medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Article

Immune evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features

Authors: Hitoshi Dejima[1,17], Xin Hu[2,17], Runzhe Chen[3,17], Jiexin Zhang[4,17], Junya Fujimoto[1,17], Edwin R. Parra[1], Cara Haymaker[1], Shawna Hubert[3], Dzifa Duose[1], Luisa M. Solis[1], Dan Su[5,6], Junya Fukuoka[7], Kazuhiro Tabata[7], Hoa Pharm[7], Nicholas Mcgranahan[8], Baili Zhang[1], Jie Ye[3], Lisha Ying[5,9], Latasha Little[2], Curtis Gumbs[2], Chi-Wan Chow[1], Marcos Roberto Estecio[10,11], Myrna C.B. Godoy[12], Mara B. Antonoff[13], Boris Sepesi[13], Harvey Pass[14], Carmen Behrens[3], Jianhua Zhang[2], Ara A. Vaporciyan[13], John V. Heymach[3], Paul Scheet[15], J. Jack Lee[16], P. Andrew Futreal[2], Alexandre Reuben[3*], Humam Kadara[1*], Ignacio Wistuba[1*], Jianjun Zhang[2, 3,18*]

Affiliations

Departments of [1]Translational Molecular Pathology, [2]Genomic Medicine, [3]Thoracic/Head and Neck Medical Oncology, [4]Bioinformatics & Computational Biology, [10]Epigenetics and Molecular Carcinogenesis, [11]Center of Cancer Epigenetics, [12]Thoracic Imaging, [13]Thoracic and Cardiovascular Surgery, [15]Epidemiology, [16]Biostatistics, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA.

[5]Institute of Cancer and Basic Medicine (IBMC), Chinese Academy of Sciences, Hangzhou, China.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

[6]Department of Pathology, [9]Zhejiang Cancer Research Institute, Cancer Hospital of the University of Chinese Academy of Sciences (Zhejiang Cancer Hospital), Hangzhou, China.

[7]Department of Pathology, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan.

[8]Cancer Research United Kingdom-University College London Lung Cancer Centre of Excellence, London, UK.

[14]Department of Cardiothoracic Surgery, New York University Langone Medical Center, New York, NY 10016, USA.

[17]These authors contributed equally

[18]Lead Contact

[*]Correspondence: AReuben@mdanderson.org (A.R), HKadara@mdanderson.org (H.K.), Iiwistuba@mdanderson.org (I.W.), Jzhang20@mdanderson.org (J.Z.)

SUMMARY

How anti-cancer immunity shapes early carcinogenesis of lung adenocarcinoma (ADC) is unknown. We characterized immune contexture of invasive lung ADC and its precursors by transcriptomic immune profiling, T cell receptor (TCR) sequencing and multiplex immunofluorescence. Our results demonstrated that anti-tumor immunity evolved as a continuum from lung preneoplasia, to preinvasive ADC, minimally-invasive ADC and frankly invasive lung ADC with a gradually less effective and more intensely regulated immune response including down-regulation of immune-activation pathways,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

up-regulation of immunosuppressive pathways, higher infiltration of CD4+ T cells, lower infiltration of CD8+ T cells, decreased T cell clonality, and lower frequencies of top T cell clones in later stages. Driver mutations, HLA loss, chromosomal copy number aberrations and DNA methylation changes may collectively impinge host immune responses and facilitate immune evasion as a potential mechanism underlying outgrowth of the most fit subclones in preneoplasia into dominant clones in invasive ADC.

SIGNIFICANCE

There has been a drastic increase in the detection of lung nodules, many of which are lung ADC precursors. The management of these lung nodules is controversial. We discovered that immune activation and evasion have started at preneoplastic stage and lung ADC precursors may exhibit an overall better-preserved anti-tumor immune contexture suggesting therapeutic strategies reprograming the immune microenvironment in patients with lung ADC precursors prior to further immunosuppression in invasive lung cancers may be beneficial. These findings have served as the critical scientific rationale for our immunoprevention clinical trial IMPRINT-Lung (NCT03634241) recruiting individuals diagnosed with lung nodules at high risk developing invasive lung cancers.

KEYWORDS

Lung adenocarcinoma, premalignancy, atypical adenomatous hyperplasia, adenocarcinoma *in situ*, minimally invasive adenocarcinoma, tumor evolution, early carcinogenesis, immunoediting, immune microenvironment, immunogenomics

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

INTRODUCTION

Despite significant advances in its management, lung cancer remains the leading cause of cancer death worldwide (Siegel et al., 2020; Tan et al., 2016), largely due to late diagnosis at advanced stages when cures are generally unachievable (Barnett, 2017; Borghaei et al., 2015; Siegel et al., 2020). Computed tomography (CT) scan-guided lung cancer screening has demonstrated a reduction of lung cancer mortality by 26%-61% (de Koning et al., 2020) highlighting the importance of early detection and intervention. These findings suggest that early cancer interception is crucial to reduce lung cancer incidence and mortality. Yet, to date, randomized clinical trials on primary lung cancer prevention have only produced disappointing results (de Koning et al., 2020), primarily due to our rudimentary knowledge of early phases in lung cancer development. Improved understanding of targetable molecular mechanisms underlying early lung carcinogenesis may accelerate the development of precise diagnostic as well as effective preventive and therapeutic strategies.

Lung adenocarcinoma (ADC) is the most common histological subtype of lung cancer. Recent studies has postulated that lung ADC may arise from atypical adenomatous hyperplasia (AAH), the only recognized preneoplasia to lung ADC (Aoyagi et al., 2001; Chiosea et al., 2007; Kitamura et al., 1999; Maeshima et al., 2010; Min et al., 2010; Noguchi, 2010; Seki and Akasaka, 2007), which evolves into preinvasive adenocarcinoma *in situ* (AIS) (Weichert and Warth, 2014), to micro-invasive lesion termed minimally invasive adenocarcinoma (MIA) (Aoyagi et al., 2001; Travis et al., 2011) and eventually frankly invasive ADC (Aoyagi et al., 2001; Travis et al., 2011). Early-stage lung ADCs and their precursors usually present as lung nodules with distinct radiologic

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

features called ground glass opacity (GGO). These lung nodules are often referred as indeterminate pulmonary nodules (IPN) without histologic diagnosis as the diagnostic yield from biopsy of GGO-predominate nodules is low and surgery is not the standard of care. This has subsequently led to the scarcity of appropriate materials to study the molecular profiles of lung ADC precursors (Izumchenko et al., 2015).

Carcinogenesis results from progressive accumulation of molecular abnormalities (molecular evolution) (Vogelstein et al., 2013) and escape from host immune surveillance (immunoediting) (Schreiber et al., 2011). Our recent gene expression and genomic pilot studies on lung ADC precursors have demonstrated distinct transcriptomic features (Sivakumar et al., 2017) and progressive genomic evolution along the spectrum of AAH to AIS, MIA and ADC (Hu et al., 2019). However, the extent to which immunoediting sculpts early carcinogenesis of lung ADC and the underlying genomic and epigenetic alterations associated with these immune features still remain to be determined. In the current study, we performed immune gene expression profiling, T cell receptor (TCR) sequencing and multiplex immunofluorescence (mIF) staining on a cohort of resected AAH, AIS, MIA and invasive ADC lesions and paired morphologically normal lung tissues (NL) to delineate the evolution of immune contexture, particularly T cell landscape across different stages of early lung ADC pathogenesis. We further leveraged whole exome sequencing (WES) (Hu et al., 2019) and methylation data (Hu et al., 2020) from the same cohort of IPNs to underscore the genomic and epigenetic alterations that may impinge on these immune features (**Table S1** and **Figure 1**).

RESULTS

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Progressive decrease in overall immunity mirrors evolution from preneoplasia to invasive lung adenocarcinoma

To assess dynamic changes in the immune contexture during early lung carcinogenesis, we performed immune profiling using the nCounter PanCancer Immune Profiling Panel (NanoString), which includes 770 genes from 14 different immune cell types, common checkpoint inhibitors, cancer/testis antigens, and genes covering both the adaptive and innate immune response (Chen et al., 2016) on 47 resected pulmonary nodules (n=9 for AAH, n=11 for AIS, n=21 for MIA, and n=6 for invasive ADC) and paired NL tissues (n=38). There were no differences in age (p=0.55, Kruskal-Wallis test), sex (p=0.31, fisher's exact test) or smoking status (p=0.35, fisher's exact test) between patients with pulmonary nodules of different histologic stages. In total, 291 genes were differentially-expressed (DEGs). Interestingly, changes in the majority of DEGs, regardless of their direction, exhibited a progressive pattern along the spectrum from NL, to AAH, AIS, MIA and ADC (**Table S2**). Examples of progressively increased genes included immune suppressive genes *CD47* (protection of cancer cells from immune cell killing) (Soto-Pantoja et al., 2014), *CD276* (inhibition of immune responses) (Picarda et al., 2016) and *CTLA4* (checkpoint molecule) (Pardoll, 2012), while progressively decreased genes included *ENTPD1* (expressed on tumor-specific T cells) (Bastid et al., 2013), granzyme B (*GZMB*) and perforin 1 (*PRF1*) (two cytotoxic molecules produced by T lymphocytes and natural killer cells (NK cells) (Prakash et al., 2014; Tschopp et al., 1986) (**Figure S1**). Functional pathway analysis of DEGs revealed 26 significantly de-regulated pathways associated with neoplastic evolution from NL to invasive ADC, of which, 23 were down-regulated (**Figure 2A**). On the other hand, all three up-regulated pathways (systemic lupus

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

erythematosus (SLE) in B cell signaling, T cell exhaustion signaling and PARP signaling pathways) could potentially impair immune response (Pan et al., 2020; Pantelidou et al., 2019; Wherry and Kurachi, 2015). These results indicated an overall decreased immunity in later-stage lesions.

We next de-convoluted gene expression profiling data using TIMER (Li et al., 2020) to evaluate changes in immune cell composition. As shown in **Figure 2B**, CD4+ T lymphocyte infiltration progressively increased from NL to invasive ADC. Conversely, infiltration of CD8+ T lymphocytes progressively decreased with neoplastic evolution (**Figure 2C**) leading to significantly higher CD4/CD8 ratio in later-stage lesions (**Figure 2D**). Additionally, B cell infiltration progressively increased along the spectrum from NL to invasive ADC (**Figure 2E**). Of particular interest, all 6 tertiary lymphoid structure (TLS) markers CD19, MS4A1 (CD20), CXCL13, CXCR5, CCR7 and CCL19 (Cabrita et al., 2020) (**Figure S2**) also progressively increased from NL to invasive ADC indicating a progressive TLS aggregation, which may play a role in follicular regulatory T cell-mediated CD8+ T cell exclusion (Wang et al., 2020). Indeed, these TLS markers were positively associated with CD4+ T cell infiltration, but negatively associated with CD8+ T cell infiltration (**Figure S3**). To validate these findings, we applied TIMER to RNAseq data from an independent cohort previously published by our group and observed similar results (**Figure S4**) (Sivakumar et al., 2017).

Taken together, these results suggest that immune evolution progressed as a continuum from preneoplasia to invasive lung ADC with a gradually less effective and more intensely regulated immune response.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Dynamic changes in T cell phenotype and infiltration define histologic stages of early ADC development

We next performed multiplex immunofluorescence (mIF) using antibodies against cytokeratin (CK), CD3, CD8, PD-1, PD-L1, CD68, CD45RO, GZMB and FoxP3 (**Figure 3A, Figure S5** and **Figure S6**) on a subset of pulmonary nodules (n=9 for AAH, n=10 for AIS, n=21 for MIA, and n=6 for invasive ADC) and paired NL (n =7) to assess the dynamic changes of various subtypes of T cells and their interaction with (pre-)malignant cells during early lung carcinogenesis. Densities of infiltrating activated CTLs (CD3+CD8+GZMB+) and regulatory T cells (Treg, CD3+CD8-FoxP3+) assessed by mIF were positively correlated with corresponding CD8+ and CD4+ T cell infiltrate levels derived from gene expression profiling, respectively (**Figure S7A-B**). Furthermore, Treg/activated CTL ratio closely recapitulated CD4/CD8 ratio inferred from immune gene expression profiling (**Figure S7C**). Correlations among immune subsets and between immune components and epithelial cells across IPNs of different histologic stages demonstrated that in normal tissues, densities of CK+ and CK+PD-L1+ cells were positively correlated, as were Tregs and memory T cells, cytotoxic and activated cytotoxic T cells (**Figure S8A**). For AAH, only total macrophages and activated cytotoxic T cells were positively correlated (**Figure 3B**). In AIS, PD-L1+ macrophages were positively correlated with Tregs, antigen-experienced T cells and cytotoxic T cells (**Figure 3C**). For MIA, CK cells were negatively correlated with total macrophages and memory T cells, while Tregs were positively correlated with antigen-experienced T cells and memory T cells were correlated with cytotoxic and activated cytotoxic T cells (**Figure 3D**). Finally, in ADC, PD-L1+CK+ cells were correlated with total and PD-L1+ macrophages, PD-L1+

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

macrophages were correlated with total macrophages, and Tregs were correlated with memory T cells (**Figure S8B**). Overall, relationships between cell types varied widely across stages, highlighting the dynamic nature of tumor-immune interactions during tumor evolution.

Progressively divergent TCR repertoire with neoplastic progression

Because of the central role of T cells in tumor surveillance, we next sought to investigate the T cell repertoire (Shah et al., 2011) by multiregional T cell receptor (TCR) sequencing on 13 AAH, 11 AIS, 23 MIA and 10 ADC as well as 49 NL. We first assessed the T cell diversity using Inverse Simpson Index (Kaplinsky and Arnaout, 2016). T cell diversity was positively correlated with infiltration of CD4+ T cells derived from gene expression profiling (**Figure S9A**) as well as densities of Tregs from mIF (**Figure S9B**) and increased steadily from AAH to AIS, MIA and ADC ($p<0.0001$) **(Figure 4A)** in line with higher CD4+ T cells in later stage lesions **(Figure 2B)**. We then compared the distributions of T cell clonality across histologic stages, focusing on the most expanded T cell clones (Mansfield et al., 2018; Wang et al., 2019). As shown in **Figure 4B** and **Figure S10**, the top 10, 100, 200, and 500 clones accounted for a higher frequency in the NLs, and gradually decreased in AAH, AIS, MIA and invasive ADC ($p<0.0001$) implying suppressed T cell expansion during early carcinogenesis of lung ADC.

Suppression of the T cell repertoire may occur at the preneoplastic stage

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

To further understand the T cell response at different histologic stages, we calculated T cell clonality, a metric indicating expansion and activation of T cell clones. T cell clonality was positively correlated with infiltration of CD8+ T cells (r=0.561, p=4.863e-06), activated CTLs (r=0.318, p=7.340e-03), *GZMB* expression (r=0.319, p=5.068e-02) and Th1 cytokines (r=0.473, p=3.923e-05) (**Figure S11A-D**), while negatively correlated with infiltration of CD4+ T cells (r=-0.385, p=8.993e-04) and Tregs (r=-0.477, p=2.965e-05) (**Figure S11E-F**) suggesting that T cell clonality was mainly driven by clonal expansion of activated CTLs, with Tregs inhibiting CTL responses. Comparing T cell clonality among different stages revealed the highest clonality in NL **(Figure 4C)** consistent with our previous findings (Reuben et al., 2020). T cell clonality declined from NL to AAH increasing in AIS/MIA and declining to its lowest level in ADC **(Figure 4C).** These results suggest that early ADC pathogenesis is associated with local immunosuppression in TCR repertoire that may have commenced at the preneoplastic stage.

Driver mutations affect the immune response in pre/early ADC

We next sought to explore the molecular features associated with the immune contexture observed in these lesions. We first explored the two most frequently mutated driver genes in this cohort *EGFR* and *KRAS* (Hu et al., 2019). Compared to *EGFR*-mutant lesions or double wild-type lesions, *KRAS*-mutant lesions exhibited highest CD8+ T cell infiltration and lowest CD4/CD8 T cell ratio inferred from immune gene expression; highest CTL infiltration, highest effector memory cytotoxic T cell infiltration, highest CTL/Treg ratio from mIF as well as highest T cell clonality measured by TCR sequencing **(Figure S12)**. These trends remained similar when analyzing each histologic stage (**Figure S13**). These

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

findings are consistent with our previous studies in invasive lung cancers (Reuben et al., 2020) and emphasize the interplay between oncogene mutations and immune surveillance during early pathogenesis of lung adenocarcinoma.

Chromosomal copy number variations and HLA loss may contribute to impaired T cell responses

In light of recent studies suggesting immune evasion could be facilitated by an inability to present neoantigens due to loss of HLA (McGranahan et al., 2017), we applied the LOHHLA (loss of heterozygosity in HLA) algorithm (McGranahan et al., 2017) to WES data from these lesions (Hu et al., 2019). LOH at the HLA loci was observed in 7%, 15% and 33% of AIS, MIA and ADC respectively, but in none of AAH lesions **(Figure S14A,** p=0.005, Chi-Square test). Correlation with T cell infiltration derived from immune gene expression profiling demonstrated that lesions with HLA LOH had similar CD4+ T cell infiltration, but trends of lower CD8+ T cell infiltration and higher CD4/CD8 ratio although the difference did not reach statistical significance **(Figure S14B-D**).

Chromosomal copy number variation (CNV) has been reported to impact immune microenvironment across different cancers (Davoli et al., 2017). We next investigated whether CNV affected the immune landscape of these lung ADC precursors. Correlation with T cell infiltration derived from gene expression profiling demonstrated burden of allelic imbalance (AI), a subtle form of CNV, was negatively correlated with CD8+ T cell infiltration, but positively correlated with CD4+ T cell infiltration and CD4/CD8 T cell ratio **(Figure S15A-C**). Similar trend was also observed with CNV burden, although these

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

differences did not reach statistical significance **(Figure S15D-F**). Interestingly, HLA-LOH positive lesions exhibited significantly higher AI burden and CNV burden compared to that of HLA-LOH negative lesions **(Figure S16**). One plausible explanation is that the development AI, CNV or HLA LOH resulted from chromosomal instability (CIN) and cell clones with CIN may lead to higher CNV/AI burdens as well as increased likelihood of HLA loss, which could subsequently enable these cells escaping from anti-tumor immune surveillance and developing into dominant clones in the later-stage of neoplastic lesions.

Methylation status interacts with genomic alterations and impacts the immune response during early lung cancer pathogenesis

Somatic mutations play central roles in activating anti-tumor immune responses through generating neoantigens that can be recognized by T cells (McGranahan et al., 2016; Yarchoan et al., 2017). As shown in **Figure S17**, a progressive increase in total mutation burden was observed from AAH to AIS, MIA and ADC. However, analysis of methylation data from reduced representation bisulfite sequencing (RRBS) of the same cohort of IPNs (Hu et al., 2020) revealed that a significantly larger proportion of genes exhibited promoter hypermethylation (>30% CpG methylated) in later-stage lesions, thus potentially dampening the expression of neoantigens in later stage lesions. These data suggest that promoter hypermethylation could contribute to neoantigen depletion and immune escape. Furthermore, we assessed the impact of global methylation status, using long interspersed transposable elements-1 (LINE-1) as a surrogate marker (Kankava et al., 2018; Ohka et al., 2011; Saito et al., 2010), on the immune microenvironment in these lesions. Overall, global methylation level was negatively associated with CD4+ T cell

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

infiltration and CD4/CD8 ratio derived from gene expression profiling, Tregs infiltration and Treg/CD8 ratio from mIF (**Figure S18**) suggesting decreased global methylation was associated suppressive immune contexture.

DISCUSSION

Cancer evolution is shaped by the interaction between cancer cells and host immune surveillance, a process termed immunoediting consisting of elimination, equilibrium, and escape phases (Dunn et al., 2002; Dunn et al., 2004a; Dunn et al., 2004b). It is well documented that the majority of human cancers are infiltrated with various immune cells, but often in an immunosuppressive microenvironment as remnant evidence of immunoediting (Mittal et al., 2014; O'Donnell et al., 2019). T cell immunity is significantly compromised even in stage I lung cancers (Lavin et al.; Reuben et al., 2020) indicating that these cancers have already begun evading immune surveillance. However, when and how the elimination and equilibrium phases occur over the lung cancer evolution continuum is unknown. Therefore, investigating the molecular and immune landscape of lung premalignancies is warranted to elucidate the timing of immune activation/evasion and its underlying molecular mechanisms during early lung carcinogenesis.

In the current study, we leveraged a relatively large cohort of resected lung ADC precursors with available genomic and epigenetic profiling data to characterize the dynamic changes in the immune contexture across different consecutive developmental stages by gene expression profiling, mIF and TCR sequencing. To the best of our knowledge, this is the first systemic and comprehensive study on immune evolution

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

during early carcinogenesis of lung ADC and its potential underlying molecular basis. Overall, there was a more suppressive and tightly controlled immune response, particularly T cell response, in later stage lesions highlighting that the dynamic interaction between cancer cells and host immune surveillance was overall evolving toward immune escape along with pathogenesis of lung ADC. Notably, dynamic changes in immune contexture progressed as a continuum from preneoplasia AAH to pre-invasive AIS, to micro-invasive MIA and finally frankly invasive ADC without obvious "step-wise" major leaps at transitions between different histologic stages. This is consistent with the overall progressive genomic evolution from AAH to AIS, MIA and ADC (Hu et al., 2019) suggesting that early carcinogenesis of lung ADC is a gradual process shaped by continuous interaction with host immune surveillance. Importantly, dynamic immune activation and suppression have already started at preneoplastic stage. This is in line with a similar study on lung squamous cell carcinoma (SCC) precursors by Mascaux and colleagues, which revealed evidence of both immune activation and evasion along with evolution of preinvasive SCC lesions (Mascaux et al., 2019). Taken together, these data advocate for therapy targeting the immune microenvironment in patients with lung cancer precursors to prevent development of invasive lung cancers.

