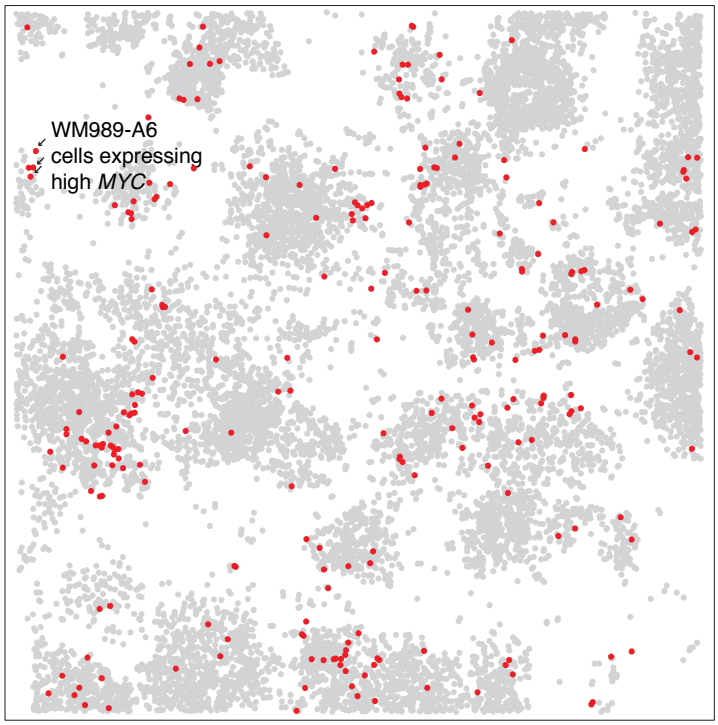
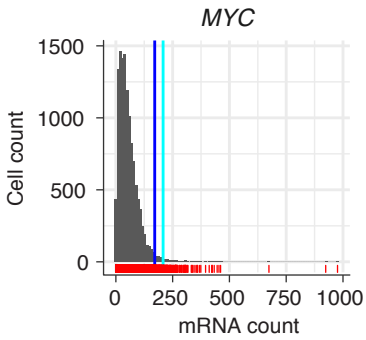
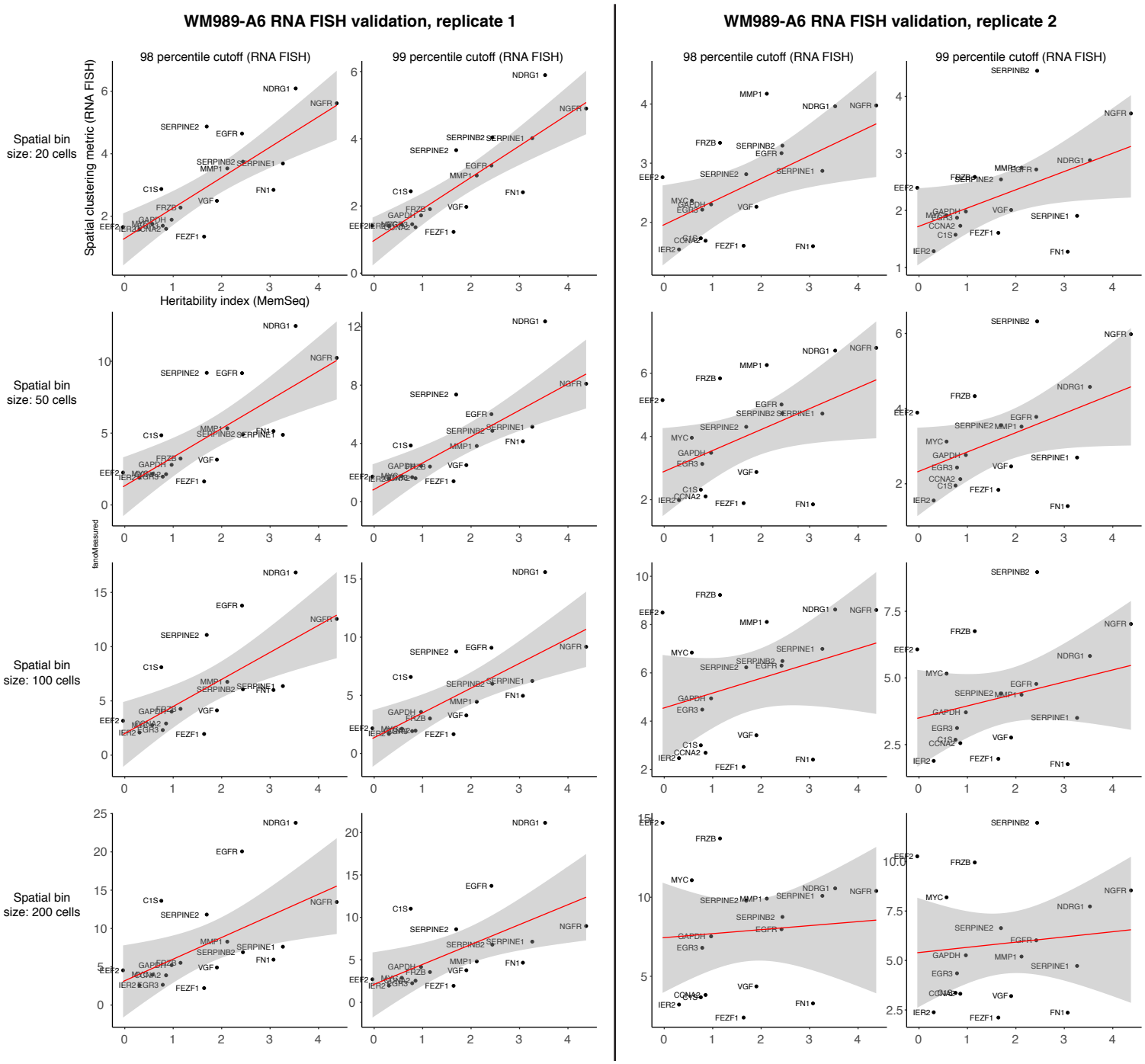
Supplementary Figure 1





