scientific reports

Check for updates

OPEN

Functional and molecular single-cell analyses implicate PRDM14 in the initiation of B cell leukemia in mice

Lauren J. Tracey[1,2], Mahmoud El-Maklizi[3,4,5], Dustin J. Sokolowski[1,2,5], Miki S. Gams[3,4], Travis Brooke-Bisschop[1], Julie Ruston[1], Christine Taylor[1], Alexandra Khozin[1], Sujeetha A. Rajakumar[3], Michael D. Wilson[1,2], Cynthia J. Guidos[3,4] & Monica J. Justice[1,2]✉

The transcription factor *Prdm14* is a potent oncogene implicated in the initiation of many cancers. PRDM14 resets and maintains the pluripotent state in normal cells, but the molecular mechanisms through which PRDM14 drives oncogenesis are poorly understood. Here, we interrogated the heterogeneity of *Prdm14*-expressing cells in a T cell lymphoblastic leukemia/lymphoma mouse model. Using mass cytometry (CyTOF) of bone marrow at a pre-leukemic timepoint, an unexpected abnormal progenitor B cell population was identified. *Prdm14*-expressing progenitor B cells demonstrated short-term self-renewal and a block in differentiation when transferred to syngeneic hosts. Consistently, aged host mice succumb to a highly penetrant B-LL. Single-cell RNA-seq analyses suggests that the expression signature of these pre-leukemia cells is more consistent with that of B-1 cells than B-2 cells. B-1 cells are a self-renewing population of unconventional B cells established during embryonic development. Overlaying the chromatin binding of transcriptional marks H3K4me1 and H3K4me3 with PRDM14 suggests that PRDM14 initiates cancers through promiscuous DNA binding, activating oncogenic pathways and skewing development towards a self-renewing B-1-like phenotype. Together, our data show that *Prdm14* can initiate premature T and B cell cancer programs when expressed in hematopoietic progenitor cells.

The pluripotency regulator PR domain-containing 14 (*Prdm14*) is a potent oncogene whose mis-expression has been implicated in cancer initiation[1]. The mechanisms *Prdm14* employs to regulate pluripotency are well understood, but how *Prdm14* impacts oncogenesis remains unclear. In normal cells, expression of *Prdm14* is restricted to pluripotent cell types. In embryonic stem cells (ESCs), *Prdm14* supports the maintenance of naïve pluripotency by promoting global DNA hypomethylation and directly repressing genes that promote cell differentiation[2–4]. *Prdm14* is essential for primordial germ cell (PGC) establishment by supporting pluripotency factor expression, reducing genome-wide DNA methylation, and indirectly altering histone methylation[5]. *Prdm14* mis-expression in mouse bone marrow (BM) causes both pre-B and pre-T cell leukemia/lymphoma (LL) that harbor other recurrent mutations seen in human leukemia[6–8]. In humans, *PRDM14* amplification and overexpression is frequently observed in a wide range of malignancies including breast cancer, testicular germ cell tumours, non-small cell lung carcinoma (NSCLC), gastric cancer and T- and B-lymphoblastic leukemia/lymphoma[1,6,9–12]. Expression of *PRDM14* is associated with metastasis, resistance to chemotherapy, and worse outcomes for patients[1]. Despite numerous links to human cancer, little is known about the molecular mechanisms driving *Prdm14*-induced cancer initiation.

Mouse models of leukemia are invaluable tools for understanding how developmental genes can be corrupted to cause cancer when aberrantly activated. A classic example of this is NOTCH1 activating mutations in T cell acute lymphoblastic leukemia (T-ALL). *NOTCH1*, a regulator of early T cell development, is mutated in over 50% of human T-ALL cases[13]and corresponding mutations are found in mice[14]. Constitutive activation of NOTCH1 in hematopoietic progenitor cells leads to fully penetrant T-ALL in mice[15].

[1]Program in Genetics and Genome Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada. [2]Department of Molecular Genetics, University of Toronto, Toronto, ON M5S 1A8, Canada. [3]Program in Developmental and Stem Cell Biology, The Hospital for Sick Children, Toronto, ON M5G 0A4, Canada. [4]Department of Immunology, University of Toronto, Toronto, ON M5S 1A8, Canada. [5]Mahmoud El-Maklizi and Dustin J. Sokolowski contributed equally. ✉email: monica.justice@sickkids.ca

Less common in mice are B cell malignancies. Mouse B cells can be broadly divided into conventional CD19^{+} B220^{+} B-2 cells, which originate from hematopoietic stem cells (HSCs) in the BM and are re-populated throughout life, and CD19^{+} B220$^{lo/-}$ B-1 cells, which are established during early embryonic development and maintain their population through self-renewal[16,17]. In one mouse model of B lymphoblastic leukemia/lymphoma (B-LL), transformation efficiency of B cells by the oncogenic BCR-ABL translocation was compared between B-1 and B-2 progenitor cells[18]. Transformed B-1 progenitor cells initiated an aggressive form of B-LL characterized by rapid onset and high tumor burden. Furthermore, B-1 progenitor cells displayed a marked resistance to apoptosis compared to transformed B-2 progenitor cells[18]. These results suggest that oncogenic mutations arising in B-1 progenitor cells can give rise to a high-risk subtype of B-LL, likely due to the intrinsic proliferative and self-renewal potential of B-1 cells[18,19]. Studying additional mouse models of B-LL will help elucidate molecular mechanisms underlying high risk subtypes of leukemia and open new avenues for therapeutic intervention.

PRDM14 was identified as an oncogene in AKXD mice, which develop both T and B-cell leukemia due to insertional mutation by murine leukemia retroviruses[7]. Subsequently, we generated mice with Mx1-cre-inducible *Prdm14* expression (R26PR;Mx1-cre mice), which showed that *Prdm14* initiates T-LL with a very short latency[20]. When *Prdm14* expression was induced from the *Rosa26* locus in hematopoietic stem and progenitor cells (HSPC), R26PR; Mx1-cre mice rapidly developed T-LL with 100% penetrance; median survival was 41 days post-polyinosinic:polycytidylic acid (pIpC), with CD8 immature single positive (ISP) T cells accumulating in the BM and thymus[20]. Similar to other T-LL mouse models and human T-ALL, activating *Notch1* mutations were prevalent[21].

Here, we performed high parameter single-cell mass cytometry time-of-flight (CyTOF) and single-cell RNA sequencing (scRNA-seq) to investigate the early stages and molecular mechanisms of *Prdm14*-driven leukemia initiation. CD8 ISP cells were present in the thymus 8–10 days after *Prdm14* induction. Surprisingly, an expanded, abnormal BM population of progenitor B lymphocytes was also present at this time. ScRNA-seq analysis carried out on *Prdm14*-expressing B cells revealed a gene expression signature similar to that of human leukemic B-1 cells[22]. These progenitor B cells expanded rapidly and retained their immunophenotype after adoptive transfer to syngeneic host mice, eventually generating a highly penetrant B-LL. Together, these data reveal a molecular mechanism through which *Prdm14* subverts B cell development to initiate B-LL that is B-1-like, in addition to a T-LL.

Results

Deep phenotyping identifies an expanded progenitor B cell population in pre-leukemic BM expressing *Prdm14*

To investigate molecular mechanisms through which *Prdm14* promotes leukemogenesis, we used time-of-flight mass cytometry (CyTOF) to deeply profile BM cells 10 days after Mx1-cre-induced expression of *Prdm14* by pIpC administration in R26PR;Mx1-cre mice as compared to pIpC-injected control R26PR mice. T-LL has not developed at this early timepoint[21], providing a pre-leukemic window to identify early molecular perturbations caused by *Prdm14*. We quantified expression of 23 markers expressed by classical HSPC, myeloid cells and immature and mature B and T cells (see Methods) from six R26PR;Mx1-cre mice and three control R26PR mice. We pre-gated on cells lacking Ly6G, a granulocyte marker, to enrich for HSPC and lymphocytes prior to performing dimensionality reduction using the unbiased t-Distributed Stochastic Neighbor Embedding (tSNE) algorithm[23]followed by FlowSOM clustering[24] of the tSNE1 and tSNE2 parameters. Unexpectedly, given our prior work documenting *Prdm14*-induced T-LL in this model, we observed that several immature and mature B cell clusters were also differentially abundant between genotypes at this early timepoint (Fig. 1A, B, Supplemental Fig. 1). Among the immature IgM^{-} B cell metaclusters (MC), MC-7 representing pro-B cells (CD24^{med} CD117^{hi} B220^{lo} CD19^{lo} CD127^{hi} CD25^{-} CD22^{-} MHC2^{-}) (Fig. 1B) was significantly more abundant in R26PR;Mx1-cre mice (Fig. 1C, D). Conversely, MC-2 and MC-5, consisting respectively of early pre-B (CD24^{hi} CD117^{-} B220^{lo} CD19^{lo} CD25^{+} CD127^{lo} CD22^{lo} MHC2^{lo}) and late pre-B (CD24^{hi} CD117^{-} B220^{lo} CD19^{lo} CD25^{-} CD127^{med} CD22^{lo} MHC2^{lo}) (Fig. 1B) cells were significantly depleted in this strain (Fig. 1C, D). These findings suggest that *Prdm14* induction partially blocked B cell development at the pro-B cell stages, a conclusion also supported by the decreased abundance of MC-3 and MC-4, IgM^{+} CD117^{-} CD25^{-} CD127^{-} subsets consisting respectively of late pre-B (CD24^{hi} B220^{med} CD19^{lo} CD22^{lo} MHC2^{lo}) and mature (CD24^{lo} B220^{hi} CD19^{hi} CD22^{hi} MHC2^{hi}) B cells (Fig. 1).

MC-8 and MC-12 IgM^{-} B cell subsets were abundant in R26PR;Mx1-cre mice, but were nearly absent in control mice, although values were not statistically significant. MC-12 was similar to pro-B (MC-7) cells but lacked B220 and also had more MHC2 (Fig. 1C, D). MC-8 was B220^{+} MHC2^{hi} but lacked Sca-1 and CD117 and expressed lower amounts of CD19, CD127 and CD5. MC-7 pro-B cells in R26PR;Mx1-cre mice expressed more of the HSPC markers Sca-1 and CD117 than in controls, but also more late/mature B cell markers CD22 and MHC2 (Fig. 1C, D), suggesting that *Prdm14* expression corrupts early steps of B cell differentiation.

The increase in CD19^{+} CD127^{+} Sca-1^{+} B220^{-} cells observed in *Prdm14*-expressing R26PR;Mx1-cre but not R26PR control mice was confirmed using flow cytometry as an orthogonal method (Supplemental Fig. 2A, B). Importantly, 98–99% of CD19^{+} CD127^{+} Sca-1^{+} B220^{-} cells expressed EGFP (Supplemental Fig. 2A), a proxy for *Prdm14* expression, demonstrating that *Prdm14* acted cell autonomously to expand this subset. *Prdm14*-induced T-LLs, like others with activating *Notch1* mutations, consist of malignant immature T cells that are variably CD24^{+} CD25$^{+/-}$ TCRβ^{lo} CD4^{-} CD8^{+} or CD24^{+} CD25$^{+/-}$ TCRβ^{lo} CD4^{+} CD8^{+}[25]. However, cells with these immunophenotypes were not detected in BM from these R26PR; Mx1-cre mice at 10 days post-pIpC treatment using flow cytometry (not shown), although immature CD24^{+}CD4^{-} CD8^{+} TCRβ$^{lo/-}$ GFP + T cell precursors were increased in the thymus (Supplemental Fig. 2C, D), suggesting that *Prdm14*-induced T cell

