TRANSCRIPTIONAL PROFILE AND CHROMATIN ACCESSIBILITY IN ZEBRAFISH MELANOCYTES AND MELANOMA TUMORS

Eva T. Kramer*, Paula M. Godoy*, Charles K. Kaufman*

*Division of Medical Oncology, Department of Medicine and Department of Developmental Biology, Washington University in Saint Louis, St. Louis, MO 63110 USA

Supporting Material Legends

Supplementary Figure 1. Evaluation of mCherry labeling in zebrafish. (A) Representative *MiniCoopR;mitfa:mCherry* in *BRAF$^{V600E/+}$/p53$^{+/lf}$/mitfa$^{+/-}$* zebrafish (source of MC_Het melanocytes) imaged with brightfield (top) and a narrow mCherry filter (bottom). (B) Close up of melanocytes in melanocyte stripe from boxed region in A with 6x zoom/1s exposure/3.4x gain (top) and 13.5x zoom/1s exposure/6.2x gain (bottom). (C) Areas of stripe and interstripe imaged at 5x zoom with brightfield/400ms exposure/1.0x gain (left column), a narrow red filter for mCherry/3-4s exposure/1.5x gain (middle column), and a GFP longpass filter/5s exposure/1.5x gain to distinguish true mCherry expression from autofluorescence. Top two rows are two areas on the fish with scales present. Bottom two rows are two nearby areas with scales removed (within dotted area for third row, whole area for bottom row), leaving mCherry+ hypodermal melanocytes. White arrows point out visible mCherry expression. (D) Three scales (top, middle, bottom rows) imaged at 5x zoom with brightfield/400ms exposure/1.0x gain (left column), a narrow red filter for mCherry/3s exposure/1.5x gain (middle column), and a GFP longpass filter/5s exposure/1.5x gain.

Supplementary Figure 2. Isolation of skin melanocytes and melanoma cells from zebrafish. (A) FACS plot of skin cells from AB* zebrafish without mCherry (top) and with mCherry (bottom) used to set gating pattern for mCherry isolation (box). (B-C) FACS plots from unpigmented (B) and pigmented (C) melanoma tumors from *BRAF^{V600E}/p53^{lf}* zebrafish used to set gating strategy to isolate EGFP+ cells (boxed). Top plots from zebrafish lacking *crestin:EGFP*, bottom plots from zebrafish with *crestin:EGFP* expression. (D-F) Representative FACS plots showing gating for isolation of *mCh*+ (D-E) and *EGFP*+ (F) populations. (G) H&E staining of *BRAF$^{V600E/+}$/p53$^{+/lf}$/mitfa$^{+/-}$/MiniCoopR;mitfa:mCherry* zebrafish. Boxed region displayed in inset. (H) H&E staining of *BRAF^{V600E}/p53^{lf}/crestin:EGFP* zebrafish. Boxed region inset shows invasion of melanoma tumor into underlying tissue.

Supplementary Figure 3. Gene expression changes in melanocytes and melanoma cells. (A) Volcano plot with the significance ($-\log_{10}$ p-value) and the $\log_2$FC between melanoma cells and melanocytes. Of 25,221 genes plotted, 1,144 genes were significantly upregulated in melanomas ($\log_2$FC > 1, p-value < 10^{-6}). and 2,984 genes were significantly upregulated in melanocytes ($\log_2$FC < -1, p-value < 10^{-6}). Gray points are not significant (NS), yellow points are significant for $\log_2$FC ($|\log_2$FC$|$ > 1, and green points are significant for both $\log_2$FC and p-value. (B) Comparing the number of genes at different degrees of upregulation in melanoma and melanocytes. (C) Volcano plot with the significance ($-\log_{10}$ p-value) and the $\log_2$FC between WT and Het melanocytes. 12 genes were upregulated in WT melanocytes ($\log_2$FC > 1, p-value < 10^{-6}) and 41 genes were upregulated in Het melanocytes ($\log_2$FC < -1, p-value < 10^{-6}). Significance follows same color scheme as in (A). (D-H) Violin plots depicting gene expression in zebrafish melanomas (blue) and melanocytes (red) of genes associated with (D) pigmentation subtype of human melanoma (paired p-value = 0.007, adjusted p-value = 0.063), (E) proliferative subtype

human melanoma (paired p-value = 0.095799, adjusted p-value = 0.383198), (F) normal-like human melanoma (paired p-value = 0.000112 and adjusted p-value = 0.001686), (G) high-immune subtype of human melanoma (paired p-value = 9.96×10^{-6} and adjusted p-value = 0.000219), and (H) human melanocytes, paired p-value = 0.000194, adjusted p-value = 0.002355. Paired p-value represents Mann-Whitney test comparing two non-normal distributions. Subtypes as defined by Jönsson *et al*., 2010.

Supplementary Figure 4. Chromatin accessibility in melanocytes and melanoma cells based on ATAC-seq. (A) Correlation plot with the proportion of peaks between each sample passing an irreproducible discovery rate (IDR) of 0.05. (B) Profile of peaks within 3kb of transcriptional start site. (C) Number of more accessible (up, blue) and less accessible (down, red) sites within each annotated genomic region. Differential accessible regions located in promoter regions: 1,373 down, 8,096 up; intronic regions: 3,734 down, 6,725 up; exons: 2,310 down, 1,437 up; regions downstream of genes: 193 down, 460 up; distal intergenic regions: 13,212 down, 18,489 up; 5' untranslated (UTR): 70 down, 68 up; 3' UTR: 354 down, 415 up. (D) Representative epigenome browser track near *mitfa* (yellow box) on zebrafish chromosome 6. Promoter region is boxed in red.

Supplementary Table 1. Differentially expressed genes, generated from RNA-seq data. Each row displays the DESeq2 normalized read counts calculated for each sample for each gene listed. Then, differential expression for each gene is reported between conditions (MA-MC and MC_WT-MC_Het) with the $\log_2$FC, $\log_2$FC standard error (lfcSE), p-value (pvalue), and adjusted p-value (padj).

Supplementary Table 2. List of samples used in sequencing experiments. For each sequencing type, the sample name is reported with the full genotype and description, the location of the tumor for melanoma samples, the number of cells collected for the sample with FACS, and the run grouping.

Supplementary Table 3. Neural crest gene lists. First tab specifies genes based on stage when it is expressed, modified from Simões-Costa and Bronner 2015. Second tab lists all genes associated with the neural crest, obtained from ZFIN.org.

Supplementary Table 4. Tab 1 lists differentially accessible sites (FDR < 0.05) in melanoma cells versus melanocytes, generated from ATAC-seq. In each row, the chromosome and start and end of the called location is reported, along with the width of the site. The mean read concentration over all samples, and specific to the MA and MC populations is reported using $\log_2$ normalized read counts, followed by the $\log_2$FC and associated p-value and false discovery rate (FDR) calculated by DESeq2. This is followed by the peak annotation. The nearest gene is then reported by chromosome, start and end sites, strand, gene ID, transcript ID, the distance between the peak and the gene annotated, the ENSEMBL ID, gene symbol, and the gene name. Tab 2 lists all peaks.

Supplementary Table 5. Combined RNA-seq and ATAC-seq data for “long” NC list. In tab 1, for each gene in the list, the associated RNA-seq data from Supplementary Table 1 is listed. Then, associated ATAC-seq data with each peak associated with the gene is listed without

statistics. Finally, differentially accessible peaks from DiffBind (Supplementary Table 4) with statistics is listed. Tab 2 has descriptions of all the columns from tab 1.