Single-cell transcriptional analysis of normal, aberrant, and malignant hematopoiesis in zebrafish

Finola E. Moore,1* Elaine G. Garcia,1* Riadh Lobbedi,1 Esha Jain,3 Qin Tang,1 John C. Moore,1 Mauricio Cortes,3 Aleksy Molodtsov,1 Melissa Kasheta,6 Christina C. Luo,1 Amaris J. Garcia,1 Ravi Mylvaganam,1 Jeffrey A. Yoder,7 Jessica S. Blackburn,8,9,10 Ruslan I. Sadreyev,4,12 Craig J. Ceol,6 Trista E. North,5 and David M. Langenau1,2,3,13

Hematopoiesis culminates in the production of functionally heterogeneous blood cell types. In zebrafish, the lack of cell surface antibodies has compelled researchers to use fluorescent transgenic reporter lines to label specific blood cell fractions. However, these approaches are limited by the availability of transgenic lines and fluorescent protein combinations that can be distinguished. Here, we have transcriptionally profiled single hematopoietic cells from zebrafish to define erythroid, myeloid, B, and T cell lineages. We also used our approach to identify hematopoietic stem and progenitor cells and a novel NK-lysin 4+ cell type, representing a putative cytotoxic T/NK cell. Our platform also quantified hematopoietic defects in rag2E450fs1 mutant fish and showed that these fish have reduced T cells with a subsequent expansion of NK-lysin 4+ cells and myeloid cells. These data suggest compensatory regulation of the innate immune system in rag2E450fs1 mutant zebrafish. Finally, analysis of Myc-induced T cell acute lymphoblastic leukemia showed that cells are arrested at the CD4+/CD8+ cortical thymocyte stage and that a subset of leukemia cells inappropriately reexpress stem cell genes, including bmi1 and cmyb. In total, our experiments provide new tools and biological insights into single-cell heterogeneity found in zebrafish blood and leukemia.

Cell differentiation is a complex process that results in the production of widely heterogeneous and functionally distinct cell populations. The individual cell fate decisions that create cell heterogeneity are best understood at single-cell resolution, especially within blood where hematopoiesis culminates in the production of a variety of functionally diverse cells. Blood cell lineages and leukemia cell subfractions from mouse and human have been commonly identified using cell surface antibodies and FACS. Building off these approaches, investigators have used increasingly complex strategies to define blood cell heterogeneity, including multiparameter flow cytometry and mass cytometry (Irish et al., 2006; Kotecha et al., 2008; Bendall et al., 2011, 2014; Fathman et al., 2011; Gibbs et al., 2011; Amir el et al., 2013; Lacayo et al., 2013; Litjens et al., 2013; Shalek et al., 2013, 2014; Sen et al., 2014; Levine et al., 2015). Single-cell transcriptional profiling has also been used to assess blood cell heterogeneity and leukemia (Flatz et al., 2011; Guo et al., 2013; Moignard et al., 2013; Riddell et al., 2014; Saadatpour et al., 2014; Fan et al., 2015; Wilson et al., 2015). For instance, Flatz et al. (2011) used single-cell quantitative PCR (RfPCR) to identify new subsets of CD8+ T cells and studied their response to different vaccines. Heterogeneity within leukemia and acquisition of novel gene programs have also been described using this approach (Guo et al., 2013; Saadatpour et al., 2014). Despite the plethora of techniques used to assess blood cell heterogeneity in mice and humans, these single-cell transcriptional profiling approaches have yet to be widely adapted to other experimental models, including the zebrafish.

Zebrafish have emerged as a robust model for studying hematopoiesis, immunity, and cancer. Best known for its utility in forward genetic screens, the zebrafish has contributed immensely to our understanding of blood development. For example, a forward genetic screen identified ferroportin as a
novel iron exporter (Donovan et al., 2000), and mutations in this gene were subsequently found to be a common cause of inherited disorders of iron overload in humans (Pietrangelo, 2004). Zebrafish have also become a facile and powerful model for discovering novel drugs that affect blood and leukemia growth. For example, North et al. (2007) identified prostaglandin as a potent inducer of hematopoietic stem cells. Di-methyl PGE2 is currently in Phase II clinical trials to improve transplantation of human umbilical cord blood (Goessling et al., 2011;Cutler et al., 2013). Beyond normal hematopoiesis, immune-compromised zebrafish have been developed as models of severe combined immunodeficiency (Wienholds et al., 2002;Jima et al., 2009;Petrie-Hanson et al., 2009;Tang et al., 2014). Finally, a wide range of zebrafish blood malignancies has been developed including T cell acute lymphoblastic leukemia (T-ALL; Langenau et al., 2003, 2005;Chen et al., 2007;Feng et al., 2007;Frazer et al., 2009;Gutierrez et al., 2011). Using these models and chemical screening approaches, investigators have discovered new pathways and novel drugs that differentiate or kill leukemia cells (Yeh et al., 2009;Ridges et al., 2012;Blackburn et al., 2014;Gutierrez et al., 2014).

