



This is a repository copy of *Somatic alteration burden involving non-cancer genes predicts prognosis in early-stage non-small cell lung cancer.*

White Rose Research Online URL for this paper:
<http://eprints.whiterose.ac.uk/148909/>

Version: Published Version

Article:

Wang, D. orcid.org/0000-0003-0068-1005, Pham, N.-A., Freeman, T.M. et al. (12 more authors) (2019) Somatic alteration burden involving non-cancer genes predicts prognosis in early-stage non-small cell lung cancer. *Cancers*, 11 (7). 1009. ISSN 2072-6694

<https://doi.org/10.3390/cancers11071009>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
<https://creativecommons.org/licenses/>

Takedown






If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

Article

Somatic Alteration Burden Involving Non-Cancer Genes Predicts Prognosis in Early-Stage Non-Small Cell Lung Cancer

Dennis Wang ^{1,2,*} , Nhu-An Pham ¹, Timothy M. Freeman ² , Vibha Raghavan ¹, Roya Navab ¹, Jonathan Chang ¹ , Chang-Qi Zhu ¹, Dalam Ly ³ , Jiefei Tong ⁴, Bradly G. Wouters ^{1,5}, Melania Pintilie ¹, Michael F. Moran ⁴ , Geoffrey Liu ^{1,5,6}, Frances A. Shepherd ^{1,6} and Ming-Sound Tsao ^{1,5,7,*}

¹ Princess Margaret Cancer Centre, University Health Network, Toronto, ON M5G 2M9, Canada

² NIHR Sheffield Biomedical Research Centre, University of Sheffield, Sheffield S10 2HQ, UK

³ Toronto General Research Institute, University Health Network, Toronto, ON M5G 2C4, Canada

⁴ Program in Molecular Structure and Function, Hospital for Sick Children, Toronto, ON M5G 0A4, Canada

⁵ Department of Medical Biophysics, University of Toronto, Toronto, ON M5G 1L7, Canada

⁶ Department of Medicine, University of Toronto, Toronto, ON M5G 2C4, Canada

⁷ Department of Laboratory Medicine and Pathobiology, University of Toronto, Toronto, ON M5S 1A1, Canada

* Correspondence: dennis.wang@sheffield.ac.uk (D.W.); ming.tsao@uhn.ca (M.-S.T.)

Received: 29 June 2019; Accepted: 15 July 2019; Published: 19 July 2019



Abstract: The burden of somatic mutations and neoantigens has been associated with improved survival in cancer treated with immunotherapies, especially non-small cell lung cancer (NSCLC). However, there is uncertainty about their effect on outcome in early-stage untreated cases. We posited that the burden of mutations in a specific set of genes may also contribute to the prognosis of early NSCLC patients. From a small cohort of 36 NSCLC cases, we were able to identify somatic mutations and copy number alterations in 865 genes that contributed to patient overall survival. Simply, the number of altered genes (NAG) among these 865 genes was associated with longer disease-free survival (hazard ratio (HR) = 0.153, $p = 1.48 \times 10^{-4}$). The gene expression signature distinguishing patients with high/low NAG was also prognostic in three independent datasets. Patients with a high NAG could be further stratified based on the presence of immunogenic mutations, revealing a further subgroup of stage I NSCLC with even better prognosis (85% with >5 years survival), and associated with cytotoxic T-cell expression. Importantly, 95% of the highly-altered genes lacked direct relation to cancer, but were implicated in pathways regulating cell proliferation, motility and immune response.

Keywords: cancer genomics; prognosis; mutation burden; copy number; patient stratification; cancer immunology; oncogenic pathways; patient-derived xenograft

1. Introduction

There have been extensive studies evaluating prognostic markers for early stage non-small cell lung cancer (NSCLC) patients. These studies included genes/proteins that are overexpressed in tumor compared to non-tumor lung tissue, mutations in specific “driver” oncogenes and tumor suppressor genes, and gene expression profiles [1–3]. To date, no single gene/protein has been shown consistently to be prognostic especially in large multi-institutional patient cohorts [4–7]. While many prognostic mRNA signatures have been identified by various investigators, few of these signatures have provided meaningful insight into the biological mechanisms that drive their prognostic significance [8,9]. More recently, whole genome and targeted sequencing studies have observed a relationship between the mutational load of tumors and prognosis in NSCLC [10–14]. However, these

studies have yet to show consistent association between mutation burden and survival across multiple cohorts. Since some non-synonymous mutations generated neoantigens that induced an anti-tumor immune response [15–18], they have been reported to predict improved progression-free survival in NSCLC patients treated with immunotherapies [19–21]. However, these effects of mutation burden and neoantigens are not well characterized in early-stage NSCLC cases where immunotherapies are currently not given.

We reported previously the genomic, transcriptomic and proteomic profiles of 36 NSCLC patients from their primary tumors and established xenografts [22–25]. Given NSCLC tumors are genomically heterogeneous [26], we postulated that mutations found in both primary and xenograft tumors are more likely to correlate with patient outcomes. We applied an integrative genomics approach to identify prognostic somatic alterations in the 36 NSCLC patients and validated their predictive performance in multiple, much larger and independent patient cohorts (Figure 1). These results provide new insights into genes previously not implicated in carcinogenesis itself, but that have an impact on patient outcome.

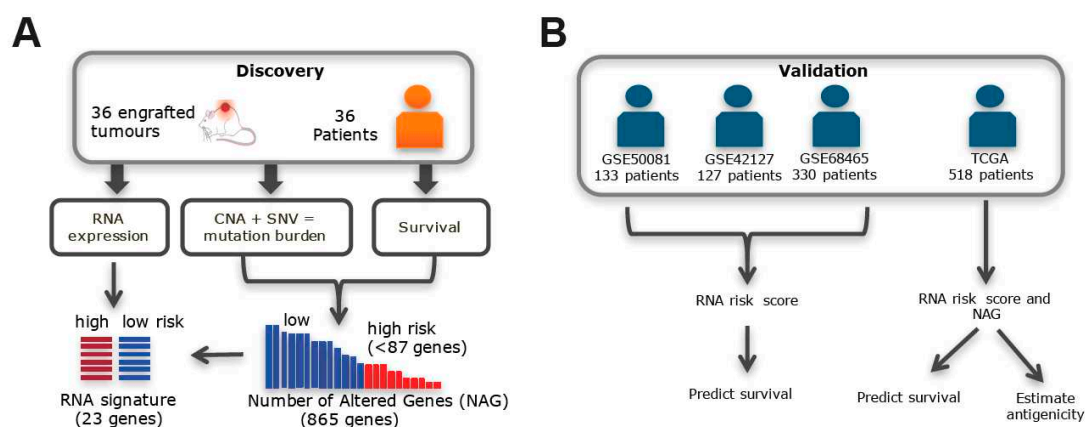


Figure 1. Translation of genomic information from patient-derived xenografts (PDXs) to predict patient outcome. **(A)** Somatic copy number aberrations (CNA), single nucleotide variants (SNV) and gene expression (RNA) were profiled in PDXs. The number of alterations within 865 genes (NAG) and the associated 23 genes’ expression signature were selected using the survival of patients corresponding to the PDXs. **(B)** The expression signature was used to calculate a risk score and predict overall survival (OS) in three independent cohorts of NSCLC patients with data accessioned in the Gene Expression Omnibus web-based data repository (GSE50081, GSE42127 and GSE68465). Survival prediction was also tested for 518 NSCLC stage I patients from The Cancer Genome Atlas (TCGA), by using both the expression signature and NAG. The mutations found in the 865 gene panel of the TCGA samples were also assessed for antigenicity.

