



**HAL**  
open science

## **Somatic profile in lung cancers is associated to reproductive factors in never-smokers women: Results from the IFCT-1002 BioCAST study**

C. Fontaine-Delaruelle, J. Mazières, J. Cadranel, B. Mastroianni, C. Dubos-Arvis, P. Dumont, I. Monnet, E. Pichon, M. Locatelli-Sanchez, A. Dixmier, et al.

### ► To cite this version:

C. Fontaine-Delaruelle, J. Mazières, J. Cadranel, B. Mastroianni, C. Dubos-Arvis, et al.. Somatic profile in lung cancers is associated to reproductive factors in never-smokers women: Results from the IFCT-1002 BioCAST study. *Respiratory Medicine and Research*, 2020, 77, pp.58 - 66. 10.1016/j.resmer.2020.01.003 . hal-03491122

**HAL Id: hal-03491122**

**<https://hal.science/hal-03491122>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## **Somatic Profile in Lung Cancers Is Associated to Reproductive Factors in Never-Smokers Women: Results from the IFCT-1002 BioCAST study**

Clara Fontaine-Delaruelle<sup>1</sup>, M.D., Julien Mazières<sup>2</sup>, M.D., Ph.D., Jacques Cadranel<sup>3</sup>, M.D., Ph.D., Bénédicte Mastroianni<sup>4</sup>, M.D., Catherine Dubos-Arvis<sup>5</sup>, M.D., Patrick Dumont<sup>6</sup>, M.D., Isabelle Monnet<sup>7</sup>, M.D., Eric Pichon<sup>8</sup>, M.D., Myriam Locatelli-Sanchez<sup>1</sup>, M.D., Adrien Dixmier<sup>9</sup>, M.D., Bruno Coudert<sup>10</sup>, M.D., Séverine Fraboulet<sup>11</sup>, M.D., Pascal Foucher<sup>12</sup>, M.D., Eric Dansin<sup>13</sup>, M.D., Nathalie Baize<sup>14</sup>, M.D., Michel Vincent<sup>15</sup>, M.D., Pascale Missy<sup>16</sup>, Ph.D., Franck Morin<sup>16</sup>, M.Sc., Denis Moro-Sibilot<sup>16,17</sup>, M.D., Ph.D., Sébastien Couraud<sup>1,18</sup>, M.D., for the IFCT10-02 BioCAST investigators.

1. Service de Pneumologie Aigue Spécialisée et Cancérologie Thoracique, Hôpital Lyon Sud, Institut de Cancérologie des Hospices Civils de Lyon, Pierre Bénite, France
2. Service de pneumologie, Hôpital Larrey, Centre Hospitalier Universitaire, Université Paul Sabatier, Toulouse, France
3. Service de Pneumologie, Hôpital Tenon, AP-HP, Paris, France
4. Service de pneumologie, Institut de cancérologie des Hospices Civils de Lyon, Hôpital Louis Pradel, Bron, France
5. UCP d'oncologie thoracique, Centre de lutte contre le cancer François Baclesse, Caen, France
6. Service de Pneumologie, Centre Hospitalier de Chauny, Chauny, France
7. OncoThoParisEst, Service de Pneumologie, CHI Créteil, UPEC, Créteil, France
8. Service de pneumologie, CHRU de Tours, Hôpital Bretonneau, Tours, France
9. Service de pneumologie et oncologie thoracique, Centre hospitalier régional d'Orléans, Orléans, France
10. Oncologie Médicale, Centre GF Leclerc, Dijon, France
11. Service de pneumologie, Hôpital Foch, Suresnes, France
12. Fédération d'Oncologie Thoracique, CHU Dijon-Bourgogne, Hôpital du Bocage, Dijon, France
13. Département de Cancérologie Générale, Centre Oscar Lambret, Lille, France
14. Unité Transversale de Thérapeutiques Innovantes en Oncologie Médicale (UT-TIOM), CHU d'Angers, Angers, France

15. Service de pneumologie et cancérologie thoracique, Centre Hospitalier Saint Joseph et Saint Luc, Lyon ; et Minapath Développement Insavalor, Villeurbanne, France
16. Intergroupe Francophone de Cancérologie Thoracique (IFCT), Paris, France.
17. Clinique de pneumologie et oncologie thoracique, CHU Grenoble-Alpes, La Tronche, France
18. EMR 3738 Ciblage Thérapeutique en Oncologie, Faculté de Médecine Lyon Sud, Université Lyon 1, Oullins, France

**Corresponding author:**

Sébastien Couraud, Service de Pneumologie Aigue Spécialisée et Cancérologie Thoracique, Hôpital Lyon Sud, 165 chemin du grand Revoyet, 69495 Pierre bénite CEDEX, France; tel: +33 478 864 401; fax: +33 478 864 410; e-mail: sebastien.couraud@chu-lyon.fr

**Words (manuscript):** 2625 words

**References:** 41

**Figures:** 1

**Tables:** 5

**Additional files:** Yes

## **ABSTRACT (197 words)**

### **Background**

Lung cancer in women is on the rise, with a higher proportion occurring in lifelong never-smokers. Lung cancer in never-smokers (LCINS) exhibits a high frequency of driver oncogene alterations. In this study, we aimed to investigate whether exposure to reproductive factors in women with LCINS may modulate the molecular pattern.

### **Methods**

All newly diagnosed LCINSs were included in a prospective, observational study (IFCT-1002 BioCAST). Each patient responded to a questionnaire including reproductive factors. Biomarker test results were also collected.

### **Results**

260 women were included in this analysis, and 166 alterations were characterized. *EGFR* mutation frequency proved greater among patients with late menarche (74% in age >14 vs. 40% and 41% for 12-14 and ≤12 years, respectively;  $p=0.020$ ) and tended to decrease with increasingly late age at menopause. In multivariate analysis, *EGFR* mutation frequency increased by 23% per increment of 1 year of age at menarche ( $p=0.048$ ), and by 9% for each year at age at first birth ( $p=0.035$ ). *ALK* alteration frequency was greater in women with high parity (50% in ≥5 vs. 12% and 7% for 1-4 and nulliparity, respectively;  $p=0.021$ ).

### **Conclusion**

In a cohort of women LCINSs, female hormonal factors appear to impact molecular pattern.

**Trial registration:** The IFCT-1002 BioCAST study was registered on the US National Institutes of Health website, at [www.clinicaltrials.gov](http://www.clinicaltrials.gov), under the identifier NCT01465854 (date of registration: November 7, 2011)

**Keywords:** Lung cancer; Female hormonal factors; Oncogenic drivers; Never-smoker

**List of abbreviations**

LCINS: lung cancer in never-smokers

EGFR: epidermal growth factor receptor

ER: estrogen receptor

IFCT: French Cooperative Thoracic Intergroup

OC: oral contraceptive

HRT: hormone replacement therapy

BMI: body mass index

EE: ethinylestradiol

AOR: adjusted odds ratio

NSCLC: non-small cell lung cancer

## **BACKGROUND**

Lung cancer in never-smokers (LCINS) constitutes an interesting model to assess risk factors for lung cancer excluding active smoking (1). Since LCINS exhibits an original pattern of driver gene mutation or alteration, this population is also of interest for evaluating the association between molecular abnormalities and risk factor exposure (2,3).

A high proportion of LCINS occurs in women (1,4). Within our French LCINS IFCT-1002 BioCAST cohort, 81% of LCINS cases were female (2). The high prevalence may suggest that female hormones and/or reproductive factors play a role in lung carcinogenesis, which is further supported by many biological and epidemiological findings. It is also well-demonstrated that lung cancer cells express estrogen receptors (ER) more frequently than normal lungs, and that this expression is associated with patient outcomes (5). Data regarding molecular pattern in terms of reproductive and hormonal factor exposure in women is sparse, but a potential association between *EGFR* mutation and ER expression has been suggested (6–8).

The IFCT-1002 BioCAST study is a nationwide cohort aiming to collect LCINS cases over 1 year in France. Risk factor exposure was recorded via a standardized questionnaire investigating several variables regarding reproductive factors and the use of estrogen-based treatment. In this paper, we aim to examine whether exposure to reproductive factors and/or hormonal intake in women with LCINS modulates the somatic profile.

## **PATIENTS AND METHODS**

### ***Population***

We utilized data from the IFCT-1002 BioCAST cohort. The analysis was restricted to women who completed the questionnaire and were tested for at least one biomarker. The study design has been reported elsewhere, as is the case for the cohort overall description (2). In short, BioCAST is a prospective, multicenter, observational study designed to describe the clinical, pathological, and molecular epidemiology of LCINS in a French population. It enrolled consecutive, newly diagnosed NSCLC patients claiming to be never-smokers (less than 100 cigarettes in their lifetime). Patients were surveyed using a standardized questionnaire during a phone interview with a study team member. This 17-page questionnaire requested information on demographics, and exposure to risk factors including hormonal, and reproductive factors (women only). Additional medical data was collected directly from participating physicians.

