An attractor state zone precedes neural crest fate in melanoma initiation

Alicia M. McConnell[1,2,3], Maggie H. Chassé[1,2,3], Haley R. Noonan[1,2,3,4], Jeffrey K. Mito[1,2,3,5], Julia Barbano[1], Erika Weiskopf[1], Irene F. Gosselink[1], Meera Prasad[1,2,3], Song Yang[1,2,3], Phammela Abarzua[6], Christine G. Lian[6], George F. Murphy[6], Cole Trapnell[7], and Leonard I. Zon[1,2,3*]

[1]Stem Cell Program and Division of Hematology/Oncology, Children's Hospital Boston, Howard Hughes Medical Institute, Boston, MA 02115, USA.

[2]Harvard Stem Cell Institute, Boston, MA 02115, USA.

[3]Harvard Medical School, Boston, MA 02115, USA.

[4]Biological and Biomedical Sciences Program, Harvard Medical School, Boston, Massachusetts 02115, USA.

[5]Department of Pathology, Brigham and Women's Hospital, Boston, MA 02215, USA.

[6]Program in Dermatopathology, Department of Pathology, Brigham and Women's Hospital, Boston, MA 02215, USA.

[7]Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA.

*Corresponding author: zon@enders.tch.harvard.edu

The field cancerization theory suggests that a group of cells containing oncogenic mutations are predisposed to transformation[1, 2]. We previously identified single cells in *BRAF^{V600E};p53$^{-/-}$* zebrafish that reactivate an embryonic neural crest state before initiating melanoma[3-5]. Here we show that single cells reactivate the neural crest fate from within large fields of adjacent abnormal melanocytes, which we term the "cancer precursor zone." These cancer precursor zone melanocytes have an aberrant morphology, dysplastic nuclei, and altered gene expression. Using single cell RNA-seq and ATAC-seq, we defined a distinct transcriptional cell attractor state for cancer precursor zones and validated the stage-specific gene expression initiation signatures in human melanoma. We identify the cancer precursor zone driver, ID1, which binds to TCF12 and inhibits downstream targets important for the maintenance of melanocyte morphology and cell cycle control. Examination of patient samples revealed precursor melanocytes expressing ID1, often surrounding invasive melanoma, indicating a role for ID1 in early melanomagenesis. This work reveals a surprising field effect of melanoma initiation *in vivo* in which tumors arise from within a zone of morphologically distinct, but clinically covert, precursors with altered transcriptional fate. Our studies identify novel targets that could improve early diagnosis and prevention of melanoma.

Melanoma incidence has risen sharply over the past 30 years; however multiple studies show high disagreement among pathologists in the diagnoses of early potential precursor melanocyte lesions[6-9]. The cancer attractor state theory presented by Stuart Kauffman 50 years ago hypothesizes that stable gene expression profiles arise from complex gene expression networks to reinforce malignant cell phenotypes[10-12]. However, due to the difficulty of visualization and

40 detection, the earliest stages in melanoma initiation have not been extensively studied in either
41 human or animal models on a morphological or transcriptional level.

42 Over 80% of melanomas contain activating mutations in the MAPK pathway, such as
43 $BRAF^{V600E}$, which are often found alongside tumor suppressor loss-of-function mutations[13, 14]. We
44 developed a zebrafish melanoma model in which melanocytes express $BRAF^{V600E}$ and are
45 deficient for *p53*. Despite the massive melanoma-prone cancerized field generated,
46 *BRAF*^{V600E}*;p53*$^{-/-}$ fish only develop one to three melanomas in their lifetime, with almost no
47 additional genetic mutations[3, 15]. In contrast to conventional dogma that melanomas arise from
48 melanocytic nevi, there is growing recognition that most melanomas arise de novo, despite the fact
49 that 70-90% of nevi carry the $BRAF^{V600E}$ mutation[13]. Thus, many potential melanoma precursors
50 may be clinically covert and are molecularly undefined. Moreover, the disconnect between the
51 acquisition of oncogenic mutations and the initiation of tumorigenesis indicates that additional
52 pathways are required to drive tumor initiation.

53 Melanoma arises due to malignant transformation of melanocytes, which are
54 embryonically derived from the neural crest. We previously developed a neural crest reporter,
55 *crestin:EGFP,* and showed that individual *BRAF*^{V600E}*;p53*$^{-/-}$ melanocytes reactivate the neural crest
56 progenitor state to form a tumor[3, 4]. Here we identify a novel intermediate step in melanoma
57 initiation and visualize melanoma arising from within morphologically and transcriptionally
58 aberrant cancer precursor zone melanocytes, which are driven by ID1 inhibition of TCF12. These
59 cancer precursor zones define cells that are transitioning to the cancer attractor state, which drives
60 and reinforces the epigenetic change of normal melanocytes into stable neural crest reactivated
61 melanoma cells.

62

63 **Melanoma forms in cancer precursor zones**

64 In order to characterize the earliest stages of melanoma formation, we generated *p53* deficient,
65 *crestin:EGFP* zebrafish in which melanocytes express $BRAF^{V600E}$, are marked red by
66 *mitfa:mCherry,* and are devoid of pigment due to a *tyrosinase* CRISPR knockout
67 (*mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$;crestin:EGFP;mcr:Empty;mitfa:mCherry;tyr$^{-/-}$*, referred to as
68 *BRAF;p53*)[3, 16, 17]. We performed time course imaging and found that prior to neural crest
69 reactivation, regions of hundreds of morphologically aberrant melanocytes express high levels of
70 *mitfa:mCherry*, which we term the cancer precursor zone (Fig. 1a, Extended Data Fig. 1a-c). We
71 closely followed a cohort of 13 fish that exhibit varying kinetics of CPZ formation and neural crest
72 reactivation with an average onset of 11- and 15-weeks post-fertilization (wpf), respectively (Fig.
73 1b, Extended Data Fig. 1d). Individual *crestin* positive cells subsequently arose from within a
74 cancer precursor zone to form a tumor (Fig. 1a, Extended Data Fig 1b,c, Extended Movie 1). While
75 not all cancer precursor zones reactivate *crestin:EGFP* and form a tumor, once the neural crest
76 program is reactivated, every fish will inevitably go on to form a melanoma (Fig. 1b). We
77 correlated our imaging with histologic analysis of mitfa and crestin across melanoma initiation
78 stages, confirming that individual *crestin* positive cells give rise to tumors (Fig. 1c). We confirmed
79 cancer precursor zone formation around *crestin* patches in two additional zebrafish melanoma
80 models without p53 inactivation, *casper;mcr:BRAF^{V600E}* fish and *casper;mcr:NRAS^{Q61R}* zebrafish
81 suggesting CPZs form in multiple melanoma genotypes and are not dependent on p53 loss
82 (Extended Data Fig. 1e, 1f).

83

84 **Cancer precursor zones have aberrant morphology**

85 To characterize how cancer precursor zone cells lead to neural crest reactivation, we
86 evaluated the anatomical location of CPZs. In line with previous studies, CPZs arise across the
87 entire fish body with more frequent localizations in the head, dorsal fin, and tail base[21] (Extended
88 Data Fig. 2a). As anatomic position did not play major a role in CPZ formation, we next analyzed
89 if CPZ size or if *mitfa:mCherry* intensity drove neural crest reactivation. CPZs exhibited a range
90 of sizes (0.54±0.27mm^2) and *mitfa:mCherry* intensities (105.6±27.4 a.u.) prior to neural crest
91 reactivation (Extended Data Fig. 2b,c). These data indicate neural crest reactivation is a
92 consequence of cell morphology and/or cell state rather than anatomic locations or physical
93 features of the zone. To address this, we imaged *BRAF;p53* cancerized field melanocytes and CPZ
94 melanocytes for morphology. Despite the presence of oncogenic mutations, the cancerized field
95 melanocytes appear morphologically normal, with proper dendrite formation and spacing
96 throughout the skin (Extended Data Fig. 3a, left). However, upon CPZ formation, the melanocytes
97 exhibit aberrant morphology with loss of dendrites (Extended Data Fig. 3a, right). Histologic
98 analysis shows that cancer precursor zone melanocytes encompassing *crestin:EGFP*+ cells are
99 highly proliferative with dysplastic nuclei (Extended Data Fig. 3b,c). Our zebrafish model reveals
100 a novel intermediate step between cancerized field melanocytes and tumor initiation, where
101 morphologically abnormal cancer precursor zone melanocytes encompass the cell undergoing
102 neural crest progenitor state activation as it transforms.
103

104 **Neural crest reactivation upon transplant**

105 A hallmark of malignancy is the ability of a tumor cell to engraft upon transplantation. To
106 determine if cancer precursor zones have malignant potential, we FACS isolated normal
107 melanocytes, cancerized field melanocytes, cancer precursor zone cells, small *crestin*+ patch cells,

108 or tumors (Extended Fig. 4a). These cells are enrobed in Matrigel and injected under the skin of
109 irradiated recipient fish; a limit dilution assay was performed at 14 days post-transplant. Cancer
110 precursor zone cells engrafted at about half the rate of *crestin*+ cells from small patches or tumors
111 while normal or cancerized field melanocytes did not engraft (Fig. 1d, Extended Data Fig. 4b).
112 Small *crestin* patch cells engrafted at a similar rate to tumor cells, suggesting that neural crest
113 reactivated cells have similar malignant potential. Intriguingly, we found that engrafted cells from
114 cancer precursor zones subsequently expressed *crestin:EGFP*, even though FACS isolated cancer
115 precursor zone cells were *crestin:EGFP*- upon transplantation (Extended Data Fig. 4c). These data
116 reveal that cancer precursor zone cells have malignant potential due to their ability to reactivate
117 the neural crest state and engraft but have a significantly decreased tumor initiating potential.
118

119 **Melanoma initiation atlas**

120 *BRAF;p53* zebrafish only develop one to three melanomas over their lifetime with minimal
121 additional genetic mutations[15], indicating that transcriptional or epigenetic changes are key drivers
122 of melanoma initiation. To explore this, we performed single-cell RNA-seq (scRNA-seq). Cells
123 isolated from *BRAF;p53* zebrafish skin containing either morphologically normal control
124 cancerized field melanocytes or small *crestin* patches with surrounding cancer precursor zones
125 were run through the inDrop scRNA-seq platform, doublets removed, then partitioned into 18
126 clusters and annotated using known markers (Fig 2a, Extended Data Fig. 5a-c). Epithelial and
127 immune gene signatures largely overlap between the control and *crestin*+ sample; however, the
128 melanocyte clusters differentially segregate into cancerized field melanocytes, cancer precursor
129 zone, and *crestin*+ melanoma populations. Pseudotime analysis highlights cellular reprogramming
130 from cancerized field melanocytes to crestin+ melanoma with CPZ melanocytes as a transitory

131 population (Fig. 2b, Extended Fig. 5d). We then re-clustered selected melanocytes and assessed
132 differential gene expression. Despite the oncogenic mutations, control cancerized field
133 melanocytes (mel) highly express expected terminally differentiated melanocyte markers, *dct,*
134 *pmela,* and *tryp1b* (Fig. 2c). Cancer precursor zone (CPZ) cells begin to express neural crest
135 markers and *crestin*+ cells no longer express differentiated melanocyte markers and instead
136 resemble embryonic neural crest cells (Fig. 2c, 2d). These analyses reveal that the transformation
137 of a terminally differentiated melanocyte occurs through a transcriptional shift through a cancer
138 precursor zone and towards a neural crest progenitor state.

139 To verify these signatures in melanoma initiation, we performed RNA-seq on FACS
140 isolated *mitfa:mCherry*+ control cancerized field melanocytes and compared them to
141 *mitfa:mCherry*+ cancer precursor zone melanocytes and *mitfa:mCherry+/crestin:EGFP*+ cells
142 isolated from small, medium, and large *crestin* patches in *BRAF;p53* zebrafish. CPZ induction did
143 not cause a global impact on gene expression relative to cancerized field with only 502 genes
144 (2.8% [n=17,483], $\log_2$FC±2, p-value <0.001) altering in expression, indicating this transcriptional
145 switch is specific to progression in melanomagenesis (Extended Data Fig. 6a). Similarly, the
146 transition from cancerized field to patch and tumor led to differential expression of 1,438 (8.2%)
147 and 1,670 (9.6%) genes, respectively (Extended Data Fig. 6b, 6c). Due to these changes in gene
148 expression, PCA analysis shows that the sorted melanocytes cluster according to stage, with the
149 cancer precursor zone melanocytes overlapping both the cancerized field and *crestin* patch clusters
150 (Extended Data Fig. 6d). This suggests that the cancer precursor zone cells represent a transition
151 state between the cancerized field and neural crest reactivated cells.

152

153 **Stage-specific chromatin states define initiation**

154 The distinct gene expression signatures separating melanocytes, cancer precursor zones, and
155 *crestin* positive cells indicates that melanoma initiation occurs through unique transcriptional cell
156 states. Differential expression analysis of both the scRNA-seq and RNA-seq data revealed stage-
157 specific expression signatures (Fig. 2e, Extended Data Fig. 7a). Cancer precursor zone cells have
158 increased expression of BMP signaling (red), as well as neural plate border and pre-migratory
159 neural crest transcription factors (green). These expression profiles were correlated with gene
160 expression at various stages of human melanoma formation from two published datasets (Extended
161 Data Fig. 7b, 7c)[22, 23].

