

CORRESPONDENCE

Open Access



Multi-omics and clustering analyses reveal the mechanisms underlying unmet needs for patients with lung adenocarcinoma and identify potential therapeutic targets

Ken Asada^{1,2*†}, Syuzo Kaneko^{1,2†}, Ken Takasawa^{1,2†}, Kouya Shiraishi³, Norio Shinkai^{1,2}, Yoko Shimada³, Satoshi Takahashi^{1,2}, Hidenori Machino^{1,2}, Kazuma Kobayashi^{1,2}, Amina Bolatkan^{1,2}, Masaaki Komatsu^{1,2}, Masayoshi Yamada⁴, Mototaka Miyake⁵, Hirokazu Watanabe⁵, Akiko Tateishi⁶, Takaaki Mizuno^{3,6,7}, Yu Okubo⁸, Masami Mukai⁹, Tatsuya Yoshida⁶, Yukihiro Yoshida⁸, Hidehito Horinouchi⁶, Shun-Ichi Watanabe⁸, Yuichiro Ohe⁶, Yasushi Yatabe¹⁰, Takashi Kohno^{3*} and Ryuji Hamamoto^{1,2*}

Abstract

Background The cancer genome contains several driver mutations. However, in some cases, no known drivers have been identified; these remaining areas of unmet needs, leading to limited progress in cancer therapy. Whole-genome sequencing (WGS) can identify non-coding alterations associated with the disease. Consequently, exploration of non-coding regions using WGS and other omics data such as ChIP-sequencing (ChIP-seq) to discern novel alterations and mechanisms related to tumorigenesis have been attractive these days.

Methods Integrated multi-omics analyses, including WGS, ChIP-seq, DNA methylation, and RNA-sequencing (RNA-seq), were conducted on samples from patients with non-clinically actionable genetic alterations (non-CAGAs) in lung adenocarcinoma (LUAD). Second-level cluster analysis was performed to reinforce the correlations associated with patient survival, as identified by RNA-seq. Subsequent differential gene expression analysis was performed to identify potential druggable targets.

Results Differences in H3K27ac marks in non-CAGAs LUAD were found and confirmed by analyzing RNA-seq data, in which mastermind-like transcriptional coactivator 2 (*MAML2*) was suppressed. The down-regulated genes whose expression was correlated to *MAML2* expression were associated with patient prognosis. WGS analysis revealed

[†]Ken Asada, Syuzo Kaneko and Ken Takasawa contributed equally to this work.

*Correspondence:

Ken Asada
ken.asada@riken.jp
Takashi Kohno
tkkohno@ncc.go.jp
Ryuji Hamamoto
rhamamot@ncc.go.jp

Full list of author information is available at the end of the article



somatic mutations associated with the H3K27ac marks in the *MAML2* region and high levels of DNA methylation in *MAML2* were observed in tumor samples. The second-level cluster analysis enabled patient stratification and subsequent analyses identified potential therapeutic target genes and treatment options.

Conclusions We overcome the persistent challenges of identifying alterations or driver mutations in coding regions related to tumorigenesis through a novel approach combining multi-omics data with clinical information to reveal the molecular mechanisms underlying non-CAGAs LUAD, stratify patients to improve patient prognosis, and identify potential therapeutic targets. This approach may be applicable to studies of other cancers with unmet needs.

Keywords Genomics, Epigenomics, Lung adenocarcinoma, Multi-omics analysis, Biomarkers, Therapeutic target

Introduction

Lung cancer is one of the most frequently diagnosed cancers and the second most common cause of death worldwide. Non-small cell lung cancer (NSCLC) accounts for approximately 85% of all lung cancers. Genetic alterations can drive cancer, therefore, genetic testing using next-generation sequencing (NGS) to identify targeted mutations in lung cancer can facilitate strategic decisions regarding cancer therapy [1]. However, the discovery of genes altered in the coding regions of cancer is expected to reach a plateau. In other words, newly identified genes related to tumorigenesis may essentially be rediscoveries or already reported in cancer research. Therefore, cancer genome studies have gradually shifted from whole-exome sequencing (WES) to whole-genome sequencing (WGS) analysis and analyses of non-coding regions revealed novel mutations, highlighting the feasibility and benefits of WGS. Non-coding or regulatory regions are cis-regulatory elements that include promoters, enhancers, insulators, and 5'- and 3'-untranslated regions (UTRs) as locus control regions. Changes in DNA sequences or functional dysregulation in regulatory regions cause cancer. Thus, focusing on non-coding regions is highly beneficial for cancer studies. As non-coding regions do not code proteins, methods other than genomic sequencing are more desirable. Specifically, multi-omics analysis methods, including chromatin immunoprecipitation sequencing (ChIP-seq), can facilitate genome-wide DNA structure profiling effectively and elucidate cancer traits.

Genomic alterations differ according to race. According to The Cancer Genome Atlas (TCGA) lung adenocarcinoma (LUAD) database, mutations occur in *KRAS* (32.2%), *EGFR* (11.3%), and *BRAF* (7.9%), which are three of the four alterations with available molecular targeting medicines for lung cancer. However, the frequency of *EGFR* mutations is higher in the East Asian LUAD population (~50%) [2]. This suggests that cancer research using defined cohort datasets is crucial for race-based medicine or precision oncology and will contribute to better decision-making for cancer treatment. Furthermore, driver mutations in 30–50% of patients with NSCLC, including those from East Asian and Caucasian

populations, have not yet been identified [3], leading to limited progress in cancer therapy.

DNA methylation is an epigenetic marker found in the promoter region throughout the gene body and the levels of the DNA methylation are associated with gene expression. Another epigenetics, enhancer activity is an epigenetic landscape, which is characterized by histone modifications associated with chromatin structure, have potential clinical implications. In this study, we conduct multi-omics analysis of patients with non-CAGA LUAD using WGS to identify genomic alterations, ChIP-seq to examine histone modifications, RNA sequencing (RNA-seq) to analyze gene expression, DNA methylation to identify epigenetic modifications, and clinical information to characterize clinical features, reveal the onset of cancer and discover potential therapeutic targets.

