Review article:

DISTRIBUTION OF POLYMORPHIC VARIANTS OF CYP2A6 AND THEIR INVOLVEMENT IN NICOTINE ADDICTION

Luis A. López-Flores¹, Gloria Pérez-Rubio¹, Ramcés Falfán-Valencia¹*

¹ Instituto Nacional de Enfermedades Respiratorias Ismael Cosio Villegas
* Corresponding autor: Ramcés Falfán-Valencia, Ph. D.
Phone +52 55 5487 1700 ext. 5152. E-mail: rfalfanv@iner.gob.mx

http://dx.doi.org/10.17179/excli2016-847

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/).

ABSTRACT

Tobacco consumption has become a major public health issue, which has motivated studies to identify and understand the biological processes involved in the smoking behavior for prevention and smoking cessation treatments. CYP2A6 has been identified as the main gene that codifies the enzyme that metabolizes nicotine. Many alleles have been identified after the discovery of CYP2A6, suggesting a wide interethnic variability and a diverse smoking behavior of the allele carrying individuals. The main purpose of this review is to update and highlight the effects of the CYP2A6 gene variability related to tobacco consumption reported from diverse human populations. The review further aims to consider CYP2A6 in future studies as a possible genetic marker for the prevention and treatment of nicotine addiction. Therefore, we analyzed several population studies and their importance at addressing and characterizing a population using specific parameters. Our efforts may contribute to a personalized system for detecting, preventing and treating populations at a higher risk of smoking to avoid diseases related to tobacco consumption.

Keywords: CYP2A6, ethnic differences, genetic polymorphism, tobacco consumption, nicotine metabolism, nicotine addiction

INTRODUCTION

Tobacco consumption has become an epidemic affecting more than 1,000 million people worldwide and is considered the main cause of avoidable death causing approximately 6 million premature deaths each year (WHO, 2011). Thus, tobacco consumption is a public health issue combined with economic losses; it has motivated studies for identifying and understanding the biological processes involved in the smoking behavior for prevention and smoking cessation treatments (Bierut et al, 2014).

Among cigarette compounds, nicotine is responsible for causing dependence by stimulating the smoker and allowing the other compounds to access the body, causing chronic harmful effects and tobacco-related diseases. Tobacco consumption is caused by environmental, psychosocial and genetic factors (Bierut, et al., 2014). Previous studies have identified genes encoding proteins that influence nicotine addictive behavior due to their effect on the cerebral neurotransmission pathways (Al Koudsi and Tyndale, 2005; Arinami et al., 2000; Verde Rello and Santiago Dorrego, 2013). Moreover, some gene prod-
ucts are involved in nicotine response as receptors and metabolizers (Hukkanen et al., 2005; Verde Rello and Santiago Dorrego, 2013). Pérez-Rubio et al. (2015) have reviewed this subject in detail. In the metabolizers group, the CYP2A6 gene plays an important role. Its protein product (by the same name) is the principal enzyme responsible for nicotine metabolism to cotinine and other sub-products in the human body. However, more than 45 alleles have been discovered suggesting wide interindividual and inter-ethnic variety.

The main purpose of this review is to update and highlight the effects of the genetic polymorphisms of CYP2A6 related to tobacco consumption reported in diverse human populations. Moreover, we aimed to consider these polymorphisms in future studies as possible genetic markers for the prevention and treatment of nicotine addiction.

**BIOLOGICAL FUNCTION OF CYP2A6 IN SMOKING**

The CYP2A6 enzyme belongs to the enzyme superfamily known as the cytochrome P450 system (CYP450), also classified in the drug metabolizing enzymes group. These enzymes are found in the endoplasmic reticulum of the cells of some tissues in the body, particularly in the liver. Moreover, they are phase I enzymes responsible for metabolizing more than 80% of drugs such as xenobiotics and endogen products in the body (Evans and Relling, 2004; Ingelman-Sundberg, 2004).

CYP2A6 is found mainly in the liver and is approximately 5-10% of the total CYP450 content (Shimada et al., 1996; Yamano et al., 1990), although it has also been found to be expressed in the nasal mucosa, trachea, bronchi, lungs, sinuses and several brain regions (Bhagwat et al., 2000; Chiang et al., 2012; Crawford et al., 1998; Ding and Kaminsky, 2003; Macé et al., 1998).

CYP2A6 has demonstrated to be involved in the metabolism of some endogen and exogen substrates such as precarcinogenic and carcinogenic compounds, as well as some toxins and drugs including nicotine.

CYP2A6 has been particularly important in tobacco consumption because of its involvement in nicotine metabolism. Nicotine is the main compound in tobacco responsible for the development of cigarette addiction by stimulating nicotinic cholinergic receptors (nAChR) that release neurotransmitters in the brain and cause a pleasant sensation in the smoker. The nicotine availability in the body is mediated by biological factors, mainly those related to its metabolism. Smokers tend to consume the same amount of nicotine each day to acquire the desired effects by modulating their smoking behavior to adjust nicotine availability for the purpose of regulating nicotine levels in the body (Benowitz, 1992).

It has been reported that CYP2A6 is the main enzyme involved in the nicotine oxidation to cotinine. CYP2A6 catalyzes approximately 80% (55-92%) of this reaction via C-oxidation in addition to other metabolic pathways for nicotine and its metabolites (Benowitz and Jacob III, 1994; Messina et al., 1997; Nakajima et al., 1996). Some other enzymes of CYP450 contribute to a lesser degree to nicotine metabolism such as CYP2B6, CYP2A13, CYP2D6 and CYP2E1 (mentioned in the order of relevance).

These biological products interact directly with nicotine to affect physiological brain processes (e.g., nAChR) and are inactivated and removed from the body (e.g., CYP450 enzymes) making their genes ideal candidates for altering smoking behavior (Malaiyandi et al., 2005).

**CYP2A6 GENETIC VARIANTS**

The CYP2A6 gene has a 6 kb extension length, and it is composed of 9 exons, which encode for a 494 amino-acid product (Fernandez-Salguero et al., 1995). It is located in the chromosomal band 19q13.2, where other CYP450 gene subfamilies (CYP2B, CYP2F, CYP2G, CYP28, and CYP2T) are also present. The CYP2A subfamily cluster includes the CYP2A6, CYP2A7, and CYP2A13 genes and other pseudogenes (Hoffman et al., 2001).
The **Human CYP-Allele Nomenclature Database (HCAND)** ([http://www.cypalleles.ki.se](http://www.cypalleles.ki.se)) was created in 1999 for the identification and characterization of CYP2A6 alleles (and other CYP450 genes). This database consists of a committee for unifying and assigning nomenclature for the already discovered alleles and alleles to be discovered in the future (Sim and Ingelman-Sundberg, 2010, 2013).

To date, 42 well-characterized alleles and some haplotypes that are uncharacterized (“CYP2A6 allele nomenclature,” 2014) have been identified. These alleles are determined according to the origins of their mutation(s), such as gene conversion, gene deletion, gene duplication and/or single nucleotide polymorphism (SNP). The gene mutations are summarized in Figure 1.