162 To dissect the chromatin landscape, we performed ATAC-seq on sorted melanocytes from
163 various stages of initiation. We filtered called peaks for direct overlap or within a 5000 bp distance
164 of annotated genes for differential accessibility analysis (multiple tests corrected with FDR <0.05).
165 Like RNA-seq, ATAC-seq showed cellular reprogramming of specific loci. Only 999 genes (2.8%
166 [n=35,592], $\log_2$FC±2.0, p-value <0.001) had differentially accessible regions (DAR) in cancer
167 precursor zone melanocytes relative to cancerized field. While neural crest reactivation in patches
168 and tumors led to more chromatin reprogramming (1,117 DAR in patch; 8,612 DAR tumor), all
169 melanoma initiation stages showed focal chromatin changes specific to cancer progression
170 (Extended Data Fig. 8a-c). Promoter and enhancer regions associated with differentially expressed
171 genes contained peaks that opened in a stage-specific manner (Fig. 2f, Extended Data Fig. 9a-g).
172 The chromatin accessibility findings correspond with the gene expression data and further indicate
173 that a transcriptionally driven cancer attractor state is present at the earliest stages of melanoma
174 initiation.

175

176 **Cancerized field lacks MAPK activation**

177 Most human melanomas arise de novo, with only a small fraction of nevi progressing to
178 malignancy[24]. However, 70-90% of benign nevi carry the BRAF^{V600E} mutation and activation of
179 the MAPK pathway does not always directly correlate with BRAF mutations in many
180 melanomas[24-27]. Despite the presence of constitutively active BRAF in our zebrafish model, the
181 expression levels of MAPK pathway members and downstream targets are upregulated exclusively
182 in cancer precursor zone and *crestin*+ cells (Extended Data Fig. 10a). Negative regulators of
183 MAPK, *spry2*, *spry4*, *dusp2*, and *dusp6*, show little to no expression and closed chromatin in the
184 cancerized field melanocytes (Extended Data Fig. 10a, 10b). ATAC-seq analysis on isolated
185 melanocytes revealed that the enhancer regions around MAPK target genes are not open in control
186 cancerized fields but become open in cancer precursor zones (Extended Data Fig. 10b).
187 Furthermore, phospho-ERK immunohistochemistry showed no staining in cancerized field
188 melanocytes, low staining in *mitfa*^{med} cancer precursor zones, and high staining in *mitfa*^{high} cancer
189 precursor zones (Extended Data Fig. 10c). We considered the possibility that the *mitfa*^{high} state
190 increases the amount of BRAF^{V600E} expression due to the *mitfa:BRAF*^{V600E} transgene; however,
191 RNA-seq on isolated melanocytes shows no change in BRAF^{V600E} expression between the
192 cancerized field and cancer precursor zone stages, only increasing at the patch stage (Extended
193 Data Fig. 10d). Together these data indicate that the MAPK pathway is not active until the cancer
194 precursor zone stage and is further activated during tumor initiation.
195

196 **ID1 identifies and drives precursor zones**

197 Given that cancer precursor zones form prior to and have a distinct transcriptional state from
198 *crestin* patches, we hypothesized these states are independently regulated. Since cancer precursor
199 zones had high expression of BMP pathway members and downstream targets (Fig. 2e), we used

200 the MiniCoopR system to overexpress ID1 in melanocytes[16]. ID1 overexpression significantly
201 increased cancer precursor zones with no increase in *crestin* patches at 6 weeks of age (Fig. 3a,
202 3b). We confirmed upregulation of ID1 expression in cancer precursor zones by RNAscope in situ
203 hybridization[28]. ID1 is highly expressed in CPZ melanocytes but not in adjacent cancerized field
204 cells (Extended Data Fig. 11a, 11b). These data indicate ID1 and BMP signaling are strong drivers
205 of cancer precursor zone formation. Unlike CPZs, *crestin*+ patches exhibit high expression of IRF3
206 pathway members and downstream targets (Extended Data Fig. 11c). Indeed, IRF3 overexpression
207 caused a significant increase in the number of *crestin* patches without a change in cancer precursor
208 zone formation at 6 weeks post-fertilization (Fig. 3a, 3b). While ID1 or IRF3 overexpression did
209 not cause an earlier incidence of tumor formation, both led to a significant increase in the number
210 of tumors per fish (Fig. 3c, 3d). Overall, these results provide evidence that the stages of melanoma
211 initiation can be independently regulated, and perturbation of these stages lead to an increase in
212 tumor formation.

213 To verify the role of ID1 and IRF3 in human melanoma, we performed
214 immunofluorescence staining on patient samples (Fig. 3e). The percent of melanocytes expressing
215 ID1 and the intensity of ID1 staining was significantly increased in patient samples with atypical
216 melanocyte proliferative (potentially precursor) zones, melanoma in situ, or invasive melanoma
217 compared to those in normal skin (Fig. 3f, Extended Data Fig. 11d). Interestingly, ID1+ cells were
218 frequently found in the surrounding epithelial borders of invasive melanoma. In addition to the
219 ability of ID1 to initiate cancer precursor zones in the zebrafish, the appearance of ID1 in over
220 80% of atypical melanocytic hyperplasia fields indicates that ID1 is a major driver of cancer
221 precursor zones. Additionally, the percent of melanocytes with nuclear IRF3 staining increased at
222 both early and late stages of initiation, with a striking increase in staining in invasive melanoma

223 samples (Fig. 4e, 4f). These data validate the role for ID1 and IRF3 in melanomagenesis and
224 identify clinically covert ID1+ precursor melanocyte zones in humans.
225
226 **ID1 inhibits TCF12 to drive initiation**
227 ID1 is overexpressed in over 20 types of cancer, including melanoma, where strong ID1 expression
228 was significantly associated with increased tumor thickness and decreased patient survival[29, 30].
229 ID1 is a transcriptional repressor that prevents basic helix-loop-helix (bHLH) transcription factors
230 from binding to DNA[31]. We performed IP-MS on A375 melanoma cells overexpressing V5-tagged
231 ID1 or Clover control. The bHLH factor TCF12 was the most significantly pulled down protein
232 with ID1 (Fig. 4a, Extended Data Fig. 12a). When performing ATAC-seq on sorted melanocytes
233 isolated from ID1-overexpressing zebrafish, we found significant enrichment of the TCF12 motif
234 (ACAGCTG) under chromatin peaks that decreased with ID1 overexpression, indicating ID1 is a
235 repressor of TCF12 signaling (HOMER, p=1e-20). Predicted downstream targets of TCF12 are
236 significantly decreased in cancer precursor zones where ID1 is most highly expressed (Extended
237 Data Fig. 12b)[32]. Next, we performed RNA-seq on sorted melanocytes from zebrafish
238 overexpressing ID1 and identified potential downstream targets in the context of melanoma
239 initiation that were downregulated either with ID1-overexpression or in cancer precursor zones
240 (Extended Data Fig. 12c-d). TCF12 motifs were found under open chromatin peaks in the
241 promoter/enhancer regions associated with these genes (Fig. 4b). Strikingly, many of these genes
242 are important in the maintenance of dendritic processes and cell adhesion, such as *rhogb*, *parvg*,
243 *lamc1*, and *enah*. During cancer precursor zone formation, the zebrafish melanocytes attain an
244 aberrant morphology, including a loss of dendrites, which is likely achieved via a decrease in these
245 targets. When performing gene ontology analysis on downregulated genes containing TCF12

246 motifs, pathways arose such as neural crest cell migration (-6.002 LogP) and pigment cell
247 differentiation (-4.547 LogP), including genes such as *ednrba*, *slc45a2*, and *mlpha*.

248 To confirm these findings and identify targets of TCF12, we performed chromatin
249 immunoprecipitation following V5-tagged ID1 or Clover overexpression in immortalized human
250 primary melanocytes (PMEL[33]) harboring $BRAF^{V600E}$. ID1 overexpression led to a marked
251 reduction in TCF12 occupancy on chromatin and loss of TCF12 motif enrichment relative to
252 control (Fig. 4c, 4d). This marked decrease in TCF12 occupancy focally affected neuronal genes
253 and melanocyte development genes. In an unbiased analysis, gene ontology analysis of peaks
254 within genes with annotated TCF12 motifs highlighted an enrichment of neuronal genes, including
255 cell body, synapse regulation, and somatodendritic compartment regulation in PMEL cells
256 (Extended Data Fig. 12e). Specifically, ID1 overexpression led to loss of TCF12 occupancy at
257 known neuronal genes, and tumor suppressor genes, DLGAP4 and UNC5A. High DLGAP4
258 expression is a known positive prognostic biomarker in melanoma and regulates synaptic signaling
259 [34-36]. Further, UNC5A is a netrin receptor that controls neurite outgrowth in the developing nervous
260 system but loss of UNC5A has established roles in malignant transforming mechanisms such as
261 anchorage independent growth and escape of apoptosis in tissues outside the nervous system[37, 38].
262 In addition to known markers of the neural lineage, TCF12 occupancy was lost at melanocyte
263 markers, such as PKNOX2 and PRKCZ. PKNOX2 is an identified tumor suppressor gene in solid
264 tumors, such as gastric cancer, with known roles in neural crest differentiation[39, 40]. Further,
265 PKNOX2 has been shown to be uniquely enriched in human melanocytes relative to other skin
266 cells and functions in melanin production[41]. Similarly, PRKCZ, a member of the protein kinase C
267 family, drives melanin production in melanocytes[42]. As such, loss of PRKCZ highlights that CPZs
268 may represent unpigmented clinically covert precursor lesions.

269 To explore how modulation of the ID1/TCF12 axis affected cell viability in primary
270 melanocytes and transformed melanoma cells, we measured cell proliferation in human PMEL and
271 A375 melanoma cells. Overexpression of ID1 in PMEL cells led to a significant increase in cell
272 proliferation relative to Clover control, likely due to the loss of development and cell cycle
273 regulators (Fig. 4e, 4f; $p<0.01$). Highlighting the importance of TCF12 loss in driving the
274 malignant program, we observed a significant loss of cell proliferation and viability in A375
275 human melanoma cells following TCF12 overexpression relative to Clover control (Fig. 4g,
276 $p<0.001$). These data suggest restoring TCF12 activity can reverse malignant potential in both
277 dysplastic melanocytes and transformed malignant cells via ID1 or TCF12 overexpression,
278 respectively. Finally, we overexpressed TCF12 in zebrafish using the MiniCoopR system and
279 found a significant decrease in the number of CPZs formed and tumor onset compared to control
280 (Fig. 4h, 4i). In total, these data support the hypothesis that ID1 binds to and inhibits TCF12,
281 preventing it from activating downstream targets thereby enabling morphological and
282 transcriptional changes to form cancer precursor zones.
283

284 **Discussion**

285 Our studies support a model in which a cancerized field of morphologically normal melanocytes
286 expand, become dysplastic, and reactivate a neural crest progenitor program, culminating in
287 tumorigenesis. The data presented here add new insights to our understanding of melanomagenesis
288 by visualizing the earliest stages in initiation, revealing that a group of melanocytes undergo
289 independent, transcriptionally driven cell state changes before one cell becomes malignant. These
290 findings are in line with cell of origin studies that show the transformation of interfollicular and
291 differentiated melanocytes in the mouse tail[43]. Gene expression analysis and immunofluorescence

292 staining show that cancer precursor zones in zebrafish resemble atypical melanocytic proliferative
293 lesions that constitute presumed precursor fields that often surround melanomas in patient samples.
294 The strong correlation between species suggests that melanoma initiates through transcriptionally
295 driven cell state changes, with ID1 inhibition of TCF12 and IRF3 as drivers of this process. We
296 demonstrate that ID1 represses downstream targets of TCF12, which are key to the maintenance
297 differentiated melanocytes and cell cycle control.

298 The field cancerization theory posits that pro-tumorigenic mutations increase the malignant
299 transformation potential of cells. Potential precursor zones of frank malignancy have been
300 identified in several different organs, including Barrett's esophagus, ductal carcinoma in situ,
301 prostatic intraepithelial neoplasia, the uterine cervix, and dysplastic lesions in melanoma[44-47]. The
302 distinct transcriptional and chromatin states we identified during melanoma initiation are
303 reminiscent of the cancer attractor state theory, which posits that cancer cells sit in a low energy
304 basin of attraction that drives and reinforces epigenetic change[10, 11, 48]. The tight, distinct clustering
305 of *crestin*+ melanocytes in our single cell RNA-seq analysis supports the cancer attractor state
306 theory. The finding that a cancer precursor zone encompasses the initiating cell of melanoma
307 suggests that other cancers may also initiate from within a surrounding group of precursor cells
308 and drive towards a cancer attractor state.

