SUPPLEMETAL FILES

Supplemental File S1 – Additional information on recorded data and exposure measurements

Supplemental Table S2 – Pathological and immune-histochemical features of the BioCAST cohort

Supplemental Table S3 – Tumor sample features and lung cancer stage

Supplemental Table S4 – Domestic pollution exposure according to gender

Supplemental Table S5 – Familial and personal medical history

Supplemental Table S6 – Reproductive factors and exposure to exogenous hormones among women

Supplemental Table S7 – Main molecular and clinicopathological features of patients harboring more than one biomarker mutation

Supplemental file S1 – Additional information on recorded data and exposure measurements

Here we aimed to provide additional information regarding data collected through patients' and physicians' questionnaires.

-Body mass index

Patients self-reported administrative height (in meters) and usual (before any unexpected weight loss) weight (in kilograms). Body mass index (BMI) was calculated as weight (kg) divided by the square of the height (m). BMI was then classified into four categories according to the WHO global database on BMI adult classification[1]: underweight (<18.50kg/m2), normal range (18.50 - 24.99kg/m2), pre-obese (25.00 - 29.99kg/m2), and obese (≥30.00kg/m2).

-Geographical origin

We asked for patients' country of birth as well as for that of their parents and grandparents. We then categorized patients and their families according to the UN's classifications of countries by major area and region of the world.[2] If the region was the same for the entire family, we categorized the patient under that region. In case of differences in the family, we considered the modal category to be correct. If there were two equal modal categories, we added the patient's region of birth to calculate a new mode, and thus the correct category. In one case only, this process failed to properly define the patient's origin, and we had to consider the major area instead of the region.

-Education level

There were seven categories: unschooled, graduated primary school, graduated secondary school, graduated high school (French *Baccalauréat*), and achieved 3, 5, or 8 (or more) years in university. The French *Certificat d'Aptitude Professionnel* and *Brevet d'Etudes Professionnelles* were considered high school graduation equivalents. Those who did not graduate from a given level were categorized in the level below.

-Passive smoke exposure

[1] Available at http://apps.who.int/bmi/index.jsp?introPage=intro_3.html, last accessed May 30, 2013.

[2] Available at http://esa.un.org/unpd/wpp/Excel-Data/country-classification.pdf, last accessed July 3, 2013

Domestic passive smoke exposure was defined as “living with at least one smoker during at least 1 year, in the same house.” We also asked about exposure before the age of 18 years old (usually considered childhood).

Passive smoking exposure at the workplace was defined as “working with at least one smoker, in the same closed room, for at least 1 year.” We retained only those who reported working with at least one smoker over the entire day (exposures during breaks or part-time days were excluded). For both domestic and workplace exposure, we asked each patient for the number of index smokers, cigarette consumption of each index smoker (in pack-years[3]), and the duration of exposure (in year).

We calculated cumulative time of exposure to passive smoking by summing years of exposure to each index-smoker. “Overall years of exposure” was calculated by summing cumulative time of exposure (years) in domestic and workplace settings. Then, we divided the results into tertiles in order to set up a categorical variable.

-Social status

There were five different categories: employed, unemployed, retired, disability, and other. Monks were classified as employed irrespective of their age. Stay-at-home spouses were classified as unemployed.

-Personal history of cancer

Two cancers occurring at the same site but separated by at least 5 years were considered two different diseases. All cancer sites, non-melanoma skin cancers included, were considered. For each cancer, we asked for the year of diagnosis and the kind of treatment undergone. Physicians were also consulted for this information. Both answers were used in order to complete the database.

-Familial history of lung cancer

Family history of lung cancer was considered only for biological first-degree relatives (spouses were excluded).

-Oral contraceptive (OC) drug and post-menopause hormone replacement therapy (HRT) intake

The questionnaire was followed with an annex sheet listing all oral contraceptive drugs and post-menopause hormone replacement therapies (HRT) approved in France (kindly provided by the French Agency for Drugs Safety, ANSM). Female patients reported the drug(s) used during their lifetimes.

