View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by DigitalCommons@The Texas Medical Center

Texas Medical Center Library
DigitalCommons@The Texas Medical Center

UT GSBS Dissertations and Theses (Open Access)

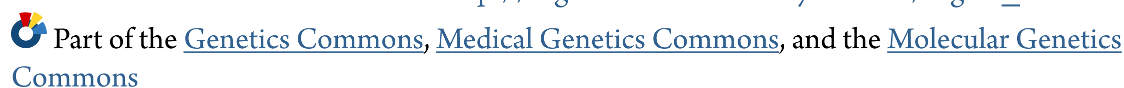
Graduate School of Biomedical Sciences

5-2010

Xenobiotic Metabolism Genes and Clubfoot

Amy M. Sommer

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations


Part of the Genetics Commons, Medical Genetics Commons, and the Molecular Genetics Commons

Recommended Citation

Sommer, Amy M., "Xenobiotic Metabolism Genes and Clubfoot" (2010). *UT GSBS Dissertations and Theses (Open Access)*. Paper 35.

This Thesis (MS) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

XENOBIOTIC METABOLISM GENES AND CLUBFOOT

by

Amy Michelle Sommer

APPROVED:

____
Supervisory Professor
Jacqueline T. Hecht, PhD

____
Susan H. Blanton, PhD

____
Stephen P. Daiger, PhD

____
James E. Hixson, PhD

____
Jennifer Hoskovec, MS, CGC

APPROVED:

____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

XENOBIOTIC METABOLISM GENES AND CLUBFOOT

A

THESIS

Presented to the Faculty of
The University of Texas
Health Science Center at Houston
and
The University of Texas
M. D. Anderson Cancer Center
Graduate School of Biomedical Sciences
in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

by

Amy Michelle Sommer, B.S.
Houston, Texas

May, 2010

ACKNOWLEDGEMENTS

First and foremost I would like to thank Dr. Hecht, my thesis advisor, for giving me the opportunity to work on this project. Her guidance throughout the thesis process allowed me to submit a project that I am extremely proud of. I could not have done any of this without her. I would also like to thank my thesis committee, Dr. Susan Blanton, Dr. Stephen Daiger, Dr. James Hixson and Jennifer Hoskovec for their enthusiasm throughout this project. For his help in statistics, I would like to thank Dr. Syed Hashmi. Additionally, I would like to thank Claire Singletary and Sarah Jane Noblin, the Genetic Counseling Program Director and Assistant Program Director, for their endless support and advice over the past eighteen months. For Carrie, Deanna, Gina, Justine and Leslie, my amazing classmates, I cannot explain how precious my time getting to know you has been; I will treasure it forever. Finally, I would especially like to thank my family and boyfriend, Michael, for their unyielding encouragement and love.

XENOBIOTIC METABOLISM GENES AND CLUBFOOT

Publication No.________

Amy Michelle Sommer, B.S.

Supervisory Professor: Jacqueline T. Hecht, PhD

Idiopathic or isolated clubfoot is a common orthopedic birth defect that affects approximately 135,000 children worldwide. It is characterized by equinus, varus and adductus deformities of the ankle and foot. Correction of clubfoot involves months of serial manipulations, castings and bracing, with surgical correction needed in forty percent of cases. Multifactorial etiology has been suggested in numerous studies with both environmental and genetic factors playing an etiologic role. Maternal smoking during pregnancy is the only common environmental factor that has consistently been shown to increase the risk for clubfoot. Moreover, a positive family history of clubfoot and maternal smoking increases the risk of clubfoot twenty-fold. These findings suggest that genetic variation in smoking metabolism genes may increase susceptibility to clubfoot. Based on this reasoning, we interrogated eight candidate genes, chosen based on their involvement in phase 1 and 2 cigarette smoke metabolism. Twenty-two SNPs and two null alleles in eight genes (*CYP1A1, CYP1A2, CYP1B1, CYP2A6, EPHX1, NAT2, GSTM1* and *GSTT1*) were genotyped in a dataset composed of nonHispanic white and Hispanic multiplex and simplex families. Only one SNP in *CYP1A1*, rs1048943, had significantly altered transmission in the aggregate and multiplex NHW datasets (p=0.003 and p=0.009). Perturbation of *CYP1A1* by rs1048943 polymorphism causes an increase in the amount of harmful, adduct forming metabolic intermediates. A significant gene interaction between *EPHX1* and *NAT2* was also found (p=0.007). This interaction may affect the metabolism of harmful metabolic intermediates. Additionally, marginal interactions were found for other xenobiotic genes and these interactions may play a contributory role in clubfoot. Importantly, for *CYP1A2*, significant maternal (p=0.03; RR=1.24; 95% CI: 1.04-1.44) and fetal (p=0.01; RR=1.33; 95% CI: 1.13-1.54) genotypic effects were identified, suggesting that both maternal and fetal genotypes impact normal limb development. No association was found for maternal

smoking status and tobacco metabolism genes. Together, these results suggest that xenobiotic metabolism genes may play a contributory role in the etiology of clubfoot regardless of maternal smoking status and may impact foot development through perturbation of tobacco metabolic pathways.

TABLE of CONTENTS

List of Figures	vii
List of Tables	viii
Introduction	1
Materials and Methods	35
Results	40
Discussion	46
Appendix	56
References	77
Vita	90

LIST of FIGURES

Figure 1. Infant with clubfoot 2

Figure 2. Illustrations of embryonic development of the limbs (32-56 days) 5

Figure 3. Illustrations of positional changes of the developing limbs of human embryos 7

Figure 4. Clubfoot casting 9

Figure 5. Xenobiotic metabolism pathway 21

Figure 6. Metabolism of BaP 22

Figure 7. Regulation of CYP1 enzymes 23

Figure 8. Xenobiotic metabolism pathway highlighting the role of genes in this study 34

Figure 9. Optimized Takara *Ex Taq* Polymerase PCR protocol 36

Figure 10. PCR-PAGE results for *GSTT1* and *GSTM1* null allele assay 37

LIST of TABLES

Table 1. Compilation of clubfoot incidence, gender prevalence and laterality in various populations 3

Table 2. Proposed inheritance patterns for clubfoot 13

Table 3. Risk of clubfoot by smoking and family history 16

Table 4. Odds ratio for clubfoot in mothers who smoke during pregnancy 17

Table 5. Relative expression of human CYP1, 2 and 3 subfamilies in normal extrahepatic tissues at the level of mRNA in comparison with the liver 28

Table 6. CYP450 functional alleles 29

Table 7. Acetylation activity of common *NAT2* SNPs 32

Table 8. Common NAT2 slow acetylator alleles 32

Table 9. Composition of clubfoot dataset by ethnicity and family history 35

Table 10. Smoking metabolism genes: SNP location, alleles and ethnic frequencies 41

Table 11. Results of single SNP association analysis by family history 42

Table 12. Gene-gene interactions 43

Table 13. Results of log-linear regression modeling of genotypes for mothers and children 44

Table 14. Predicted transcription factor binding sites for 5' associated SNPs 44

Table 15. Results for *GSTM1* and *GSTT1* analysis 45

Table 16. Functional effects of significant smoking metabolism gene interactions

16A. Interactions between phase I genes 51

16B. Interactions between phase I and phase II genes 52

INTRODUCTION

Idiopathic talipes equinovarus, or isolated (non-syndromic) clubfoot is a common birth defect that has been recognized and described for centuries. All over the world, since the time of the earliest forms of language and writing, clubfoot has been etched, painted and written into human history. The clubfoot deformity can be seen in the tombs of the ancient Egyptians (1). Archeologic evidence shows that the ancient Aztec tribes in Mexico recognized clubfoot as a deformity and treated it with splints (1). In what is today the European continent, Hippocrates scribed one of the first formal descriptions of clubfoot as early as 300 BC (1). From the tenth century, clubfoot is mentioned in an Indian prayer book by Yajur-Veda (2).

Throughout the Middle Ages, clubfoot was considered a punishment from God for mothers who had lived a sinful life, or the result of a mother sitting too long with crossed legs (2). Beginning in the middle of the seventeenth century, writings on clubfoot began to describe its etiology, pathology, and options for treatment and/or surgical intervention (1). In 1842, the first Danish publications on clubfoot by Eschrict reported clubfoot as a consequence of developmental arrest in which the normal rotation of the foot is inhibited (2). During and since that time, numerous hypotheses on the etiology of clubfoot have been proposed and will be discussed in more detail in the following sections. Interestingly, however, the true etiology, pathology and most appropriate treatment techniques have not yet been unanimously accepted.

Historical review reveals that clubfoot has been a part of human life for hundreds of years. Its diagnosis, description and treatment have changed as more has been learned about clubfoot and as technology and medical procedures have advanced. Currently the causes of clubfoot are still unknown, but diagnosis and treatments for the condition have improved. Therefore, those affected by clubfoot continue to benefit from the ongoing research as it allows for continuing modification and improvement of their care.

Definition of Idiopathic Talipes Equinovarus (Clubfoot)

Clubfoot can be considered a combination of three abnormalities in foot development. These abnormalities are equinus, varus and adduction of the foot and ankle.

The severity of each deformity can vary among individuals and all three deformities interact, such that correction of one abnormality requires correction of the other two (1). In 50-75% of cases, clubfoot is identified as an isolated anomaly (3, 4). It is these cases, termed idiopathic talipes equinovarus (referred to as clubfoot for the duration of this paper) that are of interest in the current study.

Figure 1. Infant with clubfoot

CURE Clubfoot Worldwide (cureclubfoot.org)

In the remaining 25-50% of cases, clubfoot is seen in conjunction with other birth defects or developmental abnormalities. When clubfoot is found as part of a group of anomalies, it is considered to be syndromic. Defects in the hands and eyes, cleft palate, micrognathia, spinal cord defects, developmental delay and motor delay are the most commonly associated findings (1). Syndromic clubfoot can be the result of a chromosome abnormality, teratologic events, or it can develop postnatally due to neurologic disorders and myopathies (1). Clubfoot is seen as a feature in a variety of syndromes, such as arthrogryposis, nail-patella syndrome, congenital constriction bands, lead poisoning, diastrophic dwarfism, Gordon syndrome and Mobius syndrome (1). Because clubfoot can be a feature in various syndromes that all have different etiologies, identification of idiopathic clubfoot is important for an accurate determination of recurrence risks, prognosis and treatment.

BIRTH PREVALENCE

While the birth prevalence of clubfoot is often simplified to 1 affected per 1,000 live births, studies have shown that it can actually vary approximately 10-fold between different populations (Table 1). The highest prevalence of clubfoot is 6.8 per 1,000 live births and is seen in Polynesian populations (5). The lowest prevalence is 0.57 per 1,000 live births and is seen in oriental populations (5). In European, Australian or American Caucasian

populations the birth prevalence consistently ranges from 0.73 to 1.24 per 1,000 (6-8). The prevalence in populations of Hispanic descent, both US- and foreign-born, is 0.762 and 0.759 per 1,000, respectively (6). The African-American population has a slightly lower birth prevalence of clubfoot with 0.68 children affected per 1,000 births (6). In Texas, the birth prevalence of clubfoot is not significantly different between Caucasians, US-born Hispanics and foreign-born Hispanics (6).

The prevalence of clubfoot also varies between the sexes. All studies that have evaluated the sex ratio have found that males are affected more often than females. The ratio of affected males to affected females ranges from 1.7:1 in the Maori and Polynesian populations in New Zealand to 2.4:1 in the South Australian population (8, 9). Studies of

Table 1. Compilation of clubfoot incidence, gender prevalence and laterality in various populations

Reference	Population	Incidence per 1,000	N	M:F	Laterality
Alberman, 1965[a]	European	3	36	1.6:1	-
Wynne-Davies, 1965	European/Caucasian	1.24	144	2.1:1	-
Ching, 1969	Hawaiians	6.8	-	-	-
	Caucasian	1.12	-	-	-
	Unmixed Orientals	0.567	-	-	-
Cartlidge, 1984	Caucasian	-	120	2.2:1	B = 49%; U = 57% R and 43% L
	Polynesian	-	118		B = 41%; U = 54% R and 46% L
Chapman, 2000	Polynesian (Maori or New Zealand)	-	-	1.7:1	U = 61.3% R and 38.7% L
Moorthi, 2005[b]	Caucasian	0.725	255	2:1	B = 49.3%; U = 56.6% R and 43.4% L
	Black	0.683	67		
	Foreign-born Hisp	0.759	159		
	US-born Hisp	0.762	177		
	Overall	0.740	1354		
Byron-Scott, 2005	South Australian	1.1	231	2.4:1	B = 45%; U = 58% R and 42% L
Cardy, 2007	UK	-	194	2.1:1	B = 50%; U = 57% R and 43% L

[a]Other defects seen in 33%; 13.3% had a sib with severe malformations

[b]Found POR to be similar in all groups

European Caucasian populations of European descent have consistently found a 2:1 male to female ratio (6, 7, 10, 11).

Clubfoot can be bilateral or unilateral, with either the right or the left foot affected (Table 1). Bilateral clubfoot occurs in 41% to 57% of cases (6, 8-11). When the clubfoot deformity is unilateral there seems to be a slight predominance of right-sided cases (54-61%) versus left-sided cases (39-46%) (6, 8-11). The variation in the data for laterality is likely due to the ascertainment methods used for each study and differences among the populations studied. In general, approximately half of all clubfoot cases are unilateral, with the right foot affected more often than the left foot.

LIMB DEVELOPMENT

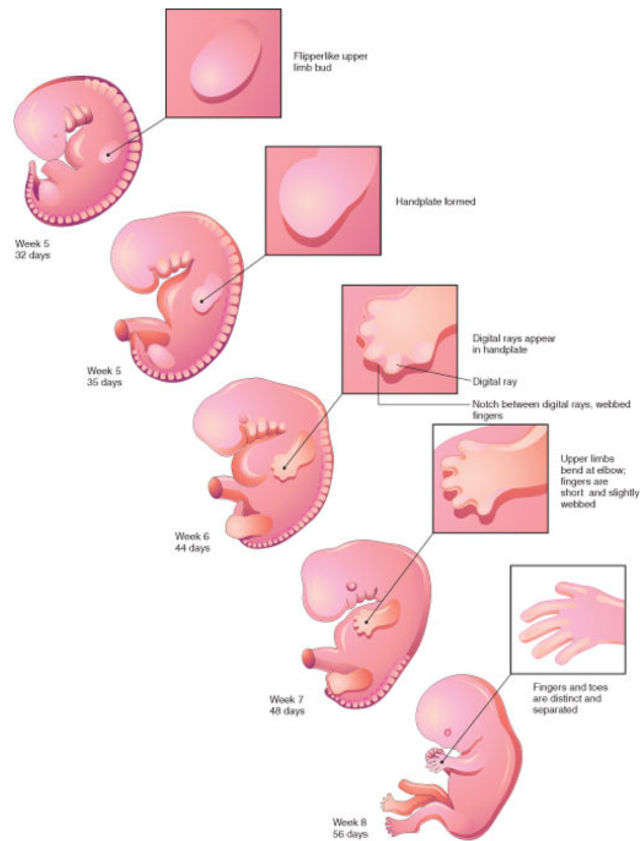
It is important to examine and understand normal limb development so that it can be compared to the aberrant limb development seen in clubfoot (Figure 2). Because the clubfoot has all of the same components as a normal foot, observations of when the development between the clubfoot and normal foot diverge are of particular importance to understanding its etiology. By identifying differences during development, researchers can better identify when a change in development occurs, what may be causing the deformity, and how it might be prevented or better treated.

Normal Limb Development

The development of the lower limb as a unique and identifiable part of the human body begins in the embryonic period of morphogenesis, after fertilization, cleavage, gastrulation and neurulation have all been completed (12, 13). The formation of the limbs is a result of cell proliferation, cell differentiation and patterning (12). The process begins 28 days after fertilization, at which time the limb bud only consists of loose mesenchymal tissue surrounded by epithelial ectoderm (13). The lower limb bud erupts opposite the five lumbar and first sacral somites and lengthens at the progress zone (PZ) (14). The undifferentiated and proliferative state of the PZ is maintained by the apical ectodermal ridge (AER) (14).

The cells that make up the lower limb are patterned into three axes: anteroposterior (great toe to fifth digit), dorsoventral (top of foot to bottom of foot) and proximal-distal (ankle to toe) (13). Various cell groups and numerous developmental genes are responsible for interacting and coordinating this pattern through inhibition and activation at different times during development (13). For example, mesenchymal cells within the zone of polarizing activity (ZPA) are responsible for producing diffusible morphogens that polarize the cells along the limb based on a concentration gradient (14).

Figure 2. Illustrations of embryonic development of the limbs (32-56 days)



This figure was published in The Developing Human, 8th ed., Moore & Persaud, The Limbs, 367, Copyright Elsevier (2008).

As development progresses, the homogeneous mesenchymal tissue differentiates. The center of the limb bud forms a chondrogenic core that is surrounded by dense vasculature. This process occurs in a proximal to distal pattern and represents the beginning of the formation of the skeleton in the leg. In the eighth week all of the future skeletal components of the leg, except the distal phalanges, exist but are made of cartilage, and apoptosis occurs which allows for separation of individual digits. Amongst the cartilage and dense vascular beds, nerves begin to grow into the lower limb and muscle tissue develops. From the end of the embryonic period to just before birth, the cartilage skeletal structures ossify to form bone (13).

The mechanism for vasculature development in the limbs is less well described. Initially there are multiple arteries that innervate the lower limb (13). The arteries form one

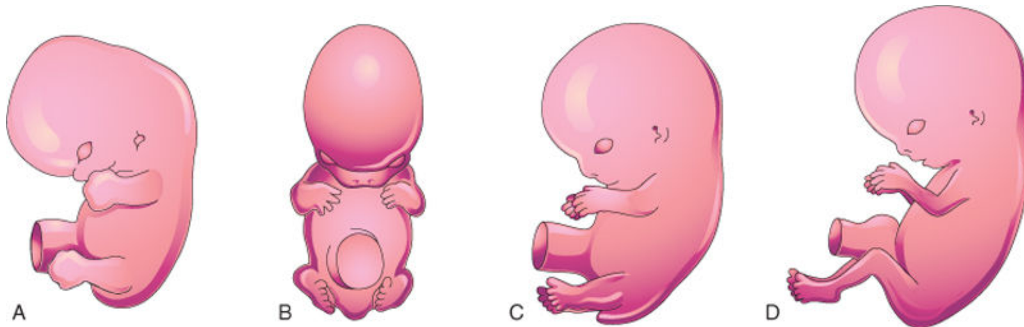
single axial artery that comes from the umbilical artery and becomes the gluteal, sciatic, proximal politeal and distal peroneal arteries (13). Further artery formation occurs secondary to axial artery formation (13). The adult vasculature pattern is achieved by the eighth week (14).

The mesenchymal cells that will become the leg muscles migrate and surround the chondrogenic core of the limb bud in the fifth week (12). In a proximal to distal pattern the mass of cells develop into individual muscles until the eighth week, when most of the adult muscles are formed and fetal movement can begin (13). During this time the cells undergo differentiation and interact with other cells to form myoblasts, myotubes, sarcomeres and myofibrils (13). Most muscle cells are developed prior to birth and all are formed within the first year of life (13).

As the tissues in the lower limbs differentiate and the limb bud grows along all its axes, it also begins to rotate (Figure 3). In fact, early in development a fetus sits with its lower limbs in a clubfoot position (2). Before the seventh week the preaxial border sits cranially and the postaxial border sits caudally (12). In the seventh week the leg extends ventrally and then the lower limbs rotate medially 90° (12). In the eighth week of development the feet display plantar flexion with adduction of both the forefoot and midfoot (2). In the beginning of the third month there is plantar flexion, adduction and supination which slowly resolves until the fourth month when only adduction of the forefoot and supination persist (2). By the twelfth week of gestation the embryo is fully formed and will grow and mature for the remaining time in the womb. The gross morphology of the lower limb has been summarized by Boehm into 4 stages:

1. I (eight weeks): The foot is 90° equinus and adducted.
2. II (nine to ten weeks): The foot is 90° equinus, adducted and supinated.
3. III (ten to eleven weeks): The foot dorsiflexes at the ankle. Mild equinus and significant supination persist. The first metatarsal retains adduction.
4. IV (twelve weeks): The foot pronates to a position of midsupination (will not be completed until after birth) and equinus positioning is resolved (14).

Figure 3. Illustrations of positional changes of the developing limbs of human embryos



A. 48 days; B. 51 days; C. 54 days; D. 56 days.
This figure was published in The Developing Human, 8th ed., Moore & Persaud, The Limbs, 367, Copyright Elsevier (2008).

Limb development is a complex process and perturbation of any aspect could potentially result in anomalous development and malformations. Consequently, there have been numerous hypotheses proposed regarding the etiology of clubfoot. Many of these hypotheses are based on comparisons between the anatomy and development of the clubfoot and the normal foot. To make this comparison, one must also be familiar with the functional and structural differences in the clubfoot.

Clubfoot Anatomy and Development

Interestingly, all normal elements of the foot are present in the clubfoot deformity (2). It is the relative position, size and shape of the bones, as well as contractures and hypoplasia of various muscles, ligaments and joints that comprise the deformity (1, 2). It is still unclear precisely which abnormalities are primary defects and which may be secondary malformations (1, 2). However, it is known that the clubfoot deformity is comprised of abnormal equinus, varus and adduction of the ankle and foot, which can frequently be associated with a cavus deformity. The equinus deformity refers to the plantar-flexed positioning of the foot, which is due to ankle joint equinus, inversion of the talocalcaneonavicular complex and plantar flexion of the forefoot. The varus deformity describes the inward rotation of the hindfoot, primarily at the talocalcaneonavicular joint. Adduction is a consequence of the medial displacement of the talonavicular and anterior

subtalar joint which causes the foot to turn inward. The cavus component refers directly to the forefoot plantar flexion that can accompany the other three deformities (1).

The severity of clubfoot differs among individuals and these findings are variable, meaning they are not all seen in all cases (1, 2). The severity of the clubfoot deformity may, therefore, be attributed to the degree of the displacement and malformation of the bony structures, while resistance to corrections is often a consequence of the changes and rigidity in the soft tissues (1). With a better understanding of the normal and abnormal structures of the foot, methods for treatment have been theorized, implemented and refined over the last century.

TREATMENT AND PROGNOSIS

All treatments of clubfoot seek to correct both the displacement of the bones and the function of the soft tissues in order to return the ankle and foot to its normal positioning. The goal is to give the patient a "functional, pain-free, normal-looking, plantigrade foot with good mobility, without calluses and requiring no modified shoes" (Ponseti, 1996). "A totally normal foot is not attainable" (Ponseti, 1996). There are multiple methods and procedures in place, which involve serial manipulations, castings and bracing, with 40% of patients requiring more invasive, surgical intervention (3, 15). Ideally, treatment begins soon after birth (2, 16). Depending on the severity of the clubfoot and tendency for relapse, correction could continue through a child's fourth year of life (16). Therefore, correction of clubfoot can be physically, emotionally and financially overwhelming for an individual and their family members.

Non-surgical Treatment

Almost all orthopedic practitioners believe that initial treatment of clubfoot should be non-surgical (1, 2, 17). Non-surgical methods utilize manipulation of the limb to stretch the soft tissues, muscles, tendons, ligaments and joint capsules (2). Short-term, or daily manipulations use bandages or splints while longer-term manipulation requires plaster casting (2) (Figure 4). The most common form of non-surgical correction is serial casting, which can be accomplished via multiple methods (17).

Figure 4. Clubfoot casting

CURE Clubfoot Worldwide (cureclubfoot.org)

In 1939, Kite proposed a method of casting that sought to correct each component of the clubfoot deformity separately, in a sequential procedure (16). Kite's casting method lasted up to 2 years in some cases and 50-75% of cases still required subsequent soft tissue release surgery for complete correction (16).

