의학박사 학위논문

Investigating clinico-molecular and immunological evolution of lung adenocarcinoma using a pseudotime analysis

유사시간분석을 이용한 폐선암의 임상분자적 및 면역학적 변화 연구

2021년 8월

서울대학교 융합과학기술대학원

분자의학 및 바이오제약학과

이 현 종

Investigating clinico-molecular and immunological evolution of lung adenocarcinoma using a pseudotime analysis

지도교수 이 동 수

이 논문을 의학박사 학위논문으로 제출함

2021년 4월

서울대학교 융합과학기술대학원

분자의학 및 바이오제약학과

이 현 종

이현종의 의학박사 학위논문을 인준함

2021년 7월

위 원 장 김 태 유

부위원장 이 동 수

위 원 천 기 정

위 원 최 홍 윤

위 원 최 준 영

Abstract

Investigating clinico-molecular and immunological evolution of lung adenocarcinoma using a pseudotime analysis

Hyunjong Lee

Department of Molecular Medicine and Biopharmaceutical Sciences,

Graduate School of Convergence Science and Technology,

Seoul National University

Lung adenocarcinoma (LUAD) is the most common histological type among lung malignancies. Because molecular features of lung cancer have been evaluated as a cross-section study, the course of biological progression of lung cancer has not been modeled. The progression of lung cancer has been clinically and pathologically evaluated by staging, histological features, and glucose metabolism on ^{18}F-fluorodeoxyglucose positron emission tomography (FDG PET), while they hardly model the biological progression in terms of evolution of molecular profiles. Here, pseudotime trajectories reflecting the biological

progression of LUAD were estimated from publicly available datasets. The evolution of tumor characteristics was interrogated along the pseudotime trajectories.

A pseudotime trajectory was constructed in lung cancer dataset from the Cancer Genome Atlas (TCGA) using “Phenopath” tool. Genes associating with pseudotime were selected and gene ontology analysis was performed. A pseudotime trajectory was estimated also in LUAD samples of NSCLC radiogenomics dataset using lasso regression. Correlation analyses were performed between clinical factors including TNM stage and pseudotime. FDG PET images of subjects were collected from the Cancer Imaging Archive (TCIA). Imaging parameters including standardized uptake value (SUV) were obtained from region-of-interests drawn with adaptive tumor margin in LifeX software. Correlation analyses were performed between imaging parameters and pseudotime. Immune profiles were obtained using “xCell” tool. Gene mutation data were downloaded from genomic data commons (https://gdc.cancer.gov/). Correlation analyses were performed between enrichment scores of immune cells, tumor mutation burden, and pseudotime. Immunologic factors associating with lymph node metastasis and glucose metabolism were explored using differentially expressed genes, gene ontology, and comparison analyses.

A pseudotime trajectory was successfully estimated in TCGA dataset. In LUAD samples, molecular profiles related to natural killer cell activity were downregulated along pseudotime. Those related to cell division were upregulated along pseudotime. A pseudotime trajectory was estimated in LUAD samples of NSCLC radiogenomics dataset with lasso regression model constructed from TCGA datasets. In TCGA dataset, there was a significant difference of pseudotime in each T stages and overall TNM stages, whereas no difference in N and M stages. Overall survival was significantly different between early pseudotime group and late pseudotime group. Maximal SUV and solute carrier family 2 member 1 gene (SLC2A1) expression showed a positive correlation with pseudotime. Cell enrichment in tumor immune microenvironment was changed according to the pseudotime. Type 1 helper

T (Th1) cells showed a positive correlation, whereas M2 macrophages showed a negative correlation with pseudotime. In early pseudotime group, triggering receptor expressed on myeloid cells-1 gene (TREM-1) was upregulated in adenocarcinoma patients with lymph node metastasis than in those without lymph node metastasis. In late pseudotime group, negative regulation of mononuclear cell migration was selected as a significantly downregulated function in samples with high FDG uptake.

In conclusion, the pseudotime trajectories of lung cancer were estimated according to transcriptomic profiles. Clinical stage and SUV demonstrated significant association with pseudotime, showing feasibility of new scale evaluating molecular progression of lung cancer. Among tumor immune microenvironment, Th1 cells and M2 macrophages showed positive and negative correlation with pseudotime, respectively. Immunologic factors were revealed to associate with lymph node metastasis in early phase of disease and glucose metabolism in late phase of disease. The present study represented evolution of tumor characteristics in LUAD using pseudotime analysis.

Keywords: lung adenocarcinoma, metastasis, immune, metabolism, pseudotime

Student number: 2015-26015

Contents

Abstract ····· i

Contents ····· iv

List of Figures ····· vi

List of Abbreviations ····· ix

Introduction ····· 1

Staging of lung adenocarcinoma ····· 1

FDG PET in lung adenocarcinoma ····· 1

Immune microenvironment in lung adenocarcinoma ····· 2

Pseudotime analysis ····· 2

Purpose ····· 4

Materials and Methods ····· 5

Pseudotime estimation ····· 5

Genetic feature analysis ····· 7

Pseudotime prediction in NSCLC radiogenomics dataset ····· 7

Clinical feature analysis ····· 9

Glucose metabolism analysis ····· 9

Immune profile analysis ····· 9

Features associating with lymph node metastasis and glucose metabolism ····· 10

Statistical analysis ····· 11

Results ····· 12

Part I. Pseudotime estimation and validation ····· 12

Temporal evolution of genetic features ····· 12

Temporal evolution of clinical features ····· 16

Temporal evolution of glucose metabolism in TCGA-LUAD dataset ····· 21

Temporal evolution of glucose metabolism in LUAD samples of NSCLC radiogenomics dataset ····· 24
Part II. Temporal evolution of immune profiles in lung adenocarcinoma ····· 27
Temporal evolution of immune profiles in TCGA-LUAD dataset ····· 27
Temporal evolution of immune profiles in LUAD samples of NSCLC radiogenomics dataset ····· 36
Part III. Immunological factors associating with lymph node metastasis and glucose metabolism ····· 41
TREM-1 associating with lymph node metastasis ····· 41
Mononuclear cell associating with glucose metabolism ····· 47
Discussion ····· 54
Conclusions ····· 60
References ····· 61
국 문 초 록 ····· 68

List of Figures

Figure 1. Overall scheme of the study. 6

Figure 2. Result of cross-validation to select lambda with the least error for lasso regression. 8

Figure 3. PCA plots of TCGA lung cancer cohorts. 13

Figure 4. Genes upregulated and downregulated along pseudotime. 14

Figure 5. Gene ontology analysis with genes upregulated and downregulated along pseudotime. 15

Figure 6. A heatmap visualizing clinical factors of each sample with top 10 genes associated along pseudotime in TCGA-LUAD dataset. 17

Figure 7. Boxplots visualizing difference of pseudotime according to TNM stages in TCGA-LUAD dataset. 18

Figure 8. Kaplan-Meier curves in TCGA-LUAD dataset. 19

Figure 9. Boxplots visualizing difference of pseudotime according to TNM stages in LUAD samples of NSCLC radiogenomics dataset. 20

Figure 10. Scatter plots showing correlation between SUV and pseudotime in TCGA-LUAD dataset. 22

Figure 11. Scatter plots showing correlation of SLC2A1 expression with pseudotime and SUV in TCGA-LUAD dataset. 23

Figure 12. Scatter plots showing correlation between SUV and pseudotime in LUAD samples of NSCLC radiogenomics dataset. 25

Figure 13. Scatter plots showing correlation of SLC2A1 expression with pseudotime and SUV in LUAD samples of NSCLC radiogenomics dataset. 26

Figure 14. A volcano plot representing immune cells associating with pseudotime in TCGA-LUAD dataset. 28

Figure 15. A heatmap representing immune and stromal cells associating with pseudotime in TCGA-LUAD dataset. 29
Figure 16. Scatter plots showing correlation between immune cell enrichment scores and pseudotime in TCGA-LUAD dataset. 30
Figure 17. Scatter plots showing correlation between PD-L1 expression, TMB, and pseudotime in TCGA-LUAD dataset. 31
Figure 18. Difference and association with pseudotime of Th1 cell enrichment in each mutation group. 32
Figure 19. Difference and association with pseudotime of M2 macrophages enrichment in each mutation group. 33
Figure 20. Difference and association with pseudotime of PD-L1 expression in each mutation group..... 34
Figure 21. Difference and association with pseudotime of TMB in each mutation group. . 35
Figure 22. A volcano plot representing immune cells associating with pseudotime in LUAD samples of NSCLC radiogenomics dataset..... 37
Figure 23. A heatmap representing immune and stromal cells associating with pseudotime in LUAD samples of NSCLC radiogenomics dataset..... 38
Figure 24. Scatter plots showing correlation between immune cell enrichment scores and pseudotime in LUAD samples of NSCLC radiogenomics dataset. 39
Figure 25. A scatter plot showing correlation between PD-L1 expression and pseudotime in LUAD samples of NSCLC radiogenomics dataset..... 40
Figure 26. A mosaic plot representing difference of subjects with lymph node metastasis in early pseudotime and late pseudotime groups. 42
Figure 27. Volcano plots representing DEGs between samples without lymph node metastasis and those with lymph node metastasis. 43
Figure 28. Gene ontology analysis with DEGs between samples without lymph node

metastasis and those with lymph node metastasis. 44
Figure 29. Violin plots representing difference of immune cell enrichment scores. 45
Figure 30. Differences of TREM-1 expression and monocytic MDSC enrichment. 46
Figure 31. A scatter plot showing weak positive correlation between SUV and pseudotime in LUAD samples of NSCLC radiogenomics dataset. 48
Figure 32. Mosaic plots representing difference of subjects with EGFR mutation according to FDG uptake. 49
Figure 33. Mosaic plots representing difference of subjects with different histological grade according to FDG uptake. 50
Figure 34. Gene ontology analysis with DEGs between samples with high FDG uptake and those with low FDG uptake. 51
Figure 35. Violin plots representing difference of immune cells according to FDG uptake. 52
Figure 36. Scatter plots showing correlation of Th1 cell enrichment with pseudotime and SLC2A1 expression in LUAD sample of NSCLC radiogenomics dataset. 53

List of Abbreviations

Full name	Abbreviations
Lung adenocarcinoma	LUAD
Lymph node	LN
^{18}F-fluorodeoxyglucose	FDG
Positron emission tomography	PET
Tumor immune microenvironment	TIME
Programmed cell death protein 1	PD-1
Programmed death ligand 1	PD-L1
Tumor-associated macrophages	TAM
RNA-sequencing	RNA-seq
The Cancer Genome Atlas	TCGA
Non-small cell lung cancer	NSCLC
Lung squamous cell carcinoma	LUSC
Principal component analysis	PCA
Gene ontology	GO

Disease-free survival	DFS
Overall survival	OS
Differentially expressed gene	DEG
Tumor mutation burden	TMB
False discovery rate	FDR
The Cancer Imaging Archive	TCIA
Standardized uptake value	SUV
Metabolic tumor volume	MTV
Total lesion glycolysis	TLG
Solute carrier family 2 member 1	SLC2A1
Triggering receptor expressed on myeloid cells-1	TREM-1
Type 1 helper T cell	Th1 cell
Myeloid-derived suppressor cells	MDSC
Integrin subunit alpha M	ITGAM
Natural killer	NK

Introduction

Staging of lung adenocarcinoma

Lung adenocarcinoma (LUAD) is the most frequent histological type among lung malignancies (1, 2). Diagnosis and treatment of LUAD is very important in terms that lung malignancy is the leading cause of cancer-related deaths not only in Korea but also in other countries (3, 4). Since the first revision of the lung cancer TNM staging in 1973, TNM stage is known as the most basic and critical factor to evaluate status of disease. This staging system is supported by clinical observation that prognosis worsens as the size of tumor and extent of metastasis increase (5). Therefore, scholars have been commonly used TNM staging system as a reference of disease progression in previous studies. However, there is limitation to investigate natural progression of tumor based on TNM staging as it is evaluated at a timing of initial diagnosis. Referring to the current TNM staging system, patients with lymph node (LN) or distant metastasis are classified into high stage despite small-sized tumor. In clinical fields, early T stage patients with LN metastasis or high T stage patients without are not uncommon.

FDG PET in lung adenocarcinoma

^{18}F-fluorodeoxyglucose positron emission tomography (FDG PET) is an essential imaging tool in diagnosing lung cancer. It is well established knowledge that FDG PET provides better accuracy in staging and more information in therapeutic response assessment (6, 7). Especially, FDG uptake of tumor is known to be increased according to stage of lung cancer (8, 9). On the contrary to this knowledge, there have been several cases with low FDG uptake in tumor of high T stage or high FDG uptake in tumor of low T stage. These inconsistent clinical experiences raise questions which factors affect FDG uptake in the early phase of

cancer progression.

Immune microenvironment in lung adenocarcinoma

The tumor immune microenvironment (TIME) plays a crucial role in tumor progression and metastasis. Among immune cells, tumor-associated macrophages (TAMs) exert various functions in lung cancer by differentiating into different subtypes, M1 and M2 macrophages: M1 macrophages mainly contribute to anti-tumor activity and M2 macrophages to pro-tumor activity (10). CD4+ and CD8+ T cells are revealed to associate with prognosis of LUAD (11, 12). Programmed cell death protein 1 (PD-1) / programmed death ligand 1 (PD-L1) pathway is the most well-known pathway to modulate immune response via regulating activity of T cells (13). Pembrolizumab, a PD-1 inhibitor, showed improvement of prognosis in lung cancer patients (14). Thus, characterization of TIME is important to explore therapeutic target and predict response of immunotherapy (15, 16). There have been a few studies to report difference of TIME according to the stage of lung cancer. Zeni et al. showed that expression of interleukin-10 in TAMs was higher in late stage of lung cancer (17). Okita et al. documented that expression of PD-L1 was higher in subjects with LN metastasis or high pathological stage (18). However, there have been no study to investigate evolution of TIME along continuous progression of disease.

Pseudotime analysis

Pseudotime analysis, also called trajectory inference analysis, is a spotlighted method to explore change of cell or tissue characteristics based on transcriptomic expression (19). It provides a numerical scale to reflect where a cell or tissue is in the course of disease, other than TNM staging system. It is possible to investigate evolution of tumor characteristics based on the continuous and quantified time series and even validate conventional staging system. There are over 70 methods for pseudotime analysis. Among them, Phenopath is a

recently developed analytic tool for pseudotime estimation with linear methods (20). It showed good stability, good code assurance, and fair accuracy (21). Although there have been several studies to apply pseudotime analysis in lung cancer samples, scopes of those studies were limited to only single cell RNA-sequencing (RNA-seq) data from small numbers of patients (22, 23). Pseudotime analysis in large numbers of subjects may provide a model to explore the course of biological progression of lung cancer.

Purpose

In this study, it is aimed to reveal evolution of tumor characteristics along the pseudotime trajectories, focused on four aspects: genetic features, clinical features, glucose metabolism and immune profiles. Pseudotime trajectories were estimated in the LUAD cohorts from the Cancer Genome Atlas (TCGA) and non-small cell lung cancer (NSCLC) radiogenomics dataset. Association between genetic features, clinical features, glucose metabolism, immune profiles and pseudotime was analyzed. In addition, immunological factors associating with LN metastasis and glucose metabolism were investigated.

Materials and Methods

Pseudotime estimation

The overall study scheme is described in Figure 1. A pseudotime trajectory was constructed in two publicly available datasets; TCGA-LUAD and TCGA-lung squamous cell carcinoma (LUSC). They were obtained using “TCGAbiolinks” package in R. Legacy data of gene expression quantification were downloaded by “GDCdownload” function. There were 600 LUAD samples and 553 LUSC samples. RNA-seq data were normalized by log 2 transformation. Highly variable genes were selected using “DESeq2” package in R. First, estimates of variance and coefficient of variation were calculated. Subsequently, a regression line was fitted. Finally, genes showing significant deviation from regression line were selected by chi-squared test with a threshold of 0.001 as p-value. 8589 genes were selected as HVGs among total 21022 genes. A pseudotime trajectory was generated using “Phenopath” package in R (20). An input data was a gene expression matrix of highly variable genes from TCGA-LUAD and TCGA-LUSC datasets. We choose an evidence lower bound as 10^{-6} and compute it thinned by 2 iterations.