[3] One pack-year was defined as smoking one standard pack of cigarettes (n=20) per day during 1 year.

We consulted the VIDAL® drug dictionary (VIDAL, Issy Les Moulineaux, France) in order to assess whether OC drugs contained ethinyl-estradiol (EE). If so, we checked if it was a continuous or a sequential dose regimen. Taking this variable into account, we calculated the total dose of EE delivered every 28 days. We then calculated the cumulative EE dose (in micrograms) by multiplying the dose delivered every 28 days by the reported duration of exposure (in years). For cumulative duration, we summed the durations of use for each patient.

We did the same with oral HRTs that contained estradiol. Cumulative estradiol doses were expressed in milligrams per 28 days. For non-oral HRTs, we only considered the transdermal route. Other local routes, such as vaginal route, were not taken into account. Finally, we calculated total cumulative estradiol intake by summing the doses from these two deliveries for each patient.

-Reproductive factors

We asked for age at menarche, at menopause, and at first live birth. We also asked for the number of pregnancies, of live births (parity), and of ovarian hormonal stimulations for pregnancy purpose. Finally, we asked about intake of certain specific drugs, such as digitalis and aromatase inhibitors.

-Cooking habits

Patients who were used to fried or stir-fried cuisine (at least one dish per week for at least 1 year) were asked to report the number of dishes prepared, as well as duration of exposure. We calculated the cooking dish-year (CDY) as follows: one cooked dish per day during 1 year = one CDY.[4] We calculated fried and stir-fried CDY separately and then summed them. The presence of a hood in the kitchen was recorded for each home. We retained only the following indicator: “has ever lived (at least one year) in a house with a hood.”

-Exposure to indoor air pollution from heating and cooking

We asked patients to declare fuels used in their homes for heating and cooking (for at least 1 year). Patients “ever exposed” were those exposed for at least 1 year of their lifetime to a domestic non-solid

[4] Yu ITS, Chiu Y-L, Au JSK, et al. Dose-response relationship between cooking fumes exposures and lung cancer among Chinese nonsmoking women. *Cancer Res*. 2006;66(9):4961–7.

fuel for heating or cooking. We distinguished solid and non-solid fuels for heating and cooking.[5] Solid (traditional) fuels were: wood, coal, garbage burning, and wood pellets. Modern (non-solid) fuels were: gas, kerosene, electricity, air, oil, water, stones, and geothermal. Homes with no heating and/or cooking system were considered as "never exposed". If heating fuels were not recorded, we looked for the heating system recorded (if any): stove, fireplace, and firebox were considered as "solid fuel exposure," while oven, boiler, radiator, central heating, electricity, and gas were not. If two energy sources or systems were recorded simultaneously (for example, wood followed by electricity if the owner upgraded the house facilities), we considered that patient to have been "ever exposed" to the listed solid fuel for the recorded duration (even if we can logically assume that a part of this duration was spent without solid fuel exposure). We also included the duration of patients' stays in given domiciles (at least 1 year) in the duration of exposure. Each patient's cumulative exposure was calculated by adding durations of stay with solid fuels for either cooking or heating. We then used the duration of life between birth and diagnosis to calculate the percent of patients' lifetime exposed to solid fuels. A cut-off value of 50% was considered relevant to identify high exposure.

-Home postal code

We recorded postal codes (with corresponding dates of residence) for further analysis on environmental atmospheric pollution and radon exposure. If the postal code was not recorded but the town name was, we searched for the postal code on the French National Addresses Service website[6]. If both the postal code and town name were missing, we reported the postal code of the main town in the department (if recorded).

-Occupational status and exposure

[5] Lissowska J, Bardin-Mikolajczak A, Fletcher T, et al. Lung cancer and indoor pollution from heating and cooking with solid fuels: the IARC international multicentre case-control study in Eastern/Central Europe and the United Kingdom. *Am. J. Epidemiol*. 2005;162(4):326–33.

