Supplemental Information

Constitutive Ras Signaling and *Ink4a/Arf* Inactivation Cooperate During the Development of B-ALL in Mice

Tomasz Sewastianik[1,2], Meng Jiang[1,3], Kumar Sukhdeo[4,5], Sanjay S. Patel[6], Kathryn Roberts[7], Yue Kang[1], Ahmad Alduaij[8], Peter S. Dennis[1], Brian Lawney[9], Ruiyang Liu[1], Zeyuan Song[1], Jessie Xiong[10], Yunyu Zhang[11], Madeleine E. Lemieux[12], Geraldine S. Pinkus[6], Jeremy N. Rich[4], David M. Weinstock[11], Charles G. Mullighan[7], Norman E. Sharpless[10], and Ruben D. Carrasco[1,6,†]

[1]Department of Oncologic Pathology, Dana-Farber Cancer Institute, Boston, MA, USA; [2]Department of Experimental Hematology, Institute of Hematology and Transfusion Medicine, Warsaw, Poland; [3]Department of Surgical Oncology, The Fourth Affiliated Hospital of Harbin Medical University, Harbin, China; [4]Department of Stem Cell Biology and Regenerative Medicine, Cleveland Clinic, Cleveland, OH, USA; [5]Department of Pathology, Case Western Reserve University, Cleveland, OH, USA; [6]Department of Pathology, Brigham & Women's Hospital, Boston, MA, USA; [7]Department of Pathology, St Jude Children's Research Hospital, Memphis, TN, USA; [8]Pathology and Laboratory Medicine Institute, Cleveland Clinic Abu Dhabi, Abu Dhabi, United Arab Emirates; [9]Center for Computational Cancer Biology, Dana-Farber Cancer Institute, Boston, MA, USA; [10]Lineberger Comprehensive Cancer Center, School of Medicine, University of North Carolina, Chapel Hill, NC, USA; [11]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA, USA; [12]Bioinfo, Plantagenet, ON, Canada.

Corresponding author[†]: Ruben D. Carrasco, address: Dana-Farber Cancer Institute, 450 Brookline Ave JF215H, Boston, MA 02215; phone: 617-582-8159; fax: 617-582-8761; email: ruben_carrasco@dfci.harvard.edu

Supplemental Methods

Antibodies used for IHC, IF, Flow cytometry and WB

IHC: Cyclin D2 (M-20, Santa Cruz Biotechnology), B220/CD45R (RA3-6B2, BD Bioscience), CD138 (281-2, BD Bioscience), BCL6 (N-3, Santa Cruz Biotechnology), CD3 (A0452, DAKO), CD5 (53-7.3, BD Bioscience), CD10 (EPR5904, LifeSpan Biosciences, Inc.), IgM (BA-2020, Vector Lab), TdT (005, Supertechs, Inc.), Ki-67 (VPK451, Vectror Lab), Arf (5-C3-1, Santa Cruz Biotechnology), Myc (Y69, Abcam); IF: CD19 (6D5, Biolegend); Flow cytometry: B220/CD45R (553087, BD Biosciences); WB: $p16^{Ink4a}$ (M-156, Santa Cruz Biotechnology), Cyclin D2 (M-20, Santa Cruz Biotechnology), and Actin (C-11, Santa Cruz Biotechnology).

Analysis of microarray gene expression data

Arrays were corrected for background, normalized, and log2-transformed using the rma function of the affy Bioconductor package[1-4]. Present/absent calls were made using the mas5calls function of the affy package. Probe sets present in >20% of samples and for which the interquartile range was >log2(1.2) were retained for further analysis. The Bioconductor limma package[5] was used to identify probe sets significantly up- or downregulated. For significance comparisons between premalignant groups (n=2 each), a combination of *P* value <0.01 and fold-change >2 as the cut-off were used. A more stringent adjusted *P* value of <0.01 (Benjamini-Hochberg method[6]) and a fold-change >2 were used for comparisons between premalignant samples (n=2) and $CD19^{Cre/+};Kras^{G12D/+};Ink4a/Arf^{1/+}$ tumors (n=10).

To compare gene expression profiles of $CD19^{Cre/+};Kras^{G12D/+};Ink4a/Arf^{1/+}$ tumors and subpopulations of normal mouse B-cells[7], we used probe sets common to both array platforms and corrected for batch effects with ComBat[8]. Probe sets were filtered after batch correction to retain only those that passed filtering criteria described above for both data sets. In total, 9490 probe sets were retained for unsupervised hierarchical clustering based on Pearson's correlation coefficient. Determination of

differential gene expression between tumors and normal pre-B and pro-B-cells was carried out using the 8962 probe sets that satisfied filtering criteria for these arrays (i.e. excluding mature and immature B-cells as well as premalignant and Cre-only samples). In the latter case, we used an adjusted *P* value of <0.05 and fold-change of >2 as our combined cut-off. Genes satisfying these criteria were used as novel gene signatures. Signature gene symbols were mapped from mouse to human via HomoloGene (April 2012 download) for use with human ALL gene expression profiles and were included with the "Hallmark" gene sets (v.5.1) downloaded from the Broad Institute's MSigDB website (http://software.broadinstitute.org/gsea/msigdb/collections.jsp). These combined gene sets were used for gene set enrichment analysis (GSEA)[9]. Affymetrix Human Genome U133A arrays for series GSE12995[10] were downloaded from the Gene Expression Omnibus (GEO) and processed as above. We used 10 samples from each of the 5 following ALL subtypes: TCF3-PBX1, hyperdiploid, MLL, BCR-ABL, and TEL-AML1. Data series GSE11877[11,12] was downloaded from GEO as a MAS5-summarized scaled-expression matrix and further log2-transformed and quantile normalized before use. GSEA for human arrays was carried out using Signal-to-Noise as the metric with base 10 values (not log2 transformed). GSEA for murine premalignant and tumor samples were run on values pre-ranked by log2 fold-change. GSEA signature heat maps showed log10(FDR) x sign(NES) values such that correlation with phenotype A in an A vs B comparison was positive (NES >0) and correlation with phenotype B is negative (NES <0). Heat maps were generated using the R package pheatmap[13].

