

ARTICLE

DOI: 10.1038/s41467-017-01018-0

OPEN

Genomic landscape associated with potential response to anti-CTLA-4 treatment in cancers

Chan-Young Ock^{1,2}, Jun-Eul Hwang^{1,3}, Bhumsuk Keam², Sang-Bae Kim¹, Jae-Jun Shim^{1,4}, Hee-Jin Jang^{1,5}, Sarang Park¹, Bo Hwa Sohn¹, Minse Cha¹, Jaffer A. Ajani⁶, Scott Kopetz⁶, Keun-Wook Lee⁷, Tae Min Kim², Dae Seog Heo² & Ju-Seog Lee¹

Immunotherapy has emerged as a promising anti-cancer treatment, however, little is known about the genetic characteristics that dictate response to immunotherapy. We develop a transcriptional predictor of immunotherapy response and assess its prediction in genomic data from ~10,000 human tissues across 30 different cancer types to estimate the potential response to immunotherapy. The integrative analysis reveals two distinct tumor types: the mutator type is positively associated with potential response to immunotherapy, whereas the chromosome-instable type is negatively associated with it. We identify somatic mutations and copy number alterations significantly associated with potential response to immunotherapy, in particular treatment with anti-CTLA-4 antibody. Our findings suggest that tumors may evolve through two different paths that would lead to marked differences in immunotherapy response as well as different strategies for evading immune surveillance. Our analysis provides resources to facilitate the discovery of predictive biomarkers for immunotherapy that could be tested in clinical trials.

¹Department of Systems Biology, UT MD Anderson Cancer Center, Houston, TX 77054, USA. ²Department of Internal Medicine, Seoul National University Hospital, Seoul 03080, Korea. ³Department of Hematology-Oncology, Chonnam National University Medical School, Gwangju 61469, Korea. ⁴Department of Internal Medicine, School of Medicine, Kyung Hee University, Seoul 02447, Korea. ⁵Department of Molecular Oncology, The Graduate School of Medicine, Seoul National University, Seoul 08826, Korea. ⁶Department of Gastrointestinal Medical Oncology, UT MD Anderson Cancer Center, Houston, TX 77054, USA. ⁷Department of Internal Medicine, Seoul National University Bundang Hospital, Seongnam 13620, Korea. Correspondence and requests for materials should be addressed to J.-S.L. (email: jlee@mdanderson.org)

Understanding the interaction between cancer cells and the immune system has led to novel strategies for treating cancer^{1–3}. The administration of tumor-infiltrating lymphocytes (TILs), interleukin-2, and vaccinations targeting tumor-specific antigens has prompted the treatment of cancer via host immune modulation^{4, 5}. A recent strategy targeting immune checkpoints such as CTLA-4 and PD-1/PD-L1 has showed striking clinical benefit^{6–8}. However, the overall response rates of advanced solid cancers to checkpoint inhibitors have been only modest (18–38%)^{7, 8} with prolonged responses being even less common. Furthermore, marked response to immune checkpoint therapies have been limited to a subset of tumor lineages^{9–11}, suggesting that differences in organ physiology and molecular characteristics of various cancers may play a role in the efficacy of treatment response.

As seen in earlier studies demonstrating that therapeutic targets were reliable predictive biomarkers^{12, 13}, recent studies reported that tumor PD-L1 expression or its amplification was significantly associated with better response in patients undergoing anti-PD-1/PD-L1 therapies^{11, 14, 15}, although not all responders had high PD-L1 expression. Recent studies have shown that interferon-gamma target genes such as *CXCL9*, *CXCL10*, *IDO1*, *IFNG*, *HLA-DRA*, and *STAT1* are indicative of response to immunotherapy in many cancers^{16–19}. Moreover, TILs as well as PD-1 expression in TILs were also correlated with clinical outcomes¹⁴, indicating that a better understanding of the immunologic landscape could lead to the identification of useful biomarkers for immunotherapy increasing the spectrum of patients able to benefit^{20, 21}. Interestingly, recent small-scale genomic studies demonstrated significant correlation of mutational burden with response to immunotherapy^{22, 23}, suggesting that genomic alterations may dictate clinical outcomes of immunotherapies, as they do in targeted therapies. However, this contention has not been thoroughly tested in large cohorts of cancer patients across multiple cancer lineages.

In the current study, we aim to assess the potential benefit of immunotherapy across different cancer lineages and identify potential genetic markers associated with benefit of immunotherapy by developing a transcriptional profile from interventional studies integrated with unbiased systematic analysis of genomic data from The Cancer Genome Atlas (TCGA) project.

Results

Immune signature predicting response to immunotherapy. Gene expression data from a randomized phase II trial of immunotherapy with MAGE-A3 antigen in malignant melanoma without prior treatment for metastases other than isolated limb perfusion were used for analysis^{24, 25}. The tumor samples were obtained before the immunotherapy and clinical responders were defined by objective responders (complete and partial) according to RECIST 1.0²⁶ and patients showing stable disease (>4 months) or mixed response with unequivocal tumor shrinkage. In the current analysis, we identified 105 genes significantly associated with response to immunotherapy ($P < 0.005$ and 1.5-fold difference, Fig. 1a and Supplementary Data 1) and constructed a prediction model based on the Bayesian compound covariate predictor algorithm²⁷. When patients were stratified according to Bayesian probability (cutoff = 0.5), responders were well separated from non-responders (AUC = 0.83, CI: 0.72–0.93, $P < 0.001$, Fig. 1d). We next sought to determine whether the predictor could also identify potential responders to different immunotherapy like anti-CTLA-4 antibodies. When applied to data from a mouse mesothelioma model treated with anti-CTLA-4 antibodies²⁸, our model reliably separated responders from

non-responders (AUC: 0.99, $P < 0.001$, 90% sensitivity, 90% specificity) (Fig. 1b, e). We next sought to determine if predictor can identify responders in clinical setting when applied to gene expression data from melanoma tissues of patients treated with ipilimumab²⁹. Consistent with results from mouse model, our model reliably separated responders from non-responders (AUC: 0.7, $P = 0.02$) (Fig. 1c, f). Furthermore, patients classified as responders by predictor showed significantly favorable clinical outcome in both overall survival and progression-free survival ($P = 0.009$ and $P = 0.03$, respectively, Fig. 1g, h). Taken together, our data strongly suggest that the Bayesian probability of the immune signature (IS), hereafter referred to as the IS score, is associated with response to different immunotherapy approaches including MAGE-A3 antigen-based immunotherapy and anti-CTLA-4 immune checkpoint inhibitors. The prediction of responder by IS score has a good performance compared with other candidates of immune biomarker such as interferon-gamma signature^{16, 17} or cytolytic activity³⁰ (Supplementary Figs. 1–3). IS score was not well associated with response to treatment with anti-PD-1 antibody in melanoma ($N = 27$) and renal cell carcinoma ($N = 10$), suggesting potential limitation of IS score predicting response to different immunotherapies (Supplementary Figs. 4, 5). However, it is worthwhile to point out that all other immune biomarkers failed to identify responders in these cohorts, indicating that lack of association might be due to small sample size. Pathway enrichment analysis of 105 genes showed activation of immune signaling pathways (Supplementary Fig. 6). In good agreement with predicted outcomes of anti-CTLA-4 antibody treatment, the CTLA-4 pathway was significantly activated, strongly supporting the notion that IS scores are associated with immunotherapy response at the biological and molecular levels. Consistent with pathway enrichment analysis, gene network analysis identified many pro-inflammatory cytokines and related transcription factors as potential upstream regulators activated in responder patients (Supplementary Data 2). On the contrary, anti-inflammatory cytokine *IL10* and negative regulators of cytokine signaling such as *SOCS1* and *SOCS3* were activated in non-responder patients^{31, 32} (Supplementary Fig. 7A). Interesting, same analysis revealed that *MYC* is activated in non-responders (Supplementary Fig. 7B). This is in good agreement with previous study demonstrating that *MYC* is negative regulator of immune response³³.

Distribution of IS score in TCGA pan-cancer cohort. Having found that the IS score reflected response to anti-CTLA-4 immunotherapies, we applied the IS to gene expression data from TCGA pan-cancer data including samples of 30 tumor types ($N = 9081$, Supplementary Data 3) to estimate the potential response rate of each cancer lineages to immunotherapy (Supplementary Fig. 8). As expected, cancers arising from lymphoproliferative tissues, such as diffuse large B-cell lymphoma, thymoma, and acute myeloid leukemia, had the highest IS scores, further supporting the notion that the signature reliably reflects immune activity in cancer tissues. When stratified into two subcategories (potential responder: >0.5 and non-responder: <0.5), kidney clear cell carcinoma (KIRC), lung adenocarcinoma (LUAD), and cervical and endocervical cancer (CESC) had the highest median IS scores indicative of large proportion of potential responders (Fig. 2a, b). The proportion of potential responders to immunotherapy highly varied within each type of solid cancer ranging from 0.5 to 65.9%. Interestingly, among the solid cancers, skin cutaneous melanoma (SKCM) had a relatively high proportion of predicted responders (33.7%) even though the median IS score was not top-ranked, because IS scores in SKCM were skewed to a high level.