Despite the clear advantages of the zebrafish model for studying hematopoiesis and leukemia, the lack of lineage-specific cell surface antibodies remains a major hurdle for the field. Rather, analysis of heterogeneity has been largely limited to morphological assessment of blood cells after cytospin or by FACS that can discriminate cells based on size and granularity (Traver et al., 2003). Fluorescent transgenic reporter lines provide a more detailed understanding of blood development by labeling specific cell lineages. For example, Page et al. (2013) delineated different stages of B cell development in adult zebrafish using a dual fluorescent transgenic Tg(rag2:dsRed); Tg(IgM:eGFP) line; yet these experiments could not distinguish between mature T lymphocytes, myeloid cells, and erythroid cells within the same animal. These experiments illustrate the state of our field, where reliance on identifying blood cell lineages is limited by the precision with which transgenic promoters label cells and by the availability of fluorophores that can be distinguished by FACS or confocal imaging.

Here, we developed a transcriptional profiling approach that robustly characterizes single-cell heterogeneity in a wide range of blood cell types. Using the Fluidigm single-cell quantitative (qPCR) platform, we systematically categorized the major blood cell lineages. We have also characterized hematopoietic stem and progenitor cells (HSPCs) and NK-lysin 4+ cells (nkl.4), representing a putative cytotoxic T/NK cell population. Our platform also impartially assessed hematopoietic defects in T cell–deficient rag2−/− zebrafish. Finally, our work revealed that zebrafish Myc-induced T-ALL cells are arrested at the immature CD4+/CD8+ double positive stage and that only a subset of leukemia cells reactivate stem cell genes, including bmi1 and cmyb. In total, our experiments provide facile and robust methods to transcriptionally profile zebrafish blood lineages at single-cell resolution. We also provide new insights into the biology of hematopoiesis in the zebrafish model.

RESULTS
Identifying diverse blood cell types using single-cell qPCR
To transcriptionally profile blood cell types in adult zebrafish, we first optimized a set of 96 primer pairs, comprising 50 genes that include well-known markers for specific blood cell lineages, candidate markers for undefined cell types of interest, and housekeeping genes. qPCR was completed using the Fluidigm BioMark microfluidics platform (Table S1). To substantiate calls made by qPCR, multiple primers for each gene were analyzed when possible. Not unexpectedly, 90% of primers for the same gene clustered immediately next to one another when assessed by row distance matrix analysis (n = 27 of 30 genes). BioMark results were also highly reproducible as assessed by technical replicates of bulk cDNA and replicate analysis of single cells completed on different days (r² = 0.93; n = 69 single cells analyzed). Single cells from WT whole-kidney marrow (WKM), the site of hematopoiesis in adult zebrafish, were isolated by FACS and transcriptionally profiled. Data were then subjected to unsupervised hierarchical clustering, which identified four major gene expression clusters that comprised erythroid, myeloid, B, and T lymphoid cells (Fig. 1A, gene order is the same for all heat maps and is provided in Table S2). Weighted gene co-expression network analysis (WGCNA) independently revealed four major clusters of genes that correlate with specific blood lineages (Fig. 1B and C; and Fig. S1). Violin plots showed the distribution of cells expressing each gene transcript, allowing independent assessment of cells assigned to specific cell lineages (Fig. 1D). As expected, the majority of cells characterized as erythroid by hierarchical clustering analysis expressed erythroid-specific genes, including β A1 globin (ba1), solute carrier family 4 (anion exchanger; slc4a1a/band3), and spectrin β erythrocytotic (sptb; P = 5.67 × 10⁻1⁶⁸ and 8.88 × 10⁻⁶⁴, 1.57 × 10⁻⁵³, respectively, by ANOVA). Similarly, cells in the myeloid cluster expressed high levels of myeloid-specific peroxidase (mpx), lysosome C (lyz), and carboxypeptidase a5 (cpa5; P = 4.31 × 10⁻⁷³, 2.70 × 10⁻⁷¹, and 1.67 × 10⁻⁶², respectively, by ANOVA). Single-cell qPCR was also able to distinguish T from B cells within the same animal. This is particularly important because standard flow cytometry using forward and side scatter or morphological analysis of cytospins cannot discriminate these populations within bulk marrow cells of the same zebrafish (Traver et al., 2003). T cells expressed T cell receptor α (tcra), lymphocyte-specific protein-tyrosine kinase (lck), and interleukin 7 receptor (il7r/cd127; P = 3.24 × 10⁻¹⁶, 2.64 × 10⁻³⁰, and 4.48 × 10⁻²⁴, respectively, by ANOVA), whereas B cells commonly expressed paired box 5 (pax5), cd37, and cd79a (P = 1.45 × 10⁻⁴³, 1.19 × 10⁻⁸², and 1.30 × 10⁻⁵⁷, respectively, by ANOVA).

