Spatially resolved transcriptomics reveals the architecture of the tumor/microenvironment interface

Miranda V. Hunter[1]*, Reuben Moncada[2]*, Joshua M. Weiss[1,3], Itai Yanai[2 †], Richard M. White[1,4 †]
[1] Cancer Biology and Genetics, Memorial Sloan Kettering Cancer Center, New York, NY.
[2] Institute for Computational Medicine, NYU Langone Health, New York, NY.
[3] Weill Cornell/Rockefeller/Sloan Kettering Tri-Institutional MD-PhD Program, Memorial Sloan Kettering Cancer Center, New York, NY.
[4] Lead Contact
*Equal contribution
[†] Correspondence: itai.yanai@nyulangone.org (I.Y); whiter@mskcc.org (R.M.W)

SUMMARY
During tumor progression, cancer cells come into contact with novel non-tumor cell types, but it is unclear how tumors adapt to these new environments. Here, we integrate spatially resolved transcriptomics, single-cell RNA-seq, and single-nucleus RNA-seq to characterize tumor-microenvironment interactions at the tumor boundary. Using a zebrafish model of melanoma, we identify a distinct “interface” cell state where the tumor contacts neighboring tissues. This interface is composed of specialized tumor and microenvironment cells that upregulate a common set of cilia genes, and cilia proteins are enriched only where the tumor contacts the microenvironment. Cilia gene expression is regulated by ETS-family transcription factors, which normally act to suppress cilia genes outside of the interface. An ETS-driven interface is conserved in human patient samples, suggesting it is a conserved feature of human melanoma. Our results demonstrate the power of spatially resolved transcriptomics in uncovering mechanisms that allow tumors to adapt to new environments.

KEYWORDS
melanoma, zebrafish, tumor microenvironment, transcriptomics, cilia, cell-cell interaction

INTRODUCTION

As tumors grow and invade into new tissues, they come into contact with various new cell types, but it is poorly understood how these cell-cell interactions allow for successful invasion and tumor progression. In melanoma, these interactions can occur between the tumor cells and a diverse number of cell types. In many cases, the tumor cells interact directly with stromal cells such as fibroblasts[1] or immune cells[2]. However, increasing evidence suggests that the repertoire of such interactions is considerably broader, and can include cell types including adipocytes[3] and keratinocytes[4]. Many of these cell interactions can influence tumor cell behavior[1–4].

There are likely at least two levels of cell-cell interactions that are relevant to cancer: “microenvironmental” interactions in which the tumor cell directly interacts with adjacent non-tumor cells, and “macroenvironmental” interactions, in which the tumor cell indirectly interacts with more distant cells. The microenvironment is increasingly appreciated to play a

major role in cancer phenotypes, including proliferation, invasion, metastasis and drug resistance[5,6]. However, it is debatable whether every cell type that a tumor interacts with is truly part of the microenvironment, since the mechanisms by which these cells influence tumor cell behavior are often unclear. This uncertainty is compounded by the fact that tumor cells themselves are highly heterogeneous[7], making it challenging to determine which subset of tumor cells are directly interacting with surrounding non-tumor cells. The macroenvironment may also influence tumor progression, since the tumor cells can interact with other cells in the body at a distance, as recently demonstrated for metabolic coupling between melanoma cells and distant cells in the liver[8].

A better understanding of the nature of these cell-cell interactions requires high resolution imaging and analyses of genes expressed by tumor cells as they interact with different cell types. While bulk and single-cell RNA-sequencing approaches have improved our ability to understand cell-cell interactions, these techniques require dissociation of the tissue of interest, resulting in a loss of spatial information. Thus, a comprehensive understanding of how tumor and surrounding cells interact *in situ* is lacking, at least in part due to the limitations of current RNA-sequencing technologies.

Spatially resolved transcriptomics (SRT) has recently emerged as a way to address the limitations of both bulk and single-cell RNA-seq, by preserving tissue architecture while still profiling the genes expressed by the cell or tissue at high resolution. Current SRT techniques typically either use spatially-barcoded probes to capture and sequence mRNA from tissue sections[9,10], or multiple rounds of *in situ* hybridization, sequencing, and imaging to computationally reconstruct the transcriptional landscape of the cell[11,12]. *In situ* hybridization-based SRT techniques allow the user to profile the transcriptional landscape of the cell at cellular or even subcellular resolution, whereas the resolution of techniques that capture and sequence mRNA from sections is limited by the diameter of each capture spot on the SRT array (for example, 55 μm with the current 10X Genomics Visium SRT technology). However, we recently developed a technique to integrate capture probe-based SRT and scRNA-seq to map the transcriptomic and cellular architecture of tumors[13]. This provides a unique opportunity to understand the mechanisms that are driving the cell-cell interactions that occur between the tumor and its immediately adjacent microenvironment.

Here, we integrate SRT, single-cell RNA-seq, and single-nucleus RNA-seq to characterize the transcriptional landscape of melanoma cells as they interact with the immediately adjacent microenvironment. Using a zebrafish model of melanoma, we construct a spatially-resolved gene expression atlas of transcriptomic heterogeneity within tumors and surrounding tissues. We discover a histologically invisible but transcriptionally distinct “interface” region where tumors contact neighbouring tissues, composed of cells in specialized tumor-like and microenvironment-like states. We uncover enrichment of cilia genes and proteins at the tumor boundary, and find that ETS-family transcription factors regulate cilia gene expression specifically at the interface. We further demonstrate that this distinct “interface” transcriptional

state is conserved in human melanoma, providing a conserved mechanism that presents new opportunities for halting melanoma invasion and progression.

RESULTS

Spatially resolved transcriptomics reveals the architecture of the melanoma/microenvironment interface

To investigate the transcriptional landscape of tumor-microenvironment cell interactions *in situ*, we processed frozen sections from 3 adult zebrafish with large, invasive *BRAF*^{V600E}-driven melanomas[14] for capture probe-based spatially resolved transcriptomics (SRT), using the 10X Genomics Visium platform (**Fig. 1a-c**). Although the size of the tissue section used is limited by the size of the Visium array (6.5 mm^2), zebrafish allow us the unique advantage that a transverse section through an adult fish (~5 mm diameter) fits in its entirety on the array (**Fig. 1c**). Zebrafish are thus one of the only vertebrate animals that can be used to study both the tumor and all surrounding tissues in their intact forms, without any need for dissection. Our SRT dataset contained transcriptomes for 7,281 barcoded array spots across three samples, encompassing 17,317 unique genes (**Fig. 1d-e** and **S1**). We detected approximately 1,000 - 15,000 transcripts (unique molecular identifiers, UMIs) and 500 - 3,000 unique genes per spot, with somewhat fewer UMIs/genes detected in sample C (**Fig. S1**). Visium array spots within the tumor region typically contained more UMIs than spots in the rest of the tissue (**Fig. S1d,f**), likely at least in part due to higher density of cells within the tumor region (**Fig. S1g**).

We first combined our expression matrices using an anchoring framework to identify common cell states across different datasets[15] (**Fig. S2a**). After community-detection based clustering on our integrated dataset, we inferred the identities of 13 distinct clusters. When we projected the cluster assignments back onto the tissue coordinates (**Fig. 1d**) and onto the UMAP embeddings for each spot (**Fig. 1e**), we found complex spatial patterns in the data that strongly recapitulated tissue histology. Our Visium data captured multiple microenvironment cell types (muscle, liver, brain, skin, pancreas, heart, intestine, gills) in addition to the *BRAF*^{V600E}-driven melanomas (**Fig. 1d-f**). We validated our cluster assignments by plotting onto the Visium array the expression of marker genes that should be expressed exclusively in the tumor (*BRAF*^{V600E}), muscle (*pvalb4*), heart (*kcn6a*) and nervous system (*mbpa*), and observed that expression of these marker genes was restricted to the expected regions of the tissue (**Fig. 1f**).

To further characterize the transcriptional architecture of the microenvironment, we asked whether we could leverage publicly available, annotated gene sets to uncover spatially-organized patterns of biological activity across the tissue. To this end, we computed the mean expression of genes associated with all zebrafish Gene Ontology (GO) terms, and measured the distance between spots that highly express these genes, reasoning that shorter distances between spots may represent underlying spatial organization of these biological pathways across the tissue. We then compared this distribution to that of a null distribution of distances between random spots, allowing us to identify GO terms with spatially coherent, non-random expression patterns. Applying this to the tumor region of our samples, we identified several GO terms displaying interesting spatial expression patterns related to tissue structure

(GO: extracellular structure organization; p = 2.3x10^{-8}) and the immune system (GO: macrophage migration, p = 7.1x10^{-4}), among others (**Figs. 1g** and **S3a**; **Video S1**). We performed the same analysis on the microenvironment, and found several notable spatially-organized pathways that function in tumor growth and invasion (GO: lipid import into cell, p = 1.2x10^{-96}; GO: IMP biosynthetic process, p = 2.0x10^{-40}) (**Figs. 1h** and **S3b**; **Video S2**). Together, these data validate our spatially resolved transcriptomics workflow and demonstrate the existence of discrete tumor and microenvironment regions within our SRT dataset.

The tumor-microenvironment interface is transcriptionally distinct from the surrounding tissues

We noticed in all of the samples a transcriptionally distinct cluster of array spots that localized to the border between the tumor and the adjacent microenvironment (**Fig. 1c-e**), in which specific biological pathways were upregulated (**Fig. 1h**). This "interface" cluster was present in all three samples (**Fig. S2**). Interestingly, the tissue in this interface region appeared largely indistinguishable from the surrounding microenvironment (muscle) (**Fig. 2a**), despite it being transcriptionally distinct (**Fig. 1e**). We thus hypothesized that this interface cluster represented the region in which the tumor was contacting neighboring tissues. To get a better sense of the transcriptional profile of the interface cluster, we computed the correlation between the averaged transcriptomes of each SRT cluster across all 3 samples. We found that the transcriptional profile of the interface cluster was more correlated with the tumor (R = 0.33) than with muscle (R = 0.06) (**Fig. 2b**), despite the fact that the tissue in this region histologically resembles muscle with few tumor cells visible (**Fig. 2a**).

We next sought to identify genes that may differentiate the interface from muscle (to which it is most similar histologically) and from tumor (to which it is most similar transcriptionally). We found a number of genes that were upregulated specifically in the interface cluster, including, interestingly, a number of uncharacterized genes (**Fig. 2c**). To identify gene expression programs that are enriched specifically at the interface and provide further evidence for the interface as a transcriptionally distinct tissue region, we performed non-negative matrix factorization (NMF) on all microenvironment spots (including both interface and muscle clusters) across all samples (see Methods). When we projected the NMF factor scores onto each spot, we found that multiple factors were specifically enriched at the interface between the tumor and the microenvironment (factors 7 and 8; **Figs. 2d** and **S4**). To investigate the biology underlying the genes contributing to each factor, we looked for significantly enriched GO terms among the top 150 genes contributing to each factor (**Figs. 2e** and **S4**). This revealed several factors enriched in muscle-specific genes, as expected (**Fig. S4**), and that the interface factors were enriched in genes functioning in membrane-bound organelles and protein targeting to organelles (**Figs. 2e** and **S4**). Some factors were enriched across all three samples (e.g. factor 2, **Fig. S4**), whereas some were only enriched in one or two of the samples (e.g. factors 4, 11, **Fig. S4**). These differences may be due to different tissue types present across the three samples. Together, these data uncover a distinct "interface" region bordering the tumor, which histologically resembles the microenvironment, transcriptionally resembles tumor, but expresses distinct gene modules that may contribute to tumor-microenvironment cell interactions.

The tumor-microenvironment interface is composed of specialized cell states

Our SRT results so far detail a transcriptionally distinct "interface" region where tumors contact the microenvironment. However, spatially resolved transcriptomics data is limited in resolution by the diameter of each spot on the Visium array (55 µm with current technology). Thus, each array spot probably captures transcripts from multiple cells. As the interface region is, by nature, likely a mixture of tumor and microenvironment cells, we performed single-cell RNA-seq (scRNA-seq) on tumor and non-tumor cells from 3 adult zebrafish with large melanomas (**Fig. 3a**) in order to better define the cell states present in the interface. We detected approximately 10,000-75,000 transcripts and 1,000-5,000 unique genes per cell (**Fig. S5a,c**). As expected, our scRNA-seq data contained tumor cells as well as various non-tumor cell types, including erythrocytes, keratinocytes, and several types of immune cells (**Fig. 3b**). We did not identify a muscle cell cluster in our scRNA-seq dataset, likely due to the fact that adult skeletal muscle is composed of multinucleated muscle fibers that cannot be isolated and encapsulated for droplet-based scRNA-seq.

Consistent with our SRT results, clustering of our scRNA-seq data revealed a distinct "interface" cell cluster (**Fig. 3b-d**), which we identified based on the fact that cells in this cluster significantly upregulated the same genes that were upregulated in our SRT interface cluster ($p = 1.83 \times 10^{-26}$; **Fig. 3c**). The distinct clustering of the interface population was not due to the presence of a significant number of cell doublets within this cluster (**Fig. S5g**). Strikingly, UMAP and principal component analysis of the interface cluster revealed two distinct cell populations (**Figs. 3d** and **3f**), one expressing tumor markers such as *BRAF*^{V600E} and the other expressing muscle genes such as *ckba*, with other genes such as the centromere gene *stra13* upregulated in both populations (**Fig. 3e-f**). This result suggests that the transcriptionally distinct "interface" region we identified in our SRT data is actually composed of at least two similar, but distinct cell states: a "tumor-like interface" and a "muscle-like interface". The interface region may not be limited to only tumor-like and muscle-like cell states; however, since zebrafish melanomas frequently invade into muscle, this likely contributes to the presence of muscle-like interface cells in our data.