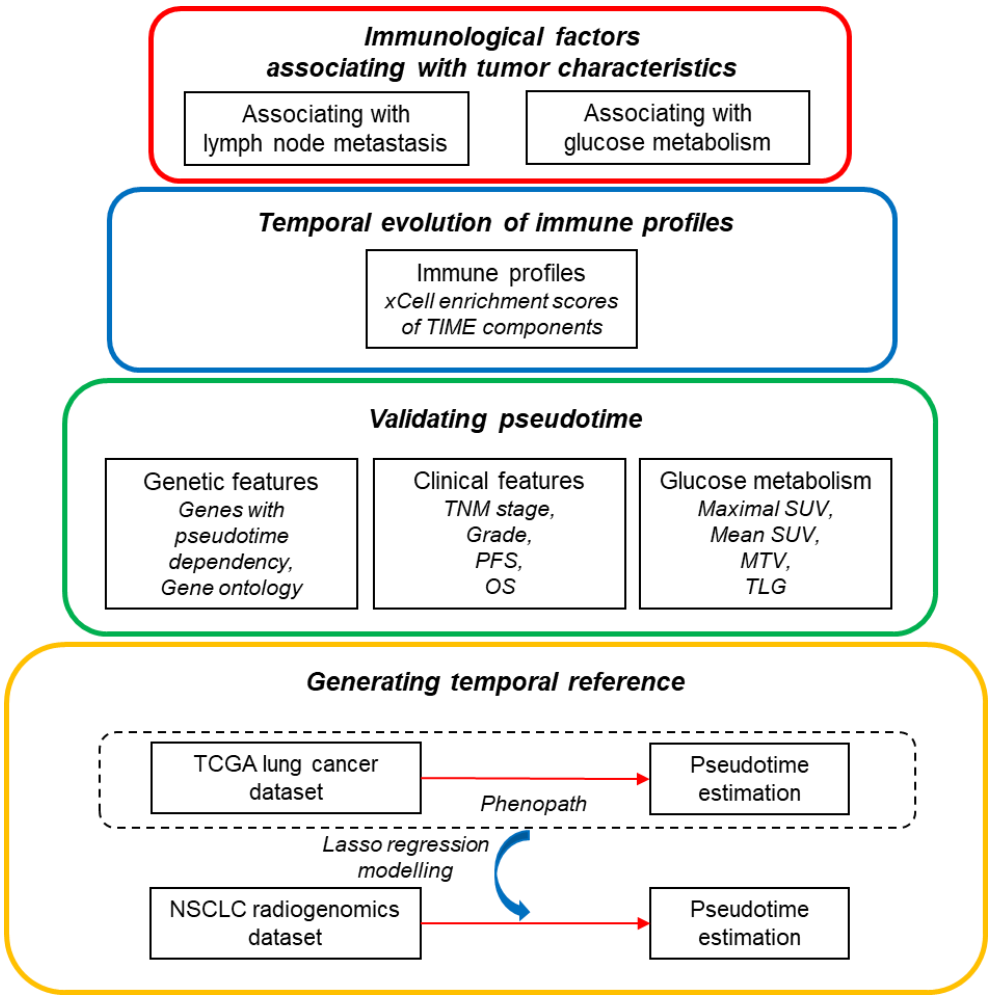

Figure 1. Overall scheme of the study.

Genetic feature analysis

Principal component analysis (PCA) was performed to visualize temporal evolution of genetic characteristics of LUAD and LUSC, using "PCA" function included in "factoextra" package in R. Phenopath analysis provided four output values; alpha: degree of differential expression, beta: degree of covariate-pseudotime interaction, lambda: degree of pseudotime dependency, z: estimates of pseudotime. Bayesian significant test was applied to select genes showing significant pseudotime dependency (non-zero lambda) and significant covariate-pseudotime dependency (non-zero beta). Gene ontology (GO) analysis was conducted on genes showing significant pseudotime dependency to investigate which functions were upregulated or downregulated along the pseudotime trajectory, using "enrichGO" function included in "clusterProfiler" package in R.

Pseudotime prediction in NSCLC radiogenomics dataset

LUAD samples of NSCLC radiogenomics dataset were employed to perform validation analysis. A RNA-seq dataset (GSE103584) was downloaded from Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/) (24). There were 96 LUAD samples in NSCLC radiogenomics dataset. A lasso regression model was used to estimate pseudotime in LUAD samples of NSCLC radiogenomics dataset based on results of TCGA dataset. Two hundred genes were selected from genes which showed significant association between pseudotime: top 100 genes in the positive correlation group, top 100 genes in the negative correlation group. An expression matrix of those 200 genes was constructed from TCGA dataset. It was divided into two groups with 2:1 ratio, training and test data. A lambda with the least error was selected via cross-validation (Figure 2). A lasso regression model was obtained. An alpha was 1. The model was applied to LUAD samples of NSCLC radiogenomics dataset to predict a pseudotime trajectory.

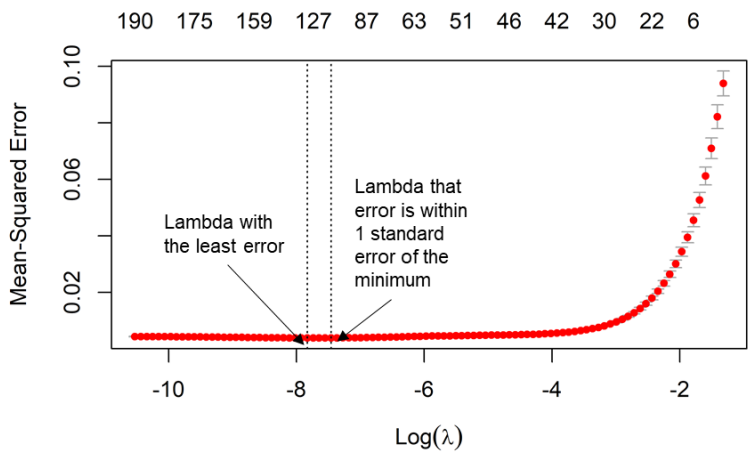

Figure 2. Result of cross-validation to select lambda with the least error for lasso regression.

For lasso regression modelling, a lambda with the least error was selected via cross-validation.

Clinical feature analysis

Clinical data of TCGA dataset were downloaded from cBioPortal (http://www.cbioportal.org/) using "cgdsr" package in R. TNM stage, disease-free status (DFS), overall survival (OS), and duration of DFS/OS were selected as representative clinical factors. A heatmap was plotted to visualize association between genes, clinical factors, and pseudotime using "Complexheatmap" package in R. Pseudotime of each TNM stage group was compared each other using t-test or ANOVA test. Survival curves were plotted with a median value of pseudotime as a threshold. Clinical data of NSCLC radiogenomics dataset were downloaded from The Cancer Imaging Archive (TCIA, https://www.cancerimagingarchive.net/). Pseudotime of each TNM stage group was compared each other using Wilcoxon rank-sum test or Kruskal-Wallis test.

Glucose metabolism analysis

FDG PET images of LUAD subjects from both TCGA dataset and NSCLC radiogenomics dataset were downloaded from TCIA. There were 16 and 93 samples with both RNA-seq data and FDG PET images in LUAD samples of TCGA dataset and NSCLC radiogenomics dataset, respectively. Tumor margins were delineated using Nestle adaptive threshold method provided by "LifeX" software (25, 26). Beta was 0.3. Maximal standardized uptake value (SUV), mean SUV, and metabolic tumor volume (MTV) were obtained from the region-of-interests. Total lesion glycolysis (TLG) was calculated from mean SUV and MTV. Correlation coefficients between FDG PET parameters, solute carrier family 2 member 1 (SLC2A1) expression, and pseudotime were calculated by Spearman and Pearson correlation test in TCGA dataset and NSCLC radiogenomics dataset, respectively.

Immune profile analysis

In both of TCGA dataset and NSCLC radiogenomics dataset, enrichment scores of 64

immune and stroma cell types were estimated using “xCellAnalysis” function in “xCell” package in R (27). Correlation coefficients between enrichment scores and pseudotime were calculated by Pearson correlation test. False discovery rate (FDR) was calculated from p-values with Bonferroni method. Volcano plot, heatmap, and scatter plot were represented to describe association between enrichment scores of immune cells and pseudotime. Expression of PD-L1 and tumor mutation burden (TMB) are well-known indicators for immune profiles of tumor (13, 28, 29). In TCGA dataset, gene mutation data were downloaded from genomic data commons (https://gdc.cancer.gov/), followed by constructing mutation annotation format file using “read.maf” function included in “maftools” package (30). TMB was estimated using “tmb” function included in “maftools” package. Correlation coefficients between expression of PD-L1, TMB, and pseudotime were calculated by Pearson correlation test. Difference of immune profiles were investigated between samples with epidermal growth factor receptor gene (EGFR) or Kirsten rat sarcoma gene (KRAS) mutation and those without, using ANOVA. Correlation between enrichment of immune cells, expression of PD-L1, TMB, and pseudotime was explored in each mutation group, respectively.

Features associating with lymph node metastasis and glucose metabolism

An additional analysis was performed to explore which factors associate with lymph node metastasis in early and late pseudotime of LUAD. A median value of pseudotime was set as the threshold to classify samples into two group: early pseudotime and late pseudotime group. Samples without lymph node metastasis (N0) was defined as “No LN metastasis” group, those with lymph node metastasis (N1, N2, N3) as “LN metastasis” group. First, chi-square test was conducted to reveal whether there is difference in possibility of lymph node metastasis between two groups, early pseudotime and late pseudotime. Subsequently, differentially expressed genes (DEGs) between “No LN metastasis” group and “LN metastasis” group were obtained using “limma” package in R. The analysis was performed

respectively in early pseudotime and late pseudotime group. Genes with FDR below 0.1 were selected as DEGs. A volcano plot was drawn using "EnhancedVolcano" package in R. GO analysis was conducted on DEGs, using "enrichGO" function included in "clusterProfiler" package in R. TMB was compared using t-test and immune cell enrichment scores using Wilcoxon rank-sum test due to non-normality. Expression of triggering receptor expressed on myeloid cells-1 (TREM-1) expression and enrichment scores of monocytic myeloid-derived suppressor cells (MDSCs) were also compared using t-test. Marker genes for monocytic MDSC were selected as integrin subunit alpha M (ITGAM), CD33, and CD14 based on a previous study (31). Enrichment scores of monocytic MDSC were estimated using "gsva" function included in "GSVA" package in R.

An additional analysis was performed to explore which factors associate with glucose metabolism in early and late pseudotime of LUAD. As described above, a median value of pseudotime was set as the threshold to classify samples into two group: early pseudotime and late pseudotime group. A median value of maximal SUV was set as the threshold to classify samples into two groups: high FDG uptake and low FDG uptake group. First, chi-square test was conducted to reveal whether there is difference in mutation profiles including status of EGFR mutation and KRAS mutation between two groups, early pseudotime and late pseudotime. Subsequently, DEG and GO analysis was performed with same methods as previously described, except 0.001 as a threshold of p-value. Immune cell enrichment scores were also compared using Wilcoxon rank-sum test.

Statistical analysis

All statistical analysis were performed using R software (v4.0.4, R Foundation for Statistical Computing, Vienna, Austria). A p-value of < 0.05 was considered statistically significant.

Results

Part I. Pseudotime estimation and validation

Temporal evolution of genetic features

TCGA-LUAD and TCGA-LUSC datasets were employed to estimate a pseudotime trajectory in bulk data of lung cancer. In PCA, LUAD and LUSC samples seem to be in same position at the beginning of pseudotime (Figure 3). As pseudotime passes, LUAD and LUSC samples are clearly discriminated in PCA plot. We investigated genes regulated along pseudotime in total lung cancer, LUAD, and LUSC samples, respectively (Figure 4a, b). 603 genes showed significantly positive correlation with pseudotime in total lung cancer samples, 2594 genes negative correlation in total cancer samples, 178 genes positive correlation in LUAD samples, 853 genes negative correlation in LUAD samples, 479 genes positive correlation in LUSC samples, and 647 genes negative correlation in LUSC samples, respectively. GO analysis was performed to uncover which biological pathways are related (Figure 5). In total lung cancer samples, molecular functions related to cell division are upregulated along pseudotime. In LUAD samples, those related to natural killer (NK) cell function are downregulated along pseudotime. Those related to cell division such as nucleosome assembly and DNA packaging are upregulated along pseudotime as in total lung cancer samples.

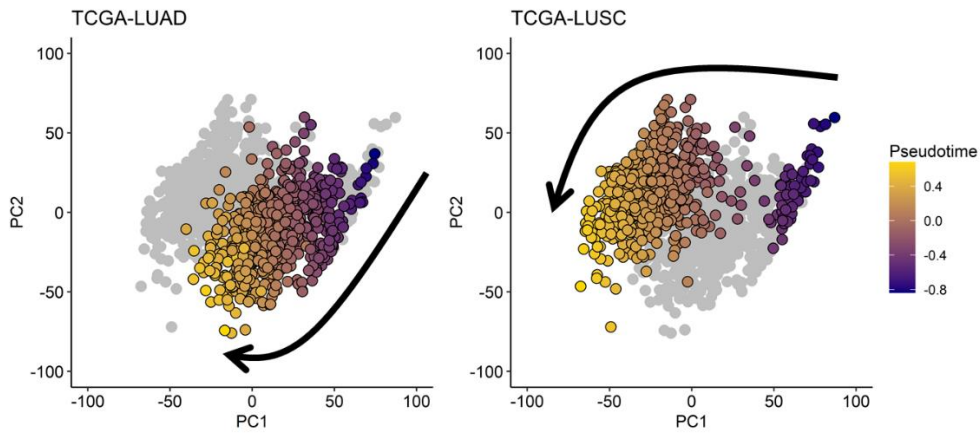

Figure 3. PCA plots of TCGA lung cancer cohorts.

Principal components analysis was performed in TCGA-LUAD and TCGA-LUSC samples. Adenocarcinoma samples (TCGA-LUAD) were plotted in the left panel and squamous cell carcinoma (TCGA-LUSC) in the right panel. Both of two histological types located in same position at the beginning of pseudotime and differentiate along pseudotime as black arrows.

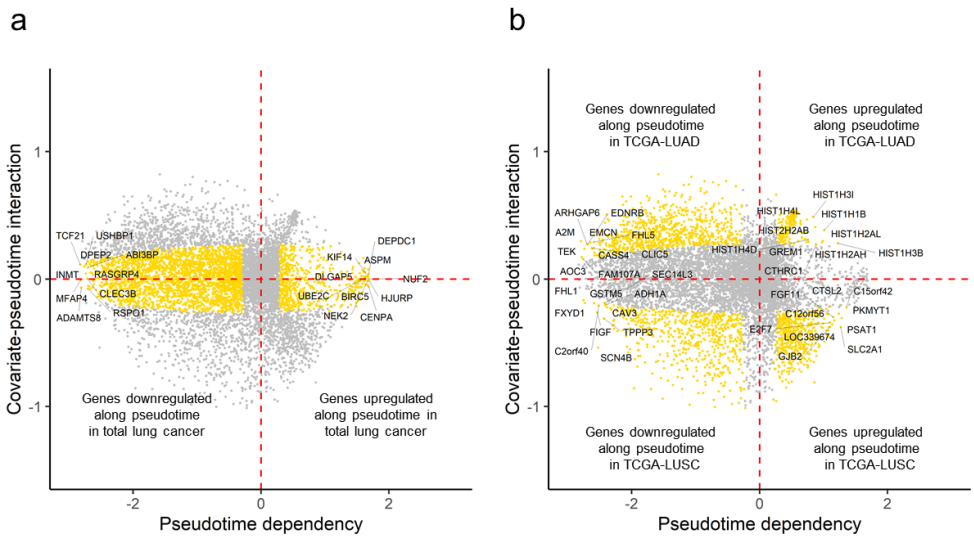

Figure 4. Genes upregulated and downregulated along pseudotime.

Genes upregulated and downregulated along pseudotime were selected and plotted with yellow dots. Top 10 genes with high pseudotime dependency were annotated. **(a)** 3197 genes showed significant correlation with pseudotime in total lung cancer. **(b)** 1031 genes showed significant correlation in LUAD samples, and 1126 genes in LUSC samples.

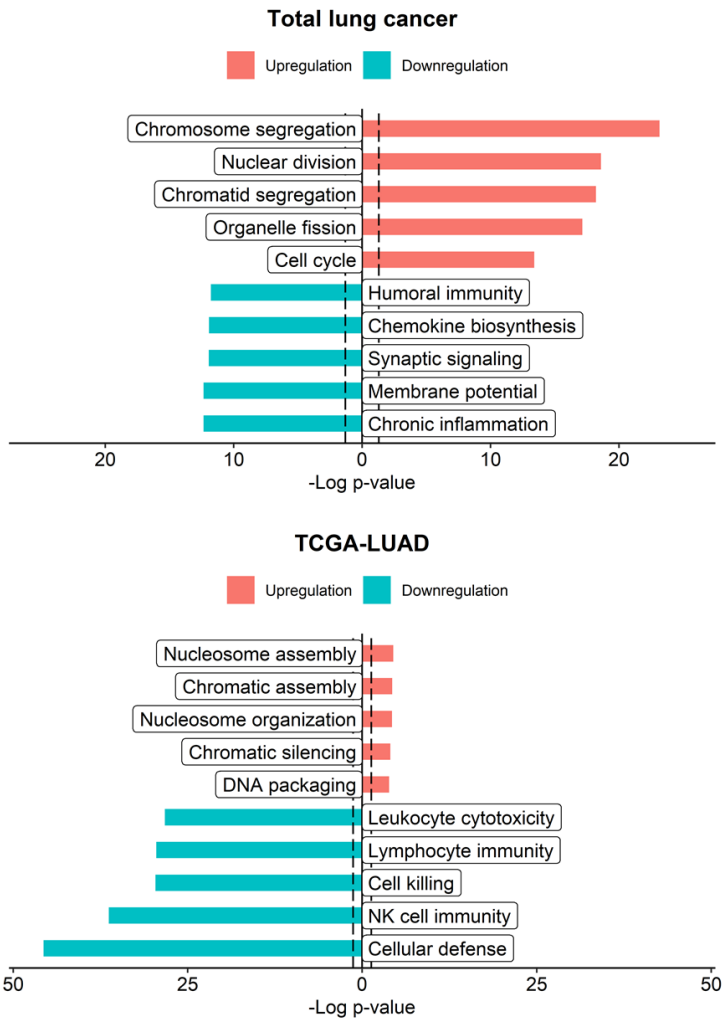

Figure 5. Gene ontology analysis with genes upregulated and downregulated along pseudotime.

Gene ontology analysis was performed in genes associating with pseudotime. In total lung cancer samples, molecular functions related to cell division are upregulated along pseudotime (upper). In LUAD samples, those related to natural killer cell activity are downregulated along pseudotime and those related to cell division upregulated (lower).

