

# SCIENTIFIC REPORTS



OPEN

## Mutations in DNA repair genes are associated with increased neoantigen burden and a distinct immunophenotype in lung squamous cell carcinoma

Young Kwang Chae<sup>1,2</sup>, Jonathan F. Anker<sup>1</sup>, Michael S. Oh<sup>1</sup>, Preeti Bais<sup>3</sup>, Sandeep Namburi<sup>3</sup>, Sarita Agte<sup>1</sup>, Francis J. Giles<sup>1,2</sup> & Jeffrey H. Chuang<sup>3,4</sup>

Deficiencies in DNA repair pathways, including mismatch repair (MMR), have been linked to higher tumor mutation burden and improved response to immune checkpoint inhibitors. However, the significance of MMR mutations in lung cancer has not been well characterized, and the relevance of other processes, including homologous recombination (HR) and polymerase epsilon (POLE) activity, remains unclear. Here, we analyzed a dataset of lung squamous cell carcinoma samples from The Cancer Genome Atlas. Variants in DNA repair genes were associated with increased tumor mutation and neoantigen burden, which in turn were linked with greater tumor infiltration by activated T cells. The subset of tumors with DNA repair gene variants but without T cell infiltration exhibited upregulation of TGF- $\beta$  and Wnt pathway genes, and a combined score incorporating these genes and DNA repair status accurately predicted immune cell infiltration. Finally, high neoantigen burden was positively associated with genes related to cytolytic activity and immune checkpoints. These findings provide evidence that DNA repair pathway defects and immunomodulatory genes together lead to specific immunophenotypes in lung squamous cell carcinoma and could potentially serve as biomarkers for immunotherapy.

Immune checkpoint inhibitors have reshaped the landscape of treatment for multiple cancers, including squamous cell carcinoma (SCC) of the lung and other types of non-small cell lung cancer (NSCLC)<sup>1,2</sup>. These treatments inhibit immune regulatory molecules such as cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), programmed death 1 (PD-1), and programmed death ligand 1 (PD-L1), which normally function to suppress immune cell activity<sup>3,4</sup>. Blocking immune checkpoints with therapeutic antibodies can augment the anti-tumor immune response, thereby providing the mechanistic basis for immunotherapy. In NSCLC, treatment with immune checkpoint inhibitors has yielded dramatic results with improved clinical response in comparison to standard chemotherapy in certain subpopulations of patients<sup>5,6</sup>.

The specificity of the immune response promoted by these therapies is dependent on neoantigens, which are immunogenic cancer-related peptides formed by distinct somatic mutations in tumor cells<sup>7</sup>. The unique epitopes of these neoantigens are able to elicit a tumor-specific immune response<sup>8</sup>, which can then be amplified by the immune-activating actions of immunotherapy. Neoantigens have been associated with improved clinical response to inhibitors of CTLA-4<sup>9,10</sup>, PD-1<sup>11</sup>, and PD-L1<sup>12</sup>. In many solid tumors, deleterious mutations in DNA repair genes can drive a substantial increase in the number of neoantigens<sup>13</sup>. Deficient DNA repair has accordingly been associated with improved clinical responses to PD-1 blockade. Specifically, insufficiencies in mismatch repair (MMR) conferred greater clinical benefit with pembrolizumab in patients with colorectal cancer<sup>14</sup>, as well

<sup>1</sup>Northwestern University Feinberg School of Medicine, Chicago, IL, 60611, USA. <sup>2</sup>Robert H. Lurie Comprehensive Cancer Center of Northwestern University, Chicago, IL, 60611, USA. <sup>3</sup>The Jackson Laboratory for Genomic Medicine, Farmington, CT, 06030, USA. <sup>4</sup>Department of Genetics and Genome Sciences, University of Connecticut Health, Farmington, CT, 06032, USA. Young Kwang Chae, Jonathan F. Anker and Michael S. Oh contributed equally. Correspondence and requests for materials should be addressed to Y.K.C. (email: [young.chae@northwestern.edu](mailto:young.chae@northwestern.edu))

as in a study of multiple solid tumor types<sup>15</sup>. These results have now led to the landmark FDA approval for PD-1 inhibitors in MMR-deficient tumors, which represents a paradigm-altering shift towards oncologic treatments centered on molecular profile<sup>15</sup>.

Several other DNA repair pathways have been implicated in contributing to neoantigen load. In an analysis of NSCLC patients, mutations in *POLD1*, *POLE*, and *MSH2* were identified in tumors with the highest neoantigen burden<sup>11</sup>, which in turn correlated with improved response to PD-1 inhibitors. Further, endometrial cancers with polymerase epsilon (*POLE*) mutations contained increased neoantigen burden and PD-L1 expression<sup>16</sup>, and cases of exceptional responders to immunotherapy have been reported with these mutations<sup>17</sup>. Similarly, alterations in the homologous recombination (HR) apparatus, such as *BRCA1* and *BRCA2* mutations, were associated with higher neoantigen load and increased overall survival after anti-PD-1 treatment<sup>18</sup>.

Though DNA repair mutations have been shown to be relevant to immunotherapy response in a variety of solid tumors, limited data exists detailing the importance of these pathways in lung cancer. We had previously utilized datasets from The Cancer Genome Atlas (TCGA) to demonstrate that DNA repair status was strongly associated with tumor neoantigen burden and immune cell infiltration in lung adenocarcinoma<sup>19</sup>. We hypothesized that a comparable relationship would be elucidated in squamous cell carcinoma (SqCC) of the lung, and that mutations in DNA repair pathways could thus function as biomarkers predictive of response to immune checkpoint blockade.

## Results

**Tumors with DNA repair pathway mutations have increased mutational and neoantigen burden.** To study the effect of DNA repair gene mutations on tumor mutation burden (TMB) in lung SqCC, we analyzed 178 annotated samples from TCGA<sup>20</sup>. We evaluated tumors for somatic variants in genes related to MMR, HR, or in *POLE*, and identified changes predicted to be deleterious by the SIFT<sup>21</sup> and CADD v1.4<sup>22</sup> scoring systems. Tumors with defects in MMR and HR had a significantly higher number of overall mutations (Student's t-test,  $p < 0.0001$  for both; Fig. 1a). Within HR genes, *BRCA1* and *BRCA2* were the most commonly mutated (7.9% of tumors), and were associated with increased TMB (Student's t-test,  $p < 0.0001$ ) (Supplementary Fig. S1a). Tumors with multiple DNA repair gene variants had corresponding increases in TMB. For example, tumors with 1 affected gene had an average of  $293.8 \pm 27.0$  tumor mutations, while those with 3–5 affected genes had  $815.8 \pm 248.6$  mutations (Fig. 1b). *POLE* variants were rare ( $n = 8$ ) but were also associated with increased TMB (Student's t-test,  $p = 0.010$ ). There was no difference in smoking history between tumors with low and high TMB (Supplementary Fig. S1c).

