medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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 evolution from preneoplasia to invasive lung adenocarcinomas and underlying molecular features

Hitoshi Dejima[1]*, Xin Hu[2]*, Runzhe Chen[3]*, Jiexin Zhang[4]*, Junya Fujimoto[1]*, Edwin R. Parra[1], Cara Haymaker[1], Shawna Hubert[3], Dzifa Duose[1], Solis Soto[1], Dan Su[5,6], Junya Fukuoka[7], Hoa Pham[7], Nicholas Mcgranahan[8], Baili Zhang[1], 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#]

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.

[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.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.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 July 14, 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 The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.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.

[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.

* Co-first authors

[#] Correspondence should be addressed to Alexandre Reuben (AReuben@mdanderson.org), Humam Kadara (HKadara@mdanderson.org), Ignacio Wistuba (iiwistuba@mdanderson.org) or Jianjun Zhang (jzhang20@mdanderson.org).

ABSTRACT

The evolution of anti-tumor immune surveillance during initiation and progression of preneoplasia into invasive lung adenocarcinoma (ADC) and its underlying molecular changes are largely unknown. To fill this void, we characterized the immune contexture of invasive lung ADC (n=12) and its precursors of consecutive developmental stages including preneoplasia atypical adenomatous hyperplasia (AAH, n=22), adenocarcinoma *in situ* (AIS, n=16), minimally invasive adenocarcinoma (MIA, n=28) as well as paired normal lung tissues (NL, n=53) by transcriptomic profiling of 770 genes of nCounter PanCancer Immune Profiling Panel (Nanostring), T cell receptor (TCR) sequencing and multiplex immunofluorescence (mIF). Our results demonstrated that anti-tumor immunity evolved as a continuum from AAH to AIS, MIA and invasive lung ADC with a gradually less effective and more intensely regulated immune response evidenced by down-regulation of immune activation pathways, up-regulation of immunosuppressive pathways,

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

higher infiltration of CD4+ T cells, lower infiltration of CD8+ T cells, increased CD4/CD8 ratio, decreased T cell clonality, lower frequencies of top T cell clones in later stages. Further correlation of these immune features with exome sequencing and methylation data demonstrated that the immune response could be impacted by oncogene mutation status, HLA loss, chromosomal copy number aberrations and DNA methylation aberrations suggesting that only cells in preneoplasia with the ideal combination of these molecular features enabling rapid proliferation and evasion from host immune could outgrow into the dominant clones in invasive lung ADCs. Importantly, the immune activation and evasion have started at preneoplastic stage advocating for immunotherapy in patients with lung ADC precursors for prevention of invasive lung cancers.

INTRODUCTION

Despite significant advances in its management, lung cancer remains the leading cause of cancer death worldwide (1, 2), largely due to late diagnosis at advanced stages when cures are generally unachievable (1, 3, 4). Computed tomography (CT) scan-guided lung cancer screening has demonstrated a reduction of lung cancer mortality by 26%-61% (5) 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 (5), 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 (6-12), which evolves into preinvasive adenocarcinoma *in situ* (AIS) (13), to micro-invasive lesion termed minimally invasive adenocarcinoma (MIA) (8, 14) and eventually frankly invasive ADC (8, 14). Early-stage lung ADCs and their precursors usually present as lung nodules with distinct radiologic 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 (15).

Carcinogenesis results from progressive accumulation of molecular abnormalities (molecular evolution) (16) and escape from host immune surveillance (immunoediting) (17). Our recent gene expression and genomic pilot studies on lung ADC precursors have demonstrated distinct transcriptomic features (18) and progressive genomic evolution along the spectrum of AAH to AIS, MIA and ADC (19). 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

whole exome sequencing (WES) (19) and methylation data from the same cohort of IPNs (Hu et al, bioRxiv, 2020) to underscore the genomic and epigenetic alterations that may impinge on these immune features (**Supplementary Table 1** and **Fig. 1**).

RESULTS

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 (20) on 47 resected pulmonary nodules (n=9 for AAH, n=11 for AIS, n=21 for MIA, and n=6 for invasive ADC) and 38 paired NL tissues. There were no differences in age (p=0.55, Kruskal test), sex (p=0.31, fisher exact test) or smoking status (p=0.35, fisher exact test) between patients with pulmonary nodules of different histologic stages. In total, 291 genes were differentially-expressed (DEGs) (**Supplementary Table 2**). Interestingly, changes in the majority of DEGs, regardless of their direction, exhibited a progressive pattern along the spectrum from NL, AAH, AIS, MIA and up to ADC (**Supplementary Fig. 1**). Examples of progressively increased genes included immune suppressive genes *CD47* (protection of cancer cells from immune cell killing) (21), *CD276* (inhibition of immune responses) (22) and *CTLA4* (checkpoint molecule) (23), while progressively decreased genes included *ENTPD1* (expressed on tumor-specific T cells)

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

(24), granzyme B (*GZMB*) and perforin 1 (*PRF1*) (two cytotoxic molecules produced by T lymphocytes and natural killer cells (NK cells) (25, 26) (**Supplementary Fig. 2A-F**). These results suggest that immune evolution may have occurred in premalignant phases of lung carcinogenesis and progressed as a continuum from preneoplasia to invasive lung ADC with a gradually less effective and more intensely regulated immune response. 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 (**Fig. 2A**). All 3 up-regulated pathways, systemic lupus erythematosus (SLE) in B cell signaling, T cell exhaustion signaling and PARP signaling pathways could potentially impair immune response (27-29). Taken together, these results indicated an overall decreased immunity in later-stage lesions.

We next de-convoluted gene expression profiling data using TIMER (30) to evaluate changes in immune cell composition. As shown in **Fig. 2B**, CD4+ T lymphocyte infiltration progressively increased from NL to invasive ADC. Conversely, infiltration of CD8+ T lymphocytes progressively decreased with neoplastic evolution (**Fig. 2C**) leading to significantly higher CD4/CD8 ratio in later-stage lesions (**Fig. 2D**). Similarly, B cell infiltration (**Fig. 2E**) and tertiary lymphoid structure signatures progressively increased along the spectrum from NL to invasive ADC (**Supplementary Fig. 3**). To validate these findings, we applied TIMER to RNA sequencing data from an independent cohort previously published by our group and observed similar results (**Supplementary Fig. 4A-D**) (18). Further studies are warranted to understand the subtypes and functions of these B cells although concomitant upregulation of IL-10 (**Supplementary Fig. 5**) and SLE in

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

B cell signaling pathway (**Fig. 2A**) could indicate their negative impact on anti-tumor immunity in later-stage lesions.

To further evaluate changes in T cell subsets, we next compared gene expression of cytokines associated with Th1, Th2, Th17 and Treg differentiation and function (31, 32). Th1 cytokines were attenuated in AIS and MIA when compared with AAH and returned to baseline levels in ADC (**Supplementary Fig. 6**). Conversely, Th2 and Treg cytokines were overall higher in later stages (**Supplementary Fig. 6**), potentially suggesting helper T cell reprogramming to pro-tumor phenotypes along with neoplastic evolution.

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 (**Fig. 3A, Supplementary Fig. 7** and **Supplementary Fig. 8**) 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 (**Supplementary Fig. 9A-B**). Furthermore, mIF-derived Treg/activated CTLs ratio closely recapitulated CD4/CD8 ratio from immune gene expression profiling (**Supplementary Fig. 9C**).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 (**Supplementary Fig. 10A**). For AAH, only total macrophages and activated cytotoxic T cells were positively correlated (**Fig. 3B**). In AIS, PD-L1+ macrophages were positively correlated with Tregs, antigen-experienced T cells and cytotoxic T cells (**Fig. 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 (**Fig. 3D**). Finally, in ADC, PD-L1+CK+ cells were correlated with total and PD-L1+ macrophages, PD-L1+ macrophages were correlated with total macrophages, and Tregs were correlated with memory T cells (**Supplementary Fig. 10B**). 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 (33) by multiregional T cell receptor (TCR) sequencing on 158 spatially-separated specimens from 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 (34). T cell diversity was positively correlated with infiltration of CD4+ T cells derived from gene expression profiling (**Supplementary Fig. 11A**) as well as densities of Tregs from mIF (**Supplementary Fig. 11B**) and increased steadily from AAH to AIS, MIA and ADC

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

(p<0.0001) **(Fig. 4A)** in line with higher CD4+ T cells in later stage lesions **(Fig. 2B)**. We then compared distributions of T cell clones across histologic stages, focusing on the most expanded T cell clones (35, 36). As shown in **Fig. 4B** and **Supplementary Fig. 12**, 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 T cell repertoire may occur at the preneoplastic stage

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) (**Supplementary Fig. 13A-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) (**Supplementary Fig. 13E-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 **(Fig. 4C)** consistent with our previous findings (25). Clonality declined from NL to AAH increasing in AIS/MIA and declining to its lowest level in ADC **(Fig. 4C).** These results suggest that early ADC pathogenesis is associated with local immunosuppression that may have commenced at the preneoplastic stage.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 probed the two most frequently mutated driver genes in this cohort *EGFR* and *KRAS* (19). 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 from TCR sequencing **(Supplementary Fig. 14A-F)**. These trends remained similar when analyzing each histologic stage (**Supplementary Fig. 15A-F**). These findings are consistent with our previous studies in invasive lung cancers (37) 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 (26), we applied the LOHHLA (loss of heterozygosity in HLA) algorithm (38) to WES data from these lesions (19). LOH at the HLA loci was observed in 7%, 15% and 33% of AIS, MIA and ADC respectively, but in none of AAH lesions **(Supplementary Fig. 16A,** p=0.005, Chi-Square test). Correlation with T cell infiltration derived from immune gene expression profiling demonstrated that

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

lesions with HLA LOH had similar CD4+ T cell infiltration, but trends of lower CD8+ T cell infiltration and higher CD4/CD8 ratio **(Supplementary Fig. 16B-D**).

Chromosomal copy number variation (CNV) has been reported to impact immune microenvironment across different cancers (39). 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 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 **(Supplementary Fig. 17A-C**). Similar trend was also observed with CNV burden, although these differences did not reach statistical significance **(Supplementary Fig. 17D-F**). Interestingly, HLA-LOH positive lesions exhibited significantly higher AI burden and CNV burden compared to that of HLA-LOH negative lesions **(Supplementary Fig. 18A-B**). 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 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 (40, 41). As shown in **Supplementary Fig. 19A-B**, a progressive increase in both MHC-I or MHC-II associated

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

predicted neoantigens 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, bioRxiv, 2020) revealed that a significantly larger proportion of genes associated with predicted neoantigens exhibited promoter hypermethylation (>30% CpG methylated) in later-stage lesions, thus potentially dampening expression of neoantigens in later stage lesions (**Supplementary Fig. 19C-D**). 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 (42-44), on the immune microenvironment in these lesions. Overall, global methylation level was negatively associated with CD4+ T cell infiltration and CD4/CD8 ratio derived from gene expression profiling, Tregs infiltration and Treg/CD8 ratio from mIF (**Supplementary Fig. 20A-D**) 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 (45-47). 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 (48, 49). The T cell immunity is significantly compromised even in stage I lung cancers (37, 50) indicating that these cancers have already begun evading the anti-tumor immune surveillance. However, when

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 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 (19) 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

evolution of preinvasive SCC lesions (51). 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 (19) and RRBS data (Hu, et al BioRxiv, 2020) available from the same cohort of pulmonary nodules provided the opportunity to depict the genomic and epigenetic features associated with the distinct immune landscape. Overall, oncogene mutations, HLA loss, CNV burden, AI burden and methylation status were all found to associate with the immune contexture, similarly to advanced cancers (52, 53). 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(19), primarily by the host immune system. In the early phases of carcinogenesis, the 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

genomic and epigenetic alterations on immune response is intertwined. For example, methylation may directly affect immune response by regulating expression of immune genes (54) or potential neoantigens (55) or indirectly by increasing DNA vulnerability for development of CNV and somatic mutations that can subsequently influence the immune microenvironment (56, 57). Global hypomethylation is known to associate with CIN (58) and increased rate of somatic mutations (59) and indeed, global methylation level was negatively associated with TMB and CNV burden in this cohort of lesions (Hu, et al BioRxiv, 2020). Since high CNV burden is associated with a cold immune microenvironment (39) while high TMB can increase tumor immunogenicity and facilitate immune recognition and elimination of cancer cells (60), 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 the 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 predicted neoantigen 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 (61-64), resulting in an overall cold immune microenvironment in invasive ADCs (lower CD8+ T cell infiltration, lower CTL, lower T cell clonality, higher CD4+ T cell infiltration and higher 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 (65, 66),

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

many of which are lung ADC precursors. However, the 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 (18, 19) that preinvasive lung ADC precursors were molecularly simpler, and therefore theoretically easier to eradicate. In 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.

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 (China) from 2014 to 2017 as described previously (19). 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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) (67), 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 (68, 69). 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 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 (70). The DEGs, along with log ratios were then evaluated with Ingenuity Pathway Analysis (IPA) software (Quiagen, Hilden, Germany) (71) 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 (72). 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, 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 (73).

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 (**Supplementary Fig. 7A**). The selected high magnification field was total 1.6725 mm^2 in size (669×500μm per field) (**Supplementary Fig. 7B**). Each field with panel 1 and panel 2 were overlapped with sequential sections. The target areas were analyzed by in 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 (**Supplementary Fig. 7C**). The individual cells were recognized by DAPI nuclei staining (**Supplementary Fig. 7D**). Panel 1 contained the co-localization patterns

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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+) (**Supplementary Fig. 8**). 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 (37, 74).

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:

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

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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 (19) (Hu, et al, Bioxiv, 2020). Somatic mutations, allelic imbalance (AI), copy number variations (CNV), oncogene mutations, global methylation data were 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 (75). This approach takes into account the base

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

qualities of aligned reads, observed insert sizes, as well as the ethnicity-dependent prior probabilities of each allele (75). Tumor purity and ploidy were estimated using ASCAT (76). We then applied LOHHLA (Loss Of Heterozygosity in Human Leukocyte Antigen) algorithm (38) 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 $p < 0.01$.

Prediction of neoantigens

All peptides with 9-12 amino acids that span missense or stop gain mutations were extracted as mutant peptides to identify candidate peptides binding to MHC Class I or II molecules. The affinity of 9-12 peptides binding to MHC Class I molecules were predicted using the NetMHCPan3.0 binding algorithm (77). The affinity of 9-12 mer peptides binding to MHC Class II molecules were predicted using the NetMHCIIPan3.1 binding algorithm (78). The threshold for strong binding peptides is defined as half-maximum inhibitory concentration (IC50) affinity value <50 nM or a rank percentage score <0.5%; the threshold for weak binding peptides is defined as IC50 <500 nM or a rank percentage score <2%. HLA-A, HLA-B, HLA-C are included in MHC class I molecules, and HLA-DR, HLA-DP and HLA-DQ are included in MHC Class II molecules.

Statistical Analysis

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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 more than two levels) were applied. The FDR method was used for multiple testing adjustment of p-values(79). 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.

COMPETING 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

ACKNOWLEDGMENTS

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).

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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 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 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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 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, 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

Supplementary Figure 1. The dynamic changes in differentially expressed genes from preneoplasia to invasive lung adenocarcinoma. Gene expression profiling was performed using the nCounter PanCancer Immune Profiling Panel. The 291 differentially expressed genes (DEG) were identified with one-way ANOVA (*FDR* < 0.1%). Vertical scale corresponds to log ratio between means of ADC and N samples. Red: down-regulated genes in later stages. Green: up-regulated in later stages. N: normal lung tissues. AAH: typical adenomatous hyperplasia. AIS: adenocarcinoma in situ. MIA: minimally invasive adenocarcinoma. ADC: invasive adenocarcinoma.

Supplementary Figure 2. Example genes differentially expressed from preneoplasia to invasive lung adenocarcinoma. *CD47* (**A**), *CD276* (**B**), *CTLA4* (**C**), *ENTPD1* (**D**), *GZMB* (**E**), *PRF1* (**F**). 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.

Supplementary Figure 3. Expression of tertiary lymphoid structure (TLS) signature genes from preneoplasia to invasive lung adenocarcinoma. (**A**) Gene expression profiling was performed using the nCounter PanCancer Immune Profiling Panel (Nanostring). Principal component analysis was performed to 12 TLS signature chemokines (*CCL2*, *CCL3*, *CCL4*, *CCL5*, *CCL8*, *CCL18*, *CCL19*, *CCL21*, *CXCL9*, *CXCL10*, *CXCL11*, and *CXCL13*) and the PC1/PC2, representing the most variability across all samples, was used to derive the 12-chemokine TLS signature score. The difference of 12-chemokine TLS signature score among different stages was evaluated using the Kruskal-Wallis H test. (**B**) The correlation between 12 chemokine TLS signature

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

score and B cell infiltration. B cell infiltration was derived from gene expression profiling using TIMER. The correction coefficient (rho) was assessed by Spearman's rank correlation test.

