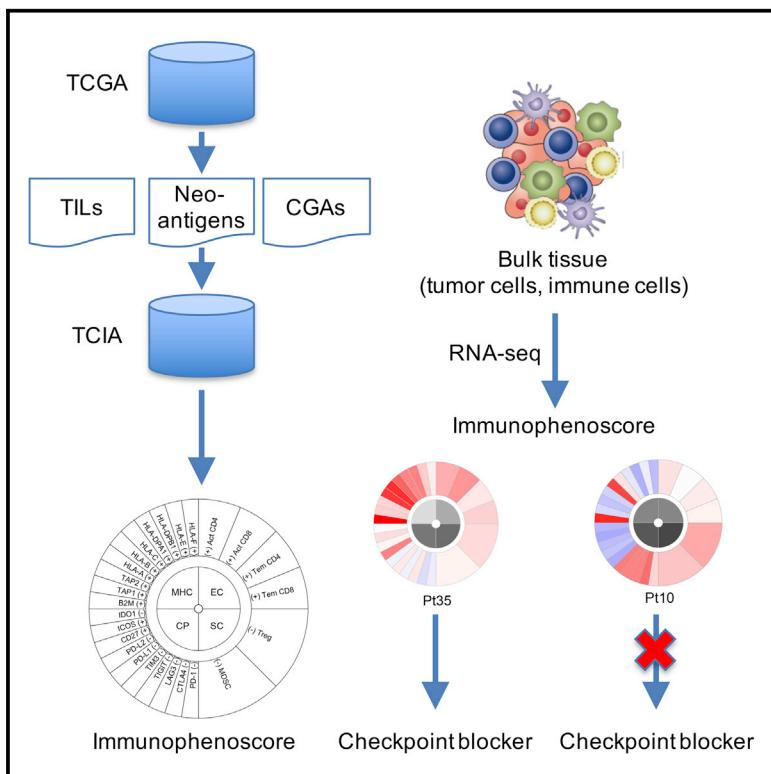


Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade

Graphical Abstract



Authors

Pornpimol Charoentong, Francesca Finotello, Mihaela Angelova, ..., Dietmar Rieder, Hubert Hackl, Zlatko Trajanoski

Correspondence

zlatko.trajanoski@i-med.ac.at

In Brief

Analyses of publicly available cancer genomics data with respect to immunologically relevant parameters can provide important insights but are challenging. Charoentong et al. analyzed data for 20 solid cancers and developed TCIA, a web-accessible resource that allows researchers to mine the data for immunological insights.

Highlights

- Web-accessible resource provides a view of the tumor-immune interface in 20 solid cancers
- The database includes charted antigenomes and cellular compositions of the immune infiltrates
- Cancer genotypes determine tumor immunophenotypes and tumor escape mechanisms
- Immunophenoscore is a predictor of response to checkpoint blockade in melanoma

Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade

Pornpimol Charoentong,^{1,2} Francesca Finotello,^{1,2} Mihaela Angelova,^{1,2} Clemens Mayer,¹ Mirjana Efremova,¹ Dietmar Rieder,¹ Hubert Hackl,¹ and Zlatko Trajanoski^{1,3,*}

¹Biocenter, Division of Bioinformatics, Medical University of Innsbruck, 6020 Innsbruck, Austria

²Co-first author

³Lead Contact

*Correspondence: zlatko.trajanoski@i-med.ac.at

<http://dx.doi.org/10.1016/j.celrep.2016.12.019>

SUMMARY

The Cancer Genome Atlas revealed the genomic landscapes of human cancers. In parallel, immunotherapy is transforming the treatment of advanced cancers. Unfortunately, the majority of patients do not respond to immunotherapy, making the identification of predictive markers and the mechanisms of resistance an area of intense research. To increase our understanding of tumor-immune cell interactions, we characterized the intratumoral immune landscapes and the cancer antigenomes from 20 solid cancers and created The Cancer Immunome Atlas (<https://tcia.at/>). Cellular characterization of the immune infiltrates showed that tumor genotypes determine immunophenotypes and tumor escape mechanisms. Using machine learning, we identified determinants of tumor immunogenicity and developed a scoring scheme for the quantification termed immunophenoscore. The immunophenoscore was a superior predictor of response to anti-cytotoxic T lymphocyte antigen-4 (CTLA-4) and anti-programmed cell death protein 1 (anti-PD-1) antibodies in two independent validation cohorts. Our findings and this resource may help inform cancer immunotherapy and facilitate the development of precision immuno-oncology.

INTRODUCTION

In the past decade, driven by technological advances and novel mechanistic insights, we are witnessing two major advances in cancer research and treatment. First, the emergence of next-generation sequencing (NGS) technologies and comprehensive and coordinated efforts (e.g., The Cancer Genome Atlas [TCGA]) enabled the characterization of multi-dimensional maps of genomic changes in common cancers. And second, basic research in cancer immunology paved the way for the development and approval of checkpoint blockers. These drugs,

which augment T cell activity by blocking cytotoxic T lymphocyte antigen-4 (CTLA-4), programmed cell death protein 1 (PD-1), or PD-1 ligand (PD-L1), show remarkable clinical effects. Analysis of long-term data of patients who received anti-CTLA-4 antibodies in unresectable or metastatic melanoma shows a plateau in the survival curve after 3 years (Schadendorf et al., 2015), suggesting curative potential. Over and above, efficacy of anti-PD-1 antibodies has been shown not only in melanoma, but also in nine different tumor types (Wolchok, 2015). There is currently a rapid pace of development of checkpoint blockers evident from more than 150 clinical trials with monotherapies or combination therapies (Wolchok, 2015). However, there is disparity in response rates across and within tumor types, suggesting the existence of intrinsic immune resistance, as well as evidence for acquired immune resistance (Pitt et al., 2016).

With the development of the immunotherapies with checkpoint blockers as well as other immunotherapeutic strategies including therapeutic vaccines and engineered T cells (Schumacher and Schreiber, 2015), the tumor-immune cell interaction came into focus. The investigation of tumor-immune cell interaction poses considerable challenges, because of the evolving nature of these two ecosystems: the development of cancer, which can be seen as evolutionary process, and the immune system, with a number of innate and adaptive immune cell subpopulations, some of which show phenotypic plasticity and possess memory. Using genomic data and bioinformatics tools, it is now possible to computationally dissect tumor-immune cell interactions (Hackl et al., 2016). These immunogenomic analyses can provide information on the two crucial characteristics of the tumor microenvironment: (1) composition and functional orientation of the infiltrated immune cells and (2) expression of the cancer genome, i.e., the repertoire of tumor antigens including two classes of antigens, neoantigens that arise from somatic mutations and cancer-germline antigens (CGAs).

Previous studies have used genomic data from the TCGA to characterize neoantigens and their association with survival (Brown et al., 2014) or with cytolytic activity estimated using the expression of two genes (Rooney et al., 2015). However, these studies did not characterize the cellular composition of the intratumoral immune infiltrates. More recently, gene signature was used to analyze infiltration of B cells, T cells, and macrophages and the prognostic relevance of these subpopulations

(Iglesia et al., 2016). Similarly, RNA expression data corrected for tumor purity were used to estimate infiltration of B cells, CD4⁺ T cells, CD8⁺ T cells, neutrophils, macrophages, and dendritic cells (Li et al., 2016). However, although such analyses of few major cell types are helpful for identifying clinical associations, higher resolution of the TIL landscape is required in order to dissect tumor-immune cell interactions and identify prognostic and predictive markers. We have previously shown that the tumor infiltrates are composed of at least 28 different types (Bindea et al., 2013), some of which are generally associated with favorable prognosis, whereas others like regulatory T cells (Tregs) are immunosuppressive. Over and above, we described the importance of memory and cytotoxic cells in controlling the growth and recurrence of tumors in colorectal cancer (CRC) (Galon et al., 2006) and showed that a prognostic marker termed immunoscore based on the quantification of these two subpopulations more accurately predicts survival than the standard TNM staging (Mlecnik et al., 2011). Thus, it is of utmost importance to provide a comprehensive view of the intratumoral immune landscape including memory cells, cytotoxic cells (CD8⁺ T cells, natural killer [NK] cells, and NK T [NKT] cells), as well as immunosuppressive cells (Tregs and myeloid-derived suppressor cells [MDSCs]).

We have recently developed an analytical strategy to characterize the cellular composition of the immune infiltrates and examined colorectal cancer (CRC) data sets from the TCGA (Angelova et al., 2015). Our approach is based on the use of metagenes, i.e., non-overlapping sets of genes that are representative for specific immune cell subpopulations and are neither expressed in CRC cell lines nor in normal tissue. The expression of these sets of metagenes is then used to analyze statistical enrichment using gene set enrichment analysis (GSEA). The advantage of the metagene approach is the robustness of the method due to two characteristics: (1) the use of a set of genes instead of single genes that represent one immune subpopulation, because the use of single genes as markers for immune subpopulations can be misleading as many genes are expressed in different cell types; and (2) the assessment of relative expression changes of a set of genes in relation to the expression of all other genes in a sample. Thus, the calculations are less sensitive to noise resulting from sample impurity or sample preparation compared with the deconvolution methods.

Here, we further developed our metagene approach by defining a set of pan-cancer metagenes for 28 immune cell subpopulations and scaled up the analyses to solid cancers. We carried out immunogenomic characterization of the TCGA data for 20 solid cancers with >8,000 tumor samples, and provide a comprehensive view of the cellular composition of the intratumoral immune infiltrates. Additionally, we derived also cancer antigens as well as genetic characteristics (tumor heterogeneity and clonality) for individual samples in order to enable integrative analyses of both immune features as well as genetic features of the tumors. We then developed a web-accessible database TCIA (The Cancer Immunome Atlas) with the results of our analyses (<https://tcia.at/>). To demonstrate the utility of the resource, we carried out integrative analyses and revealed cellular profiles that were predictors of survival for distinct cancers and geno-

type-immunophenotype relationships. We further demonstrate the value of the TCIA by using a machine learning approach to identify determinants of immunogenicity and propose a scoring scheme for solid cancers: the immunophenoscore. Validation with two cohorts treated with anti-CTLA-4 and anti-PD-1 antibodies showed that the immunophenoscore is a superior predictor of response to checkpoint blockers in patients with melanoma.

RESULTS

High-Resolution Genomic Analyses of the Tumor-Immune Interface

An overview of the analytical strategy is shown in Figure 1A (for details, see *Supplemental Information*). We mined the TCGA data to extract the following information: (1) TILs, which were estimated from gene expression data using two approaches, GSEA and deconvolution; (2) cancer antigenomes comprising neoantigens and CGAs; (3) tumor heterogeneity; and (4) immunophenotypes, which were defined using TILs and predetermined sets of genes (MHC molecules, immunostimulators, and immunoinhibitors).

