Analysis of ultra-deep targeted sequencing reveals mutation burden is associated with gender and clinical outcome in lung adenocarcinoma

Dakai Xiao[1,2,3,*], Hui Pan[1,2,*], Fuqiang Li[4], Kui Wu[4], Xin Zhang[1,3], Jianxing He[1,3]

[1]Department of Thoracic Surgery, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[2]Research Center for Translational Medicine, The First Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

[3]Guangzhou Institute of Respiratory Disease and State Key Laboratory of Respiratory Disease, Guangzhou, China

[4]BGI-Shenzhen, Shenzhen, China

[*]These authors contributed equally to this work

Correspondence to: Jianxing He, **e-mail:** Drjianxing.he@gmail.com

ABSTRACT

Gender-associated difference in incidence and clinical outcomes of lung cancer have been established, but the biological mechanisms underlying these gender-associated differences are less studied. Recently we have characterized the genomic landscape of lung adenocarcinoma derived from Chinese population (Reference [1]). In this study we evaluated the clinical significance of mutation burden in lung adenocarcinoma and found that the male tumors harbored statistically greater burden of genetic alterations than female counterparts (Male median 3 (range 0–34) vs female median = 2 (0–24), male to female ratio = 1.636, 95% CI = 1.343–1.992) after adjustment of age at surgery, stage, smoking status. Kaplan-Meier survival analysis revealed that greater burden of genetic alterations was associated with worse overall survival. Moreover, multivariable analysis demonstrated mutation burden was an independent prognostic factor for the patients. Taken together, our analysis demonstrated gender disparity of mutation burden and their prognostic value in lung adenocarcinoma. This gender difference in mutation burden might provide an explanation for the distinct difference in the clinical outcomes between sexes in lung adenocarcinoma.

INTRODUCTION

Lung cancer is the most frequently diagnosed cancer and leading cause of cancer-related death among men. In contrast, among women lung cancer is the third most frequently diagnosed cancer and second leading cause of cancer–related death worldwide [2]. There were estimated 1,241,600 newly diagnosed cases of lung cancer among men and 583,100 newly diagnosed cases among women, and moreover men accounted for 70% of lung cancer-related deaths [2]. The gender-associated difference in clinical outcome of lung cancer has also been well confirmed in several population-based studies [3–6]. Although the gender difference in lung cancer is diminishing in North American and some European countries, this difference remains in other areas [7]. Such gender disparities in incidence and clinical outcomes of lung cancer have been attributed to the differences in smoking habit, environmental exposure, genetic variants, lifestyle and sex hormones activity between sexes [8, 9]. However, the biological mechanisms underlying the gender-associated differences in incidence and clinical outcomes are still less explored.

With the advances in cancer genomic sequencing, cancer is considered as a genomic disease resulting from progressive accumulation of genetic aberrations in the process of initiation and dissemination of tumor, and a small proportion of mutations have been identified as potential driver mutations, which are likely associated with initiation and progression of cancer [10–12].

Recently we have characterized the genetic landscape of lung adenocarcinoma in Chinese population and identified the significantly altered genes in this lethal disease using whole genome or whole exome sequencing [1]. Based-on the results from genomic sequencing, we thus developed a customized cancer panel targeting 51 genes which are closely associated with lung cancer. Although aberrations in well-known driver genes have been linked to clinical outcomes of patients, the prognostic value of molecular genotyping is limited by the low prevalence of some driver mutations [13]. Since the progression of cancer is accompanied by the accumulation of distinct genetic alterations, we test if the mutation burden in the primary tumors could predict the prognosis of the patients with lung adenocarcinoma. In this study, we found that male tumors harbored higher burden of genetic alterations than female counterparts, and greater burden of genetic alterations was also associated with worse clinical outcomes. The gender-associated difference in mutation burden will help gain insight into the biological mechanism of gender disparity in clinical outcome of lung adenocarcinoma.

With the increasing number of patients whose tumors are sequenced with targeted sequencing technology, this finding will provide the prognostic value of sequencing data other than choosing suitable targeted therapy for the patients with lung adenocarcinoma.