Supplementary Figure 4. 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.

Supplementary Figure 5. Expression of IL-10 from AAH to ADC. Gene expression profiling was performed using the nCounter PanCancer Immune Profiling Panel (Nanostring). The difference of expression among different stages was evaluated using the Kruskal-Wallis H test.

Supplementary Figure 6. The dynamic changes of cytokines relevant to different helper T cells (Th) and regulatory T cells (Tregs). (**A**) The heatmap represent expression of cytokines relevant to Th1, Th2, Th17 and Tregs in different stages. (**B**) The average expression of marker cytokines relevant to Th1, Th2, Th17 and Tregs in each stage. NL: Normal lung tissue, AAH: Atypical adenomatous hyperplasia, AIS: Adenocarcinoma in situ, MIA: Minimally invasive adenocarcinoma, ADC: Invasive adenocarcinoma.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

Supplementary Figure 7. 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).

Supplementary Figure 8. 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 (CD3+CD45RO+), memory CTL (CD3+CD8+CD45RO+), regulatory T cell (CD3+CD8-FoxP3+), memory/regulatory T cell (CD3+CD45RO+FoxP3+) in panel 2 (right).

Supplementary Figure 9. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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.

Supplementary Figure 10. 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.

Supplementary Figure 11. 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.

Supplementary Figure 12. 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.

Supplementary Figure 13. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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.

Supplementary Figure 14. 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.

Supplementary Figure 15. 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 (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.

Supplementary Figure 16. 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

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

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.

Supplementary Figure 17. 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.

Supplementary Figure 18. 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 (purple) and without (green) HLA-LOH. Willcoxon rank-sum test was used to assess the differences.

Supplementary Figure 19. Impact of promoter methylation on predicted neoantigen burden from preneoplasia to invasive lung adenocarcinoma. The predicted MHC Class I neoantigen (IC50 < 500nM) burden (**A**) and predicted MHC Class II neoantigen (IC50 < 500nM) burden (**B**) in in AAH (blue), AIS (red), MIA (green) and invasive ADC (purple). The total number of mutations associated with predicted MHC Class I

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

neoantigens (IC50 < 500nM) (**C**) and MHC Class II neoantigens (IC50 < 500nM) (**D**) in each histologic stage are shown as purple boxes. The total number of mutations associated with predicted MHC Class I neoantigens (IC50 < 500nM) (**C**) and MHC Class II neoantigens (IC50 < 500nM) (**D**) from genes without promoter methylation (<30% CpG sites methylated) are shown as blue boxes.

Supplementary Figure 20. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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

1. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.
2. Tan WL, Jain A, Takano A, Newell EW, Iyer NG, Lim WT, et al. Novel therapeutic targets on the horizon for lung cancer. Lancet Oncol. 2016;17(8):e347-e62.
3. Barnett R. Lung cancer. Lancet. 2017;390(10098):928.
4. Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in advanced nonsquamous non-small-cell lung cancer. N Engl J Med. 2015;373(17):1627-39.
5. de Koning HJ, van der Aalst CM, de Jong PA, Scholten ET, Nackaerts K, Heuvelmans MA, et al. Reduced lung-cancer mortality with volume CT screening in a randomized trial. N Engl J Med. 2020;382(6):503-13.
6. Chiosea S, Jelezcova E, Chandran U, Luo J, Mantha G, Sobol RW, et al. Overexpression of Dicer in precursor lesions of lung adenocarcinoma. Cancer Res. 2007;67(5):2345-50.
7. Kitamura H, Kameda Y, Ito T, Hayashi H. Atypical adenomatous hyperplasia of the lung: implications for the pathogenesis of peripheral lung adenocarcinoma. Am J Clin Pathol. 1999;111(5):610-22.
8. Aoyagi Y, Yokose T, Minami Y, Ochiai A, Iijima T, Morishita Y, et al. Accumulation of losses of heterozygosity and multistep carcinogenesis in pulmonary adenocarcinoma. Cancer Res. 2001;61(21):7950-4.
9. Noguchi M. Stepwise progression of pulmonary adenocarcinoma—clinical and molecular implications. Cancer Metastasis Rev. 2010;29(1):15-21.
10. Seki M, Akasaka Y. Multiple lung adenocarcinomas and AAH treated by surgical resection. Lung Cancer. 2007;55(2):237-40.
11. Maeshima AM, Tochigi N, Yoshida A, Asamura H, Tsuta K, Tsuda H. Clinicopathologic analysis of multiple (five or more) atypical adenomatous hyperplasias (AAHs) of the lung: evidence for the AAH-adenocarcinoma sequence. J Thorac Oncol. 2010;5(4):466-71.
12. Min JH, Lee HY, Lee KS, Han J, Park K, Ahn MJ, et al. 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. 2010;69(1):123-6.
13. Weichert W, Warth A. Early lung cancer with lepidic pattern: adenocarcinoma in situ, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma. Curr Opin Pulm Med. 2014;20(4):309-16.
14. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol. 2011;6(2):244-85.
15. Izumchenko E, Chang X, Brait M, Fertig E, Kagohara LT, Bedi A, et al. Targeted sequencing reveals clonal genetic changes in the progression of early lung neoplasms and paired circulating DNA. Nat Commun. 2015;6:8258.
16. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-58.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

17. Schreiber RD, Old LJ, Smyth MJ. Cancer immunoediting: integrating immunity's roles in cancer suppression and promotion. Science. 2011;331(6024):1565-70.
18. Sivakumar S, Lucas FAS, McDowell TL, Lang W, Xu L, Fujimoto J, et al. Genomic landscape of atypical adenomatous hyperplasia reveals divergent modes to lung adenocarcinoma. Cancer Res. 2017;77(22):6119-30.
19. Hu X, Fujimoto J, Ying L, Fukuoka J, Ashizawa K, Sun W, et al. Multi-region exome sequencing reveals genomic evolution from preneoplasia to lung adenocarcinoma. Nat Commun. 2019;10(1):2978.
20. Chen PL, Roh W, Reuben A, Cooper ZA, Spencer CN, Prieto PA, et al. Analysis of immune signatures in longitudinal tumor samples yields insight into biomarkers of response and mechanisms of resistance to immune checkpoint blockade. Cancer Discov. 2016;6(8):827-37.
21. Soto-Pantoja DR, Terabe M, Ghosh A, Ridnour LA, DeGraff WG, Wink DA, et al. CD47 in the tumor microenvironment limits cooperation between antitumor T-cell immunity and radiotherapy. Cancer Res. 2014;74(23):6771-83.
22. Picarda E, Ohaegbulam KC, Zang X. Molecular pathways: targeting B7-H3 (CD276) for human cancer immunotherapy. Clin Cancer Res. 2016;22(14):3425-31.
23. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. Nat Rev Cancer. 2012;12(4):252-64.
24. Bastid J, Cottalorda-Regairaz A, Alberici G, Bonnefoy N, Eliaou J, Bensussan A. ENTPD1/CD39 is a promising therapeutic target in oncology. Oncogene. 2013;32(14):1743-51.
25. Tschopp J, Masson D, Stanley KK. Structural/functional similarity between proteins involved in complement- and cytotoxic T-lymphocyte-mediated cytolysis. Nature. 1986;322(6082):831-4.
26. Prakash MD, Munoz MA, Jain R, Tong PL, Koskinen A, Regner M, et al. Granzyme B promotes cytotoxic lymphocyte transmigration via basement membrane remodeling. Immunity. 2014;41(6):960-72.
27. Pantelidou C, Sonzogni O, De Oliveria Taveira M, Mehta AK, Kothari A, Wang D, et al. 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. 2019;9(6):722-37.
28. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015;15(8):486-99.
29. Pan L, Lu MP, Wang JH, Xu M, Yang SR. Immunological pathogenesis and treatment of systemic lupus erythematosus. World J Pediatr. 2020;16(1):19-30.
30. Li T, Fu J, Zeng Z, Cohen D, Li J, Chen Q, et al. TIMER2.0 for analysis of tumor-infiltrating immune cells. Nucleic Acids Res. 2020;49(W1):W509-14.
31. Knochelmann HM, Dwyer CJ, Bailey SR, Amaya SM, Elston DM, Mazza-McCrann JM, et al. When worlds collide: Th17 and Treg cells in cancer and autoimmunity. Cell Mol Immunol. 2018;15(5):458-69.
32. Lança T, Silva-Santos B. The split nature of tumor-infiltrating leukocytes: Implications for cancer surveillance and immunotherapy. Oncoimmunology. 2012;1(5):717-25.
33. Shah W, Yan X, Jing L, Zhou Y, Chen H, Wang Y. A reversed CD4/CD8 ratio of tumor-infiltrating lymphocytes and a high percentage of CD4+ FOXP3+ regulatory T cells

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

are significantly associated with clinical outcome in squamous cell carcinoma of the cervix. Cell Mol Immunol. 2011;8(1):59-66.

34. Kaplinsky J, Arnaout R. Robust estimates of overall immune-repertoire diversity from high-throughput measurements on samples. Nat Commun. 2016;7:11881.

35. Wang X, Zhang B, Yang Y, Zhu J, Cheng S, Mao Y, et al. Characterization of distinct T cell receptor repertoires in tumor and distant non-tumor tissues from lung cancer patients. Genomics Proteomics Bioinformatics. 2019;17(3):287-96.

36. Mansfield AS, Ren H, Sutor S, Sarangi V, Nair A, Davila J, et al. Contraction of T cell richness in lung cancer brain metastases. Sci Rep. 2018;8(1):2171.

37. Reuben A, Zhang J, Chiou SH, Gittelman RM, Li J, Lee WC, et al. Comprehensive T cell repertoire characterization of non-small cell lung cancer. Nat Commun. 2020;11(1):603.

38. McGranahan N, Rosenthal R, Hiley CT, Rowan AJ, Watkins TBK, Wilson GA, et al. Allele-specific HLA loss and immune escape in lung cancer evolution. Cell. 2017;171(6):1259-71.e11.

39. Davoli T, Uno H, Wooten EC, Elledge SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science. 2017;355(6322).

40. Yarchoan M, Johnson III BA, Lutz ER, Laheru DA, Jaffee EM. Targeting neoantigens to augment antitumour immunity. Nat Rev Cancer. 2017;17(4):209-22.

41. McGranahan N, Furness AJ, Rosenthal R, Ramskov S, Lyngaa R, Saini SK, et al. Clonal neoantigens elicit T cell immunoreactivity and sensitivity to immune checkpoint blockade. Science. 2016;351(6280):1463-9.

42. Ohka F, Natsume A, Motomura K, Kishida Y, Kondo Y, Abe T, et al. The global DNA methylation surrogate LINE-1 methylation is correlated with MGMT promoter methylation and is a better prognostic factor for glioma. PLoS One. 2011;6(8):e23332.

43. Kankava K, Kvaratskhelia E, Burkadze G, Kokhreidze I, Gogokhia N, Abzianidze E. Line-1 methylation in blood and tissues of patients with breast cancer. Georgian Med News. 2018(276):107-12.

44. Saito K, Kawakami K, Matsumoto I, Oda M, Watanabe G, Minamoto T. Long interspersed nuclear element 1 hypomethylation is a marker of poor prognosis in stage IA non-small cell lung cancer. Clin Cancer Res. 2010;16(8):2418-26.

45. Dunn GP, Bruce AT, Ikeda H, Old LJ, Schreiber RD. Cancer immunoediting: from immunosurveillance to tumor escape. Nat Immunol. 2002;3(11):991-8.

46. Dunn GP, Old LJ, Schreiber RD. The immunobiology of cancer immunosurveillance and immunoediting. Immunity. 2004;21(2):137-48.

47. Dunn GP, Old LJ, Schreiber RD. The three Es of cancer immunoediting. Annu Rev Immunol. 2004;22:329-60.

48. O'Donnell JS, Teng MW, Smyth MJ. Cancer immunoediting and resistance to T cell-based immunotherapy. Nat Rev Clin Oncol. 2019;16(3):151-67.

49. Mittal D, Gubin MM, Schreiber RD, Smyth MJ. New insights into cancer immunoediting and its three component phases—elimination, equilibrium and escape. Curr Opin Immunol. 2014;27:16-25.

50. Lavin Y, Kobayashi S, Leader A, Amir E-aD, Elefant N, Bigenwald C, et al. Innate immune landscape in early lung adenocarcinoma by paired single-cell analyses. Cell.169(4):750-65.e17.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

51. Mascaux C, Angelova M, Vasaturo A, Beane J, Hijazi K, Anthoine G, et al. Immune evasion before tumour invasion in early lung squamous carcinogenesis. Nature. 2019;571(7766):570-5.
52. Thorsson V, Gibbs DL, Brown SD, Wolf D, Bortone DS, Yang T-HO, et al. The immune landscape of cancer. Immunity. 2018;48(4):812-30. e14.
53. Fridman WH, Zitvogel L, Sautès–Fridman C, Kroemer G. The immune contexture in cancer prognosis and treatment. Nat Rev Clin Oncol. 2017;14(12):717.
54. Liu M, Zhou J, Chen Z, Cheng AS. Understanding the epigenetic regulation of tumours and their microenvironments: opportunities and problems for epigenetic therapy. J Pathol. 2017;241(1):10-24.
55. Serrano A, Castro-Vega I, Redondo M. Role of gene methylation in antitumor immune response: implication for tumor progression. Cancers (Basel). 2011;3(2):1672-90.
56. Bakhoum SF, Cantley LC. The multifaceted role of chromosomal instability in cancer and its microenvironment. Cell. 2018;174(6):1347-60.
57. Porta-Pardo E, Godzik A. Mutation drivers of immunological responses to cancer. Cancer Immunol Res. 2016;4(9):789-98.
58. Eden A, Gaudet F, Waghmare A, Jaenisch R. Chromosomal instability and tumors promoted by DNA hypomethylation. Science. 2003;300(5618):455.
59. Chen RZ, Pettersson U, Beard C, Jackson-Grusby L, Jaenisch R. DNA hypomethylation leads to elevated mutation rates. Nature. 1998;395(6697):89-93.
60. Jung H, Kim HS, Kim JY, Sun JM, Ahn JS, Ahn MJ, et al. DNA methylation loss promotes immune evasion of tumours with high mutation and copy number load. Nat Commun. 2019;10(1):4278.
61. Rosenthal R, Cadieux EL, Salgado R, Al Bakir M, Moore DA, Hiley CT, et al. Neoantigen-directed immune escape in lung cancer evolution. Nature. 2019;567(7749):479-85.
62. Neal JT, Li X, Zhu J, Giangarra V, Grzeskowiak CL, Ju J, et al. Organoid modeling of the tumor immune microenvironment. Cell. 2018;175(7):1972-88. e16.
63. Paulson K, Voillet V, McAfee M, Hunter D, Wagener F, Perdicchio M, et al. Acquired cancer resistance to combination immunotherapy from transcriptional loss of class I HLA. Nature communications. 2018;9(1):1-10.
64. Jeschke J, Bizet M, Desmedt C, Calonne E, Dedeurwaerder S, Garaud S, et al. DNA methylation–based immune response signature improves patient diagnosis in multiple cancers. J Clin Invest. 2017;127(8):3090-102.
65. Bellomi M, Veronesi G, Trifirò G, Brambilla S, Bonello L, Preda L, et al. Computed tomography-guided preoperative radiotracer localization of nonpalpable lung nodules. Ann Thorac Surg. 2010;90(6):1759-64.
66. McWilliams A, Tammemagi MC, Mayo JR, Roberts H, Liu G, Soghrati K, et al. Probability of cancer in pulmonary nodules detected on first screening CT. N Engl J Med. 2013;369(10):910-9.
67. Schroeder A, Mueller O, Stocker S, Salowsky R, Leiber M, Gassmann M, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. BMC Mol Biol. 2006;7:3.
68. Cesano A. nCounter(®) PanCancer immune profiling panel (NanoString Technologies, Inc., Seattle, WA). J Immunother Cancer. 2015;3:42.

medRxiv preprint doi: https://doi.org/10.1101/2020.07.11.20142992; this version posted July 14, 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.

69. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString nCounter system. Curr Protoc Mol Biol. 2011;Chapter 25:Unit25B.10.

