2017

# Identification of SNPs Potentially Associated with Asthma Using a Di-isocyanate Luciferase Reporter Assay

Wentao Jiang

# Identification of SNPs potentially associated with di-isocyanate asthma using a luciferase reporter assay

**Wentao Jiang**

**Thesis submitted to the**
**Davis College of Agriculture, Natural Resources and Design**
**at West Virginia University**
**in partial fulfillment of the requirements**
**for the degree of**

**Master of Science**
**in**
**Genetics and Developmental Biology**

**Jianbo Yao, Ph.D., Chair**
**Daniel G. Panaccione, Ph.D.**
**Vagner A. Benedito, Ph.D.**

**Morgantown, West Virginia**
**2017**

**Keywords: SNPs, Asthma, Transcription factor, Luciferase assay**

# ABSTRACT

**Identification of SNPs potentially associated with di-isocyanate asthma using a luciferase reporter assay**
**Wentao Jiang**

Occupational Asthma (OA) is a common lung disorder that can be caused or aggravated by exposures and conditions in the work places. Di-isocyanate asthma (DA) is one type of OA. Despite years of research, there are no reliable markers to predict risk or susceptibility for DA. Previous studies have identified 23 candidate single nucleotide polymorphisms (SNPs) that are potentially associated with genes that are related with DA. In this study, we tested the functional relevance of the 23 candidate SNPs (mostly intronic) in several DA-related genes. We hypothesized that some of these SNPs may affect the binding of relevant transcription factors, causing altered expression of these genes. We inserted short DNA fragments (~30 bp) containing the SNPs and long DNA fragments containing multiple SNPs into the pGL3-promoter firefly luciferase reporter vector and transfected them into adenocarcinomas human alveolar basal epithelial cells (A549 cell line). The expression levels of different SNP constructs were measured by luciferase reporter assay. Our results showed that 3 SNPs (rs11571537, rs2287231 and rs2446824) that are located in ATF3, TAR1 and CDH17 genes, respectively, had significant difference in luciferase activities between risk and non-risk alleles (C vs. T for rs11571537, A vs. G for rs2287231, and T vs. C for rs2446824). Our data indicate that the SNPs we found may affect the expression of ATF3, TAR1 and CDH17 genes, which may play important roles in the development of the DA.

# Acknowledgements

Foremost, I would like to express my sincere gratitude to the chair of my advisory committee Dr. Jianbo Yao, for the continuous support of my master study and research, for his patience, motivation, encouragement, and immense knowledge. His guidance helped me in all the time of research and writing of this thesis.

I acknowledge the support by my thesis committee members Dr. Daniel G. Panaccione and Dr. Vagner A. Benedito for their enthusiasm on teaching and insightful comments on my thesis. Certainly their suggestions have made me a better speaker, writer and thinker.

I would like to thank my colleagues in the lab, Mingxiang and Prasanthi, for the stimulating discussions, for the friendly lab environment, and for all the fun we have had in the past year.

Last but not the least, I would like to thank my girlfriend and my parents, for supporting me throughout my life. I wish I could show them just how much I love and appreciate them.

# Dedication

To Mom and Dad

# Table of Contents

# List of Tables

# List of Figures

# Introduction

Occupational Asthma (OA) is one type of asthma that is work-related. It is the most common lung disorder in many developed countries (Meyer et al., 1999； Contreras, Rousseau and Chan-Yeung, 1994； Reijula and Patterson, 1994). This disease can be caused or aggravated by exposures and conditions in the workplaces. OA has two main types: immunologic and non-immunologic. It can be differentiated by the appearance after a latency period. Di-isocyanate asthma (DA) is one type of immunological asthma caused by isocyanate chemicals. Di-isocyanate chemicals are a family of chemical building blocks mainly used to make polyurethane products. Toluene di-isocyanate (TDI) and methylene diphenyl di-isocyanate (MDI) are the main di-isocyanates. TDI is used in the production of flexible foams. MDI is used in the production of a variety of polyurethane products like elastomers, sealants, adhesives, and coatings (McDonald, 2000). The isocyanates have become the most common agent causing OA, and nearly 10 percent OA cases are due to the isocyanate chemicals (Chan-Yeung and Malo, 1995).

SNPs are the most common form of genomic variations. They occur once every several hundred base pairs throughout the human genome. The vast majority of SNPs are neutral allelic variants. However, the few that do influence a phenotype in a measurable way are important for understanding the underlying genetics of human health. SNPs can be divided into three types. They may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions. Synonymous and

nonsynonymous SNPs are SNPs in the coding region. Synonymous SNPs do not affect the protein sequence while nonsynonymous SNPs change the sequence of the protein. SNPs that are not in protein-coding regions are called eSNPs (expression SNPs). They may still affect gene splicing, transcription factor binding, messenger RNA degradation, or the sequence of non-coding RNA (Katkoori, 2008).

Several genes, as well as environmental factors, have shown interactions with OA (Mapp, 2003). OA is a good model for investigating the gene-environment interactions (Park and Frew, 2002). To find the gene interaction with a specific disease, linkage analysis is the way for identifying genetic regions which may cause disease. Using candidate loci linkage disequilibrium mapping, finding single nucleotide polymorphisms (SNPs) on suspicious genes might be helpful to understand mechanisms of asthma (Palmer and Cookson, 2001). A preliminary study recruited workers with occupational asthma caused by di-isocyanates at occupational health clinics in Canada and Spain. The next-generation sequencing was performed in 91 workers with confirmed DA (cases) and compared to 293 subjects with sequencing data from the 1,000 genomic (1KG) control data set and 143 significant SNPs associated with DA were found (Lummus et al., 2017).

In this study, we examined top-ranked 23 SNPs from the preliminary study (Table 1) to evaluate their effort on gene expression. We hypothesized these SNPs may contain potential transcriptional elements that may affect transcriptional factor binding, thus affecting gene expression. The candidate genes are known to have strong relationships with di-isocyanates induced occupational asthma (Table 2). The main

purpose of this study was to identify SNPs that may affect the expression level of these genes using a luciferase reporter assay.

# Literature Review

## *1. Occupational Asthma*

### History and Definition

Occupational asthma (OA) is a common lung disorder. It is one of the most common occupational lung diseases in many industrialized countries (Meredith and Nordman, 1996). OA is worth to study because it has the potential to provide the information about the effect of genetic, environmental and behavioral interactions in onset of asthma (Maestrelli, 2004). Although OA was recognized in the early 18$^{th}$ century, the definition of OA has met with difficulty until more recently. From textbook, it briefly defines: occupational asthma is a disease characterized by variable airflow limitation and/or airway hyperresponsiveness due to causes and conditions attributable to a particular occupational environment and not to stimuli encountered outside the workplace (Chan-Yeung, 1995). There are two types of OA; they are differentiated by the appearance after a latency period-immunologic and nonimmunologic. Immunologic OA appears after a variable period of worker exposure in causal agent and acquires immunologically mediated sensitization. Immunologic asthma can be classified into that caused by high-molecular or low-molecular-weight compounds. The high-molecular-weight (>5000 DA) compounds are an animal and plant proteins or polysaccharides such as wheat flour and animal dander that cause an Immunoglobulin E (IgE)-dependent immune response. The low-molecular-weight (<5,000 DA)

compounds are chemicals such as di-isocyanates and western red cedar that can also initiate an immune response after repeated inhalation. While non-immunologic OA appears without a latent period, it happens after exposure to high concentration of workplace irritant. This reaction also called Irritant-induced asthma clinically (Bardana, 1999) (Tarlo and Broder, 1989). The airway dysfunction syndrome (RADS) is typical irritant-induced asthma, which may occur after a single exposure to high levels of an irritating vapor, fume, or smoke (Brooks, Weiss and Bernstein, 1985).

**Epidemiology and Pathophysiology**

In recent years, many large population-based studies of OA and information on the occupation have been carried out. Such investigations point the risk of OA and contribute to studies of OA. OA accounted for 26% of all work-related respiratory disease reported to the Surveillance of Work and Occupational Respiratory Disease (SWORD) program in the UK (Bakerly et al., 2008). In the United States, the analysis of 1978 social security disability data point around 15% of asthma is attributed to workplace exposure (Blanc, 1987). Moreover, six communities in Canada showed the frequency of possible OA was as high as 36.1% (Allen, 2002).

Inhaled agents (High-molecular-weight or Low-molecular-weight) in a workplace environment could lead to asthma by sensitization, normally, by creating airway inflammation or by irritant reflex pathways. Until now, more than 250 agents have been defined as causing immunologic OA (Chan-Yeung and Malo, 1994). The pathophysiology of immunologic OA can be divided as IgE-Dependent and IgE

Independent. Most high-molecular-weight agents such as flour and animal or plant protein induce asthma by producing IgE antibodies. Some specific low-molecular-weights agents such as platinum salts and acid anhydrides also induce IgE antibodies, may acting as haptens, working with body protein to form functional antigens (Baur and Czuppon, 1995) (Biagini et al., 1985). The unique IgE antibodies on the surface of different types of cells (Mast cell, Eosinophil, Macrophage, Neutrophil, dendritic cell, etc.) combine with allergens giving rise to a cascade of events which result in cell inflammation or form inflammatory mediators causing inflammation reaction. Other low-molecular-weight agents such as isocyanates and plicatic acid can cause OA without consistently induce IgE antibodies, the pathophysiology is still unclear (Sastre, Vandenplas and Park, 2003). The mechanism of RADS is unknown. It has a postulation that extensive denudation of the epithelium results in airway inflammation (Leroyer et al., 1998).

**Diagnosis and Management**

The diagnosis of OA has to establish a relation between asthma and work (Anees, 2003) (Moscato, Malo and Bernstein, 2003). It should be based on a compatible history and the presence of variable airflow limitation and lung volume. Making diagnosis need the presence of both intermittent respiratory symptoms such as cough, chest tightness, wheezing and physiologic evidence of reversible airways obstruction or hyperresponsiveness. A non-specific bronchial hyperreactivity test is one way be used to help diagnose occupational asthma. The IgE test is performed to evaluate whether

the subject is allergic to specific substances (de Groene et al., 2012). A spirometer also can use to measure timed expired and inspired volumes can help to diagnose occupational asthma. It is important to realize that no single test can diagnosis OA. The best way to manage OA is the removal of the worker from further exposure and making a substituted environment. If the substitution is not possible, changing the ventilation system in the workplace or ongoing maintenance of engineering control are necessary. Anti-inflammatory agents like corticosteroids, LKTRA or mast cell stabilizers can be used depending on the severity of the case.

## *2. SNPs and GWAS study*

**SNPs**

Single-nucleotide polymorphism (SNP) is a variation in a single nucleotide of a DNA sequence at a specific position. It is occurring when a single nucleotide in the genome or other shared sequence differs between members of a species or paired chromosomes in an individual. The nucleotides change from purine to purine called transition and from purine to pyrimidine called transversion. For example, two DNA sequence fragments from different individuals, ACAGCTA to ACACCTA, contain one single nucleotide difference. Most common SNPs have two alleles. In this case, there are two alleles: G and C. There is always different frequency between two alleles, the one with lower frequency at a locus in a specific position called minor allele, while the one with a higher frequency called major allele. Single nucleotides can be changed

(substitution), removed (deletions) or added (insertion) to a polynucleotide sequence. SNPs may be located on the coding, non-coding or intergenic region. Even if SNPs locate on the coding region, it does not mean amino acid sequence change. This type of SNPs can encode the same amino acid sequence, amking it a synonymous SNP. If a different amino acid is produced, the SNPs is termed nonsynonymous. SNPs may also generate stop codons, in this case, they are termed nonsense. SNPs that do not locate in protein-coding regions may still have consequences for gene splicing, transcription factor binding, or the sequence of non-coding RNA. SNPs occur once in every 300 nucleotides on average, which means there are nearly 10 million SNPs in the human genome. Normally, these variations are found in the DNA between genes. They can act as molecular markers, helping scientists locate genes that are associated with the disease. When SNPs occur within a gene or in a regulatory region near a gene, they may play a more direct role in disease by affecting the gene's function.

**GWAS**

For years, researchers have tried to use genetic markers to identify genetic risk factors for common and complex human disease. Genome-wide association study (GWAS) is a powerful tool for the examination of a genome-wide set of genetic variants in different individuals to see if any variant is associated with a phenotype or trait. The ultimate goal of GWAS study is to use genetic risk factors to make predictions about one individual is at risk and to identify the biological cornerstone of disease susceptibility for developing prevention even treatment strategies (Bush and Moore,

2012). SNPs are ideal markers for GWAS study, because SNPs are notably a type of common genetic variation; it is easy to find a relation between genotype and phenotype.

How to capture common variation? There is a hypothesis named common disease/ common variant (CD/CV) that states common disorders are likely influenced by genetic variation that is also common in the population (Bush and Moore, 2012). Over a five-year test of this CD/CV hypothesis, for most of the common diseases, this hypothesis is true. Based on this hypothesis, a project named Human Haplotype Map (HapMap) builds a systematic approach to interrogate the common variation in the human genome; the project included 11 human population with genotypes of 1.6 million SNPs (Altshuler et al., 2010). Because of the HapMap genotype data, the examination of linkage disequilibrium (LD) is available. LD is a property of SNPs on a stretch of genomic sequence that describes the degree to which an allele of one SNP is inherited or correlated with an allele of another SNP, or even more than two SNPs within a population. There are two common approaches to measuring LD: D' and $r^2$; the equations is showing below:

$$D' = \left\{ \frac{P_{AB}P_{ab} - P_{Ab}P_{aB}}{\min(P_A P_b, P_a P_B)} \; if \; P_{AB}P_{ab} - P_{Ab}P_{aB} > 0 \right\}$$
$$\left\{ \frac{P_{AB}P_{ab} - P_{Ab}P_{aB}}{\min(P_A P_B, P_a P_b)} \; if \; P_{AB}P_{ab} - P_{Ab}P_{aB} < 0 \right\}$$

$$r^2 = \frac{(P_{AB}P_{ab} - P_{Ab}P_{aB})^2}{P_A P_B P_a P_b}$$

P represents the frequency, if $r^2$ is high, it means two or more SNPs convey similar information and only one of SNPs needs to be genotyped to capture the allelic variation. LD can prevent genotyping SNPs provide redundant information.

Combining all the information from HapMap and LD analysis, and other details such as indirect association, the chip-based microarray technology makes possible assaying one million or more SNPs for GWAS research. The two most famous platforms for GWAS are Illumina (San Diego, CA) and Affymetrix (Santa Clara, CA). However, it is important to note that the genomic variation measurement changing rapidly, more secrets of human disease phenotype will find its linking line with genotype.

The primary approach of GWAS is the case-control setup. It compares two large groups of individuals, one healthy control group and one case group with the disease. All individuals in each group are genotyped for the majority of common known SNPs. If the allele frequency is significantly different between the case and the control, the SNPs then need to do more investigations. In such setups, the odds ratio will show the effect size. The odds of disease for individuals having a specific allele and the odds of disease for individuals who do not have that same allele divide together. The odds ratio is higher than one if the allele frequency in the case group is much higher than in the control group, and vice versa. Additionally, a P-value for the significance of the odds ratio is typically calculated using a simple chi-squared test. Finding odds ratios that are significant is prove of the SNPs is associated with the disease.

From the development of GWAS, we can tell this study has a huge impact on the field of human genetics and the knowledge of the common human disease. In the future, large data management system and cheap sequencing technologies will help GWAS go up to a higher level.

