Transcriptional profile and chromatin accessibility in zebrafish melanocytes and melanoma tumors

Eva Kramer
Paula Godoy
Charles Kaufman

Follow this and additional works at: https://digitalcommons.wustl.edu/open_access_pubs
Transcriptional profile and chromatin accessibility in zebrafish melanocytes and melanoma tumors

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

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

*Corresponding author: Division of Medical Oncology, Departments of Medicine and Developmental Biology, Washington University School of Medicine, 4518 McKinley Avenue, St Louis, MO 63110, USA. Email: ckkaufman@wustl.edu

Abstract

Transcriptional and epigenetic characterization of melanocytes and melanoma cells isolated from their in vivo context promises to unveil key differences between these developmentally related normal and cancer cell populations. We therefore engineered an enhanced Danio rerio (zebrafish) melanoma model with fluorescently labeled melanocytes to allow for isolation of normal (wild type) and premalignant (BRAF<sup>V600E</sup>-mutant) populations for comparison to fully transformed BRAF<sup>V600E</sup>-mutant, p53 loss-of-function melanoma cells. Using fluorescence-activated cell sorting to isolate these populations, we performed high-quality RNA- and ATAC-seq on sorted zebrafish melanocytes vs. melanoma cells, which we provide as a resource here. Melanocytes had consistent transcriptional and accessibility profiles, as did melanoma cells. Comparing melanocytes and melanoma, we note 4128 differentially expressed genes and 56,936 differentially accessible regions with overall gene expression profiles analogous to human melanocytes and the pigmentation melanoma subtype. Combining the RNA- and ATAC-seq data surprisingly revealed that increased chromatin accessibility did not always correspond with increased gene expression, suggesting that though there is widespread dysregulation in chromatin accessibility in melanoma, there is a potentially more refined gene expression program driving cancerous melanoma. These data serve as a resource to identify candidate regulators of the normal vs. diseased states in a genetically controlled in vivo context.

Keywords: zebrafish; melanocytes; melanoma; RNA-seq; ATAC-seq

Introduction

Human melanoma is notable as one of the most highly mutagenized cancers, making comparisons between patient samples difficult given variation in potentially irrelevant passenger somatic mutations in addition to the background of an outbred heterogeneous population (Hodis et al. 2012). Genetically engineered tumor models allow for comparisons of premalignant and cancer cells between highly related individuals (e.g., siblings, cousins) with defined driver and tumor suppressor mutations. A zebrafish model combining a BRAF<sup>V600E</sup> mutation, present in over 50% of human melanomas, with a loss of function (if) in p53 develops one to three melanomas in its lifetime (Patton et al. 2005). Despite high genetic relatedness, the Tg(BRAF<sup>V600E</sup>/p53<sup>-/-</sup>) tumors still display genomic heterogeneity but without clear functional consequences in this model (Yen et al. 2013). High heterogeneity in melanoma can lead to variable responses to currently available therapies (Reuben et al. 2017). To continue to identify novel melanoma therapies, we must delve deeper into the transcriptional and epigenetic differences that exist between the normal and diseased states.

Many studies have utilized next-generation sequencing technologies to evaluate melanoma subtypes based on clinico pathological characteristics (Cancer Genome Atlas Network 2015; Hayward et al. 2017; Tsoi et al. 2018; Rabbie et al. 2019, Cisarova et al. 2020; Durante et al. 2020), major mutations (Cancer Genome Atlas Network 2015; Travnickova et al. 2019), and drug resistance and survival (Rambow et al. 2018; Garg et al. 2021). Prior studies comparing melanocytes and cutaneous melanoma cells have reported on coding mutations, gene expression changes only, or have used human or zebrafish cell lines rather than focusing on in vivo animal models (Yen et al. 2013; Haltaufderhyde and Oancea 2014; Kaufman et al. 2016; Badal et al. 2017; Kunz et al. 2018, Venkatesan et al. 2018; Marie et al. 2020; Wouters et al. 2020). Single-cell RNA-seq experiments so far generally focus on the heterogeneity within cell populations, rather than a comparison between melanocytes and melanoma cells, or they do not evaluate the epigenetic landscape (Ennen et al. 2015; Tirosh et al. 2016; Gan et al. 2018; Rambow et al. 2018; Baron et al. 2020; Li et al. 2020; Belote et al. 2021; Travnickova and Patton 2021). Thus, we sought to characterize the epigenetic and transcriptional differences between zebrafish nonmalignant, precancerous, and malignant melanocytes in a genetically defined melanoma context.