[6] Available at http://www.laposte.fr/sna/rubrique.php3?id_rubrique=59, last accessed May 30, 2013.

We only considered occupations held for at least 1 year. We used the 2008 edition of the International Standard Classification of Occupations (ISCO-2008) from International Labour Organization[7] and the 2008 edition of the French classification of activities (NAF-2008) from the French National Institute of Statistics and Economic Studies.[8] Both were used at their 4th levels. In both cases, if the code was not fully recorded (*i.e.,* not to the 4th level), we used free text in comment fields to complete it. The ISCO-2008 does not acknowledge housewives, the unemployed, the disabled, or students. However, the NAF recognizes housewives under the code 9820. French "*gendarmes*" were considered as policemen, and not military personnel.

Occupational exposure to agents carcinogenic to the lungs was assessed via a specific 71-item questionnaire. Each item asked about exposure to a specific carcinogen and a specific activity. The number of years and frequency of exposure (Linkert scale from 1 to 5) were recorded. Then, a previously published algorithm was applied to the dataset in order to define a probability of exposure to each carcinogen agent for each task the patient performed.[9]

-Personal history of disease

Patient's personal history of disease was compiled from both the patients' and the physicians' questionnaires. In case of a missing value on one, we used the other to complete the record. In case of inconsistent responses, we always kept the positive answer, regardless who gave it. For example, if a patient reported having a personal history of tuberculosis and the physician reported otherwise, we considered the patient to have had a history of tuberculosis. In case of disagreement on data such as dates, we kept the patient's answers. We searched both questionnaires for: human immunodeficiency virus infection, tuberculosis, pertussis, bronchiectasis, emphysema, asthma, chronic obstructive pulmonary disease (COPD), asbestos and pleural plaques, silicosis, lung fibrosis, other chronic lung

[7] Available at http://www.ilo.org/public/english/bureau/stat/isco/isco08/index.htm, last accessed May 30, 2013.

[8] Available at http://www.insee.fr/en/methodes/default.asp?page=nomenclatures/naf2008/naf2008.htm, last accessed May 30, 2013.

[9] Bourgkard E, Wild P, Gonzalez M, et al. Comparison of exposure assessment methods in a lung cancer case-control study: performance of a lifelong task-based questionnaire for asbestos and PAHs. *Occup Environ Med.* 2013 Sep 18. doi: 10.1136/oemed-2013-101467. [Epub ahead of print].

disease, chronic arthritis, lupus or other connective or vasculitis disease, Crohn's disease or other chronic inflammatory bowel disease, solid organ transplantation, and liver cirrhosis. There were also two free text fields labelled "other personal disease."

Unfortunately, personal history of pneumonia was not systematically recorded. However, we asked both patients and physicians for "other lung diseases" and "other diseases." We were thus able to use answers to those fields to create a new "pneumonia" variable. If pneumonia (or broncho-pneumonia, but not bronchitis) was noted, we considered it (whether recorded by physician or patient). If no fields mentioned pneumonia (on either questionnaire), we considered the patient not to have a history of pneumonia. Finally, if both questionnaires provided a missing value, we reported the missing value. Lung "congestions" (n=2) and interstitial pneumonia (n=1) were considered to be pneumonia.

-**Molecular analysis**

We consulted the "Catalogue Of Somatic Mutations In Cancer" (COSMIC) database[10] for mutation definitions and used Amino-Acid code (AA) linkage for coding sequence identification (CDS). However, we did not extrapolate a CDS code from AA code alone.

