Single-cell CRISPR screening characterizes transcriptional deregulation in T-cell acute lymphoblastic leukemia

by Sarah Meyers, Olga Gielen, Jan Cools, and Sofie Demeyer

Received: December 27, 2023.
Accepted: May 17, 2024.


Publisher’s Disclaimer.
E-publishing ahead of print is increasingly important for the rapid dissemination of science. Haematologica is, therefore, E-publishing PDF files of an early version of manuscripts that have completed a regular peer review and have been accepted for publication. E-publishing of this PDF file has been approved by the authors. After having E-published Ahead of Print, manuscripts will then undergo technical and English editing, typesetting, proof correction and be presented for the authors’ final approval; the final version of the manuscript will then appear in a regular issue of the journal. All legal disclaimers that apply to the journal also pertain to this production process.
Single-cell CRISPR screening characterizes transcriptional deregulation in T-cell acute lymphoblastic leukemia

Authors
Sarah Meyers\textsuperscript{1,2,3}, Olga Gielen\textsuperscript{1,2,3}, Jan Coools\textsuperscript{1,2,3,*}, Sofie Demeyer\textsuperscript{1,2,3,*}

Affiliations
\textsuperscript{1} Center for Human Genetics, KU Leuven, Leuven, Belgium.
\textsuperscript{2} Center for Cancer Biology, VIB, Leuven, Belgium.
\textsuperscript{3} Leuven Kanker Instituut (LKI), KU Leuven – UZ Leuven, Leuven, Belgium.

Short title: Single-cell CRISPR screening in T-ALL

*Corresponding authors
Jan Coools (jan.coools@kuleuven.be), Sofie Demeyer (sofie.demeyer@kuleuven.be)

Author contributions
SM conceptualized the research, performed experiments, analyzed and visualized the data and wrote the manuscript. OG performed experiments. JC conceptualized the research, analyzed data and wrote the manuscript. SD conceptualized the research, performed bioinformatic and other data analyses and wrote the manuscript.

Data availability
All data are available in the main text or the supplementary files. Fastq files and count data of single-cell CRISPR screening experiments have been deposited in NCBI’s Gene Expression Omnibus and are accessible through GEO Series accession number GSE222378. (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE222378)

Acknowledgments
This study was supported by the KU Leuven Genomics Core facility and VIB Flow Core facility.

Funding
This work was funded by FWO (11L9122N) and KU Leuven PhD fellowships to SM, Foundation Against Cancer postdoctoral fellowship to SD, VIB Tech Watch Technology Implementation fund (JC) and KU Leuven funding (C14/18/104, JC)
Competing interests

Authors declare that they have no competing interests.
ABSTRACT

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive type of leukemia caused by accumulation of multiple genetic alterations in T-cell progenitors. However, for many genes it remains unknown how their mutations contribute to disease development. Therefore, we performed two single-cell CRISPR screens in primary pro-T cells ex vivo to study the transcriptional impact of loss-of-function alterations in T-ALL and correlate this with effects on cell fitness. The various perturbations were clustered based on their effects on E2F/MYC or STAT/NOTCH signatures, which play a defining role in driving T-cell proliferation. Many of the perturbations resulted in positive effects on the STAT and NOTCH signatures and were predicted to behave as haploinsufficient tumor suppressors in T-ALL. Additionally, Spi1 was identified as an essential gene for pro-T cell survival, associated with deregulation of the MYC signature and epigenetic consequences. In contrast, Bcl11b was identified a strong tumor suppressor gene in immature T lymphocytes, associated with deregulation of NF-kB and JAK/STAT signaling. We found a correlation between BCL11B expression level and JAK/STAT pathway mutations in T-ALL patients and demonstrated oncogenic cooperation between Bcl11b inactivation and JAK3 hyperactivation in pro-T cells. Altogether, these single-cell CRISPR screens in pro-T cells provide fundamental insights in the mechanisms of transcriptional deregulation caused by genetic alterations in T-ALL.
Introduction

T-cell acute lymphoblastic leukemia (T-ALL) is an aggressive form of leukemia characterized by high white blood cell counts and infiltration of immature T cells into the bone marrow and other tissues. Sequencing studies of patient cohorts have identified many genetic alterations in T-ALL patients, often in major signaling pathways regulating T-cell development such as NOTCH, JAK/STAT, PI3K/AKT, RAS and the T-cell receptor signaling pathway. While frequent oncogenic events such as NOTCH1 mutations have been extensively studied, the contribution of other mutations, for example in epigenetic regulators such as CTCF, remains poorly understood.

Many of the current in vitro models for T-ALL have their limitations, since immortalized leukemic cell lines intrinsically have a complex genetic background with numerous mutations and structural abnormalities. In recent years, a pro-T cell model was developed to culture murine primary T-cell precursors ex vivo. This model has a wild-type genetic background, facilitating the study of gene functions and oncogenic cooperation in T-ALL in a clean way. Pro-T cells are generated by isolation of murine lineage-negative bone marrow cells and culturing these primary cells ex vivo in the presence of mDLL4 (stimulating NOTCH), mIL7 (stimulating JAK/STAT) and mSCF (stimulating c-KIT). This drives their differentiation into immature double-negative T lymphocytes, which were shown to represent a suitable model for studying T-cell development and transformation. We have previously shown that recurrent mutations in T-ALL can drive cytokine-independent proliferation of pro-T cells, demonstrating that this model effectively simulates leukemic transformation ex vivo.

Because many of the recurrent genetic alterations in T-ALL patients remain largely uncharacterized, we used single-cell CRISPR screening to study a number of potential tumor suppressor genes in pro-T cells. This technique combines CRISPR/Cas9 screening with single-cell RNA-sequencing (scRNA-seq) and allows the interrogation of multiple genetic alterations while studying the full transcriptome associated with each perturbation. Furthermore, gRNA enrichment or depletion can be assessed to evaluate the impact of each perturbation on cell fitness. Specific measures are required for detection of the gRNA in a single cell, either through a unique barcode or by directly sequencing the gRNA using a modified construct with a poly-A tail or capture sequence. Single-cell CRISPR screening has been applied previously in T cells by Zhou et al., where they studied different transcription factors that drive the T-cell developmental trajectory. In general, single-cell sequencing is increasingly used to study the heterogeneity of leukemia at single-cell resolution.

Next to a number of major T-cell regulators that are known to play a role in T-ALL development, we targeted multiple genes that are recurrently mutated in patients but for which their exact function
remains to be elucidated. In a first experiment, we perturbed 17 known transcriptional regulators, to enable optimization of the methodology and identification of the driving gene signatures. We subsequently designed a larger library targeting 42 recurrently mutated genes, with the aim of characterizing their transcriptional profiles, providing fundamental insight in the function of these genes and the affected downstream signaling pathways.

**Methods**

**Experimental design**

Single-cell CRISPR screening was performed on ex vivo cultured pro-T lymphocytes to study recurrent mutations in T-ALL. Custom gRNA libraries were designed and retrovirally transduced in pro-T cells, with subsequent FACS sorting and cell harvesting at multiple timepoints. Samples were loaded onto the 10X Genomics Chromium for single-cell encapsulation and mRNA and gRNA extraction. Transcriptome data were analyzed to obtain gene expression signatures per perturbation and gRNA counts were used to infer proliferative effects.

**Cell culture and transduction**

Pro-T cell cultures were established as described previously from C57BL/6J (Charles River Laboratories, Saint-Germain-Nuelles, France) Cas9 knock-in transgenic mice (024858, Jackson Laboratories, Bar Harbor, ME, USA). Lineage-negative hematopoietic stem cells were purified from the bone marrow of C57BL/6 mice using the EasySep Mouse Hematopoietic Progenitor Cell Isolation Kit (StemCell Technologies). The use of mice for the purpose of establishing pro-T cell cultures was approved and supervised by the KU Leuven ethical committee (ECD 030/2023, approval date 16th March 2023). Pro-T cells were maintained in non-tissue culture treated plates pre-coated with anti-human IgG1-Fc mAb (Abcam ab1927). Pre-coated wells were washed twice and subsequently coated with 2 μg/mL mDLL4-Fc in RPMI-1640. Pro-T cells were seeded at 500000 cells/mL in medium supplemented with 20% FBS with mIL7 (20 ng/mL, PeproTech), mSCF (20 ng/mL, PeproTech) and primocin (100 μg/mL, Invivogen). Cells were transduced by retroviral spinfection at 2500 rpm for 90 minutes at 30°C, with 0.5 μL/mL polybrene (10mg/mL, Sigma Aldrich TR-1003-G) per million cells. Next, cells were incubated at 37°C for 2 hours, washed with PBS and seeded into freshly coated wells. For proliferation assays using gRNAs or shRNAs, cells were seeded at 500000 cells/mL and cell number, viability and fluorescent signal were monitored using a MACSQuant VYB (Miltenyi) flow cytometer. For drug treatments, 50000 cells were seeded in 100 μL RPMI-1640 medium with 20% FBS with mIL7, mSCF and primocin in 96-well plates. The D300e digital dispenser (TECAN) was used for dispensing specific concentrations of ruxolitinib phosphate (MedChemExpress) in triplicate. Cell
viability was evaluated after 60 hours using the ATPlite luminescence assay (PerkinElmer) on a Victor
X4 plate reader (PerkinElmer).

