Oncogenic cooperation between TCF7-SPI1 and NRAS(G12D) requires β-catenin activity to drive T-cell acute lymphoblastic leukemia

Quentin Van Thillo1,2,3,17, Jolien De Bie1,2,4,17, Janith A. Seneviratne5,6, Sofie Demeyer1,2,3, Sofia Omari5,6, Anushree Balachandran5,6, Vicki Zhai5,6, Wai L. Tam7, Bram Sweron1,2, Ellen Geerdens1,2,3, Olga Gielen1,2, Sarah Provost1,2, Heidi Segers3,8,9, Nancy Boeckx8,10, Glenn M. Marshall5,6,11, Belamy B. Cheung5,6, Kiyotaka Isebe12, Itaru Kato12, Junko Takita12, Timothy G. Amos13, Ira W. Deveson13,14, Hannah McCalmont5,6, Richard B. Lock5,6, Ethan P. Oxley15, Maximilian M. Garwood15, Ross A. Dickins15, Anne Uyttebroeck3,8,9, Daniel R. Carter5,6,16, Jan Cools1,2,3✉ & Charles E. de Bock5,6✉

Spi-1 Proto-Oncogene (SPI1) fusion genes are recurrently found in T-cell acute lymphoblastic leukemia (T-ALL) cases but are insufficient to drive leukemogenesis. Here we show that SPI1 fusions in combination with activating NRAS mutations drive an immature T-ALL in vivo using a conditional bone marrow transplant mouse model. Addition of the oncogenic fusion to the NRAS mutation also results in a higher leukemic stem cell frequency. Mechanistically, genetic deletion of the β-catenin binding domain within Transcription factor 7 (TCF7)-SPI1 or use of a TCF/β-catenin interaction antagonist abolishes the oncogenic activity of the fusion. Targeting the TCF7-SPI1 fusion in vivo with a doxycycline-inducible knockdown results in increased differentiation. Moreover, both pharmacological and genetic inhibition lead to down-regulation of SPI1 targets. Together, our results reveal an example where TCF7-SPI1 leukemia is vulnerable to pharmacological targeting of the TCF/β-catenin interaction.

1Department of Human Genetics, KU Leuven, Leuven, Belgium. 2Center for Cancer Biology, VIB, Leuven, Belgium. 3Leuvense Kanker Instituut (LKI), KU Leuven – UZ Leuven, Leuven, Belgium. 4Center for Human Genetics, UZ Leuven, Leuven, Belgium. 5Children’s Cancer Institute, UNSW Sydney, Lowy Cancer Research Centre, Sydney, NSW, Australia. 6School of Women’s and Children’s Health, UNSW Sydney, Sydney, NSW, Australia. 7Technology Innovation Lab, VIB, Gent, Belgium. 8Department of Oncology, KU Leuven, Leuven, Belgium. 9Department of Pediatric Hemato-Oncology, UZ Leuven, Leuven, Belgium. 10Department of Laboratory Medicine, UZ Leuven, Leuven, Belgium. 11Kids Cancer Centre, Sydney Childrens’ Hospital, Randwick, NSW, Australia. 12Department of Pediatrics, Graduate School of Medicine, Kyoto University, Kyoto, Japan. 13Kinghorn Centre for Clinical Genomics, Garvan Institute of Medical Research, Sydney, NSW, Australia. 14St Vincent’s Clinical School, Faculty of Medicine, UNSW Sydney, Sydney, NSW, Australia. 15Australian Centre for Blood Diseases, Monash University, Melbourne, VIC, Australia. 16School of Biomedical Engineering, University of Technology, Sydney, NSW, Australia. 17These authors contributed equally: Quentin Van Thillo, Jolien De Bie. ✉✉email: jan.cools@kuleuven.be; cdebock@ccia.org.au
cell acute lymphoblastic leukemia or T-ALL is a hematopoietic malignancy of T-cell precursor cells that primarily affects young children. Chromosomal rearrangements are present in the majority of T-ALL patients, resulting in the ectopic expression of transcription factors, including T-cell acute lymphocytic leukemia protein 1 (TAL1), T-cell leukemia homeobox protein 1 and 3 (TLX1 and TLX3) and NK2 homeobox 1 (NKX2-1). Other typical pathways involved in the pathogenesis of T-ALL include the NOTCH1, Janus kinase (JAK)/signal transducer of activation and transcription (STAT) and RAS signaling pathways. Often multiple cooperating hits are needed to induce leukemia.

Seki et al. recently reported that up to 4% of pediatric T-ALL patients carry gene fusions involving the SPI1 locus on chromosome 11 (also known as PU.1) with different fusion partners, and that patients harboring such SPI1 fusions represent a distinct subgroup of T-ALL cases with significantly shorter overall survival. While SPI1 (PU.1) is an important transcription factor in all hematopoietic cells, TCF7 (located on the short arm of chromosome 5) encodes TCF1, a T-cell specific transcriptional activator that is highly expressed in the earliest thymic progenitors and controls chromatin accessibility of T-cell regulatory elements. Dysregulation of PU.1 contributes to several types of leukemia and forced overexpression of SPI1 in T-cell progenitors is known to inhibit T-cell differentiation. Conversely, TCF7 knock-out also leads to a T-cell differentiation block.

In normal T-cell development, NOTCH-signaling plays an important role in downregulating PU.1 and upregulating TCF1 expression during progression from the double negative (DN) 2 to DN3 T-cell stage. To evaluate the role of SPI1 fusion genes in T-ALL, Seki et al. expressed distinct fusion transcripts in mouse stem/progenitor cells and documented an increased T-cell proliferation and a maturation block at the DN3 stage upon SPI1 or PI1 fusion expression. Constitutive expression of TCF7-SPI1 alone, however, was not found to cause leukemia development in a mouse bone marrow transplant model. Significantly, whilst TCF7-SPI1 fusions are insufficient for frank T-cell leukemia development, these clinical cases are often coincident with NRAS mutations suggesting the potential for oncogenic cooperation. Here we show that TCF7-SPI1 fusions and the NRAS (G12D) activating mutation cooperate to form an immature T-ALL in vivo that is absent when each is expressed alone. Moreover, using a genetic approach we show this cooperation absolutely requires the presence of the β-catenin interaction domain in the TCF7-SPI1 fusion with genetic loss of the fusion resulting in immunophenotypic differentiation. We also provide proof-of-principle that this is phenocopied through small molecule antagonism of the interaction between β-catenin and TCF1 of the TCF7-SPI1 fusion.