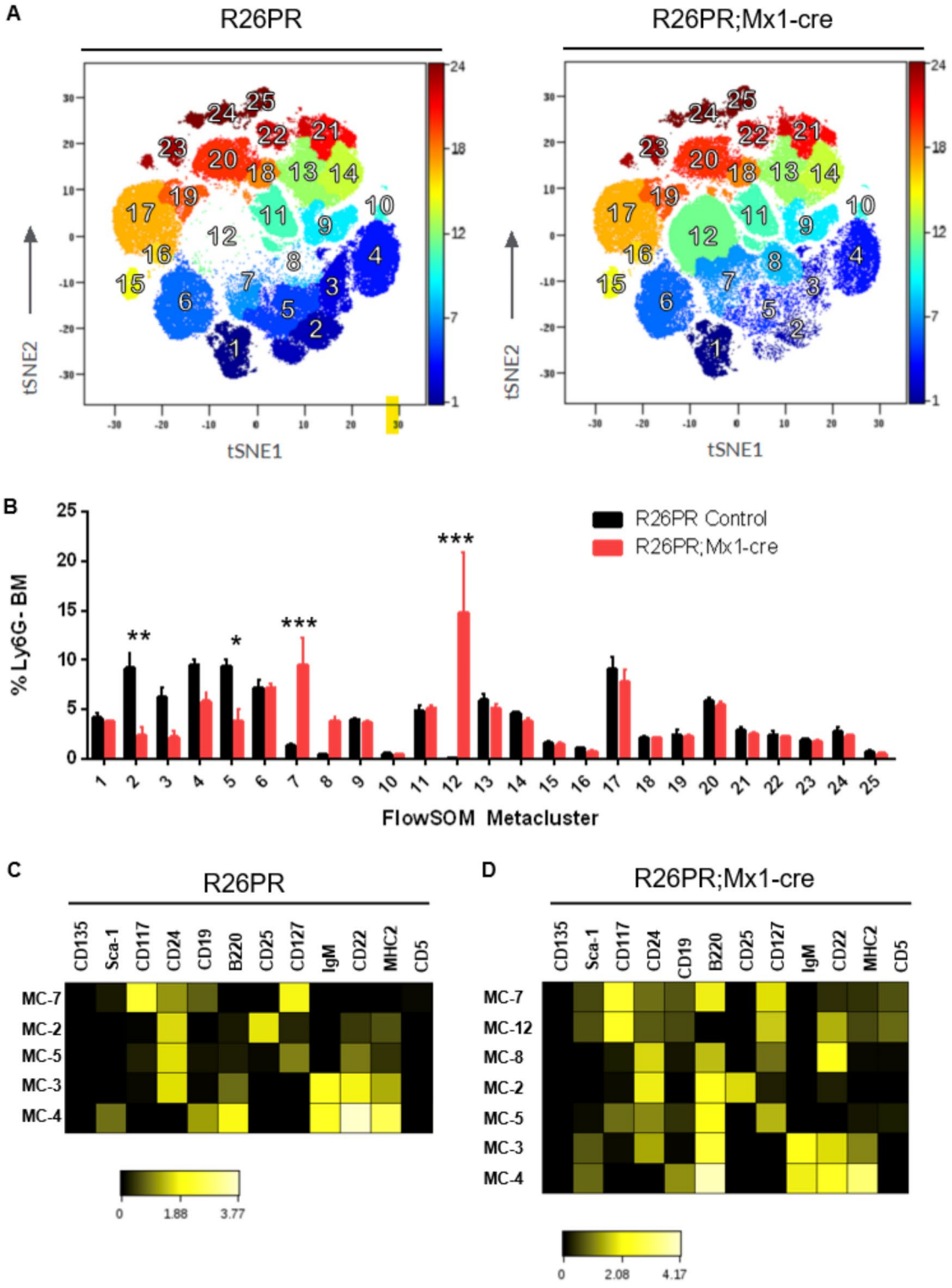


Fig. 1. CyTOF analysis identifies unique B cell clusters that have varying abundance depending on genotype. (**A**) tSNE plots of Ly6G$^-$ BM cells from control (R26PR; WT) and R26PR;Mx1-cre mice 10 days after pIpC injection during the pre-leukemic phase. Plots were coloured by the FlowSOM Metacluster ID. Colour legend is shown on the right y-axis of each plot. (**B**) Quantification of the percent Ly6D$^-$ BM present in each FlowSOM Metacluster (MC) per mouse, per genotype. Multiple t-tests were used and the False Discovery Rate (Q) value was less than 0.05 for MC-2, MC-5, MC-7 and MC-12. * $p < 0.02$, ** $p < 0.002$, ***$p < 0.0002$. (**C** and **D**) Heatmap showing intensities of the indicated markers in each MC from a representative R26PR control sample (**C**) or a representative R26PR;Mx1-cre sample (**D**). Color scales represent the ratio of the arcsinh-transformed marker medians for each cluster scaled to the columns' minimum arcsinh-transformed value. MC were arranged according to inferred maturation status (least to most mature).

leukemogenesis initiates in the thymus. Thus, CyTOF enabled the identification of obscured abnormal B cells present in *Prdm14*-expressing R26PR;Mx1-cre BM.

Prdm14-expressing pre-leukemic B cells initiate leukemia/lymphoma when transferred into syngeneic mice

We hypothesized that the expanded B cells were pre-malignant; however, this possibility could not be assessed because T-LL develops within a month of *Prdm14* induction in R26PR;Mx1-cre mice. Therefore, we used an adoptive transfer system to test this hypothesis. We sorted 50, 000 GFP$^+$ CD11b$^-$ CD8β$^-$ Sca-1$^+$ CD19$^+$ CD127$^+$ BM cells (referred to as GFP$^+$ CD19$^+$ Sca-1$^+$ CD127$^+$ for simplicity) from R26PR;Mx1-cre (CD45.2) donor mice 8 days post-pIpC and injected them into the right femur of *B6.Rag2.CD45.1* host mice (Fig. 2A, Supplemental Fig. 3A). In addition, we injected one host mouse with unsorted R26PR;Mx1-cre BM cells (referred to as total BM) from each donor into one host mouse to assess the leukemogenic potential of unsorted cells (Fig. 2A). We co-injected all mice with total BM cells from *B6.Rag2.CD45.2/CD45.1* (referred to as Rag2 KO) as a positive control for HSPC and pro-B cell engraftment (not shown). Host mice were sacrificed four weeks post-injection and the right femur (injection site), distant BM (left femur & tibias), and thymus were analyzed via flow cytometry.

All host mice that received GFP$^+$ CD19$^+$ Sca-1$^+$ CD127$^+$ BM cells showed successful engraftment of CD45.2$^+$ cells and extensive expansion within the BM (Fig. 2B, C, Supplemental Fig. 3B). Cells from the total donor BM migrated to the thymus, whereas the sorted cells did not (Fig. 2C). In the thymus of all three mice who received total donor BM, 50% of the donor origin GFP$^+$ CD45.2$^+$ cells were CD8$^+$ (Supplemental Fig. 3D), representing early stages of T-LL. Host right femurs showed a 20-to-70-fold expansion of sorted cells relative to the number of injected cells, compared to a 7-to-19-fold expansion of total BM and a 0.1-to-5-fold expansion for co-engrafted cells (Fig. 2C, Supplemental Table 1), demonstrating the significant competitive advantage of GFP$^+$ CD19$^+$ Sca-1$^+$ CD127$^+$ cells. Importantly, 90–96% of CD45.2 single positive donor cells at the injection site (right femurs; $n = 11$) retained their sorted immunophenotype of GFP$^+$ CD19$^+$ CD127$^+$ Sca-1$^+$ (Fig. 2D, E), demonstrating that these cells proliferated and expanded. These cells expressed no or low levels of B220, which is unusual for CD19$^+$ B cells (Fig. 2D). Host mice who received total BM showed a different cellular profile: in 2/3 of their right femurs over 75% of CD45.2 single positive cells were GFP$^+$ CD19$^+$ CD127$^+$ Sca-1$^+$, demonstrating this population's ability to expand from unsorted BM (Fig. 2E, Supplemental Fig. 3C). Together, these data support the idea that *Prdm14* expression induces expansion of abnormal progenitor B cells, which are blocked in differentiation and have the capacity for short-term self-renewal in secondary transplants.

To determine the capacity of the progenitor B cells in R26PR; Mx1-cre mice to initiate LL over time, we carried out an in vivo cell transfer experiment before aging the host mice until they showed signs of illness. The previous experiment demonstrated that sorted cells were able to home to BM beyond the injection site, thus IV injections were used for simplicity going forward. GFP$^+$ CD19$^+$ Sca-1$^+$ CD127$^+$ BM cells were sorted from R26PR;Mx1-cre mice ($n = 4$) 8 days post-pIpC and injected into syngeneic *Rag2-/-* host mice ($n = 4$ hosts per donor) (Fig. 3A). Overall, 14/16 host mice were sacrificed due to illness by 225 days post-transfer with a median survival of 156 days post-transfer (Fig. 3B). Moribund mice had grossly enlarged spleens (Fig. 3C) and complete blood counts (CBCs) showed a significant increase in white blood cells (WBC; $4.0–194.0 \times 10^3$/ μL, mean 39.0 ± 58.0) and lymphocytes ($1.5–24.8 \times 10^3$/ μL, mean 9.28 ± 10.6) compared to non-injected, age-matched *Rag2*$^{-/-}$ control mice (WBCs: $0.6–3.7 \times 10^3$/ μL, mean 2.0 ± 1.2; lymphocytes: $0.4–1.9 \times 10^3$/ μL, mean 0.94 ± 0.5) (Fig. 3D). Conversely, CBCs in moribund mice showed a significant decrease in red blood cells (RBC) ($4.5–10.0 \times 10^6$/ μL, mean 7.4 ± 1.7) and hemoglobin (Hgb) (9.4–14.7 x g/L, mean 11.5 ± 1.5) compared to non-injected age-matched *Rag2*$^{-/-}$ control mice (RBC: $8.4–10.2 \times 10^6$/ μL, mean 9.4 ± 1.8; Hgb: 11.2–16.3 x g/L, mean 14.0 ± 1.8) (Fig. 3D).

BM from moribund host mice was analyzed by flow cytometry. Indeed, 30–60% of live BM cells were GFP$^+$ CD19$^+$ CD127$^+$, compared to 0–0.01% of live CD19$^+$ CD127$^+$ BM cells in *Rag2*$^{-/-}$ control mice (Fig. 3E, F). GFP$^+$ CD19$^+$ CD127$^+$ cells expressed CD43 but were B220$^-$ or B220^{lo} (Fig. 3G), consistent with an abnormal B cell immunophenotype. Control *Rag2*$^{-/-}$ spleens showed the expected white pulp hypoplasia due to B and T cell lymphopenia. In contrast, moribund host mice injected with GFP$^+$ CD19$^+$ CD127$^+$ *Prdm14*-expressing cells showed marked infiltration of lymphoblast cells throughout BM, spleen, liver, kidney, and the meninges of the brain, consistent with metastatic LL (Fig. 3H). Thus, *Prdm14*-expressing progenitor B cells initiated aggressive LL when transferred to host mice.

We corroborated these results using a second mouse model in which FLAG-tagged *Prdm14* (FLPR) expression was restricted to CD19$^+$ B cells by mating R26FLPR mice to CD19-cre transgenic mice (R26FLPR;CD19-cre). Aged R26FLPR; CD19-cre mice, but not R26FLPR control mice, succumb to an infiltrative LL with a progenitor B cell immunophenotype within 14 months of life (Supplemental Figs. 4 and 5). LL cells express PRDM14 but not activated NOTCH1 (Supplemental Fig. 5). All R26FLPR;CD19-cre mice accumulated CD19$^+$ CD127$^+$ Sca-1$^+$ CD24$^+$ IgM$^-$ B220$^-$ cells in their BM (4–40% of live BM) by 6 months of age (Supplemental Fig. 6). Taken together, aberrant expression of *Prdm14* in B cells leads to a completely penetrant and infiltrative LL with a phenotype similar to that identified in the adoptive transfer experiment.

Pre-leukemic progenitor B cells express proto-oncogenes but not DNA damage response genes

To determine how *Prdm14* expression impacts early hematopoietic development and to identify mis-regulated genes and pathways in pre-leukemic B cells, we performed scRNA-seq on three pre-leukemic (PL) cell populations from the BM of R26PR;Mx1-cre mice 8 days post-pIpC: Total GFP$^-$, total GFP$^+$, and sorted CD19$^+$ CD127$^+$ Sca-1$^+$ B cells (including GFP$^-$ and GFP$^+$ cells). For comparison, we also sorted a fourth control subset of CD19$^+$ CD127$^+$ Sca-1$^+$ B cells from the BM of 8 days post-pIpC control mice for scRNA-seq (Supplemental Fig. 7). We analyzed 2000–3000 cells from each of the four subsets after quality control and integrated single