To confirm our ability to discern specific blood cell lineages by transcriptional profiling, hematopoietic cells from
Figure 1. Single-cell qPCR delineates four major blood cell lineages from adult zebrafish marrow. (A) Unsupervised hierarchical clustering after gene expression analysis of single blood cells isolated from the whole kidney marrow. Heat map shows high transcript expression in red and low/absent expression in blue. Four major clusters were identified, including the following: erythroid (red), myeloid (green), B cells (light blue), and T cells (dark blue). Select lineage-specific genes are shown on the right. n = 166 WKM cells combined from two animals. (B) WGCNA-identified genes that were expressed together in single cells. Each row and column corresponds to a specific gene primer set. The heat map represents WGCNA adjacency dissimilarity values in the range between 0 (low topological overlap, light yellow) and 1 (high topological overlap, dark red). (C) The genes in each lineage module from WGCNA are noted. (D) Violin plots show the distribution of gene expression of single cells. Cells types were assigned based on hierarchical clustering and assessed for transcript expression of well-known blood cell lineage genes. By ANOVA, ba1 (P = 5.67 × 10^{-168}), slc4a1a/band3 (P = 8.88 × 10^{-404}), sptb (P = 1.57 × 10^{-45}), mpx (P = 4.31 × 10^{-71}), lyz (P = 2.70 × 10^{-77}), cpa5 (P = 1.67 × 10^{-62}), pax5 (P = 1.45 × 10^{-43}), cd37 (P = 1.19 × 10^{-51}), cd79a (P = 1.30 × 10^{-72}), tcra (P = 3.24 × 10^{-116}, lck (P = 2.64 × 10^{-20}), and il7r/cd127 (P = 4.48 × 10^{-24}).
adult fluorescent transgenic zebrafish and peripheral blood were compared. Erythroid cells were isolated by FACS from the peripheral blood of WT animals and the WKM of Tg(gata1:dsRed) animals (n = 44 cells). Neutrophils comprise a distinct subset of myeloid cells in the WKM and were isolated from Tg(mpx:GFP) animals (n = 48 cells). B cells and mature T cells were also isolated from WKM of Tg(rag2:dsRed) (n = 49 cells) and Tg(lck:GFP) animals (n = 83 cells), respectively. Transgenic-labeled cell populations were represented at relative frequencies as previously reported (Fig. 2 A) and isolated to high purity after FACS (>92% purity, >95% viability; Langenau et al., 2003, 2004; Traver et al., 2003; Lin et al., 2005; Mathias et al., 2006; Ma et al., 2012). Transcriptional profiling revealed that cells from each transgenic line largely clustered together after unsupervised hierarchical clustering (Fig. 2B). Principal component analysis independently confirmed that transgenic and unlabeled WKM cells clustered tightly within each of the four major lineages (Fig. 3 A). Finally, ≥94% of transgene-expressing cells were classified into their expected blood cell lineages after unsupervised hierarchical clustering with unlabeled WKM cells (Fig. 3 B and Fig. S2), confirming the specificity of transgenic lines, the high sort purity after FACS, and the accuracy of the qPCR calls. Violin plots confirmed distinct gene expression associated with each lineage (Fig. 3 C). Collectively, our results show that single-cell qPCR with our panel of markers reproducibly identifies distinct classes of blood cells within the WKM.

**Gene expression analysis of hematopoietic stem and precursor cells (HSPCs)**

Having used our platform to identify common blood cell lineages, we next wanted to characterize rare subpopulations of cells, including hematopoietic stem and precursor cells (HSPCs). Tg(CD41:GFP)low cells comprise 0.7% of the marrow and are HSPCs as determined by both gene expression and cell transplantation studies (Bertrand et al., 2008; Ma et al., 2011). Single blood cells were isolated from the GFPlow population of Tg(CD41:GFP) WKM (n = 85 cells), and then compared with normal WKM (n = 166 cells). After unsupervised hierarchical clustering, the CD41:GFP low cells clustered into two distinct populations (Fig. 4, A and B). The most prominent group clustered outside any of the previously defined lineages and expressed high levels of known stem cell genes including **Figure 2. Single cells isolated from lineage-restricted transgenic reporter lines segregate together after hierarchical clustering.** (A) FACS analysis (left) and cytospins of sorted transgenic blood populations (right). Bars, 25 µm. (B) Unsupervised hierarchical clustering after single-cell gene expression analysis. Cell identity is shown above the heat map: peripheral blood cells (pink, n = 20 cells) and sorted fluorescent WKM cells from gata1 :dsRed (red, n = 24 cells), mpx:GFP (green, n = 48 cells), rag2:dsRed (light blue, n = 49 cells), or lck:GFP (dark blue, n = 83 cells). Cells shown are from a single representative animal. Yellow shading denotes cells that failed to cluster into known lineages described by our qPCR panel.
LIM domain only 2 (lmo2), runt-related transcription factor 1 (runx1), and myeloblastosis viral oncogene homologue (cmyb; Fig. 4 A). As expected, these HSPCs largely failed to express differentiated blood cell lineage markers including ba1, mpx, pax5, and tcrα (Fig. 4, A and D). A second group of CD41 :GFPlow cells loosely clustered with myeloid cells and failed to express the stem cell lineage markers including lmo2, runx1, and cmyb (Fig. 4 A). These cells also did not express prominent cell lineage markers for erythroid, myeloid, B, or T cells (Fig. 4 D), suggesting they are likely precursor cells that are transiting into mature cell lineages. In support of this interpretation, principal component analysis showed that the CD41 :GFPlow cells cluster tightly together, independently of the other four dominant cell subfractions identified in the WKM (Fig. 4 C). Collectively, these data demonstrate that our approach can characterize rare HSPCs and identify previously undefined heterogeneity within CD41 :GFPlow cells.

