ImmunoTherapy scoring function

The statistics I propose for $s(\bullet)$ will be the proportion vector {a, b, c, d, e} where a is the proportion of progenitor cells, b, the proportion for effector cells, c, for terminal exhausted, d, for the cycling cells and e, for other cells.

So $s(P_i)$ will be $\{a_i, b_i, c_i, d_i, e_i\}$ where i stands for the index of the knocked out gene yielding the corresponding gene expression distribution P_i.

Let P_0 correspond to the proportion vector $\{a_0, b_0, c_0, d_0, e_0\}$ and Q be the proportion vector $\{a_q, b_q, c_q, d_q, e_q\}$. The proposed score function will be given by:

$$\frac{\exp(\Sigma_{i=0}^{n}\frac{1}{var_i})}{\exp((\Sigma_{i=1}^{n}\frac{L_i}{var_i}) - \frac{M}{var_0})} \quad \text{where}$$

L_i is the L_1 loss between P_i and Q given by

$$|a_i - a_q| + |b_i - b_q| + |c_i - c_q| + |d_i - d_q| + |e_i - e_q|$$

M is the L_1 loss between P_0 and Q,

$$|a_0 - a_q| + |b_0 - b_q| + |c_0 - c_q| + |d_0 - d_q| + |e_0 - e_q|$$

n is the number of genes

var_0 is the variance of M and

var_i where $i = 1,..n$ is the variance of L_i

Dividing M and L_i by respective variance will ensure that quantities with more uncertainties can be given less weight in the calculation of the score.

The variance of L_i is the sum of the variances, $var(a_i) + var(b_i) + var(c_i) + var(d_i) + var(e_i)$

Similarly the variance of M is calculated by taking the sum $var(a_0) + var(b_0) + var(c_0) + var(d_0) + var(e_0)$ from the data representing P_0

The resulting score is the pseudo-inverse of a weighted average of the knockout losses minus the loss of unperturbed distribution, P_0 with Q where each loss is weighted by the inverse of the corresponding variance. Less loss in the knockout will yield greater the score and thus indicative of better perturbations. Taking the exponent will help in avoiding negative values in the denominator thus being supportive to the purpose of the scoring function. Finally greater M with less L_i will suggest that the model is performing well irrespective of the unperturbed distribution not already close to Q.