Original Article

Low-depth whole genome sequencing reveals copy number variations associated with higher pathologic grading and more aggressive subtypes of lung non-mucinous adenocarcinoma

Zheng Wang[1], Lin Zhang[1], Lei He[1], Di Cui[1], Chenglong Liu[1], Liangyu Yin[1], Min Zhang[2], Lei Jiang[2], Yuyan Gong[3], Wang Wu[3], Bi Liu[3], Xiaoyu Li[3], David S Cram[3], Dongge Liu[1]

[1]Department of Pathology, Beijing Hospital, National Center of Gerontology, Beijing 100730, China; [2]Department of Radiology, Beijing Hospital, National Center of Gerontology, Beijing 100730, China; [3]Berry Genomics Corporation, Beijing 102206, China

Correspondence to: Dongge Liu. Department of Pathology, Beijing Hospital, Beijing 100730, China. Email: 13661275182@163.com; David S Cram. Berry Genomics Corporation, Beijing 102206, China. Email: david.cram@berrygenomics.com.

Abstract

Objective: Histology grade, subtypes and TNM stage of lung adenocarcinomas are useful predictors of prognosis and survival. The aim of the study was to investigate the relationship between chromosomal instability, morphological subtypes and the grading system used in lung non-mucinous adenocarcinoma (LNMA).

Methods: We developed a whole genome copy number variation (WGCNV) scoring system and applied next generation sequencing to evaluate CNVs present in 91 LNMA tumor samples.

Results: Higher histological grades, aggressive subtypes and more advanced TNM staging were associated with an increased WGCNV score, particularly in CNV regions enriched for tumor suppressor genes and oncogenes. In addition, we demonstrate that 24-chromosome CNV profiling can be performed reliably from specific cell types (<100 cells) isolated by sample laser capture microdissection.

Conclusions: Our findings suggest that the WGCNV scoring system we developed may have potential value as an adjunct test for predicting the prognosis of patients diagnosed with LNMA.

Keywords: Lung adenocarcinoma; lung non-mucinous adenocarcinoma (LNMA); histological grading; TNM staging; copy number variations (CNVs); whole genome copy number variation (WGCNV) score

Submitted Mar 04, 2020. Accepted for publication May 13, 2020.
doi: 10.21147/j.issn.1000-9604.2020.03.05
View this article at: https://doi.org/10.21147/j.issn.1000-9604.2020.03.05

Introduction

Based on recommendations from the International Association for the Study of Lung Cancer (IASLC), the American Thoracic Society (ATS), the European Respiratory Society (ERS) and the 2015 World Health Organization (WHO), novel classification concepts were introduced for lung adenocarcinoma, including adenocarcinoma *in situ* (AIS) and minimally invasive adenocarcinoma (MIA). Moreover, invasive adenocarcinoma (IA) is now classified as lepidic, acinar, papillary, micropapillary or solid subtypes according to their predominant growth pattern (1-3). This classification system has subsequently been shown to be important for predicting prognosis and survival (4-7). In some studies and clinical practices, a three tiered architectural grading system has been applied, with grade 1 inclusive of AIS, MIA and lepidic predominant IA patterns, grade 2 inclusive of papillary and acinar predominant IA patterns and grade 3 inclusive of micropapillary and solid predominant IA patterns (8-11).

Chromosome instability is a common cytogenetic event in the evolution of most cancers, including lung adenocarcinomas (12-15). In recent molecular studies of lung adenocarcinoma, both oncogenic mutations and copy

number variations (CNVs) were observed in the early stages of tumor development with a tendency for CNVs to increase in frequency as tumorigenesis progresses (16-19). Based on these studies and similar observations of progressive genetic changes in other organ-specific carcinomas (20,21), we hypothesize that molecular subtyping may provide additional diagnostic and prognostic value in patients with lung adenocarcinoma.

To test this hypothesis, we selected lung non-mucinous adenocarcinoma (LNMA) as a model system to evaluate the clinical utility of CNV analysis. First, we established an accurate technique for whole genomic CNV (WGCNV) detection in laser capture micro-dissected adenocarcinoma samples, and second, developed a new scoring algorithm to assess the pathological significance of these WGCNV changes. Utilizing this approach for 91 LNMA patients, we further explored correlations between pathologic grading and subtyping of adenocarcinoma cells and WGCNV scores, and then assessed the diagnostic and prognostic value of WGCNV scoring.

Materials and methods

Patients

The retrospective study (*Figure 1*) was approved by Beijing Hospital Ethics Committee (approval number 2020BJYYEC-065-02). Patients provided written informed consent for collection and analysis of tissue samples. Pathologic diagnostic and subtypes determination for LNMA were based on 2011 IASLC/ATS/ERS/2015 WHO criteria (22-24). In addition, the 8th edition of the American Joint Committee for TNM stage was adopted to assess cancer staging. Cases were enriched for AIS, MIA and the different subtypes of IA in surgically resected samples collected from the Department of Pathology, Beijing Hospital between January 2017 and March 2018. Malignant pleural effusion (MPE) and metastatic nodules (MN) originating from LNMA were selected as stage IV samples. Patients were excluded from the study if they had either a history of neoadjuvant therapy, other lung cancer surgeries or discrepancies of diagnosis existing among

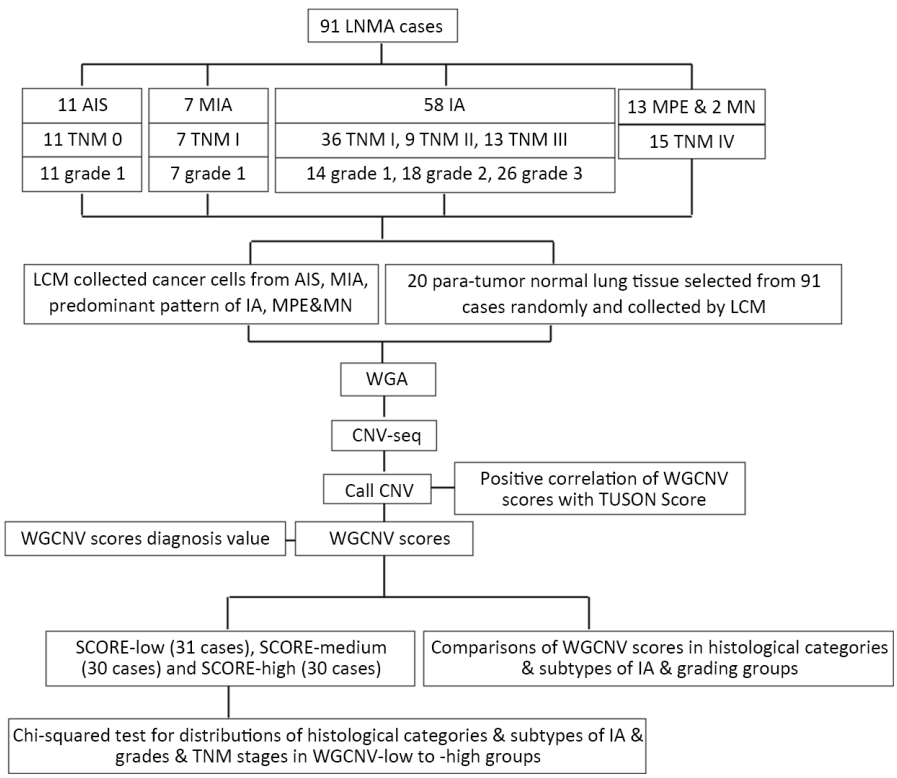


Figure 1 Study design. LNMA, lung non-mucinous adenocarcinoma; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; IA, invasive adenocarcinoma; MPE, malignant pleural effusion; MN, metastatic nodules; LCM, laser capture microdissection; WGCNV, whole genome copy number variation; WGA, whole genome amplification; seq, sequencing; TUSON, Tumor Suppressor and Oncogene.

pathologists/radiologists.

Histological and radiological evaluation

All pathological hematoxylin and eosin (H&E) stained slides were reviewed independently by two pulmonary pathologists (co-authors Z.W, and D.G.L). In addition, images of the enhanced chest CT scan were assessed by two chest radiologists (M.Z and L.J). The predominant histological subtype grading system was adopted in this study using a three tiered system and divided into grade 1 (AIS, MIA, lepidic predominant IA), grade 2 (acinar or papillary predominant IA) and grade 3 (solid or micropapillary predominant IA) (4,7). For all cases, pathologists and radiologists were blinded to the results of WGCNV scores.

