**Supplement Text: Description of the participant studies**

**1) CAncer de PUlmon en Asturias Study (CAPUA)** (1-5)

The CAPUA study is a hospital-based case-control study conducted in Asturias, Spain by the University of Oviedo. Lung cancer cases were recruited in three main hospitals of Asturias, following an identical protocol from 2002 to 2012. Eligible cases were incident cases of histologically confirmed lung cancer between 30 and 85 years of age and residents in the geographical area of each participating hospital. Controls were selected from patients admitted to those hospitals with diagnoses unrelated to the exposures of interest and individually matched by ethnicity, gender, age (± 5 years) and hospital. Epidemiologic data were collected personally through computer-assisted questionnaires by trained interviewers during the first hospital admission. Structured questionnaires collected information on sociodemographic characteristics, recent and prior tobacco use, environmental exposure (air pollution and passive smoking), diet, personal and family history of cancer, and occupational history from each participant. Peripheral blood samples (or mouthwash samples when they refused to donate blood) were collected from all participants. Coding of histology was based on 2001 WHO/IASLC. Genomic DNA was extracted based on standard protocol.

**2) Environment and Genetics in Lung Cancer Etiology (EAGLE)** (6-8)

The EAGLE Study is a large population-based case-control study conducted in the Lombardy region of Italy. Between April 2002 and February 2005, primary lung cancer cases (n=2098) were identified from 13 hospitals, which covered approximately 80% of incident lung cancer cases in the catchment area, and consists of 216 municipalities, including five cities (Milan, Monza, Brescia, Pavia, and Varese) and surrounding towns and villages. Inclusion criteria for both cases and controls were Italian nationality between the ages of 35 and 79 years, official residents of the municipalities, and had no severe disease that could impede participation. Case response rate was 86.6%. Approximately 95% of cases were confirmed pathologically or cytologically, and the remaining 5% were confirmed based on clinical history and imaging. Detailed histologic classification was recorded for all cases.

Controls were randomly selected from the Lombardy Regional Health Service database, which contains demographic information for virtually all Italians from the catchment area, and were frequency-matched to cases based on sex, 5-year age group, and area of residence. Family physicians for the potential controls were asked to verify the absence of lung cancer history or any advanced diseases that would impede participation. At study completion, 2120 controls were enrolled with an overall participation rate of 72.4%. The study protocol was approved by the Institutional Review Board of the US National Cancer Institute and the involved institutions in Italy. Informed consent was obtained for all subjects prior to study participation.

**3) Epidemiology & Genetics of Lung Cancer Study (EGLC), Mayo Clinic** (9,10)

Primary lung cancer patients diagnosed at Mayo clinic were identified and recruited between January 1997 and September 2008. Detailed tobacco history and clinical information were obtained through a structured questionnaire and medical record. All patients were followed by mailed questionnaires, beginning 6 months after diagnosis and then annually. Vital status was verified through the Mayo Clinic registration database, next-of-kin reports, death certificates, the Mayo Clinic Tumor Registry, and the Social Security Death Index website. A peripheral blood was collected from each consented patient. Genotyping for genome-wide association study was done with the Illumina HumanHap 370k and 610k BeadChips (Illumina, San Diegom CA, USA).

**4) Fred Hutchinson Cancer Research Center Molecular Epidemiology of Lung Cancer (Ancillary study to CARET)** (11,12)

The Carotene and Retinol Efficacy Trial (CARET) was a randomized, double-blind, placebo-controlled trial of the cancer prevention efficacy and safety of a daily combination of 30 mg of beta-carotene and 25,000 IU of retinyl palmitate in 18,314 persons at high risk for lung cancer. CARET began in 1985, and the intervention was halted in January 1996, 21 months ahead of schedule, with the twin conclusions for definitive evidence of no benefit and substantial evidence of a harmful effect of the intervention on both lung cancer incidence and total mortality. CARET continued to follow and collect endpoints on their participants through 2005. Pathology reports and medical records were reviewed to confirm cancer endpoints, and death certificates obtained to capture cause of death. During the active intervention phase of CARET, serum, plasma, whole blood, and lung tissue specimens were collected on participants. These biospecimens make up the CARET Biorepository. For the OncoArray Project, CARET provided DNA extracted from whole blood of lung cancer cases and controls matched on age at baseline (+/- 4 years), sex, race, baseline smoking status, history of occupational asbestos exposure (asbestos VS heavy smoker), and year of enrollment (2-year intervals).