**Genes related to occupational asthma**

Genetic factors that relate with occupational asthma have begun to be explored for a long time (Mapp et al., 1994). Understanding the interaction between genetic factors and the environment is the best way to characterize OA. The first pool or gene family found associated with OA is major histocompatibility complex genes on chromosome 6p. It encodes the HLA class II molecules that requires binding of antigen to a T-cell receptor to start a series cascade reaction for antibody response. The major histocompatibility complex class II proteins are the important factors for the specific response to occupational agent, such as di-isocyanates, acid anhydrides, natural rubber latex, western red cedar and animal proteins (Horne et al., 2000) (Rihs et al., 2002) (Jeal et al., 2003). One research study indicates that about 40% of OA in the population examined could be attributed to an HLA-DRβ1*07 phenotype (Jeal et al., 2003). In general, HLA class II molecules provide evidence of a specific immunologic response in asthma induced by low-molecular-weight reagents.

The second pool of genes is the superfamily of glutathione S-transferase (GST). It comprises a family of eukaryotic and prokaryotic phase II metabolic isozymes best known for their ability to catalyze the conjugation of the reduced form of glutathione

to xenobiotic substrates for the purpose of detoxification. This family mainly function is protecting cells from oxidative stress products. One research indicates that patients who exposed to TDI for 10 or more years have a lower frequency of the GSTP1 Val/Val genotype compared to the subjects who had asthma. Data suggest that homozygosity for the GSTP1*Val allele confers protection against TDI-induced asthma and airway hyperresponsiveness (Mapp et al., 2002). Moreover, besides GST, N-acetyltransferase genotypes can slow down acetylation genotypes posed a 7.77 fold greater risk of asthma, especially TDI-induced asthma (Wikman et al., 2002). Because OA is not a simple clinical disease but has to consider multiple factors, efforts should be made to use the genetic information carefully.

## 3. Eukaryotic Transcription Regulation

Transcription regulation means the controls during the conversion from DNA to RNA, thereby orchestrating gene activities. In a eukaryote, the phenotypic differences that distinguish the many types of cells are majorly due to differences in the expression of genes that code for proteins. The transcription processes play an important role in the regulation of gene expression. Six potential control points during transcription by RNA polymerase II are from the following series: activation of gene structure (open chromatin), initiation and elongation, processing transcription, transport to the cytoplasm from the nucleus, translation, degradation and turnover to mRNA. Eukaryotes have three RNA polymerases: Pol I, Pol II, and Pol III. Each of these three

polymerases has specific targets and activities. They can work with each other or be regulated by independent mechanisms. The mechanisms can be grouped into three areas: (1) Gates for polymerases access to genes. This includes the functions of histone remodeling enzymes, enhancers and repressors, transcription factors, and many other complexes. (2) The engine of elongation. It can help to escape the promoter complex and making a productive elongation. (3) The brake of transcription. Many factors can control when and how termination occurs. All three of this system working accordingly and making transcription accomplish successfully. Unlike prokaryotic system, eukaryotes have a restrictive basal state which requires the recruitment of other factors. This difference is largely due to the compaction of the eukaryotic genome by winding DNA around histones to form higher order structures.

**Relationship of Chromatin Structure to Transcription**

In eukaryotes, the DNA is not "naked," instead, it highly packed as chromatin to fit in the nucleus. This process is accomplished by wrapping DNA on an octamer structure protein called histone. The one single histone wrapped with 146 base pair of DNA is nucleosome. It is basic units of chromatin. With these highly packed characteristics, eukaryotes genes are at default status "silence." To activate gene expression, that is, initiation transcription, the chromatin structure should be opened. The change of structure precedes the act of transcription and indicates that the gene can be transcribed. Chromatin remodeling is the general process of inducing changes in chromatin structure. It requires repositioning or displacing histones through recruiting

energy. There are several types of chromatin remodeling: (1) Histone slides along DNA. This can alter rotationally and the translational position of the particular sequence on the nucleosome. (2) The spacing may change between two octamer histones. (3) The octamer histones may displace entirely thereby releasing “naked” DNA. The histone structure has a pair of tails on the end of core histone, which can be modified by histone acetyltransferases (HATs), histone methyltransferases (HMTs), and histone deacetylases (HDACs), etc. These enzymes can add or remove covalent modifications such as methyl groups, acetyl groups, phosphates, and ubiquitin. Usually, histone acetylation is associated with transcription activation, and deacetylation is associated with repression. As for methylation, generally, represents a feature of inactive chromatin. However, some methylation also can activate transcription, such as di- or tri-methylation of H3K4 is associated with transcriptional activation. Histone phosphorylation also can affect chromatin structure, linking with transcriptional repairing, chromosome condensation, and cell-cycle progression.

**Activators and Repressors**

The initiation of transcription requires a complicated protein-protein interaction system that features transcription factors bound at enhancers with the basal apparatus that assembles at the promotor including the RNA polymerase. The transcription factors may stimulate transcription or delay transcription, the one with a positive effect called activators, the one with a negative effect called repressors.

Activators can be classified into three types by its function. The first class of activator is true activator which contacts with basal apparatus at promotor either directly or indirectly by co-activator. They all function on DNA or chromatin template. The second class is the anti-repressor. It functions to unwind chromatin from closed status to open status by recruiting the histone modifier enzymes or the chromatin remodeling complexes. It can only function on chromatin template. The third class is architectural proteins that can bend strand DNA and bring two or more functional areas together. Regarding repressor, there are four possible illustrations of the mechanism. First, is sequestering an activator in the cytoplasm, the repressor will mask activator pass through the cytoplasm. The second one is repressor in nucleus masks activators which have already bind on promotor. Third is repressor masking and holding in the cytoplasm until it is released to enter the nucleus. The last one is competing with the enhancer.

In many cases, activators may not bind directly at the promotor area because they lack a transcription activation domain. However, they can bind another protein that has a transcription activation domain. An activation domain works by making protein-protein interaction contacts with general transcription factors that promote assembly of basal apparatus. Basal apparatus contains basal factor: $TF_{II}D$, $TF_{II}B$, or $TF_{II}A$. All components required for efficient transcription are basal apparatus, RNA polymerase, activators, and co-activators. This component consists of ~40 proteins. They will combine to form a big complex.

## *4. Luciferase Assay*

Luciferase is a class of oxidative enzymes that produce bioluminescence. Using this characteristic, researchers found a powerful and adaptable tools for cell biology research-luciferase assay. It is commonly using on eukaryotic gene expression at transcription level due to its convenience, relatively inexpensive and instantaneously quantitative measurement. Luciferase assay technology can be applied to many areas.

It is important to find optic-reporter genes for the research of gene expression and cellular events during the gene expression. Typically, a reporter gene should be cloned with a DNA sequence of interest into an expression vector thereby transfect to the specific cells. A good reporter gene is easy to be identified and can be measured quantitatively. For a long time finding of bioluminescence, the luciferase from firefly (*Photinus Pyralis*) and renilla (*Renilla reniform*) are well known in detail and commonly using for dual-reporter assays. The reason to use dual-reporter assay is to improve experimental accuracy. The “dual reporter” represents two individual reporter enzymes within a single system measured and expressed simultaneously. Because typically the experimental condition should be the same, the co-transfection of two enzymes can act as an internal control and serve a baseline response. Normalizing the activity of the experimental reporter to the activity of the internal control minimizes

experimental variability caused by differences in cell viability or transfection efficiency. Through the dual reporter system, the external influence of experiment can be largely removed. Also, because of the difference between Firefly and Renilla luciferases enzyme structures and substrates requirements, makes it possible to selectively discriminate between their bioluminescence reactions. Firefly luciferase is a 61 kDa monomeric protein, and Renilla luciferase is a 36 kDa monomeric protein. They both do not require post-translational processing for enzymatic activities (Wood et al., 1984) (de Wet et al., 1985). Thus, they can function as a reporter immediately. The reactions of both Firefly and Renilla luciferase show in Figure 1.

Luciferase assay has been developed for many research applications. The most common application of luciferase assay is analyzing the function of cis-acting genetic elements, especially promoters. Deletion or mutation will be made on promoter regions and then measured by luciferase assay. For example, the human factor VIII gene promoter region was found through luciferase assay, which contains around 200bp (Figueiredo and Brownlee, 1995). The second application is focused on SNPs researches. Through comparing two different allele luciferase activities, the functional allele can be recognized. The real power of luciferase technology comes from its ability to be used as cellular readouts for virtually any signaling event. By coupling a response element controlled by the signaling event of interest to luciferase, researchers can point the intracellular events due to quantified analysis. Luciferase assay can also help research of RNA interference (RNAi). Once the reporter assay cell line established, the signaling pathway of specific molecules can be detected.

Those applications mentioned before are just a few of the broad range of luciferase assay application. It is clear that the luciferase assay is a powerful and versatile tool for gene expression studies as well as studies of other cellular components and events.

## Objectives of the Study

A next-generation DNA sequencing analysis of di-isocyanate asthma cases compared to control was made by a preliminary study. Twenty-three out of one hundreds and forty-three SNPs were identified for further study. The SNPs are contained in seven loci: CDH17 (10), ATF3 (6), FAM71A (2), PITPNC1, TACR1, ZBTB16, LOC101929565. Based on the locations of the 23 SNPs, many of them are located near each other. So we also designed two haplotypes long fragments containing those linked SNPs as two new samples. In total, we have twenty-three SNPs and two long fragments that contain several SNPs for this study. We hypothesized that some of those risk SNPs will affect the binding of the relevant transcription factors which may cause the expression changes of DA-related genes. The specific goal is to determine if any of the risk SNPs are located in putative transcriptional elements that may affect the binding of the relevant transcription factors which may change expression of genes related to di-isocyanate induced occupation asthma using luciferase reporter assay.

# Material and Method

## *SNP fragment design and annealing*

Forty-six pairs of DNA oligonucleotides were designed and ordered from oligo making company (Integrated DNA Technologies, Inc. IA) according to twenty-three SNPs show on Table 1. For each SNP, two pairs of oligonucleotides around 25 nanomole DNA oligonucleotides having XhoI and KpnI restriction sites that contain risk allele or non-risk allele were designed, an example of fragment design is shown in Figure 2. Each oligo was spun down by mini centrifuge for 1 minute and diluted to 100 nM with molecular biology grade water. Forty-six pairs of oligonucleotides then were incubated at 4℃ for one hour. For oligonucleotides annealing, each pair of oligonucleotides were added 10 ul into PCR tube separately. PCR tubes later were incubated in a PCR machine at 95℃ for 5 minutes and then cooled down to room temperature (25℃) over 45 minutes. All tubes were kept in 4℃ fridges for storage until use. Table 4 lists all 23 SNP insert sequences.

## *PCR amplification and gel purification*

The long DNA region contains 8 different SNPs (rs2513787:8:94114852, rs2446824:8:94115346, rs2446823:8:94115384, rs149630836:8:94115608, rs2513788:8:94115668, rs2513789:8:94115885, rs2513790:8:94115919, rs2446821:8:94116251) that are in linkage disequilibrium and located in 5’ end of the CDH17 gene was amplified by PCR. The primers were designed according to the

human genome. The DNA sample of an OA patient with heterozygotes was selected as PCR template. PCR amplification was carried out with the reaction mixture composed of 1 µl of diluted DNA sample of OA patient, 2.5 µl of gene-specific forward primer (5 µM) and reverse primer (5 µM) (see Table 3 for the list of primer sequences), 1.5 µl of MgCl2 (25 mM), 0.5 µl of dNTP (10 mM), 5 µl of 5x PCR buffer, and 0.25 µl of Go Taq® DNA polymerase (5 u/µl, Promega) with the final volume adjusted to 25 µl with $ddH_2O$. Amplification was carried out in a thermal cycler with 5 minutes of denaturation at 94°C, followed by 35 cycles of 30 seconds of denaturing at 94°C, 1 minute and 30 seconds of annealing at 59°C, 1 minute of extension at 72°C with a final extension at 72°C for 10 minutes. A negative control was always included and the amplified products were mixed with loading buffer and separated on a 1% agarose gel and stained with 1:10,000 diluted RGB nucleic acid stain (Phenix). The gel that contained PCR amplified products was cut by scalpel accordingly and purified by using QIAquick® Gel Extraction Kit (QIAGEN, Valencia, CA) following the manufacturer's instructions. The purified products then were diluted by 30 ul ddH2O, and the concentration was measured by NanoDrop™ 2000 Spectrophotometer.

## *TA cloning and double digestion*

The pGEM®-T Easy Vector (Promega, Madison, WI, USA) was used as the vector for cloning PCR amplified products. Insert and vector were incubated with 2X Rapid Ligation Buffer and T4 ligase (Promega, 3 u/µl) overnight at 4°C. The Escherichia coli competent cells (Novagen) (> 1.5 x 108 CFU/µg) were transformed

with 1 µl of ligation mixture using a heat shock method involving incubation on ice for 5 minutes, followed by a heat shock for 30 seconds in a 42°C water bath, and then on ice for another 2 minutes. The transformation mixture was plated on LB 1% carbenicillin agar plates. Plates were incubated overnight at 37°C. Four clones were selected and inoculated in 3 ml of LB broth with 1% carbenicillin overnight in shaking incubator at 37°C with 250 rpm. Plasmids were isolated from the overnight grown culture using a QIAprep miniprep plasmid isolation kit (QIAGEN, Valencia, CA) following the manufacturer's instruction. The isolation products later were sent to DNA sequencing with T7 and SP6 primers. The double digestion performed by adding 2.5 ul of each XhoI and KpnI enzyme. Five µl of restriction buffer 1.1 was used according to the information from Double Digest Finder (www.neb.com). All the enzymes, buffer 1.1 and 2.5 µg of the TA cloning sample was added into 1.5 ml Eppendorf tube with the final volume adjusted to 50 µl with dd$H_2O$ (For each of TA cloning sample). The tubes later were sent to 37℃ incubators for 3 hours. After 3 hours incubation, the digestion samples were added with loading buffer and separated on a 1% agarose gel and stained with 1:10,000 diluted RGB nucleic acid stain (Phenix). The double digestion result was observed by comparing 1kb ladder (NEB, Ipswich, MA, USA). The restriction products again were purified by using QIAquick® Gel Extraction Kit (QIAGEN, Valencia, CA) following the instruction. The resulting sample was finally stored at 4℃.

## *Cloning of expression reporter constructs*

The pGL3-promoter Firefly Luciferase reporter vector (Promega, Madison, WI, USA) was selected as the final reporter assay vector through comparing with other similar reporter vectors. The detailed information of pGL3-promoter Firefly Luciferase reporter vector is shown in Figure 3. Then, 0.25 µg of pGL3 reporter vector was digested by 2.5 µl of each XhoI and KpnI enzyme and 5 µl of restriction buffer 1.1 into the total volume adjusted to 50 µl with $ddH_2O$ and then kept into 37℃ incubators for 3 hours. The product was added with loading buffer and separated on a 1% agarose gel and stained with 1:10,000 diluted RGB nucleic acid stain (Phenix). All the SNPs fragments and the PCR amplified product were used as inserts for cloning. The ligations were made with 3:1 molar ratio for the insert ratio to the vector. Two µl of T4 DNA Ligase Buffer (10X) (Promega) was used to each of reactions and 1 µl of T4 ligase (Promega, three u/µl) was added at the last steps of reaction with the final volume adjusted to 20 µl for each of reactions. All the cloning samples were incubated in 4℃ overnight. Escherichia coli competent cells (Novagen) (> 1.5 x 108 CFU/µg) were transformed with 1 µl of ligation mixture using a heat shock method involving incubation on ice for 30 minutes, followed by a heat shock for 30 seconds in a 42°C water bath, and then on ice for another 2 minutes. The transformation mixture was plated on LB 1% carbenicillin agar plates. Plates were incubated overnight at 37°C. Eight clones of each plate were selected and restricted on LB 1% carbenicillin agar plates for inoculation after PCR checking. Simultaneously, eight clones of each plate also were used as a template for cloning PCR to check if cloning were a success or not.

