

1 **TITLE: Functional *in vivo* characterization of zebrafish *sox10***
2 **enhancers in melanoma and neural crest**

3

4 **RUNNING TITLE: Melanoma and neural crest *sox10* enhancers**

5

6 **Authors:** Rebecca L. Cunningham¹, Eva T. Kramer¹, Sophia K. DeGeorgia¹, Shayana
7 Seneviratne², Vadim Grigura¹, and Charles K. Kaufman¹

8

9 ¹Division of Medical Oncology, Department of Medicine and Department of Developmental
10 Biology, Washington University in Saint Louis, 4518 McKinley Ave., St. Louis, MO 63110, USA

11 ²School of Arts and Sciences, Washington University in Saint Louis, 1 Brookings Drive, St. Louis,
12 MO, 63130

13

14 Corresponding author e-mail address: ckkaufman@wustl.edu

15 Keywords: melanoma, cancer initiation, epigenetics, cancer models, ATAC-Seq, *sox10*

16

1 **Abstract**

2 The re-emergence of a neural crest transcriptional program, including *Sox10* upregulation,
3 is a key step in melanoma initiation in humans and zebrafish. We hypothesize that epigenetic
4 regulation of *sox10* modulates melanoma initiation. ATAC-Seq analysis of zebrafish melanoma
5 tumors identifies recurrently open chromatin domains near *sox10*. Reporter constructs for each
6 putative *sox10* enhancer were examined in zebrafish embryos for neural crest activity and in stable
7 transgenic lines for melanoma activity. One element, *peak5* (23 kilobases upstream of *sox10*),
8 drives *EGFP* reporter expression in a subset of neural crest cells, Kolmer-Agduhr neurons, and
9 early melanoma patches and tumors with high specificity. A ~200 bp region, conserved across the
10 *Cyprinidae* family (fish), is required for *peak5* activity in neural crest and melanoma, and contains
11 dimeric SoxE binding sites essential for neural crest activity. Our work identifies a novel
12 melanoma transcriptional enhancer, expanding our knowledge of epigenetic regulation of neural
13 crest identity in melanoma.

14

1 **Introduction**

2 Melanoma is a potentially deadly cancer of transformed melanocytes, which are pigment-
3 producing cells derived from embryonic neural crest. Early detection and treatment of melanoma
4 is critical because of high mortality associated with metastatic melanoma. The difficulty of treating
5 metastatic melanoma underscores the necessity of defining how this cancer initiates and the
6 urgency of developing new therapeutics (Lo and Fisher, 2014). Encouragingly, chromatin
7 modifiers and other proteins that shape the epigenetic landscape offer promising novel therapeutic
8 targets that may expand the treatment repertoire beyond immunotherapy and MAPK-targeting
9 strategies (Tanaka et al., 2015; Badner et al., 2017), and combat the ability of melanomas to evade
10 treatment due to the high mutational burden, heterogeneous mutations, and rapid evolution of these
11 tumors (Gallagher et al., 2014; Fontanals-Cirera et al., 2017).

12 To effectively treat melanoma in its earliest stages requires a thorough understanding of
13 mechanisms that drive melanoma initiation. The transgenic *BRAF^{V600E}/p53^{-/-}* zebrafish has been
14 established as a powerful melanoma model that is highly reflective of the human disease and can
15 be utilized to study early stages of melanoma initiation (Patton et al., 2005). In this model, the
16 human *BRAF^{V600E}* oncogene is expressed under the control of the melanocyte-specific promoter of
17 the *mitfa* gene. Within *BRAF^{V600E}/p53^{-/-}* zebrafish, all melanocytes thus have the potential to
18 become melanoma, but only one to three melanomas typically develop from the many thousands
19 of melanocytes in any single fish, presenting an excellent model in which to study how cancer
20 initiates from a field of cancer-prone cells.

21 Recent work in zebrafish has demonstrated that reactivation of a neural crest program
22 (NCP) is an early step in melanoma initiation (Kaufman et al., 2016). A single melanoma cell can
23 be visualized in live zebrafish by expression of the *crestin:EGFP* transgene (Kaufman et al., 2016).

1 Not only does *crestin:EGFP* expression uniquely enable visualization of cancer before a large
2 tumor is visible, expression of *crestin* also indicates a re-emergence of aspects of neural crest
3 transcriptional identity in these cells because *crestin* is exclusively expressed in the embryonic
4 neural crest of zebrafish (Luo et al., 2001). Neural crest cells (NCCs) are a transient and migratory
5 embryonic cell population that give rise to a variety of cell types, including melanocytes. Prior
6 studies have noted that NCCs and cancer cells share many cell biological characteristics, including
7 the ability to proliferate, change cell adhesion properties, and migrate (Maguire et al., 2015). For
8 example, epithelial-to-mesenchymal transition occurs to permit both NCC migration and cancer
9 metastasis (Kerosuo and Bronner-Fraser, 2012) Furthermore, melanoma cells can behave as NCCs
10 after transplantation into chick embryos, demonstrating their plasticity and connection to their
11 developmental origins (Kulesa et al., 2006).

12 While *crestin* is expressed in the neural crest, this gene is zebrafish-specific and
13 functionally uncharacterized; therefore, *crestin* itself does not represent a future therapeutic target
14 in human melanoma. However, *crestin* is one of a suite of neural crest genes that are upregulated
15 in melanoma in zebrafish and humans (Kaufman et al., 2016). The re-emergence of an NCP
16 transcriptional program in melanoma is supported by the expression of other neural crest related
17 genes in tumors such as *sox10* and *dlx2a*. Therefore, applying knowledge of developmental
18 mechanisms in NCCs may enhance our understanding of melanoma initiation and treatment of
19 NCC derived cancers.

20 One highly upregulated neural crest gene in melanoma in both zebrafish and humans is
21 *sox10*, a transcription factor necessary for neural crest development (Kaufman et al., 2016;
22 Rönnstrand and Phung, 2013; Kelsh, 2006). Furthermore, in a zebrafish *BRAF^{V600E}* melanoma
23 model, a *Nras* mouse melanoma model, and in human melanoma cell culture, abrogation of *sox10*

1 expression affects melanoma onset and maintenance, respectively (Kaufman *et al.*, 2016;
2 Shakhova *et al.*, 2012). Likewise, overexpression of *sox10* in melanocytes increases the rate of
3 melanoma onset in zebrafish (Kaufman *et al.*, 2016). Further demonstrating an NCP activation in
4 melanoma, regulatory elements near neural crest genes, such as *sox10*, are activated in melanoma
5 cells lines and NCCs, but not generally in other adult tissues or selected cancer cells lines
6 (Kaufman *et al.*, 2016). Importantly, this epigenetic phenomenon is conserved in human
7 melanomas (Kaufman *et al.*, 2016).

8 Since *sox10* is critical to establish neural crest identity, the development of melanocytes,
9 and the maintenance and progression of melanoma (Kaufman *et al.*, 2016; Shakhova *et al.*, 2012;
10 Cronin *et al.*, 2013), epigenetic regulation of *sox10* may be critical for re-establishing neural crest
11 identity in melanoma initiation. Therefore, identifying and functionally characterizing enhancer
12 elements that control *sox10* expression during this oncogenic transition will broaden our
13 understanding of epigenetic changes that occur to trigger melanoma initiation and activation of an
14 NCP-like identity from within a cancerized field.

15 In this study, we functionally screen putative enhancers elements, identified as accessible
16 or "open" regions of chromatin using ATAC-Seq on zebrafish melanomas, near the *sox10* gene for
17 transcriptional activating activity in embryonic neural crest as well as in melanoma tumors. This
18 screening approach allowed us to identify and focus initially on one 669 bp enhancer, termed
19 *peak5*, that is highly and specifically active in adult zebrafish melanoma. We further show that
20 *peak5* is an active embryonic enhancer in a subset of the NCCs and a select population of Kolmer-
21 Agduhr neurons in the central nervous system. We also identify a ~200 bp nucleotide sequence
22 centered within *peak5* that is conserved between members of the fish *Cyprinidae* family and is
23 necessary and sufficient for neural crest activity. Furthermore, mutational analysis of a dimeric

1 SoxE binding site in the conserved core of *peak5* uncovers transcription factor binding sites that
2 are necessary for robust *peak5* neural crest activity. Together, these data enhance our
3 understanding of epigenetic changes that drive the re-emergence of neural crest identity in
4 melanoma and suggest a model wherein auto-regulation of *sox10* in a feed-forward loop may
5 contribute to the formation of melanoma.

6

7 **Results**

8 *Identification of putative sox10 enhancers*

9 To identify putative regulatory elements surrounding the *sox10* locus, ATAC-Seq was
10 performed on bulk melanoma tumors isolated from *Tg(BRAF^{V600E};crestin:EGFP);p53^{-/-}* zebrafish
11 (**Figure 1A**). ATAC-Seq on melanoma tumor cells showed consistent regions of open chromatin
12 and identified 11 regions of open chromatin around and within the *sox10* locus, including a putative
13 *sox10* minimal promoter encompassing exon 1 of *sox10* (**Figure 1B; Figure S1**). These open
14 regions of chromatin were also consistent with ATAC-Seq peaks in zebrafish melanoma cell lines
15 (Kaufman et al., 2016) (**Figure S1**). To assess if these ATAC-Seq peaks are functional enhancers
16 in neural crest and melanoma cells, we performed *in vivo* reporter assays in *Tg(BRAF^{V600E}); p53^{-/-}*
17 zebrafish embryos (**Figure 1C**). Putative non-intronic enhancers that were present in at least 4 out
18 of 5 tumor ATAC-Seq samples were cloned upstream of the mouse *beta-globin* basal promoter
19 driving *EGFP* (Tamplin et al., 2011), within a Tol2 vector (394 from the Tol2 Kit) to enable mosaic
20 integration into the zebrafish genome.

21 Each putative enhancer construct was injected into single-cell *Tg(BRAF^{V600E}); p53^{-/-}*
22 embryos. A *crestin:mCh* construct was also co-injected with each putative enhancer construct to
23 serve as a positive injection control and to co-label NCCs. Embryos were screened at 1 dpf for co-

1 labeling of EGFP and mCh, indicating that the enhancer functions in the neural crest. Larva were
2 ranked as either “no expression,” “weak expression,” or “strong expression” for both *EGFP* and
3 *mCh*. “Strong expression” was characterized by 5 or more EGFP positive cells where as “weak
4 expression” was characterized by fewer than 5 EGFP positive cells. Nine out of eleven putative
5 enhancers drove *EGFP* expression in at least a subset of NCCs, as determined by cell shape,
6 location, and co-labeling with *crestin:mCh* (**Figure S2**). One peak of interest, *peak5*, cloned as a
7 669 bp element (**Figure S3**), exhibited particularly robust activity in NCCs (**Figure 2A-B**). The
8 putative *sox10* minimal promoter, cloned as a 351 bp element and potential positive control,
9 exhibited the most robust *EGFP* expression throughout the nervous system and in NCCs (**Figure**
10 **2C**). A region of closed chromatin approximately 76 kb upstream of *sox10*, as determined by
11 ATAC-Seq, was also cloned upstream of *betaglobin:EGFP* as a negative control and did not drive
12 neural crest *EGFP* expression (**Figure 2D**). In subsequent analyses, we noted that this control
13 region also exhibits sequence similarity to another region on Chromosome 3 (Chr3: 45240516-
14 45241216). An additional negative control (Negative Control A) cloned from a region 10.7 KB
15 upstream of the *sox10* also exhibited minimal reporter activity (**Figure S2**). Collectively, ATAC-
16 Seq analysis of zebrafish melanoma tumors reliably identifies transcriptional regulatory regions
17 near *sox10* that function as enhancers with neural crest cell activity in transgenic reporter studies.
18
19 *peak5 is active in embryonic neural crest and Kolmer-Agduhr neurons*

20 Enhancer reporter constructs injected into zebrafish embryos are randomly integrated
21 throughout the genome, and thus subject to position effects (Udvadia and Linney, 2003). In
22 addition to examining numerous F0 transgenic embryos for consistent spatial and temporal
23 expression as above, we further characterized the expression pattern of activity of *peak5*, a region
24 showing particularly robust neural crest activity in zebrafish embryos, by generating 6 independent

1 stable transgenic lines. To generate stable lines, F0 injected embryos were raised to maturity and
2 then outcrossed to *Tg(BRAF^{V600E}); p53^{-/-}* zebrafish. Subsequent F1 embryos were then screened
3 for *EGFP* expression at 1 dpf. All 6 identified stable transgenic lines exhibited commonalities in
4 expression patterns, detailed below (**Figure 3**). One stable line, *peak5_115A* exhibited *EGFP*
5 signal in muscle. Given that *sox10* is not endogenously expressed in muscle, we reasoned that this
6 muscle activity was due to position effects, such as enhancer trapping, and as a result did not use
7 this line for subsequent analyses.

