Molecularly distinct models of zebrafish *Myc*-induced B cell leukemia

Chiara Borga[1*], Clay A. Foster[1*], Sowmya Iyer[2,3*], Sara P. Garcia, David M. Langenau[2,3]†, J. Kimble Frazer[1]†

[1]Section of Pediatric Hematology-Oncology, Department of Pediatrics, University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104; [2]Department of Pathology and Center for Cancer Research, Massachusetts General Hospital Research Institute, Boston, MA 02114; [3]Harvard Stem Cell Institute, Boston, MA 02114

[*]Co-first authorship contribution

†Equal senior authorship contribution

Corresponding authors:

J. Kimble Frazer, University of Oklahoma Health Sciences Center, Basic Sciences Education Bldg., 941 Stanton L Young, Rm. 229 Oklahoma City, OK 73104, USA. Phone: 405-271-5311 ext. 42375. Fax: 405-271-1025. Email: Kimble-Frazer@ouhsc.edu

David M. Langenau, Harvard Medical School and Massachusetts General Hospital, 149 13th Street, office #6012, Charlestown, Massachusetts 02129, USA. Phone: 617-643-6508. Email: dlangenau@mgh.harvard.edu

Conflict of interest disclosure statement: Authors declare no conflict of interests.

Abstract:

Zebrafish models of T cell acute lymphoblastic leukemia (T-ALL) have been studied for over a decade, but curiously, robust zebrafish B cell ALL (B-ALL) models had not been described. Recently, our laboratories reported two seemingly closely-related models of zebrafish B-ALL. In these genetic lines, the primary difference is expression of either murine or human transgenic c-MYC, each controlled by the zebrafish *rag2* promoter. Here, we compare ALL gene expression in both models. Surprisingly, we find that B-ALL arise in different B cell lineages, with *ighm*$^{+}$ vs. *ighz*$^{+}$ B-ALL driven by murine *Myc* vs. human *MYC*, respectively. Moreover, these B-ALL types exhibit signatures of distinct molecular pathways, further unexpected dissimilarity. Thus, despite sharing analogous genetic makeup, the ALL types in each model are markedly different, proving subtle genetic changes can profoundly impact model organism phenotypes. Investigating the mechanistic differences between mouse and human c-MYC in these contexts may reveal key functional aspects governing MYC-driven oncogenesis in human malignancies.

Zebrafish are a valuable leukemia model due to highly conserved hematopoietic and oncogenic pathways, facile genetics, and ease of use in chemical genetic screens. However, until recently, robust zebrafish B-cell leukemia models had not been described[1, 2]. The first transgenic zebrafish leukemia model was created 15 years ago and targeted murine c-*Myc* (*mMyc*) to thymocytes of AB strain zebrafish, leading to the rapid development of T-cell acute lymphoblastic leukemia (T-ALL)[3]. Additional genetic models were subsequently developed that result in induction of T-ALL[4], but B-cell leukemia models lagged behind[5, 6].

In two recent reports published in *Leukemia*, our groups independently demonstrated the development of zebrafish B-ALL using transgenic expression of *mMyc* or human c-*MYC* (*hMYC*) controlled by the zebrafish recombination activating gene 2 (*rag2*) promoter[1, 2]. Both models shared genetic and phenotypic features, but there were also key differences including strain background and species differences in the MYC transgene that was used to generate each model. Here, we compare and contrast these models, making the important finding that zebrafish develop at least four molecularly distinct ALL types, including cortical thymocyte-arrested *cd4*$^{+}$/*cd8*$^{+}$ T-ALL, *ighm*$^{+}$ B-ALL, *ighz*$^{+}$ B-ALL, and biphenotypic T/B-ALL.

Thirteen ALLs were purified from *rag2:hMYC;lck:eGFP* double-transgenic fish[2]. As previously reported, these leukemias had heterogeneous GFP expression, with T-ALL being exclusively GFP^{hi}, B-ALL exclusively GFP^{lo}, and other fish harboring mixed-ALL with both GFP^{hi} and GFP^{lo} cells, representing simultaneous T- and B-ALL, respectively[2]. Leukemias were subjected to RNA-seq transcriptomic profiling and compared with transplanted leukemias generated from single ALL clones described by Garcia et al., 2018[1]. Principal Component Analysis clearly distinguished *mMyc*-induced T-ALL from B-ALL, with the single *mMyc*-induced biphenotypic B/T-ALL clustering between these samples (Fig. 1A). The eight *hMYC*-