To estimate TILs, we first built a compendium of genes related to specific immune cells using gene expression profiles from 37 studies comprising 366 microarrays (Figure 1B). A subset of these genes, which are representative of specific immune cell subpopulations and are neither expressed in cancer cell lines nor in normal tissue, were then selected based on correlation analysis (782 in total). We then used gene set enrichment analysis (GSEA) (Subramanian et al., 2005) with this gene set to decompose cellular profiles from RNA sequencing data from bulk tissue for individual samples. Additionally, we used an independent method to identify fractions of immune subpopulations based on a deconvolution approach (CIBERSORT; Newman et al., 2015). We built a model to transform RNA-sequencing data to microarray data considering the TCGA samples for which microarray and RNA-sequencing data were available (n = 550; see *Experimental Procedures*). We present here only the results from the GSEA method, which depicts a more comprehensive picture of the tumor-suppressive or tumor-promoting roles of TILs. However, we make both GSEA and deconvolution data available on the TCIA website.

It has been previously shown that the tumor purity can be used in selecting genes informative for deconvolving TILs (Li et al., 2016). We therefore corrected the mutational load for tumor purity and analyzed associations of individual TIL subpopulations estimated using the GSEA and tumor purity (Figure S1). As can be seen, there is a low to weak correlation suggesting that the GSEA method is robust to varying degrees of tumor purity, mainly because GSEA measures the overexpression level of a particular gene by comparing the ranks of the genes in the gene list with those of all other genes.

To chart the antigenome for each sample, we used RNA-sequencing data to derive expression levels of CGAs. For the identification of neoantigens, we assembled an analytical pipeline as previously described (Angelova et al., 2015). The predicted HLA alleles and the mutated expressed peptides for each sample were used as input for the algorithm NetMHCpan

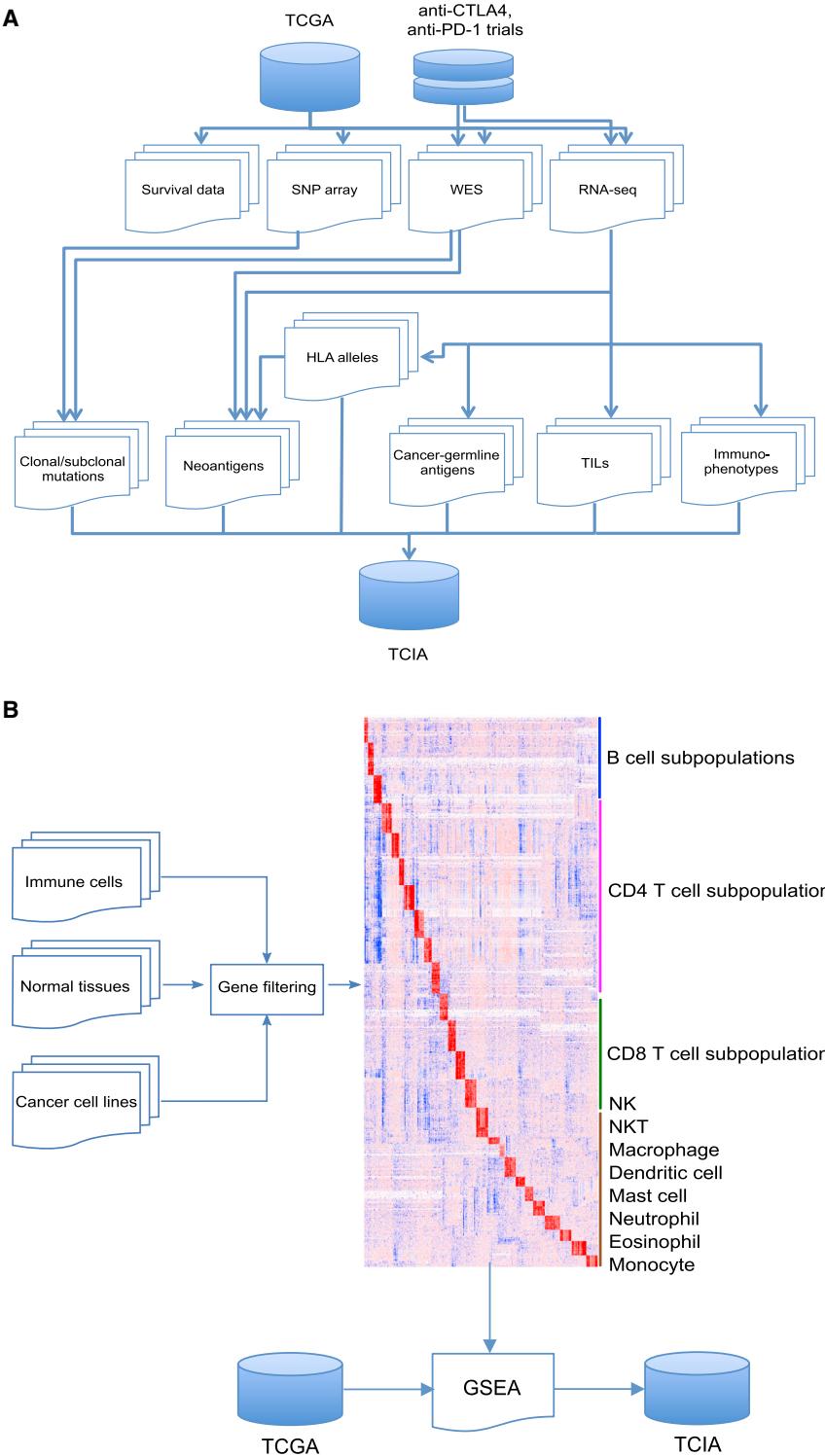


Figure 1. Strategy for Pan-cancer Immuno-genomic Analyses

(A) The scheme shows the immunogenomic analyses and the types of data used for the analyses. The results are deposited in a web-accessible database, The Cancer Immunome Atlas (TCIA) (<https://tcia.at/>).

(B) Immune-related signatures are derived from expression profiles of purified immune cells, normal cells, and cancer cell lines, and used for the gene set enrichment analysis (GSEA) of the TCGA RNA-sequencing data.

be present that do not elicit T cell responses against a given neoantigen, and these cells may have selective advantage and outgrow other clones. We therefore estimated also tumor heterogeneity using exome sequencing data and SNP arrays data to calculate cancer cell fractions (CCF). Finally, driver genes were assigned from a recently published study (Rubio-Perez et al., 2015), whereas clonal and subclonal origins of the neoantigens were estimated as previously described (Landau et al., 2013).

Using this strategy and the analytical pipelines we developed, we analyzed TCGA data for 8,243 samples and 20 solid tumors. Hereafter colon (COAD) and rectal (READ) cancers are considered as a single entity (CRC). Additionally, available exome-sequencing and RNA-sequencing data from recent clinical studies using antibodies against CTLA-4 (Van Allen et al., 2015) and PD-1 (Hugo et al., 2016) were analyzed using the same procedures. Analyses of 1 petabyte of primary and processed genomic data resulted in 50 gigabytes of structured immunogenomic data.

We then developed a web-accessible relational database TCIA (<https://tcia.at/>) and deposited the results of the analyses including immune infiltrates calculated with both GSEA and the deconvolution method, expression of predefined immune subsets, CGAs, HLA allele data, neoantigens, tumor heterogeneity, and clinical data. The database allows queries for the above features for user-defined cancers and subgroups of patients, as

well as a free download of the data for further analyses. Due to privacy concerns, access to HLA allele data is provided only to researchers who have access to the primary sequence data from the TCGA.

(Nielsen et al., 2007) to estimate their binding affinities and predict neoantigens.

Tumor mutational heterogeneity adds another layer of complexity (Gerlinger et al., 2012): within a tumor, clones may