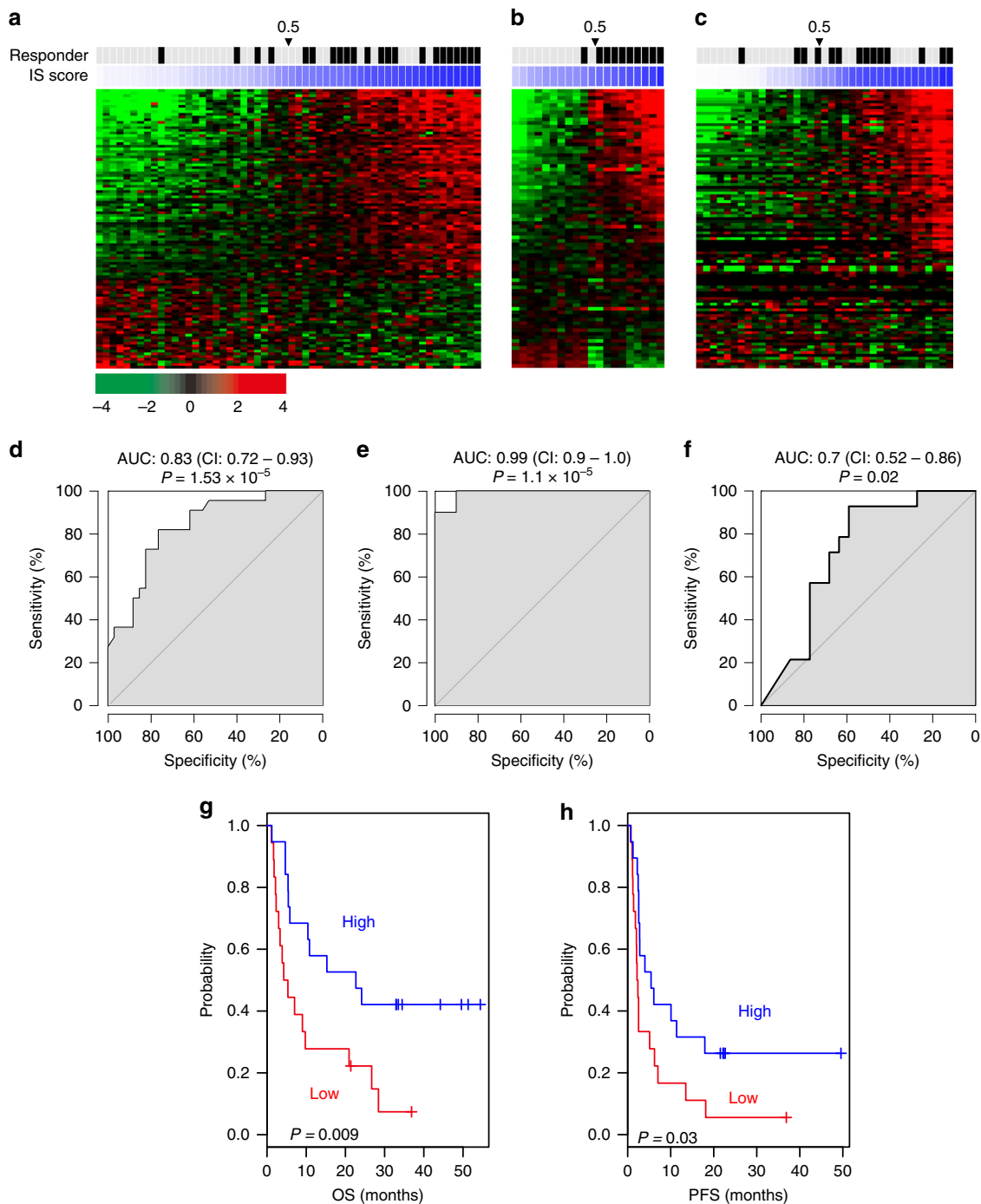


Fig. 1 Immune signature reflecting response to immunotherapy from human and mouse cancer tissues. **a** Expression patterns of genes significantly associated with response to immunotherapy in training cohort. Pretreatment biopsies from patients with metastatic melanoma were used to generate gene expression data. Genes whose expression is significantly different between responders and non-responders were selected (105 genes, $P < 0.005$ and 1.5-fold difference). The data are presented in matrix format, with rows representing the individual gene and columns representing each tissue. Each cell in the matrix represents the expression level of a gene feature in an individual tissue. Red and green reflect high-expression and low-expression levels, respectively, as indicated in the scale bar (log₂ transformed scale). Immune signature (IS) scores are presented as color index in blue (0–1 scale). **b** Expression patterns of immune signature genes and IS scores from mouse mesothelioma model treated with anti-CTLA-4 antibodies. **c** Expression patterns of immune signature genes and IS scores from human melanoma tissues treated with ipilimumab. **d–f** Receiver operating characteristics (ROC) analysis of IS scores from each prediction. Robustness of IS scores identifying responders to immunotherapy was estimated by area under curve (AUC) from ROC analysis. **d** Human melanoma treated with MAGE-A3 antigen, **e** mouse mesothelioma model treated with anti-CTLA-4 antibodies, **f** human melanoma treated with ipilimumab. (CI: 95% confident interval of AUC). **g, h** Kaplan–Meier plots of overall survival (OS) and progression-free survival (PFS) of advanced melanoma patients treated with ipilimumab. Patients were stratified according to IS scores (high >0.5). See also Supplementary Figs. 1–7 and Supplementary Data 1, 2

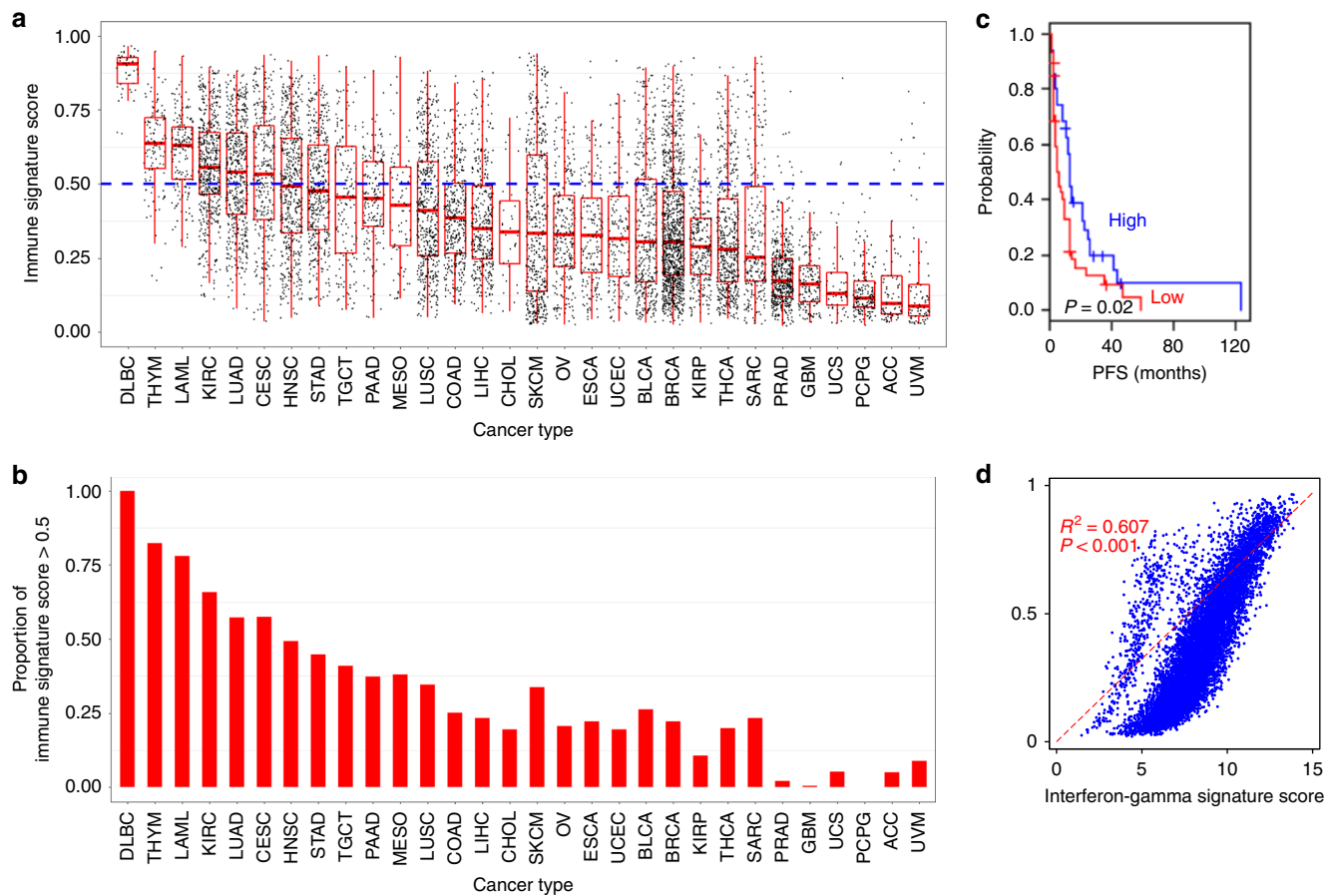


Fig. 2 Immune signature scores and potential responders to immunotherapy across major cancer types. **a** Immune signature (IS) score calculated by Bayesian probability of immune signature is plotted according to cancer types. Each black dot represents IS score, and red lines in the box represent upper 75%, median, and lower 25% values of each cancer type. Blue line represents IS score of 0.5. **b** Proportion of potential responders to immunotherapy (IS score > 0.5) is shown according to cancer types. **c** Kaplan–Meier plots of progression-free survival (PFS) of advanced melanoma patients in TCGA treated with immunotherapy. Of 472 patients with melanoma, only 78 patients treated with immunotherapy were included in analysis. Patients were stratified according to IS scores (high > 0.5). **d** Association between IS score and interferon-gamma signature score. Scatter plot and fitted dash line showed the significant association between IS score and interferon-gamma signature score. Abbreviation of cancer type was referred from The Cancer Genome Atlas tag. See also Supplementary Figs. 8–11 and Supplementary Data 3, 4.

IS score was significantly correlated with progression-free survival of 78 SKCM patients who were received immunotherapy in TCGA ($P = 0.024$, Fig. 2c and Supplementary Data 4). Moreover, IS score was significantly correlated with interferon-gamma score ($R^2 = 0.607$, $P < 0.001$), which can predict the responders of anti-PD-1 antibody in the previous studies^{16, 17}. PD-L1 mRNA expression and PD-1 mRNA expression, which were also proposed to be related with the responder of anti-PD-1/PD-L1 inhibition^{11, 14, 21}, were significantly correlated with IS scores (Supplementary Fig. 9) even though these are not components of the IS score, further supporting the notion that IS scores reflect underlying biology that determines the outcomes of immunotherapy.

Consistent with previous indications that patients with immunogenic tumors had a favorable survival outcome³⁴, patients with high IS scores (>0.5) showed significantly favorable overall survival in bladder cancer (BLCA) although they were not treated for immunotherapy (Supplementary Fig. 10).