The study was conducted in 75 centers throughout France, from November 1, 2011, to January 31, 2013, and sponsored by the French Cooperative Thoracic Intergroup (IFCT).

### ***Reproductive Factors and Hormonal Intake***

Data collected included age at menarche, age at menopause, age at first live birth, number of pregnancies and live births (parity). We also record the oral contraceptive (OC) drug(s) and post-menopause hormone replacement therapies (HRTs) used during lifetimes (See Supplemental File 1). All were self-reported.

### ***Tumor Somatic Mutation Analysis***

Each participating physician was asked to systematically order tests for somatic mutations in *EGFR* and *KRAS*, as well as *ALK* alterations, as routinely performed in

France (9,10). Investigator sites were also encouraged to request *BRAF*, *HER2*, and *PI3KCA* mutation analyses. Assuming most mutations to be mutually exclusive, centers were allowed to forgo further mutation testing if one mutation was found (11). The final results were collected for each patient. Taking into account the most frequent mutations among never-smokers, we considered patients with no established mutation after testing for at least *EGFR*, *KRAS*, and *ALK* to be “wild type/unknown”.

### **Statistics**

Categorical variables were expressed in percentages. Proportion comparison was conducted via the Chi-squared test or with Fisher’s exact test. We employed the one-sample Kolmogorov-Smirnov test to assess the plausibility of the normal distribution assumption for continuous variables. Normally distributed continuous variables were expressed in mean and standard deviation. Comparison of means was performed via the two-sided Student’s t-test. Differences in continuous variable distribution between two independent samples were assessed using the Mann–Whitney U test. Continuous variables were categorized according to relevant categories or tertile categories. We applied a binary logistic regression model to assess the risk of mutation for each considered gene. We generated a model adjusted for age (continuous), total years of passive smoking (continuous), and body mass index (BMI) (continuous). For *EGFR* mutations, we performed two additional multivariate models, setting all reproductive variables as continuous. The first model was adjusted for age at diagnosis, BMI, total duration of passive smoking exposure (3), age at menopause, parity, number of pregnancies, and age at first birth. The second model was additionally adjusted for family history of lung cancer (number of relatives),



personal history of cancer (n), total duration of exposure to cooking oil fumes (years), exposure to fumes of solid fuels for cooking/heating (% of lifetime). All variables were continuous.

Missing values were reported as such, and all tests were two-sided. All statistics were performed using the SPSS V20 software (IBM SPSS Statistics, New York, USA).

## **RESULTS**

### ***Population***

Among the 384 patients included in IFCT-1002 BioCAST, 260 were analyzed in this study (Fig 1). The main descriptive results of the population are displayed in Table 1. Differences between men and women were previously reported (2). Data regarding cancer and diagnosed somatic alterations is reported in Table 2. We observed a 44% prevalence for *EGFR* mutations (n=109), 13% for *ALK* alterations (n=19), 7% for *KRAS* mutations (n=15), 5% for *BRAF* mutations (n=8), 4% for *HER2* mutations (n=6), and 2% for *PI3K* (n=3). Six patients (3%) exhibited at least two alterations in two different genes.

### ***Mutation Frequency According to Reproductive Factors and Hormonal Intake***

Table 3 reports mutation rates for *EGFR*, *KRAS*, *ALK*, *BRAF*, and *HER2* regarding relevant reproductive variables.

We found an increase in *EGFR* mutation frequency for patient exhibiting late menarche (74% in age >14 vs. 40% and 41% for 12-14 and ≤12 years, respectively; p=0.020). By contrast, a non-significant descending trend of *EGFR* mutations was

observed with increasing age at menopause and increasingly long duration of sex life (Additional files 1-3). Although not significant, we determined a higher rate of *EGFR* mutations in patients with older age at first birth (62% for those >30y) compared to other categories (40% for 20-30y, 41% for ≤20y; p=0.216) (Additional file 4). We found no significant trend or association with regard to *HER2* mutations.

*ALK* alteration frequency was greater in patients with the highest parity (50% in ≥5; p=0.021), and marginally so with increased number of pregnancies (31% in ≥5; p=0.075) (Additional files 5-6).

*KRAS* mutation frequency increased with rising age at menopause (2% in age category ≤45y, 8% in 45-55y, and 24% in >55y, respectively; p=0.024), but was not associated with age at menarche and marginally with sex life duration. We also observed a non-significant *KRAS* mutation frequency decrease associated with rising number of pregnancies and parity. We found the same trends with regard to *BRAF* mutations. In particular, *BRAF* mutation frequency proved higher in cases of long sex life duration (3% in ≤40y category and 13% in >40y, respectively; p=0.024).

Use of OC alone (used vs. never used / duration of use), use of HRT alone (used vs. never used / duration of use) (Table 3), and dose of ethinylestradiol (EE) (divided into 3 tertile categories) showed no association with any mutational pattern (not shown).

### ***Logistic Regression Analysis***

Logistic regression analysis results are presented in Table 4. Interestingly, significant associations in univariate analysis as well as tendencies all remained significant after adjustment.

Thus, the risk of *EGFR* mutation was five times greater for patients with menarche after 14 years ([95%CI 1.5-17.0];  $p=0.008$ ). *EGFR* mutation frequency also proved 2.7 times greater [0.9-8.1] ( $p=0.071$ ) for patients of the highest age at first birth (>30y).

The risk of *ALK* alteration was significantly greater in case of high parity (adjusted odds ratio (AOR)=20.6 [95%CI 1.4-301.5];  $p = 0.027$ ) and marginally elevated in case of high pregnancy.

*KRAS* mutations were more frequent in case of menopause after 55 years (AOR=16.0 [1.5-167.6];  $p=0.021$ ), and increased risk of *BRAF* mutations was observed in correlation with a sex life exceeding 40 years (AOR=7.8 [95%CI 1.4-44.5];  $p=0.021$ ).

In our additional adjusted analysis using reproductive variables, *EGFR* mutation frequency was associated significantly with age at menarche, with an AOR of 1.231 ([95%CI 1.002-1.512];  $p=0.048$ ), as well as age at first birth (AOR=1.094 [95%CI 1.006-1.189];  $p=0.035$ ), both as continuous variables (for an increment of 1 year in each variable). These associations remained significant after a more comprehensive adjustment for lifestyle variables (Table 5). No significant associations were found with these models in other biomarkers due to the low mutation numbers (not shown).

## **DISCUSSION**

Female reproductive factors were associated significantly with specific patterns of mutation in never-smoker women with lung cancer. In particular, *EGFR* mutations were associated significantly with increasing age at menarche and age at first birth. In addition, *ALK* alterations were associated positively with both parity and number of pregnancies.

Many studies report gender differences in clinical presentation and biology of lung cancer. The epidemiology appears different in women, with an increasing incidence worldwide, as compared to men<sup>1</sup>, and with a highest proportion of lung cancer occurring in lifelong never-smokers (1,4). Moreover, lung cancer prognosis appears slightly more favorable for women than men (12), while female lung cancer exhibits a highest frequency of *EGFR* mutations (13). These differences suggest a potential role of hormonal factors in lung carcinogenesis of women (8,14). Indeed, early (and/or long) exposure to hormones may favor the emergence of hormone-dependent cancers, by directly promoting tumorigenesis and/or by exacerbating a pre-existing cancer. Moreover, early menopause leads to early senescence and thus to greater risk of malignancies. Epidemiologically, the role of hormones is the subject of many studies. In the American NIHAARP cohort (185,017 women), late menarche was a protective factor, while early menopause was identified as a risk factor for lung cancer (15). The link between parity and age at first birth remains unclear (1,15). Biologically, ERs are known to be expressed in normal lung tissue, as well as in lung tumors in both men and women (5,8). Several *in vitro* and *in vivo* studies yielded evidence supporting the hypothesis that estrogens stimulate non-small cell lung cancer (NSCLC) cell proliferation through ER-mediated signaling, while anti-estrogens inhibit NSCLC cell growth. That said, data remains contradictory regarding the role of the two ER isoforms ( $\alpha$  and  $\beta$ ), the range of antibodies used, and the absence of a validated threshold or score (7,8,16–18). Similarly, there is evidence concerning the frequent expression of progesterone receptors in NSCLC tumor cells (18). Interactions between the ER and EGFR pathways have been investigated. EGFR protein expression is down-regulated in response to estrogens and up-regulated in response to an-

---

<sup>1</sup> WHO/IARC Globocan 2012 database; available at [http://globocan.iarc.fr/Pages/fact\\_sheets\\_cancer.aspx](http://globocan.iarc.fr/Pages/fact_sheets_cancer.aspx); last accessed on 03/16/2017.

ti-estrogens, suggesting a reciprocal control between EGFR and ER pathways (5). Estrogen directly stimulates transcription of estrogen-responsive genes in lung cell nucleus, but also trans-activating growth factor signaling pathways, such as the EGFR pathway (19). A meta-analysis of five studies also found a positive association between high ER $\beta$  expression in lung cancer cells and *EGFR* mutation (20). No correlation was determined between ER $\alpha$  expression and *EGFR* mutation by contrast to findings by Mazieres *et al.* and Raso *et al.*(7,18). Nose *et al.* also reported a correlation between *EGFR* mutation and ER $\beta$ , and proposed that ER expression might be associated with a response to EGFR tyrosine kinase inhibitor (6,21). However, the IFCT-1003 LADIE trial (NCT01556191), aiming to test anti-estrogen in addition to EGFR TKI in EGFR mutated or naïve post-menopausal women with advanced stage lung adenocarcinoma, showed no benefit of the combo (22).

