



Published in final edited form as:

Cancer Discov. 2022 April 01; 12(4): 1088–1105. doi:10.1158/2159-8290.CD-21-0887.

Deconvolving clinically relevant cellular immune crosstalk from bulk gene expression using CODEFACS and LIRICS stratifies melanoma patients to anti-PD-1 therapy

Kun Wang^{**1}, Sushant Patkar^{**1,2}, Joo Sang Lee^{1,3}, E. Michael Gertz¹, Welles Robinson^{1,2}, Fiorella Schischlik¹, David R. Crawford^{1,4}, Alejandro A. Schäffer¹, Eytan Rupp^{#,1}

¹Cancer Data Science Laboratory, National Cancer Institute, NIH, Bethesda, MD

²Department of Computer Science, University of Maryland, College Park, MD

³Department of Artificial Intelligence & Department of Precision Medicine, School of Medicine, Sungkyunkwan University, Suwon, Republic of Korea

⁴Department of Cell Biology and Molecular Genetics, University of Maryland, College Park, MD

Abstract

The tumor microenvironment (TME) is a complex mixture of cell types whose interactions affect tumor growth and clinical outcome. To discover such interactions, we developed CODEFACS (COntident DEconvolution For All Cell Subsets), a tool deconvolving cell-type-specific gene expression in each sample from bulk expression, and LIRICS (LIgand Receptor Interactions between Cell Subsets), a statistical framework prioritizing clinically relevant ligand-receptor interactions between cell types from the deconvolved data. We first demonstrate the superiority of CODEFACS versus the state-of-the-art deconvolution method, CIBERSORTx. Second, analyzing the TCGA, we uncover cell-type-specific ligand-receptor interactions uniquely associated with mismatch repair deficiency across different cancer types, providing additional insights into their enhanced sensitivity to anti-PD1 therapy compared to other tumors with high neoantigen burden. Finally, we identify a subset of cell-type-specific ligand-receptor interactions in the melanoma TME that stratify survival of patients receiving anti-PD1 therapy better than some recently published bulk transcriptomics-based methods.

[#] Corresponding author and contact information: Eytan Rupp, National Institutes of Health, Building 15-C1; Bethesda, MD 20892 USA, Telephone: 240-858-3169, eytan.rupp@nih.gov.

^{**} these authors contributed equally to this work and are co-first authors

Author contributions

K.W., S.P. A.A.S. and E.R. conceptualized and designed the project. A.A.S. and E.R. supervised the project. K.W. and S.P. implemented the software and performed the data analyses. K.W., S.P., A.A.S. and E.R. wrote the original draft with input from all authors. J.L., E.M.G., W.R., F.S., and D.R.C. contributed to data analyses, visualization, review and editing. K.W. and S. P share first authorship on this work.

Competing interests

The authors declare no competing interests.

Competing interests

The authors declare no competing interests.

Keywords

deconvolution; gene expression; immunotherapy; tumor microenvironment; ligand-receptor interaction; cell-cell interaction; microsatellite instability

Introduction

The importance of the tumor microenvironment (TME) in cancer has been recognized since the late 1800s(1). The recent success of immune checkpoint blockade has further sparked interest in studying the various cellular states of cell types and their interactions in the TME, aiming to find biomarkers of treatment response and new treatment opportunities(2). One key step in elucidating the different cellular states of cell types *de novo* is the characterization of the molecular profiles of each cell type in *each patient's tumor sample*. Fluorescence-activated cell sorting (FACS) and single-cell RNA sequencing have emerged as effective tools to address this challenge(3). The use of FACS is limited because it can only measure expression of a few protein markers simultaneously. The use of single-cell RNA sequencing has been limited due to its cost and the scarcity of fresh tumor biopsies. Since bulk tumor gene expression from preserved biopsies accompanied by clinical outcome metadata is abundant, computational methods that can effectively extract cell-type-specific expression from such data in each individual sample could be very helpful. If successful, such deconvolution methods could be used to identify clinically relevant cellular states of cell types and prioritize cell-cell interactions associated with patient survival and response to treatment. Furthermore, they may be readily applied to interrogate the troves of large bulk expression datasets that are in the public domain, to further advance what we can learn from them.

Recovery of cell-type-specific gene expression profiles from bulk gene expression is a computationally challenging problem because the standard formulation requires one to solve a large system of underdetermined linear equations. This problem is closely related to the problem of compressed sensing in signal processing(4). Several recent studies have developed a variety of algorithms to address this challenge. DeMix(5) was designed to estimate individual-specific expression for three cell components provided prior reference samples of two of these cell components. ISOpure(6) has aimed to derive sample-specific tumor cell expression only, with the assumption that the observed bulk gene expression profile is a mixture of predefined stromal and immune cell expression profiles that are shared across all the samples. Building on this work, Fox et al extended ISOpure to predict individual-specific non-tumor cell expression by subtracting cancer cell expression profiles from the bulk mixtures in a two-cell type model(7). More recently, Newman et al (8) developed CIBERSORTx, the first approach that aims to predict the *sample-specific* gene expression of all cell types composing it by employing a set of novel deconvolution heuristics. As a proof of concept, Newman et al showed that CIBERSORTx can accurately reconstruct the cell-type-specific expression of genes in each input sample under certain modelling assumptions. This groundbreaking work has, however, some notable limitations: (1) The number of genes whose cell-type-specific expression can be reconstructed in each sample is quite small, especially for low-abundance cell types, and (2) CIBERSORTx does

not provide confidence estimates of the predictions made. Confidence estimates on the expression predictions could be useful in most deconvolution applications because of the absence of ground truth data.

Starting from the work of Newman et al, we introduce a new deconvolution algorithm and software, CODEFACS (COntident DEconvolution For All Cell Subsets) that markedly advances our ability to successfully reconstruct cell-type-specific gene expression of each bulk sample. CODEFACS receives as input bulk gene expression profiles of tumor samples and either pre-computed estimates of abundance of expected tumor, immunological and stromal cell types in each sample, or their prototypical molecular signatures, which serve as seeds for estimating the abundance of each cell type in each sample. CODEFACS then predicts the cell type-specific gene expression profiles in each sample. It is a heuristic approach aimed at maximizing the number of genes in each cell type whose expression across the samples can be estimated as confidently predicted. Using 15 benchmark datasets where the ground-truth is known, we show that CODEFACS robustly improves over CIBERSORTx, both in terms of gene coverage and the individual gene expression estimation accuracy. We focus our comparisons on CIBERSORTx because it is widely used (cited more than 350 times as measured by Google Scholar since its publication in 2019) and to our knowledge, it is the only widely used package that makes sample-specific estimates of gene expression for each cell type. The essential methodological contributions of CODEFACS that distinguish CODEFACS from CIBERSORTx are: (i) a principled, quantitative definition of the confidence with which the expression of a (gene, cell type) has been predicted (which is highly important, if not essential, in any application) and (ii) two new methods of iterative refinement to leverage the high confidence predictions to improve predictions for genes predicted with low confidence at an earlier stage, implemented with a modular software engineering design.

We additionally introduce LIRICS (LIgand Receptor Interactions between Cell Subsets), that integrates the output of CODEFACS with a database of prior immunological knowledge that we curated to infer the cell-type-specific ligand-receptor pairs that are likely to interact in each sample. These data can then be analyzed in conjunction with any sample-associated clinical annotations (e.g., response to treatment) to systematically prioritize clinically relevant immune related interactions between any pair of cell types in each patient's cancer cohort. This opens the possibility of identifying phenotypic/clinical associations of individual interactions in large clinical tumor expression cohorts.

Building on the enhanced coverage and accuracy of CODEFACS, we applied it to reconstruct the cell-type-specific transcriptomes of ~8000 tumor samples from 21 cancer types in the Cancer Genome Atlas (TCGA). Analyzing these fully deconvolved TCGA expression datasets using LIRICS, we find a shared repertoire of cell-type-specific ligand-receptor interactions unique to the TME of mismatch repair deficient (MMRD) tumors from different tissues of origin. These interactions shed additional insights into the very high response rates of MMRD tumors to anti-PD1 treatment compared to other highly mutated tumors. Finally, using machine learning techniques, we identify a subset of intercellular TME interactions that stratify survival outcomes of melanoma patients receiving anti-PD1 therapy better than some recently published transcriptomics-based methods.

In summary, CODEFACS and LIRICS effectively build upon the statistical power of large bulk RNA-seq datasets and prior immunological knowledge to characterize sample specific cell-type-specific gene expression and learn more about the association of different tumor-immune interactions with different clinical measures. Notably, the potential scope of applications of both CODEFACS and LIRICS goes beyond studying the TME, as these tools can be applied to study any disease of interest given bulk gene expression data and relevant reference signatures of cell types involved.

Results

Overview of CODEFACS and LIRICS

CODEFACS is designed to characterize the tumor microenvironment by reconstructing the cell-type-specific transcriptomes of each sample from bulk expression. It takes as input the bulk RNA-seq expression values of a cohort of tumor samples and either the estimations of the cell fractions of a pre-defined set of cell types in each sample or their cell-type-specific molecular signature profiles, derived based on reference datasets or from the literature.

CODEFACS then employs a heuristic approach that sequentially executes three modules: (*module 1*) *high resolution deconvolution*, (*module 2*) *hierarchical deconvolution* and (*module 3*) *imputation* (See Figure. 1A and Supplementary Note 1). In module 1, we perform a high-resolution deconvolution, which extends the CIBERSORTx algorithm. In module 2 (hierarchical deconvolution), bulk expression is modeled as a mixture of two components: a specific cell type of interest and all the remaining cell types. The expression for the cell type of interest is predicted by removing the estimated expression in the second component (using high-resolution deconvolution from module 1) from the bulk mixture. In module 3 – imputation-based deconvolution, we impute the cell-type-specific expression of a specific gene based on the predicted cell-type-specific expression of other high-confidence genes that are co-expressed with that gene in the bulk. Each module is designed to overcome the shortcomings of its predecessor based on their respective modeling assumptions.

A key component of CODEFACS is its *confidence ranking system*, which receives cell-type-specific expression predictions from the different modules and labels them as high or low-confidence estimations. (For an overview of the rationale behind the confidence ranking system, see Methods). Genes whose expression is determined with high confidence in each module are added to the output set, while low confidence predictions are continued to be processed in subsequent modules. The final output of CODEFACS consists of two items: **(a)** a three-dimensional gene expression matrix, where each entry represents the predicted gene expression a gene in a given cell-type in a specific sample, and **(b)** a two-dimensional matrix of confidence scores ranging from [0,1] representing which gene-cell-type pairs have the most confident predictions (Figure. 1A, **Output**).

Given fully deconvolved gene expression data from CODEFACS, one can use LIRICS (Ligand Receptor Interactions between Cell Subsets) (Figure. 1B) to infer which cell-type-specific ligand-receptor pairs are likely to interact in each sample. Specifically, LIRICS takes the output of CODEFACS and processes it in three steps: (*Step 1*) The first step queries a database of all plausible ligand-receptor interactions between any two cell types

A and B, which we have systematically assembled and curated from the literature. This database is publicly available as part of LIRICS (see Supplementary Note 2, Supplementary Tables 1–3) and serves as a repository of all prior immunological knowledge. (*Step 2*) In the second step, given the deconvolved expression profiles of cell type A and cell type B in each bulk tumor sample, LIRICS denotes as ‘active’ or ‘likely to occur’ (‘1’) the interactions where both the ligand and receptor are over-expressed in the relevant cell-types in that sample, or otherwise ‘inactive’ (‘0’). A ligand or receptor is overexpressed in each cell type if its expression exceeds the median expression in that cell type (Supplementary Note 2). (*Step 3*) In the third step, a Fisher’s enrichment analysis is performed to test the association of the activity of specific ligand-receptor interactions with any relevant phenotypes of interest (e.g., treatment response, mutational subtype, etc.) (Figure. 1B).

Evaluating the performance of CODEFACS versus numerous different benchmarks

To assess the accuracy of CODEFACS, we generated 15 benchmark datasets (see Methods) by merging publicly available single-cell RNA-seq (9,10) and FACS sorted purified RNA-seq(11). Thereafter, we applied CODEFACS to deconvolve these generated bulk datasets and define the accuracy of its predictions by computing the Kendall correlation between the predicted and ground truth expression in each cell type across individual samples (the Kendall correlation provides a less inflated measure of accuracy by accounting for ties in the data). In the main text, we show the results obtained on three of these benchmark bulk datasets: one derived from a FACS-sorted lung cancer data(11), one from a single-cell melanoma RNA-seq dataset (9) and from a single-cell glioblastoma (GBM) RNA-seq dataset(10), respectively. Each bulk sample from these datasets represents a biopsy from a patient. We show that CODEFACS predicts the cell-type-specific expression of many more genes than CIBERSORTx (with Kendall’s correlation = 0.3) (Figure 2A–C), and its predictions are overall more accurate (Figure 2D–F). Furthermore, the estimated cell type specific expression profiles from CODEFACS recapitulate known tumor cell subtypes derived from single cell data (Supplementary Figure S1 demonstrates with the GBM dataset as an example). The results for all the remaining 12 benchmark datasets, created via artificial mixing of single-cell profiles and simulation of batch effects between bulk and single-cell expression data (See Methods and Supplementary Table S4 for more details), further show the superiority of CODEFACS (Supplementary Figure S2A–L and S3A–L). The comparisons based on different cutoffs (the baseline of 0.3, Kendall’s correlation = 0.4 and Kendall’s correlation = 0.5) also support our conclusion (Supplementary Figure S2A–L, S3A–L, S4A–O and S5A–O). Overall, we observe that the more abundant the cell type is, the better CODEFACS can predict its cell-type-specific gene expression (Supplementary Figure S6).

