

Anatomic position determines oncogenic specificity in melanoma

Joshua M. Weiss^{1,2,3}, Miranda V. Hunter², Nelly M. Cruz², Arianna Baggiolini⁴, Mohita Tagore², Yilun Ma^{1,2,3}, Sandra Misale⁵, Michelangelo Marasco⁵, Theresa Simon-Vermot², Nathaniel R. Campbell^{1,2,6,7}, Felicity Newell⁸, James S. Wilmott⁹, Peter A. Johansson⁸, John F. Thompson^{9,10,11}, Georgina V. Long^{9,10,12}, John V. Pearson⁸, Graham J. Mann^{9,13,14}, Richard A. Scolyer^{9,10,11,15}, Nicola Waddell^{8,16}, Emily D. Montal², Ting-Hsiang Huang², Philip Jonsson^{17,18,19}, Mark T.A. Donoghue¹⁷, Christopher C. Harris¹⁷, Barry S. Taylor^{17,18,19}, Tianhao Xu⁶, Ronan Chaligné⁶, Pavel V. Shliha^{20,21}, Ronald Hendrickson²¹, Achim A. Jungbluth²², Cecilia Lezcano²², Richard Koche²³, Lorenz Studer⁴, Charlotte E. Ariyan²⁴, David B. Solit^{17,19,25}, Jedd D. Wolchok^{17,25,26,27}, Taha Merghoub²⁷, Neal Rosen⁵, Nicholas K. Hayward⁸, Richard M. White^{2,28*}

¹ Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program, New York, NY, 10065, USA

² Department of Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

³ Cell and Developmental Biology Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, 10065, USA

⁴ Developmental Biology, The Center for Stem Cell Biology, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

⁵ Molecular Pharmacology Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

⁶ Computational and Systems Biology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

⁷ Physiology, Biophysics & Systems Biology Graduate Program, Weill Cornell Graduate School of Medical Sciences, New York, NY, 10065, USA

⁸ QIMR Berghofer Medical Research Institute, Brisbane, QLD, 4006, Australia

⁹ Melanoma Institute Australia, The University of Sydney, Sydney, NSW, 2006, Australia

¹⁰ Faculty of Medicine and Health, The University of Sydney, Sydney, NSW, 2050, Australia

¹¹ Royal Prince Alfred Hospital, Sydney, NSW, 2050, Australia

¹² Royal North Shore Hospital, Sydney, NSW, 2065, Australia

¹³ John Curtin School of Medical Research, Australian National University, ACT, 2601, Australia

¹⁴ Centre for Cancer Research, Westmead Institute for Medical Research, The University of Sydney, Westmead, Sydney, NSW, 2528, Australia

¹⁵ New South Wales Health Pathology, Sydney, NSW, 2099, Australia

¹⁶ School of Medicine, The University of Queensland, Brisbane, QLD, 4072, Australia

¹⁷ Human Oncology & Pathogenesis Program, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

¹⁸ Department of Epidemiology and Biostatistics, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

¹⁹ Marie-Josée and Henry R. Kravis Center for Molecular Oncology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

²⁰ Department of Biochemistry and Molecular Biology, VILLUM Center for Bioanalytical Sciences, University of Southern Denmark, Odense 5230, Denmark

²¹ Microchemistry and Proteomics Core Facility, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

²² Department of Pathology, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

²³ Center for Epigenetics Research, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

²⁴ Department of Surgery, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

²⁵ Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY, 10065, USA

²⁶ Parker Institute for Cancer Immunotherapy, San Francisco, California, 94129, USA

²⁷ Weill Cornell Medicine, New York, NY, 10065, USA

²⁸ Lead Contact

*Correspondence: whiter@mskcc.org

Summary

Oncogenic alterations to DNA are not transforming in all cellular contexts^{1,2}. This may be due to pre-existing transcriptional programs in the cell of origin. Here, we define anatomic position as a major determinant of why cells respond to specific oncogenes. Cutaneous melanoma arises throughout the body, whereas the acral subtype arises on the palms of the hands, soles of the feet, or under the nails³. We sequenced the DNA of cutaneous and acral melanomas from a large cohort of human patients and found a specific enrichment for BRAF mutations in cutaneous melanoma but CRKL amplifications in acral melanoma. We modeled these changes in transgenic zebrafish models and found that CRKL-driven tumors predominantly formed in the fins of the fish. The fins are the evolutionary precursors to tetrapod limbs, indicating that melanocytes in these acral locations may be uniquely susceptible to CRKL. RNA profiling of these fin/limb melanocytes, compared to body melanocytes, revealed a positional identity gene program typified by posterior HOX13 genes. This positional gene program synergized with CRKL to amplify IGF signaling and drive tumors at acral sites. Abrogation of this CRKL-driven program eliminated the anatomic specificity of acral melanoma. These data suggest that the anatomic position of the cell of origin endows it with a unique transcriptional state that makes it susceptible to only certain oncogenic insults.

Introduction

During development, cells express lineage-specific and position-specific gene programs. These coordinated programs ensure that cells pattern appropriately and fulfill the unique role they have at a given anatomic location. Transcriptional programs unique to a particular anatomic site are referred to as positional identity⁴. The extent to which positional identity determines the response to oncogenes is unclear. Across cancers, the anatomic location of the tumor is associated with distinct molecular and clinical subtypes⁵⁻⁸. This is exemplified by melanoma, a cancer categorized by its anatomic origin at cutaneous sites (skin of face, chest, back), acral sites (glabrous skin of the palms and soles of the hands and feet^{3,8,9}), mucosal sites, or within the uveal tract. Even within the skin, melanocytes residing in different anatomic sites or substructures (i.e. follicular/interfollicular or scale/interscale regions) can have different susceptibilities to transformation^{9,10,11}. Highlighting these differences, acral melanoma, compared to other forms of cutaneous melanoma, has a unique genetic profile^{5,6,12-14}, a lower response rate to both targeted and immunotherapy, and a worse overall survival^{8,15,16}. Using data from human patients and transgenic zebrafish models, we identify genetic drivers unique to this anatomically restricted type of melanoma and find that the positional identity gene program in the cell-of-origin determines the competence to respond to those oncogenes.

Results

Acral genes drive tumors to acral sites

To discern genetic differences between acral and cutaneous melanoma, we performed DNA sequencing of 100 acral and 839 cutaneous melanoma patients using the MSK-IMPACT platform^{17,18}, a focused sequencing panel (Fig. 1a). We calculated the log2-fold ratio of genetic alterations to generate an “acral enrichment score” (Fig. 1a). Consistent with previous reports^{5,6}, we found BRAF mutations as the most common mutation in cutaneous melanoma (Extended Data Fig. 1a). In contrast, acral melanoma has a significantly higher frequency of copy number alterations (CNAs) and a lower overall mutational burden (Fig. 1a-b, Extended Data Fig. 1a-b), consistent with prior analyses^{5,6,12-14}. We also performed RNA-sequencing (RNA-seq) on an independent set of 61 acral and 53 cutaneous patient samples. Amplification of CRKL and GAB2 were amongst the top acral enriched genes identified by a combined analysis of DNA and RNA-sequencing (Fig. 1b-c, Extended Data Fig. 1c). These were of particular interest to us since CRKL and GAB2 are both oncogenic signaling proteins that form a complex with receptor tyrosine kinases (RTKs) to amplify downstream signaling (Extended Data Fig. 1d)¹⁹. They have

previously been implicated in other cancers²⁰⁻²², but limited investigation has been performed in melanoma^{23,24}. Other frequently altered genes included NF1 and TERT (Fig. 1a-b, Extended Data Fig. 1a-b), which significantly co-occur with alterations in CRKL and GAB2 across cancer (Extended Data Fig. 1e). This led to the hypothesis that alterations in CRKL, GAB2, NF1, and TERT may synergize to drive acral compared to cutaneous melanoma (Extended Data Fig. 1d). For example, we identified an acral melanoma patient at Sloan Kettering who had amplification of CRKL and GAB2, deletion of NF1, and an activating promoter mutation in TERT in both primary and metastatic sites (Extended Data Fig. 1f-h).

Since there are few available animal models of acral melanoma²⁵, we used a rapid transgenic zebrafish system to model these potential drivers *in vivo*. Using the melanocyte-specific *mitfa* promoter, we created mosaic (F0) transgenic animals expressing acral melanoma drivers (CRKL, GAB2, TERT and NF1-/-) and then compared this to a previously developed cutaneous melanoma model (BRAF^{V600E};p53-/-) or to wildtype melanocytes (Fig. 1d, Extended Data Fig. 2a-f). We developed criteria to detect tumors in both acral and cutaneous melanoma models by imaging and histology (Supplemental Table 1). Melanocytes in all models were labeled with *mitfa*:GFP, and darkly pigmented cells with dendritic morphology on H&E, RNA expression of *tyrp1a*, and immunofluorescence (IF) for *sox10* indicate that tumors generated by both models are melanocyte lineage (Fig. 1d, Extended Data Fig. 2e,g, Extended Data Fig. 3). We validated the expression and knockout of the transgenes using a combination of genotyping, CRISPR-seq, qPCR, RNA-seq, western blot, and IHC (Extended Data Fig. 2a-g, Extended Data Fig. 3). While wildtype melanocytes never formed melanomas, we found that 68% of the animals expressing the acral genes resulted in melanoma and 100% of the animals expressing the cutaneous genes resulted in melanoma by 1 year (Fig. 1f).

Extensive research across fields ranging from paleontology to genetics has demonstrated an evolutionary link between the morphogenesis of fish fins and tetrapod limbs, which include the acral sites of the hands and feet^{26,27}. Although fins and limbs are structurally distinct, the genes involved in fin and limb development are well conserved²⁶. This led us to ask whether the acral versus cutaneous driver genes would yield differences in anatomic distribution of the tumors. We monitored the fish and calculated the relative proportion of tumors arising on the head, body, and fins. To visually depict the anatomic distribution of tumors, we utilized the ternary plot, which compresses 3-dimensional categorical data into 2-dimensional space (Fig. 1e, Extended Data Fig. 2h)²⁸. This revealed that the acral model developed a significantly higher proportion of fin tumors than the cutaneous model (53% versus 30%, $p < 0.0001$) (Fig. 1e, Extended Data Fig. 2h, Supplemental Table 3).

Because mosaic (F0) transgenics can be variable, we also created stable “WT” (*mitfa*:GFP), acral melanoma (*mitfa*:CRKL;GAB2;TERT;NF1-/-), and cutaneous melanoma (*mitfa*:BRAF^{V600E};p53-/-) lines and assessed their phenotypes in the same way (Extended Data Fig. 4a-c). In the acral stable lines, the preponderance of fin tumors was even more striking. Whereas in the mosaic acral fish 53% of tumors (n=110/206) arose on the fins, we found that 84.2% (n=96/114) of tumors from the acral stable line were fin tumors (Extended Data Fig. 4c, Supplemental Table 3). In contrast, only 39.7% (n=77/194) of tumors arose on the fins in the cutaneous stable line (Extended Data Fig. 4c, Supplemental Table 3). The anatomic position of the tumors did not shift over the 1-year observation period. At 3 days post-fertilization, melanocytes in the acral melanoma stable line demonstrated a greater total melanocyte area in the tailfin than melanocytes in the wild-type stable line (Extended Data Fig. 2i). By 6 weeks post-fertilization the acral melanoma model developed fin hyperpigmentation and a significantly greater expansion of the melanocyte population in the fins than in the body (Extended Data Fig. 4d-g).

CRKL amplification drives acral melanoma

To determine which of the acral driver genes were sufficient for fin positioning, we expressed each one individually (Fig. 2a). CRKL was the only genetic driver sufficient to form tumors without any additional genetic alterations (Fig. 2b-c), with 62% of CRKL-alone tumors in the fins (Fig. 2d-e, Supplemental Table 3) indicating that fin melanocytes were more efficiently transformed by CRKL compared to other locations. Tumor initiation in CRKL-only tumors was slower compared to all 4 genes together, suggesting that GAB2/TERT/NF1-/- accelerated progression (Fig. 2c). We then investigated whether any of the 4 putative genetic drivers were necessary for tumor progression by removing CRKL, GAB2, TERT, or NF1 loss from the full 4-gene acral model (Fig. 2f). While tumors still formed after withdrawal of GAB2, TERT, or NF1 loss, tumors never formed after withdrawal of CRKL (Fig. 2g-i). Withdrawal of GAB2, TERT, or NF1 loss from the acral model still demonstrated fin specificity (Fig. 2h-i, Supplemental Table 3). To ensure these results were not an artifact of mosaic F0 transgenesis, we also created a stable line for CRKL alone (*mitfa*:CRKL) (Extended Data Fig. 4a). Similar to what we observed in the acral stable line, the enrichment of fin tumors was even greater in the CRKL stable line. While 62% (n=37/60) of tumors arose on the fins in the mosaic CRKL fish, we found that 75.8% (n=25/33) of tumors from CRKL-only stable transgenic fish were fin tumors (Extended Data Fig. 4b-c, Supplemental Table 3). Together, these data demonstrate that CRKL is a key driver of acral melanoma, and that fin melanocytes (analogous to human hand/feet melanocytes) are more susceptible to this oncogene than melanocytes at other locations.

Melanocytes have a positional identity

We hypothesized that this susceptibility was due to intrinsic differences in the gene program of these melanocytes, and that this was due to anatomic positioning. We performed RNA-sequencing of body versus fin melanocytes, along with surrounding microenvironment, using a combination of skin dissection and FACS sorting for GFP (Fig. 3a). As expected, the GFP+ melanocyte population (compared to the GFP- population) showed enrichment for melanocyte markers such as *mitfa*, *pmela*, *tyr*, and *sox10* as well as the transgenes (GFP, CRKL, etc.) (Fig. 3d, Extended Data Fig. 5a-c). Unsupervised hierarchical clustering showed that the samples clustered first by cell lineage (melanocyte or not), second by anatomic location (fin vs body), and only third by genotype (CRKL vs WT) (Fig. 3b). Along the same lines Principal Component Analysis (PCA) showed clustering by cell lineage on PC1 and by anatomic position on PC2 (Extended Data Fig. 5d). This data suggests that anatomic location contributes a greater amount of transcriptional variation than the genotype and can shape the response to oncogenes such as CRKL.

We performed pathway analysis comparing fin versus body melanocytes, which found that 11 out of 30 of the top pathways in fin melanocytes were related to limb development or anatomic position (Fig. 3c, Extended Data Fig. 5e-f, Supplemental Table 4). This indicated that the differences in the transcriptional programs of fin versus body melanocytes were largely associated with positional identity. We found a significant upregulation of limb-specific HOX genes in the fin melanocytes (*hoxa13a*, *hoxa13b*, *hoxb13a*, *hoxc13a*, *hoxc13b*, *hoxd13a*) whereas the body melanocytes upregulated axial-specific HOX genes (*hoxb5a*, *hoxb6a*, *hoxb7a*, *hoxb8a*, *hoxb9a*) (Fig. 3d, Extended Data Fig. 5g-h). Posterior HOX13 genes are master regulators of limb development²⁶, whereas the more anterior HOXB5-9 regulate many segments of axial body development²⁹. Because FACS sorting followed by bulk RNA-seq could result in the presence of contaminating microenvironmental cells, we validated our bulk RNA-seq by performing single-cell RNA-sequencing (Extended Data Fig. 6a-b). Similar to bulk sequencing, there was a significant upregulation of posterior HOX13 genes such as *hoxc13b* in fin versus body melanocytes (Extended Data Fig. 6c-d). Upregulation of HOX13 genes was

also observed in other fin cell types as well, including fibroblasts, keratinocytes, and xanthophores (Extended Data Fig. 6e). Clustering based on expression of all detected HOX genes from bulk RNA-seq separated samples by anatomic location, demonstrating that positional identities of fin versus body melanocytes are associated with a distinct HOX code (Extended Data Fig. 6f).

Human melanoma has a positional identity

To understand if these anatomic position gene programs were conserved in patients, we performed RNA-seq on human acral versus cutaneous melanoma (Fig. 4a-b). Similar to the zebrafish, 6 out of 30 of the top pathways in acral melanoma were related to limb development or positional identity (Fig. 4b, Extended Data Fig. 7a-b, Supplemental Table 5). Human acral melanoma also had significantly greater expression of HOXA13, HOXB13, HOXD13 as well as other key homeobox transcription factors that regulate limb development, such as TBX4, DLX4, and HAND2 (Fig. 4c-d)³⁰. PCA of all samples based on just HOX gene expression shows clustering by melanoma subtype (Extended Data Fig. 7c). To validate HOXB13 expression on the protein level, we used an acral and cutaneous melanoma tissue microarray (TMA) with n=32 acral patient samples and n = 14 cutaneous patient samples (Fig. 4e, Extended Data Fig. 7d-h). 90% of acral samples and 57.1% of cutaneous samples were positive for HOXB13 (Fig. 4e, Extended Data Fig. 7d-e). Acral tumors staining positively for HOXB13 included both primary (n=8/9) and metastatic (n=19/21) samples, indicating that positional identity may be retained at distant metastatic sites (Fig. 4e, Extended Data Fig. 7d-e). N = 4/7 of HOXB13 positive cutaneous melanoma samples were originally from distal limbs, suggesting that some limb cutaneous melanomas may also have an “acral-like” positional identity. In contrast to the RNA-seq data which showed higher expression of CRKL in acral melanoma (Fig 1c), IHC staining showed a similar expression pattern with 53.3% of acral samples and 64.3% of cutaneous samples staining positive (Extended Data Fig. 7d-h). This discrepancy may reflect differences in post-transcriptional regulation of CRKL between the two subtypes or a limitation of antibody based approaches in measuring expression from disparate metastatic sites, which may be influenced by local microenvironmental factors. Because many genes other than HOX genes are differentially expressed between acral and cutaneous melanoma, we also performed HOMER motif enrichment analysis on the RNA-seq data to identify which transcription factors regulate the broader set of differentially expressed genes. A motif for HOXB13 was one of the top motifs indicating that HOX genes may regulate many of the transcriptional differences between the two subtypes (Fig. 4f).

