Supplement

Evaluation of current methods used to analyze the expression profiles of ABC transporters yields an improved drug-discovery database

Josiah N. Orina[1], Anna Maria Calcagno[1], Chung-Pu Wu[1], Sudhir Varma[2], Joanna Shih[3], Min Lin[4], Gabriel Eichler[2], John Weinstein[2≠], Yves Pommier[2], Suresh V. Ambudkar[1], Michael M. Gottesman[1]* and Jean-Pierre Gillet[1]

Methods

TLDA data processing

TLDA cards were analyzed with RQ Manager Software (Applied Biosystems, Foster City, CA). Seven normalization methods (median normalization, 1 house-keeping gene (HK), 2-HK genes, 3-HK genes, 4-HK genes, 5-HK genes, 6-HK genes) were evaluated to identify the optimum normalization procedure for the TLDA data. The 6 HK genes used were 18S, ACTB, ATP2B4, B2M, GADPH, and HPRT1. All possible combinations of the housekeeping genes were evaluated. Median normalization was performed as follows: the median cycle threshold (C_T) value over all genes on each TLDA card was subtracted from the C_T value for each reaction well on that card such that after normalization, the median C_T value on each card was zero. For normalization using multiple HK genes, the C_T values of the multiple HK genes were averaged. The average HK C_T value on each TLDA card was then subtracted from the C_T value for each reaction well on that card.

The intraclass coefficient (ICC) defined as $\sigma^2_g / (\sigma^2_g + \sigma^2_e)$ was used to evaluate the reliability of each normalization method, where σ^2_g represented the gene to gene C_T variability (within a particular cell-line sample) and σ^2_e represented the variance within a particular gene between replicate cell-line samples. A good normalization method results in highly reproducible C_T values for replicate samples, yielding a small value of σ^2_e relative to σ^2_g and hence high ICC values. Each normalization method was assessed using data gathered from duplicates of 5 cell lines (IGROV1, OVCAR4, OVCAR5, OVCAR8, and ADR-RES). An ICC value was calculated for each of the 5 samples under each normalization method. As shown in **Suppl. Figure 1 and Suppl. Table 1**,

the median normalization was found to be as good as or better than normalization based on the best single and combined house-keeping genes, and the all subsequent data from TLDA cards was median normalized.

Drug database

Negative log10(GI50) values for 1429 compounds with known 2D structure that have been tested 4 or more times on the NCI-60 cell lines by the NIH Developmental Therapeutics Program were used in this study. This set also includes 118 compounds with known mechanisms of action. The database is available from the web-tool CellMiner at http://discover.nci.nih.gov/cellminer/.

Correlation of ABC transporter expression and drug sensitivity

For each drug/gene pair where at-least 8 cell lines have both gene expression from TLDAs as well as drug –log10 (GI50) values, we calculated the Pearson correlation and p-value (using Fisher's Z transform). Correlations for the Sybr green qRT-PCR data and p values are given in **Suppl. Table 2** while **Suppl. Table 3 (note both worksheets)** contains the Pearson correlation and p-values for the TLDA gene expression data.

Normalization of gene expression data from the TLDA and BioMark 48.48 Dynamic Array

Median-normalized data for TLDA and BioMark profiling are provided in **Suppl. Table 4.**

Statistical analysis for BioMark 48.48 Dynamic Array

Assays were performed in triplicate for each gene and each cell line. The CV (coefficient of variation) was calculated for the intra-card variability (**Suppl. Table 5**) and inter-card variability (**Suppl. Table 6**) for the raw data with C_T values <25. C_T values

>25 are reported as N.A. since the gene is not detectable. Median normalization was used to normalize each array. After the normalization, the C_T values of the triplicate of each gene were averaged to represent the intensity of that gene. Pearson correlations were calculated for the gene expression profiles derived from the TLDA and BioMark 48.48 Dynamic Array normalized data sets (**Suppl. Table 7**).

