Comprehensive and Integrated Genomic Characterization of Human Immunome in Cancer

Yongsheng Li¹, Todd Triplett¹, Brandon Burgman^{1,2}, Ming Sun³, Daniel J. McGrail⁴, Dan Qi⁵, Sachet Shukla⁶, Erxi Wu^{1,5,7}, Catherine J. Wu⁶, Anna Capasso^{1,2}, S. Gail Eckhardt^{1,2}, George Georgiou^{2,8,9}, Bo Li^{10*}, Nidhi Sahni^{3,11,12*}, and S. Stephen Yi^{1,2,9,13*}

¹Department of Oncology, The University of Texas at Austin, Dell Medical School, Livestrong Cancer Institutes, Austin, TX 78712, USA

²Institute for Cellular and Molecular Biology (ICMB), College of Natural Sciences, The University of Texas at Austin, Austin, TX 78712, USA

³Department of Bioinformatics and Computational Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁴Department of Systems Biology, The University of Texas MD Anderson Cancer Center, Houston, TX 77030, USA

⁵Neuroscience Institute and Department of Neurosurgery, Baylor Scott & White Health, Temple, TX 76502, USA

⁶Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, MA 02215, USA

⁷Departments of Surgery and Pharmaceutical Sciences, Texas A & M University Health Science Center, Colleges of Medicine and Pharmacy, Temple, TX 76508, USA

⁸Department of Chemical Engineering, The University of Texas at Austin, Austin, TX 78712, USA

⁹Department of Biomedical Engineering, Cockrell School of Engineering, The University of Texas at Austin, Austin, TX 78712, USA

¹⁰Department of Bioinformatics and Immunology, UT Southwestern Medical Center, Dallas, TX 75390, USA

¹¹Department of Epigenetics and Molecular Carcinogenesis, The University of Texas MD Anderson Cancer Center, Smithville, TX 78957, USA

¹²Quantitative and Computational Biosciences, Baylor College of Medicine, Houston, TX 77030, USA

¹³Oden Institute for Computational Engineering and Sciences (ICES), The University of Texas at Austin, Austin, TX 78712, USA

*Correspondence: bo.li@utsouthwestern.edu (B.L.), nsahni@mdanderson.org (N.S.), stephen.yi@austin.utexas.edu (S.Y.)

Abstract

Genetic alterations in immune-related pathways are common hallmarks of cancer. However, to realize the full potential of immunotherapy, a comprehensive understanding of immune networks and how mutations impact network structure and functional output across cancer types is instrumental. Herein we systematically interrogated somatic mutations that could express neoantigens and alter immune responses in cancer patients compared to wild-type controls. To do so, we developed a network-based immunogenomics model (NIPPER) with scoring systems to prioritize critical genes and mutations eliciting differential HLA binding affinity and alternate responses to immunotherapy. These mutations are enriched in essential protein domains and often alter tumor infiltration by immune cells, affecting T cell receptor repertoire and B cell clonal expansion. Furthermore, we devised an interactome network propagation framework integrated with drug associated gene signatures to identify potential immunomodulatory drug candidates. Together, our systems-level analysis results help interpret the heterogeneous immune responses among patients, and serve as a resource for future functional studies and targeted therapeutics.

Significance

Cancer cells induce specific immune-related pathway perturbations by mutations, transcriptional dysregulation, and integration of multi-omics data can help identify critical molecular determinants for effective targeted therapeutics.

Introduction

Immunotherapy has been considered as a promising strategy for treatment of various types of cancer (1,2). Although these immunotherapy methods have made remarkable success, only a minority of patients are observed to respond to the treatment with cytotoxic T lymphocyte antigen-4 (CTLA-4) or programmed death receptor-1 (PD-1) blockade (3,4). There is a clinical need to identify predictive biomarkers and to understand the potential mechanisms of immunotherapy resistance.

With the development of high-throughput sequencing technology, recent evidence has pointed to a number of predictive biomarkers, such as tumor mutation burden (TMB) (5,6), neoantigen burden (7), PD-L1 or PD-L2 mRNA expression (8,9), epigenetic markers and immune cell infiltration profiles (10,11). With these predictive biomarkers in hand, machine learning-based methods are being built to predict the immunotherapy response. However, it is challenging to integrate different biomarkers. Moreover, increasing studies are beginning to reveal the underlying mechanisms of immunotherapy resistance. The expression of immune-related genes (11), insufficient immune cell infiltration (2), oncogenic pathway activity (12) as well as metabolism dysregulation (13) have been related to therapy resistance. But additional insights are still needed for a comprehensive understanding of the mechanisms of immunotherapy resistance.

In addition, genes do not function in isolation but rather interact extensively with each other, and the genetic abnormality is not restricted to the gene product that it encodes (14). The emerging tools of network- or pathway-based medicine offer a valuable platform to explore the predictive biomarkers as well as to understand the dysregulation pattern of cancer-related pathways. Previous studies by The Cancer Genome Atlas (TCGA) have systematically analyzed the alteration landscape of cancer-related signaling pathways (15). However, we still lack the knowledge about the alteration pattern of immune-related pathways across cancer type.

To further refine both the genomic and transcriptomic derived biomarkers for immune response, we first charted the detailed perturbation landscape of immune-related pathways. A network-based immunogenomics model with scoring systems was

proposed to prioritize genes and mutations that could express neoantigens and alter immune responses in cancer patients. We identified mutations eliciting differential HLA binding affinity and these mutations are enriched in essential protein domains and often alter tumor infiltration by immune cells. Furthermore, using a network propagation framework integrated with drug associated gene signatures we discovered potential immunomodulatory drug candidates. These results can help elucidate the functionally relevant mechanisms of immune-related pathway alterations and might inform potential immunotherapy treatment options.

Materials and Methods

Human Immunome

Human immune-related genes were obtained from the ImmPort project (<http://www.immport.org>) (16). In total, 1811 genes in 17 immune-related pathways were obtained. In addition, we also obtained genes that are identified to be essential to immune response from recent literature (17,18). These genes were defined as essential immunotherapy genes and grouped into the 18th pathway. In total, 2,273 immune-related genes were obtained.

TCGA Patient Cohort

The results in our analysis are based upon datasets generated by TCGA Research Network (<http://cancergenome.nih.gov/>). We analyzed 33 different TCGA projects, each project represents a specific cancer type, including KIRC, kidney renal clear cell carcinoma; KIRP, kidney renal papillary cell carcinoma; KICH, kidney chromophobe; LGG, brain lower grade glioma; GBM, glioblastoma multiforme; BRCA, breast cancer; LUSC, lung squamous cell carcinoma; LUAD, lung adenocarcinoma; READ, rectum adenocarcinoma; COAD, colon adenocarcinoma; UCS, uterine carcinosarcoma; UCEC, uterine corpus endometrial carcinoma; OV, ovarian serous cystadenocarcinoma; HNSC, head and neck squamous carcinoma; THCA, thyroid carcinoma; PRAD, prostate adenocarcinoma; STAD, stomach adenocarcinoma; SKCM, skin cutaneous melanoma; BLCA, bladder urothelial carcinoma; LIHC, liver hepatocellular carcinoma; CESC,

cervical squamous cell carcinoma and endocervical adenocarcinoma; ACC, adrenocortical carcinoma; PCPG, pheochromocytoma and paraganglioma; SARC, sarcoma; LAML, acute myeloid leukemia; PAAD, pancreatic adenocarcinoma; ESCA, esophageal carcinoma; TGCT, testicular germ cell tumors; THYM, thymoma; MESO, mesothelioma; UVM, uveal melanoma; DLBC, lymphoid neoplasm diffuse large B-cell lymphoma; CHOL, cholangiocarcinoma.