For cloning PCR, forward primers were all forward oligonucleotides of each SNPs, reverse primer was designed as a universal primer that located at the shared sequence of pGL3 reporter vector. Each pair of primers were specifically generated a short fragment around 205 bp for the 23 SNPs and 1496 bp for long fragments if the ligations were succeeded. PCR amplification was carried out in a thermal cycler with 5 minutes of denaturation at 94°C, followed by 35 cycles of 30 seconds of denaturing at 94°C, 1 minute and 30 seconds of annealing at 59°C, 1 minute of extension at 72°C with a final extension at 72°C for 10 minutes. Figure 4 shows some examples of how plates were designed as eight pieces for eight clones and the cloning PCR results on 1% agarose gel. The verified clones were inoculated the next day and cultured in 3 ml of LB broth with 1% carbenicillin overnight in shaking incubator at 37°C with 250 rpm. Plasmids were isolated from the overnight grown culture using a QIAprep miniprep plasmid isolation kit (QIAGEN, Valencia, CA) and QIAfilter plasmid midi kit (QIAGEN, Valencia, CA) following manufacturer's instructions. The plasmids isolated with the mini kit were then sent to DNA sequencing.

## *Cell Culture and Transfection*

A549 cells were cultured in DMEM–Dulbecco's Modified Eagle Medium (Promega, Madison, WI, USA), supplemented with 10% fetal calf serum (Life Technologies). A549 cells were digested by trypsin for 5 minutes at 37℃ when it covered plate surface nearly 100%. Then A549 cells were plated in 96 well plates one day before transfection. The amount of each cell in each well was adjusted according

to transfection protocol. Transfections was carried out with Lipofectamine® LTX Reagent (Invitrogen, CA). Each of the Pgl3-Promoter Vector containing the each specific SNP (180 ng) was co-transfected with 20 ng of pRL Renilla Luciferase Control Reporter Vectors pRL-CMV, (Promega, Madison, WI, USA). All the processes of cell culture and transfection were conducted under sterile condition.

## *Luciferase Assay*

Luciferase activity was measured using the Dual-Luciferase Assay System (Promega, Madison, WI) following the manufacturer's instructions. Briefly, after transfection, cells were incubated at 37 °C in a $CO_2$ incubator for 48 hours. Cells were then lysed in the buffer provided by the manufacturer and mixed on a shaker for 15 min at room temperature. The firefly and renilla luciferase activities of the cell lysates were measured separately on a Synergy HT 96 microplate luminometer (Biotech, Winooski, Vermont, USA).

## *Statistics analysis*

All experiments were performed at least three times. To normalize luciferase activities, the renilla luciferase activities were measured as normalization reference. Shapiro-Wilk W test was made to make sure that the residuals of expression data are normally distributed across their treatment means at $\alpha$=0.05. Comparisons were performed using Fisher's least significant difference (LSD) test following one-way analysis of variance (ANOVA) for the different groups. We also did Benjamini-

Hochberg adjustment (Q=0.1) for all the normalized data in order to control the false discovery rate. Results are expressed as means ± S.D.

# Results

## *Cloning of haplotypes of CDH17-1469bp long fragments*

The DNA sample of an OA heterozygous patient was used as a template for PCR amplification of the CDH17-1469 bp long fragments. The primers were designed to flank the two sides of the target region. The sequences of primers are shown in Table 3. The size of PCR product was around 1500 bp indicating that the PCR worked. The PCR product was purified and used as the insert for TA cloning. Because the amplification of the CDH17-1469 bp fragment was based on a heterozygous template, two different haplotypes were expected. To identify two different haplotypes, plasmids from four colonies were isolated for DNA sequencing. The two different haplotypes of CDH17-1469bp were verified by sequencing analysis (CDH17-1469-OA11-2: C allele for rs2513787 and G for rs2446821; CDH17-1469-OA11-4: An allele for rs2513787 and T for rs2446821). For further reporter assay construct cloning, the sticky restriction sites of two haplotypes CDH17-1469-bp long fragments were made through double digestion (Figure 4).

## *Cloning of the reporter assay construct*

The 23 SNP short fragments and two haplotypes CDH17-1469-bp long fragments were cloned into the pGL3-promoter Firefly Luciferase reporter vector (Figure 5A). To ensure cloning success before plasmid isolation, colony PCR for each fragment was performed. An example of colony PCR result is shown in Figure 5B (Other 46 samples were also made); the CDH17-1469-bp long fragments colony PCR

result is shown in Figure 5C. The expected band sizes of PCR products means cloning of the reporter assay construct was successful. Isolations of all plasmids were made and sent for DNA sequencing. All the sequences of clones prepared in pGL3-promoter Firefly Luciferase reporter vector were confirmed by DNA sequencing.

## *Quantitative analysis of Luciferase assay*

The A549 cell line was cultured in DMEM with fragments transfection. The 23 short fragments and the two haplotypes of the CDH17-1469-bp long fragments were transfected with the pRL renilla luciferase control reporter vectors into the A549 cell line. The luciferase activities were measured 48 hours after transfection. Statistical analyses were used according to normalized data: 23 SNPs normalized luciferase activity values indicated each SNP normalized luciferase activity data was normally distributed through the Shapiro-Wilk W test with treatment means at a $\alpha$=0.05. CDH17-1469-OA11 long fragments normalized luciferase activity values also were normally distributed. With normally distributed data points, an one-way analysis of variance (ANOVA) is appropriate.

## *SNPs associated with di-isocyanate asthma*

To determine if any SNP affected transcription factor binding, we used Fisher's least significant difference (LSD) test following by one-way analysis of variance (ANOVA) and Benjamini-Hochberg adjustment (Q=0.1) for all the normalized data. We found 3 SNPs out of 23 SNPs (T vs. C for rs11571537, A vs. G for rs2287231, and A vs. G for rs2446824) have significant difference in normalized luciferase activities

between risk and non-risk alleles (*p < 0.025, *p < 0.01, *p < 0.01), as Figure 6A, Figure 6B and Figure 6C show respectively. The two haplotypes of CDH17-1469-OA11 showed no significant difference between each other, but both haplotypes showed significant difference expression level (*p < 0.01) compare to the pGL-3-promoter control group (Figure 6D). We were unable to find any significant difference for the rest of the 20 SNPs (Figure 7). These findings support the hypothesis that some SNPs out of 23 SNPs are associated with di-isocyanate asthma.

# Discussion

Occupational asthma is an inflammatory disease caused by environmental and genetic factors. Some studies identified and characterized the interaction between SNPs and inflammatory disease (Kim et al., 2009) (Tokuhiro et al., 2003). Based on GWAS research before, one hundred and forty-three SNPs were identified that show association with DA, of which all but one SNP were located in intronic regions. From transcriptomic analysis that used available transcription factor (TF) datasets from relevant lung-derived tissues and cell lines, including ChIP-seq datasets for TFs, regulatory histone markers, and DNase-seq (open chromatin), the top-ranked 23 SNPs were identified based on the number of datasets they overlap and the biological relevance of the associated SNPs (Lummus et al., 2017). Based on the locations of the 23 SNPs, many of them are located near to each other, so we also designed two haplotypes long fragments containing those linked SNPs as two new samples. The possible roles of these variants on gene expression and protein function are under investigation in separate study. In this study, out of these 23 SNPs, we identified 3 SNPs that may be considered as genetic factors related with DA.

The SNP rs2287231 is located upstream of the TAR1 gene. TAR1 was found in mitochondria, coded within the 25S rRNA gene on the opposite strand. It may be involved in mtDNA stability or mitochondrial gene expression regulation at the post-

transcriptional level (Coelho, 2002). It seems like TAR1 gene has no interaction with OA. Nevertheless, the TAR1 protein may interact with other proteins as a co-activator.

The SNP rs2446824 is located downstream of CDH17 gene. This gene belongs to the cadherin superfamily that encodes calcium-dependent, membrane-associated glycoproteins. In human, CDH17 gene encodes cadherin 17 protein, which is a component of the gastrointestinal tract and pancreatic ducts. This protein may play an important role in the morphological organization of liver and intestine (Ncbi.nlm.nih.gov, 2017). Perhaps the shape of pancreatic ducts has some influence on OA. A GWAS study also indicated that SNP (rs2514805) near the CDH17 genes on chromosome 8 has a significant association with DA (Yucesoy et al., 2015). This evidence suggests CDH17 gene expression may be affected by SNPs near the CDH17 gene. Although both SNP rs2287231 and SNP rs2446824 are not located on promoter regions, transcription regulation is not a simple process of binding transcription factors with promotor. The recruitment of proteins that contain different function is a critical step for a successful transcription. Therefore, SNPs around a gene may also affect the regulation of gene expression.

The SNP rs11571537 locates on the first intron of activating transcription factor 3 (ATF3) gene. ATF3 is a member of the ATF/cyclic AMP response element-binding family of a transcription factor. The ATF family represents a large group of basic-region leucine zipper (bZIP) transcription factors. The common feature of the ATF family is they all can use bZIP to form homodimers or heterodimers with other bZIP

that contains other proteins to recognize specific DNA binding in the basic region (Hai, 2006). Since ATF3 most times has to work with other proteins, it is hard to define it as an activator or repressor (Hai et al., 1999). Studies point out negative feedback regulation of ATF3 as the key to protect acute inflammatory syndromes through limiting pro-inflammatory cytokine expression, and ATF is critical in the suppression of inflammatory responses to infection and allergy (Whitmore et al., 2007). The most remarkable research that connects with our study is in a mouse model of ovalbumin allergen-induced asthma. ATF3 expression was shown to be significantly increased (Gilchrist et al., 2008). From these data, it is clear that the ATF3 gene plays an important role in the inflammatory disease. More specifically, it has a connection with OA. ATF3 acts like a switch for stimulating transcription processes by preventing the access of inhibitory co-factors to gene promoter. It is possible that alternative splicing of the ATF3 gene can be physiologically critical in the regulation of target genes.

In this study, we only studied the SNPs at the gene expression level without protein expression level research. The magnitude of difference in transcriptional processes cannot be defined. Based on these data, further research on DA can determine if the identified risk SNPs are located in any putative transcription elements that may affect the binding of the relevant transcription factors. The transcription element search software [(TESS) http://www.cbil.upenn.edu/tess/] can analyze the promoter sequence of this three genes, and the electrophoretic mobility shift assay (EMSA) can measure

the capacity of any identified transcription elements to bind candidate transcription factors and to investigate the influence of variations on transcription factor binding.

In addition, the long fragment CDH17-1469bp-OA11 that contains eight SNPs is a test model that we may use for the future research. According to the description of eukaryotic transcription regulation (Krebs et al., 2014), we believe the transcription of a specific gene can be regulated by more than one SNP. A haplotype, more specifically, probably decides the amount of gene expression more than just one SNP. Even these two haplotypes in this study did not show a significant difference. The higher expression than original pGL3-promoter reporter vector still indicates expression level changed because of the participation of eight SNPs. The more groups of haplotypes will be tested in the future.

In conclusion, our data indicate that these three SNPs (rs11571537, rs2287231, and rs2446824) located inside of ATF3 gene, upstream of TAR1 gene and downstream of CDH17 gene, respectively, cause a significant difference in luciferase activities between risk and non-risk alleles (C vs. T for rs11571537, A vs. G for rs2287231, and T vs. C for rs2446824) and may play important roles in the development of the DA.

# Tables and Figures

**Table 1. Top ranked SNPs associated with di-isocyanate-induced occupational asthma (OA) based on sequence data analysis**

| No. | Marker | SNP | Chr | Position | Minor Allele | Major Allele | Risk Allele | Non-Risk Allele | $X^2$ - $log_{10}$ P DA centro | $X^2$ - $log_{10}$ P 1KG centro |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | SNV* | rs1001304 | 1 | 212732686 | T | C | C | T | 1.744854 | 3.501355 |
| 2 | SNV* | rs72756369 | 1 | 212770271 | A | T | A | T | 0.831740 | 3.466039 |
| 3 | SNV* | rs11571537 | 1 | 212787025 | C | T | C | T | 1.192560 | 3.320238 |
| 4 | SNV* | rs11571559 | 1 | 212794550 | T | C | T | C | 1.147106 | 4.057164 |
| 5 | SNV* | rs11571563 | 1 | 212794873 | G | T | G | T | 1.147106 | 3.552191 |
| 6 | SNV* | rs74138575 | 1 | 212794997 | A | G | A | G | 1.147106 | 4.057164 |
| 7 | SNV* | rs75465959 | 1 | 212795196 | A | G | A | G | 1.147106 | 4.057164 |
| 8 | Deletion | rs147978008 | 1 | 212796008 | - | ACA | - | ACA | 1.147106 | 4.057164 |
| 9 | SNV* | rs17019510 | 1 | 212816729 | G | A | G | A | 1.299957 | 5.337089 |
| 10 | SNV* | rs2287231 | 2 | 75449129 | G | A | A | G | 2.449367 | 4.366918 |
| 11 | SNV* | rs2446824 | 8 | 95127574 | T | C | T | C | 1.877903 | 4.523108 |
| 12 | SNV* | rs2446823 | 8 | 95127612 | G | T | G | T | 1.877903 | 4.523108 |
| 13 | Deletion | rs149630836 | 8 | 95127836 | - | TCAGTAG | - | TCAGTAG | 1.956805 | 4.317313 |
| 14 | SNV* | rs2513788 | 8 | 95127896 | G | T | G | T | 1.877903 | 4.523108 |
| 15 | SNV* | rs2513789 | 8 | 95128113 | A | G | A | G | 1.877903 | 4.523108 |
| 16 | SNV* | rs2513790 | 8 | 95128147 | C | T | C | T | 1.877903 | 4.523108 |
| 17 | SNV* | rs2446821 | 8 | 95128479 | T | G | T | G | 1.877903 | 4.523108 |
| 18 | SNV* | rs2513791 | 8 | 95128529 | A | C | A | C | 1.877903 | 4.523108 |
| 19 | SNV* | rs117579120 | 8 | 95159359 | A | C | A | C | 1.557530 | 3.422331 |
| 20 | SNV* | rs2251996 | 8 | 95160288 | C | G | C | G | 1.877903 | 4.793522 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 21 | SNV* | rs1672692 | 11 | 113945609 | A | G | A | G | 2.323669 | 4.370613 |
| 22 | SNV* | rs62084077 | 17 | 65386837 | A | G | A | G | 1.583248 | 3.026485 |

*****Single Nucleotide Variant

**&**p values for DA+ occupational asthma cases (n=91) compared to diisocyanate exposed asymptomatic (DA) control subjects (n=53) or to controls subjects found from the 1000 genomes (1KG) database (n=?).