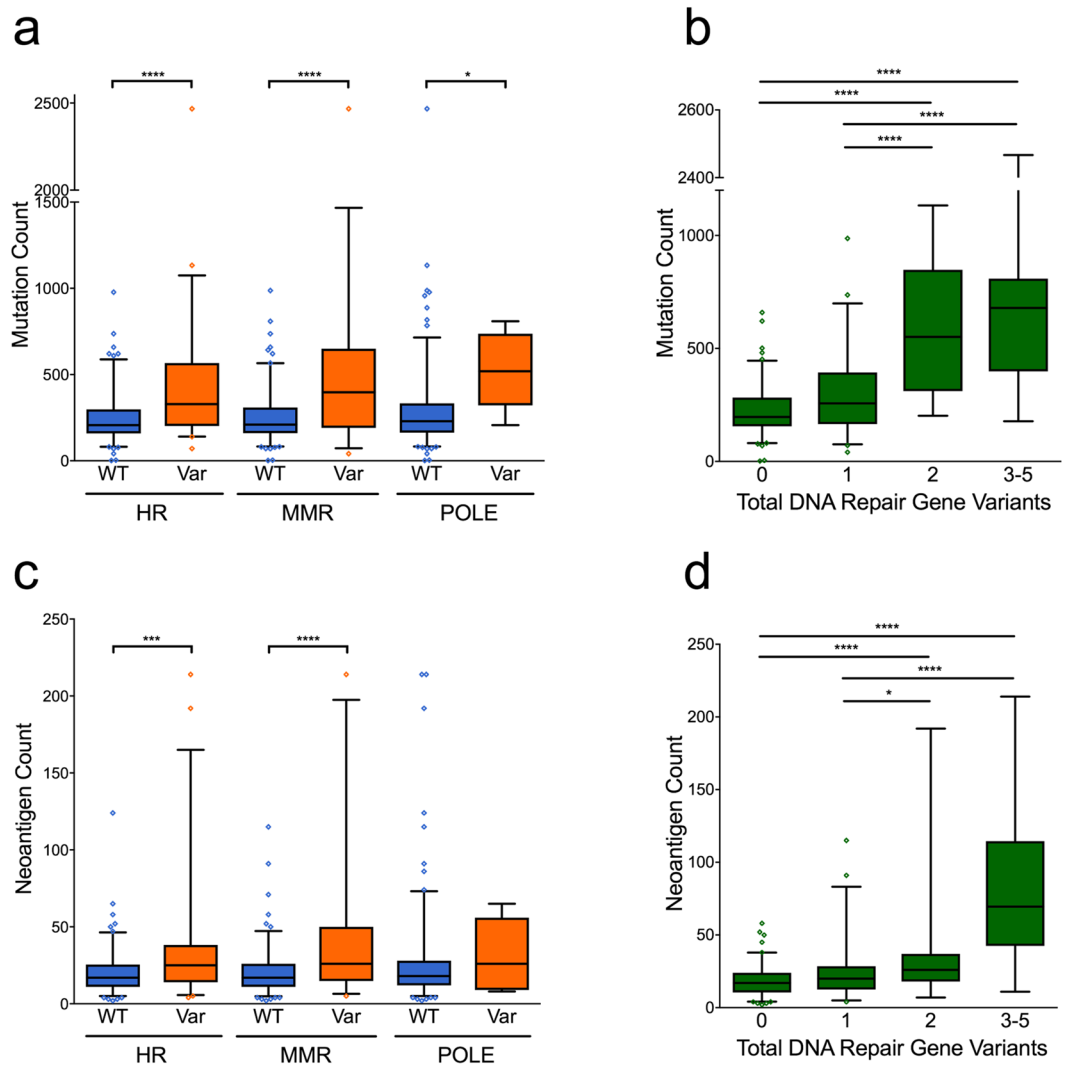
We next calculated neoantigen burden in the samples by filtering total non-synonymous mutations based on predicted MHC binding affinity as a surrogate for immunogenicity. Strong putative binders were then further selected by assessing for poor immunogenicity of the non-mutated parental epitope. This total predicted neoantigen burden was significantly greater in tumors with somatic variants in HR (Student's t-test,  $p = 0.0003$ ) and MMR ( $p < 0.0001$ ), but not *POLE* ( $p = 0.538$ ) (Fig. 1c). Neoantigen burden was also positively associated with the number of affected DNA repair genes (Fig. 1d).

**High mutation burden is associated with increased tumor infiltration by activated T cells.** To determine the association between TMB and tumor-infiltrating lymphocytes (TILs), we divided TCGA samples into high- and low-mutation groups based on the median mutation count of 232. We then assessed immune cell infiltration from gene expression data as previously described<sup>23</sup> (see Methods). Tumors with high TMB were more likely to be infiltrated by activated CD4+ (proportion Z-score,  $p = 0.013$ ) and activated CD8+ ( $p = 0.036$ ) T cells (Fig. 2a), but this finding was not statistically significant after adjusting for multiple comparisons (adjusted  $p = 0.38$  and  $0.52$ , respectively). Infiltration by activated CD4+ and CD8+ T cells was positively correlated with infiltration by CD4+ effector memory T cells, type 2 helper cells, memory B cells, myeloid dendritic cells, and myeloid-derived suppressor cells (Fig. 2b).

**DNA repair gene variants are not associated with T cell infiltration, possibly due to compensatory immunosuppressive signals.** Presence of variants in HR, MMR, and *POLE* did not predict infiltration by CD4+, CD8+, or total activated T cells (Fig. 3a). To explain this finding, we categorized the samples into a  $2 \times 2$  classification scheme based on DNA repair status and activated T cell infiltration (Fig. 3b). This grouping delineated four types of tumor: DNA repair variant absent without T cell infiltration (group I), DNA repair variant present without T cell infiltration (group II), DNA repair variant absent with infiltration (group III), and DNA repair variant present with infiltration (group IV).

We then evaluated the mRNA expression of genes previously shown to impair T cell infiltration, including those related to transforming growth factor beta (TGF- $\beta$ ) and the  $\beta$ -catenin/Wnt pathway<sup>24</sup>. Both group I and II tumors, which lack activated T cells, had significantly increased expression of TGF- $\beta$  genes. Specifically, there were significant differences when comparing group II and IV tumors in regards to *TGFB1* (one-way ANOVA with Tukey's test for multiple comparisons,  $p = 0.002$ ) *TGFB3* ( $p = 0.049$ ), and *WNT2* ( $p = 0.029$ ) (Fig. 3c and Supplementary Fig. S2a). There was no significant association between groups based on *VEGF-A* expression (Fig. 3c and Supplementary Fig. S2a). We also performed a multivariate logistic regression analysis to determine the influence of these expression levels on group identity. When compared to group IV as a baseline, *TGFB1* and *WNT2* were significant predictors for groups I and II, while decreased levels of *WNT7A* was a significant predictor for group III (Table 1).

Presence of somatic variants in genes related to antigen presentation, including *HLA-A*, *HLA-B*, *HLA-C*, *B2M* ( $\beta$ -2-microglobulin), *TAP2*, and *PSMB8* (LMP-7), did not predict tumor classification (Supplementary Fig. S2b). Variants in the  $\beta$ -catenin/Wnt pathway were similarly not associated with tumor groups (Supplementary Fig. S2b).

