Discovery of targets for immune-metabolic antitumor drugs identifies Estrogen Related Receptor Alpha

Supplementary Notes

1. Comparison of BipotentR immune sub-module with macrophage polarization

BipotentR bulk-tumor sub-module uses to proinflammatory signature of BipotentR that combines recently published 32 key immune response biomarkers (1). As an alternative to the proinflammatory signature, we evaluated a signature of macrophage polarization. Specifically, we estimated M1 and M2 macrophage levels in TCGA using Cibersort (2) and used their difference (M1 - M2) as an inferred level of macrophage polarization in a tumor. We then associated polarization with the activity of 672 TFCRs in TCGA tumors by replacing PC1 with M1/M2 in the bulkRNA sub-module (Methods 2.4.2.1) and recalculated the new associations $\beta_{bulkRNA}(polarization)$ and their significance $P_{bulkRNA}(polarization)$ for each TFCRs. Finally, we compared top-30 significant predicted TFCRs with top-30 TFCRs predicted by BipotentR bulkRNA sub-module. Among the top 30 TFCRs associated with polarization, 11 overlapped with the top-30 TFCR associated with PC1 (**Fig S18B**). Though the procedure identified macrophage regulators such as MAF and EZH1, it failed to identify important regulators of other immune cells such as BATF, EMOES, and IRF4. Although our initial application was focused on general immune regulators, future studies could be designed to focus on identifying regulators of each immune cell separately.

2. Benchmark BipotentR against BARTWeb

We benchmarked BipotentR against BARTWeb (3), using a strategy used to benchmark BipotentR against LISA. Specifically, BARTWeb-predicted TFCRs were obtained by inputting the union of the four energy metabolism pathways. The top-38 BARTWeb-predicted TFCR were then evaluated against BipotentR-predicted TFCRs (n=38). To evaluate these two TFCR sets, we compared how strongly their knockout suppressed energy metabolism genes in the KnockTF database (4). We observed knockout of BipotentR TFCRs suppresses energy metabolism genes more strongly than knockout of BARTWeb-predicted TFCRs (**Fig S2D**).

3. ESRRA activity in patient tumors correlates with antigen presentation, immune cell infiltration, and macrophage polarization

We analyzed 33,000 tumor transcriptomes compiled from several cohorts (5), including TCGA and PRECOG datasets (6). ESRRA activity in tumors was quantified as the weighted sum of expression of ESRRA targets

(described in Methods 2.8). In tumors with low ESRRA activity, tumor energy metabolism was significantly downregulated across several cancer types in TCGA (**Fig S3C**; Methods). In these tumors, cytokine interaction pathways were upregulated (**Fig S3D**), consistent with the *in vitro* induction of macrophage-polarizing cytokines upon ESRRAi. We, therefore, asked whether macrophage polarity was also shifted in such tumors. We analyzed how macrophage polarity (estimated using gene expression signature (7)) relates to ESRRA activity in tumors, and found that macrophage polarity was markedly correlated with ESRRA activity within tumors across most cancer types (**Fig S3A**), strongly suggesting that M1 macrophage polarization upon ESRRAi seen *in vivo* in mice may be clinically relevant in most cancer types.

Antigen presentation genes were upregulated in tumors with low ESRRA activity across several cancer types (**Fig S3E**). Expression of antigen presentation genes differs with immune cell expresses, therefore, expression of antigen pathway genes in bulk tumors from TCGA could be confounded by immune cell infiltrates. To ensure the association between ESRRA activity and antigen presentation genes is due to this confounding effect, we performed two analyses. First, we analyzed single-cell data from patients' tumors. We studied cancer cells of patient tumors from a single-cell cohort (Jerby-Arnon et al. (8)). In particular, we adopted the unsupervised clustering approach from Puram et. al 2017 to cluster the cancer cells, which divided the cancer cells into two clusters with significantly different ESRRA activities (**Fig S3F, G**). The cluster with low ESRRA activity showed up-regulation of MHC genes compared to other clusters (**Fig S3H**).