In the 1940's, in response to the onset of painful outcomes resulting from treatment with surgical techniques, Ignacio Ponseti modified the previous casting method to include serial manipulation, casting, tenotomy of the Achilles tendon and bracing (16, 17). The Ponseti method avoids bone and joint surgery by beginning treatment with gentle manipulations and plaster casting followed by splinting for up to 4 years to prevent relapses (18). This method boasts nearly a 90% success rate after an average treatment time of 49 months (8.6 weeks of casting) with 70-90% of individuals undergoing an Achilles tenotomy and approximately 50% requiring anterior tibial tendon transfers (16-18). Relapses are rare after a child turns 5-years-old and extremely rare after 7-years (18).

Ideally an individual would begin casting approximately 24-hours after birth and have their casts changed every few days for a duration of 6-8 weeks (2). Castings are often continued beyond 8-weeks as long as improvement is detected (1). However, if evaluation at approximately 2-3 months reveals that the foot is not completely corrected, it is unlikely that it will respond to further casting or other non-operative techniques (1). Although safer than surgery, non-operative techniques are generally only successful in 15-50% (17). There is also an increased risk for pressure sores, fractures and abnormal structural changes, such as rocker-bottom feet (17).

Surgical Treatment

The goal of surgery for the correction of clubfoot is to reestablish normal relationships between the bones of the foot and ankle and balance the correction of the surrounding soft tissue to prevent relapse (17). Surgical treatment of clubfoot can include

soft-tissue release, tendon transfers and bony operations (1). Surgical options for clubfoot correction became more widely available in the 1800's, when anesthesia and aseptic techniques were introduced (16). Soft-tissue releases were the most common surgical procedures performed and resulted in good outcomes in 45% of cases (16). Soft-tissue release operations were promoted between the 1970's and 1990's, but some studies have shown short- and long-term consequences associated with this type of correction (16). Complications such as overcorrection, neurovascular injuries, joint stiffness, arthritis, muscle weakness, pain and residual deformity have been reported in a series of studies (16, 17). To minimize the risks of surgery, surgeons tend to postpone surgery until an infant is 6 to 12-months of age because the foot is bigger and easier to operate on, and it decreases the risk imposed by anesthesia (17).

Unfortunately, most cases (89.2%) of clubfoot require some type of corrective surgery (8). In a review of 200 unselected clubfeet, Dangelmajor (1961) found that 60% of cases required soft tissue or bony surgery and that each foot had an average of 2.7 operations (1). The study also found that the active treatment time for patients who underwent surgery was 8 4/12 years, with 45% of individuals attaining a good outcome (1). Today invasive surgery is recommended only after serial manipulation and casting techniques have been attempted and have failed, or have produced inadequate results; although failure of correction has not been adequately described (16, 17).

Prognosis

Isolated clubfoot is not lethal; however, there are varying degrees of severity that can drastically affect an individual's prognosis and, ultimately, quality-of-life (8). In general, the more severe the deformity the more difficult it is to correct (1). Milder cases tend to correct within 2-3 months and are less likely to relapse (1). In one study, 35% of cases were corrected with non-operative techniques with no relapse at the 7-year follow-up (1). Bilateral cases tend to be more severe and harder to correct than unilateral cases (2, 4, 11). Studies have found that up to 35% of bilateral cases require multiple operations while only 22.6% of unilateral cases require multiple operations for satisfactory correction (8). In contrast, Canto *et al*., 2008, reported that bilateral cases do not have a worse prognosis than

unilateral cases. It is likely that this difference in observations is caused by a difference in description and classification of severity and prognosis.

In a population of patients treated specifically by the Ponseti method 55/104 (53%) of feet had no relapse and 54% of patients reported their results as excellent (18). Of the patients who relapsed, 25/104 (24.0%), 10/104 (9.6%) and 3/104 (2.9%) had a second, third and fourth relapse, respectively (18). Of the individuals who did not rate their results as excellent, 20% rated them as good, 14% as fair and 12% as poor (18). In a subsequent study consisting of 70 individuals, 59% of patients said that their corrected clubfoot was never painful and 72% had no limitation of activity (18).

It has been suggested that other factors, such as the time of initiation of treatment, the skill of the treating physician, the nature of the treatment, the duration and intensity of treatment and the cooperation of the parents, can also influence the prognosis (2). For example, when treatment is started after 2-months of age, there is an increase in poor outcomes (2). Of note, there has been no correlation found between family history of idiopathic clubfoot and severity of the clubfoot (11). However, one study found that 38.5% of cases with a poor result at follow-up had a family history of clubfoot while only 19.4% of cases had a family history in the whole series (2, 11). Therefore, the relationship between a family history of clubfoot and the severity, or worse prognosis, is still unclear.

ETIOLOGY AND RISK FACTORS FOR CLUBFOOT

Although clubfoot is a common and well-studied birth defect, its cause(s) and risk factors have not yet been identified. Many etiologies of nonsyndromic clubfoot have been hypothesized and include vascular obstruction, abnormal muscle development, intrauterine growth restriction and neurological abnormalities (1, 2, 19, 20). Maternal factors such as age, education and parity have also been considered, but have not consistently been shown to be associated with clubfoot (6, 8, 21, 22). In addition, segregation analyses and twin studies have pointed to a genetic etiology (6, 9, 23, 24).

Hypothesized Etiologies

While there are many theories about the causes of clubfoot, these theories can be divided into a few main hypotheses. These hypotheses are based on various studies and seek to explain the occurrence of clubfoot in all populations.

The oldest hypothesis is that there are mechanical forces that act on the developing fetus in the womb and cause deformity of the foot and ankle (2, 25). These forces can be a result of compression by the uterus, oligohydramnios, twin gestations, restriction caused by the umbilical cord or amniotic bands (2). A mechanical etiology was presented by Hippocrates and has not been definitively disproven to date (2). In fact, mechanical factors may be responsible for some cases of clubfoot, but are rare and cannot explain the majority of cases (2).

A second hypothesis is that there is arrest in fetal development that inhibits the normal rotation of the lower limb (1, 2). This hypothesis stems from the findings from Eschricht (1842) and Bohm (1929) that the feet are in a clubfoot-like position in normal development and rotate toward pronation later in development (2). The cause of the developmental arrest has not yet been determined, but the mechanism is considered to act similarly to teratogenic effects, which are affected by individual susceptibility, timing and duration of the event (1).

There have also been suggestions that malformed muscles, tendons, and bones are responsible for the development of the clubfoot phenotype (2). One study suggests that there is a relationship between clubfoot and embryonic arterial accidents (19). In contrast to these hypotheses, studies looking at the structural malformations of the clubfoot, such as muscular imbalances, have found a range of abnormalities that lack a consistent pattern or presentation (2, 21).

Finally, it is hypothesized that clubfoot is genetic and can be inherited within families through genes. Support for this theory comes from a higher prevalence among relatives of those affected than in the general population, increased concordance among monozygotic twins compared to dizygotic twins and numerous segregation analyses (2). The evidence for a genetic component to the development of idiopathic clubfoot is discussed in more detail in the following paragraphs.

Evidence for a Genetic Etiology

One piece of evidence supporting the involvement of genetic factors in the development of clubfoot comes from twin studies. By analyzing concordance between twin pairs, two twin studies in Caucasian populations both found that genes likely play a role in clubfoot development (23, 26). Idelberger (1939) studied 40 monozygotic twin pairs and 143 dizygotic twin pairs and found 32.5% concordance among the monozygotic twins and 2.9% concordance among the dizygotic twins (23). Therefore, if one child in a monozygotic twin pair has clubfoot, there is a 33% risk for the second twin to also have clubfoot (23). In dizygotic pairs, the risk is approximately 3% (23). These results were mirrored by a study performed in a Danish population of 12 monozygotic twins, 22 dizygotic twins of the same sex and 18 dizygotic twins of opposite sex (26). Pairwise concordance for monozygotic twins was 0.17 (95% CI: 0.02-0.48), giving a recurrence risk of approximately 17% (26). Pairwise concordance for dizygotic twins of the same sex was 0.09 (95% CI: 0.01-0.32) and for all dizygotic twins was 0.05 (95% CI: 0.006-0.18), giving a recurrence risk of approximately 9.1% and 5%, respectively (26).

There are many studies that have described the likely pattern of inheritance for

Table 2. Proposed inheritance patterns for clubfoot

Reference	Findings
Wynne-Davies, 1965	Multifactorial
Reimann, 1967	Irregular dominant/low penetrance or polymeric
Wang *et al.*, 1988	One major gene with additional contribution of multifactorial inheritance
Rebbeck *et al.*, 1993	Single gene, Mendelian inheritance, two alleles, incomplete dominance, with unmeasured factors shared among siblings
Lochmiller *et al.*, 1998	Major locus additive model
Andrade *et al.*, 1998	Recessive mixed gender-specific model with reduced penetrance/major autosomal locus with additional polygenic component
Chapman *et al.*, 2000	Single dominant gene with reduced penetrance
Engell *et al.*, 2006	Genetic component with predominant nongenetic factors
Kruse *et al.*, 2008	Multifactorial threshold model

clubfoot (Table 2). It is unlikely that the inheritance is solely autosomal dominant or autosomal recessive because there are affected children who do not have affected parents, and the prevalence of the deformity is the same among parents, siblings and children of probands (2). It is also unlikely that the inheritance is X-linked, either recessive or dominant, because there is transmission from father to son (2). This is, in fact, precisely what segregation analyses have concluded. By analyzing affected individuals and their pedigrees, researchers have found that a mixed model, involving multiple genes and other factors, best describes the segregation patterns seen in the families studied (7, 24, 26-30).

Pedigree analysis has also elucidated the prevalence and recurrence risks of clubfoot in families with an affected relative. In 1984, Cartlidge reported a positive family history in 54% and 30% of probands in the Polynesian and Caucasian populations, respectively. This is higher than other reports for Caucasian families, which find that 24% of cases had a first-, second- or third-degree relative with clubfoot (10, 30). In European Caucasian populations, 2.14% to 2.9% of affected individuals also had an affected first-degree relative, giving clubfoot a recurrence risk of 17 to 20 times higher than the risk for the general population (21, 23). This risk decreases when individuals are more distantly related (7, 10, 30). Second-degree relatives have a risk six times higher than the population risk while third-degree relatives have a risk close to the general population (23). In general, the risk of recurrence for siblings is approximately 3%; 2% for siblings of affected males and 5% for siblings of affected (31).

Many studies have found that the risk for clubfoot is increased further for males when they have an affected female relative (2, 21, 28). This finding may be explained by the Carter effect. The Carter effect describes a phenomenon in which one sex requires a greater genetic contribution in order to develop a condition (28). This can be seen in families if there is a discrepancy in the susceptibility to a condition and a tendency for one sex to transmit the disorder more frequently than the other (28). This effect is seen for clubfoot, as 59% of children born to mothers with clubfoot versus 37% of children born to fathers with clubfoot are also born with clubfoot (28). Females are 5.6 times more likely than males to transmit clubfoot to their children (28). The lowest prevalence of clubfoot is in daughters of men with clubfoot while the highest prevalence is in sons of females with

clubfoot (28). The Carter effect supports a multifactorial threshold model of inheritance, for which females require a greater genetic load to be affected (28).

Environmental Risk Factors

Numerous environmental risk factors have been evaluated for a relationship with the development of clubfoot. A study using birth certificates and birth defect registry information for 134 children with isolated clubfoot found male gender, death of a preterm sibling and being married were independently related to clubfoot (22). This was the first report of a relation of death of a preterm sibling and being married with clubfoot (22). The increased risk for males is consistent with the predominance of male cases and a multifactorial model (28). The study authors suggest that the relationship between clubfoot and marital status may be the result of a diagnostic bias, while the relationship with death of a preterm sibling may stem from overreporting by case mothers and underreporting by control mothers (22). No significant associations were found for other parameters, such as low parity, seasonality, young maternal age and breech prenatal position (22). Additionally, this study did not offer support for the intrauterine constraint theory (22).

Other studies have also found a lack of association between various environmental factors and clubfoot. A study with 285 cases of idiopathic clubfoot by Lochmiller, et al. (1997) found no support for variation in prevalence of clubfoot based on months/season, maternal gravidity, amniotic levels during pregnancy or breech presentation (30). Wynne-Davies (1965) found no association between parental age nor birth order.

Early amniocentesis is a procedural risk factor that has inconsistently been shown to increase the risk of clubfoot (32-34). The CEMAT group found that the incidence of clubfoot is significantly increased (p=0.0001) from 0.1% in the midtrimester amniocentesis group to 1.3% in the group that underwent early amniocentesis (before 13-weeks gestational age) (32). They also found that there was a significant increase in amniotic-fluid leakage in women who underwent early amniocentesis versus those who had midtrimester amniocentesis (p=0.0007) (32). When they examined what week of pregnancy the cases had undergone amniocentesis, they found that the highest incidence of clubfoot (2.1%) was seen when amniocentesis was performed in the 12th week of pregnancy (32). Their conclusions were that early amniocentesis, when performed between 11 weeks + 0 days and 12 weeks +

6 days, is associated with an increased risk of clubfoot (32). This was true whether or not a woman experienced amniotic-fluid leakage, however, if a woman experienced leakage the incidence of clubfoot was found to be as high as 15% (32). Similar incidences have been reported by two subsequent studies that have compared the risks of chorionic villus sampling (CVS) to early amniocentesis (35, 36).

Maternal smoking is the only single environmental risk factor shown to be significantly associated with clubfoot and was first suggested by Alderman et al. in 1991 (22). This study found an OR of 2.22 (95% CI: 0.7-6.4) for mothers who smoked 1-10 cigarettes per day and an OR of 2.90 (95% CI: 0.8-9.2; *p*-value for trend=0.3) for mothers who smoked greater than 10 cigarettes per day (22). Therefore, a greater exposure to cigarette smokes increases the association between smoking and clubfoot. The association between smoking and clubfoot is also stronger for affected females, OR = 2.28 (95% CI 0.68-7.66), compared to affected males, OR = 1.16 (95% CI 0.53 – 2.55) (10). Because females are the less frequently affected sex and, theoretically, require a greater genetic and environmental load, this finding is consistent with the Carter Effect and the findings by Kruse et. al., 2008.

Table 3. Risk of clubfoot by smoking and family history*

Risk Factors	OR (95% CI)
Smoking	1.34 (95% CI: 1.04-1.72)
Family History	6.52 (95% CI: 2.95-14.41)
Smoking + Family History	20.3 (95% CI: 7.90-52.17)

* Created from Honein, et al., 2000

Honein et al. (2000) evaluated smoking and family history of clubfoot in a cohort of 346 infants from the Atlanta Birth Defects Case-Control Study (ABDCCS) and 3,029 infants without birth defects. The study found that both maternal smoking and family history, individually, are associated with an increased prevalence of clubfoot with an OR of 1.34 (95% CI: 1.04-1.72) and 6.52 (95% CI: 2.95-14.41), respectively (Table 3) (37). Furthermore, a history of maternal smoking and a family history of clubfoot resulted in an OR of 20.3 (95% CI: 7.90-52.17) (37).

In 2002, Skelly, et al. confirmed that there is a strong association between maternal smoking and clubfoot by analyzing 239 cases of clubfoot and 365 controls in Washington State. The study found that the risk of clubfoot for mothers who smoked at any time during

pregnancy was 2.2 times that of controls (95% CI: 1.5-3.3) (38). In addition, the risk was correlated with the number of cigarettes smoked per day, resulting in a 3.9-fold (95% CI: 1.6-9.15) risk for women who smoked 20 or more cigarettes a day (38).

Lastly, in 2008, Dickinson, et al., supported the hypothesis that maternal smoking is associated with an increased risk of clubfoot with an OR of 1.79 (95% CI: 1.17-2.74) (39). This study also found a stronger association between maternal smoking and female fetuses versus male fetuses, but did not support a dose dependent association in any dataset (39).

Together these results imply an increased risk of clubfoot for children of mothers who smoke during pregnancy (Table 4). The risk for clubfoot is increased further for children of mothers who smoke during pregnancy and have a positive family history, supporting an etiology with environmental and genetic interactions.

Table 4. Odds ratio for clubfoot in mothers who smoke during pregnancy

Study	Cases of Clubfoot	Smoking OR (95% CI)
Alderman, et al., 1991	175	2.6 (95% CI: 1.6-4.0)
Honein, et al., 2000	346	1.34 (95% CI: 1.04-1.72)
Skelly, et al., 2002	239	2.2 (95% CI: 1.5-3.3)
Dickinson, et al., 2008	443	1.49 (95% CI: 1.15-1.92)

FINDING DISEASE-SUSCEPTIBILITY GENES IN MULTIFACTORIAL TRAITS

Finding genes responsible for susceptibility to multifactorial diseases is important because identification of causative factors can provide information about the pathophysiology of the disease, increase the understanding of human development and biology, improve diagnosis and help identify improved therapies (40). Multifactorial inheritance, by definition, involves the complex interaction of multiple genes and environmental factors. Therefore, many of the methods used to identify genes responsible for disorders with simple Mendelian inheritance are not as useful for common, more complex disorders. The methods for identifying disease genes range from positional cloning to genome-wide association studies (GWAS) and utilize different statistical methods and

technologies (41-43). However, all methods are based on the principles of genomic structure and segregation.

Genetic Markers and Variation

The genome contains infinite variability within a population. This variation is created by changes in the coding regions of genes, as well as intragenic and intergenic sequences that makes up the genetic code (44). When these variants are present in at least 1% of the population, they are considered normal variants or polymorphisms. Researchers can identify and analyze the inheritance of unique variations within an individual's genome to look for genes that may play a role in a genetic susceptibility to a disease, condition or trait. Over the years, many different types of genetic markers have been identified and utilized for scientific research (45). Restriction fragment length polymorphisms (RFLP), variable number of tandem repeats (VNTR or minisatellites) and short tandem repeats (STRP or microsatellites) became widely used in the 1980's (45). These markers consist of variable lengths of DNA sequence that can be detected through polymerase chain reaction (PCR). There are a few thousand VNTRs in the human genome, while there are >100,000 microsatellites that cover the majority of the genome (45). Microsatellites tend to be highly heterozygous and amenable to analysis using high throughput technology (45).

At the beginning of the twenty-first century, single nucleotide polymorphisms (SNPs) became widely used and were found to be the most prevalent type of variation in the genome, occurring every 300 base pairs (45). Today, more than three million SNPs have been described (45). These markers are useful tools for identifying genetic variants that can affect susceptibility to a disease in a population (45). SNPs are found in coding and non-coding regions within genes and between genes and can be used in linkage and association analyses (46). They can be causative or indicative of a disease based on their interaction or effect on gene function or regulation (46). Identification of SNPs, improvements in high through-put technology and reduction in costs and labor have allowed for genome-wide association studies (GWAS) which can analyzed an entire genome for association with a disease susceptibility or trait (45).

Heritability

Past and present methods of evaluating a genetic etiology of a condition or disease trait have started by establishing the heritability of a condition (41, 43). Heritability measures how much of the phenotypic variation in a condition is caused by genetic variation (41). Heritability can be established through twin studies and segregation analyses (41). Twin studies are based on the premise that monozygotic twins share 100% of their genes while dizygotic twins share 50% of their genes, like non-twin siblings. Both monozygotic and dizygotic twins share an environment during fetal development. Therefore, if a trait has a genetic component, and is not purely the result of environmental factors, there should be a greater concordance between monozygotic twins when compared to dizygotic twins (41). Segregation analysis seeks to characterize genes with a major role in pathogenesis by proposing a mode of inheritance and genetic parameters and determining the model that best fits the segregation patterns within a family or group of families (41). By studying twins and families with a condition or disease trait, information about whether or not a condition is genetic and how that condition is most likely inherited (autosomal dominant, autosomal recessive, X-linked recessive, multifactorial, etc.) can be deduced. Once the etiology of a trait is determined to have a genetic component, more analyses can be undertaken to identify which genes are responsible for pathogenesis.

Linkage and Association

Linkage analysis is one approach to localize a disease susceptibility gene by narrowing the chromosome region through observation of recombination events within families (41, 42). Multiplex families are required for linkage analysis. The result of a linkage analysis is translated into a physical genetic distance, which can be used to create a genome-wide genetic linkage map for identification of susceptibility genes at multiple loci (41). Linkage analysis is useful when the gene of interest is unknown. This method is underpowered when there are many low penetrant genes involved and resolution is hindered if there are few generations within the families being studied (41, 42).

Association analysis is a second approach, which utilizes linkage disequilibrium (LD) to identify a marker in close proximity to a disease susceptibility allele (43). A marker is in LD with a disease allele when there is a small probability of crossover between the two

loci, which means the marker and the allele may be close together on a chromosome (43). Association studies use case-control, simplex or extended families. This method is generally more powerful than linkage analysis and offers a more narrow genomic interval within which the gene of interest may be contained (42).

Gene Discovery

The traditional route used to identify causative genes for Mendelian disorders has been positional cloning by linkage analysis (42). This method is able to find a location of the gene likely causative for the phenotype within the genome when the gene is not yet known. Positional cloning utilizes families with a condition of interest to perform linkage analyses and/or association tests to map a gene to a small interval within the genome (42, 43). Once an area is defined, the genes within the interval can be evaluated to determine if their biological function has relevance to the observed trait (42). Researchers can then look at the most probable genes and scan for disease-causing mutations (42).

Positional cloning has had only limited success for multifactorial disease because of the weak relationship between any one locus and the observed phenotype (42). Linkage analysis, population-based association studies, and chromosomal deletions, duplications or rearrangements can help identify a genomic region of interest, but often the interval is large and contains many hundreds of genes (42). Therefore, most research on complex disease susceptibility has focused on testing plausible candidate genes through linkage and association approaches (41, 42). This method requires knowledge about the biology of complex disease and the function of candidate genes in an organism.

The candidate gene approach uses information about the biology of the disease, including biochemical pathways, tissue expression profiles, differential expression studies and animal models to identify genes with a likely involvement in the susceptibility to the disease state (42). Once identified, the genes can be prioritized for the ones most likely to play a significant role in pathogenicity (42). The segregation patterns of polymorphic markers in flanking high-priority candidate genes can be analyzed to look for disease-associated variation (42).

Using the candidate gene approach, studies have begun to examine whether there is an association between SNP variants in and around genes and SNP haplotypes with

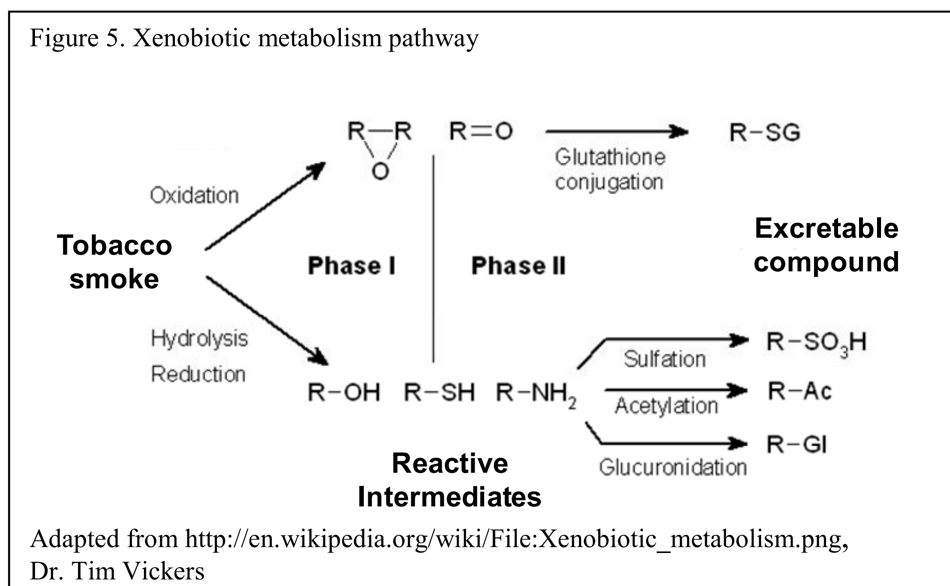
clubfoot. Genes involved in limb development and apoptosis have been examined and appear to play a modest role in the etiology of clubfoot (47-49). Few studies have been performed on smoking metabolism genes (J. T. Hecht, et al., 2007). Based on the consistent association between smoking and clubfoot and the previous *NAT2* association with clubfoot, a systematic interrogation of smoking metabolism genes is needed.