Analysis of whole-exome sequencing data

Lane-specific sequence files in fastQ format were aligned with BWA-mem[14] to the Ensembl GRCm38.75 reference genome and further processed using the Genome Analysis Toolkit (GATK)[15] best practices for targeted exome sequencing. This included duplicate-read marking, base-quality score recalibration, and local *de novo* indel realignment. Read groups were assigned and lane-specific files were merged according to GATK best-practice recommendations.

BAM-format files were loaded into the GATK's Mutect2[16] tool for variant-calling analysis. Control samples were processed using Mutect2's recommendations for creating a panel of normals (PON); variant calling was performed individually for each sample and the resulting VCF files were combined using GATK's CombineVariants tool for each phenotype (LG = low-grade B-cell lymphoproliferative disorder and B-ALL = precursor B-cell acute lymphoblastic leukemia). The variants found in the PON were used when analyzing variants in the tumor-derived samples (i.e., normal_panel argument). All calls to Mutect2 methods were made with default arguments.

Resultant sample-specific VCF files were merged and annotated using Ensembl Variant Effect Predictor (VEP)[17] and stored in a Gemini[18] database. Prior to annotation, variants were limited to those that passed Mutect2's standard quality filters (PASS status). Further, we examined potentially high-impact variants annotated as stop-gain, frameshift, splice-acceptor, splice-donor, stop-lost, start-lost, and missense.

Remaining variants were subsequently analyzed to determine recurrent, well-covered sites featuring a high percentage of reads supporting the alternate allele. Analysis was limited variant loci found in a majority of samples (>5 for ALL, >3 for LG), having a minimum read depth of 40 reads, and with alternate allele percentage of >75%. These high-quality sites were cross-referenced to those determined in the PON analysis to remove false-positive somatic variants. Indeed, many such variants were also found in the PON (often in a single sample), apparently not removed by Mutect2 in the initial enumeration of somatic variants. Following this filtration, 36 (B-ALL) and 35 (LG) sites remained with seemingly high-quality recurrent somatic variants. Visual inspection (via IGV) was performed on these sites, revealing that the "normal" samples harbored similar mutations; such sites were not included in the PON set due to lower sequencing coverage.

Supplemental Figure Legends

Supplemental Figure 1. Conditional loss of *Ink4a/Arf* and activation of $Kras^{G12D}$ expression in $CD19^{+}$ B-cells.

(A) Southern blot of *PstI*-digested genomic DNA from the indicated tissues and mice at 8 weeks of age. The wild-type (WT) or lox-*Ink4a/Arf* (*Cdkn2a*) allele migrates at 9.0 kb, and the recombined *Ink4a/Arf* null allele (KO) corresponds to the shorter 4.6 kb band. Note that excision of the locus only occurs in the presence of Cre recombinase.

(B) Immunoblots of $p16^{Ink4a}$ expression in $CD19^{+}$ splenocytes from mice at 8 weeks of age. Actin was used as loading control.

(C) RT-PCR/RFLP analysis of *Lox-STOP-Lox* $Kras^{G12D}$ allele recombination in testicles or $CD19^{+}$ splenocytes from mice at 8 weeks of age. PCR-amplified cDNA was untreated (-) or digested with *HindIII* (+). Recombination of the *Lox-STOP-Lox* $Kras^{G12D}$ allele introduces a *HindIII* site into the $Kras^{G12D}$ transcript that results in a released fragment (TG) upon digestion with restriction enzyme.

Supplemental Figure 2. Histologic and immunohistochemical evaluation of premalignant mice.

(A) and **(B)** Histological and immunohistochemical analyses of CD3, B220, and TdT expression in lymph nodes (A) and bone marrows (B) from one additional mouse of each genotype. Positive control (insert): TdT staining of thymus from the same mouse. Scale bars: A = 1 mm, B = 50 µm.

Supplemental Figure 3. Gene set enrichment analysis of $CD19^{+}$ B-cells from premalignant mice.

(A), **(B)**, and **(C)** Mountain plots documenting pathways upregulated in $CD19^{+}$ B-cells from mice of specified genotypes compared to $CD19^{Cre/+}$ controls. Top three hallmark signatures from each genotype (A). Selected 'curated' (MSigDB C2) signatures (B) and (C).

Supplemental Figure 4. Histologic and immunohistochemical evaluation of low-grade B-cell lymphoproliferative disorder from *CD19$^{Cre/+}$;Kras$^{G12D/+}$* mice.

Representative histological and immunohistochemical analyses of B220, BCL6, TdT, CD3, CD5, and CD138 expression in a nodular lymphoid infiltrate in lungs from additional *CD19$^{Cre/+}$* and *CD19$^{Cre/+}$;Kras$^{G12D/+}$* mice. Positive control: TdT staining of thymus from the same animal. Scale bars: 50 µm.

