

## Library Design and HT Sequencing are Crucial to Obtain Quality Results from Pooled shRNA Library Screening

### Overview

Previously, we attempted a genome-wide RNAi-knockdown screening to identify signaling molecules in FAS-induced HeLa cell model with a commercially-available lentiviral-based pooled shRNA library constructed with 200,000 shRNA. The results were characterized by a high percentage of both false positives and false negatives results with only about 4% of putative positives correlating to modulators in the known apoptotic signal transduction pathways. These findings indicated a potentially useful but problematic screening. In an effort to optimize the screening process and assess the effectiveness of pooled shRNA libraries, we designed an improved Collecta shRNA Lentiviral Library to perform a similar screen

### Description of Screening Problem—Few True Positives, Lots of False Positives

Gene knock down screening with a commercially-available pooled lentiviral-based shRNA library constructed with 200,000 shRNAs was undertaken to identify genes modulating FAS-induced apoptosis of HeLa D98/AH2 cells. The library contained 4-5 shRNA inserts targeting each of the 47,000 transcripts represented on the Affymetrix Human Genome U133+2 GeneChip array. Each of the shRNA sequences were selected to be compatible with one of the probe sequences on the Genome U133+2 array with the purpose of using hybridization to this array as a means to detect shRNA sequences expressed by cell populations.

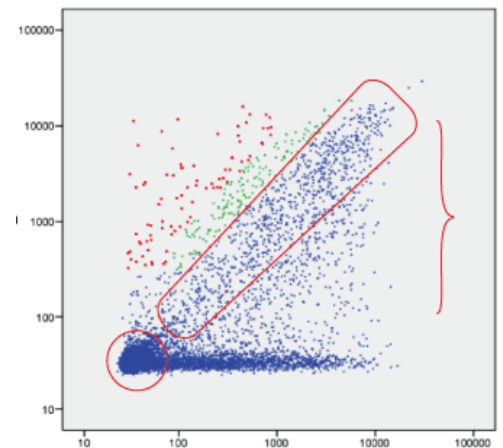
In duplicate, HeLa cells ( $1 \times 10^8$  cells) were transduced with the packaged 200K shRNA library (MOI = 0.3), then treated with anti-Fas receptor antibodies that activate the FAS-pathway and induce apoptosis, and grown for seven days. In control cells the FAS pathway was not induced. The entire pool of siRNA inserts was amplified by two rounds of PCR using loop-specific and vector-specific primers from 5 ug of cDNA from control (untreated) and Fas-treated cells. Amplified DNA was hybridized to the Affymetrix HG U133+2 arrays (Figure on right).

The selection protocol was efficient and resulted in about 95% cell death. However, despite this stringency, we identified approximately 2,500 shRNA sequences that were significantly enriched in the treated vs. control populations. However, less than 50% of these overlapped between the two independent experiments, indicating an abundance of false positives. Among the positives, there were 20 apoptosis-related gene candidates and three well-known apoptosis-related genes (bax, bcl-x<sub>L</sub>, and FasL) which indicate the potential of the approach. However, the excessively high number of false positives (background) makes it impossible to select real positives based on just the experimental results. In addition, several well-known genes involved in activation of Fas-induced apoptosis, such as CASP8, BID, and DR4 were not revealed which indicates a high percentage of false negatives.

### Analysis of 200K Screening Results—Library Problem

The results of the screening and hybridization with the shRNA library constructed with 200,000 shRNAs generated too many false positives to confidently identify shRNA sequences enriched in the target population. To address this, we decided to repeat the RNAi screen with a lower complexity shRNA library and looked at using a more quantitative method, HT sequencing,

FAS-Treated Cells



Transduced Control Cells

### Results of Screen with 200K shRNA Library

The shRNA sequences from amplified genomic DNA isolated from both FAS-treated and untreated HeLa cells was hybridized to the Affymetrix HG U133+2. The plot above shows the intensity (which should correlate to the quantity) of each hybridized shRNA sequence for the control cells (x-axis) and treated cells (y-axis). Most of the cells apoptose with treatment so the shRNA sequences are not seen in the treated population and appear at the bottom of the graph. shRNA sequences present at equivalent concentrations in both populations fall on the diagonal. Sequences above the diagonal were found at enriched concentrations in the treated vs. control populations, indicating that suppression of these genes confers resistance to FAS-activated apoptosis. Approximately 2,500 shRNAs were seen at 2-fold or higher concentrations in the treated cells (outside of the red delineated range on the diagonal). Also, a large number of shRNA did not show up in either population (red circle at the base of both axes).

to assess the abundance of shRNA in the sample population. We found that we could use high-throughput (HT) sequencing to address both of these points.