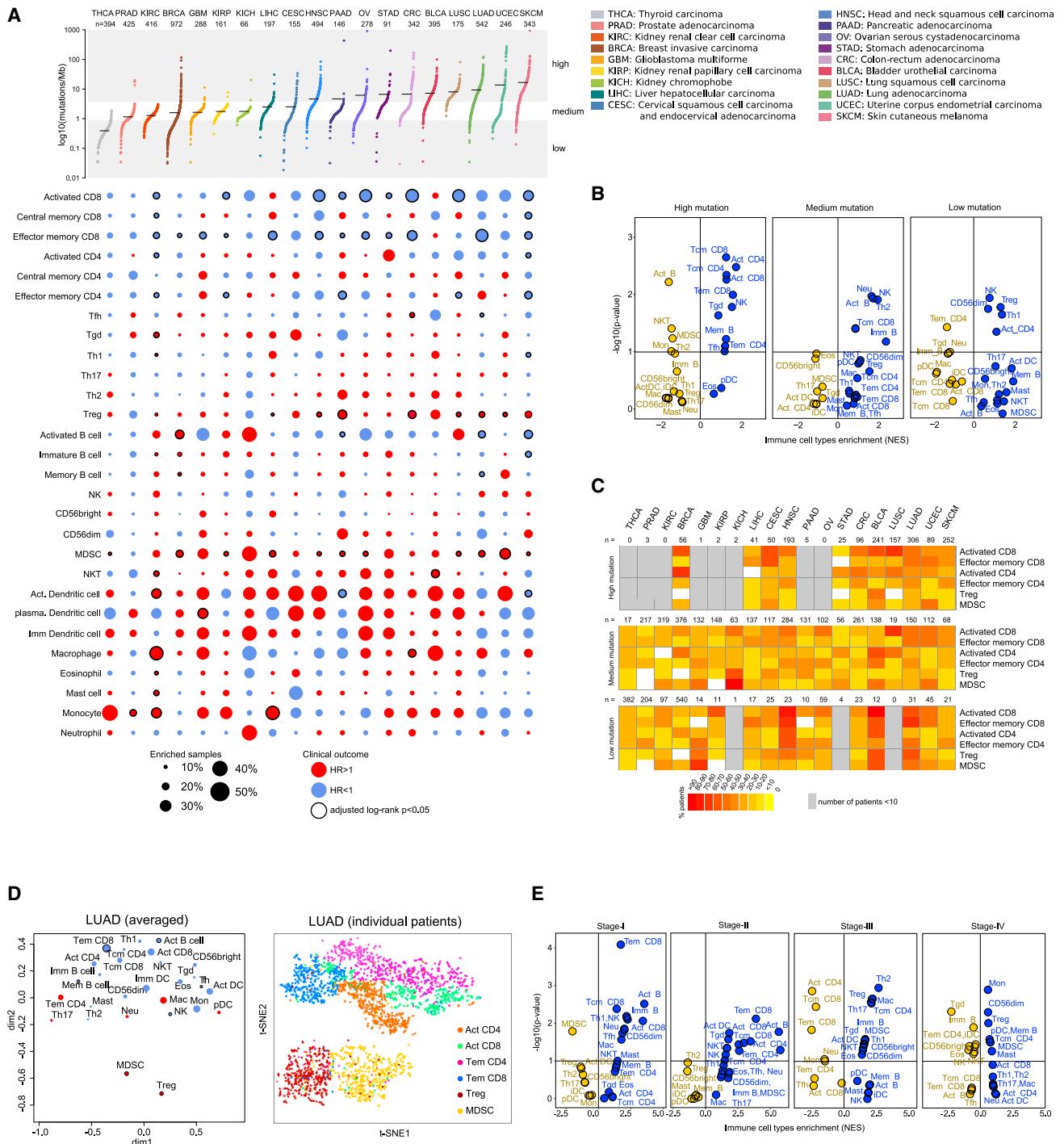


Figure 2. Cellular Characterization of Immune Infiltrates in Solid Cancers

(A) Immune subpopulations across 19 solid cancers. Cancers are sorted according to the mutational load and the immune cell subpopulations according to adaptive and innate immunity, respectively. Lower panel shows results from ssGSEA as bubble plot, where the size of the circles gives the percentage of patients with NES > 0 and q value (FDR) < 0.1, and the color indicates the good (blue; HR < 1) or bad (red; HR > 1) outcome (overall survival). The border indicates adjusted log rank p < 0.05.

(B) Volcano plots for the enrichment (blue) and depletion (yellow) of immune cell types across cancers for tumors with high, intermediate, and low mutational load calculated based on the NES score from the GSEA.

(C) Fraction of samples in which selected immune subpopulations were enriched in all cancers with high, intermediate, and low mutational load.

(legend continued on next page)

Cellular Characterization of Immune Infiltrates Reveals Prognostic Cell Types in Distinct Cancers

Using our GSEA strategy, we estimated 28 subpopulations of TILs including major types related to adaptive immunity: activated T cells, central memory (Tcm), effector memory (Tem) CD4⁺ and CD8⁺ T cells, gamma delta T (T $\gamma\delta$) cells, T helper 1 (Th1) cells, Th2 cells, Th17 cells, regulatory T cells (Treg), follicular helper T cells (Tfh), activated, immature, and memory B cells, as well as cell types related to innate immunity, such as macrophages, monocytes, mast cells, eosinophils, neutrophils, activated, plasmacytoid, and immature dendritic cells (DCs), NK cells, natural killer T (NKT) cells, and MDSCs.

The results of the cellular characterization of the immune infiltrates using GSEA showed heterogeneity across cancers and within individual cancer entities and associations of specific TIL subpopulations with survival (Figure 2A). The cellular profiles associated with survival differed between cancers. In general, the infiltration of many TIL subpopulations related to adaptive immunity was associated with good prognosis including activated CD8⁺ T cells, Tem and Tcm CD8⁺ cells, and Tem CD4⁺ cells, whereas MDSCs and Tregs were associated with bad prognosis. The most common cell type significantly associated with good prognosis was Tem CD8⁺ (15 of 19 cancers) followed by activated CD8⁺ cells (10 of 19 cancers) and Tem CD4⁺ (9 of 19 cancers). Cell types significantly associated with bad prognosis were MDSCs (13 of 19 cancers) and Tregs (10 of 19 cancers). The cellular profiles and the clinical information for specific cancers can be viewed at <http://tcia.at> and downloaded for further analyses.

Mutational Load and Tissue Context Determine Cellular Composition of Immune Infiltrates

The vision of precision oncology implies patient-specific genomic profiling and testing of targeted therapeutics on the subset of patients carrying the relevant genetic lesions, rather than on patients selected based on histologic tumor subtypes (Garraway and Lander, 2013). We asked what the consequences of this genome-centric view on the immune landscapes are and analyzed the impact of the mutational load. The tumors were classified into three groups: tumors with high mutational load (upper quartile), intermediate mutational load (two intermediate quartiles), and low mutational load (lower quartile) (Figure 2A). Tumors with high mutational load were enriched with activated T cells and Tem cells, and were depleted with immunosuppressive Tregs and MDSCs, whereas tumors with low mutational load showed opposite enrichments and depletions (Figure 2B). However, within each mutational load group, the composition of the TILs was divergent for specific cancers as shown in Figure 2C for selected cell types. The tissue context was also observable using correspondence analysis for all immune subpopulations and all cancers (Figure S2). Similar trends were observed also within tumor types (Figure S3).

We then analyzed individual cancer entities and the TILs using dimensional reduction technique and frequently observed separation of the subpopulation related to immune suppression (MDSCs, Tregs) from the subpopulations related to the effector function (activated T cells, Tcm cells, Tem CD4⁺ and CD8⁺ cells). An example of this is shown in Figure 2D for lung adenoma (LUAD) for all cell types in the LUAD cohort, as well as for individual tumors for selected subpopulations.

The progression of the tumor across cancers was also characterized by distinct immune cell patterns. For example, Tem CD8⁺ cells were enriched in stage I and stage II tumors and depleted in stage III and stage IV tumors (Figure 2E). In contrast, Tregs and MDSCs were depleted in early-stage tumors and enriched in late-stage tumors. In general, the enrichment of the TIL subpopulations related to adaptive immunity was decreasing from stage I to stage IV, whereas enrichment of TILs related to innate immunity was increasing from stage I to stage IV. Hence the cellular composition of the immune infiltrates across solid cancers during progression is shifting toward immunosuppressive phenotype.

In summary, these analyses showed that both the genomic profiles and the specific tissue context contribute to the cellular composition of the immune infiltrates. Furthermore, the results support the notion of the evolving nature of the immune landscape during tumor progression.

CGAs Are Associated with Infiltration of CD4⁺ and CD8⁺ T Cells

Tumor antigens that have the potential to elicit immune responses that are strictly tumor specific are CGAs and neoantigens. CGAs are proteins that are normally expressed by germline cells but have aberrant expression in tumor cells, whereas neoantigens result from a mutation or rearrangement of a gene-coding sequence. For virus-associated tumors, epitopes derived from viral open reading frames also contribute to the pool of neoantigens. We therefore characterized these two antigens classes in 20 solid cancers.

The expression of all CGAs (Table S2) in all cancers can be retrieved from <http://tcia.at>. Because many of these CGAs have low tumoral specificity, we selected 60 CGAs that were previously reported to be transcriptionally silent in normal, non-germline tissues (Rooney et al., 2015) and analyzed the expression and the putative antigens that might be targeted by T cells. In a previous study there was no clear positive association between cytolytic activity and the number of expressed CGAs (Rooney et al., 2015). Here we identified a number of CGAs that were significantly correlated with activated, Tem, and/or Tcm CD8⁺/CD4⁺ cells for each solid cancer (Figure 3A).

Because these CGAs are candidates for vaccination, we also predicted the peptides from these CGAs that bind to HLA molecules, leading to 5,775 unique peptides. We then estimated the antigen load by calculating a CGA expression score

(D) Visualization of the immune infiltrates (averaged normalized enrichment score [NES]) in lung adenocarcinoma (LUAD) using two-dimensional coordinates from multidimensional scaling (MDS) (left panel) and for individual patients and selected cell types based on two-dimensional coordinates from t-distributed stochastic neighbor embedding (t-SNE) (right panel).

(E) Volcano plots for the enrichment (blue) and depletion (yellow) of immune cell types across cancers for tumor stage I to IV calculated based on the NES score from the GSEA.

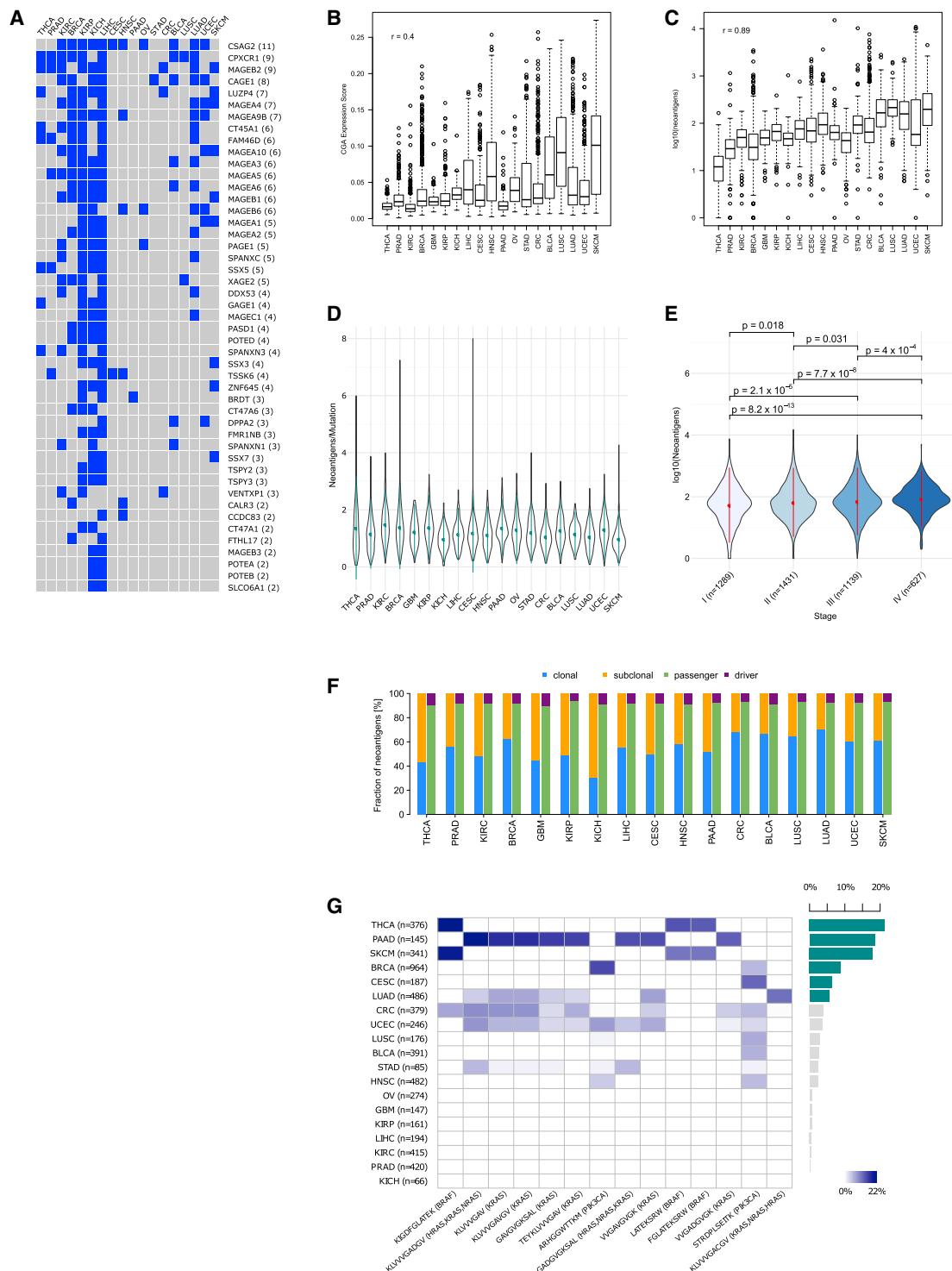


Figure 3. Antigenomes in Solid Cancers

(A) Cancer-germline antigens (CGAs) that are associated with CD8⁺ and/or CD4⁺ T cells. Blue squares show significant association with the corresponding cancer type (Spearman rank correlation > 0.3; adjusted p < 0.1). CGAs are sorted according to the number of cancers with significant associations.