We found *EGFR* mutation frequency to be associated with shorter estrogen exposure. These results might be explained by the role of estrogen in the down-regulation of EGFR expression, although not consistent with findings between higher ER expression and *EGFR* mutation frequency (5,20). Nonetheless, we can hypothesize that an *EGFR* mutation may interact with normal negative control between ER and EGFR and thus enhance tissue ER expression (Additional file 7). At last, the association between *EGFR* mutation and shorter lifetime estrogen exposure is probably more complex, additionally entailing imbalance between the two ER isoforms as well as progesterone receptors, and the role of genomic polymorphisms (23). Our findings may furthermore be put into perspective by taking into account the decreasing secular trends in terms of age at menarche worldwide. This decrease seems to have reached a plateau in Europe and the US (24,25), but not yet in Asia (26–29), where women still display a higher age at menarche. This observation and our results may

partially explain ethnic disparities regarding *EGFR* mutation frequency (4,30). Finally, we can only hypothesize regarding the role of high estrogen/progesterone exposure in senescent maternal tissues in order to explain the association between later age at first birth and *EGFR* mutations.

We found *ALK* alterations to be consistently associated with high parity and an elevated number of pregnancies. Interestingly, although not expressed in normal lungs, *ALK* plays a key role in early embryogenesis of nervous system development (31). Thus, *ALK* expression is high during embryogenesis and down-regulated during later pregnancy and after birth (31). We might hypothesize that *ALK*<sup>+</sup> fetal cells circulating in maternal blood lead to fetal cell microchimerism, promoting *ALK*-dependent tumorigenesis. Therefore, repeated pregnancies may increase this risk of fetal chimerism tumorigenesis (33,34). We have learned from *ALK*<sup>+</sup>, anaplastic, large-cell lymphomas that a normal anti-*ALK* immune response exists in adults (35,36). Thus, repeated pregnancies and maternal immune system exposure to *ALK*<sup>+</sup> fetal cells may increase the immune tolerance of *ALK*<sup>+</sup> tumor cells in multiparous women (Additional file 8) (37). Such assumptions are supported by the findings by Dagogo-Jack *et al.*, who observed a high frequency of *ALK* rearrangements in their series of lung cancers diagnosed during pregnancy or the peripartum period (n=6/8) (38).

We found a correlation between long sex life duration and *BRAF* mutations, suggesting a potential role of estrogen impregnation. This is consistent with findings in thyroid cancer, with estrogen being known to stimulate the RAS/RAF/MAP kinase signaling pathway in this cancer (39). Moreover, a study showed that *BRAF* mutations were correlated with ER and progesterone receptor expression in thyroid cancer cells (40).

Our study exhibits certain limitations. Except for *EGFR*, the absolute number of mutations/alterations was low (N=19 *ALK* fusions; N=15 *KRAS* mutations) or very low (N=8 *BRAF* mutations; N=6 *HER2* mutations), thus limiting our statistical power and ability to conduct comprehensive multivariate analysis, while reproductive factors are possibly correlated with each other and synergistic. In addition, reporting of reproductive factors was declarative and possibly entailed recall bias. That said the low rate of missing values and high quality of questionnaire responses for other variables indicate that this bias is probably weak. Finally, we performed multiple statistic tests in this analysis. Thus, the risk of alpha risk inflation is significant. In this analysis, we used 5 independent variables suggesting that the p value threshold should be 0.01 ( $0.05 / 5$ ) instead of 0.05 using the Bonferonni's correction. However, our study is exploratory and this kind of correction is usually not performed (41). Our study also exhibits certain strengths. To our knowledge, our cohort is the only one to comprehensively assess five driver oncogenes with several detailed reproductive factors in women. In addition, these findings were observed in the setting of never-smoker patients, thereby ruling out the potential confounding role of active smoking.

## **CONCLUSION**

Although our findings are observational and not explanatory, female reproductive factors appear to impact molecular pattern in never-smoker women with lung cancer. Our findings provide a unique and comprehensive overview of the relationship between female reproductive factors and lung cancer molecular pattern and may advance new conjectures in order to explain gender and ethnic differences regarding lung cancer epidemiology. Our findings also underline the significance of gender effects, with possible therapeutic impact, in patients with targetable oncogenic driver. For instance, we found a negative association between duration to estrogen

exposure and *EGFR* mutation frequency. This observation is in line with the findings of the IFCT-1003 LADIE trial showing that anti-estrogen may be interesting in *EGFR* WT patient but not in *EGFR* mutated patients.

## **Declarations**

### ***Ethics approval and consent to participate***

The Sud-Est IV (Lyon, France) ethics committee approved the study protocol on September 13, 2011. The Advisory Committee on Information Processing for Health Research (CCTIRS) authorized the use of a computerized database on September 8, 2011, and the National Commission for Data Protection (CNIL) was consulted on September 23, 2011, in accordance with French law.

All participants signed an informed consent form before any study procedures begin.

### ***Consent for publication***

Not applicable

### ***Availability of data and material***

All data generated or analyzed during this study are available from the corresponding author upon reasonable request.

### ***Declaration of interests***

Dr. Fontaine-Delaruelle reports non-financial support from Boehringer Ingelheim, outside the submitted work.

Dr. Cadranel reports grants and personal fees from Astra Zeneca, Pfizer, Novartis, Boehringer Ingelheim; personal fees from Roche, BMS, Lilly, outside the submitted work.



Dr. Fraboulet reports personal fees and non-financial support from Novartis; non-financial support from Lilly, outside the submitted work.

Dr. Dansin reports personal fees from BMS and Lilly; grants from Roche, outside the submitted work.

Dr. Moro-Sibilot reports personal fees from Roche, Eli Lilly, Pfizer, Novartis, Astra Zeneca, BMS, MSD, Boehringer Ingelheim, outside the submitted work.

Dr. Couraud reports grants and personal fees from Astra Zeneca, Boehringer Ingelheim, Lilly, Pfizer, Roche; grants from Pierre Fabre, during the conduct of the study; grants from BMS, Novartis, Merck, Amgen; grants and personal fees from Astra Zeneca, Pfizer, Lilly, Laidet Medical; personal fees and non-financial support from Roche, MSD, Chugai, Boehringer Ingelheim; grants and non-financial support from Sysmex Inostics; personal fees from Exact Science; non-financial support from Eformed, outside the submitted work.

Other authors have nothing to disclose.

### ***Funding sources***

The IFCT-1002 BioCAST study was supported by research grants from Astra-Zeneca, Boehringer Ingelheim, Lilly, Pfizer, Pierre-Fabre, and Roche. The funding sources played no role in the design, analysis, and interpretation of results, and thus the authors were independent from the funding sources.

### ***Authors' contributions***

SC, FM, DMS, PM designed the study, all the authors participated in the data acquisition, SC, CFD, JM, JC analyzed and interpreted the data and wrote the manuscript; the final version was revised and approved by all the authors. No writing assistance was provided to the authors.

**Table 1.** Main Characteristics of the Analyzed Population

<b>Variable</b>	<b>Category</b>	<b>n=260</b>	<b>%</b>
<b>Sex</b>	Women	260	100%
<b>Age at diagnosis</b>	Mean±SD	69.8±11.1	
<b>Origin in terms of global region</b>	Europe	237	91%
	Africa	11	4%
	Asia	8	3%
	Caribbean	4	2%
<b>Education level</b>	High school and above	108	42%
	Secondary school	58	23%
	None / primary school	91	35%
	Missing	3	
<b>Marital status</b>	Married / common-law	152	58%
	Divorced/separated	28	11%
	Widowed	64	25%
	Single	15	6%
	Other	1	0%
<b>Body mass index (kg/m<sup>2</sup>)</b>	Median/IQR	23.9/5.7	
	Underweight/normal	153	60%
	Pre-obese	69	27%
	Obese	34	13%
	Missing	4	
<b>Alcohol intake (standard glasses per day)</b>	0-1	249	96%
	≥2	10	4%
	Missing	1	
<b>Exposure to passive smoking (at home/workplace)</b>	Yes	180	70%
	No	79	30%
	Missing	1	
	Mean duration* (±SD), in years	31.5±21.7	
<b>Exposure to cooking oils (cooking dish year)</b>	<10 CDY	195	78%
	≥10 CDY	54	22%
	Missing	11	
<b>Exposure to solid fuels (at home)</b>	≥50% of lifetime	50	25%
	Missing	7	
<b>Family history of lung cancer (≥2 relatives)</b>	Yes	63	24.3%
	Missing	1	
<b>Personal history</b>	Lung infection**	90	35%
	Chronic bronchial disease***	36	14%
	Cancer (≥1 cancer diagnosis)	47	18%
<b>Occupational exposure</b>	≥1 carcinogen	21	8%
<b>Age at first birth</b>	Mean±SD	23.9±4.0	
	Missing	8	
<b>Number of pregnancies</b>	Mean±SD	2.6±1.7	
	Missing	9	
<b>Parity</b>	Mean±SD	2.2±1.4	

	Nulliparity	27	10.4%
	Missing	7	
<b>Age at menarche</b>	Mean±SD	13.0±1.7	
	Missing	20	
<b>Menopause</b>	Menopausal	239	92%
	Not reached	20	8%
	Missing	1	
<b>Age at menopause</b>	Mean±SD	49.3±5.6	
	Missing	25	
<b>OC</b>	Never	150	58%
	At least once	109	42%
	Missing	1	
	Mean duration (±SD), in years	10.5±9.0	
	Mean cumulat. EE dose (mg)	9.8±7.1	
<b>HRT (oral/local routes)</b>	Never	193	75%
	At least once	65	25%
	Missing	2	
	Mean duration (±SD), in years	10.0±8.6	

Abbreviations: EE, ethinylestradiol; SD, standard deviation; IQR, interquartile range; CDY, cooking dish year; OC, oral contraceptive; HRT, hormone replacement therapy.