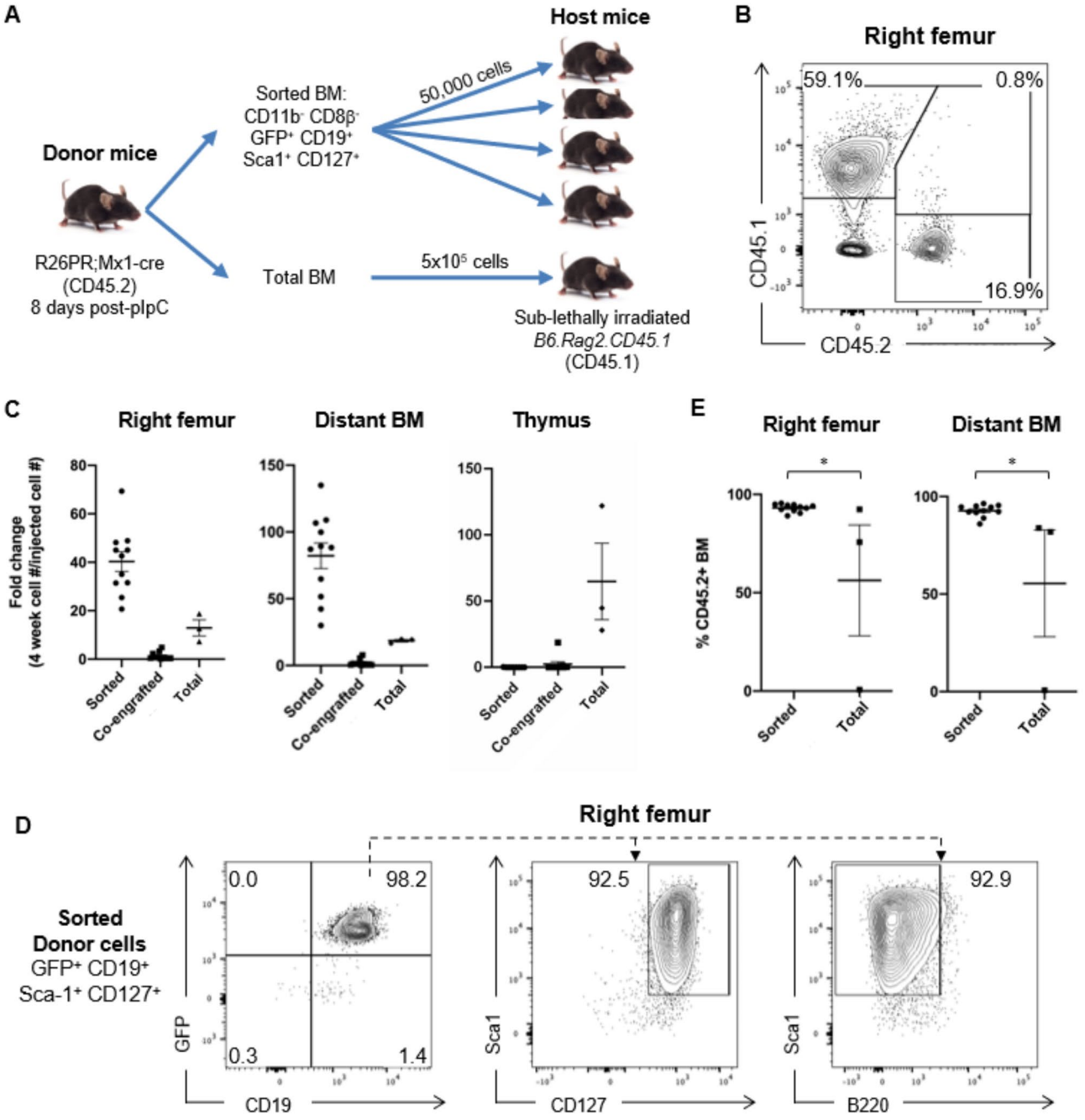
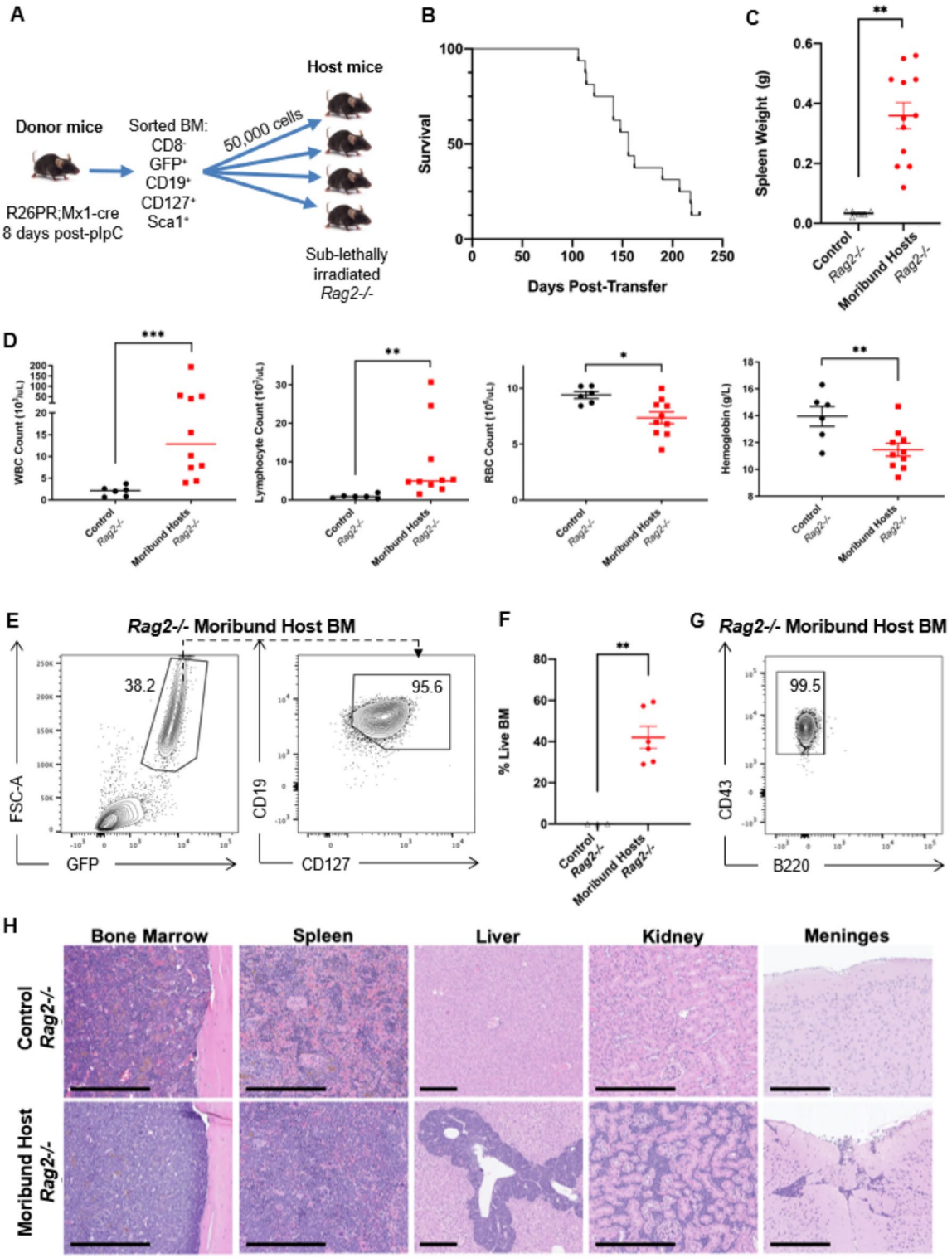


Fig. 2. *Prdm14*-expressing GFP$^+$ CD19$^+$ CD127$^+$ cells expand rapidly and self-renew in secondary hosts. (**A**) Schematic showing design of interfemoral adoptive transfer experiments. GFP$^+$ CD19$^+$ CD127$^+$ BM cells sorted from R26PR;Mx1-cre CD45.2 + mice ($n = 3$) 8 days post-pIpC were injected into the right femurs of sub-lethally irradiated *B6.Rag2.CD45.1* host mice ($n = 11$). Sorted cells were co-transferred with 1×10^5 control donor Rag2 KO cells as a read out for engraftment (not shown). Total (unsorted) R26PR; Mx1-cre BM cells (5×10^5) from each donor were injected into one *B6.Rag2.CD45.1* host mice per experimental donor mouse for comparison. (**B**) Representative flow analysis of right femur (injection site) of mice who received GFP$^+$ CD19$^+$ CD127$^+$ BM cells. Gate shows the percentage among live single cells. (**C**) Quantification of the number of CD45.2 single positive (sorted and total BM) or CD45.1/2 double positive (co-engrafted) cells in the right femur, distant BM, or thymus, represented as the fold change between the number of cells present 4-weeks after transfer compared to the number of cells injected on the day of transfer. (**D**) Representative flow analysis of the right femur from host mice that received GFP$^+$ CD19$^+$ CD127$^+$ BM cells. The left plot shows GFP vs. CD19 expression pre-gated on CD45.1$^-$ CD45.2$^+$ live singlets. The middle and right plots show Sca-1 vs. CD127 (middle) or B220 (right) gated on CD45.2$^+$ GFP$^+$ cells. (**E**) Graphs show the percentage of GFP$^+$ CD19$^+$ CD127$^+$ Sca-1$^+$ cells as a percentage of CD45.1$^-$CD45.2$^+$ cells in the right femur or distant BM (left femur & tibias). Significance was determined using an unpaired t-test. * $p < 0.05$; n.s., not significant.



cell-data analysis. A total of 24 clusters were derived from unsupervised clustering, and visualized using uniform manifold approximation and projection (UMAP; Fig. 4A). Using a standard differential gene expression workflow, we identified the most significantly up-regulated genes (FDR-adjusted p-value < 0.05, percent of cells expressing marker > 25%, fold change > 1.5). Clusters were annotated with cell type labels based on their gene expression profiles (Fig. 4A).

As expected, certain clusters were comprised of almost exclusively pre-leukemic GFP⁻ cells (e.g. erythroid clusters 4 and 16), whereas pre-leukemic GFP⁺ cells were scattered mainly through T and B cell clusters (Fig. 4A). Consistent with NOTCH1-driven T-LL lymphoblast cells that accumulate in R26PR; Mx1-cre mice, Cluster 18 was comprised of >99% pre-leukemic GFP⁺ cells and had a gene expression pattern similar to CD8⁺ ISPs, as evidenced by the top marker genes for cluster 18 including pre-T cell receptor α chain (PTCRA), TCRβ, and H2-

◀**Fig. 3**. *Rag2*$^{-/-}$ mice injected with sorted BM from R26PR;Mx1-cre mice succumb to a leukemia/lymphoma-like disease. (**A**) Intravenous adoptive transfer experimental outline. Pre-leukemic R26PR; Mx1-cre (CD45.2) BM from experimental donor animals ($n = 4$ donor mice) at 8 days post-pIpC was sorted for CD8$^-$ GFP$^+$ CD19$^+$ CD127$^+$ Sca1$^+$ cells. 50,000 sorted cells were injected into the tail vein of sub-lethally irradiated *Rag2*$^{-/-}$ host mice ($n = 16$ host mice) and aged. (**B**) Kaplan-Meier survival curve of aged *Rag2*$^{-/-}$ host mice injected with sorted cells. Host mice had a median survival of 145 days post-transfer ($n = 14$). (**C**) Spleen weights after dissection of moribund host *Rag2*$^{-/-}$ ($n = 12$) and age-matched control *Rag2*$^{-/-}$ ($n = 6$) mice. Significance was determined using an unpaired t-test. $^{**}p < 0.01$. (**D**) White blood cells (WBC), red blood cells (RBC), lymphocytes, and hemoglobin in peripheral blood of moribund host *Rag2*$^{-/-}$ ($n = 10$) and age-matched control *Rag2*$^{-/-}$ ($n = 6$) mice. Significance was tested using an unpaired t-test for RBC and hemoglobin counts, and error bars represent the mean ± standard error of the mean (SEM). Significance was tested using a Kolmogorov-Smirnov test for WBC and lymphocyte counts as these data are not normally distributed. Error bars represent the median value. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{**}p < 0.001$ (**E**) Representative plots from flow cytometry analysis of moribund host *Rag2*$^{-/-}$ BM cells ($n = 6$). (**F**) Quantification of GFP$^+$ CD19$^+$ CD127$^+$ cells in moribund host BM. $^{**}p < 0.001$. (**G**) Representative image of GFP$^+$ CD19$^+$ CD127$^+$ cells expressing CD43 but not B220 in moribund host BM. (**H**) Representative images of hematoxylin and eosin-stained sections of BM, spleen, liver, kidney, and brain tissues from moribund host mice and age-matched *Rag2-/-* controls. All images were taken with a 10X objective. Scale bar, 250μM.

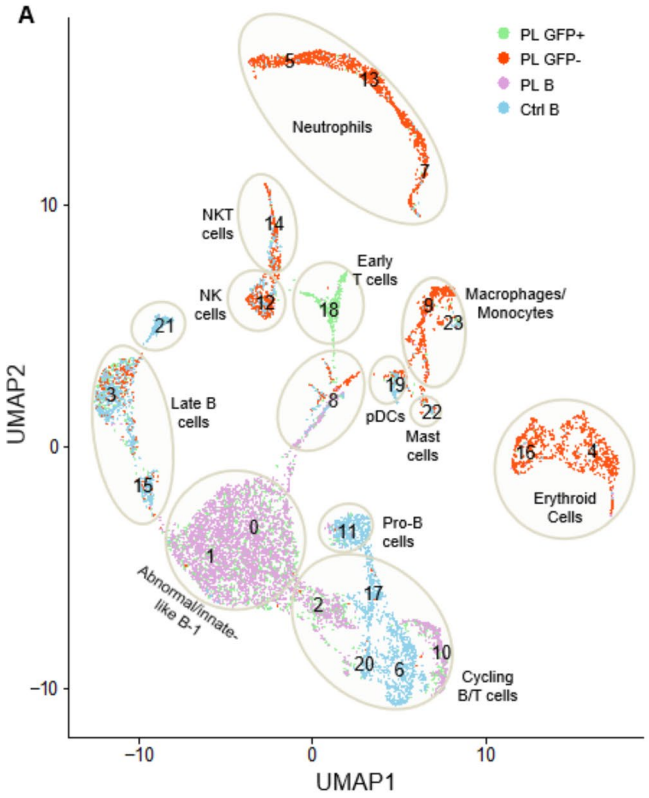
T3 (Supplemental Table 2). Cluster 18 contained cells that expressed high levels of *Notch1* and NOTCH1 target genes *Hes1* and *Cdk4*[26], and thus represents an early population of T-LL cells present in the BM.

The majority of pre-leukemic B cells fell into clusters 0 and 1, which shared 347 cell-type marker genes relative to all other clusters. Many early B cell genes were identified as marker genes for these clusters, including *Ebf1*, *Pax5*, *CD127*, *Vpreb1* and *Igll1* (Fig. 4B, C). Interestingly, a striking number of AP-1 transcription factors were expressed in these clusters, namely *Fos*, *Fosb*, *Jun*, *Junb*, and *Jund*. These genes are expressed in innate lymphoid cells within the hematopoietic system, but are not typically co-expressed with early B cell genes (Fig. 4B, C)[27]. AP-1 proteins, which homo- and hetero-dimerize to act as transcription factors, are rapidly activated in response to stimuli to affect proliferation, apoptosis, differentiation, and malignant transformation[28]. *Fos* and *Jun* have been widely studied as proto-oncogenes[28,29], but have never been connected to *Prdm14* in cancer. Beyond AP-1 proto-oncogenes, *Myb* and *Cbfa2t3*, known *Prdm14* target genes that have been implicated in cancer initiation, were expressed in clusters 0 and 1. Furthermore, comparing the cell markers from clusters 0 and 1 to 3162 human gene expression data sets[30], we found that these cell markers strongly enriched for EZH2 inhibition in cancer cells, subpopulations of BM-derived cancer stem cells, tumor subtyping, and FOXM1 inhibition in cancer cells (Supplementary Fig. 8). These results show that the enriched marker genes in clusters 0 and 1 reflect an oncogenic gene signature and could be classified as abnormal/innate-like B cells.

Pre-leukemic B cells have a CD19$^+$B220$^-$ immunophenotype, consistent with that of unconventional B-1 cells. B-1 cells differ from conventional B-2 cells with respect to their origin, self-renewal capacity, location in the body and surface protein expression (Fig. 5A). To further explore the nature of these cells, we compared the DEGs from clusters 0 and 1 with a 30-gene signature that distinguishes human B-1 and B-2 cell leukemias. 9/20 B-1 signature genes were expressed in pre-leukemic B cells from clusters 0 and 1 (*Cd72*, *Cd79b*, *Bcl7a*, *Bach2*, *Vpreb3*, *Akap12*, *Igll1*, *Cd19*, *Lef1*), including 2 of the top 3 most significantly up-regulated genes (*Vpreb3*, *Igll1*) (Fig. 5B)[22]. Consistently, none of the B-2 signature genes were up-regulated in clusters 0 and 1, while 4/10 were significantly downregulated (*Tyrobp*, *Ahnak*, *S100a4*, *Anxa2*)(Fig. 5B). Clusters 0 and 1 also expressed known B-1 cell genes (*Foxo1*, *CD93*, *Siglecg*), further supporting their classification as B-1-like cells (Fig. 4B, Supplemental Table 2)[27,31,32].