Based on this, we separated the interface cluster into two subclusters, and confirmed that the two subclusters express anti-correlated levels of tumor markers such as *BRAF*^{V600E}, *mitfa*, and *pmela*, and muscle markers such as *ckba*, *neb* and *ak1* (**Fig. 3f**). A common set of genes, including many genes related to the microtubule cytoskeleton and cell proliferation such as *stra13*, *stmn1a*, *plk1*, and *haus4*, were upregulated in both subclusters (**Fig. 3d-f**). Both the tumor-like and muscle-like interface cell states were present in both scRNA-seq samples (**Fig. S5c,f**). The presence of putative "muscle" cells in the interface is particularly notable, in light of the fact that adult skeletal muscle is composed of multinucleated muscle fibers that we were unable to isolate in our scRNA-seq workflow due to their size, evidenced by the lack of a muscle cell cluster in our dataset (**Fig. 3b**). This could suggest the presence of mono-nucleated muscle cells, or a hybrid tumor-muscle cell state at the invasive front. Previous work suggests that tumor and microenvironment cells can fuse to create a hybrid cell state that contributes to tumor

heterogeneity and metastasis[16]. Together, these data suggest that the interface region is composed of specialized tumor-like and microenvironment-like cell states.

Interface cell states are distinct from neighboring tissues

Our results so far indicate that we have uncovered an "interface" cell state localized to where the tumor contacts neighboring tissues. However, our scRNA-seq dataset does not contain a muscle cell cluster due to the fact that muscle fibers cannot be encapsulated for scRNA-seq. This makes it difficult to assess whether the specialized muscle cell state found in the interface is truly distinct from muscle that is not in proximity to the tumor. Thus, to effectively compare the interface cell state(s) to other microenvironment cell types/states that cannot be captured with scRNA-seq, we validated our scRNA-seq results by performing single-nucleus RNA-seq (snRNA-seq) on nuclei extracted from 3 adult zebrafish, all with large transgenic melanomas. Although snRNA-seq captures only nascent transcripts in the nucleus, which contains only 10-20% of the cell's mRNA[17], scRNA-seq and snRNA-seq typically recover the same cell states/types, albeit sometimes in different proportions[18]. After quality control and filtering, our dataset encompassed transcriptomes for 10,748 individual nuclei (**Fig. 4a**). We detected an average of 3800 UMIs and 1350 unique genes per nucleus (**Fig. S6a-b**). Overall gene expression was lower in our snRNA-seq dataset compared to our scRNA-seq dataset, likely due to the relatively low number of transcripts found in the nucleus in general (**Fig. S6c**). Dimensionality reduction and clustering revealed 12 distinct clusters encompassing tumor and non-tumor cell types, including muscle, keratinocytes, liver, various immune cells and erythrocytes (**Fig. 4a**). We also identified an "interface" cluster in our snRNA-seq dataset (**Fig. 4a**). We identified the interface cluster based on the fact that nuclei in this cluster strongly upregulated genes that were strongly upregulated in the interface cluster in our scRNA-seq dataset, including *stmn1a* (**Figs. 3d** and **4b**), *stra13* (**Figs. 3e** and **4b**), *plk1* (**Figs. 3e** and **4b**), and *haus4* (**Figs. 3f** and **4b**), and that the interface cluster from our snRNA-seq dataset clustered with the interface cluster from our scRNA-seq dataset when the two datasets were integrated[15,19] (**Fig. 4c**).

To interrogate the types of nuclei present in the interface cluster in our snRNA-seq dataset, we performed dimensionality reduction (UMAP) and clustering on the nuclei from the interface cluster, which identified 5 discrete subclusters (**Fig. 4d**). Similar to our scRNA-seq dataset, within the interface cluster in our snRNA-seq dataset we identified subclusters of nuclei that upregulated tumor- or muscle-specific genes (**Fig. 4d-e**). The interface cluster in our snRNA-seq dataset also contained other subclusters that did not express tumor- or muscle-specific genes. Nuclei in these subclusters expressed genes related to other cell types in our snRNA-seq dataset, including immune cells (*ctss2.1*), liver (*fabp10a*) and digestive system (*ela2*) (**Fig. 4d-e**). This is in line with recent work showing that melanomas can reprogram microenvironmental cells such as liver cells even when not in physical contact[8]. However, similar to our scRNA-seq and SRT datasets, there were many genes that were specifically upregulated across the interface subclusters that were not upregulated in any other cell type in the snRNA-seq dataset, further suggesting that the "interface" cell state is a distinct transcriptional entity (**Fig. 4f**).

Since our snRNA-seq dataset contained more cells/nuclei and a greater breadth of cell types than our scRNA-seq dataset, we integrated our snRNA-seq data with our Visium SRT data using our recently developed multimodal intersection analysis (MIA) method[13] to confirm the presence of tumor-like and microenvironment-like cell states within the interface region (see Methods; **Fig. S7a-b**). Notably, our MIA results suggested that the interface regions in our SRT dataset were enriched in cell types including muscle, macrophages, and tumor, in line with our scRNA-seq and snRNA-seq results (**Fig. S7b**). The cluster that was most significantly enriched in the interface region was the muscle-like interface cell state (**Fig. S7b**), in accordance with the histology of our SRT samples that showed that the interface region closely resembles the surrounding muscle (**Fig. 2a**). Together, these results suggest that the interface is composed of cells in various tumor-like or environment-like cell states, which together upregulate a unique gene program that may contribute to tumor growth into new tissues.

As our SRT results suggest that the interface cell state may be modulated by direct cell-cell interactions between tumor and microenvironment cells, we used NicheNet[20] to computationally infer interactions between interface cells and the rest of the cells in our snRNA-seq dataset by identifying potential ligands expressed by interface cells and receptors and target genes in the other cell types (**Fig. S8a-c**). As the NicheNet model is currently designed to work with human genes, we performed this analysis with the human orthologs of the zebrafish genes in our dataset (see Methods). The top ligand predicted to be active in interface nuclei was *HMGB2* (**Fig. S8a**), of which there are two zebrafish orthologs: *hmgb2a* and *hmgb2b*. These genes were highly expressed in the interface clusters across our snRNA-seq, scRNA-seq and SRT datasets (**Fig. S8d**). Interestingly, *HMGB2* expression has been reported to be correlated with tumor aggressiveness[21,22]. The predicted receptors for *HMGB2* were *AR*, *ITPR1* and *CDH1* (fish orthologs: *ar*, *itpr1a*, *itpr1b*, *cdh1*) (**Fig. S8b**). Of these potential receptors, *cdh1* was the most highly expressed in general across the 3 datasets (**Fig. S8e**). *cdh1* was expressed in various microenvironment cell types, including intestinal cells, keratinocytes, and also in some interface cell states (**Fig. S8e**). *cdh1* (E-cadherin) is a core component of adherens junctions along with α-catenin and β-catenin[23]. Interestingly, HMGB2 and β-catenin have been reported to cooperate to promote melanoma progression[24]. These data demonstrate one of likely many signaling interactions that occur between interface cells and other cells adjacent to the tumor. Taken together, our results suggest that we have identified a putative “interface” cell state in each of our SRT, scRNA-seq, and snRNA-seq datasets, composed of tumor and microenvironment cells which upregulate a common gene program that may contribute to tumor growth into surrounding tissues.

Cilia genes and pathways are upregulated at the interface

To gain further insight into the biological processes underlying the specialized “interface” region identified in our SRT and scRNA-seq data, we performed pre-ranked gene set enrichment analysis (GSEA), using differentially expressed genes in the scRNA-seq interface cluster, to identify conserved pathways that may be active in interface cells. We noticed that many cilia-related pathways were enriched in the combined interface cluster (**Fig. 5a**). This

enrichment of cilia-related pathways occurred in both the muscle-like and tumor-like interface cell states (**Fig. 5b**). Cilia-related GO terms were also enriched in the SRT interface (**Fig. 5c**), as were GO terms related to membrane-bound organelles in the genes contributing to NMF factor 7, which localized to the interface (**Fig. 2d-e**). Several recent studies have implicated cilia in an important role in melanoma initiation and progression, although the mechanism by which cilia mediate melanoma progression is unclear [25–29]. To further investigate a role for cilia at the tumor-microenvironment interface, we scored each cell from our scRNA-seq dataset for relative enrichment of cilia genes, using the "gold standard" SYSCILIA gene list[30], and quantified a significant upregulation of cilia genes in both interface cell states in our scRNA-seq data (**Fig. 5d**). Although cilia genes generally were expressed at relatively low levels in our snRNA-seq dataset, in line with the overall lower expression of most genes in our snRNA-seq data relative to our scRNA-seq data (**Fig. S6c-d**), the most highly upregulated cilia genes in the scRNA-seq interface cluster were also upregulated across the tumor-like and muscle-like cell states in the snRNA-seq interface cluster relative to the tumor and muscle clusters (**Fig. 5e**). Furthermore, we quantified a clear enrichment of cilia genes such as *ran*, *tubb4b, tuba4l*, and *gmnn* specifically in tumor-like and muscle-like interface cells in our snRNA-seq dataset (**Fig. 5f**). Together, these results suggest an important role for cilia at the tumor-microenvironment interface.

The tumor-microenvironment interface is ciliated

Interestingly, previous studies have shown that human and mouse melanomas are not ciliated, although they express cilia genes[25–29]. To reconcile these models, we stained sections through adult zebrafish with invasive *BRAF*^{V600E}-driven melanomas for acetylated tubulin, a common cilia marker[27]. Strikingly, we found that although the bulk of the tumor was not ciliated as expected (**Fig. 5g, center** and **S9**), there was a specific enrichment of cilia at the invasive front of the tumor, where it contacts the muscle. We observed long, acetylated tubulin-positive projections that were often found in the extracellular space spanning tumor and adjacent muscle cells (**Fig. 5g, left, 5h,** and **S9**). These projections were not found in the bulk of the tumor (**Fig. 5g, center** and **S9**) or in muscle that was not adjacent to the tumor (**Fig. 5g, right** and **S9**). We could not conclusively determine whether these cilia originated in tumor cells, muscle cells or both cell types, an interesting area that awaits further study. These data suggest that although the bulk of primary melanomas is not ciliated, cilia are enriched at the tumor-microenvironment interface, where they may facilitate growth of the tumor into surrounding tissues.

ETS-family transcription factors regulate cilia gene expression at the interface

To identify potential regulators of gene expression within the interface, we performed HOMER motif analysis[31] to identify conserved transcription factor (TF) binding motifs enriched in genes differentially expressed in the interface. When we performed *de novo* motif enrichment analysis on genes differentially expressed in the SRT interface compared to normal muscle, the top-ranked motif was the highly conserved ETS DNA-binding domain, containing a core GGAA/T sequence (p = 1x10^{-22}; **Fig. 6a**). The ETS domain was also the top-ranked motif enriched in genes differentially expressed in the SRT interface compared to all other SRT spots (p = 1x10^{-15}), and was the second-ranked motif enriched in genes differentially expressed in the

interface cluster identified in our scRNA-seq dataset (p = 1x10^{-13}; **Fig. 6a**) and in genes differentially expressed in our snRNA-seq interface cluster (p = 1x10^{-13}; **Fig. 6a**). Furthermore, ETS motifs were frequently enriched in both the tumor-like and muscle-like interface subclusters in our scRNA-seq dataset, along with, notably, motifs for RFX-family transcription factors which regulate ciliogenesis[32] (**Fig. 6b**). Although ETS-family transcription factors have not been widely studied in melanoma, they have been reported to function in melanoma invasion[33] and phenotype switching[34], and are aberrantly upregulated in many types of solid tumors[35]. Interestingly, zebrafish ETS-family transcription factors were downregulated in the interface in each of our scRNA-seq, snRNA-seq, and SRT datasets (**Fig. 6c-e**).

To identify potential biological processes that could be regulated by ETS transcription factors at the tumor-microenvironment interface, we investigated putative target genes containing an ETS motif in their promoter. We queried the zebrafish genome for genes with an ETS motif within 500 bp of the transcription start site, filtered these genes to include only those differentially expressed in the tissue/cell state of interest, and performed GSEA on the resulting target gene lists. Surprisingly, within the ETS-target genes in both the SRT and scRNA-seq interface clusters, we again found an enrichment of pathways related to cilia (**Fig. 6g**). As ETS TFs are downregulated specifically in the interface (**Fig. 6d-f**), this suggests that ETS-family TFs may act as a transcriptional repressor of cilia genes. ETS TFs can act as transcriptional activators and/or repressors depending on gene and context (**Table S1**). In support of this model, when we scored each cell in the interface for relative expression of both ETS genes and ETS-target genes, the two were strongly anti-correlated (R = -0.625, p = 9.02x10^{-27}, **Fig. 6h**). Collectively, these data suggest that ETS-family transcription factors act as transcriptional repressors of cilia genes in cells at the interface between tumors and the microenvironment, where upregulation of cilia may contribute to tumor-microenvironment cell interactions.

Interface cells are conserved in human melanoma

Finally, we investigated whether “interface” cells, distinguished by upregulation of cilia genes and downregulation of ETS transcription factors, are also found in human melanoma. We analyzed a published scRNA-seq dataset of 4,645 cells isolated from 19 patients, comprising tumor (malignant) cells in addition to several other microenvironment cell types[36] (**Fig. S10a**). We scored each cell for expression of the human orthologs of the genes upregulated by more then 1.5-fold in our scRNA-seq interface cluster (**Fig. S10b**), and classified cells that upregulated these genes as an “interface” population (**Fig. S10c**). Similar to our snRNA-seq results, “interface” cells were found across all the cell types in the human melanoma dataset (**Fig. S10b-c**). These cells significantly upregulated cilia genes (p = 3.8x10^{-5}; **Fig. S10d**), and slightly downregulated ETS genes (p = 2.3x10^{-2}, **Fig. S10e**). Furthermore, we noticed that expression of ETS and cilia genes was inversely correlated across all the cells and cell types in this dataset (R = -0.1462; p = 1.31x10^{-23}; **Fig. S10d-e**). This suggests that a transcriptionally distinct “interface” cell state is conserved in human melanoma. Identifying which human melanoma subtypes (e.g., BRAF, NRAS, c-KIT, etc) in which an interface cell state is found awaits larger datasets of freshly isolated tumors subjected to scRNA-seq and/or SRT. Together, our results suggest that cell-cell interactions at the tumor-microenvironment interface are

accomplished by a subset of specialized tumor and muscle cells, which together upregulate a conserved common gene program characterized by upregulation of cilia genes and downregulation of ETS transcription factors.

