SMOKING AND NICOTINE ALTER UGT1A EXPRESSION

A THESIS SUBMITTED TO
THE DEPARTMENT OF MOLECULAR BIOLOGY AND GENETICS AND
GRADUATE SCHOOL OF ENGINEERING AND SCIENCE OF
BILKENT UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE
OF MASTER OF SCIENCE

By Gizem ÖLMEZER

August 2011

To my one and only sister,

Sinem

ABSTRACT

SMOKING AND NICOTINE ALTER UGT1A EXPRESSION

Gizem ÖLMEZER

MSc. in Molecular Biology and Genetics

Supervisor: Assist. Prof. Dr. Özlen KONU

August 2011

The expression and activity of enzymes taking role in drug metabolism are important as in the case of phase II glucuronidation enzymes; namely UDP-glucuronosyltransferases (UGTs). Previously, it has been identified that smoking upregulates the expression of UGT enzymes in oral mucosa. We asked whether smoking induces UGT1A expression in other tissues and re-analyzed publically available datasets run with samples from smokers and non-smokers. It was observed that UGT1A enzymes were overexpressed in several types of epithelial cells of smokers.

30% of nicotine metabolism is performed by UGT enzymes; however, whether UGT1A expression is modulated by nicotine, the addictive component of tobacco smoke, is not known. For this purpose, the expression levels of UGT1A isoforms were measured using Real-Time PCR in nicotine treated SW620 colorectal cancer cells. Our findings showed that nicotine's effect on UGT1A expression was isoform specific; and the magnitude of modulation differed among isoforms. Furthermore, the upregulation of UGT1A enzymes could only be observed in serum-deprived SW620 cells. In summary, nicotine metabolism enzymes are regulated by both smoking *in vivo* and nicotine *in vitro*. Nevertheless, enhanced xenobiotic metabolism may result in chemoresistance, which is undesirable for cancer patients. Therefore, before drug therapy cancer patients might be analyzed in terms of their smoking status and UGT1A expression patterns.

Keywords: Nicotine, Smoking, UDP-glucuronosyltransferases (UGTs), Serum Starvation, Colon Cancer

ÖZ

SİGARA VE NİKOTİN UGT1A GEN İFADESİNİ DEĞİŞTİRİYOR

Gizem ÖLMEZER

Moleküler Biyoloji ve Genetik Yüksek Lisansı

Tez danışmanı: Yrd. Doç. Dr. Özlen KONU

August 2011

İlaç metabolizmasında görev alan genlerin ifade ve aktiviteleri, ilaçların etkilerini değerlendirmekte büyük önem taşır. UGT genleri faz II metabolizmasında yer alan enzimlerdir ve sağlıklı bireylerin ağız mukozasında sigara içimiyle ifadelerinin arttığı gösterilmiştir. Bu çalışmada, sigaranın UGT1A gen ifadeleri üzerindeki etkileri başka dokularda da geçerli midir sorusu sorulmuştur. Bu amaçla, GEO veritabanında yer alan mikrodizin çalışmaları tekrar analiz edilmiş ve UGT1A genlerinin birçok epitel dokuda sigara içimiyle arttığı gösterilmiştir.

Nikotin metabolizmasının %30'u UGT genleri tarafından yürütülmekte; fakat, gen düzenlenmesinde nikotinin rolü olup olmadığı bilinmemektedir. Bu amaçla UGT1A genlerinin ifadeleri izoform olarak mRNA seviyesinde ve Gerçek Zamanlı Polimeraz Zincir Reaksiyonu ile nikotin ile muamele edilmiş SW620 kolorektal kanser hücrelerinde tespit edilmiştir. Sonuçlar, UGT1A genlerinin serumdan uzak büyütülen SW620 hücrelerinde izoforma bağlı olarak nikotinle beraber arttığını ve bu artışın her izoformda farklı olduğunu göstermiştir. Özet olarak, sigara ve nikotin UGT1A gen ifadesini düzenlemektedir. Zenobiyotik metabolizmasının düzeni vücuda alınan her ilaca verilen cevabı etkilediği için, bu mekanizmaların anlaşılması büyük önem taşımaktadır. İlaç tedavisinden önce hastanın sigara alışkanlığı ve UGT1A izoform ifadesi özellikleri açısından analiz edilebillir.

Anahtar sözcükler: Nikotin, UGT enzimleri, Sigara, Kolon kanseri

ACKNOWLEDGEMENTS

I would like to convey my thanks to Assist. Prof. Dr.Özlen Konu for her supervision and extensive knowledge in bioinformatics. I am also thankful to her because she set a priority as me learning as much as I can.

I would like to thank Assist. Prof. Dr. Ali Osmay Güre, who has supported me since the first day I stepped into Bilkent University. I would like to thank Assist. Prof. Dr. Michelle Adams for her valuable comments on the thesis. I would also like to thank Assist. Prof. Dr.İhsan Gürsel for his moral support at all times.

I am especially indebted to Ceren Sucularlı, Muammer Üçal from whom I learned a lot and had an invaluable time with. In addition, Ahmet Raşit Öztürk, Melike Öztürk and Onur Kaya never let me feel alone, which I am thankful for.

I would like to thank Nilüfer Sayar, without who this project cannot reach to this point. She tried her best to be there when I needed her and made my days precious. Gurbet Karahan, Sinem Özcan, Şükrü Atakan and Ece Akhan, who have been very special friends for me, are among those I would like to thank to, as well.

My special thanks are for my fiancee, İsmail Tanyeli and his family. They took me under their wings when all I needed was a shelter.
I am very grateful to my family, who raised me with unconditional love and respect. They supported me at every step I take and did their best for my education.

I would like to thank TÜBİTAK for supporting me with BIDEB-2210 scholarship during my M.Sc. research period.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Glucuronidation

In order to maintain its stability and homeostasis, a cell must eliminate non-membrane associated substrates originating both endo- and exogenously. For this purpose, in mammals, many chemical defense pathways have evolved, one of which is glucuronidation [1]. Glucuronidation is a process of conjugating glucuronic acid to the lipid-soluble toxics; this results in formation of water-soluble "glucuronides", which, then, can be easily transported to excretory system. This crucial function has been performed by Endoplasmic Reticulum (ER) resident enzymes, UDP-glucuronosyltransferases (UGTs) [2].

UGT enzymes catalyze glucuronidation of a variety of detrimental substances for removal from the body. Among the lipid-soluble substrates of UGTs, there are bile acids, bilirubin, hormones, steroids, free fatty acids and xenobiotics such as therapeutic drugs, environmental pollutants, and carcinogens; including nicotine and tobacco smoke. Since almost all of the glucoronides generated are free of any biological activity in the body, the process of glucuronidation is regarded as an essential part of detoxification; responsible for the elimination of one-tenth of all drugs [3]. As an example for emphasizing their necessity, genetic mutation of *UGT1A1* isoform results in jaundice, a disease caused by increased levels of bilirubin in blood [4].

1.2 UGT Superfamily

In mammals, there are two UGT subfamilies that belong to UGT Superfamily: UGT1A, located on chromosome 2q37 [5] and UGT2, found on chromosome 4q13 [6]. There are 50% sequence similarity among the proteins encoded by these two families while the similarity in enzymes belonging to each family reaches up to 60% [7].

Human UGT1A locus contains 9 functional genes, spanning a region of 200kb (**Figure 1.1**). Each gene has a variable Exon 1 while sharing the remaining four exons, Exons 2-5 [8]. These variable exons are positioned tandemly in the genome and are followed by the common exons.

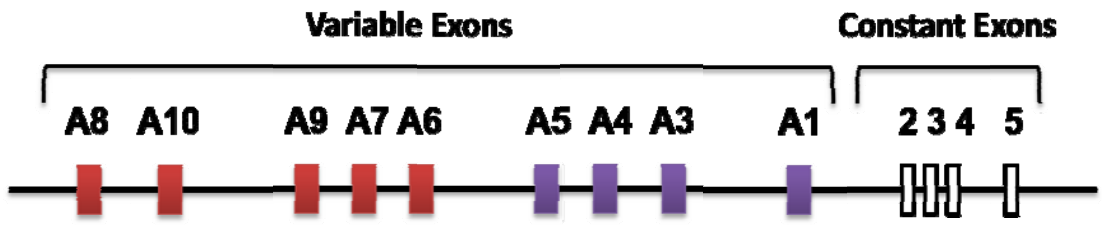

Figure 1.1 Genomic organization of human UGT1A locus on chromosome 2. The red boxes indicate phenol-type *UGT1A* variable exons while purple boxes demonstrate bilirubin-type *UGT1A* variable exons. Constant exons are shown as white boxes.

Based on the sequence similarity in Exon I, UGT1A superfamily is divided into two groups; namely, phenol-type and bilirubin type UGT1A exons. UGT1A1-5 belong to bilirubin-type group while phenol-type group involves UGT1A7-10 [9].

Constant exon of UGT1A family (Exon 2-5) encodes 245 amino acids at the C terminal, comprising the UDP-binding domains, which are identical in all UGT1A proteins [8,10]. On the other hand, aglycone-binding domain found at the N-terminal end (260 amino acids) is encoded uniquely by each UGT1A isoform, suggested to be specifying their substrates [11-12]. Nevertheless, the exact principles determining the substrate specificity is not revealed yet.

UGT2 subfamily, on the other hand, bears multiple gene loci and encodes 8 functional proteins; namely, UGT2A1, UGT2B4, UGT2B7, UGT2B10, UGT2B11, UGT2B15, UGT2B17, and UGT2B28 [13]. Six exons comprise a functional gene in this subfamily.

1.3 Regulation of UGT1A Gene Expression

What is striking about genomic organization of UGT1A gene cluster is that there are multiple exon cassettes tandemly located and that show sequence similarity preceding a number of constant exons [8]. The phenomenon of multiple variable exons has firstly been identified in gene family of neural protocadherins (Pcdh), expressed in a cell-specific manner [14]. Combining different exons with the same exon cassettes can be named as an alternative splicing event, which promotes tissue and cell-specific gene expression. Indeed, it is logical to think that genes expressed in

mammalian central nervous system, which is a quite sophisticated network, may require such a genomic organization to be differentially expressed in a cell and developmental-stage specific manner.

Zhang *et al.* (2004)[14] has conducted a genome-wide research for the determination of gene clusters with similar genomic organization, bringing up UDP glucuronosyltransferases (UGT1), plectin, neuronal nitric oxide synthase (NOS1), and glucocorticoid receptor (GR) gene clusters with more than 10 alternative exons.

In UGT1A gene cluster, each variable exon shares common motifs in their coding areas [15] while regulatory regions at 5'end diverge from each other. Accordingly, Tukey and Strassburg (2000)[16] has proposed that formation of functional UGT1A isoforms requires the binding of RNA Polymerase II to the corresponding promoter regions located at 5'flanking sequence of each variable Exon I. Then, each Exon I is independently spliced into the upcoming constant exon, generating a distinct UGT1A isoform. Indeed, Zhang *et al*. has shown the presence of CpG islands at 5'ends of each exon [14].

According to this, Zhang *et al.* (2004)[14] speculates that alternating promoters of individual Exon 1 of UGT1A1 isoforms make the generation of divergent isoforms possible. In addition, independent promoter activation of each exon enables the cell to respond to different environmental and homeostatic stimuli. Moreover, each Exon 1, having a distinct regulatory region, may change the stability of the corresponding mRNA and affect its posttranscriptional modifications [14].

1.4 Conservation of UGT Superfamily among Species

To what extent the UGT genes are conserved is very well identified in a study by Zhang *et al*. (2004)[14]. A sequence analysis via VISTA shows that both the variable and common exons are highly conserved among human, mouse and rat UGT gene clusters [14]. Nevertheless, no conservation of intergenic sequences could be observed, except the unique intronic region between Exon 1 and 2.

What is also conserved between mouse, human and rat is the unique genomic organization of the cluster. However, the number of the variable exons shows difference between the species.

The coding common exons of human UGT1 isoforms show 90% amino acid sequence similarity between three species, in addition to full similarity in length of the exons [14].

The classification of variable exons as bilirubin type and phenol type is also conserved in rat. In human, bilirubin type exons show an approximately 69% amino acid similarity while the sequence identity of amino acids is 73% in the phenol type exons [14].

In case of human UGT1A cluster, bilirubin type and phenol type exons appear to have duplicated from two variable exons. In addition, these ancestral exons are expected to have duplicated from the one common exon [14].

There is an extensive study focusing on Ugt superfamily in zebrafish [17]. This study showed that there are three subfamilies in this species; namely, Ugt1, Ugt2 and Ugt5. Ugt1 and Ugt2 clusters have the same genomic organization as their counterparts in human: Each variable Exon 1 is independently conjugated with constant exons.

Phylogenetic comparisons in report of Huang and Wu (2010)[17] proposed that the ancestral Ugt1 and Ugt2 clusters have gone through a lineage-specific gene loss and duplication events resulting in "zebrafish" specific isozymes.

Ugt5 subfamily, on the other hand, is a novel subfamily appearing specifically in teleosts and amphibians. This novel cluster is thought to have formed by retrotransposition preceding gene duplication [17].

1.5 Tissue Specific Expression

As the main resort of detoxification in the body, UGTs are mainly expressed in liver [2,18]. Nevertheless, there are certain UGT isoforms (UGT1A7, 1A8, 1A10) whose expression could only be detected in human gastrointestinal and kidney tissues as they contribute to drug metabolism, as well [19].

In addition, Nakamura *et al.* (2008)[20] has precisely identified mRNA expression of UGT isoforms in human lung, bladder and steroid-related tissues in an isoform-specific manner. In addition, the same study has demonstrated the isoformic expression in certain human cancer cell lines, showing unexpected overexpression or suppression. Consequently, each isoform will be discussed in terms of expression independently, as follows:

1.5.1 UGT1A1

Formerly, UGT1A1 isoform was detected to be expressed in liver in high amounts [21]. In addition, it was determined that it is expressed in other gastrointestinal organs; namely, bile ducts, small intestine, stomach and colon as well as kidney and testis [20,22-23].

Nakamura *et al.* (2008)[20] has conducted a study where various human cancer cell lines were analyzed in terms of expression of the UGT1A isoforms. This study has additionally showed that 1A1 isoform was expressed in human liver cancer cell lines HepG2 and Huh7, and human colon cancer cell lines, Caco-2 and LS180, in high quantities. UGT1A1 expression was low in the human kidney cancer cell line HK-2.

Particularly interesting was that high amounts of UGT1A1 mRNA was detected also in human breast cancer cell line MCF-7, and human ovary cancer cell line OMC-3. In addition, 1A1 expression was present in low amounts in human breast cancer cell line MDA-MB-435, ovary carcinoma cell line, and human adrenocortical carcinoma cell line H295R even though it is not expressed in the normal adrenal gland tissue, or breast and ovary [20].

1.5.2 UGT1A3

The expression pattern of UGT1A3 is quite similar to that of UGT1A1 isoform. Both are expressed in detoxifying organs; such as, liver, small intestine, colon, bile ducts, stomach and bladder [8,24-25]. Moreover, UGT1A3 was detected in the same cancer tissues as in UGT1A1 [20]. Nevertheless, expression of UGT1A3 is relatively lower, implying the possibility that UGT1A1 can compensate for UGT1A3.

1.5.3 UGT1A4

UGT1A4 is an isoform, expressed highly in many tissues. It is a hepatic isoform besides being expressed in colon and bile ducts [21,23]. In addition, Kaivosaari *et al.* (2007) [26] encountered its slight expression in trachea, kidney and small intestine. Nakamura then added bladder and ovary into this list, as well [20].

In case of cancer cells, UGT1A4 also shows a similar expression to that of UGT1A1 [20].

1.5.4 UGT1A5

This unique isoform is different from all others in that its expression could not be detected in most of the tissues for a long time [21,27]. Then, in 2002, Collier *et al.)*[28] has detected the mRNA transcripts of UGT1A5 in placenta. Afterwards, in 2005 Finel *et al.*[29] has proved the presence of a high interindividual variability as in the case of UGT1A1, causing a discrepancy in the results. Moreover, Nakamura *et al* (2008)[20] has shown its presence in gastrointestinal tract, kidney, bladder, and uterus and UGT1A5 was marginally detected in the other tissues, as well [20]. However, it is important to note that it is not expressed in liver.

When it comes the carcinoma tissues analyzed in Nakamura's study[20], UGT1A5 was highly expressed in colon cancer cell line LS180, breast cancer cell lines MDA-MB-435 and MCF-7, and ovarian cancer cell line OMC-3 [20].

1.5.5 UGT1A6

mRNA transcript of UGT1A6 isoform was detected in most of the detoxifying tissues analyzed. The normal tissues in which UGT1A6 is expressed can be given as: liver, stomach, biliary tissue, small intestine, colon, kidney and bladder [24,30].

In addition, UGT1A6 expression was found in liver cancer cell lines HepG2 and HuH7, renal carcinoma cell line HK-2, colon cancer cells Caco-2 and LS180 [20]. Furthermore, lung carcinoma cell line A549 has constitutive UGT1A6 expression [31]. Interestingly, even though UGT1A6 is not expressed in normal steroid-related tissues, it is found in the carcinoma cell lines related with those tissues, such as,

adrenal carcinoma cell line H295R, breast carcinoma cell line MCF-7, uterus cancer cell line HeLa, adrenal gland carcinoma cell line Ishikawa, and ovarian cancer cell line OMC-3 [20]. This deregulated expression can stem from the fact that UGT1A6 expression is regulated by hormones and ligands activating pregnane X receptor (PXR) and constitutive androstone receptor (CAR) [31-32].

1.5.6 UGT1A7

The expression of UGT1A7 isoform can be defined as extrahepatic. In more detail, UGT1A7 was identified in esophagus [27] and stomach [22].

In addition, it is stated that UGT1A7 is not expressed in colon in the reports of Strassburg *et al.* (1999) [27]. Nakamura *et al.*(2008)[20], on the other hand, additionally detected the presence UGT1A7 isoform in colon as well as small intestine, kidney and bladder [20].

The cancer cells in which UGT1A7 expression was observed are as follows: liver cancer cell line HuH7, renal carcinoma cell line HK-2, colon cancer cells Caco-2 and LS180, breast carcinoma cell line MCF-7, ovarian cancer cell line OMC-3 [20].

1.5.7 UGT1A8

UGT1A8 is again an extrahepatic isoform [16,33] but it is expressed in the small intestine, colon, kidney, bladder and trachea [20,34]. Its expression was found in jejunum and ileum, as well [33]. Additionally, it was identified in esophagus [27]. Moreover, the cancer cells showing high expression of UGT1A8 are LS180, H295R, MDA-MB-435, MCF-7, and OMC-3 cells [20].

1.5.8 UGT1A9

Even though UGT1A9 belongs to the group of phenol-type UGT1A isoforms, members of which are not expressed in liver, this isoform shows a high expression in liver [10]. In addition, colon is a tissue with UGT1A9 expression [21]. Additionally, it is expressed in esophagus [27]. However, where UGT1A9 abundantly expressed besides liver is kidney [35]. Expectedly, the carcinoma cell line with UGT1A9 expression is Hk-2 cells, derived from kidney [20].

Ohno and Nakajin (2009)[34]claim that mRNA of UGT1A9 is also present in adrenal glands, small intestine, and colon; while Nakamura *et al.* (2008)[20] adds to this list, bladder and testes [20,34].

1.5.9 UGT1A10

Initially, UGT1A10 isoform was identified in certain members of gastrointestinal tract; such as, stomach, intestine and colon, in addition in biliary epithelium, but not in liver [21,36-37]. Then, it also was found in esophagus [27].

The expression of UGT1A10 was recently found in liver, kidney, ovary, and uterus even though it is extremely low [20].

UGT1A10 was found to be highly expressed in LS180 and MCF-7 cells, in addition to H295R, MDAMB-435, HuH7, and OMC-3 cells [20].

1.6 Nicotine

Nicotine is a pyridine alkaloid. When administered, it increases the dopamine levels in the brain, triggering an addiction [38]. Possessing this ability, it is defined as the addictive compound in cigarettes.

Nicotine is absorbed from the alveoli of lungs very rapidly and at the cessation of smoking, the blood concentrations of nicotine reaches to its highest level [39-40]. Total nicotine clearance from body takes approximately 2 hours.

1.7 Nicotine Metabolism

Nicotine removal from body is mainly carried out in liver by the enzymes, CYP2A6, UDP-glucuronosyltransferases (UGT), and flavin-containing monooxygenase 3 (FMO3) [38].

In mammals, most of nicotine (approximately 75%) is metabolized to cotinine by the action of CYP2A6. Moreover, cotinine is further metabolized to trans-3′-hydroxycotinine (3HC) and less-harmful side-products again by the oxidation reactions catalyzed by CYP2A6 enzymes [41].

Additionally, nicotine is metabolized by UGT enzymes, which constitutes 3-5% of nicotine metabolism. This reaction generates an *N*-quaternary glucuronide and enables its excretion through urinary system. In addition, other major metabolites of nicotine, which are cotinine and trans-3′-hydroxycotinine (3HC), are subjected to glucuronidation by UGTs, as well. Overall, 31% of all nicotine metabolites found in urine are composed of phase II glucuronidated compounds depending on the activity of ER-resident UGT enzymes [42]. Eventually, the glucuronidated nicotine metabolites are removed from the body by urinary system. 8-10% of nicotine administered remains as unchanged and excreted in urine, as well.

Among the UGT1A isoforms, a potential role for the isoforms UGT1A4 and 1A9 has been suggesting for the glucuronidation of nicotine and its abundant metabolite, Cotinine *in vitro* [43].

1.8 Real-Time PCR Analysis

Each gene has its own unique pattern of expression, varying from cell to cell or as a response to differing physiological conditions [44]. It is of importance to analyze the gene expressions in molecular biology; for example, when assessing the response of cells to external/internal stimuli, to different developmental stages or for diagnostic purposes [44].

Reverse Transcription of whole cellular mRNA content, followed by PCR has been a robust and reliable technique for gene expression analyses since 1993 [45]. In addition, the advent of Real-Time PCR enhanced this protocol and enabled collection of data during amplification by means of fluorescence emission [46].

PCR reaction can be divided into four distinct stages [47-48]. Firstly, linear ground stage takes place in which PCR starts and fluorescence emission is not high enough to be detected. This is followed by the stage at which exponential amplification takes place and fluorescence emission exceeds a certain threshold cycle (also known as Ct). The threshold cycle enables us to determine the amount of starting DNA template since the more the template, the sooner the Ct is reached [49]. The third stage is when log-linear amplification of the products occurs and a steep increase in fluorescence is

observed. Eventually, reaction components become scarce, exponential accumulation of the products cease and a plateau phase is reached.

In order to collect data during amplification, there are two well-established methods used, i.e., gene-specific probes and double strand-binding agents [50]. Probe-based systems rely on the 5'-3' nuclease activity of *Taq* polymerase and after they recognize and bind their target on the single strand, polymerase causes their hydrolysis during elongation and a consequent fluorescence emission [46].

DNA binding dyes are used based on their ability to emit fluorescence when they intercalate with double strand DNA. As the amplicons accumulate, fluorescence intensity is increased proportionally [51]. Generally, SYBR Green or Ethidium Bromide is preferred for this purpose [52]. In cases when sequence specific, intron-spanning primer couples are used, SYBR Green has been shown to work quite efficiently [45]. However, since these dyes fail to specifically recognize the target amplicons, the false positive rate might be high [53]. In order to prevent this, PCR products should be run on gel [54] and dissociation curves should be checked whether there are unrelated melting peaks [55].

Real-time PCR data can be given in several reporting methods, including absolute quantification and relative quantification. Absolute quantification reports the results by relating Ct values with input copy number linearly via a calibration curve. The calibration curves are plotted using known concentrations of DNA, e.g., recombinant DNA, genomic DNA or RT-PCR product [56]. This method is based on the assumption that all standards and samples bear approximately equal efficiencies [57].

Relative quantification, on the other hand, as the name implies, presents the gene expression levels relative to an internal standard [58]. The internal references are generally house-keeping/reference genes, which show steady-state expression in all samples [45]. There are numerous mathematical methods designed to assess the relative expression of genes to an endogenous reference. An established mathematical model is presented in the paper of Pfaffl *et al.* (2001)[59]. This method calculates the gene expressions as a relative fold change between target genes and reference, including the efficiency values of each primer. The fold change is calculated as follows:

Fold change = $(E_{target})^{\Delta Ct_{Target\ (control-sample)}} / (E_{reference})^{\Delta Ct_{Reference\ (control-sample)}}$

Here, the formula assumes that Ct value of the reference gene does not show deviation among control and sample; thus, ΔCtReference is set as 0. The formula requires an efficiency value for each primer couple together with the enzyme used. Ideally, the amplification efficiency in each reaction is assumed to be 1, implying that the amplicon concentration doubles itself in each cycle at log phase [60]. Nevertheless, in actuality, every reaction has its own efficiency value and the failure in calculating that may overestimate the results [61].

In order to calculate the efficiency of the primers, a widely used formula is as follows:

$E = \text{Magnitude of dilution}^{[-1/\text{slope of the dilution curve}]}$

The slopes are derived from the calibration curve that is plotted against serially diluted samples and the corresponding Ct values run with the same enzyme and primer couple. Eventually, the statistical significance of the results should be determined by standard parametric tests; such as ANOVA and t-tests [62].

1.9 *SW620* Colon Adenocarcinoma Cell Line as a Model System

SW620 colon adenocarcinoma cell line is a metastatic form of a primary adenocarcinoma cell line SW480, which was initially isolated from a 51-year old Caucasian male [63]. Within a year, a metastasis was observed from the colon to abdomens and SW620 cell line was isolated from a lymph node. SW620 cells were shown to be highly tumorigenic when injected to nude mice [64].

This cell line is a good model for studying UGT1A expression since colon is the second organ where UGTs are expressed most. In addition, there is a study displaying that Pregnane X Receptor (PXR), a UGT1A expression regulator, responds to the drug irinotecan in SW620 cells [65]. Irinotecan is a widely used drug in treatments of colon cancer. When SW620 cells are treated with irinotecan, the expression of PXR as well as UGT1A1, UGT1A9 and UGT1A10 are upregulated. Thus, SW620 establishes a good model for analyzing the effect of nicotine on colon cancer cells in terms of UGT1A expression.

2. AIM OF THE STUDY

Chemical defense pathways are evolved to protect cells from toxic xenobiotics. It is especially of critical importance in organs exposed to xenobiotics most. There are two detoxification routes for biochemical modification of the compounds, namely, phase I drug metabolism involving oxidation/reduction and hydrolysis reactions and phase II drug metabolism indicating conjugation [2]. Glucuronidation, as a phase II drug metabolic pathway, is responsible for conjugating glucuronic acid to its numerous lipophilic substrates, converting them into hydrophilic molecules. In cell, glucuronidation is predominantly catalyzed by a family of UDP-glucosyltransferases (UGTs) [2,66-68]. UGTs take a central role in elimination of one tenth of all drugs [3]. That's why it is important to understand how their expression changes in response to which drug.

Each year, over 1 million people are expected to develop colorectal cancer [69]. In addition to genetic factors, the life style of the individual affects the possibility of the advent of the disease [70]. Colon acts as the primary barrier that initially and directly contacts with the any toxic and carcinogenic compound that we are daily exposed to in our lives (by smoking, food, medication, etc.). That's why; colonic cells should bear effective defense pathways against them, including glucuronidation.

Liver is classified as the main resort for the elimination of toxic substances in the body, originating either endo- or exogenously [2,18]. Nevertheless, the metabolism of a drug, when administered orally, starts at the epithelial cells of the gastrointestinal (GI) tract, which is called first-pass metabolism [71]. The epithelial cells lining the GI tract possess their own metabolic pathways and start the metabolizing reactions during absorption of the chemical. Among the organs, contributing both to absorption and detoxification of the drugs are skin, intestine, oral and nasal epithelium, as well as lung and kidney [72]. When the compound reaches to liver, its extensive metabolism begins and then, it joins to systemic circulation of the body so that it can exert its biological activity on target tissue.