Results

Identification of a TCF7-SPI1 positive and NRAS mutant T-ALL case. Peripheral blood cells from patient X09, a 6-year-old boy newly diagnosed with an immature CD4/CD8 double negative T-ALL, underwent whole-genome and RNA sequencing, revealing a translocation between chromosomes 5 and 11 and co-occurring mutations in NRAS, NOTCH1, Krüppel-like factor 9 (KLF9), Cyclin-dependent kinase inhibitor 2A (CDKN2A), Sidekick cell adhesion molecule 1 (SDK1), T-cell receptor gamma (TRG), and transducin (β-like)-like 1 X-linked receptor 1 (TBL1XR1) (Fig. 1a). The t(5;11) translocation breakpoint occurred at TCF7 intron 4–5 (Chr. 5) and SPI1 intron 2–3 (Chr. 11) generating a predicted fusion transcript connecting exons 1–4 from TCF7 to exons 3–5 of SPI1 that was validated using both long-read nanopore and Sanger sequencing (Fig. 1b, c). At the protein level, the predicted fusion product brings together the first 182 amino acids from TCF1 (TCF7), including the N-terminal β-catenin binding region, and the terminal amino acids 52 to 271 of PU.1 (SPI1), resulting in a 402 amino acid long protein that still retains the DNA binding domain regions of PU.1 (Fig. 1d).

Single-cell RNA sequencing identifies that high SPI1 expression is associated with NRAS, stem cell, and Wnt/β-catenin signaling cell signatures. Single-cell RNA sequencing was carried out within six primary T-ALL samples (Supplementary Data 1–4) including patient X09 to identify transcriptomic determinants of this T-ALL subtype. Clustering a total of 13,848 cells from the pooled samples identified 22 cell clusters representing either malignant T cells or normal cells including NK-T cells, B cells, monocytes, and red blood cells (Supplementary Fig. 1a–c). Focusing on the 11,620 malignant T cells only, patients with TCF7-SPI1 fusions were distinct from the other T-ALL samples and displayed significantly higher SPI1 expression (Fig. 2a–c, Supplementary Fig. 1d). Since SPI1 is the 3′ partner of the TCF7-SPI1 fusion, and 10× short read RNA-sequencing is biased, we confirmed that SPI1 expression levels were correlated with TCF7-SPI1 fusion expression through long read nanopore sequencing (Supplementary Fig. 1e–h, Supplementary Data 5). Gene expression analysis also showed that SPI1 expression in fusion positive cases correlated with higher levels of hematopoietic stem markers like CD34 and lower levels of mature T-cell markers like CD3D and CD1E (Fig. 2d–g) reconciling with the immature phenotype by flow cytometry analysis.

We next sought to determine how fusion positive cases compared to other defined T-ALL subtypes at the transcriptome level. We found fusion cases clustered in close proximity to the early T-cell precursor (ETP) and immature Lin domain only 2/ Lymphoblastic Leukemia Associated Hematopoiesis Regulator 1 (LMO2/LYL1) cases of the Liu et al. dataset (Supplementary Fig. 2a, b), suggesting fusion cases have a transcriptome similar to ETP-ALL despite their immunophenotype. Therefore, using published transcriptome data, we next generated gene signatures from ETP- and non-ETP-ALL patients (Supplementary Data 6, 7). The resulting Liu and Zhang ETP-ALL signatures were validated using single-cell RNA-sequencing data of normal developing human thymocytes, confirming highest enrichment in cells within the designated ETP region. Conversely, the Liu nonETP-ALL and Zhang T-ALL signatures had the highest enrichment in T cells designated at double positive (DP) and qPT-entry regions (Supplementary Fig. 2c–k). Applying these signatures to our data, the three TCF7-SPI1 fusion positive cases were positively enriched for an ETP-ALL signature and negatively enriched for the complementary T-ALL signature (Fig. 2h, i, Supplementary Fig. 3a–f).

Combining additional bulk RNA-sequencing data of these 6 patients with the cohort described by Seki et al. further revealed that patients harboring both the TCF7-SPI1 fusion and NRAS mutations had a higher ETP-ALL signature compared to those with the TCF7-SPI1 fusion alone (Supplementary Fig. 4). We also used GSEA analysis to determine signaling pathways associated with SPI1 expression, which revealed a positive enrichment for myeloid, stem, and Wnt/β-catenin gene signatures (Fig. 2i, Supplementary Data 8). Collectively, these data highlight the fusion positive transcriptomes are enriched for an ETP signature and there is a strong link between TCF7-SPI1 fusion status, RAS activation and the Wnt/β-catenin pathway.

TCF7-SPI1 requires mutant NRAS to cause in vitro transformation. A recent study identified recurrent translocations similar to our TCF7-SPI1 fusion and these often harbored mutations in either NRAS (3/7 cases) or KRAS (1/7 cases). This same study
showed that expression of SPI1 fusions using an in vivo bone marrow transplant model led to a block in T-cell differentiation but not acute leukemia. Given the high frequency of co-occurring mutations with NRAS (OR 5.833, p value 0.0067, Supplementary Data 9) and association with NRAS signaling (Fig. 2j) we sought to investigate whether the TCF7-SPI1 fusion cooperates with oncogenic NRAS to drive T-ALL. We generated two sets of retroviral vectors that would express TCF7-SPI1 and/or NRAS(G12D) from a single vector using a viral P2A sequence, in either a constitutive or inducible manner as has been previously described (Fig. 3a, b). Western blot analysis confirmed efficient constitutive or inducible expression of both the TCF7-SPI1 (TCF1-PU.1) fusion and/or NRAS(G12D) upon transduction of Ba/F3 cells and Ba/F3-Cre cells respectively (Fig. 3c, d). The transformation potential was first determined in wild-type and Cre-positive Ba/F3 cells expressing NRAS(G12D) alone or in combination with TCF7-SPI1. Only when NRAS(G12D) was present, were the cells able to grow in an interleukin-3 (IL3)-independent manner, demonstrating that the NRAS(G12D) protein provides the essential proliferation and survival signals to lymphoid cells (Fig. 3e, f). Using a more physiological relevant Cre-positive pro-T cell culture model in combination with our inducible vector system, all transduced cells grew in the presence of Delta-like ligand 4 (Dll4), interleukin-7 (IL7) and stem cell factor (Scf). Notably, inducible expression of NRAS(G12D) + /− TCF7-SPI1 provided growth and survival signals that could transform cells to Scf- and to a lesser extent IL7-independent growth, but all remained dependent on Notch signaling via Dll4 (Fig. 3g–j). Consequently, these results show that NRAS(G12D) alone is sufficient and necessary for transformation of these two in vitro cellular systems to cytokine-independent growth.
TCF7-SPI1 and NRAS(G12D) expression leads to aggressive T-cell leukemia in a mouse bone marrow transplant model. We next sought to determine whether the combination of TCF7-SPI1 (TCF1-PU.1) and NRAS(G12D) cooperates in vivo to drive T-ALL. To this end, HSPCs isolated from the bone marrow of CD2-Cre positive C57BL/6J mice were retrovirally transduced with the different lox66/71 vectors (Fig. 3b–d) to restrict expression of TCF7-SPI1 with or without NRAS(G12D) within developing T cells (Fig. 4a). Wild-type recipient mice that received TCF7-SPI1 only transduced cells did not develop leukemia within the 200 days post injection observation period. Recipient mice that received TCF7-SPI1+NRAS(G12D) developed leukemia (median disease-free survival = 66 days) with a similar latency to those that received NRAS(G12D) alone (median disease-free survival =
Fig. 2 Single-cell RNA sequencing identifies that high SPI1 expression is associated with NRAS, stem cell, and Wnt/β-catenin signaling cell signatures. a–c UMAP plot of 11,620 malignant T-cells from 6 T-ALL samples with cells colored by their patient of origin (a), TCF7-SPI1 and TCF7-SPI1 + NRAS mutation status (b), level of SPI1 expression (c), the scalebar represents normalized expression values. d Heat map for all single cells across 6 T-ALL patients highlighting top 25 and bottom 25 genes correlated with SPI1 expression. e–g UMAP plot for expression levels of CD3D (e) CD1E (f) and, CD34 (g) highlighting differential expression between fusion-positive and fusion-negative T-ALL cases, h, i UMAP plot of cells colored by the average expression of the ETP-ALL (h) or nonETP-ALL (i) signature derived from Liu et al.5, the scalebar represents averaged and scaled gene signature expression values. j Associated GSEA analysis pathways associated with SPI1 expression highlighting enrichment of stem, Wnt/β-catenin, and myeloid signature. Full annotation and associated statistic of each pathway provided in Supplementary Data 8.

