ERJ Express. Published on July 16, 2009 as doi: 10.1183/09031936.00120808

GENETIC ASSOCIATION STUDY OF *CYP1A1* POLYMORPHISMS IDENTIFIES RISK HAPLOTYPES in NON-SMALL CELL LUNG CANCER

Casey Maree Wright[1,2*], Jill Everland Larsen[1,2*], Maree Louise Colosimo[1], Jeremy John Barr[1], Lu Chen[1], Rebecca Elizabeth McLachlan[1], Ian Anthony Yang[1,2], Rayleen Veronica Bowman[1,2], Kwun Meng Fong[1].

[1]Department of Thoracic Medicine, The Prince Charles Hospital, Brisbane, Australia
[2]School of Medicine, The University of Queensland, Brisbane, Australia
*The first two authors have contributed equally to the work.

Corresponding author:
Casey Maree Wright
Room 2, Level 1 Clinical Sciences Building,
The Prince Charles Hospital,
Rode Road, Chermside, QLD, 4032.
Email: c.wright@uq.edu.au
Telephone: +61 7 3139 4110
Fax Number: +61 7 3139 4957

Short Title: *CYP1A1* risk haplotypes identified for NSCLC.

Abbreviations:
NSCLC – Non-small cell lung cancer; PAH – Polyaromatic hydrocarbon; CYP1A1 – Cytochrome P450 family 1, subfamily A, polypeptide 1; OR – Odds Ratio; UTR – untranslated region; TPCH – The Prince Charles Hospital; PCR – Polymerase Chain Reaction

Sources of Funding:
This work was supported by The Prince Charles Hospital Foundation, Queensland Smart State Fellowship (KMF), National Health and Medical Research Council (NHMRC) Practitioner Fellowship (KMF), Queensland Cancer Fund Clinical Research Fellowship (KMF), and NHMRC Career Development Award (IAY).

Abstract
Lung cancer remains a leading cause of disease globally with smoking being the largest single cause. Phase I enzymes including cytochrome P450 family 1, subfamily A, polypeptide 1 (CYP1A1), are involved in the activation of carcinogens such as polycyclic aromatic hydrocarbons (PAHs) to reactive intermediates, which are capable of binding covalently to DNA to form DNA adducts, potentially initiating the carcinogenic process. The present study aimed to investigate the association of *CYP1A1* polymorphisms and haplotypes with lung cancer risk.
We performed a case-control study of 1040 non-small cell lung cancer (NSCLC) cases and 784 controls on three variants, *CYP1A1*2A* (rs4646903 3801T>C), *CYP1A1*2C* (rs1048943 2455A>G Ile462Val (Exon 7)), and *CYP1A1*4* (rs1799814 2453C>A Thr461Asn (Exon 7)) using PCR-RFLP methods.
*CYP1A1*2A* and *CYP1A1*2C* variants were significantly over-represented in NSCLC cases compared with controls whilst the *CYP1A1*4* variant was under-represented. *CYP1A1* haplotypes (allele order: *CYP1A1*4-CYP1A1*2C-CYP1A1*2A*) *C-G-C* and *C-G-T* were associated with an increased risk of lung cancer, while *A-A-T* was associated with decreased lung cancer risk in this population.
This study identifies risk haplotypes for *CYP1A1* in NSCLC and confirms that *CYP1A1* polymorphisms are a minor risk factor for NSCLC.
Key Words: Lung cancer, NSCLC, *CYP1A1*, cytochrome P450, haplotype.

Introduction

Lung cancer is a leading cause of cancer death internationally, with smoking being the largest single cause. Smoking is responsible for 85-90% of lung cancers [1] yet fewer than 20% of life-long smokers will develop lung cancer suggesting that other factors including genetics may play a role [1]. Phase I enzymes (mainly cytochrome P450) metabolically activate carcinogens such as polycyclic aromatic hydrocarbons (PAHs) and N-nitrosamines to reactive intermediates [2]. These intermediates are capable of binding covalently to DNA to form DNA adducts, potentially initiating the carcinogenic process. Two functionally important non-synonymous polymorphisms have been described for the *CYP1A1* gene, a base substitution at codon 462 in exon 7, resulting in an isoleucine to valine amino acid substitution (*CYP1A1*2C*; rs1048943 2455A>G Ile462Val (Exon 7)) and a point mutation (thymine/cytosine) at the *Msp1* site in the 3'UTR (*CYP1A1*2A*; rs4646903 3801T>C) [3, 4]. A third polymorphism substituting threonine for asparagine (*CYP1A1*4*; rs1799814 2453C>A Thr461Asn (Exon 7)) has also been reported two bases upstream of *CYP1A1*2C*, however its functional effects have yet to be fully elucidated.

