






# IRF4 drives clonal evolution and lineage choice in a zebrafish model of T-cell lymphoma

Stella Amanda<sup>1,7</sup>, Tze King Tan<sup>1,7</sup> , Jolynn Zu Lin Ong<sup>1</sup>, Madelaine Skolastika Theardy<sup>1</sup>, Regina Wan Ju Wong<sup>1</sup>, Xiao Zi Huang<sup>1</sup>, Muhammad Zulfaqar Ali<sup>1</sup>, Yan Li<sup>2</sup> , Zhiyuan Gong<sup>2</sup> , Hiroshi Inagaki<sup>3</sup> , Ee Yong Foo<sup>4</sup>, Brendan Pang<sup>4</sup>, Soo Yong Tan<sup>4</sup>, Shinsuke Iida<sup>5</sup> & Takaomi Sanda<sup>1,6</sup>  <sup>✉</sup>

IRF4 is a master regulator of immunity and is also frequently overexpressed in mature lymphoid neoplasms. Here, we demonstrate the oncogenicity of IRF4 *in vivo*, its potential effects on T-cell development and clonal evolution using a zebrafish model. IRF4-transgenic zebrafish develop aggressive tumors with massive infiltration of abnormal lymphocytes that spread to distal organs. Many late-stage tumors are mono- or oligoclonal, and tumor cells can expand in recipient animals after transplantation, demonstrating their malignancy. Mutation of *p53* accelerates tumor onset, increases penetrance, and results in tumor heterogeneity. Surprisingly, single-cell RNA-sequencing reveals that the majority of tumor cells are double-negative T-cells, many of which express *trc-γ* that became dominant as the tumors progress, whereas double-positive T-cells are largely diminished. Gene expression and epigenetic profiling demonstrates that *gata3*, *mycb*, *lrrm1*, *patl1* and *psip1* are specifically activated in tumors, while genes responsible for T-cell differentiation including *id3* are repressed. IRF4-driven tumors are sensitive to the BRD inhibitor.

<sup>1</sup>Cancer Science Institute of Singapore, National University of Singapore, 117599 Singapore, Singapore. <sup>2</sup>Department of Biological Sciences, National University of Singapore, 117543 Singapore, Singapore. <sup>3</sup>Department of Pathology and Molecular Diagnostics, Nagoya City University Graduate School of Medical Sciences, Nagoya 467-8601, Japan. <sup>4</sup>Department of Pathology, Yong Loo Lin School of Medicine, National University of Singapore, 117599 Singapore, Singapore. <sup>5</sup>Department of Hematology and Oncology, Nagoya City University Graduate School of Medical Sciences, Nagoya, Aichi 467-8601, Japan. <sup>6</sup>Department of Medicine, Yong Loo Lin School of Medicine, National University of Singapore, 117599 Singapore, Singapore. <sup>7</sup>These authors contributed equally: Stella Amanda and Tze King Tan. ✉email: [takaomi\\_sanda@nus.edu.sg](mailto:takaomi_sanda@nus.edu.sg)

The transcription factor IRF4, a member of the interferon regulatory factor (IRF) family, is an essential regulator of immunity<sup>1–3</sup>. The expression of IRF4 is highly restricted to hematopoietic and adipocytic lineages, and its level culminates in terminally differentiated lymphocytes, such as plasma cells and activated T-cells<sup>4</sup>. In contrast to other IRF family genes, IRF4 expression is not induced by type I or II interferon but by diverse stimuli, such as T-cell receptor (TCR) signaling in T-cells, IL-4 and CD40 in B-cells, and Toll-like receptor signaling in macrophages<sup>5,6</sup>. IRF4 is required for normal functioning and developmental processes in these cells<sup>1,4,7–11</sup>, particularly as a critical mediator of TCR signaling.

IRF4 has also been implicated as an oncogene in various mature lymphoid neoplasms<sup>12–22</sup>. IRF4 activates oncogenic pathways by transcriptionally regulating downstream target genes, such as MYC, and also coordinately regulate transcriptional program with NF- $\kappa$ B<sup>18,19,22–24</sup>. Functional studies have demonstrated that IRF4 and its binding partners are selective dependency genes; that is, malignant cells are highly dependent on these genes for their proliferation and survival<sup>18,19,23,25</sup>. These in vitro studies suggest that IRF4 can be a driver oncogene; however, the transgenic animal models necessary to investigate the oncogenic property of IRF4 in vivo have not yet been established.

Here, we establish the IRF4-driven lymphoma animal model demonstrating the emergence of highly-invasive tumors that infiltrate multiple organs. Single-cell RNA-sequencing (scRNA-seq) analysis reveal preferential expansion of malignant clones, which can expand in the recipient animals. Gene expression and epigenetic profiling demonstrate that tumors are characterized by the expression of orthologues of human *TCR- $\gamma$* , *GATA3*, and *MYC*.

## Results

**Overexpression of wild-type IRF4 induces aggressive tumors in zebrafish.** We established Tg(*lck:IRF4*) transgenic zebrafish that overexpressed human *IRF4* and a fluorescent marker (*mCherry*) genes (Fig. 1A). *IRF4* DNA sequence, which includes the IRF superfamily domain and DNA recognition site, is highly conserved among human, mouse and zebrafish (Supplementary Fig. 1A). To analyze the oncogenic ability of *IRF4* in different lineages and stages of lymphocytes, we decided to utilize zebrafish *lck* promoter because *lck* is expressed both in the immature and mature stages of T- and B-lymphocytes in mice (Supplementary Fig. 1B) and zebrafish (Supplementary Fig. 1C), as also reported by others<sup>26–28</sup>. In this setting, tumors arise from the cells where IRF4 expression is oncogenic. To compare the phenotype, we also established the *rag2* promoter-driven *IRF4*-transgenic line because *rag2* expression is more dominantly observed in CD4 + CD8 + double-positive (DP) stage of T-cells (Supplementary Fig. 1B, C). Thus, these systems enable us to detect tissue- and stage-specific oncogenicity of *IRF4*.

After we confirmed the successful integration of the transgenes (*IRF4* and *mCherry*) in somatic tissue of the same animal and actual expression of IRF4 (Supplementary Fig. 1D–G), we monitored the phenotype of transgenic animals. Interestingly, Tg(*lck:IRF4*) zebrafish showed expansion of *mCherry* signals into the gills (defined as “early-stage”) by 16 weeks in multiple F0 founders (Fig. 1B; Supplementary Fig. 1H) and F1 offspring (Fig. 1C, D). By 32 weeks, the signals had spread throughout the whole body (defined as “late-stage”). In control animals that expressed only the marker gene, Tg(*lck:mCherry*), the *mCherry* signals disappeared after 16 weeks post-fertilization due to spontaneous regression of the thymus (Fig. 1D; Supplementary Fig. 1H). Because multiple independent F0 animals showed the same phenotype, we concluded that the tumors were induced by

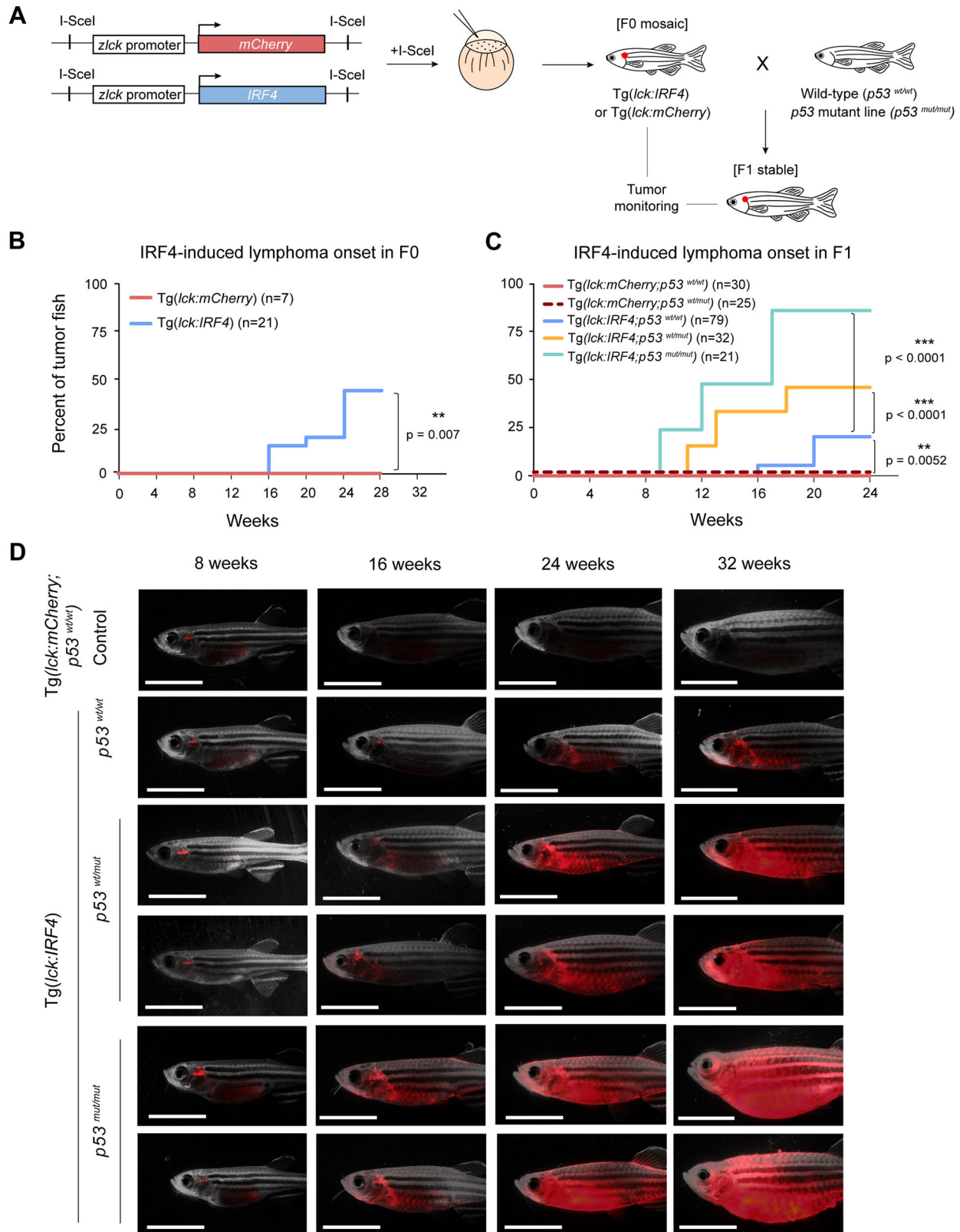
*IRF4* overexpression and were not due to the nonspecific effects of random integration of transgenes. Of note, the overall tumor penetrance was lower and tumor onset was later in F1 animals (Fig. 1C) than in F0 animals (Fig. 1B) because F0 fish that developed more aggressive tumors before sex maturation could not be maintained as founder lines. Flow cytometry analysis suggested that the tumor cells were lymphocytes (Supplementary Fig. 1I). In contrast, *rag2* promoter-driven *IRF4* transgenic zebrafish did not cause any tumors (Supplementary Fig. 1J, K).

In human lymphomas, mutation of *p53* gene often coincides with *IRF4* overexpression<sup>29–31</sup>. Thus, we crossed Tg(*lck:IRF4*) with *p53*-mutant fish that expressed the transactivation-dead *p53* variant mutation (*M214K*) similar to that observed in human cancers<sup>32</sup>. Strikingly, 45% of the heterozygous-mutant Tg(*lck:IRF4;p53<sup>wt/mut</sup>*) and 85% of the homozygous-mutant Tg(*lck:IRF4;p53<sup>mut/mut</sup>*) fish developed tumors by 30 weeks, compared to only 20% in fish with wild-type Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) (Fig. 1C, D). Of note, *p53*-mutant animals expressing only *lck-mCherry*, Tg(*lck:mCherry;p53<sup>wt/mut</sup>*), did not develop any tumors within 30 weeks (Fig. 1C). Thus, the significant difference in tumor onset and penetrance demonstrated the synergy between the mutation of *p53* and *IRF4* overexpression to induce tumorigenesis. This result also indicated that the presence of one mutant *p53* allele is able to accelerate tumorigenesis, suggesting a potential oncogenic effect of mutant *p53* besides loss of wild-type function, as reported in human cancers<sup>33</sup>.

**IRF4-driven zebrafish tumors recapitulate invasive human T-cell lymphoma.** We next examined the clinicopathological features of *IRF4*-induced tumors. Histopathological analysis of tumors using whole-body sections from the fish revealed massive infiltration of abnormal cells into the skin, spinal cord, gill, skeletal muscle, and kidney (Fig. 2A), as well as the gastrointestinal tract, testis, and liver (Supplementary Fig. 2A). The mutation of *p53* resulted in more aggressive tumor infiltration into multiple organs at a younger age. In particular, tumor cells infiltrated epidermis, dermis, and subcutaneous tissue (Fig. 2B, C; Supplementary Fig. 2B). These findings recapitulated invasive human T-cell lymphoma, in particular, several types of peripheral T-cell lymphomas in which tumor cells often aggressively infiltrate into the cutaneous region<sup>34–36</sup>.

