of the data obtained in this thesis will require more intensive evaluation.

Functional studies with the developed T-ALL models are an essential aspect lacking in this work. This thesis claimed to address the impaired translatability of previous findings for the treatment of T-ALL due to genetically low-complex leukemia models. Especially regarding successful pre-clinical NOTCH inhibition, clinical studies indicated poor response and benefit for treated patients[200,201,282]. Therefore, this work provided genetically complex T-ALL, covering clinically relevant mutation patterns. With an increasing understanding of T-ALLs biology, new drug candidates have been proposed (reviewed in [220]), which can be tested within the T-ALL models presented here. In this regard, targeting *Bcl11b* might be a promising candidate. Its haploinsufficiency is well-known and recapitulated in the groups targeted by a sgRNA[234]. Therefore, its knockdown was beneficial for leukemia formation, but the minor expression was still necessary for survival. Thus, targeting the remaining *Bcl11b* expression could depict an exciting therapeutic candidate. In addition, genetic rescue experiments that repair the introduced genetic aberrations or large-scale CRISPR-Cas9 library screenings could help identify novel targets *in vitro* and directly test them in appropriate *in vivo* models. Although the developed model has not proven a differential drug response based on its genetics so far, the performed characterization within this work founded a tool to address all these questions.

6.5 Conclusion and future perspective

In this dissertation, a mouse and a vector model were developed to modulate different entities of T cell acute lymphoblastic leukemia. Overexpression of specific oncogenes and introduction of defined genetic missense and nonsense alterations in T-ALL-associated genes reflected the cooperative evolution of numerous genetic lesions transforming T cells *in vivo*. Thus, complex disease phenotypes have been modeled that exhibit disease-relevant genetic alterations and reflect various mechanistic transformation processes. This multiplex approach overcame the long latency for spontaneous mutations required for leukemia development in transgenic mice.

Following the objective of this work, a mouse model was generated that tightly expressed Cre and Cas9 during T cell development, allowing the targeted manipulation of the differentiation process. This resource allowed the modeling of three phenotypically and transcriptomically distinct groups of T-ALL, ranging from an immature $CD3^{-}$DN/$CD4^{+}$ phenotype towards the $CD3^{low}$DP stage and mature $CD3^{+}CD8^{+}$ leukemia cells. The phenotypically conserved leukemia re-induced the disease after secondary transplantation with a slightly changed immune phenotype. This proved the presence of leukemic stem cells and highlighted the plasticity of T-ALL.

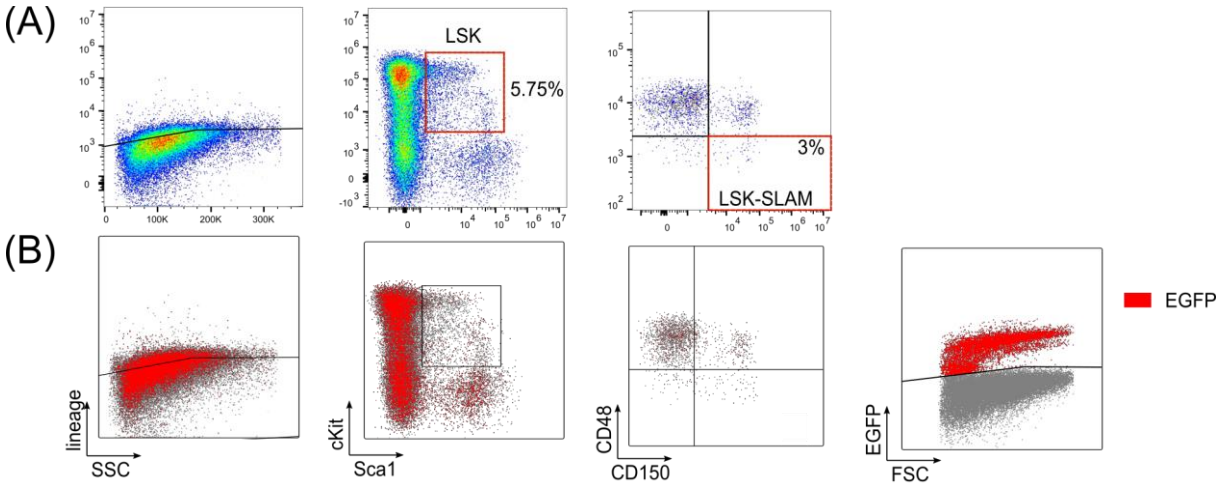
The careful design and evaluation of the corresponding sgRNAs resulted in high on-target activity with up to 7 different intended mutations within one animal. An off-target activity of Cas9 was not assessed within this thesis and represents a mandatory next step. With this

achievement, the genetic complexity of these models surpasses previous attempts in modeling T-ALL, paving the way to test new drugs in a genetically more realistic disease context. *Bcl11b*, *Notch1*, *Cdkn2a* and *Pten* mutations emerged as the main drivers for leukemia growth and migration. At the same time, the introduced 'type A' vectors encoding *Tlx1.NRAS*^{G12D}, *Tlx1.PIK3CD*^{E1021K} and *Tal1.Lmo2* mainly shaped the developmental and immunophenotypic stage in the obtained T-ALL. Further experiments must determine how the different cooperating genetic aberrations rely on each other and if this facilitates potential therapeutic targets. Derived cell lines grew sufficiently in co-culture with OP9/OP9-DL1 cells and featured a stable immune phenotype, but with slight changes compared to the primary T-ALL. They provide a reliable cell source to screen for new therapeutic targets.

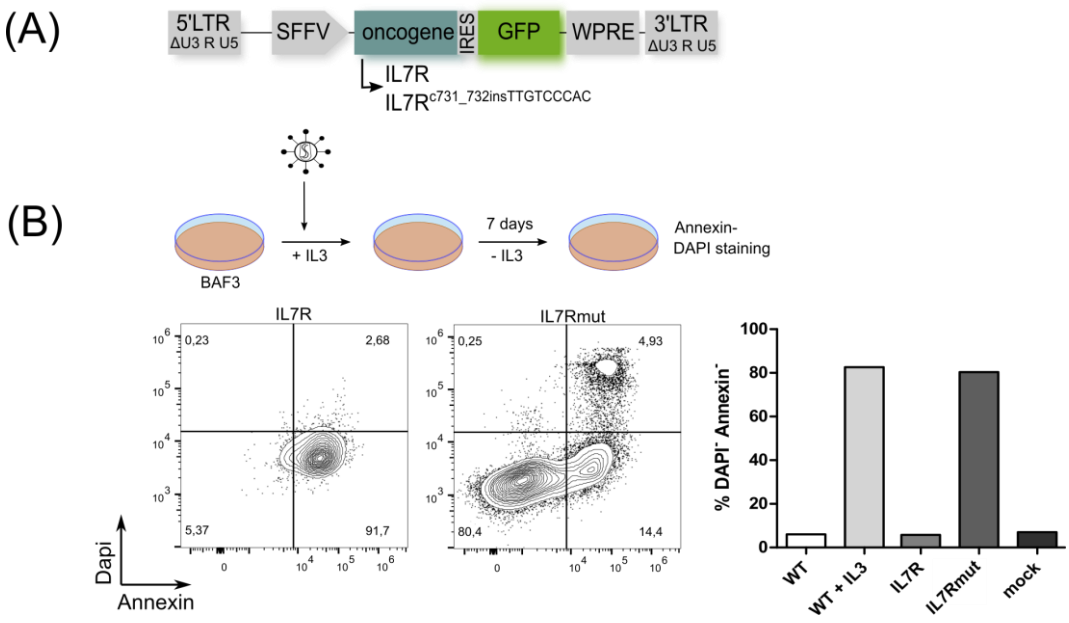
Due to its ease of use and modularity, the described model system would allow the mimicking of a broad spectrum of different T-ALL subgroups. This approach further provides the establishment of a cell bank with defined phenotypes and genotypes. The exclusive availability of patient genomic data has been challenging for predicting drug sensitivity[203], whereas transcriptional data have been evaluated as more promising[204]. This work indicated transcriptional heterogeneity within experimental groups, and may therefore help unravel novel therapeutic strategies and understand the mechanisms behind drug resistance. The described transcriptional data sets require further analysis to define definitive changes and predict reliable susceptibilities. Endogenous *Cas9* in *LSL.Cas9 x Lck-cre*-derived T-ALL holds the potential to perform CRISPR-Cas9 screens on the obtained cell lines *in vitro* and directly validate findings *in vivo* by retransplantation of the corresponding T-ALL. Thus, extensive functional studies can be addressed in future studies.

7 Appendix

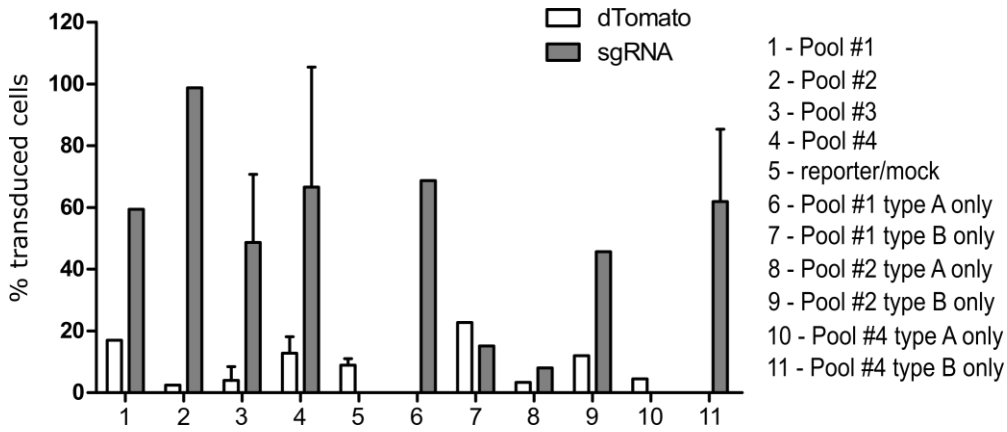
7.1 Supplementary Figures



Supplementary Figure 1: Differential marker expression on lineage-negative BM. (A) Composition of lineage-negative BM from *LSL.Cas9 x Lck-cre* mice. Cell were first gated for the absence of lineage-marker (Lin$^-$), before gating on cKit$^+$ and Sca1$^+$ cells. Further gating was performed for CD48 and CD150 markers. Hematopoietic stem cell population (HSPC) was identified within the LSK positive fraction (Lin$^-$ cKit$^+$ Sca1$^+$) and the LSK-SLAM immunophenotype (LSK$^+$, CD48$^-$ CD150$^+$). (B) EGFP expression of lineage-negative BM cells and its distribution along its different subpopulations.

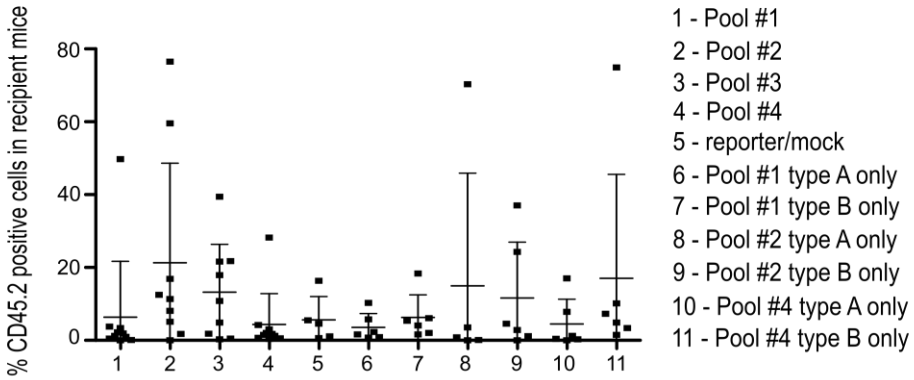


Supplementary Figure 2: Comparing immortalization potential of WT and mutant *IL7R* genes in BAF3 cell. Vector design of lentiviral *IL7R* WT and mutant *IL7R*$^{c.731\_732insTTGTCCCAC}$ (IL7Rmut). (B) BAF3 cells were transduced in RPMI supplemented with mIL3 with the respective lentiviral vectors. Afterwards, cells were starved for 7 days from mIL3, stained with Annexin and DAPI and analyzed by flow cytometry. Exemplary flow cytometry gating summarized percentages of living cells (DAPI$^-$Annexin$^-$) as graph; n = 1.

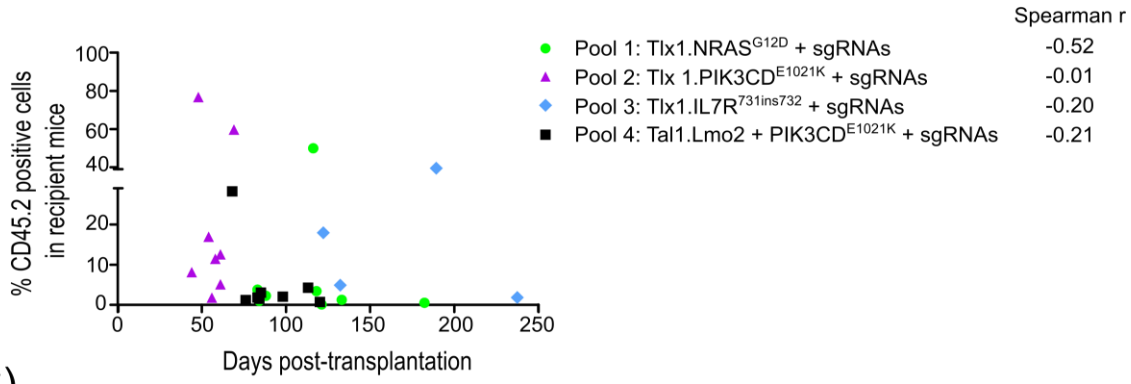


Supplementary Figure 3: Transduction rates in transplanted lineage-negative BM cells. Percentage of cells, expressing dTomato from the type A inverse vector construct and Thy1.1 from the sgRNA-encoding vectors at the day of transplantation.

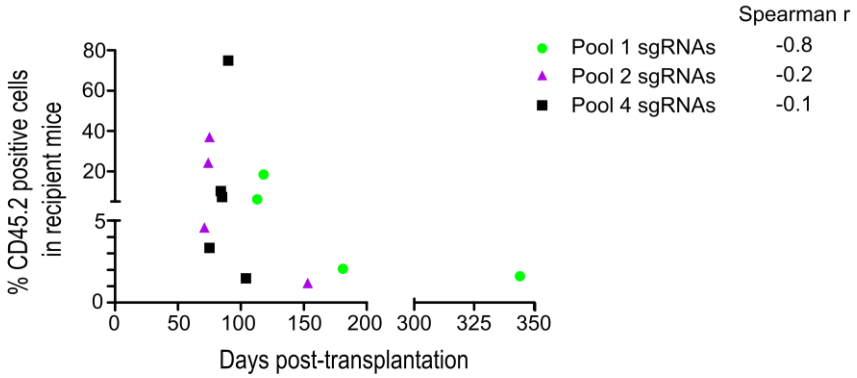
(A)



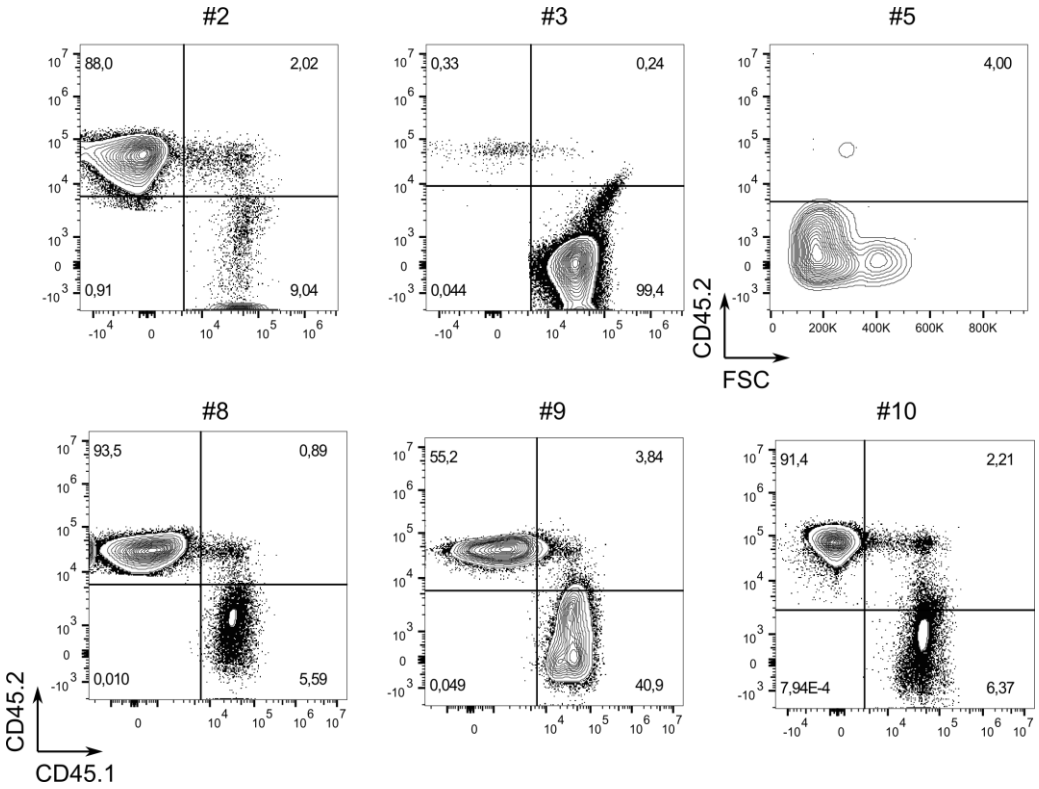
(B)