HOX13 genes regulate IGF signaling

We hypothesized that oncogenic drivers such as CRKL could synergize with transcriptional targets of limb-specific HOX13 genes to drive acral melanoma. To identify these targets, we analyzed published HOX13 ChIP-seq data from developing mouse limb bud tissue³¹ and used Cistrome-GO³² to perform pathway analysis for genes with promoters or enhancers occupied by HOX13 (Fig. 5a). We found that IGF/insulin signaling, a key regulator of limb development and regeneration, was amongst the top bound pathways³³ (Fig. 5b, Extended Data Fig. 8a-c), with binding of HOXA13 and HOXD13 at promoters of both IGF1 and IGF2. These HOX13 binding sites were associated with H3K27 acetylation peaks, indicating that HOX13 genes may regulate IGF expression (Extended Data Fig. 8d-e). Consistent with these data, re-examination of the fin versus body melanocyte bulk RNA-seq showed elevated expression of *igf1*, *igf2a*, and *igf2b* in fin compared to body melanocytes (Extended Data Fig. 8f-h).

Since the mouse ChIP-seq experiment was performed on whole limb bud tissue, we also wanted to determine whether this HOX13/IGF relationship specifically existed in human melanoma cells. We performed western blots to identify HOX13+ patient-derived melanoma cell

lines (SKMEL-1176 and SKMEL-1206 which are from known acral patients, and SKMEL-1088 which is from an unknown site) (Extended Data Fig. 8p). We then performed CUT&RUN using an antibody for HOXA13 on those 3 melanoma cell lines. HOMER motif analysis correctly found the HOX13 motif to be enriched in the CUT&RUN data from all 3 lines (Extended Data Fig. 8i). Cistrome-GO pathway analysis identified insulin receptor substrate binding and other pathways related to insulin/IGF signaling were also amongst the top 20 enriched pathways (8j). Examination of individual genes showed significant HOXA13 binding at the IGF1R promoter, along with a peak approximately 2.8kb 5' to the IGF2 promoter as well as a peak 50kb 3' to IGF2 (Extended Data Fig. 8k-l). Other IGF signaling related genes, such as IRS1, IRS2, and IGFBP3, demonstrated peaks near the transcription start site, suggesting HOX13 regulation of multiple components of the pathway (Extended Data Fig. 8m-o). In order to determine whether these peaks regulate expression of IGF2 and IGF1R, we performed siRNA knockdown of HOX13 genes in the SKMEL-1176 and SKMEL-1206 acral lines and validated knockdown by western blot (Fig. 5c-d, Extended Data Fig. 8q-r). This resulted in a significant decrease in IGF2 ligand in both cell lines (Fig. 5c-d), which was orthogonally validated by ELISA for secreted IGF2 (Fig. 5e). HOX13 knockdown also caused a decrease in pIGF1R and IGF1R expression although the effect was more modest and varied between cell lines (Extended Data Fig. 8q-r). Supporting the CUT&RUN data, HOXB13^{HI} acral cells had higher levels of pIGF1R and total IGF1R, indicating greater overall IGF pathway activation (Extended Data Fig. 8p). We also observed pathway activation in patient tumors, whereby 96.7% of samples from the acral melanoma TMA stained positive for pIGF1R (Extended Data Fig. 7d-h). Together this data suggests a direct role for HOX13 in regulating IGF signaling in melanoma.

CRKL amplifies HOX13/IGF signaling

CRKL is an adapter protein that builds protein complexes that then mediate downstream signaling through pathways such as IGF1R/PI3K^{34,35}, raising the hypothesis that CRKL was amplifying the HOX13/IGF axis to promote acral melanoma. To test this, we performed a phospho-RTK array of in WM3918 cells with or without overexpression of CRKL (Extended Data Fig. 9a). CRKL overexpression specifically increased the phosphorylation of the insulin and IGF1 receptors (INSR/IGF1R) (Extended Data Fig. 9b). We confirmed this effect in 2/3 additional cell lines (SKMEL-1176, and 293T) in which CRKL overexpression increased pIGF1R expression (Extended Data Fig. 9b). To characterize the protein complex formed by CRKL and identify its binding partners, we performed IP-mass spec in WM3918 human melanoma cells overexpressing CRKL with a V5-tag (Extended Data Fig. 9c-f). CRKL bound multiple subunits of PI3K, including PIK3CB, PIK3CD PIK3R2, which are the major downstream mediators of IGF/IGF1R signaling (Extended Data Fig. 9d-f, Supplemental Table 8). In addition, we also pulled down many GTPase-activating proteins (GAPs) of Rac/Rho GTPases, such as ARAP1³⁶, ASAP1³⁶, ARHGAP32, and guanine nucleotide exchange factors (GEFs) of Rac/Rho, such as C3G, DOCK1³⁷, DOCK5, and the DOCK co-activators, ELMO1³⁷ and ELMO2 (Extended Data Fig. 9d-f, Supplemental Table 8). DOCK1 and ELMO1 have been previously shown to activate PI3K signaling specifically through PIK3CB³⁷ and ARAP1³⁶ is activated by PIP3, the product of PI3K. In contrast to this heavily enriched PI3K-centered mechanism, we found relatively few members of the MAPK pathway (SOS1/SOS2) (Extended Data Fig. 9d-f, Supplemental Table 8). To further examine the effect of CRKL overexpression on signaling, we performed detailed histological analysis of our zebrafish transgenic models (Extended Data Fig. 3). We stained n=9 fish (3 wild-type, 3 acral, 3 cutaneous) using H&E as well as antibodies for GFP, CRKL, pIGF1R, pERK, and pS6 (Extended Data Fig. 3). In the acral zebrafish tumors, there was a clear and striking overlap between GFP (indicating the presence of the transgenic melanocytes), CRKL, pIGF1R, and pS6 (Extended Data Fig. 3). Interestingly, the acral fish tumors also stained for pERK (Extended Data Fig. 3), yet we saw little evidence for sensitivity to MAPK inhibition (as discussed below). In contrast, the cutaneous fish tumors were negative for

CRKL, faintly positive for pIGF1R (similar to control animals), strongly positive for pERK (as expected), and only 1 animal was positive for pS6 (Extended Data Fig. 3). Given CRKL's well known ability to mediate RTK signaling, these data imply that somatic amplification of CRKL in acral melanoma synergizes with the pre-existing HOX13/IGF positional identity program, and this is likely mediated via direct binding to PI3K family members.

Acral melanoma depends on CRKL/HOX13/IGF

This data raised the possibility that IGF signaling, downstream of the intrinsic HOX13 program, might synergize with CRKL and explain why that oncogene is enriched at acral sites. We utilized the increased tailfin melanocyte area in CRKL fish for high throughput genetic and pharmacologic experimentation (Fig. 5f-j). We created a CRISPR vector which allowed us to knockout all 6 HOX13 genes (*hoxa13a*, *hoxa13b*, *hoxb13a*, *hoxc13a*, *hoxc13b*, *hoxd13a*) or 3 IGF genes (*igf1*, *igf2a*, *igf2b*) in a melanocyte-specific manner by co-injecting a *mitfa*:Cas9 plasmid (Fig. 5f, Extended Data Fig. 10a). All CRISPR targets were confirmed through Surveyor assays (Extended Data Fig. 10b-d). As a positive control, we first injected sgRNAs against CRKL itself and found that this decreased fin melanocyte expansion (Fig. 5g,i). sgRNAs against the fish IGF genes (*igf1*, *igf2a*, *igf2b*) caused a significant decrease in the total area occupied by tailfin melanocytes, indicating that IGF signaling is necessary for CRKL-mediated melanocyte expansion at acral sites (Fig. 5g,i). Similarly, knockout of all 6 HOX13 genes caused a significant decrease in tailfin melanocyte expansion (Fig. 5h,j) indicating that HOX13 genes are necessary for CRKL-mediated melanocyte expansion at acral sites. As a negative control, we used a sgRNA to knockout *hoxb7a* (an anterior HOX gene not expressed in the fin melanocytes), and consistent with our hypothesis, this had no effect on the CRKL-induced phenotype (Fig. 5h,j). While we did not test the contribution of HOX13 genes to tumor incidence in adult fish, we did investigate the contribution of insulin/IGF signaling by overexpressing a previously validated dominant negative IRS2 (dnIRS2) to block insulin/IGF signaling in the CRKL-driven acral melanoma model^{38,39} (Fig. 5k-l). While this modestly slowed tumor initiation (Extended Data Fig. 10e, Supplemental Table 3), most importantly it led to a robust loss of fin specificity in the acral melanoma model (Fig. 5k-l, Supplemental Table 3).

Finally, we tested whether pharmacologic inhibition of the IGF/PI3K pathway would block the fin phenotype of the acral model. Treatment with INSR/IGF1R antagonists (BMS-754807 and NVP-AEW541) led to a significant decrease in the total melanocyte area in the tailfin³⁹ (Extended Data Fig. 11a). Similarly, PI3K inhibition with LY294002 caused a robust decrease in the total melanocyte area in the tailfin (Extended Data Fig. 11b). In contrast, treatment with the RAF/MEK inhibitor CH5126766 (Extended Data Fig. 11b) or MEK inhibitors trametinib, pimasetib, refametinib, or the SOS1 inhibitor, BI-3406, only led to a very modest or insignificant rescue of the tailfin phenotype (Extended Data Fig. 11b-d). This data is consistent with our observation that only a few MAPK members (SOS1/SOS2) were enriched in our IP-mass spec data, in contrast to the multiple PI3K family members, indicating a greater dependence on PI3K signaling (Extended Data Fig. 9c-f). Another possibility is that the chromatin state of acral melanocytes is not optimal for responding to activation of MAPK signaling, as we recently demonstrated⁴⁰. Together, these data demonstrate that CRKL-mediated transformation depends on a HOX13/IGF/PI3K-associated limb positional gene program (Extended Data Fig. 12).

Discussion

What are the critical factors necessary for a cell to become a cancer? Many studies have documented phenotypically normal cells with classic oncogenic driver mutations^{1,2}, suggesting that tumorigenesis depends not just on mutations but also on a particular cell state. Here we demonstrate that positional identity, a transcriptional program linked to a particular anatomic location⁴, is a key determinant of the transforming potential of oncogenes in melanoma. Through

a combination of DNA and RNA-sequencing of human patients, we identified CRKL amplification as specifically enriched in acral melanoma, a subtype defined by its anatomic position on the hands and feet³. Due to the evolutionary relationship between zebrafish fins and human hands and feet^{26,27}, we were able to establish the first *in vivo* model that shows that acral melanocytes were more susceptible to CRKL transformation than melanocytes at other anatomic sites. Mechanistically we found that fin melanocytes have a limb positional identity defined by HOX13 genes, which drive higher levels of IGF signaling. We also observed this in human acral melanoma. CRKL potentiates this HOX13/IGF axis by forming a protein complex with multiple PI3K family members, the major downstream signaling pathway enacted by IGF/IGF1R. This leads to a model in which somatic amplifications of CRKL synergize with an intrinsic limb positional identity program to drive melanoma most efficiently at acral anatomic sites (Extended Data Fig. 12).

Our genomic analysis of acral and cutaneous melanoma is consistent with previous analyses, which find a high frequency of gene amplifications for GAB2/PAK1, CCND1, TERT, MDM2, and CDK4^{5,6,12-14}. However, our analysis is the first to report amplification of CRKL as a driver of acral melanoma. Supporting our results, there is an independent acral genomic analysis in press also finding a relatively high frequency of CRKL amplifications⁴¹.

This study provides a conceptual framework to understand previous work that has documented spatial differences in susceptibility to tumorigenesis. Regional differences have been observed in melanocytes from different locations⁹, and between follicular and interfollicular melanocytes in their ability to respond to oncogenic drivers^{10,11}. Pigmented melanocytes in the scale region of the mouse tail, but not the interscale region, give rise to tumors despite both populations expressing BRAF^{V600E} and loss of PTEN¹¹. While our study demonstrates how cell intrinsic positional programs can determine oncogenic specificity in tumorigenesis, we do not rule out the important role of the microenvironment in this process. Our scRNA-seq data suggests that not just melanocytes, but also microenvironmental cells in the fins express higher levels of HOX13 genes (Extended Data Fig. 6e). Whether the positional identity of microenvironmental cells also contributes to tumorigenesis is an area of future investigation.

Positional identity can address a longstanding question, which is why do cancers arising at different locations often possess distinct genetic and clinical characteristics? Examples include not just other skin cancers like basal and squamous cell carcinoma², but also colon cancer⁷, gastric cancer⁴², cholangiocarcinoma⁴³, and glioma⁴⁴. While the HOX13-IGF-CRKL mechanism we propose may be specific to subsets of patients with acral melanoma, it is likely that oncogenes other than CRKL can interact with the HOX13 program. For example, we and others have also identified significant enrichment in genes such as PAK1, CDK4 and YAP1 in acral melanoma^{12,13} (Fig.1a-b), and how they intersect with positional identity is an important area for future study. The concept that oncogenes interact with location-specific programs in the cell of origin may apply to a variety of cancer types, which may lead to new opportunities in the treatment of cancer. Future studies to comprehensively characterize the positional dependencies of oncogenes will be a valuable resource in elucidating the origins of cancer and treating cancer based on its positional identity.

Main text references

- 1 Tang, J. et al. The genomic landscapes of individual melanocytes from human skin. *Nature*, doi:10.1038/s41586-020-2785-8 (2020).
- 2 Fowler, J. C. et al. Selection of oncogenic mutant clones in normal human skin varies with body site. *Cancer Discovery*, doi:10.1158/2159-8290.Cd-20-1092 (2020).
- 3 Reed, R. *New concepts in surgical pathology of the skin*. 89-90 (Wiley, 1976).

- 4 Wang, K. C., Helms, J. A. & Chang, H. Y. Regeneration, repair and remembering identity: the three Rs of Hox gene expression. *Trends Cell Biol* 19, 268-275, doi:10.1016/j.tcb.2009.03.007 (2009).
- 5 Curtin, J. A. et al. Distinct Sets of Genetic Alterations in Melanoma. *N Engl J Med* 353 (2005).
- 6 Hayward, N. K. et al. Whole-genome landscapes of major melanoma subtypes. *Nature* 545, 175-180, doi:10.1038/nature22071 (2017).
- 7 Petrelli, F. et al. Prognostic Survival Associated With Left-Sided vs Right-Sided Colon Cancer: A Systematic Review and Meta-analysis. *JAMA Oncol* 3, 211-219, doi:10.1001/jamaoncol.2016.4227 (2017).
- 8 Rabbie, R., Ferguson, P., Molina-Aguilar, C., Adams, D. J. & Robles-Espinoza, C. D. Melanoma subtypes: genomic profiles, prognostic molecular markers and therapeutic possibilities. *J Pathol* 247, 539-551, doi:10.1002/path.5213 (2019).
- 9 Belote, R. L. et al. Human melanocyte development and melanoma dedifferentiation at single-cell resolution. *Nat Cell Biol* 23, 1035-1047, doi:10.1038/s41556-021-00740-8 (2021).
- 10 Moon, H. et al. Melanocyte Stem Cell Activation and Translocation Initiate Cutaneous Melanoma in Response to UV Exposure. *Cell Stem Cell* 21, 665-678 e666, doi:10.1016/j.stem.2017.09.001 (2017).
- 11 Kohler, C. et al. Mouse Cutaneous Melanoma Induced by Mutant BRAF Arises from Expansion and Dedifferentiation of Mature Pigmented Melanocytes. *Cell Stem Cell* 21, 679-693 e676, doi:10.1016/j.stem.2017.08.003 (2017).
- 12 Newell, F. et al. Whole-genome sequencing of acral melanoma reveals genomic complexity and diversity. *Nat Commun* 11, 5259, doi:10.1038/s41467-020-18988-3 (2020).
- 13 Yeh, I. et al. Targeted Genomic Profiling of Acral Melanoma. *J Natl Cancer Inst* 111, 1068-1077, doi:10.1093/jnci/djz005 (2019).
- 14 Liang, W. S. et al. Integrated genomic analyses reveal frequent TERT aberrations in acral melanoma. *Genome Res* 27, 524-532, doi:10.1101/gr.213348.116 (2017).
- 15 Klemen, N. D. et al. Survival after checkpoint inhibitors for metastatic acral, mucosal and uveal melanoma. *J Immunother Cancer* 8, doi:10.1136/jitc-2019-000341 (2020).
- 16 Shoushtari, A. N. et al. The efficacy of anti-PD-1 agents in acral and mucosal melanoma. *Cancer* 122, 3354-3362, doi:10.1002/cncr.30259 (2016).
- 17 Cerami, E. et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* 2, 401-404, doi:10.1158/2159-8290.CD-12-0095 (2012).
- 18 Zehir, A. et al. Mutational landscape of metastatic cancer revealed from prospective clinical sequencing of 10,000 patients. *Nat Med* 23, 703-713, doi:10.1038/nm.4333 (2017).
- 19 Luo, L. Y. & Hahn, W. C. Oncogenic Signaling Adaptor Proteins. *J Genet Genomics* 42, 521-529, doi:10.1016/j.jgg.2015.09.001 (2015).
- 20 Bentires-Alj, M. et al. A role for the scaffolding adapter GAB2 in breast cancer. *Nat Med* 12, 114-121, doi:10.1038/nm1341 (2006).
- 21 Cheung, H. W. et al. Amplification of CRKL induces transformation and epidermal growth factor receptor inhibitor resistance in human non-small cell lung cancers. *Cancer Discov* 1, 608-625, doi:10.1158/2159-8290.CD-11-0046 (2011).
- 22 Hemmeryckx, B. et al. Crkl enhances leukemogenesis in BCR/ABL P190 transgenic mice. *Cancer Research* 61, 1398-1405 (2001).
- 23 Chernoff, K. A. et al. GAB2 amplifications refine molecular classification of melanoma. *Clin Cancer Res* 15, 4288-4291, doi:10.1158/1078-0432.CCR-09-0280 (2009).
- 24 Horst, B. et al. Gab2-mediated signaling promotes melanoma metastasis. *Am J Pathol* 174, 1524-1533, doi:10.2353/ajpath.2009.080543 (2009).
- 25 Eshiba, S. et al. Stem cell spreading dynamics intrinsically differentiate acral melanomas from nevi. *Cell Rep* 36, 109492, doi:10.1016/j.celrep.2021.109492 (2021).