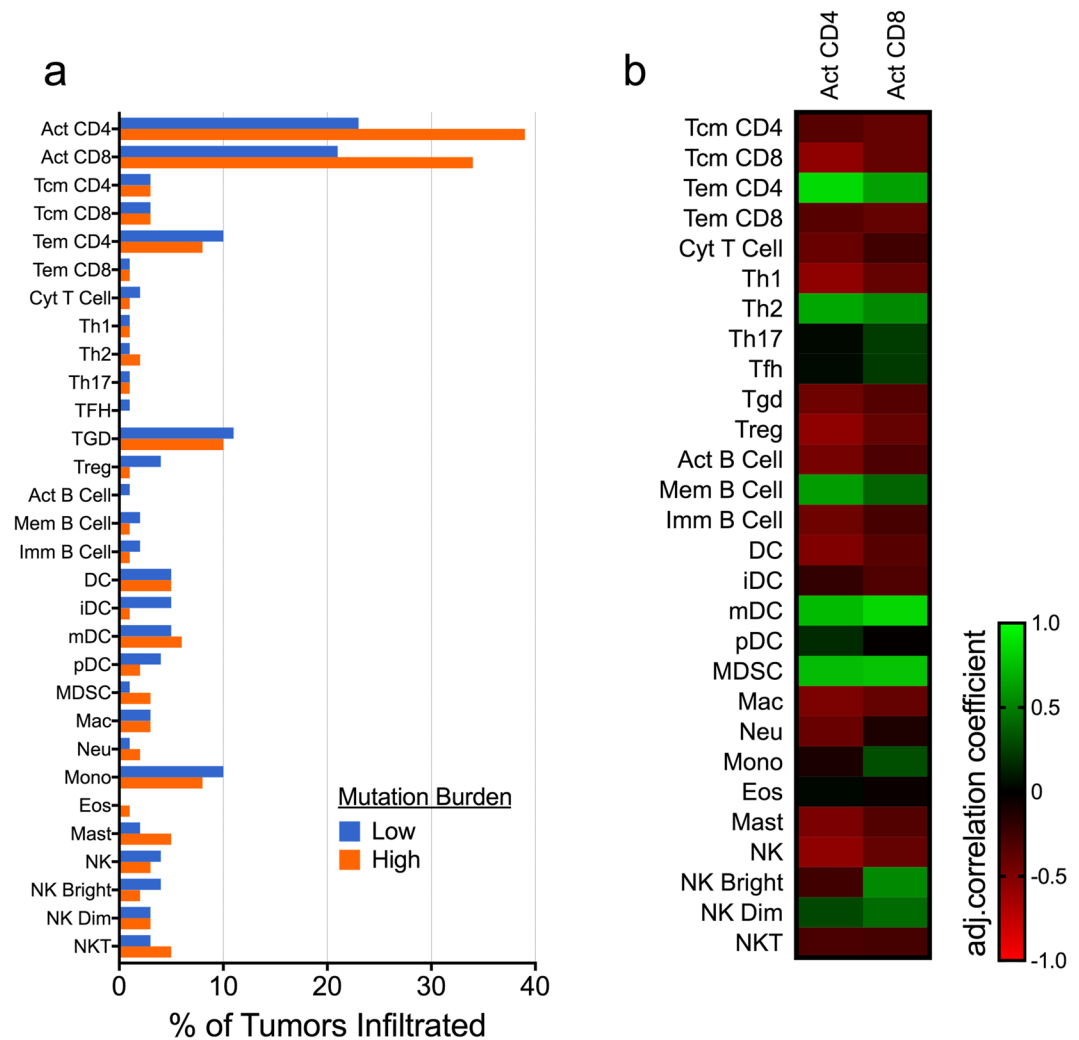


**Figure 1.** DNA repair gene variants are associated with increased mutation and neoantigen count. **(a)** Presence of somatic variants in homologous recombination (HR), mismatch repair (MMR) or polymerase epsilon (POLE) were associated with increased mutation burden. **(b)** Mutation count increases with higher number of DNA repair gene variants. **(c)** Neoantigen burden similarly was associated with DNA repair gene variants and **(d)** with the number of affected genes. Statistical analysis completed with Student's t-test **(a,c)** and one-way ANOVA with Tukey's test for multiple comparisons **(b,d)**, \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , \*\*\*\* $p < 0.0001$ .

**A combined score incorporating DNA repair status, *TGFB1*, and *WNT2* predicts T cell infiltration.** As a proof of principle, we sought to compute a score based on the above data that could better predict T cell infiltration. Given the results of our logistic regression model, we conferred 1 point each for an expression z-score greater than the median for *TGFB1* and *WNT2*. We then added another point for absence of any DNA repair gene variants. Thus, a score of 3 indicates intact DNA repair pathways and high TGF- $\beta$ /Wnt2, while a score of 0 indicates presence of DNA repair pathway variants and low TGF- $\beta$ /Wnt2. We hypothesized that low scores would lead to increased T cell infiltration. Using a threshold of  $\leq 1$ , tumors with low scores indeed demonstrated a significantly increased likelihood of CD4<sup>+</sup> (false discovery rate-adjusted proportion Z-score,  $p < 0.0003$ ) and CD8<sup>+</sup> ( $p = 0.012$ ) activated T cell infiltration (Fig. 3d).

We compared the expression of *TGFB1* and *WNT2* in tumors with and without activated T cells infiltration, after stratifying tumors based on DNA repair gene status (Supplementary Fig. S3a,b). *TGFB1* gene expression was significantly associated with activated CD4 and CD8 T cell infiltration regardless of DNA repair status, while *WNT2* was associated with T cell infiltration only in tumors with wild-type DNA repair genes (Supplementary Fig. S3a,b). Furthermore, we found that the combined score exhibited stronger correlations with immune cell infiltration than did either TMB or neoantigen burden (Supplementary Fig. S3c).

**Mutation and neoantigen burden are associated with increased expression of pro-inflammatory and immune checkpoint markers.** We next utilized RNA-Seq expression data from TCGA to identify any potential immune signature related to increased tumor mutations and neoantigens. We first assessed associations