Molecular and Clinical Datasets across Cancer Types

Somatic mutations were obtained from the publicly available TCGA MAF file (“MC3”) which covers 10,224 patients (19). This dataset was directly downloaded from Synapse under the number of syn7214402. Six calling methods were applied and numbers of filters were applied. All these mutations were subjected to ANNOVAR(20) and we obtained the conservation and MetaSVM score for each mutation (21).

RNA-seq data were obtained from the TCGA project via the R-package “TCGAbiolinks” (22), which was specifically developed for integrative analysis with GDC data. We downloaded the Fragments Per Kilobase of transcript per Million mapped reads (FPKM)-based gene expression for 33 types of cancer. In addition, we also obtained the raw read counts for each gene in these samples.

The clinical information for patients of 33 cancer types were downloaded from TCGA project via the R-package “TCGAbiolinks”, including the survival status, stages, grades, survival time.

Tumor Purity Estimation and Immune Cell Deconvolution

ESTIMATE (Estimation of STromal and Immune cells in MAlignant Tumor tissues using Expression data) was used to estimate the immune score, stromal score and tumor purity of each patients across 33 cancer types (23). The FPKM-based gene expression profiles were used as input of this method. In addition, TIMER (<https://cistrome.shinyapps.io/timer/>) was used to estimate the abundances of member cell types in a mixed cell population, using gene expression data (7).

Pathway Mutation Burden Score

Pathway Mutation Burden (PMB) score was defined to evaluate whether the immune pathways were likely to mutate in a specific cancer type. This score considers the proportion of mutated genes as well as the coverage of patients in each cancer type. For an immune pathway i in cancer type j , the score is defined as following:

$$PMB(P_{ij}) = \frac{m_{ij}}{m_i} * \frac{n_{ij}}{n_j}$$

Where m_{ij} is the number of mutated immune genes in pathway i ; m_i is the total number of genes in pathway i ; n_{ij} is the number of patients of cancer j that with gene mutations in pathway i ; and n_j is the total number of patients sequenced in cancer j .

Immune Response-related Scoring System

Here, we considered three immune response-related scores estimated from gene expression. The immune score that represented the infiltration of immune cells in tumor tissue was assessed based on ESTIMATE (23). The MHC score was formulated from the gene expression levels of the “core” MHC-I set (including HLA-A, HLA-B, HLA-C, TAP1, TAP2, NLRC5, PSMB9, PSMB8 and B2M) (24). The FPKMs of genes were first log-transformed and then median-centered. The mean expression levels of these core genes was defined as the MHC score. The CYT score was a quantitative measure of immune cytolytic activity based on transcript levels of two key cytolytic effectors, granzyme A (GZMA) and perforin (PRF1), found in previous studies (25).

Classification of Patients with Different Immune Responses

To investigate whether the immune-related scores could be used to classify the patients with different immune responses to immunotherapy, we obtained gene expression profiles of patients treated with MAGE-A3 immunotherapy in metastatic melanoma patients (25). The three types of immune response-related scores were calculated for each patient. Next, we used the Receiver Operating Characteristic (ROC) Curve Analysis to evaluate the effects of these scores. This process was performed by using ‘ROCR’ package in R program.

Identification of Protein Regions Associated with Immune Response

The domainXplorer algorithm was used to identify the protein regions that are associated with immune scores (26). The protein functional regions were defined as protein domains in Pfam (27) or intrinsically disordered regions identified by Foldindex (28). In addition, potential domains identified by AIDA (29) were also included in our analysis. Here, we only focused on immune genes-related functional regions.

In brief, domainXplorer used three statistical tests to evaluate whether mutations in a functional protein region were correlated with the immune response-related scores (26). The first linear model was defined as follow:

$$ImS = \beta_0 + \beta_1 * T + \beta_2 * D$$

Where “*ImS*” is the immune score of each patient, “*T*” is the tissue of origin of each patient (here is the cancer type), and “*D*” indicates whether the functional region is mutated or not. In the next step, Wilcox rank sum test was used to compare the immune response-related scores in patients with mutations in the functional region being analyzed against those with mutations in other regions of the same immune gene (Test-2) or those without mutations in the immune gene at all (Test-3). Protein regions with $p_1 < 0.05$, $p_2 < 0.05$ and $p_3 < 0.05$ were identified as to be associated with immune response. In addition, we also identified the protein regions that were correlated with the proportion of immune cell infiltration based on the same method.

Differential Expression Analysis

To identify differentially expressed genes in each cancer type, we downloaded raw RNA-seq counts for all tumor samples. We then used limma (30) to identify the genes and considered the tumor purity as a factor. Genes with p-values less than 0.05 and four-fold changes in expression were identified as differentially expressed genes in each cancer type.

To evaluate the similarity of cancer types based on the differentially expressed immune genes, we calculated the Jaccard index for each pair of cancer

$$Jaccard\ index\ (C_i, C_j) = \frac{DEG_i \cap DEG_j}{DEG_i \cup DEG_j}$$

Where DEG_i and DEG_j are the differentially expressed immune genes in cancer type i and j .

Gene Set Enrichment Analysis

Gene Set Enrichment Analysis was used to identify the enriched immune pathways by differentially expressed genes in each cancer type (31). First, genes were ranked by negative log10 of the differential expression analysis-derived FDR multiplied by the sign of the logFC (log fold change). The “weighted” enrichment statistics were calculated for enrichment or depletion of each immune pathway in specific cancer type.

Prediction of Neoantigens and Functional Screen of Mutations

The somatic mutations were collected from TCGA project and we obtained the HLA alleles for tumor samples from TCIA (<https://tcia.at/home>) (32). Next, NetMHCpan was employed for neoantigen prediction based on the information on mutations and HLA alleles (33). All the wild-type and mutated peptides with 8-11 amino acids were extracted and predicted for the binding affinity. The strong binding is defined as $IC_{50} < 150$ nM and weak binding with $150 \text{ nM} < IC_{50} < 500$ nM based on the previous study (34).

Network-based Prioritization of Immunome-related Genes and Drug Discovery

Human protein-protein network was integrated into our analysis to prioritize immune dysregulated genes. First, we extracted the subnetwork formed by immune-related genes from an integrated functional network (35). The main component of this subnetwork consists 6,060 interactions among 1,202 immune-related genes. Next, the immune genes were ranked based on the correlation between gene expression and immune scores, MHC scores and CYT scores. Top 50 genes with higher positive- or negative-correlation with immune, MHC and CYT scores were selected as seed genes. In addition, the genes which mutations are associated with immune-related scores were also integrated as seed genes. To simulate the propagation of the genetic alterations through the network, we employed the random walk with restart (36). When an immune gene in the network is ranked in top 50, a value “1” is initially assigned to the gene. Then the genetic alterations are propagated along the neighbors such that the genetic

alteration probability was calculated for all genes in the network according to the following equation:

$$P_{t+1} = \alpha AP_t + (1 - \alpha)P_0$$

Where P_0 is the initial binary probability vector, A denotes a degree-normalized adjacency matrix of the immune subnetwork, and α determines the diffusion degree of the genetic influence through the network. We used the optimal value ($\alpha = 0.7$) in our analysis (37). After numbers of iteration, we obtained the final probability for each gene in the network. Genes with greater score than average probability were identified as immune dysregulated genes for each cancer type (38).