70. Pounds S, Morris SW. Estimating the occurrence of false positives and false negatives in microarray studies by approximating and partitioning the empirical distribution of p-values. Bioinformatics. 2003;19(10):1236-42.

71. Kramer A, Green J, Pollard J, Jr., Tugendreich S. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics. 2014;30(4):523-30.

72. Parra ER, Villalobos P, Behrens C, Jiang M, Pataer A, Swisher SG, et al. 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. 2018;6(1):48.

73. Francisco-Cruz A, Parra ER, Tetzlaff MT, Wistuba, II. Multiplex Immunofluorescence Assays. Methods Mol Biol. 2020;2055:467-95.

74. Reuben A, Gittelman R, Gao J, Zhang J, Yusko EC, Wu CJ, et al. TCR repertoire intratumor heterogeneity in localized lung adenocarcinomas: an association with predicted neoantigen heterogeneity and postsurgical recurrence. Cancer Discov. 2017;7(10):1088-97.

75. Shukla SA, Rooney MS, Rajasagi M, Tiao G, Dixon PM, Lawrence MS, et al. Comprehensive analysis of cancer-associated somatic mutations in class I HLA genes. Nat Biotechnol. 2015;33(11):1152-8.

76. Van Loo P, Nordgard SH, Lingjaerde OC, Russnes HG, Rye IH, Sun W, et al. Allele-specific copy number analysis of tumors. Proc Natl Acad Sci U S A. 2010;107(39):16910-5.

77. Nielsen M, Andreatta M. NetMHCpan-3.0; improved prediction of binding to MHC class I molecules integrating information from multiple receptor and peptide length datasets. Genome Med. 2016;8(1):33.

78. Andreatta M, Karosiene E, Rasmussen M, Stryhn A, Buus S, Nielsen M. Accurate pan-specific prediction of peptide-MHC class II binding affinity with improved binding core identification. Immunogenetics. 2015;67(11-12):641-50.

79. Benjamini Y, Hochberg Y. Controlling the false discovery rate - a practical and powerful approach to multiple testing. J R Stat Soc B. 1995;57(1):289-300.

Figure 1

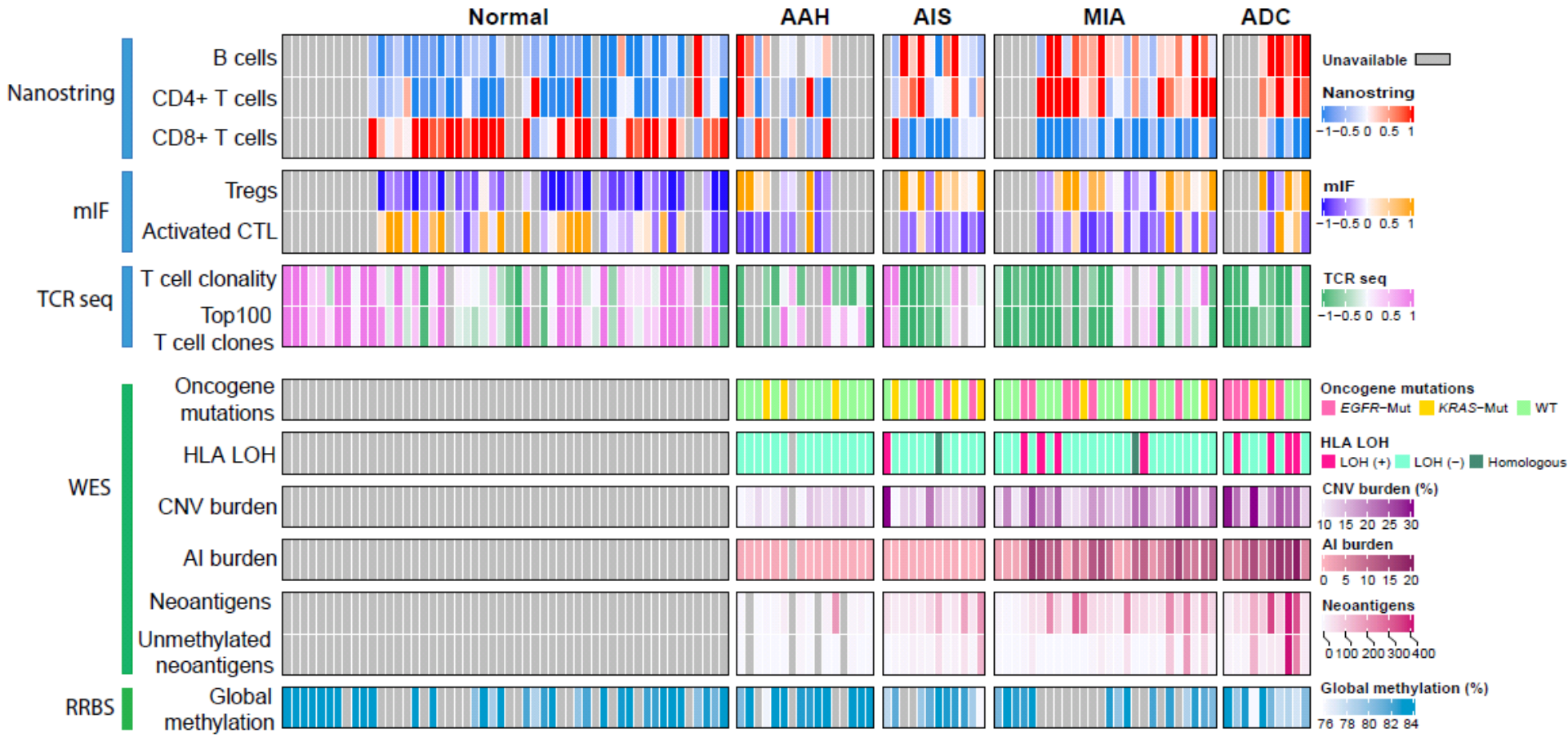

Normal
AAH
AIS
MIA
ADC
Nanostring
B cells
CD4+ T cells
CD8+ T cells
mIF
Tregs
Activated CTL
TCR seq
T cell clonality
Top100
T cell clones
WES
Oncogene mutations
HLA LOH
CNV burden
AI burden
Neoantigens
Unmethylated neoantigens
RRBS
Global methylation
Unavailable
Nanostring
−1 −0.5 0 0.5 1
mIF
−1 −0.5 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

Figure 2A

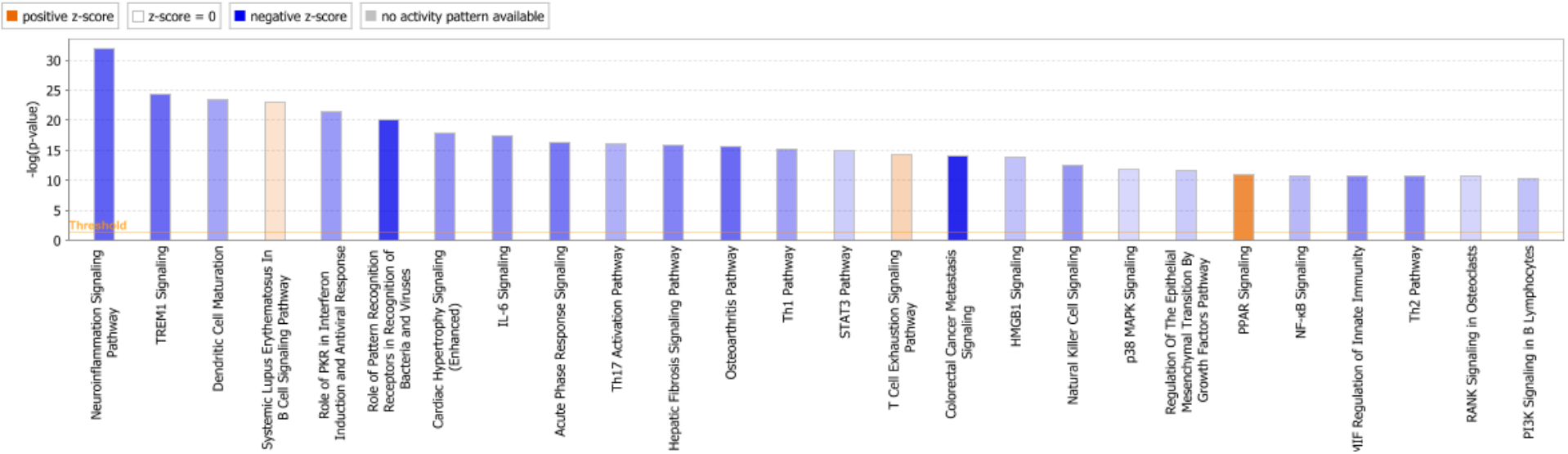

positive z-score
z-score = 0
negative z-score
no activity pattern available
-log(p-value)
30
25
20
15
10
5
0
Threshold
Neuroinflammation Signaling Pathway
TREM1 Signaling
Dendritic Cell Maturation
Systemic Lupus Erythematosus In B Cell Signaling Pathway
Role of PKR in Interferon Induction and Antiviral 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

Figure 2

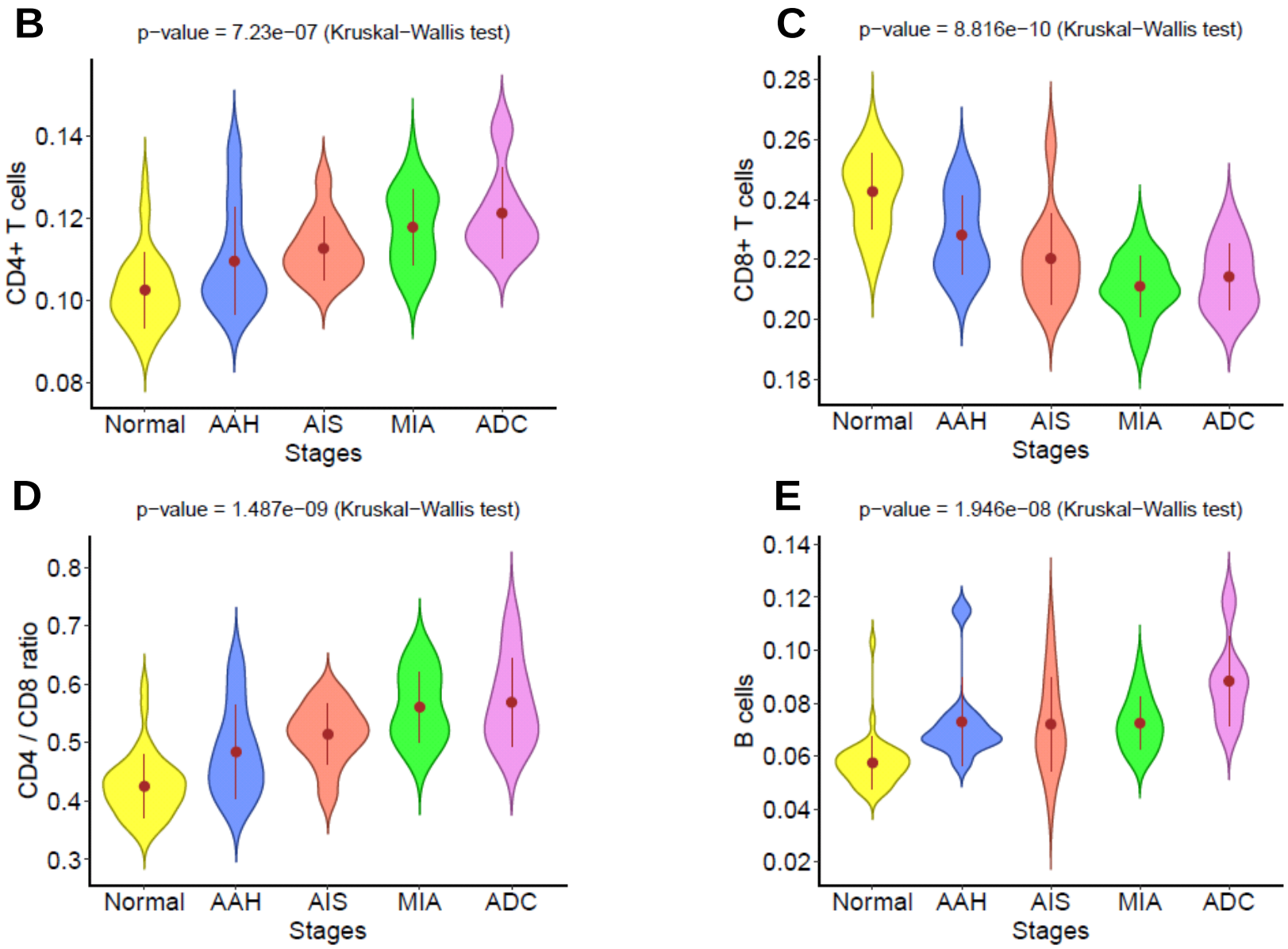
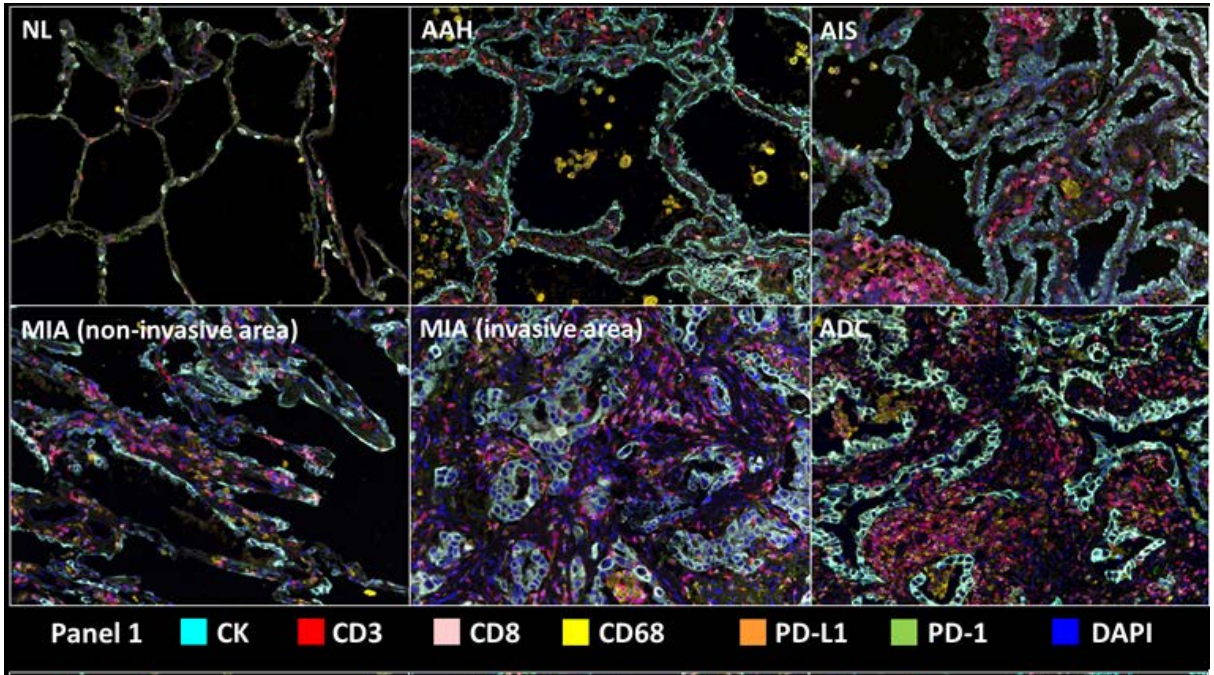
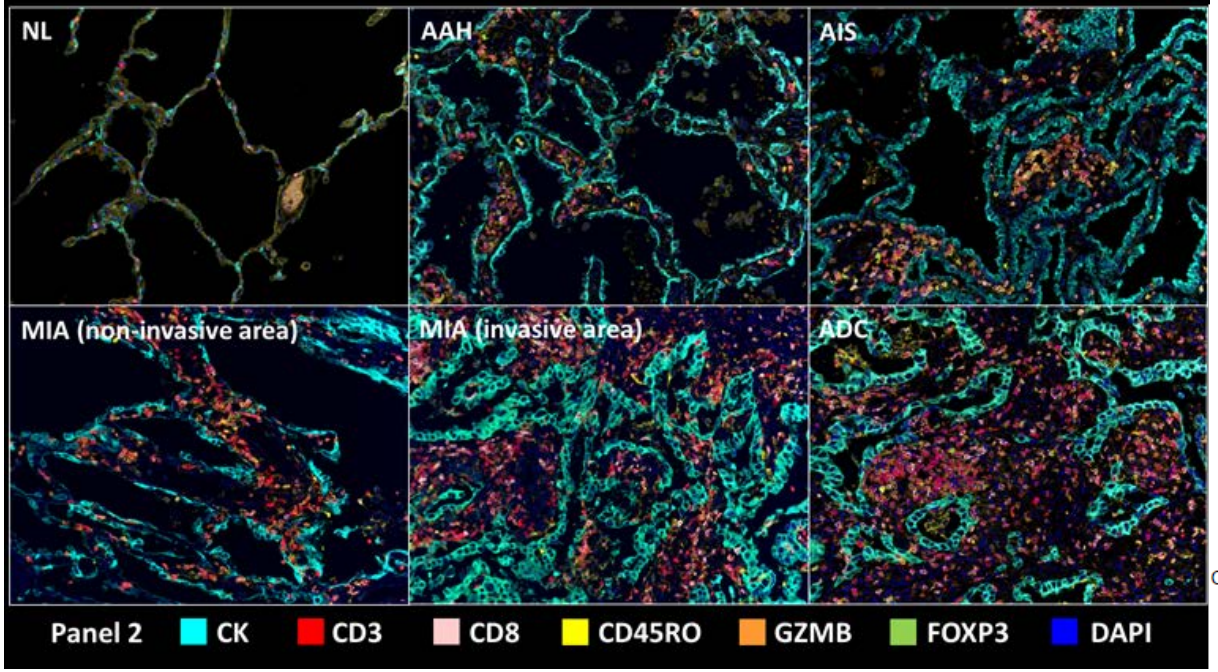
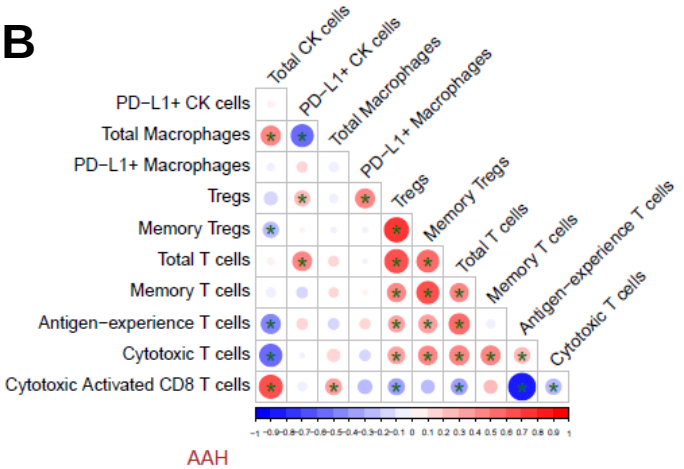
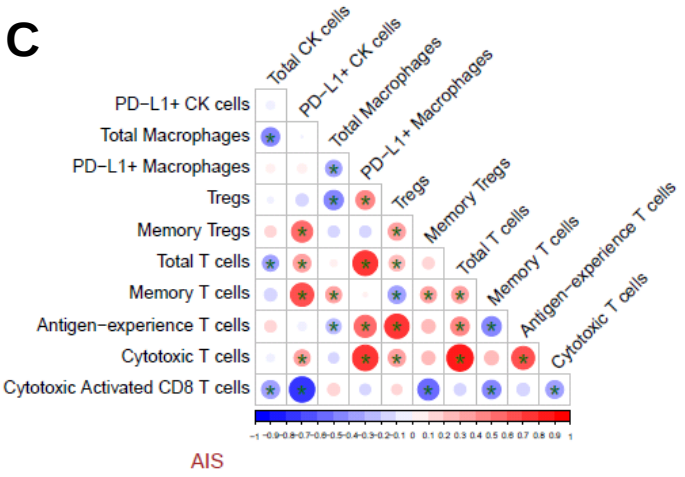

