**Blood-based Genomic and Cellular Determinants of Response to Neoadjuvant PD-1 Blockade in Patients with Non-Small-Cell Lung Cancer**

Xi Zhang[1,2#], Rui Chen[1], Wenqing Li[1], Shengchao Zhang[1], Mengju Jiang[1], Guodong Su1[1], Yuru Liu[1], Yu Cai[1], Wuhao Huang[3], Yuyan Xiong[1,2], Shengguang Wang[3#]

1 School of Life Science, Northwest University, Xi'an, Shaanxi 710069, China

2 Key Laboratory of Resource Biology and Biotechnology in Western China, Ministry of Education, School of Medicine, Northwest University, Xi'an, Shaanxi 710069, China

3 Department of Lung Cancer, Tianjin Medical University Cancer Institute and Hospital, National Clinical Research Center for Cancer, Key Laboratory of Cancer Prevention and Therapy, Tianjin's Clinical Research Center for Cancer, Tianjin Lung Cancer Center, Tianjin 300060, China.

# Corresponding authors

**Abstract**

**Background:** Despite the improved survival observed in PD-1/PD-L1 blockade therapy, there still is a lack of response to the anti-PD1 therapy for a large proportion of cancer patients across multiple indications, including non-small cell lung cancer (NSCLC)

**Methods:** Transcriptomic profiling was performed on 57 whole blood samples from 31 NSCLC patients and 5 healthy donors, including both responders and non-responders received anti-PD-1 Tislelizumab plus chemotherapy, to characterize differentially expressed genes (DEGs), signature pathways, and immune cell subsets regulated during treatment. Mutations of oncogenic drivers were identified and associated with therapeutic outcomes in a validation cohort with 1661 cancer patients. These multi-level biomarkers were validated and compared across different methods, external datasets and multiple computational tools.

**Results:** NSCLC patients examined and achieved pathological complete response (pCR) were considered as responders or non-responders otherwise. Expression of hundreds DEGs (FDR $p<0.05$, fold change<-2 or >2) was changed in blood during neoadjuvant anti-PD-1 treatment, as well as in lung cancer tissue as compared to normal samples. Enriched PD-1-mediated pathways and elevated cell abundances of CD8 T cells and regulatory T cells were exclusively observed in responder blood samples. In an independent validation cohort of 1661 pan-cancer patients, a panel of 4 top ranked genetic alterations (PTCH1, DNMT3A, PTPRS, JAK2) identified from responders in discovery cohort were found positively associated with the overall survival ($p<0.05$).

**Conclusion:** These findings suggest peripheral blood-based biomarkers and cell subsets could be utilized to define the response to neoadjuvant PD-1 blockade in NSCLC patients and a set of

novel gene mutations is strongly associated with the therapeutic outcome of cancer immunotherapy.

**Key words:** anti-PD-1 blockade, biomarker, RNA-seq, non-small cell lung cancer

## Introduction

Given the significant benefit checkpoint inhibitors (ICIs) immunotherapy achieved in cancer, it is emerging as a "common denominator" in the past decade [1]. To date, several anti-PD-1 antibodies have been approved as monotherapy or neoadjuvant immunotherapy in combination with chemotherapy or other drugs [2]. It is proved that anti-PD-1 agents with chemotherapy could restore anti-tumor activities of multiple immune cell subsets, leading to an increased overall survival [3]. Despite the impressive success of PD-1 blockade therapy, the clinical benefit of this treatment is always and only limited to a small subset of patients [4]. Advanced NSCLC is one of the first pioneers becoming a common therapeutic focus for therapy targeting PD-1 or its ligand PD-L1 [5, 6]. Although the combination of anti-PD-1 therapy with chemotherapy have showed more encouraging results in the up-front treatment of NSCLC [7], the overall response rate to PD-1/PD-L1 blockade treatment in NSCLC is still low. Taking anti-PD-1 antibody Pembrolizumab as example, the objective response rate for the unselected population with NSCLC was only 19% and the median overall survival was 12 months [8].

The outcome of anti-PD-1 treatment for NSCLC patents is associated with many factors, such as PD-L1 expression level, tumor mutation burden, mRNA expression of certain biomarker genes, neo-antigens, and the composition of tumor-infiltrating immune cells [9]. Many clinical studies as well as review articles suggested PD-L1 expression testing should become (if not yet) a companion diagnostic (CDx) assay for patient who is newly diagnosed NSCLC, while other biomarkers should be evaluated in the future [10, 11]. A couple of drug-specific CDx PD-L1

immunohistochemistry (IHC) assays were validated and recently approved to determine eligibility of patients with NSCLC for anti-PD-1 treatment [12, 13]. In addition to IHC assay which requires tissue collection, the demand of treatment outcome determinants identified from non-invasive blood test is increasing. Actually a number of new biomarkers predictive of the response to ICIs were investigated, including DNA mutational variant in NK cells [14], CD28 levels in tumor-infiltrating CD8-T cells [15], metabolite from on-treatment serum [16], local or peripheral immune cell clusters [17, 18], or combination of T cell properties and energy metabolism markers in blood [19]. The proposed biomarkers for a successful response to anti-PD-1/PD-L1 treatments mostly fall in two kinds, either molecular signatures determined by genetic tests (such as tumor mutational burden [20, 21], and mismatch-repair deficiency [22]) or signatures of the immune response (such as tumor-infiltrating lymphocytes [23] and peripheral blood immune cells [24]) .

In the present study, a comprehensive analysis workflow was constructed to identify multi-level biomarker signatures using the transcriptomic profiles of whole blood from a discovery cohort and followed by validations in published datasets or an independent cohort. This workflow utilized well-recognized computational tools including *in silico* deconvolution approaches developed to profile tumor-infiltrating immune cells in tissue [25, 26] and validated on PBMC samples in our prior studies [27-29]. This workflow also adapted an upgraded variant-calling method to rate and report cancer-related DNA mutations called from RNA-sequencing (RNA-seq) data [30, 31]. Together, this study identified and validated gene expression biomarkers, peripheral immune cell sets and a panel of DNA mutational markers that are expected to be predictive or responding to neoadjuvant anti-PD-1 treatment in NSCLC patients.

**Result**

**Responder DEGs are potential gene expression biomarkers that alter during neoadjuvant anti-PD-1 treatment**

Differential expression analysis was performed on the transcriptomic profiles between all pre- and post-treatment blood samples to identify differentially expressed genes (DEGs) in response to the treatment. In total 927 protein-coding genes were identified as significant DEGs (FDR p <0.05 and Log2(fold change) >1 or <-1) in comparison of unpaired samples and 1211 DEGs were identified from paired samples (**Figure 2A, Supplementary Table S1**). Notably, most of the DEGs (876 out of 927) identified from unpaired samples are also detected in pairwise comparison although the latter one used less samples (**Figure 2A, Supplementary Figure S1A**). In terms of the comparison of paired samples, 2298 and 1025 DEGs were identified from responders and non-responders respectively while 834 Common DEGs are shared in both (**Figure 2B, Supplementary Figure S1B**). As annotated in Figures 2A and 2B, 8 top ranked genes were selected (FDR p <0.05 and Log2(fold change) >1.5 or <-1.5) from the Common DEG lists described above and their mRNA expression levels were examined by qRT-PCR for validation (**Figure 2C**). We next looked up the expression profiles of these genes in cancer tissues database (GENT2). Interestingly, 7 out of 8 gene showed consistent alteration from tumor to normal lung tissue versus pre- to post-treatment blood samples (p<0.01) (**Figure 2D**). A strong agreement to this result is observed in a larger lung cancer database (LCE) with meta-analysis across multiple independent cohorts (**Supplementary Figure S2**).

Next, we sought to test our hypothesis that these DEGs may serve as gene expression biomarkers during neoadjuvant anti-PD-1 treatment. Firstly only the significant DEGs from responders, either exclusively from responders (Unique DEGs) or Common DEGs, showed overlay with published transcriptional signatures that either correlating with PD-L1 expression [32] or

responding to anti-PD1 therapy [33] in cancer patients (**Figure 3A**). In consistent with the RT-PCR results of Common DEGs (**Figure 2C**), a cross-method validation is also seen in 5 representative genes from Unique DEGs (**Figure 3B**). Secondly, it is notable that the Common DEGs are more significantly regulated (higher fold change and lower p value) than the Unique DEGs (**Figure 3B**). However, the Common DEGs change displayed a low correlation (correlation coefficient $R^2$ = 0.29) between responders and non-responders (**Figure 3C**), suggesting the immune activation boosted by anti-PD-1 may stimulate different molecular mechanisms in responders and non-responders. Lastly, the sample similarity matrix (**Figure 3D**) as well as hierarchical clustering (**Figure 3E, Supplementary Figure S3**) results were generated using Unique DEGs and Common DEGs. In both cases, only Unique DEGs organized pre-treatment samples into a distinct group from the post-treatment samples and its gene expression profile changed significantly across pre-, post-treatment and HC samples.

**PD-1-mediated pathways and regulatory T cells are distinct signatures in responders during the treatment**

We also asked which pathways and immune cell responses account for a successful response to the treatment. Unlike the striking difference observed in the numbers of DEGs identified from responders vs. non-responders (**Figure 2B**), the pathways and GOs identified from these two groups are quite comparable. Not only the total numbers are close but also a majority of them are commonly seen in two groups, at least in KEGG pathways (211 out of 246 and 237) and GOs (2144 out of 2634 and 2728) (**Figure 4A, Supplementary Figure S4 & S5**). Uniquely, the result in Reactome database showed about one quarter (215 out of 817) of the total pathways (**Figure 4A**) and 12 out of the top 20 pathways are exclusive in responders (**Figure 4B**) and non-responders (**Figure 4C**). Interestingly, several key pathways related with PD-1 regulatory

network are overrepresented and only overrepresented in responders, including signaling of PD-1, CD3, TCR, CD28, IL-1 and IFN-γ. Gene set enrichment analysis (GSEA) analyses were performed across all available collections and the top ranked gene sets were compared in distinct-directional classes (up and down regulations) (**Supplementary Figure S6**). All GSEA results suggested substantial changes in immune cell subsets such as a decrease in monocytes compared to plasmacytoid DCs and an increase in CD4+ T cells compared to myeloid cells (**Figure 5A & 5B, Supplementary Figure S7**).

The immune cell abundance change was further elucidated by *in silico* leukocyte deconvolution approach adopted from our previous work (AImmune) [27-29] as well as a classic machine learning tool CIBERSORTx [26] (**Supplementary Figure S8 & S9**). In agreement with the GSEA results, a significant composition decrease in monocytes and a sturdy increase of DCs were observed in AImmune (**Figure 5C, Supplementary Figure S8**) while a substantial increase of resting memory CD4+ and CD8+ T lymphocytes were demonstrated in CIBERSORTx (**Figure 5D, Supplementary Figure S9**). While the changes of the cell subsets above were actually seen in both responders and non-responders, the abundance of macrophages, CD8 T cells and regulatory T cells were elevated and only elevated in responders (**Supplementary Figure S8, Figure 5E**).