Temporal evolution of clinical features

A heatmap visualizes clinical factors of each sample with top 10 genes upregulated and those downregulated along pseudotime in LUAD samples (Figure 6). Notably, histone coding genes showed upregulation along pseudotime. Boxplots represent association between TNM stage and pseudotime in LUAD samples (Figure 7a-d). There was significant difference of pseudotime in each T stages ($p < 0.001$), especially in T1-T2 (mean: -0.08013 vs. 0.03092, $p < 0.001$) and T1-T3 (mean: -0.08013 vs. 0.08310, $p < 0.001$). Pseudotime in difference N and M stages showed no difference. There was significant difference of pseudotime in each overall TNM stage ($p < 0.001$), especially in IA-IB (mean: -0.09326 vs. 0.02969, $p < 0.001$), IA-IIB (mean: -0.09326 vs. 0.05374, $p < 0.001$), and IA-IIIA (mean: -0.09326 vs. 0.03749, $p < 0.001$). Kaplan-Meier curves were plotted for survival analysis (Figure 8a, b). OS was significantly different between early pseudotime group and late pseudotime group ($p = 0.015$).

Pseudotime was estimated in the NSCLC radiogenomics dataset based on lasso regression model from TCGA dataset. Association between clinical factors and pseudotime was validated in LUAD samples. There was mild tendency of increasing T stage along pseudotime, especially in early T stages (Figure 9a, $p = 0.097$). There was no association between N/M stage and pseudotime (Figure 9b, c). Histological grade showed association with pseudotime (Figure 9d, $p = 0.017$). There was no significant association between overall TNM stage and pseudotime (Figure 9e). However, in early stages, association with pseudotime was partially observed.

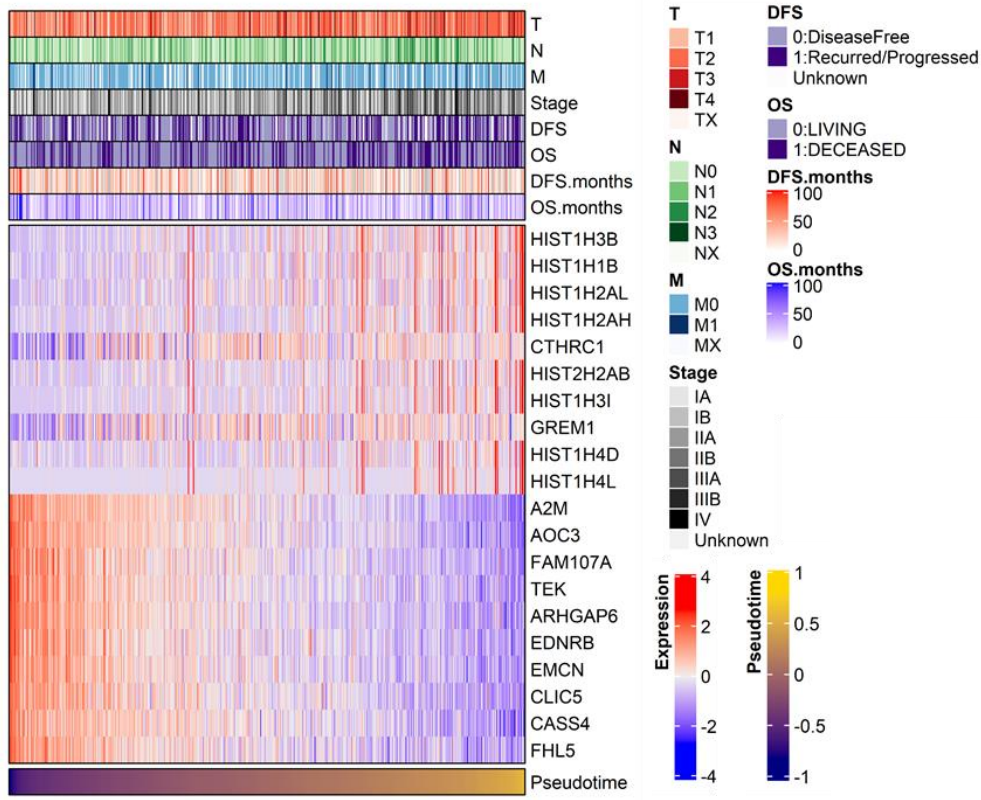

Figure 6. A heatmap visualizing clinical factors of each sample with top 10 genes associated along pseudotime in TCGA-LUAD dataset.

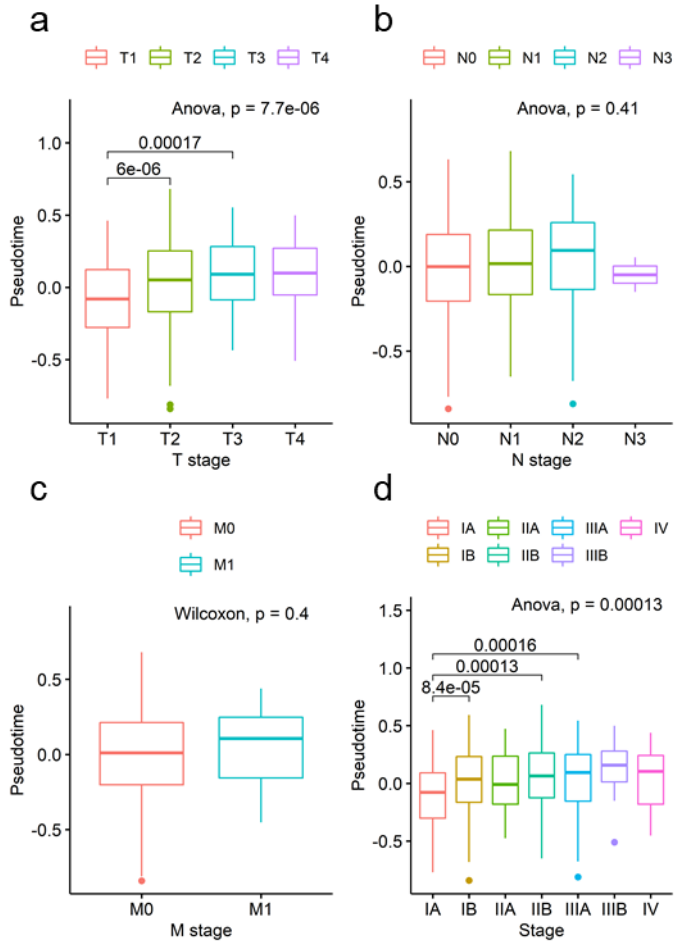

Figure 7. Boxplots visualizing difference of pseudotime according to TNM stages in TCGA-LUAD dataset.

(a) There was significant difference of pseudotime in each T stages. **(b, c)** Pseudotime in each N and M stage showed no difference. **(d)** There was significant difference of pseudotime in each overall TNM stage.

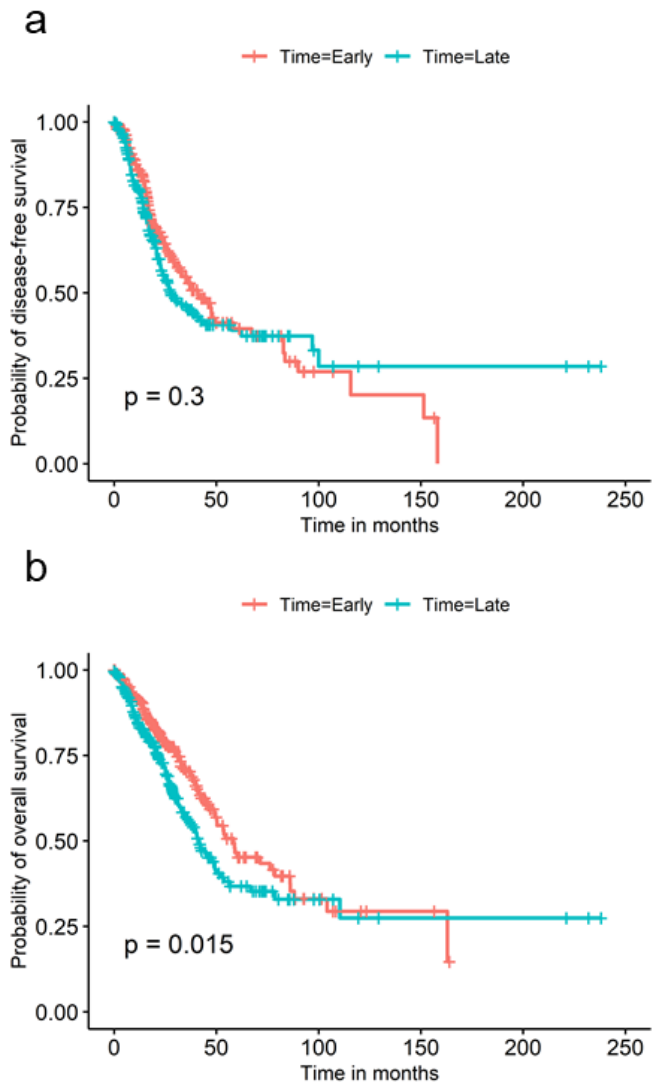

Figure 8. Kaplan-Meier curves in TCGA-LUAD dataset.

Survival analysis was performed in two groups: early pseudotime group and late pseudotime group. **(a)** There was no significant difference in disease-free survival. **(b)** Overall survival was significantly different between two groups.

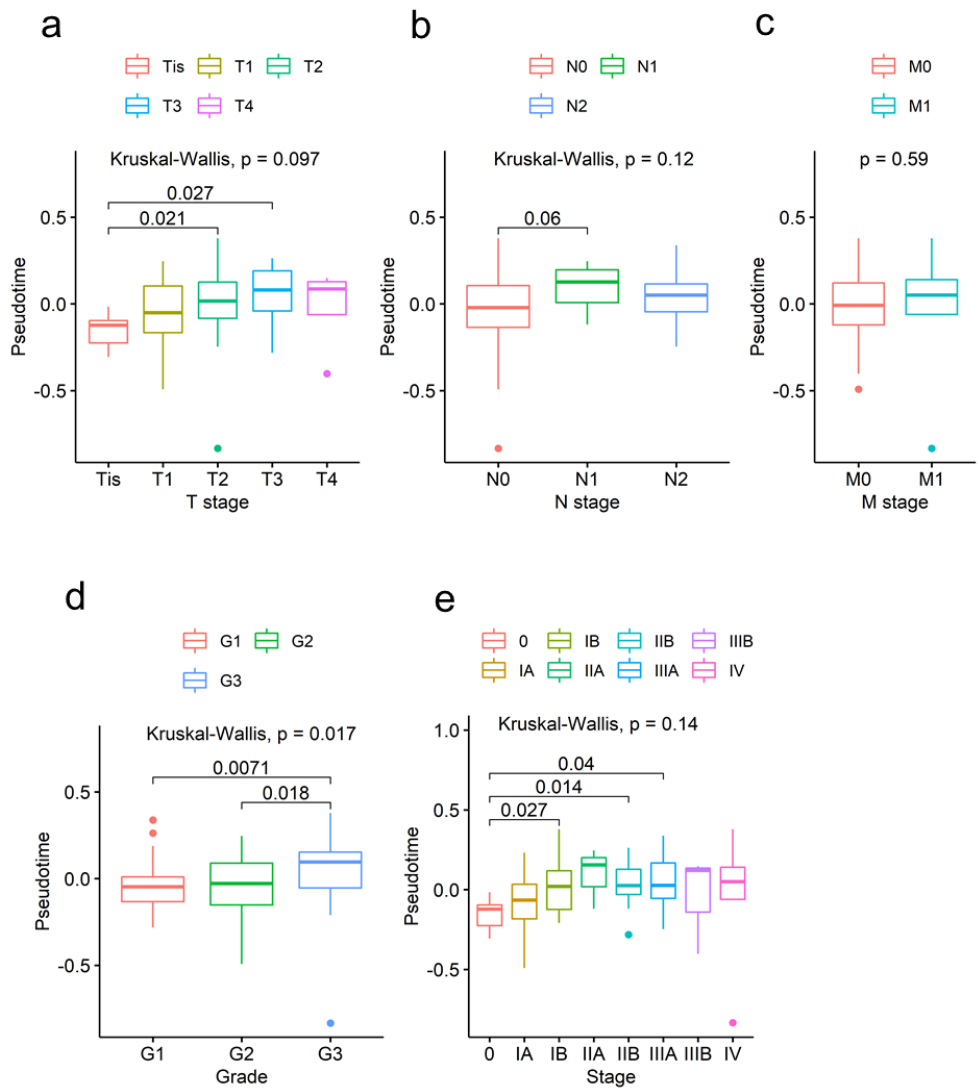

Figure 9. Boxplots visualizing difference of pseudotime according to TNM stages in LUAD samples of NSCLC radiogenomics dataset.

(a) There was mild tendency of increasing T stage along pseudotime. **(b, c)** Pseudotime in each N stage and M stage showed no difference. **(d)** There was significant difference of pseudotime in different histological grades. **(e)** There was no significant association between overall TNM stage and pseudotime.

Temporal evolution of glucose metabolism in TCGA-LUAD dataset

The Spearman correlation test was performed to reveal temporal change of glucose metabolism in LUAD. There was significant positive correlation between maximal SUV and pseudotime (Figure 10, rho = 0.518, p = 0.042). There was significant positive correlation between mean SUV and pseudotime (Figure 10, rho = 0.517, p = 0.049). However, MTV and TLG showed no association with pseudotime. There was tendency of high SLC2A1 expression in late pseudotime without statistical significance (Figure 11a, rho = 0.444, p = 0.087). However, there was no correlation between maximal SUV and expression of SLC2A1 (Figure 11b, rho = 0.409, p = 0.117).

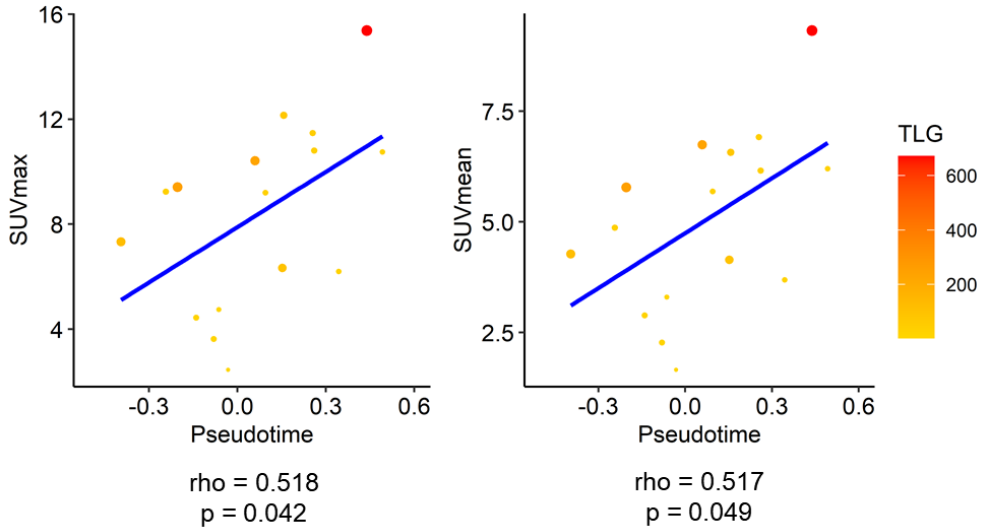

Figure 10. Scatter plots showing correlation between SUV and pseudotime in TCGA-LUAD dataset.

A size of dot represents metabolic tumor volume. Color of dot represents total lesion glycolysis. Maximal SUV showed positive correlation with pseudotime (left). Mean SUV showed positive correlation with pseudotime (right).

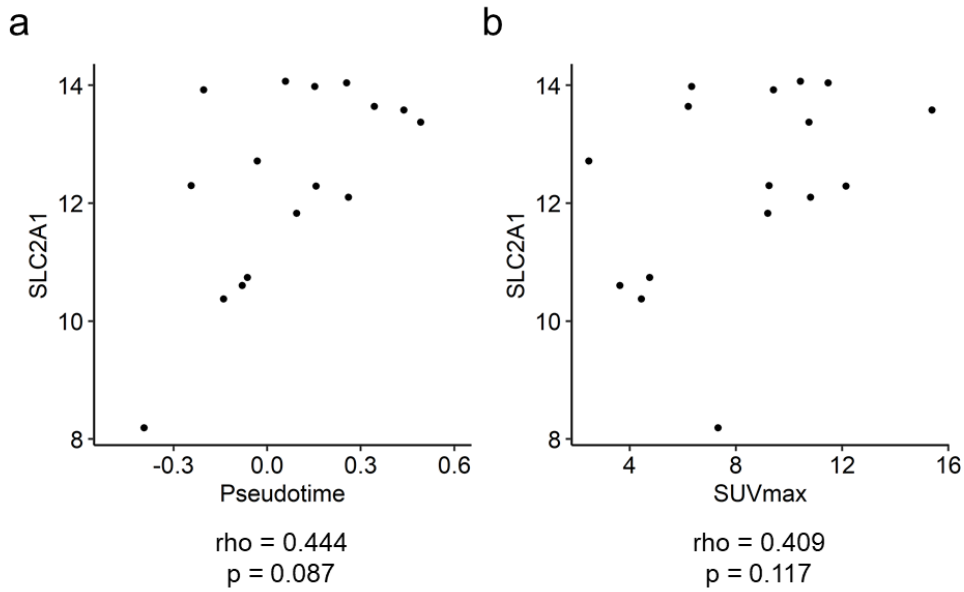

Figure 11. Scatter plots showing correlation of SLC2A1 expression with pseudotime and SUV in TCGA-LUAD dataset.

(a) Expression of SLC2A1 increased along pseudotime without statistical significance. **(b)** There was no significant correlation between SLC2A1 and maximal SUV.