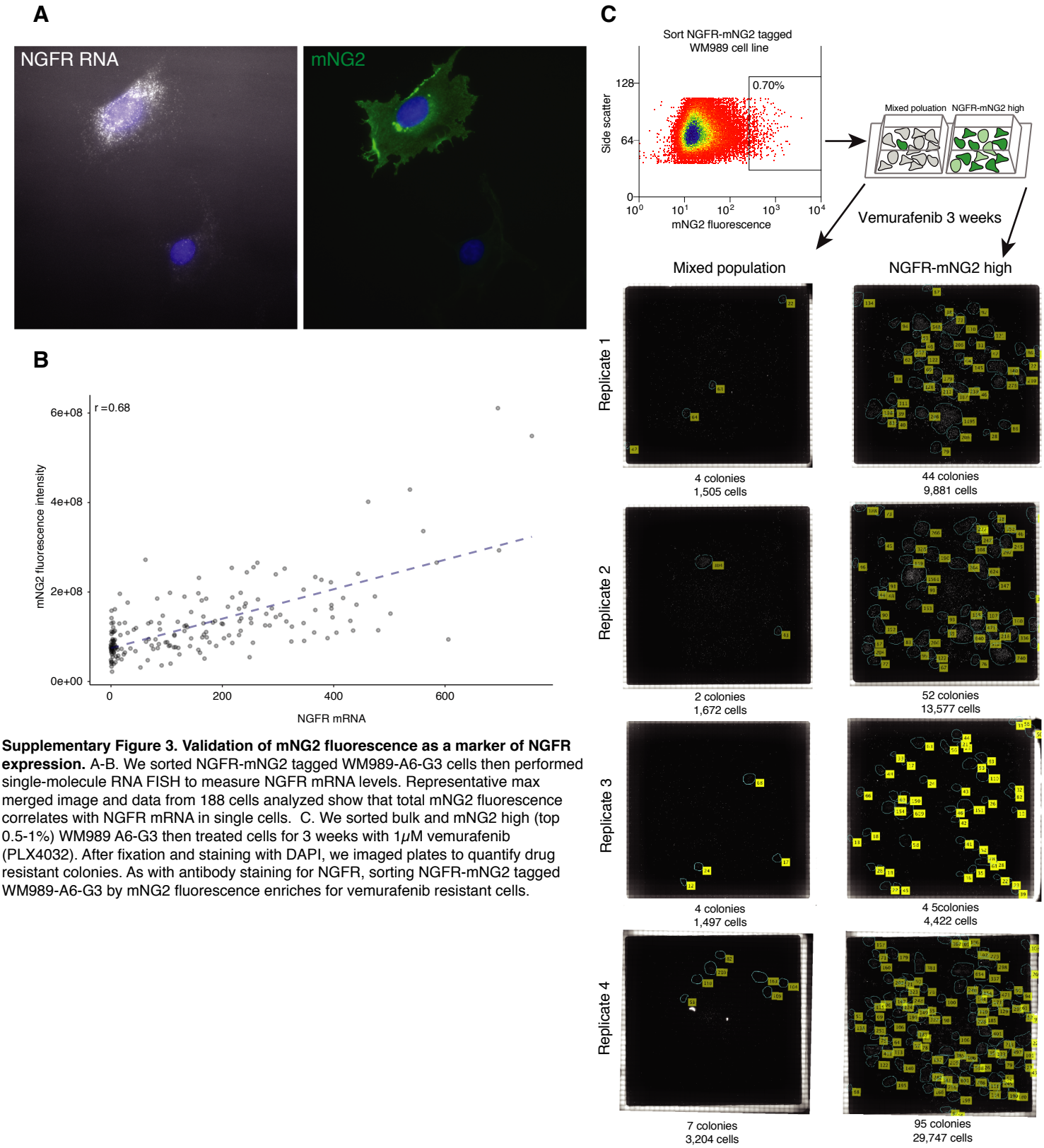
Supplementary Figure 1. MYC expresses in a highly variable but not heritable fashion. A. We measured expression of MYC in cells grown on culture dishes for 2 weeks in WM989-A6 cells, and labeled the top 2% of cells red (rest of cells labeled gray). The degree of spatial clustering was minimal, reflecting the low heritability of the high MYC expression level cellular state. B. Histogram of MYC mRNA levels across individual cells, showing the large amount of variability in MYC expression. The rightmost cutoff marks the top 1% of cells, the left cutoff marks the top 2%.