8 At 1 dpf, 4 out of 6 stable lines exhibited *EGFP* expression in pre-migratory NCCs dorsally
9 located on the posterior trunk (**Figure 3A, C, D, E**). To confirm that *peak5* is active in NCCs and
10 cells that would typically express *sox10* we crossed *Tg(peak5:betaglobin:EGFP)* to either a
11 *Tg(crestin:mCh)* line (**Figure 4A'**) or the previously published *Tg(sox10(7.2):mRFP)* line
12 (Kucenas et al., 2008) (**Figure 4B'**). Epifluorescence imaging of double transgenic embryos at 1
13 dpf showed most *EGFP* positive cells in the dorsal posterior region co-localize with mCh or mRFP
14 at 1 dpf (**Figure 4A-B'**), illustrating that *peak5* is active in NCCs. In addition to a subset of NCCs,
15 6 out 6 *peak5* transgenic lines express *EGFP* within a subset of ventral Kolmer-Agduhr (KA)
16 neurons in the spinal cord, which contact cerebrospinal-fluid (**Figure 3**), as identified by cell shape
17 and localization (**Figure S4**). By 5 dpf, *EGFP* expression was consistently observed in KA neurons
18 across all lines, and in the heart in 3 out 6 stable lines (**Figure 3**).

19 *EGFP* expression in adult fish was most prominent by epifluorescence in cells within
20 the tail, tip of the dorsal fin (**Figure 5A-C**), and peripheral nervous system, particularly observable
21 within the maxillary barbel, a sensory organ, which contains peripheral nerves and neural crest
22 derived Schwann cells (**Figure 5D**). Together, the similar expression patterns of *EGFP* between

1 different stable transgenic lines suggests that our *peak5* transgenic zebrafish reflect endogenous
2 activity of *peak5* in isolation.

3
4 *peak5 is active in early melanoma patches and tumors*

5 Given that *peak5* is active in NCCs, we next asked if *peak5* is an active enhancer in
6 melanoma. Excitingly, in every transgenic line assessed for tumor growth (Lines A, B, and C),
7 *EGFP* was expressed in nearly all melanoma tumors (**Figure 6; Table 4**) (n = 55/59 *EGFP*+

8 tumors in 51 animals). Not only is *peak5* active in raised tumors, it is also active in small, unraised,
9 “preclinical” melanoma precursor lesions. Patches of *EGFP* positive cells, tracked over time,
10 drastically grew (**Figure 7**) (n = 9/15 *EGFP*+ patches tracked in 13 animals), similar to
11 *crestin:EGFP* positive patches (Kaufman et al., 2016). The remaining 6/15 patches did not
12 obviously expand but remained at least stable over a 2-3 month time period observed. *EGFP*
13 localization in precursor lesions and tumors is particularly striking because *EGFP* is either
14 minimally or not detectably expressed by epifluorescence in adult melanocytes, relative to patches
15 and tumors, in the lines studied. Of note, differences in expression levels of *EGFP* in the body of
16 the fish may be due to differences in the copy number of the transgene. Overall, *EGFP* expression
17 in patches and tumors indicates that *peak5* is an active enhancer in melanoma, and furthermore,
18 represents a *sox10* regulatory element that is active during early stages of melanoma. To our
19 knowledge, this is the first time a single enhancer has been shown to be active specifically in early
20 melanoma patches and tumors *in vivo*. In addition to *peak5*, we also generated several stable lines
21 to assess activities of the *sox10* minimal promoter, *peak1*, and *peak8* in melanoma. As predicted,
22 the *sox10* minimal promoter drove *EGFP* expression in melanoma patches and tumors across 6
23 stable lines (n = 38/46 *EGFP*+ tumors in 42 animals) (**Figure S5A-C’; Table 4**). Unlike *peak5*
24 transgenic lines, *peak1* and *peak8* stable lines expressed *EGFP* embryonically, but did not express

1 *EGFP* in nearly all melanoma tumors (n = 0/6 *EGFP*+ tumors in 6 *peak1* transgenic animals from
2 1 stable line; n = 1/26 *EGFP*+ tumors in 25 *peak8* transgenic animals from 3 independent stable
3 lines) (**Figure S5D-G'**; **Table 4**). These *peak1* and *peak8* data highlight the specificity of the
4 activity of *peak5* from within the *sox10* super-enhancer and demonstrate that not all sub-elements
5 are sufficient to drive *EGFP* expression in melanoma. These data also show that *peak5* is an
6 enhancer upstream of the *sox10* minimal promoter that is specifically active in melanoma, both in
7 precursor lesions and tumors.

8
9 *A conserved sequence within peak5 is necessary for activity in NCCs and melanoma*

10 We next asked if within *peak5* there is a minimal sequence that is necessary for activity in
11 neural crest and melanoma. We hypothesized that any nucleotide conservation within *peak5*
12 between zebrafish and other animals would identify a more critical regulatory target region.
13 Nucleotide sequence conservation of enhancers between zebrafish and mammals is rare (Ritter et
14 al., 2010); accordingly, we did not observe any detectable colinear sequence conservation of *peak5*
15 in humans or mice. However, past sequence comparison of the zebrafish exomes to the exomic
16 sequence of its nearest relative, the common carp (*Cyprinus carpio*), revealed high conservation
17 of synteny and homologous coding sequences (Henkel et al., 2012). Thus, we reasoned that
18 evolutionary important non-coding sequences may also be conserved. Aligning the nucleotide
19 sequence around the *sox10* locus in both zebrafish and carp using Nucleotide BLAST revealed
20 discrete regions of conservation, including *peaks 2, 4, 5, 8, 13* and the minimal promoter, as well
21 as most of the *sox10* coding region (**Figure 8A**). Interestingly, these conserved regions correspond
22 to enhancer assays that result in the most robust NCC expression (**Figure S2**).

23 Within *peak5*, we identified a 192 bp region that is conserved between carp and zebrafish,
24 as well as other selected members of the *Cyprinidae* family (*Sinocyclocheilus grahami*,

1 *Pimephales promelas*, *Carassius auratus*, and *Oxygymnocypris stewartii*) (**Figure 8B**). Regions
2 of conservation of *peak5* between species other than carp were located on whole-genome shotgun
3 contigs through Nucleotide BLAST. Therefore, to verify that these regions of conservation are
4 indeed near *sox10* in these other species, we aligned each identified scaffold to the zebrafish *sox10*
5 locus. We identified proximity to the *sox10* locus and/or synteny with other putative *sox10*
6 regulatory regions for all other examined species except *Pimephales promelas* (**Figure S6**).

7 To test if this *peak5* conserved sequence is required for *peak5* activity, we deleted the
8 sequence within the reporter assay plasmid containing the wild-type *peak5* sequence (**Figure 8C**).
9 Interestingly, compared to full-length wild-type *peak5* sequence, which exhibits activity in NCCs
10 and in KA neurons embryonically, embryos injected with the 192 bp conserved sequence deletion
11 construct (*peak5Δ192:betaglobin:EGFP*) exhibited diminished *peak5* activity in NCCs, while
12 maintaining strong EGFP localization in KA neurons (**Figure 8D-H'**). This suggests that this core
13 conserved sequence controls *peak5* activity within NCCs. Injection of only the *peak5* conserved
14 sequence (*peak5_conserved:betaglobin:EGFP*) resulted in labeling of only NCCs, albeit at weaker
15 levels compared to full-length *peak5*, demonstrating that the conserved sequence is sufficient for
16 *peak5* activity in NCCs and is not sufficient for labeling of KA neurons (**Figure 8D, 8I-J'**).

17 In accordance with these mosaic analyses, embryos from *peak5Δ192* stable transgenic lines
18 exhibit EGFP localization in KA neurons, but no clear localization in NCCs located in the dorsal
19 posterior region of the trunk (**Figure 9A-A'**; **Figure S7**). In contrast, NCCs in *peak5_conserved*
20 stable transgenic embryos are EGFP positive and KA neurons are not labeled (**Figure 9B-B'**;
21 **Figure S7**). In adult stable transgenic *peak5_conserved* zebrafish, cells within the maxillary barbel
22 are labeled, similar to wild-type *peak5* stable lines (**Figure 9D**); however, we did not observe
23 EGFP positive barbels in *peak5Δ192* stable transgenic animals (**Figure 9C**). Furthermore, the

1 majority of tumors in adult *peak5Δ192* stable transgenic animals are not EGFP positive (n = 21/23
2 EGFP- tumors in 22 animals from 3 different stable transgenic lines) (**Figure 9E-F; Table 4**).
3 These data suggest that the conserved sequence we identified is not only necessary for *peak5*
4 activity in a subset of NCCs, but also is necessary for robust *peak5* activity in melanoma.

5

6 *SoxE TFBS regulate peak5 activity in NCCs*

7 We next asked which TFBS within this conserved region of *peak5* are functionally
8 necessary for activity. Out of a list of predicted TFBS, some of which have roles in NCC
9 development, our attention was drawn to predicted dimeric SoxE binding sites (**Figure 10A**) as it
10 was previously noted that such dimeric SoxE binding sites have been shown to exhibit functional
11 relevance in mouse *Sox10* enhancers (Antonellis et al., 2008). Furthermore, dimeric SOXE TFBS
12 with 3-5 bp spacers are enriched in human melanoma cell putative enhancers identified by DNase
13 hypersensitivity site mapping (Huang et al., 2015). We therefore mutated both of these predicted
14 SoxE binding sites within the *peak5* WT sequence to functionally test their necessity (**Figure 10B-**
15 **C**). Mutated plasmid was co-injected with *crestin:mch* and scored for activity in NCCs and KA
16 neurons. Compared to *peak5* wild-type injected embryos, mutation of the SoxE binding sites result
17 in drastically diminished *EGFP* expression in NCCs, but KA *EGFP* expression is maintained
18 (**Figure 10D-J**). This mutational analysis confirms that the dimeric SoxE binding sites are
19 functional and likely important for the transcriptional regulatory activity of *peak5* in the neural
20 crest.