Quantifying blood cell deficiencies in rag2-hypomorphic zebrafish

We next used our approach to define alterations in blood cell lineage differentiation in mutant zebrafish. rag2\textsuperscript{E450fs} mutant zebrafish contain a hypomorphic mutation in the plant homeodomain; similar mutations in human are associated with severe combined immunodeficiency and Omenn Syndrome (Villa et al., 1999; Tang et al., 2014). Bulk WKM analysis previously revealed that rag2\textsuperscript{E450fs} mutant zebrafish have reduced expression of mature T and B cell–specific genes, reduced lymphocyte cell counts as assessed by cytospin analysis, and variable effects on T and B cell receptor recombination (Tang et al., 2014); yet definitive loss of either lymphocyte cell type has not been reported. Single-cell gene expression studies of rag2\textsuperscript{E450fs} WKM cells revealed a striking lack of mature T cells (Fig. 5, A and B; and Fig. S3; \(P = 0.0001\), Fisher’s exact test). rag2\textsuperscript{E450fs} mutant fish had no alteration in the proportion of B cells contained within the marrow when compared with WT control animals (Fig. 5, A and B; and Fig. S3; \(P = 0.335\), Fisher’s exact test). These single-cell gene expression studies are consistent with the lack of T cell receptor β (tcrβ) recombination, but largely normal IgM rearrangements in rag2\textsuperscript{E450fs} mutant fish (Tang et al., 2014).

Cytotoxic innate immune cells, including NK cells and cytotoxic T lymphocytes (CTLs), have long been postulated to exist in zebrafish, but have yet to be identified.
rag2-deficient mice have an expansion of NK cells (Shinkai et al., 1992; Wang et al., 1996), we reasoned that rag2E450fs mutant zebrafish might have elevated numbers of cytotoxic innate immune cells. NK-lysins are ancient antimicrobial proteins that are functionally similar to Granulysin, a protein produced by both CTLs and NK cells in mammals (Andersson et al., 1995; Peña et al., 1997). Of the four NK-lysin genes in zebrafish, NK-lysin 4 (nkl.4) is up-regulated after response to viral infection, suggesting nkl.4 could be a marker for cytotoxic T/NK cells in the zebrafish (Pereiro et al., 2015). Thus, nkl.4 was included in our panel of 48 experimental genes.
these cells are distinct but related to T cells. Given the lack of other mature lineage marker expression (Fig. S3 B), we posit that these cells define a novel blood cell type and are likely related to cytotoxic T/NK cells. Myeloid cells were also greatly expanded in this model (Fig. 5, A and B; P = 0.0001, Fisher’s exact test), suggesting that compensatory mechanisms are turned on in rag2E450fs mutant zebrafish. Counts were based on single-cell PCR gene expression and assignments based on hierarchical clustering defined in Fig. 3. *, P = 0.0081; **, P = 0.0001, by Fisher’s exact test. The numbers of cells per lineage are noted within the bar graph. n = 166 WT WKM cells and n = 112 rag2E450fs WKM cells combined from two animals. (C) Violin plots show the distribution of gene expression in nkl.4+ cells. (D) Genes expressed within each cell type.

Figure 5. rag2E450fs mutant zebrafish have reduced T cells and expanded myeloid and nkl.4+ cells. (A) Principal component analysis showing loss of T cells from the WKM of rag2E450fs mutant zebrafish and expansion of nkl.4+ cells (black). WT and rag2E450fs WKM were included in PCA. For visualization, cells from rag2E450fs WKM are shown in color, and WT WKM cells are greyed circles. (B) Quantification of blood cell types from the marrow of WT and rag2E450fs mutant zebrafish. Counts were based on single-cell PCR gene expression and assignments based on hierarchical clustering defined in Fig. 3. *, P = 0.0081; **, P = 0.0001, by Fisher’s exact test. The numbers of cells per lineage are noted within the bar graph. n = 166 WT WKM cells and n = 112 rag2E450fs WKM cells combined from two animals. (C) Violin plots show the distribution of gene expression in nkl.4+ cells. (D) Genes expressed within each cell type.

Analysis of differentiation arrest and reactivation of stem cell genes in T cell acute lymphoblastic leukemia

T-ALLs are heterogeneous, comprising many molecular subtypes arrested at different stages of T cell maturation (Ferrando et al., 2002; Ferrando and Look, 2003; van Grotel et al., 2008). For example, the zebrafish Myc-induced T-ALL model has been previously suggested to mimic a common and aggressive form of human leukemia with cells arrested at the CD4+/CD8+ cortical thymocyte stage of development (Langenau et al., 2005; Blackburn et al., 2012). Single cells were isolated from normal rag2:dsRed+ thymocytes and compared with three T-ALLs that had similar latency and high leukemia-propagating cell (LPC) frequency (Blackburn et al., 2014; Fig. 6, A and C; and Table S3). After single-cell transcriptional profiling, rag2:dsRed+ thymocytes were comprised of 11.7% double-negative (DN) cells, 53.2% double-positive (DP) cells, and single-positive CD4+ or CD8+ cells (10.4 and 24.7%, respectively; Fig. 6 D and Fig. S4). In contrast, >87% of T-ALL cells coexpressed both CD4 and CD8, showing
that leukemias were arrested at the immature double-positive (DP) stage (Fig. 6 D; P = 0.0001, Fisher’s exact test).

