Comparative Genotype and Allele Frequency Analysis of CYP1A1 Polymorphisms (rs4646903 and rs1048943) in Bangladeshi Shipyard Workers with Occupational Heavy Metal Exposure

An Observational Comparative Genotype and Allele Frequency Study

by

MD. Abir Abdullah

22126005

Afifa Binte Rahman

22126078

Nahiyan Raisa Era

22326068

Neelanjana Rahman

21226004

A thesis submitted to the Department of Microbiology, School of Life Sciences in partial fulfillment of requirements for the degree of Bachelor of Science in Microbiology

School of Life Sciences

Brac University

August 2026

Declaration

It is hereby declared that:

1. The thesis submitted is our own original work while completing a degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

MD. Abir Abdullah

Afifa Binte Rahman

Nahiyan Raisa Era

Neelanjana Rahman

Approval

The thesis titled "Comparative Genotype and Allele Frequency Analysis of CYP1A1 Polymorphisms (rs4646903 and rs1048943) in Bangladeshi Shipyard Workers with Occupational Heavy Metal Exposure and Healthy Controls" submitted by

1. MD. Abir Abdullah (22126005)
2. Afifa Binte Rahman (22126078)
3. Nahiyan Raisa Era (22326068)
4. Neelanjana Rahman (21226004)

of Summer, 2026, has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Science in Microbiology on 13 August, 2026.

Examining Committee & Signature:

Supervisor:

Dr. M. Mahboob Hossain

Professor, Department of Microbiology, School of Life Sciences, Brac University

Co-Supervisor:

Dr. Md. Aminul Haque

Associate Professor, School of Pharmacy, Brac University

Chairperson:

Dr. Nadia Sultana Deen

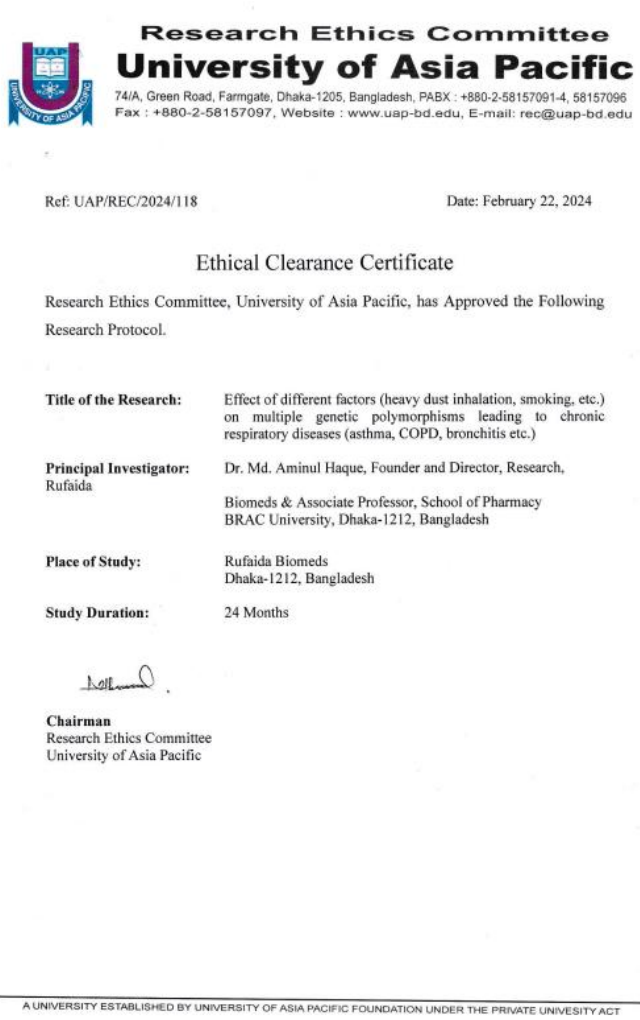
Associate Professor and Chairperson,

Department of Microbiology, School of Life Sciences, Brac University

Ethics Statement

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki. Ethical clearance was obtained from the University of Asia Pacific Research Ethics Committee, reference number UAP/REC/2024/118. Written informed consent was obtained from every participant before sample collection, participation was voluntary, and participants were free to withdraw at any stage. Peripheral venous blood was the only material collected, and no clinical intervention was performed. All samples were anonymised by numerical code; no personally identifying information appears in this thesis, and genotype results were not disclosed to employers or any third party.

Ethical Clearance Certificate:



Research Ethics Committee
University of Asia Pacific
74/A, Green Road, Farmgate, Dhaka-1205, Bangladesh, PABX : +880-2-58157091-4, 58157096
Fax : +880-2-58157097, Website : www.uap-bd.edu, E-mail: rec@uap-bd.edu

Ref: UAP/REC/2024/118 Date: February 22, 2024

Ethical Clearance Certificate

Research Ethics Committee, University of Asia Pacific, has Approved the Following Research Protocol.

Title of the Research: Effect of different factors (heavy dust inhalation, smoking, etc.) on multiple genetic polymorphisms leading to chronic respiratory diseases (asthma, COPD, bronchitis etc.)

Principal Investigator: Dr. Md. Aminul Haque, Founder and Director, Research, Rufaida
Biomeds & Associate Professor, School of Pharmacy
BRAC University, Dhaka-1212, Bangladesh

Place of Study: Rufaida Biomeds
Dhaka-1212, Bangladesh

Study Duration: 24 Months

Chairman
Research Ethics Committee
University of Asia Pacific

A UNIVERSITY ESTABLISHED BY UNIVERSITY OF ASIA PACIFIC FOUNDATION UNDER THE PRIVATE UNIVESITY ACT

Abstract

Workers in the ship-breaking and metal-scrapping yards of Bangladesh are chronically exposed to a mixture of heavy metals that includes the Group 1 human carcinogens arsenic, cadmium, and chromium. Susceptibility to the genotoxic consequences of such exposure is modulated by inherited variation in xenobiotic-metabolizing enzyme genes, among which CYP1A1 is central: it is induced through the aryl hydrocarbon receptor by both polycyclic aromatic hydrocarbons and several of these metals, and it bioactivates pro-carcinogens into DNA-reactive intermediates. The study determined and compared the first CYP1A1 genotype and allele frequency data for a Bangladeshi ship-breaking and metal-scrapping workforce. Allele frequencies of two functional CYP1A1 polymorphisms, rs4646903 and rs1048943, were compared between 55 workers with ten years of occupational heavy metal exposure and 55 healthy unexposed controls. Genomic DNA was extracted from peripheral blood by spin column and genotyped by two-tube allele-specific Amplification Refractory Mutation System PCR (ARMS-PCR) with agarose gel electrophoresis, using two allele-specific forward primers and one common reverse primer per locus. Allele distributions were compared by chi-square, and the association between genotype and group membership was estimated by multivariable logistic regression adjusted for age and sex under codominant, dominant, recessive, over-dominant, and additive models.

Allele distributions differed significantly between the two groups at both loci (rs4646903: OR 0.492, 95% CI 0.282-0.859, P = 0.0121; rs1048943: OR 0.446, 95% CI 0.253-0.786, P = 0.0048). At rs4646903 the G-containing genotypes were associated with higher odds of worker status (additive OR 2.855, P = 0.0048), whereas at rs1048943 the C-containing genotypes were associated with lower odds (additive OR 0.302, P = 0.0037). No worker carried the rs1048943 CC genotype, rendering the recessive and CC-versus-TT contrasts inestimable. Age, but not sex, was an independent predictor of group membership in every model.

Because both polymorphisms are germline variants that cannot be induced by adult exposure, and because no clinical outcome was measured, these findings are interpreted as associations with exposure-group membership within this sample.

Keywords: CYP1A1; rs4646903; rs1048943; single nucleotide polymorphism; ARMS-PCR; occupational exposure; heavy metals; ship-breaking; genotype frequency; Bangladesh

Dedication

While closing this chapter,
with gratitude to God,
we dedicate this work to our parents.
To everyone who has supported us throughout our journey,
to each other, for carrying this journey together,
to the shipyard workers who carry the weight of their work,
and to the victims of environmental disasters
whose stories remind us why this work matters.

Acknowledgement

We thank Dr. Nadia Sultana Deen, Associate Professor and Chairperson, School of Life Sciences, Brac University, for institutional support, and the laboratory officers and technical staff of the School of Life Sciences for their assistance with reagents, equipment and troubleshooting during the optimisation of the genotyping assay.

We express our sincere gratitude to our supervisor, Dr. M. Mahboob Hossain, Professor, School of Life Sciences, Brac University, for his continuous guidance, patience and critical insight throughout this work.

We are deeply grateful to our co-supervisor, Dr. Md. Aminul Haque, Associate Professor, School of Pharmacy, Brac University, both for his academic guidance and for facilitating our access to the laboratory of Rufaida Biomeds, where the greater part of the practical work of this study was carried out under the collaboration between Brac University and Rufaida Biomeds. We also thank Research Assistant Sharif Hossain for his hands-on support during sample processing and genotyping.

We are grateful to the workers of the shipyards and junkyards who took part in this study, and to the healthy volunteers who consented to participate. Without their willingness to give their time and a blood sample, this study could not have been carried out.

Finally, we thank our families for their support and encouragement throughout the course of this degree.