RESULTS

Mutational signature in lung adenocarcinoma patients

Recently we have characterized the genetic landscape of lung adenocarcinoma using whole-genome or whole exome sequencing. In the discovery set, we identified a set of genes that are recurrently mutated in Chinese patients with lung adenocarcinoma [1]. To further confirm the mutated genes in lung adenocarcinoma patients, we performed a hybrid-recapture and ultra-deep DNA sequencing on a set of 51 selected genes that were closely associated with the progression of lung cancer (Gene list was available in Supplementary Table S1).

Through the ultra-deep targeted sequencing, we identified total 962 alterations from 335 cases including missense, frameshift and synonymous mutations (Figure 1), among which, missense mutation (632/962, 65%) was the major type of genetic alterations. The median number of alterations per patient was 2 (range, 0–34). Most of patients (293/335,87.5%) had at least one alteration. Forty-two patients (12.5%) had no alteration, 79 patients (23.6%) had one alteration, 74 patients (22.1%) had 2 alterations, 66 patients (19.7%) had 3 alterations, and 74 patients (22.1%) had 4 or more alterations (Table 1).

The association of mutation burden with clinical features in lung adenocarcinoma patients

Targeted sequencing for a set of cancer genes is a strategy to screen actionable targets for the individual patients. In our cohort, clinical significance of individual genetic alteration was evaluated, only variants in TP53, STK11, BRAF, MET, LRP1B and MRC2 were found to be associated with worse overall survival [1]. However, the predictive value of these genes is limited by the low frequency of these genetic alterations.

Besides identifying actionable targets, we also tried to determine the clinical significance of mutation burden in lung adenocarcinoma with analysis of targeted sequencing data. To determine the association of mutation burden with clinical feature of lung adenocarcinoma, we firstly analyzed the number of genetic alterations in different groups stratified according to clinical features such as age at surgery, gender, TNM stage and smoking status. Univariate analysis with negative binomial regression model showed that number of genetic alterations was statistically significantly associated with gender and smoking status (Table 2). Male tumors harbored more total alterations than female tumors (Table 2, male median 3 (range 0–34) vs female median 2 (range 0–24), Male to female ratio = 1.636, 95% CI = 1.343–1.992). And tumors from smokers also had more total genetic alterations than those from non-smokers (Table 2, Smokers median 3 (range 0–34) vs non-smokers median 2 (range 0–32), $p = 0.00139$). Further multivariate analysis revealed that male tumors harbored statistically significantly higher burden of genetic alterations than female counterparts (Table 2, Multivariable male to female ratio = 1.524, 95% CI = 1.189–1.956) even after adjustment of age at surgery, stage, smoking status while there was no significant difference in mutation burden between smokers and non-smokers. Interestingly, as missense mutation was the major type of genetic alteration in our cohort, we also found that the male tumors harbored greater burden of missense mutation than female tumors after adjusting for other clinical features(Supplementary Table S2).

Tobacco smoking is a mutagen for lung cancer, and most of smokers were male in our cohort. To further test whether the gender-associated difference in mutation burden was independent of smoking status, we also compared the number of genetic alterations between sexes among non-smokers. This subgroup analysis showed that in non-smokers, male tumors also harbored more total alterations (Table 3. male to female ratio = 1.486, 95% CI = 1.124–1.968, $p = 0.0055$)) and missense mutations (Supplementary Table S3. Male-to-female ratio = 1.481, 95% CI = 1.099–1.996, $p = 0.0099$) than female tumors. However, among smokers, there was no statistically difference in mutation burden between sexes probably

due to the limited size of female smokers ($n = 9$) (Table 3 and S3). Taken together, these results indicated that there was a greater burden of genetic alterations among men compared to women, and also suggested tobacco smoking exposure alone could not fully explain the gender–associated difference in mutation burden. Other exogenous or endogenous factors might be involved in the mutational processes.