**Single-cell CRISPR screening**

Pro-T cells with constitutive Cas9 expression were transduced with the CRISPR library with an
efficiency around 10%. Transduced cells were sorted 48 hours post-transduction using an Influx cell
sorter (BD Biosciences) or FACSaria Fusion (BD Biosciences). Sorted cells were cultured for a total of
14 days, with harvesting at day 3, 7 and 14. At each timepoint, cell suspensions of 1000 cells/μL in
PBS + 0.04% bovine serum albumin were loaded onto the Chromium controller (10X Genomics) for
single-cell encapsulation. Sequencing libraries were prepared according to manufacturer’s directions
using the Chromium Next GEM single-cell 3’ reagents with feature barcoding technology. Libraries
were analyzed with a High Sensitivity DNA kit on a Bioanalyzer (Agilent Technologies) and sequenced
on a NovaSeq 6000 instrument (llumina).

**Statistical analysis and data visualization**

Graphpad Prism (version 9.2.0) was used for statistical analyses. Data was displayed as mean +/-
standard deviation. Comparisons between two groups were performed via unpaired two-tailed
Student t-tests. One-way analysis-of-variance (ANOVA) was used for comparisons of multiple groups
in proliferation assays, using Tukey’s correction for multiple testing. Prior to ANOVA, a Shapiro-Wilk
normality test was performed to test for Gaussian distribution of residuals and Brown-Forsythe test
for homogeneity of variance. P values were indicated by ****p < .0001, ***p < .001, **p < .01, *p <
.05 or not significant (ns). Significance of overlap between two lists was calculated via the
hypergeometric probability distribution and visualized in Venn diagrams. Box plots show a line at the
median and box from Q1 to Q3, with whiskers ranging from minimum to maximum values. FlowJo
(version 10.8.1) was used for analysis of flow cytometry data. Custom R (version 4.1.1) scripts were
used for analysis of single-cell and bulk RNA-sequencing data, as well as for data visualization, as
further elaborated on in supplemental methods. Webgestalt21 and the Broad GSEA software (version
4.2.1)22,23 were used for analysis of differentially expressed genes, and i-cisTarget24 for prediction of
cis-regulatory motifs.

**Results**

**Transcriptional phenotyping of perturbations in pro-T cells by single-cell CRISPR screening**

We first performed a CRISPR screen using a gRNA library targeting 17 important T-cell regulators with
a known role in T-ALL (Table S1). Five gRNAs were designed per gene, together with three non-
expressed genes as negative controls (NT1 - 3), resulting in a library of 96 gRNAs (Table S1). The 10X
Genomics Chromium 3’ RNA-seq chemistry with feature barcoding technology was used to detect gRNAs via their capture sequence (Fig. S1a). Different capture sequence configurations were evaluated experimentally first in Ba/F3 cells and subsequently in the pro-T cell model (Fig. S1b).

Pro-T cells were transduced with the gRNA library with a low multiplicity of infection (<10%) to avoid introducing multiple gRNAs in a single cell. The transduced cells were subsequently selected by fluorescence-activated cell sorting (FACS) (Fig S1a, c) and samples were harvested at day 3, 7 and 14 post-transduction (Fig. 1a). After filtering bad quality cells and doublets, the dataset contained 4773, 3606 and 9672 cells for each respective timepoint, with 45% to 54% of cells containing a single gRNA (Fig. 1b). At day 3, a median of 112 good-quality cells (interquartile range [IQR]: 97-133) per target gene and a median of 21 cells (IQR: 16-27) per gRNA were sequenced (Fig. 1c). At this timepoint, we observed an almost equal distribution of gRNA abundance across the cell population (Fig. 1d, Fig. S1d). Since the cells were sequenced at multiple timepoints, the impact of each perturbation on cell proliferation could be studied, thereby identifying essential genes and tumor suppressors. Cells with perturbation of Il7r, Notch1, Myc, Kit and Spi1 had a significant proliferative disadvantage, in agreement with the fact that pro-T cells are dependent on JAK/STAT, NOTCH and KIT stimulation. In contrast, cells with Pten inactivation had a strong growth advantage (Fig. 1d, e), which was expected since PTEN is a well-described tumor suppressor. Cells with Pten gRNA had become the dominant clone by day 14, covering over 51% of the cell population and outcompeting clones with other perturbations. Comparison of the trends of all gRNAs over time with the NT cells also identified Fos, Jun, Ptprc, E2f1 and Runx1 as mild tumor suppressor genes (Fig. 1e).

**Perturbation of major transcriptional regulators identifies essential gene expression signatures**

To study the transcriptome profiles at single-cell level, the cDNA libraries were sequenced with 108400, 160011 and 59678 mean reads per cell for day 3, 7 and 14, respectively. This resulted in a median number of detected transcripts per cell of 4285 (IQR: 3406 – 5264), 5106 (IQR: 4186 – 5889) and 3649 (IQR: 2948 – 4390), respectively (Fig. 2a). We observed efficient CRISPR-mediated downregulation for the majority of target genes (Fig. 2b, S2). However, downregulation of gene expression is not a perfect measure for knockdown efficiency as CRISPR editing could result in intact mRNA generating a non-functional protein. Indeed, for some genes, such as Myc and Il7r, we did not observe downregulation of the gene itself, but we did detect changes in expression of known downstream target genes (Fig. S1e, S2). Cis-regulatory motif analysis of the significant differentially expressed genes also demonstrated enrichment of these transcription factor binding motifs (Fig. 2c). In summary, combined analysis of the expression of the gene targeted by the gRNA, the expression
of downstream targets and a regulatory motif analysis provides a robust analysis to evaluate perturbation efficiency in silico.

Cell cycle phase was estimated based on the expression of cell cycle marker genes. A strong correlation was observed between proliferation based on cell number (Fig. 1e) and predicted phase distribution based on marker gene expression (Fig. 2d). In cells with Myc, Kit, Notch1 and Spi1 gRNA, a decreased fraction of cells resided in the G2M/S phase, while the number of cells in the stationary G1 phase was elevated. In contrast, cells with Pten, E2f1, Ptprc, Cdkn2a, Jun and Runx1 inactivation showed a higher proportion of cells in the G2M/S phases, indicative of more active cycling (Fig. 1e, 2d).

Based on the scRNA-seq data, we aimed to identify core gene expression signatures driving the response of pro-T cells to CRISPR perturbation. Since T lymphocytes are highly dependent on STAT and NOTCH signaling, with MYC as downstream effector, we focused on these pathways in particular. Enrichment analyses were performed on the ranked lists of differentially expressed genes of the pseudo-bulk transcriptomes, using a selection of JAK/STAT, NOTCH and MYC gene sets.\textsuperscript{7,26,27} Significant negative enrichment of multiple relevant gene sets was observed in cells with the respective perturbations. The signatures described by Bornschein et al.\textsuperscript{7} ranked amongst the most significant enriched gene sets and were presumed to describe the behavior of pro-T cells most accurately (Table 1, Fig. 2e).