21
22

23 **Discussion**

1 Melanoma initiating cells express aspects of a neural crest program that is maintained as a
2 tumor progresses (Kaufman et al., 2016). One gene within the neural crest program, *sox10*, is a
3 critical regulator of NCC development and is also upregulated in both zebrafish and human
4 melanomas. Indeed, human melanomas immunostain positive for Sox10 in most cases, even more
5 often than the key melanocyte lineage regulator MITF (Clevenger et al., 2014; Ordóñez, 2014;
6 Willis et al., 2015). Past work demonstrated that modulation of *sox10* expression in melanocytes,
7 which normally lowly express *sox10*, affects melanoma onset rates (Clevenger et al., 2014;
8 Kaufman et al., 2016; Shakhova et al., 2012), illustrating that *sox10* plays a key role during
9 melanoma initiation. Therefore, deciphering the epigenetic regulation of *sox10* may elucidate
10 earlier mechanisms of melanoma initiation that drive NCP activation. Ongoing studies continue to
11 decipher the embryonic function of *sox10* in zebrafish (Chong-Morrison et al., 2018; Dutton et al.,
12 2001; Kelsh and Eisen, 2000), but to date, regulation of *sox10* expression in melanoma is poorly
13 understood. In this study, we screened individual putative *sox10* regulatory elements for activity
14 within NCCs and melanoma in zebrafish. From this analysis, we identified *peak5*, an enhancer that
15 is active in embryonic neural crest and melanoma, including precursor lesions.

16 Using ATAC-Seq data generated from zebrafish melanoma tumors, we identified putative
17 regulatory regions of open chromatin surrounding and within the *sox10* locus. Injection of reporter
18 constructs harboring these putative enhancers into embryos revealed that 9 out of 11 displayed
19 some level of activity within NCCs, as compared to injected negative controls, indicating their
20 potential as active enhancers in melanoma. Generation of multiple stable lines revealed that *peak5*,
21 a 669 bp sequence from 22.8 kb upstream of the *sox10* transcriptional start site, is active
22 embryonically in a subset of NCCs and KA neurons. In zebrafish adults, *peak5* is at most
23 minimally active in melanocytes in *in vivo* EGFP reporter assays, and is highly active in melanoma

1 patches and tumors. Furthermore, we identified a region within *peak5* that is conserved between
2 members of the *Cyprinidae* family and mediates the NCC activity of *peak5*. Within this conserved
3 region, dimeric SoxE binding sites are critically necessary for *peak5* activity within the neural
4 crest. Together, *peak5* serves as a proxy enhancer to understand how transcriptional regulation of
5 *sox10* and potentially other melanoma enriched genes influences melanoma initiation.

6 The identification of *crestin* as an early marker of melanoma provided a live marker of the
7 earliest detectable events of melanoma initiation and confirmed that the NCP activation is an early
8 step in the process (Kaufman et al., 2016). This study expands upon this knowledge by identifying
9 a regulatory element that can also be utilized as an *in vivo* specific reporter of melanoma initiation.
10 Importantly, unlike *crestin*, *sox10* is conserved across all vertebrate species; therefore, identifying
11 *sox10* enhancers that are active in melanoma in zebrafish could lead to the identification of human
12 *SOX10* enhancers that influence melanoma onset. While we were unable to identify stretches of
13 colinear conserved sequence between the zebrafish and human/mouse upstream regions of *SOX10*,
14 predicting the regulatory function of enhancers based on sequence conservation alone is often not
15 possible, such as in the case of *RET* enhancer elements, which function faithfully between human
16 and zebrafish in reporter assays, but are not conserved in an obvious way at the primary nucleotide
17 level (Fisher, 2006).

18 Further, we were intrigued to find that while zebrafish *sox10* enhancers are not conserved
19 in a colinear sequence with potential human *SOX10* enhancers, several zebrafish enhancers are
20 conserved by nucleotide sequence between members of the *Cyprinidae* family. Conservation of
21 *peaks 2, 4, 5, 8, 13* and the *sox10* minimal promoter between zebrafish and carp suggest an
22 evolutionary pressure to maintain these regulatory sequences. Interestingly, the highest region of
23 conservation for each peak is oriented near the center of the ATAC-Seq peak, including *peak5*. As

1 the conserved region of *peak5* controls activity in NCCs, but not activity in KA neurons, regions
2 of conservation in *peaks 2, 4, 8, and 13* may represent core sequences that control NCC activity of
3 the enhancer. Future work to dissect the role of these core conserved enhancers in NCC
4 development and melanoma will test this hypothesis.

5 This region of high conservation within *peak5* also leads to the hypothesis that the key
6 TFBS controlling *peak5* activity in NCCs and melanoma are also conserved. Identification of SoxE
7 dimeric binding sites points toward potential transcription factors, most notably Sox10, that may
8 bind and activate *peak5*. Similar to deletion of the conserved sequence within the context of full-
9 length *peak5*, mutation of both SoxE TFBS lessens *peak5* activity in NCC, but does not affect
10 *peak5* activity within KA neurons. Therefore, members of the SoxE family (Sox10, Sox9, and
11 Sox8) are candidate transcription factors that may bind and activate *peak5*. Dimeric SOXE binding
12 sites are overrepresented in human melanoma cells in regions of open chromatin (Huang et al.,
13 2015) and have also been observed near promoters of genes regulated by Sox10 (Mollaaghbab
14 and Pavan, 2003), suggesting that Sox10 may bind the dimeric sites in *peak5*. Mutation of the
15 dimerization domain within Sox10 in mice also causes defects in melanocyte development,
16 underscoring the importance of dimeric Sox10 binding (Schreiner et al., 2007).

17 Interestingly, dimeric SoxE sites have also been shown to be necessary for activity of two
18 mouse *Sox10* enhancers, MCS4 and MCS7 (Antonellis et al., 2008). Since zebrafish *sox10*
19 enhancers are not conserved by sequence with mouse *Sox10* enhancers, it will be interesting to test
20 if either mouse MCS4 or MCS7 *Sox10* enhancers have overlapping embryonic and adult activity
21 patterns with *peak5* in zebrafish. Moreover, examining the other functional TFBS present in *peak5*
22 in zebrafish may help predict and identify a *peak5*-equivalent enhancer in humans that is depend
23 on the same or similar complement of TFBS.

1 One caveat of determining the functionality of a putative enhancer is the potential to
2 identify false positives. In zebrafish, random genome integration of enhancers using the Tol2
3 system can result in enhancer trapping (Udvadia and Linney, 2003). Thus, we generated multiple
4 stable lines for *peak5*, which exhibited similar expression patterns, suggesting that our lines reflect
5 endogenous spatial and temporal activity of *peak5* embryonically and in melanoma. Variations
6 amongst transgene integration site in different stable lines leading to subtle position effects in
7 addition to genomic instability in melanoma tumors, may also explain why we occasionally
8 observe EGFP negative tumors in stable lines that predominately express EGFP in tumors (*peak5*
9 n = 4/59 EGFP negative tumors) and a small number of EGFP positive late-stage tumors in *peak8*
10 (n = 1/26) and *peak5Δ192* (n = 2/23) stable lines.

11 Finally, we propose a model by which *peak5* contributes to increased *sox10* expression
12 regulation in melanoma initiation and progression (**Figure 11**). In this model, the *peak5* region is
13 open and enhances *sox10* expression in early melanoma precursor lesions and, through a feed-
14 forward mechanism, autoregulates increasing *sox10* expression by binding at the dimeric SoxE
15 sites in *peak5*. Again, while stretches of conserved, noncoding DNA are most often not readily
16 identifiable by comparison between humans/mouse and zebrafish, the presence of such dimeric
17 SoxE sites in *bona fide* murine NCC enhancers (Antonellis et al., 2008) and putative human
18 melanoma enhancers (Huang et al., 2015) is consistent with this potentially being a general
19 mechanisms by which Sox10 target genes are upregulated in melanoma. Given that *peak5* is not
20 active in all cells that express *sox10* during embryogenesis, other *sox10* enhancers likely also
21 contribute to *sox10* regulation during melanoma initiation. Together, developmentally incorrect
22 re-activation of *sox10* enhancers in zebrafish and human melanomas suggests that understanding
23 the activity and contribution of individual enhancers within this locus will illuminate earlier

1 initiating steps of melanoma. This study identifies one of these enhancers, *peak5*, that is active
2 embryonically and in melanoma, and lays the groundwork to further dissect the role of other *sox10*
3 enhancers in zebrafish and higher vertebrates in melanoma initiation.

4 **Methods and Materials**

5 *Zebrafish Lines and Rearing Conditions*

6 Zebrafish were reared in accordance with Washington University IACUC animal protocols in the
7 Washington University Zebrafish Consortium Facility. Adult zebrafish were crossed either as pairs
8 or groups, and embryos were raised in egg water (5 mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂,
9 0.33 mM MgSO₄) at 28.5°C. Larvae were staged at days post fertilization (dpf) and adults were
10 staged at months post fertilization (mpf). The following strains of transgenic zebrafish were used
11 in this study: AB*, *Tg(BRAF^{V600E}/p53^{-/-}/crestin:EGFP)* (Kaufman et al., 2016),
12 *Tg(BRAF^{V600E});p53^{-/-}* (Patton et al., 2005), *Tg(sox10(7.2):mRFP)* (Kucenas et al., 2008),
13 *Tg(crestin:mcherry)*, *Tg(peak5:betaglobin:EGFP)*, *Tg(peak5Δ192:betaglobin:EGFP)*,
14 *Tg(peak5_conserved:betaglobin:EGFP)*, *Tg(sox10minimalpromoter:EGFP)*,
15 *Tg(peak1:betaglobin:EGFP)*, and *Tg(peak8:betaglobin:EGFP)*.

16 17 *Isolation of Bulk Tumors*

18 Following humane euthanasia, grossly visible bulk melanoma tumors from *Tg(BRAF^{V600E}/*
19 *crestin:EGFP);p53^{-/-}* zebrafish were excised with a razor, manually sheared, and incubated in fresh
20 0.9X PBS with 2.5 mg/mL Liberase for up to 30 minutes to dissociate cells. Fetal bovine serum
21 addition terminated the digestion, and the cells were passed through a 40 mm filter. Cells were
22 centrifuged at 2000 x g for 5 minutes at 4°C. Supernatant was removed and the cells were
23 resuspended in 500 µL of 0.9X PBS and kept on ice.

1
2 *ATAC-Seq*
3 50,000 cells per sample underwent tagmentation reaction with Nextera Tn5 transposase using the
4 Illumina Nextera kit and purified with a Qiagen MinElute reaction kit (Buenrostro et al., 2015).
5 The DNA was then PCR amplified for 9 cycles to add indexing primers. SPRI AMPure beads
6 enriched for fragments under ~600 bp. The DNA was cycled again with the 9-cycle protocol,
7 followed by cleanup with SPRI AMPure beads. The DNA was quantified with a Qubit DNA High
8 Sensitivity assay and analyzed for quality and size distribution on an Agilent TapeStation with a
9 High Sensitivity D5000 ScreenTape. Samples were pooled for a 10 nM final overall
10 concentration. Sequencing was performed with an Illumina HiSeq 2500 system with 2 x 50 bp
11 read length by the Washington University in St. Louis School of Medicine Genome Technology
12 Access Center.

13
14 *ATAC-Seq Data Analysis*
15 Reads were demultiplexed then trimmed with Cutadapt (adaptor sequence 5'-
16 CTGTCTCTTATACACATCT-3' for both reads) and checked with FastQC to ensure quality. The
17 reads were aligned to the GRCz10/danRer10 genome using BWA-MEM and sorted using
18 SAMtools. Duplicate reads removed with Picard tools using the following parameters:
19 ASSUME_SORTED=true, VALIDATION_STRINGENCY=LENIENT. The files were indexed
20 with SAMtools then filtered for high quality alignments using the following parameters: -f 3, -F
21 4, -F 8, -F 256, -F 1024, -F 2048, -q 30. MACS2 was then used to identify peaks with the callpeak
22 command with parameters -g 1.4e9, -q 0.05, --nomodel, --shift -100, --extsize 200. Differential
23 peak accessibility was assessed with the MACS2 bdgdiff function. HOMER was then used to

1 annotate the peaks (annotatePeaks.pl) and identify enriched transcription factor motifs using
2 findMotifsGenome.pl (-size 200, -mask parameters).