The clinical significance of mutation burden in lung adenocarcinoma

To investigate the clinical significance of mutation burden in lung adenocarcinoma patients, we performed Kaplan-Meier survival analysis and revealed that the patients whose tumors harbored greater burden of genetic alterations (> 4 total genetic alterations) was correlated with poorer overall survival (Figure 2A, log rank test, $p = 0.0198$). The mean survival time for the patients whose tumor harbored ≤ 4 and > 4 total alterations was 61 months and 48 months, respectively (Table 4). Of note, we also found the greater burden of missense mutations (> 3missense mutations), was also associated with worse overall survival (Figure 2B, log rank test, $p = 0.0145$). The mean survival time for the patients whose tumor harbored ≤ 3 and > 3 missense mutations was 61 months and 47 months, respectively. These data suggested that mutation burden was associated with overall survival of the lung adenocarcinoma patients.

In addition, univariate survival analysis with log rank test also showed that TNM stage and gender were correlated with overall survival (Table 4, $p < 0.001$ and $p = 0.013$, respectively), while smoking status was not associated with survival of the lung adenocarcinoma patients (Table 4). To examine the impact of the mutation burden on the survival of smokers and non-smokers, we also performed Kaplan-Meier analysis of mutation burden in smokers and non-smokers, respectively. The subanalyses revealed that higher burden of genetic alterations were tended to be correlated with worse overall survival at the margin of statistical significance among non-smokers and smokers (Figure S1A–S1D).

To further determine the prognostic value of mutation burden, we thus performed a multivariable analysis with Cox proportional hazard regression model. The results showed total genetic alterations > 4 (Table 4, HR = 1.81, 95% CI = 1.05–3.13, $p = 0.034$) was an independent risk factor for the clinical outcome of the patients with lung adenocarcinoma, and similarly, greater burden of missense mutation was also an independent risk factor for the survival (Supplementary Table S4, HR = 1.99, 95% CI = 1.12–3.52, $p = 0.019$). These results suggested that higher burden of genetic alterations represented an independent risk factor for the lung adenocarcinoma patients. Furthermore, consistent with previous reports [3–6, 14], female patients with lung adenocarcinoma had a better overall survival than male patients (Table 4, mean survival 63.0 vs 55.0 months for women and men, respectively, $p = 0.013$), and gender was an independent prognostic factor for lung adenocarcinoma patients (Table 4, HR = 2.04, 95% CI = 1.24~3.35, $p = 0.005$).

DISCUSSION

In present study we characterized the burden of genetic alterations and assessed their prognostic value in patients with lung adenocarcinoma and found that there was different mutation burden between male and female tumors, and higher burden of genetic alterations was associated with worse clinical outcome. We also showed the evidence to support that mutation burden is an independent prognostic factor for lung adenocarcinoma patients. These analyses demonstrated for the first time a
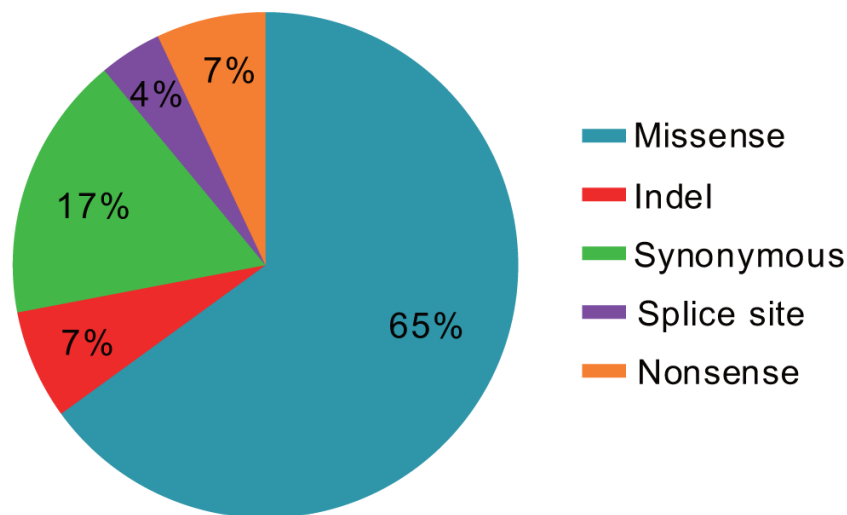

Figure 1: Pie chart showing the different types of genetic alterations found in 335 patients with lung adenocarcinoma.