Next, we integrated drug associated gene signatures to identify the drug candidates for each cancer type. We first obtained the expression signature genes from The Connectivity Map (also known as CMAP) (39). For each drug, there are two types of signature genes, one is the up-regulated genes and the other one is the down-regulated genes. We used hypergeometric test to evaluate whether the drug-perturbed gene signatures were enriched in cancer-specific immune gene sets. For each cancer type, we obtained the p-value and then all the p-values across 33 cancer types were combined to a combined p-value.

$$p_{combined} = \frac{\sum w * z(p)}{\sqrt{\sum w^2}},$$

Where w is the weight for the individual p-values, and z is the z-score of p-values. Here, we set the same weights for all individual p-values. The drugs with adjusted p-values less than 0.01 were identified. We obtained the drugs that had up-regulated/down-regulated gene signatures enriched for genes positively/negatively correlated with immune scores.

Survival Analysis

The patients were divided into two groups based on the median expression of specific immune gene. Log-rank test was used to evaluate the survival difference between two groups. Genes with p-values less than 0.05 were considered as significant.

Results

Genome and Transcriptome-based Immune Response Prediction

Tumor mutational burden (TMB), a measurement of the overall number of genetic alterations observed in a cancer sample, has been suggested as a potentially prognostic marker for immunotherapy (40,41). In addition, several gene signatures have also been proposed to predict the immune response of patients (Figure 1A). We first evaluated whether the three immune-related scores (immune score, MHC score and cytolytic activity) could be used to predict the immune response of tumor patients. Based on the gene expression data in response to treatment with MAGE-A3 immunotherapy of metastatic melanoma patients (42), we found that the patients that responded to immunotherapy had significantly higher immune scores ($p=0.003$), MHC scores ($p=1.7e-4$) and cytolytic activity ($p=0.007$, Figure 1B). Moreover, we found that these three immune-related scores indeed distinguished the responder from non-responder patients with area under the curve of the receiver operating characteristic (AUROC) from 0.71 to 0.79 (Figure 1C). To further validate this, we analyzed another gene expression dataset for PD-1 immune checkpoint blockade (43). We also found that patients which were responsive to the PD-1 immunotherapy exhibited higher immune-related scores (Figure S1A). The AUROCs for classification ranged from 0.52 to 0.85 (Figure S1B), which was similar to the results obtained from gene signatures in previous studies (44). These results suggest that the immune-related scores can potentially reflect the immune milieu of tumors in patients and help identify specific factors that may modulate these environments.

We next explored the generalizability of these scoring systems in predicting patient immune responses across cancer types. We first calculated the three immune-related scores (immune score, MHC score and cytolytic activity) for all tumor patients from the TCGA project (Figure 1D). We found that the distributions of the immune-related scores were diverse across 33 cancer types. Specifically, patients in several cancer types, such as lymphoid neoplasm diffuse large B-cell lymphoma (DLBC) and kidney renal clear cell carcinoma (KIRC), exhibited higher immune-related scores. These results

suggest that immunotherapy might be suitable for these cancer types. Moreover, although the three immune-related scores appeared to be correlated with each other, the Pearson Correlation Coefficients (PCC) ranged from 0.51-0.79 (Figure 1E). This observation implied that these scores might be complementary with each other, therefore integrating these scores would facilitate the identification of novel candidate gene targets for immunotherapy.

Immune Mutational Burden Analysis Identifies Critical Pathways and Genes

It has been suggested that the mutations in immune pathways could impact tumor-immune interactions (45,46). It is unclear, however, how the immune mutational burden (IMB) is associated with patient outcome. Moreover, it seems that the mutation burdens across pathways are varied (15). Therefore it remains elusive if and how mutations in immune-related pathways contribute to higher mutation burden. To systematically investigate the global functional genomic landscape of immune-related genes, we manually curated 2,273 immune-related genes from literature (16,17). These genes were further classified into 18 immune-related pathways (Figure 2A). Next, we analyzed whole exomes from 100 melanoma patients treated with CTLA-4 blockade (47). We first calculated the number of mutations (tumor mutation burden, TMB) and the number of mutations located in immune-related genes (defined as immune mutation burden, IMB). We further compared TBM and IMB between responders and non-responders. The mutation burden and immune burden of responders were significantly higher than those of non-responders (Figure S2A, $p < 0.05$). In addition, mutation burden and immune burden could predict immune response with similar power (Figure S2B, AUCs=0.68). Furthermore, we calculated the IMB for each patient in TCGA cohort. We found that the cancer types that are likely to respond well to immunotherapy had higher proportion of immune-related gene mutations, such as skin cutaneous melanoma (SKCM) and kidney cancer (Figure S2C). These results suggest that the immune burden is a useful indicator of immune response; target sequencing of immune-related genes would help predict the immune response of patients.

Next, we defined a Pathway Mutation Burden (PMB) score to evaluate to what extent each immune-related pathway is perturbed by genomic alterations. This score is

calculated based on the mutation frequency and the proportion of mutated genes in each pathway. We found that the essential genes identified by CRISPR-Cas9 KO screens (17) had higher PMB scores across cancer types (Figure 2B). Moreover, we observed another pathway-cytokine receptors (CRs) with higher PMB scores in LUSC, SKCM and DLBC (Figure 2B). We next focused on the CR mutations in LUSC (Figure 2C). Consistent with previous studies, the mutations in CR pathway exhibited a mutual exclusive pattern (48,49). Moreover, the mutations in CR pathway had higher functional impact scores than other mutations evaluated by metaSVM (21) (Figure S2D). The top-1 gene (PLXNA4) ranked by mutation frequency was mutated in approximately 19% of lung cancer patients. PLXNA4 was shown previously to promote tumor progression and tumor angiogenesis by activating VEGF and FGF signaling (50). We further found that the expression of PLXNA4 was correlated with the patients' survival in LUSC (Figure S2E, $p=0.017$). These results suggest that the mutations in cytokine receptors pathway might play critical roles in cancer and the proposed PMB index could help prioritizing key mutations in cancer pathways.

Widespread Transcriptional Alterations of Human Immunome across Cancer Types

Next, we interrogated if and to what extent the human immunome is altered in cancer at the gene expression level. We identified differentially expressed genes across 18 cancer types with more than five normal samples. As tumor purity may affect the identification of differentially expressed genes (51), we first estimated the tumor purity of each sample (Figure S3A) and identified the differentially expressed genes using tumor purity as a co-variable in the regression model. We found that the immune-related genes were significantly perturbed in cancer when compared to other coding genes (Figure 3A and Figure S3B, p -values <0.05 for all cancer types, Fisher's exact test). The Odd Ratios (ORs) of immune genes vs non-immune genes were higher than 1.0 in all cancer types. These results indicate that these genes could play a role in promoting tumor growth.

An increasing number of studies have demonstrated that cancer types with similar tissue origin show similar transcriptome and regulation (52,53). Next, we calculated the

Jaccard index for each pair of cancer based on the similarity of differentially expressed immune-related genes (see details in Methods). We found that cancer lineages with similar tissue origin were clustered together (Figure 3B), such as lung adenocarcinoma (LUAD) and lung squamous cell carcinoma (LUSC), cholangiocarcinoma (CHOL) and liver hepatocellular carcinoma (LIHC), and three types of kidney cancer. These results indicate that these cancer types show similar immunome perturbations.

