Closeout Document 1
Final Progress Report-Renewal– 2013-2019

Title Page

PI: Dr. David Bernstein, Professor
University of Cincinnati
College of Medicine
Internal Medicine, Department of Allergy and Immunology
3255 Eden Ave
Cincinnati OH 45267-0563
(513) 558-4701
David.bernstein@uc.edu

Institution: University of Cincinnati
Sponsored Research Services
University Hall, Suite 530
51 Goodman Drive
P.O. Box 210222
Cincinnati OH 45221-0222
(513) 556-4358

Project Title: Genetic Susceptibility for Occupational Asthma

Co-investigators:
Lummus, Zana L., University of Cincinnati (2006-2018)
Yucesoy, Berran, University of Cincinnati (2013-2015)
Project Director and Sponsors : NIOSH, CDC

Project Dates: 09/01/2013 – 08/31/2019

Grant Number: R01 OH008795-08 Revised

Date of Report: January 16, 2020

Table of Contents

List of Terms and Abbreviations

A549, A549 human lung carcinoma epithelial cell line
BEAS 2B, BEAS 2B epithelial virus transformed normal human lung/bronchus cell line
BWA, Burrows-Wheeler transform
CRISPR, clustered regularly interspaced short palindromic repeats
CRISPRi, CRISPR interference
DA, diisocyanate-induced asthma
DAPA, DNA affinity purification assay
DI, diisocyanate chemical
EMSA, electrophoretic mobility shift assay
HDI, hexamethylene diisocyanate
1KG, 1000 genomes project
IMR-90, IMR-90 normal human fetal lung fibroblast cell line
GATK, Genome Analysis Toolkit (tools for NGS using MapReduce programming)
GWAS, genome-wide association study
MAF, minor allele frequency
MDI, methylene diphenyl diisocyanate
NanoLC-MS/MS, Nano liquid chromatography, followed by tandem mass spectrometry analysis
NGS, next generation sequencing
NR, non-risk allelic oligonucleotide probe
OA, occupational asthma
R, risk allele oligonucleotide probe
rSNPs, non-coding regulatory SNPs
SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SIC, specific inhalation challenge
SNP, single nucleotide polymorphism
SNV, single nucleotide variant
TDI, toluene diisocyanate
TF, Transcription Factor
tSNP, tagging SNP

**Title:** Genetic Susceptibility for Occupational Asthma
**Investigator:** David I. Bernstein, University of Cincinnati, College of Medicine, Department of Internal Medicine, Division of Immunology and Allergy, Cincinnati, OH 45267-0563 (Email: Bernstdd@ucmail.uc.edu).
**Affiliation:** University of Cincinnati College of Medicine
**State:** OH
**Telephone:** (513) 558-3612
**Award Number:** R01 OH008795-08 Revised
**Start and End Dates:** 09/01/2013 – 08/31/2019
**Program Area:** Occupational Respiratory Disease

**Final Report Abstract**

Diisocyanates cause occupational asthma (DA) in 5-15% of exposed workers. Specific bronchial sensitization to diisocyanates are associated with airway inflammation and remodeling. Unlike IgE dependent asthma, biomarkers of IgE dependent sensitization have not been consistently demonstrated. Due to inherent toxicity, the DA phenotype is likely attributable to innate immune responses, oxidative stress, and epithelial cell barrier injury.

We have investigated DA associated genotypes, using candidate gene studies, GWAS and Next Generation Sequencing (NGS) to identify gene variants associated with DA. DNA was collected from consented workers with specific inhalation challenge (SIC) confirmed DA recruited in Canada and Spain. The initial NIOSH R01 grant award (2006-2012) identified candidate variants (snps) linked to Th2 cytokines, antioxidant enzymes, and HLA class I and II antigens. GWAS was performed revealing novel DA associated SNPs in 74 DA subjects compared with 824 healthy controls using Omni-2.5 and Omni-5 SNP microarrays; 11 SNPs exceeded genome-wide significance with the strongest association with rs12913832 SNP on chromosome 15, mapping to HERC2 gene ($p=6.94\times10^{-14}$). Disease associated SNPs were detected near ODZ3 and CDH17 genes (rs908084, $p=8.59\times10^{-9}$ and rs2514805, $p=1.22\times10^{-8}$, respectively). We prioritized 38 SNPs with suggestive genome-wide significance ($p <1\times10^{-6}$) with 17 SNPs mapping to PITPNC1, ACMSD, ZBTB16, ODZ3, and CDH17 gene loci. We replicated a Korean GWAS study showing two DA associated SNP variants (rs10762058, rs7088181) on the CTNNA3 gene. CTNNA codes for α-T-catenin, a protein that complexes with E-cadherin to form the epithelial junctional complex (EJC).

In this renewal (2013-2019), high-throughput methods were used to identify DA associated tagging SNPs (tSNP) within E-cadherin, α, β and γ catenin genes in DA workers (n=150) and 150 asymptomatic workers (AWs) exposed to HDI or MDI . We focused on SNPs within or near TACR1/LOC105374811 (chromosome 2) and CTNNA3/DNAJC12 (chromosome 10).

Based on a previous GWAS study, 14 DA associated loci and flanking regions were selected for NGS (6,996,180 base pairs). NGS was performed in 91 workers with DA (cases) and in 293 subjects from the 1,000 genomic (1KG) control data set. NGS yielded 143 SNP associations with DA ($\chi^2$; $p \leq 10^{-3}$) (range $10^{-3.006}$- $10^{-5.504}$, nearly all located in intronic regions. Analysis of transcription factor (TF) and ChIP- seq datasets identified overlap of 70 SNPs with TF binding regions of relevant lung cell lines (e.g., A549 and IMR90). The top 22 DA associated SNPs were contained in 7 loci: CDH17 (10), ATF3 (6), FAM71A (2), PITPNC1, TACR1, ZBTB16, LOC101929565. 22 top-ranked SNPs associated with DA ($X^2$ $p< 1 \times 10^{-3}$). Electrophoretic mobility shift assays (EMSA) identified oligonucleotide-protein binding for risk and non-risk SNPs to nuclear extracts of A549, BEAS 2B, and IMR-90 lung cell lines. EMSA detected A549 nuclear extract binding to 10 variants, with 8 displaying preferential binding to non-risk alleles (rs1001304, rs2287231,rs2513789, rs2513791, rs11571537, rs14798008) or risk alleles (rs2513788, rs2513790). Bound proteins were purified by DNA affinity precipitation assays (DAPA) and eluted proteins analyzed by mass spectrometry (MS), demonstrating that rs14798008 bound H1 histones. A luciferase reporter assay, in A549 cells, identified allele-dependent mRNA transcription. Four SNPs exhibited allele-dependent increases in gene expression (rs2287231, rs2513789, rs11571537 and rs2446824).