Identification of the molecular basis underlying immune activation and evasion may provide novel insights for understanding tumor-immune interactions and establishing biomarkers to select patients who may benefit from immunoprevention. WES and RRBS data available from the same cohort of pulmonary nodules provided the opportunity to depict the genomic and epigenetic features associated with the distinct immune landscape (Hu et al., 2020; Hu et al., 2019). Overall, oncogene mutations, HLA loss, CNV

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

burden, AI burden and methylation status were all found to associate with the immune contexture, similarly to advanced cancers (Fridman et al., 2017; Thorsson et al., 2018). For example, *KRAS* mutations correlated with more active immune response, while *EGFR* mutations, HLA loss, higher CNV burden and decreased global methylation status correlated with a cold immune microenvironment. However, associations with any single genomic or epigenetic feature were weak, suggesting heterogenous but convergent evolution towards immune escape during early lung carcinogenesis.

Our previous work suggested that early lung ADC carcinogenesis may predominantly follow the clonal sweep model, whereby certain subclones in early-stage preneoplasia turn into dominant clones in later-stage diseases while unfit subclones are eliminated (Hu et al., 2019), primarily by the host immune system. In the early phases of carcinogenesis, stochastic genomic and epigenetic alterations lead to heterogeneous subclones with various combinations of molecular features that define the distinct biology of each subclone including survival ability under selective immune pressure. The impact of genomic and epigenetic alterations on immune response is intertwined. For example, methylation may directly affect immune response by regulating expression of immune genes (Liu et al., 2017) or potential neoantigens (Serrano et al., 2011) or indirectly by increasing DNA vulnerability for development of CNV and somatic mutations that can subsequently influence the immune microenvironment (Bakhoum and Cantley, 2018; Porta-Pardo and Godzik, 2016). Global hypomethylation is known to associate with CIN (Eden et al., 2003) and increased rate of somatic mutations (Chen et al., 1998) and indeed, global methylation level was negatively associated with TMB and CNV burden in this cohort of lesions (Hu et al., 2020). Since high CNV burden is associated with a cold

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

immune microenvironment (Davoli et al., 2017) while high TMB can increase tumor immunogenicity and facilitate immune recognition and elimination of cancer cells (Jung et al., 2019), the impact of global hypomethylation (associated with both high CNV burden and TMB) on anti-cancer immunity is further complicated. In the end, the outgrowth of subclones is determined by the accumulated effects of all molecular aberrations. Only the cells with the ideal combination of molecular features enabling their rapid proliferation and escape from immune attack will survive and outgrow into the dominant clones in invasive cancers. For example, though invasive lung ADCs tend to harbor a high mutation burden that could lead to more active anti-tumor immune response, they also exhibited a higher CNV burden, higher likelihood of losing HLA and decreased global methylation, all of which are associated with a cold tumor immune microenvironment (Jeschke et al., 2017; Neal et al., 2018; Paulson et al., 2018; Rosenthal et al., 2019), resulting in an overall cold immune microenvironment in invasive ADCs (lower CD8+ T cell infiltration, CTLs, and T cell clonality, with higher CD4+ T cell infiltration and Tregs).

With the increasing implementation of CT-guided screening and advent of high-resolution diagnostic CT scans, there has been a drastic increase in the detection of IPNs (Bellomi et al., 2010; McWilliams et al., 2013), many of which are lung ADC precursors. However, their management remains controversial. Although surgical resection could potentially offer cure in a large proportion of these patients, surgery-associated morbidity and high cost have called surgical resection into question. Additionally, ~25% patients often present with multifocal diseases, which complicates surgery. We have previously demonstrated (Hu et al., 2019; Sivakumar et al., 2017) that preinvasive lung ADC precursors were molecularly simpler, and therefore theoretically easier to eradicate. In

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the current study, we showed that lung ADC precursors may exhibit an overall better-preserved anti-tumor immune landscape. In light of these findings, therapeutic strategies reprograming the tumor immune microenvironment in patients with lung ADC precursors prior to further immunosuppression in invasive lung cancers may be beneficial. The lung cancer immunoprevention clinical trial IMPRINT-Lung (NCT03634241) recruiting individuals with high-risk IPNs is currently underway to validate this hypothesis.

ACKNOWLEDGEMENTS

This study was supported by the MD Anderson Khalifa Scholar Award, the National Cancer Institute of the National Institute of Health Research Project Grant (R01CA234629-01), the AACR-Johnson & Johnson Lung Cancer Innovation Science Grant (18-90-52-ZHAN), the MD Anderson Physician Scientist Program, the MD Anderson Lung Cancer Moon Shot Program, Sabin Family Foundation Award, Duncan Family Institute Cancer Prevention Research Seed Funding Program, the Cancer Prevention and Research Institute of Texas Multi-Investigator Research Award grant (RP160668) and the UT Lung Specialized Programs of Research Excellence Grant (P50CA70907), Cancer Prevention and Research Institute of Texas (CPRIT) grant RP150079. We thank Sally Boyd, Jinzhen Chen, and Rong Yao, Eric Sisson and Stan Bujnowski for providing excellent technical support for high-performance cluster resource http://hpcweb.mdanderson.edu/citing.html. We thank MD Anderson Cancer Center's Epigenomics Profiling Core and its Science Park Next-Generation Sequencing Core (supported by CPRIT Core Facility Support Award #RP120348) for performing RRBS profiling.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

AUTHOR CONTRIBUTIONS

J.Z., I.W., H.K., and A.R. conceived, designed, and directed the study. H.D., X.H., R.C., J.F., L.L., C.G., C.C., and J.Y. performed the experiments. X.H., Jiexin Z., N.M., S.H., J.L., and Jianhua Z. conducted bioinformatic analyses. J.F., E.P., C.H., D.D., L.S., D.S., J.F., H.P., B.Z., L.Y., performed pathologic assessment and tissue processing. M.E., M.G., M.A., B.S., H.P., C.B., A.V., J.H., P.S., and P.A.F participated in data interpretation and clinical correlation. J.Z., R.C., H.D., H.K., A.R., and X.H. wrote the manuscript with comments from all authors.

DECLARATION OF INTERESTS

Dr. Zhang reports research funding from Merck, Johnson and Johnson, and consultant fees from BMS, Johnson and Johnson, AstraZeneca, Geneplus, OrigMed, Innovent outside the submitted work. The other authors declare no competing financial interests. Dr. Kadara reports funding form Johnson and Johnson and from Janssen pharmaceuticals.

FIGURE LEGEND

Figure 1. The immune evolution from preneoplasia to invasive lung adenocarcinoma and associated genomic and epigenomic features. Infiltration of B cells, CD4+ T cells, CD8+ T cells inferred from immune gene expression using TIMER; regulatory T cells (Treg) and activated cytotoxic T lymphocytes (CTL) measured by

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

multiplex immunofluorescence (mIF); T cell clonality and frequency of the top 100 T cell clones by TCR sequencing are shown in upper panel. Genomic alterations from whole exome sequencing (WES) including *EGFR*/*KRAS* mutations, HLA loss, copy number variation (CNV) burden, allelic imbalance (AI) burden, total number of mutations associated with predicted neoantigens, total number of mutations associated with predicted neoantigens without promoter methylation; global methylation status accessed by reduced representation bisulfite sequencing (RRBS) are shown in bottom panel. AAH: typical adenomatous hyperplasia. AIS: adenocarcinoma in situ. MIA: minimally invasive adenocarcinoma. ADC: invasive adenocarcinoma.

Figure 2. Progressively changes of immune cell infiltrations from preneoplasia to invasive lung adenocarcinoma. (**A**) Significantly enriched functional pathways based on the 291 differentially expressed genes by Ingenuity pathway analysis (IPA®; Ingenuity Systems) software. Pathways with –log (p-value) > 10 (p-values are obtained from Fisher's right-tailed exact test) and an absolute z-score > 0.5 are shown. Pathways that were predicted to be inhibited (negative Z scores) in later stages are in blue and pathways that were predicted to be activated (positive Z scores) in later stages are in orange. The heights of the bars indicate the significance of the enrichment (-log (p-value)) and the scales of the orange or blue colors represent the predicted directionality. Fractions of immune cells including CD4+ T cells (**B**), CD8+ T cells (**C**), CD4/CD8 ratio (**D**) and B cells (**E**) were estimated using TIMER based on the gene expression using nCounter PanCancer Immune Profiling Panel. Error bars indicate 95% confidence intervals and solid point represent mean value in each stage. The difference of cell fraction among different stages was evaluated using the Kruskal-Wallis H test. NL: Normal lung tissue,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

AAH: Atypical adenomatous hyperplasia, AIS: Adenocarcinoma in situ, MIA: Minimally invasive adenocarcinoma, ADC: Invasive adenocarcinoma.

Figure 3. Dynamic changes of various immune cells and their interaction with epithelial cells across preneoplasia to invasive lung adenocarcinoma. (**A**) Representative multiplex immunofluorescence (mIF) images each histologic stage analyzed by panel 1 and panel 2 markers. The correlation between immune cell subtypes and CK+ epithelial cells measured by mIF in AAH (**B**), AIS (**C**) and MIA (**D**). Significant correlation was marked with * ($p<0.05$). Red circles indicate positive correlations and blue circles indicate negative correlations. The size of circles is proportional to the spearman's correlation co-efficient between each pair of cells. AAH: Atypical adenomatous hyperplasia, AIS: Adenocarcinoma in situ, MIA: Minimally invasive adenocarcinoma.

Figure 4. Dynamic change in T cell repertoire from preneoplasia to invasive lung adenocarcinoma. T cell diversity (**A**) and T cell clonality (**C**) in normal lung, AAH, AIS, MIA and invasive ADC. (**B**) Distribution of T cell clones with frequency of top 1 (brown), top 2 to 10 (black), top 11 to 100 (orange), top 101 to 200 (purple), top 201 to 500 (green), top 501 to 1000 (red) and beyond 1000 (blue) in normal lung, AAH, AIS, MIA and invasive ADC lesions.

STAR METHODS

Patients and tissue processing

Specimens were collected from 53 patients presenting with pulmonary nodules, who underwent surgical resection at Nagasaki Hospital (Japan) or Zhejiang Cancer Hospital

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(China) from 2014 to 2017 as described previously (Hu et al., 2019). The study was approved by the Institutional Review Boards (IRB) at MD Anderson Cancer Center, Nagasaki University Graduate School of Biomedical Sciences and Zhejiang Cancer Hospital. Hematoxylin and eosin (HE) slides of each case were reviewed by experienced lung cancer pathologists to confirm the diagnosis before further analyses.

DNA and RNA extraction

After pathologic assessment, manual macro-dissection was performed to enrich premalignant cells or cancer cells for DNA or RNA extraction. DNA or RNA was isolated using the AllPrep® DNA/RNA FFPE Kit (Qiagen, Hilden, Germany) according to manufacturer's instructions. Finally, the DNA samples were quantified by NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and RNA was quantified using the RNA High sensitivity kit on the Qubit 3.0 fluorometer (Thermo Fisher Scientific, USA). RNA quality and integrity were evaluated with RNA integrity number (RIN) (Schroeder et al., 2006), concentration (ng/µl), and size (nt) using RNA Screen Tape in 4200 Tape Station System (Agilent Technologies, USA).

Gene profiling of immune cells using nCounter® platform

The nCounter® PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA, USA), which contains 770 genes (including 730 immune associated genes and 40 housekeeping genes) using NanoString nCounter Analysis was applied for gene expression profiling as previously described (Cesano, 2015; Kulkarni, 2011). The data were imported into the nSolver 4.0 software (NanoString Technologies) for quality control (QC). After the QC was performed with default setting, background correction was done

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

with negative controls, and data was normalized by using the geometric mean of the 6 positive controls, and 40 housekeeping genes. We further performed quantile-normalization and log2 transformation to stabilize the variance. A one-way ANOVA test was applied to identify differentially expressed genes (DEGs) in different stages. We modeled the p-values using a beta-uniform mixture (BUM) model, combined with false discovery rate (FDR) to determine a cutoff for p-values (Pounds and Morris, 2003). The DEGs, along with log ratios were then evaluated with Ingenuity Pathway Analysis (IPA) software (Quiagen, Hilden, Germany) (Kramer et al., 2014) to identify pathways that are enriched by these DEGs. We performed a Core Analysis with species set to human and tissue set to lung. IPA identifies the top canonical pathways associated with the list of DEGs by the Fisher's exact test to ascertain enrichment. IPA also calculates a z-score to predict activation status of the pathway by comparing input genes and the stored activity pattern.

Multiplex immunofluorescence staining

All samples were confirmed to be appropriate for multiplex immunofluorescence (mIF) analysis based on tissue quality and availability using the HE staining of same lesion by the experienced pathologist. The manual mIF staining was performed on unstained slides of FFPE samples using the Opal 7-Color fIHC Kit (Akoya Biosciences, USA) as previously described (Parra et al., 2018). Eight immune markers were placed in two 6-antibody panels before the stained slides were scanned by Vectra multispectral microscope (Akoya Biosciences, USA). Panel 1 contained pancytokeratin (AE1/AE3; epithelial marker; dilution 1:300; Dako, Carpinteria, CA), PD-L1 (clone E1L3N, dilution 1:100; Cell Signaling Technology, Beverly, MA), PD1 (clone EPR4877-2, dilution 1:250; Abcam, Cambridge,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

MA), CD3 (T lymphocyte marker; dilution 1:100; Dako), CD8 (cytotoxic T cell marker; clone C8/144B, dilution 1:20; Thermo Fisher Scientific, Waltham, MA), and CD68 (macrophage marker; clone PG-M1, dilution 1:450; Dako). Panel 2 contained pancytokeratin, CD3, CD8, CD45RO (memory T cell marker; clone UCHL1, ready to use; Leica Biosystems, Buffalo Grove, IL), Granzyme B (cytotoxic lymphocyte marker; clone F1, ready to use; Leica Biosystems), and FoxP3 (regulatory T cell marker; clone 206D, dilution 1:50; BioLegend, SanDiego, CA). Human tonsil FFPE tissues were used with each individual antibody on same fluorochrome assay to build the spectral library, and were also used with and without primary antibodies as positive and negative (autofluorescence) controls, respectively. The stained slides were scanned with Vectra 3.0 multispectral microscope system (Akoya Biosciences, USA) under fluorescent illumination. These mIF assays followed our previous publication (Francisco-Cruz et al., 2020).

Multispectral analysis

The five individual fields for multispectral analysis were selected from areas of interest in a scanned low magnification (×10) image on Phenochart1.0.9 (Akoya Biosciences, USA). The five high magnification (×20) fields for mIF analysis were carefully selected by experienced pathologists after comparing with HE slides to capture malignant and premalignant cell cluster and various elements of heterogeneity. The corresponding normal fields were selected in the farthest field of tumor periphery with morphological normal tissue on the same slide (**Figure S5A**). The selected high magnification field was total 1.6725 mm^2 in size (669×500μm per field) (**Figure S5B**). Each field with panel 1 and panel 2 were overlapped with sequential sections. The target areas were analyzed by in

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Form 2.4.4 software (Akoya Biosciences, USA). The area was divided into two compartments: the epithelial compartment (the alveolar wall and septal or malignant cell nests) and alveolar air space or tumor stroma compartment (**Figure S5C**). The individual cells were recognized by DAPI nuclei staining (**Figure S5D**). Panel 1 contained the co-localization patterns as follows: PD-L1 expressing CK+ cells (CK+PD-L1+), total T lymphocytes (total CD3+), cytotoxic T lymphocytes (CD3+CD8+), antigen experienced T lymphocytes (CD3+PD-1), antigen experienced cytotoxic T lymphocytes (CD3+CD8+PD-1+), total macrophages (total CD68+), and PD-L1 expressing macrophages (CD68+ PD-L1+). Panel 2 contained the co-localization patterns as follows: total T lymphocytes (total CD3+), cytotoxic T lymphocytes (CD3+CD8+), activated cytotoxic T lymphocytes (CD3+CD8+GB+), memory T lymphocytes (CD3+CD45RO+), effector/memory T lymphocytes (CD3+CD8+CD45RO+), regulatory T lymphocytes (CD3+FOXP3+CD8-), and memory/regulatory T lymphocytes (CD3+CD45RO+FOXP3+) (**Figure S5**). The percentage of the total nucleated cell density from each individual cell phenotyping population was used for further analysis.

TCRβ amplification and sequencing

Immunosequencing of the CDR3 regions of human TCRβ chains was performed using the protocol of ImmunoSeq (Adaptive Biotechnologies, hsTCRβ Kit) and T cell clonality and diversity were calculated as described previously (Reuben et al., 2017; Reuben et al., 2020).

Briefly, T cell clonality is a metric of T cell proliferation and reactivity, and it is defined as 1-Pielou's evenness and is calculated on productive rearrangements by:

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

$$1+\frac{\sum_{i}^{N} p_i \log_2\left(p_i\right)}{\log_2\left(N\right)}$$

Where p_i is the proportional abundance of rearrangement i, and N is the total number of rearrangements. Clonality ranges from 0 to 1: values approaching 0 indicate a very even distribution frequency of different clones (polyclonal), whereas values approaching 1 indicate a distinct asymmetric distribution in which a few activated clones are present at high frequencies (monoclonal).

We also applied Inverse Simpson in order to observe T cell diversity, the sum over all observed rearrangements of the square fractional abundances of each rearrangements using productive templates.

$$\frac{1}{\sum_{i=1}^{S} p_i^2}$$

Where p_i is the proportional abundance of rearrangement i, and S is the total number of rearrangements. Inverse Simpson ranges from 1 to infinite, where a sample with little variation or abundance has a value approaching 1, and a maximally diverse and evenly distributed sample has a value approaching infinite.

Analysis of genomic and methylation profiling data

Genomic and methylation data were generated and processed in previous studies as described (Hu et al., 2020; Hu et al., 2019). Somatic mutations, allelic imbalance (AI), copy number variations (CNV), oncogene mutations, global methylation data were

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

directed from above mentioned studies. The following additional analyses were performed on the genomic data.

Detection of allele-specific HLA loss

Class I HLA alleles for each HLA gene was inferred by POLYSOLVER using a two-step Bayesian classification approach (Shukla et al., 2015). This approach takes into account the base qualities of aligned reads, observed insert sizes, as well as the ethnicity-dependent prior probabilities of each allele (Shukla et al., 2015). Tumor purity and ploidy were estimated using ASCAT (Van Loo et al., 2010). We then applied LOHHLA (Loss Of Heterozygosity in Human Leukocyte Antigen) algorithm (McGranahan et al., 2017) to detect allele-specific HLA loss in each sample. In brief, logR and BAF across each HLA gene loci was obtained by binning the coverage at mismatch positions between homologous HLA alleles, and HLA haplotype specific copy numbers were then calculated based on logR and BAF value from the corresponding bin adjusted by tumor purity and ploidy. The median value of binned allelic copy number was used to determined LOH, where a copy number of < 0.5 indicated allele loss and AI was determined if $p < 0.01$.

Statistical Analysis

Different statistical models were applied to assess the association among immune data, genomic data and methylation data. When assessing association between two variables, different tests were applied depending on the types of variables. For association between two continuous variables, spearman's rank correlation test was used. For association between one continuous variable and one categorical variable, Wilcoxon rank-sum test (categorical variable with two levels) and Kruskal-Wallis test (categorical variable with

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

more than two levels) were applied. The FDR method was used for multiple testing adjustment of p-values (Benjamini and Hochberg, 1995). All p-values are calculated with two-sided test, and $p<0.05$ was considered to be statistically significant.

DATA AVAILABILITY

The data from WES has been deposited at European Genome-phenome Archive (EGA), which is hosted by The European Bioinformatics Institute (EBI) and the Centre for Genomic Regulation (CRG) under the accession code: EGAS00001004960 [https://www.ebi.ac.uk/ega/datasets/EGAD00001004960]. All other data may be found within the main manuscript or supplementary Information or available from the authors upon request.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

SUPPLEMENTAL INFORMATION

SUPPLEMENTAL FIGURE LEGENDS

Figure S1. Example genes differentially expressed from preneoplasia to invasive lung adenocarcinoma. Gene expression of *CD47* (**A**), *CD276* (**B**), *CTLA4* (**C**), *ENTPD1* (**D**), *GZMB* (**E**), *PRF1* (**F**) was performed using the nCounter PanCancer Immune Profiling Panel (Nanostring). The difference in expression of each gene among different stages was evaluated using the Kruskal-Wallis H test. AAH: typical adenomatous hyperplasia. AIS: adenocarcinoma in situ. MIA: minimally invasive adenocarcinoma. ADC: invasive adenocarcinoma.