SMOKING METABOLISM

Xenobiotic metabolism of cigarette smoke and PAH adduct formation

Metabolism of the components of cigarette smoke is accomplished through the xenobiotic metabolism pathway (Figure 5) (50). The pathway involves biotransformation of a lipid-soluble xenobiotic compound by functionalization and/or conjugation reactions into polar, water-soluble metabolites that can be excreted (50, 51). This pathway consists of two phases, which are denoted phase I and phase II. Phase I is characterized by the

Figure 5. Xenobiotic metabolism pathway



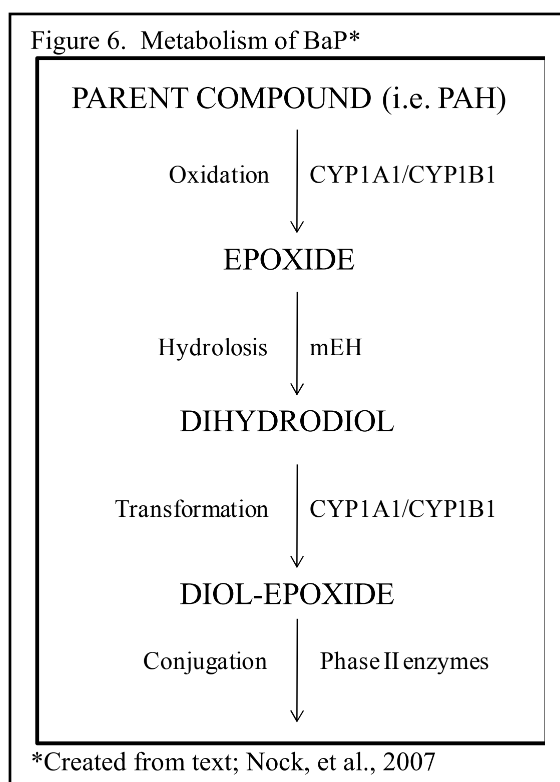
Adapted from http://en.wikipedia.org/wiki/File:Xenobiotic_metabolism.png, Dr. Tim Vickers

functionalization reactions and utilizes enzymes capable of dehydrogenation/hydrogenation, oxidation, hydrolysis, reduction and mono-oxygenation (50). Phase II consists of the conjugation reactions and utilizes enzymes capable of glucuronidation, sulphation, acetylation, GSH-conjugation and methylation (50). Biotransformation can detoxify a compound or create a more toxic intermediate metabolite (50). The effects of the

intermediate metabolite are dictated by the type of environmental exposure (parent compound) and by the effectiveness/activity level of both phases of xenobiotic metabolism.

Cigarette smoke is one type of environmental exposure and consists of more than 4,000 chemical compounds, including dioxins, dioxin-like compounds and other AhR agonists (52, 53). The main toxins in cigarette smoke are the polycyclic aromatic hydrocarbons (PAHs) (53, 54). Individuals are exposed to PAHs everyday through fossil fuel combustion, forest fires and car exhaust; however, the greatest exposure to PAHs comes from cigarette smoke (55, 56). Metabolism of these toxins occurs primarily in the liver, which expresses numerous drug-metabolizing enzymes (DME) (50, 57, 58).

Figure 6. Metabolism of BaP*



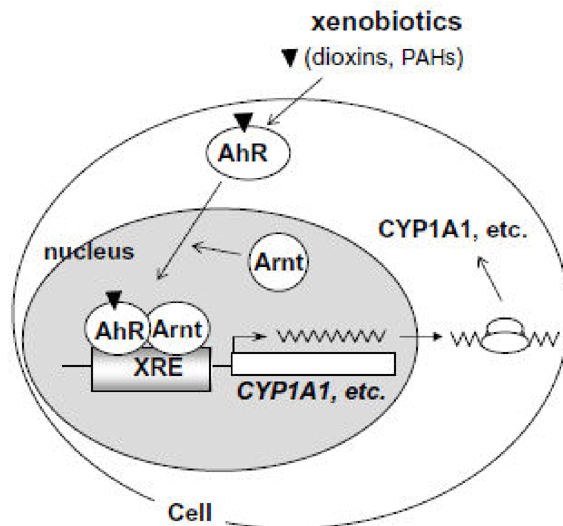
*Created from text; Nock, et al., 2007

PAHs from cigarette smoke, such as Benzo[a]pyrene (BaP) and 7,12-dimethylbenz[a]anthracene (DMBA) form DNA and protein adducts (56). PAH-DNA adducts and dioxins from cigarette smoke can cause mutagenesis and teratogenesis (52, 59). The mechanism for xenobiotic metabolism and adduct formation is specific to a unique compound, however, benzo(a)pyrene (BaP) can be used as a prototypic PAH to discuss the general metabolism of all lipophilic xenobiotics (Figure 6) (57). Once in the body, BaP is initially metabolized by *CYP1A1* or *CYP1B1* to an epoxide, such as benzo(a)pyrene-7,8-epoxide (60, 61). The compound is then hydrolyzed by the microsomal epoxide hydrolase (mEH) enzyme to a dihydrodiol (59). *CYP1A1* or *CYP1B1* can transform the intermediate compound to a highly reactive diol-epoxide that can covalently bind DNA (59). Not surprisingly, it has been shown that smoking appears to be associated with an increase in DNA adduct levels (59). Additionally, increased metabolism of PAH-diol-epoxide forms with decreased capacity to conjugate these reactive intermediates was associated with an increased level of adducts (59).

Regulation of CYP1 enzymes

PAHs and dioxins also induce the expression of xenobiotic metabolism enzymes by binding to the aryl hydrocarbon receptor (AhR) (Figure 7) (50, 56). The AhR is a transcription factor that acts as a xenobiotic sensor for a number of different hydrocarbons, including PAHs (56, 62). BaP and DMBA are two compounds that can diffuse across the cell membrane and bind to AhR as ligands (52, 56). AhR-ligand binding causes a transformational shift, which, in turn, exposes a nuclear localization sequence that allows the receptor-ligand complex to enter the nucleus (56). In the nucleus the complex binds the AhR nuclear translocator (Arnt), which then activates the transcription of cellular detoxification enzymes (52, 56). *CYP1A1*, *CYP1A2* and *CYP1B1* can all be induced through the AhR cascade (52, 62, 63).

Figure 7. Regulation of *CYP1* enzymes



Reprinted from Cancer Letters, 252 /2, Masanori Kitamura & Ayumi Kasai, Cigarette smoke as a trigger for the dioxin receptor-mediated signaling pathway, 184-194 (2007), with permission from Elsevier.

Effects of Smoking During Pregnancy

Damaging PAHs and other lipophilic substrates have been shown to cross the placenta and can form adducts to maternal and fetal tissues as well as in the DNA of human trophoblast cells (57, 64-66). Therefore, it is important for both maternal and fetal health that these toxic tobacco smoke metabolites can be converted to less damaging substances (67). The human placenta plays an important role in the oxidation of several xenobiotics (57, 66).

In their review of the transcriptional regulation of xenobiotic metabolism genes, Pavek and Dvorak (2008) highlight the role of cytochrome P450 enzymes in the placenta. They note that many CYP enzymes are expressed in placental trophoblast cells, but not all have a detectable enzymatic activity (57, 62). The amount of CYP enzymes expressed in the placenta is greatest in the first-trimester, during embryogenesis and organogenesis, and decline throughout the second- and third-trimesters (62). *CYP1A1* is expressed and can be induced in placental cells throughout pregnancy while *CYP1A2* expression is only detected in the first-trimester placenta and *CYP2A6* expression has not been detected in the placenta at any time during pregnancy (57, 62). Basal *CYP1A1* expression is inhibited during pregnancy by 30% in the liver and up to 60% in extrahepatic organs (66). Interestingly, there is an observed increase in the concentration of AhR and Arnt within placental trophoblast cells (62). This is consistent with the ability to induce *CYP1A1* expression with exposure to cigarette smoke(57). *CYP1B1*, although regulated by the same cascade as *CYP1A1*, is not inducible by maternal cigarette smoking in the placenta (57). As is seen in maternal cells, placental *CYP1A1* plays a role in the bioactivation of PAHs to reactive intermediates that form DNA adducts in placental and fetal tissues (62). The inductory mechanism for *CYP1A1* is functional in the fetal liver at day 21 of pregnancy in rats (66). Smoking-induced elevations in *CYP1A1* activity have been consistently associated with adverse birth outcomes, such as premature birth, IUGR and structural abnormalities (62). Mice exposed to PAHs had abnormal vasculature in the placenta that significantly reduced arterial surface area and volume of the fetal arterial vasculature (56).

An individual's ability to metabolize xenobiotics can be greatly affected by the genetic variation in their drug metabolizing genes. Numerous xenobiotic genes and variation within these genes have been studied. Polymorphisms affecting the activity of *CYP1A1*, *GSTM1* and *GSTT1* have been shown to be detrimental to the growth and development of a fetus when the fetus is exposed to cigarette smoke (68-70). Additionally, BaP has been shown to differentially impact the incidence of congenital malformations based on maternal and fetal genotype (71). Therefore, genetic variation in xenobiotic metabolism genes may help explain the increase in adverse effects among some individuals.

Xenobiotic metabolism and genetic variation

An interaction between metabolism genes and cigarette smoking has been suggested in previous studies (68). Consequently, the observed increased risk for clubfoot due to the interaction between genes and smoking may be caused by variation within the genes involved in smoking metabolism.

Cytochrome P450

The cytochrome P450 (CYP450) superfamily genes are phase I mono-oxygenases that are anchored in the endoplasmic reticulum (50, 72). They are key players in phase I of xenobiotic metabolism and can catalyze oxidation and reduction reactions (63). Oxidation and reduction can convert xenobiotics to water-soluble compounds, a process known as detoxification, or it can increase the toxicity of a compound by creating an active metabolite that is a target for phase II conjugation reactions (51, 63).

Humans have 57 cytochrome P450 genes, which are divided into 18 families and 43 subfamilies (62). Cytochrome P450 enzymes are labeled based on a set of standard nomenclature (72, 73). The enzymes are named first by a number representing their family, followed by a letter representing the subfamily and, finally, a second number that identifies the individual enzyme (72). Fifteen of these genes are known to play an important role in phase I of the metabolism of xenobiotic compounds and are from the CYP1, CYP2 and CYP3 families (62, 73). Smoking is an important environmental factor that influences CYP450 activity (63). *CYP1A1*, *CYP1A2* and *CYP1B1* are all known to play a role in the metabolism of compounds found in tobacco smoke, such as polycyclic aromatic hydrocarbons (PAHs) and dioxins (62). *CYP1A1*, *CYP1A2*, *CYP1B1* and *CYP2A6* expression has been identified in key organs, such as the placenta, liver and lungs (Table 6) (62).

Variation within these genes has been identified. In general, the most penetrant CYP450 genetic alterations are deletions, missense mutations and splicing defects (63). There have also been a few examples of mutations in the 5'- and 3'-untranslated regions (UTR) that affect phenotypes (63). By looking at the variation within these key CYP450 genes, the effects of specific polymorphisms on enzyme activity and smoking metabolism can be studied and hypothesized.

CYP1A1 is known as one of the most important detoxification enzymes because it has a broad substrate specificity and wide distribution throughout the body (62). *CYP1A1* can be induced by cigarette smoking (74). *CYP1A1* activates and detoxifies environmental PAHs and aromatic and heterocyclic amines from cigarette smoke (63). It is also known to produce highly carcinogenic intermediate metabolites through oxidation of PAHs (62). Specifically, induction of *CYP1A1* catalyzes the activation of BaP to DNA-bound adducts (57).

Study of the variation in *CYP1A1* has offered some insight into its role in xenobiotic metabolism. The *CYP1A1* Ile462Val polymorphism has higher enzymatic activity (Table 7) (75, 76). The *CYP1A1**2A allele had higher adduct levels and percent aberrant cells in the presence of significant environmental tobacco smoke exposure (77).

CYP1B1 is differentially expressed between tissues but is primarily extrahepatic, with the highest constitutive levels of mRNA detected in the uterus, heart, brain, lung, skeletal muscle and kidney (62, 63). Like CYP1A, CYP1B isoenzymes metabolize various polycyclic aromatic hydrocarbons and, additionally, are involved in the metabolism of endogenous compounds (62). Because of its localization within the body, *CYP1B1* is not believed to play a major role in the overall clearance of drugs and more likely plays a critical role in tissue-specific metabolism of certain compounds (63).

Specific variants in *CYP1B1* have been described (Table 7). Five common missense SNP mutations have been identified and form 7 haplotypes consisting of one or more of these SNPs (63). One of these haplotypes, *CYP1B1**7 has a significantly decreased ability to metabolize BaP (78). Additionally, the 432Leu allele has a slightly higher activity in metabolizing BaP-7,8-dihydrodiols but slightly lower activity in metabolizing the parent compound, BaP when compared to *CYP1B1* 432Val (59). *CYP1B1* polymorphisms seem to have an impact on DNA adduct level in populations exposed to low levels of tobacco smoke (77). One allele, *CYP1B1**3, seems to increase an individual's susceptibility to DNA adduct formation, although it was not a significant increase (77).

CYP1A2 is a hepatic form of CYP450s (57). It is constitutively expressed but can be induced by PAHs and dioxins (62). There have not been any common polymorphisms identified with functional effects and there have been only a few rare variations described (*CYP1A2**7 and *CYP1A2**11) (63).

CYP2A6 is found in the liver and metabolizes a variety of tobacco-related precarcinogens (74, 79). *CYP2A6* is the primary human nicotine C-oxidase and is responsible for 70-80% of nicotine metabolism (79). *CYP2A6* is not inducible by the same cascade as the CYP1 family (62). *CYP2A6* is highly polymorphic but does not have functionally important polymorphisms in Caucasians (79). One allele, *CYP2A6**2 is functionally significant in Caucasians, but it is very rare (79). One of the most important variant alleles is a gene deletion and is seen almost exclusively in Asian populations (79).

Variations in P450 genes result in different rates of drug metabolism among individuals and several known variants produce unusual metabolites that may have harmful effects (51). Therefore, variations in *CYP1A1*, *CYP1A2*, *CYP1B1* or *CYP2A6* may modify (either increase or decrease) the effect of harmful compounds found in cigarette smoke and their metabolites.

Epoxide Hydrolase

Epoxide hydrolases (EH), such as *EPHX1*, also activate and detoxify exogenous compounds, including PAHs, during phase I of xenobiotic metabolism (80). As discussed previously, EHs metabolize reactive epoxides to less-harmful dihydrodiol derivatives and can activate PAH (80, 81). *EPHX1* is the gene that encodes the microsomal epoxide hydrolase (mEH), which is located primarily in the endoplasmic reticulum (81). The gene, located on 1q42.1, is 20,271 bp and consists of 9 exons and 8 introns (81, 82). mEH is constitutively expressed in diverse cell types and is found in different organs throughout the body (fetal liver, adrenals, kidneys, lungs, gut and placenta) (83, 84). *EPHX1* is also inducible and has polymorphic variants (81).

In 1985, researchers hypothesized that a mutation in *EPHX1* could lead to an increase in toxic arene oxide metabolites and result in a variety of biological insults, including birth defects (85). Epoxides are 3-membered strained organic configurations of oxygen that can cause cellular damage and genetic mutations and have been shown to produce birth anomalies (57, 82). It has since been shown that certain polymorphic variants in *EPHX1* decrease its activity by 40%, inhibiting effective biotransformation of exogenous compounds (81). Currently there are 2 known polymorphisms in the coding region of the *EPHX1* gene (80). One polymorphism is located in exon 3 at the amino acid residue 113

and can be seen as a histidine or a tyrosine and the other is in exon 4 at residue 139 and can be a histidine or an arginine (86). The wildtype haplotype contains a tyrosine in amino acid position 113 and a histidine in amino acid position 139 (113Y/139H) and has been shown to have an approximately 2-fold increase in activity when compared to variant forms, specifically in the epoxide to dihydrodiol reaction (80, 87).

Because mEH plays an important role in the metabolism of reactive epoxides, these *EPHX1* variants have been evaluated for association with adduct levels. Caucasian individuals who are homozygous for the arginine allele at amino acid position 139 have been shown to have decreased levels of DNA adducts compared to wildtype (59, 77). Individuals heterozygous at the 139 amino acid position trended toward a significant decrease in adducts when the three allele combinations where compared (77). The highest adduct levels were seen in individuals who were homozygous for the histidine allele at amino acid position 139. Therefore, the adduct levels increase with the number of histidine alleles present. Additionally, individuals homozygous for the histidine allele or heterozygous at position 139 who also had a tyrosine allele at position 113 have been shown to have significantly higher PAH-DNA adducts (59). Based on these studies, genetic variation in *EPHX1* may contribute to the impact of an individual's exposure to harmful cigarette smoke toxins and metabolic intermediates.

Glutathione S-Transferase

Glutathione S-Transferases (GSTs) belong to a multifunctional enzymatic system that catalyze detoxification and activation reactions through the conjugation of biologically active electrophiles to endogenous tripeptide glutathione, predominantly in the liver (57, 88). There are 4 classes of GSTs, alpha, mu, theta and pi and all play a role in furthering the biotransformation of metabolites from phase I reactions (88). Of note, GSTP (pi) is the only GST to be purified and cloned from the human placenta and represents 85% of the GST activity in the placenta as early as the first trimester (57).

Table 5. Relative expression of human CYP1, 2 and 3 subfamilies in normal extrahepatic tissues at the level of mRNA in comparison with the liver*

Enzyme	Small intestine	Kidney	Lung	Placenta	Liver	Main organ of expression	Inducer/enhancer
CYP1A1	+	+	[a]+++, +/-	++/+	++	Extrahepatic	PAH, cigarette smoking, PCB
CYP1A2	-	-	+/-	[b]+/-	+++	Liver	Smoking
CYP1B1	+	++/+	++/+	+	+	Extrahepatic (fetal tissues)	None known
CYP2A6	-	-	++/+	+/-	+++	Liver	Unknown

[a] = expressed in the first trimester

[b] = smokers, +/- mRNA expression in non-smokers

+++ = organ with high expression; ++ = organ with moderate expression; + = low expression; - = undetectable expression; +/- = controversial expression or reports

*Created from Pasanen, 1999; Pavek & Dvorak, 2008

Table 6. CYP450 functional alleles*

Gene	Haplotype	SNP ID**	Nucleotide change	Amino acid change	Phenotype
CYP1A1	*CYP1A1*2A*	rs4646903	3698T>C	5' near gene	Higher adduct levels and % aberrant cells when exposed to cigarette smoke
	*CYP1A1*2B*	**rs1048943**; rs4646903	**2454A>G**; 3698T>C	**Ile462Val**	Increased enzymatic activity
	*CYP1A1*2C*	**rs1048943**	**2454A>G**	**Ile462Val**	Increased enzymatic activity
CYP1B1	*CYP1B1*1*	WT	none	none	Wildtype
	*CYP1B1*2*	**rs10012**; rs1056827	**142C>G**; 355G>T	**Arg48Gly**; Ala119Ser	-
	*CYP1B1*3*	**rs1056836**	**4326C>G**	**Leu432Val**	Decreased met of BaP, increased met of BaP-7,8-dihydrodiols; increased susceptibility to DNA adduct formation
	*CYP1B1*4*	rs1800440	4390A>G	Asn435Ser	-
	*CYP1B1*5*	**rs10012;** **rs1056836**	**142C>G;** **4326C>G**	**Arg48Gly;** **Leu432Val**	-
	*CYP1B1*6*	**rs10012**; rs1056827; **rs1056836**	**142C>G**; 355G>T; **4326C>G**	**Arg48Gly**; Ala119Ser; **Leu432Val**	-
	*CYP1B1*7*	**rs10012**; rs1056827; **rs1056836**; rs4986888	**142C>G**; 355G>T; **4326C>G**; 4360C>G	**Arg48Gly**; Ala119Ser; **Leu432Val**; Ala443Gly	Significantly decreased ability to metabolize BaP

* Created from http://www.cypalleles.ki.se; Lamba, et al., 2002; Schwarz, et al., 2005; Georgiadis, et al., 2004; Nock, et al., 2007; Akilillu, et al., 2005; Rodriguez-Antona & Ingelman-Sundberg, 2006

**SNPs in this study in bold

The GST mu (GSTM) and GST theta (GSTT) classes have been studied because of their role in detoxification of activated nicotine metabolites and xenobiotics (57). They are known for their involvement in the detoxification of epoxides created by the CYP450s (89). There are 5 sub-classes of GSTM (*GSTM1*-5), which cluster on chromosome 1p13 (90, 91). There is a large difference in the expression of GSTM between different tissue types (88). The most commonly expressed GSTM is *GSTM1* (88). The *GSTM1* subclass is of particular interest because of its prominent null allele. Only about 40-60% of individuals in the population express *GSTM1* and, for those who do not express the gene, there is an increased susceptibility to DNA-adduct formation and cytogenetic damage (92-94). GSTT has two sub-classes, designated *GSTT1* and *GSTT2* (95, 96). These enzymes are also found in the liver, but have widespread expression (88). Similarly to the *GSTM1* locus, *GSTT1* has a null allele that can be found in 10-40% of individuals, depending on the population (88). The null alleles for both *GSTM1* and *GSTT1* cause an absence of enzyme activity, and possibly increasing the amount of active metabolites in the body (88, 97). Therefore, *GSTM1* and *GSTT1* activity may impact the effects of harmful intermediates created by phase I enzymes on a fetus (57).

N-Acetyltransferase

N-Acetyltransferases, *NAT1* and *NAT2*, are xenobiotic enzymes whose genes are located on chromosome 8p22 (http://www.ncbi.nlm.nih.gov/entrez/dispomim.cgi?id=612182). There are 20 described allelic variants in these two genes that are caused by single nucleotide substitutions or insertions/deletions (47, 57). Because of their known role in phase II of the biotransformation of tobacco smoke, *NAT1* and *NAT2* were the first xenobiotic metabolism candidate genes considered for association analysis in clubfoot (47, 98, 99).

NAT2 is of particular interest in this study because of its "slow acetylator" phenotype (100). The slow acetylation phenotype is associated with a 10-20% reduction in *NAT2* protein levels and can cause an increase in harmful adduct levels (47, 100, 101). Approximately 40-70% of European and Northern American individuals have the "slow acetylator" phenotype while it is found in only 10-30% of Asian individuals (100). Many functional alleles can cause the "slow acetylator" phenotype and have been shown to be

recessively inherited (100). In total, there are 7 point mutations in *NAT2* coding regions, 5 of which cause amino acid changes (100). These mutations and the associated "slow acetylator" alleles are presented in Tables 8 and 9. The "slow acetylator" phenotype is observed in individuals who are homozygous for the G191A, C282T and T341C polymorphisms (102, 103). Conversely, individuals homozygous for the A803G polymorphism have been classified as "rapid acetylators" (104). There have been no homozygotes identified for the C418T, G590A and G857A polymorphisms (100). Almost all (99%) of haplotypes in Caucasian slow acetylators share either a T341C or C282T polymorphism (*NAT2**5 (A, B or C), *NAT2**6A, *NAT2**7B or *NAT2**13) (100).