Table of Contents

List of Tables

List of Figures

List of Acronyms

AhR	Aryl hydrocarbon receptor
ARMS-PCR	Amplification Refractory Mutation System Polymerase Chain Reaction
ARNT	Aryl hydrocarbon receptor nuclear translocator
BPDE	Benzo[a]pyrene-7,8-diol-9,10-epoxide
CI	Confidence interval
CYP1A1	Cytochrome P450 family 1 subfamily A member 1
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleoside triphosphate
EDTA	Ethylenediaminetetraacetic acid
FDR	False discovery rate
HWE	Hardy-Weinberg equilibrium
IARC	International Agency for Research on Cancer
ILO	International Labour Organization
MAF	Minor allele frequency
NE	Not estimable
NTC	No-template control
OR	Odds ratio
PAH	Polycyclic aromatic hydrocarbon
PCR	Polymerase chain reaction
PPE	Personal protective equipment
RFLP	Restriction fragment length polymorphism
ROS	Reactive oxygen species
SNP	Single nucleotide polymorphism
SPSS	Statistical Package for the Social Sciences
XRE	Xenobiotic response element

Chapter 1

Introduction

1.1 Background

Bangladesh's ship-breaking industry is one of the world's most dangerous workplaces. Bangladesh is the leading ship-recycling country in the world, with around 200 large ocean-going vessels being recycled each year (UNCTAD, 2020; Hasan et al., 2024), and ship-breaking is one of the most hazardous jobs in the world, according to the International Labour Organization (Alam & Faruque, 2014). The industry is based in the Sitakunda coastal belt near Chattogram, but the same processes of thermal cutting of steel and welding, scraping of anti-corrosion coatings and manual dismantling of vessel and machinery parts are being done in other shipyards and metal-scrapping junkyards of the country, where toxic dust, metal fumes and chemical leachates are directly released into the working environment. Environmental surveys of the ship-breaking sites in Bangladesh have reported high levels of lead, cadmium, chromium, arsenic, nickel and manganese in soil and water, with water showing the highest level for the following order: Cr > Pb > As > Cd (Rahman et al., 2022). The International Agency for Research on Cancer (IARC) has classified cadmium, chromium and arsenic as Group 1 human carcinogens, which act by inducing the production of reactive oxygen species (ROS), oxidative damage to DNA and inhibition of DNA repair (Tchounwou et al., 2012; Balali-Mood et al., 2021). Biomonitoring of shipyard workforces elsewhere has verified that this work environment results in a measurable internal dose of combustion-derived carcinogens and of metals (Lee et al., 2003). Although the industry is large and the contamination is well documented, there is no systematic genetic or biological health monitoring of this workforce in Bangladesh.

1.2 Inter-individual Susceptibility and Single Nucleotide Polymorphisms

Biological damage is not always the same for workers in the same chemical environment. This inter-individual variability is significantly regulated by the inherited differences in the genes coding for xenobiotic-metabolizing enzymes, and polymorphisms in these genes have been found to affect the level of DNA damage observed in workers exposed to metals

(Pinto et al., 2025). Single nucleotide polymorphisms (SNPs), which are single base-pair differences in the genome that occur about once every 1,000 bases, are the most common type of human genetic variation and can affect the expression of genes, the catalytic activity of enzymes, and the stability of proteins and, hence, an individual's ability to metabolize environmental toxicants (Brookes, 1999).

1.3 The Aryl Hydrocarbon Receptor Pathway and CYP1A1

CYP1A1 (chromosome 15q22-q24) is a phase I monooxygenase that metabolizes polycyclic aromatic hydrocarbons (PAHs) to reactive metabolites that form covalent DNA adducts and is transcriptionally induced by the aryl hydrocarbon receptor (AhR), the cell's major sensor for environmental xenobiotics (Nebert et al., 2004; Androutsopoulos et al., 2009). Most tissues have little basal expression of the enzyme, and it is only when inducing ligands are present that the enzyme is functionally significant. This is directly relevant to the shipyard, where workers are co-exposed to PAH-generating thermal processes and to arsenic, cadmium and chromium, each of which has been shown to modulate CYP1A1 expression through the same AhR-dependent pathway (Elbekai & El-Kadi, 2004, 2007). The worker with the genotype that is more likely to express the enzyme or to have higher catalytic activity may thus experience a disproportionately high genotoxic response to the same external exposure.

1.4 The rs4646903 and rs1048943 Polymorphisms

Two functional CYP1A1 SNPs have been studied extensively as susceptibility loci: rs4646903 (m1, MspI; T3801C or T6235C; CYP1A1*2A) is a T-C transition in the non-coding 3' flanking. region that does not alter the protein sequence but creates an MspI restriction site and has been associated with increased mRNA inducibility and elevated microsomal activity (Zeng et al., 2016); and rs1048943 (m2; A2455G or A4889G; CYP1A1*2C; Ile462Val) is a substitution of valine for isoleucine at codon 462 in the haem-binding domain, which is reported to metabolize PAH substrates with approximately two-fold higher catalytic efficiency (Mutar & Oraibi, 2021). Both variants have been linked to various malignancies in meta-analyses in Asian populations (Zeng et al., 2016; Zhu et al., 2016; Barek et al., 2023). Both were genotyped, as the two loci control complementary aspects of the same enzyme, the amount of enzyme produced and the rate at which it acts.

1.5 Rationale of the present study

There is one conceptual issue that dictates what such a study can say. Both variants are germline SNPs, which means they are passed on at conception and found in all nucleated cells and are not affected by any exposure that occurs during adult working life. Somatic damage, DNA adducts, oxidative base lesions, strand breaks and chromosomal aberrations are caused by heavy metals and PAHs; they are accumulated in exposed tissue but do not change one inherited allele into another; a change in the frequency of alleles in the germline of a population requires selection acting over generations, not a single working lifetime. It is therefore not possible to hypothesise, nor test, that these polymorphisms were caused by occupational exposure in the present study. It looks at the other way around: the inherited metabolic background that each worker brings to the exposure. Since CYP1A1 is mostly silent in the absence of inducing ligands, a high-inducibility or high-activity genotype is of little importance in an unexposed individual but becomes important in a chronically exposed individual, in which the rate of activation of the pro-carcinogens is determined. Genotype and exposure are thus not mutually exclusive explanations, but rather components of a single risk.

1.6 Interpretive Framework and Scope of Inference

It is therefore important to note that any difference in genotype or allele frequencies between an exposed and unexposed group cannot be attributed to the exposure alone, and a defined interpretive framework is needed before any such difference is discussed. Four explanations are acceptable: sampling variation, which may be large if the number of individuals in the genotype classes is small; population stratification, such as ship-breaking and metal-scrapping, recruits migrant workers from specific districts, and the two recruitment pools may have different ancestral backgrounds; recruitment and occupational selection, including the healthy-worker survivor effect, whereby individuals who can tolerate the yard environment remain employed, while those who become symptomatic leave; and technical factors, such as allele dropout or inadequate discrimination of alleles, which are guarded against by Hardy-Weinberg equilibrium testing and repeat genotyping. Exposure-induced alteration of the germline is not one of the acceptable explanations, and

any odds ratio reported in this study should be interpreted as a measure of association between genotype and group membership, rather than as an estimate of risk.

Thus, the unexposed control group is not a pre-exposure baseline but rather a locally recruited reference distribution, as allele frequencies of CYP1A1 vary significantly across South Asian, East Asian, European, and Latin American populations, and estimates from other regions cannot be validly transferred (Garte et al., 2001; Alvarado et al., 2021). It also provides an internal quality control: testing Hardy-Weinberg equilibrium checks the genotyping and the sampling before any comparison is interpreted. Both groups had age and sex recorded, and these were used as covariates in the multivariable analysis; hence, the genotype estimates reported are adjusted for these two variables. The worker group was selected based on at least 10 years of occupational exposure, but individual exposure measures, exact duration of exposure, job category, use of personal protective equipment, and biological exposure markers (e.g., blood lead or urinary cadmium) or lifestyle factors (e.g., tobacco or betel quid use) were not collected. The study is therefore unable to control for these factors or to measure gene-environment interaction. However, it is important to note that none of these unmeasured variables can have influenced the genotype itself, as genotype is assigned at meiosis, prior to any subsequent behaviour or exposure, and thus would be indispensable confounders in a study relating genotype to disease but would not bias the estimation of genotype frequency. The first suggestion for further work is their collection. It was also expected that a rare genotype class may be observed at zero frequency, and the corresponding regression contrast will be inestimable due to complete separation and will be reported as such rather than forced to a numerical value, with 55 participants per group.

Similarly, there was no clinical outcome assessed. No one in either group had a diagnosis of lung cancer or any other malignancy; no incident disease was determined, and there was no longitudinal follow-up. The cancer literature cited above is background to the study and does not constitute a finding of the study but rather the functional plausibility of the two loci. None of the estimates reported here are estimates of cancer risk, and the proper interpretation of the work is descriptive and hypothesis-generating, not etiologic.

1.7 Research Gap

Data of this sort has definite value within these limits. They represent primary population data for an occupational cohort that has never been genotyped at these loci; they provide the control allele frequencies for any adequately powered future case-control or biomarker study in this population; they provide an in-house, ARMS-PCR assay that can be easily transferred to occupational screening in a resource-limited setting; and they quantify the proportion of an at-risk workforce carrying the high-activity metabolic profile. There are no published studies that have characterised the frequency of CYP1A1 rs4646903 and rs1048943 in a defined Bangladeshi occupational cohort with known heavy metal exposure, nor compared them with a locally recruited unexposed cohort genotyped on the same platform in the same laboratory, although genotype data for general Bengali population samples are available in international variant databases. Characterisation of this distribution, thus, is not only a scientific issue but also an occupational health priority.

1.8 Aim and Objectives of the Study

The present study was conducted to find out and compare the genotype and allele frequencies of two CYP1A1 polymorphisms (rs4646903, m1, and rs1048943, m2; Ile462Val) among 55 Bangladeshi shipyard and junkyard workers with chronic occupational heavy metal exposure and 55 healthy unexposed controls using allele-specific Amplification Refractory Mutation System PCR (ARMS-PCR) with agarose gel electrophoresis and to investigate the association between genotype and group membership after adjusting for age and sex. The allele-specific primers were designed on the strand opposite to the coding sequence, and so the genotypes are reported using either A/G for rs4646903 (A = reference T allele, G = variant C allele) or T/C for rs1048943 (T = reference A allele, C = variant G allele), both of which are equivalent.

The specific objectives were:

1. To isolate the genomic DNA from peripheral venous blood of both groups and assess its amplifiable quality.
2. To develop and optimise a two-tube allele-specific ARMS-PCR assay for each locus.
3. To calculate genotype and allele frequencies and the minor allele frequency for each SNP individually for the exposed and control groups.