Figure S2. Expression of tertiary lymphoid structure (TLS) marker genes from preneoplasia to invasive lung adenocarcinoma. Gene expression of B cell marker *CD19* (**A**) and *MS4A1* (*CD20)* (**B**); follicle formation markers *CXCL13* (**C**) and *CXCR5* (**D**); T lymphocyte homing markers *CCL19* (**E**) and *CCR7* (**F**) was performed using the nCounter PanCancer Immune Profiling Panel (Nanostring). The difference in expression of each gene among different stages was evaluated using the Kruskal-Wallis H test. AAH: typical adenomatous hyperplasia. AIS: adenocarcinoma in situ. MIA: minimally invasive adenocarcinoma. ADC: invasive adenocarcinoma.

Figure S3. Expression of TLS marker genes was associated with B cells, CD4+ lymphocytes and CD8+ lymphocytes. Gene expression profiling was performed using the nCounter PanCancer Immune Profiling Panel (Nanostring). Infiltration of B cells, CD4+ T cells and CD8+ T cells was estimated using TIMER based on the gene expression using nCounter PanCancer Immune Profiling Panel. The association between

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

the expression of each TLS marker gene and B cells, CD4+ T cells and CD8+ T cells were evaluated using xxx test.

Figure S4. Progressively changes of immune cell in filtration normal lung, preneoplasia and invasive lung adenocarcinoma. Immune cell fractions including CD4+ T cells (**A**), CD8+ T cells (**B**), CD4 /CD8 ratio (**C**), B cells (**D**) were estimated using TIMER based on previously published RNA sequencing data from an independent cohort (GSE102511). Error bars indicate 95% confidence intervals and solid point represent mean value in each stage. The difference of cell fraction among stages was evaluated using the Kruskal-Wallis H test. NL: Normal lung tissue, AAH: Atypical adenomatous hyperplasia, ADC: Invasive adenocarcinoma.

Figure S5. Quantitative image analysis of multiplex immunofluorescence image data. (**A**) After immunofluorescent staining, high magnification fields of interest were selected in both malignant (premalignant) and non-malignant regions. The non-malignant fields at the farthest regions from malignant (premalignant) regions with morphologically normal histology were selected. (**B**) Representative images from Panel 1 and Panel 2. (**C**) The selected fields were divided into two areas: epithelium and alveolar space. (**D**) Individual cells were recognized by DAPI (nuclear staining).

Figure S6. Subtyping immune and epithelial cells by multiplex immunofluorescence. Cell subtypes were defined as PD-L1 expressing epithelial cells (AE1/AE3+PD-L1+), T lymphocytes (CD3+), antigen-experienced T cells (CD3+PD-1+), cytotoxic T lymphocytes (CTL, CD3+CD8+), antigen-experienced CTL (CD3+CD8+PD-1+), macrophages (CD68+) and PD-L1 expressing macrophages (CD68+PD-L1+) in panel 1 (left); and activated CTL (CD3+CD8+granzyme B+), memory T cell

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(CD3+CD45RO+), memory CTL (CD3+CD8+CD45RO+), regulatory T cell (CD3+CD8-FoxP3+), memory/regulatory T cell (CD3+CD45RO+FoxP3+) in panel 2 (right).

Figure S7. The correlation between immune cell subtypes measured by multiplex immunofluorescence and T cell subtypes inferred from immune gene expression. Fractions of immune cells including CD8+ T cells (**A**), CD4+ T cells (**B**) and CD4/CD8 ratio (**C**) estimated using TIMER based on the gene expression from nCounter PanCancer Immune Profiling Panel (y-axis) were correlated to activated CTL (CD3+CD8+granzyme B+), regulatory T cell (CD3+CD8-FoxP3+) and Treg/activated CTL ratio respectively (x-axis). All fraction and ratio were log2 transformed for visualization. The correction coefficient (rho) was assessed by Spearman's rank correlation test.

Figure S8. Various subtypes immune cells and their interaction with epithelial cells in normal lung tissues and invasive lung adenocarcinoma. The correlation between immune cell subtypes and CK+ epithelial cells measured by mIF in normal lung (**A**) and invasive lung adenocarcinoma (**B**). Significant correlation was marked with * ($p<0.05$). Red circles indicate positive correlations and blue circles indicate negative correlations. The size of circles is proportional to the spearman's correlation co-efficient between each pair of cells. ADC: Invasive adenocarcinoma.

Figure S9. Correlation between T cell diversity and regulatory T cells. T cell diversity was positively correlated with infiltration of CD4+ T cells derived from immune gene expression using TIMER (**A**), regulatory T cells from multiplex immunofluorescence (mIF) (**B**). The correction coefficient (rho) was assessed by Spearman's rank correlation test.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Figure S10. Progressive decrease of top T cell clone frequencies from preneoplasia to invasive lung adenocarcinoma. Frequencies of (**A**) top 10, (**B**) tope 100, (**C**) top 200 and (**D**) top 500 T cell clones in the Normal (yellow), AAH (blue), AIS (red), MIA (green) and ADC (purple). The difference between different stages were assessed by the Kruskal-Wallis H test.

Figure S11. Correlation between T cell clonality and different T cell subtypes. T cell clonality was positively correlated with infiltration of CD8+ T cells inferred from gene expression by TIMER (**A**), activated cytotoxic T cells (**B**) by mIF, expression of GZBM (**C**) and Th1 cytokines (INFG, IL12A and IL12B) (**D**) but negative correlated with infiltration of CD4+ T cells derived from gene expression by TIMER (**E**) and regulatory T cells by mIF (**F**). The correction coefficient (rho) was assessed by Spearman's rank correlation test.

Figure S12. The impact of oncogene mutations on T cell features. Comparison of infiltration of CD8+ T cells inferred from gene expression by TIMER (**A**), cytotoxic T cells (CTL) (**B**) by mIF, CD4/CD8 ratio inferred from gene expression by TIMER (**C**), Effector memory cytotoxic T cells by mIF (**D**), CTL/Treg ratio (**E**) by mIF, T cell clonality by TCR sequencing (**F**) in lesions with *EGFR* mutation (pink), *KRAS* mutation (orange) and wildtype for both *KRAS* and *EGFR* (green). The difference was assessed by Wilcox-rank sum test.

Figure S13. The impact of oncogene mutations on T cell features by different histologic stages. Comparison of infiltration of CD8+ T cells inferred from gene expression by TIMER (**A**), cytotoxic T cells (CTL) **B**) by mIF, CD4/CD8 ratio inferred from gene expression by TIMER (**C**), Effector memory cytotoxic T cells by mIF (**D**), CTL/Treg ratio (**E**) by mIF, T cell clonality by TCR sequencing (**F**) in lesions with *EGFR* mutation

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(pink), *KRAS* mutation (orange) and wildtype for both *KRAS* and *EGFR* (green) in AAH, AIS, MIA and ADC. The difference was assessed by Wilcox-rank sum test. AAH: Atypical adenomatous hyperplasia, AIS: Adenocarcinoma in situ, MIA: Minimally invasive adenocarcinoma, ADC: Invasive adenocarcinoma.

Figure S14. Loss of heterozygosity of HLA (HLA LOH) in different histologic stages from preneoplasia to invasive lung adenocarcinoma and its potential impact on immune contexture. (**A**) The proportion of AAH, AIS, MIA and ADC lesions had evidence of HLA-LOH. Chi-squared test were used to assess the difference among different histologic stages. The difference of infiltration of CD4+ T cells (**B**), CD8+ T cells (**C**) and CD4/CD8 ratio (**D**) inferred from gene expression by TIMER between lesions with (purple) and without (green) HLA-LOH.

Figure S15. The potential impact of chromosomal copy number changes on immune infiltration. The correlation between allelic imbalance (AI) burden (number of AI events) and infiltration of CD4+ T cells (**A**), CD8+ T cells (**B**), CD4/CD8 ratio (**C**) inferred from gene expression by TIMER. The correlation between copy number variation (CNV) burden (normalized as the number of genes with CNV) and infiltration of CD4+ T cells (**D**), CD8+ T cells (**E**), CD4/CD8 ratio (**F**) inferred from gene expression by TIMER. The correction coefficient (rho) was assessed by Spearman's rank correlation test.

Figure S16. The relationship between HLA loss and chromosomal copy number variations. (**A**) Comparison of allelic imbalance (AI) burden (number of AI events) in lesions with (purple) and without (green) HLA-LOH. (**B**) Comparison of copy number variation (CNV) burden (normalized as the percent of genes with CNV) in lesions with

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

(purple) and without (green) HLA-LOH. Willcoxon rank-sum test was used to assess the differences.

Figure S17. Promoter methylation and mutation burden from preneoplasia to invasive lung adenocarcinoma. The number of all nonsynonymous mutations in each histologic stage is shown as purple boxes and the number of nonsynonymous mutations from genes without promoter methylation (<30% CpG sites methylated) is shown as blue boxes.

Figure S18. The potential impact of global methylation and immune infiltration. The correlation between global methylation levels (using LINE-1 as a surrogate marker) with CD4+ T cells (**A**), CD4/CD8 ratio (**B**) inferred from immune gene expression by TIMER as well as Tregs (**C**) and Treg/CD8 ratio (**D**) measured by mIF. The correction coefficient (rho) was assessed by Spearman's rank correlation test.

SUPPLEMENTAL TABLES

Table S1. Clinical characteristics and availability of different immune and molecular data

Table S2. Statistics of Differentially Expressed Genes

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

REFERENCES

Aoyagi, Y., Yokose, T., Minami, Y., Ochiai, A., Iijima, T., Morishita, Y., Oda, T., Fukao, K., and Noguchi, M. (2001). Accumulation of losses of heterozygosity and multistep carcinogenesis in pulmonary adenocarcinoma. Cancer research *61*, 7950-7954.

Bakhoum, S. F., and Cantley, L. C. (2018). The Multifaceted Role of Chromosomal Instability in Cancer and Its Microenvironment. Cell *174*, 1347-1360.

Barnett, R. (2017). Lung cancer. The Lancet *390*, 928.

Bastid, J., Cottalorda-Regairaz, A., Alberici, G., Bonnefoy, N., Eliaou, J., and Bensussan, A. (2013). ENTPD1/CD39 is a promising therapeutic target in oncology. Oncogene *32*, 1743-1751.

Bellomi, M., Veronesi, G., Trifirò, G., Brambilla, S., Bonello, L., Preda, L., Casiraghi, M., Borri, A., Paganelli, G., and Spaggiari, L. (2010). Computed tomography-guided preoperative radiotracer localization of nonpalpable lung nodules. The Annals of thoracic surgery *90*, 1759-1764.

Benjamini, Y., and Hochberg, Y. (1995). Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. J R Stat Soc B *57*, 289-300.

Borghaei, H., Paz-Ares, L., Horn, L., Spigel, D. R., Steins, M., Ready, N. E., Chow, L. Q., Vokes, E. E., Felip, E., Holgado, E.*, et al.* (2015). Nivolumab versus Docetaxel in Advanced Nonsquamous Non-Small-Cell Lung Cancer. N Engl J Med *373*, 1627-1639.

Cabrita, R., Lauss, M., Sanna, A., Donia, M., Larsen, M. S., Mitra, S., Johansson, I., Phung, B., Harbst, K., and Vallon-Christersson, J. (2020). Tertiary lymphoid structures improve immunotherapy and survival in melanoma. Nature *577*, 561-565.

Cesano, A. (2015). nCounter(®) PanCancer Immune Profiling Panel (NanoString Technologies, Inc., Seattle, WA). Journal for immunotherapy of cancer *3*, 42.

Chen, P. L., Roh, W., Reuben, A., Cooper, Z. A., Spencer, C. N., Prieto, P. A., Miller, J. P., Bassett, R. L., Gopalakrishnan, V., Wani, K.*, et al.* (2016). Analysis of Immune Signatures in Longitudinal Tumor Samples Yields Insight into Biomarkers of Response and Mechanisms of Resistance to Immune Checkpoint Blockade. Cancer Discov *6*, 827-837.

Chen, R. Z., Pettersson, U., Beard, C., Jackson-Grusby, L., and Jaenisch, R. (1998). DNA hypomethylation leads to elevated mutation rates. Nature *395*, 89-93.

Chiosea, S., Jelezcova, E., Chandran, U., Luo, J., Mantha, G., Sobol, R. W., and Dacic, S. (2007). Overexpression of Dicer in precursor lesions of lung adenocarcinoma. Cancer research *67*, 2345-2350.

Davoli, T., Uno, H., Wooten, E. C., and Elledge, S. J. (2017). Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science *355*.

de Koning, H. J., van der Aalst, C. M., de Jong, P. A., Scholten, E. T., Nackaerts, K., Heuvelmans, M. A., Lammers, J.-W. J., Weenink, C., Yousaf-Khan, U., and Horeweg, N. (2020). Reduced lung-cancer mortality with volume CT screening in a randomized trial. New England Journal of Medicine *382*, 503-513.

Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002). Cancer immunoediting: from immunosurveillance to tumor escape. Nature immunology *3*, 991-998.

Dunn, G. P., Old, L. J., and Schreiber, R. D. (2004a). The immunobiology of cancer immunosurveillance and immunoediting. Immunity *21*, 137-148.

Dunn, G. P., Old, L. J., and Schreiber, R. D. (2004b). The three Es of cancer immunoediting. Annu Rev Immunol *22*, 329-360.

Eden, A., Gaudet, F., Waghmare, A., and Jaenisch, R. (2003). Chromosomal instability and tumors promoted by DNA hypomethylation. Science *300*, 455.

Francisco-Cruz, A., Parra, E. R., Tetzlaff, M. T., and Wistuba, II (2020). Multiplex Immunofluorescence Assays. Methods in molecular biology (Clifton, NJ) *2055*, 467-495.

Fridman, W. H., Zitvogel, L., Sautès–Fridman, C., and Kroemer, G. (2017). The immune contexture in cancer prognosis and treatment. Nature reviews Clinical oncology *14*, 717.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Hu, X., Estecio, M. R., Chen, R., Reuben, A., Wang, L., Junya, F., Ying, L., Junya, F., Chow, C.-W., McGranahan, N.*, et al.* (2020). Evolution of the DNA methylome from precancer to invasive lung adenocarcinoma. medRxiv, 2020.2007.2011.20142745.

Hu, X., Fujimoto, J., Ying, L., Fukuoka, J., Ashizawa, K., Sun, W., Reuben, A., Chow, C.-W., McGranahan, N., and Chen, R. (2019). Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. Nature communications *10*, 2978.

Izumchenko, E., Chang, X., Brait, M., Fertig, E., Kagohara, L. T., Bedi, A., Marchionni, L., Agrawal, N., Ravi, R., Jones, S.*, et al.* (2015). Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nature communications *6*, 8258.

Jeschke, J., Bizet, M., Desmedt, C., Calonne, E., Dedeurwaerder, S., Garaud, S., Koch, A., Larsimont, D., Salgado, R., and Van den Eynden, G. (2017). DNA methylation–based immune response signature improves patient diagnosis in multiple cancers. The Journal of clinical investigation *127*, 3090-3102.

Jung, H., Kim, H. S., Kim, J. Y., Sun, J. M., Ahn, J. S., Ahn, M. J., Park, K., Esteller, M., Lee, S. H., and Choi, J. K. (2019). DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat Commun *10*, 4278.

Kankava, K., Kvaratskhelia, E., Burkadze, G., Kokhreidze, I., Gogokhia, N., and Abzianidze, E. (2018). Line-1 Methylation in Blood and Tissues of Patients with Breast Cancer. Georgian Med News, 107-112.

Kaplinsky, J., and Arnaout, R. (2016). Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. Nature communications *7*, 11881-11881.

Kitamura, H., Kameda, Y., Ito, T., and Hayashi, H. (1999). Atypical adenomatous hyperplasia of the lung: implications for the pathogenesis of peripheral lung adenocarcinoma. American journal of clinical pathology *111*, 610-622.

Kramer, A., Green, J., Pollard, J., Jr., and Tugendreich, S. (2014). Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics (Oxford, England) *30*, 523-530.

Kulkarni, M. M. (2011). Digital multiplexed gene expression analysis using the NanoString nCounter system. Current protocols in molecular biology *Chapter 25*, Unit25B.10.

Lavin, Y., Kobayashi, S., Leader, A., Amir, E.-a. D., Elefant, N., Bigenwald, C., Remark, R., Sweeney, R., Becker, C. D., Levine, J. H.*, et al.* Innate Immune Landscape in Early Lung Adenocarcinoma by Paired Single-Cell Analyses. Cell *169*, 750-765.e717.

Li, T., Fu, J., Zeng, Z., Cohen, D., Li, J., Chen, Q., Li, B., and Liu, X. S. (2020). TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic acids research.

Liu, M., Zhou, J., Chen, Z., and Cheng, A. S. (2017). Understanding the epigenetic regulation of tumours and their microenvironments: opportunities and problems for epigenetic therapy. J Pathol *241*, 10-24.

Maeshima, A. M., Tochigi, N., Yoshida, A., Asamura, H., Tsuta, K., and Tsuda, H. (2010). Clinicopathologic analysis of multiple (five or more) atypical adenomatous hyperplasias (AAHs) of the lung: evidence for the AAH-adenocarcinoma sequence. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer *5*, 466-471.

Mansfield, A. S., Ren, H., Sutor, S., Sarangi, V., Nair, A., Davila, J., Elsbernd, L. R., Udell, J. B., Dronca, R. S., Park, S.*, et al.* (2018). Contraction of T cell richness in lung cancer brain metastases. Scientific reports *8*, 2171.

Mascaux, C., Angelova, M., Vasaturo, A., Beane, J., Hijazi, K., Anthoine, G., Buttard, B., Rothe, F., Willard-Gallo, K., Haller, A.*, et al.* (2019). Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature *571*, 570-575.

McGranahan, N., Furness, A. J., Rosenthal, R., Ramskov, S., Lyngaa, R., Saini, S. K., Jamal-Hanjani, M., Wilson, G. A., Birkbak, N. J., Hiley, C. T.*, et al.* (2016). Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science *351*, 1463-1469.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

McGranahan, N., Rosenthal, R., Hiley, C. T., Rowan, A. J., Watkins, T. B. K., Wilson, G. A., Birkbak, N. J., Veeriah, S., Van Loo, P., Herrero, J.*, et al.* (2017). Allele-Specific HLA Loss and Immune Escape in Lung Cancer Evolution. Cell *171*, 1259-1271.e1211.

McWilliams, A., Tammemagi, M. C., Mayo, J. R., Roberts, H., Liu, G., Soghrati, K., Yasufuku, K., Martel, S., Laberge, F., and Gingras, M. (2013). Probability of cancer in pulmonary nodules detected on first screening CT. New England Journal of Medicine *369*, 910-919.

Min, J. H., Lee, H. Y., Lee, K. S., Han, J., Park, K., Ahn, M. J., and Lee, S. J. (2010). Stepwise evolution from a focal pure pulmonary ground-glass opacity nodule into an invasive lung adenocarcinoma: an observation for more than 10 years. Lung cancer (Amsterdam, Netherlands) *69*, 123-126.

Mittal, D., Gubin, M. M., Schreiber, R. D., and Smyth, M. J. (2014). New insights into cancer immunoediting and its three component phases—elimination, equilibrium and escape. Current opinion in immunology *27*, 16-25.

Neal, J. T., Li, X., Zhu, J., Giangarra, V., Grzeskowiak, C. L., Ju, J., Liu, I. H., Chiou, S.-H., Salahudeen, A. A., and Smith, A. R. (2018). Organoid modeling of the tumor immune microenvironment. Cell *175*, 1972-1988. e1916.

Noguchi, M. (2010). Stepwise progression of pulmonary adenocarcinoma—clinical and molecular implications. Cancer and Metastasis Reviews *29*, 15-21.

O'Donnell, J. S., Teng, M. W., and Smyth, M. J. (2019). Cancer immunoediting and resistance to T cell-based immunotherapy. Nature reviews Clinical oncology *16*, 151-167.

Ohka, F., Natsume, A., Motomura, K., Kishida, Y., Kondo, Y., Abe, T., Nakasu, Y., Namba, H., Wakai, K., Fukui, T.*, et al.* (2011). The global DNA methylation surrogate LINE-1 methylation is correlated with MGMT promoter methylation and is a better prognostic factor for glioma. PLoS One *6*, e23332.

Pan, L., Lu, M. P., Wang, J. H., Xu, M., and Yang, S. R. (2020). Immunological pathogenesis and treatment of systemic lupus erythematosus. World J Pediatr *16*, 19-30.

Pantelidou, C., Sonzogni, O., De Oliveria Taveira, M., Mehta, A. K., Kothari, A., Wang, D., Visal, T., Li, M. K., Pinto, J., Castrillon, J. A.*, et al.* (2019). PARP Inhibitor Efficacy Depends on CD8(+) T-cell Recruitment via Intratumoral STING Pathway Activation in BRCA-Deficient Models of Triple-Negative Breast Cancer. Cancer Discov *9*, 722-737.

Pardoll, D. M. (2012). The blockade of immune checkpoints in cancer immunotherapy. Nature Reviews Cancer *12*, 252-264.

Parra, E. R., Villalobos, P., Behrens, C., Jiang, M., Pataer, A., Swisher, S. G., William, W. N., Jr., Zhang, J., Lee, J., Cascone, T.*, et al.* (2018). Effect of neoadjuvant chemotherapy on the immune microenvironment in non-small cell lung carcinomas as determined by multiplex immunofluorescence and image analysis approaches. J Immunother Cancer *6*, 48.