For estimation of relative fractions of immune cells in each tumor, we used CIBERSORT to infer relative RNA fractions of 22 different immune cells³⁵. Not surprisingly, fraction of CD8⁺ T cells and M1 macrophages were most significantly associated with IS scores (Supplementary Fig. 11), further supporting that IS scores well reflect infiltrated active immune cells in tumor mass.

Association of IS scores with molecular subtypes of cancers.

We next assessed the association of IS scores with molecular subtypes defined by TCGA studies^{36–43}. In SKCM³⁶, although IS scores were only modestly associated with four mutation subtypes, they were significantly associated with platform subtypes (Supplementary Fig. 12). IS scores were significantly higher in the immune-high mRNA subtype and normal-like methylation subtype. When tumors in the immune-high mRNA subtype were further stratified by methylation subtype, IS scores were significantly higher in the normal-like subtype than in other subtypes ($P = 4.1 \times 10^{-8}$, Fig. 3a). Interestingly, IS scores were lower in the RAS subtype than in other subtype (Supplementary Fig. 12).

A TCGA study classified thyroid cancer (THCA) into BRAF-like and RAS-like subtypes³⁷. Consistent with observations in SKCM, IS scores were significantly lower in the RAS-like subtype than in the BRAF-like subtype (Supplementary Fig. 13, $P = 3.5 \times 10^{-19}$). Among platform subtypes in THCA, the C1 methylation subtype was most significantly associated with IS scores, whereas the distribution of IS scores was skewed toward high in the follicular methylation subtype (Supplementary Fig. 13). When BRAF-like subtypes were further stratified according to methylation subtype, C1 and follicular subtypes were more significantly associated with higher IS scores than other methylation subtypes ($P = 1.2 \times 10^{-30}$, Fig. 3b). In head and neck squamous cell

carcinoma (HNSC), most of the molecular subtypes were significantly associated with IS scores (Supplementary Fig. 14). IS scores were significantly higher in the C3 copy number

alteration (CNA) subtype, hypermethylation subtype, mesenchymal mRNA subtype, and C3 miRNA subtype than in all the other subtypes. This association was independent of human

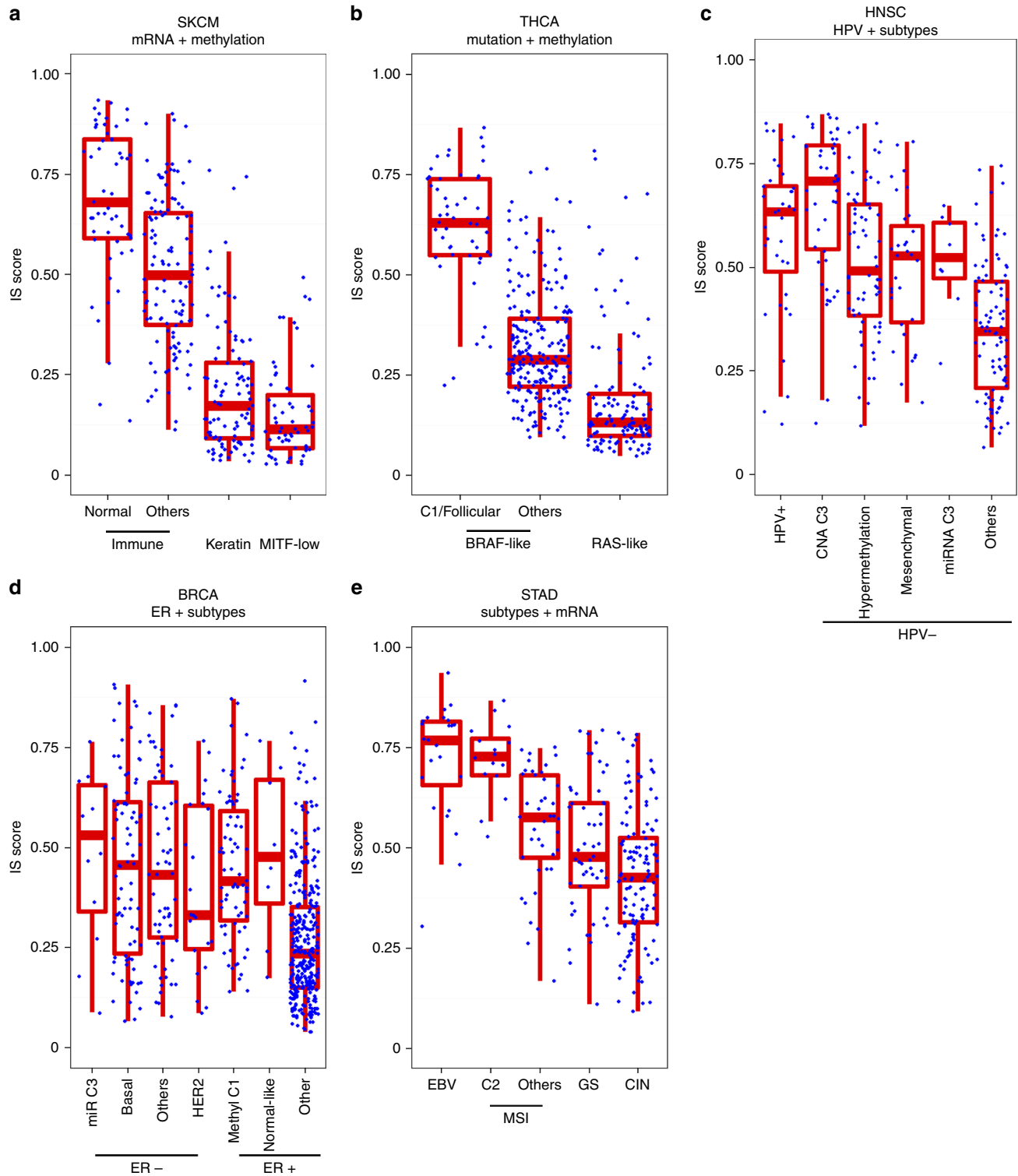


Fig. 3 Molecular subtypes of cancers associated with potential response to immunotherapy. **a** Melanoma tumors were grouped according to mRNA subtypes first and immune-high subtype was further stratified by methylation subtypes (Normal-like or others). **b** Thyroid tumors were grouped according to mutation subtypes first and BRAF-like subtype was further stratified by methylation subtypes (C1/Follicular or Others). **c** Head and neck squamous cell carcinoma tumors were grouped by HPV-infection status first and HPV-negative tumors were further stratified by molecular subtypes. **d** Breast tumors were grouped by ER status first and tumors were further stratified by molecular subtypes. **e** Stomach tumors were grouped by molecular subtypes first and MSI tumors were further stratified by mRNA subtype (C2 or others). Blue dot represents IS score, and red lines in the box represent upper 75%, median, and lower 25% values of molecular subtype. See also Supplementary Figs. 12–19

papillomavirus (HPV) status because IS scores remained high in these subtypes when HPV-negative tumors only were analyzed (Supplementary Fig. 15). IS scores in each sensitivity subtype remained high when HPV-negative tumors were subsequently stratified into different subtypes (Fig. 3c), suggesting that molecular mechanisms driving sensitivity to immunotherapy might be different in each sensitive subtype.

A TCGA study revealed a similarity between lung squamous cell carcinoma (LUSC) and HNSC^{38, 39}. In good agreement with this, IS scores were significantly higher in secretory mRNA subtypes of LUSC that is highly related to mesenchymal mRNA subtypes in HNSC (Supplementary Fig. 16, top). In LUAD⁴⁰, IS scores were significantly higher in the proximal inflammation mRNA subtype and CIMP-intermediate methylation subtype than in other subtypes (Supplementary Fig. 16, bottom). In BLCA⁴¹, IS scores were highest in the infiltrated/mesenchymal mRNA subtype (Supplementary Fig. 17), suggesting potential role of signaling events governing epithelial to mesenchymal transition in cancer immunity. In BRCA⁴², IS scores were significantly higher in estrogen receptor (ER)-negative tumors (Supplementary Fig. 18). When the BRCA subtype was further stratified, HER2 mRNA subtype had the lowest IS scores among ER-negative tumors, whereas the C1 methylation subtype and normal-like mRNA subtype had higher IS scores than other ER-positive tumors (Fig. 3d). The majority of the ER-positive showed much lower IS scores, indicating that in addition to genomic and epigenetic alterations, ER is a major determinant of cancer immunity in BRCA. In stomach adenocarcinoma (STAD)⁴³, the Epstein-Barr virus (EBV) subtype had the highest IS scores among four molecular subtypes (Supplementary Fig. 19). In the microsatellite instability (MSI) subtype, a substantial number of tumors were the C2 mRNA subtype and C2 subtype had significantly higher IS scores than others (Fig. 3e). Most interestingly, subtypes with higher IS scores in different cancers were associated with low genomic CNAs (i.e., the C6 iCluster subtype of LUAD, C3 CNA subtype of HNSC, C1 CNA subtype of BRCA, and low CNA subtype of STAD), indicating a potential connection of genomic instability to tumor immunogenicity that may govern clinical outcomes of immunotherapies.

Association of IS scores with genomic characteristics of cancer.

Since previous small-scale clinical studies indicated a potential association of mutation burdens with immunotherapy response^{22, 23}, we tested the association of IS scores with mutation rates in TCGA data set ($N = 6162$) (Supplementary Fig. 20, top). The number of predicted neoantigens³⁰ was significantly associated with the mutation rates regardless of mutation types (Supplementary Fig. 20, bottom). Interestingly, a global analysis of all tumors showed a significant positive correlation between the non-synonymous mutation rate and the IS score ($R^2 = 0.017$, $P < 0.001$, Supplementary Fig. 21, top). In particular, this association was more significant in colorectal adenocarcinoma (COAD), STAD, and BRCA (Fig. 4a and Supplementary Fig. 22, top).

Because our analysis indicated a potential association of chromosomal instability (CIN) with IS scores, we assessed the global association of CIN with IS scores ($N = 8637$) by generating a “CIN score”⁴⁴. As expected, CIN scores accurately reflected the overall CIN of tumors (Supplementary Fig. 23). Most interestingly, CIN scores had a significant negative correlation with IS scores ($R^2 = 0.095$, $P < 0.001$, Supplementary Fig. 21, bottom) to a greater degree than mutation rates. Furthermore, the trends of negative correlation were observed in most cancers (Fig. 4b and Supplementary Fig. 22, bottom), strongly suggesting that CIN might be a more important predictor of clinical outcomes of immunotherapy than mutation rates.