**Section 1**
**Significant (Key) Findings**

1. This is the first GWAS in a predominately Caucasian population exposed to diisocyanates and the first study to employ NGS in the investigation of OA and DA. . NGS yielded 143 significant SNP associations with DA ($\chi^2$; $p \leq 10^{-3}$) (range $10^{-3.006}$- $10^{-5.504}$), of which all but one SNP were located in intronic regions.

2. This study advances knowledge of genetic mechanisms of occupational asthma by identification of a role for non-coding genetic variants associated with diisocyanate-induced asthma. Results from recent studies suggest that as many as 93% of disease- and trait-associated SNPs are located in regulatory regions (1, 2), indicating that many SNPs might act by affecting the binding of transcription factors (TFs). We identified 5 potential regulatory allele-dependent SNPs for DA. Four variants exhibited functional activity for dysregulation of gene transcription. . Our results demonstrate that many DA-associated genetic variants likely act by modulating gene transcription

3, The fifth variant, in the promoter region of FAM71A (rs14798008), showed non-risk allele preferential binding to H1 histones, that modulate chromosomal accessibility of regulatory proteins to DNA. *FAM71A is* the coding gene for a Golgi-resident Rab2B specific binding protein GARI-L4 (also called the FAM71A protein), with an important role in the maintenance of the compacted Golgi morphology of mammalian cells (3). No other function has been reported for this protein, and this is the first report of allele-dependent binding of H1 histones to the *FAM71A* promoter.

**Translations of Findings**
Currently the treatment for subjects diagnosed with DA is supportive asthma therapies and cessation of exposure. It appears likely that a reduction in the incidence and severity of DA could be achieved by genetic screening at some time in the future. Workers with validated high-risk genotypes and work-related asthma could be targeted for more intensive exposure control measures in the workplace and enhanced medical surveillance measures. This approach could reduce or prevent development of severe asthma. Further research is needed to determine a functional role for the above-mentioned gene polymorphisms in the pathogenesis of DA. In particular, screening methods are needed that can distinguish true causal disease variants from other variants showing an association with DA by way of linkage to the causal disease variants. It may also be possible to correlate the genetic risk factors identified with specific job descriptions and disease phenotypes in DA.

**Outcomes/Impact**

1) **Potential outcomes** (findings that could impact workplace risk if used)

These studies were directed at meeting the goals and intent of the NIOSH programmatic Research to Practice (r2p) initiative which calls for extramural NIOSH-funded proposals that transfer or translate innovative research findings to effective prevention practices which can be adapted to workplace practices that will reduce illness and injury. Consistent with the r2p initiative, our results have identified & validated single or multiple SNPs that in the future can be utilized as susceptibility markers for OA caused by reactive chemical agents and which may identify asymptomatic workers at greatest risk for development of airways inflammatory responses and, ultimately, occupational asthma. Further, our studies are likely to have a major impact on human health in that, for the first time, a large DNA databank of extremely well-phenotyped DA cases and control subjects exposed to a variety of low molecular weight chemicals from both Canadian and Spanish background populations has been established. The DNA bank will continue to grow and is very likely to serve as a valuable resource for NIOSH and other investigators pursuing investigations of susceptibility genes for OA.

2) **Intermediate outcomes** (how findings,results or recommendations have been used by others to influence practices, legislation, safety management program, training, etc.)
**UNKNOWN**

3) **End outcomes** (how findings have contributed to documented reductions in work-related morbidity, mortality, and/or exposure)
**UNKNOWN**

**Section 2: Scientific Report**

**Background**

Occupational Asthma (OA) is the most common occupational lung disorder (4). Diisocyanate (DI) chemicals, are used in urethane products, as spray paint hardening agents and adhesives, and are leading causes of OA (5). Despite years of research, there are no reliable markers to predict risk or susceptibility for diisocyanate asthma (DA). In our multicenter candidate gene association study funded by this NIOSH RO1, we replicated findings from a GWAS study of Korean workers, demonstrating that two SNP variants (rs10762058, rs7088181) of the CTNNA3 gene (coding for α-catenin) were associated with DA in a Caucasian worker population (6). Alpha-T-catenin, β- and γ-catenin are cytoplasmic anchorage proteins that complex with E cadherin to form the epithelial junctional complex, critical in maintaining epithelial integrity . These findings led us to postulate that genetic variants may alter expression of junctional proteins, resulting in a dysfunctional epithelial barrier, rendering airways of exposed workers susceptible to pro-inflammatory and oxidative effects of DIs. Thus, dysfunctional epithelial proteins could enhance susceptibility to DA among workers with respiratory exposure to DIs. We proposed the following specific aims to test the hypothesis that fine mapping of CTNNA3 loci and genotyping additional proteins of the epithelial junctional complex will reveal functional genetic variants that define susceptibility for DA among DI exposed workers.

*Specific Aim 1***:** Sequence informative regions of CTTNA3 (and other DA-associated genes identified by GWAS or Specific Aim 2) to identify functional genotype associations with DA.

*Specific Aim 2:* Identify E cadherin and catenin gene polymorphisms associated with DA

*Specific Aim 3:* Define the functional relevance of replicable candidate SNPs associated with DA (and initiate studies, using CRISPR interference techniques, to distinguish between the true causal disease risk loci, & other loci in linkage equilibrium with the disease causal loci).