Paulson, K., Voillet, V., McAfee, M., Hunter, D., Wagener, F., Perdicchio, M., Valente, W., Koelle, S., Church, C., and Vandeven, N. (2018). Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. Nature communications *9*, 1-10.

Picarda, E., Ohaegbulam, K. C., and Zang, X. (2016). Molecular pathways: targeting B7-H3 (CD276) for human cancer immunotherapy. Clinical Cancer Research *22*, 3425-3431.

Porta-Pardo, E., and Godzik, A. (2016). Mutation Drivers of Immunological Responses to Cancer. Cancer Immunol Res *4*, 789-798.

Pounds, S., and Morris, S. W. (2003). Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. Bioinformatics *19*, 1236-1242.

Prakash, M. D., Munoz, M. A., Jain, R., Tong, P. L., Koskinen, A., Regner, M., Kleifeld, O., Ho, B., Olson, M., and Turner, S. J. (2014). Granzyme B promotes cytotoxic lymphocyte transmigration via basement membrane remodeling. Immunity *41*, 960-972.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Reuben, A., Gittelman, R., Gao, J., Zhang, J., Yusko, E. C., Wu, C. J., Emerson, R., Zhang, J., Tipton, C., Li, J., *et al.* (2017). TCR Repertoire Intratumor Heterogeneity in Localized Lung Adenocarcinomas: An Association with Predicted Neoantigen Heterogeneity and Postsurgical Recurrence. Cancer Discov *7*, 1088-1097.

Reuben, A., Zhang, J., Chiou, S. H., Gittelman, R. M., Li, J., Lee, W. C., Fujimoto, J., Behrens, C., Liu, X., Wang, F., *et al.* (2020). Comprehensive T cell repertoire characterization of non-small cell lung cancer. Nat Commun *11*, 603.

Rosenthal, R., Cadieux, E. L., Salgado, R., Al Bakir, M., Moore, D. A., Hiley, C. T., Lund, T., Tanić, M., Reading, J. L., and Joshi, K. (2019). Neoantigen-directed immune escape in lung cancer evolution. Nature *567*, 479-485.

Saito, K., Kawakami, K., Matsumoto, I., Oda, M., Watanabe, G., and Minamoto, T. (2010). Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. Clin Cancer Res *16*, 2418-2426.

Schreiber, R. D., Old, L. J., and Smyth, M. J. (2011). Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science *331*, 1565-1570.

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., and Ragg, T. (2006). The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC molecular biology *7*, 3.

Seki, M., and Akasaka, Y. (2007). Multiple lung adenocarcinomas and AAH treated by surgical resection. Lung cancer (Amsterdam, Netherlands) *55*, 237-240.

Serrano, A., Castro-Vega, I., and Redondo, M. (2011). Role of gene methylation in antitumor immune response: implication for tumor progression. Cancers (Basel) *3*, 1672-1690.

Shah, W., Yan, X., Jing, L., Zhou, Y., Chen, H., and Wang, Y. (2011). A reversed CD4/CD8 ratio of tumor-infiltrating lymphocytes and a high percentage of CD4+ FOXP3+ regulatory T cells are significantly associated with clinical outcome in squamous cell carcinoma of the cervix. Cellular & molecular immunology *8*, 59-66.

Shukla, S. A., Rooney, M. S., Rajasagi, M., Tiao, G., Dixon, P. M., Lawrence, M. S., Stevens, J., Lane, W. J., Dellagatta, J. L., Steelman, S., *et al.* (2015). Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nature biotechnology *33*, 1152-1158.

Siegel, R. L., Miller, K. D., and Jemal, A. (2020). Cancer statistics, 2020. CA Cancer J Clin *70*, 7-30.

Sivakumar, S., Lucas, F. A. S., McDowell, T. L., Lang, W., Xu, L., Fujimoto, J., Zhang, J., Futreal, P. A., Fukuoka, J., Yatabe, Y., *et al.* (2017). Genomic Landscape of Atypical Adenomatous Hyperplasia Reveals Divergent Modes to Lung Adenocarcinoma. Cancer research *77*, 6119-6130.

Soto-Pantoja, D. R., Terabe, M., Ghosh, A., Ridnour, L. A., DeGraff, W. G., Wink, D. A., Berzofsky, J. A., and Roberts, D. D. (2014). CD47 in the tumor microenvironment limits cooperation between antitumor T-cell immunity and radiotherapy. Cancer research *74*, 6771-6783.

Tan, W. L., Jain, A., Takano, A., Newell, E. W., Iyer, N. G., Lim, W. T., Tan, E. H., Zhai, W., Hillmer, A. M., Tam, W. L., and Tan, D. S. W. (2016). Novel therapeutic targets on the horizon for lung cancer. The Lancet Oncology *17*, e347-e362.

Thorsson, V., Gibbs, D. L., Brown, S. D., Wolf, D., Bortone, D. S., Yang, T.-H. O., Porta-Pardo, E., Gao, G. F., Plaisier, C. L., and Eddy, J. A. (2018). The immune landscape of cancer. Immunity *48*, 812-830. e814.

Travis, W. D., Brambilla, E., Noguchi, M., Nicholson, A. G., Geisinger, K. R., Yatabe, Y., Beer, D. G., Powell, C. A., Riely, G. J., Van Schil, P. E., *et al.* (2011). International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. Journal of thoracic oncology : official publication of the International Association for the Study of Lung Cancer *6*, 244-285.

Tschopp, J., Masson, D., and Stanley, K. K. (1986). Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature *322*, 831-834.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

Van Loo, P., Nordgard, S. H., Lingjaerde, O. C., Russnes, H. G., Rye, I. H., Sun, W., Weigman, V. J., Marynen, P., Zetterberg, A., Naume, B.*, et al.* (2010). Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A *107*, 16910-16915.

Vogelstein, B., Papadopoulos, N., Velculescu, V. E., Zhou, S., Diaz, L. A., Jr., and Kinzler, K. W. (2013). Cancer genome landscapes. Science *339*, 1546-1558.

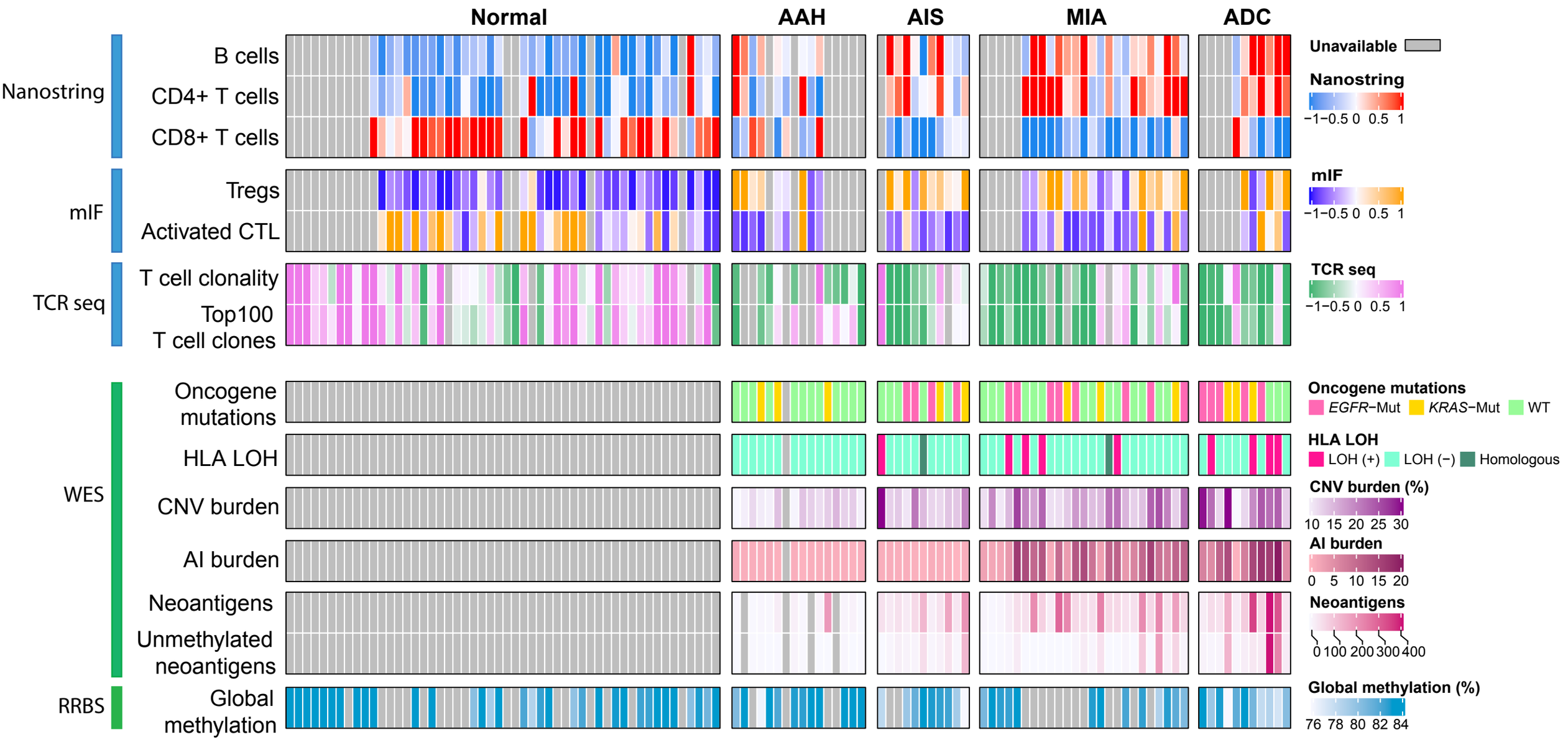
Wang, J., Jiang, D., Zheng, X., Li, W., Zhao, T., Wang, D., Yu, H., Sun, D., Li, Z., Zhang, J.*, et al.* (2020). Tertiary Lymphoid Structure and CD8 T Cell Exclusion in Minimally Invasive Adenocarcinoma. medRxiv, 2020.2008.2003.20166991.

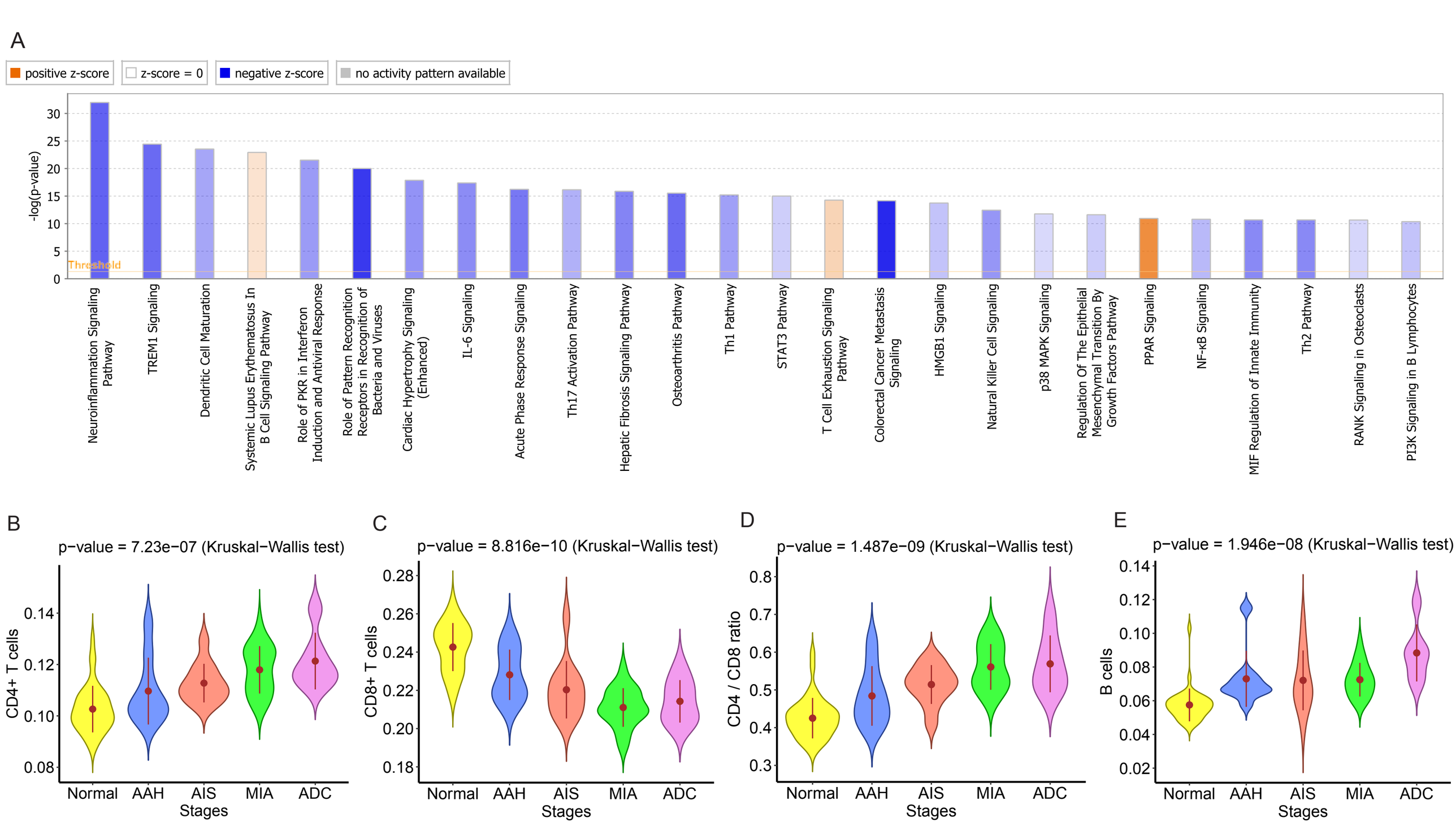
Wang, X., Zhang, B., Yang, Y., Zhu, J., Cheng, S., Mao, Y., Feng, L., and Xiao, T. (2019). Characterization of Distinct T Cell Receptor Repertoires in Tumor and Distant Non-tumor Tissues from Lung Cancer Patients. Genomics Proteomics Bioinformatics *17*, 287-296.

Weichert, W., and Warth, A. (2014). Early lung cancer with lepidic pattern: adenocarcinoma in situ, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma. Current opinion in pulmonary medicine *20*, 309-316.

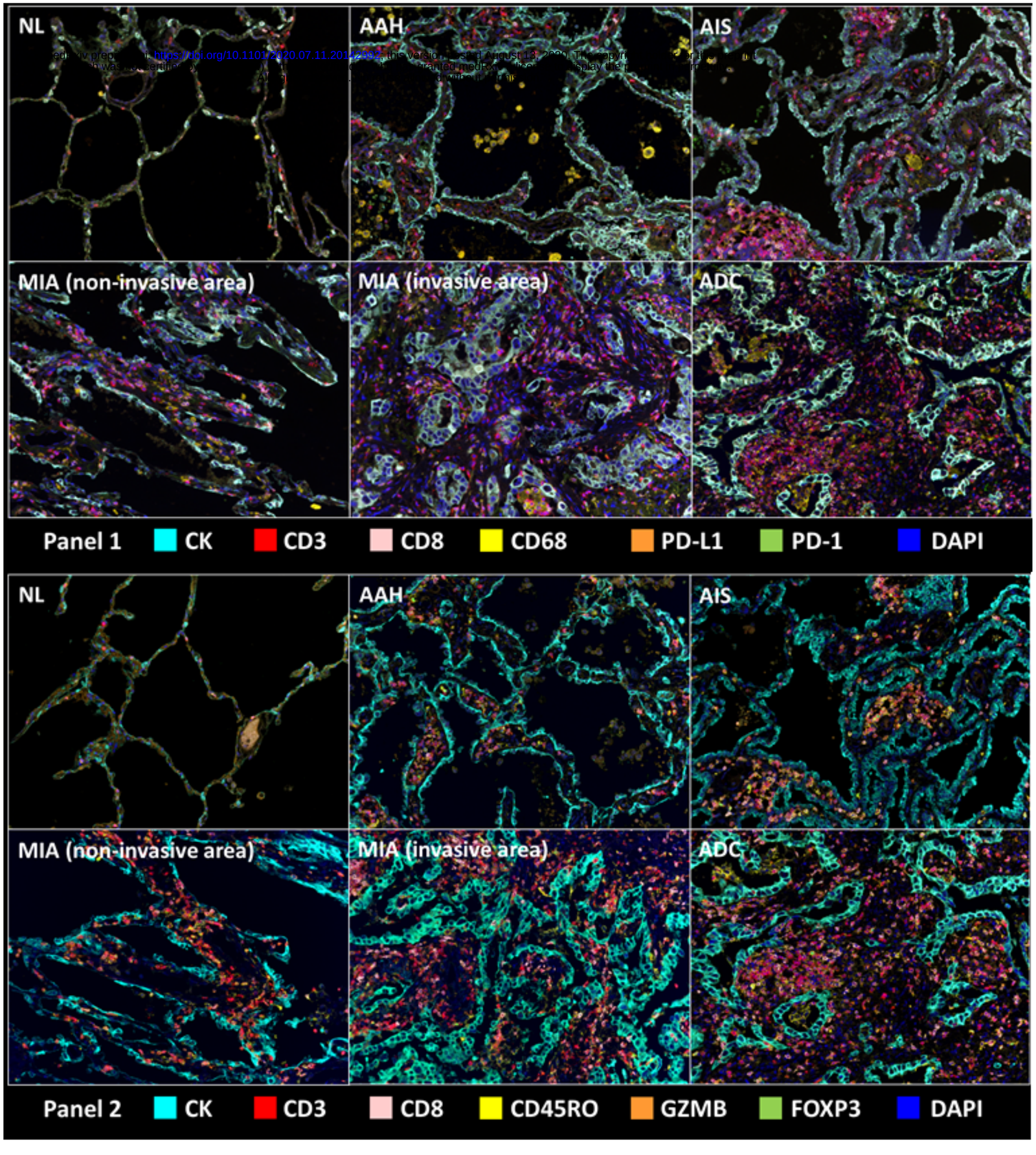
Wherry, E. J., and Kurachi, M. (2015). Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol *15*, 486-499.

Yarchoan, M., Johnson III, B. A., Lutz, E. R., Laheru, D. A., and Jaffee, E. M. (2017). Targeting neoantigens to augment antitumour immunity. Nature Reviews Cancer *17*, 209.