2. Results

2.1. Somatic Alteration Burden in PDX Is Correlated with Patient Survival

A total of 164–765 genes were characterized with recurrent somatic SNVs or copy number aberration (CNA) in 36 NSCLC patients (Figure 2A). To investigate the importance of these alterations, we correlated their burden to the corresponding patient survival. The comparison of disease-free survival (DFS) and overall survival (OS) of patients with these tumors with ≤ 343 alterations (first quartile) to patients corresponding to patient-derived xenografts (PDXs) with >343 alterations revealed that the group with the higher number of alterations had significantly better OS (HR = 0.366, 95% CI = 0.140–0.952, $p = 0.040$) and DFS (hazard ratio (HR) = 0.294, 95% CI = 0.103–0.840, $p = 0.022$). The comparison of CNAs alone with DFS showed a weaker and statistically non-significant association (HR = 0.546, 95% CI = 0.185–1.61, $p = 0.272$); this also applies to OS (HR = 0.532, 95% CI = 0.197–1.43, $p = 0.202$). Similarly, the number of somatic mutations alone was not significantly associated with DFS (HR = 0.568, 95% CI = 0.179–1.79, $p = 0.335$) or OS (HR = 1.41, 95% CI = 0.407–4.76, $p = 0.592$).

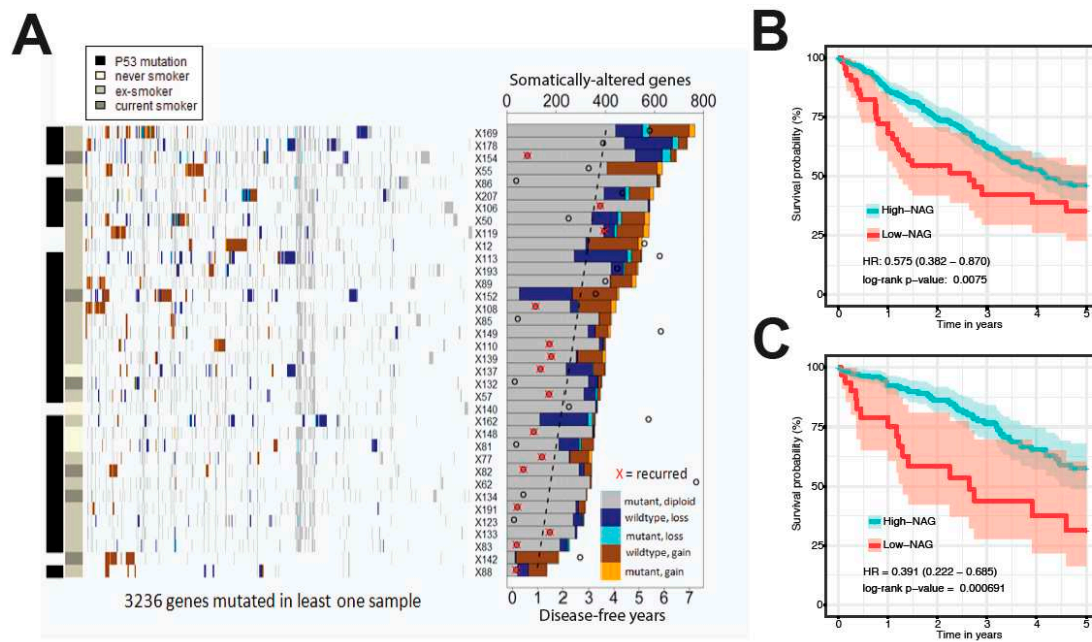


Figure 2. Association between somatic alterations found in PDXs and OS of patients. (A) A heatmap of SNVs and CNAs across PDX models in genes that have been altered at least once. The bar graph describes the total number of altered genes in each PDX (labeled by their ID preceded by “X”). Different combinations of SNV point mutations and CNAs are described by the side bar graph. Circles on the bar graph show the disease-free years to last follow-up, and the red “X” denotes the recurrence of disease for each PDX’s matched patient, while the dashed line shows the best fit running through them. Genes were ordered by hierarchical clustering across the samples. Kaplan–Meier curves show significant differences, as evidenced by the hazard ratio (HR), its confidence interval and log-rank test p -value, in the overall survival (OS) between high and low NAG in (B) all TCGA NSCLC patients and (C) only in stage I patients.

Across the 36 tumors, a total of 3236 unique genes with somatic alterations detected in more than a single patient were used to perform a penalized regression that was most associated with OS. The regression model highlighted 865 genes with a non-zero contribution to the overall risk of death (Supplementary Table S1). Remarkably, only 47 of these genes (5.4%) were identified in the Cancer Gene Census [27]. The 36 patients were then divided into two groups based on the number of altered genes (NAG) among the 865 genes. Patients with a high NAG had better survival (HR = 0.153, 95% CI = 0.051–0.459, $p = 1.48 \times 10^{-4}$) than patients who had a low NAG (Supplementary Figure S1A) and their associated clinical-pathological features (summarized in Supplementary Table S2 and individually listed in Supplementary Table S3).

2.2. Validation Using the Gene Expression Signature Associated with Somatic Alteration Burden

We hypothesized that the prognostic differences of NAG may be recapitulated in the differentially-expressed gene signature between the high and low NAG groups, and this may be used as a proxy of NAG for further validation in publicly-available independent patient cohorts accessioned in the Gene Expression Omnibus web-based data repository (GSE42127, GSE50081 and GSE68465 described in Supplementary Table S2). We identified 79 genes with differentially-expressed mRNA (>2 absolute fold change, $p < 0.05$) between the two groups in our cohort of 36 patients based on the NAG (Supplementary Table S4). Using mRNA expression profiles of the 79 genes from 127 NSCLC patients in GSE42127, the expression scores associated with survival were calculated. Of the 79 genes, only 23 genes had non-zero coefficients when we used penalized regression for fitting to OS in this cohort. The resulting NAG expression signature was significantly prognostic for three independent early-stage cohorts with mRNA expression data (GSE42127 HR 0.23, 95% CI = 0.125–0.422, $p = 2.3 \times 10^{-7}$; GSE50081 HR =

0.562, 95% CI = 0.332–0.952, $p = 0.029$; GSE68465 HR = 0.518, 95% CI = 0.338–0.794, $p = 2.2 \times 10^{-3}$ (Supplementary Figure S1B–D)).

2.3. Validation in TCGA Datasets

To test the alteration burden together with the expression signature as a prognostic classifier for NSCLC, we attempted to predict the overall survival of 221 lung adenocarcinoma (LUAD) and 173 lung squamous cell carcinoma (LUSC) cases from TCGA. We counted the NAG among the 865 genes and calculated the mRNA expression risk score for each TCGA patient tumor, then classified the patients using the same cut-offs that were used to classify our cohort of 36 patients. All patients with a high NAG (at least 87 altered genes) or a low expression risk score had a relatively better prognosis (Figure 2B; HR = 0.575, 95% CI = 0.382–0.870, $p = 0.0075$); this also held true for stage I patients (Figure 2C; HR = 0.391, 95% CI = 0.222–0.685, $p = 6.9 \times 10^{-4}$). Multivariate survival analyses of stage I patients adjusted for age, sex, smoking history and histology confirmed that the number of alterations was an independent predictor of good OS (HR = 0.407, 95% CI = 0.211–0.787, $p = 7.7 \times 10^{-3}$, Table 1).

Table 1. Multivariate survival model of NAG score with clinical-pathological factors of TCGA NSCLC stage I patients.

Variation	HR	95% CI	Wald Test p -Value
Age (>65 vs. ≤65)	1.04	0.62–1.73	0.89
Sex (F vs. M)	0.85	0.51–1.43	0.54
Tobacco (smoker vs. never)	1.75	0.74–4.12	0.2
Histology (adeno vs. squamous)	0.96	0.52–1.79	0.91
Overall NAG score (low vs. high burden)	2.46	1.27–4.75	0.0077 *

NAG, number of altered genes; HR, hazard ratio; CI, confidence interval; * designates significance at $p \leq 0.05$.