Results

Mutational landscape in non-CAGAs LUAD samples

An overview of this study is shown in Fig. S1. After extracting LUAD, those with sufficient specimens to perform WGS and at least one epigenetic analysis (ChIP-seq and DNA methylation) were used for subsequent analysis ($N=184$). Approximately 40% of the cohort was patients with non-CAGA. Additionally, we analyzed TCGA LUAD dataset and the analysis revealed that approximately 46% samples might be non-CAGA samples. Detailed information on the dataset are provided in the Table S1-3 and Supplementary Methods.

Previous studies have classified LUAD into driver gene mutation and driver mutation-negative subtypes based on various criteria. The definition of driver mutations in this study was based on three criteria detailed in the Supplementary Methods. Samples that did not contain any of the mutations described in the Supplementary Methods and Table S3 were categorized as driver mutation-negative (hereafter referred to as non-clinically actionable genetic alterations (non-CAGAs)) in this study.

The global landscape of somatic mutations is shown in Fig. S2A (bin of 1 kb). The mutations in each sample, including non-coding mutations, are summarized in Table S4. Profiling of copy number alterations (CNAs) showed that chromosomes (Chr) 1, 5, 7, and 8 tended

to have more gain (Fig. S2B). One patient showed multiple hetero losses in Chr4, 7, 11, 12, and 13, and the other patient showed duplications in Chr8, 10, 13, and 14 (Fig. S2C and D). Allele-specific copy number analysis detected patients with a copy neutral loss of heterozygosity (LOH) (Fig. S2E). Hetero losses and copy neutral LOH were detected as a clonal event, whereas duplications were observed as a subclonal event. However, not all samples had apparent CNAs. Figure S2C – E shows examples of representative data for CNA in non-CAGA samples. Notably, we recently reported that 1.15% of non-CAGA LUAD cases exhibit chromosomal rearrangement around *ERBB2*. This structural variation was linked to the super enhancer formation that was associated with *ERBB2* overexpression. We further demonstrated that *ERBB2* is a feasible of druggable target in non-CAGA LUAD patients [4].

Genomic, epigenomic, and transcriptomic differences between adjacent normal and tumor samples

Although we identified the CNAs in a few cases, a study reported that LUAD is generally rich in somatic mutation compared to SV [5]. We therefore performed a comprehensive study of epigenomic alterations using a combination of WGS, ChIP-seq, RNA-seq, and DNA methylation analyses to investigate the non-coding mutations in enhancer regions, with the aim of revealing the molecular mechanisms underlying non-CAGAs. Here, we compared enhancer activity between normal and non-CAGA samples and found that enhancer activity at the mastermind-like transcriptional coactivator 2 (*MAML2*) genomic locus (chr11:95976598–96343195) was ablated in a significant fraction of samples (Table S5). MAML is a coactivator of Notch and the MAML complex induced Notch-dependent target genes, including *c-MYC*, *p21*, *ERBB2*, *CCND3*, *HES1*, *HEY1*, and *NFKB1*. *MAML2* has conserved domains, forms stable DNA-binding complexes, and regulates Notch and Wnt/ β -catenin signaling pathways by promoting β -catenin turnover independent of Notch signaling [6]. Several fusion genes, such as *YAPI-MAML2*, *MECT1-MAML2*, and *CRTC1/3-MAML2* have been identified; among these, *CRTC1-MAML2* is an oncogenic driver in mucopidermoid carcinoma (MEC) [7]. Here, the suppressed enhancer activity was coupled with the down-regulation of *MAML2* expression. A comparison of the matched normal adjacent to tumor and tumor tissues samples showed that *MAML2* was suppressed in tumors (Fig. 1A and B). To examine if somatic mutations are associated with enhancer activity, patients with mutations in any of genomic regions of *FAM78B* (95768953–95789782), *CEP57* (CEP57:95790498–95832693), *MTMR2* (95832880–95924107), and *CCDC82* (96352773–96389912) genomic regions, which are neighboring genes

of *MAML2*, were examined alongside those of *MAML2*. One hundred and seven of 184 patients had mutations in the *MAML2* gene locus, whereas other genes showed relatively fewer mutations (Fig. 1C). H3K27ac ChIP-seq peak of matched normal and tumor samples with genomic mutations are shown in Fig. 1D. Only two cases were mutated in the coding regions (R60Q and R422L) of the *MAML2* gene, indicating that most detected alterations were non-coding regions. Therefore, we further examined whether genomic alterations in those regions affect enhancer activity. We extracted a complete dataset that included WGS, ChIP-seq, and RNA-seq ($N=113$). Forty-nine of the 113 samples had null mutations, whereas 64 samples had alterations at the *MAML2* locus. Enhancer activity was suppressed in the mutated samples, which was associated with *MAML2* gene expression (Fig. 1E and F). Analysis of the TCGA non-CAGA-like LUAD dataset revealed that *MAML2* was also down-regulated in tumors compared its expression in a relatively large number of normal tissues adjacent to the tumors (Fig. 1G). This suggests that *MAML2* expression was down-regulated in patients with non-CAGA LUAD in both cohorts.