74 days). Both were confirmed acute leukemias with secondary recipients succumbing to leukemia with a significantly decreased latency (Fig. 4b).

Although NRAS(G12D) alone and TCF7-SPI1-P2A-NRAS (G12D) both gave rise to leukemia with a similar latency, a clear difference was observed in disease presentation and immunophenotype. Mice expressing NRAS(G12D) alone within developing T cells had an enlarged thymus but rarely had increased white blood cell (WBC) counts or leukemic invasion of the spleen. In contrast, mice expressing both NRAS(G12D) and TCF7-SPI1 with leukocytosis and splenomegaly but no enlarged thymus (Fig. 4c). This phenotype remained unchanged after secondary transplantation (Fig. 4d). The resulting leukemias were further phenotyped using flow cytometry with NRAS(G12D) leukemic thymic cells having a more differentiated CD4+/CD8+ immunophenotype and TCF7-SPI1 + NRAS(G12D) leukemic cells a CD4−/CD8−, (intracellular) CD3+ and CD117+ immature immunophenotype (Fig. 4e). Subsequently, an in vivo limiting dilution assay was performed to evaluate the frequency of leukemic stem cells (LSCs). This showed a significantly higher LSC frequency in the TCF7-SPI1-P2A-NRAS(G12D) mice (1/1243) compared to the NRAS(G12D) only condition (1/312092) (Supplementary Fig. 5). TCF7-SPI1-induced immature leukemia requires upstream β-catenin activity. Given the identification of Wnt/β-catenin signaling signatures in association with SPI1 expression in the single-cell data and the presence of a β-catenin binding domain at the N-terminal region of the fusion protein, we next determined whether the N-terminal β-catenin binding domain of TCF1 within the fusion protein is also important in leukemia development. TCF proteins are well known to be effectors of canonical Wnt signaling together with β-catenin27. Furthermore, β-catenin has previously been suggested to be critical for LSC self-renewal and more recently linked to maintaining LSC stemness and SPI1 expression28–30. We, therefore, set up a new bone marrow transplant where we selectively removed the first 55 amino acids of TCF1 (TCF7) that comprise the β-catenin binding site (Fig. 5a). Similar to the leukemia that develops in NRAS(G12D) only mice (Fig. 4), recipient mice that received Δβ-Cat-TCF7-SPI1 only had an enlarged thymus, limited splenic invasion, and a more mature CD4+CD8− immunophenotype with similar disease latency (median DFS = 67.5 days) (Fig. 5b–d). This indicates that the deletion of the β-catenin binding domain abolished the action of the fusion leading to a NRAS(G12D) only disease.

To further determine whether loss of β-catenin binding was responsible for the altered disease presentation, we generated a series of GAL4 DNA binding domain fusion constructs, containing the N-terminal TCF7 part of the TCF7-SPI1 fusion or a truncated Δβ-Cat-TCF7, and performed luciferase reporter gene expression assays to assess transcriptional activity in HEK293T cells in the presence or absence of the GSK3β inhibitor CHIR-99021 to activate upstream Wnt signaling and increase β-catenin levels. The activation of Wnt signaling resulted in increased luciferase activity only for the full length but not for the Δβ-Cat-TCF7 construct (Fig. 5e, Supplementary Fig. 6a) and this was phenocopied in the melanoma Mel888 cell line, where β-catenin is constitutively active due to a S37F mutation31 (Fig. 5f).

We observed a similar difference between TCF7-SPI1 and Δβ-Cat-TCF7-SPI1 fusion constructs using Lambda B1 or Fes promoter luciferase reporter assays that both have endogenous SPI1 binding sites32 (Supplementary Fig. 6b). The luciferase activity induced by full length SPI1 is lower than the activity induced by the fusion TCF7-SPI1 and is at a similar level as for the Δβ-Cat-TCF7-SPI1 fusion (Fig. 5g). Taken together, these data demonstrate that the TCF7-SPI1 fusion induced immature T-ALL phenotype is driven in part by upstream β-catenin activity.

Genetic and small molecule antagonism of the TCF7-SPI1 fusion results in leukemia phenotype differentiation. To further characterize the function of the TCF7-SPI1 fusion in leukemic maintenance and survival, we undertook an inducible knockdown strategy using an all in one Tet-On vector system to express short hairpin RNAs (Fig. 6a)33. X09 PDX cells were transduced with constructs encoding either a shRNAmir targeting the fusion transcript (shSPI1_885) or control Renilla luciferase (shREN_713), sorted and expanded to reach >90% mCherry positivity in vivo, then divided to receive either doxycycline or control chow at 0.1–1% human CD45 in the peripheral blood (Fig. 6b). At an end stage determined a priori, there was a 50% decrease in fusion expression but no decrease in leukemia burden in the peripheral blood or reduction of spleen weight upon doxycycline-induced knockdown of the fusion (Supplementary Fig. 7a, b). Strikingly, there was a significant increase in CD4 and CD8 and loss of CD117 (cKIT) cell surface expression with concomitant changes at the mRNA level (Fig. 6c, d, Supplementary Fig. 7b). This increased phenotypic differentiation was further supported by RNA-seq analysis, where knockdown of the fusion significantly decreased the ETP-ALL signatures (Fig. 6e, Supplementary Fig. 7c).