Principal component analysis also revealed that T-ALLs are molecularly distinct from both normal rag2:dsRed mature thymocytes and Ick:GFP mature T cells (Fig. 6 E). Remarkably, T-ALL cells transcribed genes not normally expressed in thymocytes. WGCNA identified stem cell genes that were specifically found in T-ALLs (Fig. 6 F). For instance, the early T cell maturation genes rag1 and gata3 were highly expressed in a large fraction of T-ALL cells, reflecting the strong developmental arrest at the DP stage (Fig. 6 G). Leukemias also showed inappropriate activation of β-catenin (Fig. 6 G), reminiscent of recent studies that showed Wnt pathway contributes to LPC self-renewal in T-ALL (Guo et al., 2008; Giambra et al., 2015). bmi1, a well-known driver of self-renewal in both normal and malignant blood (Dik et al., 2005; Hosen et al., 2007; Saudy et al., 2014), was expressed in a subfraction of T-ALL cells, but was not expressed in normal thymocytes (Fig. 6 G; P = 0.0001, Fisher’s exact test). cmyb is required for the proliferation and differentiation of HSPCs (Mucenski et al., 1991; Sandberg et al., 2005). MYB is also required for T-ALL growth and maintenance (Lahortiga et al., 2007; Mansour et al., 2014). cmyb was highly expressed in zebrafish T-ALL cells (Fig. 6 G; P = 0.0001, Fisher’s exact test). In fact, cmyb and bmi1 were coexpressed in 18.4% of T-ALL cells, but not normal T cells (n = 46 of 250 T-ALL cells and n = 0 of 160 normal T cells, P = 0.0001, Fisher’s exact test). Intriguingly, ~17% of leukemic cells were LPCs as assessed by limiting dilution cell transplantation (Fig. 6 A and Table S3). These data suggest that bmi1+/cmyb+ T-ALL cells may represent the self-renewing LPCs. Collectively, our results demonstrate that Myc-induced T-ALLs are arrested at immature cortical thymocyte stage and that a minority of tumor cells aberrantly reexpress stem cell genes.

**DISCUSSION**

Multiparameter flow cytometry and mass cytometry have richly profiled single-cell heterogeneity in blood and leukemia (Irish et al., 2006; Kotecha et al., 2008; Bendall et al., 2011, 2014; Gibbs et al., 2011; Lacayo et al., 2013; Litjens et al., 2013). Although these approaches have revolutionized how we analyze diversity in hematopoietic cells, they require a large set of validated antibodies specific to a given lineage or cell subtype. Yet, antibody reagents are sorely lacking in zebrafish. More recently, cell heterogeneity has also been defined by differences in gene expression between single cells. Single-cell qPCR using the Fluidigm platform has assessed novel transcriptional networks in blood, T cell diversity in response to vaccination, and self-renewal programs in AML (Flatz et al., 2011; Guo et al., 2013; Moignard et al., 2013). For instance, Moignard et al. (2013) examined transcription factor expression using single-cell qPCR analysis of mouse hematopoietic cells and uncovered novel gene regulatory networks between gata1, gfi1, and gfi2. Guo et al. (2013) also used single-cell qPCR to assess differentiation hierarchies in normal blood cells and revealed gene expression changes associated with self-renewal in leukemic cells. Our work has optimized a simple qPCR transcriptional profiling approach using the Fluidigm platform to assess heterogeneity in zebrafish blood. This approach is advantageous in that it allows analysis of lineage-restricted genes to define cell states, can detect low-level expressed genes, including transcription factors, and is facile in both its implementation and data analysis. Moreover, using this approach, we have refined gene expression of novel blood cell types and uncovered new biology underlying hematopoiesis in the zebrafish model.

Capitalizing on the power of single-cell qPCR to identify blood cell types within the marrow, we assessed if we could identify quantitative losses of specific cell types in hematopoietic mutant zebrafish. Our single-cell transcriptional analysis uncovered that rag2<sup>△E450</sup> mutant zebrafish lack T cells but have a largely intact B repertoire, in agreement with previous work (Tang et al., 2014). The rag2<sup>△E450</sup> mutant also had a significant increase in myeloid cells, similar to that suggested for rag1 mutant zebrafish (Petrie-Hanson et al., 2009) and contain a previously undescribed cell type that express nkl4. NK-lysins are functionally similar to Granulysin, which is produced by CTLs and natural killer (NK) cells in mammals (Andersson et al., 1995; Peña et al., 1997), rag2 deficiency in mice leads to NK cell expansion as a result of the lack of inhibition by normal T cells (Shinkai et al., 1992; Wang et al., 1996), suggesting the expansion of nkl4+ cells observed in rag2<sup>△E450</sup> zebrafish could also represent cytotoxic T/NK cells. Collectively, our single-cell qPCR approach provides a robust platform to quantify blood cell deficiencies in mutant zebrafish, bypassing the need for transcriptional analysis of bulk marrow and inference of cell losses based on bulk RNA expression changes (Tang et al., 2014; Pereiro et al., 2015).

Our work also characterized rare CD41:GFP<sup>+</sup> HSPCs in the normal WKM. In agreement with findings in the zebrafish embryo, adult HSPCs express lmo2, runx1, and cmyb (Burns et al., 2005; Bertrand et al., 2008). Other stem cell genes are likely further restricted to less differentiated precursors, or even to hematopoietic stem cells (HSCs). For instance, bmi1, though expressed in mouse HSCs (Hosen et al., 2007), is not commonly expressed in zebrafish CD41:GFP<sup>+</sup> HSPCs, reflecting the rarity of HSCs even within the progenitor compartment. Single-cell qPCR also captured additional heterogeneity within CD41:GFP<sup>+</sup> cells, identifying a group of transitioning HSPCs that failed to express cmyb, runx1, or other lineage-restricted genes. Heterogeneity within CD41:GFP<sup>+</sup> cells is expected given that imaging and FACS experiments have demonstrated variable coexpression of CD41:GFP<sup>+</sup> cells with other progenitor markers (Bertrand et al., 2008; Tamplin et al., 2015).