To test whether CODEFACS confidence scores can filter out potentially noisy predictions we quantified how well these scores align with the Kendall scores that measure the prediction accuracy with the ground truth for each (gene, cell type) pair, as described above. We quantified this using two metrics: Spearman correlation and a classification Area Under the ROC Curve (AUC). As evident, this correlation is strong and positive (Figure 2G depicts the results for the FACS sorted lung cancer benchmark dataset, and Supplementary Figure S7A–N depicts the results for the remaining benchmark datasets).

To perform a classification-based quantification, we grouped the genes in each cell type into two classes based on the correlation between their predicted and actual expression, as *informative* (Kendall score ≥ 0.1 and p-value ≤ 0.05) and *uninformative* (Kendall score < 0.1 or p-value > 0.05). We then tested whether the confidence scores could be used to classify genes into these two classes for each cell type. We found that the confidence score could effectively filter out uninformative predictions (Figure 2H depicts the results for the FACS sorted lung cancer benchmark dataset, and Supplementary Figure S8A–N, depicts for the remaining benchmark datasets). Supplementary Figure S9A–O visualizes the correlation between predicted and measured gene expression. This figure reveals that the predictions are generally less accurate for lowly expressed genes compared to highly expressed genes. Supplementary Figure S10A–O shows scatter plots of the confidence score (X-axis) and the measured expression (Y-axis). Confidence levels mildly increase with expression levels; the Spearman correlation coefficients have min = 0.07, median=0.23, max=0.75) across the 15 data sets.

To illustrate the biological information that may be gleaned from further analysis of the results using the confidence scores we performed a KEGG pathway enrichment analysis over the highest-confidence genes (confidence score = 1) in TCD8 cells of the SKCM benchmark dataset 1 and found 58 significantly (q-value < 0.1) enriched biological pathways, including the T cell receptor signaling pathway, Th1 and Th2 cell differentiation, PD-L1 expression and PD-1 checkpoint pathway in cancer, chemokine signaling pathway etc.. However, after permuting the confidence score profile across genes, none of those findings recur. The rationale to identify the enriched pathways in high confidence genes is that high-confidence genes are expected to be substantially expressed on specific types of cells, and thus cell-type-specific pathways are expected to be enriched in those genes.

We used the package clusterProfiler(12) to identify KEGG pathways that are significantly (q-value < 0.1) enriched for high-confidence genes. We found that the high-confidence genes involved in significantly enriched pathways exhibit higher expression. Supplementary Figure S11 shows that the difference in expression comparing the median expression values of all relevant high confidence genes to the median expression values of all other genes is significant, across all cancer types (two-sided Wilcoxon test, $p < 1e-14$, Bonferroni correction).

To test the robustness of CODEFACS to variations in cell fraction estimation, we performed deconvolution on two benchmark datasets (SKCM dataset 1 and GBM dataset 1) by providing the true cell fractions instead of predicted cell fractions as input to CODEFACS. The performance of CODEFACS for the two scenarios are not substantially different as Supplementary Figure S12A, B shows. Additionally, we revealed that the overall performance of CODEFACS is not strongly affected by random perturbations in cell fractions (Supplementary Figure S13A, B).

In addition to testing on real data sets, we simulated 3000 cells using R package “dyngen”(13) and performed cell clustering using R package “Rtsne”. Three cell clusters (see Supplementary Figure S14A, B) are defined as proxy cell types. After one fifth of random cells from each cell type are selected for cell type signature derivation, all other cells

are randomly assigned to sixteen proxy individuals to generate a bulk mixture dataset (1035 genes and 16 individual samples) with ground truth cell-type-specific expression (1035 genes and 3 cell types) and cell fractions (16 individual samples and 3 cell types). Thereafter the performance of CODEFACS and CIBERSORTx are compared based on this benchmark dataset. We conclude that CODEFACS is superior to CIBERSORTx based on this simulated benchmark dataset (see Supplementary Figure S15 and S16).

Finally, to evaluate CODEFACS on real bulk tumor data, we applied it to deconvolve bulk expression data from 21 cancer types (~8000 RNA-seq samples) in TCGA. To infer the cellular abundance of each cell type in each sample which is required as input for CODEFACS, we made use of matched bulk methylation data available for these samples and methylation-based reference signature profiles of distinct cell types. These include 11 cell type signatures (macrophages/dendritic cells--CD14+, B cells--CD19+, CD4+T cells, CD8+ T cells, T regulatory cells, NK cells--CD56+, endothelial cells, fibroblasts, neutrophils, basophils, eosinophils and tissue-specific tumor cells) obtained from MethylCIBERSORT(14). Reassuringly, we found high Spearman correlations between the resulting predicted tumor cell fraction and the tumor purity estimates derived from matched mutation and copy number data (based on ABSOLUTE(15)) for the same samples in each of 10 different cancer types (Spearman correlation: min=0.72, max=0.88, avg=0.8; Supplementary Figure S17A–J). The correlations are consistently high among the comparisons to other reported tumor purity estimates (based on ESTIMATE, LUMP, IHC and CPE(16,17); Supplementary Figure S18A–N, S19A–O, S20A–O, and S21A–O).

We then asked if CODEFACS can recover the cell-type-specific gene expression signature of different cell types in a given cancer type. To this end, we computed the Spearman correlation between (a) the mean deconvolved gene expression of the top confidently deconvolved genes in a given cell type (confidence score > 0.95) and (b) the mean expression of these genes, which we derived from completely independent single-cell expression data of the same cancer type (Methods). We find that (a) and (b) are substantially correlated (Figure 2I depicts results for the TCGA-LUAD (lung adenocarcinoma) dataset as an example and Supplementary Figure S22A–H for the remaining cancer types that have publicly available scRNA-seq data). The concordance level is higher for cell types that are abundant (e.g., tumor cells and fibroblasts) and decreases for less abundant cell types. Additionally, we observed that tumor cells have the largest fraction of genes whose expression is predicted with high confidence, with the highest in thyroid cancer (THCA, ~67.4% of all genes; Supplementary Figure S23A). Furthermore, seven KEGG pathways are significantly enriched (adjusted p-value < 0.1) with highly confident genes in tumor cells (confidence score > 0.95) across the 21 cancer types (Supplementary Figure S23B). Those pathways mostly involve mRNA surveillance, spliceosome, DNA replication, and mismatch repair.

We additionally assessed the correlation between mean confidence levels for tumor cell expression and the mean aneuploidy scores (data borrowed from (18)) across cancer types in the TCGA (Supplementary Figure S24A–C), finding that they are not significantly correlated. As an additional quality control test of our pseudo-bulk mixtures, Supplementary Fig. S25 shows a scatter plot of measured mean gene expression in malignant LUAD

single cells versus CODEFACS predicted gene expression for malignant cells in bulk TCGA LUAD samples.

Inference of cellular crosstalk linked with enhanced sensitivity of mismatch repair deficient tumors to anti-PD1 therapy

In normal cells, DNA is constantly repaired in response to DNA damage or DNA replication errors(19). However, defects in specific DNA repair pathways in cancer cells may result in the accumulation of many somatic mutations resulting in hypermutated tumors (TMB > 10 nonsynonymous mutation/Mb)(20–22). One cause of hypermutability is a mismatch repair deficiency (MMRD) that leads to the accumulation of insertions and deletion mutations in microsatellite regions of the genome due to uncorrected DNA replication polymerase slippage events. This is known as microsatellite instability (MSI)(23,24).

Solid tumors with mismatch repair deficiency were shown to be sensitive to immune checkpoint blockade (ICB) therapy irrespective of tumor type, leading the FDA to approve MSI as the first cancer type agnostic biomarker for patients receiving anti-PD1 treatment(25). Prior work has led to the prevailing hypothesis that elevated tumor mutation burden in mismatch repair deficient tumors leads to more neoantigens, and thus is more likely to activate a host immune response against tumor cells(23,26–28).

However, not all tumor types with elevated tumor mutation burden have similar response rates to anti-PD1(29,30) and recent studies have revealed that T cells recognize and respond to only a few neoantigens per tumor (31–34). In the TCGA collection, there are three solid tumor types with a significant association between hypermutability and survival benefit in patients. These three are the MSI enriched solid tumor types (Supplementary Figure S26A, B, borrowed from (18,35)), MSI tumors highlighted as red dots in Supplementary Figure S26A. Supplementary Figure S26C log-rank test p-value = 0.00084), but not in other tumor types (Supplementary Figure S26D, log-rank test p-value = 0.4). Most importantly, immunogenic indel mutations generated by MSI tumors need to escape nonsense mediated mRNA decay to be presented by the antigen presentation system and activate immune responses(36,37).

Despite these barriers, mismatch repair-deficient solid tumors have very high response rates to anti-PD1 therapy. We therefore searched for other tumor immune interactions that are possibly unique to the TME of MMRD tumors and could contribute to their strong immune responses. One key lead supporting this possibility comes from the work of (38), which shows that mismatch repair deficient tumors have elevated unfolded protein response (UPR). Recent work has shown that tumor cells with elevated UPR can initiate intercellular communication to spread the UPR to other cells in the TME, including immune cells (39)

To identify cell-cell interactions that are differentially active between MMRD tumors and non-MMRD tumors (red vs black dots in Supplementary Figure S26A, data borrowed from (35), we applied CODEFACS to deconvolve the bulk gene expression of all solid tumors from TCGA and integrated their predicted cell-type-specific gene expression levels with LIRICS. The top 50 interactions likely to occur in mismatch repair deficient solid tumors (ordered by FDR adjusted p-value) based on LIRICS are shown in a network in Figure 3A.

The complete enrichment table listing results for all plausible ligand-receptor interactions are provided in Supplementary Table S5. An odds ratio > 1 , implies the interaction is more likely to occur in mismatch-repair deficient solid tumors vs non-mismatch repair deficient solid tumors; odds ratio < 1 implies enrichment in the opposite direction. The top 50 interactions shown in Fig. 3A were tumor type-agnostic and less likely to occur in other hypermutated (but not mismatch repair deficient) tumors as depicted in Fig. 3B and Supplementary Figure S27A–F.

For each deconvolved solid tumor sample in the TCGA, we additionally computed a sample-specific FDR weighted interaction enrichment score linked with mismatch repair deficiency (MMRD crosstalk score; Methods), its non-synonymous tumor mutation burden and fraction of indel mutations, with the goal of comparing their values between MMRD samples (shown as red dots in Figure 3C, D and Supplementary Figure S26A) and non-MMRD samples (shown as black dots in Figure 3C, D and Supplementary Figure S26A) (Methods). The x-axis in Figure 3C records the non-synonymous mutation whereas the x-axis in Figure 3D records the fraction of indel mutations. The y-axis records those MMRD crosstalk scores. Overall, as expected, the MMRD crosstalk scores are strongly associated with mismatch repair deficiency status (Figure 3E; Wilcoxon test p-value $< 2.2\text{E-}16$). However, the associations of these scores with the non-synonymous tumor mutation burden or fraction of indel mutations are weak (Figure 3C Spearman's rho: 0.06, p-value: $2.4\text{E-}7$; Figure 3D Spearman's rho: -0.009 , p-value: 0.49).

Lastly, we fit a multivariable logistic regression model to the mismatch-repair deficiency status data across TCGA using as features the sample-specific MMRD crosstalk scores, the non-synonymous mutation burden and the fraction of frame-shifted/indel mutations. Figure 3F plots the estimated regression coefficients after transforming the values taken by each feature into Z scores. Overall we see that the sample-specific MMRD score is significantly associated with MMRD status, independent of other features, thus suggesting that the inferred ligand-receptor interaction repertoire is an additional unique feature of mismatch repair deficiency and could potentially be contributing to the exceptional response rates of MMRD solid tumors to anti-PD1 therapy.

The top 50 interactions from the differential analysis (Figure 3A) include the well-established PDL1-PD1 checkpoint interaction between tumor cells and CD8+ T cells, but additionally, numerous T-cell co-stimulatory interactions such as the 41BBL-41BB interaction between tumor cells and CD8+ T cells and ULBP2-NKG2D between tumor cells and CD4+ T cells. They also encompass several chemotaxis related interactions involved in trafficking of lymphocytes in and around the tumor mass, such as the CXCL9-CXCR3 chemokine interaction between macrophages and CD4+ T cells and CCL3/4/5-CCR5 interactions between various immune and stromal cell-types. These observations are supported by recently emerging results analyzing differences in the TME of 34 mismatch repair deficient vs 28 mismatch repair proficient colorectal cancers at a single-cell resolution (40). In this study, the analysis of core cell-type-specific gene expression programs upregulated specifically in mismatch repair deficient tumors revealed elevated expression levels of genes such as *PDI*, *CXCL13*, *GZMB* and *PRFI* in T cells indicative of chronic T cell stimulation (41–43) and elevated expression levels of T cell attracting chemokines

in malignant and myeloid cells. For instance, the T cell attracting chemokine *CXCL9* is among the top upregulated genes in cells of myeloid lineage and Pelka et al. (40) suggested that expression of *CXCL9*, *CXCL10*, *CXCL11* all of which bind to CXCR3 on T cells creates a positive feedback loop for increased T cell activity. Increased CXCR3 binding to its ligands has also been previously associated with response to anti-PD-1 treatment (44). These findings overlap with our differential analysis using LIRICS, wherein we observe a significant enrichment of *CXCL9*-CXCR3 chemokine interaction between macrophages and CD4⁺ T cells in mismatch repair deficient solid tumors across the TCGA collection (fold change: 2.58, FDR adjusted p-value: 2.38×10^{-10} , displayed in Figure 3A among the top 50 most significantly enriched ligand receptor interactions), showing that this phenomenon is not specific to colorectal mismatch repair deficient tumors.