**Identification of multiple cell populations derived from different stages of lymphocyte differentiation.** To determine the specific lineages and stages of tumor cells, we performed scRNA-seq analysis for one normal sample from Tg(*lck:mCherry*) at 3 months old (“control”), five non-tumor preleukemic samples from Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) at 3 months old (“P<sub>IRF4</sub>”), three early-stage and six late-stage tumor samples from Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) (“E<sub>IRF4</sub>” and “L<sub>IRF4</sub>”), and five late-stage tumor samples from Tg(*lck:IRF4;p53<sup>wt/mut</sup>*) fish (“L<sub>MUT</sub>”) (Fig. 3A). We sorted all the *mCherry*-positive cells obtained from the whole body by flow cytometry and then performed single-cell separation and labeling, followed by library construction and sequencing analysis. After normalization of the mRNA expression values, we selected *mCherry*-positive cells to exclude contaminated cells and then analyzed different cell populations for individual samples or merged data. An average of 5,497 cells were profiled per sample comprising an average of 4,082 UMIs/cell (~3% duplet rate). Zebrafish *lck* gene was expressed in most of the cells (Supplementary Fig. 3A).

As expected, the control sample contained different stages of T-cells, which were classified into: 1) CD4/CD8 double-negative (DN), 2) CD4/CD8 double-positive (DP), 3) CD4 single-positive (4SP), and 4) CD8 single-positive (8SP), based on the expression



of *cd4-1*, *cd8a* and other marker genes that have been known to be expressed in zebrafish<sup>37-41</sup> (Fig. 3B; Supplementary Data 1). The largest population among all *lck-mCherry* positive cells in the control sample was DP T-cells (Fig. 3B). A fraction of DN cells also showed either single expression of *trgc* or *trdc* or co-expression of *trgc* and *trdc*, which are the orthologues of human TCR- $\gamma$  and TCR- $\delta$  constant regions (Supplementary Fig. 3B),

indicating that they are the intermediate stage of DN cells or committed  $\gamma\delta$  T-cells, respectively, as reported previously in zebrafish<sup>42</sup>. 8SP cells may include intrathymic single-positive (ISP) T-cells, although this population has not been characterized in zebrafish. Together, our results demonstrated that *lck* promoter-driven transgenes are indeed expressed in various stages of normal T-cells and B-cells.

**Fig. 1 Overexpression of wild-type IRF4 induces tumors in zebrafish.** **A** Schematic diagram of the experimental procedure used to overexpress the human *IRF4* gene under the zebrafish *lck* promoter. **B** Tumor onset in FO animals was analyzed in the control Tg(*lck:mCherry*) ( $n = 7$ ) and Tg(*lck:IRF4*) transgenic ( $n = 21$ ) groups;  $p = 0.007$  according to the Gehan-Breslow-Wilcoxon test. **C** Tumor onset in F1 animals was analyzed in the control Tg(*lck:mCherry;p53<sup>wt/wt</sup>*) ( $n = 30$ ), Tg(*lck:mCherry;p53<sup>wt/mut</sup>*) ( $n = 25$ ), Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) ( $n = 79$ ), Tg(*lck:IRF4;p53<sup>wt/mut</sup>*) ( $n = 32$ ) and Tg(*lck:IRF4;p53<sup>mut/mut</sup>*) ( $n = 21$ ) groups.  $p = 0.0052$  for Tg(*lck:mCherry;p53<sup>wt/wt</sup>*) or Tg(*lck:mCherry;p53<sup>wt/mut</sup>*) vs Tg(*lck:IRF4;p53<sup>wt/wt</sup>*);  $p < 0.0001$  for Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) vs Tg(*lck:IRF4;p53<sup>wt/mut</sup>*); and  $p < 0.0001$  for Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) vs Tg(*lck:IRF4;p53<sup>mut/mut</sup>*) according to the Gehan-Breslow-Wilcoxon test. **D** Representative microscopy images of F1 animals: control Tg(*lck:mCherry;p53<sup>wt/wt</sup>*), Tg(*lck:IRF4;p53<sup>wt/wt</sup>*), Tg(*lck:IRF4;p53<sup>wt/mut</sup>*) and Tg(*lck:IRF4;p53<sup>mut/mut</sup>*) zebrafish at 8, 16, 24, and 32 weeks postfertilization. Panels show merged fluorescence and brightfield images. Scale bar = 4 mm. Source data are provided as a Source Data file.

**The expansion of DN and TCR- $\gamma$ -positive T-cells during tumorigenesis.** We then used a merged data to visualize the distribution of each cell population at different stages of tumor progression (Fig. 3C). After the sample processing and selection for mCherry-positive cells, we annotated the cell populations using the aforementioned criteria used to analyze the control sample. In marked contrast to the control sample, as the tumor progressed, DN T-cells expanded while DP population was largely diminished (Fig. 3C, D). Surprisingly, within DN T-cells, there was the preferential expansion of the cells expressing *trgc* and majority of these cells were not accompanied with *trdc* expression (Fig. 3D, line; Supplementary Fig. 3C), suggesting that they may be the intermediate stages of DN T-cells or aberrant *trgc*-expressing T-cells. Additionally, several genes known to be expressed in the  $\gamma\delta$  T-cells, such as *sox4a* and one of *granzymes* (*gzmk*)<sup>34,43–46</sup>, were highly expressed in tumor samples (Supplementary Fig. 3D). Those *trgc*-positive cells further expanded in many tumor samples at the late stage while other populations of  $\gamma\delta$  T-cells stayed throughout tumor progression (Supplementary Fig. 3E). They accounted for approximately 75% of the total T-cells, which was a marked contrast to the percentage in control or pre-leukemic samples (2–10%) (Fig. 3D). We validated this result using independent qRT-PCR analysis of 20 additional tumor samples, which confirmed the upregulation of *trgc* expression in all tumor samples compared to that in control (Supplementary Fig. 3F). It is noteworthy that *trgc* is not expressed in the existing zebrafish T-cell acute lymphoblastic leukemia (T-ALL) model driven by *rag2-myc*<sup>47</sup> (Supplementary Fig. 3G). Hence, the expansion of TCR- $\gamma$ -positive CD4-/CD8-DN T-cells is unique to and is a common phenotype in our zebrafish model.

**IRF4-driven tumor cells possess malignant features.** We next examined if the proliferating cells are indeed malignant cells. We transplanted FACS-sorted mCherry-positive cells from Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) fish into immunocompromised zebrafish (*rag2<sup>E450fs</sup>*-mutant strain) and monitored the engraftment and expansion of mCherry-positive cells. Representative images from three independent experiments with different donor fish are shown (Fig. 4A). This analysis demonstrated that the donor cells from different tumors were able to be engrafted at a high frequency (Fig. 4B) and expanded to the whole body within a month. A serial transplantation assay further revealed that these cells could expand and survive for a long time (Supplementary Fig. 4A) and many tumors were able to be maintained over 1 year. Histopathological examinations revealed that the cells primarily infiltrated the skin, muscle, kidney, liver and spinal cord, which was similar to the observations in the original tumors, except the thymus and gills (Supplementary Fig. 4B). These results indicated that the transplanted cells contained self-renewing cells, which can generate tumor cells for a long term.

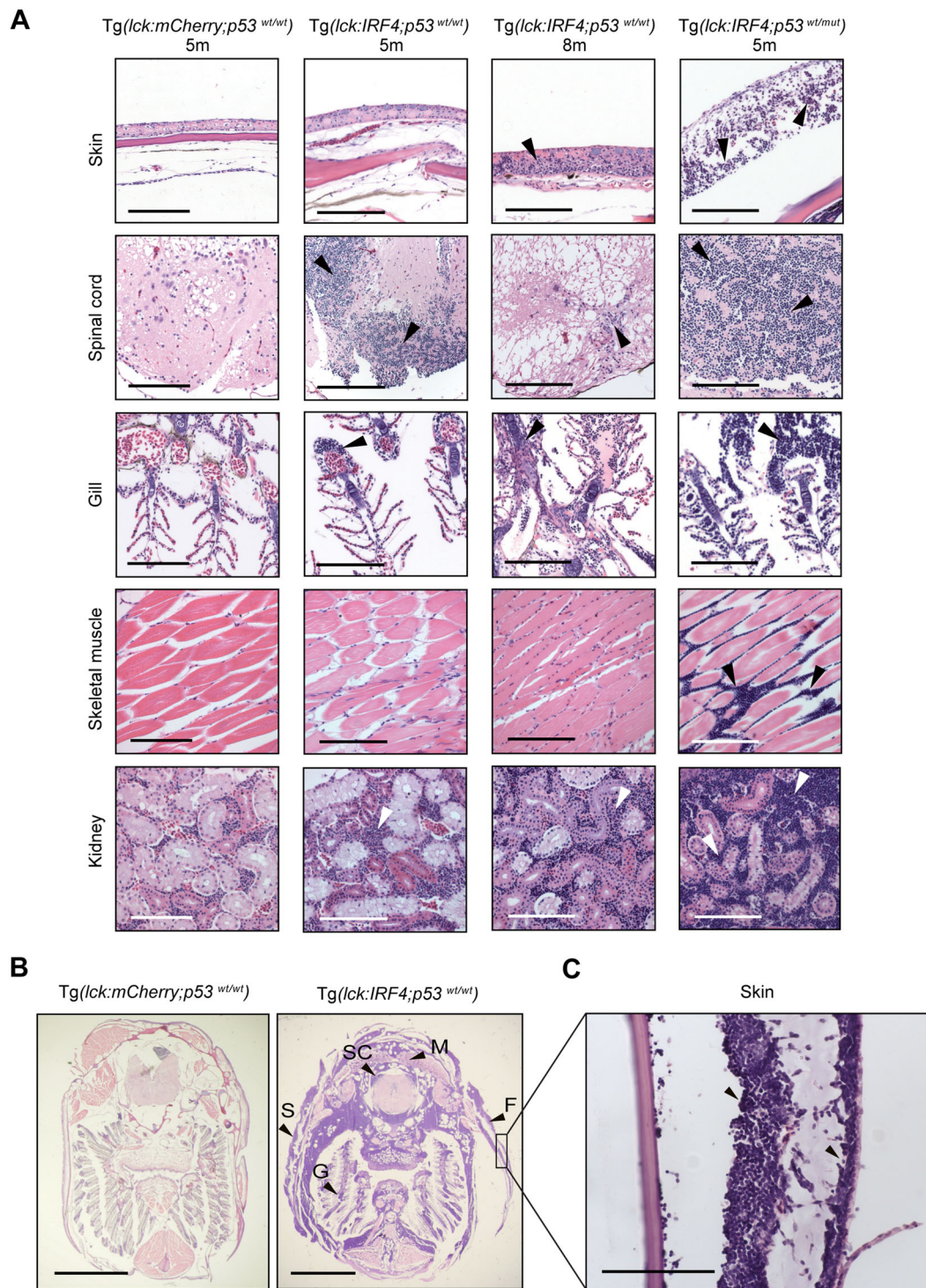
Furthermore, scRNA-seq analysis showed that the T-cell population in the preleukemic and early-stage tumors expressed multiple *trgv* genes (Fig. 4C, left), which are the variable genes of zebrafish TCR- $\gamma$  gene locus (Supplementary Fig. 4C, D), while

those in the late-stage samples expressed only one or a few *trgv* genes at high level (middle), indicating that the tumors were polyclonal at the early stage and became oligo- or monoclonal at the late stage. In the *p53*-mutant background, expression of multiple *trgv* genes was observed, even at the late stage (right). To validate this finding, we performed qRT-PCR for transcripts between TCR- $\gamma$  constant and different variable regions from independent tumor samples (Supplementary Fig. 4E). This analysis revealed that among eight Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) tumor samples that expressed the *trgc* gene, 50% (4 of 8) of them expressed only one *trgv* transcript (Fig. 4D). In contrast, the majority of tumors expressed multiple *trgv* transcripts in the *p53*-mutant background. This result further suggests that the mutation of *p53* confers a strong proliferative or survival advantage to multiple different clones at the early stage of tumorigenesis, thereby contributing to tumor heterogeneity.

**IRF4-driven tumors are characterized by high expression of *mycb* and *gata3*.** We next aimed to elucidate potential mechanisms by which IRF4 induces tumorigenesis. We performed one-to-one differential gene expression analysis using scRNA-seq data between DN populations of different tumor stages, and selected genes, which were significantly upregulated ( $p$  value  $< 0.05$ ) (Fig. 5A). For example, before the lymphoma phenotype was evident in the preleukemic stage, one of the earliest genes to be upregulated was *s100a11* (Fig. 5A, top), which overexpression has been reported to promote cell proliferation, antiapoptotic, and tumor metastasis in various cancers<sup>48,49</sup>. Interestingly, several genes that have been known to be involved in T-cell tumorigenesis, including *mycb* and *gata3*, the orthologues of human *MYC* and *GATA3*, were specifically upregulated in tumor samples (Fig. 5A, middle). On the other hand, multiple metabolic enzymes (e.g., *aldoaa*) were significantly upregulated in late-stage tumors with the *p53* mutation (Fig. 5A, bottom).