2.4. Immunogenic Mutations Correlate with the Best Survival and Cytotoxic T-Cell Signature

Given that alteration burden is associated with good survival outcome, we further hypothesized that some of the mutations in the 865 genes may induce an immune response towards the tumor through the expression of immunogenic peptides or neoantigens. We examined whether the somatic variants within the 865 genes found in TCGA NSCLC patients could encode for tumor-specific antigens presented by MHC Class I molecules. A subset of 86 patients had at least one expressed mutated peptide with high MHC I binding affinity against both HLA-A alleles, and this immunogenic group had significantly better OS (Supplementary Figure S2A; HR = 0.536, CI = 0.341–0.8419, $p = 0.0068$). Among NSCLC stage I patients, 47 were classified as having immunogenic mutations and had significantly better OS (Figure 3A; HR = 0.266, CI = 0.1068–0.6619, $p = 0.0044$), even after adjusting for other clinical factors (Table 2). In comparison, patients stratified using all possible neoantigens, beyond those in the panel of 865 genes, had no significant difference ($p < 0.01$) in overall survival (Supplementary Table S5). By combining this immunogenicity classifier with the alteration burden based on NAG, we were able to identify patients with extremely good prognosis. Stage I patients with both immunogenic and high NAG tumors had significantly better survival than patients with either low or high NAG alone (Figure 3B; HR = 0.0807, CI = 0.02315–0.2813, $p = 7.8 \times 10^{-5}$).

The RNA expression profiles from these two groups of patients were analysed further for immune components. Immune cell population estimates from TIMER [28] revealed moderately, but not significantly higher proportions of B-cells, dendritic cells, and CD8 T-cells in high NAG patient tumors with immunogenic peptides compared to those without neoantigens (Figure 3C). Focusing specifically on markers of cytotoxic T-cells in Figure 3D, we also found in high NAG tumors with immunogenic peptides significantly higher ($p < 0.05$) expression of CD8A, granzyme B (GZMB) and perforin (PRF1). Previously, cytokines CCL5, CXCL9, CXCL10 and IL16 were found to be associated with cytotoxic T-cells [29]. We found CCL5, CXCL9 and CXCL10 to be significantly more highly expressed in our group with immunogenic peptides. In contrast, there were no consistent differences in immune cell

populations or cytotoxic T-cell markers between high and low NAG patients without immunogenic peptides (Supplementary Figure S2B,C).

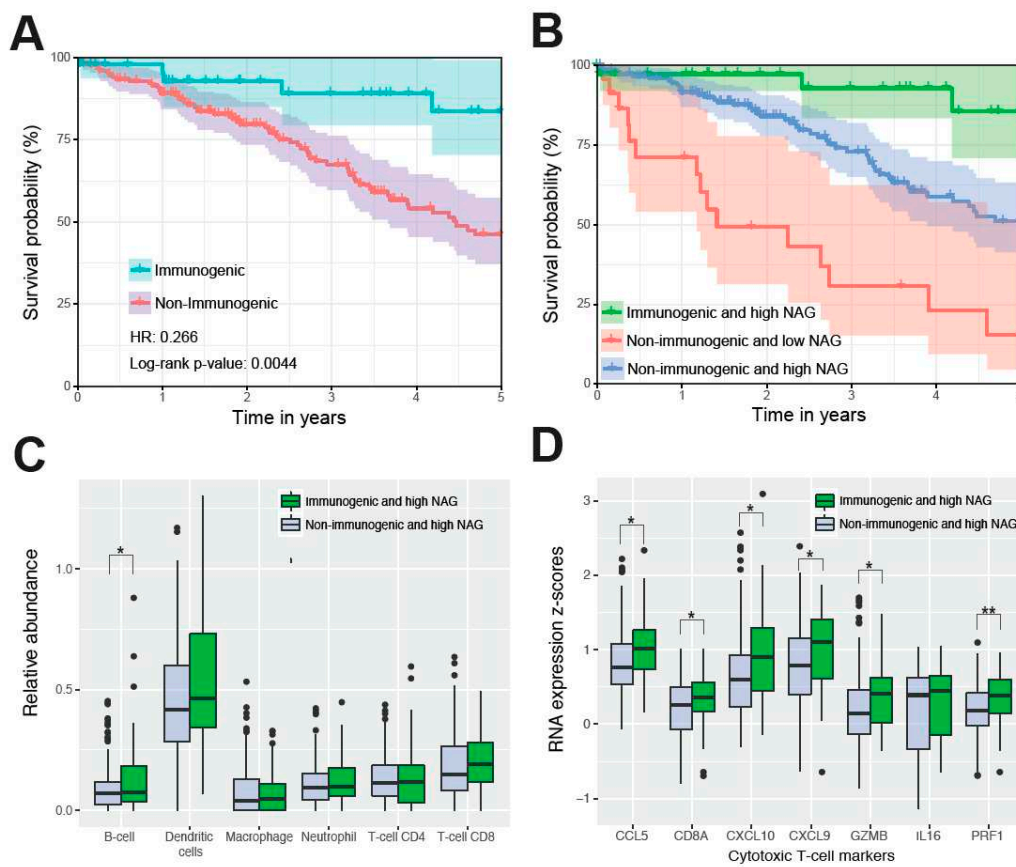


Figure 3. TCGA NSCLC patients stratified by immunogenic neoantigens. (A) Stage I patients (47 with neoantigens, 171 without neoantigens) are grouped into those with immunogenic neoantigens and those with none. Hazard ratios and log-rank *p*-values compare immunogenic patient tumors to non-immunogenic patient tumors. (B) The overall survival differences of stage I patients were classified based on the presence of immunogenic neoantigens and the number of altered genes. Immunogenic and high-NAG patient tumors (green, 37 patients) vs. non-immunogenic and low-NAG patient tumors (red, 23 patients) show HR = 0.0807, *p* = 7.8×10^{-5} . Immunogenic and high-NAG tumors (green) vs. non-immunogenic and high-NAG patient tumors (blue, 147 patients) show HR = 0.229, *p* = 0.013. Non-immunogenic and high-NAG tumors (blue) vs. non-immunogenic and low-NAG patient tumors (red) show HR = 0.309, *p* = 7.9×10^{-5} . High NAG patients with and without neoantigens were contrasted based on the relative abundance of immune cell types estimated by the TIMER algorithm (C), and the RNA expression of cytotoxic T-cell markers (D). Significant differences for each component are marked (*t*-test *p*-value * <0.05; ** <0.01).

Table 2. Multivariate survival model of the immunogenicity factor with clinical-pathological factors of TCGA NSCLC stage I patients.

Variation	HR	95% CI	Wald Test <i>p</i> -Value
Age of diagnosis (≤65 y vs. >65 y)	0.929	0.545–1.583	0.7865
Gender (male vs. female)	0.871	0.524–1.447	0.5929
Smoking history (last 15 y; yes vs. no)	0.751	0.434–1.302	0.3085
Histology (adeno vs. squamous cell)	0.677	0.392–1.169	0.1615
Immunogenicity (≥1 neoantigen vs. 0 neoantigens)	0.296	0.119–0.740	0.00919 *

* designates significance at *p* ≤ 0.05.

2.5. Prognostic Alterations Enriched in Genes Involved in Extracellular Signaling

To investigate the alternate or additional mechanisms by which alteration burden might be impeding cancer progression, pathway enrichment analyses were performed on the 865 genes. The most enriched pathways involved BARD1 signaling, Janus kinase activity, integrin signaling and extracellular matrix interactions (Figure 4A; hyper-geometric test $p < 0.01$). Within these pathways, a number of genes were altered more frequently across TCGA cases in the group of low-risk patients with better survival (Figure 4B). Inactivating alterations, which deactivate Wnt signaling, were most frequent in *CTNNB1* (23.1%) and *WNT5A* (22.8%), along with copy gains in the tumor suppressors *NHERF1* (16.4%) and *NF2* (13.1%). The high frequency of copy gains occurring in *ITGA10* (20.9%), *COL20A1* (17.9%) and *COL8A1* (26.5%) suggested that cell adhesion to the extracellular matrix may be affected. These highly-altered genes and pathways have been known to control specific cellular functions such as proliferation, motility and adhesion.

In order to evaluate the effect of these pathway alterations on cell functionality, we selected three NSCLC cell lines (H1573, A549, HCC827) with different levels of NAG and expression risk scores (Figure 5A). The three cell lines had different patterns of expression for the NAG genes we found frequently altered in TCGA cases (Figure 5B). Cell line H1573 had 45 NAG, while A549 and HCC827 had 13 and 18, respectively. HCC827 had the fastest rate of proliferation over 80 h of growth and the highest number of invading cells across a Matrigel membrane after 48 h (Figure 5C,D). In contrast, H1573, which had the highest NAG, had a much slower rate of proliferation and invasion ($p = 0.0145$). Furthermore, HCC827 moved the fastest in a 24-h motility assay compared to the other two cell lines (Figure 5E).

