scRNA-seq of Melanoma subpopulations that rapidly escape MAPK pathway inhibition reveals upregulation of ATF4 stress signalling in escapees

BIO-463
Max Carvalho

Cancer research has been at the forefront of medical research for a while now, and new therapies are being uncovered yearly. However, one of the main limitations of cancer treatments is the tumors' capability to acquire resistance to these drugs. In this study, we will analyse the proliferation of treated and untreated melanoma cells to uncover the particularities of escapee cells. To do so, we will analyse the single-cell RNA seq data obtained from melanoma cells driven by the BRAF V600E mutation that were treated using MAPK inhibitor dabrafenib.
This pipeline effectively integrates data preprocessing, normalisation, clustering, and visualisation to analyse scRNA-seq data.

Data reference: Yang, C., Tian, C., Hoffman, T.E. et al. *Melanoma subpopulations that rapidly escape MAPK pathway inhibition incur DNA damage and rely on stress signalling*. **Nat Commun** 12, 1747 (2021).

Methods:

1. **Data**: Sparse matrices representing scRNA-seq data for dabrafenib-treated cells and untreated cells are loaded. Cell barcodes as columns and genes as rows with the values being the expression levels. Then we examine the data, confirming that it is very sparse, with most genes not expressed in most cells. And total cell expression in the form of a positively skewed gaussian distribution (Fig1a and b).

2. **Seurat setup**: Seurat objects for both treated and untreated datasets are created, incorporating basic filtering thresholds for cells and features. Then the two seurat objects are merged to facilitate the analysis and compare the two groups. We add metadata to the merged seurat object indicating the proliferative status of each cell. To finalise the setup, cells are classified into 4 groups: UT_Q, UT_P, Escapees, Non-Escapees based on their treatment status and proliferation. After this we make a violin plot of the number of genes expressed by the cells and the number of cells expressing genes above the selected thresholds (Fig2)

3. **Seurat processing**: Before conducting the analysis, a log-normalisation is applied to the RNA assay of the seurat object to standardise gene expression across cells in order to make the data suitable for comparison between groups. Now variable features can be identified across the dataset, focusing on genes that show significant variation across cells.

4. **Dimensionality reduction**: with a preprocessed seurat object, PCA can be conducted to reduce dimensionality, with an elbow plot to visualise which principal components are significant. Here we see that the first 20 components are enough to capture most of the variance as the 20th component has almost 0 standard deviation. Then we use t-SNE with the first 20 Principle Components to further reduce the dimension to 2. This t-SNE is visualised in Fig3.

5. **Differential gene expression analysis**: first the seurat object is prepared for DGEA by joining all layers and resetting the cell IDs, then the marker genes can directly be extracted. Here we wish to compare the overlap between two marker gene populations: escapees vs non-escapees and escapees vs untreated proliferating cells. First, the marker genes of these populations are extracted, then they are filtered to keep the upregulated genes (log change >0) and above a certain significance threshold (adjusted p-value < 0.005 here).

6. **Biological significance**: finally the the two populations of differentially expressed genes can be compared by making a venn diagram (Fig 3.b). Then we can compare the overlapping populations to the list of genes targeted by ATF4 to see to what extent the escapee populations rely on stress signalling promoted by ATF4.

Discussion:

Sc-RNA-seq data can be tricky to work with as it is very sparse and there is only one reading per cell, thus creating some error. However, after some filtering and preprocessing using seurat, we can plot the t-distributed stochastic neighbor embedding (t-SNE) of our treated and untreated sc-RNA data onto a two dimensional plane (Fig 3a). As we can see, the escapees are not very numerous, at the edge of the treated population cluster and pointing towards the untreated population. Thus we can infer that the escapee population can still be distinguished from the rest of the treated population solely looking at their gene expression profile.

In addition to this, we can identify some new genes and pathways involved in the escape from the drug by conducting a differential gene expression analysis of the treated escapees vs non escapees and treated escapees vs untreated proliferating cells. The intersection of these differentially expressed populations yielded 337 common upregulated genes* with 7 ATF4 target genes in the top 40 (Fig. 4b). ATF4 is expressed when the cell is under stress, leading to the upregulation of the stress response genes. Thus the ATF4 pathway to escape dabrafenib can be targeted in future treatments to break the escape of melanoma cells and prevent the cancer from persisting.

In summary, we have seen that escapees tend to avoid the drug-induced BRAF inhibition by, among other things, upregulating the ATF4 pathway. Thus we could envision a drug that targets this pathway to suppress persistent melanomas.

* The 40 common upregulated genes obtained by the writers of the paper were also present in the 337 genes I have found, however, despite similar processing, I have slightly different results.