B
p−value = 7.23e−07 (Kruskal−Wallis test)
CD4+ T cells
0.14
0.12
0.10
0.08
Normal
AAH
AIS
MIA
ADC
Stages
C
p−value = 8.816e−10 (Kruskal−Wallis test)
CD8+ T cells
0.28
0.26
0.24
0.22
0.20
0.18
Normal
AAH
AIS
MIA
ADC
Stages
D
p−value = 1.487e−09 (Kruskal−Wallis test)
CD4 / CD8 ratio
0.8
0.7
0.6
0.5
0.4
0.3
Normal
AAH
AIS
MIA
ADC
Stages
E
p−value = 1.946e−08 (Kruskal−Wallis test)
B cells
0.14
0.12
0.10
0.08
0.06
0.04
0.02
Normal
AAH
AIS
MIA
ADC
Stages

Figure 3A


NL
AAH
AIS
MIA (non-invasive area)
MIA (invasive area)
ADC
Panel 1
CK
CD3
CD8
CD68
PD-L1
PD-1
DAPI


NL
AAH
AIS
MIA (non-invasive area)
MIA (invasive area)
ADC
Panel 2
CK
CD3
CD8
CD45RO
GZMB
FOXP3
DAPI


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
AAH


C
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
AIS

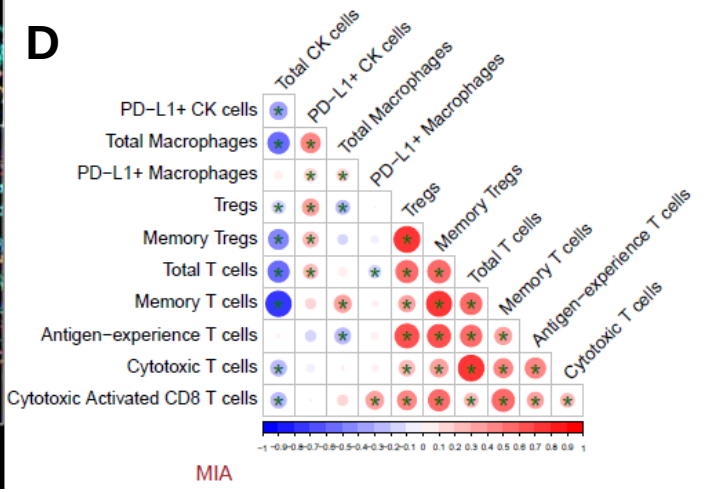
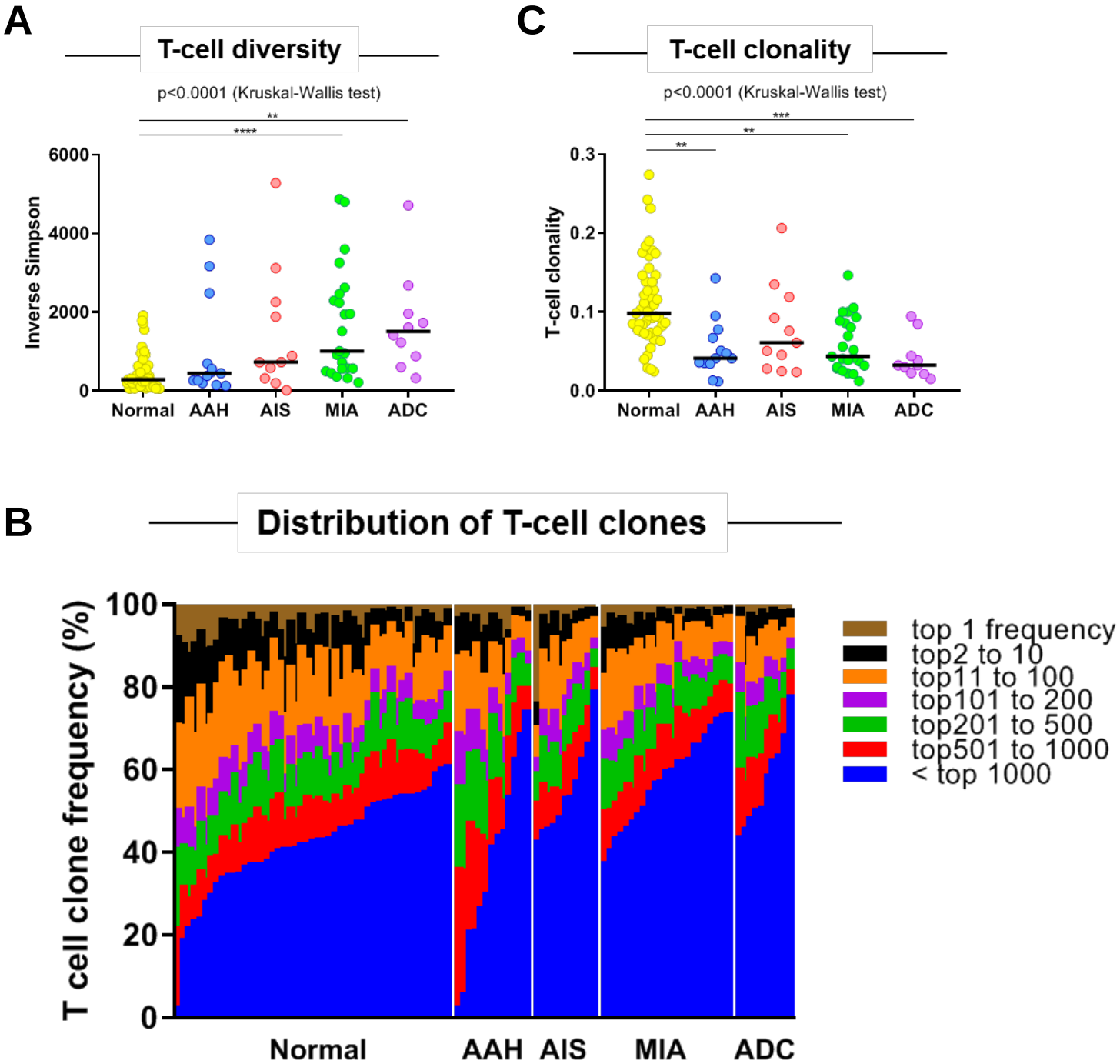
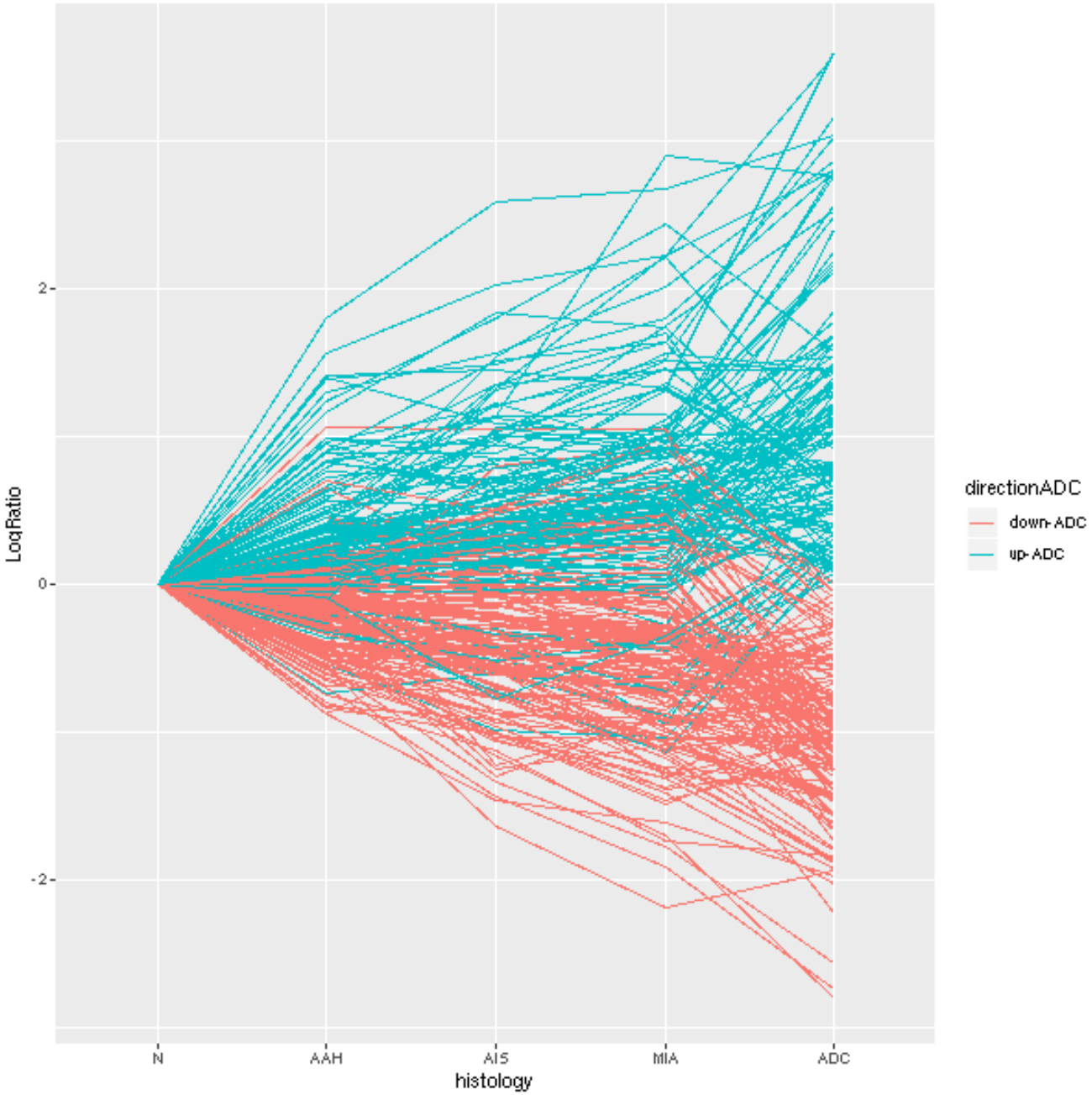

D
Total CK cells
PD-L1+ CK cells
Total Macrophages
PD-L1+ Macrophages
Tregs
Memory Tregs
Total T cells
Memory T cells
Antigen-experience T cells
Cytotoxic T cells
Cytotoxic Activated CD8 T cells
MIA

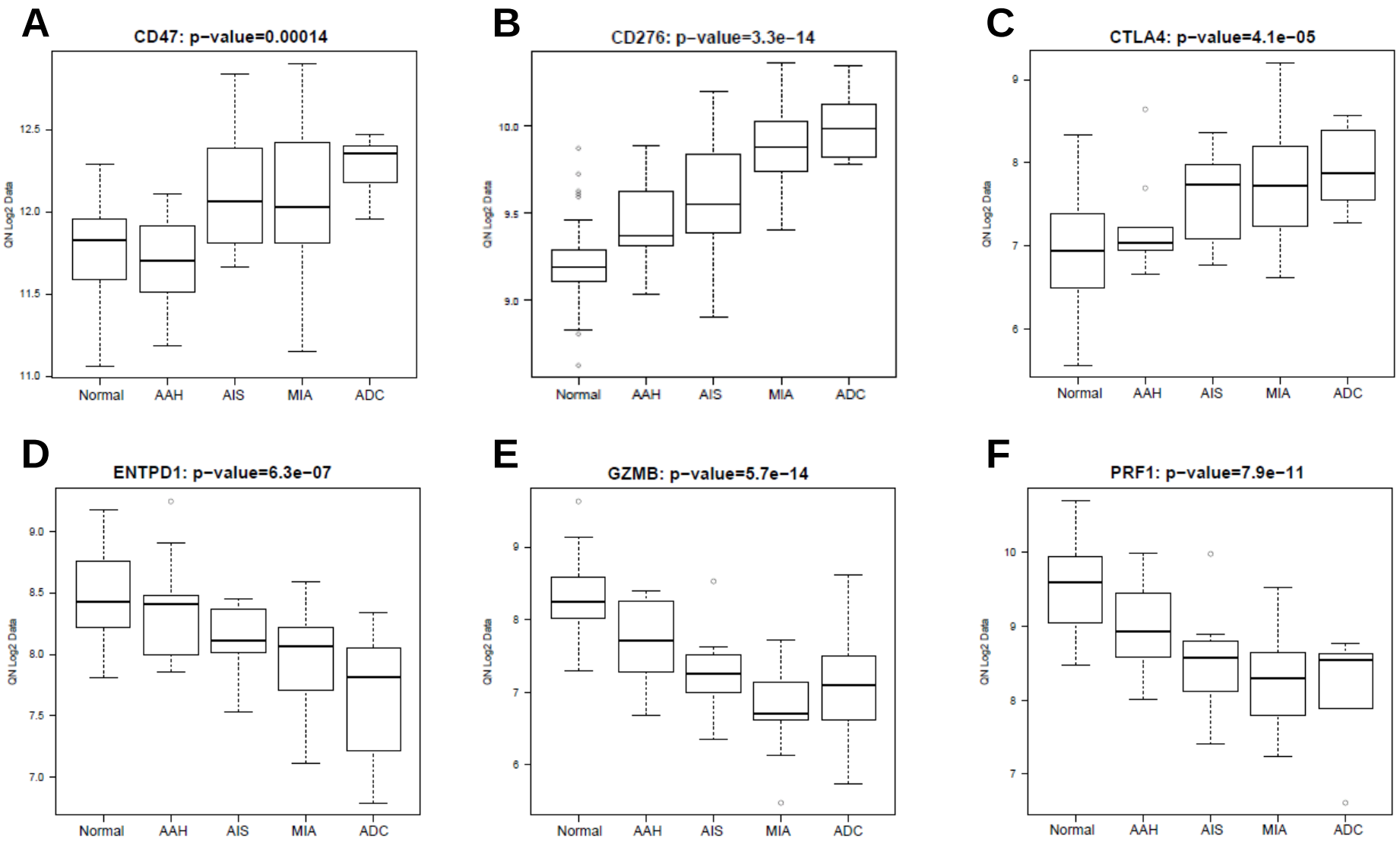
Figure 4


A
T-cell diversity
p<0.0001 (Kruskal-Wallis test)
**
****
Inverse Simpson
6000
4000
2000
0
Normal
AAH
AIS
MIA
ADC
C
T-cell clonality
p<0.0001 (Kruskal-Wallis test)
***
**
**
T-cell clonality
0.3
0.2
0.1
0.0
Normal
AAH
AIS
MIA
ADC
B
Distribution of T-cell clones
T cell clone frequency (%)
100
80
60
40
20
0
Normal
AAH
AIS
MIA
ADC
top 1 frequency
top2 to 10
top11 to 100
top101 to 200
top201 to 500
top501 to 1000
< top 1000