4. To use Hardy-Weinberg equilibrium as an internal test of genotyping and sampling.
5. To compare the distribution of alleles between the two groups using chi-squared and to present ORs and 95% CI.
6. To estimate the association between genotype and group membership using multivariable logistic regression with worker status as the dependent variable and age and sex as covariates under the codominant, dominant, recessive, overdominant and additive models and to report a zero-count cell in the contingency table as "inestimable".
7. To interpret the results in the context of the design outlined in Section 1.6, to clearly outline the boundaries of the inferences that can be drawn from the design, and to provide a foundation for future, sufficiently powered, and fully covariate-adjusted investigation.

1.9 Novelty of the Present Study

Initial report of frequencies of CYP1A1 genotype and allele, CYP1A1 rs4646903 and CYP1A1 rs1048943, and frequencies of those in a high-exposure workforce that has never been genetically characterised: shipbreaking and metal scrapping.

Occupation: Heavy metal occupation CYP1A1; variation is positional, and the functional impact of genotype is enhanced by AhR-mediated effects of metals instead of being independent of them.

Provides frequencies on controls provided, necessary to conduct future case-control and genotoxicity-biomarker experiments in this group, and adds data on the South Asians to a literature that is largely European and East Asian biomarker samples.

Provides a clear interpretive map on the issue of genotype comparisons between exposure-defined groups, identifying what such a comparison can reveal and what it cannot.

Chapter 2

Materials and Methods

2.1 Approach, Participants, and Ethical Considerations

This study is a case-control type research study conducted with 55 Bangladeshi individuals who were exposed to heavy metals in various shipyards and junkyards in Dhaka for more than 10 years. This included 55 volunteers (cases) and 55 age- and gender-matched healthy controls.

2.2 Collection of Samples

Around 5 mL of whole blood was drawn from each participant using EDTA tubes through venipuncture, provided by Rufaida Biomeds. Whole blood samples were frozen at -20 degC before extraction of genomic DNA.

2.3 Procedures Carried out in the Laboratory

The genomic DNA was extracted using the spin-column extraction method, and the methodology followed in the laboratory is as follows: Peripheral blood was collected in five mL tubes by venipuncture and transported to 15 mL tubes containing ethylenediaminetetraacetic acid (EDTA)-Na2 as an anticoagulant. The blood samples were collected, and the genomic DNA was extracted with FavorPreptm Blood Genomic DNA Extraction Kit (Favorgen Biotech Corporation, Taiwan) according to the manufacturer's protocol. The isolated DNA was also stored at -20 degrees Celsius for subsequent treatments.

Genotyping of CYP1A1 rs4646903 and CYP1A1 rs1048943 was carried out by the technique, amplification and refractory mutation system polymerase chain reaction (ARMS-PCR). The list of the external and internal forward and reverse primers was taken from the study (Yousefian et al., 2022) to recognize the homozygous and heterozygous alleles.

The PCR reaction mixture (10 µL) was prepared with the following (Table 2.1):

Table 2.1

Composition of the ARMS-PCR reaction mixture

Components	Volume
Master mix (2X)	5 µL
forward primer (1µM)	1 µL
reverse primer (1 µM)	1 µL
ddH2O	2 µL
Template DNA	1 µL

The primers employed for amplifying the CYP1A1(rs4646903) polymorphism were (Table 2.2):

Table 2.2

Allele-specific and common primers for CYP1A1 rs4646903

Primer name	Sequence
CYP1A1rs4646903F(G)	5'-GAGAATCGTGTGAGCCCG-3'
CYP1A1rs4646903F(A)	5'-GAGAAATCGTGTGAGCCCA-3'
CYP1A1rs4646903R	5'-CAGAAATCCTATAGGTGG-3'

The primers employed for amplifying the CYP1A1(rs1048943) polymorphism were (Table 2.3):

Table 2.3

Allele-specific and common primers for CYP1A1 rs1048943

Primer name	Sequence
CYP1A1rs1048943F(T)	5'-GAAGTGTATCGGTGAGACAT-3'
CYP1A1rs1048943F(C)	5'-GAAGTGTATCGGTGAGACAC-3'
CYP1A1rs1048943R	5'-ACCAGACCAGGTAGACAGAG-3'

PCR amplification was performed under the following conditions:

For CYP1A1(rs4646903):

- The initial denaturation step is carried out at 95degC for 5 minutes.
- The sample is then denatured at the same temperature for 40 seconds, annealed for 50 seconds at 50 degrees Celsius, and extended for 30 seconds at 72 degrees Celsius for 30 cycles.
- Finally, the last extension step was carried out at 72 degrees Celsius for 5 min.

Following electrophoresis of the ARMS PCR result, two types of bands can be detected in the gel:

Based on the following band patterns, the CYP1A1(rs4646903) and CYP1A1(rs1048943) polymorphisms were identified:

The T and C alleles of CYP1A1 (rs1048943) were evaluated by visualizing the band on the agarose gel at 220 bp for both the cases and controls and the amplicons were confirmed by the presence of the amplification products on 2 % agarose gel by electrophoresis along with 50 bp DNA ladder.

The A and G alleles of the CYP1A1(rs4646903) gene were evaluated by observing the band on the agarose gel at 269 bp, and the amplicons were validated by amplification products on 2% agarose gel by electrophoresis using 100 bp DNA ladder.

2.4 Quantitative Analysis of Data

Data were analyzed using IBM SPSS Statistics version 25.0 (SPSS Inc., Chicago, IL, USA). Genotype and Allele frequencies were expressed as counts and percentages for each

group, and allele counts were calculated from genotype counts assuming that each individual contributes two alleles (110 alleles per group). The minor allele frequency was determined as the frequency of the less common allele in the control group.

The observed number of each genotype was compared with the number of each genotype predicted by the Hardy-Weinberg equilibrium ($p^2 + 2pq + q^2 = 1$), where p and q are the observed allele frequencies, by a chi-square goodness-of-fit test. In a study like this, departure from equilibrium is initially thought of as a potential sign of genotyping error or non-random sampling, and then a population-genetic phenomenon.

The chi-square test was used to test differences in the distribution of alleles between the exposed and control groups, and the strength of association was expressed as an odds ratio with a 95% confidence interval on the 2 x 2 table of allele counts. Multivariate binary logistic regression was then used to test the association between genotype and group membership, using worker status as the dependent variable (worker = 1, control = 0) and age, sex, and genotype as covariates. Genotype was coded based on five genetic models, which are different assumptions about the mode of inheritance:

- **Codominant:** Each variant genotype is tested individually against the reference homozygote, as well as an overall test with 2 degrees of freedom.
- **Dominant:** Heterozygotes and variant homozygotes are grouped and compared to the reference homozygote, with one variant copy being enough.
- **Recessive:** Variant homozygotes compared to heterozygotes and reference homozygotes pooled, where two copies of the variant are needed.
- **Over-dominant:** Heterozygotes vs both homozygote groups combined, to test for any effect due to heterozygosity per se.
- **Additive (log-additive):** Genotype coded 0, 1 or 2, estimating a per-allele effect in a single degree of freedom.

All five models were fitted, rather than a single model, as the mode of inheritance at these loci is not known and reporting only the model that gave the smallest P value would be selective reporting. Contrasts that could not be estimated due to complete separation in the logistic model are reported as not estimable and are not assigned a value. Two-sided P

values < 0.05 were considered significant. Since fourteen tests were conducted at the two loci (genotype-distribution and allelic association tests plus five genetic models for each SNP), the results of the Bonferroni procedure (family-wise error rate) and Benjamini-Hochberg procedure (false discovery rate) are presented.

Chapter 3

Results

3.1 Sample Characteristics

110 participants were analysed: 55 occupationally exposed shipyard workers and 55 healthy controls. There were no missing genotype data for either polymorphism (N = 110, 0 missing). All participants had age and sex recorded and these were used as covariates in the adjusted analyses (Section 3.6).

3.2 ARMS-PCR Amplification and Identification of Genotype

Both CYP1A1 polymorphisms were genotyped using Amplification Refractory Mutation System Polymerase Chain Reaction (ARMS-PCR). For each polymorphism, two allele-specific forward primers, which differed at the 3′ terminal nucleotide, were used in combination with a common reverse primer, so that each DNA sample was amplified in two parallel allele-specific reactions (see Section 2.3). Amplification products were separated by agarose gel electrophoresis and visualised by UV light, using commercial DNA molecular-weight markers on each gel to estimate the size of the fragments (100 bp DNA ladder for rs4646903 and 50 bp DNA ladder for rs1048943). Genotypes were determined by the pattern of allele-specific amplification: amplification in one reaction was considered to be homozygous for the allele, and amplification in both reactions was considered to be heterozygous for the allele. The allele-specific product for rs4646903 was detected at ~269 bp and representative AA, AG, and GG amplification patterns are displayed in Figure 3.1. The allele-specific product for rs1048943 was seen at ~220 bp, with representative TT and TC amplification patterns displayed in Figure 3.2. There were

no genotyping failures for both polymorphisms in all 110 samples (55 workers and 55 controls).

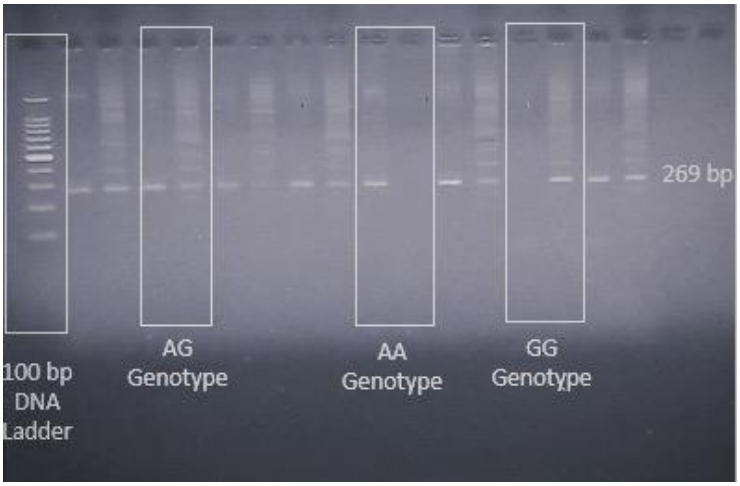


Figure 3.1.