To identify differences between control B cells and pre-leukemic B cells contained within clusters 0 and 1, we carried out differential expression analysis. Gene ontology (GO) enrichment analysis identified Cellular Response to DNA Damage as the most enriched biological process in genes downregulated in pre-leukemic B cells compared to controls (Fig. 5C), which included chromosome stability genes *Xrcc6* and *Mcm4*. Notably, *Xrcc6* is expressed at high levels in pre-pro-B cells, but its expression is reduced in the abnormal/innate-like B-1 cell clusters. Together, these results suggest that the mis-regulation of known oncogenes and DNA damage response genes may facilitate cancer development in the abnormal B cells that expand in *Prdm14*-expressing B-LL.

We interrogated the molecular programs involved in the development of these abnormal pre-leukemic B cells using pseudotime analysis, which helps to infer a lineage trajectory across single cells, identifying key molecular programs driving cellular phenotypes. Pseudotime analysis in B cell-associated scRNA-seq clusters 0, 1, 2, 11 and 17 identified one lineage beginning at cluster 11 (pro-B) and ending at cluster 1 (abnormal/innate-like B-1) which was associated with differential expression of 30 transcription factors (Fig. 5D, E), many of which have been previously associated with cancer related biological pathways (Fig. 5F). Two AP-1 transcription factors, *Jun* and *Junb*, were identified as key genes involved in the lineage progression to abnormal/innate-like B-1 cells, supporting a role for AP-1 in tumour identity (Fig. 5E). *Klf4*, a reprogramming gene in induced pluripotent stem cell (iPSC) generation[33] and a target of *PRDM14* in human PGC-like cells[34], was also identified as a key transcription factor defining the pseudotime axis (Fig. 5E). Collectively, these data show that *Prdm14*-expressing pre-leukemic B cells aberrantly express proto-oncogenes, but not DNA damage response genes, supporting the idea that these cells are early cancer initiating cells for B-lineage leukemia.



B

Pathway	Gene	Average logFC	FDR-adjusted p-value
Early B cell Genes	*Igll1*	0.912832922	<8E-300
	Vpreb1	0.911531692	<8E-300
	Ebf1	0.899172442	<8E-300
	Il7r	0.758761151	<8E-300
	Pax5	0.474472607	1.7071E-204
AP-1 Genes	*Jund*	0.679096649	<8E-300
	Jun	0.677011752	1.6237E-292
	Fos	0.589996018	7.2931E-191
	Fosb	0.445612954	2.109E-132
	Junb	0.262506517	2.58918E-67
Prdm14 Targets	*Myb*	0.556216178	1.4318E-258
	Cbfa2t3	0.448501288	8.5038E-215
B-1 cell Genes	*Siglecg*	0.49469997	1.9572E-254
	Cd93	0.257358886	6.20774E-84
	Foxo1	0.307300841	1.89971E-79

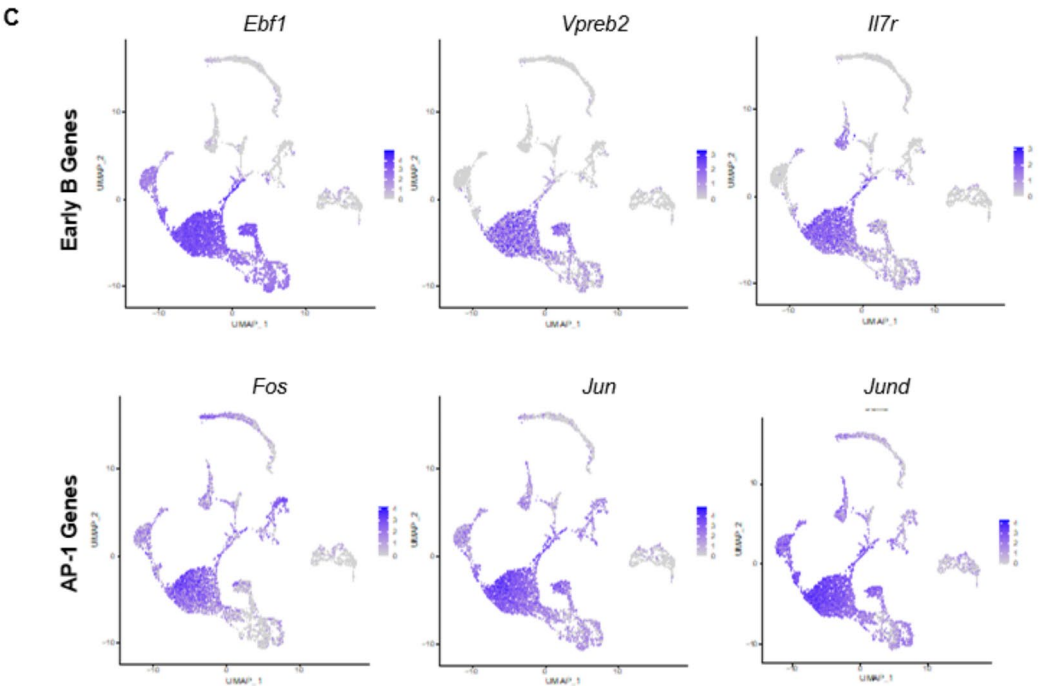


Fig. 4. Clustering analysis of aggregated scRNA-seq samples visualized with a UMAP plot identifies 24 clusters, including abnormal/innate-like B-1-like clusters. (**A**) Unsupervised graph-based clustering was used on the aggregated samples, and the UMAP dimensionality reduction algorithm was used to display the clustering output. Single cells from each sample are colour-coded. Cell types were annotated based on gene expression. (**B**) Significant and representative marker genes that define Cluster 0 are shown based on relevant classifications. The average log2 fold change (FC) across all cells in the cluster and the FDR-adjusted p-value are indicated for each gene. (**C**) Heat maps overlaid onto the UMAP plot highlight gene expression patterns for early B cell and AP-1 genes.

PRDM14 binds DNA adjacent to genes that are mis-regulated in pre-leukemic cells

Prdm14 binds DNA in pluripotent cells to modulate gene expression. Accordingly, *Prdm14* may be acting in a similar manner to directly regulate gene expression in expanded B progenitor cells. To test this hypothesis, a FLAG-PRDM14 protein was used to map DNA binding sites in CD8β$^-$ CD19$^+$ CD127$^+$ Sca-1$^+$ cells sorted from R26FLPR;Mx1-cre mice 8 days post-pIpC injection (referred to as pre-leukemic B cells) using Cleavage Under Targets and Release Using Nuclease (CUT&RUN). H3K4me1 and H3K4me3 sites were also identified to determine putative enhancer and promoter locations, respectively. An IgG antibody was used as a negative control for all experiments Fig. 6A).

Sample replicates were highly correlated (>0.90 for PRDM14 peaks, >0.98 for H3K4me1 peaks, and >0.99 for H3K4me3 peaks), independently of batch (Supplemental Fig. 9). We obtained 13,518 peaks associated with PRDM14 binding, 242,150 peaks associated with H3K4me1 binding, and 42,752 peaks associated with H3K4me3 binding. PRDM14, H3K4me1 and H3K4me3 peaks were assigned to nearby genes and annotated genomic features. As expected, 70% of H3K4me3 peaks mapped to promoter regions within 3 kb from a gene transcriptional start sites (TSS) in the mm10 genome annotation (Fig. 6B). PRDM14 peaks mapped consistently to promoters, introns, and distal intergenic regions (Fig. 6B). In pluripotent cells, PRDM14 peaks overlap predominantly with H3K4me1-enriched enhancer regions 10–100kb away from TSSs[2]. In sorted pre-leukemic B cells, 17% of PRDM14 peaks overlapped with H3K4me1-bound enhancers. Furthermore, 21% of PRDM14 peaks overlapped with H3K4me3-bound promoters, highlighting a novel association of PRDM14 and promoters (Fig. 6C).

Previous studies have identified a canonical PRDM14 binding motif that is nearly identical in both humans and mice[2,35]. We ran motif enrichment analysis on the shared peaks between published PRDM14 ChIP-seq data from ESCs[2] and our PRDM14 CUT&RUN dataset from pre-leukemic B cells and found the canonical PRDM14 binding sequence was the most enriched motif identified (Supplemental Fig. 10A). Interestingly, PRDM14 binds to many non-canonical motifs in pre-leukemic B cells (Supplemental Fig. 10A). We hypothesized that chromatin accessibility could be playing a role in promiscuous PRDM14 binding in pre-leukemic B cells. Indeed, a 10-fold reduction in H3K4me1 and H3K4me3 marks at canonical PRDM14 binding sites is seen in pre-leukemic B cells compared with canonical PRDM14 binding sites in ESCs (Supplemental Fig. 10B, C), suggesting that inaccessible chromatin is impacting the ability of PRDM14 to bind its canonical motif in pre-leukemic B cells.

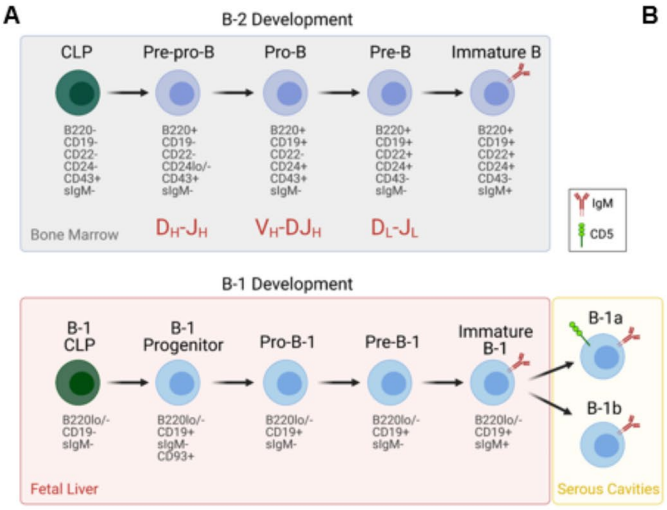
To determine if PRDM14 may directly regulate the mis-expressed genes in pre-leukemic B cells, CUT&RUN peaks were assigned to genes based on proximity mapping +/- 3 kb of a gene's TSS. Peak-associated genes were then overlapped with marker gene sets that defined scRNA-seq clusters and tested for enrichment. The abnormal/innate-like B-1 clusters' marker genes had the strongest enrichment ($p < 6.09 \times 10^{-39}$) for genes bound by PRDM14 within 10 kb of their TSS compared to all other clusters, followed by the pro-B cell cluster's marker genes (Fig. 6D). A similar enrichment was seen when comparing PRDM14 peaks that overlapped with either H3K4me1 ($p < 7.45 \times 10^{-24}$) or H3K4me3 ($p < 1.83 \times 10^{-20}$). For example, a distinct PRDM14 binding peak was observed upstream of the *Jund* TSS, which overlapped with H3K4me1 (Fig. 6A). In contrast, PRDM14 binding was not significantly associated with marker genes for other cell types such as neutrophils ($p > 0.01$) or macrophages/monocytes ($p > 0.01$), as expected (Fig. 6D). Together, these data suggest that overexpression of PRDM14 in pre-leukemic cells results in ectopic DNA binding interactions that, directly or indirectly, lead to mis-regulated expression of genes that drive oncogenesis.

Discussion

Collectively, this work provides the first demonstration that *Prdm14* can initiate B-LL and suggests a mechanism to explain how *Prdm14* hijacks normal B cell development to initiate cancer. Complementary single-cell techniques employed after *Prdm14* mis-expression in hematopoietic progenitor cells revealed a heterogeneous pool of expanded cells, including pre-T and unconventional progenitor B cells. A highly aggressive T-LL is predominant after PRDM14 expression in the BM of mature mice, yet our data show that the expansion of progenitor B cells, coupled with lower expression of DNA damage response genes, may lead to unregulated growth, accumulation of genetic mutations, and ultimately a B-LL that is masked by the early T-LL (Fig. 7). Single cell sequencing and immunophenotyping show that these cells have a signature that is similar to that seen in B-1 type leukemias. PRDM14 binds DNA adjacent to mis-regulated genes in the pre-leukemic B cells, suggesting PRDM14 has a direct role in regulating gene expression.

CyTOF analysis identified an abnormal progenitor B population that expands in the BM of R26PR;Mx1-cre mice during the 10 day post-pIpC pre-leukemic timepoint. Previous immunophenotyping analyses in pre-leukemic mice were limited by the use of targeted small flow cytometry panels[21]. Here, the immunophenotype of pre-leukemic cells was described using a single 26-parameter CyTOF panel that included HSPC, lymphoid, and myeloid lineage markers which enabled more in-depth immune profiling than was previously possible. Expanded CD127$^+$BM cells expressed high levels of the B cell marker CD19, but not B220. This phenotype is typical of early B-1 cells but is rare in wild-type BM, which contains predominantly conventional B-2 cells[36]. B-1 cells are a minor population of B cells derived from the fetal liver, which have distinct functions from B-2 cells, and are predominantly found in the intestine, peritoneal and pleural cavities in adult mice[16]. B-1 cells maintain their population throughout an organism's life through self-renewal and undergo clonal expansions, a key feature of leukemia and lymphoma[18]. In contrast, B-2 cells are progeny of BM HSCs, which are repopulated throughout life from these stem cells[17]. A population of common lymphoid progenitor-like cells that can give rise to B-1 cells has been described in 2.5 week old mouse BM, but this population decreases in adults[36].