Laser capture microdissection (LCM)

LCM was used to isolate pure predominant histological subtype tissue from each LNMA patient for WGCNV analysis. One H&E staining slide was selected based on the presence of the adenocarcinoma *in situ* components in AIS, the invasive components in MIA, a predominant histological subtype in IA and the adenocarcinoma component in MPE or MN samples. Formalin-fixed paraffin-embedded (FFPE) histological and cytological blocks corresponding to the selected H&E staining slides were then cut as 4–5 μm tissue sections (25,26). The individual sections were deposited on FrameSlides PEN-Membrane slides (Leica MicroDissect GmbH, Herborn, Germany) and H&E stained. LCM was performed with a Leica LMD 6500 (Leica Microsystems, Wetzlar, Germany). An average of 3–4 pieces containing homogeneous histological subcategory or subtype were collected by LCM into the caps of 200 μL Eppendorf tubes. As controls, 20 para-cancerous matching lung tissue samples (non-tumor lung tissue) were also randomly collected from the 91 enrolled LNMA cases.

CNV sequencing (CNV-seq) of germline genomic DNA

To identify background germline CNVs, CNV-seq was performed following previously published protocols (27,28). In brief, 50 ng genomic DNA extracted from peripheral blood lymphocytes was fragmented and purified using the QIAquick PCR Purification Kit (QIAGEN). After an end repair step, the sequencing adapter was ligated to the input DNA. To distinguish different libraries, index sequences were introduced by a PCR step. Next-generation sequencing (NGS) of purified libraries was performed using the Illumina HiSeq 2500 sequencer, and approximately 5 million single end 36 bp reads were obtained for CNV analysis.

Whole genome amplification (WGA) and CNV-seq of LCM cells

A centrifugation step was performed to ensure tissue samples collected by LCM were placed down at the bottom of the Eppendorf tubes. Samples (test and control) were subjected to WGA and sequencing library construction following previously published protocols (27,28). Next, the libraries were purified by AMPure XP beads (Beckman Coulter) and the concentration measured by Qubit 3.0 Fluorometer with Qubit dsDNA HS assay kits (Thermo Fisher Scientific, Waltham, USA). Finally, sequencing was performed using the HiSeq 2500 to obtain approximately 5 million 36 bp single end reads for CNV analysis.

CNV calling

The 20 batched test para-cancerous control tissues were internally compared with each other as reference using the data processing and analysis algorithms described previously (29). To analyze chromosomal abnormalities, we applied our previously reported pipeline for identifying, mapping and quantitating chromosomal aneuploidies, CNVs or mosaic variants. The copy number of different chromosomal regions was calculated using the fused lasso algorithm (30). Considering that cells of multiple genotypes may be present in one sample, a more relaxed threshold was used with copy number >2.3 marked as a duplication and a copy number <1.7 marked as a deletion.

Determination of resolution of CNV detection from WGA templates

To evaluate whether the WGA step affected copy number calls, we firstly collected 8 germline DNA samples harboring different known CNVs, which were previously detected by CNV-seq with the 50 ng genomic DNA input. We then performed WGA and CNV-seq from 100 pg of the corresponding genomic DNA and compared the CNVs detected by two different methods (*Supplementary Figure S1*). Validation experiments showed that both deletions and duplications larger than 2 Million bases (Mb) were accurately identified by CNV-seq from WGA templates,

whereas several other smaller CNVs not detected in the original genomic DNA samples were additionally detected and these were considered to be false positives. The sensitivity and specificity for detecting CNVs larger than 2 Mb was 100% (9/9) and 75% (9/12), respectively, whereas both sensitivity and specificity was 100% for detecting CNVs larger than 5 Mb. Therefore, for CNV calling, we defined the resolution of CNV detection at 5 Mb. Thus, in this study, only CNVs larger than 5 Mb in size were analyzed and scored.

WGCNV scoring and essential gene scoring algorithm

A scoring algorithm was developed to evaluate the cytogenetic abnormality of each sample (*Supplementary Figure S2*, selected example). The value was calculated by the formula below. The summed value was named "WGCNV score".

$$\text{WGCNV score} = \sum \frac{\text{CNV length}}{\text{Located chromosome length}} \times |\text{CN}_{\text{test}} - \text{CN}_{\text{ref}}|$$

The presence of CNVs involving tumor suppressor genes (TSGs) and oncogenes (OGs) was screened using Tumor Suppressor and Oncogene (TUSON) Explorer software. This method quantifies the likelihood that a gene within the CNV interval is a TSG or OG according to the pattern of mutational signature in tumors. Genes were predicted as TSGs or OGs based on the combined p value and q value of the selected parameters using the Liptak method (31) and these genes were scored using a transformation value equal to −log(q value) range from 0 to 100.

Statistical analysis

One-way analysis of variance (ANOVA)/analysis of covariance (AN(C)OVA) was used as the statistical method to compare the means of two or more groups of samples. To further determine significant differences between every two groups of samples, we used the Tukey's test for multiple comparisons. Chi-squared test was also used to determine whether there was a significant difference in distribution of low to high WGCNV scores between two groups. Pearson's correlation coefficient was used to measure the association between the WGCNV score and TUSON score and statistical significance was judged by the P value of the correlation test. A receiver operating characteristic (ROC) curve was used to elucidate any diagnostic value of the WGCNV score, a binary classifier created by plotting the true positive rate (sensitivity) against the false positive rate (1−specificity) at various threshold settings. The best cut-off was set by balancing sensitivity and specificity. The area under the ROC curve (AUC) was used as a measure of how well a threshold can diagnostically distinguish between two classes of samples. All P values were based on two-tailed statistical analysis and $P<0.05$ was considered as statistically significant. P values for all Tukey's test and Chi-squared test comparisons in this study are shown in *Supplementary Table S1*. Statistical analyses were conducted using R software (Version 3.2.5; R Foundation for Statistical Computing, Vienna, Austria).

Results

Clinicopathologic characteristics of LNMA patients

A total of 91 LNMA (tumor) and 20 para-cancerous (non-tumor) lung tissue specimens were collected by LCM from 91 patients (*Figure 1*). Of these specimens, 76 of 91 (84%) were surgical resection samples, including 11 (12%, 11/91) AIS, 7 (8%, 7/91) MIA and 58 (64%, 58/91) IA. In addition, 13 of 91 (14%) were MPE and 2 of 91 (2%) were MN biopsy samples. The predominant histological subtypes of 58 IA included 14 lepidic, 8 acinar, 10 papillary, 10 solid and 16 micropapillary subtypes. Of the 91 patients, 11 (12%) were in stage 0, 43 (47%) in stage I, 9 (10%) in stage II, 13 (14%) in stage III and 15 (17%) in stage IV. There were approximately equal numbers of females (47, 52%) and males (44, 48%) with an average age of 62 (range, 24–81) years. All patients were of Chinese ethnicity, with 30 (33%) classified as smokers (11 former smokers and 19 current smokers) and 61 (67%) were non-smokers. Patient clinicopathologic characteristics data are shown in *Supplementary Table S2*.

WGCNV scores versus histologic subtypes and predominant subtype grading

The presence of confounding germline CNVs was adjusted by subtraction of control germline CNVs (pool of 20 control samples) from the test CNVs, thus only the somatic CNVs were used for calculating the WGCNV scores. The WGCNV score algorithm was then applied to evaluate the extent and pattern of WGCNVs in tumor tissue from the 91 LNMA samples. The median WGCNV score of the 91 LNMA test samples was 2.70 (range, 0–9.95) whereas the

median WGCNV score of the para-cancerous control samples was 0.36 (range, 0–0.80) (*Supplementary Figure S3*). The median WGCNV scores of histological subcategories, subtypes, grades and TNM stages are shown in *Table 1*. Trends of positive correlation were observed in the median WGCNV scores among LMNA histological subtypes lepidic, acinar, papillary, solid and micropapillary and also among the three tiered grades. There were statistically significant differences between the WGCNV scores of lepidic and solid (P=0.026), lepidic and micropapillary (P<0.001), acinar and micropapillary (P=0.031) subtypes groups, as well as between grade 1 and grade 2 groups (P=0.004), grade 1 and grade 3 groups (P<0.001) and grade 2 and grade 3 groups (P=0.002) (*Figure 2A*,*B*). However, lepidic and acinar, lepidic and papillary, acinar and papillary and, acinar and solid patterns showed no significant differences in WGCNV scores.