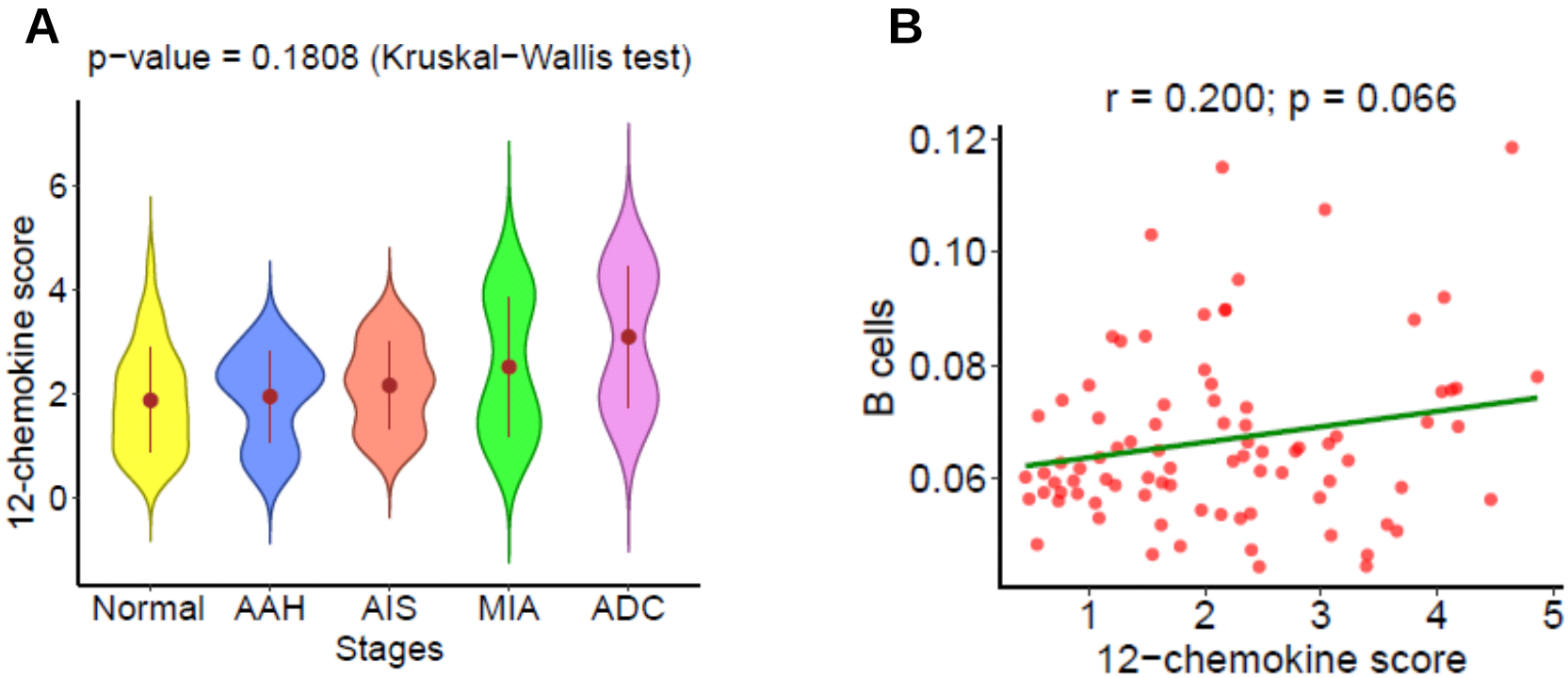
Supplementary Figure 1



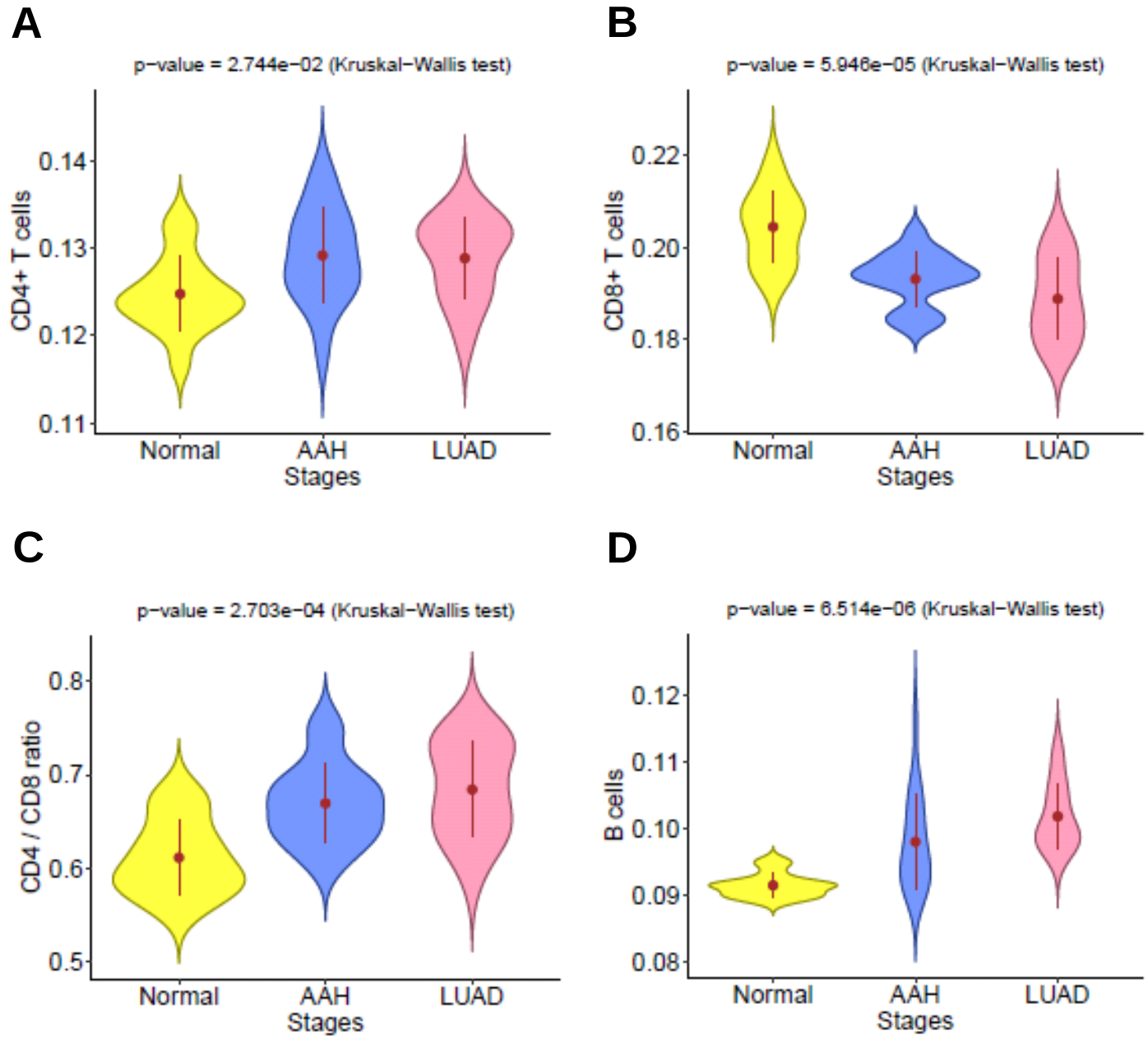
Supplementary Figure 2



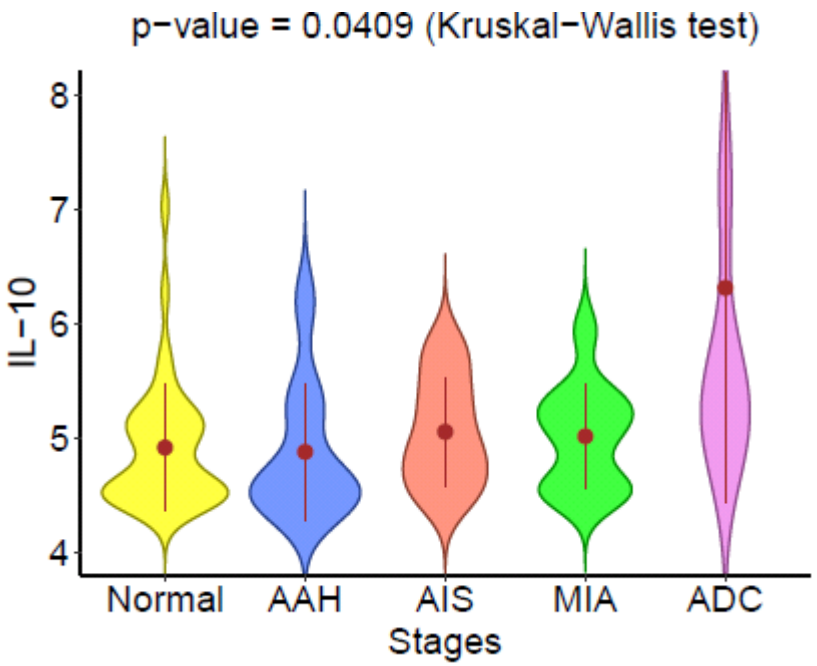
Supplementary Figure 3



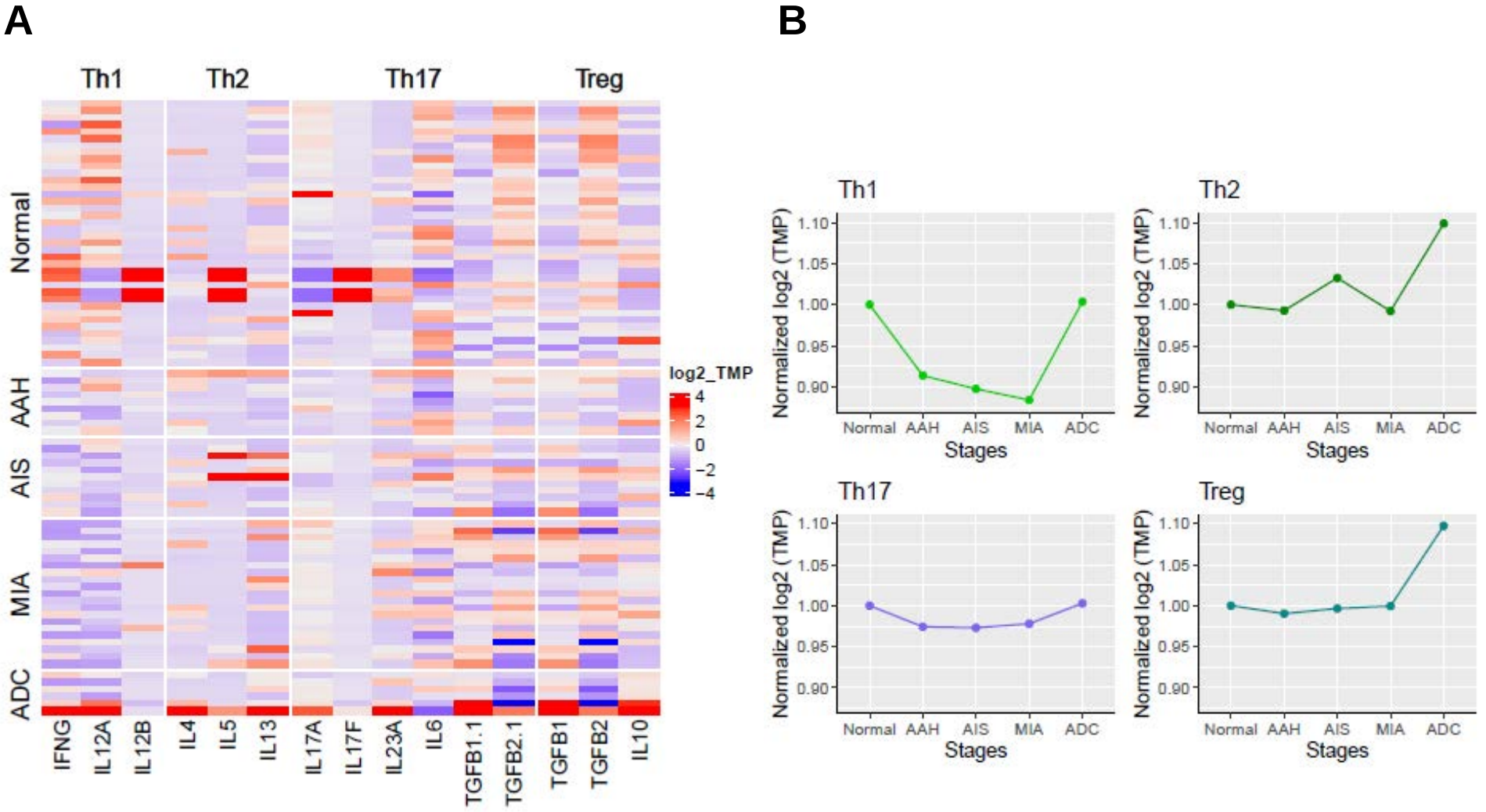
Supplementary Figure 4



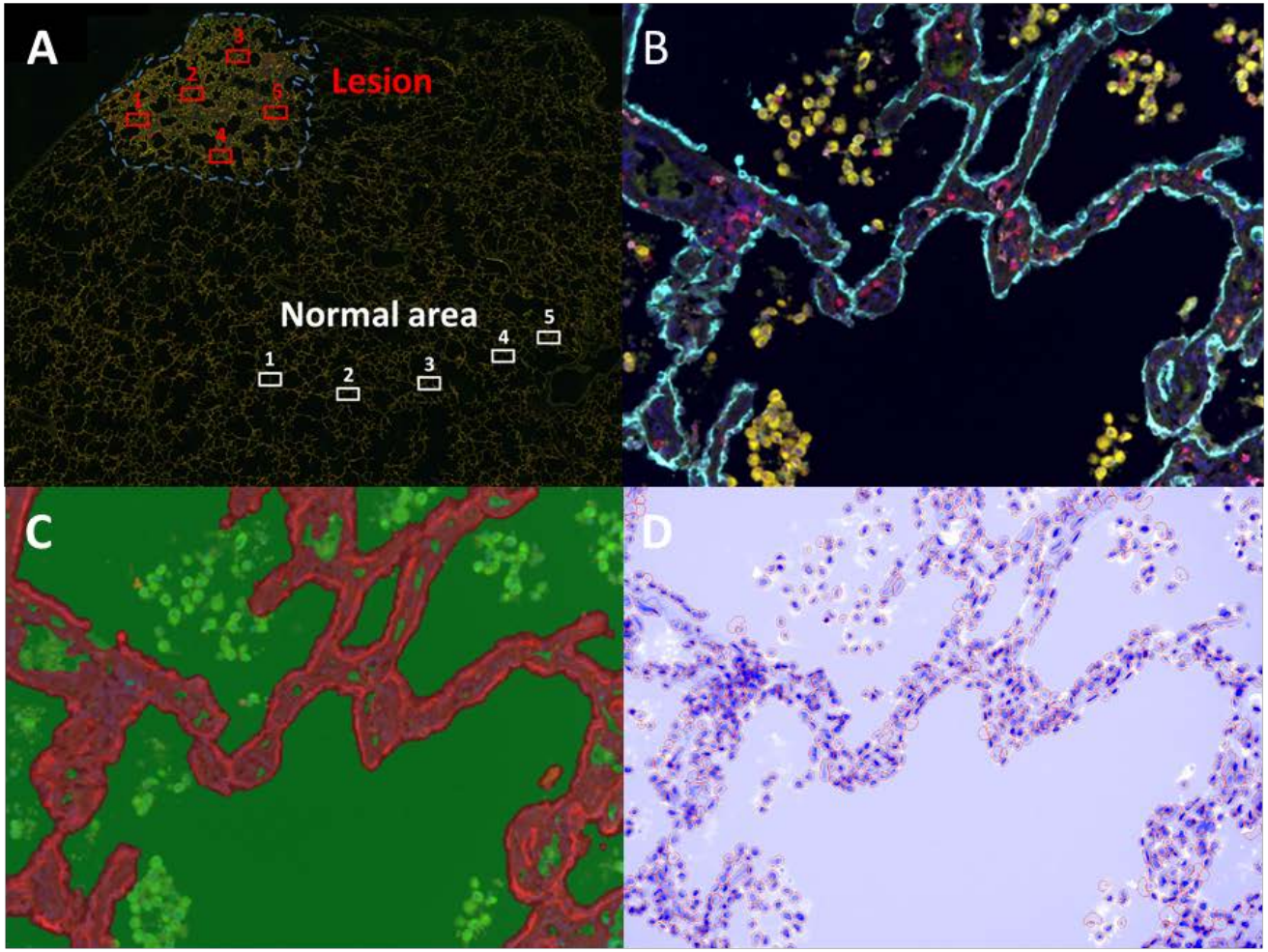
Supplementary Figure 5



Supplementary Figure 6