In addition, we ranked all genes based on their differential expression and used gene set enrichment analysis (GSEA) to explore which immune-related pathways were likely to show transcriptional perturbation (31,54). 16/18 immune-related pathways were significantly altered in at least one cancer type (Figure 3C). The antigen processing and presentation, cytokines and cytokines receptors pathways were perturbed in more than 16 cancer types. Moreover, we found that these immune-related pathways showed consistent trend (enriched or depleted) across cancer types. For instance, antigen processing and presentation pathway was likely to be activated in breast invasive carcinoma and glioblastoma (BRCA and GBM, Figure 3C and Figure S3C). In addition, cytokines and cytokine receptors pathways were repressed in BRCA and GBM (Figure 3E-F and Figure S3D, FDR<0.001).

Critical Regions of Immune Genes Associated with Immune Response

Growing evidence suggests that mutations in certain genes influence the immune response (26). Identification of these genes may provide valuable insights for improving immunotherapy. Thus, we explored 'domainXplorer' to identify critical regions of immune genes that are associated with immune response in cancer (26). Given the importance of the proposed immune-related scores in predicting immune response, we next determined whether cancer mutations in immunity candidate driver regions were significantly correlated with these scores. Our analysis yielded a total of 209 protein regions in 145 immune-related genes that are potential cancer immunity drivers (Figure 4A). Specifically, mutations in seven genes (UBXN1, LTBP2, STAT1, NGFR, NRAS, MET and LTBP4) were significantly correlated with all of three types of immune-related scores. LTBP4, a member of the latent TGF β binding protein gene family, had been linked to several human diseases (55). We found that patients with EGF_CA domain

mutations exhibited significantly higher immune scores, MHC scores and cytolytic activity (Figure 4B, $p < 0.05$). Moreover, we also identified mutations of STAT1 significantly correlated with these scores, which was consistent with previous observation that STAT1 played an important role in the innate immune response (56). These results collectively expand the catalog of potential cancer immune drivers and highlight the importance of taking into account the protein structural context to identify patients that are likely to clinically benefit from immunotherapy.

Somatic Mutations Associated with Immune Cell Infiltration

Our above analyses identified candidate immune driver protein regions that are potentially correlated with immune response. However, we are still lack of knowledge about the potential mechanism. An emerging role in treatment response has been attributed to immune cell infiltration in human tumors (57). Thus, we next explored to what extent the mutations in these regions were related with immune cell infiltration. We first identified the protein regions whose mutations were significantly correlated with the abundance of six tumor-infiltrating immune cells (TIIC) subsets (B cells, CD4 T cells, CD8 T cells, macrophages, neutrophils, and dendritic cells) (58). We also identified the proteins that were associated with B cell receptor (BCR) diversity and CPK (TCR/BCR CDR3s per kb of TCR/BCR reads) used to evaluate the clonotype diversity of T and B cells (59). We found that both MHC and CYT scores showed significant association with the expansion of tumor-infiltrating B cells after correction for tumor purity (Figure S4A). In addition, MHC score was weakly associated with TCR diversity, potentially due to clonal expansion of tumor antigen-specific T cell populations. CYT score on the other hand was strongly correlated with CPK (Figure S4B). We found that the protein regions that were correlated with immune score, MHC score and cytolytic activity were significantly overlapped with those correlated with TIIC abundance (Figure 4C). Especially, there were a higher number of protein regions correlated with CD8 T cells, neutrophils, and dendritic cells.

Specifically, we found that seven genes, whose mutations correlated with all three types of immune-related scores, were associated with at least one type of immune cell abundance (Figure 4D). For example, patients with LTBP4 mutations exhibited

significantly higher abundance of CD8 T cells and neutrophils (Figure 4E). CD8+ T cells are crucial mediators of anti-tumor immune response and the targets of checkpoint blockade (60), while neutrophils are part of the innate immune system whose roles in cancer have been mixed (61). Moreover, we found that not only the genomic variants in LTBP4 were correlated with immune cell abundance, but patients with copy number alterations also exhibited significantly different abundance (Figure S4C). These results suggest that the genomic alterations might contribute to immunotherapy response through affecting the immune cell infiltration in cancer.

Critical Immune Genes Enriched for Neoantigens in Cancer

Increasing studies have demonstrated that mutation-derived neoantigens form a major ‘active ingredient’ of successful cancer immunotherapies (62,63). Next, we predicted the neoantigens that were derived from mutations of TCGA project based on NetMHC (64). We found that the immune-related genes tended to generate significantly more neoantigens compared with coding genes in general (Figure 5A, $p=1.18e-4$, Fisher’s exact test). Especially we found that the mutations in PIK3CA, KRAS, HRAS and NRAS were likely to generate neoantigens. Approximate 25% of the mutations in PIK3CA and HRAS could generate neoantigens (Figure 5B). This result is consistent with the observation that RAS plays a role in immunotherapy (65,66).

Next, we specifically focused on mutations in PIK3CA and EGFR that give rise to neo-antigens. Based on the viability-based functional screen in two cell lines (Ba/F3 and MCF10A) (67), we found that the mutations that encode neoantigens were likely to be activating mutations in both cell lines (Figure 5C-5D). For example, we identified 46 missense mutations that could generate neoantigens in PIK3CA. Among these mutations, 41 mutations were in the functional dataset. We found that 40 mutations were activating mutations and one was inactivating mutations. Moreover, we identified a mutation cluster (including R38C/H, E81K, R88Q, R93W, K111E and R115L) in PIK3CA (Figure 5E). For EGFR, we identified 17 mutations that could generate neoantigens. In total, 14 mutations were screen for function and 11 mutations were activating, two were neutral and 1 inconclusive in two cell lines (Figure 5D). We also found seven mutations formed a cluster in 3D (including R108K, R252/P/C/H and A189V/D, Figure 5E). These

results suggest that these mutations may play a role in the growth of tumors and are less likely to be selected during immune editing and undergo antigen escape.

Prioritization of Candidate Driver Genes Based on Immunome Network

Our above results indicate widespread genomic and transcriptome perturbations in the cancer immunome network. An integrated landscape of these perturbations offers a basis for further advancing our understanding of the immune regulation in cancer. Thus, we proposed a Network-based Integrative model to Prioritize the Potential immune drivERs (NIPPER) (Figure 6A). Briefly, we first identified the genes whose mutations or expression were significantly correlated with immune scores, MHC scores and CYT scores. These genes were then mapped to an immune-related protein-protein interaction (PPI) network, which consisted of 6,060 interactions among 1,202 immune-related genes. Next, the mutational effects were propagated in the PPI network. Finally, the genes with driver probability greater than expected were identified as immune response-related signature genes. This process was repeated for all cancer types. Here, the genes positively/negatively correlated with immune-related scores were analyzed, separately.

Based on the proposed network-based model, we identified 39 genes as signature genes in 33 cancer types (Figure 6B). Patient with high expression of these genes were likely to response to immunotherapy. These genes were involved in 11 immune-related pathways, and 8 genes (such as STAT1, B2M, TAP1 and TAP2) had been identified as essential genes in immune response. Moreover, we found that these candidate immune response-related genes were clustered together as modules in the PPI network (Figure 6C). Seven of the 8 essential immune genes were linked in a dense module, and STAT1 was clustered together with more novel immune response-related genes; many of these were enzymes that play critical roles in modulating signaling by the T cell receptor (e.g. LCK, FYN, PTPN11). Interestingly, we identified one candidate gene-PDIA3, which linked 7 known essential genes. The expression of this gene was correlated with the patients' survival in several cancer types, including CESC, LUAD, HNSC and GBM (Figure 6D and Figure S5). These observations suggest that this gene

may play critical roles in immunotherapy, which is consistent with the conclusion of one recent study (68).