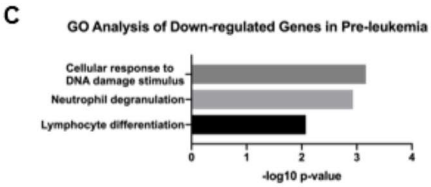
The first B-1 progenitor cells are found at mouse embryonic day (E) 8.25 in the hemogenic endothelium of the yolk sac, and soon after, in the embryonic precursor of the aorto-gonadal mesenteric (AGM) region, well before the emergence of the first fetal HSCs[16–18]. In the early embryo, *Prdm14* orchestrates genome-wide transcriptional and epigenetic reprogramming that is essential to re-establish a pluripotent cell state in PGCs

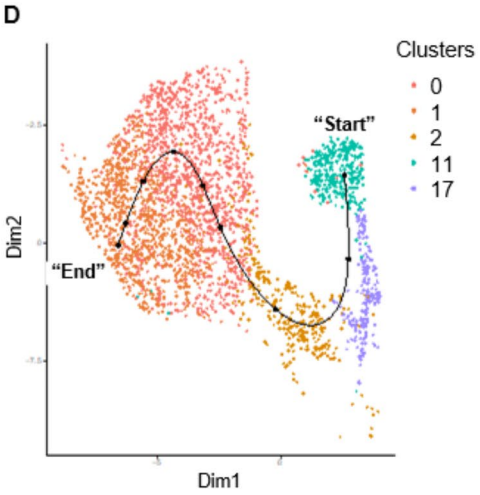


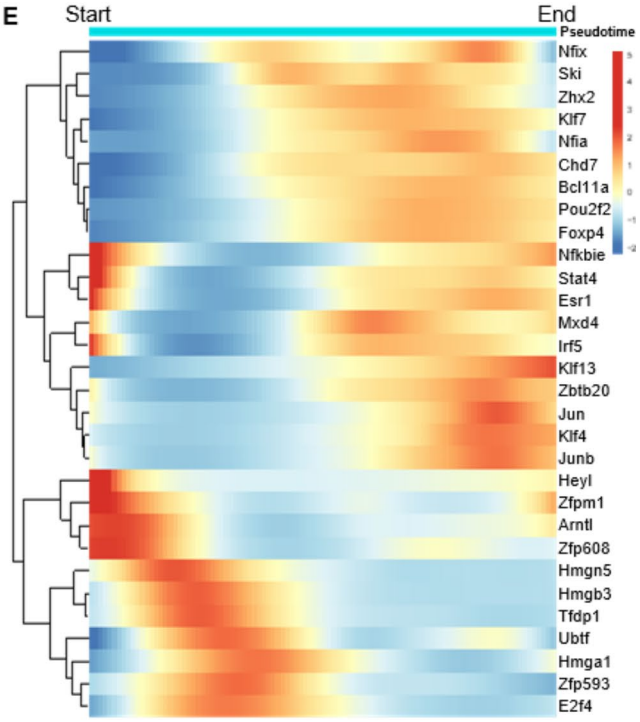
B

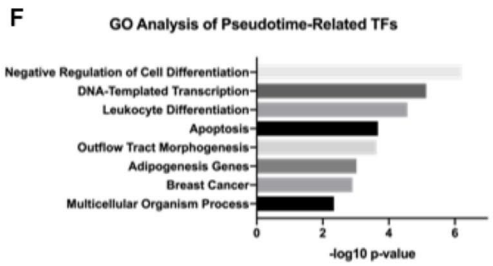
Rank (most B-1-like)	Gene symbol
1	RASGRP1
2	LEF1
3	CD19
4	POLM
5	LGR5
6	IGLL1
7	CDC25B
8	LIG4
9	POU2AF1
10	AKAP12
11	IRF4
12	VPREB3
13	BACH2
14	BCL7A
15	BCL2L1
16	CD79B
17	TRAF5
18	CERK
19	NEIL1
20	CD72

Rank (most B-2-like)	Gene symbol
1	MN1
2	FLT3
3	TYROBP
4	AHNAK
5	BST2
6	ID2
7	S100A4
8	IFITM1
9	IRF8
10	ANXA2









after their migration from the AGM to the embryonic gonad[5,40,41]. PRDM14's role in pluripotency and self-renewal raises intriguing questions about its potential normal role in the embryonic origin and self-renewal of B-1 cells as well. Further, it will be pertinent to evaluate how inducing expression of *Prdm14* in mouse HSPCs perinatally, when higher numbers of B-1 progenitor cells are present, influences B-LL initiation.

When expression is restricted to B cells, *Prdm14* can subvert B cell development. Six-month old R26FLPR;CD19-cre mice showed a striking accumulation of immature B cells, which expressed CD19 and CD24, but not B220, consistent with a B-1 cell lineage. These cells also expressed early lymphoid markers CD127 and Sca-1 and lacked mature cell markers (Supplemental Fig. 6). The similarities between the immunophenotypes of expanded immature B cells found in both the Mx1-cre and CD19-cre models was surprising given the difference in cre expression patterns. While Mx1-cre is expressed in HSPCs and more mature cell types including B cells and T cells[20,38], CD19-driven cre is not expressed until the pre-B cell stage[39]. Expression of *Prdm14* could

◀**Fig. 5**. Pseudotime analysis identifies regulators of differentiation, including AP-1 genes, that play a role in pre-leukemic B-1-cell identity. (**A**) B cell-specified common lymphoid progenitors (CLPs) progress through multiple stages of development to become immature B cells after undergoing V(D)J antigen receptor heavy (H) and light (L) chain recombination. B-2 cell development occurs in the bone marrow, whereas B-1 cell development occurs in the fetal liver. Early B-1 progenitor cells have been identified in the yolk sac as early as E8.25. Immature B-1 cells differentiate into CD5 + B-1a or CD5- B-1b cells in the periphery, predominantly in serous cavities (e.g. pleural, peritoneal cavities). Expression of defining cell surface markers are indicated below each cell type. Cell surface markers expressed during B-1 cell development are not well defined. sIgM, surface immunoglobulin M. Created with BioRender.com. (**B**) B-1 (left) or B-2 (right) signature gene sets from Fitch et al.[22]. Expression of signature genes in scRNA-seq data from *Prdm14*-expressing pre-leukemic B cells is indicated, with significant differentially expressed genes highlighted in green (up-regulated) or orange (down-regulated). (**C**) Gene Ontology analysis of down-regulated genes in sorted pre-leukemia cells relative to immunophenotype-matched controls. (**D**) Pseudotime analysis was restricted to one manifold which included B cell-associated clusters 0, 1, 2, 11, and 17. One lineage was identified, beginning at cluster 11 and ending at cluster 1. (**E**) Thirty transcription factors were significantly differentially expressed along the pseudotime axis. (**F**) GO analysis of the 30 transcription factors identified enrichment of various biological processes.

possibly reprogram this pre-B cell into an immature, pro-B-1-like cell. Reprogramming of a more mature cell type may explain the difference in disease latency between the T-ALL and B-LL that develop in the R26PR;Mx1-cre mouse model. Indeed, R26PR;Mx1-cre mice aged after pIpC injection develop T-ALL with a median survival of 41 days[20], whereas mice injected with CD19 + CD127 + Sca-1 + cells from R26PR;Mx1-cre mice have a median survival of 156 days (Fig. 3). We previously demonstrated that the T-ALL that develops in R26PR;Mx1-cre mice is driven by activating *Notch1* mutations, with factors present in normal lymphocytic precursors catalyzing RAG recombination to generate an active Notch intracellular domain[21]. It is likely that DNA damage must accumulate in progenitor B cells that mis-express *Prdm14*, allowing for the expression of additional oncogenes that would drive B-cell leukemogenesis, including the AP-1 transcription factors, along with the loss of tumour suppressors. The time it takes for B-LL to develop is more typical of leukemia development in mice, as most cancers require more than one genetic "hit" to transform a cell[1]. Thus, two steps, chromatin modifications to allow for transformation to a more progenitor-like state, and accumulation of mutations that allow for the expression of oncogenes and loss of tumor suppressor genes likely need to take place to transform B cells. These steps will likely take much more time compared to a cell that is poised to acquire a potent mutation such as constitutive *Notch1* activation. Even so, previous work suggests that *Prdm14* requires factors present in immature cells to transform B cells: first, it requires an ETO family member such as CBFA2T2 or CBFA2T3 to function and initiate T-ALL[1,43], and second, its expression in more mature T-cells cannot initiate T-ALL[20].

While the majority of mouse B-lineage leukemias are thought to be derived from B-2 cells, several studies have described B-LL with B-1 cell origin. One study demonstrated that B-1 cells are more susceptible to transformation by the BCR-ABL1 oncogenic translocation than B-2 cells and can initiate disease more rapidly[18]. It has been proposed that the innate self-renewal capacity of B-1 cells contributes to a predisposition for transformation[37]. *Prdm14* mis-expression has previously been suggested as a mechanism for cancer initiation due to its powerful reprogramming ability[1]. In human fibroblasts, *PRDM14* expression enhances the reprogramming efficiency of cells into iPSCs and can replace the Yamanaka factor KLF4 entirely[35]. A similar mechanism has been reported in adult BM-derived HSPCs subjected to ectopic *Lin28b* expression, which become reprogrammed into progenitor cells able to recapitulate fetal-like hematopoiesis, including B-1 cell specification[42]. Sorted *Prdm14*-expressing progenitor B cells are functional cancer initiating cells, as they initiate cancer when transferred to host mice. Interestingly, a *NUP98-PHF23* transgenic mouse model of progenitor B-1 ALL has an immunophenotype similar to the B-1-like cells observed in R26PR;Mx1-cre mice[31]. In the model, pro-B-1 ALL cells were Lineage$^-$ CD93$^+$ CD19$^+$ B220$^{lo/-}$. Although CD93 was not included in our CyTOF analysis, we saw clear expression of CD93 in sorted pre-leukemic B cells by scRNA-seq. Few B-1-specific genes have been described in the literature, but lack of B220 expression combined with expression of *Foxo1*, *Cd93*, and *Siglecg* are consistent with a B-1-like gene expression signature[31,32]. Additionally, when the gene expression profile of PRDM14-expressing pre-leukemic B cells was compared to published human B-1 leukemia gene signatures, there was a 45% overlap (Fig. 4C). Conversely, there was no overlap with the B-2 leukemia signature genes, suggesting the mouse pre-leukemic B cells are more B-1-like in nature[22].

Sorted progenitor B cells in R26PR;Mx1-cre mice at the pre-leukemic timepoint overexpressed proto-oncogenes including *Myb*, *Cbfa2t3*, and AP-1 genes relative to sorted progenitor B cells in R26PR control mice. Our previous work demonstrated the dependency of PRDM14-expressing hematopoietic tumours on *Cbfa2t3*: absence of *Cbfa2t3* was sufficient to prevent the development of any type of leukemia in these mice[43]. These data suggest that *Cbfa2t3* is required for both B and T cell LL initiation when *Prdm14* is mis-expressed. AP-1 family members play important roles in cellular proliferation, differentiation, apoptosis, oncogenesis and inflammation[28]. Overexpression of *Fos* is associated with osteosarcomas and endometrial carcinoma in humans, and *Jun* overexpression is linked to aggressive lymphomas[28]. Interestingly, *Fos* overexpression has been implicated in murine B-1 cell proliferation. Forced *Fos* expression from a B cell-specific promoter led to a 3-to-4-fold expansion of B-1 cells[44]. Thus, the overexpression of *Fos* could partly explain the expansion of B-1-like cells in *Prdm14*-expressing BM cells.

The sorted pre-leukemic cells also displayed decreased expression of DNA damage response genes. Genomic instability has already been associated with *Prdm14* mis-expression in a mouse BM transduction leukemia model[8]. Mice who received *Prdm14*-transduced BM succumbed to leukemias of pre-B and pre-T cell origin. These