Supplementary Figure 2



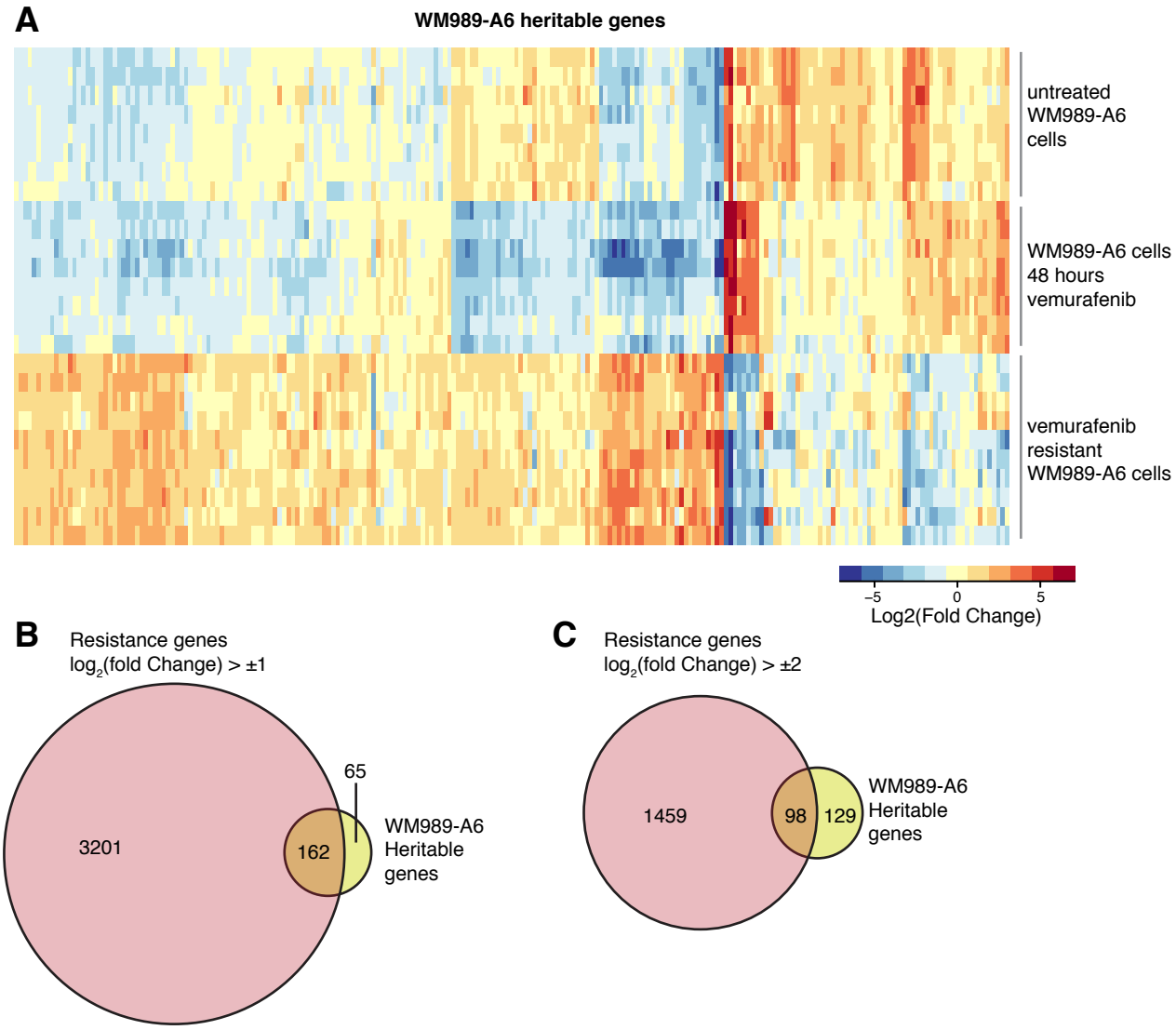
Supplementary Figure 2. Spatial single cell RNA FISH analysis validation of MemorySeq. We performed analysis of single molecule RNA FISH for the indicated genes as per Fig. 2. Shown are two biological replicates (left and right). Within each replicate, we analyzed different cutoffs for cells to be considered to be in the high expression level state (either the top 1 percent of cells or the top 2 percent of cells). We also increased the bin size for each neighborhood for analysis, which is quantified as the number of neighboring cells included in each bin. We quantified the skewness across MemorySeq clones vs. the Spatial clustering metric, which is the Fano factor (variance/mean) in the number of positive cells per bin across all bins of indicated size. We found the correspondence decreased once the bin size reached 100, and that using the top 1 percent of cells also led to a tighter correlation. The decreased correlation at higher bin size likely reflects the fact that at larger bin sizes, more clusters would merge together, thus decreasing apparent Fano factor across bins. (To see this, take the limiting case of just two bins, in which the variance between the two would be relatively minimal.)