DNA methylation is another epigenetic mechanism that is beneficial for revealing the underlying mechanisms in cancer. The EPIC array includes Functional Annotation of the Mammalian Genome (FANTOM) 5 and Encyclopedia of DNA Elements (ENCODE) enhancer regions for DNA methylation detection, which promotes the study of regulatory regions. Therefore, we decided to use the EPIC array in this study to assess DNA methylation. Analysis using matched normal and tumor samples revealed that 18 of the 102 probes exhibited substantially different patterns in the *MAML2* region (Fig. S3A). Similarly, 46 of 102 probes had different methylation levels in normal to non-CAGA samples. Sixteen probes overlapped in both analyses (Fig. S3B, left, Venn diagram; right; summary), indicating that these methylation sites may be potential diagnostic markers. We identified low DNA methylation levels in normal samples but high DNA methylation levels in tumor samples, which were inversely correlated with *MAML2* gene expression. This finding agrees with those of a previous report, in which high DNA methylation levels in the *MAML2* region suppressed gene expression [8]. To investigate how DNA methylation is regulated in *MAML2*, we examined the expression levels of known DNA methyltransferases (DNMTs) and demethylation-related enzymes. In this study, elongator complex protein 3 (ELP3), which plays a role in paternal genome demethylation, and tet methylcytosine dioxygenase (TET) 2, which is involved in the TET dioxygenase-mediated oxidation of 5-methylcytosine (5mC) pathways, were down-regulated in matched tumor and non-CAGA samples (Fig. S3C and D). However, we

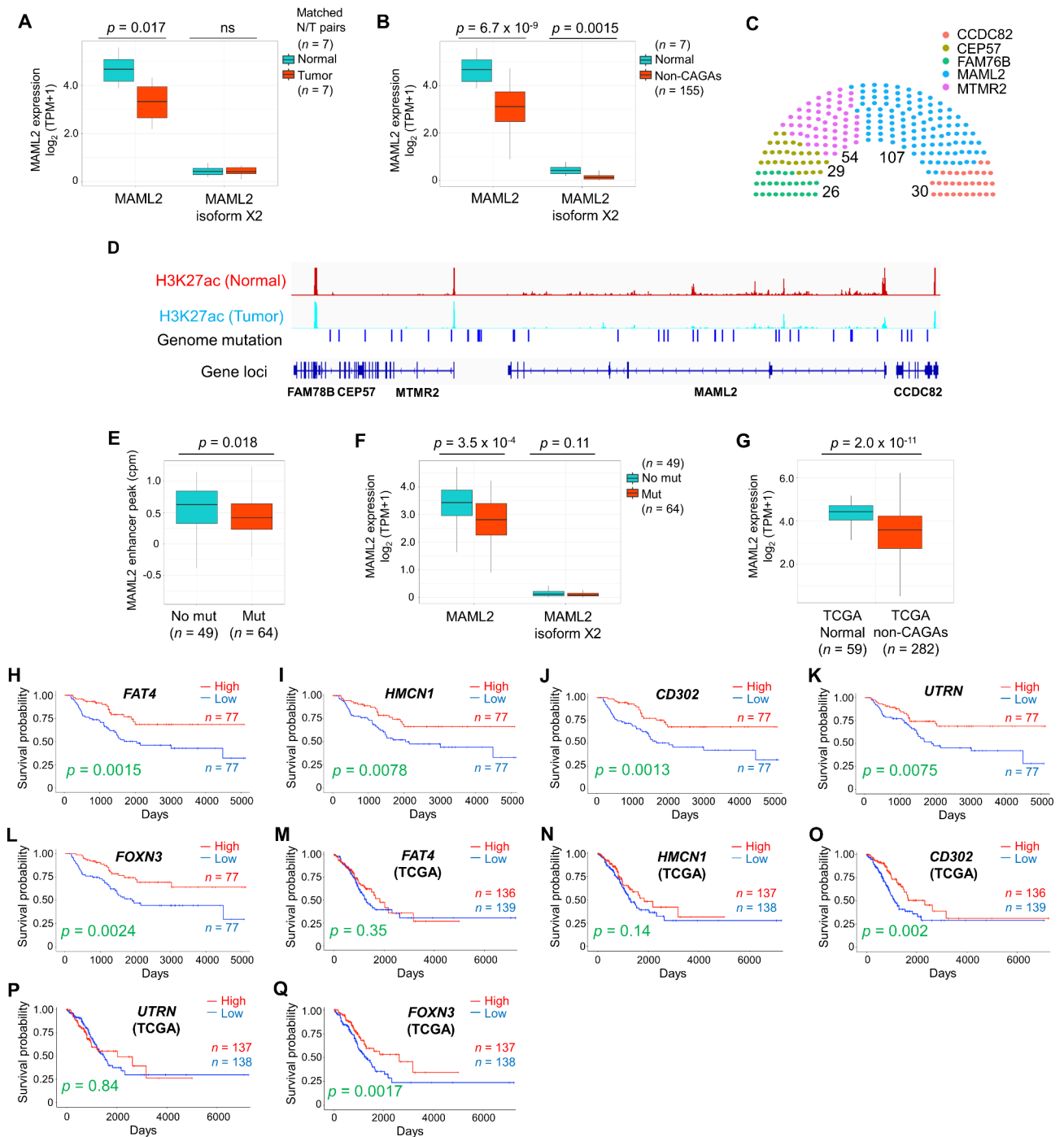


Fig. 1 Genetic and epigenetic analysis in non-CAGA lung adenocarcinoma samples. **A** MAML2 expression in matched normal and tumor tissue samples. **B** MAML2 expression in normal and non-CAGA samples. **C** Number of patients with mutations in Chr11 (*FAM78B*:95768953-95789782, *CEP57*:95790498-95832693, *MTMR2*:95832880-95924107, *MAML2*:95976598-96343195, and *CCDC82*:96352773-96389912). Somatic mutations and small insertions and deletions (INDELs) were analyzed, patients with at least one mutation were counted. **D** H3K27ac ChIP-seq peak of matched normal and tumor samples with genomic mutations. **E** Enhancer peak of H3K27ac in null and mutated samples at the *MAML2* locus. **F** MAML2 expression analysis using the non-CAGA-like TCGA LUAD dataset. **G** MAML2 expression analysis using the non-CAGA-like TCGA LUAD dataset. **H-L** Kaplan–Meier estimates of overall survival (OS) in patient with non-CAGA LUAD. A total of 154 patients were divided in half. **H** OS of *FAT4*. **I** OS of *HMCN1*. **J** OS of *CD302*. **K** OS of *UTRN*. **L** OS of *FOXN3*. **M-Q** Kaplan–Meier estimates of OS using the non-CAGA-like TCGA LUAD dataset. **M** OS of *FAT4*. **N** OS of *HMCN1*. **O** OS of *CD302*. **P** OS of *UTRN*. **Q** OS of *FOXN3*