3
4 *Cloning putative enhancers*
5 All putative enhancer elements except for *peak13* were PCR amplified from genomic DNA from
6 AB* zebrafish larvae or *Tg(BRAF^{V600E});p53^{-/-}* larvae with high-fidelity DNA polymerases. *peak13*
7 was amplified from BAC CHORI-211:212I14 (a gift from the Johnson Lab). Genomic locations
8 for each putative enhancer in the Zv10 build of the zebrafish genome are listed in **Table 1** and
9 primers used to amplify each sequence and total cloned sizes of each element are listed in **Table**
10 **2**. PCR amplified enhancer elements were cloned into pENTR5' vectors (ThermoFisher). Anytime
11 Phusion (NEB) was utilized to amplify a putative enhancer, adenines were added to the end of the
12 PCR product using *Taq* polymerase (Promega) before performing a TOPO reaction. Using
13 plasmids from the Tol2 kit, all putative enhancers were placed upstream of a basal promoter,
14 mouse *beta-globin* (Tamplin et al., 2011) via Gateway LR reactions with the following
15 combination of Gateway vectors: p5E-enhancer, pME-*betaglobin:EGFP*, p3E-polyA, and
16 pDestTol2pA. To generate a *sox10* minimal promoter plasmid, a *sox10* minimal promoter was
17 PCR amplified from genomic DNA from AB* larvae with primers (**Table 1 and 2**) containing
18 overlapping nucleotides to sequences on either side of the AgeI site in pENTR-EGFP2 (Addgene
19 #22450), similarly described in (Quillien et al., 2017). The PCR product was then cloned into
20 pENTR-EGFP2 and digested with AgeI, through a Gibson Assembly reaction using an in-house
21 Gibson mixture (gifted by the Solnica-Krezel lab, Washington University in St. Louis; NEB
22 Phusion, NEB Taq DNA Ligase, NEB T5 Endonuclease). This pME-
23 *sox10minimalpromoter:EGFP* was then used in a Gateway LR reaction with p5E-MCS, p3E-
24 polyA, and pTol2Destp2A to generate the final reporter construct. All PCR amplified sequences

1 were Sanger sequenced by Genewiz. The list of plasmids used in this study can be found in **Table**
2 **3**. The conserved *peak5* sequence was PCR amplified from the previously cloned *peak5* sequence,
3 cloned into a pENTR5' vector, and placed upstream of *betaglobin:EGFP* through a Gateway
4 reaction, as described above.

5
6 *Screening putative enhancers and generation of transgenic zebrafish lines*
7 One cell stage *Tg(BRAF^{V600E});p53^{-/-}* embryos were injected with 20 ng/µL of an enhancer reporter
8 plasmid in addition to 15 ng/µL of *crestin:mch*, and 20 ng/µL of Tol2 transposase mRNA. 1 dpf
9 larvae were screened for EGFP and mCh localization in the neural crest on a stereomicroscope.
10 Embryos were scored as “strong expression” if 5 or more cells were EGFP and mCh positive.
11 Three technical and biological replicates of injections were performed for each putative enhancer.
12 At least 30 embryos scored as positive from each replicate were raised to adulthood. Founders
13 were identified by out-crossing F0 adults, and progeny were screened from 1 dpf - 5 dpf for EGFP
14 localization and subsequently raised to adulthood and screened for *EGFP* expression in melanoma
15 patches and tumors (**Table 4**).

16
17 *Imaging*
18 For live-imaging, larvae were anesthetized in Tricaine, embedded in 0.8% agarose, and imaged
19 with a Nikon SMZ18 fluorescent dissecting microscope. For imaging adults, fish were
20 anesthetized in Tricaine and imaged with a Nikon SMZ18 fluorescent dissecting microscope.
21 Images were processed with Photoshop and ImageJ. The Photomerge function in Photoshop was
22 used to stitch together tiled images of adult zebrafish.

23
24 *Evolutionary conserved sequence identification*

1 To generate the Dot-Matrix view, sequences surrounding the *sox10* locus in both zebrafish
2 (Chromosome 3: 1976387-2026723, danRer10) and carp (scaffold: LG6, Chromosome 6:
3 16126287-16150639), were aligned with optimization for more dissimilar sequences in BLAST
4 (<http://blast.ncbi.nlm.nih.gov>). BLAST with optimization for more dissimilar sequences with
5 discontiguous megablast was also used to compare the zebrafish *peak5* sequence to selected
6 members of the *Cyprinidae* family (**Table 5**). The nucleotide collection or whole-genome shotgun
7 contigs databases were chosen as the search set. The most conserved regions of *peak5* nucleotide
8 sequences from each species were then aligned with M-Coffee
9 (<http://tcoffee.crg.cat/apps/tcoffee/do:mcoffee>). The fasta_aln file was downloaded and Boxshade
10 (https://embnet.vital-it.ch/software/BOX_form.html) was then used to shade the alignment using
11 RTF_new as the output format and “other” as the input sequence format. TFBS predictions were
12 identified with FIMO (<http://meme-suite.org/tools/fimo>) using the JASPAR core 2016 vertebrate
13 list. SoxE dimeric binding sites were predicted with JASPAR
14 (http://jaspar.genereg.net/search?q=&collection=CORE&tax_group=vertebrates), searching
15 explicitly for Sox10 motifs using an 80% relative profile score threshold.

16
17 *Plasmid Mutagenesis and Deletion Analysis*

18 The Q5 Site-Directed Mutagenesis Kit (NEB) was used according to the manufacturer’s
19 instructions to either delete the *peak5* conserved 192 base pair sequence or mutate SoxE binding
20 sites. The correct deletion or mutations were confirmed through Sanger Sequencing at Genewiz.
21 20 ng/µL of *peak5* full-length wild type, *peak5*Δ192, or *peak5* full-length SoxE site mutated
22 plasmids were co-injected with 15 ng/µL of *crestin:mch* and 20 ng/µL of Tol2 transposase mRNA
23 into embryos derived from a *Tg(BRAF^{V600E});p53^{-/-}* in-cross. Only embryos that expressed *mCh*,
24 ensuring successful injection, were scored on 1 dpf for expression of *EGFP*. EGFP+ embryos that

1 had identifiable cell type labeling were categorized into the following cell-localization categories:
2 neural crest, KA neurons, both the neural crest and KA neurons, or ectopic. The scorer was blinded
3 to treatment groups when scoring. Three technical and biological replicates were performed for
4 the conserved deletion experiment and two technical and biological replicates were performed for
5 the SoxE TFBS mutation experiment. Stable lines were identified and characterized as described
6 above (**Table 4**).

7
8 *Statistical Analyses*

9 Chi-squared analyses were used to compare *peak5* conserved sequence deletion and SoxE TFBS
10 mutagenesis injection experiments in GraphPad Prism 8.

11

12 **Author Contributions**

13 R.L.C., E.T.K, and C.K.K. designed research, and R.L.C, E.T.K., S.K.D, V.G., and S.S. performed
14 research. R.L.C. and C.K.K. analyzed data, and R.L.C. and C.K.K. wrote the manuscript.

15 **Acknowledgements**

16 Research reported in this publication was supported in part by the National Cancer Institute of the
17 National Institutes of Health under award number R01CA240633. The content is solely the
18 responsibility of the authors and does not necessarily represent the official views of the National
19 Institutes of Health. R.L.C. was supported by the National Science Foundation Graduate Research
20 Fellowship (DGE-1745038). C.K.K. was supported by the Cancer Research Foundation Young
21 Investigator Award. E.T.K was supported by T32 GM007067, and S.K.D is supported by T32
22 GM007067. We thank members of the Kaufman lab and Souroullas lab for helpful discussions;

1 the Washington University Zebrafish Consortium; the Washington University in St. Louis School
2 of Medicine Genome Technology Access Center; Bo Zhang in the Department of Developmental
3 Biology and the Center for Regenerative Medicine at Washington University School of Medicine
4 for assistance with ATAC-Seq analysis; S. Johnson lab members for the BAC plasmid
5 (Washington University in St. Louis); S. Kucenas (University of Virginia) for the
6 *Tg(sox10:mRFP)* line; M. Bagnall (Washington University in St. Louis) for identifying the
7 Kolmer-Agduhr neurons.

8 **Competing Interests**

9 C.K.K. reports no competing financial interests.
10

11

12

13

14

15

Figures

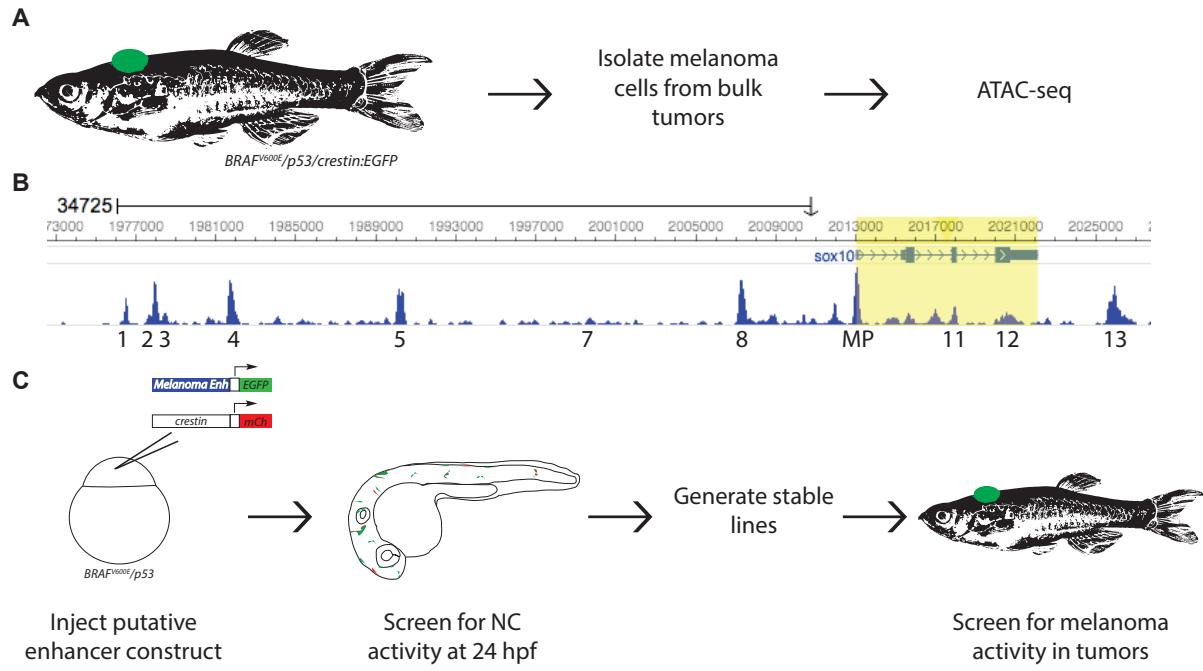


Figure 1: ATAC-Seq identifies putative *sox10* enhancers in zebrafish melanoma tumors

A) Schematic for obtaining zebrafish melanoma tumor cells for ATAC-Seq. **B)** Regions of open chromatin surrounding the *sox10* locus. Peak numbers are annotated below. MP = minimal promoter. **C)** Schematic for screening putative enhancers for activity in NCCs and melanoma.