induced ALLs that were largely GFP^{hi} clustered with known T-ALLs (2, 6, 8-11, 13, 14), while three primarily GFP^{lo} ALL clustered near the *mMyc* B-ALLs (3-5). Two *hMYC*-induced ALLs with substantial populations of both GFP^{lo} and GFP^{hi} cells grouped near the *mMyc* biphenotypic B/T leukemia. Hierarchical clustering using the top 100 positively- and negatively-correlated genes from PC2 confirmed that these genes defined B and T lymphocytes, respectively (Fig. 1B), with B-ALLs expressing *cd79b*, *syk*, *pax5*, *blnk* and *efb1*, while T-ALLs expressed *cd8b*, *lck*, *runx3* and *gata3*. As previously reported, the biphenotypic B/T-ALL and mixed *hMYC*-induced ALLs expressed both T- and B-cell lineage genes[1, 2]. PC2 up-regulated genes were enriched for B cell signaling pathways when independently assessed by GSEAsig (Supplemental Tables 1 and 2).

To examine clonality and maturation in *mMyc*- and *hMYC*-induced ALL, we next analyzed the expression of constant and variable regions of the T cell receptor β (*tcrβ*) and immunoglobulins μ and ζ (*ighm*, *ighz*; Fig. 2A). Every *mMyc* and *hMYC* T-ALL exhibited *tcrβ* expression, with V(D)J recombination occurring in most samples as determined by expression of specific variable regions[1]. Conversely, *mMyc* and *hMYC* B-ALL did not recombine or express *tcrβ*, but expressed constant regions of *ighm* or *ighz*. Ig variable regions were not detected, indicating V(D)J rearrangement likely had not occurred in these leukemias and suggesting B-ALLs arrest at the early pro-B cell stage. As expected, *hMYC* mixed-ALL contained distinct T- and B-ALL clones expressing both *tcr* and *ig* mRNAs, with their relative expression correlating well with the percentage of GFP^{hi}/T-ALL vs. GFP^{lo}/B-ALL cells found in each sample (Fig. 2A). Intriguingly, *mMyc* B-ALL expressed exclusively *ighm* while *hMYC* B-ALL favored *ighz* expression, indicating that *mMyc* and *hMYC* might be oncogenic in distinct B cell lineages.

To further explore differences between these models, we next identified genes uniquely-expressed by T-ALL, *mMyc*/*ighm*$^{+}$ B-ALL, or *hMYC*/*ighz*$^{+}$ B-ALL (Fig. 2B). As expected T-ALLs expressed known T cell lineage markers, yet *mMyc*/*ighm*$^{+}$ and *hMYC*/*ighz*$^{+}$ B-ALLs were transcriptionally distinct. *mMyc*/*ighm*$^{+}$ B-ALL expressed *gfi1ab*, *zfhx3*, *notch1a*, *nf1b*, and *gtf3aa*. By contrast, *hMYC*/*ighz*$^{+}$ B-ALL expressed higher *cd79a*, *cd83*, *mef2cb*, and *jak2a* levels. To further test for differences in these two molecular subtypes of B-ALL, we next performed GSEAsig using these same differentially-regulated genes. From this analysis, we uncovered that *mMyc*/*ighm*$^{+}$ B-ALLs exhibited significant enrichment for pathways regulating ribosome biogenesis and RNA binding (Fig. 2C and Supplemental Tables 3 and 4). By contrast, *hMYC*/*ighz*$^{+}$ B-ALLs were enriched for intracellular signaling, protein binding, and germinal center B cell maturation pathways. In support of our findings, Liu et al. recently reported the identification of molecularly and biological distinct *ighz*$^{+}$ and *ighm*$^{+}$ B cell lineages using *rag2:mCherry; cd79b:GFP* transgenic zebrafish[7]. In the context of normal B cell development, *ighz*$^{+}$ B cells were mCherry^{hi}/GFP^{lo} while *ighm*$^{+}$ B cells were mCherry^{hi}/GFP^{hi}. Overall, these results demonstrate that *mMyc*/*ighm*$^{+}$ and *hMYC*/*ighz*$^{+}$ B-ALLs are not subtle B cell leukemia variants, but rather distinct malignancies that arise in different B cell types with vastly different molecular pathway signatures.

