doi:10.4149/neo_2020_200203N103

Comparative molecular profiling of distant metastatic and non-distant metastatic lung adenocarcinoma

Zhiming CHEN[1], Gang CHEN[2], Zhongqi WANG[3], Jianing YU[4], Henghui ZHANG[5], Beibei MAO[4], Haitao MA[1,*]

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Soochow University, Suzhou, China; [2]Department of Thoracic Surgery, Huashan Hospital, Shanghai, China; [3]Department of Medical Oncology, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China; [4]Beijing Genecast Biotechnology Co, Beijing, China; [5]Institute of Infectious Diseases, Beijing Key Laboratory of Emerging Infectious Diseases, Beijing Ditan Hospital, Capital Medical University, Beijing, China

**Correspondence: mht7403@163.com*

Received February 3, 2020 / Accepted July 21, 2020

Most lung cancer deaths are caused by a distant disseminated disease rather than primary tumors. Understanding the biology behind distant metastasis (DM) is crucial for the effective prediction and reduction of recurrence rates. Genome-wide analysis of the tumor provides a new way to explore the pathogenesis and molecular diagnosis of metastasis in lung adenocarcinoma. In our study, a total of 215 eligible lung adenocarcinoma patients were enrolled. The DNA was extracted from formalin-fixed paraffin-embedded (FFPE) samples from the primary tumors of these patients. Comprehensive molecular profiling was performed using a panel covering the exome of lung cancer-associated driver genes based on targeted next-generation sequencing. Tumor gene alterations were analyzed to investigate the differences in molecular features between lung adenocarcinomas with or without DM. Patients with DM of lung adenocarcinoma had significantly more variations in overall copy number (defined as Copy Number Alteration (CNA) load and Copy Number Instability (CNI) score). Interestingly, the study of the relationship between copy number variation and other molecular features verified that the degree of copy number variation has a positive correlation with mutations of DNA damage repair pathway (DDR). Thus, the additional analysis further revealed that metastatic patients accumulated more mutations in the DDR pathway, suggesting that impaired function of the DDR pathway and copy number variations play important roles in the invasion process of cancer cells. A comprehensive genetic analysis of lung adenocarcinoma revealed significant genomic changes between DM and non-DM patients. This finding may shed new light on the elucidation of lung cancer invasion mechanisms, and provide potential predictors for metastatic lung cancer.

Key words: distant metastasis, CNA load, CNI score, DDR pathway, lung adenocarcinoma

Lung cancer is the most common malignant cancer in the world and is still one of the main causes of cancer-related deaths [1, 2]. Distant metastasis (DM) causes more than 90% of non-small cell lung cancer (NSCLC) deaths [3]. Although some patients were not diagnosed with distant metastasis at the time of initial consultation, distant recurrence was found in these patients soon after surgery. Currently, the diagnosis of DM in cancer patients mainly depends on imaging techniques [4, 5, 6]. However, these techniques cannot be used for the early diagnosis or occasional monitoring of metastasis.

Metastasis is a multistep cell-biological process with the initial step of local invasion, then intravasation into the circulatory system, and ultimately colonization at distant sites [7]. Several studies have discovered the risk factors related to metastases from lung cancer. In general, a solid histologic appearance, tumor size, and higher tumor stages are important for DM [8, 9]. Regarding the molecular profiles, Wang et al. identified differentially expressed genes of primary squamous cell lung carcinoma patients with or without subsequent DM [10]. Many other individual proteins, such as CD44, E-cadherin, and KIF1C, are differentially expressed in brain metastasis and primary lung tumor specimens [11–13]. However, as most of the studies focused on searching the distinctive characteristics at the RNA or protein level, the findings still lack evidence for genetic alterations critical for metastasis.

Next-generation sequencing (NGS) technologies provide an unprecedented high-throughput level of genetic information and are widely used for the mutational analysis of tumors for clinical and research applications [14, 15]. Targeted NGS allows the sequencing of thousands of genes with high read depth and has become a powerful tool for detecting complex and heterogeneous gene mutations [16, 17].

In the present study, targeted NGS was performed to compare the genetic aberrations in the primary tumors of patients with or without DM from two independent cohorts of lung adenocarcinoma. The comparative profiles provide an improved understanding of the genetic mechanism underlying lung cancer metastasis, identifying potential predictive factors for distant metastasis of lung adenocarcinoma.

Patients and methods

Patient information and sample collection. The study population consisted of 54 patients in cohort 1 and 161 patients in cohort 2 with lung adenocarcinoma. The patients in cohorts 1 and 2 were recruited from The First Affiliated Hospital of Soochow University and Huashan Hospital, respectively. Clinically recorded information, such as age, sex, tumor histology, and pathologic stage, was collected. Patients with other malignancies prior to or at the time of their lung cancer diagnosis or non-lung adenocarcinoma were excluded. All patients provided written informed consent for molecular analysis of the tissue samples.

DNA extraction and sequencing. Formalin-fixed paraffin-embedded (FFPE) tissue specimens of the primary tumors and matched whole blood DNA were collected from each patient for analysis. DNA was isolated from FFPE tissue specimens with the black PREP FFPE DNA Kit (Analytik Jena, Germany) according to the manufacturers' instructions. FFPE sample matched blood lymphocytes were isolated by centrifugation of whole blood at 1600×g for 10 min at room temperature. Tiangen whole blood DNA kits (Tiangen, Beijing, PRC) were used to extract DNA from FFPE sample matched peripheral blood lymphocytes according to the manufacturer's instructions. Genomic DNA was sheared into 150–200 bp fragments with a Covaris M220 Focused-Ultrasonicator (Covaris, Massachusetts, USA). Fragmented DNA libraries were constructed with a KAPA HTP Library Preparation Kit (Illumina Platform, KAPA Biosystems, Massachusetts, USA) according to the manufacturer's instructions. DNA libraries were captured with a designed 1.6 M panel of the NimbleGen SeqCap EZ Library (Roche, Wisconsin, USA), which included major tumor-related genes. The captured samples were then subjected to Novaseq 6000 processing for paired-end sequencing.

Variant calling. We used VarScan2 (v2.4.2) to call somatic SNVs and indels on tumor samples and matched blood samples. Paired gDNA samples were used as a control to distinguish somatic mutations from inherited germline variation. The following filters were applied: i) number of mutant allele reads >2; ii) coverage in normal >50 and coverage in tumor >100; iii) mutant allele frequency >2%; iv) nonsynonymous SNVs and indels; v) located in exon regions; and vi) allele frequency <0.5% in the exac03 database.

We used CNVkit (v0.9.2) to obtain the log_2 *copy ratio* from the tumor samples for each patient and each gene. A panel of blood healthy control samples was used for reference construction. A gene was defined as copy number gain (log_2 *copy ratio* >0.5) or loss (log_2 *copy ratio* <–$\log_2$ 4/3) only if the number of target intervals was greater than or equal to 5. CNA load was calculated as the number of copy number varied genes/per Mb for each patient.

CNI score calculation. For each target interval, the read counts, corrected by GC content and target interval length, were converted to and were transformed into Z-scores based on a baseline established by the healthy population data (n=30). Ignoring the target intervals where the Z-scores is less than the 95 percentiles plus twice the absolute standard deviation of the healthy group, and Z-scores of the remaining intervals were summed as the CNI score [18].

Statistical analysis. The Mann-Whitney U test was applied for comparisons of continuous variables between two groups. Fisher's exact test was used for comparisons of mutations and CNV frequencies between defined patient groups. A p-value <0.05 was considered statistically significant. The error bars represent the mean ± SD.