**Studies and Results**

***Specific Aim 2 (performed in grant year 5 prior to Specific Aim 1)***

These studies have been performed , using microarray technology (Illumina Goldengate assay) for E-cadherin/catenin complex genes, with DNA samples obtained in the following diisocyanate-exposed worker groups: 1) 136 workers with DA confirmed by a specific inhalation challenge test; 2) 144 asymptomatic HDI-exposed workers; and 3) 108 asymptomatic MDI-exposed workers. 384 SNPs were identified within E-cadherin (CDH1), α-T catenin (CTNNA3), alpha E-Catenin (CTNNA1), beta-catenin (CTNNB1), catenin Gamma (JUP), Catenin Delta-1 (CTNND1) genes. Tag SNPs were selected with r2≥0.80, minor allele frequency ≥ 0.1 to tag the region of interest and included a 25kb flanking region on either side of each gene. In addition, we used FESD V2 and Regulome to ensure that regulatory regions (e.g., promoter regions, coding exons, 5'UTRs and 3'UTRs) are represented. There were 349 unique samples that were genotyped after 39 samples were excluded due to high missing call rates, heterogeneity or duplications. After exclusion of subjects, gene association analyses, 119 workers with confirmed diisocyanate asthma (DA) were compared with 230 isocyanate-exposed asymptomatic control subjects. Key variables included specific isocyanate exposure (HDI, MDI, TDI), smoking status (current vs. ex-smoker vs. no smoking history), and gender. Table 1 lists significant associations with DA after adjustment for relevant variables (i.e., smoking status, gender and specific isocyanate exposure) in a logistic regression model. The results suggested that genetic variants within or near TACR1/LOC105374811 (chromosome 2) and CTNNA3/DNAJC12 (chromosome 10) might be associated with OA. These regions were included in the sequence study for completion of Specific Aim 1.

**Table 1. Analysis of Diisocyanate asthma cases vs. controls after adjusting for gender, smoking and specific isocyanate exposure (TDI,MDI,HDI)**

| chr | region | genes | SNPs | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | rs | pos | A/B | Q.2 | Pexact | P1df | P2df | OR |
| 10 | 69342534 to 69713031 | SIRT1 CTNNA3 HERC4 DNAJC12 TRS-TGA1-1 | rs7081671 | 69442534 | C/A | 0.139 | 0.637 | 0.00736 | 0.0259 | 1.14 (1.04,1.25) |
| | | | rs3125309 | 69482437 | G/A | 0.353 | 0.62 | 0.00519 | 0.0065 | 1.10 (1.03,1.18) |
| | | | ***rs1113976*** | ***69517922*** | ***A/G*** | ***0.274*** | ***0.887*** | ***0.000123*** | ***9.26e-05*** | ***1.16 (1.07,1.25)*** |
| | | | rs717976 | 69551141 | G/C | 0.378 | 0.905 | 0.000172 | 0.000329 | 1.14 (1.06,1.22) |
| | | | rs7081610 | 69613031 | A/C | 0.316 | 0.0486 | 0.0108 | 0.0278 | 1.09 (1.02,1.17) |
| 2 | 75326968 to 75587574 | TACR1 LOC105374812 LOC105374811 LOC105374810 | rs2193405 | 75426968 | T/A | 0.276 | 0.0334 | 0.0392 | 0.000741 | 1.09 (1.00,1.18) |
| | | | ***rs11684394*** | ***75444073*** | ***C/T*** | ***0.354*** | ***0.0847*** | ***0.003*** | ***0.00076*** | ***1.12 (1.04,1.20)*** |
| | | | rs6546958 | 75452466 | C/T | 0.434 | 0.166 | 0.00492 | 0.000456 | 0.90 (0.97,0.84) |
| | | | rs4852364 | 75457783 | C/T | 0.431 | 0.419 | 0.024 | 0.0216 | 0.92 (0.99,0.86) |
| | | | rs1158816 | 75458678 | G/T | 0.396 | 0.193 | 0.00305 | 0.00141 | 1.11 (1.04,1.20) |
| | | | rs7585802 | 75474706 | T/G | 0.429 | 0.908 | 0.00693 | 0.0198 | 1.10 (1.03,1.17) |
| | | | rs2216103 | 75487574 | A/G | 0.354 | 0.265 | 0.0253 | 0.0741 | 1.09 (1.01,1.17) |
| 10 | 68807082 to 69079853 | CTNNA3 LRRTM3 | rs12411525 | 68907082 | C/A | 0.381 | 0.281 | 0.0126 | 0.0417 | 0.92 (0.98,0.85) |
| | | | ***rs9651325*** | ***68979853*** | ***G/A*** | ***0.0946*** | ***0.746*** | ***0.00573*** | ***0.0178*** | ***1.17 (1.05,1.31)*** |
| 1 | 212638909 to 212838909 | ATF3 FAM71A LOC105372909 LOC105372908 LOC105372907 LOC102723761 LOC101929565 | ***rs1195472*** | ***212738909*** | ***T/C*** | ***0.461*** | ***0.82*** | ***0.00651*** | ***0.0106*** | ***1.10 (1.03,1.18)*** |
| 8 | 95058146 to 95258146 | CDH17 LOC105375648 LOC105375647 | ***rs13263145*** | ***95158146*** | ***G/C*** | ***0.327*** | ***0.797*** | ***0.0172*** | ***0.04*** | ***0.91 (0.98,0.85)*** |
| 10 | 68493916 to 68700060 | CTNNA3 LRRTM3 LOC101928961 | ***rs10733830*** | ***68593916*** | ***C/T*** | ***0.301*** | ***0.346*** | ***0.0251*** | ***0.0804*** | ***0.92 (0.99,0.85)*** |
| | | | rs5019668 | 68600060 | C/T | 0.115 | 0.581 | 0.0318 | 0.0662 | 0.89 (0.99,0.80) |
| 5 | 137842510 to 138042510 | HSPA9 ETF1 SNORD63 LOC105379193 | ***rs11747974*** | ***137942510*** | ***G/A*** | ***0.179*** | ***0.335*** | ***0.0405*** | ***0.0979*** | ***0.91 (1.00,0.83)*** |
| 10 | 68627662 to 68827662 | CTNNA3 LRRTM3 LOC101928961 | ***rs11591465*** | ***68727662*** | ***T/C*** | ***0.189*** | ***0.853*** | ***0.0497*** | ***0.132*** | ***1.09 (1.00,1.19)*** |