To further analyze molecular mechanism, we also performed a differential epigenetic profiling by chromatin-immunoprecipitation-sequencing (ChIP-seq) analysis for an active histone mark (H3K27ac) for multiple bulk tumor and normal thymus samples and performed a differential binding analysis (Supplementary Data 2). Density plots and metagene plots (Fig. 5B) represent three groups of regions: i) specifically activated in tumors (Group I), ii) commonly activated both in tumors and normal thymus (Group II), and iii) specifically activated in the normal thymus (Group III), based on statistical analysis (FDR  $< 5\%$  &  $\log_2$  fold-change  $> 1$ ). Regions activated in tumors (Group I + II) were enriched for DNA binding motifs recognized by IRF family transcription factors (Supplementary Fig. 5A). In fact, Group I included *mycb* and *gata3*, which were associated with high level of H3K27ac signals in multiple tumor samples but not in the normal thymus (Supplementary Fig. 5B). We also identified super-enhancers, clusters of active enhancers, which are often present at genes that characterize cell type<sup>50,51</sup>. This analysis demonstrated that many genes involved in T-cell development, including *bcl11ba*, *ikzf1*, *lck*, *runx3* and *tcf7*, were associated with super-enhancers in multiple tumors and normal thymus samples (Supplementary Fig. 5C and Supplementary



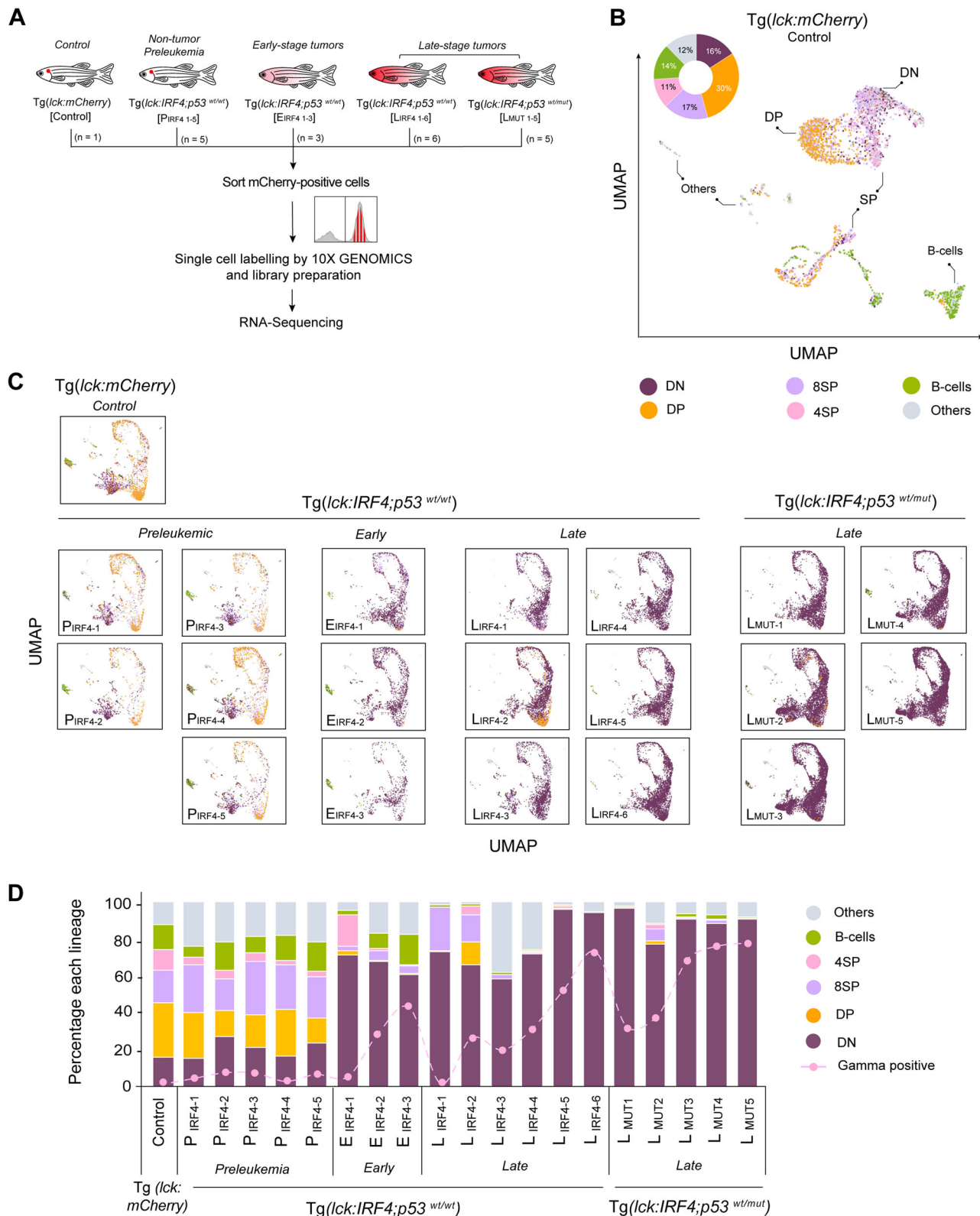


**Fig. 2** *IRF4*-driven zebrafish tumors recapitulate invasive human T-cell lymphoma. **A** Histopathological examination of H&E-stained tissue sections of representative samples from control *Tg(lck:mCherry;p53<sup>wt/wt</sup>)* ( $n = 8$ ), *Tg(lck:IRF4;p53<sup>wt/wt</sup>)* ( $n = 8$ ) and *Tg(lck:IRF4;p53<sup>wt/mut</sup>)* ( $n = 4$ ) zebrafish at 5 or 8 months. Similar findings were observed in multiple independent animals as shown for each sample. Tumor cells are indicated by white or black arrowheads. Scale bar = 100  $\mu\text{m}$ . **B** Transverse sections of representative samples from *Tg(lck:mCherry;p53<sup>wt/wt</sup>)* ( $n = 8$ ) and *Tg(lck:IRF4;p53<sup>wt/wt</sup>)* ( $n = 8$ ). F, fin; G, gill; M, muscle; S, skin; and SC, spinal cord. Similar findings were observed in multiple independent animals as shown for each sample. Scale bar = 2 mm. **C** High magnification of skin lesions from *Tg(lck:IRF4;p53<sup>wt/wt</sup>)* fish. Scale bar = 100  $\mu\text{m}$ .

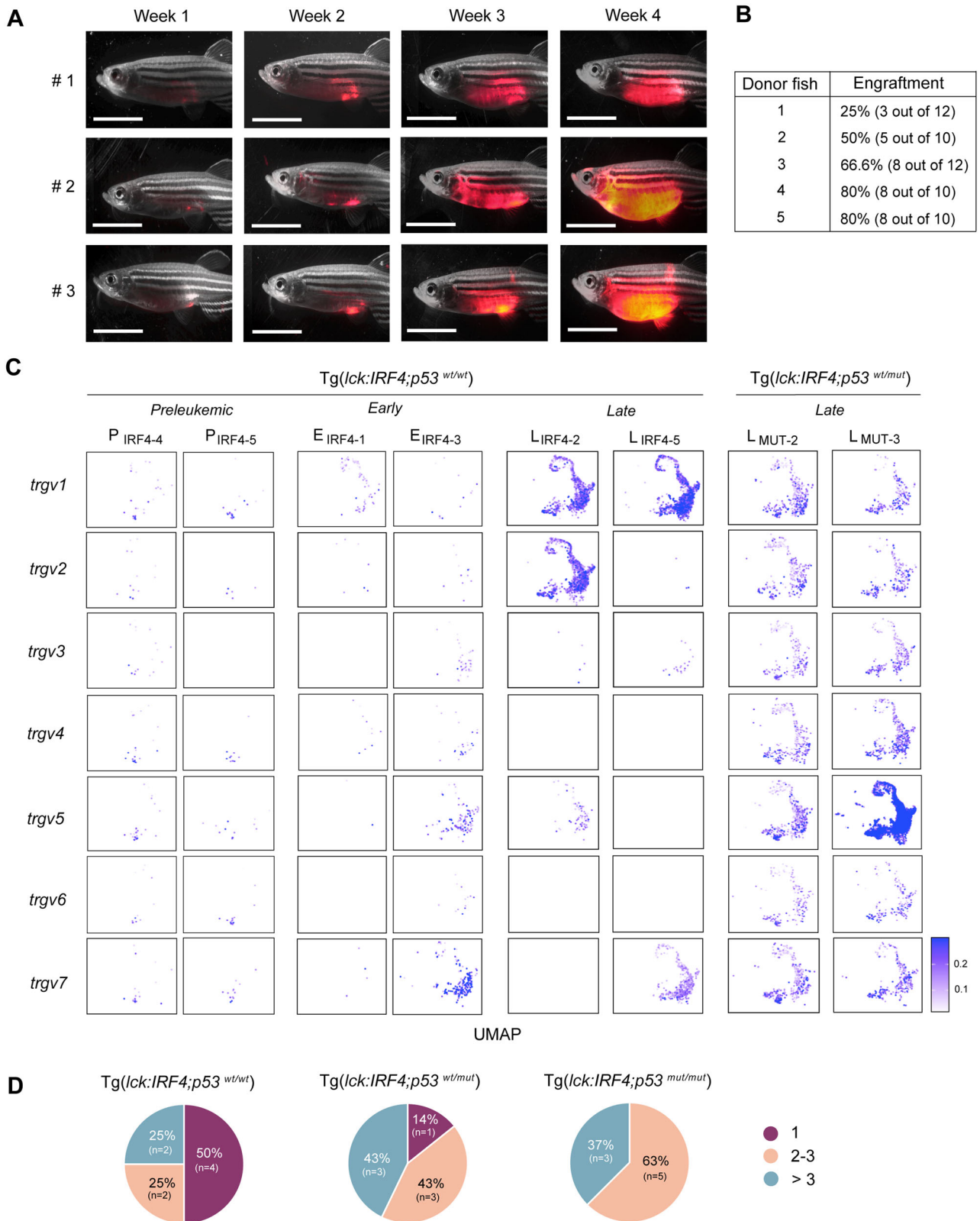
Data 3). Notably, super-enhancers at *mycb/fam49ba*, *gata3* and PI3K kinase subunit genes (*pik3cd* and *pik3c2b*) were found to be more frequent in tumors ( $\geq 3$  out of 4 tumors, and  $1 \leq$  out of 3 thymus samples), suggesting their requirement in tumorigenesis (Supplementary Fig. 5D). We performed independent validation by

qRT-PCR of 20 additional samples, which demonstrated the upregulation of *mycb* and *gata3* (Supplementary Fig. 5E).

Notably, besides *mycb* and *gata3*, eight genes (*h2afx*, *lrrn1*, *midn*, *ncaldb*, *patl1*, *psip1a*, *sec22ba*, and *sla2*) which were not well studied in cancers, were commonly selected by gene