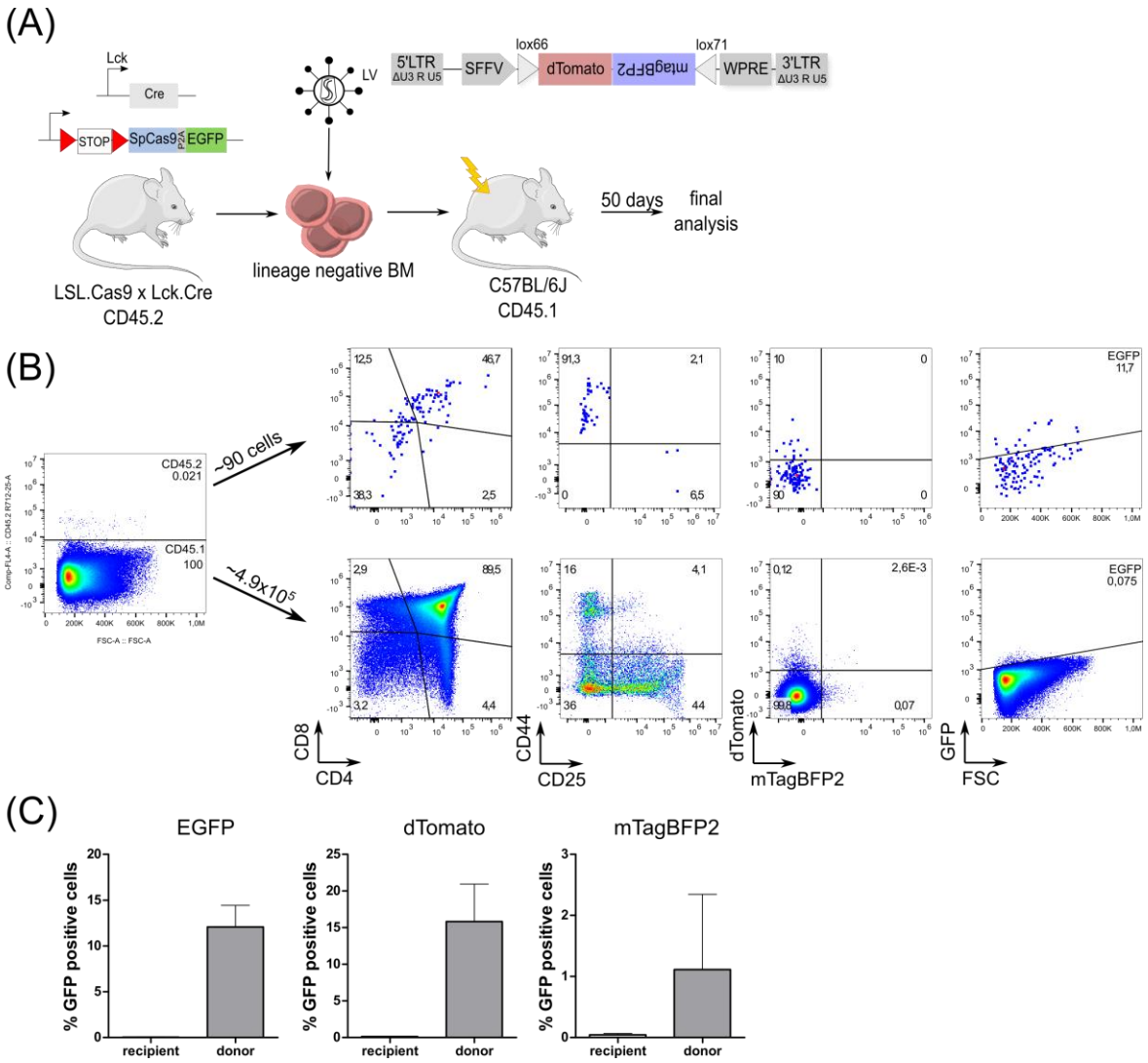
(C)



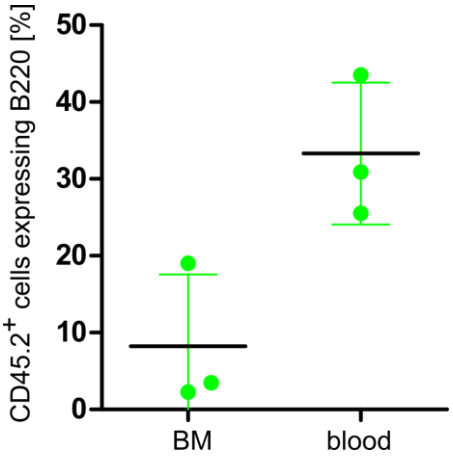
Supplementary Figure 4: Leukemia development and its relation to animals' chimerism after transplantation. (A) Chimerism of CD45.2$^+$ cells five weeks after transplantation for each transplanted group. Experimental groups compromise transplantation of transduced bone marrow with both primary oncogenes and sgRNAs (pool 1 – 4), and the transplantation of only the primary oncogene or only their sgRNAs as control groups. (B+C) Relation between first measured chimerism of transplanted CD45.2$^+$ cells in the blood and the duration for leukemia development for the experimental main groups (B) and sgRNA control groups (C). Spearman r summarizes the correlation between both parameters. Values < |0.2| represent very week correlation, < |0.4| a week correlation, < |0.6| a moderate and > |0.6| a stronger one.



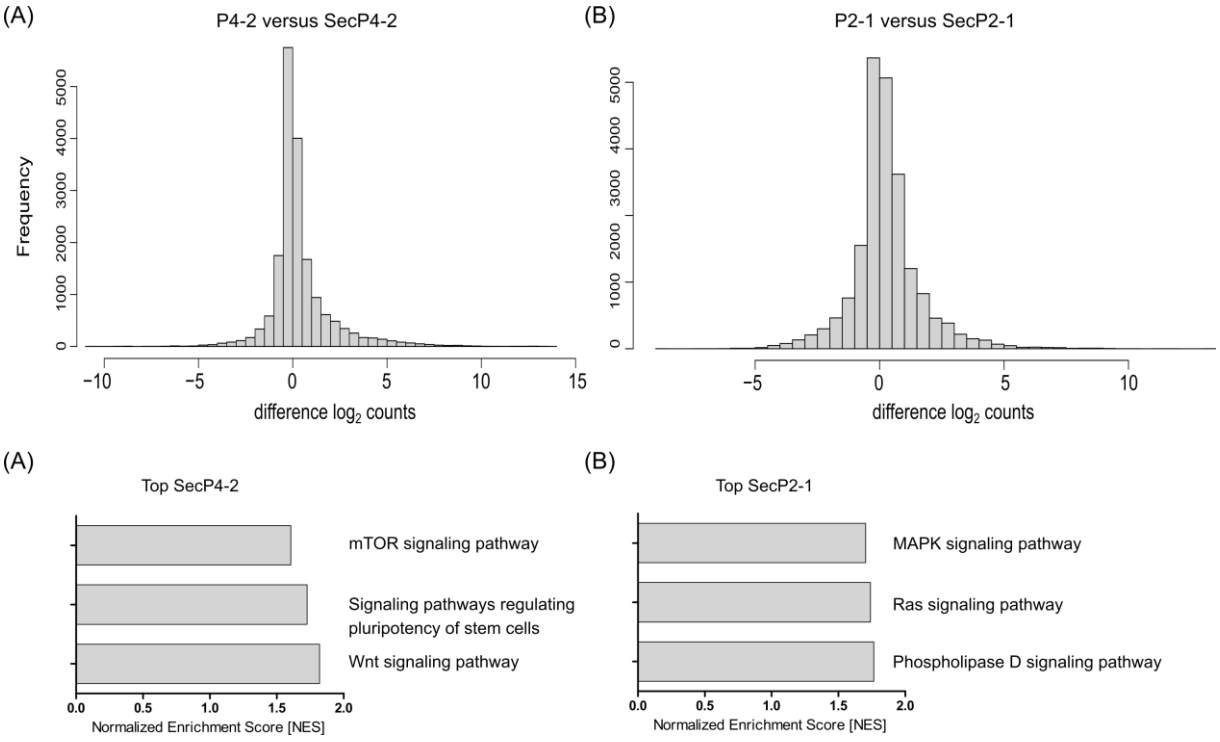
Supplementary figure 5: Final chimerism of CD45.1$^+$ and CD45.2$^+$ cells in the spleen of pool 3 transplanted animals. The spleen of all leukemia-diseased animals was stained for CD45.1 and CD45.2 surface markers and analyzed by flow cytometry.



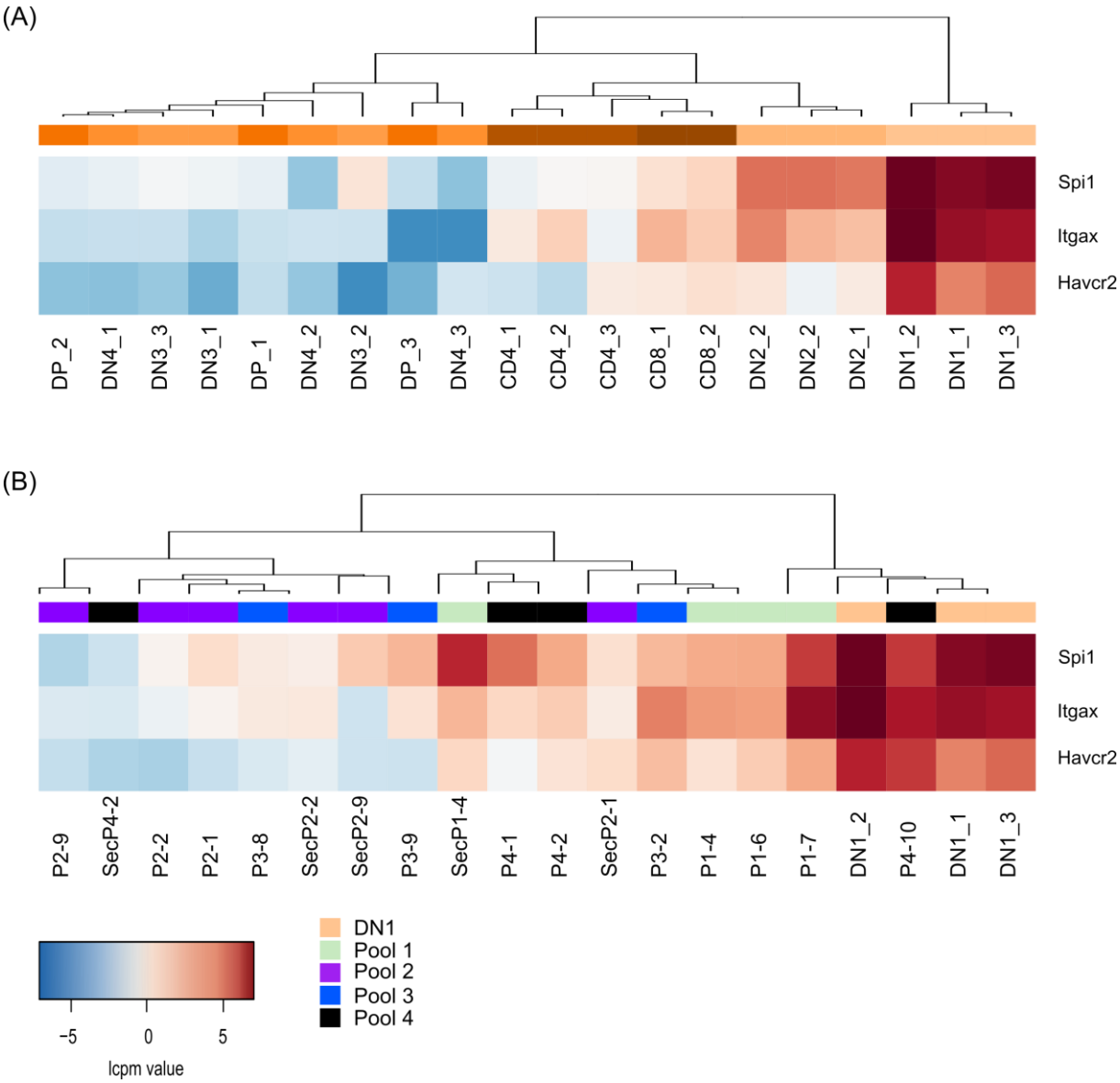
Supplementary Figure 6: Transplantation of transduced bone marrow with a reporter vector failed due to low chimerism. (A) Schematic overview of the co-transduction of lineage-negative bone marrow (BM) cells from *LSL.Cas9 x Lck-cre* mice with an inverse reporter vector, harboring a *dTomato* in sense and a *mTagBFP2* in antisense orientation. After 50 days, animals were sacrificed and thymus was analyzed for reporter expression. (B) Both, donor- (CD45.2$^+$) and recipient-derived (CD45.1$^+$) thymocytes were analyzed for the T cell surface markers CD4, CD8, CD25 and CD44. Additionally measured expression of dTomato, mTagBFP2 and GFP. (C) Quantification of EGFP, dTomato and mTagBFP2 expressing cells in donor- (CD45.2$^+$) and recipient-derived (CD45.1$^+$) thymocytes. No statistics was performed on the small cell numbers.



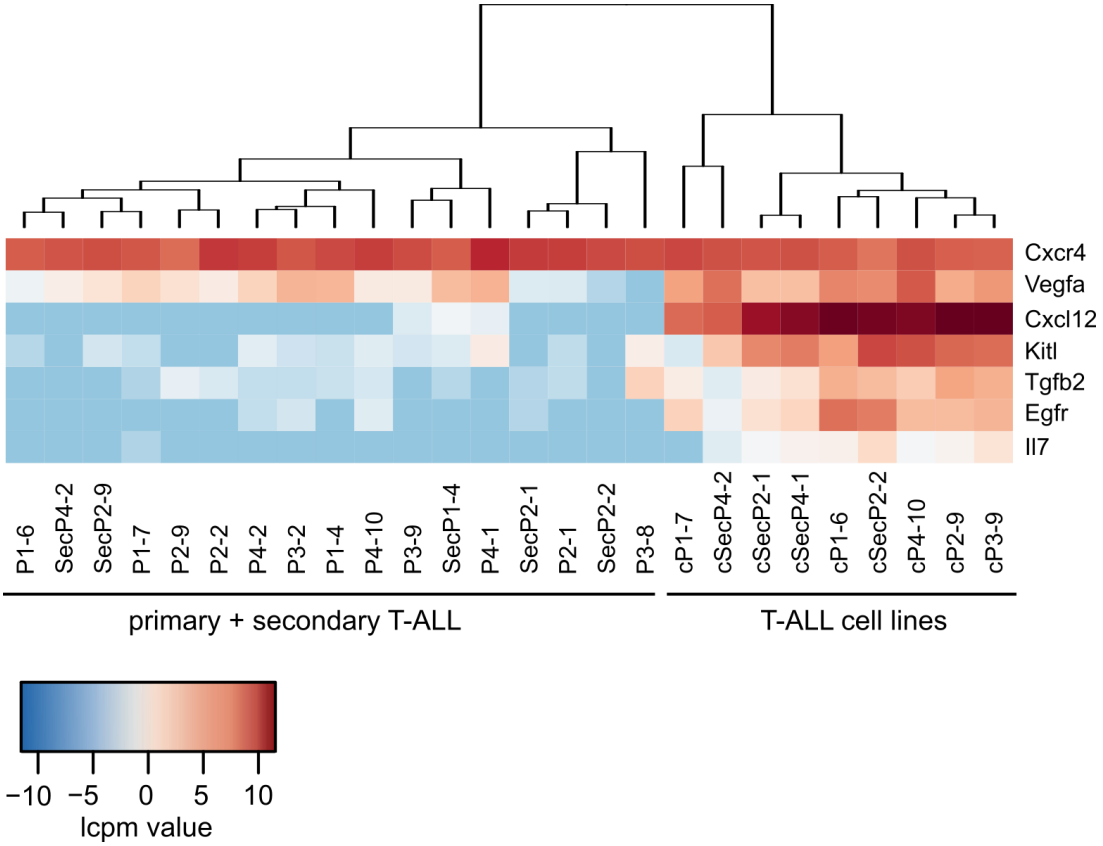
Supplementary Figure 7: B220-expressing cells in secondary P1-4 leukemia. Percentage of B220-expressing cells in the fraction of CD45.2$^+$ transplanted cells in the bone marrow (BM) and the peripheral blood.