***Specific Aim 1***

1. In 2012, a GWAS study was performed using intramural research funding received from the University of Cincinnati for an interdisciplinary study between the PI and Co-investigators at the Cincinnati Children's Hospital Medical Center (Drs. Kenneth Kaufman and John Harley), which led to the discovery of new DA associated genetic variants (7). Genetic loci containing these SNPs were included in our sequencing study (see item 2 below). Genome-wide SNP genotyping was performed on genomic DNA from 88 cases using Omni-2.5 SNP microarray (Illumina, San Diego, CA). The chip contains ~2.5M SNP markers with an average call frequency of > 99% and is unbiased with respect to coding and noncoding regions of the genome. 832 self-reported Caucasian healthy individual controls recruited from the Cincinnati Genomic Control Cohort

(8) were run on the Omni-5 microarray which contains ~4.3 million SNPs. A total of ~2.4 million markers were common to both arrays and used for analysis.

We identified 10 SNPs that exceeded genome-wide significance, with the strongest association being for the rs12913832 SNP located on chromosome 15, which has been mapped to the HERC2 gene ($p=6.94x10^{-14}$). Strong evidence was also found for SNPs near the ODZ3 and CDH17 genes on chromosomes 4 and 8 (rs908084, $p=8.59x10^{-9}$ and rs2514805, $p=1.22x10^{-8}$, respectively). We also prioritized 38 suggestive genome-wide significant SNPs ($p <1\times10^{-6}$). Among them, 17 SNPs map to the PITPNC1, ACMSD, ZBTB16, ODZ3, and CDH17 gene loci. Functional genomics data indicate that two of the suggestive SNPs (rs2446823 and rs2446824) are located within putative binding sites for the CEBPA/B and HNF4A transcription factors (TF), respectively.

**Top Ranked Genes**

| Entrez Gene ID | Gene Symbol | Gene Name | Number of SNPs | P-Values |
|---|---|---|---|---|
| 8924 | HERC2 | HECT and RLD domain containing E3 ubiquitin protein ligase 2 | 1 | $6.94 x10^{-14}$ |
| 55714 | ODZ3 | teneurin transmembrane protein 3 | 3 | $8.59x 10^{-9}$ - $5.64x10^{-7}$ |
| 1015 | CDH17 | cadherin 17, liver-intestine cadherin | 4 | $1.22x10^{-8}$ - $4.74x10^{-7}$ |
| 130013 | ACMSD | aminocarboxymuconate semialdehyde decarboxylase | 1 | $6.35x10^{-7}$ |
| 26207 | PITPNC1 | phosphatidylinositol transfer protein, cytoplasmic 1 | 6 | $6.33x10^{-7}$ – $7.82x10^{-7}$ |
| 7704 | ZBTB16 | zinc finger and BTB domain containing 16 | 5 | $1.68x10^{-7}$ – $7.03x10^{-7}$ |
| **Suggestive genes** | | | | |
| 64067 | NPAS3 | neuronal PAS domain protein 3 | 7 | $2.42x10^{-6}$- 0.012711 |
| 5578 | PRKCA | protein kinase C, alpha | 4 | $2.55x10^{-5}$- 0.000185 |
| 6539 | SLC6A12 | solute carrier family 6 (neurotransmitter transporter), member 12 | 3 | $1.47x10^{-6}$- $1.67x10^{-5}$ |
| 6869 | TACR1 | tachykinin receptor 1 | 20 | $4.08x10^{-9}$- 0.006198 |

2. Next generation sequencing (9) was performed at the Cincinnati Children's Hospital Medical Center DNA Core. Fourteen loci were selected for sequencing categorized based on relative level of significance: 1) loci containing disease associated SNPs of Genome-wide significance ($p <10^{-7}$) - HERC2, ODZ3, CDH17, PITPNC1 and ZBTB16; 2) loci containing disease associated SNPs of moderate statistical significance ($p < 10^{-6}$) - NPAS3, PRKCA, SLC6A12, TACR1; 3) loci of lower significance ($p > 10^{-6}$) - CTNNA3, GALNT14, HDAC9, KCND3, and ATF3. Data were obtained for 91 exposed subjects with confirmed diisocyanate asthma (cases) and comparator data from the publicly available 1,000 genomic (1KG) control data set. Sequence data analysis yielded 142 statistically significant SNP associations with DA, with 1 of the 142 significant SNPs mapping to a coding region, and that coding SNP was in an uncharacterized region of DNA. This result strongly suggests that the significant SNPs located in non-coding regions, if functionally relevant, are likely associated with regulatory transcription binding sites.

3. Results from recent studies suggest that as many as 93% of disease- and trait-associated SNPs are located in regulatory regions (1, 2), indicating that many SNPs might act by affecting the binding of transcription factors (TFs). We therefore used functional genomics data to identify SNPs located in likely regulatory regions in cell types relevant to DA. Transcriptomic analysis was performed on the 70-top disease associated SNPs in collaboration with Dr. Matthew Weirauch, a bioinformatic consultant at Cincinnati Children's Hospital. Using an algorithm, transcription factor datasets were evaluated for overlap with gene variants and ranked based upon likelihood of having regulatory properties in relevant lung cell types. Analysis

focused on available datasets from lung cell lines such as A549 and IMR90, and lung-derived tissues. Data sets were used with ChIP-seq datasets for TFs, regulatory histone marks, and DNase-seq (open chromatin). SNPs were ranked based on number of overlaps demonstrated by ChIP-seq peaks for different datasets of varying lung cell types and biologic relevance of associated genes. Twenty-one top ranked SNPs were identified based on the algorithm for further investigation.