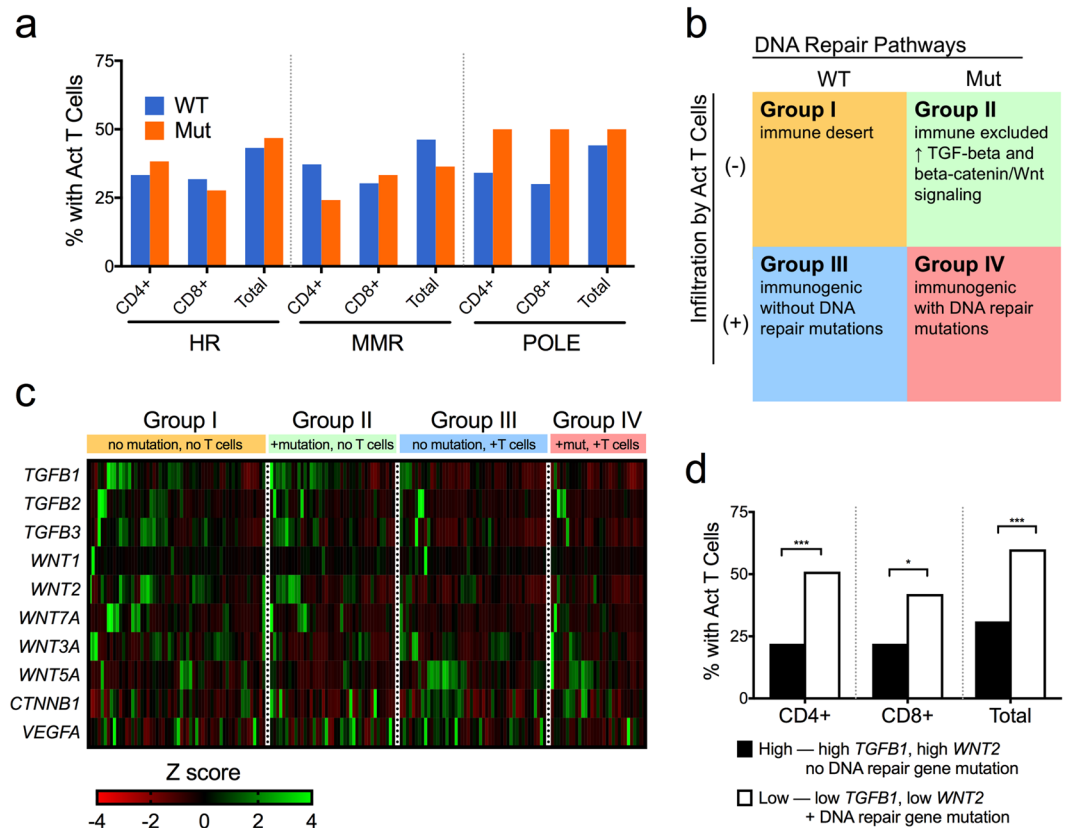


**Figure 2.** Activated T cell infiltration is increased in high-mutation tumors and associated with a specific immunophenotype. **(a)** Tumors with higher mutation burden were more likely to contain activated (Act) CD4+ and CD8+ activated T cells, though this difference was not significant after adjusting for multiple comparisons. **(b)** Activated CD4+ and CD8+ T cells were both positively correlated with CD4+ effector memory T cells (Tem), type 2 helper cells (Th2), memory (Mem) B cells, myeloid dendritic cells (mDC), and myeloid-derived suppressor cells (MDSC). Additional cell types analyzed include central memory T cells (Tcm); follicular helper T cells (Th1); regulatory T cells (Tregs); gamma-delta T cells (Tgd); dendritic cells (DC); immature dendritic cells (iDC); plasmacytoid dendritic cells (pDC); macrophages (Mac); neutrophils (Neu); monocytes (Mono); eosinophils (Eos); mast cells (Mast); natural killer cells (NK); CD56 bright NK cells (NK Bright); CD56 dim NK cells (NK Dim); and natural killer T cells (NKT).

between tumor mutation burden and mRNA levels of an immunomodulatory gene set<sup>23</sup>. SqCC samples with increased TMB had significantly increased expression of genes associated with immune activity, including *GZMA*, *GZMB*, and *PRF1* (Fig. 4a,b). They had higher levels of *IFNG* (interferon- $\gamma$ ) and the interferon-stimulated chemokine *CXCL9*. Meanwhile, high TMB tumors demonstrated decreased expression of *TGFB1* and *PRDM1* (Fig. 4a,b). Genes related to M1 and M2 tumor-associated macrophages (TAMs) did not have either higher or lower expression in these tumors (Fig. 4a).

In addition, we assessed for a comprehensive set of genes encoding immune-related cytokines and cytokine receptors. Expression levels of the cytokines *CCL8*, *CCL17*, *TNFSF13B*, *XCL1*, and *XCL2* were significantly increased in high TMB tumors (see Supplementary Fig. S4). Expression of the cytokines *CCL17*, *CCL22*, *CX3CL1*, *CXCL12*, *TNF*, and *TSLP*, as well as the cytokine receptors *CX3CR1*, *IL17RA*, and *IL31RA* were decreased in these tumors.

We subsequently classified SqCC samples by neoantigen load and immunophenotypic group and assessed the resulting immune signature. High-neoantigen tumors exhibited increased gene expression of the pro-inflammatory markers *GZMA*, *GZMB*, *PRF1*, *CD8A*, *EOMES*, *CXCL9*, and *IFNG*, as well as the immune checkpoint marker *LAG3* (Fig. 4c and Supplementary Table S1). The immunophenotypic groups (*i.e.* the classification based on DNA repair status and T cell infiltration) likewise demonstrated unique immune signatures