Machine learning guided discovery of cell-type-specific ligand-receptor interactions that predict patient response to ICB therapy

It has been previously demonstrated in pan-cancer clinical trials that differences in mutation burden result in differences in survival upon receiving immune checkpoint blockade treatments leading to the recent FDA approval of TMB as a biomarker for anti-PD1 therapy (45). However open questions remain on the contribution of additional factors within individual cancer types (46,47). Hence to find additional tumor immune interactions predictive of response to immune checkpoint blockade therapy we chose to focus on studying TME differences between hypermutated tumors (> 10 non-synonymous mutations/Mb) and non-hypermutated tumors in TCGA data. We focus on melanoma, which has one of the best response rates to ICB, and where there are many independent publicly available bulk expression datasets of patients receiving anti-PD1 treatment. Starting from the deconvolved TCGA-SKCM dataset as our training set (N=469), we trained a genetic algorithm to find cell-type-specific ligand receptor interactions that are predictive of hypermutation in melanomas (see Methods, Figure 4A). For all analyses in this study, the definition of hypermutated tumor was set to any tumor with > 10 non-synonymous mutations per Megabase based on previous papers (20,21,48). In melanoma, these excess mutations primarily arise from UV radiation induced DNA damage and not replication slippage and mismatch repair deficiency (21). We term the interactions identified by the algorithm in this pretext task as *Melanoma hypermutation-Specific Functional Interactions (MSFI)*, and the network formed by these interactions is displayed in Figure 4B.

Having identified the MSFI interactions, we applied CODEFACS to deconvolve the bulk expression data of pre-treatment samples from the three largest publicly available melanoma datasets where patients received anti-PD1 treatment (either monotherapy or in combination with anti-CTLA4; Methods)(49–51). We then employed LIRICS (step 1 and step 2) to the respective deconvolved expression of each of these checkpoint datasets, and *without any additional training*, simply quantified the number of MSFI interactions that are active in each of these patients' tumor samples, which we denote as the tumor's *MSFI score*. Remarkably, we find that the MSFI score of each sample can robustly stratify patients into those that are likely to respond to ICB vs those that are unlikely to respond. Figure 4C, plots progression-free survival differences between low risk group (MSFI score > median value; N = 120) and high risk group (MSFI score ≤ median value; N = 124). The log rank test p-value

is: 0.00057. Figure 4D plots overall survival differences between the same two groups. The log rank test p-value is 0.0031. Supplementary Figure S28A–H plots the progression free survival and overall survival differences between the same two groups for each dataset separately. Importantly, a multivariable Cox-proportional hazards model with MSFI scores and TMB of each patient receiving anti-PD1 (wherever TMB data were available) shows that the MSFI score remains significantly associated with improved progression free survival (p-value: 0.013) and overall survival (p-value: 0.0258) after accounting for differences in mutation burden (Figure 4E, F), suggesting that the genetic algorithm learns how to predict immunotherapy outcomes without overfitting to TMB.

In addition, we plot Kaplan-Meier survival curves depicting the stratification performance of 7 other previously published bulk transcriptomics-based scores (52–60) besides MSFI based on their respective median values (Figure 5A–H and 6A–H). For TIDE (59) and MPS (60) scores, a patient is put into the high risk group if that patient's TIDE and MPS scores are > respective median values. This is because these scores were previously linked with immune resistance. Overall, for patients receiving anti-PD1 monotherapy, the MSFI score achieved the best survival stratification results, whereas for patients that are receiving or have previously received anti-CTLA4 in addition to anti-PD1, the IS score (58) performed best.

We further report the performance of the MSFI score in classifying (partial/complete responses) vs (stable/progressive) disease as determined by post-treatment radiological review. The MSFI score achieves an average AUC (Area Under receiver operator Curve) of 0.63 for this task across different datasets and treatment groups (The AUCs for each dataset and treatment group are reported in Supplementary Table S6). A similar performance could not be achieved if the placement of the ligand and receptor between interacting cell-types in the MSFI network was swapped (average AUC 0.58) or by randomly shuffling the interaction activity profiles (average AUC ~ 0.5), testifying that the selected cell-cell interactions are best predictive of response to ICB. AUCs for the seven other scores we benchmarked against are also provided in Supplementary Table S6. Overall, for this task cytotoxic signature score (55) achieved the best average performance (average AUC: 0.68).

We note that the performance levels of some previously published scores on more recent larger bulk expression datasets, where the original RNAseq reads have been uniformly aligned and normalized as described in the Methods, is different than what was expected from the original publications (Supplementary Table S6), pointing to the growing awareness of the sensitivity of expression-based predictors to batch effects, the normalization and alignment methods used, highlighting the need to pre-process the raw sequencing data uniformly in making meaningful comparisons (see Discussion).

We next examine the interactions in the MSFI network (Figure 4B). Supplementary Table S7 reports the number of times each interaction from the MSFI network was selected by the genetic algorithm over multiple training runs on the TCGA data (Methods). Overall, we find an over-representation of cell-type-specific co-stimulatory interactions from immunological literature (hypergeometric test p-value < 0.05, See Methods)(61–71), highlighting the importance of co-stimulation in mediating successful anti-tumor

immune responses. Additionally, the MSFI network includes various cytokine/chemokine interactions involved in trafficking of specific NK, T and B cell subsets to the tumor site.

Taken together, these results suggest that while recognition of melanoma tumor neoantigens serves as an initiating trigger, the activation of additional, cell type-specific, co-stimulatory signals in the TME are required to further enhance an effective host immune response upon immune checkpoint blockade treatments.

Discussion

This study presents a new computational tool, CODEFACS and a supporting immune interactions framework, LIRICS, that enable an (averaged) ‘virtual single-cell’ characterization of the TME from bulk tumor expression data. Applying these tools across the TCGA, we systematically identify, for the first time, cell-type-specific ligand-receptor pairs that are likely to interact directly in the TME of specific patient populations and prioritize those that are associated with patients’ response to ICB. Specifically, we identified a shared core of intercellular TME interactions unique to DNA mismatch repair deficient solid tumors that are associated with improved patient survival and enhanced sensitivity to immune checkpoint blockade therapy vs other highly mutated tumors. Finally, focusing on melanoma, we show that one can bootstrap on the large deconvolved data resource from TCGA using machine learning techniques to discover cell-cell interactions within the TME that successfully stratify patients’ responses to immune checkpoint blockade.

The heightened cellular crosstalk unique to the TME of mismatch repair deficient tumors suggests that T cells are being actively recruited to the tumor mass and activated by multiple co-stimulatory signals, in addition to immunogenic neo-antigens, only to be kept in balance by other immunoregulatory mechanisms such as the PD1-PDL1 checkpoint interaction between CD8+ T cells and tumor cells. This suggests that when this interaction is blocked by anti-PD1 treatment, additional co-stimulatory interactions such as the 41BBL-41BB or ULBP2-NKG2D interaction between tumor and T cells are likely to lead to the observed enhanced response of MMRD tumors to immune checkpoint blockade therapy. These results further support the investigation of the unfolded protein response in tumor cells or infiltration of specific immune cell subsets such as (NKG2D+, 41BB+, PD1+) T cells in the TME of patients. Notably, recent preclinical studies have shown that combination therapies dually targeting specific immune checkpoint and co-stimulatory interactions rescue anti-tumor immune responses in low TMB and highly immuno-suppressive settings(72–77). Currently, several clinical trials to assess the safety and efficacy of such combinations in humans are in progress (78–81).

While we have provided a toolkit to prioritize clinically relevant cell-cell interactions from bulk tumor expression, there are limitations that should be noted and potentially further improved upon in the future. First, several recent publications have reported discrepancies between different RNA-seq expression quantification pipelines based on the reference transcriptome version used, choice of method and normalization (e.g.: alignment-based vs alignment free) (82–86). This can potentially affect reproducibility of findings. Hence, whenever possible, we recommend that all bulk RNA-seq datasets are homogeneously

pre-processed using the same RNA-seq quantification method and reference transcriptome before the application of our tools. Indeed, following this notion, in this study we have pre-processed all bulk RNA-seq datasets in a uniform manner, using STAR (v2.7.6a) + RSEM (v1.3.3) and GENCODE v23 human genome annotation (Methods).

Second, CODEFACS itself has several limitations: (1) it requires prior information about the cell type composition of the input tumors, or alternatively, knowledge of the pertaining cell-types' gene expression or methylation signatures that can be used to infer their abundances, and its accuracy depends on the accuracy of the latter. (2) It is predictive typically for cell types with higher abundance and its performance deteriorates for lowly-abundant cell types. However, the confidence scores provided help by allowing the user to rank genes in each cell-type by the quality of predictions. During the revision of our manuscript, we became aware of another software package called BayesPrism(87) that solves the problem of estimating gene expression in malignant cells taking as input read counts from a bulk mixture of malignant and non-malignant cells. BayesPrism uses a Bayesian approach and Markov Chain Monte Carlo methods, which are completely different from the methods in CODEFACS. We found that the results of BayesPrism are comparable to CODEFACS, but very dependent on the input sample size. Because the problems solved are different and because of the BayesPrism dependence on sample size, we elected not to present a full head-to-head comparison.

Third, LIRICS is currently restricted to well-defined protein-protein interactions and does not consider the spatial localization of cells in the TME. The inclusion of the latter with the advent of forthcoming spatial transcriptomics data is likely to lead to considerably more informative interaction inference approaches. In the supplementary note, we have described how LIRICS fundamentally differs from previously published single cell-based approaches both in terms of its inputs and the problem being solved. We have further provided a brief walkthrough of how interested users can use LIRICS to conduct a meta-enrichment analysis of their own (like we demonstrated with the TCGA) with the example of the three large melanoma immune checkpoint blockade datasets studied in this work. In case one would like to do a more exploratory analysis, we provide a walkthrough of how this can be achieved via a joint analysis of single cell and corresponding deconvolved bulk data.

Although this work focuses on studying the tumor microenvironment, the tools presented here can be applied to prioritize important cell-cell interactions in noncancerous tissues under a variety of normal and disease states. One interesting application that we envision is the characterization of clinically relevant intercellular interactions occurring at the maternal-fetal interface using corresponding bulk gene expression data and pregnancy outcome information, whose elucidation may help treat and mitigate preeclampsia and other pregnancy related complications. One can also use our tools to study bulk gene expression data from pre-malignant tissue samples and compare them against malignant samples to elucidate cell-cell interaction dynamics on the way to malignancy. Finally, one can deconvolve expression data from autoimmune disorders to learn more about the underlying immune interactions. In sum, the computational tools developed and presented here offer a cost-effective way to study immune responses at a cell-type-specific resolution

from the widely abundant bulk gene expression in both health and disease, complementing first-line single-cell technologies in situations where the latter are less applicable.

Methods

Data collection and pre-processing

Single-cell RNA-seq datasets: To benchmark the performance of CODEFACS, we first set out to obtain publicly available single-cell RNA-seq datasets where both tumor and non-tumor cells were successfully isolated. This search led us to the identification of nine such single-cell RNA-seq datasets from the literature, each from a different cancer type. Collection of additional single-cell datasets was frozen after Dec 2019. Details of the single-cell datasets that we curated for this study in Supplementary Table S8. For each dataset sequenced on the SmartSeq2 platform, the log-normalized transcript counts for each gene in each sequenced cell were made publicly available by the original authors. For the application of deconvolution, these counts were transformed back to the Transcripts Per Million (TPM) scale. For datasets sequenced on the 10x platform, unique molecular identifier (UMI) counts for each gene were made publicly available and were scaled by the library size of each cell and multiplied by a factor of 1 million to get expression values in the TPM scale.

Summary of the CODEFACS algorithm and confidence ranking system

rationale: The CODEFACS algorithm is a greedy algorithm that is executed in three stages: 1. high resolution deconvolution with recursive splitting (module 1), 2. Hierarchical deconvolution (module 2), and 3. Imputation (module 3). Each of these stages aim to accurately estimate a gene's expression in a given cell type and sample based on some assumptions.

The fundamental rationale behind the confidence ranking system is that we expect to have high confidence in the estimated expression of a gene in a cell type if its expression distribution across samples satisfies as many stage-specific modeling assumptions as possible. This intuition is captured by computing multiple rankings of gene-cell type pairs based on different stage-specific confidence features that have been engineered (see Supplementary Table S9), and subsequently averaging over these rankings.

The aim of the confidence ranking cut-offs defined at each stage was to help the CODEFACS algorithm heuristically converge to a solution by forcing a distinction between gene-cell type pairs whose estimated expression levels are satisfactory (i.e., belonging to a high confidence set) or require further refinement (i.e., belonging to a low confidence set). The “greediness” of the algorithm arises from the fact that we only focus on refining the expression of gene-cell type pairs belonging to the low confidence set and that we choose cut-offs as conservatively as possible. After execution of all stages, a single confidence score is assigned to each gene-cell type pair. The sequential confidence building design of our approach are inspired by e.g., the quality scores that are assigned to reads generated from sequencing machines. The final confidence score (0–1) indicates the confidence level of each predicted gene in each cell type. We tend to assign higher confidence scores to those

genes in a specific cell type whose predicted expression levels are highly correlated with the high confidence gene set. A detailed description is provided in Supplementary Note 1.