Table 7. Acetylation activity of common *NAT2* SNPs*

SNP	Mutation	Amino acid change	N-acetylation activity	
			in vivo	*in vitro*
rs1041983	C282T	Silent (Y94)	↓	normal
rs1801280	T341C	I114T	↓	↓
rs1799929	C481T	Silent (L161)	unknown	normal
rs1799930	G590A	R197Q	unknown	↓
rs1799931	G857A	G286E	unknown	controversial

↓ = decreased N-acetylation activity level

*Created from U.A. Meyer & U.M. Zanger, 1997

Table 8. Common *NAT2* slow acetylator alleles*

Mutations[a]	**Allele Designation**
T341C, C481T, A803G	*NAT2*5B*
C282T, G590A	*NAT2*6A*
T341C, C481T	*NAT2*5A*
T341C, A803G	*NAT2*5C*
C282T, G857A	*NAT2*7B*
C282T	*NAT2*13*
G191A	*NAT2*14A*

[a]Presumed inactivating mutations in bold

*Created from U.A. Meyer & U.M. Zanger, 1997

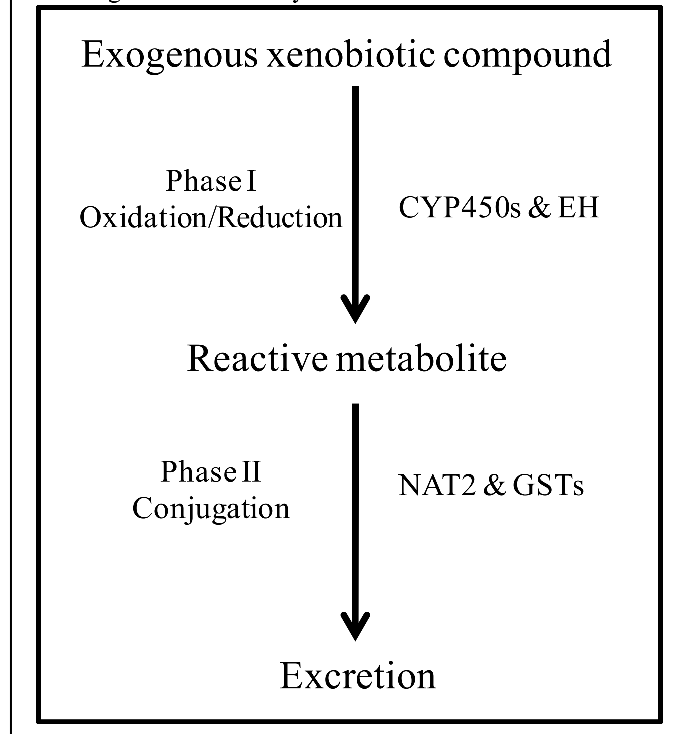
A study by Hecht, et al. (2007) suggested that the slow-acetylator phenotype of *NAT2* may play a role in the development of clubfoot. This is supported by the finding that NAT enzymes are found in the human placenta and can bioactivate certain arylamines into compounds that can be toxic to the fetus (57). The study found that the T341C polymorphism was transmitted more often in Hispanic simplex cases (only the proband is affected) (47). Additionally, the G590A normal SNP and haplotype was transmitted less often in the Hispanic clubfoot population (47). These findings add support for the interaction between smoking metabolism genes in the lipid-soluble xenobiotic pathway and clubfoot.

Together, studies have delineated the importance of *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2A6*, *EPHX1*, *GSTT1*, *GSTM1* and *NAT2* in the metabolism of cigarette smoke. The genes in this study interact in the sense that they all play a role in a pathway responsible for metabolizing cigarette smoke, both through activating and detoxifying reactions. Because the level of genotoxic damage in individuals is the result of complex gene-environment and gene-gene interactions, genetic variants within multiple genes may interact and cause phenotypes that are less able to detoxify exogenous compounds or create toxic intermediate metabolites (77). Specifically, individuals with polymorphisms that increase the activity in activating reactions and decrease the activity of inactivating reactions are more susceptible to the effects of genotoxic compounds (77). Therefore, a mutation in any one of these genes may have a large impact on the efficiency of biotransformation of toxic compounds and the interaction of multiple mutations may increase the susceptibility even further.

In summary, clubfoot is a common congenital anomaly that has been recognized for hundreds of years. Despite this attention, very little is known about the etiology. Recent studies have implicated an interaction between maternal smoking and genetic variation as a possible cause of clubfoot. Maternal smoking during pregnancy is an important and prominent risk factor to study. In the United States in 2007, the overall rate of cigarette smoking during pregnancy was 13.2% and the rate for NHW women (18.1%) was more than six times the rate for Hispanic women (2.8%) (105). This study will analyze the variation within specific smoking metabolism genes involved in both phases I and II of cigarette smoke metabolism (Figure 8) to determine if this variation, in conjunction with maternal smoking, could have an etiologic role in the development of clubfoot in NHW and Hispanic

populations. Findings from this study may offer insight into the etiology, mechanism and risk of clubfoot in the nonHispanic white and Hispanic populations.

Figure 8. Xenobiotic metabolism pathway highlighting the role of genes in this study



MATERIALS AND METHODS

IRB Approval

This study was reviewed and approved by the Committee for the Protection of Human Subjects at the University of Texas Health Science Center (HSC-MS-09-0328) and all collaborating centers (see below).

Clubfoot Study Samples and Sample Preparation

Families in the study data set were identified through a proband at six orthopedic centers: Shriners Hospital for Children of Houston and Los Angeles Texas Scottish Rite Hospital for Children of Dallas, University of Iowa and University of British Columbia. The diagnosis of clubfoot was based on the presence of adducted forefoot, varus hindfoot, and ankle equinus deformities and determined by either examination and/or by review of medical records. Only patients with isolated clubfoot were included in the study. All patients with syndromic clubfoot or with multiple malformations were excluded.

Family history and exposure information were obtained by interview with the proband's mother and/or by chart review. Ethnicity for each family was recorded based on self-reporting. Only nonHispanic white (NHW) and Hispanic families were included in this study. Two-generation pedigrees were constructed for all families. Pedigrees were extended to include all affected individuals if a positive family history was reported. Probands were recorded as having a positive or negative family history, which was used in the analysis. Blood or saliva samples were collected on all available family members. DNA was extracted from the blood or saliva using either the Roche DNA Isolation Kit for Mammalian Blood (Roche, Basel, Switzerland) or the Oragene Purifier for saliva (DNA Genotek, INC., Kanata, Ontario, Canada) following the manufacturer's protocol.

Table 9. Composition of clubfoot dataset by ethnicity and family history

Family type	Families			Total
	Total	NHW	Hispanic	Individuals
Multiplex	*242*	*149*	*92*	*853*
Simplex	*377*	*149*	*226*	*923*
Total	**619**	**298**	**318**	**1776**

The study dataset consisted of 1,776 individuals from 619 families (Table 10). Families were considered to be multiplex or simplex based on the presence or absence of a family history of clubfoot, respectively. The sample consisted of 242 multiplex families (149 NHW and 92 Hispanic) and 377 simplex families (149 NHW and 226 Hispanic).

Gene and SNP Identification and Genotyping

Candidate genes were selected following a thorough literature search of relevant publications. Hundreds of enzymes are known to play a role in xenobiotic metabolism, however only a subset of these enzymes are well-characterized and an even smaller subset have been shown to be specifically involved in the metabolism of tobacco smoke. Only genes known to metabolize compounds in cigarette smoke and that interact in multiple steps of a common general pathway were included in this study.

SNPs in the *CYP1A1*, *CYP1A2*, *CYP1B1*, *CYP2A6*, *EPHX1* genes were identified using NCBI and Ensembl websites. SNPs in the *NAT2* gene were identified previously (47). SNPs were selected based on a standard set of criteria including: heterozygosity >0.3, inter- and intragenic positions, coverage of the gene and tagging ability. SNPs with a higher heterozygosity that caused a missense mutation and/or tagged for multiple SNPs were preferred. Many SNPs in the target genes had low heterozygosity. Information about the SNPs identified for this study is presented in Table 11. Once identified, SNPs were genotyped using TaqMan Genotyping Assays (Applied Biosystems, Foster City, CA) and detected on a 7900 HT Sequence Detection System (Applied Biosystems). Results that could not be interpreted for any given sample and/or SNP were re-genotyped. All genotype data was entered into a Progeny database and checked for incompatibility with Pedcheck. Pedigrees with conflicting genotyping results that could not be resolved were eliminated from subsequent analyses.

Figure 9. Optimized Takara *Ex Taq* Polymerase PCR protocol

PCR Step	Temp. (°C)	Duration	Number of repetitions
Primary denaturation	95	2 minutes	1
Denaturation	94	1 minute	30
Annealing	64	1 minute	
Extension	72	1 minute	
Final elongation	72	5 minutes	1

GSTM1 and *GSTT1* were genotyped to identify individuals with wild type or null alleles following the protocol of Arand *et al.* (1996). The *GSTM1* and

GSTT1 alleles were amplified simultaneously via Takara *Ex Taq* Polymerase PCR (Takara Bio USA). The PCR reaction protocol was modified by an addition of 0.5 µL of MgCl per reaction and by using an optimized annealing temperature of 64˚C. The modified Takara *Ex Taq* Polymerase PCR protocol is shown in Figure 9. The Arand method combines primers for *GSTM1*, *GSTT1* and *ALB* as an internal positive control into one assay (106). The *ALB* product is 350 bp in length while the *GSTM1* and *GSTT1* products are 215 bp and 480 bp long, respectively (106). Amplified samples were run on a 2% agarose gel and scored according to the presence of the wild type or null allele for both *GSTM1* and *GSTT1* (Figure 10). Individuals with an absence of *GSTM1* or *GSTT1* were scored as having null alleles (106). Individuals showing the presence of *GSTM1* or *GSTT1* were identified as having at least one allele. Discrimination of individuals being heterozygous for *GSTM1* and *GSTT1* could not be achieved using this assay.

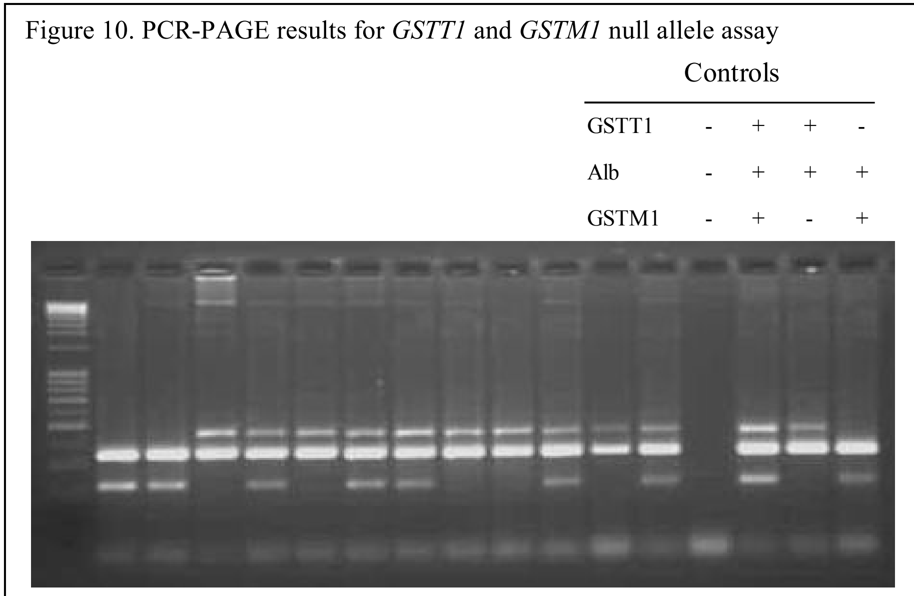


Figure 10. PCR-PAGE results for *GSTT1* and *GSTM1* null allele assay

Statistical Analysis

Allele frequencies and Hardy-Weinberg equilibrium (HWE) were calculated using SAS (v9.1). SNPs found to be out of HWE (P<0.001) were identified and excluded from the subsequent analyses. Chi-squared (X^2) analysis was performed to identify any differences in allele frequencies between the NHW and Hispanic populations. Pair-wise linkage disequilibrium (LD) values (D’ and r^2) were calculated using GOLD (107).

Linkage and/or association were tested using multiple analytic methods to extract the greatest amount of information from the data. Both parametric and non-parametric linkage analyses were performed using Merlin (108). Pedigree Disequilibrium Test (PDT) was

performed in order to include data pertaining to individuals in larger pedigrees (109). PDT uses all of the informative data in a pedigree, is valid when population substructure is present and remains powerful if there is misclassification of unaffected individuals (109). The genotype-pedigree disequilibrium test (geno-PDT) was also used to include the information from families with multiple affected individuals (110). Geno-PDT tests for patterns of association at the genotypic level (110). Association in the Presence of Linkage (APL) tests for association and 2-SNP haplotypes within a gene and can use all genotype information even when a parental genotype is missing (111). Generalized estimating equations (GEE) was used to detect gene-gene and gene-environment interactions (112).

Log-linear regression models were used to evaluate the independent effects of maternal and child genotypes. Information on maternal, paternal and child genotypes were coded and analyzed using LEM software (113). To prevent any violations of the assumption of independence between each unit of analysis (each triad), only one triad was selected per family (where, child = proband). For each SNP, log-likelihoods were computed for the full models (including both maternal and child genotypes) and compared to the log-likelihoods computed for partial models (including either the maternal genotype or the child genotype only). The resulting two log-likelihood ratios (LLR) were considered to be statistically significant at $p<0.05$.

Chi-square analysis was used to evaluate the relationship between smoking and clubfoot in the presence of null *GSTM1* or *GSTT1* alleles. The relationship between *GSTM1* and *GSTT1* null allele status and smoking/exposure status were analyzed for probands and mothers of probands. Additionally, maternal and paternal null allele status was analyzed to determine if more mothers than fathers of probands possess the null allele genotype.

Protein function analyses and identification of transcription factor binding sites

In silico analyses of significant exonic missense mutations were performed using SNPs3D and Polyphen to estimate the effect of the ancestral and alternate alleles on protein function (114, 115). The ancestral and alternate allele sequences were obtained from the NCBI Entrez SNP Database (www.ncbi.nlm.nih.gov).

In silico analyses of the overtransmitted SNP sequence present in a potential regulatory region were performed using Alibaba2, Patch and Transcription Element Search

Software (TESS) (116-118). The ancestral and alternate allele sequences were obtained from the NCBI Entrez SNP Database (www.ncbi.nlm.nih.gov).

RESULTS

Twenty-two SNPs and two null alleles in eight genes, *CYP1A1, CYP1A2, CYP1B1, CYP2A6, EPHX1, NAT2, GSTM1* and *GSTT1,* were genotyped in our NHW and Hispanic families (Table 11). All SNPs had call rates of >95%. All SNPs were in HWE in the NHW group (Supp. Table 1A). In the Hispanic subset, rs1048943 (*p*=0.02), rs2470893 (*p*=0.0004) and rs4105144 (*p*=0.01) were not in HWE and excluded from further analysis (Supp. Table 1B).

Allele frequencies for 16 of the 22 SNPs differed between the NHW and Hispanic groups (Table 11); therefore the data was stratified by ethnicity. In addition, the data was further stratified by the presence or absence of family history (FH) of clubfoot. The LD plots for each of the 22 SNPs are similar between ethnicities and between unaffected and affected individuals (Supp. Table 2).

In the NHW aggregate group, SNPs in four genes (*CYP1A1* (*p*=0.003), *CYP1A2* (*p*=0.03), *CYP1B1* (*p*=0.05) and *EPHX1* (*p*=0.05)) showed altered transmission (Table 12). rs1048943 in *CYP1A1* (*p*=0.009) also showed altered transmission in the NHW multiplex families as did the rs2234922 in *EPHX1* (*p*=0.05). Two SNPs in *CYP1B1* demonstrated altered transmission in the NHW simplex families as did one SNP in *NAT2*. None of the two-SNP haplotypes showed altered transmission in this group (Supp. Table 3A).

For the Hispanic group, no SNPs showed altered transmission in the aggregate group (Table 12). rs7250713 (*p*=0.01) in *CYP2A6* and rs360063 (*p*=0.04) in *EPHX1* showed altered transmission in the multiplex family subset, whereas rs1456432 (*p*=0.03) in *CYP1A1* and rs360063 (PDT: *p*=0.03; APL: *p*=0.01) in *EPHX1* were altered in the simplex families. None of the two-SNP haplotypes showed altered transmission (Supp. Table 3B).

Strong evidence for a gene interaction was seen only in the NHW group between rs105740 in *EPHX1* and rs1799929 in *NAT2* (*p*=0.007) (Table 13). Suggestive evidence for interactions was found for SNPs in *CYP2A6* and SNPs in *CYP1B1*, *EPHX1* and *NAT2*. There was minimal evidence for gene interactions in the Hispanic group.

Regression modeling was used to evaluate the independent effects of maternal and child genotypes. Only two SNPs showed evidence of a genotypic effect (Table 14). For rs11854147, a significant maternal genotypic effect (*p*= 0.03) was found with a relative risk

Table 10. Smoking metabolism genes: SNP location, alleles and ethnic frequencies

Gene	SNP	Bp #	Alleles[a]	Location	Type[b]	cDNA Change	Protein Change	MAF	HCF[c]
CYP1A1 15q22-24 5.99kb	rs2470893	72806502	G/A	Upstream	-	-	-	0.30	**0.13**
	rs1048943	72800038	A/G	Exon 7	M	1384A>G	Ile462Val	0.04	**0.35**
	rs1456432	72790104	G/A	Downstream	-	-	-	0.16	**0.44**
CYP1A2 15q22-qter 7.76kb	rs2472299	72820453	G/A	Upstream	-	-	-	0.28	0.27
	rs2470890	72834479	C/T	Exon 7	S	1548T>C	Asn516Asn	0.37	**0.68**
	rs11854147	72839824	C/T	Downstream	-	-	-	0.33	**0.61**
CYP1B1 2p22-p21 8.55kb	rs4646429	38160439	A/G	Upstream	-	-	-	0.32	0.31
	rs10012	38155894	G/C	Exon 2	M	142C>G	Arg48Gly	0.32	0.33
	rs1056836	38151707	G/C	Exon 3	M	1294C>G	Leu432Val	0.42	**0.25**
	rs163084	38144420	T/C	Downstream	-	-	-	0.20	**0.14**
CYP2A6 19q13.2 6.90kb	rs4105144	46050464	C/T	Upstream	-	-	-	0.32	**0.25**
	rs7250713	46047035	C/G	Intron 2	-	-	-	0.40	**0.34**
	rs7246742	46037235	G/T	Downstream	-	-	-	0.13	**0.18**
EPHX1 1q42.1 20.29kb	rs2854450	224079200	C/T	Upstream	-	-	-	0.20	0.18
	rs1051740	224086256	T/C	Exon 3	M	337T>C	Tyr113His	0.30	**0.41**
	rs2234922	224093029	A/G	Exon 4	M	416A>G	His139Arg	0.17	**0.08**
	rs360063	224102932	G/A	Downstream	-	-	-	0.44	0.48
NAT2 8p23.1-p21.3 9.97kb	rs1041983	18302075	C/T	Exon 2	S	282C>T	Tyr94Tyr	0.33	0.31
	rs1801280	18302134	T/C	Exon 2	M	341T>C	Ile114Thr	0.44	**0.32**
	rs1799929	18302274	C/T	Exon 2	S	481C>T	Leu161Leu	0.42	**0.32**
	rs1799930	18302383	G/A	Exon 2	M	590G>A	Arg197Gln	0.30	**0.18**
	rs1799931	18302650	G/A	Exon 2	M	857G>A	Gly286Glu	0.04	**0.13**
GSTM1			WT/null	-	null	-	null	0.47	0.56
GSTT1			WT/null	-	null	-	null	0.20	0.13

MAF = minor allele frequency in nonHispanic White sample; HCF = Hispanic corresponding frequency to NHW minor allele

WT = wildtype

[a] Ancestral allele/Alternate allele

[b] Type of Mutation: M = Missense; S = Synonymous

[c] Values in bold = HCF significantly different from MAF [P<0.01]

Table 11. Results of single SNP association analysis by family history

A. NHW

Gene	dbSNP	All			Multiplex			Simplex	
		PDT	GENO-PDT	APL	PDT	GENO-PDT	APL	PDT	GENO-PDT
CYP1A1	rs1048943	**0.003**	**0.003**	0.12	**0.009**	**0.009**	0.71	0.17	0.17
CYP1A2	rs2472299	0.85	0.29	**0.03**	0.61	0.22	0.27	0.21	0.43
CYP1B1	rs1056836	0.53	0.76	0.14	0.92	0.93	0.81	0.13	0.14
CYP1B1	rs163084	0.10	0.14	**0.05**	0.40	0.59	0.40	**0.05**	**0.04**
EPHX1	rs2234922	**0.05**	0.09	0.13	**0.05**	0.06	0.06	0.73	0.68
NAT2	rs1799931	0.12	0.07	0.55	0.59	0.70	0.41	**0.01**	**0.01**

* results for ($p \leq 0.05$)

B. Hispanic

Gene	dbSNP	All			Multiplex			Simplex	
		PDT	GENO-PDT	APL	PDT	GENO-PDT	APL	PDT	GENO-PDT
CYP1A1	rs1456432	0.44	0.68	0.52	0.39	0.70	0.37	**0.03**	0.12
CYP2A6	rs7250713	0.37	0.17	0.74	0.88	0.39	**0.01**	0.25	0.20
EPHX1	rs360063	0.62	0.13	0.28	0.42	0.08	**0.04**	**0.03**	0.15

* results for ($p \leq 0.05$)

of 1.24 (95% CI: 1.04-1.44). A significant fetal genotypic effect (p= 0.01), with a relative risk of 1.33 (95% CI: 1.13-1.54), was found for rs2470890.

The single SNP association analysis detected altered transmission of four nonsynonymous exonic SNPs. To assess whether these SNPs are expected to affect the overall protein stability and function, we utilized SNPs3D (SNPs3D.org) and Polyphen (http://genetics.bwh.harvard.edu/pph/) prediction models. rs1799930 is predicted to either affect the *NAT2* protein stability by altering bond and interaction strength (SNPs3D.org) or to be benign. rs1048943, in *CYP1A1*, is predicted to be a harmless alteration, but is surrounded by SNPs that may affect protein structure stability. rs1056836, in *CYP1B1* is predicted to be deleterious to the protein stability or to be benign while the rs2234922 variant in *EPHX1* is benign.

Table 12. Gene-gene interactions

A. NHW

Gene 1 SNP 1	Gene 2 SNP 2	*P*-Value
EPHX1 rs1051740	*NAT2* rs1799929	0.007
CYP1B1 rs1056836	*CYP2A6* rs4105144	0.02
EPHX1 rs1051740	*NAT2* rs1801280	0.03
EPHX1 rs360063	*CYP2A6* rs4105144	0.04
NAT2 rs1799930	*CYP2A6* rs4105144	0.04
EPHX1 rs2234922	*CYP2A6* rs7246742	0.05
EPHX1 rs1051740	*CYP2A6* rs7250713	0.05

B. Hispanic

Gene 1 SNP 1	Gene 2 SNP 2	*P*-Value
CYP1B1 rs1056836	*NAT2* rs1799929	0.04
CYP1A1 rs1456432	*CYP1A2* rs2472299	0.04
CYP1B1 rs1056836	*NAT2* rs1801280	0.04
CYP1B1 rs163084	*NAT2* rs1799929	0.05
CYP1B1 rs1056836	*CYP1A2* rs2472299	0.05

We also found evidence of association for SNPs that are located 5' upstream of *CYP1A2* and *CYP2A6* in potential regulatory regions. Three transcription factor binding site (TFBS) prediction algorithms were used to assess whether these SNPs could play a role in gene regulation (Table 15). One SNP, rs2472299 in *CYP1A2*, showed evidence for association in the NHW group and marginal evidence of gene-gene interactions in the Hispanics. All three algorithms predicted that the alternate rs2472299 allele eliminates a glucocorticoid receptor (GR) TFBS. Evidence for gene-gene interaction was found for rs4105144 in *CYP2A6* in the NHW group and the alternate allele was predicted to create a TFBS, although the type of TFBS differed (Table 15).