The wild-type allele that is considered as a reference is **CYP2A6*1A** (Yamano et al., 1990). **CYP2A6*1B** consists of a 58 bp gene conversion with **CYP2A7** in the **CYP2A6** 3’ UTR region (Ariyoshi et al., 2000; Yamano et al., 1990). *In vitro* and *in vivo* assays have shown that the **CYP2A6** 3’ UTR region has more enzymatic activity and RNA stabilization than the reference allele (Ho et al., 2009; Wang et al, 2006; Yoshida et al., 2002). However, several haplotypes contain this gene conversion and other noncoding genetic changes and have been designated **CYP2A6*1B1-B17** (Ariyoshi et al., 2000; Haberl et al., 2005; Mwenifumbo et al., 2007, 2008, 2010; Nakajima et al., 2006; Pitarque et al., 2004; Yamano et al., 1990). Moreover, it has been determined that other wild-type alleles, named **CYP2A6*1D-L**, that have genetic changes in coding, noncoding and regulatory regions (Mwenifumbo et al., 2008, 2010; Nakajima et al, 2004; Pitarque et al., 2004; von Richter et al., 2004; Yamano et al., 1990).

**CYP2A6*2** consists of a missense mutation of 1799T>A, causing an amino acid change of Leu160His in the enzyme. Thus, the protein does not incorporate the heme group, inactivating the enzyme for *in vitro* and *in vivo* assays (Benowitz et al., 1995; Hadidi et al., 1997; Oscarson et al., 1999b; Yamano et al., 1990).

**CYP2A6*3** is presumed to be a gene conversion of **CYP2A6** to **CYP2A7**, which results in an inactive enzyme; however, the methodology for their detection, function, and allelic frequency has been controversial (Fernandez-Salguero et al., 1995; Oscarson et al., 1998; Yamano et al., 1990).

![Figure 1: CYP2A6 location and mutations in the gene. Mutations are reported by the Human CYP-Allele Nomenclature Database (http://www.cypalleles.ki.se) and include insertion, deletion, CNV and SNP. Ref Seq NG_008377.1](http://www.cypalleles.ki.se)
CYP2A6*4 consists of a homologous unequal crossover with CYP2A7 on several positions (CYP2A6*4A-F), which leads to a whole gene deletion causing loss of enzymatic activity (Kitagawa et al., 1999; Nakajima et al., 2000, 2001; Nunoya et al., 1999).

CYP2A6*5 contains a missense mutation, 6582G>T, creating a Gly479Val amino acid change and resulting in a lack of enzyme function (Oscarson et al., 1999b).

CYP2A6*6 contains a missense mutation, 6582G>T, which creates an Arg128Gln amino acid change causing lower enzymatic activity of the enzyme (Kitagawa et al., 2001).

CYP2A6*7 contains a missense mutation, 6558G>A, which produces an Ile471Thr amino acid change that decreases the enzymatic activity to metabolize some substrates for in vivo and in vitro assays (Ariyoshi et al., 2001; Uno et al., 2013; Xu et al., 2002).

CYP2A6*8 contains a missense mutation, 6600G>T, creating an amino acid change in Arg485Leu; however, the mutation’s effect apparently does not change the enzymatic activity (Xu et al., 2002).

CYP2A6*9 contains a point mutation, -48T>G, on the TATA box located in the gene promoter, which results in a decrease of more than 50% of the enzymatic activity for in vitro and in vivo assays (Pitarque et al., 2001; Yoshida et al., 2003). There have been identified two subtypes of this allele: CYP2A6*9A, which contains an additional -1013A>G point mutation (von Richter et al., 2004), and CYP2A6*9B, which contains the point mutations -1680A>G, -1301A>C, -1289G>A, 1620T>C, 1836G>T, 6354T>C and 6692C>G (Haberl et al., 2005).

CYP2A6*10 contains two point mutations, similar to the CYP2A6*7 and CYP2A6*8 alleles. These point mutations decrease the enzymatic activity dramatically and make it completely inactive for some substrates (Xu et al., 2002).

CYP2A6*11 contains a missense mutation, 3391T>C, which results in the amino acid change of Ser224Pro, decreasing the enzymatic activity (Daigo et al., 2002).
substitutions at 51G>A, 5684T>C and 5702T>C. CYP2A6*18C also has the point mutations at -1680A>G, -1579T>C, -1464A>T, -1301A>C, -1289G>A, -1013A>G, 1620T>C, 5668A>T and 6692C>G. The enzymatic activity of this allele tends to be specific according to the substrate (Fukami et al., 2005a; Haberl et al., 2005).

CYP2A6*19 is produced by the point mutations at 5668A>T, 6354T>C, and 6558T>C and a gene conversion at the 3’UTR with CYP2A7, which correspond to the amino acid changes of Tyr392Phe and Ile471Thr, decreasing the enzymatic activity (Fukami et al., 2005a).

CYP2A6*20 has a frameshift mutation at nucleotides 2140 and 2141, which displaces the frameshift from the codon 196 to stop prematurely at 220 codons. In addition, it has three point mutations at 51G>A, 5684T>C and 6692C>G. These mutations produce a loss-of-function enzyme (Fukami et al., 2005a).

CYP2A6*21 is the result of two point mutations: 51G>A and 6573A>G, which produce an amino acid change Arg203Cys; this decreases the enzymatic activity to 19 % compared to the reference allele (Ho et al., 2009; Mwenifumbo et al., 2008).

CYP2A6*22 has two subvariants, among which CYP2A6*24A has the following point mutations: -1301A>C, -1289G>A, -1013A>G, 1620T>C, 5668A>T, 6692C>G, 2296C>T, 2483G>A, 4071delA, 4636A>C, 5668A, 6586T>C, 6692C>G and 7160A>G which result in the amino acid changes Phe118Leu and Arg203Cys; these decreases the enzymatic activity to 19 % compared to the reference allele (Ho et al., 2009, 2008).

stop prematurely at 5th exon (Mwenifumbo et al., 2008), resulting in a loss-of-function enzyme (Ho et al., 2009; Mwenifumbo et al., 2008).

**CYP2A6*28** has two subvariants that share the following mutations: -1269T>C, 51G>A, 656G>T, 1620T>C, 4681T>G, 5668A, 5738C>T, 5745A>G, 5750G>C, 6354T>C, 6361C>A and 7160A>G, as well as a gene conversion in the 3’UTR of 58 bp. However, **CYP2A6*28A** also carries the point mutations 6385G>T, 6389C>G, 6390T>C, on the other hand **CYP2A6*28B** has additionally the following mutations: 1381_1382CT>TC, 6960_6961insGAAAAG and 1481_1486delCTCTCT. These mutations cause the amino acid change of Asn418Asp and Glu419Asp, however, the enzymatic activity is the same as the reference allele (Mwenifumbo et al., 2008).

According to the HCAND, **CYP2A6*29**, 30, 32 and 33 alleles are now in evaluation phase (“CYP2A6 allele nomenclature,” 2014).

**CYP2A6*31** has two subvariants which share the following mutations: -1289G>A, -1013A>G, -492_-470delCCCCTTCCTGA-GACCCCTTAACCInsATCCATAATGTGG A ATCTG, 16A>C, 51G, 1339C>G, 1620T>C, 2721G>A, 2994T>C, 3255A>G, 3315C>T, 5668A, 6692C>G and 7160A>G. However, **CYP2A6*31A** also has the following point mutations: 7568C>T, 6458A>T, 6558T>C, gene conversion at 3’UTR, 6782C>G, 6835C>A, 6999T>C and 7160A>G that change the amino acid of Asn438Tyr and Ile471Thr (Al Koudsi et al., 2010). These mutations produce a protein that has the amino acid change Asn438Tyr which decrease the enzymatic activity according to in vitro and in vivo assays (Al Koudsi et al., 2010).