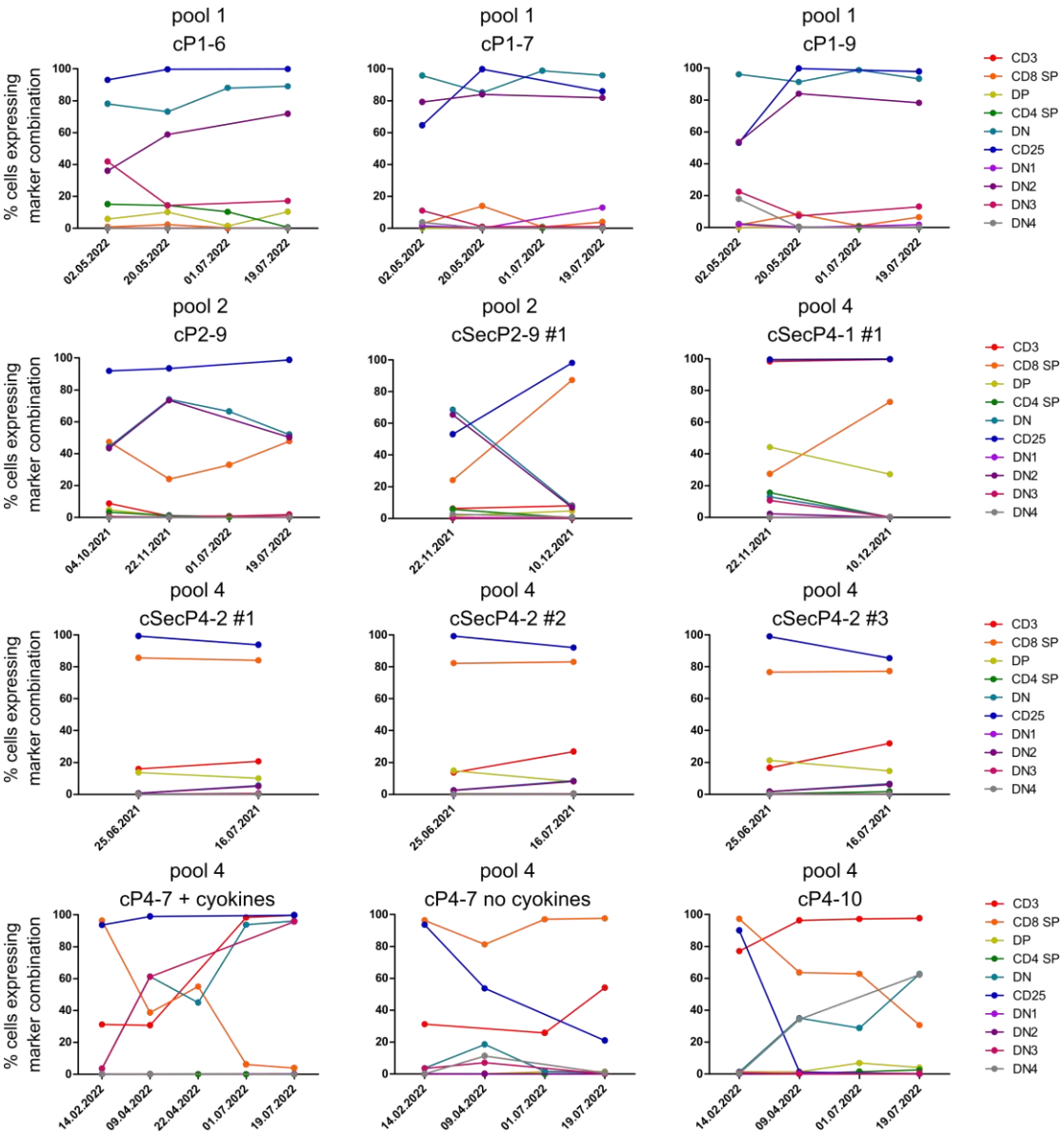
Supplementary Figure 8: Secondary T-ALL outliers downregulated different signaling pathways. (A) The arithmetic difference of the normalized and logarithmized counts per million mapped reads was calculated between all expressed genes from the primary T-ALL and the secondary T-ALL. Histogram summarizing the calculated value highlighted no change in gene expression for most genes (difference ~ 0) and a tendency towards more upregulated genes within the primary T-ALL. (B) Gene set enrichment analysis for deregulated KEGG-pathways on all positively enriched genes with a $\log_2$ difference ≥ 1. $p < 0.05$, false discovery rate > 0.05 for each identified pathway



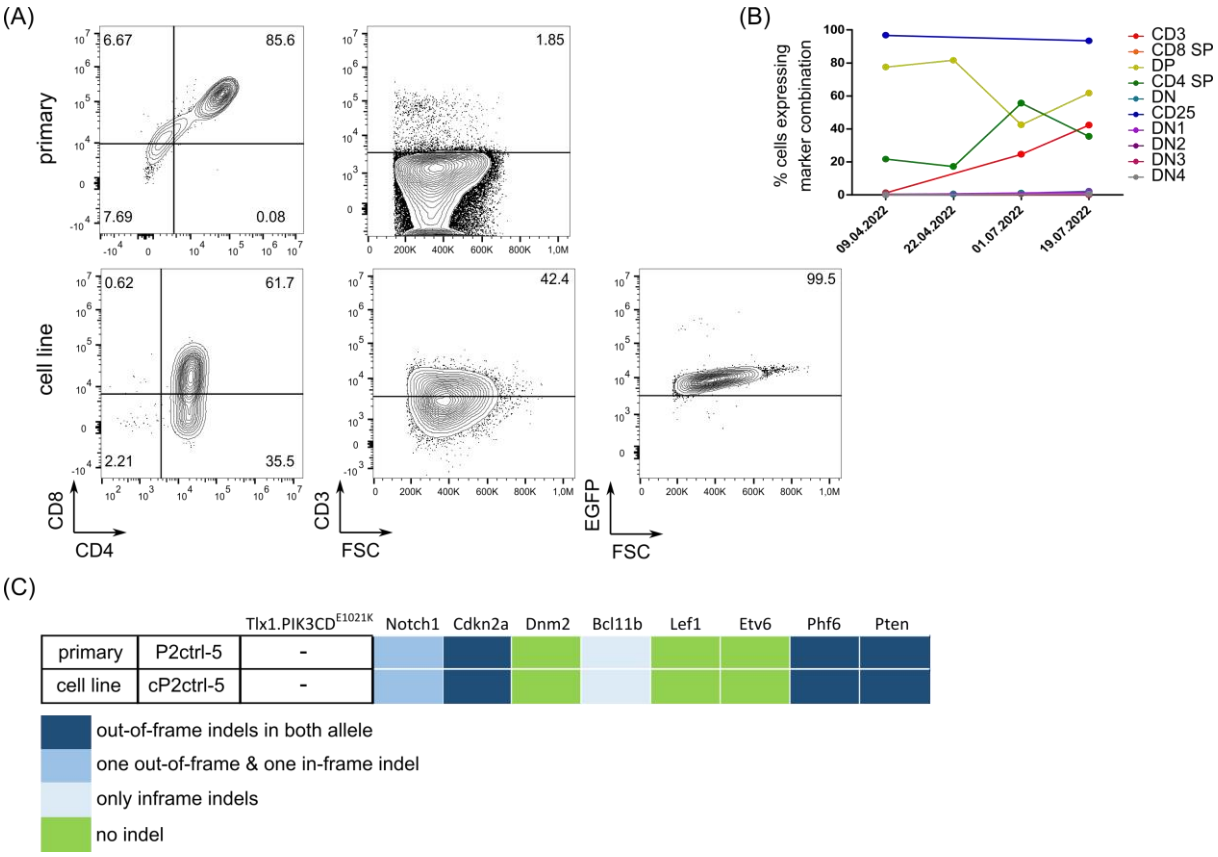
Supplementary Figure 9: Expression of the stemness regulator genes *Spi1*, *Itgax* and *Havcr2*. Unsupervised hierarchical clustering analysis on the expression of *Spi1*, *Itgax* and *Havcr2* ([267]) within healthy developing thymocytes (A) and the obtained T-ALL samples (B) visualized as heatmap. Depicted were normalized and $\log_2$-logarithmized counts per million mapped reads (lcpm). Color legends assign T-ALL samples to the transduced and transplanted lentivirus pool. DN – double-negative ($CD4^-CD8^-$), DP – double-positive ($CD4^+CD8^+$).



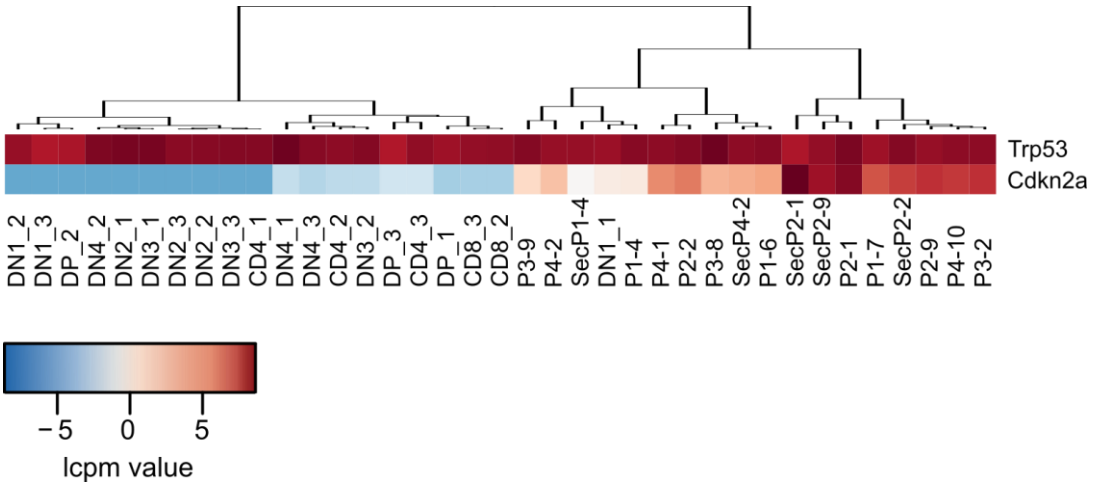
Supplementary Figure 10: Expression of genes involved in the interaction of the BM niche and leukemia cells. Unsupervised hierarchical clustering analysis on the expression of chosen genes interacting within the BM niche[278,281]. Comparison of the primary T-ALL samples and derived cell lines indicated a significant upregulation of those genes through the *in vitro* cultivation. Depicted are normalized and log_2-logarithmized counts per million mapped reads (lcpm).



Supplementary Figure 11: Stability of culturing T-ALL cell lines' immune phenotype. Cell lines that were obtained from primary or secondary (Sec) T-ALL samples were cultivated in co-culture with OP9/OP9-DL1 cells in RPMI, supplemented with 5 – 10 ng/mL mIL2 and mIL7. Progression of the percentages of $CD25^+$, $CD3^+$, $CD4^+$, $CD8^+$ single-positive (SP), $CD4^+CD8^+$ double-positive and $CD4^-CD8^-$ double-negative cells over time. DN cells were further categorized as DN1 ($CD44^+CD25^-$), DN2 ($CD44^+CD25^+$), DN3 ($CD44^-CD25^+$) and DN4 ($CD44^-CD25^-$).



Supplementary Figure 12: Characterization of the P2ctrl-5 sample. (A) Immune phenotype staining for flow cytometry on the primary P2ctrl-5 T-ALL sample and the obtained cell line (cP2ctrl-5). (B) Progression of the percentages of $CD25^+$, $CD3^+$, $CD4^+$, $CD8^+$ single-positive (SP), $CD4^+CD8^+$ double-positive and $CD4^-CD8^-$ double-negative cells over time. DN cells were further categorized as DN1 ($CD44^+CD25^-$), DN2 ($CD44^+CD25^+$), DN3 ($CD44^-CD25^+$) and DN4 ($CD44^-CD25^-$). (C) Presence of insertions and deletions (indel) within Cas9-targeted genes, verified by NGS analysis for primary T-ALL and TIDE analysis for the obtained cell line. Color legend is depicted.



Supplementary Figure 13: Expression of *Trp53* and *Cdkn2a* genes in. Unsupervised hierarchical clustering analysis on the expression of *Trp53* and *Cdkn2a* genes compared between healthy thymocyte populations and T-ALL samples. Depicted are normalized and log_2-logarithmized counts per million mapped reads (lcpm).

7.2 Supplementary Tables

Supplementary Table 1: Summary of primary transplanted animals

experimental group	number	ExpTime	end point disease
Pool 1 Tlx1.NRAS^{G12D} + sgRNAs	1	133	T-ALL
	2	121	T-ALL, lymphoma
	3	370	>1 year in experiment
	4	182	T-ALL
	5	116	T-ALL
	6	88	T-ALL
	7	84	T-ALL
	8	118	T-ALL
	9	83	T-ALL
	10	190	T-ALL, lymphoma
Pool 2 Tlx1.PIK3CD^{E1021K} + sgRNAs	1	61	T-ALL
	2	61	T-ALL
	3	308	sick
	4	56	undefined leukemia
	5	12	engraftment failure
	6	54	T-ALL
	7	69	T-ALL
	8	48	T-ALL
	9	58	T-ALL
	10	44	T-ALL
Pool 3 Tlx1.IL7R^{731ins732} + sgRNAs	1	28	engraftment failure
	2	188	T-ALL
	3	237	recipient AML
	4	375	>1 year in experiment
	5	166	recipient T-ALL
	6	97	BM failure
	7	38	engraftment failure
	8	132	T-ALL
	9	122	T-ALL
	10	189	T-ALL
Pool 4 Tal1.Lmo2 + PIK3CD^{E1021K} + sgRNAs	1	98	T-ALL
	2	68	T-ALL
	3	370	>1 year in experiment
	4	217	infection
	5	76	T-ALL
	6	84	T-ALL
	7	120	T-ALL
	8	83	T-ALL
	9	113	T-ALL
	10	85	T-ALL

experimental group	number	ExpTime	end point disease
Tlx1.NRAS^{G12D} only	1	88	Mixed lineages
	2	368	>1 year in experiment
	3	216	infection
	4	368	>1 year in experiment
	5	368	>1 year in experiment
	6	368	>1 year in experiment
Pool 1 only sgRNAs	1	187	recipient T-ALL
	2	220	BM failure
	3	>340	
	4	118	T-ALL
	5	113	T-ALL
	6	181	T-ALL
Tlx1.PIK3CD^{E1021K} only	1	19	engraftment failure
	2	383	>1 year in experiment
	3	383	>1 year in experiment
	4	383	>1 year in experiment
	5	383	>1 year in experiment
	6	383	>1 year in experiment
Pool 2 only sgRNAs	1	76	engraftment failure
	2	74	T-ALL
	3	287	chimerism < 0.1%
	4	71	T-ALL
	5	153	T-ALL
	6	75	T-ALL
Tal1.Lmo2 only	1	>225	
	2	>225	
	3	>225	
	4	>225	
	5	>225	
	6	172	BM failure
Pool 4 only sgRNAs	1	89	undefined lymphoma
	2	85	T-ALL
	3	90	Mixed lineages
	4	84	T-ALL
	5	104	T-ALL
	6	75	T-ALL
reporter	1	25	engraftment failure
	2	153	BM failure
	3	120	recipient T-ALL/AML
	4	166	BM failure
	5	203	recipient T-ALL
	6	41	engraftment failure
reporter 2	1	52	sacrified for thymus
	2	52	sacrified for thymus
	3	52	sacrified for thymus

Supplementary Table 2: Summary of primary transplanted animals

experimental group	number	ExpTime	end point disease
Pool 1 secondary P1-2	1	16	engraftment failure
	2	28	sick
	3	28	sick
Pool 1 secondary P1-4	1	41	T-ALL
	2	41	T-ALL
	3	41	T-ALL
Pool 2 secondary P2-1	1	13	T-ALL
	2	13	T-ALL
	3	13	found dead
Pool 2 secondary P2-2	1	13	T-ALL
	2	15	engraftment failure
	3	15	engraftment failure
Pool 2 secondary P2-9	1	12	T-ALL
	2	12	T-ALL
	3	12	T-ALL
Pool 2 secondary P2-10	1	15	T-ALL
	2	14	T-ALL
	3	13	T-ALL

experimental group	number	ExpTime	end point disease
Pool 3 secondary P3-2	1	38	BM failure
	2	>180	
	3	57	T-ALL
Pool 3 secondary P3-9	1	15	T-ALL
	2	15	T-ALL
	3	89	BM failure
Pool 4 secondary P4-1	1	16	T-ALL
	2	15	T-ALL
	3	16	T-ALL
Pool 4 secondary P4-2	1	13	T-ALL
	2	13	T-ALL
	3	13	T-ALL

Supplementary Table 3: Detailed summary of detected allele variations by NGS

The whole table can also be find attached in a hard drive.