309

310 **Author Contributions**

311 A.M.M., M.H.C., and L.I.Z. conceived of the study; L.I.Z. and G.F.M. supervised the research;
312 A.M.M. and M.H.C. designed and performed experiments with the assistance of J.B., E.W., I.F.G.,
313 E.L., C.B., and P.A.; H.R.N. performed ID1 overexpressing RNA-seq experiment and advised on
314 experimental design for all studies; J.K.M. assisted with histology and pathologic assessment of

315 fish and human samples; M.P. and S.Y. performed computational analysis for all sequencing
316 experiments; C.G.L. provided human samples and pathologic analysis.
317

318 **Acknowledgements**

319 We would like to thank Christian Lawrence, Shane Hurley, Li-Kun Zhang, Kara Maloney, and
320 Hannah Nations for expert fish care. The authors thank many of their colleagues for their
321 discussions and critical reading of the manuscript. We thank Dana-Farber/Harvard Cancer Center
322 for the use of the Specialized Histopathology Core, which provided histology and
323 immunohistochemistry service. The authors would like to thank the Neurobiology Imaging
324 Facility (NIF) at Harvard Medical School for performing the RNAscope. Dana-Farber/Harvard
325 Cancer Center is supported in part by an NCI Cancer Center Support Grant #P30CA06516. We
326 thank Ross Tomaino and the Harvard Medical School Taplin Mass Spectrometry Facility for their
327 assistance in performing and analyzing our mass spectrometry experiments. This work was
328 supported by R01CA103846-14, American Cancer Society Postdoctoral Fellowship PF-18-150-
329 01-DCC (A.M.M.), and K00CA253749 (M.H.C.). L.I.Z. is a Howard Hughes Medical Institute
330 Investigator.
331
332

333 **Methods**

334 *Generation of BRAF;p53 zebrafish*

335 *mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$;crestin:EGFP;mcr:Empty;mitfa:mCherry;tyr$^{-/-}$* zebrafish (referred
336 to as *BRAF;p53*) were generated by injecting *mcr:Empty* and *mitfa:mCherry* at 25 ng/μL, 5
337 mg/mL Cas9 protein (PNA Bio CP02), a gRNA targeting *tyr* (GGACTGGAGGACTTCTGGGG)

338 at 50 ng/μL, and Tol2 mRNA at 20 ng/μl into the single cell stage of *mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$;crestin:EGFP* embryos[49]. These fish were originally developed to be used in combination with
340 MiniCoopR (mcr), which is a transgenic tool that permits selective mis-expression of any gene in
341 melanocytes, thereby allowing us to assess the ability of each gene to promote or inhibit melanoma
342 initiation[16]. Injection of the MiniCoopR vector mosaically rescues melanocyte development via an
343 *mitfa* minigene, allowing melanomas to form. The MiniCoopR vector also contains the open
344 reading frame of any candidate gene driven specifically in melanocytes by the *mitfa* promoter,
345 which enables us to assess its oncogenic potential. For the morphological and molecular
346 characterization experiments, a MiniCoopR vector containing only the *mitfa* minigene was
347 injected to rescue melanocyte formation (mcr:Empty).
348 *mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$;crestin:EGFP* zebrafish were generated by crossing
349 *mitfa:BRAF^{V600E};p53$^{-/-}$;crestin:EGFP*[3] zebrafish with *mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$* zebrafish.
350 *mitfa:mCherry* was cloned by Gateway reaction using the zebrafish *mitfa* promoter and the Tol2
351 mCherry middle entry plasmids. The *tyrosinase* gRNA was synthesized using SP6 in vitro
352 transcription[17, 50]. In brief, the *tyr* oligo template (CCTCCATACGATTTAGGTGACACT
353 ATAGGACTGGAGGACTTCTGGGGGTTTTAGAGCTAGAAATAGCAAG) and the constant
354 oligonucleotide (AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGC
355 CTTATTTTAACTTGCTATTTCTAGCTCTAAAAC) were annealed and filled in with T4 DNA
356 polymerase (New England BioLabs, M0203S). The product was PCR amplified, gel purified, and
357 transcribed using MEGAscript T7/SP6 (ThermoFisher Scientific, AM1333). gRNAs were cleaned
358 up using Direct-zol RNA Miniprep kit (Zymo Research, R2051).

359 To generate the *casper;mcr:NRAS*^{Q61R} or *casper;mcr:BRAF*^{V600E} zebrafish, *mcr:NRAS*^{Q61R} or
360 *mcr:BRAF*^{V600E} was injected into *casper;crestin:EGFP* embryos along with Tol2 mRNA and
361 *mitfa:mCherry* at a concentration of 25 ng/µL.
362 This study was performed in strict accordance with the recommendations in the Guide for the Care
363 and Use of Laboratory Animals of the National Institutes of Health. The animal research protocol
364 was approved by the Institutional Animal Care and Use Committee of Boston Children's Hospital.
365 All zebrafish used in this study were maintained and euthanized under the guidelines of the
366 Institutional Animal Care and Use Committee of Boston Children's Hospital.
367
368 *Histology*
369 Fish were euthanized and fixed in 4% paraformaldehyde overnight at 4°C. Paraffin embedding,
370 sectioning, Hematoxylin and Eosin (H&E) staining were performed according to standard
371 techniques by the Brigham & Women's Hospital Pathology Core. Immunohistochemistry was
372 performed with 5 µm thick formalin-fixed, paraffin-embedded tissue sections using the Leica
373 Bond III automated staining platform and the Leica Biosystems Refine Detection Kit. Antibody
374 mCherry from Thermo Fisher, catalog number M11217, clone 16D7, was run at 1:200 dilution
375 with EDTA antigen retrieval and Goat Anti-Rat IgG secondary antibody from Vector Labs catalog
376 number PI-9401-.5. Antibody GFP from Abcam, catalog number 6556, polyclonal, was run at
377 1:800 dilution with EDTA antigen retrieval. Antibody PCNA from Cell Signaling Technology,
378 catalog number 2586, clone PC(10), was run at 1:16000 dilution with citrate antigen retrieval.
379 Antibody Phospho ERK (p44/42 MAPK), clone D13.14.4E, from Cell Signaling Technologies,
380 catalog number 4370, was run at 1:150 dilution with citrate antigen retrieval.

381 All human tissue samples were derived from the Pathology Archives of the Brigham and Women's
382 Hospital with full institutional review board approval. Patient consent for experiments was not
383 required because de-identified pathological specimens of human samples are discarded material
384 by our institution, and thus the studies were exempt. Immunofluorescence studies were performed
385 on paraffin-embedded sections of formalin-fixed tissue. Tissue sections cut at 5 µm intervals were
386 deparaffinized, rehydrated, and heated with Target Antigen Retrieval Solution (Dako, Agilent
387 Technologies) in a pressure cooker. Sections were blocked with 10% animal serum for 30 minutes
388 before incubation with primary antibodies. Samples were incubated overnight at 4°C with the
389 following primary antibodies: mouse IgG1 anti-IRF3 (1:100; SCBT, sc-33641), mouse IgG2a anti-
390 ID1 (1:500; SCBT, sc133104), mouse IgG2b anti-MART1 (1:1, Biolegend, 917902). Sections
391 were then treated with 0.1% Sudan Black (Abcam) for 10 minutes to remove autofluorescence.
392 The following secondary antibodies were used: goat anti-mouse IgG2b AF488 (1:2000;
393 Invitrogen, A-21141), goat anti-mouse IgG2a AF594 (1:2000; Invitrogen, A-21135), goat anti-
394 mouse IgG1 AF647 (1:2000; Invitrogen, A-21235). Slides were mounted with Floromount G with
395 NucBlue.
396
397 *In Situ Hybridization (RNAscope)*
398 RNAscope Multiplex Fluorescent Assay[28] (Biotechne) was performed on formalin-fixed paraffin
399 embedded cancerized field, cancer precursor zone, patch, and tumor sections. Melanoma stages
400 were assigned by H&E and mCherry or GFP expression (IHC methods described above). Protocol
401 was followed according to manufacturer's instructions except protease 3 was used instead of
402 protease 4. Each sample was hybridized with zebrafish id1 (#517531, C1 probe, red) and zebrafish
403 crestin (#534061, C2 probe, far red) alongside a DAPI stain to mark nuclei. RNAscope was

404 provided by the Neurobiology Imaging Facility (NIF) at Harvard Medical School. Stained slides
405 were imaged with a 40x objective on a Nikon Eclipse Ti-2 spinning disk confocal microscope. All
406 images were acquired using NIS-Elements (Nikon) and minimally processed using Imaris.
407
408 *Imaging*
409 Zebrafish were anesthetized with 4% MS-222 (Pentair, TRS1) and imaged on a Nikon SMZ18
410 Stereomicroscope or a Nikon C2si Laser Scanning Confocal using a 10x objective. Histologic
411 sections were imaged on a Nikon Eclipse Ti-2 Spinning Disk Confocal using a 40x or 100x
412 objective. Maximum intensity projections of Z stacks or three-dimensional re-constructions are
413 presented here. Images were minimally processed using Photoshop, FIJI, or Imaris. Multiple tiled
414 images of adult zebrafish and histology were stitched together by using the automated Photomerge
415 function in Photoshop. Given the variability in the rescue specific to each animal, number of
416 zones/patches per fish were weighted for *mitfa:mCherry* rescue.
417
418 *FACS Cell Isolation*
419 *mitfa:BRAF^{V600E};p53$^{-/-}$;mitfa$^{-/-}$;crestin:EGFP;mcr:Empty;mitfa:mCherry;tyr$^{-/-}$* zebrafish from the
420 same cohort were categorized by stage: control cancerized field, cancer precursor zone, small
421 *crestin* patch, medium *crestin* patch, and large *crestin* patch/tumor. The zones and patches were
422 manually dissected and individually chopped for 1-2 minutes. The finely chopped tissue was
423 suspended in 3 mLs of TrypLE Express (ThermoFisher Scientific, 12605028) and incubated at
424 37°C shaking at 300 rpm for 30 minutes. Samples were filtered through a 40 µm filter and washed
425 with 5 mL of FACS buffer, consisting of DPBS (ThermoFisher Scientific, 14190144), 10x
426 Penicillin Streptomycin (ThermoFisher Scientific, 15140122), and 2% heat-inactivated FBS

427 (ThermoFisher Scientific, A3840001). Samples were centrifuged at 500 rcf for 5 minutes and
428 resuspended in 150 µL FACS buffer. Samples were stained with SYTOX blue (ThermoFisher,
429 S34857) immediately before sorting for live (SYTOX blue negative), mCherry and/or EGFP
430 positive cells on a FACSAria II (BD Biosciences). FACS cell isolation was used for bulk and
431 single cell RNA-seq, ATAC-seq, and transplantation experiments.

432
433
434 *scRNA-seq*
435 Cells isolated from the skin of 6 control fish and 4 fish with small *crestin* patches were FACS
436 sorted for *mitfa:mCherry*, as described above. 1360 melanocytes were combined with 29,000
437 mCherry negative skin cells in the control and 6224 melanocytes were combined with 24,000
438 mCherry negative skin cells in the *crestin* patch sample to yield 2 samples each with 30,000 cells.
439 These samples were run through the inDrops platform by the Single Cell Core at Harvard Medical
440 School, as previously described[53, 54]. Libraries were run on an Illumina HiSeq 4000 with paired
441 end 150bp. inDrops single-cell RNA-seq analysis follows the instruction as described in
442 https://github.com/indrops/indrops. inDrops Library v3 requires manual demultiplex raw bcl into
443 different samples. Zebrafish Bowtie transcriptome index was built based on Ensembl GRCz11
444 genome sequence and gene annotation[55]. The output data matrix contains the raw count of each
445 gene for each cell barcode. Analysis was performed using the Seurat and monocle 3 packages in
446 R[56-58].

447

448 *RNA-seq*

449 RNA was extracted from 5,000 sorted melanocytes from 3-4 control or ID1-overexpressing fish
450 per stage using RNeasy Micro Kit (Qiagen, 74004). Ultralow input RNA-seq was performed using
451 the SMART-Seq v4 Ultra Low Input RNA kit for Sequencing (Clontech, 634888) and Nextera XT

452 DNA Library Preparation Guide (Illumina, FC-131-1024). Libraries were run on an Illumina
453 HiSeq 4000 with paired end 150bp. Quality control of RNA-Seq datasets was performed by
454 FastQC and Cutadapt to remove adaptor sequences and low-quality regions[59, 60]. The high-quality
455 reads were aligned to Ensembl build GRCz11 of zebrafish genome using Tophat 2.0.11 without
456 novel splicing form calls[61]. Transcript abundance and differential expression were calculated with
457 Cufflinks 2.2.1[62]. Differential expression analysis was performed using DESeq2 in R[46]. Pathway
458 analysis was performed using Metascape and graphed using GraphPad Prism 7 (GraphPad
459 Software)[63].
460
461 *ATAC-seq*
462 5,000 sorted melanocytes from 3-4 control or ID1-overexpressing fish at each stage were lysed
463 and subjected to "tagmentation" reaction and library construction as previously described[64].
464 Libraries were run on an Illumina HiSeq 4000 with paired end 150bp. All zebrafish ATAC-Seq
465 datasets were aligned to build version Ensembl build GRCz11of the zebrafish genome using
466 Bowtie2 (version 2.2.1) with the following parameters: --end-to-end, -N0, -L20[55]. We used the
467 MACS2 (version 2.1.0) peak finding algorithm to identify regions of ATAC-Seq peaks and derive
468 the normalized tracks, with the following parameter --nomodel --shift -100 --extsize 200 --
469 SPMR[65]. A q-value threshold of enrichment of 0.05 was used for all datasets. The motifs enriched
470 in ATAC-seq peaks of interest were analyzed using the findMotifsGenome program in the
471 HOMER package[66]. The corresponding zebrafish genome sequences are used as background
472 sequences in motif search. The top known HOMER motifs and de novo motifs with q-value less
473 than *** are calculated. HOMER motif results are presented as a heatmap of the percent of targets
474 containing the indicated motif minus background. For example, 41.8% of new or increased target

475 peaks in large *crestin*+ samples contain a RUNX motif, as compared to 20.7% of background
476 peaks. Therefore, the heatmap indicates 21.1% of target peaks are enriched for a RUNX motif.
477
478 *Transplantation*
479 FACS isolated *mitfa:mCherry*+/*crestin:EGFP*- cancer precursor zone cells or
480 *mitfa:mCherry*+/*crestin:EGFP*+ cells from small *crestin* patches or tumors were enrobed in 3 μL
481 Matrigel (Corning, 356234) and injected at concentrations of 5,000, 3,000, 1,000, or 500 cells
482 under the skin of *casper* recipient fish irradiated sub-lethally with 30 Gy split over 2 days. Isolated
483 cells from each donor went into 3-7 recipient fish. Fish were imaged as described above and limit
484 dilution calculations and curves were generated using the Extreme Limiting Dilution Analysis
485 software (http://bioinf.wehi.edu.au/software/elda/)[49] on data collected 14 days post-transplant.
486
487 *IP-MS*
488 The fusion (no stop) human ID1 or Clover open reading frames were cloned into the pcDNA3.2
489 V5 destination vector (Invitrogen, 12489019) to create V5 tagged ID1 or Clover constructs.
490 Constructs were transiently transfected into A375 human melanoma cells using
491 Lipofectamine®3000 (Invitrogen, L3000001) in 10 cm^2 plates with three independent replicates.
492 48 hours after transfection, cytoplasmic and nuclear fractions were isolated using the NE-PER™
493 Nuclear and Cytoplasmic Extraction Kit (ThermoFisher, 78833) and lysed per protocol. Anti-V5
494 (Clone V5-10, Sigma, V8012) was conjugated to protein G beads (ThermoFisher, 10004D). IP'd
495 proteins were eluted and submitted for mass spectrometry using the Taplin Mass Spectrometry
496 Facility at Harvard University. Proteins were included in analysis if all replicates had greater than
497 3 peptides.