A global analysis of the differentially expressed genes to identify common regulatory elements using iCisTarget34 revealed that the significantly downregulated genes were highly enriched for SPI1 binding motif (Fig. 6f). To confirm the requirement of β-catenin for the oncogenic function of TCF7-SPI1 and to confirm the effect of fusion knockdown on differentiation, we made use of the small molecule β-catenin/TCF antagonist PKF 118-31032,35. PKF 118-310 completely abrogated the luciferase signal generated by TCF7-GAL4 or TCF7-SPI1 in Mel888 cells phenocopying the genetic removal of the β-catenin binding domain (Supplementary Fig. 7d). Moreover, similar to the iCisTarget results after knockdown of the fusion, treatment with PKF-118-310 in three TCF7-SPI1 fusion positive PDX samples also showed an enrichment of SPI1 motifs in the downregulated genes and decreased cKIT expression (Fig. 6f, Supplementary Fig. 7e, f).

However, ex vivo treatment with PKF 118-310 did not show specific cytotoxicity toward fusion positive T-ALL cases compared to fusion-negative T-ALL cases (Supplementary Fig. 7g).

Taken together, these data demonstrate that the interaction of β-catenin with TCF1 (TCF7) is important for increasing the
oncogenic function of the TCF7-SPI1 fusion and that its inhibition leads to downregulation of SPI1 targets and phenotypic differentiation of the leukemic cells.

Discussion
In 2017, Seki et al. described a distinct subset of pediatric T-ALL cases with dismal prognosis, characterized by an aberrant expression of SPI1 fusion genes. Seven out of 181 investigated patients carried translocations encompassing the SPI1 locus on chromosome 11 with TCF7 or STMN1 as typical fusion partners. Even though TCF7 appeared to be a stronger activator of PU.1 than STMN1, the authors were unable to induce T-cell leukemia in a mouse model by constitutively expressing TCF7-SPI1 on its own. This suggests that additional proliferative and oncogenic signals are needed for leukemic transformation. In our center we identified a patient with a CD4/CD8 double negative T-ALL, characterized by the TCF7-SPI1 fusion in combination with a...
NRAS(G12D) mutation. The frequency of mutations in RAS signaling in pediatric T-ALL is estimated to approximate 15%, but in the cohort of Seki et al., NRAS mutations were found in almost half of the cases with a SPI1 fusion.36

In this current work, we were able to demonstrate that the TCF7-SPI1 fusion can cooperate with NRAS(G12D) to induce an aggressive, CD4/CD8 double negative T-cell leukemia in vivo. Concordant with Seki et al., TCF7-SPI1 alone did not induce T-ALL in our bone marrow transplant model. Expressing NRAS (G12D) alone using our conditional retroviral expression system limiting NRAS(G12D) to developing T cells caused a mature T-cell leukemia/lymphoma characterized by thymus enlargement. Other research groups have also reported on the development of hematologic diseases such as myeloproliferative diseases and T-ALL when expressing oncogenic RAS mutations in mouse progenitor cells.37,38 However, combining both oncogenic alterations in our conditional bone marrow transplant model resulted in an aggressive acute T-ALL characterized by hyperleukocytosis, splenomegaly and the absence of thymus enlargement. Although no difference in disease latency was observed, the altered disease presentation when both TCF7-SPI1 and NRAS(G12D) are expressed was also found to cause a significantly higher frequency of leukemic stem cells. The combination also resulted in a more immature immunophenotype than mutated NRAS alone with a double negative leukemia (CD4− and CD8−), similar to the immunophenotype of the patient X09. This suggests that the TCF7-SPI1 fusion provides a transcriptional program to block differentiation in the early thymic progenitor stage of development. Recently, it was shown that the TCF7-SPI1 fusion blocks differentiation of murine T cells in the DN3a stage.39

This is also in part supported by our earlier findings,40 where the order of mutual acquisition predicted that the translocation leading to the TCF7-SPI1 fusion occurred prior to acquisition of the NRAS mutation. Hence, the observed immature T-ALL phenotype in our experimental bone marrow transplant models is the result of the fusion blocking T-cell differentiation and the NRAS(G12D) mutation providing the necessary proliferative signals for leukemia development. This model was also recently described for a TCF3-HLF fusion protein in B-ALL which cooperates with ETS factors and also leads to a block in differentiation.41 Our data, therefore, posit that the observed occurrence of SPI1 fusions with RAS mutants is driven by a positive selective pressure in the development of T-ALL.

We also demonstrated that the TCF7-SPI1 fusion is dependent on β-catenin. Genetically, the TCF7-SPI1 construct lacking the β-catenin binding domain was sufficient to alter the disease presentation in vivo to be equivalent to NRAS(G12D) alone with a more mature CD4+CD8+ immunophenotype. This was further demonstrated using PKF 118-310, which specifically inhibits the interaction between β-catenin and TCF1 and led to significant downregulation of SPI1 target genes in three separate TCF7-SPI1 fusion PDX cases.

In normal T-cell development, the controlled expression of the SPI1 transcription factor plays an important role in delaying the onset of T-cell commitment and possibly contributing to the expansion of early T cells.42 In T-ALL, the mutually exclusive and ectopic expression of transcription factors such as TLX1, TLX3, and TAL1 is a key feature of T-ALL and often results from chromosomal translocations.43 In this current work, we demonstrate that ectopic expression of SPI1 can occur via chromosomal translocations and a unique situation where β-catenin hijacks the SPI1 transcriptional program. This is in line with the recent finding that expression of SPI1 is important for maintaining the ‘stemness’ of T-ALL leukemic stem cells in a regulatory loop involving both HAVCR2 and β-catenin.44 Given these findings, we speculate that the ectopic and deregulated expression of SPI1 may be a more general feature of immature T-ALL that either occurs by chromosomal translocations or potentially by cryptic mutations in cis-regulatory regions as has been recently described for TAL1.44