A higher degree of invasiveness and a more advanced TNM stage are associated with a more complex spectrum of CNVs

Based on the trends observed with increased WGCNV scores and higher levels of pathological subtypes and grades, we further hypothesized that LNMA tumor samples with a higher degree of invasiveness may harbor larger CNV types or more complex CNV patterns. To test this hypothesis, we conducted more in depth comparisons of WGCNV scores from the different LNMA histological subcategories groups (from AIS, MIA, to IA and MPE&MN) based upon the degree of invasiveness (*Figure 2C*). While histological subcategories with a higher degree of invasiveness were associated with higher median WGCNV scores, the median WGCNV scores between the

Table 1 WGCNV scores of diversified groups of pathologic subcategories, subtypes, grades and TNM stages

Variables	% (n/N)			WGCNV scores [median (range)]	Significance*
	WGCNV score-low (0–1.74)	WGCNV score-medium (1.75–4.23)	WGCNV score-high (4.24–9.95)		
Pathological subcategories (n/N)					
AIS (11/91)	100 (11/11)	0 (0/11)	0 (0/11)	0.25 (0–1.63)	a
MIA (7/91)	71.4 (5/7)	28.6 (2/7)	0 (0/7)	1.12 (0–2.50)	a
IA (58/91)	20.7 (12/58)	37.9 (22/58)	41.4 (24/58)	3.50 (0–9.95)	b
MPE&MN (15/91)	20.0 (3/15)	40.0 (6/15)	40.0 (6/15)	3.51 (0–9.20)	b
Histological subtypes in IA (n/N)					
Lepidic type (14/58)	35.7 (5/14)	57.1 (8/14)	7.1 (1/14)	2.02 (0–5.31)	a
Acinar type (8/58)	25.0 (2/8)	50.0 (4/8)	25.0 (2/8)	2.28 (1.36–5.80)	a
Papillary type (10/58)	30.0 (3/10)	30.0 (3/10)	40.0 (4/10)	3.82 (0.31–8.86)	a
Solid type (10/58)	20.0 (2/10)	30.0 (3/10)	50.0 (5/10)	4.69 (1.30–9.95)	ab
Micropapillary type (16/58)	0 (0/16)	25.0 (4/16)	75.0 (12/16)	5.22 (2.99–9.46)	b
Subtype grades in resection (n/N)					
Grade 1 (32/76)	65.6 (21/32)	31.3 (10/32)	3.1 (1/32)	0.73 (0–5.31)	a
Grade 2 (18/76)	27.8 (5/18)	38.9 (7/18)	33.3 (6/18)	2.57 (0.31–8.86)	b
Grade 3 (26/76)	7.7 (2/26)	26.9 (7/26)	65.4 (17/26)	5.22 (1.30–9.95)	b
8th AJCC TNM stages (n/N)					
Stage 0 (11/91)	100.0 (11/11)	0.0 (0/11)	0.0 (0/11)	0.25 (0–1.63)	a
Stage I (43/91)	37.2 (16/43)	30.2 (13/43)	32.6 (14/43)	2.14 (0–8.95)	b
Stage II (9/91)	0 (0/9)	77.8 (7/9)	22.2 (2/9)	3.11 (2.03–9.46)	c
Stage III (13/91)	7.7 (1/13)	30.8 (4/13)	61.5 (8/13)	4.59 (0.31–9.95)	bc
Stage IV (15/91)	20.0 (3/15)	40.0 (6/15)	40.0 (6/15)	3.59 (0–9.20)	bc

WGCNV, whole genome copy number variation; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; IA, invasive adenocarcinoma; MPE&MN, malignant pleural effusion & metastatic nodule biopsy; AJCC, American Joint Committee on Cancer. *, Chi-squared test for distribution, P<0.05 is statistically significant, and same letters indicate no significant difference.

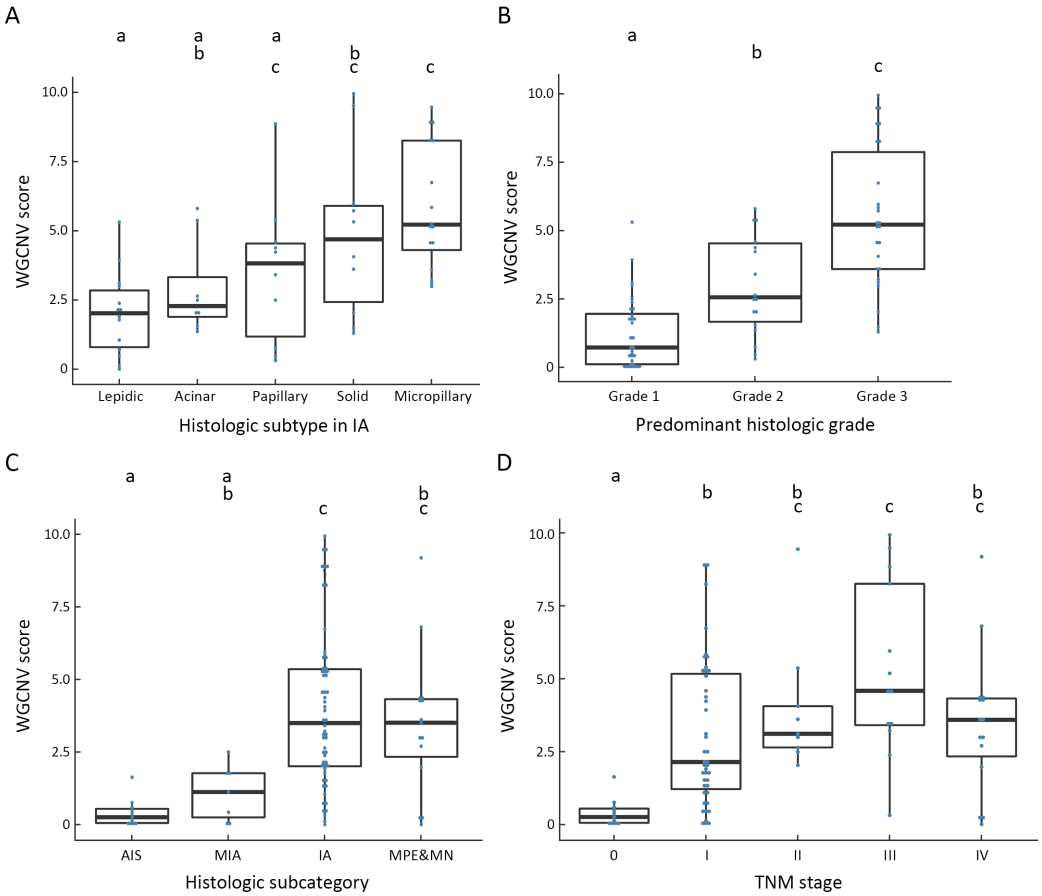


Figure 2 Positive trends between whole genome copy number variation (WGCNV) scores with different histologic subtypes, the predominant histologic grade system, histological subcategories and TNM stages in lung non-mucinous adenocarcinoma (LNMA). (A) Box-plot for correlation between WGCNV scores and different histologic subtypes in invasive adenocarcinoma (IA). A positive correlation trend is seen between the median WGCNV scores with five subtypes, from lepidic, acinar, papillary, solid to micropapillay groups; (B) Box-plot for correlation between WGCNV scores and the predominant histologic grade system. A positive correlation trend is seen between median WGCNVs with grade 1–3. Grade 1, AIS, MIA or lepidic predominant IA; Grade 2, acinar or papillary predominant IA; Grade 3, micropapillary or solid predominant IA; (C) Box-plot for correlation between WGCNV scores and LNMA histologic subcategories. A positive correlation trend is seen between median WGCNVs with adenocarcinoma *in situ* (AIS), minimally invasive adenocarcinoma (MIA), IA and malignant pleural effusion & metastatic nodule biopsy (MPE&MN) samples; similar median WGCNV scores were found between IA and MPE&MN group; (D) Box-plot for the correlation between WGCNV scores and TNM stages in LNMA. A positive correlation trend is seen between median WGCNVs with stage 0–III, but not in stage IV. Letters (a,b,c) above the box-plot diagrams are results of Tukey's test comparisons. Same letters indicate no significant difference between the two groups WGCNV scores.

MPE&MN and IA group were considered similar. The WGCNV scores were significantly higher among the IA group and the MPE&MN groups compared to the AIS group (P<0.001 and P=0.010, respectively) and in the IA group compared to the MIA group (P=0.019). However, there was no significant difference in WGCNV scores between AIS and MIA groups (P=0.930), MIA and MPE&MN groups (P=0.148), IA and MPE&MN groups (P=0.878).

Trends of positive correlation were also observed between the median WGCNV scores and stage groups, from 0 to III, but not in stage IV group (*Figure 2D*). The WGCNV scores were significantly higher for stages I–IV compared to stage 0 (P<0.05) and for stage I compared to stage III (P=0.025). However, there were no significant differences of WGCNV scores between stage I and II (P=0.831), I and IV (P=0.983), II and IV (P=0.984), III and IV (P=0.222) groups.

Higher pathologic subcategories, subtypes, grades and TNM stages of LNMA tend to distribute into higher WGCNV scores groups

The WGCNV scores of 91 tumor samples were evenly distributed between 0 and 9.95 (refer to scatter diagram of the 91 tumor samples distribution in *Supplementary Figure S3*). For a more global view, we first sorted the 91 test cases based on their WGCNV scores and then equally divided them into three groups as score-low (31 cases, 0–1.74), score-medium (30 cases, 1.75–4.23) and score-high (30 cases, 4.24–9.95).