In mammals, there are two subfamilies of UGT Superfamily, each located on a different chromosome. In this study, however, we focused on merely the expression

of UGT1A family since these enzymes are primarily involved in the metabolism of xenobiotics while UGT2 family is mostly responsible for the metabolism of endogenous compounds [70].

UGT genes are mainly expressed in liver in an isoform-specific manner [20]. It is documented that 1A1, 1A3, 1A4, 1A6, 1A9 are the isoforms expressed abundantly in liver, taking roles in hepatic drug metabolism [3,73]. However, there are studies showing that the other detoxifying organs with a different profile of UGT isoform expression can well catalyze glucuronidation of numerous drugs [74]. To be more precise, UGT1A1, and UGT1A10 are expressed highly in intestine [34] while the primary tissues expressing 1A7 and 1A8 are members of GI tract [20]. Kidney is the organ where UGT1A9 is expressed at most [34].

Nicotine is one of the most effective addictive compounds in tobacco [75]. There have been numerous research studies ongoing regarding to effects of nicotine on cells. So far, it has been known that nicotine promotes the proliferation of tumor cells *in vivo;* however, a study showing that nicotine causes transformation of normal cells into tumor cells is absent [76].

Smoking has been related with many diseases, such as cardiovascular diseases as well as cancer. Smoking has been associated with UGT1A expression via changes in transcription factors [77]. According to this study, smoking upregulates an oxidative stress response transcription factor Nrf2 and UGT1A4 and UGT1A6 are coordinately expressed with this regulator in small airway epithelia. This initial report initiated us to look at a more generalized set of microarray datasets that contain information on smoker and non-smokers to generalize the previous findings.

Tobacco smoke contains a multitude of chemicals one of which is nicotine. Nicotine is the addictive component in tobacco; and has additionally been associated with lung cancer previously. Nevertheless, the crosstalk between nicotine and colon cancer cells, especially SW620 cells, has not been elucidated in detail. Our previous studies showed that nicotine upregulates proliferation in SW620 cells when serum starved [78]. Kaya (2009)[78] has shown that nicotine increases survival of SW620 cells and a plethora of mRNA transcripts identified by microarrays were involved in this effect. Among the transcripts that are responsive to nicotine but not serum

replenishment are the probesets that belong to UGT1A isoforms. This prompted the present study for identification of isoform expression of UGT1A in SW620 cells under serum starvation.

The present thesis thus has the rationale of revealing the connection between UDP-glucuronosyltransferases with smoking *in vivo* and with nicotine in SW620 colon adenocarcinoma cells. The methods used for accomplishing these aims included microarray expression analysis of publically available datasets obtained from GEO and qRT-PCR studies of UGT1A isoforms using primers published in Nakamura *et al.* (2008) [20].

For this purpose, there are certain questions that we build our experiments on as shown below:

- Does smoking upregulate UGT1A expression *in vivo*?

- What are potential reference genes for studying effects of nicotine on gene expression?

- Is there a change in expression of genes related with cell proliferation when SW620 cells are administered with 1μM Nicotine for 48h?

- Does 1μM Nicotine treatment for 48h upregulate UGT1A isoform expression when cells are grown under physiological serum levels (10%) or serum starved (0.1%)?

-Could there be a relationship between sequence and expression of UGT1A genes?

3. MATERIALS AND METHODS

3.1 Cell Line

The cell line of interest, SW620 colon adenocarcinoma cell line, is supplied by American Type Culture Collection (ATCC; Manassas, USA) with an ATCC number of CCL-227. SW620 colon adenocarcinoma cells are derived from a 51 years-old male and are metastatic. This cell line is kindly presented by Assist. Prof. Dr. Sreeparna Banerjee, METU, Ankara.

3.2 Cell Preservation Protocols

The base medium for SW620 colon adenocarcinoma cells is formulated as DMEM (FG0415) (Biochrom AG, Berlin, Germany); added with 10% Fetal Bovine Serum (CH30160) (Hyclone, Logan, USA) and 1% Penicillin/Streptomycin mixture (SV30010) (Hyclone, Logan, USA).

Cells are grown as supplemented with 12-14mL of complete media in 75-T flasks (Greiner Bio One, Frickenhausen, Germany) in the 37^{o}C incubator with 5% CO_2. In three days interval, cells are passaged into new flasks as described in Kaya (2009) [78].

For cryopreservation, cells were detached with trypsinization and centrifuged. The pellet, then, is resuspended in freezing media (i.e., complete media supplemented with 5% (v/v) DMSO; Sigma-Aldrich, St. Louis, MO, U.S.A) and kept in liquid nitrogen tanks for long time storage. Cell culturing experiments for SW620 cells were performed by Onur Kaya [78].

3.3 Nicotine and Serum Starvation Treatment

To see the effects of nicotine treatment in cells grown in either complete media or serum deprived media, 1x10^{6} SW620 cells were seeded onto 75-T flasks. After 24h, in which cells attach on the surface, media were replaced with serum starved media (prepared as supplementing DMEM with 0.1%FBS and 1%Penicillin/Streptomycin solution) for induction of quiescence.

Quiescent cells were transferred to complete media with different concentrations of nicotine; namely, 10ηM, 100ηM, 1μM and 10μM for 48 hours [78].

Alternatively, another set of a condition was prepared by growing the cells either with complete media, complete media containing 1μM Nicotine, serum-deprived media (0.1%FBS containing media) or serum-deprived media including 1μM nicotine for 48 hours. Each condition was designed so that there are biological replicas. Nicotine and serum starvation treatments were performed by Onur Kaya [78].

3.4 Total RNA Isolation

Treated cells were collected from the flasks on the 2nd day of the treatments by cell scraper as dissolved in ice-cold 1X PBS solution. Following, cells were centrifuged at 1500rpm for 5 min at 4°C and the pellets are stored at -80°C till RNA isolation [78].

Total RNA isolation is carried out with Promega SV Total RNA isolation kit (Z3100) (Madison, USA) according to the manufacturer's protocol. The RNA concentration was measured by NanoDrop ND-1000 spectrophotometer (Thermo Fisher Scientific, Wilmington, DE, USA). Total RNA was used for microarray analysis and cDNA synthesis. The integrity of the RNA samples for microarray analysis was assessed by Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). RNA isolation and quality controls were performed by Onur Kaya [78].

3.5 SW620 Nicotine Treatment Microarray Experiment

5μg RNA samples were hybridized with Affymetrix U133 plus 2 GeneChips [78]. Amplification, labeling and hybridizations were performed at the Genomics Core Facility of Bilkent University by the facility technician under the supervision of Assoc. Prof. Dr. Işık Yuluğ according to manufacturer's protocols.

3.6 Microarray Analysis Methods

BRB array tools has been chosen for data analysis of microarray results [79]. Each .cel file was imported to Microsoft Office Excel as a BRB Project. During import of data, no specific filtering criteria were employed.

Normalization:

Normalization was carried out by justRMA Normalization option integrated in BRB-Array tools. No preferences or filters were used for normalization to be able to see whole data and see the big picture.

Class Comparison:

In order to set a table of differentially expressed genes versus corresponding intensities, unpaired-class comparison tool of BRB Array tools with a p-value of 0.05 was used. After annotating each array as either smoker or non-smoker, differentially expressed genes were explored by Class Comparison among these two groups. BRB-array tools make use of univariate parametric tests (t/F) to define differentially expressed genes among two groups. Class comparison analyses also provide a false discovery rate value (FDR). This allows multiple test correction for the probesets analyzed from microarray analyses.

Cluster and Treeview:

Clustering was carried out by Cluster 3.0 software program while the image was generated by Java TreeView software [80]. For this purpose, differentially expressed genes among all the groups derived from unpaired-class comparison tool of BRB Array tools (with a P-value of 0.05) were uploaded on Cluster 3.0. Then, median gene and array centering was performed. Average linkage method was chosen and un-centered correlation coefficient was used.

3.7 Re-analysis of Public Datasets for UGT1A probesets

Publicly available datasets were gathered by searching through Gene Expression Omnibus (GEO) website and using "Smoke or smoking or Nicotine" as keywords[81]. In order to analyze as many datasets as possible, Affymetrix HG-U133A, HG-U133 plus 2 and HuGeneFL Human Full Length Arrays were collected.

According to Ensembl database, there are five probesets in Affymetrix HG-U133A and HG-U133 Plus 2 arrays that give hits on all UGT1A isoforms. These probesets are:

- 204532_x_at

- 206094_x_at
- 207126_x_at
- 208596_s_at
- 215125_s_at

In addition, there is extra probeset in the same arrays matching with UGT1A8 and UGT1A9 genes. This probeset is 221305_s_at.

Also, there is a probeset that is assigned only to UGT1A6 isoform in addition to those given above; which is 232654_s_at. However, this probeset is present only in Affymetrix HG-133 plus 2 arrays.

Furthermore, in HuGeneFL Human Full Length array, there is only one probeset that recognizes all UGT1A isoforms; namely, J04093_s_at.

We have analyzed several microarray results obtained in [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array, [HG-U133A] Affymetrix Human Genome U133A Array or [Hu6800] Affymetrix Human Full Length HuGeneFL Array **Table 5.1**, **Table 5.2**, **Table 5.3** and **Table 5.4**. To do so, we have used BRB-Array Tools as described above in **Section 3.6**. We have normalized the .cel data downloaded from the webpage of each array in GEO database with justRMA option of BRB Array tools. Each array was annotated as either smoker or non-smoker and a class comparison was done between these two groups as setting p value as 0.05. The results of each re-analysis were shown below labeled with GSE numbers and short explanation of the clinical sample.

3.8 Statistical Analysis and Plots of GEO Expression Values using GraphPad

Initially, the normalized values for each probeset in a dataset were obtained as non-log values. The geometric mean and standard deviation of each group were calculated and uploaded to GraphPad (GraphPad Software, Inc.). This program plots bar graphs according to the mean and SD given. The statistical significance of alteration in probesets was assessed depending on the p-value and FDR-value determined by BRB-Array tools.

3.9 cDNA synthesis

cDNA synthesis was done with RevertAid First Strand cDNA synthesis kit (K1622) (MBI Fermentas, Ontario, Canada) according to manufacturers' protocol.

3.10 Validation of Microarray Analysis with qRT-PCR

For quantitative Real-time PCR experiments, Maxima SYBR Green qPCR Master Mix (K0251) obtained from MBI Fermantas (Ontario, Canada) was used with Applied Biosystems 7500 Real-time PCR systems (California, USA).

3.10.1 Oligonucleotides:

The oligonucleotides for qRT-PCR analysis of UGT1A isoforms were adapted from the study of Nakamura *et al.* (2008) [20] for comparability. They were synthesized by Iontek Inc. (Istanbul, Turkey). For all UGT1A primer couples, same reverse primer was used. Oligonucleotides to check for the proliferative capacity of SW620 cells treated with nicotine also were ordered from Iontek Inc. (Istanbul, Turkey). The oligo sequences of the primers, their expected amplicon size and annealing temperatures are shown on **Table 3. 1,**

,Table 3. and **Table 3.3**.

Table 3. 1 The oligonucleotides used for UGT1A expression analysis. Adapted from Nakamura *et al.* (2008) [20].

Gene name	Forward Primer	Reverse Primer	Amplicon size	Annealing Temp
UGT1A1	5'CCT TGC CTC AGA ATT CCT TC 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	361bp	58°C
UGT1A3	5'GTT GAA CAA TAT GTC TTT GGT CT 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	595bp	58°C
UGT1A4	5'CCT GCT GTG TTT TTT TGG AGG T 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	431bp	58°C

UGT1A5	5'TGT CCT ACC TTT GCC ATG CTG 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	274bp	58°C
UGT1A6	5'CAA CTG TAA GAA GAG GAA AGA C 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	97bp	58°C
UGT1A7	5'CCC CTA TTT TTT CAA AAA TGTCTT 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	261bp	58°C
UGT1A8	5'GGT CTT CGC CAG GGG AATAG 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	423bp	58°C
UGT1A9	5'GAA CAT TTA TTA TGC CAC CG 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	275bp	58°C
UGT1A10	5'CTC TTT CCT ATG TCC CCA ATG A 3'	5' ATT GAT CCC AAA GAG AAA ACC AC 3'	364bp	58°C

,**Table 3.2** The oligonucleotides used as reference genes in SW620 cells.

Gene name	**Forward Primer**	**Reverse Primer**	**Amplicon size**	**Annealing Temp**
RPL3	5' CCT TCT CTG TGG CAC GCG CT 3'	5' CCG CTT CGT CTG CAC CAG CA 3'	284bp	60°C
RPL10	5' ACC TGG GGC GGA AAA AGG CA 3'	5' GGT GGA GCC GCA CCC GGA TA 3'	174bp	60°C
GAPDH	5'GGC TGA GAA CGG GAA GCT TGT CAT 3'	5' CAG CCT TCT CCA TGG TGG TGA AGA 3'	142bp	60°C
PPIA	5' CGT GTG CTA TTA GCC ATG GT 3'	5' CAT TAT GGC GTG TGA AGT C 3'	229bp	60°C
RPLP0	5'TCATCCAGCAGGT GTTCGAC C 3'	5' AGA CAA GGC CAG GAC TCG TT 3'	194bp	60°C

Table 3.3 The oligonucleotides used for the analysis of proliferative capacity in SW620 cells.

Gene name	Forward Primer	Reverse Primer	Amplicon size	Annealing Temp
ANLN	5' TAA AGC AGG TGA TTG TTC GG 3'	5' GTT CTT CAT CAA CAC AGC AG 3'	180bp	60°C
MKI67	5' GTG TCA AGA GGT GTG CAG AA 3'	5' GCC TTA CTT ACA GAA TTC AC 3'	197bp	60°C

3.10.2 Quantitative Real-time PCR

The primer pairs targeting the genes in interest to analyze the change in their expression as a response to nicotine and serum starvation treatments are given in *Table 3.1, 3.2* and *3.3.* For each reaction, 12.5 µl 2X SBYR Green Master Mix (MBI Fermantas, Ontario, Canada) was added with 0.75 µl 10pmol/ µl forward and reverse primers, 0.05µl 10nM ROX solution as a reference dye and total volume was completed to 24 µl with distilled water. Reaction volume was completed to a total of 25 µl with 1:2 diluted cDNA templates. The reaction was as follows: initial denaturation at 95ºC for 10 min, followed with 40 cycles of 30s of 95 ºC, 30s of 58 ºC, 30s of 72 ºC and final extension at 72 ºC for 10 min. Dose-dependent nicotine samples grown at 10%FBS were run with two technical duplicates while serum starved samples were run with three biological replicates and two technical duplicates.

Melting curves and Ct values were automatically obtained from the software of Applied Biosystems 7500 Real-time PCR systems (California, USA). Each melting curve was analyzed to see whether negative controls were free of product and samples resulted in a single amplicon. For each primer pair, an efficiency value was calculated by running a reaction with 2-fold serially diluted cDNA and formula as shown below:

$$E=2^{[-1/\text{slope of the dilution curve}]}$$

Fold changes in gene expression was calculated by normalizing according to the expression of a reference gene, i.e., GAPDH. GAPDH was chosen among five potential reference genes, as one of the most stable ones across nicotine treatment

,Table 3.. The mathematical formula used for determination of fold changes was [59]:

$$(\text{Etarget})^{\Delta \text{CtTarget (control-sample)}} / (\text{Ereference})^{\Delta \text{CtReference (control-sample)}}$$

Following the runs, each PCR product was run on 2% Agarose Gel Electrophoresis.

3.10.3 ANOVA Analysis of UGT1A isoforms from qRT-PCR

The fold changes normalized to GAPDH levels of each control and treated samples were uploaded to Minitab Statistical Software [82] and One-way ANOVA (Unstacked) option was used to calculate DF, F-value and p-value.

3.11 General Solution Recipes

10X PBS Stock solution: 80g NaCl, 2g KCl, 11.5g Na2HPO4.7H2O, and 2g KH2PO4 are weighed and dissolved in ddH_2O to a total volume of 1 liter and autoclave steriled.

1X PBS Working solution: 10X PBS was diluted with 1:10 ratio and pH is adjusted to 7.4. Working solution should be autoclaved for one more time after dilution and filtered before use in tissue culture.

50X Tris-acetic acid-EDTA (TAE): 242gr Tris Base and 18.6g EDTA are dissolved in ddH2O. They are added with 57.1mL Glacial Acetic acid and volume is completed to 1 liter.

4. RESULTS

4.1 Re-analysis Microarray Datasets with Smoker Status from GEO

In order to answer the question whether smoking induces the expression of UGT1A genes *in vivo*, we performed a large-scale re-analysis of publicly available datasets that contain information on past and/or smoking status.

As given in Ensembl database, probesets assigned for UGT1A isoforms are uniform, implying that they do not differentiate each isoform. The reason is that since RNA degradation begins at 5'end, probesets are designed to hit on 3'ends. Because UGT1A isoforms share the same common 3'ends, it is not possible to discriminate among them using microarray experiments from Affymetrix. Consequently, we can only observe the changes in expression of UGT1A isoforms as a whole, and the alterations occurring on each gene are not well reflected on the results. In other words, an increase in the expression of a certain isoform may cover a decrease in the other. Nevertheless, the significant changes in expression can be observed anyway. Furthermore, there are specific probesets that additionally recognize the certain isoforms UGT1A8 and UGT1A9. In addition, there is a specific probeset hitting on UGT1A6 gene only in HG133 Plus 2.0 Array.

After normalization and class comparison of the data, a list of genes that are differentially expressed among smokers and non-smokers was obtained using the class comparison tool ($p<0.05$). Afterwards, the probesets that are annotated for UGT1A isoforms in the corresponding chip type were surveyed whether they are differentially expressed, as well.

Below, a short summary of each dataset chosen and the corresponding intensities of each UGT1A probeset in that array if significantly expressed ($p<0.05$) were shown. In addition, the geometric mean of probeset intensities was represented with a bar graph for each array.

GSE4635: Large airway bronchial epithelial cells

This array was designed to analyze the distinct transcriptomic profiles in large airway bronchial epithelial cells, differentiating smokers from non-smokers. There were 5 healthy non-smoker and 5 healthy smoker samples, whose RNA were hybridized with [HG-U133A] Affymetrix Human Genome U133A Array. Samples for RNA were isolated from right mainstem bronchus of individuals by means of fiberoptic bronchoscopy [83]. Based on the knowledge that, airway epithelium is one of the tissues that might be exposed to nicotine most, the findings might implicate to the relationship between smoking and UGT1A expression.

Class comparison tool of BRB-Array tools has brought up 906 genes that are significantly changed ($p<0.05$) among smokers and non-smokers. In this gene list, there were probesets annotated to UGT1A isoforms, as well **(Table 4.1**). According to this data, UGT1A expression was significantly upregulated ($p<0.05$) in airway epithelium of smokers by an average of 2 fold change constant as compared to non-smokers (**Figure 4.1).** In addition, UGT1A8 and UGT1A9 expression did not change among smokers and non-smokers **(Figure 4.2).** Indeed expression level of these probesets was very low making the detection of their differential expression difficult.

Table 4.1 The geometric mean of intensities corresponding to common UGT1A probesets in GSE4635 array ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0004475	0.19	217.39	411.29	0.53	204532_x_at
0.0016715	0.443	214.49	382.03	0.56	206094_x_at
0.0012313	0.365	225.28	424.74	0.53	207126_x_at
0.0208181	1	149.26	241.53	0.62	208596_s_at
4.44E-05	0.0707	99.82	250.35	0.4	215125_s_at

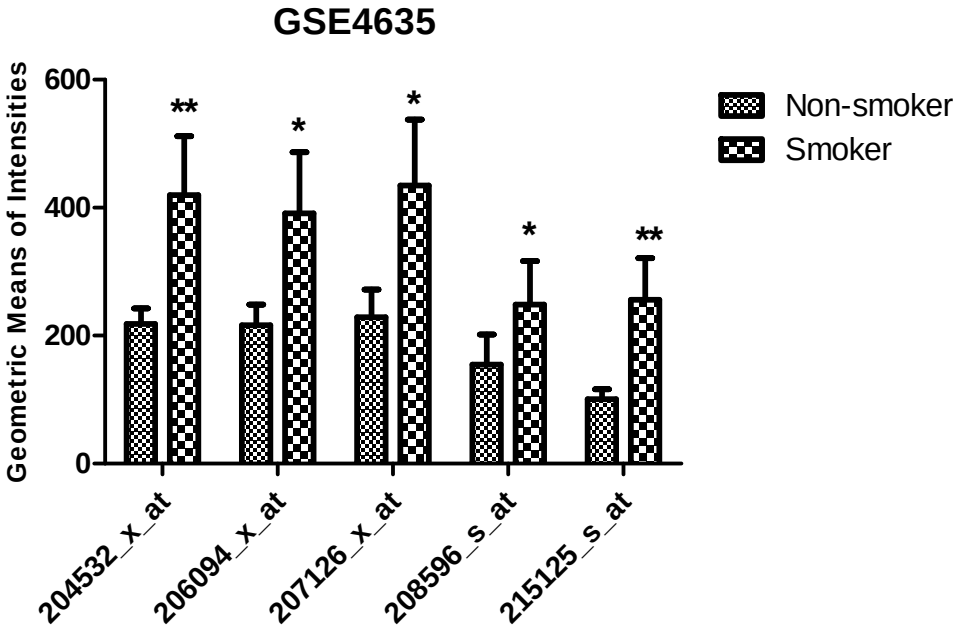

Figure 4.1 UGT1A expression in large airway epithelium of smokers and non-smokers according to GSE4635 array (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

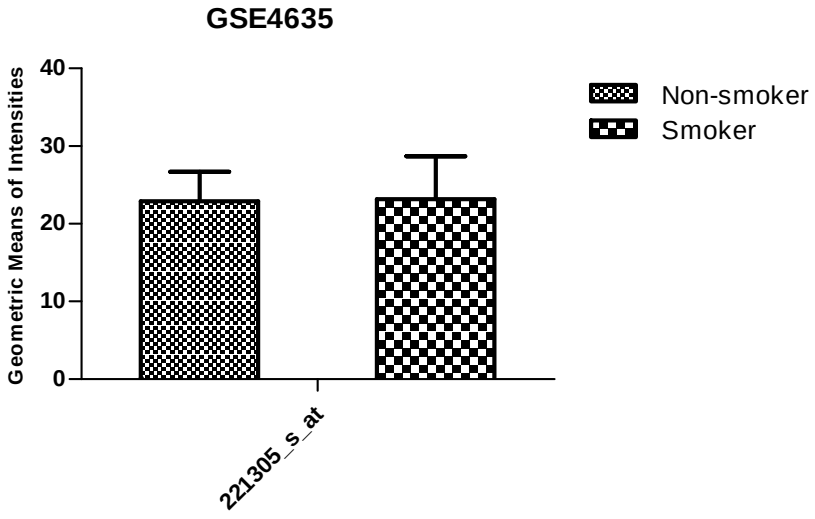

Figure 4.2 UGT1A8-1A9 expression in large airway epithelium of smokers and non-smokers according to GSE4635 array (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

GSE4498: Small airway bronchial epithelial cells

This dataset is focused on specific arrays originally conducted as a part of superseries GSE5060 [84]. Overall, GSE5060 superseries aim to analyze how cigarette smoking affects neuroendocrine cells in human airway epithelium. Based on this question, they have analyzed both large and small airways of phenotypically normal smokers and non-smokers, in addition to patients with early COPD and COPD. Since we try to assess the effects of smoking on healthy individuals, we excluded the subseries analyzing patients with COPD. Moreover, all subseries were not hybridized with the same platform. Thus, we decided to analyze each subseries individually.

GSE4498 is composed of small airway epithelial samples from 10 healthy non-smokers and 12 healthy smokers obtained via fiberoptic bronchoscopy. The platform used is [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Among 6543 genes with significance of $p<0.05$, there were common UGT1A probesets **(Table 4.2)** and UGT1A6 **(Table 4.3)** and UGT 1A8- 1A9 probesets, as well **(Table 4.4).** Accordingly, smokers have at least twice the expression of UGT1A isoforms when compared with non-smokers **(Figure 4.3)** where FDR was less than 0.05. UGT1A6 probeset detected a very low expression in non-smokers, which again was doubled in smokers **(Figure 4.4).** UGT1A8 and UGT1A9 probesets detected very low expression, which was difficult to distinguish from base-line background expression **(Figure 4.4).**

Table 4.2 The geometric mean of intensities corresponding to common UGT1A probesets in small airway bronchial epithelial cells from GSE4498 ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
1.00E-07	0.000342	112.2	338.11	0.33	204532_x_at
5.00E-07	0.0013	121.27	332.65	0.36	206094_x_at
1.00E-07	0.000342	119.88	354.75	0.34	207126_x_at

4.00E-07	0.00109	120.92	416.67	0.29	208596_s_at
7.30E-06	0.00849	79.82	261.01	0.31	215125_s_at

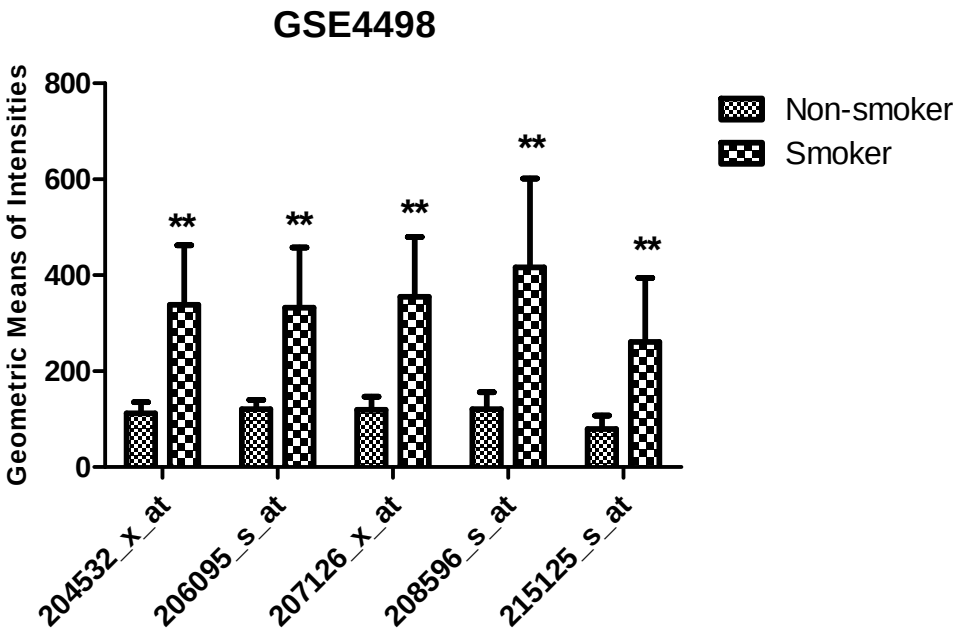

Figure 4.3 UGT1A expression in small airway epithelium of smokers and non-smokers according to GSE4498 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Table 4.3 The geometric mean of intensities corresponding to UGT1A6 probeset in GSE4498 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0025036	0.159	16.76	25.89	0.65	232654_s_at

Table 4.4 The geometric mean of intensities corresponding to UGT1A8 and 1A9 probeset in GSE4498 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0196277	0.328	10.75	11.99	0.9	221305_s_at

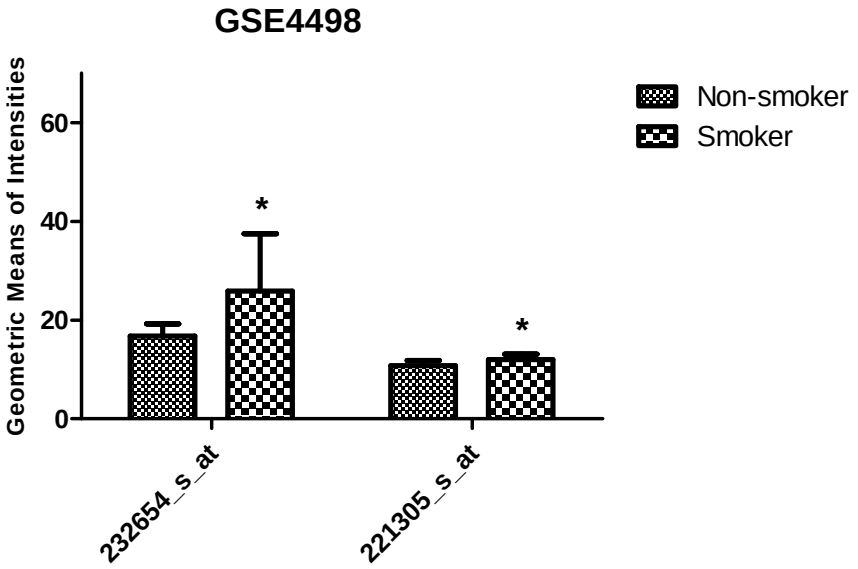

Figure 4.4 UGT1A6 (232654_s_at), UGT1A8 and 1A9 (221305_s_at) expression in small airway epithelium of smokers and non-smokers according to GSE4498 (*$p<0.05$, **$p<0.001$). Results were given as geometric mean ± SD.