did not observe the up-regulation of DNA methyltransferases, suggesting that demethylation mechanisms play pivotal roles in patients with non-CAGA LUAD. Notably, these mechanisms could be tissue-specific, as DNMT3B is involved in breast cancer [8]; however, *ELP3* and *TET2* were associated with patients with non-CAGA LUAD.

MAML2-dependent signaling pathways and genes related to clinical outcomes

To investigate whether known Notch and Wnt/ β -catenin targeted genes are associated with *MAML2*, we selected seven Notch targeted genes (*BCL2*, *CCND3*, *CDKN1*, *ERBB2*, *HERDUP1*, *HES1*, *HEY1*), seven Wnt/ β -catenin targeted genes (*CD44*, *CTNBN1*, *FNI*, *MMP7*, *PMP22*, *SMYD3*, *VEGFA*), and two common genes (*CCND1* and *MYC*) that are expressed in lung cancer (Table S6). We identified genes such as *BCL2* (XM_047437733 and NM_000633), *CDKN1* (NM_001374511), *CD44* (NM_001001390, XM_005253238, and XM_006718390), *PMP22* (XM_047436306 and NM_153322) were down-regulated, whereas *ERBB2* were overexpressed in non-CAGA samples (Fig. S4A). Next, to identify prognostic biomarkers and potential therapeutic targets associated with *MAML2*, we performed a correlation analysis and identified the top 15 positively and negatively correlated genes against *MAML2* (Table S7). The RNA-expression levels of these genes were compared to those in normal samples; all 15 genes were significantly down-regulated and most were up-regulated in response to *MAML2* down-regulation (Fig. S4B and C). Kaplan–Meier survival analysis revealed poor prognosis in the subgroups with low *FAT4* (XM_011532237; isoform X1), *HMCN1* (XM_011510038; isoform X1), *CD302* (NM_014880; isoform 1 precursor), *UTRN* (NM_007124; isoform 1), and *FOXN3* (NM_001085471; isoform 1) expression (Fig. 1H–L, Table S8). To validate our findings, we performed survival analysis using a Korean dataset (GSE8894), because the previously published paper showed that the genetic backgrounds of the Japanese and Korean populations were the closest among the populations analyzed [9]. Consistent with our earlier results, there was a tendency for low expression levels of marker genes that were associated with poor prognosis (Fig. S5). Some of the genes did not show statistical significance, possibly due to the smaller sample size of the Korean dataset ($N=61$) compared to our dataset ($N=154$) and/or the presence of samples with driver mutations (if any), which could affect the results, as genomic information was unavailable in the Korean cohort. To further investigate whether these prognostic marker genes were specific to the Asian cohort, we performed survival analysis using the non-CAGA-like TCGA LUAD dataset, the results of which also revealed significant differences in survival according to *CD302* and *FOXN3* expression (Fig. 1M–Q). In

summary, *CD302* and *FOXN3* are prognostic markers in non-CAGA, Korean, and non-CAGA-like TCGA datasets, independent of ethnicity or race.

Although we identified prognostic markers and potential therapeutic targets, the molecular mechanisms related to *MAML2* are unclear. Therefore, to explore the underlying mechanisms, we performed weighted gene correlation network analysis (WGCNA) to identify clusters of highly co-expressed hub genes (Fig. S6). We chose power 10 as the lowest possible power term where topology fits a scale free network (Fig. S6A left and right) and constructed a gene dendrogram to detect modules by hierarchical clustering (Fig. S6B). The *PCNX1* gene had the most gene connections at 48, and exhibited greatest co-expression with the *RNLS* gene, followed by *FTO* with 19 connections, suggesting that these two genes are hub genes identified in *MAML2*-associated subgroups and may orchestrate the signaling pathways (Fig. S6C).

Identification of potential therapeutic target genes via unsupervised learning

Next, we examined whether the prognostic marker genes identified were commonly expressed in all samples. Approximately 30% of patients had a common gene expression profile and were subclassified into either the high expression group or low expression group (Fig. S7A and B). Then, Kaplan–Meier survival analysis was conducted to determine whether commonly expressing subgroups showed improved patient stratification for overall survival (OS) compared with those analyzed using the expression of each gene. This approach failed to achieve better patient stratification (Fig. S7C and D). However, heatmap analysis of the prognostic genes enabled clustering of the samples (Fig. S7E, I – III in samples and A and B in genes), indicating that patient stratification related to OS could be improved. Therefore, we aimed to re-cluster patients by reinforcing existing correlations between the expression levels of prognostic genes and survival associations, thereby inflating the association of these components with survival. The results of the survival analysis for each gene were regarded as a first-level cluster, and patients with low expression were labeled -1 , whereas those with high expression were labeled 1 (Fig. 2A). Using these labels, hierarchical or non-hierarchical K-means clustering was performed to obtain second-level cluster labels (Fig. S7F–H and Supplementary Method). The Elbow method was used to determine the optimal number of clusters for K-means analysis (Fig. S7G), and the clustering result were plotted (Fig. S7H). Based on the aforementioned results, we performed second-level patient stratification related to prognosis using these labels. Here, we achieved the optimal classification using labels obtained from the hierarchical clustering of *CD302*, *FAT4*, and *FOXN3* genes (Fig. 2B), rather than