**Table 2. SNPs for which transcription factor binding sites have been identified in cell lines**

| No. | SNP | Gene | DNA strand | Direction To Gene | Associations with Known | Transcription Factors identified |
|---|---|---|---|---|---|---|
| 1 | rs1001304 | LOC101929565 | + | downstream | 6000 bp upstream of ATF3; lots of TFs and histone marks - potential | P130,RBL2, RUNX1, POLR2A |
| 2 | rs72756369 | ATF3 | + | Inside gene | | P130 |
| 3 | rs11571537 | ATF3 | + | Inside gene | inside ATF3 1st | SPI1, POLR2A |
| 4 | rs11571559 | ATF3 | + | downstream | | SpI1,RUNX1, |
| 5 | rs11571563 | ATF3 | + | downstream | inside RUNX peak, | RUNX1, POLR2A |
| 6 | rs74138575 | ATF3 | + | downstream | inside RUNX peak, | RUNX1, POLR2A |
| 7 | rs75465959 | ATF3 | + | downstream | inside RUNX peak, but no predicted RUNX binding site | RUNX1, POLR2A,MAFK |
| 8 | rs147978008 | FAM71A | + | upstream | | SPI1, POLR2A |
| 9 | rs17019510 | FAM71A | + | downstream | | USF1 |
| 10 | rs2287231 | TACR1 | - | upstream | | SPI1 |
| 11 | rs2446824 | CDH17 | - | downstream | | SPI1, CEPBP |
| 12 | rs2446823 | CDH17 | - | downstream | | SPI1, POLR2A |
| 13 | rs149630836 | CDH17 | - | downstream | Risk allele is a 6 bp deletion | RELA, SPI1, EP300, CEPBP,FOSL2 |
| 14 | rs2513788 | CDH17 | - | downstream | inside CDX2 ChIP peak, and predicted to affect CDX2 | RELA, SPI1, EP300, CEPBP,FOSL2 |
| 15 | rs2513789 | CDH17 | - | downstream | | SPI1, CEPBP,FOSL2 |
| 16 | rs2513790 | CDH17 | - | downstream | | P130 |
| 17 | rs2446821 | CDH17 | - | downstream | | SPI1,P130 |
| 18 | rs2513791 | CDH17 | - | downstream | | P130 |
| 19 | rs117579120 | CDH17 | - | Inside gene | | SPI1, EP300 |
| 20 | rs2251996 | CDH17 | - | Inside gene | potentially right | RUNX1 |
| 21 | rs1672692 | ZBTB16 | + | Inside gene | ~10 bases away from | RUNX1 |
| 22 | rs62084077 | PITPNC1 | + | Inside gene | inside 1st intron; lots of TFs binding here; inside MAX | SPI1,RUNX1,NR3C1, POLR2A, RUNX3 |

**Table 3. Primers used in the study.**

| PCR method | Primer name | Primer sequences |
|---|---|---|
| Regular PCR | 1LOC-A-F | CCATCTGGTTAACCAAAGAGGGCTTCCAGGAGC |
| | 1LOC-G-F | CCATCTGGTTAACCAAGGAGGGCTTCCAGGAGC |
| | 2ATF3-A-F | CAGTGTGGAGCCTCTTACTGTCTTGTTCATTCC |
| | 2ATF3-T-F | CAGTGTGGAGCCTCTTTCTGTCTTGTTCATTCC |
| | 3ATF3-C-F | CACCACAGATCCCCGGCTGAGAGGAATGCCCAC |
| | 3ATF3-T-F | CACCACAGATCCCCGGTTGAGAGGAATGCCCAC |
| | 4ATF3-C-F | CTGTTTCATGTCTAAACGTGTGTCTCATTGTTC |
| | 4ATF3-T-F | CTGTTTCATGTCTAAATGTGTGTCTCATTGTTC |
| | 5ATF3-G-F | CTTGACCTCATGTATGGTTCCTTTAAATGATCC |
| | 5ATF3-T-F | CTTGACCTCATGTATGTTTCCTTTAAATGATCC |
| | 6ATF3-A-F | CAGCACTTCTTTCAGGAGATAATTTATGTTCCC |
| | 6ATF3-G-F | CAGCACTTCTTTCAGGGGATAATTTATGTTCCC |
| | 7ATF3-A-F | CTGATGTGTATGGTAAATGTAAATGAAGGTCTC |
| | 7ATF3-G-F | CTGATGTGTATGGTAAGTGTAAATGAAGGTCTC |
| | 8FAM71A-Del-F | CTCCCATTTAATCCTCACAACTGCTACCCCCC |
| | 8FAM71A-Ins-F | CTCCCATTTAATCCTCACAACAACTGCTACCCCCC |
| | 9FAM71A-A-F | CACACCACATGCCCAGATAGAGATGTACCTGTC |
| | 9FAM71A-G-F | CACACCACATGCCCAGGTAGAGATGTACCTGTC |
| | 10TACR1-A-F | CCACTTGGCTGATGAAAGACAACATTGGCAGGC |
| | 10TACR1-G-F | CCACTTGGCTGATGAAGGACAACATTGGCAGGC |
| | 11CDH17-A-F | CCATTTTTTCCATTGCAACATTTGGGCAGATGC |
| | 11CDH17-G-F | CCATTTTTTCCATTGCGACATTTGGGCAGATGC |
| | 12CDH17-A-F | CCAGAAGCAAAAGTTCATCTGACTGAGGGAAAC |
| | 12CDH17-C-F | CCAGAAGCAAAAGTTCCTCTGACTGAGGGAAAC |
| | 13CDH17-del-F | CGAATTGCCTACTATCTCAGTAGCCCACCATC |
| Regular PCR | 13CDH17-Ins-F | CGAATTGCCTACTATCTCAGTAGTCAGTAGCCCACCATC |
| | 14CDH17-G-F | CGCCCATCCTCCTTGTGTTATGGGCCCAGATGC |
| | 14CDH17-T-F | CGCCCATCCTCCTTGTTTTATGGGCCCAGATGC |

| PCR method | Primer name | Primer sequences |
|---|---|---|
| | 15CDH17-A-F | CTGTTTTTTCATTGTCAATGTTATCTAAGTCAC |
| | 15CDH17-G-F | CTGTTTTTTCATTGTCGATGTTATCTAAGTCAC |
| | 16CDH17-C-F | CTTCCAAGCACATACTCAAAAGAATCACTTGAC |
| | 16CDH17-T-F | CTTCCAAGCACATACTTAAAAGAATCACTTGAC |
| | 17CDH17-A-F | CGGTCCTCCTTCAGAAAAACATCAGCTTCCCAC |
| | 17CDH17-C-F | CGGTCCTCCTTCAGAACAACATCAGCTTCCCAC |
| | 18CDH17-A-F | CAGTCAGCTCAGGCCAACATAACAAAATACCAC |
| | 18CDH17-C-F | CAGTCAGCTCAGGCCACCATAACAAAATACCAC |
| | 19CDH17-A-F | CGAGAAACTCTTGGGAAACATCCAGAGTGAAGC |
| | 19CDH17-C-F | CGAGAAACTCTTGGGACACATCCAGAGTGAAGC |
| | 20CDH17-C-F | CAAGGAGCAAGTGGTGCTCTACGTGAACAGGTC |
| | 20CDH17-G-F | CAAGGAGCAAGTGGTGGTCTACGTGAACAGGTC |
| | 21ZBTB16-A-F | CCTTGTCATATGTAAAAAGGGAATCATTGTGGC |
| | 21ZBTB16-G-F | CCTTGTCATATGTAAAGAGGGAATCATTGTGGC |
| | 22PITPNC1-A-F | CAATCTCCCAGGTGGCATATTGCTAGAGGAGAC |
| | 22PITPNC1-G-F | CAATCTCCCAGGTGGCGTATTGCTAGAGGAGAC |
| | 23ATF3-del-F | CTCAGGAGGCTGAGGCAGAAGGATGGCTTGAC |
| | 23ATF3-Ins-F | CTCAGGAGGCTGAGGCAGAGAAGGATGGCTTGAC |
| | pGL3-R280 | TGGTGGCTTTACCAACAGTAC |
| | pGL3-R364 | CCTTATGCAGTTGCTCTCCAG |
| | CDH17-1496-F | GTGGCAGCTTCTCATGTGAAC |
| | CDH17-1496-R | TGGTATTTTGTTATGGTGGCCTG |
| | CDH17-1496-KpnI | GGCCGGTACCTTGTTGCAAGGATTCGGCGAC |
| | CDH17-1496-XhoI | CCGGCTCGAGCTGAGCTGACTGATACAGAATC |

## Table 4. Sequence of oligos used for preparation of SNP fragments

| SNP | Variant | Design | SNP | Variant | Design |
|---|---|---|---|---|---|
| 1LOC | 1LOC-A | CCATCTGGTTAACCAAAGAGGGCTTCCAGGAGC<br>CATGGGTAGACCAATTGGTTTCTCCCGAAGGTCCTCGA<br>GCT | 13CDH17 | CDH17-Del | CGAATTGCCTACTATCTCAGTAGCCCACCATC<br>CATGGCTTAACGGATGATAGAGTCATCGGGTGGTAGAGCT |
| | 1LOC-G | CCATCTGGTTAACCAAGGAGGGCTTCCAGGAGC<br>CATGGGTAGACCAATTGGTTCCTCCCGAAGGTCCTCGA<br>GCT | | CDH17-Ins | CGAATTGCCTACTATCTCAGTAGTCAGTAGCCCACCATC<br>CATGGCTTAACGGATGATAGAGTCATCAGTCATCGGGTGG<br>TAGAGC |
| 2ATF3 | 2ATF3-A | CAGTGTGGAGCCTCTTACTGTCTTGTTCATTCC<br>CATGGTCACACCTCGGAGAATGACAGAACAAGTAAGG<br>AGCT | 14CDH17 | 14CDH17-G | CGCCCATCCTCCTTGTGTTATGGGCCCAGATGC<br>CATGGCGGGTAGGAGGAACACAATACCCGGGTCTACGAG<br>CT |
| | 2ATF3-T | CAGTGTGGAGCCTCTTTCTGTCTTGTTCATTCC<br>CATGGTCACACCTCGGAGAAAGACAGAACAAGTAAGG<br>AGCT | | 14CDH17-T | CGCCCATCCTCCTTGTTTTATGGGCCCAGATGC<br>CATGGCGGGTAGGAGGAACAAAATACCCGGGTCTACGAG<br>CT |
| 3ATF3 | 3ATF3-C | CACCACAGATCCCCGGCTGAGAGGAATGCCCAC<br>CATGGTGGTGTCTAGGGGCCGACTCTCCTTACGGGTGA<br>GCT | 15CDH17 | 15CDH17-A | CTGTTTTTTCATTGTCAATGTTATCTAAGTCAC<br>CATGGACAAAAAAGTAACAGTTACAATAGATTCAGTGAGC<br>T |
| | 3ATF3-T | CACCACAGATCCCCGGTTGAGAGGAATGCCCAC<br>CATGGTGGTGTCTAGGGGCCAACTCTCCTTACGGGTGA<br>GCT | | 15CDH17-G | CTGTTTTTTCATTGTCGATGTTATCTAAGTCAC<br>CATGGACAAAAAAGTAACAGCTACAATAGATTCAGTGAGC<br>T |
| 4ATF3 | 4ATF3-C | CTGTTTCATGTCTAAACGTGTGTCTCATTGTTC<br>CATGGACAAAGTACAGATTTGCACACAGAGTAACAAG<br>AGCT | 16CDH17 | 16CDH17-C | CTTCCAAGCACATACTCAAAAGAATCACTTGAC<br>CATGGAAGGTTCGTGTATGAGTTTTCTTAGTGAACTGAGC<br>T |
| | 4ATF3-T | CTGTTTCATGTCTAAATGTGTGTCTCATTGTTC<br>CATGGACAAAGTACAGATTTACACACAGAGTAACAAG<br>AGCT | | 16CDH17-T | CTTCCAAGCACATACTTAAAAGAATCACTTGAC<br>CATGGAAGGTTCGTGTATGAATTTTCTTAGTGAACTGAGCT |
| 5ATF3 | 5ATF3-G | CTTGACCTCATGTATGGTTCCTTTAAATGATCC<br>CATGGAACTGGAGTACATACCAAGGAAATTTACTAGGA<br>GCT | 17CDH17 | 17CDH17-A | CGGTCCTCCTTCAGAAAAACATCAGCTTCCCAC<br>CATGGCCAGGAGGAAGTCTTTTTGTAGTCGAAGGGTGAG<br>CT |
| | 5ATF3-T | CTTGACCTCATGTATGTTTCCTTTAAATGATCC | | 17CDH17-C | CGGTCCTCCTTCAGAACAACATCAGCTTCCCAC |

| | | | | | |
|---|---|---|---|---|---|
| | | CATGGAACTGGAGTACATACAAAGGAAATTTACTAGGAGCT | | | CATGGCCAGGAGGAAGTCTTGTTGTAGTCGAAGGGTGAGCT |
| 6ATF3 | 6ATF3-A | CAGCACTTCTTTCAGGAGATAATTTATGTTCCCCATGGTCGTGAAGAAAGTCCTCTATTAAATACAAGGGAGCT | 18CDH17 | 18CDH17-A | CAGTCAGCTCAGGCCAACATAACAAAATACCACCATGGTCAGTCGAGTCCGGTTGTATTGTTTTATGGTGAGCT |
| | 6ATF3-G | CAGCACTTCTTTCAGGGGATAATTTATGTTCCCCATGGTCGTGAAGAAAGTCCCCTATTAAATACAAGGGAGCT | | 18CDH17-C | CAGTCAGCTCAGGCCACCATAACAAAATACCACCATGGTCAGTCGAGTCCGGTGGTATTGTTTTATGGTGAGCT |
| 7ATF3 | 7ATF3-A | CTGATGTGTATGGTAAATGTAAATGAAGGTCTCCATGGACTACACATACCATTTACATTTACTTCCAGAGAGCT | 19CDH17 | 19CDH17-A | CGAGAAACTCTTGGGAAACATCCAGAGTGAAGCCATGGCTCTTTGAGAACCCTTTGTAGGTCTCACTTCGAGCT |
| | 7ATF3-G | CTGATGTGTATGGTAAGTGTAAATGAAGGTCTCCATGGACTACACATACCATTAACATTTACTTCCAGAGAGCT | | 19CDH17-C | CGAGAAACTCTTGGGACACATCCAGAGTGAAGCCATGGCTCTTTGAGAACCCTGTGTAGGTCTCACTTCGAGCT |
| 8FAM71A | 8FAM71A-Del | CTCCCATTTAATCCTCACAACTGCTACCCCCCCATGGAGGGTAAATTAGGAGTGTTGACGATGGGGGGAGCT | 20CDH17 | 20CDH17-C | CAAGGAGCAAGTGGTGCTCTACGTGAACAGGTCCATGGTTCCTCGTTCACCACGAGATGCACTTGTCCAGAGCT |
| | 8FAM71A-Ins | CTCCCATTTAATCCTCACAACAACTGCTACCCCCCCATGGAGGGTAAATTAGGATGTGTGTTGACGATGGGGGGAGCT | | 20CDH17-G | CAAGGAGCAAGTGGTGGTCTACGTGAACAGGTCCATGGTTCCTCGTTCACCACCAGATGCACTTGTCCAGAGCT |
| 9FAM71A | 9FAM71A-A | CACACCACATGCCCAGATAGAGATGTACCTGTCCATGGTGTGGTGTACGGGTCTATCTCTACATGGACAGAGCT | 21ZBTB16 | 21ZBTB16-A | CCTTGTCATATGTAAAAAGGGAATCATTGTGGCCATGGGAACAGTATACATTTTTCCCTTAGTAACACCGAGCT |
| | 9FAM71A-G | CACACCACATGCCCAGGTAGAGATGTACCTGTCCATGGTGTGGTGTACGGGTCCATCTCTACATGGACAGAGCT | | 21ZBTB16-G | CCTTGTCATATGTAAAGAGGGAATCATTGTGGCCATGGGAACAGTATACATTTCTCCCTTAGTAACACCGAGCT |
| 10TACR1 | 10TACR1-A | CCACTTGGCTGATGAAAGACAACATTGGCAGGCCATGGGTGAACCGACTACTTTCTGTTGTAACCGTCCGAGCT | 22PITPNC1 | 22PITPNC1-A | CAATCTCCCAGGTGGCATATTGCTAGAGGAGACCATGGTTAGAGGGTCCACCGTATAACGATCTCCTCTGAGCT |
| | 10TACR1-G | CCACTTGGCTGATGAAGGACAACATTGGCAGGCCATGGGTGAACCGACTACTTCCTGTTGTAACCGTCCGAGCT | | 22PITPNC1-G | CAATCTCCCAGGTGGCGTATTGCTAGAGGAGACCATGGTTAGAGGGTCCACCGCATAACGATCTCCTCTGAGCT |
| 11CDH17 | 11CDH17-A | CCATTTTTTCCATTGCAACATTTGGGCAGATGC | 23ATF3 | 23ATF3-Del | CTCAGGAGGCTGAGGCAGAAGGATGGCTTGAC |

| | | | | | |
|---|---|---|---|---|---|
| | | CATGGGTAAAAAAGGTAACGTTGTAAACCCGTCTACGAGCT | | | CATGGAGTCCTCCGACTCCGTCTTCCTACCGAACTGAGCT |
| | 11CDH17-G | CCATTTTTTCCATTGCGACATTTGGGCAGATGCCATGGGTAAAAAAGGTAACGCTGTAAACCCGTCTACGAGCT | | 23ATF3-Ins | CTCAGGAGGCTGAGGCAGAGAAGGATGGCTTGACCATGGAGTCCTCCGACTCCGTCTCTTCCTACCGAACTGAGCT |
| 12CDH17 | 12CDH17-A | CCAGAAGCAAAAGTTCATCTGACTGAGGGAAACCATGGGTCTTCGTTTTCAAGTAGACTGACTCCCTTTGAGCT | | | |
| | 12CDH17-C | CCAGAAGCAAAAGTTCCTCTGACTGAGGGAAACCATGGGTCTTCGTTTTCAAGGAGACTGACTCCCTTTGAGCT | | | |