Next, we validated these genes in another cohort that underwent treatment with immunotherapy (24). We found that the expression of these genes could classify the responder and non-responder patients with an AUC of 0.713 (Figure 6E). Moreover, the average expression of these genes were significantly associated with improved overall survival (Figure 6F, log-rank $p=0.02$).

Mutation-Perturbed HLA Binding and Potential Immune-related Drugs

Next, we evaluated mutations in the genes identified by NIPPER in different cancers. We found that the mutations in these genes were likely to decrease the binding affinity of HLAs (Figure 7A, $p<0.001$, Wilcox's rank sum test). Specifically, we identified 165 mutations in 30 genes that induced the weaker binding of HLAs (Figure 7B). These mutations were distributed in various types of cancer. We identified six mutations in PIK3R1 (encoding p85 α) that caused the weak binding of HLAs (Figure 7B and 7D). It had been demonstrated that both loss-of-function and gain-of-function mutations in this gene lead to immunosuppression (70). Consistently, we identified two mutations (N564D and K567E) that increased growth of cancer cell lines (Figure 7D). These results suggest that our network-based method helps identify candidate driver mutations by inducing the decreased binding of HLAs.

Identification of candidate small molecule drugs is critical for the immunotherapy in cancer. Herein we identified candidate drugs that could increase the expression of identified immune gene signatures based on GSEA analysis (Figure 6A). We obtained more than 6,000 small molecules and their perturbation gene signatures from The Connectivity Map (71). We identified 49 small molecules that significantly perturbed the expression of immune gene signatures ($p<0.05$, Figure 7E). Interestingly, we identified propofol as the top one small molecule for increased immune response. Evidence had shown that propofol could possibly induce a favorable immune response in terms of preservation of IL-2/IL-4 and CD4+/CD8+ T cell ratio in the perioperative period for breast cancer (72). In vitro and in vivo studies had also suggested that propofol could independently reduce migration of cancer cells and metastasis (73). Taken together,

these results suggest that the network-based integrative model not only can identify immune response-related gene signatures, it also helps prioritize candidate drugs for immunotherapy in cancer.

Discussion

Systematically understanding the regulation of immune systems opens more avenues for cancer immunotherapy with a potent clinical efficacy. By integration of pan-cancer omics datasets, we have examined human tumor correlations in the context of immune-related pathways to an extent that is not previously possible. High-throughput gene expression data have been widely used to investigate differentially expressed genes in various types of cancer (74,75). Strikingly, we observed extensive immune-gene signatures with widespread perturbation in cancer compared with normal samples. These deregulated immune-related genes are enriched in ‘antigen processing and presentation’, ‘cytokines’ and ‘cytokines receptors’ pathways. Cytokines are critical molecular messengers for cells of the immune system to communicate with one another, which can directly stimulate immune effector cells and enhance tumor cell recognition. Increasing studies have demonstrated that cytokines have broad anti-tumor activity (76,77). Our analysis revealed that there are widespread perturbations of cytokine pathways in more than 80% of cancer types. A more detailed understanding of cytokine pathway regulation will provide new opportunities for improving cancer immunotherapy.

It is still unclear which factors are key players in regulating the expression of immune-related genes. Thus, we systematically analyzed two types of genetic regulation (including CNVs and eQTLs) in immune genes (Figure S6). We found that a number of immune-related genes recurrently dysregulated in various types of cancer. For example, up-regulation of BIRC5 is correlated with CNV amplification (78); and down-regulation of CCL14 is correlated with CNV deletion across cancer types. Moreover, eQTLs analysis revealed a hotspot locus that modulates the expression of DEFB1 (Figure S7), which might function through perturbation of CTCF binding. While a number of immune-related genes show correlation between copy number variations/mutations and gene expression, the relationship is weak and could not fully explain immune transcriptome perturbations. Thus, it is likely that other genetic (such as

post-transcriptional regulation) or epigenetic variables (such as DNA methylation or histone modification) also contribute to the transcriptome perturbations in cancer.

Although several studies have proposed that the immune score, MHC score and CYT score are correlated with immune response, we found that these scores are complementary with each other. Thus, we proposed an integrated network model to predict the critical genes for immune response. This method identified several well-known immune response-related genes, such as B2M (79), TAP1 (80) and TAPBP (81), as well as several novel genes. HLA binding analysis also showed that these mutations are likely to induce the decreased binding of HLAs. Especially, we identified 165 candidate driver mutations that potentially play central roles in immune response. The mutations in PIK3R1 were validated in cell lines and were found to induce cancer cell growth. Moreover, we predicted candidate small molecules for immunotherapy based on whether the treatment could induce the expression of these signature immune genes. In particular, propofol was identified as the top-ranked drug candidates for immunotherapy. Propofol is the most commonly used in clinical anesthesia and propofol exhibits a good inhibitory effect on tumor recurrence and metastasis (82). Our current study found that propofol likely induces the expression of immune response-related genes, which may contribute to improved efficacy of immunotherapy.

Taken together, our integrative analysis of the human immunome across cancer types reveals several candidate gene and mutation markers that may be possible to use for predicting effective immune response prior to immunotherapy. Future studies will also need to evaluate the potential molecular functions of our predictive gene and mutation markers by low-throughput experiments.

Disclosure of Potential Conflicts of Interest

No potential conflicts of interest were disclosed by the authors.

Authors' Contributions

Conception and design: Y.L. and S.Y.

Development of methodology: Y.L., N.S. and S.Y.

Acquisition of data: Y.L., S.S., C.J.W., B.L. and S.Y.

Analysis and interpretation of data (e.g., statistical analysis, biostatistics, computational analysis): Y.L., B.L., N.S. and S.Y.

Writing, review, and/or revision of the manuscript: Y.L., T.T., D.J.M., E.W., A.C., S.G.E., G.G., B.L., N.S. and S.Y.

Administrative, technical, or material support (i.e., reporting or organizing data, constructing databases): T.T., B.B., M.S., D.Q., S.S., A.C.

Study supervision: S.Y.

Acknowledgments

The authors would like to acknowledge Kevin Drew and Edward M. Marcotte from University of Texas at Austin for helpful discussions. The authors are grateful to contributions from TCGA Research Network Analysis Working Group, and also acknowledge the Biomedical Research Computing Facility at UT Austin and Texas Advanced Computing Center (TACC) for high-performance computing assistance.

Grant Support

This work was supported by National Institutes of Health grant [K22CA214765 to S.Y.], Komen Foundation grant [CCR19609287 to S.Y.], Young Investigator Grant by Breast Cancer Alliance [to N.S.], Liz Tilberis Early Career Award by Ovarian Cancer Research Alliance [Grant# 649968 to N.S.], Alfred P. Sloan Scholar Research Fellowship [FG-2018-10723 to N.S.], Cancer Prevention and Research Institute of Texas grants RR170079 [to B.L.] and RR160093 [to S.G.E.]. N.S. is a CPRIT Scholar in Cancer Research with funding from the Cancer Prevention and Research Institute of Texas (CPRIT) New Investigator Grant RR160021. D.J.M. is supported by National Cancer Institute grant K99CA240689. A.C. is supported by Department of Defense grant CA191245.

