Mutat Res Gen Tox En 826 (2018) 15–24

Contents lists available at ScienceDirect

Mutat Res Gen Tox En

journal homepage: www.elsevier.com/locate/gentox

Interactive potential of genetic polymorphism in Xenobiotic metabolising and DNA repair genes for predicting lung cancer predisposition and overall survival in North Indians

Apurva Bhardwaj[a], Charu Bahl[a], Siddharth Sharma[a,*], Navneet Singh[b], Digamber Behera[b]

[a] Department of Biotechnology, Thapar University, Patiala, Punjab-147002, India
[b] Department of Pulmonary Medicine, Post Graduate Institute of Medical Education and Research (PGIMER), Sector 14, Chandigarh, India

ARTICLE INFO

Keywords:
Xenobiotics
DNA repair
Polymorphisms
Lung cancer

ABSTRACT

Introduction: Cancer, a multi-step, multifactorial and multi-gene disease, not only damages the genomic integrity of the cell but also hinders the DNA repair mechanisms of the body. Gene-gene and gene environment interactions amongst the genetic polymorphisms together modulate the susceptibility towards a cancer. We have studied the high order gene interactions between the genetic polymorphism of detoxifying genes (*CYP1A1, Ahr, XRCC and GST1*) that play a key role in the metabolism of the xenobiotics and have been proved to be prognostic markers for lung cancer

Methods: 237 cases and 250 controls have been genotyped using PCR-RFLP technique. In order to find out the association, unconditional logistic regression approach was used and to analyse high order interactions MDR and CART was used.

Results: In the MDR analysis, the best model was one factor model which included *GSTM1* (CVC 10/10, Prediction error = 0.43, $p < .001$). The best three factor model comprised of *XRCC1* 632, *XRCC1* 206, *GSTM1* (CVC 10/10, Prediction error = 0.45, $p < .0001$). The CART analysis exhibited that Node 1 carrying mutant type of GSTM1 imposed the highest risk towards lung cancer (OR = 11.0, 95%C.I. = 6.05-20.03, $p = .000001$). Wild type of *GSTM1* when combined with mutant type of *CYP1A1* M2 and *XRCC1* 632, an 8 fold risk towards lung cancer was observed (95%C.I. = 4.07-16.29, $p = .00001$). The high order interactions were used to predict the prognosis of lung cancer patients. Of all the genetic variants, XRCC1 632, GSTM1 and AhR rs2066853 was the most important determinant of overall survival of lung cancer patients

Conclusion: Through the study we introduced the concept of polygenic approach to get an insight about the various polymorphic variants in determining cancer susceptibility. Lesser number of subjects were found in the high risk subgroups. Further studies with larger sample size are required to warranty the above findings.

1. Introduction

Genetic epidemiology is often undermined due to the interplay of several interactions occurring amongst the genes and between the gene and environment. Molecular studies reveal that genetic interactions, across different verticals, leading to mutations and alterations are few of the common factors responsible for multi-locus and multifactorial diseases such as cancer, neurodegenerative diseases and other lifestyle diseases such as hypertension and diabetes. [1]. Literature from the past is also indicative of the fact that gene–gene and gene environment interactions are not only difficult to detect, but also due to the multi dimensionally invasive nature of these diseases, they pose several challenges in understanding the diagnosis, prognosis and treatment of these diseases. Cancer, a multi-step, multifactorial and multi-gene disease, causes significant damage to the genomic integrity of the cell, wherein this damage is at times compromised by the DNA repair mechanisms of the body. Lung cancer, succumbing one in every 10 smokers to death in the world, is a cancer with poor prognosis and high morbidity and is variedly influenced by an individual's genetic polymorphism, environmental factors and several cross-talk genetic interactions [2].