Table 13. Results of log-linear regression modeling of genotypes for mothers and children

Gene	SNP	RR[a] (95% CI)		LRT[b] *p* -value	
		Child	Mother	Child	Mother
CYP1A2	rs2470890	1.33 (1.13-1.54)	1.23 (1.02-1.45)	0.01	0.06
	rs11854147	1.19 (0.99-1.39)	1.24 (1.04-1.44)	0.09	0.03

[a] Relative Risk
[b] Log-likehood ratio test

Table 14. Predicted transcription factor binding sites for 5' associated SNPs

Gene/SNP	Location	Alibaba2 Alleles		Patch Alleles		TESS Alleles	
		Ancestral	Alternate	Ancestral	Alternate	Ancestral	Alternate
CYP1A2 / rs2472299	8.7 kb upstream	GR	None	GR, AR	None	GR, AR	None
CYP2A6 / rs4105144	2.3 kb upstream	None	PU.1	None	None	Bcd, Ft2.2	LEF

GR = Glucocorticoid receptor; AR = Androgen receptor; Bcd = Bicoid; LEF = Lymphoid Enhancer Factors

Because *GSTM1* and *GSTT1* null alleles play a significant role in the variation of xenobiotic metabolism between individuals, we assessed the relationship between null genotypes and the risk for clubfoot. Maternal smoking status was known for 276 NHW probands (yes=65; no=211) and 326 Hispanic probands (yes=15; no=311). Of the individuals genotyped for the *GSTM1* and *GSTT1* alleles, 195 NHW probands had smoking data available (yes=48; no=147) and 251 Hispanic probands had smoking data available (yes=9; no=242). Smoking during pregnancy was reported by 13% of the mothers. Forty-seven percent of individuals genotyped were homozygous for the *GSTM1* null allele and 18% for *GSTT1* null allele, which is consistent with previous reports (88). The data were analyzed in aggregate and by ethnicity. In the NHW group, smoking was evaluated as a covariate for the individual alleles. This analysis could not be performed in the Hispanic subset because of limited information. For the aggregate and NHW groups, there was no evidence for an association between *in utero* exposure to maternal smoking and *GSTM1* or *GSTT1* genotype (Table 16). There was also no relationship between maternal smoking and the mother's genotype and the risk for clubfoot for either gene. In addition, there was no difference in the percentage of null mothers versus null fathers for either gene for either ethnic group.

Table 15. Results for *GSTM1* and *GSTT1* analysis

Dataset	**Genotype**	***GSTM1***		***GSTT1***	
		X^2	***p*-value**	X^2	***p*-value**
All	Proband + Smoking exposure	0.42	0.52	0.05	0.83
All	Maternal + Smoking exposure	0.04	0.84	0.07	0.80
All	Maternal + Paternal	2.12	0.15	1.03	0.31
NHW	Maternal + Paternal	2.15	0.14	0.30	0.58
Hisp	Maternal + Paternal	0.50	0.48	0.60	0.44

DISCUSSION

Numerous studies throughout history have sought to explain the etiology and identify the risk factors for isolated clubfoot. Both genes and environmental factors are speculated to contribute to clubfoot, although the exact roles of each still need to be defined. Candidate gene analysis is beginning to uncover etiologic gene pathways for clubfoot. For example, apoptotic genes, such as *Casp8/10* and *Casp3*, as well as *HOX* genes important for embryologic limb development have recently been reported to be associated with clubfoot (48, 49). While other genes contributing to clubfoot are yet to be discovered, maternal smoking remains the only common environmental risk factor that has consistently been shown to increase the risk of clubfoot (22, 37, 38). In addition, the risk of clubfoot is significantly increased for women who smoke during pregnancy and have a positive family history (37). These findings suggest that genetic variation in smoking metabolism genes may increase susceptibility to clubfoot. Based on this reasoning, we interrogated eight candidate genes, chosen based on their involvement in cigarette smoke metabolism (63, 80, 98). SNPs in six of these genes were chosen based on their frequency, functionality and location within the gene. In addition, two null alleles in *GSTM1* and *GSTT1* were assessed because of their known role in phase II of tobacco metabolism (57, 88).

Considering the strength of the association between clubfoot and smoking from previous population-based studies, there was surprisingly minimal evidence for a role of variation in these eight genes. The strongest evidence for association was for *CYP1A1* (rs1048943; p=0.003) in the NHW dataset in the single SNP analysis. The variant is a missense mutation (1384A>G) in exon 7 that changes an isoleucine to a valine at amino acid position 462 (www.ncbi.nlm.nih.gov/) and confers higher phase I enzymatic activity, which may increase exposure to harmful, adduct forming, metabolic intermediates (75, 76). Another *CYP1A1*-related SNP, rs1456432*,* gave minimal evidence of altered transmission in the Hispanic group. This SNP is located 9.1 kb downstream of *CYP1A1* and therefore, may play a role in regulation of the gene by affecting an enhancer region or stabilization of the mRNA (119). One other significant finding was the interaction between rs1051740 in *EPHX1* and rs1799929 in *NAT2* (p=0.007). *EPHX1*'s major role in phase I of tobacco smoke metabolism is hydrolysis of PAH (80, 86). As previously discussed, the rs1051740 variant

has decreased activity and has been associated with decreased DNA adduct formation (59, 77). *NAT2* plays an important role in phase II reactions of tobacco smoke metabolism (98). The rs1799929 variant allele in *NAT2* codes for a synonymous amino acid change in exon 2, which could decrease the rate of *NAT2* translation (120). Additionally, *NAT2* is known to have variants with decreased activity, though the *in vivo* effect of the rs1799929 on enzyme activity is not known and may or may not confer a "slow acetylator" phenotype (100).

While SNP associations do not imply causation, they can indicate that variation in, near and/or in linkage disequilibrium with these SNPs may be causative. There are many ways SNPs can cause or can be related to SNPs that cause the observed associations. One interpretation of our results is that perturbation of phase I reactions of xenobiotic metabolism that increase enzyme activity and adduct formation may play a role in the development of clubfoot. Compounds from cigarette smoke form DNA and protein adducts, which can cause mutagenesis and teratogenesis (52, 56, 59). Additionally, simultaneous increase in phase I activity with a perturbation in phase II activity may also play a role. Because phase I xenobiotic metabolism genes create harmful metabolic intermediates in the normal biotransformation pathway, it may be that an increase in phase I enzyme activity and a decrease in phase II degradation of these intermediates increases the concentration of harmful compounds to damaging levels that can interfere with fetal development. Some of the metabolic intermediates known to be produced by phase I xenobiotic metabolism are active oxygen species and DNA or protein binding adducts (50). While it has been suggested that harmful oxygen species and adducts can interfere with normal fetal development, the role of these compounds in the pathogenesis of clubfoot is unknown (57, 121, 122). Our results suggest that an increase in harmful metabolic intermediates could contribute to abnormal foot development or rotation of the foot.

Our results also identified other marginal associations and gene interactions that could potentially be important. Many of these SNP variants are known or predicted to alter the function of the gene and may impact the efficiency of the xenobiotic metabolism pathway by perturbing the activity of phase I and/or phase II. While these associations and interactions are marginal, they suggest that many gene variants and/or interactions could be important and may perturb different parts of the xenobiotic metabolism pathway, and

thereby contribute to clubfoot causality. Additional studies are needed to further explore these potential interactions.

One such association involves SNPs in *CYP1A2* (rs2472299, rs1056836) and rs163084 in *CYP1B1,* which had marginal evidence for association in the NHW dataset only. The rs2472299 polymorphism is 8.7 kb upstream of the *CYP1A2* gene and is predicted to abolish a glucocorticoid receptor or androgen receptor transcription factor-binding site. These results suggest that the variant may alter transcription of *CYP1A2* and decrease the efficiency of the metabolic pathway (116-118). Previous studies that have considered the effects of SNPs in *CYP1B1* and have suggested that rs1056836 alone may alter *CYP1B1* function by decreasing metabolism of benzo(a)pyrene (BaP) metabolites (BaP-7,8-dihydrodiols) and increasing metabolism of the parent compound (BaP), which would increase DNA adduct formation (59, 77). *CYP1A2* and *CYP1B1* both play a role in phase I xenobiotic metabolism, and one possible effect of increased phase I enzymatic activity is a build up of intermediate metabolites. Therefore, harmful metabolic intermediates may have a greater impact due to a more substantial "exposure" (59). For example, BaP, a prototypic PAH, is first metabolized to an epoxide by *CYP1A1* and *CYP1B1* (60, 61) (Figure 17). The end-product of phase I metabolism of BaP is a highly reactive diol-epoxide, which can form DNA adducts (Nock, et al., 2007).

Additionally, an intronic SNP in *CYP2A6*, rs7250713, had altered transmission in Hispanic multiplex families. Intronic SNPs have been associated with a dysregulation of splice-variant expression and have been shown to produce truncated/non-functional protein, which could alter pathways or gene activity and play a role in disease pathogenesis (123, 124). However, because SNPs are markers for susceptibility and rs7250713 is not near an exonic-intronic junction, our results most likely indicate that variation in *CYP2A6* or nearby genes may play a role in the etiology of clubfoot.

rs1799931 in *NAT2* also had marginal evidence of association in the single SNP analyses only in NHW simplex cases. This SNP encodes a missense mutation that changes a glycine to a glutamine in the mature protein (www.ncbi.nlm.nih.gov/snp/) and is predicted to affect protein stability (SNPs3D.org). The activity of this SNP *in vivo* is unknown and the activity *in vitro* is controversial, showing decreased activity in Chinese hamster ovary (CHO) cells and unstable protein formation with maximum enzymatic velocity not significantly different

from wildtype in E. coli (100, 125, 126). Our result supports the finding by Hecht, et al., 2007, which found suggestive evidence for an association with rs1799931 in both NHW and Hispanic simplex families. However, in that study a more significant association was found for the Hispanic simplex families suggesting that this variant may be a greater risk factor in the Hispanic population. The current results suggest that variation in *NAT2* may also be an important risk factor for the NHW population. A larger Hispanic dataset is needed to determine whether this is indeed a risk factor in that group.

In addition to the significant interaction between *EPHX1* and *NAT2*, a marginal interaction was found between rs1051740 in *EPHX1* and rs1801280 in *NAT2*, which causes a “slow acetylator” phenotype (100, 102, 103). Interestingly, we observed a marginal interaction between rs1056836 in *CYP1B1* and the same *NAT2* SNPs seen in the interactions with *EPHX1*, rs1799929 and rs1801280. *EPHX1* and *CYP1B1* interact through sequential reactions in phase I metabolism of PAH and *NAT2* is a well-characterized phase II enzyme with variants that have been shown to cause an increase in harmful adduct levels (59, 100, 101). These results also support the findings by Hecht, et al., 2007 and provide additional evidence that *NAT2* may play a role in clubfoot. In addition, these results support the conclusions that perturbation of phase I and phase II enzymatic activity may lead to an increased risk of clubfoot. Other marginal but potentially interesting interactions were found for different phase I xenobiotic metabolism genes. These interactions involved *CYP1B1* with *CYP1A1*, *CYP1A2* and *CYP2A6*, and multiple SNPs in *EPHX1* and *CYP2A6*. These results are intriguing when the functional effect of the SNPs and the possible effect on the metabolic pathway are considered (Table 17). In general, we found interactions between polymorphisms with variants that are known to alter the activity of the phase I enzyme or could affect regulation of the gene. Although the these findings need to be further explored, collectively these results support the hypothesis that disruption of normal activity in both phase I and II of xenobotic metabolism can increase the concentration of harmful intermediates and may play a role in clubfoot.

GSTM1 and *GSTT1* also play a role in phase II of tobacco smoke metabolism and the null alleles of both genes cause absence of enzymatic activity (57). Loss of activity causes an increased susceptibility to DNA adducts and could increase harmful metabolic intermediates (88, 92-94, 97). Interestingly, our results provide no support for a relationship

between a homozygous null allele at either locus and an increased risk of clubfoot, even in the presence of smoking. However, these results must be carefully interpreted because heterozygotes cannot be discriminated from homozygous wildtype individuals. The GST null alleles likely act as a recessive system in which both null alleles must be present (88, 97). Based on this assumption, the phenotype would be anticipated only in the homozygous null individuals who have little or no residual enzymatic activity. Therefore, our results suggest that *GSTM1* and *GSTT1* null alleles, whether they are present in the fetus or the mother, are not associated with an increased risk of clubfoot when a mother smokes during pregnancy.

Because both the maternal and fetal smoking metabolism genes may affect the risk for clubfoot, we evaluated whether each SNP possessed an independent maternal genetic effect (maternal genotype confers a risk) or an independent inherited genetic effect (fetal genotype confers a risk). The maternal effects calculation allows us to compare the role of the maternal and/or the fetal genotype on the risk for clubfoot and to obtain relative risks for each genotype (113). For *CYP1A2*, we found a significant deleterious effect for rs11854147 (p=0.03; RR=1.24; 95% CI: 1.04-1.44) in the mother and for rs2470890 (p=0.01; RR=1.33; 95% CI: 1.13-1.54) when in the fetus. This suggests that either rs11854147 or rs2470890 can increase the risk of clubfoot but the mechanisms differ based on either whether it is a maternal or fetal genotype. Differences in the gene expression, induction and enzyme activity between maternal and fetal tobacco metabolism genes have been reported and these results would support evidence that variants in tobacco metabolism genes have different consequences when they are of maternal and/or fetal origin (57, 62, 71). Differentiating risk based on maternal and child genotypes as well as interactions between the two genotypes will be important for elucidating the role of smoking in the etiology of clubfoot. Future studies need to evaluate maternal and fetal combinations to determine genotypic effects and to whether the information can be used to predict pregnancy outcomes.

Interestingly, although an association between xenobiotic metabolism genes and clubfoot was found, we were unable to confirm the association between maternal smoking and clubfoot reported in other studies (22, 37-39). This may partly be explained by the fact that only 13% of women reported smoking during pregnancy in this dataset. In epidemiological studies that have found an association between maternal smoking and

clubfoot, an average of 30% of mothers reported smoking during their pregnancy (37-39). Additionally, our dataset consists of a majority of Hispanic probands (54%) while epidemiological studies looked predominantly at NHW probands (average of 81% NHW). Therefore, there are clear differences between our dataset and the epidemiological datasets, which may explain why we did not observe more significant results. Because smoking does not appear to be as great a risk factor in our population, we would not expect to see strong associations between xenobiotic metabolism genes and clubfoot. Evaluation of the role of xenobiotic metabolism genes in a dataset for which smoking is found to increase the risk of clubfoot may reveal stronger associations between these genes and clubfoot and should be explored in future studies.

Alternatively, our results could suggest that xenobiotic metabolism genes play a role in clubfoot that is independent of smoking because these genes are known to have other roles in metabolizing compounds not found in cigarette smoke, such as prostaglandins and other endogenous hormones (127). For example, *CYP1A1*, *1A2* and *1B1* play a role in the metabolism of estradiols, which are important in maintaining a pregnancy and have been shown to increase blood flow in coronary arteries (127, 128). Estradiols may also increase blood flow in other organs, such as the placenta, and therefore, increased metabolism of these hormones may contribute to a decreased availability of oxygen and nutrients to a developing fetus. Additionally, *CYP2A6* metabolizes approximately 1% of over-the-counter and prescription drugs as well as many environmental toxins (127). Therefore, xenobiotic metabolism genes may impact the risk of clubfoot by altering the metabolism of multiple exogenous and endogenous compounds besides tobacco smoke.

If tobacco smoke exposure is not affecting the fetus through the xenobiotic metabolism pathway, it may play a role in the etiology of clubfoot by other mechanisms. For example, early amniocentesis also increases the risk of clubfoot suggesting that a common teratogenic mechanism may be common to both exposures. The simplest explanation would be vascular insufficiency and hypoxia that would deprive the fetus of blood flow and necessary nutrients. Maternal smoking and nicotine exposure in mice specifically reduces blood flow and increases vascular resistance in the uterus (129-132). Moreover, mice exposed to PAHs have been shown to have abnormal vasculature in the placenta that significantly reduces arterial surface area and volume of the fetal arterial

Table 16. Functional effects of significant smoking metabolsim gene interactions*

A. Interactions between phase I genes

Gene 1	SNP 1	Functional effect	Gene 2	SNP 2	Functional effect	*p*-Value	Pop.
CYP1B1	rs1056836	Variant has increased metabolism of BaP and decreased metabolism of BaP-7,8-dihydrodiols; increased risk for adducts[a, b]	*CYP2A6*	rs4105144	5' of gene; may affect TFBS[d]	0.02	NHW
EPHX1	rs360063	3' of gene; possible enhancer region	*CYP2A6*	rs4105144	5' of gene; may affect TFBS[d]	0.04	NHW
CYP1A1	rs1456432	3' of gene; possible enhancer region	*CYP1A2*	rs2472299	5' of gene; may affect TFBS[d]	0.04	H
EPHX1	rs2234922	Variant has decreased activity and decreased adducts [a]	*CYP2A6*	rs7246742	3' of gene; possible enhancer region	0.05	NHW
EPHX1	rs1051740	Variant has decreased activity and decreased adducts [a]	*CYP2A6*	rs7250713	Intronic	0.05	NHW
CYP1B1	rs1056836	Variant has increased metabolism of BaP and decreased metabolism of BaP-7,8-dihydrodiols; increased risk for adducts[a, b]	*CYP1A2*	rs2472299	5' of gene; may affect TFBS[d]	0.05	H

Table 16. Functional effects of significant smoking metabolsim gene interactions* (cont.)

B. Interactions between phase I and phase II genes

Gene 1	SNP 1	Functional effect	Gene 2	SNP 2	Functional effect	*p*-Value	Pop.
EPHX1	rs1051740	Variant has decreased activity and decreased adducts [a]	*NAT2*	rs1799929	Activity unknown *in vivo* [c]	0.007	NHW
EPHX1	rs1051740	Variant has decreased activity and decreased adducts [a]	*NAT2*	rs1801280	Decreased activity = "slow acetylator" [c]	0.03	NHW
NAT2	rs1799930	Decreased activity *in vitro* [c]	*CYP2A6*	rs4105144	5' of gene; may affect TFBS[d]	0.04	NHW
CYP1B1	rs1056836	Variant has increased metabolism of BaP and decreased metabolism of BaP-7,8-dihydrodiols; increased risk for adducts[a, b]	*NAT2*	rs1799929	Activity unknown *in vivo* [c]	0.04	H
CYP1B1	rs1056836	Variant has increased metabolism of BaP and decreased metabolism of BaP-7,8-dihydrodiols; increased risk for adducts[a, b]	*NAT2*	rs1801280	Decreased activity = "slow acetylator" [c]	0.04	H
CYP1B1	rs163084	3' of gene; possible enhancer region	*NAT2*	rs1799929	Activity unknown *in vivo* [c]	0.05	H

*published literature listed in references

[a] Nock et al., 2007; [b] Georgiadis, et al., 2004; [c] Meyer & Zanger, 1997; [d] AliBaba2, Patch, TESS

Variant = alternate allele

TFBS = Transcription factor binding site

vasculature (56). Reduction in vascular efficiency and an increased susceptibility to hypoxia could cause abnormal limb development and has been shown to cause transverse limb defects in prolonged (30-60 minutes) cases of anoxia (139). Therefore, variable degrees of hypoxia may increase the risk of other limb abnormalities, the most mild of which would be clubfoot. Interestingly, aborted fetuses with homozygous alpha-thalassemia, a genetic condition known to cause severe fetal anemia and hypoxia, experience transverse limb defects with the feet more severely affected than the hands (133). These effects may be specific to the developing foot depending on the timing of the hypoxic event, the intricate and complicated nature of the developing foot and/or a strategy for a fetus to conserve nutrients and oxygen to more essential body parts and organs.

The targeted effect of both maternal smoking and early amniocentesis is possible reduction in fetal movements that are necessary for joint, vascular and soft tissue development (134, 135). Decreased fetal movement leads to or contributes to joint and limb anomalies in mouse models and humans (134, 136). Therefore, the association between maternal smoking and clubfoot might be explained by the direct effect of tobacco on *in utero* fetal lower limb movement. Although there is no clear relationship between maternal tobacco smoking and early amniocentesis, they both confer an increased risk for clubfoot. It is possible that they share a common mechanism that is responsible for an increased risk of clubfoot, although the exact mechanism(s) requires additional study.

The results of this study are important but must be carefully interpreted until larger validation studies can be undertaken. As previously discussed, smoking data was reported by 13% (n=80) of our mothers and is consistent with that reported by pregnant women in the general population (105). However, this translates into a relatively small number of smoking mothers upon which to base our analyses and limits the ability to detect an effect that is not large. This is particularly striking for the GST null allele analyses (137, 138). Additionally, this study focused on only eight of the most important tobacco smoke metabolism enzymes. There are likely hundreds of genes involved in the overall biotransformation of the compounds found in tobacco smoke, which were not considered in this study. While the genes in this study are the most likely candidates, we cannot rule out that other genes play a major role or interact with the genes in this study and contribute to the clubfoot phenotype.

Previous epidemiologic and molecular research has shown that maternal smoking is associated with an increased risk of clubfoot and variation in smoking metabolism genes can cause an increase in harmful metabolic intermediates that may lead to a variety of birth defects (37, 38, 100, 121, 122). To better understand this relationship and to identify the possible role of smoking metabolism in the etiology of clubfoot, we examined the relationship between polymorphic variants in smoking metabolism genes and clubfoot in a dataset of well characterized multiplex and simplex clubfoot families for which prenatal exposure to cigarette smoke was known. Interestingly, no association was found between maternal smoking and clubfoot. This may indicate that the genes in this study play a role that impacts the developing fetus regardless of cigarette smoke exposure or they play a larger role in the development of clubfoot for populations in which smoking is found to be a risk factor. However, our results suggest that there is an association between *CYP1A1* and an interaction between *EPHX1* and *NAT2* xenobiotic metabolism genes and an increased risk for clubfoot. Additionally, we observed multiple marginal associations and gene interactions in other xenobiotic metabolism genes, suggesting that these genes may also play a role in the etiology of clubfoot. The genes in this study are likely to interact in pathways that affect fetal development, possibly by altering blood flow to the developing fetus, causing changes in hormonal metabolism, inducing fetal hypokinesia or some other mechanism. Further studies are needed to better delineate the role of xenobiotic metabolism genes during pregnancy and the effects of polymorphisms on the developing fetus.

Appendix A

Supplementary Tables

Supplementary Table 1. Hardy-Weinberg equilibrium results

A. NHW

Gene	SNP	No.	PIC	Het.	X^2	*p*-value
CYP1A1	rs1048943	439	0.08	0.08	0.02	0.88
	rs1456432	434	0.23	0.29	2.32	0.13
	rs2470893	436	0.33	0.43	0.14	0.71
CYP1A2	rs1185414	430	0.34	0.44	0.01	0.92
	rs2470890	433	0.36	0.48	0.19	0.66
	rs2472299	429	0.32	0.42	0.41	0.52
CYP1B1	rs10012	428	0.34	0.42	0.22	0.64
	rs1056836	439	0.37	0.53	2.80	0.09
	rs163084	441	0.27	0.34	0.97	0.32
	rs4646429	435	0.34	0.43	0.01	0.91
CYP2A6	rs4105144	423	0.34	0.41	1.56	0.21
	rs7246742	443	0.20	0.21	0.59	0.44
	rs7250713	440	0.37	0.45	2.15	0.14
EPHX1	rs1051740	430	0.33	0.44	1.31	0.25
	rs2234922	438	0.25	0.28	0.36	0.55
	rs2854450	439	0.27	0.31	0.08	0.77
	rs360063	436	0.37	0.48	0.14	0.71
NAT2	rs1041983	433	0.35	0.44	0.01	0.92
	rs1799929	429	0.37	0.48	0.02	0.88
	rs1799930	427	0.33	0.41	0.34	0.56
	rs1799931	429	0.07	0.07	0.37	0.54
	rs1801280	432	0.37	0.50	0.13	0.72
GSTM1	-	392	0.37	0.00	392.00	<.0001
GSTT1	-	393	0.27	0.00	393.00	<.0001

Supplementary Table 1. Hardy-Weinberg equilibrium results (cont.)