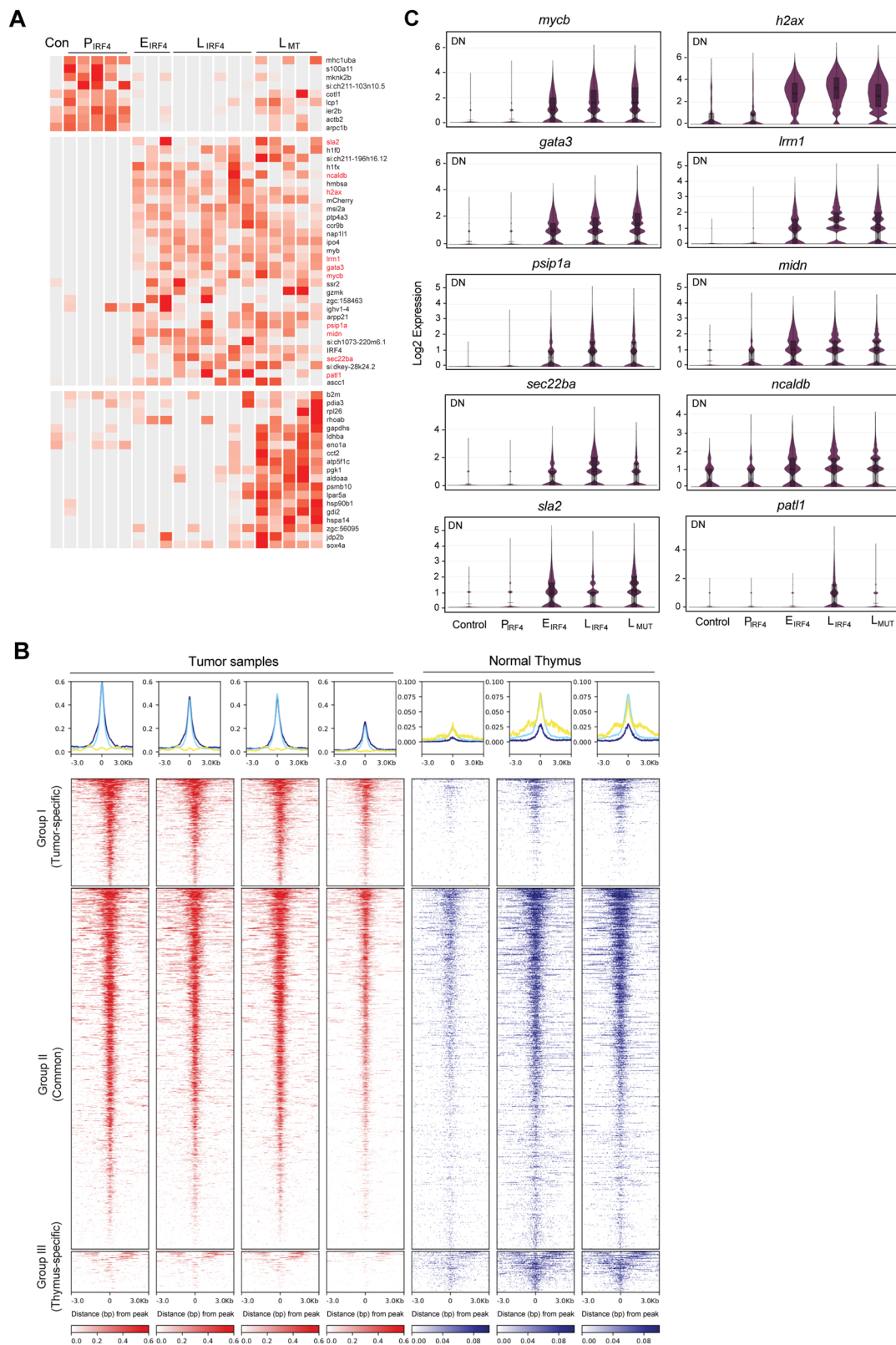


**Fig. 3 Identification of multiple cell populations derived from different stages of lymphocyte differentiation.** **A** Schematic diagram of the sample preparation for scRNA-seq analysis. **B** UMAP plot showing T and B lymphocytes from one control zebrafish. We first selected T and B-cell clusters and then performed a subclustering analysis. Six populations were then defined based on marker gene expression. **C** UMAP plots showing aggregated cells from 20 independent zebrafish. Six populations were annotated using the same criteria used in **(B)**. **D** Bar charts showing the percentage of different cell populations in each sample. Source data are provided as a Source Data file.



**Fig. 4 IRF4-driven tumor cells possess malignant features.** **A** Representative microscopy images of three recipient fish at 1, 2, 3, and 4 weeks after the first transplantation. Panels show merged fluorescence and brightfield images. Scale bar = 4 mm. **B** Table showing the engraftment rate of five independent donor tumor fish. **C** Gene expression pattern of the *tcr-γ* variable regions in eight representative samples. The cells in the UMAP plot are colored by the marker gene expression level. **D** Pie charts showing the percentage of mono-, oligo-, and polyclonal tumors analyzed by qRT-PCR in three different genotype settings. Source data are provided as a Source Data file.





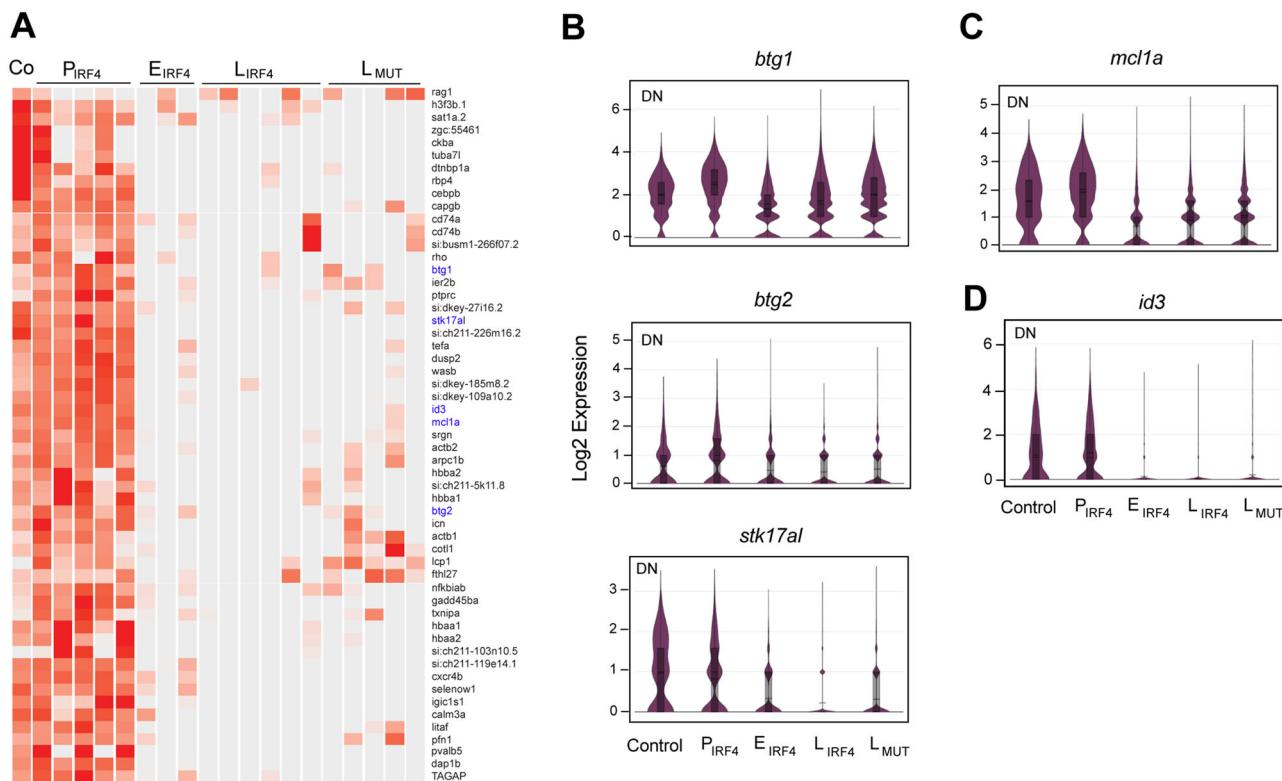
expression and differential epigenetic analyses (Fig. 5A, middle; Fig. 5C; Supplementary Fig. 5B). In particular, expression of *lrn1*, *pat1* and *psip1a* were highly specific to tumor cells; 0% of normal thymus and less than 3% preleukemic samples were positive for all three genes, while up to 46% of tumor samples were positive (Supplementary Fig. 5F). The orthologues of five genes, *LRRN1*, *PSIP1*, *PATL*, *H2AFX*, and *MIDN*, were significantly upregulated

in human  $\gamma\delta$  T-cell lymphoma (GD-TCL) samples as compared to normal  $\gamma\delta$  T-cells (Supplementary Fig. 5G)<sup>52</sup>, suggesting their involvement in T-cell lymphomagenesis.

**Downregulation of p53 targets and upregulation of anti-apoptotic genes.** We next focused on genes which were significantly downregulated ( $p$ -value < 0.05) in the DN population of



**Fig. 5 IRF4-driven tumors are characterized by high expression of *myc*b and *gata*3.** **A** Heatmap showing the significantly upregulated genes from one-to-one differential gene analysis of scRNA-seq data between DN populations of different tumor stages. **B** Density plots (bottom) show the distribution of H3K27ac signals in tumor and normal thymus samples. All significant peaks (FDR < 0.05) were first detected and then classified into three classes of regions (Groups I-III). The color scale represents the intensity of signals and distances (3,000 bps) from the center (0) of peaks. Metagene plots (top) show the distribution of H3K27ac signals from the center for Group I ( $\log_2FC \geq 1$ ) (blue), Group II ( $-1 < \log_2FC < 1$ ) (light blue) and Group III ( $\log_2FC \leq -1$ ) (yellow) regions. **C** Violin plots showing the expression of ten genes in DN cell populations across tumor stages analyzed by scRNA-seq (total cell number after merging of biologically independent samples: Control  $n = 557$  cells;  $P_{IRF4} n = 2,222$  cells;  $E_{IRF4} n = 3,365$ ;  $L_{IRF4} n = 15,280$  cells;  $L_{MT} n = 28,205$  cells). Black vertical line indicates max and min values. Box indicates the third and first quartiles. Horizontal lines in the box indicate mean and median values. Source data are provided as a Source Data file.



**Fig. 6 Downregulation of p53 targets and *id3*.** **A** Heatmap showing the significantly downregulated genes from one-to-one differential gene analysis of scRNA-seq data between DN populations of different tumor stages. **B-D** Violin plots showing the expression of *btg1*, *btg2* and *stk17a1* (**B**), *mcl1a* (**C**), *id3* (**D**) in DN cell populations across tumor stages analyzed by scRNA-seq (total cell number after merging of biologically independent samples: Control  $n = 557$  cells;  $P_{IRF4} n = 2,222$  cells;  $E_{IRF4} n = 3,365$ ;  $L_{IRF4} n = 15,280$  cells;  $L_{MT} n = 28,205$  cells). See Fig. 5C legend for the details of violin plots. Source data are provided as a Source Data file.

tumor samples (Fig. 6A). This list included genes associated with tumor suppressor and apoptotic pathways. Several genes known to be downstream of p53 pathway, namely *btg1*, *btg2*, and *stk17a1*<sup>53-55</sup>, were downregulated to an indistinguishable level with that in tumor cells with p53 mutation (Fig. 6A, top; Fig. 6B), suggesting that p53 activity might be lost during tumorigenesis even in the wild-type setting. However, p53 gene was still expressed and also not mutated in these samples based on PCR and sequencing analysis, indicating that p53 might be functionally suppressed via a different mechanism. It is also noteworthy that one of antiapoptotic BH3 family genes, *mcl1a*, was also downregulated in early-stage tumors although its expression slightly recovered in late-stage of tumors (Fig. 6C). This suggested that pro-apoptotic signal might be dominant and thus the tumor cells underwent apoptosis at the early stage while they became resistant to apoptosis at the late stage.

**IRF4 overexpression affects cell differentiation and lineage choice.** Interestingly, one of the most significantly downregulated

genes in DN cells was *id3*, which was expressed in preleukemic samples but almost completely suppressed in tumor samples (Fig. 6A, top; Fig. 6D). This gene was also found in Group III by ChIP-seq analysis (Fig. 5B; Supplementary Fig. 6A). Previous reports showed that conditional knockout of *id3* in mice under *Lck-Cre* impaired  $\gamma\delta$  T-cell development and led to the expansion of the immature stage of  $\gamma\delta$  T-cells<sup>56,57</sup>, similar to the phenotype observed in our zebrafish model. Additionally, we performed one-to-one differential gene analysis of the DP, 4SP, and 8SP between each tumor stage. This analysis revealed that although these populations do not undergo abnormal expansion like DN cells, their expression profiles still deviated from the normal pattern. For example, the expression level of *cd4-1*, *cd8a/b*, *rorc*, and *rag1*, which are normally upregulated in the DP stage<sup>58</sup>, was lower in the tumor samples (Supplementary Fig. 6B). Of note, a forced overexpression of IRF4 in an IRF4-negative human T-ALL cell line (HPB-ALL), which exhibits DP stage immunophenotype, significantly inhibited cell growth and induced apoptosis (Supplementary Fig. 6C, D), suggesting that IRF4 overexpression at

this stage of thymocytes is potentially lethal and thus IRF4-positive DP cells cannot expand. These results indicated that although cells are differentiating in the  $\alpha\beta$ -lineage continuum, *IRF4* overexpression interfered with the differentiation process.

**IRF4-driven T-cell tumors are sensitive to a small-molecule BRD inhibitor.** Lastly, we tested a potential therapeutic option in our zebrafish lymphoma model. Gene expression analysis indicated that *mycb* was predominantly upregulated in IRF4-driven T-cell tumors. Recently, accumulating studies show that *MYC*-overexpressing cancers are sensitive to small-molecule BRD inhibitors, such as JQ1<sup>59–61</sup>. BRD proteins interact with acetylated histones and promote transcriptional initiation and elongation processes. Genes highly activated in tumor cells are known to be strongly inhibited by JQ1 in various cancers<sup>60–64</sup>. Strikingly, the fish tumors showed a remarkable response after JQ1 treatment (Fig. 7A). The tumor volumes were significantly reduced compared to those of the control fish treated with DMSO (Fig. 7B). To further confirm if genes activated in tumors are sensitive to JQ1, we performed an RNA-seq analysis for three JQ1-treated and three DMSO-treated samples at Day3 (Supplementary Data 4). As expected, many of Groups I and II genes that represent genes highly activated in tumor cells (from Fig. 5B) were downregulated by JQ1 treatment, as demonstrated by the gene set enrichment analysis (Fig. 7C). In fact, seven of 10 shortlisted genes (from Fig. 5A, red) were significantly downregulated after the treatment (Fig. 7D), further suggesting that their expressions are associated with tumor maintenance. Together, our data indicated that the IRF4-driven tumors depended on the transcriptional activity and also demonstrated the applicability of our transgenic zebrafish model for drug testing.