Sample	Gene	REF	ALT	Predicted Effect	Details	Read Depth	Allele Freq		WT count
P1-1	Bcl11b	AGGTGA	AGAGGGTGA,AGAAGTGA	frameshift_variant & conservative_inframe_insertion	c.170_171insTT|p.Cys58fs; c.170_171insCCT|p.Thr57dup	3148	264	250	2616
	Notch1	CCATG	GCCTG,AGATTG	frameshift_variant&missense_variant	7338_7342delCATGGinsCAATCT|p.Met2447fs; c.7340_7342delTGGinsGGC|p.MetVal2447ArgLeu	2822	280	259	2241
	Lef1	TATGA	TTCCCTTCATGG,TCTCATGA	frameshift_variant&missense_variant	352_355delATGAinsTCCCTTCATGG|p.Met118fs; 351_352insCTC|p.Asn117_Met118insLeu	3803	304	280	3188
	Cdkn2a	CTG	CTCC,CTTG	frameshift_variant	c.294_295insA|p.Ser99fs; c.230_231insA|p.Ala78fs	1563	147	130	1278
	Dnm2	CGGT	CGGGG,CGGGCGT	frameshift_variant & disruptive_inframe_insertion	c.72delTinsGG|p.Gln25fs; c.71_72insGCG|p.Gly24_Gln25insArg	2053	196	115	1729
	Phf6	GAG	GAAG	frameshift_variant	c.267dupA|p.Ala90fs	2780	233		2523
P1-2	Lef1			unquantified insertion	+3AS	3461	135		
	Bcl11b			unquantified insertion	+6/+2AS	1952	102		
	Etv6			unquantified indel		2811	126		
	Dnm2			unquantified insertion		2075	64		
P1-4	Bcl11b	AGGTGAGTAGATCAG	AGGAGG,AGTGGTGAGTAGATCAG	frameshift_variant&disruptive_inframe_deletion	c.170_171insCA|p.Cys58fs; c.158_169delCTGATCTACTCAinsCCT|p.Asp54_Thr57delinsSer	2496	1311	1135	28
	Lef1	TAT	TAAAT,TCCAT	frameshift_variant	c.352_353insAA|p.Met118fs; c.351_352insCC|p.Met118fs	2798	1442	1296	40
	Etv6	CAAAGAGG	CAGGAGG,CAGGTGAAGAGG	frameshift_variant	c.169_170insGGTG|p.Lys57fs; c.298_299insGGTG|p.Lys100fs;	2530	1416	1023	30
	Dnm2	CGGTCA	CGGCCCTCT,CGGGTCA	missense_variant&disruptive_inframe_insertion	c.71dupG|p.Gln25fs; c.72_74delTCAinsCCCTCT|p.Gln25delinsProLeu	2423	1246	998	22
	Phf6	AGC	AGGGGAGGC	disruptive_inframe_insertion	c.268_269insGGGAGG|p.Gly89_Ala90insGlyGlu	1906	1840		36
P1-5	Phf6	AGC	AGGGC	frameshift_variant	c.268_269insGG|p.Ala90fs	2887	714		2166
P1-6	Bcl11b	GGT	GAAAAT,GGGCCCCGT	missense_variant&conservative_inframe_insertion	c.169_170delACinsATTTT|p.Thr57delinsIlePhe; c.169_170insCGGGGC|p.Thr57_Cys58insGlyPro	2349	1292	904	129
	Notch1	ACCATG	AATTG,ACCAGCTG	frameshift_variant	c.7338_7342delCATGGinsCAAT|p.Met2447fs; c.7339_7340insGC|p.Met2447fs	2355	1278	881	138
	Lef1	CAATATGAAC	CAAAGATAG,CAATCCATGAAC	frameshift_variant	c.351_357delTATGAACinsAGATAG|p.Asn117fs; c.351_352insCC|p.Met118fs	3080	1527	1354	144
	Cdkn2a	CCGCGCTGCGTC	CC	frameshift_variant	c.290_299delGACGCAGCGC|p.Arg98fs	1516	1457		58
P1-7	Bcl11b	GG	GCC,GAGGGCCCCCT	missense_variant&conservative_inframe_insertio	c.170delCinsGG|p.Thr57fs; |c.170delCinsAGGGGGCCCT|p.Thr57delinsLysGlyAlaLeu	2071	909	876	261
	Notch1	CATGGGTGGGACCATGGAGGATGGCAGTGATGTGGGCAG	CATGGGTCATG,CTGGAGGATGGCAGTGATGTGGGCAG	frameshift_variant	c.7340_7352delTGGTCCCACCCAT|p.Met2447fs; c.7315_7346delCTGCCCACATCACTGCCATCCTCCATGGTCCCinsCATG|p.Leu2439fs	1673	814	616	221
	Cdkn2a	CGCTG	CCGAGGGGTCCC	frameshift_variant	c.294_297delCAGCinsGGGACCCCTCG|p.Ser99fs	1485	1007		443
	Dnm2	CGGTC	CGAGGTTC,CGGCACTC	disruptive_inframe_insertion	c.71delGinsAGGT|p.Gly24delinsGluVal; c.71_72insCAC|p.Gly24_Gln25insThr	2197	955	851	246
P1-8	Bcl11b	AGGTGAGTA	AGT,AGAAAGTGAGTA	disruptive_inframe_insertion	c.164_170delTACTCACinsA|p.Leu55_Thr57delinsHis; c.170_171insTTT|p.Thr57_Cys58insPhe	988	302	350	30
	Notch1	ATG	ATTTG,AGTTTTG	frameshift_variant	c.7339_7340insAAAC|p.Met2447fs; c.7338_7339insAA|p.Met2447fs	865	268	183	28
	Lef1	TATG	TATGCCCTG,TTGG	frameshift_variant	c.354_355insCCCTG|p.Asn119fs; c.156_157insCCCTG|p.Asn53fs	1049	256	329	30
	Cdkn2a	CTG	CTTG,CTGATG	frameshift_variant & conservative_inframe_insertion	c.294_295insA|p.Ser99fs; c.229_230insCAT|p.Ala77_Ala78insSer	618	289	185	21
	Etv6	ACCAAAGAGGAT	AGGAAGAGGAT,ACCTACGGAT	frameshift_variant	c.296_298delCCAinsGG|p.Thr99fs; c.169_173delAAAGAinsTAC|p.Lys57fs	1039	302	267	29
	Dnm2	CGGTC	CGGCCC,CGGGC	heterozygous_frameshift	c.72T>G|p.Gly24Gly; c.72delTinsCC|p.Gln25fs	812	268	182	24
	Phf6	GAGCA	GAAAGCA,GAGGTCTCCA	frameshift_variant	c.27_28insAA|p.Ala10fs; c.28_29insGTCTC|p.Ala10fs	864	511	316	29
P1-9	Bcl11b	GGT	GTAGCCCCGTCGT,GTTAACCGT	frameshift_variant & disruptive_inframe_insertion	c.170_171insGACGGGGCTA|p.Cys58fs; c.170_171insGGTTAA|p.Thr57_Cys58insValAsn	2236	894	973	334
	Notch1	GACCATG	GCCTG,GACCACGTG	frameshift_variant	c.7338_7343delCATGGTinsCAGG; c.7339_7340insCG|p.Met2447fs	1738	747	685	248
	Lef1	ATA	ATGGCT,ATGGAGTTA	disruptive_inframe_insertion	c.352delAinsGGCT|p.Met118delinsGlyLeu; c.351_352insGGAGTT|p.Asn117_Met118insGlyVal	3184	1506	1155	490
	Cdkn2a	CGCTGCGT	CGCGA	disruptive_inframe_insertion	c.227_231delACGCAinsTC|p.Asp76_Ala77delinsVal	2131	1546		545
	Dnm2	GGT	GGCCTC,GAGAGT	disruptive_inframe_insertion	c.72delTinsCCTC|p.Gly24_Gln25insLeu; c.70_71insAGA|p.Gly24delinsGluSer	2712	1263	1067	324
P1-10	--								
P2-1	Bcl11b	AGGTGA	AGGAGTAGATCAGGA,AGGGGGTGA	conservative_inframe_insertion	c.167_169delTCAinsTCCTGATCTACT|p.Thr57delinsLeulleTyrSer; c.169_170insCCC|p.Thr57_Cys58insPro	2150	1183	847	81
	Pten	TGTGT	TGGTGT,TGTGGGT	frameshift_variant	c.745dupG|p.Val249fs; c.747_748insGG|p.Cys250fs	2784	1267	1233	172
	Notch1	CC	CAAG	frameshift_variant	c.7341delGinsCTT|p.Met2447fs	2103	1146		910
	Notch1	ATG	AGGTG	frameshift_variant	c.7339_7340insCC|p.Met2447fs	2081	600		1271
	Cdkn2a	CGCGCTGCG	CGCG,CGCGCTTGCG	frameshift_variant	c.294_295insA|p.Ser99fs; c.230_231insA|p.Ala78fs	1617	689	661	89
	Etv6	CTGACCAAAGAGGAT	CTGACCAGGG,CAAGAGGAT	frameshift_variant & disruptive_inframe_insertion	c.299_306delAAGAGGATinsGGG|p.Lys100fs; c.293_298delTGACCA|p.Leu98_Lys100delinsGln	2674	1275	1240	138
	Dnm2	GGTC	GAAGAGTC,GGGGGCCTAATTC	frameshift_variant & disruptive_inframe_insertion	c.70_71insAAGA|p.Gly24fs; c.71_72insGGGCCTAAT|p.Gly24_Gln25insGlyLeulle	2320	1248	942	113
P2-2	Bcl11b	ACAGGTGAGT	ACAGGAAGTGAGT,ACAGGATGTGAGT	conservative_inframe_insertion	c.169_170insCTT|p.Thr57_Cys58insSer; c.169_170insCAT|p.Thr57_Cys58insSer	1285	510	421	92
	Pten	GTGT	GAATGT,GTGGGT	frameshift_variant	c.745_746insAA|p.Val249fs; c.747_748insGG|p.Cys250fs	1448	537	573	97
	Notch1	GTGGGACCATG	GTG,GTGGGACCATCCTCTG	frameshift_variant	c.7338_7345delCATGGTCC|p.Met2447fs; c.7338_7339insAGAGG|p.Met2447fs	1196	420	429	67
	Lef1	ATAT	ATTAT,ATGGAT	frameshift_variant	c.351dupT|p.Met118fs; c.351_352insGG|p.Met118fs	1584	616	538	119
	Cdkn2a	CGCTG	CG,CGCTTG	frameshift_variant & disruptive_inframe_insertion	c.294_296delCAG|p.Ser99del; |c.294_295insA|p.Ser99fs	951	364	423	78
	Etv6	CAAAGAGGA	CAGGTC,CAGAAGGGCTAAGAGGA	frameshift_variant & disruptive_inframe_insertion	c.298_299insGAAGGGCT|p.Lys100fs; c.299_305delAAGAGGAinsGGTC|p.Lys100_Asp102delinsArgSer	1495	729	415	117
	Dnm2	CGGTC	CGGGGGTC	disruptive_inframe_insertion	c.71_72insGGG|p.Gly24dup	943	778		70

Sample	Gene	REF	ALT	Predicted Effect	Details	Read Depth	Allele Freq		WT count
P2-6	Bcl11b	GGTG	GTTA,GAAAAGTG	frameshift_variant & disruptive_inframe_insertion	c.170_171insTTTT|p.Cys58fs; c.168_170delCACinsTAA|p.Thr57Asn	1554	761	684	92
	Pten	TGT	TCCGT	frameshift_variant	c.746_747insCC|p.Cys250fs	1266	1097		146
	Notch1	ATG	AGGCCTTG	frameshift_variant	c.7339_7340insAGGCC|p.Met2447fs	1190	982		168
	Cdkn2a	GCTG	GCTCCCTG,GCCCTTG	conservative_inframe_insertion	c.230_231insAGGG|p.Ala78fs; c.295_296insAGG|p.Ser99delinsLysGly (c.231_232insAGG|p.Ala77_Ala78insArg)	1288	560	614	87
P2-7	Bcl11b	AGGTGA	AGGTGGTGA	conservative_inframe_insertion	c.167_168insCAC|p.Leu56_Thr57insThr	1936	1628		223
	Pten	ATTGCCTGTGTGTGGT	AGT,ATTGCCTGTGTGGT	frameshift_variant	c.739_751delTTGCCTGTGTGTG|p.Leu247fs; c.750_751delTG|p.Cys250fs	2650	1130	1268	123
	Notch1	CCATG	GCAGTG	frameshift_variant	c.7338_7342delCATGGinsCACTGC|p.Met2447fs	2582	2278		242
	Cdkn2a	CTG	CTTG	heterozygous_frameshift_variant	c.230_231insA|p.Ala78fs	1776	246		896
P2-8	Bcl11b	GGT	GAGAGT	disruptive_inframe_insertion	c.170_171insTCT|p.Thr57dup	2040	360		1264
	Pten	TGTGT	CGAGCATGT	frameshift_variant	c.744_745delTGinsCGAGCA|p.Val249fs	2855	583		1682
	Notch1	ACCA	AGGGGA	frameshift_variant	c.7340_7342delTGGinsTCCCC|p.Met2447fs	1515	327		953
	Cdkn2a	GC	TG	missense_Variant	c.293_294delGCinsCA|p.Arg98Pro; c.229_230delGCinsCA|p.Ala77Gln	1550	532		974
	Etv6	TGACCAAAG	TG	frameshift_variant	c.296_302delCCAAAGA|p.Thr99fs; c.167_173delCCAAAGA|p.Thr56fs	2335	139		1859
	Dnm2	TCCATCGGTC	CCCCTTCGGCCGTC	frameshift_variant	c.64_70delTCCATCGinsCCCCTTCGGCC|p.Ser22fs	1976	366		1252
P2-9	Bcl11b	CAGGTGA	CAGGGGGAGGTGA,CAGGTGGTGA	conservative_inframe_insertion	c.169_170insCCTCCC|p.Thr57_Cys58insSerPro; c.167_168insCAC|p.Leu56_Thr57insThr	2091	979	1076	16
	Pten	TGTGT	TGGGAGGGTGT,TGTGGGT	frameshift_variant & disruptive_inframe_insertion	c.747_748insGG|p.Cys250fs; c.745_746insGGAGGG|p.Pro248_Val249insGlyArg	2608	1113	1314	50
	Notch1	TGGGACCATGGAG	TG,TGGGACCAAGGCCTGGAG	frameshift_variant	c.7335_7345delCTCCATGGTCC|p.Ser2446fs; c.7339_7340insGGCCT|p.Met2447fs	1946	1028	830	18
	Lef1	TATG	TCC,TCCATG	frameshift_variant	c.351_352insCC|p.Met118fs; c.352_354delATGinsCC|p.Met118fs;	2444	1220	1178	33
	Cdkn2a	GCTG	GCCTG,GCTTG	frameshift_variant	c.295_296insG|p.Ser99fs; c.294_295insA|p.Ser99fs	1788	726	981	52
	Etv6	AAAGAGG	AGGAAGAGG,ACTGTTAAGAGG	frameshift_variant	c.298_299insCTGTT|p.Lys100fs; c.298_299insGG|p.Lys100fs	2877	1363	1437	40
	Dnm2	CGGTC	CGGGGTC,CGGCGTTC	frameshift_variant & disruptive_inframe_insertion	c.70_71dupGG|p.Gln25fs; c.71_72insCGT|p.Gly24_Gln25insVal	2274	1128	1094	27
P2-10	Bcl11b	AGGTGA	AGCAAGTGA,AGCTTGTGA	disruptive_inframe_insertion	c.170_171insTTG|p.Thr57_Cys58insCys; c.170_171insAAG|p.Thr57_Cys58insSer	2083	779	730	73
	Pten	GCCATTGCCTGTGT	GCCATTGCCTGGGGGT,GGGGGT	frameshift_variant	c.736_746delCCATTGCCTGTinsGGG|p.Pro246fs; c.746delTinsGGG|p.Val249fs	3352	1759	540	1
	Notch1	GGACCATGGAGGATGGCAGTG	GGACCAACCGGTGGAGGATGGCAGTG,GGACCACCGCTTGGAGGATG	frameshift_variant	c.7339_7340insCCGGT|p.Met2447fs; c.7339_7340insAGCGG|p.Met2447fs	1571	286	501	64
	Cdkn2a	CTGCG	CCTGGTG,CTAGCCTGAGGG	frameshift_variant	c.292_295delCGCAinsCACCAG|p.Arg98fs; c.292_295delCGCAinsCACCAG|p.Arg98fs;	2650	918	1019	1
P2ctrl-5	Bcl11b	CAGGT	CAGCCGCT,CT	disruptive_inframe_deletion	c.169_170delACinsAGCGG|p.Thr57delinsSerGly; c.170_172delCCT|p.Thr57_Cys58delinsSer	3788	1881	1633	239
	Pten	CTGTGT	CGTGGGGGGGT,CTGGGGGT	frameshift_variant	c.744_746delTGTinsGTGGGGGG|p.Val249fs; c.746delTinsGGG|p.Val249fs	4512	1366	2255	31
	Notch1	ACCATG	CCTTTCCTG,ACCAACCTTG	frameshift_variant & disruptive_inframe_insertion	c.7339_7340insAGGT|p.Met2447fs; c.7338_7343delCATGGTinsCAGGAAAGG|p.Met2447_Val2448delinsArgLysGly	3241	1630	1270	195
	Cdkn2a	CTG	CGTG,CTTG	frameshift_variant	c.295_296insC|p.Ser99fs; c.230_231insA|p.Ala78fs	2634	1218	1179	179
	Phf6	AGC	AGGC,ACCGC	frameshift_variant	c.267_268insCC|p.Ala90fs; c.28dupG|p.Ala10fs	3321	1664	1435	192
P4-1	Pten	GTG	GGGCTG	disruptive_inframe_insertion	c.745_746insGGC|p.Val249delinsGlyLeu	2246	1725		482
	Notch1	CCATG	CGGTTTG,CCACTCTG	frameshift_variant & disruptive_inframe_insertion	c.7338_7341delCATGinsCAAACC|p.Met2447fs; c.7339_7340insGAG|p.Met2447delinsArgVal	1561	746	587	164
P4-2	Pten	GCCTGTGT	GGGGGT,GCCTGTTCACCCGT	frameshift_variant & disruptive_inframe_insertion	c.742_746delCCTGTinsGGG|p.Pro248fs; c.746_747insTCACCC|p.Val249_Cys250insHisPro	2521	1032	800	351
	Notch1	GGGACCATG	GGGACCGGCTGTG,GGGACCAATG	frameshift_variant	c.7338_7340delCATinsCACAGCC|p.Met2447fs; c.7339_7340insT|p.Met2447fs	1782	794	584	226
	Cdkn2a	CGCTG	CGCTCCT	frameshift_variant	c.294delCinsAGG|p.Ser99fs	1561	1020		366
P4-5	Pten	GTG	GCCGATG,GAGTAGATTTTG	disruptive_inframe_insertion	c.745_746insCCGA|p.Val249fs; c.745_746insAGTAGATTT|p.Val249delinsGluTerIleLeu	3752	1346	1331	711
	Notch1	GTA	GTAGTGGTAATTTA	frameshift_variant	c.7358_7359insTAAATTACCAC|p.Thr2454fs	3343	1171		2153
	Notch1	CCATG	CGGGGTGTG,CCACAAGATCTTG	frameshift_variant	c.7338_7341delCATGinsCACACCCC|p.Met2447fs; c.7339_7340insAGATCTTG|p.Met2447fs	3399	1333	1180	354
	Cdkn2a	CGCGCTGCG	CGCGCTGCCTGCG	frameshift_variant	c.292_293insGCAG|p.Ser99fs	2177	1249		514
P4-6	Pten	TGT	TGTGAGT	frameshift_variant	c.747_748insAGTG|p.Cys250fs	2892	2066		800
	Lef1	TAT	TCTCAT,TCCTCAT	frameshift_variant & disruptive_inframe_insertion	c.351_352insCTC|p.Asn117_Met118insLeu; c.351_352insCCTC|p.Met118fs	3507	1561	1442	474
P4-7	Notch1	ACCATG	ACTTG,ACCAGGTGTG	frameshift_variant	c.7338_7341delCATGinsCAA|p.Met2447fs; c.7339_7340insCACC|p.Met2447fs	2256	1174	826	181
	Cdkn2a	CTG	CTTG	frameshift_variant	c.294_295insA|p.Ser99fs	2064	536		1519
P4-8	Pten	GTGTG	GTAC,GGAACTGTG	frameshift_variant	c.745_746insGAAC|p.Val249fs; c.747_749delGTGinsAC|p.Cys250fs	2798	1312	1025	413
	Notch1	CATG	CTTTG,CATGCTTTTAGTTG	frameshift_variant	c.7338_7340delCATinsCAAA|p.Met2447fs; c.7337_7338insCAACTAAAAG|p.Met2447fs	1370	625	501	217
	Cdkn2a	GCT	GCCT	frameshift_variant	c.295_296insG|p.Ser99fs	1855	911		606
P4-9	Pten	TGTG	TGCGGGGTG,TGTTG	frameshift_variant	c.745_746insCGGGG|p.Val249fs; c.746dupT|p.Cys250fs	2560	925	1299	270
	Notch1	CCAT	CCCTCTT,CCAAAT	frameshift_variant	c.7339_7340insTT|p.Met2447fs; c.7339_7340delATinsAAGAG	2399	1090	1021	217
	Cdkn2a	CTG	CTCC,CTTG	frameshift_variant	c.294_295insA|p.Ser99fs; c.294delCinsGG|p.Ser99fs	1834	812	768	218