Supplementary Figure 7



A. Area selection

B. Composite image

CK CD3 CD8

CD68 PD-1 PD-L1

DAPI

C. Tissue Segmentation

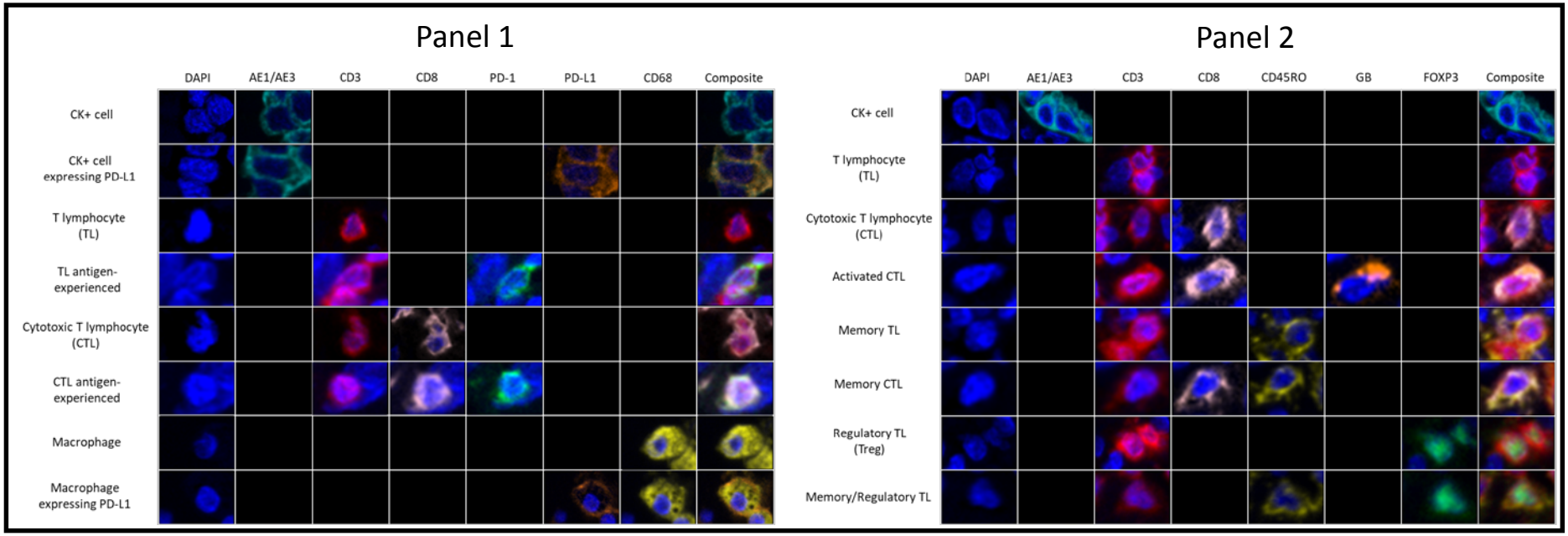
Epithelium

Alveolar space

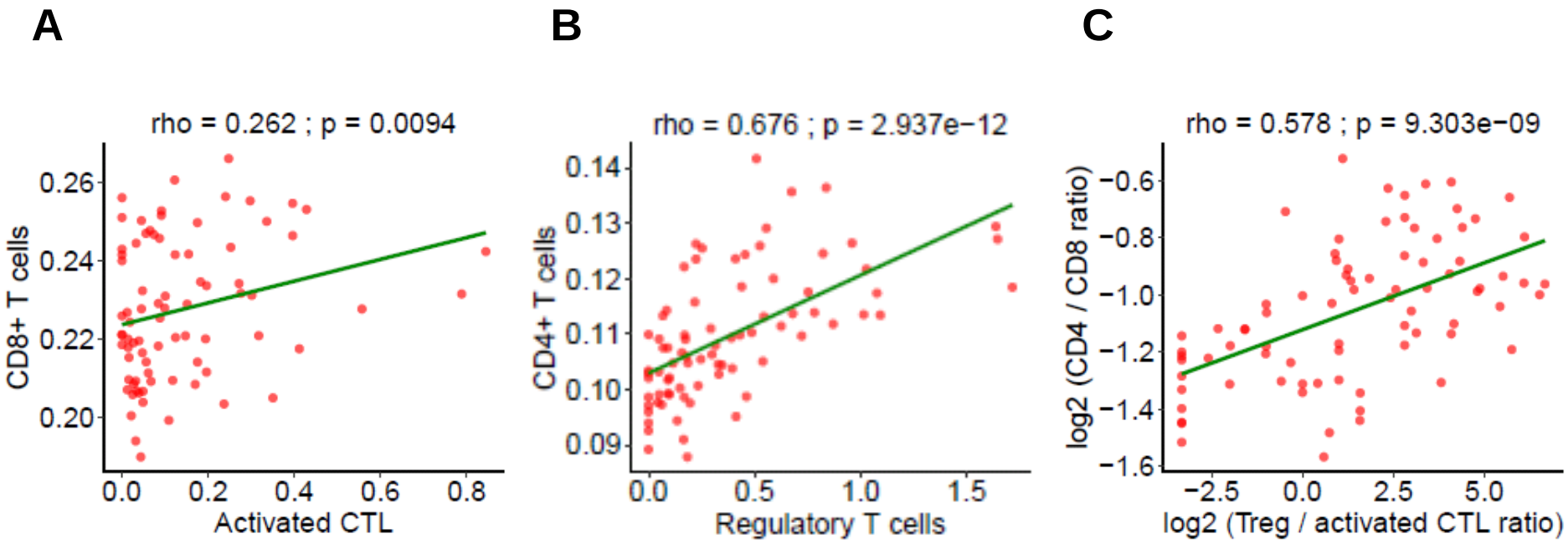
D. Cell Segmentation

DAPI

Supplementary Figure 8

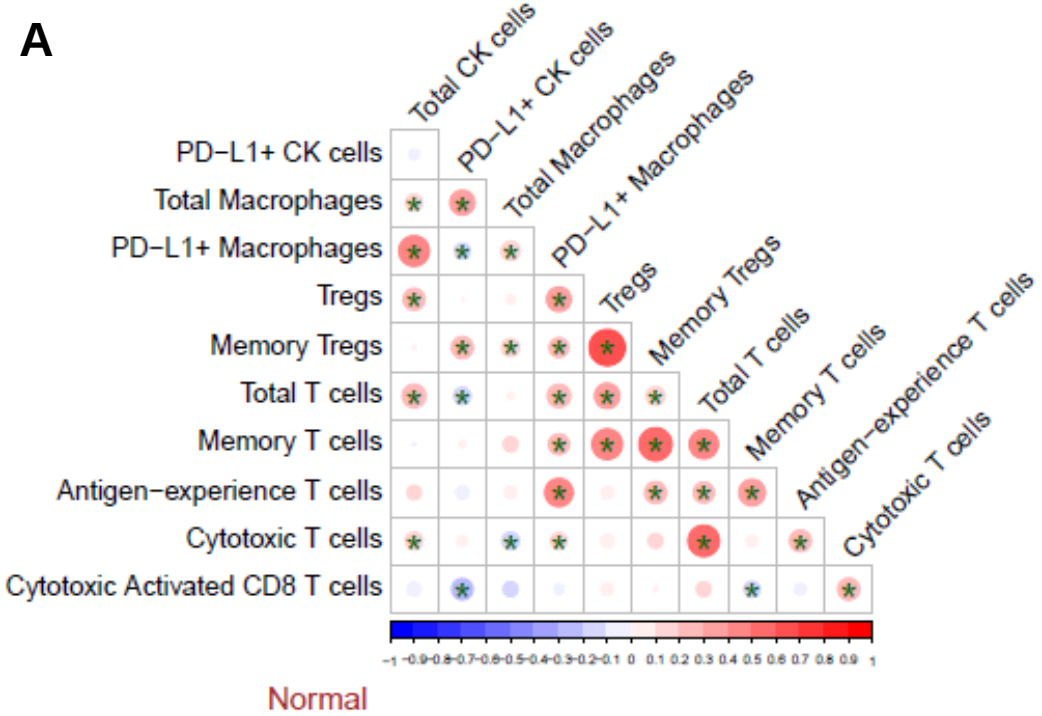


Supplementary Figure 9