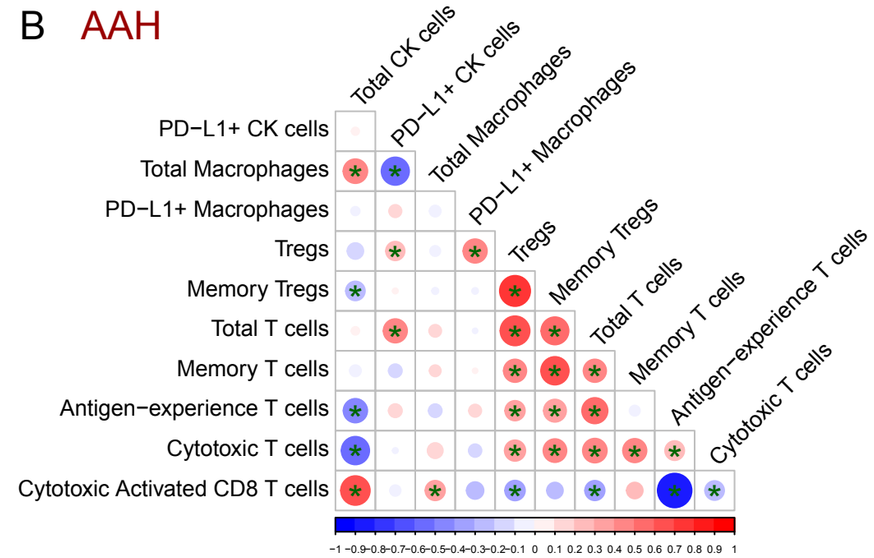

Normal
AAH
AIS
MIA
ADC
Nanostring
B cells
CD4+ T cells
CD8+ T cells
mIF
Tregs
Activated CTL
TCR seq
T cell clonality
Top100 T cell clones
WES
Oncogene mutations
HLA LOH
CNV burden
AI burden
Neoantigens
Unmethylated neoantigens
RRBS
Global methylation
Unavailable
Nanostring
−1 −0.5 0 0.5 1
mIF
−1 −0.5 0 0.5 1
TCR seq
−1 −0.5 0 0.5 1
Oncogene mutations
EGFR−Mut
KRAS−Mut
WT
HLA LOH
LOH (+)
LOH (−)
Homologous
CNV burden (%)
10 15 20 25 30
AI burden
0 5 10 15 20
Neoantigens
0 100 200 300 400
Global methylation (%)
76 78 80 82 84


A
positive z-score
z-score = 0
negative z-score
no activity pattern available
-log(p-value)
Threshold
Neuroinflammation Signaling Pathway
TREM1 Signaling
Dendritic Cell Maturation
Systemic Lupus Erythematosus In B Cell Signaling Pathway
Role of PKR in Interferon Induction and Antiviral Response
Role of Pattern Recognition Receptors in Recognition of Bacteria and Viruses
Cardiac Hypertrophy Signaling (Enhanced)
IL-6 Signaling
Acute Phase Response Signaling
Th17 Activation Pathway
Hepatic Fibrosis Signaling Pathway
Osteoarthritis Pathway
Th1 Pathway
STAT3 Pathway
T Cell Exhaustion Signaling Pathway
Colorectal Cancer Metastasis Signaling
HMGB1 Signaling
Natural Killer Cell Signaling
p38 MAPK Signaling
Regulation Of The Epithelial Mesenchymal Transition By Growth Factors Pathway
PPAR Signaling
NF-κB Signaling
MIF Regulation of Innate Immunity
Th2 Pathway
RANK Signaling in Osteoclasts
PI3K Signaling in B Lymphocytes
B
p−value = 7.23e−07 (Kruskal−Wallis test)
CD4+ T cells
Normal
AAH
AIS
MIA
ADC
Stages
C
p−value = 8.816e−10 (Kruskal−Wallis test)
CD8+ T cells
D
p−value = 1.487e−09 (Kruskal−Wallis test)
CD4 / CD8 ratio
E
p−value = 1.946e−08 (Kruskal−Wallis test)
B cells

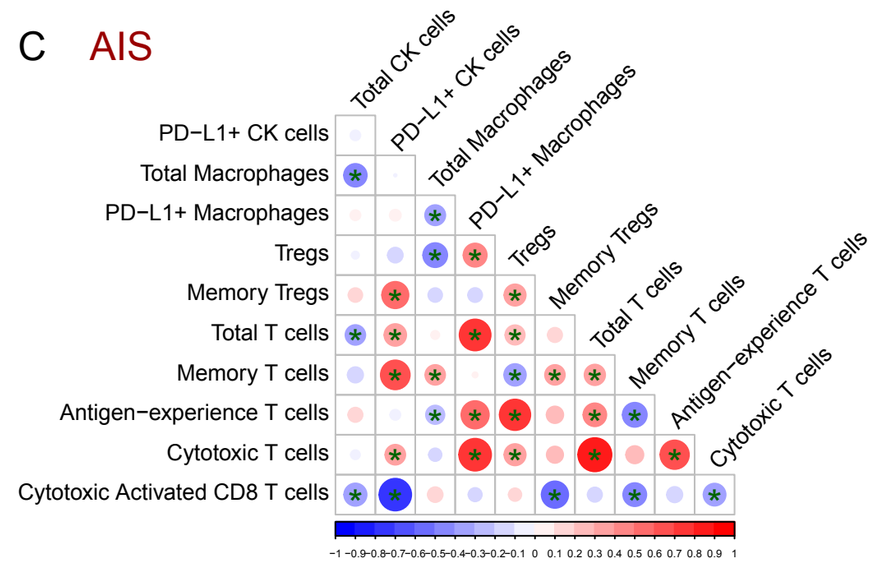
A



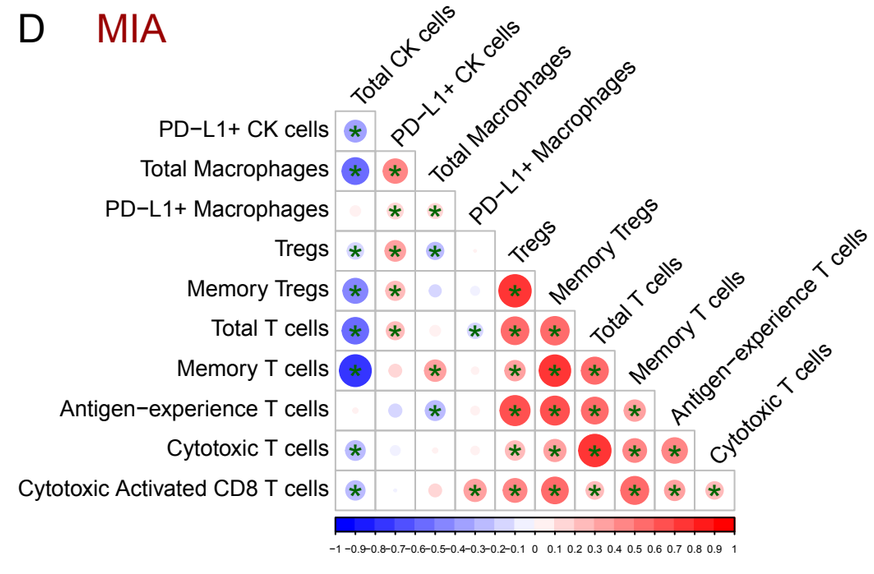
B AAH



C AIS

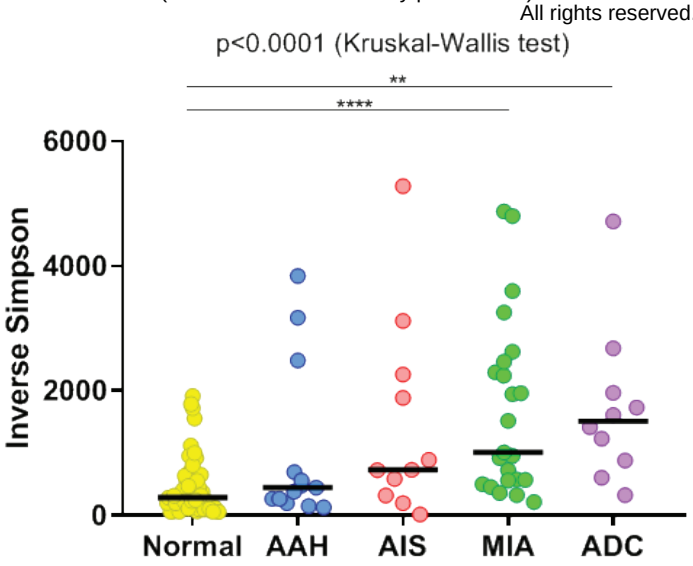


D MIA

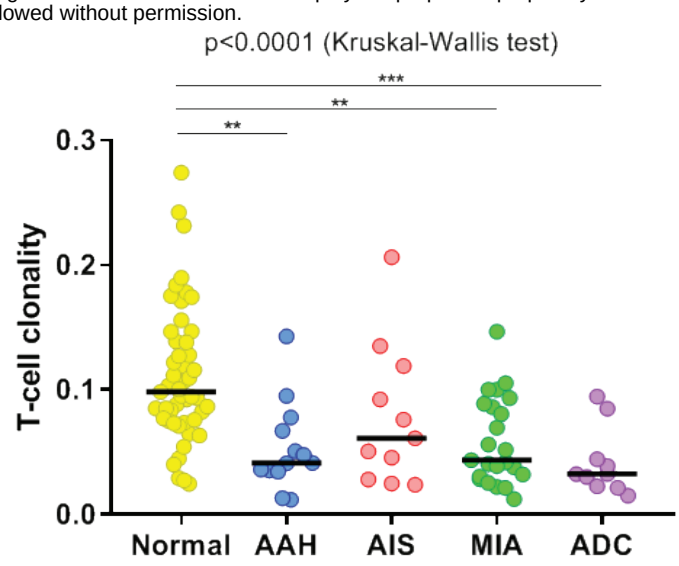


medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted August 13, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. All rights reserved. No reuse allowed without permission.

A



C



B

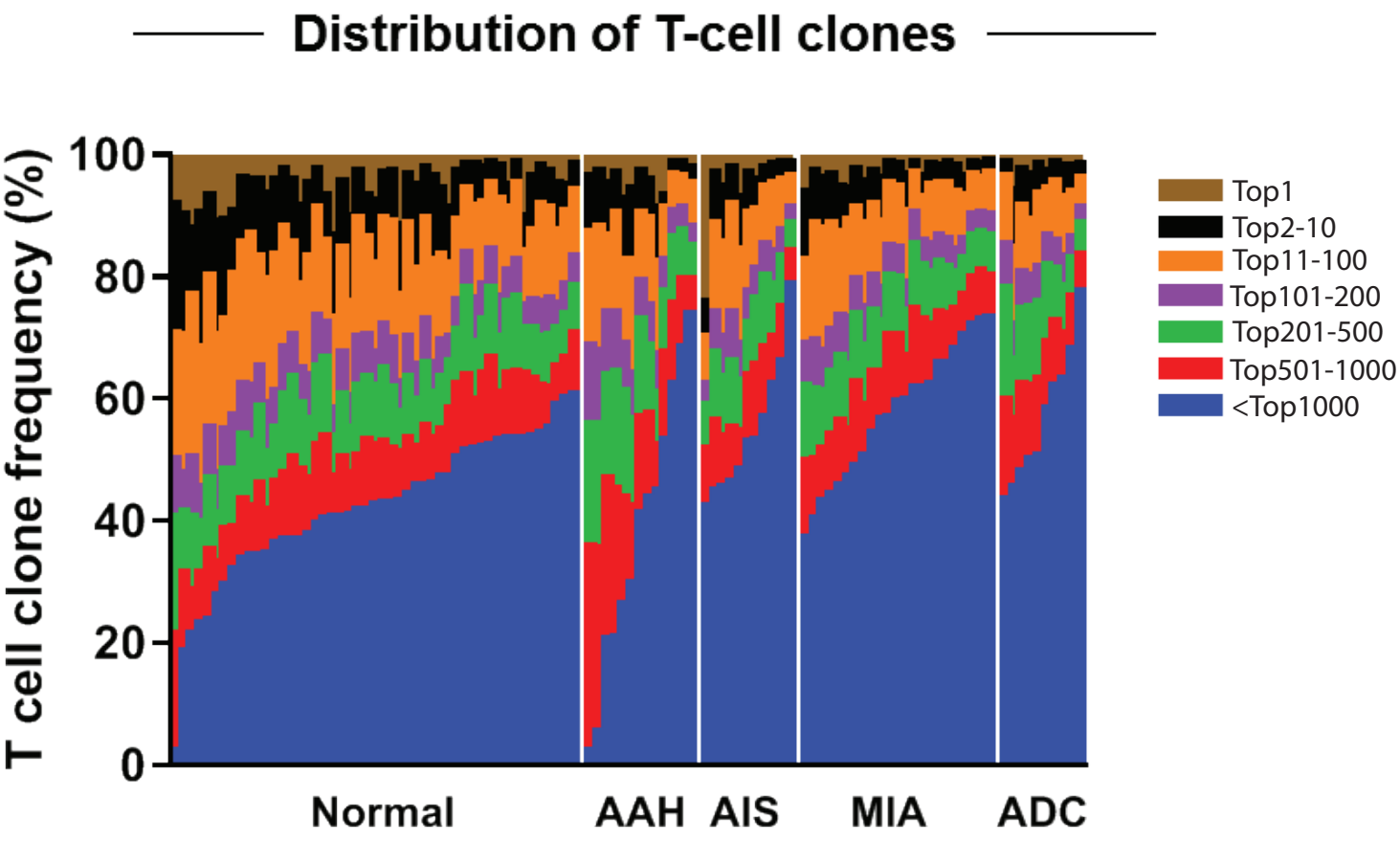


Figure S1

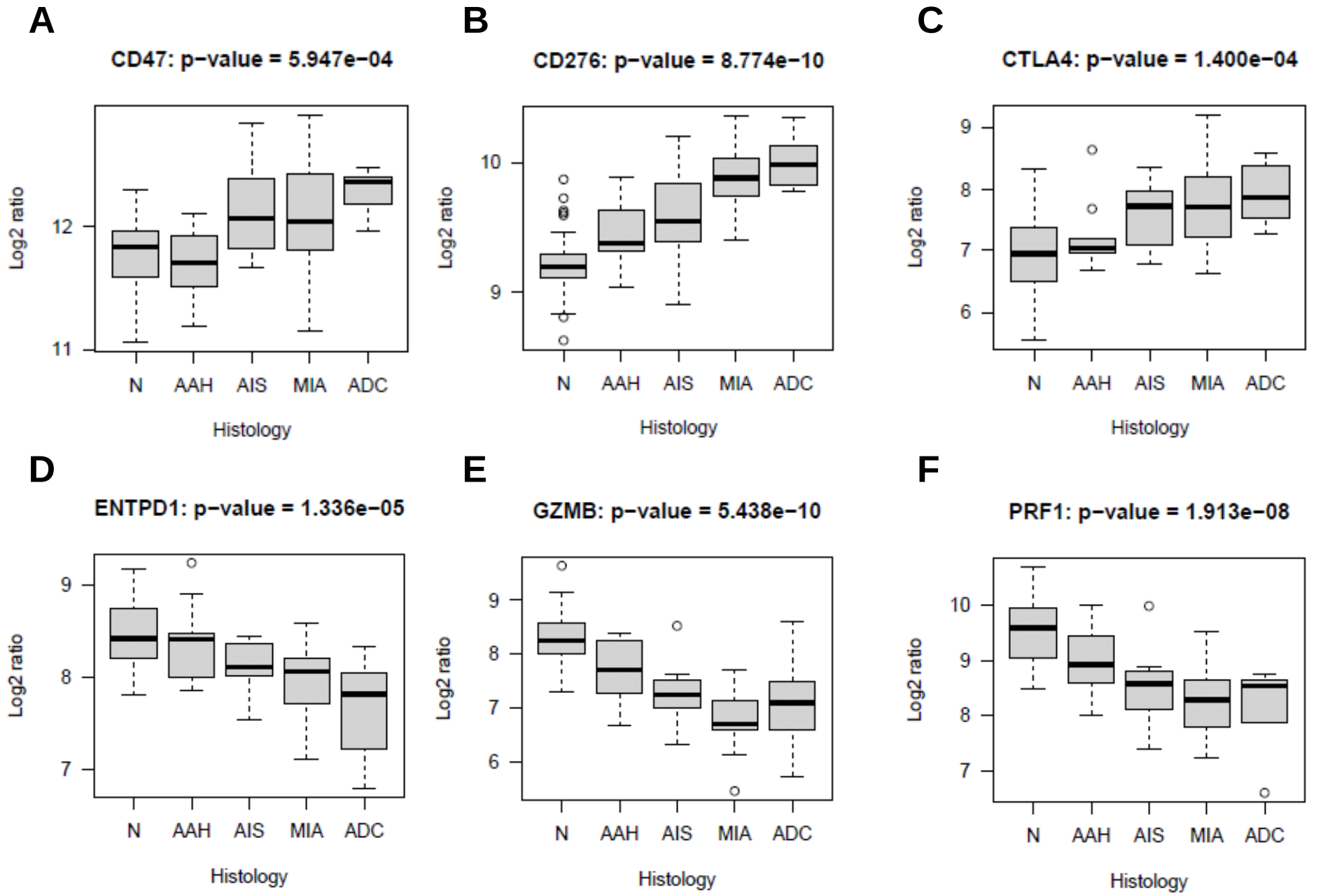

A
CD47: p−value = 5.947e−04
Log2 ratio
11
12
N
AAH
AIS
MIA
ADC
Histology
B
CD276: p−value = 8.774e−10
Log2 ratio
9
10
N
AAH
AIS
MIA
ADC
Histology
C
CTLA4: p−value = 1.400e−04
Log2 ratio
6
7
8
9
N
AAH
AIS
MIA
ADC
Histology
D
ENTPD1: p−value = 1.336e−05
Log2 ratio
7
8
9
N
AAH
AIS
MIA
ADC
Histology
E
GZMB: p−value = 5.438e−10
Log2 ratio
6
7
8
9
N
AAH
AIS
MIA
ADC
Histology
F
PRF1: p−value = 1.913e−08
Log2 ratio
7
8
9
10
N
AAH
AIS
MIA
ADC
Histology

Figure S2

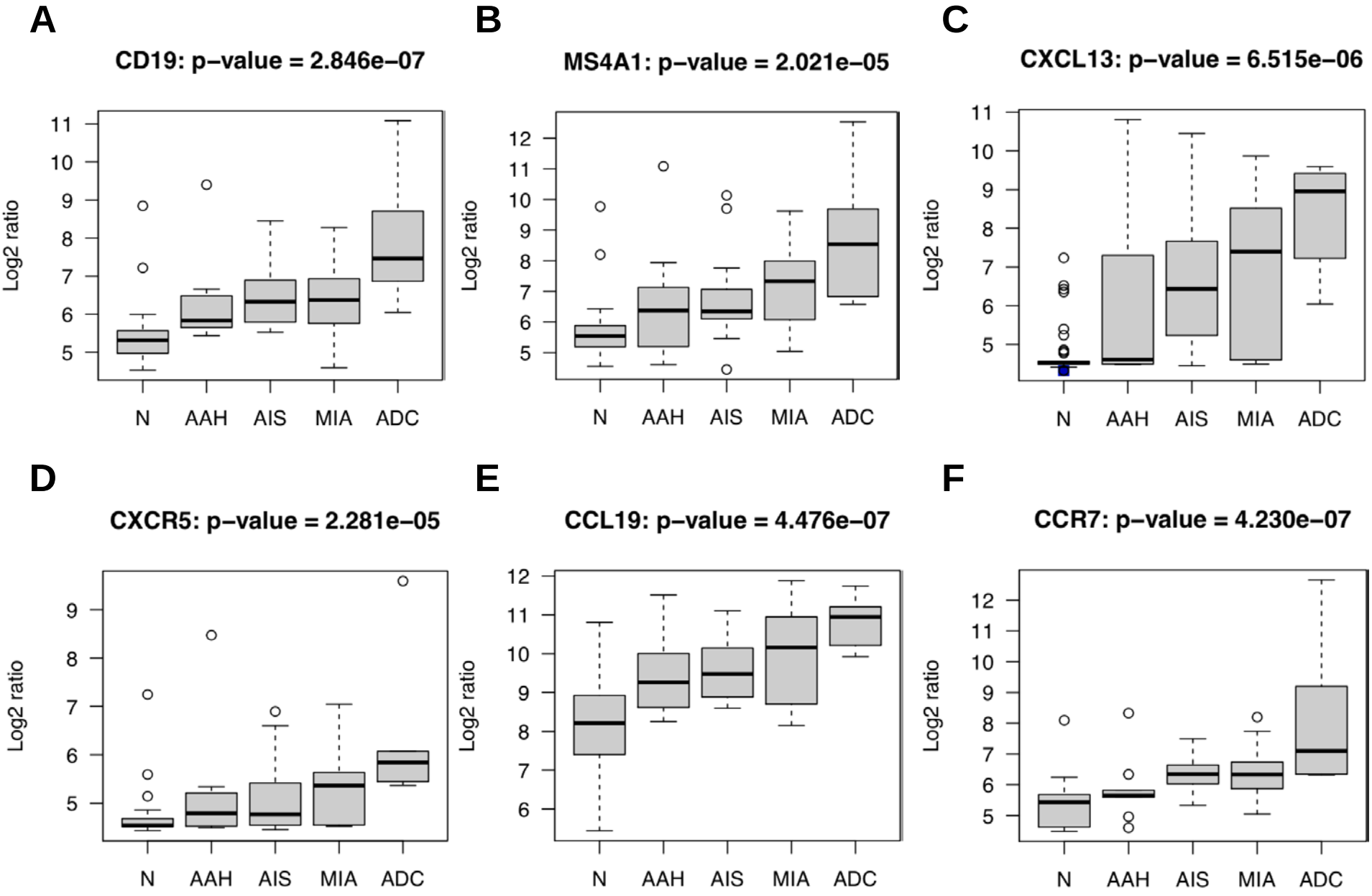

A
CD19: p−value = 2.846e−07
B
MS4A1: p−value = 2.021e−05
C
CXCL13: p−value = 6.515e−06
D
CXCR5: p−value = 2.281e−05
E
CCL19: p−value = 4.476e−07
F
CCR7: p−value = 4.230e−07
Log2 ratio
N
AAH
AIS
MIA
ADC