**CYP2A6*34** originated by the unequal crossover between **CYP2A6** and **CYP2A7**, which resulted in a hybrid allele at the 5’UTR and exons 1-4 belonging to **CYP2A7** and from the 5th to 9th exon belonging to **CYP2A6** (di Iulio et al., 2009). The enzymatic activity has not yet been evaluated, however could be similar to **CYP2A6*12** and it might be a loss-of-function allele.

**CYP2A6*35** has two subvariants that share the following mutations: -1301A>C, -1289G>A, 1620T>C, 6458A>T, 6782C>G, 7160A>G and a gene conversion at the 3’UTR with **CYP2A7**. However, **CYP2A6*35A** also has the following point mutations: -1013A>G, 720G>A, 1137C>G, 2483G>A, 3225A>G, 6218A>G, 6282A>G, 6293T>C and 6354T>C; on the other hand, **CYP2A6*35B** has the following point mutations: -745A>G, 22C>T, 4084delA, 6835C>A and 6999T>C (Al Koudsi et al., 2010). These mutations produce a protein that has the amino acid change Asn438Tyr which decrease the enzymatic activity according to in vitro and in vivo assays (Al Koudsi et al., 2010).

**CYP2A6*36** is produced by the following mutations: -1301A>C, -1289G>A, -745A>G, 22C>T, 1620T>C, 4084delA, 6584A>T, 6558T>C, gene conversion at 3’UTR, 6782C>G, 6835C>A, 6999T>C and 7160A>G that change the amino acid of Asn438Tyr and Ile471Thr (Al Koudsi et al., 2010). There are no assays that prove their enzymatic activity.

**CYP2A6*37** is a result of the missense mutation 5023A>G, which produces the amino acid change of Tyr351His (Bloom et al., 2011). An in silico assay classified the SNP as harmful, suggesting a decreased enzymatic activity (Bloom et al., 2011).

**CYP2A6*39** was described by Pilinguian et al. (2014) as a missense mutation of 468G>A; however, the HCAND (“CYP2A6 allele nomenclature,” 2014) adds the point mutations 171C>A, 1779G>A and 5717C>T, which cause the amino acid change...
Val68Met. The enzymatic activity of this allele is reported as decreased to half of the reference allele (Piliguian et al., 2014).

CYP2A6*40 has the missense mutation 1767C>G (Piliguian et al., 2014), but later the HCAND (“CYP2A6 allele nomenclature,” 2014) added the point mutations 144G>A, 3492C>T and 5738C>T, which modifies the amino acid to Ile149Met. The enzymatic activity is reported to be half of the reference allele (Piliguian et al., 2014).

CYP2A6*41 contains the missense mutation 3515G>A according to Piliguian et al. (2014), but the HCAND (“CYP2A6 allele nomenclature,” 2014) added the mutations 51G>A and 507C>T, which changed the amino acid to Arg265Gln. This allele was shown to have a minimal alteration in expression and a normal enzymatic activity (Piliguian et al., 2014).

CYP2A6*42 is the result of a missense mutation 3524T>C according to Piliguian et al. (2014), but the HCAND (“CYP2A6 allele nomenclature,” 2014) added the mutations 51G>A and 5684T>C, which made the amino acid change Ile268Thr that decreases the expression and enzymatic activity on in vivo and in vitro assays (Piliguian et al., 2014).

CYP2A6*43 has the missense mutation 4406C>T, which makes the amino acid change to Thr303Ile and shows decreased expression and enzymatic activity in in vivo and in vitro assays (Piliguian et al., 2014).

CYP2A6*44 carries the missense mutation 5661G>A according to Piliguian et al. (2014), but later the HCAND (“CYP2A6 allele nomenclature,” 2014) added the mutations 51G>A, 5738C>T, 5745A>G and 5750G>C, which modify the amino acids to Glu390Lys, Asn418Asp and Glu419Asp. These mutations have been shown to reduce the enzymatic activity and the reference allele expression to one-third (Piliguian et al., 2014).

CYP2A6*45 has a missense mutation at 6531T>C according to Piliguian et al. (2014), but later the HCAND (“CYP2A6 allele nomenclature,” 2014) added the point mutations 51G>A and 4464G>A, which change the amino acid Leu462Pro. The mutations have been proven (as in CYP2A6*44) to reduce the enzymatic activity and the reference allele expression to one-third (Piliguian et al., 2014).

There are two CYP2A6 gene duplications: CYP2A6*1X2A originated by an unequal crossover from the 8th to 9th exon with CYP2A6*4D as a reciprocal product (Rao et al., 2000). CYP2A6*1X2B also originated by an unequal crossover of CYP2A7 from 5.2 to 5.6 kb downstream of the stop codon with CYP2A6*4B as the reciprocal product (Fukami et al., 2007). Its enzyme activity has been shown to increase in in vivo assays (Rao et al., 2000).

CYP2A6 has been suggested as a highly polymorphic gene because it is located in a small chromosomal region that contains several genes and some unequal crossover events, point mutations and genetic conversions between CYP2A6 and CYP2A7 (Hoffman et al., 1995). These facts, plus evolutionary forces such as genetic drift and natural selection, may have resulted in this genetic variability, which spread among human populations (Ingelman-Sundberg, 2004, 2005). This genetic variability, similar to other CYP450 genes, could explain the metabolic response to exogenous compounds such as nicotine and other drugs that ranges from 20-40 % (Ingelman-Sundberg, 2001).

CYP2A6 genotypes can be classified according to their alleles and their enzymatic activity, which is referred to as the metabolism range of 3-hydroxycotinine/cotinine (Dempsey et al., 2004), as mentioned below:

- Ultrarapid metabolizers. Individuals who have an enzymatic activity >100 % of normal; they contain more than two functional alleles (the CYP2A6*1X2 allele).
- Normal metabolizers. Individuals who have an enzymatic activity of 100 % (normal); they contain two functional alleles.
- Intermediate metabolizers. Individuals who have an enzymatic activity of 100 % (normal); they contain two functional alleles.
- Slow metabolizers. Individuals who have an enzymatic activity ≤50% of normal can contain a functional and a loss-of-function allele or even two loss-of-function alleles.

EFFECTS OF CYP2A6 GENETIC VARIANTS ON TOBACCO CONSUMPTION

The genetic variability of CYP2A6 directly influences the range of nicotine metabolism in the body, which can indirectly affect the reinforcing and aversive nicotine properties in the brain and can change the individual risk of nicotine dependence. To prove the effect of CYP2A6 variants on tobacco consumption, several studies were conducted that included family, twin and non-related subject designs. These studies associate an allele with the amount of metabolized nicotine or another variable related to tobacco consumption.

Smokers who carry some CYP2A6 alleles show a different smoking behavior compared to carrying the wild-type allele, suggesting that smokers regulate their smoking behavior to obtain the desired nicotine levels in their body (Malaiyandi et al., 2005; Rao et al., 2000; Schoedel et al., 2004).