Figure 1

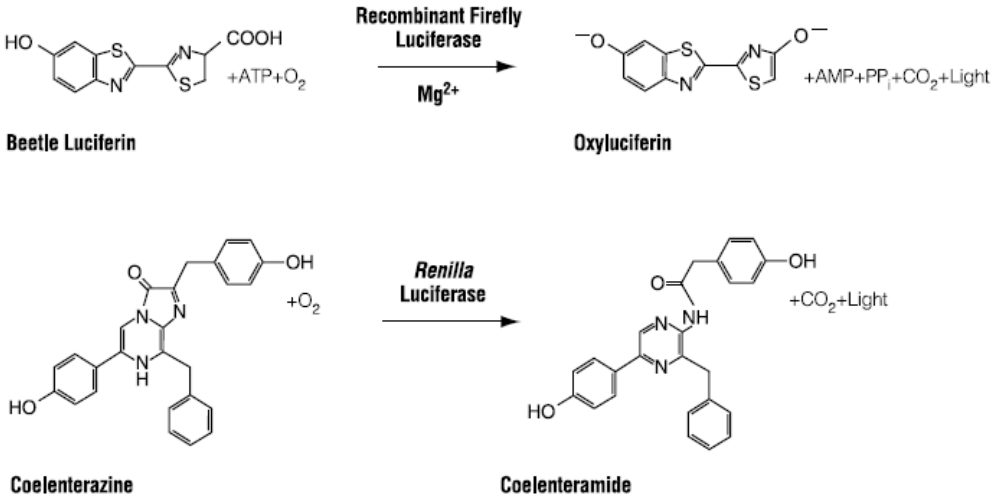



**Figure 1. Bioluminescent reactions catalyzed by firefly and Renilla luciferases.** Photon emission is achieved through oxidation of beetle luciferin in a reaction that requires ATP, Mg2+, and O2. This assay chemistry generates a "flash" of light that rapidly decays after the substrate and enzyme are mixed. The luminescent reaction catalyzed by Renilla luciferase utilizes O2 and coelenterate luciferin. (Protocol of Dual-Luciferase reporter assay system)

Figure 2

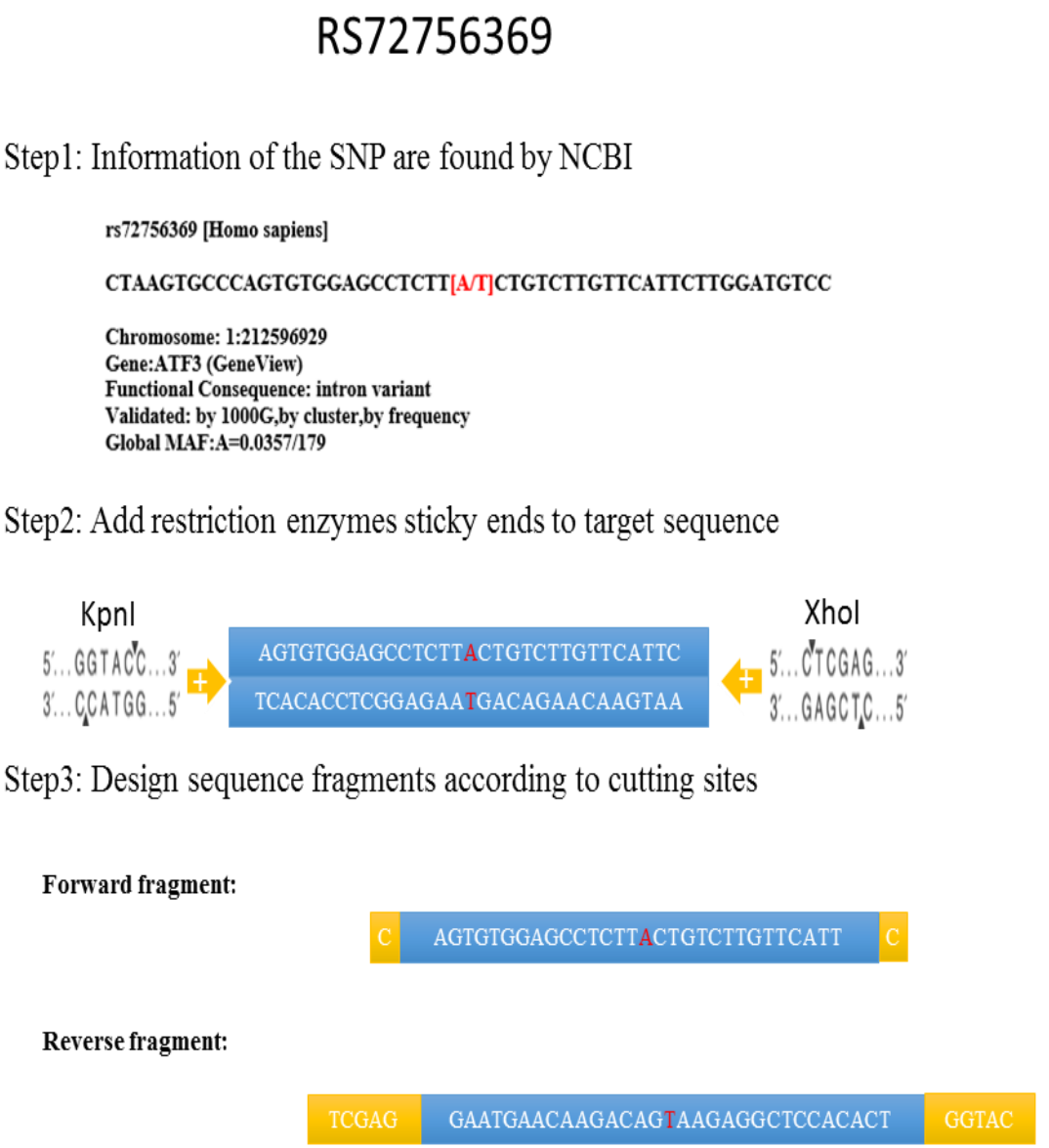



**Figure 2. An example of an SNP fragment design.** According to the information from dbSNP of NCBI (https://www.ncbi.nlm.nih.gov/snp), the name, location, sequence and detail information of an SNP are showed on top. The two restriction sites were added at each end of two alleles with sticky ends. The blue parts are the two alleles of one SNP. The yellow parts at each sequence ends are restriction sites; the yellow middle parts are the SNP variant. The sequences followed by red parts are fragments that need to order from the company.

Figure 3

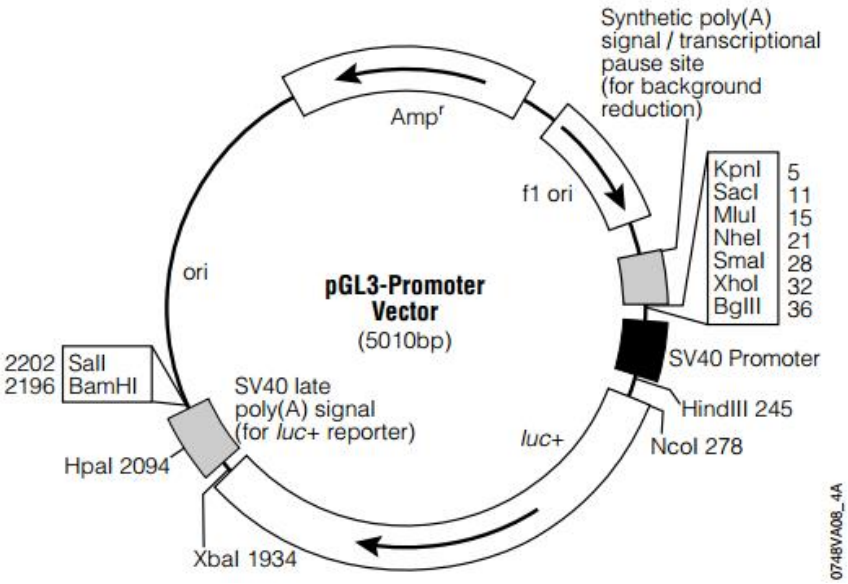



**Figure 3. The pGL3-Promoter Vector circle map.** The luc+ can encode the modified firefly luciferase; Ampr, gene conferring ampicillin resistance in Escherichia coli; f1 ori, origin of replication derived from filamentous phage; ori, origin of plasmid replication in E. coli. Arrows within luc+ and the Ampr gene indicate the direction of transcription; the arrow in f1 ori indicates the direction of ssDNA strand synthesis.

Figure 4

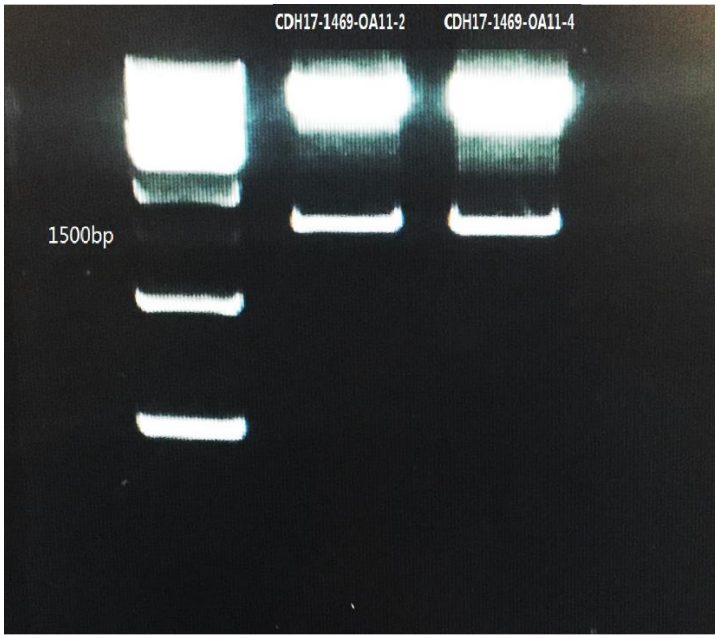



**Figure 4. Double digestion of CDH17-1469-OA11-2 and CDH17-1469-OA11-4.** The double digestion was analyzed on a 1% agarose gel. The two bands around 1500 bp proved the CDH17-1469-OA11-2 and CDH17-1469-OA11-4 were digested successfully from pGEM®-T Easy Vector and ready to be used as inserts.

Figure 5

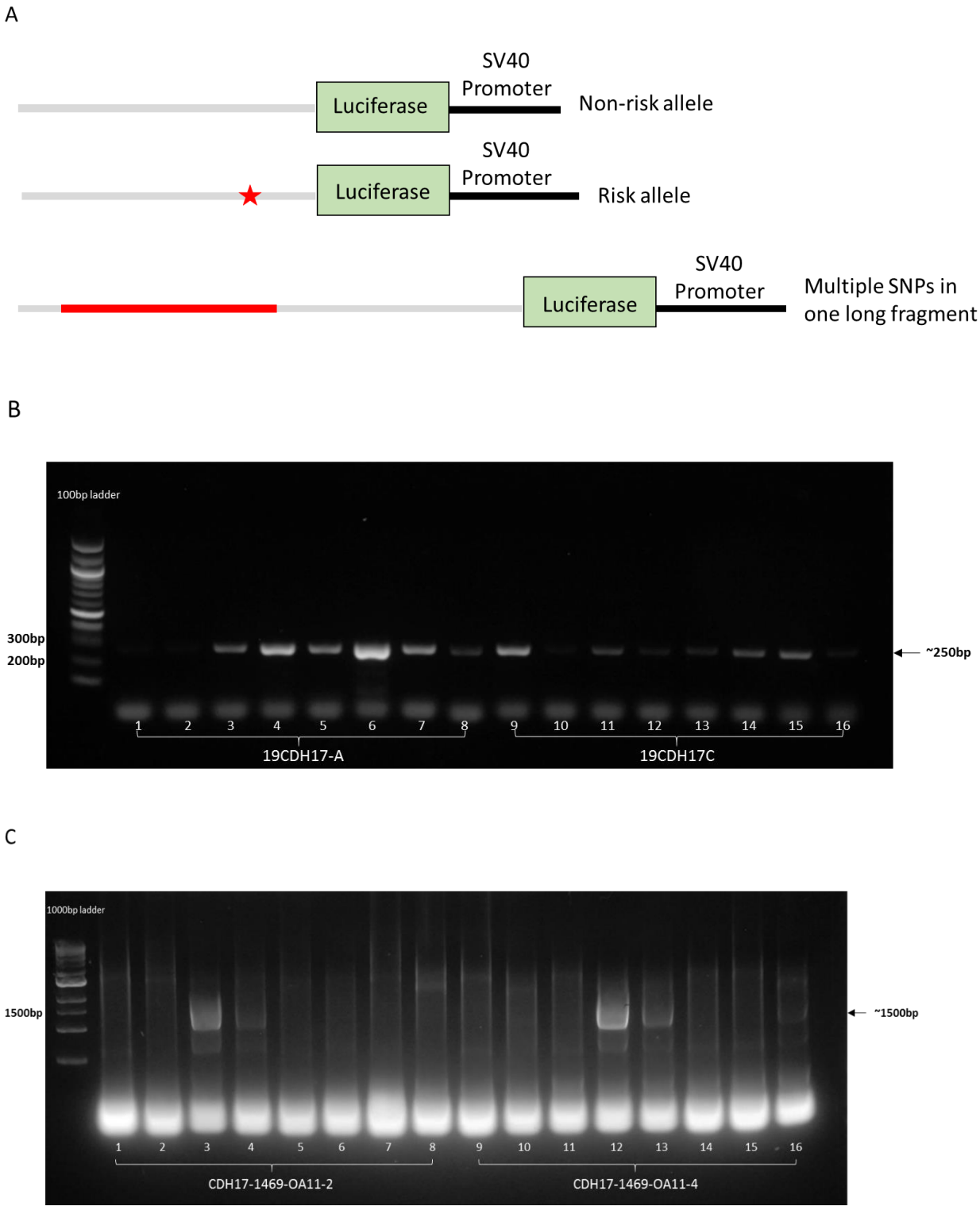
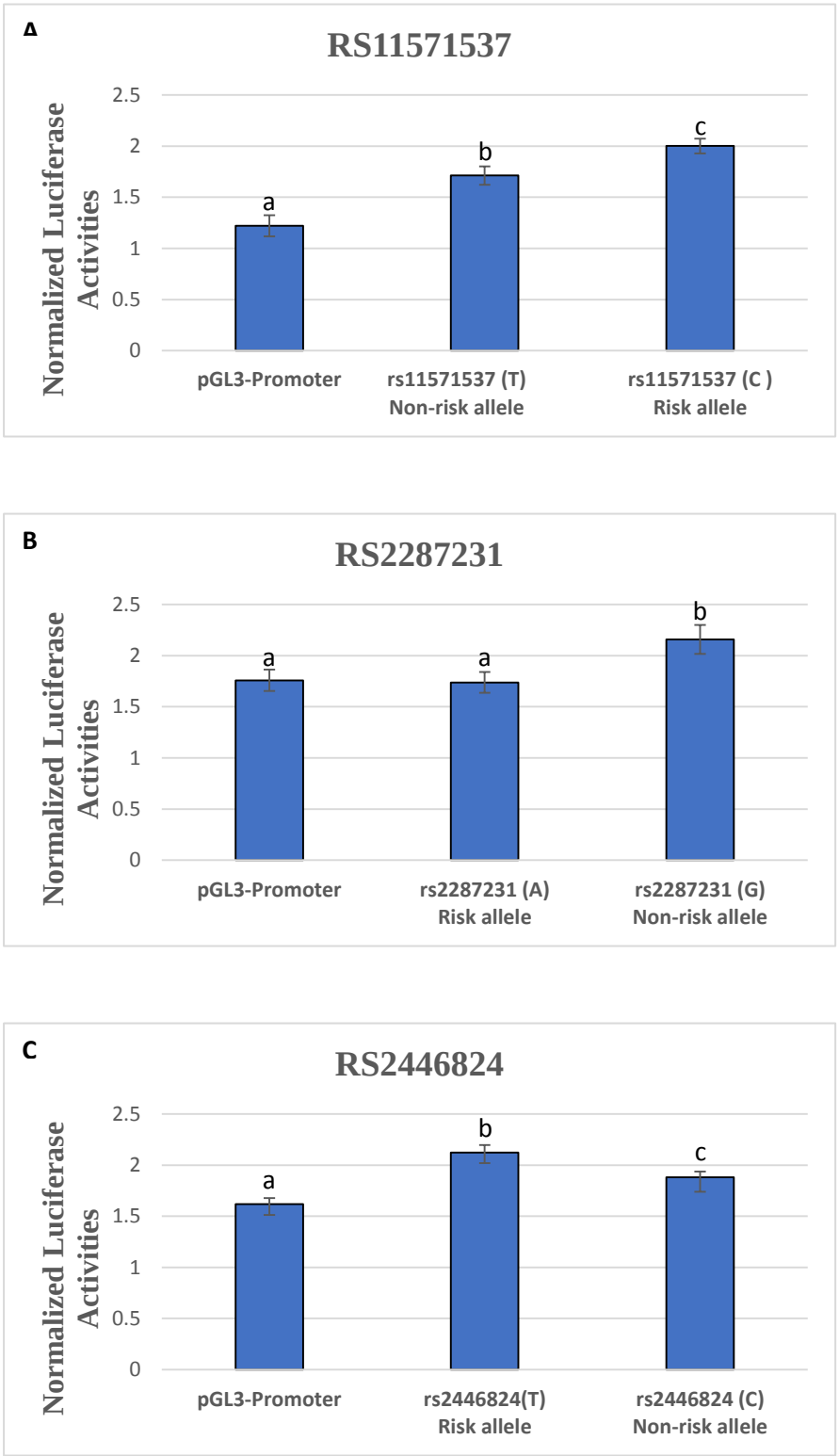