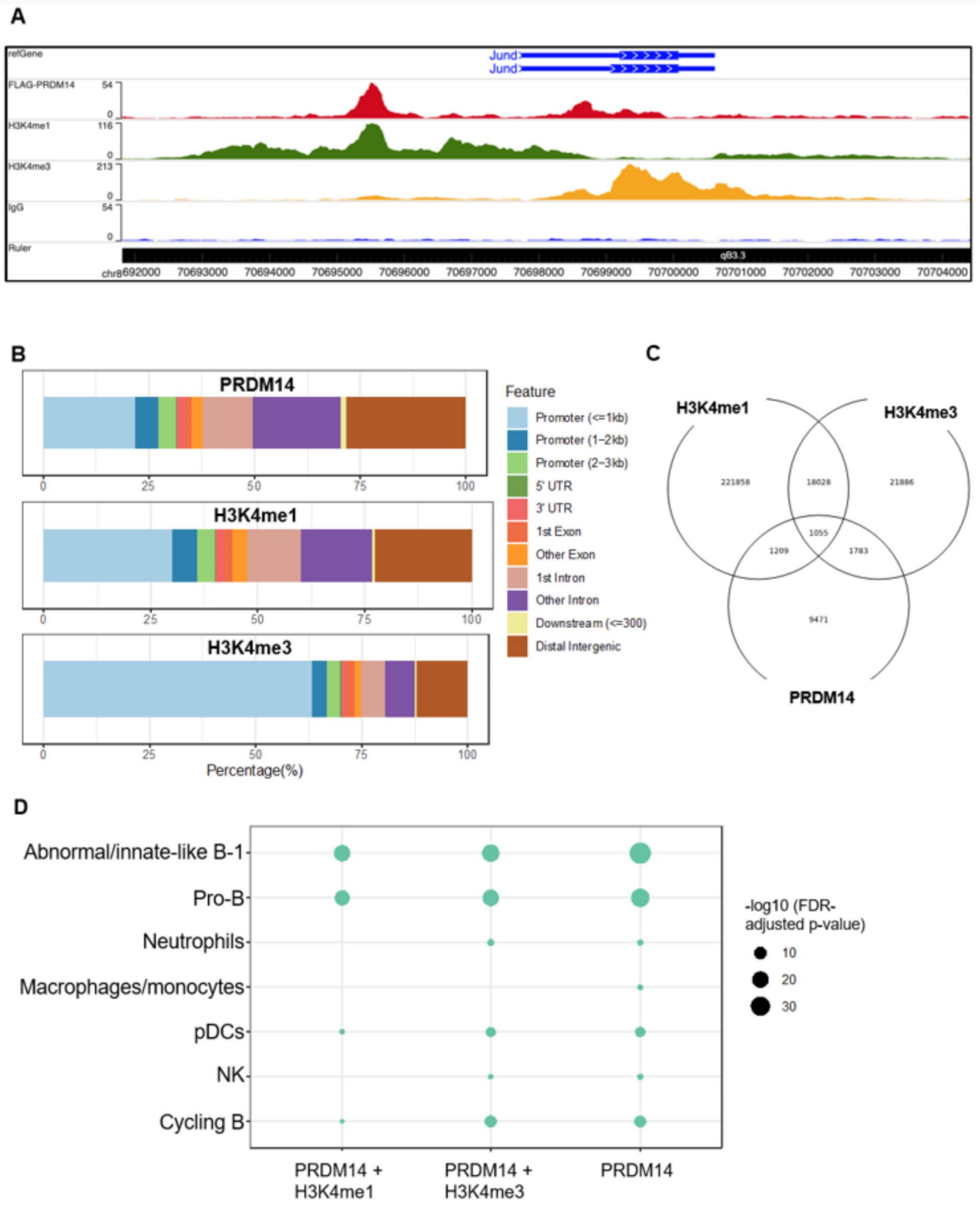


Fig. 6. CUT&RUN identifies PRDM14 binding sites that overlap with methylated H3K4 and show enrichment for genes mis-regulated in pre-leukemic pro-B and B-1-like cells. (**A**) Snapshot of CUT&RUN results visualized on the WashU Epigenome Browser at the *Jund* locus. Each track represents the mapped reads from a given antibody's immunoprecipitation. An IgG antibody was used as a negative control. The Ruler track shows the genomic position on mouse chromosome 8. (**B**) Mapped reads were annotated based on their position relative to an adjacent gene. Untranslated region is abbreviated UTR. (**C**) Pie chart showing the overlap of peaks between three antibodies. (**D**) CUT&RUN peak-adjacent genes were overlapped with marker genes that defined scRNA-seq clusters. Enrichment (FDR-adjusted p-value) is indicated by the size of the circle for PRDM14 peaks alone or combined with H3K4me1/3 and select clusters. The abnormal/innate-like B-1 clusters' marker genes had the strongest enrichment of all clusters examined for genes bound by PRDM14 within 10 kb of their TSS. Pro-B cells also showed high enrichment.

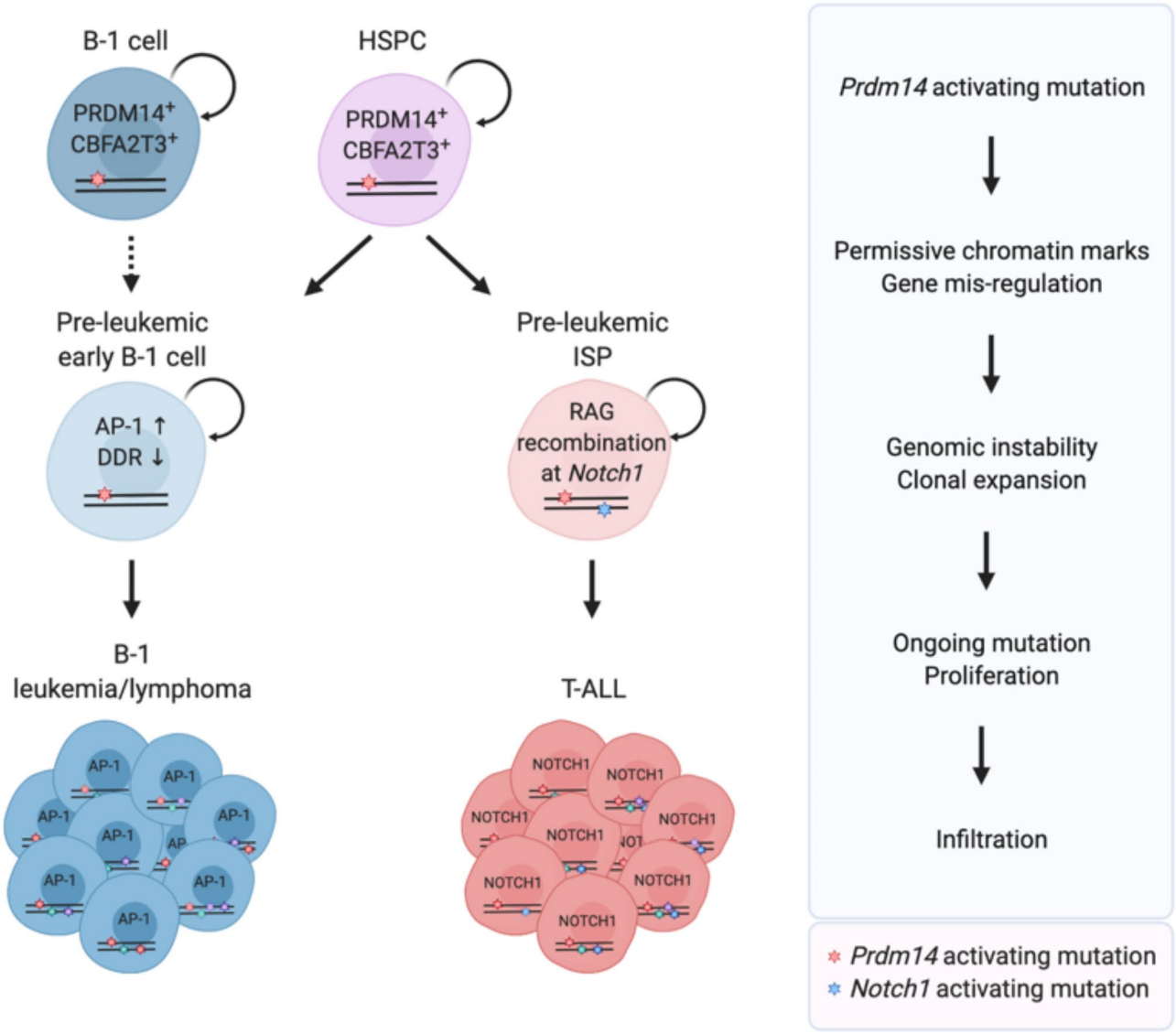


Fig. 7. Model of *Prdm14*-mediated cancer initiation. Upon mis-expression in adult HSPCs, PRDM14 interacts with CBFA2T3 to establish cancer initiating cells that initiate leukemia. This complex could promote permissive chromatin modifications (e.g. H3K4me1/3), gene mis-regulation and genomic instability through recruiting chromatin modifiers to PRDM14 target genes. RAG-dependent *Notch1* mutations are acquired in the thymus. In parallel, pre-leukemic B-1-like cells overexpress proto-oncogenes, such as AP-1 genes, while mis-regulating DNA damage response genes and begin expanding in the BM. Cancer initiation involves multiple downstream steps, including continual clonal expansion and DNA damage leading to uncontrolled tumor growth. Both pre-leukemic cell types initiate infiltrative malignancies. *Prdm14* expression in a mature B-1 cell may promote reprogramming directly into a pre-leukemic B-1 cell but has not been tested. CBFA2T3, which is expressed in mature B-1 cells, would be poised to interact with PRDM14 when it becomes aberrantly expressed. Stars represent individual mutations. +, gene and protein expression present; DDR, DNA damage response; HSPC, hematopoietic stem and progenitor cell; ISP, immature single positive; T-ALL, T cell acute lymphoblastic leukemia. Created in BioRender.com.

tumours had acquired numerous amplifications and deletions, with deletions recurring in tumour suppressor and B cell development genes. Microarray of BM cells prior to leukemia onset showed that genes involved in chromosomal stability and the DNA damage response, such as *Xrcc6*, were expressed at lower levels[8], consistent with the decreased expression of *Xrcc6* found here. A failure to upregulate the DNA damage response pathway is a critical component of tumorigenesis as it results in the accumulation of additional mutations that can drive tumour growth, subvert apoptosis, and block differentiation[45]. Increased genomic instability in R26PR;Mx1-cre mice contributes to the accumulation of activating *Notch1* mutations that drive rapid T-ALL onset[21]. The impact on DNA damage repair could explain why *PRDM14* is so frequently amplified and overexpressed in multiple human cancers[1]. Furthermore, high expression of *PRDM14* is correlated with worse prognoses in breast cancer and non-small cell lung carcinoma[11,46], and is associated with higher-grade carcinomas in cervical cancer[47].

That *Prdm14* can initiate two types of cancer in a single transgenic mouse model suggests that this powerful pluripotency regulator should be examined further for its roles in cancer initiation and progression.

Methods

Mouse strains/animal care

Mice were primarily housed at The Centre for Phenogenomics (TCP) except for the mice used for intra-femoral adoptive transfer experiments, which were housed at the Hospital for Sick. Children Laboratory Animal Science (LAS) facility. Both facilities are accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International. All experimental protocols, strains and breeding strategies were pre-approved by the SickKids Animal Care Committee in accordance with the Canadian Council on Animal Care guidelines, and all methods were carried out in strict adherence to these guidelines. The results of experiments reported here adhere to the reporting requirements of the ARRIVE guidelines. Mice were housed in a pathogen-free environment and were maintained on a C57BL/6J congenic genetic background: B6.*Gt(ROSA)26Sor*$^{tm1(LSL-Prdm14-IRES-EGFP)Jus}$, abbreviated "R26PR", B6.*Gt(ROSA)26-Sor*$^{tm(LSL-3XFLAG-Prdm14-P2A-EGFP)Jus}$, abbreviated "R26FLPR", B6.Tg(*Mx1*-cre)1Cgn/J abbreviated "Mx1-cre" (from Dr. Margaret A. Goodell, Baylor College of Medicine, Houston, Texas, USA), *B6.129P2(C)-Cd19*$^{tm1(cre)Cgn}$*/J*, abbreviated "CD19-cre" (purchased from The Jackson Laboratory), *B6(Cg)-Rag2*$^{tm1.1Cgn}$*/J*, abbreviated "*Rag2*$^{-/-}$ or "*Rag2*$^{-/-}$ (CD45.2)" (purchased from The Jackson Laboratory), *B6.129S6-Rag2*[1] *N12* mice (purchased from Taconic) expressing the B6 allele of *Ptprca* encoding CD45.2 were back-crossed to *B6.SJL*$^{Ptprca\ Pepcb}$*/BoyJ* (JAX002014) mice to generate *B6.Rag2*$^{-/-}$ mice expressing only the SJL allele of *Ptprca* encoding CD45.1 or expressing both CD45.1 and CD45.2. The latter strains are referred to as *B6.Rag2.CD45.1* and *B6.Rag2.CD45.1/CD45.2*, respectively. Polyinosinic:polycytidylic acid (pIpC) administration to activate Mx1-cre was performed at 8 weeks of age with one intraperitoneal injection of 250µg.

All mice were monitored for signs of cancer including enlarged lymph nodes and abdomen, lethargy, and laboured breathing. Mice were euthanized via CO2 inhalation. Tissues from moribund mice were immediately dissected and processed for protein, flow cytometry, and/or blood analyses. Standard complete blood count (CBC) analysis was carried out by the Clinical Phenotyping Core of The Centre for Phenogenomics. Samples for pathology analysis were fixed in 10% neutral buffered formalin (NBF) overnight. Tissues were paraffin embedded and sectioned, and slides stained with hematoxylin and eosin (H&E). Images were acquired with an Eclipse E100 (Nikon: Tokyo, Japan) microscope and DS-Qi1Mc (Nikon) camera.

CyTOF

Single-cell suspensions of BM from pIpC-injected R26PR; Mx1-cre and R26PR (control) mice were treated with red blood cell lysis solution (Miltenyi Biotec: Bergisch Gladbach, Germany) according to the manufacturer's instructions. Lysed BM cells ($1–2 \times 10^6$ cells per mouse) were stained with a cocktail of 23 metal-tagged antibodies (see Table 1) and analyzed on a Helios instrument with a wide-bore injector according to the manufacturer's protocols as previously described[48]. The Helios software (v6.7.1014) was used to generate and normalize FCS 3.0 data files, which were then uploaded to CytoBank (Beckman Coulter Enterprise license) where manual pre-gating was performed to remove debris, dead cells, doublets, and Ly6G$^-$cells using standard techniques. After performing dimensionality reduction using the t-Distributed Stochastic Neighbor Embedding (tSNE) algorithm (iterations = 3,000; perplexity = 30; theta = 0.5; equal sampling of 95,000 cells yielding Kullback-Leibler divergence of 5.9)[23], we clustered (hierarchical consensus) the tSNE1 and tSNE2 parameters using FlowSOM as described[24] to produce 25 metaclusters (MC) from 256 clusters after 10 iterations. We then compared MC abundance between *Prdm14*-induced and control groups using an unpaired t-test. P-values were then corrected for multiple testing using the Benjamini and Hochberg False Discovery Rate (FDR) method in Prism (GraphPad). Significance was determined with an FDR-adjusted P-value (Q-value) < 0.05).

Flow cytometry

Single-cell suspensions from thymus, spleen, and BM ($1–2 \times 10^6$ cells per sample) were treated with 5 µg/ml of anti-FcRII/FcRIII (2–4.G2) to block Fc receptors, prior to staining with fluorochrome-conjugated antibodies for 30' at RT in the dark (see Table 2), as previously described[49]. Samples were analyzed on a LSR II CFI (BD) or LSR Fortessa (BD) flow cytometer in the Flow Cytometry Facility at the SickKids' Research Institute. For sorting experiments, all cell preparations were done under sterile conditions. To enrich for HSPC, CD11b$^+$ cells were magnetically depleted using an autoMACS (Miltenyi Biotec) per manufacturer's instructions. Cells were sorted on a MoFlo Astrios (Beckman Coulter: Brea, California, United States) in the Flow Cytometry Facility at the Hospital for Sick Children. Data collection and analyses were performed with FACSDiva (BD Biosciences) and FlowJo v10 (BD) software. In some cases, fluorescence minus one (FMO) controls were used to set gates for markers with a continuum of expression.