**A panel of mutational markers was validated through an independent cohort to predict sensitivity to PD-1 blockade**

The Cancer-Related Analysis of Variants Toolkit (CRAVAT) was applied in the RNA-seq data from our discovery cohort (**Figure 1**) in order to identify somatic mutations that predict survival benefit for anti-PD-1 treatment (**Supplementary Figure S10**). Top 50 mutated genes (classifier gene-level p values <0.05 and FDR = 0.05, ranked by CHASM score) were identified from

responders, non-responders and healthy donors, with minority of the genes are shared by two or more groups (**Figure 6A**). The scores generated from CHASM (driver classifier) and VEST (pathogenicity classifier) are all closed to 1, suggesting these mutations are confidently classified as cancer drivers and likely are pathogenic (**Figure 6B**). Three out of all 39 mutational markers (TNFAIP3, BRCA1, ASXL1) identified exclusively from non-responders in our NSCLC cohort were shared with the top 50 mutational markers from progressive disease patients of an independent cervical cancer cohort (GEO repository: GSE205247) (**Figure 6C**). It is worth noting 3 shared genes here are highly ranked as the top 1st, 3rd and $4^{th}$ mutations in the discovery cohort (bold in **Figure 6B**).

In another much larger validation cohort ("tmb_mskcc_2018") with 1661 pan-cancer patients [34], 8 out of all 20 top-ranked gene mutations from the discovery cohort were detected patients who all received PD-1 or CTLA-4 blockade treatment (**Figure 6D**). The Kaplan-Meier curves illustrates the significantly difference of overall survival ($p<0.01$) across 9 patient groups as defined by 8 mutational markers. The overall median survival among patients with no mutation is 16 months, while the median survivals (if available) are shorter (10-15 months) in patients with non-responder mutations and much longer (23-41 months) in patients with responder mutations (**Figure 6D**). Moreover, the mean survival was consistently extended (41 vs 17 months in Res. vs Non-res. for all patients, 31 vs 14 months in Res. vs Non-res. for anti-PD1/PD-L1 treated patients) in a patient group defined by a combination of all 4 genetic alterations (PTCH1, DNMT3A, PTPRS, JAK2) identified from responder (**Figure 6E**). No significant difference was observed in patient groups defined by a combination of non-responder mutations (**Supplementary Figure S11**).

**Discussion**

Given the majority of tumor patients fail respond to anti-PD-1/PD-L1 therapy [35], a lot of effort made by clinical trials led to the identification of predictive biomarkers to select treatment responders and the discovery of underlying mechanisms during treatment for improving current treatment modalities [36, 37]. Although PD-1/PD-L1 IHC expression test has been widely evaluated as a predictive biomarker, as an invasive tissue test it is still not implemented as a routine diagnostic test for anti–PD-1/PD-L1 therapy worldwide [38]. Part of the reason is lack of universal standard of antibody for detecting PD-L1 expression and no mention it is still controversial whether PD-1/PD-L1 expression is an inadequate determinant of the treatment.

The overall hypothesis of the present study is the blood-based molecular signatures that differently expressed in blood samples from responders versus non-responders, and thus provide additive prognostic value in patient selection. In order to test this hypothesis, the present study employed an integrated workflow and identified multi-level response biomarkers from a discovery cohort, including gene expression biomarker, immune cell subsets and DNA mutational markers. The discovering results were then further validated by different methods (RNA-seq and qRT-PCR), multiple databases (GENT and LCE), innovative tools (AImmune and CIBERSORT) as well as independent cohort (tmb_mskcc_2018 etc.) By comparing responders and non-responders during the treatment, our data provided new evidences suggesting key driving genes, targetable pathways, peripheral immune cells, and most importantly mutational determinants which are detectable by blood test and of predictive value towards the therapy outcome and clinical efficacy.

At the beginning of this study, a group of Common DEGs that changed during the treatment were found consistently altered in overall comparisons of both responders and non-responders (**Figure 2A & 2B**). These genes were ranked as top DEGs in an earlier version of the discovery

cohort and thus were selected for RT-PCR validation and database investigation. Our finding of these genes was significant as them were, on one hand, and regulated by similar folds between RNA-seq and qRT-PCR (**Figure 2C**); on the other hand, their expression levels were also changed from tumor to normal tissue sample (**Figure 2D & Supplementary Figure S2**). This signature group contains a hematopoietic transcription regulator (GATA2) [39], a metabolism mediator (ANKRD22) [40, 41], and a transcription factor involved in differentiation control (LHX4) [42], which is consistent with the current understandings that immunological and metabolism mechanisms are enriched in post-treatment and responder patients [43]. Moreover, several of them were proven to regulate the oncogenesis of lung cancer or respiratory diseases. For instance, ANKRD22 as we found down-regulated in NSCLC patients after therapy, was reported to promote the progression of NSCLC by enhancing cell proliferation [44]. The expression of HBG1/2 is induced in normal lung tissue after smoking and is downregulated again when tumors form in smokers [45]. More importantly, a recent report also determined HBG1 was consistently upregulated in blood samples of patients with stage III/IV NSCLC from the 2$^{nd}$ anti-PD-1 antibody treatment through to the 5$^{th}$ treatment [46]. This observation seems dependent on stage and histology as both protein and mRNA levels of HBG1/2 were found down-regulated in stage I NSCLC adenocarcinomas [47]. Because these gene were idented from non-responders as well, these findings suggested the anti-PD-1 treatment may change transcription levels in cancer-relevant genes that not necessarily indicate a positive response.

More importantly, the present study identified Unique DEGs of responders which together with the Common DEGs showed exclusive overlays with two sets of published signature genes that are related with PD-L1 expression or anti-PD-1 therapy response. The first set of genes are the top 100 (out of the total 1325) transcriptional correlates of PD-L1 expression [32]. The second

set of genes were the 100 immune-positive genes used for predicting melanoma patient response to anti-PD-1 therapies as identified by integrating MHC I-association prediction with the TCGA transcriptomic data from pre- or on-treatment skin cutaneous melanoma samples (anti-PD-1 or combination of anti-PD-1 and anti-CTLA-4) [33]. The overlays between these published signature genes with newly identified DEGs from responders (**Figure 3A)** are not high (likely due to differences between study subjects, cancer types and sample types) but significantly greater than the DEGs only in non-responders. This result clearly proposed the biomarker value in responder DEGs rather than non-responder DEGs, which was further confirmed by sample similarity and hierarchical clustering (**Figure 3D & 3E)**. The predictive values of these newly identified markers for immunotherapy are either novel or only recently reported in other cancer indications. The top 1$^{st}$ ranked DEGs identified in responders in this study is DAAM2, which was recently proved to predict high sensitivity to chemotherapy, anti-EGFR therapy and immunotherapy for in pancreatic cancer patients [48]. The other DEG PHF24 as we identified down-regulated in responders here were found highly co-expressed with gene LAMP2 which confers prognostic value to indicate worse survival of esophageal cancer patients [49].

Given PD-1/PD-L1 pathway is a central mediator of immunosuppression, PD-1 pathway blockade therapy is known to crosstalk and fine-tune with T cell activation and differentiation as well as other immune cell functions. When analyzing the Reactome pathways and GOs regulated in responders, it makes a lot sense to only see PD-1 signaling and PD-1-regulated signaling cascades (such as CD3, TCR, CD28, IL-1 and IFN-γ) in responders but not in non-responders (**Figure 4B)**. In contrast, PD-L1 expression and PD-1 checkpoint pathway in cancer is a KEGG pathway regulated in both responders and non-responders, although the significance is much weaker in the latter one (**Supplementary Figure S4**). Our finding is supported by several studies

either concluded the expression of PD-1 or PD-L1 might not be qualified as a consistent biomarker for NSCLC patients received combination of chemotherapy plus immunotherapy [50], or questioned the practice of dichotomizing the range of PD-L1 expression levels is adequate for patient stratification [51]. Our GSEA analysis and immune cell abundance analysis (**Figure 5**) further sustained our findings in pathway and GO analyses, which are also backed up by a broad range of published studies. Firstly, since monocyte-to-DC differentiation attenuates CD8+ T cell response and other T-cell populations expanding [52], this differentiation was found significantly higher during successful PD-1 blockade therapy [53-55]. In line with that, our results observed DCs and macrophages were amplified in responders while mononuclear phagocytes as their precursor cells responded oppositely. Secondly, CD4 T cell memory was another cell subset we found expended during the treatment. Notably a longitudinal study recently reported that higher percentages of effector, CD62Llow CD4+ T cells, was detected in peripheral blood of responder patients with non–small lung cancer [56]. It was proposed that the abundance of certain CD4+ T cell subsets in peripheral blood before treatments is an effective predictor of responses in NSCLC patients [57]. At last, regulatory T cell (Treg cell) was found significantly elevated only in responders here, suggesting a specific role of Treg cell during PD-1 mediated immune suppression that may determine the outcome of anti-PD-1 therapy. PD-1 blockade may re-activate both CD8+ T cells and Treg cells but our finding confirmed that an effective therapy requires a skewed balance in favor of the effector CD8+ T cells rather than Treg cells [58, 59]. Summarizing, our data above proposed several novel DEGs and immune cell subsets that may serve as monitoring markers in peripheral blood of NSCLC patients. Given the fact these markers are, at least partially, independent of the PD-L1 expression levels in tissue, they are expected to bring additional values in guiding clinical decision-making.

In the last part of this study, several cancer driver genes were found significantly mutated in discovery cohort and led to extended (responder mutations) or reduced (non-responder mutations) survivals in validation cohorts (**Figure 6**). Individual diver gene mutations were previously demonstrated to promote (such as KRAS [60], TP53 [61], PTCH1 [62]) or weaken (such as JAK1/2[63], EGFR[64], PTEN [65]) the response to immune checkpoint inhibitor therapy in cancer patients. On one hand, the role of somatic mutation is consistent across cancer indications. An typical example is TNFAIP3 mutation indicates low responses to PD-1 inhibitor in NSCLC and cervical cancer patients as showed in the present study, which is also supported by a study on melanoma [66]. Another example is that the association of PTCH1 mutation with improved outcome of PD-1 blockade was seen in both colorectal cancer [62] and NSCLC (the present study). On the other hand, the role of a somatic mutation might vary across different cancer types. For example, JAK2 is associated with acquired resistance in melanoma [63] while in this NSCLC study it marks a better PD-1 blockade response. The overall impact of genome-wide mutation is to increase therapeutic sensitivity to immunotherapy as large-scale clinical trials revealed that high tumor mutational burden (TMB) level is associated with T-cell–mediated response and thus correlates with improved survivals to PD-1 plus CTLA-4 blockade in NSCLC patients [21, 67]. In line with this finding, our study highlights four newly identified mutations (PTCH1, DNMT3A, PTPRS, JAK2) as one panel is robustly predictive of clinical benefit across multiple cancer types with CTLA4 or PD-1/PD-L1 blockade therapy (**Figure 6E**).