Firstly, when investigating the distributions of WGCNVs according to four pathological subcategories (*Table 1*), it was found that all 11 AIS samples (100%) fell in the score-low group. Similarly, 5 of 7 MIA samples (71.4%) fell in the score-low group and 2 of 7 (28.6%) fell in the score-medium groups. In contrast, the majority of IA and MPE&MN samples (58/73, 79%) fell into either the score-medium (28/58, 48%) or score-high (30/58, 52%) groups. Chi-squared test confirmed statistically significant differences for the distributions in IA *vs.* AIS (P<0.001), in MPE&MN *vs.* AIS (P<0.001), in MIA *vs.* IA (P=0.010), and in MIA *vs.* MPE&MN (P=0.040).

Secondly, we further investigated the distributions of histological subtypes in IA and three tiered grades of LNMA in three WGCNV score groups. Significant differences of the distributions into WGCNV score low to high groups were observed not only among lepidic, acinar, papillary and solid cells with micropapillary subtypes, but also among grade 3 samples compared to grade 1 (P=0.004) and grade 2 (P<0.001) samples. All these observations indicated that LNMA samples with a higher grading in the pathologic subcategories, histological subtypes or grading system, had tendencies to distribute into higher WGCNV score groups (*Table 1*). However, we found some exceptions where there were differences in distributions within one histological subtype or grade group, for example, 10 cases with predominant papillary type IA, and 3, 5, and 2 cases were distributed into the low, medium and high WGCNV score groups, respectively.

Thirdly, we also investigated the distribution of TNM stages in the WGCNV score from low to high groups (*Table 1*). All stage 0 (11/11, 100%) samples fell into the score-low CNV group whereas the 43 stage I samples almost equally fell into the three SCORE CNV groups. In contrast, the majority (7/9, 77.8%) of stage II samples fell into the score-medium CNV group and a minority (2/9, 22.2%) fell into the score-high. There were significant differences among the distributions from stage 0 *vs*. I–IV (P=0.001, P<0.001, P<0.001, P<0.001, respectively), stage I and stage II (P=0.018) but not in stage I and III (P=0.082) or stage I and IV (P=0.471) and stage III and IV (P=0.461) group.

Association of TUSON scores with histological LNMA subtypes and grading

To assess the overall effect of CNVs on tumorigenesis, we further evaluated the full spectrum of tumor CNVs in each of the 91 samples according to specimen histologic grade/subtype (*Figure 3A*). For CNV deletions, the TSG scores in IA and MPE&MN samples were generally higher than OG scores. However, for CNV duplications there was an opposite trend. The median TUSON score was 0 (−143.2–1.6) for AIS samples, 8.3 (−57.7–61.6) for MIA samples, 134.53 (−70.3–534.6) for IA samples and 116 (0–397.3) for MPE&MN samples. In regard to the predominant histological grade system, the TUSON scores of grade 2 and grade 3 were significantly higher than those of grade 1 (P=0.004 and P<0.001, respectively) (*Figure 3B*). TUSON scores were significant higher in IA samples than in AIS samples (P=0.005) and in MIA samples (P=0.033). There was no significant difference in TUSON scores between MPE&MN and IA samples (P=0.874). There was a general trend for selective enrichment for TSG in deletion CNVs and for OG in duplication CNVs and these may all contribute to tumorigenesis (*Figure 3C*). By Pearson correlation (R=0.53, P<0.001), there was a moderate association of TUSON and WGCNV scores, indicating that a majority of the total CNVs detected affected expression of tumorigenesis genes or oncogenes (*Figure 3D*).

Assessment of diagnostic value of WGCNV scores

WGCNV scores of all the 111 samples in this study (91 cancerous and 20 para-cancerous lung tissue specimens) were calculated to determine the differential diagnostic value for LNMA from non-tumor lung tissue (LNMA differentiated diagnosis). By ROC curve analysis of WGCNV scores we were able to define a cut-off value of 0.02 to separate LNMA and para-cancerous non-tumor lung samples. The diagnostic sensitivity and specificity was 95.0% and 97.8%, and the positive and negative predictive value of this classifier was 99% and 90%, respectively (*Figure 4A*). The AUC was 0.981, indicating WGCNV

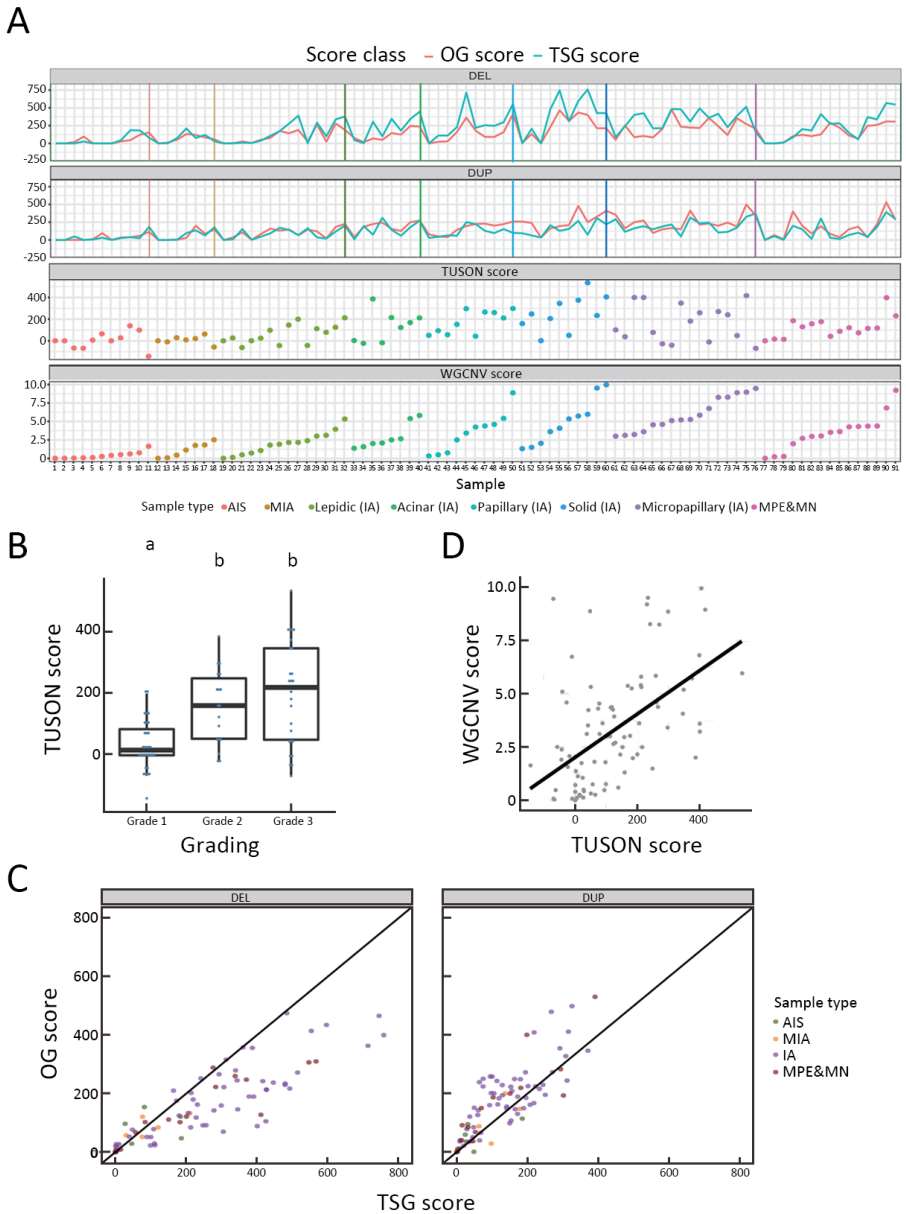


Figure 3 Distribution and score of OGs or TSGs on tumor CNV. (A) In the first and second row of the grid diagram, TSGs and OGs scores in loss and gain CNV regions for all 91 tumor samples were separately calculated. TSGs score was generally higher than OGs in deletion areas in IA and MPE&MN tissues; a contrary trend was shown in duplication areas. This trend was not seen in AIS and MIA. The third and fourth row of the grid diagram show TUSON score and WGCNV score of all 91 tumor samples, dots of different colors represent different types of samples. There is a rising trend in TUSON score and WGCNV score based upon the degree of invasiveness from AIS, MIA to IA and MPE&MN; (B) Box-plot of TUSON score of samples from different gradings. TUSON scores in grade 1 were significantly lower than that in the grade 2 and grade 3. Letters (a,b,c) above the box-plot diagrams are results of Tukey's test comparisons. Same letters indicate no significant difference between the two groups WGCNV scores.; (C) Scatter plot of TSGs and OGs scores in loss and gain CNV regions. There is a trend for selective enrichment for TSG in deletion CNVs and for OG in duplication CNVs, DEL, deletion; DUP, duplication; (D) Moderate correlation was shown in Pearson correlation analysis for WGCNV score and TUSON score (r=0.53, P<0.001, Pearson correlation). OG, oncogene; TSG, tumor suppressor gene; CNV, copy number variation; IA, invasive adenocarcinoma; MPE&MN, malignant pleural effusion & metastatic nodule biopsy; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; TUSON, Tumor Suppressor and Oncogene; WGCNV, whole genome copy number variation.

scores are a very good separator of cancerous and non-cancerous tissue. To further evaluate whether the WGCNV scores could distinguish AIS and MIA samples from IA and MPE&MN samples, a cut-off value of 1.85 was set to be the threshold to classify the two groups. The sensitivity and specificity was 78.1% and 94.4%, respectively, and the positive and negative predictive value of this classifier was 98% and 52%, respectively. The AUC was 0.896 (*Figure 4B*).