Ethnic differences have been reported in the distribution of *CYP1A1*2C* and *CYPA1*2A* genotypes in lung cancer subjects with few reports available for *CYP1A1*4*. An over-representation of the valine allele (*CYP1A1*2C*) among lung cancer cases has been reported in Asian and Caucasian populations [5-7]. While relatively frequent in Asian populations (0.18-0.25) [8, 9] the valine allele is quite rare in Caucasian control populations occurring in approximately 7-13% of people [7, 10]. In a previous study, we did not show any interactive effects between *CYP1A1*2C* and polymorphisms of the Phase II enzymes *GSTP1, GSTT1* or *GSTM1*. However, risk genotypes for *MPO* and *CYP1A1*2C* interacted to increase the overall risk of NSCLC (OR 2.88; 95%CI 1.70–5.00; *P*=0.001) [11]. Similarly, the *CYP1A1*2A* variant has been strongly associated with increased lung cancer risk in Asian populations especially in relation to tobacco smoking and in combination with polymorphisms of Phase II enzymes *GSTM1* and *GSTT1* [12, 13]. In contrast, studies in Caucasian populations have not clearly established an association between *CYP1A1* polymorphisms and increased lung cancer risk [14-19]. Relatively frequent incidences of *CYP1A1*2A* variants have been reported in Asian (39%) [20] and Caucasian control populations (18-23%) [14, 16, 21]. Conversely, few studies report population frequencies for *CYP1A1*4* in lung cancer subjects. Cascorbi et. al [6] has reported variant allele frequencies of 3% in Caucasian lung cancer cases and controls. In contrast Song et. al, detected no polymorphic sites in an Asian population [13] supporting similar polymorphism studies of Japanese liver subjects [10]. Similar frequencies (2-3%) have been observed in healthy Caucasian populations [22, 23]. A recent pooled analysis by Hung et al. [24] reported an increase in lung cancer susceptibility by more than two-fold for *CYP1A1*2A* and *CYP1A1*2C* in non-smokers. The authors suggested these polymorphisms may be implicated in lung carcinogenesis at low levels of tobacco exposure possibly in combination with phase II enzymes including GSTM1 [24], although this has not been replicated by other groups [21].

Many studies investigating the relationships between *CYP1A1* polymorphisms and lung cancer have been limited by small sample numbers leading to a lack of statistical power. Pooled analyses to increase sample size have tried to address this issue. Conflicting results between groups may be due to population differences (i.e. ethnicity) or failure to control for other potential confounders including age and gender. C*YP1A1* haplotype studies have potential to determine whether combinations of *CYP1A1* (*CYP1A1*2A*, *CYP1A1*2C*, *CYP1A1*4*) together confer a greater risk of lung cancer than single polymorphisms. One recent study of 200 case-matched controls from an Indian population, showed that only one haplotype (*C-G-C*) was significantly associated with increased lung cancer risk, although the

study was limited by its small sample numbers [3]. Few haplotype association studies for *CYP1A1* have been conducted in a Caucasian population. A study by Han et al [25] investigated the genotype frequencies of 13 SNPs found in the promoter region of CYP1A1 in 21 Caucasian individuals. Subsequent functional studies identified two CYP1A1 haplotypes (2923C<T-2875G<A-3777T<G and 2923C<T-3777T<G-4553G<A) demonstrating moderate increases in basal activity compared with the wildtype CYP1A1 constructs (1.38 and 1.50 respectively $P<0.05$). These were considered unlikely to be of functional significance considering the magnitude of differed CYP1A1 expression in response to benzo[a]pyrene and cigarette smoke extract [25].

To further investigate the role of *CYP1A1* polymorphism variation (*CYP1A1*2C*, *CYP1A1*2A* and *CYP1A1*4*) in lung cancer risk we performed *CYP1A1* haplotype analyses in a large sample of Australian lung cancer cases aiming to identify risk modifying *CYP1A1* haplotypes.

Materials and Methods

Study population

This study population has been described previously [11, 26]. Cases were subjects with confirmed primary lung cancer treated at The Prince Charles Hospital between 1980 and 2007 (n = 1040). Controls consisted of subjects with chronic obstructive pulmonary disease (COPD) treated at the same hospital (n=506) or healthy smokers (n = 278) who, at the time of recruitment to this study, did not have a doctor-diagnosis of lung cancer. The ethnicity of the study population was >99% Caucasian. The study was approved by the Ethics Committee at The Prince Charles Hospital. Cases and controls gave informed written consent for use of resected lung tissue or peripheral blood. Demographics for cases and controls were checked by a research nurse or the treating physician against patient medical charts or the hospital lung cancer database (Table 1).

Sample Preparation and Genotyping

DNA from cases diagnosed with NSCLC was extracted from peripheral blood or resected cryopreserved normal lung tissue as described previously [26]. Cases and controls from our previous study of *CYP1A1*2C* were included [11, 26]. DNA from control subjects was extracted from peripheral blood. In 592 cases, DNA was extracted from more than one source (blood-lymphocyte and fresh-frozen normal lung tissue). In these cases both sources were genotyped with identical results in all cases, reinforcing the reproducibility of our methods. PCR-based restriction fragment length polymorphism (RFLP) methods were used to analyse *CYP1A1*2A* [27] and *CYP1A1*4* [6] polymorphisms. Approximately 10% of samples were randomly selected for repeat genotyping by PCR-RFLP to test reproducibility.

DNA sequencing and single base extension genotyping

To confirm accuracy of PCR-RFLP methods, each polymorphism had two representative samples per genotype confirmed by DNA sequencing. Samples were purified using the Wizard PCR Cleanup and Gel extraction kit (Promega, Madison, USA) and sequenced at the Australian Genome Research Facility (AGRF, Brisbane, Australia) for DNA sequencing using BigDye Terminator V3.1 chemistry (Applied Biosystems, California, USA). Sequencing primers were identical to those used in PCR amplification. Sequences were visualised using Chromas V1.4 (Conor McCarthy, School of Biomolecular and Physical Sciences, Griffith University, Brisbane, Australia).

In addition, we also used an independent genotyping method, iPLEX single base extension, to genotype a subset (89%) of subjects genotyped by PCR-RFLP methods. All experiments were performed by the AGRF. Briefly genomic DNA was amplified using primers with 10-mer tags, designed to amplify a 75-150bp amplicon. Following this, shrimp alkaline phosphatase (SAP) was added to each reaction to remove any unincorporated dNTPs by cleavage of phosphates from dNTP groups. An iPLEX reaction master mix consisting of primer, enzyme, buffer and mass-modified nucleotides was then added and samples placed in a thermocycler to allow addition of nucleotides to the polymorphic site, producing allele specific base extension products of differing sizes. Products were then run on a MALDI-TOF mass spectrometer to determine product size.