References

1. Liu XS, Mardis ER. Applications of Immunogenomics to Cancer. *Cell* **2017**;168:600-12
2. Tumei PC, Harview CL, Yearley JH, Shintaku IP, Taylor EJ, Robert L, *et al.* PD-1 blockade induces responses by inhibiting adaptive immune resistance. *Nature* **2014**;515:568-71
3. Schadendorf D, Hodi FS, Robert C, Weber JS, Margolin K, Hamid O, *et al.* Pooled Analysis of Long-Term Survival Data From Phase II and Phase III Trials of Ipilimumab in Unresectable or Metastatic Melanoma. *J Clin Oncol* **2015**;33:1889-94
4. Valsecchi ME. Combined Nivolumab and Ipilimumab or Monotherapy in Untreated Melanoma. *N Engl J Med* **2015**;373:1270
5. Rizvi NA, Hellmann MD, Snyder A, Kvistborg P, Makarov V, Havel JJ, *et al.* Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **2015**;348:124-8
6. Snyder A, Makarov V, Merghoub T, Yuan J, Zaretsky JM, Desrichard A, *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **2014**;371:2189-99
7. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **2016**;351:1463-9
8. Topalian SL, Hodi FS, Brahmer JR, Gettinger SN, Smith DC, McDermott DF, *et al.* Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N Engl J Med* **2012**;366:2443-54
9. Yearley JH, Gibson C, Yu N, Moon C, Murphy E, Juco J, *et al.* PD-L2 Expression in Human Tumors: Relevance to Anti-PD-1 Therapy in Cancer. *Clin Cancer Res* **2017**;23:3158-67
10. Peng D, Kryczek I, Nagarsheth N, Zhao L, Wei S, Wang W, *et al.* Epigenetic silencing of TH1-type chemokines shapes tumour immunity and immunotherapy. *Nature* **2015**;527:249-53
11. Herbst RS, Soria JC, Kowanetz M, Fine GD, Hamid O, Gordon MS, *et al.* Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **2014**;515:563-7
12. Spranger S, Bao R, Gajewski TF. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. *Nature* **2015**;523:231-5
13. Yang W, Bai Y, Xiong Y, Zhang J, Chen S, Zheng X, *et al.* Potentiating the antitumour response of CD8(+) T cells by modulating cholesterol metabolism. *Nature* **2016**;531:651-5
14. Barabasi AL, Gulbahce N, Loscalzo J. Network medicine: a network-based approach to human disease. *Nat Rev Genet* **2011**;12:56-68
15. Sanchez-Vega F, Mina M, Armenia J, Chatila WK, Luna A, La KC, *et al.* Oncogenic Signaling Pathways in The Cancer Genome Atlas. *Cell* **2018**;173:321-37 e10
16. Bhattacharya S, Andorf S, Gomes L, Dunn P, Schaefer H, Pontius J, *et al.* ImmPort: disseminating data to the public for the future of immunology. *Immunologic research* **2014**;58:234-9
17. Patel SJ, Sanjana NE, Kishton RJ, Eidizadeh A, Vodnala SK, Cam M, *et al.* Identification of essential genes for cancer immunotherapy. *Nature* **2017**;548:537-42
18. Miao D, Margolis CA, Gao W, Voss MH, Li W, Martini DJ, *et al.* Genomic correlates of response to immune checkpoint therapies in clear cell renal cell carcinoma. *Science* **2018**;359:801-6
19. Zhang Y, Kwok-Shing Ng P, Kucherlapati M, Chen F, Liu Y, Tsang YH, *et al.* A Pan-Cancer Proteogenomic Atlas of PI3K/AKT/mTOR Pathway Alterations. *Cancer cell* **2017**;31:820-32 e3
20. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research* **2010**;38:e164
21. Dong C, Wei P, Jian X, Gibbs R, Boerwinkle E, Wang K, *et al.* Comparison and integration of deleteriousness prediction methods for nonsynonymous SNVs in whole exome sequencing studies. *Hum Mol Genet* **2015**;24:2125-37

22. Colaprico A, Silva TC, Olsen C, Garofano L, Cava C, Garolini D, *et al.* TCGAbiolinks: an R/Bioconductor package for integrative analysis of TCGA data. *Nucleic acids research* **2016**;44:e71
23. 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
24. Lauss M, Donia M, Harbst K, Andersen R, Mitra S, Rosengren F, *et al.* Mutational and putative neoantigen load predict clinical benefit of adoptive T cell therapy in melanoma. *Nature communications* **2017**;8:1738
25. Rooney MS, Shukla SA, Wu CJ, Getz G, Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **2015**;160:48-61
26. Porta-Pardo E, Godzik A. Mutation Drivers of Immunological Responses to Cancer. *Cancer Immunol Res* **2016**;4:789-98
27. Finn RD, Bateman A, Clements J, Coggill P, Eberhardt RY, Eddy SR, *et al.* Pfam: the protein families database. *Nucleic acids research* **2014**;42:D222-30
28. Prilusky J, Felder CE, Zeev-Ben-Mordehai T, Rydberg EH, Man O, Beckmann JS, *et al.* FoldIndex: a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics* **2005**;21:3435-8
29. Xu D, Jaroszewski L, Li Z, Godzik A. AIDA: ab initio domain assembly server. *Nucleic acids research* **2014**;42:W308-13
30. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **2015**;43:e47
31. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* **2005**;102:15545-50
32. Charoentong P, Finotello F, Angelova M, Mayer C, Efremova M, Rieder D, *et al.* Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep* **2017**;18:248-62
33. Jurtz V, Paul S, Andreatta M, Marcatili P, Peters B, Nielsen M. NetMHCpan-4.0: Improved Peptide-MHC Class I Interaction Predictions Integrating Eluted Ligand and Peptide Binding Affinity Data. *J Immunol* **2017**;199:3360-8
34. Wu J, Zhao W, Zhou B, Su Z, Gu X, Zhou Z, *et al.* TSNAdb: A Database for Tumor-specific Neoantigens from Immunogenomics Data Analysis. *Genomics Proteomics Bioinformatics* **2018**;16:276-82
35. Menche J, Sharma A, Kitsak M, Ghiassian SD, Vidal M, Loscalzo J, *et al.* Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* **2015**;347:1257601
36. Kohler S, Bauer S, Horn D, Robinson PN. Walking the interactome for prioritization of candidate disease genes. *American journal of human genetics* **2008**;82:949-58
37. Hofree M, Shen JP, Carter H, Gross A, Ideker T. Network-based stratification of tumor mutations. *Nature methods* **2013**;10:1108-15
38. Shin D, Lee J, Gong JR, Cho KH. Percolation transition of cooperative mutational effects in colorectal tumorigenesis. *Nature communications* **2017**;8:1270
39. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, *et al.* The Connectivity Map: using gene-expression signatures to connect small molecules, genes, and disease. *Science* **2006**;313:1929-35
40. Yarchoan M, Hopkins A, Jaffee EM. Tumor Mutational Burden and Response Rate to PD-1 Inhibition. *The New England journal of medicine* **2017**;377:2500-1