**Spi1 is an essential gene linked to MYC transcriptional activity**

PU.1, encoded by Spi1, is an essential transcriptional regulator responsible for regulation of both myeloid and lymphoid development, in a time- and dose-dependent manner. In the T-cell lineage, Spi1 is typically only expressed in early progenitors, and is responsible for delaying commitment and maintaining a multipotent state while allowing strong proliferation of immature T-cell precursors. Upon commitment, Spi1 is silenced, which is associated with upregulation of specific T-lineage genes and further progression through T-cell development.\textsuperscript{28,29} Our CRISPR data identified Spi1 as an essential gene for pro-T cell proliferation (Fig. 1e). Independent proliferation assays were performed to confirm the negative impact of Spi1 inactivation on cell fitness (Fig. 3a, b), while in contrary, Spi1 overexpression had no impact on proliferation (Fig. 3c). To complement the scRNA-seq data, bulk RNA-seq was performed on cells with Spi1 inactivation. We found 3097 significant differentially expressed genes upon Spi1 loss, with a significant overlap between the single-cell and bulk datasets (Fig. 3d, S3a, Table S2).
Spi1 inactivation was associated with a significant negative enrichment of MYC target genes in both the bulk data as well the single-cell dataset (Fig. 3e, f). Furthermore, cell cycle regulators, glycolysis and oxidative phosphorylation were downregulated, as well as E2F target genes (Fig. 3e). Single-Cell regulatory Network Inference and Clustering (SCENIC) was used to infer regulons associated with Spi1 perturbation based on the bulk data. This analysis also showed downregulation of Myc-related regulons, as well as downregulation of the E2f and Cebpβ regulons, the latter being a known SPI1 binding partner (Fig. 3g, Fig. S3b, c). Since Myc and its downstream targets are crucial for pro-T cell survival and since the E2F transcription factor family is heavily involved in cell cycle regulation, downregulation of these signatures is in agreement with the observed negative impact of Spi1 inactivation on cell proliferation. In contrast to the negative effects on Myc, regulons associated with epigenetic regulation as well as regulons associated with Ets1 and other ETS factors were upregulated, indicating an impact on the epigenetic landscape and the ETS transcription factor family upon Spi1 loss (Fig. 3g, Fig. S3d, e).

**Characterizing the transcriptomic impact of recurrent mutations in T-ALL**

We next selected 42 poorly characterized genes that are recurrently mutated in T-ALL patients with predicted loss-of-function alterations. The CRISPR library was designed with five gRNAs per gene, 11 NT gRNAs and 10 positive controls, resulting in a total of 231 gRNAs (Table S3). Single cells were sequenced at three timepoints (3, 7 and 14 days after transduction), generating a dataset of 18032, 19718 and 16135 cells passing filtering at each respective timepoint, with a single gRNA detected in 5280, 8123 and 6084 cells, respectively (Fig. S4a). Other quality metrics were comparable to those of the first CRISPR screen (Fig. S4b, c, d). For the majority of perturbations, downregulation of the target gene was observed, indicating efficient perturbation (Fig. S2, S4e).

In general, we aimed to investigate whether the transcriptomic impact of different perturbations converges on common downstream pathways. The data of both CRISPR screens were merged, with regression of batch and cell cycle effects. Enrichment analyses were performed based on comparison of the pseudo-bulk transcriptomes using the previously defined signatures as reference (i.e. NOTCH, STAT, E2F and MYC) (Fig. 4a, b, S4f). We observed a large group of perturbations (cluster A) with upregulation of both STAT and NOTCH, and downregulation of the E2F signature. Within this cluster, the effect on MYC transcriptional activity was variable, but often resulted in downregulation of the MYC signature. Similar to direct perturbation of Myc (Fig. S4f), these perturbations were commonly associated with upregulation of NOTCH and STAT signatures. Cluster A contained multiple genes for which their function in T-ALL remains unclear, such as Tsyp12, Trapr, Zbtb7a and Pbm1 (Fig. 4a). These genes clustered alongside multiple known epigenetic regulators, including Ccfl, Arid1a,
Smarca4 and Esth2, suggesting that these factors may impact the transcriptome of pro-T cells in a similar way.

By quantifying the number of cells over time, we analyzed the impact of each gRNA on cell proliferation (Fig. 4c, S4g). While many of the perturbations had a negative effect on cell growth, only few had a positive effect. Good correspondence was found between cell proliferation and cell cycle phase as predicted by the single-cell RNA-seq data (Fig. 4d). Loss of Dnm2, Med12, Gata3, Smarca4, Arid1a, Ncor1 and Cctf caused the strongest growth disadvantage, identifying these genes as dependency factors in pro-T cells. Data from the DepMap cancer dependency portal also showed negative CRISPR scores for those genes in lymphoid cell lines (Fig. 4e, S5). Using our single-cell data, we could link inactivation of these genes with a strong negative effect on MYC and E2F signatures (Fig. 4a). In contrast, we observed a growth advantage for cells with Pten, Cdkn2α, Runx1, Bcl11b, Crebbp, Bcor, Setd2, Dnmt3a and Trp53 gRNAs (Fig. 4c). These genes were thus identified as tumor suppressors in pro-T cells and many are indeed known as such. Inactivation of these tumor suppressor genes was associated with upregulation of either the MYC/E2F signatures (e.g. Pten, Runx1, Cdkn2α, Ptprc, Jun) or the STAT signature (e.g. Crebbp, Trp53, Bcl11b, Setd2, Dnmt3a) (Fig. 4a, b).

We next applied SCENIC to study regulons associated with the different perturbations. We observed considerable variation in the expression of regulons of several ETS family members, such as Spi1, Fli1, Ets1 and Erg. Most strikingly, the Ets1 regulon, as well as the Ets1 gene, were upregulated specifically in cells with Spi1 gRNA and not in any of the other perturbations (Fig. 4f, Fig. S4h, S6). The Spi1 regulon was downregulated in cells with Spi1 and Pten gRNA, and upregulated in cells with Ptpr2, Gata3, Crebbp and Arid1a loss. We observed that the Cebpb regulon was active in about half of the perturbations, while the Cebpregulon is active in roughly the other half of the cells, suggesting an antagonistic function (Fig. S6). Interestingly, we found the Erg regulon upregulated in cells with a growth disadvantage (i.e. Arid1a, Cctf, Dnm2, Kit, Kmt2c, Kmt2d, Myc and Spi1 inactivation) (Fig. 4f). Upregulation of Erg seems to interfere with cell proliferation, and it was previously described that Erg knockout leads to a highly proliferative phenotype.

Bcl11b is a tumor suppressor and cooperates with activated JAK/STAT signaling.

We next focused on Bcl11b, which was identified as one of the strongest tumor suppressor genes in pro-T cells in our CRISPR screen (Fig. 4c). To validate the results of the screen, we performed independent Bcl11b inactivation by two gRNAs, confirming the increase in cell proliferation (Fig. 5a) and the direct effect on BCL11B protein level (Fig. 5b). Similarly, Bcl11b knockdown by shRNA also
resulted in increased cell proliferation, although at a more modest rate as compared to CRISPR perturbation (Fig. S7a). In contrast, Bcl11b overexpression significantly impaired cell growth, confirming the growth-suppressing role of BCL11B (Fig. 5c).

Single-cell RNA-seq indicated that loss of Bcl11b in pro-T cells was associated with increased STAT and E2F signatures (Fig. 4a, b). For in-depth characterization of the impact of Bcl11b inactivation, bulk RNA-sequencing was performed. We found 2690 significant differentially expressed genes, with significant overlap with the single-cell dataset (Fig. 5d, S7b, Table S2). These data were integrated with a publicly available ChIP-seq dataset to define high-confidence direct BCL11B target genes (Fig. 5e). Enrichment analysis of these direct target genes demonstrated significant upregulation of inflammatory signatures, NF-kB and STAT signaling (Fig. 5f). Cis-regulatory motif analysis showed that RELA, STAT2 and SPI motifs were positively enriched, while EZH2, FOXO1 and SOX11 motifs were negatively enriched (Fig. 5g). In addition to enrichment of the NF-kB pathway and RELA motif, we found increased P65 protein level and phosphorylation, as well as elevated expression of the Rela and Nfkb2 regulons (Fig. 5h, S7c). Additionally, several genes of the BCL2 family (Bcl2a1, Bim, Mcl1) were upregulated (Table S2), which are negative regulators of apoptosis and known NF-kB target genes. Besides NF-kB activation, the data showed significant upregulation of STAT signaling induced by Bcl11b loss (Fig. 5f, g, h). Together, these data identified Bcl11b as a strong tumor suppressor in pro-T cells, associated with deregulated NF-kB and STAT signaling.