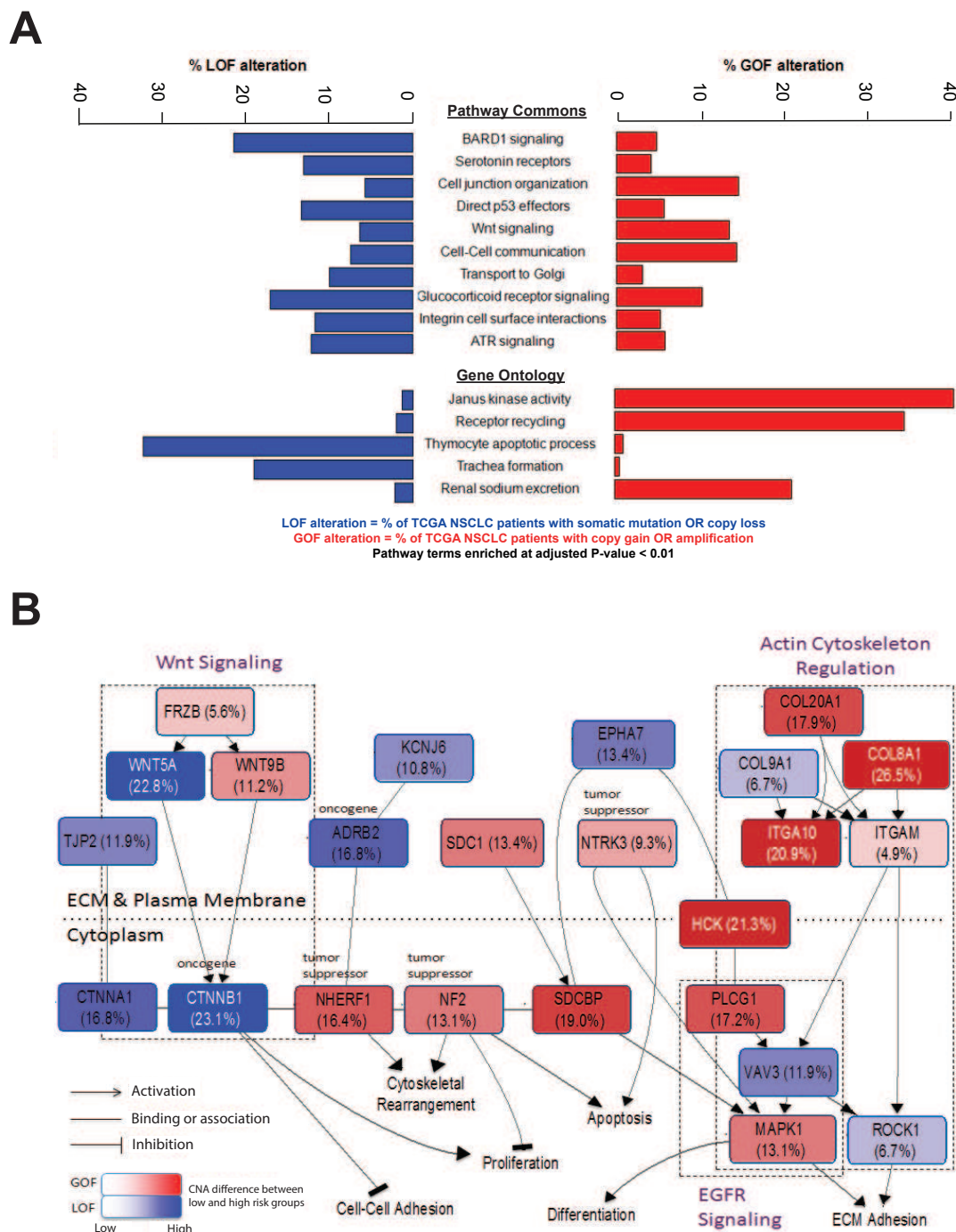


Figure 4. Enriched pathways and biological processes for genes in NAG. (A) Pathways in Pathway Commons and Gene Ontology were enriched for NAG genes. Bar plots show the proportions of NAG genes within each pathway that were classified as gain of function (GOF) or loss of function (LOF) alterations in TCGA cases. (B) A diagram of enriched pathways identified from NAG and their associated expression signature genes. The main pathways that emerged included Wnt signaling, integrin signaling and actin cytoskeleton regulation. Blue nodes identify copy number loss or inactivating mutation indicating loss of function (LOF), and red nodes indicate copy number gain indicating gain of function (GOF). Percentages in brackets represent the difference in alteration frequency between patients of low and high mortality risk, respectively, as classified by the NAG and the NAG expression score for TCGA NSCLC tumors. The color intensity reflects the proportion of tumors with the gene alteration. ECM = extracellular matrix.

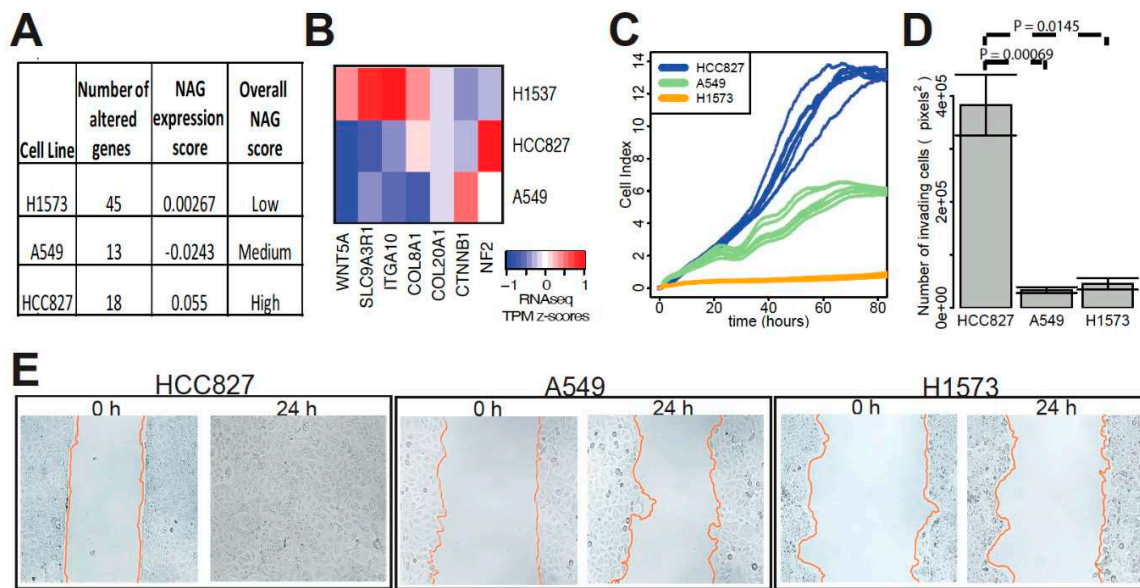


Figure 5. Functional impact of altered and differentially-expressed genes on cancer cells. **(A)** The number of genes (out of 865 genes) with somatic SNV or CNA was tallied for routinely-used lung cancer cell lines. Three lung cancer cell lines with differing overall NAG (based on quartiles) and different risk scores for NAG associated expression signature were compared to each other in functional assays. **(B)** A heatmap of RNA expression level scaled across cell lines for NAG genes commonly altered in TCGA is shown for the three lung cancer cell lines. **(C)** Proliferation of the three cultured cell lines was estimated by electrical impedance (cell index) over time. **(D)** Invasion of cells across trans-well membranes coated with Matrigel was counted. **(E)** Cell motility was measured by a time-lapse wound-healing assay. The rates of wound area travelled by cells were 15.42 $\mu\text{m}/\text{hr}$ for HCC827, 2.64 $\mu\text{m}/\text{h}$ for A549 and 1.98 $\mu\text{m}/\text{h}$ for H1573.