**5) Harvard Lung Cancer Study (LCS)** (13)

Study subjects will derive from the existing collection of the parent cohort. The project has been successful in assembling a large, well-characterized population for molecular epidemiologic investigations of lung cancer and extended follow-up. The cohort has identified newly diagnosed lung cancer patients of all stages, administered detailed questionnaires including prospective collection of demographic, occupational, environmental, and dietary information,smoking exposure data including second hand smoking, drawn blood samples for DNA, RNA, and plasma, and banked tumor and non-involved lung tissues. To date, 11,164 participants (cases and controls) have been enrolled at Massachusetts General Hospital (MGH) and DFCI with a success rate of recruitment of 85%.

**6) Japan lung cancer study** (14,15)

Eighty-seven small cell carcinoma patients with clinical stages IA and IV tumors were subjected to treatment at the National Cancer Center Hospital in Tokyo, Japan, from 2004 to 2006. Clinical information was obtained by attending physicians and nurses. All the patients were Japanese and were diagnosed with small cell carcinoma according to the WHO classification. Written informed consent was obtained from all patients for the use of blood cells for the analysis of genetic polymorphisms in association with clinical information. This study was approved by the institutional review boards of the National Cancer Research Center. From each patient, a 20 ml whole-blood sample was obtained, and genomic DNA was extracted from whole-blood cells. Genotyping for the two SNPs was performed by TaqMan methods as previously described.

**7) Kentucky Lung Cancer Research Initiative (LCRI)**

The Kentucky Lung Cancer Research Initiative is a study conducted by the Markey Cancer Center and the University of Kentucky using a population-based, case-control framework to study the extraordinarily high rates of lung cancer in Southeastern, Appalachian Kentucky. Cancer cases were recruited from the Kentucky Cancer Registry at the time of diagnosis and controls were recruited from a random digit dialing process from the same region. Study accrual began in January 5, 2012 and completed on September 5, 2014 and 520 subjects were recruited in a 4:1 ratio of controls: cases from Appalachian Kentucky. Of the 520 subjects recruited, 231 are included in the Oncoarray analysis, including all 93 cancer cases, and 123 controls.  Newly diagnosed lung cancer cases and controls underwent blood, toenail (for trace element analysis), urine, buffy coat, water, soil, and radon collection, residence GPS mapping, as well as an extensive epidemiologic, occupational, and health history questionnaire. Clinical Trials.gov Identifier: NCT01648166.

**8) Liverpool Lung Project (LLP)**

The Liverpool Lung Project (LLP) (16) is a case-control and cohort study which has recruited over 13,000 individuals since 1996 from the Liverpool region in the UK. Detailed epidemiological and clinical data is collected with associated specimens (i.e. tumour tissue, blood, plasma, sputum, bronchial lavage and oral brushings). The participants have completed a detailed lifestyle questionnaire at recruitment, with repeat questionnaires at intervals; updated data on clinical outcome and hospital events are collected through the Health and Social Care Information Centre (including Office of National Statistics mortality data, Cancer Registry and Health Episode Statistics). The Liverpool Lung cancer risk prediction model has been developed and validated utilising this dataset (17-19), as well as the development of LLPi risk model for lung cancer incidence (20). The project is registered on the UK National Institute for Health Research (NIHR) lung cancer portfolio and has all the required ethical approvals and sponsorship arrangements in place. The lung tumours were reviewed by the reference pathologist.