The importance of the null and decreased biological function alleles is explained in smokers (carrying these alleles) who exhibit less time smoking (Liu et al., 2011; Malaiyandi et al., 2006b), fewer cigarettes smoked (Audrain-McGovern et al., 2007; Fujieda et al., 2004; Malaiyandi et al., 2005; Minematsu et al., 2006; O’Loughlin et al., 2004; Pan et al., 2015; Rao et al., 2000; Schoedel et al., 2004; Thorgeirsson et al., 2010), fewer aspirations per cigarette (Strasser et al., 2011), later smoking onset (Gu et al., 2000; O’Loughlin et al., 2004; Schoedel et al., 2004) and less nicotine dependence (Wassenaar et al., 2011). Likewise, it has been reported that these individuals respond better to replacement nicotine therapy (Lerman et al., 2010; Malaiyandi et al., 2006b). Moreover, they tolerate withdrawal symptoms (Kubota et al., 2006) better and have higher rates of quitting smoking spontaneously (Chenoweth et al., 2013; Malaiyandi et al., 2006b; Minematsu et al., 2003; Ray et al., 2009). These alleles have also been associated with lung, bladder, nasopharyngeal, esophageal and oral cancer (Fujieda et al., 2004; Hosono et al., 2015; Islam et al., 2013; Ito et al., 2015; Kumondai et al., 2016; Lourembam et al., 2015; Miyamoto et al., 1999; Song et al., 2009; Tamaki et al., 2011; Tan et al., 2001; Tiwawech et al., 2006; Topcu et al., 2002; Wassenaar et al., 2015).

On the other hand, some reports do not prove the association between null and decreased CYP2A6 alleles related to tobacco consumption (London et al., 1999; Loriot et al., 2001; Sabol and Hamer, 1999; Schulz et al., 2001; Tiitonen et al., 2000; Zhang et al., 2001). This lack of association may occur because of several factors such as designing the population stratification (comparing populations where there are substructures between cases and controls) and population ethnicity, lack of detailed phenotypic evaluation, indefinite comorbidities, different genotyping methods, examination of different allelic variants, inconsistency in smoking history and differences in symptoms of nicotine dependence among smokers (Lerman and Niaura, 2002; O’Loughlin et al., 2004).

Detecting the alleles of CYP2A6 can allow us to characterize different smoking behaviors and smoking-related diseases among individuals (Fujieda et al., 2004), due to their role in nicotine metabolism and the metabolism of certain carcinogenic compounds. This could have an implication on public health by reducing the harmful effects related to smoking and developing a personalized smoking cessation according to their individual genotype (Liu et al., 2011; Schoedel et al., 2004). However, these alleles have a specific distribution in worldwide populations.

POPULATION DISTRIBUTION OF CYP2A6 ALLELES

The CYP2A6 allele distribution has an interethnic pattern. Knowing the individual differences in nicotine metabolism may allow us to answer the following questions: Why do some people become regular smokers after in-
itial exposure, while others experience negative reactions and discontinue use? Why do some people smoke in greater quantities than others? Why do different individuals not respond the same way to drug therapies to quit smoking? Why do some individuals develop smoking related diseases faster than others? (O’Loughlin et al., 2004; Schoedel et al., 2004; Swan et al, 1997, 2005).

Therefore, we present the allele frequencies in reported populations where tobacco consumption, cancer and nicotine metabolism were studied and in cohorts and general population studies, which involve CYP2A6. The number of reports of each allele according to a population group is summarized in Figure 2. For more practical reasons, we only showed the frequency without distinction for alleles with subvariants, except for the wild-type allele CYP2A6*1 whose frequency was not completely calculated because some studies assign an unidentified genotype to the reference allele.

The wild-type allele subvariants are distributed in a particular way on worldwide populations. CYP2A6*1A is found in almost all populations studied such as Caucasian populations (Spanish, British, French, Swedish and Serbian), which range from 57-67% (Djordjevic et al., 2010, 2013; Gambier et al., 2005; Huang et al., 2005; Nakajima et al., 2006, 2004; Oscarson, et al., 1999a; Soriano et al., 2011), Africans, and Ethiopians, which reported an allele frequency of 34.8% (Aklillu et al., 2014). However, in African Americans, Ghanaians and Namibians the frequencies were between 66.5-80.5% (Gyamfi et al., 2005; Nakajima et al., 2004; Takeshita et al., 2006) while it was the opposite in Asians. Asians living in the UK report a frequency of 64.1% while East and South Asian populations (Chinese, Japanese, Korean, Malaysian, Thai, Indian, Bangladeshi and Sri Lankan) report a frequency of 27-52% (Ariyoshi et al., 2002; Djordjevic et al., 2013; Islam et al., 2013; Ito et al, 2015; Iwahashi et al., 2004; Kwon et al., 2001; Mahavorasirikul et al., 2009; Nakajima et al., 2001, 2006; Nurfadhlina et al., 2006; Oscarson et al., 1999a; Peamkrasatam et al., 2006; Takeshita et al., 2006; Topcu et al., 2002; Yoshida et al., 2002). In the Middle East, the Turkish report a frequency of 23.9-69.7% (Takeshita et al., 2006; von Richter et al. (2004). The American

![Figure 2: CYP2A6 alleles reported in tobacco consumption related studies. Number of CYP2A6 alleles that have reported frequencies in population and association studies related to tobacco/nicotine consumption and cancer related to tobacco consumption (in the population study).](image-url)