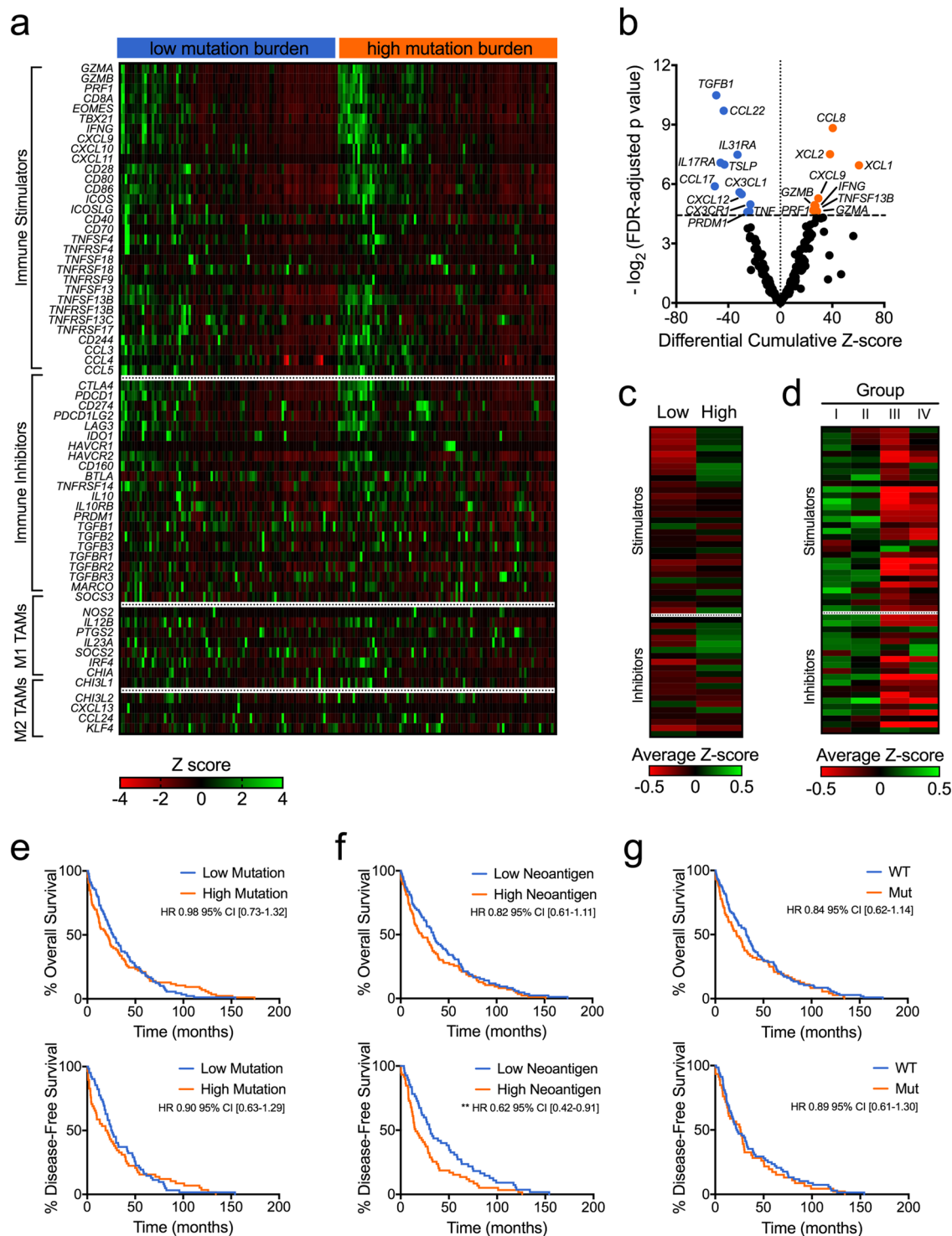


**Figure 3.** DNA repair gene variants and immune signals together predict infiltration by activated T cells. (a) DNA repair gene variants did not lead to a change in activated T cell infiltration. (b) These findings suggested that the tumors could be divided into four immunophenotypic groups based on DNA repair status and T cell infiltration. (c) A heat map of immunosuppressive genes showed differential mRNA expression among these groups. (d) A combined score incorporating DNA repair gene variants as well as *TGFB1* and *WNT2* expression was significantly associated with activated T cell infiltration (proportion Z-score, \*FDR-adjusted p-value < 0.05, \*\*\*p < 0.001).

	Group I			Group II			Group III		
	Coef.	SE	p-value	Coef.	SE	p-value	Coef.	SE	p-value
<i>APC</i>	-0.166	0.235	0.480	0.104	0.228	0.649	-0.279	0.225	0.215
<i>CTNNB1</i>	-0.019	0.229	0.935	0.114	0.230	0.620	-0.015	0.217	0.944
<i>WNT1</i>	2.240	1.226	0.068	2.130	1.242	0.086	2.117	1.225	0.084
<i>WNT2</i>	0.970	0.410	0.018*	0.967	0.425	0.023*	0.331	0.437	0.449
<i>WNT3A</i>	0.141	0.235	0.547	-0.146	0.291	0.615	-0.032	0.237	0.891
<i>WNT5A</i>	0.232	0.277	0.401	-0.607	0.329	0.065	0.078	0.246	0.751
<i>WNT7A</i>	-0.323	0.303	0.286	-0.501	0.316	0.113	-0.828	0.402	0.039*
<i>TGFB1</i>	1.030	0.409	0.012*	1.341	0.421	0.001**	0.736	0.418	0.079
<i>TGFB2</i>	-0.232	0.179	0.195	-0.242	0.200	0.226	-0.318	0.219	0.145
<i>TGFB3</i>	0.097	0.355	0.785	-0.098	0.385	0.799	-0.034	0.390	0.930
<i>VEGFA</i>	0.203	0.234	0.385	0.321	0.235	0.172	0.312	0.216	0.148

**Table 1.** Predictors of immunophenotypic group based on logistic regression analysis. A multinomial logistic regression model showed that *WNT2* and *TGFB1* had the strongest influence on determining group identity, using Group IV as a baseline. This table lists regression coefficients (Coef.) and standard errors (SE), \*p < 0.05, \*\*p < 0.01.

(Fig. 4d and Supplementary Table S2). Notably, there was no significant difference in genes related to cytolytic activity, such as *GZMA*, *GZMB*, and *PRF1* (Supplementary Table S2). These groups did differ in expression levels of immune activation genes, including *CD28*, *CD80*, *CD86*, *ICOS*, *ICOSLG*, and *CD40*. They also had significant differences in regards to the immune checkpoint markers *CTLA4*, *PDCD1*, and *HAVCR2* (TIM-3).



**Figure 4.** Mutation burden is associated with a unique immune and clinical profile. **(a)** High mutation burden was associated with an immune signature that includes increased expression of genes related to cytolytic and interferon- $\gamma$  signaling. **(b)** A volcano plot shows differential mRNA expression in low versus high mutation tumors (Student's t-test with FDR-adjusted p-values). Unique immune signatures were associated with **(c)** neoantigen load and **(d)** immunophenotypic group. **(e)** Mutation burden was not associated with significant differences in overall and disease-free survival. **(f)** High neoantigen burden was not associated with overall survival, but did lead to worse disease-free survival **(g)** Presence of DNA repair gene variants did not affect overall or disease-free survival. Kaplan-Meier plots shown, \*\* $p < 0.01$  based on log-rank test.

Clinical analysis of this SqCC cohort has previously classified tumors into four subtypes that reflect their underlying biologic processes<sup>25</sup>. Mutation burden alone did not have strong association with any subtype (Supplementary Fig. S5a). However, MMR-variant tumors were more likely to have a secretory subtype (false discovery rate-adjusted proportion Z-score,  $p = 0.029$ ) (Supplementary Fig. S5b).

**DNA repair status is not associated with overall survival.** Mutation burden has been shown to affect both treatment response<sup>26</sup> as well as intrinsic survival<sup>27</sup> in multiple cancer types. In our cohort, overall survival was not affected by high mutation burden (HR 0.98 with 95% confidence interval 0.73–1.32) or neoantigen burden (HR 0.82 [0.61–1.11]) (Fig. 4e,f). Disease-free survival was not increased in the high TMB group (HR 0.90 [0.63–1.29]), but it was significantly improved in the high neoantigen group (HR 0.64 [0.44–0.93],  $p = 0.007$ ). Presence of DNA repair variants was not associated with change in overall survival (HR 0.84 [0.62–1.14]) or disease-free survival (HR 0.89 [0.61–1.30]) (Fig. 4g). The immunophenotypic group also did not significantly affect survival outcomes (Supplementary Fig. S6a,b).

## Discussion

Defects in DNA repair pathways have been shown to strongly affect the tumor immune profile and consequently the clinical response to immunotherapy. We here sought to characterize the role of DNA repair gene mutations in shaping immunological characteristics in lung squamous cell carcinoma.

Our results overall support findings seen in other cancer types, and demonstrate that mutations in DNA repair pathways are associated with high tumor mutation burden (TMB). We also used mutation data to predict presence of tumor-specific neoantigens, which may be more relevant to anti-tumor immunity<sup>28</sup>, and showed that presence of DNA repair gene variants was associated with high neoantigen load. DNA repair status could thus serve as a surrogate marker for identifying patients with increased TMB. Recent clinical trials have demonstrated greater efficacy of immune checkpoint inhibitors in patients with NSCLC and high TMB<sup>26,29</sup>. Though targeted next-generation sequencing has been shown to be a viable method for measuring TMB<sup>30</sup>, assessing for mutation load remains a resource-intensive process. Use of more limited gene panels, such as one focused on DNA repair pathways, may be more practical for widespread clinical implementation.