- 26 Nakamura, T., Gehrke, A. R., Lemberg, J., Szymaszek, J. & Shubin, N. H. Digits and fin rays share common developmental histories. *Nature* 537, 225-228, doi:10.1038/nature19322 (2016).
- 27 Shubin, N. H., Daeschler, E. B. & Jenkins, F. A., Jr. The pectoral fin of *Tiktaalik roseae* and the origin of the tetrapod limb. *Nature* 440, 764-771, doi:10.1038/nature04637 (2006).
- 28 Xu, B., Feng, X. & Burdine, R. D. Categorical data analysis in experimental biology. *Dev Biol* 348, 3-11, doi:10.1016/j.ydbio.2010.08.018 (2010).
- 29 Philippidou, P. & Dasen, J. S. Hox genes: choreographers in neural development, architects of circuit organization. *Neuron* 80, 12-34, doi:10.1016/j.neuron.2013.09.020 (2013).
- 30 Petit, F., Sears, K. E. & Ahituv, N. Limb development: a paradigm of gene regulation. *Nat Rev Genet* 18, 245-258, doi:10.1038/nrg.2016.167 (2017).
- 31 Sheth, R. et al. Distal Limb Patterning Requires Modulation of cis-Regulatory Activities by HOX13. *Cell Rep* 17, 2913-2926, doi:10.1016/j.celrep.2016.11.039 (2016).
- 32 Li, S. et al. Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. *Nucleic Acids Res* 47, W206-W211, doi:10.1093/nar/gkz332 (2019).
- 33 Chablais, F. & Jazwinska, A. IGF signaling between blastema and wound epidermis is required for fin regeneration. *Development* 137, 871-879, doi:10.1242/dev.043885 (2010).
- 34 Dhupkar, P., Zhao, H., Mujoo, K., An, Z. & Zhang, N. Crk II silencing down-regulates IGF-IR and inhibits migration and invasion of prostate cancer cells. *Biochem Biophys Res* 8, 382-388, doi:10.1016/j.bbrep.2016.10.009 (2016).
- 35 Zhang, J. et al. CRKL Mediates p110beta-Dependent PI3K Signaling in PTEN-Deficient Cancer Cells. *Cell Rep* 20, 549-557, doi:10.1016/j.celrep.2017.06.054 (2017).
- 36 Tanna, C. E., Goss, L. B., Ludwig, C. G. & Chen, P. W. Arf GAPs as Regulators of the Actin Cytoskeleton-An Update. *Int J Mol Sci* 20, doi:10.3390/ijms20020442 (2019).
- 37 Fritsch, R. et al. RAS and RHO families of GTPases directly regulate distinct phosphoinositide 3-kinase isoforms. *Cell* 153, 1050-1063, doi:10.1016/j.cell.2013.04.031 (2013).
- 38 Ye, L., Robertson, M. A., Mastracci, T. L. & Anderson, R. M. An insulin signaling feedback loop regulates pancreas progenitor cell differentiation during islet development and regeneration. *Dev Biol* 409, 354-369, doi:10.1016/j.ydbio.2015.12.003 (2016).
- 39 Zhang, Y. M. et al. Distant Insulin Signaling Regulates Vertebrate Pigmentation through the Sheddase Bace2. *Dev Cell* 45, 580-594 e587, doi:10.1016/j.devcel.2018.04.025 (2018).
- 40 Baggiolini, A. et al. Developmental chromatin programs determine oncogenic competence in melanoma. *Science* 373, eabc1048, doi:10.1126/science.abc1048 (2021).
- 41 Farshidfar, F. et al. Integrative molecular and clinical profiling of acral melanoma identifies LZTR1 as a key tumor promoter and therapeutic target. *bioRxiv*, doi:10.1101/2021.04.20.440286 (2021).
- 42 Kim, K. et al. Clinicopathologic characteristics of early gastric cancer according to specific intragastric location. *BMC Gastroenterol* 19, 24, doi:10.1186/s12876-019-0949-5 (2019).
- 43 Razumilava, N. & Gores, G. J. Cholangiocarcinoma. *The Lancet* 383, 2168-2179, doi:10.1016/s0140-6736(13)61903-0 (2014).
- 44 Tang, Q. et al. Anatomic mapping of molecular subtypes in diffuse glioma. *BMC Neurol* 17, 183, doi:10.1186/s12883-017-0961-8 (2017).

Acknowledgements

J.M.W. was supported by the Kirschstein-NRSA predoctoral fellowship (F30) from the NIH under award number F30CA236442 and by a predoctoral fellowship (T32) by the Cell and Developmental Biology Program at Weill Cornell Graduate School supported by the NIH under award number T32GM008539. N.R.C was supported by the Kirschstein-NRSA predoctoral fellowship (F30) from the NIH under award number F30CA220954. J.M.W., Y.M., and N.R.C were supported by a Medical Scientist Training Program grant from the NIH under award number T32GM007739. A.B. was supported by the Swiss National Science Foundation Postdoc mobility fellowships P2ZHP3_171967 and P400PB_180672. M.T. was supported by the Marie-Josée Kravis Women in Science Endeavor Postdoctoral Fellowship. E.D.M. was supported by the Predoctoral to Postdoctoral Fellow Transition Award (F99/K00) from the NIH under award number K00CA223016. R.A.S., G.V.L., J.F.T., G.J.M. and N.K.H. are supported by an Australian National Health and Medical Research Council Program Grant. R.A.S., N.K.H. and G.V.L. are supported by the Australian National Health and Medical Research Council Fellowship program. The authors acknowledge support from colleagues at their respective institutions and support from the Cameron Family and the Ainsworth Foundation is also gratefully acknowledged. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA008748, Ludwig Collaborative and Swim Across America Laboratory, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; Parker Institute for Cancer Immunotherapy, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; Department of Medicine, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA; Weill Cornell Medicine, New York, NY 10065, USA. This work was supported by awards from the Melanoma Research Alliance, The Debra and Leon Black Family Foundation, NIH Research Program Grants R01CA229215 and R01CA238317, NIH Director's New Innovator Award DP2CA186572, The Pershing Square Sohn Foundation, The Mark Foundation, The American Cancer Society, The Alan and Sandra Gerry Metastasis Research Initiative at the Memorial Sloan Kettering Cancer Center, The Harry J. Lloyd Foundation, Consano and the Starr Cancer Consortium (all to R.M.W.). Thank you to Lambros T. Koufariotis for technical assistance with the alignment of human RNA-seq data. We acknowledge the use of the Integrated Genomics Operation Core, funded by the NCI Cancer Center Support Grant (CCSG, P30 CA08748), Cycle for Survival, and the Marie-Josée and Henry R. Kravis Center for Molecular Oncology.

Author Contributions

Conceptualization: J.M.W. and R.M.W.; Transgenic fish: J.M.W., T.S.V.; Zebrafish RNA-seq analysis: J.M.W., M.V.H., N.R.C.; Zebrafish scRNA-seq: J.M.W., R.C., M.V.H., Y.M., T.X. IP-mass spec: J.M.W., P.V.S., R.H., S.M.; Human pathology specimen analysis: J.M.W., A.A.J., C.L., C.E.A.; ChIP-seq/C&R analysis: M.T., R.K.; Melanocyte sorting: J.M.W., Y.M., E.D.M.; Zebrafish drug treatments: J.M.W., N.M.C., S.M., N.R.; Human CUT&RUN: A.B., L.S.; Human cell line overexpression and knockdown studies: J.M.W., T-H.H, S.M., M.M., N.R.; Human RNA-seq analysis: J.M.W., F.N., J.S.W., P.A.J., J.F.T., G.V.L., J.V.P, G.J.M., R.A.S., N.W., N.K.H. Human acral cell line supply: J.D.W., T.M., C.E.A.; IMPACT analysis: J.M.W., B.S.T, D.B.S., P.J., M.T.A.D., C.C.H.; Manuscript writing: J.M.W., R.M.W.

Declaration of Interests

S.M. consulted for Boehringer-Ingelheim. J.F.T. has received honoraria for advisory board participation from BMS Australia, MSD Australia, GSK and Provectus Inc, and travel support from GSK and Provectus Inc. R.A.S. has received fees for professional services from Qbiotics, Novartis, MSD Sharp & Dohme, NeraCare, AMGEN Inc., Bristol-Myers Squibb, Myriad Genetics, GlaxoSmithKline. G.V.L. is consultant advisor for Aduro Biotech Inc, Amgen Inc, Array Biopharma inc, Boehringer Ingelheim International GmbH, Bristol-Myers Squibb, Evaxion

Biotech A/S, Hexel AG, Highlight Therapeutics S.L., Merck Sharpe & Dohme, Novartis Pharma AG, OncoSec, Pierre Fabre, QBiotech Group Limited, Regeneron Pharmaceuticals Inc, SkylineDX B.V., Specialised Therapeutics Australia Pty Ltd. J.V.P. and N.W. are equity holders and Board members of genomiQa PTY LTD. P.J. is currently employed by Celsius Therapeutics. B.S.T. reports receiving Honoria and research funding from Genentech and Illumina and advisory board activities for Boehringer Ingelheim and Loxo Oncology, a wholly owned subsidiary of Eli Lilly. B.S.T. is currently employed by Loxo Oncology. L.S. is co-founder and consultant of BlueRock Therapeutics. D.B.S. has consulted with/received honoraria from Pfizer, Loxo Oncology, Lilly Oncology, Vivideon Therapeutics, Q.E.D. Therapeutics, and Illumina. J.D.W. is a consultant for Amgen; Apricity; Arsenal; Ascentage Pharma; Astellas; Boehringer Ingelheim; Bristol Myers Squibb; Eli Lilly; F Star; Georgiamune; Imvaq; Kyowa Hakko Kirin; Merck; Neon Therapeutics; Polynoma; Psioxus, Recepta; Trieza; Truvax; Sellas. J.D.W. has grant and research support from Bristol Meyers Squibb and Sephora. J.D.W. has equity in Tizona Pharmaceuticals; Imvaq; Beigene; Linneaus, Apricity, Arsenal IO; Georgiamune. T.M. is a consultant for Leap Therapeutics, Immunos Therapeutics and Pfizer, and co-founder of Imvaq therapeutics. T.M. has equity in Imvaq therapeutics. T.M. reports grants from Bristol Myers Squibb, Surface Oncology, Kyn Therapeutics, Infinity Pharmaceuticals, Peregrine Pharmaceuticals, Adaptive Biotechnologies, Leap Therapeutics and Aprea. T.M. is inventor on patent applications related to work on oncolytic viral therapy, alphavirus-based vaccines, neo-antigen modeling, CD40, GITR, OX40, PD-1 and CTLA-4. N.R. is on the SAB and receives research funding from Chugai, on the SAB and owns equity in Beigene, and Fortress. N.R. is also on the SAB of Daiichi-Sankyo, Astra-Zeneca-MedImmune, and F-Prime, and is a past SAB member of Millenium-Takeda, Kadmon, Kura, and Araxes. N.R. is a consultant to Novartis, Boehringer Ingelheim, Tarveda, and Foresight and consulted in the last three years with Eli Lilly, Merrimack, Kura Oncology, Araxes, and Kadman. N.R. owns equity in ZaiLab, Kura Oncology, Araxes, and Kadman. N.R. also collaborates with Plexxikon. R.M.W. is a paid consultant to N-of-One Therapeutics, a subsidiary of Qiagen. R.M.W. receives royalty payments for the use of the casper line from Carolina Biologicals.

Data and Resource Availability

Information and requests for resources and reagents should be directed to and will be fulfilled by Richard M. White (whiter@mskcc.org). All plasmids generated in this study will be either deposited with Addgene or available upon request. All fish lines will be made available through Zebrafish International Resource Center (ZIRC) or by request. All cell lines are available by request. Human RNAseq data that support the findings of this study have been deposited in the European Genome-phenome Archive (EGA) and are available under study accession EGAS00001001552 (<https://ega-archive.org/studies/EGAS00001001552>) with dataset accession EGAD00001006439 (<https://www.ebi.ac.uk/ega/datasets/EGAD00001006439>). Patient sample information and differential expression tables are made available in the Supporting Information (Supplemental Table 5). Zebrafish fin versus body bulk RNA-seq experiment is available via the NCBI GEO repository under identifier code GSE158538, with bulk RNA-seq counts and differential expression tables in the Supporting Information (Supplemental Table 4). Zebrafish fin versus body single cell RNA-seq experiment is available via the NCBI GEO repository under the identifier code GSE181748. Original source data for ChIP-seq analysis is from³¹ and can be found at GEO GSE81358. The pathway analysis is made available in the Supporting Information (Supplemental Table 6). HOXA13 Cut & Run data is available via the NCBI GEO repository under identifier code GSE181768. MACS2 peak scores and pathway analysis is made available in the Supporting Information (Supplemental Table 7). CRKL IP mass-spec data is available on via <http://www.proteomexchange.org> with identifier PXD027968. and is also available in the Supporting Information (Supplemental Table 8). Matlab script for tailfin image analysis and R scripts for RNA-seq, ChIP-seq pathway

analysis, Cut&Run, and IP-mass spec analysis is available on Github (https://github.com/jmweiss18/Weiss_Nature_2022)

Fig. 1: Acral versus cutaneous melanoma driver genes lead to anatomically distinct tumors in transgenic zebrafish.

(a-b) MSK-IMPACT targeted sequencing of biologically independent samples from n=100 acral and n=839 cutaneous melanoma patients. (a) The acral enrichment score was calculated by dividing the frequency a gene is altered in acral melanoma by the frequency in cutaneous melanoma. This ratio was then log2 transformed for visualization. Positive scores indicate genes enriched in acral melanoma and negative scores indicate genes enriched in cutaneous melanoma. (b) Frequency of gene amplifications compared by melanoma subtype using a two-sided Fisher's exact test. (c) RNA sequencing on a separate cohort of biologically independent samples from n=61 acral and n=53 cutaneous melanoma patients comparing expression of CRKL and GAB2. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extend to the largest/smallest value up to 1.5 x the interquartile range. FDR = 0.05 and adjusted p-values calculated using a two-sided Wald test with DESeq2. (d) Transgenic zebrafish models. The WT melanocyte model expresses no oncogenic drivers. The acral melanoma model has overexpression of human CRKL, GAB2, TERT and knockout of zebrafish NF1 orthologues, *nf1a* and *nf1b*. The cutaneous melanoma model expresses BRAF^{V600E} in a p53^{-/-} genetic background. All transgenes are driven by the melanocyte-specific promoter *mitfa*, and melanocytes are marked with *mitfa*:GFP. See Extended Data Fig. 2a for more details. (e) Ternary diagram portraying the percentage of tumors arising in the head, body, or fins. Anatomic distribution was compared using a chi-squared test. See Extended Data Figure 2h for histogram representation. See Supplemental Table 3 for fish and tumor numbers. (f) Tumor-free survival of WT melanocytes (n=94), acral (n=230), cutaneous (n=182) transgenic zebrafish models. P-values generated by log-rank Mantel-Cox test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Fig. 2: CRKL drives melanoma to acral sites.

(a) The 4 genes (CRKL, GAB2, TERT, NF1 KO) used to drive the acral melanoma model were introduced separately to assess sufficiency in tumorigenesis. (b) Representative images of zebrafish for each genotype. (c) Tumor-free survival of indicated genotypes. Number of fish for each genotype indicated in figure key. P-values generated by log-rank Mantel-Cox test. (d) Pie chart showing the percentage of CRKL-alone melanomas arising at each anatomic location. The number of tumors is listed under the pie chart. (e) Ternary diagram portraying the percentage of tumors forming in the head, body, or fins of indicated genotypes. A chi-squared test was performed to compare the anatomic distribution between the different transgenic models. (f) Each one of the 4 acral drive genes was removed to assess which driver genes were necessary for tumorigenesis. (g) Representative images of zebrafish for each genotype. Right panel shows pie charts with the anatomic distribution of tumors. Number of tumors are indicated. (h) Ternary diagram comparing the indicated genotypes. A chi-squared test was used to compare anatomic distribution between genotypes. (i) Tumor-free survival of indicated genotypes. Number of fish for each genotype indicated in figure key. P-values generated by log-rank Mantel-Cox test. See Supplemental Table 3 for fish and tumor numbers across all replicates and experimental conditions. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Fig. 3: Positional identity gene programs determine the response to CRKL.