It is important to point out the cell abundance analysis tool AImmune employed here is developed in-house and there is still room for improvement. Although this tool underwent robust testing and validation in our previous studies [27-29], the current version is only designed to cover up to 10 immune cell subsets from peripheral blood. This coverage limit is partially due to

its nature as AImmune was trained and developed to profile circulating blood samples while other computational immune profiling tools including CIBERSORTx [26, 68] and TIMER [69] are often employed to enumerate tumor-infiltrating cells in tissue-dependent microenvironments. Further development and fine-tuning of AImmune tool are required and it is expected to observe more effector T cell subsets detected in responders, such as CD62Llow CD4+ T cells, CD4 Th2 cells and Treg cells [56, 57, 70]. Other limitations of this study include its relatively small sample size, which decreased statistical power and flexibility in drawing statistical conclusion. Given the high molecular heterogeneity among different NSCLC subtypes, the number of patients sequenced in this study might be too small to provide strong statistical power or drawn informative conclusions. It is encouraged that our current findings and results should be further confirmed by a larger cohort with long-term follow-up and time series analysis.

In summary, the integrated analyses employed in the current study identified blood-based molecular and cellular signatures that are associated with the responsiveness and overall survival of PD-1/PD-L1 blockade therapy plus chemotherapy.

## Method

### Study cohort and overall workflow

In order to characterize blood-based determinants of positive response to PD-1 blockade treatment, this study built a discovery cohort collecting 57 whole blood samples from 31 patients visited Tianjin Cancer Hospital from 2020 to 2021 and 5 healthy donors (**Figure 1**). Five blood samples were collected from healthy donors and included as control samples in the discovery cohort. All patients were diagnosed with non-small cell lung cancer (**Table 1**) and underwent treatment of anti–PD-1 monoclonal antibody Tislelizumab [71, 72], in combination with

standard chemotherapy according to the histology. The majority of the patients were male (86.7%), over 60-year-old (66.7%), ever-smokers (80%), and classified as lung squamous cell carcinoma (SqCC) (76.7%) (**Table 1**). The protocol was approved by the local Ethics Committee and the Institutional Review Board of Norwest University (approval number: 200402001) and all subjects provided written informed consent.

In this study, 12 patients who achieved pathological complete response (pCR) were considered as responders, 17 patients categorized as non-responders and 2 patients were not examined to determine their responsiveness status. Briefly, gene expression biomarkers responding to the treatment were identified, validated by RT-PCR and compared with published transcriptional signatures correlated with PD-1 expression or therapy response [32, 33] and lung cancer tissue datasets (GENT2: https://gent2.appex.kr; LCE: https://lce.biohpc.swmed.edu). Pathways, gene ontology items and gene sets altered during treatment were investigated via KOBAS-i portal (http://kobas.cbi.pku.edu.cn/) [73]. Secondly, peripheral immune cell subsets activated or inactivated by the therapy were characterized and valeted by two independent immune profiling tools (AImmune; CIBERSORTx: https://cibersortx.stanford.edu). Finally, DNA mutational markers were reported by CRAVAT (https://www.cravat.us/CRAVAT) and then compared and validated through independent validation cohorts contain cancer patients received immune checkpoint inhibitor therapy [34, 74] (Gene Expression Omnibus, GSE44227; cBioPortal , tmb_mskcc_2018).

**AImmune analysis and computational analysis**

A novel *in silico* leukocyte deconvolution method, named AImmune, is a computational approach developed by integrating our established immune cell profiling [30] with published single cell RNA-seq data obtained from NSCLC blood samples (1071 qualified cells from one

patient, GSE127471) and healthy blood samples (8369 and 7687 cells from two donors, 10X Genomics). Briefly, with the additional marker genes included, more than 30 candidate marker genes for each cell subsets in peripheral immune cell subsets (CD4-T cells, CD8-T cells, B cells, Monocytes, DCs, NK cells and NKT cells) were selected based on their expression patterns across immune cell subsets [68]. The pairwise similarity statistic of all cell subsets was computed (data not shown) between all pairs of the candidate marker genes within the normalized RNA-seq profiles (FPKM) from whole blood samples. Using the criteria (average Pearson correlation factor >0.60, $p<0.01$), 10-20 selected marker genes were identified as our final marker genes. The raw cell abundance score was calculated as the sum of the simple averages of the marker genes' log2 expression, which allows comparison of cell composition across subject groups. This approach also tested a novel deconvolution model (unpublished) built by DNN (deep neural networks) algorithms and then trained by pseudo-bulk samples obtained by randomly subsampling of published scRNA-seq data [75]. Machine learning-based feature extraction (marker gene selection) was integrated for model optimization. Most of the computational analysis procedure was coded by common Python packages; scRNA-seq data was processed by R package scanpy; machine learning model was developed and tested with Python library Tensorflow. All computational analysis were performed and visualized using R version 3.6.1 or Python version 3.7.9.

**RNA isolation, transcriptomic profiling, and qRT-PCR**

Buffycoat or PBMCs were isolated from whole blood. Total RNA was isolated using Trizol (Invitrogen, Waltham, MA, USA) and the purity and concentration were verified using a NanoDrop ND-1000 instrument (ThermoFisher Scientific, Waltham, MA, USA). The integrity of the RNA was assessed by a 2100 Bioanalyzer gel image analysis system (Agilent, Santa Clara,

CA, USA). Only RNA samples with an integrity number of 7 or higher were processed for RNA-seq. At least 100 ng mRNA per sample was submitted for library preparation.

Qualified RNA samples were then enriched and synthesized into two strand cDNA for library preparation. RNA-seq libraries were constructed using the TruSeq RNA Sample Prep Kit (Illumina, San Diego, CA, USA). The libraries from qualified RNA samples were sequenced in the 150 nt paired-end mode on an Illumina HiSeq 6000 platform at Novogene Bioinformatic Technology (Tianjin, China). After quality filtering (FastQC, quality value >5), over 30 billion clean reads were obtained in each library and then used for down-stream analysis.

PCR primers were designed top-ranked genes (10 for each comparison) as obtained by differential gene analysis (**Supplemental Table S2**). Real-time quantitative PCR was performed in real-time PCR systems (Bio-Rad, Hercules, CA, USA). The 2−ΔΔCT method for relative gene expression analysis was performed to calculate the mRNA levels of individual gene. Two or three replicates were measured for each sample.

**Bioinformatics and variant calling**

All raw RNA-seq reads were filtered by R package trim_fastq to remove adapters, rRNA and low-quality reads. The QC criteria included: removing bases below Phred quality 20, containing over two “N”, or shorter than 75. The output reads were then indexed by aligner STAR and mapped to reference genome by BAM. Normalized read counts were generated and compared between groups to generate DEGs using R package DESeq2. Another R package countToFPKM was employed to produce FPKM for AImmune analysis. Genes that displayed at least two-fold difference in gene expression between comparison groups (false discovery rate [FDR]<0.05) were considered significant DEGs and carried forward in the analysis. Differentially expressed RNAs were illustrated in a volcano plot. Hierarchical clustering was performed to show the gene

expression patterns and similarities among samples. Pathway and gene ontology (GO) enrichment analysis was carried out via an integrated platform KOBAS 3.0 [35]. GSEA analysis was carried out by searching the established MSigDB gene-set collections (C7). CIBERSORTx analysis was performed following the instruction from the portal (https://cibersortx.stanford.edu). Differences of mRNA levels and cell abundance scores were evaluated using independent *t*-tests or paired *t*-tests if pairwise samples were given. CHAVAT [30, 31].

Variant calling was performed using the HaplotypeCaller that plug-in in the Genome Analysis Toolkit v4.0 (GATK). First, RNA reads were aligned to the reference genome using the STAR aligner, then the MarkDuplicates was used to clean up data. The gatk BaseRecalibrator and gatk ApplyBQSR were used to adjust the mass fraction of original bases, detect the system errors in mass fraction, and reduce the false positives. We used only variants marked with PASS in the VCF file and filtered the variant calls with the VariantFiltration tool. The Cancer-Related Analysis of Variants Toolkit (CRAVAT) is a well-recognized informatics toolkit used in this study for variant calling from VCF files [30, 31]. This tool covers multi-level mutational analysis functions including mutation mapping and quality control, impact prediction and extensive annotation. Two Random Forest filters were employed in CRAVAT for predicting mutation impact, namely Cancer-Specific High-Throughput Annotation of Somatic Mutations (CHASM) and Variant Effect Scoring Tool (VEST). CHASM is a classifier that classifies if a mutation is an oncogenic driver while VEST rates if a mutation is pathogenic or benign.

A p value of $<0.05$ was considered statistically significant. Bioinformatics and statistics analyses were performed and visualized using R version 3.6.1 or Python version 3.7.9.

**Funding**

This work was supported by National Natural Science Foundation of China (82071863).

## Author contributions

XZ designed the workflow, analyzed and illustrated the data, drafted and finalized the manuscript, and supervised all aspects of the study. RC built the data-analysis pipeline. WL processed samples and performed experiments. SZ and YC reviewed the manuscript and provided advice. MJ performed AImmune analysis. GS, YL and WH aided in sample collecting and processing. YX aided in experiment design, supervised experiment, and reviewed the manuscript. SW recruited patients, collected clinical information, supervised all clinical data analysis, and reviewed the manuscript.

## Acknowledgments

We thank all patients and donors who donated blood samples in this study.

## Conflicts of Interest Statement

The authors have declared that no conflict of interest exists.

## Reference

1. Topalian, S.L., J.M. Taube, and D.M. Pardoll, *Neoadjuvant checkpoint blockade for cancer immunotherapy.* Science, 2020. 367(6477).
2. Twomey, J.D. and B. Zhang, *Cancer Immunotherapy Update: FDA-Approved Checkpoint Inhibitors and Companion Diagnostics.* AAPS J, 2021. 23(2): p. 39.
3. Murciano-Goroff, Y.R., A.B. Warner, and J.D. Wolchok, *The future of cancer immunotherapy: microenvironment-targeting combinations.* Cell Res, 2020. 30(6): p. 507-519.
4. Havel, J.J., D. Chowell, and T.A. Chan, *The evolving landscape of biomarkers for checkpoint inhibitor immunotherapy.* Nat Rev Cancer, 2019. 19(3): p. 133-150.
5. De Giglio, A., et al., *The Landscape of Immunotherapy in Advanced NSCLC: Driving Beyond PD-1/PD-L1 Inhibitors (CTLA-4, LAG3, IDO, OX40, TIGIT, Vaccines).* Curr Oncol Rep, 2021. 23(11): p. 126.
6. Sgambato, A., et al., *Anti PD-1 and PDL-1 Immunotherapy in the Treatment of Advanced Non-Small Cell Lung Cancer (NSCLC): A Review on Toxicity Profile and its Management.* Curr Drug Saf, 2016. 11(1): p. 62-8.
7. Mathew, M., et al., *Combining chemotherapy with PD-1 blockade in NSCLC.* Pharmacol Ther, 2018. 186: p. 130-137.
8. Peters, S., K.M. Kerr, and R. Stahel, *PD-1 blockade in advanced NSCLC: A focus on pembrolizumab.* Cancer Treat Rev, 2018. 62: p. 39-49.
9. Santini, F.C. and M.D. Hellmann, *PD-1/PD-L1 Axis in Lung Cancer.* Cancer J, 2018. 24(1): p. 15-19.