An effective immune response requires not only the immunogenic impetus provided by tumor mutations, but also the ability of immune cells to infiltrate the tumor parenchyma<sup>24,31</sup>. In order to evaluate tumor infiltration, we utilized a comprehensive set of immune “metagenes” previously validated to estimate immune cell subpopulations<sup>23</sup>. In tumors with greater mutation load, there was a higher percentage of tumors infiltrated by activated CD4+ and CD8+ T cells. Though this difference was not statistically significant when adjusted for multiple comparisons, high TMB tumors had increased mRNA expression of genes that indicate T cell cytolytic activity, such as *GZMA* and *PRF1*<sup>32</sup>. High-mutation and high-neoantigen tumors also had significantly elevated expression of IFN- $\gamma$ -inducible chemokines such as *CXCL9*, which promotes trafficking of activated T cells<sup>33</sup>. Furthermore, infiltration by activated T cells correlated with an increased presence of effector memory CD4+ T cells and myeloid dendritic cells, both of which play important roles in T cell stimulation and functionality<sup>34,35</sup>. These data together suggest that a real association between increased mutation load and T cell infiltration exists in lung SqCC.

Despite the relationship between mutation load and T cell infiltration, we did not find a direct association between DNA repair gene variants and tumor immunophenotype. Notably, there was a sizeable sub-group of tumors with repair gene variants (many with high mutation burden) that nonetheless did not exhibit T cell infiltration. This corresponds to prior categorizations of tumors into distinct categories based on the cancer-immunity cycle<sup>24</sup>: “immune desert” tumors with low neoantigen burden, “immune excluded” tumors that suppress T cell infiltration despite adequate neoantigens, and “inflamed” tumors with increased T cell activity. Our group of DNA repair gene-variant tumors without infiltrated T cells (*i.e.* “group II”) could thus represent an immune excluded phenotype.

Several factors have been postulated as contributing to immune exclusion. TGF- $\beta$  represents a family of cytokines that regulate immune activity and have been demonstrated to reduce functionality of TILs<sup>36,37</sup>. High levels of TGF- $\beta$  therefore represent a mechanism for tumors to impair immune cell infiltration despite the presence of adequate neoantigens. The  $\beta$ -catenin/Wnt pathway has also been implicated as contributing to decreased TILs and abrogating efficacy of immunotherapies<sup>38</sup>. In order to assess their potential role in lung SqCC, we looked at mRNA expression of TGF- $\beta$ ,  $\beta$ -catenin, and 5 Wnt genes known to be overexpressed in NSCLC<sup>39</sup>. We found that expression levels of genes related to TGF- $\beta$  and the Wnt pathway were significantly increased in our putative immune excluded group. Deficiencies in the antigen presenting apparatus have been linked to low anti-tumor immune activity<sup>40</sup>, but mutations in relevant genes (including HLA molecules and  $\beta$ -2-microglobulin) were not predictive of immunophenotypic groups. HLA defects have been linked specifically to escape mechanisms in the context of acquired resistance to immunotherapy<sup>41</sup>, and they thus may not have relevance in a broader set of tumors.

By combining DNA repair status with expression of *TGFB1* and *WNT2* genes, we were able to calculate a multifactorial score that better predicted infiltration by activated T cells. This score is a crude measure but serves as a proof of concept demonstrating that DNA repair gene aberrations, when adjusted for tumor microenvironment factors, can help identify inflamed tumors. This finding does exist at odds with studies showing that MMR status alone can predict TIL presence and subsequent response to immunotherapies<sup>15</sup>. The use of somatic variants with unknown functional significance may have limited our analysis. However, it may also be that lung SqCC biology exhibits a greater propensity to create an immunosuppressive tumor milieu, especially when compared to colorectal<sup>42</sup>, endometrial<sup>16</sup>, and other cancer types whose immune responsiveness have been closely tied to DNA repair deficiencies.

In contrast to the results in SqCC, our prior study in lung adenocarcinoma did observe a direct relationship between DNA repair gene variants and activated T cell infiltration<sup>19</sup>. Important clinical differences have previously been observed between SqCC and adenocarcinoma; in particular, many recent advances in NSCLC treatment have had less impact on SqCC<sup>1,43</sup>. Nevertheless, these two subtypes frequently are grouped together, and few studies have explored the potential differences in their response to immune checkpoint inhibitors<sup>6</sup>. Our data imply that DNA repair status and TMB, when used as individual biomarkers, may be less predictive of immunophenotype in SqCC when compared to adenocarcinoma. Comprehensive immune signatures may instead be necessary to sufficiently capture relevant immune parameters in SqCC. In this dataset, we found significant differences between low- and high-mutation tumors in regards to immune-related gene expression, including in genes related to cytolytic activity and IFN- $\gamma$ -inducible pro-inflammatory factors.

Increased mutation burden did not affect survival outcomes, but high neoantigen burden was associated with worse disease-free survival, possibly due to greater clinical relevance of tumor-specific neoantigens<sup>7</sup>. Meanwhile, presence of DNA repair gene variants did not relate to patient survival. This discrepancy seems to again imply that DNA repair status may have limited utility in SqCC. Variants in MMR genes, however, were associated with a secretory subtype, which interestingly has been defined by enhanced immune response<sup>25</sup>. It is possible that any clinical effect of DNA repair deficiencies would only be uncovered after treatment with immunotherapy. This SqCC cohort did not select for any treatment type, which may be relevant to several results. For example, we speculate that the link between neoantigen load and T cell infiltration might be stronger in a dataset obtained from immunotherapy responders. Similarly, high neoantigen load would be expected to lead to improved survival in those patients, while in these tumors, an increased number of neoantigens may represent more aggressive disease.

Identifying patients who will respond to immunotherapy, and tailoring treatments to their specific tumor biology, will rely on the ability to accurately assess the immune parameters of their tumors. To our knowledge, this work represents the largest analysis of the tumor immune profile in lung SqCC. Our data overall provide evidence that DNA repair pathway variants are closely associated with mutation and neoantigen burden in lung SqCC, and together with other immune-related signals are important factors in determining the tumor immunophenotype. Due to direct relevance of these parameters on the efficacy of immunotherapies, defects in DNA repair should be evaluated further as predictive biomarkers for these rapidly-developing treatments.