Temporal evolution of glucose metabolism in LUAD samples of NSCLC radiogenomics dataset

The Pearson correlation test was performed to reveal temporal change of glucose metabolism in the NSCLC radiogenomics dataset. There was significant positive correlation between maximal SUV and pseudotime (Figure 12, $r = 0.259$, $p = 0.005$). There was significant positive correlation between mean SUV and pseudotime (Figure 12, $r = 0.227$, $p = 0.029$). However, MTV and TLG showed no association with pseudotime. There was marginally positive correlation between expression of SLC2A1 and pseudotime (Figure 13a, $r = 0.202$, $p = 0.053$). In addition, there was no significant correlation between maximal SUV and expression of SLC2A1 (Figure 13b, $r = 0.070$, $p = 0.504$).

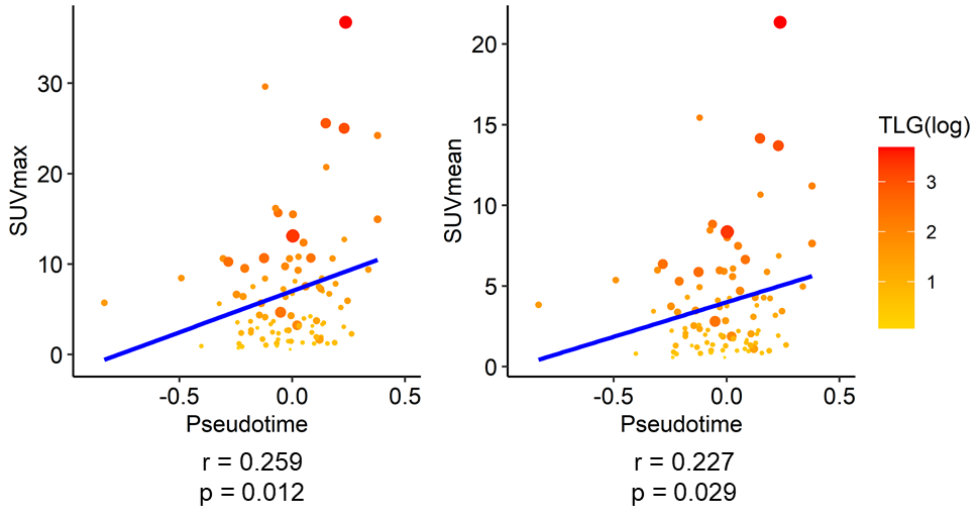

Figure 12. Scatter plots showing correlation between SUV and pseudotime in LUAD samples of NSCLC radiogenomics dataset.

A size of dot represents metabolic tumor volume. Color of dot represents total lesion glycolysis as log scale. Maximal SUV showed weakly positive correlation with pseudotime (left). Mean SUV showed weakly positive correlation with pseudotime (right).

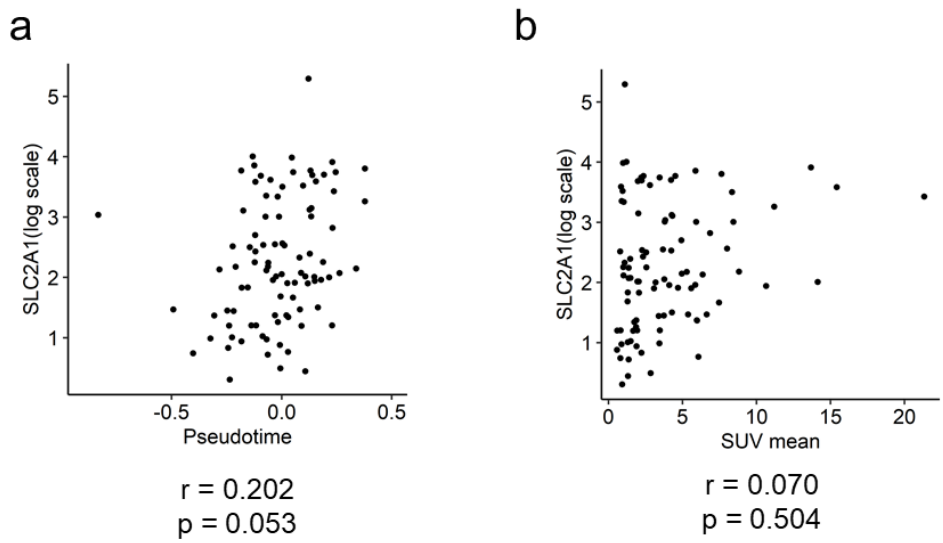

Figure 13. Scatter plots showing correlation of SLC2A1 expression with pseudotime and SUV in LUAD samples of NSCLC radiogenomics dataset.

(a) Expression of SLC2A1 showed weakly positive correlation with pseudotime despite of statistical insignificance. **(b)** Expression of SLC2A1 showed no significant correlation with maximal SUV.

Part II. Temporal evolution of immune profiles in lung adenocarcinoma

Temporal evolution of immune profiles in TCGA-LUAD dataset

A volcano plot and a heatmap demonstrate immune and stroma cells associating with pseudotime in TCGA dataset. (Figure 14, 15). Among them, type 1 helper T (Th1) cells showed positive correlation (Figure 16a, $r = 0.524$, $p < 0.001$) and M2 macrophages negative correlation (Figure 16b, $r = -0.545$, $p < 0.001$). PD-L1, the most representative immunotherapy target in lung cancer, showed weakly negative correlation with pseudotime (Figure 17a, $r = -0.289$, $p < 0.001$). TMB showed weakly positive correlation with pseudotime (Figure 17b, $r = 0.243$, $p < 0.001$).

Enrichment scores of Th1 cells, those of M2 macrophages, expression of PD-L1, and TMB were compared between groups with different driver mutations: EGFR mutation, KRAS mutation and other mutations. Although there was no difference in enrichment of M2 macrophages and expression of PD-L1, there were significant differences in enrichment of Th1 cells and TMB (Figure 18-21, $p = 0.023$ and $p = 0.011$, respectively). Th1 cells and TMB showed significantly positive correlation with pseudotime in all mutation groups (Figure 18, Figure 21).

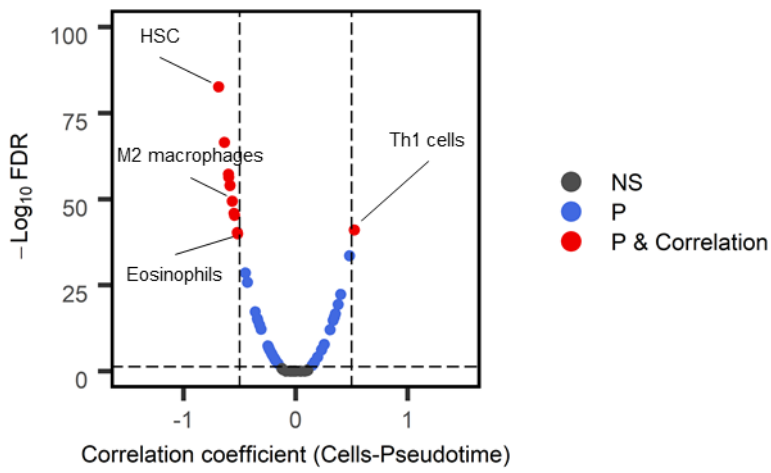

Figure 14. A volcano plot representing immune cells associating with pseudotime in TCGA-LUAD dataset.

Cell types of correlation coefficients above 0.5 were plotted as red dots. Cell types of FDR below 0.05 and correlation coefficients below 0.5 were plotted as blue dots. Among them, immune cells were annotated.

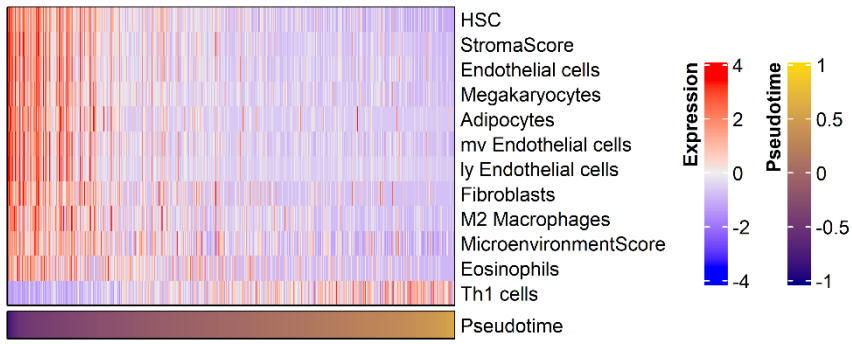

Figure 15. A heatmap representing immune and stromal cells associating with pseudotime in TCGA-LUAD dataset.

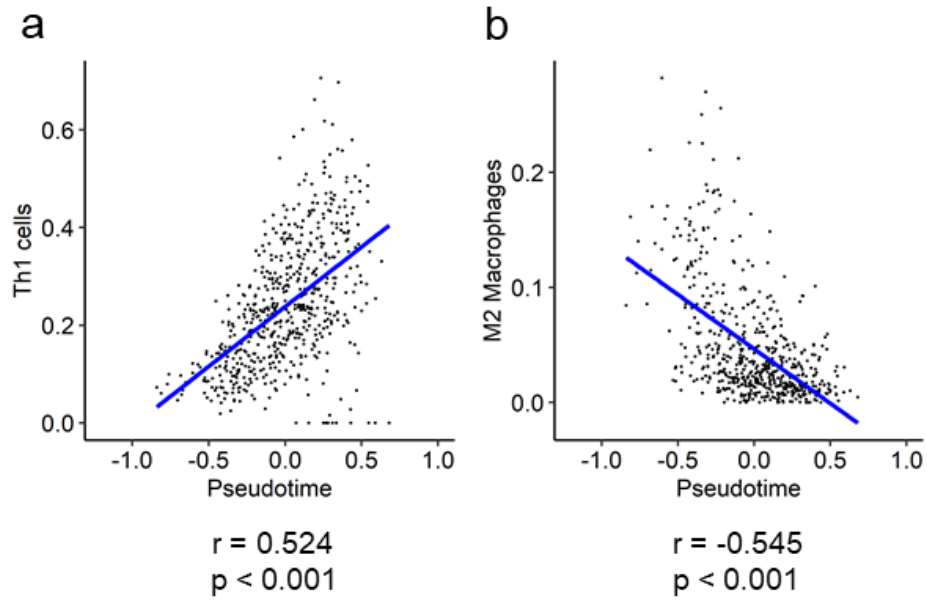

Figure 16. Scatter plots showing correlation between immune cell enrichment scores and pseudotime in TCGA-LUAD dataset.

(a) Th1 cells showed positive correlation with pseudotime. **(b)** M2 macrophages showed negative correlation with pseudotime.

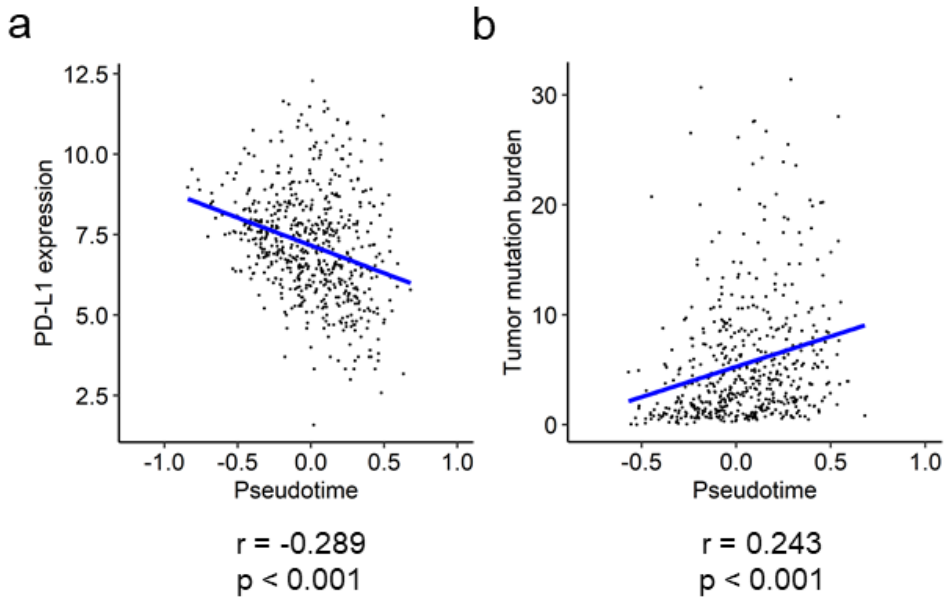

Figure 17. Scatter plots showing correlation between PD-L1 expression, TMB, and pseudotime in TCGA-LUAD dataset.

(a) PD-L1 expression showed weakly negative correlation with pseudotime. **(b)** TMB showed weakly positive correlation with pseudotime.

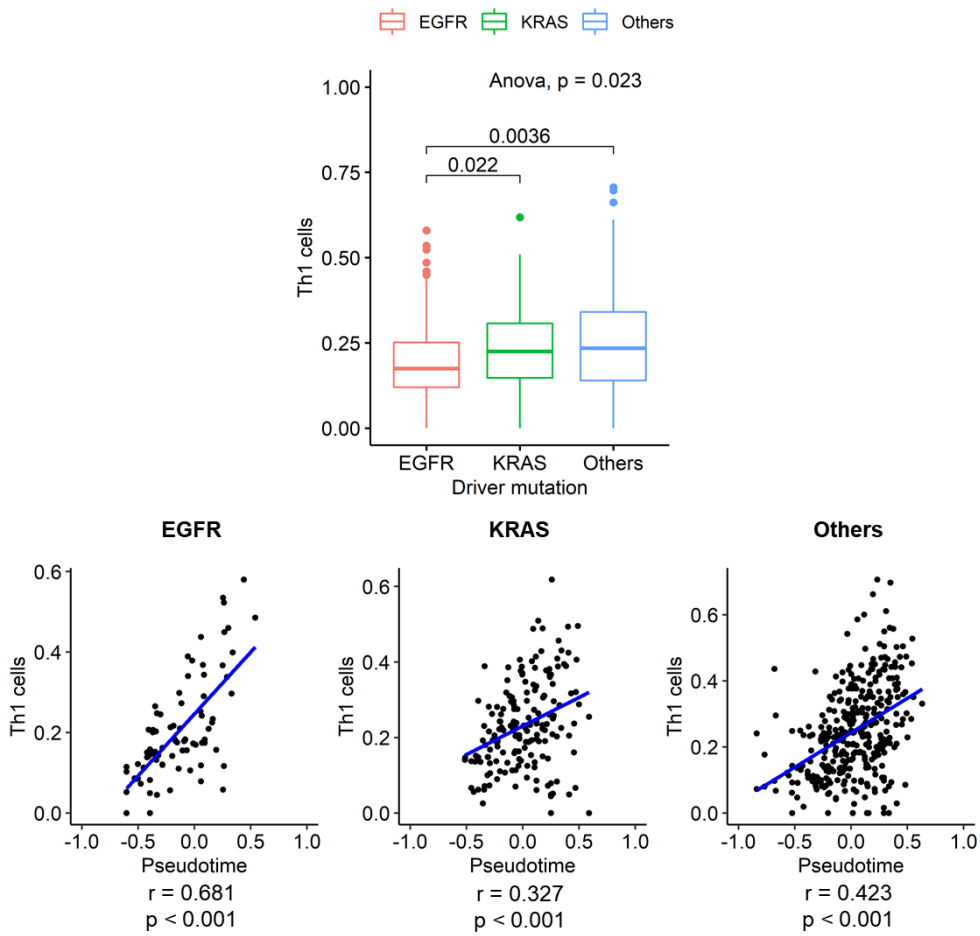

Figure 18. Difference and association with pseudotime of Th1 cell enrichment in each mutation group.

There was significant difference in enrichment of Th1 cells between different driver mutation groups (upper). In all mutation groups, enrichment of Th1 cells increased along pseudotime (lower).

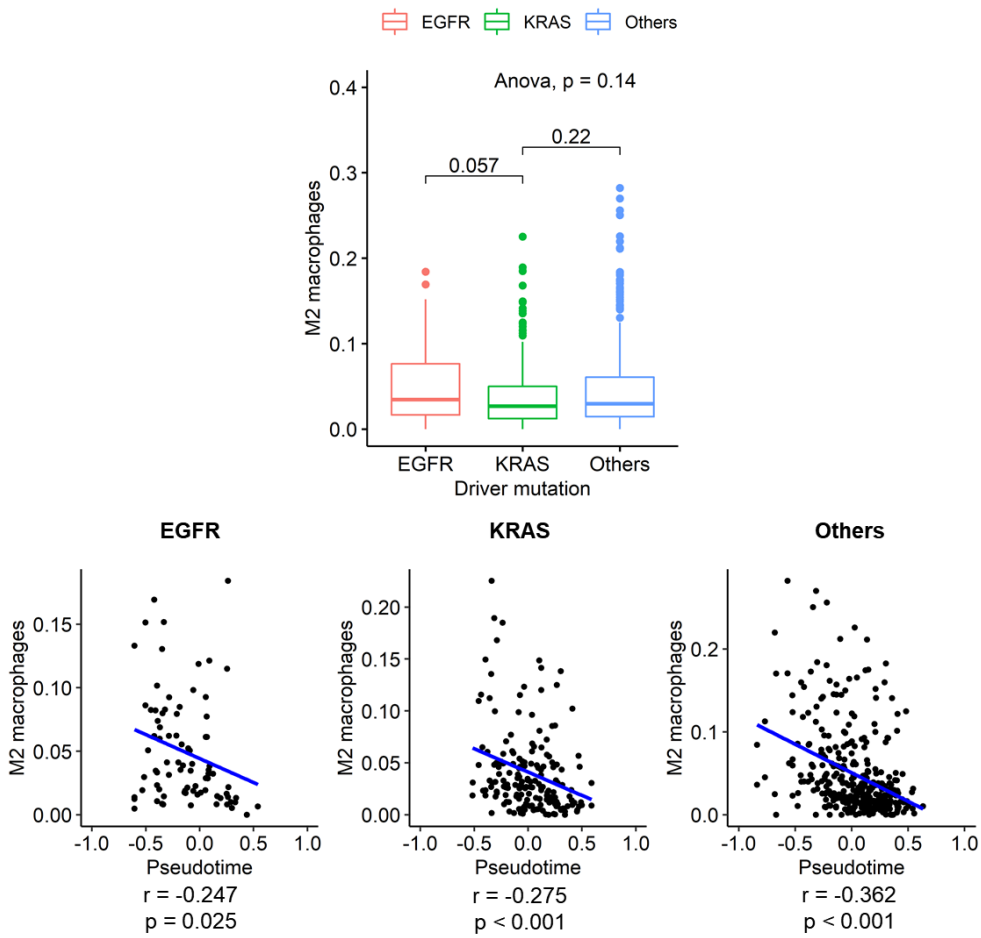

Figure 19. Difference and association with pseudotime of M2 macrophages enrichment in each mutation group.