Guide to some of the Supplementary Information concerning methods of

CODEFACS: The methods for CODEFACS and LIRICS are complex and have many stages. At the suggestion of a reviewer, we provide here some cross-reference to the Supplementary Information that should aid the reader to find specific information about different stages of the algorithms and some additional evaluations of performance not covered in Supplementary Figure S1–S28. Supplementary Note 1 describes in detail the algorithms in CODEFACS. A schematic description of CODEFACS module 1 can be found in Supplementary Figure S29. An evaluation of how confidence of gene expression predictions changes from module 1 to module 2 to module 3 can be found in Supplementary Figure S30A–O. Supplementary Figure S31 shows a schematic of the optional cell fraction estimation step in CODEFACS module 1. Supplementary Figure S32 shows a schematic of the optional batch correction step if CODEFACS module 1. Supplementary Figure S33 shows a schematic of the recursive splitting method in CODEFACS module 1.

Supplementary Note 1, part 3 delves further into the confidence ranking system, which relies on the assumption that the expression levels of functionally related genes in the same cell type are correlated and that this correlation can be used incrementally to boost confidence in the prediction of gene expression levels within a cell type. The correlation of gene expression levels of different genes within a cell type is quantified in Supplementary Figure S34A–C.

Supplementary Note 1, part 4 describes CODEFACS module 2, which is illustrated schematically in Supplementary Figure S35. Part 5 describes how confidence values are changed in module 2. Part 6 describes CODEFACS module 3, which is illustrated schematically in Supplementary Figure S36. Part 7 describes how confidence values are changed in CODEFACS module 3. Part 8 summarizes the three modules together and includes Supplementary Figure S37A–F that evaluates some of the algorithmic choices in the recursive splitting and sliding window steps of module 1. Part 8 also includes Supplementary Figure S38A–O that quantifies how gene expression prediction performance improves from module 1 to module 2 to module 3.

Bulk RNA-seq datasets: Gene expression and matching bulk tumor methylation data from fresh frozen tumor biopsies in TCGA were downloaded from (<http://xena.ucsc.edu/>) (88). In addition, publicly available bulk expression data from formalin fixed paraffin embedded tumor biopsies of melanoma patients receiving immune checkpoint blockade treatment were downloaded from (49–51). All bulk RNA-seq datasets were collected such that they have a sufficiently large sample size to reliably perform complete deconvolution of expression profiles (> 4 times the number of cell-types involved) (8). Collection of datasets was frozen after Dec 2019. Details on bulk RNA-seq datasets deconvolved in this study can be found in Supplementary Table S10. To maintain consistency with the pipeline used for preprocessing TCGA data, bulk gene expression levels in immune checkpoint blockade datasets were re-quantified using STAR v2.7.6a and RSEM v1.3.3 (89) with GENCODE v23 human genome annotation (90). Furthermore, to mitigate technical biases,

between-sample scaling factors were estimated using the TMM method implemented in edgeR (91) and TPM values in each sample were further rescaled by these scaling factors (92).

Bulk RNA-seq datasets used for benchmarking: To evaluate the performance of CODEFACS, we generated different pseudo-bulk RNA-seq datasets from mixing experiments with single-cell data. Each sample in each benchmark dataset has matching cell type specific gene expression profiles derived from averaging single-cell RNA-seq profiles of individual cells from the same sample and same cell type. For each pseudo-bulk sample, to derive the ground truth cell fraction we normalize the number of cells for each cell type by dividing by the total number of cells. These profiles serve as the ground truth for the evaluation of deconvolution performance. To avoid any circularity in our validations, for each of the single-cell datasets involved, single-cell data from 4 randomly chosen patients were separated from the rest. These data were used to derive reference gene expression signatures for each cell type. The mixing experiments were then performed on single-cell data of the remaining patients that were hidden from the reference signature derivation process. The datasets and the mixing experiment used to generate the pseudo-bulk samples are listed in Supplementary Table S4. In addition, we obtained a FACS sorted lung cancer dataset which includes purified RNA-seq for four cell types (11) and matched bulk RNA-seq for each sample.

To inject technical noise (e.g. PCR amplification bias or degradation of mRNA in preserved biopsies (92)) into the pseudo-bulk datasets, the following general procedure was followed:

For each bulk sample

1. Multiply the TPM count values of a random subset of genes with a random non-negative scalar value.
2. Re-normalize the count values of all genes in that sample to add up to a million (thereby satisfying the definition of TPM).

With the general procedure described above, here is how we generated each noisy replicate dataset with batch effects:

- For each sample in SKCM dataset 2, SKCM dataset 5, GBM dataset 2 and GBM dataset 5: artificially perturb the TPM values of 500 randomly chosen genes with randomly chosen multiplicative scalars in the range [50,100] to generate SKCM dataset 3, SKCM dataset 6, GBM dataset 3 and GBM dataset 6, respectively.
- For each sample in SKCM dataset 2, SKCM dataset 5, GBM dataset 2 and GBM dataset 5: artificially perturb the TPM values of 3000 randomly chosen genes with randomly chosen multiplicative scalars in the range [0,100] to generate SKCM dataset 4, SKCM dataset 7, GBM dataset 4 and GBM dataset 7, respectively.

For reproducibility, we fixed the seed in our implementations. The codes to generate these datasets are made available in the code repository. Additionally, we also generated datasets with noise injected into the cell fraction estimates and found that the overall

performance of CODEFACS is not strongly affected by random perturbations in cell fractions (Supplementary Fig. S13A, B).

Deconvolving bulk RNA-Seq datasets used for benchmarking with

CIBERSORTx: To perform high resolution deconvolution of bulk expression datasets used in benchmarking experiments using CIBERSORTx, we relied on their online portal (<https://cibersortx.stanford.edu/>) (8). We ran their impute cell expression module in high resolution analysis mode. To run this analysis, three inputs are required: bulk gene expression profiles (Mixture file), the cell type specific signature matrix (Signature Matrix file) and the names of genes whose expression needed to be deconvolved (we fed the names of all genes). In addition, Batch correction mode (B-mode) was check-marked to adjust for any batch effects between the signature expression matrix and bulk expression profiles. The outputs obtained are high resolution cell type specific expression values of each gene of interest in each cell type and sample.

Deconvolving immune checkpoint blockade (ICB) and TCGA bulk RNA-seq datasets using CODEFACS:

For the application of CODEFACS, molecular profiles of signature genes of each cell type of interest are needed to estimate the relative cell fractions in the bulk. We used single-cell-expression-derived signatures as priors to deconvolve the melanoma ICB datasets. To derive these signatures from single-cell data, we first obtained the class labels of each cell type of interest. These data are publicly available for each single-cell dataset we collected (9,10,93–98). Hence, we primarily used these labels in our study (unless further refinement of labels into specific cell subtypes of interest is needed for a specific usage). With a collection of single-cell expression profiles and matching cell type labels as input, we used CIBERSORT online tool to derive a cell-type-specific signature matrix. Thereafter, we applied CODEFACS to ICB datasets with default parameter settings and batch correction requirement specified. For TCGA deconvolution, we first estimated cell fractions based on bulk methylation and then applied CODEFACS to corresponding bulk gene expression for the 21 cancer types which have both types of data available. The main motivation to use DNA methylation to generate initial cell fraction estimates in TCGA is that the methylation signatures from MethylCIBERSORT (14) cover exactly the same set of cell types for all 21 cancer types in TCGA we analyzed (See Supplementary Table S11). In contrast, cell types covered by expression-based signatures (derived from publicly available single cell data) across cancer types are very heterogeneous due to design differences of the original studies. In addition, DNA methylation-based signatures are considered to be more stable marks of cellular identity compared to dynamic RNA-expression-derived signatures (99). The methylation-based cell type signatures were obtained from MethylCIBERSORT(14). We applied CODEFACS to TCGA datasets with default parameters settings and without batch correction option specified. For the details of CODEFACS algorithm and hyperparameter settings, see Supplementary Note 1.

Guide to some of the Supplementary Information concerning methods of

LIRICS: Supplementary Note 2

Supplementary Note 2 describes the methods of LIRICS and includes a tutorial related to the MMRD analysis in Results above. We begin with a description of the literature search and

curation of the ligand-receptor data used in LIRICS. Second, we summarize LIRICS steps 1, 2, and 3. Third, we compare the problems that LIRICS solves to what can be done with other methods such as CellPhoneDb. Fourth and finally, we give a detailed tutorial whose two key outputs of predicted ligand-receptor interactions are illustrated in Supplementary Figure S39A, B and S40A, B.

Defining sample specific MMRD crosstalk scores: The MMRD-specific ligand-receptor interaction enrichment score for a given tumor sample j was defined as:

$$MMRD\ crosstalk\ score_j = mean_i(-\log_{10}FDR[i] \times \mathbf{1}_{ij})$$

Where $\mathbf{1}_{ij}$ is an indicator variable that takes value 1 if ligand-receptor interaction i tends to occur with greater frequency in mismatch repair deficient tumors and is observed as 'activated' in the given sample j (0 otherwise) (See LIRICS step 2).

Machine learning for discovery of cell-cell interactions predictive of clinical responses to immune checkpoint blockade: The genetic algorithm is a randomized heuristic search algorithm designed to select optimal features for a prediction task given some user-defined fitness function for training (100). For the genetic algorithm, the training set is from TCGA, and the test sets are the three ICB data sets. Overall, the features of interest are ligand-receptor interactions between cell-types, and the prediction task is predicting response to ICB treatment. The fitness function guides the search for predictive cell-cell interaction features, which predict the hypermutation status of a sample (TMB > 10 mutations per Mb) based on the total number of these features that are observed as activated in a given (non-treated, TCGA melanoma) sample. Prediction accuracy is quantified by the area under the ROC curve (AUC). The overall feature-space consists of 4288 possible ligand-receptor interactions between 10 cell types in the melanoma TME. To minimize the risk of overfitting to the TCGA dataset and to aid in faster convergence of the genetic algorithm during training, the size of the feature space is first reduced by filtering out complex and harder to measure interactions involving receptors/ligands encoded by multiple genes, interactions among deconvolved cell types that are not shared between train and test datasets and interactions with fold change in frequency of activation between hypermutated vs non-hypermutated TCGA melanoma samples ≥ 1 . A total of 769 ligand-receptor interactions passed these filters and compose the search space of the genetic algorithm.

The genetic algorithm then proceeds by randomly selecting subsets of features from the set of 769 interactions, defined as the seed population. These subsets are iteratively modified to maximize the fitness function specified above, following standard evolutionary algorithms practices, adding or deleting features aiming to gradually increase the fitness function until a (local) optimum is converged upon. Since the fitness function landscape is often non-convex and the training process is stochastic, we repeat the training process 500 times, each with a randomly chosen seed population. We then choose the most frequently selected features over all 500 solutions to approximate a solution that is likely to be closer to the global optimum solution and less likely to overfit. For this aggregate feature selection one can estimate

the empirical p-value associated with each feature being selected more than the observed number of times by random chance. This is done by repeatedly shuffling 1000 times the values of the binary solution vector returned by the genetic algorithm for each training run. Given 1000 random permutations of 500 solutions, one can then derive a null distribution representing the number of times a feature is selected across all 500 solutions by random chance. The empirical p-value for each feature j is then estimated as:

$$\frac{1}{1000} \sum_{i=1}^{1000} \mathbf{1}_{\text{null frequency}_{ij} > \text{observed frequency}_{ij}}$$

We found that features with empirical p-value < 0.01 were selected > 100 times by the genetic algorithm. Overall, 103 interactions were selected > 100 times and constitute the final MSFI network displayed in Figure 4B. The downstream results are qualitatively similar for more stringent thresholds. The genetic algorithm was implemented in R using the `genalg` package using the following default hyperparameter settings. (Population size = 200, number of training epochs = 100, mutation probability = 0.001, elitism = True). A more detailed description of the pros and cons of this approach vs other feature selection approaches has been provided in Supplementary Note 2.

Hypergeometric test for over representation of costimulatory interactions in MSFI network

The overrepresentation p-value for costimulatory interactions is calculated as follows. Let $S = 103$ be the number of MSFI interactions selected by the genetic algorithm. The foreground $F = 103$ is the number of all plausible interactions between deconvolved cell types common to both train and test sets and costimulatory. The background $B = 2004 - 103$ is the number of all plausible interactions between deconvolved cell types common to both train and test sets but not co-stimulatory. The overlap O , between the foreground and the selection is $= 11$. The hypergeometric enrichment p-value for having O costimulatory interactions selected by random chance is then calculated as $1 - \text{phyper}(O-1, F, B, S)$ (The phyper function is implemented in R stats package).

Data and Code availability

All deconvolved expression data generated in this study, codes and tools (CODEFACS and LIRICS) are available via ZENODO repository: <https://zenodo.org/record/5790343>.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research is supported in part by the Intramural Research Program of the National Institutes of Health, National Cancer Institute, Center for Cancer Research. This work utilized the computational resources of the NIH HPC Biowulf cluster. The results here are in part based upon data generated by the TCGA Research Network: <https://www.cancer.gov/tcga>. We would additionally like to acknowledge Dr. Noam Auslander, Dr. Chi-Ping Day, Dr. Di Wu and other members of the Cancer Data Science Laboratory for their helpful feedback on this work.