Adoptive transfers

Mice were anesthetized by isoflurane inhalation before intra-femoral and tail-vein injections. For adoptive transfer experiments involving intra-femoral injection, viable CD11b$^-$ CD8β$^-$ GFP$^+$ Sca-1$^+$ CD19$^+$ CD127$^+$ BM cells (referred to as GFP$^+$ CD19$^+$ Sca-1$^+$ CD127$^+$ for simplicity) were sorted from three R26PR;Mx1-cre CD45.2$^+$ donor animals 8 days post-pIpC. Fifty thousand cells were co-injected together with 10^5 *B6.Rag2.CD45.1/CD45.2* cells (referred to as Rag2 KO). BM cells were injected orthotopically into the BM cavity of the right femurs of sub-lethally irradiated (650-cGy) *B6.Rag2.CD45.1* host mice as described[50]. Total BM cells (5×10^5) from each R26PR;Mx1-cre donor mouse were mixed with 10^5 Rag2 KO BM cells and co-injected into one host mouse per experimental donor mouse for comparison. All mice were euthanized 4 weeks after intra-femoral injection via CO2 inhalation and the right femur, distant BM (left femur and both tibias), and thymus were harvested immediately for flow cytometry.

Metal Tag	Markers	Clone	Manufacturer	Catalogue #
89Y	CD45	30-F12	Fluidigm (South San Francisco, California, United States)	3089005B
115In	Ly6C	1A8	Biolegend (San Diego, California, United States)	127,602
141Pr	CD44	IM7	Biolegend	103,002
143Nd	CD4	RM4-5	Biolegend	100,561
144Nd	MHC2	M5/114.15.2	Biolegend	107,637
147Sm	CD5	53 – 7.3	Inhouse hybridoma	
148Nd	CD11b	M1/70	Biolegend	101,202
149Sm	CD19	1D3	BD (Franklin Lakes, New Jersey)	553,783
150Nd	CD24	M1/69	Biolegend	101,829
151Eu	IgM	II/41	BD	553,435
152Sm	CD3	145-2C11	Biolegend	100,331
154Sm	CD22	Cy34.1	Inhouse hybridoma	
155Gd	CD16/32	2.4G2	Inhouse hybridoma	
160Gd	SiglecF	E50-2440	BD	552,125
162Dy	CD135	A2F10.1	eBioscience (San Diego, California, United States)	14–1351-82
163Dy	CD8	H35-17.2	eBioscience	14–0083–85
164Dy	Sca-1	E13-161.7	Biolegend	122,502
166Er	CD117	2B8	Fluidigm	3166004B
167Er	Ly6G	RB6-8C5	Biolegend	108,402
169Er	TCR	H57-597	Biolegend	109,235
170Er	CD25	PC61.5	Biolegend	102,002
175Lu	CD127	A7R34	Biolegend	135,029
176Yb	B220	RA3-6B2	Biolegend	103,249
191Ir	DNA1		Fluidigm	201192B
193Ir	DNA2		Fluidigm	201192B
195Pt	Viability		Fluidigm	201,064

Table 1. Antibodies used for the CyTOF experiments.

For adoptive transfer experiments involving intravenous injection, 50,000 viable GFP$^+$ CD19$^+$ CD127$^+$ BM cells were sorted from four R26PR;Mx1-cre (CD45.2) mice 8 days post-pIpC and injected into the tail vein of sub-lethally irradiated (650-cGy) anesthetized *Rag2*$^{-/-}$ host mice using a 28-gauge micro-fine insulin syringe. Host mice were weighed twice a week and euthanized when signs of extreme lethargy and cachexia became apparent. An equal number of male and female mice were used for this experiment.

Single cell RNA-sequencing

The 10X Genomics Chromium platform was used for single cell RNA-sequencing (scRNA-seq) at the Princess Margaret Genomics Centre and all steps were performed at the same time for all samples. Briefly, determination of the viability and density of sorted cells were carried out by a Countess Automated Cell Counter (ThermoFisher Scientific). Capture and encapsulation of cells (approximately 2000 per sample) was followed by cDNA amplification and library preparation for sequencing. Deep sequencing was performed on the Illumina NextSeq 500. Following sequencing, raw FASTQ files were aligned to the appropriate genome (mm10) using the STAR aligner (STAR v2.5.2b), which aligns reads simultaneously to the genome and the transcriptome. Both types of alignments were used to accurately determine whether a read could be confidently associated with a transcript and/or a gene. Accessory programs for the alignment stage include SAMTOOLS (v1.3.1) and BEDTOOLS (v2.26.0). The final number of cells per sample used for downstream analysis were: PL GFP$^+$, 1846; PL GFP$^-$, 2783; CD11b$^-$ CD19$^+$ CD127$^+$ Sca-1$^+$ PL B, 2798 (sorted from R26PR; Mx1-cre BM 8 days post-pIpC injection); and CD11b$^-$ CD19$^+$ CD127$^+$ Sca-1$^+$ Ctrl B, 1974 (sorted from R26PR control BM).

The CELLRANGER (v3.0.2) pipeline was used to obtain two types of gene-barcode matrices. The first matrix was an unfiltered gene-barcode matrix. This matrix contained every barcode from the fixed list of known barcode sequences, including background and non-cellular barcodes. The next matrix type was the filtered gene-barcode matrix. The filtered matrix contained only the detected cellular barcodes. In the final filtered matrix, each row corresponded to a gene, and each column corresponded to a cell barcode sequence. The filtered gene barcode matrix was loaded into R using the "Read10X" function in Seurat (v 4.0) and processed using the "process_dgTmatrix_lists" function in scMappR, which functions as a wrapper for the Seurat-SCTransform pipeline. We chose to integrate across datasets with the "integration anchors" pipeline to avoid batch correction, and because each scRNA-seq run had diverse cell compostions. Mitochondrial content was calculated for every cell, and cells with abnormally high mitochondrial content (MT) using a mean (%MT) + 2sd (%MT) were filtered. Then, counts were normalized using scTransform while controlling for mitochondrial content. Principal Component Analysis (PCA) was computed based on all variable genes and dimensionality reduction based on the Uniform Manifold Approximation Projection algorithm (UMAP) was completed using the first 20 PCs. Cell-type markers

Antibody	Manufacturer	Clone
7-AAD	BioLegend	N/A
B220-APC	BioLegend	RA3-6B2
B220-APC-Cy7	BioLegend	RA3-6B2
B220-BV605	BioLegend	RA3-6B2
CD117-APC-eFluor780	eBioscience	2B8
CD11b-Biotin	eBioscience	M1/70
CD11b-BV711	BD Biosciences	M1/70
CD127-BV421	BioLegend	A7R34
CD19-BUV395	BD Pharmingen	1D3
CD19-PE	eBioscience	MB19-1
CD19-PE-Cy7	BD Biosciences	1D3
CD24-BV480	BD Biosciences	M1/69
CD25-BV786	BD Biosciences	PC61.5
CD4-BUV737	BD Biosciences	GK1.5
CD43-PE	eBioscience	eBioR2/60
CD45.1-PE-Cy7	eBioscience	A20
CD45.2-APC-eFluor780	eBioscience	104
CD5-APC	eBioscience	53 – 7.3
CD5-BV605	BD Biosciences	53 – 7.3
CD8-APC-eFluor780	Thermo Fisher Scientific (Waltham, Massachusetts, United States)	53 – 6.7
CD8b-PerCP-eFluor710	eBioscience	H35-17.2
GL7-PerCP-Cy5.5	BioLegend	GL7
IC-IgM-APC	eBioscience	II/41
IgD-PE-Cy7	BioLegend	11–26 C.2 A
LIVE/DEAD™ Fixable Blue	Thermo Fisher Scientific	N/A
PI	Millipore Sigma (Burlington, Massachusetts, United States)	N/A
Sca1-PE	BD Pharmingen	E13-161.7
Sca1-PE-Cy7	BioLegend	E13-161.7
sIgM-APC	eBioscience	II/41
sIgM-BV650	BD Biosciences	R6-60.2
TCRβ-PE-Cy7	BD Biosciences	H57-597
Zombie Aqua	BioLegend	N/A
Zombie UV	BioLegend	N/A

Table 2. Antibodies used for the flow cytometry experiments.

were measured using the Wilcoxon's test in the "FindMarkers" function in Seurat V4. Lastly, cell-type labels were predicted using multiple gene-set enrichment methods against the PanglaoDB and CellMarker databases. While cell-type label predictions helped inform the final cell-type assignment, cell-type assignment was also determined based on a literature review of the top markers within each cell-type.

These scRNA-seq data were leveraged to identify cell lineage trajectories towards the pre-leukemic B cell population using the "slingshot" R package and function. Pseudotime trajectories were computed using default petameters except for setting the "extend" variable to "n". Lineages were traced from using all B cells and abnormal cell-types (clusters 0, 1, 2, 11 and 17), and cluster 11 was selected as the starting cluster. For each lineage, the tradeSeq R package was used to identify which genes have cell-type specific expression patterns associated with that lineage. Specifically, the "evaluateK" function was used to measure the minimum number of knots required to accurately fit a negative-binomial general adversarial model (nb-GAM). An nb-GAM was generated for every gene and lineage using the "fitGAM" function. With each model fit, the association of each gene with pseudotime was tested using the "associationTest" function (FDR-adjusted p-value < 0.05, absolute fold-change > 1.5). The smoothed gene expression patterns of associated genes were plotted by applying the "predictSmooth" function to the normalized scRNA-seq data and plotting these smoothed scRNA-seq data using the "pheatmap" package and function. Individual genes were plotted using the "FeaturePlot" and "VlnPlot" functions within Seurat. Pathway enrichment of all gene sets was completed using gprofiler, with FDR correction and all detected genes as the genomic background.

CUT&RUN

$CD8\beta^{-}$ $CD19^{+}$ $CD127^{+}$ $Sca\text{-}1^{+}$were cells sorted from R26FLPR; Mx1-cre or R26FLPR control mice 8 days post-pIpC and Cleavage Under Targets and Release Using Nuclease (CUT&RUN) was carried out as previously described[51]. Cleaved DNA was extracted with phenol-chloroform, and libraries were generated using the

NEBNext® Ultra II DNA Library Prep Kit (New England BioLabs: Ipswich, Massachusetts, United States) per manufacturer's instructions.

Libraries were sequenced at the Ontario Institute for Cancer Research (OICR) on an Illumina NovaSeq S1 using 2 × 51 bp paired-end sequencing to a depth of 10 million reads per sample. Fastq files were checked for initial quality scores with fastqc. Cut-and-run reads were processed based on methods (https://github.com/Henikoff/Cut-and-Run)[52]. Trimmed reads were aligned to the mouse GRCm38 genome and sacCer3 genome using bowtie2 with the "very sensitive local", "no mixed", "no discordant", "no unal", "p", and "x" parameters. Reads aligning to both genomes were filtered to not be part of spike-in calibration. After converting aligned .sam files to bed files using "samtools" and "bedtools" the "spike_in_calibration.csh" script was applied, with a scale of 1000, minimum length of 50, and maximum length of 800. Calibrated bedGraph files outputted by spike_in_calibration.csh were then converted to bigwig files using the "bedGraphToBigWig" function in "kentUtils" with default parameters.

Spike-in calibrated bigwig files were used to call peaks using the "derfinder" R package. Bigwig signal files were enumerated using the "fullCoverage" function with default parameters and the "regionMatrix" function with a "cutoff = 5" and "L = 30". Correlation matrices of spike-in and edgeR-log2(CPM) normalized counts were generated using the "rcorr" function in the "Hmisc" R package and plotted using the "ComplexHeatmap" R package. Peaks were computed using edgeR, where the average expression of each counted region was compared to the IgG negative control. In the "DGEList" function in edgeR, the "lib.size" parameter was fixed to be the average library size across the IgG and antibody (i.e., Prdm14, H3K4me1, H3K4me3), as there will be inherently fewer reads sequenced in the IgG. With the edgeR object generated and library sized fixed to account for calling peaks, differentially bound regions were computed between each antibody and the IgG using the "model.matrix", "estimateDisp", "glmQLFit", and "glmQLFTest" functions in edgeR (FDR-adjusted p-value < 0.05, |fold-change| > 2). Volcano plots and MA plots were generated using "ggplot2".

Peaks called for Prdm14, H3K4me1, and H3K4me3 were overlapped using "intervene" using a 15% reciprocal overlap as a cutoff. Peaks were annotated to genes against the "TxDb.Mmusculus.UCSC.mm10.knownGene" dataset using the "annotatePeak" function in the ChIPseeker R package using default parameters and a promoter cutoff of 3 kb from the transcription start site. The distribution of genomic regions was plotted using the "plotAnnoBar" function in Chipseeker using default parameters. Next, the enrichment of PRDM14 peaks, H3K4me1 peaks, and H3K4me3 peaks were computed against cell-type markers detected in our scRNA-seq data (FDR-adjusted p-value < 0.05, pct cell-type expression > 25%). The enrichment of peaks was computed onto each cell type using the "cellmarker_enrich" function in the scMappR R package, which uses a fisher's exact test between peaks from cut-and-run and the cell-type markers computed our scRNA-seq data. Enriched cell-types were ranked by odds ratio and p-value before being plotted as dotplots using ggplot2.