The populations are grouped according to possible ancestral and/or geographical origin: African (Canadian, American, Ghanaian, Ethiopian, Namibian, “African” and “Black” populations), Caucasian (German, Canadian, American, Spanish, Finish, French, Hungarian, English, Serbian, Swedish, “Caucasian” and “Whites”), East Asian (Chinese, Korean, Japanese, Malay-
mestizo populations such as Brazilian and Ecuadorian show a frequency of 71.7% (Rossini et al., 2006) and 61.7% (Soriano et al., 2011), respectively. CYP2A6*1B frequency is higher in East and South Asian populations and is reported as 26.7-54.2% (Ariyoshi et al., 2002; Djordjevic et al., 2013; Islam et al., 2013; Ito et al., 2015; Iwahashi et al., 2004; Kwon et al., 2001; Lourembam et al., 2015; Mahavorasirikul et al., 2009; Nakajima et al., 2001, 2006; Nurfadhлина et al., 2006; Oscarson et al., 1999a; Peamkrasatam et al., 2006; Schoedel et al., 2004; Takeshita et al., 2006; Tiwawech et al., 2006; Topcu et al., 2002; Yoshida et al., 2002; Yusof and Gan, 2009). In Caucasians (Nonspecific, North American, Spanish, British, French, Swedish and Serbian) a frequency of 27.6-33.5% (Bloom et al., 2011; Djordjevic et al., 2013, 2010; Gambier et al., 2005; Haberl et al., 2005; Huang et al., 2005; Nakajima et al., 2006, 2004; Oscarson et al., 1999a; Schoedel et al., 2004; Soriano et al., 2011) is reported, but it is lower in Turkish populations (25.9-26.7% (Takeshita et al., 2006; von Richter et al., 2004)). African populations such as African American, Ghanaian and Namibian show lower frequencies (11.2-19.8% (Gyamfi et al., 2005; Ho et al., 2009; Mwenifumbo et al., 2008; Nakajima et al., 2006, 2004; Schoedel et al., 2004; Takeshita et al., 2006)), except for Ethiopian populations (31.3% (Aklillu et al., 2014)). American mestizo populations (Brazilian and Ecuadorian) have a frequency of 26.4-31.2% (Rossini et al., 2006; Soriano et al., 2011; Vasconcelos et al., 2005) while the Native Canadian and Alaskan Yupik show a frequency of 55-65.3% (Rossini et al., 2006; Soriano et al., 2011; Vasconcelos et al., 2005). CYP2A6*1D showed a higher frequency in the African nonspecific population (>50% (Nakajima et al., 2006)) than in the Ethiopian population (29.4% (Aklillu et al., 2014)). In the Caucasian population (Nonspecific, North American and Swedish) the frequency is 26.7-40% (Bloom et al., 2011; Nakajima et al., 2006; von Richter et al., 2004) and in Turkish populations it is 32.3% (von Richter et al., 2004). Asian populations (Japanese and Korean) show a frequency of 10-20% (Nakajima et al., 2006). CYP2A6*1F has been reported only in nonspecific Caucasian (1.8% (Nakajima et al., 2004)) and Turkish (2.2% (Takeshita et al., 2006)) populations, but was not found in African and Japanese populations (Nakajima et al., 2004; Takeshita et al., 2006). CYP2A6*1H has a higher frequency in African populations (North American and Namibian) (12.3-13.3% (Nakajima et al., 2004; Takeshita et al., 2006)) than in Caucasian populations (1.2% (Nakajima et al., 2004)), but it was not reported in Turkish and Japanese populations (Takeshita et al., 2006). CYP2A6*1J has been found higher in Caucasian populations such as nonspecific Caucasian (>11% (Nakajima et al., 2006)), North American (7.9% (Bloom et al., 2011)) and Swedish (3.1% (von Richter et al., 2004)), followed by nonspecific African (9.8% (Nakajima et al., 2006)), Turkish (5.2% (von Richter et al., 2004)), Japanese (4.5% (Nakajima et al., 2006)) and Korean (1.6% (Nakajima et al., 2006)). CYP2A6*1K has been reported as nonexistent in Caucasian, African, Japanese and Korean populations (Nakajima et al., 2006).

CYP2A6*2 has a higher frequency in Caucasian populations such as nonspecific Caucasian (1.1-5.3% (Audrain-McGovern et al., 2007; Benowitz et al., 2006; Haberl et al., 2005; Malaiyandi et al., 2006a, b; Nakajima et al., 2006, 2004; Paschke et al., 2001, 2000; Schoedel et al., 2004; Xu et al., 2002), English populations such as British (2.5% (Huang et al., 2005)) and Canadian (3.4% (O’Loughlin et al., 2004)); and in Central Europe populations such as German, French and Spanish it has been reported from 2.3-3% (Bourian et al., 2000; Loriot et al., 2001; O’Loughlin et al., 1999b)). In North of Europe populations such as Finish and Swedish, a frequency of 1.1-3% (Oscarson et al., 1998, 1999b) was reported. In African populations (Afro American and Ethiopian), the frequency is less than 1% (Malaiyandi et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004) and is not reported in Ghanaian populations (Gyamfi et al., 2005). The frequency
is different in American populations: American populations, Canadian natives and Alaskan Yupik have a frequency less than 1% (Binnington et al., 2012; Nowak et al., 1998; Schoedel et al., 2004). However, the mestizo populations such as Brazilian (1.6-1.7%) (Rossini et al., 2006; Vasconcelos et al., 2005) and Chilean (2.3-7%) (Cáceres et al., 2012; Roco et al., 2012) show a frequency higher than Hispanics (Benowitz et al., 2006). In East and Southeast Asian populations such as Chinese, Korean, Japanese, Malaysian and Thai, CYP2A6*2 is non-existent (Huang et al., 2005; Kitagawa et al., 1999; Malaiyandi et al., 2005; Nakajima et al., 2006; Nurfadhilina et al., 2006; Oscarson et al., 1999b; Schoedel et al., 2004). A minimum range of 1% (Nurfadhilina et al., 2006) was reported in Indians. The Iranian population had a frequency higher than 2.2% (Emamghoreishi et al., 2008; Heravi et al., 2010), unlike the Turkish population where CYP2A6*2 was not found (Cok et al., 2001).

The methodologies for the detection, function and allelic frequency of CYP2A6*3 have been controversial (Fernandez-Salguero et al., 1995; Oscarson et al., 1998; Yamano et al., 1990); however, we present the reported frequency. The highest frequency (13.8% (Nowak et al., 1998)) was reported in Canadian natives, followed by Turkish (12% (Cok et al., 2001)), German (1.4%) (Bourian et al., 2000) and Indian (1.1%) (Nurfadhilina et al., 2006) populations. However, in African (Nakajima et al., 2006; Paschke et al., 2001). Caucasian (Nakajima et al., 2006; Paschke et al., 2001), Spanish (Oscarson et al., 1999a), Chilean (Cáceres et al., 2012), Iranian (Heravi et al., 2010), Korean (Kwon et al., 2001; Nakajima et al., 2006; Yoshida et al., 2002), Chinese (Nurfadhilina et al., 2006; Oscarson et al., 1999b), Malaysian (Nurfadhilina et al., 2006) and Japanese (Nakajima et al., 2001, 2006; Yoshida et al., 2002) (except a report of 0.2% (Minematsu et al., 2003)) populations, this allele was not found.