Table 1: The frequency of genetic alterations (total mutation and missense mutations) identified in 335 patients with lung adenocarcinoma

Number of Total mutations	Number of patients (%)	Number of missense mutations	Number of patients (%)
0	42 (12.54)	0	72 (21.49)
1	79 (23.58)	1	98 (29.25)
2	74 (22.09)	2	86 (25.67)
3	66 (19.70)	3	40 (11.94)
4	27 (8.06)	> 3	39 (11.64)
> 4	47 (14.03)		

Table 2: Univariate and multivariate analysis with negative binomial regression comparing the mutation counts by variables in patients with lung adenocarcinoma ($n = 335$)

	No. of patients	No. of total mutations median (range)	Univariate analysis		Multivariate analysis	
			p value	Male-to-Female Ratio	p value	Male-to-Female Ratio
Gender				(95%CI Ratio)		(95%CI Ratio)
Female	152	2 (0–24)	(reference)		(reference)	
Male	183	3 (0–34)	**9.64E–07**	1.636 (1.343–1.992)	**0.000819**	1.524 (1.189~1.956)
Age						
< 65	228	2 (0–34)	(reference)		(reference)	
≥ 65	107	2 (0–32)	0.616		0.381	
Smoking						
No	199	2 (0–32)	(reference)		(reference)	
Yes	105	3 (0–34)	**0.00139**		0.361	
NA	31					
Stage						
I	82	2 (0–32)	(reference)		(reference)	
II	69	2 (0–22)	0.378		0.691	
III	154	2 (0–24)	0.893		0.532	
IV	29	2(0–34)	0.541		0.317	
NA	1					

NA, not applicable.

gender difference in burden of recurrent mutation in lung adenocarcinoma. This finding also provided a biological explanation for gender difference in clinical outcome of lung adenocarcinoma in our cohort

In this study, the mutation burden was obtained through ultra-deep targeted sequencing in a set of selected genes rather than missense mutation obtained through whole genomic sequencing. There are several reasons we focused on the burden of these selected genes: first, 1) although thousands of mutations have been identified in cancer genome sequencing, most of these mutations are 'passengers' [15], only a handful was recognized as driver mutations which are believed to be associated with cancer progression. 2) mutation recurrence has proven to be a

Table 3: Negative binomial regression comparing the counts of total mutations in smokers and non-smokers with lung adenocarcinoma

		No. of patients	No. of total mutations Median (Range)	*p* value	Male to female ratio (95% CI ratio)
Non-smoker	Female	128	2 (0–24)	(reference)	1.486 (1.124~1.968)
	Male	71	3 (0–32)	**5.52E–03**	
Smoker	Female	9	2 (1–5)	(reference)	1.496 (0.824–2.733)
	Male	96	3 (0–34)	0.18562	

powerful tool for the identification of cancer gene [16]. Most of genes in this panel have been identified to be recurrently mutated in Chinese population, suggesting that these mutations are highly potential drivers, not passenger mutations in cancer progression. 3) Targeted sequencing of a panel of cancer genes can be rapidly performed in several sequencing platform and has been applied to estimate the mutational burden [17].

It is interesting to note that Gupta et al.[18] also found the gender-associated difference in mutation burden in cutaneous melanoma through the analysis of exomic sequences. However, they found that greater burden of missense mutation was associated with improved survival although men harbored greater burden of mutation and exhibited higher incidence and poorer outcome than women [18]. They provided an 'immune fitness' hypothesis to interpret the inconsistency between mutation burden and clinical outcomes. Similar to melanoma, lung adenocarcinoma was also one type of mutagen-induced tumors and harbored highest number of mutations