GSE10006: Human airway bronchial epithelial cells

This array is another study concerning the changes at the transcriptomic level in the airway epithelium of smokers and non-smokers [85]. Nevertheless, the dataset is comprised of samples from both healthy and sick individuals. Therefore, we excluded the samples from patients with COPD, leaving 58 healthy samples behind.

With significance of $p<0.05$, there were 11240 genes that were differentially expressed; among which there were UGT1A genes **(Table 4.5).** Furthermore,

probesets annotated for UGT1A6 **(Table 4.6)** and UGT1A8-1A9 significantly changed, as well **(Table 4.7).**

Analysis of GSE10006 further showed that human airway epithelial cells bear a defense mechanism against smoking. This notion was supported by close to three fold upregulation of UGT1A common probesets **(Figure 4.5)** as well as UGT1A6 (less than two fold). However, 1A8 and 1A9 specific probes **(Figure 4.6)** did not exhibit differential expression at the level of FDR=0.05; furthermore, the expression of UGT1A8 and UGT1A9 was low based on the given microarray probesets.

Table 4.5 The geometric mean of intensities corresponding to common UGT1A probesets in GSE10006 ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
< 1e-07	< 1e-07	189.71	505.94	0.37	204532_x_at
1.00E-07	5.26E-05	172.08	405.18	0.42	206094_x_at
< 1e-07	< 1e-07	167.94	428.43	0.39	207126_x_at
1.00E-07	5.26E-05	183.89	537.32	0.34	208596_s_at
1.00E-07	5.26E-05	130.91	401.78	0.33	215125_s_at

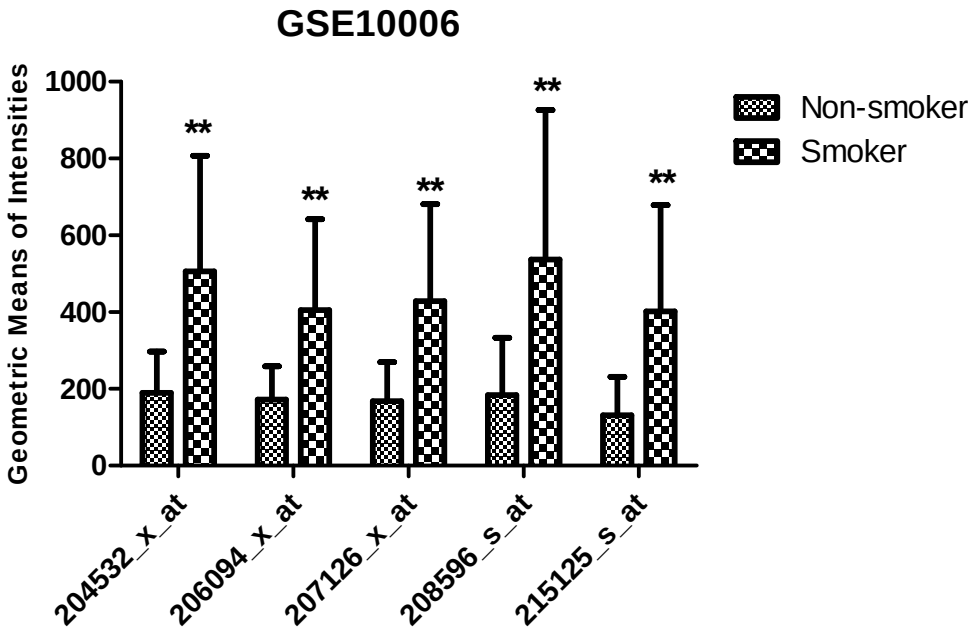

Figure 4.5 UGT1A expression in small airway epithelium of smokers and non-smokers according to GSE10006 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Table 4.6 The geometric mean of intensities corresponding to UGT1A6 probeset in GSE10006 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.000199	0.014	21.47	31.28	0.69	232654_s_at

Table 4.7 The geometric mean of intensities corresponding to UGT1A8-1A9 probeset in GSE10006 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0136421	0.141	14.23	17.44	0.82	221305_s_at

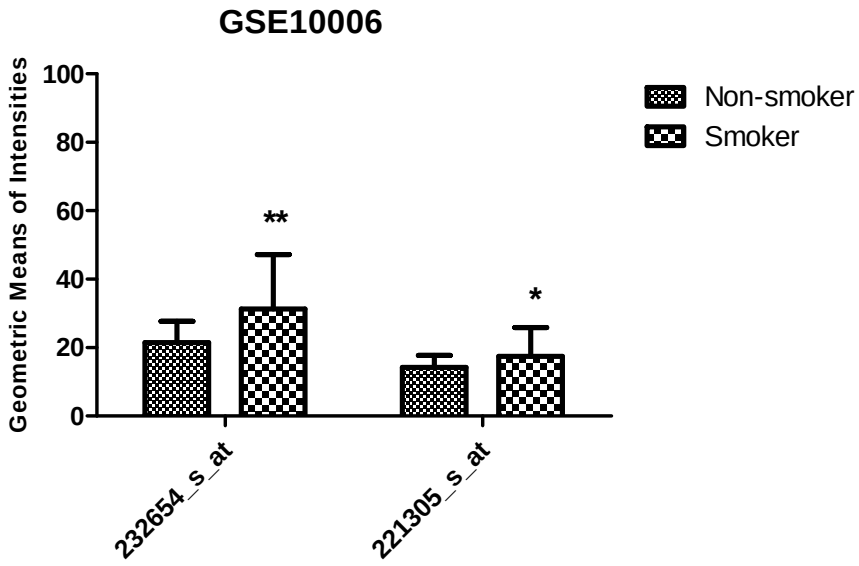

Figure 4.6 UGT1A6 (232654_s_at), UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of smokers and non-smokers according to GSE10006 (* $p<0.05$, **$p<0.001$). Results were given as geometric mean ± SD.

GSE17913: Oral Mucosa

GSE 17913, an extensive dataset, is of big importance in that it is one of the pioneering studies in literature showing that the expression of UGT enzymes was regulated by cigarette smoking [86]. This study was conducted on 39 healthy smokers and 40 healthy non-smokers. The RNA samples belong to oral mucosa of the individuals and were obtained by buccal biopsy. The platform used is [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array.

BRB analysis shows that there are 7956 genes that are differentially expressed between smokers and non-smokers ($p<0.05$). Moreover, the signal intensities of all probes assigned to UGT1A probes were significantly changed at the FDR<0.05 **(Table 4.8; Figure 4.7).** In addition, probesets annotated for UGT1A8-1A9 were upregulated, as well **(Table 4.9; Figure 4.8).**

Furthermore, all probesets showed a less than two-fold upregulation when compared with non-smokers. In summary, oral mucosa cells were shown to be responsive to cigarette smoke in terms of UGT1A induction.

Table 4.8 The geometric mean of intensities corresponding to common UGT1A probesets in GSE17913 ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
5.00E-07	0.000804	459.2	692.61	0.66	204532_x_at
4.00E-07	0.000683	331.76	487.24	0.68	206094_x_at
2.00E-07	0.000421	439.84	667.39	0.66	207126_x_at
2.50E-06	0.00258	315.65	483.31	0.65	215125_s_at
3.55E-05	0.0139	269.32	397.65	0.68	208596_s_at

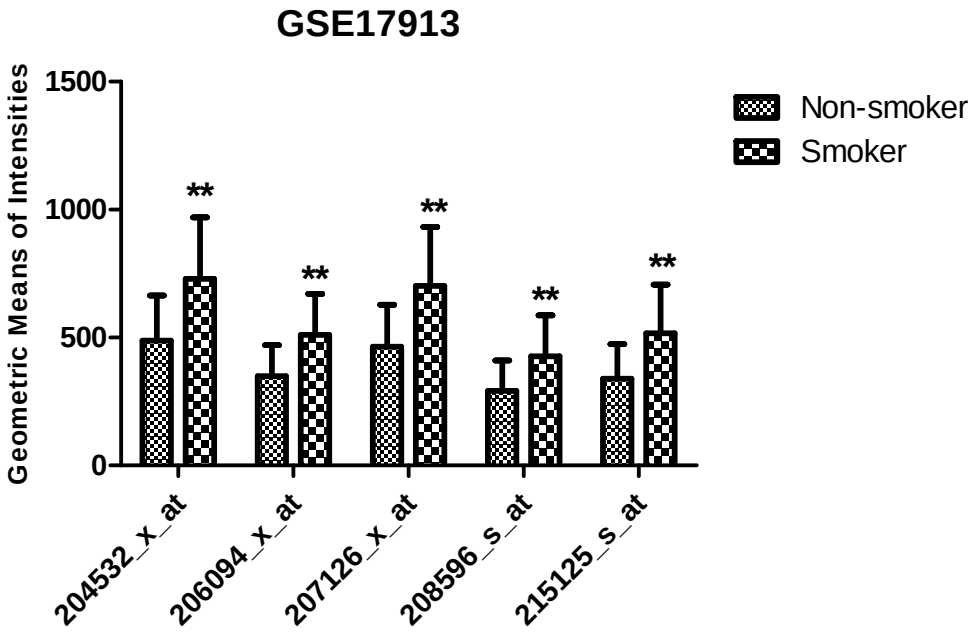

Figure 4.7 UGT1A expression in oral mucosa of smokers and non-smokers according to GSE17913 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Table 4.9 The geometric mean of intensities corresponding to UGT1A8-1A9 probeset in GSE17913 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
5.22E-05	0.016	62.88	97.56	0.64	221305_s_at

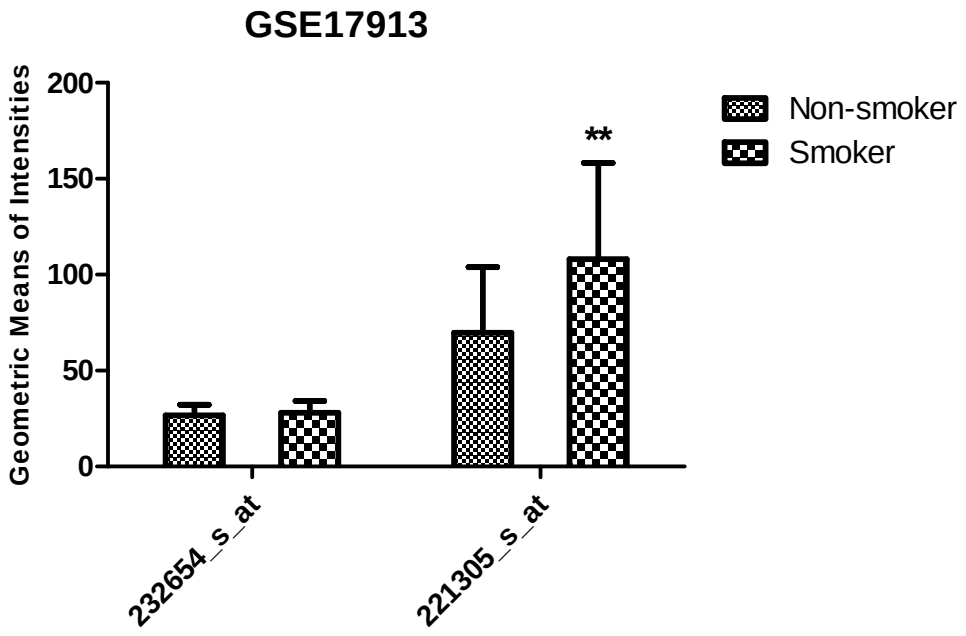

Figure 4.8 UGT1A6 (232654_s_at), UGT1A8 and 1A9 (221305_s_at) expression in oral mucosa of smokers and non-smokers according to GSE17913 (* $p<0.05$, **$p<0.001$). Results were given as geometric mean ± SD.

GSE 16149: Oral mucosa

GSE16149, on the other hand, is another dataset generated with buccal mucosa samples of individuals [87]. In this array, [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array was used to detect the alterations that have occurred in global gene expression patterns of buccal mucosa of 4 smokers and 4 non-smokers. An interesting point is that there are two biological replicates from each individual

originating from either right or left cheek of the patient. Therefore, there are 8 RNA samples from each group. In addition, there are two samples not having a replicate, making a 9 sample for each group. It should be noted that we collectively used each replicate as a single sample, ignoring a distinction as right or left cheek. Future analyses might consider testing right or left cheek separately or using a repeated measures design for statistical analysis.

With significance of $p<0.05$, there were 2755 genes that were differentially expressed; among which there was a UGT1A probeset **(Table 4.10).** However, since the expression level measured by this probeset is around the basal level, it was not considered significantly modified by smoking (**Table 4.11**; **Figure 4.9**; **Figure 4.10)**. Although UGT1A8 and UGT1A9 probeset was expressed in oral mucosa it was not differentially expressed as in the previous dataset (**Figure 4.8**). Small size of this dataset might lead to such a discrepancy.

Table 4.10 The geometric mean of intensities corresponding to common UGT1A probesets in GSE16149 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0368421	0.883	13.6	10.44	1.3	206094_x_at

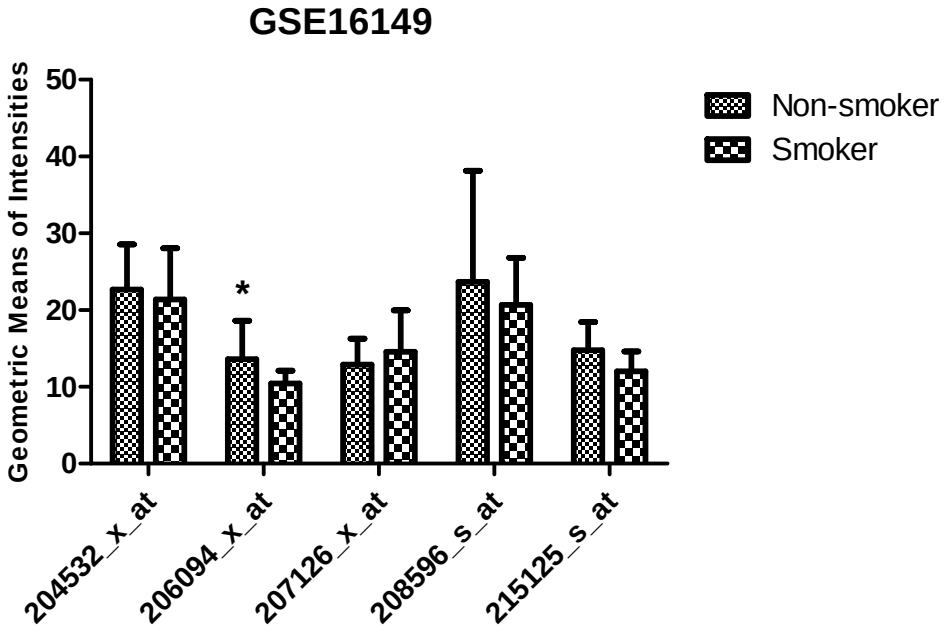

Figure 4.9 UGT1A expression in oral mucosa of smokers and non-smokers according to GSE16149 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

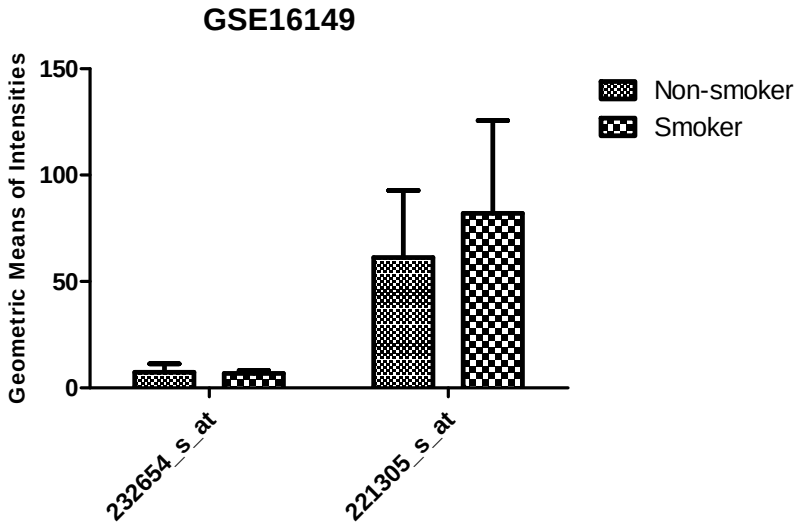

Figure 4.10 UGT1A6 (232654_s_at), UGT1A8 and 1A9 (221305_s_at) expression in oral mucosa of smokers and non-smokers according to GSE16149 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

GSE 8987: Nasal and Buccal Epithelial Cells

This dataset is composed of two different array sets conducted with two different platforms and tissues [88]. Since nose and mouth are two primary organs exposed to tobacco, the alterations in the gene expression of epithelial cells lining in these organs are of importance. Regarding this, the research team has isolated RNA from both nasal epithelial cells and buccal epithelial cells of current and never smokers. Total RNA isolated from mouth of 5 current and 5 non-smokers were run on [HG-U133A] Affymetrix Human Genome U133A Array. Moreover, RNA from nose epithelium of 7 current and 8 non-smokers were run on [HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array. Therefore, we have separated this dataset into two as GSE8987-Nasal Epithelium and GSE8987-Buccal Epithelium.

Nasal Epithelial Cells

BRB analysis has shown that there were UGT1A1 probesets that were significantly expressed besides 493 other genes ($p<0.05$) **(Table 4.11).** FDR analysis did not show significance, however, most likely due to small sample size of the study. The common UGT1A probesets showed a less than two-fold upregulation when compared with non-smokers **(Figure 4.11).** However, UGT1A8-1A9 probesets, although expressed, were not among those **(Figure 4.12).**

Table 4.11 The geometric mean of intensities corresponding to common UGT1A probesets in GSE8987 ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0050582	1	383.65	507.27	0.76	204532_x_at
0.0058929	1	292.89	391.25	0.75	206094_x_at
0.0325089	1	397.4	496.5	0.8	207126_x_at
0.0062479	1	184.77	257.47	0.72	215125_s_at

0.0013462	1	276.31	482.27	0.57	208596_s_at

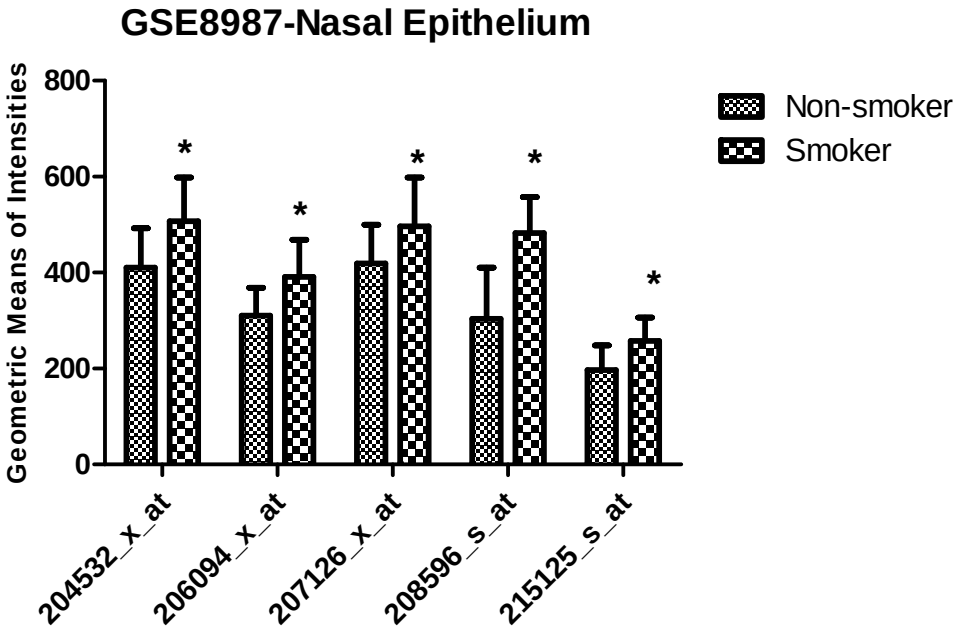

Figure 4.11 UGT1A expression in nasal epithelium of smokers and non-smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

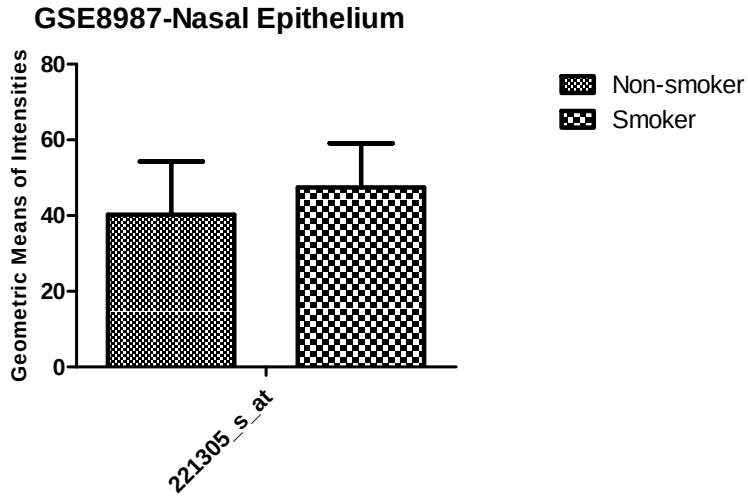

Figure 4.12 UGT1A8 and 1A9 (221305_s_at) expression in nasal epithelium of smokers and non-smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Buccal Epithelial Cells

In the second part of the experiment, total RNA isolated from buccal epithelial cells was hybridized with [HG-U133A] Affymetrix Human Genome U133A Array. Interestingly none of the UGT1 probesets or UGT1A8-UGT1A9 probesets, which were expressed at very low levels, were significantly altered in buccal epithelial cells while there were genes affected in nasal epithelial cells **(Figure 4.13 and Figure 4.14).**

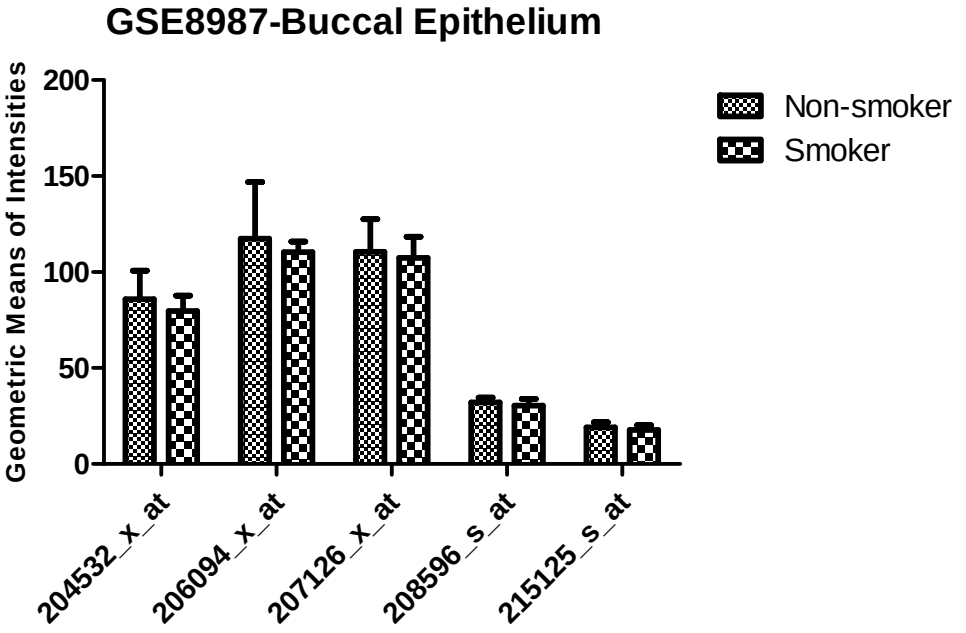

Figure 4.13 UGT1A expression in buccal epithelium of smokers and non-smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

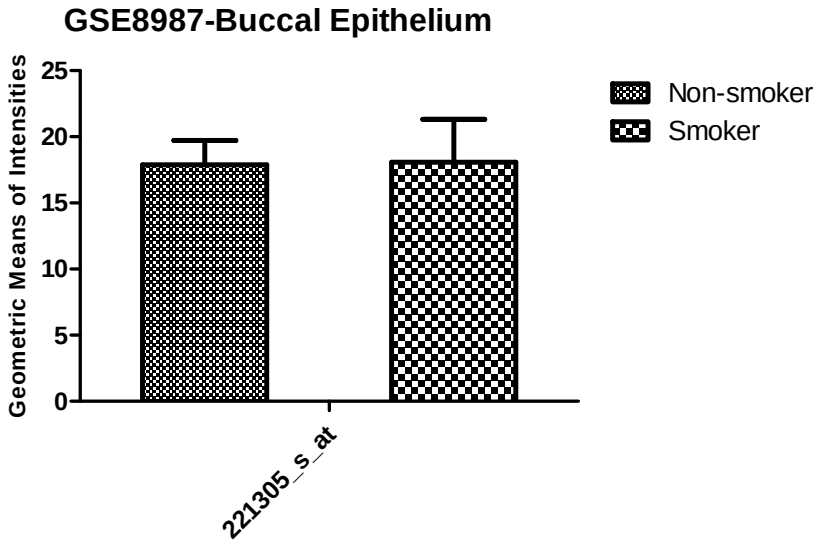

Figure 4.14 UGT1A8 and 1A9 (221305_s_at) expression in buccal epithelium of smokers and non-smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

GSE7895: Human airway bronchial epithelial cells

GSE7895 is one of the arrays with detailed smoking history, indicating whether patients were former, current and never-smokers [89]. This dataset tested the reversible and irreversible effects of tobacco on healthy individuals using bronchial epithelial cells isolated with fiberoptic bronchoscopy. Since long term effects of smoking can be reversible after quitting, this array contains samples of never, former and current smokers. The number of samples is as follows: 21 never smokers, 31 former smokers and 52 current smokers. Total RNA isolated from bronchial epithelium of individuals were hybridized with [HG-U133A] Affymetrix Human Genome U133A Array.

Smoker vs Non-smoker Comparison

When the arrays were grouped as smokers (including current and former) and non-smokers, with significance of $p<0.05$, there were 3141 genes that were differentially expressed among which there were common UGT1A probesets **(Table 4.12)** and the UGT1A8-1A9 probeset **(Table 4.13)**. All of these probesets except one had an FDR level less than 0.1 indicating 10% false discovery rate.