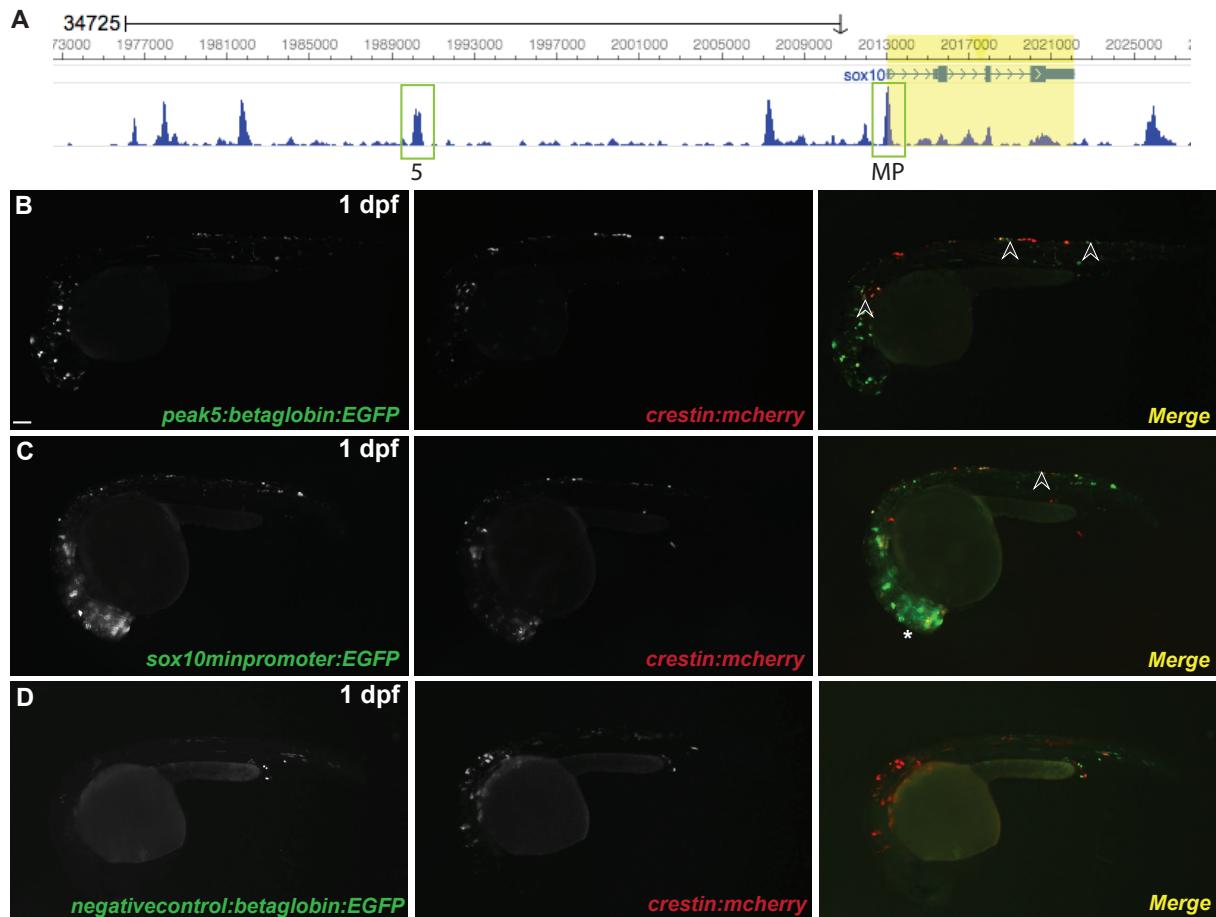


Figure 2: *peak5* is an active neural crest enhancer

A) Location of *peak5* and the *sox10* minimal promoter (MP) based on ATAC-Seq. **B)** *peak5*-driven EGFP expression is mosaically present in NCCs (arrowheads) in F0 injected embryos, as indicated by co-localization with *crestin:mcherry* at 1 dpf. **C)** The *sox10* minimal promoter in F0 embryos is active in both NCCs (arrowheads) and the CNS (asterisk). **D)** F0 embryos injected with a negative control (~76 kb upstream of the *sox10* TSS) exhibit only limited expression of EGFP not localized to NCCs.

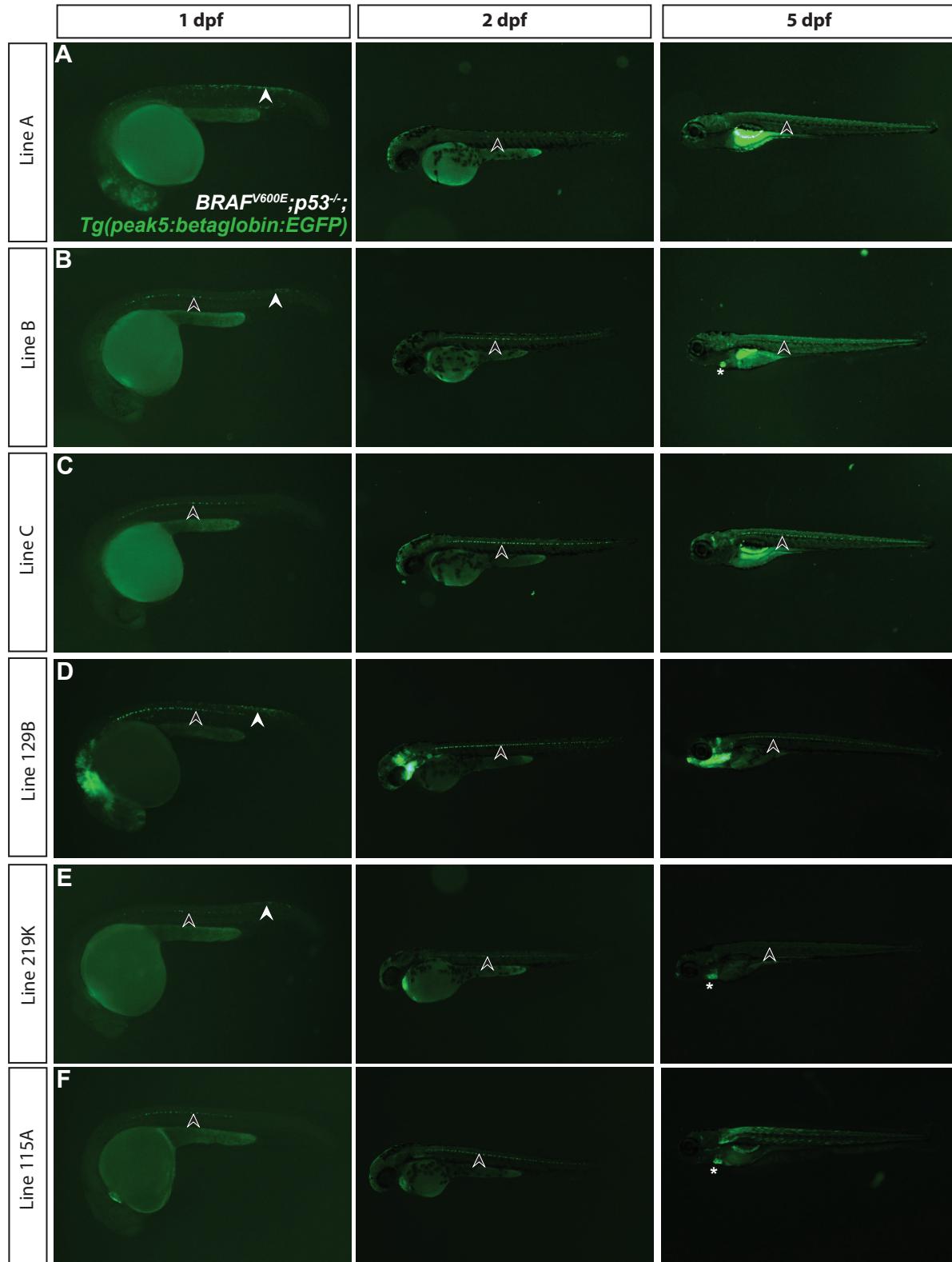


Figure 3: *peak5* stable transgenic lines exhibit similar expression patterns

Figure 3 (cont) Multiple transgenic lines derived from independent founder (F0) parental fish were established that transmit the *peak5:betaglobin:EGFP* reporter through the germline. White arrowheads indicate NCCs. Black arrowheads point to KA neurons. Asterisks indicate heart EGFP localization in the heart. **A)** Stable line A exhibits strong EGFP localization in NCCs at 1 dpf. At 2 dpf, some NCCs are labeled, and KA neurons are also faintly visible. At 5 dpf, the most prominently labeled cells are KA neurons and cells near or within the swim bladder. **B)** At 1 dpf, Line B exhibits EGFP localization in posterior premigratory NCCs and strong localization in KA neurons. By 3 dpf and 5 dpf, EGFP localization is primarily within KA neurons and the heart. **C)** Line C mainly exhibits EGFP localization in KA neurons at 1 dpf, 2 dpf, and 3 dpf. **D)** In Line 129B, NCCs, KA neurons, and cells in the cranial CNS are strongly labeled with EGFP at 1 dpf. Labeling of some NCCs is maintained at 2 dpf, and by 5 dpf EGFP is primarily localized within KA neurons and parts of the head. **E)** Line 219K exhibits EGFP localization in some premigratory NCCs, KA neurons, and the heart at 1 dpf. At 5 dpf, EGFP is mainly localized in KA neurons and the heart. **F)** Line 115A displays signs of ectopic *EGFP* expression. At 1 dpf, EGFP localization is mainly visible in KA neurons and the heart. By 2 dpf, EGFP is also localized in the muscles, which is more prominent at 5 dpf. KA neurons and the heart are also labeled at 2 dpf and 5 dpf.

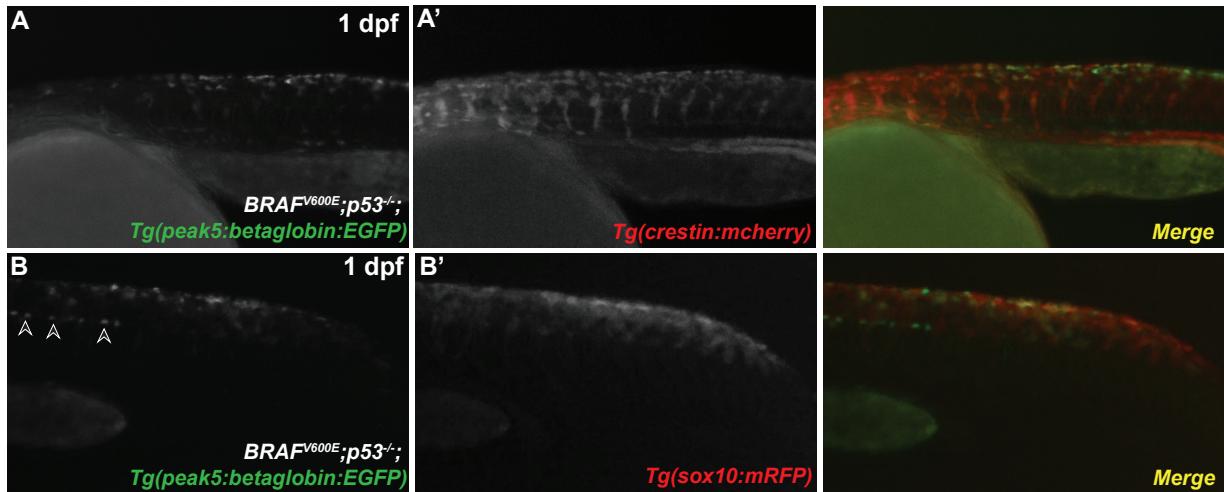


Figure 4: EGFP positive cells co-label with *crestin:mcherry* and *sox10(7.2):mRFP*

A) *peak5* (Stable Line A) is active in cells with NCC morphology and NCC localization at 1 dpf. **A'**) These EGFP positive cells co-label with *Tg(crestin:mCh)* positive cells. **B)** *peak5* (Stable Line 129B) also exhibits EGFP localization in NCCs and **B')** co-labels with *Tg(sox10(7.2):mRFP)* positive cells. Arrowheads indicate EGFP positive KA neurons.

