Mathematical Formulations

Competition assay and derivation of log2RI values

In previous works, we generated a dataset consisted of a number of gene knockdowns (using shRNA) in *Eμ-myc; p19$^{Arf-/-}$* lymphoma cells and their responses to a large number of chemotherapeutic and targeted drugs [1,2]. The responses were measured in an *in vitro* competition assay (for more details, refer to Material and Methods and Supp Fig. S3) and recorded as an enrichment/depletion score of the shRNA population variant (to be termed 'subpopulation' in the context of a heterogeneous mixture) relative to the parental lymphoma subpopulation. Specifically, we quantified the enrichment/depletion using a log2 transform of a resistance index (RI), which was previously derived in [1]. Here we revisit the derivation for log2RI, in order to further extend its calculation for two subpopulations (plus the parental subpopulation).

We can define RI with value λ as the likelihood in survival of the fluorophore-labeled shRNA-containing infected cells relative to the uninfected parental cells. In other words, for every one cell out of n uninfected cells that survives drug treatment, λ labeled cells out of n labeled cells survives the same treatment. For a total live cell population of T (= numbers of labeled + unlabeled parental cells), with L_u as percentage of untreated labeled cells out of T, there are TL_u labeled cells and $T(1-L_u)$ unlabeled parental cells. The percentage of labeled cells following treatment can be described as,

$$L_t = \frac{\frac{TL_u\lambda}{n}}{\frac{T(1-L_u)}{n} + \frac{TL_u\lambda}{n}} \tag{1}$$

Solving for λ and log2 transform the resulting expression,

$$\text{log2RI} = \log_2 \lambda = \log_2 \left(\frac{L_t - L_t L_u}{L_u - L_t L_u} \right) \tag{2}$$

Thus, a positive log2RI represents an enrichment of the shRNA-containing labeled subpopulation, whereas negative a depletion. We can use (2) to solve for projected tumor composition with a single labeled subpopulation (L_t) following drug treatment, given the starting tumor composition (L_u) and a log2RI value.

We can now extend this to examine a heterogeneous population containing two differently labeled shRNA-containing cells, L_1 and L_2, with the corresponding RI values as λ_1 and λ_2.

$$L_{1,t} = \frac{\frac{TL_{1,u}\lambda_1}{n}}{\frac{T(1-L_{1,u}-L_{2,u})}{n} + \frac{TL_{1,u}\lambda_1}{n} + \frac{TL_{2,u}\lambda_2}{n}}, L_{2,t} = \frac{\frac{TL_{2,u}\lambda_2}{n}}{\frac{T(1-L_{1,u}-L_{2,u})}{n} + \frac{TL_{1,u}\lambda_1}{n} + \frac{TL_{2,u}\lambda_2}{n}} \tag{3}$$

Solving for λ_1 and λ_2, we have the following log2RI values,

$$\text{log2RI}_1 = \log_2 \left(\frac{L_{1,t} - L_{1,u}L_{1,t} - L_{2,u}L_{1,t}}{L_{1,u} - L_{1,u}L_{1,t} - L_{1,u}L_{2,t}} \right), \text{log2RI}_2 = \log_2 \left(\frac{L_{2,t} - L_{1,u}L_{2,t} - L_{2,u}L_{2,t}}{L_{2,u} - L_{2,u}L_{1,t} - L_{2,u}L_{2,t}} \right) \tag{4}$$

We used (4) to solve for projected tumor composition ($L_{1,t}, L_{2,t}$) following drug combination treatment, given the starting tumor composition ($L_{1,u}, L_{2,u}$) and log2RI values ($\text{log2RI}_1, \text{log2RI}_2$). Since we only know the log2RI values for single drugs to single shRNAs, we used a predicted $\hat{\text{log2RI}}$ (see next section) for combination of drugs with a heterogeneous tumor.

Drug combination optimization model

The *in vitro* efficacy dataset [1,2] we generated previously is a $m \times n$ matrix, $\mathbf{R}$, consisting of log2RI values with rows and columns representing drugs and single shRNA population variants, respectively. We have previously shown that many drug combinations act as linear averages in heterogeneous tumor cells [2]. As such, the linearity observed based on our previous experimental evidence allows us to formulate a linear function in calculating a drug combination log2RI value. Specifically, the matrix $\mathbf{R}$ can be multiplied by a population vector $\mathbf{p}$ containing values of the subpopulation proportions and a binary vector $\mathbf{d}$ representing the drug choice (i.e. a value of one means the drug is chosen for the combination; zero means the drug is not chosen). In our drug combination dosing experiments, the combination was dosed such that each drug contribute equally to a combined LD80-90 cell kill. As such, in a N-drug combination, each drug

have a contribution of $1/N$ towards the cumulative cell kill. Taken together, the log2RI value of the drug combination is,

$$\hat{\log 2\text{RI}} = \left(\frac{1}{N}\mathbf{d}\right)^T \mathbf{Rp} \tag{5}$$

