# Cancer archetypes co-opt and adapt the transcriptional programs of existing cellular states

Maayan Baron[1], Isabella S. Kim[2], Reuben Moncada[1], Yun Yan[1], Nathaniel R. Campbell[3], Richard M. White[2], and Itai Yanai[1,*]

[1] Institute for Computational Medicine, NYU School of Medicine, New York, NY USA
[2] Cancer Biology & Genetics, Memorial Sloan Kettering Cancer Center, New York, NY USA
[3] Weill Cornell/Rockefeller/Sloan-Kettering Tri-Institutional MD-PhD Program, Memorial Sloan Kettering Cancer Center, New York, NY USA

[*] Correspondence to: itai.yanai@nyumc.org

## ABSTRACT

Tumors evolve as independent systems comprising complex survival-ensuring functions, however the nature of these distinct processes and their recurrence across cancers is not clear. Here we propose that melanoma cancer-cells can be classified to three 'archetypes' that co-opt the neural crest, mature melanocytes, and stress gene expression programs, respectively, have a unique subclonal structure, and are conserved between zebrafish and human melanomas. Studying the natural history of a zebrafish melanoma tumor at the single-cell level, we found that one archetype exclusively exhibits the signature of the Warburg effect, suggesting that a shifting balance in energy production occurs differentially in the tumor. Deconvolving bulk human melanomas, we found that patients with a dominant fraction of the neural crest archetype show worse survival rates, indicating a clinical relevance for the composition of archetypes. Finally, we provide evidence that extending our approach to other cancer types can reveal universal and cancer-specific archetypes.

# INTRODUCTION

Cancer is an evolutionary process occurring at the level of an individual's cells – rather than individuals within a species where it is traditionally studied – by which driver mutations accumulate and increase tumor 'fitness' (Nowell, 1976). Indeed, a universal feature of cancer is the genomic accumulation of mutations to a recurring set genes, including *p53*, *BRCA1/2*, and telomerase (Goode, Ulrich, & Potter, 2002; Hollstein, Sidransky, Vogelstein, & Harris, 1991; Vogelstein & Kinzler, 2004). All tumors also show some degree of genetic heterogeneity (McGranahan & Swanton, 2017; Navin et al., 2011; Wang et al., 2014), which may be usefully classified into mutational signatures (Alexandrov et al., 2013). Less well understood, however, is the functional consequences of such heterogeneity – whether stemming from cooperation or competition among the clones of a tumor. Traditionally, tumor heterogeneity is interpreted as a snapshot in the rise to fixation of driver alleles and their linked sites (Hanahan & Weinberg, 2000; Vogelstein et al., 2013), while evidence for the cooperation model is not abundant, though one pioneering study examining sub-clonal heterogeneity did find strong evidence for the notion of clonal cooperation during tumorigenesis (Marusyk et al., 2014). Distinguishing between these two explanations of tumor heterogeneity thus constitutes an important problem in cancer biology.

Single-cell RNA-Seq (scRNA-Seq) has revolutionized the study of tumors by enabling an examination of the composing cell types, defining both the microenvironment and the cancer cells themselves. Exploring the tumor microenvironment at the single-cell level, has revealed exhaustion program of T-cells in melanoma (Tirosh, Izar, et al., 2016), diversity of tissue resident and normal immune cells and the trajectories of activation of T-cells shaped by TCR diversity in breast cancer (Azizi et al., 2018), and a highly complex structure containing 52 stromal cell types in lung cancer (Lambrechts et al., 2018). Transcriptional programs of cancer cells have also been identified in human melanoma, glioblastoma, oligodendroglioma, breast and head and neck cancer (Chung et al., 2017; Patel et al., 2014; Puram et al., 2017; Tirosh, Izar, et al., 2016). For example in glioblastoma, multiple transcriptional programs co-exist according to classical, proneural, neural, and mesenchymal cell states (Patel et al., 2014). However, the plasticity of these transcriptional states is not known. Moreover, the relationship to the

hypothesized cancer stem cell theory (Meacham & Morrison, 2013) has not been established and thus it is not clear if transcriptional states are hierarchically related to one another.

The genetics of melanoma have been well characterized in terms of recurring genes and pathways across nevus and invasive bulk tumors (Akbani et al., 2015; Tsao, Chin, Garraway, & Fisher, 2012). However, in order to understand how heterogeneity is distributed across the cells that compose the tumor, it necessary to study the natural history of this process. A system allowing for sampling the cells of individual tumors over time is thus required. This is difficult to achieve in human patients, however a zebrafish model for cancer has been introduced in which the expression of the human $BRAF^{V600E}$ gene in a *p53*-deficient background leads to 100% fish developing melanoma (Figure 1A) (White et al., 2011). These tumors resemble the human disease at histological (Patton et al., 2005), genomic (Yen et al., 2013), and gene expression levels (Kansler et al., 2017), suggesting that studying these tumors can reveal universal principles of melanoma.

Here, we study melanoma at the level of the transcriptomes of thousands of cells by repeatedly sampling a melanoma tumor over the course of its evolution without treatment. Analyzing the resulting dataset led us to introduce 'archetypes' as a theory for understanding cancer cell-types. An archetype is a cell type that co-opts an existing transcriptional program and adapts it over the progression of the tumor. We propose that each melanoma archetype manifests a distinct function – the neural crest, differentiated, and stress transcriptional programs – in distinct anatomic locations in the tumor. We find t very same archetypes in human melanoma suggesting their universality to melanomas. Over the natural history of the tumor, we detected changes in how the archetypes adapt their co-opted transcriptional program. In particular, we detected activation of the Warburg effect – in terms of turning 'on' glycolysis and turning 'off' oxidative phosphorylation – perhaps due to its central location in the tumor, revealing that the Warburg effect may occur separately for distinct tumor archetypes. Reanalyzing tumors from the TCGA project we found that patients show a worse prognosis if their tumors are dominated by the neural crest archetype. Finally, we report evidence for the recurrence of one of the detected archetypes in breast cancer and pancreatic cancer tumors. Collectively, our results lead to the

hypothesis that tumors of each cancer type will be composed of a coherent set of archetypes, as well as that universal archetypes may span all cancers.

## RESULTS

### Single-cell RNA-Seq on zebrafish melanoma reveals eight distinct cell-types

To understand the natural history of this process, a system is required that allows for sampling individual tumor cells over time and space, which is difficult to achieve in human patients. To overcome this limitation, we employed an isogenic zebrafish model for melanoma which arises from expression of the human *BRAF*$^{V600E}$ gene in a *p53*-deficient background (Figure 1A) (White et al., 2011). 100% of these fish develop melanoma that resembles the human disease at histological (Patton et al., 2005), genomic (Yen et al., 2013), and gene expression levels (Kansler et al., 2017). To begin to study the composition and dynamics of tumorigenesis, we used this experimental system to examine transcriptomes in individual cells using scRNA-Seq.