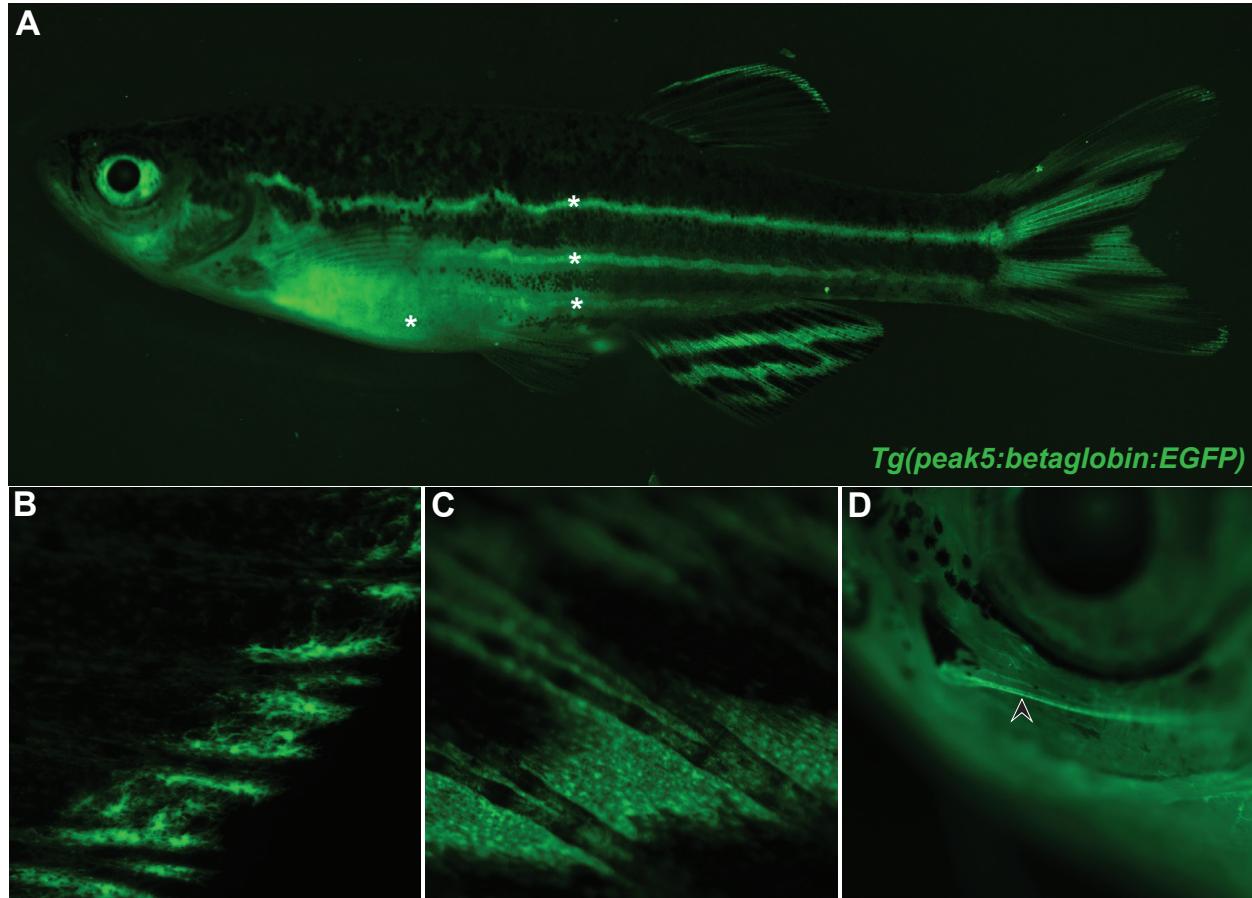


Figure 5: *peak5* is active in regions of fins and the peripheral nervous system in adults

A) Whole animal image of *peak5* active in stable transgenic line A. Asterisks (*) indicate regions of auto-fluorescence; a long pass filter set allows discrimination between the yellow hue of autofluorescent tissue and EGFP-expressing cells. **B)** EGFP positive cells in the tip of the dorsal fin. **C)** EGFP positive cells in the caudal fin. **D)** EGFP positive cells peripheral nervous system cells within the barbel (arrowhead).



Figure 6: *peak5* is active in melanoma tumors across multiple stable lines

A-B) Bright field images of tumors, circled, in *Tg(peak5:betaglobin:EGFP)* stable line A in the *Tg(BRAF^{V600E};p53^{-/-})* zebrafish background. **A'-B')** Tumors are EGFP positive and EGFP is highly expressed in tumors compared to any low EGFP expression elsewhere the fish. **C-D)** Bright field images of *Tg(peak5:betaglobin:EGFP)* stable line B with tumors, circled. **C'-D')** Tumors are EGFP positive. **E-F)** Bright field images of *Tg(peak5:betaglobin:EGFP)* stable line C with tumors, circled. **E'-F')** Tumors are EGFP positive.

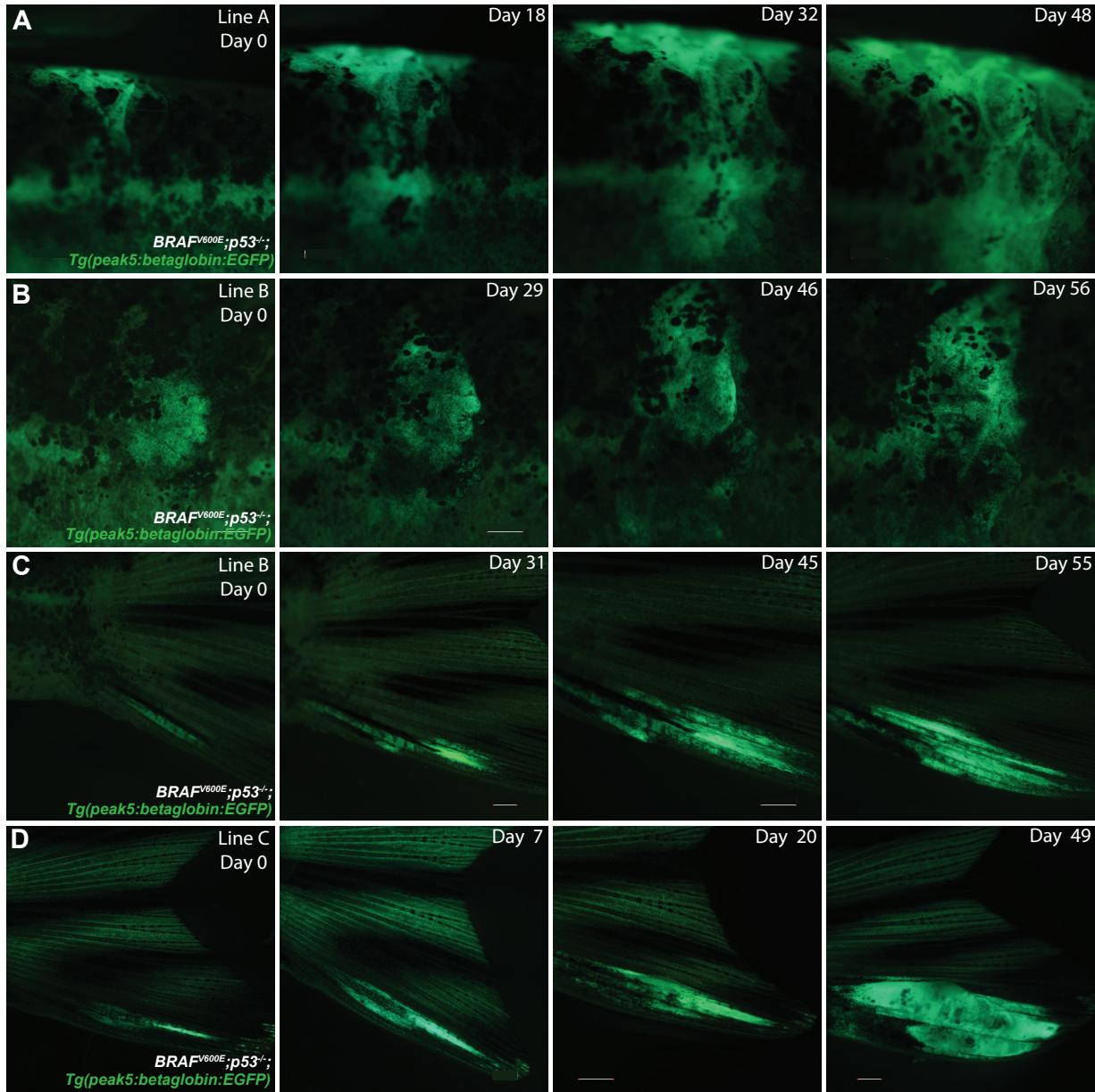


Figure 7: *peak5* is active in melanoma precursor lesions

A) *Tg(peak5:betaglobin:EGFP)* stable line A zebrafish with dorsally located melanoma precursor lesion that grows over time. **B)** *Tg(peak5:betaglobin:EGFP)* stable line B zebrafish with a melanoma precursor lesion in a scale that expands in size over time. **C)** *Tg(peak5:betaglobin:EGFP)* stable line B zebrafish with precursor lesion located on the tail that grows over time. **D)** *Tg(peak5:betaglobin:EGFP)* stable line C zebrafish with a precursor lesion on the tail that grows into a tumor. Day 0 denotes the first day the precursor lesion was observed.

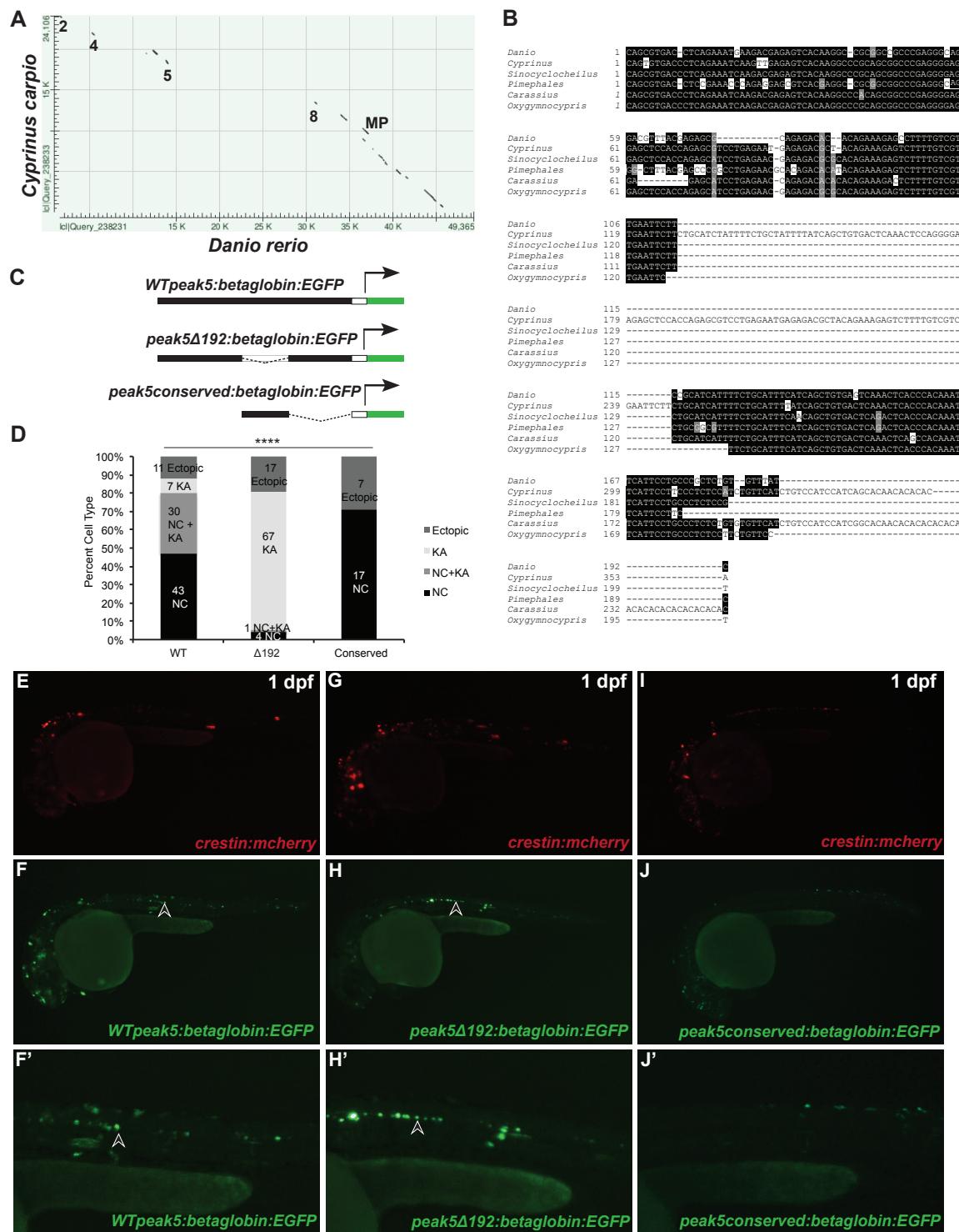


Figure 8: A conserved region of *peak5* is necessary for neural crest activity

A) Dot-matrix view of the alignment of the carp (*Cyprinus carpio*) *sox10* coding and surrounding non-coding genomic region of *sox10* compared to the same genomic region in zebrafish (*Danio rerio*)