To validate these findings in T-ALL patient samples, we used a publicly available RNA-seq dataset of 264 T-ALL samples. We divided the patients in three groups based on BCL11B expression levels (Fig. 6a, S8) and found that T-ALL cases with the lowest BCL11B expression also showed significant enrichment of NF-kB and STAT signaling (Fig. 6b). In concordance, we found significantly reduced Bcl11b expression in T-ALL patients with alterations in the JAK/STAT signaling pathway (Fig. 6c). This two-way correlation led us to hypothesize that BCL11B inactivation could have a synergistic effect in combination with activating JAK/STAT alterations. Indeed, Bcl11b perturbation in combination with overexpression of the activating JAK3<sup>M511I</sup> mutant was able to drive cytokine-independent growth of pro-T cells, while single mutants did not have transforming potential (Fig. 6d). Treatment with ruxolitinib demonstrated that pro-T cells with loss of Bcl11b had decreased sensitivity towards pharmacological JAK targeting (Fig. 6e), in agreement with a direct effect of BCL11B on STAT transcriptional activity.

Discussion
This work describes a comprehensive study of a number of recurrently mutated genes in T-ALL, which typically harbor loss-of-function mutations but remain poorly characterized. The aim was to determine whether these genes have tumor suppressive characteristics and to correlate this with specific effects on gene expression. Single-cell CRISPR screening enabled us to study the effects of gene perturbation on proliferation as well as transcriptional regulation. An ex vivo model of primary pro-T cells was used, as these immature cells are likely closely related to the cells of origin of T-ALL. We initially validated the methodology by targeting well-characterized T-cell regulators, achieving efficient downregulation of the majority of genes, as well as significant differential expression of known downstream target genes, pathways and enriched cis-regulatory motifs. Furthermore, these data enabled us to define the core transcriptional signatures driving pro-T cell growth.

We subsequently performed a CRISPR screen targeting 42 genes which are frequently altered in T-ALL patients. By assessing cell number over time, these perturbations were linked with their effects on cell proliferation. Interestingly, we observed upregulation of the STAT signature by the majority of perturbations. It is well-established that deregulated JAK/STAT signaling is an important driver of T-ALL, as illustrated by the numerous mutations this pathway (i.e. in JAK3, JAK1, IL7R, PTPN2 etc.) that are detected in T-ALL patients. Based on our CRISPR data, we could hypothesize that many genetic mutations eventually lead to deregulated STAT signaling, thereby contributing to disease development. Moreover, we identified a large cluster of genes exhibiting upregulation of both the STAT and NOTCH signatures, with a smaller subcluster characterized by additional downregulation of MYC and associated with reduced cell proliferation. As MYC is a master transcriptional regulator in T cells, it is not surprising that its downregulation is associated with strong fitness effects. For instance, our data showed a strong growth disadvantage for cells with Ctcf inactivation, coinciding with MYC downregulation. This is in line with previous data illustrating a role for CTCF in chromatin looping of the MYC enhancer to the MYC promoter. Remarkably, many of the genes in cluster A encode epigenetic regulators, including Arid1a and Smarca4 as part of the SWI/SNF complex and Kdm6a, Kmt2c and Asxl2 as regulators of histone (de)methylation. Our data suggest that other genes in this cluster, such as Tspyl2 or Zbtb7a, may have similar transcriptional effects. In addition to genes in cluster A, we identified Spi1 as an essential gene for cell survival and linked this dependency to MYC. A relationship between SPI1 and MYC was previously described in myeloid cells where SPI1 perturbation leads to downregulation of MYC targets due to loss of SPI1 binding to MYC enhancers. Additionally, we observed upregulation of Ets1, as well as regulons associated with other ETS factors. These data complement work from the Rothenberg lab who previously characterized the role of several transcription factors, including PU.1, in mouse T-cell development.
Several of the genes in our screen were here identified as essential genes in pro-T cells, despite harboring loss-of-function mutations in T-ALL patients. Our findings are in line with CRISPR data from the DepMap database where many of those genes were similarly recognized as essential in lymphoid leukemia and lymphoma cell lines.\textsuperscript{34} Many of these genes thus seem to behave as haploinsufficient tumor suppressor genes in T-ALL, with complete inactivation by CRISPR not being tolerated. This is likely explained by the fact that many loss-of-function mutations in T-ALL patients are either heterozygous or are point mutations leading to specific changes in protein function, whereas CRISPR typically causes complete knock-out. Still, we demonstrated that also mild knockdown by shRNA had similar effects on pro-T cell proliferation as compared to CRISPR perturbation. It was previously shown that T-ALL patients have a complex genetic landscape and can carry up to 20 different mutations, with many of these alterations showing either significant co-occurrence or mutual exclusivity.\textsuperscript{2,44} Single mutations are often not sufficient for driving leukemia development in mouse models and multiple cooperating mutations are required to initiate disease.\textsuperscript{6–8,10,45,46} We hypothesize that this genetic complexity is an important factor which shapes the disease. As it was previously shown that the order of mutational acquisition in ALL is not random\textsuperscript{47,48}, we reason that the timing of each mutation is crucial and some phenotypes may only manifest in a background where earlier driving mutations are present.

We identified \textit{Bcl11b} as a strong tumor suppressor gene in pro-T cells and demonstrated a direct connection between effects on cell proliferation and gene expression. There is conflicting data on the role of BCL11B that most likely reflects the results obtained in various model systems using various technologies. Our CRISPR screen allowed us to study the inactivation of \textit{Bcl11b} in primary immature T cells. In physiological conditions, \textit{Bcl11b} expression initiates around the commitment phase and pushes progenitor cells forward through T-cell development.\textsuperscript{49} BCL11B insufficiency has been described to lead to loss of commitment, reduced sensitivity to NOTCH stimulation and allows reprogramming of T cells into NK or myeloid fates.\textsuperscript{49,50} In T-ALL, \textit{BCL11B} is deleted in 3\% of patients and shows mutations in around 10\% of cases, often co-occurring with aberrant TLX1 expression. In other cases, a t(5;14)(q35;q32) translocation juxtaposes the BCL11B regulatory region to the TLX3 coding sequence, thereby inactivating \textit{BCL11B}.\textsuperscript{49,51} In our dataset, as well as publicly available RNA-seq data from T-ALL patients, significant upregulation of the NF-kB and STAT pathway was observed in cells with low \textit{Bcl11b} levels. Moreover, we also found low expression of \textit{BCL11B} in T-ALL patient samples with JAK/STAT pathway alterations, indicating that reduced levels of \textit{BCL11B} might cooperate with JAK/STAT pathway activation. Activating JAK3 mutations are frequent events in T-ALL and it was previously demonstrated that JAK3 signaling can cooperate with other factors, such as \textit{Suz12} inactivation and \textit{Il7r} mutations.\textsuperscript{3,6,46,52} In the current study, we found that inactivation of
Bcl11b enhanced the oncogenic effects of mutant JAK3 and drove transformation to cytokine-independent growth. In addition, cells with Bcl11b inactivation showed decreased sensitivity to JAK inhibition, irrespective of JAK3 mutant status. These data, in combination with the single-cell RNA-seq data, indicate that BCL11B functions as an important negative regulator of STAT transcriptional activity in immature T-cells. The absence of Bcl11b results in increased activation of STAT transcriptional activity, making the cells more vulnerable to JAK pathway mutations, yet simultaneously reducing their sensitivity to JAK kinase inhibitors.

Our data demonstrate that single-cell CRISPR screening is a powerful technique for parallel characterization of the transcriptional alterations resulting from numerous gene perturbations. In this way, we characterized Bcl11b as a potent tumor suppressor gene in T-ALL with a link to JAK/STAT and Spi1 as an essential gene associated with direct effect on regulation of Myc. These data offer insight in the function of uncharacterized genes and can be linked with other characteristics such as effects on proliferation or cell cycle.
REFERENCES

### TABLES

Table 1: Gene set enrichment analysis for cells with *Il7r* and *Stat5b* gRNA, *Notch1* gRNA and *Myc* gRNA. Normalized enrichment scores (NES) and significance values were calculated for the MSigDB Hallmark gene sets, KEGG pathways and others.