We processed ~3k cells from a tumor biopsy using the inDrop system (Klein et al., 2015), thus comprehensively capturing the set of cell-types present in the tumor. After quality control and filtering (see STAR Methods), we were left with 1957 cells, each with an average of approximately 2,500 transcripts and 1,000 genes detected. Studying the transcriptomes, we detected eight cell-types which form distinct clusters when visualized using t-distributed stochastic neighbor embedding (tSNE) (Van Der Maaten & Hinton, 2008). To annotate these clusters, we first identified the cancer cell population by the detection of *BRAF*$^{V600E}$ transcripts. The rest of the clusters were annotated using known markers: keratinocytes, fibroblasts, erythrocytes, natural killer cells, neutrophils, macrophages, other immune cells, and the cancer cell population (Figure 1B and S1). For example, a cluster was annotated as keratinocytes since it was enriched in the expression of keratin 4 (*krt4*) and other genes (see Table S1).

### Principal component analysis on human and zebrafish malignant cells reveals three conserved cancer cell-types

We next examined the heterogeneity of the cancer cells using principal components analysis (PCA), and found a triangle-shaped arrangement of transcriptomes, with a concentration of cells

near the vertices (Figure 1C). Studying the genes that most contribute to PC1 and PC2, we found that the *sox2*, *dct*, and *jun* genes exhibit expression each restricted to one of the three vertices (Figure 1D). Furthermore, we found the same configuration of transcriptional programs when examining two additional zebrafish melanoma tumors (Figure S2).

These vertices may be unique to the zebrafish system or general to melanoma neoplasm. To test for their universality, we studied previously reported human melanoma scRNA-Seq data by Tirosh et al. (Tirosh, Izar, et al., 2016) who described the *MITF* and *AXL* transcriptional programs of metastatic cancer cells, respectively corresponding to proliferating (Carreira et al., 2005) and invasive features (Müller et al., 2014). Since our zebrafish tumor was induced by the *mitfa* promoter, we focused on the human cells annotated with the *MITF*-transcriptional program and found that these have a strikingly similar pattern of expression to the zebrafish tumor expression (Figure 1E). This correspondence suggests that the transcriptional programs that we observe in zebrafish are a conserved aspect of melanoma.

**Cancer cell-types as archetypes**

Recent work has shown that a population of individuals in 'morphospace' often reveals geometrical shapes, whose vertices may be interpreted as 'archetypes', each capturing an idealized state of the population (Shoval et al., 2012). The notion of archetypes has also been extended to the study of bulk cancer samples (Hart et al., 2015). Here, we invoke archetypes to distinguish cancer cell-types within an individual tumor, by identifying the cells most closely associated with the vertices of the triangle shape in the PC analysis (Figure 1C and STAR Methods). We proceeded to characterize each archetype by identifying the most uniquely-expressed genes among the 500 cells closest to the corresponding vertex (Figure 2A). We found that archetype 1 is enriched for the expression of neural crest genes, such as *sox2* and *sox10*, suggesting that these cells co-opt this progenitor transcriptional program. In contrast, archetype 2 is enriched with the expression of genes associated with mature melanocytes, such as *dct*, *tyrp1b*, and *pmela*, indicating the co-option of the differentiated melanocyte transcriptional program. To test for the distinction between archetypes 1 and 2, we compared their expression profiles to a dataset that measured the gene expression changes that accompany human melanocyte differentiation from pluripotent stem cells (Mica, Lee, Chambers, Tomishima, & Studer, 2013).

As expected, we found that archetypes 1 and 2 best correlate in the expression to neural crest cells and mature melanocytes, respectively (Figure 2B), further supporting the notion that these archetypes capture distinct stages of the melanocyte differentiation process.

We found that archetype 3 is enriched in the expression of genes such as *jun, fosb, fosab*, and eight 'heat-shock protein' genes, all involved in an apparent stress-activated transcriptional program. Cells expressing this program have also been observed in other single-cell tumor analyses (Puram et al., 2017; Tirosh, Izar, et al., 2016), where they were found to be in a distinct – though uncharacterized – tumor region (Tirosh, Izar, et al., 2016). To test if archetype 3 also occupies a distinct region in the zebrafish melanoma tumor, we invoked a spatial transcriptomics approach. We sectioned a zebrafish tumor and placed it on a DNA microarray where each spot queried the entire polyA complement of the transcriptomes of the surrounding cells. Delineating three concentric rings in the tumor, we found that expression of archetype 3 genes is higher in the outer ring of the tumor (Figure 2C, $P<10^{-10}$, Wilcoxon rank-sum test, Figure S3). This suggests that the cells of this archetype are enriched at the outer edges of the tumor, thus relating the function of the archetype 3 cells to their spatial organization. Enriched expression of the stress module in the outer ring also controls for the possibility that the stress-induced signals follows from the cell dissociation protocol (Van Den Brink et al., 2017). Together, our inference of the three archetypes suggests that cancer cells co-opt existing transcriptional programs corresponding to neural-crest cells, mature melanocytes, and a stress response.

## Copy number variations distinguish among the archetypes

Beyond transcriptional programs, scRNA-Seq data allows us to study copy number variations (CNVs) at the level of individual cells. As previously introduced (Patel et al., 2014) and used (Darmanis et al., 2017; Kim et al., 2018; Tirosh, Izar, et al., 2016; Tirosh, Venteicher, et al., 2016), by mapping gene expression levels to their chromosomal locations, large-scale CNVs can be inferred. Analyzing our scRNA-Seq data, we found striking evidence for at least two distinct clones (Figure 3A). Both clones were found to lack one or two copies of chromosome 3. We detected an amplification and a loss of chromosome 19 and 16, respectively, in clone A, while clone B cells showed an amplification of chromosomes 10 and 23. These observations may be driven by the transcriptional levels and not actual DNA changes, and so we performed three

controls to mitigate this possibility. First, we performed bulk whole-genome sequencing on a sample of the same single-cell suspension used to generate the scRNA-Seq data and detected a pattern of deletions and amplifications that is consistent with the RNA-inferred CNVs (Figure 3A, lower panel). Second, we excluded from the RNA-CNV analysis the 235 genes whose expression is archetype-specific (Figure 2A) in order to eliminate the effect of uniquely expressed genes. Finally, we studied two zebrafish melanoma cell lines cultured in two conditions and found that while gene expression of the cells clusters by condition, the RNA-CNV analysis retains the strain-specific pattern (Figure S4), thus indicating the robustness of the RNA-CNV approach. Together these results support the existence of two genetically distinct clones present in the tumor.

Projecting clones A and B cells to the same PCA-space, we found that each clone localized to a specific region (Figure 3B). Clone A cells matched archetype 1, while clone B cells matched both archetypes 2 and 3. This dichotomy of two clones with distinct archetype suggests that the clones are adapted for different functions. Since our data indicates that archetype 1 (clone A) cells are enriched for the interior of the tumor (Figure 2C), we hypothesized that a fundamental functional distinction between clones A and B is the increased interaction of clone B cells with the micro-environment. We sought to test this by studying the zebrafish melanoma cell lines in two conditions: 1. *in vitro*, where cells are cultured in a dish and do not interact with non-cancer cell types and 2. *in vivo*, where cells were injected into zebrafish embryos resulting in a tumor, whose cells are naturally exposed to other cell types (Figure 3C). For each of the two cell lines (ZMEL1 and ZMEL2), we found that cells from the *in vitro* condition clustered with archetype 1 (clone A) while cells from the *in vivo* condition clustered with archetype 3 (clone B). This supports the hypothesis that clones A and B may indeed reflect functions involved in proliferation and tumor-microenvironment interaction, respectively. These results also highlight the notion that cancer clones may specialize in a particular archetype and achieve a distinct genotype over time (Figure 3A-B), though the same genotype may also adopt different archetypes through transcriptional plasticity (Figure 3C-D).

