1) Viral titration of target cells- aim for 2-10% transduction efficiency to limit multiple integrations per cell.

2) Sort barcode positive cells 2-3 days following transduction and plate a certain amount that will yield around 2000-5000 unique barcodes/clones (based on optimisation). Limiting this starting number of barcodes is important (explained in point 7). For reference, we find that plating 50,000 of our AML cells leads to around 5000 unique barcodes. This allows for some cell death following the sort, plus the expansion period prior to the sort.

3) Expand the cells to allow for multiple population doublings. We expand ours for 5-7 days. You need enough cells to allow for all ~5,000 of your starting barcodes to enter all arms of experiments (e.g vehicle vs. drug, across different mice). You also want representation of each barcode (i.e At least 20-fold representation (20 cells) of each barcode to account for stochastic loss during replating for in vitro experiments, engraftment during transplant experiments). We also set up technical replicates for every experiment, to ask whether a certain cellular behaviour is intrinsic or stochastic (e.g the same or different barcodes growing at the end).

4) Following expansion period, take a sample for single cell RNA sequencing on the day you start experiments for your “baseline” sample and carefully cryopreserve ready for future analysis. This is the most important sample. Also take 500,000 cells for population-based DNA barcoding.

5) Apply desired evolutionary pressure to your cells.

6) At the experimental endpoint, cryopreserve samples for single cell RNA seq and lyse samples for barcode PCR.
7) **Important:** on any given scRNA-seq run, you are limited by how many cells you can sample, both practically and cost wise. We generally run 20-25,000 cells at Baseline on the 10X. The reason we aim to only start with 2000-5000 barcodes is so that when we sample our 20,000 cells at baseline, most barcodes are present in at least 10 cells. This allows you to confidently identify transcriptional signatures of interest. If you started with 50,000 barcodes, it is cost prohibitive to sample all of these cells on the 10X (you would need to sample 500,000). Therefore, if one of these 50,000 barcodes dominated at the endpoint, it might not be represented at the start, and you would never be able to assign a transcriptional programme to that clone.