Since its initial report in 2013, Assay for Transposase Accessible Chromatin with high-throughput sequencing (ATAC-seq) has become a widely used method to provide a sensitive assessment of genomic accessibility (Buenrostro et al. 2013). The combination of epigenetic and transcriptional approaches has been used to decipher regulators in hematopoiesis and leukemia...
(Corces et al. 2016), find loci influencing pancreatic \(\alpha\)- and \(\beta\)-cell differentiation (Ackermann et al. 2016), identify regions influencing gestational duration (Sakabe et al. 2020), characterize intermediate states in melanoma cell cultures (Wouters et al. 2020), among many other applications. Still, a comparison of the epigenetic and transcriptional profiles of sorted normal melanocytes to melanoma tumors in a native, in vivo biological context utilizing an animal model has been limited, thus diminishing our ability to dissect mechanisms governing melanoma initiation.

Animal models represent a powerful tool to understand disease initiation and progression, discover therapeutic targets, and test drugs. Utilizing the Tg(BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\) zebrafish model, we probed the natural biological context of normal melanocytes and related disease development, specifically that of melanoma cancer, using RNA- and ATAC-seq ( Patton et al. 2015; White et al. 2011; Kaufman et al. 2016). Furthermore, we compared the transcriptional profiles of zebrafish melanocytes (MC) and melanoma cells (MA) to existing human RNA-seq datasets classifying subtypes of melanoma (Jönsson et al. 2010; Harbst et al. 2012; Cancer Genome Atlas Network 2015; Cirenajwis et al. 2015;Nsengimana et al. 2015; Lauss et al. 2016; Badal et al. 2017; Gan et al. 2018; Kunz et al. 2018) to further confirm the relevance of these animal models. We present a transcriptomic and genome-wide chromatin accessibility analysis of precancerous MC and fully transformed MA in the most widely used zebrafish melanoma model to be used as a resource for identifying pathways, genes, and loci which differentiate melanoma from normal pigment cells.

**Materials and methods**

**Zebrafish husbandry**

Zebrafish were raised in the Washington University Zebrafish Consortium in accordance with animal protocols and the Washington University IACUC. Pair or harem crosses generated embryos, which were then raised at 28.5°C. The following wild type (WT), mutant, and transgenic strains were utilized: AB\(^+\), Tg(MiniCoopR, mitfa:mCherry), Tg(mitfa:BRAF\(V600E\)), p53\(\text{yt}^+/\text{lf}^+\), Tg(crestin:EGFP), Tg(mitfa:BRAF\(V600E\)), p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\), Tg(MiniCoopR, mitfa:mCherry), Tg(mitfa:BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\)/Tg(MiniCoopR, mitfa:mCherry) and Tg(mitfa:BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\).

**Transgenic fish generation**

First, we injected WT AB\(^+\) zebrafish with MiniCoopR mitfa:mCherry. Fish with germine transmission of Tg(MiniCoopR, mitfa:mCherry) (source of MC_WT) were then crossed to Tg(mitfa:BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\) zebrafish and scored for presence of mcCh+ melanocytes. These Tg(mitfa:BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\) melanocytes and Tg(mitfa:BRAF\(V600E\))/p53\(\text{yt}^+/\text{lf}^+\), mitfa\(\text{yt}^-/\text{lf}^-\)/Tg(crestin:EGFP) melanocyte cells from sorted skin samples were utilized to set the gating for isolating mcCh+ melanocytes. Briefly, we gated first to exclude debris and doublets, then gated for the desired fluorescent marker based on comparison to a nonfluorescent sample.