Discussion

In this present study, we developed a new scoring algorithm to assess WGCNV changes and found positive correlation trends of median WGCVN scores with the diversified LNMA histological subcategories (from AIS, MIA to IA and MPE&MN), five histologic subtypes (from lepidic, acinar, papillary, solid to micropapillary subtype) of IA, and three tiered predominant histologic subtype grades. In addition, both a higher degree of invasiveness and histological grade in LNMA were associated with a more complex spectrum of CNVs. Based on statistical analyses of tumor samples from LNMA patients, there was a general trend for low histological pathologic subcategories, subtypes, grading and TNM stages with lower WGCNV scores and a trend for higher histological pathologic subcategories, subtypes, grading and TNM stages with higher WGCNV scores. For both IA and MPE&MN samples, TSG and OG scores based on a subset of deletions and duplications, respectively, displayed a general trend for selective enrichment.

More recently, some research studies have concluded that chromosomal instability is a driver not only for tumor evolution but also for metastasis (19,32). Higher CNV levels were associated with cancer recurrence and death. In this study, we revealed significant correlations between WGCNV scoring with important pathological diagnostic and clinical prognostic criteria, suggesting that the WGCNV scoring algorithm developed here may have potential clinical value as a more objective biomarker for predicting prognosis as well as diagnostic value to differentiate diversified histological subcategories in LNMA patients. For the first time, to our knowledge, the findings reported here build on the correlations between evaluation for the number of copy number aberrations and chromosome abnormality score with pathological morphological diagnostic and clinic-pathological prognostic parameters in LNMA.

Although a positive correlation was found between WGCNV score levels and the degree of invasiveness in LNMA (from AIS, MIA to IA and MPE&MN), there were no significant differences between AIS and MIA groups for median WGCNV score levels and the distribution in low to high score groups. Recently, the clinical and prognostic meaning of the IASLC/ATS/ERS and WHO classification (as AIS, MIA and IA) was challenged (33-35), because the

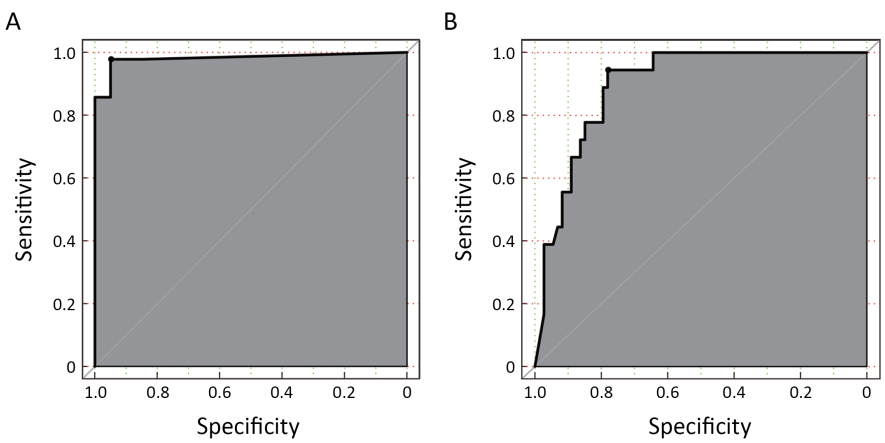


Figure 4 Diagnostic value of WGCNV score. ROC curve shows diagnostic performance of WGCNV score in distinguishing LNMA from para-cancerous non-tumor lung samples [cut-off value, 0.02; sensitivity, 95.0%; specificity, 97.8%; AUC=0.981]. (A) and in distinguishing AIS, MIA from IA and MPE&MN samples [cut-off value, 1.85; sensitivity, 78.1%; specificity, 94.4%; AUC=0.896] (B). WGCNV, whole genome copy number variation; ROC, receiver operating characteristic; LNMA, lung non-mucinous adenocarcinoma; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; IA, invasive adenocarcinoma; MPE&MN, malignant pleural effusion & metastatic nodule biopsy; AUC, the area under the ROC curve.

results of individual center studies and large meta-analyses showed no significant differences in the survival rates between AIS and MIA patients with lung adenocarcinoma. Based on our finding that chromosome abnormalities mentioned above were not significantly different between AIS and MIA groups, this challenge to the classification may have a genuine basis.

Although there is still no internationally accepted grading system for lung adenocarcinoma, the new histological classification has at least given the impetus to develop the grading systems based more on both architectural and nuclear morphology (4,8). From this study, a positive correlation trend and significant differences of median WGCNV score levels were identified amongst the three-tiered predominant subtype grades. Additionally, the predominant histological grading system adopted in this study was supported by our findings, and thus the WGCNVs scoring algorithm we developed may prove to be a useful ancillary system for molecular grading of LNMA.

A diversified distribution in WGCNV-low to -high score groups from one histological subtype was found in our data. For example, of the 10 cases with predominant papillary type IA, 3, 5 and 2 cases distributed into the low, medium and high WGCNV score groups, respectively. In one study, the papillary subtype IA was separated into 3 different types on the basis of the degree of papillary architectural and cytological pleomorphism and patient follow-up showed that this subclassification of papillary was associated with a prognostic impact with type 3 having a worse prognosis (36). According to the 2011 IASLC/ATS/ERS/2015 WHO classification, the cribriform and fused glands are regarded as acinar subtype in LNMA (9), but some reports suggest that the cribriform and fused glands should be distinguished from the acinar type as a subtype with worse prognosis (36-38). We reviewed the H&E staining slides from the 10 predominant papillary IA cases, and found 2 high WGCNV score cases had the morphological characteristics of type 3 papillary tumor previously proposed (36). This finding suggested that the tumor morphologic heterogeneity may be linked to different molecular abnormalities. Thus establishing a WGCNV scoring system may be more objective and practical than a simple architectural grading system and thus more useful as a prognostic evaluator in LNMA patients. Nonetheless, WGCNV scores need to be further assessed in association with survival analysis in a larger number of patients to determine prognostic value.

Based on the previous suggestions of the relationship between chromosomal instability and tumor metastasis (32,39-41), it has been proposed that the invasive ability of tumor cells may be enhanced by a rising degree of increasing CNV changes, and specific cell subclones with the most severe extensive CNVs changes may obtain the capacity to invade out of organs and form local or distant metastasis. In general, our findings underlined the positive correlation trend with median WGCNV scores among AIS, MIA and IA within the lung and supported our hypothesis. The exception was for the MPE&MN samples, which showed almost the same score with IA, even in the distributions analysis of the low to high score groups. Also, similar to the results of histologic subcategories, in the analysis of the relationship between WGCNV scores and TNM stages, the median WGCNV scores were generally increasing with higher TNM stages. Thus, within the organ (lung), the hypothesis about the relationship between invasion and WGCNV scoring was verified. Further research is therefore needed to confirm our findings and determine whether different mechanisms of tumor aggression exist within and outside of an affected organ.

Frequent CNV accumulation may be the most striking characteristic of cancer genomes, though ostensibly random, they may follow a pattern. Based on our results, we identified a non-random pattern of somatic CNVs which gradually increased with the degree of invasion (*Figure 3A*). Scores of TSG and OG were basically the same either in loss or gain CNVs for AIS, MIA and predominant lepidic IA, whereas a trend of selective enrichment for TSG in deletion CNVs and for OG in duplication CNVs was more pronounced in other IA and MPE&MN samples, and such changes are likely to promote tumorigenesis. Whether the non-random accumulation of CNV mainly occurs in clonal or subclonal lines or not needs to be verified and the relationship of CNV and oncogenic mutation enrichment in different clonal lines remains obscure.

Our molecular studies suggested that the WGCNV score may have diagnostic value for LNMA patients. ROC curve analysis indicated that WGCNV scores may be valuable in differentiated diagnosis of LNMA from non-tumor lung tissue (AUC=0.981), and of AIS and MIA from IA and MPE&MN (AUC=0.896). WGA of LCM samples combined with CNV-seq was shown to be a relatively simple technique for detecting CNV changes in less than 100 cells, and may add diagnostic value to guide treatment options. In addition, in cases where limited biopsy or low

tumor content cytological samples are all that is available, it will be possible to obtain additional information to give a more precise pathologic diagnosis.