Since our results revealed a correlation of two genomic alterations types with IS scores, we next integrated non-synonymous mutation rates with CIN scores ($N = 5989$) to assess the interplay of two genomic alterations in cancer immunity. When two data sets were integrated, tumors were clearly separated into three major groups: tumors with high mutational burden and low CIN (mutator or M type), those with low mutational burden but high CIN (chromosome-unstable or C type), and those not otherwise specified (NOS) (Fig. 4c). Consistent with previous observation, M-type tumors had high IS scores, whereas C-type tumors had low IS scores. IS scores were significantly higher in MSI-high tumors than in MSI-low or microsatellite-stable tumors (Supplementary Fig. 24, top), consistent with MSI tumors having high mutation rates and relatively low CNAs (Supplementary Fig. 24, bottom) as well as markedly increased responses to anti-PD-1 immunotherapy²³. Furthermore, the proportion of M type was well correlated with IS score in each cancer type with the exception of KIRC (Fig. 5). Although MSI-H tumors have highest average IS scores among MSI subtypes, some of them have much higher IS scores, suggesting additional layer of regulatory mechanisms. Further analysis of gene expression data from MSI-H subtypes indicate that several interleukins (*IL4*, *IL15*, and *IL21*) are more active in tumors with high IS scores (Supplementary Fig. 25).

Because CNA can be influenced by tumor purity in tumor tissues⁴⁵, we estimated the potential impact of tumor purity in our analysis by examining the correlation of CIN scores with histologically assessed tumor purity. The correlation between CIN scores and tumor purity was only modest (Supplementary Fig. 26, top). Interestingly, non-synonymous mutation rates were also modestly correlated with tumor purity, suggesting that the correlation was not specific to CIN scores. Furthermore, the significance is not markedly altered by reanalysis of integrated data with adjusted CIN scores (Supplementary Fig. 26, bottom), strongly indicating a minimum impact of tumor purity in our analysis. To further validate insignificant contribution of tumor purity to CIN and IS scores, we adopted previously established genomic approach, consensus measurement of purity estimations (CPE)⁴⁶, for estimation of tumor purity that use gene expression, copy number alterations, and methylation data. As seen with IHC data, the correlation between CIN scores and tumor purity was modest (Supplementary Fig. 27, top) and the significance is not markedly altered by reanalysis of integrated data with adjusted CIN scores (Supplementary Fig. 27, bottom). Not surprisingly, IS scores are positively correlated with high stromal fraction in tumor mass (Supplementary Fig. 28), probably reflecting higher infiltration of immune cells.

Somatic mutations positively associated with IS scores. We next examined the association of IS scores with somatic mutations in 373 genes that have been designated drivers in previous studies⁴⁷ (Fig. 6a and Supplementary Data 5, 6). Strikingly, the majority of the significantly mutated genes were positively correlated with IS scores, suggesting that some of them might contribute host immunity. Interestingly, three of the significant genes were *MUC4*, *MUC17*, and *MUC7*, members of the mucin family that were previously identified as tumor antigens^{48–50}, and this finding supports our hypothesis. In contrast to tumor antigens, some of the positively correlated mutations might be selected under host immunogenic pressure as part of cancer cells’ mechanisms to evade immune surveillance. Mutations in *HLA-A*, *-B*, and *-C*, *B2M*, and *CASP8* might fall into this category since *CASP8* is an executor of ligand-mediated apoptosis⁵¹ and *HLA-A*, *-B*, and *-C* and *B2M* encode major antigen presenting machinery to immune cells⁵². Any loss-of-function mutations would give a significant

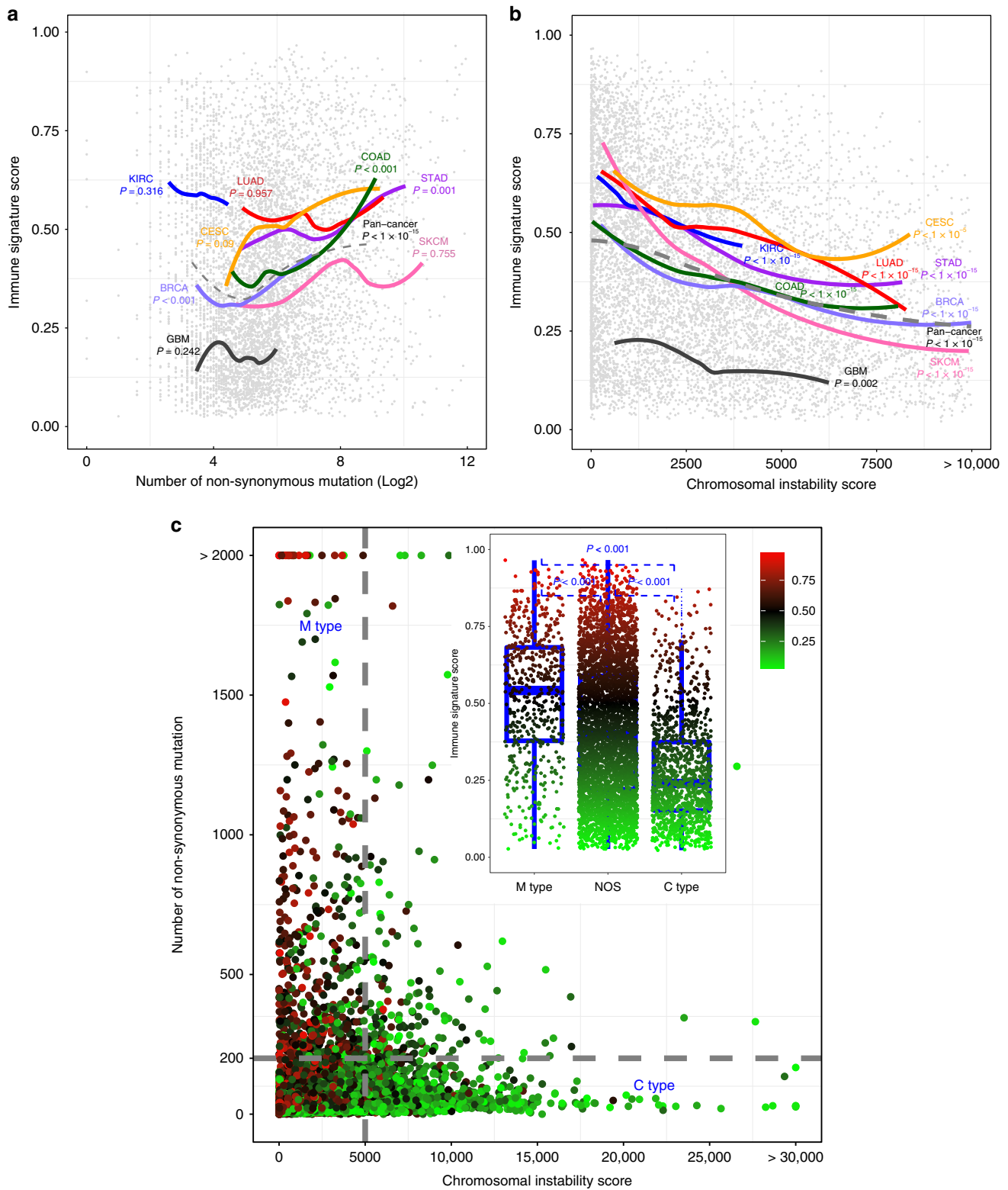


Fig. 4 Two types of tumors distinct in genomic alterations and response to immunotherapy. **a** Scatter plots between mutation rates and IS scores. Dotted line represents global regression curve between mutation rates and IS scores in all cancers. Solid lines represent local regression curves between mutation rates and IS scores in each cancer as indicated. **b** Scatter plots between chromosome instability (CIN) scores and IS scores. Dotted line represents global regression curve between two scores in all cancers. Solid lines represent local regression curves between two scores in each cancer as indicated. CIN scores were defined by sum of square of gene-level gistic 2 value. **c** Scatter plots between mutation rates and CIN scores. Size and colors of dots represents IS scores as indicated in reference index. The tumor with high mutational burden (M type) is defined by number of non-synonymous mutation more than 200 as described in previous study, whereas the tumor with high chromosomal instability (C-type) is defined by CIN score more than 5000. Otherwise, tumors are classified as not otherwise specified (NOS). IS score according to M type or C type is summarized inside the graph. Blue lines in the box represent upper 75%, median, and lower 25% values of subtype. See also Supplementary Figs. 20–28