## Natural history of clonal gene expression during tumor progression

To study changes over time in the gene expression of the two clones, we exploited the zebrafish system's ability to repeatedly query the same tumor. We thus performed three additional weekly biopsies for the same analyzed tumor to produce a 4-week period time-course dataset (Figure 4A). We detected both clones in similar proportions at each biopsy (Figure 4B). In order to identify functional properties that change over time in the tumor, we queried for KEGG pathways that contain genes significantly varying across the biopsies (see STAR Methods). For clone A, we found that glycolysis genes increase in expression over time ($P < 10^{-7}$, hypergeometric test, Figure 4C), suggesting activation of the Warburg effect (Warburg, 1956) of a shifting balance in energy production. Consistently, oxidative phosphorylation genes exhibited a correlated decrease in expression ($P < 10^{-10}$, hypergeometric test, Figure 4C). Furthermore, this effect was restricted to clone A (Figure 4D), suggesting that clone A / archetype 1's internal location in the tumor (Figure 2C) may selectively activate the Warburg effect.

In clone B, we found that MAPK signaling significantly increases in expression ($P < 10^{-5}$, hypergeometric test, Figure 4D), while glycolysis genes decrease, suggesting adaptation beyond the co-opted differentiation program. Expanding this analysis, Figure 4E shows the expression profiles of 18 KEGG pathways over time, revealing the distinct clonal trajectories, with the exception of FoxO signaling and focal adhesion which increase in expression for both. Overall, this analysis suggests that signaling pathways and metabolic pathways are selected for dynamic expression in clone A and B, respectively. We hypothesized that as the clones diverge into these complementary functions, they would resemble the idealized expression of functions of cells growing under *in vitro* or *in vivo* conditions. To test this, we again turned to our zebrafish melanoma cell lines (Figure 3C) and identified for each KEGG pathway the level of enrichment for differential expression between the *in vitro* and *in vivo* conditions. We found that pathways that increase in clone A are generally significantly enriched for expression in the *in vivo* condition. In contrast, clone B adaptive pathways are enriched for increased expression in the *in vitro* condition ($P<10^{-2}$, Wilcoxon rank-sum, Figure 4E). We conclude from this analysis that clone-specific changes to gene expression implicate distinct and independent functional requirements.

## Archetype frequency is correlated with clinical prognosis

To study the effect of the archetypes on clinical outcome we extended our analysis to the TCGA-Skin Cutaneous Melanoma dataset which contains bulk RNA-Seq data for 472 melanoma tumors (Akbani et al., 2015). For each tumor, we inferred the fraction of each archetype by deconvolving the expression profile using the expression of archetype markers (see STAR Methods). As expected, we found that most bulk tumors are composites of all three archetypes (Figure 5A), and archetype 3 was not a dominant archetype in any of the tumors. We distinguished four groups of tumors according to their archetype frequencies: Group 1 containing tumors comprising exclusively archetype 1 (neural crest), group 2 containing tumors comprising all archetypes but dominated mostly by archetype 1, group 3 containing tumors comprising all archetypes but dominated mostly by archetype 2 (mature melanocyte) and group 4 containing tumors comprising only archetype 2 and 3 (stress) (Figure 5A).

Studying the gene expression profiles of the bulk tumors using PCA revealed a triangle shape. Interestingly, archetype frequency is correlated with PC2: high PC2 scores corresponded to patients with high fractions of archetype 1 and low PC2 corresponded to patients with high fraction of archetype 2 (Figure 5B). However, other aspects available from the clinical data such as gender, invasiveness, and survival did not correlate with the PCs (Figure S5). We next asked if patients with tumors consisting of different archetype frequencies experienced significantly different clinical outcomes. Focusing on the two groups of tumors that exhibited a mixed combination of archetypes (groups 2 and 3) we found that the group with higher fractions of archetype 1 has a significantly worse prognosis than the group with higher fractions of archetype 2 ($P<0.05$). This result indicates that while a mixture of archetypes is a feature of most tumors, heterogeneity in their frequencies may be a predictor of tumor fate.

## The stress archetype is conserved in breast and pancreatic cancers

To study the generality of the notion of archetypes across cancer types we applied our analytic approach to two previously published scRNA-Seq tumor datasets: triple negative breast cancer (TNBC) (Kim et al., 2018) and pancreatic cancer adenocarcinoma (PDAC) (Moncada et al., 2018). For each cancer type, we analyzed the transcriptomes of only the cancer cells: 388 single

cells, from one TNBC patient and 462 single cells from another PDAC patient, both before treatment. Studying these cancer cells using PCA, we found a triangle-shaped distribution of cells for each cancer type (Figure 6), reminiscent to that found for melanoma (Figure 1). This recurrent distribution of cells of a tumor according to their gene expression indicates that transcriptional heterogeneity across cancer cells is not unique to melanoma.

Querying for the expression of key melanoma archetype genes in the TNBC and PDAC datasets, we found that the expression of *JUN* is restricted to a single vertex of the triangle-shaped distribution in both cancers (Figure 6A,B, left panels). Since *JUN* was identified as a marker gene for the stress archetype (Figure 1D and E, right panel), we asked if expression of the entire gene signature of that archetype (Table S2) is also enriched for expression in this vertex. Indeed, we detected expression of the stress archetype signature for each of the tumors in the vertex where *JUN* is enriched for expression (Figure 6A and B, right panels). These results provide evidence for the generality of the archetype concept across cancers and moreover that the stress archetype may be conserved across diverse cancer types.

# DISCUSSION

Here we have studied the gene expression of individual cancer cells in zebrafish and human melanomas. We detected three recurring gene expression patterns across melanoma cancer cells which we refer to as 'archetypes' because they appear to be co-opted cellular states. Over the temporal progression of the tumor, we detected the presence of all three archetypes in similar frequencies (Figure 4), though their spatial locations in the tumor were distinct (Figure 2). An analysis of the tumor heterogeneity revealed that archetype 1 was encoded by a unique clone while a second clone encoded two archetypes (Figure 3), suggesting a greater transcriptional plasticity for that clone. In this section, we discuss the notion of archetypes as a coherent theory for understanding cancer cell-types, the relationships between the archetype notion and central concepts in cancer biology, and the potential clinical relevance of archetypes.

We defined a cancer archetype as a transcriptional program that is co-opted from that of an existing cell-type or cell-state. Strikingly, we found that cancer cells between human and zebrafish melanomas exhibit the same transcriptional programs. We further provided evidence that one of the archetypes – that of the stress cell state (archetype 3) – is present across other cancers (Figure 6). The other two archetypes account for two states (pluripotent and differentiated, respectively) along the development of melanocytes, the melanoma cell of origin. Tsoi *et al.* also detected that bulk melanoma tumors show different states along the trajectory of differentiation (Tsoi et al., 2018). Together, these results led us to propose that the archetypes present in a particular cancer type will include those that co-opt states along the development of the cell-of-origin of that cancer. This may result from those particular states being more accessible from the particular epigenetic state of the cell that initiates tumorigenesis. It will be interesting to further explore the dynamics by which archetype adapt over time. Since in our study we examined a relatively later stage in tumorigenesis in which the tumor is already fully formed, we did not gain purview into the origin of archetypes (Figure 4). It should also be possible to provide a dictionary of all cancer archetypes – whereby each tumor is a collection of a few of these – allowing for a deeper functional understanding of tumors of any given cancer type.