Antigen presentation promotes immune infiltration into tumors. If antigen presentation simulated by ESRRAi is clinically relevant, tumors deficient in ESRRA activity should have elevated immune infiltration. Indeed, tumors with low ESRRA expression had high immune infiltrations across most cancer types (**Fig S3B**, $P < 2E\text{-}16$ controlled for cancer types). Finally, we found that low ESRRA activity in tumors is associated positively with proinflammatory factors, and negatively with anti-inflammatory factors in both TCGA (**Fig S4A**) and PRECOG data (**Fig S4B**). Together, the patient tumor data demonstrate the potential clinical relevance of ESRRAi in enhancing antigen presentation, immune cell infiltration, and macrophage polarization in multiple cancer cohorts and cancer types.

4. ESRRA activity associated with immune infiltration, antigen presentation, and cytokine activation in bladder cancer cohort.

We validated the associations of tumor ESRRA activity with immune infiltrations, antigen presentation, and cytokine activation in a tumor cohort where immune infiltration was measured by immunohistochemistry (9). We first estimated ESRRA activity in 300 tumors of bladder cancer patients using their tumor transcriptomes. The tumor ESRRA activities were negatively associated with higher tumor $CD8^{+}$ T

infiltrations: immune-inflamed tumors with the highest CD8$^+$ T infiltration have the lowest ESRRA activity (**Fig S4C**), followed by immune excluded tumors, while CD8$^+$ T deficient tumors showed the highest ESRRA activity. We also applied dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP), which automatically projected tumors into two clusters, with a pronounced difference in ESRRA activity (**Fig S4D**, P< 3.9E-08). The cluster with low ESRRA activity showed enrichment in tumors with high immune infiltration (**Fig S4E**). This cluster also exhibited a remarkable upregulation of antigen presentation genes and cytokines (**Fig S4F**), particularly cytokines known to polarize macrophages towards M1 (**Fig S4G**). Thus tumors were seen in two distinct states: one with low ESRRA activity exhibiting upregulated antigen presentation, cytokines, and immune infiltration but low energy metabolism; and the other with high ESRRA activity exhibiting downregulated antigen presentation, cytokines, and immune infiltration but high energy metabolism. Overall, these results recapitulate the association of immune activation and infiltration with ESRRA inhibition in tumors.

5. ESRRA activity correlates with upregulation of antigen presentation and cytokines in cancer cell lines

We analyzed data from ~600 CCLE cancer cell lines. We first quantified ESRRA activity in cell lines using their transcriptomes (as the weighted sum of expression of ESRRA targets Methods 2.8), then compared cell lines with low ESRRA activity with those with high ESRRA activity. Cell lines with a low ESRRA activity group showed downregulation of energy metabolism across several cancer types (**Fig S7B**) as well as upregulation of immune pathways, such as cytokine-interactions and viral infection (**Fig S7C**). These cell lines also showed upregulation of multiple cytokines, particularly those that recruit macrophages (such as CCL2 and CCL5) and shift macrophage polarity toward activated M1 macrophage (IL1B and IL6) (**Fig S7D**). We also observed these cell lines upregulated antigen presentation genes across most cancer-types (**Fig S7E**). This data was consistent with our observations that ESRRA suppression upregulates immune pathways, particularly cytokines and antigen presentation.

CCLE mapped genes using a generic reference, but alleles of MHC genes differ between individuals, so MHC mapping should be done relative to individual reference genotype. To mitigate the potential effects of this confounding factor on the ESRRA association that we observed with MHC genes, we performed the following analysis regarding MHC-I alleles. MHC-I (HLA-A/B/C) genes are co-expressed with each other; therefore, expression of HLA-C in a sample could be predicted from expression of HLA-A and HLA-B in the sample. However, the predicted HLA-C expression in a sample would deviate from HLA-C expression estimated by the aligner if the aligner failed to map reads to HLA-C because the HLA-C allele of the sample was different from the reference allele. In such samples, aligner-estimated HLA-C expression would be much lower than predicted. To identify such samples, we calculate D = aligner(HLA-C) - predicted(HLA-C),

and found indeed D is more often negative (distribution of D shown in **Fig S7F** is heavier for negative tail than positive tail), suggesting the mapping problem. We filtered out samples with extreme negative D (bottom 20 percentile) while calculating the correlation between HLA-C and ESRRA activity. We continued to observe a significant correlation between HLA-C and ESRRA activity (**Fig S7G**). We repeated similar procedures for HLA-A and HLA-B and also observed a significant correlation with ESRRA activity (**Fig S7G**). The data support a correlation between antigen presentation and ESRRA that is not confounded by misalignment.