498

499 *ChIP-Seq*

500 PMEL/hTERT/CDK4(R24C)/p53DD/BRAF(V600E) cells (courtesy of D. Fisher lab[33]) were
501 transiently transfected with V5-Clover or V5-ID1, as detailed above, in 15 cm^2 plates with three
502 independent replicates. Cells were crosslinked in 11% formaldehyde, lysed, and sheared as
503 described[67]. Solubilized chromatin was immunoprecipitated with 10 µg TCF12 antibody (SCBT,
504 sc357). Antibody-chromatin complexes were purified and libraries were prepared as described[67].
505 All ChIP-Seq datasets were aligned to Ensembl build version GRCh38 of the human genome using
506 Bowtie2 (version 2.2.1) with the following parameters: --end-to-end, -N0, -L2086[55]. MACS2
507 (version 2.1.0) peak finding algorithm was used to identify regions of ChIP-Seq peaks, with a q-
508 value threshold of enrichment of 0.05 for all datasets[65]. ChipSeeker is utilized to annotate ChIP-
509 seq peaks to neighboring genes according to Ensembl gene annotation[68]. The parameters are
510 defined as proximal promoter: 500 bp upstream – 50 bp downstream of TSS; distal promoter: 2 k
511 bp upstream – 500 bp downstream of TSS; enhancer: 100 k bp from TSS. The genome-wide
512 occupancy profile figures were generated by deeptools2 with peaks centered[69]. HOMER analysis
513 was performed to confirm transcription factor binding under peaks with hg38 genome utilized as
514 a random set of background peaks for motif enrichment[66]. To count TCF12 motifs enriched in
515 called peaks, mapped TCF12 motifs across the hg38 genome from ENCODE enrichment was
516 used[70]. Reactome pathway analysis was completed with Gene Ontology[71, 72].

517

518 *Cell Viability*

519 The fusion human TCF12 open reading frame was cloned into pcDNA3.2 V5 destination vector
520 to create V5 tagged TCF12. A375 or PMEL cells were seeded in opaque-walled 96-well plates.

521 5,000 cells were transiently transfected with V5-ID1 (A375), V5-TCF12 (PMEL) or a V5-clover
522 control (A375 and PMEL) for 48 hours. Cell viability was then measured using CellTiter-Glo
523 (Promega #G7570) and luminescence was read on Synergy Neo plate reader. Experiments were
524 performed in technical triplicate for each biological replicate (n=3).
525
526 *Statistics*
527 A two-way ANOVA with Tukey's multiple comparisons test was used to compare the weighted
528 number of cancer precursor zones or *crestin* patches and the immunofluorescence quantitation.
529 The calculations were performed using GraphPad Prism 7 (GraphPad Software).

530 1. Curtius K, Wright NA, Graham TA. An evolutionary perspective on field cancerization.
531 Nat Rev Cancer. 2018;18(1):19-32. Epub 2017/12/09. doi: 10.1038/nrc.2017.102. PubMed
532 PMID: 29217838.
533 2. Slaughter DP, Southwick HW, Smejkal W. Field cancerization in oral stratified
534 squamous epithelium; clinical implications of multicentric origin. Cancer. 1953;6(5):963-8. Epub
535 1953/09/01. doi: 10.1002/1097-0142(195309)6:5<963::aid-cncr2820060515>3.0.co;2-q.
536 PubMed PMID: 13094644.
537 3. Kaufman CK, Mosimann C, Fan ZP, Yang S, Thomas AJ, Ablain J, Tan JL, Fogley RD,
538 van Rooijen E, Hagedorn EJ, Ciarlo C, White RM, Matos DA, Puller AC, Santoriello C, Liao
539 EC, Young RA, Zon LI. A zebrafish melanoma model reveals emergence of neural crest identity
540 during melanoma initiation. Science. 2016;351(6272):aad2197. Epub 2016/01/30. doi:
541 10.1126/science.aad2197. PubMed PMID: 26823433; PMCID: PMC4868069.
542 4. White RM, Cech J, Ratanasirintrawoot S, Lin CY, Rahl PB, Burke CJ, Langdon E,
543 Tomlinson ML, Mosher J, Kaufman C, Chen F, Long HK, Kramer M, Datta S, Neuberg D,
544 Granter S, Young RA, Morrison S, Wheeler GN, Zon LI. DHODH modulates transcriptional
545 elongation in the neural crest and melanoma. Nature. 2011;471(7339):518-22. Epub 2011/03/25.
546 doi: 10.1038/nature09882. PubMed PMID: 21430780; PMCID: PMC3759979.
547 5. McConnell AM, Mito JK, Ablain J, Dang M, Formichella L, Fisher DE, Zon LI. Neural
548 crest state activation in NRAS driven melanoma, but not in NRAS-driven melanocyte expansion.
549 Dev Biol. 2019;449(2):107-14. Epub 2018/06/09. doi: 10.1016/j.ydbio.2018.05.026. PubMed
550 PMID: 29883661; PMCID: PMC6281797.
551 6. Elmore JG, Barnhill RL, Elder DE, Longton GM, Pepe MS, Reisch LM, Carney PA,
552 Titus LJ, Nelson HD, Onega T, Tosteson ANA, Weinstock MA, Knezevich SR, Piepkorn MW.

553 Pathologists' diagnosis of invasive melanoma and melanocytic proliferations: observer accuracy
554 and reproducibility study. Bmj. 2017;357:j2813. Epub 2017/07/01. doi: 10.1136/bmj.j2813.
555 PubMed PMID: 28659278; PMCID: PMC5485913.
556 7. Elmore JG, Elder DE, Barnhill RL, Knezevich SR, Longton GM, Titus LJ, Weinstock
557 MA, Pepe MS, Nelson HD, Reisch LM, Radick AC, Piepkorn MW. Concordance and
558 Reproducibility of Melanoma Staging According to the 7th vs 8th Edition of the AJCC Cancer
559 Staging Manual. JAMA Netw Open. 2018;1(1). Epub 2018/12/18. doi:
560 10.1001/jamanetworkopen.2018.0083. PubMed PMID: 30556054; PMCID: PMC6294444.
561 8. Shoo BA, Sagebiel RW, Kashani-Sabet M. Discordance in the histopathologic diagnosis
562 of melanoma at a melanoma referral center. J Am Acad Dermatol. 2010;62(5):751-6. Epub
563 2010/03/23. doi: 10.1016/j.jaad.2009.09.043. PubMed PMID: 20303612.
564 9. Davis LE, Shalin SC, Tackett AJ. Current state of melanoma diagnosis and treatment.
565 Cancer Biol Ther. 2019;20(11):1366-79. Epub 2019/08/02. doi:
566 10.1080/15384047.2019.1640032. PubMed PMID: 31366280; PMCID: PMC6804807.
567 10. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a
568 gene network dynamics and developmental perspective. Semin Cell Dev Biol. 2009;20(7):869-
569 76. Epub 2009/07/15. doi: 10.1016/j.semcdb.2009.07.003. PubMed PMID: 19595782; PMCID:
570 PMC2754594.
571 11. Kauffman S. Differentiation of malignant to benign cells. J Theor Biol. 1971;31(3):429-
572 51. Epub 1971/06/01. doi: 10.1016/0022-5193(71)90020-8. PubMed PMID: 5556142.
573 12. Li Q, Wennborg A, Aurell E, Dekel E, Zou JZ, Xu Y, Huang S, Ernberg I. Dynamics
574 inside the cancer cell attractor reveal cell heterogeneity, limits of stability, and escape. Proc Natl

575 Acad Sci U S A. 2016;113(10):2672-7. Epub 20160229. doi: 10.1073/pnas.1519210113.
576 PubMed PMID: 26929366; PMCID: PMC4790994.
577 13. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E,
578 Auclair D, Li L, Place C, Dicara D, Ramos AH, Lawrence MS, Cibulskis K, Sivachenko A, Voet
579 D, Saksena G, Stransky N, Onofrio RC, Winckler W, Ardlie K, Wagle N, Wargo J, Chong K,
580 Morton DL, Stemke-Hale K, Chen G, Noble M, Meyerson M, Ladbury JE, Davies MA,
581 Gershenwald JE, Wagner SN, Hoon DS, Schadendorf D, Lander ES, Gabriel SB, Getz G,
582 Garraway LA, Chin L. A landscape of driver mutations in melanoma. Cell. 2012;150(2):251-63.
583 Epub 2012/07/24. doi: 10.1016/j.cell.2012.06.024. PubMed PMID: 22817889; PMCID:
584 PMC3600117.
585 14. Regad T. Molecular and cellular pathogenesis of melanoma initiation and progression.
586 Cell Mol Life Sci. 2013;70(21):4055-65. Epub 2013/03/28. doi: 10.1007/s00018-013-1324-2.
587 PubMed PMID: 23532409.
588 15. Yen J, White RM, Wedge DC, Van Loo P, de Ridder J, Capper A, Richardson J, Jones D,
589 Raine K, Watson IR, Wu CJ, Cheng J, Martincorena I, Nik-Zainal S, Mudie L, Moreau Y,
590 Marshall J, Ramakrishna M, Tarpey P, Shlien A, Whitmore I, Gamble S, Latimer C, Langdon E,
591 Kaufman C, Dovey M, Taylor A, Menzies A, McLaren S, O'Meara S, Butler A, Teague J, Lister
592 J, Chin L, Campbell P, Adams DJ, Zon LI, Patton EE, Stemple DL, Futreal PA. The genetic
593 heterogeneity and mutational burden of engineered melanomas in zebrafish models. Genome
594 Biol. 2013;14(10):R113. Epub 2013/10/24. doi: 10.1186/gb-2013-14-10-r113. PubMed PMID:
595 24148783; PMCID: PMC3983654.
596 16. Ceol CJ, Houvras Y, Jane-Valbuena J, Bilodeau S, Orlando DA, Battisti V, Fritsch L, Lin
597 WM, Hollmann TJ, Ferré F, Bourque C, Burke CJ, Turner L, Uong A, Johnson LA, Beroukhim

598 R, Mermel CH, Loda M, Ait-Si-Ali S, Garraway LA, Young RA, Zon LI. The histone
599 methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset.
600 Nature. 2011;471(7339):513-7. Epub 2011/03/25. doi: 10.1038/nature09806. PubMed PMID:
601 21430779; PMCID: PMC3348545.
602 17. Jao LE, Wente SR, Chen W. Efficient multiplex biallelic zebrafish genome editing using
603 a CRISPR nuclease system. Proc Natl Acad Sci U S A. 2013;110(34):13904-9. Epub
604 2013/08/07. doi: 10.1073/pnas.1308335110. PubMed PMID: 23918387; PMCID: PMC3752207.
605 18. Raj B, Wagner DE, McKenna A, Pandey S, Klein AM, Shendure J, Gagnon JA, Schier
606 AF. Simultaneous single-cell profiling of lineages and cell types in the vertebrate brain. Nat
607 Biotechnol. 2018;36(5):442-50. Epub 2018/04/03. doi: 10.1038/nbt.4103. PubMed PMID:
608 29608178; PMCID: PMC5938111.
609 19. McKenna A, Findlay GM, Gagnon JA, Horwitz MS, Schier AF, Shendure J. Whole-
610 organism lineage tracing by combinatorial and cumulative genome editing. Science.
611 2016;353(6298):aaf7907. Epub 2016/05/28. doi: 10.1126/science.aaf7907. PubMed PMID:
612 27229144; PMCID: PMC4967023.
613 20. Brombin A, Simpson DJ, Travnickova J, Brunsdon H, Zeng Z, Lu Y, Young AIJ,
614 Chandra T, Patton EE. Tfap2b specifies an embryonic melanocyte stem cell that retains adult
615 multifate potential. Cell Rep. 2022;38(2):110234. Epub 2022/01/13. doi:
616 10.1016/j.celrep.2021.110234. PubMed PMID: 35021087; PMCID: PMC8764619.
617 21. Weiss JM, Hunter MV, Cruz NM, Baggiolini A, Tagore M, Ma Y, Misale S, Marasco M,
618 Simon-Vermot T, Campbell NR, Newell F, Wilmott JS, Johansson PA, Thompson JF, Long GV,
619 Pearson JV, Mann GJ, Scolyer RA, Waddell N, Montal ED, Huang TH, Jonsson P, Donoghue
620 MTA, Harris CC, Taylor BS, Xu T, Chaligné R, Shliaha PV, Hendrickson R, Jungbluth AA,

621 Lezcano C, Koche R, Studer L, Ariyan CE, Solit DB, Wolchok JD, Merghoub T, Rosen N,
622 Hayward NK, White RM. Anatomic position determines oncogenic specificity in melanoma.
623 Nature. 2022;604(7905):354-61. Epub 2022/04/01. doi: 10.1038/s41586-022-04584-6. PubMed
624 PMID: 35355015; PMCID: PMC9355078.
625 22. Shain AH, Joseph NM, Yu R, Benhamida J, Liu S, Prow T, Ruben B, North J, Pincus L,
626 Yeh I, Judson R, Bastian BC. Genomic and Transcriptomic Analysis Reveals Incremental
627 Disruption of Key Signaling Pathways during Melanoma Evolution. Cancer Cell. 2018;34(1):45-
628 55.e4. Epub 2018/07/11. doi: 10.1016/j.ccell.2018.06.005. PubMed PMID: 29990500; PMCID:
629 PMC6319271.
630 23. Smith AP, Hoek K, Becker D. Whole-genome expression profiling of the melanoma
631 progression pathway reveals marked molecular differences between nevi/melanoma in situ and
632 advanced-stage melanomas. Cancer Biol Ther. 2005;4(9):1018-29. Epub 2005/10/28. doi:
633 10.4161/cbt.4.9.2165. PubMed PMID: 16251803.
634 24. Kato S, Lippman SM, Flaherty KT, Kurzrock R. The Conundrum of Genetic "Drivers" in
635 Benign Conditions. J Natl Cancer Inst. 2016;108(8). Epub 2016/04/10. doi: 10.1093/jnci/djw036.
636 PubMed PMID: 27059373; PMCID: PMC5017937.
637 25. Levin DB, Wilson K, Valadares de Amorim G, Webber J, Kenny P, Kusser W. Detection
638 of p53 mutations in benign and dysplastic nevi. Cancer Res. 1995;55(19):4278-82. Epub
639 1995/10/01. PubMed PMID: 7671235.
640 26. Uribe P, Andrade L, Gonzalez S. Lack of association between BRAF mutation and
641 MAPK ERK activation in melanocytic nevi. J Invest Dermatol. 2006;126(1):161-6. Epub
642 2006/01/19. doi: 10.1038/sj.jid.5700011. PubMed PMID: 16417232.