**9) Toronto Lung Cancer Study** (21,22)

The lung cancer study was conducted in two recruitment periods in Greater Toronto area. The first recruitment period was between 1997 and 2002. In brief, lung cancer patients were recruited at the hospitals in the network of University of Toronto and Lunenfeld-Tanenbaum Research Institute. All subjects were interviewed based on a standard questionnaire and information was collected on lifestyle risk factors, occupational history; medical and family history. The second recruitment period was between 2008 to 2013. Lung cancer cases were recruited at the hospitals in the network of the University of Toronto. Controls were selected randomly from individuals registered in the family medicine clinics databases and were frequency matched with cases on age and sex. All subjects were interviewed, and information on lifestyle risk factors, occupational history and medical and family history was collected using a standard questionnaire. Tumors were centrally reviewed by the reference pathologist (a member of the International Association for the Study of Lung Cancer (IASLC) committee) and a second pathologist in the University Health Network. If the reviews conflicted, a consensus was arrived at after discussion. Coding of histology was based on 2001 WHO/IASLC. Genomic DNA was extracted based on standard protocol.

**10) Total Lung Cancer (TLC): Molecular Epidemiology of Lung Cancer Survival** (23)

This study is case only study of 458 lung cancer patients recruited for Moffitt Cancer Center’s Total Cancer Care™ (TCC) protocol between April 2006 and August 2010. Total Cancer Care™ is a multi-institutional observational study of cancer patients that prospectively collects self-reported demographic and clinical data, medical record information and blood samples for research purposes. All patients used in this cohort were recruited from the Thoracic Oncology Clinic at the Moffitt Cancer Center.

**References**

1. Marin MS, Lopez-Cima MF, Garcia-Castro L, Pascual T, Marron MG, Tardon A. Poly (AT) polymorphism in intron 11 of the XPC DNA repair gene enhances the risk of lung cancer. Cancer Epidemiol Biomarkers Prev **2004**;13(11 Pt 1):1788-93.

2. Lopez-Cima MF, Gonzalez-Arriaga P, Garcia-Castro L, Pascual T, Marron MG, Puente XS*, et al.* Polymorphisms in XPC, XPD, XRCC1, and XRCC3 DNA repair genes and lung cancer risk in a population of northern Spain. BMC Cancer **2007**;7:162 doi 10.1186/1471-2407-7-162.

3. Fernandez-Rubio A, Lopez-Cima MF, Gonzalez-Arriaga P, Garcia-Castro L, Pascual T, Marron MG*, et al.* The TP53 Arg72Pro polymorphism and lung cancer risk in a population of Northern Spain. Lung Cancer **2008**;61(3):309-16 doi 10.1016/j.lungcan.2008.01.017.

4. Gonzalez-Arriaga P, Pascual T, Garcia-Alvarez A, Fernandez-Somoano A, Lopez-Cima MF, Tardon A. Genetic polymorphisms in MMP 2, 9 and 3 genes modify lung cancer risk and survival. BMC Cancer **2012**;12:121 doi 10.1186/1471-2407-12-121.

5. Lopez-Cima MF, Alvarez-Avellon SM, Pascual T, Fernandez-Somoano A, Tardon A. Genetic polymorphisms in CYP1A1, GSTM1, GSTP1 and GSTT1 metabolic genes and risk of lung cancer in Asturias. BMC Cancer **2012**;12:433 doi 10.1186/1471-2407-12-433.

6. Landi MT, Consonni D, Rotunno M, Bergen AW, Goldstein AM, Lubin JH*, et al.* Environment And Genetics in Lung cancer Etiology (EAGLE) study: an integrative population-based case-control study of lung cancer. BMC Public Health **2008**;8:203.

7. Landi MT, Chatterjee N, Yu K, Goldin LR, Goldstein AM, Rotunno M*, et al.* A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. Am J Hum Genet **2009**;85(5):679-91. Erratum in: Am J Hum Genet. **2011**;88(6):861. Metsapalu, Andres [corrected to Metspalu, Andres].

8. Consonni D, Pierobon M, Gail MH, Rubagotti M, Rotunno M, Goldstein A*, et al.* Lung cancer prognosis before and after recurrence in a population-based setting. J Natl Cancer Inst **2015**;107(6):djv059 doi 10.1093/jnci/djv059.

9. Yang P, Allen MS, Aubry MC, Wampfler JA, Marks RS, Edell ES*, et al.* Clinical features of 5,628 primary lung cancer patients: experience at Mayo Clinic from 1997 to 2003. Chest **2005**;128(1):452-62 doi 10.1378/chest.128.1.452.