A central concept in cancer biology posits that hallmarks are universal neoplastic features (Hanahan & Weinberg, 2000). In our analysis, we captured evidence for the appearance of one hallmark, the Warburg effect: over time, the glycolysis pathway was activated while oxidative phosphorylation decreased in expression (Figure 4). Interestingly, however, we observed this pattern only in archetype 1, suggesting that distinct hallmarks may accumulate across the archetypes rather than in all of them. This observation, together with the result that archetypes co-opt distinct transcriptional programs, leads us to propose that cooperation among the archetypes in the tumor is vital to its overall functional. Previous work has indeed revealed a form of cooperation across cancer cell lines as they form a tumor *in vivo* (Marusyk et al., 2014). The coexistence of the archetypes in tumors also accounts for the observed genetic heterogeneity universal to cancers (Anderson et al., 2011; Lawrence et al., 2013; Mroz et al., 2013; Sottoriva et al., 2013). While natural selection leads to an increase in the frequency of certain alleles, substantial variation may also be maintained as cooperating coexisting archetypes. Previous work has also suggested a hierarchical model for cancer cells in the tumor, with cancer stem cells occupying a key role in the generation of cancer cell-types (Meacham & Morrison, 2013). Our results indicating the presence of distinct clones (Figure 3) suggest that – while one of our archetypes co-opts a stem cell module – these cells are not hierarchically related to other cells in the tumor.

The notion of archetypes as tumor building blocks has clinical implications. Since different drugs likely lead to distinct impacts across the different archetypes, a tumor may resist a particular drug by such transcriptional plasticity. Indeed, we observed that two archetypes may follow from the same clone (Figure 3), suggesting that if a tumor is depleted of the function of one archetype, it may be able to recover it using transcriptional plasticity followed by gradual evolution. By defining the drugs that best target each specific archetype, it may be possible to develop a more powerful chemotherapy consisting of a combination of drugs that match the archetypes composing the particular tumor. Collectively, we have proposed here a theory for studying a tumor through the archetypes that compose it. Future insight may follow by uncovering the spatial-temporal nature of cancer archetypes in the tumor and their specific interaction with the microenvironment.

**Acknowledgments.** We thank Naftalie Senderovich, and Anna Yeaton for work on the initial pilot of the project and Dalia Barkley for the constructive feedback. We thank Matt Maurano and Megan Hogan for assistance with sequencing. We also thank the NYU Langone Genome Technology Center for assisting with the whole genome sequencing.

**Author contributions.** I.Y. conceived the project. M.B. led the collection and sequencing of the single-cell RNA-Seq data collection, with contributions from I.S.K. and N.R.C.. R.M. contributed the spatial transcriptomics data. M.B. provided the WGS and Y.Y. contributed to its analysis. M.B. led the analysis of the data, with significant contribution from I.Y.. I.Y. and R.M.W. provided project coordination. I.Y. and M.B. drafted the manuscript on which all authors commented.

**Declaration of interests.** The authors declare no competing interests.

**Data and materials availability:** The complete data that support the findings of this study have been deposited in NCBI GEO database with the accession code GSE115140.

## References

Akbani, R., Akdemir, K. C., Aksoy, B. A., Albert, M., Ally, A., Amin, S. B., … Zou, L. (2015). Genomic Classification of Cutaneous Melanoma. *Cell*, *161*(7), 1681–1696. https://doi.org/10.1016/j.cell.2015.05.044

Alexandrov, L. B., Nik-Zainal, S., Wedge, D. C., Aparicio, S. A. J. R., Behjati, S., Biankin, A. V., … Stratton, M. R. (2013). Signatures of mutational processes in human cancer. *Nature*, *500*(7463), 415–421. https://doi.org/10.1038/nature12477

Anderson, K., Lutz, C., Van Delft, F. W., Bateman, C. M., Guo, Y., Colman, S. M., … Greaves, M. (2011). Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*, *469*(7330), 356–361. https://doi.org/10.1038/nature09650

Azizi, E., Carr, A. J., Plitas, G., Mazutis, L., Rudensky, A. Y., Pe'er, D., … Kadaveru, K. (2018). Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment Resource Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. *Cell*, *174*, 1–16. https://doi.org/10.1016/j.cell.2018.05.060

Carreira, S., Goodall, J., Aksan, I., La Rocca, S. A., Galibert, M.-D., Denat, L., … Goding, C. R. (2005). Mitf cooperates with Rb1 and activates p21Cip1 expression to regulate cell cycle progression. *Nature*, *433*(7027), 764–9. https://doi.org/10.1038/nature03269

Chung, W., Eum, H. H., Lee, H. O., Lee, K. M., Lee, H. B., Kim, K. T., … Park, W. Y. (2017). Single-cell RNA-seq enables comprehensive tumour and immune cell profiling in primary breast cancer. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms15081

Darmanis, S., Sloan, S. A., Croote, D., Mignardi, M., Chernikova, S., Samghababi, P., … Quake, S. R. (2017). Single-Cell RNA-Seq Analysis of Infiltrating Neoplastic Cells at the Migrating Front of Human Glioblastoma. *Cell Reports*, *21*(5), 1399–1410. https://doi.org/10.1016/j.celrep.2017.10.030

Goode, E. L., Ulrich, C. M., & Potter, J. D. (2002). Polymorphisms in DNA repair genes and associations with cancer risk. *Cancer Epidemiology, Biomarkers & Prevention : A Publication of the American Association for Cancer Research, Cosponsored by the American Society of Preventive Oncology*, *11*(12), 1513–1530. https://doi.org/Published December 2002

Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. *Cell*, *100*(1), 57–70. https://doi.org/10.1007/s00262-010-0968-0

Hart, Y., Sheftel, H., Hausser, J., Szekely, P., Ben-Moshe, N. B., Korem, Y., … Alon, U. (2015). Inferring biological tasks using Pareto analysis of high-dimensional data. *Nature Methods*, *12*(3), 233–235. https://doi.org/10.1038/nmeth.3254

Hollstein, M., Sidransky, D., Vogelstein, B., & Harris, C. C. (1991). P53 Mutations in Human Cancers. *Science*, *253*(5015), 49–53. https://doi.org/10.1126/science.1905840

Kansler, E. R., Verma, A., Langdon, E. M., Simon-Vermot, T., Yin, A., Lee, W., … White, R. M. (2017). Melanoma genome evolution across species. *BMC Genomics*, *18*(1). https://doi.org/10.1186/s12864-017-3518-8

Kim, C., Gao, R., Sei, E., Brandt, R., Hartman, J., Hatschek, T., … Navin, N. E. (2018). Chemoresistance Evolution in Triple-Negative Breast Cancer Delineated by Single-Cell Sequencing. *Cell*, *173*(4), 879–893.e13. https://doi.org/10.1016/j.cell.2018.03.041