643 27. Venesio T, Chiorino G, Balsamo A, Zaccagna A, Petti C, Scatolini M, Pisacane A,
644 Sarotto I, Picciotto F, Risio M. In melanocytic lesions the fraction of BRAF V600E alleles is
645 associated with sun exposure but unrelated to ERK phosphorylation. Mod Pathol.
646 2008;21(6):716-26. Epub 2008/04/15. doi: 10.1038/modpathol.2008.41. PubMed PMID:
647 18408659.
648 28. Wang F, Flanagan J, Su N, Wang LC, Bui S, Nielson A, Wu X, Vo HT, Ma XJ, Luo Y.
649 RNAscope: a novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues.
650 J Mol Diagn. 2012;14(1):22-9. Epub 2011/12/15. doi: 10.1016/j.jmoldx.2011.08.002. PubMed
651 PMID: 22166544; PMCID: PMC3338343.
652 29. Straume O, Akslen LA. Strong expression of ID1 protein is associated with decreased
653 survival, increased expression of ephrin-A1/EPHA2, and reduced thrombospondin-1 in
654 malignant melanoma. Br J Cancer. 2005;93(8):933-8. Epub 2005/09/29. doi:
655 10.1038/sj.bjc.6602792. PubMed PMID: 16189525; PMCID: PMC2361652.
656 30. Ling MT, Wang X, Zhang X, Wong YC. The multiple roles of Id-1 in cancer progression.
657 Differentiation. 2006;74(9-10):481-7. Epub 2006/12/21. doi: 10.1111/j.1432-0436.2006.00083.x.
658 PubMed PMID: 17177845.
659 31. Alani RM, Young AZ, Shifflett CB. Id1 regulation of cellular senescence through
660 transcriptional repression of p16/Ink4a. Proc Natl Acad Sci U S A. 2001;98(14):7812-6. Epub
661 2001/06/28. doi: 10.1073/pnas.141235398. PubMed PMID: 11427735; PMCID: PMC35424.
662 32. Matys V, Kel-Margoulis OV, Fricke E, Liebich I, Land S, Barre-Dirrie A, Reuter I,
663 Chekmenev D, Krull M, Hornischer K, Voss N, Stegmaier P, Lewicki-Potapov B, Saxel H, Kel
664 AE, Wingender E. TRANSFAC and its module TRANSCompel: transcriptional gene regulation

665 in eukaryotes. Nucleic Acids Res. 2006;34(Database issue):D108-10. Epub 2005/12/31. doi:
666 10.1093/nar/gkj143. PubMed PMID: 16381825; PMCID: PMC1347505.
667 33. Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim
668 R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML,
669 Fisher DE, Sellers WR. Integrative genomic analyses identify MITF as a lineage survival
670 oncogene amplified in malignant melanoma. Nature. 2005;436(7047):117-22. Epub 2005/07/08.
671 doi: 10.1038/nature03664. PubMed PMID: 16001072.
672 34. Uhlén M, Fagerberg L, Hallström BM, Lindskog C, Oksvold P, Mardinoglu A,
673 Sivertsson Å, Kampf C, Sjöstedt E, Asplund A, Olsson I, Edlund K, Lundberg E, Navani S,
674 Szigyarto CA, Odeberg J, Djureinovic D, Takanen JO, Hober S, Alm T, Edqvist PH, Berling H,
675 Tegel H, Mulder J, Rockberg J, Nilsson P, Schwenk JM, Hamsten M, von Feilitzen K, Forsberg
676 M, Persson L, Johansson F, Zwahlen M, von Heijne G, Nielsen J, Pontén F. Proteomics. Tissue-
677 based map of the human proteome. Science. 2015;347(6220):1260419. doi:
678 10.1126/science.1260419. PubMed PMID: 25613900.
679 35. Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, Ellrott K, Shmulevich
680 I, Sander C, Stuart JM. The Cancer Genome Atlas Pan-Cancer analysis project. Nat Genet.
681 2013;45(10):1113-20. doi: 10.1038/ng.2764. PubMed PMID: 24071849; PMCID: PMC3919969.
682 36. Uhlen M, Zhang C, Lee S, Sjöstedt E, Fagerberg L, Bidkhori G, Benfeitas R, Arif M, Liu
683 Z, Edfors F, Sanli K, von Feilitzen K, Oksvold P, Lundberg E, Hober S, Nilsson P, Mattsson J,
684 Schwenk JM, Brunnström H, Glimelius B, Sjöblom T, Edqvist PH, Djureinovic D, Micke P,
685 Lindskog C, Mardinoglu A, Ponten F. A pathology atlas of the human cancer transcriptome.
686 Science. 2017;357(6352). doi: 10.1126/science.aan2507. PubMed PMID: 28818916.

687 37. Thiebault K, Mazelin L, Pays L, Llambi F, Joly MO, Scoazec JY, Saurin JC, Romeo G,
688 Mehlen P. The netrin-1 receptors UNC5H are putative tumor suppressors controlling cell death
689 commitment. Proc Natl Acad Sci U S A. 2003;100(7):4173-8. Epub 20030324. doi:
690 10.1073/pnas.0738063100. PubMed PMID: 12655055; PMCID: PMC153067.
691 38. Williams ME, Lu X, McKenna WL, Washington R, Boyette A, Strickland P, Dillon A,
692 Kaprielian Z, Tessier-Lavigne M, Hinck L. UNC5A promotes neuronal apoptosis during spinal
693 cord development independent of netrin-1. Nat Neurosci. 2006;9(8):996-8. Epub 20060709. doi:
694 10.1038/nn1736. PubMed PMID: 16829956.
695 39. Han X, Chen H, Huang D, Chen H, Fei L, Cheng C, Huang H, Yuan GC, Guo G.
696 Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell
697 RNA-sequencing. Genome Biol. 2018;19(1):47. Epub 20180405. doi: 10.1186/s13059-018-
698 1426-0. PubMed PMID: 29622030; PMCID: PMC5887227.
699 40. Zhang L, Li W, Cao L, Xu J, Qian Y, Chen H, Zhang Y, Kang W, Gou H, Wong CC, Yu
700 J. PKNOX2 suppresses gastric cancer through the transcriptional activation of IGFBP5 and p53.
701 Oncogene. 2019;38(23):4590-604. Epub 20190211. doi: 10.1038/s41388-019-0743-4. PubMed
702 PMID: 30745575; PMCID: PMC6756047.
703 41. Reemann P, Reimann E, Ilmjärv S, Porosaar O, Silm H, Jaks V, Vasar E, Kingo K, Kõks
704 S. Melanocytes in the skin--comparative whole transcriptome analysis of main skin cell types.
705 PLoS One. 2014;9(12):e115717. Epub 20141229. doi: 10.1371/journal.pone.0115717. PubMed
706 PMID: 25545474; PMCID: PMC4278762.
707 42. Selzer E, Okamoto I, Lucas T, Kodym R, Pehamberger H, Jansen B. Protein kinase C
708 isoforms in normal and transformed cells of the melanocytic lineage. Melanoma Res.
709 2002;12(3):201-9. doi: 10.1097/00008390-200206000-00003. PubMed PMID: 12140376.

710 43. Köhler C, Nittner D, Rambow F, Radaelli E, Stanchi F, Vandamme N, Baggiolini A,
711 Sommer L, Berx G, van den Oord JJ, Gerhardt H, Blanpain C, Marine JC. Mouse Cutaneous
712 Melanoma Induced by Mutant BRaf Arises from Expansion and Dedifferentiation of Mature
713 Pigmented Melanocytes. Cell Stem Cell. 2017;21(5):679-93.e6. Epub 2017/10/17. doi:
714 10.1016/j.stem.2017.08.003. PubMed PMID: 29033351.
715 44. Bostwick DG, Cheng L. Precursors of prostate cancer. Histopathology. 2012;60(1):4-27.
716 Epub 2012/01/04. doi: 10.1111/j.1365-2559.2011.04007.x. PubMed PMID: 22212075.
717 45. Kaz AM, Grady WM, Stachler MD, Bass AJ. Genetic and Epigenetic Alterations in
718 Barrett's Esophagus and Esophageal Adenocarcinoma. Gastroenterol Clin North Am.
719 2015;44(2):473-89. Epub 2015/05/30. doi: 10.1016/j.gtc.2015.02.015. PubMed PMID:
720 26021206; PMCID: PMC4449457.
721 46. Virnig BA, Tuttle TM, Shamliyan T, Kane RL. Ductal carcinoma in situ of the breast: a
722 systematic review of incidence, treatment, and outcomes. J Natl Cancer Inst. 2010;102(3):170-8.
723 Epub 2010/01/15. doi: 10.1093/jnci/djp482. PubMed PMID: 20071685.
724 47. Shain AH, Yeh I, Kovalyshyn I, Sriharan A, Talevich E, Gagnon A, Dummer R, North J,
725 Pincus L, Ruben B, Rickaby W, D'Arrigo C, Robson A, Bastian BC. The Genetic Evolution of
726 Melanoma from Precursor Lesions. N Engl J Med. 2015;373(20):1926-36. Epub 2015/11/13.
727 doi: 10.1056/NEJMoa1502583. PubMed PMID: 26559571.
728 48. Lian CG, Murphy GF. The Genetic Evolution of Melanoma. N Engl J Med.
729 2016;374(10):994-5. Epub 2016/03/11. doi: 10.1056/NEJMc1515834. PubMed PMID:
730 26962742.
731 49. Kwan KM, Fujimoto E, Grabher C, Mangum BD, Hardy ME, Campbell DS, Parant JM,
732 Yost HJ, Kanki JP, Chien CB. The Tol2kit: a multisite gateway-based construction kit for Tol2

733 transposon transgenesis constructs. Dev Dyn. 2007;236(11):3088-99. Epub 2007/10/17. doi:
734 10.1002/dvdy.21343. PubMed PMID: 17937395.
735 50. Gagnon JA, Valen E, Thyme SB, Huang P, Akhmetova L, Pauli A, Montague TG,
736 Zimmerman S, Richter C, Schier AF. Efficient mutagenesis by Cas9 protein-mediated
737 oligonucleotide insertion and large-scale assessment of single-guide RNAs. PLoS One.
738 2014;9(5):e98186. Epub 2014/05/31. doi: 10.1371/journal.pone.0098186. PubMed PMID:
739 24873830; PMCID: PMC4038517.
740 51. Baron CS, Mitchell O, Avagyan S, Menard R, Yang S, Robertson AL, Potluri R,
741 Shendure J, Madelaine R, McKenna A, Zon LI. Leukemia-derived apelin selects endothelial
742 niche clones to promote tumorigenesis. bioRxiv. 2024:2024.09.09.612077. doi:
743 10.1101/2024.09.09.612077.
744 52. Avagyan S, Henninger JE, Mannherz WP, Mistry M, Yoon J, Yang S, Weber MC, Moore
745 JL, Zon LI. Resistance to inflammation underlies enhanced fitness in clonal hematopoiesis.
746 Science. 2021;374(6568):768-72. Epub 2021/11/05. doi: 10.1126/science.aba9304. PubMed
747 PMID: 34735227.
748 53. Klein AM, Mazutis L, Akartuna I, Tallapragada N, Veres A, Li V, Peshkin L, Weitz DA,
749 Kirschner MW. Droplet barcoding for single-cell transcriptomics applied to embryonic stem
750 cells. Cell. 2015;161(5):1187-201. Epub 2015/05/23. doi: 10.1016/j.cell.2015.04.044. PubMed
751 PMID: 26000487; PMCID: PMC4441768.
752 54. Zilionis R, Nainys J, Veres A, Savova V, Zemmour D, Klein AM, Mazutis L. Single-cell
753 barcoding and sequencing using droplet microfluidics. Nat Protoc. 2017;12(1):44-73. Epub
754 2016/12/09. doi: 10.1038/nprot.2016.154. PubMed PMID: 27929523.