CYP2A6*4 may be the most widely studied allele in all populations. East Asian populations such as Japanese (11.2-25.6% (Ariyoshi et al., 2002; Fujieda et al., 2004; Fukami et al., 2006; Ito et al., 2015; Iwashashi et al., 2004; Minematsu et al., 2003, 2006; Nakajima et al., 2001, 2006; Schoedel et al., 2004; Takeshita et al., 2006; Tamaki et al., 2011; Xu et al., 2002; Yoshida et al., 2002)), Chinese (4.9-14% (Gu et al., 2005; Liu et al., 2011; Nurfadhilina et al., 2006; Oscarson et al., 1999a; Schoedel et al., 2004; Song et al., 2009; Tan et al., 2001; Xu et al., 2002; Yuan et al., 2016)), Korean (9.4-11%) (Djordjevic et al., 2013; Fukami et al., 2006; Kwon et al., 2001; Nakajima et al., 2006; Yoshida et al., 2002), Thai (4-14.2% (Mahavorasirikul et al., 2009; Peamkrasatam et al., 2006; Tiwawech et al., 2006)), Malaysian (7-16.7% (Nurfadhilina et al., 2006; Yusof and Gan, 2009)) and Vietnamese (11.8% (Veiga et al., 2009)) showed the highest frequency among populations. However, some Asian populations that reside in a foreign country such as those living in the UK report a frequency of less than 1% (Benowitz et al., 2006; Huang et al., 2005). However, ethnic populations in Asian countries show no differences in the frequencies such as in Japan with the Shimane (18.2%), Tottori (16.9%), Fukoka (20.6%) and Ehime (25.9%) (Takeshita et al., 2006) and in China with the Han (7.9%), Uighur (15%), Bouyei (0%) and Tibetan (2%) (Pang et al., 2015). South Asian populations such as Bangladeshi (4.7-11.2% (Islam et al., 2013)), Sri Lankan (2.8-9.6% (Topcu et al., 2002)) and Indian (1.5-8.9% (Krishnakumar et al., 2012; Nurfadhilina et al., 2006) show a high frequency. Middle East populations such as Turkish (2.2% (Takeshita et al., 2006)) and Iranian (0.9-2.5% (Emamghoreishi et al., 2008; Heravi et al., 2010)) show a lower frequency. However, the distribution is different in Caucasian populations: Nonspecific Caucasians had a frequency lower than 3% (Audrain-McGovern et al., 2007; Benowitz et al., 2006; Fukami et al., 2006; Malaiyandi et al., 2006a, b; Nakajima et al., 2006, 2004; Rao et al., 2000; Schoedel et al., 2004; Xu et al., 2002); Atlantic Europe populations such as Spanish (0.5-4% (Oscarson et al., 1999b; Soriano et al., 2011)) and French (3.8%
also have this allele; English populations such as British (0.3 % (Huang et al., 2005)), North American (1.6 % (Bloom et al., 2011)) and Canadian (0.2 % (O’Loughlin et al., 2004)) had a lower frequency; North of Europe populations such as Finish (1 % (Oscarson et al., 1999b)) and Swedish (1.1 % (Djordjevic et al., 2013)) also report this allele. Southeastern Europe populations such as Serbian report a frequency of 2.9 % (Djordjevic et al., 2010) and Tatar from Russia reported a frequency range from 6.8-16.9 % (Korytina et al., 2014). African populations had a frequency less than 2 % and are reported as follows: Nonspecific African ≤1.9 % (Ho et al., 2009; Nakajima et al., 2006; Schoedel et al., 2004), African American (0.5-0.6 % (Fukami et al., 2006; Nakajima et al., 2004)), Ghanaian (2 % (Gyamfi et al., 2005)), Ethiopian (0.6 % (Aklillu et al., 2014)) and Namibian (0 % (Takeishi et al., 2006)). American populations show a variable frequency: Amerindian populations as Alaskan Yupik (14.5 % (Binnington et al., 2012)) have a higher frequency than Canadian natives (1 % (Schoedel et al., 2004)) and more than mestizo populations as Brazilian (0.5 % (Vasconcelos et al., 2005)), Ecuadorian (7.1 % (Soriano et al., 2011)), Chilean (3.7-4 % (Cáceres et al., 2012; Roco et al., 2012)) and Hispanics (0 % (Benowitz et al., 2006)). On the other side of the world, the Māori native population from New Zealand report a high frequency (9.6 % (Lea et al., 2008)) of this allele and it is the only population reported in Oceania.

CYP2A6*5 is found in a higher frequency in East and Southeast Asian populations such as the Vietnamese had the highest frequency (14.6 % (Veiga et al., 2009)) compared with the Chinese (Liu et al., 2011; Nurfadhlina et al., 2006; Oscarson et al., 1999b; Schoedel et al., 2004), Korean (Djordjevic et al., 2013; Kwon et al., 2001; Nakajima et al., 2006; Yoshida et al., 2002) and Malaysian (Nurfadhlina et al., 2006) populations with a frequency less than 1.5 %. However, this allele is not found in the Japanese (Nakajima et al., 2001, 2006; Schoedel et al., 2004; Yoshida et al., 2002); South Asians such as Indians show a frequency of less than 1 % (Krishnakumar et al., 2012; Nurfadhlina et al., 2006). Canadian populations such as the natives (0.5 % (Schoedel et al., 2004)) and Caucasians (0.1 % (Schoedel et al., 2004)) showed the minimum frequency. However, this allele was not found in the African, Caucasian, Middle East and American mestizo populations (Aklillu et al., 2014; Djordjevic et al., 2013, 2010; Gyamfi et al., 2005; Huang et al., 2005; Malaiyandi et al., 2006a; Nakajima et al., 2006; Oscarson et al., 1999b; Rossini et al., 2006).

CYP2A6*6 has been found only in a Japanese population study at a low frequency (0.4 % (Kitagawa et al., 2001)). It not has been found in African (Gyamfi et al., 2005; Nakajima et al., 2006), Caucasian (Malaiyandi et al., 2006a; Nakajima et al., 2006), Asian (Nakajima et al., 2006; Yoshida et al., 2002) or Canadian native populations (Schoedel et al., 2004).

CYP2A6*7 has a higher frequency in East and Southeast Asian populations such as Japanese (6.3-13 %) (Fujieda et al., 2004; Fukami et al., 2005a; Minematsu et al., 2006; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et al., 2002; Yoshida et al., 2002), Chinese (2.2-13.8 % (Liu et al., 2011; Mwenifumbo et al., 2005; Nurfadhlina et al., 2006; Schoedel et al., 2004; Xu et al., 2002; Yuan et al., 2016)), Korean (3.6-11.1 % (Djordjevic et al., 2013; Fukami et al., 2005a; Mwenifumbo et al., 2005; Nakajima et al., 2006; Yoshida et al., 2002)), Taiwanese (10 % (Mwenifumbo et al., 2005)), Thai (5-6.4 % (Mahavorasirikul et al., 2009; Peamkrasatam et al., 2006)) and Malaysian (4.3 % (Nurfadhlina et al., 2006; Yusof and Gan, 2009)) populations. It has also been reported in a lower frequency in Caucasian (≤0.3 % (Fukami, et al., 2005a; Malaiyandi et al., 2006a; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et al., 2002)) and Māori natives from New Zealand (1 % (Lea et al., 2008)). On the other hand, in Indian, African, Canadian native and Alaskan Yupik populations,
this allele is not found (Binnington et al., 2012; Fukami et al., 2005a; Gyamfi et al., 2005; Mwenifumbo et al., 2005; Nakajima et al., 2006; Nurfadhлина et al., 2006; Schoedel et al., 2004).

\textit{CYP2A6}*8 has a specific frequency in Asian populations such as Malaysian (4.2-5 \%) (Nurfadhлина et al., 2006; Yusof and Gan, 2009), Chinese (≤3.6 \%) (Nurfadhлина et al., 2006; Schoedel et al., 2004; Xu et al., 2002)), Japanese and Korean (≤2.2 \%) (Djordjevic et al., 2013; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et al., 2002; Yoshida et al., 2002)), Indian (0.9 \%) (Nurfadhлина et al., 2006), Thai (≤0.5 \%) (Mahavorasirikul et al., 2009; Pemkhrasatam et al., 2006)) and is the lowest in Taiwanese (0.2 \%) (Mwenifumbo et al., 2005)). However, it has not been found in African, Caucasian, Canadian natives and Alaskan Yupik populations (Binnington et al., 2012; Gyamfi et al., 2005; Malaiyandi et al., 2006a; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et al., 2002).