## Discussion

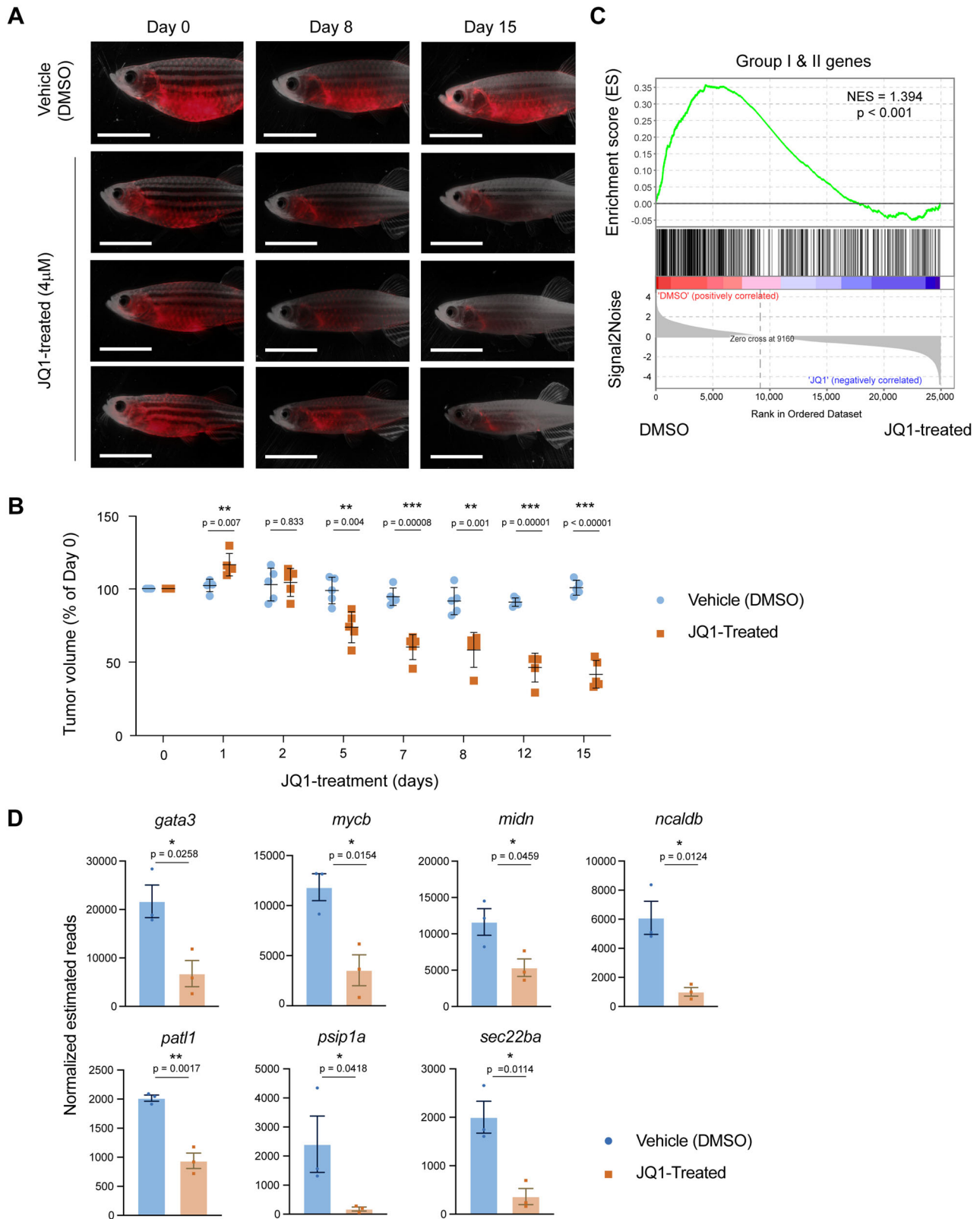
IRF4 has long been implicated as an oncogene in lymphoid malignancies. Since this gene was cloned at the breakpoint of a chromosomal translocation, t(6;14)(p25;q32), involving the *IgH* gene locus in MM cells<sup>65</sup>, a plethora of in vitro and clinical studies have reported the abnormal overexpression and oncogenic roles of IRF4 in various mature lymphoid neoplasms<sup>13–16,18–22,25</sup>. IRF4 directly induces *MYC* expression and also coordinately regulate the transcription program with NF- $\kappa$ B pathway<sup>19,22</sup>. However, the oncogenicity of *IRF4* in vivo has not been reported.

Here, we demonstrated that overexpression of wild-type IRF4 can cause the development of aggressive T-cell lymphoma. Although our model resembled human GD-TCL based on TCR- $\gamma$  expression, it also possesses several common features that are generally observed in mature T-cell neoplasms, such as high level of expression of *GATA3* and *MYC*, and the infiltration of tumor cells into the surrounding tissues and distal organs, in particular, cutaneous regions. This finding was different from the phenotype reported for the *rag2-Myc*-induced zebrafish T-ALL model, in which tumor cells typically spread through the vessels and showed the expansion of DP T-cells<sup>66–68</sup>. Of note, a previous study reported that overexpression of IRF4 under the *E $\mu$*  promoter did not cause tumors in mice<sup>69</sup>. Similarly, the *rag2*-promoter-driven *IRF4*-transgenic zebrafish did not cause any tumors in our study. This discrepancy is likely due to the different promoter systems that can be activated in different stages and lineages of lymphocytes. In general, *rag2* promoter is more active in the DP stage of T-cells and our study also showed that *IRF4* overexpression results in downregulation of *rag1/2* expression. The *E $\mu$*  promoter is more active in immature B-cells, as shown in a *MYC*-induced lymphoma model using the same promoter that predominantly developed pre-B lymphomas and mixed pre-B and

B-cell lymphomas<sup>70</sup>. In contrast, the zebrafish *lck* promoter is active both in immature precursor lymphocytes and mature T-cells. LCK is also an immediate downstream target of the TCR, while IRF4 mediates the T-cell activation process through TCR signaling<sup>7,9–11</sup>. Thus, there is a positive feed-forward loop between IRF4 and TCR-LCK. Additionally, the kinetics and expression pattern of *IRF4* expression are positively correlated with the strength of TCR signaling, which in turn, affects normal T-cell population expansion<sup>10,11,71</sup>. Therefore, a specific cellular context that possesses active TCR signaling including LCK expression may be required for the full oncogenic ability of IRF4 in vivo.

However, the expansion of DN T-cells, many of which expressed TCR- $\gamma$ , was initially unexpected. Our analysis suggested that tumor cells are likely derived from the intermediate stage of DN T-cells that can generate both  $\gamma\delta$  T-cells and  $\alpha\beta$  T-cells. Although this phenotype may be specific to the experimental setting, our study provides an important implication: overexpression of IRF4 in the DN stage affects cell differentiation and lineage choice. A potential reason for this outcome can be explained by several consequential factors. First, our and other studies indicated that IRF4 transcriptionally upregulates several molecules downstream of the TCR signaling pathway so that the cells can respond strongly upon TCR activation. It has been known that the lineage fate of T-cells ( $\alpha\beta$  vs  $\gamma\delta$ ) can be influenced by the strength of TCR signaling at DN stage, in which stronger signal favors the development of the  $\gamma\delta$  lineage<sup>72,73</sup>. Thus, overexpression of IRF4 prior to this stage might affect the lineage choice. Similarly, several genes responsible for T-cell differentiation such as *rag1* and *id3* were downregulated in DN stage of tumor samples. Thus, IRF4-overexpressing cells may not be able to further rearrange TCR- $\alpha$  or  $\beta$  and instead be committed to  $\gamma\delta$  lineage. Second, the DP stage of T-cells normally undergo positive and negative selections, by which the cells that lack or have excess activity of TCR signaling undergo apoptosis. In a physiological condition, IRF4 expression is strictly restricted to mature T-cells, which have successfully passed the DP checkpoint. Hence, ectopic overexpression of IRF4 in DP stage may cause excess TCR signals, which potentially triggers apoptosis at this stage of thymocytes development. Therefore, IRF4 overexpression at this stage could be lethal. The fact that *rag2* promoter-driven system does not cause tumors, but *rag2-myc* does also supports this possibility. This finding also supports the fact that IRF4 has not been implicated as an oncogene in T-ALL, which mainly proliferate at late DN and DP stage of T-cells. In fact, overexpression of IRF4 in DP stage T-ALL cell lines caused apoptosis. Similar finding has been reported in B-cell acute lymphoblastic leukemia and chronic lymphocytic leukemia cases, which implicated a tumor-suppressive role of IRF4<sup>23,74–76</sup>. Lastly,  $\gamma\delta$  T-cells are relatively more frequent in amphibians, reptiles or teleost fish, including zebrafish; and the intermediate stage of DN T-cells express TCR- $\gamma$  alone as previously reported<sup>42</sup>. Thus, high activation of TCR- $\gamma$ -mediated signaling at the intermediate stage of DN T-cells might result in the preferential expansion of this type of T-cells.

Additionally, we observed that IRF4-driven tumor cells exhibited several important malignant features. Pre- or malignant cells undergo clonal competition during tumorigenesis, which ultimately leads to mono- or oligoclonal proliferation of T-cells. These cells were able to be engrafted and survive for a long time in the recipient animals, suggesting the existence of a leukemia-initiating cell population. Furthermore, the mutation of *p53* confers an additional advantage, by which the cells can resist apoptotic cell death and thus collaborate with IRF4 overexpression. In this setting, the clones may bypass the severe competition and thus exhibit tumor heterogeneity. Our study also demonstrated that heterozygous mutation of *p53* in the



transactivation domain is able to accelerate tumorigenesis. It has been reported that the mutant p53 functions as a gain-of-function by interacting with other transcription factors, besides the loss of wild-type protein<sup>33,77</sup>. We would like to investigate these aspects in a future study.

In conclusion, we established an *IRF4*-driven zebrafish lymphoma model. Our results with a small-molecule BRD inhibitor

also demonstrated the feasibility of this model for drug testing. Thus, faced with the pressing need to understand the pathogenesis and find effective therapies for T-cell lymphoma, our work provides the necessary animal model system, which would be a useful platform for further mechanistic studies as well as for testing therapeutic options to improve the treatment of this disease.



**Fig. 7 IRF4-driven T-cell tumors are sensitive to a small-molecule BRD inhibitor. A** Representative microscopy images of Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) transgenic fish treated with vehicle (DMSO;  $n = 5$ ) or JQ1 (4  $\mu\text{M}$ ;  $n = 5$ ). Panels show merged fluorescence and brightfield images. Scale bar = 4 mm. **B** Quantification of tumor volume for five independent animals on days 1, 2, 5, 7, 8, 12 and 15 compared to that on day 0. Integrated fluorescence intensity was quantified by ImageJ (NIH) software and normalized to body size; \* $p < 0.05$ , \*\* $p < 0.01$ , and \*\*\* $p < 0.001$  by two-tailed unpaired Student's  $t$ -tests. Bars indicate mean and SD values. Day 1  $p = 0.07$ ; day 2  $p = 0.83$ ; day 5  $p = 0.04$ ; day 7  $p = 7.7 \times 10^{-5}$ ; day 8  $p = 0.001$ ; day 12  $p = 1 \times 10^{-5}$ ; day 15  $p = 1.7 \times 10^{-5}$ . **C** Gene set enrichment analysis plots showing the overall pattern of gene expression changes of Group I + II genes upon JQ1 treatment. Each solid bar represents one gene within the gene set. The normalized enrichment scores (NES) by dataset permutations and  $p$ -values are indicated. **D** Bar charts showing the expression level of seven selected genes in control and JQ1-treated samples ( $n = 3$  independent animals for each group). The data represent the mean  $\pm$  SEM (shown as bars) of replicate samples; \* $p < 0.05$  and \*\* $p < 0.01$  by two-tailed unpaired Student's  $t$ -tests. Source data are provided as a Source Data file.

## Methods

**Zebrafish studies.** All zebrafish protocols (BR17-0351, R17-0353, BR18-1349, R18-1350, BR18-0526, R19-1182 BR21-0082, R21-0084) were approved by the Institutional Animal Care and Use Committee of the National University of Singapore and were performed according to their recommendations. The zebrafish *lck* promoter construct was kindly provided by Dr. Hui Feng (Boston University, Boston). Animals used in this study were of *Danio rerio* species with the specific strains explained, as follows. The *p53*-mutant zebrafish line (*M214K* mutant) was kindly provided by Dr. Thomas Look (Dana-Farber Cancer Institute, Boston)<sup>28</sup>. An immunocompromised strain (*rag2<sup>E450fs</sup>* mutant) was established and kindly provided by Dr. David Langenau's laboratory<sup>78</sup>. We established other transgenic animals using the Singapore wild-type strain, as follows. The human *IRF4* or *mCherry* gene was cloned downstream of the *lck* or *rag2* promoter into a backbone vector. The plasmids were resuspended in buffer and linearized using the I-SceI enzyme (New England Biolabs). The solution was injected into one-cell-stage zebrafish embryos to establish the Tg(*lck:IRF4*) line. Tg(*lck:mCherry*) animals injected with only *lck-mCherry* were established as a control line. Tg(*rag2-IRF4*) was also established as comparison for Tg(*lck:IRF4*).