(B) CGA expression score across cancers. Cancers are sorted according to the mutational load.

(C) Neoantigen load across cancers.

(D) Neoantigen frequencies for solid cancers.

(legend continued on next page)

([Experimental Procedures](#)). Interestingly, there was positive correlation of the CGA score with mutational load across cancers ($r = 0.4$, $p < 2.2e-16$) ([Figure 3B](#)), suggesting additional relationships between methylation phenotypes and mutational load as observed in CRC ([Angelova et al., 2015](#)).

These results suggest a tissue-dependent role of CGAs in spontaneous immunity and reinvigorate the use of CGA-based therapeutic vaccines, perhaps in combination with epigenetic drugs to increase the expression of selected CGAs.

Neoantigen Landscape Is Diverse and Sparse

Because neoantigens are attractive candidates for developing therapeutic cancer vaccines, we analyzed the neoantigens landscape in order to identify potential candidates for off-the-shelf vaccination. The neoantigen landscape in solid tumors was composed of 933,954 expressed neoantigens (911,548 unique) originating from 893,960 somatic point mutations. As expected, the number of neoantigens correlated with the mutational load ([Figure 3C](#)). The median neoantigen frequency, i.e., the number of neoantigens per mutation, varied across cancers and ranged from 0.93 (skin cutaneous melanoma [SKCM]) to 1.43 (kidney renal cell carcinoma [KIRC]) ([Figure 3D](#)). It should be noted that these differences across cancers might be attributed to the differences in mutational processes, which are more likely to lead to nonsynonymous mutations (e.g., C > A transversions compared with C > T transitions; [Lawrence et al., 2013](#)). The associations of the neoantigen load as well as of the CGA load and selected genes and immune cell subpopulations were highly variable between different cancer entities, suggesting that the immune response is likely governed by combined effects of both antigen classes.

Somewhat unexpectedly, there was only a slight increase of the neoantigen load from stage I to stage IV across all cancers (all adjusted pairwise $p < 0.032$, two-sided Dunn's post hoc tests on ranked sums; [Figure 3E](#)). Furthermore, as the cellular composition of the immune infiltrates during progression is shifting toward immunosuppressive phenotype, this raises an issue of whether immunotherapies that depend on the adaptive immune response can be effective in later stages.

The fraction of neoantigens derived from driver genes for all solid tumors was 7.6% (ranging between 7.0% for lung squamous cell carcinoma [LUSC] and 10.6% for glioblastoma [GBM]) ([Figure 3F](#)). Hence the bulk of neoantigens had its origin in passenger genes. On average, 56% of the neoantigens were of clonal origin with varying proportions of clonal to subclonal fractions across cancers ([Figure 3F](#)). The fractions of neoantigens with clonal origin ranged from 31% (kidney chromophobe [KICH]) to 71% (LUAD) ([Table S3](#)).

The neoantigens were infrequently shared between patients ([Figure 3G](#); see also [Figure S4](#) for HLA-stratified shared neoantigens). From the total of 911,548 unique neoantigens, only 24 were shared in at least 5% of patients in one or more cancer types. These shared neoantigens represent identical peptides

originating from one or more genes. As expected, the most frequent neoantigens were induced by mutations in driver genes like BRAF, RAS, and PIK3CA. Among these, only two peptides were shared in more than 15% of patients in one cancer type: KIGDFGLATEK was shared in thyroid carcinoma (THCA) and SKCM, and KLVVVGADGV was shared in PAAD. The neoantigen KIGDFGLATEK originates from BRAF^{V600E} mutation, which is present in a large fraction of THCA ([Cancer Genome Atlas Research Network, 2014](#)) and SKCM ([Cancer Genome Atlas, 2015](#)) tumors. The neoantigen KLVVVGADGV originates from the p.G12D mutation of KRAS, which is shared across a large fraction of PAAD patients ([Witkiewicz et al., 2015](#)).

Thus, the antigenome landscape was diverse between and within cancers, and with respect to the neoantigens highly sparse. Furthermore, the results suggest that T cell responses are not only directed against neoantigens, but also against CGAs.

Genotypes of the Tumors Determine Immunophenotypes and Tumor Escape Mechanisms

The immunogenomic analyses of the CRC data in our previous study revealed that the immunophenotypes in the hypermutated compared with the non-hypermutated tumors were characterized by increased enrichment of effector T cells ([Angelova et al., 2015](#)), likely to be a consequence of the higher antigen burden. We asked the question how are the immunophenotypes related to other genomic features describing the complexity of the tumor genome, like tumor heterogeneity (high versus low) and antigenicity (high versus low).

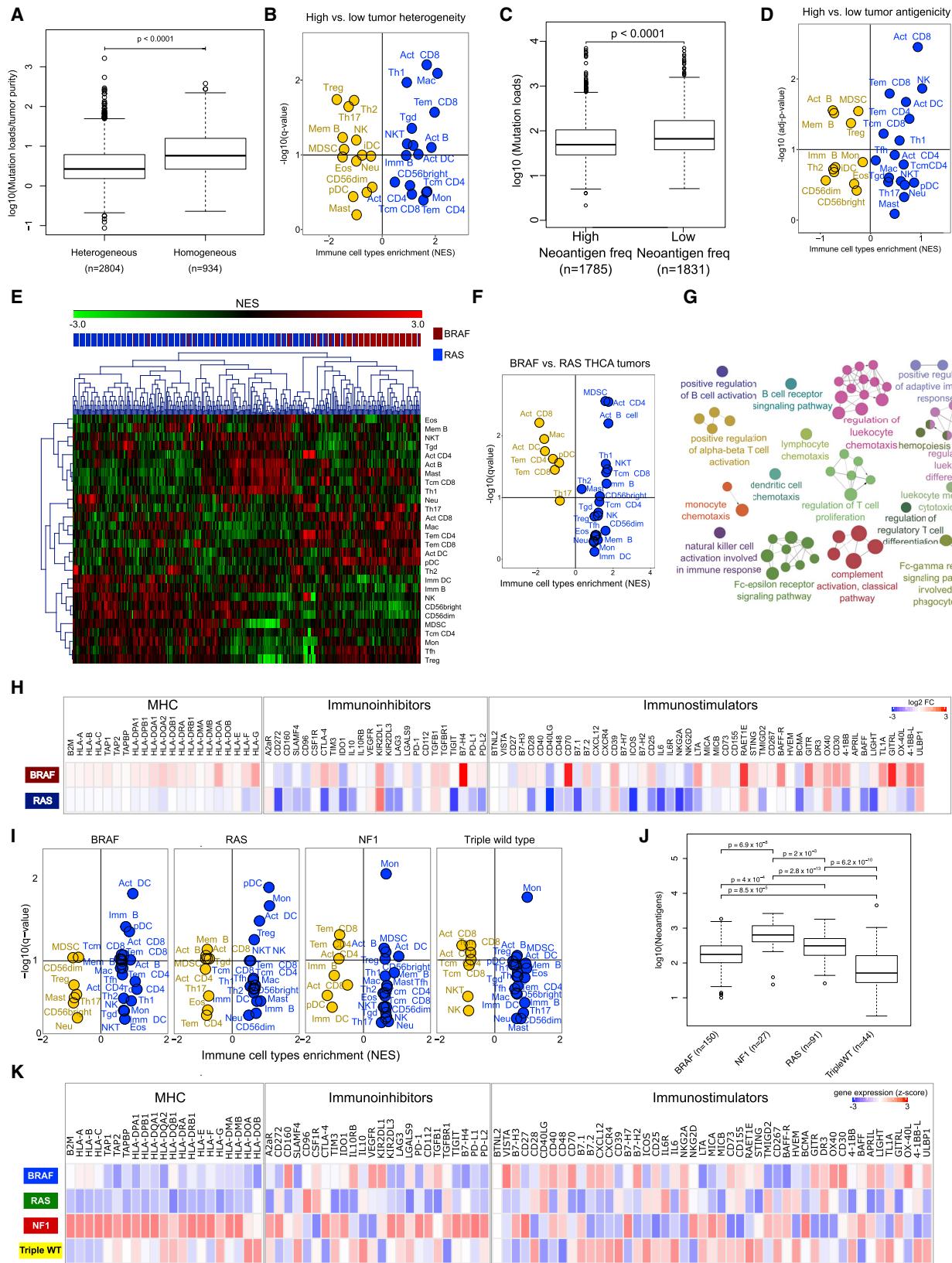
Tumors that were more heterogeneous were enriched with activated T cells and Tem cells, and depleted of immunosuppressive cells, despite slightly lower mutational load ([Figures 4A and 4B](#)). Tumors with higher neoantigen frequencies showed similar cellular patterns of immune infiltrates ([Figures 4C and 4D](#)). We then asked whether there is a genotype-immunophenotype relationship also for the gene-specific view of the genomic landscape. We selected two cancers with the lowest and highest mutational load, i.e., THCA and SKCM, and analyzed the immunophenotypes for their distinct genotypes: BRAF and RAS subtypes for THCA, and BRAF, RAS, NF1, and triple-negative wild-type for SKCM. Additionally, we examined the expression of three classes of molecules that are involved in tumor escape mechanisms: MHC molecules (class I, class II, and non-classical), which may be downregulated to avoid recognition by T cells, immunostimulators (e.g., OX40), which may be downregulated to avoid immune destruction, and immunoinhibitory genes (e.g., CTLA-4), which may be upregulated to enable tumor escape.