Figure S3

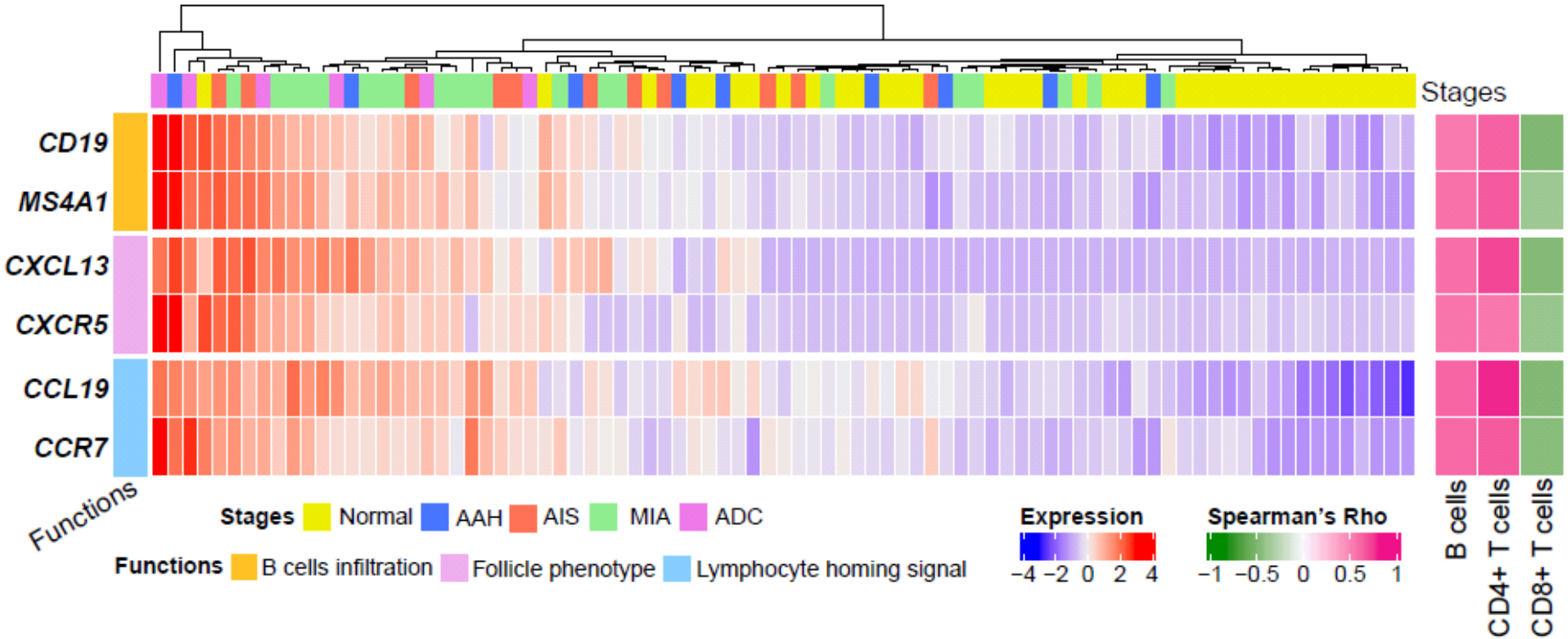

Stages
CD19
MS4A1
CXCL13
CXCR5
CCL19
CCR7
Functions
Stages Normal AAH AIS MIA ADC
Functions B cells infiltration Follicle phenotype Lymphocyte homing signal
Expression
−4 −2 0 2 4
Spearman's Rho
−1 −0.5 0 0.5 1
B cells
CD4+ T cells
CD8+ T cells

Figure S4

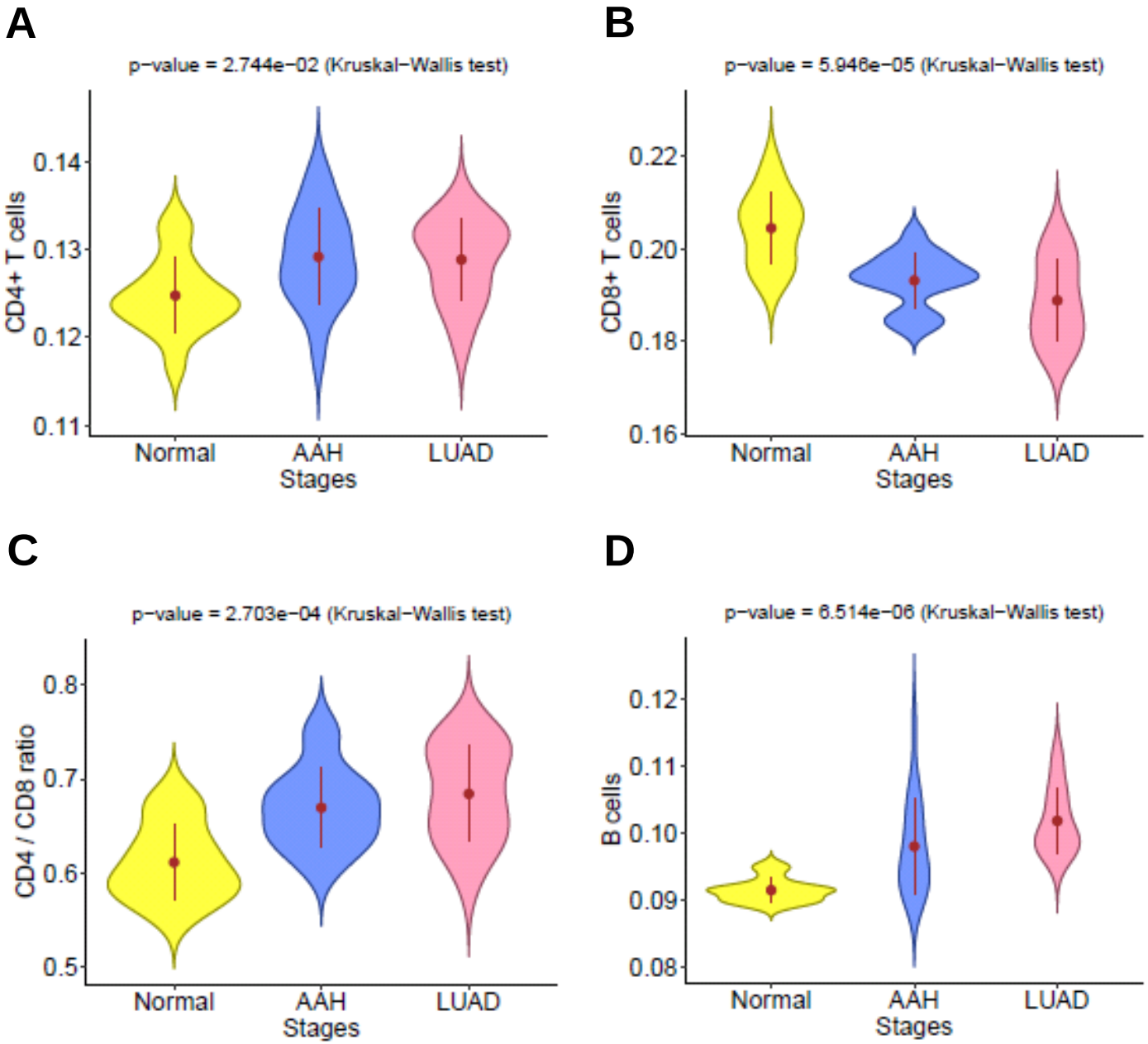
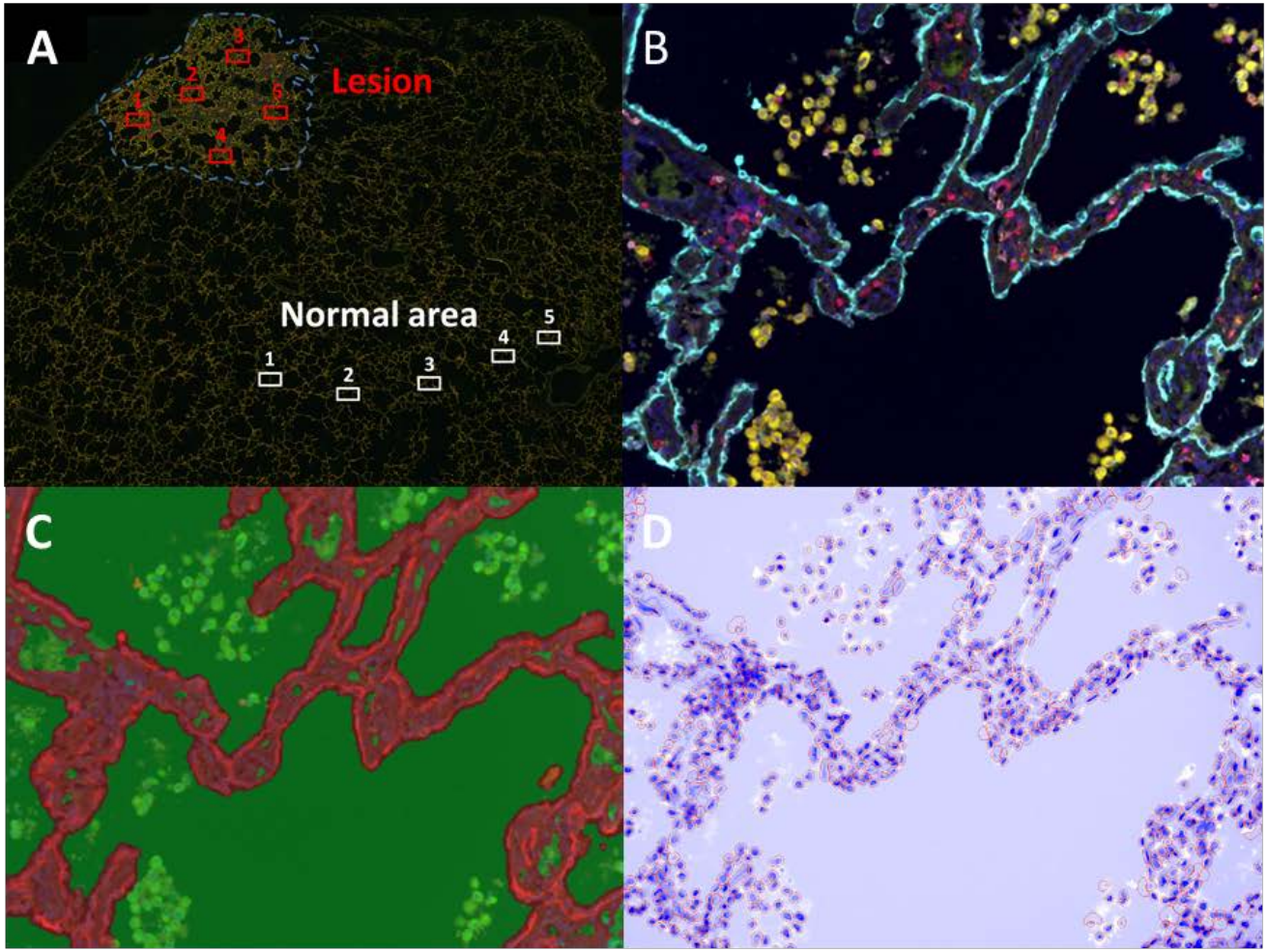

A
p−value = 2.744e−02 (Kruskal−Wallis test)
CD4+ T cells
0.14
0.13
0.12
0.11
Normal
AAH
LUAD
Stages
B
p−value = 5.946e−05 (Kruskal−Wallis test)
CD8+ T cells
0.22
0.20
0.18
0.16
Normal
AAH
LUAD
Stages
C
p−value = 2.703e−04 (Kruskal−Wallis test)
CD4 / CD8 ratio
0.8
0.7
0.6
0.5
Normal
AAH
LUAD
Stages
D
p−value = 6.514e−06 (Kruskal−Wallis test)
B cells
0.12
0.11
0.10
0.09
0.08
Normal
AAH
LUAD
Stages

Figure S5



A. Area selection

B. Composite image

CK CD3 CD8

CD68 PD-1 PD-L1

DAPI

C. Tissue Segmentation

Epithelium

Alveolar space

D. Cell Segmentation

DAPI

Figure S6

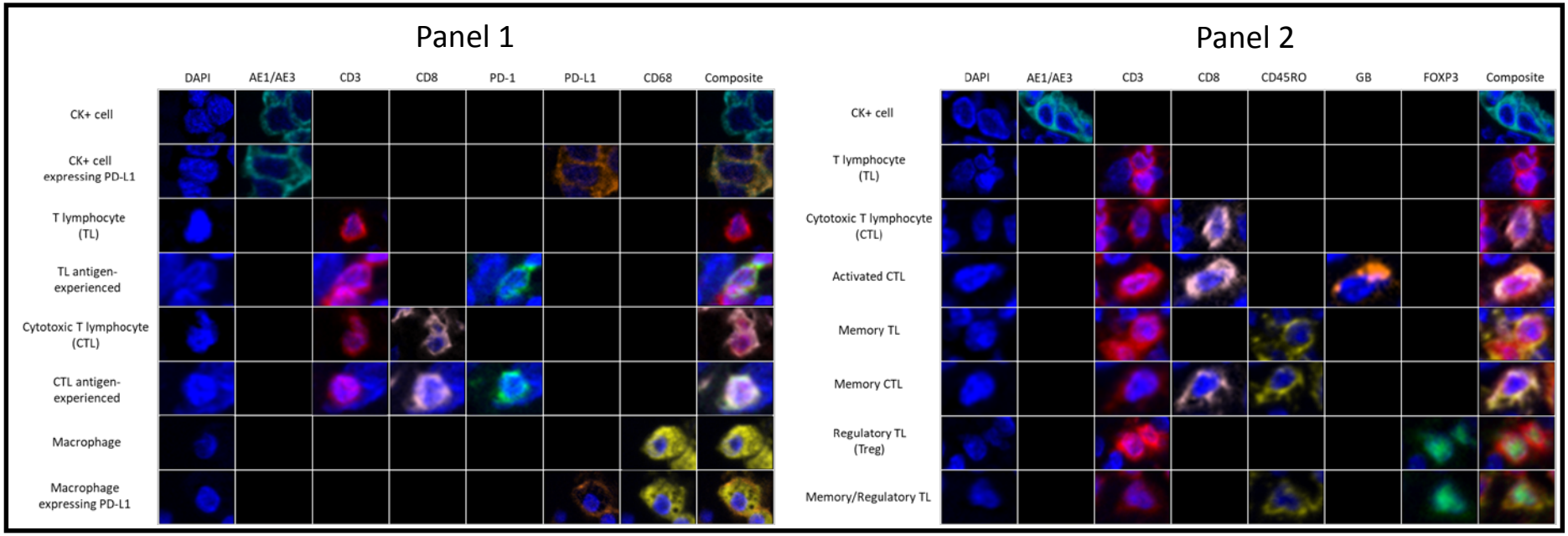

Panel 1
DAPI
AE1/AE3
CD3
CD8
PD-1
PD-L1
CD68
Composite
CK+ cell
CK+ cell expressing PD-L1
T lymphocyte (TL)
TL antigen-experienced
Cytotoxic T lymphocyte (CTL)
CTL antigen-experienced
Macrophage
Macrophage expressing PD-L1
Panel 2
DAPI
AE1/AE3
CD3
CD8
CD45RO
GB
FOXP3
Composite
CK+ cell
T lymphocyte (TL)
Cytotoxic T lymphocyte (CTL)
Activated CTL
Memory TL
Memory CTL
Regulatory TL (Treg)
Memory/Regulatory TL

Figure S7

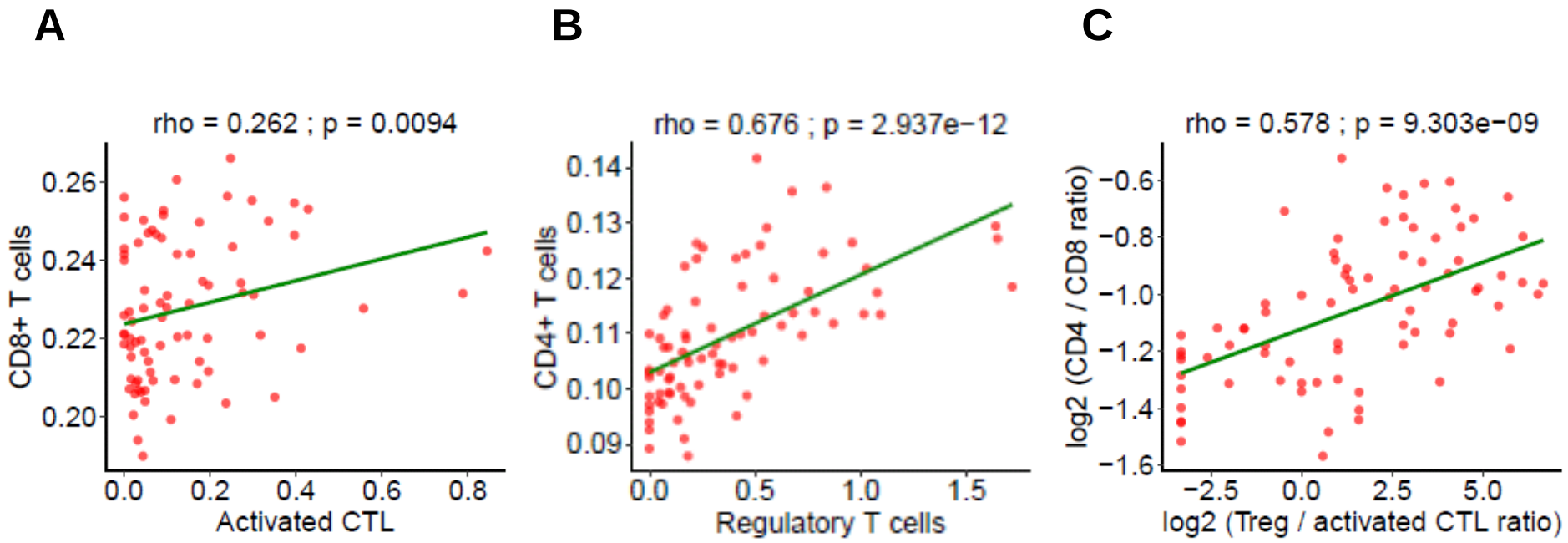

A
rho = 0.262 ; p = 0.0094
CD8+ T cells
0.26
0.24
0.22
0.20
0.0
0.2
0.4
0.6
0.8
Activated CTL
B
rho = 0.676 ; p = 2.937e−12
CD4+ T cells
0.14
0.13
0.12
0.11
0.10
0.09
0.0
0.5
1.0
1.5
Regulatory T cells
C
rho = 0.578 ; p = 9.303e−09
log2 (CD4 / CD8 ratio)
−0.6
−0.8
−1.0
−1.2
−1.4
−1.6
−2.5
0.0
2.5
5.0
log2 (Treg / activated CTL ratio)

Figure S8

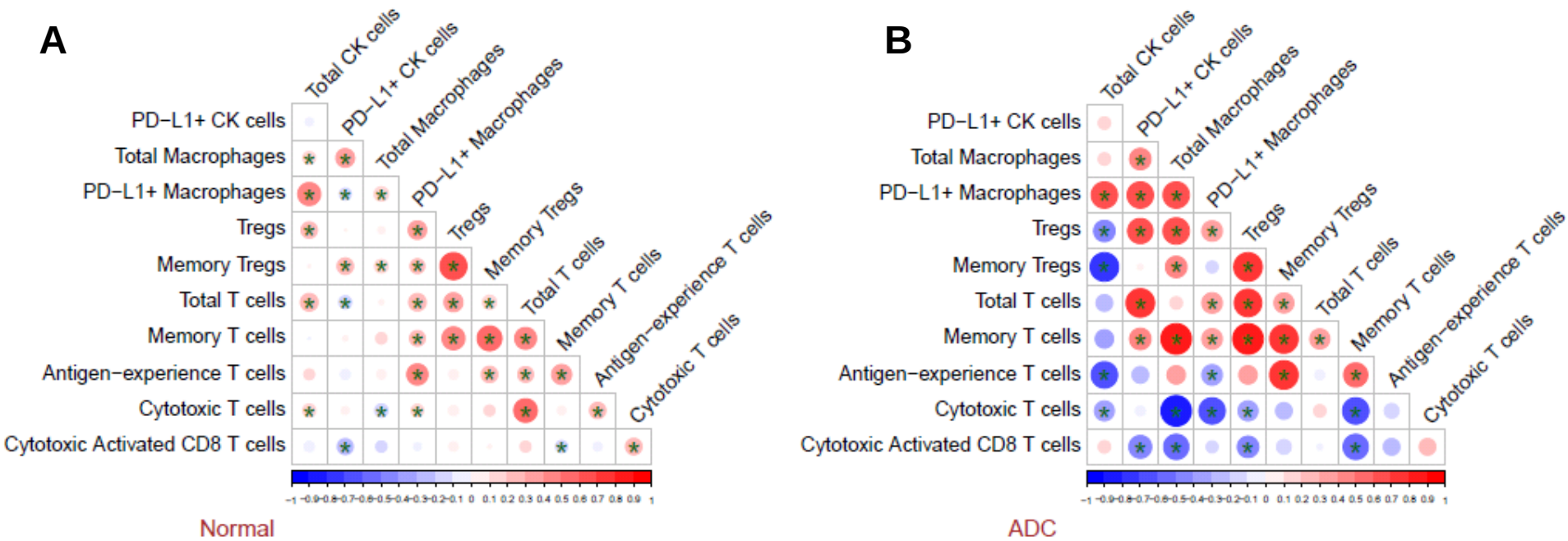

A
Total CK cells
PD−L1+ CK cells
Total Macrophages
PD−L1+ Macrophages
Tregs
Memory Tregs
Total T cells
Memory T cells
Antigen−experience T cells
Cytotoxic T cells
Cytotoxic Activated CD8 T cells
−1 −0.9 −0.8 −0.7 −0.6 −0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
Normal
B
Total CK cells
PD−L1+ CK cells
Total Macrophages
PD−L1+ Macrophages
Tregs
Memory Tregs
Total T cells
Memory T cells
Antigen−experience T cells
Cytotoxic T cells
Cytotoxic Activated CD8 T cells
−1 −0.9 −0.8 −0.7 −0.6 −0.5 −0.4 −0.3 −0.2 −0.1 0 0.1 0.2 0.3 0.4 0.5 0.6 0.7 0.8 0.9 1
ADC

