2015

Generating CRISPR-Cas9 genome-engineered human embryonic stem cell to model a genetic mechanism of asthma

https://hdl.handle.net/2144/16194

Boston University
GENERATING CRISPR-CAS9 GENOME-ENGINEERED HUMAN EMBRYONIC STEM CELLS TO MODEL A GENETIC MECHANISM OF ASTHMA

by

SEAN J. MCMANUS

B.S., Villanova University, 2012

Submitted in partial fulfillment of the requirements for the degree of Master of Science

2015
Approved by

First Reader

Jean L. Spencer, Ph.D.
Instructor of Biochemistry

Second Reader

Paul H. Lerou, M.D.
Medical Director of Neonatal Intensive Care Unit, Brigham and Women's Hospital
ACKNOWLEDGMENTS

I would like to thank Dr. Paul Lerou from the Department of Pediatric Newborn Medicine at Brigham and Women’s Hospital for allowing me to conduct research in his laboratory. I appreciate your patience while bringing me up to speed on various laboratory techniques and your sage advice about life at the bench. I am grateful for your affording me the independence needed in science to approach problems with a creative mindset, while also leveraging your experience to guide me in the right direction. The laboratory has been a large part of my life over the past year, and my growth both scientifically and personally would not have been possible without your steady encouragement to keep pushing the envelope both in my research interests and in my pursuit of a medical career. I would especially like to thank Dr. Jeanne Carroll and Dr. Junjie Lu for their incredible support and guidance on a day-to-day basis. I would also like to thank Dr. Junning Wang for her expertise in any and all cell culture matters, and Dr. Bryan Gorman for bringing me up to speed on the technicalities of image analysis and MATLAB programming. Finally, I would say a special thank you to my family members who have been my rock throughout my unique journey.
GENERATING CRISPR-CAS9 GENOME-ENGINEERED HUMAN EMBRYONIC STEM CELLS TO MODEL A GENETIC MECHANISM OF ASTHMA

SEAN J. MCMANUS

ABSTRACT

Asthma is a major public health epidemic that presents a heavy burden on those who suffer from the disease. Little is currently understood about the genetic signature that distinguishes one type of asthma from another. Recently, the single nucleotide polymorphism (SNP) rs968567 was found to have a high degree of association in asthmatic patients (Sharma et al., 2014). This particular SNP is in the promoter region of the FADS2 gene that synthesizes the enzyme delta-6-desaturase (D6D). D6D mediates the formation of pro-inflammatory factors that lead to exacerbation of asthmatic symptoms. We engineered a novel, customized CRISPR-Cas9 construct to induce the SNP rs968567 in the HUES9 human embryonic stem cell (hESC) line. Our results show success in generating the custom CRISPR-Cas9 construct for use in stem cells, while efficiency in expressing the desired mutation in our cell line is currently being optimized. Disease modeling in the genomic era of medicine provides an opportunity for the development of personalized medical treatment. Future projects aim to differentiate stem cell lines edited with our CRISPR-Cas9 construct to lung progenitor cells to study the cellular phenotype of this mutation in context of asthma pathogenesis.
# TABLE OF CONTENTS

**TITLE** ............................................................................................................................................... i

**COPYRIGHT PAGE** ............................................................................................................................... ii

**READER APPROVAL PAGE** .................................................................................................................. iii

**ACKNOWLEDGMENTS** ............................................................................................................................ iv

**ABSTRACT** ............................................................................................................................................. v

**TABLE OF CONTENTS** .......................................................................................................................... vi

**LIST OF TABLES** ..................................................................................................................................... viii

**LIST OF FIGURES** ................................................................................................................................... ix

**LIST OF ABBREVIATIONS** ..................................................................................................................... x

**INTRODUCTION** ...................................................................................................................................... 1

  * Prevalence of Asthma ................................................................. 1
  * Asthma Diagnosis and Symptoms ................................................... 3
  * Genetics of Asthma ............................................................. 4
  * Role of Desaturase Enzymes in Asthma ........................................... 6
  * CRISPR-Cas9 Genome Engineering ................................................... 9
  * Specific Aims/Objectives ............................................................ 12

**METHODS** .......................................................................................................................................... 13

  * CRISPR-Cas9 Customized Construct ................................................ 13
## LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Surveyor Mutation Detection Kit Hybridization Schedule</td>
<td>27</td>
</tr>
</tbody>
</table>
LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Metabolic pathway of n-3 and n-6 PUFAs</td>
<td>28</td>
</tr>
<tr>
<td>2</td>
<td>CRISPR-Cas9 mechanism of action</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>CRISPR-Cas9 customized construct and FADS2 gene sequence</td>
<td>30</td>
</tr>
<tr>
<td>4</td>
<td>FACS results for HUES9 cells post-nucleofection</td>
<td>31</td>
</tr>
<tr>
<td>5</td>
<td>Surveyor mutation detection gel</td>
<td>32</td>
</tr>
<tr>
<td>6</td>
<td>Cell culture flow chart</td>
<td>33</td>
</tr>
<tr>
<td>7</td>
<td>Sanger sequencing results for HUES9 cell line</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td>Immunofluorescence imaging of definitive endoderm</td>
<td>35</td>
</tr>
<tr>
<td>9</td>
<td>Immunofluorescence imaging of anterior foregut endoderm</td>
<td>36</td>
</tr>
</tbody>
</table>
LIST OF ABBREVIATIONS

AA.................................................................................................Arachidonic Acid
ALA .................................................................................................\(\alpha\)-Linoleic Acid
C........................................................................................................Cytosine
Cas9...............................................................................................CRISPR-Associated Protein 9
CCR5..............................................................................................Cysteine-Cysteine Chemokine Receptor 5
CD4\(^+\)..............................................................................................Cluster of Differentiation 4
CDC .................................................................................................Centers for Disease Control and Prevention
CRISPRs .................................................................Clustered Regularly Interspaced Short Palindromic Repeats
DAPI.................................................................................................4',6-Diamidino-2-Phenylindole
D5D.................................................................................................\(\delta\)-5-Desaturase
D6D.................................................................................................\(\delta\)-6-Desaturase
DHA..............................................................................................Docosahexaenoic Acid
DNA.................................................................................................Deoxyribonucleic Acid
E8.................................................................................................Essential 8 Medium
eQTL...............................................................................................Expression Quantitative Trait Locus
eSNPs..............................................................................................Expression-Associated Variants
FACS..............................................................................................Fluorescence-Activated Cell Sorting
FADS1............................................................................................Fatty Acid Desaturase 1
FADS2............................................................................................Fatty Acid Desaturase 2
\(\text{F}_{\text{ENO}}\)........................................................................................Fractional Exhaled Nitric Oxide
FITC-A ................................................................. Fluorescein Isothiocyanate - Area
FoxA2 ...................................................................... Forkhead Box Protein A2
FSC-A ...................................................................... Forward-Scattered Light – Area
FSC-H ...................................................................... Forward-Scattered Light - Height
GFP ........................................................................ Green Fluorescent Protein
GWAS ................................................................. Genome-Wide Association Study
HDR ................................................................. Homology-Directed Repair
hESC ....................................................................... Human Embryonic Stem Cell
IgE ........................................................................ Immunoglobulin E
Indel ........................................................................ Insertion or Deletion
IL-5 ........................................................................ Interleukin-5
IL-13 ........................................................................ Interleukin-13
LA ........................................................................ Linoleic Acid
LB ........................................................................ Lysogeny Broth
NHEJ ................................................................. Nonhomologous End-Joining
Oct4 ........................................................................ Octamer-Binding Transcription Factor 4
P1-5 ........................................................................ Events 1-5
PAM ........................................................................ Protospacer Adjacent Motif
PBS ........................................................................ Phosphate-Buffered Saline
PCR ........................................................................ Polymerase Chain Reaction
PE-A ........................................................................ Phycoerythrin - area
PUFAs ........................................................................ Polyunsaturated Fatty Acids
RNA .......................................................... Ribonucleic Acid
ROCK ........................................................ Rho-Associated Protein Kinase
rpm ............................................................. Rotations per Minute
sgRNA ..................................................... Short Guide RNA
SNP .......................................................... Single Nucleotide Polymorphism
Sox2 .................................................... (Sex-Determining Region Y)-Box 2
Sox17 .................................................... (Sex-Determining Region Y)-Box 17
SSC-A .......................................................... Side-Scattered Light - Area
ssODN ................................................... Single-Stranded Oligodeoxynucleotide
T ................................................................. Thymine
Th1 ............................................................ T-Helper Type 1
Th2 ............................................................ T-Helper Type 2
INTRODUCTION