Supplemental Figure 5. Histologic and immunohistochemical evaluation of precursor B-ALL in *CD19$^{Cre/+}$;Ink4a/Arf$^{L/+}$* and *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice.

Representative histologic and immunohistochemical analyses of TdT expression (in-frame) in lymph nodes, BMs, and livers of additional *CD19$^{Cre/+}$*, *CD19$^{Cre/+}$;Ink4a/Arf$^{L/+}$*, and *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice. Scale bars: black – 1 mm, white - 10 µm.

Supplemental Figure 6. Analysis of mutations and copy number variations in mouse tumors.

(A) Summary of genomic profiles of ten B-ALLs from *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice. Integer-value recurrence of CNAs across the samples in segmented data (y axis) is plotted for each probe evenly aligned along the x axis in chromosomal order. Because, *Ink4a/Arf* deletion was relatively small and focal, it was omitted by the algorithm used for data segmentation.

(B) Analysis of the probe specific for the *Ink4a/Arf* confirms the deletion.

(C) Representative immunohistochemical analyses of Myc expression in lymph nodes of mice with pre-B-ALL with (n=3) and without (n=3) chromosome 15 amplification from *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice. Scale bars: 20 µm.

(D) Mean coverages of WES for low-grade B-cell lymphomas (n=6; LG) from *CD19$^{Cre/+}$;Kras$^{G12D/+}$* mice and B-ALLs (n=10; ALL) from *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice as well as two controls per group (c1 and

c2). Dotted lines represent mean for tumor group. Percentiles for the coverage depth are shown below the graph.

Supplemental Figure 7. Gene expression analysis of murine B-ALL.

Mountain plots documenting nine pathways most highly upregulated in *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* B-ALL cells compared to *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* premalignant CD19^{+} B-cells.

Supplemental Figure 8. Gene expression analysis of human B-ALL.

Heatmap of leading edge genes from GSEA comparing the relative enrichment of genes up-regulated in *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* tumors relative to normal pre-B-cells ("Tumor vs Pre-B up" signature) in human BCR-ABL B-ALL samples versus all other B-ALLs.

Supplemental Table 1. Differential gene expression analysis of CD19^{+} B-cells from 8-week-old engineered mice and of B-ALL tumors from *CD19$^{Cre/+}$;Kras$^{G12D/+}$;Ink4a/Arf$^{L/+}$* mice.

Supplemental References

1. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res*. 2003;31(4):e15.

2. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics*. 2003;4(2):249-264.

3. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics*. 2003;19(2):185-193.

4. R Development Core Team. R: A Language and Environment for Statistical Computing. Vienna, Austria R Foundation for Statistical Computing; 2010.

5. Smyth GK. limma: Linear Models for Microarray Data. In: Gentleman R, Carey VJ, Huber W, Irizarry RA, Dudoit S, eds. Bioinformatics and Computational Biology Solutions Using R and Bioconductor. New York, NY: Springer New York; 2005:397-420.

6. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)*. 1995;57(1):289-300.

7. Krivtsov AV, Feng Z, Lemieux ME, et al. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. *Cancer Cell*. 2008;14(5):355-368.

8. Johnson WE, Li C, Rabinovic A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics*. 2007;8(1):118-127.

9. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A*. 2005;102(43):15545-15550.

10. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009;360(5):470-480.

11. Kang H, Chen IM, Wilson CS, et al. Gene expression classifiers for relapse-free survival and minimal residual disease improve risk classification and outcome prediction in pediatric B-precursor acute lymphoblastic leukemia. *Blood*. 2010;115(7):1394-1405.

12. Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116(23):4874-4884.

13. Kolde R. pheatmap: Pretty Heatmaps. Vol. 2017; 2015:Implementation of heatmaps that offers more control over dimensions and appearance.

14. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589-595.

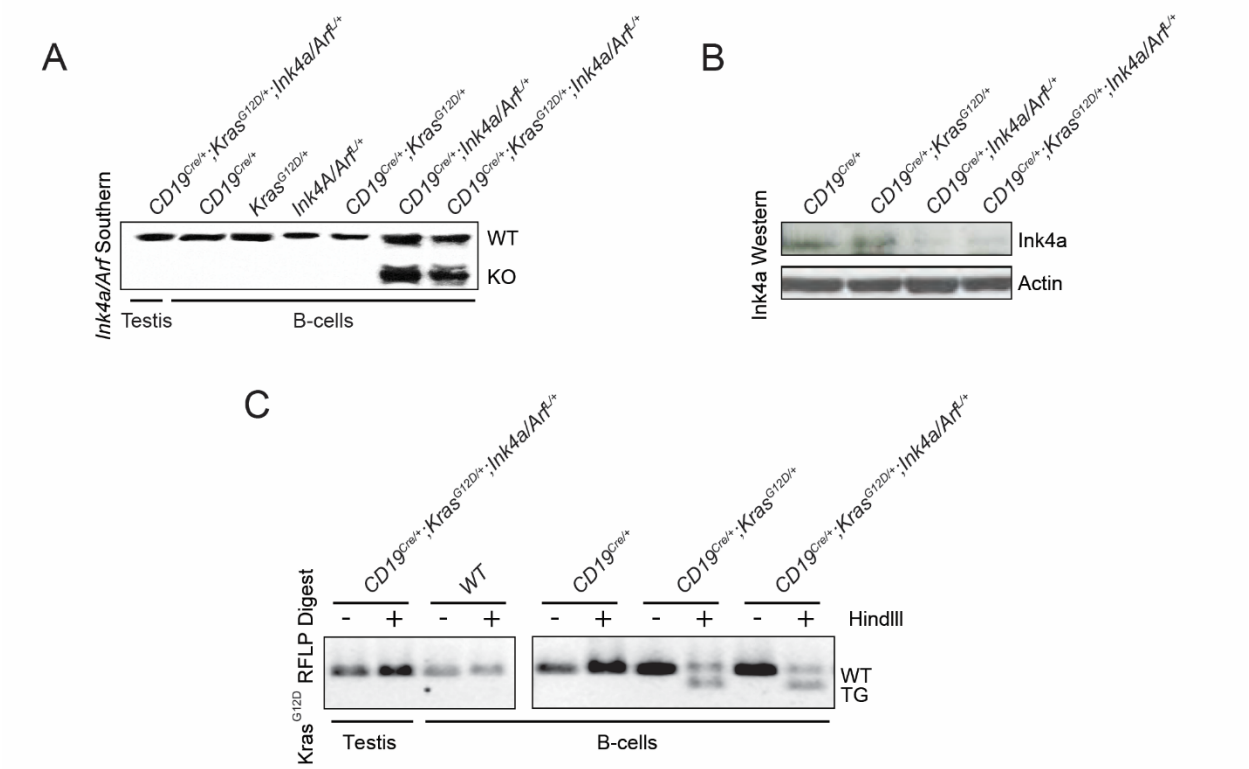
15. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 2011;43(5):491-498.

16. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213-219.

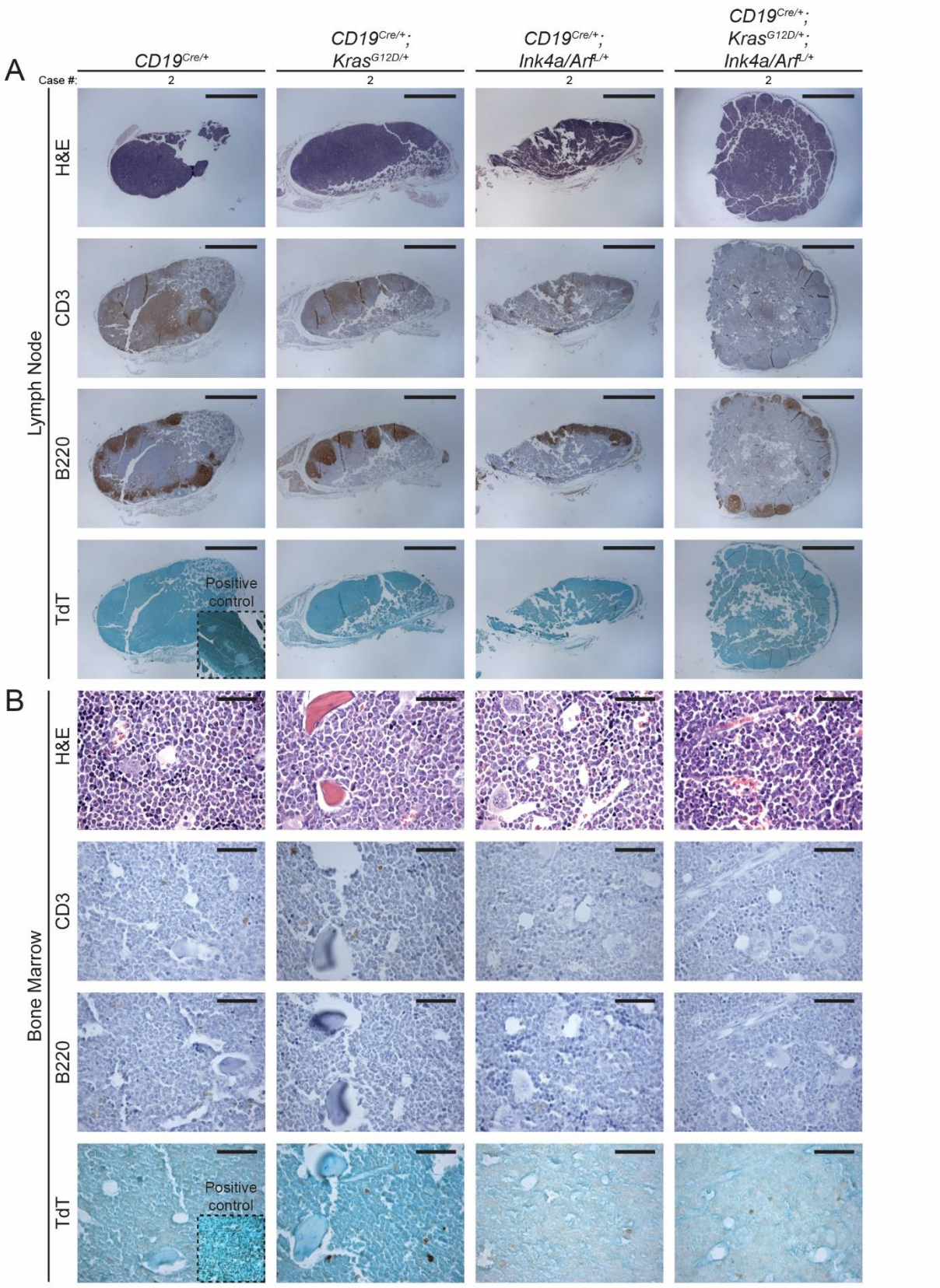
17. McLaren W, Gil L, Hunt SE, et al. The Ensembl Variant Effect Predictor. *Genome Biol*. 2016;17(1):122.