Representative ARMS-PCR amplification patterns on agarose gel for the CYP1A1 rs4646903. Lane M, 100 bp DNA molecular-weight ladder. Boxed regions are representative samples with AG, AA and GG genotype. Allele specific amplification products were detected at ~269 bp (as shown on the right of the gel). Each genotype panel consists of the two allele-specific reactions for one sample.

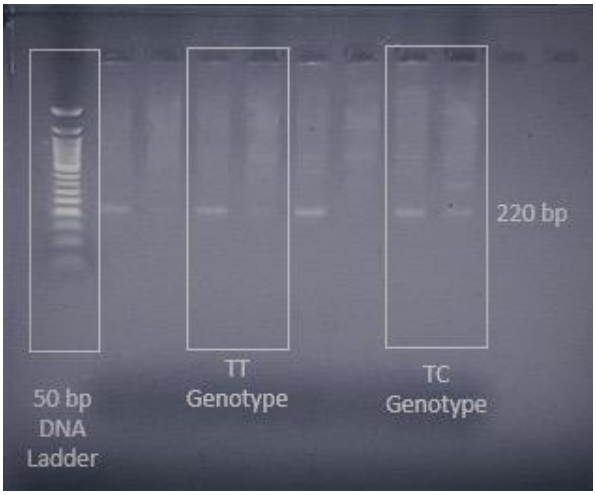


Figure 3.2.

Representative ARMS-PCR amplification patterns on agarose gel for the CYP1A1 rs1048943. Lane M, 50 bp DNA molecular-weight ladder. Boxed areas represent samples

with the TT and TC genotype as labelled. Allele-specific amplification products were seen at about 220 bp as shown on the right of the gel. Each genotype panel consists of the two allele-specific reactions for one sample. The CC genotype was not found on this gel.

The primary genotype data set for this study was the genotype assigned by ARMS-PCR for all 110 participants. These genotype calls were used to calculate the genotype and allele frequencies, minor allele frequency, and Hardy-Weinberg equilibrium in controls, and served as the input for all subsequent association analyses, which included the 3 × 2 genotype-distribution chi-square test, allelic association test, five genetic models (codominant, dominant, recessive, over-dominant, and additive), and age and sex-adjusted logistic regression models. It is important to note that ARMS-PCR and agarose gel electrophoresis only provide information on genotype identity; the question of whether genotype and allele distributions differ between occupationally exposed workers and healthy controls is only answered by the statistical analyses described in the sections below.

3.3 CYP1A1 rs4646903

Genotype distribution in controls (n = 55): AA 10 (18.18%), AG 30 (54.55%), GG 15 (27.27%); in workers (n = 55): AA 2 (3.64%), AG 28 (50.91%), GG 25 (45.45%). Allele frequencies: controls A 45.45% / G 54.55%; workers A 29.09% / G 70.91%. Minor allele frequency in controls = 0.4545 (allele A). Control genotypes were not deviated from Hardy-Weinberg equilibrium ($\chi^2 = 0.550$, df = 2, P = 0.759). The 3×2 genotype-distribution test was significant ($\chi^2 = 7.902$, df = 2, P = 0.0192; minimum expected count = 6.00, assumption met; Cramér's V = 0.268), shown in Table 3.1. Allelic association was statistically significant ($\chi^2 = 6.299$, df = 1, P = 0.0121; OR for A vs G = 0.492, 95% CI 0.282-0.859). The codominant overall test, dominant, recessive and additive models were significant, while the AG-vs-AA codominant contrast and over-dominant model were not significant, as shown in Table 3.2.

Table 3.1

Genotype distribution of CYP1A1 rs4646903, and 3×2 genotype-distribution test

Genotype	Control, n (%)	Worker, n (%)	Expected (Ctrl/Wkr)
AA	10 (18.18)	2 (3.64)	6.0 / 6.0
AG	30 (54.55)	28 (50.91)	29.0 / 29.0
GG	15 (27.27)	25 (45.45)	20.0 / 20.0
Total	55 (100)	55 (100)	55 / 55

Note. Pearson $\chi^2 = 7.902$, df = 2, P = 0.0192. Cramér's V = 0.268. Allele A = 50 (45.45%) ctrl / 32 (29.09%) wkr; allele G = 60 (54.55%) ctrl / 78 (70.91%) wkr. MAF (ctrl) = 0.4545. HWE (ctrl): $\chi^2 = 0.550$, df = 2, P = 0.759.

Table 3.2

Adjusted genetic models for CYP1A1 rs4646903 (reference genotype AA)

Model	Comparison	OR	95% CI	P
Codominant (overall)	AG, GG vs AA	-	-	0.0223
Codominant	AG vs AA	4.679	0.706-31.019	0.1099
Codominant	GG vs AA	11.421	1.622-80.420	0.0145
Dominant	AG+GG vs AA	6.529	1.062-40.127	0.0428
Recessive	GG vs AA+AG	2.966	1.232-7.138	0.0153
Over-dominant	AG vs AA+GG	0.637	0.278-1.458	0.2859
Additive	per G allele	2.855	1.377-5.920	0.0048

Note. Adjusted for age and sex.

3.4 CYP1A1 rs1048943

Genotype distribution in controls: TT 13 (23.64%), TC 35 (63.64%), CC 7 (12.73%); in workers: TT 26 (47.27%), TC 29 (52.73%), CC 0 (0.00%). Allele frequencies: controls T

55.45%/C 44.55%; workers T 73.64% / C 26.36%. MAF (controls) = 0.4455 (allele C). Control genotypes did not deviate from HWE ($\chi^2 = 4.561$, df = 2, P = 0.1022). The 3×2 genotype-distribution test was significant; because two cells had expected counts below 5 (minimum = 3.50), exact tests were used: exact P = 0.0017 (asymptotic P = 0.0026; Fisher's exact = 12.179, exact P = 0.0015; Cramér's V = 0.329) - Table 3.3. Allelic association was significant ($\chi^2 = 7.945$, df = 1, P = 0.0048; OR for C vs T = 0.446, 95% CI 0.253-0.786). Adjusted genetic models are in Table 3.4. No worker carried CC genotype and the recessive model and CC-vs-TT codominant contrast were not co. The dominant and additive and the TC-vs-TT codominant models were significant, while the overall codominant and over-dominant models were not.

Table 3.3

Genotype distribution of CYP1A1 rs1048943, and 3×2 genotype-distribution test

Genotype	Control, n (%)	Worker, n (%)	Expected (Ctrl/Wkr)
TT	13 (23.64)	26 (47.27)	19.5 / 19.5
TC	35 (63.64)	29 (52.73)	32.0 / 32.0
CC	7 (12.73)	0 (0.00)	3.5 / 3.5
Total	55 (100)	55 (100)	55 / 55

Note. Pearson $\chi^2 = 11.896$, df = 2, asymptotic P = 0.0026, exact P = 0.0017 (two cells had expected count < 5). Fisher's exact = 12.179, exact P = 0.0015. Cramér's V = 0.329. Allele T = 61 (55.45%) ctrl / 81 (73.64%) wkr; allele C = 49 (44.55%) ctrl / 29 (26.36%) wkr. MAF (ctrl) = 0.4455. HWE (ctrl): $\chi^2 = 4.561$, df = 2, P = 0.1022.

Table 3.4

Adjusted genetic models for CYP1A1 rs1048943 (reference genotype TT)

Model	Comparison	OR	95% CI	P
Codominant (overall)	TC, CC vs TT	-	-	0.0776
Codominant	TC vs TT	0.363	0.151-0.874	0.0238

Codominant	CC vs TT	NE	NE	0.9989[a]
Dominant	TC+CC vs TT	0.317	0.132-0.759	0.0100
Recessive	CC vs TT+TC	NE	NE	0.9989[a]
Over-dominant	TC vs TT+CC	0.495	0.215-1.139	0.0984
Additive	per C allele	0.302	0.135-0.678	0.0037

Note. Adjusted for age and sex. NE = not estimable (zero CC-genotype workers). [a]Artefact of model non-convergence; not a null finding.

3.5 Comparison Between SNPs

Genotypes which contain G at rs4646903 were associated with higher odds of being a worker (allelic, dominant, recessive, additive, GG-vs-AA codominant); genotypes which contain C at rs1048943 were associated with lower odds (allelic, dominant, additive, TC-vs-TT codominant). Both over-dominant models and the rs4646903 AG-vs-AA contrast were non-significant (Table 3.5).

Table 3.5

Comparison of adjusted genetic-model results for rs4646903 and rs1048943

Analysis	rs4646903 (ref. AA)	rs1048943 (ref. TT)
Genotype-distribution	χ^2=7.902, df=2, P=0.0192	χ^2=11.896, df=2, exact P=0.0017
Allelic association	OR 0.492 (0.282-0.859), P=0.0121	OR 0.446 (0.253-0.786), P=0.0048
Codominant (overall)	P=0.0223	P=0.0776
Codom. - het vs ref	OR 4.679 (0.706-31.019), P=0.1099	OR 0.363 (0.151-0.874), P=0.0238
Codom. - var homo vs ref	OR 11.421 (1.622-80.420), P=0.0145	NE

Dominant	OR 6.529 (1.062-40.127), P=0.0428	OR 0.317 (0.132-0.759), P=0.0100
Recessive	OR 2.966 (1.232-7.138), P=0.0153	NE
Over-dominant	OR 0.637 (0.278-1.458), P=0.2859	OR 0.495 (0.215-1.139), P=0.0984
Additive	OR 2.855 (1.377-5.920), P=0.0048	OR 0.302 (0.135-0.678), P=0.0037

3.6 Adjusted Logistic Regression

For each genetic model, a logistic regression was performed with worker status (worker = 1, control = 0) as an outcome and age, sex and genotype as covariates. Age was significant in all ten models (rs4646903: Adjusted OR 1.084-1.096/year, P = 0.0008-0.0016; rs1048943: OR 1.074-1.088/year, P = 0.0012-0.0054). Sex (female = reference) was not significant in any model (rs4646903: OR 2.92-3.44; rs1048943: OR 2.55-3.79; all 95% CI crossed 1). The omnibus model test was significant in all ten models ($P < 0.001$), and represents the combined effect of age, sex and genotype (not the effect of the genotype term alone, which is reported separately above).