When we compare the arrays as smokers and non-smokers, there was a significant increase in the intensities of common probesets (some close to two fold) and UGT1A8 and 1A9 specific probesets **(Figure 4.15 and Figure 4.16).** However, the expression level of UGT1A8-9 probeset was again at a very low level and the differential expression detected by class comparison is not reliable due to low expression.

Table 4.12 The geometric mean of intensities corresponding to common UGT1A probesets in GSE7895 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
9.10E-06	0.00155	172.17	258.66	0.67	204532_x_at
2.26E-05	0.00286	168.51	242.98	0.69	206094_x_at
1.02E-05	0.00169	196.91	303.28	0.65	207126_x_at
0.009617	0.158	112.31	145.39	0.77	208596_s_at
1.22E-05	0.00181	78.05	135.78	0.57	215125_s_at

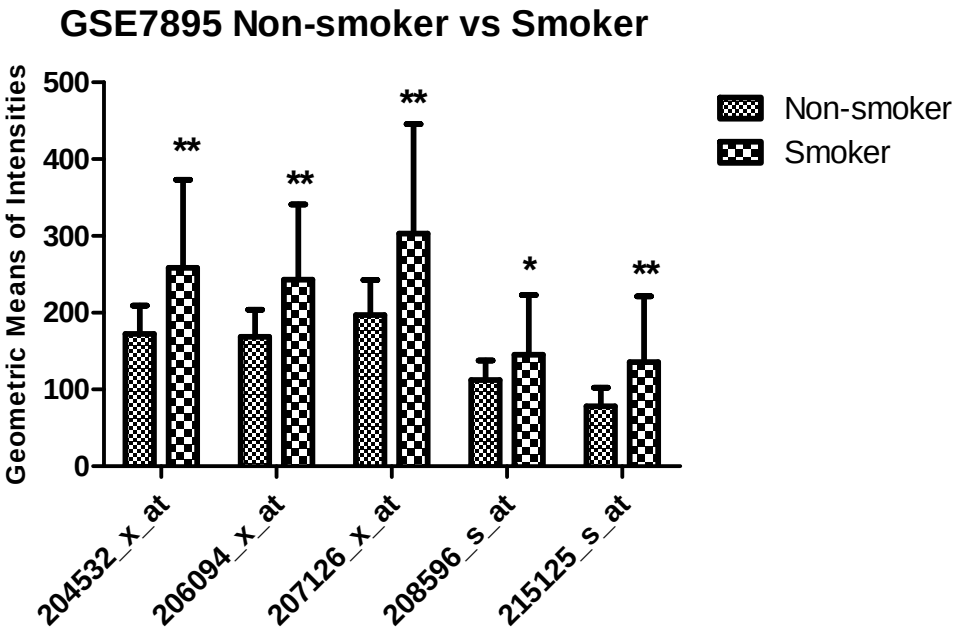

Figure 4.15 UGT1A expression in airway epithelium of smokers and non-smokers according to GSE7895 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Table 4.13 The geometric mean of intensities corresponding to UGT1A8-1A9 probeset in GSE7895 ($p<0.05$)

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
0.0025816	0.0742	18	20.41	0.88	221305_s_at

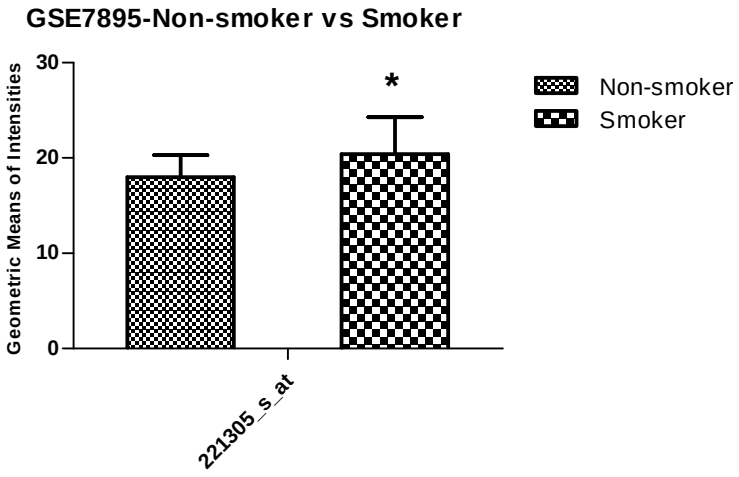

Figure 4.16 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of smokers and non-smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Never vs Current vs Former Comparison

When arrays were grouped based on their smoking status, there were 4841 genes that were differentially expressed among current, former and never smokers; among these probesets there were the common UGT1A probesets ($p<0.05$; FDR<0.05) **(Table 4.14; Figure 4.17).** Although, UGT1A8-1A9 probeset was reported as differentially expressed; the fold change is very low while the expression of this probeset was not different from the background levels on the array **(Table 4.15; Figure 4.18)**.

A comparison between former, current and never smokers show that the levels of UGT1A genes as a whole were highest at current smokers. The fold difference between former and never smokers were not as much as the alteration between current smokers and the rest. This implies that the effects of cigarette smoke with regard to UGT1A expression are reversible in case when the individual quit smoking.

Table 4.14 The geometric mean of intensities corresponding to common UGT1A probesets in GSE7895 ($p<0.05$)

Parametric p-value	FDR	Current Smoker (Geom mean of intensities)	Former Smoker (Geom mean of intensities)	Never Smoker (Geom mean of intensities)	Probe set
< 1e-07	< 1e-07	306.71	194.37	172.17	204532_x_at
< 1e-07	< 1e-07	282.5	188.7	168.51	206094_x_at
< 1e-07	< 1e-07	362.86	224.48	196.91	207126_x_at
< 1e-07	< 1e-07	172.86	108.75	112.31	208596_s_at
< 1e-07	< 1e-07	170.85	92.36	78.05	215125_s_at

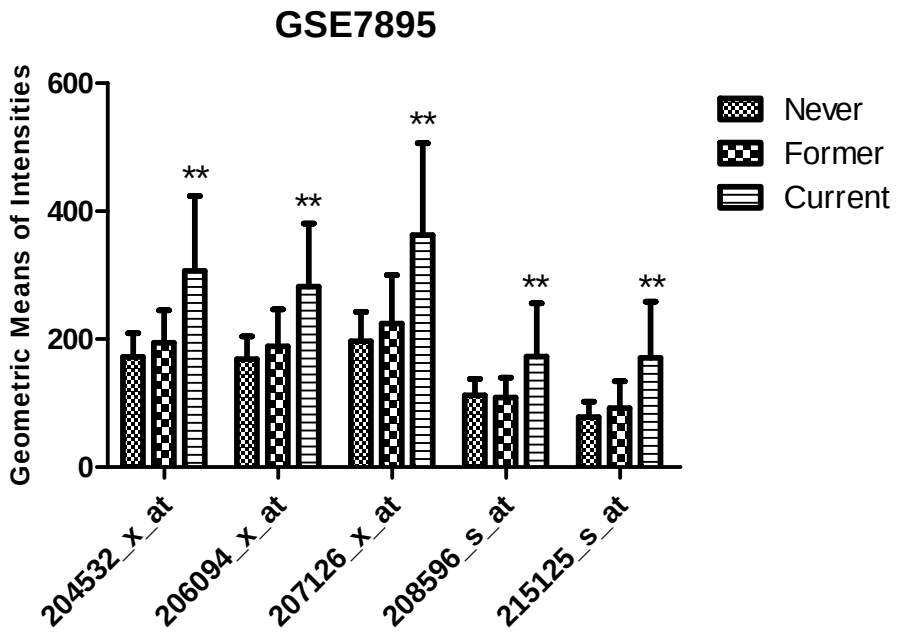

Figure 4.17 UGT1A expression in airway epithelium of current, former and non-smokers according to GSE7895 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Table 4.15 The geometric mean of intensities corresponding to UGT1A8-1A9 probeset in GSE7895 ($p<0.05$)

Parametric p-value	FDR	Current Smoker (Geom mean of intensities)	Former Smoker (Geom mean of intensities)	Never Smoker (Geom mean of intensities)	Probe set
0.0102216	0.107	20.31	20.59	18	221305_s_at

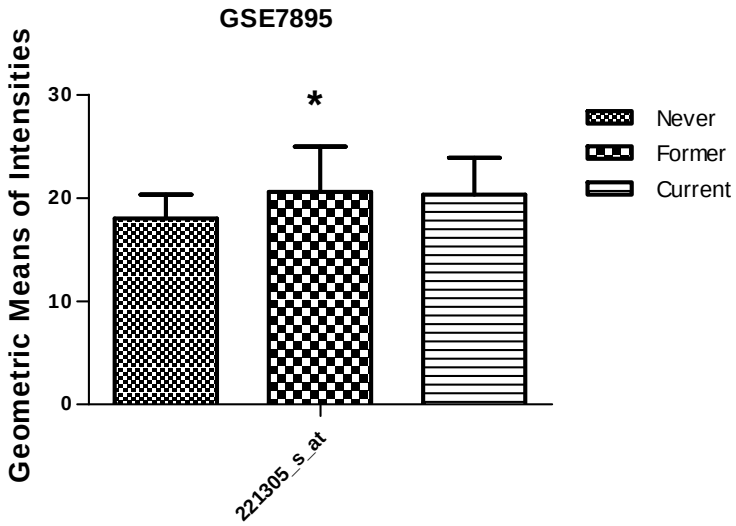

Figure 4.18 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of current, former and never smokers according to GSE8987 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

GSE 994: Human airway epithelial cells

GSE994 dataset included arrays that were run with samples from intra-pulmonary airway epithelial cells of 34 current smokers, 23 never smokers and 18 former smokers [90]. The samples were obtained by brushing. The arrays run with [HG-U133A] Affymetrix Human Genome U133A Array were analyzed with BRB-Array tools in two parts.

Smoker vs Non-smoker Comparison

When smokers and non-smokers were compared, with significance of $p<0.05$, there were 3827 genes that were differentially expressed. Among these all common UGT1A probesets except one (208596_s_at) were significant at FDR<0.05 **(Table 4.16).**

GSE994 showed that the highest expression of UGT1A genes was present at airways of smokers in comparison with non-smokers **(Figure 4.19).** Again UGT1A8-A9 probeset was not expressed to detectable levels by the microarray probeset **(Figure 4.20).**

Table 4.16 The geometric mean of intensities corresponding to common UGT1A probesets in GSE994 ($p<0.05$).

Parametric p-value	FDR	Non-smokers (Geom mean of intensities)	Smokers (Geom mean of intensities)	Fold-change	Probe set
3.00E-06	0.000743	99.28	148.98	0.67	204532_x_at
0.0003131	0.0176	104.19	139.25	0.75	206094_x_at
3.56E-05	0.00422	111.33	166.21	0.67	207126_x_at
0.0108195	0.142	74.38	94.51	0.79	208596_s_at
5.30E-06	0.00111	46.98	83.41	0.56	215125_s_at

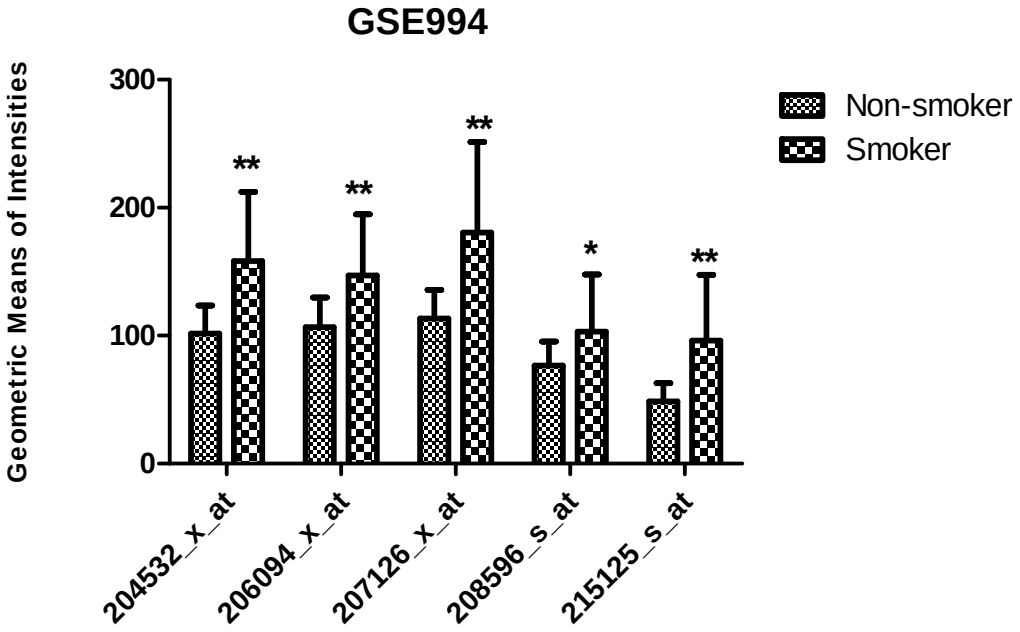

Figure 4.19 UGT1A expression in airway epithelium of smokers and non-smokers according to GSE994 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

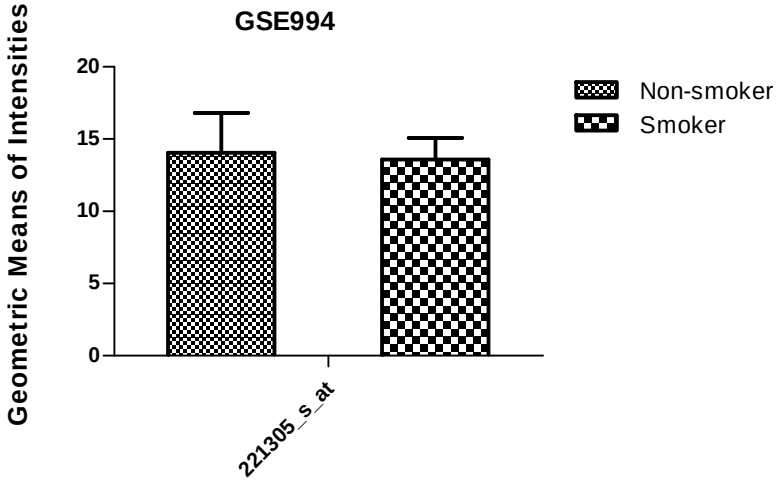

Figure 4.20 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of smokers and non-smokers according to GSE994 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

Former vs Current vs Never Comparison

When the arrays were grouped according to their status of smoking, there were 3369 genes that were differentially expressed with significance at $p<0.05$ among former, current and never smokers. In this list there were UGT1A genes with FDR levels less than 0.05 **(Table 4.17).** The results indicated that current smokers had relatively and significantly higher expression when compared with either the former or never smokers suggesting that the effects of smoking on UGT1A expression were reversible **(Figure 4.21).** Furthermore, the upregulation of genes do not include UGT1A8-1A9 genes as implicated in the graph since this probeset was not detectably measured **(Figure 4.22).**

Table 4.17 The geometric mean of intensities corresponding to common UGT1A probesets in GSE994 ($p<0.05$)

Parametric p-value	FDR	Current Smoker (Geom mean of intensities)	Former Smoker (Geom mean of intensities)	Never Smoker (Geom mean of intensities)	Probe set
< 1e-07	< 1e-07	169.16	117.21	99.28	204532_x_at
1.00E-06	0.000237	156.06	112.28	104.19	206094_x_at
< 1e-07	< 1e-07	194.42	123.61	111.33	207126_x_at
0.0002031	0.0142	106.21	75.8	74.38	208596_s_at
< 1e-07	< 1e-07	102.5	56.51	46.98	215125_s_at

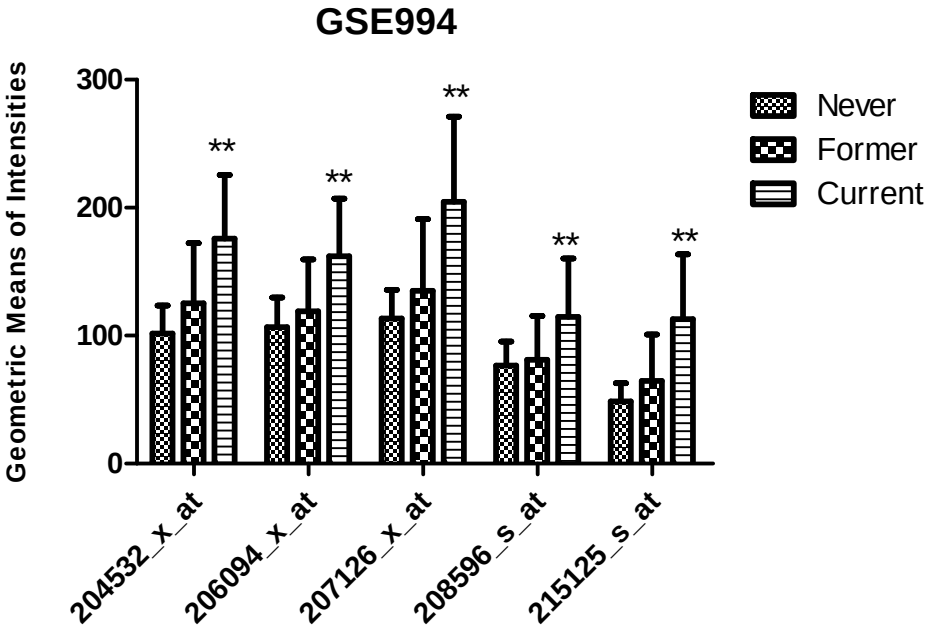

Figure 4.21 UGT1A expression in airway epithelium of current, former and never smokers according to GSE994 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

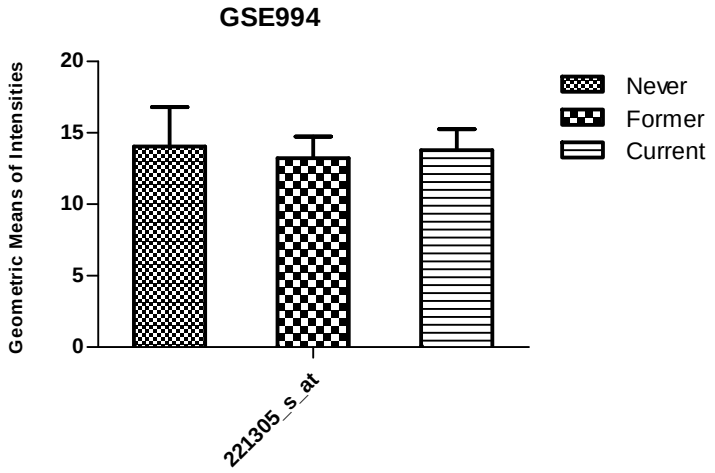

Figure 4.22 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of current, former and never smokers according to GSE994 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

GSE10072: Lung adenocarcinoma cells

This study was different among the others in that it focused on the lung adenocarcinoma cells [91]. There were 58 tumor tissue derived cells and 49 non-tumor tissue derived cells that were analyzed. These cells were isolated from 28 current and 26 former smokers in addition to 20 never smokers. RNA from the samples was run on [HG-U133A] Affymetrix Human Genome U133A Array. A comparison between non-smokers and smokers showed that no UGT1A genes were significantly changed ($p<0.05$). Furthermore, a comparison between former and current smokers showed that no UGT1A genes were significantly changed, as well ($p<0.05$). However, the dataset includes tumors and non-tumor adjacent samples from the same patient thus a paired significance analysis might also be performed in the future.

GSE19027 Human airway bronchial epithelial cells

This study defined the genes that may identify the transition to lung cancer. Therefore, they analyzed smokers and non-smokers with and without lung cancer in addition to healthy individuals [92]. Since the arrays belonging to healthy individuals are derived from GSE994, and to prevent re-use of the same samples, we have excluded these arrays from analysis and normalized the rest. Moreover, one array was excluded due to an error while importing the data (GSM470852_s54p103B1.CEL).

The chip type in this dataset is [HG-U133A] Affymetrix Human Genome U133A Array. There were 3 never smokers, 17 current smokers without cancer, 8 former smokers without cancer, 9 current smokers with cancer and 12 former smokers with cancer analyzed. The samples isolated were bronchial epithelial cells obtained via bronchoscopy.

The class comparison among smokers and non-smokers were done by grouping the samples into two as healthy and persons with cancer.

Comparison among healthy individuals

With significance level of $p<0.05$ there were 5687 genes differentially expressed among which there were UGT1A genes **(Table 4.18).** In all common UGT1A probesets, a significant increase in current smokers was observed while the expression in former and never smoker were similar to each other (FDR<0.16**;**

Figure 4.23). When the probeset specific for UGT1A8-1A9 genes was surveyed, no significant fold-difference was observed **(Figure 4.24).**

Table 4.18 The geometric mean of intensities corresponding to common UGT1A probesets in GSE19027 ($p<0.05$)

Parametric p-value	FDR	Current Smoker (Geom mean of intensities)	Former Smoker (Geom mean of intensities)	Never Smoker (Geom mean of intensities)	Probe set
0.0022436	0.101	364.66	236.63	209.92	204532_x_at
0.0119296	0.125	366.16	248.11	243.97	206094_x_at
0.0023689	0.101	402.23	241.46	245.55	207126_x_at
0.0309416	0.164	195.13	129.19	115.13	208596_s_at
0.0044015	0.104	180.44	93.93	88.23	215125_s_at

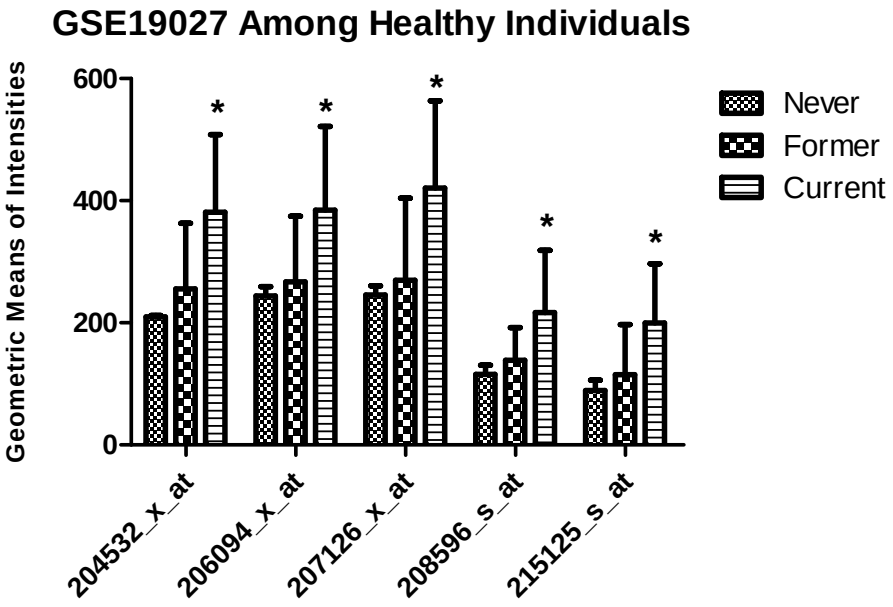

Figure 4.23 UGT1A expression in airway epithelium of healthy current, former and non-smoker according to GSE19027 (* $p<0.05$,**$p<0.001$). Results were given as geometric mean ± SD.

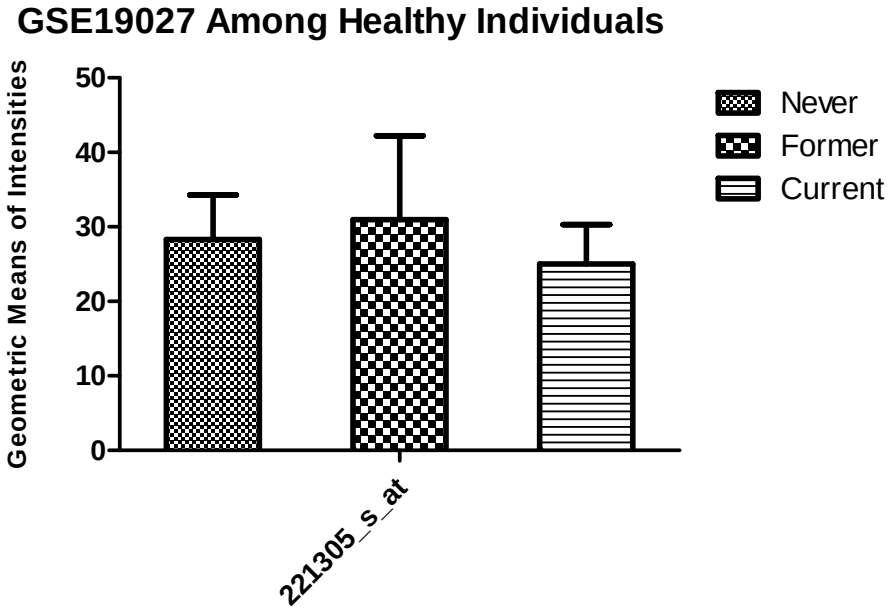

Figure 4.24 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of healthy current, former and never smokers according to GSE19027 (* $p<0.05$, **$p<0.001$). Results were given as geometric mean ± SD.

Non-smoker vs Smoker Comparison among healthy individuals

With a significance level of $p<0.05$ there were 763 genes that were differentially expressed among which there were no UGT1A genes. It was an unexpected result; however, it may have stemmed from the relatively small number of samples in non-smokers' group.

Comparison among cancer patients

With a significance level of $p<0.05$ there were 910 genes differentially expressed among which there were UGT1A genes **(Table 4.19)** although the FDR level was high possibly due to low sample size. A fold change of 1.5 was observed among current and former smokers with lung cancer **(Figure 4.25).** The high and significant difference between current and former smokers was not due to expression changes in UGT1A8-1A9 **(Figure 4.26).**

Table 4.19 The geometric mean of intensities corresponding to common probesets in GSE19027 ($p<0.05$)

Parametric p-value	FDR	Current Smoker Geom mean of intensities)	Former Smoker (Geom mean of intensities)	Fold-change	Probe set
0.0054486	0.917	378.69	221.66	1.71	204532_x_at
0.0150579	1	360.49	242.7	1.49	206094_x_at
0.0157856	1	407.62	236.4	1.72	207126_x_at
0.0427523	1	159.86	91.67	1.74	215125_s_at

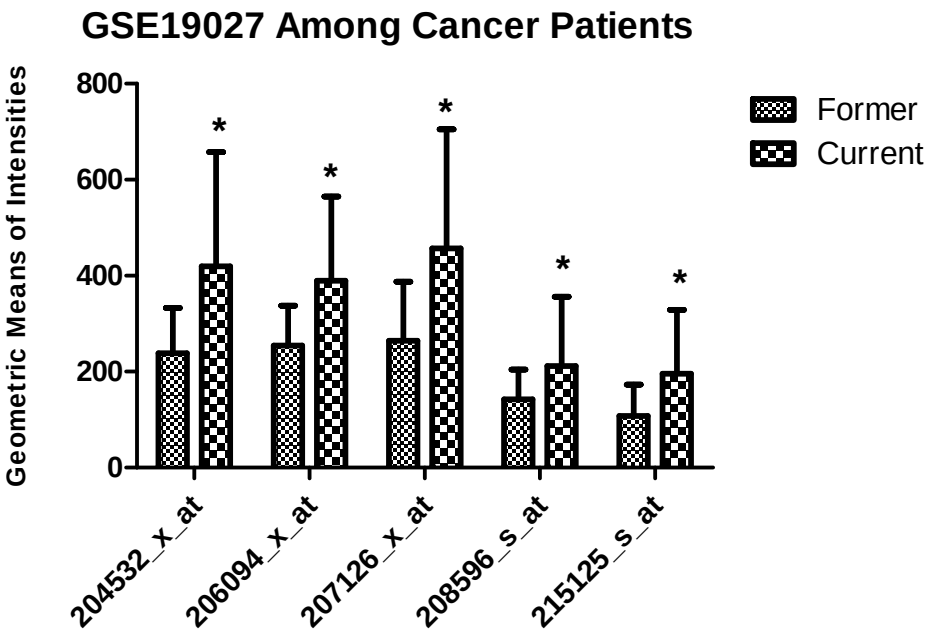

Figure 4.25 UGT1A expression in airway epithelium of current and former smokers with lung cancer according to GSE19027 (* p<0.05,**p<0.001). Results were given as geometric mean ± SD.