The different THCA genotypes were associated with specific immunophenotypes ([Figure 4E](#)) and showed distinct cellular patterns of immune infiltrates ([Figure 4F](#)) in spite of comparable neoantigen burden ($p > 0.05$, two-sided Wilcoxon rank sum test). Analysis of the differentially expressed genes with respect to

(E) Neoantigen load for different tumor stages (Kruskal-Wallis test followed by two-sided Dunn's pairwise post hoc tests on rank sums with Benjamini-Hochberg adjustment of p values).

(F) Fractions of neoantigens and their origin.

(G) Shared neoantigens in solid tumors. Shown are only neoantigens shared in at least 5% of the tumors.



(legend on next page)

the immune-related Gene Ontology (GO) terms highlighted the pathways, which might explain the different effects on the immune system including chemotaxis, T cell differentiation, T cell proliferation, T cell activation, or B cell receptor signaling pathway (Figure 4G). The expression levels of MHC, immunostimulatory, and immunoinhibitory molecules were also associated with the genotypes (Figure 4H). These results suggest that BRAF-mutated THCA tumors employ different tumor escape mechanisms compared with RAS-mutated tumors in THCA: BRAF tumors were infiltrated with immunosuppressive cells, whereas RAS tumors downregulated MHC molecules as well as immunomodulatory molecules.

Similarly, the SKCM genotypes were also associated with distinct immunophenotypes (Figure 4I), with BRAF tumors enriched with effector T cells, whereas other genotypes were enriched with immunosuppressive cells. It is noteworthy that for the SKCM cohort the neoantigen burden differed (all adjusted pairwise $p \leq 0.002$, two-sided Dunn's post hoc tests on ranked sums; Figure 4J), and therefore the relative contribution of the mutational load can have an additional impact. The four SKCM genotypes were also associated with varying expression levels of MHC and immunomodulatory molecules (Figure 4K).

The results of this study suggest that the genotypes of the tumor determine the immunophenotypes and the tumor escape mechanisms. This was evident at the high-level view of the genomic landscape (e.g., mutational load, tumor heterogeneity) as well as at the gene-specific view (e.g., BRAF- or RAS-mutated tumors).

Machine Learning Identifies Major Determinants of Tumor Immunogenicity in Solid Cancers

The results of this work showed not only highly heterogeneous TILs, but also varying ratios of different T cell subsets including suppressive ones. These observations raise the question of the underlying molecular mechanisms that explain the differences in immunogenicity of the tumors. The question can be reduced to the notion of sources of immunogenic differences, which can be divided into two categories: tumor-intrinsic factors and tumor-extrinsic factors. Tumor-intrinsic factors include the mutational load, the neoantigen load, the neoantigen frequency, the expression of immunoinhibitors and immunostimulators (e.g., PD-L1), and HLA class I molecule alterations. Tumor-extrinsic factors include chemokines that regulate T cell traf-

ficking, infiltration of effector TILs and immunosuppressive TILs, and soluble immunomodulatory factors (cytokines) (Gajewski et al., 2006).

In a previous study, an association of the cytolytic activity and the neoantigen load was reported (Rooney et al., 2015). Here, we asked the question what are the determinants of the tumor immunogenicity, and considered all tumor-intrinsic and tumor-extrinsic factors mentioned earlier. The number of single parameters that determine the immunogenicity of the tumors and the various combinations thereof render it difficult to identify a set of representative parameters even with large cohorts like the TCGA ones. We therefore employed machine learning techniques and exploited the large-scale genomics data. We reasoned that the immunogenicity of the tumor can be represented by the cytolytic activity (estimated using the expression of granzyme A (*GZMA*) and perforin (*PRF1*) according to Rooney et al., 2015), because this is the ultimate effector mechanism in the cancer immunity cycle. For each cancer type, we used a random forest classification approach, which is based on a multitude of decision trees, including 127 parameters (Table S4) to separate tumors with high cytolytic activity from tumors with low cytolytic activity. For individual cancer types the most predictive features were identified using the mean decrease of accuracy over all cross-validated predictions (Figure 5A). For each of the studied cancers, the analysis revealed only immune-related factors, which we classified into four categories: (1) infiltration of activated CD8⁺/CD4⁺ T cells and Tem CD8⁺/CD4⁺ cells; (2) infiltration of immunosuppressive cells (Tregs and MDSCs); (3) expression of MHC class I, class II, and non-classical molecules; and (4) expression of certain co-inhibitory and co-stimulatory molecules (Figure 5A).

To visualize the information, we constructed an immunophenogram that includes these four categories (Figure 5B). We then calculated an aggregated score, immunophenoscore, based on the expression of the representative genes or gene sets comprising four categories: MHC molecules, immunomodulators, effector cells (activated CD8⁺ T cells and CD4⁺ T cells, Tem CD8⁺ and Tem CD4⁺ cells), and suppressor cells (Tregs and MDSCs) (see Experimental Procedures). Multivariate analysis showed that the immunophenoscore was associated with survival in 12 solid cancers, of which 4 were significant: KIRC, SKCM, breast cancer (BRCA), and bladder cancer (BLCA) (Figure 5C).

Figure 4. Genotypes and Immunophenotypes in Solid Cancers

- (A) Mutational load and tumor heterogeneity (two-sided Wilcoxon rank sum test).
- (B) Immune infiltrates in tumors. Shown is a volcano plot for tumors with high and low heterogeneity calculated based on the NES score from the GSEA.
- (C) Mutational load and neoantigen frequency (two-sided Wilcoxon rank sum test).
- (D) Immune infiltrates in tumors. Shown is a volcano plot for tumors with high and low antigenicity calculated based on the NES score from the GSEA.
- (E) Hierarchical clustering of immune cell composition for BRAF- and RAS-mutated THCA tumors.
- (F) Volcano plot for BRAF- and RAS-mutated TCHA tumors calculated based on the NES score from the GSEA.
- (G) Gene ontology (GO) analysis of the differentially expressed genes for BRAF- and RAS-mutated TCHA tumors using ClueGO (Bindea et al., 2009).
- (H) Expression of MHC and immunomodulatory molecules in BRAF- and RAS-mutated TCHA tumors. Expression values were compared with normal tissue (\log_2 -fold changes are color coded according to the legend).
- (I) Volcano plots for SKCM genotypes calculated based on the NES score from the GSEA.
- (J) Mutational load for SKCM genotypes (Kruskal-Wallis test followed by two-sided Dunn's pairwise post hoc tests on rank sums with Benjamini-Hochberg adjustment of p values).
- (K) Expression of MHC and immunomodulatory molecules for SKCM genotypes. Expression values are represented by Z score calculated across all SKCM tumors and color coded according to the legend.

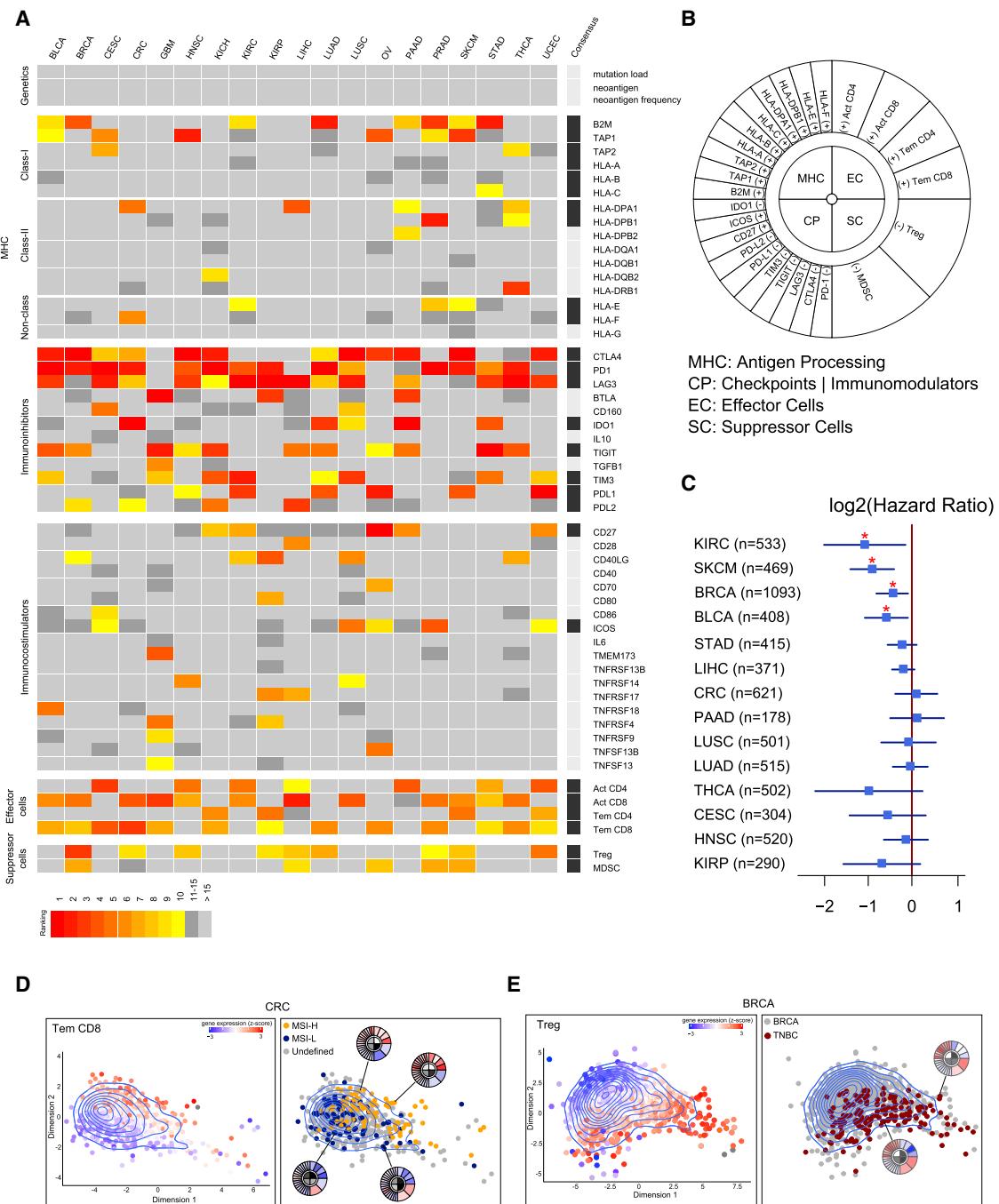


Figure 5. Determinants of Immunogenicity in Solid Cancers

(A) Major parameters determining immunogenicity in solid cancers revealed using random forest approach.