Western blot

Protein lysates were resolved on a 4–12% polyacrylamide gradient gel (Bio-Rad: Hercules, California, United States) and transferred to a polyvinylidene difluoride (PVDF) membrane (Bio-Rad) for blotting. Membranes were blocked in 5% skim milk for 1 h and subsequently incubated with primary antibodies diluted in 5% skim milk for 2 h. After washing with PBST, membranes were incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies diluted in 5% skim milk for 1 h. Membranes were washed again with PBST, and chemiluminescence Clarity Western ECL Substrate (Bio-Rad) was added for signal detection with Amersham Hyperfilm ECL (GE Healthcare: Chicago, Illinois, United States). The following primary antibodies were used: FLAG (M2, Sigma), GAPDH (D16H11, Cell Signaling Technology: Danvers, Massachusetts, United States), Cleaved Notch1 (Val1744, Cell Signaling Technology). The following secondary antibodies were used: Peroxidase-AffiniPure Goat Anti-Mouse IgG Light Chain Specific (Jackson ImmunoResearch: West Grove, Pennsylvania, United States), Peroxidase-AffiniPure Goat Anti-Rabbit IgG (H + L) (Jackson ImmunoResearch).

Data analysis & statistics

For mouse survival analyses, Kaplan-Meier curves were compared between two groups using a log-rank (Mantel-Cox) test for statistical significance in Prism (GraphPad, version 9.3). An unpaired t-test was performed to identify significant changes between two groups in quantified flow cytometry, complete blood count, and organ weight datasets in Prism (GraphPad). Unpaired t-tests corrected for multiple testing in Prism (GraphPad) were used to determine the FDR-adjusted p-value (q-value) when more than two groups were being compared. The WashU Epigenome Browser (v53.8.0) was used to visualize CUT&RUN results (http://epigenomegateway.wustl.edu/browser/)[53].

Data availability

The CUT&RUN datasets generated and analysed during the current study are available in the Array Express repository (E-MTAB-13159), "https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13159". The scRNA-seq datasets generated and analysed during the current study are available in the Array Express repository (E-MTAB-13160), "https://www.ebi.ac.uk/biostudies/arrayexpress/studies/E-MTAB-13160". All other datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 30 December 2024; Accepted: 4 March 2025
Published online: 14 March 2025

References

1. Tracey, L. J. & Justice, M. J. Off to a bad start: Cancer initiation by pluripotency regulator PRDM14. *Trends Genet.* 1–12. https://doi.org/10.1016/j.tig.2019.04.004 (2019).
2. Ma, Z., Swigut, T., Valouev, A., Rada-Iglesias, A. & Wysocka, J. Sequence-specific regulator Prdm14 safeguards mouse ESCs from entering extraembryonic endoderm fates. *Nat. Struct. Mol. Biol.* **18**, 120–128 (2011).
3. Tsuneyoshi, N. et al. PRDM14 suppresses expression of differentiation marker genes in human embryonic stem cells. *Biochem. Biophys. Res. Commun.* **367**, 899–905 (2008).
4. Chan, Y. S. et al. A PRC2-Dependent repressive role of PRDM14 in human embryonic stem cells and induced pluripotent stem cell reprogramming. *Stem Cells*. **31**, 682–692 (2012).
5. Yamaji, M. et al. Critical function of Prdm14 for the establishment of the germ cell lineage in mice. *Nat. Genet.* **40**, 1016–1022 (2008).
6. Dettman, E. J. et al. Prdm14 initiates lymphoblastic leukemia after expanding a population of cells resembling common lymphoid progenitors. *Oncogene* **30**, 2859–2873 (2011).
7. Dettman, E. J. & Justice, M. J. The zinc finger SET domain gene Prdm14 is overexpressed in lymphoblastic lymphomas with retroviral insertions at Evi32. *PLoS One*. **3**, 1–9 (2008).
8. Simko, S. J., Voicu, H., Carofino, B. L. & Justice, M. J. Mouse lymphoblastic leukemias induced by aberrant Prdm14 expression demonstrate widespread copy number alterations also found in human ALL. *Cancers (Basel)*. **4**, 1050–1066 (2012).
9. Nishikawa, N. et al. Gene amplification and overexpression of PRDM14 in breast cancers. *Cancer Res.* **67**, 9649–9657 (2007).
10. Ruark, E. et al. Identification of nine new susceptibility loci for testicular cancer, including variants near DAZL and PRDM14. *Nat. Genet.* **45**, 686–689 (2013).
11. Zhang, T. et al. High expression of PRDM14 correlates with cell differentiation and is a novel prognostic marker in resected non-small cell lung cancer. *Med. Oncol.* **30**, 1–7 (2013).
12. Li, X. et al. Integrated analysis of the role of PR/SET domain 14 in gastric cancer. *BMC Cancer*. **24**, 1–13 (2024).
13. Weng, A. P. et al. Activating mutations of NOTCH1 in human T cell acute lymphoblastic leukemia. *Sci. (80-).* **306**, 269–271 (2004).
14. Ashworth, T. D. et al. Deletion-based mechanisms of Notch1 activation in T-ALL: key roles for RAG recombinase and a conserved internal translational start site in Notch1. *Blood* **116**, 5455–5464 (2010).
15. Aster, J. C. et al. Essential roles for Ankyrin repeat and transactivation domains in induction of T-Cell leukemia by Notch1. *Mol. Cell. Biol.* **20**, 7505–7515 (2000).
16. Yoshimoto, M. et al. Embryonic day 9 yolk sac and intra-embryonic hemogenic endothelium independently generate a B-1 and marginal zone progenitor lacking B-2 potential. *Proc. Natl. Acad. Sci. U S A*. **108**, 1468–1473 (2011).
17. Baumgarth, N. A Hard(γ) look at B-1 cell development and function. *J. Immunol.* **199**, 3387–3394 (2017).
18. Montecino-Rodriguez, E., Li, K., Fice, M. & Dorshkind, K. Murine B-1 B cell progenitors initiate B-Acute lymphoblastic leukemia with features of High-Risk disease. *J. Immunol.* **192**, 5171–5178 (2014).
19. Hardy, R. R. B-1 B cell development. *J. Immunol.* **177**, 2749–2754 (2006).
20. Carofino, B. L., Ayanga, B. & Justice, M. J. A mouse model for inducible overexpression of Prdm14 results in rapid-onset and highly penetrant T-cell acute lymphoblastic leukemia (T-ALL). *Dis. Model. Mech.* **6**, 1494–1506 (2013).
21. Carofino, B. L., Ayanga, B., Tracey, L. J., Brooke-Bisschop, T. & Justice, M. J. PRDM14 promotes RAG-dependent Notch1 driver mutations in mouse T-ALL. *Biol. Open.* **5**, 645–653 (2016).
22. Fitch, B. et al. Human pediatric B-cell acute lymphoblastic leukemias can be classified as B-1 or B-2-like based on a minimal transcriptional signature. *Exp. Hematol.* **90**, 65–71 (2020).
23. Amir, E. D. et al. ViSNE enables visualization of high dimensional single-cell data and reveals phenotypic heterogeneity of leukemia. *Nat. Biotechnol.* **31**, 545–552 (2013).
24. Van Gassen, S. et al. FlowSOM: using self-organizing maps for visualization and interpretation of cytometry data. *Cytom Part. A.* **87**, 636–645 (2015).
25. Suliman, S. et al. Notch3 is dispensable for thymocyte β-selection and notch1-induced T cell leukemogenesis. *PLoS One* **6**, 1–13 (2011).
26. Zhou, B. et al. Notch signaling pathway: architecture, disease, and therapeutics. *Signal. Transduct. Target. Ther.* **7**, 1–33 (2022).
27. Heng, T. S. P. et al. The immunological genome project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
28. Gazon, H., Barbeau, B., Mesnard, J. M. & Peloponese, J. M. Hijacking of the AP-1 signaling pathway during development of ATL. *Front. Microbiol.* **8**, 1–13 (2018).
29. Van Dam, H. & Castellazzi, M. Distinct roles of Jun:Fos and Jun:ATF dimers in oncogenesis. *Oncogene* **20**, 2453–2464 (2001).
30. Sokolowski, D. J. et al. Differential expression enrichment tool (DEET): an interactive atlas of human differential gene expression. *NAR Genomics Bioinforma.* **5**, 1–17 (2023).
31. Gough, S. M. et al. Progenitor B-1 B-cell acute lymphoblastic leukemia is associated with collaborative mutations in 3 critical pathways. *Blood Adv.* **1**, 1749–1759 (2017).
32. Montecino-Rodriguez, E. et al. Distinct genetic networks orchestrate the emergence of specific waves of fetal and adult B-1 and B-2 development. *Immunity* **45**, 527–539 (2016).
33. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
34. Sybirna, A., Tang, W. W. C., Dietmann, S., Gruhn, W. H. & Surani, M. A. A critical but divergent role of PRDM14 in human primordial germ cell fate revealed by inducible degrons results. *Nature Communication* **11**, 1–18 (2019).
35. Chia, N. Y. et al. A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
36. Barber, C. L., Montecino-Rodriguez, E. & Dorshkind, K. Reduced production of B-1-specified common lymphoid progenitors results in diminished potential of adult marrow to generate B-1 cells. *Proc. Natl. Acad. Sci.* **108**, 13700–13704 (2011).
37. Hayakawa, K. et al. Early generated B-1–Derived B cells have the capacity to progress to become mantle cell Lymphoma–like neoplasia in aged mice. *J. Immunol.* **201**, 804–813 (2018).
38. Kuhn, R., Schwenk, F., Aguet, M. & Rajewsky, K. Inducible gene targeting in mice author. *Adv. Sci.* **269**, 1427–1429 (2010).
39. Rickert, R. C., Roes, J. & Rajewsky, K. B lymphocyte-specific, Cre-mediated mutagenesis in mice. *Nucleic Acids Res.* **25**, 1317–1318 (1997).
40. Magnúsdóttir, E. et al. A tripartite transcription factor network regulates primordial germ cell specification in mice. *Nat. Cell. Biol.* **15**, 905–915 (2013).
41. Grabole, N. et al. Prdm14 promotes germline fate and Naive pluripotency by repressing FGF signalling and DNA methylation. *EMBO Rep.* **14**, 629–637 (2013).
42. Yuan, J., Nguyen, C. K., Liu, X., Kanellopoulou, C. & Muljo, S. A. Lin28b reprograms adult bone marrow hematopoietic progenitors to mediate fetal-like lymphopoiesis. *Sci. (80-).* **335**, 1195–1200 (2012).
43. Tracey, L. J. et al. The pluripotency regulator PRDM14 requires hematopoietic regulator CBFA2T3 to initiate leukemia in mice. *Mol. Cancer Res.* https://doi.org/10.1158/1541-7786.mcr-18-1327 (2019).
44. Mori, S. et al. Effect of c-fos overexpression on development and proliferation of peritoneal B cells. *Int. Immunol.* **16**, 1477–1486 (2004).
45. Hanahan, D. & Weinberg, R. A. Hallmarks of cancer: the next generation. *Cell* **144**, 646–674 (2011).

46. Moelans, C. B. et al. Genomic evolution from primary breast carcinoma to distant metastasis: few copy number changes of breast cancer related genes. *Cancer Lett.* **344**, 138–146 (2014).
47. Steenbergen, R. D. M. et al. Methylation-specific digital karyotyping of HPV16E6E7-expressing human keratinocytes identifies novel methylation events in cervical carcinogenesis. *J. Pathol.* **231**, 53–62 (2013).
48. Dervovic, D. et al. In vivo CRISPR screens reveal Serpinb9 and Adam2 as regulators of immune therapy response in lung cancer. *Nat. Commun.* **14**, 1–21 (2023).
49. Boudil, A. et al. IL-7 coordinates proliferation, differentiation and Tcra recombination during thymocyte β-selection. *Nat. Immunol.* **16**, 397–405 (2015).
50. Rajakumar, S. A. et al. B cell acute lymphoblastic leukemia cells mediate RANK-RANKL-dependent bone destruction. *Sci. Transl Med.* **12**, 1–13 (2020).
51. Meers, M. P., Bryson, T. D., Henikoff, J. G. & Henikoff, S. Improved CUT&RUN chromatin profiling tools. *Elife* **8**, 1–16 (2019).
52. Skene, P. J. & Henikoff, S. An efficient targeted nuclease strategy for high-resolution mapping of DNA binding sites. *Elife* **6**, 1–35 (2017).
53. Li, D., Hsu, S., Purushotham, D., Sears, R. L. & Wang, T. WashU epigenome browser update 2019. *Nucleic Acids Res.* **47**, W158–W165 (2019).

Acknowledgements

CyTOF and flow cytometry were performed at the SickKids' Centre for Advanced Single Cell Analysis and SickKids-UHN Flow Cytometry Facility, respectively, both supported by the SickKids Foundation. Mouse experiments were performed at The Centre for Phenogenomics or the Hospital for Sick Children Laboratory Animal Science (LAS) facility. The authors would also like to thank Angie Griffin for technical assistance.

Author contributions

MJJ supervised the study. LJT and MJJ conceived of and designed all experiments. CJG designed CyTOF, flow cytometry and adoptive transfer experiments. LJT, MJJ, DJS, and MDW designed CUT&RUN experiments. LJT, ME-M, TBB, MSG, AK, and JR performed experiments. SR performed intra-femoral adoptive transfer experiments. TBB, CT and JR maintained mouse colonies and performed tissue preparations. LJT, ME-M, CJG, and MJJ analyzed and interpreted data. DJS analyzed scRNA-seq and CUT&RUN data. LJT wrote the manuscript. LJT, ME-M, DJS, MDW, CJG, and MJJ reviewed and revised the manuscript. All authors read and approved the final manuscript.

Funding

MJJ's work was supported by an NIH Grant (R01CA163849). CJG's work was supported by a Canadian Institutes of Health Research Project Grant (FRN 165973) and a Canadian Foundation for Innovation Fund John Evans Fund Leaders' Infrastructure Grant (#37562). MSG was supported by a SickKids' Restracomp Studentship, and ME-M received a NSERC Vanier Canada Graduate Scholarship from the National Sciences and Engineering Council of Canada. MDW and DS were supported in part by NSERC grant RGPIN-2019-07014 to MDW. DS was also supported by NSERC CGS M, PGS D and the Ontario Genomics-CANSSI Ontario Postdoctoral Fellowships in Genome Data Science.

Declarations

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/s41598-025-93043-z.

Correspondence and requests for materials should be addressed to M.J.J.