DISCUSSION

Here, we combined spatially resolved and single-cell and single-nucleus transcriptomics approaches to characterize how tumor cells interface with new tissues in their surrounding environment, revealing key regulators of how this interface is formed. We analyzed a total of 25,563 transcriptomes encompassing expression of 20,589 unique genes from 7,281 spatial array spots, 2,889 zebrafish cells, 10,748 zebrafish nuclei, and 4,645 human cells. Our results identified a series of spatially-patterned gene modules, some of which specifically localize to the interface between tumors and surrounding tissue. We showed that the interface is composed of specialized tumor and muscle cell states, which are distinguished by upregulation of cilia genes and proteins. We further show that ETS transcription factors regulate expression of cilia genes at the interface, and that a distinct "interface" cell population is conserved in human melanoma patient samples. Together, our results reveal a novel "interface" transcriptional state that may mediate melanoma growth into surrounding tissues.

Our results identify a novel role for ETS-family transcription factors in mediating cilia gene expression at the interface. Interestingly, in recent years cilia have been implicated in multiple facets of melanoma biology, but their role in melanoma progression is still unclear. The bulk of melanomas are not ciliated[27,28] (**Fig. 5g-h** and **S9**), and in fact, the "ciliation index" is gaining prominence as a diagnostic tool to distinguish melanomas from benign nevi[25,29]. Furthermore, cilia disassembly has recently been implicated in melanoma metastasis[26], in which deconstruction of cilia, regulated by EZH2, drives metastasis. The paradox is that while most melanoma cells are not ciliated, many melanomas still express cilia genes (**Fig. 5d-e**). Our data adds a new layer onto this complexity, in that we find that not only are cilia genes upregulated specifically at the interface between tumor and microenvironment, but more importantly that only cells at that interface express high levels of cilia proteins. This raises the still not fully answered question of what role cilia play in various steps in melanoma progression. In primary melanoma growth, it is clear that most cells are unciliated, and that EZH2 acts to suppress those genes. Loss of cilia via EZH2 increases metastasis in these models via enhanced Wnt/β-catenin signaling[26]. Our finding that most melanoma cells do not have cilia is consistent with this finding, but yet we find a specific subset of cells at the interface that upregulate cilia genes and protein, and these cells appear to be present in human melanoma as well.

How to reconcile these seemingly conflicting pieces of data? Our data would suggest that intact cilia may be most important when they are first encountering new, heterotypic cell types in the neighboring environment. We can envision several different possibilities to why cilia are upregulated specifically at this interface. First, this upregulation of cilia genes and proteins at the interface may be transient, induced by heterotypic cell-cell interactions between tumor and muscle. Primary cilia are critical signalling hubs for the cell, and regulate signaling pathways such as Hedgehog and TGF-β/BMP[37], all of which are important in cancer progression[38] and

cell-cell communication. Our NicheNet analysis suggests that there may be distinct ligand/receptor pairs, including HMG family proteins, that may mediate such signaling. A second possibility is that the primary cilia are acting as mechanotransducers, and play a role in directional migration of the cells as they invade into new tissues. For example, seminal work on primary cilia demonstrated that cilia can orient in the direction of migration in 3T3 cells[39], which has been also seen in the context of wound healing[40,41]. Finally, it is possible that the emergence of cilia at the interface is actually acting as a barrier to systemic metastatic dissemination, and that heterotypic interactions between melanoma and muscle might be restraining progression. It is notable that our zebrafish melanomas metastasize at a low rate, and in fact skeletal muscle (where we most easily see the interface) is a rare site of metastasis in humans, consistent with this possibility. A major endeavor for future studies will be to delineate how cilia act at each step of tumorigenesis, what signaling nodes are most critical, and whether they act as a barrier or enabler of metastasis.

Our results uncovered a novel role for ETS-family transcription factors in melanoma, as potential transcriptional repressors of cilia genes. Although most ETS TFs can function as transcriptional activators, at least four ETS TFs are known to have repressor activity[42] (**Table S1**). Despite the fact that ETS TFs have a well-characterized role in several types of solid tumors, their role in melanoma has not been studied in depth, although a recent study found that ETS TFs induce a UV damage signature that correlates with increased mutational burden in human melanoma[43]. ETS TFs broadly function in various facets of tumorigenesis, including DNA damage, metabolism, self-renewal, and remodelling of the microenvironment[35]. However, most if not all of these situations have been found to be induced by aberrant upregulation of ETS genes. Conversely, we found a role for downregulation of ETS TFs specifically where tumors contact surrounding tissues. It is still unclear what triggers this downregulation of ETS genes in such a spatially-restricted region. Despite their role as transcription factors, ETS proteins also participate in a wide range of protein-protein interactions, and their activity is regulated through phosphorylation as a result of signaling cascades[44]. MAPK signaling has been reported to regulate ETS[45], and the MAPK pathway is frequently activated in melanoma[46]. It is unclear if MAPK or other signaling pathways display spatially-restricted patterns of activation within tumors and/or the microenvironment, but the advent of SRT techniques will help to address these questions.

Although it was not a focus of our study, our SRT dataset also uncovered spatially-organized transcriptomic heterogeneity within the tumor itself (**Figs. 1f-g**). In recent years, the advent of single-cell transcriptomics approaches has identified a substantial degree of transcriptomic heterogeneity in most if not all types of cancer[47]. Tumor heterogeneity often increases as tumors progress, and may be a predictor of poor clinical outcomes as it is believed to be a major contributor to drug resistance[48]. Investigating the underlying cause and complex clonal relationships within different tumor cell subtypes has proven to be challenging for many reasons, one of which being a lack of information regarding the spatial patterning of this heterogeneity. Our dataset acts as a proof-of-principle for the use of spatially resolved transcriptomics in

identifying spatially-organized tumor heterogeneity, and lays the groundwork for future studies using our dataset or others to explore the basis of this spatial patterning.

Our study is, to our knowledge, the first spatially-resolved gene expression atlas of the interface between the tumor and its environment. Although we uncovered many genes, pathways and gene modules that exhibit novel spatial patterns within the tumor and/or environment, there are likely many more interesting biological phenomena in our dataset that we have yet to identify. Recently, deep learning methods have been applied to histopathology images to uncover spatially-resolved predictions of molecular alterations, mutations, and prognosis[49,50]. A logical next step would be extension of these approaches to integrate deep learning and pattern recognition algorithms with SRT data, to identify interesting spatial patterns of gene expression and also predict transcriptomes based on histopathology. Ultimately, integration of transcriptomics, histopathology, and deep learning techniques will allow us to expand the utility of both SRT and histological datasets and broaden our understanding of cancer cell interactions *in vivo*.

ACKNOWLEDGEMENTS
We thank S. Selvaraj and B. Dabovic from the NYU Experimental Pathology Core for technical assistance with imaging for Visium experiments, R. Luther and M. Hogan from the Maurano lab at NYU for assistance with sequencing, R. Chaligne and the MSK Single Cell Research Initiative for assistance with snRNA-seq, and the members of the Yanai and White labs for useful discussions and technical support. M.V.H was funded by a postdoctoral fellowship from the Canadian Institutes of Health Research. J.M.W. was supported by a NIH Kirschstein-NRSA predoctoral fellowship (F30CA236442), a NIH predoctoral fellowship (T32GM008539) from the Cell and Developmental Biology Program at Weill Cornell Graduate School, and a NIH Medical Scientist Training Program grant (T32GM007739). R.M.W was funded by grants 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 Alan and Sandra Gerry Metastasis Research Initiative at Memorial Sloan Kettering Cancer Center, The Harry J. Lloyd Foundation, Consano, and the Starr Cancer Consortium.

AUTHOR CONTRIBUTIONS
M.V.H, R.M, I.Y and R.M.W conceived the study. M.V.H and R.M performed all experiments and data analysis. J.M.W generated transgenic fish. M.V.H and R.M wrote the manuscript and all authors provided feedback before submission.

DECLARATION OF INTERESTS
R.M.W is a paid consultant to N-of-One Therapeutics, a subsidiary of Qiagen. R.M.W is on the scientific advisory board of Consano, but receives no income for this. R.M.W receives royalty payments for the use of the *casper* zebrafish line from Carolina Biologicals.

FIGURE LEGENDS

Figure 1. Spatially resolved transcriptomics reveals the transcriptional architecture of melanoma and its surrounding microenvironment.

a. Schematic showing the spatially resolved transcriptomics (SRT) experiment workflow. **b.** Images of zebrafish with *BRAF*^{V600E}-driven melanomas used for SRT. The region where the fish were sectioned is highlighted. **c.** H&E staining of cryosections used for SRT. **d.** Visium array spots colored by clustering assignments of the integrated dataset (see Methods). **e.** UMAP embedding of SRT spots from all three samples colored by cluster assignments of the integrated dataset (see Methods). **f.** The expression of select marker genes (*BRAF*^{V600E}, tumor; *pvalb4*, muscle; *kcn6a*, heart; *mbpa*, nervous system) from the SRT data projected over tissue space (left), with images of the corresponding histology from the indicated region of the SRT array (right). **g-h.** Average, standardized expression of annotated genes for gene ontology (GO) terms displaying spatially-coherent expression patterns in the tumor (g) and microenvironment (h) regions in each sample. *p*-values represent the comparison between the distance between spots expressing that GO term genes and a null-distribution of distances between random spots (Wilcoxon's Rank Sum test, see Methods).

Figure 2. The tumor-microenvironment interface is transcriptionally distinct from the surrounding microenvironment.

a. Interface and muscle-annotated cluster spots projected onto tissue image. Insets show the tissue underlying the interface spots (1) and muscle spots (2). **b.** Correlation matrix between average expression profile of SRT clusters across all three datasets. Clusters are ordered by hierarchical clustering of the Pearson's correlation coefficients (see Methods) and bubble sizes correspond to *p*-value (-log10) of correlation, with *p*-values < 10^{-3} omitted. Clustering of tumor and interface together is highlighted in the dendrogram (red). **c.** Volcano plot of differentially expressed genes between the interface cluster versus the muscle and tumor clusters. *p*-values were obtained from the Wilcoxon's rank sum test. **d.** Non-negative matrix factorization (NMF) of the microenvironment spots (muscle and interface clusters). Shown are the standardized factor scores for interface-specific NMF factor 7, projected onto microenvironment spots. Arrows denote areas with higher factor scores. **e.** Enriched GO terms for the top 150 scoring genes in NMF factor 7.

Figure 3. The tumor-microenvironment interface is composed of specialized tumor and muscle cells.

a. Schematic showing scRNA-seq experiment workflow. **b.** UMAP dimensionality reduction plot for 2,889 cells sequenced as in (a). Cluster/cell type assignments are labelled and colored. **c.** Expression score per cell (scRNA-seq) for average expression of interface marker genes from the SRT interface cluster. **d.** Inset of the outlined interface cluster in (b) showing the two interface subclusters. **e.** Heatmap showing expression of the top 50 genes upregulated in the tumor cell cluster (top, orange) and interface cell cluster (bottom, yellow). Selected genes are labelled. **f.** Principal component analysis of cells in the interface cluster, scored for expression of the tumor marker *BRAF*^{V600E}, the muscle marker *ckba*, and the centromere gene *stra13*. Scores for principal components 1 and 2 are plotted. Cells are labelled by standardized expression of the indicated genes. **g.** Dot plot showing expression of tumor and muscle markers. The size of

each dot corresponds to the percentage of cells in that cluster expressing the indicated gene, and the colour of each dot indicates the expression level.

Figure 4. Single-nucleus RNA-seq demonstrates that the interface cell states are distinct from the rest of the microenvironment.
a. snRNA-seq cluster assignments plotted in UMAP space. **b.** Expression of marker genes from the scRNA-seq interface cluster in the snRNA-seq dataset. **c.** Integrated UMAP of the snRNA-seq and scRNA-seq datasets (labelled, top plot) showing colocalization of the two interface clusters (bottom plot). **d.** Subcluster assignments and expression of marker genes from the snRNA-seq interface cluster. **e.** Dotplot showing expression of microenvironment cell-type specific genes within the interface subclusters. **f.** Heatmap showing expression of the top 100 genes upregulated across all of the interface subclusters.

Figure 5. Cilia genes and proteins are enriched at the tumor-microenvironment interface.
a. Waterfall plot showing the top and bottom 250 GO cellular component terms by normalized enrichment score (NES) in the scRNA-seq interface cluster, with cilia GO terms labelled in red. **b.** Cilia-related GO term enrichment scores within the tumor-like and muscle-like interface cell states from the scRNA-seq dataset. **c.** Cilia-related GO term enrichment scores for the scRNA-seq and SRT interface clusters. **d.** Relative expression of fish SYSCILIA genes in the scRNA-seq interface cluster. **e.** Relative expression of the top 25 SYSCILIA genes upregulated in the scRNA-seq interface cluster across the snRNA-seq clusters. **f.** Normalized expression of selected cilia genes across the snRNA-seq interface clusters. **g.** Immunofluorescent images of sections through adult zebrafish with invasive melanomas, stained for GFP (tumor cells), acetylated tubulin (cilia) and Hoescht (nuclei), showing the tumor-muscle interface (left), center of the tumor (middle), and distant muscle (right). Arrows denote cilia at the interface. Scale bars, 100 μm. **h.** Inset of region highlighted in (g) (left). Scale bars, 25 μm. ***, $p < 0.001$ (Wilcoxon rank sum test with Bonferroni's correction).

Figure 6. ETS transcription factors regulate cilia gene expression at the interface in zebrafish and human melanoma.
a. Results from HOMER *de novo* motif analysis of differentially expressed genes in the SRT, scRNA-seq, and snRNA-seq interface clusters. **b.** Top 10 enriched motifs from HOMER known motif analysis of the scRNA-seq tumor-like (left) and muscle-like (right) interface cell states. **c-e.** Relative expression of zebrafish ETS genes across the clusters in the scRNA-seq (c), SRT (d) and snRNA-seq (e) datasets. **f.** Normalized enrichment score of cilia-related pathways enriched in ETS-target genes within the SRT and scRNA-seq interface clusters. **g.** Scatter plot comparing ETS gene expression scores per cell and ETS-target gene expression scores per cell in the scRNA-seq interface cluster. Pearson's correlation coefficient (R) is indicated. **, $p < 0.01$, ***, $p < 0.001$ (Wilcoxon rank sum test with Bonferroni's correction).