If all biomarkers – *EGFR*, *KRAS*, *ALK*, *PI3KCA*, *HER2*, and *BRAF* – were not tested in a given patient, we considered that patient: (i) "mutated" if one biomarker presented with a mutation; (ii) "multiple" if we found at least one mutation in at least two different genes; (iii) "wild-type" if at least *EGFR*, *KRAS*, and *ALK* were wild-type (regardless of whether *PI3KCA*, *HER2*, and *BRAF* were missing); (iv) "missing" if at least *EGFR*, *KRAS*, or ALK were missing (regardless all other results).

-Staging:

We used the 7th edition of TNM staging of lung tumors was used[11].

-Pathological reports

Final pathological reports were collected. The 2004 WHO classification for lung tumors[12] and the IASLC / ERS / ATS classification for lung adenocarcinoma were used to categorize each tumor disease[13].

---

[10] Available at http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/, last accessed May 30, 2013.

[11] Groome PA, Bolejack V, Crowley JJ, et al. The IASLC Lung Cancer Staging Project: validation of the proposals for revision of the T, N, and M descriptors and consequent stage groupings in the forthcoming (seventh) edition of the TNM classification of malignant tumours. *J Thorac Oncol*. 2007;2(8):694–705.

[12] Travis W, Brambilla E, Mûller-Hemerlinck H. Pathology and genetics of Tumours of of the Lung, pleura, thymus and Heart. Lyon: IARC Press; 2004

[13] Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, 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(2):244–85.

Supplemental Table S2 – Pathological and immune-histochemical features of the BioCAST cohort

	Men		Women		P value	All		KBP-CPHG 2010, ref. 14
	N=65	%	N=319	%		*N*=384	%	N=5347[a]
Adenocarcinoma	**55**	**85%**	**272**	**85%**	0.893 *	**327**	**85%**	**50.6%[b]**
ADC in situ	5	9%	8	3%		13	4%	
Non-mucinous	*2*	*4%*	*7*	*3%*		*9*	*3%*	
Mucinous	*1*	*2%*	*-*	*-*		*1*	*<1%*	
Mixed	*2*	*4%*	*-*	*-*		*2*	*1%*	
In situ NOS	*-*	*-*	*1*	*<1%*		*1*	*<1%*	
Minimally-invasive ADC	1	2%	6	2%		7	2%	
Non-mucinous	*-*	*-*	*3*	*1%*		*3*	*1%*	
Mucinous	*1*	*2%*	*3*	*1%*		*4*	*1%*	
Invasive ADC	34	62%	199	73%		233	71%	
Lepidic	*2*	*4%*	*14*	*5%*		*16*	*5%*	
Acinar	*9*	*16%*	*45*	*17%*		*54*	*17%*	
Papillary	*3*	*6%*	*27*	*10%*		*30*	*9%*	
Micropapillary	*-*	*-*	*2*	*1%*		*2*	*1%*	
Solid predominant with mucin production	*6*	*11%*	*24*	*9%*		*30*	*9%*	
Invasive NOS	*14*	*26%*	*87*	*32%*		*101*	*31%*	
Variant of invasive ADC	-	-	8	3%		8	2%	
Invasive mucinous adenocarcinoma	*-*	*-*	*5*	*2%*		*5*	*2%*	
Colloid	*-*	*-*	*3*	*1%*		*3*	*1%*	
ADC NOS	15	27%	51	19%		66	20%	
Squamous cell carcinoma	**6**	**9%**	**23**	**7%**	0.574 *	**29**	**8%**	**33.2%**
Large cell carcinoma	**2**	**3%**	**15**	**5%**	0.562 *	**17**	**4%**	**12.5%**
LCC with NE component	-	-	10	67%		10	59%	
Lymphoepithelial LCC	-	-	1	7%		1	6%	
LCC NOS	2	100 %	4	27%		6	35%	
Adenosquamous carcinoma	**1**	**2%**	**4**	**1%**		**5**	**1%**	**1.5%[c]**
Sarcomatoid carcinoma	**1**	**2%**	**3**	**1%**		**4**	**1%**	**2.2%[d]**
Carcinoma NOS	**-**	**-**	**2**	**1%**		**2**	**1%**	
TTF-1 Immunohistostaining								
TTF-1 not required (missing)	3	5%	32	10%		35	9%	
TTF-1 negative	15	23%	43	14%	0.077	58	15%	
TTF-1 positive	47	72%	244	77%		291	76%	
Adenocarcinoma	44		226			270		
Large cell carcinoma	1		11			12		
Other and NOS	2		7			9		