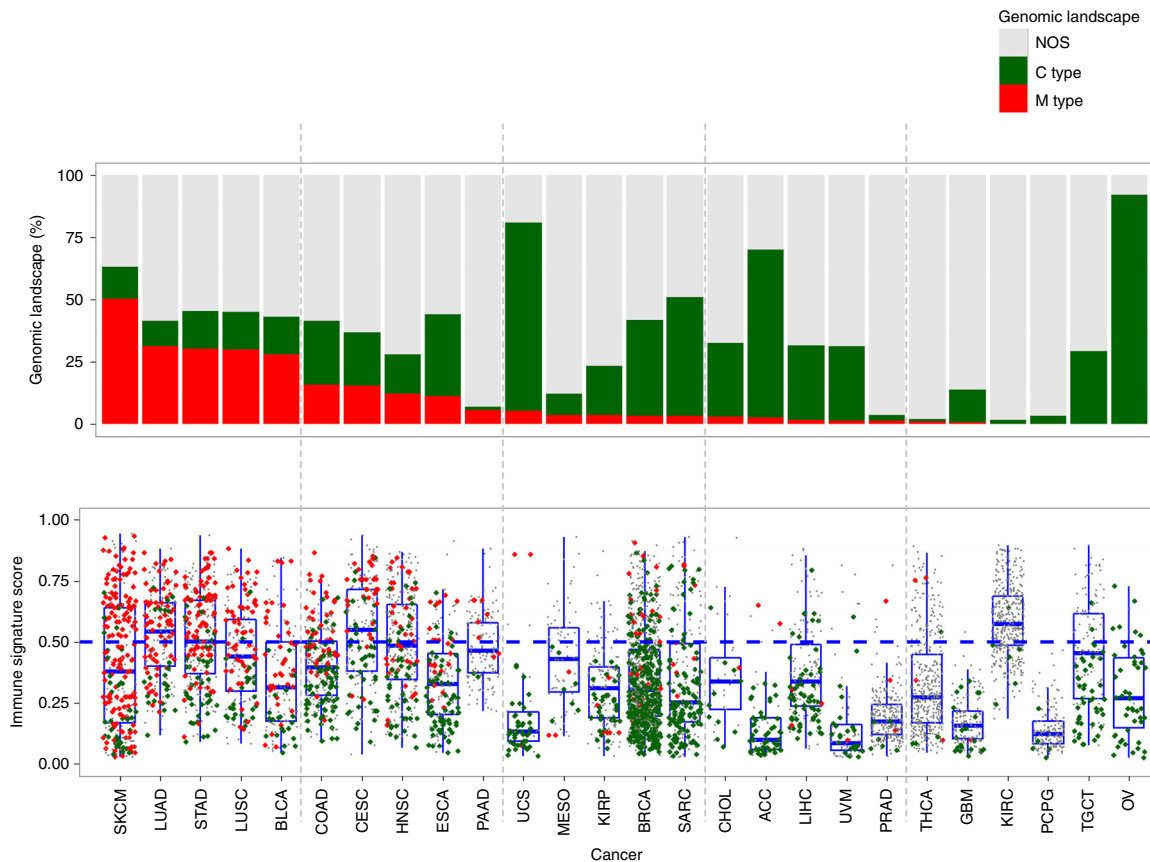


Fig. 5 Distribution of genomic cancer types in each cancer. (top) A systematic overview of proportion of mutator type tumors (M type, red), chromosome-instable tumors (C type, green), and not otherwise specified (NOS) (bottom) immune signature (IS) score in each cancer. Blue lines in the box represent upper 75%, median, and lower 25% values of cancer type. Abbreviation of cancer type was referred from The Cancer Genome Atlas tag name

advantage to cancer cells to evade immune surveillance. In good agreement, tumors with mutations in these genes represent typical M type characteristics (Fig. 6b–d).

Copy number alteration negatively associated with IS scores.

We next examined the association of IS scores with previously identified copy number-dependent drivers (87 amplification and 123 deletion)^{47,53} (Supplementary Fig. 29, Supplementary Data 5, 7, 8). In contrast to mutations, the majority of the significantly amplified genes were negatively correlated with IS scores (Fig. 7a). Likewise, the majority of the deleted genes also had a significant negative association with IS scores (Fig. 7b), suggesting that this type of genetic event is not prone to stimulate host immunity and that activated or suppressed genes may play a role in the suppression of host immunity. Consistent with our observations in BRCA (Fig. 3d), the amplification of *ERBB2* (*HER2*) was significantly associated with low IS scores. Amplified genes negatively associated IS scores include well-known driver oncogenes such as *MYC* and *E2F3* while deleted genes with negative association include well-known tumor suppressor genes such as *RB1*, *TP53*, and *PTEN*. Importantly, recent study demonstrated that loss of *PTEN* is indeed significantly associated with resistant to immunotherapy with anti-PD-1 antibodies in melanoma⁵⁴, strongly suggesting that many of identified candidates might play key roles in host immunity to cancer cells. Interestingly, expression of *HLA-A*, *HLA-B*, and *HLA-C* had a significant negative correlation with CIN scores in tumors with amplified genes or deleted genes (Fig. 7c), suggesting that some of the copy number-altered genes might be involved in the suppression of antigen presentation in cancer cells either alone or in combination with

other genes. Further supporting this notion, the expression of HLA genes was further reduced in tumors with co-amplified *MYC* and *FGFR1* (Fig. 7c).

Association of IS scores with viral presence. Not surprisingly, EBV-positive STAD and HPV-positive HNSC tumors were significantly associated with higher IS scores ($P < 0.001$, Supplementary Fig. 30). However, hepatitis B virus positivity was not associated with IS scores in liver hepatocellular carcinoma (LIHC), or other cancers (Supplementary Fig. 30, bottom right).

Discussion

In the current study, we generated IS scores based on response to different immunotherapy approaches in patients and in a model system and applied them to major cancer types. The analysis revealed two distinct types of tumors (M type and C type) that differ in their potential response to immunotherapy. Our analysis suggested that tumors evolve through two major paths that have different mechanisms for activating driver genes and may account for difference in immunotherapy response as well as strategies for evading immune surveillance.

While initially uncovered by analyzing the data from a vaccine immunotherapy approach, several lines of evidences strongly support that IS and IS scores are applicable to other types of immunotherapy. First, IS scores reliably identified responders to immunotherapy in a mouse model treated with anti-CTLA-4 antibodies. Second, pathway enrichment analysis identified the CTLA-4 pathway as one of the key pathways activated in the signature. Further supporting this finding, the iCOS-iCOSL pathway that is activated in the signature was recently

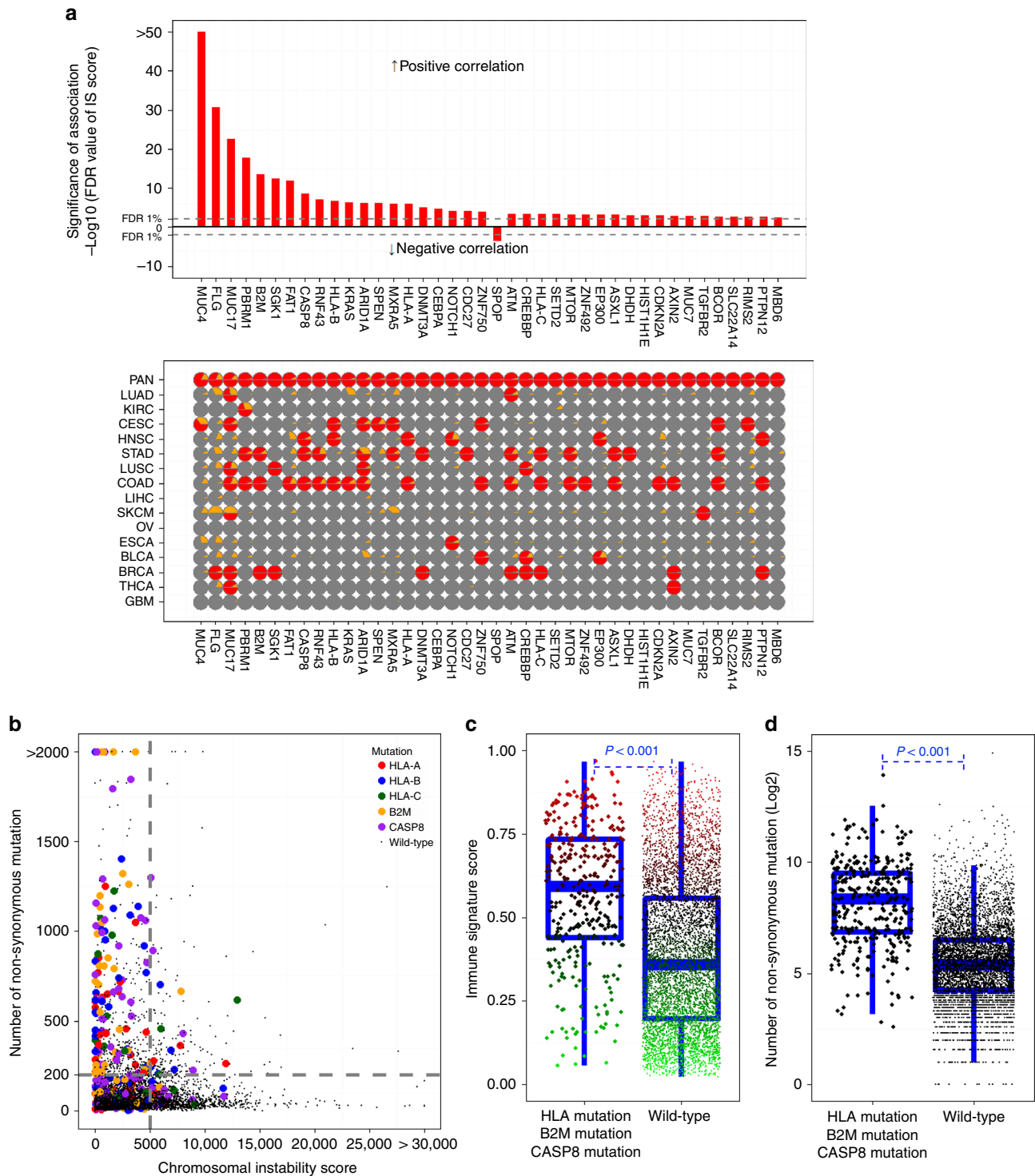


Fig. 6 Somatic mutations associated with immune signature scores. **a** Top: the significant pan-cancer-wide association of mutation rates with immune signature (IS) scores. Red bars represent significance (FDR) of association of mutation rate with IS score in each gene. Dashed line indicates 1% threshold of FDR. Bottom: the significance of association in each cancer type. Red or gray indicate significance or insignificance, respectively, and yellow indicates the frequency of mutations in each gene. **b** Scatter plots between mutation rates and chromosomal instability scores in all tumors highlighted with mutations in *HLA-A*, *HLA-B*, *HLA-C*, *B2M*, and *CASP8*. **c**, **d** IS score **c** and number of mutation **d** in tumors with and without mutations in *HLA-A*, *HLA-B*, *HLA-C*, *B2M*, and *CASP8*. Blue lines in the box represent upper 75%, median, and lower 25% values of subtype. See also Supplementary Data 5

identified as a pharmacodynamic marker for anti-CTLA-4 therapy⁵⁵. Third, most importantly, IS scores can identify responder patients with melanoma after treatment with ipilimumab²⁹. Forth, IS scores is significantly correlated with interferon-gamma score that is predictive markers for anti-PD-1 therapy in gastric

and head and neck cancer. Furthermore, IS scores were significantly associated with expression of PD-1 and PD-L1 in TCGA data. Fifth, gene network analysis identified many pro-inflammatory cytokines as activated upstream regulators in responder patients while it identified anti-inflammatory cytokines

and negative regulators of cytokine signaling as activated regulators in non-responder patients. Moreover, it also identified MYC as negative regulator of immune activity. Indeed, recent

study demonstrated that MYC is negative regulator of immune³³. Finally, IS scores predicted that MSI tumors would have strong responses to immunotherapy which is supported by clinical trial