3.7 Multiple-Testing Correction

Fourteen tests were evaluated (genotype-distribution and allelic tests, and five genetic models per SNP, including the codominant model as the overall test). Bonferroni threshold = 0.05/14 = 0.00357. The only test that survived was the rs1048943 genotype-distribution test (exact P = 0.0017). Benjamini-Hochberg FDR (q = 0.05) resulted in 9 of 13 rankable tests being significant (Table 3.6).

Table 3.6

Multiple-testing correction (m = 14)

SNP	Test	Raw P	Bonf. adj. P	Bonf. sig.	FDR sig.
943	Genotype-distribution (exact)	0.0017	0.0232	Yes	Yes
943	Additive	0.0037	0.0518	No	Yes
903	Additive	0.0048	0.0675	No	Yes
943	Allelic	0.0048	0.0675	No	Yes
943	Dominant	0.0100	0.1397	No	Yes
903	Allelic	0.0121	0.1691	No	Yes
903	Recessive	0.0153	0.2136	No	Yes
903	Genotype-distribution	0.0192	0.2693	No	Yes
903	Codominant (overall)	0.0223	0.3123	No	Yes
903	Dominant	0.0428	0.5999	No	No
943	Codominant (overall)	0.0776	1.0000	No	No
943	Over-dominant	0.0984	1.0000	No	No
903	Over-dominant	0.2859	1.0000	No	No
943	Recessive	NE	-	-	N/A

903 = rs4646903; 943 = rs1048943. FDR = Benjamini-Hochberg false discovery rate.

Integrated Interpretation

Direction and support: The genotype-distribution test, allelic test, and 4 genetic models showed that the GG genotype/G allele were more common among workers at rs4646903. The TT genotype/ T allele was more common among workers (or equivalently, the C-containing genotypes were more common among controls) and was supported by the genotype-distribution test, allelic test, and 3 estimable models at rs1048943. The two loci are thus negatively associated.

Multiple testing: The rs1048943 genotype-distribution test was the only test that passed strict Bonferroni correction (m = 14); 9 of 13 rankable findings passed FDR. When the two genotype-distribution tests were added to the correction family, the single test that survived Bonferroni at m = 12, it was the rs1048943 additive model; did not change, only the threshold.

Model consistency: The association seems to be focused on the GG genotype for rs4646903 (AG vs. AA and over-dominant were not significant). The recessive model and the CC-vs-TT contrast were non-estimable for rs1048943, presumably because there were no CC among workers, which is why the codominant model omnibus test was not significant, even though the TC-vs-TT component was.

Small/zero cells: The rs4646903 AA reference genotype (n = 12 total, 2 workers) produced wide CIs (e.g., 1.622-80.420). Two comparisons were non-estimable due to zero CC-genotype workers at rs1048943, and exact testing was performed for the genotype-distribution test (confirmed, not overturned, by the exact result).

After adjustment: Genotype was significant in rs4646903 dominant, recessive, additive, and GG-vs-AA codominant models, and in rs1048943 dominant, additive, and TC-vs-TT codominant models, suggesting that the associations are independent of these two covariates, although causation cannot be excluded and unmeasured confounding cannot be ruled out.

Chapter 4

Discussion

4.1 Main Findings

The genotype and allele distributions of both CYP1A1 polymorphisms were significantly different between the 55 shipyard workers and 55 controls (rs4646903 genotype distribution P = 0.0192; rs1048943 exact P = 0.0017). Hardy-Weinberg equilibrium was met in control genotypes. Worker status was consistently associated with age, but not with sex. Only one out of 14 tests remained significant after Bonferroni correction. The two loci

are associated in opposite directions and these are associations with exposure-group membership not disease in a design where genotype cannot be a consequence of exposure.

4.2 rs4646903

The G allele and GG genotype were consistently linked to increased odds of being a worker, and was not related to heterozygosity. The small size of the AA reference group (2 workers) resulted in wide confidence intervals and the recessive model was not significant by Fisher's exact test, unadjusted (P = 0.0738) but was significantly adjusted (P = 0.0153) showing sensitivity to analytic approach. The higher variant frequency among workers could be due to a difference in the background population between recruitment pools, chance (due to low counts), or an unmeasured factor other than the causal effect of exposure on genotype.

4.3 rs1048943

The C allele / C containing genotypes were linked to decreased odds of being a worker. The genotype-distribution test, which was confirmed by exact testing, was the only Bonferroni-surviving result in the study, but this is an omnibus test and does not itself identify which genotype is responsible. This is because no worker had CC and thus the recessive model and CC-vs-TT contrast were non-estimable, and not because there was no association for that genotype.

4.4 Discordant Direction of the Two Loci

The most impressive aspect of the results is that the two polymorphisms are in opposite directions in the same sample. The variant carrying genotypes were enriched among workers at rs4646903 and depleted at rs1048943. This is hard to explain by any one biological explanation, for two reasons.

Firstly, the two loci are not independent, but are in the same gene. Hayashi et al. (1991) showed that the MspI polymorphism in the 3′-flanking region is genetically linked to the amino acid substitution in the haem-binding region of exon 7, thus defining the haplotype structure: m1 alone (CYP1A1*2A), m1 in combination with m2 (2B), and m2 alone (2C). If the two variants are in positive LD, enrichment of one variant allele in a group should be accompanied by enrichment, not depletion, of the other variant allele. The opposite

directions observed, therefore, indicate either that the linkage disequilibrium between the two loci is weak in this population, which was not tested here, or that at least one of the two associations is not a true difference in underlying allele frequency. Second, both variants are reported in the functional literature to increase, not decrease, CYP1A1 activity or inducibility. One might expect the associations to shift together if they were due to a common underlying process related to metabolic phenotype. These differences are more likely to reflect chance or differences between the two study groups or technical variation in genotype testing than a clear biological effect. This discordance is presented here as a substantive observation rather than being obscured, because it is one of the most compelling internal reasons for considering the findings to be exploratory.

4.5 Comparison with Previous Studies

4.5.1 Occupational studies of CYP1A1 and PAH exposure

There are no previous studies that compared CYP1A1 genotype frequencies of occupationally exposed workers with unexposed controls using the design adopted here for either of the two polymorphisms (rs4646903 or rs1048943). In the large occupational literature on these polymorphisms, the genotype is not used as a variable to differentiate between exposed and unexposed people, but rather as a modifier of the association between exposure and a biological measurement. That distinction is significant and it sets the boundaries for the current comparison and the direction of future research.

The most directly relevant body of work is that of workers exposed to PAHs, primarily in coke ovens and other combustion industries, in whom internal dose can be measured by urinary 1-hydroxypyrene and genotoxic effect by measurement of DNA adducts. Alexandrie et al. (2000) found that polymorphisms of CYP1A1 and GSTM1 influenced urinary 1-hydroxypyrene levels after PAH exposure, indicating that the genotype does influence the internal handling of these compounds in humans as well as in experimental systems. In coke-oven workers, PAH-DNA adduct levels have been reported to be higher in individuals with the exon 7 Ile/Val variant than in those with the reference Ile/Ile genotype, and this effect was observed only in the higher exposure group. That last observation is the critical one for the current study: the genotype had a detectable effect

only when exposure was sufficiently high to be important, that is, the gene-environment interaction posited in Section 1.5.

The present study shares the population of interest, namely, healthy workers under chronic occupational exposure to combustion products and metals, and the genotypes studied. It has no internal dose or genotoxic effect. It thus sets up the genotype distribution of such a workforce without being able to show that the genotype had any consequence within it. The present frequencies, read in conjunction with the coke-oven literature, take on their significance as a necessary preliminary to the knowledge of what difference it makes, which would be the second step of a programme of work. The differences observed suggest that there is a relationship between genetic variation of CYP1A1 and occupational group status in the study population. Furthermore, the two studies differed in the nature of the occupational exposure, which makes it impossible to compare the genetic effects of the exposures. However, the uniformity of the evidence for a possible association between genetic variability in CYP1A1 and biological responses to occupational exposures suggests that the study of these polymorphisms in occupationally exposed populations is of interest.

4.5.2 Comparing with a Japanese Population Study

The results of the present study can be compared with the results of a Japanese case-control study by Kiyohara et al. (2012), which examined the relationship between the CYP1A1 polymorphisms and the susceptibility to lung cancer. The study involved 462 lung cancer patients and 379 controls, and focused on polymorphisms of CYP1A1 (rs4646903, rs1048943) and other genes related to carcinogen metabolism and DNA repair. Both CYP1A1 polymorphisms were significantly associated with lung cancer risk, with reported odds ratios of 1.72 (95% CI: 1.25-2.38) for rs4646903 and 1.40 (95% CI: 1.02-1.92) for rs1048943. The authors also found that certain combinations of potentially at-risk genotypes of CYP1A1 rs4646903 and other genes were linked to significantly higher risk of lung cancer, indicating that interactions between multiple genes may play a role in susceptibility. Biological relevance of CYP1A1 in lung cancer is associated with its role in the metabolism of carcinogens. Cigarette smoke is a major risk factor for lung cancer and the carcinogenic compounds in tobacco smoke are metabolically activated and detoxified by several enzymatic pathways. Variations in these pathways may affect the way that a

person responds to exposure to a carcinogen and thus to the damage to cells and the development of cancer. In this context, Kiyohara et al. (2012) studied genes that are associated with carcinogen metabolism, such as the gene for CYP1A1. Their results thus confirm the possible involvement of genetic polymorphisms in carcinogen metabolism and DNA-repair genes in lung cancer susceptibility.