Prevalence of Asthma

Asthma has become a steadily growing epidemic, especially in the Western world, and is now a major public health concern for people of any gender, age, or race. As the scientific community continues to expand its knowledge of environmental cues that exacerbate asthma and common genetic links among asthmatic patients, the push to find a cure for asthma is greater than ever. As of now, there are no preventative medical measures that can be taken to stop asthma from developing in either children or adults, and there is no cure for asthma once an individual has developed it.

A variety of studies have examined asthma from a public health perspective in an effort to quantify the burden asthma places on both the United States and global populations. From 2001 to 2010 the number of Americans diagnosed with asthma increased from 7.3% to 8.4% (25.7 million Americans total) (Akinbami et al., 2012). Asthma was responsible for 3,447 deaths in the United States in 2007 (Akinbami et al., 2011). On a global scale, asthma affects approximately 300 million individuals (Croisant, 2014). While asthma may sometimes be dismissed as a manageable burden, a 2012 study by the Asthma and Allergy Foundation of America determined that asthma is the number one cause of missed school days in the United States among school-aged children (Meng et al., 2012). From a financial standpoint, the Centers for Disease Control and Prevention (CDC) estimated that asthma accounts for $56 billion in expenses and lost productivity of employees (CDC 2011 as cited in Croisant, 2014). This burden is projected to grow at an
alarming rate as the World Health Organization predicts the number of diagnosed
asthmatic patients will increase by 100 million by the year 2025 (World Health

Childhood asthma is a subset of asthma that is particularly poorly understood. A
variety of studies have examined a host of environmental factors that influence a child’s
susceptibility to asthma and other atopic diseases. For instance, a study from 2011
demonstrated that children raised on a farm had a significantly lower incidence of atopic
disease as compared to a metropolitan cohort (Ege et al., 2011). Another study by Ferry
et al. (2014) demonstrated that children who lived in households with carpet before the
age of 2 were 1.4 times more likely to be diagnosed with asthma between ages 2 and 6
than those who were eventually diagnosed with asthma but did not grow up in a
household with carpet. Furthermore, if a child experienced a serious respiratory illness
and grew up in a carpeted household all before the age of 2 years, then his or her risk of
developing asthma between ages 2 and 6 increased 3.2-fold (Ferry et al., 2014).

Asthma has also been recently linked to early childhood exposures to diesel
exhaust particle, otherwise known as traffic pollution (Codispoti et al., 2015). Children
who are exposed to diesel exhaust particle by 1 year of age show a positive correlation
with developing aeroallergen sensitization by ages 2 and 3 (Codispoti et al., 2015). This
erly aeroallergen sensitization offers an interesting avenue for future research, since it
was shown to be a risk factor for patients developing asthma at some point in their life
(Settipane and Settipane, 2000). This finding by Settipane and Settipane in a 23-year
follow-up with students from Brown University led them to describe the “hygiene
hypothesis” for developing asthma. The hypothesis says that of the two primary immunological pathways, T-helper 1 (Th1) and T-helper 2 (Th2), the Th1 pathway has slowly faded with the onset of the era of immunizations and public hygiene awareness. Settipane and Settipane (2000) cited the results of this study to support their hypothesis that the repressed Th1 pathway allows the Th2 pathway to predominate in the form of increased immunoglobulin E (IgE) levels, eosinophilia, atopy, and airway hyperresponsiveness. However, the hygiene hypothesis is not universally supported, as recent research points to asthma being triggered by alternative factors such as diet, obesity, infection, and T-cell skewing (Woodruff et al., 2009; Yang and Schwartz, 2012).

**Asthma Diagnosis and Symptoms**

Asthma is a chronic inflammatory respiratory disease that manifests itself in the form of varying degrees of bronchoconstriction, plasma excretion, and mucus secretion in the airway (Barnes, 1996). As a result, the heterogeneity of asthma is what makes this disease so difficult to diagnose and to treat adequately (Busse, 2011). A combination of clinical history, family history, and patient response to medication typically dictates the course of treatment (Busse, 2011). The most common way to diagnose childhood asthma is through an assessment of symptom severity, frequency of exacerbations, and degree of interference with normal activity (Potter, 2010). One quantitative measure that physicians use to diagnose asthma is through an airflow reversibility measurement (Busse, 2011). If, after administration of a short-acting bronchodilator, a patient demonstrates an increase of at least 12% in their forced expiratory volume, then the patient is diagnosed with
asthma contingent upon supporting clinical symptoms (Busse, 2011). One way in which clinicians are able to stratify between moderate and severe asthma is on the basis of fractional exhaled nitric oxide (FENO) (Fitzpatrick et al., 2006). Patients with severe asthma were found to have 2-fold higher levels of FENO than those with moderate asthma (Fitzpatrick et al., 2006).

Asthma symptoms are associated with airway hyperresponsiveness and include coughing, wheezing, and breathlessness (Kaufman, 2011). Lymphocytes, eosinophils, and mast cells that have leaked into the airway lumen are primarily responsible for triggering the inflammatory response underlying asthma (Barnes, 1996). In fact, the presence of eosinophils in the airway lumen of asthmatic patients is so predominant that asthma was first referred to as “chronic eosinophilic bronchitis” as far back as 1916 (Barnes, 1996). This chronic inflammatory state, exacerbated by bouts of acute inflammation in response to environmental stressors, results in the shedding of airway epithelium, which triggers airway hyperresponsiveness (Barnes, 1996). Studies examining the unique microenvironment of airway epithelium in the post-shedding state at the level of electron microscopy have added insight into the pathophysiology of asthma. For instance, an earlier study revealed that shedding of airway epithelium in asthmatic patients creates a microenvironment conducive to production of type III and type V collagen, subsequently causing subepithelial fibrosis of the airway (Djukanović et al., 1990).

Genetics of Asthma
Understanding the genetic signature of asthma is a new frontier in the field of asthma research; challenges in this field include a lack of reproducibility across studies, along with hundreds of potential candidate genes (Yang and Schwartz, 2012). While progress is being made in terms of defining environmental triggers that predict asthma susceptibility at a young age, our knowledge of the genetic regulation of asthma has lagged behind. Asthma was first hypothesized to have a genetically inherited component by Dr. Bonnie Sibbald in the late 1970s. Ironically, her study in 1980 (Sibbald et al., 1980) failed to statistically prove the presence of a common genetic defect among patients with either asthma or wheezy bronchitis, but the study piqued the interest of the scientific community with respect to a genetic link among patients with similar obstructive airway conditions. Subsequent twin studies have shown that asthma heritability is 36%-70%. This heritability is likely due to complex genetic and environmental interactions involving more than 100 genes with diverse impacts from association with immunoregulation to epithelial structure and function (Sibbald et al., 1980).