[a] Non-small cell lung cancers only; [b] includes adenocarcinoma and bronchiolo-alveolar subtypes; [c] includes “mixed” category; [d] includes the “other” category

ADC: adenocarcinoma; NOS: not otherwise specified; LCC: large cell carcinoma; NE: neuroendocrine; TTF-1: thyroid transcription factor -1; *Computed between the corresponding histologic subtype versus other.

Supplemental Table S3 – Tumor sample features and lung cancer stage

		Lung cancer in never smokers (BioCAST)			Lung cancer in ever smokers (KBP-CPHG 2010 ref.14)
		***N*=384**	**%**	**%**	**N=6246**
Sample type	Biopsy	347	90%		
	Cytology	37	10%		
Sample origin	Primitive tumor	263	69%		
	Under bronchial endoscopy	*148*	*56%*		
	Under surgery	*46*	*18%*		
	Under CT scan	*69*	*26%*		
	Node biopsy	25	7%		
	Under mediastinoscopy or EBUS	*19*	*76%*		
	Susclavicular node	*6*	*24%*		
	Metastasis biopsy	88	23%		
	Bone metastasis	*15*	*17%*		
	Pleura	*50*	*57%*		
	Other metastasis	*23*	*26%*		
	Other or NOS	8	2%		
Stage	Stage I	35	9%	15%	*17%*
	Stage IA	*23*	*6%*		
	Stage IB	*12*	*3%*		
	Stage II	22	6%		
	Stage IIA	*8*	*2%*		
	Stage IIB	*14*	*4%*		
	Stage III	45	12%	12%	*25%*
	Stage IIIA	*30*	*8%*		
	Stage IIIB	*15*	*4%*		
	Stage IV	278	73%	73%	*59%*
	Missing	4			
Metastasis	Bones[a]	119	43%[b]		
	Lung[a]	96	35%[b]		
	Pleura[a]	95	34%[b]		
	Brain[a]	63	23%[b]		
	Liver[a]	44	16%[b]		
	Adrenals[a]	26	9%[b]		
	Other[a]	39	14%[b]		

[a] may cumulate

[b] among Stage IV only

NOS: not otherwise specified; CT: computed tomography; EBUS: endo-bronchial ultrasound

Supplemental Table S4 – Domestic pollution exposure according to gender

		Men		Women		*P*-value	All	
		N	**%**	***N***	**%**		***N***	**%**
Cooking oil exposure								
Never exposed		47	82%	164	59%	*0.001*	211	63%
Ever exposed		10	18%	113	41%		123	37%
Exposure dose (CDY)	<10	3	33%	41	40%	*NC*	44	40%
	10 to 29	3	33%	51	50%		54	49%
	≥30	3	33%	10	10%		13	12%
	Missing	1	-	11	-	-	12	-
Missing		8	-	42	-	-	50	-
Exposure to solid fuels for cooking or heating								
Never exposed		12	21%	54	20%	*0.819*	66	20%
Ever exposed		44	79%	215	80%		259	80%
Exposure dose (% lifetime)	≤50%	30	68%	162	75%	*0.323*	192	74%
	>50%	14	32%	53	25%		67	26%
Missing		9	-	50	-	-	59	-

CDY: cooking dish-year; NC: not computable; all p-values were computed using a chi-squared test