Targeting the fusion in vivo with an inducible knockdown system did not have a profound effect on leukemia burden. However, decreasing expression of the fusion did lead to increased differentiation, providing evidence the differentiation block observed in murine systems is directly relevant to clinical samples in maintaining leukemic cells in an immature state. Induction of fusion knockdown at an earlier time point before reaching 1% human CD45 in the peripheral blood might have resulted in an eventual exhaustion of T-ALL through differentiation even in the presence of the NRAS mutation but could not be answered in our in vivo experiment conditions where our ethical end point was predetermined. Of interest is the comparison with acute promyelocytic leukemia (APL), which was the paradigm of differentiation therapy. In recent years, it has become clearer that by targeting the PML-RARA fusion, other mechanisms, such as senescence, are also contributing to the clearance of APL cells.45 In many other leukemias, e.g., IDH mutant AML, mutations can cause a differentiation block, but are dispensable for disease maintenance.46 Here we show that such a therapy might be relevant for the first time in T-ALL upon the development of improved TCF-β-catenin interaction inhibitors that may also be able to eliminate leukemic stem cells. In those cases that also have NRAS mutation, combinatorial treatments might be required. Interestingly, during the revision of this manuscript, Gocho et al. identified T-ALL cases that are sensitive to dasatinib, some of which were positive for TCF7-SPI1 without a NRAS mutation.47 What remains unclear is the precise mechanism by which mutated NRAS and the TCF7-SPI1 fusion cooperate to drive T-ALL. NRAS might exert its oncogenic function through the MAPK pathway. On the other hand, RAS is known to activate the PI3K/AKT pathway, thereby inhibiting GSK3β and thus activating β-catenin.47 Independent of the mechanism, we do show for the first time that antagonism of the TCF/β-catenin interaction is sufficient to induce differentiation, providing a
Fig. 4 Conditional co-expression of TCF7-SPI1 fusion and NRAS(G12D) within developing T cells cooperate to generate an immature T-cell acute lymphoblastic leukemia. a Schematic representation of the conditional bone marrow transplant (BMT) model. b Survival curve showing the disease-free survival with no significant difference in disease latency between NRAS only and TCF7-SPI1-P2A-NRAS (p = 0.7429). Secondary transplantations (dashed lines) have significantly shorter disease latencies (p = 0.0008 for NRAS only; p < 0.0001 for TCF7-SPI1-P2A-NRAS). Log-rank (Mantel–Cox) test. c White blood cell (WBC), spleen weight and thymus weight at end stage for the primary BMT. n = 4 biologically independent animals for TCF7-SPI1, 8 for NRAS only (for the WBC graph n = 9) and n = 9 biologically independent animals for TCF7-SPI1 + NRAS in the three graphs. d Equivalent WBC, spleen, and thymus data for secondary transplantation. e Each point represents a different mouse, n = 5 biologically independent animals for NRAS only and 6 for TCF7-SPI1 + NRAS; the mean is shown with standard deviation. P values are indicated and were calculated by a one-way ANOVA with post hoc Dunnett T3 multiple comparisons test for (c) and a two-tailed Mann–Whitney test for (d). e Immunophenotyping of two different mice from (b) for each condition with spleen or thymus cells. Flow cytometry for cyCD3, CD4, CD8, CD117 are shown after gating on the viable and GFP+ cells. Distinguishing markers between the two conditions and their respective percentages are highlighted in red.
unique proof of principle that, in this instance, it is possible to directly target an oncogenic transcription factor.

In conclusion, we have shown that TCF7-SPI1 cooperates with NRAS (G12D) to drive an aggressive immature T-cell leukemia in vivo and that it is possible to target this fusion protein by inhibiting protein-protein interactions between TCF1 and β-catenin. Patients with the TCF7-SPI1 fusion could thus potentially benefit from an improved pharmacological targeting of the TCF/β-catenin interaction.

Methods
Patient sample collection and storage. Fresh bone marrow and peripheral blood samples were collected from newly diagnosed ALL patients in the Pediatric Hemato-Oncology Department of the University Hospitals Leuven at time of...
**Fig. 5** The N-terminal β-catenin binding site of the TCF7-PU.1 fusion protein is essential for transcriptional activity and ETP-ALL development in vivo.

- **a** Schematic representation of the protein domains resulting from the entire TCF7-SP1I fusion gene (above) or without (below) the β-catenin binding domain (CTNNB1) at the N-terminus (first 55 amino acids).
- **b** Kaplan-Meier curve for disease-free survival with the indicated inducible constructs. Log-rank (Mantel–Cox) test, \( p = 0.0589 \). Spleen and thymus weights at end stage for mice in (b). Dots represent different mice (\( n = 6 \) biologically independent animals for Δβ-Cat-TCF7-SP1I and 5 for TCF7-SP1I); the mean is shown with standard deviation. P values are indicated and calculated by a two-tailed Mann–Whitney test. 
- **c** Flow cytometry staining of a thymus of a Δβ-Cat-TCF7-SP1I-P2A-NRAS(G12D) mouse. Staining for CD4 and CD8 is shown after gating on GFP-positive cells. 
- **d-g** Dual luciferase assay as described in “Methods”. e HEK293T cells with GAL4 constructs. Over-night treatment with CHIR-99021 at a dose of 1 \( \mu \text{M} \). f Mel888 cells with GAL4 constructs. g Mel888 cells with SPI1 reporter gene containing the L1B promoter or mutant. Results were normalized to the average GAL4 signal. P values (e–g) are indicated, one-way ANOVA with post hoc Dunnett T3 multiple comparisons test. Mean with standard deviation is shown. Dots represent different samples, \( n = 3 \) independent experiments per condition. Repeat experiments of (e) and (f) are shown in Supplementary Fig. 5.
primers according to the manufacturer’s protocol (GoScript cDNA Synthesis Kit, Promega). Real-time quantitative was performed using GoScript SYBR master mix kit (Promega) with the QuantStudio 3 PCR system. (Applied Biosystem). Quality control, primer efficiency, and data analysis were carried out using qbase+ software (v3.2; Biogazelle). All gene expression was normalized using two housekeeping reference genes. The primer sequences for the qPCR experiments in Supplementary Fig. 7 can be found in Supplementary Data 12.

Dose-response curve. Leukemic cells were seeded into a 96-well plate at 300,000 cells/mL and the drug was added using a D300e digital dispenser (Tecan) in an increasing dose. After 24 h, cell proliferation was measured using the ATPlite luminescence system on a Victor multilabel plate reader (PerkinElmer). Results were normalized to the DMSO condition. Results were analyzed with a non-linear regression. The analysis of the J188 PDX mouse was undertaken independently by the lab of Itaru Kato following the same protocol.
Bulk RNA-seq analysis of patient cohorts. RNA-seq data were retrieved for 123 T-ALL patients from the National Bioscience Database Center (NBDC) of the Japan Science and Technology Agency (JST), under the accession JGAS00000000090 and approved by the University of New South Wales institutional review board human ethics committee (HREAP HC200562). Patients and/or their guardians provided informed consent in accordance with the Declaration of Helsinki. RNA-seq data of PKF 118-310 or DMSO treated PDX cells and sorted mCherry knockdown RNA samples were polyA-enriched and sequenced at BGI Genomics. Paired-end reads were then aligned to the human genome (GRCh38) using STAR (v.2.7.5c)\(^55\) in two-pass mode, whilst also retaining junction/chimeric reads for future fusion transcript detection. Aligned reads were then counted using HTSeq (v0.12.1)\(^66\) using GENCODE (v22) gene annotations. Gene-level read counts were additionally normalized to transcripts per million (TPM) using a custom R script (zenodo repository DOI: 10.5281/zenodo.4756105). Transcripts were then detected using arriba (v1.2.0)\(^72\) with previously generated STAR outputs. Differential gene expression analysis were conducted under biological replicates using the DESeq2 R package (v1.22.0)\(^73\), where differentially expressed genes were considered as those genes with a fold-change >1 or <−1 and Benjamini-Hochberg adjusted \(p\)-value < 0.05 (Supplementary Data 13). Gene set enrichment analyses (GSEA) were performed using the Zhang and Liu ETP-ALL signature gene sets. Normalized enrichments scores, \(p\)-values, and FDR \(q\)-values were indicated and were calculated with a hypergeometric distribution. All results can be found in Supplementary Data 15–22.