There was no significant difference in enrichment of M2 macrophages between different driver mutation groups (upper). In all mutation groups, enrichment of M2 macrophages decreased along pseudotime (lower).

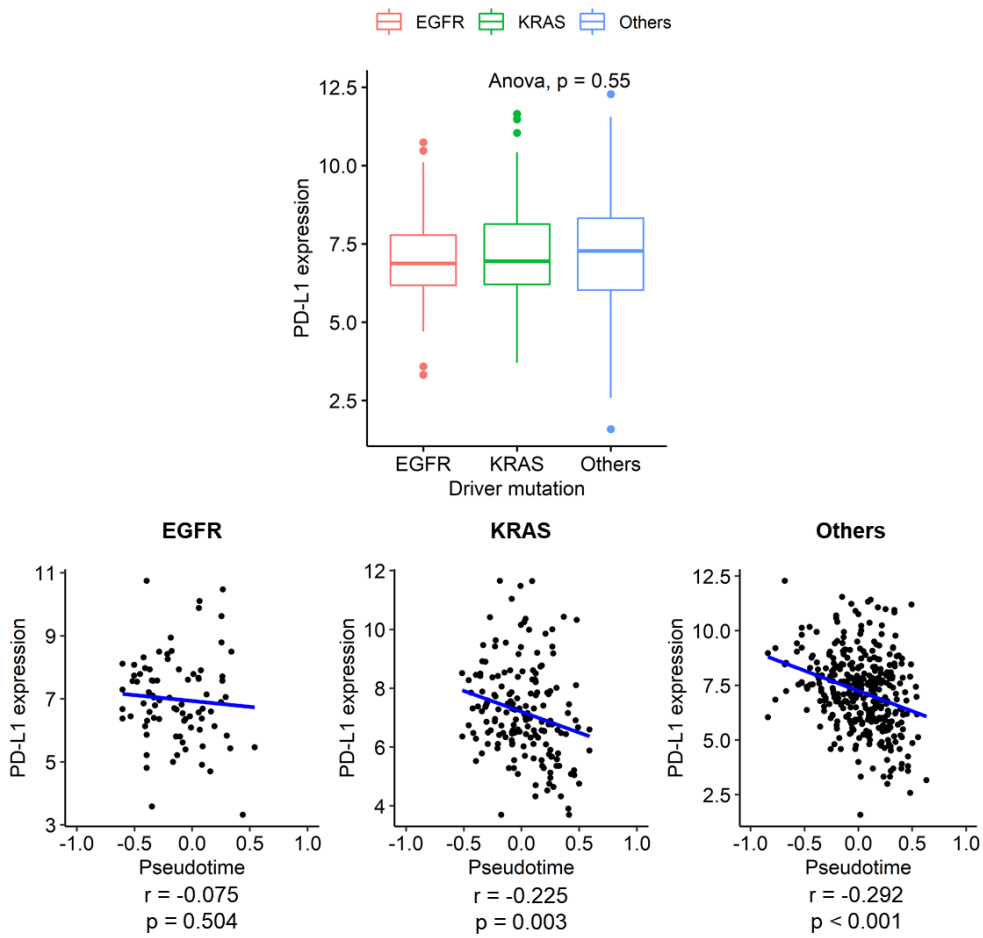

Figure 20. Difference and association with pseudotime of PD-L1 expression in each mutation group.

There was no significant difference in enrichment of PD-L1 expression between different driver mutation groups (upper). In KRAS and others mutation groups, PD-L1 expression decreased along pseudotime (lower).

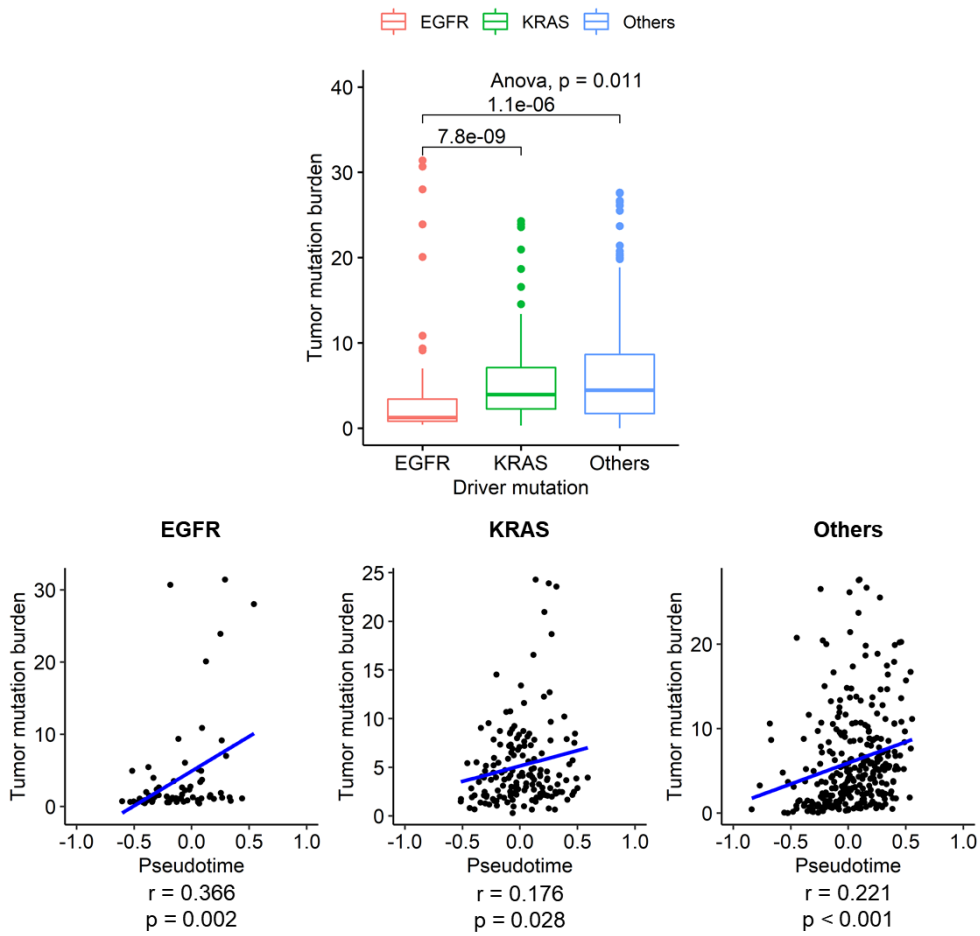

Figure 21. Difference and association with pseudotime of TMB in each mutation group.

There was significant difference in TMB between different driver mutation groups (upper). In all mutation groups, TMB increased along pseudotime (lower).

Temporal evolution of immune profiles in LUAD samples of NSCLC radiogenomics dataset

A volcano plot and a heatmap demonstrate immune and stroma cells associating with pseudotime in LUAD samples of NSCLC dataset. (Figure 22, 23). Among them, Th1 cells showed positive correlation (Figure 24a, $r = 0.444$, $p < 0.001$) and M2 macrophages negative correlation (Figure 24b, $r = -0.367$, $p < 0.001$). PD-L1 showed no significant correlation with pseudotime (Figure 25, $r = 0.041$, $p = 0.698$). There was no significant difference in enrichment scores of Th1 cells, those of M2 macrophages and expression of PD-L1 between groups with different driver mutations: EGFR mutation, KRAS mutation and other mutations.

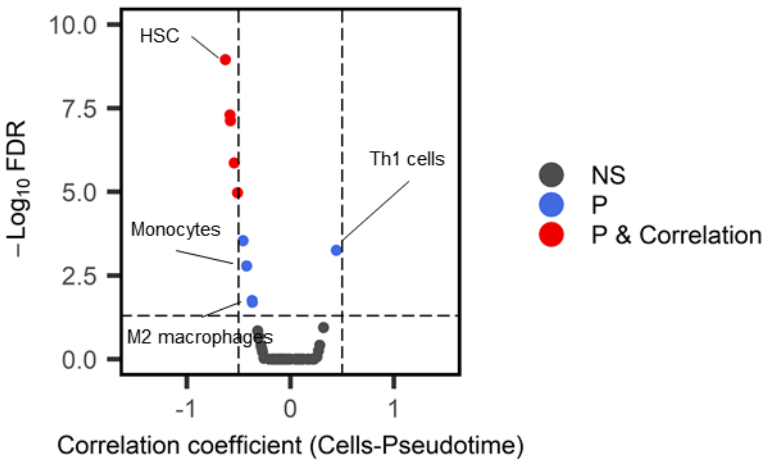

Figure 22. A volcano plot representing immune cells associating with pseudotime in LUAD samples of NSCLC radiogenomics dataset.

Cell types of correlation coefficients above 0.5 were plotted as red dots. Cell types of FDR below 0.05 and correlation coefficients below 0.5 were plotted as blue dots. Among them, immune cells were annotated.

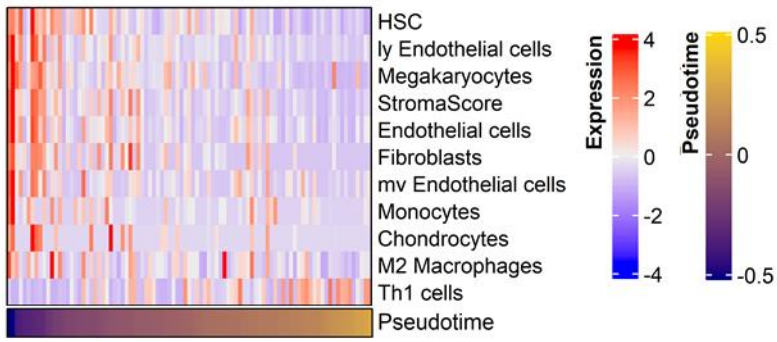

Figure 23. A heatmap representing immune and stromal cells associating with pseudotime in LUAD samples of NSCLC radiogenomics dataset.

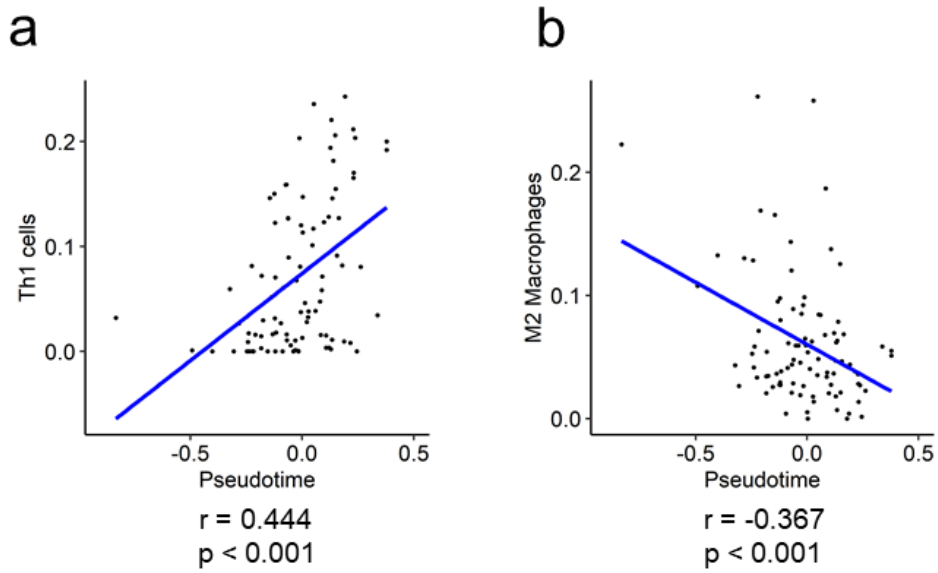

Figure 24. Scatter plots showing correlation between immune cell enrichment scores and pseudotime in LUAD samples of NSCLC radiogenomics dataset.

(a) Th1 cells showed positive correlation with pseudotime. **(b)** M2 macrophages showed negative correlation with pseudotime.

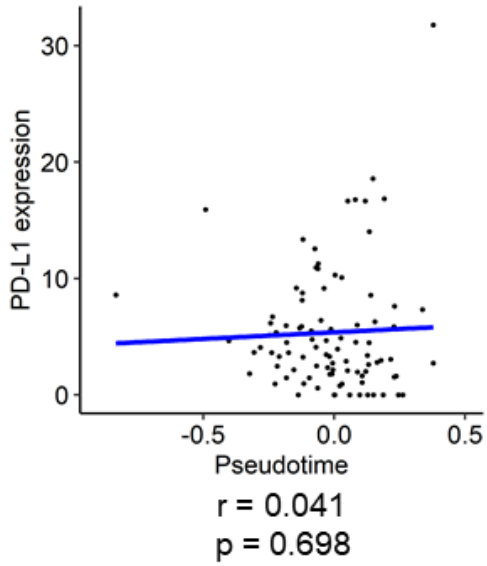

Figure 25. A scatter plot showing correlation between PD-L1 expression and pseudotime in LUAD samples of NSCLC radiogenomics dataset.

PD-L1 expression showed no significant correlation with pseudotime.

Part III. Immunological factors associating with lymph node metastasis and glucose metabolism

TREM-1 associating with lymph node metastasis

In TCGA-LUAD dataset, there was no significant difference in possibility of lymph node metastasis between early pseudotime group and late pseudotime group (Figure 26, $p = 0.066$). Volcano plots represent DEGs between samples without LN metastasis and those with LN metastasis in early pseudotime group and late pseudotime group, respectively (Figure 27). Notably, triggering receptor expressed on myeloid cells-1 (TREM-1) was highly expressed in samples with LN metastasis in early pseudotime group. Among GO terms, acute inflammatory response and humoral immune response were selected as significantly upregulated functions in samples with LN metastasis in early pseudotime group (Figure 28). In early pseudotime group, monocytes were enriched higher in samples with LN metastasis than those without LN metastasis (Figure 29a, $W = 7184.5$, $p = 0.008$). There was tendency of high macrophages and M2 macrophages in samples with LN metastasis despite statistical significance (Figure 29a, $W = 7657$, $p = 0.051$; $W = 7756$, $p = 0.071$; respectively.). In late pseudotime group, CD4+ T cells, CD8+ T cells, and B cells were enriched higher in samples without LN metastasis than those with LN metastasis (Figure 29b, $W = 12321$, $p < 0.001$; $W = 11581.5$, $p = 0.015$; $W = 11831.5$, $p = 0.005$; respectively).

Expression of TREM-1 was compared between samples with and without LN metastasis in early pseudotime group. Additionally, gene set enrichment scores for monocytic myeloid-derived stem cells were compared. Both of those were higher in samples with LN metastasis than those without (Figure 30, mean: 11.28313 vs. 10.45068, $p < 0.001$; mean: 0.306 vs. 0.158, $p = 0.039$).

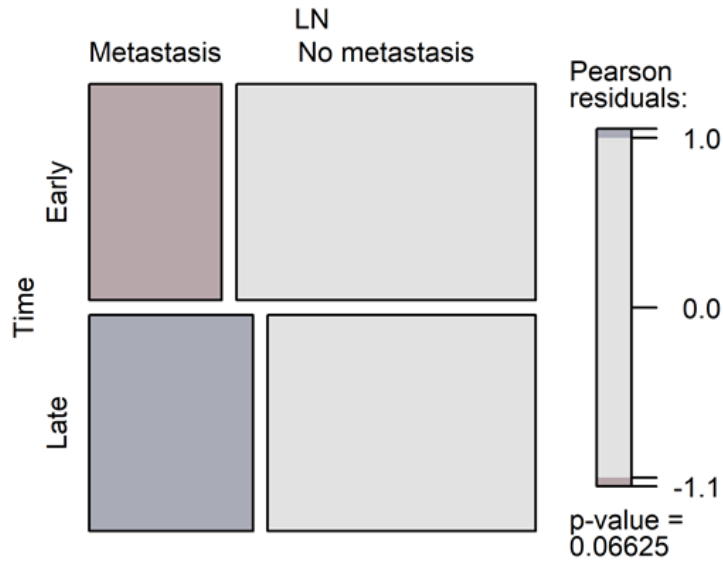

Figure 26. A mosaic plot representing difference of subjects with lymph node metastasis in early pseudotime and late pseudotime groups.

There was no significant difference in possibility of lymph node metastasis between early pseudotime group and late pseudotime group in TCGA-LUAD dataset.