Klein, A. M., Mazutis, L., Akartuna, I., Tallapragada, N., Veres, A., Li, V., … Kirschner, M. W. (2015). Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells. *Cell*, *161*(5), 1187–1201. https://doi.org/10.1016/j.cell.2015.04.044

Lambrechts, D., Wauters, E., Boeckx, B., Aibar, S., Nittner, D., Burton, O., … Thienpont, B. (2018). Phenotype molding of stromal cells in the lung tumor microenvironment. *Nature Medicine*, pp. 1–13. https://doi.org/10.1038/s41591-018-0096-5

Lawrence, M. S., Stojanov, P., Polak, P., Kryukov, G. V., Cibulskis, K., Sivachenko, A., … Getz, G. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*, *499*(7457), 214–218. https://doi.org/10.1038/nature12213

Marusyk, A., Tabassum, D. P., Altrock, P. M., Almendro, V., Michor, F., & Polyak, K. (2014). Non-cell-autonomous driving of tumour growth supports sub-clonal heterogeneity. *Nature*, *514*(7520), 54–58. https://doi.org/10.1038/nature13556

McGranahan, N., & Swanton, C. (2017). Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell*. https://doi.org/10.1016/j.cell.2017.01.018

Meacham, C. E., & Morrison, S. J. (2013). Tumour heterogeneity and cancer cell plasticity. *Nature*. https://doi.org/10.1038/nature12624

Mica, Y., Lee, G., Chambers, S. M., Tomishima, M. J., & Studer, L. (2013). Modeling Neural Crest Induction, Melanocyte Specification, and Disease-Related Pigmentation Defects in hESCs and Patient-Specific iPSCs. *Cell Reports*, *3*(4), 1140–1152. https://doi.org/10.1016/j.celrep.2013.03.025

Moncada, R., Chiodin, M., Devlin, J. C., Baron, M., Simeone, D., & Yanai, I. (2018). Building a tumor atlas: integrating single-cell RNA-seq data with spatial transcriptomics in pancreatic cancer. *BioRxiv*. https://doi.org/10.1101/254375

Mroz, E. A., Tward, A. D., Pickering, C. R., Myers, J. N., Ferris, R. L., & Rocco, J. W. (2013). High intratumor genetic heterogeneity is related to worse outcome in patients with head and neck squamous cell carcinoma. *Cancer*, *119*(16), 3034–3042. https://doi.org/10.1002/cncr.28150

Müller, J., Krijgsman, O., Tsoi, J., Robert, L., Hugo, W., Song, C., … Peeper, D. S. (2014). Low MITF/AXL ratio predicts early resistance to multiple targeted drugs in melanoma. *Nature Communications*, *5*(1), 5712. https://doi.org/10.1038/ncomms6712

Navin, N., Kendall, J., Troge, J., Andrews, P., Rodgers, L., McIndoo, J., … Wigler, M. (2011). Tumour evolution inferred by single-cell sequencing. *Nature*, *472*(7341), 90–95. https://doi.org/10.1038/nature09807

Nowell, P. C. (1976). The clonal evolution of tumor cell populations. *Science (New York, N.Y.)*, *194*(4260), 23–8. https://doi.org/10.1126/science.959840

Patel, A. P., Tirosh, I., Trombetta, J. J., Shalek, A. K., Gillespie, S. M., Wakimoto, H., … Bernstein, B. E. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science*, *344*(6190), 1396–1401. https://doi.org/10.1126/science.1254257

Patton, E. E., Widlund, H. R., Kutok, J. L., Kopani, K. R., Amatruda, J. F., Murphey, R. D., … Zon, L. I. (2005). BRAF mutations are sufficient to promote nevi formation and cooperate with p53 in the genesis of melanoma. *Current Biology*, *15*(3), 249–254. https://doi.org/10.1016/j.cub.2005.01.031

Puram, S. V., Tirosh, I., Parikh, A. S., Patel, A. P., Yizhak, K., Gillespie, S., … Bernstein, B. E. (2017). Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. *Cell*, *171*(7), 1611–1624.e24. https://doi.org/10.1016/j.cell.2017.10.044

Shoval, O., Sheftel, H., Shinar, G., Hart, Y., Ramote, O., Mayo, A., … Alon, U. (2012). Evolutionary trade-offs, pareto optimality, and the geometry of phenotype space. *Science*, *336*(6085), 1157–1160. https://doi.org/10.1126/science.1217405

Sottoriva, A., Spiteri, I., Piccirillo, S. G. M., Touloumis, A., Collins, V. P., Marioni, J. C., …

Tavaré, S. (2013). Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(10), 4009–14. https://doi.org/10.1073/pnas.1219747110

Tirosh, I., Izar, B., Prakadan, S. M., Wadsworth, M. H., Treacy, D., Trombetta, J. J., … Garraway, L. A. (2016). Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. *Science*, *352*(6282), 189–196. https://doi.org/10.1126/science.aad0501

Tirosh, I., Venteicher, A. S., Hebert, C., Escalante, L. E., Patel, A. P., Yizhak, K., … Suvà, M. L. (2016). Single-cell RNA-seq supports a developmental hierarchy in human oligodendroglioma. *Nature*, *539*(7628), 309–313. https://doi.org/10.1038/nature20123

Tsao, H., Chin, L., Garraway, L. A., & Fisher, D. E. (2012). Melanoma: From mutations to medicine. *Genes and Development*. https://doi.org/10.1101/gad.191999.112

Tsoi, J., Robert, L., Paraiso, K., Galvan, C., Sheu, K. M., Lay, J., … Graeber, T. G. (2018). Multi-stage Differentiation Defines Melanoma Subtypes with Differential Vulnerability to Drug-Induced Iron-Dependent Oxidative Stress. *Cancer Cell*, *33*(5), 890–904.e5. https://doi.org/10.1016/j.ccell.2018.03.017

Van Den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., … Van Oudenaarden, A. (2017). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods*. https://doi.org/10.1038/nmeth.4437

Van Der Maaten, L., & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605. https://doi.org/10.1007/s10479-011-0841-3

Vogelstein, B., & Kinzler, K. W. (2004). Cancer genes and the pathways they control. *Nature Medicine*. https://doi.org/10.1038/nm1087

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., & Kinzler, K. W. (2013). Cancer genome landscapes. *Science*. https://doi.org/10.1126/science.1235122

Wang, Y., Waters, J., Leung, M. L., Unruh, A., Roh, W., Shi, X., … Navin, N. E. (2014). Clonal evolution in breast cancer revealed by single nucleus genome sequencing. *Nature*, *512*(7513), 155–160. https://doi.org/10.1038/nature13600

Warburg, O. (1956). On respiratory impairment in cancer cells. *Science (New York, N.Y.)*, *124*(3215), 269–70. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/13351639

White, R. M., Cech, J., Ratanasirintrawoot, S., Lin, C. Y., Rahl, P. B., Burke, C. J., … Zon, L. I. (2011). DHODH modulates transcriptional elongation in the neural crest and melanoma.