**Figure 5. Cloning of the reporter constructs and colony PCR. A.** Cloned luciferase reporter constructs containing the SNPs. Gray boxes, the 3'UTR used in the reporter constructs that contain the target SNPs or the long SNPs fragment. Red star representsthe risk allele of each SNP that may affect expression level change. Red box,

the long fragment that contains multiple SNPs, represented a haplotype. **B.** The colony PCR amplification using eight different single colonies that possibly have cloned report construct as templates showed single band around 250 bp, indicating that the cloning of report constructs are of good quality and ready for DNA sequencing. All the 46 clones were tested before DNA sequencing. **C.** The colony PCR amplification using eight different single colonies that possibly have cloned reporter construct as templates showed single band around 1500 bp, indicating that the cloning of reporter constructs are of good quality and ready for DNA sequencing.

Figure 6


A
RS11571537
Normalized Luciferase Activities
2.5
2
1.5
1
0.5
0
a
b
c
pGL3-Promoter
rs11571537 (T)
Non-risk allele
rs11571537 (C )
Risk allele
B
RS2287231
Normalized Luciferase Activities
2.5
2
1.5
1
0.5
0
a
a
b
pGL3-Promoter
rs2287231 (A)
Risk allele
rs2287231 (G)
Non-risk allele
C
RS2446824
Normalized Luciferase Activities
2.5
2
1.5
1
0.5
0
a
b
c
pGL3-Promoter
rs2446824(T)
Risk allele
rs2446824 (C)
Non-risk allele

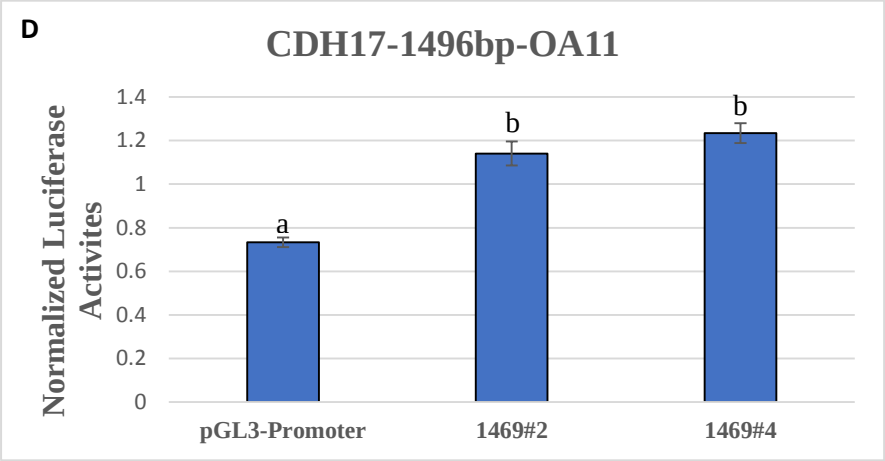
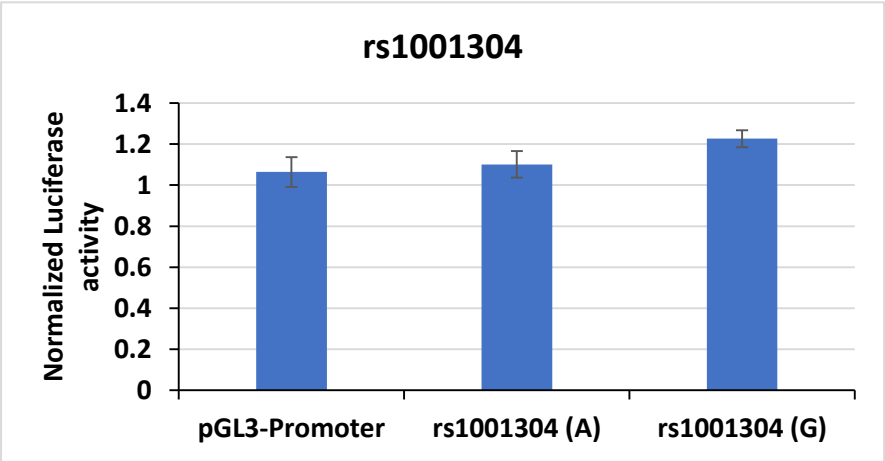
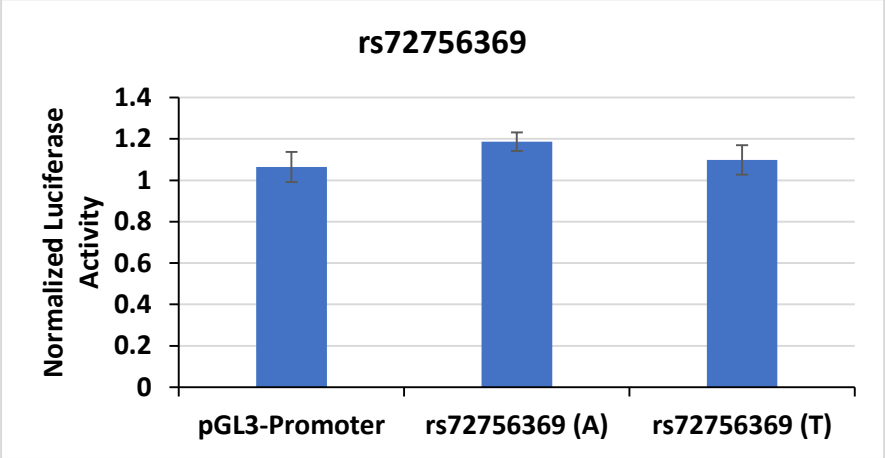
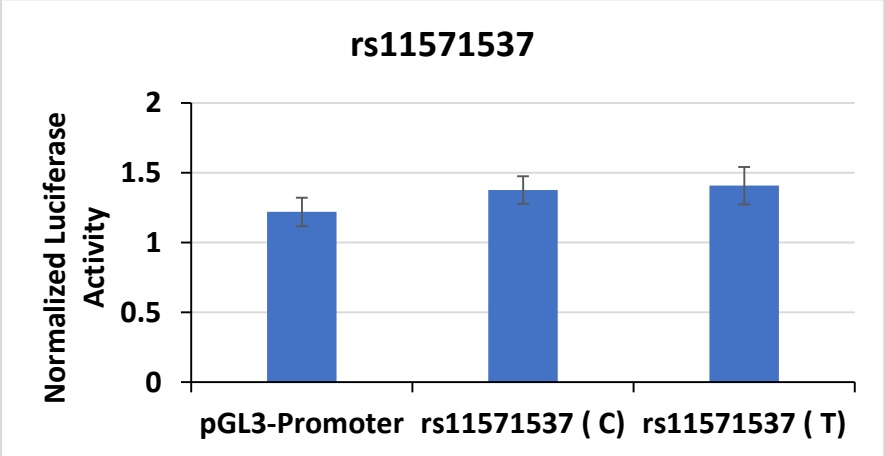


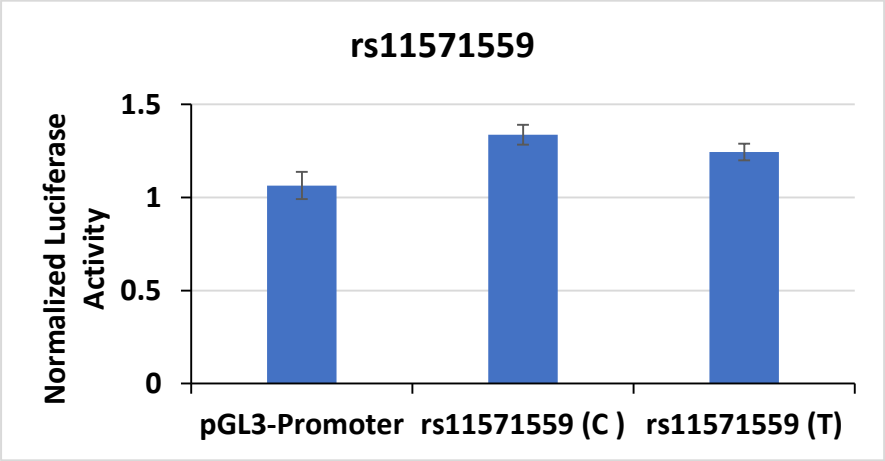

**Figure 6. Normalized luciferase activities of SNPs and CDH17-1469 bp long fragments.** **A.** Normalized luciferase activities of SNP rs11571537 between risk and non-risk alleles (C vs. T). Significant difference between two alleles was observed by Fisher's least significant difference test following one-way analysis of variance (Difference [D], 0.2880018; 95% Confidence interval [CI], 0.038-0.540, p = .025). **B.** Normalized luciferase activities of SNP rs2287231 between risk and non-risk alleles (A vs. G). Significant difference between two alleles was observed by Fisher's least significant difference test following one-way analysis of variance (Difference [D], 0.4206311; 95% Confidence interval [CI], 0.088-0.752, p = .01). **C.** Normalized luciferase activities of SNP rs2446824 between risk and non-risk alleles (T vs. C). Significant difference between two alleles was observed by Fisher's least significant difference test following one-way analysis of variance (Difference [D], 0.2407542; 95% Confidence interval [CI], 0.059-0.423, p = .01). **D.** Normalized luciferase activities of CDH17-1496bp-OA11 between two haplotypes. No significant difference between two haplotypes was observed by Fisher's least significant difference test following one-way analysis of variance. Different letters indicate statistical difference ($p<0.05$).

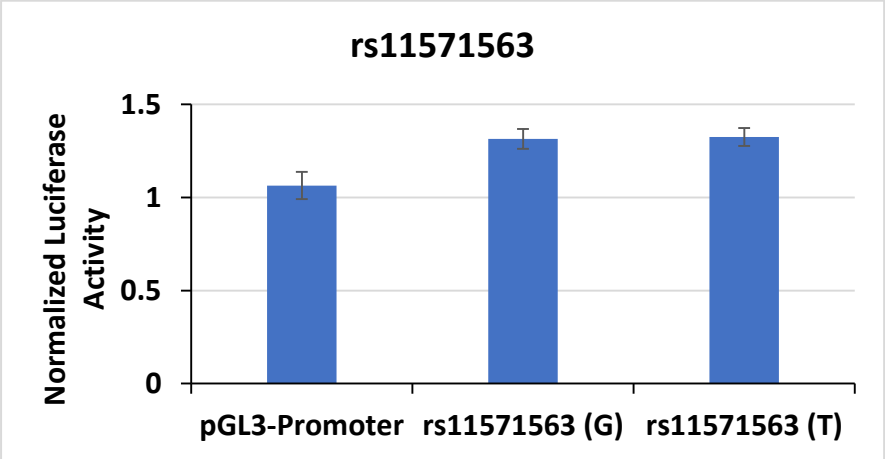
Figure 7

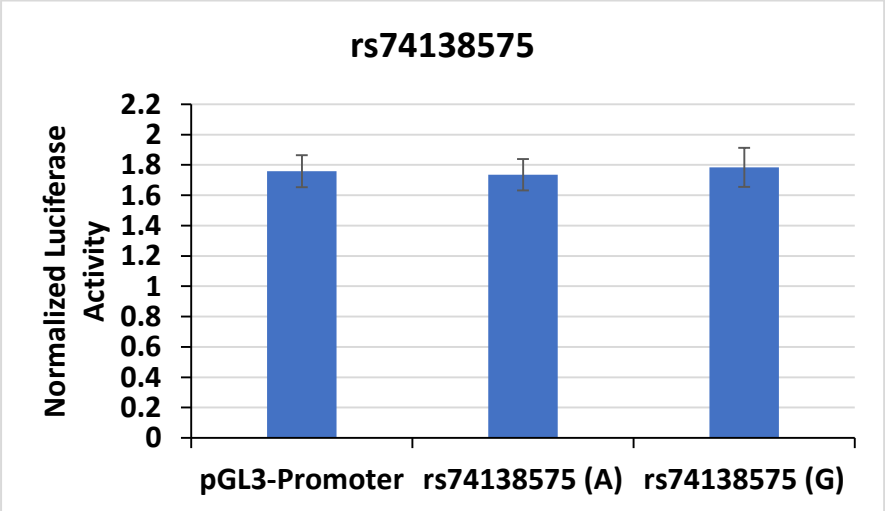

rs1001304
Normalized Luciferase activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs1001304 (A)
rs1001304 (G)