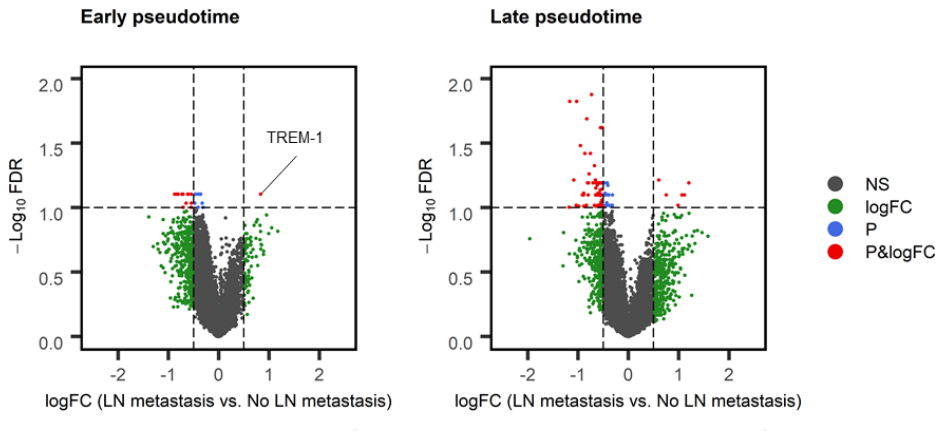

Figure 27. Volcano plots representing DEGs between samples without lymph node metastasis and those with lymph node metastasis.

DEGs were plotted with red dots in early pseudotime group (left) and late pseudotime group (right). TREM-1 was highly expressed in samples with lymph node metastasis in early pseudotime group.

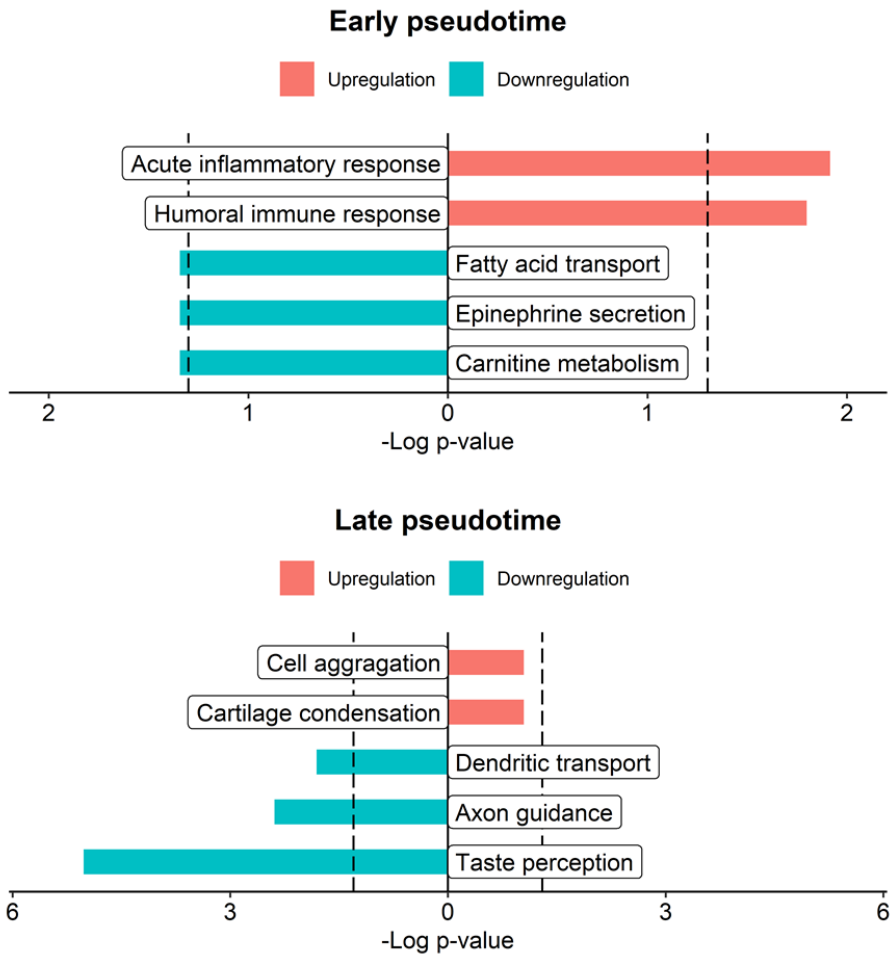

Figure 28. Gene ontology analysis with DEGs between samples without lymph node metastasis and those with lymph node metastasis.

Acute inflammatory response and humoral immune response were selected as significantly upregulated functions in samples with LN metastasis in early pseudotime group.

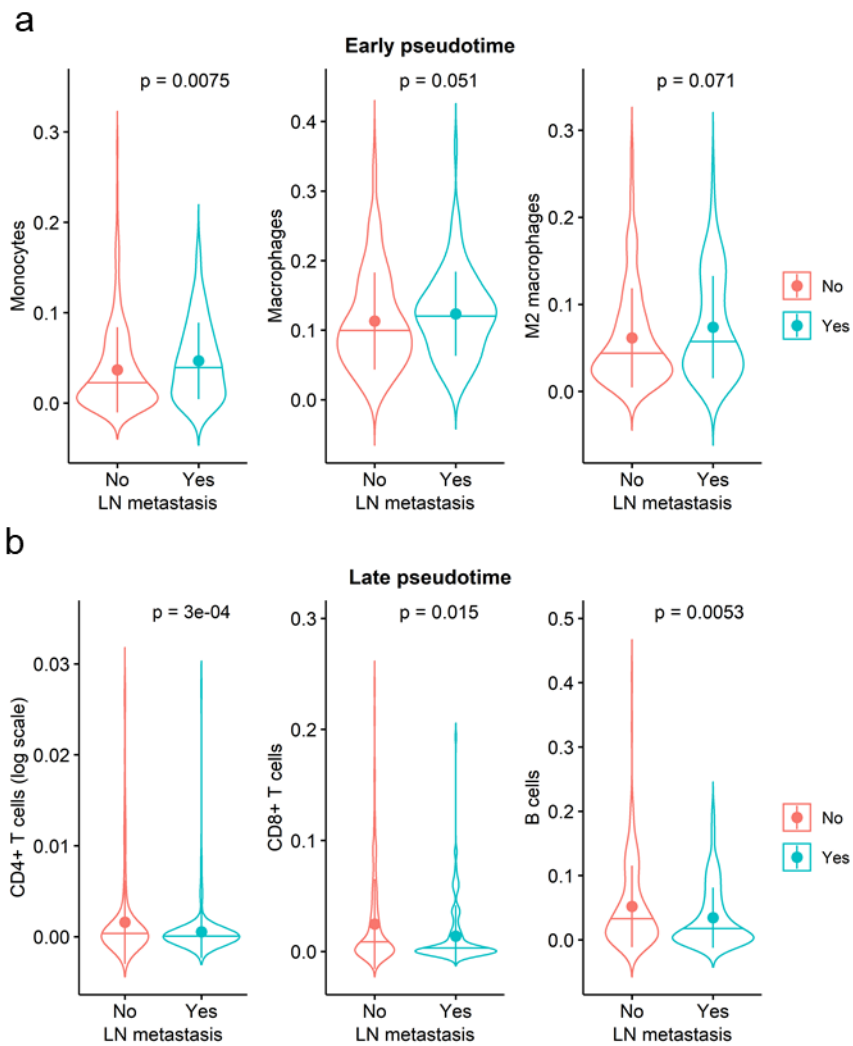

Figure 29. Violin plots representing difference of immune cell enrichment scores.

(a) In early pseudotime group, monocytes were enriched higher in samples with lymph node metastasis than those without lymph node metastasis. **(b)** In late pseudotime group, CD4+ T cells, CD8+ T cells, and B cells were enriched higher in samples without lymph node metastasis than those with lymph node metastasis.

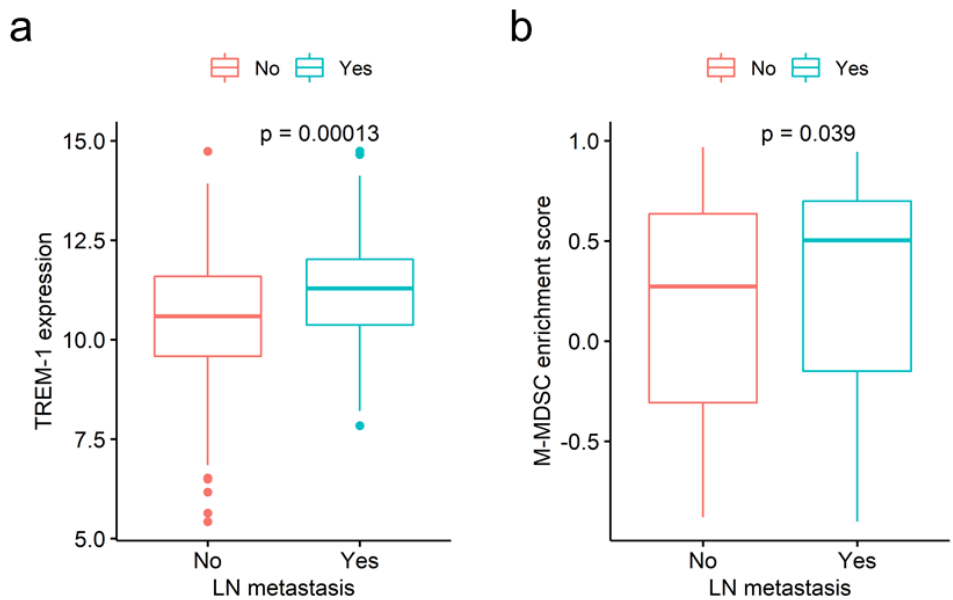

Figure 30. Differences of TREM-1 expression and monocytic MDSC enrichment.

(a) In early pseudotime group, TREM-1 expression was higher in samples with lymph node metastasis than those without lymph node metastasis. **(b)** In early pseudotime group, monocytic MDSC were enriched higher in samples with lymph node metastasis than those without lymph node metastasis.

Mononuclear cell associating with glucose metabolism

In LUAD samples of NSCLC radiogenomics dataset, there were samples with high FDG uptake in the early pseudotime and samples with low FDG uptake in the late pseudotime (Figure 31). In early pseudotime group, there was lesser subjects with EGFR mutation than those with wildtype EGFR in low FDG uptake group despite statistical insignificance (Figure 32a, $p = 0.079$). On the contrary, there was no differences in proportion of subjects with EGFR mutation between high FDG uptake group and low FDG uptake group in late pseudotime group (Figure 32b, $p = 0.817$). In early pseudotime group, a proportion of subjects with grade 3 was higher in high FDG uptake group than in low FDG uptake group (Figure 33a, $p = 0.050$). In late pseudotime group, a proportion of subjects with grade 3 was higher in high FDG uptake group and that with grade 1 was higher in low FDG uptake group (Figure 33b, $p = 0.016$). Among GO terms, negative regulation of keratinocyte proliferation was selected as significantly downregulated function in samples with high FDG uptake in early pseudotime group. Cell cycle arrest and negative regulation of mononuclear cell migration were selected as significantly downregulated function in samples with high FDG uptake in late pseudotime group (Figure 34). Enrichment scores of monocytes and Th1 cells were higher in those with FDG uptake than those with low FDG uptake in early and late pseudotime group, respectively (Figure 35, $W = 378$, $p = 0.012$; $W = 379$, $p = 0.028$, respectively). In addition, enrichment scores of Th1 cells showed positive correlation with maximal SUV (Figure 36a, $r = 0.442$, $p < 0.001$).

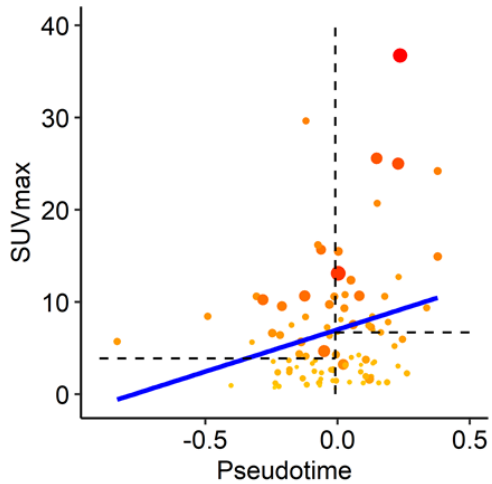

Figure 31. A scatter plot showing weak positive correlation between SUV and pseudotime in LUAD samples of NSCLC radiogenomics dataset.

There were samples with high FDG uptake in the early pseudotime (the left upper quadrant of scatter plot) and samples with low FDG uptake in the late pseudotime (the right lower quadrant of scatter plot). The vertical dashed line is a median value of pseudotime. The horizontal dashed line is a median value of maximal SUV in early and late pseudotime group.

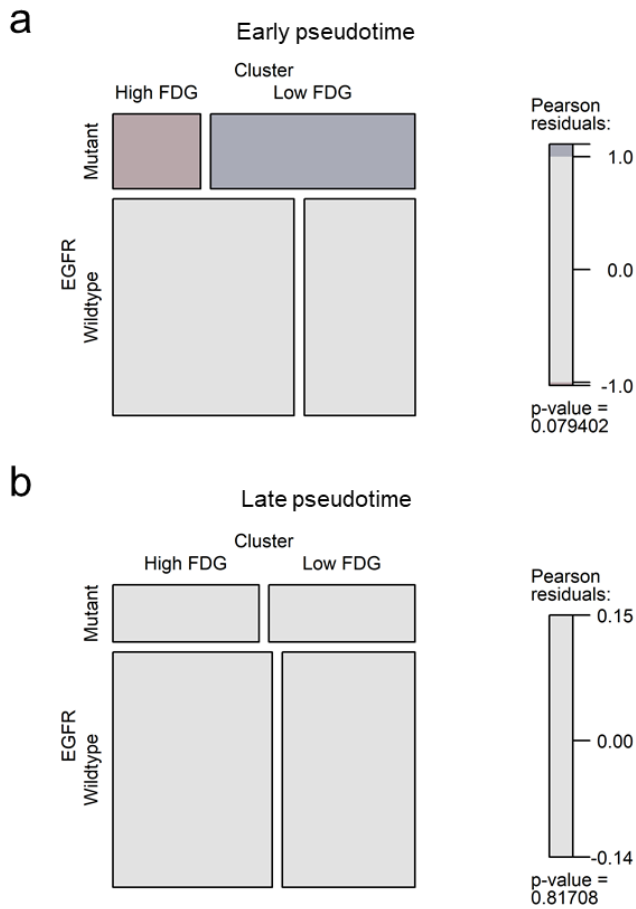

Figure 32. Mosaic plots representing difference of subjects with EGFR mutation according to FDG uptake.

(a) In early pseudotime group, there was lesser subjects with EGFR mutation than those with wildtype EGFR in low FDG uptake group despite statistical insignificance. **(b)** In late pseudotime group, there was no differences in proportion of subjects with EGFR mutation between high FDG uptake group and low FDG uptake group.

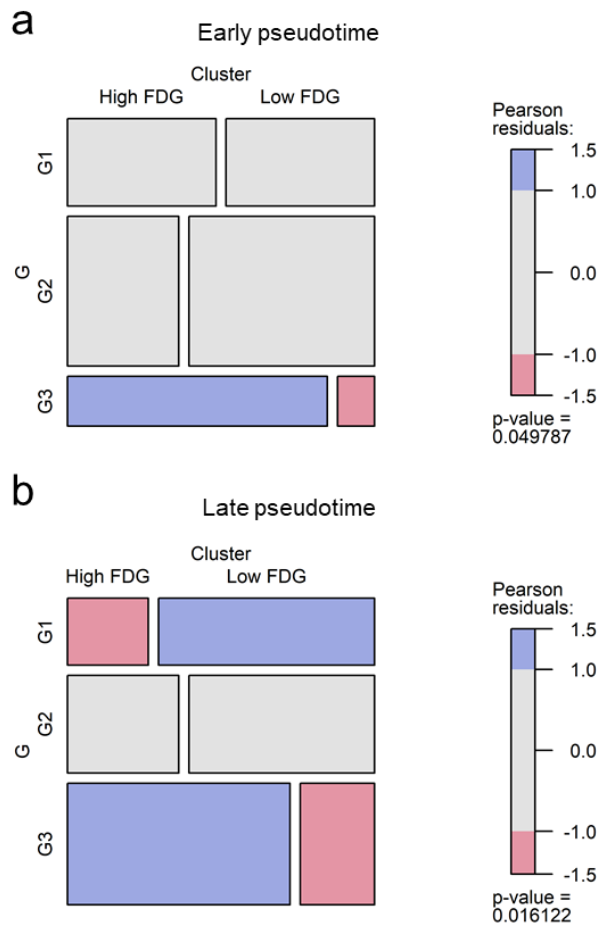

Figure 33. Mosaic plots representing difference of subjects with different histological grade according to FDG uptake.

(a) In early pseudotime group, a proportion of subjects with G3 was higher in high FDG uptake group than in low FDG uptake group. **(b)** In late pseudotime group, a proportion of subjects with G3 was higher in high FDG uptake group and that with G1 was higher in low FDG uptake group.