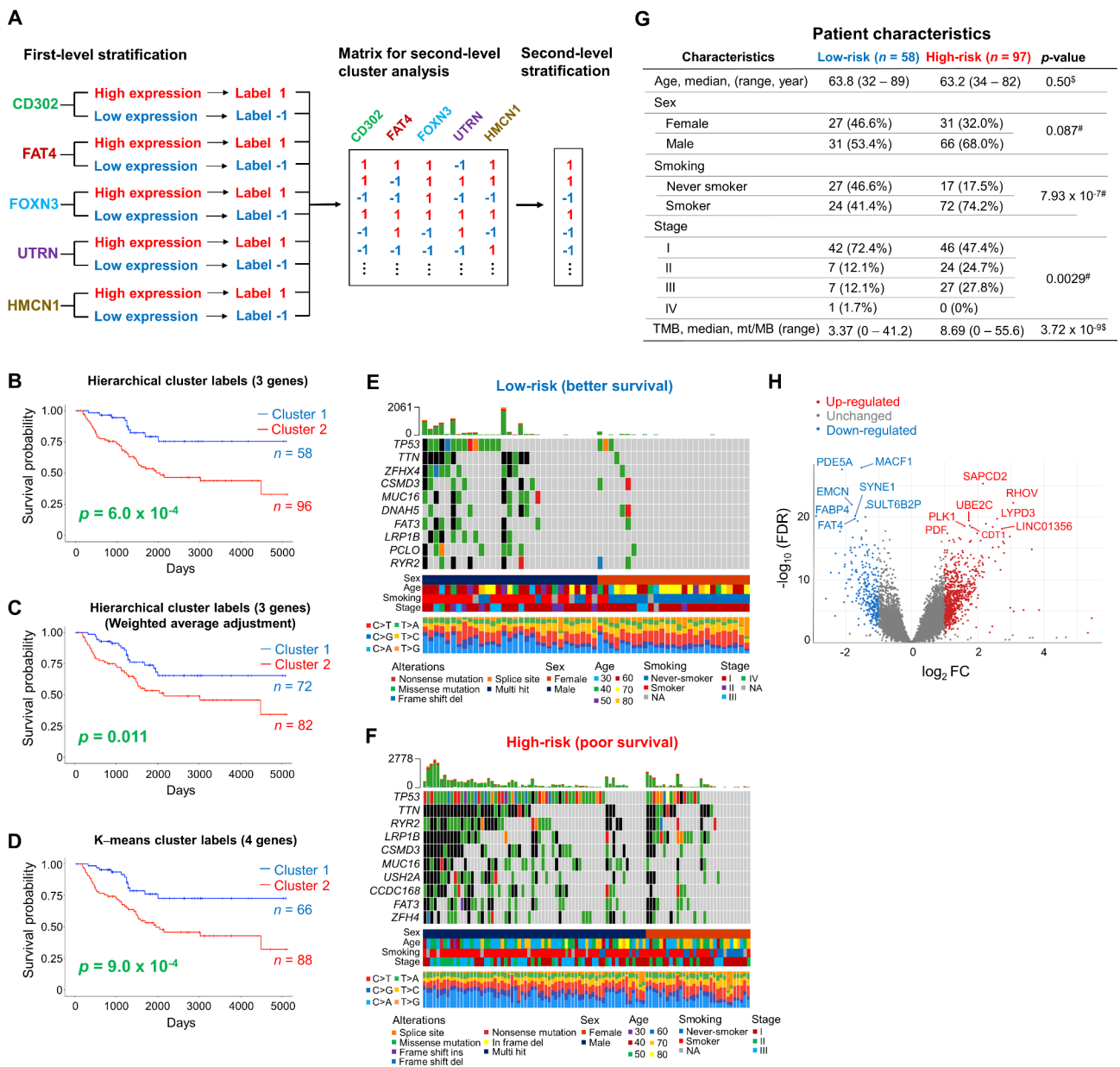


Fig. 2 Second-level cluster analysis to improve patient stratification. **A** Workflow of the analysis. **B–D** Kaplan–Meier estimates of overall survival (OS) with secondlevel cluster analysis. **B** OS was assessed using hierarchical labels. Second-level cluster labels were obtained from the survival analysis of *CD302*, *FAT4*, and *FOXN3* genes. **C** OS was assessed using hierarchical labels. Second-level cluster labels were obtained from the survival analysis of *CD302*, *FAT4*, and *FOXN3* genes with weighted average adjustment. **D** OS was assessed using K–means labels. Second-level cluster labels were obtained from the survival analysis of *CD302*, *FAT4*, *FOXN3*, and *UTRN* genes. **E–F** Mutation profiles of high-risk (poor survival) and low-risk (better survival) subgroups. **E** Low-risk subgroup (cluster 1) from **B**. **F** High-risk subgroup (cluster 2) from **B**. **G** Patient characteristics in the two subgroups. § represents p-values obtained from the Mann–Whitney U test. # represents p-values obtained from Fisher’s exact test. **H** Volcano plot of the genes differentially expressed between high-risk and low-risk subgroups. Up-regulated genes in the poor survival subgroup are represented in red and down-regulated genes are represented in blue.

hierarchical clustering with weighted average adjustment and K-means cluster labels (Fig. 2C and D, Table S9 and 10).

Genomic features and patient characteristics were examined in both groups. The high-risk group accumulated more mutations than the low-risk group (Fig. 2E and F). For example, *TP53* mutations were found in 72% of high-risk patients, whereas less than half of patients in the low-risk subgroup had *TP53* mutations (29%). Other recurrently mutated genes such as *TTN* and *RYR2* were also highly mutated in the high-risk group. A comparison of patient characteristics between the two groups revealed that the occurrence of smoking status, advanced cancer stage, and high tumor mutation burden (TMB) was greater in the high-risk group (Fig. 2G).