First, we looked at the library. When the library itself was hybridized to the Affymetrix HG U133+2 arrays, only approximately half the shRNA sequences were detectable. This may be a result of weak hybridization for this sequence or the sequence is missing from the library. To assess this, we attempted to directly amplify by PCR 40 shRNA sequences that were not detected in the packaged library with the array. 36 of these were not detected, indicating that they are missing from the library. Thus, although the library was constructed with 200,000 shRNAs, it seems that approximately 40% of the shRNA are missing. This indicates that there is probably also a large variability in the remaining sequences with some being very abundant and others in low numbers.

In looking at the array hybridization data, two other problems also appeared to occur. (1) There was some cross hybridization of signals which contributed to the false positive background, and (2) there was a limited dynamic range from the lowest to the highest enrichment levels of approximately 100-fold. Below this level, there may be some hybridization but the shRNA sequence was undetectable and, at the high-end, the signals plateau so any enrichment in shRNA between control and treated populations is undetectable.

### New Library to Improve Screening Effectiveness

To address the shRNA representation, signal cross hybridization, and the low dynamic range of array readouts, we constructed a new library where each shRNA insert also contained a uniquely identifiable sequence (a “barcode”) that was compatible with Illumina GAIIx sequencing technology so it could be detected by HT sequencing. We also made the library with a pool of just 38,000 shRNA designed to target 8,000 well-annotated genes. This smaller size library could be more practical to screen exhaustively and ensure that all shRNA are interrogated in the assay.

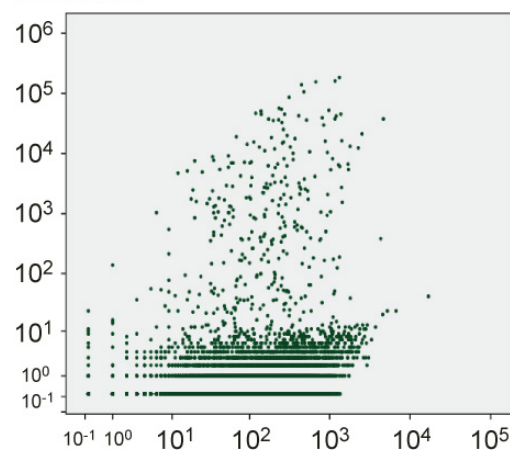
Analysis of the Collecta 38K shRNA Library by HT sequencing demonstrated that, after ligation and packaging, all sequences were present. Further, that at least 95% of the shRNA fell within a range where there was less than a 10-fold difference from the least to the most abundant (see figure)—although at least 5 orders of magnitude in range were easily discerned.

### Results and Conclusion with Improved Quality Library

Screening with the Collecta 38K shRNA Library found approximately 350 shRNAs enriched in the treated population. Of these, 150 (or 43%) were known modulators of apoptosis and fell into the expected pathways related to NF- $\kappa$ B and p53 pathways.

The results demonstrate that careful RNAi knockdown screening of a targeted set of genes with a well designed pooled library can provide a comprehensive interrogation of the genes responsible for a particular biological response, in this case, FAS-induced apoptosis. The marked difference in the results of the Collecta 38K shRNA screen compared with the 200K shRNA emphasize the importance of using representation and complexity of the library in ensuring a screening produces usable results.

FAS-Treated Cells



Transduced Control Cells

**Results of Screen with Collecta's 38K shRNA Library** Bar-codes corresponding to individual shRNA sequences from amplified genomic DNA isolated from both treated and untreated populations were counted by HT sequencing using a Illumina system. The plot above shows the abundance of of each shRNA in each population (x-axis for controls and y-axis for treated cells). As with the previous screen, the non-functional shRNA sequences fall to the bottom of the graph (not represented in the treated population). As is clear, the number of shRNA seen in the treated population is much smaller than with the previous screen with the 200K shRNA library. Approximately 350 shRNAs are seen to be enriched in the treated population.