Recent evidence suggests that epigenetics may play a strong role in how asthma is inherited. For instance, a mother who has asthma is much more likely to pass down asthma to her children than a father who has asthma (Moffat et al., 1998). The epigenetic control of this parent-of-origin phenomenon was demonstrated in mice by feeding pregnant mice either a high- or low-methylation diet and observing the phenotype of their offspring (Hollingsworth et al., 2008). Hollingsworth’s group showed that mice fed with a high-methylation diet had progeny that demonstrated airway hyperresponsiveness,
airway inflammation, and increased serum IgE levels. This phenotype was reversible upon supplementing with the demethylating agent 5-aza-deoxycytidine. With continued evidence of DNA methylation affecting asthmatic symptoms, it is possible that DNA methylation levels from nasal or buccal samples and peripheral blood cells could be used as a biomarker for drug targets (Yang and Schwartz, 2012).

Building on the hygiene hypothesis and its predilection toward the Th2 pathway mediating asthmatic symptoms, a recent study managed to stratify asthmatic patients according to their Th2 molecular phenotype (Woodruff et al., 2009). This study classified patients based on levels of interleukin-13 (IL-13) and interleukin-5 (IL-5) as “Th2-high” and “Th2-low.” The Th2-high group showed classical symptoms of severe asthma with elevated serum IgE levels, high eosinophil count, high degree of airway hyperresponsiveness, and thickening of the reticular basement membrane. Interestingly, only Th2-high patients saw clinical benefit from use of inhaled corticoid steroids. Though asthma can have myriad genetic signatures, this study provides evidence that classifying patients as Th2-high or Th2-low allows a clinician to stratify patients and tailor treatment accordingly. Understanding the complex nature of genetic susceptibility to asthma may help personalize treatment based on an individual’s genetic profile.

**Role of Desaturase Enzymes in Asthma**

A recent genome-wide association study identified the fatty acid desaturase 2 (FADS2) enzyme, involved in the biosynthesis of polyunsaturated fatty acids (PUFAs), as strongly linked to symptomatic asthma patients (Sharma et al., 2014). PUFAs are
categorized into the two subfamilies n-3, originating from α-linoleic acid (ALA), and n-6, originating from linoleic acid (LA) (Calder, 2008). ALA and LA cannot be synthesized by the human body, which requires them to be consumed through one’s diet (Calder, 2008). As such, consumption of ALA and LA in the Western diet accounts for 95%-98% of PUFAs in the human body (Calder and Burdge, 2004). Common foods in which LA is found in large amounts include vegetable oils (corn, soybean, and sunflower oils) and products made from margarine. ALA is more commonly found in nuts, flaxseed, and flaxseed oil. Given the evolution of the Western diet over the past half century to feature foods higher in LA content, the ratio of LA:ALA in the body has significantly increased (Calder and Burdge, 2004).

PUFAs are metabolized from ingested ALA and LA through the addition of double bonds by desaturase enzymes and the elongation of the fatty acid chain (Figure 1) (Calder, 2008). Ultimately, ALA is converted to docosahexaenoic acid (DHA) of the n-3 family, and LA is metabolized to arachidonic acid (AA) of the n-6 family (Calder, 2008). AA is important in the pathophysiology of asthma because it is the precursor to pro-inflammatory eicosanoids including prostaglandins, thromboxanes, and leukotrienes (Calder, 2008). As such, increased levels of AA in the body necessarily yield higher levels of pro-inflammatory eicosanoids which exacerbate asthma and other chronic inflammatory conditions (Martinelli et al., 2008). The two enzymes responsible for the synthesis of both DHA and AA are δ-5 desaturase (D5D) and δ-6 desaturase (D6D) (Martinelli et al., 2008). More specifically, D6D occurs upstream of D5D in the biosynthetic pathway of PUFAs and converts LA to γ-linoleic acid in the n-6 family and
ALA to stearidonic acid in the n-3 family (Calder, 2008; Martinelli et al., 2008). D5D later converts dihomo-γ-linoleic acid to AA in the n-6 family and eicosatetraenoic acid to eicosapentaenoic acid in the n-3 family (Calder, 2008; Martinelli et al., 2008). D5D is encoded on the fatty acid desaturase 1 (FADS1) gene, while D6D is encoded on the FADS2 gene.

Many groups have shown a strong correlation between increased desaturase activity and inflammatory diseases. For instance, a study by Martinelli et al. (2008) demonstrated an increase in D6D activity among patients with coronary artery disease compared with healthy individuals. Martinelli’s group examined the ratio of AA:LA on red blood cell membranes to establish a difference in desaturase activity between the two cohorts. This difference was explained by the inflammatory state of the arterial wall during atherosclerosis (Martinelli et al., 2008). Upregulation of D6D activity has been linked to metabolic diseases, especially type II diabetes (Kroger et al., 2012). However, the identification of the particular SNPs that cause an increase in D6D activity require further study. One thought as to why D6D might be implicated in the pathogenesis of asthma is that the fatty acid composition of cell membranes could impact the effectiveness of glucose transporters and affect insulin sensitivity (Storlien et al., 1996). A different study by Lattka et al. (2010) analyzed the FADS2 gene at a transcriptional level and found that the SNP rs968567, a cytosine (C) to thymine (T) transition mutation 299 base pairs upstream of the transcription start site, yielded a 2-to 3-fold increase in luciferase reporter gene assay. In other words, the C to T SNP resulted in a 2-to 3-fold
increase in expression of the FADS2 gene. Similarly, the C to T mutation resulted in increased DNA-binding ability of nuclear proteins (Lattka et al., 2010).

A recent study by Sharma’s group employed expression quantitative trait locus (eQTL) mapping to investigate 6,706 cis-acting expression-associated variants (eSNPs) in cluster of differentiation 4 (CD4+) lymphocytes in hopes of pinpointing genetic variants associated with asthma (Sharma et al., 2014). This study clearly showed associations with asthma for eSNPs in the FADS2 gene. The C allele on chromosome 11q demonstrated an increased risk of asthma in this study. Of the two eSNPs found to be of statistically significant importance on the FADS2 gene, the SNP rs968567 in the conserved promoter region explained 30% of the variance of expression of this gene, while SNP rs174611 explained 11% of the variance. Further evidence was found to support FADS2 gene expression as an important asthmatic marker by demonstrating increased FADS2 mRNA expression in CD4+ lymphocytes in the study. Additionally, the eSNPs rs968567 and rs174611 were found in segments of DNA showing histone modifications conducive to enhancer activity (Sharma et al., 2014).

**CRISPR-Cas9 Genome Engineering**

The implementation of the clustered regularly interspaced short palindromic repeats-CRISPR-associated protein 9 (CRISPR-Cas9) genome editing system in mammalian cells has made a remarkable impact on how scientists are able to precisely target sequences of interest in the genome. CRISPRs were first found in *E. coli* by Ishino’s group in 1987 as part of the bacterial adaptive immune system (Garneau et al.,
In the last decade scientists began to realize the potential for harnessing this bacterial defense strategy for use in mammalian cells to edit the genome. The CRISPR-Cas9 system improves upon previous genome editing techniques, such as zinc finger nucleases and transcription activator-like effector nucleases, because it demonstrates significantly fewer off-target mutations, and it only requires the engineering of a short guide RNA (sgRNA), instead of entire proteins, which can be costly and inefficient (Jiang et al., 2013).