Heterogenous in its clinical presentation, lung cancer has a diverse range of prognostic variables. A complex mesh of risk factors such as environmental pollution, tobacco/cigarette smoke, oxidizing agents, alkylating agents together with the genetic variability predisposes an individual to the risk of lung cancer [3]. The human body is equipped with a robust detoxification system that carries out the metabolism of

* Corresponding author.
E-mail address: siddharthsharma.phd@thapar.edu (S. Sharma).

https://doi.org/10.1016/j.mrgentox.2017.12.006
Received 8 September 2017; Received in revised form 18 October 2017; Accepted 15 December 2017
Available online 21 December 2017

the xenobiotics that includes both carcinogenic and mutagenic agents entering the body through different sources. The carcinogenic agents include compounds specifically found in tobacco smoke such as PAH (polycyclic aromatic hydrocarbons), NNK 4-(methyl nitrosamino)-1-(3-pyridyl)-1-butanone, nitroso compounds and aromatic amines. These compounds are processed *via* a cascade of metabolising reactions, where in the Phase I reactions the procarcinogens are activated by a myriad of enzymes that include cytochromes *(CYP1A1, CYP1A2, CYP1B1, CYP2A6, CYP2E1* and *CYP3A4)* followed by Phase II reactions where the activated carcinogen forms an adduct with the DNA to produce somatic mutations serving as a point for carcinogenesis. During the Phase I reactions, AhR, a ligand induced transcription factor along with the aryl hydrocarbon receptor nuclear translocator (ARNT) forms a heterodimer and interacts with the target genes (such as *CYP* genes) of the phase I reaction. Upon getting exposed to carcinogens, the AhR upregulates the CYP1A1 activity and therefore serves as a significant biomarker for smoke induced lung carcinoma and the disease prognosis [4]. During the Phase II reactions the pre-carcinogens along with glutathione (transferases by nature; are encoded as EC 2.5.1.18) are removed with the help of hydrolases enzymes. [5]. The intervention of the DNA repair system in case of cancer is unpredictable and is deemed to safeguard the cells. The repair system has several proteins such as the XRCC1 (a scaffold protein) that plays a critical role in the BER pathway, but due to the genetic variability within, its behaviour modulates and becomes a prospective cancer susceptible gene. Therefore, these essential genes that dominantly regulate the xenobiotic metabolism of the body, placed on different loci and portray diverse polymorphism.

The classic candidate gene approach, which takes in account the knowledge and function of the SNP, has backed research findings but does not hold true for several cancer association studies [6]. Data suggests that, the disease being multi-factorial, cross-talks between the genes present on different loci is a crucial element in the architecture of cancer prognosis. The evaluation of the risk of the cancer for a population remains obsolete due to small sampling groups, inaccurate statistical methods and inability to study complex interactions between the multiple genes and environmental risk factors. Proven in history, curse of dimensionality answers the low weightage conferred to single gene polymorphism in studying complex diseases such as cancer. To undermine the principle that backs the susceptibility for cancer, it is important to study the cumulative effect of the intra-gene and gene-environment interaction. With the introduction of Multifactor Dimensionality Reduction (MDR) and classification and regression tree (CART), the high order intra-gene and gene-environment interactions can be studied, validating the study of genetic polymorphism in cancer progression.

Across different studies, researchers have studied high order gene interactions and the association of different polymorphisms with the susceptibility of cancers such as lung cancer, bladder cancer and others. However, in this study, we have used high-end statistical and analytical methods (such as MDR and CART) to minimize the chances of ambiguity. To add robustness to the study, we have evaluated the high order gene interactions between the genetic polymorphism of detoxifying genes and DNA Repair genes (*CYP1A1, Ahr, XRCC and GST1*) that play a key role in the metabolism of the xenobiotics and have been proved to be prognostic markers for lung cancer [7]