Results

Somatic mutations in patients with lung adenocarcinoma. According to the NCCN guidelines (version 1.2020) for Non-Small Cell Lung Cancer, we defined DM patients as those who had M1a or M1b or M1c (stage IVA and IVB) disease, non-DM patients as those who had M0 (stage I–III) disease. To investigate the molecular characteristics essential for DM, we first conducted targeted NGS on 54 primary tumors from patients with (n=26) or without (n=28) DM. The clinical characteristics of the patients are summarized in Table 1. In total, 1587 mutations were detected in 54 patients, producing an average of 29.4 mutations (1–517 mutations) per patient. 1587 mutations comprised 417 synonymous mutations and 1170 nonsynonymous mutations. However, no significant difference in the number of mutations was observed between the two groups (Figures 1A, 1B). Meanwhile, we observed an enrichment of C>T in both groups, denoted as signature 1 in the COSMIC Mutational Signature Framework (Figures 1C, 1D).

To compare the overall mutational landscape of patients with or without DM, we assessed the mutation frequencies of NSCLC-associated driver genes in all samples from the DM and non-DM patients. Generally, mutations in EGFR were the most recurrent event, present in DM samples from 18/26 (69%) of patients and in non-DM samples from 13/28 (46%) of patients. We also confirmed other frequently mutated driver genes, including TP53 in 29 patients (54%), ARID1A in 9 patients (17%), ERBB2 in 8 patients (15%), and PTEN in 7 patients (13%) (Figure 1E). However, regarding specific genes, the results of Fisher's exact test showed no significant associations between single gene mutations and DM.

Copy number variation is distinct between DM and non-DM lung adenocarcinoma patients. We next sought to determine if gene copy number conveyed metastatic informa-

Table 1. Clinical characteristics of patients in cohort 1.

Characteristics	Numbers of the patients (%)	DM	Non-DM	p-value (Fisher's exact test)
Gender				
Male	31 (57.4)	16	15	0.59
Female	23 (42.6)	10	13	
Age (years) (range from 34–78)				
>58	25 (46.3)	14	11	0.41
≤58	29 (53.7)	12	17	
Stage				
I	13 (24.1)	0	13	
II	8 (14.8)	0	8	
III	7 (13.0)	0	7	
IV	26 (48.1)	26	0	
T classification				
T1–T2	39 (72.2)	15	24	0.03
T3–T4	15 (27.8)	11	4	
Lymph node metastasis				
N0	19 (35.2)	4	15	0.02
N1–3	30 (55.6)	17	13	
Unknown	5 (9.2)	5	0	
Sites of Distant metastasis				
Brain	13 (24.1)	13	0	
Liver	1 (1.8)	1	0	
Bone	4 (7.4)	4	0	
Pleura	8 (14.8)	8	0	
Smoking status				
Smoking	29 (53.7)	14	15	>0.9999
Non-smoking	25 (46.3)	12	13	

tion. We first mapped the copy number variated landscape of the 54 patients. Copy number analysis uncovered deletions of tumor suppressor genes, including CDKN2B in 5 patients (9%) and CDKN2A in 5 patients (9%). Amplifications were observed for RICTOR in 21 patients (39%), RAC1 in 13 patients (24%), EGFR in 14 patients (26%), CCND1 in 14 patients (26%), and HRAS in 8 patients (15%) (Supplementary Figure S1).

To describe the degree to which the tumor genome copy number is altered, we evaluated the number of varied genes/Mb (CNA load) in the two groups. The median CNA load of 54 patients was 4.3 (range from 0–27.8). We noted markedly higher CNA load in patients with DM compared to patients without DM (Figure 2A). We then defined the high-load CNA group as the one with a CNA load at or above the median and a low-load CNA group as the one with a CNA load below the median. The results of Fisher's exact test showed that more patients with distant metastases had high CNA loads (Figure 2B).

The copy number instability (CNI) score, which was established by Weiss et al., is another measure of total copy number variation [18]. To confirm the association between copy number variation and DM of lung adenocarcinoma, we also estimated the CNI scores of primary tumors from 54 patients (median: 1435.505, ranged from 45.2–4348). As shown in Figure 2C, the patients in the DM group exhibited higher CNI scores. Accordingly, the proportion of patients with high CNI scores (≥ median) in the DM group was also higher than that in the non-DM group (Figure 2D), indicating the positive correlation between copy number variation and DM.

To determine the generality of our findings, an external cohort (cohort 2) consisting of 161 lung adenocarcinoma patients was used to further validate the relationship between copy number variation and DM. The baseline characteristics are listed in Supplementary Table 1. In agreement with the results of cohort 1, both CNA load and CNI score were much higher in patients of the DM group (Figures 2E, 2F). Meanwhile, patients with a high CNA load (≥6.6) and high CNI score (≥ median CNI, 1462.68) had a higher prevalence of metastatic disease (Figures 2G, 2H). Additional analysis revealed that high-grade T stage and lymph node metastasis were helpful for DM (Table 1), while the clinical characteristics such as gender, age, smoking status, T stage, and lymph node metastasis, had no impact on either CNA load or CNI score (Supplementary Figure S2).