The CRISPR-Cas9 system is comprised of CRISPR sequences, a sgRNA, and a Cas9 nuclease used to induce a DNA double-stranded break at the point of interest (Cong et al., 2013). Jinek’s group showed that the Cas9 nuclease is made up of a RuvC domain which is responsible for cleaving the non-target strand and an HNH domain which is the primary domain responsible for cleaving the DNA (Jinek et al., 2014). When the CRISPR-Cas9 system is used the DNA repairs itself by either nonhomologous end-joining (NHEJ) or homology-directed repair (HDR) (Mali et al., 2013). When either the RuvC or HNH domain is inactivated, the ratio of repair by NHEJ to HDR increases while toxicity decreases (Mali et al., 2013). The sgRNA that is part of this CRISPR-Cas9 system is approximately 20 base pairs and is chosen so that it binds by means of nucleotide base pairing with the DNA sequence of interest. Pattanayak’s group showed that the shorter, less active sgRNAs are more specific and less tolerant of mutations than the longer sgRNAs (Pattanayak et al., 2013). The protospacer adjacent motif (PAM) sequence of DNA is just downstream of the target sequence and is important for binding of the Cas9 nuclease (Cong et al., 2013).
The CRISPR-Cas9 system has wide-ranging impacts from designing phage-resistant bacteria used in industrial fermentation processes to administering functional genes to patients in a clinical setting (Hsu et al., 2014). This new technique is also very useful for genetic screening in that it now allows one to inactivate both copies of a gene in diploid mammalian cells which was previously very difficult to achieve (Wang et al., 2014). Disease modeling is another application for which the CRISPR-Cas9 system shows tremendous potential because researchers can insert or delete certain genes with ease to study treatments for specific diseases. For instance, Lombardo’s group showed that in lymphocytes if you are able to inactivate the cysteine-cysteine chemokine receptor 5 (CCR5) receptor by NHEJ, then it may be a potential method for avoiding HIV infection (Lombardo et al., 2007). A group at MIT was recently able to utilize the CRISPR-Cas9 system to correct in mice a mutation in fumarylacetoacetate hydrolase, an enzyme playing a role in the catabolism of tyrosine that leads to hereditary tyrosinemia (Yin et al., 2014). Another study demonstrated the ability to rescue mice progeny from cataracts caused by a one base pair deletion in the critical Crygc gene by injecting zygotes with a CRISPR-Cas9 construct and correcting the mutation (Wu et al., 2013).

With respect to the research presented in the following pages, the CRISPR-Cas9 system was used to design a custom construct to induce the SNP rs968567 so we could model in human embryonic stem cells (hESCs) the C to T mutation in the conserved FADS2 promoter region which upregulates desaturase activity and is thought to be present in genetically inherited forms of asthma.
Specific Aims/Objectives

The first objective of this study was to design a customized CRISPR-Cas9 construct that could be successfully inserted into a hESC line of interest. The CRISPR-Cas9 construct provides a way to induce a double-stranded break in the DNA at a precise location (Figure 2). This break then allows the cell to repair by either HDR or NHEJ. The second objective was to design a single-stranded oligodeoxynucleotide (ssODN) to direct DNA repair after cleavage by our CRISPR-Cas9 construct such that a C to T transition mutation occurs in the highly conserved FADS2 promoter region 299 base pairs upstream from the transcription start site. This mutation is known as the SNP rs968567 and has been shown to have a strong correlation with inheritance of asthma. The third objective of this research was to successfully generate a cell line with the mutation of interest in a hESC line. This cell line could then be utilized to study the effect of this FADS2 mutation on lung development through differentiation of the hESCs to lung progenitor cells. The beauty of disease modeling in embryonic stem cells instead of mature lung cells is that one can take a stem cell line with this FADS2 polymorphism and direct it to become almost any other cell type in the human body. This allows one to study the effects of this mutation on a variety of other organ systems.
METHODS

CRISPR-Cas9 Customized Construct

In order to induce the targeted SNP into our HUES9 hESC line, the CRISPR-Cas9 construct was engineered using the protocol described by Ran et al. (2013). This SNP results in a C to T transition mutation. At this locus 299 base pairs upstream of the transcription start site in the FADS2 promoter region, 92.75% of the population has a cytosine base pair, and the remaining 7.25% has the disease-associated thymine base pair. The sgRNA sequence of the CRISPR construct was chosen so that it hybridized with a 20 base pair region spanning the SNP of interest and meeting the qualification of proximity to a PAM sequence (Figure 3). This sequence was selected using the CRISPR design tool (http://crispr.mit.edu) for its proximity to the target site and limited predicted off-target effects. This allowed the Cas9 nuclease to cleave at the desired base pair. The ssODN was designed to span the target region with homology arms of approximately 60 base pairs on either side. The ssODN differed from the target sequence by one base pair (the SNP of interest). CRISPR plasmid pSpCas9(BB)-2A-GFP (PX458) from the Feng Zhang lab was obtained through Addgene (plasmid #48138; Cambridge, MA). The sgRNA sequence was cloned into the CRISPR vector which also carried Cas9 and green fluorescent protein (GFP) for positive selection. Successful cloning was confirmed by Sanger sequencing performed by GeneWiz (Cambridge, MA). The plasmid co-expressing the sgRNA, Cas9 nuclease, and GFP was amplified in lysogeny broth (LB)-ampicillin culture medium. Plasmid DNA was then purified in preparation for delivery of the CRISPR-Cas9 construct into the HUES9 hESC line by nucleofection.
**CRISPR-Cas9 Delivery by Nucleofection**

HUES9 hESCs were cultured for nucleofection in three wells of a 6-well plate on Matrigel (Corning, Tewksbury, MA) in 2 mL Essential 8 Basal Medium (E8) and E8 Supplement (Life Technologies, Grand Island, NY). Medium was replaced every 24 hours in each well. Cells were grown to about 80% confluency before carrying out the following Amaxa nucleofection procedure (Lonza, Walkersville, MD). A quantity of 10 µM of Rho-associated protein kinase (ROCK) inhibitor (Y-27632) was added to the medium in HUES9 wells 3 hours before electroporation. Cells were dissociated from the substrate using Accutase (StemCell Technologies, Vancouver, Canada), which ensured single-cell dissociation. Dissociated cells were counted under the microscope, and 8 x 10^5 cells were resuspended in 100 µL of Nucleofector Solution 1. A combination of 2.5 µg of plasmid DNA and 2.5 µg of ssODN repair template P was added to the 100 µL cell suspension. The 100 µL solution was transferred to an Amaxa nucleofector cuvette and allowed to run through the B-016 program in the nucleofector (settings are proprietary technology). After nucleofection, 0.5 mL E8 was added to each of the three nucleofector cuvettes and the resulting solution was evenly plated in three wells of a 6-well plate. The other three wells of the 6-well plate were seeded with HUES9 cells without nucleofection to serve as a negative control for future fluorescence-activated cell sorting (FACS). Twenty-four hours after seeding nucleofected cells, the medium was replaced with E8 medium with no ROCK inhibitor.
Fluorescence-Activated Cell Sorting (FACS) of GFP\(^+\) HUES9 Cells

Forty-eight hours after nucleofection, fluorescence microscopy was used to confirm the transfection of the CRISPR-Cas9 construct into the cells by presence of a green signal. After visual confirmation, cells were dissociated from each well using Accutase and combined into one 15 mL centrifuge tube. FACS was carried out courtesy of the on-site Brigham and Women’s Hospital FACS Core using their standard procedures. Post-FACS cells (GFP\(^+\)) were determined to have a concentration of 1.5 x 10\(^4\) cells/mL. In a 6-well plate, 300 µL of this cell suspension was plated in each of three wells and 1000 µL was plated in one well. Post-FACS medium made up of E8, 10 µM ROCK inhibitor, and 1X penicillin/streptomycin was used to culture the cells for the first two days post-plating. Medium was changed every 24 hours, and on day three the medium was switched to standard E8 medium.