*Nature*, *471*(7339), 518–522. https://doi.org/10.1038/nature09882

Yen, J., White, R. M., Wedge, D. C., Van Loo, P., de Ridder, J., Capper, A., … Futreal, P. A. (2013). The genetic heterogeneity and mutational burden of engineered melanomas in zebrafish models. *Genome Biology*, *14*(10). https://doi.org/10.1186/gb-2013-14-10-r113

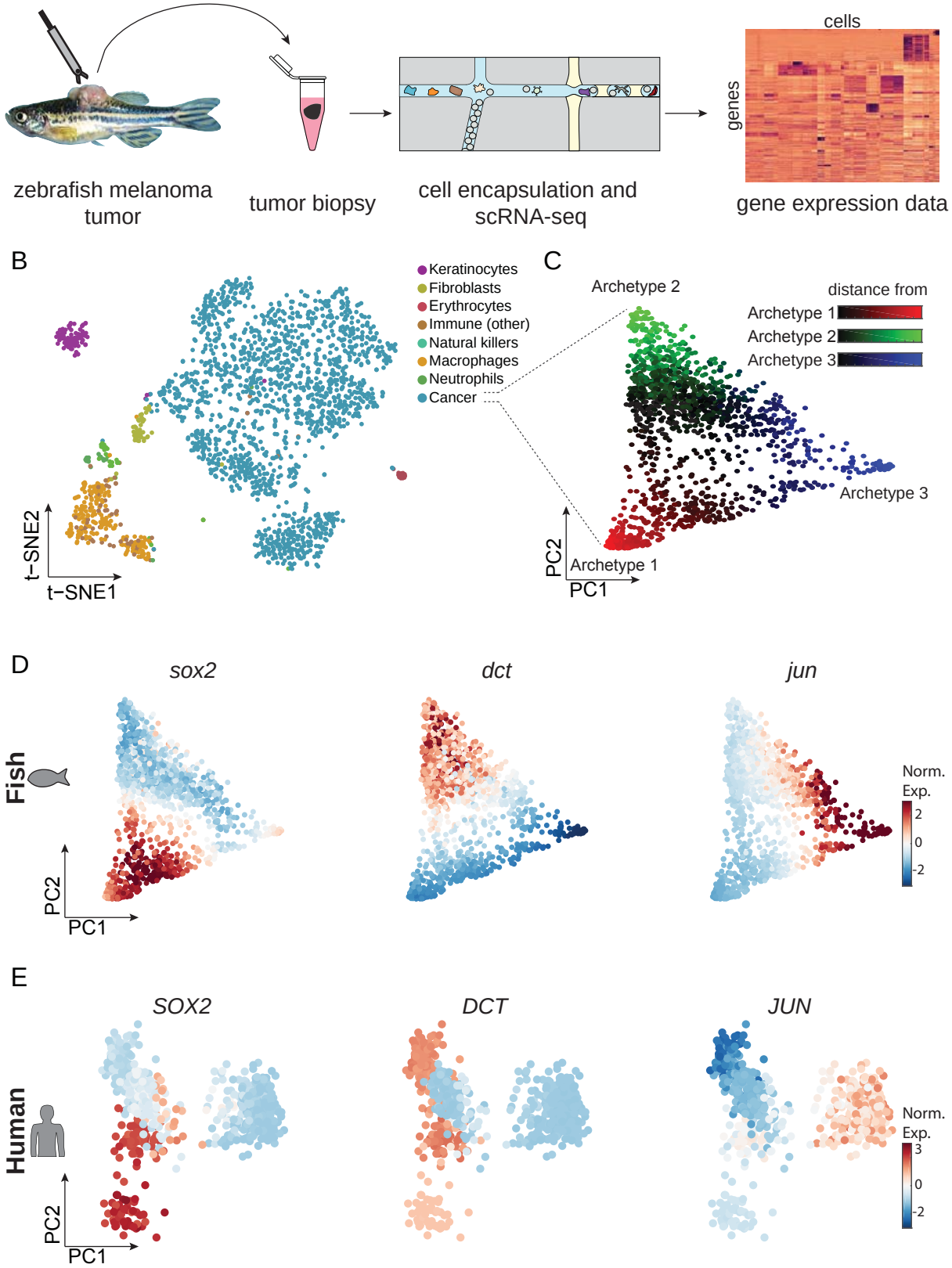


**Figure 1. Single-cell RNA-Seq on zebrafish melanoma**

(A) Tumor biopsy was processed using scRNA-Seq, producing a gene expression matrix.
(B) tSNE analysis of 1957 tumor cells from the biopsy. Color indicates the inferred cell-type.
(C) PCA on the tumor cancer cells. Color indicates distance from the three vertices.
(D) Gene expression levels for the indicated genes, mapped onto the PCA shown in (C).
(E) Analysis of human melanoma data indicating expression levels across the cancer cells for the indicated genes.