(a) Schematic illustrating the fin versus body RNA-seq experiment. Body skin and fins were isolated from WT melanocyte and acral melanoma fish by dissection and then FACS sorted for GFP+ melanocytes and GFP- microenvironmental cells. (b) Unsupervised clustering using the 500 most variable genes across all samples, which clustered by cell lineage, anatomic position, and then genotype. (c) Waterfall plot representation of GSEA pathway analysis of WT fin vs body melanocytes showing the top 2500 enriched pathways and highlighting pathways related to limb development and positional identity. Limb development pathways were identified using the search terms “limb” or “appendage.” Positional-identity related pathways were identified using the search terms “morphogenesis,” “pattern”, and “regionalization”. Only pathways with FDR < 0.05 are highlighted. The specific pathways highlighted in the plot can be found in Extended Data Fig. 5e and Supplemental Table 4. (d) Heatmap representing expression of transgenes, melanocyte markers, and HOX genes across all samples.

Fig. 4: Human acral versus cutaneous melanoma has a positional identity gene program.

(a) RNA-seq performed on n=61 acral and n=53 cutaneous melanoma human patient samples. (b) Waterfall plot representation of GSEA pathway analysis of human acral vs cutaneous melanoma showing the top 2000 enriched pathways and highlighting pathways related to limb development and positional identity. GSEA analysis is controlled for disease stage. Only pathways with FDR < 0.05 are highlighted. The specific pathways highlighted in the plot can be found in Extended Data Fig. 7a and Supplemental Table 5. (c) Volcano plot showing differentially expressed genes between acral and cutaneous melanoma samples. Genes with FDR-adjusted p-value < 0.05 indicated in blue. P-values calculated using DESeq2. N=61 acral and n=53 cutaneous melanoma human patient samples. (d) Boxplots showing differences in gene expression between acral melanoma and cutaneous melanoma samples. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extend to the largest/smallest value up to 1.5 x the interquartile range. FDR=0.05 and adjusted p-values calculated using a two-sided Wald test with DESeq2. (e) Acral melanoma tissue microarray (TMA) underwent IHC for HOXB13. Representative images of 3 samples are shown. A total of n=30 samples were independently scored for extent and intensity, which are represented by pie charts. (f) HOMER known motif enrichment analysis on genes upregulated in acral vs cutaneous melanoma (log2 fold change cut-off \pm 1.5, two-sided Wald test with padj < 0.05, \pm 500bp of transcription start site (TSS). * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Fig. 5: CRKL synergizes with HOX13/IGF signaling to drive acral melanoma.

(a) HOX13 ChIP-seq performed on developing limb buds from E11.5 mouse embryos³¹. (b) Waterfall plot representation of 5625 top enriched GSEA pathways regulated by HOXD13 binding, highlighting limb and IGF-related pathways with FDR < 0.05. Specific pathways can be found in Extended Data Fig. 8b and Supplemental Table 6. (c) Western blot for HOXB13 and IGF2 in human acral melanoma cell lines after siRNA knockdown of all 4 HOX13 genes (HOXA13, HOXB13, HOXC13, HOXD13). For western blot source data, see Supplementary Fig. 1). (d) Western blot quantification of (c). Data was analyzed using two-sided student's t-test on n=4 biological replicates. Error bars = SEM. (e) ELISA for IGF2 on conditioned media after siRNA knockdown of all 4 HOX13 genes. Data was analyzed using a two-sided student's t-test on n=3 biological replicates. Error bars = SEM. (f) Schematic representation of tailfin melanocyte assay. (g-h) *mitfa*:CRKL stable line fish were injected with indicated plasmids to knockout specified gene targets. At 3-days post-fertilization tailfins were imaged for melanocyte area. Representative images are provided. (i-j) Quantification of (g-h). Data is pooled over n=3 biological replicates and the number of fish per condition is indicated. Data was analyzed using a two-sided student's t-test. Error bars = SEM. (k) Schematic of dnIRS2 transgene used to block insulin/IGF signaling in zebrafish melanocytes. Representative images of acral melanoma model with or without overexpression of dnIRS2-GFP. Pie charts demonstrate the anatomic distribution of each genotype. Number of tumors is indicated. (l) Ternary plot showing the anatomic distribution of tumors with indicated genotypes. P-values generated by chi-squared test. See Supplemental Table 3 for fish and tumor numbers. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 1: Identification of acral versus cutaneous melanoma genes.

Related to Fig. 1. (a-b) MSK-IMPACT of biologically independent samples from n=100 acral and n=839 cutaneous melanoma patients. Two-sided Fisher's exact test was used to compare the frequency of the most recurrently mutated and deleted genes by melanoma subtype. Both coding and promoter mutations were counted for TERT. (c) RNA-seq of biologically independent samples from n=61 acral and n=53 cutaneous melanoma patient. Boxplots compare indicated genes by subtype. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extends from the box maxima/minima to the largest/smallest value no further than $1.5 \times \text{IQR}$ (interquartile range) from the box maxima/minima. Data beyond the end of the whiskers are plotted individually as outliers. FDR=0.05 and adjusted p-values were calculated using a two-sided Wald test with DESeq2. (d) Schematic detailing predicted synergistic interaction between putative driver genes in acral melanoma. (e) TCGA pan-cancer analysis shows significant co-occurrence of CRKL, GAB2, NF1, and TERT alterations. Data analyzed with one-sided Fisher's exact test and p-values adjusted for FDR=0.05. (f) Clinical course of an acral melanoma patient. (g) WGS copy number profile showing copy number changes in GAB2, CRKL, and NF1. (h) Putative drivers of patient's acral melanoma. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 2: A transgenic zebrafish model of acral melanoma.

Related to Fig. 1. (a) Plasmids used to create the WT melanocyte model, acral melanoma model, and cutaneous melanoma models. (b) Genotyping PCR to confirm integration of plasmids into zebrafish genome. Representative images of n=3 independent replicates. For gel source data, see Supplementary Fig. 3. (c) qPCR to validate RNA expression of transgenes. N=3 biological independent replicates. Error bars = SEM. (d) Western blot was performed on WT and acral melanoma fish for human CRKL and GAB2 to validate transgene expression. N=3 biological replicates are shown. For western blot source data, see Supplementary Fig. 4. (e) RNA-seq on embryos at 5 days post-fertilization FACS sorted for GFP+ melanocytes. Log normalized expression of melanocyte markers and acral transgenes are indicated. N=4 biologically independent replicates. P-values calculated using a two-sided Wald test with DESeq2. Error bars = SEM. (f) CRISPR-seq was performed on the predicted sgRNA cut locus of zebrafish *nf1a* and *nf1b* to sensitively detect Cas9-mediated editing. Reference genome as well as the two most commonly altered reads are shown. The right panel is a heatmap with the frequency of reference and edited sequences in WT melanocyte and acral melanoma models. Variants displayed for both *nf1a* and *nf1b* are frameshift mutations in exon 1 leading to a predicted loss of function. (g) Immunofluorescence on transverse sections of WT melanocyte model, tumor-bearing acral melanoma model, and tumor-bearing cutaneous melanoma model for GFP and SOX10. GFP labels all tumor cells and SOX10 is used as a melanocyte lineage marker. Asterisks indicate blood vessels with autofluorescence. Representative of n=3 biological replicates. (h) Histogram representation of the ternary plot portrayed in Fig. 1e showing the percentage of tumors forming in the head, body, or fins of acral fish and cutaneous fish melanoma models. Data represents n=3 biological replicates, which range from n=43 to n=141 fish per replicate. See Supplemental Table 3 showing the exact number of fish and corresponding percentages for each replicate. To compare overall anatomic distribution between the two genotypes, a chi-squared test was performed. To compare the frequency of tumors at each anatomic location, a student's two-sided t-test was performed. Error bars = SEM. Ternary plot represents the same data presented in the histogram. (i) WT melanocyte vs acral melanoma model compared for melanocyte area in tailfin at 3-days post-fertilization. Data represents n=29 WT fish and n=41 acral fish pooled from n=4 biological replicates. Each point represents the tailfin melanocyte area of a different animal. P-values generated by a two-sided student's t-test. Error bars = SEM. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 3: Histological profiling of a zebrafish model of acral melanoma.

Related to Fig. 1. n=3 WT, acral, and cutaneous melanoma transgenic fish were used for histological profiling. Images of the fish used for histology are shown on top with a red line indicating anatomic region used for sectioning and profiling. For WT fish, head, body, and fin regions were used as a negative control. For acral fish, profiled tumors were located on the fins. For cutaneous fish, profiled tumors were located on the body and head. The markers used include H&E, GFP (marking melanocytes in all 3 models), CRKL (expressed only in the acral model), pIGF1R, pERK, and pS6. See methods for details regarding antibody, concentration, and staining procedure.

Extended Data Fig. 4: Stable (germline) transgenic acral lines show enhanced anatomic specificity.

Related to Fig. 1. (a) F0 transgenic fish shown in Fig 1d and Fig 2b were outcrossed to generate stable germline transgenics. Representative images of each transgenic line are shown with arrows to indicate the location of tumors. (b) Tumor frequency of each stable line at 1 year post fertilization. (c) Ternary plot showing the anatomic distribution of tumors for each stable transgenic line. P-values generated by chi-squared test. See Supplemental Table 3 for a full list of fish and tumor numbers across all replicates and experimental conditions. (d) Adult WT melanocyte and acral melanoma models were dissected to isolate body skin and fins and then analyzed via flow cytometry. The images show representative differences in pigment patterning between the two models. The histogram shows the percentage of total cells that were GFP+ grouped by anatomic location and genotype. Data represents n=3 biological replicates. Each replicate represents the pooling of n=2 male and n=2 female fish. P-values generated by student's two-sided t-test. Error bars = SEM. (e) The ratio of fin melanocytes to body melanocytes was compared between the two models using the data from (d). Data represents n=3 biological replicates. P-values generated by student's two-sided t-test. Error bars = SEM. (f) Gating strategy of the flow cytometry data. (g) Representative contour plots of melanocyte frequency by genotype and location. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 5: Fin melanocytes have a limb positional identity.

Related to Fig. 3. (a-c) Validation of successful isolation of melanocytes for RNA-seq. (a) Volcano plots comparing melanocytes (GFP+ sample) to their microenvironment (GFP- sample) across locations and transgenic models. Melanocyte markers are labeled. Genes with FDR-adjusted p-value < 0.05 indicated in blue. P-values calculated using DESeq2. (b) GSEA showing the list of the top pathways enriched in melanocytes (GFP+ sample) compared to their microenvironment (GFP- sample). Colors indicate p-value adjusted for FDR = 0.05. (c) Log normalized counts for the expression of transgenes across all samples. EGFP and *mitfa* expression are high in all melanocyte samples and CRKL, GAB2, TERT, and Cas9-mCherry expression is high only in acral melanoma model melanocytes. N=3 biologically independent replicates. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extends to the largest/smallest value no further than $1.5 * \text{IQR}$ (interquartile range) from the box maxima/minima. Data beyond the end of the whiskers are plotted individually as outliers. (d) Principal component analysis (PCA) for all samples showing principal components 1 (PC1) and 2 (PC2). The percent of transcriptional variation captured by each principal component is indicated. (e) GSEA pathway analysis comparing fin vs body melanocytes from the WT melanocyte model listing the top enriched pathways in fin melanocytes. Limb development and positional identity-related pathways are highlighted. See Supplemental Table 4 for a full list of pathways. (f) GSEA barcode plot showing enrichment of genes in the GO: Appendage Development pathway, generated with weighted kolmogorov smirnov (WKS) testing. NES and FDR=0.05 adjusted p-values are indicated. (g-h) Volcano plots comparing fin melanocytes vs body melanocytes from the (g) WT melanocyte model and (h) acral melanoma model. Genes with FDR-adjusted p-value < 0.05 indicated in blue. P-values calculated using DESeq2.

Extended Data Fig. 6: Zebrafish fin melanocytes express higher levels of HOX13 genes.

Related to Fig. 3. (a) Schematic illustrating the fin versus body scRNA-sequencing experiment. Similar to the bulk RNA-seq experiment in Fig. 3a, body skin and fins were isolated from acral melanoma fish by dissection and then FACS sorted for GFP+ melanocytes and GFP-microenvironmental cells. This generated n=4 samples (body melanocytes, body TME, fin melanocytes, fin TME) that each underwent scRNA-seq. (b) UMAP pooled for all 4 samples highlighting the various cell types captured by scRNA-seq. (c) Violin plot comparing *hoxc13b* expression in body and fin melanocytes. Data represents log normalized counts and was compared by two-sided Wilcoxon rank sum test. (d) Violin plot comparing total expression of all HOX13 genes (*hoxa13a*, *hoxa13b*, *hoxb13a*, *hoxc13a*, *hoxc13b*, *hoxd13a*). To avoid single-cell drop out of more lowly expressed HOX13 genes, log normalized counts were summed together to calculate total expression per cell. Data was analyzed by two-sided Wilcoxon rank sum test. (e) Violin plot comparing the total expression of all HOX13 genes across all fin and body cell types. Data was analyzed by two-sided Wilcoxon rank sum test. (f) Zebrafish hox genes detected by bulk RNA-seq were used to perform unsupervised clustering and visualized with a heatmap. Samples clustered by anatomic location, then lineage, and then genotype in that order, indicating the hox genes predominantly associate with anatomic position. Particular hox genes that are differentially expressed across all body and fin samples are indicated with a black box. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 7: Human melanoma has a positional identity defined by HOX genes.

Related to Fig. 4. (a) GSEA pathway analysis for top enriched pathways in human acral melanoma and is controlled for disease stage. Limb development pathways and positional identity-related pathways are highlighted. See Supplemental Table 5 for a full list of pathways. (b) GSEA barcode plot showing enrichment for the GO: Appendage Development pathway in acral melanoma, generated with WKS testing. NES and FDR=0.05 adjusted p-values are indicated. (c) PCA of all samples based on the expression of just HOX genes. Color indicates the combination of melanoma subtype and disease stage. (d) N=3 acral melanoma tissue microarrays of n=32 samples total from Sloan Kettering was used for staining for H&E, CRKL, HOXB13, and pIGF1R. (e) Extent and intensity of each sample was independently scored by a Sloan Kettering dermatopathologist and is represented in a pie chart. 2 samples were removed from the analysis due to heavy pigmentation, resulting in n=30 samples total used for IHC quantification analysis. Extent scores range 0-4 and intensity scores range 0-3. (f) N=3 cutaneous melanoma tissue microarrays of n=14 total samples from Sloan Kettering was used for staining for H&E, CRKL, HOXB13, and pIGF1R. (g) Extent and intensity of each sample was independently scored by a Sloan Kettering dermatopathologist and is represented in a pie chart. Extent scores range 0-4 and intensity scores range 0-3. (h) CRKL, HOXB13, and pIGF1R antibodies were optimized on human tissues chosen for their expression characterized by the Human Protein Atlas. CRKL is a ubiquitously expressed protein with relatively high expression in the colonic epithelia and lower expression in liver. HOXB13 is expressed in prostate epithelia with low expression in liver and skin (anatomic source unknown). pIGF1R demonstrates higher levels in prostate epithelia, colon epithelia, tonsil, and lower levels in skin. See methods for details regarding antibody, concentration, and staining procedure.

Extended Data Fig. 8: HOX13 regulates insulin/IGF signaling.

Related to Fig. 5. (a-e) Analysis of HOXA13, HOXD13, H3K27ac ChIP-seq data analyzed from Sheth et al., 2016³¹ performed on developing limb buds from E11.5 mouse embryos. (a) Waterfall plot representation of 5594 top enriched GSEA pathways regulated by HOXA13, highlighting limb and insulin/IGF-related pathways. Limb development pathways were identified using the search terms “limb” or “appendage.” Insulin/IGF signaling pathways were identified using the search terms “insulin” or “IGF”. Only pathways with FDR < 0.05 are highlighted. (b-c) Histogram showing the significantly enriched insulin/IGF pathways regulated by HOXA13 and HOXD13. NES and p-value are indicated. (d-e) Integrated genome browser tracks for HOXA13 and HOXD13 ChIP-seq binding and H3K27 acetylation near the transcription start site of IGF1 and IGF2. A full list of pathways can be found in Supplemental Table 6. (f) GSEA pathway analysis for the top pathways enriched in fin melanocytes versus body melanocytes from the zebrafish acral melanoma model. IGF-related pathways are indicated in red. A full list of pathways can be found in Supplemental Table 4. (g) GSEA barcode plot comparing acral model fin vs body melanocytes showing enrichment of genes in the GO: Regulation of Multicellular Organisms Growth pathway, generated weighted WKS testing. NES and FDR=0.05 adjusted p-values are indicated. (h) Boxplot showing log normalized counts of zebrafish *igf1*, *igf2a*, and *igf2b* expression in all melanocyte samples from bulk RNA-seq. N=3 independent biological replicates. P-values calculated using DESeq2 and adjusted for FDR = 0.05. Box minima=25th percentile, centre=50th percentile, maxima=75th percentile. Whiskers extends from the box maxima/minima to the largest/smallest value no further than 1.5 * IQR (interquartile range). Data beyond the end of the whiskers are plotted individually as outliers. (i-o) Cut & Run was performed on HOX13 expressing human melanoma cell lines, SKMEL-1088, SKMEL-1176, SKMEL-1206, using an antibody against HOXA13 and IgG as a negative control. (i) HOMER known motif analysis comparing HOXA13 vs IgG peaks identified significant enrichment for HOXB13 in all 3 melanoma cell lines. (j) GSEA pathway analysis using Cistrome-GO was performed, and the top enriched pathways are listed highlighting pathways related to insulin/IGF signaling. Insulin/IGF signaling pathways were identified using the search terms “insulin” and “IGF.” Insulin/IGF-related pathways were identified based on literature known to operate directly downstream of insulin/IGF signaling. A full list of pathways can be found in Supplemental Table 7. (k-o) IGV plot showing peaks along IGF1R, IGF2, IRS1, IRS2, and IGFBP3 for all 3 cell lines. The transcription start site (TSS) is indicated. (p) Patient derived melanoma cell lines were analyzed for HOXB13, pIGF1R, and total IGF1R expression by western blot. Actin was used as a loading control. Cell lines derived from acral and cutaneous tumors are labeled. SKMEL-1088 is a melanoma cell line of unknown anatomic origin. Images are representative of n=3 biological replicates. (q) SKMEL-1176 and SKMEL-1206 acral melanoma cell lines were transfected with siRNAs targeting all 4 human HOX13 genes (HOXA13, HOXB13, HOXC13, HOXD13) or a non-targeting control. Cells were grown in standard media conditions. 96 hours post-transfection, cells were collected and analyzed by western blot. HOXB13 was used to validate siRNA knockdown and then also probed for pIGF1R and IGF1R. Actin was used as a loading control. Blots are representative of n=4 biological replicates. (r) Western blot quantification of the n=4 biological replicates from (q). Data was analyzed using a two-sided student's t-test. Error bars = SEM. For western blot source data, see Supplementary Fig. 5. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 9: CRKL amplifies IGF signaling in human melanoma cells

Related to Fig. 5.