Sample	Gene	REF	ALT	Predicted Effect	Details	Read Depth	Allele Freq		WT count
P4-10	Pten	TGTGT	TGGGGGTGT,TGTCCGT	frameshift_variant	c.745_746insGGGG|p.Val249fs; c.746_747insCC|p.Cys250fs	3376	1287	1613	319
	Notch1	GACC	GAAAT	frameshift_variant	c.7341_7342delGGinsATT|p.Met2447fs	4008	1838		2057
	Notch1	ATG	AACATTG	frameshift_variant	c.7339_7340insATGT|p.Met2447fs	3641	1385		2199
	Cdkn2a	GCT	GCACCGG,GCCTCCT	frameshift_variant	c.295_296insGGAG|p.Ser99fs; c.295delAinsCCGGT|p.Ser99fs	1628	777	593	220
P3-2	Bcl11b	GGT	GAGACTAGT	disruptive_inframe_insertion	c.170_171insTAGTCT|p.Thr57_Cys58insSerLeu	1786	1671		101
	Ptpn2	AATG	AG	frameshift_variant	c.644_645delAT|p.His215fs	2816	2498		313
	Cdkn2a	GGA	GCGA	frameshift_variant	c.310_311insG|p.Pro104fs	1948	845		1096
	Dnm2	CGGTC	CGGGGGTC,CGTGTTTTTGTC	frameshift_variant & disruptive_inframe_insertion	c.70_71insTGTTTTT|p.Gly24fs; c.71_72insGGG|p.Gly24dup	1472	695	660	91
P3-3	--								
P3-5	--								
P3-8	Notch1	CATG	CAACTATG,CACCTCCTG	frameshift_variant	c.7339_7340insTAGT|p.Met2447fs; c.7339_7340insGGAGG|p.Met2447fs	1667	801	477	66
	Cdkn2a	CTG	CGGTCGTG	frameshift_variant	c.295_296insCGACC|p.Ser99fs	1544	1347		161
	Etv6	CAAAGAGGAT	CAGGGAGAAAGAGGAT	disruptive_inframe_insertion	c.298_299insGGGAGA|p.Thr99_Lys100insArgGlu	1611	1409		178
	Dnm2	ACGCCTTCAGCTCCATCGGTCAGAGCT	ACGCCTTCAGCTCCATCCAGAACCGTCAGAGCT,AG	frameshift_variant & disruptive_inframe_insertion	c.54_79delCGCCTTCAGCTCCATCGGTCAGAGCTinsG|p.Asp18fs; c.70delGinsCAGAACC|p.Gly24delinsGlnAsnArg	1620	765	772	49
	Phf6	GAGC	GAAGC,GAGCGCGC	frameshift_variant	c.267dupA|p.Ala90fs; c.29_30insGCGC|p.Thr11fs	2728	1458	1153	47
P3-9	Notch1	GGACCATGGAGG	GGGAGG,GGACCAG	frameshift_variant & conservative_inframe_deletion	c.7335_7339delCTCCA|p.Ser2446fs; |c.7338_7343delCATGGT|p.Met2447_Val2448del	1430	622	520	238
	Lef1	ATGATGCCCAATATG	AA	frameshift_variant	c.341_354delTGATGCCCAATATGinsA|p.Met114fs	2543	1941		587
P3-10	Cdkn2a	GCT	GCTGTGCCCT	frameshift_variant	c.294_295insAGGGCAC|p.Ser99fs	2262	667		1586
	Cdkn2a	GC	GGCG	frameshift_variant	c.293delGinsCGC|p.Arg98fs	2438	784		1637

7.3 References

1. Seita, J. & Weissman, I. L. Hematopoietic stem cell: self- renewal versus differentiation. *WIREs Syst. Biol. Med.* **2**, 640–653 (2010).
2. Nakamura-Ishizu, A., Takizawa, H. & Suda, T. The analysis, roles and regulation of quiescence in hematopoietic stem cells. *Dev.* **141**, 4656–4666 (2014).
3. Yoshihara, H. *et al.* Thrombopoietin/MPL Signaling Regulates Hematopoietic Stem Cell Quiescence and Interaction with the Osteoblastic Niche. *Cell Stem Cell* **1**, 685–697 (2007).
4. Méndez-Ferrer, S. *et al.* Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* **466**, 829–834 (2010).
5. Metcalf, D. ASH 50th anniversary review Hematopoietic cytokines. *Blood* **111**, 485–491 (2009).
6. Ding, L., Saunders, T. L., Enikolopov, G. & Morrison, S. J. Endothelial and perivascular cells maintain haematopoietic stem cells. *Nat. 2012 4817382* **481**, 457–462 (2012).
7. Gary Gilliland, D. & Griffin, J. D. The roles of FLT3 in hematopoiesis and leukemia. *Blood* **100**, 1532–1542 (2002).
8. De Graaf, C. A. & Metcalf, D. Thrombopoietin and hematopoietic stem cells. *Cell Cycle* **10**, 1582–1589 (2011).
9. Oguro, H., Ding, L. & Morrison, S. J. SLAM Family Markers Resolve Functionally Distinct Subpopulations of Hematopoietic Stem Cells and Multipotent Progenitors. *Cell Stem Cell* **13**, 102–116 (2013).
10. Lécuyer, E. & Hoang, T. SCL: From the origin of hematopoiesis to stem cells and leukemia. *Exp. Hematol.* **32**, 11–24 (2004).
11. Yokomizo, T. *et al.* Hlf marks the developmental pathway for hematopoietic stem cells but not for erythro-myeloid progenitors. *J. Exp. Med.* **216**, 1599–1614 (2019).
12. Chambers, J. & Rabbitts, T. H. LMO2 at 25 years: A paradigm of chromosomal translocation proteins. *Open Biol.* **5**, (2015).
13. Vicente, C., Conchillo, A., García-Sánchez, M. A. & Odero, M. D. The role of the GATA2 transcription factor in normal and malignant hematopoiesis. *Crit. Rev. Oncol. Hematol.* **82**, 1–17 (2012).
14. Alharbi, R. A., Pettengell, R., Pandha, H. S. & Morgan, R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* **27**, 1000–1008 (2013).
15. Lampreia, F. P., Carmelo, J. G. & Anjos-Afonso, F. Notch Signaling in the Regulation of Hematopoietic Stem Cell. *Curr. Stem Cell Reports* **3**, 202–209 (2017).
16. Cabezas-Wallscheid, N. *et al.* Vitamin A-Retinoic Acid Signaling Regulates Hematopoietic Stem Cell Dormancy. *Cell* **169**, 807-823.e19 (2017).
17. Shao, L. *et al.* Hematopoietic stem cell senescence and cancer therapy-induced long-term bone marrow injury. *Transl. Cancer Res.* **2**, 397–411 (2013).
18. Kondo, M., Weissman, I. L. & Akashi, K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* **91**, 661–672 (1997).
19. Akashi K, Traver D, Miyamoto T & IL, W. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* **404**, 193–197 (2000).
20. Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science (80-.).* **351**, 1–22 (2016).
21. Muller-Sieburg, C. E., Cho, R. H., Karlsson, L., Huang, J. F. & Sieburg, H. B. Myeloid-biased hematopoietic stem cells have extensive self-renewal capacity but generate diminished lymphoid progeny with impaired IL-7 responsiveness. *Blood* **103**, 4111–4118 (2004).
22. Nestorowa, S. *et al.* A single-cell resolution map of mouse hematopoietic stem and progenitor cell differentiation. *Blood* **128**, e20–e31 (2016).
23. Wilson, N. K. *et al.* Combined Single-Cell Functional and Gene Expression Analysis Resolves Heterogeneity within Stem Cell Populations. *Cell Stem Cell* **16**, 712–724 (2015).

24. Hosokawa, H. & Rothenberg, E. V. How transcription factors drive choice of the T cell fate. *Nat. Rev. Immunol.* **21**, 162–176 (2021).

25. Allman, D. *et al.* Thymopoiesis independent of common lymphoid progenitors. *Nat. Immunol. 2003 42* **4**, 168–174 (2003).

26. Zhou, W. *et al.* Single-Cell Analysis Reveals Regulatory Gene Expression Dynamics Leading to Lineage Commitment in Early T Cell Development. *Cell Syst.* **9**, 321-337.e9 (2019).

27. Hosokawa, H. & Rothenberg, E. V. Cytokines, Transcription Factors, and the Initiation of T-Cell Development. *Cold Spring Harb. Perspect. Biol.* **10**, 1–20 (2018).

28. Erarslan- Uysal, B. *et al.* Chromatin accessibility landscape of pediatric T- lymphoblastic leukemia and human T- cell precursors. *EMBO Mol. Med.* **12**, 1–14 (2020).

29. Gegonne, A. *et al.* Immature CD8 Single-Positive Thymocytes Are a Molecularly Distinct Subpopulation , Selectively Dependent on BRD4 for Their Differentiation Article Immature CD8 Single-Positive Thymocytes Are a Molecularly Distinct Subpopulation , Selectively Dependent on . *Cell Rep.* **24**, 117–129 (2018).

30. Seet, C. S. *et al.* Generation of mature T cells from human hematopoietic stem and progenitor cells in artificial thymic organoids. *Nat. Methods* **14**, 521–530 (2017).

31. Xiong, J., Armato, M. A. & Yankee, T. M. Immature single-positive CD8+ thymocytes represent the transition from Notch-dependent to Notchindependent T-cell development. *Int. Immunol.* **23**, 55–64 (2011).

32. Mitchell, J. L., Seng, A. & Yankee, T. M. Ikaros, Helios, and Aiolos protein levels increase in human thymocytes after β selection. *Immunol. Res.* **64**, 565–575 (2016).

33. Yui, M. A. & Rothenberg, E. V. Developmental gene networks: A triathlon on the course to T cell identity. *Nat. Rev. Immunol.* **14**, 529–545 (2014).

34. Brodeur, J. F., Li, S., Damlaj, O. & Dave, V. P. Expression of fully assembled TCR–CD3 complex on double positive thymocytes: synergistic role for the PRS and ER retention motifs in the intra-cytoplasmic tail of CD3ε. *Int. Immunol.* **21**, 1317–1327 (2009).

35. Klein, L., Kyewski, B., Allen, P. M. & Hogquist, K. A. Positive and negative selection of the T cell repertoire: what thymocytes see (and don't see). *Nat. Rev. Immunol.* **14**, 377–391 (2014).

36. Germain, R. N. t-cell development and the CD4-CD8 lineage decision. *Nat. Rev. Immunol.* **2**, 309–322 (2002).

37. Rothenberg, E. V., Moore, J. E. & Yui, M. A. Launching the T-cell-lineage developmental programme. *Nat. Rev. Immunol.* **8**, 9–21 (2008).

38. de Pooter, R. & Zúñiga-Pflücker, J. C. T-cell potential and development in vitro: the OP9-DL1 approach. *Curr. Opin. Immunol.* **19**, 163–168 (2007).

39. Holmes, R. & Zuñiga-Pflücker, J. C. The OP9-DL1 system: Generation of T-lymphocytes from embryonic or hematopoietic stem cells in vitro. *Cold Spring Harb. Protoc.* **4**, 1–12 (2009).

40. Koniaeva, E. *et al.* Conditional Immortalization of Lymphoid Progenitors via Tetracycline-Regulated LMO2 Expression. *Hum. Gene Ther.* **31**, 183–198 (2020).

41. Deftos, M. L. & Bevan, M. J. Notch signaling in T cell development. *Curr. Opin. Immunol.* **12**, 166–72 (2000).

42. Mazzucchelli, R. & Durum, S. K. Interleukin-7 receptor expression: Intelligent design. *Nat. Rev. Immunol.* **7**, 144–154 (2007).

43. So, L. & Fruman, D. A. PI3K signalling in B- and T-lymphocytes: new developments and therapeutic advances. *Biochem. J.* **442**, 465–481 (2012).

44. Lapinski, P. E. & King, P. D. Regulation of Ras signal transduction during T cell development and activation. *Am. J. Clin. Exp. Immunol.* **1**, 147–153 (2012).

45. Wuebbles, A. A. *et al.* Stimulation of p21 ras upon T -cell activation. **327**, 324–327 (1990).

46. Koch, U. *et al.* Delta-like 4 is the essential, nonredundant ligand for Notch1 during thymic T cell lineage commitment. *J. Exp. Med.* **205**, 2515–2523 (2008).

47. Palomero, T. *et al.* Mutational loss of PTEN induces resistance to NOTCH1 inhibition in T-cell leukemia. *Nat. Med.* **13**, 1203 (2007).

48. Radtke, F., Fasnacht, N. & MacDonald, H. R. Notch Signaling in the Immune System. *Immunity* **32**, 14–27 (2010).

49. King, B. *et al.* The ubiquitin ligase FBXW7 modulates leukemia-initiating cell activity by regulating MYC stability. *Cell* **153**, 1552 (2013).

50. Liu, Y. *et al.* The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* **49**, 1211–1218 (2017).

51. Oliveira, M. L., Akkapeddi, P., Ribeiro, D., Melão, A. & Barata, J. T. IL-7R-mediated signaling in T-cell acute lymphoblastic leukemia: An update. *Adv. Biol. Regul.* **71**, 88–96 (2019).

52. Lodewijckx, I. & Cools, J. Deregulation of the interleukin-7 signaling pathway in lymphoid malignancies. *Pharmaceuticals* **14**, (2021).

53. de Bock, C. E. *et al.* HOXA9 cooperates with activated JAK/STAT signaling to drive leukemia development. *Cancer Discov.* **8**, 616–631 (2018).

54. Vanden Bempt, M. *et al.* Cooperative Enhancer Activation by TLX1 and STAT5 Drives Development of NUP214-ABL1/TLX1-Positive T Cell Acute Lymphoblastic Leukemia. *Cancer Cell* **34**, 271-285.e7 (2018).

55. Singer, A. & Bosselut, R. CD4/CD8 coreceptors in thymocyte development, selection, and lineage commitment: Analysis of the CD4/CD8 lineage decision. *Adv. Immunol.* **83**, 91–131 (2004).

56. Hwang, J. R., Byeon, Y., Kim, D. & Park, S. G. Recent insights of T cell receptor-mediated signaling pathways for T cell activation and development. *Exp. Mol. Med.* **52**, 750–761 (2020).

57. Arozarena, I., Calvo, F. & Crespo, P. Ras, an actor on many Stages: Posttranslational modifications, Localization, and Site-Specified Events. *Genes and Cancer* **2**, 182–194 (2011).

58. Atsaves, V., Leventaki, V., Rassidakis, G. Z. & Claret, F. X. AP-1 transcription factors as regulators of immune responses in cancer. *Cancers (Basel).* **11**, (2019).

59. Alessi, D. R. *et al.* Characterization of a 3-phosphoinositide-dependent protein kinase which phosphorylates and activates protein kinase Bα. *Curr. Biol.* **7**, 261–269 (1997).

60. Huang, J. & Manning, B. D. A complex interplay between Akt, TSC2, and the two mTOR complexes. *Biochem. Soc. Trans.* **37**, 217 (2009).

61. Laplante, M. & Sabatini, D. M. mTOR signaling in growth control and disease. *Cell* **149**, 274 (2012).

62. Sun, K. *et al.* The PI3K/AKT/mTOR signaling pathway in osteoarthritis: a narrative review. *Osteoarthr. Cartil.* **28**, 400–409 (2020).

63. Xiao, L., Wang, Y. C., Li, W. S. & Du, Y. The role of mTOR and phospho-p70S6K in pathogenesis and progression of gastric carcinomas: An immunohistochemical study on tissue microarray. *J. Exp. Clin. Cancer Res.* **28**, 1–9 (2009).

64. Leal, P. *et al.* AKT/mTOR substrate P70S6K is frequently phosphorylated in gallbladder cancer tissue and cell lines. *Onco. Targets. Ther.* **6**, 1373 (2013).

65. Bardelli, V. *et al.* T-cell acute lymphoblastic leukemia: Biomarkers and their clinical usefulness. *Genes (Basel).* **12**, 1–20 (2021).

66. Martelli, A. M. *et al.* The key roles of PTEN in T-cell acute lymphoblastic leukemia development, progression, and therapeutic response. *Cancers (Basel).* **11**, (2019).

67. Durinck, K. *et al.* Novel biological insights in T-cell acute lymphoblastic leukemia. *Experimental Hematology* vol. 43 625–639 (2015).

68. Follini, E., Marchesini, M. & Roti, G. Strategies to overcome resistance mechanisms in t-cell acute lymphoblastic leukemia. *International Journal of Molecular Sciences* vol. 20 (2019).

69. Zhang, J. *et al.* The genetic basis of early T-cell precursor acute lymphoblastic leukaemia. *Nature* **481**, 157–163 (2012).

70. Ferrando, A. Can one target T-cell ALL? *Best Pract. Res. Clin. Haematol.* **31**, 361–366 (2018).

71. Cauwelier, B. *et al.* Molecular cytogenetic study of 126 unselected T-ALL cases reveals high incidence of TCRβ locus rearrangements and putative new T-cell oncogenes. *Leukemia* **20**, 1238–1244 (2006).

72. Chen, B. *et al.* Identification of fusion genes and characterization of transcriptome features in T-cell acute lymphoblastic leukemia. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 373–378 (2017).

73. Tremblay, M. *et al.* Modeling T-cell acute lymphoblastic leukemia induced by the SCL and LMO1 oncogenes. *Genes Dev.* **24**, 1093 (2010).

74. Noronha, E. P. *et al.* The profile of immunophenotype and genotype aberrations in subsets of pediatric T-cell acute lymphoblastic leukemia. *Front. Oncol.* **9**, 1–10 (2019).