Mutational analyses. Somatic mutations previously identified by targeted sequencing in Seki et al.\(^74\) were first obtained. Somatic mutational burdens in the remaining 6 patients were derived from WGS data and annotated using the ensemble variant value < 0.05 (Supplementary Data 13). Gene set enrichment analyses (GSEA) were performed using the Zhang and Liu ETP-ALL signature gene sets. Normalized enrichments scores, \(p\)-values, and FDR \(q\)-values were indicated and were calculated with a hypergeometric distribution. All results can be found in Supplementary Data 15–22.

Single-cell RNA sequencing data analysis. Bone marrow samples from four T-ALL patients (X09, XB37, XB41, XB47) were previously profiled using the 10x Genomics Chromium\(^\text{TM}\) 3′v3 single-cell gene expression platform\(^82\). In addition, samples from two T-ALL patients (STJALL030263, STJALL301201) were sourced from the St. Jude Biorepository and profiled using 10x Genomics Chromium\(^\text{TM}\) 3′v3 chemistry. Reads were demultiplexed and aligned to the human reference genome (GRCh38), followed by unique molecular identifier (UMI) counting using the cellRanger v3.0.2 software suite\(^63\). Raw UMI count matrices were used as input for the Seurat v3 package (v3.2.3) for downstream processing of the data\(^83\). In droplet-based scRNA-seq methods, droplets can often contain ambient RNA and be misinterpreted as a cell, therefore we eliminated these putative empty droplets using the DropletUtils package (v1.4.3), set at a false discovery rate (FDR) of 0.01 (1%)\(^84\). Furthermore, we filtered out “poor quality” cells that contained either: high proportions of ribosomal reads (>0.6), high proportions of mitochondrial reads (>0.1) or a low number of expressed genes (<200)\(^85\). All cells were then integrated and batch-corrected using scran while regressing out features in the data associated with technical variation, such as; the number of genes expressed, the number of reads per droplet, read mapping quality, and uniform manifold approximation and projection (UMAP) based methods without Harmony. Principle components with the highest variance were used to semi-supervised cell clustering using the Louvain algorithm\(^62,66\), wherein the optimal clustering resolution was determined using the elbow criterion (v0.4.1)\(^68\). Dimensional reduction was then performed using UMAP to visualize transcriptional variation among single cells\(^69,70\). SP11 expression correlations were computed using median-scaled log-transformed & normalized gene expression values from the “data” slot of the Seurat object. Correlation estimate was calculated according to the Spearman correlation coefficient. Expression heatmaps were generated for highly correlated and anti-correlated genes and visualized using the ComplexHeatmap package (v2.0.0)\(^71\). Gene-set enrichment analysis was undertaken on a dataset ranked according the SP11 Pearson correlation coefficient, and enrichment statistics were evaluated for gene-sets of the Molecular Signatures Database (MSigDB v7.4) using the gsea package (v1.41.0)\(^72,73\). Gene signature expression scores of Zhang et al.\(^32\) and Liu et al.\(^74\) signatures were computed for each cell using the AddModuleScore function of the Seurat3 v3 package\(^2,63\) and scores were then re-scaled using z-score normalization. Processed single-cell RNA-seq data of the human fetal thymic cell atlas were directly retrieved from the zenodo data repository under the accession 10.5281/zenodo.3572422. These data contained normalized count matrices, embeddings, and cell type annotations made in the original study\(^26\). Cells corresponding to either “T” or “innate T” cell types in “level 1” meta data annotations as well as early thymic progenitor (ETP) cells were subsetted. Gene signature expression scores were computed as described above. Two-sided pairwise \(t\) tests were performed between cell groups, followed by
Benjamin-Hochberg (BH) adjustment of resultant p values to correct for multiple hypothesis testing.

**Amplification of the TCF7-SPI1 fusion from 10x cdNA libraries.** A total of 0.1 ng of cdNA from 10x cdNA libraries for SJTALL03201_01 was amplified using 2× KAPA HiFi HotStart ReadyMix and 0.4 μM Biotinylated TCF7 Exon 4 forward primer and 0.4 μM Read 1 reverse primer (Supplementary Data 14) using 1 cycle of initial denaturation 95°C for 5 min, followed by 15 cycles of 98°C denaturation for 20 s, 63°C annealing for 30 s, 72°C extension for 2 min; and then a final 72°C extension for 5 min. PCR reactions underwent a 0.9X SPRIselect single sided size selection purification and eluted with 20 μl of nuclease-free water. Biotinylated PCR products were then captured using M-280 Streptavidin Dynabeads and washed 3× with 10 mM Tris-HCl (pH 7.5) 1 mM EDTA 2 M NaCl wash buffer prior to final resuspension in 60 μl nuclease free water. A second nested PCR was then carried out using 5 μl of the streptavidin bead resuspension using 2× KAPA HiFi HotStart ReadyMix and 0.4 μM Nested TCF7 Exon 4 forward primer and 0.4 μM Read 1 reverse primer using 1 cycle of initial denaturation 95°C for 5 min, followed by 30 cycles of 98°C denaturation for 20 s, 63°C annealing for 30 s, 72°C extension for 2 min; and then a final 72°C extension for 5 min. The resulting nested PCR underwent a 0.9X SPRIselect single sided size selection purification and eluted with 20 μl of nuclease free water. Amplification was confirmed by TapeSe\-
tation (Agilent) analysis and Sanger Sequencing (Supplementary Fig. 3) prior to nanopore sequencing.