METHODS

Zebrafish husbandry

Zebrafish lines were maintained at 28.5°C in a dedicated aquatics facility with a 14 hours on/10 hours off light cycle. *casper* [51] fish were used for all experiments. Fish were anesthetized with Tricaine (MS-222) at a stock concentration of 4 g/L (pH 7.0). All zebrafish experiments and procedures were carried out in compliance with institutional protocols for vertebrate animals, and were approved by the Memorial Sloan Kettering Cancer Center IACUC (protocol #12-05-008).

Generation of transgenic fish

Transgenic tumor-bearing fish were generated using the *miniCoopR* system as previously described[52,53]. Briefly, *casper* fish with the genotype *mitfa-BRAF*^{V600E}*; p53*$^{-/-}$*; mitfa*$^{-/-}$ were incrossed, and the resulting 1-cell stage embryos were injected with plasmids containing *mitfa-MITF* and *mitfa-GFP*. Fish were raised to adulthood (4-6 months) and screened for the presence of pigmented, GFP-positive tumors.

Spatially resolved transcriptomics

Sample preparation

Adult tumor-bearing fish were euthanized on ice and washed in 1X PBS. After dissection of the head and tail, the remaining tissue was equilibrated in cold OCT for 2 minutes, before transfer to a tissue mold filled with fresh OCT for snap-freezing in liquid nitrogen-chilled isopentane. Tissue blocks were stored at -80°C. For cryosectioning, both the tissue block and the Visium slide were equilibrated inside the cryostat for 15-30 minutes at 16°C before sectioning. Transverse sections through the entire fish were cut at a thickness of 10 µm and immediately placed on the Visium array slide (Visium Spatial Gene Expression Slides, 10X Genomics). Array slides containing sections were stored at -80°C for a maximum of one week before use.

Fixation, staining, imaging and construction of cDNA libraries

Samples were processed according to the Visium Spatial Gene Expression User Guide (10X Genomics) and all reagents were from the Visium Spatial Gene Expression Kit (10X Genomics). Briefly, sections were fixed in chilled methanol for 30 min at -20°C, stained with hematoxylin and eosin, and mounted in 85% glycerol for imaging. Imaging was performed on a Leica SCN400 F whole-slide scanner at 40X magnification. After imaging, sections were permeabilized at 37°C for 45 minutes. After permeabilization, the on-slide reverse transcription (RT) reaction was performed at 53°C for 2 hours. Permeabilization time and RT reaction length were determined using the Visium Spatial Tissue Optimization Kit (10X Genomics). Second strand synthesis was subsequently performed on-slide for 15 minutes at 65°C. All on-slide reactions were performed in a thermocycler with a metal slide adaptor plate. Following second strand synthesis, samples were transferred to tubes for cDNA amplification and cleanup. Library quality was assayed using a Bioanalyzer High Sensitivity chip (Agilent).

Sequencing

10X Genomics Visium libraries were pooled, denatured, and diluted to a loading concentration of 1.8 pM with 1% PhiX control, followed by paired-end sequencing on an Illumina NextSeq 500 to a depth of approximately 110 - 180 million paired reads per sample. Sequencing parameters:

Read1 28 cycles. i7 10 cycles, i5 10, Read2 120 cycles. Sequencing data was processed using the Space Ranger pipeline v.1.0.0 (10X Genomics).

Dimensionality reduction and clustering

SRT data was processed using R version 3.6.3, Seurat version 3.1.4[15], Python version 3.6, and MATLAB 2019b. Data was normalized using SCTransform[54]. The 3 SRT datasets were integrated using the Seurat SCTransform integration workflow, using 3000 integration features and including all common genes between the 3 datasets. Principal component analysis[55] and UMAP dimensionality reduction[56] were done using default parameters. Initial clustering was done using the FindClusters function implemented in the Seurat R package with the resolution parameter = 0.8. Tissue types of each cluster were inferred and clusters were further refined by plotting clusters onto the associated histology images and identifying marker genes using the Wilcoxon's Rank Sum test. Expression scores for ETS and cilia gene sets were calculated using the Seurat function AddModuleScores with default parameters. A list of cilia genes was obtained from the SYSCILIA gold standard list[30]. A list of ETS genes was obtained from[57].

Identification of genes enriched in the SRT interface

To identify genes that were enriched at the interface in the SRT data, we had to first take into account the likelihood of a mixture of tumor and muscle cells being present at the interface (related to Figure 2c). The fold-change in expression of a given gene *X* in the interface relative to the tumor or the muscle was computed by comparing the expression of gene *i* in the interface versus the maximum expression in either the tumor or the muscle (whichever is higher). Similarly, when calculating the *p*-value for specificity for each gene being expressed in the interface, we used a Wilcoxon's test to compare the expression of gene *X* in the interface versus the expression in the tumor or the muscle (again, in whichever SRT cluster that gene is more highly expressed in). Results were plotted as a volcano plot using the EnhancedVolcano package in R, where genes whose fold-change *p*-value from a Wilcoxon's test is < 0.05, and whose log2(fold-change) in expression is greater than 1 is highlighted in red.

Non-negative matrix factorization (NMF)

After normalization and integration of SRT data (see *Dimensionality reduction and clustering*), negative values in the integrated expression matrix were set to zero. NMF was performed with a rank of 11. The optimal number of ranks was estimated using the function nmfEstimateRank[58] based on the first rank for which cophenetic starts decreasing[59] and for which RSS presents an inflection point[60]. Factor scores were first z-scored across factors prior to plotting onto array spots.

Analysis of gene ontology (GO) terms with spatially coherent expression patterns

GO term annotations for *Danio rerio* were downloaded from Biomart[61]. For each GO term, the average expression of genes annotated for that GO term was computed. We defined spots that highly express this GO term as spots whose expression level for these genes is above the mean plus two standard deviations (we required the number of these spots to be at least 5 to proceed with the analysis). We then computed the Euclidean distance between these spots.

Next, we computed the Euclidean distance between the same number of random spots, and repeated this computation 100 times to generate a null distribution of distances. We then compared the GO term spot distances to the null distribution using Wilcoxon's rank sum test to compute a *p*-value.

Correlation between SRT spots and SRT clusters

For computing the correlation across SRT clusters, we first computed the average expression of each tissue cluster in the integrated expression matrix of our three datasets. We then used the union of the ~1000 variably expressed genes in each individual dataset to obtain a list of ~2300 total variably expressed genes. We then used these genes to compute the Pearson's correlation and associated *p*-values.

GSEA and pathway analysis

Lists of differentially expressed genes for pathway analysis were created using the Seurat function FindMarkers using the Wilcoxon rank sum test. Ribosomal genes and genes with *p*-values above 0.05 were removed. Zebrafish genes were converted to their human orthologs using DIOPT[62], keeping only human orthologs with a DIOPT score > 6. In cases where there were multiple zebrafish orthologs for one human gene, the gene with the highest log fold change in expression was used. Pathway analysis and GSEA[63] was done using the fgsea R package[64], using the MSigDB[65] Gene Ontology (GO)[66,67] biological processes and GO cellular component human genesets.

HOMER motif analysis

Motif analysis was performed using HOMER[31], using the function findMotifs.pl. Motifs of lengths 8, 10, and 16 were queried within +/- 500 bp of the TSS of differentially expressed genes. Target genes containing the motif of interest were found by filtering the list of differentially expressed genes to contain only those with the desired motif. JASPAR[68] was used to annotate motifs.

Multimodal intersection analysis (MIA)

To determine cell type enrichment in tissue regions we used multimodal intersection analysis (MIA)[14], which uses the hypergeometric cumulative distribution to determine the statistical significance of the overlap between cell type specific gene sets and tissue region specific gene sets. We used the intersect between all genes in the SRT count matrix and all genes in the snRNA-seq count matrix as the gene background to calculate the *p*-value. In parallel, we tested for cell type depletion by computing $-\log10(1 - p)$.

Single-cell RNA-seq

Sample preparation

Adult tumor-bearing fish were dissected to obtain only the tumor and surrounding tissues (i.e head and tail were removed). Tissue was minced with a fresh scalpel and incubated in 0.16 mg/mL liberase (Sigma-Aldrich #5401020001) in 0.9X PBS for 15 minutes at RT. Tissue was then further dissociated by repeated pipetting with a wide-bore P1000, followed by incubation for an additional 15 minutes at RT. After adding 500 uL FBS to stop the dissociation reaction,

samples were filtered through a 70 µm filter and centrifuged at 500*g* for 5 minutes at RT. The resulting pellet was resuspended in DMEM supplemented with 2% FBS, and cells were sorted at room temperature to remove debris and doublets, using a BD FACSAria III cell sorter (BD Biosciences). Equal numbers of GFP+ (tumor) and GFP- (microenvironment) cells were collected.

Cell encapsulation and library preparation

Equal numbers of sorted GFP+ (tumor) and GFP- (microenvironment) cells were centrifuged at 300*g* for 5 minutes at RT, and resuspended in DMEM + 10% FBS. Droplet-based scRNA-seq was performed using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (10X Genomics) and Chromium Single Cell 3' Chip G (10X Genomics). Approximately 10,000 cells from two fish were split encapsulated in a single v3 reaction. GEM generation and library preparation were performed according to manufacturer's instructions.

Sequencing

10X scRNA-Seq libraries were pooled, denatured, and diluted to a concentration of 1.8 pM with 1% PhiX prior to paired-end sequencing on a NextSeq 500. Each library (corresponding to approximately 5000 cells) were sequenced to a depth of 550M paired-end reads. Sequencing parameters: Read1 28 cycles, index read 8 cycles, Read2 132 cycles. Sequencing data was aligned to our reference zebrafish genome using Cell Ranger v5.0.1 (10X Genomics).

Analysis

Data was processed using R version 3.6.3 and Seurat version 3.1.4[15]. Cells with fewer than 200 unique genes or >20% mitochondrial reads were filtered out. Expression data was normalized using SCTransform[54]. Datasets were integrated using the Seurat SCTransform integration workflow, with 3000 integration anchors and including all genes expressed in both datasets (15,154 genes). Principal component analysis[55], UMAP dimensionality reduction[56], HOMER analysis[31], GSEA, and pathway analysis were performed as described above. Cluster annotations were performed using the Seurat function FindAllMarkers, in conjunction with marker genes used in previous analyses[69]. Doublets were detected using the doubletFinder R package[70], using 15 principal components.

Single-nucleus RNA-seq

Sample preparation

Adult tumor-bearing fish were dissected to obtain only the tumor and surrounding tissues (i.e head and tail were removed). The tissue was then ground in a Dounce homogenizer on ice in Nuclei EZ Prep Lysis Buffer (Sigma-Aldrich #NUC101). The nuclear suspension was then spun down at 4°C (500g, 5 min). After resuspending the pellet in 1 mL wash buffer (250 mM sucrose, 50 mM citric acid, 1% BSA, 20 mM DTT, 0.2U/µL RNAse inhibitor), the sample was again spun at 4°C for 5 min at 500g. The pellet was resuspended in 1 mL wash buffer and subsequently sorted at 4°C to isolate individual nuclei, using a BD FACSAria III cell sorter (BD Biosciences). Approximately equal numbers of GFP+ (tumor) and GFP- (microenvironment) nuclei were collected.

Cell encapsulation and library preparation

Equal numbers of sorted GFP+ (tumor) and GFP- (microenvironment) nuclei were centrifuged at 600*g* for 5 minutes at 4°C. Droplet-based snRNA-seq was performed using the Chromium Single Cell 3' Library & Gel Bead Kit v3 (10X Genomics) and Chromium Single Cell 3' Chip G (10X Genomics). Approximately 12,000 nuclei were encapsulated in a single v3 reaction. GEM generation and library preparation were performed according to manufacturer's instructions.

Sequencing

10X snRNA-Seq libraries were pooled, denatured, and diluted to a concentration of 1.8 pM with 1% PhiX prior to paired-end sequencing on a NovaSeq 6000. Sequencing parameters: Read1 26 cycles, Read2 70 cycles, index read 8 cycles. Sequencing depth was approximately 200 million reads per 10,000 nuclei. Sequencing data was aligned to our reference zebrafish genome using CellRanger (10X Genomics).

Analysis

Data was processed using R version 3.6.3 and Seurat version 3.1.4[15]. Nuclei with fewer than 200 unique genes, more than 1 million UMIs, predicted doublets[70] and/or >20% mitochondrial reads were filtered out. Expression data was normalized using SCTransform[54]. PCA[55], UMAP[56], and HOMER analysis[31] were performed as described above. Potential doublets were detected with doubletFinder[70] and were filtered out before downstream analyses. Cluster annotations were performed using the Seurat function FindAllMarkers, in conjunction with marker genes used in previous analyses[69]. Modelling of ligand-receptor interactions was performed using NicheNet and the nichenetR R package[20], with the combined interface cluster as the "sender" cell population and all other cells as "receiver", using a cutoff of 0.1 for determining expressed genes and 0.5 for ligand-target scores. For NicheNet analysis, Zebrafish genes were converted to human as described above, using DIOPT[62], keeping only human orthologs with a DIOPT score > 6. In cases where there were multiple zebrafish orthologs for one human gene, the gene with the highest log fold change in expression was used.