\*Exposed only.

\*\*Pertussis or tuberculosis or pneumonia.

\*\*\*Asthma or emphysema or COPD or bronchiectasis.

**Table 2.** Cancer Data and Somatic Alterations

<b>Variable</b>	<b>Category</b>	<b>n=260</b>	<b>%</b>
<b>Histological subtype</b>	Squamous cell carcinoma	18	7%
	Adenocarcinoma	225	87%
	Large-cell carcinoma	10	4%
	Adenosquamous carcinoma	3	1%
	Sarcomatoid carcinoma	2	<1%
	Carcinoma NOS	2	<1%
	<b>TNM stage at diagnosis</b>	Stage I	22
Stage II		15	6%
Stage III		33	13%
Stage IV		188	73%
Missing		2	
<b>Somatic aberrations</b>	<i>BRAF</i>	8	5%*
	<i>EGFR</i>	109	44%*
	<i>KRAS</i>	15	7%*
	<i>HER2</i>	6	4.1%*
	<i>ALK</i>	19	13%*
	<i>PI3KCA</i>	3	2%*
	No mutation*	59	28%*
Multiple	6	3%*	

Abbreviations: NOS, not otherwise specified.

\*Tested for at least EGFR, KRAS, and ALK, and optionally HER2, BRAF, and PIK3.

**Table 3.** Univariate Analysis of Somatic Mutation Rate Regarding Reproductive Factors and Hormonal Intake

Biomarker	EGFR		P	KRAS		P	ALK		P	BRAF		P	HER2		P	
	WT	Mt		WT	Mt		WT	Mt		WT	Mt		WT	Mt		
Variable	Category	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)	N (%)		
Age at first birth	≤20	30 (60%)	20 (40%)	39 (91%)	4 (9%)	21 (78%)	6 (22%)	35 (97%)	1 (3%)	29 (94%)	2 (6%)	0.216	0.316	0.262	0.717	0.607
	20-30	85 (57%)	65 (43%)	124 (95%)	6 (5%)	81 (89%)	10 (11%)	97 (95%)	5 (5%)	90 (97%)	3 (3%)					
	≥30	8 (38%)	13 (62%)	14 (88%)	2 (13%)	7 (78%)	2 (22%)	11 (92%)	1 (8%)	8 (100%)	0 (0%)					
	Total	123 (56%)	98 (44%)	177 (94%)	12 (6%)	109 (86%)	18 (14%)	143 (95%)	7 (5%)	127 (96%)	5 (4%)					
Number of pregnancies	None	15 (63%)	9 (38%)	15 (83%)	3 (17%)	14 (93%)	1 (7%)	11 (92%)	1 (8%)	12 (92%)	1 (8%)	0.561	0.236	0.075	0.531	0.744
	1 to 4	106 (54%)	89 (46%)	155 (93%)	11 (7%)	97 (88%)	13 (12%)	127 (95%)	7 (5%)	110 (96%)	4 (4%)					
	≥5	17 (63%)	10 (37%)	24 (96%)	1 (4%)	11 (69%)	5 (31%)	18 (100%)	0 (0%)	18 (95%)	1 (5%)					
	Total	138 (56%)	108 (44%)	194 (93%)	15 (7%)	122 (87%)	19 (13%)	156 (95%)	8 (5%)	140 (96%)	6 (4%)					
Parity	None	15 (58%)	11 (42%)	17 (85%)	3 (15%)	14 (93%)	1 (7%)	13 (93%)	1 (7%)	13 (93%)	1 (7%)	0.728	0.260	0.021	0.753	0.727
	1 to 4	116 (55%)	94 (45%)	168 (93%)	12 (7%)	107 (88%)	15 (12%)	136 (95%)	7 (5%)	121 (96%)	5 (4%)					
	≥5	8 (67%)	4 (33%)	10 (100%)	0 (0%)	3 (50%)	3 (50%)	8 (100%)	0 (0%)	7 (100%)	0 (0%)					
	Total	139 (56%)	109 (44%)	195 (93%)	15 (7%)	124 (87%)	19 (13%)	157 (95%)	8 (5%)	141 (96%)	6 (4%)					
Age at menarche	≤12	61 (59%)	43 (41%)	81 (92%)	7 (8%)	50 (85%)	9 (15%)	63 (93%)	5 (7%)	67 (97%)	2 (3%)	0.020	0.937	0.892	0.266	0.415
	12-14	68 (60%)	45 (40%)	89 (93%)	7 (7%)	58 (87%)	9 (13%)	75 (97%)	2 (3%)	56 (93%)	4 (7%)					
	>14	5 (26%)	14 (74%)	17 (94%)	1 (6%)	9 (90%)	1 (10%)	14 (100%)	0 (0%)	13 (100%)	0 (0%)					
	Total	134 (57%)	102 (43%)	187 (93%)	15 (7%)	117 (86%)	19 (14%)	152 (96%)	7 (4%)	136 (96%)	6 (4%)					
Age at menopause	≤45	27 (56%)	21 (44%)	41 (98%)	1 (2%)	19 (79%)	5 (21%)	30 (97%)	1 (3%)	29 (97%)	1 (3%)	0.932	0.024	0.424	0.433	0.624
	45-55	81 (58%)	58 (42%)	108 (92%)	9 (8%)	74 (89%)	9 (11%)	84 (94%)	5 (6%)	76 (94%)	5 (6%)					
	>55	14 (61%)	9 (39%)	13 (76%)	4 (24%)	10 (83%)	2 (17%)	14 (88%)	2 (13%)	10 (100%)	0 (0%)					
	Total	122 (58%)	88 (42%)	162 (92%)	14 (8%)	103 (87%)	16 (13%)	128 (94%)	8 (6%)	115 (95%)	6 (5%)					
Sex life duration	≤40y	88 (56%)	68 (44%)	126 (93%)	10 (7%)	77 (87%)	12 (13%)	100 (97%)	3 (3%)	86 (95%)	5 (5%)	0.396	0.463	0.860	0.024	0.684
	>40y	31 (63%)	18 (37%)	32 (89%)	4 (11%)	23 (85%)	4 (15%)	26 (87%)	4 (13%)	27 (96%)	1 (4%)					
	Total	119 (58%)	86 (42%)	158 (92%)	14 (8%)	100 (86%)	16 (14%)	126 (95%)	7 (5%)	113 (95%)	6 (5%)					
Duration of OC and/or HRT use	0-10y	93 (57%)	71 (43%)	126 (91%)	12 (9%)	82 (89%)	10 (11%)	105 (95%)	5 (5%)	89 (95%)	5 (5%)	0.402	0.366	0.538	0.957	0.427
	≥10y	35 (51%)	34 (49%)	57 (95%)	3 (5%)	35 (85%)	6 (15%)	44 (96%)	2 (4%)	42 (98%)	1 (2%)					
	Total	128 (55%)	105 (45%)	183 (92%)	15 (8%)	117 (88%)	16 (12%)	149 (96%)	7 (4%)	131 (96%)	6 (4%)					

Abbreviations: WT, wild type; Mt, mutated; OC, oral contraceptive; HRT, hormone replacement therapy.