Our single-cell transcriptional platform also enabled analysis of differentiation states and reinitiation of stem cell programs in a subset of leukemia cells. For example, we found that Myc-induced T-ALL cells were arrested at the immature CD4/CD8 DP stage. Similarly, zebrafish T-ALLs showed ab-
Figure 6. Zebrafish Myc-induced T-ALLs are arrested at the CD4⁺/CD8⁺ cortical thymocyte stage and reexpress stem cell genes. (A) Merged bright field and fluorescent images of Tg(rag2:dsRed) zebrafish (top left) or fluorescent leukemic fish. LPC frequency for each individual T-ALL is denoted in bottom left corner of each image, and clone name is denoted in the top left corner. The white arrowhead designates thymus. Bars, 2.5 mm. (B) May-Grunwald staining of representative T-ALL cells (left) and normal WKM (right; red blood cell, red arrow; myeloid cell, green arrow, and lymphocyte, blue arrow). Bars, 25 µm. (C) Kaplan-Meier analysis of leukemia onset in transplanted zebrafish. Number designations correspond to leukemias described by Blackburn et al. (2014). (D) Percentage of T cells that express CD4 and CD8 in normal rag2:dsRed⁺ thymic T cells and fluorescent-labeled T-ALLs. ***, P = 0.0001, by Fisher’s exact test. (E) Principal component analysis identified three distinct clusters of cells that include the following: mature T cells (Tg(lck:GFP) dark blue, n = 83 cells), thymocytes (Tg(rag2:dsRed) light blue, n = 77 cells from a single animal), and T-ALLs (pink, n = 250 cells, combined from three independent leukemias). (F) WGCNA of T-ALLs revealed that stem cell genes are reexpressed in T-ALLs and expressed together. Each row and column corresponds to a specific gene.
normally high expression of genes that affect differentiation in normal hematopoiesis, including cmyb and gata3. These transcription factors are highly expressed in a large cohort of T-ALL patient samples (Minegishi et al., 1997; Clappier et al., 2007; Lahortiga et al., 2007). Intriguingly, knockdown of MYB causes human T-ALL cells to partially differentiate (Lahortiga et al., 2007), suggesting the high expression of cmyb could contribute to the maintenance of zebrafish T-ALL maturation arrest. **Bmi1** confers self-renewal properties to a variety of solid and hematological tumors, including multiple subtypes of T-ALL (Jacobs et al., 1999; Lessard and Sauvageau, 2003; Savitoğlu et al., 2012; Saudy et al., 2014). Cells that expressed both bmi1 and cmyb comprised 18.4% of zebrafish T-ALL cells and correlated well with LPC frequency as assessed by limiting dilution cell transplantation, raising the interesting possibility that these genes define LPCs in the zebrafish model. The expression pattern of stem cell genes, such as assessed by limiting dilution cell transplantation, raising the interesting possibility that these genes define LPCs in the zebrafish model. The expression pattern of stem cell genes, such as bmi1 and cmyb, in a minority of tumor cells demonstrates the power of single-cell analysis in studying rare cell subtypes and provides new testable hypotheses about leukemia LPCs.

Looking to the future, we envision that expanded primer sets will be developed for use with the Fluidigm platform, allowing further subclassification of blood and leukemia cell types. Importantly, we also anticipate that future studies will use single-cell RNA sequencing to investigate blood cell heterogeneity. RNAseq has become a powerful technology to assess single-cell transcriptomes and identified novel gene networks in embryonic stem cells, retinal cells, blood and development (Shalek et al., 2013; Fan et al., 2015; Klein et al., 2015; Macosko et al., 2015; Satija et al., 2015; Wilson et al., 2015). Because single-cell RNA sequencing often only captures transcriptional changes confined to highly expressed genes, analysis of low expressed genes including transcription factors, as assessed in our qPCR analysis, is often incomplete. Thus, we believe that combining our qPCR profiling approach with RNA sequencing will likely permit anchoring of large-scale, single-cell RNA-Seq data. This approach is similar to those pioneered in mouse and human, where antibody labeling and FACS delineate distinct blood cell populations, which can then be further refined by RNA sequencing approaches. For example, Wilson et al. (2015) used FACS to enrich for mouse HSPCs, which were then assessed by RNAseq to identify previously unappreciated heterogeneity in this cell population and to discover new gene networks that affect self-renewal. Although qPCR profiling will complement RNAseq approaches, it cannot replace FACS to functionally analyze single cells caused by the destructive nature of RNA extraction. Our approach provides a robust assay to define a wide range of blood cell types after transcriptional profiling by qPCR, obviating the need to identify single-cell heterogeneity using cell surface antibodies and FACS.

In total, our work provides the first analysis of gene expression at the single-cell level in zebrafish blood development and has identified novel cell types not previously defined in the zebrafish model. We have demonstrated the utility of single-cell qPCR in zebrafish as an accessible and facile tool to study heterogeneous blood cell populations.

### MATERIALS AND METHODS

#### Zebrafish

All experiments were completed with IACUC approval from the Massachusetts General Hospital (2011N000127).