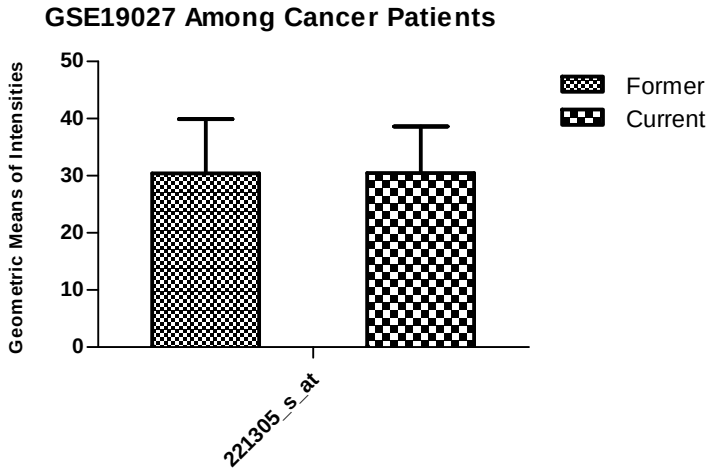

Figure 4.26 UGT1A8 and 1A9 (221305_s_at) expression in airway epithelium of current and former smokers with lung cancer according to GSE19027 (* p<0.05, **p<0.001). Results were given as geometric mean ± SD.

GSE3212: Human alveolar macrophages

In this dataset, the aim was to comprehend the gene expression changes in alveolar macrophages in healthy smokers and non-smokers. Cells were obtained by bronchoalveolar lavage [93]. This dataset included RNA from 5 healthy smokers and 5 non-smokers; which were hybridized on [Hu6800] Affymetrix Human Full Length HuGeneFL Array. When the data were analyzed there were 283 genes differentially expressed with a significance value of $p<0.05$; among which there was no UGT1A probeset. There was considerable expression of common UGT1A probesets though. There are studies proving the presence of UGT1A activity and their inducibility by polycyclic aromatic hydrocarbon in alveolar macrophages [94]. However, our results proposed that their levels are not changed by smoking.

GSE8823: Human alveolar macrophages

GSE 8823 focused on alveolar macrophages of healthy smokers and non-smokers isolated by bronchoalveolar lavage [95]. The platform used was [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. Dataset combined samples of 11 non-smokers and 13 smokers. BRB-analysis showed that among 6033 genes, no UGT1A genes were significantly changed ($p<0.05$) even though they are expressed.

GSE13896: Human alveolar macrophages

This dataset was generated to ask the question whether alveolar macrophage polarization was related with smoking and COPD [96]. The arrays carrying samples from patients with COPD were excluded from the analysis. In addition, in this dataset, there were arrays showing redundancy with GSE8823. Those were excluded as well. The platform used was [HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Array. The class comparison among smokers and non-smokers showed that there was no UGT1A gene that was differentially expressed ($p<0.05$). It is important to note that common probesets detected a considerable level of expression.

GSE12585: Human peripheral lymphocytes

This subset series was part of a larger study, collected under the name of GSE12587. The superseries GSE12587 combined both *in vivo* and *in vitro* data. *In vivo* data were composed of human peripheral lymphocytes isolated from smokers. *In vitro* data, on the other hand, was generated by culturing the isolated human peripheral lymphocyte cells with or without smoke-condensate. We will be focusing on GSE12525, the subseries with *in vivo* data. The basic comparison underlying this study was between heavy and light smokers. The platform used was [HG-U133A] Affymetrix Human Genome U133A Array and there were 10 heavy and 13 light smokers attending this study. Among 922 genes, no UGT1A genes were identified as significantly changed ($p<0.05$).

When we analyzed this array, we could not see a significant change in UGT1A expression although there is considerable expression. The reason may be that UGT1A expression may not be that responsive to the level of smoking.

4.2 Statistical analysis of UGT1A probesets from SW620 Nicotine Treatment Experiment

Previous studies performed by Onur Kaya [78] showed that nicotine increased the cell survival as evidenced by an MTT assay even at a lower dose of 10nm but was effective also in 1 and 10μM. A microarray experiment was also performed [78] and the results of these arrays were now re-analyzed in the context of UGT1A probesets using BRB tools. When the probesets assigned to UGT1A locus were surveyed individually, it was seen that they were upregulated significantly in serum starved and nicotine treated cells ($p<0.05$; and FDR<0.05 except for 221305_s_at) **(Table 4.20).** Furthermore, this upregulation could not be observed in cells that were nicotine treated and grown at 10% FBS containing medium **(Table 4.21).** At this point, verification of the data and more detailed analysis of the results were needed to speculate that nicotine induced proliferation and UGT1A expression under serum starvation.

Table 4.20 The geometric mean of intensities corresponding to common UGT1A probesets in serum starved and nicotine treated SW620 cells according to the microarray performed in our lab

Parametric p-value	FDR	0.1% FBS Control	0.1%FBS 1uM Nicotine	Fold-change	Probe set
0.0002279	0.0102	348.45	639.51	0.54	204532_x_at
0.0006852	0.0195	252.23	444.32	0.57	206094_x_at
0.0002348	0.0103	328.15	577.7	0.57	207126_x_at
1.30E-06	0.000397	189.43	565.42	0.34	208596_s_at
5.10E-06	0.00102	107.11	385.29	0.28	215125_s_at
0.0357341	0.201	25.87	19.68	1.31	221305_s_at

When nicotine treated cells from serum starved and serum replenished group were analyzed using the BRB Array Tools class comparison tool, it became evident that UGT1A expression induction was unique to nicotine treated starved SW620 cells (**Table 4.21**).

Table 4.21 The geometric mean of intensities corresponding to common UGT1A probesets in 48h Nicotine-treated SW620 cells according to the microarray performed in our lab

Parametric p-value	FDR	0.1% FBS Control	0.1%FBS 1uM Nicotine	10% FBS Control	10% FBS 1uM Nicotine	Probe set
4.77E-05	0.00113	348.45	639.51	255.26	267.17	204532_x_at
6.38E-05	0.00135	252.23	444.32	193.44	182.78	206094_x_at
2.30E-06	0.000163	328.15	577.7	222.69	213.72	207126_x_at

2.00E-07	3.38E-05	189.43	565.42	150.35	145.34	208596_s_at
2.20E-06	0.00016	107.11	385.29	65.8	63.15	215125_s_at

For further demonstration of how common UGT1A probesets were changed drastically when serum starved cells are treated with 1μM nicotine, we have plotted a dot graph **(Figure 4.27).** Furthermore, a heatmap was plotted using Cluster 3.0 and Treeview **(Figure 4.28).**

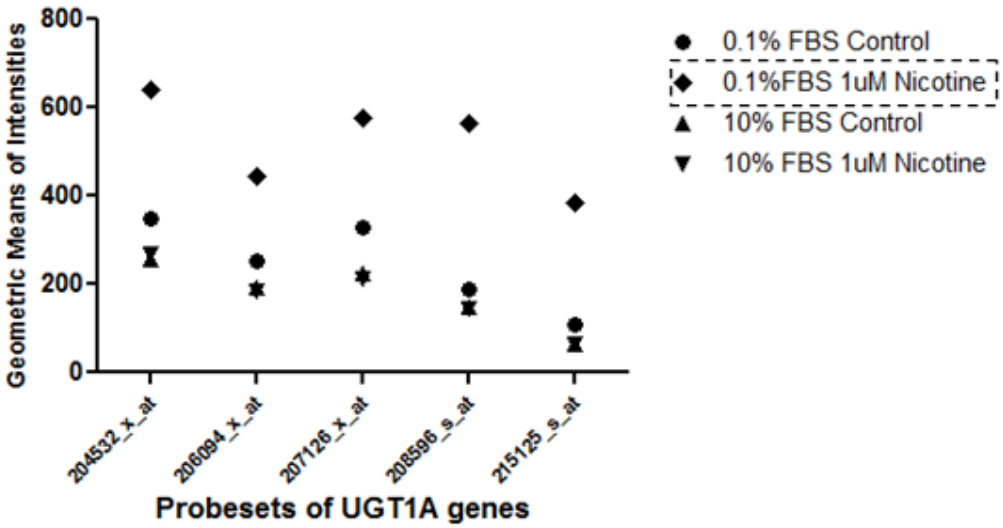

Figure 4.27 Mean intensities of each UGT1A probeset in 48h Nicotine-treated and untreated SW620 cells (starved or serum-replenished).

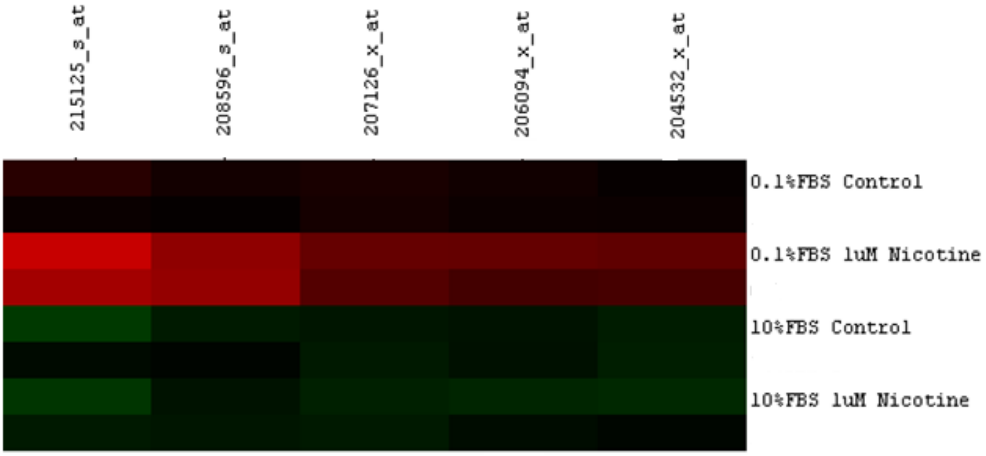

Figure 4.28 Treeview image of UGT1A probesets in 48h nicotine-control treatment in SW620 experiment ($p<0.001$). Green and red colors indicate a reduction and induction of expression, respectively, when each gene is normalized to its median.

4.3 Microarray Validation by qRT-PCR

4.3.1 Melt Curve Analysis

At the end of each qRT-PCR run, it is essential to check the melt curves and agarose gels for the validity and reliability of each reaction. We have analyzed each PCR run as below before inclusion in analyses:

- Dissociation curve of each sample should be checked. Each sample should present a curve specific to a particular amplicon centered on the same axis between 80°C-95°C, and indicating a Tm of a double strand DNA with sizes of at least 80bp. Samples with an additional peak of primer-dimers are included only if the dimer curve is negligible relative to the target curve.

- Each sample should be run on Agarose gel and checked with respect tothe presence of the expected target amplicon and the relatively higher intensity of the target band to other bands, if any.

Depending on these parameters, each PCR reaction was individually evaluated and included in the analyses, accordingly (see for examples: Figures 4.29-4.31).

As **Figure 4.29** shows, all the results in this reaction have a concrete dissociation peak on one axis.

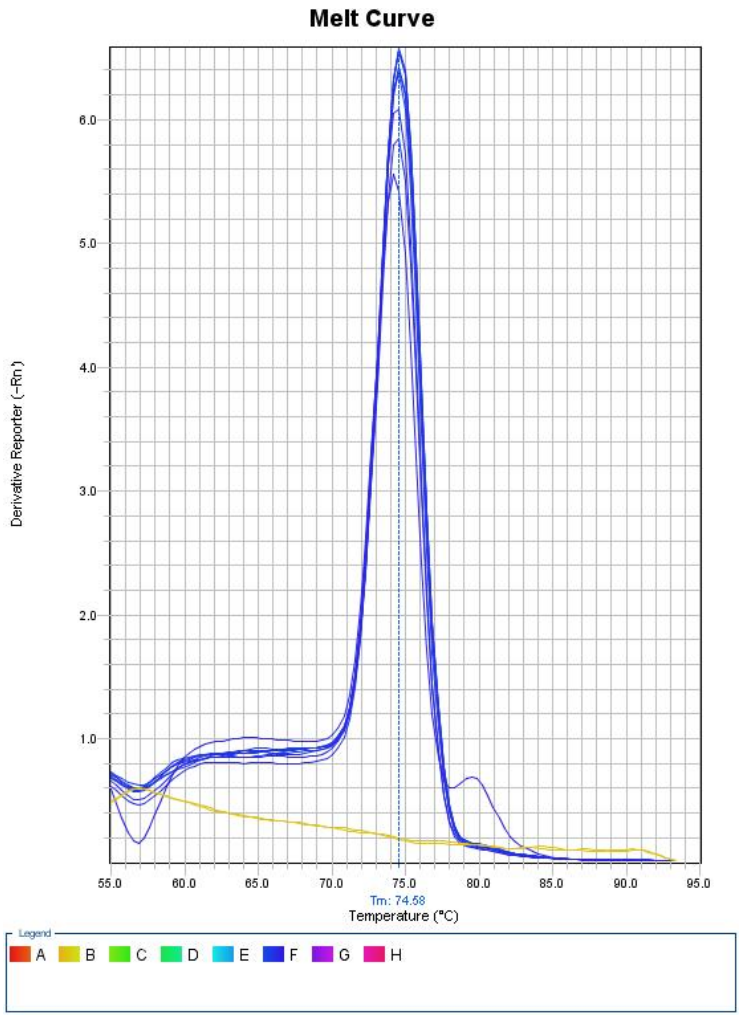

Figure 4.29 A demonstration melt curve belonging to samples that all pass the inclusion criteria. Blue curves indicate the samples while red curve belongs to negative control.

Figure 4.30 shows a gel photo including a sample with higher intensity band for primer-dimer. This sample was excluded, as expected.

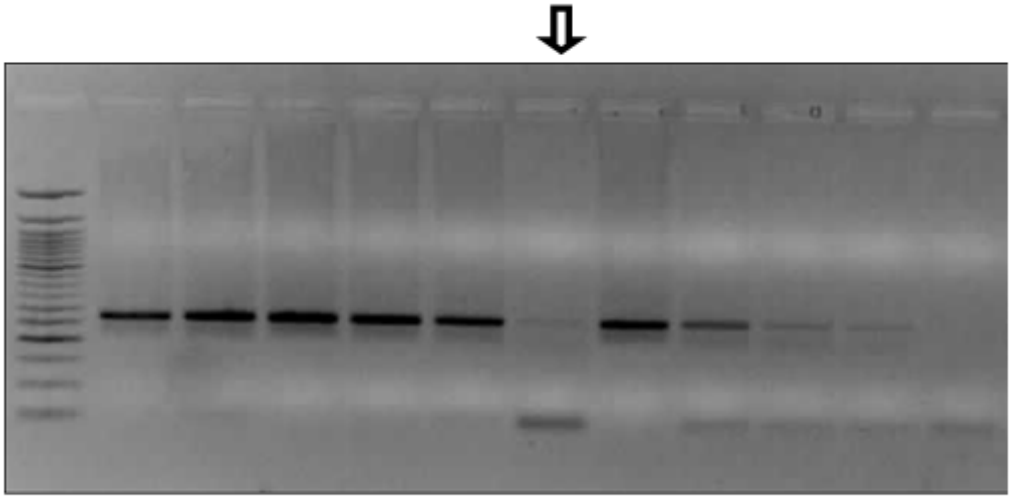

Figure 4.30 Arrows indicate the samples to be excluded for analysis in a demonstration gel photo.

Figure 4.31 demonstrates a reaction with some samples that are excluded. The sample with longer size, as indicated with a band higher than the other was excluded, as expected. Furthermore, the sample with additional band was excluded, too, since it was not possible to assign which band was responsible for Ct value detected.

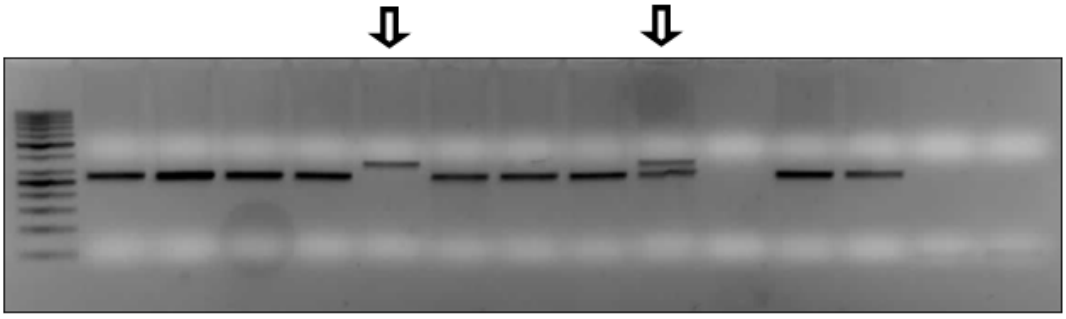

Figure 4.31 Arrows indicate the samples to be excluded for analysis in a demonstration gel photo.

4.3.2 Primer Efficiencies

The expression of UGT1A isoforms was studied using qRT-PCR analysis based on primers reported by Nakamura *et al.* (2008)[20]. Before use, the primer efficiency of each gene was calculated. Since efficiency calculations require at least five Ct values

in a row, the reactions were run with cDNA in which there are abundant expression of the genes. cDNA profile was determined according to Nakamura *et al.* (2008)[20]. E values for each primer couple were given in **Table 4.22.**

Most likely because amplicon size is large or simply due to its low expression the efficiency calculated for UGT1A7 could not be reliably assessed. That's why; its E value was set as 2. New primer pairs should be tried to assess the expression of UGT1A7 gene.

Table 4.22 Efficiency values calculated for each primer couple with the same kit used.

Gene name	Efficiency (E)	cDNA used for efficiency reactions
UGT1A1	1.86	MCF-7
UGT1A3	1.99	A pool of HepG2, Huh-7,Caco-2, PLC, NCIH69, SNU-475 cDNA
UGT1A4	2.28	SW620
UGT1A5	2.19	SW620
UGT1A6	1.86	MCF-7
UGT1A7	2*	-
UGT1A8	1.99	A pool of HepG2, Huh-7,Caco-2, PLC, NCIH69, SNU-475 cDNA
UGT1A9	2.03	A pool of HepG2, Huh-7,Caco-2, PLC, NCIH69, SNU-475 cDNA
UGT1A10	1.95	A pool of HepG2, Huh-7,Caco-2, PLC, NCIH69, SNU-475 cDNA
ANLN	2.1	SW620

MKI67	2.27	SW620

4.3.3 Selection of a Reference Gene for Nicotine Treatment of Starved Cells

As indicated in the paper of Pfaffl *et al.* (2004) [45], for a successful analysis with relative expression calculations in qRT-PCR, the stability of internal standards are important. Taking this criterion into account, we have determined the expression levels of five potential internal standard candidates, which have been used as reference genes in other studies. GAPDH is the most utilized reference gene in molecular biology. PPIA gene was shown to be a reliable standard gene for breast cancer tissues [97]. Furthermore, some ribosomal proteins (RPL) have been shown as the best housekeeping gene in mice exposed to oxidative stress [98].

As can be seen in **Figure 4.32**, the expression levels of GAPDH were very close in both the control and nicotine treated samples, indicating its strength as a reference gene. In addition, Ct values of PPIA in both samples were not divergent, either. On the other hand, ribosomal proteins were consistently downregulated under nicotine treatment, which raises an interesting point that warrants further study. Accordingly, GAPDH and PPIA were found to be stable and thus were considered as two probable endogenous references for serum starved and nicotine treated SW620 cells. Based on this observation, we have normalized the expression of target genes to GAPDH in each sample. Future analyses might consider using a combination of reference genes.

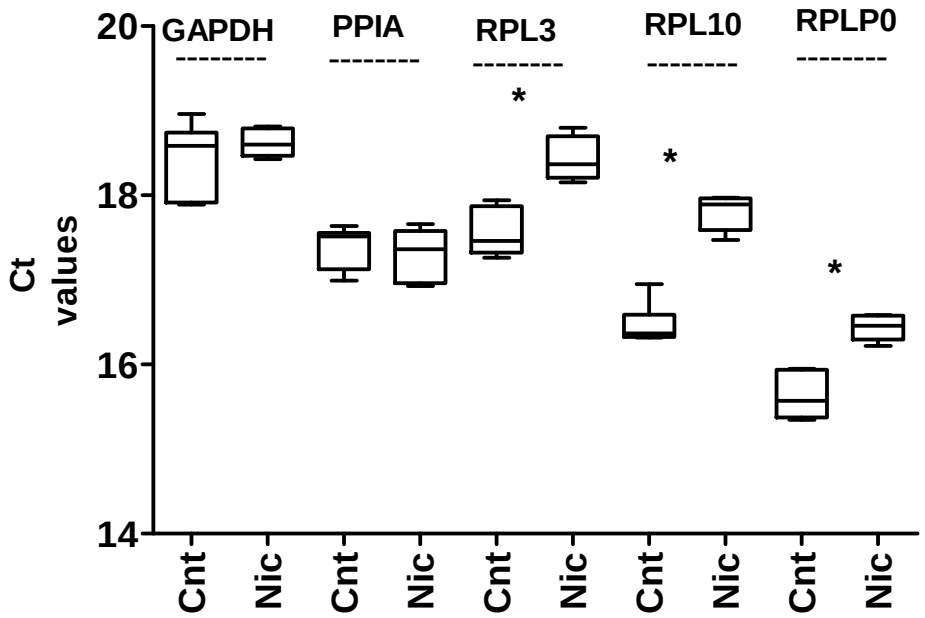

Figure 4.32 Expression values of GAPDH, PPIA, RPL3, RPL10, RPLP0 in serum starved SW620 cells. Results were given as Ct values and asteriks indicated significance at $p<0.05$ level.

4.3.4 qRT-PCR results as analyzed by the modified delta-delta-ct method

With the purpose of further verifying the results obtained from microarray analysis as shown above, we performed a series of quantitative real-time PCR (qRT-PCR) experiments.

As a starting point, two genes representatives of growth induction were chosen to be analyzed. In case when ANLN and MKI67 mRNA levels were upregulated significantly by nicotine, this would mean that cells were triggered for proliferation.

When cDNA of cells that were incubated with different concentrations of nicotine in the presence of 10% FBS, these two genes did not show any alterations in expression suggesting that only starved cells responded to nicotine by increasing cell proliferation response **(Figure 4.33 and Figure 4.34)**. This was confirmatory of the MTT experiments done previously[78].

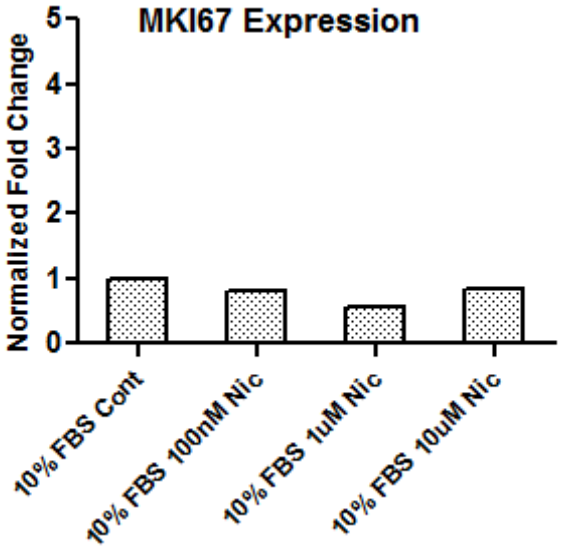

Figure 4.33 Analysis of cell proliferation in Nicotine-treated SW620 cells in 10%FBS media via Quantitative RT-PCR of MKI67 levels.

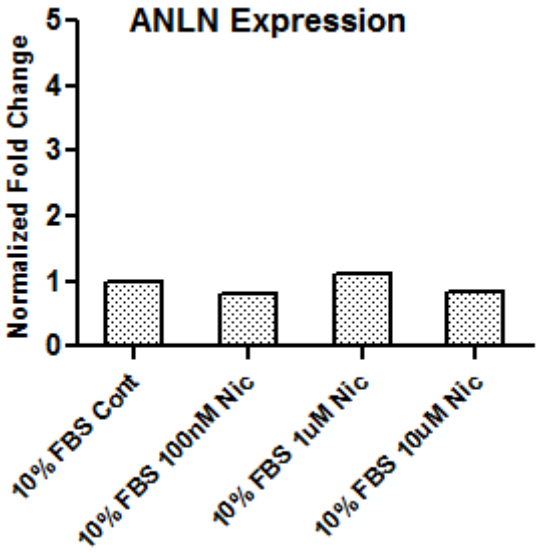

Figure 4.34 Analysis of cell proliferation in Nicotine-treated SW620 cells in 10%FBS media via Quantitative RT-PCR of ANLN levels.

All together, these showed that cells grown under normal conditions were not responsive to increasing concentrations of nicotine within 48h.

On the other hand, three sets of cDNA from cells that were either serum starved or serum starved and 1μM Nicotine treated were analyzed in terms of ANLN and MKI67 levels. A significant increase in the expression of both genes were observed when cells were exposed to 1μM Nicotine ($p<0.05$) **(Figure 4.35).**

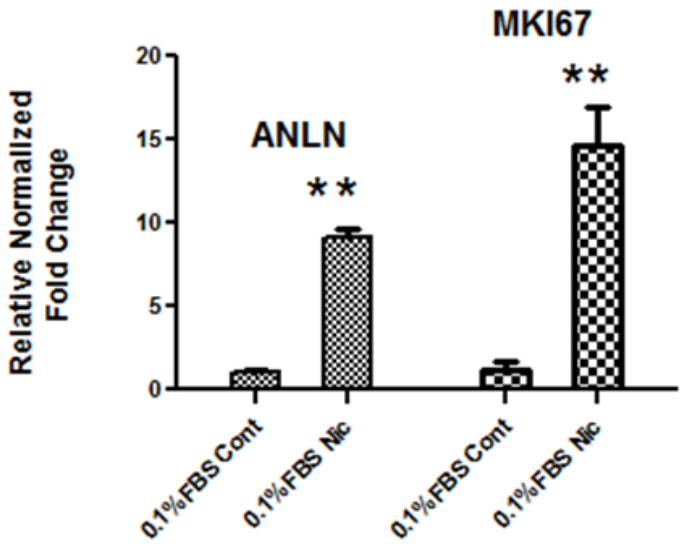

Figure 4.35 Analysis of cell proliferation in Nicotine-treated SW620 cells in 0.1%FBS media via quantitative RT-PCR of ANLN and MKI67 levels. Results were given as mean ± SD (DF=5, F-value= 541.41, $p<0.001$ for ANLN, DF=5, F value=79.91, $p<0.001$ for MKI67).

To sum up, we have seen that SW620 cells kept in the presence of 10% FBS did not respond to changing nicotine levels in terms of proliferation [78] and this also was now verified at the transcriptomic level using qRT-PCR on a dose-curve of nicotine. Moreover, when cells were serum starved 1μM nicotine upregulated the genes related with growth induction. ANLN and MKI67 have been shown to be related with proliferation in literature [99-100]. This picture was consistent with what our microarray analyses have shown. From this point, we expected to see the significant change in glucuronidation genes as the BRB analyses have pointed out based on the probeset expression values. In addition, it was important to ask whether the deregulation of UGT1A locus occurs in an isoform-specific manner. Isoform specificity could not be assessed using microarrays in our case thus a more specific qRT-PCR study has been essential part of our study.