In summary, although zebrafish B cell leukemia models were lacking for many years, our analyses reveal two highly-divergent types of B-ALL. This is surprising, as both models utilize the same promoter (*rag2*) to regulate a near-identical oncoprotein, c-Myc/MYC, with the only differences being the MYC transgene species of origin and the genetic backgrounds upon which the models were developed. Yet, despite the high molecular similarity of both models, these B-ALL subtypes also show unique gene expression signatures when compared to one another,

104 which likely reflects differences in both their lineage (*ighm* vs. *ighz*) and potential differences in
105 MYC transcriptional targets expressed by the early developmental stages of these distinct pro-B
106 cell populations. Our new analysis of these models reconciles the perceived differences in the
107 manuscripts published by our groups, identifying four molecularly distinct ALL subtypes in
108 zebrafish: cortical *cd4*$^{+}$/*cd8*$^{+}$ T-ALL, biphenotypic B/T ALL, *ighm*$^{+}$ B-ALL, and *ighz*$^{+}$ B-ALL.
109 Developing a wider array of leukemia models and refining mechanisms that drive their growth,
110 aggression, and stem cell frequency will surely lead to new insights into human disease.

111 **Figure Legends:**

112 **Figure 1. Gene expression differences define T- vs. B-lineages in ALL of both the**
113 ***rag2:mMyc* and *rag2:hMYC;lck:eGFP* models.** (**A**) Principal component analysis of RNAseq
114 expression profiles of previously-classified *mMyc*-induced T- (blue, n=8), B- (red, n=2), or
115 biphenotypic ALL (pink, n=1) and compared with 13 unknown *hMYC*-induced ALL (black). (**B**)
116 Heat map and hierarchical clustering using the top 100 positively- and negatively-correlated
117 genes from PC2. B cell-specific genes are denoted in blue and T cell-specific genes in red at the
118 right.

119

120 **Figure 2. Identification of two molecularly-distinct B-ALL types arising independently in**
121 **either the *ighm*$^{+}$ or *ighz*$^{+}$ B cell lineages.** (**A**) T cell receptor beta and Ig heavy chain expression
122 in individual ALLs. *tcrβ* recombination is denoted by grey-shaded boxes (left column), with
123 percentage of GFP^{hi} cells in each *rag2:hMYC;lck:eGFP* ALL noted in right column.
124 Histograms depict expression of *tcrβ* and *igh* constant regions by each ALL. Not available (NA).
125 (**B**) Heatmap showing expression of genes differentially expressed in T-ALL, *mMyc*/*ighm*$^{+}$ B-
126 ALL, and *hMYC*/*ighz*$^{+}$ B-ALL. (**C**) Gene set enrichment analysis using genes positively
127 correlated with each ALL molecular subtype. T-ALL (blue), *mMyc*/*ighm*$^{+}$ B-ALL (red), and
128 *hMYC*/*ighz*$^{+}$ B-ALL (orange). Complete geneset and GSEAsig results are provided in Supp.
129 Tables 3 and 4.

References:

1. Garcia EG, Iyer S, Garcia SP, Loontiens S, Sadreyev RI, Speleman F, *et al.* Cell of origin dictates aggression and stem cell number in acute lymphoblastic leukemia. *Leukemia* 2018 2018/04/18.

2. Borga C, Park G, Foster C, Burroughs-Garcia J, Marchesin M, Shah R, *et al.* Simultaneous B and T cell acute lymphoblastic leukemias in zebrafish driven by transgenic MYC: implications for oncogenesis and lymphopoiesis. *Leukemia* 2018 2018/08/15.