75. Zuurbier, L. *et al.* The significance of PTEN and AKT aberrations in pediatric T-cell acute lymphoblastic leukemia. *Haematologica* **97**, 1405 (2012).

76. Dai, Y. T. *et al.* Transcriptome-wide subtyping of pediatric and adult T cell acute lymphoblastic leukemia in an international study of 707 cases. *Proc. Natl. Acad. Sci. U. S. A.* **119**, (2022).

77. Roberts, K. G. & Mullighan, C. G. Genomics in acute lymphoblastic leukaemia: Insights and treatment implications. *Nat. Rev. Clin. Oncol.* **12**, 344–357 (2015).

78. Iacobucci, I. & Mullighan, C. G. Genetic Basis of Acute Lymphoblastic Leukemia. *J. Clin. Oncol.* **35**, 975 (2017).

79. McMahon, C. M. & Luger, S. M. Relapsed T Cell ALL: Current Approaches and New Directions. *Curr. Hematol. Malig. Rep.* 83–93 (2019).

80. Saadeh, S. S. & Litzow, M. R. Hematopoietic stem cell transplant in adults with acute lymphoblastic leukemia: the present state. *Expert Rev. Hematol.* **11**, 195–207 (2018).

81. Gottesman, M. M., Pastan, I. & Ambudkar, S. V. P-glycoprotein and multidrug resistance. *Curr. Opin. Genet. Dev.* **6**, 610–617 (1996).

82. Ferrando, A. A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).

83. Andrieu, G. P. *et al.* PRC2 loss of function confers a targetable vulnerability to BET proteins in T-ALL. *Blood* **138**, 1855–1869 (2021).

84. Ragon, B. K. *et al.* Buparlisib, a PI3K inhibitor, demonstrates acceptable tolerability and preliminary activity in a phase I trial of patients with advanced leukemias. *Am. J. Hematol.* **92**, 7 (2017).

85. Chiarini, F. *et al.* Activity of the novel dual phosphatidylinositol 3-kinase/mammalian target of rapamycin inhibitor NVP-BEZ235 against T-cell acute lymphoblastic leukemia. *Cancer Res.* **70**, 8097–8107 (2010).

86. Brady, S. W. *et al.* The genomic landscape of pediatric acute lymphoblastic leukemia. *Nat. Genet. 2022* **10**, 1–14 (2022).

87. Le Noir, S. *et al.* Extensive molecular mapping of TCRα/δ- and TCRβ-involved Chromosomal translocations reveals distinct mechanisms of oncogene activation in T-ALL. *Blood* **120**, 3298–3309 (2012).

88. Belver, L. & Ferrando, A. The genetics and mechanisms of T cell acute lymphoblastic leukaemia. *Nat. Rev. Cancer* **16**, 494–507 (2016).

89. Yu, S. *et al.* The TCF-1 and LEF-1 Transcription Factors Have Cooperative and Opposing Roles in T Cell Development and Malignancy. *Immunity* **37**, 813–826 (2012).

90. Mendes, R. D. *et al.* PTEN microdeletions in T-cell acute lymphoblastic leukemia are caused by illegitimate RAG-mediated recombination events. *Blood* **124**, 567–578 (2014).

91. McRae, H. M. *et al. PHF6 regulates hematopoietic stem and progenitor cells and its loss synergizes with expression of TLX3 to cause leukemia*. *Blood* vol. 133 (2019).

92. Wendorff, A. A. *et al.* Phf6 loss enhances HSC self-renewal driving tumor initiation and leukemia stem cell activity in T-All. *Cancer Discov.* **9**, 436–451 (2019).

93. Smith, S. *et al.* LIM Domain Only-2 (LMO2) induces T-cell leukemia by two distinct pathways. *PLoS One* **9**, (2014).

94. De Bie, J. *et al.* Single-cell sequencing reveals the origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia. *Leukemia* **32**, 1358–1369 (2018).

95. Alexander, T. B. *et al.* The genetic basis and cell of origin of mixed phenotype acute leukaemia. *Nature* vol. 562 (Springer US, 2018).

96. Bellavia, D. *et al.* Constitutive activation of NF- k B and T-cell leukemia / lymphoma in Notch3 transgenic mice. **19**, 3337–3348 (2000).

97. Kelliher, M. A., Seldin, D. C. & Leder, P. Tal-1 induces T cell acute lymphoblastic leukemia accelerated by casein kinase IIalpha. *EMBO J.* **15**, 5160 (1996).

98. Vanden Bempt, M. *et al.* Cooperative Enhancer Activation by TLX1 and STAT5 Drives Development of NUP214-ABL1/TLX1-Positive T Cell Acute Lymphoblastic Leukemia. *Cancer Cell* **34**, 271-285.e7 (2018).

99. Booth, C. A. G. *et al.* Ezh2 and Runx1 Mutations Collaborate to Initiate Lympho-Myeloid Leukemia in Early Thymic Progenitors. *Cancer Cell* **33**, 274-291.e8 (2018).

100. Gao, J. *et al.* Therapeutic targeting of Notch signaling and immune checkpoint blockade in a spontaneous, genetically heterogeneous mouse model of T-cell acute lymphoblastic leukemia. *DMM Dis. Model. Mech.* **12**, 1–10 (2019).

101. Maude, S. L. *et al.* Efficacy of JAK/STAT pathway inhibition in murine xenograft models of early T-cell precursor (ETP) acute lymphoblastic leukemia. *Blood* **125**, 1759–1767 (2015).

102. Govaerts, I. *et al.* PSEN1-selective gamma-secretase inhibition in combination with kinase or XPO-1 inhibitors effectively targets T cell acute lymphoblastic leukemia. *J. Hematol. Oncol.* **14**, 1–14 (2021).

103. Andreas Reinisch1, Daniel Thomas1, M. Ryan Corces1, Xiaohua Zhang1, Dita Gratzinger2, Wan-Jen Hong1, Katharina Schallmoser3,5, Dirk Strunk4,5, and R. M. A Humanized Ossicle-niche Xenotransplantation Model Enables Improved Human Leukemic Engraftment. *Nat. Med.* **22**, 812–821 (2016).

104. Rhein, M. *et al.* Leukemias induced by altered TRK-signaling are sensitive to mTOR inhibitors in preclinical models. *Ann. Hematol.* **90**, 283–292 (2011).

105. Habets, R. A. *et al.* Safe targeting of T cell acute lymphoblastic leukemia by pathology-specific NOTCH inhibition. *Sci. Transl. Med.* **11**, (2019).

106. Yokoyama, K. *et al.* Lymphoid neoplasia: In vivo leukemogenic potential of an interleukin 7 receptor a chain mutant in hematopoietic stem and progenitor cells. *Blood* **122**, 4259–4263 (2013).

107. Schwarzer, A. *et al.* Targeting Rac-Gtpases Reverses Stroma-Induced Resistance to Notch and mTOR-Inhibition in Acute T-Cell Leukemia. *Blood* **126**, 1326–1326 (2015).

108. Heckl, D. *et al.* Generation of mouse models of myeloid malignancy with combinatorial genetic lesions using CRISPR-Cas9 genome editing. *Nat. Biotechnol.* **32**, 941–946 (2014).

109. Maetzig, T., Galla, M., Baum, C. & Schambach, A. Gammaretroviral vectors: Biology, technology and application. *Viruses* **3**, 677–713 (2011).

110. Yamashita, M. & Emerman, M. Retroviral infection of non-dividing cells : Old and new perspectives. **344**, 88–93 (2006).

111. Naldini, L. *et al.* In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. *Science (80-.).* **272**, 263–267 (1996).

112. Mitchell, R. S. *et al.* Retroviral DNA integration: ASLV, HIV, and MLV show distinct target site preferences. *PLoS Biol.* **2**, (2004).

113. Lewinski, M. K. *et al.* Retroviral DNA integration: Viral and cellular determinants of target-site selection. *PLoS Pathog.* **2**, 0611–0622 (2006).

114. Schröder, A. R. W. *et al.* HIV-1 Integration in the Human Genome Favors Active Genes and Local Hotspots. *Cell* **110**, 521–529 (2002).

115. Finkelshtein, D., Werman, A., Novick, D., Barak, S. & Rubinstein, M. LDL receptor and its family members serve as the cellular receptors for vesicular stomatitis virus. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 7306–7311 (2013).

116. Bokhoven, M. *et al.* Insertional Gene Activation by Lentiviral and Gammaretroviral Vectors. *J. Virol.* **83**, 283–294 (2009).

117. Hacein-Bey-Abina, S. *et al.* Insertional oncogenesis in 4 patients after retrovirus-mediated gene therapy of SCID-X1. *J. Clin. Invest.* **118**, 3132 (2008).

118. Zufferey, R. *et al.* Self-Inactivating Lentivirus Vector for Safe and Efficient In Vivo Gene Delivery. *J. Virol.* **72**, 9873–9880 (1998).

119. Modlich, U. *et al.* Insertional transformation of hematopoietic cells by self-inactivating lentiviral and gammaretroviral vectors. *Mol. Ther.* **17**, 1919–1928 (2009).

120. Zheng, C. & Baum, B. J. Evaluation of Promoters for Use in Tissue-Specific Gene Delivery. in *Gene Therapy Protocols* 205–219 (Humana Press, 2008).

121. Verhoeyen, E. & Cosset, F.-L. Surface-engineering of lentiviral vectors. *J. Gene Med.* **6**, S83–S94 (2004).

122. Kim, H., Kim, M., Im, S.-K. & Fang, S. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab. Anim. Res.* **34**, 147 (2018).

123. Nagy, A. Cre recombinase: The universal reagent for genome tailoring. *Genesis* **26**, 99–109 (2000).

124. Sternberg, N. & Hamilton, D. Bacteriophage P1 site-specific recombination. I. Recombination between loxP sites. *J. Mol. Biol.* **150**, 467–486 (1981).

125. Hoess, R. H., Ziese, M. & Sternberg, N. P1 site-specific recombination: nucleotide sequence of the recombining sites. *Proc. Natl. Acad. Sci. U. S. A.* **79**, 3398–402 (1982).

126. Dragatsis, I. & Zeitlin, S. A method for the generation of conditional gene repair mutations in mice. *Nucleic Acids Res.* **29**, 1–9 (2001).

127. Albert, H., Dale, E. C., Lee, E. & Ow, D. W. Site-specific integration of DNA into wild-type and mutant lox sites placed in the plant genome. *Plant J.* **7**, 649–659 (1995).

128. Zhang, Z. & Lutz, B. Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. *Nucleic Acids Res.* **30**, 5–9 (2002).

129. Garneau, J. E. *et al.* The CRISPR/cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature* **468**, 67–71 (2010).

130. Miao, J. *et al.* Targeted mutagenesis in rice using CRISPR-Cas system. *Cell Res.* **23**, 1233–1236 (2013).

131. Shan, Q., Wang, Y., Li, J. & Gao, C. Genome editing in rice and wheat using the CRISPR/Cas system. *Nat. Protoc.* **9**, 2395–2410 (2014).

132. Min, Y. L. *et al.* CRISPR-Cas9 corrects Duchenne muscular dystrophy exon 44 deletion mutations in mice and human cells. *Sci. Adv.* **5**, 1–12 (2019).

133. Bjursell, M. *et al.* Therapeutic Genome Editing With CRISPR/Cas9 in a Humanized Mouse Model Ameliorates α1-antitrypsin Deficiency Phenotype. *EBioMedicine* **29**, 104–111 (2018).

134. Hou, P. *et al.* Genome editing of CXCR4 by CRISPR/cas9 confers cells resistant to HIV-1 infection. *Sci. Rep.* **5**, 1–12 (2015).

135. Takeda, H. *et al.* CRISPR-Cas9-mediated gene knockout in intestinal tumor organoids provides functional validation for colorectal cancer driver genes. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 15635–15644 (2019).

136. Chen, C. H. *et al.* Dual inhibition of PIK3C3 and FGFR as a new therapeutic approach to treat bladder cancer. *Clin. Cancer Res.* **24**, 1176–1189 (2018).

137. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nat. Protoc.* **8**, 2281–2308 (2013).

138. Chylinski, K., Makarova, K. S., Charpentier, E. & Koonin, E. V. Classification and evolution of type II CRISPR-Cas systems. *Nucleic Acids Res.* **42**, 6091–6105 (2014).

139. Jinek, M. *et al.* A Programmable Dual-RNA – Guided. **337**, 816–822 (2012).

140. Brandsma, I. & Gent, D. C. Pathway choice in DNA double strand break repair: Observations of a balancing act. *Genome Integr.* **3**, 1 (2012).

141. Miyaoka, Y. *et al.* Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci. Rep.* **6**, 1–12 (2016).

142. Michlits, G. *et al.* Multilayered VBC score predicts sgRNAs that efficiently generate loss-of-function alleles. *Nature Methods* vol. 17 (Springer US, 2020).

143. Schwank, G. *et al.* Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13**, 653–658 (2013).

144. Frangoul, H. *et al.* CRISPR-Cas9 Gene Editing for Sickle Cell Disease and β-Thalassemia. *N. Engl. J. Med.* **384**, 252–260 (2021).

145. Xu, H. *et al.* Sequence determinants of improved CRISPR sgRNA design. *Genome Res.* **25**, 1147–1157 (2015).

146. Stemmer, M., Thumberger, T., Del Sol Keyer, M., Wittbrodt, J. & Mateo, J. L. CCTop: An intuitive, flexible and reliable CRISPR/Cas9 target prediction tool. *PLoS One* **10**, 1–11 (2015).

147. Moreno-Mateos, M. A. *et al.* CRISPRscan: designing highly efficient sgRNAs for CRISPR-Cas9 targeting in vivo. *Nat. Methods* **12**, 982–988 (2015).

148. Allen, F. *et al.* Predicting the mutations generated by repair of Cas9-induced double-strand breaks. *Nat. Biotechnol.* **37**, 64–82 (2019).

149. Labuhn, M. *et al.* Refined sgRNA efficacy prediction improves largeand small-scale CRISPR-Cas9 applications. *Nucleic Acids Res.* **46**, 1375–1385 (2018).

150. Adams, F. F. *et al.* An optimized lentiviral vector system for conditional RNAi and efficient cloning of microRNA embedded short hairpin RNA libraries. *Biomaterials* **139**, 102–115 (2017).

151. Schambach, A. *et al.* Woodchuck hepatitis virus post-transcriptional regulatory element deleted from X protein and promoter sequences enhances retroviral vector titer and expression. *Gene Ther.* **13**, 641–645 (2006).

152. Ng, P. K. S. *et al.* Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell* **33**, 450-462.e10 (2018).

153. Schambach, A. *et al.* Overcoming promoter competition in packaging cells improves production of self-inactivating retroviral vectors. *Gene Ther.* **13**, 1524–1533 (2006).

154. Yang, Y. *et al.* Inducible, High-Level Production of Infectious Murine Leukemia Retroviral Vector Particles Pseudotyped with Vesicular Stomatitis Virus G Envelope Protein. *Hum. Gene Ther.* **6**, 1203–1213 (1995).

155. Baron, Y. *et al.* Improved alpharetrovirus-based Gag.MS2 particles for efficient and transient delivery of CRISPR-Cas9 into target cells. *Mol. Ther. - Nucleic Acids* **27**, 810–823 (2022).

156. Hennet, T., Hagen, F. K., Tabak, L. A. & Marth, J. D. Genetics T-cell-specific deletion of a polypeptide N-acetylgalactosaminyl-transferase gene by site-directed recombination. *Proc. Natl. Acad. Sci. U. S. A.* **92**, 12070–12074 (1995).

157. Brinkman, E. K., Chen, T., Amendola, M. & Van Steensel, B. Easy quantitative assessment of genome editing by sequence trace decomposition. *Nucleic Acids Res.* **42**, (2014).

158. Gordon, A. No Title. FASTQ/A short-reads pre-processing tools http://hannonlab.cshl.edu/fastx_toolkit/ (2010).

159. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).

160. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. (2012)

161. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2;

iso-3. *Fly (Austin).* **6**, 80–92 (2012).

162. Smyth, G. K. *et al.* RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. *F1000Research* **5**, 1–30 (2018).

163. Wang, J., Vasaikar, S., Shi, Z., Greer, M. & Zhang, B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **45**, W130–W137 (2017).

164. Carow, B., Gao, Y., Coquet, J., Reilly, M. & Rottenberg, M. E. lck -Driven Cre Expression Alters T Cell Development in the Thymus and the Frequencies and Functions of Peripheral T Cell Subsets . *J. Immunol.* **197**, 2261–2268 (2016).

165. Fernald, K. & Kurokawa, M. Evading apoptosis in cancer. *Trends Cell Biol.* **23**, 620–633 (2013).

166. Milella, M. *et al.* PTEN: Multiple Functions in Human Malignant Tumors. *Front. Oncol.* **5**, (2015).

167. Di Cristofano, A., Pesce, B., Cordon-Cardo, C. & Pandolfi, P. P. Pten is essential for embryonic development and tumour suppression. *Nat. Genet. 1998 194* **19**, 348–355 (1998).

168. Guo, W. *et al.* Multi-genetic events collaboratively contribute to Pten-null leukaemia stem-cell formation. *Nature* **453**, 529 (2008).

169. Lewis, S. M., Williams, A. & Eisenbarth, S. C. Structure-function of the immune system in the spleen. *Sci. Immunol.* **4**, (2019).

170. Doench, J. G. *et al.* Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. *Nat. Biotechnol.* **34**, 184 (2016).

171. Herbst, F. *et al.* Extensive methylation of promoter sequences silences lentiviral transgene expression during stem cell differentiation in vivo. *Mol. Ther.* **20**, 1014–1021 (2012).

172. Winiarska, M. *et al.* Selection of an optimal promoter for gene transfer in normal B cells. *Mol. Med. Rep.* **16**, 3041 (2017).

173. Kustikova, O. S. *et al.* Dose finding with retroviral vectors: Correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. *Blood* **102**, 3934–3937 (2003).