The same cDNA set used in ANLN and MKI67 expression experiments were utilized for UGT1A expression determination. The expression levels of each gene were normalized according to the endogenous levels of GAPDH mRNA and relative fold changes in each gene are represented with two bar graphs as shown below:

qRT-PCR with UGT1A isoforms in serum deprived cells showed there was a change in expression levels of all UGT1A isoforms to relative degrees **(Figure 4.36 and Figure 4.37).** However, while some of the changes were statistically significant, some were not. In addition, most of the results belonging to UGT1A8 primer fail to pass the exclusion criterions. It may be because there are PCR inhibitors for UGT1A8 in SW620 cDNA since the E value of UGT1A8 could be determined in a pool of cDNA from different cell lines. Therefore, their results could not be presented. For the assessment of statistical significance, we also considered multiple test correction criterions and calculated a Benferroni p-value, which was 0.00625. p values under 0.00625 were defined as significant.

According to **Figure 4.36**, UGT1A1, UGT1A3, UGT1A4 and UGT1A6 genes were the isoforms that were very significantly upregulated with 1μM Nicotine exposure in serum starved SW620 cells ($p<0.0065$). Relative fold changes of UGT1A1 and UGT1A3 were 10 and 13, respectively, and were statistically significant with p-values of $p<0.001$ (DF=5, F-value= 514.24 for UGT1A1, DF=5, F value=88.39 for UGT1A3). UGT1A6 upregulation was a significant increase with a fold change of 22 (DF= 5, F-value= 85.29 p-value=0.001) (**Figure 4.36**). Furthermore, UGT1A5 was another isoform that is being upregulated, but not significantly according to Benferroni correction (DF=5, F-value=19.92 p-value=0.011) **(**Figure 4.36**).** The upregulation of UGT1A4 was highest among the others with almost 90-folds; suggesting that it has a major role in drug metabolism carried out in SW620 cells (DF=5, F-value=49.83 p-value=0.002). In addition, UGT1A7 levels seemed to increase with nicotine but not significantly and with high variability **(Figure 4.37)** (DF= 3, F-value= 1.25 p-value=0.379). This gene requires new primer couples.

As a last note, UGT1A9 and UGT1A10 expression levels were very low as measured by the qRT-PCR experiments also complementing the findings from microarray probeset results. The changes in these genes were not significant **(Figure 4.37)**

(DF=3, F-value=0.02 p-value=0.903 for UGT1A9) (DF=5, F-value=1.09 p-value=0.356 for UGT1A10).

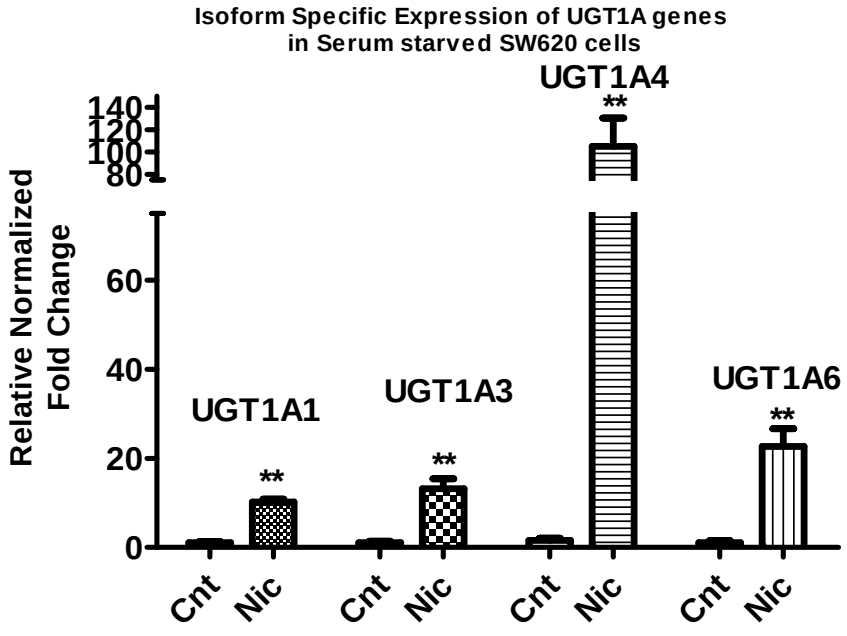

Figure 4.36 The expression levels of UGT1A isoforms normalized to GAPDH mRNA levels. Results were given as mean ± SD. (*indicates the significance corrected with Benferroni)

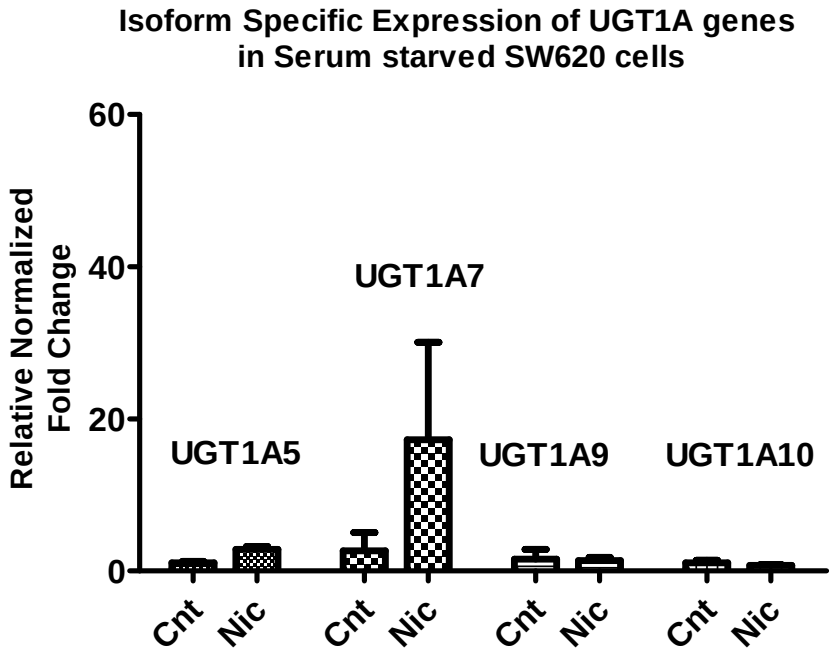

Figure 4.37 The expression levels of UGT1A isoforms normalized to GAPDH mRNA levels. Results were given as mean ± SD.

In order to verify that upregulation of UGTs as a response to nicotine treatment was only observed in cells that were serum starved, we have checked the expression levels of UGT1A6 in normally grown cells. The reason for choosing UGT1A6 is that it was one of the isoforms most affected under nicotine treatment. The result of this experiment demonstrated that UGT1A6 was overexpressed in normally grown SW620 cells with 48h nicotine treatment **(Figure 4.38**).

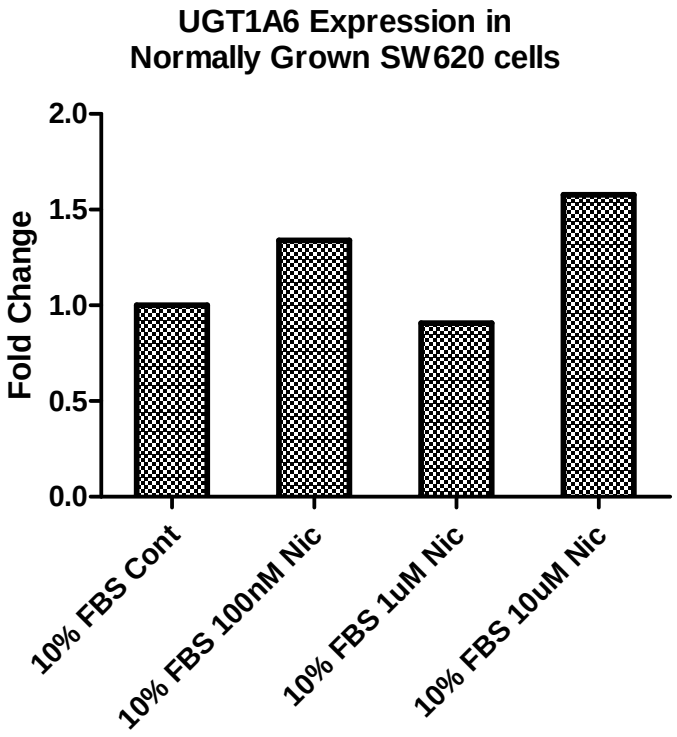

Figure 4.38 The expression level of UGT1A6 normalized to GAPDH mRNA levels in SW620 cells grown with 10%FBS-containing medium.

These results further confirm the microarray results, indicating that modulation of UGTs with 48h nicotine treatment is specific to serum-deprived SW620 cells.

4.4 *in silico* Promoter Analysis of UGT1A isoforms

Observing that with nicotine addition, only certain UGT1A isoforms were being upregulated while no change could be observed in particular isoforms has raised an important question. In addition, the magnitude of upregulation was different among the overexpressed isoforms. When these particular isozymes were surveyed in terms of their substrate specificity, nothing common for these isozymes could be found. Moreover, their relation with nicotine glucuronidation was absent, as well. At this point, an important question was asked regarding the association between UGT1A expression and sequence.

The starting point was analysis of Exon 1 and promoter region of 300bp. Each isoform was aligned via ClustalW software to see whether there was a clustering of the isozymes [101]. Accordingly, there was a grouping of genes into two based on the sequence similarity in the corresponding regions (**Figure 4.39**). The important thing

was that these two groups could well discriminate the genes deregulated with nicotine from the others.

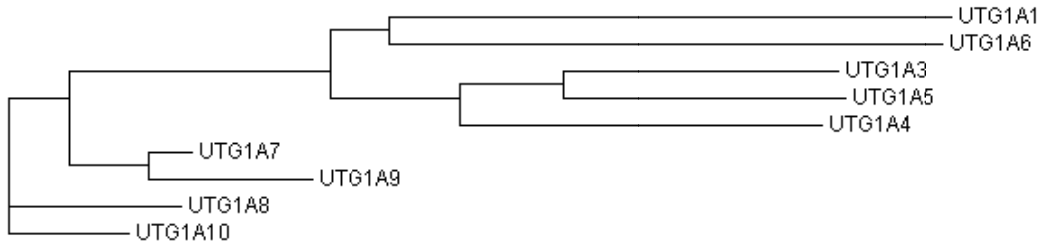

Figure 4.39 Alignment based on promoter region (300nt upstream) and Exon 1 of each UGT1A isofom

Furthermore, when the alignment was carried out with sequences within promoter regions, this clustering of genes was maintained; further supporting the relation between sequence and expression of UGTs (**Figure 4.40**).

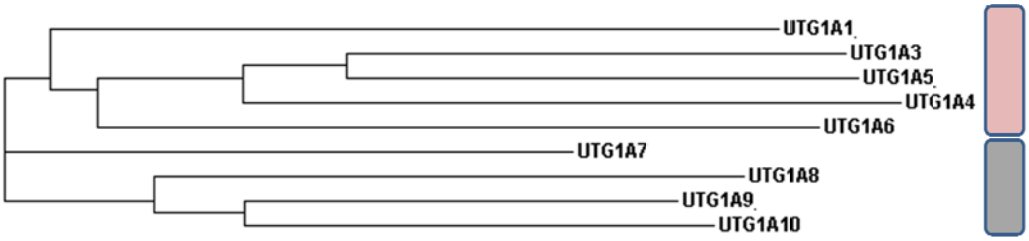

Figure 4.40 Alignment based on promoter region (300nt upstream) of each UGT1A isofom. Pink box indicates the genes that are upregulated with nicotine administration while grey box displayed the ones without change in gene expression.

The promoter regions carry certain motifs, enabling the binding of relevant transcription factor. Therefore, the presence of motifs, found in deregulated UGT1A isoforms and absent in the others could answer many questions. Based on this notion, we analyzed the promoter regions of all UGT1A isoforms with MEME software [102]. As shown in **Figure 4.41**, there are common and specific motifs, located in the promoters of each UGT1A isoform.

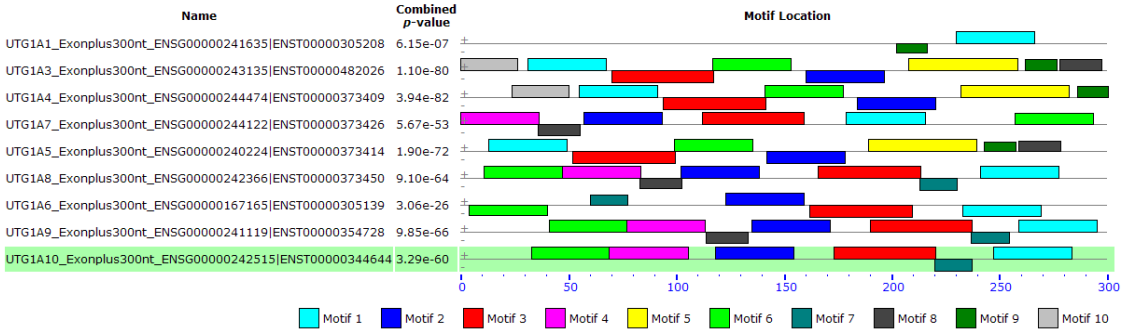

Figure 4.41 MEME result of promoter regions belonging to all UGT1A isoforms

These results suggest that expression and sequence of UGT1A isoforms might be associated and warrants further analysis.

5. CONCLUSION AND DISCUSSION

Here in this report, we showed that UGT1A isoforms were responsive to smoking and to nicotine, using *in vivo* public microarray data and our *in vitro* SW620 colon cancer nicotine exposure model, respectively. *In vivo* datasets from GEO repository were carefully selected and analyzed to assess whether smokers have higher UGT1A expression and also whether there is a difference between current and former smokers. For the first time we showed that, in multiple epithelial tissues, smokers significantly overexpress UGT1A gene clusters. For all microarray evaluation steps, a uniform bioinformatics tool and a set of protocols were used. Being integrated software which enables statistical analysis and visualization, BRB array tools has been chosen for data analysis of microarray results. Another reason for preference of this analytical software was that it had an extensive network allowing access to several genomic websites and analysis tools. BRB Tools also provides a raw p-value and its associated FDR value, allowing for multiple test correction. The use of Affymetrix microarrays with common probesets helped us to incorporate three different platforms.

We focused on arrays performed on healthy individuals but not the COPD patients since we cannot be sure that changes in UGT1A expression merely stem from smoking phenotype. However, a few cancer datasets also were analyzed. The overall scheme of arrays analyzed included a combination of number of arrays conducted in airway epithelium, oral and nasal mucosa, in members of immune system and 2 arrays with samples from lung adenocarcinoma.

The overall scheme from re-analysis of the above mentioned studies were summarized in **Table 5.1**, **Table 5.2**, **Table 5.3** and **Table 5.4**.

Table 5.1 Datasets run on various tissues isolated from healthy individuals

Array Name	Platform	Origin of tissue/cells	Smokers (n)	Non-smokers (n)	Citation	Result
GSE4635	Affymetrix Human Genome U133A Array	Large airway epithelium	5 (all current)	5	PMID: 19357784	UGT1A expression of smokers was significantly upregulated ($p<0.05$) by an average of 2 fold change constant as compared to non-smokers. Expression of UGT1A8-1A9 was very low.
GSE4498	Affymetrix Human Genome U133 Plus 2.0 Array	Small airway epithelium	12	10	PMID: 17115125 PMID: 19106307	Smokers have at least twice the expression of UGT1A isoforms when compared with non-smokers (FDR<0.05). UGT1A6 probeset detected a very low expression in non-smokers, which again was doubled in smokers**.** UGT1A8 and UGT1A9 probesets detected

						very low expression.
GSE10006	Affymetrix Human Genome U133 Plus 2.0 Array	Large airway epithelium	38	20	PMID: 18832735	Approximately three fold upregulation of UGT1A common probesets were observed in smokers. UGT1A6 was upregulated by less than two fold constant. However, 1A8 and 1A9 expressions were low and changes were not significant (FDR>0.05).
GSE17913	Affymetrix Human Genome U133 Plus 2.0 Array	Oral mucosa	39	40	PMID: 20179299	Common UGT1A probesets were significantly changed with less than two-fold upregulation at the FDR<0.05 in smokers. In addition, probesets annotated for UGT1A8-1A9 were upregulated (FDR<0.05).
GSE16149	Affymetrix Human Genome U133 Plus 2.0 Array	Buccal Mucosa	9	9	PMID: 20576139	One common probeset detected a very low expression, whose changes are not significant

						(p>0.05). UGT1A8-1A9 probeset levels were high but did not show difference among smokers and non-smokers.
GSE8987-Nose	Affymetrix Human Genome U133A 2.0 Array	Nose	7	8	PMID: 18513428	The common UGT1A probesets showed a less than two-fold upregulation when compared with non-smokers**.** UGT1A8-1A9 probesets were expressed, but not alter significantly.
GSE8987-Mouth	Affymetrix Human Genome U133A Array.	Mouth	5	5	PMID: 18513428	Although expressed, the expression levels of common probesets did not show a significant change as well as UGT1A8-1A9 probesets.
GSE3212	Affymetrix Human Full Length HuGeneFL Array	Alveolar Macrophages	5	5	PMID: 16520944	The probeset for UGT1A detected epxression at moderate levels;yet, there was not a change among smokers and non-smokers.

GSE8823	Affymetrix Human Genome U133 Plus 2.0 Array	Alveolar Macrophages	13	11	PMID: 18587056	Although expressed, the expression levels of common probesets did not show a significant change. UGT1A8-1A9 probesets detected very low expression.
GSE13896	Affymetrix Human Genome U133 Plus 2.0 Array	Alveolar Macrophages	22	13	PMID: 19635926	Although expressed, the expression levels of common probesets did not show a significant change. UGT1A8-1A9 probesets detected very low expression.

Table 5.2 Datasets utilizing samples from current, former and never smokers

Array Name	Platform	Origin of tissue/cells	Current Smokers (n)	Former Smokers (n)	Never smokers (n)	Citation	Result
GSE7895	Affymetrix Human Genome U133A Array	Bronchial epithelium	52	31	21	PMID: 17894889	There was a significant change in levels of common probesets and the highest value belonged to current smokers ($p<0.05$; FDR<0.05). UGT1A8-1A9 expression was very low and not significantly changed.
GSE994	Affymetrix Human Genome U133A Array	Intra-pulmonary airway epithelium	34	18	23	PMID: 15210990	All common UGT1A probesets except one (208596_s_at) were significantly changed at FDR<0.05 and highest in current smokers.

Table 5.3 Datasets utilizing samples from cancer and healthy patients

Array Name	Platform	Origin of tissue/cells	Current Smokers (n)	Former Smokers (n)	Never smokers (n)	Citation	Result
GSE10072	Affymetrix Human Genome U133A Array	Lung adenocarcinoma cells and non-tumor tissues	16 healthy, 24 tumor	17 healthy, 18 tumor	13 healthy, 16 cancer	PMID: 18297132	There was not a change in UGT1A expression among smokers and non-smokers regardless of their health status. The highest expression was detected in smokers with lung cancer ($p<0.05$).
GSE19027	Affymetrix Human Genome U133A Array	Bronchial epithelium	17 healthy, 9 cancer	8 healthy, 12 cancer	3 healthy	PMID: 20689807	When heathy individuals were analyzed, all common UGT1A probesets showed a significant increase in current smokers was observed while the expression in former and never smoker were similar to each other (FDR<0.16).

							A comparison among cancer patients showed that there was a fold change of 1.5 among current and former smokers with lung cancer.

Table 5.4 A dataset comparing heavy and light smokers

Array Name	Platform	Origin of tissue/cells	Heavy Smokers	Light Smokers	Citation	Result
GSE12585	Affymetrix Human Genome U133A Array	Lymphocytes	10	13	None	Common UGT1A probesets detected a considerable expression, which did not change among heavy and light smokers. The expression of UG1A8-1A9 was almost at basal level.

Overall, the re-analysis of publically available data showed us that the epithelial cells in close contact with exogenous compounds showed a response to smoking by induction of UGT1A expression. By the analysis of UGT1A8-1A9 specific probes, we hypothesize that either of these genes might not be expressed or detected by the given microarray probeset thus is hard to evaluate with respect to their contribution to drug metabolism in oral, nasal and airway epithelia. Even though the picture showed difference in two datasets (GSE16149, GSE8987) the discrepancy may result from the low number of samples used in these arrays. The reason why these datasets have contradicting results with all the other may be because UGT expression and glucuronidation could be restricted to certain parts of a tissue [103] and the later datasets have isolated the samples from unrelated part of the tissue. Moreover, the interindividual variance in UGT1A expression, which has been shown before [104-105] could be a reason for this contradiction.

Arrays with current smoking status of the patients demonstrated that UGT1A upregulation in airway epithelia can be reversible as it is in the case with former smokers.

The members of the immune system do not show a response to smoking in terms of UGT1A expression. Yet, their expression was detected in all datasets. In fact, there is a study revealing the expression of UGTs in bovine alveolar macrophages and bronchial epithelia [94]. Furthermore Tochigi (2005)[106] showed that UGT1A enzymes are induced in an isoform-specific manner in peritoneal macrophages in rat as a response to polycyclic aromatic hydrocarbons. It can be proposed that UGTs in immune system may be expressed in a constitutive manner but not in an inducible state. This in turn supports the notion that upregulation of UGT1A enzymes requires a close contact of cells with the xenobiotics and the participation of the corresponding tissue in drug metabolism.

When it comes to carcinoma cells; however, one microarray experiment presented a picture in which cancer cells were not responsive to smoking. On the contrary, the other array showed that there was a significant difference among current and former smokers with cancer. In order to comment on these contradicting results, the part of the tissue from which the samples were collected should be known precisely. The

reason is that, as stated above, drug metabolism could be carried out by restricted cells in a tissue. In addition, it has been demonstrated that drug metabolism enzymes were downregulated in certain cancer types, which in turn results in elevated resistance to chemotherapeutic drugs [107]. Furthermore, in some cancerous cells drug metabolism enzymes are upregulated as in the case of healthy individuals [92]. However, there is a common outcome of these two datasets, which is smoker cancer patients have the highest UGT1A expression ($p<0.05$; $FDR<0.05$). This picture is consistent with expectations, further supporting the relation between smoking and UGTs.

Previous studies reported higher risk of colon cancer with smoking [108-109]. In addition, Hübner *et al.* (2009)[77]showed that smoking upregulates UGT1A4 and UGT1A6 expression in small airway epithelia of healthy individuals [77]. Nicotine, on the other hand, is the major addictive component in cigarettes. Moreover, 30% of nicotine metabolism is merely carried out by UGT enzymes. Therefore, testing whether nicotine is one of the driving forces in the relationship between smoking and UGT expression is logical.

Wong *et al.* (2007)[110], has shown that nicotine upregulates proliferation of HT-39 colon adenocarcinoma cells in a dose-dependent manner [110]. With the purpose of experimenting whether nicotine had the same effect on SW620 colon adenocarcinoma cells, our group performed an MTT analysis [78]. What was observed is that nicotine triggered the proliferation of SW620 cells only when they are serum starved [111].

Based on this outcome, a microarray performed in our lab was re-analyzed exclusively for the UGT1A probesets used also for the re-analysis study in regard to smoking status. Functional analysis of the data has shown that glucuronidation was one of the pathways upregulated when SW620 cells were serum starved and treated with 1uM nicotine ($p<0.05$) but not when treated with 10% serum replenishment[78]. This suggested that increase in UGT1A activity was unique to nicotine although it is not known how nicotine induces UGT1A expression.

As mentioned before, there is a handicap of the genomic organization of UGT1A locus, disabling the analysis of the expression of each isoform by microarray.

Therefore, it was not possible to propose a specific isoform/isoforms that are being differentially regulated with nicotine and serum starvation treatments. Moreover, because UGT1A gene products share a high sequence similarity, either Northern Blot or Western Blot is not an ideal method to analyze their expression. Quantitative Real-time PCR analysis utilizing a well-designed unique primer pairs, on the other hand, is a robust approach to detect even the small changes in mRNA transcripts of each product.

With the advent of Real-time PCR technology, it is possible to quantify mRNA expression levels precisely in a given cell line, as a response to any physiological conditions in a very short time [59]. Real-time PCR is a highly preferred technique since it includes quite sensitive *Taq* polymerases, requires very small amounts of RNA template and does not require post-PCR modulations [46]. Nevertheless, the results of analysis and quantification can only be reliable when the markers of successful RT-PCR assay are completed.

Since no reference genes were stated for SW620 cells when they were nicotine treated we first studied a stable set of reference genes. Therefore we determined the expression of five internal standards and saw that GAPDH is the gene with most stable expression. Ribosomal proteins that are shown to be good reference genes in some other conditions failed in this case [98].

In addition, the efficiency values of each primer couple catalyzed by the enzyme that we used were determined individually. For this purpose, cDNA pools were generated by mixing different carcinoma cell lines and diluting them by 2 folds **Table 4.22**. Most of the efficiency results were close to 2, which were acceptable values. An efficiency value by UGT1A7 primer; however, could not be determined due to its low expression in cDNA pool analyzed, disabling the detection of five values in a row. Furthermore, a cDNA inhibitor may be present in some of the samples, disabling the annealing with these primers. Since the qRT-PCR primers should be designed to amplify small products, some of the primers we used might suffer from large amplicon inefficiencies. The primers used in this study were published in the literature and was used to make our results comparable with the previous ones and they have common reverse primers. New primer pairs specific to each isoform might

be studied for selected isoforms, e.g., UGT1A8 and UGT1A7, where amplicon size is large and/or in the presence of excessive dimer formation.

After running of each experiment, dissociation curves for each well were analyzed individually to check if there was a single peak. Furthermore, these peaks were expected to show up around 80°-90°C, indicating double stranded DNA target. Moreover, all the products were run on Agarose Gel electrophoresis to see whether negative controls were free of DNA and each sample has a unique band at the expected sizes. Overall, in most of our samples, both technical replicates were of good quality although in a few runs, one technical replicate of the sample had to be removed based on the exclusion criteria set previously. However, there were two cases in which a biological replicate had to be eliminated for analysis due to a detection failure in both replicates of either control or nicotine samples. These scenarios were observed with UGT1A7 and UGT1A9. These could stem from a technical error or due to large sizes of amplicons. What's more, we used the automatic set for treshold calculations determined by the ABI 7500 Software itself. Treshold values can be manually placed, as well. Indeed, we noticed that fold changes might change even minimally when changing the threshold.

Microarray experiments previously performed suggested that 1μM nicotine affects cell proliferation and induces a large scale transcriptomic response under starvation but not in the presence of serum[78]. 1uM nicotine has been selected since it represents the maximum amount of nicotine in smokers blood stream [112]. In the present thesis, we have checked the expression levels of two genes that are known markers of cell proliferation, ANLN and MKI6. Previously, it was shown that overexpression of ANLN was associated with growth of lung carcinoma both *in vivo* and *in vitro* [100]. In addition, it was speculated that ANLN was phosphorylated by PI3K/AKT pathway, which nicotine was known to activate [113-114]. Furthermore, a recent study showed that MKI67 were among the genes whose expression increased when human airway epithelial cells were exposed to tobacco toxins for short term [115]. Based on these, lack of upregulation of ANLN nor MKI67 with nicotine exposure in SW620 cells in the presence of serum points out that nicotine does not affect SW620 cells under 48h exposure unless when they are under stress. On the other hand, it is possible to see the statistically significant upregulation of both ANLN and MKI67 genes when cells

were treated with nicotine in presence of 0.1% FBS. This clearly shows that stress caused by quiescence enhances nicotine's effect on cells.