3. Discussion

We have shown that the number of somatic alterations among 865 genes is associated with better prognosis in early stage NSCLC patients. These prognostic alterations included not only non-synonymous mutations, but also copy number alterations, and they have a corresponding gene expression signature derived from differentially-expressed genes between high and low NAGs validated in three cohorts of early stage NSCLC. For stage I TCGA patients, we further showed that the integrative NAG score was able to classify patients into three distinct subpopulations with low NAG, high NAG and high NAG with neoantigens. These subpopulations were distributed equally in their frequency of clinical features (Supplementary Table S6). Pathway analysis of the panel of 865 genes revealed that the majority of the mutations had not been implicated in driving cancer pathogenesis, but were associated with processes relating to immune activation, tumor cell–extracellular matrix interactions and cell motility, which were supported by cell line experiments. These findings provide evidence that high-level somatic genomic aberrations involving non-cancer-associated genes may confer better prognosis through the generation of neoantigens, impairment of cellular functions involved in tumor cells and stroma interaction and motility.

Few recent investigations of mutation or neoantigen burden in early stage or untreated lung cancers provided contradictory evidence for the directionality of the relationship between burden and survival [10–12]. These studies relied on measuring burden by counting all point mutations across genomes and, thus, were sensitive to mutation calling error (Supplementary Table S7). In contrast, our burden measure of counting the number of genes with at least one alteration (point mutation, CNA, neoantigen or RNA expression) may be more robust to variant calling errors and variation in the number of variants within each gene.

Smaller targeted gene panels recently have been shown to be clinically effective at measuring overall tumor mutation burden in mostly known cancer driver genes [10,15]. Our observations indicate that mutation burden in non-driver genes may also have an effect on patient outcome. The mechanism by which alteration burden in gene sets affects cancer progression and patient survival remains unclear, but it has been hypothesized that excessive genomic instability can lead to deleterious mutations and gene losses that impede tumor growth [30,31]. Greater DNA damage, especially due to dysregulation of DNA repair mechanisms, could result in deleterious mutations or copy number losses in many of the 865 genes that are associated with cell adhesion, motility and integrin signaling, cellular functions that are likely also essential in the phenotypic manifestation of malignant tumor cells [32,33]. Integrin signaling is a major pathway contributing to cancer cell survival and has been a target for antagonists to inhibit tumor growth [34]. Therefore, the occurrence of copy loss alterations in integrin pathway genes may mitigate tumor cell malignancy and is thus a marker for less aggressive tumors in patients with good survival outcomes.

Somatic mutation burden has been correlated with neoantigen load and T-cell infiltration in NSCLC and melanoma [14,35–37], yet there is a lack of evidence for this relationship in early-stage cancers, which are predominantly untreated with therapeutics. Our assessment of neoantigen load among the 865 genes is the first to show that even in stage I NSCLC cancer without immunotherapy, the presence of neoantigens and high mutation burden is correlated with T-cell expression. Previously, mutation burden and neoantigen load were used to identify patients with good prognosis following immunotherapy treatment, such as immune checkpoint inhibitors, but there has been debate about whether both markers are needed given that they select for a similar subgroup of patients [14,15,38]. Even though all the patients we identified with immunogenic mutations (neoantigens) were also classified as high alteration burden and as having long OS (>85% surviving longer than five years), we identified a unique subgroup of patients with low alteration burden and no immunogenic mutations that had far worse OS (median survival of two years). In contrast to previous reports about the clonality of neoantigens that are prognostic [14,16], the immunogenic mutations we used for patient stratification did not show evidence of being subclonal compared to other mutations based on their variant allele fractions (Supplementary Figure S3). Our criteria for immunogenicity prediction may have been overly stringent, resulting in far fewer neoantigens found in each patient compared to other studies and may have led to our approach missing some clonal or subclonal neoantigens.

4. Materials and Methods

4.1. Patient and NSCLC PDX for Genomics Profiling

This study involved 36 patients [22–24,39] whose NSCLC tumors were surgically resected and successfully engrafted as PDX models in non-obese diabetic (NOD) severe immune deficient (SCID) mice. Short indels and single nucleotide variants (SNV) were identified from whole exome sequencing in the primary and PDX tumors. Copy number alterations were profiled using the HumanOmni 2.5 BeadChip SNP array and processed in accordance with previous reports [22–24,39]. The average log likelihood ratio (LRR) for each gene was used to determine copy gains ($LRR > 0.5$) and copy loss ($LRR < -0.5$).

4.2. TCGA Patients

Published mutation, copy number, RNA-seq and clinical data were downloaded for LUAD [40] and LUSC [41] patients on cBioPortal, using the cgdsR R package [42], and clinical data were downloaded using the FireBrowseR R package [43]. Copy number gains and losses were determined using an LRR cut-off of ± 0.5 . RNA expression data in the RNA-seq by expectation maximization (RSEM) scale from RNA-seq were transformed by the asinh function for survival analysis.

4.3. Stratification by High-Level Mutation Burden

To identify genomic alterations associated with patient survival from somatic profiles, we first defined a gene as “altered” if there was a somatic non-synonymous mutation, copy number gain or copy number loss. Genes altered only in one patient tumor were removed. The resulting alteration profiles across patients were fitted to the overall survival (OS, up to five years) of the corresponding patients. ElasticNet from the R package “glmnet” Version 1.9-8 was used for the regression fit with alpha at 0.1 and the best fitting model selected from the maximum of the deviance ratio. In the ElasticNet model, 865 genes had non-zero coefficients.

The number of somatic alterations among 865 genes (NAG) was used as the risk score for each TCGA patient. The risk scores were dichotomized at the first quartile to assign patients to one of the prognostic arms (high vs. low number of alterations). The proportions of DFS and OS were calculated using the Kaplan–Meier method, and the difference between curves was tested using the log-rank test. The OS is the time between the date of diagnosis to the date of death or last follow-up, and the DFS is the date of diagnosis to the date of death, or relapse or last follow-up. All reported hazard ratios (HR) scores and *p*-values used the Kaplan–Meier method unless reported as multivariate analyses. A Cox proportional hazards model was used to fit survival times to the number of altered genes while adjusting for other clinical factors including age, stage and smoking status. The significance for the Cox proportional hazards model was based on the Wald test. The R code for conducting this validation test on TCGA datasets can be found at https://github.com/TransAnalytics/cancer_prog_prediction.

4.4. Stratification by Gene Expression Profiles Corresponding to NAGs

The mRNA expression profiles in the 36 NSCLC tumors were measured as previously described and accessioned in the Gene Expression Omnibus (GEO) data repository (GSE68929) [22–24,39]. 79 differentially-expressed genes (absolute fold change >2 and *t*-test *p*-value <0.05) between the high and low number of genomic alterations were identified. The expression profiles of these genes from the UT Lung Spore dataset (GSE42127) [8] were fitted to the OS using ElasticNet. ElasticNet’s lambda parameter was selected from 5-fold cross-validated likelihood, and the alpha parameter was set at 0.1. The coefficients from the ElasticNet model, which formed the 23-gene prognostic classifier, were summed to create an expression risk score for each patient. An expression risk score ≥ 0.17 was associated with low NAGs, whereas an expression risk score <0.17 was associated with high NAGs. The cut-off at 0.17 was at the third quartile of risk scores, which was consistent with the first quartile cut-off for NAG. All reported survival differences were tested using the Wald test within the Cox proportional hazards model, and follow-up of patients was measured up to 5 years. Validation of NAG expression risk scores was performed on the Director’s Challenge Consortium for Molecular Classification of Lung Adenocarcinoma dataset with a total of 442 samples (GSE68465) [1] and the UHN181 cohort of 181 NSCLC patients (GSE50081) [44].

4.5. Estimating Immunogenicity

Non-synonymous mutations causing amino acid changes were identified among the 865-gene panel for TCGA patients. Protein sequences were acquired from the UniProt reviewed canonical human proteome UP000005640 FASTA file. Windows of 8–11 amino acids in length were derived by applying non-synonymous SNVs to their respective protein sequence and using all 8–11mer amino acid peptides containing the altered amino acid. The 8–11mer altered peptides for each patient were input into NetMHCpan v3.0 [45] along with both of the patient’s supertyped HLA-A alleles to calculate their predicted HLA-A binding affinity values for each altered peptide. The 4-digit HLA types for the patients were sourced from The Cancer Immunome Atlas [46], which used Optitype [47] to call the HLA-A alleles from RNA-seq FASTQ files. The HLA types were supertyped into standard categories prior to carrying out the HLA-A binding affinity prediction analysis [48].