(B) Immunophenogram for the visualization of the parameters determining immunogenicity.

(C) Results of multivariate survival analyses using the immunophenoscore for all solid cancers. Forest plots showing log₂ hazard ratio (95% confidence interval).

*Adjusted p < 0.05.

(D) Visualization based on two-dimensional coordinates from multidimensional scaling (MDS) of expression profiles from the genes and the gene sets as used in the immunophenogram in colorectal cancer (CRC). Mean expression (Z score) of immune genes of Tem CD8⁺ T cells as well as MSI and microsatellite stable (MSS) samples for CRC are indicated by colors according to the legend and the mean expression (Z scores).

(E) Visualization based on two-dimensional coordinates from MDS in breast cancer (BRCA). Tregs and triple-negative breast cancer (TNBC) samples for BRCA are indicated by colors according to the legend. Representative immunophenoscores for selected patients are shown.

The immunophenogram enables the visualization of the immunophenotypes of the tumor also at the levels of individual tumors as seen for CRC (Figure 4D) and BRCA (Figure 4E). As can be seen, subgroups of tumors like microsatellite-unstable (MSI) and triple-negative breast cancer (TNBC) are grouped with respect to distinct TILs like Tem CD8⁺ cells and Tregs, respectively. We also implemented an interactive version of the tool (see <https://tcia.at/>), which enables visualization and scoring of tumor samples using expression data for the markers we identified using the random forest approach.

Hence, using a data-driven approach, a large number of tumors (>8,000), and genomic, transcriptomic, and immunological features, we were able to identify the major determinants of tumor immunogenicity. Based on these results we propose a visualization method, the immunophenogram, and a scoring scheme, the immunophenoscore, for solid tumors.

Immunophenoscore Predicts Response to Immunotherapy with CTLA-4 and PD-1 Blockers

Because only a minority of the patients is responsive to checkpoint blockers, the identification of predictive markers and the mechanisms of resistance to immunotherapy is a subject of intense research. Although several markers have been proposed including TILs, PD-1 or PD-L1 expression, mutational load (Rizvi et al., 2015), or clonal neoantigens (McGranahan et al., 2016), none has yet been fully validated (Spencer et al., 2016).

We reasoned that the determinants of immunogenicity identified using the random forest approach might also have a predictive value, and analyzed two genomic and transcriptomic data sets from patients with melanoma treated with anti-CTLA-4 (Van Allen et al., 2015) and anti-PD-1 antibodies (Hugo et al., 2016). Using RNA-sequencing data and GSEA, we reconstructed the TIL landscape and scored the patients using the immunophenoscore. The immunophenograms of the individual patients treated with anti-CTLA-4 antibodies are shown in Figure 6A. Tumors of the responders were enriched with cytotoxic cells (CD8⁺ T cells, T γ δ cells, NK cells) and depleted of MDSCs and Tregs (Figure 6B). More importantly, the immunophenogram and the score derived from the analyses enabled stratification of patients into responders and non-responders (Figure 6C) with a superior predictive power compared with the expression of checkpoint molecules, as can be seen from the receiver operating characteristics (ROC) curve (Figure 6D).

Similar observation was made also for the patients treated with anti-PD-1 antibody. Visualization of the determinants of the immunogenicity with the immunophenogram for responders and non-responders showed distinct expression patterns in the two groups (Figure 6E). Again, tumors of the responders were enriched with cytotoxic cells (CD8⁺ T cells, T γ δ cells, NK cells) and depleted of MDSCs and Tregs, as evident in the volcano plot (Figure 6F). Finally, the immunophenoscore (Figure 6G) and the ROC curve (Figure 6H) showed the predictive value also for patients treated with anti-PD-1 antibodies.

Hence the immunophenoscore developed from a panel of immune-related genes belonging to the four classes—effector cells, immunosuppressive cells, MHC molecules, and selected immunomodulators—has predictive value in melanoma patients treated with the CTLA-4 and PD-1 blockers.

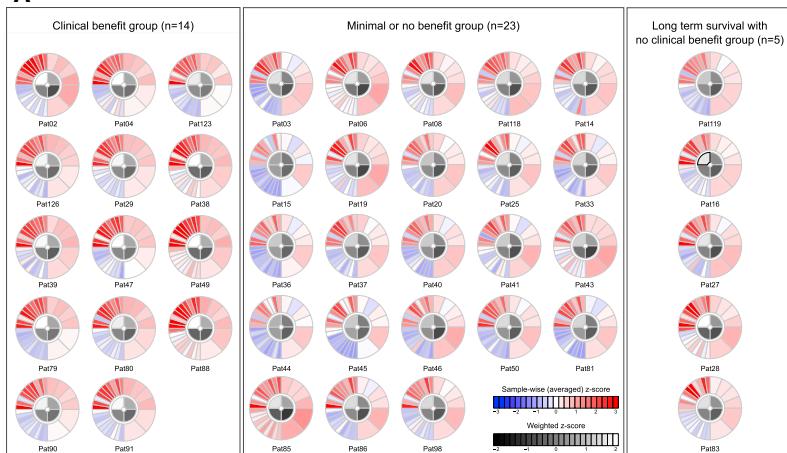
DISCUSSION

We used an analytical strategy to provide comprehensive view of 28 TIL subpopulations including effector and memory T cells and immunosuppressive cells (Tregs, MDSCs). We also developed an interactive big data resource that includes cellular compositions of TILs, neoantigens, CGAs, and tumor heterogeneity, and enables other researchers to generate testable hypotheses. Our approach to deeply mine cancer genomic data sets revealed a number of associations suggesting important biological conclusions with implications for cancer immunotherapy. Moreover, using machine learning approach, we identified major determinants of tumor immunogenicity and proposed an immunophenoscore, which is not only prognostic like the immunoscore for CRC (Mlecnik et al., 2011), but also has superior predictive power for identifying responders for treatment with CTLA-4 and PD-1 antibodies.

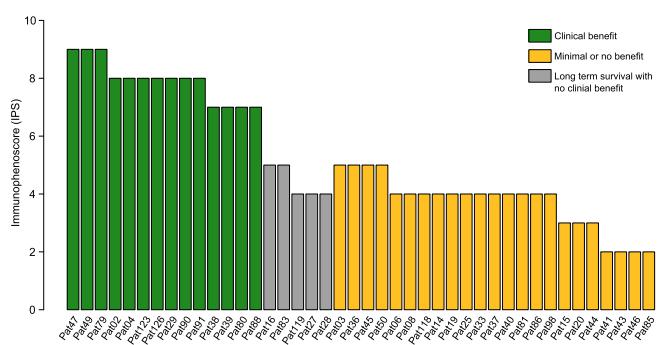
First, our intriguing observation of the association of the genotypes and the immunophenotypes and the distinct tumor escape mechanisms determined by the genotypes implicates that the interaction with the immune system is predetermined by the genetic basis of the tumors. This genotype-immunophenotype relationship was evident at the high-level view of the genomic landscape, i.e., heterogeneous versus homogeneous and antigenic versus less antigenic tumors. Strikingly, this association was observed also at the gene-specific view, i.e., at the level of mutational origin of the tumors like BRAF or RAS, as well as for tumors at both ends of the mutational load: THCA and SKCM. This raises the question of the underlying molecular mechanisms. Currently we can only speculate on the possible mechanisms, which can include modifications of different signaling pathways because of mutational changes. For example, evidence from clinical and experimental studies has demonstrated that signaling by BRAF and RAS oncogenes can present similarities but also differential effects (Oikonomou et al., 2014). The analysis of the genes differentially expressed between the two genotypes in THCA indicates several processes that might be involved in the sculpting of the immune landscape (Figure 4G).

Second, charting of the antigenome showed that both the CGAs and the neoantigens are associated with effector and memory T cells, suggesting that T cell responses are directed against both antigen classes. Moreover, the antigen load of the CGAs and neoantigens was highly variable across and within cancers. Although neoantigens are thought to be of particular relevance to tumor control (because these antigens are not affected by central T cell tolerance) (Schumacher and Schreiber, 2015), the contribution of the CGAs is still unclear. In a previous study the role of the CGAs in anti-tumor immunity was not uncovered (Rooney et al., 2015), likely because of the methods used (e.g., expression signatures of only two genes). In contrast, our comprehensive and high-resolution approach provides evidence for T cell reactivity also against CGAs. The quality of the T cell responses directed against neoantigens and CGAs has not been compared so far, and future experimental studies are needed to identify the contribution of each class of antigens. We envision that in the case of malignancies with a low mutational load that are less likely to produce immunogenic

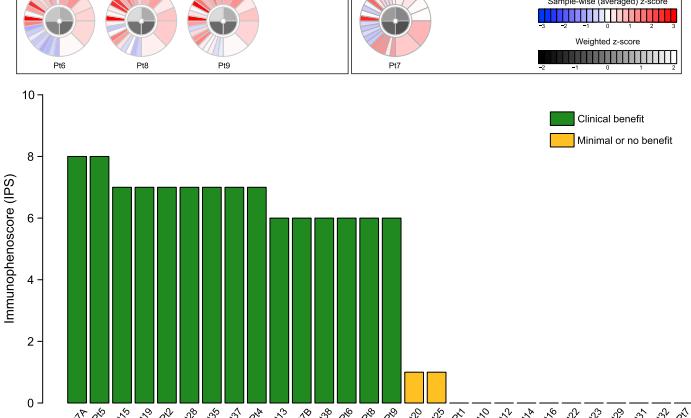
A



C



G



neoantigens, CGAs might be therapeutically exploited by combining DNA hypomethylation agents to induce expression of CGAs and checkpoint blockers to target tumor cells. This type of combination strategy was recently evaluated in a mouse model (Covre et al., 2015).