Statistical Analysis

Distributions of genotypes and demographic variables were compared between cases and controls using chi-square tests for categorical outcome variables and two-sided *t*-tests for continuous outcome variables. Odds ratios (OR) and 95% confidence intervals (CI) were estimated to measure the association between lung cancer and genotype/haplotype frequency. Standard chi-square statistics were used to determine whether the three *CYP1A1* variants were in Hardy-Weinberg equilibrium. All statistical analyses for genotypes were performed

using the SPSS software package (Version 13.0 for Windows: SPSS Inc., Chicago, IL, USA). Haplotype analyses (haplotype frequency estimations and linkage disequilibrium) were carried out using Haploview Linkage software (Version 4.0, http://www.broad.mit.edu/edu/mpg/haploview [28]). A *P* value less than 0.05 (two-tailed) was considered statistically significant. Power calculations to evaluate our ability to detect associations between NSCLC and *CYP1A1* variants among our study population were determined using $\alpha = 0.05$ and 80% power. This study had the power to detect an odds ratio of 1.39, 1.48 and 1.69 for the variant alleles of *CYP1A1*2A*, *CYP1A1*4* and *CYP1A1*2C* respectively with 80% confidence at $\alpha = 0.05$.

Results:

Participant characteristics

To ensure that observed effects between cases and controls were due to genotype frequency and not other potential confounding factors such as age, smoking history or gender, we compared mean and frequency distributions between cases and controls (Table 1). Mean age differed significantly between cases and controls ($P<0.001$) with controls being slightly younger. In contrast, there were no significant differences in gender distribution ($P=0.81$) or pack-years smoked ($P=0.47$), excluding these factors as potential study confounders. The majority of NSCLC cases were adenocarcinoma or squamous cell carcinoma histological subtypes (39% and 36%, respectively).

To confirm the validity of combining COPD subjects and healthy smokers to form one control group as we have done previously [26], we used chi-square tests to investigate whether there were significant differences in genotype frequency between the two groups. We observed no significant difference in variant frequency between COPD and healthy smokers ($P>0.05$, data not shown) validating our decision to combine these two subgroups.

PCR-RFLP

Genotype frequencies for all three variants, determined by PCR-RFLP methods, are detailed in Table 3. Variant allele frequencies were low in the control population for all SNPs. The minor allele frequencies for combined cases and controls were 4.6%, 4.5% and 11.4% for *CYP1A1*4, CYP1A1*2C* and *CYP1A1*2A* respectively. As less than five samples were classified as homozygous for the variant allele in either cases or controls for *CYP1A1*2C CYP1A1*4*, homozygous variants were combined with heterozygous genotypes for statistical analysis (Table 4). In this study, genotypes containing the variant allele of *CYP1A1*2A* occurred in 17.6% of the control population, *CYP1A1*4* in 11% and *CYP1A1*2C* in 5.5%. These values were in relative agreement with previous studies of Caucasian populations (*CYP1A1*2A* 18 – 23%; *CYP1A1*4* 2 – 3%, *CYP1A1*2C* 9 – 10%) [14, 16, 21-23]. In this study population, genotype frequencies of *CYP1A1*2A* and *CYP1A1*4* were in Hardy-Weinberg Equilibrium for cases and controls, however *CYP1A1*2C* was not (Table 2). DNA Sanger sequencing of two samples per genotype were performed for each polymorphism to confirm the accuracy of PCR-RFLP assays.

From the PCR-RFLP data, we observed that subjects carrying the *CYP1A1*2A* variant were overrepresented in NSCLC cases compared with controls (*TT* versus *CT/CC*: OR 1.43, 95%CI 1.35-1.51, $P=0.003$). We also observed a decreased risk of NSCLC for those with the *CYP1A1*4* variant (*CC* versus *CA/AA*; OR 0.64 95% CI 0.64-0.62; $P=0.005$). The wildtype genotype (*CC*) was slightly over-represented in cases versus controls (92.8% versus 89%). Data from our previous study on *CYP1A1*2C* was included in this study. The variant allele was associated with a greater than two-fold increase in the risk of lung cancer.

To investigate associations between the three polymorphisms, haplotype analyses were performed to determine whether combined polymorphisms conferred a greater lung cancer risk. Five possible haplotypes were identified amongst cases and the larger control group: where the haplotype *C-A-T* was considered to be the wildtype (haplotype order = *CYP1A1*4-*

*CYP1A1*2C-CYP1A1*2A*). The frequencies of *C-A-T, C-A-C, A-A-T, C-G-C* and *C-G-T* are presented in Table 4. Two haplotypes were significantly over-represented in NSCLC versus controls, *C-G-C* (4.1% versus 2.4%, *P*=0.0038; OR 1.76, 95% CI 1.19-2.60) and *C-G-T* (1.4% versus 0.4%, *P*=0.0026; OR 3.26, 95% CI 1.46-7.28). Conversely, *A-A-T* was significantly under-represented in cases versus controls (3.4% versus 5.4%, *P*=0.0014; OR 0.62, 95% CI 0.45-0.86). *C-A-C* showed a general increase in lung cancer risk, though this did not reach statistical significance (8.5% versus 6.9%, *P*=0.0747; OR 1.25, 95% CI 0.98-1.60). We found *CYP1A1*2A* and *CYP1A1*2C* were in linkage disequilibrium (D'=0.749; LOD = 66.5; r^2 =0.207; 95% CI = 0.67-0.81) but not *CYP1A1*4* and *CYP1A1*2C* (D'=0.01; LOD=0; r^2=0; 95% CI = 0.01-0.696) or *CYP1A1*4* and *CYP1A1*2A* (D'=0.582; LOD = 1.21; r^2 = 0.0020; 95% CI =0.17-0.82).

Single base extension genotyping

As a secondary method of confirming PCR accuracy and to address the issue of Hardy-Weinberg disequilibrium, we used single base extension genotyping methods (iPLEX, AGRF, Brisbane, Australia) to genotype cases and controls. We were able to successfully re-genotype 89% of cases and controls with iPLEX, with failure rates of 3-5% for the three genotypes. Tables 3 and 4 illustrate the sample numbers, genotype frequencies and haplotype frequencies observed using iPLEX. All variants were in Hardy-Weinberg equilibrium for our control population (Table 2) with two of three variants in Hardy-Weinberg equilibrium for cases (Table 2).

We observed similar genotype frequencies for *CYP1A1*2C, CYP1A1*4* and *CYP1A1*2A*, although relatively fewer variant homozygotes were observed for *CYP1A1*4* (Controls - 5 for PCR-RFLP vs 0 for iPLEX). Observed associations between *CYP1A1* polymorphisms and lung cancer risk remained consistent, with subjects carrying the *CYP1A1*2A* variant overrepresented in NSCLC cases compared to controls (*TT* versus CT/CC: OR 1.43, 95% CI 1.35-1.52, *P*=0.005), and carriers of the *CYP1A1*4* variant under-represented in NSCLC cases (*CC* versus *CA/AA*; OR 0.66, 95% CI 0.67-0.66; *P*=0.026). Carriers of the *CYP1A1*2C* variant were also associated with elevated lung cancer risk (*AA* versus *AG/GG*: OR 1.83; 95% CI 1.57-2.12, *P*=0.002), consistent with our previous findings. While our findings from PCR-RFLP remain consistent, we observed a decrease in statistical significance possibly due to the small decrease in sample number genotyped by iPLEX.

To ensure that our haplotype findings from PCR-RFLP still remained consistent, we also performed new haplotype analyses on the iPLEX data. Only samples with iPLEX genotype data for all three polymorphisms were included in these analyses. We observed four possible haplotypes amongst cases and controls (haplotype order = *CYP1A1*4-CYP1A1*2C-CYP1A1*2A*; *C-A–T, C-A-C, C-G-C, A-A-T*; Table 4) where *C-A-T* was considered to be the wildtype. We did not observe any of the *C-G-T* haplotype in this group. Two of these haplotypes were significantly associated with either increased (*C-G-C*; 5.3% versus 3.0%, *P*=0.0023) or decreased lung cancer risk (*A-A-T*, 3.4% versus 4.8%, *P*=0.037) with *C-A-C* showing a trend towards increased lung cancer risk (8.5% versus 6.7%, *P*=0.063). These risk haplotypes were also identified in our PCR-RFLP haplotype analyses. *CYP1A1*2A* and *CYP1A1*2C* remained in linkage disequilibrium (D'=1.0; LOD = 98.17; r^2 =0.324; 95% CI = 0.97-1.00) with *CYP1A1*4* and *CYP1A1*2C* (D'=0.991; LOD=0.67; r^2=0.002; 95% CI = 0.11-0.98) and *CYP1A1*4* and *CYP1A1*2A* (D'=1.0; LOD = 2.76; r^2 = 0.006; 95% CI =0..48-1.00) not exhibiting disequilibrium, as observed in our PCR-RFLP findings.

Discussion

We have previously shown that carriers with the valine allele *CYP1A1*2C* (Ile/Val or Val/Val genotypes) are significantly overrepresented in NSCLC compared to controls (OR 1.9; 95% CI 1.20–2.90; *P*=0.005), especially in women, those aged <64 years and those with

<46 pack years of tobacco exposure [26]. We have previously shown that the *CYP1A1*2C* variant in combination with the *MPO* risk allele confers a significantly increased risk of NSCLC (OR 2.88; 95% CI 1.70–5.00; *P*<0.0001) [11]. In this report, we present the genotype and haplotype frequencies of two additional *CYP1A1* polymorphisms in relation to lung cancer risk. Our data indicates an association between *CYP1A1*2A* and lung cancer risk and supports a recent pooled analysis of 2451 lung cancers and 3358 controls which showed a clear association between the homozygous *CYP1A1*2A* allele (CC) and lung cancer risk in Caucasians (age- and gender-adjusted OR 2.36; 95% CI 1.16-4.81) [15]. Although a meta-analysis by Houlston et al., provided little support for variations of *CYP1A1* involvement in lung cancer risk [29], a recent review investigating the role of polymorphisms in candidate genes for 18 different cancer sites, reported an increased risk of lung cancer for carriers of the *CYP1A1*2A* variant in Caucasian populations (OR 2.36; 95% 1.16–4.81; *P*=0.018) in addition to variants of *CYP1A1*2C* in Asian populations (OR 1.61; 95% CI 1.24–2.08; *P*=0.0003) [30]. Associations between *CYP1A1* variations and lung cancer risk have also been observed in never-smokers with lung cancer [24, 31].

*CYP1A1*2A* and *CYP1A1*2C* have been reported to be in linkage disequilibrium in Caucasian [18] and Asian populations [27] although not in people of African descent [32]. Linkage disequilibrium can be influenced by a variety of factors including genetic linkage, recombination, mutation rates, random drift, non-random mating and population structure. In this study, *CYP1A1*2A* and *CYP1A1*2C* were in linkage disequilibrium confirming reports from previous studies [18]. Conversely, like Cascorbi et. al, we did not observe evidence of linkage disequilibrium between *CYP1A1*2C* and *CYP1A1*4* [6]. This may be due to the close proximity of the two variants (only one base pair separates the two polymorphic sites) decreasing the rate of recombination.

The function of *CYP1A1*4* has not been clearly established, although it has been suggested that it has the greatest enzymatic efficiency amongst all of the *CYP1A1* polymorphisms [33]. In the present study, we identified very few homozygous variants in cases or controls for *CYP1A1*4*, limiting our ability to draw strong conclusions for a Caucasian population. However, we did observe a general decrease in frequency of heterozygotes in this study population compared to controls. Others have reported no clear association between lung cancer risk and *CYP1A1*4* [3].

Very few haplotype analyses studying the interactions between *CYP1A1*2A*, *CYP1A1*2C* and *CYP1A1*4* have been performed in a population of Caucasian lung cancer subjects. A recent pooled analysis studying interactions between *CYP1A1* polymorphisms and *GSTM1/GSTT1* in an Asian population, identified *G-T* and *A-C* haplotypes as being associated with lung cancer risk compared to *A-T* (OR 3.41; 95% CI 1.78-6.53 and OR 1.39; 95% CI 1.12-1.71, respectively) [34]. Among our population, the haplotypes (allele order: *CYP1A1*4-CYP1A1*2C-CYP1A1*2A*) *C-G-C* and *C-G-T* were associated with an increased risk of lung cancer while *A-A-T* was associated with a decreased risk. In an Indian population, a four-fold increased lung cancer risk (HR 3.90, 95% CI 1.00–5.10, *P*=0.025) for the *C-G-C* haplotype [3] was shown suggesting that this haplotype may be an important risk factor for NSCLC. Although the *C-A-C* haplotype has been associated with higher enzymatic activity and decreased cancer risk in prostate cancer [35], Shah et al [3], showed a trend for increased lung cancer risk although this did not reach statistical significance. The differences observed may be due to tumour type, gene expression or functional role of CYP1A1. For example, in addition to detoxification of carcinogens, CYP1A1 also has a role in the oxidative metabolism of estrogens [36] which have been implicated in the aetiology of prostate cancer [37, 38]. Chang et al, suggested that in prostate cancer, CYP1A1's role of oestrogen metabolism may be more important than carcinogen detoxification [35]. Therefore, it is

possible that the role of CYP1A1 differs between cancer types resulting in haplotype associations that vary according to tumour origin.

Interpretation of our data from this study are curtailed by the low frequency of some variant genotypes (for instance, we observed very few *A/A* genotypes for *CYP1A1*4*) limiting our ability to draw statistically valid conclusions. Subjects in our control group also have the potential to develop lung cancer in the future: a source of misclassification bias. One of the polymorphisms, *CYP1A1*2C* was not in Hardy-Weinberg equilibrium (HWE) in either controls or cases for PCR-RFLP but was in HWE for iPLEX. In addition, while *CYP1A1*2A* was in HWE using PCR-RFLP, it was out of HWE via iPLEX for NSCLC subjects. This may be due to chance (from the selection of a large subset of the original samples), decreased sensitivity from iPLEX due to differences in sample concentration (Sequenom – 10 ng/ul; PCR-RFLP – 10-40 ng/ul as determined by gel estimation techniques), non-random population (highly unlikely), or a possible association with NSCLC risk. In addition, *CYP1A1*4* and *CYP1A1*2C* are separated by only one nucleotide on exon 7 raising the possibility that misclassifications could occur. In this study, we tried to overcome this by employing published assays using enzymes that selectively digest each polymorphism irrespective of the neighbouring polymorphism, thereby decreasing the risk of misclassification bias. Finally, this study evaluated three commonly studied *CYP1A1* variants, however, it is possible that other SNPs may be in linkage disequilibrium with these variants, and may also be instrumental in determining lung cancer risk.

Technical validation of our PCR-RFLP method by iPLEX confirmed our observed associations for all *CYP1A1* polymorphisms. We observed similar genotype frequencies for *CYP1A1*2C, CYP1A1*4* and *CYP1A1*2A* in both case and control populations. Haplotype analyses also confirmed association of haplotypes *A-A-T* and *C-G-C* with decreased and increased lung cancer risk respectively. We did not observe any of the *C-G-T* haplotype in these analyses and could not confirm association with lung cancer risk. Despite the significance of these results decreasing in the iPLEX cohort, we were still able to confirm the observations obtained using PCR-RFLP, increasing our confidence in the reproducibility of genotyping techniques. Due to our high reproducibility rate, we are confident that these associations are valid and support a role for *CYP1A1* polymorphisms in altering lung cancer risk.

In conclusion, the results of the present study confirm that *CYP1A1* polymorphisms are a minor risk factor for NSCLC. While several studies have confirmed associations between these polymorphisms and lung cancer risk in the past, few have explored linkage effects between the three different *CYP1A1* polymorphisms in a Caucasian population. To the authors' knowledge, this is the first study to show that haplotypes *C-G-C*, *C-G-T* and *A-A-T* are strongly associated with lung cancer risk in Caucasians. Identification of these haplotypes may assist in risk stratification, early detection and improvement of current treatment options for subjects with lung cancer. Larger studies are required to explore these risk *CYP1A1* haplotypes further.

Acknowledgements:

The authors wish to thank the physicians, surgeons and pathologists at The Prince Charles Hospital who assisted with the project; the subjects and donors who participated in the study; Linda H Passmore for recruiting subjects and confirming clinical data; technical staff for their assistance with DNA extractions.

Table 1: Demographics of cases and controls using the PCR-RFLP method. Significant values (P<0.05) are bolded.

		NSCLC	Controls (all)[a]	P-value[b]
N		1040	784	
Gender[c]	Male	710 (68%)	531 (68%)	0.807
	Female	330 (32%)	253 (32%)	
Age	Mean (SD)	65.7 (10)	63.1 (12.7)	**<0.0001**
	Range	32-92	19-93	
Pack-years[c,d]	Mean (SD)	48 (35.1)	49.1 (33.5)	0.466
	Range	0-270	0-246	-
Histology[c]	AC	407 (39%)	-	-
	SCC	373 (36%)	-	-
	Other	260 (25%)	-	-
Tumour Stage[c]	I	422 (40.6%)	-	-
	II	202 (19.4%)	-	-
	III	231 (22.2%)	-	-
	IV	113 (10.8%)	-	-
	Unknown	72 (7%)	-	

[a]Subjects with COPD were combined with healthy smokers to form the control group
[b]Level of significance (α) was considered to be 0.05, χ^2 or t-test, cases versus all controls
[c]Presented as number (%)
[d]Pack-years (a measure of cumulative smoking exposure) was defined as the average number of packs (20 cigarettes/pack) of cigarettes smoked per day multiplied by the number of years of smoking

Table 2: Hardy-Weinberg chi-square tests for NSCLC and controls.

	PCR-RFLP		Sequenom IPLEX	
Genotypes	NSCLC[a]	Controls[a]	NSCLC[a]	Controls[a]
CYP1A1*2A	0.16	0.45	0.04	0.51
CYP1A1*2C	0.03	0.01	0.34	0.07
CYP1A1*4	0.48	0.27	0.96	0.18

[a]*P* values >0.05 were considered to be consistent with Hardy-Weinberg equilibrium

Table 3: Distribution of CYP1A1 genotypes (CYP1A1*2C, CYP1A1*2A, CYP1A1*4) amongst NSCLC subjects and controls using PCR-RFLP and Sequenom IPLEX methods. Significant associations (P <0.05) are shown in bold.

		Cases		***Controls***[a]		***Sequenom - Samples failed or not re-genotyped***			
		PCR-RFLP	**Sequenom IPLEX**	**PCR-RFLP**	**Sequenom IPLEX**	**Samples failed**	**Samples not re-genotyped**	***P*-value RFLP**[b]	***P*-value Sequenom IPLEX**[b]
*CYP1A1*2C*	A/A	929 (89.3%)	827 (89.9%)	741 (94.5%)	634 (94.2%)	53 (3.2%)	162 (9.7%)	**<0.001**	**0.008**
	A/G	103 (9.9%)	89 (9.7%)	40 (5.1%)	37 (5.5%)	2 (1.4%)	11 (7.7%)		
	G/G	8 (0.8%)	4 (0.4%)	3 (0.4%)	2 (0.3%)	1 (8.3%)	1 (8.3%)		
	Total	1040	920	784	673	56	174		
*CYP1A1*2A*	TT	797 (76.6%)	675 (77.4%)	646 (82.4%)	541 (82.3%)	67 (4.6%)	150 (10.4%)	**0.008**	**0.008**
	CT	219 (21.1%)	197 (22.6%)	128 (16.3%)	116 (17.7%)	20 (5.8%)	21 (6.1%)		
	CC	24 (2.3%)	24 (2.7%)	10 (1.3%)	8 (1.2%)	1(2.9%)	3 (8.8%)		
	Total	1040	896	784	665	88	174		
*CYP1A1*4*	CC	965 (92.8%)	859 (93.4%)	698 (89.03%)	607 (90.3%0	49 (2.9%)	159 (9.6%)	**0.003**	**0.049**
	CA	74 (7.1%)	60 (6.5%)	81 (10.33%)	65 (9.7%)	9 (5.6%)	14 (8.8%)		
	AA	1 (0. 1%)	1 (0.1%)	5 (0. 64%)	0 (0%)	0 (0%)	0 (0%)		
	Total	1040	920	784	672	58	173		

[a]Subjects with COPD were combined with healthy smokers to form the control group.
[b]Level of significance (α) was considered to be 0.05, χ^2 or *t*-test, cases versus all controls
PCR-RFLP – Polymerase Chain Reaction-Restriction Fragment Length Polymorphism.

Table 4: Distribution of CYP1A1 genotypes (CYP1A1*2C, CYP1A1*2A, CYP1A1*4) and haplotypes amongst NSCLC subjects and controls using PCR-RFLP methods. Significant associations (P <0.05) are shown in bold.

	PCR-RFLP				Sequenom IPLEX			
	NSCLC[a] n = 1040	Controls[a,b] n = 784	*P*-value	OR (95% CI)	NSCLC[a]	Controls[a,b]	*P*-value	OR (95% CI)
Genotypes								
*CYP1A1*2C*								
AA	929 (89.3%)	741 (94.5%)	**≤0.001**	2.057 (1.77-2.39)	827 (89.9%)	663 (94.2%)	**0.002**	1.829 (1.57-2.12)
AG or GG	111 (10.7%)	43 (5.5%)			93 (10.1%)	39 (5.8%)		
*CYP1A1*2A*								
TT	797 (76.6%)	646 (82.4%)	**0.003**	1.427 (1.35-1.51)	675 (75.3%)	541 (81.4%)	**0.005**	1.428 (1.35-1.52)
CT or CC	243 (23.4%)	138 (17.6%)			221 (24.7%)	124 (18.6%)		
*CYP1A1*4*								
CC	965 (92.8%)	698 (89.0%)	**0.005**	0.637 (0.64-0.62)	859 (93.4%)	607 (90.3%)		0.664 (0.67-0.66)
CA or AA[c]	75 (7.2%)	86 (11.0%)			61 (6.6%)	65 (9.7%)	**0.026**	
Haplotypes[d]								
C-A-T	1715 (82.3%)	1329(84.5%)	0.0511	1.0 (Ref.)	1526 (82.8%)	1151 (85.3%)	0.0536	
C-A-C	178 (8.5%)	109 (6.9%)	0.0747	1.251 (0.98 – 1.60)	157 (8.5%)	91 (6.7%)	0.0629	
A-A-T	72 (3.5%)	84 (5.6%)	**0.0014**	0.624 (0.45 – 0.86)	62 (3.4%)	65 (4.8%)	**0.0365**	
C-G-C	85 (4.1%)	37 (2.3%)	**0.0038**	1.761 (1.19 – 2.60)	97 (5.3%)	41 (3.0%)	**0.0023**	
C-G-T	30 (1.4%)	7 (0.4%)	**0.0026**	3.259 (1.46 – 7.28)	N/A	N/A		