Supplementary Figure 3



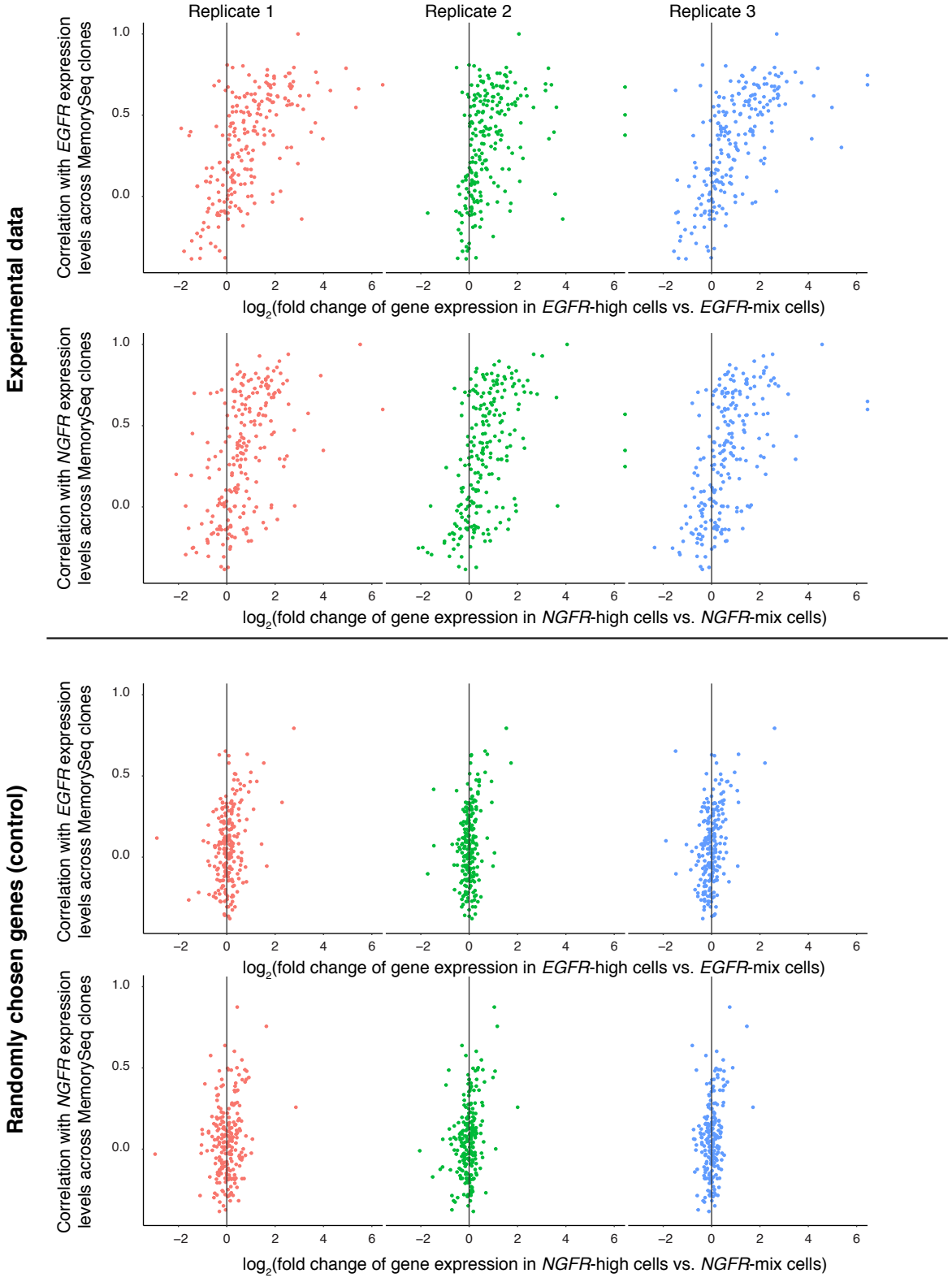
Supplementary Figure 3. Validation of mNG2 fluorescence as a marker of NGFR expression. A-B. We sorted NGFR-mNG2 tagged WM989-A6-G3 cells then performed single-molecule RNA FISH to measure NGFR mRNA levels. Representative max merged image and data from 188 cells analyzed show that total mNG2 fluorescence correlates with NGFR mRNA in single cells. C. We sorted bulk and mNG2 high (top 0.5-1%) WM989 A6-G3 then treated cells for 3 weeks with 1μM vemurafenib (PLX4032). After fixation and staining with DAPI, we imaged plates to quantify drug resistant colonies. As with antibody staining for NGFR, sorting NGFR-mNG2 tagged WM989-A6-G3 by mNG2 fluorescence enriches for vemurafenib resistant cells.

Supplementary Figure 4



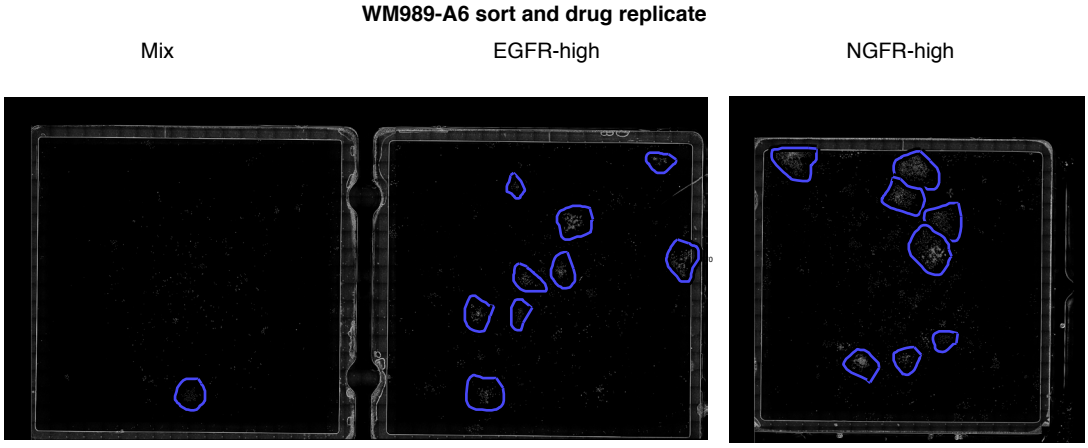
Supplementary Figure 4. Genes identified by MemorySeq and their expression levels in stably resistant melanoma cells. We wondered how the expression levels of genes identified by MemorySeq changed during the process of acquiring resistance (WM989-A6 cells) using data from Shaffer et al. 2017. A. Each column is a gene identified by MemorySeq as being heritable. Each row is a biological replicate, from bulk untreated cells, cells treated with vemurafenib for 48 hours (to account for drug-response effects) and the stably resistant cell lines. B,C. Venn diagram showing which genes were associated with resistance together with being identified as heritable by MemorySeq. We ran the analysis at two different thresholds for fold change. Either way, we found that while a large fraction of heritable genes were also resistance-associated genes (over 70% for a two-fold change cutoff), only a very small proportion of total resistance genes were identified by MemorySeq as being heritable, in agreement with the conclusions of Shaffer et al. 2017.

Supplementary Figure 5



Supplementary Figure 5. Comparison of correlation to EGFR/NGFR expression across MemorySeq clones to expression levels in EGFR/NGFR-high WM989-A6 cells. We wondered whether correlations between clones corresponded to correlations at the single cell level genome-wide. We verified this by sorting cells by EGFR or NGFR targeting antibody staining (see methods), isolating the highest expressing cells and then performing RNA-seq on both the high and total population (three biological replicates shown). Focusing on genes that were determined to be high-memory by MemorySeq, we then determined the degree of correlation with EGFR or NGFR for each of these genes across the MemorySeq clones. We found a general positive correspondence between fold-change and correlation, showing that correlations measured by MemorySeq do correspond to correlations at the single cell level genome-wide. In the bottom two rows, we performed the same analysis, but this time with randomly selected genes (same number as in top two rows), showing that the correspondences shown in the top two rows are not spurious.