Experiments used WT AB strain zebrafish, **rag2** mutant fish (Tang et al., 2014), and the following transgenic lines: **Tg(gata1:dsRed)** (Traver et al., 2003), **Tg(mpx:GFP)** (Mathias et al., 2006), **Tg(rag2:dsRed)** (Ma et al., 2012), **Tg(rag2;GFP)** (Langenau et al., 2003), **Tg(lck;GFP)** (Langenau et al., 2004), and **Tg(CD41;GFP)** (Lin et al., 2005). Leukemic fish were created by implantation of monoclonal, fluorescent T-ALLs into syngeneic CG1 recipient animals, as previously described (10^5 cells/fish; Blackburn et al., 2014). Kaplan-Meier analysis was completed on engrafted animals, and LPC frequency was evaluated at the time of single-cell sorting using the Extreme Limiting Dilution Analysis algorithm as previously described (Smith et al., 2010).

#### Sample preparation and isolation of single cells by FACS

WKM and thymocytes were isolated from 3–6-mo-old fish, placed in 5% FBS/1xPBS, triturated by pipet, and filtered through a 35-µm cell strainer. T-ALL were harvested by dissecting diseased animals in 5% FBS/1xPBS and filtering through a 40-µm cell strainer (Falcon 352340), as previously described (Smith et al., 2010; Blackburn et al., 2012). For WT and **rag2** mutant WKM, cells were isolated from two animals, sorted independently, and then combined for analysis.

All samples were single-cell sorted in the presence of cell viability stain (DAPI or PI) using the FACSArta Fusion Cell Sorter (BD). Doublets were gated out based on forward-scatter area compared with forward-scatter height. Cell purity was assessed by sorting and reanalysis, showing enrichment of fluorescent reporters to >92% and viability >95%. Single cells were sorted into 96-well plates, with each well containing 5 µl of lysis buffer (10 ml master mix was made combining 9.5 ml DNA Suspension Buffer [TEKnova T0221], 50 µl 20 U/µl SUPERase-in [Ambion AM2696], and 500 µl 10% NP-40 [AppliChem A2239]). Unlabeled WKM samples were index sorted. Plates were centrifuged, frozen on dry ice, and stored at −80°C until further use.

#### Primer optimization

Selected genes were identified using NCBI, Ensemble, and ZFIN databases. PCR primers spanned primer set. The heat map represents WCGNA adjacency dissimilarity values in the range between 0 (low topological overlap, light yellow) and 1 (high topological overlap, dark red). Red lettering denotes genes shown in G. (G) Violin plots show the distribution of transcript expression for genes involved in T cell differentiation and stem cells. ***; P = 0.0001, by Fisher’s exact test.
introns when possible, were restricted to regions that were common to all REFseq transcripts, and were verified for specificity by PrimerBLAST (Ye et al., 2012). Primer sequences are listed in Table S1.

Primers were tested on the BioMark (Fluidigm; and also ABI7500) using bulk cDNA, as well as single cells. Varying amounts of bulk cDNA, no room temperature samples, and water were used as controls. Bulk cDNA from embryos (single cell, 24 hpf, 5 dpf embryos), WT WKM, several T-ALL clones, peripheral blood, and thymus was used. Primers were ordered in plates (200 µl well volume, 25 nM synthesis scale, desalted purification, TE buffer, ambient ship format, normalized concentration to 100 µM) obtained from Invitrogen. Outer primers were prepared as follows: 100 µM forward and 100 µM reverse were mixed, and then 3.3 µl of each primer pair were pipetted into Eppendorf tube, 99.1 µl DNA suspension buffer (TEKnova T0221), such that the STA Outer Primer mix has each primer at a concentration of 166.6 nM. Inner primers were as follows for 20 µl of 5 µM stock: 10 µl of 0.25 µl forward primer (Invitrogen 100 µM), 0.25 µl reverse primer (Invitrogen 100 µM) for each primer. After priming and combining 2.5 µl 2X Assay Loading Reagent (Fluidigm PN 85000736), 8 µl 1X DNA Suspension Buffer (TEKnova PNT0211), 1 µl forward primer (100 µM), and 1 µl reverse primer (100 µM).

### Single-cell qPCR using the BioMark HD

Plates were thawed on ice and heated at 65°C for 90 s. 1 µl of Reverse Transcription Mix (100–6299; Fluidigm) was added to each well and then placed into a thermocycler for 25°C, 5 min; 42°C, 30 min; and 85°C, 5 min. Next, a preamplification step was completed using gene-specific primers (Table S1, outer primers). To each 6 µl cDNA RT reaction, 1 µl 10X PreAmp Master Mix (100–5581; Fluidigm) and 3 µl 166.6 nM specific target amplification (STA) outer primer mix was added. Plates were then placed into the thermocycler for 95°C, 5 min and 20 cycles of 96°C, 5 s and 60°C, 6 min. Unincorporated primers were removed by exonuclease treatment (for each 10 µl preamplification, 0.8 µl 20 U/µl exonuclease I [NEB M0293L], 0.4 µl 10X Exonuclease I Buffer, 2.8 µl H2O) and incubating at 37°C for 30 min, and then 80°C for 15 min. Samples were then diluted by addition of 36 µl of 1X TE. Each well now contained 50 µl of preamplified cDNA that was ready for qPCR using the BioMark HD.