Surveyor Mutation Detection Kit

The Surveyor Mutation Detection Kit (Integrated DNA Technologies, Coralville, IA) was used to scan the heteroduplex DNA created by our engineered CRISPR-Cas9 construct and to cleave the DNA using the Surveyor nuclease on the 3’ end of both known and unknown polymorphisms. We used the HotStarTaq Plus Master Mix Kit (Qiagen, Limburg, Netherlands) to amplify reference DNA from wild type HUES9 cells as well as DNA collected from a mixed cell population after nucleofection with CRISPR construct both before and after FACS. The polymerase chain reaction (PCR) products hybridized to form homo- and hetero-duplexes by running the program listed in Table 1.
in the C1000 Thermal Cycler (Bio-Rad, Waltham, MA). Once the heteroduplexes formed, PCR products were treated with the Surveyor nuclease. The reference DNA was also treated with the Surveyor nuclease to serve as a negative control. After Surveyor nuclease, Surveyor enhancer, and 0.15 M MgCl₂ were added, the solution was allowed to incubate at 42 °C for 60 minutes. The kit “stop solution” was added to terminate the reaction. A 2% agarose gel was then used to analyze the DNA fragments (reference without nuclease treatment, reference with nuclease treatment, mixed cell DNA prior to FACS, and mixed cell DNA post-FACS) to determine if Surveyor nuclease cut the 636 base pair DNA fragment into a 126 base pair fragment and a 510 base pair fragment.

**Cell Culture Subcultivation**

After the GFP⁺ sorted cells were plated, they were cultured until colonies were large enough to passage, but still isolated representing a colony generated from a single cell. Seventy-two colonies were picked and transferred to individual wells on three 24-well plates. Each well was given a unique indexing code (72 unique wells total). These cells were cultured in 0.5 mL of E8 per well. When each well was confluent they were split 1:2, with one plate serving as a maintenance plate and the other as a duplicate. These cells were maintained until DNA could be collected from each line and a stock was frozen back.

**2.6 DNA Isolation**
After the second passage onto 24-well Matrigel plates, confluent wells were selected for DNA isolation. DNA isolation was carried out according to the procedure outlined by Laird et al. (1991). A 0.5 mL quantity of lysis buffer (prepared according to protocol) was added to each well ready for harvesting and allowed to incubate in the Excella E24 Incubator Shaker series (New Brunswick, Edison, NJ) series set at 37 °C and 180 rotations per minute (rpm) for 2.5 hours. During incubation, 1.5 mL sterile microfuge tubes were labeled with a corresponding well identification code, and 200 µL DNA recovery solution (prepared according to protocol) was added to each microfuge tube. After incubation, 500 µL isopropanol was added to each well, two wells at a time, and the plate was shaken vigorously to initiate DNA precipitation. A 2.5 µL pipette tip was used to gather DNA strands and transfer them to the respective microfuge tube. After all DNA had been transferred to microfuge tubes, five or six tubes were packed into 50 mL conical tubes and taped vertically to the inside of the Hybaid Shake ‘n’ Stack hybridization oven (Thermo Scientific, Waltham, MA). The samples were allowed to spin for 24-48 hours at 37 °C and were then tested for DNA concentration. Samples with at least 30.0 ng/µL DNA were amplified using the HotStarTaq Master Mix Plus PCR Kit and sent off for sequencing to test for presence of desired FADS2 promoter mutation.

After DNA isolation and amplification, it was necessary to run a purification procedure that prepared the amplified DNA product for Sanger sequencing. The PCR purification procedure was followed according to the QIAquick Spin Kit (Qiagen, Limburg, Netherlands) by first adding 250 µL of supplied PBI buffer to 50 µL of our PCR product. This solution was transferred to a QIAquick spin column and centrifuged
for 60 seconds at 17,900 rpm. The “flow-through” was discarded, and 750 µL of supplied PE buffer was added to the QIAquick spin column. This was again centrifuged for 60 seconds at 17,900 rpm. The “flow-through” was discarded, and the column was centrifuged one last time for 60 seconds at 17,900 rpm to ensure the removal of any excess ethanol. The QIAquick spin column was placed in a sterile 1.5 mL microfuge tube, and 50 µL of supplied EB buffer was added directly to the membrane of the column. The column was allowed to sit for 60 seconds, and then it was centrifuged for 60 seconds at 17,900 rpm. DNA concentration of the now-purified PCR product was measured on the NanoDrop 2000 spectrophotometer (ThermoScientific, Waltham, MA) and sent to GeneWiz for Sanger sequencing.
RESULTS

*Engineering CRISPR-Cas9 Construct to Induce SNP rs968567 in FADS2 Gene*

Based on the recent study by Sharma et al. (2014) that demonstrated a strong association between a mutation in the FADS2 promoter region and asthmatic patients, our aim was to induce this polymorphism and model the phenotype in hESCs. The mutation of interest was SNP rs968567 which results in a C to T transition mutation in the promoter region of the FADS2 gene 299 base pairs upstream from the transcription start site. This SNP accounts for 30% of the variance in FADS2 expression, making it an intriguing candidate to study for its effects on diseases associated with chronic inflammation (Sharma et al., 2014). We utilized the new CRISPR-Cas9 RNA-guided genome engineering tool to introduce this SNP into our HUES9 hESCs (Figure 2). We designed a sgRNA sequence that was complementary to the 20 base pairs spanning the desired mutation (Figure 3). This sgRNA served as a guide to direct the Cas9 nuclease to create a double-stranded break in the FADS2 promoter region. A DNA repair template was added to encourage DNA repair by HDR with maximum efficiency and minimal off-target mutations. A FACS assay confirmed the presence of our CRISPR-Cas9 construct in HUES9 hESCs and also allowed positive selection of transfected cells (Figure 4). GFP+ cells indicated proper introduction of the CRISPR-Cas9 construct because the plasmid that was nucleofected contained the CRISPR-Cas9 construct and GFP. The GFP+ cells were used to run a Surveyor Mutation Detection assay to see if the recently introduced CRISPR-Cas9 construct was functioning properly by cutting the DNA at the desired locus. In Figure 5, a 2% agarose gel was run with control HUES9 cells, control
HUES9 cells with the Surveyor nuclease added, HUES9 cells nucleofected with CRISPR-Cas9 construct before FACS, and HUES9 cells with the CRISPR-Cas9 after FACS. Comparison of the GFP+ HUES9 cells post-FACS (lane 6) with the control HUES9 cells (lane 3) clearly indicated that the initial 636 base pair PCR product was successfully cut into a 126 base pair fragment and a 510 base pair fragment. The band in lane 3 would not be expected to be cut since there was no digestion by any nucleases.