18. Paila U, Chapman BA, Kirchner R, Quinlan AR. GEMINI: integrative exploration of genetic variation and genome annotations. *PLoS Comput Biol*. 2013;9(7):e1003153.

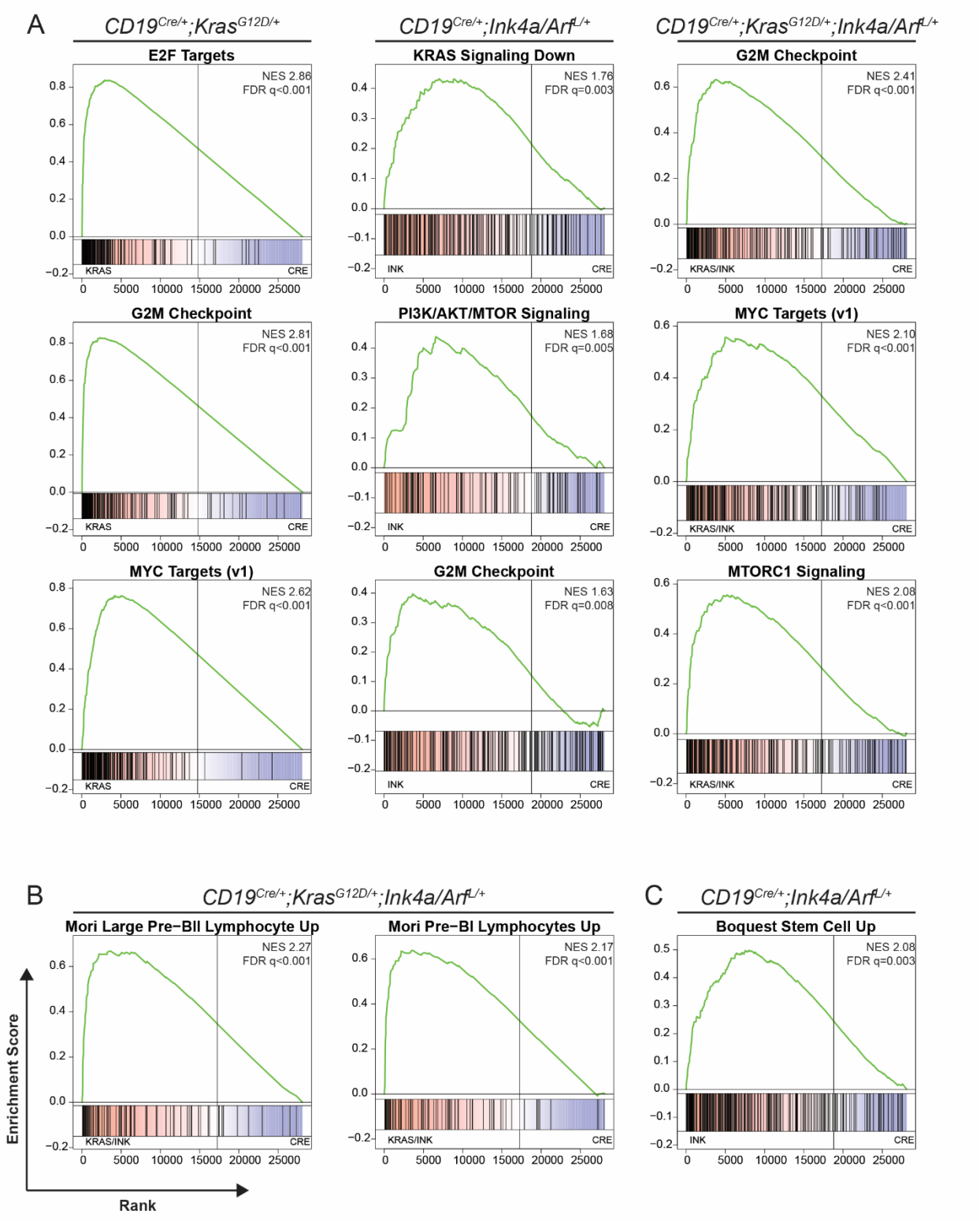
Supplemental Figures and Tables



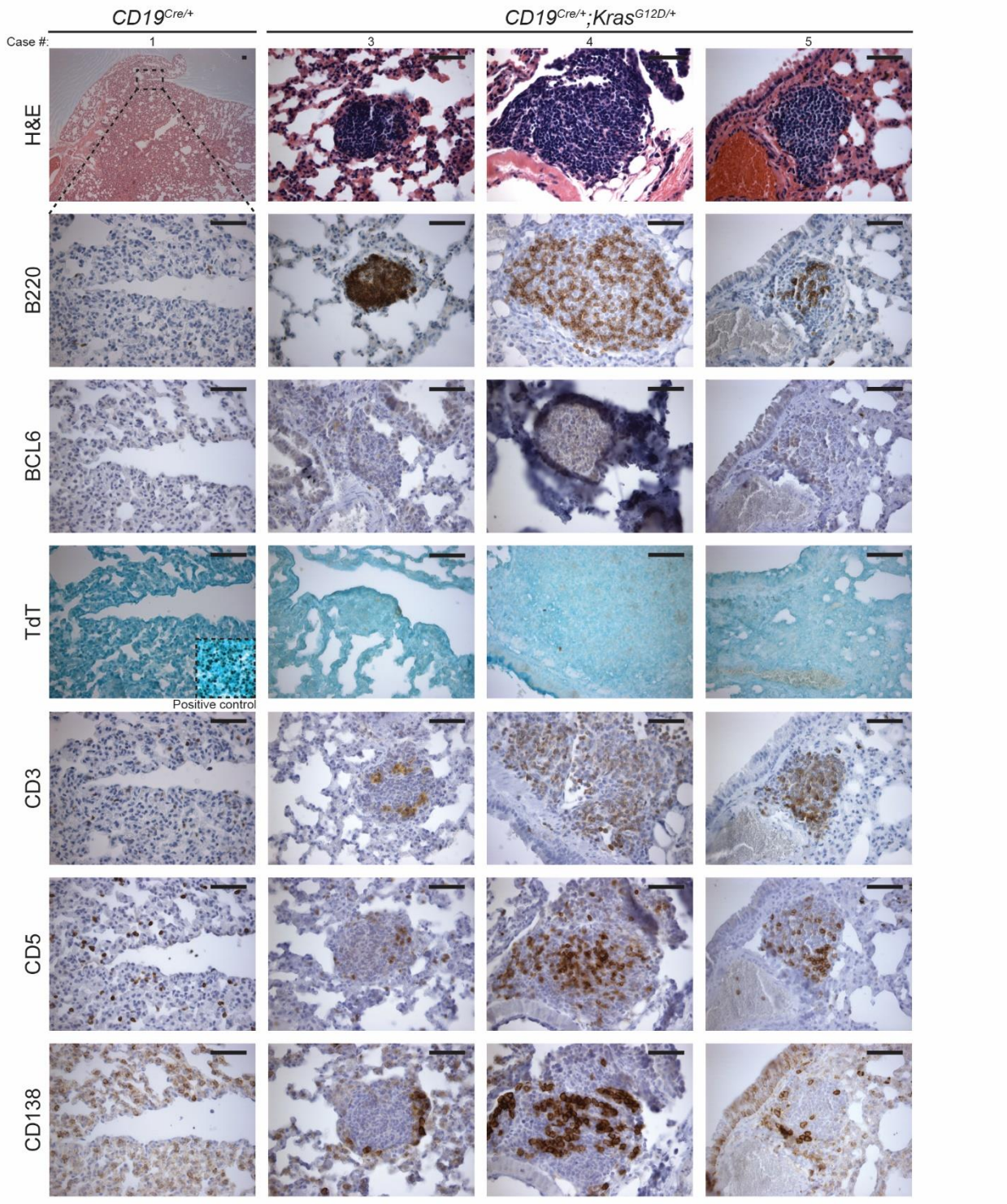
Supplemental Figure 1