<table>
<thead>
<tr>
<th>Perturbation</th>
<th>Gene set</th>
<th>NES</th>
<th>p-value</th>
<th>FDR q</th>
<th>NES</th>
<th>p-value</th>
<th>FDR q</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perturbation</td>
<td><em>Il7r</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STAT_SIGNATURE_BORNSCHEIN</td>
<td>-2.09</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td>-1.74</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
</tr>
<tr>
<td></td>
<td>HALLMARK_IL2_STAT5_SIGNALING</td>
<td>-1.73</td>
<td>&lt; 2.2e-16</td>
<td>0.001</td>
<td>-1.45</td>
<td>0.004</td>
<td>0.125</td>
</tr>
<tr>
<td></td>
<td>GOBP_ACTIVATION_OF_JANUS_KINASE_ACTIVITY</td>
<td>-1.48</td>
<td>0.03</td>
<td>0.072</td>
<td>-1.46</td>
<td>0.052</td>
<td>0.214</td>
</tr>
<tr>
<td></td>
<td>HALLMARK_IL6_JAK_STAT3_SIGNALING</td>
<td>-1.23</td>
<td>0.111</td>
<td>0.135</td>
<td>-1.14</td>
<td>0.304</td>
<td>0.597</td>
</tr>
<tr>
<td></td>
<td>JAK_STAT_CASCADE</td>
<td>-1.04</td>
<td>0.394</td>
<td>0.463</td>
<td>-1.22</td>
<td>0.228</td>
<td>0.556</td>
</tr>
<tr>
<td></td>
<td>KEGG_JAK_STAT_SIGNALING_PATHWAY</td>
<td>-1.01</td>
<td>0.445</td>
<td>0.447</td>
<td>-1.02</td>
<td>0.451</td>
<td>0.622</td>
</tr>
<tr>
<td>Perturbation</td>
<td><em>Stat5b</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>STAT_SIGNATURE_BORNSCHEIN</td>
<td>-1.82</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HALLMARK_NOTCH_SIGNALING</td>
<td>-1.61</td>
<td>0.007</td>
<td>0.017</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>GOBP_NOTCH_SIGNALING_PATHWAY</td>
<td>-1.54</td>
<td>&lt; 2.2e-16</td>
<td>0.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>REACTOME_SIGNALING_BY_NOTCH</td>
<td>-1.41</td>
<td>0.014</td>
<td>0.067</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>KEGG_NOTCH_SIGNALING_PATHWAY</td>
<td>-1.4</td>
<td>0.044</td>
<td>0.054</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BIOCARTE_NOTCH_PATHWAY</td>
<td>-1.25</td>
<td>0.181</td>
<td>0.159</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>WP_NOTCH_SIGNALING_PATHWAY</td>
<td>-1.17</td>
<td>0.23</td>
<td>0.194</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Perturbation</td>
<td><em>Myc</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HALLMARK_MYC_TARGETS_V1</td>
<td>-1.96</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>DANG_MYC_TARGETS_UP</td>
<td>-1.91</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>MYC_SIGNATURE_BORNSCHEIN</td>
<td>-1.72</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>HALLMARK_MYC_TARGETS_V2</td>
<td>-1.68</td>
<td>&lt; 2.2e-16</td>
<td>&lt; 2.2e-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>PID_MYC_ACTIV_PATHWAY</td>
<td>-1.62</td>
<td>0.003</td>
<td>0.003</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>COLLER_MYC_TARGETS_UP</td>
<td>-1.59</td>
<td>0.009</td>
<td>0.005</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>BILD_MYC_ONCOGENIC_SIGATURE</td>
<td>-1.41</td>
<td>0.019</td>
<td>0.05</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
FIGURE LEGENDS

Figure 1: Pilot experiment of single-cell CRISPR screening in ex vivo pro-T cells.
a) Experimental setup of the single-cell CRISPR screens. Pro-T cells were generated by isolation of lineage negative (Lin-) hematopoietic stem and progenitor cells (HSPCs) from C57BL/6J CAS9 knock-in mice which are cultured ex vivo in presence of mDLL4, mSCF and mIL7. Custom gRNA libraries were designed and cloned in a retroviral construct with capture sequence. After transduction and FACS selection, cells were cultured for a total of 14 days, with cell harvesting at day 3, 7 and 14 for single-cell library preparation of the mRNA and gRNAs.
b) Pie charts show percentages of cells with zero, one or multiple gRNAs, with indication of the total number of cells passing filtering parameters at each timepoint.
c) Number of cells per gene and per gRNA. The gRNA library contained five gRNAs per target gene. Box plots show the number of sequenced cells passing filter criteria per gene (upper panel) and per gRNA (lower panel). Plots show a line at the median and box from Q1 to Q3, with whiskers ranging from min to max values.
d) Distribution of cells per CRISPR target gene at day 3, 7 and 14. Each slice of the donut represents the fraction of cells with a specific perturbation. Over time, changes in the ratio are observed due to differences in proliferation rate in cells with different perturbations.
e) Evolution of the fraction of cells per perturbation over time across day 3, 7 and 14. Y-axis shows data in logarithmic scale. Non-targeting (NT) controls are indicated in grey font. Perturbations are ordered based on the number of cells at day 14.

Figure 2: Single-cell RNA-seq data demonstrate downregulation of target genes, downstream pathways and predicted cis-regulatory motifs.
a) Violin plots displaying the number of detected transcripts per cell for each timepoint.
b) Normalized expression levels of the target genes in cells with the respective gRNA. To study gRNA efficiency, all cells of the three timepoints were aggregated per perturbation to create ridge plots of the normalized expression level of each target gene. On the top of the graph is indicated for which transcript the expression level is displayed, with the purple ridge highlighting cells with the corresponding perturbation.
c) Enriched cis-regulatory motifs in cells with different perturbations. 1-cisTarget was used to predict regulatory motifs in the significant (padj < 0.05) differentially expressed genes in cells with different gRNAs, with indication of the normalized enrichment scores (NES).
d) Cell cycle phase distribution per perturbation. Based on the normalized expression levels of cell cycle marker genes, cells were classified in S, G2M or G1 phase and perturbations were ordered based on the fraction of cells in G1 phase.

e) Enrichment analyses showing enrichment scores of pro-T signature gene sets in the ranked list of differentially expressed genes in cells with *Stat5b*, *Notch1* and *Myc* perturbation, vs NT cells. Indicated on the graphs are the NES and FDR false discovery rate q-value (FDR q).

**Figure 3: Combined analysis of single-cell and bulk transcriptome of cells with Spi1 perturbation.**

a) Cell densities of BFP-positive cells (mean +/- standard deviation (SD), n = 3 replicates) over time for pro-T cells with *Spi1* gRNA, compared with two NT control gRNAs. Data were analyzed by one-way ANOVA at day 22. P values were indicated with ***p < .001, **p < .01, or not significant (ns). (*Spi1_gRNA1* vs *Spi1_gRNA3*: p=0.9989, *Spi1_gRNA1* vs *NT1_gRNA1*: p=0.0026, *Spi1_gRNA1* vs *NT1_gRNA5*: p=0.0010, *Spi1_gRNA3* vs *NT1_gRNA1*: p=0.0022, *Spi1_gRNA3* vs *NT1_gRNA5*: p=0.0009, *NT1_gRNA1* vs *NT1_gRNA5*: p=0.8374).

b) Median fluorescence intensities (MFI) of SPI1 protein levels in cells with *Spi1* gRNA vs NT by intracellular flow cytometry. Normalized MFI were calculated for AlexFluor568 emission. Three biological replicates were performed, with mean and SD shown. Comparison between groups was performed by two-sided unpaired Student t-test, p<0.0001.

c) Cell densities of GFP-positive cells (mean +/- SD, n = 3 replicates) over time for pro-T cells with SPI1 overexpression (OE) and empty vector (EV). Data were compared at day 15 by two-sided unpaired Student t-test, p = 0.9662.

d) Overlap between significant (padj <0.05) differentially expressed genes in the bulk and single-cell data for *Spi1* perturbed cells. P-values for the overlaps were calculated using the hypergeometric probability distribution.

e) Gene set enrichment analysis showing NES scores for the Hallmark gene sets in the ranked list of differentially expressed genes in cells with *Spi1* gRNA vs NT control cells.

f) Enrichment analysis for the for the MYC gene expression signature in the single-cell dataset (upper panel) and the bulk dataset (lower panel) of *Spi1* perturbation. Indicated on the graphs are the NES and FDR false discovery rate q-value (FDR q).