rerio). Numbers indicate regions where peak sequences are conserved. MP = minimal promoter. **B)** Sequence alignment of the most conserved region of *peak5* to members of the *Cyprinidae* family. **C)** Schematic of plasmids used in the experiment. **D)** Quantification of the percentage of categorizable EGFP+ embryos exhibiting NC, KA, or both NC and KA labeling, or ectopic expression. NC = neural crest. KA = Kolmer-Agduhr neurons. **** p-value < 0.0001. **E)** *crestin:mch* expression in an embryo injected with *WTpeak5* and *crestin:mch*. **F)** *WTpeak5* is active in both neural crest and **F')** KA neurons. **G)** *crestin:mch* expression in an embryo injected with *peak5Δ192* and *crestin:mch*. **H)** *peak5Δ192* is not active in NCCs, but **H')** is active in KA neurons. **I)** *crestin:mch* expression in an embryo injected with the conserved *peak5* sequence and *crestin:mch*. **J)** The conserved *peak5* sequence is active in neural crest and **J')** does not exhibit activity in KA neurons. Arrowheads indicate KA neurons.

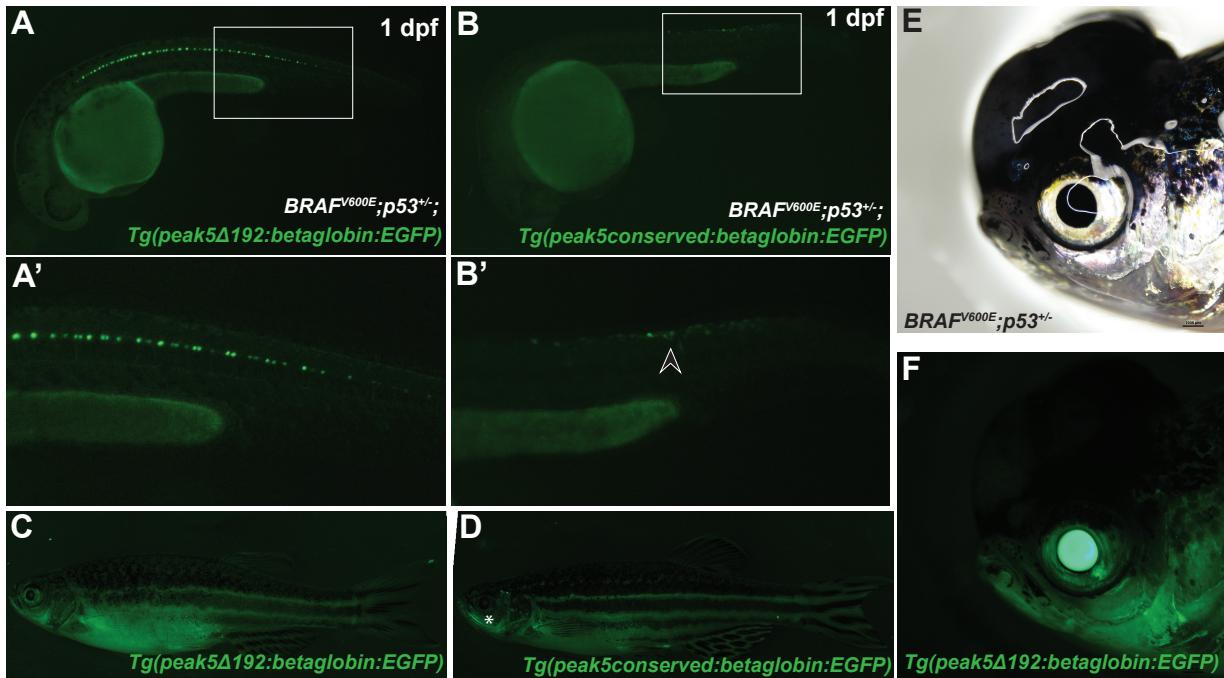


Figure 9: A conserved region of *peak5* is necessary for melanoma activity

A and A') *Tg(peak5Δ192:EGFP)* embryos exhibit EGFP localization in KA neurons, but no overt localization in NCCs in the trunk is observed. **B-B')** Contrastingly, *Tg(peak_conserved:EGFP)* embryos do not have labeled KA neurons, but exhibit EGFP positive cells in the trunk of the embryo. **C**) Little EGFP localization is observed in *Tg(peak5Δ192:EGFP)* adult fish whereas **D**) cells within the maxillary barbels (asterisk) of *Tg(peak_conserved:EGFP)* are EGFP positive. **E-F**) The majority of tumors that develop in *Tg(peak5Δ192:EGFP)* are EGFP negative.

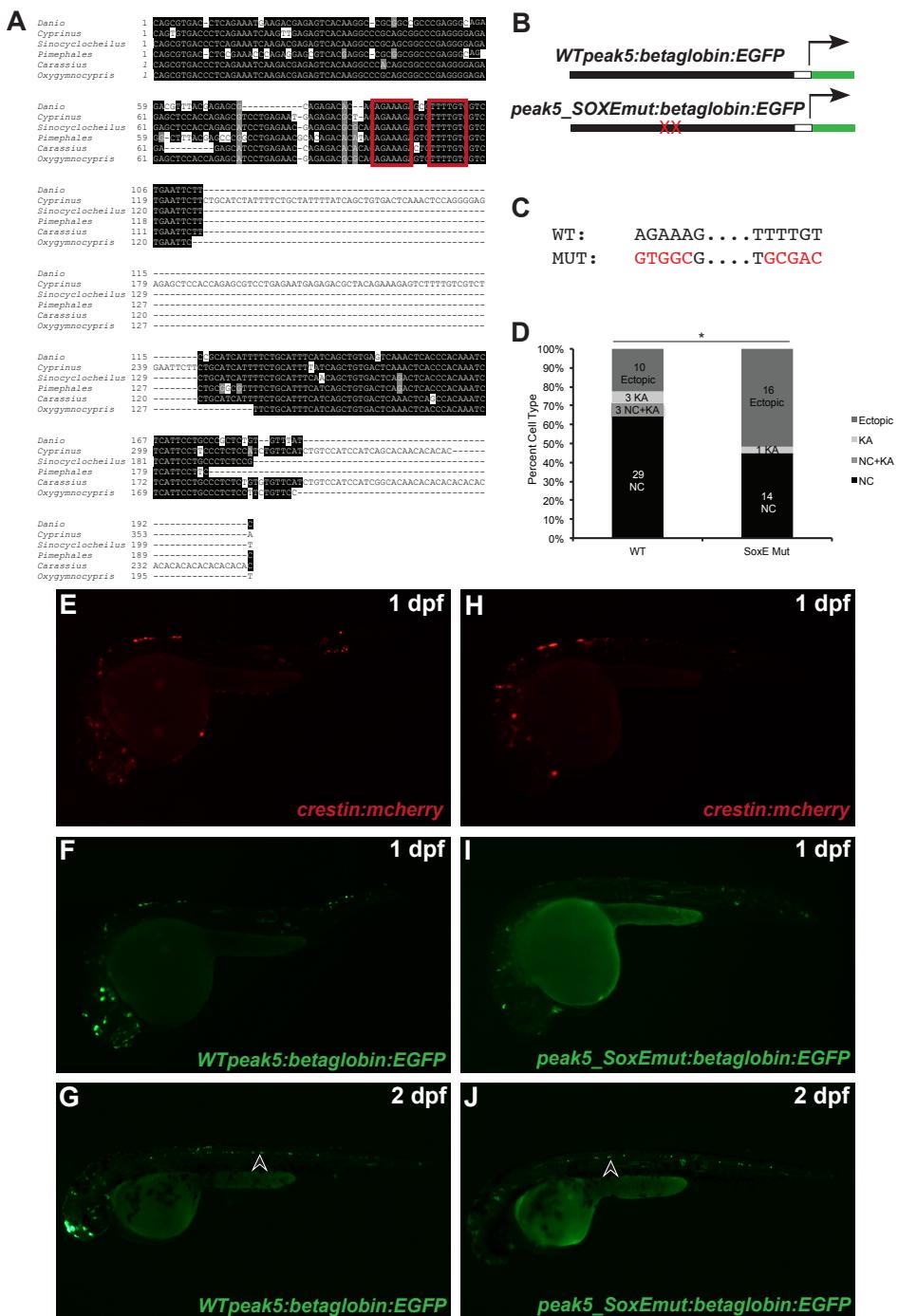


Figure 10: Mutation of dimeric SoxE TFBS affect *peak5* neural crest activity

A) Dimeric SoxE TFBS, as predicted by JASPAR (outlined in red), are present in the conserved sequence of *peak5*. **B**) Schematic of plasmids used in the experiment. Mutations to SoxE TFBS were made in the context of the full-length WT *peak5* plasmid. **C**) Five out of six nucleotides for each motif were mutated. **D**) Quantification of the percentage of categorizable EGFP+ embryos

exhibiting NC, KA, both NC and KA, or ectopic expression. NC = neural crest. KA = Kolmer-Agduhr neurons. * p-value = 0.0398. **E)** *crestin:mch* expression in an embryo injected with *WT peak 5* and *crestin:mch*. **F)** *WT peak5* is strongly active in neural crest at 1 dpf. **G)** *WT peak5* is strongly active in neural crest derived cells and KA neurons at 2 dpf. **H)** *peak5_SoxEmut* is not active in the neural crest, but **J)** is active in KA neurons. Arrowheads indicate KA neurons.