Supplemental Table S5 – Familial and personal medical history

		Men		Women		*P*-value	All	
		N	**%**	***N***	**%**		***N***	**%**
First degree relative with lung cancer	None	30	52%	104	38%	0.134	134	40%
	One	16	28%	104	38%		120	36%
	Two or more	12	21%	68	25%		80	24%
	Missing	7	-	43	-		50	-
Personal history of at least one cancer[a]		11	17%	53	17%	0.951	64	17%
Tuberculosis		3	5%	29	9%	0.234	32	8%
Pertussis		9	14%	71	22%	0.128	80	21%
Pneumonia[c]		3	5%	21	7%	0.779[b]	24	6%
Chronic bronchus disease[d]		11	17%	39	12%	0.305	50	13%

[a] all sites, non-melanoma skin included

[b] Fisher's exact test; all other are chi-squared tests;

[c] not systematically recorded, obtained from free texts;

[d] emphysema, chronic bronchitis, asthma, and bronchiectasis

Supplemental Table S6 – Reproductive factors and exposure to exogenous hormones among women

			N=319	%
Menopause reached			253	92%
	Missing		43	
Age at menopause	Median ± IQR		50 ± 7	
	Missing[a]		17	
Age at menarche	Median ± IQR		13 ± 2	
	Missing		56	
Number of live birth	Median ± IQR		2 ± 2	
	Missing[b]		11	
Age at first live birth	Median ± IQR		23 ± 5	
	Missing[¥]		43	
Oral contraceptive drugs	Never		161	58%
	Ever	All	115	42%
		≤10 years	67	65%[c]
		>10 years	36	35%[c]
		Miss	12	
	Missing		43	
Post-menopause hormone replacement therapy (oral or transdermal)	Never		205	75%
	Ever	All	70	25%
		≤10 years	46	69%[c]
		>10 years	21	31%[c]
		Miss	3	
	Missing		44	

[a] among responding women who reached menopause (n=253);

[b] nulliparous women excluded (n=31);

[c] among ever users (non-missing)

Supplemental Table S7 – Main molecular and clinicopathological features of patients harboring more than one biomarker mutation

Pt	Sex	Histo	Biomarker#1	Mutation#1	Biomarker#2	Mutation#2	Biomarker#3	Mutation#3	Untested biomarker(s)	Final diagnose (figure 2)
#1	F	LCC NE	BRAF, E15	Sub. V600E	PI3KCA, E9	NOS	PI3KCA, E20	NOS		Multiple
#2	F	SCC	KRAS, E2	Codon 12 NOS	PI3KCA, E10	Sub. E545K			HER2	Multiple
#3	F	ADC	BRAF, E15	Sub. V600E	PI3KCA, E10	Sub. E545K				Multiple
#4	F	ADC	KRAS, E2	G12V	EGFR, E20	Sub. P848L			BRAF, ALK	Multiple
#5	F	ADC	EGFR, E21	NOS	EGFR, E20	NOS				EGFR
#6	F	ADC	EGFR, E19	Del. NOS	EGFR, E18	Sub. NOS			ALK	EGFR
#7	F	ADC	EGFR, E21	L858R	EGFR, E20	Sub. T790M			HER2, ALK	EGFR
#8	F	ADC	BRAF, E15	Sub. L597L	EGFR, E19	Del. E746-A750			HER2, ALK	Multiple
#9	F	ADC	EGFR, E18	Sub. G719C	EGFR, E21	Sub. L861Q			PI3KCA, BRAF, ALK	EGFR
#10	F	ADC	EGFR, E21	Sub. L858R	ALK	2p23			PI3KCA, HER2	Multiple
#11	F	SC	EGFR, E21	Sub. L858R	EGFR, E20	Sub. S768I			HER2, PI3KCA, BRAF, ALK	EGFR

F: female; E: exon; LCC NE: large cell carcinoma with neuroendocrine component; ADC: adenocarcinoma; SCC: squamous cell carcinoma; SC: sarcomatoid carcinoma; Sub.: substitution; Del.: deletion; NOS: not otherwise specified