We defined peptides as antigenic if they had an HLA-A binding affinity below 500 nM with both HLA-A alleles. In addition, antigenic peptides' genes had to be expressed above the median expression level for that gene calculated across all patients in order for the peptides to be considered immunogenic, and the patient HLA-A gene needed to be expressed above the median HLA-A expression level across all patients. Patients were considered to have immunogenic peptides that could potentially trigger an immune response by fulfilling these three conditions: strong binding affinity, high expression of the antigenic peptide to ensure they are clonal and high expression of the MHC-I receptor. Patients were stratified based on whether or not they had above the median number of immunogenic peptides (1 or more) as outlined above, either alone, or in combination with the NAG score. A Cox proportional hazards model was used to calculate the effect size/risk factor associated with the presence/absence of these features, along with the *p*-values for the differences between patient groups.

4.6. Pathway Enrichment

Gene set enrichment was performed on the 865 genes using the database for annotation, visualization and integrated discovery (DAVID) and WebGestalt [49]. Using the DAVID API, 120 "adhesion"-, "integrin"- and "extracellular matrix"-related genes were isolated from the initial gene set. Of these, the list was refined further by filtering for those with alteration frequencies greater than 10% among TCGA patients. Pathway memberships of genes were extracted from GeneCards, KEGG and Pathway Commons [50] and visualized by Cytoscape (v3.2.0). Nodes on the edges of the network that did not connect to any of the main cancer pathways were removed (i.e., CLTC and EPN1). The genes were arranged by cellular location into either the extracellular matrix (ECM) and plasma membrane or the cytoplasm.

4.7. In Vitro Functional Assays

Cell proliferation assays were performed on A549, H1573 and HCC827 cells as described previously [51]. In brief, 5000 cells were seeded per well of an E-plate. Impedance was measured every 15 min for 80 h. Growth curves were constructed using the xCELLigence platform (ACEA Biosciences, San Diego, CA, USA). Tumor cell invasion was assessed in vitro by the reconstituted basement membrane (Matrigel) invasion assay [52], which was performed using 8- μ m polycarbonate filters coated with reconstituted basement membrane (Matrigel; BD Bio-sciences, San Jose, CA, USA). The motility of cells was measured after seeding in a plastic-bottomed 24-well dish and incubated for 12 h. FBS was added to the medium, and the migration assay was carried out for 24 h. Images were acquired every 20 min for approximately 24 h using a 10 \times phase objective.

5. Conclusions

Extensive efforts by many laboratories have attempted to identify cancer genomic features that correlate with the prognosis of NSCLC patients [3,53]. Very few studies have attempted to conduct biological validation of the prognostic genes identified in these prognostic signatures [9,54]. In contrast, Tang et al. used a systems biology approach integrating genome-wide functional (RNAi) data and genetic aberration data to derive a 12-gene prognostic signature that is also predictive for adjuvant chemotherapy benefits in NSCLC. Our study represents a translational research approach to identify genomic features of NSCLC tumor cells that could confer the prognostic implication in patient clinical outcome. We have provided evidence that the good prognostic impact of a high level of genomic aberrations was contributed by genes that are not considered oncogenic drivers in NSCLC through potential neoantigen generation and impairment of proteins that remain essential for tumor cells to manifest their malignant phenotypes.

Supplementary Materials: The following are available online at <http://www.mdpi.com/2072-6694/11/7/1009/s1>, Figure S1: Association between the number of altered genes and overall survival in four independent cohorts, Figure S2: Effect of immunogenic peptides in stage I patients, Figure S3: Distribution of variant allele frequencies (VAFs) of missense mutations used to stratify the TCGA NSCLC stage I patients. Table S1: List of 865 genes, Table S2: Summarized clinical-pathological features of NSCLC cohorts, Table S3: List of PDX models characterized by NAG levels, Table S4: Differentially-expressed genes between high and low NAG patients predictive of patient survival, Table S5: Survival analysis of neoantigens estimated from all coding mutations, Table S6: Clinical data from TCGA NSCLC stage I patients across subpopulations. Individual Excel spreadsheets contain Table S7: Comparison of mutation burden measures and their association with survival in early stage NSCLC.

Author Contributions: D.W., N.-A.P., F.A.S., B.G.W., R.N., J.F., G.L. and M.-S.T. conceived of the study and devised the experiments. D.W., N.-A.P., T.M.F., J.C., C.-Q.Z., D.L., M.P., M.F.M. and V.R. were involved in data processing, analysis and interpretation. D.W. and M.-S.T. wrote the first draft of the paper. All other co-authors contributed to the writing and approved the final manuscript.

Funding: This work was supported by grants from the Canadian Cancer Society (CCS), Grant #020527, CCS IMPACT Grant #701595, the Canadian Institutes of Health Research Foundation Grant FDN-148395 and the Princess Margaret Cancer Foundation. D.W. is supported by funding from the NIHR Sheffield Biomedical Research Centre. M.-S.T. is the M. Qasim Choksi Chair in Lung Cancer Translational Research. F.A.S. is the Scott Taylor Chair in Lung Cancer Research.

Acknowledgments: We wish to thank Jenna Sykes, Jessica Weiss and Ni Liu for assistance in data collection.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Director's Challenge Consortium for the Molecular Classification of Lung Adenocarcinoma; Shedden, K.; Taylor, J.M.G.; Enkemann, S.A.; Tsao, M.-S.; Yeatman, T.J.; Gerald, W.L.; Eschrich, S.; Jurisica, I.; Giordano, T.J.; et al. Gene expression-based survival prediction in lung adenocarcinoma: A multi-site, blinded validation study. *Nat. Med.* **2008**, *14*, 822–827. [[PubMed](#)]
2. Zhu, C.-Q.; Pintilie, M.; John, T.; Strumpf, D.; Shepherd, F.A.; Der, S.D.; Jurisica, I.; Tsao, M.-S. Understanding prognostic gene expression signatures in lung cancer. *Clin. Lung Cancer* **2009**, *10*, 331–340. [[CrossRef](#)] [[PubMed](#)]
3. Tang, H.; Wang, S.; Xiao, G.; Schiller, J.; Papadimitrakopoulou, V.; Minna, J.; Wistuba, I.I.; Xie, Y. Comprehensive evaluation of published gene expression prognostic signatures for biomarker-based lung cancer clinical studies. *Ann. Oncol.* **2017**, *28*, 733–740. [[CrossRef](#)] [[PubMed](#)]
4. Shepherd, F.A.; Domerg, C.; Hainaut, P.; Jänne, P.A.; Pignon, J.-P.; Graziano, S.; Douillard, J.-Y.; Brambilla, E.; Le Chevalier, T.; Seymour, L.; et al. Pooled analysis of the prognostic and predictive effects of KRAS mutation status and KRAS mutation subtype in early-stage resected non-small-cell lung cancer in four trials of adjuvant chemotherapy. *J. Clin. Oncol.* **2013**, *31*, 2173–2181. [[CrossRef](#)] [[PubMed](#)]
5. Ma, X.; Le Teuff, G.; Lacas, B.; Tsao, M.S.; Graziano, S.; Pignon, J.-P.; Douillard, J.-Y.; Le Chevalier, T.; Seymour, L.; Filipits, M.; et al. Prognostic and Predictive Effect of TP53 Mutations in Patients with Non-Small Cell Lung Cancer from Adjuvant Cisplatin-Based Therapy Randomized Trials: A LACE-Bio Pooled Analysis. *J. Thorac. Oncol.* **2016**, *11*, 850–861. [[CrossRef](#)] [[PubMed](#)]
6. Zer, A.; Ding, K.; Lee, S.M.; Goss, G.D.; Seymour, L.; Ellis, P.M.; Hackshaw, A.; Bradbury, P.A.; Han, L.; O'Callaghan, C.J.; et al. Pooled Analysis of the Prognostic and Predictive Value of KRAS Mutation Status and Mutation Subtype in Patients with Non-Small Cell Lung Cancer Treated with Epidermal Growth Factor Receptor Tyrosine Kinase Inhibitors. *J. Thorac. Oncol.* **2016**, *11*, 312–323. [[CrossRef](#)] [[PubMed](#)]
7. Seymour, L.; Le Teuff, G.; Brambilla, E.; Shepherd, F.A.; Soria, J.-C.; Kratzke, R.; Graziano, S.; Douillard, J.-Y.; Rosell, R.; Reiman, A.; et al. LACE-Bio: Validation of Predictive and/or Prognostic Immunohistochemistry/Histochemistry-based Biomarkers in Resected Non-small-cell Lung Cancer. *Clin. Lung Cancer* **2019**, *20*, 66–73. [[CrossRef](#)] [[PubMed](#)]
8. Tang, H.; Xiao, G.; Behrens, C.; Schiller, J.; Allen, J.; Chow, C.-W.; Suraokar, M.; Corvalan, A.; Mao, J.; White, M.A.; et al. A 12-gene set predicts survival benefits from adjuvant chemotherapy in non-small cell lung cancer patients. *Clin. Cancer Res.* **2013**, *19*, 1577–1586. [[CrossRef](#)] [[PubMed](#)]
9. Hai, J.; Zhu, C.-Q.; Wang, T.; Organ, S.L.; Shepherd, F.A.; Tsao, M.-S. TRIM14 is a Putative Tumor Suppressor and Regulator of Innate Immune Response in Non-Small Cell Lung Cancer. *Sci. Rep.* **2017**, *7*, 39692. [[CrossRef](#)]