Figure S9

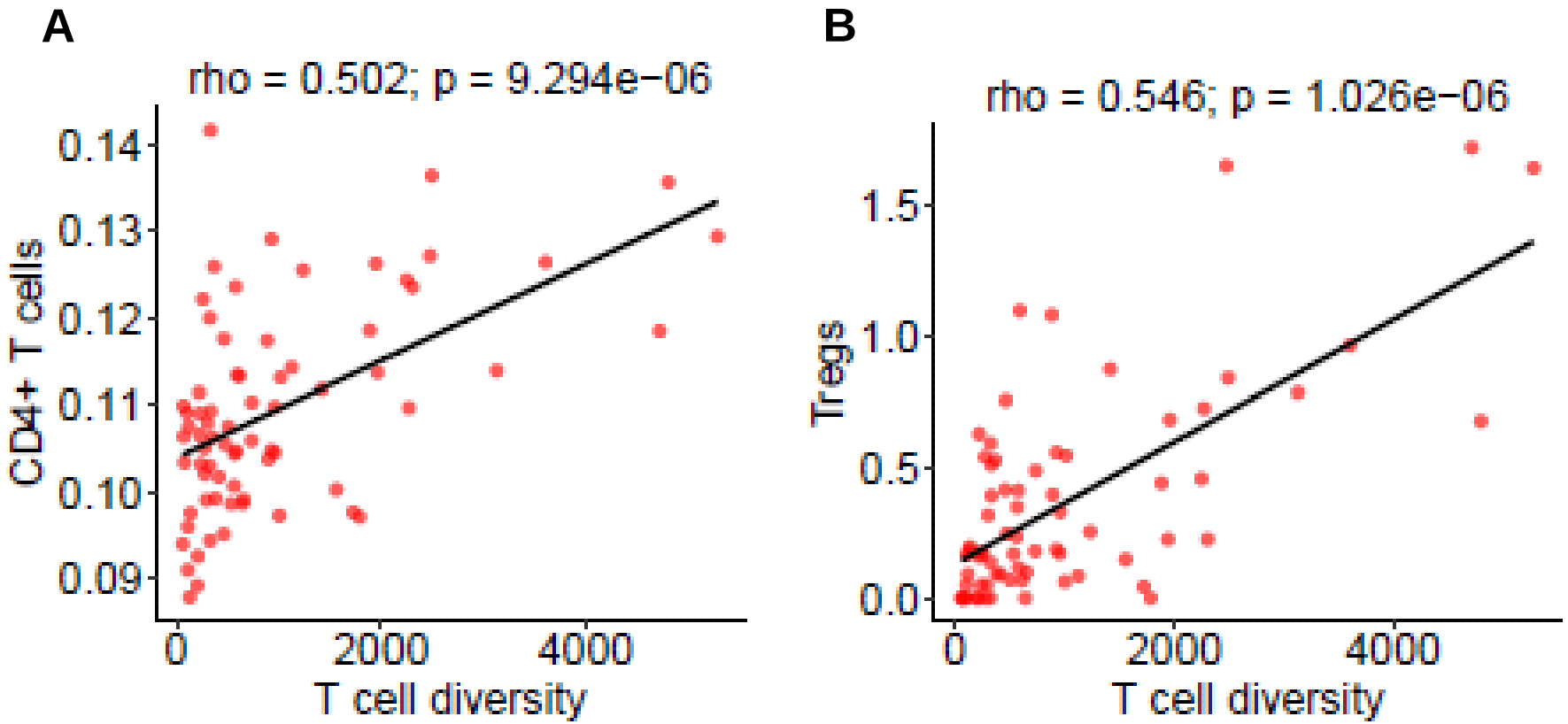

A
rho = 0.502; p = 9.294e−06
CD4+ T cells
0.14
0.13
0.12
0.11
0.10
0.09
0
2000
4000
T cell diversity
B
rho = 0.546; p = 1.026e−06
Tregs
1.5
1.0
0.5
0.0
0
2000
4000
T cell diversity

Figure S10

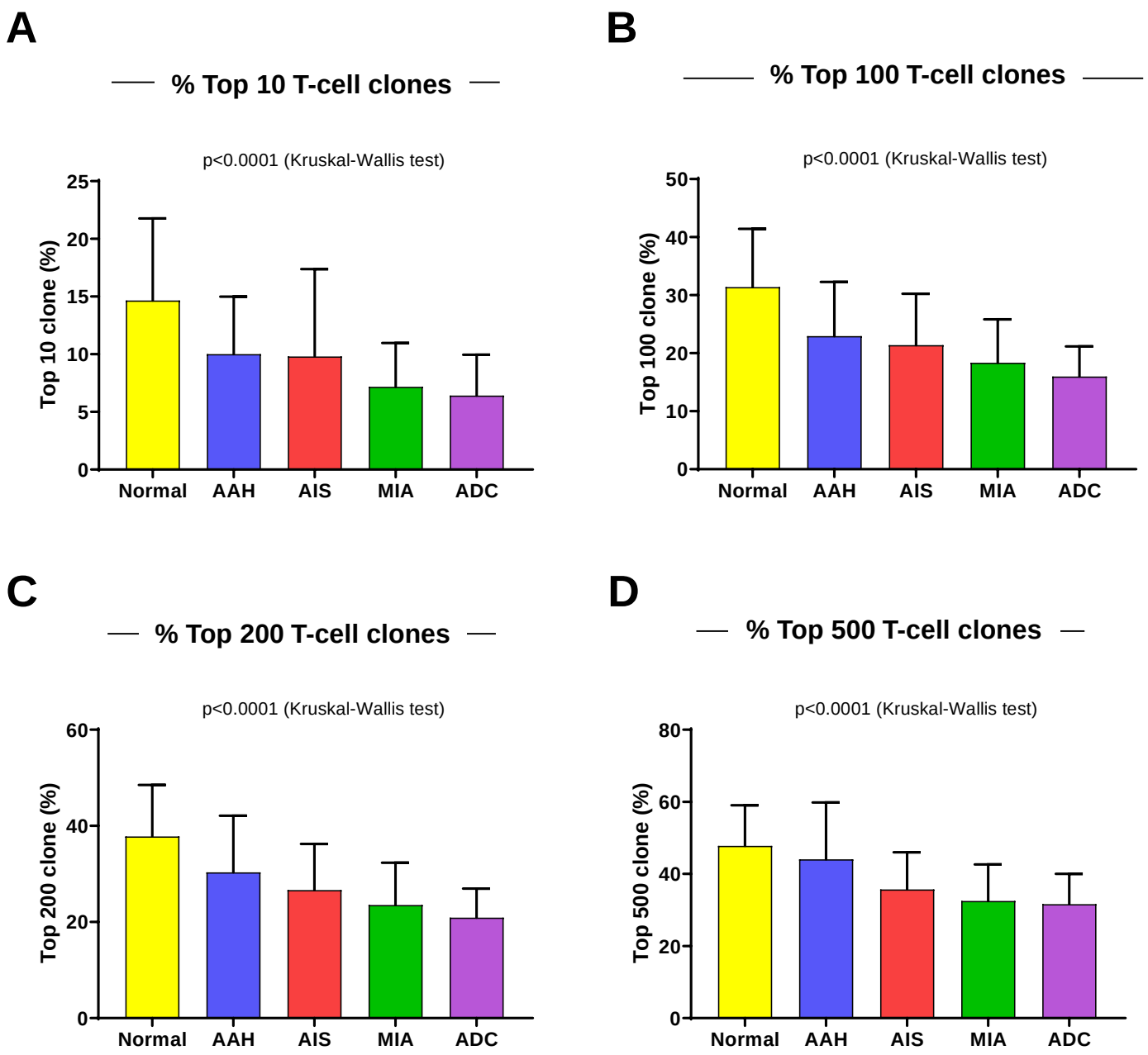

A
% Top 10 T-cell clones
p<0.0001 (Kruskal-Wallis test)
Top 10 clone (%)
25
20
15
10
5
0
Normal
AAH
AIS
MIA
ADC
B
% Top 100 T-cell clones
p<0.0001 (Kruskal-Wallis test)
Top 100 clone (%)
50
40
30
20
10
0
Normal
AAH
AIS
MIA
ADC
C
% Top 200 T-cell clones
p<0.0001 (Kruskal-Wallis test)
Top 200 clone (%)
60
40
20
0
Normal
AAH
AIS
MIA
ADC
D
% Top 500 T-cell clones
p<0.0001 (Kruskal-Wallis test)
Top 500 clone (%)
80
60
40
20
0
Normal
AAH
AIS
MIA
ADC

Figure S11

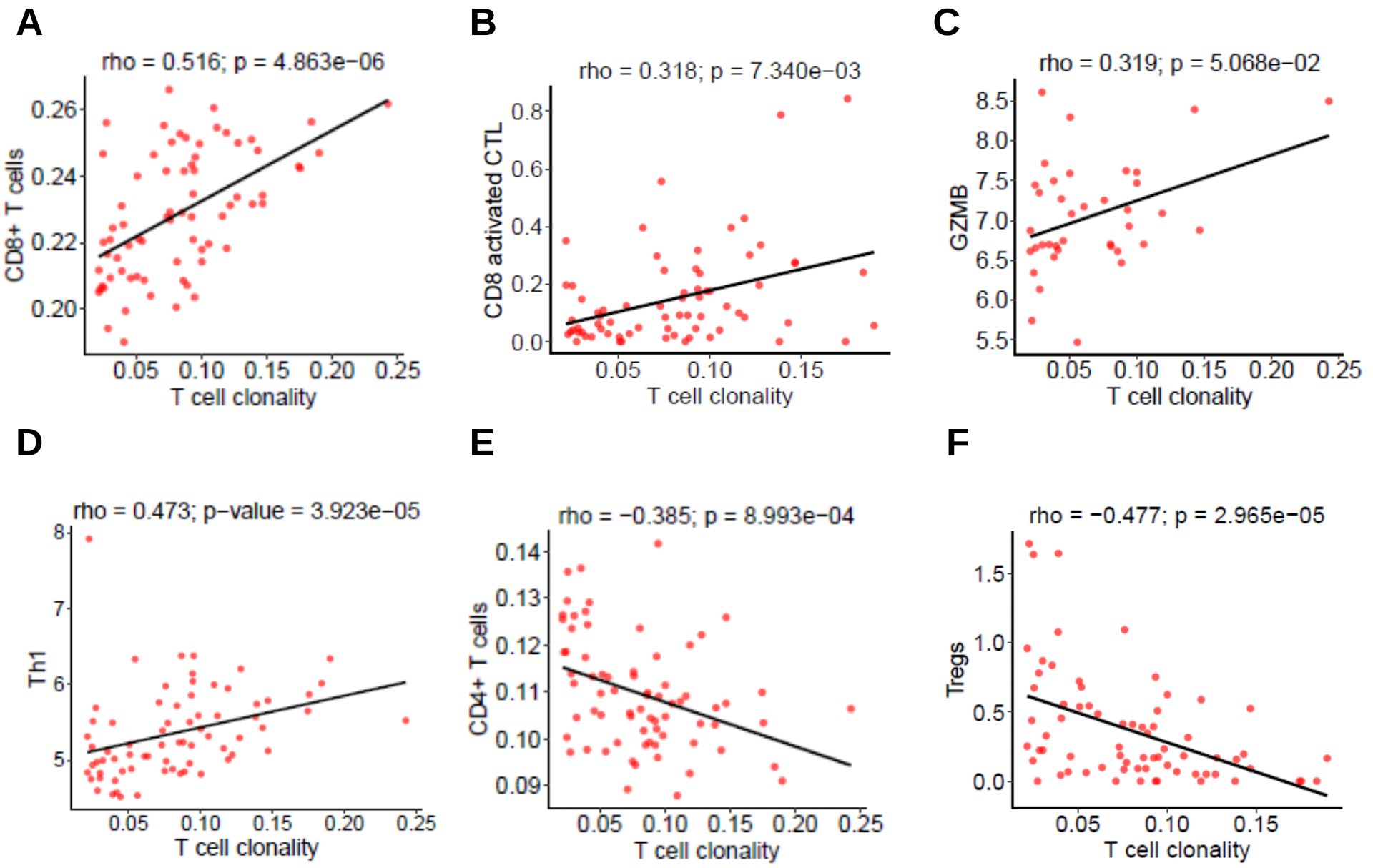

A
rho = 0.516; p = 4.863e−06
CD8+ T cells
0.20
0.22
0.24
0.26
0.05
0.10
0.15
0.20
0.25
T cell clonality
B
rho = 0.318; p = 7.340e−03
CD8 activated CTL
0.0
0.2
0.4
0.6
0.8
0.05
0.10
0.15
T cell clonality
C
rho = 0.319; p = 5.068e−02
GZMB
5.5
6.0
6.5
7.0
7.5
8.0
8.5
0.05
0.10
0.15
0.20
0.25
T cell clonality
D
rho = 0.473; p−value = 3.923e−05
Th1
5
6
7
8
0.05
0.10
0.15
0.20
0.25
T cell clonality
E
rho = −0.385; p = 8.993e−04
CD4+ T cells
0.09
0.10
0.11
0.12
0.13
0.14
0.05
0.10
0.15
0.20
0.25
T cell clonality
F
rho = −0.477; p = 2.965e−05
Tregs
0.0
0.5
1.0
1.5
0.05
0.10
0.15
T cell clonality

Figure S12

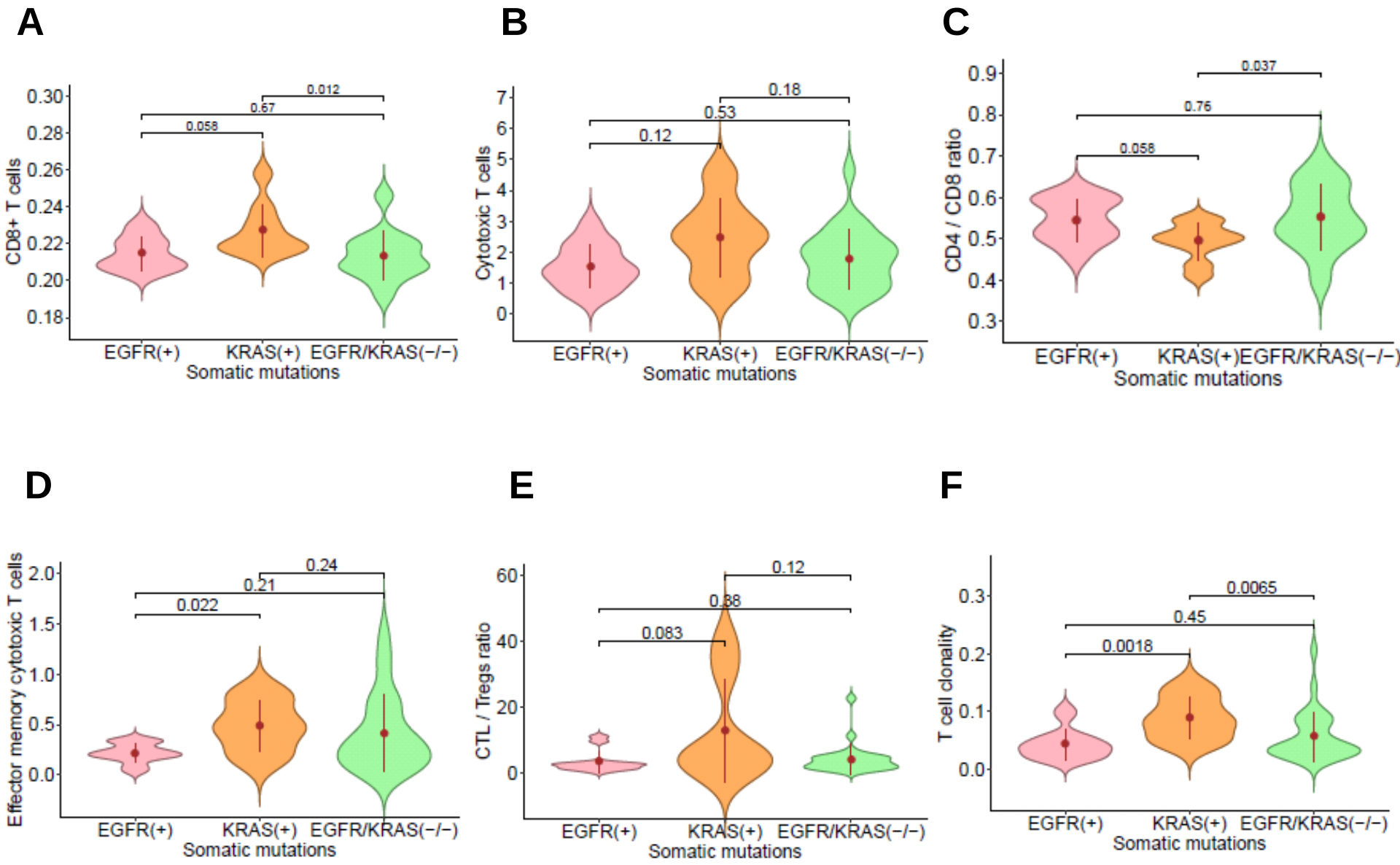

A
CD8+ T cells
0.30
0.28
0.26
0.24
0.22
0.20
0.18
0.012
0.67
0.058
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations
B
Cytotoxic T cells
7
6
5
4
3
2
1
0
0.18
0.53
0.12
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations
C
CD4 / CD8 ratio
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.037
0.76
0.058
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations
D
Effector memory cytotoxic T cells
2.0
1.5
1.0
0.5
0.0
0.24
0.21
0.022
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations
E
CTL / Tregs ratio
60
40
20
0
0.12
0.88
0.083
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations
F
T cell clonality
0.3
0.2
0.1
0.0
0.0065
0.45
0.0018
EGFR(+)
KRAS(+)
EGFR/KRAS(−/−)
Somatic mutations

Figure S13

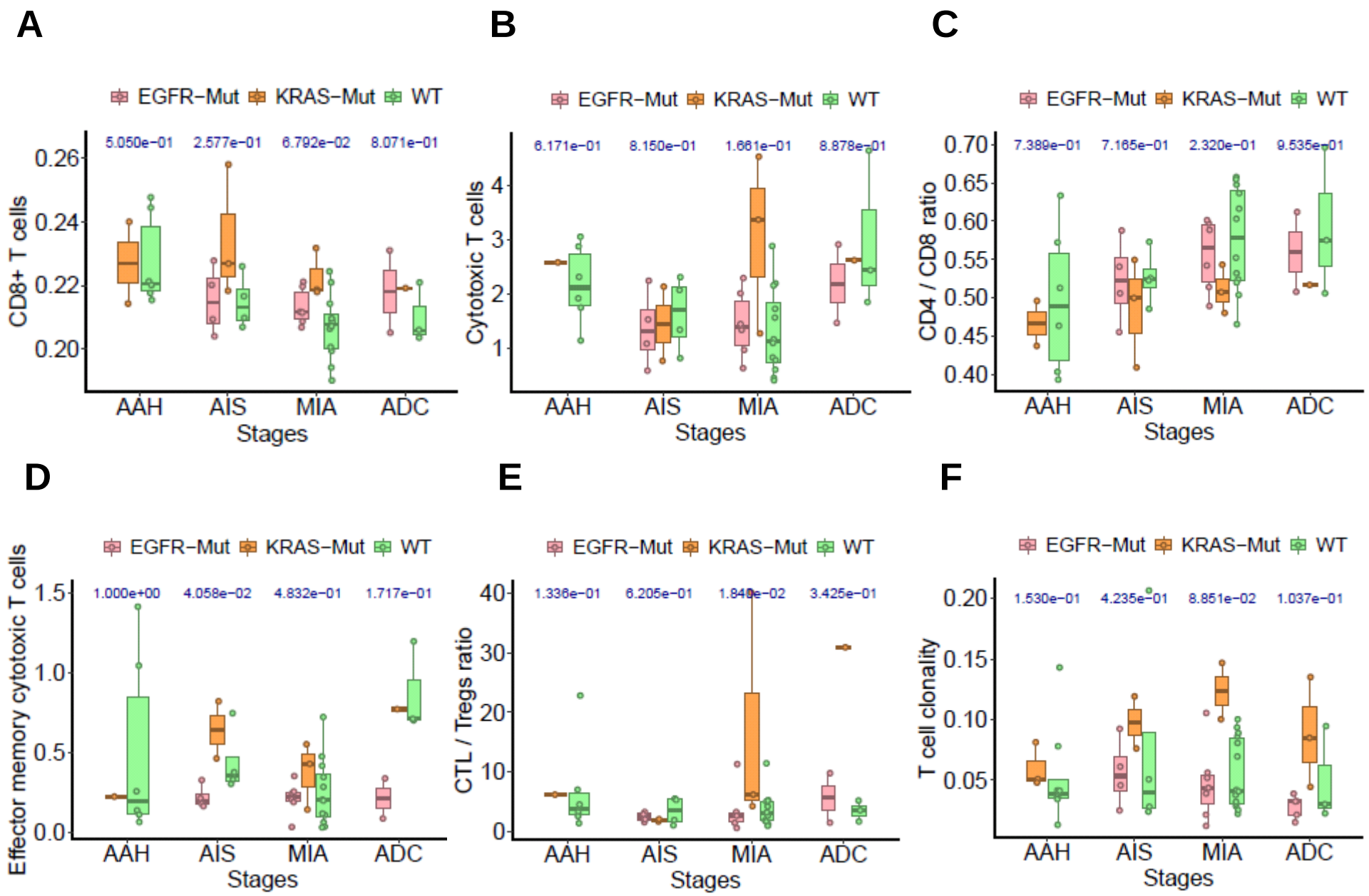

A
EGFR−Mut KRAS−Mut WT
5.050e−01 2.577e−01 6.792e−02 8.071e−01
CD8+ T cells
0.26
0.24
0.22
0.20
AAH AIS MIA ADC
Stages
B
EGFR−Mut KRAS−Mut WT
6.171e−01 8.150e−01 1.661e−01 8.878e−01
Cytotoxic T cells
4
3
2
1
AAH AIS MIA ADC
Stages
C
EGFR−Mut KRAS−Mut WT
7.389e−01 7.165e−01 2.320e−01 9.535e−01
CD4 / CD8 ratio
0.70
0.65
0.60
0.55
0.50
0.45
0.40
AAH AIS MIA ADC
Stages
D
EGFR−Mut KRAS−Mut WT
1.000e+00 4.058e−02 4.832e−01 1.717e−01
Effector memory cytotoxic T cells
1.5
1.0
0.5
0.0
AAH AIS MIA ADC
Stages
E
EGFR−Mut KRAS−Mut WT
1.336e−01 6.205e−01 1.840e−02 3.425e−01
CTL / Tregs ratio
40
30
20
10
0
AAH AIS MIA ADC
Stages
F
EGFR−Mut KRAS−Mut WT
1.530e−01 4.235e−01 8.851e−02 1.037e−01
T cell clonality
0.20
0.15
0.10
0.05
AAH AIS MIA ADC
Stages