(a) Phospho-RTK array performed on WM3918 human melanoma cell line with or without overexpression of CRKL. Array tests phosphorylation status of 49 RTKs. For pRTK array source data, see Supplementary Fig. 2. (b) MeWo, SKMEL-1176, and 293T cells were transduced and selected to overexpress CRKL or an empty vector (EV) as a negative control. Western blot for CRKL, pIGF1R, and total IGF1R was then performed to compare basal levels of IGF signaling. Actin was used as a loading control. For western blot source data, see Supplementary Fig. 6. Quantification of western blots shown. Data represents n=3 biological replicates and was analyzed using a two-sided student's t-test. (c) Schematic representation of CRKL with a V5 tag. CRKL is composed of a SH2 (Src homology) domain that bind proteins with phosphotyrosines and two SH3 domains that bind to proline-rich proteins. (d) Human melanoma cell line WM3918 overexpressing CRKL with a V5 tag was compared to CRKL overexpressing cells without the V5 tag. N=5 control replicates and n=6 CRKL-V5 biological replicates were used for each condition. Lysates were immunoprecipitated using a V5 antibody and then underwent mass spectrometry, yielding 57 significant interactors as defined by logFC > 2 and corrected p-value < 0.05. CRKL itself was among these significant interactors. Significant interactors with CRKL were organized based on whether they contain a canonical CRKL SH2-binding domain (pY-x-x-P), canonical CRKL SH3 binding motif (Ψ -P- Ψ -L/V/P/A/I-P- Ψ -K), known SH3 binding motif (proline-rich sequence), or no identified binding motif. Genes are color coded by their contribution to PI3K and MAPK signaling. (e) Heatmap showing increased detection of significant interactors in the CRKL-V5 vs control group, expressed as normalized counts. Raw data can be found in Supplemental Table 8. (f) Volcano plot showing log2 fold enrichment of CRKL-V5 IP vs. control IP (CRKL with no V5 tag). * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 10: CRKL-mediated melanocyte tailfin expansion depends on a HOX13/IGF positional program.

Related to Fig. 5. (a) Plasmids used to knockout the human CRKL transgene, and zebrafish IGF ligands (*igf1*, *igf2a*, *igf2b*) zebrafish HOX13 genes (*hoxa13a*, *hoxa13b*, *hoxb13a*, *hoxc13a*, *hoxc13b*, *hoxd13a*) and orthologue of HOXB7, *hoxb7a*, in a mosaic manner in CRKL stable line zebrafish. *mitfa*:Cas9 ensures melanocyte specificity in CRISPR editing. This plasmid also contains a fluorescent heart marker *myl17*:GFP, which was used to determine which fish had successful plasmid integration to be used in further downstream image analysis. (b) Surveyor validation demonstrating targeted editing of the human CRKL transgene using 3 different sgRNAs. For gel source data, see Supplementary Fig. 7a. (c) Surveyor validation demonstrating targeted editing zebrafish HOX genes using 6 different sgRNAs. Note, HOX13-targeting plasmids only target HOX13 genes (not *hoxb7a*) and *hoxb7a*-targeting plasmid only targets *hoxb7a*. For gel source data, see Supplementary Fig. 8. (d) Surveyor validation demonstrating targeted editing of all zebrafish IGF ligands using 3 different sgRNAs. For gel source data, see Supplementary Fig. 7b. (e) Tumor-free survival comparing acral to acral dnIRS2 genotypes. Number of fish for each genotype indicated in figure key. P-values generated by log-rank Mantel-Cox test. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 11: HOX13 synergizes with CRKL to drive acral melanoma through IGF/PI3K signaling.

Related to Fig. 5. (a-d) Acral melanoma model imaged for melanocyte tailfin area at 3-days post-fertilization after indicated pharmacologic treatment. Representative images and quantification provided. P-value generated with two-sided student's t-test. Error bars = SEM. (a) Insulin/IGF1 receptor antagonists BMS-754807 at 7.5 μ M and NVP-AEW541 at 60 μ M compared to 0.1% DMSO control. Data represents n=35 DMSO-treated fish, n=31 BMS-treated fish, and n=26 NVP-treated fish pooled over n=3 biological replicates. (b) PI3K inhibitor LY294002 at 15 μ M and RAF/MEK inhibitor CH5126766 at 1 μ M compared to 0.1% DMSO control. Data represents n=45 DMSO-treated fish, n=43 PI3K inhibitor-treated fish, and n=21 RAF/MEK inhibitor-treated fish pooled over n=3 biological replicates. (c) MEK inhibitors pimerasertib at 1 μ M, refametinib at 1 μ M, and trametinib at 200nM compared to 0.1% DMSO control. Data represents n=46 DMSO-treated fish, n=48 pimasertib-treated fish, and n=61 refametinib-treated fish, and n=44 trametinib-treated fish pooled over n=4 biological replicates. (d) SOS1 inhibitor, BI-3406, at 1 μ M, MEK inhibitor refametinib at 1 μ M, or combined BI-3406 (1 μ M) + refametinib (1 μ M) treatment, compared to 0.1% DMSO control. Data represents n=43 DMSO-treated fish, n=31 SOS1 inhibitor-treated fish, n=44 MEK inhibitor-treated fish, and n=35 combined SOS1/MEK inhibitor-treated fish pooled over n=3 biological replicates. * p-value < 0.05, ** p-value < 0.01, *** p-value < 0.001, **** p-value < 0.0001.

Extended Data Fig. 12: Anatomic position determines oncogenic specificity in melanoma.

Related to Figure 5. Melanocytes at different anatomic locations have different positional identities determined by HOX genes. Fin/limb melanocytes have higher expression of posterior HOX13 genes. HOX13 drives higher expression of IGF ligands and IGF1R, resulting in greater IGF signaling and increases the vulnerability of fin melanocytes to CRKL-mediated transformation. CRKL synergizes with a HOX13 program in fin melanocytes by forming a complex with PI3K, the primary downstream mediator of IGF1R, thereby amplifying IGF signaling. GAB2, another commonly amplified gene in acral melanoma, does not directly bind to CRKL, but participates in the same complex as CRKL through their shared binding and amplification of PI3K signaling. This results in tumor phenotypes with sensitivity to IGF1R or PI3K inhibition, but only modest effects from MAPK inhibition.

Methods

Zebrafish

Zebrafish Husbandry

Fish stocks were kept under standard conditions at 28.5°C under 14:10 light:dark cycles, pH (7.4), and salinity-controlled conditions. Animals were fed standard zebrafish diet consisting of brine shrimp followed by Zeigler pellets. The animal protocols described in this manuscript are approved from the Memorial Sloan Kettering Cancer Center (MSKCC) Institutional Animal Care and Use Committee (IACUC), protocol number 12-05-008. All anesthesia was performed using Tricaine-S (MS-222, Syndel USA, 712 Ferndale, WA) with a 4g/L, pH 7.0 stock. Both male and female zebrafish were utilized in equal proportions for all experiments utilizing adult fish. Sex determination in embryos is not possible at 3 days post fertilization (dpf). Embryos were collected from natural mating and incubated in E3 buffer (5mM NaCl, 0.17 mM KCl, 0.33 mM CaCl₂, 0.33 mM MgSO₄) at 28.5°C.

Generating Zebrafish Transgenic Lines

One-cell-stage embryos of indicated genotype were injected with indicated plasmid (see below) along with 20pg of tol2 mRNA. The total amount of plasmid per injection did not exceed 40pg. Embryos were grown to adulthood at 2 months post-fertilization (2mpf) and then screened for melanocyte rescue. To create stable lines, these fish were outcrossed with casper (*mitfa*^{-/-}, *mpv17*^{-/-})⁴⁵ or BRAF/p53 caspers fish for two generations (F1 and F2) before in-crossing for F3 stable line generation. The following stable lines were generated in this study: “WT” stable line (MiniCoopR-GFP), acral melanoma stable line (*mitfa*:CRKL, *mitfa*:GAB2, *mitfa*:TERT, *mitfa*:Cas9-mCh; zU6:gRNA-nf1a, *mitfa*:Cas9-mCh; zU6:gRNA-nf1b, MiniCoopR-GFP), CRKL stable line (*mitfa*:CRKL, MiniCoopR-GFP), cutaneous melanoma stable line (*mitfa*:BRAF^{V600E}; p53^{-/-}; MiniCoopR-GFP, *crestin*:tdTomato). Monitoring for tumor-free survival began at 2mpf and ended at 1 year. For imaging of tailfin melanocytes, embryos at 3dpf were utilized. Fish utilized for FACS and RNA-seq were 6mpf.

Zebrafish Transgenic Lines

Zebrafish strains used in these studies included casper (*mitfa*^{-/-}, *mpv17*^{-/-})⁴⁵ and casper p53^{-/-} with *mitfa*:hBRAF^{V600E} (“BRAF/p53 casper”). The following transgenic lines were all generated using one-cell injection of 20pg of tol2 mRNA and indicated plasmids. The WT melanocyte model was generated by injecting casper fish with 5pg of MiniCoopR-eGFP and 25pg of an empty vector/394b plasmid, which is a plasmid backbone with no transgene. This was used as filler to ensure all conditions were injected with the same total amount of DNA. The cutaneous melanoma model was generated by injecting BRAF/p53 casper fish with 5pg of MiniCoopR-eGFP and 25pg of empty vector/394 plasmid. The acral melanoma model was generated by injecting casper fish with 5pg of each of the following plasmids: MiniCoopR-eGFP, *mitfa*:hsCRKL, *mitfa*:hsGAB2, *mitfa*:hsTERT, *mitfa*:Cas9-mCherry; zU6:*nf1a*-gRNA, of *mitfa*:Cas9-mCherry; zU6:*nf1b*-gRNA. For transgenic models evaluating the role of a single transgene (CRKL, GAB2, and TERT), 5pg of MiniCoopR-eGFP and 25pg of the indicated transgene was used. The NF1 KO melanoma model was generated by injecting 5pg of MiniCoopR-eGFP and 12.5pg of *mitfa*:Cas9-mCherry; zU6-*NF1a*-gRNA and 12.5pg of *mitfa*:Cas9-mCherry; zU6-*NF1b*-gRNA. For transgenic models evaluating the removal of a given transgene from acral melanoma model, 5ng was used for all indicated plasmids. For transgenic models evaluating the effect of dnIRS2 on the acral melanoma model, 5pg was used for all indicated plasmids. *crestin*:tdTomato reporter was also utilized to aid in identifying tumors to monitor tumor-free survival^{46,47}. For experiments to knockout hCRKL, IGF, and HOX genes in the *mitfa*:CRKL stable line, 10pg of *mitfa*:Cas9/395 along with 10pg along with the following indicated sgRNA plasmids: zU6:gRNA-hCRKL, zU6:gRNA-igf, zU6:hox13-1, zU6:hox13-2, zU6:hox7ba, or zU6:gRNA-NT. Genotypes were regularly monitored by PCR.

Humans

MSK-IMPACT DNA Sequencing

MSK-IMPACT testing for patients with advanced cancer was ordered by the treating physician to identify clinically relevant genomic alterations that could potentially inform treatment decisions. Patients undergoing MSK-IMPACT testing signed a clinical consent form or, in >85% of cases, enrolled on an institutional IRB-approved research protocol (MSKCC; NCT01775072) permitting return of results from clinical sequencing and broader genomic characterization of banked specimens for research. All MSK-IMPACT testing that was not submitted for reimbursement by insurance was paid for using institutional and philanthropic funds. Following consent, either archival or new tumor samples were obtained and blood was drawn as a source of matched normal (germline) DNA. To ensure uniform nomenclature of tumor types, tumors were annotated according to an institutional classification system, OncoTree (<http://www.cbioportal.org/oncotree/>). To minimize selection bias, patients of all ages, ethnicities, and stages were used for data recruitment. As noted in population characteristics above, this led to a similar distribution of sex, age, and disease stage between acral and cutaneous melanoma samples.

WES and WGS of an Individual Acral Melanoma Patient

The female patient was enrolled at Memorial Sloan-Kettering Cancer Center (MSKCC) and consented on a protocol approved by MSKCC and analyzed as part of a previous acral melanoma study¹⁵. Samples were obtained in accordance with standard biopsy or surgical procedures. The protocol for WES and WGS of acral melanoma patient was approved by MSKCC IRB 12-245.

RNA-sequencing of Human Melanoma Patients

For human RNA-seq sequencing analysis, fresh-frozen tissue samples were obtained from the biospecimen bank of Melanoma Institute Australia (MIA) and all samples were accrued prospectively with written informed patient consent. The protocol for the study was approved by the Sydney Local Health District Ethics Committee (Protocol No X15-0454 (prev X11-0289) & HREC/11/RPAH/444 and Protocol No X17-0312 (prev X11-0023) & HREC/11/RPAH/32) and cases were also approved by institutional ethics committees of Melanoma Institute of Australia and QIMR Berghofer Medical Research Institute (HREC approval P452 & P2274). Details of patient samples can be found in Supplemental Table 5. The acral RNA-seq samples have been previously published in¹³.

Acral and cutaneous melanoma Tissue Microarrays

The human acral melanoma tissue microarrays were made from patients consented to IRB protocol 06-107 at Memorial Sloan Kettering Cancer Center. Punch biopsies were taken from viable tumor areas and embedded in paraffin.

Cell Lines and Virus Preparation

WM3918 melanoma cell line was grown with Dulbecco's Modified Eagle Medium (Gibco #11965) supplemented with 10% FBS (Seradigm), 1X penicillin/streptomycin/glutamine (1X PSG) (Gibco #10378016). SKMEL-1152, SKMEL-1136, SKMEL-1094, SKMEL-1128, SKMEL-1088, SKMEL-1176, and SKMEL-1206 are patient-derived human melanoma cell lines established at Memorial Sloan Kettering Cancer Center. SKMEL-1152, SKMEL-1136, SKMEL-1094, and SKMEL-1128 are derived from cutaneous melanoma patients and SKMEL-1176 and SKMEL-1206 are derived from acral melanoma patients. SKMEL-1088 is derived from a human melanoma of unknown subtype. MSK-IMPACT sequencing was performed to identify their putative genetic drivers. WM3918 was authenticated using Short Tandem repeat profiling

performed at the Wistar institute and MSK-IMPACT at MSKCC. MeWo and 293T were authenticated using Short Tandem repeat profiling at ATCC. SKMEL-1176, SKMEL-1206, SKMEL-1088, SKMEL-1152, SKMEL-1136, SKMEL-1094, SKMEL-1128 were authenticated at MSKCC using Short Tandem repeat profiling as well as MSK-IMPACT. All cell lines were confirmed mycoplasma negative using a commercially available mycoplasma kit. No commonly misidentified cell lines were used in this study. Patient derived cell lines were grown in RPMI 1640 Medium (Invitrogen #11875093) supplemented with 20% FBS (Seradigm), 1X

penicillin/streptomycin/glutamine (Gibco #10378016). All cell lines were kept in a sterile 37°C,

5% CO₂ incubator.

To make CRKL-overexpressing cell line, CRKL retrovirus was produced by transfecting 293T packaging cells with CRKL-pWzl, VSVg and pCL-Ampho plasmids. WM3918 human melanoma cell line was transduced with CRKL retroviral particles plus polybrene and then selected with 10 µg/mL blasticidin. To make CRKL-V5-overexpressing cell lines and empty vector (EV) control cell lines, CRKL-V5 and EV lentivirus was produced by transfecting 293T packaging cells with pLenti6.3-CRKL-V5 or pLenti6.3-EV and Ready-to-Use Lentiviral Packaging Plasmid Mix (Cellecra #CPCP-K2A). WM3918, SKMEL-1176, MeWo, and 293T cell lines were transduced with CRKL lentivirus particles plus polybrene and then selected with 10µg/mL blasticidin. CRKL-pWzl, and pLenti6.3-CRKL-V5 was a kind gift from professor William Hahn.

Method Details

Acral Enrichment Score Analysis of MSK-IMPACT DNA Sequencing

Only melanoma patients clearly indicated as having either acral or cutaneous melanoma were analyzed. We selected the 14 most frequently observed copy number alterations and 7 most frequently mutated genes in acral melanoma and compare to the frequency observed in cutaneous melanoma. We also included well described melanoma drivers, such as deletion of NF1⁴⁸ and mutations in PTEN⁴⁹. To calculate the acral enrichment score we used the following equation $\log_2(\text{acral frequency}/\text{cutaneous frequency})$. Statistical differences in the frequency of copy number alterations and mutations were determined using a Fisher's exact test.