B. Hispanic

Gene	Locus	No.	PIC	Het.	X^2	p-value*
CYP1A1	rs1048943	416	0.35	0.40	5.19	**0.02**
	rs1456432	416	0.37	0.50	0.10	0.75
	rs2470893	415	0.20	0.19	12.55	**0.00**
CYP1A2	rs1185414	422	0.36	0.44	2.33	0.13
	rs2470890	414	0.34	0.41	1.07	0.30
	rs2472299	419	0.32	0.42	0.76	0.38
CYP1B1	rs10012	408	0.34	0.44	0.02	0.88
	rs1056836	415	0.30	0.37	0.00	0.96
	rs163084	421	0.21	0.25	0.38	0.54
	rs4646429	418	0.33	0.43	0.04	0.84
CYP2A6	rs4105144	402	0.31	0.33	6.13	**0.01**
	rs7246742	419	0.26	0.31	0.14	0.71
	rs7250713	409	0.35	0.42	1.12	0.29
EPHX1	rs1051740	418	0.37	0.46	0.97	0.32
	rs2234922	421	0.13	0.15	0.99	0.32
	rs2854450	416	0.25	0.30	0.38	0.54
	rs360063	421	0.37	0.51	0.14	0.71
NAT2	rs1041983	412	0.33	0.44	0.35	0.56
	rs1799929	406	0.34	0.44	0.10	0.76
	rs1799930	405	0.26	0.32	1.50	0.22
	rs1799931	409	0.20	0.22	0.38	0.54
	rs1801280	409	0.34	0.45	0.55	0.46
GSTM1	-	399	0.37	0.00	399.00	<.0001
GSTT1	-	399	0.20	0.00	399.00	<.0001

No. = number of individuals

PIC = Polymorphic information content

*Bolded values are not in HWE and were excluded from our analyses

Supplementary Table 2. Linkage disequilibrium (D') for SNPs in smoking metabolism genes by chromosome [a,b]

Chromosome 1 (*EPHX1*)

	rs1051740	rs2234922	rs2854450	rs360063
rs1051740		0.083	0.191	0.594
rs2234922	0.073		0.076	1.000
rs2854450	0.422	0.014		0.349
rs360063	0.539	0.951	0.238	

Chromosome 2 (*CYP1B1*)

	rs163084	rs1056836	rs10012	rs4646429
rs163084		0.942	1.000	1.000
rs1056836	0.949		0.982	0.982
rs10012	1.000	0.987		1.000
rs4646429	1.000	0.987	1.000	

Chromosome 8 (*NAT2*)

	rs1041983	rs1801280	rs1799929	rs1799930	rs1799931
rs1041983		0.984	0.983	0.971	1.000
rs1801280	0.979		1.000	0.960	1.000
rs1799929	0.967	0.995		0.957	1.000
rs1799930	0.962	0.987	0.987		1.000
rs1799931	1.000	1.000	1.000	1.000	

Chromosome 15 *(CYP1A1, CYP1A2)*

	rs1456432	rs1048943	rs2470893	rs2472299	rs2470890	rs11854147
rs1456432		1.000	1.000	0.634	0.259	0.211
rs1048943	1.000		0.999	0.252	0.640	0.619
rs2470893	1.000	1.000		0.842	0.870	0.889
rs2472299	0.721	0.272	0.861		0.989	0.981
rs2470890	0.285	0.694	0.904	1.000		0.901
rs11854147	0.281	0.508	0.887	0.987	0.922	

Chromosome 19 (*CYP2A6*)

	rs7246742	rs7250713	rs4105144
rs7246742		0.190	1.000
rs7250713	0.117		0.989
rs4105144	0.815	0.987	

Supplementary Table 2. Linkage disequilibrium (D') for SNPs in smoking metabolism genes by chromosome [a,b] (cont.)

B. Hispanic

Chromosome 1 (*EPHX1*)

	rs1051740	rs2234922	rs2854450	rs360063
rs1051740		0.559	0.769	0.120
rs2234922	0.418		0.109	1.000
rs2854450	0.585	0.227		0.311
rs360063	0.123	1.000	0.503	

Chromosome 2 (*CYP1B1*)

	rs163084	rs1056836	rs10012	rs4646429
rs163084		1.000	1.000	1.000
rs1056836	0.948		0.564	0.842
rs10012	0.926	0.694		0.982
rs4646429	1.000	1.000	0.994	

Chromosome 8 (*NAT2*)

	rs1041983	rs1801280	rs1799929	rs1799930	rs1799931
rs1041983		0.973	0.973	0.951	0.975
rs1801280	1.000		0.991	1.000	1.000
rs1799929	0.979	0.988		1.000	1.000
rs1799930	0.906	0.898	0.894		0.777
rs1799931	1.000	1.000	0.941	1.000	

Chromosome 15 *(CYP1A1, CYP1A2)*

	rs1456432	rs1048943	rs2470893	rs2472299	rs2470890	rs11854147
rs1456432		1.000	1.000	0.873	0.711	0.405
rs1048943	1.000		1.000	0.944	0.881	0.839
rs2470893	1.000	1.000		0.804	0.956	0.946
rs2472299	0.906	0.951	0.833		1.000	1.000
rs2470890	0.719	0.861	0.914	1.000		0.980
rs11854147	0.510	0.848	0.899	1.000	0.987	

Chromosome 19 (*CYP2A6*)

	rs7246742	rs7250713	rs4105144
rs7246742		0.069	0.746
rs7250713	0.065		0.989
rs4105144	0.807	0.974	

[a] D': 0.3-0.6 = Light gray; 0.6-0.8 = Gray; 0.8-1.0 = Dark gray

[b] D' of affected individuals above diagonal line; D' of unaffecteds below gray line

Supplementary Table 3. Two-SNP haplotypes by ethnicity

A. NHW population

Gene	SNPs	Haplotype	Freq.	*p*-value
CYP1A1	rs1456432 rs1048943	11	0.839	0.437
		21	0.113	
		22	0.048	
	rs1456432 rs2470893	11	0.543	0.792
		12	0.303	
		21	0.153	
		22	0.001	
	rs1048943 rs2470893	11	0.649	0.658
		12	0.306	
		21	0.046	
		22	0.000	
CYP1A2	rs2472299 rs2470890	11	0.002	0.106
		12	0.285	
		21	0.632	
		22	0.081	
	rs2472299 rs11854147	11	0.004	0.089
		12	0.283	
		21	0.661	
		22	0.051	
	rs2470890 rs11854147	11	0.608	0.348
		12	0.023	
		21	0.055	
		22	0.314	
CYP1B1	rs163084 rs1056836	11	0.197	0.279
		12	0.006	
		21	0.225	
		22	0.572	
	rs163084 rs10012	11	0.198	0.179
		12	0.006	
		21	0.484	
		22	0.312	

B. Hispanic population

Gene	SNPs	Haplotype	Freq.	*p*-value
CYP1A1	rs1456432 rs1048943	11	0.570	0.124
		12	0.002	
		21	0.098	
		22	0.330	
	rs1456432 rs2470893	11	0.438	0.609
		12	0.129	
		21	0.433	
	rs1048943 rs2470893	11	0.535	0.416
		12	0.130	
		21	0.333	
		22	0.001	
CYP1A2	rs2472299 rs2470890	11	0.004	0.761
		12	0.280	
		21	0.313	
		22	0.403	
	rs2472299 rs11854147	11	0.001	0.694
		12	0.281	
		21	0.387	
		22	0.331	
	rs2470890 rs11854147	11	0.308	0.434
		12	0.006	
		21	0.077	
		22	0.609	
CYP1B1	rs163084 rs1056836	11	0.119	0.342
		12	0.006	
		21	0.120	
		22	0.754	
	rs163084 rs10012	11	0.122	0.366
		12	0.004	
		21	0.548	
		22	0.326	

Supplementary Table 3. Two-SNP haplotypes by ethnicity (cont.)

A. NHW population

Gene	SNPs	Haplotype	Freq.	*p*-value
CYP1B1	rs163084 rs4646429	11	0.202	0.295
		12	0.005	
		21	0.479	
		22	0.315	
	rs1056836 rs10012	11	0.405	0.273
		12	0.008	
		21	0.275	
		22	0.312	
	rs1056836 rs4646429	11	0.410	0.260
		12	0.007	
		21	0.268	
		22	0.315	
	rs10012 rs4646429	11	0.677	0.794
		12	0.002	
		21	0.002	
		22	0.319	
CYP2A6	rs7246742 rs7250713	11	0.245	0.344
		12	0.045	
		21	0.314	
		22	0.125	
	rs7246742 rs4105144	11	0.249	0.833
		12	0.040	
		21	0.561	
		22	0.149	
	rs7250713 rs4105144	11	0.681	0.337
		12	0.147	
		21	0.128	
		22	0.044	
EPHX1	rs1051740 rs2234922	11	0.245	0.368
		12	0.045	
		21	0.585	
		22	0.125	
	rs1051740 rs2854450	11	0.249	0.824
		12	0.040	
		21	0.561	
		22	0.149	

B. Hispanic population

Gene	SNPs	Haplotype	Freq.	*p*-value
CYP1B1	rs163084 rs4646429	11	0.122	0.497
		12	0.002	
		21	0.576	
		22	0.300	
	rs1056836 rs10012	11	0.211	0.945
		12	0.030	
		21	0.464	
		22	0.295	
	rs1056836 rs4646429	11	0.229	0.973
		12	0.009	
		21	0.472	
		22	0.291	
	rs10012 rs4646429	11	0.667	0.458
		12	0.002	
		21	0.029	
		22	0.301	
CYP2A6	rs7246742 rs7250713	11	0.123	0.671
		12	0.070	
		21	0.533	
		22	0.274	
	rs7246742 rs4105144	11	0.175	0.823
		12	0.013	
		21	0.576	
		22	0.236	
	rs7250713 rs4105144	11	0.653	0.608
		12	0.004	
		21	0.097	
		22	0.246	
EPHX1	rs1051740 rs2234922	11	0.380	0.118
		12	0.017	
		21	0.542	
		22	0.061	
	rs1051740 rs2854450	11	0.360	0.501
		12	0.036	
		21	0.451	
		22	0.153	

Supplementary Table 3. Two-SNP haplotypes by ethnicity (cont.)

A. NHW population

Gene	SNPs	Haplotype	Freq.	*p*-value
EPHX1	rs1051740 rs360063	11	0.057	0.993
		12	0.233	
		21	0.365	
		22	0.345	
	rs2234922 rs2854450	11	0.681	0.345
		12	0.147	
		21	0.128	
		22	0.044	
	rs2234922 rs360063	11	0.418	0.313
		12	0.409	
		21	0.003	
		22	0.171	
	rs2854450 rs360063	11	0.319	0.965
		12	0.490	
		21	0.106	
		22	0.085	
NAT2	rs1041983 rs1801280	11	0.423	0.863
		12	0.240	
		21	0.002	
		22	0.335	
	rs1041983 rs1799929	11	0.249	0.516
		12	0.410	
		21	0.337	
		22	0.004	
	rs1041983 rs1799930	11	0.008	0.964
		12	0.658	
		21	0.287	
		22	0.048	
	rs1041983 rs1799931	12	0.656	0.917
		21	0.039	
		22	0.305	
	rs1801280 rs1799929	11	0.015	0.298
		12	0.411	
		21	0.572	
		22	0.002	

B. Hispanic population

Gene	SNPs	Haplotype	Freq.	*p*-value
EPHX1	rs1051740 rs360063	12	0.227	0.380
		22	0.290	
		11	0.169	
		21	0.314	
	rs2234922 rs2854450	11	0.766	0.146
		12	0.157	
		21	0.047	
		22	0.031	
	rs2234922 rs360063	11	0.478	0.350
		12	0.441	
		21	0.001	
		22	0.079	
	rs2854450 rs360063	11	0.361	0.307
		12	0.449	
		21	0.116	
		22	0.073	
NAT2	rs1041983 rs1801280	11	0.302	0.384
		12	0.378	
		21	0.006	
		22	0.313	
	rs1041983 rs1799929	11	0.391	0.601
		12	0.292	
		21	0.310	
		22	0.007	
	rs1041983 rs1799930	11	0.009	0.861
		12	0.678	
		21	0.169	
		22	0.143	
	rs1041983 rs1799931	11	0.004	0.744
		12	0.683	
		21	0.125	
		22	0.188	
	rs1801280 rs1799929	11	0.009	0.608
		12	0.298	
		21	0.691	
		22	0.002	

Supplementary Table 3. Two-SNP haplotypes by ethnicity (cont.)

A. NHW population

Gene	SNPs	Haplotype	Freq.	*p*-value
NAT2	rs1801280 rs1799930	11	0.002	0.605
		12	0.421	
		21	0.291	
		22	0.285	
	rs1801280 rs1799931	12	0.425	0.737
		21	0.038	
		22	0.537	
	rs1799929 rs1799930	11	0.296	0.284
		12	0.292	
		21	0.002	
		22	0.410	
	rs1799929 rs1799931	11	0.037	0.403
		12	0.550	
		22	0.413	
	rs1799930 rs1799931	12	0.298	0.880
		21	0.038	
		22	0.664	

B. Hispanic population

Gene	SNPs	Haplotype	Freq.	*p*-value
NAT2	rs1801280 rs1799930	11	0.010	0.385
		12	0.302	
		21	0.174	
		22	0.514	
	rs1801280 rs1799931	12	0.311	0.489
		21	0.130	
		22	0.560	
	rs1799929 rs1799930	11	0.174	0.185
		12	0.523	
		21	0.010	
		22	0.292	
	rs1799929 rs1799931	11	0.126	0.524
		12	0.572	
		21	0.001	
		22	0.301	
	rs1799930 rs1799931	11	0.003	0.968
		12	0.179	
		21	0.128	
		22	0.690	

Supplementary Table 4. Results of all single SNP association analyses by family history

A. NHW

Gene	dbSNP	All			Multiplex			Simplex		
		PDT	GENO-PDT	APL	PDT	GENO-PDT	APL	PDT	GENO-PDT	APL
CYP1A1	rs1048943	0.003	0.003	0.12	0.009	0.009	0.71	0.17	0.17	0.06
	rs1456432	0.33	0.24	0.67	0.36	0.35	0.86	0.73	0.68	0.33
	rs2470893	0.45	0.60	0.55	0.45	0.74	0.71	0.82	0.61	0.54
CYP1A2	rs11854147	0.86	0.44	0.07	0.48	0.32	0.28	0.40	0.71	0.18
	rs2470890	0.68	0.93	0.06	0.94	0.99	0.16	0.38	0.70	0.21
	rs2472299	0.85	0.29	0.03	0.61	0.22	0.27	0.21	0.43	0.07
CYP1B1	rs10012	0.36	0.12	0.37	0.30	0.11	0.38	1.00	0.94	0.69
	rs1056836	0.53	0.76	0.14	0.92	0.93	0.81	0.13	0.14	0.05
	rs163084	0.10	0.14	0.05	0.40	0.59	0.40	0.05	0.04	0.05
	rs4646429	0.36	0.18	0.40	0.32	0.15	0.46	0.91	0.97	0.63
CYP2A6	rs4105144	0.54	0.63	0.62	0.43	0.54	0.23	0.90	0.98	0.49
	rs7246742	0.47	0.55	0.37	0.51	0.36	0.51	0.75	0.56	0.44
	rs7250713	0.13	0.27	0.67	0.16	0.28	0.94	0.54	0.84	0.58
EPHX1	rs1051740	0.60	0.75	0.66	0.16	0.39	0.40	0.22	0.47	0.85
	rs2234922	0.05	0.09	0.13	0.05	0.06	0.06	0.73	0.68	0.93
	rs2854450	0.91	0.64	0.75	0.94	0.68	0.58	0.92	0.85	0.75
	rs360063	0.85	0.98	0.85	0.89	0.97	0.82	0.91	0.98	0.53
NAT2	rs1041983	0.52	0.41	0.91	0.54	0.42	0.52	0.83	0.91	0.80
	rs1799929	0.82	0.91	0.29	0.73	0.88	0.59	0.33	0.66	0.28
	rs1799930	0.41	0.49	0.90	0.71	0.56	0.18	0.35	0.66	0.40
	rs1799931	0.12	0.07	0.55	0.59	0.70	0.41	0.01	0.01	0.04
	rs1801280	0.79	0.97	0.48	0.86	0.98	0.78	0.38	0.70	0.42

Supplementary Table 4. Results of all single SNP association analyses by family history (cont.)

B. Hispanic

Gene	dbSNP	All			Multiplex			Simplex		
		PDT	GENO-PDT	APL	PDT	GENO-PDT	APL	PDT	GENO-PDT	APL
CYP1A1	rs1048943	0.52	0.77	0.17	0.12	0.32	0.07	0.16	0.43	0.68
	rs1456432	0.44	0.68	0.52	0.39	0.70	0.37	0.03	0.12	0.21
	rs2470893	0.05	0.15	0.61	0.14	0.39	0.14	0.19	0.36	0.88
CYP1A2	rs11854147	0.39	0.33	0.60	0.11	0.29	0.32	0.52	0.53	0.97
	rs2470890	0.66	0.12	0.51	0.50	0.22	0.35	0.14	0.29	0.16
	rs2472299	0.42	0.67	0.93	0.73	0.82	0.96	0.14	0.31	0.89
CYP1B1	rs10012	0.38	0.65	0.88	0.45	0.58	0.41	0.65	0.88	0.78
	rs1056836	0.65	0.53	0.94	0.84	0.62	0.95	0.65	0.81	0.94
	rs163084	0.23	0.34	0.84	0.29	0.36	0.72	0.55	0.79	1.00
	rs4646429	0.46	0.52	0.80	0.83	0.75	0.92	0.30	0.59	0.86
CYP2A6	rs4105144	0.56	0.85	0.97	0.58	0.82	0.06	0.31	0.56	0.24
	rs7246742	0.44	0.58	0.82	0.74	0.90	0.59	0.40	0.26	0.95
	rs7250713	0.37	0.17	0.74	0.88	0.39	0.01	0.25	0.20	0.34
EPHX1	rs1051740	0.95	0.29	0.75	0.74	0.48	0.58	0.58	0.37	0.44
	rs2234922	0.78	0.72	0.50	0.39	0.57	0.68	0.71	0.48	0.52
	rs2854450	1.00	0.94	0.27	0.75	0.70	0.07	0.71	0.79	0.77
	rs360063	0.62	0.13	0.28	0.42	0.08	0.04	0.03	0.15	0.01
NAT2	rs1041983	0.70	0.54	0.56	0.66	0.86	0.78	0.91	0.56	0.63
	rs1799929	0.89	0.98	0.32	0.92	0.89	0.38	0.77	0.95	0.10
	rs1799930	0.45	0.65	0.84	0.59	0.46	0.27	0.61	0.78	0.71
	rs1799931	0.51	0.73	0.96	0.67	0.54	0.64	0.60	0.84	0.80
	rs1801280	0.84	0.96	0.20	0.93	0.89	0.66	0.70	0.89	0.11

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity

A. NHW population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
CYP1A1	rs1456432	*CYP1A2*	rs2472299	0.19
	rs1456432		rs2470890	0.82
	rs1456432		rs11854147	0.26
	rs1048943		rs2472299	0.33
	rs1048943		rs2470890	0.30
	rs1048943		rs11854147	0.74
	rs2470893		rs2472299	0.56
	rs2470893		rs2470890	0.96
	rs2470893		rs11854147	0.28
CYP1A1	rs1456432	*CYP2A6*	rs7246742	0.09
	rs1456432		rs7250713	0.29
	rs1456432		rs4105144	0.44
	rs1048943		rs7246742	0.91
	rs1048943		rs7250713	0.64
	rs1048943		rs4105144	0.11
	rs2470893		rs7246742	0.85
	rs2470893		rs7250713	0.36
	rs2470893		rs4105144	0.53
CYP1A2	rs2472299	*CYP2A6*	rs7246742	0.72
	rs2472299		rs7250713	0.84
	rs2472299		rs4105144	0.80
	rs2470890		rs7246742	0.31
	rs2470890		rs7250713	0.53
	rs2470890		rs4105144	0.16
	rs11854147		rs7246742	0.81
	rs11854147		rs7250713	0.62
	rs11854147		rs4105144	0.80

B. Hispanic population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
CYP1A1	rs1456432	*CYP1A2*	rs2472299	0.04
	rs1456432		rs2470890	0.67
	rs1456432		rs11854147	0.05
	rs1048943		rs2472299	0.01
	rs1048943		rs2470890	0.07
	rs1048943		rs11854147	0.05
	rs2470893		rs2472299	0.76
	rs2470893		rs2470890	0.79
	rs2470893		rs11854147	0.74
CYP1A1	rs1456432	*CYP2A6*	rs7246742	0.93
	rs1456432		rs7250713	0.36
	rs1456432		rs4105144	0.49
	rs1048943		rs7246742	0.68
	rs1048943		rs7250713	0.07
	rs1048943		rs4105144	0.07
	rs2470893		rs7246742	0.49
	rs2470893		rs7250713	0.69
	rs2470893		rs4105144	0.82
CYP1A2	rs2472299	*CYP2A6*	rs7246742	0.78
	rs2472299		rs7250713	0.51
	rs2472299		rs4105144	0.69
	rs2470890		rs7246742	0.89
	rs2470890		rs7250713	0.95
	rs2470890		rs4105144	0.46
	rs11854147		rs7246742	0.73
	rs11854147		rs7250713	0.66
	rs11854147		rs4105144	0.21

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
CYP1B1	rs163084	*CYP1A1*	rs1456432	0.31
	rs163084		rs1048943	0.19
	rs163084		rs2470893	0.84
	rs1056836		rs1456432	0.43
	rs1056836		rs1048943	0.19
	rs1056836		rs2470893	0.36
	rs10012		rs1456432	0.90
	rs10012		rs1048943	0.27
	rs10012		rs2470893	0.07
	rs4646429		rs1456432	0.79
	rs4646429		rs1048943	0.40
	rs4646429		rs2470893	0.09
CYP1B1	rs163084	*CYP1A2*	rs2472299	0.65
	rs163084		rs2470890	0.73
	rs163084		rs11854147	0.90
	rs1056836		rs2472299	0.43
	rs1056836		rs2470890	0.77
	rs1056836		rs11854147	0.52
	rs10012		rs2472299	0.44
	rs10012		rs2470890	0.56
	rs10012		rs11854147	0.33
	rs4646429		rs2472299	0.23
	rs4646429		rs2470890	0.52
	rs4646429		rs11854147	0.42
CYP1B1	rs163084	*CYP2A6*	rs7246742	0.93
	rs163084		rs7250713	0.57
	rs163084		rs4105144	0.93
	rs1056836		rs7246742	0.14
	rs1056836		rs7250713	0.11
	rs1056836		rs4105144	0.02
	rs10012		rs7246742	0.54
	rs10012		rs7250713	0.44
	rs10012		rs4105144	0.51
	rs4646429		rs7246742	0.82
	rs4646429		rs7250713	0.24
	rs4646429		rs4105144	0.35

B. Hispanic population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
CYP1B1	rs163084	*CYP1A1*	rs1456432	0.94
	rs163084		rs1048943	0.45
	rs163084		rs2470893	0.36
	rs1056836		rs1456432	0.52
	rs1056836		rs1048943	0.46
	rs1056836		rs2470893	0.38
	rs10012		rs1456432	0.62
	rs10012		rs1048943	0.68
	rs10012		rs2470893	0.04
	rs4646429		rs1456432	0.72
	rs4646429		rs1048943	0.45
	rs4646429		rs2470893	0.06
CYP1B1	rs163084	*CYP1A2*	rs2472299	0.25
	rs163084		rs2470890	0.58
	rs163084		rs11854147	0.41
	rs1056836		rs2472299	0.05
	rs1056836		rs2470890	0.76
	rs1056836		rs11854147	0.60
	rs10012		rs2472299	0.52
	rs10012		rs2470890	0.87
	rs10012		rs11854147	0.93
	rs4646429		rs2472299	0.84
	rs4646429		rs2470890	0.50
	rs4646429		rs11854147	0.60
CYP1B1	rs163084	*CYP2A6*	rs7246742	0.49
	rs163084		rs7250713	0.81
	rs163084		rs4105144	0.55
	rs1056836		rs7246742	0.32
	rs1056836		rs7250713	0.52
	rs1056836		rs4105144	0.22
	rs10012		rs7246742	0.46
	rs10012		rs7250713	0.45
	rs10012		rs4105144	0.52
	rs4646429		rs7246742	0.39
	rs4646429		rs7250713	0.50
	rs4646429		rs4105144	0.58