The main limitations of introducing a WGCNV scoring system into a routine pathology laboratory are the time and resources needed to identify and assess the CNVs and derive a score. For this purpose, a next generation sequencing platform would be required to perform CNV sequencing which normally takes 2–3 days for sample preparation, library construction, sequencing and data analysis. Further, staff training would be required so that the molecular and histology results could be interpreted and integrated into a final report, so that the information can be used to guide treatment. In addition, the whole procedure would add additional cost to the patient for molecular assessment of the tumor.

There were also some limitations of this study that may have influenced the association studies with WGCNV scores. First, in some cases there were only small sample numbers for assessing the relationship between WGCNV scores with histological subtypes and TNM stages, and this may have introduced statistical bias in some comparisons. Thus, further studies on a larger cohort of LNMA patients are still needed to establish more confidently some of the associations reported in this study. Second, for obtaining pure histological subtype material, the LCM technique was used in our study to examine the predominant tumor type. However, in cases whether there is significant tumor heterogeneity, it would be interesting to compare WGCNV scores between different tissue regions. Third, for the purpose of avoiding poor DNA quality and reducing test failure, FFPE histologic and cytologic cell blocks less than 2 years old were selected and so no meaningful patient survival data linking with WGCNV scores could be obtained. Fourth, our research only focused on LNMA samples, and so additional studies are necessary to determine if similar associations are found in lung mucinous adenocarcinoma and other types of lung cancer or in other organ-specific tumors. Fifth, it will be important to perform CNV and oncogenic mutation analyses in parallel to identify a more holistic set of genetic markers. Lastly, genome-wide low-depth analysis cannot determine other genetic changes such as genomic doubling reported in lung adenocarcinoma.

Conclusions

In this cohort study of LNMA patients, we showed a positive correlation of WGCNV score with morphological diagnostic and clinicopathological prognostic parameters. Accordingly, we propose that WGCNV scoring may have intrinsic value as a robust molecular parameter in pathological diagnosis and could be used to better evaluate the prognosis of LNMA patients and, potentially other patients with lung adenocarcinoma.

Acknowledgements

This work was supported by grants from Beijing Hospital Key Research Program (121 Research Program, No. BJ-2019-195).

Footnote

Conflicts of Interest: The authors have no conflicts of interest to declare.

References

1. Travis WD, Brambilla E, Noguchi M, et al. International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. J Thorac Oncol 2011;6:244-85.
2. Travis WD, Brambilla E, Noguchi M, et al. Diagnosis of lung adenocarcinoma in resected specimens: Implications of the 2011 International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification. Arch Pathol Lab Med 2013;137:685-705.
3. Travis WD, Brambilla E, Nicholson AG, et al. The 2015 World Health Organization classification of lung tumors: Impact of genetic, clinical and radiologic advances since the 2004 classification. J Thorac Oncol 2015;10:1243-60.
4. Kadota K, Suzuki K, Kachala SS, et al. A grading system combining architectural features and mitotic count predicts recurrence in stage I lung adenocarcinoma. Mod Pathol 2012;25:1117-27.
5. Tsuta K, Kawago M, Inoue E, et al. The utility of the proposed IASLC/ATS/ERS lung adenocarcinoma subtypes for disease prognosis and correlation of driver gene alterations. Lung Cancer 2013;81:371-6.
6. Mäkinen JM, Laitakari K, Johnson S, et al.

Histological features of malignancy correlate with growth patterns and patient outcome in lung adenocarcinoma. Histopathology 2017;71:425-36.
7. Ujiie H, Kadota K, Chaft JE, et al. Solid predominant histologic subtype in resected stage I lung adenocarcinoma is an independent predictor of early, extrathoracic, multisite recurrence and of poor post-recurrence survival. J Clin Oncol 2015;33:2877-84.
8. Sica G, Yoshizawa A, Sima CS, et al. A grading system of lung adenocarcinomas based on histologic pattern is predictive of disease recurrence in stage I tumors. Am J Surg Pathol 2010;34:1155-62.
9. Travis WD. Pathology of lung cancer. Clin Chest Med 2011;32:669-92.
10. Zombori T, Furák J, Nyári T, et al. Evaluation of grading systems in stage I lung adenocarcinomas: A retrospective cohort study. J Clin Pathol 2018;71: 135-40.
11. Warth A, Muley T, Meister M, et al. The novel histologic International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society classification system of lung adenocarcinoma is a stage-independent predictor of survival. J Clin Oncol 2012;30:1438-46.
12. Xia S, Huang CC, Le M, et al. Genomic variations in plasma cell free DNA differentiate early stage lung cancers from normal controls. Lung Cancer 2015;90:78-84.
13. Bowcock AM. Invited review DNA copy number changes as diagnostic tools for lung cancer. Thorax 2014;69:495-6.
14. Qiu ZW, Bi JH, Gazdar AF, et al. Genome-wide copy number variation pattern analysis and a classification signature for non-small cell lung cancer. Genes Chromosomes Cancer 2017;56:559-69.
15. Micke P, Edlund K, Holmberg L, et al. Gene copy number aberrations are associated with survival in histologic subgroups of non-small cell lung cancer. J Thorac Oncol 2011;6:1833-40.
16. Campbell JD, Alexandrov A, Kim J, et al. Distinct patterns of somatic genome alterations in lung adenocarcinomas and squamous cell carcinomas. Nat Genet 2016;48:607-16.
17. de Bruin EC, McGranahan N, Mitter R, et al. Spatial and temporal diversity in genomic instability processes defines lung cancer evolution. Science 2014;346:251-6.
18. Zhang J, Fujimoto J, Zhang J, et al. Intratumor heterogeneity in localized lung adenocarcinomas delineated by multiregion sequencing. Science 2014;346:256-9.
19. Jamal-Hanjani M, Wilson GA, McGranahan N, et al. Tracking the evolution of non-small-cell lung cancer. N Engl J Med 2017;376:2109-21.
20. Guinney J, Dienstmann R, Wang X, et al. The consensus molecular subtypes of colorectal cancer. Nat Med 2015;21:1350-6.
21. Bell DW, Ellenson LH. Molecular genetics of endometrial carcinoma. Annu Rev Pathol 2019; 14:339-67.
22. Borczuk AC. Assessment of invasion in lung adenocarcinoma classification, including adenocarcinoma *in situ* and minimally invasive adenocarcinoma. Mod Pathol 2012;25(suppl 1):S1-10.
23. Weichert W, Warth A. Early lung cancer with lepidic pattern: adenocarcinoma *in situ*, minimally invasive adenocarcinoma, and lepidic predominant adenocarcinoma. Curr Opin Pulm Med 2014;20:309-16.
24. Warth A, Cortis J, Fink L, et al. Training increases concordance in classifying pulmonary adenocarcinomas according to the novel IASLC/ATS/ERS classification. Virchows Arch 2012;461:185-93.
25. Longuespée R, Alberts D, Pottier C, et al. A laser microdissection-based workflow for FFPE tissue microproteomics: Important considerations for small sample processing. Methods 2016;104:154-62.
26. Golubeva Y, Salcedo R, Mueller C, et al. Laser capture microdissection for protein and NanoString RNA analysis. Methods Mol Biol 2013;931:213-57.
27. Liang D, Peng Y, Lv W, et al. Copy number variation sequencing for comprehensive diagnosis of chromosome disease syndromes. J Mol Diagn 2014;16:519-26.
28. Wang L, Cram DS, Shen J, et al. Validation of copy number variation sequencing for detecting chromosome imbalances in human preimplantation embryos. Biol Reprod 2014;91:37.
29. Wang Y, Chen Y, Tian F, et al. Maternal mosaicism is a significant contributor to discordant sex chromosomal aneuploidies associated with non-invasive prenatal testing. Clin Chem 2014;60: 251-9.