Since all drugs responses for each shRNA population variant are based on a LD80-90 dosage with respect to the parental, minimizing shRNA subpopulations relative to the parental also means a minimization of the overall heterogeneous population. As such, the objective of finding the optimal drug combination for minimizing subpopulations relative to the parental subpopulation can be formulated as an optimization problem. The optimization consists of finding the values of $\mathbf{d}$ that minimizes (5), given the population composition and the total number of drugs in a combination. Since N is fixed, the drug contribution term can be ignored in formulating the objective function. Moreover, since the solutions only consist of a binary decision of whether to include or not to include a drug in the drug combination, we can thus formulate the problem specifically as a binary integer programming problem, as follows,

$$\begin{aligned} \text{given} \quad & \mathbf{R}, \mathbf{p}, N \\ & \mathbf{R} \in \mathbb{R}^{m \times n}, \mathbf{p} \in \mathbb{Q}_+^n, \mathbf{e}^T\mathbf{p} = 1, N \in \mathbb{N} \\ \text{minimize} \quad & \mathbf{d}^T\mathbf{Rp} \\ \text{subject to} \quad & \mathbf{e}^T\mathbf{d} = N \\ & \mathbf{d} \in \{0,1\}^m \end{aligned} \tag{6}$$

where $\mathbf{e}$ is a vector of ones. There are several assumptions in this model, including no interactions among subpopulations and additive drug combinations. Within the scope of drugs and population variants examined in this study, these assumptions are supported by our previous experiments [1,2]. In addition, we assume there are no significant genetic drifts in the tumor population that greatly diversify the population to affect drug responses beyond those conferred by the initial heterogeneous tumor population. Under the timescale of our experiments, we have not observed any drifts in drug response.

This computational model thus allows us to explore optimal N drug combinations for a heterogeneous population $\mathbf{p}$. In this study, as proof of concept, we examined three-component population (i.e. two RNAi-produced subpopulations + parental subpopulation) with two drug combinations ($N = 2$). Future works will explore more complex heterogeneous tumor compositions across the full range of N drug combinations.

The simulations were performed using MATLAB R2012b (Mathworks) with the optimization solved using `bintprog` from the MATLAB Optimization Toolbox.

Descriptive model of therapeutic efficacy comparison

We developed a descriptive model to explore the efficacy difference between Vin/SAHA and IRT/CBL combinations in mice transplanted with three-component shChk2/shBok/parental tumor over varying subpopulation proportions. Since the tumor consists of three subpopulations, all possible tumor compositions can be represented using a ternary plot, with each of the corners representing a homogeneous tumor consisting of one of the subpopulations. Assuming a combined survival metric as a linear combination of the survival metric of individual subpopulations and subpopulation proportions, we can estimate the survival metric $S_t^{\mathbf{P}}$ of a given heterogeneous tumor composition $\mathbf{p}$ treated with therapy t as follows,

$$\hat{S}_t^{\mathbf{P}} = \mathbf{S}_t^T\mathbf{p} \tag{7}$$

Since all mice eventually relapse and there are no censored data, we can utilize the mean absolute/relative survival, which corresponds to the area under the survival curve, as our survival metrics. We first determined the tumor-free survival ($\mathbf{S}$) for mice transplanted with a homogeneous tumor treated with either Vin/SAHA or IRT/CBL (Supp Fig. S9). The relative tumor-free survival of a mice with heterogeneous tumor for each drug combination can also be calculated through normalization of days to the median survival of mice with homogeneous tumor (Supp Fig. S10). We explored the efficacy difference between the two drug combinations both in terms of relative and absolute tumor-free survival.

For relative tumor-free survival,

$$\Delta \hat{S}^{\mathbf{P}}_{\text{Vin/SAHA - IRT/CBL}} = 7.17 p_{\text{shChk2}} + 3.00 p_{\text{shBok}} - 0.73 p_{\text{parental}} \tag{8}$$

For absolute tumor-free survival,

$$\Delta \hat{S}^{\mathbf{P}}_{\text{Vin/SAHA - IRT/CBL}} = 4.17 p_{\text{shChk2}} + 0.00 p_{\text{shBok}} - 3.73 p_{\text{parental}} \tag{9}$$

A matrix of possible tumor compositions and resulting tenary plots were generated in MATLAB.

References

[1] Jiang H, Pritchard JR, Williams RT, Lauffenburger DA, Hemann MT. A mammalian functional-genetic approach to characterizing cancer therapeutics. Nat Chem Biol. 2011;7:92–100.

[2] Pritchard JR, Bruno PM, Gilbert LA, Capron KL, Lauffenburger DA, Hemann MT. Defining principles of combination drug mechanisms of action. Proc Natl Acad Sci U S A. 2013;110:E170–9.