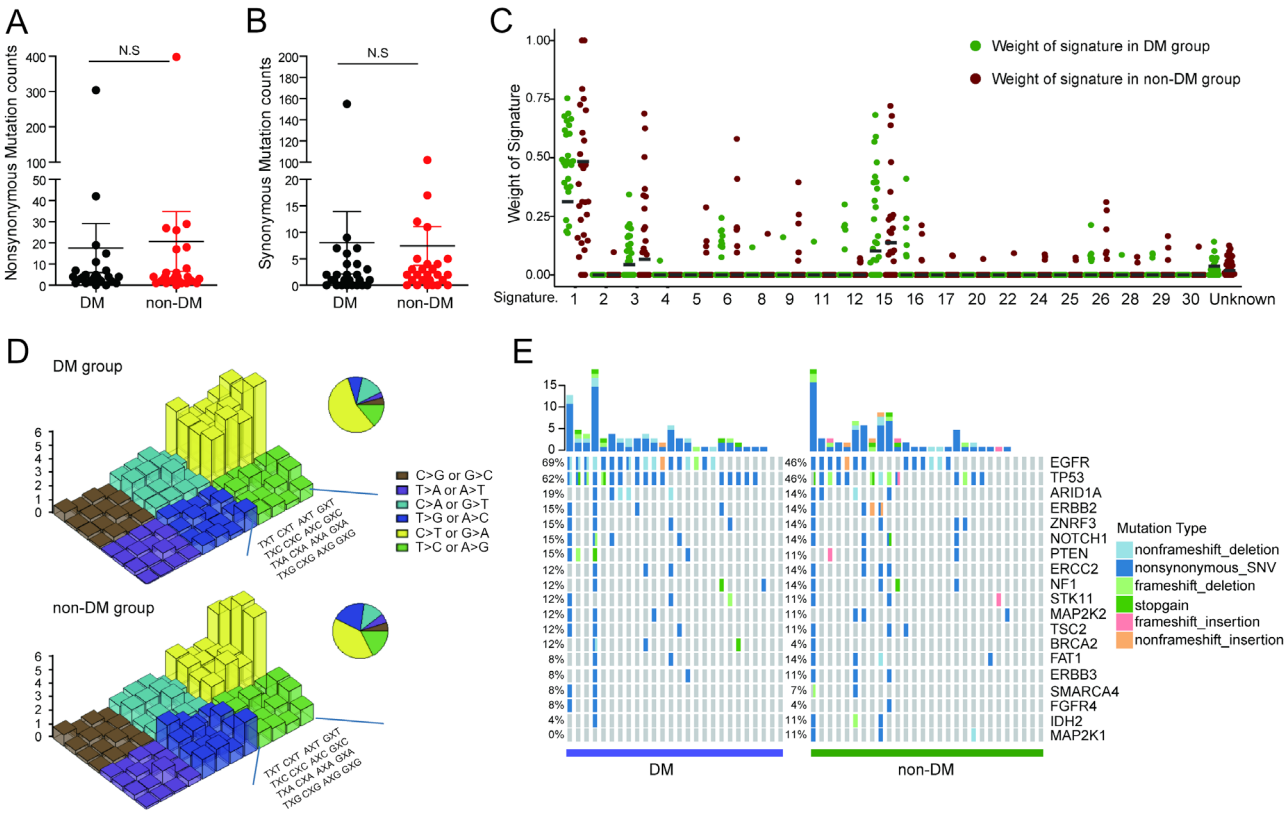


Figure 1. SMs in patients with lung adenocarcinoma. A, B) Comparison of the mutation number for DM and non-DM patients in cohort 1. C) The weight of decomposed signatures for each of the DM (green dots) and non-DM patients (red dots) of cohort 1. D) Lego plots display the frequencies of base substitutions within specific trinucleotide mutational contexts for DM samples (up) and non-DM samples (down) of cohort 1. E) Somatic mutation landscape of lung cancer-associated driver genes in the DM and non-DM patients of cohort 1. Samples are displayed as columns.

Relationships between copy number variation and other genetic features of lung adenocarcinoma. We next investigated which molecular properties of cancer cells might influence copy number variation using cohort 2 which comprised relatively more patients. As CNI and CNA load had a strong correlation with each other (Spearman R=0.91, p<0.0001), we divided patients into two groups according to the CNI score. We found that TP53, ATR, and EGFR were enriched for mutations in patients with high CNI scores that were ≥ the median level (Figures 3A, 3B). In addition, both CNA load and CNI score were higher in patients with TP53 or EGFR mutations than in wild-type patients (Figures 4A–4D).

ATR belongs to the DNA damage repair pathway. Nuclear EGFR (nEGFR) has also been reported to be involved in DNA repair through associations with various molecules e.g. DNA-PK and histones [19, 20]. Therefore, we wanted to determine whether the mutation of the DDR pathway contributes to the copy number variation levels. Statistical comparison showed a significant association between copy number variation and DDR pathway mutation. CNA loads and CNI scores were lower in patients without any detectable mutations in the 33 genes (Supplementary Table S2) of the DDR pathway (DDR WT group). Inversely, the DDR mutated group, with 1 or more than 1 mutated gene, had a higher CNA load and CNI score (Figures 4E, 4F), consistent with a putative role for this pathway in genomic stability conservation. In addition to the DDR pathway, we also assessed the impact of other oncogenic signaling pathways on CNA load and CNI score. As shown in Figures 4G and 4H, only mutations of the cell cycle pathway interfered with copy number variation.

DDR alterations have been reported to be prevalent in metastatic samples from individual datasets [21–24], indicating the role of the DDR pathway in tumor metastasis. However, it is still unknown whether the enrichment of DDR pathway mutations also exists in the primary tumors of patients with DM. Therefore, we examined the relationship of DDR pathway mutation in primary tumors with distant metastasis of lung adenocarcinoma. In a very high proportion (72%, 116/161) of the primary specimens, at least one mutation was identified in genes involved in the DDR pathway. Notably, many more genes in the DDR pathway were mutated in DM patients than in non-distant metastatic patients (Supplementary Figure S3A). Meanwhile, a larger proportion of the DDR mutated subgroup was observed in DM patients (Supplementary Figure S3B).

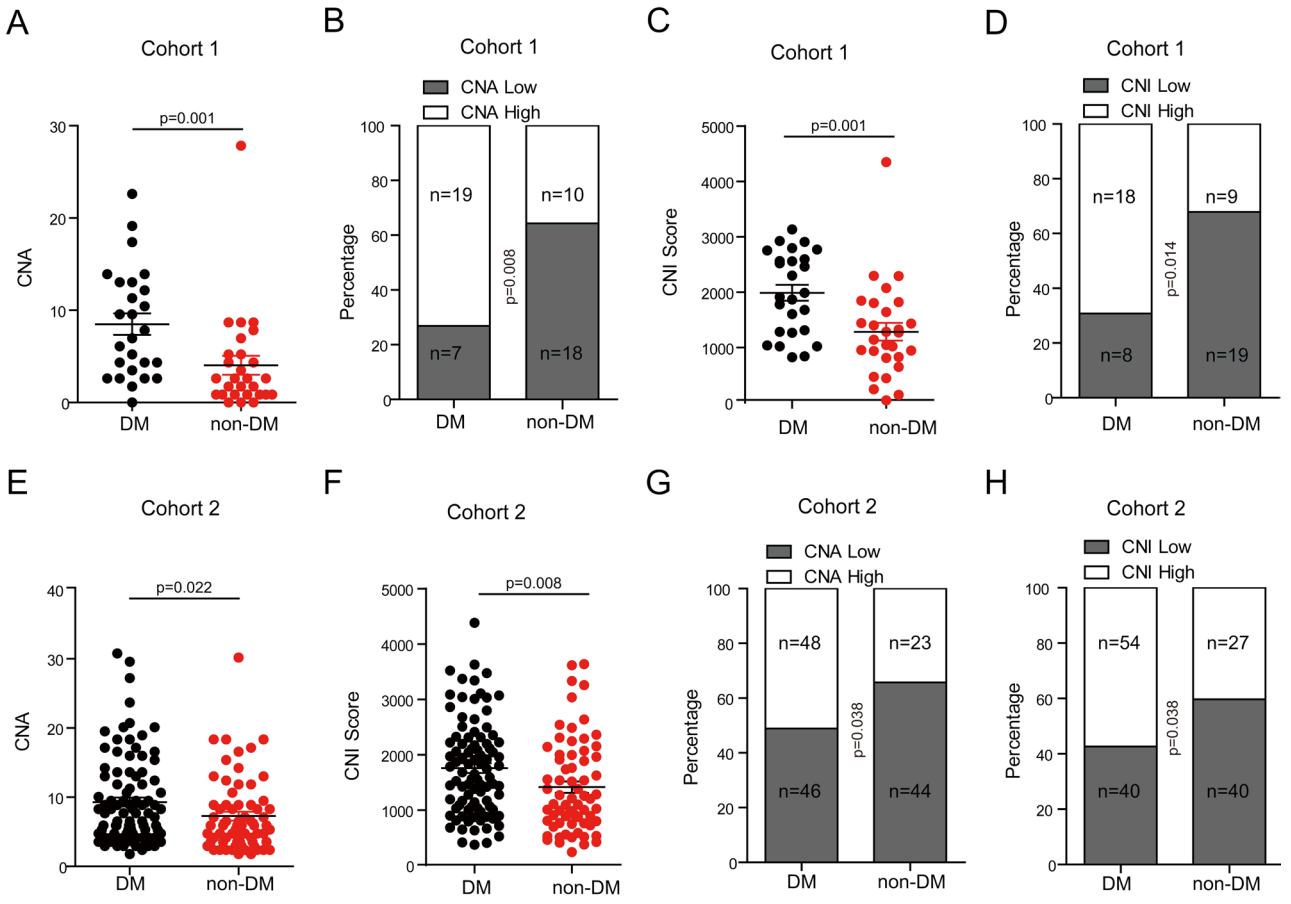


Figure 2. CNA load and CNI score are distinct in DM and non-DM patients of lung adenocarcinoma. A, C) Comparison of CNA load (A) or CNI score (C) in DM and non-DM patients of cohort 1. B, D) The bar plots illustrate the frequency of patients with high and low CNA load (B) or CNI score (D) in DMs versus non-DMs in cohort 1. E, F) The scatterplots show the range of CNA load (E) or CNI scores (F) for tumors from DM and non-DM patients in cohort 2. G, H) The bar plots illustrate the frequency of patients with high and low CNA load (G) or CNI score (H) in DMs versus non-DMs in cohort 2.