Likewise, the current study revealed significant association for both polymorphisms with worker status; both SNPs were significantly associated, but in opposite directions and with different interpretations.

The two studies share the same finding that both of the genetic polymorphisms of CYP1A1, rs4646903 and rs1048943, were statistically significantly associated with their respective studies, suggesting that genetic variation of CYP1A1 may play a role in the susceptibility or biological response of individuals. CYP1A1 is involved in the metabolism of environmental carcinogens, especially polycyclic aromatic hydrocarbons, and genetic polymorphisms in this gene may account for inter-individual variability in xenobiotic metabolism. The Japanese study also found that the risk of lung cancer was significantly higher when multiple at-risk genotypes were combined, such as the combination of CYP1A1 rs4646903, indicating that the effects of the CYP1A1 variants may be compounded by other genetic factors.

Therefore, the results of the two studies should not be interpreted as indicating the same result. Kiyohara et al. focused on lung cancer susceptibility, while the current study focused on the difference in the distribution of the CYP1A1 genotype and alleles between occupationally exposed shipyard workers and healthy controls. Thus, the important associations found in the current study should not be interpreted as evidence that the variants identified increase or decrease the risk of lung cancer. Instead, they suggest that the frequency of these inherited genetic variants was different between the two study groups. In addition, the Japanese study doesn't prove that the polymorphisms were caused by occupational exposure. Likewise, the present study cannot conclude that the rs4646903 or rs1048943 variants were caused by exposure to heavy metals, since these are germline polymorphisms. The observed associations could be due to differences in genetic susceptibility or population characteristics and not necessarily due to exposure-induced genetic changes. Further, the two studies differ in terms of sample size, population

background, exposure characteristics, and study outcome, which should be taken into account when comparing the two studies.

Overall, the results of the present study support the previous evidence that the polymorphisms of the CYP1A1 gene (rs4646903 and rs1048943) could be biologically relevant and associated with differences in susceptibility or response to environmental factors. However, since the current study evaluated the status of the occupational group and not the occurrence of the disease, additional studies with larger populations, quantitative evaluation of heavy metal exposure and clinical outcomes would be needed to determine whether these polymorphisms have any implications for disease susceptibility among occupationally exposed workers.

4.5.3 Comparing this with a Bangladeshi Population Study

The results of the present study are especially pertinent in the light of the results of Islam et al. (2013) who studied the association of CYP1A1 polymorphisms with lung cancer susceptibility in a Bangladeshi population. They conducted a case-control study of 106 lung cancer patients and 116 controls and studied three variants of CYP1A1, including the two polymorphisms studied in the present study, rs4646903 and rs1048943. The authors reported significantly elevated risk of lung cancer in people with heterozygous, mutant and combined heterozygous-plus-mutant genotypes of rs4646903 after accounting for age, sex and smoking. A significant association was also found for the heterozygous and combined heterozygous-plus-mutant genotype of rs1048943. The authors concluded that the CYP1A1 2B variants represented by rs4646903 and rs1048943 were associated with increased lung cancer risk in the study population. An important finding of the study was that the associations between CYP1A1 polymorphisms and lung cancer were dependent on the level of tobacco-smoking exposure. This finding indicates that genetic variation in CYP1A1 may play a role in individual susceptibility to carcinogen exposure when there is a significant exposure to tobacco smoking.

The present study also found significant associations for both CYP1A1 polymorphisms, but the study population and outcome were different, and both polymorphisms were statistically significantly associated in the present study, although the genotype patterns and direction of association were different from those reported in the lung-cancer study.

The agreement between the two studies is important because both studies were done in Bangladeshi populations and examined the same two CYP1A1 polymorphisms. In the present study, differences in the distribution of these same variants were found between occupationally exposed shipyard workers and healthy controls, while Islam et al. showed that these variants were associated with a disease outcome, namely lung cancer, particularly in people with large tobacco-smoking exposure. This similarity suggests the biological importance of CYP1A1 variation in Bangladeshi population exposed to environmental or occupational agents. The results should not be interpreted as proof that the same genetic variants will have the same effect in different exposure scenarios. The context of exposure is especially significant. Islam et al. (2013) mainly studied tobacco-smoking exposure and found that the impact of CYP1A1 polymorphisms on lung cancer susceptibility was more pronounced at high levels of smoking exposure. The present study, on the other hand, included shipyard workers who were occupationally exposed to heavy metals and PAHs. Thus, the large genotype differences found in the present study could be a result of differences in genetic susceptibility or biological response to the occupational exposure environment, and not a result of differences in susceptibility to lung cancer. Results suggest a model in which genetic variation can modify individual susceptibility given environmental exposures. Propose that CYP1A1 polymorphisms are potential genetic markers that should be studied in Bangladeshi populations who are exposed to environmental and occupational carcinogenic and/or toxic agents.

4.5.4 Allele frequencies across populations

The frequencies of CYP1A1 alleles are highly variable among populations, as shown in a systematic compilation by Garte et al. (2001) of metabolic gene polymorphism frequencies in control populations worldwide, and in more recent regional surveys like that of Alvarado et al. (2021) in the central coastal population of Peru. The variant allele at the MspI locus is always more frequent in East and South Asian populations than in populations of European descent, and the exon 7 variant exhibits a similar, but less steep, gradient. This variation is why it was necessary to use a locally recruited control group in the present study instead of comparing the frequencies with published frequencies from other locations.

In the present control group, the minor allele frequencies were 0.4545 for the A allele of the rs4646903 and 0.4455 for the C allele of the rs1048943. Both are high compared to the minor allele frequencies typically observed for these loci in South Asian reference populations, and the two are also very similar to each other. It is recommended that a formal comparison of the present control frequencies with published South Asian and reference-panel values be undertaken prior to submission for publication, which would greatly enhance the report. If the local frequencies are very far from the published range, the first question to ask would be technical and not population-genetic, as explained in Section 4.10.

4.5.5 Allele nomenclature and the direction of comparison

Before any allele-for-allele comparison with the literature is completed, a caution on the nomenclature is necessary. In this thesis, the genotypes are reported as rs4646903 A/G and rs1048943 T/C, while in the published literature, these variants are usually reported in the sense of the gene or complementary DNA as rs4646903 T/C (m1) and rs1048943 A/G (Ile/Val). The two systems are formally equivalent, adenine pairs with thymine and guanine pairs with cytosine, and reporting in the meta-analytic literature of rs4646903 in A/G form is not unusual. However, the strand convention should be explicitly checked against the primer design and the reference sequence before comparing the direction of association to previous publications, as an accidental strand reversal would reverse the direction of effect reported in every association, but would not be detected by any internal inconsistencies in the data set.

4.6 Expression, activity and prediction of genotype for CYP1A1

The present study measured genotype and not expression. It is helpful, however, to state what the published expression literature suggests about the relationship between these genotypes and the enzyme phenotype, both because this is the biological premise on which the study rests and because this is the area in which the premise is not as secure as is sometimes assumed.

Functionality of these variants is based mainly on studies in mitogen stimulated peripheral blood lymphocytes, where catalytic activity of CYP1A1 is determined by ethoxyresorufin-O-deethylase (EROD) activity and gene expression by messenger RNA abundance. Basal

EROD activity was not different between individuals with different CYP1A1 genotypes, but inducible enzyme activity was about threefold higher in individuals homozygous for the rare MspI variant (Cosma et al., 1993). In the same volunteers, Landi et al. (1994) studied the relationship between genotype and mRNA expression and enzyme activity and found that both basal and induced EROD activity was significantly elevated in carriers of a variant allele at one or both sites. Crofts et al. (1994) studied the transcriptional and post-transcriptional consequences of both the exon 7 substitution and the MspI polymorphism and found increased EROD activity to be primarily associated with the exon 7 substitution (valine for isoleucine in the active site) and not with the MspI variant outside the coding region.

This last distinction matters for the interpretation of the two loci in the present study. The exon 7 variant has a directly demonstrable structural mechanism: it changes an amino acid in the haem-binding domain, which forms the catalytic centre of the enzyme, and Hayashi et al. (1991) mapped the substitution to exactly this area. The MspI variant is located in the non-coding 3′-flanking region and cannot have such a mechanism and, if it does have an effect, it must be either through regulation and stability of the transcripts or as a marker in linkage disequilibrium with another regulatory variant in the gene. The functional evidence for the MspI variant is also less strong and less consistent than the evidence for the exon 7 variant.

This has two implications for the current study. The first is that even if the difference in genotype frequency between two groups were entirely real, it would not justify any claim about the metabolic phenotype of those groups, since the genotype-phenotype relationship at these loci is only partial. The second is that the logical next step for this research programme is to measure the expression of CYP1A1 messenger RNA (mRNA) in peripheral blood lymphocytes by reverse transcription quantitative PCR (RT-qPCR) and genotype in the same individuals. This would directly assess whether workers exposed to the same environmental factors who have the variant genotypes exhibit the higher inducibility predicted by the susceptibility hypothesis.

4.7 Possible Biological Explanations and Their Limits

CYP1A1 is the gene that encodes an enzyme responsible for metabolism of polycyclic aromatic hydrocarbons and related xenobiotics, and both polymorphisms examined in this study have been shown to affect enzyme activity or inducibility. Biological plausibility suggests that the individual's ability to metabolise the xenobiotic compounds they are exposed to in a shipyard may be modified by variation in the activity of the enzyme, and heavy metals have been reported to influence the expression of CYP1A1 through aryl hydrocarbon receptor signalling (Sadikovic et al., 2006; Vakharia et al., 2001). The mechanistic chain leading to such variation that might eventually be important is well characterised: AhR-mediated induction of CYP1A1, bioactivation of benzo[a]pyrene via the diol-epoxide pathway to a species that forms covalent adducts at guanine, and, where such adducts escape repair prior to replication, the characteristic transversion mutations associated with PAH-driven carcinogenesis.