**Imaging**

Images were acquired using a Nikon SMZ-18 with RiD2 color camera. For whole adult zebrafish images, multiple images were taken at 0.75x magnification and then merged in Adobe Lightroom 2020. Magnification of additional images for Supplementary Figure 1 specified in figure legend.

**RNA library preparation and sequencing**

RNA was isolated from cells using the Machery Nagel Nucleospin XS kit (Fischer Scientific cat No. NCO389511). Quantity and quality of RNA were assessed using a High Sensitivity RNA ScreenTape on an Agilent 2200 TapeStation. Sample preparation for sequencing was performed by the Genome Technology Access Center (GTAC). Briefly, cDNA was generated using a Clontech SMARTer cDNA amplification kit as per manufacturer’s recommendation. 1× 50 base pair single-end sequencing was performed on an Illumina HiSeq 3000.

**RNA-seq analysis**

Demultiplexed fastq files were provided by the GTAC. Gene set enrichment analysis (GSEA) was performed using the GSEA software (Mootha et al. 2003; Subramanian et al. 2005). Quality of sequenced reads was assessed by FastQC (Anders 2010). Reads were aligned to GRCz11/danRer11 by STAR with the following parameters: --outSAMType BAM SortedByCoordinate --outFilterType BySJout --outSAMunmapped Within --outFilterIntronMotifs RemoveNoncanonical --quantMode Tran scriptomeSAM --outSAMstrandField intronMotif (Dobin et al. 2013). Transcriptome quantification was performed by RSEM using the danRer11 NCBI RefSeq gene annotations (Li and Dewey 2011). RNA-seq reads were normalized with DESeq2 (Anders and Huber 2010). Boxplots were made using ggplot2 (Wickham 2016), heatmaps with either ggplot2 or the heatmap function (Kolde 2015), and principal component analysis (PCA) was performed using the plotPCA function in the DESeq2 package (Anders and Huber 2010).
ATAC-seq library preparation and sequencing

Up to 50,000 cells per sample were tagmented with Nextera Tn5 transposase using the Illumina Nextera kit and purified with a Qiagen MinElute reaction kit per methods modified from Buenrostro et al. (2015). The DNA was then PCR amplified to add indexing primers in nine cycles. SPRI AMPure beads enriched for fragments under ~600 bp. The library was then amplified again with the nine-cycle protocol, followed by cleanup with SPRI AMPure beads. The DNA library was quantified with a Qubit DNA High Sensitivity assay and analyzed for quality and size distribution on an Agilent 2200 TapeStation with a High Sensitivity D5000 ScreenTape. Samples were pooled at a 10 nM final concentration. Sequencing was run with an Illumina HiSeq 2500 system with 2 × 50 bp read length by the Washington University in St Louis School of Medicine GTAC.

ATAC-seq analysis

Demultiplexed fastq files were provided by GTAC. Quality of fastq files was assessed by FASTQC (Andrews 2010). Adapter sequences were removed using CutAdapt (Martin 2011), and reads were aligned to GRCh31/danRer11 using Burrows–Wheeler Aligner using default parameters (Li and Durbin 2009). BAM files were sorted using samtools, and reads with Mapping Quality (MAQ) scores < 30 and improperly paired were removed using bam tools filters with the following commands: -mapQuality “>30” -isProperPair “true” (Li et al. 2009; Barnett et al. 2011). We used the MarkDuplicates command from the Genome Analysis Tool Kit to remove duplicates (McKenna et al. 2010). BAM files were converted to the BED format using bedtools bamtoBed command (Quinlan and Hall 2010) and used as input for peak calling with MACS2 (Zhang et al. 2008). We used the callpeak command with the following parameters: -g 1e9, -nomodel, -extsize 100, -shift 50, -keep-dup 999, -call-summits. Output bedGraph files were sorted, clipped, and converted to BigWig format using the bedGraphToBigWig command (Kent et al. 2010). Quality of peaks was assessed by ChiPQC (Carroll et al. 2014). The irreproducible discovery rate (IDR) was performed between replicates using the idr command (Li et al. 2011). Differentially accessible regions were called using DiffBind (Stark and Brown 2011). All motif analyses were performed using Homer (Heinz et al. 2010).