Supplement Figure Legends

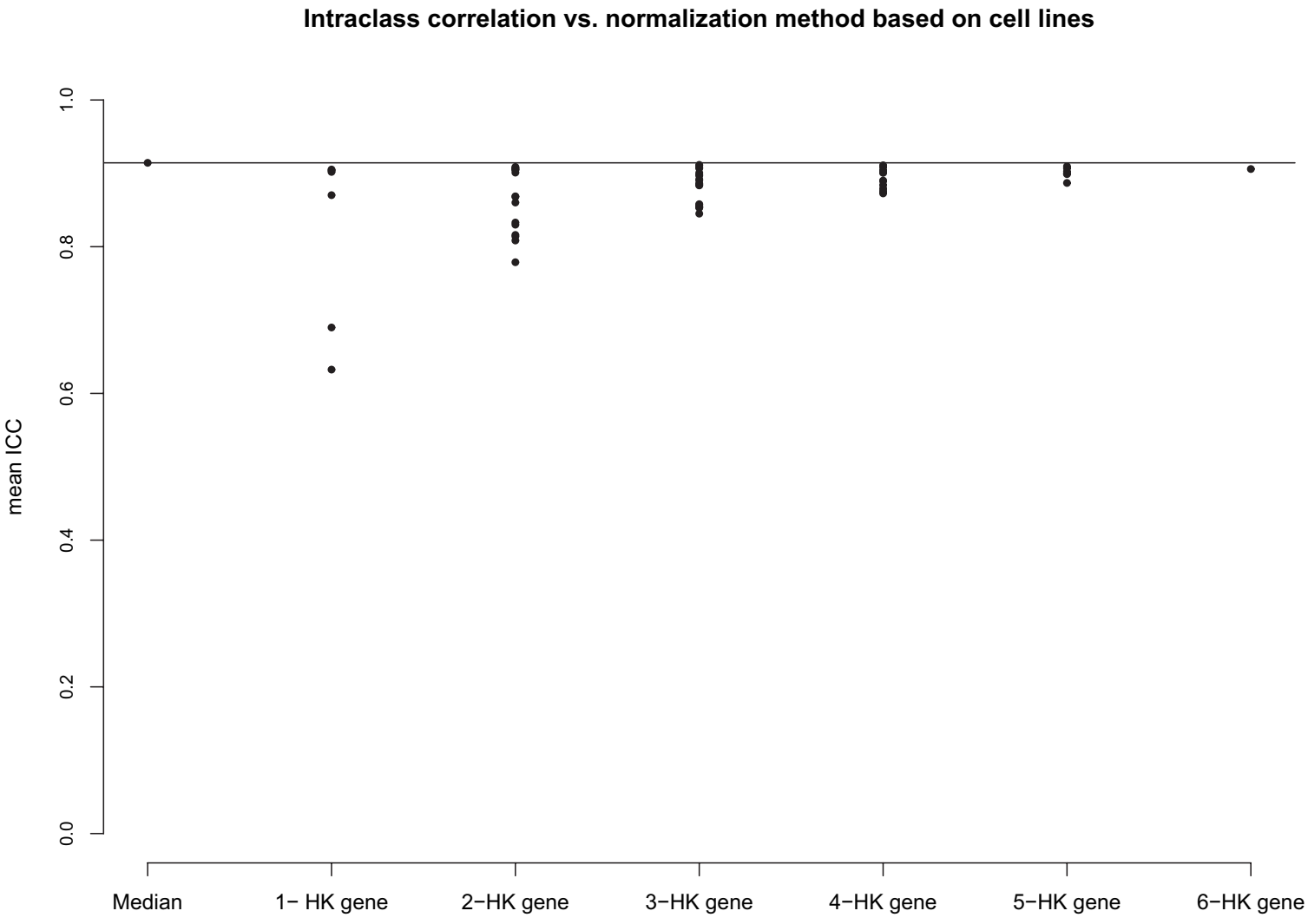
Suppl. Figure 1: Intraclass correlation vs. normalization method based on cell lines

Seven normalization methods (median normalization, 1 house-keeping gene (HK), 2-HK genes, 3-HK genes, 4-HK genes, 5-HK genes, 6-HK genes) were evaluated to identify the optimum normalization procedure for the TLDA data. The 6 HK genes used were 18S, ACTB, ATP2B4, B2M, GADPH, and HPRT1. The mean intraclass coefficient (ICC), defined as $\sigma^2_g / (\sigma^2_g + \sigma^2_e)$ where σ^2_g represented the gene to gene C_T variability (within a particular cell-line sample) and σ^2_e represented the variance within a particular gene between replicate cell-line samples was used to evaluate the reliability of each normalization technique.

Suppl. Figure 2: Structure of compounds which display ABC transporter-mediated drug resistance

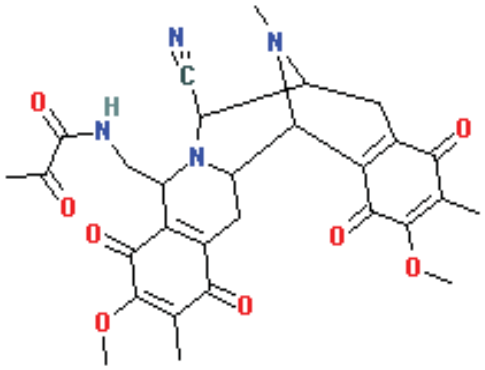
Structures for **(A)** Saframycin A, **(B)** Sparoxomycin A1 and **(C)** NSC265473 are given.

Supplemental Figure 1



Supplemental Figure 2

A

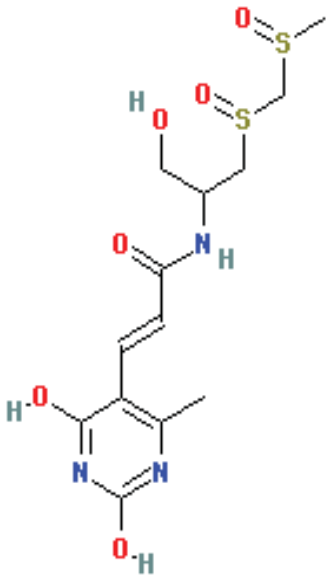


Saframycin A

NSC325663

Formula: C29H30N4O8

B

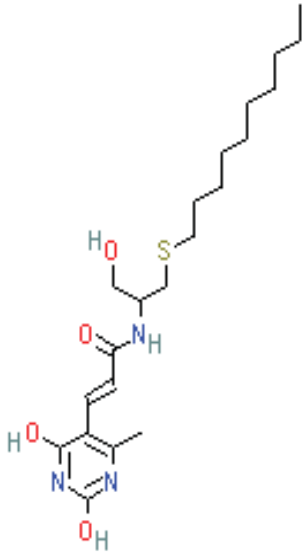


Sparoxomycin A1:

NSC251819

Formula: C13H19N3O6S2

C



NSC265473

Formula: C21H35N3O4S