References:

1. Paget S The distribution of secondary growths in cancer of the breast. *Lancet* 1889;133(3421):571–3.
2. Smyth MJ, Ngio SF, Ribas A, Teng MW. Combination cancer immunotherapies tailored to the tumour microenvironment. *Nat Rev Clin Oncol* 2016;13(3):143–58. [PubMed: 26598942]
3. Wagner A, Regev A, Yosef N. Revealing the vectors of cellular identity with single-cell genomics. *Nat Biotechnol* 2016;34(11):1145–60. [PubMed: 27824854]
4. Candès EJ, Wakin MB. An introduction to compressive sampling. *IEEE Sig Proc Mag* 2008;25(2):21–30.
5. Ahn J, Yuan Y, Parmigiani G, Suraokar MB, Diao LX, Wistuba II, et al. DeMix: deconvolution for mixed cancer transcriptomes using raw measured data. *Bioinformatics* 2013;29(15):1865–71. [PubMed: 23712657]
6. Quon G, Haider S, Deshwar AG, Cui A, Boutros PC, Morris Q. Computational purification of individual tumor gene expression profiles leads to significant improvements in prognostic prediction. *Genome Med* 2013;5(3):29. [PubMed: 23537167]
7. Fox NS, Haider S, Harris AL, Boutros PC. Landscape of transcriptomic interactions between breast cancer and its microenvironment. *Nat Commun* 2019;10:3116. [PubMed: 31308365]
8. Newman AM, Steen CB, Liu CL, Gentles AJ, Chaudhuri AA, Scherer F, et al. Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37(7):773–82. [PubMed: 31061481]
9. Jerby-Arnon L, Shah P, Cuoco MS, Rodman C, Su MJ, Melms JC, et al. A cancer cell program promotes T cell exclusion and resistance to checkpoint blockade. *Cell* 2018;175(4):984–97 e24. [PubMed: 30388455]
10. Neftel C, Laffy J, Filbin MG, Hara T, Shore ME, Rahme GJ, et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* 2019;178(4):835–49 e21. [PubMed: 31327527]
11. Gentles AJ, Hui ABY, Feng WG, Azizi A, Nair RV, Bouchard G, et al. A human lung tumor microenvironment interactome identifies clinically relevant cell-type cross-talk. *Genome Biol* 2020;21(1):107. [PubMed: 32381040]
12. Yu GC, Wang LG, Han YY, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics* 2012;16(5):284–7. [PubMed: 22455463]
13. Cannoodt R, Saelens W, Deconinck L, Saeys Y. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. *Nat Commun* 2021;12:3942. [PubMed: 34168133]
14. Chakravarthy A, Furness A, Joshi K, Ghorani E, Ford K, Ward MJ, et al. Pan-cancer deconvolution of tumour composition using DNA methylation. *Nat Commun* 2018;9:3220. [PubMed: 30104673]
15. Carter SL, Cibulskis K, Helman E, McKenna A, Shen H, Zack T, et al. Absolute quantification of somatic DNA alterations in human cancer. *Nat Biotechnol* 2012;30(5):413–21. [PubMed: 22544022]
16. Yoshihara K, Shahmoradgoli M, Martinez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. *Nat Commun* 2013;4:2612. [PubMed: 24113773]
17. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nat Commun* 2015;6:8971. [PubMed: 26634437]
18. Taylor AM, Shih J, Ha G, Gao GF, Zhang X, Berger AC, et al. Genomic and functional approaches to understanding cancer aneuploidy. *Cancer Cell* 2018;33(4):676–89 e3. [PubMed: 29622463]
19. Drake JW, Charlesworth B, Charlesworth D, Crow JF. Rates of spontaneous mutation. *Genetics* 1998;148(4):1667–86. [PubMed: 9560386]
20. Loeb LA. Mutator Phenotype May Be Required for Multistage Carcinogenesis. *Cancer Res* 1991;51(12):3075–9. [PubMed: 2039987]
21. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SAJR, Behjati S, Biankin AV, et al. Signatures of mutational processes in human cancer. *Nature* 2013;500(7463):415–421. [PubMed: 23945592]

22. Muzny DM, Bainbridge MN, Chang K, Dinh HH, Drummond JA, Fowler G, et al. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 2012;487(7407):330–7. [PubMed: 22810696]
23. Funkhouser WK, Lubin IM, Monzon FA, Zehnbauser BA, Evans JP, Ogino S, et al. Relevance, pathogenesis, and testing algorithm for mismatch repair-defective colorectal carcinomas a report of the association for molecular pathology. *J Mol Diagn* 2012;14(2):91–103. [PubMed: 22260991]
24. Cortes-Ciriano I, Lee S, Park WY, Kim TM, Park PJ. A molecular portrait of microsatellite instability across multiple cancers. *Nat Commun* 2017;8:15180. [PubMed: 28585546]
25. Boyiadzis MM, Kirkwood JM, Marshall JL, Pritchard CC, Azad NS, Gulley JL. Significance and implications of FDA approval of pembrolizumab for biomarker-defined disease. *J Immunother Cancer* 2018;6:137. [PubMed: 30514386]
26. Galon J, Costes A, Sanchez-Cabo F, Kirilovsky A, Mlecnik B, Lagorce-Page C, et al. Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 2006;313(5795):1960–4. [PubMed: 17008531]
27. Llosa NJ, Cruise M, Tam A, Wicks EC, Hechenbleikner EM, Taube JM, et al. The vigorous immune microenvironment of microsatellite instable colon cancer is balanced by multiple counter-inhibitory checkpoints. *Cancer Discov* 2015;5(1):43–51. [PubMed: 25358689]
28. Le DT, Uram JN, Wang H, Bartlett BR, Kemberling H, Eyring AD, et al. PD-1 blockade in tumors with mismatch-repair deficiency. *New Engl J Med* 2015;372(26):2509–20. [PubMed: 26028255]
29. Gurjao C, Tsukrov D, Imakaev M, Luquette LJ, Mirny LA. Limited evidence of tumour mutational burden as a biomarker of response to immunotherapy. *bioRxiv* 2020:2020.09.03.260265.
30. Yarchoan M, Hopkins A, Jaffee EM. Tumor mutational burden and response rate to PD-1 inhibition. *New Engl J Med* 2017;377(25):2500–1. [PubMed: 29262275]
31. Robbins PF, Lu YC, El-Gamil M, Li YF, Gross C, Gartner J, et al. Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat Med* 2013;19(6):747–52. [PubMed: 23644516]
32. Bassani-Sternberg M, Braunlein E, Klar R, Engleitner T, Sinitcyn P, Audehm S, et al. Direct identification of clinically relevant neoepitopes presented on native human melanoma tissue by mass spectrometry. *Nat Commun* 2016;7:13404. [PubMed: 27869121]
33. Roh W, Chen PL, Reuben A, Spencer CN, Prieto PA, Miller JP, et al. Integrated molecular analysis of tumor biopsies on sequential CTLA-4 and PD-1 blockade reveals markers of response and resistance. *Sci Transl Med* 2017;9(379):eaah3560. [PubMed: 28251903]
34. Kalaora S, Wolf Y, Feferman T, Barnea E, Greenstein E, Reshef D, et al. Combined analysis of antigen presentation and T-cell recognition reveals restricted immune responses in melanoma. *Cancer Discov* 2018;8(11):1366–75. [PubMed: 30209080]
35. Bonneville R, Krook MA, Kautto EA, Miya J, Wing MR, Chen HZ, et al. Landscape of microsatellite instability across 39 cancer types. *JCO Precis Oncol* 2017;1:PO.17.00073.
36. Lindeboom RG, Vermeulen M, Lehner B, Supek F. The impact of nonsense-mediated mRNA decay on genetic disease, gene editing and cancer immunotherapy. *Nat Genet* 2019;51(11):1645–51. [PubMed: 31659324]
37. Litchfield K, Reading JL, Lim EL, Xu H, Liu P, Al-Bakir M, et al. Escape from nonsense-mediated decay associates with anti-tumor immunogenicity. *Nat Commun* 2020;11:3800. [PubMed: 32733040]
38. McGrail DJ, Garnett J, Yin J, Dai H, Shih DJH, Lam TNA, et al. Proteome instability is a therapeutic vulnerability in mismatch repair-deficient cancer. *Cancer Cell* 2020;37(3):371–86 e12. [PubMed: 32109374]
39. Rodvold JJ, Chiu KT, Hiramatsu N, Nussbacher JK, Galimberti V, Mahadevan NR, et al. Intercellular transmission of the unfolded protein response promotes survival and drug resistance in cancer cells. *Sci Signal* 2017;10(482):eaah7177. [PubMed: 28588081]
40. Pelka K, Hofree M, Chen JH, Sarkizova S, Pirl JD, Jorgji V, et al. Spatially organized multicellular immune hubs in human colorectal cancer. *Cell* 2021;184(18):4734–52 e20. [PubMed: 34450029]
41. Thommen DS, Koelzer VH, Herzig P, Roller A, Trefny M, Dimeloe S, et al. A transcriptionally and functionally distinct PD-1+ CD8+ T cell pool with predictive potential in non-small-cell lung cancer treated with PD-1 blockade. *Nat Med* 2018;24(7):994–1004. [PubMed: 29892065]

42. Gros A, Robbins PF, Yao X, Li YF, Turcotte S, Tran E, et al. PD-1 identifies the patient-specific CD8(+) tumor-reactive repertoire infiltrating human tumors. *J Clin Invest* 2014;124(5):2246–59. [PubMed: 24667641]
43. Simoni Y, Becht E, Fehlings M, Loh CY, Koo SL, Teng KWW, et al. Bystander CD8+ T cells are abundant and phenotypically distinct in human tumour infiltrates. *Nature* 2018;557(7706):575–9. [PubMed: 29769722]
44. Chow MT, Ozga AJ, Servis RL, Frederick DT, Lo JA, Fisher DE, et al. Intratumoral activity of the CXCR3 chemokine system is required for the efficacy of anti-PD-1 therapy. *Immunity* 2019;50(6):1498–512 e5. [PubMed: 31097342]
45. Marabelle A, Fakih M, Lopez J, Shah M, Shapira-Frommer R, Nakagawa K, et al. Association of tumour mutational burden with outcomes in patients with advanced solid tumours treated with pembrolizumab: prospective biomarker analysis of the multicohort, open-label, phase 2 KEYNOTE-158 study. *Lancet Oncol* 2020;21(10):1353–65. [PubMed: 32919526]
46. Subbiah V, Solit DB, Chan TA, Kurzrock R. The FDA approval of pembrolizumab for adult and pediatric patients with tumor mutational burden (TMB) ≥ 10 : a decision centered on empowering patients and their physicians. *Ann Oncol* 2020;31(9):1115–8. [PubMed: 32771306]
47. Strickler JH, Hanks BA, Khasraw M. Tumor mutational burden as a predictor of immunotherapy response: is more always better? *Clin Cancer Res* 2021;27(5):1236–41. [PubMed: 33199494]
48. Cancer Genome Atlas Research N, Weinstein JN, Collisson EA, Mills GB, Shaw KR, Ozenberger BA, et al. The Cancer Genome Atlas Pan-Cancer analysis project. *Nat Genet* 2013;45(10):1113–20. [PubMed: 24071849]
49. Riaz N, Havel JJ, Makarov V, Desrichard A, Urba WJ, Sims JS, et al. Tumor and microenvironment evolution during immunotherapy with Nivolumab. *Cell* 2017;171(4):934–49 e16. [PubMed: 29033130]
50. Gide TN, Quek C, Menzies AM, Tasker AT, Shang P, Holst J, et al. Distinct immune cell populations define response to anti-PD-1 monotherapy and anti-PD-1/Anti-CTLA-4 combined therapy. *Cancer Cell* 2019;35(2):238–55 e6. [PubMed: 30753825]
51. Liu D, Schilling B, Liu D, Sucker A, Livingstone E, Jerby-Arnon L, et al. Integrative molecular and clinical modeling of clinical outcomes to PD1 blockade in patients with metastatic melanoma. *Nat Med* 2019;25(12):1916–27. [PubMed: 31792460]
52. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med* 2018;24(10):1545–9. [PubMed: 30127394]
53. Ayers M, Lunceford J, Nebozhyn M, Murphy E, Loboda A, Kaufman DR, et al. IFN- γ -related mRNA profile predicts clinical response to PD-1 blockade. *J Clin Invest* 2017;127(8):2930–40. [PubMed: 28650338]
54. Cristescu R, Mogg R, Ayers M, Albright A, Murphy E, Yearley J, et al. Pan-tumor genomic biomarkers for PD-1 checkpoint blockade-based immunotherapy. *Science* 2018;362(6411):eaar3593. [PubMed: 30309915]
55. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* 2017;355(6322):eaaf8399. [PubMed: 28104840]
56. Steiniche T, Rha SY, Chung HC, Georgsen JB, Ladekar M, Nordsmark M, et al. T-cell-inflamed gene expression profile (GEP) and PD-L1 expression in patients (pts) with esophageal cancer (EC). *J Clin Oncol* 2019;37(4_suppl):26.
57. Fehrenbacher L, Spira A, Ballinger M, Kowanetz M, Vansteenkiste J, Mazieres J, et al. Atezolizumab versus docetaxel for patients with previously treated non-small-cell lung cancer (POPLAR): a multicentre, open-label, phase 2 randomised controlled trial. *Lancet* 2016;387(10030):1837–46. [PubMed: 26970723]
58. Ock CY, Hwang JE, Keam B, Kim SB, Shim JJ, Jang HJ, et al. Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers. *Nat Commun* 2017;8(1):1050. [PubMed: 29051489]
59. 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(10):1550–8. [PubMed: 30127393]