\textit{CYP2A6}*9 is the most widely studied decreased function allele. Asian populations show the highest frequency: East Asian populations (Chinese, Korean and Japanese) had a frequency between 15-20 \% (Benowitz et al., 2006; Djordjevic et al., 2013; Fujieda et al., 2004; Liu et al., 2011; Minematsu et al., 2006; Nakajima et al., 2006; Pitarka et al., 2001; Schoedel et al., 2004; Yoshida et al., 2003; Yuan et al., 2016), followed by South Asian population as Thai (12.1-20.4 \%) (Mahavorasirikul et al., 2009; Pemkhrasatam et al., 2006)) and Malaysian (10.4 \%) (Yusof and Gan, 2009)); Middle East populations as Turkish (6.9-7.2 \%) (Pitarka et al., 2001; von Richter et al., 2004), Iranian (12.4 \%) (Emamghoreishi et al., 2008) and some ethnic groups among them (Sepehr et al., 2004) such as Turkomans (14 \%), Turks (5 \%) and Zoroastrian Persian (4 \%) showed a high frequency. Mediterranean European (Spanish (Soriano et al., 2011) and Serbian (Djordjevic et al., 2010)), North European (Swedish (Djordjevic et al., 2013; Pitarka et al., 2001)) and Central European (Hungarian (Fiatal et al., 2016)) populations show a frequency range of 5-8 \%, which is the same as Caucasian (Aurdin-McGovern et al., 2007; Benowitz et al., 2006; Haberl et al., 2005; Malaiyandi et al., 2006a, b; Nakajima et al., 2006, 2004; Schoedel et al., 2004) and North American Caucasian populations (Bloom et al., 2011; O’Loughlin et al., 2004). A different pattern occurs in African populations; African and African American show a frequency of 7-10 \% (Ho et al., 2009; Mwenifumbo et al., 2008; Nakajima et al., 2006, 2004; Schoedel et al., 2004), but in populations with a more conserved African component as Ghanaian (5.7 \%) (Gasamfi et al., 2005)) and Ethiopian (2.8 \%) (Aklilu et al., 2014)) the frequency is lower. Amerindian populations such as Canadian natives (15.5 \%) (Schoedel et al., 2004)) and Alaskan Yupik (8.9 \%) (Binnington et al., 2012)) had a heterogeneous frequency, which was the same as American mestizo populations such as Brazilian (5.7 \%) (Vasconcelos et al., 2005)), Ecuadorian (10.3 \%) (Soriano et al., 2011)), Mexican (16.4 \%) (Svyryd et al., 2015)) and other Hispanics (7.1 \%) (Benowitz et al., 2006)). The only population in Oceania to report this allele is the native population Māori from New Zealand with 19 \% (Lea et al., 2008).

\textit{CYP2A6}*10 has been reported to be higher among Asian populations as Japanese (1-4.3 \%) (Fujieda et al., 2004; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et al., 2002; Yoshida et al., 2002), Chinese (0.4-4.3 \%) (Liu et al., 2011; Mwenifumbo et al., 2005; Nurfadhлина et al., 2006; Schoedel et al., 2004), Korean (0.5-4.2 \%) (Djordjevic et al., 2013; Mwenifumbo et al., 2005; Nakajima et al., 2006; Yoshida et al., 2002), Malaysian (4.3 \%) (Nurfadhлина et al., 2006), Taiwanese (4.1 \%) (Mwenifumbo et al., 2005) and Thai (1.6-2.4 \%) (Mahavorasirikul et al., 2009; Pemkhrasatam et al., 2006); and the Alaskan Yupik population with 1.9 \% (Binnington et al., 2012). However, Indian (Nurfadhлина et al., 2006), Caucasian (Malaiyandi et al., 2006a; Mwenifumbo et al., 2005; Nakajima et al., 2006; Schoedel et al., 2004; Xu et et al.,
CYP2A6*11 is not well-reported. However, it has been reported in a minimum frequency in Japanese and Korean populations at 0.5-0.7 % (Fujieda et al., 2004; Nakajima et al., 2006). The few analyzed African and Caucasian populations did not report this allele.

CYP2A6*12 has been reported at the highest frequency among the studied populations in Hispanics (3.5-4.7 % (Benowitz et al., 2006; Koontz et al., 2009)) and Mexican (3.5 % (Borrego-Soto et al., 2015)) population. The frequency decreases in Caucasians (≤3 % (Audrain-McGovern et al., 2007; Benowitz et al., 2006; Haberl et al., 2005; Koontz et al., 2009; Malaiyandi et al., 2006a, b; Nakajima et al., 2006; Schoedel et al., 2004)), Spanish (2.2 % (Oscarson et al., 2002)) and Canadian populations (1.1 % (O’Loughlin et al., 2004)). Then, in Asians, similar to in Middle East populations such as Iranians, it is reported in 1.3 % (Emamghoreishi et al., 2008), but in East Asia only Japanese had a minimum frequency (0.8 % (Nakajima et al., 2006; Schoedel et al., 2004)) while Chinese and Korean populations (Benowitz et al., 2006; Koontz et al., 2009; Nakajima et al., 2006; Oscarson et al., 2002; Schoedel et al., 2004) did not report any frequency of the allele. Lastly, Amerindian populations such as Canadian natives and Alaskan Yupik reported a very low frequency (0.4-0.5 % (Binnington et al., 2012; Schoedel et al., 2004)). In African populations, some studies show a low frequency (0.4 %) (Ho et al., 2009; Schoedel et al., 2004) and others do not show any frequency (Benowitz et al., 2006; Koontz et al., 2009; Mwenifumbo et al., 2008; Nakajima et al., 2006).

CYP2A6*13 has been reported in Japanese (1.1-1.5 % (Kiyotani et al., 2002; Nakajima et al., 2006)) and Koreans (0.2 % (Nakajima et al., 2006)). It was not found in Caucasians and Africans (Kiyotani et al., 2002; Nakajima et al., 2006).

CYP2A6*14 has a higher frequency in Caucasian (3.5-5.2 % (Haberl et al., 2005; Kiyotani et al., 2002; Nakajima et al., 2006)) than in African (0.9-1.4 % (Mwenifumbo et al., 2008; Nakajima et al., 2006)), but was not reported in Asians (Japanese and Korean) (Kiyotani et al., 2002; Nakajima et al., 2006).

CYP2A6*15 has been reported at a minimum frequency in Japanese (1.5-2.2 % (Kiyotani et al., 2002; Nakajima et al., 2006)) and Korean (1.2 % (Nakajima et al., 2006)), but not in Caucasian and African populations (Kiyotani et al., 2002; Mwenifumbo et al., 2008; Nakajima et al., 2006).

CYP2A6*16 has been reported in Caucasian (0.3-3.6 %) and African (0-1.7 %) populations (Kiyotani et al., 2002; Mwenifumbo et al., 2008; Nakajima et al., 2006), but not in Japanese or Korean populations (Kiyotani et al., 2002; Nakajima et al., 2006).

CYP2A6*17 has a frequency of approximately 7.3-10.5 % in African populations (Fukami et al., 2004; Ho et al., 2009; Mwenifumbo et al., 2008; Nakajima et al., 2006), but not in Caucasian, Korean, Japanese or Alaskan Yupik (Binnington et al., 2012; Fukami et al., 2004; Nakajima et al., 2006) populations.

CYP2A6*18 has a higher frequency range in Caucasian (0.3-2.2 % (Fukami et al., 2005a; Haberl et al., 2005; Nakajima et al., 2006)) than in Korean (0.3-0.5 % (Fukami et al., 2005a; Nakajima et al., 2006)), but has not been found in Japanese and African populations (Fukami et al., 2005a; Nakajima et al., 2006).

CYP2A6*19 has been reported in a low frequency in Korean (1-1.4 % (Djordjevic et al., 2013; Fukami et al., 2005a; Nakajima et al., 2006)) and Japanese (0.5 %), but not in Caucasian nor African (Fukami et al., 2005a; Nakajima et al., 2006) populations.