Samples were prepared for qPCR by combining 3 µl of sample with 3.5 µl 2X Sso Fast EvaGreen Supermix with low ROX (172–5212; Bio–Rad Laboratories), 0.35 µl 20X DNA Binding Dye Sample Loading Reagent (100–7609; Fluidigm), and 0.15 µl H2O. 5 µM inner primer mix was prepared by combining 2.5 µl 2X Assay Loading Reagent (85000736; Fluidigm), 2 µl 1X DNA Suspension Buffer (TEKnova T0211), 0.25 µl forward primer (Invitrogen 100 µM), 0.25 µl reverse primer (Invitrogen 100 µM) for each primer. After priming the 96.96 Dynamic Array Chip for Gene Expression or integrated fluidic circuit (IFC; BMK–M–96.96; Fluidigm), 5 µl primers and 5 µl samples were loaded onto IFC and mixed in the Fluidigm IFC controller HX. IFC was then loaded into the BioMark HD following manufacturer’s protocol. Thermal cycling was completed using protocol GE 96 × 96 PCR melt v2.pcl (2,400 s at 70°C, 30 s at 60°C, 60 s at 95°C, 30 cycles of 5 s at 96°C, 20 s at 60°C, melting curve 60–95°C, 1°C/3 s).

### Statistical analysis and display of single-cell data

Ct's were recovered from the BioMark HD. The quality threshold was set to 0.65 and a linear derivative as baseline correction was used. Limit of detection was set to the Ct of 28; Ct's for qPCR reactions that failed the quality threshold were converted to 28. Ct's were converted to log2 expression by subtracting Ct's from 28. Sample wells that failed to express housekeeping genes (EF1a and β-actin) or other lineage-specific markers were deemed to lack a cell and were eliminated from further analysis.

Unsupervised hierarchical clustering used Pearson rank coefficient and was completed using GENE-E package (Broad Institute). All samples were included in a single large hierarchical clustering heat map. The principal component analysis (PCA) was conducted using percomp method in R. Loading plots for the first three principal components were used to visualize the relative positioning of samples in the multidimensional space of gene expression. Correlation network analysis was performed using WGCNA package (Langfelder and Horvath, 2008; Saadatpour et al., 2014). Expression values were converted to an adjacency matrix constructed from Pearson correlation coefficients as follows: power adjacency function, \( a_{ij} = |r(x_i,x_j)|^\alpha \) to define the connection strength between pairs of genes \( x_i \) and \( x_j \). Soft power threshold (\( \alpha \)) is applied to network construction to emphasize high correlations and minimize low correlations. Based on the adjacency matrix, topological overlap matrix (TOM) and its modules were calculated and used to construct the dendogram of primers. The TOM was plotted as a heatmap. To obtain more informative topology, power threshold was modified to 4 for WT WKM (Fig. 1), and 3 for T-ALLs (Fig. 6). For confirmation, hierarchical clustering, three-dimensional PCA, one-way ANOVA, violin plots, and correlation values were also generated using the SINGuLAR R package, with limit of detection set to the Ct of 28.

Fisher’s exact test was performed on cell numbers for Figs. 5 B, 6 D, and 6 G. For Fig. 6 G, Fisher’s exact test was performed on cell numbers by establishing cut-off of log2 of detection set to the Ct of 28.

For Fig. S4 shows thymocyte maturation states as assessed for Fig. 6 (D and G). Table 2 shows the order of genes and primer pairs used to generate heat maps shown in Figs. 1, 2, and 4 and Figs. S2 and S3.

**Online supplemental material.** Fig. S1 shows the complete heat map with primer identifiers for the Weighted Gene Co-Expression Analysis shown in Fig. 1 B. Fig. S2 shows unsupervised hierarchical clustering of cells from unlabeled WKM and transgenic reporter lines. Fig. S3 shows unsupervised hierarchical clustering of WT and rag2^{flk106} WKM. Fig. S4 shows thymocyte maturation states as assessed by single-cell qPCR. Table S1 is a list of primer sequences for
outer preamplification and inner qPCR reactions. Table S2 lists the primer order and complete gene name for heat maps depicted in Figs. 1, 2, 4 and Figs. S2 and S3. Table S3 shows limiting dilution cell transplantation analysis and LPC frequency for the Myc-induced T-ALLs shown in Fig. 6. Online supplemental material is available at http://www.jem.org/cgi/content/full/jem.20152013/DC1.

ACKNOWLEDGMENTS

We thank Daniel Haber, David Miyamoto, and Katherine Broderick for technical assistance and use of the BioMark HD. We would like to acknowledge the technical advice and assistance of Ken Livak and Chris Simollardes of Fluidigm for help in optimizing qPCR protocols. We thank Damon Magnuski, Michiha Sharpe, David Adamovich, and Liz Millett for statistical analysis and thoughtful review of this manuscript.

This work was supported by Alex’s Lemonade Stand Foundation (D.M. Langenau), The Live Like Bella Foundation for Childhood Cancer (D.M. Langenau), American Cancer Society (D.M. Langenau), the Massachusetts General Hospital (MGH) Howard Goodman Fellowship (D.M. Langenau), and National Institutes of Health (NIH) grant R24DK016761. FE. Moore is supported by NIH grant 5F32DK098875-03. Flow cytometry and sorting services were supported by MGH Pathology ONY Flow Cytometry Core shared instrumentation grant S10RR023440-01A. The authors declare no competing financial interests.

Submitted: 28 December 2015
Accepted: 17 March 2016

REFERENCES