g) Heat maps showing expression levels of genes within SCENIC regulons, in cells with *Spi1* gRNA vs NT control cells. Small gene regulatory networks were determined based on based on co-expression patterns and transcription factor motifs.
Figure 4: Perturbations can be clustered based on their gene expression profiles

a) Heatmap of pro-T gene signature clusters NOTCH, STAT, E2F and MYC. The data of both CRISPR screens were merged and the NES scores were calculated for each perturbation. Unsupervised hierarchical clustering was used to cluster perturbations based on similarity of the NES scores. For the blank boxes, no NES score could be calculated.

b) Volcano plots of the gene sets E2F, MYC, NOTCH and STAT for cells with perturbation of Bcl11b, Ctcf, Sp11 and Arid1a. Dotted lines are thresholds for significant (padj <0.05) differentially expressed ($|\log2Fc|>0.5$) genes.

c) Evolution of the fraction of cells per perturbation over time at day 3, 7 and 14 of screen 2. Y-axis shows data in logarithmic scale. NT controls are indicated in grey font. Perturbations are ordered based on the number of cells at day 14.

d) Cell cycle phase distribution per perturbation in screen 2. Based on normalized expression levels of cell cycle marker genes, cells were classified in S, G2M or G1 phase and perturbations were ordered based on the fraction of cells in G1 phase.

e) Average CRISPR scores from the DepMap database for all genes in lymphoid leukemia and lymphoma cell lines. Negative scores mean negative impact on cell proliferation, while genes with positive scores cause a proliferative advantage. Genes on x-axis were ordered by increasing gene effect score.

f) Enrichment of SCENIC regulons per perturbation. Stacked violin plot displaying the distribution of AUCell values for each regulon and each perturbation. AUCell values represent enrichment of each predicted regulon in each single cell. The color intensity is a measure for the median expression of each regulon in each perturbation. This figure shows a selection of perturbations and regulons, with complete data displayed in supplementary figure S6.

Figure 5: Bcl11b inactivation drives cell proliferation and upregulation of NF-kB and STAT signaling.

a) Cell densities of BFP-positive cells (mean +/- SD, n = 3 replicates) over time for pro-T cells with Bcl11b gRNAs, compared with two NT control gRNAs. Data were analyzed by one-way ANOVA at day 22. P values were indicated with **p < .01, *p < .05 or not significant (ns).

\begin{align*}
\text{NT1}_{-}\text{gRNA1} \text{vs } \text{NT1}_{-}\text{gRNA5}: p=0.9993, \text{NT1}_{-}\text{gRNA1} \text{vs Bcl11b}_{-}\text{gRNA1}: p=0.0037, \text{NT1}_{-}\text{gRNA1} \text{vs Bcl11b}_{-}\text{gRNA5}: p=0.0367, \text{NT1}_{-}\text{gRNA5} \text{vs Bcl11b}_{-}\text{gRNA1}: p=0.0043, \text{NT1}_{-}\text{gRNA5} \text{vs Bcl11b}_{-}\text{gRNA5}: p=0.0434, \text{Bcl11b}_{-}\text{gRNA1} \text{vs Bcl11b}_{-}\text{gRNA5}: p=0.3575)
\end{align*}

b) Protein levels of BCL11B as measured by intracellular flow cytometry. Normalized median fluorescence intensities (MFI) were calculated for AlexFluor555 emission. Three biological replicates were performed, with mean and SD shown. Comparisons between groups were performed by two-sided unpaired Student t-test, p<0.0001.
c) Cell densities of GFP-positive cells (mean +/- SD, n = 3 replicates) over time for pro-T cells with Bcl11b overexpression (OE) and empty vector (EV). Data were compared at day 10 by two-sided unpaired Student t-test, p=0.0004.

d) Overlap between significant (padj <0.05) differentially expressed genes in the bulk and single-cell data for cells with Bcl11b inactivation compared to NT cells. P-values for the overlaps were calculated using the hypergeometric probability distribution.

e) Overlap between ChIP-seq peaks obtained from a publicly available dataset of BCL11B binding regions, and the significant differentially expressed genes in the bulk RNA-seq data. P-values for the overlaps were calculated using the hypergeometric probability distribution.

f) Gene set enrichment analysis on the set of direct Bcl11b target genes. NES scores and p-values are shown for the Hallmark gene sets.

g) Predicted cis-regulatory binding motifs in the significant (padj <0.05) up- and downregulated genes in cells with Bcl11b inactivation.

h) Heat maps showing expression levels of genes within SCENIC regulons in cells with Bcl11b gRNA vs NT control cells.

Figure 6: Bcl11b levels are correlated with STAT pathway activation and cause transformation of pro-T cells.

a) Normalized expression level of BCL11B in publicly available RNA-seq data of T-ALL patients. Patients were subdivided in three groups based on expression level of BCL11B (i.e. average – high – low).

b) Gene set enrichment analysis comparing patients with low BCL11B expression with those with high expression level. NES scores and p-values are shown for the Hallmark gene sets.

c) Normalized expression level of BCL11B in publicly available RNA-seq data of T-ALL patients with alterations in the JAK/STAT signaling pathway compared to patients without mutations in this pathway. The p-value was calculated with an unpaired Student t-test.

d) Cell densities (mean +/- SD, n = 3 replicates) of pro-T cells over time with the following conditions: JAK3M511I + NT, JAK3M511I + gBcl11b, EV + gBcl11b and EV + NT. Left panel shows cells cultured in presence of IL7 and right panel shows IL7-deprived cells. Data were analyzed by one-way ANOVA at day 14 (left panel) or day 7 (right panel). P values were indicated with ****p < .0001, ***p < .001, **p < .01, *p < .05 or not significant (ns). (Left panel: JAK3M511I + NT vs JAK3M511I + Bcl11b gRNA: p=0.0063, JAK3M511I + NT vs EV + NT: p=0.9981, JAK3M511I + NT vs EV + Bcl11b gRNA: p=0.0035, JAK3M511I + Bcl11b gRNA vs EV + NT: p=0.0078, JAK3M511I + Bcl11b gRNA vs EV + Bcl11b gRNA: p=0.9634, EV + NT vs EV + Bcl11b gRNA: p=0.0043). (Right panel: JAK3M511I + NT vs JAK3M511I + Bcl11b...
gRNA: p<0.0001, JAK3<sup>M511I</sup> + Bcl11b gRNA vs EV + NT: p<0.0001, JAK3<sup>M511I</sup> + Bcl11b gRNA vs EV + Bcl11bgRNA: p<0.0001).

e) Dose-response curve for pro-T cells with JAK3<sup>M511I</sup> and Bcl11b gRNA treated with ruxolitinib phosphate for 60h. Relative viability was calculated by normalizing the data based on the DMSO condition of each group.
a.

b.

b.

b.

b.

b.

f.

f.

f.

f.

f.
SUPPLEMENTARY DATA

Single-cell CRISPR screening characterizes transcriptional deregulation in T-cell acute lymphoblastic leukemia

Sarah Meyers$^{1,2,3}$, Olga Gielen$^{1,2,3}$, Jan Cools$^{1,2,3,*}$, Sofie Demeyer$^{1,2,3,*}$

$^1$ Center for Human Genetics, KU Leuven, Leuven, Belgium.
$^2$ Center for Cancer Biology, VIB, Leuven, Belgium.
SUPPLEMENTARY METHODS

Cell culture
Ba/F3 pro-B cells (DSMZ) were cultured in RPMI-1640 medium supplemented with 10% fetal bovine serum (FBS) and interleukin 3 (10ng/mL, PeproTech).

gRNA library design
gRNAs were designed using the CRISPick tool\(^1\), with compatible BpiI/BbsI overhangs at the 5’ end for cloning in a retroviral vector containing a 10x Genomics-compatible capture sequence (CS) for gRNA capture (CS1: GCTTTAAGGCGGTCCTAGCAA, CS2: GCTCACCTATTAGCGGCTAAGG).\(^2\) 100 μM of forward and reverse gRNA oligos were annealed under following conditions; 95°C for 3 minutes, 25°C for 1 minute and holding at 4°C. Annealed gRNAs were pooled, diluted to 0.04 μM and ligated into the BpiI-digested vector backbone. This construct was electroporated in Endura electrocompetent cells (Lucigen) according to the manufacturer’s protocols. A minimum of 1000 bacterial colonies per gRNA in the library were harvested and DNA was isolated using the Compactprep Plasmid Midi kit (Qiagen) according to manufacturer’s directions.