Histology

Three Tg(mita:BRAFV600E)/p53⁻/⁻/mitfa⁺/mitfa:mCherry zebrafish and three Tg(mita:BRAFV600E), p53⁻/⁻/crestin:EGFP) zebrafish were euthanized and each whole fish was placed in a 50 ml conical with 40 ml 10% formalin for fixation and shipped to HistoWiz for further preparation. Briefly, zebrafish were decalcified and embedded, then, five sagittal sections along the midline were stained with hematoxylin and eosin.

Results

Isolating melanocyte and melanoma populations from adult zebrafish skin

To compare the transcriptional and chromatin accessibility states of normal (WT) and precancerous melanocytes (MC; harboring the BRAFV600E mutation and p53 If mutation) to fully transformed melanoma (MA) cells, we generated a stable transgenic zebrafish line with fluorescently labeled melanocytes expressing mCherry driven by the melanocyte-specific melanocyte inducing transcription factor a (mitfa) promoter (Ceol et al. 2011; White et al. 2011; Kaufman et al. 2016). A stable MiniCoopr; mitfa:mCherry transgenic line was generated in WT AB⁺ zebrafish using Tol2-mediated insertion (Figure 1, A and B) and then crossed to Tg(BRAFV600E)/p53⁻/⁻/mitfa⁺/⁻ zebrafish (Ceol et al. 2011) to generate Tg(BRAFV600E⁺)/p53⁻/⁻/mitfa⁺/⁻/Tg(MiniCoopr; mitfa:mCherry) zebrafish to investigate the impact of overactivation of BRAF in melanocytes (Figure 1, A and C and Supplementary Figure S1). Zebrafish heterozygous at the key BRAF and p53 melanoma driver loci had thicker and darker melanocyte stripes, with minor melanocyte expansion into the interstripe as compared with WT AB⁺ fish (Figure 1, B and C and Supplementary Figure S1), consistent with the melanocyte pattern seen with homozygous BRAFV600E shown previously (Patton et al. 2005). The presence of the mitfa minigene in the MiniCoopR backbone containing the mitfa:mCherry transgene did not appear to alter the normal strip- ping pattern (whereas the presence of the BRAF and p53 alterations did, comparing Figure 1, B and C), and we detected upregulation of mitfa in the MA vs. both MC populations (Supplementary Table S1), consistent with prior expression studies (Kaufman et al. 2016).

Using FACS, we carefully isolated mCherry-positive (mCh⁺) melanocytes from Tg(MiniCoopr; mitfa:mCherry) (MC, WT) and Tg(BRAFV600E⁺)/p53⁻/⁻/mitfa⁺/⁻/Tg(MiniCoopr; mitfa:mCherry) (MC, Het) zebrafish skin (Supplementary Figure S2, A, D, and E). We could visualize mCh⁺ melanocytes in both isolated scales and in the underly ing hypodermis when overlying scales were removed (Supplementary Figure S2, C and D), and as we stripped the skin in its entirety after euthanasia, our melanocyte population likely includes both scale associated and hypodermal sources. mCh⁺ melanocyte cells were similarly ~1–5% abundant during FACS from the skin of AB⁺ or BRAFV600E⁺/p53⁻/⁻/mitfa⁺/⁻ zebrafish (Supplementary Figure S2). We also performed FACS on previously described Enhanced Green Fluorescent Protein (EGFP)-positive (GFP⁺) melanoma tumors from Tg(BRAFV600E)/p53⁻/⁻/crestin:EGFP (MA) zebrafish to isolate melanoma cells (Figure 1D and Supplementary Figure S2, B, C, F, and H; Kaufman et al. 2016). As previously described, these melanoma tumors are identifiable as raised, highly cellular masses that are locally invasive (Supplementary Figure S2, G and H) and readily differentiated from normal skin regions, both grossly (Figure 1, B–D) and on histological sections (Supplementary Figure S2, G and H). Zebrafish crestin is expressed in embryonic neural crest (NC) and is re-expressed with previously reported massive upregulation of crestin in melanoma (Kaufman et al. 2016). We then performed RNA- and ATAC-seq with each cell population (see Materials and methods; Figure 1A and Supplementary Table S2).