**Genetic variants considered most likely to function by altering gene regulatory mechanisms in disease-relevant cells**

| Type of Marker* | Variant | Chr: position | Gene | genomic region | Flanking Genes (Left: :Right) | Risk Allele | Non-Risk Allele | $\chi^2$P ** | Number of Data-sets*** |
|---|---|---|---|---|---|---|---|---|---|
| SNV | rs1001304 | 1:212732686 | none | intergenic | LOC10192956: :ATF3 | C | T | $3.15x10^{-4}$ | 50 |
| SNV | rs72756369 | 1: 212770271 | ATF3 | intronic | | A | T | $3.42x10^{-4}$ | 27 |
| SNV | rs11571537 | 1:2127870251 | ATF3 | intronic | | C | T | $4.78x10^{-4}$ | 12 |
| SNV | rs11571559 | 1:212794550 | ATF3 | promoter | | T | C | $8.77x10^{-5}$ | 4 |
| SNV | rs11571563 | 1:2127948731 | none | intergenic | ATF3: :FAM71A | G | T | $2.80x10^{-4}$ | 5 |
| SNV | rs74138575 | 1:2127949971 | none | intergenic | ATF3: :FAM71A | A | G | $8.77x10^{-5}$ | 6 |
| SNV | rs75465959 | 1:2127951961 | none | intergenic | ATF3: :FAM71A | A | G | $8.77x10^{-5}$ | 6 |
| Deletion | rs147978008 | 1:212796008 | FAM71A | promoter | | - | ACA | $8.77x10^{-5}$ | 3 |
| SNV | rs17019510 | 1:212816729 | none | intergenic | FAM71A: :Loc100129948 | G | A | $4.6x10^{-6}$ | 7 |
| SNV | rs2287231 | 2:75449129 | none | intergenic | TACR1: :GAPDHP59 | A | G | $4.30x10^{-5}$ | 17 |
| SNV | rs2446824 | 8:95127574 | none | intergenic | RPL34P18: CDH17 | T | C | $3.00x10^{-5}$ | 32 |
| SNV | rs2446823 | 8:95127612 | none | intergenic | RPL34P18: CDH17 | G | T | $3.00x10^{-5}$ | 32 |
| Deletion | rs149630836 | 8:95127836 | none | intergenic | RPL34P18: CDH17 | - | TCAGTAG | $4.82x10^{-5}$ | 35 |
| SNV | rs2513788 | 8:95127896 | none | intergenic | RPL34P18: CDH17 | G | T | $3.00x10^{-5}$ | 32 |
| SNV | rs2513789 | 8:95128113 | none | intergenic | RPL34P18: CDH17 | A | G | $3.00x10^{-5}$ | 16 |
| SNV | rs2513790 | 8:95128147 | none | intergenic | RPL34P18: CDH17 | C | T | $3.00x10^{-5}$ | 23 |
| SNV | rs2446821 | 8:95128479 | none | intergenic | RPL34P18: CDH17 | T | G | $3.00x10^{-5}$ | 27 |
| SNV | rs2513791 | 8:95128529 | none | intergenic | RPL34P18: CDH17 | A | C | $3.00x10^{-5}$ | 22 |
| SNV | rs117579120 | 8:95159359 | CDH17 | intronic | | A | C | $3.78x10^{-5}$ | 21 |
| SNV | rs2251996 | 8:95160288 | CDH17 | intronic | | C | G | $1.61x10^{-5}$ | 22 |
| SNV | rs1672692 | 11:113945609 | ZBTB16 | intronic | | A | G | $4.26x10^{-5}$ | 8 |
| SNV | rs62084077 | 17:65386837 | PITPNC1 | intronic | | A | G | $9.41x10^{-4}$ | 34 |

*SNV = Single Nucleotide Variant
** $X^2$ P values for the DA association results for the DA subjects compared to the 1000 genomes database controls
*** The total number of datasets each SNP overlaps

***Specific Aim 3***

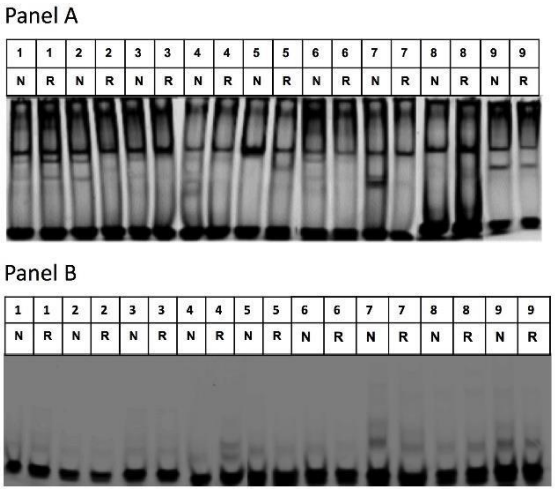
1. Two methodologies were used to test the functional relevance of the 21 top-ranked candidate SNPs associated with DA, the electrophoretic mobility shift assay (EMSA) and the luciferase reporter assay.

2. EMSA was performed in Dr. Bernstein's laboratory in the Department of Medicine. To prepare nuclear extracts for EMSA studies, three lung-derived cell lines were cultured: A549 epithelial cells, IMR90 fibroblasts

and BEAS 2B epithelial cells. SNP oligonucleotide sequences were obtained from the NCBI dbSNP database (https://www.ncbi.nlm.nih.gov/snp) and used to prepare 5'IMR700dye-labeled 35 bp duplexed oligonucleotide probes to test for nuclear protein binding. Both risk and non-risk alleles (n=42) were tested (probe sequences shown in Table 3. Binding reactions were carried out and the products were electrophoresed on native 5% TBE (Tris-Borate-EDTA) polyacrylamide gels, and then imaged using an Odyssey® CLx imager. These experiments revealed that 9 of the 21 tested SNPs bound to nuclear proteins, with 7 displaying genotype-dependent binding (Fig 1)

**Fig 1. Allele-dependent protein-DNA interactions in A549 cells.**

**A.** EMSA binding reactions of A549 nuclear extract proteins by oligonucleotide probes containing non-risk (N) or risk (R) alleles of genetic variants: (1) rs75465959, (2) rs1001304, (3) rs2446824, (4) rs11571537, (5) rs2513790, (6) rs2287231, (7) rs14978008, (8) rs2513788, (9) rs2513789. Variants 5 and 8 show stronger binding to the risk allele, variants 2, 4, 6, 7, and 9 show stronger binding to the non-risk allele, and genetic variants 1 and 3 show comparable risk and non-risk binding. **B.** Oligonucleotide controls for gels shown in Panel A.