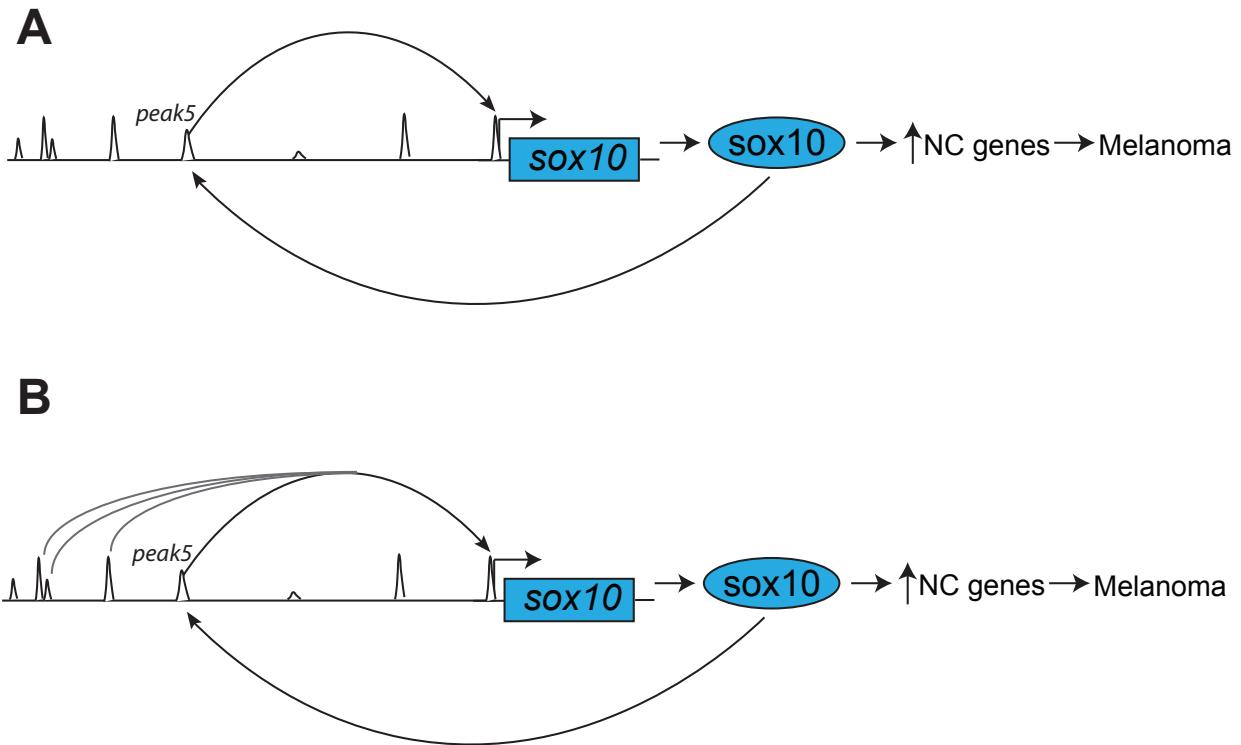


Figure 11: Model of *peak5* regulation of *sox10* expression during melanoma initiation

A) In our model, *peak5* activates *sox10* expression in NCCs during development and then again in melanoma precursor cells, likely through dimeric SoxE TFBS. This leads to upregulation of neural crest related genes and melanoma onset. **B)** *peak5* is most likely not the only *sox10* enhancer that regulates *sox10* expression. Multiple elements within the *sox10* super-enhancer may coordinate together to upregulate *sox10* expression during melanoma initiation.

References

- Antonellis, A., Huynh, J.L., Lee-Lin, S.-Q., Vinton, R.M., Renaud, G., Loftus, S.K., Elliot, G., Wolfsberg, T.G., Green, E.D., McCallion, A.S., et al. (2008). Identification of neural crest and glial enhancers at the mouse Sox10 locus through transgenesis in zebrafish. *PLoS Genet.* *4*, e1000174.
- Bradner, J.E., Hnisz, D., and Young, R.A. (2017). Transcriptional Addiction in Cancer. *Cell* *168*, 629–643.
- Buenrostro, J.D., Wu, B., Chang, H.Y., and Greenleaf, W.J. (2015). ATAC-seq: A Method for Assaying Chromatin Accessibility Genome-Wide. *Curr. Protoc. Mol. Biol.* *109*, 21.29.1-21.29.9.
- Chong-Morrison, V., Simões, F.C., Senanayake, U., Carroll, D.S., Riley, P.R., and Sauka-Spengler, T. (2018). Re-purposing Ac/Ds transgenic system for CRISPR/dCas9 modulation of enhancers and non-coding RNAs in zebrafish (Developmental Biology).
- Clevenger, J., Joseph, C., Dawlett, M., Guo, M., and Gong, Y. (2014). Reliability of immunostaining using pan-melanoma cocktail, SOX10, and microphthalmia transcription factor in confirming a diagnosis of melanoma on fine-needle aspiration smears. *Cancer Cytopathol.* *122*, 779–785.
- Cronin, J.C., Watkins-Chow, D.E., Incao, A., Hasskamp, J.H., Schönewolf, N., Aoude, L.G., Hayward, N.K., Bastian, B.C., Dummer, R., Loftus, S.K., et al. (2013). SOX10 ablation arrests cell cycle, induces senescence, and suppresses melanomagenesis. *Cancer Res.* *73*, 5709–5718.
- Dutton, K.A., Pauliny, A., Lopes, S.S., Elworthy, S., Carney, T.J., Rauch, J., Geisler, R., Haffter, P., and Kelsh, R.N. (2001). Zebrafish colourless encodes sox10 and specifies non-ectomesenchymal neural crest fates. *Dev. Camb. Engl.* *128*, 4113–4125.
- Fisher, S. (2006). Conservation of RET Regulatory Function from Human to Zebrafish Without Sequence Similarity. *Science* *312*, 276–279.
- Fontanals-Cirera, B., Hasson, D., Vardabasso, C., Di Micco, R., Agrawal, P., Chowdhury, A., Gantz, M., de Pablos-Aragoneses, A., Morgenstern, A., Wu, P., et al. (2017). Harnessing BET Inhibitor Sensitivity Reveals AMIGO2 as a Melanoma Survival Gene. *Mol. Cell* *68*, 731-744.e9.
- Gallagher, S.J., Mijatov, B., Gunatilake, D., Tiffen, J.C., Gowrishankar, K., Jin, L., Pupo, G.M., Cullinane, C., Prinjha, R.K., Smithers, N., et al. (2014). The epigenetic regulator I-BET151 induces BIM-dependent apoptosis and cell cycle arrest of human melanoma cells. *J. Invest. Dermatol.* *134*, 2795–2805.
- Henkel, C.V., Dirks, R.P., Jansen, H.J., Forlenza, M., Wiegertjes, G.F., Howe, K., van den Thillart, G.E.E.J.M., and Spaink, H.P. (2012). Comparison of the exomes of common carp (*Cyprinus carpio*) and zebrafish (*Danio rerio*). *Zebrafish* *9*, 59–67.

Huang, Y.-H., Jankowski, A., Cheah, K.S.E., Prabhakar, S., and Jauch, R. (2015). SOXE transcription factors form selective dimers on non-compact DNA motifs through multifaceted interactions between dimerization and high-mobility group domains. *Sci. Rep.* *5*, 10398.

Kaufman, C.K., Mosimann, C., Fan, Z.P., Yang, S., Thomas, A.J., Ablain, J., Tan, J.L., Fogley, R.D., van Rooijen, E., Hagedorn, E.J., et al. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* *351*, aad2197.

Kelsh, R.N. (2006). Sorting out Sox10 functions in neural crest development. *BioEssays News Rev. Mol. Cell. Dev. Biol.* *28*, 788–798.

Kelsh, R.N., and Eisen, J.S. (2000). The zebrafish colourless gene regulates development of non-ectomesenchymal neural crest derivatives. *Dev. Camb. Engl.* *127*, 515–525.

Kerosuo, L., and Bronner-Fraser, M. (2012). What is bad in cancer is good in the embryo: importance of EMT in neural crest development. *Semin. Cell Dev. Biol.* *23*, 320–332.

Kucenas, S., Takada, N., Park, H.-C., Woodruff, E., Broadie, K., and Appel, B. (2008). CNS-derived glia ensheathe peripheral nerves and mediate motor root development. *Nat. Neurosci.* *11*, 143–151.

Kulesa, P.M., Kasemeier-Kulesa, J.C., Teddy, J.M., Margaryan, N.V., Seftor, E.A., Seftor, R.E.B., and Hendrix, M.J.C. (2006). Reprogramming metastatic melanoma cells to assume a neural crest cell-like phenotype in an embryonic microenvironment. *Proc. Natl. Acad. Sci. U. S. A.* *103*, 3752–3757.

Lo, J.A., and Fisher, D.E. (2014). The melanoma revolution: from UV carcinogenesis to a new era in therapeutics. *Science* *346*, 945–949.

Luo, R., An, M., Arduini, B.L., and Henion, P.D. (2001). Specific pan-neural crest expression of zebrafish Crestin throughout embryonic development. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* *220*, 169–174.

Maguire, L.H., Thomas, A.R., and Goldstein, A.M. (2015). Tumors of the neural crest: Common themes in development and cancer. *Dev. Dyn. Off. Publ. Am. Assoc. Anat.* *244*, 311–322.

Mollaaghbab, R., and Pavan, W.J. (2003). The importance of having your SOX on: role of SOX10 in the development of neural crest-derived melanocytes and glia. *Oncogene* *22*, 3024–3034.

Ordóñez, N.G. (2014). Value of melanocytic-associated immunohistochemical markers in the diagnosis of malignant melanoma: a review and update. *Hum. Pathol.* *45*, 191–205.

Patton, E.E., Widlund, H.R., Kutok, J.L., Kopani, K.R., Amatruda, J.F., Murphey, R.D., Berghmans, S., Mayhall, E.A., Traver, D., Fletcher, C.D.M., et al. (2005). BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol. CB* *15*, 249–254.

- Quillien, A., Abdalla, M., Yu, J., Ou, J., Zhu, L.J., and Lawson, N.D. (2017). Robust Identification of Developmentally Active Endothelial Enhancers in Zebrafish Using FANS-Assisted ATAC-Seq. *Cell Rep.* *20*, 709–720.
- Ritter, D.I., Li, Q., Kostka, D., Pollard, K.S., Guo, S., and Chuang, J.H. (2010). The importance of being cis: evolution of orthologous fish and mammalian enhancer activity. *Mol. Biol. Evol.* *27*, 2322–2332.
- Rönnstrand, L., and Phung, B. (2013). Enhanced SOX10 and KIT expression in cutaneous melanoma. *Med. Oncol. Northwood Lond. Engl.* *30*, 648.
- Schreiner, S., Cossais, F., Fischer, K., Scholz, S., Bösl, M.R., Holtmann, B., Sendtner, M., and Wegner, M. (2007). Hypomorphic Sox10 alleles reveal novel protein functions and unravel developmental differences in glial lineages. *Dev. Camb. Engl.* *134*, 3271–3281.
- Shakhova, O., Zingg, D., Schaefer, S.M., Hari, L., Civenni, G., Blunschi, J., Claudinot, S., Okoniewski, M., Beermann, F., Mihic-Probst, D., et al. (2012). Sox10 promotes the formation and maintenance of giant congenital naevi and melanoma. *Nat. Cell Biol.* *14*, 882–890.
- Tamplin, O.J., Cox, B.J., and Rossant, J. (2011). Integrated microarray and ChIP analysis identifies multiple Foxa2 dependent target genes in the notochord. *Dev. Biol.* *360*, 415–425.
- Tanaka, M., Roberts, J.M., Qi, J., and Bradner, J.E. (2015). Inhibitors of emerging epigenetic targets for cancer therapy: a patent review (2010-2014). *Pharm. Pat. Anal.* *4*, 261–284.
- Udvadia, A.J., and Linney, E. (2003). Windows into development: historic, current, and future perspectives on transgenic zebrafish. *Dev. Biol.* *256*, 1–17.
- Willis, B.C., Johnson, G., Wang, J., and Cohen, C. (2015). SOX10: a useful marker for identifying metastatic melanoma in sentinel lymph nodes. *Appl. Immunohistochem. Mol. Morphol. AIMM* *23*, 109–112.