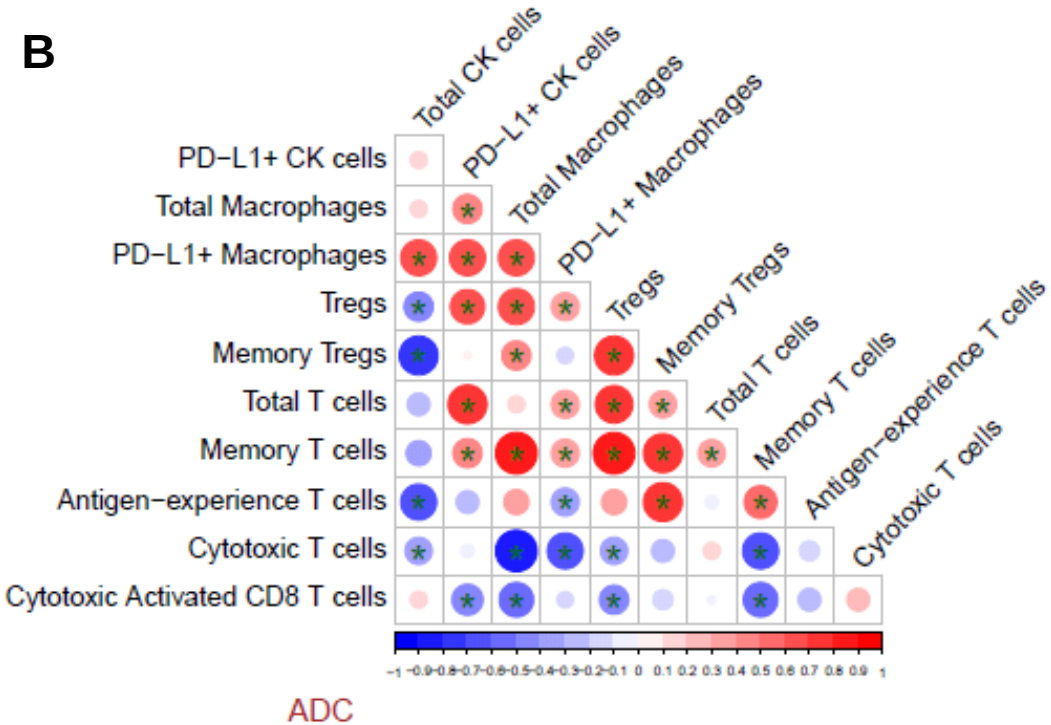


Supplementary Figure 10

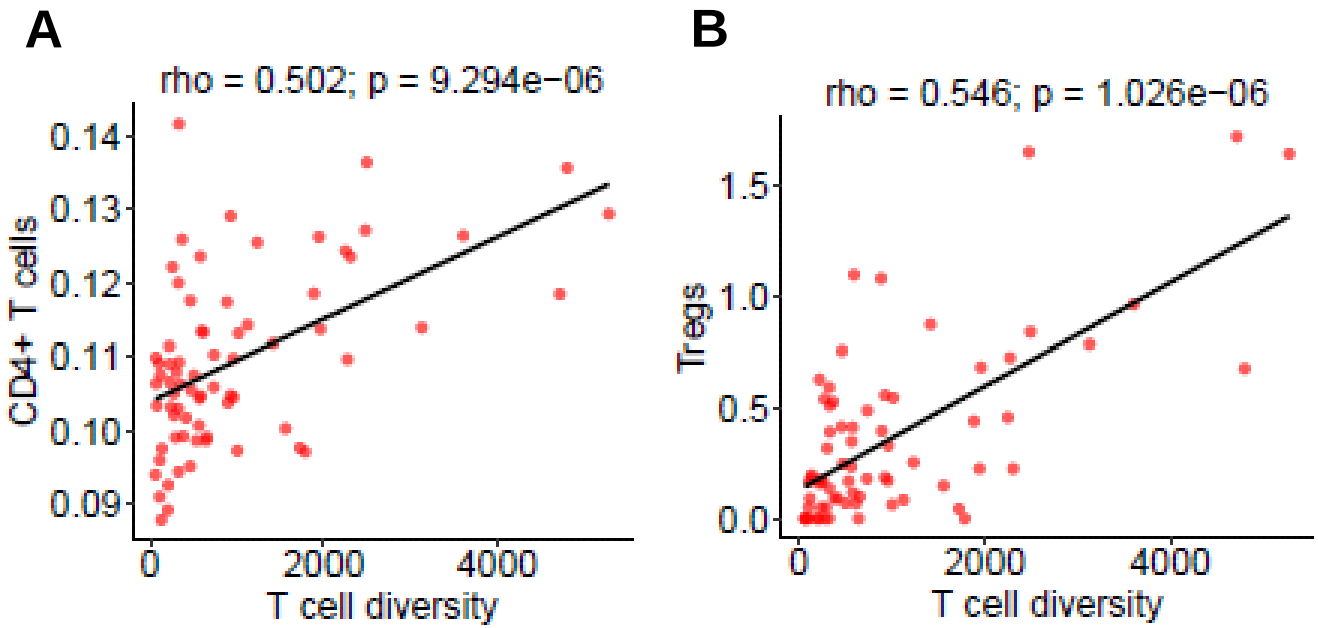
A



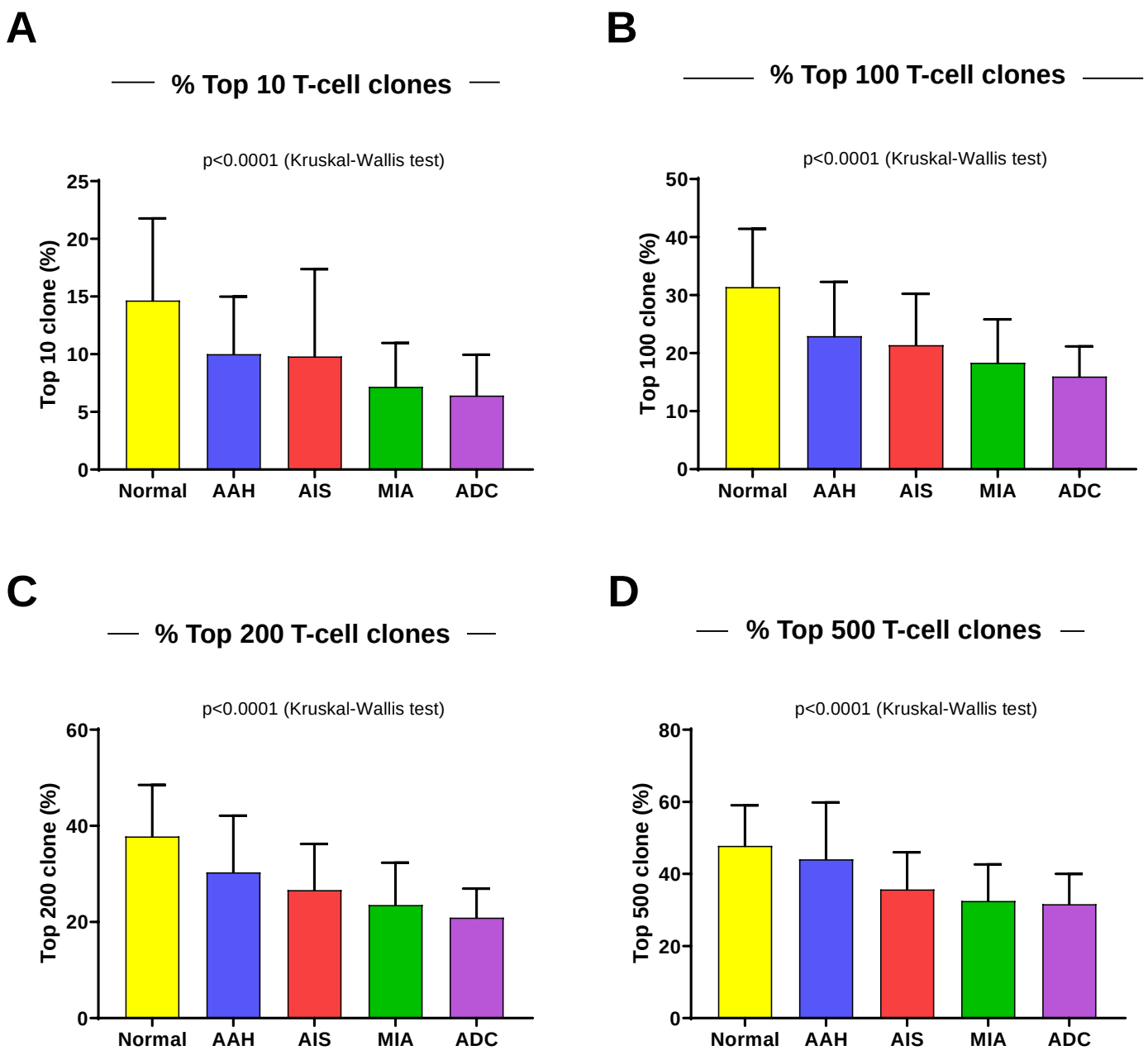
B



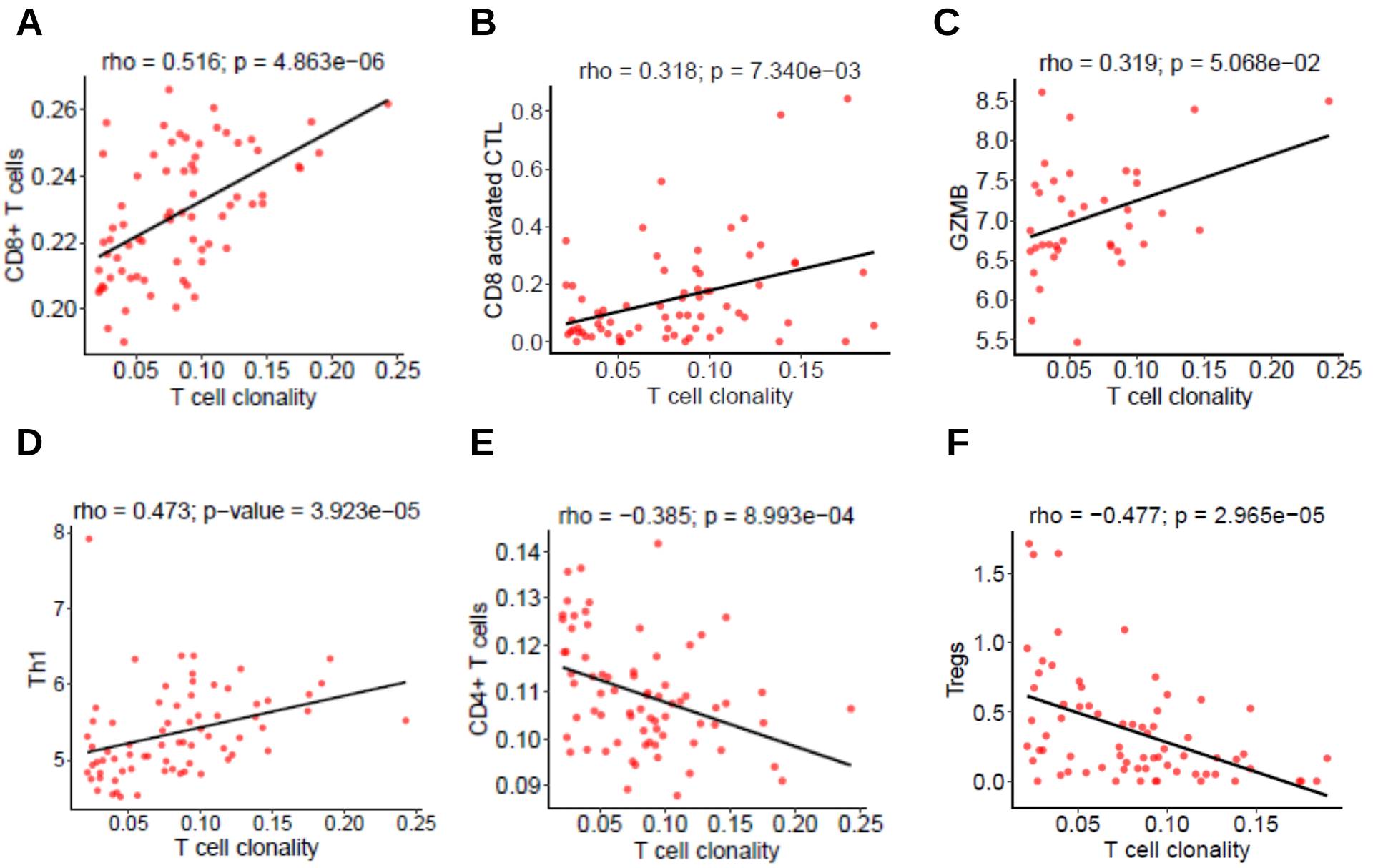
Supplementary Figure 11



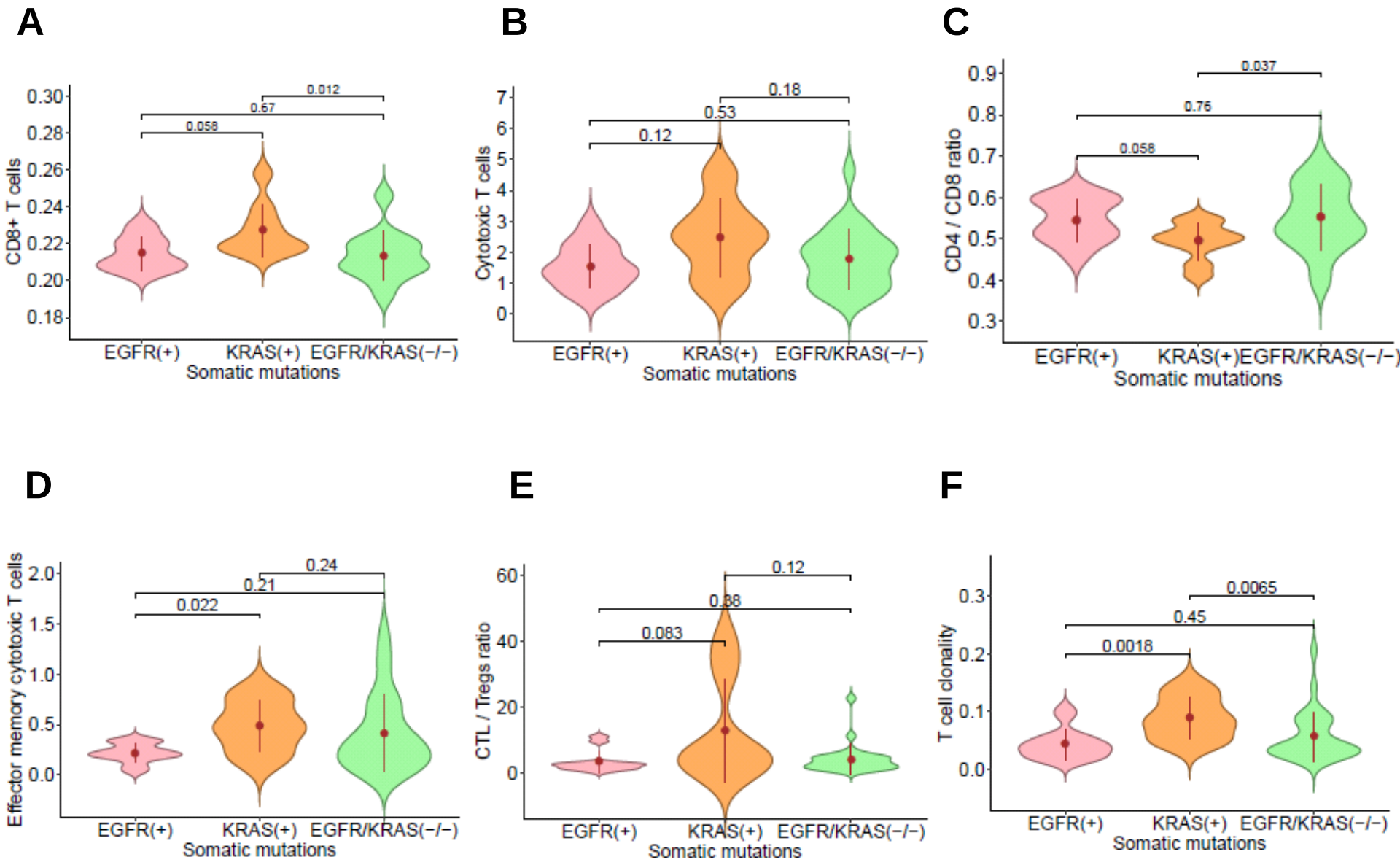
Supplementary Figure 12



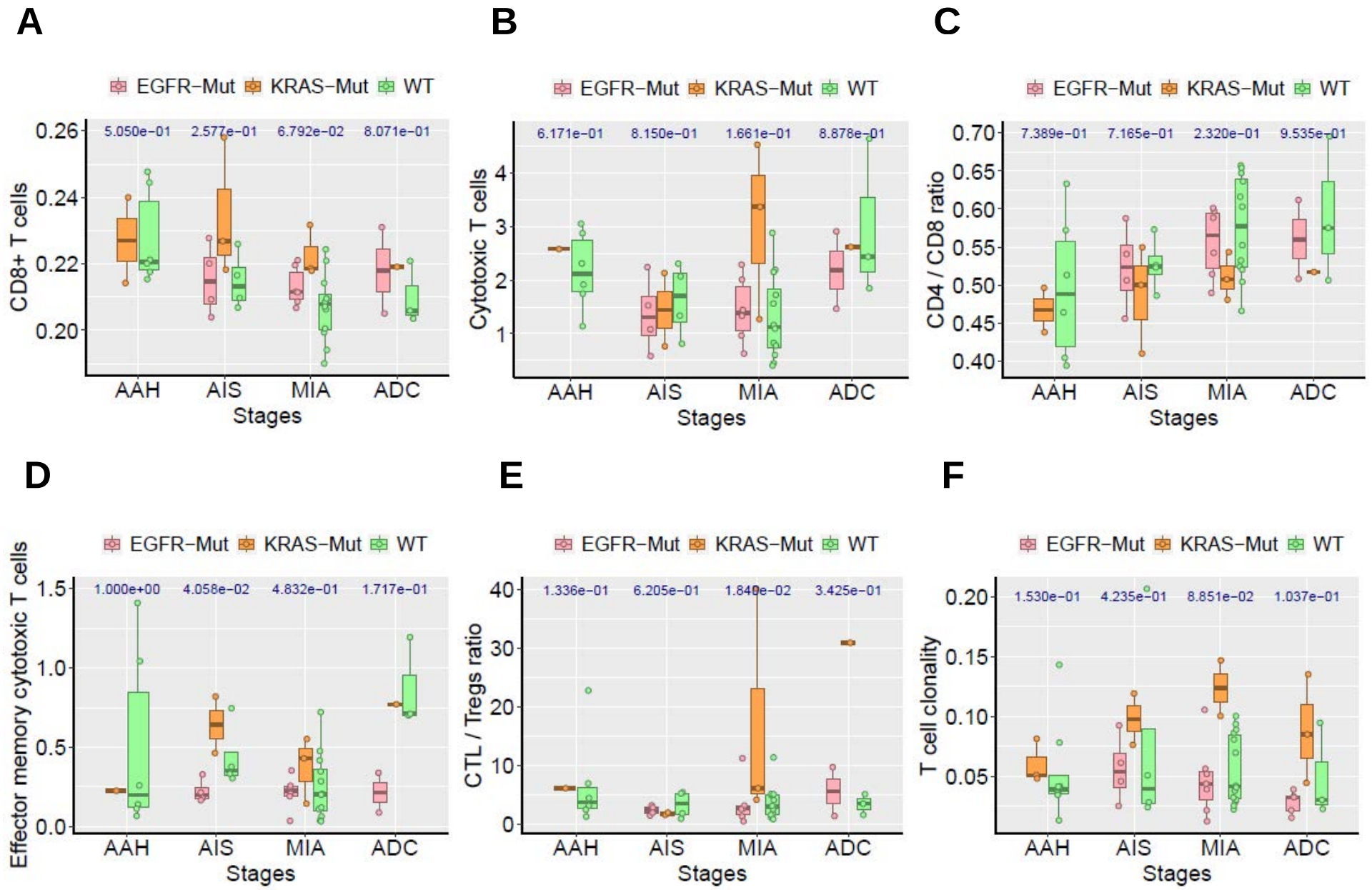
Supplementary Figure 13



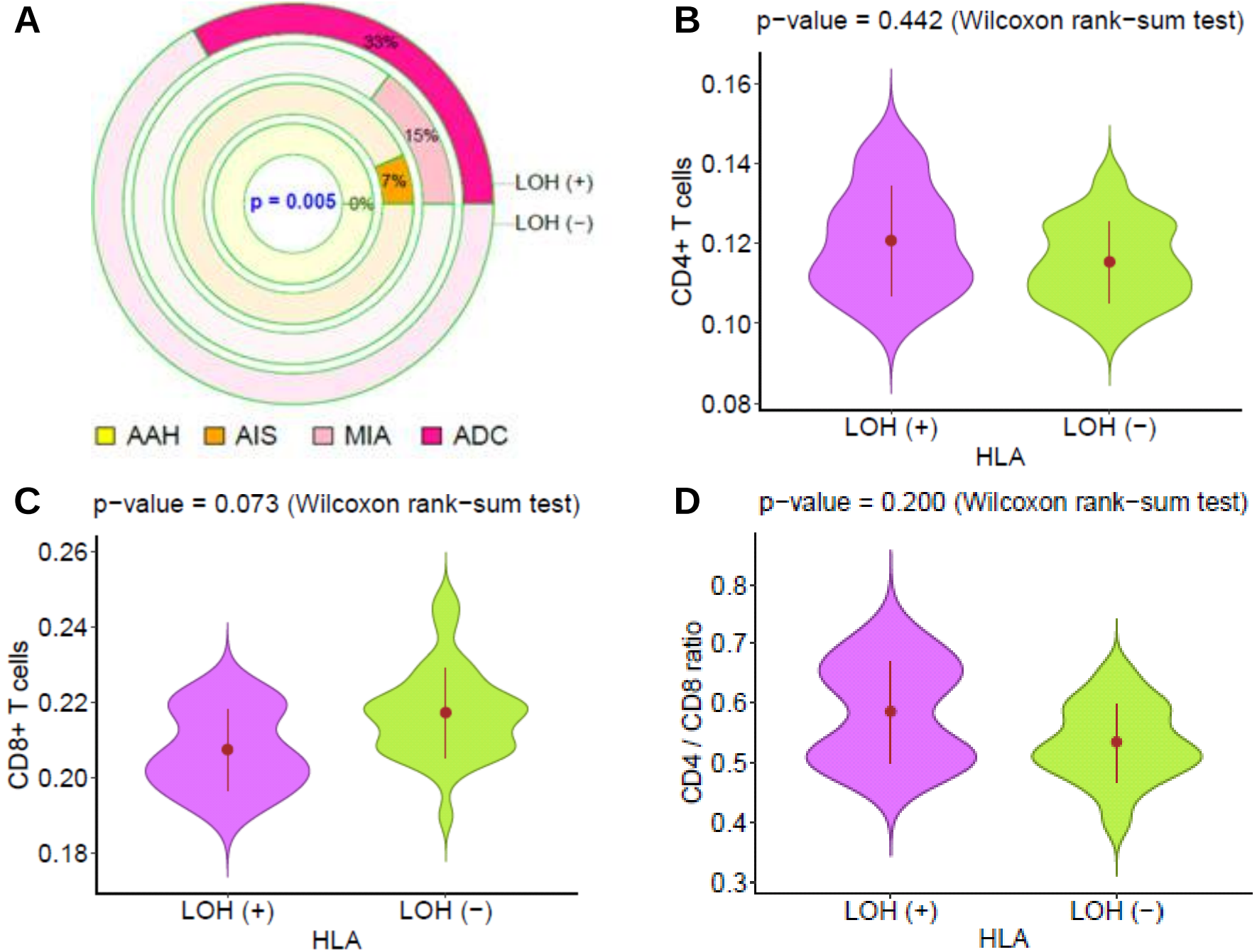