755 55. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
756 2012;9(4):357-9. Epub 2012/03/06. doi: 10.1038/nmeth.1923. PubMed PMID: 22388286;
757 PMCID: PMC3322381.
758 56. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, 3rd, Hao Y,
759 Stoeckius M, Smibert P, Satija R. Comprehensive Integration of Single-Cell Data. Cell.
760 2019;177(7):1888-902.e21. Epub 2019/06/11. doi: 10.1016/j.cell.2019.05.031. PubMed PMID:
761 31178118; PMCID: PMC6687398.
762 57. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell
763 transcriptomic data across different conditions, technologies, and species. Nat Biotechnol.
764 2018;36(5):411-20. Epub 2018/04/03. doi: 10.1038/nbt.4096. PubMed PMID: 29608179;
765 PMCID: PMC6700744.
766 58. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ,
767 Mikkelsen TS, Rinn JL. The dynamics and regulators of cell fate decisions are revealed by
768 pseudotemporal ordering of single cells. Nat Biotechnol. 2014;32(4):381-6. Epub 20140323. doi:
769 10.1038/nbt.2859. PubMed PMID: 24658644; PMCID: PMC4122333.
770 59. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads.
771 2011. 2011;17(1):3. Epub 2011-08-02. doi: 10.14806/ej.17.1.200.
772 60. Andrews S. FastQC: a quality control tool for high throughput sequence data2010.
773 61. Trapnell C, Pachter L, Salzberg SL. TopHat: discovering splice junctions with RNA-Seq.
774 Bioinformatics. 2009;25(9):1105-11. Epub 2009/03/18. doi: 10.1093/bioinformatics/btp120.
775 PubMed PMID: 19289445; PMCID: PMC2672628.
776 62. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL,
777 Wold BJ, Pachter L. Transcript assembly and quantification by RNA-Seq reveals unannotated

transcripts and isoform switching during cell differentiation. Nat Biotechnol. 2010;28(5):511-5. Epub 2010/05/04. doi: 10.1038/nbt.1621. PubMed PMID: 20436464; PMCID: PMC3146043.

63. Zhou Y, Zhou B, Pache L, Chang M, Khodabakhshi AH, Tanaseichuk O, Benner C, Chanda SK. Metascape provides a biologist-oriented resource for the analysis of systems-level datasets. Nat Commun. 2019;10(1):1523. Epub 2019/04/05. doi: 10.1038/s41467-019-09234-6. PubMed PMID: 30944313; PMCID: PMC6447622.

64. Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. 2013;10(12):1213-8. Epub 2013/10/08. doi: 10.1038/nmeth.2688. PubMed PMID: 24097267; PMCID: PMC3959825.

65. Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. Model-based analysis of ChIP-Seq (MACS). Genome Biol. 2008;9(9):R137. Epub 2008/09/19. doi: 10.1186/gb-2008-9-9-r137. PubMed PMID: 18798982; PMCID: PMC2592715.

66. Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell. 2010;38(4):576-89. doi: 10.1016/j.molcel.2010.05.004. PubMed PMID: 20513432; PMCID: PMC2898526.

67. Noonan HR, Thornock AM, Barbano J, Xifaras ME, Baron CS, Yang S, Koczirka K, McConnell AM, Zon LI. A chronic signaling TGFb zebrafish reporter identifies immune response in melanoma. eLife. 2024;13:e83527. doi: 10.7554/eLife.83527.

799 68. Kondili M, Fust A, Preussner J, Kuenne C, Braun T, Looso M. UROPA: a tool for
800 Universal RObust Peak Annotation. Sci Rep. 2017;7(1):2593. Epub 20170601. doi:
801 10.1038/s41598-017-02464-y. PubMed PMID: 28572580; PMCID: PMC5453960.
802 69. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar
803 F, Manke T. deepTools2: a next generation web server for deep-sequencing data analysis.
804 Nucleic Acids Res. 2016;44(W1):W160-5. Epub 20160413. doi: 10.1093/nar/gkw257. PubMed
805 PMID: 27079975; PMCID: PMC4987876.
806 70. An integrated encyclopedia of DNA elements in the human genome. Nature.
807 2012;489(7414):57-74. doi: 10.1038/nature11247. PubMed PMID: 22955616; PMCID:
808 PMC3439153.
809 71. Consortium TGO, Aleksander SA, Balhoff J, Carbon S, Cherry JM, Drabkin HJ, Ebert D,
810 Feuermann M, Gaudet P, Harris NL, Hill DP, Lee R, Mi H, Moxon S, Mungall CJ,
811 Muruganugan A, Mushayahama T, Sternberg PW, Thomas PD, Van Auken K, Ramsey J,
812 Siegele DA, Chisholm RL, Fey P, Aspromonte MC, Nugnes MV, Quaglia F, Tosatto S, Giglio
813 M, Nadendla S, Antonazzo G, Attrill H, dos Santos G, Marygold S, Strelets V, Tabone CJ,
814 Thurmond J, Zhou P, Ahmed SH, Asanitthong P, Luna Buitrago D, Erdol MN, Gage MC, Ali
815 Kadhum M, Li KYC, Long M, Michalak A, Pesala A, Pritazahra A, Saverimuttu SCC, Su R,
816 Thurlow KE, Lovering RC, Logie C, Oliferenko S, Blake J, Christie K, Corbani L, Dolan ME,
817 Drabkin HJ, Hill DP, Ni L, Sitnikov D, Smith C, Cuzick A, Seager J, Cooper L, Elser J, Jaiswal
818 P, Gupta P, Jaiswal P, Naithani S, Lera-Ramirez M, Rutherford K, Wood V, De Pons JL,
819 Dwinell MR, Hayman GT, Kaldunski ML, Kwitek AE, Laulederkind SJF, Tutaj MA, Vedi M,
820 Wang S-J, D'Eustachio P, Aimo L, Axelsen K, Bridge A, Hyka-Nouspikel N, Morgat A,
821 Aleksander SA, Cherry JM, Engel SR, Karra K, Miyasato SR, Nash RS, Skrzypek MS, Weng S,

822 Wong ED, Bakker E, Berardini TZ, Reiser L, Auchincloss A, Axelsen K, Argoud-Puy G, Blatter
823 M-C, Boutet E, Breuza L, Bridge A, Casals-Casas C, Coudert E, Estreicher A, Livia Famiglietti
824 M, Feuermann M, Gos A, Gruaz-Gumowski N, Hulo C, Hyka-Nouspikel N, Jungo F, Le Mercier
825 P, Lieberherr D, Masson P, Morgat A, Pedruzzi I, Pourcel L, Poux S, Rivoire C, Sundaram S,
826 Bateman A, Bowler-Barnett E, Bye-A-Jee H, Denny P, Ignatchenko A, Ishtiaq R, Lock A, Lussi
827 Y, Magrane M, Martin MJ, Orchard S, Raposo P, Speretta E, Tyagi N, Warner K, Zaru R, Diehl
828 AD, Lee R, Chan J, Diamantakis S, Raciti D, Zarowiecki M, Fisher M, James-Zorn C,
829 Ponferrada V, Zorn A, Ramachandran S, Ruzicka L, Westerfield M. The Gene Ontology
830 knowledgebase in 2023. Genetics. 2023;224(1). doi: 10.1093/genetics/iyad031.
831 72. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, Davis AP, Dolinski
832 K, Dwight SS, Eppig JT, Harris MA, Hill DP, Issel-Tarver L, Kasarskis A, Lewis S, Matese JC,
833 Richardson JE, Ringwald M, Rubin GM, Sherlock G. Gene Ontology: tool for the unification of
834 biology. Nature Genetics. 2000;25(1):25-9. doi: 10.1038/75556.
835

836 **FIGURE LEGENDS**
837

838 **Figure 1 | Cancer precursor zone formation precedes neural crest reactivation in zebrafish.**
839 **a,** Melanoma initiation in zebrafish occurs in stages, beginning with the formation of a
840 cancerized field of rescued melanocytes, which expand to form a cancer precursor zone. Neural
841 crest reactivation occurs in cells within the cancer precursor zone and tumor formation follows,
842 with all tumor cells becoming *crestin:EGFP*+. Scale bar: 1 mm. **b,** Swimmers plot of 13
843 zebrafish during melanoma initiation show variable onset of CPZ and neural reactivation (#
844 represents fish time course in Fig 1a; * represents fish time course in Extended Data Fig. 1c). **c**,
845 Transverse sections of zebrafish skin stained for mCherry and EGFP at the cancerized field,

846 cancer precursor zone, neural crest reactivation, and tumor formation stages. Scale bars, 100 μm.
847 **d,** Limit dilution curves at 14 days post-transplant (dpt). Confidence interval estimate is
848 displayed as 1/(engraftment cell frequency).
849
850 **Figure 2 | Single cell expression atlas of melanoma initiation. a**, UMAP of cancerized fields
851 (935 cells) or small *crestin:EGFP*+ patches with surrounding cancer precursor zone (1251 cells).
852 **b,** UMAP of Pseudotime analysis with cancerized field melanocytes set as the trunk. **c**,
853 Melanocyte and neural crest gene expression in cancerized field control melanocyte (mel),
854 cancer precursor zone (CPZ), and *crestin*+ melanoma cell subsets. **d,** Pseudotime analysis of
855 genes identified as stage specific in scRNA-seq. **e**, Differentially expressed genes categorized
856 into signaling (red), immune (blue), and neural crest (green). **f**, Open chromatin peaks in
857 promoter or enhancer regions of stage-specific genes.
858
859 **Figure 3 | Stage-specific transcription factors induce cancer precursor zone formation**
860 **through ID1 and neural crest reactivation via IRF3. a**, Images showing *mitfa:mCherry* and
861 *crestin:EGFP* fluorescence in zebrafish overexpressing ID1, IRF3, compared to an empty
862 control. White dashed line outlines the cancer precursor zones or patches. Gut autofluorescence
863 can be seen outside of the dashed lines. **b**, Violin plot showing the weighted number of cancer
864 precursor zones or *crestin*+ patches per fish. Each dot represents one fish; n = 35-65. ns = not
865 significant; $^{**}p \leq 0.01$ **c,** Stacked bar graph representing the number of tumors in ID1 and IRF3
866 overexpressing zebrafish compared to empty control (n=15 fish per cohort). **d,** Kaplan-Meier
867 curve of melanoma incidence. ID1 or IRF3 overexpression does not accelerate melanoma onset
868 compared to an empty control. n.s. = not significant. **e,** Immunofluorescent staining of

869 melanocytes (MART1), ID1, and IRF3 in human skin samples. Scale bars, 10μm. **f**, Percent of
870 melanocytes that are positive for cytoplasmic and/or nuclear ID1 and nuclear IRF3. Each dot
871 represents one patient (n = 5-9). **$p \leq 0.01$; ****$p \leq 0.0001$.
872
873
874 **Data Figure 4 | ID1 sequesters TCF12 from chromatin to drive cancer precursor zone**
875 **formation through loss of development and signaling regulatory genes. a**, Volcano plot
876 analysis of IP-MS data comparing proteins bound to ID1 vs Clover pull down. **b**, Heatmap of
877 TCF12 targets identified by their decreased expression in the context of ID1 overexpression and
878 the presence of a TCF12 motif in the promoter/enhancer region. **c,** Heatmaps depicting ChIP-seq
879 peak structure following Clover (left, gray) or ID1 (right, purple) overexpression in
880 PMEL(BRAF^{V600E}) cells. A kb window is centered on the peak center highlighting peaks lost
881 after ID1 overexpression. **d,** Quantification of peaks with TCF12 motifs following Clover
882 (n=805) or ID1 (n=336) overexpression. TCF12 motif is only enriched in Clover overexpressing
883 cells (p = 1e-17). **e,** IGV tracks of TCF12 motif containing peaks in clover (gray) or ID1 (purple)
884 overexpressing cells. Purple bar indicates peak called by MACS. **f,g.** Bar graph representing
885 viable cells relative to non-transfected controls following Clover or ID1 overexpression in
886 PMEL(BRAF^{V600E}) (**f**) or TCF12 overexpression in A375 metastatic melanoma cells (**g**). ID1
887 overexpression increases proliferation in primary melanocytes while TCF12 decreases
888 proliferation in transformed malignant cells. ***, p<0.01; ****, p<0.001. **h,** Violin plot showing
889 the weighted number of cancer precursor zones following TCF12 overexpression relative to
890 empty control. Each dot represents one fish; n = 12. ns ****$p \leq 0.001$. **i,** Kaplan-Meier curve of

891 melanoma incidence. Melanoma incidence is delayed in TCF12 overexpressing fish that develop
892 CPZs. *, $p<0.05$.
893
894 **EXTENDED DATA FIGURE LEGENDS**
895
896 **Extended Data Figure 1 | Cancer precursor zone formation precedes neural crest**
897 **reactivation in zebrafish. a**, Melanoma initiates in a stage-specific manner, beginning with the
898 formation of an *mitfa*^{high} *crestin:EGFP*- cancer precursor zone. The neural crest progenitor state
899 is reactivated from within the cancer precursor zone and tumor formation occurs following this
900 reactivation. **b,** Melanoma initiation stages in a zebrafish with a dorsal tumor. Scale bar: 1mm**. c**
901 , Images of melanoma initiation for zebrafish marked * in Fig. 1a. Scale bar: 1mm. **c,** Violin plot
902 of CPZ and patch onset. CPZ have a median onset of 11 wpf (circle, left) while patch occur at
903 15wpf (square, right). Median is shown in solid black line. Each dot is a fish; n=25. **d-e,** Cancer
904 precursor zone encompassing *crestin* reactivation is seen in *mcr:BRAF*^{V600E} (**d**) and
905 *mcr:NRAS*^{Q61R} (**e**).
906
907 **Extended Data Figure 2 | Neural crest reactivation is not dependent on anatomic position,**
908 **size, or *mitfa:mCherry* intensity. a,** CPZ anatomic location across the zebrafish body. CPZ
909 incidence does not correlate with anatomic location. Size of the dot represents the count of CPZs
910 at that location, n=64. **b, c.** Violin plots of cancer precursor zone area (**b**) or mean mitfa:mCherry
911 intensity within that area (**c**). Size or mitfa:mCherry intensity does not predict which CPZs will
912 reactivate neural crest (square; mean area: 103.2±30.0; mean intensity: 0.49±0.26) compared to