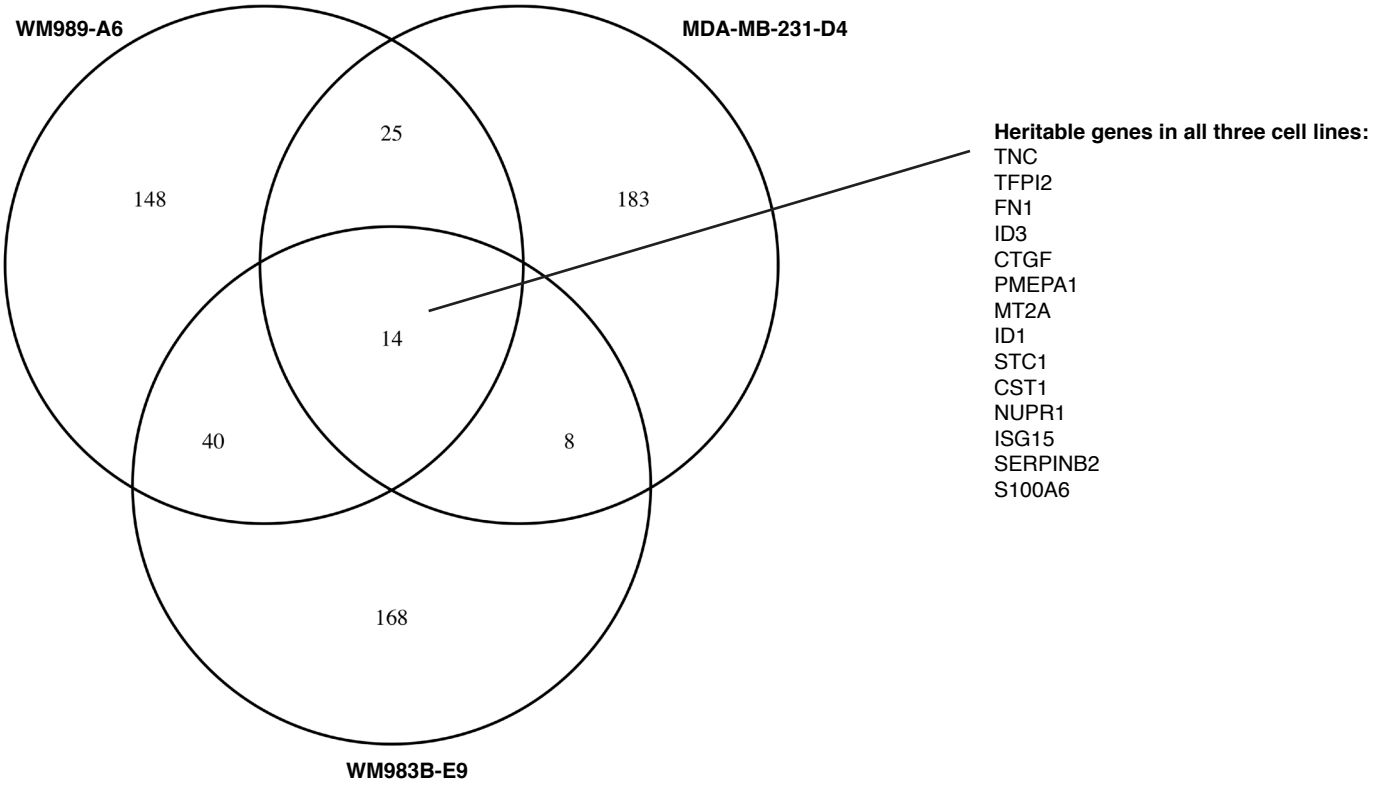
Supplementary Figure 6



Supplementary Figure 6. Replicate of NGFR sorted cells grown in trametinib. Biological replicate of the data shown in Fig. 3B. We used fluorescence activated cell sorting to isolate the NGFR-High subpopulation of WM989-A6 cells, cultured for 8-16 hours and then subjected to trametinib treatment at 10nM (MEK inhibitor) for 3 weeks. Image shows number of resistant colonies (circled) along with number of cells within the resistant colony as indicated.