WES/WGS Analysis of an Acral Melanoma Patient

Tissue was selected by the pathologist to limit the amount of necrotic tissue and adjacent normal tissue was collected for DNA extraction of germline DNA. An H&E slide was made to confirm normal or malignant tissue and <50% necrosis, which was reviewed by the pathologist. The patient had both a primary tumor sample, which was formalin-fixed and paraffin embedded (FFPE) and an in-transit metastasis, which was fresh frozen by placing in a vial and submerging into a liquid nitrogen container. Paired tumor/normal whole-exome and whole-genome sequencing (WES and WGS, respectively) libraries were constructed and sequenced on the Illumina HiSeq using V3 reagents. WES and WGS data was aligned to the genome with BWA-MEM^{50,51} and mutations called using MuTect⁵². Coverage-based copy-number analysis was performed using custom scripts for WGS and using the allele-specific method FACETS for the WES data⁵³.

RNA-seq Analysis of Human Melanoma Patients

Fresh-frozen tumor RNA was extracted using the AllPrep® DNA/RNA/miRNA Universal kit (Qiagen #80224) and were quantified using the Qubit® RNA HS Assay (Q32852, Life Technologies). The TruSeq RNA library prep kit (Illumina, San Diego, California, USA) was

used to prepare libraries from RNA and these were sequenced with 100bp paired-end reads using Illumina HiSeq2000 or HiSeq2500 platforms. RNA-seq reads were trimmed for adapter sequences using Cutadapt (version 1.9) and aligned with STAR (version 2.5.2a)⁵⁴ to the GRCh37 assembly using the gene, transcript, and exon features of Ensembl (release 70) gene model. Gene expression counts were estimated using RSEM (version 1.2.30)⁵⁵. Differential expression was calculated with DESeq2⁵⁶ using the output of the quantMode and GeneCounts feature of STAR. The vst function was used to generate log2 transformed normalized counts. Differential expression for Figure 5D-F, including HOMER analysis used comparison of all acral vs cutaneous melanoma samples. Differential expression used for pathway analysis for Figure 5B-C used comparison of all acral vs cutaneous melanoma samples normalized by specimen type (i.e. primary tumor or lymph node metastasis). Pathway and Gene Ontology (GO) analysis were performed with GSEA using FGSEA-multilevel⁵⁷. Known motif analysis was performed with the HOMER⁵⁸ function findMotifs.pl, using the human genome (GRCh37) and searching for motifs of lengths 8, 10, and 16 within \pm 500bp of the TSS of differentially expressed genes. Motifs were annotated using JASPAR⁵⁹.

Co-Occurrence of CRKL, GAB2, NF1, and TERT in Cancer

The frequency of alterations occurring across multiple human cancer types was analyzed in cBioPortal¹⁷ looking at the following TCGA studies: Firehose Legacy⁶⁰, PanCancer Atlas⁶¹, Cell 2017⁶²⁻⁶⁵, Nature 2014⁶⁶⁻⁶⁸, Nature 2012⁶⁹⁻⁷¹, Cell 2015^{48,72,73}, Nature 2008⁷⁴, Nature 2015⁷⁵, Nature 2013^{76,77}, Cancer Cell 2014⁷⁸, NEJM 2013⁷⁹, Nature 2011⁸⁰, Cell 2013⁸¹, Cell 2014⁸². The odds ratio for co-occurrence of alterations was calculated and used to generate a p-value adjusted for FDR = 0.05.

HOXA13/HOXD13 ChIP-seq Pathway Analysis

ChIP-seq analysis was performed from raw reads of publicly available GEO datasets from Sheth et al, 2016 (GSE81358). Paired end reads from the following samples were processed: GSM2151011 (H3K27ac), GSM2151013 (HOXA13), GSM2151014 (HOXD13), GSM2151016 (H3K27ac input DNA), GSM2151017 (HOXA13 input DNA), GSM2151018 (HOXD13 input DNA). SRA files were converted to FASTQ format using the SRA toolkit. Raw reads were checked for sequence quality, adapter content, overrepresented sequences and Kmer content using FASTQC (Babraham Bioinformatics). Adapters and low quality sequences were filtered out using Trimmomatic⁸³ and filtered reads were mapped to the mm9 genome using Bowtie2⁸⁴. Mapped reads were analyzed using Samtools⁸⁵ and the sorted bam files were processed using Deeptools⁸⁶ to generate input normalized bigwig files. MACS v1.4⁸⁷ was used to perform peak calling using input DNA for each antibody as control. Pathway analysis of enriched peaks was performed using Cistrome-GO³².

HOXA13 Cut&Run of Human Melanoma Cell Lines

Sample Preparation

We performed Cut&Run for the melanoma cell lines SKMEL-1088, SKMEL-1766, and SKME-1206. For Cut&Run we used 100'000 cells per condition and it was performed as described in¹¹³. We used antibodies against HOXA13 (Invitrogen, #PA5-76440, 1:100), H3K27ac (Active Motif, #39034, 1:100), IgG (abcam, #ab6709, 1:100) and we added a "no antibody" condition as additional negative control. In brief, 100,000 melanoma cells were collected per condition. Cells were harvested and bound to concanavalin A-coated magnetic beads after 8min incubation at RT on a rotator. Cell membranes were permeabilized with digitonin and the different antibodies were incubated overnight at 4°C on a rotator. Beads were washed and incubated with pA-MN. Ca²⁺-induced digestion occurred on ice for 30min and stopped by chelation. DNA was isolated using an extraction method with phenol and chloroform as described in Skene et al.⁸⁸

Sequencing

Immunoprecipitated DNA was quantified by PicoGreen and the size was evaluated by Agilent BioAnalyzer. When possible, fragments between 100 and 600 bp were size selected using aMPure XP beads (Beckman Coulter catalog # A63882) and Illumina libraries were prepared using the KAPA HTP Library Preparation Kit (Kapa Biosystems KK8234) according to the manufacturer's instructions with up to 1.5ng input DNA and 12-16 cycles of PCR. Barcoded libraries were run on the NovaSeq 6000 in a PE100 run, using the NovaSeq 6000 SP Reagent Kit (200 cycles) (Illumina). An average of 13 million paired reads were generated per sample.

Analysis

Reads were trimmed and filtered for quality ($q=15$) and adapter content using version 0.4.5 of TrimGalore (https://www.bioinformatics.babraham.ac.uk/projects/trim_galore). Reads were aligned to human assembly hg38 with version 2.3.4.1 of bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/index.shtml>) and duplicates were collapsed to one read using MarkDuplicates in version 2.16.0 of Picard Tools. CUT&RUN target enrichment was assessed using MACS2 (<https://github.com/taoliu/MACS>) with FDR=0.1 and fold change of 2 over the matched IgG control background. Depth-normalized read density profiles were created using the BEDTools suite (<http://bedtools.readthedocs.io>). A global peak atlas was created by first removing blacklisted regions (<http://mitra.stanford.edu/kundaje/akundaje/release/blacklists/hg38-human/hg38.blacklist.bed.gz>) then merging all overlapping peaks. Peak-gene associations were created by assigning peaks using linear genomic distance to transcription start sites. Motif signatures were obtained using Homer v4.5 (<http://homer.ucsd.edu>). Pathway analysis of enriched peaks (from MACS2) was performed using Cistrome-GO with the following parameters: peak number to use: all, half-decay distance: automatic, FDR cutoff: 0.2.

CRKL IP-mass spec

Two cell lines were used for IP-MS/MS: Human melanoma cell line WM3918 with V5-tagged CRKL overexpression and a control cell line overexpressing CRKL without a V5-tag. Both conditions were immunoprecipitated using a V5 antibody (Invitrogen #R960-25). Because only the CRKL-V5 and not the wildtype CRKL cell line expressed the V5-tag, this approach enabled detection of proteins that interact with CRKL and remove background proteins from the analysis.

Sample preparation

Two cell lines were used for IP-MS/MS: Human melanoma cell line WM3918 with V5-tagged CRKL overexpression and a control cell line with wildtype (no V5-tag) CRKL overexpression. Cells were grown in Dulbecco's Modified Eagle Medium (Gibco #11965) supplemented with 10% FBS (Seradigm), 1X penicillin/streptomycin/glutamine (1X PSG) (Gibco #10378016). For each cell line a total of 100×10^6 cells were collected for six replicates. Cells were washed twice with ice cold PBS, and then scraped off the plate and snap frozen using a mix of ethanol and dry ice. For each cell line IP was performed in six replicates. 100×10^6 cells were later thawed

at 37°C for and each replicate was lysed for 30 sec 5 minutes lysed in 2ml of ice cold lysis

buffer, which contained 50mM EPPS pH 7.5 (Fisher Scientific #16052-06-5), 150mM NaCl, 1% Triton X-100 (Fisher # PI85111), cOmplete EDTA-free Protease Inhibitor Cocktail (1 tablet per 20mL of lysis buffer; Millipore Sigma #4693132001), 1:100 of sigma phosphatase inhibitor Cocktails 2 and 3 (Millipore Sigma #P5726; #P0044), cocktails and 250U/ μ L of benzonase (Millipore Sigma #E1014). Lysates were incubated on ice for 5 min to allow DNA digestion,

centrifuged at 20,000K g for 5 minutes, to remove insoluble material and filtered through acroprep 1.0um glass filter plate (PALL #8231) at 2000g for 1 minute. The concentration of protein was then estimated by BCA Protein Assay (ThermoFisher #23225). and immunoprecipitation was performed in 2ml deep well plates (Agilent #201379-100) with using 1mg of protein material, 5.75ug of V5 antibody (Invitrogen #R960-25) bound to 5.75μL of Protein G Sepharose (Millipore Sigma #17-0618-02). The final volume of IP reaction was 300μL.

The incubation was performed at 4°C shaking at 1100 rpm for 1 hour. The beads were then

transferred to OF 1100 filter plate (Orochem #1100) and washed 5 times with ice cold 50mM EPPS pH 7.5, 150mM NaCl using vacuum manifold. 18μL of 10mM EPPS pH 8.5 with 20ng/μL trypsin, 10ng/μL LysC was added to the beads in each well and digestion was performed for 2

hours at 37°C at 2000rpm. The partial digest was then collected into a 96- well PCR plate and

left overnight at room temperature to complete digestion. 4μL s of 22g/L 11 plex TMT tags (ThermoFisher #A34808) were added to each sample. (6 replicates of IP for V5-tagged CRKL overexpression sample and 5 replicates of IP from wildtype control CRKL overexpression (no V5-tag) cell line was combined in the TMT 11 plex experiment). The samples were then pulled and 20μL of the combined sample was set aside, while the rest was fractionated into 8 fractions using High pH Reversed-Phase Peptide Fractionation Kit (ThermoFisher #84868), as suggested by manufacturer. The fractions were concatenated into 4 fractions as follows: 1st and 5th fractions, 2nd and 6th, 3rd and 7th, and 4th and 8th and so on. 1μL of DMSO was added to each sample and evaporated in a speed vac. 1μL of DMSO was added to each sample to prevent complete evaporation and then resuspended in 20 μL of 0.1%TFA (Millipore Sigma #302031).

Data acquisition

5μL of unfractionated sample and every fraction was analyzed by EASY-nLC 1200 System (ThermoFisher) with using an EASY-Spray column (ThermoFisher #ES903) with 2μm particle size, 75μm diameter, and 500mm length in direct injection mode. The samples were separated using the following gradient at 300nl/min of buffer A (0.1% formic acid in water) and buffer B (0.1%formic acid in acetonitrile): 0%–5% in for 10 minutes, 5%–25% in for 92 minutes, and 25%–50% in for 18 minutes. The column was then washed with 95% buffer B (0.1%formic acid in acetonitrile) for 10 minutes at 400nl/min. Eluting peptides were analyzed on an Orbitrap Fusion Lumos mass spectrometer (ThermoFisher #IQLAAEGAAPFADBMBHQ) using MS3 SPS with the settings recommended by the instrument manufacturer for TMT11 plex analysis. Cycle time was set at 3 seconds and exclusion time was set at 45 seconds.

Data analysis

Data were analyzed in Proteome Discoverer 3.1 software. A database search was performed with Sequest HT search engine using Human UniProt database containing only reviewed entries and canonical isoforms (retrieved on 06/14/2019). Oxidation (M) was set as a variable modification, while TMT6plex was set as fixed modification. A maximum of two missed cleavages were permitted. The precursor and fragment mass tolerances were 10 ppm and 0.6 Da, respectively. Peptide spectrum matches (PSMs) were validated by percolator Percolator^{89,90} with a 0.01 posterior error probability (PEP) threshold. Only PSMs with isolation interference < 25% and at least 5 MS2 fragments matched to peptide sequence among the 10 selected for

MS3 were considered. The quantification results of PSMs were combined into protein-level quantitation using MSstatsTMT R package¹¹⁶. Only proteins with at least 3 peptides were reported. Significant interactors were defined by having an adjusted p-value < 0.05 and fold change > 2 compared to control. Proteins were then analyzed for previously described SH2 and SH3 binding motifs^{91,92}.

Plasmid Construction

The following plasmids were constructed using the Gateway Tol2kit.

mitfa:hsCRKL/394
mitfa:hsGAB2/394
mitfa:hsTERT/394
mitfa:Cas9-mCherry;zU6:*NF1a*:gRNA/394
mitfa:Cas9-mCherry;zU6:*NF1b*:gRNA/394
mitfa:dnIRS2-GFP/394
empty-vector/394
empty-vector/395
mitfa:Cas9/395
zU6:gRNA-hCRKL/394
zU6:gRNA-hox13-1/394
zU6:gRNA-hox13-2/394
zU6:gRNA-hoxb7a/394
zU6:gRNA-igf/394
zU6:gRNA-NT/394

cDNAs in tol2-compatible pENRT223.1 plasmids were ordered through Horizon Discovery for the following genes.

hsCRKL (Clone ID: 100000145)
hsGAB2 (Clone ID: 56810)
hsTERT (Clone ID: 100061944)

sgRNA sequences

nf1a: GGCGCACAAGCCCGTGGGAAT
nf1b: GGCGCAGAAGCCCGTGGAGT
hoxa13a: GGGCAATCACAACCAAGTGA
hoxa13b: GGATGATATGAGCAAAAACA
hoxb13a: GCGAGGATTCAGGACCAGGG
hoxc13a: CCGTGATATGACGACTTCGC
hoxd13a: GGCTCTGGCTCCTTCACGTT
CRKL-1: CGCGGACGAGGAACATACCG
CRKL-2: CAACCGCCGTTTTAAGATCG
CRKL-3: GTCGGTGTCCGAGAACTCGC
igf1: TCTAGCGGTCATTTCTTCCA
igf2a: TGCATCTTGCCGAAAAACGG
igf2b: GAAACTGTCTGTTCTCGAGC
Non-target (NT): AACCTACGGGCTACGATACGCGG

Zebrafish dnIRS2-GFP was a generous gift from ⁵⁶ and then further cloned into pENTR/D-TOPO vector.

Zebrafish Genotyping by PCR

Tail clips from adult zebrafish were placed in microcentrifuge tubes or thermal cycler plates containing 50 µl of 50 mM NaOH. Samples were boiled at 95°C for 30min, then cooled down

with 5 ml of 1M Tris-HCL (pH=8.0). A 1:10 dilution of the supernatant was used in PCR. DNA was PCR amplified with Promega GoTag green mastermix (Promega #M7123). PCR amplicon sizes for CRKL, GAB2, TERT, and Cas9 are 221bp, 125bp, 296bp, and 207bp. All transgenes have the same forward primer. GTTGAACGCAAGTTTGTACA going off the *mitfa* promoter. Primers used for each gene are listed below and listed in Supplemental Table 2.