Beyond these biological insights, the results from this study also have important implications for cancer immunotherapy with checkpoint blockers as monotherapy, as combination therapy with targeted agents, and for therapeutic vaccination. Most importantly, we propose an immunophenoscore, i.e., a panel of immune genes for classification of patients likely to respond to therapy with antibodies targeting CTLA-4 and PD-1 with superior performance. The immunophenoscore we developed was derived in an unbiased manner using the TCGA data and machine learning, but it reflects current understanding of the categories of genes that determine immunogenicity of the tumors: effector cells, immunosuppressive cells, MHC molecules, and immuno-modulators. The immunophenoscore is similar to the conceptual immunogram that was recently proposed to represent the status of the immune system (Blank et al., 2016). Another advantage of the immunophenoscore is that it represents a standardized value because Z scores are used, and is therefore more robust compared with the use of expression values. However, because presently only limited data are available, additional studies are required to validate the immunophenoscore. Notably, the method can be further improved by optimizing the immunophenoscore for specific cancers. Finally, for routine applications, other techniques for gene expression profiling like microarrays and qPCR can be used instead of RNA-sequencing.

Because only a minority of the patients is responsive to checkpoint blockers, there are major efforts to develop therapeutic strategies to overcome resistance by using a combination of checkpoint blockers and targeted agents. Hence precision immuno-oncology, i.e., the amalgamation of immunotherapy and precision oncology, appears as a promising approach to treat patients. Supported by insights from genomic profiling studies and the availability of a large number of drugs and molecular targets, the precision oncology hypothesis was postulated, i.e., that cancer treatment decisions could be guided by molecular profiling irrespective of the tissue of origin. Because the efficacy of checkpoint blockers was shown in several cancers, one could adopt this tissue-agnostic approach also for precision immune oncology. However, the results of our analyses show that the tissue context is an important determinant of the cellular compo-

sition of the immune infiltrates, and hence determines the anti-tumor response. Consequently, stratification of patients in a precision immuno-oncology framework requires genomics-based information and characterization of the immune infiltrates.

With respect to neoantigen-based therapeutic vaccination, the sparsity of the neoantigen space advocates against the development of off-the-shelf vaccines. A notable exception represents vaccine for THCA, SKCM, or PAAD, which could target 20% of the patients. Thus, personalized cancer vaccination strategy is required in which whole-exome NGS is performed to identify somatic mutations, followed by bioinformatics analyses to identify neoantigens, and synthesis of peptide- or DNA/RNA-based vaccines. Viability of such personalized cancer vaccination strategy was recently demonstrated in clinical studies in melanoma patients (Robbins et al., 2013; van Rooij et al., 2013).

Finally, with the large number of ongoing studies with checkpoint blockers either as monotherapy or combination therapy, we expect that the immunogenomic data amount will continuously increase. Thus, we strongly believe that the TCIA in its current form and future incorporation of additional data sets represents an important contribution to the field and will enhance the identification of mechanistic insights of the complex tumor-immune cell interactions.

EXPERIMENTAL PROCEDURES

Identification of TIL Subpopulations, Cancer-Germline Antigens, Neoantigens, Tumor Heterogeneity, and Clonality of Mutations

Genomic and clinical data for 20 solid tumors were downloaded via the TCGA data portal (*Supplemental Experimental Procedures*). We used single sample gene set enrichment analysis (ssGSEA) (Barbie et al., 2009) to identify immune cell types that are over-represented in the tumor microenvironment (*Supplemental Experimental Procedures*). Additionally, a deconvolution approach was applied using the tool CIBERSORT (Newman et al., 2015) and a custom model to modify RNA-sequencing data from TCGA to be used as input of the deconvolution algorithm (*Supplemental Experimental Procedures*). Cancer-germline antigens, neoantigens, tumor heterogeneity, and clonality of mutations were characterized as previously described (Angelova et al., 2015) (see also *Supplemental Experimental Procedures*).

Identification of Determinants of Tumor Immunogenicity, Immunophenogram, and Immunophenoscore

For each patient the cytolytic activity was calculated as the mean of the GZMA and PRF1 expression levels [$\log_2 (\text{TPM}+1)$] as previously defined (Rooney et al., 2015). For each cancer type, patients were divided into two groups based on median cytolytic activity. A random forest classifier (Breiman, 2001) separating the group of patients with TILs exhibiting higher cytolytic

Figure 6. Immunophenoscores and Response to Checkpoint Blockade

- (A) IPS and response to blockade with anti-CTLA-4 antibody (data from Van Allen et al., 2015). Shown are immunophenograms for individual patients. HLAs were highly upregulated compared with the mean expression within the tumor ($z > 0$; showed in red within the upper left quadrant termed MHC), whereas many checkpoints were downregulated compared with the mean expression within the tumor ($z < 0$; showed in blue in the lower left quadrant termed CP) in many patients of this study.
- (B) Volcano plot for the enrichment and depletion of immune subsets in the tumor calculated based on the NES score from the GSEA.
- (C) IPSs for the cohort.
- (D) Receiver operating characteristics for IPS for the cytolytic activity (mean expression of GZMA and PRF1 as suggested by Rooney et al., 2015), mutational load, for the expression of PD-1, PD-L1, and CTLA4. AUCs with 95% bootstrap confidence interval are provided.
- (E) IPS and response to blockade with anti-PD-1 antibody (data from Hugo et al., 2016). Shown are immunophenograms for individual patients.
- (F) Volcano plot for the enrichment and depletion of immune subsets in the tumor calculated based on the NES score from the GSEA.
- (G) IPSs for the cohort.
- (H) Receiver operating characteristics for IPS, for the cytolytic activity (mean expression of GZMA and PRF1 as suggested by Rooney et al., 2015), mutational load, for the expression of PD-1, PD-L1, and CTLA4. AUCs with 95% bootstrap confidence interval are provided.

activity from the group of patients with TILs exhibiting lower cytolytic activity was trained using the R package *randomForest* with 10,000 trees and included mutational load per megabase, number of neoantigens, fraction of neoantigens per mutations, expression of MHC-related molecules, expression of immunomodulatory factors, and mean expression of the respective immune genes for each of the 28 immune cell types as independent variables. Analysis of the association between TILs and MHC molecules showed varying associations. The mean decrease of accuracy over all out-of-bag cross-validated predictions was used to rank the predictors.

The immunophenogram was constructed similarly to the recently proposed metabogram (Hakimi et al., 2016). From the results of the random forest approach, a list of consensus determinants (Figure 5A) that includes 20 single factors (MHC molecules, immunoinhibitors, and immunostimulators) and six cell types (activated CD4⁺ T cells, activated CD8⁺ T cells, effector memory CD4⁺ T cells and effector memory CD8⁺ T cells, Tregs, and MDSCs) was selected for further analyses. For each determinant a sample-wise Z score from gene expression data was calculated. For the six cell types, an average Z score from the corresponding metagenes was calculated. The determinants were then divided into four categories—effector cells (activated CD4⁺ T cells, activated CD8⁺ T cells, effector memory CD4⁺ T cells, and effector memory CD8⁺ T cells), suppressive cells (Tregs and MDSCs), MHC-related molecules, and checkpoints or immunomodulators—and color-coded in the outer part of the wheel (red: positive Z score, blue: negative Z score). The Z scores of the determinants included in the particular category were positively weighted with one (each MHC molecule, inducible T cell costimulator (ICOS), CD27, activated CD4⁺ T cells, activated CD8⁺ T cells, effector memory CD4⁺ T cells, and effector memory CD8⁺ T cells) and negatively weighted with one (PD-1, CTLA4, lymphocyte-activation gene 3 [LAG3], T cell immunoreceptor with Ig and ITIM domains [TIGIT], T cell immunoglobulin and mucin-domain containing-3 [TIM3], PD-L1, PD-L2, Tregs, and MDSCs). The weighted averaged Z score was then calculated by averaging the Z scores within the respective category leading to four values, which were subsequently grayscaled. The immunophenoscore (IPS) was calculated on an arbitrary 0–10 scale based on the sum of the weighted averaged Z score of the four categories, whereby the sum of the Z scores ≥ 3 were designated as IPS10 and the sum of the Z scores ≤ 0 were designated as IPS0.

To determine the predictive power, we stratified patients into responders and non-responders, and performed an ROC validation including the IPS, cytolytic activity, CTLA-4 expression, PD1 expression, and PD-L1 expression as independent variables. As predictive value, the area under the curve (AUC) from the ROC analyses was used (R package *RCCR*). For each of the analyzed patients the corresponding immunophenogram is provided in the TCIA. A tool for construction of the immunophenogram on other expression data was implemented in the TCIA, and a respective R code is available at github (<https://github.com/MayerC-imed/Immunophenogram>).

Statistical Analyses

Sample sizes from available TCGA data were considered adequate because sufficient power using equivalent tests was observed in a previous study (Angelova et al., 2015). To test for differential expression across two groups (tumor and normal), we used the R package *DESeq2* on raw count data. The p values were adjusted for multiple testing based on the false discovery rate (FDR) according to the Benjamini-Hochberg approach. For comparison of two patient groups, two-sided Student's t test was used where stated, otherwise the non-parametric two-sided Wilcoxon-rank sum test was used. For comparisons among multiple-patient groups, one-way ANOVA and Tukey's honest significant difference (HSD) post hoc tests were used where stated, otherwise the non parametric Kruskal-Wallis test followed by two-sided Dunn's pairwise post hoc tests on rank sums with Benjamini-Hochberg adjustment of p values using the R package *PMCMR* were used. Normality of the distributions was tested with Shapiro-Wilk test, and for normal distributed data the variance within each group of data was estimated and tested for equality between groups by a two-sided F-test. Distributions of data are shown either as individual data points, as box-and-whisker plots, or as violin plots. Association of CGAs with CD8⁺ or CD4⁺ T cells was done using Spearman rank correlation, and p values were adjusted according to the Benjamini-Hochberg method.

Overall survival analyses were performed using the R package *survival*, and the patients were dichotomized based on median expression (normalized enrichment score [NES]) or divided in two or more groups by specified parameters. Kaplan-Meier estimator of survival was used to construct the survival curves. Log rank tests (corresponding to a two-sided z test) were used to compare overall survival between patients in different groups, and hazard ratio (HR) (95% confidence interval) was provided for comparison of two groups. The p values were adjusted for multiple testing based on the false discovery rate (FDR) according to the Benjamini-Hochberg method. Patients for each cancer were divided in two groups based on median immunophenoscore, and cox regression analyses were performed and illustrated as forest plot showing log₂(HR) and 95% confidence interval. Proportional hazard assumptions were tested. Analysis and visualization of Gene Ontology terms associated to differentially expressed genes was performed with ClueGO (Bindea et al., 2009).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, four figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2016.12.019>.

AUTHOR CONTRIBUTIONS

Z.T. conceived the project. P.C. developed the GSEA method and analyzed the data. F.F. and M.A. analyzed the neoantigens and CGAs. M.A. and M.E. estimated tumor heterogeneity and clonality of mutations. D.R. organized and managed the data transfer and storage. C.M. and D.R. developed the database. P.C. and H.H. developed the random forest approach and the immunophenogram, and analyzed the data. H.H. and Z.T. interpreted the results. Z.T. wrote the manuscript.