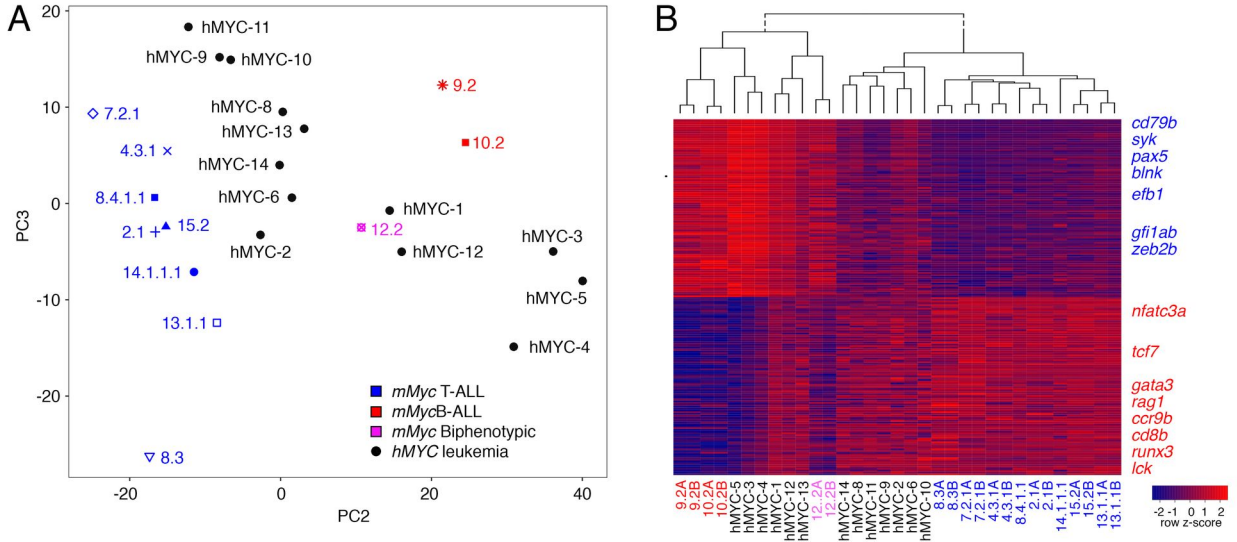
3. Langenau DM, Traver D, Ferrando AA, Kutok JL, Aster JC, Kanki JP, *et al.* Myc-induced T cell leukemia in transgenic zebrafish. *Science* 2003 Feb 7; **299**(5608): 887-890.

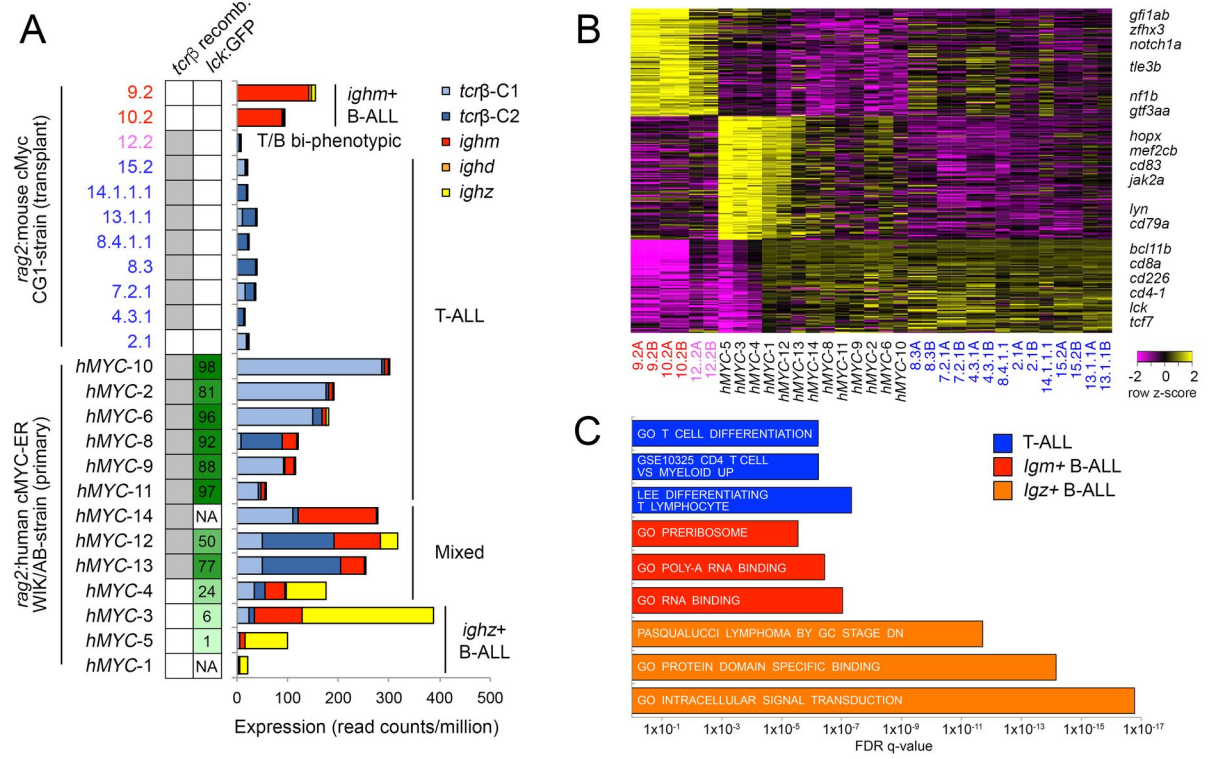
4. Frazer JK, Meeker ND, Rudner L, Bradley DF, Smith AC, Demarest B, *et al.* Heritable T-cell malignancy models established in a zebrafish phenotypic screen. *Leukemia* 2009 Oct; **23**(10): 1825-1835.