Figures:

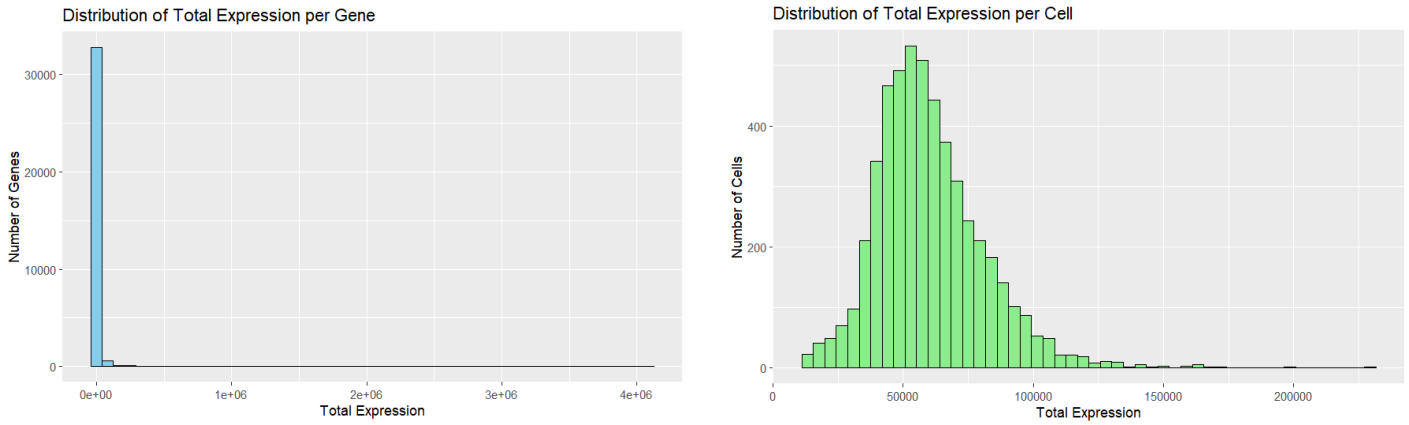

Fig1. First look at the sequencing data. a looks at the distribution of total expression per gene for this sparse data (light blue). b shows the distribution of total gene expression per cell (green).

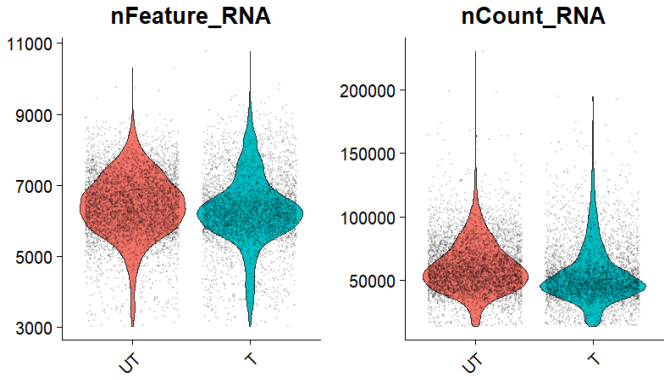

Fig2. Violin plot of the rows (genes) and columns (cell barcodes) of our T and UT matrices after filtering. On the left, the violin plots show the number of genes that the cells express and on the right, the number of cells expressing over 800 genes. In red we have the untreated cells and in blue we have the dabrefenib treated cells

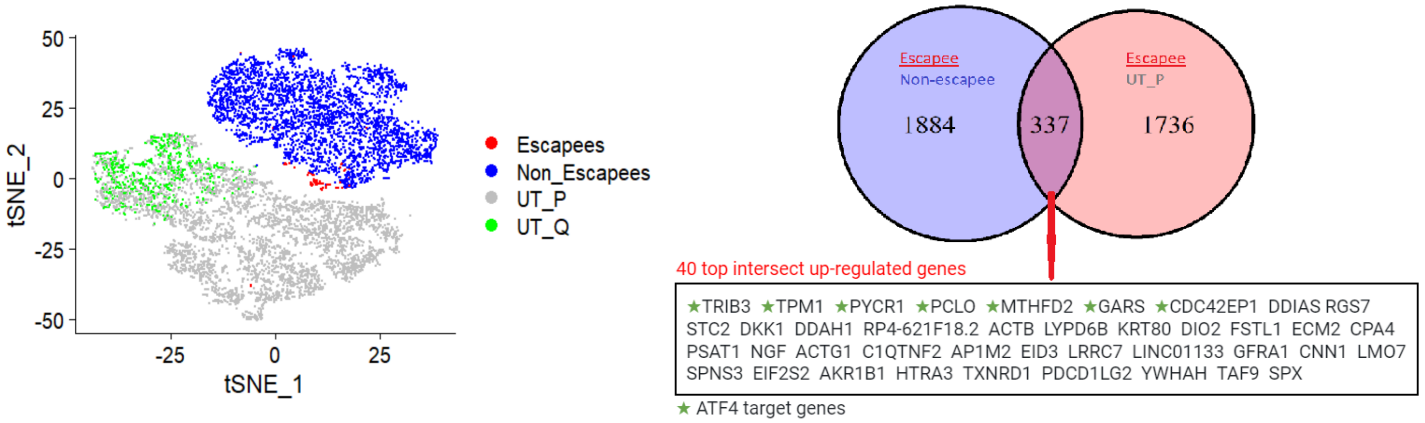

Fig. 3 scRNA-seq reveals upregulation of ATF4 stress signalling in escapees. a Co-visualization of untreated (bottom left) and treated (top right) scRNA-seq datasets on a single t-SNE plot, showing escapees in orange and non-escapees in blue . Untreated proliferative cells (UT-P) and quiescent cells (UT-Q) were coloured grey and green, respectively. Escapees can be identified as a small orange peninsula in the treated condition. b Venn diagram of differentially expressed genes as described in the text. Genes labelled with a green star are ATF4 target genes.