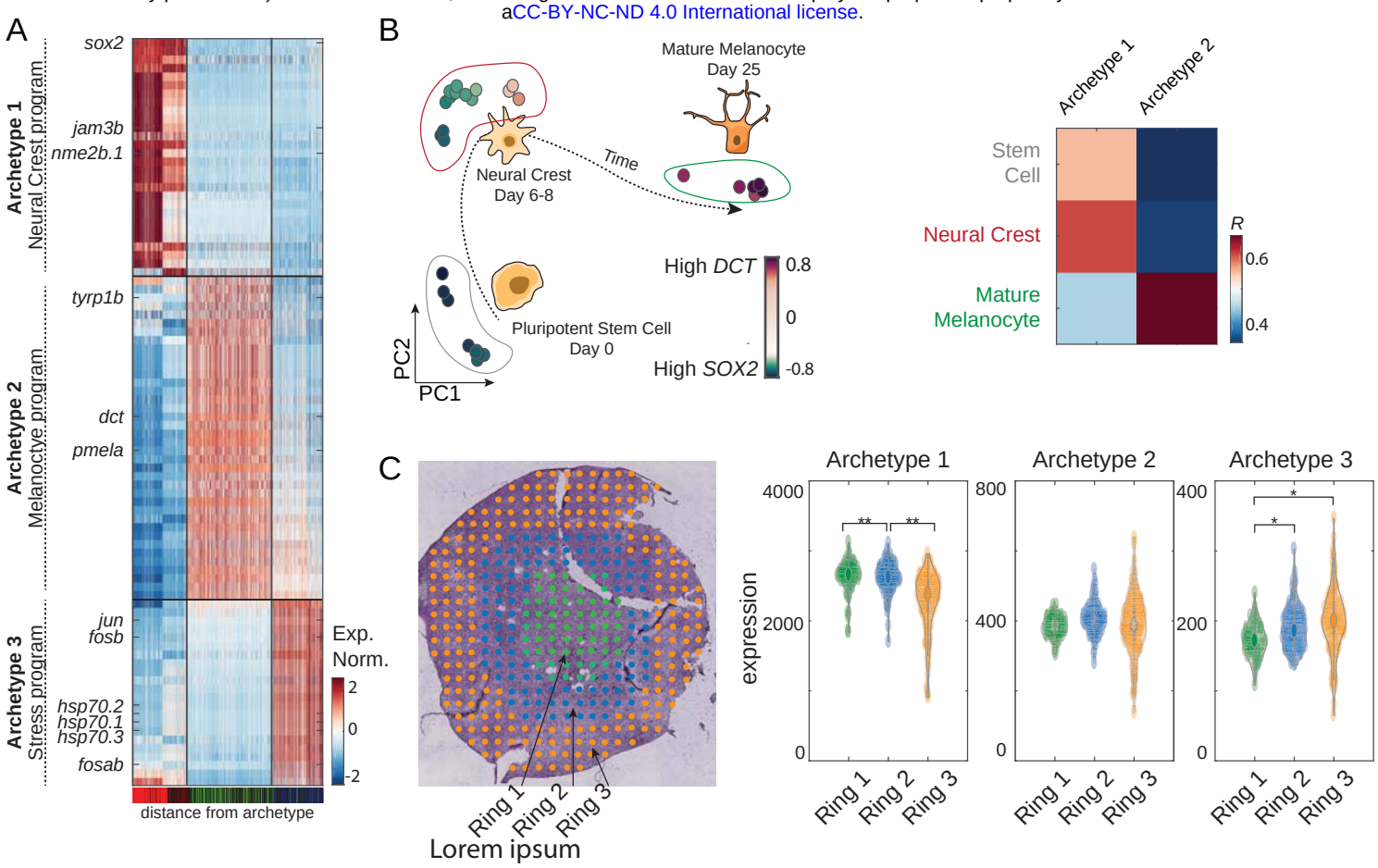


**Figure 2. Functional and structural analysis of the cancer archetypes**

(A) Expression levels of the differentially expressed genes across the cancer cells ordered by archetype proximity in Fig. 1C.

(B) PCA of bulk melanocyte differentiation from previously reported data (left panel). The color indicates expression levels of SOX2 (red) and DCT (green). Heatmap showing the Pearson's correlation levels between the human Archetype 1 & 2 and the developmental transcriptomes of stem cells, neural crest, and mature melanocytes (right panel).

(C) Spatial transcriptomics on a zebrafish p53-/BRAFV600E tumor. The micrograph shows H & E staining with three concentric rings indicated. Dots correspond to spatial transcriptomics array spots. The violin plots indicate the expression levels of neural-crest, mature melanocytes and stress transcription programs genes, respectively, across the three rings. *, P<10-10; **, P<10-20.

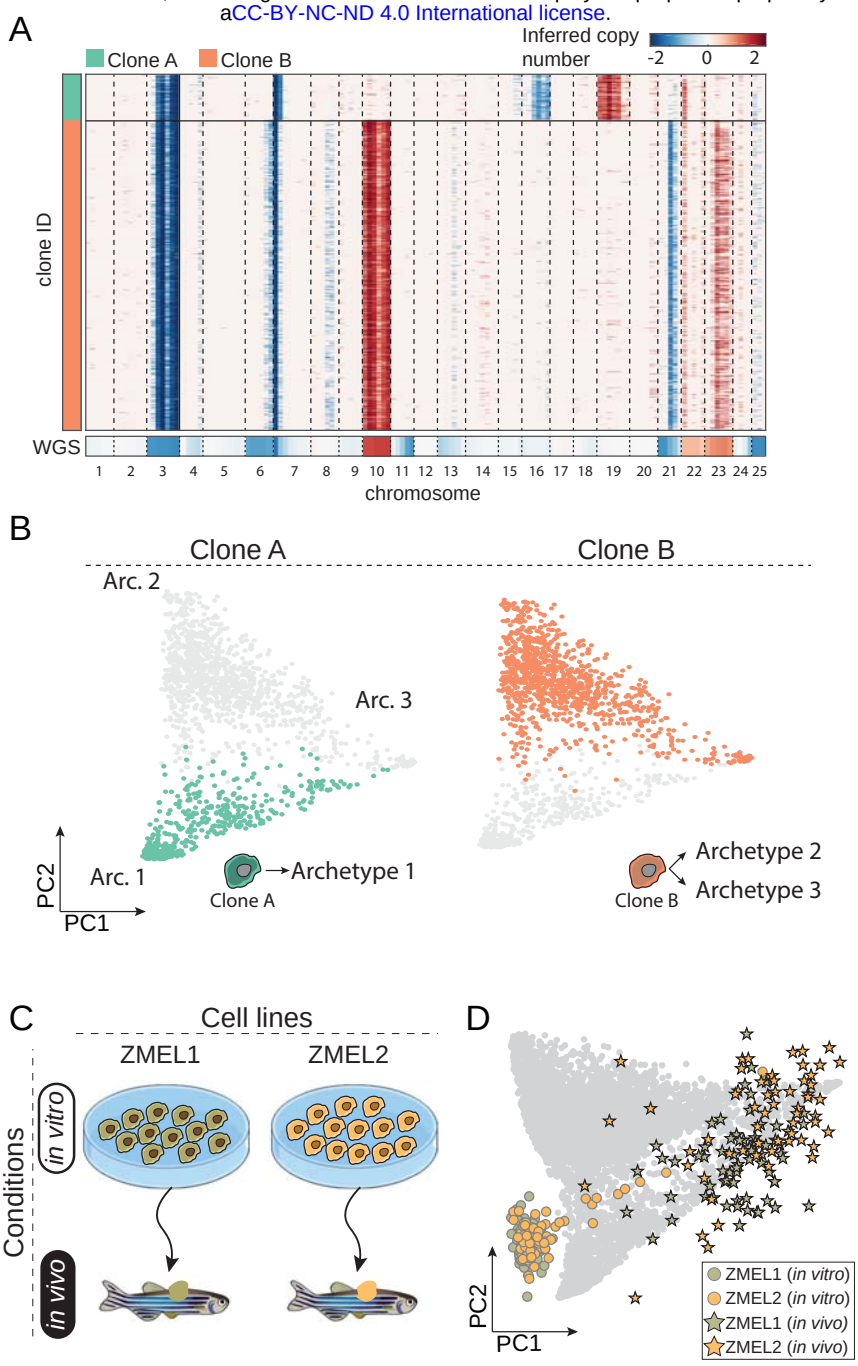


**Figure 3. Genetic and non-genetic determinants of the transcriptional programs**

(A) RNA and DNA CNV analysis. RNA-CNVs were inferred from the single-cell RNA-Seq data. Red and blue indicate chromosomal additions and deletions, respectively. Two clones (clusters) were identified. Inferred CNVs were globally validated by whole genome sequencing copy number analysis (lower bar).

(B) The locations of the cells of each clone (cluster), identified in (A), are mapped onto the PCA of cancer cells (shown in Fig. 2C).

C) Schematic of two zebrafish melanoma cell lines cultured both in vitro and in vivo.

D) Projection of the expression of the cells represented in (C) onto the PCA shown in Figure 1C.