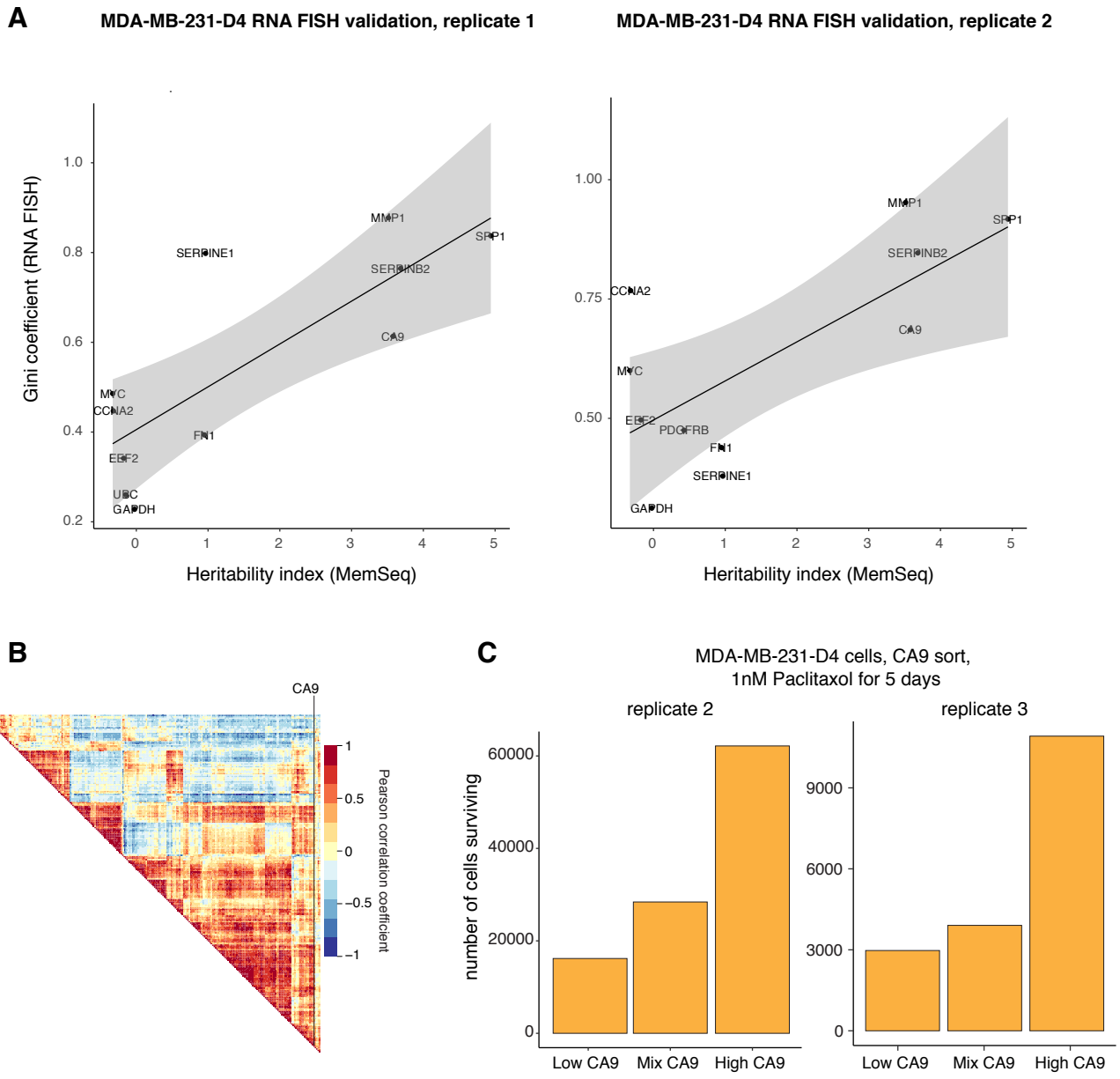
Supplementary Figure 7

Overlapping heritable gene sets across WM989-A6, MDA-MB-231-D4, and WM983B-E9:



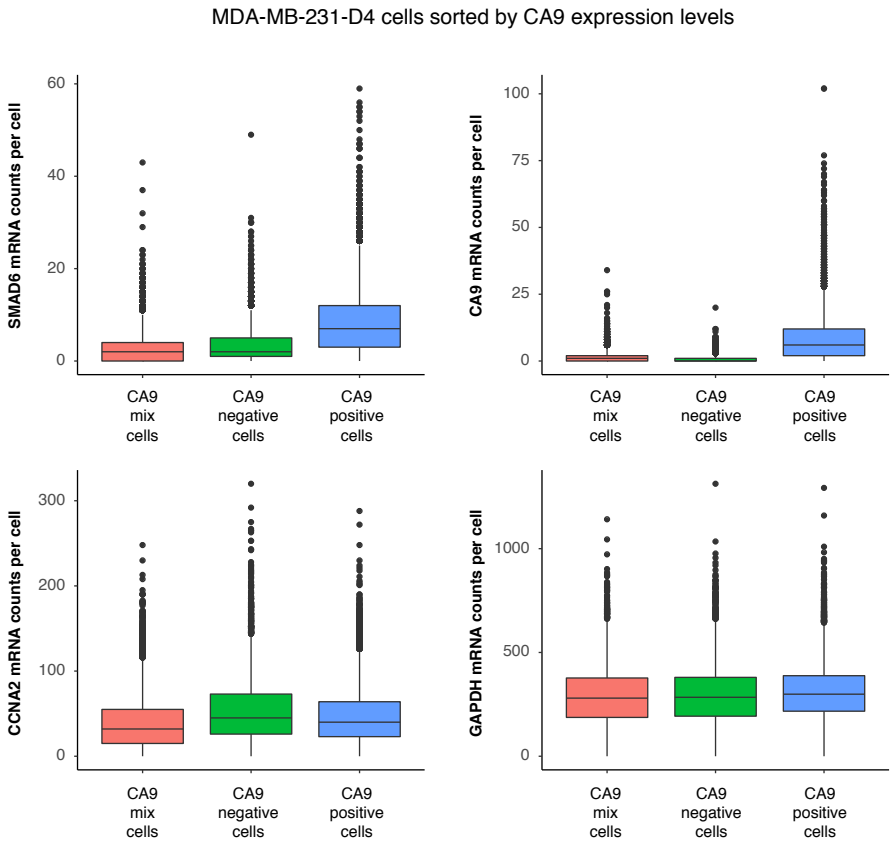
Supplementary Figure 7. Overlap of MemorySeq hits across cell lines. We wondered if genes identified by MemorySeq as heritable were the same across multiple cell lines. We found that most heritable genes were distinct between cell lines, perhaps reflecting the potential for cell-type specific heritable rare cell expression programs. There was a set of 14 genes (listed in figure) that were identified as heritable in all 3 cell lines.