After verifying that there is an induction of growth with nicotine in serum starved cells, we desired to analyze the expression levels of each UGT1A isoform in those cells. Significantly deregulated isoforms after Bonferroni correction were: UGT1A1, UGT1A3, UGT1A4 and UGT1A6 ($p<0.05$). However, UGT1A5, UGT1A7, UGT1A9, UGT1A10, were not significantly changed using the primers previously published. These results were novel in that UGT1A expression has not been analyzed in detail in SW620 cells. In addition, specific circumstances (e.g. serum starvation) promoting nicotine's short term effects have not been shown before, as well. It should be noted that the expression values for UGT1A7, UGT1A9, and UGT1A10 were quite low thus it is difficult to clearly comment on their status. Therefore, new primers could be designed to confirm these findings. In addition, qRT-PCR optimization might be useful to reduce the dimer formation and increase duplicate reliability in the qRT-PCR runs.

Furthermore, a qRT_PCR run with normally grown SW620 cells for UGT1A1 and UGT1A6 expression analysis showed that there was no change in their expression with increasing concentrations of nicotine. All together, these data has certain consequences, which can be stated as:

- UGT1A isoforms did not respond to varying concentrations of nicotine at 2 days in the presence of serum.

- Nicotine exposure elevated mRNA levels of particular UGT1A isoforms only in starved SW620 cancer cells .

- There was a fine tunning effect of nicotine on UGT1A expression, modulating the levels in upregulation of each isoform.

In broader context, it can be stated that:

- The drug responsiveness of UGT enzymes was promoted under serum starvation conditions.

- Since modulations in UGT1A enzyme levels might affect the response to other drugs, smoking status of cancer patients together with enzyme expression patterns might be evaluated before use of therapeutics.

We have confirmed the expression of most of the UGT1A genes in a colon cancer cell line as indicated in previous reports [2,20] In addition, UGT1A5 was detected to be the one with the highest expression in SW620 colon adenocarcinoma cells.

UGT1A4 and UGT1A9 are the isoforms thought to be the principle isoforms for nicotine clearance by glucuronidation [43]. These reactions predominantly take place in liver. Studies with intestinal and liver microsomes show that N-glucuronidation of nicotine in intestine constitutes less than 2% of the reactions taking place in liver [72]. This is strong evidence pointing out that the isoforms responsible for nicotine glucuronidation are not expressed highly in intestine. In addition, UGT1A7, A8 and A10 are considered to be the potential enzymes of colonic glucuronidation since they are not expressed in liver [21]. However, their expression status in SW620 cells has not been identified before. Re-analysis of microarray datasets with samples with known smoking status indicated that probeset specific to UGT1A9 was expressed in low levels in bronchial and airway epithelial tissue and can not be assessed accurately. Similarly, UGT1A9 does not seem to be affected with nicotine administration in SW620 cells. These suggest that either new primers needed to be studied in this context or changes might be due to posttranscriptional modifications.

The isoforms that are upregulated by nicotine within 48h in serum starved SW620 cells belong to bilirubin-type UGT1A enzymes except UGT1A6. Nevertheless, a common substrate for all these isoforms is absent. There are potential hypotheses to understand the coordinate upregulation of these particular isozymes. First of all, fine tuning by nicotine may result from the sequence similarity in that group of exons. In deed, *in silico* promoter analysis of all isoforms showed that the one group of exons shares a high sequence similarity, including all the genesupregulated in our experiments. In addition, this may stem from a simultaneous regulation by the same regulator. This hypothesis can be explained as follows:

UGT enzymes are expressed in their host tissues both in a constitutive and inducible manner. The constitutive expression of UGTs is carried out by specific regulatory

proteins in that tissue. As an example, hepatocyte nuclear factor 1 (HNF1) is necessary for whole-time expression of UGT1A1, UGT1A6, UGT1A7 in liver [116-117].

By comparison, there are regulatory receptors responding to hormones, xenobiotics and stress signals mediating UGT1A transcription and enzyme capacity [118]. Following a ligand binding, these regulators bind to the corresponding response elements (antioxidant response elements (ARE) or/and xenobiotic response element (XRE)) located in the promoter regions of UGTs inducing their transcription. Among those, nuclear receptors constitutive androstane receptor (CAR) and pregnane X receptor (PXR) are found [119-120]. In addition, the nuclear receptor peroxisome proliferator-activated receptor α (PPARα), the aryl-hydrocarbon receptor (AhR) and nuclear factor erythroid 2-related factor 2 (Nrf2) control UGT1A expression at trancriptional level, as well.

Senekeo *et al.* has shown that UGT1A1, UGT1A3, UGT1A4 and UGT1A6 can be indcucable through PPARα regulatory element in human liver microsomoses [121]. In addition, Barbier *et al.* (2003)[122] showed that UGT1A9 transcription was induced by PPAR α ligands.

Sugatani *et al.* (2008)[123] claim that CAR ligands activate UGT1A1 isoform. In addition, it has been shown that there are two CAR activators regulating UGT1A5 and UGT1A6 transcription in addition to UGT1A1's in only liver [124].

UGT1A1, UGT1A3, UGT1A4 and UGT1A6 are known to be regulated by aryl hydrocarbon receptor (AHR) agonists [31,125-127]. Furthermore, UGT1A9 has been shown to be induced by AHR-agonists in Caco-2 cells [128]. These suggest that AHR and PPARα might underlie the effects of nicotine and could be tested experimentally in future studies. These outcomes also are interesting since there have been reports correlating AHR with upregulation of proliferation. One study suggests that when overexpressed, AHR elevates E2F/DP2 activation, promoting cell proliferation in human lung carcinoma cell A459 [129]. In addition studies showing that mice with constitutive AHR expression develop stomach tumors [130]. Considering nicotine has a proliferative effect on starved SW620 cells, the relationship between UGT1A expression and proliferation could be further studied with respect to potential transcription factor regulation.

PXR and its variants predominantly regulate UGT1A1, UGT1A3 and UGT1A4 genes in HepG2 and Caco-2 cells. What is striking about this nuclear receptor is that nicotine is one of the ligands activating PXR [131].

It has been shown that induction of UGT1A8 and UGT1A10 requires simultaneous modulation by AHR and Nrf2 [132]. But it should be noted that there is an extensive crosstalk between all regulatory elements inducing UGT1A transcription in a coordinated manner [133].

What's more MEME analysis of the promoter regions of UGT1A genes presented a bunch of motifs, some of which are specifically present in certain genes. In case when these motifs are analyzed further, a part of the puzzle of fine tuning by nicotine can be revealed.

One other hypothesis is that epigenetic regulators may modulate UGT1A expression coordinately. It is possible that there may be a common epigenetic regulator that acts on UGT1A1, UGT1A3, UGT1A4, UGT1A5 and UGT1A6 as a response to nicotine. Indeed Gagnon *et al.* (2006) [134] showed that UGT1A1 is subjected to epigenetic regulation [134] . Furthermore, it has been documented that nicotine interacts with methyltransferases [135].

Each UGT isoform has its distinct substrate specificity, being determined by the unique Exon I. Some substrates are precisely selective for a certain UGT enzyme. For example, individuals with homozygous mutant UGT1A1 allele suffer from hyperbilirubinemia, which is defined as the deficiency in bilirubin metabolism [4]. This clearly indicates the critical role of UGT1A1 in bilirubin metabolism. By comparison, most of the endo and exogenous substrates of UGT1A family overlap with each other.

Since another UGT member can compensate for the activity of a member with low activity, it is not possible to assign a specific isoform for the metabolism of a specific drug [136]. Accordingly, it is not possible to speculate that exposure to a certain drug is responsible for the upregulation of a specific isoform's expression.

Our results indicate that the drug responsiveness of UGT enzymes was promoted under serum starvation conditions only. Serum starvation causes cells to undergo quiescence, a stage at which cells try to keep their ROS (reactive oxygen species) levels lows [137]. Nrf2 is a well-known regulator transcription factor for UGT1A genes. As Kalthoff *et al.* (2010)[132] has shown that Nrf2 was upregulated in response to increased ROS levels, this transcription factor may well be induced after serum starvation. Our microarray and re-analysis of GEO datasets have shown that probesets annotated to Nrf2 were not among the ones significantly changed ($p<0.05$). However, Nrf2 could be additionally regulated by posttranslational mechanisms.

In broader context, our report states that cells become more responsive to nicotine when they are deprived of serum. In physiology, it may have an implication as follows: As known, tumor tissues are mass of highly heterogenic cell populations. In this mass structure there are cells in contact with extracellular matrix, supplied with nutrients, growth factors and O_2. However, there are also cells closed within this mass, surrounded with cells and deprived of nutrients as well as serum. Nicotine may not induce growth in cells facing outside within 48h; nevertheless, cells found in this mass may have been well induced by nicotine within a short time.

Nicotine clearance from the body is an important criterion for assessing the effect of this compound. The rate of nicotine metabolism shows interindividual variance and is affected by several factors; such as, diet, age, gender, medication, and smoking [38]. Moreover, it has been shown that lung cancer incidences are lower in slow nicotine metabolizers [138]. This observation points out the importance of nicotine metabolism enzymes in to what extend this compound exerts its effects on cells.

Studies indicate that fast metabolism is related with higher exposure of cells to nicotine and its effects. Since upregulation of nicotine metabolism is also associated with smoking, [139] any factor that promotes these metabolic pathways forms a risk.

Another implication of our results is that smoking may deregulate drug metabolism. How xenobiotic metabolism is affected by smoking is critical because its promotion or suppression may modulate the response to other drugs. In other words, enhanced xenobiotic metabolism may result in a resistance to chemopreventive drugs and therapies, which is undesirable. Indeed, heavy smokers were shown to bear increased

glucuronidation catabolism of paracetamol [140] In addition, the correlation between Camptothecin resistance and upregulation of UGT1A activity has been shown in colon cancer cells [141].

Irinotecan, a camptothecin compound, is used in systemic treatment of metastatic colorectal cancers [142-145]. The mode of action of irinotecan is that, by the inhibition of Topoisomerase I, DNA adducts accumulate resulting in cell death [104]. The active metabolite of irinotecan, SN-38, is inactivated by the actions of UGT enzymes; namely, UGT1A1, UGT1A6, UGT1A7 and UGT1A9 [146-149].

In literature, it has been well documented that the anticancer response and toxicity of irinotecan change widely from individual to individual. The efficacy and possible toxicity of irinotecan drug has been correlated with the polymorphisms in UGT1A gene cluster [104-105]. In addition, Gagnon *et al.* (2006)[134] has shown that epigenetic silencing in the promoter region of UGT1A1 gene is correlated with lower activity and thus, irinotecan toxicity. Furthermore, cigarette smoking decreases the exposure of cells to irinotecan and the efficiency of its related treatment [150].

The overall scheme shows that polymorphisms, smoking and epigenetic mechanisms causing low activity of UGT enzymes may bring about toxic effects of the drug in addition to an anticancer response in the patients [104]. The reason speculated for this phenotype is that lower metabolic activity causes higher plasma concentrations of the drugs [151]. Consequently, these drugs can stay at plasma for longer and affect cells more efficiently.

Our finding showed that UGT1A1 and UGT1A6, the isoforms most affected by irinotecan, also were highly affected by nicotine exposure in SW620 cells. This suggests that irinotecan and nicotine might use similar isoforms for glucuronidation thus nicotine exposure might predict partially the drug response.

In a study, where the transcriptome of methotrexate-resistant breast cancer cells is compared with that of sensitive cells, UGT1A6 was found to be significantly overexpressed in MTX-resistant cancer cells [152]. The increase in mRNA transcript and enzymatic activity of UGT1A6 has been well defined and proposed to be the reason for the reduced cytotoxicity by MTX. de Almagro *et al.* (2011)[152] showed that

when UGT1A6 expression was abolished by siRNA treatment, on the other hand, the resistance to MTX was not reduced. This is very good evidence showing that UGT1A isoforms could exhibit redundancy in their substrate specificity. Moreover, an increase in sensitivity could only be observed when whole UGT1A family was targeted with a siRNA, verifying the hypothesis [153].

Besides, deregulation of UGT enzymes may have critical toxicological and physiological consequences for the patient. In case when the expression of an isoform is changed due to an exposure of a drug, the bioavailability of various nutrients and drugs, which are subjected to glucuronidation, would be changed, as well [152]. Therefore, the possibility of crossresistance between nicotine and other chemotherapeutic drugs should be taken into account in tumor treatments.

In conclusion, UGT1A expression when analyzed in an isoform specific manner can provide a more powerful profile of a drug's metabolism rather than analyzing a single isoform.

Of great importance there are studies showing that there is a gender difference in terms of response to smoking. It is indicated that smoker women are more prone to lung, oral cavity and oropharynx cancer than smoker men [154-156]. In addition, women are shown to be faster metabolizers than men [157-158]. In fact, xenobiotic metabolism was shown to be affected by factors, such as sex, as well. For example, since women generate higher amounts of estrogen, estradiol metabolizing enzymes, among which there are UGT1A isoforms, should be upregulated in women [86,159]. This can explain why smoking women might be more prone to developing cancer than smoking men. Additionally, expression of UGT enzymes has been shown to be determined by sex [160].

In introduction, tissue specific expression section aims to precisely present the expression profiles of all UGT1A isoforms tissue by tissue. However, it is clearly seen that their expression can sometimes be detected and sometimes not, pointing out the fact that there is interindividual variability. This variability may be the result of genetic polymorphisms [161], epigenetic modifications [134] or any error related with the techniques used. Moreover, it was shown that UGT expression was restricted to certain cell types or regions in an organ by immunohistochemical studies. For

example UGT1A expression has been shown to be restricted to the proximal and distal convoluted tubules, the loops of Henle, and the collecting ducts in kidney [103]. Therefore, rather than tissues, efforts should be spent to analyze the individual's expression patterns.

Based on what was presented in this report, it is possible to propose that cancer patients should be analyzed in terms of their gender, smoking status and UGT1A expression in advance to any drug therapy.

6. FUTURE PERSPECTIVES

Finding out what regulates the expression of the isoforms should be the primary task following the data presented in this thesis. It is clear that regulation of UGT1A gene cluster is sensitive to serum levels and hence, starvation conditions either depleting growth factors and/or inducing quiescence/apoptosis might lead to production of a transcription factor responsible for induction of UGT1A transcription. However, it is important to note that serum starvation by itself does not induce UGT1A transcription but only in combination with nicotine. Future studies should include whether in the presence of absence of this TF nicotine is effective in inducing transcription of UGT1A gene cluster.

Due to redundancy in substrate specificity of UGT1A isoforms, it is hard to discriminate the contribution of each isoform in the metabolism of a substrate [152]. In addition, no inhibitory chemicals or antibodies are available for the isoforms [136]. Future studies should include using liver or colon microcosms to see the activity of these enzymes. Western blot or immunohistochemistry studies might also be helpful in defining isoform specificity in response to nicotine. Furthermore, *in vivo* analysis of UGT1A expression could be carried out.

Other drugs with a similar induction profile could be explored to prevent any crossresistance cases that can be observed in drug therapies.

Making a new setup of experiments including chronic nicotine treatment to observe the effects of this compound in long term would be another route to follow.

7. REFERENCES

1 Dutton, G. J. Commentary: Control of UDP-glucuronyltransferase activity. *Biochem Pharmacol* **24**, 1835-1841 (1975).

2 Tukey, R. H. & Strassburg, C. P. Human UDP-glucuronosyltransferases: metabolism, expression, and disease. *Annu Rev Pharmacol Toxicol* **40**, 581-616, doi:10.1146/annurev.pharmtox.40.1.581 (2000).

3 Williams, J. A. *et al.* Drug-drug interactions for UDP-glucuronosyltransferase substrates: a pharmacokinetic explanation for typically observed low exposure (AUCi/AUC) ratios. *Drug Metab Dispos* **32**, 1201-1208, doi:10.1124/dmd.104.000794
dmd.104.000794 [pii] (2004).

4 Bosma, P. J. Inherited disorders of bilirubin metabolism. *J Hepatol* **38**, 107-117, doi:S0168827802003598 [pii] (2003).

5 Harding, D., Jeremiah, S. J., Povey, S. & Burchell, B. Chromosomal mapping of a human phenol UDP-glucuronosyltransferase, GNT1. *Ann Hum Genet* **54**, 17-21 (1990).

6 Monaghan, G. *et al.* Isolation of a human YAC contig encompassing a cluster of UGT2 genes and its regional localization to chromosome 4q13. *Genomics* **23**, 496-499, doi:S0888-7543(84)71531-X [pii]
10.1006/geno.1994.1531 (1994).

7 Burchell, B. Genetic variation of human UDP-glucuronosyltransferase: implications in disease and drug glucuronidation. *Am J Pharmacogenomics* **3**, 37-52, doi:030106 [pii] (2003).

8 Ritter, J. K. *et al.* A novel complex locus UGT1 encodes human bilirubin, phenol, and other UDP-glucuronosyltransferase isozymes with identical carboxyl termini. *J Biol Chem* **267**, 3257-3261 (1992).

9 Burchell, B. *et al.* The UDP glucuronosyltransferase gene superfamily: suggested nomenclature based on evolutionary divergence. *DNA Cell Biol* **10**, 487-494 (1991).

10 Wooster, R. *et al.* Cloning and stable expression of a new member of the human liver phenol/bilirubin: UDP-glucuronosyltransferase cDNA family. *Biochem J* **278 (Pt 2)**, 465-469 (1991).

11 Mackenzie, P. I. *et al.* The UDP glycosyltransferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmacogenetics* **7**, 255-269 (1997).

12 Li, Q., Lou, X., Peyronneau, M. A., Straub, P. O. & Tukey, R. H. Expression and functional domains of rabbit liver UDP-glucuronosyltransferase 2B16 and 2B13. *J Biol Chem* **272**, 3272-3279 (1997).

13 Guillemette, C. Pharmacogenomics of human UDP-glucuronosyltransferase enzymes. *Pharmacogenomics J* **3**, 136-158, doi:10.1038/sj.tpj.6500171 6500171 [pii] (2003).

14 Zhang, T., Haws, P. & Wu, Q. Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. *Genome Res* **14**, 79-89, doi:10.1101/gr.1225204 1225204 [pii] (2004).

15 Wu, Q. *et al.* Comparative DNA sequence analysis of mouse and human protocadherin gene clusters. *Genome Res* **11**, 389-404, doi:10.1101/gr.167301 (2001).

16 Strassburg, C. P. *et al.* Polymorphic gene regulation and interindividual variation of UDP-glucuronosyltransferase activity in human small intestine. *J Biol Chem* **275**, 36164-36171, doi:10.1074/jbc.M002180200 M002180200 [pii] (2000).

17 Huang, H. & Wu, Q. Cloning and comparative analyses of the zebrafish Ugt repertoire reveal its evolutionary diversity. *PLoS One* **5**, e9144, doi:10.1371/journal.pone.0009144 (2010).

18 King, C. D., Rios, G. R., Green, M. D. & Tephly, T. R. UDP-glucuronosyltransferases. *Curr Drug Metab* **1**, 143-161 (2000).

19 Bock, K. W. Functions and transcriptional regulation of adult human hepatic UDP-glucuronosyl-transferases (UGTs): mechanisms responsible for interindividual variation of UGT levels. *Biochem Pharmacol* **80**, 771-777, doi:S0006-2952(10)00335-7 [pii] 10.1016/j.bcp.2010.04.034 (2010).

20 Nakamura, A., Nakajima, M., Yamanaka, H., Fujiwara, R. & Yokoi, T. Expression of UGT1A and UGT2B mRNA in human normal tissues and

various cell lines. *Drug Metab Dispos* **36**, 1461-1464, doi:dmd.108.021428 [pii]
10.1124/dmd.108.021428 (2008).

21 Strassburg, C. P., Manns, M. P. & Tukey, R. H. Expression of the UDP-glucuronosyltransferase 1A locus in human colon. Identification and characterization of the novel extrahepatic UGT1A8. *J Biol Chem* **273**, 8719-8726 (1998).

22 Strassburg, C. P., Nguyen, N., Manns, M. P. & Tukey, R. H. Polymorphic expression of the UDP-glucuronosyltransferase UGT1A gene locus in human gastric epithelium. *Mol Pharmacol* **54**, 647-654 (1998).

23 Ritter, J. K., Crawford, J. M. & Owens, I. S. Cloning of two human liver bilirubin UDP-glucuronosyltransferase cDNAs with expression in COS-1 cells. *J Biol Chem* **266**, 1043-1047 (1991).

24 Strassburg, C. P., Oldhafer, K., Manns, M. P. & Tukey, R. H. Differential expression of the UGT1A locus in human liver, biliary, and gastric tissue: identification of UGT1A7 and UGT1A10 transcripts in extrahepatic tissue. *Mol Pharmacol* **52**, 212-220 (1997).

25 Mojarrabi, B., Butler, R. & Mackenzie, P. I. cDNA cloning and characterization of the human UDP glucuronosyltransferase, UGT1A3. *Biochem Biophys Res Commun* **225**, 785-790, doi:S0006-291X(96)91251-6 [pii]
10.1006/bbrc.1996.1251 (1996).

26 Kaivosaari, S. *et al.* Nicotine glucuronidation and the human UDP-glucuronosyltransferase UGT2B10. *Mol Pharmacol* **72**, 761-768, doi:mol.107.037093 [pii]
10.1124/mol.107.037093 (2007).

27 Strassburg, C. P. *et al.* Regulation and function of family 1 and family 2 UDP-glucuronosyltransferase genes (UGT1A, UGT2B) in human oesophagus. *Biochem J* **338 (Pt 2)**, 489-498 (1999).

28 Collier, A. C. *et al.* UDP-glucuronosyltransferase activity, expression and cellular localization in human placenta at term. *Biochem Pharmacol* **63**, 409-419, doi:S0006295201008905 [pii] (2002).

29 Finel, M. *et al.* Human UDP-glucuronosyltransferase 1A5: identification, expression, and activity. *J Pharmacol Exp Ther* **315**, 1143-1149, doi:jpet.105.091900 [pii]
10.1124/jpet.105.091900 (2005).
30 Harding, D., Fournel-Gigleux, S., Jackson, M. R. & Burchell, B. Cloning and substrate specificity of a human phenol UDP-glucuronosyltransferase expressed in COS-7 cells. *Proc Natl Acad Sci U S A* **85**, 8381-8385 (1988).
31 Munzel, P. A., Lehmkoster, T., Bruck, M., Ritter, J. K. & Bock, K. W. Aryl hydrocarbon receptor-inducible or constitutive expression of human UDP glucuronosyltransferase UGT1A6. *Arch Biochem Biophys* **350**, 72-78, doi:S0003-9861(97)90485-9 [pii]
10.1006/abbi.1997.0485 (1998).
32 Bock, K. W. & Kohle, C. UDP-glucuronosyltransferase 1A6: structural, functional, and regulatory aspects. *Methods Enzymol* **400**, 57-75, doi:S0076-6879(05)00004-2 [pii]
10.1016/S0076-6879(05)00004-2 (2005).
33 Cheng, Z., Radominska-Pandya, A. & Tephly, T. R. Cloning and expression of human UDP-glucuronosyltransferase (UGT) 1A8. *Arch Biochem Biophys* **356**, 301-305, doi:S0003-9861(98)90781-0 [pii]
10.1006/abbi.1998.0781 (1998).
34 Ohno, S. & Nakajin, S. Determination of mRNA expression of human UDP-glucuronosyltransferases and application for localization in various human tissues by real-time reverse transcriptase-polymerase chain reaction. *Drug Metab Dispos* **37**, 32-40, doi:dmd.108.023598 [pii]
10.1124/dmd.108.023598 (2009).
35 McGurk, K. A., Brierley, C. H. & Burchell, B. Drug glucuronidation by human renal UDP-glucuronosyltransferases. *Biochem Pharmacol* **55**, 1005-1012, doi:S0006-2952(97)00534-0 [pii] (1998).
36 Strassburg, C. P., Manns, M. P. & Tukey, R. H. Differential down-regulation of the UDP-glucuronosyltransferase 1A locus is an early event in human liver and biliary cancer. *Cancer Res* **57**, 2979-2985 (1997).

37 Mojarrabi, B. & Mackenzie, P. I. The human UDP glucuronosyltransferase, UGT1A10, glucuronidates mycophenolic acid. *Biochem Biophys Res Commun* **238**, 775-778, doi:S0006-291X(97)97388-5 [pii]
10.1006/bbrc.1997.7388 (1997).

38 Benowitz, N. L., Hukkanen, J. & Jacob, P., 3rd. Nicotine chemistry, metabolism, kinetics and biomarkers. *Handb Exp Pharmacol*, 29-60, doi:10.1007/978-3-540-69248-5_2 (2009).

39 Hukkanen, J., Jacob, P., 3rd & Benowitz, N. L. Metabolism and disposition kinetics of nicotine. *Pharmacol Rev* **57**, 79-115, doi:57/1/79 [pii]
10.1124/pr.57.1.3 (2005).

40 Tutka, P., Mosiewicz, J. & Wielosz, M. Pharmacokinetics and metabolism of nicotine. *Pharmacol Rep* **57**, 143-153 (2005).

41 Benowitz, N. L. & Jacob, P., 3rd. Metabolism of nicotine to cotinine studied by a dual stable isotope method. *Clin Pharmacol Ther* **56**, 483-493 (1994).

42 Chen, G. *et al.* Glucuronidation genotypes and nicotine metabolic phenotypes: importance of functional UGT2B10 and UGT2B17 polymorphisms. *Cancer Res* **70**, 7543-7552, doi:0008-5472.CAN-09-4582 [pii]
10.1158/0008-5472.CAN-09-4582 (2010).

43 Kuehl, G. E. & Murphy, S. E. N-glucuronidation of nicotine and cotinine by human liver microsomes and heterologously expressed UDP-glucuronosyltransferases. *Drug Metab Dispos* **31**, 1361-1368, doi:10.1124/dmd.31.11.1361
31/11/1361 [pii] (2003).

44 Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative C(T) method. *Nat Protoc* **3**, 1101-1108 (2008).

45 Pfaffl, M., Meyer, H. H. & Sauerwein, H. Quantification of insulin-like growth factor-1 (IGF-1) mRNA: development and validation of an internally standardised competitive reverse transcription-polymerase chain reaction. *Exp Clin Endocrinol Diabetes* **106**, 506-513, doi:10.1055/s-0029-1212025 (1998).

46 Wong, M. L. & Medrano, J. F. Real-time PCR for mRNA quantitation. *Biotechniques* **39**, 75-85, doi:05391RV01 [pii] (2005).

47 Tichopad, A., Dilger, M., Schwarz, G. & Pfaffl, M. W. Standardized determination of real-time PCR efficiency from a single reaction set-up. *Nucleic Acids Res* **31**, e122 (2003).

48 Cha, R. S. & Thilly, W. G. Specificity, efficiency, and fidelity of PCR. *PCR Methods Appl* **3**, S18-29 (1993).

49 Heid, C. A., Stevens, J., Livak, K. J. & Williams, P. M. Real time quantitative PCR. *Genome Res* **6**, 986-994 (1996).

50 Ginzinger, D. G. Gene quantification using real-time quantitative PCR: an emerging technology hits the mainstream. *Exp Hematol* **30**, 503-512, doi:S0301472X02008068 [pii] (2002).

51 Wittwer, C. T., Herrmann, M. G., Moss, A. A. & Rasmussen, R. P. Continuous fluorescence monitoring of rapid cycle DNA amplification. *Biotechniques* **22**, 130-131, 134-138 (1997).

52 Morrison, T. B., Weis, J. J. & Wittwer, C. T. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24**, 954-958, 960, 962 (1998).

53 Simpson, D. A., Feeney, S., Boyle, C. & Stitt, A. W. Retinal VEGF mRNA measured by SYBR green I fluorescence: A versatile approach to quantitative PCR. *Mol Vis* **6**, 178-183, doi:v6/a24 [pii] (2000).