6. Compare BipotentR with alternative approaches of applying modules and sub-modules serially

We also examined whether BipotentR's predictive power to detect bipotent regulators could be improved by using modules and sub-modules as filters. Below we describe comparative analyses: in which we first apply sub-modules serially and then apply modules serially.

i) Using scRNA sub-module as a filter: In immune-module of BipotentR, we calculated the integrated score for each TFCR defined as averaged bulk and scRNA estimates, and then apply Wald's test to the integrated score to determine its significance. We examined BipotentR predictive power improves if the scRNA sub-module is used as a filter in the immune module. Specifically, in the immune-module, we first removed TFCR significant in the scRNA sub-module (adjusted p value<0.05). Then, we applied bulk-RNA sub-module to the remaining TFCRs and identified 142 TFCRs as significant (adjusted p-value < 0.05). On the other hand, original immune module identified 150 TFCRs significant based on adjusted p-value of the integrated score (averaged bulk and scRNA estimates). This is because the application of Wald's test on the integrated scores enabled thresholding of adjusted p-value < 0.05 once in the immune-module. Expectedly, sets of TFCRs identified by the modified and original immune modules overlapped strongly with 119 TFCRs held in common (**Fig S18C**, Fisher test: p-value < 2.2e-16).

ii) Applying two modules serially. We removed TFCRs using the regulator module, and applied the immune module to the rest. This serial application identified the exact same 38 TFCRs identified by BipotentR. This is because dominating filtering criteria of the current implementation is the effect size. Our estimates of effect size account for variation. For example, we estimated the effect size for a TFCR at the 95% confidence interval, and therefore a large variation shrinks the effect size.

7. Compare BipotentR with Firth's regression for robustness to class imbalance problem

BipotentR regulation module uses a GLMM to predict genes of an input pathway. Since the number of pathway genes is much smaller than non-pathway genes, it could cause the problem of class imbalance in GLMM. We examined class imbalance affects p-value estimates of BipotentR using Firth's regression, a regression approach developed to mitigate class imbalance problem. For the analysis, we choose to examine the TCA pathway without loss of generalizability. We applied Firth's penalized regression and estimated the significance of TFCR potential to bind genes in the TCA pathway. We compared the significance of all TFCRs estimated from Firth's regression with those from BipotentR (regulation module), and observed no significant difference in p values between the two methods (KS test: p-value = 0.494), suggesting p-value estimates from the two approaches were similar.

Reference

1. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang T-H, et al. The Immune Landscape of Cancer. Immunity. 2018;48:812-830.e14.

2. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015;12:453–7.

3. Ma W, Wang Z, Zhang Y, Magee NE, Feng Y, Shi R, et al. BARTweb: a web server for transcriptional regulator association analysis. NAR Genom Bioinform. 2021;3:lqab022.

4. Feng C, Song C, Liu Y, Qian F, Gao Y, Ning Z, et al. KnockTF: a comprehensive human gene expression profile database with knockdown/knockout of transcription factors. Nucleic Acids Res. 2020;48:D93–100.

5. Fu J, Li K, Zhang W, Wan C, Zhang J, Jiang P, et al. Large-scale public data reuse to model immunotherapy response and resistance. Genome Med. 2020;12:21.

6. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, et al. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. Cell. 2018;173:291-304.e6.

7. Jiang P, Gu S, Pan D, Fu J, Sahu A, Hu X, et al. Signatures of T cell dysfunction and exclusion predict cancer immunotherapy response. Nat Med. 2018;24:1550–8.

8. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su M-J, Melms JC, et al. A Cancer Cell Program Promotes T Cell Exclusion and Resistance to Checkpoint Blockade. Cell. 2018;175:984-997.e24.

9. Mariathasan S, Turley SJ, Nickles D, Castiglioni A, Yuen K, Wang Y, et al. TGFβ attenuates tumour response to PD-L1 blockade by contributing to exclusion of T cells. Nature. 2018;554:544–8.