174. Wang, J. *et al.* NrasG12D/+ promotes leukemogenesis by aberrantly regulating hematopoietic stem cell functions. *Blood* **121**, 5203–5207 (2013).

175. Zenatti, P. P. *et al.* Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nat. Genet.* **43**, 932–941 (2011).

176. Shum, T. *et al.* Constitutive Signaling from an Engineered IL7 Receptor Promotes Durable Tumor Elimination by Tumor-Redirected T Cells. *Cancer Discov.* **7**, 1238–1247 (2017).

177. Thouenon, R., Moreno-Corona, N., Poggi, L., Durandy, A. & Kracker, S. Activated PI3Kinase Delta Syndrome—A Multifaceted Disease. *Front. Pediatr.* **9**, 597 (2021).

178. Angulo, I. *et al.* Phosphoinositide 3-kinase δ gene mutation predisposes to respiratory infection and airway damage. *Science* **342**, 866 (2013).

179. Tan-Pertel, H. T. *et al.* Notch Signaling Enhances Survival and Alters Differentiation of 32D Myeloblasts. *J. Immunol.* **165**, 4428–4436 (2000).

180. Warmuth, M., Kim, S., Gu, X. J., Xia, G. & Adrián, F. Ba/F3 cells and their use in kinase drug discovery. *Curr. Opin. Oncol.* **19**, 55–60 (2007).

181. Degryse, S. *et al.* Mutant JAK3 signaling is increased by loss of wild-type JAK3 or by acquisition of secondary JAK3 mutations in T-ALL. *Blood* **131**, 421–425 (2018).

182. Marks, D. I. *et al.* T-cell acute lymphoblastic leukemia in adults: Clinical features, immunophenotype, cytogenetics, and outcome from the large randomized prospective trial (UKALL XII/ECOG 2993). *Blood* **114**, 5136–5145 (2009).

183. Bornschein, S. *et al.* Defining the molecular basis of oncogenic cooperation between TAL1 expression and Pten deletion in T-ALL using a novel pro-T-cell model system. *Leukemia* **32**, 941–951 (2018).

184. Fasseu, M. *et al.* p16INK4A tumor suppressor gene expression and CD3ϵ deficiency but not pre-TCR deficiency inhibit TAL1-linked T-lineage leukemogenesis. *Blood* **110**, 2610 (2007).

185. Wang, J. *et al.* Endogenous oncogenic Nras mutation initiates hematopoietic malignancies in a dose- and cell type-dependent manner. *Blood* **118**, 368 (2011).

186. Durinck, K. *et al.* Novel biological insights in T-cell acute lymphoblastic leukemia. *Experimental Hematology* (2015)

187. Ferrando, A. A. *et al.* Gene expression signatures define novel oncogenic pathways in T cell acute lymphoblastic leukemia. *Cancer Cell* **1**, 75–87 (2002).

188. Shaw, T. I. *et al.* Integrative network analysis reveals USP7 haploinsufficiency inhibits E-protein activity in pediatric T-lineage acute lymphoblastic leukemia (T-ALL). *Sci. Reports 2021 111* **11**, 1–12 (2021).

189. Hurlstone, A. & Clevers, H. T-cell factors: Turn-ons and turn-offs. *EMBO J.* **21**, 2303–2311 (2002).

190. De Braekeleer, E. *et al.* ETV6 fusion genes in hematological malignancies: A review. *Leuk. Res.* **36**, 945–961 (2012).

191. Liu, Z. *et al.* Structural and functional insights into the human börjeson-forssman- lehmann syndrome-associated protein PHF6. *J. Biol. Chem.* **289**, 10069–10083 (2014).

192. Silva, A. *et al.* Overexpression of wild-type IL-7R a promotes T-cell acute lymphoblastic leukemia / lymphoma. *Blood* **138**, 1040–1052 (2021).

193. Lapidot, T. *et al.* A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nat. 1994 3676464* **367**, 645–648 (1994).

194. Demarest, R. M., Ratti, F. & Capobianco, A. J. It's T-ALL about Notch. *Oncogene 2008 2738* **27**, 5082–5091 (2008).

195. Winter, S. S. *et al.* Bone marrow stroma-supported culture of T-lineage acute lymphoblastic leukemic cells predicts treatment outcome in children: a Pediatric Oncology Group study. *Leuk. 2002 166* **16**, 1121–1126 (2002).

196. Zweier-Renn, L. A., Riz, I., Hawley, T. S. & Hawley, R. G. The DN2 Myeloid-T (DN2mt) Progenitor is a Target Cell for Leukemic Transformation by the TLX1 Oncogene. *J. bone marrow Res.* **1**, 1–19 (2013).

197. Pal, D. *et al.* Long-term in vitro maintenance of clonal abundance and leukaemia-initiating potential in acute lymphoblastic leukaemia. *Leukemia* **30**, 1691–1700 (2016).

198. Anandakrishnani, R., Varghese, R. T., Kinneyi, N. A. & Garner, H. R. Estimating the number of genetic mutations (HITS) required for carcinogenesis based on the distribution of somatic mutations. *PLoS Comput. Biol.* **15**, 1–19 (2019).

199. Marks, D. I. & Rowntree, C. Management of adults with T-cell lymphoblastic leukemia. *Blood* **129**, 1134–1142 (2017).

200. Deangelo, D. J. *et al.* A phase I clinical trial of the notch inhibitor MK-0752 in patients with T-cell acute lymphoblastic leukemia/lymphoma (T-ALL) and other leukemias. *J. Clin. Oncol.* **24**, 6585 (2006).

201. Papayannidis, C. *et al.* A Phase 1 study of the novel gamma-secretase inhibitor PF-03084014 in patients with T-cell acute lymphoblastic leukemia and T-cell lymphoblastic lymphoma. *Blood Cancer J.* **5**, e350-3 (2015).

202. Takebe, Naoko; Nguyen, Dat; Yang, S. X. Targeting Notch signaling pathway in cancer: Clinical development advances and challenges. *Pharmacol. Ther.* **23**, 1–7 (2014).

203. Frismantas, V. *et al.* Ex vivo drug response profiling detects recurrent sensitivity patterns in drug-resistant acute lymphoblastic leukemia. *Blood* **129**, e26–e37 (2017).

204. Cappelli, L. V. *et al.* Endothelial-leukemia interactions remodel drug responses uncovering T-ALL vulnerabilities. *Blood* (2022)

205. Thielemans, N. *et al.* TAL1 cooperates with PI3K/AKT pathway activation in T-cell acute lymphoblastic leukemia. *Haematologica* **107**, 2304–2317 (2022).

206. de Boer, J. *et al.* Transgenic mice with hematopoietic and lymphoid specific expression of Cre. *Eur. J. Immunol.* **33**, 314–325 (2003).

207. Messina, M. *et al.* Prognostic and therapeutic role of targetable lesions in B-lineage acute lymphoblastic leukemia without recurrent fusion genes. *Oncotarget* **7**, 13886 (2016).

208. Bagger, F. O. *et al.* BloodSpot: a database of gene expression profiles and transcriptional programs for healthy and malignant haematopoiesis. *Nucleic Acids Res.* **44**, D917–D924 (2016).

209. Janbandhu, V. C., Moik, D. & Fässler, R. Cre recombinase induces DNA damage and tetraploidy in the absence of LoxP sites. *Cell Cycle* **13**, 462–470 (2014).

210. Haapaniemi, E., Botla, S., Persson, J., Schmierer, B. & Taipale, J. CRISPR–Cas9 genome editing induces a p53-mediated DNA damage response. *Nat. Med. 2018 247* **24**, 927–930 (2018).

211. Shi, J. & Petrie, H. T. Activation Kinetics and Off-Target Effects of Thymus-Initiated Cre Transgenes. *PLoS One* **7**, 46590 (2012).

212. Luche, H. *et al.* In vivo fate mapping identifies pre-TCRα expression as an intra- and extrathymic, but not prethymic, marker of T lymphopoiesis. *J. Exp. Med.* **210**, 699 (2013).

213. Platt, R. J. *et al.* CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell* **159**, 440–455 (2014).

214. Nonami, A. *et al.* Identification of novel therapeutic targets in acute leukemias with NRAS mutations using a pharmacologic approach. *Blood* **125**, 3133 (2015).

215. Kleppe, M., Mentens, N., Tousseyn, T., Wlodarska, I. & Cools, J. MOHITO, a novel mouse cytokine-dependent T-cell line, enables studies of oncogenic signaling in the T-cell context. *Haematologica* **96**, 779 (2011).

216. ten Hacken, E. *et al.* In vivo modeling of CLL transformation to Richter's syndrome reveals convergent evolutionary paths and therapeutic vulnerabilities. *Blood Cancer Discov.* (2022)

217. Della Gatta, G. *et al.* Reverse engineering of TLX oncogenic transcriptional networks identifies RUNX1 as tumor suppressor in T-ALL. *Nat. Med.* **18**, 436 (2012).

218. Ferrando, A. A. *et al.* Prognostic importance of TLX1 (HOX11) oncogene expression in adults with T-cell acute lymphoblastic leukaemia. *Lancet* **363**, 535–536 (2004).

219. Dear, T. N. *et al.* The Hox11 gene is essential for cell survival during spleen development. *Development* **121**, 2909–2915 (1995).

220. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and molecular biology of T-ALL. *Blood* **129**, 1113–1123 (2017).

221. De Keersmaecker, K. *et al.* The TLX1 oncogene drives aneuploidy in T cell transformation. *Nat. Med.* **16**, 1321–1327 (2010).

222. Durinck, K. *et al.* Characterization of the genome-wide TLX1 binding profile in T-cell acute lymphoblastic leukemia. *Leukemia* **29**, 2317–2327 (2015).

223. De Keersmaecker, K., Versele, M., Cools, J., Superti-Furga, G. & Hantschel, O. Intrinsic differences between the catalytic properties of the oncogenic NUP214-ABL1 and BCR-ABL1 fusion protein kinases. *Leukemia* **22**, 2208–2216 (2008).

224. Iwasa, H., Han, J. & Ishikawa, F. Mitogen-activated protein kinase p38 defines the common senescence-signalling pathway. *Genes to Cells* **8**, 131–144 (2003).

225. Xu, Y., Li, N., Xiang, R. & Sun, P. Emerging roles of the p38 MAPK and PI3K/AKT/mTOR pathways in oncogene-induced senescence. *Trends Biochem. Sci.* **39**, 268 (2014).

226. Miyauchi, H. *et al.* Akt negatively regulates the in vitro lifespan of human endothelial cells via a p53/p21-dependent pathway. *EMBO J.* **23**, 212 (2004).

227. Mallette, F. A., Calabrese, V., Ilangumaran, S. & Ferbeyre, G. SOCS1, a novel interaction partner of p53 controlling oncogene-induced senescence. *Aging (Albany. NY).* **2**, 445–52 (2010).

228. Owens, B. M., Hawley, T. S., Spain, L. M., Kerkel, K. A. & Hawley, R. G. TLX1/HOX11-

mediated disruption of primary thymocyte differentiation prior to the CD4+CD8+ double-positive stage. *Br. J. Haematol.* **132**, 216–229 (2006).

229. Krutikov, K., Zheng, Y., Chesney, A., Huang, X. & Vaags, A. K. Ectopic TLX1 Expression Accelerates Malignancies in Mice Deficient in DNA-PK. *PLoS One* **9**, 89649 (2014).

230. Dadi, S. *et al.* TLX Homeodomain Oncogenes Mediate T Cell Maturation Arrest in T-ALL via Interaction with ETS1 and Suppression of TCRα Gene Expression. *Cancer Cell* **21**, 563–576 (2012).

231. Guidos, C. J., Williams, C. J., Wu, G. E., Paige, C. J. & Danska, J. S. Development of CD4+CD8+ thymocytes in RAG-deficient mice through a T cell receptor beta chain-independent pathway. *J. Exp. Med.* **181**, 1187–95 (1995).

232. Cramer, S. D. *et al.* Mutant IL-7Rα and mutant NRas are sufficient to induce murine T cell acute lymphoblastic leukemia. *Leuk. 2018 328* **32**, 1795–1882 (2018).

233. Van Thillo, Q. *et al.* Oncogenic cooperation between TCF7-SPI1 and NRAS(G12D) requires β-catenin activity to drive T-cell acute lymphoblastic leukemia. *Nat. Commun.* **12**, 1–15 (2021).

234. Kamimura, K. *et al.* Haploinsufficiency of Bcl11b for suppression of lymphomagenesis and thymocyte development. *Biochem. Biophys. Res. Commun.* **355**, 538–542 (2007).

235. Lykke-Andersen, S. & Jensen, T. H. Nonsense-mediated mRNA decay: An intricate machinery that shapes transcriptomes. *Nat. Rev. Mol. Cell Biol.* **16**, 665–677 (2015).

236. Klatt, D. *et al.* Competitive sgRNA Screen Identifies p38 MAPK as a Druggable Target to Improve HSPC Engraftment. *Cells* **9**, 1–19 (2020).

237. Patterson, L. J. *et al.* The transcription factors Scl and Lmo2 act together during development of the hemangioblast in zebrafish. *Blood* **109**, 2389–2398 (2007).

238. Vagapova, E. R., Spirin, P. V, Lebedev, T. D. & Prassolov, V. S. The Role of TAL1 in Hematopoiesis and Leukemogenesis. *Acta Naturae* **10**, 15–23 (2018).

239. Condorelli, G. L. *et al.* T-cell-directed TAL-1 expression induces T-cell malignancies in transgenic mice. *Cancer Res.* **56**, 5113–5119 (1996).

240. Larson, R. C. *et al.* Protein dimerization between Lmo2 (Rbtn2) and Tal1 alters thymocyte development and potentiates T cell tumorigenesis in transgenic mice. *EMBO J.* **15**, 1021 (1996).

241. Newrzela, S. *et al.* Resistance of mature T cells to oncogene transformation. *Blood* **112**, 2278–2286 (2008).

242. Hsu, H. L., Cheng, J. T., Chen, Q. & Baer, R. Enhancer-binding activity of the tal-1 oncoprotein in association with the E47/E12 helix-loop-helix proteins. *Mol. Cell. Biol.* **11**, 3037 (1991).

243. Wadman, I. A. *et al.* The LIM-only protein Lmo2 is a bridging molecule assembling an erythroid, DNA-binding complex which includes the TAL1, E47, GATA-1 and Ldb1/NLI proteins. *EMBO J.* **16**, 3145–3157 (1997).

244. Kassouf, M. T., Chagraoui, H., Vyas, P. & Porcher, C. Differential use of SCL/TAL-1 DNA-binding domain in developmental hematopoiesis. *Blood* **112**, 1056–1067 (2008).

245. Bernard, M. *et al.* Helix-loop-helix (E2-5, HEB, TAL1 and Id1) protein interaction with the TCRαδ enhancers. *Int. Immunol.* **10**, 1539–1549 (1998).

246. Bain, G. *et al.* E2A deficiency leads to abnormalities in alphabeta T-cell development and to rapid development of T-cell lymphomas. *Mol. Cell. Biol.* **17**, 4782 (1997).

247. ElOmari, K. *et al.* Structural Basis for LMO2-Driven Recruitment of the SCL: E47bHLH Heterodimer to Hematopoietic-Specific Transcriptional Targets. *Cell Rep.* **4**, 135–147 (2013).

248. Draheim, K. M. *et al.* A DNA-binding mutant of TAL1 cooperates with LMO2 to cause T cell leukemia in mice. *Oncogene* **30**, 1252 (2011).

249. Aplan, P. D. *et al.* An scl gene product lacking the transactivation domain induces bony abnormalities and cooperates with LMO1 to generate T-cell malignancies in transgenic mice. *EMBO J.* **16**, 2408–2419 (1997).

250. McCormack, M. P. *et al.* Requirement for Lyl1 in a model of Lmo2-driven early T-cell precursor

ALL. *Blood* **122**, 2093–2103 (2013).

251. McCormack, M. P. *et al.* The Lmo2 oncogene initiates leukemia in mice by inducing thymocyte self-renewal. *Science (80-.).* **327**, 879–883 (2010).

252. Chervinsky, D. S. *et al.* Disordered T-Cell Development and T-Cell Malignancies in SCL LMO1 Double-Transgenic Mice: Parallels with E2A-Deficient Mice. *Mol. Cell. Biol.* **19**, 5025–5035 (1999).

253. Adler, S. H. *et al.* Notch Signaling Augments T Cell Responsiveness by Enhancing CD25 Expression. *J. Immunol.* **171**, 2896–2903 (2003).

254. Dutton, R. W., Bradley, L. M. & Swain, S. L. T cell memory. *Annu. Rev. Immunol.* **16**, 201–23 (1998).

255. Sckisel, G. D. *et al.* Differential phenotypes of memory CD4 and CD8 T cells in the spleen and peripheral tissues following immunostimulatory therapy. *J. Immunother. Cancer* **5**, 1–11 (2017).

256. Zhang, J. A., Mortazavi, A., Williams, B. A., Wold, B. J. & Rothenberg, E. V. Dynamic transformations of genome-wide epigenetic marking and transcriptional control establish T cell identity. *Cell* **149**, 467 (2012).

257. Karlsson, M. *et al.* A single–cell type transcriptomics map of human tissues. *Sci. Adv.* **7**, (2021).

258. Lestringant, V. *et al.* Optical genome mapping, a promising alternative to gold standard cytogenetic approaches in a series of acute lymphoblastic leukemias. *Genes, Chromosom. Cancer* **60**, 657–667 (2021).

259. Ihry, R. J. *et al.* P53 inhibits CRISPR-Cas9 engineering in human pluripotent stem cells. *Nat. Med.* **24**, 939–946 (2018).

260. Wang, H. *et al.* One-Step Generation of Mice Carrying Mutations in Multiple Genes by CRISPR/Cas-Mediated Genome Engineering. *Cell* **153**, 910 (2013).

261. Sharma, V. M. *et al.* Notch1 Contributes to Mouse T-Cell Leukemia by Directly Inducing the Expression of c -myc . *Mol. Cell. Biol.* **26**, 8022–8031 (2006).