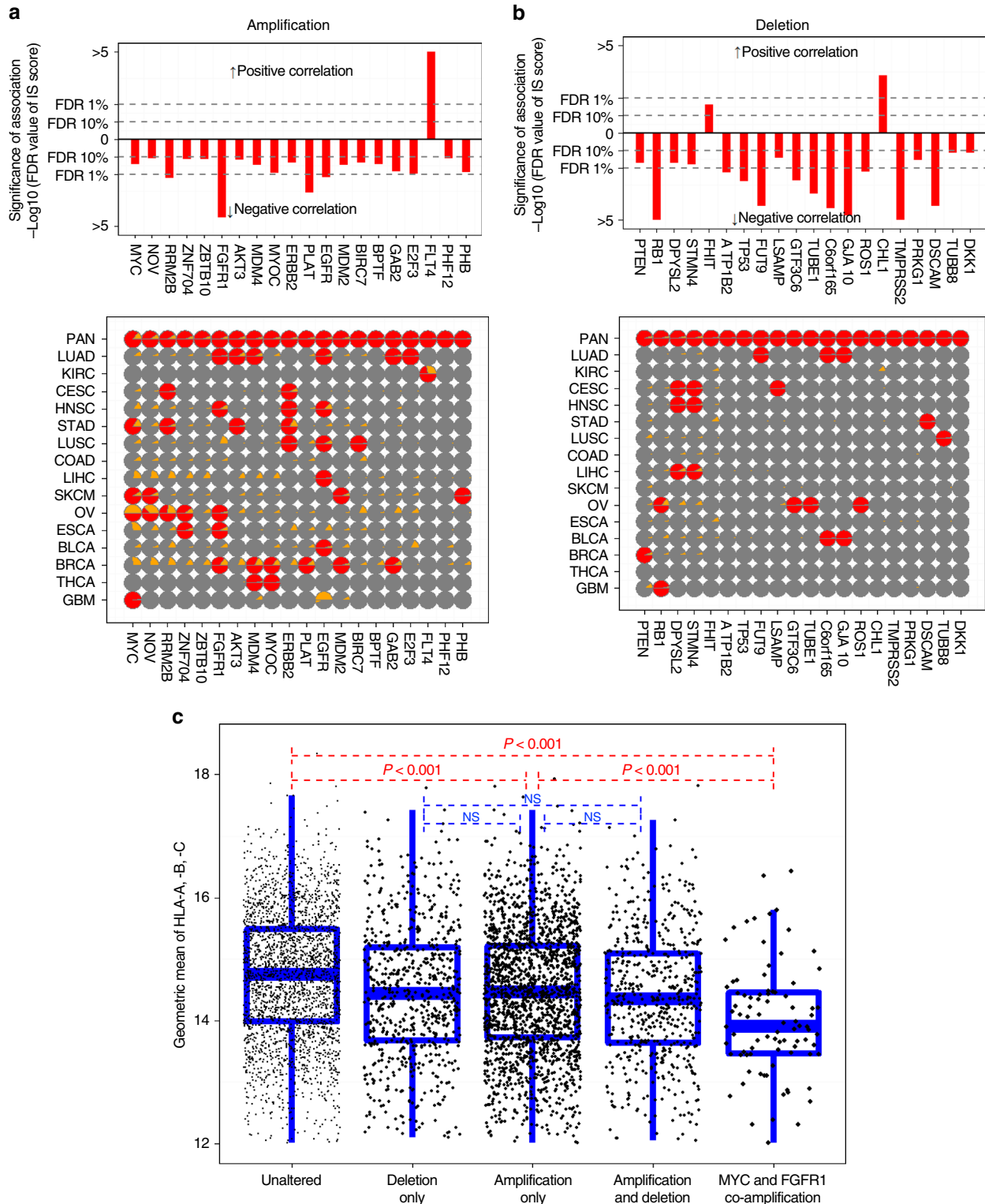


Fig. 7 Somatic copy number alterations associated with immune signature scores. **a, b** Top: the significant pan-cancer-wide association of amplified **a** and deleted **b** genes with immune signature (IS) scores. Red bars represent significance (FDR) of association of amplification with IS scores in each gene. Bottom: The significance of association in each cancer type. Red or gray indicate significance or insignificance, respectively, and yellow indicates the frequency of amplification in each gene. **c** Mean expression of HLA-A, HLA-B, and HLA-C in not significantly altered, deleted only, amplified only, both amplified and deleted, and MYC and FGFR1 co-amplified tissues. Blue lines in the box represent upper 75%, median, and lower 25% values of subtype. See also Supplementary Fig. 29 and Supplementary Data 6, 7

data²³. Taken together, these observations strongly suggest that IS scores well reflect underlying biology that may play key roles in clinical outcomes.

Although immunotherapy has led to great enthusiasm for treatment of a subset of cancer types, including melanoma, non-small cell lung cancer, and kidney cancer⁵⁶, its clinical effects have been disappointing in other tumor lineages. While recent studies using genomic approaches have begun to shed light on genomic alterations associated with the benefits of immunotherapy^{21, 22, 57}, the underlying biology of predicting benefit of immunotherapy is poorly understood. Systematic integration of somatic mutations and CNAs in connection with our IS predictor of response to immunotherapy revealed two distinct types of tumors (Supplementary Fig. 31). M-type tumors are rich in somatic mutations, low in CNAs, and likely to be sensitive to immunotherapy. Some of mutated gene products such as mucins may provide highly immunogenic antigens and may be accountable for high IS scores. In contrast to M-type tumors, C-type tumors are high in CNAs, low in mutations, and likely to be resistant to immunotherapy. This finding is in good agreement with recent study showing that high copy number alteration is potential predictive marker for immunotherapy⁵⁸. In another study analyzing of samples from clinical trials with CTLA-4 and PD-1 blockade treatments, copy number loss is associated with resistance to immunotherapy⁵⁹. Molecular mechanisms of resistance to immunotherapy is currently unknown. A lack of neoantigen production due to low mutation rates may account for the insensitivity of C-type tumors to immunotherapy, however loss of key immune mediators is also likely to contribute.

Evasion of immune surveillance is necessary for cancer cells to survive and grow⁶⁰. Two tumor types may adopt different strategies to evade immune surveillance. M-type tumors have frequent mutations in genes involved in antigen presentation. Mutations in *HLA-A*, *-B*, and *-C* and *B2M* might arise under selective pressure to evade host immunity. Likewise, mutations in *CASP8*, which is a key mediator of apoptosis⁵¹, give an advantage to cancer cells to become insensitive to T-cell-mediated cell death. Similar findings were also observed in previous study using immune cytolytic score as immune activity in tumor mass³⁰. While interesting, these associations should be interpreted with caution and need to be validated in prospective studies.

Amplified and deleted genes in C-type tumors are significantly associated with lower expression of *HLA-A*, *HLA-B*, and *HLA-C* genes, suggesting that they may suppress the expression of these antigen-presenting genes to evade immune surveillance and may account for low IS scores in C-type tumors. In good agreement with our analysis that identified *PTEN* as a key modulator of tumor immunity, recent study showed that *PTEN* play roles in T-cell activation and loss of *PTEN* is significantly associated with resistance of melanoma to immunotherapy⁵⁴. Likewise, *MYC* was predicted to be negative regulator of tumor immunity. Recent study also demonstrated that *MYC* inhibits T-cell activation by upregulating *CD47* and *PD-L1*⁶¹, further supporting validity of our approaches. Therefore, it is important to determine in future experiments whether other amplified or deleted genes are secondary therapeutic targets that can improve the efficacy of immunotherapy. The proportions of M-type tumors were generally well correlated with IS score in many cancer types. However, KIRC tumors had very high IS scores and proportion of M-type tumors was low (Fig. 5), suggesting that a high mutation rate does not fully account for IS scores.

IS scores are clearly associated with clinical subtypes of cancers. As expected, tumors with viral infection had high likelihood of response to immunotherapy. Our analysis also showed that the benefits of immunotherapy may not be limited to viral infected tumors. In HNSC, IS scores in the C3 CNA and mesenchymal mRNA subtypes were higher than or almost equal to IS scores in HPV-positive tumors. In STAD, the C2 mRNA subtype had equally high IS scores

with EBV-positive tumors. Furthermore, a subset analysis of SKCM, THCA, and BRCA tumors showed a significant difference in IS scores among clinical subtypes. In good agreement with a previous study showing a high response rate of basal type breast cancer (24%) to anti-PD-L1 antibody⁶², the basal subtype had high IS scores in our analysis. These results indicate that subtype-specific biomarkers would improve the efficacy of immunotherapy in future trials.

The results of our study should be further validated in a prospective cohort of patients receiving immunotherapy. Although IS scores were validated in melanoma patients treated with anti-CTLA-4 antibodies, we cannot rule out the possibility that IS scores are more specific to tumor vaccines and probably to melanoma. Moreover, our result should be interpreted carefully when it applied to other cancer types as IS score is mostly validated in melanoma. Differences in genetic makeup of cancer cells and tumor microenvironment might have substantial influence on IS score in other cancer types. This should be further tested and validated in future studies with data from prospectively collected samples. As not all patients with high IS scores have greater benefit of immunotherapy, more clinical factors should be incorporated to prediction models for improvement of accuracy. As tumor tissues in TCGA are relatively in the early stages, our results should be interpreted with caution since later stages of tumors may have different composition of immune cells.

In the current study, we showed that the potential benefit of immunotherapy highly varies across cancer lineages and revealed global subtypes of tumors and genomic alterations significantly associated with the potential benefit of immunotherapy. Our findings could lead to opportunities to discover new biomarkers for immunotherapy that can identify subsets of patients who could derive greater benefit from immunotherapy.