3. Dual Luciferase Reporter Assays

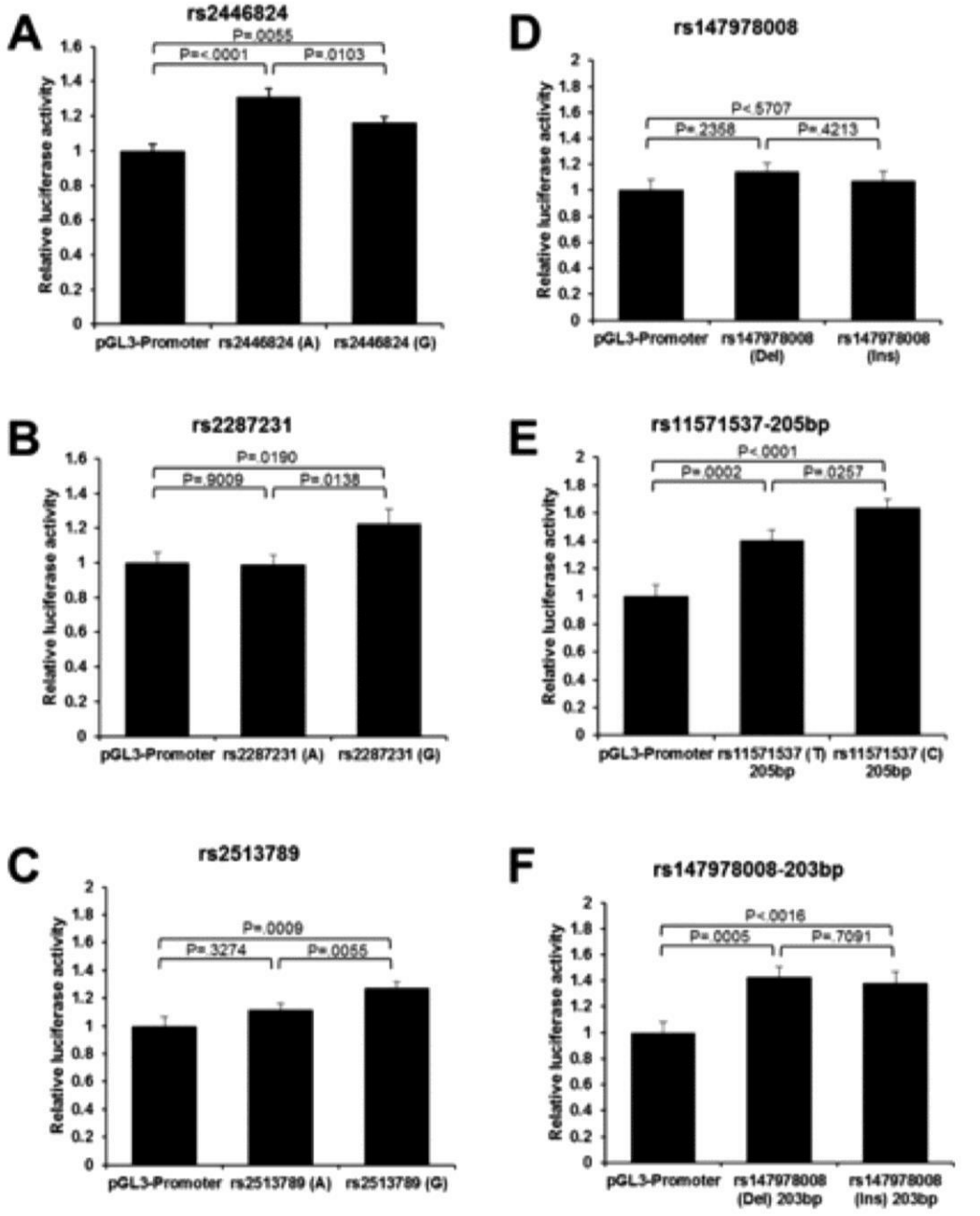
These studies were performed to identify allele dependent effects of prioritized SNPs on gene transcription, in the laboratory of Dr. Yao, West Virginia University. Luciferase reporter constructs were cloned using the pGL3-promoter vector (Promega, Madison, WI, USA) as a backbone. To create fragments for insertion into this backbone, two complementary oligonucleotides were designed so that annealing of a sense and antisense oligonucleotide formed a 31bp DNA fragment with a KpnI site at the 5'end and a XhoI site at the 3'end. DNA fragments were cloned into the vector upstream of the SV40 promoter. Both risk and non-risk allele constructs were produced for each SNP. Larger luciferase reporter constructs for SNP rs147978008 and SNP rs11571537 were made by PCR amplification of DNA fragments containing the SNPs (203 bp for rs147978008 and 205 bp for rs11571537) using a genomic DNA sample heterozygous for both SNPs and primers containing KpnI and XhoI sites. All constructs were verified by DNA sequencing.

Luciferase reporter assays were performed in A549 cells transfected with a pGL3 reporter construct and the pRL-CMV vector (The pRL Vectors are wild type Renilla luciferase control reporter vectors, which provide constitutive expression of Renilla luciferase and can be used in combination with a firefly luciferase vector to cotransfect mammalian cells). Three independent experiments were performed and eight replicate samples were assayed in each experiment. Relative luciferase activity was calculated by normalizing the luciferase activity of the empty pGL3-Promoter vector to1. Student's t-test was performed to determine differences between the risk and non-risk alleles of each SNP.

Three genetic variants showed significant genotype-dependent differences in luciferase reporter activity ($P < .05$), with the rs2446824 risk allele and the rs2513789 and rs2287231 non-risk alleles driving enhanced expression (Fig 2. A, B, C). Since rs11571537 and rs147978008 both demonstrated differential allelic binding by EMSA, we retested expanded oligonucleotide sequences surrounding these variants (203 bp for rs147978008

and 205 bp for rs11571537). The rs147978008-203bp also failed to show an allelic difference in gene transcription activity, whereas the rs11571537-205bp non-risk allele showed enhanced luciferase activity (see Fig 2, E). Of the nine genetic variants with EMSA binding, significant genotype-dependent luciferase reporter activity was identified in four (rs11571537, rs2446824, rs2287231, rs2513789).