As expected, GFP⁺ melanoma tumor cells and mCh⁺ melanocyte samples clustered based on crestin, EGFP, and mCherry normalized read counts by Euclidean distance. GFP⁺ samples show high crestin expression supporting the fidelity of the crestin:EGFP transgene as a faithful reporter of endogenous crestin expression (Figure 1E). Although melanocytes had some reads aligning to crestin, crestin expression was 109 times higher in melanoma (39,232 average reads) than in melanocytes (359 average reads), consistent with previously reported massive upregulation of crestin in melanoma tumors (White et al. 2011; Kaufman et al. 2016; Figure 1E).

ATAC-seq and RNA-seq libraries from sorted cells passed established Encode QC metrics for library quality and read depth. Samples for RNA-seq were sequenced to an average of 29,103,353 reads, of which 9,897,358 (34%) aligned to the transcriptome (Figure 1F). Of all genes with at least 1 count per million (CPM) reads, of which 9,897,358 (34%) aligned to the transcriptome (Figure 1F). Of all genes with at least 1 count per million (CPM) reads, of which 9,897,358 (34%) aligned to the transcriptome (Figure 1F). Of all genes with at least 1 count per million (CPM) reads, of which 9,897,358 (34%) aligned to the transcriptome (Figure 1F). Of all genes with at least 1 count per million (CPM) reads, of which 9,897,358 (34%) aligned to the transcriptome (Figure 1F).
Transcriptomic comparison of normal melanocytes, BRAF/p53 mutant melanocytes, and melanoma cells

Transcriptional profiles of GFP+ (melanoma) were highly correlated (Pearson $R^{21} = 0.85$; Figure 2A), even from different fish. mCh+ (melanocytes) samples also had highly correlated transcriptional profiles (Figure 2A). Melanocytes from WT fish (MC_WT) and fish heterozygous for the BRAF$^{V600E}$-driver oncogene and p53 allele (MC_Het) were also highly correlated with one another with small differences in correlation coefficients between origin genotype, which is notable given the presence of the activated BRAF$^{V600E}$ oncogene in the MC_Het melanocytes ($R = 0.88–0.91$ within WT fish, $R = 0.85–0.87$ within heterozygotes, and $R = 0.85–0.87$ across genotypes; Figure 2A). Melanoma samples were similarly highly correlated ($R = 0.87–0.93$) with slightly higher correlation for multiple tumors isolated from the same fish ($R = 0.93$ for MA3A, MA3B, MA3C, and $R = 0.87–0.92$ for external replicates). Melanocytes and melanoma cells readily separated using unsupervised hierarchical clustering (correlation coefficients between 0.74 and 0.81) indicating that transcriptional profiles alter during transformation from a BRAF mutant melanocyte to a fully oncogenic melanoma cell, but remain more closely correlated in a WT (MC_WT) or premalignant (MC_Het) state.

Furthermore, PCA readily separates melanocytes from melanomas, with the first principal component accounting for 38% of the variance (Figure 2B). Principal component two (accounting for 11% of the variance) likely captures the lesser transcriptional differences between the WT and Het MC samples.

We identified 1144 genes significantly upregulated in melanomas vs. melanocytes, and 2984 genes significantly upregulated in melanocytes compared with melanoma cells (log$_2$ fold change (log$_2$FC) $> 1$ and adjusted P-value $< 10^{-6}$, Supplementary Figure S3A). Surprisingly, the number of melanoma-upregulated genes meeting increasing log fold change cutoffs dropped off more...
steeply than melanocyte-upregulated genes, indicating that more genes are downregulated in melanoma as compared with melanocytes (Supplementary Figure S3B).

Comparing WT and Het (Tg[BRAFV600E+/p53+/−]) melanocytes, there are far fewer significantly differentially expressed genes with 12 genes upregulated in WT over Het melanocytes and 41 genes upregulated in Het over WT melanocytes with log₂FC > 1 and adjusted P-value ≤ 10⁻⁶ (Supplementary Figure S3C). These results highlight the surprising similarity within melanocytes and premalignant (i.e., BRAF mutant/p53 mutant) melanocytes as compared with the more dramatic changes that occur when compared with transformed melanocytes/melanoma. To facilitate further inquiry, we provide a list of all gene expression changes and P-values between melanocytes, premalignant melanocytes, and melanoma cells (Supplementary Table S1).