Discussion

Many studies have uncovered the mechanism and potential biomarkers for DM in lung cancer. However, the relationship between DM and genomic alterations harbored by the primary tumors of lung cancer is less well studied. Herein, we showed that patients with DM of lung adenocarcinoma had significantly more gain or loss of driver genes and higher copy number instability (CNI) in two cohorts. Furthermore, the study of the relationship between copy number variation and other molecular features verified that the degree of copy number variation was positively correlated with the level of DDR pathway mutation. Thus, the additional analysis further disclosed that metastatic patients accumulated more mutations in the DDR pathway, indicating that impaired function of the DDR pathway induced copy number variations plays an important role in the invasion process of cancer cells.

Molecular profiling of tumors revealed that copy number alterations are frequently observed in lung cancer [25–27]. Previous studies have identified the amplification or deletion of specific genes that are essential for tumor metastasis, including FGFR [28], LKB1 [29], and MET [30], in different lung cancer cohorts, suggesting that copy number variation of driver genes is essential for tumor metastasis. Herein, for the first time, we adopted two parameters, CNA load and CNI score, to examine the contribution of the overall copy number variations on DM in lung adenocarcinoma [18, 25]. And we presented evidence that CNA load and CNI score are highly associated with DM, further confirming the close relationship between the overall copy number variation and tumor malignancy. The presence of DM is a poor prognostic factor in various cancer types. Concordant with our observations, several studies have reported that the high level of copy number alteration is significantly associated with poor survival and outcomes in various cancer types, including primary prostate, breast, endometrial, pancreatic, and colorectal cancer [31–33].

Both coding mutations and CNA reflect genetic instability. The DDR pathway is considered to be the guardian for

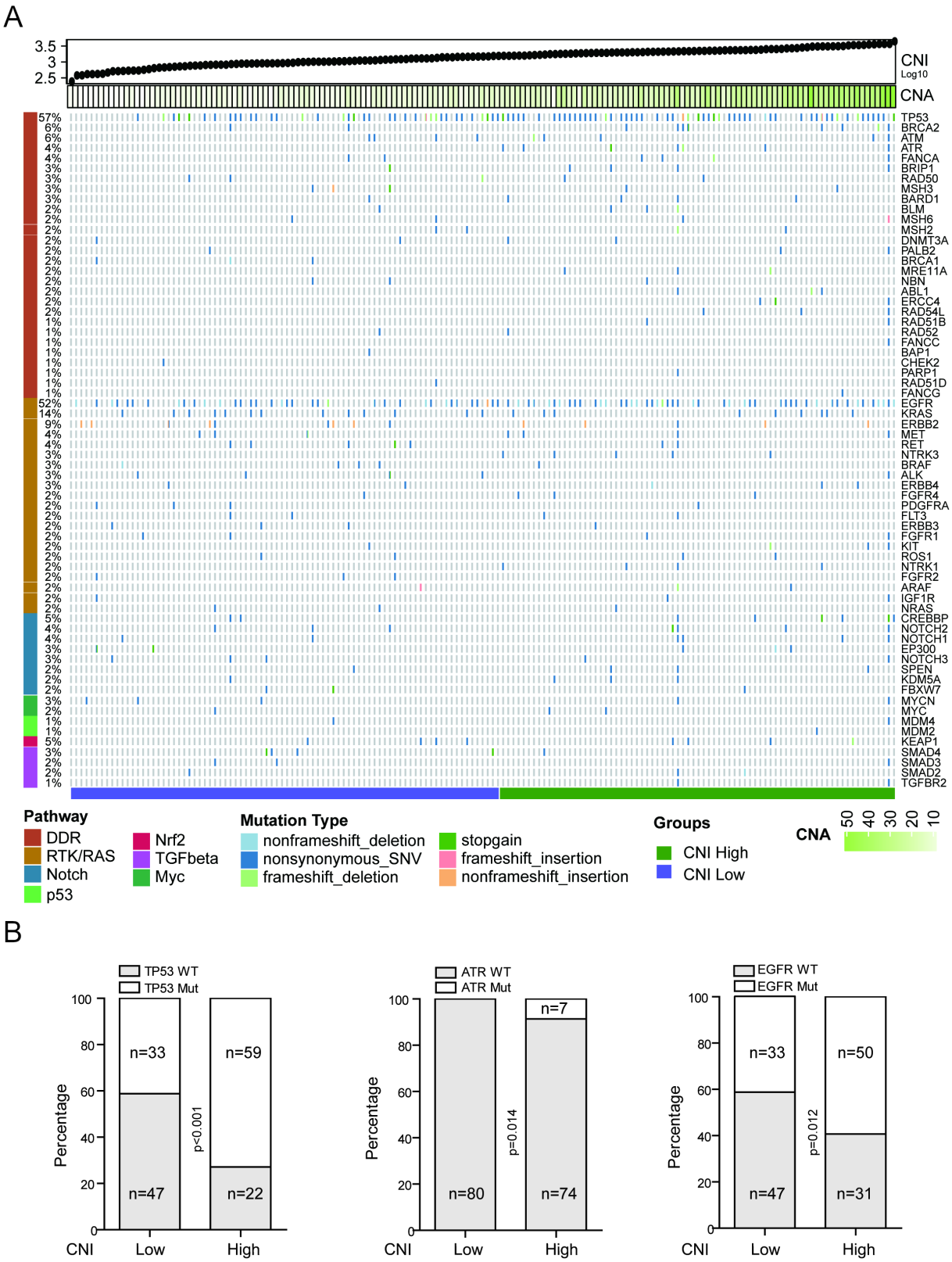


Figure 3. Mutational molecular features are different between CNI high and CNI low patients. A) The landscape of driver gene mutations in lung adenocarcinoma in 161 patients of cohort 2. B) The bar plots indicate the frequency of mutated genes in patients of cohort 2 with high and low CNI score.