**Fig 2.** Luciferase reporter assays for genetic variants showing risk (2nd bar) and non-risk (3rd bar) gene expression for variants: **A;** rs2446824-35bp, **B;** rs2287231-35bp, **C;** 2513789-35bp and **E**; rs11571537-205bp, which show allele-dependent gene expression. **D;** rs 147978008-35bp. **F;** rs147978008-203bp, which shows increased expression driven by the deletion mutant (Del).



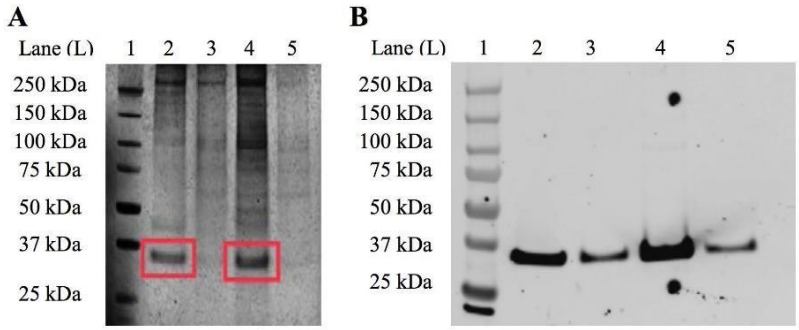
4. Identification of H1 histone binding by rs147978008.

DNA Affinity Purification Assay (DAPA) was subsequently performed to determine specific nuclear proteins bound to rs147978008 using nuclear extract proteins from epithelial cells (A549 and BEAS 2B). Binding activity for the rs147978008 non-risk oligonucleotide probe was assessed using SDS-PAGE analysis of the eluates obtained (Fig 3, A). Strong bands for both extracts were observed for proteins having a molecular size of 29 kDa. These two bands were analyzed by nano liquid chromatography coupled tandem mass spectrometry (nanoLC-MS/MS). In the BEAS 2B band, histones H1.4 (11 peptides match), H1.5 (4 peptides match) and H1.2 (3 peptides match) were clearly identified. For the A549 band, no protein identifications reached the 2 peptide standard, but two proteins were detected, histones H1.2 and H1.3, each with a single 99% confidence peptide.

We confirmed genotype-dependent binding of H1.2 by Western Blot for the risk and non-risk DAPA purified eluates of A549 and BEAS-2B cells (Fig 3, B). Regardless of the cellular source of nuclear lysate, the non-risk allele shows stronger binding of H1.2 than the risk allele, consistent with EMSA. These results indicate that rs147978008 in the promoter of *FAM71A* is bound by H1.2 in a genotype-dependent manner.

**Fig 3. Histone H1.2 binds more strongly to the rs147978008 non-risk allele probe.**

**A.** SDS-PAGE of DAPA purified rs147978008 binding proteins. Analysis of the DAPA eluates of A549 (Lanes 2 & 3) and BEAS 2B (Lanes 4 & 5) cells produced by binding to either the non-risk or risk allele of rs147978008. L1-molecular weight standards, L2-A549 extract/non-risk; L3-A549 extract/risk; L4-BEAS 2B/non-risk; L5-BEAS /risk. **B.** Western Blot of SDS-PAGE gel immunoblotted with rabbit polyclonal anti- H1.2. The lanes correspond to those shown in Panel A. The bands detected have an approximate molecular size of 29 kDa, consistent with the size of H1.2.




5. Genome wide genetic association studies (GWAS) have revealed thousands of variants that correlate with human disease and the vast majority lie in noncoding regions, implying that regulatory variation is an important component in inherited disease risk. However, finding the exact causal variant amongst other variants can be challenging due to linkage. With remaining funds from this RO1 applied to this second no cost extension and funds received from the University of Cincinnati Department of Internal Medicine, we have initiated a study to develop a new screening method, with the goal of determining which genes are regulated, and to identify true disease risk loci. We conducted a pilot study to test a new CRISPR interference model, using stable transduced epithelial lung cell lines (A549 and BEAS2B), expressing dCas9-VP64 (transcriptional activator). Single guide RNAs (gRNA) were synthesized using *in vitro* transcription to generate gRNA libraries for transfection of the stable cell lines. RNA-seq was performed on transfected and non-transfected cells to identify gene expression changes that may validate genetic variants that predict disease risk and lead to discovery of novel biomarkers of occupational asthma. Results from the screen will nominate strong causal variants, potentially identifying the variants that increase risk and demonstrate allele specific behavior related to genetic risk. The data from this study is currently being validated and analyzed.

**References**


1. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5.
2. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362-7.
3. Aizawa M, Fukuda M. Small GTPase Rab2B and Its Specific Binding Protein Golgi-associated Rab2B Interactor-like 4 (GARI-L4) Regulate Golgi Morphology. J Biol Chem. 2015;290(36):22250-61.
4. Bakerly ND, Moore VC, Vellore AD, Jaakkola MS, Robertson AS, Burge PS. Fifteen-year trends in occupational asthma: data from the Shield surveillance scheme. Occup Med (Lond). 2008;58(3):169-74.
5. Marabini A, Brugnami G, Curradi F, Severini C, Siracusa A. [The response to a specific bronchial provocation test and the evolution of occupational asthma. A longitudinal study in subjects with toluene diisocyanate-induced asthma]. Med Lav. 1994;85(2):134-41.
6. Bernstein DI, Kashon M, Lummus ZL, Johnson VJ, Fluharty K, Gautrin D, et al. CTNNA3 (alpha-catenin) gene variants are associated with diisocyanate asthma, a replication study in a Caucasian worker population. Toxicological sciences : an official journal of the Society of Toxicology. 2012.
7. Yucesoy B, Kaufman KM, Lummus ZL, Weirauch MT, Zhang G, Cartier A, et al. Genome-Wide Association Study Identifies Novel Loci Associated With Diisocyanate-Induced Occupational Asthma. Toxicol Sci. 2015;146(1):192-201.
8. Prahalad S, Ryan MH, Shear ES, Thompson SD, Giannini EH, Glass DN. Juvenile rheumatoid arthritis: linkage to HLA demonstrated by allele sharing in affected sibpairs. Arthritis and rheumatism. 2000;43(10):2335-8.
9. Bernstein DI, Lummus ZL, Kesavalu B, Yao J, Kottyan L, Miller D, et al. Genetic variants with gene regulatory effects are associated with diisocyanate-induced asthma. Journal of Allergy and Clinical Immunology. 2018;142(3):959-69.