262. Sherman, E. *et al.* INSPIIRED: A Pipeline for Quantitative Analysis of Sites of New DNA Integration in Cellular Genomes. *Mol. Ther. Methods Clin. Dev.* **4**, 39 (2017).

263. Biasco, L., Rothe, M., Büning, H. & Schambach, A. Analyzing the Genotoxicity of Retroviral Vectors in Hematopoietic Cell Gene Therapy. *Mol. Ther. - Methods Clin. Dev.* **8**, 21–30 (2018).

264. Patkar, N. *et al.* NARASIMHA: Novel Assay based on Targeted RNA Sequencing to Identify ChiMeric Gene Fusions in Hematological Malignancies. *Blood Cancer J.* **10**, 4–7 (2020).

265. Freeman, D. *et al.* Genetic Background Controls Tumor Development in Pten-Deficient Mice. *Cancer Res.* **66**, 6492–6496 (2006).

266. Hagenbeek, T. J. & Spits, H. T-cell lymphomas in T-cell-specific Pten-deficient mice originate in the thymus. *Leuk. 2008 223* **22**, 608–619 (2007).

267. Zhu, H. *et al.* T-ALL leukemia stem cell 'stemness' is epigenetically controlled by the master regulator SPI1. *Elife* **7**, (2018).

268. Guo, W. *et al.* Suppression of leukemia development caused by PTEN loss. *Proc. Natl. Acad. Sci. U. S. A.* **108**, 1409–1414 (2011).

269. Lieske, A. *et al.* A pro B cell population forms the apex of the leukemic hierarchy in Hoxa9/Meis1-dependent AML. *Leukemia* **37**, 79 (2023).

270. Maetzig, T. *et al.* A Lentiviral Fluorescent Genetic Barcoding System for Flow Cytometry-Based Multiplex Tracking. *Mol. Ther.* **25**, 606–620 (2017).

271. Indraccolo, S. *et al.* Cross-talk between tumor and endothelial cells involving the Notch3-DII4 interaction marks escape from tumor dormancy. *Cancer Res.* **69**, 1314–1323 (2009).

272. Teicher, B. A. & Fricker, S. P. CXCL12 (SDF-1)/CXCR4 pathway in cancer. *Clin. Cancer Res.* **16**, 2927–2931 (2010).

273. Duan, C. W. *et al.* Leukemia Propagating Cells Rebuild an Evolving Niche in Response to

Therapy. *Cancer Cell* **25**, 778–793 (2014).

274. Poglio, S. *et al.* Rapid childhood T-ALL growth in xenograft models correlates with mature phenotype and NF-κB pathway activation but not with poor prognosis. *Leukemia* **29**, 977–980 (2015).

275. Armstrong, F. *et al.* NOTCH is a key regulator of human T-cell acute leukemia initiating cell activity. *Blood* **113**, 1730–1740 (2009).

276. González-García, S. *et al.* IL-7R is essential for leukemia-initiating cell activity of T-cell acute lymphoblastic leukemia. *Blood* **134**, 2171–2182 (2019).

277. Liang, Z. *et al.* Silencing of CXCR4 Blocks Breast Cancer Metastasis. *Cancer Res.* **65**, 967 (2005).

278. Liang, Z. *et al.* CXCR4/CXCL12 axis promotes VEGF-mediated tumor angiogenesis through Akt signaling pathway. *Biochem. Biophys. Res. Commun.* **359**, 716–722 (2007).

279. van den Berk, L. C. J. *et al.* Disturbed CXCR4/CXCL12 axis in paediatric precursor B-cell acute lymphoblastic leukaemia. *Br. J. Haematol.* **166**, 240–249 (2014).

280. Juarez, J. *et al.* CXCR4 antagonists mobilize childhood acute lymphoblastic leukemia cells into the peripheral blood and inhibit engraftment. *Leuk. 2007 216* **21**, 1249–1257 (2007).

281. Pitt, L. A. *et al.* CXCL12-producing vascular endothelial niches control acute T cell leukemia maintenance. *Cancer Cell* **27**, 755 (2015).

282. López-Nieva, P. *et al.* More Insights on the Use of γ-Secretase Inhibitors in Cancer Treatment. *Oncologist* **26**, e298–e305 (2021).

7.4 List of abbreviations

4EBP1	4E binding protein-1
ADAM-10	a disintegrin and metalloproteinase domaincontaining protein 10
AKT	serine/threonine kinase
AML	acute myeloid leukemia
AP-1	activator protein-1
BCL11B	B cell lymphoma/leukemia 11B
B-CLL	B cell chronic lymphocyte leukemia
bHLH	basic helix-loop-helix
BM	bone marrow
bp	base pair
BRD4	bromodomain protein 4
BSA	bovine serum albumin
Cas9	CRISPR associated protein 9
CD	cluster of differentiation
CDKN2A	cyclin dependent kinase inhibitor 2A
CLP	common lymphoid progenitors
CMP	common myeloid progenitors
Cre	Cre recombinase
crRNA	CRISPR RNA
CXCL12	C-X-C motif chemokine ligand 12
CXCR4	C-X-C motif chemokine receptor 4
DAPI	4′,6-Diamidin-2-phenylindol
DE	differentially expressed
DL1	delta-like 1
DN	double-negative
DNA	deoxyribonucleic acid
DSB	double strand break
EFS	EF-1 alpha short
env	envelope
ERK1/2	extracellular-signal regulated kinases 1 and
ETV6	ETS variant transcription factor 6
EZH2	enhancer of zeste homolog 2
FACS	fluorescence-activated cell sorting
FBXW7	F-Box and WD repeat domain containing 7
FDR	false-discovery rate
floxed	flanked by loxP sites
FoxO	forkhead box O
gag	group associated antigen
GATA3	GATA binding protein 3
GMP	granulocyte-monocyte progenitors
GOI	gene-of-interest
GSEA	gene set enrichment analysis
GTP	guanosine-5'-triphosphate
HDR	homology-derived repair
HLA-DR	human leukocyte antigen – DR isotype
HOXA	Homeobox A
HSC	hematopoietic stem cell
HSCT	human stem cell transplantation
HSPC	hematopoietic stem and progenitor cell
ICN1	intracellular NOTCH1
IL2Ra	interleukin 2 receptor alpha chain, CD25
IL2Rg	interleukin 2 receptor subunit gamma

IL2	interleukin 2
IL3	interleukin 3
IL7	interleukin 7
IL7R	interleukin 7 receptor
indels	insertions and deletions
IQR	interquartile range
JAK	Janus kinase
KEGG	Kyoto Encyclopedia of Genes and Genomes
Kitl	C-Kit ligand
Lck	lymphocyte-specific protein tyrosine kinase
lcpm	logarithmized counts per million reads mapped
LEF1	Lymphoid enhancer binding factor 1
LB	Lysogeny Broth
lfc	log2-fold change
LIC	leukemia-initiating cells
LMO1/2	Lim domain only 1/2
loxP	locus of x-over, P1
LSC	leukemic stem cell
LSL	loxP-stop-loxP cassette
LT-HSC	long-term hematopoietic stem cells
LTR	long terminal repeat
LV	lentivirus
LYL	LYL1 basic helix-loop-helix (bHLH) family member
MAPK	mitogen activated protein kinase
MDM2	murine double minute 2
MEP	megakaryocyte-erythrocyte progenitors
MHC	major histocompatibility complex
MPP	multipotent progenitors
MSC	mesenchymal stroma cells
mTORC1	mammalian target of rapamycin complex 1
mTORC2	mammalian target of rapamycin complex 2
NGS	Next Generation Sequencing
NHEJ	non-homologous end-joining
NICD	Notch intracellular domain
NK	natural killer
NKX2-1	NK2 homeobox 1
NRAS	Neuroblastoma RAS viral oncogene homolog
nt	nucleotide
OIS	oncogene induced senescence
p70S6K	p70 ribosomal protein S6 kinase
PAM	protospacer adjacent motif
PBS	phosphate-buffered saline
PDK1	phosphoinositide-dependent kinase-1
PHF6	PHD finger protein 6
PI	propidium iodide
PI3K	phosphatidylinositol-3 kinase
PIK3CD	phosphatidylinositol-4,5-Bisphosphate 3-Kinase Catalytic Subunit Delta
PIP2	posphatidylinositol-3,4-bisphosphate
PIP3	phosphatidylinositol-3,4,5-trisphosphate
PLT	platelet count
PTBP2	polypyrimidine tract binding protein 2
PTEN	phosphatase and tensin homolog
PTPN2	protein tyrosine phosphatase non-receptor type 2
R	repeat sequence
RAG1	recombination activating 1
Ras	rat sarcoma virus

RBPJ	recombining binding protein suppressor of hairless
rev	regulator of virion
RNA	ribonucleic acid
RPMI	Roswell Park Memorial Institute
RTK	receptor tyrosine kinase
RT-qPCR	real-time quantitative polymerase chain reaction
RUNX1	Runt-related transcription factor 1
S6K-1	S6 kinase-1
SCF	stem cell factor
SD	standard deviation
SFFV	Spleen-Focus Forming Virus
sfGFP	superfolder green fluorescence protein
sgRNA	single-guide RNA
SIN	self-inactivating
SIN-LV	self-inactivating lentivirus
spCas9	Streptococcus pyogenes Cas9
ssGSEA	single-sample gene set enrichment analysis, single-sample gene set enrichment analysis
STAT5	signal transducer and activator of transcription
ST-HSC	short-term hematopoietic stem cell
SUZ12	Polycomb protein SUZ12
TAL1	TAL bHLH transcription factor 1
T-ALL	acute T cell lymphoblastic leukemia
TBS	Tris-buffered saline
TCRα	TCR α chain
TdT	terminal deoxynucleotidyl transferase
TF	transcription factor
Thy1.1	thymus cell antigen 1.1
TLX1	T cell leukemia homeobox 1
TPO	thrombopoietin
tracrRNA	trans-activating CIRSPR RNA
TTS	transcriptional start sites
U3	unique 3' region
U5	unique 5' region
USP7	Ubiquitin specific peptidase 7
VEGF	vascular endothelial growth factor
VSVg	vesicular stomatitis virus envelope
x g	times gravity
ZAP-70	zeta-chain-associated protein kinase 70

7.5 List of Figures

7.6 List of tables

7.7 List of publications

Klatt, D., Ha, T. C., **Schinke, M.**, Selich, A., Lieske, A., Dahlke, J., …, Schambach, A. (2020). Competitive sgRNA Screen Identifies p38 MAPK as a Druggable Target to Improve HSPC Engraftment. *Cells*, *9*(10), 1–19. https://doi.org/10.3390/cells9102194

Schinke, M., Meyer, G., Rafiei Hashtchin, A., Hetzel, M., …, Ackermann, M. (2022). Polarization of human iPSC-derived macrophages directs their immunological response to secondary pro-inflammatory stimuli. Journal of Immunology and Regenerative Medicine, 17, 100061, https://doi.org/10.1016/j.regen.2022.100061

8 Author's statement of contribution

Adrian Schwarzer and Axel Schambach supervised this project. Both contributed to the experimental design and interpretation of the results. They performed proof-reading of this thesis and suggested corrections.

I performed design and cloning of the "inverse vector" and "stop vector" (pL40C vector backbone kindly provided by Adrian Schwarzer and Felix Adams), including cloning of all fluorescence reporter proteins, transcription factors and mutant signaling proteins ($NRAS^{G12D}$ was kindly provided by Daniel Brand and Michael Morgan). The sgRNA vector backbone and the sfGFP-reporter vector were kindly provided by Dirk Heckl. The Cre-encoding vector was provided by Tobias Mätzig. All sgRNAs were designed and tested by myself, except for the *Notch1* sgRNA (kindly provided by Marc-Jens Kleppa). I performed lentiviral vector production, titration, cell cultivation and the immortalization assay (32D cells were provided by Adrian Schwarzer and BAF3 cells by Miriam Hetzel). T cell differentiation was conducted under supervision of Ekaterina Takmakova. I generated the *LSL.Cas9 x Lck-cre* mouse strain, performed breeding, genotyping and tissue preparation. I prepared proteins lysates and performed Western Blot analysis. I received help from Jenni Fleischauer when isolating BM cells. Transplantation, mouse visit, retroorbital blood collection, final analysis, blood smear and cytospin preparation and staining were performed by myself. I designed flow cytometry panels for final analysis and *in vitro* cultivation and conducted flow cytometry analysis. I extracted DNA, RNA and performed VCN analysis with an established protocol from Michael Rothe. I performed demultiplexing, trimming, mapping, indel counting and analysis of all NGS data (after initial help from Marc-Jens Kleppa). I performed RNA-Seq analysis with an established R script from Adrian Schwarzer and received help for data processing and for discussion from Anton Selich. I performed all statistical analysis and visualization with GraphPad Prism 5 and R Studio. Marc-Jens Kleppa and Ekaterina Takmakova proof-read this thesis and suggested corrections.

Next Generation Sequencing was performed at the Research Core Unit Genomics (RCUG), MHH. Quality check and sequencing was performed by Lutz Wiehlmann and Marie Dorda. RNA sequencing analysis was performed at Novogene Co., Ltd. Sequencing. Demultiplexing of data, count table generation and gene annotation was performed by employees of Novogene Co, Ltd. Cell sorting experiments were performed at the "cell sorting core facility" of the MHH by Matthias Ballmaier. Dirk Schlaudien (Fraunhofer, ITEM) scanned all cytospin and blood smear slides.

10 Acknowledgments

First of all, I would like to thank my supervisor Prof. Dr. Axel Schambach for his excellent guidance and support at each stage of my doctorate. I greatly appreciate his advice, ideas, our meetings, and the chance to pursue my research interests at the Institute of Experimental Hematology.

I would like to express my sincere gratitude to my supervisor Dr. Adrian Schwarzer, PhD for his guidance, catching motivation, and drive for this project. I greatly appreciate your constructive criticism and honest feedback. Thank you for sharing your scientific knowledge and for your solution−oriented advice. And also, thank you for the support of my political commitment at university and for enabling me to participate at various conferences.

I thank my co-supervisors Prof. Dr. Michael Heuser and Prof. Dr. Dirk Heckl for their scientific advice and constructive criticism. In this line, I would also like to thank Dirk Heckl for providing me with various vectors and expertise for different CRISPR-Cas9 applications.

I thank the MD/PhD Program “Molecular Medicine” for the opportunity to perform my research in an international and interdisciplinary environment. In this context, I would like to express my gratitude to Dr. Susanne Kruse for organizing all the lectures, exciting and interesting soft skills seminars, and her administrative help.

A big thanks go to Dr. Marc-Jens Kleppa, who always had an open ear and useful advice for any problems that I faced during my project. I gained most of my lab skills thanks to him. I am also thankful for his critical prove-reading of my thesis and for taking care of my cells during my holidays.

I would like to thank Dr. Denise Klatt for introducing me in the mouse work, which was a big part of my work. And also for her advice and the opportunity to cooperate with her for my first co-authorship.

I would like to thank Jenni Fleischauer for all the help with the mice. I really enjoyed our BM preparation sessions, where we worked nicely as a team.

I would like to thank Philippe Vollmer Barbosa for letting me join the iGUARD team. It was fascinating to see how much he is burning for this topic, and his motivation and efforts for bringing this project forward.

I would like to thank Dr. Anton Selich for all the discussions and advice for any problem I had in the lab. I am very thankful for our discussions about my RNA Seq data and your help during the analysis.

A special thanks go to Dr. Michael Rothe and all members of the 'AGRO and friends' group. They let me become part of this special group, where I instantly felt warmly welcome and valued as a scientist. Thank you for all the discussions and the easy company. In this line, thanks go to Dr. Antonella Bastone for all the input and discussions on T cells.

I would like to acknowledge Prof. Dr. Michael Morgan, Dr. Tobias Mätzig, Dr. Tengcheong Ha, Dr. Melanie Galla and Dr. Olga Kustikova for providing me with various vectors and materials.

Big thanks go to my (previous) office colleagues Larissa Nassauer, Julia Dahlke, Jenni Fleischauer, Anton Selich, Ekaterina Takmakova, Luisa Weißköppel, Friederike Mansel, and Felix Warnecke. We really had a good time together and a lot of fun! Especially during the beginning of Corona, I appreciated the social time we spent together, enabling us to feel some "normality". Another special thanks goes to Ekaterina Takmakova, for many helpful discussions on my topic, the help with my T cell differentiation, and for prove-reading my thesis.

Thanks to all of my other Experimental Hematology colleagues for the pleasant and familiar atmosphere they created in and outside the laboratory. I really appreciated our "Stable Table" and all the time and experiences we spent together. And a big thanks goes to the "Mensa Gang" for exhilarating lunchtime throughout the years.

I am very grateful to all my good friends for your support and friendship throughout the last few years.

My biggest thanks goes to Juliette! I've lost count of how many times I've received your advice and been able to improve personally and professionally as a result. You cheered me up, supported me, and I can't imagine how I would have progressed without you. Thank you! And thank you for being who you are. Someone I truly admire.

Lydia, there are too many things about you that I am grateful for to list them all here. But the most important thing is: Thank you for becoming a part of my life and for supporting me so much during the last phase of my dissertation. With you I can achieve anything!

Mein größter Dank gilt meiner Familie – meinem Bruder und im Besonderen meinen Eltern. Danke, dass ihr es mir ermöglicht habt, meinen Weg zu gehen, mir immer zur Seite steht und so selbstverständlich immer für mich da seid.

11 Declaration

Herewith, I confirm that I have written the present PhD thesis myself and independently, in compliance with “the policy of Hannover Medical School on the safeguarding of good scientific practice and procedural rules for dealing with scientific misconduct” and that I have not submitted it at any other university worldwide.

Herewith, I agree that MHH can check my thesis by plagiarism detection software as well as randomly check the primary data. I am aware that in case of suspicion, ombudsman proceedings according to § 9 of MHH ’Guidelines of Hannover Medical School to guarantee good scientific practice and dealing with scientific fraud’ will be initiated. During such proceedings, the PhD process is paused.

Hannover, 06th of April 2023