10. Roach, C., et al., *Development of a Companion Diagnostic PD-L1 Immunohistochemistry Assay for Pembrolizumab Therapy in Non-Small-cell Lung Cancer.* Appl Immunohistochem Mol Morphol, 2016. 24(6): p. 392-7.
11. Hansen, A.R. and L.L. Siu, *PD-L1 Testing in Cancer: Challenges in Companion Diagnostic Development.* JAMA Oncol, 2016. 2(1): p. 15-6.
12. Maule, J.G., et al., *Comparison of PD-L1 tumor cell expression with 22C3, 28-8, and SP142 IHC assays across multiple tumor types.* J Immunother Cancer, 2022. 10(10).
13. Batenchuk, C., et al., *A real-world, comparative study of FDA-approved diagnostic assays PD-L1 IHC 28-8 and 22C3 in lung cancer and other malignancies.* J Clin Pathol, 2018. 71(12): p. 1078-1083.
14. Trefny, M.P., et al., *A Variant of a Killer Cell Immunoglobulin-like Receptor Is Associated with Resistance to PD-1 Blockade in Lung Cancer.* Clin Cancer Res, 2019. 25(10): p. 3026-3034.
15. Kim, K.H., et al., *PD-1 blockade-unresponsive human tumor-infiltrating CD8(+) T cells are marked by loss of CD28 expression and rescued by IL-15.* Cell Mol Immunol, 2021. 18(2): p. 385-397.
16. Nie, X., et al., *Serum Metabolite Biomarkers Predictive of Response to PD-1 Blockade Therapy in Non-Small Cell Lung Cancer.* Front Mol Biosci, 2021. 8: p. 678753.
17. Araujo, B.d.L.V., et al., *Immune Cell Profiling of Peripheral Blood as Signature for Response During Checkpoint Inhibition Across Cancer Types.* Front Oncol, 2021. 11: p. 558248.
18. Frigola, J., et al., *Molecular profiling of long-term responders to immune checkpoint inhibitors in advanced non-small cell lung cancer.* Mol Oncol, 2021. 15(4): p. 887-900.
19. Hatae, R., et al., *Combination of host immune metabolic biomarkers for the PD-1 blockade cancer immunotherapy.* JCI Insight, 2020. 5(2).
20. Yarchoan, M., A. Hopkins, and E.M. Jaffee, *Tumor Mutational Burden and Response Rate to PD-1 Inhibition.* N Engl J Med, 2017. 377(25): p. 2500-2501.
21. Ricciuti, B., et al., *Association of High Tumor Mutation Burden in Non-Small Cell Lung Cancers With Increased Immune Infiltration and Improved Clinical Outcomes of PD-L1 Blockade Across PD-L1 Expression Levels.* JAMA Oncol, 2022. 8(8): p. 1160-1168.
22. Cercek, A., et al., *PD-1 Blockade in Mismatch Repair-Deficient, Locally Advanced Rectal Cancer.* N Engl J Med, 2022. 386(25): p. 2363-2376.
23. Plesca, I., et al., *Characteristics of Tumor-Infiltrating Lymphocytes Prior to and During Immune Checkpoint Inhibitor Therapy.* Front Immunol, 2020. 11: p. 364.
24. Kim, K.H., C.G. Kim, and E.C. Shin, *Peripheral blood immune cell-based biomarkers in anti-PD-1/PD-L1 therapy.* Immune Netw, 2020. 20(1): p. e8.
25. Li, B., et al., *Computational Deconvolution of Tumor-Infiltrating Immune Components with Bulk Tumor Gene Expression Data.* Methods Mol Biol, 2020. 2120: p. 249-262.
26. Newman, A.M., et al., *Determining cell type abundance and expression from bulk tissues with digital cytometry.* Nat Biotechnol, 2019. 37(7): p. 773-782.
27. Zhang, X., et al., *Cystic Fibrosis Plasma Blunts the Immune Response to Bacterial Infection.* Am J Respir Cell Mol Biol, 2019. 61(3): p. 301-311.
28. Zhang, X., et al., *CFTR-mediated monocyte/macrophage dysfunction revealed by cystic fibrosis proband-parent comparisons.* JCI Insight, 2022. 7(6).
29. Han, R.H. and X.T. Zhang, *AImmune: a new blood-based machine learning approach to improving immune profiling analysis on COVID-19 patients.* medRxiv, 2021: p. 2021.11.26.21266883.
30. Masica, D.L., et al., *CRAVAT 4: Cancer-Related Analysis of Variants Toolkit.* Cancer Res, 2017. 77(21): p. e35-e38.

31. Douville, C., et al., *CRAVAT: cancer-related analysis of variants toolkit.* Bioinformatics, 2013. 29(5): p. 647-8.
32. Banchereau, R., et al., *Molecular determinants of response to PD-L1 blockade across tumor types.* Nat Commun, 2021. 12(1): p. 3969.
33. Wu, C.C., et al., *Prediction of biomarkers and therapeutic combinations for anti-PD-1 immunotherapy using the global gene network association.* Nat Commun, 2022. 13(1): p. 42.
34. Samstein, R.M., et al., *Tumor mutational load predicts survival after immunotherapy across multiple cancer types.* Nat Genet, 2019. 51(2): p. 202-206.
35. Mathew, M., R.A. Safyan, and C.A. Shu, *PD-L1 as a biomarker in NSCLC: challenges and future directions.* Ann Transl Med, 2017. 5(18): p. 375.
36. Tunger, A., et al., *The Evolving Landscape of Biomarkers for Anti-PD-1 or Anti-PD-L1 Therapy.* J Clin Med, 2019. 8(10).
37. Niu, M., et al., *Predictive biomarkers of anti-PD-1/PD-L1 therapy in NSCLC.* Exp Hematol Oncol, 2021. 10(1): p. 18.
38. Ma, Y., et al., *Tumor Mutational Burden and PD-L1 Expression in Non-Small-Cell Lung Cancer (NSCLC) in Southwestern China.* Onco Targets Ther, 2020. 13: p. 5191-5198.
39. Castano, J., et al., *GATA2 Promotes Hematopoietic Development and Represses Cardiac Differentiation of Human Mesoderm.* Stem Cell Reports, 2019. 13(3): p. 515-529.
40. Pan, T., et al., *ANKRD22, a novel tumor microenvironment-induced mitochondrial protein promotes metabolic reprogramming of colorectal cancer cells.* Theranostics, 2020. 10(2): p. 516-536.
41. Qiu, Y., et al., *ANKRD22 is involved in the progression of prostate cancer.* Oncol Lett, 2019. 18(4): p. 4106-4113.
42. Dong, X., et al., *LIM-Homeodomain Transcription Factor LHX4 Is Required for the Differentiation of Retinal Rod Bipolar Cells and OFF-Cone Bipolar Subtypes.* Cell Rep, 2020. 32(11): p. 108144.
43. Kumar, A. and K. Chamoto, *Immune metabolism in PD-1 blockade-based cancer immunotherapy.* Int Immunol, 2021. 33(1): p. 17-26.
44. Yin, J., et al., *ANKRD22 promotes progression of non-small cell lung cancer through transcriptional up-regulation of E2F1.* Sci Rep, 2017. 7(1): p. 4430.
45. Cai, L., et al., *LCE: an open web portal to explore gene expression and clinical associations in lung cancer.* Oncogene, 2019. 38(14): p. 2551-2564.
46. Wu, J., et al., *Gene expression profiles of peripheral blood mononuclear cells from patients with advanced non-small cell lung cancer treated with anti-PD-1 monoclonal antibodies.* Journal of Clinical Oncology, 2019. 37(15_suppl): p. e14107-e14107.
47. Fahrmann, J.F., et al., *Proteomic profiling of lung adenocarcinoma indicates heightened DNA repair, antioxidant mechanisms and identifies LASP1 as a potential negative predictor of survival.* Clin Proteomics, 2016. 13: p. 31.
48. Zhang, Q., et al., *Dishevelled-Associated Activator of Morphogenesis 2 (DAAM2) Predicts the Immuno-Hot Phenotype in Pancreatic Adenocarcinoma.* Front Mol Biosci, 2022. 9: p. 750083.
49. Liu, S.P., et al., *LAMP2 as a Biomarker Related to Prognosis and Immune Infiltration in Esophageal Cancer and Other Cancers: A Comprehensive Pan-Cancer Analysis.* Front Oncol, 2022. 12: p. 884448.
50. Xu, Y., et al., *The association of PD-L1 expression with the efficacy of anti-PD-1/PD-L1 immunotherapy and survival of non-small cell lung cancer patients: a meta-analysis of randomized controlled trials.* Transl Lung Cancer Res, 2019. 8(4): p. 413-428.
51. Arfe, A., et al., *Meta-Analysis of PD-L1 Expression As a Predictor of Survival After Checkpoint Blockade.* JCO Precis Oncol, 2020. 4: p. 1196-1206.