10. Li Y, Sheu CC, Ye Y, de Andrade M, Wang L, Chang SC*, et al.* Genetic variants and risk of lung cancer in never smokers: a genome-wide association study. Lancet Oncol **2010**;11(4):321-30 doi 10.1016/s1470-2045(10)70042-5.

11. Sakoda LC, Loomis MM, Doherty JA, Neuhouser ML, Barnett MJ, Thornquist MD*, et al.* Chromosome 15q24-25.1 variants, diet, and lung cancer susceptibility in cigarette smokers. Cancer Causes Control **2011**;22(3):449-61.

12. Sakoda LC, Loomis MM, Doherty JA, Julianto L, Barnett MJ, Neuhouser ML*, et al.* Germ line variation in nucleotide excision repair genes and lung cancer risk in smokers. Int J Mol Epidemiol Genet **2012**;3(1):1-17.

13. Zhao Y, Wei Q, Hu L, Chen F, Hu Z, Heist RS*, et al.* Polymorphisms in MicroRNAs are associated with survival in non-small cell lung cancer. Cancer Epidemiol Biomarkers Prev **2014**;23(11):2503-11 doi 10.1158/1055-9965.epi-14-0389.

14. Kohno T, Sakiyama T, Kunitoh H, Goto K, Nishiwaki Y, Saito D*, et al.* Association of polymorphisms in the MTH1 gene with small cell lung carcinoma risk. Carcinogenesis **2006**;27(12):2448-54 doi 10.1093/carcin/bgl095.

15. Shiraishi K, Kohno T, Tanai C, Goto Y, Kuchiba A, Yamamoto S*, et al.* Association of DNA repair gene polymorphisms with response to platinum-based doublet chemotherapy in patients with non-small-cell lung cancer. J Clin Oncol **2010**;28(33):4945-52.

16. Field JK, Smith DL, Duffy S, Cassidy A. The Liverpool Lung Project research protocol. Int J Oncol **2005**;27(6):1633-45.

17. Cassidy A, Myles JP, van Tongeren M, Page RD, Liloglou T, Duffy SW*, et al.* The LLP risk model: an individual risk prediction model for lung cancer. Br J Cancer **2008**;98(2):270-6 doi 10.1038/sj.bjc.6604158.

18. Raji OY, Duffy SW, Agbaje OF, Baker SG, Christiani DC, Cassidy A*, et al.* Predictive accuracy of the Liverpool Lung Project risk model for stratifying patients for computed tomography screening for lung cancer: a case-control and cohort validation study. Ann Intern Med **2012**;157(4):242-50 doi 10.7326/0003-4819-157-4-201208210-00004.

19. Chen Y, Marcus MW, Niaz A, Duffy SW, Field JK. MyLungRisk: a user-friendly, web-based calculator for risk assessment of lung cancer based on the validated Liverpool Lung Project risk prediction model. International Journal of Health Promotion and Education **2014**;52(3):144-52 doi 10.1080/14635240.2014.888814.

20. Marcus MW, Chen Y, Raji OY, Duffy SW, Field JK. LLPi: Liverpool Lung Project Risk Prediction Model for Lung Cancer Incidence. Cancer Prev Res (Phila) **2015**;8(6):570-5 doi 10.1158/1940-6207.capr-14-0438.

21. Wang Y, McKay JD, Rafnar T, Wang Z, Timofeeva MN, Broderick P*, et al.* Rare variants of large effect in BRCA2 and CHEK2 affect risk of lung cancer. Nat Genet **2014**;46(7):736-41 doi 10.1038/ng.3002.

22. Zhang LR, Morgenstern H, Greenland S, Chang SC, Lazarus P, Teare MD*, et al.* Cannabis smoking and lung cancer risk: Pooled analysis in the International Lung Cancer Consortium. Int J Cancer **2015**;136(4):894-903 doi 10.1002/ijc.29036.

23. Woods NT, Monteiro AN, Thompson ZJ, Amankwah EK, Naas N, Haura EB*, et al.* Interleukin polymorphisms associated with overall survival, disease-free survival, and recurrence in non-small cell lung cancer patients. Mol Carcinog **2015**;54 Suppl 1:E172-84 doi 10.1002/mc.22275.