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
CYP1B1	rs163084	*NAT2*	rs1041983	0.30
	rs163084		rs1801280	0.10
	rs163084		rs1799929	0.19
	rs163084		rs1799930	0.16
	rs163084		rs1799931	0.62
	rs1056836		rs1041983	0.76
	rs1056836		rs1801280	0.52
	rs1056836		rs1799929	0.33
	rs1056836		rs1799930	0.32
	rs1056836		rs1799931	0.81
	rs10012		rs1041983	0.98
	rs10012		rs1801280	0.40
	rs10012		rs1799929	0.46
	rs10012		rs1799930	0.35
	rs10012		rs1799931	0.58
	rs4646429		rs1041983	0.94
	rs4646429		rs1801280	0.52
	rs4646429		rs1799929	0.50
	rs4646429		rs1799930	0.50
	rs4646429		rs1799931	0.52
EPHX1	rs1051740	*CYP1A1*	rs1456432	0.40
	rs1051740		rs1048943	0.33
	rs1051740		rs2470893	0.89
	rs2234922		rs1456432	0.32
	rs2234922		rs1048943	0.45
	rs2234922		rs2470893	0.72
	rs2854450		rs1456432	0.42
	rs2854450		rs1048943	0.08
	rs2854450		rs2470893	0.59
	rs360063		rs1456432	0.14
	rs360063		rs1048943	0.74
	rs360063		rs2470893	0.14

B. Hispanic population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
CYP1B1	rs163084	*NAT2*	rs1041983	0.67
	rs163084		rs1801280	0.06
	rs163084		rs1799929	0.05
	rs163084		rs1799930	0.82
	rs163084		rs1799931	0.72
	rs1056836		rs1041983	0.70
	rs1056836		rs1801280	0.04
	rs1056836		rs1799929	0.04
	rs1056836		rs1799930	0.71
	rs1056836		rs1799931	0.47
	rs10012		rs1041983	0.57
	rs10012		rs1801280	0.17
	rs10012		rs1799929	0.12
	rs10012		rs1799930	0.86
	rs10012		rs1799931	0.66
	rs4646429		rs1041983	0.20
	rs4646429		rs1801280	0.42
	rs4646429		rs1799929	0.44
	rs4646429		rs1799930	0.63
	rs4646429		rs1799931	0.35
EPHX1	rs1051740	*CYP1A1*	rs1456432	0.92
	rs1051740		rs1048943	0.73
	rs1051740		rs2470893	0.45
	rs2234922		rs1456432	0.87
	rs2234922		rs1048943	0.65
	rs2234922		rs2470893	0.88
	rs2854450		rs1456432	0.51
	rs2854450		rs1048943	0.59
	rs2854450		rs2470893	0.41
	rs360063		rs1456432	0.67
	rs360063		rs1048943	0.78
	rs360063		rs2470893	0.43

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus		Second Locus		*p*-Value
Gene	**SNP**	**Gene**	**SNP**	
EPHX1	rs1051740	*CYP1A2*	rs2472299	0.55
	rs1051740		rs2470890	0.42
	rs1051740		rs11854147	0.91
	rs2234922		rs2472299	0.46
	rs2234922		rs2470890	0.96
	rs2234922		rs11854147	0.80
	rs2854450		rs2472299	0.40
	rs2854450		rs2470890	0.53
	rs2854450		rs11854147	0.24
	rs360063		rs2472299	0.21
	rs360063		rs2470890	0.16
	rs360063		rs11854147	0.18
EPHX1	rs1051740	*CYP2A6*	rs7246742	0.07
	rs1051740		rs7250713	0.05
	rs1051740		rs4105144	0.07
	rs2234922		rs7246742	0.05
	rs2234922		rs7250713	0.14
	rs2234922		rs4105144	0.21
	rs2854450		rs7246742	0.99
	rs2854450		rs7250713	0.76
	rs2854450		rs4105144	0.71
	rs360063		rs7246742	0.98
	rs360063		rs7250713	0.73
	rs360063		rs4105144	0.04

B. Hispanic population

First Locus		Second Locus		*p*-Value
Gene	**SNP**	**Gene**	**SNP**	
EPHX1	rs1051740	*CYP1A2*	rs2472299	0.18
	rs1051740		rs2470890	0.98
	rs1051740		rs11854147	0.96
	rs2234922		rs2472299	0.92
	rs2234922		rs2470890	0.74
	rs2234922		rs11854147	0.96
	rs2854450		rs2472299	0.37
	rs2854450		rs2470890	0.73
	rs2854450		rs11854147	0.38
	rs360063		rs2472299	0.30
	rs360063		rs2470890	0.71
	rs360063		rs11854147	0.54
EPHX1	rs1051740	*CYP2A6*	rs7246742	0.24
	rs1051740		rs7250713	0.29
	rs1051740		rs4105144	0.18
	rs2234922		rs7246742	0.79
	rs2234922		rs7250713	0.42
	rs2234922		rs4105144	0.64
	rs2854450		rs7246742	0.89
	rs2854450		rs7250713	0.18
	rs2854450		rs4105144	0.20
	rs360063		rs7246742	0.14
	rs360063		rs7250713	0.74
	rs360063		rs4105144	0.21

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
EPHX1	rs1051740	*CYP1B1*	rs163084	0.66
	rs1051740		rs1056836	0.68
	rs1051740		rs10012	0.38
	rs1051740		rs4646429	0.15
	rs2234922		rs163084	0.84
	rs2234922		rs1056836	0.58
	rs2234922		rs10012	0.41
	rs2234922		rs4646429	0.32
	rs2854450		rs163084	0.88
	rs2854450		rs1056836	0.10
	rs2854450		rs10012	0.10
	rs2854450		rs4646429	0.13
	rs360063		rs163084	0.46
	rs360063		rs1056836	0.69
	rs360063		rs10012	0.88
	rs360063		rs4646429	0.82
EPHX1	rs1051740	*NAT2*	rs1041983	0.19
	rs1051740		rs1801280	0.03
	rs1051740		rs1799929	0.01
	rs1051740		rs1799930	0.19
	rs1051740		rs1799931	0.98
	rs2234922		rs1041983	0.24
	rs2234922		rs1801280	0.57
	rs2234922		rs1799929	0.97
	rs2234922		rs1799930	0.35
	rs2234922		rs1799931	0.38
	rs2854450		rs1041983	0.78
	rs2854450		rs1801280	0.76
	rs2854450		rs1799929	0.94
	rs2854450		rs1799930	0.82
	rs2854450		rs1799931	0.19
	rs360063		rs1041983	0.21
	rs360063		rs1801280	0.47
	rs360063		rs1799929	0.41
	rs360063		rs1799930	0.10
	rs360063		rs1799931	0.29

B. Hispanic population

First Locus Gene	First Locus SNP	Second Locus Gene	Second Locus SNP	*p*-Value
EPHX1	rs1051740	*CYP1B1*	rs163084	0.48
	rs1051740		rs1056836	0.55
	rs1051740		rs10012	0.44
	rs1051740		rs4646429	0.99
	rs2234922		rs163084	0.22
	rs2234922		rs1056836	0.85
	rs2234922		rs10012	0.11
	rs2234922		rs4646429	0.12
	rs2854450		rs163084	0.55
	rs2854450		rs1056836	0.93
	rs2854450		rs10012	0.30
	rs2854450		rs4646429	0.57
	rs360063		rs163084	0.90
	rs360063		rs1056836	0.82
	rs360063		rs10012	0.12
	rs360063		rs4646429	0.08
EPHX1	rs1051740	*NAT2*	rs1041983	0.19
	rs1051740		rs1801280	0.46
	rs1051740		rs1799929	0.40
	rs1051740		rs1799930	0.27
	rs1051740		rs1799931	0.48
	rs2234922		rs1041983	0.73
	rs2234922		rs1801280	0.59
	rs2234922		rs1799929	0.94
	rs2234922		rs1799930	0.51
	rs2234922		rs1799931	0.97
	rs2854450		rs1041983	0.42
	rs2854450		rs1801280	0.24
	rs2854450		rs1799929	0.26
	rs2854450		rs1799930	0.97
	rs2854450		rs1799931	0.69
	rs360063		rs1041983	0.89
	rs360063		rs1801280	0.65
	rs360063		rs1799929	0.48
	rs360063		rs1799930	0.87
	rs360063		rs1799931	0.53

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
NAT2	rs1041983	*CYP1A1*	rs1456432	0.96
	rs1041983		rs1048943	0.85
	rs1041983		rs2470893	0.91
	rs1801280		rs1456432	0.93
	rs1801280		rs1048943	0.61
	rs1801280		rs2470893	0.92
	rs1799929		rs1456432	0.57
	rs1799929		rs1048943	0.60
	rs1799929		rs2470893	0.99
	rs1799930		rs1456432	0.61
	rs1799930		rs1048943	0.53
	rs1799930		rs2470893	0.79
	rs1799931		rs1456432	0.94
	rs1799931		rs1048943	0.65
	rs1799931		rs2470893	0.71
NAT2	rs1041983	*CYP1A2*	rs2472299	0.85
	rs1041983		rs2470890	0.91
	rs1041983		rs11854147	0.74
	rs1801280		rs2472299	0.41
	rs1801280		rs2470890	0.61
	rs1801280		rs11854147	0.74
	rs1799929		rs2472299	0.10
	rs1799929		rs2470890	0.40
	rs1799929		rs11854147	0.20
	rs1799930		rs2472299	0.73
	rs1799930		rs2470890	0.92
	rs1799930		rs11854147	0.88
	rs1799931		rs2472299	0.49
	rs1799931		rs2470890	0.51
	rs1799931		rs11854147	0.95

B. Hispanic population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
NAT2	rs1041983	*CYP1A1*	rs1456432	0.72
	rs1041983		rs1048943	0.99
	rs1041983		rs2470893	0.46
	rs1801280		rs1456432	0.56
	rs1801280		rs1048943	0.67
	rs1801280		rs2470893	0.15
	rs1799929		rs1456432	0.50
	rs1799929		rs1048943	0.73
	rs1799929		rs2470893	0.06
	rs1799930		rs1456432	0.19
	rs1799930		rs1048943	0.48
	rs1799930		rs2470893	0.61
	rs1799931		rs1456432	0.44
	rs1799931		rs1048943	0.82
	rs1799931		rs2470893	0.92
NAT2	rs1041983	*CYP1A2*	rs2472299	0.30
	rs1041983		rs2470890	0.52
	rs1041983		rs11854147	0.36
	rs1801280		rs2472299	0.81
	rs1801280		rs2470890	0.23
	rs1801280		rs11854147	0.50
	rs1799929		rs2472299	0.51
	rs1799929		rs2470890	0.11
	rs1799929		rs11854147	0.30
	rs1799930		rs2472299	0.12
	rs1799930		rs2470890	0.52
	rs1799930		rs11854147	0.58
	rs1799931		rs2472299	0.65
	rs1799931		rs2470890	0.93
	rs1799931		rs11854147	0.92

Supplementary Table 5. Gene interactions for smoking metabolism genes by ethnicity (cont.)

A. NHW population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
NAT2	rs1041983	*CYP2A6*	rs7246742	0.99
	rs1041983		rs7250713	0.61
	rs1041983		rs4105144	0.21
	rs1801280		rs7246742	0.35
	rs1801280		rs7250713	0.34
	rs1801280		rs4105144	0.40
	rs1799929		rs7246742	0.56
	rs1799929		rs7250713	0.43
	rs1799929		rs4105144	0.26
	rs1799930		rs7246742	0.71
	rs1799930		rs7250713	0.20
	rs1799930		rs4105144	0.04
	rs1799931		rs7246742	0.16
	rs1799931		rs7250713	0.32
	rs1799931		rs4105144	0.33

B. Hispanic population

First Locus		Second Locus		*p*-Value
Gene	SNP	Gene	SNP	
NAT2	rs1041983	*CYP2A6*	rs7246742	0.19
	rs1041983		rs7250713	0.36
	rs1041983		rs4105144	0.24
	rs1801280		rs7246742	0.42
	rs1801280		rs7250713	0.45
	rs1801280		rs4105144	0.36
	rs1799929		rs7246742	0.22
	rs1799929		rs7250713	0.18
	rs1799929		rs4105144	0.12
	rs1799930		rs7246742	0.23
	rs1799930		rs7250713	0.06
	rs1799930		rs4105144	0.10
	rs1799931		rs7246742	0.16
	rs1799931		rs7250713	0.90
	rs1799931		rs4105144	0.49

Supplementary Table 6. Results of log-linear regression models for the independent effects of maternal and child genotypes

Gene	SNP	LRT[b]				RR[c]			
		Mother	*p*-value	Child	*p*-value	Mother	*p*-value	Child	*p*-value
CYP1A1	rs1456432	0.98	0.32	0.39	0.53	1.12	0.32	1.07	0.53
	rs1048943	1.03	0.31	2.73	0.10	1.14	0.31	0.79	0.10
	rs2470893	0.20	0.65	1.16	0.28	0.95	0.65	0.88	0.28
CYP1A2	rs2472299	2.70	0.10	3.02	0.08	0.84	0.10	0.84	0.08
	rs2470890	**3.62**	**0.06**	**7.77**	**0.01**	**1.23**	**0.06**	**1.33**	**0.01**
	rs11854147	**4.46**	**0.03**	**2.90**	**0.09**	**1.24**	**0.04**	**1.19**	**0.09**
CYP1B1	rs163084	0.25	0.62	2.07	0.15	0.94	0.62	1.19	0.15
	rs1056836	0.16	0.69	2.01	0.16	1.04	0.69	1.15	0.16
	rs10012	0.02	0.89	0.37	0.54	0.99	0.89	1.06	0.54
	rs4646429	0.01	0.92	0.20	0.65	1.01	0.92	1.05	0.66
CYP2A6	rs7246742	2.49	0.11	0.02	0.89	0.83	0.12	1.02	0.89
	rs7250713	0.21	0.64	0.03	0.87	1.04	0.65	0.98	0.87
	rs4105144	0.04	0.85	1.43	0.23	0.98	0.84	0.87	0.23
EPHX1	rs2854450	0.48	0.49	0.24	0.62	0.92	0.49	0.94	0.62
	rs1051740	0.03	0.87	0.05	0.82	1.02	0.87	1.02	0.82
	rs2234922	0.35	0.55	2.21	0.14	0.91	0.55	0.81	0.14
	rs360063	0.05	0.83	2.79	0.09	1.02	0.83	0.85	0.10
NAT2	rs1041983	0.49	0.49	0.05	0.82	0.93	0.48	0.98	0.82
	rs1801280	0.10	0.76	0.99	0.32	1.03	0.75	1.10	0.32
	rs1799929	0.16	0.69	1.28	0.26	0.96	0.69	0.89	0.26
	rs1799930	0.07	0.79	0.63	0.43	1.03	0.79	1.09	0.43
	rs1799931	0.32	0.57	0.42	0.52	0.91	0.57	0.89	0.52

[a] Log-likelihood

[b] Log-likehood ratio = $2*(LL_{Full}-LL_{Maternal/Child})$

[c] Relative risk

Supplementary Table 7. Nulle allele chi-squared contingency tables

GSTM1 / *GSTT1*

Proband

GSTM1

		In-utero Exposure	
		Yes	No
Allele	WT	33	209
	Null	24	183

$X^2 = 0.420$; $p = 0.517$

GSTT1

		In-utero Exposure	
		Yes	No
Allele	WT	46	321
	Null	11	71

$X^2 = 0.0469$; $p = 0.829$

Mom

GSTM1

		Maternal Smoking	
		Yes	No
Allele	WT	26	172
	Null	29	181

$X^2 = 0.0402$; $p = 0.841$

GSTT1

		Maternal Smoking	
		Yes	No
Allele	WT	46	300
	Null	9	53

$X^2 = 0.0672$; $p = 0.795$

Mom vs Dad

GSTM1

		Parent	
		Mom	Dad
Allele	WT	282	183
	Null	281	149

$X^2 = 2.12$; $p = 0.146$

GSTT1

		Parent	
		Mom	Dad
Allele	WT	473	271
	Null	90	62

$X^2 = 1.03$; $p = 0.310$

Mom vs Dad - Hispanic Only

GSTM1

		Parent	
		Mom	Dad
Allele	WT	163	85
	Null	133	60

$X^2 = 0.499$; $p = 0.480$

GSTT1

		Parent	
		Mom	Dad
Allele	WT	259	123
	Null	37	22

$X^2 = 0.600$; $p = 0.439$

Mom vs Dad - NHW Only

GSTM1

		Parent	
		Mom	Dad
Allele	WT	118	96
	Null	145	89

$X^2 = 2.15$; $p = 0.143$

GSTT1

		Parent	
		Mom	Dad
Allele	WT	212	146
	Null	51	40

$X^2 = 0.301$; $p = 0.583$

REFERENCES

1. Turco, V. J. 1981. Clubfoot. Churchill Livingstone, New York.
2. Reimann, I. 1967. Congenital idiopathic club foot with special reference to aetiology, pathogenesis, and possibilities of correction within the first years of life. Munksgaard, Køobenhavn,.
3. Gurnett, C. A., S. Boehm, A. Connolly, T. Reimschisel, and M. B. Dobbs. 2008. Impact of congenital talipes equinovarus etiology on treatment outcomes. Dev Med Child Neurol 50:498-502.
4. Bakalis, S., S. Sairam, T. Homfray, K. Harrington, K. Nicolaides, and B. Thilaganathan. 2002. Outcome of antenatally diagnosed talipes equinovarus in an unselected obstetric population. Ultrasound Obstet Gynecol 20:226-229.
5. Ching, G. H., C. S. Chung, and R. W. Nemechek. 1969. Genetic and epidemiological studies of clubfoot in Hawaii: ascertainment and incidence. Am J Hum Genet 21:566-580.
6. Moorthi, R. N., S. S. Hashmi, P. Langois, M. Canfield, D. K. Waller, and J. T. Hecht. 2005. Idiopathic talipes equinovarus (ITEV) (clubfeet) in Texas. Am J Med Genet A 132:376-380.
7. Wynne-Davies, R. 1965. Family studies and aetiology of club foot. J Med Genet 2:227-232.
8. Byron-Scott, R., P. Sharpe, C. Hasler, P. Cundy, C. Hirte, A. Chan, H. Scott, P. Baghurst, and E. Haan. 2005. A South Australian population-based study of congenital talipes equinovarus. Paediatr Perinat Epidemiol 19:227-237.
9. Chapman, C., N. S. Stott, R. V. Port, and R. O. Nicol. 2000. Genetics of club foot in Maori and Pacific people. J Med Genet 37:680-683.
10. Cardy, A. H., S. Barker, D. Chesney, L. Sharp, N. Maffulli, and Z. Miedzybrodzka. 2007. Pedigree analysis and epidemiological features of idiopathic congenital talipes equinovarus in the United Kingdom: a case-control study. BMC Musculoskelet Disord 8:62.
11. Cartlidge, I. 1984. Observations on the epidemiology of club foot in Polynesian and Caucasian populations. J Med Genet 21:290-292.

12. Moore, K. L., T. V. N. Persaud, and M. G. Torchia. 2008. The developing human : clinically oriented embryology. Saunders/Elsevier, Philadelphia, PA.
13. Lovell, W. W., R. B. Winter, R. T. Morrissy, and S. L. V. Weinstein. 1996. Lovell and Winter's pediatric orthopaedics. Lippincott-Raven, Philadelphia, PA.
14. Mooney, E., and C. Loh. 2008. Lower Limb Embryology. In eMedicine. WebMD.
15. Canto, M. J., S. Cano, J. Palau, and F. Ojeda. 2008. Prenatal diagnosis of clubfoot in low-risk population: associated anomalies and long-term outcome. Prenat Diagn 28:343-346.
16. Dobbs, M. B., and C. A. Gurnett. 2009. Update on clubfoot: etiology and treatment. Clin Orthop Relat Res 467:1146-1153.
17. Roye, D. P., Jr., and B. D. Roye. 2002. Idiopathic congenital talipes equinovarus. J Am Acad Orthop Surg 10:239-248.
18. Ponseti, I. V. 1996. Congenital clubfoot: fundamentals of treatment. Oxford University Press, New York.
19. Hootnick, D. R., E. M. Levinsohn, R. J. Crider, and D. S. Packard, Jr. 1982. Congenital arterial malformations associated with clubfoot. A report of two cases. Clin Orthop Relat Res:160-163.
20. Dunn, P. M. 1972. Congenital postural deformities: perinatal associations. Proc R Soc Med 65:735-738.
21. Wynne-Davies, R. 1964. Family Studies and the Cause of Congenital Club Foot. Talipes Equinovarus, Talipes Calcaneo-Valgus and Metatarsus Varus. J Bone Joint Surg Br 46:445-463.
22. Alderman, B. W., E. R. Takahashi, and M. K. LeMier. 1991. Risk indicators for talipes equinovarus in Washington State, 1987-1989. Epidemiology 2:289-292.
23. Idelberger, K. H. 1939. Die Ergenbnisse der Zwillingsforschung beim angeborenen Klumpfuss. Verh Dtsch Orthop Ges:272-276.
24. de Andrade, M., J. S. Barnholtz, C. I. Amos, C. Lochmiller, A. Scott, M. Risman, and J. T. Hecht. 1998. Segregation analysis of idiopathic talipes equinovarus in a Texan population. Am J Med Genet 79:97-102.
25. Miedzybrodzka, Z. 2003. Congenital talipes equinovarus (clubfoot): a disorder of the foot but not the hand. J Anat 202:37-42.

26. Engell, V., F. Damborg, M. Andersen, K. O. Kyvik, and K. Thomsen. 2006. Club foot: a twin study. J Bone Joint Surg Br 88:374-376.
27. Wang, J. H., R. M. Palmer, and C. S. Chung. 1988. The role of major gene in clubfoot. Am J Hum Genet 42:772-776.
28. Kruse, L. M., M. B. Dobbs, and C. A. Gurnett. 2008. Polygenic threshold model with sex dimorphism in clubfoot inheritance: the Carter effect. J Bone Joint Surg Am 90:2688-2694.
29. Rebbeck, T. R., F. R. Dietz, J. C. Murray, and K. H. Buetow. 1993. A single-gene explanation for the probability of having idiopathic talipes equinovarus. Am J Hum Genet 53:1051-1063.
30. Lochmiller, C., D. Johnston, A. Scott, M. Risman, and J. T. Hecht. 1998. Genetic epidemiology study of idiopathic talipes equinovarus. Am J Med Genet 79:90-96.
31. Harper, P. 2004. Practical Genetic Counseling. Hodder Arnold, London.
32. CEMAT, T. C. E. a. M.-t. A. T. G. 1998. Randomised trial to assess safety and fetal outcome of early and midtrimester amniocentesis. . Lancet 351:242-247.
33. Centini, G., L. Rosignoli, A. Kenanidis, R. Scarinci, and F. Petraglia. 2003. A report of early (13 + 0 to 14 + 6 weeks) and mid-trimester amniocenteses: 10 years' experience. J Matern Fetal Neonatal Med 14:113-117.
34. Delisle, M. F., and R. D. Wilson. 1999. First trimester prenatal diagnosis: amniocentesis. Semin Perinatol 23:414-423.
35. Nicolaides, K. H., M. L. Brizot, F. Patel, and R. Snjders. 1996. Comparison of chorion villus sampling and early amniocentesis for karyotyping in 1,492 singleton pregnancies. Fetal Diagn Ther 11:9-15.
36. Philip, J., R. K. Silver, R. D. Wilson, E. A. Thom, J. M. Zachary, P. Mohide, M. J. Mahoney, J. L. Simpson, L. D. Platt, E. Pergament, D. Hershey, K. Filkins, A. Johnson, L. P. Shulman, J. Bang, S. MacGregor, J. R. Smith, D. Shaw, R. J. Wapner, and L. G. Jackson. 2004. Late first-trimester invasive prenatal diagnosis: results of an international randomized trial. Obstet Gynecol 103:1164-1173.
37. Honein, M. A., L. J. Paulozzi, and C. A. Moore. 2000. Family history, maternal smoking, and clubfoot: an indication of a gene-environment interaction. Am J Epidemiol 152:658-665.