CRKL: TGATGTAGTGGGAGACCCGC
GAB2: AGTATTCCAGAACATCTGGG
TERT: AGGCAGGACACCTGGCGGAA
Cas9-mCherry: CATGTGCACCTTGAAGCGCA

CRISPR-seq

Genomic DNA of transgenic zebrafish was isolated via tail clip genotyping described above. DNA was PCR amplified with Phusion polymerase (NEB #M0531S), run on an agarose gel, and then gel purified using NucleoSpin® Gel & PCR Clean-up Midi (Takara # 740986.20) for deep sequencing using the CRISPR-seq platform⁹³. Primers used for each gene are listed below and listed in Supplemental Table 2. Sequencing data was aligned to the zebrafish genome (GRCz10) and analyzed with CrispRvariantsLite version 1.2⁹⁴.

nf1a: TCGGGATCGCAAAAGTGATT
CACCAAGCTCACATCTTCAA
nf1b: CACCATCTTCATCATCCTCCT
ACTACTCTCTGTCCCGTGTC

Surveyor for sgRNA Validation

Recombinant Cas9 (IDT #1081059) and sgRNA were complexed and injected into Casper embryos at the one-cell stage. Briefly, 100 µM crRNA and tracrRNA (IDT #1075928) were mixed at 1:1 ratio and incubated at 95C for 5 minutes and then cooled to room temperature. 1.11µL of sgRNA duplex, 0.9µL (9µg) of recombinant Cas9, and 2.99µL of duplex buffer (IDT #11-05-01-12) were incubated at 37C for 10minutes. After incubation 2µL of 100ng/µL tol2 mRNA, 1µL of 100ng/µL empty-vector/395 and 2µL of phenol red (Sigma-Aldrich #P0290) were added to sgRNA:Cas9 complex. This mix was then used to inject 1pL into Casper embryos at the one-cell stage, resulting in delivery of 0.9pg of Cas9 protein, 10pg of empty-vector/395, and 20pg of tol2 mRNA. At 24hpf, embryos were sorted for the GFP+ heart marker from successful 395-empty incorporation. 10 embryos were used for genomic DNA isolation using the Qiagen DNAeasy Blood & Tissue kit (Qiagen #69504). DNA was PCR amplified with Promega GoTag green mastermix (Promega #M7123). Surveyor was performed using IDT Surveyor kit (IDT #706020) following manufacturer's instructions. Primers used for each gene are listed below and listed in Supplemental Table 2:

hoxa13a: GGGTGATTCTGGAAAGCAAT
CTCCATGGGATACTGACTCT
hoxa13b: ATCCCATTGTGCAATGGAAT
CGCCAAAATATCCATAGGGC
hoxb13a: TTGACATTCTTCACCCAAGG
AGCCCTGGTAGGATATTCTT
hoxc13a: GCCAGTAGTTGTTTAAAGGG
GTCTCGGCATATTTTTCTGC
hoxc13b: TAGTGAAAGACGTTTGCGTT
ACAACTGGGACGTCCAAATA
hoxd13a: ATGCACTGAGGAATATGGAC
CTAATGAAGAGAGGCGAGGA
hoxb7a: ATATATCATCACGTGCTGCC
CTCTACATACACAGACGCAC

CRKL-1: TTCGAATAAACATGTCGTCTGC
 CGCCTATCTTGAATCTCTTGCT
 CRKL-2: TCCACTTGTCTGCTGGTGATTATG
 TTCTATCAGGGTGGTCGTATCC
 CRKL-3: CTCCACTAGCGAGAAGCTGATT
 TCTTGAAAGGAAGGTCTTCAGC
 igf1: GGATTTTCTCTCAAATCCG
 CTTAATCATGTCGACTCAGC
 igf2a: CAAAAGAACCACTCGTTCAC
 CACACGAACTGCAGTGTATC
 igf2b: CTCTCAGAGAACTTTTGCCT
 GCGTATCCAGAACGTAATGT

RT-PCR Validation of Transgene Expression

Wildtype melanocyte skin, acral melanoma model tumor, and cutaneous melanoma model tumor tissue were dissected from zebrafish and RNA was isolated with the Zymo Quick RNA Miniprep kit (#R1050) using kit protocol. cDNA was synthesized using SuperScript III First Strand Synthesis System (Life Technologies #18080-400) following the kit protocol. RT-PCR was performed using iQ Sybr Green Supermix (Biorad #1708882). RT-PCR primer sequences are listed below.

mitfa: GCCCTATGGCCCTTCTCAC
 CATCCATGAACCCAAGAATGTCA
 hsGAB2: GCGGCGACGTGGTGT
 CTTCCAGGCATAGCGCCTC
 hsCRKL: Sino Biological qPCR primer pairs (HP101145)
 hsTERT: GGAGCAAGTTGCAAAGCATTG
 TCCCACGACGTAGTCCATGTT
beta-actin: GCCAACAGAGAGAAGATGACAC
 CAGAGAGAGCACAGCCTGG

Western Blot

Zebrafish lysates were prepared by sonication in RIPA buffer (Thermo #89901) with 1X Halt Protease and Phosphatase Inhibitor Cocktail (Thermo #78441) followed by centrifugation (14,000rpm for 10min at 4°C) and collection of the supernatant. Protein concentration was quantified by Bradford (Sigma B6916-500mL) according to manufacturer's protocol. Samples were mixed with 6X reducing loading buffer (Boston BioProducts #BP-788 111R) and denatured at 95°C for 10 minutes. Samples were run on a Mini PROTEAN TGX gel (BioRad) and transferred using Turbo Mini Nitrocellulose Transfer Pack (Bio-Rad, catalog #1704158). Membranes were blocked with 5% nonfat dry milk in TBST (1X TBS + 0.1% Tween 20) for 1 hour before incubation with primary antibody in PBS overnight at 4°C. Membranes were washed with TBST and incubated with secondary antibody in 5% nonfat dry milk for 1 hour at room temperature. Membranes were washed with TBST and developed with ECL (Amersham, RPN2109) using an Amersham Imager 600 (GE) or chemiluminescence film.

Cell lines were cultured in standard media conditions (RPMI 1640 Medium (Invitrogen #11875093) supplemented with 20% FBS (Seradigm), 1X penicillin/streptomycin/glutamine (Gibco #10378016)) prior to collection. Cells were washed with cold PBS and lysed in RIPA buffer (Pierce #89901) plus phosphatase and protease inhibitors (Thermo Scientific #1861277, #1861278). Lysates were cleared by centrifugation at 14000rpm at 4°C and quantified using BCA method (Pierce #23224). Samples were prepared using LDS+Reducing agent Novex buffers (Invitrogen #NP0008, #NP0009). 10 to 20µg of lysates were loaded and run on

NuPage™ 4-12% Bis-Tris gels (ThermoFisher #NP0321BOX) followed by transfer to nitrocellulose membranes (Biorad #1620233). Membranes were incubated over night with the indicated antibodies, washed and incubated again for 45 minutes with anti-rabbit or anti-mouse secondary antibodies. Detection was performed using Immobilon Western (Millipore #WBKLS0500).

Primary antibodies are: CRKL (32H4) #3182S (Cell Signaling Technology) (used for fish lysates), CRKL (B-1) sc-365092 Lot#B0819 (Santa Cruz Biotechnology) (used for human cell lines), GAB2 (26B6) #3239 (Cell Signaling Technology), IGF-1R beta (D23H3) #9750S (Cell Signaling Technology), phospho-IGF-1R beta Y1135 (DA7A8) #3918S (Cell Signaling Technology), HOXB13 (D7N8O) #90944 (Cell Signaling Technology), IGF2 #ab170304 (Abcam), and β -Actin (AC-74) #A2228 (Sigma-Aldrich). All the primary antibodies were used at 1:1000 dilution except for IGF1 used at 1:500 and β -Actin that was used at 1:10,000 dilution. All secondary antibodies were used at 1:10,000. Secondary antibodies include goat anti-rabbit IgG #ab97051 (Abca), rabbit anti-mouse IgG #ab97046, rabbit anti-goat IgG #HAF017 (R&D Systems).

Phospho-RTK array

Human Phospho-RTK arrays (R&D Systems, ARY001B) were used to detect activated RTKs according to manufacturer's instructions. WM3918 parental and WM3918-CRKL cell lines were serum starved for 24 hours and then stimulated with 20% FBS (Seradigm) for 10 minutes before collection. Cells were washed with cold PBS and lysed using the provided lysis buffer plus phosphatase and protease inhibitors. Lysates were cleared by centrifugation at 14000rpm at 4°C and quantified using BCA method (Pierce #23224). 200 μ g lysates were incubated on membranes overnight. Membranes were subsequently washed and exposed to chemiluminescent reagent (Millipore #WBKLS0500).

HOX13 siRNA Knockdown

Three human acral melanoma cell lines including SKMEL-1088, SKMEL-1176 and SKMEL-1206 were used for knockdown of four HOX13 genes, respectively. Non-targeting Control siRNA (D-001810-01-05), HOXA13 siRNA SMARTPool (#L-011052-00-0005), HOXB13 siRNA SMARTPool (#L-012226-00-0005), HOXC13 siRNA SMARTPool (#L-017600-00-0005) and HOXD13 siRNA SMARTPool (#L-011053-00-0005) were purchased from Horizon Discovery. Briefly, 4×10^5 cells were seeded in 6-well plates and cultured with RPMI 1640 medium (ThermoFisher # 11875093) containing 10% FBS (Seradigm), 1X PSG (Gibco #10378016) for 24-hr. Media was then replaced with 1ml of fresh medium before transfection. 100nM of total siRNA (25nM of each HOX siRNA) were added into 200 μ L of Gibco Opti-MEM (ThermoFisher #31985070) and mixed with 12 μ L I DharmaFECT Duo Transfection Reagent (Horizon Discovery #T-2010-03) followed by a 10-sec agitation. After 10-min incubating at room temperature, transfection mixtures were evenly dropped into each well. Media was aspirated 6-hr later and replaced with 2ml fresh RPMI medium containing 10% FBS, 1X PSG. After 48 hours post-transfection, cells were changed to 3mL of RPMI 20% FBS 1X PSG for Western western blot or 1.5mL of serum-free RPMI 1X PSG for ELISA. Samples were collected 96 hours post-transfection for both Western western blot and ELISA.

IGF2 ELISA

SKMEL-1088 and SKMEL-1176 were prepared as described above. At 96 hours post-transfection, cell supernatants were collected and spun at 1000G for 5min to remove cell debris. IGF2 concentration was determined by using the Human IGF2 Quantikine ELISA Kit (R&D Systems #DG200) following manufacturer's instructions.

Zebrafish Imaging Tumor-free Survival Curves

Zebrafish embryos were injected at the one-cell stage with indicated plasmids. *crestin:tdTomato* reporter was also utilized to aid in identifying tumors to monitor tumor-free survival^{72,73}. All injected embryos were grown to adulthood and then screened for melanocyte rescue at 2mpf based on a combination of GFP fluorescence and pigmentation. Zebrafish were regularly monitored every 2-4 weeks for the development of new tumors. Tumors were called based on the CAMP criteria outlined in Table S1. Kaplan Meir curves were generated and analyzed in Prism version 8.4.3. Statistical differences in tumor-free survival were determined by log-rank Mantel-Cox test. All imaging was performed on a Zeiss AxioZoom V16 using Zen 2.1 software.

Analysis of Anatomic Distribution of Tumors

During monitoring of tumor-free survival, the anatomic location (head vs body vs fins) of the tumor is noted. If a fish has either >2 tumors or a tumor that encroaches on the border of multiple anatomic sites, all tumors and anatomic sites are counted. To compare the anatomic distribution of tumors, a Chi-squared test was performed using Prism software version 8.4.3. Ternary diagrams were generated using ternaryplot.com. All imaging was performed on a Zeiss AxioZoom V16 using Zen 2.1 software.

Histology of Zebrafish and Human Samples

Zebrafish were sacrificed using ice-cold water. The head and tail were separated from the body separated via dissection with a clean razor. The sample of interest was then placed in 4%PFA (Santa Cruz #30525-89-4) for 72 hours on a shaker at 4°C and then immediately embedded in paraffin blocks to generate 5-micron sections. Human acral melanoma tissue TMA was generated as described above. Human and zebrafish samples were processed and stained using the same protocols described below. All H&E and IHC staining was performed at Histowiz. Tissue was put through a heat-induced epitope retrieval process (100°C), followed by a proprietary automatic staining assay developed for the BOND RX automated stainer (Leica Biosystems Division of Leica Microsystems Inc, Buffalo Grove, Illinois). Information about the antibodies used and staining conditions is listed below. The staining process included the use of the BOND Polymer Refine Red Detection Kit (Leica Biosystems) in accordance with the manufacturer's protocols, utilizing alkaline phosphatase and Fast Red chromogen. Endogenous alkaline phosphatase activity was blocked using a levamisole solution (Vector Laboratories). After staining, sections were dehydrated and film coverslipped using a Tissue-Tek Film automated Coverslipper (Sakura Finetek USA Inc, Torrance, California). Whole slide scanning (×40 objective) was performed on an Aperio AT2 digital whole slide scanner (Leica Biosystems). Human acral melanoma TMA stains for CRKL, HOXB13, and pIGF1R were analyzed and scored by a dermatopathologist at Memorial Sloan Kettering Cancer Center. Samples were scored for extent based on the following criteria: 0 = negative, 1= 1-25% of tumor cells are positive, 2 = 26-50% of tumor cells are positive, 3 = 51-75% of tumor cells are positive, and 4 = 76-100% of tumor cells are positive. Intensity was scored, 0 = negative, 1 = weak staining, 2 = intermediate staining, 3 = strong staining. N = 2/32 samples were removed from the analysis due to heavy pigmentation obscuring detection of staining signal, leading to an analysis of n = 30 samples total.

Antibodies used and conditions:

CRKL, Sigma HPA001100, Rabbit Polyclonal, 1:50 for fish, 1:25 for human, Antigen retrieval pH9 20mins

HOXB13, Cell Signaling CST90944, Rabbit Monoclonal [D7N8O], 1:50 for fish, 1:50 for human Antigen retrieval pH9 20mins

p-IGF1R, Abcam ab39398, Rabbit Polyclonal, 1:100 for fish, 1:100 for human, Antigen retrieval pH9 20mins

GFP, Abcam ab183734, Rabbit Monoclonal [EPR14104], 1:100 for fish, Antigen retrieval pH6 20mins

p-ERK Cell Signaling CST4370, Rabbit Monoclonal [D13.14.4E], 1:100 for fish, Antigen retrieval pH6 20mins

p-S6, Cell Signaling CST4858, Rabbit Monoclonal [D57.2.2E], 1:200 for fish, Antigen retrieval pH6 20mins

IF of Zebrafish Melanoma Samples

Unstained histology slides of zebrafish samples were deparaffinized using 2x10 minutes xylene, 4 minutes 100% ethanol, 1 minute 95% ethanol, 1 minute 70% ethanol, 1 minute 50% ethanol, and then rinsed with water. Antigen retrieval was achieved by placing slides in 10mM sodium citrate pH = 6.2 by heating in the microwave for approximately 2 minutes until the solution begins to bubble. The slides then sat for 5 minutes and then were microwaved again for 40 seconds. The slides cooled down to room temperature for 30 minutes in the same buffer. The slides were blocked in a blocking solution consisting of 5% donkey serum (Millipore Sigma #S30-M), 1% BSA, 0.4% Triton X-100 (Fisher # PI85111) in PBS. 50uL of blocking buffer was added per section. The blocking solution was incubated at room temperature for 1 hour. Primary antibodies were diluted at 1:100 in blocking solution and incubated overnight at 4°C. Slides were washed 3 x 5 minutes in PBS and then applied with secondary antibody solution 1:250 in blocking solution. Hoescht (Fisher # H3570) was added to be 1:1000 dilution. The slides were washed again 3 x 5 minutes in PBS. Coverslips were mounted onto the slides using Vectashield Vibrance Antifade Mounting Medium (Vector Laboratories #H-1800-2). Slides were then imaged on a confocal microscope. Primary antibodies used were Sox10 (GeneTex #GTX128374) and GFP (Abcam #5450). Secondary antibodies used were Alexa-anti Rabbit 594 (Cell Signaling Technology #8889) and Alexa-anti Goat 488 (ThermoFisher #A32814).

Pharmacologic Treatment of Zebrafish Embryos

Zebrafish embryos were collected at 1dpf and placed in a 40-micron cell strainer (Thermo Fisher # 08-771-2) in a 6-well dish (Fisher #08-772-1B) in 6mL of E3 water. 20 embryos were used per well. Zebrafish embryos were treated with the indicated compounds: insulin/IGF1 receptor antagonists BMS-754807 (Sigma-Aldrich #BM0003-5MG) at 7.5µM and NVP-AEW541(Selleckchem # S1034) at 60µM, PI3K inhibitor LY294002 (Sigma-Aldrich #L9908-1MG) at 15µM, RAF/MEK inhibitor CH5126766 (Selleckchem # S7170) at 1µM, MEK inhibitor Pimasertib (Selleckchem #S1475) at 1µM, MEK inhibitor Refametinib (Selleckchem #S1089) at 1µM, MEK inhibitor Trametinib (Selleckchem #S2673) at 200nM, and SOS1 inhibitor BI-3406 at 1µM (MedChemExpress #HY-125817). Compound stocks were kept at 1000X in DMSO. Treatment started at 1dpf and reapplied at 2dpf. Zebrafish were imaged for tailfin area at 3dpf.

Measurement of Zebrafish Melanocyte Cell Area in Tailfin

The extent of melanocyte area in the tailfin was calculated as the area of the melanophore covering the tailfin mesenchyme. This was quantified by using MATLAB to perform background subtraction of autofluorescence and then used FIJI to threshold on GFP (from MiniCoopR-eGFP) intensity to highlight the pixels that represent melanocytes in each image. All zebrafish were imaged at 3dpf. If treated pharmacologically, treatment started at 1dpf at the indicated concentration for 48 hours and imaged at 3dpf. For experiments involving genetic perturbation

using mitfa:Cas9/395, the GFP heart marker in this plasmid was used to screen for zebrafish with successful plasmid integration and expression.

Flow Cytometry of Zebrafish Melanocytes

Zebrafish Embryos

Approximately 250 5dpf embryos were euthanized via tricaine, transferred to an Eppendorf tube (Eppendorf # 022431021), and spun at 500G x 3 minutes at room temperature to remove E3 supernatant. 400µL of trypsin (Invitrogen # 25200-114) was added and incubated for 20 minutes at 28°C. Every 5 minutes an RNase-free disposable pellet pestle (Fisher # 12-141-364) was used to mash the embryos for 2-3 minutes into a single cell solution. After digestion was complete, 500µL of phenol-free Dulbecco's Modified Eagle Medium (ThermoFisher # 21063029) with 10% FBS (Seradigm) (DMEM10) was added to each sample. The samples were centrifuged 500G x 5 minutes at room temperature and supernatant was gently aspirated. Calcein red stock (Cayman Chemicals # 20632) at 1mM was diluted 1:4000 to desired concentration of 250nM in phenol-free DMEM10 and 500µL was added to embryo pellet and incubated for 30 minutes at 28°C to be used to identify viable cells. Samples were centrifuged again 500G x 5 minutes, supernatant was removed, and pellet was resuspended in phenol-free DMEM with 2% FBS (Seradigm) (DMEM2). Samples were filtered through 40-micron cell-strainer (Thermo Fisher # 08-771-2) two times and then placed in a flow cytometry tube (Fisher # 08-771-23). 1mL of DMEM2 and 1µL of 1000X DAPI (Sigma-Aldrich #D9542-10MG) was added and analyzed on the FACS sorter (BD FACS Aria). No color and single-color controls were used to gate for GFP+/Calcein Red+ double positive cells to calculate the frequency of viable melanocytes from the bulk cell suspension. The reporter frequency of viable melanocytes in Figure S3B were calculated on the FACS sorter.