ACKNOWLEDGMENTS

The results shown here are in part based upon data generated by TCGA Research Network (<https://cancergenome.nih.gov>). This work was supported by the Austrian Science Fund (FWF): W1101-B18 (DK Molecular Cell Biology and Oncology), the Tiroler Standortagentur (Bioinformatics Tyrol), the European Union's Horizon 2020 research and innovation program under grant agreement No. 633592 (project APERIM: Advanced Bioinformatics Tools for Personalised Cancer Immunotherapy), and the Austrian National Bank (Jubiläumsfondsprojekt No. 16534).

Received: August 22, 2016

Revised: October 31, 2016

Accepted: December 7, 2016

Published: January 3, 2017

REFERENCES

- Angelova, M., Charoentong, P., Hackl, H., Fischer, M.L., Snajder, R., Krogsdam, A.M., Waldner, M.J., Bindea, G., Mlecnik, B., Galon, J., and Trajanoski, Z. (2015). Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol.* 16, 64.
- Barbie, D.A., Tamayo, P., Boehm, J.S., Kim, S.Y., Moody, S.E., Dunn, I.F., Schinzel, A.C., Sandy, P., Meylan, E., Scholl, C., et al. (2009). Systematic RNA interference reveals that oncogenic KRAS-driven cancers require TBK1. *Nature* 462, 108–112.
- Bindea, G., Mlecnik, B., Hackl, H., Charoentong, P., Tosolini, M., Kirilovsky, A., Fridman, W.H., Pagès, F., Trajanoski, Z., and Galon, J. (2009). ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* 25, 1091–1093.
- Bindea, G., Mlecnik, B., Tosolini, M., Kirilovsky, A., Waldner, M., Obenauf, A.C., Angell, H., Fredriksen, T., Lafontaine, L., Berger, A., et al. (2013).

- Spatiotemporal dynamics of intratumoral immune cells reveal the immune landscape in human cancer. *Immunity* 39, 782–795.
- Blank, C.U., Haanen, J.B., Ribas, A., and Schumacher, T.N. (2016). CANCER IMMUNOLOGY. The “cancer immunogram”. *Science* 352, 658–660.
- Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32.
- Brown, S.D., Warren, R.L., Gibb, E.A., Martin, S.D., Spinelli, J.J., Nelson, B.H., and Holt, R.A. (2014). Neo-antigens predicted by tumor genome meta-analysis correlate with increased patient survival. *Genome Res.* 24, 743–750.
- Cancer Genome Atlas Network (2015). Genomic classification of cutaneous melanoma. *Cell* 161, 1681–1696.
- Cancer Genome Atlas Research Network (2014). Integrated genomic characterization of papillary thyroid carcinoma. *Cell* 159, 676–690.
- Covre, A., Coral, S., Nicolay, H., Parisi, G., Fazio, C., Colizzi, F., Fratta, E., Di Giacomo, A.M., Sigalotti, L., Natali, P.G., and Maio, M. (2015). Antitumor activity of epigenetic immunomodulation combined with CTLA-4 blockade in syngeneic mouse models. *Oncolimmunology* 4, e1019978.
- Gajewski, T.F., Meng, Y., Blank, C., Brown, I., Kacha, A., Kline, J., and Harlin, H. (2006). Immune resistance orchestrated by the tumor microenvironment. *Immunol. Rev.* 213, 131–145.
- Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., Tosolini, M., Camus, M., Berger, A., Wind, P., et al. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* 313, 1960–1964.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Gerlinger, M., Rowan, A.J., Horswell, S., Larkin, J., Endesfelder, D., Gronroos, E., Martinez, P., Matthews, N., Stewart, A., Tarpey, P., et al. (2012). Intratumor heterogeneity and branched evolution revealed by multiregion sequencing. *N. Engl. J. Med.* 366, 883–892.
- Hackl, H., Charoentong, P., Finotello, F., and Trajanoski, Z. (2016). Computational genomics tools for dissecting tumour-immune cell interactions. *Nat. Rev. Genet.* 17, 441–458.
- Hakimi, A.A., Reznik, E., Lee, C.H., Creighton, C.J., Brannon, A.R., Luna, A., Aksoy, B.A., Liu, E.M., Shen, R., Lee, W., et al. (2016). An integrated metabolic atlas of clear cell renal cell carcinoma. *Cancer Cell* 29, 104–116.
- Hugo, W., Zaretsky, J.M., Sun, L., Song, C., Moreno, B.H., Hu-Lieskovian, S., Berent-Maoz, B., Pang, J., Chmielowski, B., Cherry, G., et al. (2016). Genomic and transcriptomic features of response to anti-PD-1 therapy in metastatic melanoma. *Cell* 165, 35–44.
- Iglesia, M.D., Parker, J.S., Hoadley, K.A., Serody, J.S., Perou, C.M., and Vincent, B.G. (2016). Genomic analysis of immune cell infiltrates across 11 tumor types. *J. Natl. Cancer Inst.* 108, djw144.
- Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152, 714–726.
- Lawrence, M.S., Stojanov, P., Polak, P., Kryukov, G.V., Cibulskis, K., Sivachenko, A., Carter, S.L., Stewart, C., Mermel, C.H., Roberts, S.A., et al. (2013). Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* 499, 214–218.
- Li, B., Severson, E., Pignon, J.C., Zhao, H., Li, T., Novak, J., Jiang, P., Shen, H., Aster, J.C., Rodig, S., et al. (2016). Comprehensive analyses of tumor immunity: implications for cancer immunotherapy. *Genome Biol.* 17, 174.
- McGranahan, N., Furness, A.J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S.K., Jamal-Hanjani, M., Wilson, G.A., Birkbak, N.J., Hiley, C.T., et al. (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* 351, 1463–1469.
- Mlecnik, B., Tosolini, M., Kirilovsky, A., Berger, A., Bindea, G., Meatchi, T., Bruneval, P., Trajanoski, Z., Fridman, W.H., Pagès, F., and Galon, J. (2011). Histopathologic-based prognostic factors of colorectal cancers are associated with the state of the local immune reaction. *J. Clin. Oncol.* 29, 610–618.
- Newman, A.M., Liu, C.L., Green, M.R., Gentles, A.J., Feng, W., Xu, Y., Hoang, C.D., Diehn, M., and Alizadeh, A.A. (2015). Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* 12, 453–457.
- Nielsen, M., Lundsgaard, C., Blicher, T., Lamberth, K., Harndahl, M., Justesen, S., Røder, G., Peters, B., Sette, A., Lund, O., and Buus, S. (2007). NetMHCpan, a method for quantitative predictions of peptide binding to any HLA-A and -B locus protein of known sequence. *PLoS ONE* 2, e796.
- Oikonomou, E., Koustas, E., Goulielmaki, M., and Pintzas, A. (2014). BRAF vs RAS oncogenes: are mutations of the same pathway equal? Differential signaling and therapeutic implications. *Oncotarget* 5, 11752–11777.
- Pitt, J.M., Vétizou, M., Daillière, R., Roberti, M.P., Yamazaki, T., Routy, B., Lepage, P., Boneca, I.G., Chamaillard, M., Kroemer, G., and Zitvogel, L. (2016). Resistance mechanisms to immune-checkpoint blockade in cancer: tumor-intrinsic and -extrinsic factors. *Immunity* 44, 1255–1269.
- Rizvi, N.A., Hellmann, M.D., Snyder, A., Kvistborg, P., Makarov, V., Havel, J.J., Lee, W., Yuan, J., Wong, P., Ho, T.S., et al. (2015). Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* 348, 124–128.
- Robbins, P.F., Lu, Y.C., El-Gamil, M., Li, Y.F., Gross, C., Gartner, J., Lin, J.C., Teer, J.K., Cliften, P., Tykkesen, E., et al. (2013). Mining exomic sequencing data to identify mutated antigens recognized by adoptively transferred tumor-reactive T cells. *Nat. Med.* 19, 747–752.
- Rooney, M.S., Shukla, S.A., Wu, C.J., Getz, G., and Hacohen, N. (2015). Molecular and genetic properties of tumors associated with local immune cytolytic activity. *Cell* 160, 48–61.
- Rubio-Perez, C., Tamborero, D., Schroeder, M.P., Antolín, A.A., Deu-Pons, J., Perez-Llamas, C., Mestres, J., Gonzalez-Perez, A., and Lopez-Bigas, N. (2015). In silico prescription of anticancer drugs to cohorts of 28 tumor types reveals targeting opportunities. *Cancer Cell* 27, 382–396.
- Schadendorf, D., Hodi, F.S., Robert, C., Weber, J.S., Margolin, K., Hamid, O., Patt, D., Chen, T.T., Berman, D.M., and Wolchok, J.D. (2015). Pooled analysis of long-term survival data from phase II and phase III trials of ipilimumab in unresectable or metastatic melanoma. *J. Clin. Oncol.* 33, 1889–1894.
- Schumacher, T.N., and Schreiber, R.D. (2015). Neoantigens in cancer immunotherapy. *Science* 348, 69–74.
- Spencer, K.R., Wang, J., Silk, A.W., Ganesan, S., Kaufman, H.L., and Mehrt, J.M. (2016). Biomarkers for immunotherapy: current developments and challenges. *Am. Soc. Clin. Oncol. Educ. Book* 35, e493–e503.
- Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., and Mesirov, J.P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* 102, 15545–15550.
- Van Allen, E.M., Miao, D., Schilling, B., Shukla, S.A., Blank, C., Zimmer, L., Sucker, A., Hillen, U., Geukes Foppen, M.H., Goldinger, S.M., et al. (2015). Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* 350, 207–211.
- van Rooij, N., van Buuren, M.M., Philips, D., Velds, A., Toebe, M., Heemskerk, B., van Dijk, L.J., Behjati, S., Hilkmann, H., El Atmoui, D., et al. (2013). Tumor exome analysis reveals neoantigen-specific T-cell reactivity in an ipilimumab-responsive melanoma. *J. Clin. Oncol.* 31, e439–e442.
- Witkiewicz, A.K., McMillan, E.A., Balaji, U., Baek, G., Lin, W.C., Mansour, J., Mollaee, M., Wagner, K.U., Koduru, P., Yopp, A., et al. (2015). Whole-exome sequencing of pancreatic cancer defines genetic diversity and therapeutic targets. *Nat. Commun.* 6, 6744.
- Wolchok, J.D. (2015). PD-1 blockers. *Cell* 162, 937.