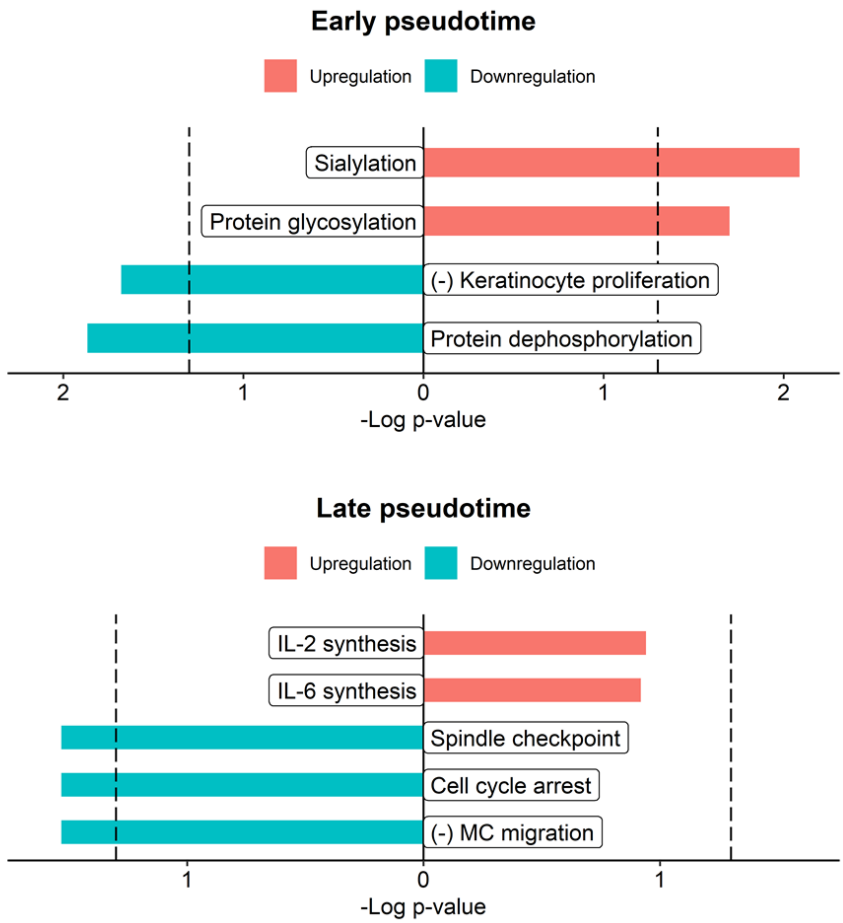

Figure 34. Gene ontology analysis with DEGs between samples with high FDG uptake and those with low FDG uptake.

Negative regulation of keratinocyte proliferation was selected as significantly downregulated function in samples with high FDG uptake in early pseudotime group (upper). Cell cycle arrest and negative regulation of mononuclear cell migration were selected as significantly downregulated functions in samples with high FDG uptake in late pseudotime group (lower).

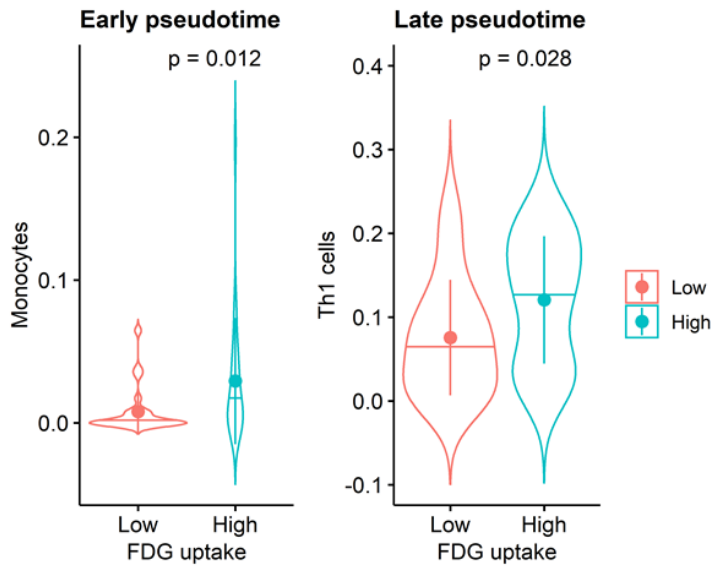

Figure 35. Violin plots representing difference of immune cells according to FDG uptake. Enrichment scores of monocytes were higher in samples with FDG uptake than those with low FDG uptake in early pseudotime (left). Enrichment scores of Th1 were higher in samples with FDG uptake than those with low FDG uptake in early pseudotime (right).

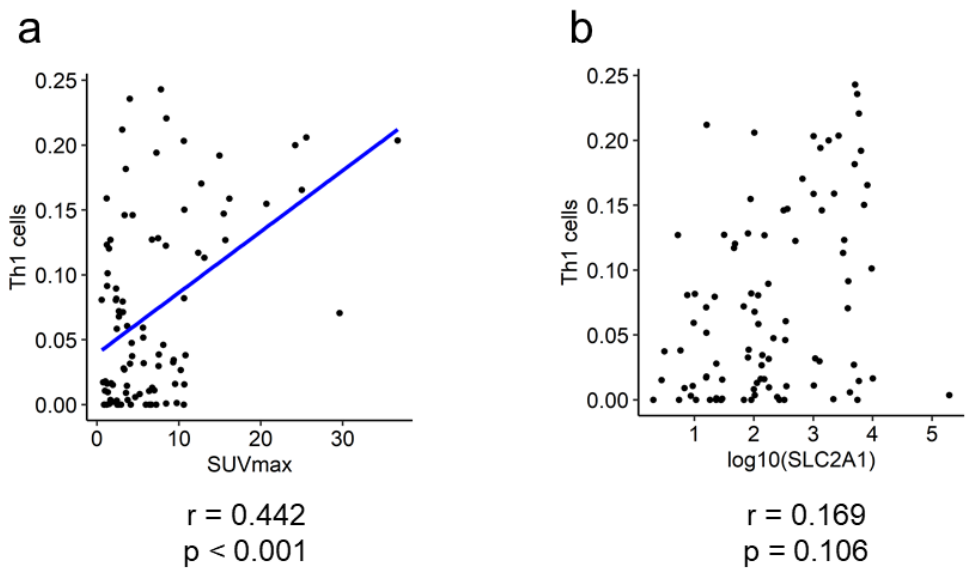

Figure 36. Scatter plots showing correlation of Th1 cell enrichment with pseudotime and SLC2A1 expression in LUAD sample of NSCLC radiogenomics dataset.

(a) Enrichment of Th1 cells showed positive correlation with maximal SUV. **(b)** Enrichment of Th1 cells showed no significant correlation with expression of SLC2A1.

Discussion

The TNM staging in lung cancer is a well-established system to evaluate disease progression status, predict prognosis, and select appropriate treatment options (32, 33). However, it is the result from cross-sectional observation via clinical / pathologic / radiologic findings at a timing of initial diagnosis. Therefore, there is limitation to investigate temporal evolution of tumor biology longitudinally based on TNM staging as a reference scale. We attempted to construct a temporal model for the biological progression from genetic profiles of a large-scale dataset. Based on the generated model, we interrogated temporal evolution of genetic features, clinical features, glucose metabolism, and immune profiles in lung adenocarcinoma.

In this study, we successfully estimated a pseudotime trajectory in TCGA-LUAD and TCGA-LUSC datasets. In PCA, LUAD and LUSC samples in early phase are revealed to have similar genetic characteristics and differentiate into LUAD and LUSC along pseudotime order. In tumorigenesis of NSCLC, molecular events such as 3p allele loss and telomerase activation are observed in most of NSCLC (34-36). Similarity of genetic characteristics in early LUAD and LUSC may be caused by the common pathogenesis mechanisms. The result implies that tumor showing specific characteristics of LUAD or LUSC has high possibility of progressed state.

In both of total lung cancer and LUAD samples, GO terms related to cell division were selected as significant upregulated molecular functions along pseudotime. Especially, histone-related genes showed high correlation with pseudotime in LUAD. These can be interpreted as the results from either large number of tumor cells or enhanced mitotic activity of tumor cell in late phase. In the same tendency, T stage and overall TNM stage demonstrated good association with pseudotime. It is well consistent with the current TNM staging system. Notably, there were significant associations in the early T stages (T1-T2, T2-T3) and early

overall stages (IA-IB, IA-IIB, IA-IIIA). In the current TNM staging system, T2-T4 stages include not only size criteria but also criteria of involving other structures such as bronchus or chest wall (37). Thus, a small size tumor with involvement of other structures can be diagnosed as high T stage. If there are lymph node metastases, it is highly likely to be classified as above stage IIIA. These characteristics of the current TNM staging system cause associations between early T-stage/overall stage and pseudotime. Probability of OS showed significant difference between early pseudotime and late pseudotime samples. It indicates that pseudotime may have clinical usability to classify patients based on prognosis as TNM staging. Of course, further study is warranted to explore clinical significance of pseudotime trajectory. In both of two datasets, maximal SUV and mean SUV demonstrated significantly positive correlation with pseudotime. These findings are consistent with previously revealed relationship between FDG uptake and stage of tumor (8, 9, 38). In brief, change of genetic features, clinical features, and glucose metabolism along pseudotime was revealed to be consistent with previous knowledges about tumor progression. Therefore, estimated pseudotime was hypothesized to be an appropriate temporal reference of disease progression.

The present study demonstrated temporal evolution of immune profiles in LUAD. It is noteworthy that enrichment score of Th1 cells represented significantly positive correlation with pseudotime. It is generally believed that Th1 cells contribute to anti-tumor response inducing cytotoxicity (39, 40). It is also remarkable that M2 macrophages showed significantly negative correlation with pseudotime. M2 macrophages exert pro-tumor activity via tissue remodeling and angiogenesis (41, 42). Briefly, anti-tumor immunity seems to strengthen along pseudotime, whereas pro-tumor immunity seems to weaken along pseudotime. These results are also consistent with a previous report documenting that proportion of high stage was larger in samples with high immune score and cytolytic score (43). In addition, NK cell function was downregulated along pseudotime. Although NK cells are known to be a good prognostic factor in lung cancer, a few studies showed inconsistent

results (44-47). Among them, Jules et al. suggested that intratumoral NK cells negatively affect the prognostic value of CD8+ T cells (47). Regarding the tendency of lesser Th1 and more M2 macrophages along pseudotime, this study supports an inhibitory role of intratumoral NK cells for anti-tumor immunity. Interestingly, PD-L1 expression showed a negative correlation with pseudotime, whereas TMB showed positive correlation in TCGA-LUAD dataset. Those are well-known biomarkers predicting response for cancer immunotherapy (29, 48). This heterogenous finding implies that response of immunotherapy may not represent any tendency according to molecular progression of lung adenocarcinoma.

Driver mutations are key factors not only to initiate tumorigenesis but also to progress disease. EGFR and KRAS mutations are the most representative driver mutations of LUAD (49). It was revealed that TMB was lower in EGFR mutation group than in other mutation groups. It is consistent with other studies that showed weak immunogenicity or low TMB of tumors with EGFR mutation (50, 51). This study showed that immunogenicity of tumor increases along disease progression regardless of driver mutations, despite low Th1 cell infiltration and TMB in EGFR mutation tumors.

In this study, there was no significant correlation between N stage and pseudotime. There was no significant difference in proportion of LN metastasis subjects between in early time group and late time group, although that in late time group was higher than that in early time group. It suggests that there is no inevitable association between molecular progression and lymph node metastasis. In clinical field, there is tendency that patients with high T stage show lymph node metastasis. However, it is not unique that patients with low T stage show multiple lymph nodes or distant metastases or patients with high T stage show no metastasis. These experiences are related to the result of this study. In early pseudotime group, TREM-1 was selected as a DEG between subjects with LN metastasis and those without LN metastasis. TREM-1 is one of receptor proteins expressed in myeloid cells including monocytes and neutrophils. It is reported to induce inflammation through pro-inflammatory gene

upregulation (52). Thus, GO terms related to inflammation were selected as significant functions upregulated in those with LN metastasis. Especially in lung cancer, TREM-1 is expressed on tumor-associated macrophages, increases invasiveness of tumor cells, and associates with poor prognosis (53). Notably, monocytes were highly enriched in samples with lymph node metastasis in early pseudotime group. TREM-1 is known to be expressed on MDSC (54). Monocytic MDSC enables immune escape and promotes tumor progression (55, 56). This study also showed that expression of TREM-1 and enrichment of monocytic MDSC were higher in samples with LN metastasis than those without, in early pseudotime. Briefly, the present study implies that immune suppression may affect lymph node metastasis in early phase of lung adenocarcinoma. In late pseudotime group, CD4+ T cells and CD8+ T cells were lowly enriched in samples with lymph node metastasis. It is supposed that reduced anti-tumor immunity of T cells may affect lymph node metastasis in late phase of lung cancer progression.

Although positive correlation between pseudotime and maximal SUV was identified, there were some heterogenous findings in NSCLC radiogenomics dataset. There were samples with relatively higher FDG uptake with maximal SUV even over 10 in early phase of lung adenocarcinoma, and samples with relatively lower FDG uptake with maximal uptake even below 3 in late phase. Clinicians often experience lung adenocarcinoma cases with FDG PET/CT findings inconsistent with general knowledge, as described above. In subgroup analyses, we explored factors associating with FDG uptake in early pseudotime group and late pseudotime group, respectively. Subjects with EGFR mutation represented low FDG uptake in early pseudotime group despite statistical insignificance. In several studies, EGFR mutants showed lower FDG uptake than EGFR wild-type due to decreased glucose metabolism (57-59). This study suggests that EGFR mutation may associate with glucose metabolism especially in early phase of lung adenocarcinoma. FDG uptake of low-grade tumor was lower than that of high-grade tumor in late pseudotime group. It may be

caused not only by high FDG uptake of poorly differentiated lung cancer but also by positive correlation between histological grade and pseudotime (38, 60).

GO analyses showed which function associates with FDG uptake in early and late pseudotime groups. In early pseudotime group, keratinocyte proliferation was revealed to associate with FDG uptake. First, it is closely linked with previously described results, larger proportion of EGFR wild type in high FDG uptake group. EGFR induces cell proliferation and accelerates tumor growth (61, 62). Thus, cell proliferative activity seems to affect FDG uptake in the early phase of lung adenocarcinoma. Secondly, it may be interpretated with results from principal component analysis. We demonstrated early phase of LUAD and LUSC shared similar genetic characteristics, implying that early stage of LUAD may show a part of squamous features. In terms that LUSC shows higher FDG uptake than LUAD, LUAD with squamous cell features is supposed to demonstrate high FDG uptake (38, 63). On the contrary, mononuclear cell migration was revealed to associate with FDG uptake in late pseudotime group. In addition, it was pointed out that Th1 cells associated with FDG uptake among immune cells infiltrated in tumor tissue. It is consistent with a previous study that Th1 cell enrichment is higher in tumor with high FDG uptake in head and neck cancer patients (64). First, high glucose metabolism may be needed for proliferation and activation of Th1 cells. Secondly, high tumor burden or enhanced metabolism of tumor may affect immunogenicity inducing Th1 cell infiltration. In short, Th1 cell is a remarkable immune cell to potentially affect glucose metabolism in late phase of lung adenocarcinoma.

There are some limitations in this study. First, FDG PET examination of subjects were performed in different institutes so that there were differences in image acquisition and reconstruction methods. However, the purpose of analyzing association between SUV and pseudotime was not to predict accurate SUV or pseudotime but to assess overall tendency of SUV along pseudotime. Furthermore, the image acquisition protocol of each sample was not identified in the obtained clinical data. Therefore, all the data were included in a single

correlation study. Further study is warranted to analyze evolution of glucose metabolism along pseudotime more accurately using FDG PET image data of same institute. Second, this study includes only RNA-seq data from primary tumor tissue of different subjects. Data of recurrent or metastatic tumor were not included in the datasets employed in this study. Although tumor characteristics of recurrent or metastatic tissue are not equal to naturally progressed malignancy due to treatment, comparison of tumor characteristics between primary tumor and recurrent/metastatic tumor in same subject can provide additional information of disease progression. Further study to explore change of tumor characteristics in same patient can support the results of this study. Third, pseudotime trajectory from RNA-seq has a limitation to apply to clinical field due to complexity of obtaining tumor tissue and analyzing transcriptomic data from each patient. To facilitate application of pseudotime in clinical situations, further study is underway to construct pseudotime trajectory from FDG PET images.

Conclusions

Taken together, pseudotime trajectories were successfully estimated in lung adenocarcinoma subjects from TCGA dataset and NSCLC radiogenomics dataset. It shows fair correlation with TNM stage, clinical outcome, and glucose metabolism, suggesting feasibility of new scale evaluating disease progression status. Th1 cells and M2 macrophages showed positive and negative correlation with pseudotime, respectively. TREM-1 was pointed out as a gene associating with lymph node metastasis in early phase of disease. Mononuclear cell migration was revealed to associate with glucose metabolism of tumor in late phase of disease. The present study demonstrated evolution of tumor biology based on transcriptomic profiles.

References

1. Barta JA, Powell CA, Wisnivesky JP. Global epidemiology of lung cancer. *Ann Glob Health.* 2019;85(1):8.

2. Lee JG, Kim HC, Choi C-M. Recent trend of lung cancer in Korea. *Tuberc Respir Dis (Seoul).* 2021;84(2):89-95.

3. Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. *CA Cancer J Clin.* 2020;70(1):7-30.

4. Jung K-W, Won Y-J, Hong S, Kong H-J, Lee ES. Prediction of cancer incidence and mortality in Korea, 2020. *Cancer Res Treat.* 2020;52(2):351-8.

5. Mountain CF, Carr DT, Anderson W. A system for the clinical staging of lung cancer. *Am J Roentgenol Radium Ther Nucl Med.* 1974;120(1):130-8.

6. Shim SS, Lee KS, Kim B-T, Chung MJ, Lee EJ, Han J, et al. Non–small cell lung cancer: prospective comparison of integrated FDG PET/CT and CT alone for preoperative staging. *Radiology.* 2005;236(3):1011-9.

7. Hicks RJ. Role of 18F-FDG PET in assessment of response in non–small cell lung cancer. *J Nucl Med.* 2009;50(Suppl 1):31S-42S.

8. Sunnetcioglu A, Arısoy A, Demir Y, Ekin S, Dogan E. Associations between the standardized uptake value of 18F-FDG PET/CT and demographic, clinical, pathological, radiological factors in lung cancer. *Int J Clin Exp Med.* 2015;8(9):15794-800.