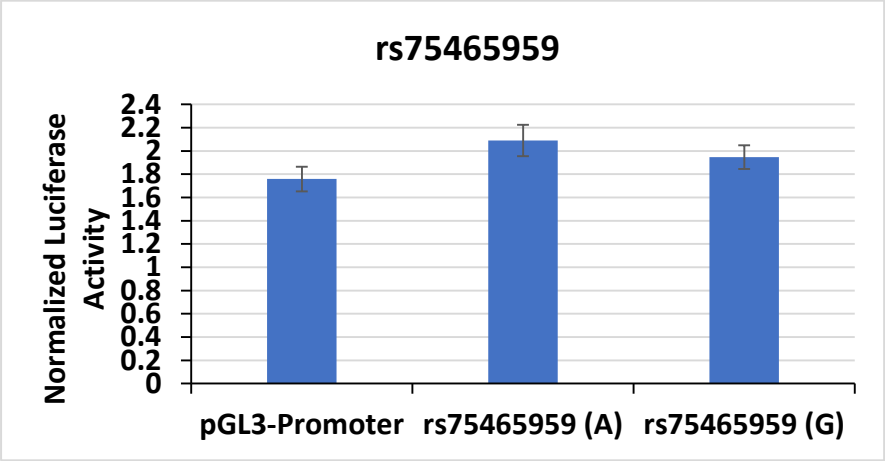

rs72756369
Normalized Luciferase Activity
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs72756369 (A)
rs72756369 (T)

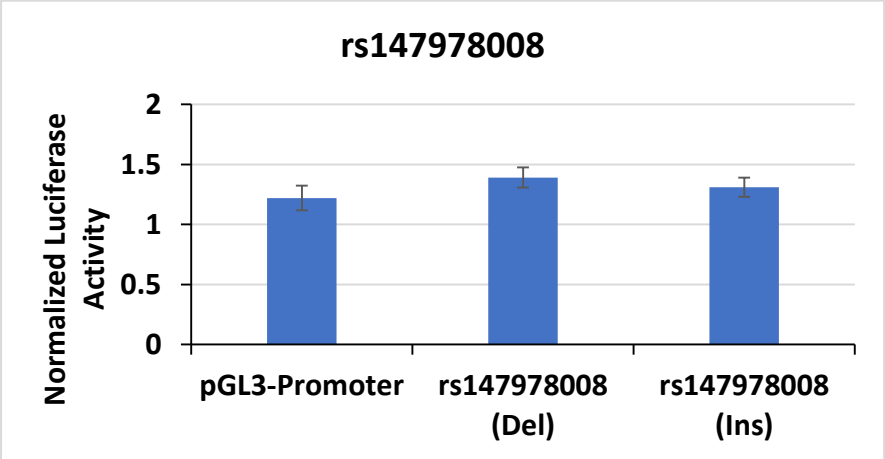

rs11571537
Normalized Luciferase Activity
2
1.5
1
0.5
0
pGL3-Promoter
rs11571537 ( C)
rs11571537 ( T)

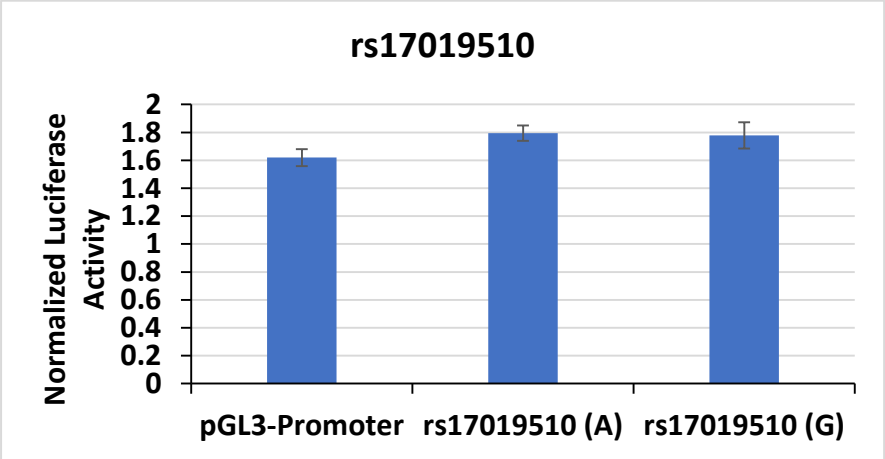

rs11571559
Normalized Luciferase Activity
1.5
1
0.5
0
pGL3-Promoter
rs11571559 (C )
rs11571559 (T)

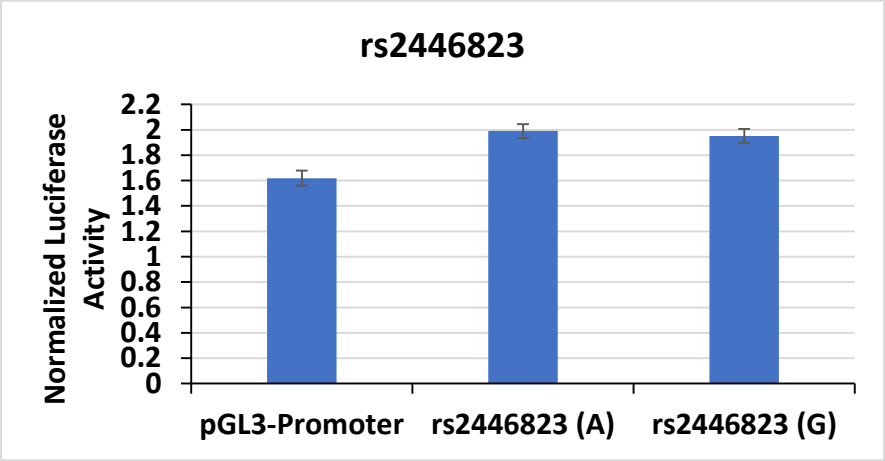

rs11571563
Normalized Luciferase Activity
1.5
1
0.5
0
pGL3-Promoter
rs11571563 (G)
rs11571563 (T)

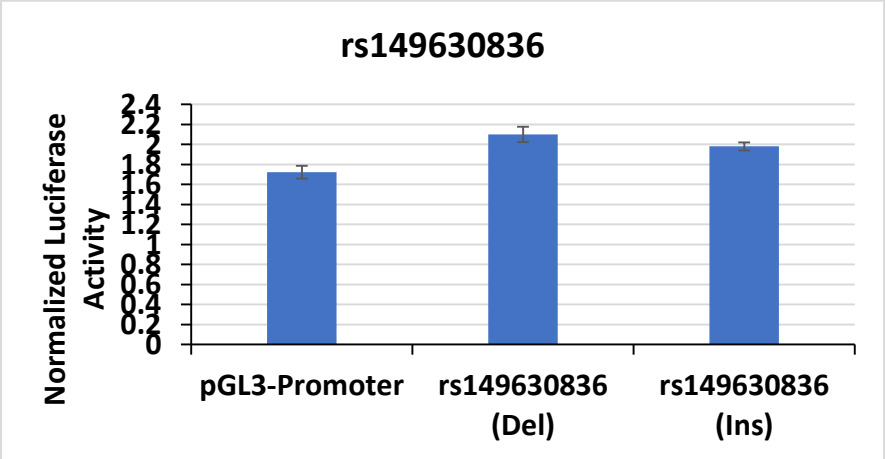

rs74138575
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs74138575 (A)
rs74138575 (G)

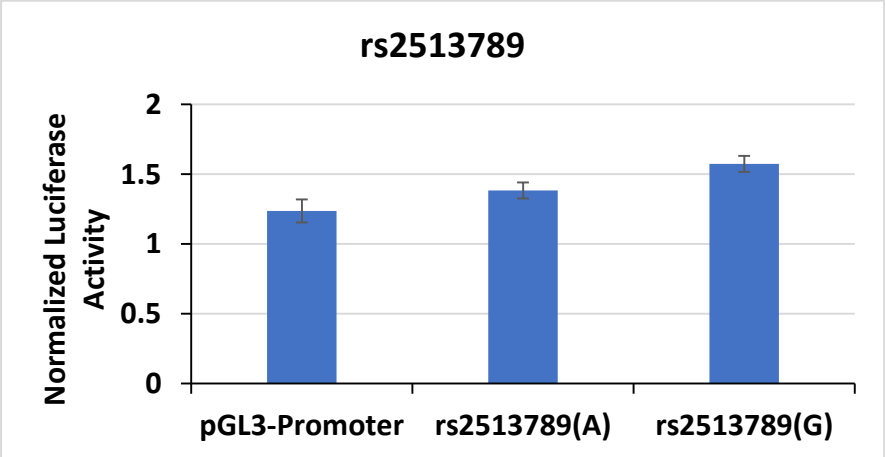

rs75465959
Normalized Luciferase Activity
2.4
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs75465959 (A)
rs75465959 (G)

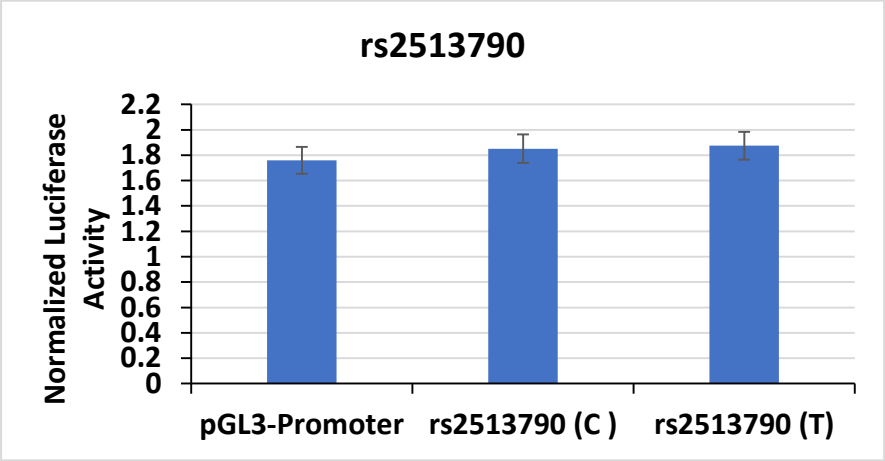

rs147978008
Normalized Luciferase Activity
2
1.5
1
0.5
0
pGL3-Promoter
rs147978008 (Del)
rs147978008 (Ins)

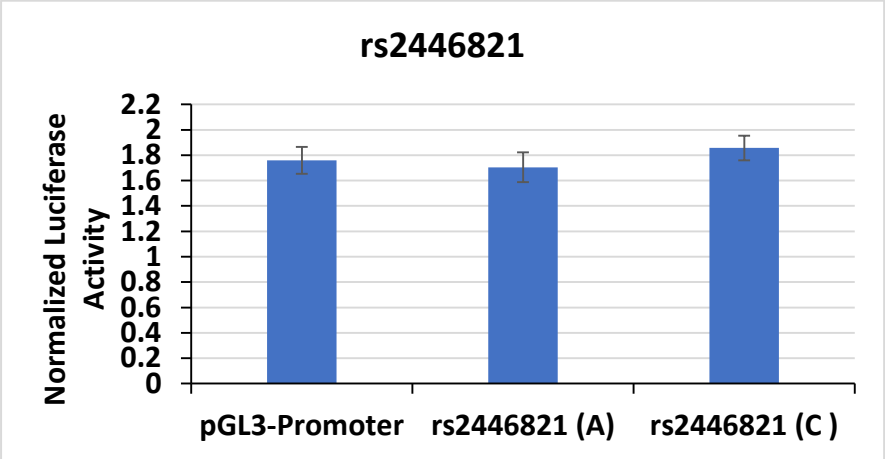

rs17019510
Normalized Luciferase Activity
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs17019510 (A)
rs17019510 (G)

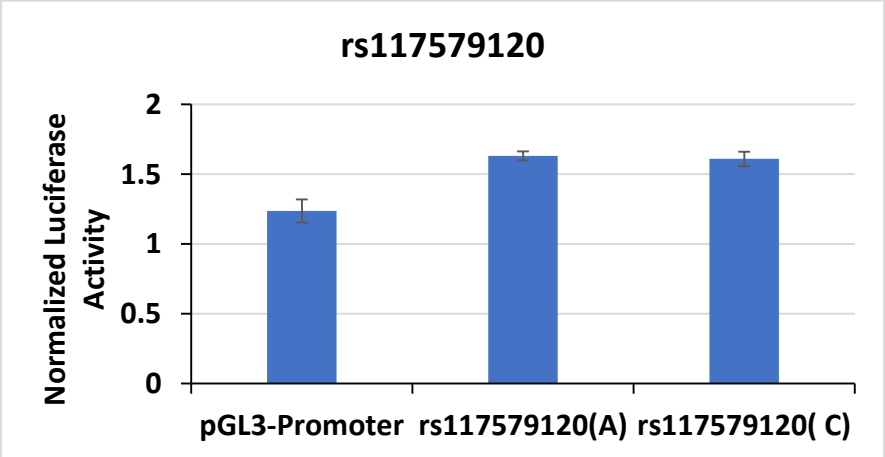

rs2446823
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs2446823 (A)
rs2446823 (G)

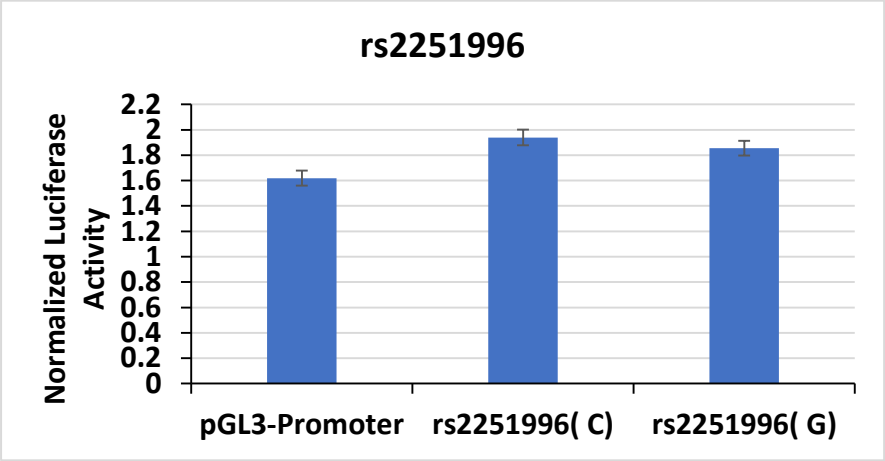

rs149630836
Normalized Luciferase Activity
pGL3-Promoter
rs149630836 (Del)
rs149630836 (Ins)

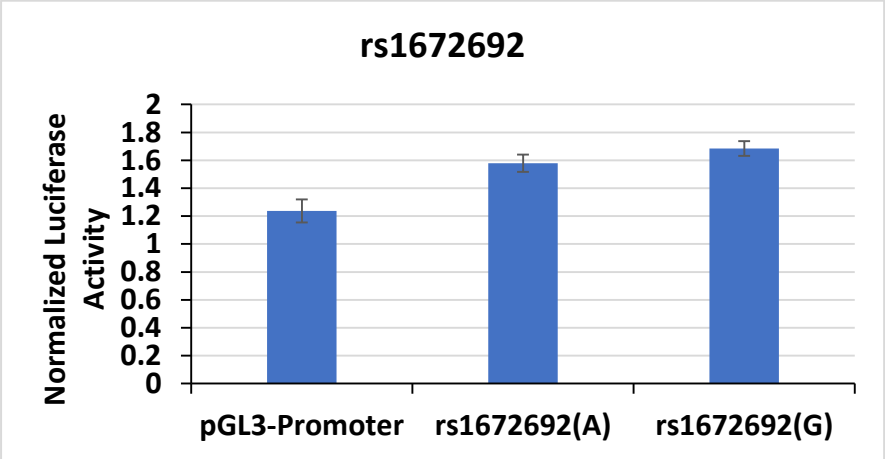

rs2513789
Normalized Luciferase Activity
2
1.5
1
0.5
0
pGL3-Promoter
rs2513789(A)
rs2513789(G)