913 CPZs that stall in the CPZ stage (circle; mean area: 0.60±0.28; mean intensity: 103.0±23.1).
914 Each point is a fish, n=20.
915
916 **Extended Data Figure 3 | Cancer precursor zones are proliferative regions of atypical**
917 **melanocytes that precede neural crest reactivation. a**, Confocal images depicting
918 morphological changes; cancerized field melanocytes are dendritic and regularly spaced,
919 whereas patches with neural crest reactivation are made of atypical non-dendritic melanocytes. **b**,
920 PCNA staining reveals that cancer precursor zones and patches with neural crest reactivation are
921 highly proliferative compared to infrequently proliferative cancerized field melanocytes. Scale
922 bars, 100 μm. **c**, H&E staining of cancer precursor zones show that nuclei are dysplastic. Scale
923 bars, 100μm.
924
925 **Extended Data Figure 4 | Malignant potential of melanoma initiation stages. a**, Schematic of
926 experiment. Precursor lesion (*mitfa:mCherry+/crestin-*), small *crestin* patch
927 (*mitfa:mCherry+/crestin+*), or tumor cells (*mitfa:mCherry+/crestin+*) were isolated via FACS
928 and subcutaneously (Sub-Q) transplanted into irradiated (IR) recipients at various densities. **b**,
929 Confidence interval table for 1/(engraftment cell frequency) in transplant studies. **c**, Time course
930 imaging of precursor lesion transplantation shows that *crestin* is reactivated upon engraftment.
931
932 **Extended Data Figure 5 | Single cell transcriptome of melanoma initiation. a,** Dot plot
933 showing the expression of cell-type specific genes used to identify clusters. The dot plot is
934 broken up by sample, with the control cancerized field (ctrl) in red and the *crestin* (cre) sample in
935 blue. **b**, UMAP visualization of cell clusters marked by sample. The control sample contains

936 cells isolated from a cancerized field and the *crestin* sample contains cells isolated from a small
937 *crestin*+ patch with a surrounding cancer precursor zone. **c,** UMAP of single cells (blue) with
938 excluded doublets (red) from the scRNA-seq. **d,** UMAP of the assigned cell identities from the
939 Pseudotime presented in Fig. 2b.
940
941 **Extended Data Figure 6 | Differential gene expression across melanoma stages. a-c**, Volcano
942 plots showing differentially expressed genes in CPZ (**a**), patch (**b**), and tumor (**c**) relative to
943 cancerized field control. Dashed lines represent a 2 $\log_2$FC (over cancerized field) and a 10e-3 q-
944 value as assigned by DeSeq2. Colored regions indicate genes that meet the logFC and q-value
945 threshold. **d**, PCA plot of cancerized field, cancer precursor zones, and patches. CPZ
946 melanocytes (yellow) represent a transitory population from cancerized field (red) to patch
947 (green and blue) and tumor (purple).
948
949 **Extended Data Figure 7 | Stage-specific gene expression during melanoma initiation in**
950 **zebrafish and humans. a**, Gene expression of stage-specific differentially expressed genes from
951 FACS-sorted melanocytes during melanoma initiation. **b**, Stage-specific expression signatures in
952 intermediate lesions, thin melanoma, and thick melanoma samples taken from human patients
953 from Shain et al. Cancer Cell, 2018. **c**, Stage-specific gene expression in human melanoma
954 samples from microarray analysis performed in Smith et al. Cancer Biol Ther 2005. Samples
955 were taken from normal skin (NS), benign nevi (BN), atypical nevi (AN), melanoma in situ
956 (MIS), vertical growth phase (VGP), metastatic growth phase (MGP), and lymph node
957 metastasis (LN).
958

959 **Extended Data Figure 8 | Differential accessible chromatin regions across melanoma stages.**
960 **a-c**, Volcano plots showing differentially accessible regions in CPZ (**a**), patch (**b**), and tumor (**c**)
961 relative to cancerized field control. Dashed lines represent a 2 $\log_2$FC (over cancerized field) and
962 a 10e-3 q-value as assigned by DeSeq2. Colored regions indicate peaks that meet the logFC and
963 q-value threshold.
964
965 **Extended Data Figure 9 | Enriched binding motifs during melanoma initiation. a-g**, ATAC-
966 seq peak tracks showing open chromatin in enhancers and/or promoters for the cancer precursor
967 zone genes *id1* (**a**), *il6st* (**b**), and *tfap2a* (**c**), as well as the *crestin* patch genes *ctnnb2* (**d**), *irf3* (**e**),
968 *sox10* (**f**), and *crestin* (**g**). Black box highlights the peak that is displayed in Figure 2d.
969
970 **Extended Data Figure 10 | MAPK pathway is not transcriptionally active in cancerized**
971 **field melanocytes. a**, MAPK pathway gene expression. CPZ = cancer precursor zone. **b**, ATAC-
972 seq peak tracks demonstrating that chromatin around MAPK pathway genes does not open until
973 the formation of a cancer precursor zone. **c**, mCherry and phospho-ERK in cancerized field
974 (left), early *mitfa* ^{med} cancer precursor zone (middle), and *mitfa* ^{high} cancer precursor zone (right).
975 Scale bars, 100 μm. **d**, Expression levels of $BRAF^{V600E}$ are unchanged between FACS-isolated
976 cancerized field (CF) and cancer precursor zone (CPZ) melanocytes, n=3-8. ****$p \leq 0.0001$.
977
978 **Extended Data Figure 11 | ID1 involved in early-stage melanoma initiation and IRF3**
979 **involved in late-stage melanoma initiation. a,b.** RNAScope in situ hybridization for id1 (red)
980 and DAPI nuclear stain in fixed and sectioned cancerized field (**a**) or CPZ (**b**) melanocytes.
981 Dashed line outlines the skin surface. Zebrafish id1 is not expressed in cancerized field

982 melanocytes but is expressed abundantly in CPZ melanocytes. **c**, Heatmap of IRF3 pathway
983 members increasing expression at later stages of initiation. Cancerized field (CF), cancer
984 precursor zone (CPZ), as well as small (S), medium (M), and large (L) *crestin* cells were
985 compared. **d**, Mean intensity of ID1in individual melanocytes in normal skin (N), atypical
986 melanocyte zones (AML), melanoma in situ (MIS), and primary melanoma (PM) samples,
987 n=109-132 melanocytes across 5-9 patient samples. $^{**}p \leq 0.01$; $^{***}p \leq 0.001$; $^{****}p \leq 0.0001$.
988
989 **Extended Data Figure 12 | ID1 inhibits TCF12, preventing binding to downstream targets**
990 **and enabling cancer precursor zone formation. a**, Heatmap of the Log2FC of ID1-bound
991 proteins over control. **b,** Expression of known TCF12 targets identified by the TRANSFAC
992 Curated Transcription Factor Targets dataset. **c**, Expression of TCF12 targets decrease in cancer
993 precursor zone and further decrease when ID1 is overexpressed. **d,** Increasing ID1 expression
994 decreases open chromatin peaks in promoter or enhancer regions of *rhogb* and *parvg*. **e,** Dot plot
995 of enriched GO terms in genes with annotated TCF12 motif lost following ID1 overexpression.
996
997

Figure 1

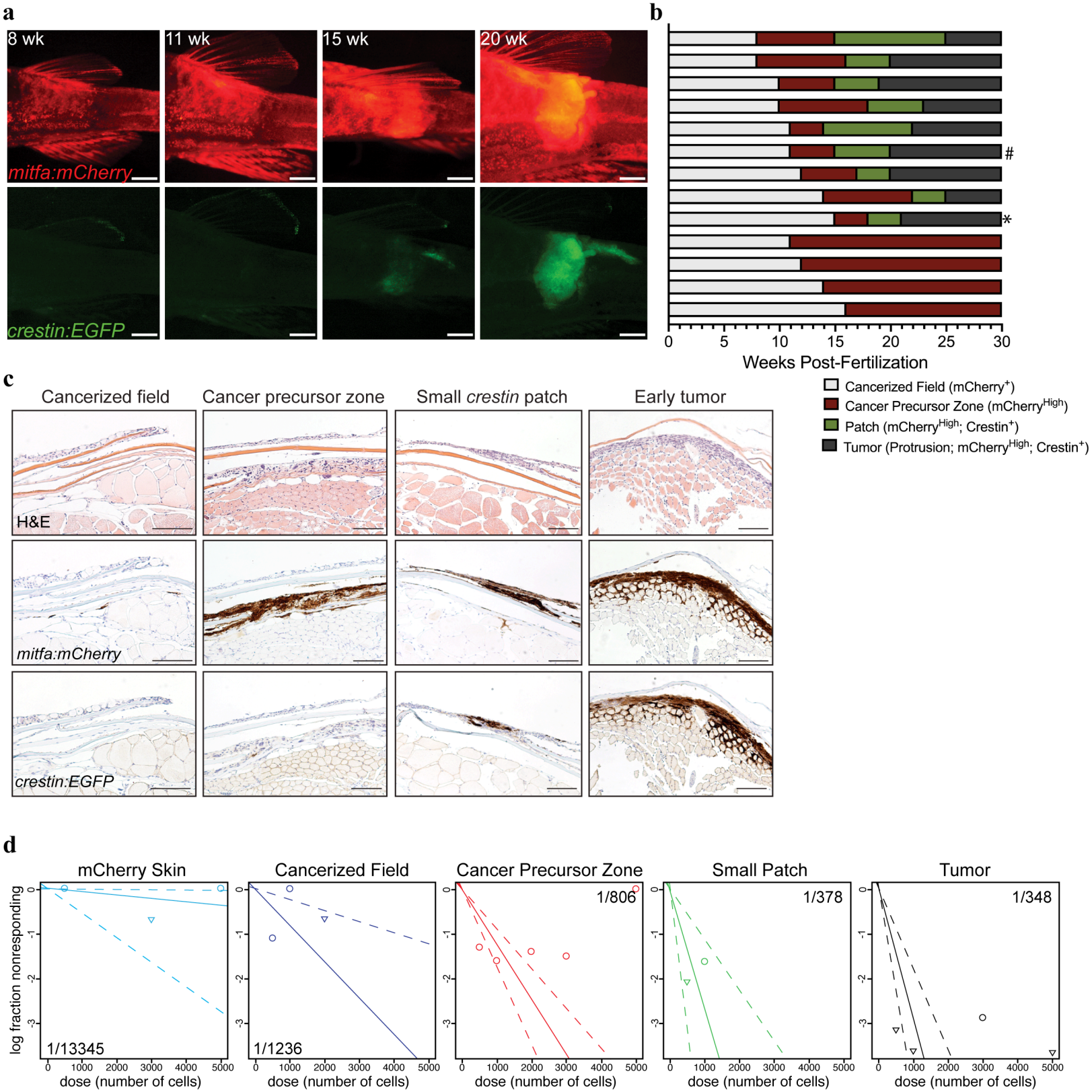

a
8 wk
11 wk
15 wk
20 wk
mitfa:mCherry
crestin:EGFP
b
#
*
0
5
10
15
20
25
30
Weeks Post-Fertilization
Cancerized Field (mCherry+)
Cancer Precursor Zone (mCherryHigh)
Patch (mCherryHigh; Crestin+)
Tumor (Protrusion; mCherryHigh; Crestin+)
c
Cancerized field
Cancer precursor zone
Small crestin patch
Early tumor
H&E
mitfa:mCherry
crestin:EGFP
d
mCherry Skin
Cancerized Field
Cancer Precursor Zone
Small Patch
Tumor
log fraction nonresponding
1/13345
1/1236
1/806
1/378
1/348
dose (number of cells)