41. Goodman AM, Kato S, Bazhenova L, Patel SP, Frampton GM, Miller V, *et al.* Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Molecular cancer therapeutics* **2017**;16:2598-608
42. Ulloa-Montoya F, Louahed J, Dizier B, Gruselle O, Spiessens B, Lehmann FF, *et al.* Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **2013**;31:2388-95
43. Hugo W, Zaretsky JM, Sun L, Song C, Moreno BH, Hu-Lieskovan S, *et al.* Genomic and Transcriptomic Features of Response to Anti-PD-1 Therapy in Metastatic Melanoma. *Cell* **2016**;165:35-44
44. 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. *Nature communications* **2017**;8:1050
45. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Ou Yang TH, *et al.* The Immune Landscape of Cancer. *Immunity* **2018**;48:812-30 e14
46. Weiss A. Human LAT mutation results in immune deficiency and autoimmunity but also raises questions about signaling pathways. *The Journal of experimental medicine* **2016**;213:1114
47. Van Allen EM, Miao D, Schilling B, Shukla SA, Blank C, Zimmer L, *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **2015**;350:207-11
48. Ciriello G, Cerami E, Sander C, Schultz N. Mutual exclusivity analysis identifies oncogenic network modules. *Genome research* **2012**;22:398-406
49. Babur O, Gonen M, Aksoy BA, Schultz N, Ciriello G, Sander C, *et al.* Systematic identification of cancer driving signaling pathways based on mutual exclusivity of genomic alterations. *Genome biology* **2015**;16:45
50. Kigel B, Rabinowicz N, Varshavsky A, Kessler O, Neufeld G. Plexin-A4 promotes tumor progression and tumor angiogenesis by enhancement of VEGF and bFGF signaling. *Blood* **2011**;118:4285-96
51. Aran D, Sirota M, Butte AJ. Systematic pan-cancer analysis of tumour purity. *Nature communications* **2015**;6:8971
52. Li Y, Sahni N, Pancsa R, McGrail DJ, Xu J, Hua X, *et al.* Revealing the Determinants of Widespread Alternative Splicing Perturbation in Cancer. *Cell reports* **2017**;21:798-812
53. Li Y, Li L, Wang Z, Pan T, Sahni N, Jin X, *et al.* LncMAP: Pan-cancer atlas of long noncoding RNA-mediated transcriptional network perturbations. *Nucleic acids research* **2018**;46:1113-23
54. Mootha VK, Lindgren CM, Eriksson KF, Subramanian A, Sihag S, Lehar J, *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nat Genet* **2003**;34:267-73
55. Lamar KM, Miller T, Dellefave-Castillo L, McNally EM. Genotype-Specific Interaction of Latent TGFbeta Binding Protein 4 with TGFbeta. *PloS one* **2016**;11:e0150358
56. Durbin JE, Hackenmiller R, Simon MC, Levy DE. Targeted disruption of the mouse Stat1 gene results in compromised innate immunity to viral disease. *Cell* **1996**;84:443-50
57. Schneider K, Marbaix E, Bouzin C, Hamoir M, Mahy P, Bol V, *et al.* Immune cell infiltration in head and neck squamous cell carcinoma and patient outcome: a retrospective study. *Acta oncologica* **2018**;1-8
58. Li T, Fan J, Wang B, Traugh N, Chen Q, Liu JS, *et al.* TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res* **2017**;77:e108-e10
59. Li B, Li T, Pignon JC, Wang B, Wang J, Shukla SA, *et al.* Landscape of tumor-infiltrating T cell repertoire of human cancers. *Nat Genet* **2016**;48:725-32
60. Ochel A, Cebula M, Riehn M, Hillebrand U, Lipps C, Schirmbeck R, *et al.* Effective intrahepatic CD8+ T-cell immune responses are induced by low but not high numbers of antigen-expressing hepatocytes. *Cellular & molecular immunology* **2016**;13:805-15

61. Yang CW, Strong BS, Miller MJ, Unanue ER. Neutrophils influence the level of antigen presentation during the immune response to protein antigens in adjuvants. *Journal of immunology* **2010**;185:2927-34
62. Schumacher TN, Schreiber RD. Neoantigens in cancer immunotherapy. *Science* **2015**;348:69-74
63. Desrichard A, Snyder A, Chan TA. Cancer Neoantigens and Applications for Immunotherapy. *Clin Cancer Res* **2016**;22:807-12
64. Lundegaard C, Lamberth K, Harndahl M, Buus S, Lund O, Nielsen M. NetMHC-3.0: accurate web accessible predictions of human, mouse and monkey MHC class I affinities for peptides of length 8-11. *Nucleic Acids Res* **2008**;36:W509-12
65. Chen N, Fang W, Lin Z, Peng P, Wang J, Zhan J, *et al.* KRAS mutation-induced upregulation of PD-L1 mediates immune escape in human lung adenocarcinoma. *Cancer Immunol Immunother* **2017**;66:1175-87
66. Dong ZY, Zhong WZ, Zhang XC, Su J, Xie Z, Liu SY, *et al.* Potential Predictive Value of TP53 and KRAS Mutation Status for Response to PD-1 Blockade Immunotherapy in Lung Adenocarcinoma. *Clin Cancer Res* **2017**;23:3012-24
67. Ng PK, Li J, Jeong KJ, Shao S, Chen H, Tsang YH, *et al.* Systematic Functional Annotation of Somatic Mutations in Cancer. *Cancer Cell* **2018**;33:450-62 e10
68. Zhuang ZM, Wang XT, Zhang L, Tao LY, Lv B. The effect of PDIA3 gene knockout on the mucosal immune function in IBS rats. *Int J Clin Exp Med* **2015**;8:6866-77
69. Timon M, Beverley PC. Structural and functional analysis of the human CD45 gene (PTPRC) upstream region: evidence for a functional promoter within the first intron of the gene. *Immunology* **2001**;102:180-9
70. Lucas CL, Chandra A, Nejentsev S, Condliffe AM, Okkenhaug K. PI3Kdelta and primary immunodeficiencies. *Nat Rev Immunol* **2016**;16:702-14
71. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, *et al.* A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. *Cell* **2017**;171:1437-52 e17
72. Woo JH, Baik HJ, Kim CH, Chung RK, Kim DY, Lee GY, *et al.* Effect of Propofol and Desflurane on Immune Cell Populations in Breast Cancer Patients: A Randomized Trial. *Journal of Korean medical science* **2015**;30:1503-8
73. Lee JH, Kang SH, Kim Y, Kim HA, Kim BS. Effects of propofol-based total intravenous anesthesia on recurrence and overall survival in patients after modified radical mastectomy: a retrospective study. *Korean journal of anesthesiology* **2016**;69:126-32
74. Liang P, Pardee AB. Analysing differential gene expression in cancer. *Nature reviews Cancer* **2003**;3:869-76
75. Brentani RR, Carraro DM, Verjovski-Almeida S, Reis EM, Neves EJ, de Souza SJ, *et al.* Gene expression arrays in cancer research: methods and applications. *Critical reviews in oncology/hematology* **2005**;54:95-105
76. Lee S, Margolin K. Cytokines in cancer immunotherapy. *Cancers* **2011**;3:3856-93
77. Dranoff G. Cytokines in cancer pathogenesis and cancer therapy. *Nature reviews Cancer* **2004**;4:11-22
78. Ghaffari K, Hashemi M, Ebrahimi E, Shirkoobi R. BIRC5 Genomic Copy Number Variation in Early-Onset Breast Cancer. *Iranian biomedical journal* **2016**;20:241-5
79. Pereira C, Gimenez-Xavier P, Pros E, Pajares MJ, Moro M, Gomez A, *et al.* Genomic Profiling of Patient-Derived Xenografts for Lung Cancer Identifies B2M Inactivation Impairing Immunorecognition. *Clinical cancer research : an official journal of the American Association for Cancer Research* **2017**;23:3203-13