Retroviral work
Retrovirus production was performed by transfection of 150 mm dishes HEK293T cells with 10 μg of the retroviral vector of interest and 10 μg packaging plasmid pK6.1MCV.ecopac.UTD, using 36 μL Genejuice transfection reagent (Merck Millipore #70767) in 600 μL serum-free RPMI-1640 medium. After 24h, medium was replaced with fresh RPMI-1640 supplemented with 10% FBS and retroviral supernatant was subsequently harvested at 48h and 72h post-transfection. Viral supernatant was filtered through a 0.45 μm filter and aliquots were frozen at -80°C. For CRISPR libraries, viral titer was calculated to determine the optimal transduction conditions to reach around 10% efficiency to avoid introduction of multiple gRNAs in a single cell. Serial dilutions of the virus stock were used to transduce 1 million cells by spinfection and transduction efficiencies were measured by flow cytometry (MACSQuant VYB, Miltenyi). The number of transducing units per mL was determined and used to calculate the optimal volume of virus stock that should be used to reach the optimal MOI.

shRNA
shRNAs were designed using the Vectorbuilder shRNA design tool (https://en.vectorbuilder.com/tool/shrna-target-design.html). Single-stranded oligonucleotides were annealed under following conditions; 95°C for 3 minutes, 25°C for 1 minute and holding at 4°C. Annealed oligos were subsequently cloned by ligation into EcoRI and XhoI sites in an MSCV-mTagBFP
vector containing a miR30 flanking cassette. A non-targeting shRNA was used as control (shCtrl: TCTCGCTTGGCGAGAGTAAGC). The other sequences used were: shBcl11b: CAGGAGAACATTGCAGGTAAA, shCtcf: GGCTTTGGGAACGGCATAATT and shSmarca4: TCGAGTCTCTACCAGCATTA. Retroviral packaging and cell transduction were performed as described earlier, and cell proliferation was monitored over time using a MACSQuant VYB (Miltenyi).

Data analysis of single-cell CRISPR experiments

CellRanger software (version 5.0.1) was used for alignment of reads to the mouse reference genome (mm10) and unique molecular identifier (UMI) counting. The resulting count matrices were further processed using the Seurat R package v4. We removed cells that expressed less than 200 genes and genes that were expressed in less than 3 cells. Furthermore, cells with high numbers of mitochondrial reads (>15%) and extreme numbers of transcripts (<1500 and >8000) were removed, as these are most likely damaged cells or doublets. Next, the data was normalized using SCTransform. Cell cycle scoring was done with Seurat based on the lists of cell cycle marker genes of Tirosh et al. Normalized gene expression levels in all single cells were displayed in ridge plots per perturbation, or via a dot plot. Differential gene expression analysis using the MAST algorithm was performed to identify differentially expressed marker genes between cells with a gRNA and NT cells. Furthermore, this data was analyzed with a number of software packets specifically designed for single cell CRISPR screening, such as Mixscape, scMAGeCK (both in RRA and LR mode), SCEPTRE and MUSIC. In an alternative approach, cells with the same perturbation were grouped and a global transcriptional profile per perturbation was determined. DESeq2 (version 1.34.0) was used to compare these pseudo-bulk transcription profiles and identify differentially expressed genes.

As we were mostly interested in the genes of importance in the pro-T cell model, we focused on the five sets identified by Bornschein et al., with which we performed gene set enrichment analyses (GSEA) on the ranked list of differentially expressed genes between the cells with a specific perturbation and the non-perturbed cells. The rank was calculated as: -sign(log2Fc)*log(padj). Based on these normalized enrichment scores (NES), an unsupervised hierarchical clustering algorithm was applied to identify perturbations with similar impact on the cells. To determine the regulatory elements involved in each of the perturbations, the pySCENIC software package (version 0.10.0), together with its command line interface (CLI), was used. First, a list of adjacencies between transcription factors and their targets was determined with the CLI version on the Flemish Super Computer (VSC). Subsequently, we used the publicly available motif database to predict regulons, after which cellular enrichment for each regulon in each cell was calculated with AUCell. The
distributions of these AUCell values for each regulon and each perturbation were then plotted with the stacked violin function in the Scanpy package.

**Bulk RNA-sequencing**

The Maxwell Simply RNA purification kit (Promega) was used to extract RNA from sorted pro-T cells, from which 500 ng of purified RNA was used to construct NGS libraries using the Lexogen Quantseq RNA sample prep kit (Illumina). Libraries were subjected to 51 bp single-end sequencing on a HiSeq4000 (Illumina). The reads were cleaned with the fastq-mcf software package (version 1.04.807) and FastQC (version 0.11.9) was used for quality control. The reads were subsequently aligned to the mouse reference genome (mm10/GRCm38) using HISAT2 (version 2.1.0) and further processed with the SAMtools software (version 1.11). The number of reads per gene was counted with the HTSeq package. The R (version 4.1.1) package DESeq2 (version 1.34.0) was used for differential gene expression analysis and the SCENIC software package (version 1.3.1) was used to determine regulons, i.e., small gene regulatory networks based on co-expression patterns and transcription factor motifs.

**Intracellular staining**

Intracellular stainings were performed using the FOXP3 Transcription Factor Staining buffer set (Invitrogen 00-5523-00). Briefly, 100 000 cells were seeded in a well of a 96-well plate, stained with fixable viability dye, washed and resuspended in fixation/permeabilization buffer for fixation for 45 minutes. Cells were subsequently stained with primary and secondary antibody in triplicate, and data were acquired on a MACSQuant VYB (Miltenyi). Primary antibodies used were anti-BCL11B (Abcam ab18465, 1/200 dilution) and anti-PU.1 (Abcam ab227835, 1/100 dilution), with secondary antibodies Goat anti-Rat IgG Secondary Antibody in Alexa Fluor 555 (Thermofisher A21434, 1/1000 dilution) and Donkey anti-Rabbit IgG Secondary Antibody in Alexa Fluor 568 (Thermofisher A10042, 1/1000 dilution). Fixable viability dyes eFluor520 (eBioscience 65-0867-14, 1/1000 dilution) and eFluor780 (eBioscience 65-0865-14, 1/1000 dilution) were used to discriminate living cells.

**Western blot**

Cell lysates were prepared in cold Cell Lysis Buffer (Cell Signaling) containing protease inhibitor (Complete – EDTA-free, Roche) and 5 mM Na3VO4. Protein concentration was determined with Protein Assay Dye Reagent Concentrate (Bio-Rad) using the Victor X4 plate reader (Perkin Elmer) according to the manufacturer’s directions. Proteins were separated on NuPAGE NOVEX Bis-Tris 4 - 12% gels (Life Technologies) and transferred to PVDF membranes. Western blot analysis was performed using primary antibodies targeting P65 (Cell Signaling 8242S, 1/1000 dilution) and phospho-P65 (Cell
Signaling 3033S, 1/1000 dilution), with anti-β-actin as loading control (Sigma-Aldrich A5441, 1/1000 dilution). Western blot detection was performed with secondary antibodies conjugated with horseradish peroxidase (GE Healthcare). Images were acquired using the VILBER FUSION FX6 imager (Analyt), bands were quantified using ImageJ software and normalized to the intensity of the loading control β-actin.

REFERENCES


SUPPLEMENTARY FIGURE LEGENDS

Fig. S1. Experimental details of the pilot CRISPR screen.

a. Vector graphic of the retroviral gRNA vector with capture sequence (CS) in the hairpin of the gRNA scaffold. This capture sequence is required for gRNA capture by the 10X Genomics 3’ feature barcoding chemistry for single-cell CRISPR screening. Expression of the gRNA is controlled by U6 promoter and expression of mTagBFP fluorescent marker by the spleen focus-forming virus (SFFV) promoter sequence. (*LTR = long terminal repeat, AmpR = ampicillin resistance marker*)

b. CD45 protein levels were measured by flow cytometry in Ba/F3 and pro-T cells expressing a Ptprc-targeting gRNA to validate the efficiency of four different capture sequence (CS) vector configurations, compared to the original vector without capture sequence and compared to the empty vector (EV).

c. Gating strategy for FACS sorting of pro-T cells transduced with the gRNA library. First, cells were discriminated from debris based on forward (FSC) versus side (SSC) scatter (P1), with subsequent gating for singlets (P2 and P3) and viable cells based on eFluor780 fixable viability dye (P4). Finally, cells with expression of the BFP fluorescent marker were selected (P5).

d. Stacked bar plot displaying the number of cells per gRNA per target gene at day 3 of the experiment. Non-targeting (NT) controls are indicated in grey font.

e. Ridge plots of normalized expression levels of perturbed genes and their downstream targets. In the left column, expression values are shown for Notch1, Myc and Il7r, with the purple ridge highlighting cells with the respective perturbation. In the right column, the expression levels of two direct downstream target genes are shown.