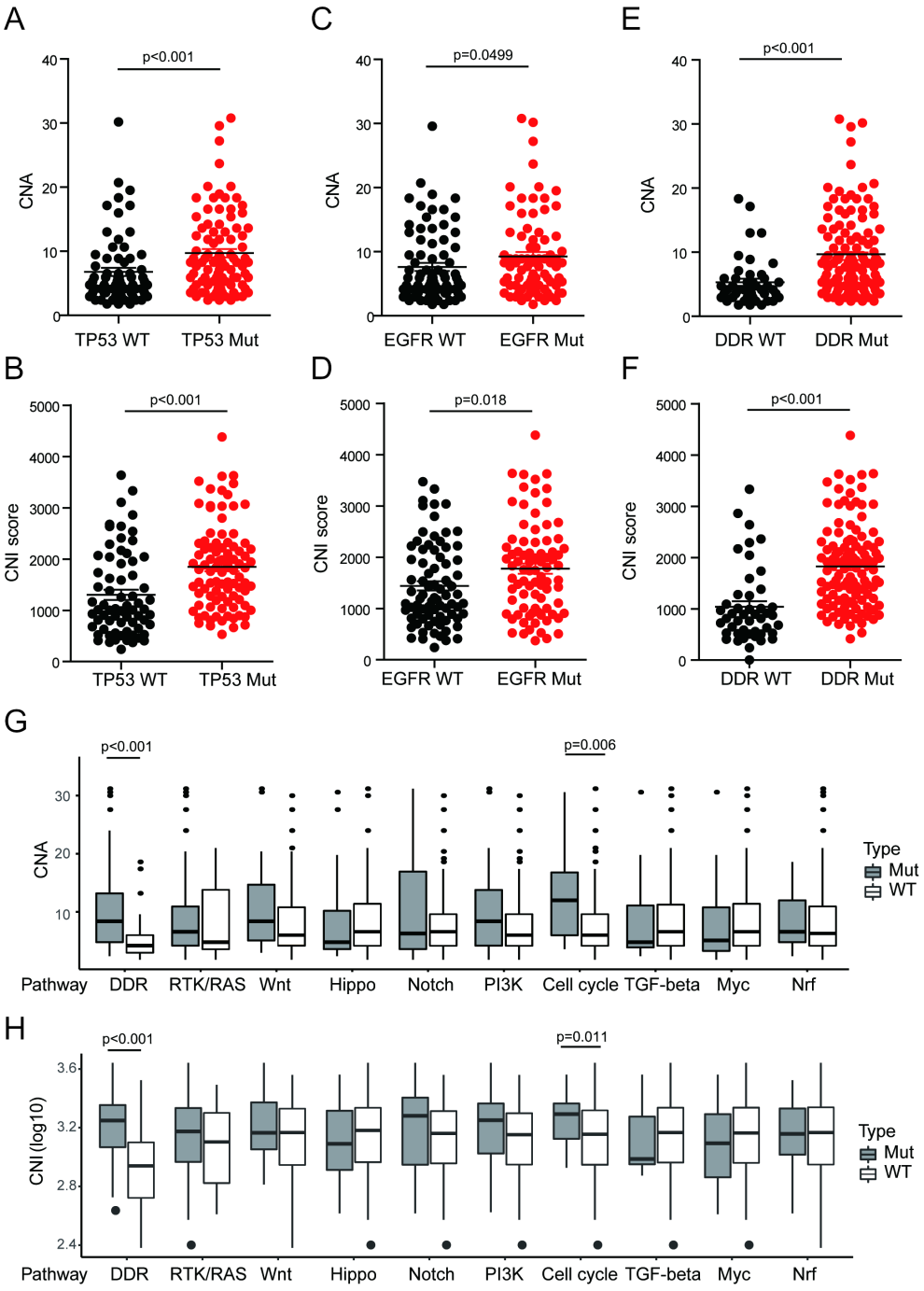


Figure 4. Relationship between copy number variation level and somatic mutations. A–D) The scatter plots show that CNA load and CNI score are higher in patients harboring TP53 (A, B) or EGFR (C, D) mutation in 161 patients of cohort 2. E, F) The scatter plots show that CNA load (E) and CNI score (F) are higher in patients with DDR pathway mutation in 161 patients of cohort 2. G, H) Analysis of the association between CNA load (G) or CNI score (H) and oncogenic signaling pathways in patients of cohort 2.

preserving the stability of genomic information [34]. Like TMB (tumor mutation burden), which has a positive correlation with the DDR pathway [35], copy number variation also obviously differed in patients with or without mutations in the DDR pathway. In agreement with our findings, Davoli et al. discovered that mutations in driver genes involved in the DDR pathway were positively correlated with SCNA levels in pan-cancer types [25]. Regarding other signaling pathways

or single genes essential for tumor development, only EGFR mutation and the cell cycle pathway were identified to affect copy number variation, probably due to that high activation of RTKs and an accelerated cell cycle can induce replication stress and DNA damage [35–37].

Based on the above discoveries, it is expected that metastatic patients have more DDR pathway mutations. Correspondingly, patients who harbored mutations in the DDR pathway more frequently had DM. Although numerous studies have investigated the role of genes involved in the DDR pathway in metastasis, most of them focused on the expression levels of these genes [38]. Comparatively less is known about how changes in the DDR pathway at the DNA level in the primary tumors affect lung cancer metastasis. Despite the fact that we did not identify single gene mutations in the DDR pathway that were significantly associated with metastasis, it is still intriguing that an accumulation of mutations in multiple genes in this pathway may facilitate the metastatic process.

In conclusion, we first revealed that overall copy number variations of driver genes were distinct in DM and non-DM patients. DDR pathway deficiency was positively related to higher copy number variation and DM. This finding may shed new light on the elucidation of lung cancer invasion, provide potential predictors and therapeutic targets for metastatic lung cancer.

Supplementary information is available in the online version of the paper.

Acknowledgments: This work was supported by the National Natural Science Foundation of China (81672281).

References

[1] RAMI-PORTA R, BOLEJACK V, CROWLEY J, BALL D, KIM J et al. The IASLC Lung Cancer Staging Project: Proposals for the Revisions of the T Descriptors in the Forthcoming Eighth Edition of the TNM Classification for Lung Cancer. J Thorac Oncol 2015; 10: 990–1003. https://doi.org/10.1097/JTO.0000000000000559

[2] SIEGEL RL, MILLER KD, JEMAL A. Cancer Statistics, 2017. CA Cancer J Clin 2017; 67: 7–30. https://doi.org/10.3322/caac.21387

[3] HERBST RS, HEYMACH JV, LIPPMAN SM. Lung cancer. N Engl J Med 2008; 359: 1367–1380. https://doi.org/10.1056/NEJMra0802714

[4] COROLLER TP, GROSSMANN P, HOU Y, RIOS VELAZQUEZ E, LEIJENAAR RT et al. CT-based radiomic signature predicts distant metastasis in lung adenocarcinoma. Radiother Oncol 2015; 114: 345–350. https://doi.org/10.1016/j.radonc.2015.02.015

[5] COSTELLOE CM, KUNDRA V, MA J, CHASEN BA, ROHREN EM et al. Fast Dixon whole-body MRI for detecting distant cancer metastasis: a preliminary clinical study. J Magn Reson Imaging 2012; 35: 399–408. https://doi.org/10.1002/jmri.22815

[6] WU J, AGUILERA T, SHULTZ D, GUDUR M, RUBIN DL et al. Early-Stage Non-Small Cell Lung Cancer: Quantitative Imaging Characteristics of (18)F Fluorodeoxyglucose PET/CT Allow Prediction of Distant Metastasis. Radiology 2016; 281: 270–278. https://doi.org/10.1148/radiol.2016151829

[7] VAN ZIJL F, KRUPITZA G, MIKULITS W. Initial steps of metastasis: cell invasion and endothelial transmigration. Mutat Res 2011; 728: 23–34. https://doi.org/10.1016/j.mrrev.2011.05.002

[8] HUNG JJ, JENG WJ, WU YC, CHOU TY, HSU WH. Factors predicting organ-specific distant metastasis in patients with completely resected lung adenocarcinoma. Oncotarget 2016; 7: 58261–58273. https://doi.org/10.18632/oncotarget.11338

[9] ZHOU H, DONG D, CHEN B, FANG M, CHENG Y et al. Diagnosis of Distant Metastasis of Lung Cancer: Based on Clinical and Radiomic Features. Transl Oncol 2018; 11: 31–36. https://doi.org/10.1016/j.tranon.2017.10.010

