

For challenge 3, a simple variation of part-a and part-b scoring function has been proposed. Here, respective original scoring functions of a knock-out gene in part-a and part-b have been scaled by a kind of P_0 summary value for the corresponding gene in unperturbed cell condition. Here are the forms of both proposed scoring functions

1. proposed scoring function of immune checkpoint blockade therapy:

$$s_{\text{checkpoint}}() = \frac{a_i}{E[a'_i | i_{\text{expr}}=0, \text{unperturbed}]}$$

where a_i is progenitor state proportion of knock-out gene i , and $E[a'_i | i_{\text{expr}}=0, \text{unperturbed}]$ is the scaling factor and the expected state proportion of gene i in the unperturbed cells when the i gene's expression value = zero. Cells with the i gene's expression value = zero in unperturbed condition have been naively assumed a 'pseudo' knock-out condition of i gene for convenience. (more comments on this assumption in the below)

2. proposed scoring function of cart t-cell therapy:

$$s_{\text{cart}}() = \frac{\frac{a_i}{E[a'_i | i_{\text{expr}}=0, \text{unperturbed}]} + \frac{b_i}{E[b'_i | i_{\text{expr}}=0, \text{unperturbed}]} - \frac{c_i}{E[c'_i | i_{\text{expr}}=0, \text{unperturbed}]} + \frac{d_i}{E[d'_i | i_{\text{expr}}=0, \text{unperturbed}]}}{d_i}$$

where the respective definition of scaling factors is the same as above and a_i = progenitor b_i = effector, c_i = terminally exhausted, and d_i = cycling state of gene i , respectively.

As a concrete example, the gene 'Tcf7' case in the given data shows respective scoring function values as below, and the proposed scoring functions give higher scoring values of Tcf7 (0.104377 → 3.654745 for checkpoint blockade, and 2.648224 → 4.712685 for CAR-T) and higher ranking position (45th → 27th and 44th → 27th respectively) in the summary table of 64 knock-out genes below (Table 1.):

$$s_{\text{part_a}}() = a_{\text{Tcf7}} = 0.104377$$

$$s_{\text{part_b}}() = \frac{a_{\text{Tcf7}}}{0.0675} + \frac{b_{\text{Tcf7}}}{0.2097} - \frac{c_{\text{Tcf7}}}{0.3134} + \frac{d_{\text{Tcf7}}}{0.3921} = \frac{0.104377}{0.0675} + \frac{0.292929}{0.2097} - \frac{0.313131}{0.3134} + \frac{0.276094}{0.3921} = 2.648224$$

$$s_{\text{checkpoint}}() = \frac{0.104377}{0.028559} = 3.654745$$

$$s_{\text{cart}}() = \frac{a_{\text{Tcf7}}}{0.028559} + \frac{b_{\text{Tcf7}}}{0.217897} - \frac{c_{\text{Tcf7}}}{0.326000} + \frac{d_{\text{Tcf7}}}{0.409562} = \frac{0.104377}{0.028559} + \frac{0.292929}{0.217897} - \frac{0.313131}{0.326000} + \frac{0.276094}{0.409562} = 4.712685$$

where the state proportion s_{Tcf7} of $\frac{s_{\text{Tcf7}}}{E[s'_{\text{Tcf7}} | \text{Tcf7}_{\text{expr}}=0, \text{unperturbed}]}$ is: state proportion $s_{\text{Tcf7}} = a_{\text{Tcf7}}$ for $s_{\text{checkpoint}}()$, and $= a_{\text{Tcf7}}, b_{\text{Tcf7}}, c_{\text{Tcf7}},$ or d_{Tcf7} for $s_{\text{cart}}()$, respectively.

$E[a'_i | i_{\text{expr}}=0, \text{unperturbed}]$ can be considered a form of P_o statistics, and can be calculated from the generated samples from the trained LDVAE models that I had used in the challenge-1 and to be confirmed by the experimental results. But, since the experimental results of 66 knock-out conditions are already given in the data set, I had applied the part-a&-b scoring functions and 2 proposed scoring functions directly to this 66KO experiment data to compare the proposed scoring functions to the originals. From the 66 knock-out gene conditions, 'Fzd1', 'P2rx7' gene conditions had been excluded because they are not present in expression data, and resulting 64 knock-out experiment data had been used and the results have been summarized in Table 1.