But the current research did not measure any of the intermediate steps in that chain. No enzyme activity, no biomarker (urinary metal concentration or 1-hydroxypyrene), no adduct measurement, and no functional outcome was obtained, only genotype frequency in relation to worker status was measured. Any biological mechanism that links CYP1A1 genotype to occupational-group membership is thus speculative in the context of this data set and not supported by the results of this study.

More basically, and as described in Section 1.5, the logic of the design limits the form that a biological explanation could take. Genotype is determined at birth and cannot be a result of an occupational exposure. On the other hand, there is no known biological mechanism by which the CYP1A1 genotype would affect the probability of an individual taking up a job in a shipyard or junkyard. Hence, there is no causal reading in either direction. The present data cannot rule out the possibility that differences in allele frequency between the recruited worker and control groups are due to factors other than exposure biology.

4.8 Genetic-Model Differences

The pattern of significance across the five genetic models for each SNP is informative about which genotype comparisons contribute to the observed associations, but should not be used to choose a single “correct” model based on the smallest P value. The five models

are not five tests of five different hypotheses, but five parameterisations of the same underlying genotype data and their results are necessarily correlated.

The dominant, recessive and additive models were significant for rs4646903, and the GG-versus-AA contrast was significant, but the AG-versus-AA contrast was not. This internal consistency suggests that the association is focused on the GG genotype, and that the apparent significance of the dominant model (AG+GG) is driven by the GG effect, and not by an independent effect of the heterozygotes. The additive model is also important because it will be identified whenever there is a large difference between the two homozygote groups, regardless of whether the heterozygote groups are distinguishable or not.

The dominant and additive models and the TC-versus-TT codominant contrast were significant, and the recessive model and the CC-versus-TT codominant contrast could not be estimated because of the lack of CC genotypes among workers. The two degrees of freedom codominant test for this SNP was not significant (P = 0.0776) but this is likely due to the non-estimable term reducing the informativeness of the combined test rather than a true lack of association; a two degree of freedom test in which one degree of freedom is non-informative is not the same as a test in which both are informative. The over-dominant model was not significant for either SNP, suggesting that the associations were not due to heterozygosity alone.

4.9 Age and Sex

In all models, age was independently associated with worker status, suggesting a true age difference between groups. Sex was not significantly associated in any model, but the strong sex imbalance in the sample reduces power to detect a true effect. This means that the association is independent of age and sex, if genotype was still significant after adjusting for these two covariates.

4.10 Hardy-Weinberg Equilibrium and Genotyping Quality

In a genotyping study, Hardy-Weinberg equilibrium testing is done for two reasons. It is a population-genetic statement that says whether or not the proportions of genotypes are compatible with random mating and lack of strong selection at the locus. It is a laboratory quality control and in this context its most immediate use is to provide the only available

check on the accuracy of the genotype calls, as the two most common genotyping errors produce characteristic and detectable distortions of the genotype distribution. Allele dropout is when one allele always fails to amplify, resulting in a loss of heterozygosity and a heterozygote deficit. Insufficiently stringent allele discrimination (in which the mismatched primer extends and produces a faint product) results in the conversion of homozygotes to apparent heterozygotes and an excess of heterozygotes. In a two-tube ARMS assay read as presence or absence of a band, both types of failure are possible and neither is apparent when looking at a single gel.

The control genotype distributions at both loci were in Hardy-Weinberg equilibrium as the test was performed. There are two points, however, which should be considered in any revision of this work. The first is the degrees of freedom: a Hardy-Weinberg goodness-of-fit test uses the same data to estimate the allele frequency as is used in the test, thus using one degree of freedom, and the correct reference distribution has one degree of freedom, not two. Testing against the 2-df distribution will give systematically higher P values and thus be less likely to pick up a true departure. The second is that the equilibrium was only tested in the control group. In a study of this design, it is customary to test both groups, and the exposed group in the study is not expected to differ from the control group in any way, so there is no methodological reason to exclude it.

It should also be noted that the heterozygote frequency 2pq is at a mathematical maximum of 0.5 at equal allele frequencies under Hardy-Weinberg proportions. Heterozygote proportions near or equal to one-half, thus warrant investigation, because they cannot be explained by an equilibrium distribution and in this kind of assay are the result of an annealing temperature that is not low enough to prevent extension of the mismatched primer. The single most valuable addition that could be made to the present dataset would be confirmation of the genotype calls in a random subset of samples by an independent method, such as Sanger sequencing, or PCR-RFLP using the MspI site that rs4646903 creates.

4.11 Multiple-Testing Considerations

The rs1048943 genotype-distribution test was the only one to remain significant after Bonferroni correction for fourteen tests; 9 of 13 rankable tests remained significant after

FDR. When the two genotype-distribution tests were combined, the results were different, demonstrating sensitivity to the test-counting convention. However, if they are in linkage disequilibrium in this sample, as reported elsewhere in a Bangladeshi population, the two SNPs' tests are not independent and Bonferroni is more conservative than its nominal threshold suggests, and the two loci may not provide fully independent evidence.

4.12 Future Directions and Perspectives

The present study is not a conclusion, but a beginning, and the significance of the frequency data reported here is more in what they can do. There are several directions below, listed in approximate order of priority.

Exposing less, not assuming exposure

The one thing that this design is missing is that exposure is not measured but rather is indicated by a group label. Biological exposure indices (e.g., blood lead, urinary cadmium, urinary 1-hydroxypyrene) and exact duration of employment and job category would make exposure a continuous variable. Only then is it possible to perform true gene-environment analysis, as defined by the coke-oven literature: to determine whether the association between internal dose and biological effect is different for each genotype.

Adding a genotoxicity endpoint

A study based on direct measurement of DNA damage by the comet assay, micronucleus frequency in lymphocytes, or urinary 8-hydroxy-2′-deoxyguanosine as a marker of oxidative DNA damage, stratified by CYP1A1 genotype, would test the susceptibility hypothesis without waiting for disease to occur. This is the design that would show the mechanism that the present study can only propose, and it is still possible in this number of staff.

Measuring expression along with genotype

The connection between these genotypes and the enzyme phenotype is incomplete and controversial as mentioned in Section 4.6. To determine if the genotypes have functional significance in this population, quantification of CYP1A1 mRNA in peripheral blood lymphocytes from the same individuals in which the genotype is determined using reverse transcription quantitative PCR would be useful. Such a study would require careful design, especially in light of the reservations of van Duursen et al. (2005) about using lymphocyte

CYP1A1 as an exposure biomarker, and would ideally include an ex vivo induction step to ensure that inducibility, not just basal expression, was measured.

Increasing the number of genes that are tested

The bioactivation is done by CYP1A1 and the detoxification is done by the phase II conjugating enzymes. The risk associated with a high-activity phase I genotype is dependent on the phase II and DNA repair capacity behind it, and the classic high-risk genotype reported in the lung cancer literature is a variant CYP1A1 genotype combined with a GSTM1 null genotype (Nakachi et al., 1993). Combined-genotype analysis using GSTM1, GSTT1, GSTP1 and NQO1, plus a DNA repair gene like XRCC1 or OGG1, would be possible and would significantly enhance the biological interpretation of the results.

Improving the genotyping platform

The main technical uncertainty that applies to the current dataset would be eliminated by independent validation of a random subset of genotypes by Sanger sequencing, and by using a format that includes an internal amplification control, like tetra-primer ARMS-PCR. The assay could be placed on a more solid footing by optimising the annealing temperature using gradient PCR for each primer set and documenting the discrimination against samples of known genotype.

Planning the ultimate study

The control allele frequencies reported here are the ones that are needed for a sample size calculation and a properly powered study can now be designed, not guessed. A study of this type should include workers and controls from the same geographical and demographic population, should be age matched, and should document the above-mentioned covariates. A prospective design (following an exposed cohort) or a case-control study nested within a cohort would be needed if a disease endpoint is ultimately of interest; a cross sectional comparison of genotype between exposure groups cannot answer that question, however large the sample.

Wider public health perspective

This study, apart from the genetic issue, shows that there is a large and dangerous workforce working without any kind of biological health surveillance. The low cost genotyping and biomonitoring approach shown here is valuable regardless of the specific

loci analyzed and could be expanded to regular monitoring for metal exposure and genotoxic effect markers. Any such programme should be presented as informing protection of the worker, not selection of the worker, as discussed in Chapter 5.

4.13 Study Limitations

Sample size was modest (55/55), resulting in wide CIs for small-cell comparisons, and the lack of CC-genotype workers at rs1048943 meant that two comparisons could not be estimated and exact testing was required. The recessive model (rs4646903) exhibited adjusted/unadjusted divergence. The majority of the findings were not significant after stringent multiple-testing correction. Allele frequencies for control samples differed from the South Asian reference panels, particularly for the rs1048943 locus. Direct assessment of possible linkage disequilibrium between the SNPs was not done. This was an observational, cross-sectional study without internal exposure biomarkers, without smoking data, and without measurement of disease outcome; findings are specific to this sample from the Bangladeshi shipyard.

4.14 Overall Implication

These results suggest that the genotype and allele frequencies of CYP1A1 rs4646903 and rs1048943 were different between this sample of shipyard and junkyard workers and healthy controls, and that the direction of association was internally consistent across multiple genetic models at each locus, but opposite between the two loci, after accounting for age and sex. These are associations with occupational-group membership in one particular sample. They do not prove that these genotype differences are due to heavy metal or other occupational exposures (which is not possible on first principles) or that either variant is associated with increased cancer risk (because cancer status was not measured). With the small sample size, the small-cell and non-estimability problems with rs1048943, the opposite direction of effect of the two loci, the few findings that survived conservative multiple-testing correction, and the systematic demographic differences between the groups, these findings should be considered preliminary and hypothesis-generating. The study provides with confidence a set of locally measured genotype and allele frequencies for a workforce that had never been characterized, a validated low cost assay by which this can be repeated, and an explicit account of the inferential limits within which such

comparisons must be read. More definitive conclusions would need to be made in a larger, independently recruited and age-matched sample with direct biomarkers of exposure and a measured biological endpoint.