**Table 4.** Results of Multivariate Analysis for Biomarker Mutation Occurrence Regarding Exposure to Reproductive Factors (Categorical Variable). For All Categories, the First Patient Group Is the Reference. Adjusted for Age (Continuous), Total Exposure to Passive Smoke in Years (Continuous), and body mass index (Continuous)

Biomarker	EGFR			KRAS			ALK			BRAF			HER2				
	Variable	Categ.	AOR	95%CI	P	AOR	95%CI	P	AOR	95%CI	P	AOR	95%CI	P	AOR	95%CI	P
Age at first birth (years)	≤20		1.00			1.00			1.00			1.00			1.00		
	20-30		1.19	[0.61-2.33]	0.605	0.54	[0.13-2.18]	0.385	0.39	[0.12-1.24]	0.111	2.61	[0.25-27.63]	0.425	0.34	[0.06-2.79]	0.352
	≥30		2.74	[0.92-8.18]	0.071	2.01	[0.28-14.15]	0.485	0.52	[0.05-5.31]	0.584	4.86	[0.23-101.61]	0.308	0.00		0.999
Pregnancies (N)	0		1.00			1.00			1.00			1.00			1.00		
	1-4		1.50	[0.62-3.61]	0.369	0.33	[0.08-1.39]	0.131	1.77	[0.21-15.09]	0.601	0.75	[0.08-7.25]	0.807	0.54	[0.05-5.46]	0.600
	≥5		1.06	[0.33-3.34]	0.927	0.12	[0.01-1.41]	0.092	8.65	[0.81-92.00]	0.074	0.00		0.998	0.69	[0.04-13.36]	0.805
Parity (N)	0		1.00			1.00			1.00			1.00			1.00		
	1-4		1.16	[0.51-2.67]	0.727	0.35	[0.09-1.46]	0.149	1.92	[0.23-16.19]	0.549	0.74	[0.08-6.98]	0.791	0.62	[0.06-6.02]	0.681
	≥5		0.71	[0.17-3.04]	0.649	0.00		0.999	20.63	[1.41-301.48]	<b>0.027</b>	0.00		0.999	0.00		0.999
Age at menarche (years)	≤12		1.00			1.00			1.00			1.00			1.00		
	12-14		1.07	[0.60-1.89]	0.824	0.88	[0.28-2.80]	0.828	0.77	[0.26-2.24]	0.628	0.40	[0.07-2.45]	0.324	3.57	[0.54-23.79]	0.189
	>14		5.11	[1.53-17.01]	0.008	0.67	[0.07-6.31]	0.726	0.00		0.999	0.00		0.999	0.00		0.999
Age at menopause (years)	≤45		1.00			1.00			1.00			1.00			1.00		
	45-55		0.91	[0.46-1.80]	0.787	4.23	[0.50-35.96]	0.186	0.31	[0.08-1.15]	0.079	1.96	[0.21-18.23]	0.556	2.18	[0.24-20.07]	0.491
	>55		0.85	[0.30-2.38]	0.755	16.01	[1.53-167.57]	0.021	0.32	[0.03-3.21]	0.332	4.94	[0.39-62.40]	0.217	0.00		0.999
Sex life duration (years)	≤40		1.00			1.00			1.00			1.00			1.00		
	>40		0.74	[0.38-1.46]	0.387	1.51	[0.43-5.34]	0.519	0.86	[0.22-3.40]	0.830	7.81	[1.37-44.48]	0.021	0.60	[0.06-5.80]	0.659

Abbreviations: AOR, adjusted odds ratio; CI, confidence interval.

**Table 5.** Results of Multivariate Analysis for EGFR Mutation Occurrence Regarding Age at Menarche and Age at First Birth (Continuous Variable) in Two Multi-Adjusted Models

Biomarker	Variable	Model A			Model B		
		AOR	95%CI	P	AOR	95%CI	P
<b>EGFR mutation</b>	Age at menarche	1.231	[1.002-1.512]	0.048	1.273	[1.023-1.585]	0.031
	Age at first birth	1.094	[1.006-1.189]	0.035	1.13	[1.033-1.236]	0.008

Abbreviations: AOR, adjusted odds ratio; CI, confidence interval.

Model A: Adjusted for age at menarche, age at diagnosis, body mass index, total duration of passive smoking exposure (years), age at menopause, parity, number of pregnancies, and age at first birth. All variables are continuous.

Model B: Additionally adjusted for family history of lung cancer (number of relatives), personal history of cancer (n), total duration of exposure to cooking oil fumes (years), exposure to fumes of solid fuels for cooking/heating (in % of lifetime). All variables are continuous.

## ACKNOWLEDGEMENTS

Authors thank M. William Lebossé and Ms. Stéphanie Labonne, who performed interviews with patients; M. Quan Tran and M. Antoine Deroy (Data Manager, IFCT); Pascale Missy (PhD, IFCT), who provided administrative support; Ms. Gabrielle Cremer for expert English rewording; the French Ligue Against Cancer; and all investigator in the 75 IFCT-1002 BioCAST participating centers; the patients and their families, who greatly contributed to this work by giving their time to prepare the questionnaire and participate in the interview.

## REFERENCES

1. Couraud S, Zalcman G, Milleron B, Morin F, Souquet P-J. Lung cancer in never smokers--a review. *Eur J Cancer Oxf Engl* 1990. 2012 Jun;48(9):1299–311.
2. Couraud S, Souquet P-J, Paris C, Do P, Doubre H, Pichon E, et al. BioCAST/IFCT-1002: epidemiological and molecular features of lung cancer in never-smokers. *Eur Respir J*. 2015 May 1;45(5):1403–14.
3. Couraud S, Debieuvre D, Moreau L, Dumont P, Margery J, Quoix E, et al. No impact of passive smoke on the somatic profile of lung cancers in never-smokers. *Eur Respir J*. 2015 May;45(5):1415–25.
4. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers--a different disease. *Nat Rev Cancer*. 2007 Oct;7(10):778–90.
5. Stabile LP, Dacic S, Land SR, Lenzner DE, Dhir R, Acquafondata M, et al. Combined analysis of estrogen receptor beta-1 and progesterone receptor expression identifies lung cancer patients with poor outcome. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2011 Jan 1;17(1):154–64.
6. Nose N, Sugio K, Oyama T, Nozoe T, Uramoto H, Iwata T, et al. Association between estrogen receptor-beta expression and epidermal growth factor receptor mutation in the postoperative prognosis of adenocarcinoma of the lung. *J Clin Oncol Off J Am Soc Clin Oncol*. 2009 Jan 20;27(3):411–7.
7. Raso MG, Behrens C, Herynk MH, Liu S, Prudkin L, Ozburn NC, et al. Immunohistochemical expression of estrogen and progesterone receptors identifies a subset of NSCLCs and correlates with EGFR mutation. *Clin Cancer Res Off J Am Assoc Cancer Res*. 2009 Sep 1;15(17):5359–68.
8. Rouquette I, Lauwers-Cances V, Allera C, Brouchet L, Milia J, Nicaise Y, et al. Characteristics of lung cancer in women: importance of hormonal and growth factors. *Lung Cancer Amst Neth*. 2012 Jun;76(3):280–5.



9. Nowak F, Soria J-C, Calvo F. Tumour molecular profiling for deciding therapy-the French initiative. *Nat Rev Clin Oncol*. 2012 Aug;9(8):479–86.
10. Beau-Faller M, Blons H, Domerg C, Gajda D, Richard N, Escande F, et al. A multicenter blinded study evaluating EGFR and KRAS mutation testing methods in the clinical non-small cell lung cancer setting--IFCT/ERMETIC2 Project Part 1: Comparison of testing methods in 20 French molecular genetic National Cancer Institute platforms. *J Mol Diagn JMD*. 2014 Jan;16(1):45–55.
11. Mounawar M, Mukeria A, Le Calvez F, Hung RJ, Renard H, Cortot A, et al. Patterns of EGFR, HER2, TP53, and KRAS mutations of p14arf expression in non-small cell lung cancers in relation to smoking history. *Cancer Res*. 2007 Jun 15;67(12):5667–72.
12. Debieuvre D, Oster J-P, Riou R, Berruchon J, Levy A, Mathieu J-P, et al. The new face of non-small-cell lung cancer in men: Results of two French prospective epidemiological studies conducted 10 years apart. *Lung Cancer Amst Neth*. 2016 Jan;91:1–6.
13. Chapman AM, Sun KY, Ruestow P, Cowan DM, Madl AK. Lung cancer mutation profile of EGFR, ALK, and KRAS: Meta-analysis and comparison of never and ever smokers. *Lung Cancer Amst Neth*. 2016 Dec;102:122–34.
14. Rivera MP. Lung cancer in women: differences in epidemiology, biology, histology, and treatment outcomes. *Semin Respir Crit Care Med*. 2013 Dec;34(6):792–801.
15. Brinton LA, Gierach GL, Andaya A, Park Y, Schatzkin A, Hollenbeck AR, et al. Reproductive and hormonal factors and lung cancer risk in the NIH-AARP Diet and Health Study cohort. *Cancer Epidemiol Biomark Prev Publ Am Assoc Cancer Res Cosponsored Am Soc Prev Oncol*. 2011 May;20(5):900–11.
16. Wu C-T, Chang Y-L, Shih J-Y, Lee Y-C. The significance of estrogen receptor beta in 301 surgically treated non-small cell lung cancers. *J Thorac Cardiovasc Surg*. 2005 Oct;130(4):979–86.
17. Ali G, Donati V, Loggini B, Servadio A, Dell'Omodarme M, Prati MC, et al. Different estrogen receptor beta expression in distinct histologic subtypes of lung adenocarcinoma. *Hum Pathol*. 2008 Oct;39(10):1465–73.
18. Mazières J, Rouquette I, Lepage B, Milia J, Brouchet L, Guibert N, et al. Specificities of lung adenocarcinoma in women who have never smoked. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2013 Jul;8(7):923–9.
19. Márquez-Garbán DC, Chen H-W, Fishbein MC, Goodglick L, Pietras RJ. Estrogen receptor signaling pathways in human non-small cell lung cancer. *Steroids*. 2007 Feb;72(2):135–43.
20. He Q, Zhang M, Zhang J, Chen Y, He J, Shen J, et al. Correlation between epidermal growth factor receptor mutations and nuclear expression of female hormone receptors in non-small cell lung cancer: a meta-analysis. *J Thorac Dis*. 2015 Sep;7(9):1588–94.