2. Material and methods

Current study recruited 237 lung cancer (LC) cases and 250 healthy controls from the Department of Pulmonary Medicine, Post Graduate Institute of Medical Education & Research (PGIMER), Chandigarh. A written informed consent was taken from each volunteer prior to blood collection. This study has been reviewed and approved by Ethics committee of PGIMER. A questionnaire having all the details about various epidemiological factors was filled by trained personnel during the recruitment process. An attempt was made to match each control on basis of age, sex and smoking parameters with respect to cases. However, despite our attempts as expected smoking came out to be more prevalent in lung cancer cases as compared to controls".The only exclusion criteria for LC patients was, they should not have any previous history of any carcinoma. The controls were healthy people who visited the hospital for health check-ups. A major attempt was to avoid sampling bias which may occur due to difference in age, sex and smoking status of the cases and controls. Smoking was quantified using pack years which is calculated by this formula: [(cigarettes or beedis (type of Indian cigarette) per day/20) X number of years smoked]. The other clinical details including histology, stage and TNM details were obtained from the medical records of the patients in the hospital.

2.1. DNA isolation

Blood collected from each subject was 5 ml and it was used to isolate the genomic DNA using the protocol given by Sodhi et al. [8]. The isolated DNA was quantified using Nanodrop and stored at −4 °C for further use.

2.2. Genotyping of GSTM, CYP1A1, XRCC and AhR genetic variants

The m1 polymorphism (rs4646903) in the 3′ noncoding region (3′-UTR) of the *CYP1A1* gene arises from a point mutation resulting in T ➡ C transition. This results in elevated induction of the enzyme, and thus, increased levels of activated intermediates. The m2 polymorphism ($Ile^{462}Val$, exon7; rs1048943) located in heme binding region results in an increase in microsomal enzyme activity. Both the *GSTM1* and *GSTT1* gene deletions in the populations render the enzymes inactivated thereby hindering the detoxification mechanism. The polymorphic variants of *XRCC* gene included the non-synonymous and synonymous variants of *XRCC1 i.e. XRCC1* $Arg^{399}Gln$, $Arg^{194}Trp$, $Pro^{206}Pro$ and $Gln^{632}Gln$. Out of the four SNPs studied for *AhR* gene, three (rs2282885, rs10250822, rs7811989) all were located in the intronic region and thus these might affect the expression or function of *AhR* gene. They may increase or decrease gene transcription and might also influence the proper splicing of RNA or yielding alternatively spliced messenger RNA variants. The non-synonymous SNP (rs2066853) resulted in substitution of arginine with lysine amino acid at 554 position which might lead to change in the primary structure of the protein and influence the function of the AhR receptor.

Genotyping of all the polymorphic sites under study was done using PCR-RFLP. The genotyping for two *GSTM1* and *GSTT1* gene variants was done by multiplex PCR approach as previously detailed by Sharma et al. [9]. For XRCC1 variants ($Arg^{399}Gln$, $Arg^{194}Trp$, $Pro^{206}Pro$, $Gln^{632}Gln$) polymorphic sites, the genotyping was carried in a similar manner as reported by Singh et al. [10]. In case of genetic variants of *CYP1A1* gene namely CYP1A1 m1 (rs4646903) and CY1PIA1 m2 (rs1048943), the protocol described by Girdhar et al. [11] was applied to find out the genotype of the subjects. The genotyping of four *AhR* variants (rs7811989, rs10250822, rs2282885, rs2066853) was done by PCR-RFLP using specific primer sequences and restriction enzymes as described previously by Bin et al. [12]. The PCR reaction (25 µl) used to amplify the desired fragment comprised of 1 × PCR buffer, 1.5 mM $MgCl_2$, with 0.5 µM of both foward and reverse primer, 200 µM of each dNTP's, 100 µg/ml bovine serum albumin (BSA) and 1U Taq polymerase (DNAzyme II DNA Polymerase, Thermo Scientific) and approximately 200 ng DNA. The amplified products were digested with their respective restriction enzymes as described above. The digested products were resolved on either agarose gel or polyacrylamide gel to find out the restriction patterns. Scoring of the patterns is done to find out the genotypic status of the sample. The genotyping of 15% samples was done twice in order to check the reproducibility of the results and it was 100%.