**Tumor monitoring.** For the tumor monitoring study, brightfield and fluorescent images were captured using a Nikon Stereomicroscope SMZ800N and DS-Ri2 camera. Image acquisition was performed using NIS-Elements D software. Zebrafish of mixed gender were monitored from 2 weeks of age weekly for 1 year or until the animals died whichever was earlier. Tumor monitoring was performed in F0 animals, control Tg(*lck:mCherry*) ( $n = 7$ ) and Tg(*lck:IRF4*) transgenic ( $n = 21$ ). Tumor monitoring was performed in F1 animals, control Tg(*lck:mCherry;p53<sup>wt/wt</sup>*) ( $n = 30$ ), Tg(*lck:mCherry;p53<sup>wt/mut</sup>*) ( $n = 25$ ), Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) ( $n = 79$ ), Tg(*lck:IRF4;p53<sup>wt/mut</sup>*) ( $n = 32$ ) and Tg(*lck:IRF4;p53<sup>mut/mut</sup>*) ( $n = 21$ ), and Tg(*rag2:IRF4*) ( $n = 13$ ). Early-stage tumors were defined by the expansion of mCherry-positive cells outside the thymus, including in the anterior musculature, orbit, nasal placode, gills, and pectoral fins. Tumors were considered late-stage when mCherry-positive cells had spread to more than 50% of the zebrafish body.

**Drug treatment.** The small-molecule BRD inhibitor (JQ1) was purchased from Selleck Chemicals and dissolved in DMSO. Zebrafish of mixed genders with the almost same tumor burden were treated with JQ1 at 4  $\mu\text{M}$  ( $n = 5$ ) or with vehicle (DMSO) ( $n = 5$ ) in individual tanks. Zebrafish were treated with JQ1 at 4  $\mu\text{M}$  in individual tanks. The treatment was conducted for the total duration of 15 days. Water containing JQ1 was changed every 48 h.

**Transplantation assay.** Tumor cells ( $n = 100,000$ ) were transplanted into the peritoneal cavity of 3-month-old *rag2<sup>E450fs</sup>* mutant fish of mixed gender, as previously described<sup>78</sup>.

**Histopathology.** Adult zebrafish were euthanized, fixed in 10% neutral-buffered formalin, decalcified with 9% EDTA (pH 7.6), dehydrated in a graded ethanol series, cleared with Histo-Clear (Sigma Aldrich), and embedded in paraffin. After rehydration using Histo-Clear and a graded series ethanol, the section slides were stained using hematoxylin and eosin (H&E). High-magnification imaging was performed, and images were acquired using Axioplan 2 Imaging (Zeiss) and AxioVision software. Macroimages were taken using an Olympus SZX fluorescent stereo zoom microscope, and Olympus DP software was used for image acquisition.

**Flow cytometry.** Cells were prepared from wild-type adult zebrafish kidney and mCherry-sorted cells from the Tg(*lck:IRF4;p53<sup>wt/wt</sup>*) fish. Cells were gated on FSC-H and FSC-A plot to eliminate doublets and then on FSC versus SSC plot to obtain hematopoietic lineage profiles for erythrocytes, granulocytes and monocytes, lymphocytes, and blood cell precursors. FSC Forward scatter, FSC-H Forward scatter height, FSC-A Forward scatter area, SSC Side scatter.

**Table 1 Primer sequences used for genotyping analysis.**

Primer ID	Primer sequence (5' -->3')
IRF4 CDS forward	CACC 5'-CACCATGAACCTGGAGGGC-3'
IRF4 CDS reverse	5'-TCATTCTTGAATAGAGGAATGGCGG-3'
mCherry Agel forward	5'-AATACCGGTATGGTGAGCAAGGG-3'
mCherry Clal reverse	5'-CCATCGATCTACTTGTACAGCTCGT-3'
rag2 forward	5'-CTGCCGGATCTCCCATGGAC-3'
rag2 reverse	5'-TCATTCTTGAATAGAGGAATGGCGG-3'

**Genotyping.** Adult fish were anesthetized in MS-222 solution and fin clipping was performed. To extract genomic DNA, the fin tissue was lysed in 50  $\mu\text{L}$  of NaOH (50 mM) and incubated at 95  $^{\circ}\text{C}$  for 1 h in a thermocycler. The reaction was then neutralized with 5  $\mu\text{L}$  of Tris-HCl (1 M) and vortexed briefly. 1  $\mu\text{L}$  of genomic DNA or 100 ng of plasmid control were subjected to PCR amplification using Accuprime™ Pfx DNA polymerase (Thermo Fisher Scientific). The PCR cycle parameters were as follows: initial denaturation at 95  $^{\circ}\text{C}$  for 3 min, and 34 cycles of (i) denaturation at 95  $^{\circ}\text{C}$  for 30 s, (ii) annealing at 59  $^{\circ}\text{C}$  for 30 s, (iii) elongation at 72  $^{\circ}\text{C}$  for 1 min 30 s, followed by a final elongation step at 72  $^{\circ}\text{C}$  for 5 min. The primers used for genotyping are shown in Table 1.

**RNA extraction and quantitative reverse-transcription PCR (qRT-PCR).** Total RNA was isolated from the sorted mCherry cells from each transgenic zebrafish using a NucleoSpin RNA kit (Macherey-Nagel). RNA was reverse-transcribed using EvoScript Universal cDNA Master Mix (Roche). qRT-PCR was performed on a QuantStudio 5 Real-Time PCR system (Thermo Fisher Scientific) using SYBR Green PCR master mix (Applied Biosystems) and analyzed using QuantStudio™ Design and Analysis software version 1.2. The primer sequences are shown in Table 2.

**Cell line and overexpression experiments.** A human cell line (HPB-ALL) was purchased from DSMZ and used from the stock confirmed by DNA fingerprinting using the PowerPlex 1.2 system (Promega) in January 2013. The cells were cultured in Roswell Park Memorial Institute (RPMI)-1640 medium (BioWest) supplemented with 10% fetal bovine serum (FBS) (BioWest). 293 T cells were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% FBS, penicillin (100 U/mL), and streptomycin (100 U/mL). All cells were cultured in a humidified incubator containing 5% CO<sub>2</sub> at 37  $^{\circ}\text{C}$ . For gene overexpression, the full-length human *IRF4* sequence were cloned into a doxycycline-inducible pCW57.1 lentiviral vector (Addgene). Cells were transduced with lentivirus produced by cotransfection into 293 T cells with pMDLg/pRRE, pRSV-Rev, and pMD2.G using FuGENE6 (Promega) and Opti-MEM (Thermo Fisher) reagents.

**Protein extraction and Western blotting.** FACS-sorted cells were homogenized in lysis buffer containing 20 mM Tris-HCl (pH 7.5), 150 mM NaCl, 1 mM Na<sub>2</sub>EDTA, 1 mM EGTA, 1% Triton, 2.5 mM sodium pyrophosphate, 1 mM  $\beta$ -glycerophosphate, 1 mM Na<sub>3</sub>VO<sub>4</sub>, and 1  $\mu\text{g}/\text{ml}$  leupeptin (Cell Signaling Technology) supplemented with 1% Halt protease inhibitor cocktail (Thermo Fisher Scientific). Protein concentrations were then quantified using a Bio-Rad protein assay kit (Bio-Rad Laboratories). A total of 15  $\mu\text{g}$  of each protein lysate from mCherry-positive cells of the Tg(*lck:IRF4*) fish were mixed with 4 $\times$  Laemmli sample buffer (Bio-Rad Laboratories) and  $\beta$ -mercaptoethanol (Sigma-Aldrich) in appropriate proportions before being incubated at 99  $^{\circ}\text{C}$  for 5 min. The samples were then separated on SDS-PAGE gels and transferred to PVDF membranes (Millipore) at 400 mA for 1.5 h, blocked with 3% skim milk (Nacalai Tesque) in Tris-buffered saline with 0.1% Tween and probed with IRF4 antibody (1:1000; Santa Cruz), cleaved PARP antibody (1: 1000; Cell Signaling Technology) and GAPDH-HRP antibody (1: 1000; Santa Cruz). Secondary detection was performed with HRP-linked antirabbit IgG (Cell Signaling Technology) and ECL™

**Table 2 Primers used for RT-PCR.**

Genes	Species	Forward primers	Reverse primers
IRF4	Human	GGCCAGCTTGTGAAAATGG	TTATGCTTGGCTCTGTGGGG
trac	Zebrafish	CTTAAAACGTCGGCTGTCCG	TGAACAAACGCTGTCTCCT
trbc	Zebrafish	CTCCGGAAAAAGTCACACTTG	AAGATGACAAGGCCATACAGTC
trgc	Zebrafish	TAGTAACTGAACCTGGGAAGGAC	TTAAGAGCGCTCTCTTTCTTTCC
trdc	Zebrafish	GTGGCCGCCGATTCTTTCTCA	TTTGTGGATGGTGGGGTGGTAGT
cd4-1	Zebrafish	TCAACACCAAGACCATCAGC	GCACATGTCCATTTACCTC
cd8a	Zebrafish	TCGGAGGTTGTGGACTTTTC	TTGTAATGGTGGGGACATCG
gata3	Zebrafish	GGTGAGATGTAGGGAGAGGAAACC	TGCCAAGACCTATAACACATCCA
mycb	Zebrafish	AAGCGCCAAAGTGGTGATCC	CACTACTTTGCCACACCTCCTCGC
trgc	Zebrafish		GAGCACAGCCACAATGAAGA
trgv1	Zebrafish	GCCCATCAAGACTGCTTTA	
trgv2	Zebrafish	AGATGGTGCTGTTGTTCTGTGA	
trgv3	Zebrafish	GCGTTCTGGGAGTGACTTTG	
trgv5	Zebrafish	CTCAGAAGGTTTGTGAAGC	
trgv6	Zebrafish	ACCAAATCTCAGGACAAGACTG	
trgv7	Zebrafish	ATGCACAAATTTAGGAGGAGA	
si:dkeyp-13d12.11	Zebrafish	ATCAAATGCCAAGTGGATGC	
si:dkeyp-13d12.24	Zebrafish	ATCAAATGGACCGTGAATC	
midn	Human	AGAGCCAGATCCGCATGTGCAA	TTCTACGGAGCCGTTTCTGCTG
pat1	Human	GCCAAACTGGAGCACGCCTATA	CACTCCGAGATGTCAACACAGC
psip1	Human	AGGCAGGAGTAGTGACAACAGC	CTCTCTGAAGGACAGGGCTGTT

Western blotting detection reagents (Fisher Scientific). Western blot images were captured using an ImageQuant LAS500 chemiluminescent image analyzer (GE Healthcare).

#### Single-cell RNA sequencing (scRNA-seq)

**Labeling and library preparation.** Samples were prepared according to the 10x Genomics Single Cell 3' v2 Reagent Kit user guide as previously described<sup>79</sup>. Briefly, mCherry-positive cells were first sorted by flow cytometry from the whole body of leukemic zebrafish. Cell viability was evaluated using trypan blue and a hemocytometer. Samples were prepared for target capture of 6,000 cells. Gel beads in emulsions (GEMs) were prepared by combining gel beads containing barcoded oligonucleotides with partitioning oil and were loaded onto the microfluidic chip. Then, the gel beads were dissolved, and cDNA libraries were constructed as described by the Single Cell 3' Reagent Kit v2 user's guide with appropriate modifications to the PCR cycle based on the calculated cDNA concentration.

**Sequencing.** The final library was amplified and sequenced using the P5 and P7 primers in an Illumina HiSeq at Novogene AIT Genomics Pte Ltd. The library was sequenced in a paired-end and non-strand-specific manner with a 150-bp read length.

**Single-cell data analysis.** Single-cell sequencing data were processed using the 10x Genomics software package Cell Ranger v3.1.0. Flow of data analysis is shown in Supplementary Fig. 7. The reads were aligned to the GRCz11/danRer11 genome, and *mCherry*, human *IRF4* (ENSG00000137265), the TCR- $\alpha$  constant (*trac*) and  $\beta$  constant (*trbc*) genes reported in Wan et al.<sup>42</sup> were manually included in the annotation. Constructs generated using the consensus *v-j-c* sequence from zebrafish *trgv1-trgv7* reported by Beetz and Steiner in Nucleotide Database (NUCCORE: <https://www.ncbi.nlm.nih.gov/nuccore>) were also manually included. Zebrafish transcriptome data were generated by filtering the Ensembl annotations for protein-coding immunoglobulin and TCR genes that were defined in the GRCz11 GTF file. Individual read counts were performed using the Cell Ranger "count" function, which counted the reads based on the number of observed cell barcodes and generated a matrix for each cell and count. Then, all the samples ( $n = 20$ ) were combined into a single aggregated version to determine gene expression using both the Cell Ranger "aggr" function and Seurat R package v3.0<sup>80</sup> and were visualized using Loupe Browser v5.0 software. QC metrics showing the number of cells after each filtering were shown in Supplementary Data 5.