### Gene expression programs

We compared gene expression profiles between our melanocyte and melanoma samples to the four gene lists from Jonsson et al. (2010) stratifying human cutaneous melanoma into proliferative, pigmentation, high immune, and normal-like subtypes. First, we assessed which subtype of human melanoma our zebrafish melanoma samples most resembled and found the highest average expression of genes associated with the pigmentation subtype (Figure 2C). This was further confirmed with GSEA detecting significant enrichment of genes associated with the human melanoma pigmentation subtype (GSEA ES = 0.4299, P-value = 0.036, q-value = 0.134; Figure 2D). Finally, comparing expression in zebrafish melanocytes to melanoma cells, genes in the pigmentation list had the most significant upregulation (paired P-value = 0.007; Supplementary Figure S3, D–H).

Previous studies have shown that melanoma reactivates aspects of an embryonic NC program, with prominent NC genes such as sox10 the subject of much interest (Shakhova et al. 2012; Cronin et al. 2013; Mohamed et al. 2013; Kaufman et al. 2016). The central gene regulatory network, or “wiring diagram,” for NC development has been well-described, and we used the NC gene regulatory network described by Kunz et al. (2018) to examine gene expression changes in each condition for NC pathway genes (Supplementary Table S3). We found sox10 and dix2a are upregulated in melanoma in agreement with previous smaller scale transcriptional analyses (black boxes, Figure 2E; Kaufman et al. 2016). Interestingly, important melanocyte-specific genes including mitfa, tyrp1b, kita, dct, tyr, and tyrp1a were upregulated in the melanoma samples, supporting that melanoma cells sustain a
continued presence of an active melanocyte identity (red boxes, Figure 2E). To further probe this finding, we assessed whether either cell population had enrichment of human melanocyte genes (Reemann et al. 2014). Indeed, both zebrafish melanocytes and melanoma cells had high expression of human melanocyte genes (Supplementary Figure S3H).

Another recurring and important question relates to the idea that the NC incorporates a spectrum of developmental time (e.g., from specification, to delamination, to migration, etc.), spatial differences (e.g., cranial, thoracic, caudal NC), and descendant lineages (e.g., glia, melanocyte, chondrocyte), each with defining genetic markers in its transcriptional program (Supplementary Table S3; Simões-Costa and Bronner 2015; Williams et al. 2019).

We reviewed our melanoma differential gene sets in relation to these NC subsets and, aside from a “strengthening” of the melanocyte program as noted above and decrease in expression of glial (mbpa/b) and chondrocyte (sox9b) factors, we did not note other clear unifying NC subsets that aligned more closely to the melanoma program (Figure 2E). Given the widespread dysregulation of gene expression in the cancerous melanoma state, this is perhaps unsurprising that the malignant phenotype would entail a broad and not entirely faithful amalgamation of melanocyte and NC transcriptional identities.

Chromatin accessibility profiles in melanocytes and melanoma cells

To characterize the genome-wide chromatin accessibility of premalignant Tg(BRAFV600E)/p53lf/lf mutant melanocytes and melanoma, we isolated premalignant melanocytes (mCh+) from four zebrafish, including three MC_Het and one sample from a Tg(BRAFV600E)/p53lf/lf/Tg(MiniCoopR; mitfa:mCherry) zebrafish (MC_Homo). Since MC_WT and MC_Het read counts were highly correlated in the RNA-seq, we utilized MC_Het zebrafish for ATAC-seq analysis. We compared these profiles to those from eight melanoma tumors sorted for crestin:EGFP expression (Figure 3A).

As in the RNA-seq analysis, melanocyte samples and melanoma samples clustered according to normal vs. malignant state (Figure 3, A and B). To assess the reproducibility of each identified accessible site within a condition, we evaluated the IDR, setting a threshold of 0.05. For MC samples, an average of 53% of sites passed IDR 0.05, and for the MA samples, the average was 33% (Supplementary Figure S4A).