Differentially expressed gene (DEG) analysis revealed 802 up-regulated and 289 down-regulated genes with a threshold of 2-fold difference and false discovery rate (FDR) < 0.05 in the high-risk subgroup (Fig. 2H). Notably, *CD302*, *FAT4*, *HMCN1*, and *UTRN* were significantly down-regulated whereas genes including *PLK1*, *UBE2C*, and *LYPD3* which are reportedly elevated in LUAD, were up-regulated (Fig. 2H, Table S11). This finding suggests that second-level stratification, followed by DEG analysis can effectively identify therapeutic target genes. According to gene ontology (GO) biological processes, the up-regulated DEGs were enriched in the mitotic cell cycle, cell cycle process, and cell cycle, whereas down-regulated DEGs were enriched in anatomical structure development, developmental process, and anatomical structure morphogenesis in the poor survival subgroup (Fig. S8A and B, Table S12). To further investigate the global signaling pathways related to the subgroups, we performed Gene Set Enrichment Analysis (GSEA) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses. Blood vessel, circulatory system development, and vasculature development were enriched in GSEA, whereas focal adhesion, platelet activation, actin cytoskeleton regulation, and vascular smooth muscle contraction were enriched in KEGG pathways (Fig. S8C-F).

WGCNA revealed that seven out of 14 modules were significantly correlated with the second-level cluster subgroups (Fig. S9A and B). Among the module eigengenes (MEs), MEblack, MEblue, MEpurple, MEmagenta, and MEbrown were top modules associated with the subgroups. *MAML2*, *CD302*, *FAT4*, *FOXN3*, and *HMCN1* were clustered as MEblack, whereas *UTRN* was clustered as MEblue (Fig. S9A and B, Table S13). *FAT4*, a human homolog of tumor suppressor gene *Fat* in *Drosophila*, modulates Wnt/ β -catenin signaling. *HMCN1* is associated with the Hippo pathway in cancer. *CD302* is associated with cancer-associated fibroblasts (CAFs) and is down-regulated in lung cancer. *UTRN* inhibits tumor growth by attenuating p38 and JNK/c-Jun signaling and

FOXN3 functions as a tumor suppressor by suppressing Wnt/ β -catenin signaling. As previously identified by WGCNA for the subgroups with high and low *MAML2* expression (Fig. S6), similar genes were discovered as hub genes as well as components of the networks (Fig. S9C, Table S14). This indicates that second-level cluster analysis enabled the re-clustering of patients by reinforcing existing correlations between the expression levels of genes and survival association for more accurate patient stratification.

Discussion

Although targeted therapies are clinically effective, a more comprehensive understanding of the cancer biology is required for precision oncology because actionable target-negative cancers have hampered progress in the field of cancer therapy for decades. Regulatory elements in non-coding regions are cis-regulatory elements that include promoters, enhancers, insulators, and 5'- and 3'-UTRs as locus control regions. Changes in DNA sequences in regulatory regions cause cancer, and histone modifications govern chromatin remodeling and enhance transcription activity. In this study, investigating the epigenomics revealed the clue of tumorigenesis and the mechanisms underlying non-CAGA LUAD patients. Later, we identified prognostic markers and potential therapeutic targets.

Here, we conducted an integrated multi-omics analysis for regulatory genomics, focusing on samples with non-CAGAs or unmet needs. Genes that were positively correlated with *MAML2* expression were considered prognostic marker genes and second-level cluster analysis demonstrated enhanced prognostic predictive power. *MAML* regulates Notch and Wnt/ β -catenin signaling pathways. *MAML2* genomic rearrangement has been clinically evaluated in MEC (<https://oncology.test-catalog.org/show/MAMLF>). *MAML2*-based therapeutic modalities could be approached through several strategies. *MAML2* regulates Notch signaling pathways and *CTCRI-MAML2* is an oncogenic fusion gene in MEC. *CTCRI-MAML2* requires AREG-EGFR signaling for MEC growth; co-targeting of Notch by DBZ and EGFR signaling by Erlotinib was an effective to anti-MEC treatment by attenuating MEC growth [7]. We found that *MAML2* was down-regulated in non-CAGA LUAD samples; therefore, rescuing *MAML2* expression serves as a potential therapeutic approach. Putative transcription factor binding sites to *MAML2* has been previously predicted using the TransFac program [8]. Thus, recruiting or enhancing binding affinity of those transcriptional factors to the promoter could induce *MAML2* up-regulation. A second therapeutic approach could involve DNA methylation targeting. *MAML2* expression negatively correlates with DNA methylation. Hence, DNA

methylase inhibition or DNA demethylase activation could induce *MAML2* up-regulation. The third therapeutic approach could involve *MAML2* expression-related prognostic marker targeting. In this case, positively and negatively correlated genes are considered potential therapeutic targets for the treatment of non-CAGA LUAD. However, further studies are needed to evaluate the efficacy and safety of these approaches.

In our study, Notch target gene *BCL2* was down-regulated, and the pro-survival and pro-apoptotic *BCL2* family proteins are attractive for the cancer treatment. *CDKNI* is also one of the Notch targeted gene. Intriguingly, previously published literature demonstrated that knockdown of *SOX9* in LUAD resulted in the up-regulation of *CDKNI*, suggesting that *CDKNI* gene might be a common target of Notch and Wnt/ β -catenin [6, 10]. A novel aspect of this study is that we regarded the poor and better survival groups as distinct clusters, and second-level cluster analysis using prognosis-related labels led to improved patient stratification. DEGs between groups demonstrated that the identified prognostic markers were down-regulated, whereas potential therapeutic targets for human cancers such as *PLK1* and *UBE2C* were up-regulated, which overexpression represses autophagy, inducing initiation, progression, and metastasis in NSCLC [11, 12].