52. Cabeza-Cabrerizo, M., et al., *Dendritic Cells Revisited.* Annu Rev Immunol, 2021. 39: p. 131-166.
53. Schetters, S.T.T., et al., *Monocyte-derived APCs are central to the response of PD1 checkpoint blockade and provide a therapeutic target for combination therapy.* J Immunother Cancer, 2020. 8(2).
54. Mayoux, M., et al., *Dendritic cells dictate responses to PD-L1 blockade cancer immunotherapy.* Sci Transl Med, 2020. 12(534).
55. Peng, Q., et al., *PD-L1 on dendritic cells attenuates T cell activation and regulates response to immune checkpoint blockade.* Nat Commun, 2020. 11(1): p. 4835.
56. Kagamu, H., et al., *CD4(+) T-cell Immunity in the Peripheral Blood Correlates with Response to Anti-PD-1 Therapy.* Cancer Immunol Res, 2020. 8(3): p. 334-344.
57. Zuazo, M., et al., *Systemic CD4 Immunity as a Key Contributor to PD-L1/PD-1 Blockade Immunotherapy Efficacy.* Front Immunol, 2020. 11: p. 586907.
58. Kumagai, S., et al., *The PD-1 expression balance between effector and regulatory T cells predicts the clinical efficacy of PD-1 blockade therapies.* Nat Immunol, 2020. 21(11): p. 1346-1358.
59. Tan, C.L., et al., *PD-1 restraint of regulatory T cell suppressive activity is critical for immune tolerance.* J Exp Med, 2021. 218(1).
60. Landre, T., et al., *Anti-PD-(L)1 for KRAS-mutant advanced non-small-cell lung cancers: a meta-analysis of randomized-controlled trials.* Cancer Immunol Immunother, 2022. 71(3): p. 719-726.
61. Sun, H., et al., *Specific TP53 subtype as biomarker for immune checkpoint inhibitors in lung adenocarcinoma.* EBioMedicine, 2020. 60: p. 102990.
62. Wang, Y., et al., *PTCH1 mutation promotes antitumor immunity and the response to immune checkpoint inhibitors in colorectal cancer patients.* Cancer Immunol Immunother, 2022. 71(1): p. 111-120.
63. Zaretsky, J.M., et al., *Mutations Associated with Acquired Resistance to PD-1 Blockade in Melanoma.* N Engl J Med, 2016. 375(9): p. 819-29.
64. Masuda, K., et al., *Efficacy of anti-PD-1 antibodies in NSCLC patients with an EGFR mutation and high PD-L1 expression.* J Cancer Res Clin Oncol, 2021. 147(1): p. 245-251.
65. Teng, J., et al., *Case Report: PTEN Mutation Induced by anti-PD-1 Therapy in Stage IV Lung Adenocarcinoma.* Front Pharmacol, 2022. 13: p. 714408.
66. Guo, W., et al., *A20 regulates the therapeutic effect of anti-PD-1 immunotherapy in melanoma.* J Immunother Cancer, 2020. 8(2).
67. Hellmann, M.D., et al., *Genomic Features of Response to Combination Immunotherapy in Patients with Advanced Non-Small-Cell Lung Cancer.* Cancer Cell, 2018. 33(5): p. 843-852 e4.
68. Newman, A.M., et al., *Robust enumeration of cell subsets from tissue expression profiles.* Nat Methods, 2015. 12(5): p. 453-7.
69. Li, T., et al., *TIMER: A Web Server for Comprehensive Analysis of Tumor-Infiltrating Immune Cells.* Cancer Res, 2017. 77(21): p. e108-e110.
70. Kamada, T., et al., *PD-1(+) regulatory T cells amplified by PD-1 blockade promote hyperprogression of cancer.* Proc Natl Acad Sci U S A, 2019. 116(20): p. 9999-10008.
71. Liu, S.Y. and Y.L. Wu, *Tislelizumab: an investigational anti-PD-1 antibody for the treatment of advanced non-small cell lung cancer (NSCLC).* Expert Opin Investig Drugs, 2020. 29(12): p. 1355-1364.
72. Desai, J., et al., *Phase IA/IB study of single-agent tislelizumab, an investigational anti-PD-1 antibody, in solid tumors.* J Immunother Cancer, 2020. 8(1).
73. Bu, D., et al., *KOBAS-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis.* Nucleic Acids Res, 2021. 49(W1): p. W317-W325.

74. Liu, X., et al., *Immune and inflammation: related factor alterations as biomarkers for predicting prognosis and responsiveness to PD-1 monoclonal antibodies in cervical cancer.* Discov Oncol, 2022. 13(1): p. 96.
75. Menden, K., et al., *Deep learning-based cell composition analysis from tissue expression profiles.* Sci Adv, 2020. 6(30): p. eaba2619.

**Table 1. General characteristics of NSCLC patients in this study (N=31)**

| Characteristics | Count (percentage) |
|---|---|
| **Gender** | |
| **Male** | 27 (87.1%) |
| **Female** | 4 (12.9%) |
| **Age** | |
| **≥60** | 20 (64.5%) |
| **<60** | 11 (35.5%) |
| **Smoking history** | |
| **Yes** | 25 (80.6%) |
| **No** | 6 (19.4%) |
| **Histology** | |
| **Squamous (SqCC)** | 26 (83.9%) |
| **Adenocarcinoma (ADC)** | 4 (12.9%) |
| **Large cell carcinoma (LCA)** | 1 (3.2%) |
| **Stage** | |
| **□A** | 10 (32.3%) |
| **□B** | 4 (12.9%) |
| **□A** | 1 (3.2%) |
| **□B** | 5 (16.1%) |
| **□A** | 5 (16.1%) |
| **IIIB** | 3 (9.7%) |
| **□A** | 1 (3.2%) |
| **Unable to stage** | 2 (6.5%) |
| **Responder status** | |
| **Responder (pCR)** | 12 (38.7%) |
| **Non-responder (non-pCR)** | 17 (54.8%) |
| **Not evaluated** | 2 (6.5%) |

pCR, pathologic complete response

## Figure legends

**Figure 1. Workflow diagram of the study to identify blood-based signatures.** NSCLC, non-small cell lung cancer; vs., versus; DEGs, differently expressed genes.

**Figure 2. Identification and validation of transcriptomic signatures altered during neoadjuvant anti-PD-1 treatment.** A) Venn diagram and volcano plot of overall differentially expressed gens (DEGs) identified in unpaired and paired comparisons of post- versus pre-treatment blood samples. Unique DEGs identified from pairwise comparison and Common DEGs identified from both pairwise and unpaired comparisons are color-coded as annotated. B) Venn diagram and volcano plot of DEGs identified in comparison of post- versus pre-treatment blood samples from non-responders. DEGs only seen in non-responders (Unique) and DEGs seen in both non-responders and responders (Common) are highlighted. The 8 Common DEGs with their names annotated in Figures 2A and 2B are genes used for qRT-PCR validation. C) Relative mRNA expressions of 8 Common DEGs validated by qRT-PCR and compared to RNA-seq results. The mean fold changes identified from both methods are provided after binary logarithmic conversion ($Log_2$(mean fold change)). Genes HBG1 and HBG2 were combined in qRT-PCR because they were detected by same primer set. D) Relative mRNA expression changes of same DEGs in normal lung and lung cancer tissues (data from GENT2 database). Dot plots and box plots, if shown, represent the individual values and the average values of the mRNA levels. N, normal lung tissue; T, lung cancer tissue; v, versus; *, $p<0.05$; **, $p<0.01$; ***, $p<0.001$.

**Figure 3. Characterization of DEGs identified from responders as gene expression biomarkers.** A) Venn diagrams showing the overlay between DEG subgroups with top ranked transcriptional signatures correlating with PD-1 expression in multiple indications including

NSCLC (blue box) or those responding to anti-PD-1 therapy in melanoma (green box). Unique DEGs were identified from either responders or non-responders while Common DEGs were identified from both patient groups. B) Volcano plot of DEGs identified in comparison of post- versus pre-treatment blood samples from responders. DEGs only seen in responders (Unique) and DEGs seen in both responders and non-responders (Common) are highlighted. The mean fold changes in RNA-seq and RT-PCR are provided (in the format of $Log_2$(mean fold change)). C) Correlation scatter plots of the fold change (log2) of Common DEGs (left, up-regulated; right, down-regulated) between responders (n=14) and non-responders (n=14). The p value and R squared (square of the correlation coefficient) were produced by a Pearson correlation analysis. The linear regression line and its equation were generated from a simple linear regression analysis. D) Sample similarity matrices from Unique or Common DEGs to visualize the mix-ups of 14 paired patients samples and 5 healthy donor samples. E) The hierarchical clustering of all study samples according to their profiles of responder Unique DEGs. The heatmap visualized the relative expression level of each DEG. Pre- and post-treatment status, responder labels are color-coded and annotated.

**Figure 4. Analysis of pathways and GOs that specifically regulated in responders during the treatment.** A) Venn diagram visualization of the significant KEGG pathways, Reactome pathways and gene ontology (GO) items regulated after treatment. The numbers of terms regulated in responders, non-responders and both are provided respectively. B & C) Bubble plots of the top 20 pathways regulated in responders (B) and non-responders (C). Bubble with bigger size stands for smaller p value and higher significance. Names of unique terms are colored in red (responder) or blue (non-responder) while the shared terms are annotated in grey text. Res, responders; Non-res, non-responders.

**Figure 5. GSEA and immune cell abundance profiling to identify key immune cell subsets responding to the treatment.**

A & B) Bubble plots of the top 20 gene sets down-regulated (A) and up-regulated (B) in responders. Bubble with bigger size stands for smaller p value and higher significance. Names of unique terms are colored in red (responder) while the shared terms are annotated in grey text. Bubble with bigger size stands for higher k/K value ratio and larger fraction of gene was matched with a certain reference gene set. C & D) Cell abundance scores as calculated by AImmune (C) and CIBERSORTx (D) of immune cell subsets are significantly changed in both responders and non-responders. E) Cell abundance scores of regulatory T cells (by CIBERSORTx) is only significantly changed in responders but not in non-responders. Dot plots and box plots show individual values and average value of the scores. Connecting lines indicate the pairwise relationship between pre- and post-treatment samples. Res, responders; Non-res, non-responders; HC, healthy control. *, $p<0.05$; **, $p<0.01$. All p values were calculated for pairwise comparisons.

**Figure 6. Identification and validation of blood-derived mutational markers to neoadjuvant anti-PD-1 treatment**

A) Venn diagram showing the overlay between the top 50 or all gene-level mutations identified from blood samples in responders, non-responders and healthy donors. B) Top 10 mutated genes exclusively identified from responders and non-responders and ranked by CHASM and VEST scores. C) Venn diagram comparing the mutated genes from non-responders of NSCLC patients and mutated genes from progression disease patients with cervical cancer (published dataset). The 3 overlaid genes are bold in table showed in B). D) Mutation status of the mutated genes in an independent pan-cancer cohort (TMB and Immunotherapy cohort, n=1661) and the Kaplan

Meier overall survival curves of pan-cancer patients grouped by 8 mutational markers individually. E) The Kaplan Meier overall survival curves of two patient groups defined by 4 responder markers combined as one panel. Curves were generated for all immunotherapy patients (upper) and patients received anti-PD1/PD-L1 therapy only. P value was generated from by Log-Rank test which compares the survival distributions in independent groups.