Methods

Genomic and clinical data sets. We used publicly available data in the current study. Gene expression data used for identification of IS and generation of IS score (accession number GSE35640²⁴), and validation of IS scores in mouse model treated with anti-CTLA-4 antibody (accession number GSE63557²⁸), in human melanoma patients treated with anti-PD-1 antibody (accession number GSE78220⁶³), human renal cell carcinoma patients treated with anti-PD-1 antibody (accession number: GSE67501⁶⁴) were obtained from Gene Expression Omnibus database (<http://www.ncbi.nlm.nih.gov/geo>). Another data set of RNA expressions regarding the validation of IS scores in human melanoma treated with anti-CTLA-4 antibody²⁹ was generously given by the authors (Van Allen EM and Garraway LA). All other data from TCGA project were obtained from TCGA data portal (<https://tcga-data.nci.nih.gov>) and cancer browser (<https://genome-cancer.ucsc.edu>). Gene-level gene expression data from RNA-seq experiments ($N = 9081$), copy number variation data ($N = 8785$), tumor purity data ($N = 8149$), somatic mutation data ($N = 6162$), clinical information data (overall survival, $N = 8522$), and microsatellite status of tumors ($N = 1103$) were included in analyses. Among TCGA data set, we excluded data for brain lower grade glioma due to indolent behavior and kidney chromophobe renal cell carcinoma due to rare incidence and far different tumor biology to other renal cell carcinoma. Altogether, samples of 30 major cancer types ($N = 9081$) were included in the final analysis (Supplementary Data 3). Somatic mutation data of *HLA-A*, *HLA-B*, and *HLA-C* genes were obtained from previous study that used the algorithm Polysolver⁶⁵. HLA somatic mutations were available in 6162 patients for our analysis. Viral presence status and number of predicted neoantigen of TCGA samples ($N = 3658$) was obtained from a previous publication³⁰. Genetic and molecular subtypes of skin cutaneous melanoma, thyroid cancer, head and neck squamous cell carcinoma, breast cancer, stomach adenocarcinoma, lung adenocarcinoma, lung squamous cell carcinoma, and bladder urothelial cell carcinoma were obtained from previous TCGA publications^{36–43}. Of 472 patients with skin cutaneous melanoma, 78 patients treated with immunotherapy which purpose is not indicated by “adjuvant” with appropriately annotated survival data were included in progression-free survival analysis.

Analysis of the data, IS scores, and CIN scores. For the number of total somatic mutations, multiple somatic mutations including non-synonymous mutation, insertion-deletion mutation, and silent mutations were respectively counted and summed, but germline mutation was excluded. Gene expression data from microarrays was normalized using a robust multiarray averaging method⁶⁶. The BRB-ArrayTools software program (<http://linus.nci.nih.gov/BRB-ArrayTools.html>) was used to analyze gene expression data⁶⁷. A heatmap was generated using the Cluster and TreeView software programs⁶⁸. Other statistical analyses were performed in the R

language (<http://www.r-project.org>) or using STATA version 12 (StataCorp LP, College Station, TX, USA). To select genes that were differentially expressed between responder and non-responder of the training cohort (GSE35640)²⁴, we applied stringent cutoff of $P < 0.005$ (Student's *t*-test) and 1.5-fold difference and identified 105 genes. The signature was used to stratify patients in a validation cohort of GSE63557²⁸, human melanoma treated with anti-CTLA-4 antibody²⁹, and TCGA data. Of 105 genes of the training set, 27, 6, and 6 genes were excluded during validation with GSE63557 data, Van Allen et al. data, and TCGA data, respectively, due to difference in number of probes in microarray platforms or RNA-seq data (Supplementary Data 1). Gene expression data for the training and test sets were re-normalized by centralizing the gene expression level across the tissues. Briefly, expression data for 105 immune signature genes in the training set were combined to form a classifier according to a Bayesian compound covariate predictor (BCCP)⁶⁹. The BCCP classifier estimated the likelihood that an individual patient had either a high immune signature or a low immune signature, according to a Bayesian probability of IS score cutoff of 0.5, which was optimized by comparing previously reported response rates to immunologic agents^{7, 9–11} and the results of the current analysis of Receiver operating characteristic to predict the responder of a training cohort and a separate validation model by which cutoff is set by maximal point of sum of sensitivity and specificity⁷⁰. To assess the degree of copy number variation which was calculated by Gistic 2.0⁴⁴, we defined "CIN score" as the summation of square of gene-level gistic 2 values. Adjusted CIN scores were computed by multiplying purity score (in range from 0 to 1) to original CIN scores.

Canonical signaling pathways enriched in IS score. Pathway analysis was carried by using Ingenuity Pathways Analysis and genes from the data set that were associated with a canonical pathway in the Ingenuity Pathways Knowledge Base were considered for the analysis. The significance of the association between immune signature and the canonical pathway was measured Fischer's exact test ($P < 0.001$). Among identified significant pathways, top 30 pathways were only reported in Supplementary Fig. 6. To estimate relative proportion of 22 types of infiltrated immune cells in tumor mass, online analytical platform CIBERSORT (<https://cibersort.stanford.edu/>) was used³⁵.

Survival analysis. Using IS score to dichotomize the patients into two subgroups (cutoff of 0.5), the prognostic significance was estimated using Kaplan–Meier plots (log-rank tests) and Cox proportional hazards regression analysis and then adjusted and stratified by cancer type. Prognostic significance of the continuous value of IS score was also calculated by Cox proportional hazards regression analysis. P -value < 0.05 was considered as a significant difference. Overall survival was gathered from TCGA clinical data, "days_to_last_follow-up" (CDE_ID: 3008273) if censored, or "days_to_death" (CDE_ID: 3165475) if dead. Progression-free survival (PFS) was measured from "days_to_drug_therapy_start" (CDE_ID: 3392465) until "days_to_drug_therapy_end" (CDE_ID: 3392470). PFS event was gathered from "therapy_ongoing" (CDE: 3103479).

Significance of IS score according to genomic alterations. The significance of global correlation between IS scores and number of mutations or CIN scores was estimated by linear regression analysis or generalized additive models (GAM) using R-Project statistical package. The significance of IS score difference according to clinicopathologic features such as the presence of virus, and mutation was estimated by Wilcoxon rank-sum test or analysis of variance (if more than three groups were compared). For each cancer type, we performed logistic analysis with IS score as the independent variable, and dichotomized status in genomic data such as higher or lower than median mutation number or CIN scores as the dependent variables. $P < 0.05$ was considered a significant difference.

To find specific mutations significantly associated with IS scores, Wilcoxon rank-sum tests were applied to the mean difference of IS score according to each mutation status (mutated versus wild-type). Likewise, significant difference of IS score by amplified or deleted genes were also identified by Wilcoxon rank-sum tests. To facilitate analysis, we limited analysis with previously recognized 373 driver genes⁴⁷ for mutation analysis and 87 amplified and 123 deleted genes⁵³ for CIN analysis. To estimate the significance of correlation in each cancer type, subgroup analysis of logistic regression was carried out to compute odds ratio (OR) of mutation rate or CIN score. False discovery rates were applied to control type I errors.

Data availability. The genomic data that support findings of this study are available from the NCBI Gene Expression Omnibus (GEO, <http://www.ncbi.nlm.nih.gov/geo/>) under accession number GSE35640, GSE63557, and GSE78220. Genomic data from TCGA project are available from the National Cancer Institute's Genomic Data Commons (<https://gdc.cancer.gov/>). All other data supporting the findings of this study are available within the article and its supplementary information files or from the corresponding author upon reasonable request.