[10] WANG N, ZHOU F, XIONG H, DU S, MA J et al. Screening and identification of distant metastasis-related differentially expressed genes in human squamous cell lung carcinoma. Anat Rec (Hoboken) 2012; 295: 748–757. https://doi.org/10.1002/ar.22441

[11] KARGI HA, KUYUCUOGLU MF, ALAKAVUKLAR M, AKPINAR O, ERK S. CD44 expression in metastatic and non-metastatic non-small cell lung cancers. Cancer Lett 1997; 119: 27–30. https://doi.org/10.1016/s0304-3835(97)00254-1

[12] THEISEN U, STRAUBE E, STRAUBE A. Directional persistence of migrating cells requires Kif1C-mediated stabilization of trailing adhesions. Dev Cell 2012; 23: 1153–1166. https://doi.org/10.1016/j.devcel.2012.11.005

[13] YOO JY, YANG SH, LEE JE, CHO DG, KIM HK et al. E-cadherin as a predictive marker of brain metastasis in non-small-cell lung cancer, and its regulation by pioglitazone in a preclinical model. J Neurooncol 2012; 109: 219–227. https://doi.org/10.1007/s11060-012-0890-8

[14] Castera L, Krieger S, Rousselin A, Legros A, Baumann JJ et al. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. European journal of human genetics. Eur J Hum Genet 2014; 22: 1305–1313. https://doi.org/10.1038/ejhg.2014.16

[15] VRIJENHOEK T, KRAAIJEVELD K, ELFERINK M, DE LIGT J, KRANENDONK E et al. Next-generation sequencing-based genome diagnostics across clinical genetics centers: implementation choices and their effects. Eur J Hum Genet 2015; 23: 1142–1150. https://doi.org/10.1038/ejhg.2014.279

[16] DILLIOTT AA, FARHAN SMK, GHANI M, SATO C, LIANG E et al. Targeted Next-generation Sequencing and Bioinformatics Pipeline to Evaluate Genetic Determinants of Constitutional Disease. J Vis Exp 2018; 134: 57266. https://doi.org/10.3791/57266

[17] SAMSTEIN RM, LEE CH, SHOUSHTARI AN, HELLMANN MD, SHEN R et al. Tumor mutational load predicts survival after immunotherapy across multiple cancer types. Nat Genet 2019; 51: 202–206. https://doi.org/10.1038/s41588-018-0312-8

[18] WEISS GJ, BECK J, BRAUN DP, BORNEMANN-KOLATZKI K, BARILLA H et al. Tumor Cell-Free DNA Copy Number Instability Predicts Therapeutic Response to Immunotherapy. Clin Cancer Res 2017; 23: 5074–5081. https://doi.org/10.1158/1078-0432.CCR-17-0231

[19] CHOU RH, WANG YN, HSIEH YH, LI LY, XIA W et al. EGFR modulates DNA synthesis and repair through Tyr phosphorylation of histone H4. Dev Cell 2014; 30: 224–237. https://doi.org/10.1016/j.devcel.2014.06.008

[20] DITTMANN K, MAYER C, KEHLBACH R, RODEMANN HP. Radiation-induced caveolin-1 associated EGFR internalization is linked with nuclear EGFR transport and activation of DNA-PK. Mol Cancer 2008; 7: 69. https://doi.org/10.1186/1476-4598-7-69

[21] GRASSO CS, WU YM, ROBINSON DR, CAO X, DHANASEKARAN SM et al. The mutational landscape of lethal castration-resistant prostate cancer. Nature 2012; 487: 239–243. https://doi.org/10.1038/nature11125

[22] HONG MK, MACINTYRE G, WEDGE DC, VAN LOO P, PATEL K et al. Tracking the origins and drivers of subclonal metastatic expansion in prostate cancer. Nat Commun 2015; 6: 6605. https://doi.org/10.1038/ncomms7605

[23] ROBINSON D, VAN ALLEN EM, WU YM, SCHULTZ N, LONIGRO RJ et al. Integrative Clinical Genomics of Advanced Prostate Cancer. Cell 2015; 162: 454. https://doi.org/10.1016/j.cell.2015.06.053

[24] TAYLOR BS, SCHULTZ N, HIERONYMUS H, GOPALAN A, XIAO Y et al. Integrative genomic profiling of human prostate cancer. Cancer Cell 2010; 18: 11–22. https://doi.org/10.1016/j.ccr.2010.05.026

[25] DAVOLI T, UNO H, WOOTEN EC, ELLEDGE SJ. Tumor aneuploidy correlates with markers of immune evasion and with reduced response to immunotherapy. Science 2017; 355: eaaf8399. https://doi.org/10.1126/science.aaf8399

[26] JAMAL-HANJANI M, WILSON GA, MCGRANAHAN N, BIRKBAK NJ, WATKINS TBK et al. Tracking the Evolution of Non-Small-Cell Lung Cancer. N Engl J Med 2017; 376: 2109–2121. https://doi.org/10.1056/NEJMoa1616288

[27] LI F, SUN L, ZHANG S. Acquirement of DNA copy number variations in non-small cell lung cancer metastasis to the brain. Oncol Rep 2015; 34: 1701–1707. https://doi.org/10.3892/or.2015.4188

[28] PREUSSER M, BERGHOFF AS, BERGER W, ILHAN-MUTLU A, DINHOF C et al. High rate of FGFR1 amplifications in brain metastases of squamous and non-squamous lung cancer. Lung Cancer 2014; 83: 83–89. https://doi.org/10.1016/j.lungcan.2013.10.004

[29] ZHAO N, WILKERSON MD, SHAH U, YIN X, WANG A et al. Alterations of LKB1 and KRAS and risk of brain metastasis: comprehensive characterization by mutation analysis, copy number, and gene expression in non-small-cell lung carcinoma. Lung Cancer 2014; 86: 255–261. https://doi.org/10.1016/j.lungcan.2014.08.013

[30] FANG L, CHEN H, TANG Z, KALHOR N, LIU CH et al. MET amplification assessed using optimized FISH reporting criteria predicts early distant metastasis in patients with non-small cell lung cancer. Oncotarget 2018; 9: 12959–12970. https://doi.org/10.18632/oncotarget.24430

[31] HIERONYMUS H, MURALI R, TIN A, YADAV K, ABIDA W et al. Tumor copy number alteration burden is a pan-cancer prognostic factor associated with recurrence and death. Elife 2018; 7: e37294. https://doi.org/10.7554/eLife.37294

[32] MIRCHIA K, SATHE AA, WALKER JM, FUDYM Y, GALBRAITH K et al. Total copy number variation as a prognostic factor in adult astrocytoma subtypes. Acta Neuropathol Commun 2019; 7: 92. https://doi.org/10.1186/s40478-019-0746-y

[33] SMITH JC, SHELTZER JM. Systematic identification of mutations and copy number alterations associated with cancer patient prognosis. Elife 2018; 7: e39217. https://doi.org/10.7554/eLife.39217

[34] BARTEK J. DNA damage response, genetic instability and cancer: from mechanistic insights to personalized treatment. Mol Oncol 2011; 5: 303–307. https://doi.org/10.1016/j.molonc.2011.07.006

[35] WANG Z, ZHAO J, WANG G, ZHANG F, ZHANG Z et al. Comutations in DNA Damage Response Pathways Serve as Potential Biomarkers for Immune Checkpoint Blockade. Cancer Res 2018; 78: 6486–6496. https://doi.org/10.1158/0008-5472.CAN-18-1814