Supplementary Figure 14



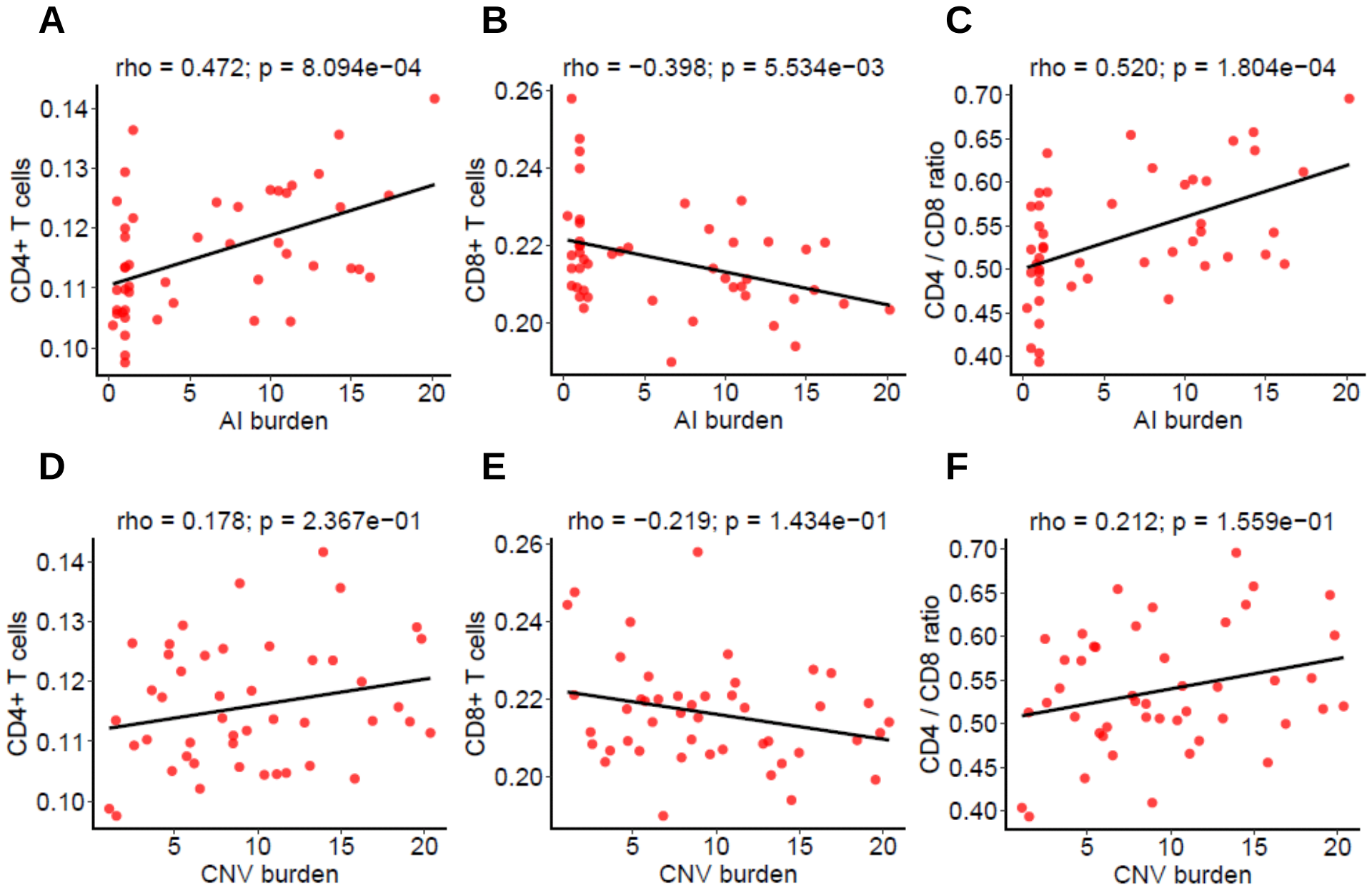
Supplementary Figure 15



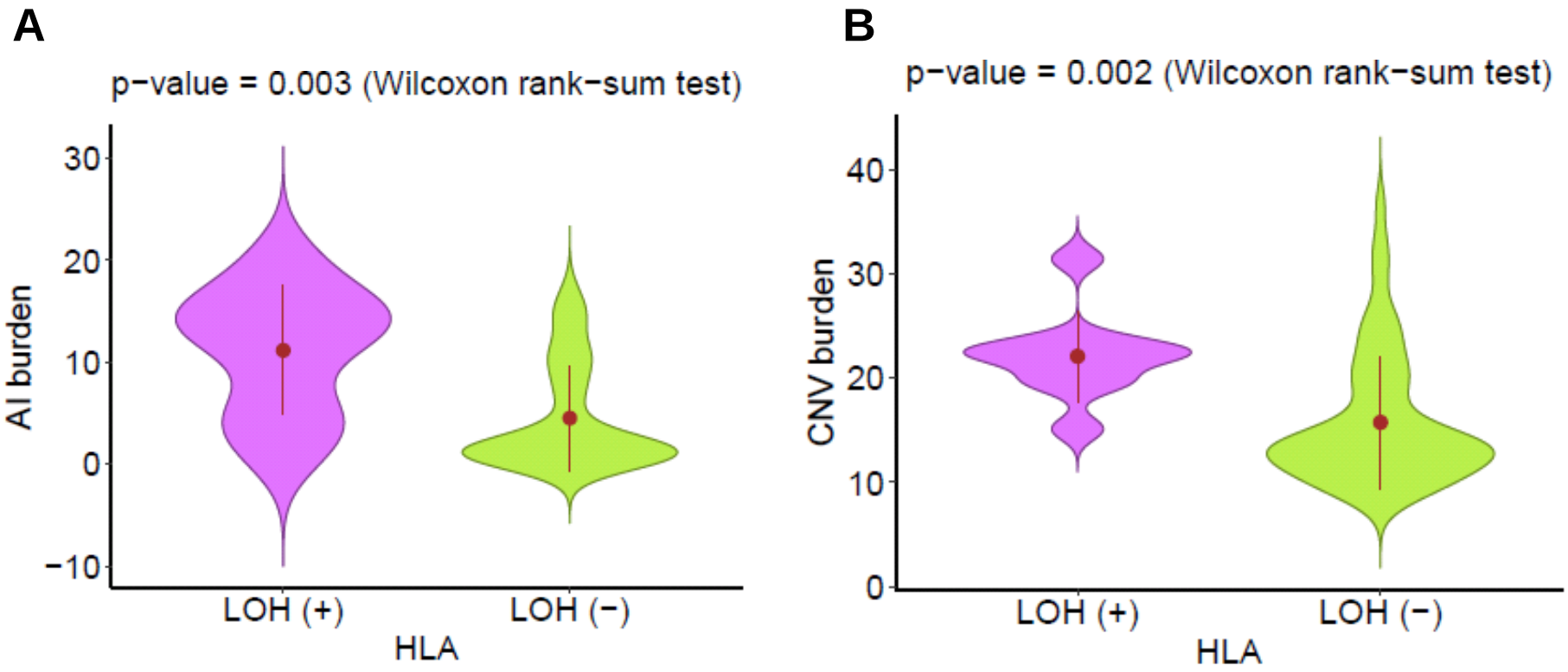
Supplementary Figure 16



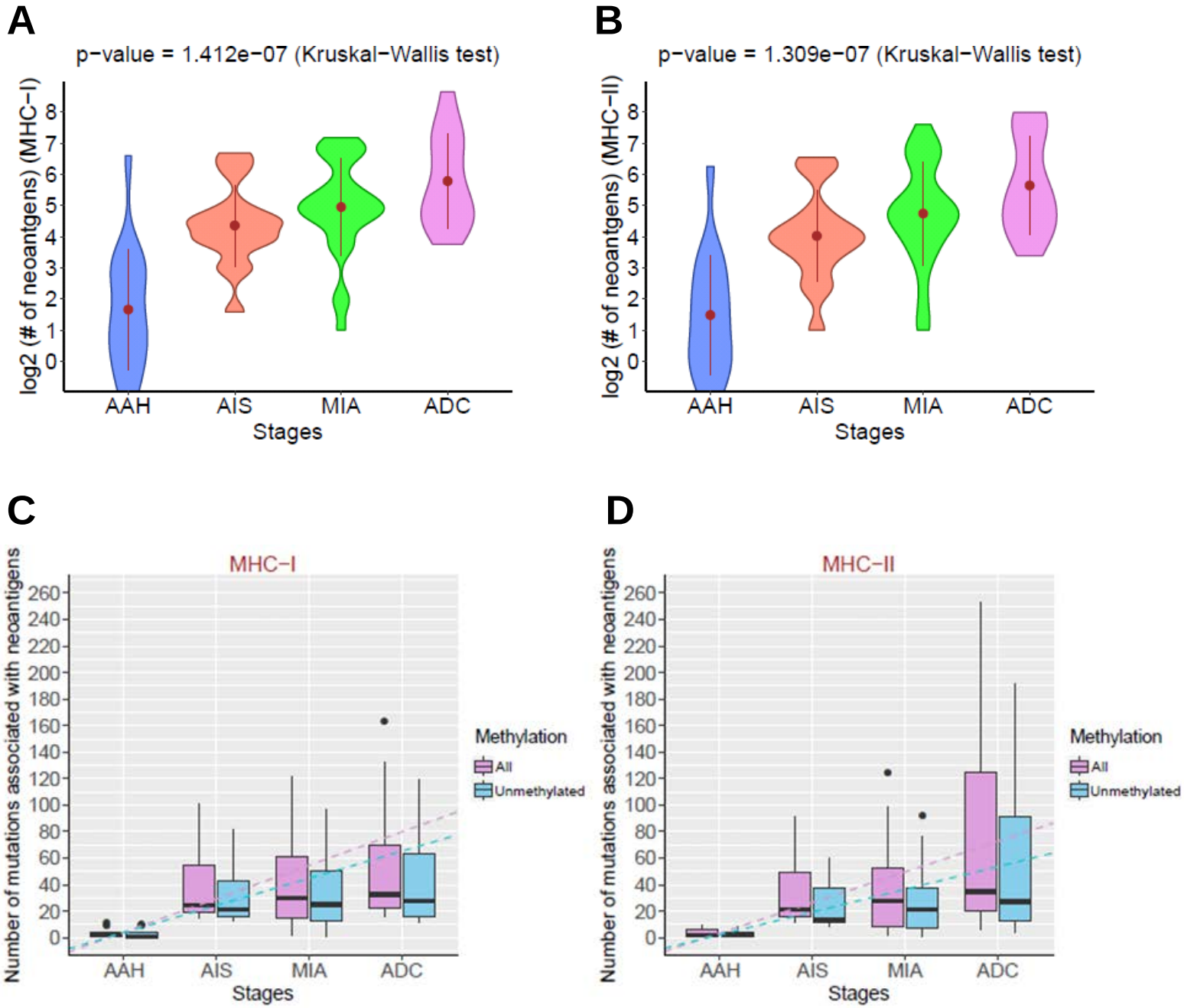
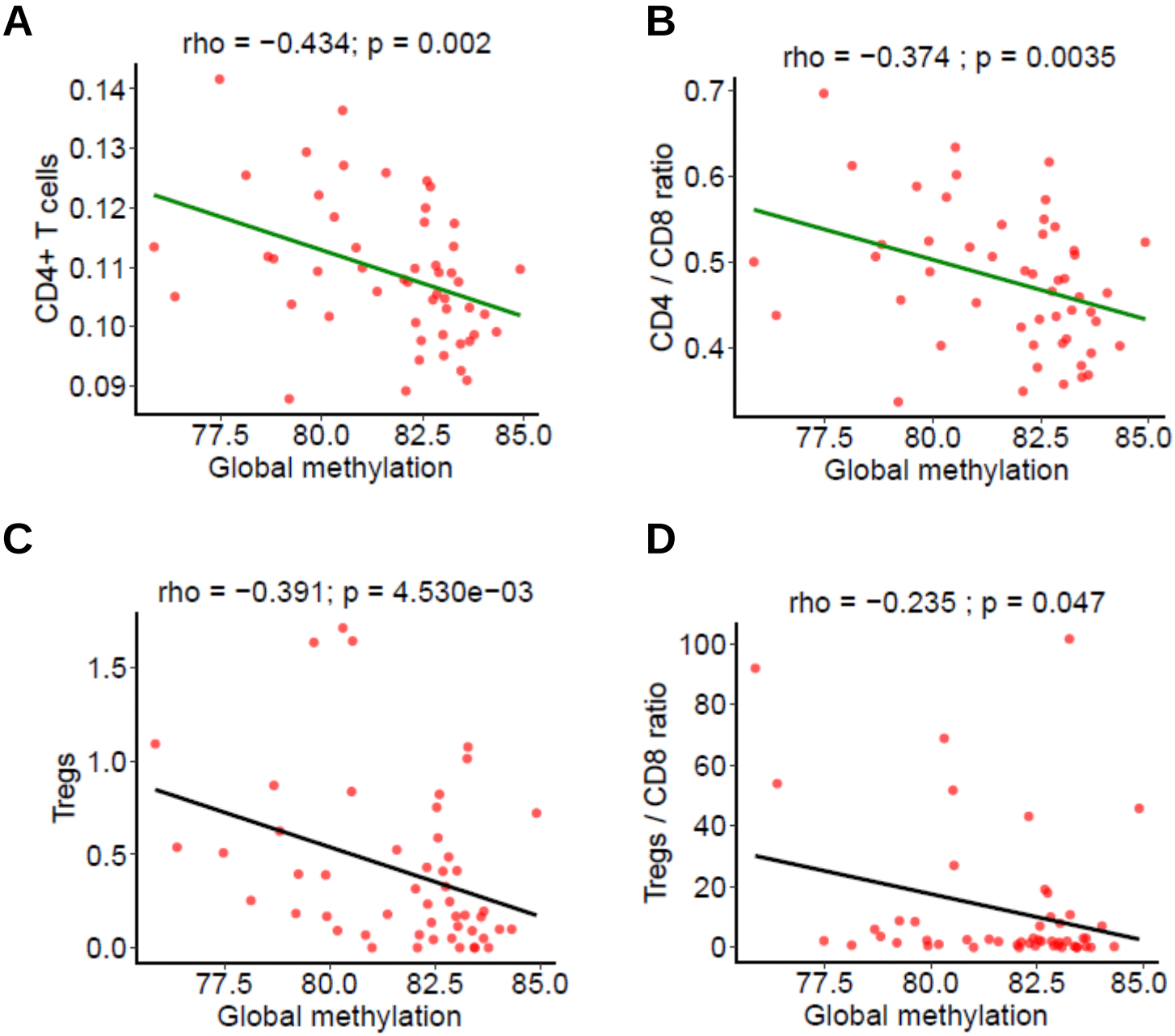
Supplemental Figure S17



Supplementary Figure 18



Supplementary Figure 19



Supplementary Figure 20