**Fig. 1**

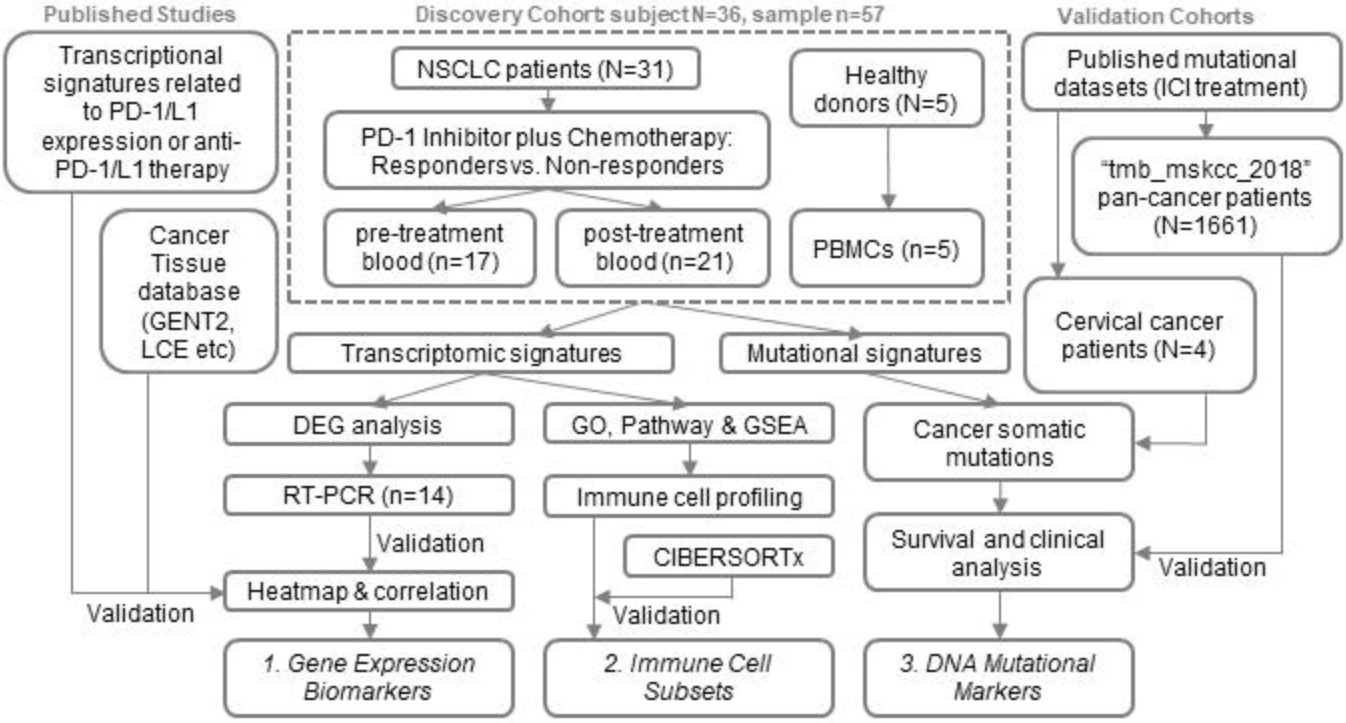

Published Studies
Discovery Cohort: subject N=36, sample n=57
Validation Cohorts
Transcriptional signatures related to PD-1/L1 expression or anti-PD-1/L1 therapy
NSCLC patients (N=31)
Healthy donors (N=5)
Published mutational datasets (ICI treatment)
PD-1 Inhibitor plus Chemotherapy: Responders vs. Non-responders
"tmb_mskcc_2018" pan-cancer patients (N=1661)
Cancer Tissue database (GENT2, LCE etc)
pre-treatment blood (n=17)
post-treatment blood (n=21)
PBMCs (n=5)
Cervical cancer patients (N=4)
Transcriptomic signatures
Mutational signatures
DEG analysis
GO, Pathway & GSEA
Cancer somatic mutations
RT-PCR (n=14)
Immune cell profiling
Validation
CIBERSORTx
Survival and clinical analysis
Validation
Heatmap & correlation
Validation
Validation
1. Gene Expression Biomarkers
2. Immune Cell Subsets
3. DNA Mutational Markers

**Fig. 2**

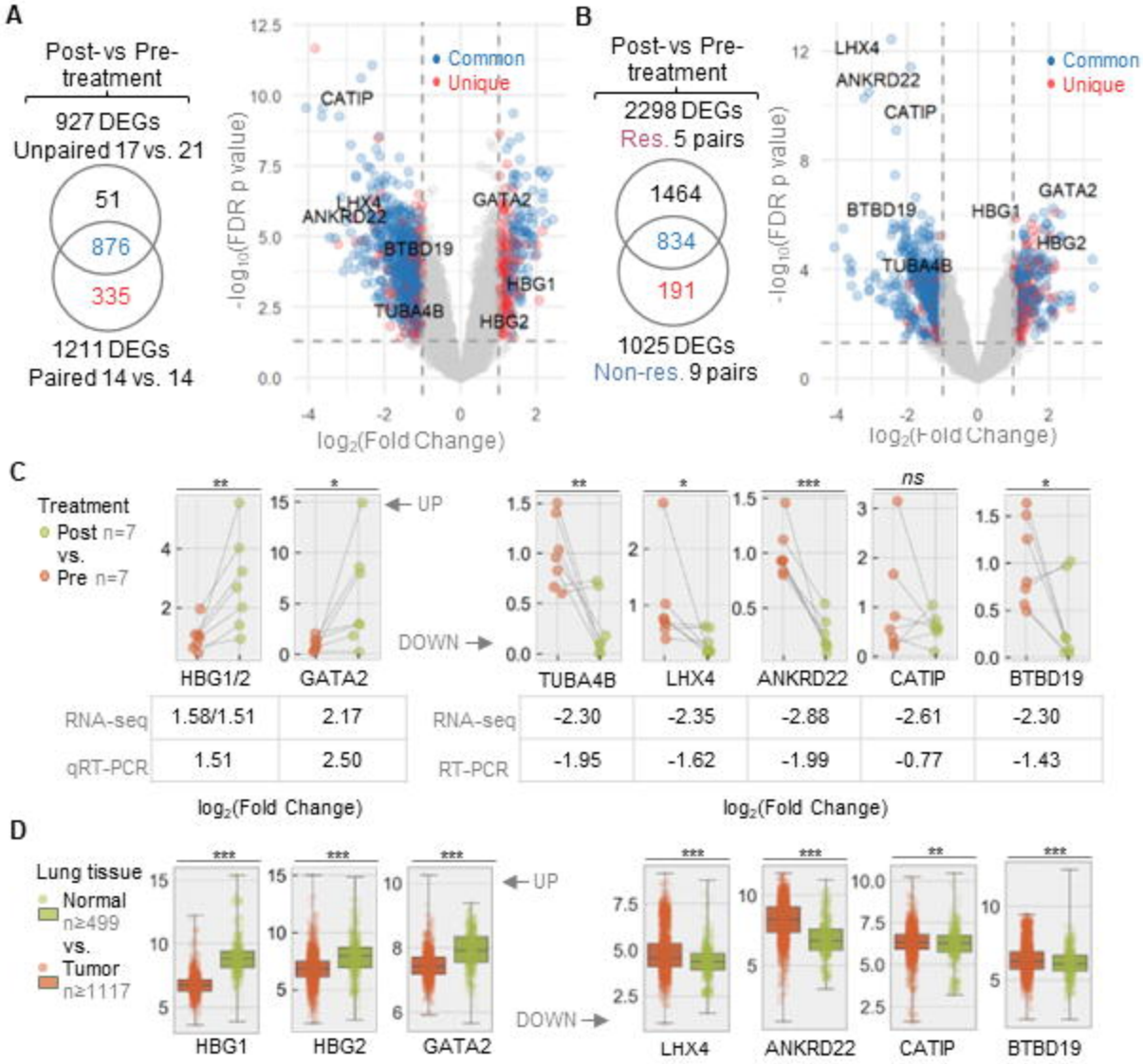
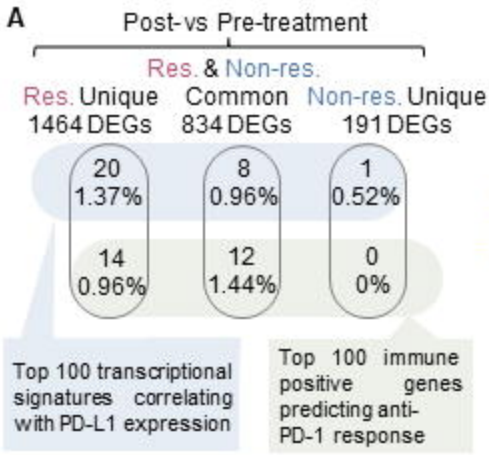
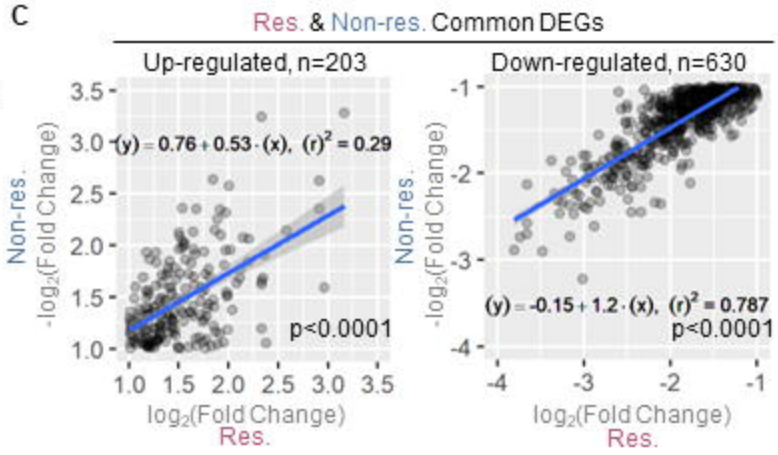
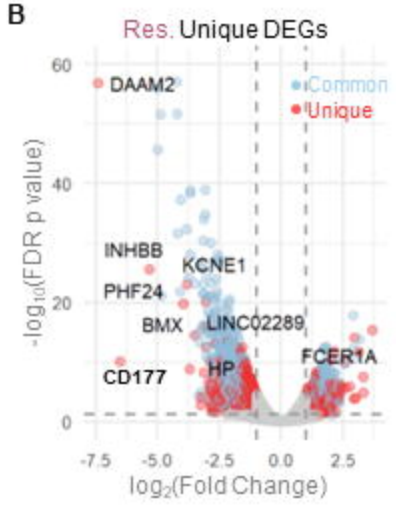

A
Post- vs Pre-treatment
927 DEGs
Unpaired 17 vs. 21
51
876
335
1211 DEGs
Paired 14 vs. 14
Common
Unique
CATIP
LHX4
ANKRD22
BTBD19
TUBA4B
GATA2
HBG1
HBG2
-log10(FDR p value)
log2(Fold Change)
B
Post- vs Pre-treatment
2298 DEGs
Res. 5 pairs
1464
834
191
1025 DEGs
Non-res. 9 pairs
C
Treatment
Post n=7
vs.
Pre n=7
UP
DOWN
HBG1/2
GATA2
TUBA4B
LHX4
ANKRD22
CATIP
BTBD19
RNA-seq 1.58/1.51 2.17
qRT-PCR 1.51 2.50
RNA-seq -2.30 -2.35 -2.88 -2.61 -2.30
RT-PCR -1.95 -1.62 -1.99 -0.77 -1.43
log2(Fold Change)
D
Lung tissue
Normal n≥499
vs.
Tumor n≥1117
HBG1
HBG2
GATA2
LHX4
ANKRD22
CATIP
BTBD19