In our ATAC-Seq analysis, 56,936 sites demonstrated differential accessibility (P < 0.05, IDR < 0.05) between melanocyte and melanoma cells. Open regions, or peaks, within 3 kilobases (kb) of transcriptional start sites (TSS) were centered at the TSS.

Figure 3 ATAC-seq analysis of melanocyte and melanoma cells. (A) Correlation plot of overall chromatin accessibility across the genome, with darker green color indicating greater correlation between samples. (B) PCA plot comparing similarity of chromatin accessibility across conditions. (C) Predicted TF binding sites, based on HOMER analysis, enriched in open chromatin domains more accessible in melanoma vs. melanoma cells. Each TF is reported with its associated TF family, as TF families frequently have similar or identical binding sites which are not differentiated by HOMER. (D) Epigenome browser tracks near and upstream of the NC and melanoma gene sox10 on chromosome 3. MC, melanocyte sample; Het, heterozygous for BRAFV600E and p53 transgenes; Homo, homozygous for BRAFV600E and p53 transgenes; MA, melanoma sample.
(Supplementary Figure S4B). Of promoter annotated peaks, 85% of differentially accessible sites were more open in melanoma as compared with melanocytes (Supplementary Figure S4C and Table S4).

**Open chromatin regions and putative transcription factor binding sites**

We examined the differentially accessible chromatin regions using HOMER de novo analysis to identify over-represented DNA motifs representing putative transcription factor (TF) binding sites (Figure 3D). In order of significance, sites associated with CCCTC-binding factor (CTCF) topped the list—a TF commonly mutated in cancer and known to normally function widely in controlling chromatin architecture (Kim et al. 2015). TF binding motif families corresponding to key NC-associated factors also rounded out the top 10 list, including the Forkhead, SOXE, and ETS families. Melanocyte and NC programs are particularly active in melanoma cells, with increased *sox10* and *mitfa* expression, respectively, seen in the RNA-seq data (Figure 2E), and this appears consistent with enrichment of SOX9/10 (SOXE family) and MITF (E/M-boxes) binding sites in more accessible regions in melanoma (Figure 3C). Finally, we noted the overall similarity in the contour of chromatin accessibility within premalignant melanocytes and melanomas as well as the numerous areas of differential accessibility as, for example, near *sox10* and *mitfa* (Figure 3D and Supplementary Figure S4D).

**Modeling gene expression as a function of peak accessibility**

To examine potential relationships between open chromatin and gene expression, we assessed the relationship between log_{2}FC for differentially expressed genes and differentially accessible promoter peaks. As expected, the majority of genes with greater expression in melanoma had more accessibility with peaks within 3 kb of the promoter (n = 2386 pairs of genes and peaks; one gene can be linked to multiple ATAC-seq peaks), while genes with less expression were less accessible at the promoter (n = 499). Interestingly, 1213 genes had greater accessibility at the promoter in melanoma, yet had lower expression. We then focused on evaluating the relationship between accessibility and

![Figure 4](https://academic.oup.com/g3journal/article/12/1/4kab379/6428538)
expression in NC genes using a "long list" of 329 genes associated with NC from ZFIN.org and a curated shortlist of 44 genes associated with specific NC developmental stages (Supplementary Tables S3 and S5). Utilizing the long list, we found a modest correlation (Pearson’s correlation coefficient $R = 0.3951$) between gene expression accessibility at promoter-proximal peaks within 3 kb of the promoter (Figure 4B). This relationship was strengthened when focusing only on peaks at the TSS (Figure 4C; $R = 0.6067$). Further, focusing specifically on genes previously shown to be involved in NC development, this association was even more pronounced (Figure 4D; $R = 0.6686$), with peaks at the promoter showing a clear correlation with expression (Figure 4E; $R = 0.8488$).