[19, 20], and displayed gender-associated differences in incidence and clinical outcome between sexes. However, Gupta et al. found no statistically significant difference in mutation burden between sexes in lung adenocarcinoma. In contrast, we found a statistically significantly higher burden of mutations among men, and higher mutation burden was correlated with worse overall survival of lung adenocarcinoma, which was also consistent with several previous studies that men have worse clinical outcomes than women [3–6, 14]. We reasoned that the difference between our study and Gupta et al. mainly relied on the type of sequencing data we analyzed. Gupta et al. [18] performed the analysis based on the exome sequencing data, most of missense mutation was considered as passenger mutation, while our analysis was performed mainly based on the genes which were recurrently mutated in lung adenocarcinoma. These genetic alterations were most likely associated with cancer progression while the vast majority of mutations generated from exomic sequences were passenger mutations, which will definitely
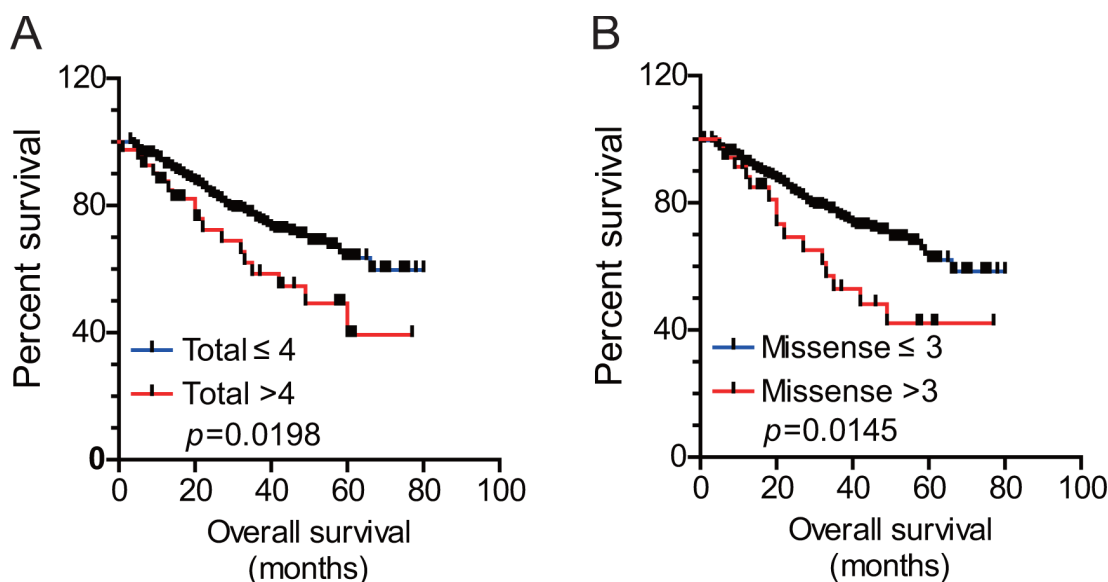
Figure 2: The burden of genetic alterations was associated with overall survival of the lung adenocarcinoma patients. Kaplan-Meier overall survival curves for patients with higher or lower burden of total mutations or missense mutation were shown in Figure 2A and 2B, respectively.

Table 4: Univariate and multivariate survival analysis of the patients with lung adenocarcinoma (*n* = 335)

	Mean Survival (months)	Univariate analysis			Multivariate Analysis		
		HR	**95% CI**	***p* value**	**HR**	**95% CI**	***p* value**
Age < 65 *vs* ≥ 65	59.0 *vs* 58.0	0.87	0.55~1.36	0.538	1.20	0.76~1.88	0.429
Gender female *vs* male	63.0 *vs* 55.0	1.71	1.12~2.61	0.013	2.04	1.24~3.35	0.005
Smoking No *vs* Yes	59.5 *vs* 56.2	0.98	0.62~1.57	0.947	0.69	0.40~1.17	0.384
Stage				< 0.0001			< 0.0001
I	69.1				(Reference)		
II	54.8				3.03	1.41~6.51	
III	56.5				3.95	1.96~7.98	
IV	31.4				8.76	3.86~19.9	
Total mutation ≤ 4 *vs* > 4	61.0 *vs* 48.4	2.19	1.13~4.24	0.020	1.81	1.05~3.13	0.034

underestimate the predominant role of driver mutations in cancer progression. These results suggested that we should be cautious to interpret the data considering the different cancer types and cancer panel detected.