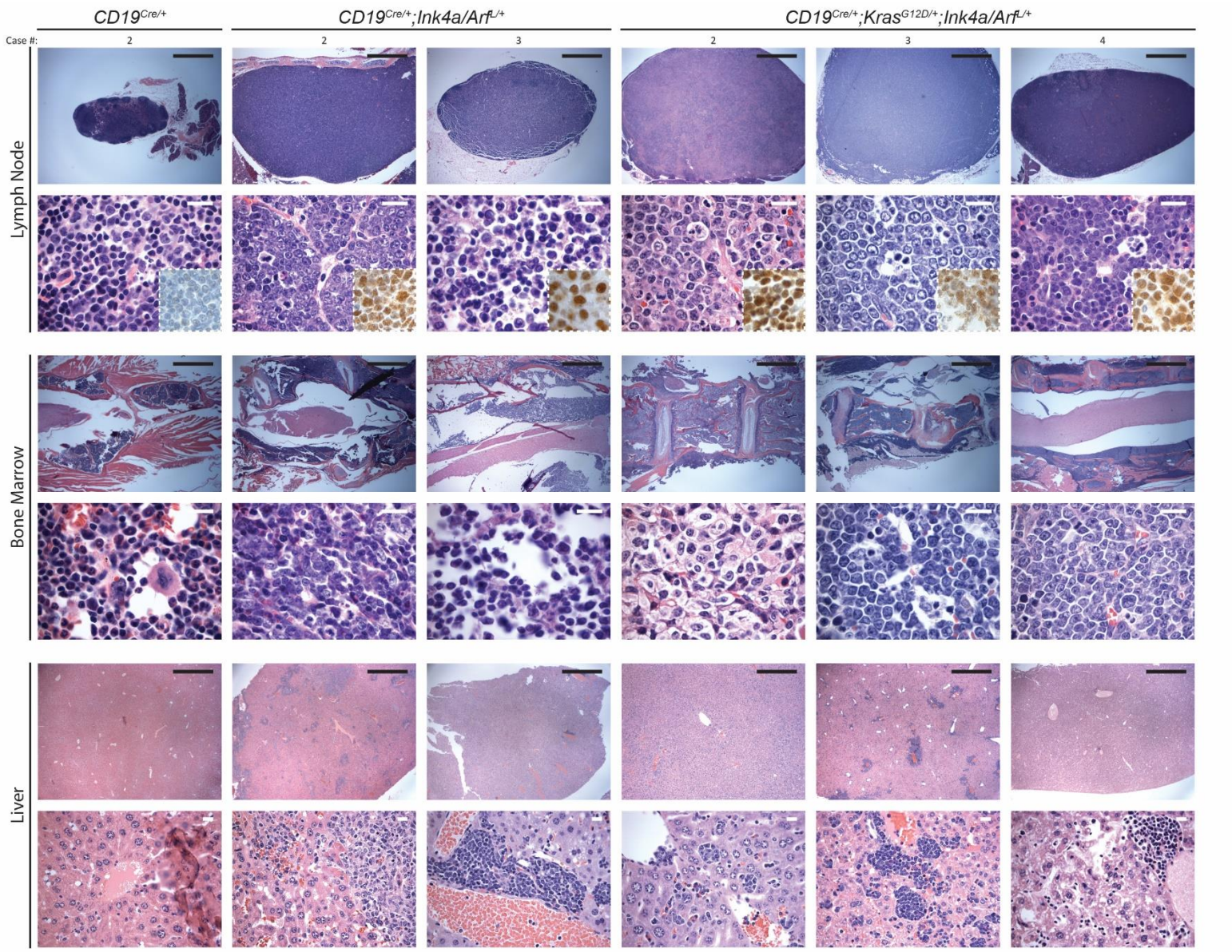
Supplemental Figure 2



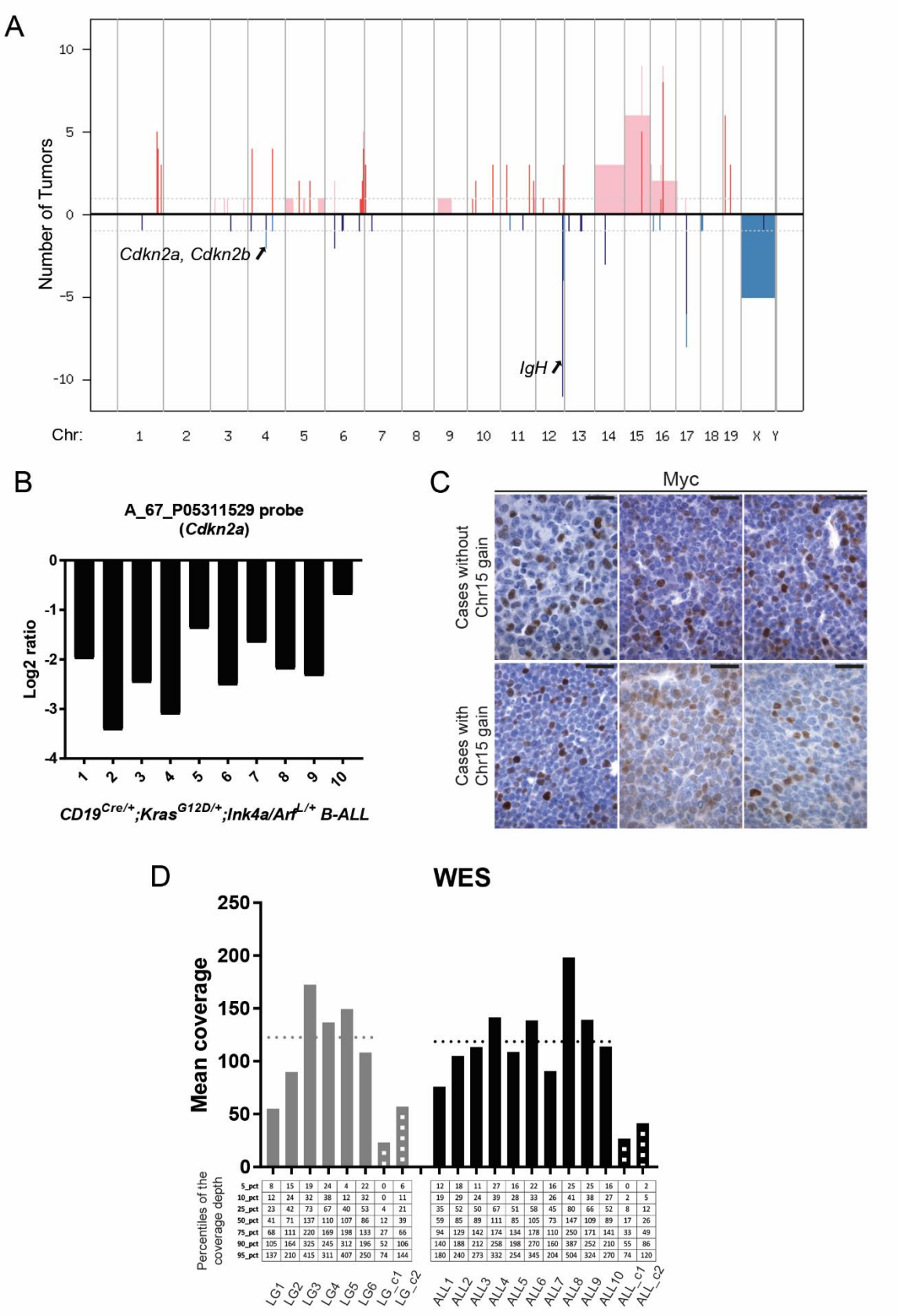
Supplemental Figure 3



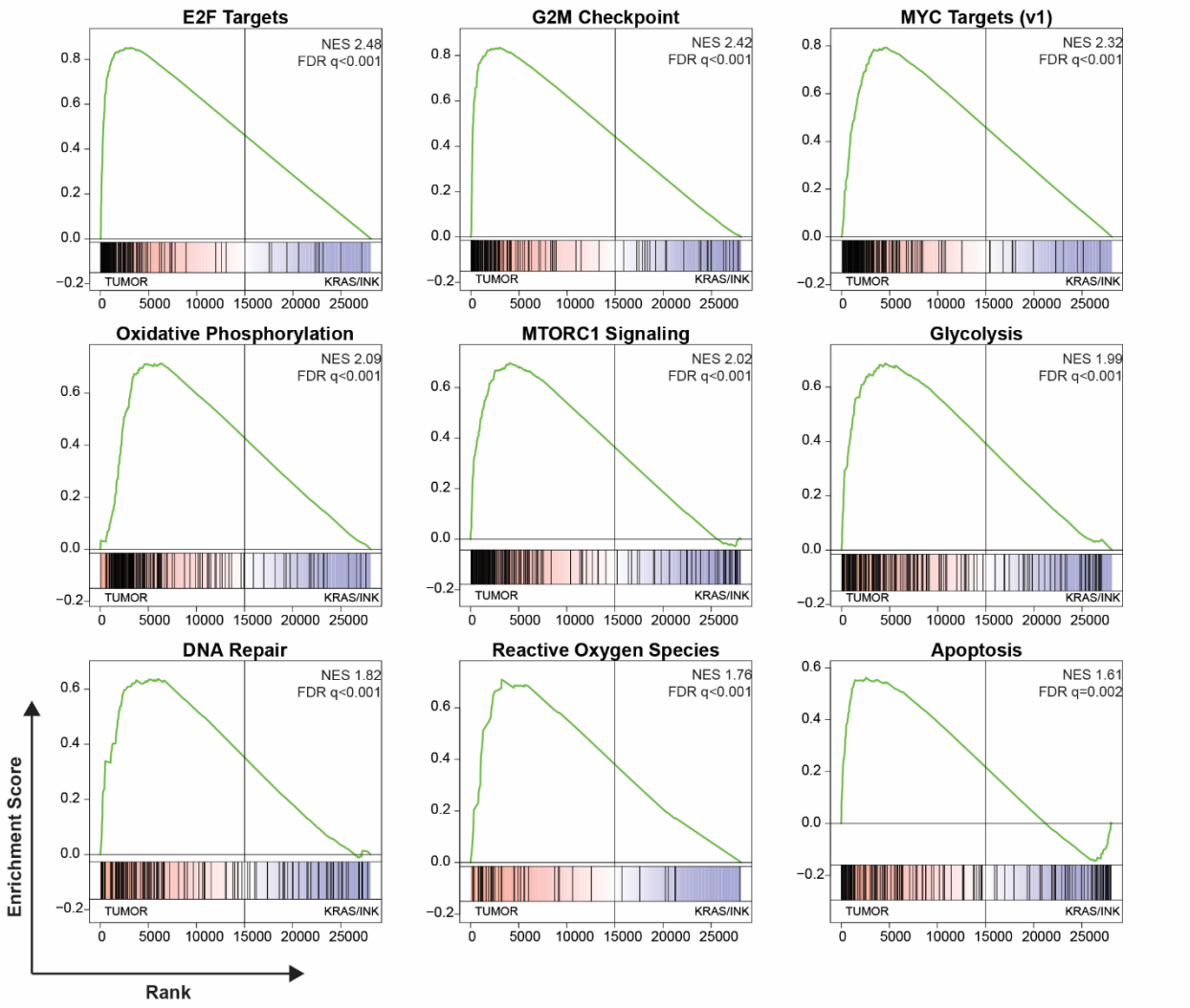