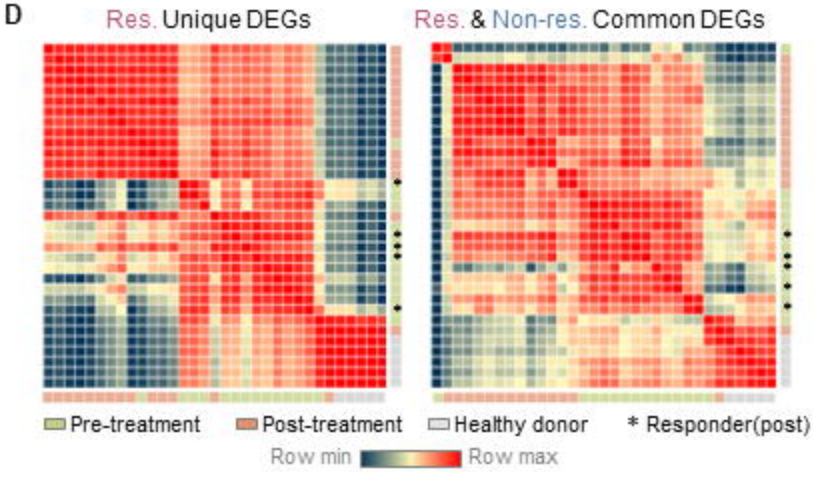
**Fig. 3**




| Res. Post- vs. Pre- treatment | | |
|---|---|---|
| $log_2$(Fold Change) | RNA-seq | qRT-PCR |
| DAAM2 | -7.39 | -5.93 |
| CD177 | -6.50 | -4.14 |
| INHBB | -5.32 | -7.84 |
| PHF24 | -3.93 | -3.60 |
| KCNE1 | -3.78 | -4.27 |



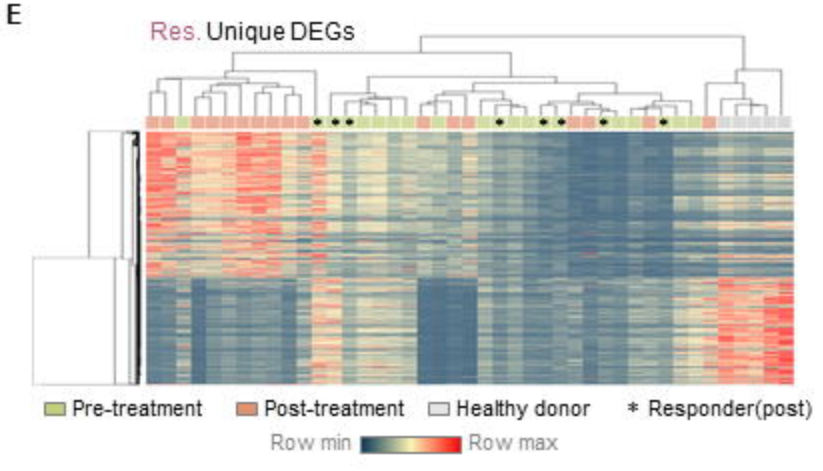



**Fig. 4**

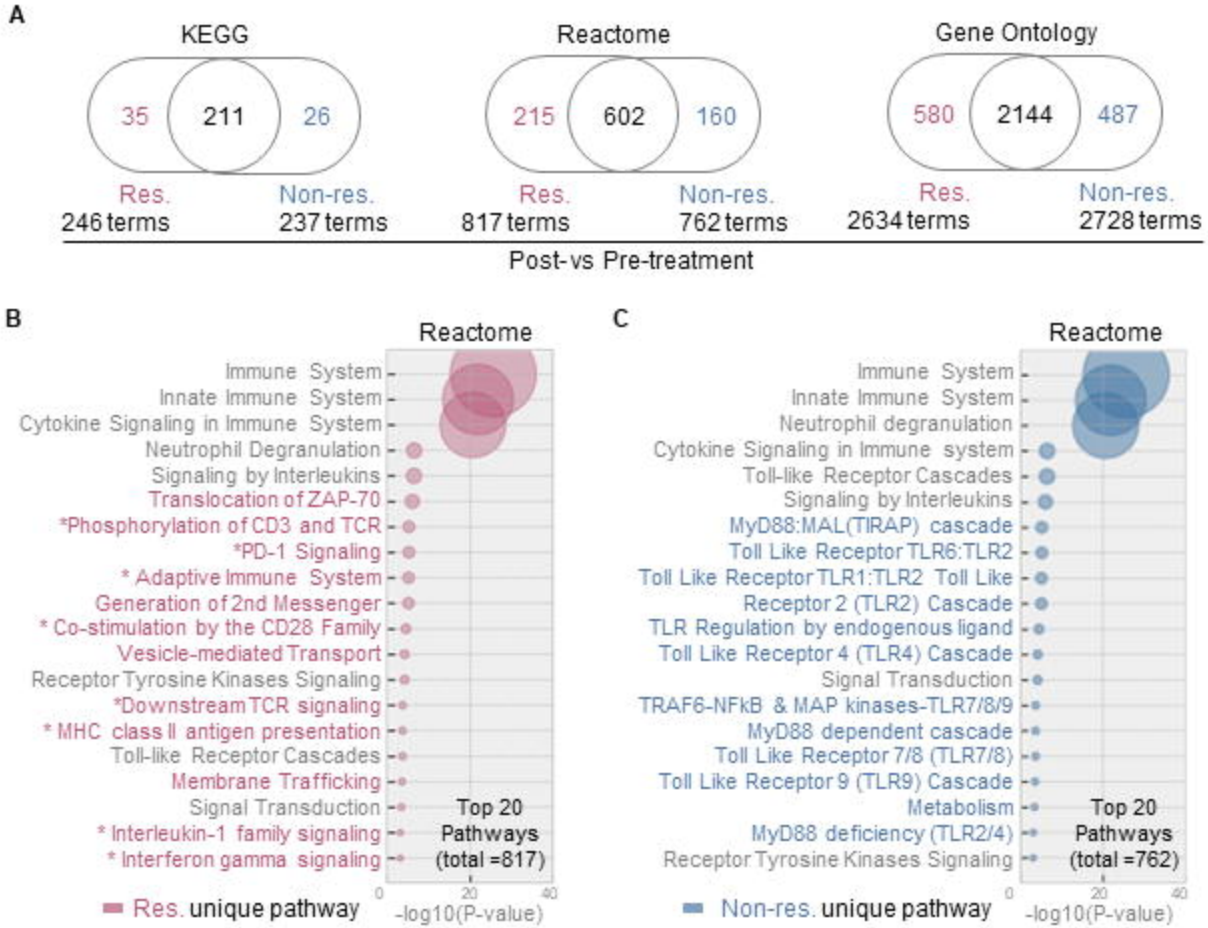

A
KEGG
35
211
26
Res.
246 terms
Non-res.
237 terms
Reactome
215
602
160
Res.
817 terms
Non-res.
762 terms
Gene Ontology
580
2144
487
Res.
2634 terms
Non-res.
2728 terms
Post- vs Pre-treatment
B
Reactome
Immune System
Innate Immune System
Cytokine Signaling in Immune System
Neutrophil Degranulation
Signaling by Interleukins
Translocation of ZAP-70
*Phosphorylation of CD3 and TCR
*PD-1 Signaling
* Adaptive Immune System
Generation of 2nd Messenger
* Co-stimulation by the CD28 Family
Vesicle-mediated Transport
Receptor Tyrosine Kinases Signaling
*Downstream TCR signaling
* MHC class II antigen presentation
Toll-like Receptor Cascades
Membrane Trafficking
Signal Transduction
* Interleukin-1 family signaling
* Interferon gamma signaling
Top 20
Pathways
(total =817)
0
20
40
-log10(P-value)
Res. unique pathway
C
Reactome
Immune System
Innate Immune System
Neutrophil degranulation
Cytokine Signaling in Immune system
Toll-like Receptor Cascades
Signaling by Interleukins
MyD88:MAL(TIRAP) cascade
Toll Like Receptor TLR6:TLR2
Toll Like Receptor TLR1:TLR2 Toll Like
Receptor 2 (TLR2) Cascade
TLR Regulation by endogenous ligand
Toll Like Receptor 4 (TLR4) Cascade
Signal Transduction
TRAF6-NFkB & MAP kinases-TLR7/8/9
MyD88 dependent cascade
Toll Like Receptor 7/8 (TLR7/8)
Toll Like Receptor 9 (TLR9) Cascade
Metabolism
MyD88 deficiency (TLR2/4)
Receptor Tyrosine Kinases Signaling
Top 20
Pathways
(total =762)
0
20
40
-log10(P-value)
Non-res. unique pathway