10. Devarakonda, S.; Rotolo, F.; Tsao, M.-S.; Lanc, I.; Brambilla, E.; Masood, A.; Olaussen, K.A.; Fulton, R.; Sakashita, S.; McLeer-Florin, A.; et al. Tumor Mutation Burden as a Biomarker in Resected Non-Small-Cell Lung Cancer. *J. Clin. Oncol.* **2018**, *36*, 2995–3006. [[CrossRef](#)]
11. Owada-Ozaki, Y.; Muto, S.; Takagi, H.; Inoue, T.; Watanabe, Y.; Fukuhara, M.; Yamaura, T.; Okabe, N.; Matsumura, Y.; Hasegawa, T.; et al. Prognostic Impact of Tumor Mutation Burden in Patients with Completely Resected Non-Small Cell Lung Cancer: Brief Report. *J. Thorac. Oncol.* **2018**, *13*, 1217–1221. [[CrossRef](#)]
12. Yu, H.; Chen, Z.; Ballman, K.; Watson, M.A.; Govindan, R.; Lanc, I.; Beer, D.G.; Bueno, R.; Chirieac, L.; Chui, M.H.; et al. Correlation of PD-L1 expression with tumor mutation burden and gene signatures for prognosis in early stage squamous cell lung carcinoma. *J. Thorac. Oncol.* **2019**, *14*, 25–36. [[CrossRef](#)]
13. Dong, Z.-Y.; Zhang, C.; Li, Y.-F.; Su, J.; Xie, Z.; Liu, S.-Y.; Yan, L.-X.; Chen, Z.-H.; Yang, X.-N.; Lin, J.-T.; et al. Genetic and Immune Profiles of Solid Predominant Lung Adenocarcinoma Reveal Potential Immunotherapeutic Strategies. *J. Thorac. Oncol.* **2018**, *13*, 85–96. [[CrossRef](#)]
14. McGranahan, N.; Furness, A.J.S.; Rosenthal, R.; Ramskov, S.; Lyngaa, R.; Saini, S.K.; Jamal-Hanjani, M.; Wilson, G.A.; Birnbak, N.J.; Hiley, C.T.; et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. *Science* **2016**, *351*, 1463–1469. [[CrossRef](#)]
15. Chalmers, Z.R.; Connelly, C.F.; Fabrizio, D.; Gay, L.; Ali, S.M.; Ennis, R.; Schrock, A.; Campbell, B.; Shlien, A.; Chmielecki, J.; et al. Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med.* **2017**, *9*, 34. [[CrossRef](#)]
16. McGranahan, N.; Rosenthal, R.; Hiley, C.T.; Rowan, A.J.; Watkins, T.B.K.; Wilson, G.A.; Birnbak, N.J.; Veeriah, S.; Van Loo, P.; Herrero, J.; et al. Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. *Cell* **2017**, *171*, 1259–1271.e11. [[CrossRef](#)]
17. Turajlic, S.; Litchfield, K.; Xu, H.; Rosenthal, R.; McGranahan, N.; Reading, J.L.; Wong, Y.N.S.; Rowan, A.; Kanu, N.; Al Bakir, M.; et al. Insertion-and-deletion-derived tumour-specific neoantigens and the immunogenic phenotype: A pan-cancer analysis. *Lancet Oncol.* **2017**, *18*, 1009–1021. [[CrossRef](#)]
18. Campbell, J.D.; Alexandrov, A.; Kim, J.; Wala, J.; Berger, A.H.; Peadarallu, C.S.; Shukla, S.A.; Guo, G.; Brooks, A.N.; Murray, B.A.; et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. *Nat. Genet.* **2016**, *48*, 607–616. [[CrossRef](#)]
19. Schumacher, T.N.; Schreiber, R.D. Neoantigens in cancer immunotherapy. *Science* **2015**, *348*, 69–74. [[CrossRef](#)]
20. 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. Cancer immunology. Mutational landscape determines sensitivity to PD-1 blockade in non-small cell lung cancer. *Science* **2015**, *348*, 124–128. [[CrossRef](#)]
21. Carbone, D.P.; Reck, M.; Paz-Ares, L.; Creelan, B.; Horn, L.; Steins, M.; Felip, E.; van den Heuvel, M.M.; Ciuleanu, T.-E.; Badin, F.; et al. First-Line Nivolumab in Stage IV or Recurrent Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **2017**, *376*, 2415–2426. [[CrossRef](#)]
22. Wang, D.; Pham, N.-A.; Tong, J.; Sakashita, S.; Allo, G.; Kim, L.; Yanagawa, N.; Raghavan, V.; Wei, Y.; To, C.; et al. Molecular heterogeneity of non-small cell lung carcinoma patient-derived xenografts closely reflect their primary tumors. *Int. J. Cancer* **2017**, *140*, 662–673. [[CrossRef](#)]
23. Wu, L.; Allo, G.; John, T.; Li, M.; Tagawa, T.; Opitz, I.; Anraku, M.; Yun, Z.; Pintilie, M.; Pitcher, B.; et al. Patient-Derived Xenograft Establishment from Human Malignant Pleural Mesothelioma. *Clin. Cancer Res.* **2017**, *23*, 1060–1067. [[CrossRef](#)]
24. Zhang, X.; Claerhout, S.; Prat, A.; Dobrecki, L.E.; Petrovic, I.; Lai, Q.; Landis, M.D.; Wiechmann, L.; Schiff, R.; Giuliano, M.; et al. A renewable tissue resource of phenotypically stable, biologically and ethnically diverse, patient-derived human breast cancer xenograft models. *Cancer Res.* **2013**, *73*, 4885–4897. [[CrossRef](#)]
25. Li, L.; Wei, Y.; To, C.; Zhu, C.-Q.; Tong, J.; Pham, N.-A.; Taylor, P.; Ignatchenko, V.; Ignatchenko, A.; Zhang, W.; et al. Integrated omic analysis of lung cancer reveals metabolism proteome signatures with prognostic impact. *Nat. Commun.* **2014**, *5*, 5469. [[CrossRef](#)]
26. Chen, Z.; Fillmore, C.M.; Hammerman, P.S.; Kim, C.F.; Wong, K.-K. Non-small-cell lung cancers: A heterogeneous set of diseases. *Nat. Rev. Cancer* **2014**, *14*, 535–546. [[CrossRef](#)]
27. Futreal, P.A.; Andrew Futreal, P.; Coin, L.; Marshall, M.; Down, T.; Hubbard, T.; Wooster, R.; Rahman, N.; Stratton, M.R. A census of human cancer genes. *Nat. Rev. Cancer* **2004**, *4*, 177–183. [[CrossRef](#)]
28. Li, T.; Fan, J.; Wang, B.; Traugh, N.; Chen, Q.; Liu, J.S.; Li, B.; Liu, X.S. TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells. *Cancer Res.* **2017**, *77*, e108–e110. [[CrossRef](#)]