Figure 2

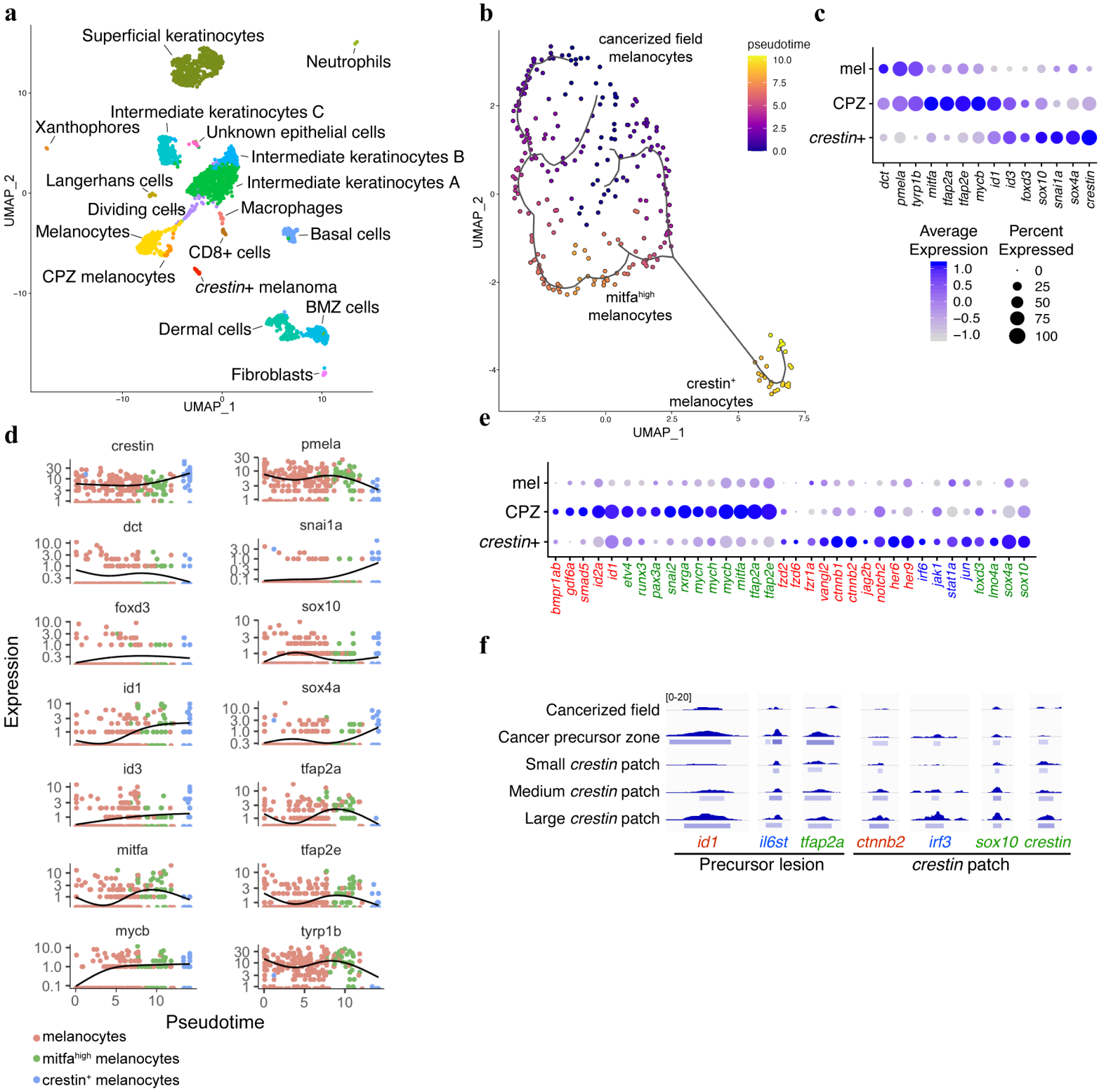

a
Superficial keratinocytes
Neutrophils
Intermediate keratinocytes C
Xanthophores
Unknown epithelial cells
Intermediate keratinocytes B
Langerhans cells
Intermediate keratinocytes A
Dividing cells
Macrophages
Melanocytes
Basal cells
CD8+ cells
CPZ melanocytes
crestin+ melanoma
BMZ cells
Dermal cells
Fibroblasts
UMAP_2
UMAP_1
b
cancerized field melanocytes
pseudotime
mitfa^high melanocytes
crestin+ melanocytes
UMAP_2
UMAP_1
c
mel
CPZ
crestin+
dct
pmela
tyrp1b
mitfa
tfap2a
tfap2e
mycb
id1
id3
foxd3
sox10
snai1a
sox4a
crestin
Average Expression
Percent Expressed
d
crestin
pmela
dct
snai1a
foxd3
sox10
id1
sox4a
id3
tfap2a
mitfa
tfap2e
mycb
tyrp1b
Expression
Pseudotime
melanocytes
mitfa^high melanocytes
crestin+ melanocytes
e
mel
CPZ
crestin+
bmpr1ab
gdf6a
smad5
id2a
id1
etv4
runx3
pax3a
snai2
rxrga
mycn
mych
mycb
mitfa
tfap2a
tfap2e
fzd2
fzd6
fzr1a
vangl2
ctnnb1
ctnnb2
jag2b
notch2
her6
her9
irf6
jak1
stat1a
jun
foxd3
lmo4a
sox4a
sox10
f
[0-20]
Cancerized field
Cancer precursor zone
Small crestin patch
Medium crestin patch
Large crestin patch
id1
il6st
tfap2a
ctnnb2
irf3
sox10
crestin
Precursor lesion
crestin patch

Figure 3

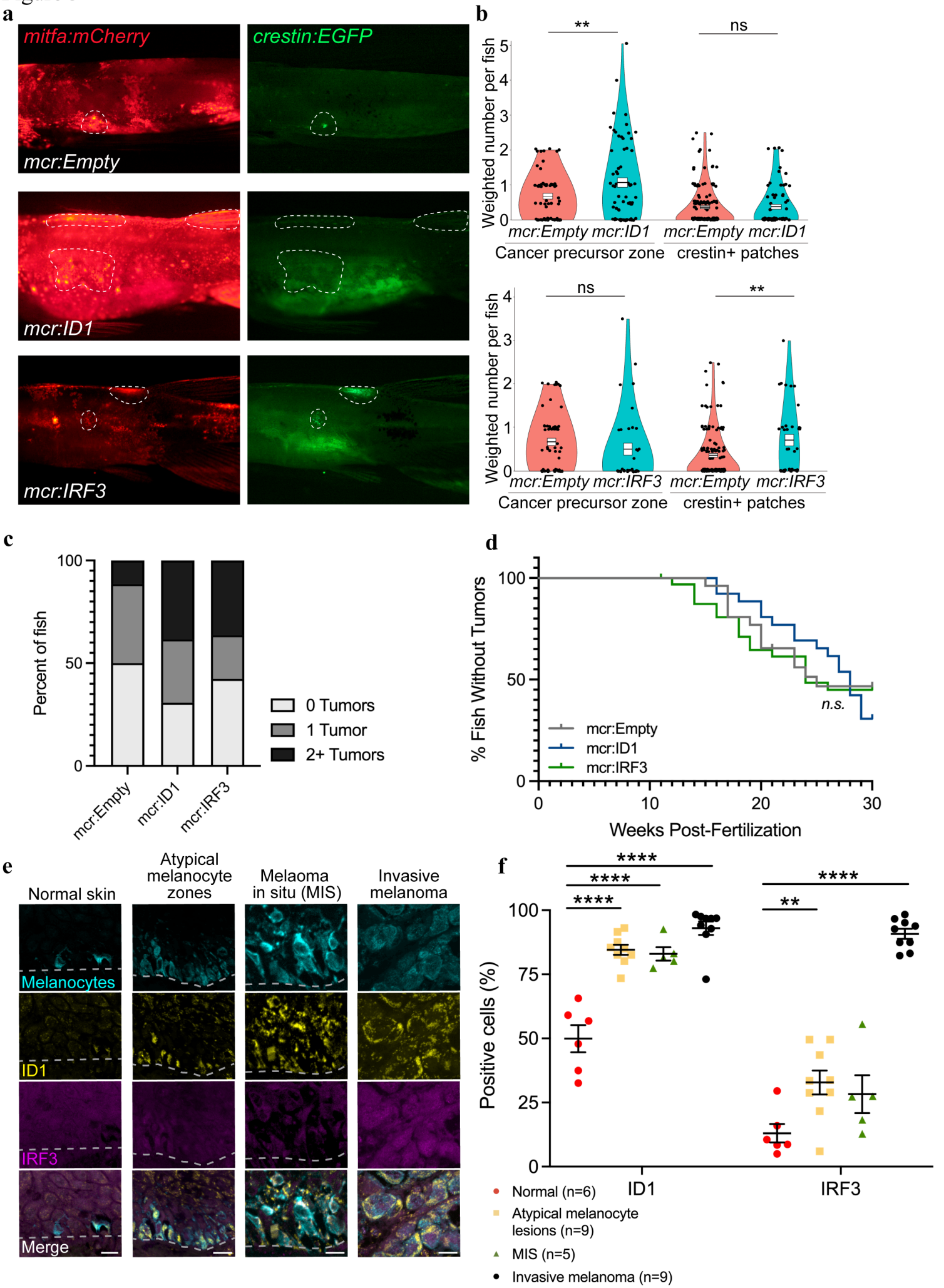
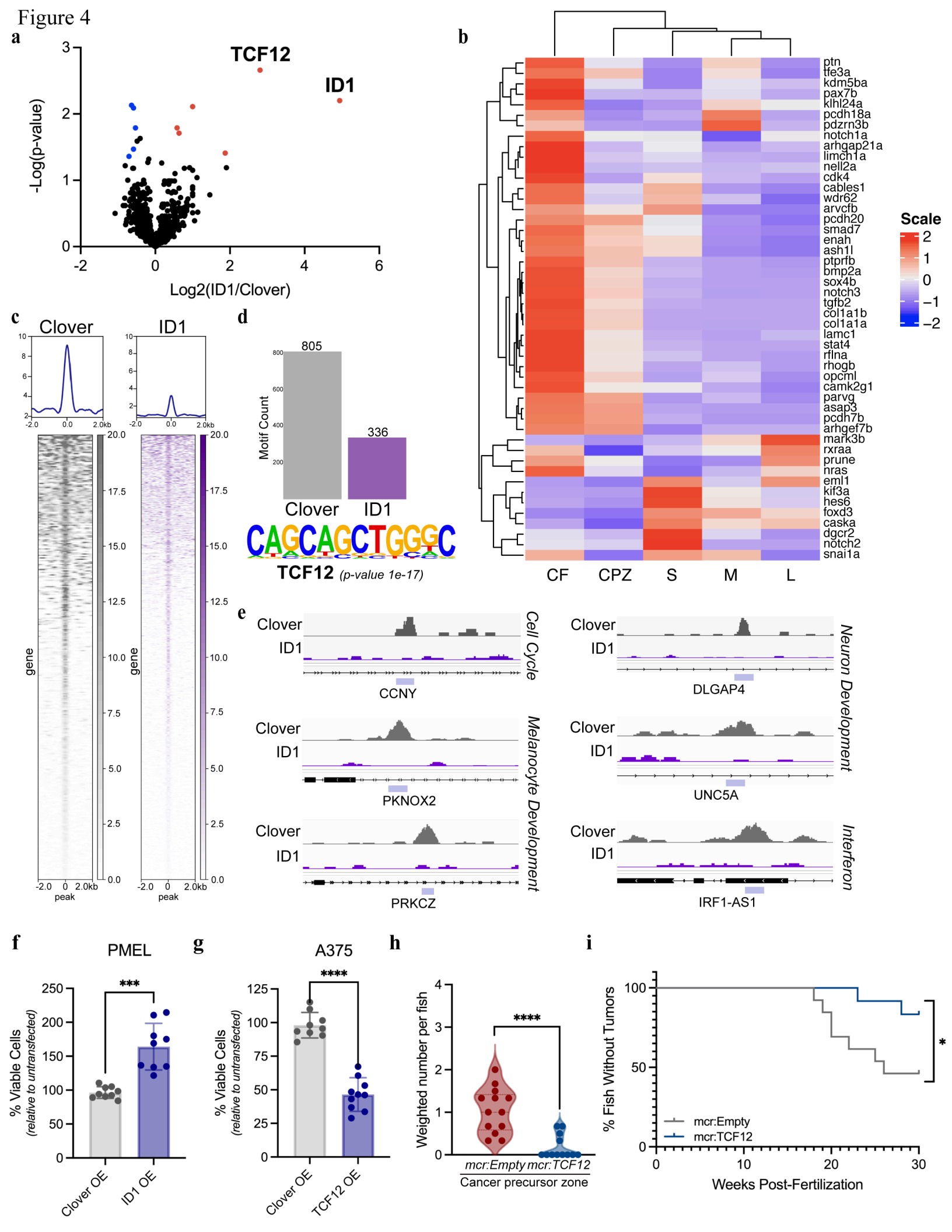

a
mitfa:mCherry
crestin:EGFP
mcr:Empty
mcr:ID1
mcr:IRF3
b
Weighted number per fish
**
ns
mcr:Empty mcr:ID1 mcr:Empty mcr:ID1
Cancer precursor zone
crestin+ patches
ns
**
mcr:Empty mcr:IRF3 mcr:Empty mcr:IRF3
c
Percent of fish
0 Tumors
1 Tumor
2+ Tumors
mcr:Empty
mcr:ID1
mcr:IRF3
d
% Fish Without Tumors
Weeks Post-Fertilization
n.s.
mcr:Empty
mcr:ID1
mcr:IRF3
e
Normal skin
Atypical melanocyte zones
Melaoma in situ (MIS)
Invasive melanoma
Melanocytes
ID1
IRF3
Merge
f
Positive cells (%)
****
**
ID1
IRF3
Normal (n=6)
Atypical melanocyte lesions (n=9)
MIS (n=5)
Invasive melanoma (n=9)

Figure 4


a
TCF12
ID1
-Log(p-value)
Log2(ID1/Clover)
b
Scale
CF
CPZ
S
M
L
c
Clover
ID1
gene
peak
d
805
336
Motif Count
Clover
ID1
TCF12 (p-value 1e-17)
e
Clover
ID1
CCNY
Cell Cycle
DLGAP4
Neuron Development
PKNOX2
Melanocyte Development
UNC5A
PRKCZ
IRF1-AS1
Interferon
f
PMEL
% Viable Cells (relative to untransfected)
Clover OE
ID1 OE
g
A375
Clover OE
TCF12 OE
h
Weighted number per fish
mcr:Empty
mcr:TCF12
Cancer precursor zone
i
% Fish Without Tumors
mcr:Empty
mcr:TCF12
Weeks Post-Fertilization