It was noteworthy that although we and other group [11] already showed that smokers had more mutations, smoking status was not correlated with the overall survival in our cohort. In fact, inconsistent results about the relationship between clinical outcome and smoking status have been reported [13, 21]. As different mutational processes cause different mutation signatures [16, 19, 22], smokers and non-smoker displayed distinct mutational patterns, we reasoned that in lung adenocarcinoma, somatic mutations could be generated by several mutational processes other than tobacco smoking exposure, such as environmental or occupational exposure, CpG deamination or off-target modification of DNA by APOBEC family which generated distinct mutation signature [16, 23]. Indeed, in our previous study, four highly confident mutational signatures have been extracted from Chinese patients with lung adenocarcinoma. However, which signature contributes to cancer progression and is associated with gender and clinical outcome remains to be investigated in large samples.

Consistent with previous study in Asia [24], we also demonstrated that female patients showed survival advantage over male, and gender was identified as a prognostic factor for survival. Woolston et al. reported lung adenocarcinoma possessed ethnic and gender specific differences in genetic pattern [25].The female patients smoked much less than male [26], and showed distinct molecular characteristics such as, higher frequency of EGFR alterations, lower frequency of KRAS mutation [27–29] and distinct genetic variants in lung cancer susceptibility loci [30]. Other than these differences, we also showed that female patients in Asia harbored higher burden of mutation than male counterparts. As we also showed that lower burden of recurrent mutation was correlated with improved clinical outcomes, this finding raise the possibility that lower burden of recurrent mutation among women, at least partially, explain the survival advantage in women. However, whether this finding also applied to other ethnic populations is still unknown. Notably, we also found there was no statistically association between mutation burden and stage, which suggested that the detailed mechanisms of mutational process in cancer progression remained to be defined.

This study has similar limitation as described previously [18] such as the targeted sequencing delineated coding sequencing of a set of genes, other genes and non-coding regions were not included. Moreover, the limited sample size of smokers (*n* = 9) among women could decrease the statistical power and diminish the importance of smoking status in survival analysis.

All together, this study demonstrated the gender difference in burden of recurrent mutations which serve as an independent prognostic factor in Chinese patients with lung adenocarcinoma. However, further studies with large sample sizes are needed to clarify whether this finding is also applied to other ethnic populations.

MATERIALS AND METHODS

Patient information and specimen collection

This study was approved by the Institutional Review Board of The First Affiliated Hospital of Guangzhou Medical University, and three hundred and thirty-five lung adenocarcinoma patients with informed consent were enrolled in this study. The samples that satisfied the following conditions were included in this study and were subjected to being sequenced as described previously [1]. Firstly, The primary tumors and adjacent normal tissues were surgically resected from lung adenocarcinoma patients and the tumor specimens were then reviewed by independent pathologists to determine the histological subtype, TNM stage and tumor cellularity. Tumors that were not confirmed as adenocarcinoma, or patients with unavailable clinical data, or neoplasms with low tumor cell content (< 50% for primary tumors) were excluded. Secondly, DNA and RNA were extracted from tissues passed clinical evaluation, and quality-control estimation was performed to remove samples of low quality. The demographic information such as gender, age at diagnosis, smoking status and other clinical information were retrieved from the electronic medical record system in our hospital, and the survival status of the patients was followed up over the phone every three months. The patient characteristics were summarized in our previous study [1].

Data analysis

The mutational profiling for the patients was performed using ultra-deep targeted sequencing and the alterations in the 51 genes for each patient were extracted from our previous study [1]. Negative binomial regression was performed to predict the mutation counts in univariate and multivariable analysis with R package. Pearson's Chi-square was used to compare the association between the mutation burden and gender, age at diagnosis, smoking status. Kaplan-Meier curve was used to estimate the overall survival with log-rank test. Multivariable analysis with Cox proportional hazard regression model was carried out to determine the prognostic significance of mutation burden. The p value < 0.05 is considered as statistically significant, and all the statistical tests were two-sided. The analyses were performed using SPSS16.0, R package and GraphPad Prism 6.0.