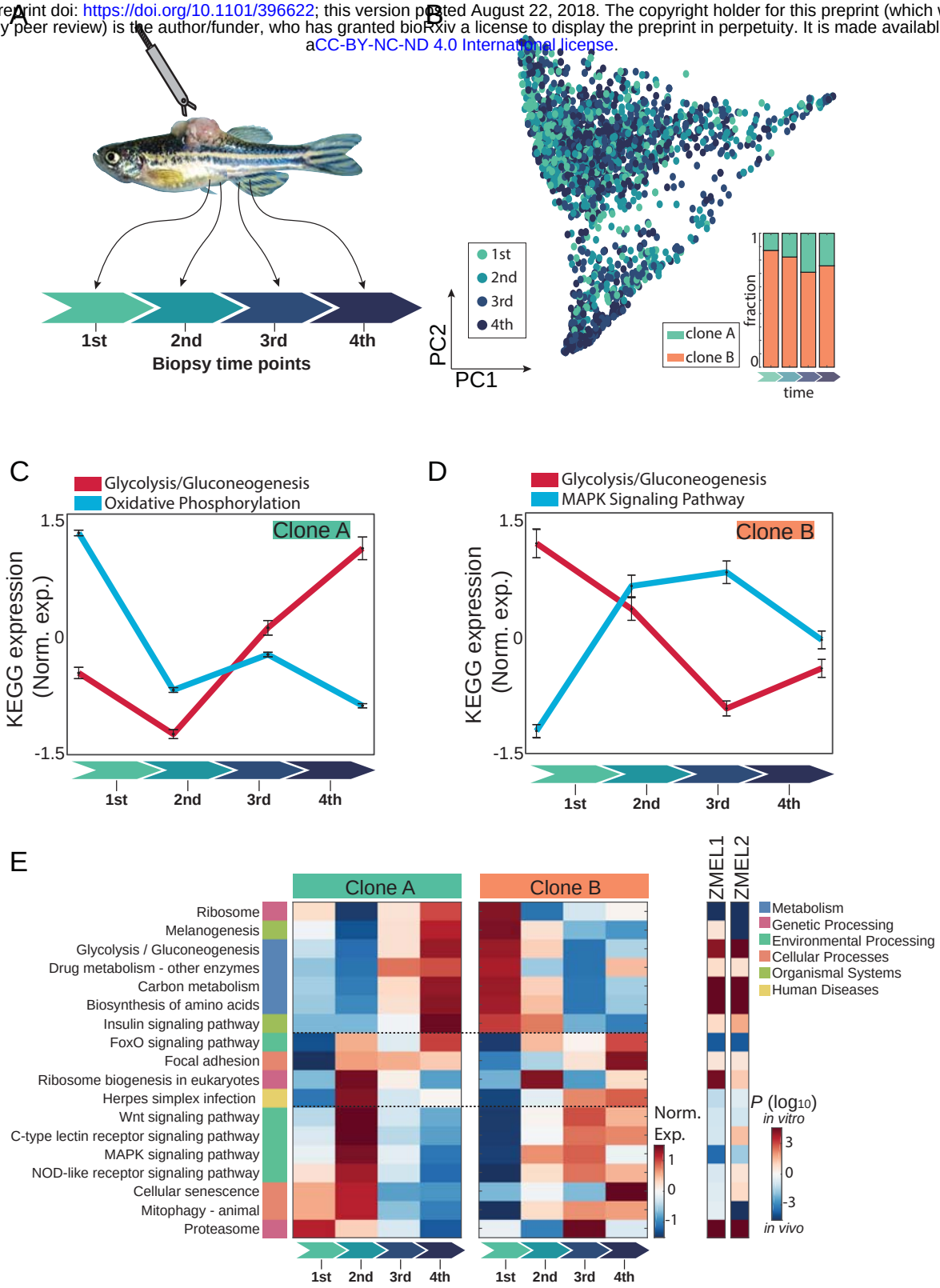


**Figure 4. The natural history of clonal gene expression during tumor progression**

(A) Four tumor biopsies were collected in one-week intervals from the same individual fish and each processed using scRNA-Seq, as described in Fig. 1A.

(B) PCA of the cancer cells, with biopsy time-point indicated by color. The inset shows the distribution of the inferred clone frequency over the biopsies.

(C-D) For the indicated KEGG pathways, the mean and standard error of the normalized expression across all cells of the indicated clone are shown.

(E) For the indicated KEGG pathways, the normalized mean expression is shown across the two clones over the four time-points. On the right, the P-values of the enrichment of expression of the genes between in vitro and in vivo culture are shown.

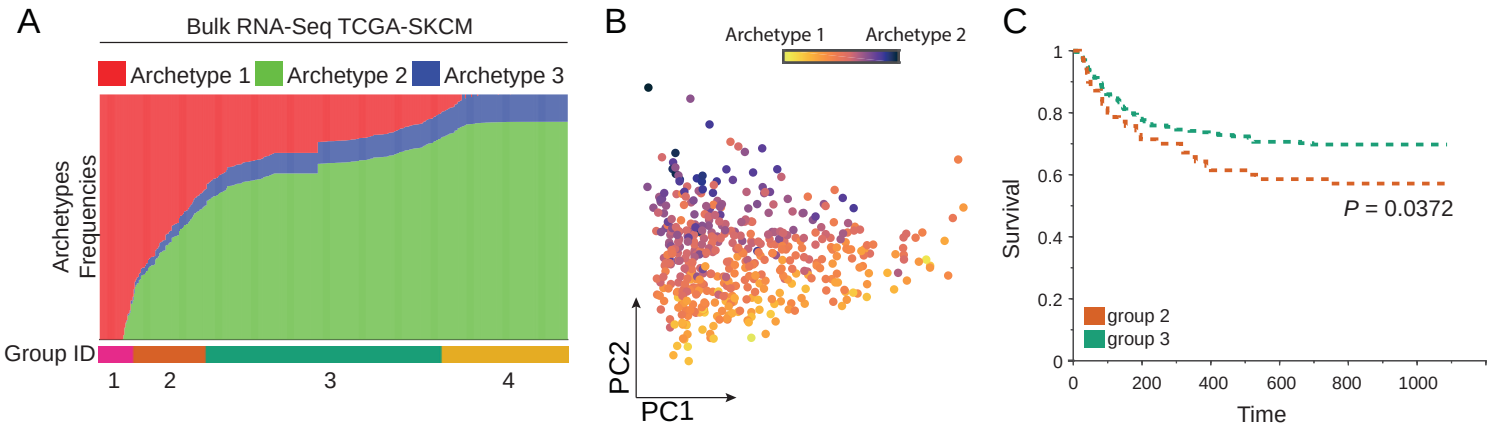


**Figure 5. Archetype frequency is correlated with clinical prognosis**

(A) Archetype frequencies in each of 472 bulk tumor from the TCGA Skin Cutaneous Melanoma project. Deconvolution was performed using expression of markers defined in Figure 2 (top panel). Samples were divided to four groups based on their archetype frequencies (lower panel).

(B) PCA of the bulk TCGA RNA-Seq data. Color indicates relative expression of Archetype 1 and 2 markers.

(C) Survival analysis of individuals with tumors high in archetype 1 (group 2) relative to those with tumors high in archetype 2 (group 3). Individuals with tumors high in Archetype 1 (group 2) showed a significantly worse prognosis compared to group 3 individuals (Kaplan-Meier estimate, P = 0.0372).


**Figure 6. The stress archetype (archetype 3) is conserved across PDAC and TNBC tumors.**

(A) PCA on cancer cells from a single PDAC tumor (Moncada et al., 2018). Color indicates normalized expression levels of JUN (left panel) and expression of the stress archetype signature (right panel).

(B) same as (A) for cells from a single TNBC tumor (C. Kim et al., 2018).