54 Wayne, C., Cook, K., Sairam, S., Hollis, B. & Thilaganathan, B. Sensitivity and accuracy of routine antenatal ultrasound screening for isolated facial clefts. *Br J Radiol* **75**, 584-589 (2002).

55 Ririe, K. M., Rasmussen, R. P. & Wittwer, C. T. Product differentiation by analysis of DNA melting curves during the polymerase chain reaction. *Anal Biochem* **245**, 154-160, doi:S0003-2697(96)99916-9 [pii] 10.1006/abio.1996.9916 (1997).

56 Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**, 169-193, doi:JME00927 [pii] (2000).

57 Souaze, F., Ntodou-Thome, A., Tran, C. Y., Rostene, W. & Forgez, P. Quantitative RT-PCR: limits and accuracy. *Biotechniques* **21**, 280-285 (1996).

58 Livak, K. J. & Schmittgen, T. D. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25**, 402-408, doi:10.1006/meth.2001.1262
S1046-2023(01)91262-9 [pii] (2001).

59 Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29**, e45 (2001).

60 Gibson, U. E., Heid, C. A. & Williams, P. M. A novel method for real time quantitative RT-PCR. *Genome Res* **6**, 995-1001 (1996).

61 Liu, W. & Saint, D. A. A new quantitative method of real time reverse transcription polymerase chain reaction assay based on simulation of polymerase chain reaction kinetics. *Anal Biochem* **302**, 52-59, doi:10.1006/abio.2001.5530
S0003269701955307 [pii] (2002).

62 Sheskin, D. *Handbook of parametric and nonparametric statistical procedures*. 2nd edn, (Chapman & Hall/CRC, 2000).

63 Leibovitz, A. *et al.* Classification of human colorectal adenocarcinoma cell lines. *Cancer Res* **36**, 4562-4569 (1976).

64 Fogh, J., Fogh, J. M. & Orfeo, T. One hundred and twenty-seven cultured human tumor cell lines producing tumors in nude mice. *J Natl Cancer Inst* **59**, 221-226 (1977).

65 Raynal, C. *et al.* Pregnane X Receptor (PXR) expression in colorectal cancer cells restricts irinotecan chemosensitivity through enhanced SN-38 glucuronidation. *Mol Cancer* **9**, 46, doi:1476-4598-9-46 [pii]
10.1186/1476-4598-9-46 (2010).

66 Burchell, B. & Coughtrie, M. W. UDP-glucuronosyltransferases. *Pharmacol Ther* **43**, 261-289, doi:0163-7258(89)90122-8 [pii] (1989).

67 Miners, J. O. & Mackenzie, P. I. Drug glucuronidation in humans. *Pharmacol Ther* **51**, 347-369, doi:0163-7258(91)90065-T [pii] (1991).

68 Wells, P. G. *et al.* Glucuronidation and the UDP-glucuronosyltransferases in health and disease. *Drug Metab Dispos* **32**, 281-290, doi:10.1124/dmd.32.3.281
32/3/281 [pii] (2004).

69 Weitz, J. *et al.* Colorectal cancer. *Lancet* **365**, 153-165, doi:S0140-6736(05)17706-X [pii]
10.1016/S0140-6736(05)17706-X (2005).

70 van der Logt, E. M. *et al.* Genetic polymorphisms in UDP-glucuronosyltransferases and glutathione S-transferases and colorectal cancer risk. *Carcinogenesis* **25**, 2407-2415, doi:10.1093/carcin/bgh251
bgh251 [pii] (2004).

71 Pond, S. M. & Tozer, T. N. First-pass elimination. Basic concepts and clinical consequences. *Clin Pharmacokinet* **9**, 1-25 (1984).

72 Kaivosaari, S., Finel, M. & Koskinen, M. N-glucuronidation of drugs and other xenobiotics by human and animal UDP-glucuronosyltransferases. *Xenobiotica* **41**, 652-669, doi:10.3109/00498254.2011.563327 (2011).

73 Burchell, B., Lockley, D. J., Staines, A., Uesawa, Y. & Coughtrie, M. W. Substrate specificity of human hepatic udp-glucuronosyltransferases. *Methods Enzymol* **400**, 46-57, doi:S0076-6879(05)00003-0 [pii]
10.1016/S0076-6879(05)00003-0 (2005).

74 Soars, M. G., Burchell, B. & Riley, R. J. In vitro analysis of human drug glucuronidation and prediction of in vivo metabolic clearance. *J Pharmacol Exp Ther* **301**, 382-390 (2002).

75 Benowitz, N. L. Drug therapy. Pharmacologic aspects of cigarette smoking and nicotine addition. *N Engl J Med* **319**, 1318-1330, doi:10.1056/NEJM198811173192005 (1988).

76 Catassi, A., Servent, D., Paleari, L., Cesario, A. & Russo, P. Multiple roles of nicotine on cell proliferation and inhibition of apoptosis: implications on lung carcinogenesis. *Mutat Res* **659**, 221-231, doi:S1383-5742(08)00043-4 [pii]
10.1016/j.mrrev.2008.04.002 (2008).

77 Hubner, R. H. *et al.* Coordinate control of expression of Nrf2-modulated genes in the human small airway epithelium is highly responsive to cigarette smoking. *Mol Med* **15**, 203-219, doi:10.2119/molmed.2008.00130 (2009).

78 Kaya, O. Investigation of the effects of Nicotine on the expression profile of SW620 colon adenocarcinoma cells using functional genomics approach. *M.Sc. Thesis submitted to Bilkent University* (2009).

79 Simon, R., Radmacher, M. D., Dobbin, K. & McShane, L. M. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst* **95**, 14-18 (2003).

80 Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci U S A* **95**, 14863-14868 (1998).

81 Edgar, R., Domrachev, M. & Lash, A. E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**, 207-210 (2002).

82 Osman, M. I. Exploring a mixture of distributions using Minitab. *Comput Biol Med* **27**, 223-232, doi:S0010-4825(97)00019-X [pii] (1997).

83 Steiling, K. *et al.* Comparison of proteomic and transcriptomic profiles in the bronchial airway epithelium of current and never smokers. *PLoS One* **4**, e5043, doi:10.1371/journal.pone.0005043 (2009).

84 Carolan, B. J. *et al.* Up-regulation of expression of the ubiquitin carboxyl-terminal hydrolase L1 gene in human airway epithelium of cigarette smokers. *Cancer Res* **66**, 10729-10740, doi:66/22/10729 [pii]
10.1158/0008-5472.CAN-06-2224 (2006).

85 Carolan, B. J., Harvey, B. G., De, B. P., Vanni, H. & Crystal, R. G. Decreased expression of intelectin 1 in the human airway epithelium of smokers compared to nonsmokers. *J Immunol* **181**, 5760-5767, doi:181/8/5760 [pii] (2008).

86 Boyle, J. O. *et al.* Effects of cigarette smoke on the human oral mucosal transcriptome. *Cancer Prev Res (Phila)* **3**, 266-278, doi:1940-6207.CAPR-09-0192 [pii]
10.1158/1940-6207.CAPR-09-0192 (2010).

87 Kupfer, D. M., White, V. L., Jenkins, M. C. & Burian, D. Examining smoking-induced differential gene expression changes in buccal mucosa. *BMC Med Genomics* **3**, 24, doi:1755-8794-3-24 [pii]
10.1186/1755-8794-3-24 (2010).

88 Sridhar, S. *et al.* Smoking-induced gene expression changes in the bronchial airway are reflected in nasal and buccal epithelium. *BMC Genomics* **9**, 259, doi:1471-2164-9-259 [pii]

10.1186/1471-2164-9-259 (2008).

89 Beane, J. *et al.* Reversible and permanent effects of tobacco smoke exposure on airway epithelial gene expression. *Genome Biol* **8**, R201, doi:gb-2007-8-9-r201 [pii]

10.1186/gb-2007-8-9-r201 (2007).

90 Spira, A. *et al.* Effects of cigarette smoke on the human airway epithelial cell transcriptome. *Proc Natl Acad Sci U S A* **101**, 10143-10148, doi:10.1073/pnas.0401422101

0401422101 [pii] (2004).

91 Landi, M. T. *et al.* Gene expression signature of cigarette smoking and its role in lung adenocarcinoma development and survival. *PLoS One* **3**, e1651, doi:10.1371/journal.pone.0001651 (2008).

92 Wang, X. *et al.* Genetic variation and antioxidant response gene expression in the bronchial airway epithelium of smokers at risk for lung cancer. *PLoS One* **5**, e11934, doi:10.1371/journal.pone.0011934 (2010).

93 Heguy, A. *et al.* Gene expression profiling of human alveolar macrophages of phenotypically normal smokers and nonsmokers reveals a previously unrecognized subset of genes modulated by cigarette smoking. *J Mol Med (Berl)* **84**, 318-328, doi:10.1007/s00109-005-0008-2 (2006).

94 Yamashiki, N., Yokota, H., Sakamoto, M. & Yuasa, A. Presence of phenol UDP-glucuronosyltransferase in bovine alveolar macrophages and bronchial epithelial cells. *Toxicology* **176**, 221-227, doi:S0300483X02001403 [pii] (2002).

95 Kazeros, A. *et al.* Overexpression of apoptotic cell removal receptor MERTK in alveolar macrophages of cigarette smokers. *Am J Respir Cell Mol Biol* **39**, 747-757, doi:2007-0306OC [pii]

10.1165/rcmb.2007-0306OC (2008).

96 Shaykhiev, R. *et al.* Smoking-dependent reprogramming of alveolar macrophage polarization: implication for pathogenesis of chronic obstructive pulmonary disease. *J Immunol* **183**, 2867-2883, doi:jimmunol.0900473 [pii]

10.4049/jimmunol.0900473 (2009).

97 Gur-Dedeoglu, B. *et al.* Identification of endogenous reference genes for qRT-PCR analysis in normal matched breast tumor tissues. *Oncol Res* **17**, 353-365 (2009).

98 Mihalick, S. M., Ortiz, D., Kumar, R., Rogers, E. & Shea, T. B. Folate and vitamin E deficiency impair cognitive performance in mice subjected to oxidative stress: differential impact on normal mice and mice lacking apolipoprotein E. *Neuromolecular Med* **4**, 197-202, doi:NMM:4:3:197 [pii]
10.1385/NMM:4:3:197 (2003).

99 de Azambuja, E. *et al.* Ki-67 as prognostic marker in early breast cancer: a meta-analysis of published studies involving 12,155 patients. *Br J Cancer* **96**, 1504-1513, doi:6603756 [pii]
10.1038/sj.bjc.6603756 (2007).

100 Suzuki, C. *et al.* ANLN plays a critical role in human lung carcinogenesis through the activation of RHOA and by involvement in the phosphoinositide 3-kinase/AKT pathway. *Cancer Res* **65**, 11314-11325, doi:65/24/11314 [pii]
10.1158/0008-5472.CAN-05-1507 (2005).

101 Thompson, J. D., Gibson, T. J. & Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Curr Protoc Bioinformatics* **Chapter 2**, Unit 2 3, doi:10.1002/0471250953.bi0203s00 (2002).

102 Bailey, T. L. & Elkan, C. Fitting a mixture model by expectation maximization to discover motifs in biopolymers. *Proc Int Conf Intell Syst Mol Biol* **2**, 28-36 (1994).

103 Gaganis, P., Miners, J. O., Brennan, J. S., Thomas, A. & Knights, K. M. Human renal cortical and medullary UDP-glucuronosyltransferases (UGTs): immunohistochemical localization of UGT2B7 and UGT1A enzymes and kinetic characterization of S-naproxen glucuronidation. *J Pharmacol Exp Ther* **323**, 422-430, doi:jpet.107.128603 [pii]
10.1124/jpet.107.128603 (2007).

104 Carlini, L. E. *et al.* UGT1A7 and UGT1A9 polymorphisms predict response and toxicity in colorectal cancer patients treated with capecitabine/irinotecan. *Clin Cancer Res* **11**, 1226-1236, doi:11/3/1226 [pii] (2005).

105 Shulman, K. *et al.* Clinical implications of UGT1A1*28 genotype testing in colorectal cancer patients. *Cancer* **117**, 3156-3162, doi:10.1002/cncr.25735 (2011).

106 Tochigi, Y., Yamashiki, N., Ohgiya, S., Ganaha, S. & Yokota, H. Isoform-specific expression and induction of udp-glucuronosyltransferase in immunoactivated peritoneal macrophages of the rat. *Drug Metab Dispos* **33**, 1391-1398, doi:dmd.105.004879 [pii]
10.1124/dmd.105.004879 (2005).

107 Spira, A. *et al.* Airway epithelial gene expression in the diagnostic evaluation of smokers with suspect lung cancer. *Nat Med* **13**, 361-366, doi:nm1556 [pii]
10.1038/nm1556 (2007).

108 Ulrich, C. M. *et al.* Epoxide hydrolase Tyr113His polymorphism is associated with elevated risk of colorectal polyps in the presence of smoking and high meat intake. *Cancer Epidemiol Biomarkers Prev* **10**, 875-882 (2001).

109 Slattery, M. L., Edwards, S., Curtin, K., Schaffer, D. & Neuhausen, S. Associations between smoking, passive smoking, GSTM-1, NAT2, and rectal cancer. *Cancer Epidemiol Biomarkers Prev* **12**, 882-889 (2003).

110 Wong, H. P. *et al.* Nicotine promotes cell proliferation via alpha7-nicotinic acetylcholine receptor and catecholamine-synthesizing enzymes-mediated pathway in human colon adenocarcinoma HT-29 cells. *Toxicol Appl Pharmacol* **221**, 261-267, doi:S0041-008X(07)00163-9 [pii]
10.1016/j.taap.2007.04.002 (2007).

111 Kaya, O. Investigation of the effects of Nicotine on the expression profile of SW620 colon adenocarcinoma cells using functional genomics approach. *M.Sc. Thesis submitted to Bilkent University* (2009).

112 Chu, M., Guo, J. & Chen, C. Y. Long-term exposure to nicotine, via ras pathway, induces cyclin D1 to stimulate G1 cell cycle transition. *J Biol Chem* **280**, 6369-6379, doi:M408947200 [pii]
10.1074/jbc.M408947200 (2005).

113 Tsurutani, J. *et al.* Tobacco components stimulate Akt-dependent proliferation and NFkappaB-dependent survival in lung cancer cells. *Carcinogenesis* **26**, 1182-1195, doi:bgi072 [pii]

10.1093/carcin/bgi072 (2005).

114 West, K. A. *et al.* Rapid Akt activation by nicotine and a tobacco carcinogen modulates the phenotype of normal human airway epithelial cells. *J Clin Invest* **111**, 81-90, doi:10.1172/JCI16147 (2003).

115 Chaudhry, I. S. *et al.* Short-term exposure to tobacco toxins alters expression of multiple proliferation gene markers in primary human bronchial epithelial cell cultures. *J Oncol* **2011**, 208563, doi:10.1155/2011/208563 (2011).

116 Metz, R. P., Auyeung, D. J., Kessler, F. K. & Ritter, J. K. Involvement of hepatocyte nuclear factor 1 in the regulation of the UDP-glucuronosyltransferase 1A7 (UGT1A7) gene in rat hepatocytes. *Mol Pharmacol* **58**, 319-327 (2000).

117 Auyeung, D. J., Kessler, F. K. & Ritter, J. K. Differential regulation of alternate UDP-glucuronosyltransferase 1A6 gene promoters by hepatic nuclear factor-1. *Toxicol Appl Pharmacol* **191**, 156-166, doi:S0041008X03002308 [pii] (2003).

118 Kiang, T. K., Ensom, M. H. & Chang, T. K. UDP-glucuronosyltransferases and clinical drug-drug interactions. *Pharmacol Ther* **106**, 97-132, doi:S0163-7258(04)00200-1 [pii]

10.1016/j.pharmthera.2004.10.013 (2005).

119 Sugatani, J. *et al.* The phenobarbital response enhancer module in the human bilirubin UDP-glucuronosyltransferase UGT1A1 gene and regulation by the nuclear receptor CAR. *Hepatology* **33**, 1232-1238, doi:S0270-9139(01)30137-4 [pii]

10.1053/jhep.2001.24172 (2001).

120 Xie, W. *et al.* Reciprocal activation of xenobiotic response genes by nuclear receptors SXR/PXR and CAR. *Genes Dev* **14**, 3014-3023 (2000).

121 Senekeo-Effenberger, K. *et al.* Expression of the human UGT1 locus in transgenic mice by 4-chloro-6-(2,3-xylidino)-2-pyrimidinylthioacetic acid (WY-14643) and implications on drug metabolism through peroxisome proliferator-activated receptor alpha activation. *Drug Metab Dispos* **35**, 419-427, doi:dmd.106.013243 [pii]

10.1124/dmd.106.013243 (2007).

122 Barbier, O. *et al.* The UDP-glucuronosyltransferase 1A9 enzyme is a peroxisome proliferator-activated receptor alpha and gamma target gene. *J Biol Chem* **278**, 13975-13983, doi:10.1074/jbc.M300749200 M300749200 [pii] (2003).

123 Sugatani, J. *et al.* Transcriptional regulation of human UGT1A1 gene expression through distal and proximal promoter motifs: implication of defects in the UGT1A1 gene promoter. *Naunyn Schmiedebergs Arch Pharmacol* **377**, 597-605, doi:10.1007/s00210-007-0226-y (2008).

124 Shelby, M. K. & Klaassen, C. D. Induction of rat UDP-glucuronosyltransferases in liver and duodenum by microsomal enzyme inducers that activate various transcriptional pathways. *Drug Metab Dispos* **34**, 1772-1778, doi:dmd.106.010397 [pii] 10.1124/dmd.106.010397 (2006).

125 Lankisch, T. O. *et al.* Aryl hydrocarbon receptor-mediated regulation of the human estrogen and bile acid UDP-glucuronosyltransferase 1A3 gene. *Arch Toxicol* **82**, 573-582, doi:10.1007/s00204-008-0347-1 (2008).

126 Erichsen, T. J. *et al.* Genetic variability of aryl hydrocarbon receptor (AhR)-mediated regulation of the human UDP glucuronosyltransferase (UGT) 1A4 gene. *Toxicol Appl Pharmacol* **230**, 252-260, doi:S0041-008X(08)00097-5 [pii] 10.1016/j.taap.2008.02.020 (2008).

127 Sugatani, J. *et al.* Transcriptional regulation of human UGT1A1 gene expression: activated glucocorticoid receptor enhances constitutive androstane receptor/pregnane X receptor-mediated UDP-glucuronosyltransferase 1A1 regulation with glucocorticoid receptor-interacting protein 1. *Mol Pharmacol* **67**, 845-855, doi:mol.104.007161 [pii] 10.1124/mol.104.007161 (2005).

128 Munzel, P. A. *et al.* Induction of human UDP glucuronosyltransferases (UGT1A6, UGT1A9, and UGT2B7) by t-butylhydroquinone and 2,3,7,8-tetrachlorodibenzo-p-dioxin in Caco-2 cells. *Drug Metab Dispos* **27**, 569-573 (1999).

129 Shimba, S., Komiyama, K., Moro, I. & Tezuka, M. Overexpression of the aryl hydrocarbon receptor (AhR) accelerates the cell proliferation of A549 cells. *J Biochem* **132**, 795-802 (2002).

130 Andersson, P. *et al.* A constitutively active dioxin/aryl hydrocarbon receptor induces stomach tumors. *Proc Natl Acad Sci U S A* **99**, 9990-9995, doi:10.1073/pnas.152706299
152706299 [pii] (2002).

131 Lamba, V. *et al.* PXR (NR1I2): splice variants in human tissues, including brain, and identification of neurosteroids and nicotine as PXR activators. *Toxicol Appl Pharmacol* **199**, 251-265, doi:10.1016/j.taap.2003.12.027
S0041008X04000602 [pii] (2004).

132 Kalthoff, S., Ehmer, U., Freiberg, N., Manns, M. P. & Strassburg, C. P. Interaction between oxidative stress sensor Nrf2 and xenobiotic-activated aryl hydrocarbon receptor in the regulation of the human phase II detoxifying UDP-glucuronosyltransferase 1A10. *J Biol Chem* **285**, 5993-6002, doi:M109.075770 [pii]
10.1074/jbc.M109.075770 (2010).

133 Kohle, C. & Bock, K. W. Coordinate regulation of human drug-metabolizing enzymes, and conjugate transporters by the Ah receptor, pregnane X receptor and constitutive androstane receptor. *Biochem Pharmacol* **77**, 689-699, doi:S0006-2952(08)00346-8 [pii]
10.1016/j.bcp.2008.05.020 (2009).

134 Gagnon, J. F., Bernard, O., Villeneuve, L., Tetu, B. & Guillemette, C. Irinotecan inactivation is modulated by epigenetic silencing of UGT1A1 in colon cancer. *Clin Cancer Res* **12**, 1850-1858, doi:12/6/1850 [pii]
10.1158/1078-0432.CCR-05-2130 (2006).

135 Satta, R. *et al.* Nicotine decreases DNA methyltransferase 1 expression and glutamic acid decarboxylase 67 promoter methylation in GABAergic interneurons. *Proc Natl Acad Sci U S A* **105**, 16356-16361, doi:0808699105 [pii]
10.1073/pnas.0808699105 (2008).

136 Liu, X., Tam, V. H. & Hu, M. Disposition of flavonoids via enteric recycling: determination of the UDP-glucuronosyltransferase isoforms responsible for

the metabolism of flavonoids in intact Caco-2 TC7 cells using siRNA. *Mol Pharm* **4**, 873-882, doi:10.1021/mp0601190 (2007).

137 Jin, K., Ewton, D. Z., Park, S., Hu, J. & Friedman, E. Mirk regulates the exit of colon cancer cells from quiescence. *J Biol Chem* **284**, 22916-22925, doi:M109.035519 [pii]

10.1074/jbc.M109.035519 (2009).

138 Benowitz, N. L., Perez-Stable, E. J., Herrera, B. & Jacob, P., 3rd. Slower metabolism and reduced intake of nicotine from cigarette smoking in Chinese-Americans. *J Natl Cancer Inst* **94**, 108-115 (2002).

139 Benowitz, N. L. & Jacob, P., 3rd. Effects of cigarette smoking and carbon monoxide on nicotine and cotinine metabolism. *Clin Pharmacol Ther* **67**, 653-659, doi:S0009-9236(00)96093-4 [pii]

10.1067/mcp.2000.107086 (2000).

140 Bock, K. W. *et al.* The influence of environmental and genetic factors on CYP2D6, CYP1A2 and UDP-glucuronosyltransferases in man using sparteine, caffeine, and paracetamol as probes. *Pharmacogenetics* **4**, 209-218 (1994).

141 Brangi, M. *et al.* Camptothecin resistance: role of the ATP-binding cassette (ABC), mitoxantrone-resistance half-transporter (MXR), and potential for glucuronidation in MXR-expressing cells. *Cancer Res* **59**, 5938-5946 (1999).

142 Saltz, L. Irinotecan-based combinations for the adjuvant treatment of stage III colon cancer. *Oncology (Williston Park)* **14**, 47-50 (2000).

143 Saltz, L. B. *et al.* Irinotecan plus fluorouracil/leucovorin for metastatic colorectal cancer: a new survival standard. *Oncologist* **6**, 81-91 (2001).

144 Cunningham, D., Maroun, J., Vanhoefer, U. & Van Cutsem, E. Optimizing the use of irinotecan in colorectal cancer. *Oncologist* **6 Suppl 4**, 17-23 (2001).

145 Rothenberg, M. L. Efficacy and toxicity of irinotecan in patients with colorectal cancer. *Semin Oncol* **25**, 39-46 (1998).

146 Iyer, L. *et al.* Genetic predisposition to the metabolism of irinotecan (CPT-11). Role of uridine diphosphate glucuronosyltransferase isoform 1A1 in the glucuronidation of its active metabolite (SN-38) in human liver microsomes. *J Clin Invest* **101**, 847-854, doi:10.1172/JCI915 (1998).

147 Ciotti, M., Basu, N., Brangi, M. & Owens, I. S. Glucuronidation of 7-ethyl-10-hydroxycamptothecin (SN-38) by the human UDP-glucuronosyltransferases encoded at the UGT1 locus. *Biochem Biophys Res Commun* **260**, 199-202, doi:10.1006/bbrc.1999.0453
S0006-291X(99)90453-9 [pii] (1999).

148 Gagne, J. F. *et al.* Common human UGT1A polymorphisms and the altered metabolism of irinotecan active metabolite 7-ethyl-10-hydroxycamptothecin (SN-38). *Mol Pharmacol* **62**, 608-617 (2002).

149 Hanioka, N. *et al.* Human liver UDP-glucuronosyltransferase isoforms involved in the glucuronidation of 7-ethyl-10-hydroxycamptothecin. *Xenobiotica* **31**, 687-699, doi:10.1080/00498250110057341 (2001).

150 van der Bol, J. M. *et al.* Cigarette smoking and irinotecan treatment: pharmacokinetic interaction and effects on neutropenia. *J Clin Oncol* **25**, 2719-2726, doi:JCO.2006.09.6115 [pii]
10.1200/JCO.2006.09.6115 (2007).

151 Iyer, L. *et al.* UGT1A1*28 polymorphism as a determinant of irinotecan disposition and toxicity. *Pharmacogenomics J* **2**, 43-47 (2002).

152 de Almagro, M. C. *et al.* UDP-glucuronosyltransferase 1A6 overexpression in breast cancer cells resistant to methotrexate. *Biochem Pharmacol* **81**, 60-70, doi:S0006-2952(10)00664-7 [pii]
10.1016/j.bcp.2010.09.008 (2011).

153 Selga, E. *et al.* Networking of differentially expressed genes in human cancer cells resistant to methotrexate. *Genome Med* **1**, 83, doi:gm83 [pii]
10.1186/gm83 (2009).

154 Risch, H. A. *et al.* Are female smokers at higher risk for lung cancer than male smokers? A case-control analysis by histologic type. *Am J Epidemiol* **138**, 281-293 (1993).

155 Henschke, C. I., Yip, R. & Miettinen, O. S. Women's susceptibility to tobacco carcinogens and survival after diagnosis of lung cancer. *JAMA* **296**, 180-184, doi:296/2/180 [pii]
10.1001/jama.296.2.180 (2006).

156 Neugut, A. I. & Jacobson, J. S. Women and lung cancer: gender equality at a crossroad? *JAMA* **296**, 218-219, doi:296/2/218 [pii]

10.1001/jama.296.2.218 (2006).

157 Kandel, D. B., Hu, M. C., Schaffran, C., Udry, J. R. & Benowitz, N. L. Urine nicotine metabolites and smoking behavior in a multiracial/multiethnic national sample of young adults. *Am J Epidemiol* **165**, 901-910, doi:kwm010 [pii]

10.1093/aje/kwm010 (2007).

158 Johnstone, E. *et al.* Determinants of the rate of nicotine metabolism and effects on smoking behavior. *Clin Pharmacol Ther* **80**, 319-330, doi:S0009-9236(06)00256-6 [pii]

10.1016/j.clpt.2006.06.011 (2006).

159 Yager, J. D. & Davidson, N. E. Estrogen carcinogenesis in breast cancer. *N Engl J Med* **354**, 270-282, doi:354/3/270 [pii]

10.1056/NEJMra050776 (2006).

160 Court, M. H. Acetaminophen UDP-glucuronosyltransferase in ferrets: species and gender differences, and sequence analysis of ferret UGT1A6. *J Vet Pharmacol Ther* **24**, 415-422, doi:366 [pii] (2001).

161 Peters, W. H., te Morsche, R. H. & Roelofs, H. M. Combined polymorphisms in UDP-glucuronosyltransferases 1A1 and 1A6: implications for patients with Gilbert's syndrome. *J Hepatol* **38**, 3-8, doi:S0168827802003069 [pii] (2003).