**Data integration.** Doublet removal was performed by integrating the outputs of DoubletFinder<sup>81</sup> and DoubletDecon<sup>82</sup>. The R package Seurat v3.0 was used to filter, integrate and merge the data<sup>83</sup>. Each sample was filtered with gene features >500 and mCherry expression  $\geq 1$  UMI. The anchors of each sample were determined by assigning each anchor a score and constructing an overall neighbor graph and computing their SNN graph using the FindIntegrationAnchors function ( $n = 3,000$ ). Later, the distance between datasets were defined by the number of anchors between datasets, the pairwise distance between datasets were computed

and clustered to determine a guide tree. These integrated multiple datasets ( $n = 20$ ) into a single merge Seurat object by using the IntegrateData function. Principal component analysis (PCA) was performed using the RunPCA function with 20 dimensions ( $\text{dims} = 20$ ), followed by uniform manifold approximation and projection (UMAP) visualization of the PCA data using the RunUMAP function. The annotation was firstly done using the control fish scRNA-seq by exploring the marker genes for each population. Cell populations were defined based on the marker gene expression (Supplementary Data 1). Because the expression pattern of hematopoietic cell markers can be different among species, we used only the genes that have been reported to be expressed in zebrafish<sup>37–41</sup>. Then, the same marker genes were applied to the rest of the individual tumor sample to classify their cell population. The single-cell plot were visualized using the DimPlot function on the UMAP reduction, and the gene expression plot was visualized using the FeaturePlot function. All these functions are available in the Seurat R package v3.0. Violin plots, and gene expression images were generated using Loupe Browser v5.0. All TCR- $\gamma$  variable genes (*trgv1-trgv7*) were calculated using AddModuleScore by summing the manually constructed *v-j-c* counts and visualized using FeaturePlot. All these functions are available in Seurat R package v3.0.

**Single-cell differential analysis.** Significant marker genes were selected by comparing one cluster to all the other clusters using Loupe Browser v5.0 with the 'locally distinguishing' method to measure the significant genes in each cluster. Loupe Browser uses the sSeq methods for differential expression analysis which employs the negative binomial exact test. Differential expression data were retrieved from the average expression of each gene per sample and the heatmap plot using ComplexHeatmap R package version 2.1.0<sup>84</sup>. The *p*-values were adjusted for multiple testing using the Benjamini-Hochberg procedure and were generated by the Loupe Browser.

**Chromatin-immunoprecipitation (ChIP)-sequencing (ChIP-seq).** Tumor cells and normal thymus samples at 3 months were sorted for mCherry-positive cells, crosslinked and processed for ChIP as described previously<sup>85</sup>. Sonicated chromatin was incubated with Dynabeads (Invitrogen) coated with H3K27ac (Abcam, #ab4729). Library construction and BGISEQ-500 sequencing of single-end, 50-bp reads were performed at BGI Biotech Solutions Co. Ltd. (Hong Kong). ChIP-seq reads were mapped to zebrafish (*Danio rerio* GRCz11) reference genome using Bowtie2 version v2.4.1 program<sup>86</sup> with default parameters. Aligned ChIP-seq reads were processed with samtools v1.12 rmdup package to remove duplicates. ChIP-Seq peaks were called with MACS2 software, version 2.1.1.2016030. Threshold for the broad peaks called is *P*-value <  $1E-3$  and the broad peaks cutoff is *P*-value <  $1E-3$  -broad-cutoff 1e-3 -p 1e-3 -broad -keep-dup 1 -f BAM -SPMR. Differential binding analysis was performed using DiffBind v3.0. The peaks of each sample were merged and consensus sites between all samples (4 tumor vs. 3 normal thymus) were compared using DESeq2 algorithm<sup>87</sup> in DiffBind R package<sup>88</sup>. The significant regions (FDR < 0.05) reported by DiffBind were selected and grouped into Group I ( $\log_2\text{FC} > 1$ ), Group II ( $1 < \log_2\text{FC} < -1$ ) and Group III ( $\log_2\text{FC} < -1$ ). 500 bp from the center of each significant region (250 bp to both ends) were extracted to perform motif enrichment analysis. Super-enhancers were identified by ROSE2

version 1.1.0 by merging peaks less than 12.5 kb and excluding signal from 2 kb of gene promoter reported in GRCz11 Ensembl annotation. Super-enhancers were called with the parameter `-s 12500 -t 2000 -mask hg19_blacklist.bed`. Closest genes were annotated for each super-enhancer and top gene were reported by ROSE2 annotation by analyzing each closest gene with highest promoter H3K27ac signal and selected to be the top gene. Motif enrichment analysis using meme-chip function from MEME Suite<sup>89</sup> by utilizing the vertebrates PWN from Transfac and HOCO-MOCO v11 motif database.

**Bulk RNA-seq.** Tumor fish were treated with JQ1 (4uM) or control DMSO for 3 days. Tumor cells were sorted by flow cytometry based on mCherry fluorescence expression. Total RNA was harvested using the miRNeasy Mini Kit (Qiagen). ERCC spike-in was added based on the total cell number. Strand-specific library construction and sequencing of paired-end, 100-bp-long reads by the BGISEQ500 were performed at the BGI Biotech Solutions (Hong Kong) Co Ltd (Hong Kong). chRNA-seq datasets of DMSO-treated control samples and JQ1-treated samples were aligned to the GRCz11 danRer11 genome Ensembl annotation with ERCC spike-in information using STAR 2.5.2a<sup>90</sup> with the parameter `outFilterMultimapNmax` set to 1. FeatureCount was used for the mapped reads in bam files to generate count tables based on the Ensembl gene annotation and ERCC annotation. Bioconductor R package DESeq2 v1.12.4<sup>87</sup> was used to analyze differential gene expression using 3 DMSO vs 3 JQ1-treated samples. RNA abundance quantification in GD-TCL patients and normal GD-T cells were quantified using the kallisto software version 0.45<sup>91</sup> with the following parameter `-b 30`.

**Gene expression in mouse hematopoietic cells.** The mRNA expression pattern in various stages and lineages of mouse hematopoietic cells were analyzed using Gene Expression Commons database<sup>92</sup>.

**Statistical analysis and reproducibility.** The significance ( $p$  values) of the tumor monitoring analysis was determined using the Gehan-Breslow-Wilcoxon test in GraphPad Prism software. For the significance of gene expression analysis of scRNA-seq data,  $p$  values adjusted for multiple testing using the Benjamini-Hochberg procedure were generated by the Loupe Browser. The significance of other analyses was determined using the Student's  $t$ -test in GraphPad Prism software.  $F$ -test was performed to analyze the variances of two populations, based on which one or two-tailed test was selected. A  $p$  value less than 0.05 was considered statistically significant. For microscopic and histopathological analyses, we showed representative images from multiple independent animals. Animal numbers are indicated in the legend.

**Reporting summary.** Further information on research design is available in the Nature Research Reporting Summary linked to this article.

## Data availability

The regular RNA-seq, ChIP-seq and scRNA-seq data generated in this study have been deposited in the Gene Expression Omnibus database: accession numbers [GSE139226](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE139226) (RNA-seq, publicly available), [GSE166644](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166644) (ChIP-seq, publicly available), [GSE166646](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166646) (scRNA-seq, publicly available), [GSE184946](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE184946) (RNA-seq, publicly available) under the super-series [GSE166650](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE166650). The dataset for human  $\gamma\delta$  T-cell lymphoma and normal  $\gamma\delta$  T-cells has been reported<sup>52</sup> and deposited under the dbGaP and GEO databases under accession number [phs001969](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=phs001969) (RNA-seq, publicly available) and [GSE107011](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE107011) (RNA-seq, publicly available). The dataset for *rag2-Myc* zebrafish has been reported<sup>47</sup> and deposited under GEO database under accession number [GSE108855](https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE108855) (RNA-seq, publicly available). Zebrafish genome and annotation version GRCz11 [[http://asia.ensembl.org/Danio\\_rerio/Info/Index](http://asia.ensembl.org/Danio_rerio/Info/Index)] is available from Ensembl database. Zebrafish raw images have been deposited in figshare (<https://doi.org/10.6084/m9.figshare.19497848>, publicly available) [<https://doi.org/10.6084/m9.figshare.19497848>]. Source data are provided with this paper.

Received: 22 April 2021; Accepted: 13 April 2022;

Published online: 03 May 2022

## References

- Mittrucker, H. W. et al. Requirement for the transcription factor LSIRF/IRF4 for mature B and T lymphocyte function. *Science* **275**, 540–543 (1997).
- Huber, M. & Lohoff, M. IRF4 at the crossroads of effector T-cell fate decision. *Eur. J. Immunol.* **44**, 1886–1895 (2014).
- Shaffer, A. L., Emre, N. C., Romesser, P. B. & Staudt, L. M. IRF4: Immunity. Malignancy! Therapy? *Clin. Cancer Res* **15**, 2954–2961 (2009).
- Sciammas, R. et al. Graded expression of interferon regulatory factor-4 coordinates isotype switching with plasma cell differentiation. *Immunity* **25**, 225–236 (2006).
- Gupta, S., Jiang, M., Anthony, A. & Pernis, A. B. Lineage-specific modulation of interleukin 4 signaling by interferon regulatory factor 4. *J. Exp. Med* **190**, 1837–1848 (1999).
- Marecki, S., Atchison, M. L. & Fenton, M. J. Differential expression and distinct functions of IFN regulatory factor 4 and IFN consensus sequence binding protein in macrophages. *J. Immunol.* **163**, 2713–2722 (1999).
- Lohoff, M. et al. Dysregulated T helper cell differentiation in the absence of interferon regulatory factor 4. *Proc. Natl Acad. Sci. USA* **99**, 11808–11812 (2002).
- Lu, R., Medina, K. L., Lancki, D. W. & Singh, H. IRF-4,8 orchestrate the pre-B-to-B transition in lymphocyte development. *Genes Dev.* **17**, 1703–1708 (2003).
- Brustle, A. et al. The development of inflammatory T(H)-17 cells requires interferon-regulatory factor 4. *Nat. Immunol.* **8**, 958–966 (2007).
- Man, K. et al. The transcription factor IRF4 is essential for TCR affinity-mediated metabolic programming and clonal expansion of T cells. *Nat. Immunol.* **14**, 1155–1165 (2013).
- Yao, S. et al. Interferon regulatory factor 4 sustains CD8(+) T cell expansion and effector differentiation. *Immunity* **39**, 833–845 (2013).
- Tsuboi, K. et al. MUM1/IRF4 expression as a frequent event in mature lymphoid malignancies. *Leukemia* **14**, 449–456 (2000).
- Chang, C. C. et al. Expression of MUM1/IRF4 correlates with clinical outcome in patients with B-cell chronic lymphocytic leukemia. *Blood* **100**, 4671–4675 (2002).
- Salaverria, I. et al. Translocations activating IRF4 identify a subtype of germinal center-derived B-cell lymphoma affecting predominantly children and young adults. *Blood* **118**, 139–147 (2011).
- Alizadeh, A. A. et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* **403**, 503–511 (2000).
- Yang, Y. et al. Exploiting Synthetic Lethality for the Therapy of ABC Diffuse Large B Cell Lymphoma. *Cancer Cell* **21**, 723–737 (2012).
- Wada, D. A. et al. Specificity of IRF4 translocations for primary cutaneous anaplastic large cell lymphoma: A multicenter study of 204 skin biopsies. *Mod. Pathol.* **24**, 596–605 (2011).
- Nakagawa, M. et al. Targeting the HTLV-I-Regulated BATF3/IRF4 Transcriptional Network in Adult T Cell Leukemia/Lymphoma. *Cancer Cell* **34**, 286–297.e210 (2018).
- Wong, R. W. J. et al. Feed-forward regulatory loop driven by IRF4 and NF-kappaB in adult T-cell leukemia/lymphoma. *Blood* **135**, 934–947 (2020).
- Feldman, A. L. et al. Recurrent translocations involving the IRF4 oncogene locus in peripheral T-cell lymphomas. *Leukemia* **23**, 574–580 (2009).
- Boddicker, R. L. et al. The oncogenic transcription factor IRF4 is regulated by a novel CD30/NF- $\kappa$ B positive feedback loop in peripheral T-cell lymphoma. *Blood* **125**, 3118 (2015).
- Weilemann, A. et al. Essential role of IRF4 and MYC signaling for survival of anaplastic large cell lymphoma. *Blood* **125**, 124–132 (2015).
- Pathak, S. et al. IRF4 is a suppressor of C-Myc induced B cell leukemia. *PLoS ONE* **6**, 1–9 (2011).
- Gopalakrishnan, R., Matta, H., Tolani, B., Triche, T. Jr. & Chaudhary, P. M. Immunomodulatory drugs target IKZF1-IRF4-MYC axis in primary effusion lymphoma in a cereblon-dependent manner and display synergistic cytotoxicity with BRD4 inhibitors. *Oncogene* **35**, 1797–1810 (2016).
- Shaffer, A. L. et al. IRF4 addiction in multiple myeloma. *Nature* **454**, 226–231 (2008).
- Langenau, D. M. et al. In vivo tracking of T cell development, ablation, and engraftment in transgenic zebrafish. *Proc. Natl Acad. Sci. USA* **101**, 7369–7374 (2004).
- Borga, C. et al. Simultaneous B and T cell acute lymphoblastic leukemias in zebrafish driven by transgenic MYC: implications for oncogenesis and lymphopoiesis. *Leukemia* **33**, 333–347 (2019).
- Borga, C. et al. Molecularly distinct models of zebrafish Myc-induced B cell leukemia. *Leukemia* **33**, 559–562 (2019).
- Tawara, M. et al. Impact of p53 aberration on the progression of Adult T-cell Leukemia/Lymphoma. *Cancer Lett.* **234**, 249–255 (2006).
- Kataoka, K. et al. Prognostic relevance of integrated genetic profiling in adult T-cell leukemia/lymphoma. *Blood* **131**, 215–225 (2018).
- Ross, C. W., Ouillette, P. D., Saddler, C. M., Shedden, K. A. & Malek, S. N. Comprehensive analysis of copy number and allele status identifies multiple chromosome defects underlying follicular lymphoma pathogenesis. *Clin. Cancer Res* **13**, 4777–4785 (2007).
- Berghmans, S. et al. tp53 mutant zebrafish develop malignant peripheral nerve sheath tumors. *Proc. Natl Acad. Sci. USA* **102**, 407–412 (2005).
- Muller, P. A. & Vousden, K. H. p53 mutations in cancer. *Nat. Cell Biol.* **15**, 2–8 (2013).
- Willemze, R. et al. The 2018 update of the WHO-EORTC classification for primary cutaneous lymphomas. *Blood* **133**, 1703–1714 (2019).