**Discussion**

Here, we present a zebrafish model allowing for efficient isolation of melanocytes and melanoma cells and provide a resource capturing transcriptional and chromatin accessibility changes occurring in vivo during de novo tumor development. Independent replicates from genetically related animals add robustness to the dataset while limiting the background genetic complexity of a fully outbred population. This combined RNA- and ATAC-seq analysis of normal, premalignant, and transformed melanocytes/melanoma offers a tool for gene and pathway discovery in melanoma biology. Moreover, the tumor diversity with pigmentation status and tumor location (Supplementary Table S2) provides a comprehensive picture of cutaneous melanoma in a BRAF-driven, MITF-high model (Travnickova et al. 2019). Based on widely used QC metrics, we conclude that our RNA- and ATAC-seq data accurately reflect the in vivo transcriptional and chromatin accessibility state of the average premalignant BRAF/p53 mutant melanocyte and melanoma cell and provide an important and broadly useful tool for further investigating transcriptional and epigenetic programs underlying these related but crucially different pre- and fully malignant states. Similar studies have shed light on the epigenetic regulation of dysregulated genes that promote melanoma, such as sox10 (Cunningham et al. 2021), or have been used to delineate developmental programs underlying erythroid differentiation (Ludwig et al. 2019). Studies have found a reactivation of an NC program in melanoma (Kaufman et al. 2016), and indeed we observed upregulation of NC genes such as crestin, sox10, tfap2a, dx2a, among others. However, when we expanded the analysis to include NC genes segregated by stage of expression in NC, there was not a clear program upregulated in melanoma (i.e., not all migratory NC markers were upregulated). This further supports a widespread dysregulation of expression programs in the disease state and begs the question of which consistent gene program is reactivated in melanoma. Nevertheless, there is an apparent relationship between accessibility at the TSS and gene expression of NC-associated genes (Figure 4E).

Interestingly, our dataset shows that in melanoma there are more genes with significant downregulation in melanoma relative to melanocytes, despite the presence of more accessible regions in the melanoma genome, supporting previous findings that points of control other than promoter accessibility, such as TF and chromatin regulatory protein abundances and activities, may be primary influencers of gene expression and thus changes in cell identity in the transition to a malignant state (Travnickova et al. 2019; Santoriello et al. 2020; Baggioni et al. 2021; Fazio et al. 2021; Terranova et al. 2021). Indeed, it has been shown that very distal enhancers can control gene expression, indicating that open regions of chromatin may not necessarily regulate the most proximal gene (Lettice et al. 2003; Amano et al. 2009; Lacomme et al. 2018). Most changes in accessibility occur in distal nonpromoter regions (Supplementary Figure S4C; Thurman et al. 2012; Pliner et al. 2018; Friman et al. 2019), and deciphering the relationship between gene expression and cis-regulatory regions has been a topic of immense focus (Ackermann et al. 2016; Gonen et al. 2018; Zhao et al. 2019; Cai et al. 2020; Cunningham et al. 2021; Panigrahi and O’Malley 2021). Our study serves as a foundation to probe open chromatin regions and potential regulatory functions.

Furthermore, though all samples had an active melanocyte-specific program, genes typically associated with melanocytes were upregulated and more accessible in melanoma, consistent with a widespread dysregulation in additional gene programs in the melanoma cell state. We anticipate these data will contribute to ongoing functional analyses testing candidates that control the crucial epigenetic and transcriptional differences driving the transition between the normal and diseased cell state.

**Data availability**

Strains are available upon request. BigWig files can be found on the Gene Expression Omnibus, accession no. GSE178803. Supplementary files submitted through figshare (https://doi.org/10.25387/g3.16841149).

**Acknowledgments**

We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant No. P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant No. UL1 TR000448 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH. We also thank the Fluorescence-Activated Cell Sorting core for their assistance with sorting our samples.

**Funding**

Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under award number R01CA240633. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health. C.K.K. was funded by the Cancer Research Foundation Young Investigator Award and R01CA240633. E.T.K. was supported by the National Institutes of Health Kirschstein National Research Service Award Institutional Research Training Grant T32# GM007067 during data collection. P.M.G. was supported by the National Science Foundation Graduate Research Fellowship (DEG-1745038).

**Conflict of interest**

The authors declare no conflict of interest.
Literature cited


Communicating editor: B. Andrews