9. Li M, Sun Y, Liu Y, Han A, Zhao S, Ma L, et al. Relationship between primary lesion FDG uptake and clinical stage at PET–CT for non-small cell lung cancer patients: An observation. *Lung Cancer.* 2010;68(3):394-7.

10. Quatromoni JG, Eruslanov E. Tumor-associated macrophages: function, phenotype, and link to prognosis in human lung cancer. *Am J Transl Res.* 2012;4(4):376-89.

11. Li J, Li X, Zhang C, Zhang C, Wang H. A signature of tumor immune microenvironment genes associated with the prognosis of non-small cell lung cancer. *Oncol Rep.* 2020;43(3):795-806.

12. Su D, Wu G, Xiong R, Sun X, Xu M, Mei Y, et al. Tumor Immune Microenvironment Characteristics and Their Prognostic Value in Non-Small-Cell Lung Cancer. *Front Oncol.* 2021;11:634059.

13. Pardoll DM. The blockade of immune checkpoints in cancer immunotherapy. *Nat Rev Cancer.* 2012;12(4):252-64.

14. Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the treatment of non–small-cell lung cancer. *N Engl J Med.* 2015;372(21):2018-28.

15. Binnewies M, Roberts EW, Kersten K, Chan V, Fearon DF, Merad M, et al. Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat Med.* 2018;24(5):541-50.

16. Taube JM, Galon J, Sholl LM, Rodig SJ, Cottrell TR, Giraldo NA, et al. Implications of the tumor immune microenvironment for staging and therapeutics. *Mod Pathol.* 2018;31(2):214-34.

17. Zeni E, Mazzetti L, Miotto D, Cascio NL, Maestrelli P, Querzoli P, et al. Macrophage expression of interleukin-10 is a prognostic factor in nonsmall cell lung cancer. *Eur Respir J.* 2007;30(4):627-32.

18. Okita R, Maeda A, Shimizu K, Nojima Y, Saisho S, Nakata M. PD-L1 overexpression is partially regulated by EGFR/HER2 signaling and associated with poor prognosis in patients with non-small-cell lung cancer. *Cancer Immunol Immunother.* 2017;66(7):865-76.

19. Trapnell C. Defining cell types and states with single-cell genomics. *Genome Res.* 2015;25(10):1491-8.

20. Campbell KR, Yau C. Uncovering pseudotemporal trajectories with covariates from single cell and bulk expression data. *Nat Commun.* 2018;9(1):2442.

21. Saelens W, Cannoodt R, Todorov H, Saeys Y. A comparison of single-cell trajectory inference methods. *Nat Biotechnol.* 2019;37(5):547-54.

22. Song Q, Hawkins GA, Wudel L, Chou PC, Forbes E, Pullikuth AK, et al. Dissecting intratumoral myeloid cell plasticity by single cell RNA-seq. *Cancer Med.* 2019;8(6):3072-85.

23. Kim N, Kim HK, Lee K, Hong Y, Cho JH, Choi JW, et al. Single-cell RNA sequencing demonstrates the molecular and cellular reprogramming of metastatic lung adenocarcinoma. *Nat Commun.* 2020;11(1):2285.

24. Bakr S, Gevaert O, Echegaray S, Ayers K, Zhou M, Shafiq M, et al. A radiogenomic dataset of non-small cell lung cancer. *Sci Data.* 2018;5:180202.

25. Nioche C, Orlhac F, Boughdad S, Reuzé S, Goya-Outi J, Robert C, et al. LIFEx: a freeware for radiomic feature calculation in multimodality imaging to accelerate advances in the characterization of tumor heterogeneity. *Cancer Res.* 2018;78(16):4786-9.

26. Nestle U, Kremp S, Schaefer-Schuler A, Sebastian-Welsch C, Hellwig D, Rübe C, et al. Comparison of different methods for delineation of 18F-FDG PET–positive tissue for target volume definition in radiotherapy of patients with non–small cell lung cancer. *J Nucl Med.* 2005;46(8):1342-8.

27. Aran D, Hu Z, Butte AJ. xCell: digitally portraying the tissue cellular heterogeneity landscape. *Genome Biol.* 2017;18(1):220.

28. Auslander N, Zhang G, Lee JS, Frederick DT, Miao B, Moll T, et al. Robust prediction of response to immune checkpoint blockade therapy in metastatic melanoma. *Nat Med.* 2018;24(10):1545-9.

29. Chan TA, Yarchoan M, Jaffee E, Swanton C, Quezada SA, Stenzinger A, et al. Development of tumor mutation burden as an immunotherapy biomarker: utility for the

oncology clinic. *Ann Oncol.* 2019;30(1):44-56.

30. Mayakonda A, Lin D-C, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res.* 2018;28(11):1747-56.

31. Elliott LA, Doherty GA, Sheahan K, Ryan EJ. Human tumor-infiltrating myeloid cells: phenotypic and functional diversity. *Front Immunol.* 2017;8:86.

32. Chansky K, Sculier J-P, Crowley JJ, Giroux D, Van Meerbeeck J, Goldstraw P. The International Association for the Study of Lung Cancer Staging Project: prognostic factors and pathologic TNM stage in surgically managed non-small cell lung cancer. *J Thorac Oncol.* 2009;4(7):792-801.

33. Collins LG, Haines C, Perkel R, Enck RE. Lung cancer: diagnosis and management. *Am Fam Physician.* 2007;75(1):56-63.

34. Fong KM, Sekido Y, Minna JD. Molecular pathogenesis of lung cancer. *J Thorac Cardiovasc Surg.* 1999;118(6):1136-52.

35. Lantuéjoul S, Salameire D, Salon C, Brambilla E. Pulmonary preneoplasia–sequential molecular carcinogenetic events. *Histopathology.* 2009;54(1):43-54.

36. Travis WD. Pathology of lung cancer. *Clin Chest Med.* 2011;32(4):669-92.

37. Brierley JD, Gospodarowicz MK, Wittekind C. TNM classification of malignant tumours. *John Wiley & Sons.* 2017.

38. de Geus-Oei L-F, van Krieken JHJ, Aliredjo RP, Krabbe PF, Frielink C, Verhagen AF, et al. Biological correlates of FDG uptake in non-small cell lung cancer. *Lung cancer.* 2007;55(1):79-87.

39. Nishimura T, Nakui M, Sato M, Iwakabe K, Kitamura H, Sekimoto M, et al. The critical role of Th1-dominant immunity in tumor immunology. *Cancer Chemother Pharmacol.* 2000;46(Suppl):S52-S61.

40. Becker JC, Andersen MH, Schrama D, thor Straten P. Immune-suppressive properties of the tumor microenvironment. *Cancer Immunol Immunother.* 2013;62(7):1137-

48.

41. Ruffell B, Affara NI, Coussens LM. Differential macrophage programming in the tumor microenvironment. *Trends Immunol.* 2012;33(3):119-26.

42. Chanmee T, Ontong P, Konno K, Itano N. Tumor-associated macrophages as major players in the tumor microenvironment. *Cancers (Basel).* 2014;6(3):1670-90.

43. Seo J-S, Kim A, Shin J-Y, Kim YT. Comprehensive analysis of the tumor immune micro-environment in non-small cell lung cancer for efficacy of checkpoint inhibitor. *Sci Rep.* 2018;8(1):14576.

44. Takanami I, Takeuchi K, Giga M. The prognostic value of natural killer cell infiltration in resected pulmonary adenocarcinoma. *J Thorac Cardiovasc Surg.* 2001;121(6):1058-63.

45. Fend L, Rusakiewicz S, Adam J, Bastien B, Caignard A, Messaoudene M, et al. Prognostic impact of the expression of NCR1 and NCR3 NK cell receptors and PD-L1 on advanced non-small cell lung cancer. *Oncoimmunology.* 2017;6(1):e1163456.

46. Platonova S, Cherfils-Vicini J, Damotte D, Crozet L, Vieillard V, Validire P, et al. Profound coordinated alterations of intratumoral NK cell phenotype and function in lung carcinoma. *Cancer Res.* 2011;71(16):5412-22.

47. Russick J, Joubert P-E, Gillard-Bocquet M, Torset C, Meylan M, Petitprez F, et al. Natural killer cells in the human lung tumor microenvironment display immune inhibitory functions. *J Immunother Cancer.* 2020;8(2):e001054.

48. Patel SP, Kurzrock R. PD-L1 expression as a predictive biomarker in cancer immunotherapy. *Mol Cancer Ther.* 2015;14(4):847-56.

49. Kris M, Johnson B, Kwiatkowski D, Iafrate A, Wistuba I, Aronson S, et al. Identification of driver mutations in tumor specimens from 1,000 patients with lung adenocarcinoma: The NCI's Lung Cancer Mutation Consortium (LCMC). *J Clin Oncol.* 2011;29(18_suppl):CRA7506.

50. Dong Z-Y, Zhang J-T, Liu S-Y, Su J, Zhang C, Xie Z, et al. EGFR mutation correlates with uninflamed phenotype and weak immunogenicity, causing impaired response to PD-1 blockade in non-small cell lung cancer. *Oncoimmunology.* 2017;6(11):e1356145.

51. Ozaki Y, Muto S, Takagi H, Watanabe M, Inoue T, Fukuhara M, et al. Tumor mutation burden and immunological, genomic, and clinicopathological factors as biomarkers for checkpoint inhibitor treatment of patients with non-small-cell lung cancer. *Cancer Immunol Immunother.* 2020;69(1):127-34.

52. Bouchon A, Dietrich J, Colonna M. Cutting edge: inflammatory responses can be triggered by TREM-1, a novel receptor expressed on neutrophils and monocytes. *J Immunol.* 2000;164(10):4991-5.

53. Ho C-C, Liao W-Y, Wang C-Y, Lu Y-H, Huang H-Y, Chen H-Y, et al. TREM-1 expression in tumor-associated macrophages and clinical outcome in lung cancer. *Am J Respir Crit Care Med.* 2008;177(7):763-70.

54. Zanzinger K, Schellack C, Nausch N, Cerwenka A. Regulation of triggering receptor expressed on myeloid cells 1 expression on mouse inflammatory monocytes. *Immunology.* 2009;128(2):185-95.

55. Lesokhin AM, Hohl TM, Kitano S, Cortez C, Hirschhorn-Cymerman D, Avogadri F, et al. Monocytic CCR2+ myeloid-derived suppressor cells promote immune escape by limiting activated CD8 T-cell infiltration into the tumor microenvironment. *Cancer Res.* 2012;72(4):876-86.

56. Tcyganov E, Mastio J, Chen E, Gabrilovich DI. Plasticity of myeloid-derived suppressor cells in cancer. *Curr Opin Immunol.* 2018;51:76-82.

57. Byun BH, Kim KM, Cheon GJ, Choe DH, Koh JS, Lee DY, et al. 18F-FDG uptake and EGFR mutations in patients with non-small cell lung cancer: a single-institution retrospective analysis. *Lung Cancer.* 2010;67(1):76-80.

58. Takamochi K, Mogushi K, Kawaji H, Imashimizu K, Fukui M, Oh S, et al.

Correlation of EGFR or KRAS mutation status with 18F-FDG uptake on PET-CT scan in lung adenocarcinoma. *PLoS One.* 2017;12(4):e0175622.

59. Minamimoto R, Jamali M, Gevaert O, Echegaray S, Khuong A, Hoang CD, et al. Prediction of EGFR and KRAS mutation in non-small cell lung cancer using quantitative 18F FDG-PET/CT metrics. *Oncotarget.* 2017;8(32):52792-801.

60. Vesselle H, Salskov A, Turcotte E, Wiens L, Schmidt R, Jordan CD, et al. Relationship between non-small cell lung cancer FDG uptake at PET, tumor histology, and Ki-67 proliferation index. *J Thorac Oncol.* 2008;3(9):971-8.

61. Herbst RS. Review of epidermal growth factor receptor biology. *Int J Radiat Oncol Biol Phys.* 2004;59(2 Suppl):21-S6.

62. Jorissen RN, Walker F, Pouliot N, Garrett TP, Ward CW, Burgess AW. Epidermal growth factor receptor: mechanisms of activation and signalling. *Exp Cell Res.* 2003;284(1):31-53.

63. Aquino SL, Halpern EF, Kuester LB, Fischman AJ. FDG-PET and CT features of non-small cell lung cancer based on tumor type. *Int J Mol Med.* 2007;19(3):495-9.

64. Han S, Oh JS, Kim JS. Immune microenvironment of the gene signature reflecting the standardised uptake value on 18 F-fluorodeoxyglucose positron emission tomography/computed tomography in head and neck squamous cell carcinoma. *Ann Nucl Med.* 2021;35(1):65-75.

국 문 초 록

유사시간분석을 이용한 폐선암의 임상분자적 및 면역학적 변화 연구

이현종

서울대학교

융합과학기술대학원

분자의학 및 바이오제약학과

폐선암은 폐의 악성 종양 중 가장 흔한 조직학적 유형이다. 폐암의 분자적 특징은 단면적으로 평가해 왔다. 따라서 폐암의 생물학적 진행 과정에 대해서는 적절히 모델화하지 못하였다. 폐암의 진행은 특히 병기 설정, 조직학적 소견, 그리고 플루오르데옥시글루코오스 양전자 단층 촬영을 이용한 포도당 대사 평가 소견에 의해 임상적으로 그리고 병리학적으로 평가해 왔다. 그러나 이러한 평가 도구들은 폐암의 분자 프로파일의 변화 측면에서 생물학적 진행을 평가하기 어렵다. 이

연구에서는 공공 접근이 가능한 데이터셋으로부터 폐선암의 생물학적 진행을 반영하는 유사시간궤적을 추정하였다. 그리고 이에 따라 종양의 특질이 어떻게 변화하는지 분석하였다.

암 유전체 아틀라스(TCGA)의 폐암 데이터셋에서 페노패쓰(Phenopath) 도구를 이용하여 폐암의 유사시간궤적을 구성하였다. 유사시간과 연관성이 있는 유전자들을 선택하였고 온톨로지 분석을 시행하였다. 비소세포폐암 영상유전체 데이터셋에서는 라쏘(Lasso) 회귀분석을 이용하여 유사시간궤적을 추정하였다. 종양 병기를 포함한 임상적 요인들과 유사시간의 상관 분석을 시행하였다. 플루오르데옥시글루코오스 양전자 단층 촬영 영상은 암 영상 아카이브(TCIA)로부터 수집하였다. 표준섭취계수를 비롯한 영상 파라미터들은 라이프엑스(LifeX) 소프트웨어에서 적응 종양 경계 방식으로 그려진 관심 영역으로부터 구했다. 영상 파라미터와 유사시간의 상관 분석을 시행하였다. 면역 프로파일은 엑스셀(xCell) 도구를 이용하여 구하였다. 유전자 돌연변이 데이터는 지노믹 데이터 커먼스(genomic data commons)로부터 다운로드하였다. 면역 세포들의 농축 점수 및 종양 돌연변이 부하와 유사시간의 상관 분석을 시행하였다. 림프절 전이 및 포도당 대사와 상관이 있는 면역학적 인자를 탐색하기 위하여 차이발현 유전자(DEG) 분석, 유전자 온톨로지 분석, 그리고 비교 분석을 시행하였다.

암 유전체 아틀라스 데이터셋에서 유사시간궤적을 성공적으로 추정하였다. 폐선암 표본에서 자연살해세포의 활성과 관련된 분자 프로파일이 유사시간에 따라 하향 조절됨을 보였다. 세포 분열과 관련된 분자 프로파일은 유사시간에 따라 상향 조절되었다. 암 유전체 아틀라스의 데이터셋에서 구한 라쏘 회귀분석 모델을 이용하여 비소세포폐암 영상유전체 데이터셋의 폐선암 샘플에서도 유사시간궤적을 추정하였다. 암 유전체 아틀라스 데이터셋에서 유사시간은 T 병기와 전체 병기에 따라 차이를 보였으나 N 병기와 M 병기에 따라서는 차이를 보이지 않았다. 전체 생존기간은 이른 유사시간 그룹과 늦은 유사시간 그룹에서 유의한 차이를 보였다. 최대 표준섭취계수와 SLC2A1 의 발현도는 유사시간과 양의 상관관계를 보였다. 종양 면역 미세 환경에서 세포의 농축 정도는 유사시간에 따라 변화하였다. 제 1 형 보조 T 세포는 유사시간과 양의 상관관계를 보였고 M2 대식세포는 음의 상관관계를 보였다. 이른 유사시간 그룹 내의 림프절 전이가 없는 폐선암 환자보다 림프절 전이가 있는 폐선암 환자에서 TREM-1 유전자가 상향 조절되었다. 늦은 유사시간 그룹 내에서는 높은 플루오르데옥시글루코오스 섭취를 보이는 표본에서 단핵구 이동의 음성조절이 유의하게 하향 조절됨을 보였다.

결론적으로 폐암의 유사시간궤적을 전사체 프로파일에 근거하여 추정하였다. 임상 병기와 표준섭취계수는 유사시간과 유의한 상관관계가 있어 폐암의 분자적 진행을 평가할 수 있는 새로운 척도로서의 가능성을

보였다. 종양 면역 미세 환경에서 제1형 보조 T 세포와 M2 대식세포는 유사시간과 각각 양의 상관관계 및 음의 상관관계를 보였다. 면역학적 인자들은 질병의 이른 단계에서 림프절 전이 여부와, 질병의 늦은 단계에서 포도당 대사와 연관성을 보였다. 이 연구는 유사시간분석을 이용하여 폐선암의 종양 특질이 어떻게 변화하는지 밝혔다.

주요어: 폐선암, 전이, 면역, 대사, 유사시간

학번: 2015-26015