Received: 23 June 2016 Accepted: 14 August 2017

Published online: 19 October 2017

References

- Dong, H. et al. Tumor-associated B7-H1 promotes T-cell apoptosis: a potential mechanism of immune evasion. *Nat. Med.* **8**, 793–800 (2002).
- Leach, D. R., Krummel, M. F. & Allison, J. P. Enhancement of antitumor immunity by CTLA-4 blockade. *Science* **271**, 1734–1736 (1996).
- Wang, E. et al. Antitumor vaccines, immunotherapy and the immunological constant of rejection. *IDrugs* **12**, 297–301 (2009).
- Rosenberg, S. A. et al. Use of tumor-infiltrating lymphocytes and interleukin-2 in the immunotherapy of patients with metastatic melanoma. A preliminary report. *N. Engl. J. Med.* **319**, 1676–1680 (1988).
- Schwartzentruber, D. J. et al. gp100 peptide vaccine and interleukin-2 in patients with advanced melanoma. *N. Engl. J. Med.* **364**, 2119–2127 (2011).
- Hodi, F. S. et al. Improved survival with ipilimumab in patients with metastatic melanoma. *N. Engl. J. Med.* **363**, 711–723 (2010).
- Topalian, S. L. et al. Safety, activity, and immune correlates of anti-PD-1 antibody in cancer. *N. Engl. J. Med.* **366**, 2443–2454 (2012).
- Hamid, O. et al. Safety and tumor responses with lambrolizumab (anti-PD-1) in melanoma. *N. Engl. J. Med.* **369**, 134–144 (2013).
- Robert, C. et al. Pembrolizumab versus ipilimumab in advanced melanoma. *N. Engl. J. Med.* **372**, 2521–2532 (2015).
- Brahmer, J. et al. Nivolumab versus docetaxel in advanced squamous-cell non-small-cell lung cancer. *N. Engl. J. Med.* **373**, 123–135 (2015).
- Garon, E. B. et al. Pembrolizumab for the treatment of non-small-cell lung cancer. *N. Engl. J. Med.* **372**, 2018–2028 (2015).
- Slamon, D. J. et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N. Engl. J. Med.* **344**, 783–792 (2001).
- Maemondo, M. et al. Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N. Engl. J. Med.* **362**, 2380–2388 (2010).
- Herbst, R. S. et al. Predictive correlates of response to the anti-PD-L1 antibody MPDL3280A in cancer patients. *Nature* **515**, 563–567 (2014).
- Ansell, S. M. et al. PD-1 blockade with nivolumab in relapsed or refractory Hodgkin's lymphoma. *N. Engl. J. Med.* **372**, 311–319 (2015).
- Muro, K. et al. Pembrolizumab for patients with PD-L1-positive advanced gastric cancer (KEYNOTE-012): a multicentre, open-label, phase 1b trial. *Lancet Oncol.* **17**, 717–726 (2016).
- Seiwert, T. Y. et al. Safety and clinical activity of pembrolizumab for treatment of recurrent or metastatic squamous cell carcinoma of the head and neck (KEYNOTE-012): an open-label, multicentre, phase 1b trial. *Lancet Oncol.* **17**, 956–965 (2016).
- Wang, E., Bedognetti, D. & Marincola, F. M. Prediction of response to anticancer immunotherapy using gene signatures. *J. Clin. Oncol.* **31**, 2369–2371 (2013).
- Weiss, G. R. et al. Molecular insights on the peripheral and intratumoral effects of systemic high-dose rIL-2 (aldesleukin) administration for the treatment of metastatic melanoma. *Clin. Cancer Res.* **17**, 7440–7450 (2011).
- Galon, J., Angell, H. K., Bedognetti, D. & Marincola, F. M. The continuum of cancer immunosurveillance: prognostic, predictive, and mechanistic signatures. *Immunity* **39**, 11–26 (2013).
- Ock, C. Y. et al. Pan-cancer immunogenomic perspective on the tumor microenvironment based on PD-L1 and CD8 T-cell infiltration. *Clin. Cancer Res.* **22**, 2261–2270 (2016).
- Rizvi, N. A. et al. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **348**, 124–128 (2015).
- Le, D. T. et al. PD-1 blockade in tumors with mismatch-repair deficiency. *N. Engl. J. Med.* **372**, 2509–2520 (2015).
- Ulloa-Montoya, F. et al. Predictive gene signature in MAGE-A3 antigen-specific cancer immunotherapy. *J. Clin. Oncol.* **31**, 2388–2395 (2013).
- Kruit, W. H. et al. Selection of immunostimulant AS15 for active immunization with MAGE-A3 protein: results of a randomized phase II study of the European organisation for research and treatment of cancer melanoma group in metastatic melanoma. *J. Clin. Oncol.* **31**, 2413–2420 (2013).
- Therasse, P. et al. New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. *J. Natl Cancer Inst.* **92**, 205–216 (2000).
- Ramaswamy, S. et al. Multiclass cancer diagnosis using tumor gene expression signatures. *Proc. Natl Acad. Sci. USA* **98**, 15149–15154 (2001).
- Lesterhuis, W. J. et al. Network analysis of immunotherapy-induced regressing tumours identifies novel synergistic drug combinations. *Sci. Rep.* **5**, 12298 (2015).
- Van Allen, E. M. et al. Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
- Rooney, M. S., Shukla, S. A., Wu, C. J., Getz, G. & Hacohen, N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* **160**, 48–61 (2015).

31. de Waal Malefyt, R., Abrams, J., Bennett, B., Figdor, C. G. & de Vries, J. E. Interleukin 10(IL-10) inhibits cytokine synthesis by human monocytes: an autoregulatory role of IL-10 produced by monocytes. *J. Exp. Med.* **174**, 1209–1220 (1991).
32. Endo, T. A. et al. A new protein containing an SH2 domain that inhibits JAK kinases. *Nature* **387**, 921–924 (1997).
33. Casey, S. C. et al. MYC regulates the antitumor immune response through CD47 and PD-L1. *Science* **352**, 227–231 (2016).
34. Gooden, M. J., de Bock, G. H., Leffers, N., Daemen, T. & Nijman, H. W. The prognostic influence of tumour-infiltrating lymphocytes in cancer: a systematic review with meta-analysis. *Br. J. Cancer* **105**, 93–103 (2011).
35. Newman, A. M. et al. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **12**, 453–457 (2015).
36. Cancer Genome Atlas N. Genomic classification of cutaneous melanoma. *Cell* **161**, 1681–1696 (2015).
37. Cancer Genome Atlas Research N. Integrated genomic characterization of papillary thyroid carcinoma. *Cell* **159**, 676–690 (2014).
38. Cancer Genome Atlas N. Comprehensive genomic characterization of head and neck squamous cell carcinomas. *Nature* **517**, 576–582 (2015).
39. Cancer Genome Atlas Research N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
40. Cancer Genome Atlas Research N. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **511**, 543–550 (2014).
41. Cancer Genome Atlas Research N. Comprehensive molecular characterization of urothelial bladder carcinoma. *Nature* **507**, 315–322 (2014).
42. Cancer Genome Atlas N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
43. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. *Nature* **513**, 202–209 (2014).
44. Mermel, C. H. et al. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol.* **12**, R41 (2011).
45. Yuan, Y. et al. Quantitative image analysis of cellular heterogeneity in breast tumors complements genomic profiling. *Sci. Transl. Med.* **4**, 157ra143 (2012).
46. Aran, D., Sirota, M. & Butte, A. J. Systematic pan-cancer analysis of tumour purity. *Nat. Commun.* **6**, 8971 (2015).
47. Lawrence, M. S. et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).
48. Wu, Y. L. et al. INSPIRE: a phase III study of the BLP25 liposome vaccine (L-BLP25) in Asian patients with unresectable stage III non-small cell lung cancer. *BMC Cancer* **11**, 430 (2011).
49. Roulois, D., Gregoire, M. & Fonteneau, J. F. MUC1-specific cytotoxic T lymphocytes in cancer therapy: induction and challenge. *Biomed. Res. Int.* **2013**, 871936 (2013).
50. Torres, M. P., Chakraborty, S., Soucek, J. & Batra, S. K. Mucin-based targeted pancreatic cancer therapy. *Curr. Pharm. Des.* **18**, 2472–2481 (2012).
51. Crowder, R. N. & El-Deiry, W. S. Caspase-8 regulation of TRAIL-mediated cell death. *Exp. Oncol.* **34**, 160–164 (2012).
52. Leone, P. et al. MHC class I antigen processing and presenting machinery: organization, function, and defects in tumor cells. *J. Natl Cancer Inst.* **105**, 1172–1187 (2013).
53. Beroukhi, R. et al. The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
54. Peng, W. et al. Loss of PTEN promotes resistance to T cell-mediated immunotherapy. *Cancer Discov.* **6**, 202–216 (2016).
55. Ng Tang, D. et al. Increased frequency of ICOS+CD4 T cells as a pharmacodynamic biomarker for anti-CTLA-4 therapy. *Cancer Immunol. Res.* **1**, 229–234 (2013).
56. Rosenberg, S. A. Progress in human tumour immunology and immunotherapy. *Nature* **411**, 380–384 (2001).
57. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
58. Davoli, T., Uno, H., Wooten, E. C., Elledge, S. J. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. *Science* **355**, eaaf8399 (2017).
59. Roh W. 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.* **9**, eaah3560 (2017).
60. Schreiber, R. D., Old, L. J. & Smyth, M. J. Cancer immunoeediting: integrating immunity's roles in cancer suppression and promotion. *Science* **331**, 1565–1570 (2011).
61. Jaiswal, S. et al. CD47 is upregulated on circulating hematopoietic stem cells and leukemia cells to avoid phagocytosis. *Cell* **138**, 271–285 (2009).
62. Gibson, J. Anti-PD-L1 for metastatic triple-negative breast cancer. *Lancet Oncol.* **16**, e264 (2015).
63. Hugo, W. et al. Genomic and transcriptomic features of response to Anti-PD-1 therapy in metastatic melanoma. *Cell* **168**, 542 (2017).
64. Ascierto, M. L. et al. Transcriptional mechanisms of resistance to anti-PD-1 therapy. *Clin. Cancer Res.* **23**, 3168–3180 (2017).
65. Shukla, S. A. et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. *Nat. Biotechnol.* **33**, 1152–1158 (2015).
66. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
67. Simon, R. et al. Analysis of gene expression data using BRB-ArrayTools. *Cancer Inform.* **3**, 11–17 (2007).
68. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
69. Oh, S. C. et al. Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer. *Gut* **61**, 1291–1298 (2012).
70. DeLong, E. R., DeLong, D. M. & Clarke-Pearson, D. L. Comparing the areas under two or more correlated receiver operating characteristic curves: a nonparametric approach. *Biometrics* **44**, 837–845 (1988).

Acknowledgements

This study was supported in part by National Institutes of Health grant CA150229, CPRIT RP170307, UT M.D. Anderson Cancer Center 2016, Institutional Research Grant (IRG), and 2016 Sister Institute Network Fund (SINF). Additional support was provided by the National Institutes of Health through a Cancer Center Support Grant to The University of Texas MD Anderson Cancer Center (CA016672). We appreciated patients and their families who generously donated their tissues to TCGA, as well as the members of TCGA who collected and disclosed valuable data. Andre Kim foundation from Seoul National University Hospital, and Korean Association of Clinical Oncology as well as Dr. Do-Youn Oh, Dr. Seock-Ah Im, and Dr. Yung-Jue Bang from Seoul National University supported Dr. Chan-Young Ock for the training program of M.D. Anderson Cancer Center to study genomic analysis to perform this study. Dr. Se-Hoon Lee from Samsung Medical Center, Dr. Jisu Oh from Cha University, Dr. Choong-kun Lee from Yonsei University, Dr. Chi Young Ok from M.D. Anderson Cancer Center, Dr. Youngil Koh, Dr. Sehhoon Park, and Dr. Jonghanne Park from Seoul National University discussed about the concept of the current study.

Author contributions

J.-S.L. designed and organized the experiment. C.-Y.O. designed and led the integrative analyses. C.-Y.O., J.-E.H., S.-B.K., J.-J.S., H.-J.J., S.P., B.H.S., M.C. and J.-S.L. performed analysis of genomic data. B.K., J.A.A., S.K., K.-W.L., T.M.K. and D.S.H. advised the comprehensive translation of the result. C.-Y.O. wrote the initial draft. B.K., J.A.A., S.K., K.-W.L., T.M.K. and J.-S.L. revised the draft. All authors read and approved the final manuscript.

Additional information

Supplementary Information accompanies this paper at doi:10.1038/s41467-017-01018-0.

Competing interests: The authors declare no competing financial interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017