Immunofluorescence and imaging

Adult *casper* zebrafish with large pigmented tumors were euthanized on ice and fixed in 4% paraformaldehyde in PBS for 72 hours at 4°C. Fish were then stored in 70% ethanol before embedding in paraffin and sectioning by Histowiz, Inc. FFPE slides were deparaffinized in xylene before several rounds of washing in 100-50% ethanol. Antigen retrieval was performed by heating slides to 95°C for 20 minutes in 10 mM sodium citrate pH 6.2 in a pressure cooker. After cooling, slides were blocked in a solution of 5% donkey serum, 1% BSA, and 0.4% Triton-X100 in PBS for 1 hour at room temperature, before overnight incubation with primary antibodies in blocking buffer. Primary antibodies used were: goat anti-GFP (abcam #ab5450, 1:200) and mouse anti-acetylated tubulin (Sigma-Aldrich #6793, 1:100). Following overnight incubation with primary antibodies, slides were washed in PBS before incubation with secondary antibodies for 2 hours at room temperature. Secondary antibodies used were: donkey anti-goat IgG conjugated to Alexa 488 (Thermo Fisher Scientific #A11055, 1:250) and

goat anti-mouse IgG conjugated to Alexa 555 (Cell Signaling Technology #4409S, 1:250). Hoechst 33342 (Thermo Fisher Scientific #H3570) was added to the secondary antibody solution at 1:1000. Slides were mounted in ProLong Glass (Thermo Fisher Scientific #P36980) and cured overnight at room temperature. Slides were imaged on a Leica SP5 upright line-scanning confocal microscope using a 40X (oil) objective. 12-bit Z-stacks were acquired at 0.3-0.5 μm steps using 3X line averaging. Maximum intensity projections were created of the Z-stacks in ImageJ. Noise was removed from each image using the ImageJ "Despeckle" function.

Re-analysis of Tirosh et al. human melanoma scRNA-seq dataset

The counts matrix was obtained from GEO (GSE72056). All analysis was done using R version 3.6.3 and Seurat version 3.1.4[19]. A Seurat object was created with default parameters, keeping all genes expressed in 3 or more cells and all cells expressing 200 or more genes, resulting in a final object containing 4,645 cells. Dimensionality reduction and clustering were performed using 15 principal components. Expression scores for interface, cilia, and ETS genes were calculated using the Seurat function AddModuleScore with default parameters. Interface marker genes were defined as the human orthologs of all genes with a log fold change > 1.5 in our zebrafish scRNA-seq interface cluster.

Statistical analysis

Statistical analysis and figure generation were performed in MATLAB (Mathworks, R2019a) and R (R Foundation for Statistical Computing, 3.6.3). Image processing and analysis was performed in MATLAB and ImageJ (NIH). Unless otherwise noted, *p*-values were calculated using the Wilcoxon rank-sum test with Bonferroni's correction for multiple groups (R functions wilcox.test and pairwise.wilcox.test). Pearson correlation coefficients and corresponding *p*-values were calculated using the R function cor.test. *p*-value abbreviations are as follows: *, $p < 0.05$, **, $p < 0.01$, ***, $p < 0.001$, n.s, not significant.

Data availability

The scRNA-seq, snRNA-seq and SRT data reported in this manuscript have been deposited to the Gene Expression Omnibus under accession number GSE159709. All code used for analysis and plotting is available at github.com/mvhunter1/Hunter_Moncada_2021.

SUPPLEMENTAL INFORMATION TITLES AND LEGENDS

Figure S1. Spatially resolved transcriptomics data statistics.

a-b. Violin plots showing number of transcripts (UMIs) detected per Visium array spot (a) and number of unique genes detected per spot (b). **c.** Images of the H&E stained tissue sections used for SRT. **d-e.** Numbers of UMIs and genes detected per spot, plotted onto the tissue array. **f.** Violin plot showing the number of UMIs per spot across the tissue types in the individual SRT samples. **g.** Approximation of the number of nuclei per Visium array spot (55 μm diameter) in the tumor and muscle regions.

Figure S2. A distinct interface cluster is found in each sample.

a. Cluster assignments plotted onto the Visium array for individual samples. **b.** Cluster assignments labelled in UMAP space for individual samples.

Figure S3. Spatial patterning of biological pathways in the tumor and microenvironment. **a-b.** Average, standardized expression of annotated genes for gene ontology (GO) terms displaying spatially-coherent expression patterns in the tumor (a) and microenvironment (b) regions in each SRT sample. P-values represent the comparison between the distance between spots expressing that GO term genes and a null-distribution of distances between random spots (Wilcoxon's Rank Sum test, see Methods).

Figure S4. Non-negative matrix factorization (NMF) analysis on SRT spots, and enriched GO terms for top genes in each NMF factor.
Non-negative matrix factorization (NMF) was performed on the integrated expression matrix of the three SRT samples, with k = 11 factors. Factor scores were then projected onto spots, and the top 150 scoring genes per factor were used for GO enrichment analysis.

Figure S5. single-cell RNA-seq data statistics. a,d. Histograms showing the number of UMIs (left) and genes (right) per cell for each scRNA-seq reaction. **b,e.** Violin plots showing the number of UMIs (left) and genes (right) per cell for the different clusters in each scRNA-seq reaction. **c,f.** Cluster assignments for each scRNA-seq reaction plotted in UMAP space. **g.** Predicted possible doublets within the integrated scRNA-seq dataset.

Figure S6. single-nucleus RNA-seq data statistics. a. Histograms showing the number of UMIs (left) and genes (right) per nucleus. **b.** Violin plots showing the number of UMIs (left) and genes (right) per nucleus across the clusters in the snRNA-seq dataset. **c-d.** Mean expression per cell/nucleus for all common genes in the scRNA-seq and snRNA-seq dataset (15,022 genes) (c) and for the fish orthologs of the published SYSCILIA genes (320 genes).

Figure S7. Deconvolution of the SRT interface region using multimodal intersection analysis (MIA). a-b. MIA maps for all snRNA-seq clusters compared to all SRT clusters (a) or only the SRT interface cluster (b). *P*-values were calculated using the hypergeometric test.

Figure S8. Computational modeling of interface cell interactions with NicheNet. a. Predicted ligands expressed by interface cells, ranked by predicted ligand activity (Pearson correlation coefficient). **b.** Predicted receptors expressed by non-interface cells for the predicted ligands in **a. c.** Predicted target genes for the ligands in **a**. **d-e.** Normalized expression of the zebrafish orthologs of *HMGB2* (*hmgb2a* and *hmgb2b*) and *CDH1* (*cdh1*) in the snRNA-seq, scRNA-seq and SRT clusters.

Figure S9. Acetylated tubulin staining within the tumor, muscle, and tumor-muscle interface regions. Immunofluorescence images of tumor (green), Hoescht (blue) and acetylated tubulin (magenta) intensity at the tumor-muscle interface (top), center of tumor (middle) and distant muscle (bottom). Scale bars, 100 μm (40X magnification).

Figure S10. An interface-like cell state is found in human melanoma. a-c. UMAP projection of human melanoma scRNA-seq data from Tirosh et al., 2016. Cell types (original annotations, a), interface marker gene expression score (b), and interface classification (c) are indicated. The cutoff for classification as an interface-like cell state is indicated in (b). **d-e.** Expression of human cilia (d) and ETS (e) genes across the cell types/states in human melanoma scRNA-seq. The expression score represents the standardized mean expression of all indicated genes per cell.

Table S1. Activator/repressor status for the human ETS genes. Activator/repressor status and fish orthologs were obtained from Uniprot and ZFIN.

Video S1. Spatial patterning of biological pathways in melanoma. 100 pathways are shown within the tumor region of sample A. The average expression of the genes annotated for the indicated pathway is shown. Blue: low expression; red: high expression.

Video S2. Spatial patterning of biological pathways in the melanoma macroenvironment. 100 pathways are shown within the interface and muscle regions of sample C. The average expression of the genes annotated for the indicated pathway is shown. Blue: low expression; red: high expression.

REFERENCES

1. Kaur, A. *et al.* sFRP2 in the aged microenvironment drives melanoma metastasis and therapy resistance. *Nature* **532**, 250–254 (2016).
2. Passarelli, A., Mannavola, F., Stucci, L. S., Tucci, M. & Silvestris, F. Immune system and melanoma biology: a balance between immunosurveillance and immune escape. *Oncotarget* **8**, 106132–106142 (2017).
3. Zhang, M. *et al.* Adipocyte-Derived Lipids Mediate Melanoma Progression via FATP Proteins. *Cancer Discov.* **8**, 1006–1025 (2018).
4. Kim, I. S. *et al.* Microenvironment-derived factors driving metastatic plasticity in melanoma. *Nat. Commun.* **8**, 14343 (2017).
5. Quail, D. F. & Joyce, J. A. Microenvironmental regulation of tumor progression and metastasis. *Nat. Med.* **19**, 1423–1437 (2013).
6. Ungefroren, H., Sebens, S., Seidl, D., Lehnert, H. & Hass, R. Interaction of tumor cells with

the microenvironment. *Cell Commun. Signal.* **9**, 18 (2011).

7. Meacham, C. E. & Morrison, S. J. Tumour heterogeneity and cancer cell plasticity. *Nature* **501**, 328–337 (2013).
8. Naser, F. J. *et al.* Isotope tracing in adult zebrafish reveals alanine cycling between melanoma and liver. *Cell Metab.* (2021) doi:10.1016/j.cmet.2021.04.014.
9. Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).
10. Vickovic, S. *et al.* High-definition spatial transcriptomics for in situ tissue profiling. *Nat. Methods* **16**, 987–990 (2019).
11. Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).
12. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S. & Zhuang, X. RNA imaging. Spatially resolved, highly multiplexed RNA profiling in single cells. *Science* **348**, aaa6090 (2015).
13. Moncada, R. *et al.* Integrating microarray-based spatial transcriptomics and single-cell RNA-seq reveals tissue architecture in pancreatic ductal adenocarcinomas. *Nat. Biotechnol.* **38**, 333–342 (2020).
14. Patton, E. E. *et al.* BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Curr. Biol.* **15**, 249–254 (2005).
15. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888-1902.e21 (2019).
16. Gast, C. E. *et al.* Cell fusion potentiates tumor heterogeneity and reveals circulating hybrid cells that correlate with stage and survival. *Sci. Adv.* **4**, eaat7828 (2018).
17. Habib, N. *et al.* Massively parallel single-nucleus RNA-seq with DroNc-seq. *Nat. Methods* **14**, 955–958 (2017).

18. Slyper, M. *et al.* A single-cell and single-nucleus RNA-Seq toolbox for fresh and frozen human tumors. *Nat. Med.* **26**, 792–802 (2020).
19. Butler, A., Hoffman, P., Smibert, P., Papalexi, E. & Satija, R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. *Nat. Biotechnol.* **36**, 411–420 (2018).
20. Browaeys, R., Saelens, W. & Saeys, Y. NicheNet: modeling intercellular communication by linking ligands to target genes. *Nat. Methods* **17**, 159–162 (2020).
21. Kwon, J.-H. *et al.* Overexpression of high-mobility group box 2 is associated with tumor aggressiveness and prognosis of hepatocellular carcinoma. *Clin. Cancer Res.* **16**, 5511–5521 (2010).
22. Fu, D. *et al.* HMGB2 is associated with malignancy and regulates Warburg effect by targeting LDHB and FBP1 in breast cancer. *Cell Commun. Signal.* **16**, 8 (2018).
23. Harris, T. J. C. & Tepass, U. Adherens junctions: from molecules to morphogenesis. *Nat. Rev. Mol. Cell Biol.* **11**, 502–514 (2010).
24. Mo, Y. *et al.* MicroRNA-329 upregulation impairs the HMGB2/β-catenin pathway and regulates cell biological behaviors in melanoma. *J. Cell. Physiol.* **234**, 23518–23527 (2019).
25. Choudhury, A., Neumann, N. M., Raleigh, D. R. & Lang, U. E. Clinical implications of primary cilia in skin cancer. *Dermatol Ther (Heidelb)* **10**, 233–248 (2020).
26. Zingg, D. *et al.* EZH2-Mediated Primary Cilium Deconstruction Drives Metastatic Melanoma Formation. *Cancer Cell* **34**, 69-84.e14 (2018).
27. Kim, J., Dabiri, S. & Seeley, E. S. Primary cilium depletion typifies cutaneous melanoma in situ and malignant melanoma. *PLoS ONE* **6**, e27410 (2011).
28. Snedecor, E. R. *et al.* Loss of primary cilia in melanoma cells is likely independent of proliferation and cell cycle progression. *J. Invest. Dermatol.* **135**, 1456–1458 (2015).

29. Lang, U. E., Love, N. R., Cheung, C., McCalmont, T. H. & Kim, J. Use of the ciliation index to distinguish invasive melanoma from associated conventional melanocytic nevi. *Am. J. Dermatopathol.* **42**, 11–15 (2020).

30. van Dam, T. J. *et al.* The SYSCILIA gold standard (SCGSv1) of known ciliary components and its applications within a systems biology consortium. *Cilia* **2**, 7 (2013).

31. 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 (2010).

32. Piasecki, B. P., Burghoorn, J. & Swoboda, P. Regulatory Factor X (RFX)-mediated transcriptional rewiring of ciliary genes in animals. *Proc Natl Acad Sci USA* **107**, 12969–12974 (2010).

33. Rothhammer, T. *et al.* The Ets-1 transcription factor is involved in the development and invasion of malignant melanoma. *Cell. Mol. Life Sci.* **61**, 118–128 (2004).

34. Wouters, J. *et al.* Robust gene expression programs underlie recurrent cell states and phenotype switching in melanoma. *Nat. Cell Biol.* **22**, 986–998 (2020).

35. Sizemore, G. M., Pitarresi, J. R., Balakrishnan, S. & Ostrowski, M. C. The ETS family of oncogenic transcription factors in solid tumours. *Nat. Rev. Cancer* **17**, 337–351 (2017).

36. Tirosh, I. *et al.* Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science* **352**, 189–196 (2016).

37. Anvarian, Z., Mykytyn, K., Mukhopadhyay, S., Pedersen, L. B. & Christensen, S. T. Cellular signalling by primary cilia in development, organ function and disease. *Nat. Rev. Nephrol.* **15**, 199–219 (2019).

38. Hassounah, N. B., Bunch, T. A. & McDermott, K. M. Molecular pathways: the role of primary cilia in cancer progression and therapeutics with a focus on Hedgehog signaling. *Clin.*

Cancer Res. **18**, 2429–2435 (2012).