CYP2A6*20 is present in African population at a frequency range between 1-1.7 % (Fukami et al., 2005b; Ho et al., 2009; Mwenifumbo et al., 2008; Nakajima et al., 2006). This allele was not reported in Caucasian, Japanese and Korean populations (Fukami et al., 2005b; Nakajima et al., 2006).
CYP2A6*21 has a higher frequency in Caucasian (0.5-2.3 %) (Al Koudsi et al., 2006; Haberl et al., 2005; Nakajima et al., 2006) than in African populations (0.6-0.7 % (Mwenifumbo et al., 2008; Nakajima et al., 2006)). It has not been reported in Japanese or Korean (Nakajima et al., 2006) populations.

CYP2A6*22 has been reported at frequency less than 0.3 % (Haberl et al., 2005; Nakajima et al., 2006) in Caucasian populations, but not in Japanese or Korean (Nakajima et al., 2006) populations.

CYP2A6*23 has a frequency range among 1-2 % (Ho et al., 2009, 2008; Mwenifumbo et al., 2008) in African populations, but was not reported in Caucasian, Japanese or Chinese populations (Ho et al., 2008).

CYP2A6*24 has a frequency range among 0.7-1.3 % in African populations (Al Koudsi et al., 2010; Ho et al., 2009; Mwenifumbo et al., 2008), but it has not been found in Caucasian, Chinese, Japanese and Taiwanese (Al Koudsi et al., 2010) populations.

CYP2A6*25-*28 has only been reported in African populations, while CYP2A6*26-27 has a frequency range of 0.7-0.9 % (Ho et al., 2009; Mwenifumbo et al., 2008) and CYP2A6*28 from 0.9-2.4 % (Ho et al., 2009; Mwenifumbo et al., 2008).

CYP2A6*31, *34 and *38 have not been studied in any population.

CYP2A6*35 has been reported at a frequency between 2.5-2.9 % (Al Koudsi et al., 2010; Ho et al., 2009) in African populations living in North American countries. In East Asian populations, such as Chinese, Japanese and Taiwanese, it has been found at a frequency of 0.5-0.8 % (Al Koudsi et al., 2010), but it has not been found in Caucasian or Alaskan Yupik (Al Koudsi et al., 2010; Binnington et al., 2012) populations.

CYP2A6*36 and *37 have been reported in Taiwanese populations at a frequency of 0.3 %, but not in African, Caucasian, Chinese nor Japanese (Al Koudsi et al., 2010) populations.

CYP2A6*39 (0.6 %), *40 (0.2 %), *41 (1.2 %), *42 (0.2 %), *43 (0.2 %), *44 (0.2 %) and *45 (0.6 %) have only been reported in the African population (Piliguian et al., 2014).

The gene duplication, CYP2A6*1X2, has been found at higher frequency in Asian populations as Asian (7.1 % (Benowitz et al., 2006)), Indian (3.5 % (Nurfadhлина et al., 2006)), Chinese (0.4-1.5 %) (Nurfadhлина et al., 2006; Schoedel et al., 2004; Xu et al., 2002), Korean (≤0.2 %) (Fukami et al., 2007; Nakajima et al., 2006) and Malaysian (0.4 %) (Nurfadhлина et al., 2006); the Hispanic population (3.5 %) (Benowitz et al., 2006) has a higher frequency than the Ecuadorian population (0.5 %) (Soriano et al., 2011); Caucasian populations have a frequency less than 2 % as in some Caucasian groups (≤1.7 %) (Benowitz et al., 2006; Fukami et al., 2007; Nakajima et al., 2006; Rao et al., 2000; Schoedel et al., 2004; Xu et al., 2002), Spanish (1.2 %) (Soriano et al., 2011), Swedish (0.8 %) (Djordjevic et al., 2013), Serbian (0.4 %) (Djordjevic et al., 2010), North American (0.3 %) (Bloom et al., 2011) and Canadian (0.2 %) (O’Loughlin et al., 2004).

In African populations, such African American (1.7 % (Fukami et al., 2007)) and Ethiopian (0.3 % (Aklillu et al., 2014)), this allele was present, but was not found in other African groups, Canadian natives or Alaskan Yupik (Binnington et al., 2012; Nakajima et al., 2006; Schoedel et al., 2004).

Most studies address the majority of a country’s population as a general population. However, a few studies focus on more specific population classifications as ethnic and regional groups. In the Asian population, it was studied among Tottori, Shimane, Ehime and Fukuoka people of the respective districts of Yonago, Izumo, Matsuyama and Kurume located in Japan (Takeshita et al., 2006). In China, it has been compared in the prevailing Han Chinese group and the Uighur, Bouyei and Tibetan ethnic groups (Pang et al., 2015). In the South of India, the frequencies of the people from Andhra Pradesh, Karnataka, Kerala and Tamil Nadu regions (Krishnakumar et al., 2012) were compared. In Iran, some eth-
nic groups such as Turkomans from the Golestan Province, Turks from the Ardabil Province and Zoroastrian Persians from Tehran (Sepehr et al., 2004) were studied. In Russia, the Tatar ethnic group was reported (Korytina et al., 2014). In African populations, the ethnic group Ovambo from Namibia was reported (Takeshita et al., 2006); and some ethnic groups such as Akan, Guan, Ewe, Ga, Nzima and Dargarti, but there were a few participants that were added in a single Ghanaian population (Gyamfi et al., 2005). In Oceania, the Māori ethnic group from New Zealand is the only population reported across the continent (Lea et al., 2008). In the American continent, the ethnic groups such as Yupik from Alaska and Canadian natives have been reported (Binnington et al., 2012; Nowak et al., 1998; Schoedel et al., 2004).

CONCLUSIONS AND PERSPECTIVES

The enzyme responsible for metabolizing nicotine is mostly encoded by the CYP2A6 gene, which is highly polymorphic. This variability is due to changes in DNA, which have generated different responses to nicotine and are reflected in the individual smoking behaviors along with other factors. It has been reported that this variability has been generated and distributed over a long time in different human populations, showing well defined ethnic patterns. Although association studies between carriers of certain variants and different smoking behaviors have been numerous with plausible results, population studies reporting frequencies of these variants are few. General population studies exhibit most reliable information than association studies with a variable frequency, because the population requirements are usually more specific, creating a population bias. However, it would be advisable to address this type of methodology in higher risk populations of smoking and those where policies to control smoking are less efficient, and where more smoking-related diseases are reported in the population. Additionally, the population must be characterized by more specific requirements, such as including the ancestry informative markers and avoiding "self-reporting" as the unique classification criteria. While there is a Human CYP-Allele Nomenclature Database in which the genetic findings of CYP2A6 are unified, it is necessary to supplement it with updated data as per the population distribution. All of this could contribute to a personalized system that could detect, prevent and treat populations at risk of smoking, and in consequence, avoid tobacco consumption related diseases.

REFERENCES

Aklilu E, Djordjevic N, Carrillo JA, Makonnen E, Bertilsson L, Ingelman-Sundberg M. High CYP2A6 enzyme activity as measured by a caffeine test and unique distribution of CYP2A6 variant alleles in Ethiopian population. OMICS. 2014;18:446–53.