[36] BARTKOVA J, REZAEI N, LIONTOS M, KARAKAIDOS P, KLETSAS D et al. Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints. Nature 2006; 444: 633–637. https://doi.org/10.1038/nature05268

[37] DI MICCO R, FUMAGALLI M, CICALESE A, PICCININ S, GASPARINI P et al. Oncogene-induced senescence is a DNA damage response triggered by DNA hyper-replication. Nature 2006; 444: 638–642. https://doi.org/10.1038/nature05327

[38] HALAZONETIS TD, GORGOULIS VG, BARTEK J. An oncogene-induced DNA damage model for cancer development. Science 2008; 319: 1352–1355. https://doi.org/10.1126/science.1140735

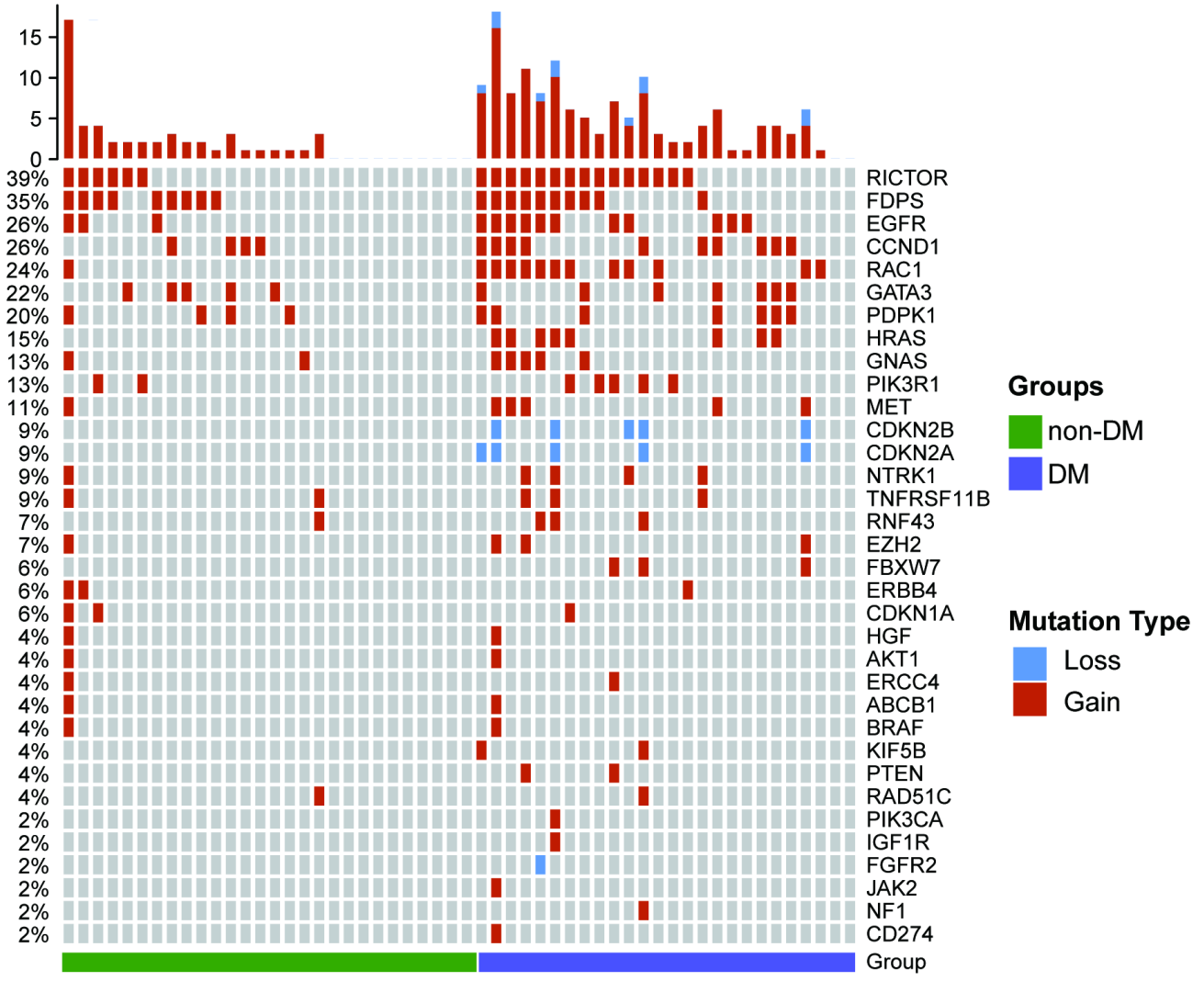
[39] KAUFFMANN A, ROSSELLI F, LAZAR V, WINNEPENNINCKX V, MANSUET-LUPO A et al. High expression of DNA repair pathways is associated with metastasis in melanoma patients. Oncogene 2008; 27: 565–573. https://doi.org/10.1038/sj.onc.1210700

https://doi.org/10.4149/neo_2020_200203N103

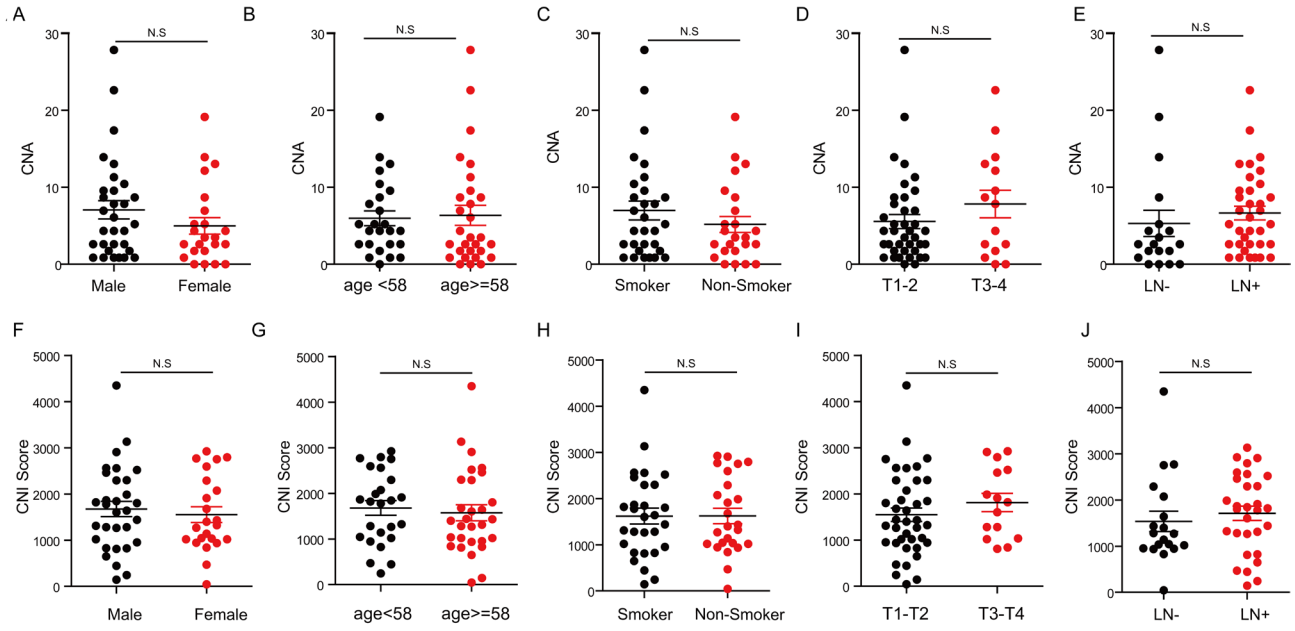
Comparative molecular profiling of distant metastatic and non-distant metastatic lung adenocarcinoma