30. Tibshirani R, Wang P. Spatial smoothing and hot spot detection for CGH data using the fused lasso. Biostatistics 2008;9:18-29.
31. Davoli T, Xu AW, Mengwasser KE, et al. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. Cell 2013;155:948-62.
32. Bakhoum SF, Ngo B, Laughney AM, et al. Chromosomal instability drives metastasis through a cytosolic DNA response. Nature 2018;553:467-72.
33. Kadota K, Villena-Vargas J, Yoshizawa A, et al. Prognostic significance of adenocarcinoma *in situ*, minimally invasive adenocarcinoma, and non-mucinous lepidic predominant invasive adenocarcinoma of the lung in patients with stage I disease. Am J Surg Pathol 2014;38:448-60.
34. Ito M, Miyata Y, Kushitani K, et al. Prediction for prognosis of resected pT1a-1bN0M0 adenocarcinoma based on tumor size and histological status: relationship of TNM and IASLC/ATS/ERS classifications. Lung Cancer 2014;85:270-5.
35. Behera M, Owonikoko TK, Gal AA, et al. Lung adenocarcinoma staging using the 2011 IASLC/ATS/ERS classification: A pooled analysis of adenocarcinoma *in situ* and minimally invasive adenocarcinoma. Clin Lung Cancer 2016;17:e57-e64.
36. Warth A, Muley T, Harms A, et al. Clinical relevance of different papillary growth patterns of pulmonary adenocarcinoma. Am J Surg Pathol 2016;40:818-26.
37. Kadota K, Yeh YC, Sima CS, et al. The cribriform pattern identifies a subset of acinar predominant tumors with poor prognosis in patients with stage I lung adenocarcinoma: a conceptual proposal to classify cribriform predominant tumors as a distinct histologic subtype. Mod Pathol 2014;27:690-700.
38. Moreira AL, Joubert P, Downey RJ, et al. Cribriform and fused glands are patterns of high-grade pulmonary adenocarcinoma. Hum Pathol 2014;45:213-20.
39. Mackinnon AC Jr., Luevano A, de Araujo LC, et al Cribriform adenocarcinoma of the lung: clinicopathologic, immunohistochemical, and molecular analysis of 15 cases of a distinctive morphologic subtype of lung adenocarcinoma. Mod Pathol 2014;27:1063-72.
40. Misumi T, Yamamoto Y, Miyachika Y, et al. DNA copy number aberrations associated with lymphovascular invasion in upper urinary tract urothelial carcinoma. Cancer Genet 2012;205:313-8.
41. Torresan C, Oliveira MM, Pereira SR, et al. Increased copy number of the DLX4 homeobox gene in breast axillary lymph node metastasis. Cancer Genet 2014;207:177-87.

Cite this article as: Wang Z, Zhang L, He L, Cui D, Liu C, Yin L, Zhang M, Jiang L, Gong Y, Wu W, Liu B, Li X, Cram DS, Liu D. Low-depth whole genome sequencing reveals copy number variations associated with higher pathologic grading and more aggressive subtypes of lung non-mucinous adenocarcinoma. Chin J Cancer Res 2020;32(3):334-346. doi: 10.21147/j.issn.1000-9604.2020.03.05

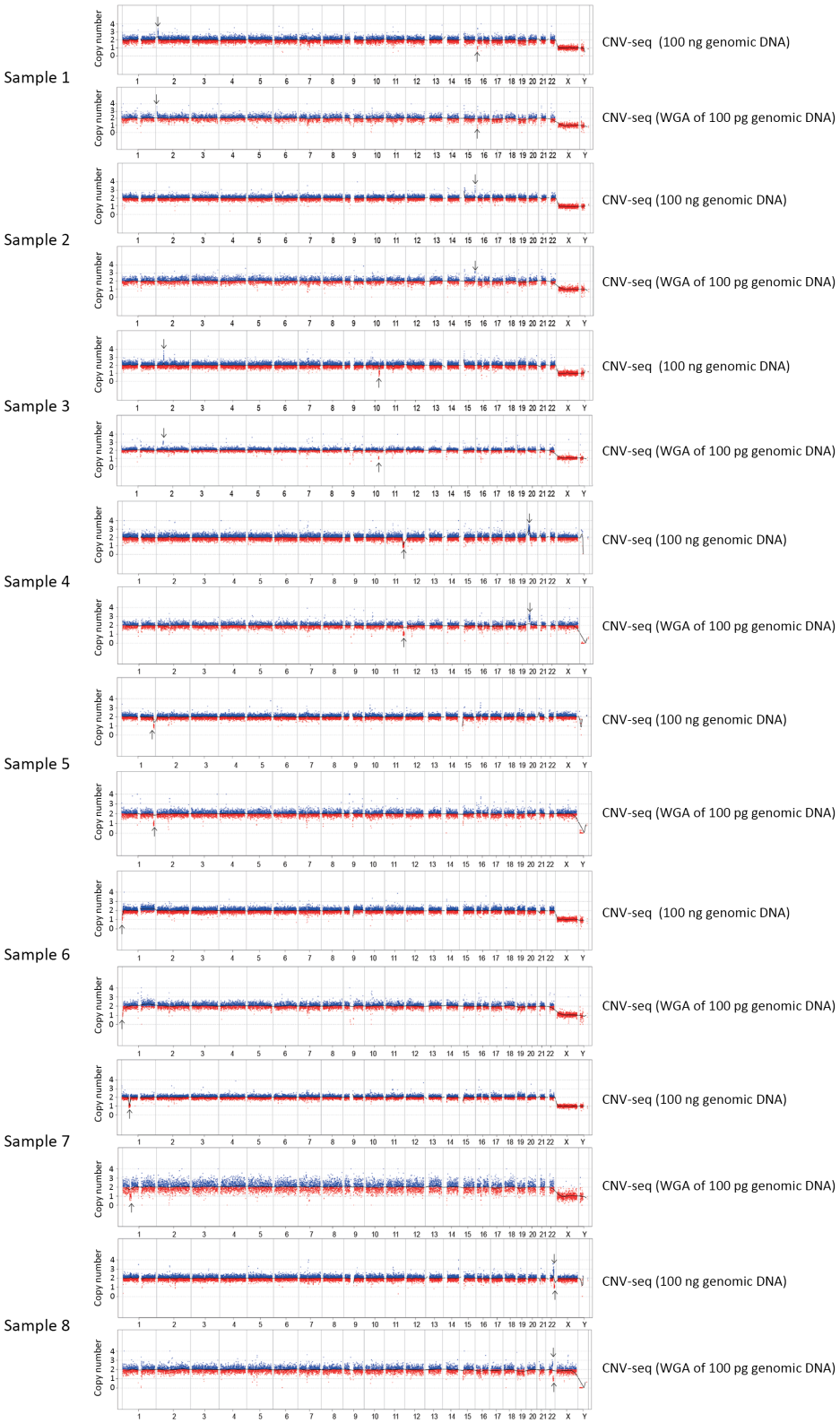
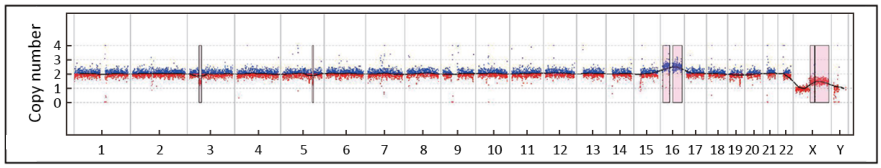


Figure S1 Sensitivity of CNV-seq for detecting CNVs. CNV, copy number variation; seq, sequencing; WGA, whole genome amplification.



Chromosome	CNV start	CNV end	CNV length	Chromosome length	CNtest	CNref	|CNtest−CNref|
chr3	43100000	55300000	12200000	194820000	1.43	2	0.57
chr5	135660000	141180000	5520000	177280000	1.43	2	0.57
chr16	1420000	32900000	31480000	78700000	2.42	2	0.42
chr16	46380000	90220000	43840000		2.48	2	0.48
chrX	66500000	88020000	21520000	148100000	1.43	1	0.43
chrX	94940000	155020000	63080000		1.44	1	0.44

$$\text{WGCNV score}=\sum\frac{\text{CNV length}}{\text{Located chromosome length}}\times|\text{CNtest}-\text{CNref}|$$

WGCNV score=0.739

Figure S2 Calculation of WGCNV scores from CNV-seq profiles. WGCNV, whole genome copy number variation; CNV, copy number variation; seq, sequencing.

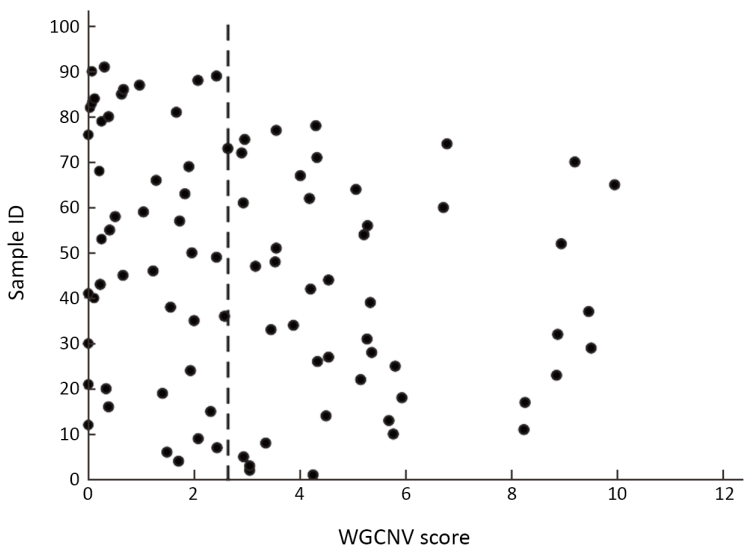


Figure S3 Distribution of WGCNV scores for LNMA tumor samples (N=91). Median is indicated by the dotted line. WGCNV, whole genome copy number variation; LNMA, lung non-mucinous adenocarcinoma.