60. Perez-Guijarro E, Yang HH, Araya RE, El Meskini R, Michael HT, Vodnala SK, et al. Multimodel preclinical platform predicts clinical response of melanoma to immunotherapy. *Nat Med* 2020;26(5):781–91. [PubMed: 32284588]
61. Billadeau DD, Leibson PJ. ITAMs versus ITIMs: striking a balance during cell regulation. *J Clin Invest* 2002;109(2):161–8. [PubMed: 11805126]
62. Staub E, Rosenthal A, Hinzmann B. Systematic identification of immunoreceptor tyrosine-based inhibitory motifs in the human proteome. *Cell Signal* 2004;16(4):435–56. [PubMed: 14709333]
63. Varin MM, Le Pottier L, Youinou P, Saulep D, Mackay F, Pers JO. B-cell tolerance breakdown in Sjogren's syndrome: focus on BAFF. *Autoimmun Rev* 2010;9(9):604–8. [PubMed: 20457281]
64. Murphy K, Weaver C. Janeway's immunobiology. Garland Science; New York, 2016.
65. Chen L, Flies DB. Molecular mechanisms of T cell co-stimulation and co-inhibition. *Nat Rev Immunol* 2013;13(4):227–42. [PubMed: 23470321]
66. Campbell KS, Purdy AK. Structure/function of human killer cell immunoglobulin-like receptors: lessons from polymorphisms, evolution, crystal structures and mutations. *Immunology* 2011;132(3):315–25. [PubMed: 21214544]
67. Pende D, Falco M, Vitale M, Cantoni C, Vitale C, Munari E, et al. Killer Ig-like receptors (KIRs): Their role in NK cell modulation and developments leading to their clinical exploitation. *Front Immunol* 2019;10:1179. [PubMed: 31231370]
68. Ward-Kavanagh LK, Lin WW, Sedy JR, Ware CF. The TNF Receptor Superfamily in Co-stimulating and Co-inhibitory Responses. *Immunity* 2016;44(5):1005–19. [PubMed: 27192566]
69. Gonçalves CM, Henriques SN, Santos RF, Carmo AM. CD6, a rheostat-type signalosome that tunes T cell activation. *Front Immunol* 2018;9:2994. [PubMed: 30619347]
70. Steri M, Orru V, Idda ML, Pitzalis M, Pala M, Zara I, et al. Overexpression of the cytokine BAFF and autoimmunity risk. *N Engl J Med* 2017;376(17):1615–26. [PubMed: 28445677]
71. Chen M, Lin X, Liu Y, Li Q, Deng Y, Liu Z, et al. The function of BAFF on T helper cells in autoimmunity. *Cytokine Growth Factor Rev* 2014;25(3):301–5. [PubMed: 24411564]
72. Chen S, Lee LF, Fisher TS, Jessen B, Elliott M, Evering W, et al. Combination of 4–1BB agonist and PD-1 antagonist promotes antitumor effector/memory CD8 T cells in a poorly immunogenic tumor model. *Cancer Immunol Res* 2015;3(2):149–60. [PubMed: 25387892]
73. Ma HS, Poudel B, Torres ER, Sidhom JW, Robinson TM, Christmas B, et al. A CD40 agonist and PD-1 antagonist antibody reprogram the microenvironment of nonimmunogenic tumors to allow T-cell-mediated anticancer activity. *Cancer Immunol Res* 2019;7(3):428–42. [PubMed: 30642833]
74. Barber A Costimulation of effector CD8+ T cells: Which receptor is optimal for immunotherapy? *MOJ Immunology* 2014;1(2):00011.
75. Melero I, Hirschhorn-Cymerman D, Morales-Kastresana A, Sanmamed MF, Wolchok JD. Agonist antibodies to TNFR molecules that costimulate T and NK cells. *Clin Cancer Res* 2013;19(5):1044–53. [PubMed: 23460535]
76. Chester C, Sanmamed MF, Wang J, Melero I. Immunotherapy targeting 4–1BB: mechanistic rationale, clinical results, and future strategies. *Blood* 2018;131(1):49–57. [PubMed: 29118009]
77. Philipson BI, O'Connor RS, May MJ, June CH, Albelda SM, Milone MC. 4–1BB costimulation promotes CAR T cell survival through noncanonical NF-kappaB signaling. *Sci Signal* 2020;13(625):eaay8248. [PubMed: 32234960]
78. Siu LL, Steeghs N, Meniawy T, Joerger M, Spratlin JL, Rottey S, et al. Preliminary results of a phase I/IIa study of BMS-986156 (glucocorticoid-induced tumor necrosis factor receptor–related gene [GTR] agonist), alone and in combination with nivolumab in pts with advanced solid tumors. *J Clin Oncol* 2017;35(15):104.
79. Tolcher AW, Sznol M, Hu-Lieskovan S, Papadopoulos KP, Patnaik A, Rasco DW, et al. Phase Ib study of utomilumab (PF-05082566), a 4–1BB/CD137 agonist, in combination with Pembrolizumab (MK-3475) in patients with advanced solid tumors. *Clin Cancer Res* 2017;23(18):5349–57. [PubMed: 28634283]
80. Cohen EEW, Pishvaian MJ, Shepard DR, Wang D, Weiss J, Johnson ML, et al. A phase Ib study of utomilumab (PF-05082566) in combination with mogamulizumab in patients with advanced solid tumors. *J Immunother Cancer* 2019;7(1):342. [PubMed: 31801624]

81. Choi Y, Shi Y, Haymaker CL, Naing A, Ciliberto G, Hajjar J. T-cell agonists in cancer immunotherapy. *J Immunother Cancer* 2020;8(2):e000966. [PubMed: 33020242]
82. Wu DC, Yao J, Ho KS, Lambowitz AM, Wilke CO. Limitations of alignment-free tools in total RNA-seq quantification. *BMC Genomics* 2018;19:510. [PubMed: 29969991]
83. Sahraeian SME, Mohiyuddin M, Sebra R, Tilgner H, Afshar PT, Au KF, et al. Gaining comprehensive biological insight into the transcriptome by performing a broad-spectrum RNA-seq analysis. *Nat Commun* 2017;8:59. [PubMed: 28680106]
84. Everaert C, Luybaert M, Maag JLV, Cheng QX, Dinger ME, Hellemans J, et al. Benchmarking of RNA-sequencing analysis workflows using whole-transcriptome RT-qPCR expression data. *Scientific Reports* 2017;7:1559. [PubMed: 28484260]
85. Teng M, Love MI, Davis CA, Djebali S, Dobin A, Graveley BR, et al. A benchmark for RNA-seq quantification pipelines. *Genome Biol* 2016;17:74. [PubMed: 27107712]
86. Robert C, Watson M. Errors in RNA-Seq quantification affect genes of relevance to human disease. *Genome Biol* 2015;16:177. [PubMed: 26335491]
87. Chu T, Danko C. Bayesian Inference of Cell Composition and Gene Expression Reveals Tumor-Microenvironment Interactions. 2020. bioRxiv 2020.01.07.897900; doi: 10.1101/2020.01.07.897900
88. Goldman MJ, Craft B, Hastie M, Repecka K, McDade F, Kamath A, et al. Visualizing and interpreting cancer genomics data via the Xena platform. *Nat Biotechnol* 2020;38(6):675–8. [PubMed: 32444850]
89. Bray NL, Pimentel H, Melsted P, Pachter L. Near-optimal probabilistic RNA-seq quantification. *Nat Biotechnol* 2016;34(5):525–7. [PubMed: 27043002]
90. Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, et al. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* 2012;22(9):1760–74. [PubMed: 22955987]
91. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26(1):139–40. [PubMed: 19910308]
92. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* 2010;11(3):R25. [PubMed: 20196867]
93. Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-cell transcriptomic analysis of primary and metastatic tumor ecosystems in head and neck Cancer. *Cell* 2017;171(7):1611–24 e24. [PubMed: 29198524]
94. Peng J, Sun BF, Chen CY, Zhou JY, Chen YS, Chen H, et al. Single-cell RNA-seq highlights intra-tumoral heterogeneity and malignant progression in pancreatic ductal adenocarcinoma. *Cell Res* 2019;29(9):725–38. [PubMed: 31273297]
95. Ma L, Hernandez MO, Zhao Y, Mehta M, Tran B, Kelly M, et al. Tumor Cell Biodiversity Drives Microenvironmental Reprogramming in Liver Cancer. *Cancer Cell* 2019;36(4):418–30 e6. [PubMed: 31588021]
96. Li H, Courtois ET, Sengupta D, Tan Y, Chen KH, Goh JJJ, et al. Reference component analysis of single-cell transcriptomes elucidates cellular heterogeneity in human colorectal tumors. *Nat Genet* 2017;49(5):708–18. [PubMed: 28319088]
97. Lambrechts D, Wauters E, Boeckx B, Aibar S, Nittner D, Burton O, et al. Phenotype molding of stromal cells in the lung tumor microenvironment. *Nat Med* 2018;24(8):1277–89. [PubMed: 29988129]
98. Karaayvaz M, Cristea S, Gillespie SM, Patel AP, Mylvaganam R, Luo CC, et al. Unravelling subclonal heterogeneity and aggressive disease states in TNBC through single-cell RNA-seq. *Nat Commun* 2018;9:3588. [PubMed: 30181541]
99. Bird A DNA methylation patterns and epigenetic memory. *Genes Dev* 2002;16(1):6–21. [PubMed: 11782440]
100. Mitchell M Genetic algorithms: An overview. 1995. Citeseer. p 31–9.

Statement of Significance

This work presents two new computational methods that can deconvolve a large collection of bulk tumor gene expression profiles into their respective cell type specific gene expression profiles and identify cell type specific ligand-receptor interactions predictive of response to immune checkpoint blockade therapy.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

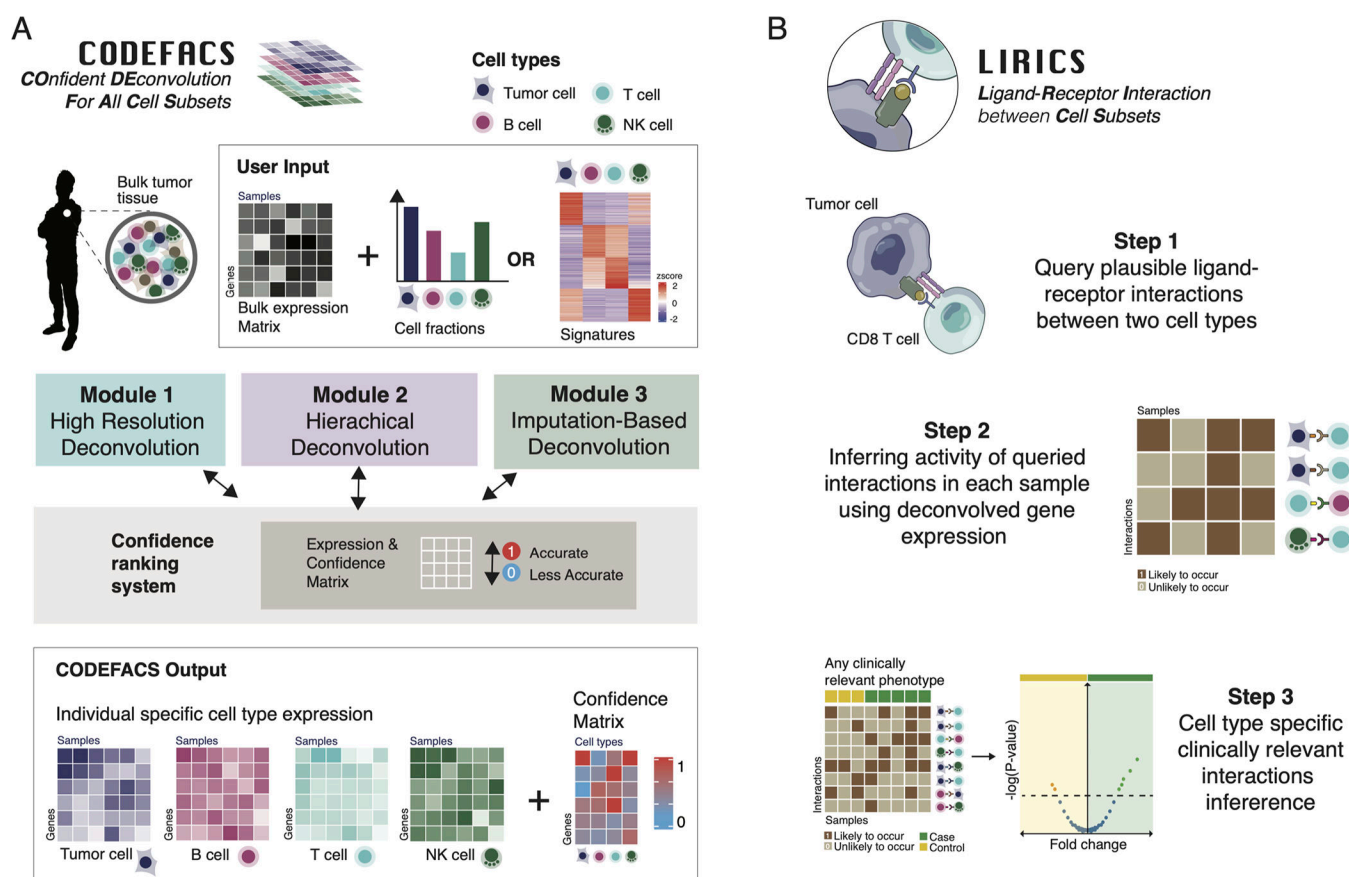


Figure 1. Overview of CODEFACS and LIRICS.