21. Nose N, Uramoto H, Iwata T, Hanagiri T, Yasumoto K. Expression of estrogen receptor beta predicts a clinical response and longer progression-free survival after treatment with EGFR-TKI for adenocarcinoma of the lung. *Lung Cancer Amst Neth*. 2011 Mar;71(3):350–5.
22. Mazieres J, Barlesi F, Rouquette I, Besse B, Monnet I, Audigier-Valette C, et al. Randomized phase II trial evaluating treatment with EGFR-TKI versus EGFR-TKI associated with anti-estrogen in women with non-squamous advanced stage NSCLC: IFCT-1003 LADIE trial. *J Clin Oncol*. 2018 May 20;36(15\_suppl):9097–9097.
23. Fang M, Toher J, Morgan M, Davison J, Tannenbaum S, Claffey K. Genomic differences between estrogen receptor (ER)-positive and ER-negative human breast carcinoma identified by single nucleotide polymorphism array comparative genome hybridization analysis. *Cancer*. 2011 May 15;117(10):2024–34.
24. Cabrera SM, Bright GM, Frane JW, Blethen SL, Lee PA. Age of thelarche and menarche in contemporary US females: a cross-sectional analysis. *J Pediatr Endocrinol Metab JPEM*. 2014 Jan;27(1–2):47–51.
25. Lalys L, Pineau J-C. Age at menarche in a group of French schoolgirls. *Pediatr Int Off J Jpn Pediatr Soc*. 2014 Aug;56(4):601–4.
26. Song Y, Ma J, Agardh A, Lau PWC, Hu P, Zhang B. Secular trends in age at menarche among Chinese girls from 24 ethnic minorities, 1985 to 2010. *Glob Health Action*. 2015;8:26929.
27. Pathak PK, Tripathi N, Subramanian SV. Secular trends in menarcheal age in India-evidence from the Indian human development survey. *PLoS One*. 2014;9(11):e111027.
28. Zong X-N, Li H, Wu H-H, Zhang Y-Q. Socioeconomic development and secular trend in height in China. *Econ Hum Biol*. 2015 Dec;19:258–64.
29. Jaruratanasirikul S, Sriplung H. Secular trends of growth and pubertal maturation of school children in Southern Thailand. *Ann Hum Biol*. 2015;42(5):447–54.
30. Caronnaux M, Souquet P-J, Meert A-P, Scherpereel A, Peters M, Couraud S. Inequalities in lung cancer: a world of EGFR. *Eur Respir J*. 2016 May;47(5):1502–9.
31. Palmer RH, Vernersson E, Grabbe C, Hallberg B. Anaplastic lymphoma kinase: signalling in development and disease. *Biochem J*. 2009 May 27;420(3):345–61.
32. Roskoski R. Anaplastic lymphoma kinase (ALK): structure, oncogenic activation, and pharmacological inhibition. *Pharmacol Res*. 2013 Feb;68(1):68–94.

33. Cirello V, Fugazzola L. Novel insights into the link between fetal cell microchimerism and maternal cancers. *J Cancer Res Clin Oncol*. 2016 Aug;142(8):1697–704.
34. Kallenbach LR, Johnson KL, Bianchi DW. Fetal Cell Microchimerism and Cancer: A Nexus of Reproduction, Immunology, and Tumor Biology. *Cancer Res*. 2011 Jan 1;71(1):8–12.
35. Ait-Tahar K, Damm-Welk C, Burkhardt B, Zimmermann M, Klapper W, Reiter A, et al. Correlation of the autoantibody response to the ALK oncoantigen in pediatric anaplastic lymphoma kinase-positive anaplastic large cell lymphoma with tumor dissemination and relapse risk. *Blood*. 2010 Apr 22;115(16):3314–9.
36. Passoni L, Gallo B, Biganzoli E, Stefanoni R, Massimino M, Di Nicola M, et al. In vivo T-cell immune response against anaplastic lymphoma kinase in patients with anaplastic large cell lymphomas. *Haematologica*. 2006 Jan;91(1):48–55.
37. Guzman-Genuino RM, Diener KR. Regulatory B Cells in Pregnancy: Lessons from Autoimmunity, Graft Tolerance, and Cancer. *Front Immunol*. 2017;8:172.
38. Dagogo-Jack I, Gainor JF, Porter RL, Schultz KR, Solomon BJ, Stevens S, et al. Clinicopathologic Features of NSCLC Diagnosed During Pregnancy or the Peripartum Period in the Era of Molecular Genotyping. *J Thorac Oncol Off Publ Int Assoc Study Lung Cancer*. 2016 Sep;11(9):1522–8.
39. Derwahl M, Nicula D. Estrogen and its role in thyroid cancer. *Endocr Relat Cancer*. 2014 Oct;21(5):T273-283.
40. Vannucchi G, De Leo S, Perrino M, Rossi S, Tosi D, Cirello V, et al. Impact of estrogen and progesterone receptor expression on the clinical and molecular features of papillary thyroid cancer. *Eur J Endocrinol*. 2015 Jul;173(1):29–36.
41. Bender R, Lange S. Adjusting for multiple testing--when and how? *J Clin Epidemiol*. 2001 Apr;54(4):343–9.

## **Figures captions**

**Fig 1.** Study flow chart.

## **Additional files**

**Additional file 1.** Somatic pattern regarding age at menarche.

**Additional file 2.** Somatic pattern regarding age at menopause.

**Additional file 3.** Somatic pattern regarding sex life duration.

**Additional file 4.** Somatic pattern regarding age at first birth.

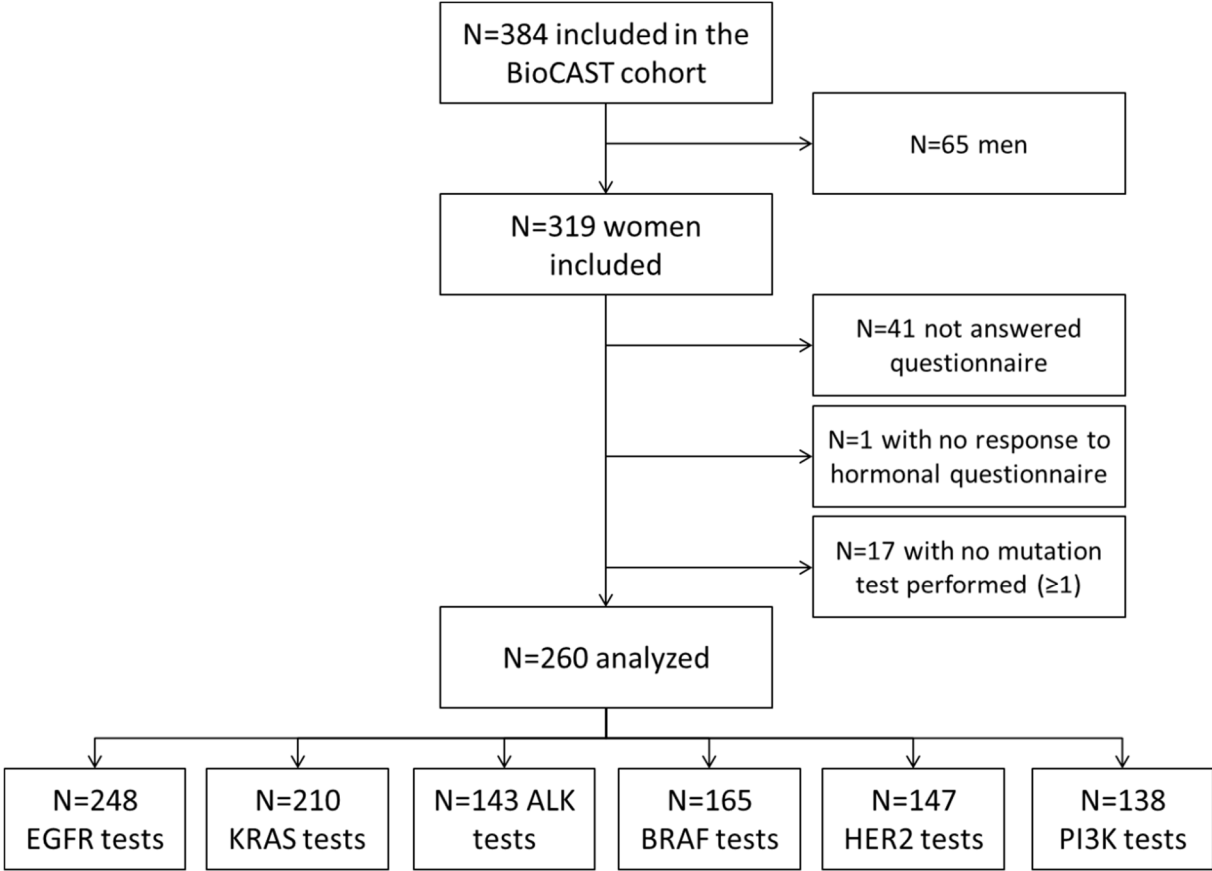
**Additional file 5.** Somatic pattern regarding number of pregnancies.