Supplementary Figure 8



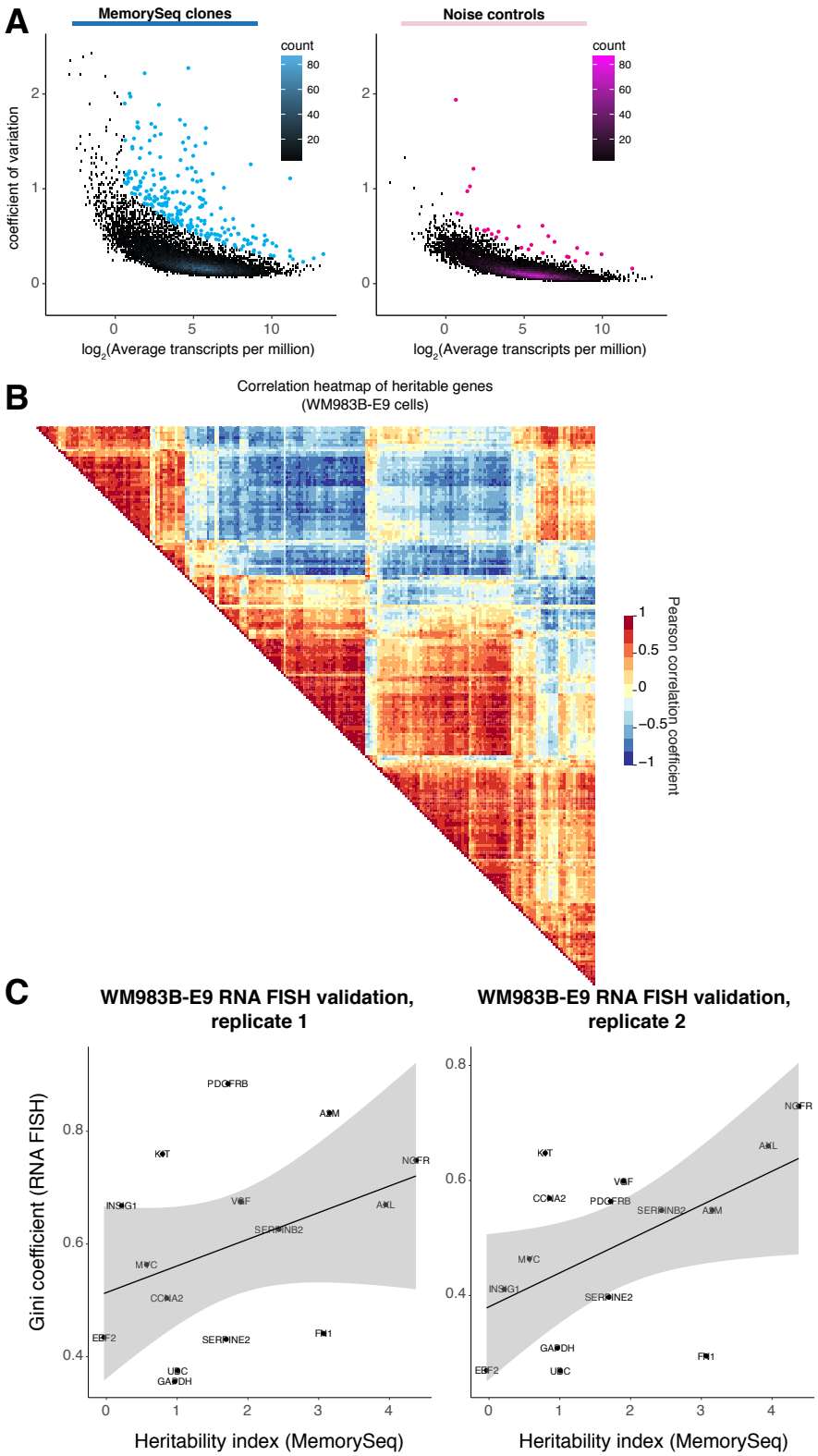
Supplementary Figure 8. Comparison of heritability by MemorySeq to Gini coefficient for MDA-MB-231-D4 cells. We plotted heritability index from MemorySeq (skewness across individual MemorySeq clones) vs. Gini coefficient as computed from single molecule RNA FISH across two biological replicates. Overall, we found a general correspondence between the two variables, as was seen for WM989-A6 cells, suggesting that MemorySeq also identifies genes that express in rare cells.