Fig. S2. Gene expression levels per perturbation.
Data from both single-cell CRISPR screens were merged, and the average expression level of each target gene was displayed in a dot plot to visualize perturbation efficiency in the cells with different gRNAs.

Fig. S3. SCENIC analysis of bulk RNA-seq after Spi1 perturbation.

a. Volcano plot of differentially expressed genes based on bulk RNA-sequencing of cells with Spi1 gRNA, compared to NT control cells. Lines are thresholds for significant (padj <0.05) upregulated and downregulated (|log2Fc|>1) genes, with indication of gene names for the top significant differentially expressed genes.

b. SCENIC heatmaps for Myc regulons in cells with Spi1 inactivation vs NT control cells.
c. SCENIC heatmaps for Cebpb regulon in cells with Spi1 inactivation vs NT control cells.
d. SCENIC heatmaps for epigenetic regulons in cells with Spi1 inactivation vs NT control cells.
e. SCENIC heatmaps for Ets factor regulons in cells with Spi1 inactivation vs NT control cells.

Fig. S4. Experimental details of the second CRISPR screen.

a. Pie charts show percentages of cells with zero, one or multiple gRNAs in the second CRISPR screen, with indication of the total number of cells passing filtering parameters at each timepoint.
b. Number of cells per gene and per gRNA. Box plots showing the number of sequenced cells passing filter criteria per gene (left panel) and per gRNA (right panel) in CRISPR screen 2. Plots show a line at the median and box from Q1 to Q3, with whiskers ranging from minimum to maximum values.
c. Violin plots displaying the number of detected transcripts in all single cells for each timepoint.
d. Stacked bar plot displaying the number of cells per gRNA per target gene at day 3 of the experiment. NT controls are indicated in grey font.
e. Normalized expression levels of the target genes in cells with the respective gRNA. Similar to the first single-cell CRISPR screen, all cells of the three timepoints were aggregated per perturbation to create ridge plots of the expression level of each target gene. On the top of the graph is indicated for which transcript the expression level is displayed, with the purple ridge highlighting the corresponding perturbation.
f. Volcano plots of the E2F, MYC, NOTCH and STAT signatures for cells with perturbation of Notch1 (left) and Myc (right). Dotted lines are thresholds for significant (padj <0.05) differentially expressed (|log2Fc|>0.5) genes.
g. Proliferation effects of Dnm2 (upper panel), Med12 (middle panel) and Ikzf1 (lower panel) gRNAs. To confirm the effects observed in the CRISPR screens, individual proliferation assays were performed using two out of the five gRNAs per target gene and compared to two NT gRNAs. The cell density as number of BFP-positive cells (mean +/- SD, n = 3 replicates) was monitored over time. Data were analyzed by one-way ANOVA at day 22. P values were indicated with ***p < .001, **p < .01, *p < .05 or not significant (ns).
(Upper panel: NT1_gRNA1 vs NT_gRNA5: p=0.8371, NT1_gRNA1 vs Dnm2_gRNA1: p=0.0018, NT1_gRNA1 vs Dnm2_gRNA4: p=0.0018, NT1_gRNA5 vs Dnm2_gRNA1: p=0.0007, NT1_gRNA5 vs Dnm2_gRNA4: p=0.0007, Dnm2_gRNA1 vs Dnm2_gRNA4: p>0.9999)
(Middle panel: NT1_gRNA1 vs NT1_gRNA5: p=0.8377, NT1_gRNA1 vs Med12_gRNA2: p=0.0037, NT1_gRNA1 vs Med12_gRNA4: p=0.0019, NT1_gRNA5 vs Med12_gRNA2: p=0.0019, Med12_gRNA1 vs Med12_gRNA4: p=0.0019, Med12_gRNA1 vs Med12_gRNA4: p=0.0019, Med12_gRNA2 vs Med12_gRNA4: p=0.0019)
p=0.0014, NT1_gRNA5 vs Med12_gRNA4: p=0.0008, Med12_gRNA2 vs Med12_gRNA4: p=0.9446)
(Lower panel: NT1_gRNA1 vs NT1_gRNA5: p=0.9165, NT1_gRNA1 vs Ikzf1_gRNA3: p=0.2853, NT1_gRNA1 vs Ikzf1_gRNA5: p=0.0143, NT1_gRNA5 vs Ikzf1_gRNA3: p=0.1204, NT1_gRNA5 vs Ikzf1_gRNA5: p=0.0063, Ikzf1_gRNA3 vs Ikzf1_gRNA5: p=0.2104)

h. Normalized expression level of Ets1. All cells of both CRISPR screens were aggregated per perturbation to create ridge plots of the normalized expression level of Ets1. The purple ridge highlights cells with Spi1 gRNA.

Fig. S5. Violin plots displaying CRISPR scores from the DepMap database for all perturbations in all lymphoid cell lines.

Negative scores mean negative impact on cell proliferation, while genes with positive scores cause a proliferative advantage. Genes on x-axis were ordered based on proliferative effects in pro-T cells.

Fig. S6. Enrichment of SCENIC regulons per perturbation.

Stacked violin plot displaying the distribution of AUCell values for each regulon and each perturbation. AUCell values represent enrichment of each predicted regulon in each single cell. The color intensity is a measure for the median expression of each regulon in each perturbation.

Fig. S7. Bcl11b inactivation results in increased cell growth and upregulation of NF-kB.

a. Cell densities of BFP-positive cells (mean +/- SD, n = 3 replicates) over time for pro-T cells with shRNAs targeting Bcl11b (left, p=0.0028), Ctf1 (middle, p=0.0039) and Smarca4 (right, p=0.0002), compared with a negative control shRNA (shCtrl). Data were compared at day 18 or 17 by two-sided unpaired Student t-test. P values were indicated with ***P < .001, **P < .01

b. Volcano plot of differentially expressed genes based on bulk RNA-sequencing of cells with Bcl11b gRNA, compared to NT control cells. Lines are thresholds for significant (padj <0.05) upregulated and downregulated (|log2Fc|>1) genes, with indication of gene names for the top significant differentially expressed genes.

c. Western blot of total P65 and phospho-P65 levels in cells with Bcl11b gRNA vs NT control cells (left panel) and quantification of their relative intensities normalized to loading control (right panel).
Fig. S8. Characteristics of T-ALL patient samples ranked according to BCL11B expression level.

Patient samples were ranked based on expression level of BCL11B and colored by ETP status (left), maturation stage (middle) or transcription factor (TF) subgroup (right). The top 20 patients with highest and lowest expression of BCL11B were selected for further analyses. Significant enrichment was observed of ETP (NES -2.03, FDR q<0.001) and pre-cortical (NES 2.33, FDR q<0.001) phenotypes in patients with lower BCL11B expression. (ETP = early T-cell precursor)
ONLINE SUPPLEMENTARY TABLES PROVIDED AS EXCEL FILES

Online Supplementary Table S1. List of gRNA sequences used in the first CRISPR screen.

Online Supplementary Table S2. Differentially expressed genes in bulk RNA-seq data from cells with *Spi1* or *Bcl11b* inactivation, compared to NT cells.

Online Supplementary Table S3. List of gRNA sequences used in the second CRISPR screen.
Figure S3.

a. Log2FC Spi1 vs NT

b. Mybl2 regulon
Mycn regulon

C. Cebpb regulon

D. Kdm4b regulon
Smarca4 regulon
Kat2a ext-regulon
Ctf ext-regulon

Expression: low high

E. Fli1 ext-regulon
Ets2 regulon
Figure S8.

BCL11B

ETP Status
- ETP
- near ETP
- not ETP
- Unevaluable
- N/A

BCL11B

Maturation stage
- Pre-cortical
- Cortical
- Post-cortical
- No pretreatment
- N/A

BCL11B

TF subgroup
- HOXA
- LMO1/2
- LMO2_LYL1
- TLX1
- TLX3
- NKX2_1
- TAL1
- TAL2
- Unknown

T-ALL patient samples

Normalized counts

Normalized counts

Normalized counts