Table S1 P value for all Tukey's test and Chi-squared test comparisons

Tukey's test	P	Tukey's test	P	Chi-squared test	P	Chi-squared test	P
Histological subtypes in IA (WGCNV score)		8th AJCC TNM stages (WGCNV score)		Histological Subtypes in IA (WGCNV score)		8th AJCC TNM stages	
Acinar–Lepidic	0.895	I–0	0.012	Acinar–Lepidic	0.491	I–0	0.001
Papillary–Lepidic	0.518	II–0	0.011	Papillary–Lepidic	0.134	II–0	<0.001
Solid–Lepidic	0.026	III–0	<0.001	Solid–Lepidic	0.057	III–0	<0.001
Micropapillary–Lepidic	<0.001	IV–0	0.016	Micropapillary–Lepidic	<0.001	IV–0	<0.001
Papillary–Acinar	0.983	II–I	0.831	Papillary–Acinar	0.671	II–I	0.018
Solid–Acinar	0.362	III–I	0.025	Solid–Acinar	0.543	III–I	0.082
Micropapillary–Acinar	0.031	IV–I	0.983	Micropapillary–Acinar	0.026	IV–I	0.471
Solid–Papillary	0.640	III–II	0.669	Solid–Papillary	0.856	III–II	0.088
Micropapillary–Papillary	0.084	IV–II	0.984	Micropapillary–Papillary	0.048	IV–II	0.148
Micropapillary–Solid	0.821	IV–III	0.222	Micropapillary–Solid	0.146	IV–III	0.461
Subtype grades in resection (WGCNV score)		Pathological subcategories (TUSON score)		Subtype grades in resection (WGCNV Score)			
Grade 2–1	0.004	MIA–AIS	1.000	Grade 2–1	0.004		
Grade 3–1	<0.001	IA–AIS	0.005	Grade 3–1	<0.001		
Grade 3–2	0.002	MPE&MN–AIS	0.106	Grade 3–2	0.072		
Pathological subcategories (WGCNV score)		IA–MIA	0.033	Pathological subcategories			
MIA–AIS	0.930	MPE&MN–MIA	0.213	MIA–AIS	0.267		
IA–AIS	<0.001	MPE&MN–IA	0.874	IA–AIS	<0.001		
MPE&MN–AIS	0.010			MPE&MN–AIS	<0.001		
IA–MIA	0.019			IA–MIA	0.010		
MPE&MN–MIA	0.148			MPE&MN–MIA	0.040		
MPE&MN–IA	0.878			MPE&MN–IA	0.989		

IA, invasive adenocarcinoma; WGCNV, whole genome copy number variation; AIS, adenocarcinoma *in situ*; MIA, minimally invasive adenocarcinoma; MPE&MN, malignant pleural effusion & metastatic nodule biopsy; AJCC, American Joint Committee on Cancer; TUSON, Tumor Suppressor and Oncogene.

Table S2 Clinicopathologic characteristics of LNMA patients

No.	Sex	Age (year)	TNM stage			TNM stage	Neoadjuvant therapy	Smoking status
			Tumor	Nodes	Metastasis			
1	M	71	TIS	N0	M0	0	No	Former
2	F	50	TIS	N0	M0	0	No	Never
3	M	64	TIS	N0	M0	0	No	Never
4	F	46	TIS	N0	M0	0	No	Never
5	M	24	TIS	N0	M0	0	No	Current
6	F	66	TIS	N0	M0	0	No	Never
7	F	53	TIS	N0	M0	0	No	Never
8	M	54	TIS	N0	M0	0	No	Never
9	M	54	TIS	N0	M0	0	No	Current
10	F	56	TIS	N0	M0	0	No	Never
11	F	58	TIS	N0	M0	0	No	Never
12	M	65	T1a	N0	M0	IA1	No	Never
13	F	53	T1a	N0	M0	IA1	No	Never
14	F	53	T1b	N0	M0	IA2	No	Never
15	F	62	T1b	N0	M0	IA2	No	Never
16	M	61	T1a	N0	M0	IA1	No	Current
17	F	56	T1a	N0	M0	IA1	No	Never
18	F	46	T1a	N0	M0	IA1	No	Never
19	F	73	T1a	N0	M0	IA1	No	Never
20	M	56	T1a	N0	M0	IA1	No	Former
21	F	50	T1a	N0	M0	IA1	No	Never
22	M	62	T1a	N0	M0	IA1	No	Never
23	M	60	T1a	N0	M0	IA1	No	Current
24	F	56	T1a	N0	M0	IA1	No	Current
25	F	53	T1a	N0	M0	IA1	No	Never
26	F	77	T1a	N0	M0	IA1	No	Never
27	M	65	T1a	N0	M0	IA1	No	Never
28	F	53	T1a	N0	M0	IA1	No	Never
29	F	60	T1a	N0	M0	IA1	No	Never
30	F	60	T1a	N0	M0	IA1	No	Never
31	F	55	T1b	N0	M0	IA2	No	Never
32	M	56	T1b	N0	M0	IA2	No	Former
33	F	56	T1b	N0	M0	IA2	No	Never
34	F	60	T1b	N0	M0	IA2	No	Never
35	M	75	T1b	N0	M0	IA2	No	Former
36	M	75	T1b	N0	M0	IA2	No	Former
37	M	68	T1b	N0	M0	IA2	No	Never
38	F	56	T1b	N0	M0	IA2	No	Never
39	F	49	T1b	N0	M0	IA2	No	Never
40	M	51	T1c	N0	M0	IA3	No	Former
41	M	71	T1c	N0	M0	IA3	No	Former

Table S2 (*continued*)

Table S2 (*continued*)

No.	Sex	Age (year)	TNM stage			TNM stage	Neoadjuvant therapy	Smoking status
			Tumor	Nodes	Metastasis			
42	M	71	T1c	N0	M0	IA3	No	Former
43	F	56	T1c	N0	M0	IA3	No	Never
44	F	56	T1c	N0	M0	IA3	No	Never
45	M	64	T1c	N0	M0	IA3	No	Current
46	M	81	T1c	N0	M0	IA3	No	Current
47	M	64	T1c	N0	M0	IA3	No	Never
48	M	64	T1c	N0	M0	IA3	No	Never
49	F	66	T1c	N0	M0	IA3	No	Never
50	F	72	T2a	N0	M0	IB	No	Never
51	M	47	T2a	N0	M0	IB	No	Never
52	M	47	T2a	N0	M0	IB	No	Never
53	M	58	T2a	N0	M0	IB	No	Former
54	M	58	T2a	N0	M0	IB	No	Current
55	M	54	T2a	N1	M0	IIB	No	Current
56	F	55	T1c	N1	M0	IIB	No	Never
57	M	63	T2a	N1	M0	IIB	No	Never
58	M	63	T2a	N1	M0	IIB	No	Never
59	F	75	T3	N0	M0	IIB	No	Never
60	M	75	T3	N0	M0	IIB	No	Former
61	M	54	T1b	N1	M0	IIB	No	Current
62	F	66	T2a	N1	M0	IIB	No	Never
63	F	79	T2a	N1	M0	IIB	No	Never
64	F	70	T4	N0	M0	IIIA	No	Never
65	F	70	T4	N1	M0	IIIA	No	Never
66	F	57	T1c	N2	M0	IIIA	No	Never
67	M	61	T3	N2	M0	IIIB	No	Former
68	F	49	T1c	N2	M0	IIIA	No	Never
69	F	49	T1c	N3	M0	IIIB	No	Never
70	M	61	T4	N0	M0	IIIA	No	Current
71	M	77	T4	N0	M0	IIIA	No	Current
72	M	65	T4	N1	M0	IIIA	No	Current
73	F	62	T4	N1	M0	IIIA	No	Never
74	M	65	T4	N2	M0	IIIB	No	Never
75	M	65	T4	N2	M0	IIIB	No	Current
76	M	70	T4	N3	M0	IIIC	No	Current
77	F	52	Tx	Nx	M1	IV	No	Never
78	F	62	Tx	NX	M1	IV	No	Never
79	M	68	Tx	Nx	M1	IV	No	Current
80	M	81	Tx	Nx	M1	IV	No	Current
81	F	46	Tx	Nx	M1	IV	No	Never
82	F	81	Tx	Nx	M1	IV	No	Never

Table S2 (*continued*)

Table S2 (*continued*)

No.	Sex	Age (year)	TNM stage			TNM stage	Neoadjuvant therapy	Smoking status
			Tumor	Nodes	Metastasis			
83	F	80	Tx	Nx	M1	IV	No	Never
84	F	74	Tx	Nx	M1	IV	No	Never
85	F	54	Tx	Nx	M1	IV	No	Current
86	F	68	Tx	Nx	M1	IV	No	Never
87	F	77	Tx	Nx	M1	IV	No	Never
88	M	77	Tx	Nx	M1	IV	No	Never
89	M	77	Tx	Nx	M1	IV	No	Never
90	M	67	Tx	Nx	M1	IV	No	Current
91	M	79	Tx	Nx	M1	IV	No	Never

LNMA, lung non-mucinous adenocarcinoma; M, male; F, female.