39. Albrecht-Buehler, G. Phagokinetic tracks of 3T3 cells: parallels between the orientation of track segments and of cellular structures which contain actin or tubulin. *Cell* **12**, 333–339 (1977).
40. Christensen, S. T., Pedersen, S. F., Satir, P., Veland, I. R. & Schneider, L. The primary cilium coordinates signaling pathways in cell cycle control and migration during development and tissue repair. *Curr. Top. Dev. Biol.* **85**, 261–301 (2008).
41. Schneider, L. *et al.* Directional cell migration and chemotaxis in wound healing response to PDGF-AA are coordinated by the primary cilium in fibroblasts. *Cell. Physiol. Biochem.* **25**, 279–292 (2010).
42. Mavrothalassitis, G. & Ghysdael, J. Proteins of the ETS family with transcriptional repressor activity. *Oncogene* **19**, 6524–6532 (2000).
43. Mao, P. *et al.* ETS transcription factors induce a unique UV damage signature that drives recurrent mutagenesis in melanoma. *Nat. Commun.* **9**, 2626 (2018).
44. Li, R., Pei, H. & Watson, D. K. Regulation of Ets function by protein–protein interactions. *Oncogene* **19**, 6514–6523 (2000).
45. Wasylyk, B., Hagman, J. & Gutierrez-Hartmann, A. Ets transcription factors: nuclear effectors of the Ras-MAP-kinase signaling pathway. *Trends Biochem. Sci.* **23**, 213–216 (1998).
46. Fecher, L. A., Amaravadi, R. K. & Flaherty, K. T. The MAPK pathway in melanoma. *Curr. Opin. Oncol.* **20**, 183–189 (2008).
47. González-Silva, L., Quevedo, L. & Varela, I. Tumor Functional Heterogeneity Unraveled by scRNA-seq Technologies. *Trends Cancer* **6**, 13–19 (2020).
48. Dagogo-Jack, I. & Shaw, A. T. Tumour heterogeneity and resistance to cancer therapies.

Nat. Rev. Clin. Oncol. **15**, 81–94 (2018).

49. Kather, J. N. *et al.* Pan-cancer image-based detection of clinically actionable genetic alterations. *Nat. Cancer* (2020) doi:10.1038/s43018-020-0087-6.
50. Fu, Y. *et al.* Pan-cancer computational histopathology reveals mutations, tumor composition and prognosis. *Nat. Cancer* (2020) doi:10.1038/s43018-020-0085-8.
51. White, R. M. *et al.* Transparent adult zebrafish as a tool for in vivo transplantation analysis. *Cell Stem Cell* **2**, 183–189 (2008).
52. Iyengar, S., Houvras, Y. & Ceol, C. J. Screening for melanoma modifiers using a zebrafish autochthonous tumor model. *J. Vis. Exp.* e50086 (2012) doi:10.3791/50086.
53. Ceol, C. J. *et al.* The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* **471**, 513–517 (2011).
54. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
55. Jolliffe, I. T. Principal component analysis and factor analysis. in *Principal Component Analysis* 115–128 (Springer New York, 1986). doi:10.1007/978-1-4757-1904-8_7.
56. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *JOSS* **3**, 861 (2018).
57. Dittmer, J. Ets Transcription Factors. in *Encyclopedia of Cancer* (ed. Schwab, M.) 1339–1343 (Springer Berlin Heidelberg, 2011). doi:10.1007/978-3-642-16483-5_2034.
58. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).
59. Hutchins, L. N., Murphy, S. M., Singh, P. & Graber, J. H. Position-dependent motif characterization using non-negative matrix factorization. *Bioinformatics* **24**, 2684–2690 (2008).

60. Frigyesi, A. & Höglund, M. Non-negative matrix factorization for the analysis of complex gene expression data: identification of clinically relevant tumor subtypes. *Cancer Inform.* **6**, 275–292 (2008).

61. Durinck, S., Spellman, P. T., Birney, E. & Huber, W. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat. Protoc.* **4**, 1184–1191 (2009).

62. Hu, Y. *et al.* An integrative approach to ortholog prediction for disease-focused and other functional studies. *BMC Bioinformatics* **12**, 357 (2011).

63. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**, 15545–15550 (2005).

64. Sergushichev, A. An algorithm for fast preranked gene set enrichment analysis using cumulative statistic calculation. *BioRxiv* (2016) doi:10.1101/060012.

65. Liberzon, A. *et al.* Molecular signatures database (MSigDB) 3.0. *Bioinformatics* **27**, 1739–1740 (2011).

66. The Gene Ontology Consortium. The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Res.* **47**, D330–D338 (2019).

67. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

68. 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**, D1284 (2018).

69. Baron, M. *et al.* The Stress-Like Cancer Cell State Is a Consistent Component of Tumorigenesis. *Cell Syst.* **11**, 536-546.e7 (2020).

70. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in

Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329-337.e4 (2019).

Figure 1. Spatially resolved transcriptomics reveals the transcriptional architecture of melanoma and its surrounding microenvironment.

a. Schematic showing the spatially resolved transcriptomics (SRT) experiment workflow. **b.** Images of zebrafish with *BRAF*^{V600E}-driven melanomas used for SRT. The region where the fish were sectioned is highlighted. **c.** H&E staining of cryosections used for SRT. **d.** Visium array spots colored by clustering assignments of the integrated dataset (see Methods). **e.** UMAP embedding of SRT spots from all three samples colored by cluster assignments of the integrated dataset (see Methods). **f.** The expression of select marker genes (*BRAF*^{V600E}, tumor; *pvalb4*, muscle; *kcn6a*, heart; *mbpa*, nervous system) from the SRT data projected over tissue space (left), with images of the corresponding histology from the indicated region of the SRT array (right). **g-h.** Average, standardized expression of annotated genes for gene ontology (GO) terms displaying spatially-coherent expression patterns in the tumor (g) and microenvironment (h) regions in each sample. *p*-values represent the comparison between the distance between spots expressing that GO term genes and a null-distribution of distances between random spots (Wilcoxon's Rank Sum test, see Methods).

Figure 1

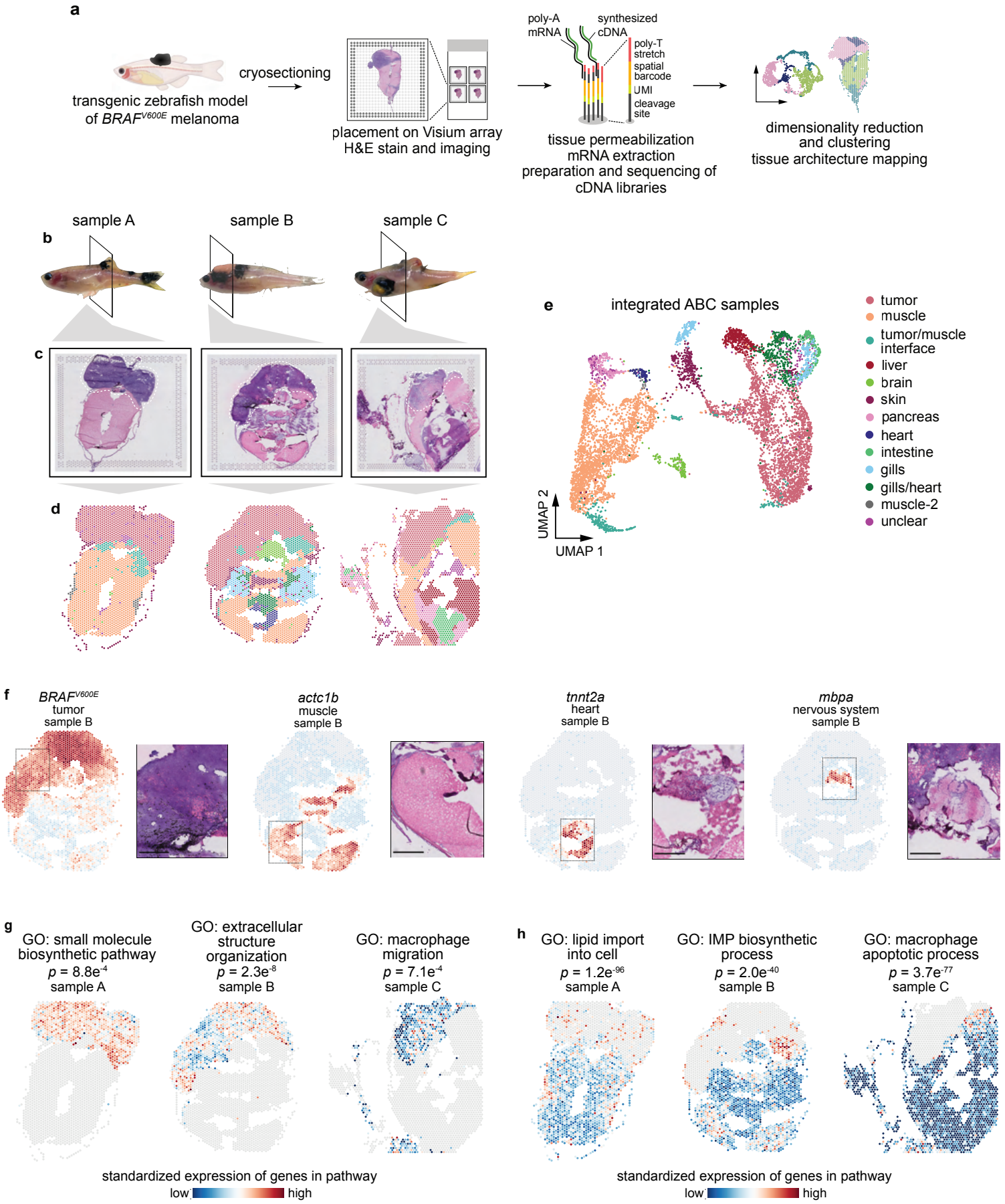

a
transgenic zebrafish model of BRAF V600E melanoma
cryosectioning
placement on Visium array
H&E stain and imaging
poly-A mRNA
synthesized cDNA
poly-T stretch
spatial barcode
UMI
cleavage site
tissue permeabilization
mRNA extraction
preparation and sequencing of cDNA libraries
dimensionality reduction
and clustering
tissue architecture mapping
sample A
sample B
sample C
b
c
d
e
integrated ABC samples
UMAP 2
UMAP 1
tumor
muscle
tumor/muscle interface
liver
brain
skin
pancreas
heart
intestine
gills
gills/heart
muscle-2
unclear
f
BRAF V600E tumor sample B
actc1b muscle sample B
tnnt2a heart sample B
mbpa nervous system sample B
g
GO: small molecule biosynthetic pathway
p = 8.8e-4
sample A
GO: extracellular structure organization
p = 2.3e-8
sample B
GO: macrophage migration
p = 7.1e-4
sample C
h
GO: lipid import into cell
p = 1.2e-96
sample A
GO: IMP biosynthetic process
p = 2.0e-40
sample B
GO: macrophage apoptotic process
p = 3.7e-77
sample C
standardized expression of genes in pathway
low
high
standardized expression of genes in pathway
low
high

Figure 2. The tumor-microenvironment interface is transcriptionally distinct from the surrounding microenvironment.

a. Interface and muscle-annotated cluster spots projected onto tissue image. Insets show the tissue underlying the interface spots (1) and muscle spots (2). **b.** Correlation matrix between average expression profile of SRT clusters across all three datasets. Clusters are ordered by hierarchical clustering of the Pearson's correlation coefficients (see Methods) and bubble sizes correspond to *p*-value (-log10) of correlation, with *p*-values < 10^{-3} omitted. Clustering of tumor and interface together is highlighted in the dendrogram (red). **c.** Volcano plot of differentially expressed genes between the interface cluster versus the muscle and tumor clusters. *p*-values were obtained from the Wilcoxon's rank sum test. **d.** Non-negative matrix factorization (NMF) of the microenvironment spots (muscle and interface clusters). Shown are the standardized factor scores for interface-specific NMF factor 7, projected onto microenvironment spots. Arrows denote areas with higher factor scores. **e.** Enriched GO terms for the top 150 scoring genes in NMF factor 7.