Zhiming CHEN[1], Gang CHEN[2], Zhongqi WANG[3], Jianing YU[4], Henghui ZHANG[5], Beibei MAO[4], Haitao MA[1,*]

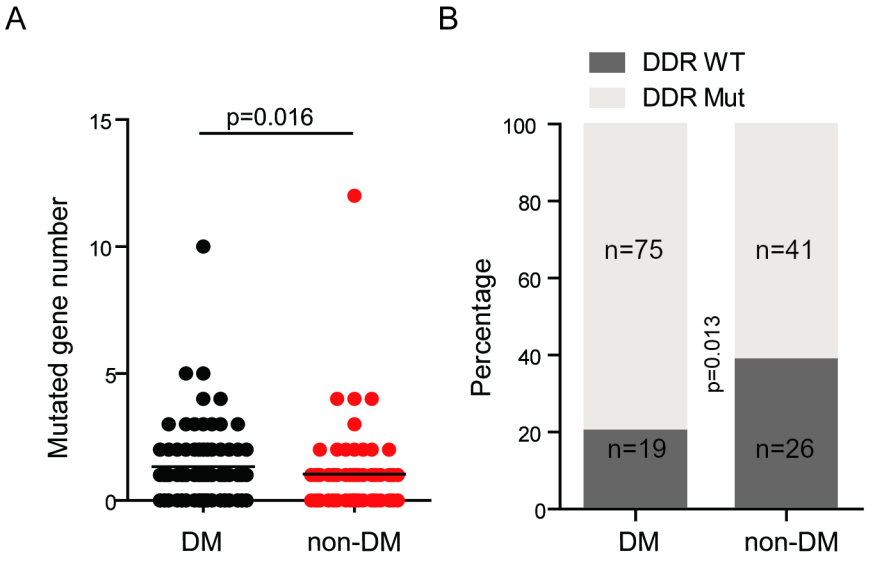
Supplementary Information



Supplementary Figure S1. Landscape illustrates copy number variation of driver genes in lung adenocarcinoma. 54 patients of cohort 1 are classified into DM and non-DM groups. Samples are distributed along the x-axis. The number of copy number varied genes in each patient is shown (columns). red: copy number gain; blue: copy number loss



Supplementary Figure S2. CNA level is not correlated with clinical features. CNA load (A–E) and CNI score (F–J) of patients in cohort 1 are identical in different groups stratified by indicated clinical characteristics. LN–: without lymph node metastasis; LN+: with lymph node metastasis



Supplementary Figure S3. Comparison of DDR pathway mutation in DM and non-DM patients. A) The scatter plots indicate the positive correlation between distant metastasis and DDR pathway mutation in cohort 2. B) Frequency of DDR pathway mutation in DM and non-DM patients of cohort 2.

Supplementary Table S1. Baseline characteristics of Cohort 2.

Factors		Cohort 2 (n=161)	DM (n=94)	Non-DM (n=67)
Age				
	median			
	<59	78 (48.4%)	45	33
	≥ 59	83 (51.6%)	49	34
Gender				
	Male	93 (57.8%)	56	37
	Female	68 (42.2%)	38	30
Stage				
	I	13 (8.1%)	0	13
	II	9 (5.6%)	0	9
	III	45 (27.9%)	0	45
	IV	94 (58.4%)	94	0

Supplementary Table S2. List of genes and pathways.

Gene	Pathway	Gene	Pathway	Gene	Pathway	Gene	Pathway	Gene	Pathway
CDKN2A	Cell-cycle	**CHEK2**	DDR	**NOTCH1**	Notch	**RICTOR**	PI3K	**NF1**	RTK/RAS
CDKN2B	Cell-cycle	**DNMT3A**	DDR	**NOTCH2**	Notch	**MTOR**	PI3K	**PTPN11**	RTK/RAS
CDKN2C	Cell-cycle	**ERCC4**	DDR	**NOTCH3**	Notch	**RPTOR**	PI3K	**KRAS**	RTK/RAS
CDKN1A	Cell-cycle	**FANCA**	DDR	**CREBBP**	Notch	**EGFR**	RTK/RAS	**HRAS**	RTK/RAS
CDKN1B	Cell-cycle	**FANCC**	DDR	**EP300**	Notch	**ERBB2**	RTK/RAS	**NRAS**	RTK/RAS
CCNE1	Cell-cycle	**FANCG**	DDR	**SPEN**	Notch	**ERBB3**	RTK/RAS	**ARAF**	RTK/RAS
RB1	Cell-cycle	**FANCI**	DDR	**KDM5A**	Notch	**ERBB4**	RTK/RAS	**BRAF**	RTK/RAS
CCND1	Cell-cycle	**NBN**	DDR	**KEAP1**	Nrf2	**MET**	RTK/RAS	**RAF1**	RTK/RAS
CCND2	Cell-cycle	**PALB2**	DDR	**CUL3**	Nrf2	**PDGFRA**	RTK/RAS	**RAC1**	RTK/RAS
CCND3	Cell-cycle	**PARP1**	DDR	**NFE2L2**	Nrf2	**FGFR1**	RTK/RAS	**MAPK1**	RTK/RAS
CDK4	Cell-cycle	**RAD51**	DDR	**MDM2**	p53	**FGFR2**	RTK/RAS	**MAP2K1**	RTK/RAS
CDK6	Cell-cycle	**RAD51B**	DDR	**MDM4**	p53	**FGFR3**	RTK/RAS	**MAP2K2**	RTK/RAS
FAT1	Hippo	**RAD51C**	DDR	**PTEN**	PI3K	**FGFR4**	RTK/RAS	**TGFBR2**	TGFbeta
NF2	Hippo	**RAD51D**	DDR	**INPP4B**	PI3K	**KIT**	RTK/RAS	**SMAD2**	TGFbeta
ABL1	DDR	**MRE11A**	DDR	**PIK3CA**	PI3K	**IGF1R**	RTK/RAS	**SMAD3**	TGFbeta
ATM	DDR	**RAD50**	DDR	**PIK3CB**	PI3K	**RET**	RTK/RAS	**SMAD4**	TGFbeta
ATR	DDR	**RAD52**	DDR	**PIK3R2**	PI3K	**ROS1**	RTK/RAS	**RNF43**	Wnt
BAP1	DDR	**RAD54L**	DDR	**PIK3R1**	PI3K	**ALK**	RTK/RAS	**AXIN1**	Wnt
BARD1	DDR	**MSH2**	DDR	**AKT1**	PI3K	**FLT3**	RTK/RAS	**AMER1**	Wnt
BLM	DDR	**MSH3**	DDR	**AKT2**	PI3K	**NTRK1**	RTK/RAS	**CTNNB1**	Wnt
BRCA1	DDR	**MSH6**	DDR	**AKT3**	PI3K	**NTRK2**	RTK/RAS	**GSK3B**	Wnt
BRCA2	DDR	**TP53**	DDR	**PPP2R1A**	PI3K	**NTRK3**	RTK/RAS	**APC**	Wnt
BRIP1	DDR	**MYC**	Myc	**STK11**	PI3K	**JAK2**	RTK/RAS		
CDK12	DDR	**MYCN**	Myc	**TSC1**	PI3K	**CBL**	RTK/RAS		
CHEK1	DDR	**FBXW7**	Notch	**TSC2**	PI3K	**ERRFI1**	RTK/RAS		