(A) **CODEFACS** takes bulk gene expression profiles and prior knowledge of the cellular composition of each sample and executes a heuristic three-step procedure to infer the deconvolved gene expression in each sample, as described in above and in more details in Supplementary Note 1. (B) **LIRICS** takes the output of CODEFACS and processes it in three steps, as described in the main text and in detail in details, see Supplementary Note 2.

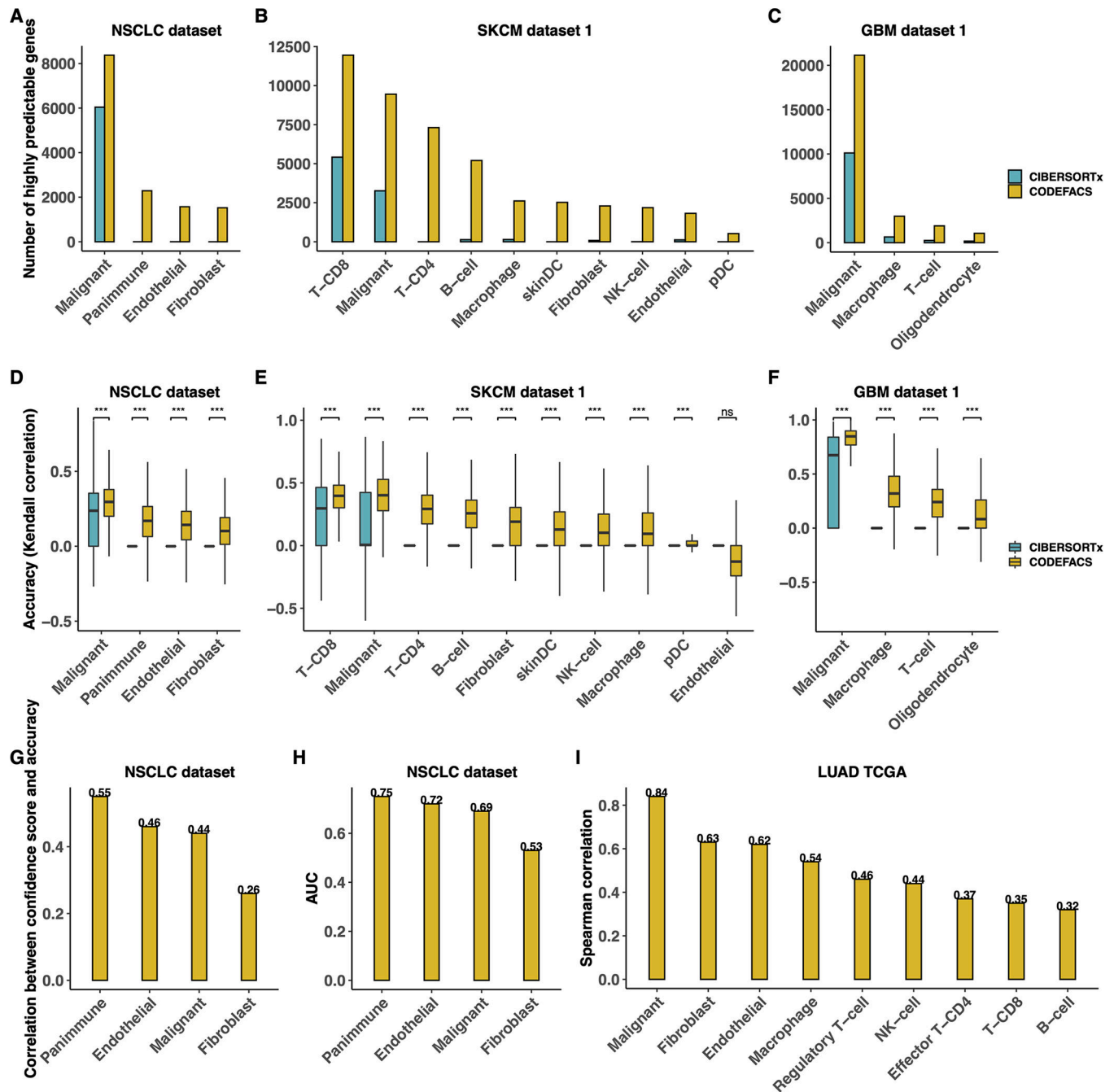
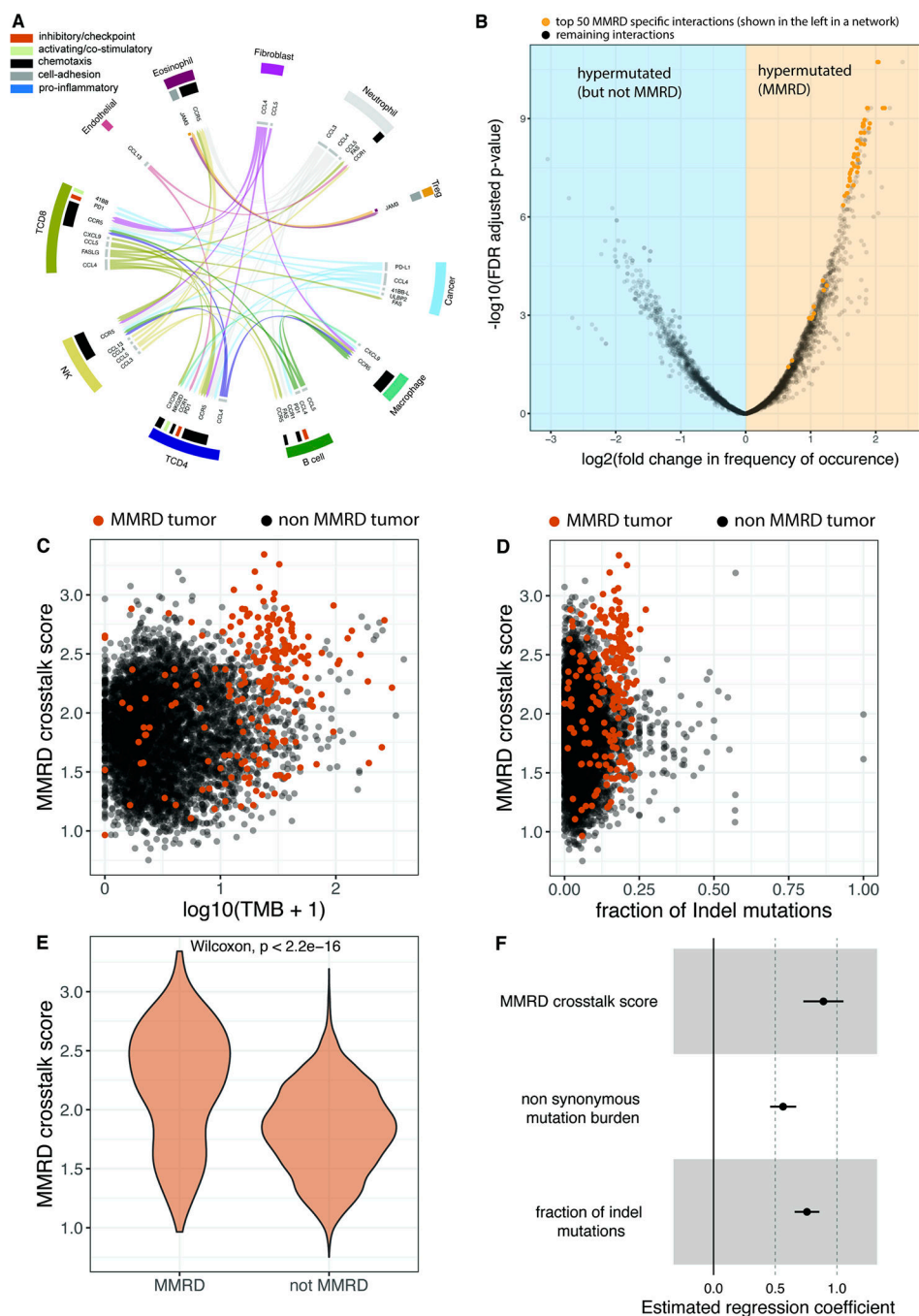


Figure 2. Evaluating the performance of CODEFACS.

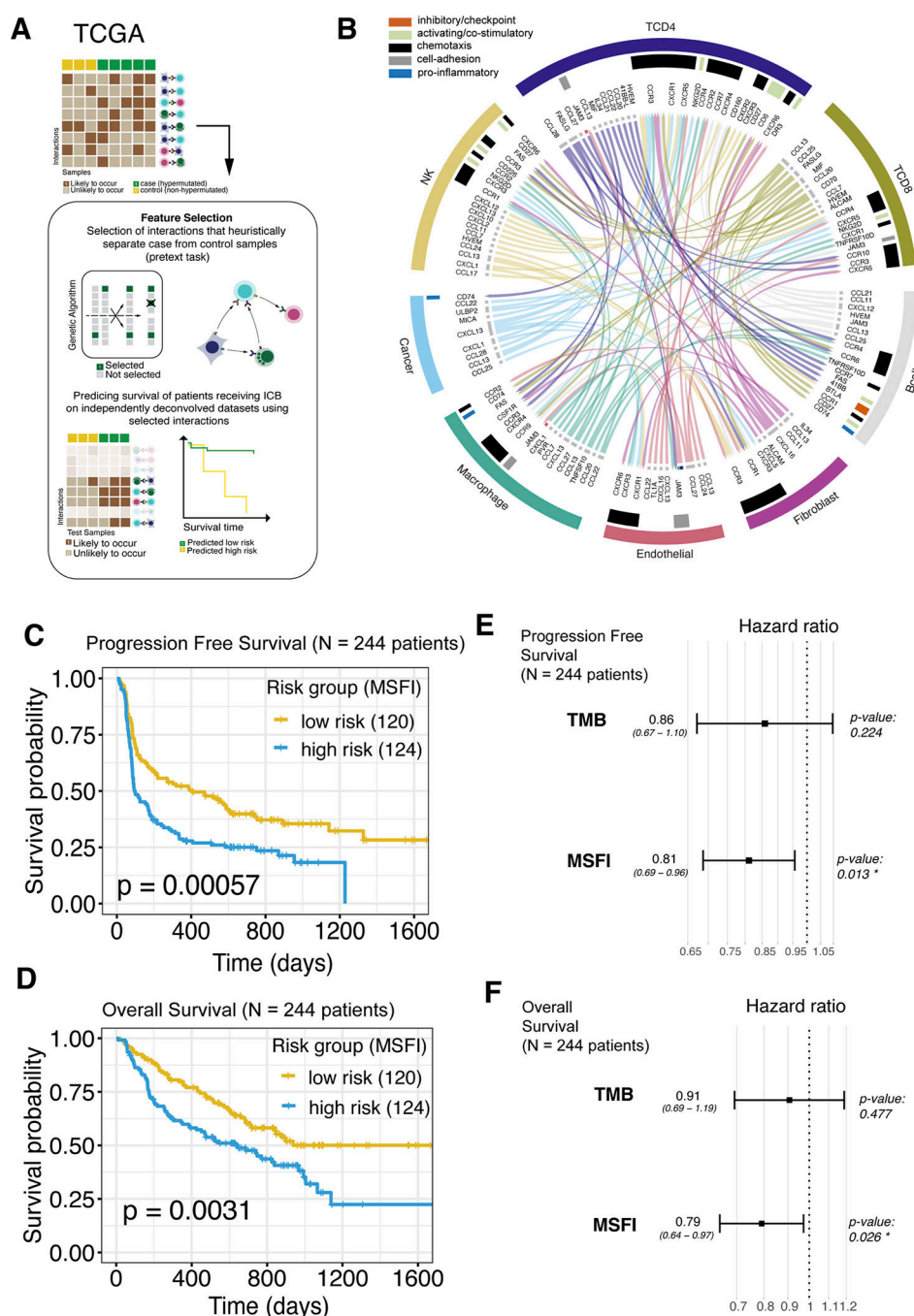
(A-C) bar plots depicting the number of genes with a prediction accuracy (Kendall correlation) ≥ 0.3 with the ground truth for each cell type, estimated from bulk-generated samples of lung cancer (NSCLC dataset; sample size = 26), melanoma (SKCM dataset 1; sample size = 28) and glioblastoma (GBM dataset 1; sample size = 24) benchmark datasets, as estimated by CODEFACS (yellow bars) and CIBERSORTx (blue bars). (D-F) boxplots depicting prediction accuracy distributions of all genes across different cell types in the lung cancer (NSCLC with sample size = 26), melanoma (SKCM with sample size = 28)

and glioblastoma (GBM with sample size = 24) benchmark datasets, using CODEFACS (yellow) and CIBERSORTx (blue). A two-sided Wilcoxon signed rank test was performed to compare the prediction accuracies of CODEFACS and that of CIBERSORTx for each cell type in each dataset. “***” denotes p-values < 0.001. **(G)** Spearman correlations between prediction accuracies and confidence scores among cell types in the lung cancer benchmark dataset (NSCLC dataset; sample size = 26). The y-axis indicates the Spearman correlation coefficient value, while the x-axis indicates the cell type. **(H)** AUCs obtained in classifying informative and uninformative predictions among cell types in lung cancer benchmark dataset (NSCLC dataset; sample size = 26). **(I)** bar plots depicting the Spearman correlations between mean deconvolved cell-type-specific expression in TCGA-LUAD, and mean cell-type-specific expression derived from publicly available single-cell datasets of LUAD. The overlapping gene set between high confidence deconvolved genes in TCGA LUAD and genes in the single cell data for each cell type are used for computing the Spearman correlation. The sizes of the gene sets for the 9 cell types are listed as follows according to the cell type order in the figure: 11765, 8342, 7881, 8790, 9132, 6743, 7448, 10379 and 7267. The y-axis indicates the Spearman correlation coefficient value, while the x-axis indicates the cell type. Details of all single-cell RNA-Seq based pseudo bulk datasets can be found in Supplementary Table S4. In addition, Supplementary Fig. S25 shows a scatter plot of measured mean gene expression in malignant LUAD single cells versus CODEFACS predicted gene expression for malignant cells in bulk TCGA LUAD samples.