35. Shimoyama, M. Diagnostic criteria and classification of clinical subtypes of adult T-cell leukaemia-lymphoma. A report from the Lymphoma Study Group (1984-87). *Br. J. Haematol.* **79**, 428–437 (1991).
36. Ishitsuka, K. & Tamura, K. Human T-cell leukaemia virus type I and adult T-cell leukaemia-lymphoma. *Lancet Oncol.* **15**, e517–e526 (2014).
37. Renshaw, S. A. et al. A transgenic zebrafish model of neutrophilic inflammation. *Blood* **108**, 3976–3978 (2006).
38. Zhang, Y. et al. In vivo interstitial migration of primitive macrophages mediated by JNK-matrix metalloproteinase 13 signaling in response to acute injury. *J. Immunol.* **181**, 2155–2164 (2008).
39. Liu, X. et al. Zebrafish B Cell Development without a Pre-B Cell Stage, Revealed by CD79 Fluorescence Reporter Transgenes. *J. Immunol.* **199**, 1706–1715 (2017).
40. Park, G. et al. Zebrafish B cell acute lymphoblastic leukemia: new findings in an old model. *Oncotarget* **11**, 1292–1305 (2020).
41. Dee, C. T. et al. CD4-Transgenic Zebrafish Reveal Tissue-Resident Th2- and Regulatory T Cell-like Populations and Diverse Mononuclear Phagocytes. *J. Immunol.* **197**, 3520–3530 (2016).
42. Wan, F. et al. Characterization of gammadelta T Cells from Zebrafish Provides Insights into Their Important Role in Adaptive Humoral Immunity. *Front Immunol.* **7**, 675 (2016).
43. Tripodo, C. et al. Gamma-delta T-cell lymphomas. *Nat. Rev. Clin. Oncol.* **6**, 707–717 (2009).
44. Foppoli, M. & Ferreri, A. J. Gamma-delta t-cell lymphomas. *Eur. J. Haematol.* **94**, 206–218 (2015).
45. Cooke, C. B. et al. Hepatosplenic T-cell lymphoma: a distinct clinicopathologic entity of cytotoxic gamma delta T-cell origin. *Blood* **88**, 4265–4274 (1996).
46. Belhadj, K. et al. Hepatosplenic gammadelta T-cell lymphoma is a rare clinicopathologic entity with poor outcome: report on a series of 21 patients. *Blood* **102**, 4261–4269 (2003).
47. Garcia, E. G. et al. Cell of origin dictates aggression and stem cell number in acute lymphoblastic leukemia. *Leukemia* **32**, 1860–1865 (2018).
48. Niu, Y. et al. LASP1-S100A11 axis promotes colorectal cancer aggressiveness by modulating TGFbeta/Smad signaling. *Sci. Rep.* **6**, 26112 (2016).
49. Sobolewski, C. et al. S100A11/ANXA2 belongs to a tumour suppressor/ oncogene network deregulated early with steatosis and involved in inflammation and hepatocellular carcinoma development. *Gut* **69**, 1841–1854 (2020).
50. Hnisz, D. et al. Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
51. Whyte, W. A. et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* **153**, 307–319 (2013).
52. Daniels, J. et al. Cellular origins and genetic landscape of cutaneous gamma delta T cell lymphomas. *Nat. Commun.* **11**, 1806 (2020).
53. Rouault, J. P. et al. BTG1, a member of a new family of antiproliferative genes. *EMBO J.* **11**, 1663–1670 (1992).
54. Rouault, J. P. et al. Identification of BTG2, an antiproliferative p53-dependent component of the DNA damage cellular response pathway. *Nat. Genet.* **14**, 482–486 (1996).
55. Mao, P. et al. Serine/threonine kinase 17A is a novel p53 target gene and modulator of cisplatin toxicity and reactive oxygen species in testicular cancer cells. *J. Biol. Chem.* **286**, 19381–19391 (2011).
56. Zhang, B., Jiao, A., Dai, M., Wiest, D. L. & Zhuang, Y. Id3 Restricts gammadelta NKT Cell Expansion by Controlling Egr2 and c-Myc Activity. *J. Immunol.* **201**, 1452–1459 (2018).
57. Lauritsen, J. P. et al. Marked induction of the helix-loop-helix protein Id3 promotes the gammadelta T cell fate and renders their functional maturation Notch independent. *Immunity* **31**, 565–575 (2009).
58. Roels, J. et al. Distinct and temporary-restricted epigenetic mechanisms regulate human alphabeta and gammadelta T cell development. *Nat. Immunol.* **21**, 1280–1292 (2020).
59. Filippakopoulos, P. et al. Selective inhibition of BET bromodomains. *Nature* **468**, 1067–1073 (2010).
60. Delmore, J. E. et al. BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904–917 (2011).
61. Puissant, A. et al. Targeting MYCN in neuroblastoma by BET bromodomain inhibition. *Cancer Disco.* **3**, 308–323 (2013).
62. Zuber, J. et al. RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524–528 (2011).
63. Lockwood, W. W., Zejnullahu, K., Bradner, J. E. & Varmus, H. Sensitivity of human lung adenocarcinoma cell lines to targeted inhibition of BET epigenetic signaling proteins. *Proc. Natl Acad. Sci. USA* **109**, 19408–19413 (2012).
64. Ott, C. J. et al. BET bromodomain inhibition targets both c-Myc and IL7R in high-risk acute lymphoblastic leukemia. *Blood* **120**, 2843–2852 (2012).
65. Iida, S. et al. Deregulation of MUM1/IRF4 by chromosomal translocation in multiple myeloma. *Nat. Genet.* **17**, 226–230 (1997).
66. Blackburn, J. S. et al. Notch signaling expands a pre-malignant pool of T-cell acute lymphoblastic leukemia clones without affecting leukemia-propagating cell frequency. *Leukemia* **26**, 2069–2078 (2012).
67. Gutierrez, A. et al. Pten mediates Myc oncogene dependence in a conditional zebrafish model of T cell acute lymphoblastic leukemia. *J. Exp. Med.* **208**, 1595 (2011).
68. Langenau, D. M. et al. Myc-induced T cell leukemia in transgenic zebrafish. *Science* **299**, 887–890 (2003).
69. Saito, T., Yamagata, T., Takahashi, T., Honda, H. & Hirai, H. ICSAT overexpression is not sufficient to cause adult T-cell leukemia or multiple myeloma. *Biochemical biophysical Res. Commun.* **260**, 329–331 (1999).
70. Harris, A. W. et al. The E mu-myc transgenic mouse. A model for high-incidence spontaneous lymphoma and leukemia of early B cells. *J. Exp. Med.* **167**, 353–371 (1988).
71. Nayar, R. et al. Graded levels of IRF4 regulate CD8+ T cell differentiation and expansion, but not attrition, in response to acute virus infection. *J. Immunol.* **192**, 5881–5893 (2014).
72. Hayes, S. M., Li, L. & Love, P. E. TCR signal strength influences alphabeta/ gammadelta lineage fate. *Immunity* **22**, 583–593 (2005).
73. Haks, M. C. et al. Attenuation of gammadeltaTCR signaling efficiently diverts thymocytes to the alphabeta lineage. *Immunity* **22**, 595–606 (2005).
74. Jo, S. H., Schatz, J. H., Acquaviva, J., Singh, H. & Ren, R. Cooperation between deficiencies of IRF-4 and IRF-8 promotes both myeloid and lymphoid tumorigenesis. *Blood* **116**, 2759–2767 (2010).
75. Pang, S. H. M. et al. PU1 cooperates with IRF4 and IRF8 to suppress pre B-cell leukemia. *Leukemia* **30**, 1375–1387 (2016).
76. Acquaviva, J., Chen, X. & Ren, R. IRF-4 functions as a tumor suppressor in early B-cell development. *Blood* **112**, 3798–3806 (2008).
77. Ignatius, M. S. et al. tp53 deficiency causes a wide tumor spectrum and increases embryonal rhabdomyosarcoma metastasis in zebrafish. *eLife* **7**, e37202 (2018).
78. Tang, Q. et al. Optimized cell transplantation using adult rag2 mutant zebrafish. *Nat. methods* **11**, 821–824 (2014).
79. MacParland, S. A. et al. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nat. Commun.* **9**, 4383 (2018).
80. Stuart, T. et al. Comprehensive Integration of Single-. *Cell Data.* *Cell* **177**, 1888–1902 e1821 (2019).
81. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337 e324 (2019).
82. DePasquale, E. A. K. et al. DoubletDecon: Deconvoluting Doublets from Single-Cell RNA-Sequencing Data. *Cell Rep.* **29**, 1718–1727 e1718 (2019).
83. 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).
84. Gu, Z., Eils, R. & Schlesner, M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* **32**, 2847–2849 (2016).
85. Sanda, T. et al. Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
86. Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat. methods* **9**, 357–359 (2012).
87. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
88. Ross-Innes, C. S. et al. Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* **481**, 389–393 (2012).
89. Bailey, T. L., Johnson, J., Grant, C. E. & Noble, W. S. The MEME Suite. *Nucleic Acids Res.* **43**, W39–W49 (2015).
90. Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
91. Bray, N. L., Pimentel, H., Melsted, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. *Nat. Biotechnol.* **34**, 525–527 (2016).
92. Seita, J. et al. Gene Expression Commons: an open platform for absolute gene expression profiling. *PLoS one.* **7**, e40321 (2012).

## Acknowledgements

We thank the members of the Sanda laboratory for discussions and their critical reviews, Vania Teoh for FACS sorting, the NUS Comparative Medicine for supports for zebrafish study, and the Duke-NUS Genome Biology Facility for performing the 10x Genomics single-cell library preparation. This research is supported by the National Research Foundation, Prime Minister's Office, Singapore, under its Competitive Research Programme (NRF-NRFF2013-02: T.S.), and by the National Medical Research Council of the

Singapore Ministry of Health (NMRC/CIRG/1491/2018 and OFLCG18May-0028: T.S.). This research is also supported by the National Research Foundation Singapore and the Singapore Ministry of Education under its Research Centres of Excellence initiative. This study was made possible in part by generous financial support from the National Cancer Institute (NCI) for the use of dataset (phs001969).

### Author contributions

S.A., M.S.T., and M.Z.A. performed the zebrafish experiments; J.Z.L.O., R.W.W.J., and X.Z.H. performed sequencing and cell line experiments; T.K. conducted the bioinformatics analyses; Y.L., E.Y. T., B.P., H.L., and S.Y.T. performed the pathological analysis; Z.G. and S.I. supervised this study; and S.A., T.K.T., and T.S. wrote the manuscript. S.A. and T.K.T. equally contributed to this work.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-30053-9>.

**Correspondence** and requests for materials should be addressed to Takaomi Sanda.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022