**NanoPore sequencing and analysis.** Amplified cdNA libraries were prepared for long-read sequencing using Oxford Nanopore Technologies (ONT) 1D adapter ligation sequencing kit (SQK-LSK109). Three samples were barcoded, pooled and sequenced on a single R9.4.1 Promethion flow cell (FLO-PRO002). MinKNOW live base-calling was performed using Guppy (v4.0.11). The resulting base-called FASTQ files were de-multiplexed by 10x cell barcodes, using a direct sequence matching strategy, followed by coverage checking and filtering (Supplementary Fig. 3). Briefly, for each barcode, all the reads spanning TCF7 reads (16 nt) were used to demultiplex the nanopore sequencing reads by scanning the first and last 200 nt of any read longer than 250 nt for a matching sequence, with <2 mismatches. Demultiplexed reads were then grouped into separate FASTQ files that were separately assembled de novo using Canu8 (version 1.8; parameters: -u -w 200 -m 8 -x -1 -g -4). The Minimap2/Racon step was used to map reads for each sample is shown in Supplementary Data 5. Reads spanning TCF7 exon #4, the fusion breakpoint and SPI exon #3 were classified as 'fusion reads', while reads spanning TCF7 exon #4 and spanning SPI exon #7 were classified as 'canonical reads'. Canonical/fusion read counts are summarized in Supplementary Data 5.

**Statistical analysis.** Statistical analyses were performed using the Prism software (Graphpad v9.0.2 (134)). Kaplan–Meier curves were used for the survival of mouse BMT’s with two-sided p values determined by log-rank (Mantel–Cox) test. For the comparison of 2 groups, an unpaired t test with Welch’s correction or Mann–Whitney test was used. Data in the luciferase experiments are expressed as mean ± standard deviation (SD). Groups were compared using the post hoc Dunnett's test from multiple comparisons tests after Brown’s ANOVA and the Student’s t-test. p values are adjusted p values. Expression analysis for different subsets of patients was done with a Tukey’s multiple comparisons test. For the analysis of the dilution assay we used the ELDA (extreme limiting dilution analysis) software.

**Code availability** Custom R-scripts used have been deposited with the Zenodo Biorepository (DOI:10.5281/zenodo.4756105).

**References**

1. De Keersmaecker, K. et al. Exome sequencing identifies mutation in CNOT3 and ribosomal genes RPL5 and RPL10 in T-cell acute lymphoblastic leukemia. *Nat. Genet.* 45, 186–190 (2013).
2. Li, Y. et al. IL-7 Receptor mutations and steroid resistance in pediatric T cell acute lymphoblastic leukemia: a genome sequencing study. *PloS Med.* 13, e1002200 (2016).
3. Degryse, S. et al. JAK3 mutants transform hematopoietic cells through JAK1 activation causing T-cell acute lymphoblastic leukemia in a bone marrow transplant mouse model. *Blood* 124, 3092–3100 (2014).
4. Atak, Z. et al. Comprehensive analysis of transcriptome variation uncovers known and novel driver events in T-cell acute lymphoblastic leukemia. *PloS Genet.* 9, e1003997 (2013).
5. Liu, Y. et al. The genomic landscape of pediatric and young adult T-lineage acute lymphoblastic leukemia. *Nat. Genet.* 49, 1211–1218 (2017).
6. de Bock, C. E. et al. HOXA9 cooperates with activated JAK/STAT signaling to drive leukemia development. *Cancer Discov.* 8, 616–631 (2018).
7. Vanden Bempt, M. et al. Cooperative enhancer activation by TLX1 and STAT5 drives development of NU1214-ABL1/TLX1-positive T cell acute lymphoblastic leukemia. *Cell Cancer* 34, 271–285 (2018).
8. Broux, M. et al. Sui2 inactivation cooperates with JAK3 mutant signaling in the development of T-cell acute lymphoblastic leukemia. *Blood* 134, 1323–1336 (2019).
9. Seki, M. et al. Recurrent SPI1 (PU.1) fusions in high-risk pediatric T-cell acute lymphoblastic leukemia. *Nat. Genet.* 49, 1274 (2017).
10. Van de Wetering, M. et al. The human T cell transcription factor-1 gene. Structure, localization, and promoter characterization. *J. Biol. Chem.* 267, 8530–8536 (1992).
11. Johnson, J. L. et al. Lineage-determining transcription factor TCF-1 initiates the epigenetic identity of T cells. *Immunity* 48, 243–257 (2018).
12. Moreau-Gachelin, F., Tavitian, A. & Tambourin, P. Spi-1 is a putative oncogene in virally induced murine erythroleukemias. *Nature* 331, 277–280 (1988).
13. Galson, D. L. et al. Mouse beta-globin DNA-binding protein B1 is identical to a proto-oncogene, the transcription factor Spi-1/PU.1, and is restricted in expression to hematopoietic cells and the testis. *Mol. Cell. Biol.* 13, 2929–2941 (1993).
14. Rosenbauer, F. et al. Acute myeloid leukemia induced by graded reduction of a lineage-specific transcription factor, PU.1. *Nat. Genet.* 36, 624–630 (2004).
15. Steidl, U. et al. Essential role of Jun family transcription factors in PU.1 expression and commitment and elution by TapeSe\-
tation (Agilent) analysis and Sanger Sequencing (Supplementary Fig. 3) prior to nanopore sequencing.