GRANT SUPPORT

CONFLICTS OF INTEREST

The authors have no conflicts of interest to disclose.

REFERENCES

1. Wu K, Zhang X, Li F, Xiao D, Hou Y, Zhu S, Liu D, Ye X, Ye M, Yang J, Shao L, Pan H, Lu N, et al. Frequent alterations in cytoskeleton remodelling genes in primary and metastatic lung adenocarcinomas. Nat Commun. 2015; 6:10131.

2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. CA Cancer J Clin. 2015; 65:87–108.

3. Ferguson MK, Skosey C, Hoffman PC, Golomb HM. Sex-associated differences in presentation and survival in patients with lung cancer. J Clin Oncol. 1990; 8:1402–1407.

4. Ferguson MK, Wang J, Hoffman PC, Haraf DJ, Olak J, Masters GA, Vokes EE. Sex-associated differences in survival of patients undergoing resection for lung cancer. Ann Thorac Surg. 2000; 69:245–249.

5. Harichand-Herdt S, Ramalingam SS. Gender-associated differences in lung cancer: clinical characteristics and treatment outcomes in women. Semin Oncol. 2009; 36:572–580.

6. Hsu LH, Chu NM, Liu CC, Tsai SY, You DL, Ko JS, Lu MC, Feng AC. Sex-associated differences in non-small cell lung cancer in the new era: is gender an independent prognostic factor? Lung Cancer. 2009; 66:262–267.

7. Devesa SS, Bray F, Vizcaino AP, Parkin DM. International lung cancer trends by histologic type: male:female differences diminishing and adenocarcinoma rates rising. Int J Cancer. 2005; 117:294–299.

8. Paggi MG, Vona R, Abbruzzese C, Malorni W. Gender-related disparities in non-small cell lung cancer. Cancer Lett. 2010; 298:1–8.

9. Patel JD. Lung cancer in women. J Clin Oncol. 2005; 23:3212–3218.

10. Maraver A, Fernandez-Marcos PJ, Herranz D, Canamero M, Munoz-Martin M, Gomez-Lopez G, Mulero F, Megias D, Sanchez-Carbayo M, Shen J, Sanchez-Cespedes M, Palomero T, Ferrando A, et al. Therapeutic effect of gamma-secretase inhibition in KrasG12V-driven non-small cell lung carcinoma by derepression of DUSP1 and inhibition of ERK. Cancer Cell. 2012; 22:222–234.

11. Govindan R, Ding L, Griffith M, Subramanian J, Dees ND, Kanchi KL, Maher CA, Fulton R, Fulton L, Wallis J, Chen K, Walker J, McDonald S, et al. Genomic landscape of non-small cell lung cancer in smokers and never-smokers. Cell. 2012; 150:1121–1134.

12. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, Cho J, Suh J, Capelletti M, Sivachenko A,

Sougnez C, Auclair D, Lawrence MS, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012; 150:1107–1120.

13. Zheng D, Wang R, Pan Y, Zheng S, Zhang Y, Li H, Cheng C, Gong R, Li Y, Shen X, Hu H, Cai D, Cheng X, et al. Prevalence and Clinicopathological Characteristics of BRAF Mutations in Chinese Patients with Lung Adenocarcinoma. Ann Surg Oncol. 2015; 22:1284–91.

14. Minami H, Yoshimura M, Miyamoto Y, Matsuoka H, Tsubota N. Lung cancer in women: sex-associated differences in survival of patients undergoing resection for lung cancer. Chest. 2000; 118:1603–1609.

15. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. Nature. 2009; 458:719–724.