**Additional file 6.** Somatic pattern regarding parity.

**Additional file 7.** Illustration of our hypothesis proposals to explain association between age at menarche/menopause/first birth and EGFR+ lung cancers.

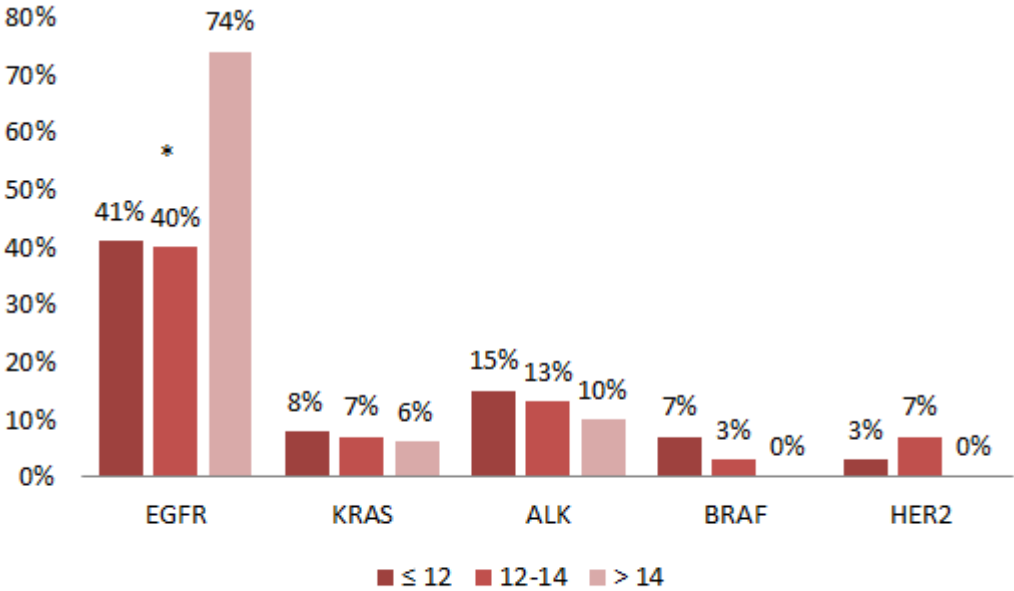
**Additional file 8.** Illustration of our hypothesis proposals to explain association between parity/pregnancy and ALK+ lung cancers.

Figure 1



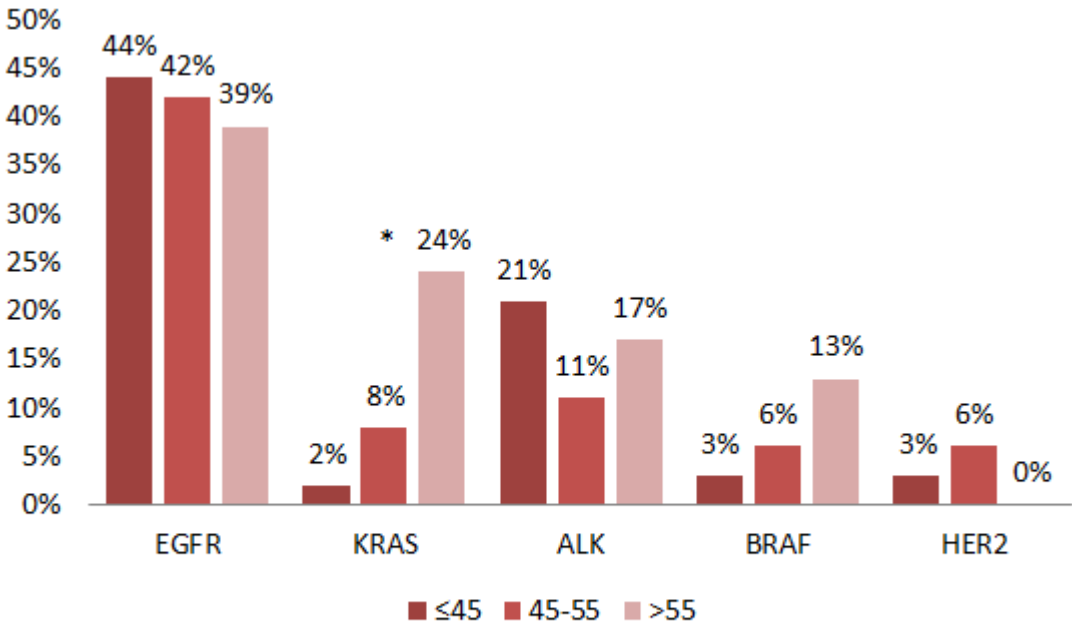
**Additional files**

**Additional file 1**



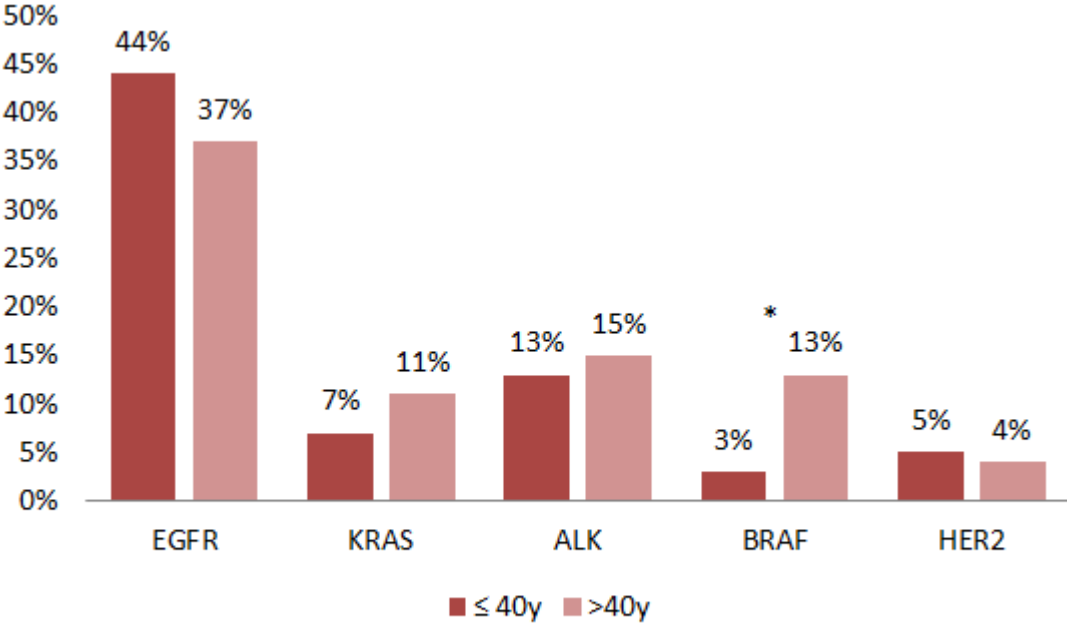
*\*indicates a significant difference (<0.05)*