**Fig. 5**

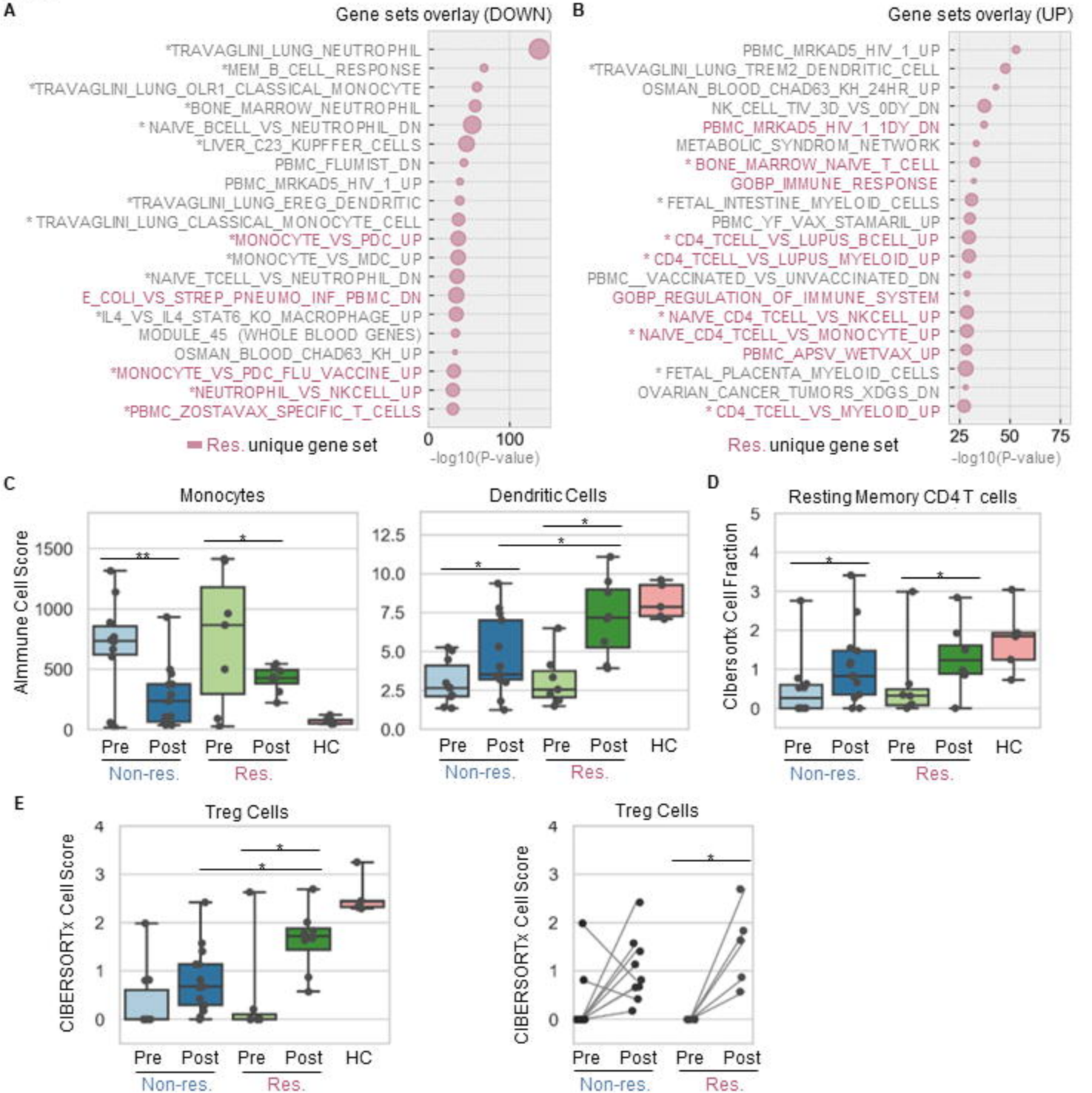
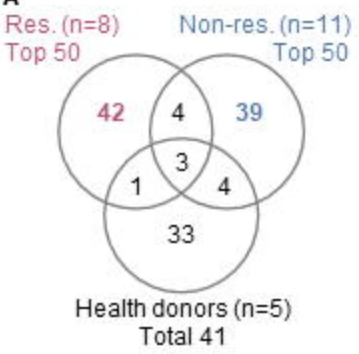

A
Gene sets overlay (DOWN)
*TRAVAGLINI_LUNG_NEUTROPHIL
*MEM_B_CELL_RESPONSE
*TRAVAGLINI_LUNG_OLR1_CLASSICAL_MONOCYTE
*BONE_MARROW_NEUTROPHIL
* NAIVE_BCELL_VS_NEUTROPHIL_DN
*LIVER_C23_KUPFFER_CELLS
PBMC_FLUMIST_DN
PBMC_MRKAD5_HIV_1_UP
*TRAVAGLINI_LUNG_EREG_DENDRITIC
* TRAVAGLINI_LUNG_CLASSICAL_MONOCYTE_CELL
*MONOCYTE_VS_PDC_UP
*MONOCYTE_VS_MDC_UP
*NAIVE_TCELL_VS_NEUTROPHIL_DN
E_COLI_VS_STREP_PNEUMO_INF_PBMC_DN
*IL4_VS_IL4_STAT6_KO_MACROPHAGE_UP
MODULE_45 (WHOLE BLOOD GENES)
OSMAN_BLOOD_CHAD63_KH_UP
*MONOCYTE_VS_PDC_FLU_VACCINE_UP
*NEUTROPHIL_VS_NKCELL_UP
*PBMC_ZOSTAVAX_SPECIFIC_T_CELLS
Res. unique gene set
0
100
-log10(P-value)
B
Gene sets overlay (UP)
PBMC_MRKAD5_HIV_1_UP
*TRAVAGLINI_LUNG_TREM2_DENDRITIC_CELL
OSMAN_BLOOD_CHAD63_KH_24HR_UP
NK_CELL_TIV_3D_VS_0DY_DN
PBMC_MRKAD5_HIV_1_1DY_DN
METABOLIC_SYNDROM_NETWORK
* BONE_MARROW_NAIVE_T_CELL
GOBP_IMMUNE_RESPONSE
* FETAL_INTESTINE_MYELOID_CELLS
PBMC_YF_VAX_STAMARIL_UP
* CD4_TCELL_VS_LUPUS_BCELL_UP
* CD4_TCELL_VS_LUPUS_MYELOID_UP
PBMC__VACCINATED_VS_UNVACCINATED_DN
GOBP_REGULATION_OF_IMMUNE_SYSTEM
* NAIVE_CD4_TCELL_VS_NKCELL_UP
* NAIVE_CD4_TCELL_VS_MONOCYTE_UP
PBMC_APSV_WETVAX_UP
* FETAL_PLACENTA_MYELOID_CELLS
OVARIAN_CANCER_TUMORS_XDGS_DN
* CD4_TCELL_VS_MYELOID_UP
Res. unique gene set
25
50
75
-log10(P-value)
C
Monocytes
Almmune Cell Score
1500
1000
500
0
Pre
Post
Pre
Post
HC
Non-res.
Res.
Dendritic Cells
12.5
10.0
7.5
5.0
2.5
0.0
D
Resting Memory CD4 T cells
Cibersortx Cell Fraction
5
4
3
2
1
0
E
Treg Cells
CIBERSORTx Cell Score
4
3
2
1
0
Treg Cells
CIBERSORTx Cell Score

**Fig. 6**

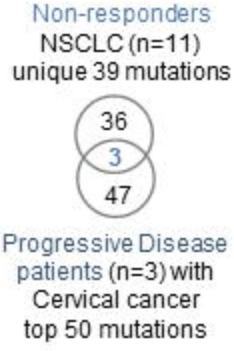
A




B

| Res. | CHASM | VEST | Non-res. | CHASM | VEST |
|---|---|---|---|---|---|
| NFIX | 0.91 | 0.99 | TNFAIP3 | 0.95 | 1.00 |
| RAB3A | 0.91 | 0.99 | TLK2 | 0.92 | 0.93 |
| CASK | 0.90 | 0.96 | BRCA1 | 0.91 | 0.84 |
| JAK2 | 0.89 | 0.93 | ASXL1 | 0.91 | 0.97 |
| CDK3 | 0.88 | 0.95 | GATA2 | 0.90 | 0.89 |
| PTPRS | 0.87 | 0.85 | MYO1C | 0.90 | 1.00 |
| DNMT3A | 0.87 | 0.89 | IDS | 0.90 | 0.97 |
| EMC3 | 0.86 | 0.99 | SRGAP2B | 0.90 | 0.87 |
| PTGER3 | 0.86 | 0.98 | KIAA1109 | 0.89 | 0.89 |
| PTCH1 | 0.86 | 0.98 | CTSB | 0.88 | 0.94 |

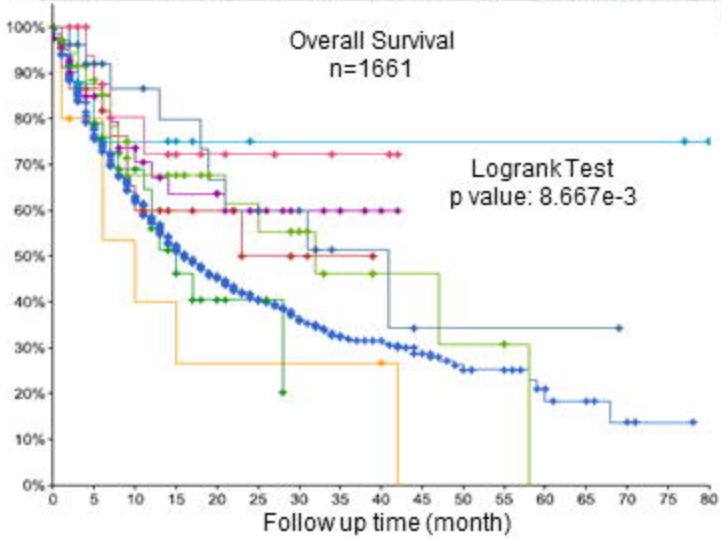
C

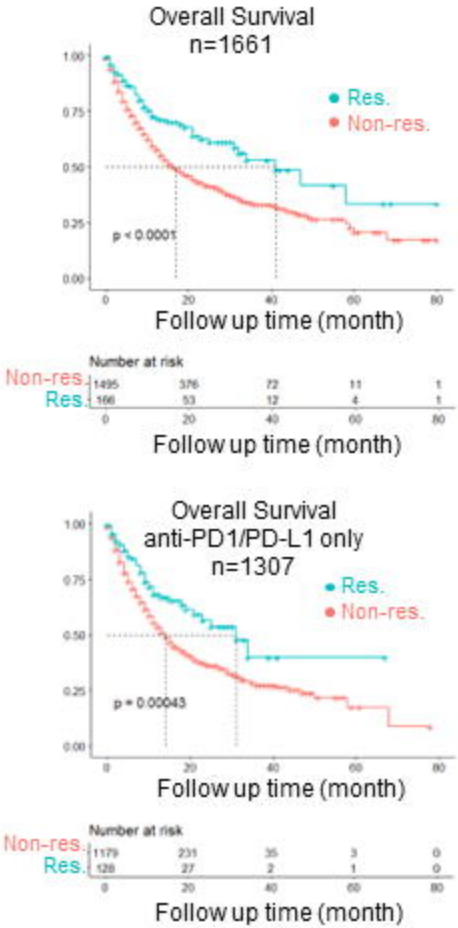



D

Validation with TMB and Immunotherapy cohort

| Mutation | Mutation # | Sample # | % | Patient # | Median Months Overall (95% CI) |
|---|---|---|---|---|---|
| Unaltered | NA | NA | NA | 1393 | 16.00 (14.00-18.00) |
| TNFAIP3 | 28 | 28 | 1.70 | 11 | 10.00 (6.00-NA) |
| BRCA1 | 61 | 56 | 3.40 | 34 | 15.00 (12.00-NA) |
| ASXL1 | 69 | 62 | 3.70 | 41 | NA |
| GATA2 | 16 | 16 | 1.00 | 8 | NA |
| JAK2 | 42 | 35 | 2.10 | 19 | NA |
| PTPRS | 90 | 78 | 4.70 | 37 | 32.00 (21.00-NA) |
| DNMT3A | 45 | 42 | 2.50 | 23 | 23.00 (9.00-NA) |
| PTCH1 | 58 | 50 | 3.00 | 26 | 41.00 (19.00 - NA) |




E