16. Martincorena I, Campbell PJ. Somatic mutation in cancer and normal cells. Science. 2015; 349:1483–1489.

17. Campesato LF, Barroso-Sousa R, Jimenez L, Correa BR, Sabbaga J, Hoff PM, Reis LF, Galante PA, Camargo AA. Comprehensive cancer-gene panels can be used to estimate mutational load and predict clinical benefit to PD-1 blockade in clinical practice. Oncotarget. 2015; 6:34221–34227. doi: 10.18632/oncotarget.5950.

18. Gupta S, Artomov M, Goggins W, Daly M, Tsao H. Gender Disparity and Mutation Burden in Metastatic Melanoma. J Natl Cancer Inst. 2015; 107.

19. Alexandrov LB, Nik-Zainal S, Wedge DC, Aparicio SA, Behjati S, Biankin AV, Bignell GR, Bolli N, Borg A, Borresen-Dale AL, Boyault S, Burkhardt B, Butler AP, et al. Signatures of mutational processes in human cancer. Nature. 2013; 500:415–421.

20. Lawrence MS, Stojanov P, Mermel CH, Robinson JT, Garraway LA, Golub TR, Meyerson M, Gabriel SB, Lander ES, Getz G. Discovery and saturation analysis of cancer genes across 21 tumour types. Nature. 2014; 505:495–501.

21. Nagy-Mignotte H, Guillem P, Vesin A, Toffart AC, Colonna M, Bonneterre V, Brichon PY, Brambilla C, Brambilla E, Lantuejoul S, Timsit JF, Moro-Sibilot D and Multidisciplinary Thoracic Oncology Group at Grenoble University H. Primary lung adenocarcinoma: characteristics by smoking habit and sex. Eur Respir J. 2011; 38: 1412–1419.

22. Alexandrov LB, Nik-Zainal S, Wedge DC, Campbell PJ, Stratton MR. Deciphering signatures of mutational processes operative in human cancer. Cell Rep. 2013; 3:246–259.

23. Burns MB, Lackey L, Carpenter MA, Rathore A, Land AM, Leonard B, Refsland EW, Kotandeniya D, Tretyakova N, Nikas JB, Yee D, Temiz NA, Donohue DE, et al. APOBEC3B is an enzymatic source of mutation in breast cancer. Nature. 2013; 494:366–370.

24. Sakurai H, Asamura H, Goya T, Eguchi K, Nakanishi Y, Sawabata N, Okumura M, Miyaoka E, Fujii Y. Survival differences by gender for resected non-small cell lung cancer: a retrospective analysis of 12,509 cases in a Japanese Lung Cancer Registry study. J Thorac Oncol. 2010; 5:1594–1601.

25. Woolston A, Sintupisut N, Lu TP, Lai LC, Tsai MH, Chuang EY, Yeang CH. Putative effectors for prognosis in lung adenocarcinoma are ethnic and gender specific. Oncotarget. 2015; 6:19483–19499. doi:10.18632/oncotarget.4287.

26. WHO. WHO Report on the Global Tobacco Epidemic, 2008 - The MPOWER package.

27. Mazieres J, Rouquette I, Lepage B, Milia J, Brouchet L, Guibert N, Beau-Faller M, Validire P, Hofman P, Fouret P. Specificities of lung adenocarcinoma in women who have never smoked. J Thorac Oncol. 2013; 8:923–929.

28. Sun Y, Ren Y, Fang Z, Li C, Fang R, Gao B, Han X, Tian W, Pao W, Chen H, Ji H. Lung adenocarcinoma from East Asian never-smokers is a disease largely defined by targetable oncogenic mutant kinases. J Clin Oncol. 2010; 28:4616–4620.

29. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M, Wistuba, II, Fong KM, Lee H, Toyooka S, Shimizu N, Fujisawa T, Feng Z, Roth JA, et al. Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst. 2005; 97:339–346.

30. Lan Q, Hsiung CA, Matsuo K, Hong YC, Seow A, Wang Z, Hosgood HD, 3rd, Chen K, Wang JC, Chatterjee N, Hu W, Wong MP, Zheng W, et al. Genome-wide association analysis identifies new lung cancer susceptibility loci in never-smoking women in Asia. Nat Genet. 2012; 44: 1330–1335.