5. He S, Jing CB, Look AT. Zebrafish models of leukemia. *Methods Cell Biol* 2017; **138:** 563-592.

6. Squiban B, Frazer JK. Danio rerio: Small Fish Making a Big Splash in Leukemia. *Curr Pathobiol Rep* 2014 Jun; **2**(2): 61-73.

7. Liu X, Li YS, Shinton SA, Rhodes J, Tang L, Feng H, *et al.* Zebrafish B Cell Development without a Pre-B Cell Stage, Revealed by CD79 Fluorescence Reporter Transgenes. *J Immunol* 2017 Sep 1; **199**(5): 1706-1715.


A
hMYC-11
hMYC-9
hMYC-10
9.2
7.2.1
hMYC-8
hMYC-13
10.2
4.3.1
hMYC-14
8.4.1.1
hMYC-6
hMYC-1
2.1
15.2
12.2
hMYC-3
hMYC-2
hMYC-12
14.1.1.1
hMYC-5
13.1.1
hMYC-4
8.3
mMyc T-ALL
mMycB-ALL
mMyc Biphenotypic
hMYC leukemia
PC3
PC2
20
10
0
-10
-20
-20
0
20
40
B
cd79b
syk
pax5
blnk
efb1
gfi1ab
zeb2b
nfatc3a
tcf7
gata3
rag1
ccr9b
cd8b
runx3
lck
9.2A
9.2B
9.2A
10.2B
hMYC-3
hMYC-4
hMYC-1
hMYC-12
hMYC-13
12.2A
12.2B
hMYC-14
hMYC-8
hMYC-11
hMYC-9
hMYC-2
hMYC-6
hMYC-10
8.3A
8.3B
7.2.1A
7.2.1B
4.3.1A
4.3.1B
8.4.1.1
2.1A
2.1B
14.1.1.1
15.2A
15.2B
13.1.1A
13.1.1B
-2 -1 0 1 2
row z-score


A
tcrβ recomb.
lck:GFP
rag2:mouse cMyc
CG1-strain (transplant)
rag2:human cMYC-ER
WIK/AB-strain (primary)
9.2
10.2
12.2
15.2
14.1.1.1
13.1.1
8.4.1.1
8.3
7.2.1
4.3.1
2.1
hMYC-10
hMYC-2
hMYC-6
hMYC-8
hMYC-9
hMYC-11
hMYC-14
hMYC-12
hMYC-13
hMYC-4
hMYC-3
hMYC-5
hMYC-1
98
81
96
92
88
97
NA
50
77
24
6
1
NA
ighm+
B-ALL
T/B bi-phenotypic
T-ALL
Mixed
ighz+
B-ALL
tcrβ-C1
tcrβ-C2
ighm
ighd
ighz
0
100
200
300
400
500
Expression (read counts/million)
B
gfi1ab
zfhx3
notch1a
tle3b
nf1b
gtf3aa
hopx
mef2cb
cd83
jak2a
lyn
cd79a
bcl11b
cd8a
cd226
cd4-1
lck
tcf7
9.2A
9.2B
10.2A
10.2B
12.2A
12.2B
hMYC-5
hMYC-3
hMYC-4
hMYC-1
hMYC-12
hMYC-13
hMYC-14
hMYC-8
hMYC-11
hMYC-9
hMYC-2
hMYC-6
hMYC-10
8.3A
8.3B
7.2.1A
7.2.1B
4.3.1A
4.3.1B
8.4.1.1
2.1A
2.1B
14.1.1.1
15.2A
15.2B
13.1.1A
13.1.1B
-2
0
2
row z-score
C
GO T CELL DIFFERENTIATION
GSE10325 CD4 T CELL VS MYELOID UP
LEE DIFFERENTIATING T LYMPHOCYTE
GO PRERIBOSOME
GO POLY-A RNA BINDING
GO RNA BINDING
PASQUALUCCI LYMPHOMA BY GC STAGE DN
GO PROTEIN DOMAIN SPECIFIC BINDING
GO INTRACELLULAR SIGNAL TRANSDUCTION
T-ALL
Igm+ B-ALL
Igz+ B-ALL
1x10-1
1x10-3
1x10-5
1x10-7
1x10-9
1x10-11
1x10-13
1x10-15
1x10-17
FDR q-value