Detection of Insertions or Deletions (Indels) in Sequenced DNA of HUES9 Cells with CRISPR-Cas9 Custom Construct

After confirming the transfection of our custom-engineered CRISPR-Cas9 construct into the HUES9 hESC line, we aimed to see if the desired SNP rs968567 was now present in the cells. We estimated that 20% of the HUES9 cell line would carry a mutation (indel or specific mutation) based on the Surveyor assay, which led us to select 72 colonies to plate on three separate 24-well plates. Based on our calculations, approximately 14 colonies would express a mutation, and we could then differentiate those colonies to lung progenitor cells to study the phenotype of the induced mutation during lung development. Figure 6 shows the cell culture subcultivation and identification process we employed to propagate the HUES9 cell line. After isolating DNA from sufficiently confluent wells and sending the DNA for Sanger sequencing at GeneWiz, we began tracking which, if any, colonies expressed the desired polymorphism (Figure 7). To date, we have sequenced 39 colonies and none of the sequenced data showed SNP rs968567 (Figure 7). Twenty colonies (51.3%) showed the presence of a
heterozygous indel, 16 colonies (41%) showed the presence of a homozygous indel, and 3 colonies (7.7%) did not have any indels present. Of the 20 homozygous indels, eight colonies expressed a homozygous mutation with an adenine base pair inserted in the same locus on the DNA strand. The insertion of this adenine was thought to be a stochastic occurrence and not the result of an inter-allelic gene conversion event during the DNA repair process.
DISCUSSION

Interpretation of Sanger Sequencing Results

Disease modeling offers a powerful tool for manipulating variables of interest and studying their effects on disease phenotypes. Genome-wide association studies (GWAS) provide a way to examine a host of genetic variants to see if they have a statistically significant correlation with a disease of interest. In the case of our study, we leveraged the recent evidence of the SNP rs968567 in the FADS2 promoter region having a high degree of association with asthma as shown in an eQTL assay (Sharma et al., 2014). We were able to generate a novel, customized CRISPR-Cas9 construct that was designed to induce a C to T mutation 299 base pairs upstream of the promoter in the FADS2 gene. However, at the time of submission we were unsuccessful in detecting this mutation in any of our sequenced DNA.

Analysis of the sequencing data from each round of sequencing showed that in the instances of a homozygous mutation, 40% of the homozygous mutations occurred with the insertion of an adenine in a position upstream of the desired mutation. The insertion occurred in the same locus upstream in all of these cases. Recent literature suggests a phenomenon called “interallelic gene conversion” that may occur when a double-stranded break is induced, such as in the CRISPR-Cas9 system (Yoshimi et al., 2013). This event takes place when the DNA does not use the supplied ssODN as its repair template, but rather uses a segment of DNA from an adjacent chromosome as its source for DNA repair. This can often be confirmed by matching the sequence of the unknown insertion with the same exact sequence somewhere on a homologous chromosome.
We do not believe that inter-allelic gene conversion was responsible for the one base pair insertion. Rather, we hypothesize that DNA repair occurred by NHEJ with errors as opposed to the desired HDR method. The insertion of the adenine occurred at the cutting point for the Cas9 nuclease, which makes this region of DNA particularly susceptible to indels. The inter-allelic gene conversion phenomenon most likely explains the homozygous nature of our observed indels. Given the 40% occurrence of the adenine insertion, we sequenced the rest of the genome to determine if there was an identical sequence of DNA with the adenine insertion that could explain its repeated occurrence. If an identical sequence appeared elsewhere in the genome, then this would strengthen the argument for inter-allelic gene conversion being responsible for the adenine insertion. We found no identical sequence in the rest of the genome with the insertion of an adenine, which leads us to conclude that the insertion of an adenine in these homozygous indels is indeed a random occurrence.

To rule out the possibility that our repair template was not made correctly, we sequenced the ssODN and found that it matched exactly what we had designed. A possible way to remedy this dilemma would be to increase the concentration of ssODN that is transfected into the HUES9 cells. If the ssODN repair template is not available in a high enough quantity for the cell to utilize during its DNA repair process, then it is natural to assume that the cell will defer to repair by NHEJ or use a nearby sequence of DNA for the repair process. One may increase the amount of ssODN by either increasing the concentration of available ssODN or improving the efficiency of nucleofection. The repair template may simply not have been nucleofected into the cells along with the
plasmid containing the CRISPR-Cas9 construct. However, improving the nucleofection efficiency may prove to be difficult because of proprietary technology that prohibits the customer from knowing the exact settings for each program. It is quite possible that the Cas9 nuclease degraded the ssODN along with the targeted DNA strand. The ssODN is essentially identical to the targeted DNA strand with the exception of a C to T mutation. This close resemblance could lead to the Cas9 nuclease degrading the ssODN instead of the DNA strand, which would abort the HDR process and transition to NHEJ.

The FACS process may also be a rate-limiting-step in terms of our lack of efficiency in finding the desired mutation in this experiment. It is possible that choosing a greater proportion of GFP+ cells and seeding more colonies (perhaps an extra 24 wells) would have resulted in a positive finding in our sequencing data. However, it may be beneficial to seek improvements and adjustments elsewhere in the process instead of simply looking at more cells as the solution. The FACS efficiency could possibly be improved next time, particularly since the cell suspension was allowed to sit for 1 hour longer than expected as a result of a long wait time at the FACS Core facilities. This prolonged time in a suboptimal environment for stem cells could cause an increase in apoptosis which would lessen the amount of GFP+ cells in the post-FACS analysis.

*Directed Differentiation of hESCs to Nkx 2.1+ Lung Progenitor Cells*

This project was designed so that it has a broad range of applications once the fundamental tools have been established. In addition to her enormous amount of support and scientific guidance throughout this project, Dr. Jeanne M. Carroll is also working on
a parallel project to establish a differentiation protocol to make Nkx 2.1+ lung progenitor cells from hESCs. The establishment of a highly efficient lung progenitor differentiation protocol will enable us to take colonies of hESCs that have the desired FADS2 mutation and differentiate them into Nkx 2.1+ lung progenitor cells. This would permit us to carefully study the phenotype of stem cells with this mutation during the differentiation process and make comparisons with wild type stem cells following the same process. A future study like this would likely lead to a better understanding of the pathophysiology of asthma and perhaps a more focused scope of important candidate genes involved in various types of asthma. Moreover, our goal of establishing the FADS2 mutation in hESCs would allow us to differentiate into a variety of other lineages including liver, heart, kidney, and neuronal cells. For example, as more is learned about the impact of this mutation, it may be useful to pivot from lung progenitor cells and study the FADS2 mutation in liver cells. Using the CRISPR-Cas9 customized construct to establish our desired mutation in hESCs gives us the flexibility to conduct this type of study in the future.