## Methods

**Data sets, mutation analysis, immune cell prediction.** A previously annotated cohort of lung squamous cell carcinoma samples ( $n = 178$ ) in The Cancer Genome Atlas (TCGA), as previously published<sup>20</sup>, was obtained from cBioPortal<sup>44</sup>. These data included DNA mutations, RNA-sequencing expression, and clinical descriptors. To identify mutations in homologous recombination, samples were assessed for mutations in the following genes: *ATR*, *ATM*, *CHEK1*, *CHEK2*, *BRCA1*, *BRCA2*, *BAP1*, *BARD1*, *FANCD2*, *FANCE*, *FANCC*, *FANCA*, *RAD50*, *RAD51*, and *PALB2*<sup>45</sup>. The mismatch repair pathway was assessed using the following gene list: *MLH1*, *MLH3*, *MSH2*, *MSH3*, *MSH4*, *MSH5*, *MSH6*, *PMS1*, *PMS2*, *PMS2L3*, *PCNA*, *EXO1*, *POLD1*, *RFC1*, *RFC2*, *RFC3*, *RFC4*, and *RFC5*<sup>45</sup>. All gene names are based on the HUGO Gene Nomenclature Committee (HGNC) database (<https://www.genenames.org/>)<sup>46</sup>. Gene transcripts were obtained using the HGNC-linked NCBI Reference Sequence (RefSeq) identifier<sup>47</sup>, with full data available at <https://www.ncbi.nlm.nih.gov/refseq>. Somatic variants in these genes were then filtered by including only those causing nonsense mutations, splicing errors, indels, or missense mutations predicted to be deleterious based on SIFT<sup>21</sup> ([sift.bii.a-star.edu.sg](http://sift.bii.a-star.edu.sg)) score  $< 0.05$  or CADD v1.4<sup>22</sup> (<https://cadd.gs.washington.edu/>) score  $> 20$ . Splicing errors were defined as a 2-basepair variant in an intron adjacent to the intron/exon junction.

Infiltration of tumor samples by specific immune cell types was performed as previously described<sup>23</sup>. In brief, expression of 812 immune “metagenes,” which were derived from 813 microarrays over 36 studies, was entered into Gene Set Enrichment Analysis (GSEA)<sup>48,49</sup> release 2.2.1. Any immune cell types with a false discovery rate (q-value) of  $\leq 10\%$  were considered as positively infiltrating into that tumor sample.

**Neoantigen prediction.** Neoantigen prediction was done using CloudNeo pipeline (<https://github.com/TheJacksonLaboratory/CloudNeo>)<sup>50</sup> on the NCI Cancer Genomics Cloud (CGC) in the Seven Bridges Genomics implementation. A list of non-synonymous mutations in maf file format was downloaded from TCGA which was converted into the vcf file format (<https://github.com/mskcc/vcf2maf/blob/master/maf2vcf.pl>). The genomic variants were translated into amino acid changes using the Variant\_Effect\_Predictor (release-83) tool<sup>51</sup> and a custom script in the CloudNeo pipeline<sup>50</sup> using the programming language R (version 3.5.1, <https://www.r-project.org>). The output of the custom tool is a list of N-amino-acid-long peptide sequences in a fasta format, such that the single peptide change is in the middle of the N-mer. The HLA prediction was done using the HLAMiner tool version v1.3<sup>52</sup>, which takes HLA allele database file which can be downloaded from the CloudNeo github (<http://cloudneo.readthedocs.io/en/latest/cloudneo/inputs.html>).

Peptides of 9 amino acids containing mutated sites were tested against 6 predicted HLA types using the CloudNeo pipeline in order to generate a neoantigen affinity score using the NetMhcpan (v3.0a) tool<sup>53</sup>. A control analysis was performed with the homologous non-mutated peptides. Neoantigens were identified as mutated peptides with strong binding affinity, as defined by IC50  $< 500$  nM, with positive gene expression and corresponding non-mutated wild-type peptides with weak MHC binding (IC50  $> 500$  nM). Supplementary Fig. S7 displays the representative schematic diagram of the CloudNeo pipeline along with the actual commands that were invoked for a sample to illustrate the parameter settings for various tools within the pipeline. Please note that we have substituted our actual project and sample path with simple strings.

**Statistical analysis.** Comparison of data was carried out using the z-score between two population proportions, unpaired two-tailed Student's or Welch's *t*-test, ANOVA analysis, and the Pearson correlation coefficient as



appropriate. These calculations were performed in GraphPad Prism version 7 (La Jolla, CA). P-values obtained by ANOVA were adjusted for multiple comparisons using Tukey's test. Adjustment for false discovery rate (FDR) was performed using Storey's q-value estimation in R<sup>54</sup>. Logistic regression analysis was also performed using R. Results are presented as percentages, means  $\pm$  standard error of the mean (SEM), Pearson correlation coefficient, or RNA-Seq z-score, as indicated. P-values are represented by \* $p < 0.05$ , \*\* $p < 0.01$ , \*\*\* $p < 0.001$ , and \*\*\*\* $p < 0.0001$ .

## Data Availability

The datasets analyzed during the current study are available in the cBioPortal repository of The Cancer Genome Atlas, found at <http://www.cbioportal.org>, last accessed 10/28/2018. Lists of primary variants for tumors are available from The Cancer Genome Atlas at the National Institutes of Health website, found at [https://tcga-data.nci.nih.gov/docs/publications/lusc\\_2012/](https://tcga-data.nci.nih.gov/docs/publications/lusc_2012/), last accessed 10/26/2018.