WGCNA identified *PCNX1* and *FTO* as hub genes in both subgroups dichotomized by *MAML2* expression and by second-level cluster labels. *PCNX1* is an evolutionarily conserved component that activates the Notch signaling. *PCNX* is a human homolog of *Drosophila pecanex* (*pcx*). Currently, the role of *PCNX* in Notch signaling remains unknown; however, in *Drosophila*, *pcx* is a component of Notch signaling and in breast cancer, *PCNX* expression is associated with post-chemotherapy patient survival [13]. *RNLS*, *PTEN*, and *ATAD1* were identified in prostate tumor [14], suggesting that *RNLS* plays a pivotal role in tumorigenesis. The other hub gene, *FTO* is a m⁶A demethylase associated with tumorigenesis in lung cancer and *FTO* down-regulation promotes epithelial-to-mesenchymal transition (EMT) by regulating Wnt/ β -catenin signaling [15]. From the perspective of targeted therapy using non-CAGAs, we suggest that the identified prognostic marker genes, the genes identified by DEG analysis, and genes in the clinically relevant modules identified by WGCNA according to second-level cluster labels all show considerable promise. We also suggest that the global molecular mechanisms underlying non-CAGAs cancer onset may involve *MAML2*-related signaling pathways such as Notch and Wnt/ β -catenin, however, we cannot exclude other possibilities and further investigation is required using gene knockout studies. Regarding the poor and better survival subgroups in non-CAGA samples divided by second-level cluster

labels, immune check point inhibitors might be an option for the poor survival subgroup because patients in this group exhibited high TMB, which is associated with a favorable response to drugs in general.

Conclusion

Abbreviations

CAF	Cancer-associated fibroblast
ChIP-seq	Chromatin-immunoprecipitation sequencing
CNA	Copy number alteration
DEG	differentially expressed gene
DNMT	DNA methyltransferase
ENCODE	Encyclopedia of DNA Elements
FANTOM	The Functional Annotation of the Mammalian Genome
GSEA	Gene Set Enrichment Analysis
GO	Gene ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
LUAD	Lung adenocarcinoma
ME	Module eigengene
MEC	Mucoepidermoid cancer
NCC	National Cancer Center
Non-CAGAs	Non-clinically actionable genetic alterations
NSCLC	Non-small cell lung cancer
OS	Overall survival
RNA-seq	RNA-sequencing
SV	Structural variation
TCGA	The Cancer Genome Atlas
TET	Tet methylcytosine dioxygenase
TMB	Tumor mutation burden
UTR	untranslated region
WES	Whole-exosome sequencing
WGCNA	Weighted gene correlation (co-expression) network analysis
WGS	Whole-genome sequencing

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12943-024-02093-w>.

Supplementary Material 1

Supplementary Material 2

Supplementary Material 3

Acknowledgements

We thank Dr. Damiano Fantini for providing helpful suggestions related to the study. We thank Drs. Erik Bergstrom, SM Islam, and Burcak Otlu at the Alexandrov Laboratory for providing technical feedback related to the analysis. We greatly thank Drs. Shinji Kosaka and Kazuya Kanemochi for their support in this study. We thank the members of the Hamamoto laboratory and a member of the Kohno laboratory, Ms. Maiko Matsuda, for their kind assistance.

Author contributions

K.A., S.K., K.T., K.S., T.K., and R.H. designed the study. K.A., S.K., K.T., K.S., N.S., M.Ma., Y.S., S.T., H.M., K.K., A.B., M.K., M.Y., A.T., T.M., Y.Ok., M.Mu., T.Y., Y.Yo., and H.H. performed the data analysis. S.W., Y.Oh., Y.Ya., T.K., and R.H. supervised this study. K.A. wrote the manuscript, T.K. and R.H. edited the manuscript. All authors contributed to the interpretation of the data and critically revised the manuscript. All authors have read and approved the final version of the manuscript.

Funding

This work was supported by JSPS KAKENHI (Grant Number JP22K07180) and Takeda Science Foundation to K.A., the Cabinet Office BRIDGE (programs for bridging the gap between R&D and the ideal society (Society 5.0) and generating economic and social value), the AMED Innovative Cancer Medical Practice Research Project (Grant Number JP22ck0106643), JST CREST (Grant Number JPMJCR1689), JSPS Grant-in-Aid for Scientific Research on

Innovative Areas (Grant Number JP18H04908), JST AIP-PRISM (Grant Number JPMJCR18Y4), and MEXT subsidy for Advanced Integrated Intelligence Platform to R.H.

Data availability

No datasets were generated or analysed during the current study.

Declarations

Ethics approval and consent to participate

All methods were performed in accordance with the ethical guidelines for medical and health research involving human subjects. Informed consent was obtained from all patients. This study was approved by the Institutional Review Board of the National Cancer Center (NCC) Hospital (2005–109, 2016–496, 2019-018). The study was conducted in accordance with the principles of the Declaration of Helsinki.

Consent for publication

All authors revised and approved the manuscript.

Competing interests

The authors declare no competing interests.

Author details

¹Division of Medical AI Research and Development, National Cancer Center Research Institute, Tokyo 104-0045, Japan

²Cancer Translational Research Team, RIKEN Center for Advanced Intelligence Project, Tokyo 103-0027, Japan

³Division of Genome Biology, National Cancer Center Research Institute, Tokyo 104-0045, Japan

⁴Department of Endoscopy, National Cancer Center Hospital, Tokyo 104-0045, Japan

⁵Department of Diagnostic Radiology, National Cancer Center Hospital, Tokyo 104-0045, Japan

⁶Department of Thoracic Oncology, National Cancer Center Hospital, Tokyo 104-0045, Japan

⁷Department of Experimental Therapeutics, National Cancer Center Hospital, Tokyo 104-0045, Japan

⁸Department of Thoracic Surgery, National Cancer Center Hospital, Tokyo 104-0045, Japan