29. McGrail, D.J.; Federico, L.; Li, Y.; Dai, H.; Lu, Y.; Mills, G.B.; Yi, S.; Lin, S.-Y.; Sahni, N. Multi-omics analysis reveals neoantigen-independent immune cell infiltration in copy-number driven cancers. *Nat. Commun.* **2018**, *9*, 1317. [[CrossRef](#)]
30. Burrell, R.A.; McGranahan, N.; Bartek, J.; Swanton, C. The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* **2013**, *501*, 338–345. [[CrossRef](#)]
31. Belvedere, O.; Berri, S.; Chalkley, R.; Conway, C.; Barbone, F.; Pisa, F.; MacLennan, K.; Daly, C.; Alsop, M.; Morgan, J.; et al. A computational index derived from whole-genome copy number analysis is a novel tool for prognosis in early stage lung squamous cell carcinoma. *Genomics* **2012**, *99*, 18–24. [[CrossRef](#)]
32. Petitjean, A.; Mathe, E.; Kato, S.; Ishioka, C.; Tavtigian, S.V.; Hainaut, P.; Olivier, M. Impact of mutant p53 functional properties on TP53 mutation patterns and tumor phenotype: Lessons from recent developments in the IARC TP53 database. *Hum. Mutat.* **2007**, *28*, 622–629. [[CrossRef](#)]
33. Martincorena, I.; Campbell, P.J. Somatic mutation in cancer and normal cells. *Science* **2015**, *349*, 1483–1489. [[CrossRef](#)]
34. Desgrosellier, J.S.; Cheresh, D.A. Integrins in cancer: Biological implications and therapeutic opportunities. *Nat. Rev. Cancer* **2010**, *10*, 9–22. [[CrossRef](#)]
35. Miller, A.; Asmann, Y.; Cattaneo, L.; Braggio, E.; Keats, J.; Auclair, D.; Lonial, S.; MMRF CoMMpass Network; Russell, S.J.; Stewart, A.K. High somatic mutation and neoantigen burden are correlated with decreased progression-free survival in multiple myeloma. *Blood Cancer J.* **2017**, *7*, e612. [[CrossRef](#)]
36. Campesato, L.F.; Barroso-Sousa, R.; Jimenez, L.; Correa, B.R.; Sabbaga, J.; Hoff, P.M.; Reis, L.F.L.; Galante, P.A.F.; Camargo, A.A. Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. *Oncotarget* **2015**, *6*, 34221–34227. [[CrossRef](#)]
37. Kinkead, H.L.; Hopkins, A.; Lutz, E.; Wu, A.A.; Yarchoan, M.; Cruz, K.; Woolman, S.; Vithayathil, T.; Glickman, L.H.; Ndubaku, C.O.; et al. Combining STING-based neoantigen-targeted vaccine with checkpoint modulators enhances antitumor immunity in murine pancreatic cancer. *JCI Insight* **2018**, *3*, 122857. [[CrossRef](#)]
38. Miller, A.; Cattaneo, L.; Asmann, Y.W.; Braggio, E.; Keats, J.J.; Auclair, D.; Lonial, S.; Russell, S.J.; Keith Stewart, A. Correlation Between Somatic Mutation Burden, Neoantigen Load and Progression Free Survival in Multiple Myeloma: Analysis of MMRF CoMMpass Study. *Blood* **2016**, *128*, 193.
39. John, T.; Kohler, D.; Pintilie, M.; Yanagawa, N.; Pham, N.-A.; Li, M.; Panchal, D.; Hui, F.; Meng, F.; Shepherd, F.A.; et al. The ability to form primary tumor xenografts is predictive of increased risk of disease recurrence in early-stage non-small cell lung cancer. *Clin. Cancer Res.* **2011**, *17*, 134–141. [[CrossRef](#)]
40. Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. *Nature* **2014**, *511*, 543–550. [[CrossRef](#)]
41. Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **2012**, *489*, 519–525. [[CrossRef](#)]
42. Jacobsen, A. *Cgdsr: R-Based API for Accessing the MSKCC Cancer Genomics Data Server (CGDS)*, version 1.3.0; CRAN: Taipei City, Taiwan, 2017.
43. Deng, M.; Brägelmann, J.; Kryukov, I.; Saraiva-Agostinho, N.; Perner, S. FirebrowseR: An R client to the Broad Institute's Firehose Pipeline. *Database* **2017**, *2017*, 160. [[CrossRef](#)]
44. Der, S.D.; Sykes, J.; Pintilie, M.; Zhu, C.-Q.; Strumpf, D.; Liu, N.; Jurisica, I.; Shepherd, F.A.; Tsao, M.-S. Validation of a histology-independent prognostic gene signature for early-stage, non-small-cell lung cancer including stage IA patients. *J. Thorac. Oncol.* **2014**, *9*, 59–64. [[CrossRef](#)]
45. Nielsen, M.; Andreatta, M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. *Genome Med.* **2016**, *8*, 33. [[CrossRef](#)]
46. Charoentong, P.; Finotello, F.; Angelova, M.; Mayer, C.; Efremova, M.; Rieder, D.; Hackl, H.; Trajanoski, Z. Pan-cancer Immunogenomic Analyses Reveal Genotype-Immunophenotype Relationships and Predictors of Response to Checkpoint Blockade. *Cell Rep.* **2017**, *18*, 248–262. [[CrossRef](#)]
47. Szolek, A.; Schubert, B.; Mohr, C.; Sturm, M.; Feldhahn, M.; Kohlbacher, O. OptiType: Precision HLA typing from next-generation sequencing data. *Bioinformatics* **2014**, *30*, 3310–3316. [[CrossRef](#)]
48. Sidney, J.; Peters, B.; Frahm, N.; Brander, C.; Sette, A. HLA class I supertypes: A revised and updated classification. *BMC Immunol.* **2008**, *9*, 1. [[CrossRef](#)]
49. Wang, J.; Duncan, D.; Shi, Z.; Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): Update 2013. *Nucleic Acids Res.* **2013**, *41*, W77–W83. [[CrossRef](#)]

50. Cerami, E.G.; Gross, B.E.; Demir, E.; Rodchenkov, I.; Babur, O.; Anwar, N.; Schultz, N.; Bader, G.D.; Sander, C. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Res.* **2010**, *39*, D685–D690. [[CrossRef](#)]
51. Radulovich, N.; Leung, L.; Ibrahimov, E.; Navab, R.; Sakashita, S.; Zhu, C.-Q.; Kaufman, E.; Lockwood, W.W.; Thu, K.L.; Fedyshyn, Y.; et al. Coiled-coil domain containing 68 (CCDC68) demonstrates a tumor-suppressive role in pancreatic ductal adenocarcinoma. *Oncogene* **2015**, *34*, 4238–4247. [[CrossRef](#)]
52. Navab, R.; Strumpf, D.; Bandarchi, B.; Zhu, C.-Q.; Pintilie, M.; Ramnarine, V.R.; Ibrahimov, E.; Radulovich, N.; Leung, L.; Barczyk, M.; et al. Prognostic gene-expression signature of carcinoma-associated fibroblasts in non-small cell lung cancer. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 7160–7165. [[CrossRef](#)]
53. Zhu, C.-Q.; Tsao, M.-S. Prognostic markers in lung cancer: Is it ready for prime time? *Transl. Lung Cancer Res.* **2014**, *3*, 149–158.
54. Hai, J.; Zhu, C.-Q.; Bandarchi, B.; Wang, Y.-H.; Navab, R.; Shepherd, F.A.; Jurisica, I.; Tsao, M.-S. L1 cell adhesion molecule promotes tumorigenicity and metastatic potential in non-small cell lung cancer. *Clin. Cancer Res.* **2012**, *18*, 1914–1924. [[CrossRef](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).