Additional Figure Legends

Figure S1 – TRCBC feature offset design for testing signal-to-noise and specificity of shRNA barcode signals. (a) Schematic of feature offset design for TRCBC array. The TRCBC array includes three perfect match (PM) features and a single mismatch (MM) feature corresponding to ~110,000 human and mouse TRC shRNA sequences. PM features 0 and 2 are offset from PM feature 1 by either a 3’C or 5’ G, respectively. A reference plasmid pool and dilution series plasmid pool consisting of the same ~45,000 shRNA barcodes were used to generate labeled probe and hybridized separately to TRCBC microarrays. The dilution series plasmid pool contained four subpools of shRNA barcodes that were present at even amounts, diluted 4-fold, 16-fold and 64-fold. Signal distributions were generated for each of the four subpools by subtracting signal intensities from shRNA barcodes in the dilution series subpool from the corresponding shRNA barcodes in the reference subpools and generating 1x, 4x, 16x and 64x distributions for each of PM0 (b), PM1 (c) and PM2 (d) features. A deconvolution algorithm was designed to measure the fraction shRNAs incorrectly assigned to the proper dilution series pool based on the above data and the results of the assignments are shown on the right of each set of PM distributions, highlighting the observation that more shRNA barcodes are assigned to the correct dilution series pool when considering PM2 feature design.

Figure S2 – Probe amount and hybridization/wash temperature determination. (a) Distribution plots of signal intensity from hybridization of 1µg, 2µg, and 3µg of probe from 78k human shRNA pool. Features on the microarray corresponding to shRNA probe sets in the 78k pool are in solid lines while dashed lines denote remaining shRNA features on the microarray not corresponding to sequences in the probe pool. (b) Distribution plots of microarray signal intensity for shRNA probes included in (solid) or excluded from (dashed) the human 78k shRNA pool. 2µg of probe was applied to the GMAP with varying hybridization and wash temperatures as indicated in the legend.

Figure S3 – Readout of high complexity shRNA pools. (a) Binned histograms, and (b) cumulative distributions for all 248,049 shRNA features on the GMAP microarray following hybridization of probe generated from the human 78k plasmid shRNA pool or the mouse 78k plasmid shRNA pool. In figures (a) and (b), shRNA features that correspond to shRNA barcodes in the plasmid pools are shown as dark red (Human 78k) or dark blue (Mouse 78k) lines while the features that do not correspond to shRNA sequences in the pools (78k-excluded) are indicated as light red or blue lines, respectively. (c) Minimum signal for inclusion in analysis is calculated for the human 78k Even pool as mean of background plus 1.96 standard deviations (95% confidence limit).

Figure S4 – Comparison of TRCBC and GMAP chips. Scatter plot depicting the signal intensities for probe generated from genomic DNA from breast cancer cells (BT-474) infected with the human 54K shRNA pool hybridized to TRCBC and GMAP arrays (Pearson’s correlation coefficient R=0.96). The amount of probe hybridized to the TRCBC array was 2µg and the GMAP array was 1.35µg.

Figure S5 – Distribution plots for GMAP and deep sequencing data. Distributions for (a) deep sequencing read number, and (b) GMAP signal intensity of the undiluted shRNAs present in the Even, 4x, 16x and 64x pools.
Figure S6 – HSPI features on the GMAP. (a) Images of GMAP chips following hybridization of probe generated from the human 78k plasmid pool in the absence of spike-in controls (left image) and the presence of 12 spike in probes (5nM) selected to represent the range of shRNA Tm’s. (b) Quantification of signal for 12 selected spike-in features averaged over 25 HSPI replicates. Final concentrations of each HSPI oligonucleotide tested in the hybridization mixtures were 0pM, 50pM, 500pM and 5nM.

Figure S7 – Comparison of 90k Plasmid and Genomic probe data. Average shRNA feature intensity data from triplicate template amplifications and GMAP hybridizations for probe generated from shRNA plasmid template and shRNA-containing A549 genomic DNA is highly correlated (Pearson correlation coefficient R=0.971).

Figure S8 – Distribution plots of data from triplicate GMAP hybridizations using probe containing the human 78k Even pool. Plots include raw data, data after GC-background (GCbg) correction, and data post-Cyclic Loess normalization.
Figure S1

(a) Probe sequence

PM Feature 0  PM Feature 1  PM Feature 2  MM Feature

shRNA 21mer

(b) Deconvolution Errors

Dilution Pool – Reference Pool

Density

PM Feature 0

Deconvolution Errors

Fraction Incorrectly Assigned

0.0 0.1 0.2 0.3 0.4 0.5

1X 0.25X 0.06X 0.016X

(c) Deconvolution Errors

PM Feature 1

Dilution Pool – Reference Pool

Density

PM Feature 1

Deconvolution Errors

Fraction Incorrectly Assigned

0.0 0.1 0.2 0.3 0.4 0.5

1X 0.25X 0.06X 0.016X

(d) Deconvolution Errors

PM Feature 2

Dilution Pool – Reference Pool

Density

PM Feature 2

Deconvolution Errors

Fraction Incorrectly Assigned

0.0 0.1 0.2 0.3 0.4 0.5

1X 0.25X 0.06X 0.016X
Figure S2

(a) Density vs. log₂ intensity for 1 μg, 2 μg, and 3 μg.

(b) Density vs. log₂ intensity for various conditions: 40H 30W, 40H 35W, 45H 30W, 45H 35W.
Figure S5

(a) Density vs. Sequence reads (log$_2$)

(b) Density vs. Array signal (log$_2$)
Figure S6

(a) 

(b) 

GMAP Signal (log2)

Spike-In #

+12 spike ins
Figure S8

- Raw
- GCbg-corrected
- GCbg-corrected + Loess-normalized

Pool-excluded features
Human 78k pool features