**Data availability** The data of bulk and single-cell RNA sequencing of 4 T-ALL patients (X09, XB37, XB41, and XB47), nanopore sequencing of X09, together with the single-cell RNA sequencing data of a PDX model of the X09 patient and sequencing data of two patients from the St-Jude cohort (SJTALL032063 and SJTALL031201), have been deposited within the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG. Accession number EGA500010056079. https://ega-archive.org/datasets/EGAD000010070010. The Seki et al. RNA-seq data are available at the National Biomedical Database Center, Japan Science and Technology Agency (ID: JGAD0000000090). The whole-genome sequencing data of the four T-ALL samples (X09, XB37, XB41, and XB47) from De Bie et al.46 can be accessed through the following link: https://ega-archive.org/datasets/EGAD000010003951. For restricted data access, restricted data access can be requested by sending an email to Sofie Demeyer (sofie.demeyer@kuleuven.be) with requests for data use outside academic research. restricted data access data access agreement. Processed single-cell data have been deposited with the Zenodo Biorepository (DOI:10.5281/zenodo.4756105). https://zenodo.org/record/4756105#.XMyppC0RqPq. For Gene Set Enrichment Analysis, Molecular Signatures Database was used (http://www.gsea-msigdb.org/gsea/msigdb/index.jsp). Source data are provided with this paper.
27. Nusse, R. & Clevers, H. Wnt/β-catenin signaling, disease, and emerging therapeutic modalities. Cell 169, 985–999 (2017).
28. Rehmsmeier, M. et al. LympoHed cell growth and transformation are suppressed by a key regulatory element of the gene encoding PUN1. Nat. Genet. 38, 27–37 (2006).
29. Guo, W. et al. Multi-genetic events collaboratively contribute to Pten-null leukemia stem-cell formation. Nature 453, 529–533 (2008).
30. Zhu, H. et al. T-ALL leukemia stem cell/stemness is epigenetically controlled by the master regulator SP1I. Elife 7, e3849 (2018).
31. Rubinfeld B. et al. Stabilization of beta-catenin by genetic defects in melanoma cell lines. Science 275, 1790–1792 (1997).
32. Roos-Weil, D. et al. A recurrent activating missense mutation in Waldenström macroglobulinemia affects the DNA binding of the ETS transcription factor SPI1 and enhances proliferation. Cancer Discov. 9, 796–811 (2019).
33. Fahlbusch, C. et al. An optimized microRNA backbone for effective single-copy RNAi. Cell Rep. 5, 1704–1713 (2013).
34. Herrmann, C., Van De Sande, B., Potier, D. & Aerts, S. cTisTarget: an integrative genomics method for the prediction of regulatory features and cis-regulatory modules. Nucleic Acids Res. 40, e14, e12 (2012).
35. Krishnamurthy, N. & Kurzrock, R. Targeting the Wnt/β-catenin pathway in cancer: Update on effectors and inhibitors. Cancer Treat. Rev. 62, 50–60 (2018).
36. Perentes, J. P. et al. RAS oncogene mutations and outcome of therapy for childhood acute lymphoblastic leukemia. Leukemia 18, 685–692 (2004).
37. Wang, J. et al. Endogenous oncogenic Nras mutation initiates hematopoietic malignancies in a dose- and cell-type dependent manner. Blood 118, 368–378 (2011).
38. Staffas, A., Karlsson, C., Persson, M., Palmqvist, L. & Bergo, M. O. Wild-type KRAS inhibits oncogenic KRAS-induced T-ALL in mice. Blood 118, 2841–2849 (2011).
39. 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).
40. Huang, Y. et al. The leukemicogenic TCF3-HLF complex rewires enhancers driving cellular identity and self-renewal conferring EP300 vulnerability. Cancer Cell 36, 630–644.e9 (2019).
41. Yui, M. A. & Rothenberg, E. V. Developmental gene networks: a triathlon on the course to T cell identity. Nat. Immunol. 14, 529–545 (2014).
42. Girardi, T., Vicente, C., Cools, J. & De Keersmaecker, K. The genetics and the molecular biology of T-ALL. Blood 129, 2–18 (2017).
43. Mansour, M. R. et al. An oncogenic super-enhancer formed through somatic mutation of a noncoding intergenic element. Science 363, 1373–1377 (2019).
44. de Thé, H., Pando, M. P. & Chen, Z. Acute promyelocytic leukemia: a paradigm for oncprotein-targeted cure. Cancer Cell 32, 552–560 (2017).
45. Lewis, A. C. & Kats, L. M. Non-genetic heterogeneity, altered cell fate and differentiation therapy. EMBO Mol. Med. 13, e12670 (2021).
46. Dillow, A. Y. & Horwitz, R. A. Targeting RAS signalling pathways in cancer therapy. Nat. Rev. Cancer 3, 11–22 (2003).
47. Potter, N. E. et al. Single-Cell mutational profiling and clonal phylogeny in cancer. Genome Res. 32, 2115–2125 (2013).
48. Xu, X. et al. Single-cell exome sequencing reveals single-nucleotide mutation characteristics of a kidney tumor. Cell 148, 886–895 (2012).
49. Teuwen, J. W. M., Huang, S., Salford, M., Lander, S. P. & Loken, M. R. Sequential generations of hematopoietic colonies derived from single nonlineage-committed CD34+CD38− progenitor cells. Blood 77, 1218–1227 (1991).
50. Li, H. Minimap2: Pairwise alignment for nucleotide sequences. Bioinformatics 34, 3094–3100 (2018).
51. Gehre, N., et al. A stromal cell free culture system generates mouse pro-T cells that can reconstitute T-cell compartments in vivo. Eur. J. Immunol. 45, 932–942 (2015).
52. Bornschein, S. et al. Defining the molecular basis of oncogetic cooperation between TAL1 expression and Pten deletion in T-ALL using a novel pro-T-cell model system. Leukemia 32, 941–951 (2018).
53. Thorne, N., Inglese, J. & Auld, D. S. Illuminating Insights into firefly luciferase and other bioluminescent reporters used in chemical biology. Chem. Biol. 17, 646–650 (2010).
54. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 13–21 (2015).
55. Dobin, A. et al. STAR: Ultrafast universal RNA-seq aligner. Bioinformatics 29, 13–21 (2015).
56. Anders, S., Pyl, P. T. & Huber, W. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169 (2015).
57. Uhlirg, S. et al. Accurate and efficient detection of gene fusions from RNA sequencing data. Genome Res. 31, 448–460 (2021).
58. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 590 (2014).
59. Segalhevich, A. A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. bioRxiv. Preprint at https://doi.org/10.1101/060012 (2016).
60. McLaren, W. et al. The ensemble variant effect predictor. Genome Biol. 17, 122 (2016).
61. Zheng, G. X. Y. et al. Massively parallel digital transcriptional profiling of single cells. Nat. Commun. 8, 14049 (2017).
62. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. 36, 411–420 (2018).
63. Moon, T. et al. Comprehensive integration of single-cell data. Cell 177, 1888–1902.e21 (2019).
64. Lun, A. et al. EmptyDrops: distinguishing cells from empty droplets in droplet-based single-cell RNA sequencing data. Genome Biol. 22, 63 (2019).
65. Illici, T. et al. Classification of low quality cells from single-cell RNA-seq data. Genome Biol. 17, 29 (2016).
66. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. Genome Biol. 20, 296 (2019).
67. Korsunsky, I. et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat. Methods 16, 1289–1296 (2019).
68. Zappa, L. & Oshlack, A. Clustering trees: a visualization for evaluating clusterings at multiple resolutions. Gigascience 7, giy083 (2018).
69. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: Uniform manifold approximation and projection. J. Open Source Softw. 3, 861 (2018).
70. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. Nat. Biotechnol. 37, 38–44 (2019).
71. Liu, Z., Ellis, R. & Schlemmer, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinformatics 32, 2847–2849 (2016).
72. Korotkevich, G., Sukhov, V. & Serghishevich, A. Fast gene set enrichment analysis. bioRxiv. Preprint at https://doi.org/10.1101/060012 (2016).
73. Singh, M. et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat. Biotechnol. 38, 1310 (2019).
74. Koren, S. et al. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. Genome Res. 27, 722–736 (2017).
75. Vaser, R., Sović, I., Nagarajan, N. & Škák, M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 27, 737–746 (2017).
76. Hu, Y. & Smyth, G. K. ELDA: Extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J. Immunol. Methods 347, 70–78 (2009).