Figure S14

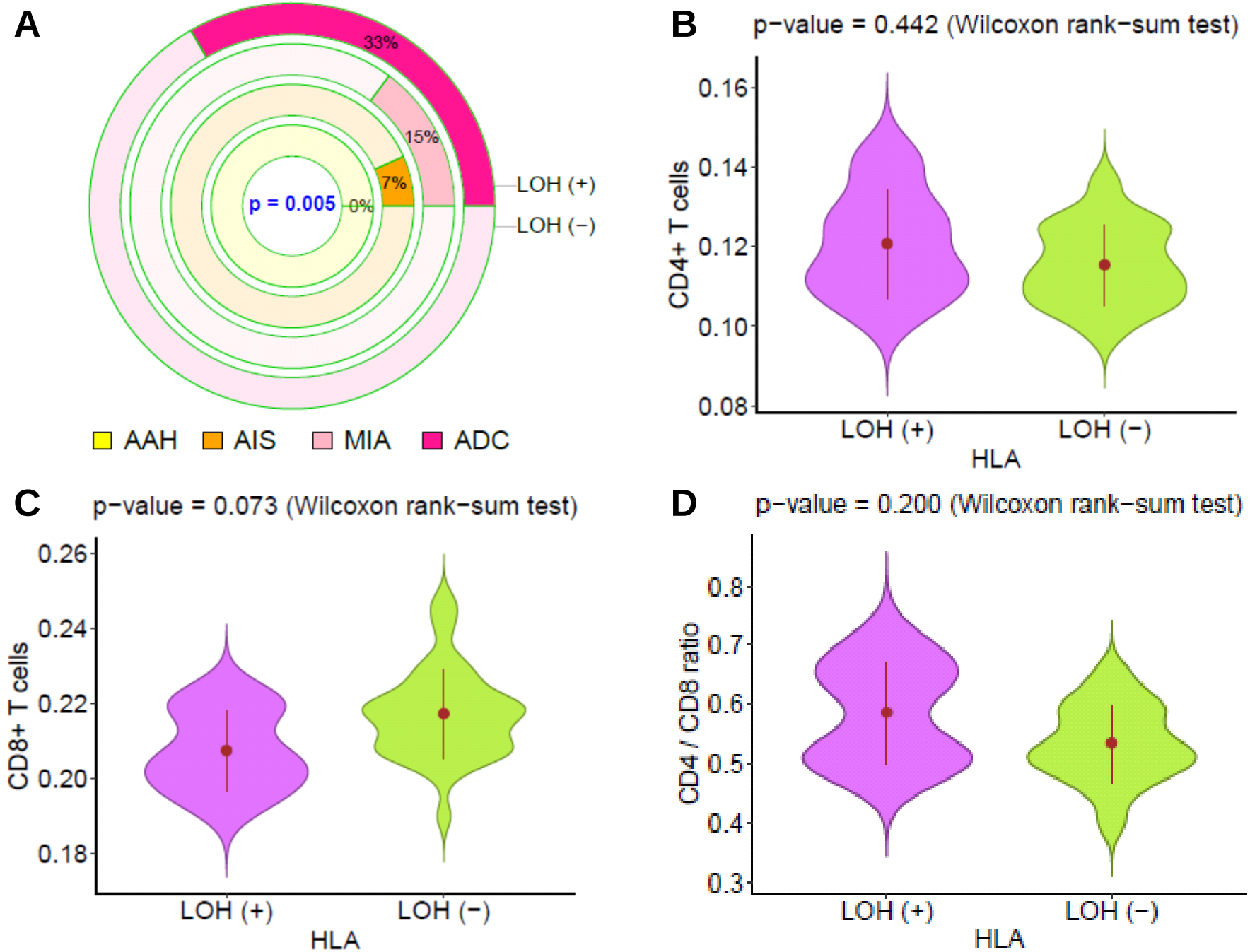

A
33%
15%
7%
0%
p = 0.005
LOH (+)
LOH (−)
AAH
AIS
MIA
ADC
B
p−value = 0.442 (Wilcoxon rank−sum test)
CD4+ T cells
0.16
0.14
0.12
0.10
0.08
LOH (+)
LOH (−)
HLA
C
p−value = 0.073 (Wilcoxon rank−sum test)
CD8+ T cells
0.26
0.24
0.22
0.20
0.18
LOH (+)
LOH (−)
HLA
D
p−value = 0.200 (Wilcoxon rank−sum test)
CD4 / CD8 ratio
0.8
0.7
0.6
0.5
0.4
0.3
LOH (+)
LOH (−)
HLA

Figure S15

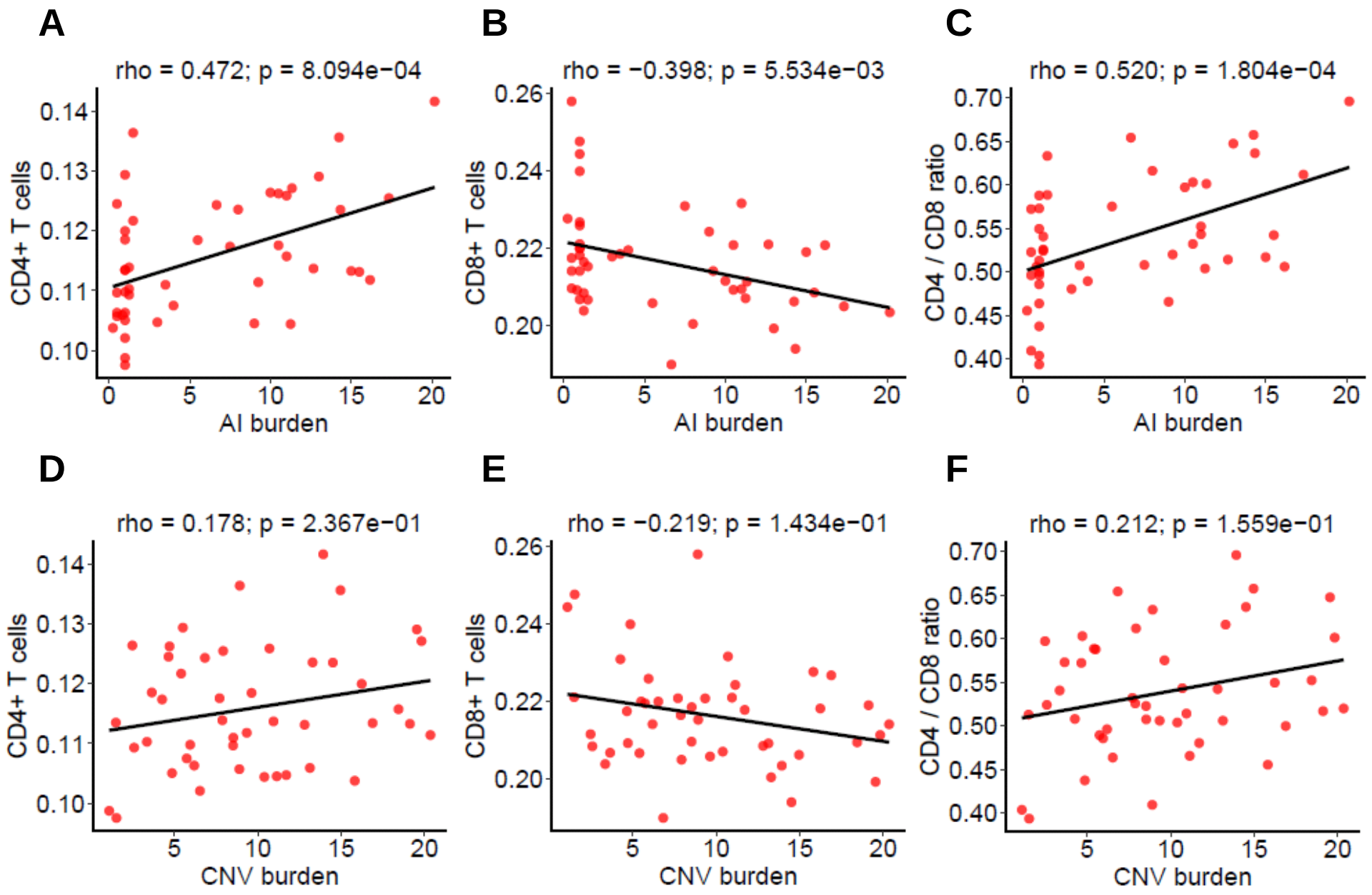

A
rho = 0.472; p = 8.094e−04
CD4+ T cells
AI burden
B
rho = −0.398; p = 5.534e−03
CD8+ T cells
AI burden
C
rho = 0.520; p = 1.804e−04
CD4 / CD8 ratio
AI burden
D
rho = 0.178; p = 2.367e−01
CD4+ T cells
CNV burden
E
rho = −0.219; p = 1.434e−01
CD8+ T cells
CNV burden
F
rho = 0.212; p = 1.559e−01
CD4 / CD8 ratio
CNV burden

Figure S16

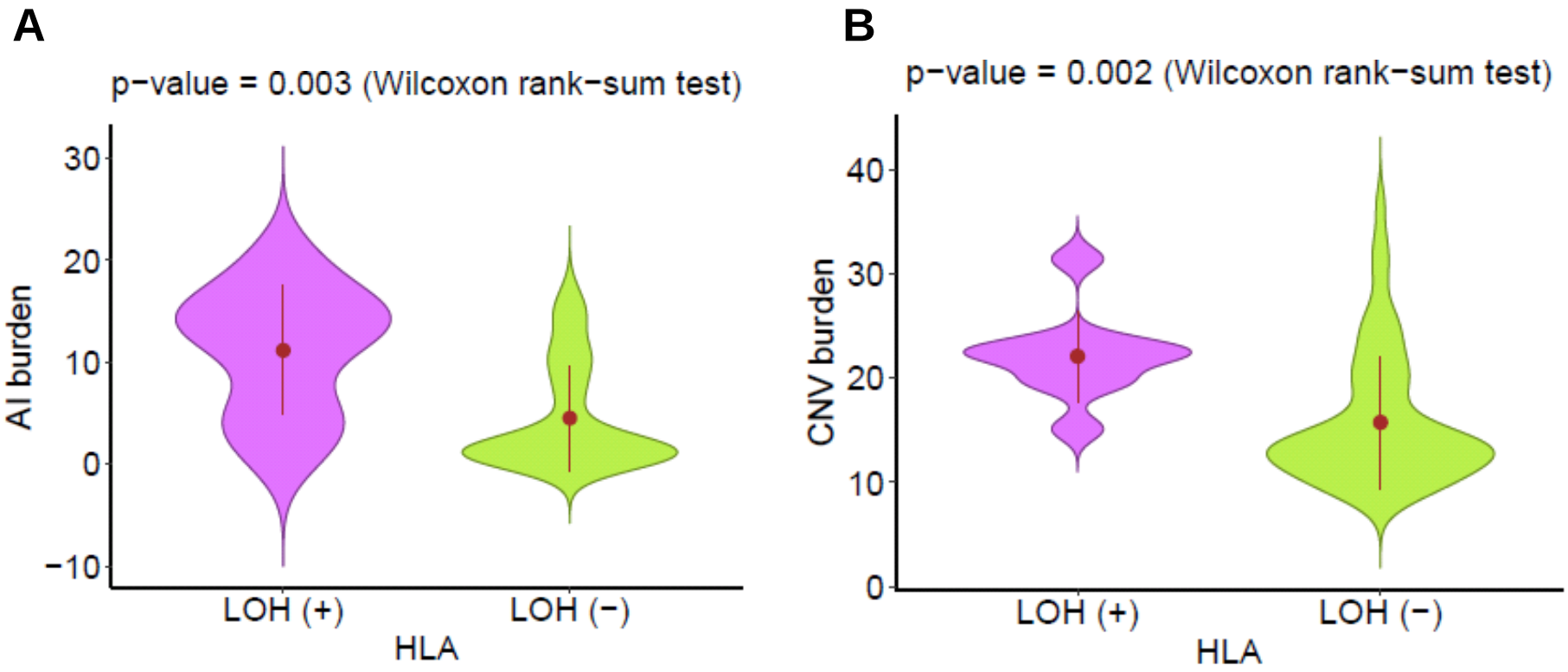

A
p−value = 0.003 (Wilcoxon rank−sum test)
AI burden
30
20
10
0
−10
LOH (+)
LOH (−)
HLA
B
p−value = 0.002 (Wilcoxon rank−sum test)
CNV burden
40
30
20
10
0
LOH (+)
LOH (−)
HLA

Figure S17

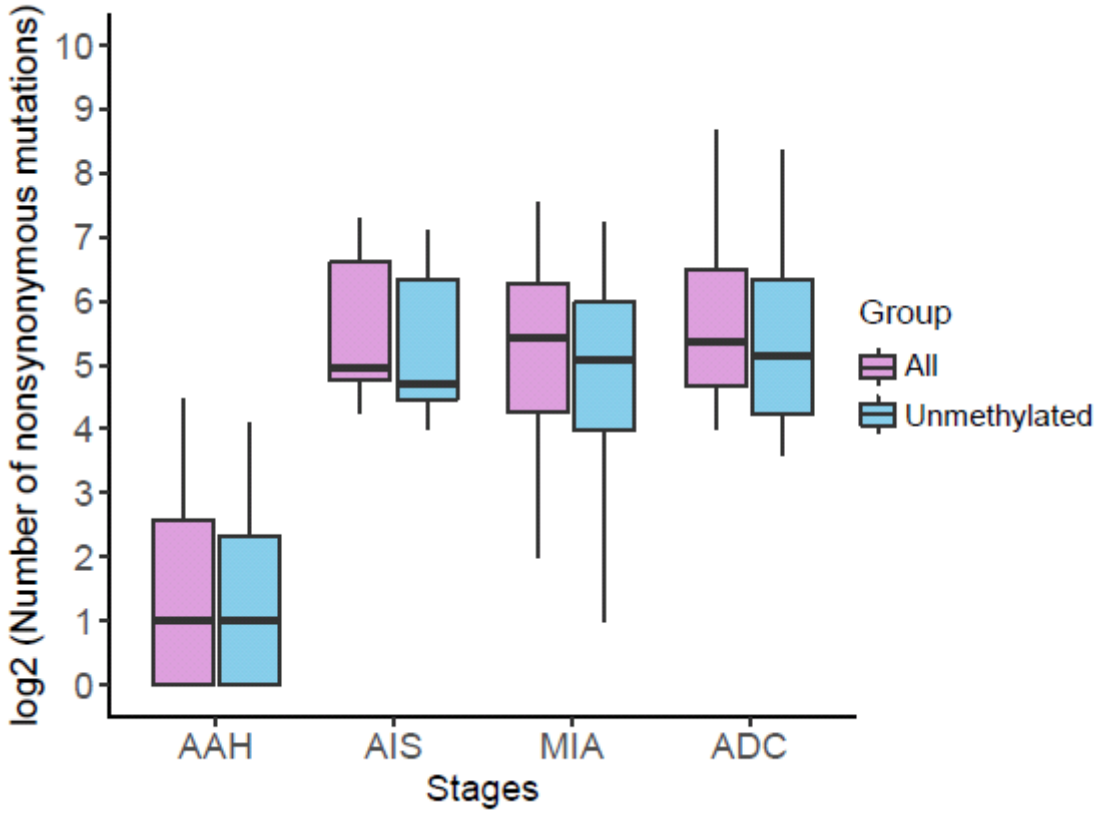
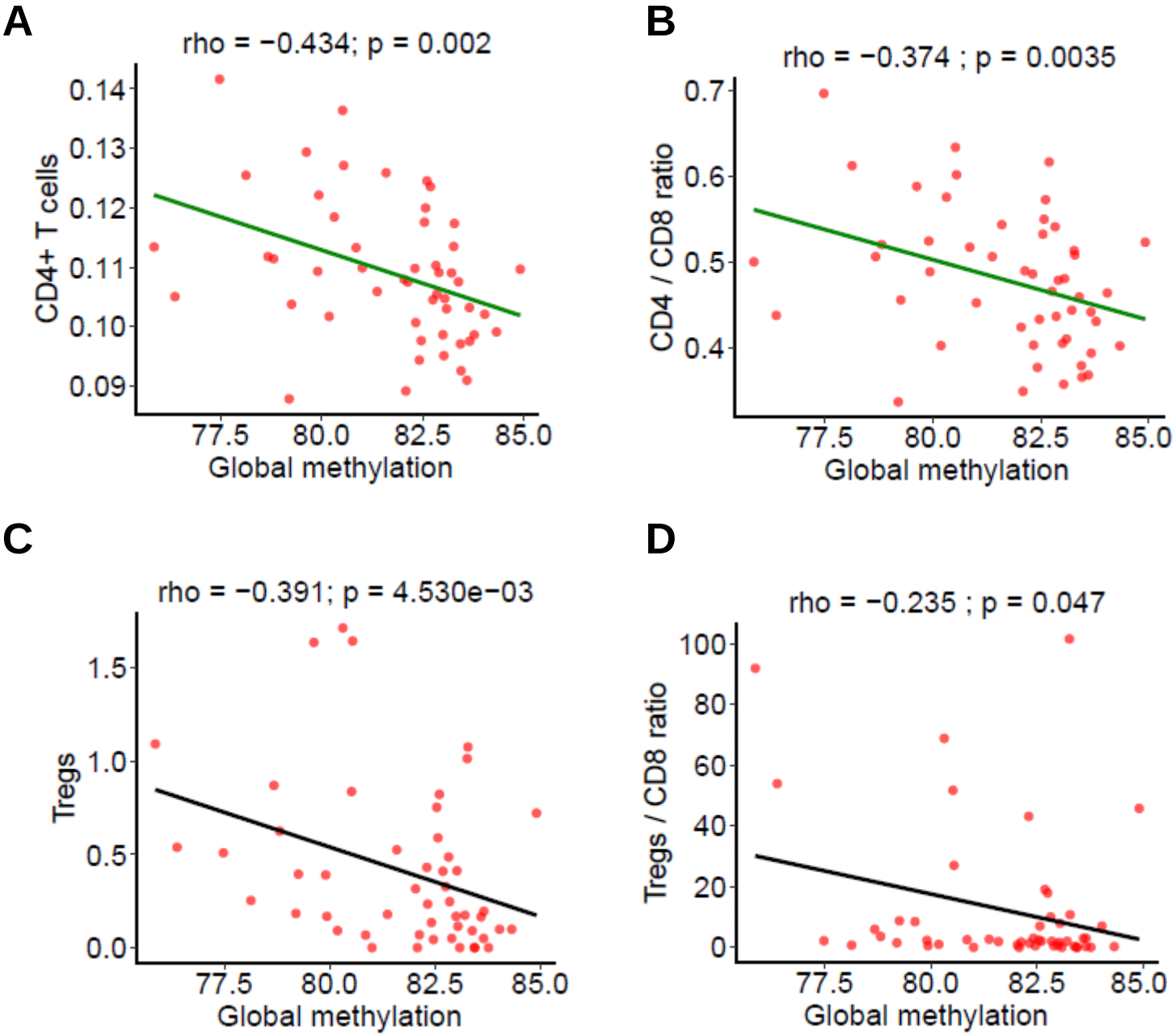

log2 (Number of nonsynonymous mutations)
10
9
8
7
6
5
4
3
2
1
0
AAH
AIS
MIA
ADC
Stages
Group
All
Unmethylated

Figure S18


A
rho = −0.434; p = 0.002
CD4+ T cells
0.14
0.13
0.12
0.11
0.10
0.09
77.5
80.0
82.5
85.0
Global methylation
B
rho = −0.374 ; p = 0.0035
CD4 / CD8 ratio
0.7
0.6
0.5
0.4
77.5
80.0
82.5
85.0
Global methylation
C
rho = −0.391; p = 4.530e−03
Tregs
1.5
1.0
0.5
0.0
77.5
80.0
82.5
85.0
Global methylation
D
rho = −0.235 ; p = 0.047
Tregs / CD8 ratio
100
80
60
40
20
0
77.5
80.0
82.5
85.0
Global methylation