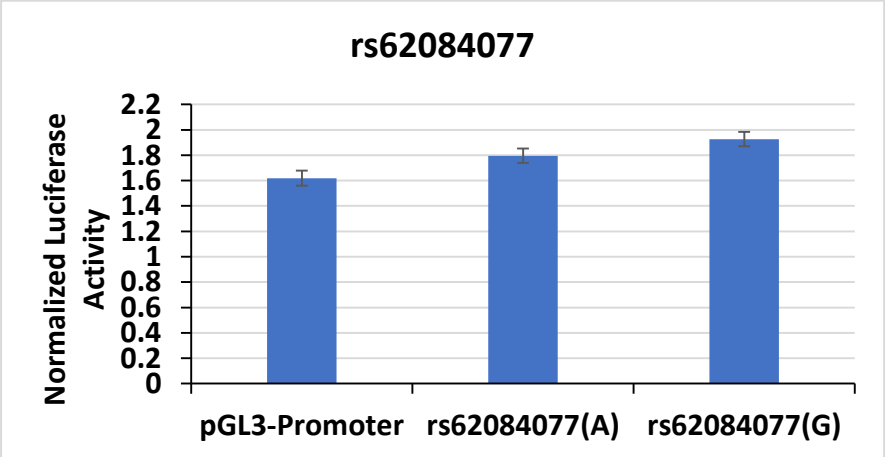

rs2513790
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs2513790 (C )
rs2513790 (T)

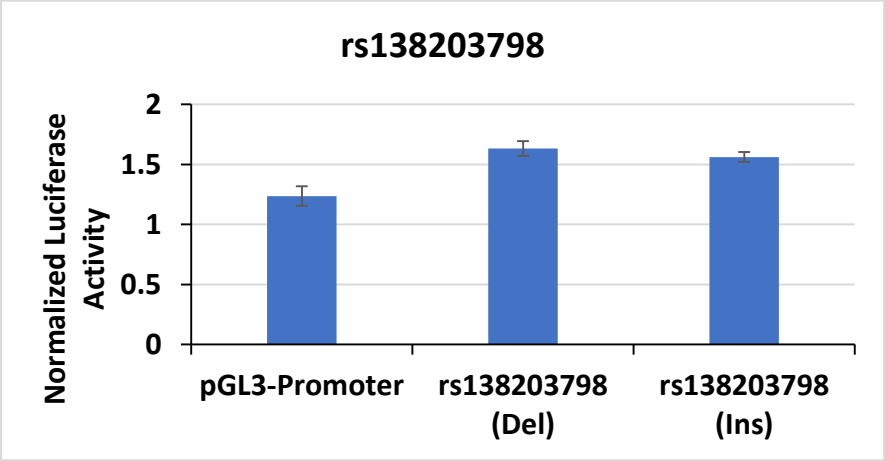

rs2446821
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs2446821 (A)
rs2446821 (C )


rs117579120
Normalized Luciferase Activity
2
1.5
1
0.5
0
pGL3-Promoter
rs117579120(A)
rs117579120( C)


rs2251996
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs2251996( C)
rs2251996( G)


rs1672692
Normalized Luciferase Activity
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs1672692(A)
rs1672692(G)


rs62084077
Normalized Luciferase Activity
2.2
2
1.8
1.6
1.4
1.2
1
0.8
0.6
0.4
0.2
0
pGL3-Promoter
rs62084077(A)
rs62084077(G)



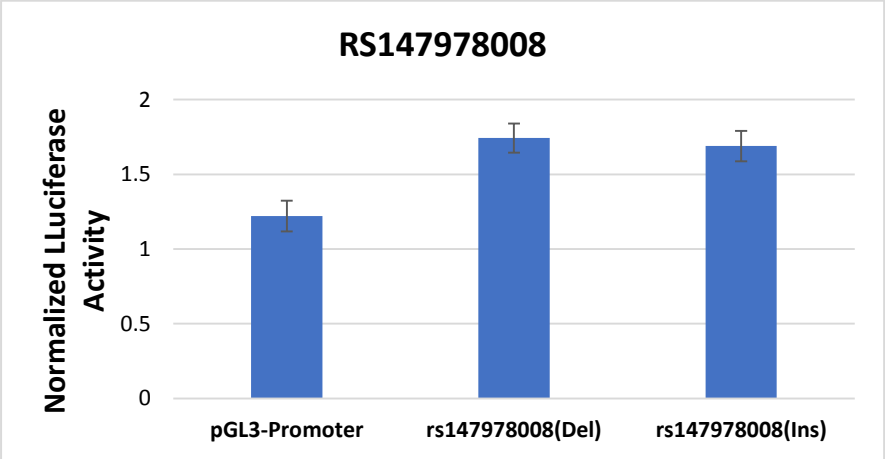




**Figure 7. Relative luciferase activities of SNPs that have no significant difference.** Relative luciferase activities of all SNPs above were analyzed by student t-test. No significant difference between each risk allele and non-risk allele was observed.

# References


Allen, G. (2002). Occupational Asthma in Adults in Six Canadian Counties. *AORN Journal*, 75(3), pp.654-657.

Anees, W. (2003). Use of pulmonary function tests in the diagnosis of occupational asthma. *Annals of Allergy, Asthma & Immunology*, 90(5), pp.47-51.

Altshuler, D., Gibbs, R., Peltonen, L., Altshuler, D. and Gibbs, R. (2010). Integrating common and rare genetic variation in diverse human populations. *Nature*, 10.1038/ nature09298(467), pp.52–58.

Bardana, E. (1999). Reactive airways dysfunction syndrome (RADS): guidelines for diagnosis and treatment and insight into likely prognosis. *Annals of Allergy, Asthma & Immunology*, 83(6), pp.583-586.

Brooks, S., Weiss, M. and Bernstein, I. (1985). Reactive Airways Dysfunction Syndrome (RADS). *Chest*, 88(3), pp.376-384.

Baur, X. and Czuppon, A. (1995). Diagnostic validation of specific IgE antibody concentrations, skin prick testing, and challenge tests in chemical workers with symptoms of sensitivity to different anhydrides. *Journal of Allergy and Clinical Immunology*, 96(4), pp.489-494.

Biagini, R., Bernstein, I., Gallagher, J., Moorman, W., Brook, S. and Gann, P. (1985). The diversity of reaginic immune responses to platinum and palladium metallic salts. *Journal of Allergy and Clinical Immunology*, 76(6), pp.794-802.

Blanc, P. (1987). Occupational Asthma in a National Disability Survey. *Chest*, 92(4), pp.613-617.

Bush, W. and Moore, J. (2012). Chapter 11: Genome-Wide Association Studies. *PLoS Computational Biology*, 8(12), p.e1002822.

Contreras, G., Rousseau, R. and Chan-Yeung, M. (1994). Occupational respiratory diseases in British Columbia, Canada in 1991. *Occupational and Environmental Medicine*, 51(10), pp.710-712.

Chan-Yeung, M. and Malo, J. (1994). Aetiological agents in occupational asthma. *European Respiratory Journal*, 7(2), pp.346-371.

Chan-Yeung, M. and Malo, J. (1995). Occupational Asthma. *New England Journal of*

*Medicine*, 333(2), pp.107-112.

Chan-Yeung, C. (1995). Assessment of Asthma in the Workplace. *Chest*, 108(4), pp.1084-1117.
Coelho, P. (2002). A novel mitochondrial protein, Tar1p, is encoded on the antisense strand of the nuclear 25S rDNA. *Genes & Development*, 16(21), pp.2755-2760.

de Wet, J., Wood, K., Helinski, D. and DeLuca, M. (1985). Cloning of firefly luciferase cDNA and the expression of active luciferase in Escherichia coli. *Proceedings of the National Academy of Sciences*, 82(23), pp.7870-7873.

Figueiredo, M. and Brownlee, G. (1995). cis-Acting Elements and Transcription Factors Involved in the Promoter Activity of the Human Factor VIII Gene. *Journal of Biological Chemistry*, 270(20), pp.11828-11838.

Groene, G., Pal, T., Beach, J., Tarlo, S., Spreeuwers, D., Frings-Dresen, M., Mattioli, S. and Verbeek, J. (2012). Workplace interventions for treatment of occupational asthma: a Cochrane systematic review: Figure 1. *Occupational and Environmental Medicine*, 69(5), pp.373-374.

Gilchrist, M., Henderson, W., Clark, A., Simmons, R., Ye, X., Smith, K. and Aderem, A. (2008). Activating transcription factor 3 is a negative regulator of allergic pulmonary inflammation. *The Journal of Experimental Medicine*, 205(10), pp.2349-2357.

Horne, C., Quintana, P., Keown, P., Dimich-Ward, H. and Chan-Yeung, M. (2000). Distribution of DRB1 and DQB1 HLA class II alleles in occupational asthma due to western red cedar. *European Respiratory Journal*, 15(5), pp.911-914.

Hai, Tsonwin & Wolfgang, C.D. & Marsee, D.K. & Allen, A.E. & Sivaprasad, Umasundari. (1999). ATF3 and stress responses. *Gene expression*. pp. 321-35.

Jeal, H., Draper, A., Jones, M., Harris, J., Welsh, K., Taylor, A. and Cullinan, P. (2003). HLA associations with occupational sensitization to rat lipocalin allergens: A model for other animal allergies?. *Journal of Allergy and Clinical Immunology*, 111(4), pp.795-799.

Katkoori, V. (2008). Clinical significance of a novel single nucleotide polymorphism in the 5' untranslated region of the Rabphillin-3A-Like gene in colorectal adenocarcinoma. *Frontiers in Bioscience*, 13(13), p.1050.

Kim, S., Cho, B., Park, C., Shin, E., Cho, E., Yang, E., Kim, C., Hong, C., Lee, J. and Park, H. (2009). Alpha-T-catenin (CTNNA3) gene was identified as a risk variant for toluene diisocyanate-induced asthma by genome-wide association analysis. *Clinical & Experimental Allergy*, 39(2), pp.203-212.

Krebs, J., Lewin, B., Kilpatrick, S. and Goldstein, E. (2014). *Lewin's genes XI*. Burlington, Mass.: Jones & Bartlett Learning.

Lummus, Z., Kaufman, K., Weirauch, M., Cartier, A., Cruz, M., Kesavalu, B., Lemiere, C., Muñoz, X., Perez-Camo, I., Quirce, S., Sastre, J., Yucesoy, B. and Bernstein, D. (2017). Next Generation DNA Sequencing of Novel Loci in Workers with Diisocyanate Asthma (DA). *Journal of Allergy and Clinical Immunology*, 139(2), p.AB187.

Leroyer, C., Perfetti, L., Cartier, A. and Malo, J. (1998). Can reactive airways dysfunction syndrome (RADS) transform into occupational asthma due to "sensitisation" to isocyanates. *Thorax*, 53(2), pp.152-153.

Meyer, J. (2001). SWORD '99: surveillance of work-related and occupational respiratory disease in the UK. *Occupational Medicine*, 51(3), pp.204-208.

McDonald, J. (2000). Reported incidence of occupational asthma in the United Kingdom, 1989-97.*Occupational and Environmental Medicine*, 57(12), pp.823-829.

Mapp, C. (2003). The role of genetic factors in occupational asthma. *European Respiratory Journal*, 22(1), pp.173-178.

Meredith, S. and Nordman, H. (1996). Occupational asthma: measures of frequency from four countries. *Thorax*, 51(4), pp.435-440.

Maestrelli, P. (2004). Natural History of Adult-Onset Asthma. *American Journal of Respiratory and Critical Care Medicine*, 169(3), pp.331-332.

Moscato, G., Malo, J. and Bernstein, D. (2003). Diagnosing occupational asthma: how, how much, how far?: Table 1. *European Respiratory Journal*, 21(5), pp.879-885.

Mapp, C., Boschetto, P., Maestrelli, P. and Fabbri, L. (2005). Occupational Asthma. *American Journal of Respiratory and Critical Care Medicine*, 172(3), pp.280-305.

Mapp, C., Saetta, M., Maestrelli, P., Ciaccia, A. and Fabbri, L. (1994). Low molecular weight pollutants and asthma: pathogenetic mechanisms and genetic factors. *European Respiratory Journal*, 7(9), pp.1559-1563.

Mapp, C., Fryer, A., Marzo, N., Pozzato, V., Padoan, M., Boschetto, P., Strange, R., Hemmingsen, A. and Spiteri, M. (2002). Glutathione S-transferase GSTP1 is a susceptibility gene for occupational asthma induced by isocyanates. *Journal of Allergy and Clinical Immunology*, 109(5), pp.867-872.

Ncbi.nlm.nih.gov. (2017). *CDH17 cadherin 17 [Homo sapiens (human)] - Gene - NCBI*. [online] Available at: https://www.ncbi.nlm.nih.gov/gene/1015 [Accessed 27 Nov. 2017].

Palmer, L. and Cookson, W. (2001). Using single nucleotide polymorphisms as a means to understanding the pathophysiology of asthma. *Respiratory Research*, 2(2), p.102.

Park, H. and Frew, A. (2002). Genetic markers for occupational asthma. *Journal of Allergy and Clinical Immunology*, 109(5), pp.774-776.

Reijula, K. and Patterson, R. (1994). Occupational Allergies in Finland in 1981–91. *Allergy and Asthma Proceedings*, 15(3), pp.163-168.

Rihs, H., Chen, Z., Ruëff, F., Cremer, R., Raulf-Heimsoth, M., Baur, X., Moneret-Vautrin, D. and Brüning, T. (2002). HLA-DQ8 and the HLA-DQ8-DR4 haplotype are positively associated with the hevein-specific IgE immune response in health care workers with latex allergy. *Journal of Allergy and Clinical Immunology*, 110(3), pp.507-514.

Sastre, J., Vandenplas, O. and Park, H. (2003). Pathogenesis of occupational asthma. *European Respiratory Journal*, 22(2), pp.364-373.

Tarlo, S. and Broder, I. (1989). Irritant-Induced Occupational Asthma. *Chest*, 96(2), pp.297-300.

Tokuhiro, S., Yamada, R., Chang, X., Suzuki, A., Kochi, Y., Sawada, T., Suzuki, M., Nagasaki, M., Ohtsuki, M., Ono, M., Furukawa, H., Nagashima, M., Yoshino, S., Mabuchi, A., Sekine, A., Saito, S., Takahashi, A., Tsunoda, T., Nakamura, Y. and Yamamoto, K. (2003). An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nature Genetics*, 35(4), pp.341-348.

Wikman, H., Piirilä, P., Rosenberg, C., Luukkonen, R., Kääriä, K., Nordman, H., Norppa, H., Vainio, H. and Hirvonen, A. (2002). N-Acetyltransferase genotypes as modifiers of diisocyanate exposure-associated asthma risk. *Pharmacogenetics*, 12(3), pp.227-233.

Wood, K., de Wet, J., Dewji, N. and DeLuca, M. (1984). Synthesis of active firefly luciferase by in vitro translation of RNA obtained from adult lanterns. *Biochemical and Biophysical Research Communications*, 124(2), pp.592-596.

Whitmore, M., Iparraguirre, A., Kubelka, L., Weninger, W., Hai, T. and Williams, B. (2007). Negative Regulation of TLR-Signaling Pathways by Activating Transcription Factor-3. *The Journal of Immunology*, 179(6), pp.3622-3630.

Yucesoy, B., Kaufman, K., Lummus, Z., Weirauch, M., Zhang, G., Cartier, A., Boulet, L., Sastre, J., Quirce, S., Tarlo, S., Cruz, M., Munoz, X., Harley, J. and Bernstein, D. (2015). Genome-Wide Association Study Identifies Novel Loci Associated With Diisocyanate-Induced Occupational Asthma. *Toxicological Sciences*, 146(1), pp.192-201.