**Additional file 2**



*\*indicates a significant difference (<0.05)*

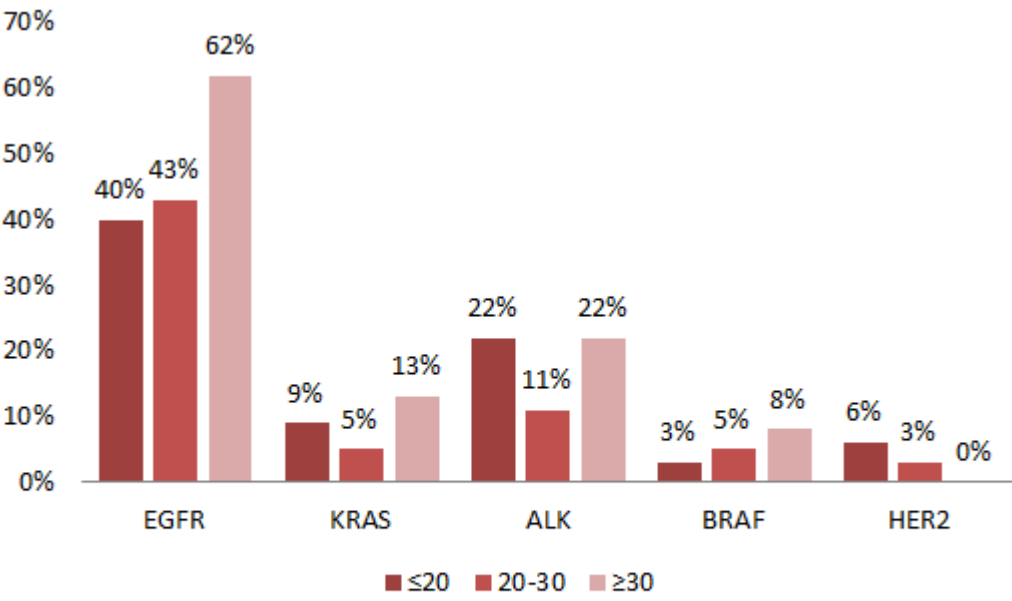
**Additional file 3**



*\*indicates a significant difference (<0.05)*

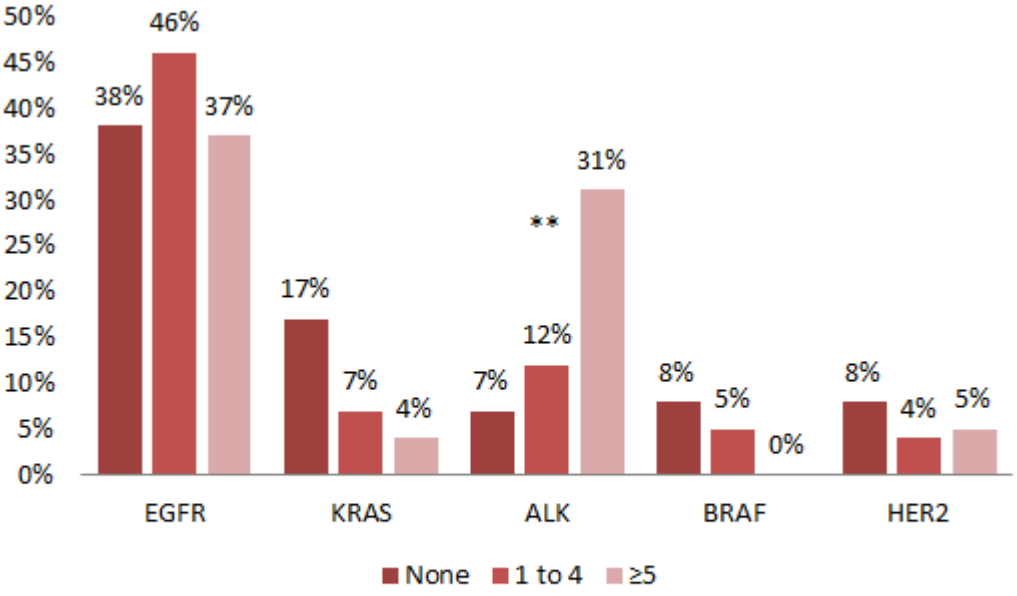


**Additional file 4**



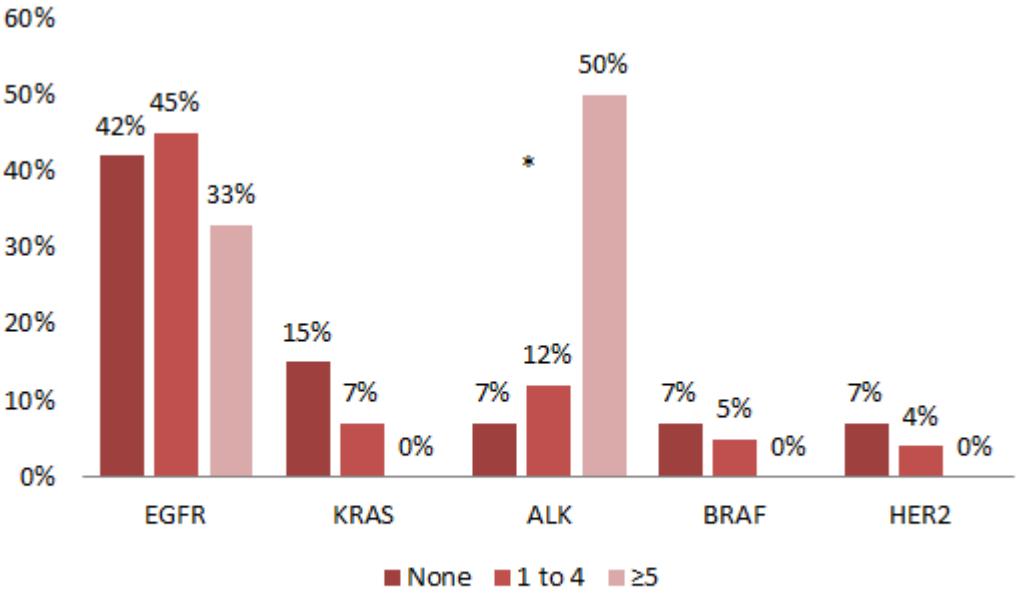
*\*indicates a significant difference (<math><0.05</math>)*

**Additional file 5**



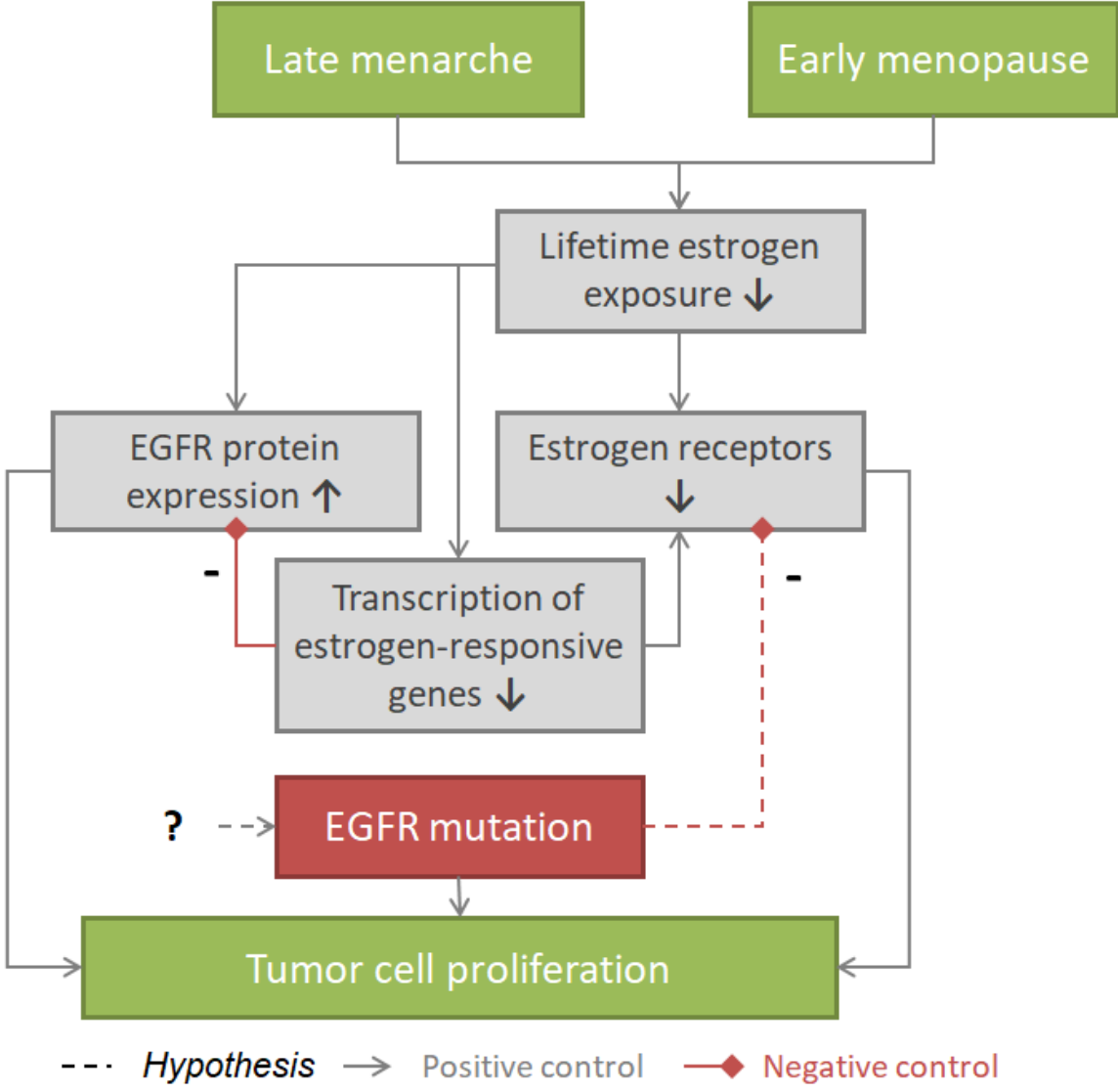
*\*\*indicates a marginally significant difference (<0.75)*

**Additional file 6**



*\*indicates a significant difference (<0.05)*

Additional file 7



**Additional file 8**