**Progress Report: Publication List**


1. **Bernstein D,** Kissling, GE, Khurana Hershey G, Yucesoy B, Johnson VJ, Cartier A, Gautrin, D, Sastre J, Boulet LP, Malo JL, Quirce S, Tarlo S, Langmeyer S, Luster MI , Lummus ZL: [2011] Hexamethylene Diisocyanate Asthma is Associated with Genetic Polymorphisms of CD14, IL-13, and IL-4RA, Journal of Allergy and Clinical Immunology, 128:418-420. **PMID: 21489615**
2. Yucesoy B, Johnson VJ, Lummus ZL, Kissling G, Fluharty K, Gautrin D, Malo JL, Cartier A, Boulet LP, Sastre J, Quirce S, Germolec D, Tarlo SM, Cruz MJ, Munoz X, Luster M, **Bernstein DI**: [2012] Genetic variants in antioxidant genes are associated with diisocyanate-induced asthma, Toxicological Sciences, 129:166-173. **PMID:22610343**
3. **Bernstein DI**, Kashon M, Lummus ZL, Johnson VJ, Fluharty K, Gautrin D, Malo JL, Cartier A, Boulet LP, Sastre J, Quirce S, Germolec D, Tarlo SM, Cruz MJ, Munoz X, Luster MI, Yucesoy B: [2013] CTNNA3 (α-catenin) gene variants are associated with diisocyanate asthma, a replication study in a Caucasian worker population. Toxicological Sciences, 131:242-6. **PMID:22977168**
4. Yucesoy B, Johnson VJ, Lummus ZL, Kashon ML, Frye B, Wang W, Marepalli R, Thompson HB, Gautrin D, Malo JL, Cartier A, Boulet LP, Sastre J, Quirce S, Tarlo SM, Germolec DR, Luster MI, **Bernstein DI**: [2014] Genetic variants in the MHC class I and class II genes are associated with diisocyanate-induced asthma, Journal of Occupational and Environmental Medicine, 56:382-387. **PMID: 24709764**
5. Lummus ZL, Wisnewski AV, **Bernstein DI:** [2011] Pathogenesis and disease mechanisms of occupational asthma. Immunology and Allergy Clinics of North America, 31:699–716. **PMID:21978852**
6. Malo J-L, L'Archeveˆque J, Lummus Z, **Bernstein DI:** [2006] Changes in specific IgE and IgG and monocyte chemoattractant protein-1 in workers with occupational asthma caused by diisocyanates and removed from exposure. Journal of Allergy and Clinical Immunology, 118:530-533. **PMID:16890787**
7. Campo P, Wisnewski AV, Lummus Z, Cartier A, Malo J-L, Boulet LP, **Bernstein D:** [2007] Diisocyanate conjugate and immunoassay characteristics influence detection of specific antibodies in HDI-exposed workers. Clinical & Experimental Allergy, 37:1095-1102. **PMID:17581205**
8. **Bernstein DI**. Genetics of occupational asthma: [2011] Current Opinion in Allergy and Clinical Immunology, 11:86-89. **PMID:21325943**
9. Ouyang B, Bernstein DI, **Lummus ZL**, Ying J, Boulet LP, Cartier A, Gautrin D, Ho SM.[2013] Interferon-γ promoter is hypermethylated in blood DNA from workers with confirmed diisocyanate asthma. Toxicol Sci.133:218-24. **PMID:** 23535363
10. Yucesoy B, Kissling GE, Johnson VJ, Lummus ZL, Gautrin D, Cartier A, Boulet LP, Sastre J, Quirce S, Tarlo SM, Cruz MJ, Munoz X, Luster MI, **Bernstein DI**:[2015] N-Acetyltransferase 2 Genotypes Are Associated With Diisocyanate-Induced Asthma. J Occup Environ Med. 57:1331-6. **PMID:** 26641831
11. Yucesoy B, Kaufman KM, Lummus ZL, Weirauch MT, Zhang G, Cartier A, Boulet LP, Sastre J, Quirce S, Tarlo SM, Cruz MJ, Munoz X, Harley JB, **Bernstein DI**: [2015] Genome-Wide Association Study Identifies Novel Loci Associated With Diisocyanate-Induced Occupational Asthma. Toxicol Sci. 146:192-201. **PMID:** 25918132.
12. Yucesoy B, Kashon ML, Johnson VJ, Lummus ZL, Fluharty K, Gautrin D, Cartier A, Boulet LP, Sastre J, Quirce S, Tarlo SM, Cruz MJ, Munoz X, Luster MI, **Bernstein DI**.[2016] Genetic variants in TNFα, TGFB1, PTGS1 and PTGS2 genes are associated with diisocyanate-induced asthma. J Immunotoxicol. 13:119-26.**PMID:** 25721048
13. Sabbioni G, Vanimireddy LR, Lummus ZL, **Bernstein DI**.[2017] Comparison of biological effects with albumin adducts of 4,4'-methylenediphenyl diisocyanate in workers.Arch Toxicol. 91:1809-1814. **PMID:** 27638504
14. **Bernstein DI**, Lummus ZL, Kesavalu B, Yao J, Kottyan L, Miller D, Cartier A, Cruz MJ, Lemiere C, Muñoz X, Quirce S, Tarlo S, Sastre J, Boulet LP, Weirauch MT, Kaufman K:[2018] Genetic variants with gene regulatory effects are associated with diisocyanate-induced asthma. J Allergy Clin Immunol. 142:959-969. **PMID:** 29969634