From the results in Table 1., for instance, some genes such as 'Hmgb1', 'Myb', 'Stat4', 'Nr4a3', 'Satb1', 'Tcf7' shows deferentially higher ranking position in scoring results (Table 1.), where they shows relatively higher position in the proposed scoring functions than in the original part-a&-b scoring functions. The genes likes 'Tpt1', 'Id2' show the opposite pattern. The higher ranking of 'Tcf7' may be consistent with the explanation of the gene from the crash course, but other genes' differential ranking may not be interpreted by myself due to my lack of domain knowledge of this field, and left undone. I wish that the challenge host would interpret it properly and confirm it whether the proposed scoring functions could improve the performance (I wish it could! ☺).

Finally, the idea of 'pseudo' knock-out gene assumption of gene-expressed-zero cells in unperturbed population serendipitously came in the process of devising a naive baseline solution to challenge1. Specifically, for the checkpoint-2 submission, a merely naive baseline solution had been devised from the assumption that when a gene has 'zero' expression value in 'unperturbed' cell, it has been considered as an instance of 'pseudo' knock-out of that gene as a prior(as a merely naive postulation). Once collecting all such cases (i.e. all the cells that show zero expression value of the corresponding gene in unperturbed condition cells) as the hypothetical 'pseudo' knock-out cases and the cell state proportion vector has been calculated by counting those cell of the respective state in the unperturbed condition cells, this naive baseline solution had been placed in the 16th position in the leaderboard of checkpoint-2 submission indicating that it is a better solution than just a random or uniform dummy baseline. This result encouraged me to conceive the idea of 'pseudo' knock-out gene assumption in unperturbed cells and try a possibly(&hopefully) beneficial role in proposing a scoring function. So, I have used it as a scaling factor of the part-a&-b scoring functions as shown above. If this 'pseudo' knock-out gene assumption would improve the performance of scoring function, and eventually efficacy of clinical trials of the cancer therapy of the target gene(s), the variations of individual patient' response to the therapy might be in some degree explained by the patient's 'pseudo' knock out state vector of the target gene(s) of the therapy in his/her normal unperturbed cell populations (another mere postulation, though ☺).

Table 1. Scoring results from 4 scoring functions(part-a, part-b, proposed $s_{\text{checkpoint}}$ and s_{cart}) and respective ranking for 64 knock-out genes (sorted by the ranking of the proposed s_{cart} scoring function)

gene	Cycling constraint	part_a score	part_a rank	part_b score	part_b rank	$s_{\text{checkpoint}}$ score	$s_{\text{checkpoint}}$ rank	s_{cart} score	s_{cart} rank
Hmgb1	1	0.071618	49	1.959266	52	0.470943	53	71.220880	1
Klf2	0	1.000000	5	14.814815	5	15.243151	1	15.243151	2
Ep300	0	1.000000	6	14.814815	6	14.995851	2	14.995851	3
Batf	0	0.666667	7	9.344741	9	10.963964	3	10.426657	4