Figure 2

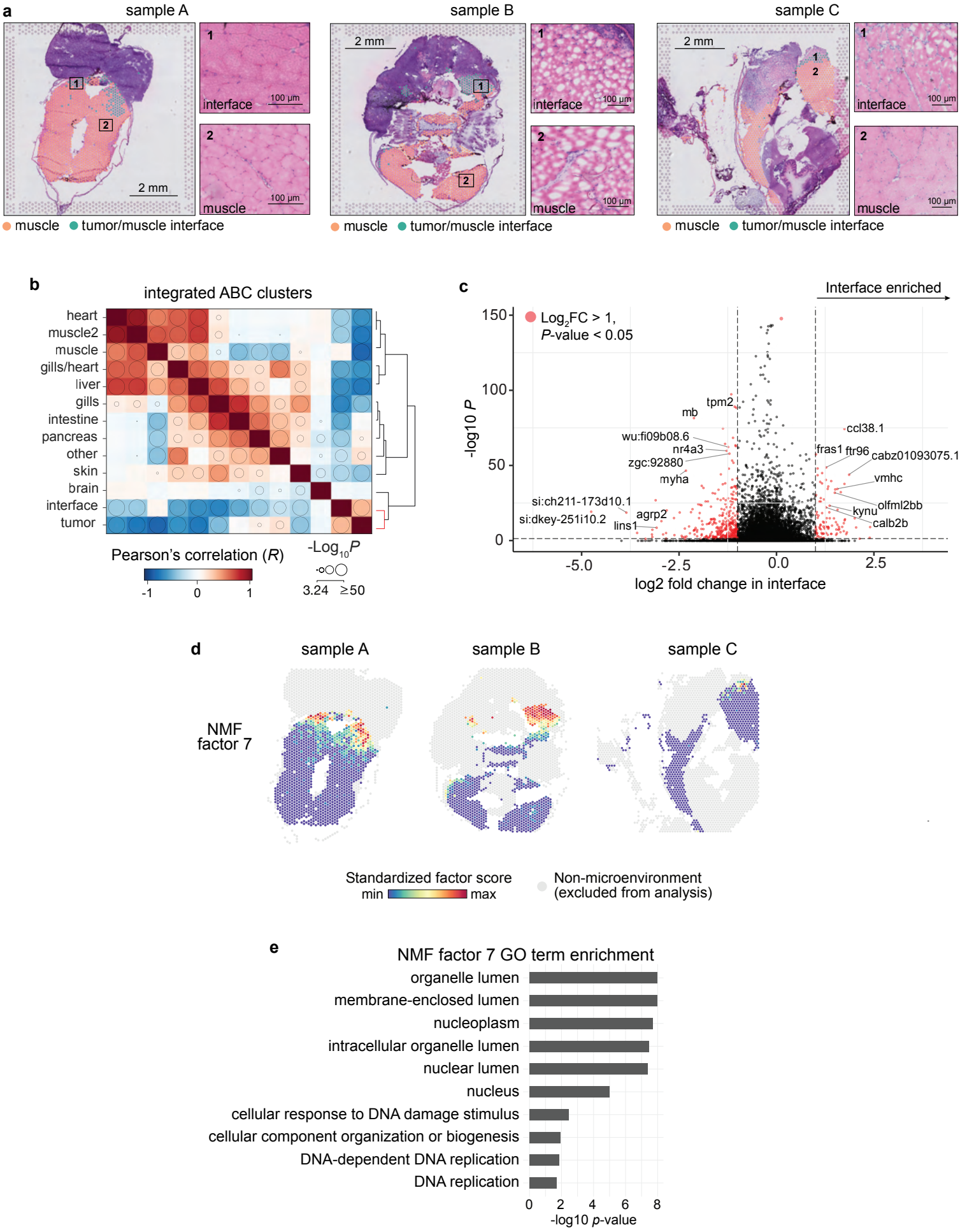

a
sample A
sample B
sample C
2 mm
interface
muscle
100 µm
muscle
tumor/muscle interface
b
integrated ABC clusters
heart
muscle2
muscle
gills/heart
liver
gills
intestine
pancreas
other
skin
brain
interface
tumor
Pearson's correlation (R)
-1
0
1
-Log10P
3.24
≥50
c
Interface enriched
Log2FC > 1,
P-value < 0.05
-log10 P
log2 fold change in interface
tpm2
mb
wu:fi09b08.6
nr4a3
zgc:92880
myha
si:ch211-173d10.1
si:dkey-251i10.2
agrp2
lins1
ccl38.1
fras1
ftr96
cabz01093075.1
vmhc
olfml2bb
kynu
calb2b
d
sample A
sample B
sample C
NMF
factor 7
Standardized factor score
min
max
Non-microenvironment
(excluded from analysis)
e
NMF factor 7 GO term enrichment
organelle lumen
membrane-enclosed lumen
nucleoplasm
intracellular organelle lumen
nuclear lumen
nucleus
cellular response to DNA damage stimulus
cellular component organization or biogenesis
DNA-dependent DNA replication
DNA replication
-log10 p-value

Figure 3. The tumor-microenvironment interface is composed of specialized tumor and muscle cells.

a. Schematic showing scRNA-seq experiment workflow. **b.** UMAP dimensionality reduction plot for 2,889 cells sequenced as in (a). Cluster/cell type assignments are labelled and colored. **c.** Expression score per cell (scRNA-seq) for average expression of interface marker genes from the SRT interface cluster. **d.** Inset of the outlined interface cluster in (b) showing the two interface subclusters. **e.** Heatmap showing expression of the top 50 genes upregulated in the tumor cell cluster (top, orange) and interface cell cluster (bottom, yellow). Selected genes are labelled. **f.** Principal component analysis of cells in the interface cluster, scored for expression of the tumor marker *BRAF*^{V600E}, the muscle marker *ckba*, and the centromere gene *stra13*. Scores for principal components 1 and 2 are plotted. Cells are labelled by standardized expression of the indicated genes. **g.** Dot plot showing expression of tumor and muscle markers. The size of each dot corresponds to the percentage of cells in that cluster expressing the indicated gene, and the colour of each dot indicates the expression level.

Figure 3

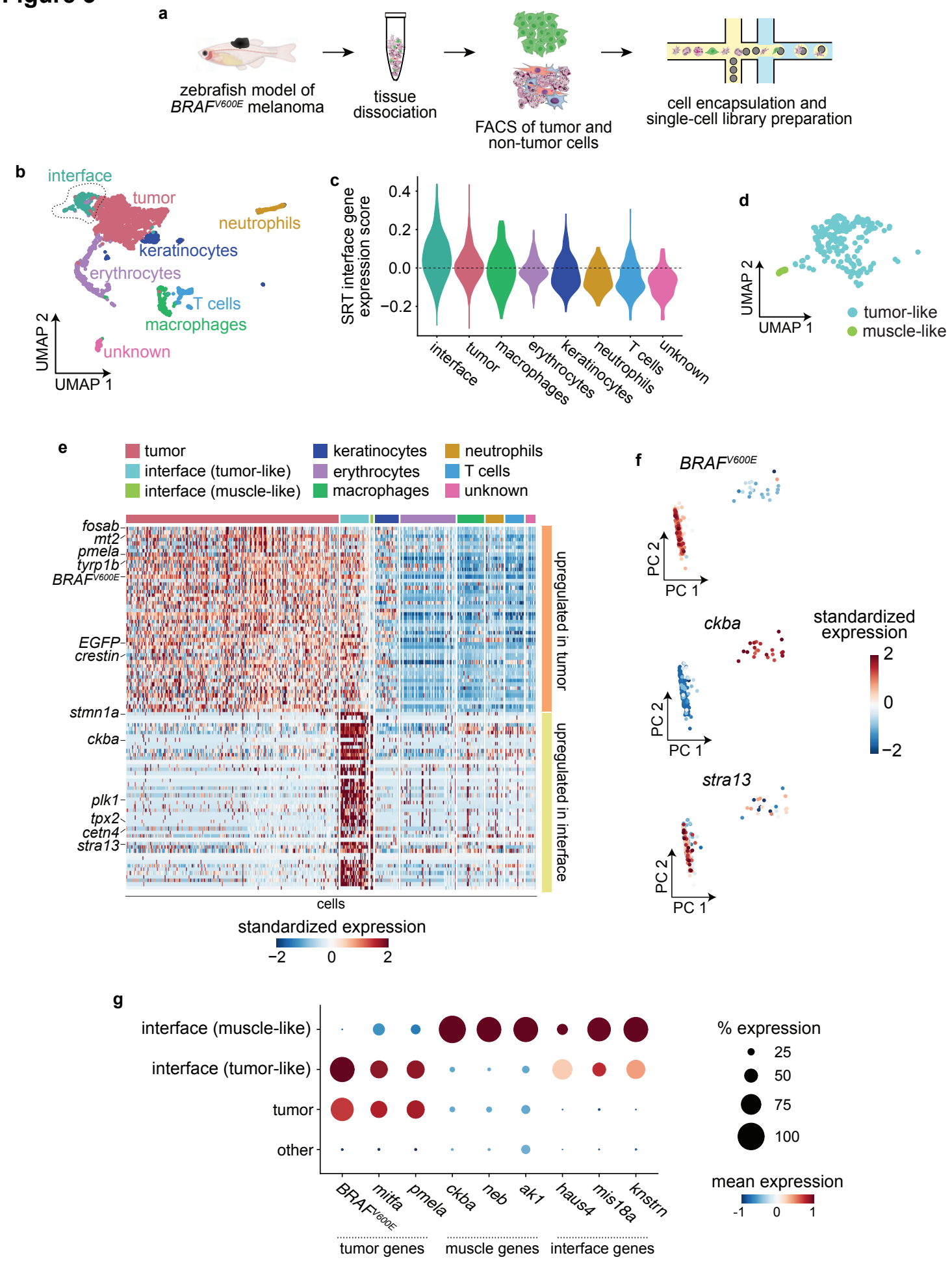

a
zebrafish model of BRAFV600E melanoma
tissue dissociation
FACS of tumor and non-tumor cells
cell encapsulation and single-cell library preparation
b
interface
tumor
neutrophils
keratinocytes
erythrocytes
T cells
macrophages
unknown
UMAP 2
UMAP 1
c
SRT interface gene expression score
0.4
0.2
0.0
−0.2
interface
tumor
macrophages
erythrocytes
keratinocytes
neutrophils
T cells
unknown
d
UMAP 2
UMAP 1
tumor-like
muscle-like
e
tumor
interface (tumor-like)
interface (muscle-like)
keratinocytes
erythrocytes
macrophages
neutrophils
T cells
unknown
fosab
mt2
pmela
tyrp1b
BRAFV600E
EGFP
crestin
stmn1a
ckba
plk1
tpx2
cetn4
stra13
upregulated in tumor
upregulated in interface
cells
standardized expression
−2
0
2
f
BRAFV600E
ckba
stra13
PC 2
PC 1
standardized expression
2
0
−2
g
interface (muscle-like)
interface (tumor-like)
tumor
other
BRAFV600E
mitfa
pmela
ckba
neb
ak1
haus4
mis18a
knstrn
tumor genes
muscle genes
interface genes
% expression
25
50
75
100
mean expression
-1
0
1

Figure 4. Single-nucleus RNA-seq demonstrates that the interface cell states are distinct from the rest of the microenvironment.

a. snRNA-seq cluster assignments plotted in UMAP space. **b.** Expression of marker genes from the scRNA-seq interface cluster in the snRNA-seq dataset. **c.** Integrated UMAP of the snRNA-seq and scRNA-seq datasets (labelled, top plot) showing colocalization of the two interface clusters (bottom plot). **d.** Subcluster assignments and expression of marker genes from the snRNA-seq interface cluster. **e.** Dotplot showing expression of microenvironment cell-type specific genes within the interface subclusters. **f.** Heatmap showing expression of the top 100 genes upregulated across all of the interface subclusters.

Figure 4

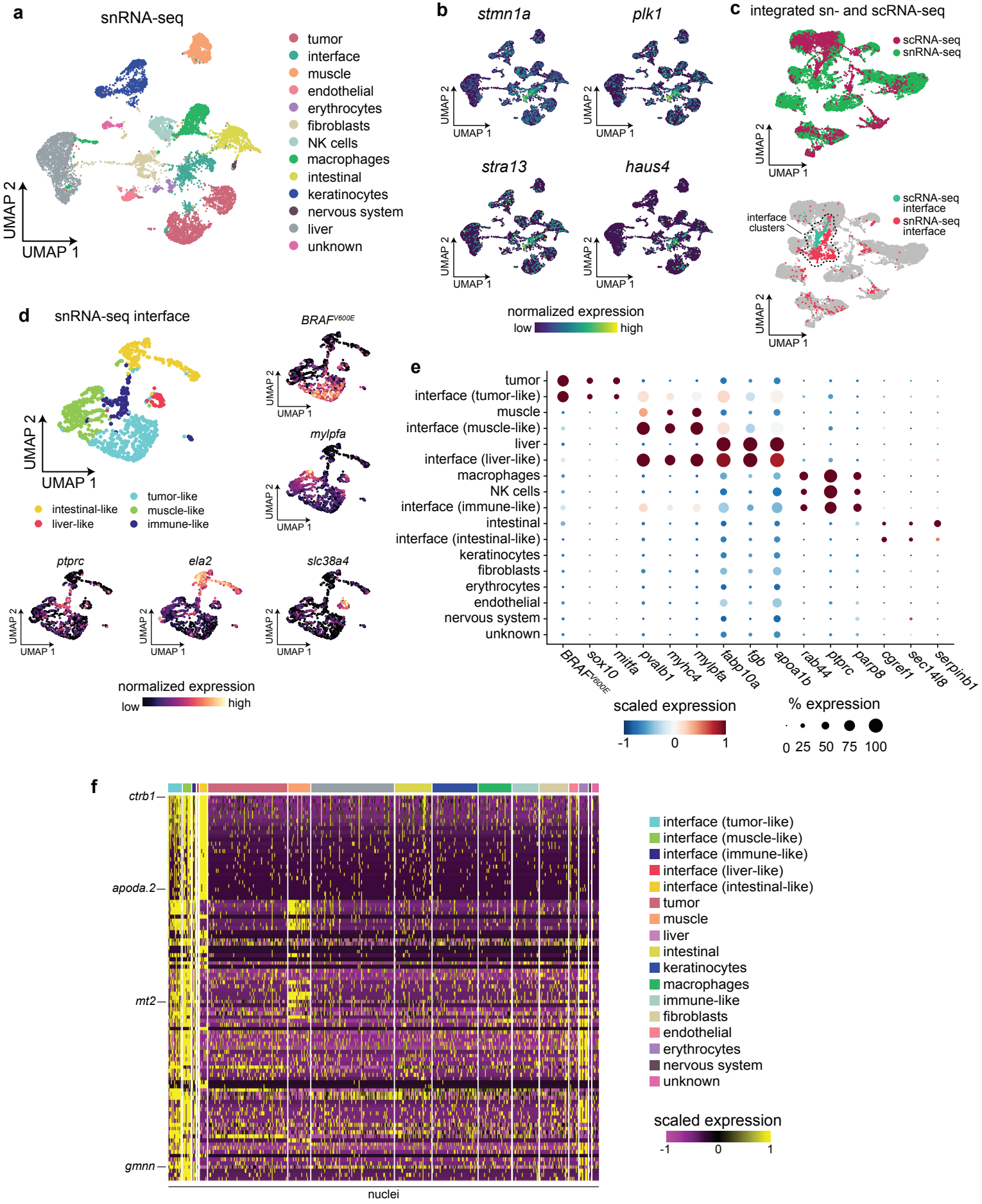

a
snRNA-seq
UMAP 2
UMAP 1
tumor
interface
muscle
endothelial
erythrocytes
fibroblasts
NK cells
macrophages
intestinal
keratinocytes
nervous system
liver
unknown
b
stmn1a
plk1
stra13
haus4
UMAP 2
UMAP 1
normalized expression
low
high
c
integrated sn- and scRNA-seq
scRNA-seq
snRNA-seq
interface clusters
scRNA-seq interface
snRNA-seq interface
d
snRNA-seq interface
intestinal-like
liver-like
tumor-like
muscle-like
immune-like
BRAFV600E
mylpfa
ptprc
ela2
slc38a4
e
tumor
interface (tumor-like)
muscle
interface (muscle-like)
liver
interface (liver-like)
macrophages
NK cells
interface (immune-like)
intestinal
interface (intestinal-like)
keratinocytes
fibroblasts
erythrocytes
endothelial
nervous system
unknown
BRAFV600E
sox10
mitfa
pvalb1
myhc4
mylpfa
fabp10a
fgb
apoa1b
rab44
ptprc
parp8
cgref1
sec14l8
serpinb1
scaled expression
-1
0
1
% expression
0 25 50 75 100
f
ctrb1
apoda.2
mt2
gmnn
nuclei
interface (tumor-like)
interface (muscle-like)
interface (immune-like)
interface (liver-like)
interface (intestinal-like)
tumor
muscle
liver
intestinal
keratinocytes
macrophages
immune-like
fibroblasts
endothelial
erythrocytes
nervous system
unknown
scaled expression
-1
0
1

Figure 5. Cilia genes and proteins are enriched at the tumor-microenvironment interface. **a.** Waterfall plot showing the top and bottom 250 GO cellular component terms by normalized enrichment score (NES) in the scRNA-seq interface cluster, with cilia GO terms labelled in red. **b.** Cilia-related GO term enrichment scores within the tumor-like and muscle-like interface cell states from the scRNA-seq dataset. **c.** Cilia-related GO term enrichment scores for the scRNA-seq and SRT interface clusters. **d.** Relative expression of fish SYSCILIA genes in the scRNA-seq interface cluster. **e.** Relative expression of the top 25 SYSCILIA genes upregulated in the scRNA-seq interface cluster across the snRNA-seq clusters. **f.** Normalized expression of selected cilia genes across the snRNA-seq interface clusters. **g.** Immunofluorescent images of sections through adult zebrafish with invasive melanomas, stained for GFP (tumor cells), acetylated tubulin (cilia) and Hoescht (nuclei), showing the tumor-muscle interface (left), center of the tumor (middle), and distant muscle (right). Arrows denote cilia at the interface. Scale bars, 100 μm. **h.** Inset of region highlighted in (g) (left). Scale bars, 25 μm. ***, $p < 0.001$ (Wilcoxon rank sum test with Bonferroni's correction).