Conclusion

In this study, genotype and allele frequencies of two alleles of CYP1A1 (rs4646903 and rs1048943) were determined in 55 shipyard and junkyard workers of Bangladesh who had been exposed to heavy metals for at least 10 years and 55 healthy controls who were not exposed to any heavy metals using a two-tube allele-specific ARMS-PCR assay developed in-house. Allele distributions were significantly different between the two groups at both loci and several age-sex adjusted genetic models were significantly associated with genotype and exposure-group membership, in opposite directions at the two loci. On the test performed, control genotype distributions did not deviate from Hardy-Weinberg equilibrium.

These findings should be interpreted as descriptive relationships with group membership in one small sample. The study does not show that occupational exposure affected genotype frequencies, nor does it show that these polymorphisms are related to disease in this population, since both are polymorphisms that cannot be induced by adult exposure. However, sampling variation, population stratification, and recruitment and selection differences between the two groups are still possible explanations that cannot be ruled out by the present design, and most individual findings were not significant after conservative correction for multiple testing.

References

Alam, S. M. R., & Faruque, A. S. G. (2014). Health hazards and risks of vulnerability of ship breaking workers: A case study on Sitakunda ship breaking industrial area of Bangladesh. ResearchGate. https://www.researchgate.net/publication/303018933

Alexandrie, A., Warholm, M., Carstensen, U., Axmon, A., Hagmar, L., Levin, J. O., Ostman, C., & Rannug, A. (2000). CYP1A1 and GSTM1 polymorphisms affect urinary 1-hydroxypyrene levels after PAH exposure. Carcinogenesis, 21(4), 669-676. https://doi.org/10.1093/carcin/21.4.669

Alvarado, A. T., et al. (2021). Prevalence of GSTM1 null and CYP1A1*2A (rs4646903) variants in the central Peruvian coastal population: Pilot study of predictive genetic biomarkers. ScienceDirect. https://www.sciencedirect.com/org/science/article/pii/S0428029625000423

Balali-Mood, M., Naseri, K., Tahergorabi, Z., Khazdair, M. R., & Sadeghi, M. (2021). Toxic mechanisms of five heavy metals: Mercury, lead, chromium, cadmium, and arsenic. Frontiers in Pharmacology, 12, Article 643972. https://doi.org/10.3389/fphar.2021.643972

Barek, M. A., Hasan, M., Shibly, A. Z., Akter, S., & Islam, M. S. (2023). Assessment of the association of CYP1A1 gene polymorphisms with the susceptibility of cervical cancer: A case-control study and meta-analysis. Gene, 879, Article 147591. https://pubmed.ncbi.nlm.nih.gov/37483787/

Brookes, A. J. (1999). The essence of SNPs. Gene, 234(2), 177-186. https://pubmed.ncbi.nlm.nih.gov/10395891/

Garte, S., et al. (2001). Metabolic gene polymorphism frequencies in control populations. Cancer Epidemiology, Biomarkers & Prevention, 10(12), 1239-1248. https://pubmed.ncbi.nlm.nih.gov/11751440/

Hasan, M. K., et al. (2024). Impact of shipbreaking industries on the Sitakunda coastal environment, Chattogram. Marine Pollution Bulletin, 209, Article 117130. https://www.sciencedirect.com/science/article/abs/pii/S0025326X24014280

Islam, M. S., Ahmed, M. U., Bin Sayeed, M. S., Al Maruf, A., Mostofa, A. G. M., Akram Hussain, S. M., Kabir, Y., Daly, A. K., & Hasnat, A. (2013). Lung cancer risk in relation

to nicotinic acetylcholine receptor, CYP2A6 and CYP1A1 genotypes in the Bangladeshi population. *Clinica Chimica Acta*, *416*, 11-19. https://doi.org/10.1016/j.cca.2012.11.011

Kiyohara, C., Horiuchi, T., Takayama, K., & Nakanishi, Y. (2012). Genetic polymorphisms involved in carcinogen metabolism and DNA repair and lung cancer risk in a Japanese population. Journal of Thoracic Oncology, 7(6), 954-962. https://doi.org/10.1097/jto.0b013e31824de30f

Mutar, T. F., & Oraibi, O. S. (2021). Sequence polymorphism in xenobiotic metabolising genes in Iraqi colorectal cancer patients. Oman Medical Journal. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8325151/

Nebert, D. W., Dalton, T. P., Okey, A. B., & Gonzalez, F. J. (2004). Role of aryl hydrocarbon receptor-mediated induction of the CYP1 enzymes in environmental toxicity and cancer. Journal of Biological Chemistry, 279(23), 23847-23850. https://pubmed.ncbi.nlm.nih.gov/15028720/

Pinto, T. G., et al. (2025). The impact of genetic polymorphisms for detecting genotoxicity in workers occupationally exposed to metals: A systematic review. Journal of Applied Toxicology. https://doi.org/10.1002/jat.4711

Rahman, M. M., et al. (2022). Toxic metal pollution and ecological risk assessment in water and sediment at ship breaking sites in the Bay of Bengal Coast, Bangladesh. Marine Pollution Bulletin, 174, Article 113274. https://www.sciencedirect.com/science/article/abs/pii/S0025326X21013084

Sadikovic, B., Haines, T. R., Murakami, K., & Bhatt, P. (2006). Transcriptional activation and posttranscriptional modification of Cyp1a1 by arsenite, cadmium, and chromium. Toxicological Sciences, 94(2), 20-34. https://pubmed.ncbi.nlm.nih.gov/17606337/

Tchounwou, P. B., Yedjou, C. G., Patlolla, A. K., & Sutton, D. J. (2012). Heavy metal toxicity and the environment. Experientia Supplementum, 101, 133-164. https://doi.org/10.1007/978-3-7643-8340-4_6

Tsoutsoulopoulou, A., Evangelou, A. M., & Bharat, B. A. (2009). Cytochrome P450 CYP1A1: Wider roles in cancer progression and prevention. BMC Cancer, 9, Article 187. https://bmccancer.biomedcentral.com/articles/10.1186/1471-2407-9-187

UNCTAD. (2020). Review of maritime transport 2020. United Nations Conference on Trade and Development.

https://unctad.org/system/files/official-document/rmt2020_en.pdf

Vakharia, D. D., Liu, N., Pause, R., Faux, S., & Bhatt, P. (2001). Modulation of aryl hydrocarbon receptor-regulated gene expression by arsenite, cadmium, and chromium. Toxicology and Applied Pharmacology, 170(1), 33-40. https://pubmed.ncbi.nlm.nih.gov/15337587/

Alexandrie, A. K., Warholm, M., Carstensen, U., Axmon, A., Hagmar, L., Levin, J. O., Ostman, C., & Rannug, A. (2000). CYP1A1 and GSTM1 polymorphisms affect urinary 1-hydroxypyrene levels after PAH exposure. Carcinogenesis, 21(4), 669-676. https://doi.org/10.1093/carcin/21.4.669

Cosma, G., Crofts, F., Taioli, E., Toniolo, P., & Garte, S. (1993). Relationship between genotype and function of the human CYP1A1 gene. Journal of Toxicology and Environmental Health, 40(2-3), 309-316. https://doi.org/10.1080/15287399309531796

Crofts, F., Taioli, E., Trachman, J., Cosma, G. N., Currie, D., Toniolo, P., & Garte, S. J. (1994). Functional significance of different human CYP1A1 genotypes. Carcinogenesis, 15(12), 2961-2963. https://pubmed.ncbi.nlm.nih.gov/8001264/

Hayashi, S., Watanabe, J., Nakachi, K., & Kawajiri, K. (1991). Genetic linkage of lung cancer-associated MspI polymorphisms with amino acid substitution in the heme-binding region of the human cytochrome P450IA1 gene. Journal of Biochemistry, 110(3), 407-411.

Kawajiri, K., Nakachi, K., Imai, K., Yoshii, A., Shinoda, N., & Watanabe, J. (1990). Identification of genetically high risk individuals to lung cancer by DNA polymorphisms of the cytochrome P450IA1 gene. FEBS Letters, 263(1), 131-133.

Landi, M. T., Bertazzi, P. A., Shields, P. G., Clark, G., Lucier, G. W., Garte, S. J., Cosma, G., & Caporaso, N. E. (1994). Association between CYP1A1 genotype, mRNA expression and enzymatic activity in humans. Pharmacogenetics, 4(5), 242-246. https://pubmed.ncbi.nlm.nih.gov/7894496/

Nakachi, K., Imai, K., Hayashi, S., & Kawajiri, K. (1993). Polymorphisms of the CYP1A1 and glutathione S-transferase genes associated with susceptibility to lung cancer in relation to cigarette dose in a Japanese population. Cancer Research, 53(13), 2994-2999.

van Duursen, M. B. M., Sanderson, J. T., & van den Berg, M. (2005). Cytochrome P450 1A1 and 1B1 in human blood lymphocytes are not suitable as biomarkers of exposure to

dioxin-like compounds: Polymorphisms and interindividual variation in expression and inducibility. Toxicological Sciences, 85(1), 703-712. https://doi.org/10.1093/toxsci/kfi089

Yousefian, M., Angaji, A., Siasi, E., Rahmani, S. A., & Khiaban, S. A. (2022). Role of CYP1A1, CYP2D6, and NOS3 gene polymorphisms in idiopathic recurrent pregnancy loss in the Iranian Azeri population: A case-control study. International Journal of Reproductive BioMedicine, 20(8), 671-682. https://doi.org/10.18502/ijrm.v20i8.11756

Zeng, W., Li, Y., Lu, E., & Ma, M. (2016). CYP1A1 rs1048943 and rs4646903 polymorphisms associated with laryngeal cancer susceptibility among Asian populations: A meta-analysis. Journal of Cellular and Molecular Medicine, 20(3), 508-519. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4727562/

Zhu, X., et al. (2016). Associations between CYP1A1 rs1048943 A>G and rs4646903 T>C genetic variations and colorectal cancer risk: Proof from 26 case-control studies. Oncotarget, 7(31), 50255-50267.

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5239481/