Zebrafish Adults

Zebrafish were euthanized via placement in ice-cold water and then dissected to separate body skin and all fins. A clean razor was used to dice the sample into small pieces that can fit through a wide bore p1000 tip (ThermoFisher #2069G). Samples were placed into a 15mL Falcon tube (Fisher 14-959-49D) with 3mL of 1X PBS (Invitrogen 14190-250) and 187.5µL of 2.5mg/mL liberase (Millipore Sigma # 05401020001) and incubated at room temperature for 30 minutes on a shaker to gently keep tissue in suspension. At 15 minutes, a wide bore p1000 tip was used to pipette up and down gently for 3 minutes to dissociate the tissue. After the 30 minutes incubation in liberase at room temperature, 250µL of FBS (Seradigm) was added and then another 3 minutes of pipetting up and down using a wide bore p1000 tip was performed. Cells were then filtered through a 40-micron cell strainer (Thermo Fisher # 08-771-2) into a 50mL conical. Samples were spun at 500G x 5 minutes at 4 °C and the supernatant was carefully aspirated using a Pasteur pipette (Fisher #13-678-20D) with low vacuum suction. The pellet was then resuspended in 500µL of PBS with 5% FBS (Seradigm) to be used as flow buffer. Cells were drawn up cells in a p1000 tip with a regular bore size and then fit on 40um filter onto the tip and place in a flow cytometry tube (Fisher # 08-771-23). 0.5µL of 1000X DAPI (Sigma-Aldrich #D9542-10MG) was added and samples were placed on ice. Samples were then FACS sorted (BD FACS Aria) for GFP-positive signal and gated based on a GFP-negative control. Each biological replicate represents the pooling of 2 males and 2 female adult zebrafish of the indicated genotype. The reporter frequency of viable melanocytes in Figure 2H were calculated on the FACS sorter. Data was analyzed using FlowJo 10.6.1.

Preparation of Zebrafish Sample for Bulk RNA-seq

Zebrafish Embryos

Samples were FACS sorted as described above into 200 μ L of Trizol (Invitrogen #15596026) in RNase-free LoBind Eppendorf tubes (Eppendorf # 022431021) and snap-frozen by placing on dry ice. Samples were shipped to Genewiz (South Plainfield, NJ), where the RNA isolation and RNA-sequencing was performed.

Zebrafish Adult Body Skin and Fins

4 WT stable line fish (2 males and 2 females) and 4 acral melanoma stable line fish (2 males and 2 females) were pooled for each biological replicate. Fish were approximately 6 months post-fertilization and acral melanoma fish had early-stage fin tumors. Samples were processed and FACS sorted as described above into 750 μ L of Trizol LS (Invitrogen # 10296010) in RNase-free LoBind Eppendorf tubes (Eppendorf # 022431021) and snap-frozen by placing on dry ice. RNA isolation was performed by following Trizol LS protocol. Glycogen (Millipore Sigma #10901393001) was used to co-precipitate the RNA. RNA quality and quantity was measured by Bioanalyzer (Agilent). RNA samples were shipped to Genewiz (South Plainfield, NJ), where RNA-sequencing was performed.

Analysis of Zebrafish Bulk RNA-seq

RNA-sequencing was performed using SMART-seq v4 Ultra Low Input RNA Kit (Clontech). Libraries were constructed using Illumina Nextera XT kit and were analyzed for concentration by Qubit and for size distribution by Agilent Bioanalyzer. Paired-end sequencing was performed on Illumina HiSeq 2500. After quality control with FASTQC (Babraham Bioinformatics) and trimming with TRIMMOMATIC⁸³ when necessary, reads were aligned to GRCz11 (Ensembl version 96) with transgenes added using STAR⁵⁴, with quality control via SeQC⁹⁵. Differential expression was calculated with DESeq2⁵⁶ using the output of the --quantMode GeneCounts feature of STAR. The rlog function was used to generate log2 transformed normalized counts. Pathway and Gene Ontology (GO) analysis were performed with GSEA using FGSEA-multilevel⁵⁷. Ortholog mapping between zebrafish and human was performed with DIOPT⁹⁶. Only orthologs with a DIOPT score greater than 6 were used for GSEA and heatmap generation. In cases of more than one zebrafish ortholog of a given human gene, the zebrafish gene with the highest average expression was selected. GSEA comparing the gene expression of GFP+ and GFP- samples was used to validate the melanocyte identity of GFP+ samples (Extended Data Fig. 7b).

Zebrafish Single-cell RNA-seq

Sample preparation

4 acral melanoma stable line fish (2 males and 2 females) approximately 6 months post-fertilization were utilized and had early-stage fin tumors. All 4 fish were pooled together and dissected into a fin and body skin sample, which were then digested and FACS sorted as described above. The fin and body skin samples were FACS sorted for GFP+ and GFP- populations, sorting 200,000 cells for each group, creating n=4 samples total for scRNA-seq. Cells were FACS sorted into a LoBind Eppendorf (Eppendorf # 022431021) tube containing 600 μ L DMEM (ThermoFisher # 21063029) supplemented with 20% FBS and 1X penicillin/streptomycin/glutamine and placed on ice. Samples were then centrifuged 300g for 5 minutes in a bucket centrifuge at 4°C. The supernatant was aspirated, and the pellet was resuspended in 40 μ L of DMEM supplemented with 10% FBS (Seradigm) and 1X penicillin/streptomycin/glutamine to ensure a cell concentration of at least 1000cells/ μ L. The viability of cells was above 80%, as confirmed with 0.2% (w/v) Trypan Blue staining (Countess II).

Cell encapsulation, library preparation, and sequencing

Droplet-based scRNA-seq was performed using the Chromium Single Cell 3' Library and Gel Bead Kit v3 (10X Genomics) and Chromium Single Cell 3' Chip G (10X Genomics). Approximately 10,000 cells were encapsulated per each of the four reactions. GEM generation and library preparation was performed according to kit instructions. Libraries were sequenced on a NovaSeq S4 flow cell. Sequencing parameters were: Read1 28 cycles, i5 10 cycles, i7 10 cycles, Read2 90 cycles. Sequencing depth was approximately 40,000 reads per cell. Sequencing data was aligned to our reference zebrafish genome using Cell Ranger version 5.0.1 (10X Genomics).

Analysis

Data was processed using R version 4.0.2 and Seurat version 4.0.3⁹⁷. Each of the four reactions were processed separately before merging into a single object. Cells with fewer than 200 unique genes were filtered out. Expression data was normalized with SCTransform⁹⁸. Principal component analysis⁹⁹ and UMAP dimensionality reduction¹⁰⁰ were performed using default parameters, with 15 principal components used for UMAP calculations. Clustering was done using the Seurat function FindMarkers with a resolution of 0.2. Clusters were annotated based on expression of zebrafish cell-type specific marker genes as done previously^{101,102}. Expression of *hox13* genes was calculated by summing the normalized expression per cell for each of *hoxa13a*, *hoxa13b*, *hoxb13a*, *hoxc13a*, *hoxc13b*, *hoxd13a*.

Quantification and Statistical Analysis

Statistical comparisons were performed with the aid of GraphPad Prism 8.4.3 and the statistical details including sample size can be found in the figure legends. A p value of >0.05 is not considered statistically significant. * indicates p<0.05, ** indicates p<0.01, *** indicates p<0.001 and **** indicates p<0.0001.

RNA-seq and ChIP-seq analysis was analyzed in R version 4.0.2 and R Studio version 1.2.5033.

ChIP-seq data was visualized in IGB (BioViz)¹⁰³.

Cut & Run data was visualized in IGV 2.10.2.

Images were captured using Zen 2.1 software.

Image analysis performed in Matlab version Update 4 (9.6.0.1150989) and FIJI version 2.0.0.

Methods references:

- 45 White, R. M. et al. Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell* 2, 183-189, doi:10.1016/j.stem.2007.11.002 (2008).
- 46 White, R. M. et al. DHODH modulates transcriptional elongation in the neural crest and melanoma. *Nature* 471, 518-522, doi:10.1038/nature09882 (2011).
- 47 Kaufman, C. K. et al. A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science* 351, aad2197, doi:10.1126/science.aad2197 (2016).
- 48 Cancer Genome Atlas, N. Genomic Classification of Cutaneous Melanoma. *Cell* 161, 1681-1696, doi:10.1016/j.cell.2015.05.044 (2015).
- 49 Dankort, D. et al. Braf(V600E) cooperates with Pten loss to induce metastatic melanoma. *Nat Genet* 41, 544-552, doi:10.1038/ng.356 (2009).

- 50 Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 25, 1754-1760, doi:10.1093/bioinformatics/btp324 (2009).
- 51 Li, H. & Durbin, R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 26, 589-595, doi:10.1093/bioinformatics/btp698 (2010).
- 52 Cibulskis, K. et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol* 31, 213-219, doi:10.1038/nbt.2514 (2013).
- 53 Shen, R. & Seshan, V. E. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res* 44, e131, doi:10.1093/nar/gkw520 (2016).
- 54 Dobin, A. et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29, 15-21, doi:10.1093/bioinformatics/bts635 (2013).
- 55 Li, B. & Dewey, C. N. RSEM- accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics* 12, 1471-2105 (2011).
- 56 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 15, 550, doi:10.1186/s13059-014-0550-8 (2014).
- 57 Korotkevich, G., Sukhov, V. & Sergushichev, A. Fast gene set enrichment analysis. *bioRxiv*, doi:10.1101/060012 (2019).
- 58 Heinz, S. et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 38, 576-589, doi:10.1016/j.molcel.2010.05.004 (2010).
- 59 Khan, A. et al. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res* 46, D260-D266, doi:10.1093/nar/gkx1126 (2018).
- 60 Grossman, R. L. et al. Toward a Shared Vision for Cancer Genomic Data. *N Engl J Med* 375, 1109-1112 (2016).
- 61 Hoadley, K. A. et al. Multiplatform analysis of 12 cancer types reveals molecular classification within and across tissues of origin. *Cell* 158, 929-944, doi:10.1016/j.cell.2014.06.049 (2014).
- 62 Cancer Genome Atlas Research Network. Electronic address, e. d. s. c. & Cancer Genome Atlas Research, N. Comprehensive and Integrated Genomic Characterization of Adult Soft Tissue Sarcomas. *Cell* 171, 950-965 e928, doi:10.1016/j.cell.2017.10.014 (2017).
- 63 Cancer Genome Atlas Research Network. Electronic address, w. b. e. & Cancer Genome Atlas Research, N. Comprehensive and Integrative Genomic Characterization of Hepatocellular Carcinoma. *Cell* 169, 1327-1341 e1323, doi:10.1016/j.cell.2017.05.046 (2017).
- 64 Robertson, A. G. et al. Comprehensive Molecular Characterization of Muscle-Invasive Bladder Cancer. *Cell* 171, 540-556 e525, doi:10.1016/j.cell.2017.09.007 (2017).
- 65 Fishbein, L. et al. Comprehensive Molecular Characterization of Pheochromocytoma and Paraganglioma. *Cancer Cell* 31, 181-193, doi:10.1016/j.ccell.2017.01.001 (2017).
- 66 Cancer Genome Atlas Research, N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* 511, 543-550, doi:10.1038/nature13385 (2014).
- 67 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* 513, 202-209, doi:10.1038/nature13480 (2014).
- 68 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* 507, 315-322, doi:10.1038/nature12965 (2014).
- 69 Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330-337, doi:10.1038/nature11252 (2012).
- 70 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* 490, 61-70, doi:10.1038/nature11412 (2012).
- 71 Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* 489, 519-525, doi:10.1038/nature11404 (2012).

- 72 Ciriello, G. et al. Comprehensive Molecular Portraits of Invasive Lobular Breast Cancer. *Cell* 163, 506-519, doi:10.1016/j.cell.2015.09.033 (2015).
- 73 Cancer Genome Atlas Research, N. The Molecular Taxonomy of Primary Prostate Cancer. *Cell* 163, 1011-1025, doi:10.1016/j.cell.2015.10.025 (2015).
- 74 Cancer Genome Atlas Research, N. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 455, 1061-1068, doi:10.1038/nature07385 (2008).
- 75 Cancer Genome Atlas, N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* 517, 576-582, doi:10.1038/nature14129 (2015).
- 76 Cancer Genome Atlas Research, N. et al. Integrated genomic characterization of endometrial carcinoma. *Nature* 497, 67-73, doi:10.1038/nature12113 (2013).
- 77 Cancer Genome Atlas Research, N. Comprehensive molecular characterization of clear cell renal cell carcinoma. *Nature* 499, 43-49, doi:10.1038/nature12222 (2013).
- 78 Davis, C. F. et al. The somatic genomic landscape of chromophobe renal cell carcinoma. *Cancer Cell* 26, 319-330, doi:10.1016/j.ccr.2014.07.014 (2014).
- 79 Cancer Genome Atlas Research, N. et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368, 2059-2074, doi:10.1056/NEJMoa1301689 (2013).
- 80 Cancer Genome Atlas Research, N. Integrated genomic analyses of ovarian carcinoma. *Nature* 474, 609-615, doi:10.1038/nature10166 (2011).
- 81 Brennan, C. W. et al. The somatic genomic landscape of glioblastoma. *Cell* 155, 462-477, doi:10.1016/j.cell.2013.09.034 (2013).
- 82 Cancer Genome Atlas Research, N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 159, 676-690, doi:10.1016/j.cell.2014.09.050 (2014).
- 83 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120, doi:10.1093/bioinformatics/btu170 (2014).
- 84 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9, 357-359, doi:10.1038/nmeth.1923 (2012).
- 85 Li, H. et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).
- 86 Ramirez, F. et al. deepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Res* 44, W160-165, doi:10.1093/nar/gkw257 (2016).
- 87 Zhang, Y. et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol* 9, R137, doi:10.1186/gb-2008-9-9-r137 (2008).
- 88 Skene, P. J., Henikoff, J. G. & Henikoff, S. Targeted in situ genome-wide profiling with high efficiency for low cell numbers. *Nat Protoc* 13, 1006-1019, doi:10.1038/nprot.2018.015 (2018).
- 89 Kall, L., Canterbury, J. D., Weston, J., Noble, W. S. & MacCoss, M. J. Semi-supervised learning for peptide identification from shotgun proteomics datasets. *Nat Methods* 4, 923-925, doi:10.1038/nmeth1113 (2007).
- 90 The, M., MacCoss, M. J., Noble, W. S. & Kall, L. Fast and Accurate Protein False Discovery Rates on Large-Scale Proteomics Data Sets with Percolator 3.0. *J Am Soc Mass Spectrom* 27, 1719-1727, doi:10.1007/s13361-016-1460-7 (2016).
- 91 Sparks, A. B. et al. Distinct ligand preferences of Src homology 3 domains from Src, Yes, Abl, Cortactin, p53bp2, PLCy, Crk, and Grb2. *PNAS* 93, 1540-1544 (1996).
- 92 Birge, R. B., Kalodimos, C., Inagaki, F. & Tanaka, S. Crk and CrkL adaptor proteins: networks for physiological and pathological signaling. *Cell Commun Signal* 7, 13, doi:10.1186/1478-811X-7-13 (2009).
- 93 Tothova, Z. et al. Multiplex CRISPR/Cas9-Based Genome Editing in Human Hematopoietic Stem Cells Models Clonal Hematopoiesis and Myeloid Neoplasia. *Cell Stem Cell* 21, 547-555 e548, doi:10.1016/j.stem.2017.07.015 (2017).

- 94 Lindsay, H. et al. CrispRVariants charts the mutation spectrum of genome engineering experiments. *Nat Biotechnol* 34, 701-702, doi:10.1038/nbt.3628 (2016).
- 95 DeLuca, D. S. et al. RNA-SeQC: RNA-seq metrics for quality control and process optimization. *Bioinformatics* 28, 1530-1532, doi:10.1093/bioinformatics/bts196 (2012).
- 96 Hu, Y. et al. An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* 12, 1471-2105 (2011).
- 97 Hao, Y. et al. Integrated analysis of multimodal single-cell data. *Cell* 184, 3573-3587 e3529, doi:10.1016/j.cell.2021.04.048 (2021).
- 98 Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol* 20, 296, doi:10.1186/s13059-019-1874-1 (2019).
- 99 Jolliffe, I. T. *Principal Component Analysis and Factor Analysis*. (Springer, 1986).
- 100 McInnes, L., Healy, J. & Melville, J. (arXiv:1802.03426, arXiv.org, 2018).
- 101 Baron, M. et al. The Stress-Like Cancer Cell State is a Consistent Component of Tumorigenesis. *Cell Systems* 11, 536-546.E537, doi:https://doi.org/10.1016/j.cels.2020.08.018 (2020).
- 102 Hunter, M. V., Moncada, R., Weiss, J. M., Yanai, I. & White, R. M. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. *Nat Commun* 12, 6278, doi:10.1038/s41467-021-26614-z (2021).
- 103 Freese, N. H., Norris, D. C. & Loraine, A. E. Integrated genome browser: visual analytics platform for genomics. *Bioinformatics* 32, 2089-2095, doi:10.1093/bioinformatics/btw069 (2016).