⁹Division of Medical Informatics, National Cancer Center Hospital, Tokyo 104-0045, Japan

¹⁰Department of Diagnostic Pathology, National Cancer Center Hospital, Tokyo 104-0045, Japan

Received: 28 April 2024 / Accepted: 16 August 2024

Published online: 02 September 2024

References

1. Waarts MR, Stonestrom AJ, Park YC, Levine RL. Targeting mutations in cancer. *J Clin Invest.* 2022;132(8):e154943. <https://doi.org/10.1172/JCI154943>.
2. Saito M, Shiraishi K, Kunitoh H, Takenoshita S, Yokota J, Kohno T. Gene aberrations for precision medicine against lung adenocarcinoma. *Cancer Sci.* 2016;107(6):713–20. <https://doi.org/10.1111/cas.12941>.
3. Sholl LM, Aisner DL, Varela-Garcia M, Berry LD, Dias-Santagata D, Wistuba II, Chen H, Fujimoto J, Kugler K, Franklin WA, et al. Multi-institutional Oncogenetic Driver Mutation Analysis in Lung Adenocarcinoma: the Lung Cancer

- Mutation Consortium Experience. *J Thorac Oncol.* 2015;10(5):768–77. <https://doi.org/10.1097/JTO.0000000000000516>.
4. Kaneko S, Takasawa K, Asada K, Shiraishi K, Ikawa N, Machino H, Shinkai N, Matsuda M, Masuda M, Adachi S, et al. Mechanism of ERBB2 gene overexpression by the formation of super-enhancer with genomic structural abnormalities in lung adenocarcinoma without clinically actionable genetic alterations. *Mol Cancer.* 2024;23(1):126. <https://doi.org/10.1186/s12943-024-02035-6>.
5. Kumar S, Warrell J, Li S, McGillivray PD, Meyerson W, Salichos L, Harmanci A, Martinez-Fundichely A, Chan CWY, Nielsen MM, et al. Passenger mutations in more than 2,500 Cancer genomes: overall molecular functional impact and consequences. *Cell.* 2020;180(5):915–e2716. <https://doi.org/10.1016/j.cell.2020.01.032>.
6. Sinha A, Fan VB, Ramakrishnan AB, Engelhardt N, Kennell J, Cadigan KM. Repression of Wnt/beta-catenin signaling by SOX9 and mastermind-like transcriptional coactivator 2. *Sci Adv.* 2021;7(8):eabe0849. <https://doi.org/10.1126/sciadv.abe0849>.
7. Ni W, Chen Z, Zhou X, Yang R, Yu M, Lu J, Kaye FJ, Wu L. Targeting notch and EGFR signaling in human mucoepidermoid carcinoma. *Signal Transduct Target Ther.* 2021;6(1):27. <https://doi.org/10.1038/s41392-020-00388-0>.
8. Lubecka K, Kurzava L, Flower K, Buvala H, Zhang H, Teegarden D, Camarillo I, Suderman M, Kuang S, Andrisani O, et al. Stilbenoids remodel the DNA methylation patterns in breast cancer cells and inhibit oncogenic NOTCH signaling through epigenetic regulation of MAML2 transcriptional activity. *Carcinogenesis.* 2016;37(7):656–68. <https://doi.org/10.1093/carcin/bgw048>.
9. Wang Y, Lu D, Chung YJ, Xu S. Genetic structure, divergence and admixture of Han Chinese, Japanese and Korean populations. *Hereditas.* 2018;155:19. <https://doi.org/10.1186/s41065-018-0057-5>.
10. Jiang SS, Fang WT, Hou YH, Huang SF, Yen BL, Chang JL, Li SM, Liu HP, Liu YL, Huang CT, et al. Upregulation of SOX9 in lung adenocarcinoma and its involvement in the regulation of cell growth and tumorigenicity. *Clin Cancer Res.* 2010;16(17):4363–73. <https://doi.org/10.1158/1078-0432.CCR-10-0138>.
11. Reda M, Ngamcherdtrakul W, Nelson MA, Siriwon N, Wang R, Zaidan HY, Bejan DS, Reda S, Hoang NH, Crumrine NA, et al. Development of a nanoparticle-based immunotherapy targeting PD-L1 and PLK1 for lung cancer treatment. *Nat Commun.* 2022;13(1):4261. <https://doi.org/10.1038/s41467-022-31926-9>.
12. Guo J, Wu Y, Du J, Yang L, Chen W, Gong K, Dai J, Miao S, Jin D, Xi S. Deregulation of UBE2C-mediated autophagy repression aggravates NSCLC progression. *Oncogenesis.* 2018;7(6):49. <https://doi.org/10.1038/s41389-018-0054-6>.
13. Al Amri WS, Allinson LM, Baxter DE, Bell SM, Hanby AM, Jones SJ, Shaaban AM, Stead LF, Verghese ET, Hughes TA. Genomic and expression analyses define MUC17 and PCNX1 as predictors of Chemotherapy response in breast Cancer. *Mol Cancer Ther.* 2020;19(3):945–55. <https://doi.org/10.1158/1535-7163.MCT-19-0940>.
14. Bhandari V, Hoey C, Liu LY, Lalonde E, Ray J, Livingstone J, Lesurf R, Shiah YJ, Vujcic T, Huang X, et al. Molecular landmarks of tumor hypoxia across cancer types. *Nat Genet.* 2019;51(2):308–18. <https://doi.org/10.1038/s41588-018-0318-2>.
15. Jeschke J, Collignon E, Al Wardi C, Krayem M, Bizet M, Jia Y, Garaud S, Wimana Z, Calonne E, Hassabi B, et al. Downregulation of the FTO m(6)A RNA demethylase promotes EMT-mediated progression of epithelial tumors and sensitivity to wnt inhibitors. *Nat Cancer.* 2021;2(6):611–28. <https://doi.org/10.1038/s43018-021-00223-7>.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.