A differentiation protocol is a difficult task to perfect, especially since an extremely high efficiency (typically above 90%) is needed in order to consider the protocol a success. Results from a disease modeling study in a cell population in which only 50% of the cells have differentiated into the lineage of interest, compared with a protocol that yields 90% of lineage-specific cells, will understandably be taken with a healthy degree of skepticism. Dr. Carroll’s lung differentiation project is based on optimizing the protocol published by Huang et al. (2015) in which they differentiated
embryoid bodies into Nkx 2.1+ lung progenitor cells. The process takes 25-50 days to generate mature lung and airway epithelial cells by going through a series of four steps. In step 1, the hESCs must be differentiated from their pluripotent state into definitive endoderm as confirmed by expression of the two genes known as SRY (sex-determining region Y)-box 17 (Sox17) and forkhead box A2 (FoxA2) (Figure 8). Next, in step 2, the definitive endoderm must be anteriorized to form anterior foregut endoderm (Figure 9). The timing is crucial because there is only a small window of time in which the cells are optimized for anterior foregut formation before proceeding onto posterior foregut endoderm. Once the anterior foregut endoderm cells have properly expanded, it is time for lung progenitor induction (step 3) followed by lung/airway epithelial maturation (step 4). This incredibly precise four-step process is mediated by manipulating media components on certain days of the differentiation process in order to turn on and off specific genes that are instrumental in forming either definitive endoderm, anterior foregut endoderm, lung progenitor cells, or mature lung/airway epithelial cells.

Overall, this study resulted in the successful engineering of a novel CRISPR-Cas9 construct designed to induce the SNP rs968567 into hESCs in an effort to model asthma in lung progenitor cells. The CRISPR-Cas9 construct will also allow us and others to study the FADS2 mutation in a variety of different cell lineages. Future efforts will revolve around improving the nucleofection and HDR efficiency in order to generate a higher yield of stem cells with our desired mutation present.
Table 1. Surveyor Mutation Detection Kit Hybridization Schedule

<table>
<thead>
<tr>
<th>Temperature</th>
<th>Time</th>
<th>Temperature Ramp</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 °C</td>
<td>10 min</td>
<td></td>
</tr>
<tr>
<td>95 °C → 85 °C</td>
<td></td>
<td>-2.0 °C/s</td>
</tr>
<tr>
<td>85 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>85 °C → 75 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>75 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>75 °C → 65 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>65 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>65 °C → 55 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>55 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>55 °C → 45 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>45 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>45 °C → 35 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>35 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>35 °C → 25 °C</td>
<td></td>
<td>-0.3 °C/s</td>
</tr>
<tr>
<td>25 °C</td>
<td>1 min</td>
<td></td>
</tr>
<tr>
<td>4 °C</td>
<td></td>
<td>HOLD</td>
</tr>
</tbody>
</table>
Figure 1. Metabolic pathway of n-3 and n-6 PUFAs. Flowchart shows the precursors, intermediates, and end products in PUFA metabolism. The enzyme D6D, encoded by FADS2, converts LA to γ-LA and has been shown to be associated with asthma. An increase in D6D activity leads to an increase in AA and subsequent increase in pro-inflammatory eicosanoids. (AA = arachidonic acid; D6D = δ-6 desaturase; FADS2 = fatty acid desaturase 2; LA = linoleic acid; PUFAs = polyunsaturated fatty acids.)
Figure 2. CRISPR-Cas9 mechanism of action. Diagram shows the theory behind how the CRISPR-Cas9 system works to induce DNA double-stranded breaks. The sgRNA/Cas9 complex binds by Watson-Crick base pairing at the desired location near the PAM sequence. The Cas9 nuclease recognizes the PAM sequence and binds to the DNA where it induces a double-stranded break at the desired base pair. DNA is then repaired by HDR with a supplied repair template so that the DNA strand re-forms with the new insertion or deletion. Adapted from Ran et al. (2013). (Cas9 = CRISPR-associated protein 9; CRISPR = clustered regularly interspaced short palindromic repeat; DNA = deoxyribonucleic acid; HDR = homology-directed repair; PAM = protospacer adjacent motif; sgRNA = small guide RNA.)
Sense Strand 5’  GTGGAAACTCGGGCGGCGGGGAACGCGGGAGGATGTGG
Antisense Strand 3’  CACCTTTTGAGCCCGCCGCCCCCCTTTGCCCTACTCC
Sense Strand 5’  AACCCGAGGCGGGGGAGCCGGAGGGGCGGCAGAGGA
Antisense Strand 3’  TTGGGCTCCGCCCCCCCTCGGCCTCCCCGCGGCTCTCCT
Sense Strand 5’  GGTGTGAGGCCTCGAGCTCCCGGGAGAGTTTTTACTGGA
Antisense Strand 3’  CCACAGCTCCGGGACTCGAGGGCCCCTCAAAAATGACCT
Sense Strand 5’  GGCAAAAAGTCCATAGGGAGGGCAGGGGATGGAGGGCGGA
Antisense Strand 3’  CCGTTTTACAGTATCGGCGCCCTCCCGACTCCCTCCCCGCCT
Sense Strand 5’  GGAAGGGGACCGCTTGGGGAGGGCACTGGGAAGGCGGA
Antisense Strand 3’  CTTCCCCCTGCGGCCAGACCCCGTGACCCTTCGGTCCCTA
Sense Strand 5’  CCTCCGCC
Antisense Strand 3’  GGAGGCGG

Red = PAM sequence
Blue = sgRNA
Green = Target allele for mutation

**Figure 3. CRISPR-Cas9 customized construct and FADS2 gene sequence.** Sequence shows a portion of the FADS2 gene in the promoter region that contains the PAM sequence, where our sgRNA will bind, and the specific base pair in which we are aiming to induce a C to T transition mutation. (Cas9 = CRISPR-associated protein 9; CRISPR = clustered regularly interspaced short palindromic repeat; FADS2 = fatty acid desaturase 2; PAM = protospacer adjacent motif; sgRNA = small guide RNA.)
Figure 4. FACS results for HUES9 cells post-nucleofection. Plot in bottom left (PE-A vs. FITC-A) shows the cohort of GFP+ cells chosen for cell culture subcultivation and DNA sequencing to confirm if desired SNP was properly induced. Cells that strongly express GFP in the FACS assay provide confirmation that nucleofection introduced the customized CRISPR-Cas9 construct into the HUES9 cell line. (Cas9 = CRISPR-associated protein 9; CRISPR = clustered regularly interspaced short palindromic repeat; DNA = deoxyribonucleic acid; FACS = fluorescence-activated cell sorting; FITC-A = fluorescein isothiocyanate - area; FSC-A = forward-scattered light - area; FSC-H = forward-scattered light - height; GFP = green fluorescent protein; P1-P5 = events 1-5; PE-A = phycoerythrin - area; SNP = single nucleotide polymorphism; SSC-A = side-scattered light - area.)
Figure 5. Surveyor mutation detection gel. A 2% agarose gel was run to determine if the custom-engineered CRISPR-Cas9 construct was properly introduced into HUES9 cells by nucleofection. Well 1: DNA ladder; Well 2: No template control (no DNA should be present); Well 3: HUES9 cells and no digestion with supplied Surveyor nuclease; Well 4: HUES9 cells with no CRISPR-Cas9 construct and digestion with supplied Surveyor nuclease; Well 5: HUES9 cells before FACS with CRISPR-Cas9 construct added by nucleofection; Well 6: HUES9 cells that were GFP+ after FACS with CRISPR-Cas9 construct added; Well 7: Same DNA ladder as Well 1. (Cas9 = CRISPR-associated protein 9; CRISPR = clustered regularly interspaced short palindromic repeat; DNA = deoxyribonucleic acid; FACS = fluorescence-activated cell sorting.)
Figure 6. Cell culture flow chart. Flow chart depicts the tracking scheme of each selected HUES9 colony through subcultivation. Cells were harvested for DNA isolation and subsequent Sanger sequencing once each well reached about 75% confluency. (DNA = deoxyribonucleic acid; FACS = fluorescence-activated cell sorting; GFP = green fluorescent protein.)
Figure 7. Sanger sequencing results for HUES9 cell line. DNA was isolated from GFP+ HUES9 cells that were confirmed to have the addition of the custom CRISPR-Cas9 construct aimed at inducing the FADS2 mutation. Sequencing results show that 51.3\% of sequenced colonies had a heterozygous indel, 41.0\% had a homozygous indel, and 7.7\% had no indel. None of the sequencing results returned data confirming the presence of the intended C \rightarrow T mutation in the FADS2 promoter region. N = 39. (Cas9 = CRISPR-associated protein 9; CRISPR = clustered regularly interspaced short palindromic repeat; DNA = deoxyribonucleic acid; FADS2 = fatty acid desaturase 2; GFP = green fluorescent protein; indel = insertion or deletion.)
Figure 8. Immunofluorescence imaging of definitive endoderm. Photomicrograph shows the expression levels of Sox17, Oct4, and Dapi (measure of DNA in cells). Oct4 is a pluripotency marker which should be lowly expressed at the definitive endoderm stage. Sox17 is upregulated in definitive endoderm. A combination of low Oct4 and high Sox17, as shown, indicates good definitive endoderm efficiency. Image captured at 10x magnification. (Dapi = 4',6-diamidino-2-phenyindole; DNA = deoxyribonucleic acid; Oct4 = octamer-binding transcription factor 4; Sox17 = SRY (sex determining region Y)-box 17.)
Figure 9. Immunofluorescence imaging of anterior foregut endoderm. Photomicrograph shows the expression levels of Sox2, FoxA2, and Dapi (measure of DNA in cells). Cells that are double-positive for Sox2 and FoxA2 indicate good anterior foregut endoderm efficiency as seen in the upper right quadrant. Image captured at 20x magnification. (Dapi = 4′,6-diamidino-2-phenylindole; DNA = deoxyribonucleic acid; FoxA2 = forkhead box A2; Sox2 = SRY (sex determining region Y)-box 2.)
LIST OF JOURNAL ABBREVIATIONS