38. Skelly, A. C., V. L. Holt, V. S. Mosca, and B. W. Alderman. 2002. Talipes equinovarus and maternal smoking: a population-based case-control study in Washington state. Teratology 66:91-100.
39. Dickinson, K. C., R. E. Meyer, and J. Kotch. 2008. Maternal smoking and the risk for clubfoot in infants. Birth Defects Res A Clin Mol Teratol 82:86-91.
40. Drayna, D. Strategies for finding disease genes. In Section on Systems Biology of Communication Disorders. LMG, NIDCD, NIH.
41. Collins, A. 2009. Approaches to the identification of susceptibility genes. Parasite Immunol 31:225-233.
42. McCarthy, M. I., D. Smedley, and W. Hide. 2003. New methods for finding disease-susceptibility genes: impact and potential. Genome Biol 4:119.
43. Iannuzzi, M. C., B. A. Rybicki, M. Maliarik, and J. Popovich, Jr. 1997. Finding disease genes. From cystic fibrosis to sarcoidosis. Thomas A. Neff Lecture. Chest 111:70S-73S.
44. Slagboom, P. E., and I. Meulenbelt. 2002. Organisation of the human genome and our tools for identifying disease genes. Biol Psychol 61:11-31.
45. Nakamura, Y. 2009. DNA variations in human and medical genetics: 25 years of my experience. J Hum Genet 54:1-8.
46. 2010. Genetic Science Learning Center In Making SNPs Make Sense. University of Utah.
47. Hecht, J. T., A. Ester, A. Scott, C. A. Wise, D. M. Iovannisci, E. J. Lammer, P. H. Langlois, and S. H. Blanton. 2007. *NAT2* variation and idiopathic talipes equinovarus (clubfoot). Am J Med Genet A 143A:2285-2291.
48. Ester, A. R., G. Tyerman, C. A. Wise, S. H. Blanton, and J. T. Hecht. 2007. Apoptotic gene analysis in idiopathic talipes equinovarus (clubfoot). Clin Orthop Relat Res 462:32-37.
49. Ester, A. R., K. S. Weymouth, A. Burt, C. A. Wise, A. Scott, C. A. Gurnett, M. B. Dobbs, S. H. Blanton, and J. T. Hecht. 2009. Altered transmission of HOX and apoptotic SNPs identify a potential common pathway for clubfoot. Am J Med Genet A 149A:2745-2752.

50. Meyer, U. A. 1996. Overview of enzymes of drug metabolism. J Pharmacokinet Biopharm 24:449-459.
51. Voet, D., Voet, J.G. 2004. Biochemistry. John Wiley & Sons, Inc.
52. Kitamura, M., and A. Kasai. 2007. Cigarette smoke as a trigger for the dioxin receptor-mediated signaling pathway. Cancer Lett 252:184-194.
53. Hecht, S. S. 1999. Tobacco smoke carcinogens and lung cancer. J Natl Cancer Inst 91:1194-1210.
54. Hoffmann, D., and I. Hoffmann. 1997. The changing cigarette, 1950-1995. J Toxicol Environ Health 50:307-364.
55. Becher, G., and A. Bjorseth. 1983. Determination of exposure to polycyclic aromatic hydrocarbons by analysis of human urine. Cancer Lett 17:301-311.
56. Detmar, J., M. Y. Rennie, K. J. Whiteley, D. Qu, Y. Taniuchi, X. Shang, R. F. Casper, S. L. Adamson, J. G. Sled, and A. Jurisicova. 2008. Fetal growth restriction triggered by polycyclic aromatic hydrocarbons is associated with altered placental vasculature and AhR-dependent changes in cell death. Am J Physiol Endocrinol Metab 295:E519-530.
57. Pasanen, M. 1999. The expression and regulation of drug metabolism in human placenta. Adv Drug Deliv Rev 38:81-97.
58. Aitken, A. E., T. A. Richardson, and E. T. Morgan. 2006. Regulation of drug-metabolizing enzymes and transporters in inflammation. Annu Rev Pharmacol Toxicol 46:123-149.
59. Nock, N. L., D. Tang, A. Rundle, C. Neslund-Dudas, A. T. Savera, C. H. Bock, K. G. Monaghan, A. Koprowski, N. Mitrache, J. J. Yang, and B. A. Rybicki. 2007. Associations between smoking, polymorphisms in polycyclic aromatic hydrocarbon (PAH) metabolism and conjugation genes and PAH-DNA adducts in prostate tumors differ by race. Cancer Epidemiol Biomarkers Prev 16:1236-1245.
60. Shimada, T., J. Watanabe, K. Kawajiri, T. R. Sutter, F. P. Guengerich, E. M. Gillam, and K. Inoue. 1999. Catalytic properties of polymorphic human cytochrome P450 1B1 variants. Carcinogenesis 20:1607-1613.
61. Shimada, T., H. Yamazaki, M. Mimura, N. Wakamiya, Y. F. Ueng, F. P. Guengerich, and Y. Inui. 1996. Characterization of microsomal cytochrome P450

enzymes involved in the oxidation of xenobiotic chemicals in human fetal liver and adult lungs. Drug Metab Dispos 24:515-522.

62. Pavek, P., and Z. Dvorak. 2008. Xenobiotic-induced transcriptional regulation of xenobiotic metabolizing enzymes of the cytochrome P450 superfamily in human extrahepatic tissues. Curr Drug Metab 9:129-143.

63. Rodriguez-Antona, C., and M. Ingelman-Sundberg. 2006. Cytochrome P450 pharmacogenetics and cancer. Oncogene 25:1679-1691.

64. Gladen, B. C., T. D. Zadorozhnaja, N. Chislovska, D. O. Hryhorczuk, M. C. Kennicutt, 2nd, and R. E. Little. 2000. Polycyclic aromatic hydrocarbons in placenta. Hum Exp Toxicol 19:597-603.

65. Shugart, L., and R. Matsunami. 1985. Adduct formation in hemoglobin of the newborn mouse exposed in utero to benzo[a]pyrene. Toxicology 37:241-245.

66. Czekaj, P., A. Wiaderkiewicz, E. Florek, and R. Wiaderkiewicz. 2005. Tobacco smoke-dependent changes in cytochrome P450 1A1, 1A2, and 2E1 protein expressions in fetuses, newborns, pregnant rats, and human placenta. Arch Toxicol 79:13-24.

67. Delpisheh, A., L. Brabin, J. Topping, M. Reyad, A. W. Tang, and B. J. Brabin. 2009. A case-control study of *CYP1A1*, *GSTT1* and *GSTM1* gene polymorphisms, pregnancy smoking and fetal growth restriction. Eur J Obstet Gynecol Reprod Biol 143:38-42.

68. Wang, X., B. Zuckerman, C. Pearson, G. Kaufman, C. Chen, G. Wang, T. Niu, P. H. Wise, H. Bauchner, and X. Xu. 2002. Maternal cigarette smoking, metabolic gene polymorphism, and infant birth weight. JAMA 287:195-202.

69. Sasaki, S., T. Kondo, F. Sata, Y. Saijo, S. Katoh, S. Nakajima, M. Ishizuka, S. Fujita, and R. Kishi. 2006. Maternal smoking during pregnancy and genetic polymorphisms in the Ah receptor, *CYP1A1* and *GSTM1* affect infant birth size in Japanese subjects. Mol Hum Reprod 12:77-83.

70. Kishi, R., F. Sata, E. Yoshioka, S. Ban, S. Sasaki, K. Konishi, and N. Washino. 2008. Exploiting gene-environment interaction to detect adverse health effects of environmental chemicals on the next generation. Basic Clin Pharmacol Toxicol 102:191-203.

71. Legraverend, C., T. M. Guenthner, and D. W. Nebert. 1984. Importance of the route of administration for genetic differences in benzo[a]pyrene-induced in utero toxicity and teratogenicity. Teratology 29:35-47.
72. 2004. Drug Metabolizing Enzymes. In Technical Resource Guide. Invitrogen, editor.
73. Nelson, D. 2008. Human P450 data.
74. Guengerich, F. P. 1992. Characterization of human cytochrome P450 enzymes. FASEB J 6:745-748.
75. Schwarz, D., P. Kisselev, A. Chernogolov, W. H. Schunck, and I. Roots. 2005. Human *CYP1A1* variants lead to differential eicosapentaenoic acid metabolite patterns. Biochem Biophys Res Commun 336:779-783.
76. Lamba, J. K., Y. S. Lin, E. G. Schuetz, and K. E. Thummel. 2002. Genetic contribution to variable human CYP3A-mediated metabolism. Adv Drug Deliv Rev 54:1271-1294.
77. Georgiadis, P., N. A. Demopoulos, J. Topinka, G. Stephanou, M. Stoikidou, M. Bekyrou, K. Katsouyianni, R. Sram, H. Autrup, and S. A. Kyrtopoulos. 2004. Impact of phase I or phase II enzyme polymorphisms on lymphocyte DNA adducts in subjects exposed to urban air pollution and environmental tobacco smoke. Toxicol Lett 149:269-280.
78. Aklillu, E., S. Ovrebo, I. V. Botnen, C. Otter, and M. Ingelman-Sundberg. 2005. Characterization of common *CYP1B1* variants with different capacity for benzo[a]pyrene-7,8-dihydrodiol epoxide formation from benzo[a]pyrene. Cancer Res 65:5105-5111.
79. Oscarson, M. 2001. Genetic polymorphisms in the cytochrome P450 2A6 (*CYP2A6*) gene: implications for interindividual differences in nicotine metabolism. Drug Metab Dispos 29:91-95.
80. Omiecinski, C. J., C. Hassett, and V. Hosagrahara. 2000. Epoxide hydrolase--polymorphism and role in toxicology. Toxicol Lett 112-113:365-370.
81. Hassett, C., K. B. Robinson, N. B. Beck, and C. J. Omiecinski. 1994. The human microsomal epoxide hydrolase gene (*EPHX1*): complete nucleotide sequence and structural characterization. Genomics 23:433-442.

82. Sandberg, M., C. Hassett, E. T. Adman, J. Meijer, and C. J. Omiecinski. 2000. Identification and functional characterization of human soluble epoxide hydrolase genetic polymorphisms. J Biol Chem 275:28873-28881.

83. Liu, S., S. P. Stoesz, and C. B. Pickett. 1998. Identification of a novel human glutathione S-transferase using bioinformatics. Arch Biochem Biophys 352:306-313.

84. Oesch, F., H. Glatt, and H. Schmassmann. 1977. The apparent ubiquity of epoxide hydratase in rat organs. Biochem Pharmacol 26:603-607.

85. Strickler, S. M., L. V. Dansky, M. A. Miller, M. H. Seni, E. Andermann, and S. P. Spielberg. 1985. Genetic predisposition to phenytoin-induced birth defects. Lancet 2:746-749.

86. Hassett, C., L. Aicher, J. S. Sidhu, and C. J. Omiecinski. 1994. Human microsomal epoxide hydrolase: genetic polymorphism and functional expression in vitro of amino acid variants. Hum Mol Genet 3:421-428.

87. Hosagrahara, V. P., A. E. Rettie, C. Hassett, and C. J. Omiecinski. 2004. Functional analysis of human microsomal epoxide hydrolase genetic variants. Chem Biol Interact 150:149-159.

88. Seidegard, J., W. R. Vorachek, R. W. Pero, and W. R. Pearson. 1988. Hereditary differences in the expression of the human glutathione transferase active on trans-stilbene oxide are due to a gene deletion. Proc Natl Acad Sci U S A 85:7293-7297.

89. Bolt, H. M., and R. Thier. 2006. Relevance of the deletion polymorphisms of the glutathione S-transferases *GSTT1* and *GSTM1* in pharmacology and toxicology. Curr Drug Metab 7:613-628.

90. Pearson, W. R., W. R. Vorachek, S. J. Xu, R. Berger, I. Hart, D. Vannais, and D. Patterson. 1993. Identification of class-mu glutathione transferase genes *GSTM1*-GSTM5 on human chromosome 1p13. Am J Hum Genet 53:220-233.

91. Zhong, S., C. R. Wolf, and N. K. Spurr. 1992. Chromosomal assignment and linkage analysis of the human glutathione S-transferase mu gene (*GSTM1*) using intron specific polymerase chain reaction. Hum Genet 90:435-439.

92. Seidegard, J., and R. W. Pero. 1985. The hereditary transmission of high glutathione transferase activity towards trans-stilbene oxide in human mononuclear leukocytes. Hum Genet 69:66-68.

93. Wiencke, J. K., K. T. Kelsey, R. A. Lamela, and W. A. Toscano, Jr. 1990. Human glutathione S-transferase deficiency as a marker of susceptibility to epoxide-induced cytogenetic damage. Cancer Res 50:1585-1590.

94. Seidegard, J., R. W. Pero, D. G. Miller, and E. J. Beattie. 1986. A glutathione transferase in human leukocytes as a marker for the susceptibility to lung cancer. Carcinogenesis 7:751-753.

95. Meyer, D. J., B. Coles, S. E. Pemble, K. S. Gilmore, G. M. Fraser, and B. Ketterer. 1991. Theta, a new class of glutathione transferases purified from rat and man. Biochem J 274 (Pt 2):409-414.

96. Hussey, A. J., and J. D. Hayes. 1992. Characterization of a human class-Theta glutathione S-transferase with activity towards 1-menaphthyl sulphate. Biochem J 286 (Pt 3):929-935.

97. Pemble, S., K. R. Schroeder, S. R. Spencer, D. J. Meyer, E. Hallier, H. M. Bolt, B. Ketterer, and J. B. Taylor. 1994. Human glutathione S-transferase theta (*GSTT1*): cDNA cloning and the characterization of a genetic polymorphism. Biochem J 300 (Pt 1):271-276.

98. Ambrosone, C. B., J. L. Freudenheim, S. Graham, J. R. Marshall, J. E. Vena, J. R. Brasure, A. M. Michalek, R. Laughlin, T. Nemoto, K. A. Gillenwater, and P. G. Shields. 1996. Cigarette smoking, N-acetyltransferase 2 genetic polymorphisms, and breast cancer risk. JAMA 276:1494-1501.

99. Sim, E., M. Payton, M. Noble, and R. Minchin. 2000. An update on genetic, structural and functional studies of arylamine N-acetyltransferases in eucaryotes and procaryotes. Hum Mol Genet 9:2435-2441.

100. Meyer, U. A., and U. M. Zanger. 1997. Molecular mechanisms of genetic polymorphisms of drug metabolism. Annu Rev Pharmacol Toxicol 37:269-296.

101. Grant, D. M., K. Morike, M. Eichelbaum, and U. A. Meyer. 1990. Acetylation pharmacogenetics. The slow acetylator phenotype is caused by decreased or absent arylamine N-acetyltransferase in human liver. J Clin Invest 85:968-972.

102. Bell, D. A., J. A. Taylor, M. A. Butler, E. A. Stephens, J. Wiest, L. H. Brubaker, F. F. Kadlubar, and G. W. Lucier. 1993. Genotype/phenotype discordance for human

arylamine N-acetyltransferase (*NAT2*) reveals a new slow-acetylator allele common in African-Americans. Carcinogenesis 14:1689-1692.

103. Cascorbi, I., N. Drakoulis, J. Brockmoller, A. Maurer, K. Sperling, and I. Roots. 1995. Arylamine N-acetyltransferase (*NAT2*) mutations and their allelic linkage in unrelated Caucasian individuals: correlation with phenotypic activity. Am J Hum Genet 57:581-592.

104. Cascorbi, I., J. Brockmoller, S. Bauer, T. Reum, and I. Roots. 1996. *NAT2**12A (803A-->G) codes for rapid arylamine n-acetylation in humans. Pharmacogenetics 6:257-259.

105. Heron, M., P. D. Sutton, J. Xu, S. J. Ventura, D. M. Strobino, and B. Guyer. Annual summary of vital statistics: 2007. Pediatrics 125:4-15.

106. Arand, M., R. Muhlbauer, J. Hengstler, E. Jager, J. Fuchs, L. Winkler, and F. Oesch. 1996. A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase *GSTM1* and *GSTT1* polymorphisms. Anal Biochem 236:184-186.

107. Abecasis, G. R., and W. O. Cookson. 2000. GOLD--graphical overview of linkage disequilibrium. Bioinformatics 16:182-183.

108. Abecasis, G. R., S. S. Cherny, W. O. Cookson, and L. R. Cardon. 2002. Merlin--rapid analysis of dense genetic maps using sparse gene flow trees. Nat Genet 30:97-101.

109. Martin, E. R., S. A. Monks, L. L. Warren, and N. L. Kaplan. 2000. A test for linkage and association in general pedigrees: the pedigree disequilibrium test. Am J Hum Genet 67:146-154.

110. Martin, E. R., M. P. Bass, J. R. Gilbert, M. A. Pericak-Vance, and E. R. Hauser. 2003. Genotype-based association test for general pedigrees: the genotype-PDT. Genet Epidemiol 25:203-213.

111. Chung, R. H., E. R. Hauser, and E. R. Martin. 2006. The APL test: extension to general nuclear families and haplotypes and examination of its robustness. Hum Hered 61:189-199.

112. Hancock, D. B., E. R. Martin, Y. J. Li, and W. K. Scott. 2007. Methods for interaction analyses using family-based case-control data: conditional logistic regression versus generalized estimating equations. Genet Epidemiol 31:883-893.
113. van Den Oord, E. J., and J. K. Vermunt. 2000. Testing for linkage disequilibrium, maternal effects, and imprinting with (In)complete case-parent triads, by use of the computer program LEM. Am J Hum Genet 66:335-338.
114. Yue, P., E. Melamud, and J. Moult. 2006. SNPs3D: candidate gene and SNP selection for association studies. BMC Bioinformatics 7:166.
115. Teng, S., E. Michonova-Alexova, and E. Alexov. 2008. Approaches and resources for prediction of the effects of non-synonymous single nucleotide polymorphism on protein function and interactions. Curr Pharm Biotechnol 9:123-133.
116. Schug, J. 2003. Using TESS to predict transcription factor binding sites in DNA sequence. In Current protocols in bioinformatics. A. Baxevanis, editor. John Wiley and Sons, Inc., Hobokem, NJ. 2.6.1-2.6.15.
117. Grabe, N. 2002. AliBaba2: context specific identification of transcription factor binding sites. In Silico Biol 2:S1-15.
118. Matys, V., O. V. Kel-Margoulis, E. Fricke, I. Liebich, S. Land, A. Barre-Dirrie, I. Reuter, D. Chekmenev, M. Krull, K. Hornischer, N. Voss, P. Stegmaier, B. Lewicki-Potapov, H. Saxel, A. E. Kel, and E. Wingender. 2006. TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res 34:D108-110.
119. Wickens, M., D. S. Bernstein, J. Kimble, and R. Parker. 2002. A PUF family portrait: 3'UTR regulation as a way of life. Trends Genet 18:150-157.
120. Nielsen, K. B., S. Sorensen, L. Cartegni, T. J. Corydon, T. K. Doktor, L. D. Schroeder, L. S. Reinert, O. Elpeleg, A. R. Krainer, N. Gregersen, J. Kjems, and B. S. Andresen. 2007. Seemingly neutral polymorphic variants may confer immunity to splicing-inactivating mutations: a synonymous SNP in exon 5 of MCAD protects from deleterious mutations in a flanking exonic splicing enhancer. Am J Hum Genet 80:416-432.

121. Izzotti, A., R. M. Balansky, C. Cartiglia, A. Camoirano, M. Longobardi, and S. De Flora. 2003. Genomic and transcriptional alterations in mouse fetus liver after transplacental exposure to cigarette smoke. FASEB J 17:1127-1129.

122. Pinorini-Godly, M. T., and S. R. Myers. 1996. HPLC and GC/MS determination of 4-aminobiphenyl haemoglobin adducts in fetuses exposed to the tobacco smoke carcinogen in utero. Toxicology 107:209-217.

123. Law, A. J., J. E. Kleinman, D. R. Weinberger, and C. S. Weickert. 2007. Disease-associated intronic variants in the ErbB4 gene are related to altered ErbB4 splice-variant expression in the brain in schizophrenia. Hum Mol Genet 16:129-141.

124. Zhang, L., R. Bacares, S. Boyar, C. Hudis, K. Nafa, and K. Offit. 2009. cDNA analysis demonstrates that the BRCA2 intronic variant IVS4-12del5 is a deleterious mutation. Mutat Res 663:84-89.

125. Deguchi, T. 1992. Sequences and expression of alleles of polymorphic arylamine N-acetyltransferase of human liver. J Biol Chem 267:18140-18147.

126. Hein, D. W., R. J. Ferguson, M. A. Doll, T. D. Rustan, and K. Gray. 1994. Molecular genetics of human polymorphic N-acetyltransferase: enzymatic analysis of 15 recombinant wild-type, mutant, and chimeric *NAT2* allozymes. Hum Mol Genet 3:729-734.

127. Zhou, S. F., J. P. Liu, and B. Chowbay. 2009. Polymorphism of human cytochrome P450 enzymes and its clinical impact. Drug Metab Rev 41:89-295.

128. Collins, P., G. M. Rosano, P. M. Sarrel, L. Ulrich, S. Adamopoulos, C. M. Beale, J. G. McNeill, and P. A. Poole-Wilson. 1995. 17 beta-Estradiol attenuates acetylcholine-induced coronary arterial constriction in women but not men with coronary heart disease. Circulation 92:24-30.

129. Shea, A. K., and M. Steiner. 2008. Cigarette smoking during pregnancy. Nicotine Tob Res 10:267-278.

130. Albuquerque, C. A., K. R. Smith, C. Johnson, R. Chao, and R. Harding. 2004. Influence of maternal tobacco smoking during pregnancy on uterine, umbilical and fetal cerebral artery blood flows. Early Hum Dev 80:31-42.

131. Bruner, J. P., and I. Forouzan. 1991. Smoking and buccally administered nicotine. Acute effect on uterine and umbilical artery Doppler flow velocity waveforms. J Reprod Med 36:435-440.
132. Clark, K. E., and G. L. Irion. 1992. Fetal hemodynamic response to maternal intravenous nicotine administration. Am J Obstet Gynecol 167:1624-1631.
133. Adam, M. P., J. Chueh, Y. Y. El-Sayed, A. Stenzel, H. Vogel, D. D. Weaver, and H. E. Hoyme. 2005. Vascular-type disruptive defects in fetuses with homozygous alpha-thalassemia: report of two cases and review of the literature. Prenat Diagn 25:1088-1096.
134. Tredwell, S. J., D. Wilson, and M. A. Wilmink. 2001. Review of the effect of early amniocentesis on foot deformity in the neonate. J Pediatr Orthop 21:636-641.
135. Habek, D. 2007. Effects of smoking and fetal hypokinesia in early pregnancy. Arch Med Res 38:864-867.
136. Yasuda, H., and B. de Crombrugghe. 2009. Joint formation requires muscle formation and contraction. Dev Cell 16:625-626.
137. Mathews, T. J. 2001. Smoking during pregnancy in the 1990s. Natl Vital Stat Rep 49:1-14.
138. 2008. Smoking During Pregnancy. In Quick Reference: Fact Sheets. March of Dimes.
139. Webster WS, Abela D. 2007. The effect of hypoxia in development. Birth Defects Res C Embryo Today 81(3):215-228.

VITA

Amy Michelle Sommer was born in Orange, California on January 26th, 1986, the daughter of Diane Michelle Sommer and David Otto Sommer. After completing her work at Vista High School in 2004, she entered the University of California, Los Angeles. She received the degree of Bachelor of Science with a major in Microbiology, Immunology and Molecular Genetics from UCLA in June 2008. In August of 2008 she entered The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences in the program of Genetic Counseling.

Permanent address:
1615 Avenida Oceano
Oceanside, CA 92056