## References

1. Reck, M. & Rabe, K. F. Precision Diagnosis and Treatment for Advanced Non-Small-Cell Lung Cancer. *N Engl J Med* **377**, 849–861 (2017).
2. Topalian, S. L., Drake, C. G. & Pardoll, D. M. Immune checkpoint blockade: a common denominator approach to cancer therapy. *Cancer Cell* **27**, 450–461 (2015).
3. Pardoll, D. M. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer* **12**, 252–264 (2012).
4. Sharma, P. & Allison, J. P. The future of immune checkpoint therapy. *Science* **348**, 56–61 (2015).
5. Herbst, R. S. *et al.* Pembrolizumab versus docetaxel for previously treated, PD-L1-positive, advanced non-small-cell lung cancer (KEYNOTE-010): a randomised controlled trial. *The Lancet* **387**, 1540–1550 (2016).
6. Brahmer, J. *et al.* Nivolumab versus Docetaxel in Advanced Squamous-Cell Non-Small-Cell Lung Cancer. *N Engl J Med* **373**, 123–135 (2015).
7. Schumacher, T. N. & Schreiber, R. D. Neoantigens in cancer immunotherapy. *Science* **348**, 69–74 (2015).
8. McGranahan, N. *et al.* Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **351**, 1463–1469 (2016).
9. Snyder, A. *et al.* Genetic basis for clinical response to CTLA-4 blockade in melanoma. *N Engl J Med* **371**, 2189–2199 (2014).
10. Van Allen, E. M. *et al.* Genomic correlates of response to CTLA-4 blockade in metastatic melanoma. *Science* **350**, 207–211 (2015).
11. 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).
12. Rosenberg, J. E. *et al.* Atezolizumab in patients with locally advanced and metastatic urothelial carcinoma who have progressed following treatment with platinum-based chemotherapy: a single-arm, multicentre, phase 2 trial. *The Lancet* **387**, 1909–1920 (2016).
13. Tomlinson, I. P., Novelli, M. R. & Bodmer, W. F. The mutation rate and cancer. *Proc Natl Acad Sci USA* **93**, 14800–14803 (1996).
14. Le, D. T. *et al.* PD-1 Blockade in Tumors with Mismatch-Repair Deficiency. *N Engl J Med* **372**, 2509–2520 (2015).
15. Le, D. T. *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* **357**, 409–413 (2017).
16. Howitt, B. E. *et al.* Association of Polymerase  $\epsilon$ -Mutated and Microsatellite-Instable Endometrial Cancers With Neoantigen Load, Number of Tumor-Infiltrating Lymphocytes, and Expression of PD-1 and PD-L1. *JAMA Oncol* **1**, 1319–1323 (2015).
17. Mehnert, J. M. *et al.* Immune activation and response to pembrolizumab in POLE-mutant endometrial cancer. *J Clin Invest* **126**, 2334–2340 (2016).
18. Strickland, K. C. *et al.* Association and prognostic significance of BRCA1/2-mutation status with neoantigen load, number of tumor-infiltrating lymphocytes and expression of PD-1/PD-L1 in high grade serous ovarian cancer. *Oncotarget* **7**, 13587–13598 (2016).
19. Chae, Y. K. *et al.* Mutations in DNA repair genes are associated with increased neo-antigen load and activated T cell infiltration in lung adenocarcinoma. *Oncotarget* **9**, 7949–7960 (2018).
20. Cancer Genome Atlas Research, N. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
21. Ng, P. C. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812–3814 (2003).
22. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics* **46**, 310–315 (2014).
23. Angelova, M. *et al.* Characterization of the immunophenotypes and antigenomes of colorectal cancers reveals distinct tumor escape mechanisms and novel targets for immunotherapy. *Genome Biol* **16**, 64 (2015).
24. Chen, D. S. & Mellman, I. Elements of cancer immunity and the cancer-immune set point. *Nature* **541**, 321–330 (2017).
25. Wilkerson, M. D. *et al.* Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. *Clin Cancer Res* **16**, 4864–4875 (2010).
26. Goodman, A. M. *et al.* Tumor Mutational Burden as an Independent Predictor of Response to Immunotherapy in Diverse Cancers. *Mol Cancer Ther* **16**, 2598–2608 (2017).
27. Miller, A. *et al.* High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood Cancer J* **7**, e612 (2017).
28. Yadav, M. *et al.* Predicting immunogenic tumour mutations by combining mass spectrometry and exome sequencing. *Nature* **515**, 572–576 (2014).
29. Hellmann, M. D. *et al.* Nivolumab plus Ipilimumab in Lung Cancer with a High Tumor Mutational Burden. *N Engl J Med* **378**, 2093–2104 (2018).
30. Rizvi, H. *et al.* Molecular Determinants of Response to Anti-Programmed Cell Death (PD)-1 and Anti-Programmed Death-Ligand 1 (PD-L1) Blockade in Patients With Non-Small-Cell Lung Cancer Profiled With Targeted Next-Generation Sequencing. *J Clin Oncol* **36**, 633–641 (2018).
31. Galon, J. *et al.* Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. *Science* **313**, 1960–1964 (2006).
32. 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).
33. Tokunaga, R. *et al.* CXCL9, CXCL10, CXCL11/CXCR3 axis for immune activation - A target for novel cancer therapy. *Cancer Treat Rev* **63**, 40–47 (2017).
34. Church, S. E., Jensen, S. M., Antony, P. A., Restifo, N. P. & Fox, B. A. Tumor-specific CD4<sup>+</sup> T cells maintain effector and memory tumor-specific CD8<sup>+</sup> T cells. *Eur J Immunol* **44**, 69–79 (2014).
35. Tran Janco, J. M., Lamichane, P., Karyampudi, L. & Knutson, K. L. Tumor-infiltrating dendritic cells in cancer pathogenesis. *J Immunol* **194**, 2985–2991 (2015).
36. di Bari, M. G. *et al.* TGF-beta modulates the functionality of tumor-infiltrating CD8<sup>+</sup> T cells through effects on TCR signaling and Spred1 expression. *Cancer Immunol Immunother* **58**, 1809–1818 (2009).
37. Massague, J. TGFbeta in Cancer. *Cell* **134**, 215–230 (2008).

38. Spranger, S., Bao, R. & Gajewski, T. F. Melanoma-intrinsic beta-catenin signalling prevents anti-tumour immunity. *Nature* **523**, 231–235 (2015).
39. Stewart, D. J. Wnt signaling pathway in non-small cell lung cancer. *J Natl Cancer Inst* **106**, djt356 (2014).
40. Cabrera, C. M. *et al.* Total loss of MHC class I in colorectal tumors can be explained by two molecular pathways: beta2-microglobulin inactivation in MSI-positive tumors and LMP7/TAP2 downregulation in MSI-negative tumors. *Tissue Antigens* **61**, 211–219 (2003).
41. Restifo, N. P. *et al.* Loss of functional beta 2-microglobulin in metastatic melanomas from five patients receiving immunotherapy. *J Natl Cancer Inst* **88**, 100–108 (1996).
42. Boland, C. R. & Goel, A. Microsatellite instability in colorectal cancer. *Gastroenterology* **138**, 2073–2087 e2073 (2010).
43. Hanna, N. *et al.* Randomized phase III trial of pemetrexed versus docetaxel in patients with non-small-cell lung cancer previously treated with chemotherapy. *J Clin Oncol* **22**, 1589–1597 (2004).
44. Cerami, E. *et al.* The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. *Cancer Discov* **2**, 401–404 (2012).
45. Chae, Y. K. *et al.* Genomic landscape of DNA repair genes in cancer. *Oncotarget* **7**, 23312–23321 (2016).
46. Yates, B. *et al.* Genenames.org: the HGNC and VGNC resources in 2017. *Nucleic Acids Res* **45**, D619–D625 (2017).
47. O’Leary, N. A. *et al.* Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* **44**, D733–745 (2016).
48. Subramanian, A. *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America* **102**, 15545–15550 (2005).
49. Mootha, V. K. *et al.* PGC-1alpha-responsive genes involved in oxidative phosphorylation are coordinately downregulated in human diabetes. *Nature genetics* **34**, 267–273 (2003).
50. Bais, P., Namburi, S., Gatti, D. M., Zhang, X. & Chuang, J. H. CloudNeo: a cloud pipeline for identifying patient-specific tumor neoantigens. *Bioinformatics* **33**, 3110–3112 (2017).
51. McLaren, W. *et al.* Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics* **26**, 2069–2070 (2010).
52. Warren, R. L. *et al.* Derivation of HLA types from shotgun sequence datasets. *Genome Med* **4**, 95 (2012).
53. Andreatta, M. & Nielsen, M. Gapped sequence alignment using artificial neural networks: application to the MHC class I system. *Bioinformatics* **32**, 511–517 (2016).
54. Storey, J. D., Bass, A. J., Dabney, A. & Robinson, D. *qvalue: Q-value estimation for false discovery rate control*, <http://github.com/jstorey/qvalue> (2015).

## Acknowledgements

JHC was supported by the National Cancer Institute of the NIH under award R21CA191848 and supplement R21CA191848-01A1S1. Research was also partially supported by the National Cancer Institute under award P30CA034196.

## Author Contributions

Y.K.C. formulated hypotheses, designed experiments, and aided in manuscript review. J.F.A. designed experiments, performed data analyses, and aided in manuscript review. M.S.O. performed data analyses and manuscript writing. P.B. and S.N. performed neo-antigen predictions. S.A. aided in data analyses. F.J.G. aided in manuscript review. J.H.C. designed experiments, performed neo-antigen predictions, and aided in manuscript review.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-019-39594-4>.

**Competing Interests:** The authors declare no competing interests.

**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) 2019