**Figure 3.**

(A) Interaction network consisting of the top 50 interactions most highly activated in TME of tumors with DNA mismatch repair deficiency. Interactions highlighted in green represent co-stimulatory interactions/having an activating effect on the target cell. Interactions highlighted in red represent checkpoint interactions/having an inhibitory effect on the target cell. Interactions highlighted in black represent pro-inflammatory/chemotaxis interactions involved in inflammatory response and immune cell trafficking to tumor sites. The ligand-expressing cell type is the sender of the signal and receptor expressing cell type

is the receiver. The functional effect of the interaction typically takes place at the receiver's end as shown in immunological literature. Hence, the functional annotation is indicated with colored sectors on the cell type at the signal receiving end. **(B) A volcano plot depicting on the x-axis the log₂ fold change in the frequency of occurrence of each cell-cell interaction in the TME of hypermutated tumors with an underlying DNA mismatch repair deficiency vs other hypermutated tumors.** The y-axis indicates the -log₁₀ FDR adjusted p-value of the observed enrichment. Highlighted in red in the scatter plot are the top 50 interactions that are most differentially active between all MMRD vs non-MMRD tumors (shown in panel A). **(C) scatter plot depicting the relationship between the MMRD cellular crosstalk scores (Y axis) and the non-synonymous mutation burden (X axis)** across all solid tumor samples in the TCGA. **(D): scatter plot depicting the relationship between the MMRD cellular crosstalk scores (Y axis) and the fraction of indel mutations (X-axis)** across all solid tumor samples in the TCGA. **(E) violin plot depicting the distribution of MMRD cellular crosstalk scores in all MMRD and non MMRD tumors in the TCGA.** **(F) depicting the estimated effect sizes of each feature when fitting a multivariable logistic regression model to MMRD status data in the TCGA.** The X axis represents the estimated logistic regression coefficient for each feature.

**Figure 4.**

(A) Overview of the machine learning analysis employed to identify cell type specific interactions that are predictive of response to immune checkpoint blockade therapy.

(B) A chord diagram of the resulting MSFI network. Each individual interaction is represented by a link from the source cell type (ligand expressing cell type) to the target cell type (receptor expressing cell type) and the color of the link represents the color of the source cell type. The ligand-expressing cell type is the sender of the signal and receptor expressing cell type is the receiver. The functional effect of the interaction typically takes

place at the receiver's end as shown in immunological literature. Hence, the functional annotation is indicated with colored sectors on the cell type at the signal receiving end. For interactions that are activating/co-stimulatory, the sector in the corresponding target cell type is highlighted in green. For inhibitory/checkpoint interactions, the sector in the target cell type is highlighted in red. Interactions involved in chemotaxis are highlighted in black and those mediating a pro-inflammatory response are highlighted in blue, cell-adhesion interactions are highlighted in grey. **(C) Kaplan-Meier plot depicting the progression free survival of the combined set of melanoma patients receiving immune checkpoint blockade (N= 244).** The patients are stratified into low-risk/high-risk groups based on the median value of MSFI score. **(D) Kaplan-Meier plots depicting the overall survival of all melanoma patients receiving immune checkpoint blockade (N= 244).** The patients are stratified into low-risk/high-risk groups based on the median value of MSFI score. **(E-F) Estimated hazard ratios of MSFI scores and TMB on progression free survival and overall survival based on multivariable Cox proportional hazards model.** Panel E depicts estimates of cox proportional hazards ratios for MSFI score and TMB for a multivariable model fitted to progression free survival outcome data. Panel F: depicts estimates of cox proportional hazards ratios for MSFI score and TMB for a multivariable model fitted to overall survival outcome data.

All Patients that received anti-PD1 monotherapy (N = 141 patients)

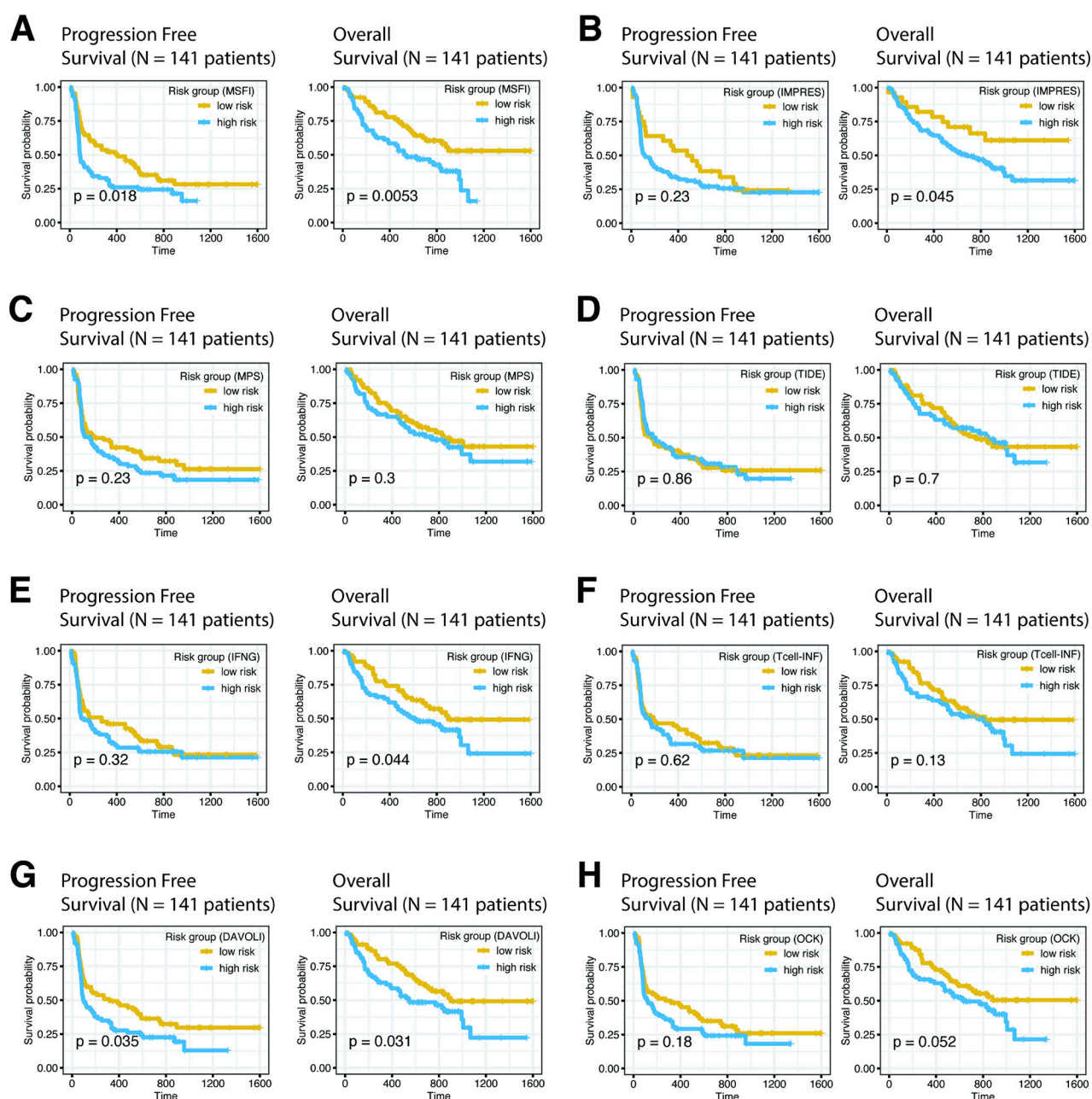


Figure 5. Survival stratification performance of MSFI score vs 7 other previously published bulk transcriptomics-based scores for melanoma patients receiving anti-PD1 only.

(A) Kaplan Meier plots showing progression free survival and overall survival differences between the low-risk groups defined by the median value of the MSFI score. (B-H) Kaplan Meier plots showing progression free and overall survival differences between low and high-risk groups defined by the median value of previously published bulk transcriptomics-based signatures. The significance of survival differences was estimated using the log-rank test. Time on the X-axis is measured in days. The signatures evaluated in each panel are: (B)

IMPRES (8) **(C)** Melanocytic Plasticity Signature (9) **(D)** TIDE (10) **(E)** IFNG signature
(11) **(F)** T cell inflamed GEP (12) **(G)** Cytotoxic Signature (13) **(H)** Immune Signature (14).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

All Patients that received anti-CTLA4 in addition to anti-PD1 (N = 103 patients)

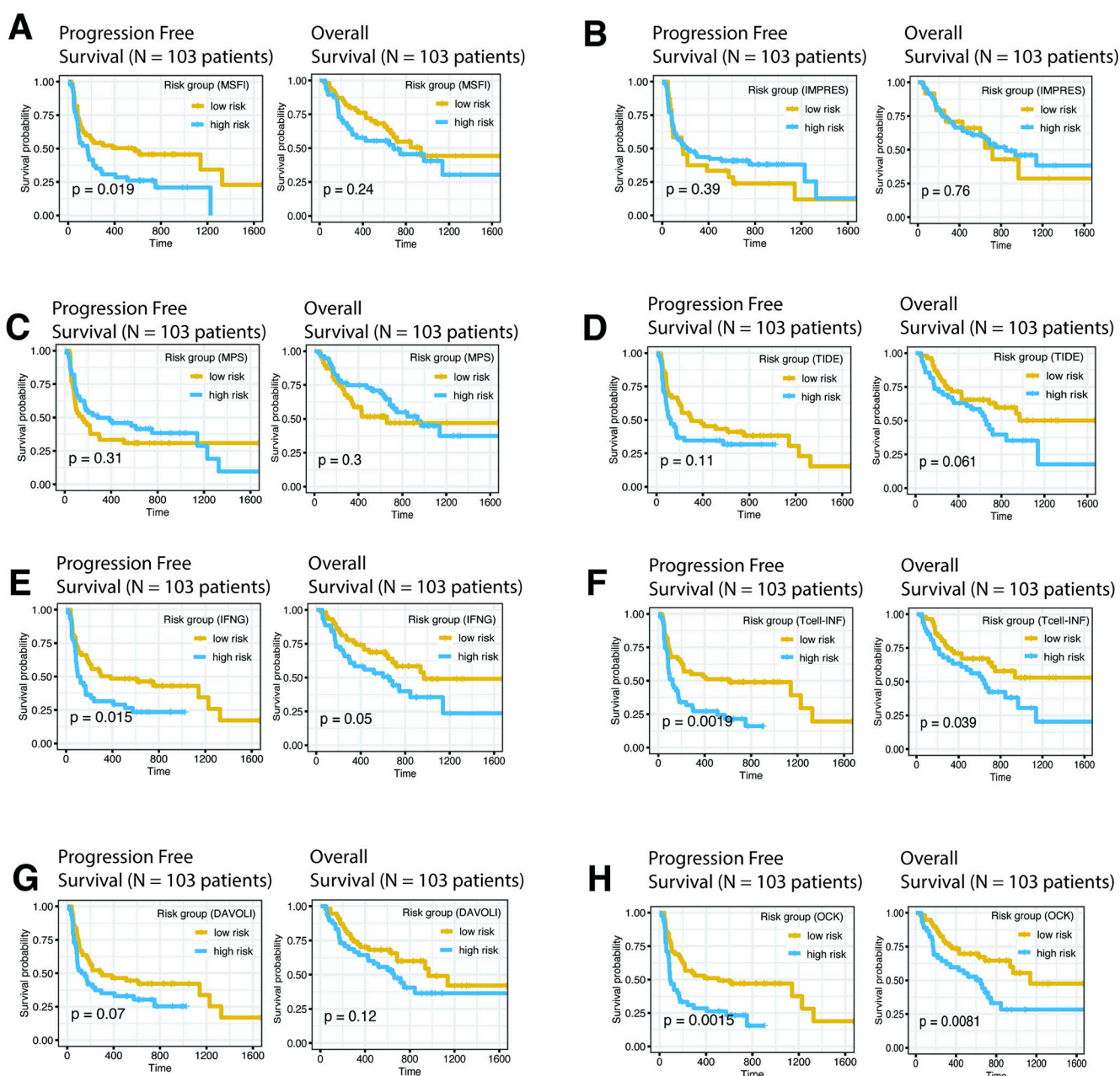


Figure 6. Survival stratification performance of MSFI score vs 7 other previously published bulk transcriptomics-based markers for melanoma patients receiving anti-CTLA4 + anti-PD1.

(A) Kaplan Meier plots showing progression free survival and overall survival differences between the low-risk groups defined by the median value of the MSFI score. (B-H) Kaplan Meier plots showing progression free and overall survival differences between low and high-risk groups defined by the median value of previously published bulk transcriptomics-based signatures. The significance of survival differences was estimated using the log-rank test. Time on the X-axis is measured in days. The signatures evaluated in each panel are: (B)

IMPRES (8) **(C)** Melanocytic Plasticity Signature (9) **(D)** TIDE (10) **(E)** IFNG signature
(11) **(F)** T cell inflamed GEP (12) **(G)** Cytotoxic Signature (13) **(H)** Immune Signature (14).

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript