CHAPTER 6

DISCUSSION

The *CYP2A13* has recently been great interesting of many molecular epidemiologists of associations between its genetic polymorphisms and lung cancer susceptibility due to its predominant expression in the respiratory tract and its high efficiency in the metabolic activation of NNK, a tobacco-specific lung carcinogen. Large inter-individual differences exist in the expression of *CYP2A13*, which likely contribute to the different susceptibility to lung cancer in smokers. Since the first report on genetic polymorphisms in the *CYP2A13* gene generated by Zhang and colleagues in 2002 at the Wadsworth center, State University of New York at Albany, many studies have announced the *CYP2A13* variant alleles with different frequency distribution in several ethnic groups (Zhang et al., 2002; Von Weymarn and Murphy, 2003; Wang et al., 2003; Cheng et al., 2004; Cuaffiez et al., 2004; 2005). To date, more than 20 mutant alleles have been detected in the *CYP2A13* gene and publicized on the Human Cytochrome P450 Allele Nomenclature (www. imm.ki.se/CYPalleles) but only three of them (578C>T, 3375C>T, and 7520C>G) have been reported to be functional SNPs which affect the enzyme activity (Zhang et al., 2002; 2004). However, no study carried out in Thai populations has been published. Therefore, the aim of the present study was to determine the frequency distribution of six SNPs of *CYP2A13* in Thai subjects from the Northern region of Thailand. This is the first report on genetic polymorphisms of *CYP2A13* gene in Thai population.

In this study, a 336 DNA samples were collected from unrelated healthy subjects (175 males and 161 females) who were born and resided in the upper Northern region of Thailand. Of the six SNPs chosen for this study, four SNPs are in coding region (74G>A in exon 1, 578C>T in exon 2, 1662G>C in exon 3 and 3375C>T in exon 5) and other two SNPs in non coding region, which are lie in 3'-untranslated region (7520C>G and 7571G>C). Several methods based on the polymerase chain reaction (PCR) amplification are available and a method suitable to the nature of each SNP is adopted for genotyping. The PCR-restriction fragment length polymorphism (PCR-RFLP), the most common technique used for genotyping, was applied to genotype four SNPs (74G>A, 1662G>C, 3375C>T and 7520C>G) of *CYP2A13*. The mismatch allele-specific amplification (mASA) and the tetra-primer PCR were developed for detected 578C>T and 7571G>C, respectively.

The common allele-specific amplification PCR depends on the differential probability of PCR primer extension by a DNA polymerase between a wild-type and mutant template. While one might imagine a priori that primer extension would be effectively blocked by a single mismatch at the ultimate or penultimate base pair of the primer/template duplex this is not generally observed under conditions commonly employed in PCR. Cha et al. in 1992 found that several different single primer/template mismatches had no effect on primer extension probability. However, when both ultimate and penultimate base pairs were mismatched, primers were, in general, effectively blocked in primer extension (Petruska et al. 1988; Mendelman et al., 1990). The mismatch allele-specific primers described here were designed to contain additional mismatch in the typing primer at third nucleotide form 3' end. Experience with this approach has taught that incorporated penultimate mismatch at

the third position from 3' end of the typing primer yields superior discrimination which gave the greatest degree of specificity (Sudo et al., 2006). Thus by creating a primer with a single mismatch with the desired mutant sequence and a double mismatch with the wild-type sequence, a desirable difference in primer extension probabilities was obtained, permitting measurement of that particular mutation allele (578T mutant allele). In addition to the ease and technical simplicity of this assay, the mismatch allele-specific PCR method requires only standard PCR equipment and electrophoresis or other common separation platform. Using the conditions reported here, are confident that other researchers can reproduce this findings of sensitivity and be able to measure this common mutations in exon 2 from any subjects.

In this study, the technique, named tetra-primer PCR assay, adopts certain principles from Ye et al. by adjusting PCR conditions including primer concentration, annealing temperature and use of the hot start technique to improve detection of the 7571G>C polymorphism in Thai subjects (data not show). The tetra-primer PCR assay presented here are less prone to contamination because they are single-tube assay and do not require transfer of amplified PCR products, and the risk for false-negative results are omitted because the assay include an internal control for PCR amplification (Ye et al., 1992; Wang et al., 1999; Ling et al., 2002; Hersberger et al., 2000; 2002; Okayama et al., 2004). For each assay, four primers were combined in a single PCR tube. Two primers that are complementary to unique sequences of *CYP2A13* (UtrF and UtrR) used for the initial amplification of the *CYP2A13* locus and two primers that are designed for the allele-specific amplification of the 7571G wild-type and the 7571C variant.

The results of genotyping of all six SNPs were also confirmed by sequencing analysis. All sequence confirmed the results of the PCR-based analysis. This is the first time that single nucleotide polymorphisms of *CYP2A13* have been detected in Thai population subjects by mismatch allele-specific amplification and tetra-primer PCR assays.

The present study showed that the 74G>A and 3375C>T were the most frequent occurring SNPs in Northern Thai population, with frequencies of 4.9% and 4.9%, respectively, followed by 7571G>C and 7520C>G, with frequencies of 4.6% and 4.2%, respectively. The least frequency was 578C>T (0.1%). However, of all 336 genomic DNA samples in this study, there were no cases of 1662G>C variant allele. It was the only one SNP that was not found variant allele in this Northern Thai population examined.

A 74G>A polymorphism located in exon 1 is missense mutation, leading to a predicted amino acid change from Arg to Gln at position 25 (Arg25Gln). This SNP was detected as heterozygotes in 33 subjects and no homozygote with allele frequency of 4.9% (95%CI 3.5-6.9). Among the thirty-three heterozygotes, 13 were males and 20 were females. This allele frequency was slightly lower than that in Japanese (7.3%) (Fujieda et al., 2003) indicating interethnic differences.

A 578C>T polymorphism in exon 2 is missense mutation, resulting in an amino acid change from an Arg residue to a stop codon at position 101 (Arg101stop) leads to the synthesis of a truncated protein presumably non functional. Zhang et al. (2003) firstly detected this variant in 2 of 31 Chinese individuals (3.2% frequency) and not in other ethnic groups. However, Cauffiez et al. (2004) reported that this deleterious variant was the most frequent (5%) variant allele in the French population

and suggested that the Arg101Stop SNP might be related to an increased risk for small cell lung carcinoma. In this study, the 578C>T mutation was detected as heterozygote in only one (male) subject, with an allele frequency of 0.1% (95%CI 0.0-1.0) suggesting that it is a rare mutation in Thai population.

A 1662G>C polymorphism in exon 3 is a missense mutation. Its cause a Gly144Arg change in the predicted amino acid sequence. However, this SNP was not detected in the present study. Zhang et al., (2002) found this variant as a rare mutation in Hispanic newborn (1 heterozygote).

A 3375C>T polymorphism in exon 5 is a missense mutation, resulting in an amino acid change from an Arg to a Cys codon at position 257 (Arg257Cys). This SNP was detected as hetorzygotes in 33 subjects and no homozygote. Among the thirty-three heterozygotes, 17 were males and 16 were females. The 3375C>T allele frequency in this study samples was 4.9% (95%CI 3.5-6.9), which is comparable with the previous studies carried out in Hispanic (5.8%) (Zhang et al., 2002), Chinese Han (5.6%) (Cheng et al., 2004) and Tunisian populations (4.2%) (Cauffiez et al., 2005). This polymorphism was found in Gabonese population at a highest allele frequency (15.3%), followed by Black (14.4%) but least frequency in White Caucasian (1.9%) (Table 2). Cauffiez et al. (2004) detected the CYP2A13 polymorphisms in French population but did not found 3375C>T in such population. Recently Zhang et al., (2002) studied the function of the Arg257Cys variant protein following its heterologous expression on CYP2A13 catalytic activity. The 257Cys variant was found to be 37 to 56% less active than the wild-type Arg257 protein toward the substrate NNK and has been suggested to provide some protection against toxic compounds in the respiratory tract. An epidemiological studies carried out in Chinese

and Japanese populations reported the significance of the 3375C>T in lung cancer, demonstration that 3375C>T polymorphism is associated with a reduced risk for smoking-related lung adenocarcinoma (Wang et al., 2003; Kiyohara et al. 2005). However, in another case-control study, no significant association was found between the CYP2A13 genetic polymorphism and the risk of developing nasopharyngeal carcinoma in a Cantonese population of Southern China (Jiang et al., 2004).

The others two SNPs lied in non-coding 3'-untranslated region, 7520C>G and 7571G>C, were also chosen for this study. The 7520C>G SNP in this study samples was detected as hetorzygotes in 22 subjects (9 males and 13 females) and homozyguos mutant in 3 subjects (2 males and 1 female), with allele frequency of 4.2% (95%CI 2.8-6.1). The 7571G>C polymorphism is also detected in 3'Untranslated region and have no amino acid change. This SNP was detected as only heterozygote in 31 subjects, with an apparent allele frequency of 4.6% (95%CI 3.2-6.6), which contrast to the results in Chinese cancer patients with a high frequency of this variant (10.9%) (Zhang et al., 2004). Among the thirty-one heterozygotes, 18 were males and 13 were females.

Zhang et al. (2004) found the 7520C>G variation in anonymous White, Black, Hispanic, and Asian newborn from New York State with the frequencies of 5.2%, 26.8%, 17.7%, and 4.3%, respectively. Whereas the allele frequencies of 7520C>G in French Caucasian, Gabonese and Tunisian population were 1.0%, 20.8%, and 7.3%, respectively (Cauffiez et al., 2005). These results indicated the inter-ethnic differences of this polymorphism. Although homozygosity of this variant was accounted for <1% of variant alleles in this study, the 3'-untranslated region polymorphism 7520C>G has been correlated with the tenfold decrease in CYP2A13

expression in human adult and fetal lung (Zhang et al., 2004). Despite being located in the non-coding region of the gene, these mutations have been reported to affect the stability of the transcript and might alter enzymatic activity (Day and Tuite 1998).

However, functional consequences of the non coding region SNPs are more difficult to predict. The two SNPs in the 3'UTR are not expected to alter RNA folding, according to an analysis using mfold, version3.1 (http://www.bioinfo.rpi.edu/applications/mfold) (Mathews et al., 1999). Further studies on the impact of these mutations on the expression of the *CYP2A13* gene are necessary, since significant inter-individual differences in the level of CYP2A proteins have been found in microsomes from fetal nasal mucosa (Gu et al., 2000).

Using five SNPs detected in 336 Thai subjects, the haplotype configuration was performed. Among the five SNPs, the following three: 74G>A, 3375C>T, and 7571G>C were simultaneously observed as heterozygotes in 18 subjects (5.4%), suggesting a haplotype of 74A, 3375T, and 7571C. These three SNPs were included in the allele of *CYP2A13*2B* which consists of nine additional SNPs (-1479T>C, -1429A>G, -1240A>G, -411G>A, 74G>A, 1757A>G, 2211T>C, 3375C>T, 6424C>T, 6432C>T, 7233T>G, 7571G>C). The two SNPs in exon 1 and exon 5 (74G>A and 3375C>T) simultaneously observed as heterozygotes in 13 subjects (3.9%) were corresponded to **2A* allele. Interestingly, The 7520C>G SNP which is constituted in **1H* and **3* were simultaneously observed as heterozygotes in 19 subjects (5.7%) and homozygotes in 3 subjects (0.9%). The further studies will be necessary to warrant that these *CYP2A13* haplotypes observed in this study associate with an incidence of smoking-related carcinoma in respect of ethnicity.