[a] Presented as number (%)
[b] Subjects with COPD were combined with healthy smokers to form the control group
[c]No AA genotypes were observed in this cohort.
[d]Haplotype order 2453C>A, 2455A<G, 3801T<C

References:

[1] Shopland D, Eyre H, Pechacek T. Smoking-attributable cancer mortality in 1991: is lung cancer now the leading cause of death among smokers in the United States? J Natl Cancer Institute. 1991;83(16):1142-48.

[2] Bouchardy C, Benhamou S, Jourenkova N, Dayer P, Hirvonen A. Metabolic genetic polymorphisms and susceptibility to lung cancer. Lung cancer (Amsterdam, Netherlands). 2001;32(2):109-12.

[3] Shah PP, Singh AP, Singh M, Mathur N, Pant MC, Mishra BN, et al. Interaction of cytochrome P4501A1 genotypes with other risk factors and susceptibility to lung cancer. Mutation research. 2008 Mar 1;639(1-2):1-10.

[4] Kawajiri K, Eguchi H, Nakachi K, Sekiya T, Yamamoto M. Association of CYP1A1 germ line polymorphisms with mutations of the p53 gene in lung cancer. Cancer research. 1996 Jan 1;56(1):72-6.

[5] Yang XR, Wacholder S, Xu Z, Dean M, Clark V, Gold B, et al. CYP1A1 and GSTM1 polymorphisms in relation to lung cancer risk in Chinese women. Cancer letters. 2004 Oct 28;214(2):197-204.

[6] Cascorbi I, Brockmoller J, Roots I. A C4887A polymorphism in exon 7 of human CYP1A1: population frequency, mutation linkages, and impact on lung cancer susceptibility. Cancer research. 1996 Nov 1;56(21):4965-9.

[7] Sugimura H, Wakai K, Genka K, Nagura K, Igarashi H, Nagayama K, et al. Association of Ile462Val (Exon 7) polymorphism of cytochrome P450 IA1 with lung cancer in the Asian population: further evidence from a case-control study in Okinawa. Cancer Epidemiol Biomarkers Prev. 1998 May;7(5):413-7.

[8] Chowbay B, Zhou S, Lee EJ. An interethnic comparison of polymorphisms of the genes encoding drug-metabolizing enzymes and drug transporters: experience in Singapore. Drug metabolism reviews. 2005;37(2):327-78.

[9] Oyama T, Mitsudomi T, Kawamoto T, Ogami A, Osaki T, Kodama Y, et al. Detection of CYP1A1 gene polymorphism using designed RFLP and distributions of CYP1A1 genotypes in Japanese. International archives of occupational and environmental health. 1995;67(4):253-6.

[10] Inoue K, Asao T, Shimada T. Ethnic-related differences in the frequency distribution of genetic polymorphisms in the CYP1A1 and CYP1B1 genes in Japanese and Caucasian populations. Xenobiotica; the fate of foreign compounds in biological systems. 2000 Mar;30(3):285-95.

[11] Larsen JE, Colosimo ML, Yang IA, Bowman R, Zimmerman PV, Fong KM. CYP1A1 Ile462Val and MPO G-463A interact to increase risk of adenocarcinoma but not squamous cell carcinoma of the lung. Carcinogenesis. 2006 Mar;27(3):525-32.

[12] Sreeja L, Syamala V, Hariharan S, Madhavan J, Devan SC, Ankathil R. Possible risk modification by CYP1A1, GSTM1 and GSTT1 gene polymorphisms in lung cancer susceptibility in a South Indian population. Journal of human genetics. 2005;50(12):618-27.

[13] Song N, Tan W, Xing D, Lin D. CYP 1A1 polymorphism and risk of lung cancer in relation to tobacco smoking: a case-control study in China. Carcinogenesis. 2001 Jan;22(1):11-6.

[14] Xu X, Kelsey KT, Wiencke JK, Wain JC, Christiani DC. Cytochrome P450 CYP1A1 MspI polymorphism and lung cancer susceptibility. Cancer Epidemiol Biomarkers Prev. 1996 Sep;5(9):687-92.

[15] Vineis P, Veglia F, Benhamou S, Butkiewicz D, Cascorbi I, Clapper ML, et al. CYP1A1 T3801 C polymorphism and lung cancer: a pooled analysis of 2451 cases and 3358 controls. International journal of cancer. 2003 May 1;104(5):650-7.