Supplemental Figure 4



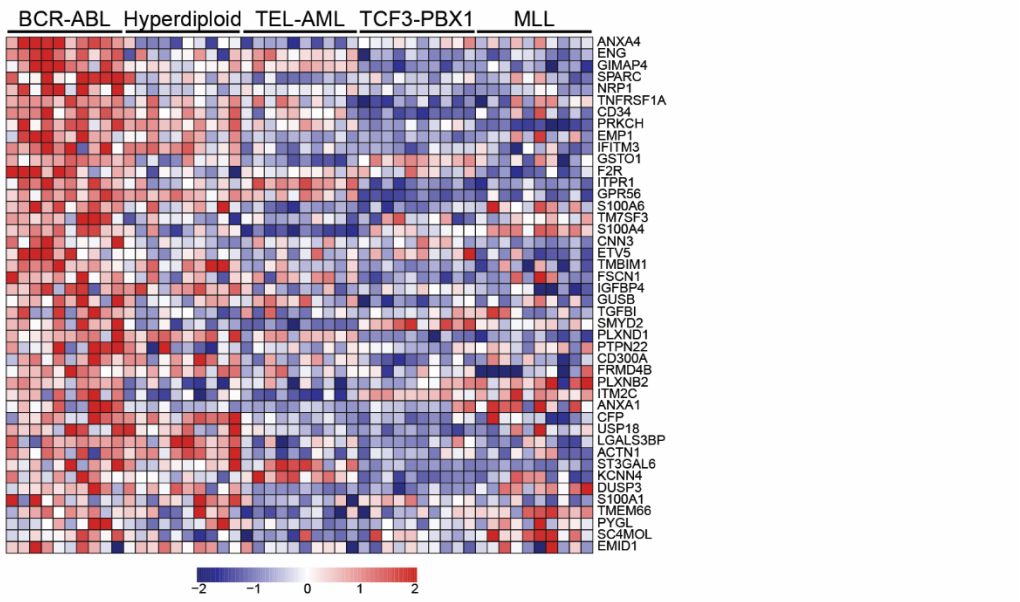
Supplemental Figure 5



Supplemental Figure 6



Supplemental Figure 7



Supplemental Figure 8