Figure 5

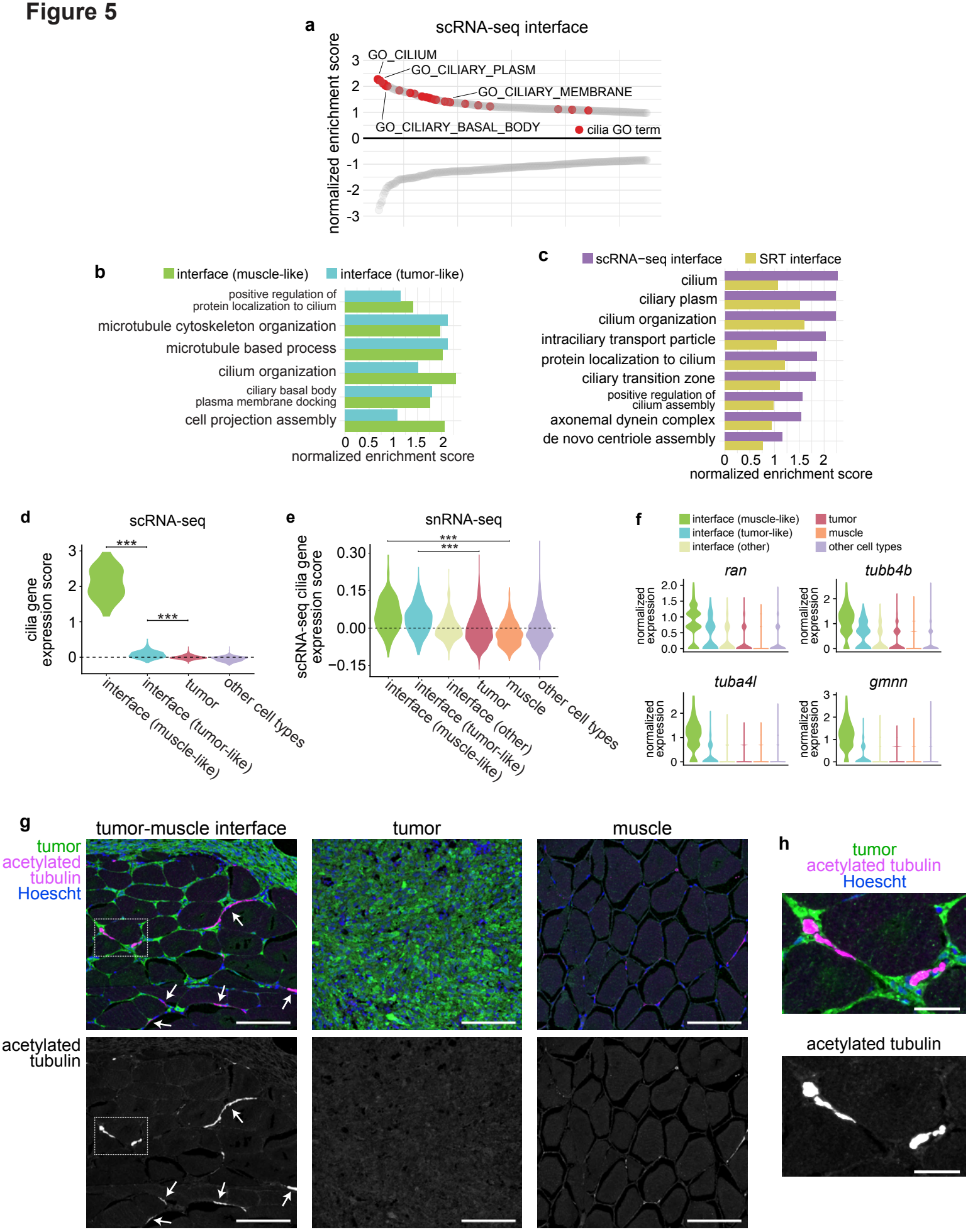

a
scRNA-seq interface
normalized enrichment score
GO_CILIUM
GO_CILIARY_PLASM
GO_CILIARY_MEMBRANE
GO_CILIARY_BASAL_BODY
cilia GO term
b
interface (muscle-like)
interface (tumor-like)
positive regulation of protein localization to cilium
microtubule cytoskeleton organization
microtubule based process
cilium organization
ciliary basal body plasma membrane docking
cell projection assembly
normalized enrichment score
c
scRNA−seq interface
SRT interface
cilium
ciliary plasm
cilium organization
intraciliary transport particle
protein localization to cilium
ciliary transition zone
positive regulation of cilium assembly
axonemal dynein complex
de novo centriole assembly
normalized enrichment score
d
scRNA-seq
cilia gene expression score
interface (muscle-like)
interface (tumor-like)
tumor
other cell types
e
snRNA-seq
scRNA-seq cilia gene expression score
interface (muscle-like)
interface (tumor-like)
interface (other)
tumor
muscle
other cell types
f
interface (muscle-like)
interface (tumor-like)
interface (other)
tumor
muscle
other cell types
ran
tubb4b
tuba4l
gmnn
normalized expression
g
tumor-muscle interface
tumor
muscle
tumor
acetylated tubulin
Hoescht
acetylated tubulin
h
tumor
acetylated tubulin
Hoescht
acetylated tubulin

Figure 6. ETS transcription factors regulate cilia gene expression at the interface in zebrafish and human melanoma.

a. Results from HOMER *de novo* motif analysis of differentially expressed genes in the SRT, scRNA-seq, and snRNA-seq interface clusters. **b.** Top 10 enriched motifs from HOMER known motif analysis of the scRNA-seq tumor-like (left) and muscle-like (right) interface cell states. **c-e.** Relative expression of zebrafish ETS genes across the clusters in the scRNA-seq (c), SRT (d) and snRNA-seq (e) datasets. **f.** Normalized enrichment score of cilia-related pathways enriched in ETS-target genes within the SRT and scRNA-seq interface clusters. **g.** Scatter plot comparing ETS gene expression scores per cell and ETS-target gene expression scores per cell in the scRNA-seq interface cluster. Pearson's correlation coefficient (R) is indicated. **, $p < 0.01$, ***, $p < 0.001$ (Wilcoxon rank sum test with Bonferroni's correction).

Figure 6

a HOMER *de novo* motif enrichment

SRT: interface vs. muscle

motif	best match	*p*-value	rank
CCGGAAGT	ETS	$1e^{-22}$	1

SRT: interface vs. all spots

motif	best match	*p*-value	rank
ACTTCCGGGA	ELF5 (ETS)	$1e^{-15}$	1

scRNA-seq: interface vs. all other cells

motif	best match	*p*-value	rank
CACTTCTCCG	NKX2-3,..., ETV6 (ETS)	$1e^{-13}$	2

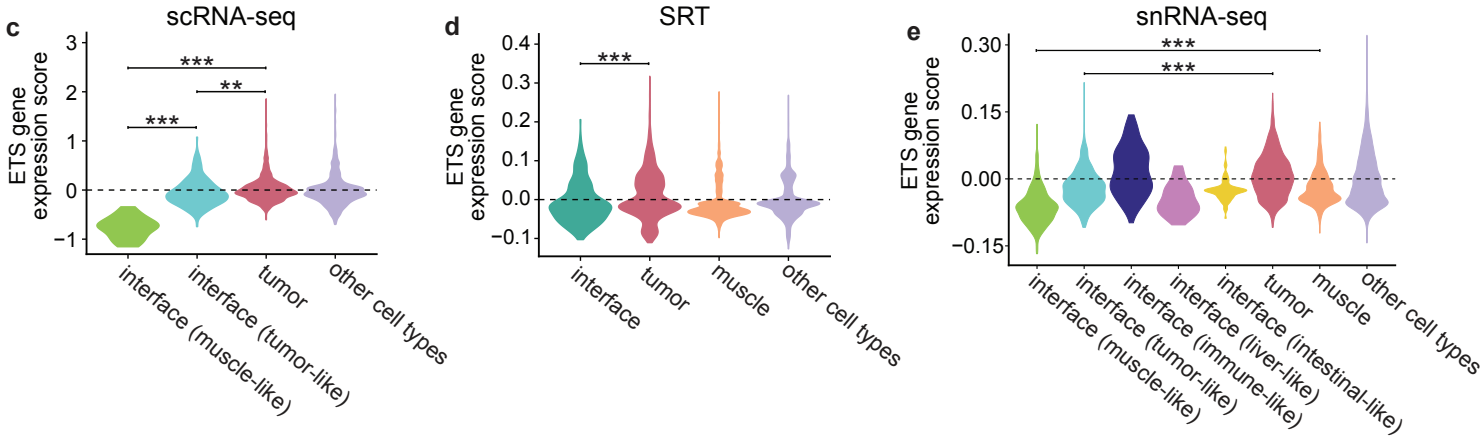
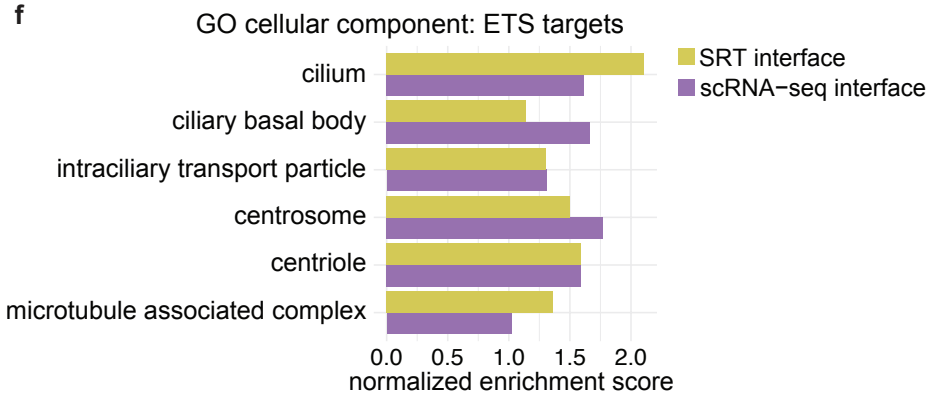
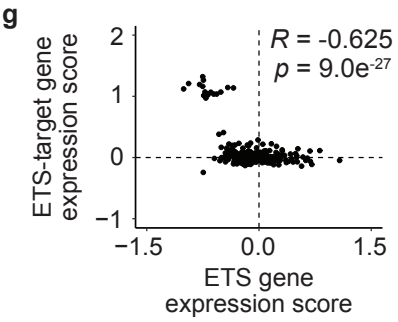
snRNA-seq: interface vs. all other nuclei

motif	best match	*p*-value	rank
TGGACGTCATCCTCAT	FOS::JUN,..., Gabpa (ETS)	$1e^{-13}$	2

b HOMER known motif enrichment

scRNA-seq interface (tumor-like)

rank	motif	best match	*p*-value
1	CACTTCCGGT	Fli1(ETS)	$1e^{-8}$
2	ACAGGAAGTG	ETS1(ETS)	$1e^{-5}$
3	CCTGTGGTTT	RUNX-AML	$1e^{-5}$
4	ACCAGGAAGT	ELF5(ETS)	$1e^{-4}$
5	ACCAGGAAGT	ELF3(ETS)	$1e^{-4}$
6	CACTTCCGGT	Elk1(ETS)	$1e^{-4}$
7	ACCGGAAGTG	ETV4(ETS)	$1e^{-4}$
8	ACTTCCTGTT	Elf4(ETS)	$1e^{-4}$
9	TGGCCCCACCCCTCTG	KLF3(Zf)	$1e^{-4}$
10	AGTGGGCGGAGC	Sp5(Zf)	$1e^{-4}$

scRNA-seq interface (muscle-like)

rank	motif	best match	*p*-value
1	CCCTAGCAACAG	Rfx5(HTH)	$1e^{-10}$
2	GTTGCCATGGCAACC	Rfx2(HTH)	$1e^{-10}$
3	CCGTTGCCATGGCAAC	RFX(HTH)	$1e^{-8}$
4	GGTTGCCATGGCAA	Rfx1(HTH)	$1e^{-7}$
5	TGTTTCCTAGCAACC	Rfx6(HTH)	$1e^{-6}$
6	GGCCGTTA	MYB(HTH)	$1e^{-5}$
7	CACTTCCGGT	Elk4(ETS)	$1e^{-5}$
8	GGTTGCCATGGCAA	X-box(HTH)	$1e^{-5}$
9	CTAACTGCCA	BMYB(HTH)	$1e^{-4}$
10	CACTTCCGGT	Fli1(ETS)	$1e^{-4}$