[16] Cote ML, Wenzlaff AS, Bock CH, Land SJ, Santer SK, Schwartz DR, et al. Combinations of cytochrome P-450 genotypes and risk of early-onset lung cancer in Caucasians and African Americans: a population-based study. Lung cancer (Amsterdam, Netherlands). 2007 Mar;55(3):255-62.
[17] Tefre T, Ryberg D, Haugen A, Nebert DW, Skaug V, Brogger A, et al. Human CYP1A1 (cytochrome P(1)450) gene: lack of association between the Msp I restriction fragment length polymorphism and incidence of lung cancer in a Norwegian population. Pharmacogenetics. 1991 Oct;1(1):20-5.
[18] Hirvonen A, Husgafvel-Pursiainen K, Karjalainen A, Anttila S, Vainio H. Point-mutational MspI and Ile-Val polymorphisms closely linked in the CYP1A1 gene: lack of association with susceptibility to lung cancer in a Finnish study population. Cancer Epidemiol Biomarkers Prev. 1992 Sep-Oct;1(6):485-9.
[19] Jacquet M, Lambert V, Baudoux E, Muller M, Kremers P, Gielen J. Correlation between P450 CYP1A1 inducibility, MspI genotype and lung cancer incidence. Eur J Cancer. 1996 Sep;32A(10):1701-6.
[20] Boyapati SM, Shu XO, Gao YT, Cai Q, Jin F, Zheng W. Polymorphisms in CYP1A1 and breast carcinoma risk in a population-based case-control study of Chinese women. Cancer. 2005 2005 Jun 1 DTBS- PubMed MEDLINE;103(11):2228-35.
[21] Wenzlaff AS, Cote ML, Bock CH, Land SJ, Santer SK, Schwartz DR, et al. CYP1A1 and CYP1B1 polymorphisms and risk of lung cancer among never smokers: a population-based study. Carcinogenesis. 2005 Dec;26(12):2207-12.
[22] Mrozikiewicz PM, Cascorbi I, Brockmoller J, Roots I. Pharmacogenetics. 1997 Aug;7(4):303-7.
[23] Bosch TM, Doodeman VD, Smits PH, Meijerman I, Schellens JH, Beijnen JH. Pharmacogenetic screening for polymorphisms in drug-metabolizing enzymes and drug transporters in a Dutch population. Molecular diagnosis & therapy. 2006;10(3):175-85.
[24] Hung RJ, Boffetta P, Brockmoller J, Butkiewicz D, Cascorbi I, Clapper ML, et al. CYP1A1 and GSTM1 genetic polymorphisms and lung cancer risk in Caucasian non-smokers: a pooled analysis. Carcinogenesis. 2003 May;24(5):875-82.
[25] Han W, Pentecost BT, Spivack SD. Functional evaluation of novel single nucleotide polymorphisms and haplotypes in the promoter regions of CYP1B1 and CYP1A1 genes. Molecular carcinogenesis. 2003 Jul;37(3):158-69.
[26] Larsen JE, Colosimo ML, Yang IA, Bowman R, Zimmerman PV, Fong KM. Risk of non-small cell lung cancer and the cytochrome P4501A1 Ile462Val polymorphism. Cancer Causes Control. 2005 Jun;16(5):579-85.
[27] Hayashi S, Watanabe J, Nakachi K, Kawajiri K. Genetic linkage of lung cancer-associated MspI polymorphisms with amino acid replacement in the heme binding region of the human cytochrome P450IA1 gene. Journal of biochemistry. 1991 Sep;110(3):407-11.
[28] Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics. 2005 Jan 15;21(2):263-5.
[29] Houlston RS. CYP1A1 polymorphisms and lung cancer risk: a meta-analysis. Pharmacogenetics. 2000 Mar;10(2):105-14.
[30] Dong LM, Potter JD, White E, Ulrich CM, Cardon LR, Peters U. Genetic susceptibility to cancer: the role of polymorphisms in candidate genes. Jama. 2008 May 28;299(20):2423-36.
[31] Taioli E, Gaspari L, Benhamou S, Boffetta P, Brockmoller J, Butkiewicz D, et al. Polymorphisms in CYP1A1, GSTM1, GSTT1 and lung cancer below the age of 45 years. International journal of epidemiology. 2003 Feb;32(1):60-3.
[32] Taioli E, Crofts F, Trachman J, Bayo S, Toniolo P, Garte SJ. Radical differences in CYP1A1 genotype and function. Toxicol Lett. 1995 May;77(1-3):357-62.

[33] Schwarz D, Kisselev P, Honeck H, Cascorbi I, Schunck WH, Roots I. Co-expression of human cytochrome P4501A1 (CYP1A1) variants and human NADPH-cytochrome P450 reductase in the baculovirus/insect cell system. Xenobiotica; the fate of foreign compounds in biological systems. 2001 Jun;31(6):345-56.
[34] Lee KM, Kang D, Clapper ML, Ingelman-Sundberg M, Ono-Kihara M, Kiyohara C, et al. Cancer Epidemiol Biomarkers Prev Preventive Oncology. 2008 May Analysis among Asian Populations.;17(5):1120-6.
[35] Chang BL, Zheng SL, Isaacs SD, Turner A, Hawkins GA, Wiley KE, et al. Polymorphisms in the CYP1A1 gene are associated with prostate cancer risk. International journal of cancer. 2003 Sep 1;106(3):375-8.
[36] Kristensen VN, Kure EH, Erikstein B, Harada N, Borresen-Dale A. Genetic susceptibility and environmental estrogen-like compounds. Mutation research. 2001 Oct 1;482(1-2):77-82.
[37] Wang Y, Sudilovsky D, Zhang B, Haughney PC, Rosen MA, Wu DS, et al. A human prostatic epithelial model of hormonal carcinogenesis. Cancer research. 2001 Aug 15;61(16):6064-72.
[38] Hsing AW. Hormones and prostate cancer: where do we go from here? Journal of the National Cancer Institute. 1996 Aug 21;88(16):1093-5.