Allergy Asthma Immunol. Res. ....................... Allergy, Asthma & Immunology Research
Allergy Asthma Proc. ........................................... Allergy & Asthma Proceedings
Am J Clin Nutr. ........................................... American Journal of Clinical Nutrition
Am. Rev. Respir. Dis. ................................... American Review of Respiratory Disease
Ann. Allergy Asthma Immunol ...................... Annals of Allergy Asthma and Immunology
Arch. Dis. Child ................................................ Archives of Disease in Childhood
Clin. Exp. Allergy ........................................... Clinical & Experimental Allergy
Curr. Opin. Lipidol. ....................................... Current Opinion in Lipidology
Immun. Inflamm. Dis. ..................................... Immunity, Inflammation, and Disease
J. Allergy Clin. Immunol ................. Journal of Allergy and Clinical Immunology
J. Clin. Invest .................................................. Journal of Clinical Investigation
J. Lipid. Res .................................................... Journal of Lipid Research
Nat. Biotechnol ...................................................... Nature Biotechnology
Nat. Protoc .............................................................. Nature Protocols
NCHS Data Brief ........................................... National Center for Health Statistics Data Brief
Nucleic Acids Res............................................................... Nucleic Acids Research
Nurs. Stand. R. Coll. Nurs.................................Nursing Standard Royal College of Nursing
Prev. Chronic. Dis......................................................Preventing Chronic Disease
REFERENCES


CURRICULUM VITAE

Sean J. McManus
smcman@bu.edu (508) 982-6009
128 Saint Botolph Street, Apt 1F
Boston, MA 02115
DOB: 1989

EDUCATION
M.S. Medical Sciences, September 2013 - May 2015
BOSTON UNIVERSITY SCHOOL OF MEDICINE, Boston, MA
B.S. Biology Minor: Bioengineering, August 2008 - May 2012
VILLANOVA UNIVERSITY, Villanova, PA

RESEARCH & PROFESSIONAL DEVELOPMENT
May 2014 - Research Assistant II, Harvard Medical School/Brigham and Women’s Hospital, Boston, MA
• Investigate the regulatory variants at established lung disease-associated loci using RNA guided genome engineering and human embryonic stem cells
• Utilize custom multi-scale imaging and informatics pipeline to analyze transcriptional dynamics in space and time during stem cell differentiation
• Serve as safety officer to lab group

May 2012 - August 2013 Oncology Clinical Research, Millennium Pharmaceuticals, Cambridge, MA
• Developed comprehensive patient profiles to better understand efficacy and safety of pipeline drugs
• Evaluated safety risks of drugs in clinical trials by examining patient data from laboratory tests
• Conducted periodic safety review by assessing, interpreting, and documenting clinical safety data
• Subject Matter Expert (SME) for TIBCO Spotfire software. Trained colleagues in department to use Spotfire as a visualization tool to aid in clinical research efforts
• Constructed company-wide database of patient baseline parameters to inform future studies about enrollment metrics by site, region, disease indication, and eligibility criteria

May 2011 - May 2012 Research Assistant, Biothermal Sciences Laboratory, Villanova, PA

43
• Studied damage to human cell lines during cryopreservation process
• Investigated effect of cooling rates on intracellular ice formation using a combination of microscopy and high-speed video cryomicroscopy techniques
• Maintained mammalian cell cultures for use in cryomicroscopy experiments
• Prepared Senior Thesis and manuscript for journal publication

June 2009 - August 2009

Physical Therapy Aide, Northeast Rehabilitation Center, North Andover, MA
• Guided patients through rehabilitation exercises under supervision of physical therapist

HONORS, AWARDS, AND MEMBERSHIPS

Fall 2012
Associate Member of Sigma Xi, Delaware Valley, PA

May 2012
1st Place International Society of Pharmaceutical Engineers Poster Competition, Delaware Valley, PA

May 2012
Senior Thesis Medallion, Villanova University, Villanova, PA

April 2012
1st Place Sigma Xi Poster Competition, Villanova University, Villanova, PA

March 2012
Member of International Society of Pharmaceutical Engineers, Boston Chapter, Boston, MA

VOLUNTEERING

October 2014 -
Science Club Volunteers, Brigham and Women’s Hospital, Boston, MA

January 2014 - May 2014
Emergency Room Volunteer, East Boston Neighborhood Health Center, Boston, MA

November 2012 - May 2013
Emergency Room Volunteer, Winchester Hospital, Winchester, MA

May 2012 - August 2013
Millennium Makes a Difference, Millennium Pharmaceuticals, Cambridge, MA
October 2012      Light the Night Walk, Leukemia and Lymphoma Society, Boston, MA

September 2012 -  Elementary School Tutor, Tutoring Plus, Cambridge, MA

April 2012        Villanova Candidates’ Day, Villanova University, Villanova, PA

February 2012 - March 2012 Villanova Club Basketball Fundraiser, Villanova University, Villanova, PA

December 2010 - 2012 Holiday Gift Drive, North Andover Youth Services, North Andover, MA

October 2008, ’10, ’11 Special Olympics Fall Festival, Villanova University, Villanova, PA

March 2011        Habitat for Humanity, Villanova University, Birmingham, AL

January ’09, ’11, ’12 Martin Luther King Day of Service, Villanova University, Bryn Mawr, PA