80. Ling A, Lofgren-Burstrom A, Larsson P, Li X, Wikberg ML, Oberg A, *et al.* TAP1 down-regulation elicits immune escape and poor prognosis in colorectal cancer. *Oncoimmunology* **2017**;6:e1356143
81. Cui D, Wang J, Zeng Y, Rao L, Chen H, Li W, *et al.* Generating hESCs with reduced immunogenicity by disrupting TAP1 or TAPBP. *Bioscience, biotechnology, and biochemistry* **2016**;80:1484-91
82. Liu J, Dong W, Wang T, Liu L, Zhan L, Shi Y, *et al.* Effects of etomidate and propofol on immune function in patients with lung adenocarcinoma. *American journal of translational research* **2016**;8:5748-55

Figure legends

Figure-1. Immune response prediction signatures in cancer. (A) Literature curated immune response prediction signatures. (B) Boxplots show the distribution of Immune score, MHC score and CYT score in immune response vs non-response patients. (C) Receiver operating characteristics (ROC) analysis of immune score, MHC score and CYT score from each prediction. (D) The distribution of immune score, MHC score and CYT score across 33 cancer types in TCGA cohort. Cancers are ranked by the median scores. (E) Scatter plots show the correlation between, immune score, MHC score and CYT score across cancer types. *** $p < 0.001$.

Figure-2. Critical mutated immune pathway and genes in cancer. (A) Human immune pathways and number of genes in each pathway. In total, 2,273 immune-related genes in 18 pathways are analyzed. (B) The PMB score for each immune-related pathway in cancer. Each dot represents a pathway in a specific cancer type. The immune score is calculated as the product of the proportion of mutated genes in the pathway and the percentage of mutated samples. Pathways are rank by the PMB score. (C) The mutations of genes in cytokine receptor pathway in LUSC. Genes were ranked by the mutation frequency.

Figure-3. Widespread transcriptome alterations of human immunome in cancer. (A) The proportion of differentially expressed immune genes and all human genes. The odds ratios and 95% confidence levels of Fisher's exact tests are shown in the right

panel. All p-values are less than 0.05 for 18 cancer types with at least five normal samples. (B) Cluster of cancer types based on the similarity (Jaccard index) of differentially expressed immune genes. Cancer types of similar tissue origin are cluster together. (C) Heat map shows the p-values and enrichment score for the differentially expressed genes enriched in immune-related pathways. Each row represents a cancer type and each column represents an immune-related pathway. The sizes of dots show the $-\log(p\text{-values})$ and the color of dots show the enrichment score. The bar plots below the heat map show the number of cancer types enriched by each pathway. (D) - (F) Gene Set Enrichment Analysis shows enrichment repression of cytokines, cytokines receptors and activated antigen processing and presentation in breast cancer. The horizontal bar in graded color from red to blue represents the rank-ordered, non-redundant list of all genes. The vertical black lines represent the projection of immune genes in each pathway onto the ranked gene list. (D) for antigen processing and presentation pathway, (E) for cytokines pathway and (F) for pathway cytokines receptors in breast cancer.

Figure-4. Critical protein-regions related to immune response in cancer. (A) The distribution of p-values for testing the association of protein regions and immune response scores. The bottom Venn plot shows the overlap of protein-regions that are related with immune score, MHC score and CYT score. The seven overlap genes were shown. (B) Representative mutated protein regions that are related with immune response in cancer. The mutations in the EGF_CA domain of LTBP4 gene are shown. The up-panel shows the mutations in the LTBP4 and the domain region is marked by green and mutations are shown as red balls. The immune score, MHC score and CYT score distributions of the samples with domain mutations, other region mutations and wild-type are shown in the bottom-panel. ***, $p < 0.01$, Wilcox's rank sum test. (C) The overlap of protein regions which mutations were related with immune response scores and immune cell infiltration. The rectangle colored are with $p < 0.05$. (D) The overlapped seven genes mutation associated with distinct immune cell infiltration. (E) The distribution of immune cell infiltration for patients with in-domain mutations, out-domain mutations and wild type of LTBP4 gene. Left for CD8 T cell and right for Neutrophil cell.

Figure-5. Neoantigen enriched in immune-related genes. (A) The left panel shows the proportion of neoantigens in immune gene compared to all coding genes ($p=1.18e-4$, Fisher's exact test). The right panel shows the distribution of binding affinity of HLAs in wild-type and mutated peptides ($p<2.2e-16$, Wilcox's rank sum test). (B) Number of neoantigens in immune-related genes. The barplots show the number of neoantigens generated from each immune-related genes. The red dot line shows the number of neoantigens/number of total mutations in each gene. (C) The mutations distribution in the PIK3CA that generate neoantigens. (D) The mutations distribution in the EGFR that generate neoantigens. (E) The left and right panels shows the 3D structure of PIK3CA and EGFR, red dots represent the mutations in the same cluster. The middle panels show the proportion of functional mutations, the green bars for mutations generated neoantigens and grey bars for all screened mutations. $**p<0.05$ for Fisher's exact test.

Figure-6. Network-based immune candidate driver identification. (A) Flowchart to identify the candidate personalized immune driver genes and the candidate drugs for each patient. For each cancer type, expression or mutations of genes that are correlated with immune response scores are mapped to an immune gene-gene interaction network. The immune response correlation is propagated via network links and immune genes are identified based on random walk method. Drug perturbed gene signatures ($n=6100$) are integrated for identification of the drugs that enriched the candidate immune driver genes ($FDR<0.01$) of the specific patient. (B) Top ranked immune response score-related genes across cancer types. Genes are ranked by the number of cancer types that show correlation between gene expression and immune response scores. (C) Top-ranked 39 genes are identified and the corresponding pathways were shown. The subnetworks formed by these 39 genes are shown in the up-panel with node colors indicate whether they are identified by the CRISPR-Cas9. (D) The expression of PDIA3 (linked more known essential immune genes) is correlated with patients' survival in GBM. (E) ROC curve for the classification of patients response to immunotherapy based on candidate genes. (F) Kaplan-Meier survival curves for patients classified by the average expression of candidate genes.

Figure-7. Candidate driver mutations and immune-response related small molecules. (A) The distribution of HLA binding affinity for wild-type (yellow) and mutated peptides (blue). *** $p < 0.001$, Wilcoxon's rank sum test. (B) The number of mutations that induced the decreased binding of HLAs across 30 genes. (C) The 3D structure of PTPRC and candidate mutations in this gene. (D) The 3D structure of PIK3R1 and candidate mutations in this gene. (E) Candidate immune-response related drugs and the corresponding p-values in cancer types predicted by network-based method.

Figure 1

Figure 2**A** **B** **C**

Altered in 284 (58.56%) of 485 samples.

Figure 3**A** **B** **D** **F** **C** **E**

Figure 4**A** **C** **B** **D** **E**

Figure 5

Figure 6

Immune pathways

Natural Killer Cell Cytotoxicity	Chemokine Receptors
BCR Signaling Pathway	Cytokine Receptors
TCR Signaling Pathway	Cytokines
Antigen Processing and Presentation	Chemokines
Immiss	TGFb Family Member
Antimicrobials

Figure 7