Supplementary Figure 9



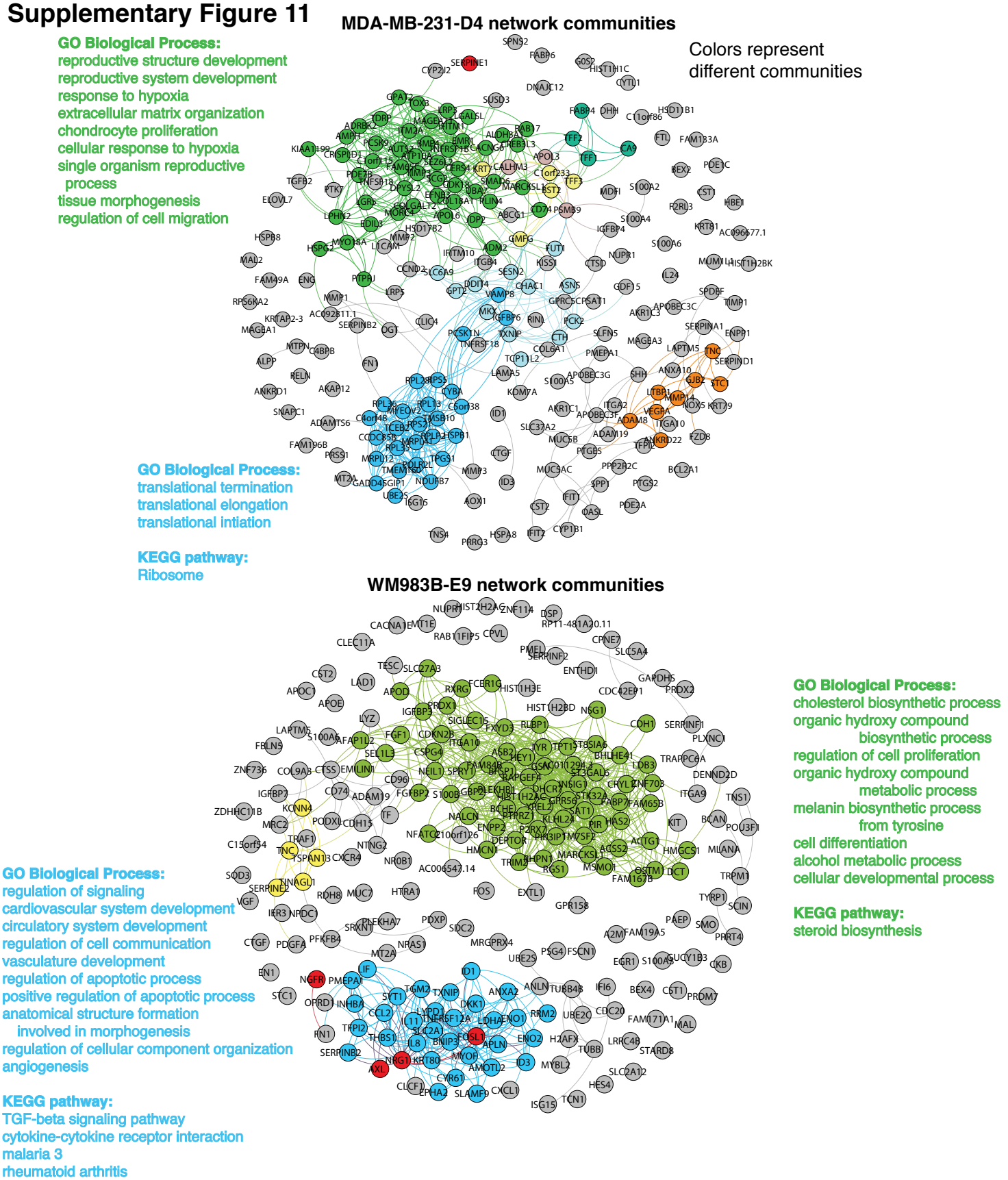
Supplementary Figure 9. Single cell RNA FISH confirms the accuracy of sorting MDA-MB-231-D4 cells by CA9. We sorted CA9-stained cells into high and low expressing subpopulations (as determined by antibody staining) as well as the unsorted mixed population. We then performed single cell RNA FISH for SMAD6, CA9, CCNA2, and GAPDH on these subpopulations. We found that the levels of CA9 mRNA were higher in the high subpopulation as expected, as was SMAD6, which was also expected based on their correlation as revealed by MemorySeq. Both housekeeping genes CCNA2 and GAPDH showed minimal differences between subpopulations.

Supplementary Figure 10



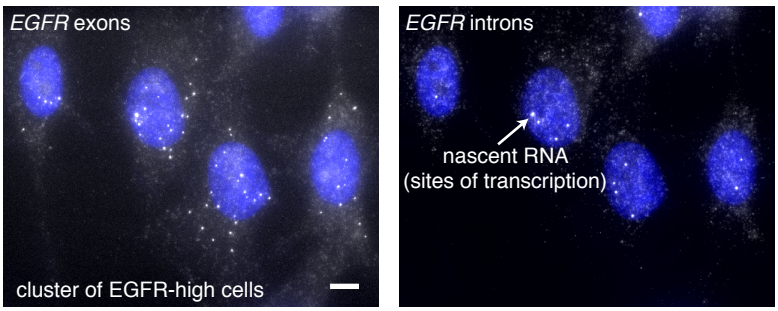
Supplementary Figure 10. Analysis of MemorySeq data for WM983B-E9 melanoma cells. A. We performed MemorySeq on the melanoma line WM983B-E9. Scatters of average transcript abundance from MemorySeq vs. the coefficient of variation across individual MemorySeq clones (left), with a technical noise control as described for WM989-A6 (right). B. Clusters of co-fluctuating genes in WM983B-E9 in genes identified by Memory-Seq as showing a high degree of heritability. C. We plotted heritability index from MemorySeq (skewness across individual MemorySeq clones) vs. Gini coefficient as computed from single molecule RNA FISH across two biological replicates. Overall, we found a general correspondence between the two variables, as was seen for WM983B-E9 cells, suggesting that MemorySeq also identifies genes that express in rare cells.

Supplementary Figure 11



Supplementary Figure 11. Community identification in MemorySeq hits for MDA-MB-231-D4 and WM983B. We performed community identification on hits detected by MemorySeq (top 2% of genes, TPM of 1.5 or greater) for MDA-MB-231-D4 and WM983B-E9 cells. Cutoffs were chosen to arrive at biologically reasonable gene sets. We performed KEGG and GO biological process analysis on genes identified as belonging to particularly communities (limited to communities that had enough genes for such an analysis to be meaningful).

Supplementary Figure 12



Supplementary Figure 12. Detection of nascent transcripts show ongoing transcription in EGFR-High expressing cells. We wondered whether EGFR-high cells (WM989-A6) stemmed from a short burst of transcription followed by dilution amongst progeny cells or from sustained transcription in this subpopulation of cells. We distinguished these possibilities by measuring nascent active transcription in single cells by targeting the intronic region of EGFR with single molecule RNA FISH probes in a separate color from the EGFR exons. We saw that in patches of EGFR-high cells, the individual cells typically also showed EGFR introns, showing that EGFR exhibited sustained transcription over multiple divisions.