gene	Cycling constraint	part_a score	part_a rank	part_b score	part_b rank	S_checkpoint score	S_checkpoint rank	S_cart score	S_cart rank
Runx3	0	1.000000	2	14.814815	2	10.423077	4	10.423077	5
Myb	1	0.419355	15	7.083245	16	8.921029	7	9.656989	6
Yy1	1	0.600000	9	9.893412	7	8.634637	9	9.474855	7
Foxm1	1	0.642857	8	9.772941	8	9.150068	6	9.456688	8
Cttnb1	1	0.534884	10	8.773632	10	8.636415	8	9.273742	9
Tbx21	0	1.000000	4	14.814815	4	9.170984	5	9.170984	10
Sp140	1	0.515152	11	8.441407	11	8.181818	10	9.025114	11
Stat4	1	0.300000	21	5.661959	20	7.081481	11	8.058213	12
Eomes	1	0.437500	13	8.250474	12	5.920178	17	7.512554	13
Oxnad1	1	0.423729	14	7.532690	15	6.266218	16	7.478992	14
Id3	1	0.391304	17	6.294318	17	6.902721	14	7.394578	15
Nr4a3	1	0.280405	25	4.968184	28	6.372850	15	7.211068	16
Rad21	1	0.400000	16	7.769369	13	5.466667	18	6.919559	17
Dvl3	1	0.347826	19	5.749617	19	5.153436	19	5.750594	18
Lef1	1	0.270000	28	5.423991	22	4.338685	23	5.717091	19
Irf9	1	0.375000	18	6.140312	18	5.031475	20	5.658134	20
Ezh2	1	0.230769	31	5.265257	23	3.518142	29	5.597452	21
Nr4a1	1	0.282609	24	4.606986	31	4.828078	21	5.484438	22
Satb1	1	0.192460	35	4.247032	35	3.998280	25	5.376193	23
Rela	1	0.228261	32	5.051039	26	3.533861	28	5.085995	24
Gsk3b	1	0.297872	23	4.931672	29	4.168840	24	4.821325	25
Sub1	1	0.111940	43	2.558986	46	0.434251	55	4.719056	26
Tcf7	1	0.104377	45	2.648224	44	3.654745	27	4.712685	27
Elf1	1	0.260274	29	5.444855	21	3.459642	30	4.699078	28
Prdm1	1	0.299479	22	4.972821	27	3.946611	26	4.525919	29
Lrp1	1	0.305556	20	4.479199	33	4.535904	22	4.488569	30
Zfp292	1	0.124601	41	4.718670	30	1.873872	42	4.443858	31
Tox	1	0.189189	36	4.519690	32	2.842011	34	4.209878	32
Egr1	1	0.274678	26	5.207423	24	2.985591	32	4.078385	33
Fzd3	1	0.194245	34	4.044082	36	2.877826	33	4.043999	34
Dkk3	1	0.187500	37	3.996422	37	2.776786	35	3.995702	35
Tcf3	1	0.146067	39	3.432727	39	2.437548	36	3.816300	36
Nr4a2	1	0.197059	33	3.387241	40	2.138298	38	3.286937	37
Fzd6	1	0.135678	40	3.255403	41	2.006420	40	3.246058	38
Il12rb2	1	0.181818	38	3.440156	38	2.311778	37	3.022731	39
Foxp1	1	0.111111	44	2.450844	47	2.118877	39	2.959286	40
Atf2	1	0.120567	42	2.756508	43	1.775331	43	2.741731	41
Rps6	0	1.000000	3	14.814815	3	7.000000	13	2.692450	42
Eef2	0	1.000000	1	14.814815	1	7.000000	12	2.692450	43
Tpt1	1	0.440000	12	7.612720	14	3.080000	31	2.692450	44
Id2	1	0.273504	27	5.089450	25	0.899917	47	2.638175	45
Crem	1	0.034166	57	2.640584	45	0.359371	60	2.440850	46
Foxo1	1	0.100000	47	2.154527	49	1.503500	45	2.165732	47
Ldhb	1	0.103093	46	2.116125	50	1.573059	44	2.165192	48
Dvl1	1	0.042017	54	2.166617	48	0.625099	51	2.153896	49
Sox4	1	0.069697	50	1.945085	53	1.040641	46	1.953211	50
Runx2	1	0.034384	56	1.979213	51	0.428141	56	1.866564	51
Litaf	1	0.071625	48	2.822769	42	0.399769	58	1.823611	52
Hif1a	1	0.032117	59	1.248649	59	0.695205	49	1.808898	53
Sp100	1	0.024476	60	1.838825	54	0.290236	61	1.734533	54
Irf2	1	0.034632	55	1.378305	57	0.451602	54	1.636493	55
Nr3c1	1	0.050667	53	1.547645	55	0.660606	50	1.567793	56
Il12rb1	1	0.032520	58	1.514775	56	0.406606	57	1.440877	57
Hmgb2	1	0.231579	30	4.360472	34	2.003158	41	1.425955	58
Dvl2	1	0.024390	61	1.342913	58	0.365124	59	1.369793	59
Arid4b	1	0.056604	51	1.248439	60	0.871698	48	1.323321	60
Zeb2	1	0.017483	62	1.029882	62	0.244900	63	1.093988	61
Ikzf3	1	0.053004	52	1.210514	61	0.499569	52	0.984636	62
Tox2	1	0.017309	63	0.708212	63	0.255192	62	0.703282	63
Arid5b	1	0.011241	64	0.371409	64	0.163673	64	0.617407	64