

CHRONIC OTITIS MEDIA: MICROBIAL ENVIRONMENT DURING VIRAL  
INFECTION AND GENETICS OF HOST SUSCEPTIBILITY

Emma Kaitlynn Allen

Charleston, SC

Bachelor of Arts in Biological Sciences, Clemson University, 2009

A Dissertation presented to the Graduate Faculty  
of the University of Virginia in Candidacy for the Degree of  
Doctor of Philosophy

Department of Biochemistry Molecular Genetics

University of Virginia

April 2014

Dr. Stefan Bekiranov, PhD, Advisor

---

Dr. Keith L. Keene, PhD

---

Dr. Ani Manichaikul, PhD

---

Dr. Jason Papin, PhD, Dean's Representative

---

Dr. Michèle M. Sale, PhD, Advisor

---

Dr. David Wotton, PhD, Chair

---

## Abstract

Otitis media (OM), inflammation of the middle ear, is an upper respiratory tract infection (URTI) caused by viral and bacterial infections. OM is the most common pediatric disease and the leading cause of pediatric health care visits, antibiotic usage and surgery. A subset of OM affected children develops chronic otitis media with effusion (COME) and/or recurrent otitis media (ROM), with significant consequences on hearing and development. The aim of this research was to find novel risk factors that increase OM susceptibility by looking at both bacterial communities during URTI and host genetics.

To explore how viral infection affects bacterial composition of the nasopharynx in the setting of URTI, two complementary approaches were used in a longitudinal examination following experimental viral challenge. Ten healthy adults were infected with a rhinovirus, and nasal lavage fluid (NLF) samples were collected before, during, and after inoculation. Microarray technology tested samples for a panel of species, and sequencing of the 16S rRNA bacterial gene was used to survey the relative abundance of bacteria down to the genus level. Pathogenic bacteria were detected in NLF samples from healthy adults, including species known to cause OM (*Streptococcus pneumoniae*, *Moraxella catarrhalis*, and *Haemophilus influenzae*). Sequencing results demonstrated that relative abundance of two genera *Neisseria* ( $P=0.02$ ) and *Propionibacterium* ( $P=0.04$ ) were significantly different between infected and not infected individuals. No significant shifts in the microbiome were detected between time points before, during, or after inoculation. This understanding of the complexity of the nasopharyngeal microbiome is important for future study design and appropriate treatment of URTIs.

To discover novel OM host risk factors, a multi-stage approach was used. First, a genome-wide association study (GWAS) of COME/ROM was conducted using a family-

based population. The most significantly associated SNP, rs1110060, is located on chromosome 15 in an intron of kinesin family member 7 (*KIF7*). Association at SNP rs10497394, an intergenic SNP on chromosome 2, was replicated in an independent family-based population of OM. Fine-mapping of associated regions of chromosomes 2 and 15 was carried out using targeted resequencing and subsequent genotyping. Six SNPs were significantly associated ( $P < 2.6 \times 10^{-4}$ ), two in the chromosome 2 region and four in the chromosome 15 region. Explorations of the functional potential of these two loci were carried out. Investigation of chromatin marks revealed an enhancer region on chromosome 2 in the human middle ear epithelial cell line, HMEEC-1. The “G” risk allele of the most significantly-associated chromosome 15 SNP, rs1110060, was found to impact expression of *IQGAP1* in adenoids ( $P = 0.026$ ). Also, both lipopolysaccharide (LPS) and TGF $\beta$  treatment were found to regulate expression of *KIF7* in HMEEC-1.

The data presented in this dissertation show the complex nature of the microbiome of the nasopharynx during URTI, and we have identified novel regions and genes involved in susceptibility to OM. Moreover, functional assays support the participation of previously unsuspected mechanisms in OM pathogenesis. In summary, this research has provided insights into OM susceptibility from both host and bacterial perspectives.

## **Acknowledgements**

Many people contributed both directly and indirectly to my success at the University of Virginia. I would like to thank those individuals who impacted me the most during this time.

First and foremost, I would like to thank my PhD advisor and mentor, Dr. Michèle Sale. Infinite thanks for her constant patience, enthusiasm for my project and passions, and her wisdom regarding both science and life that she so generously shared with me. I started on this journey unsure of a specific path, but her guidance helped me sharpen my focus. The more time I spent discussing my progressing research with her, and the process of working to solve related health issues, the more she helped me realize that a PhD was an opportunity of a lifetime. This work enables me to contribute my findings to the ongoing study of otitis media and future treatments. Not only is Michèle pivotal in my choice to become a scientist, but she has also become a close, personal friend. No matter the circumstances, she is supportive of me, my project, and my future. She is my mentor, my friend, and an inspiration. I cannot adequately express how profoundly she has impacted my life. I look ahead and see a long lasting friendship and even an amazing collaboration with Michèle in the future.

I would also like to acknowledge the members of the Sale lab who contributed and extended their valuable assistance in this project. Thank you to Paul Campolieto for his continued help and support, and also for always having an open ear when I need to vent. Thanks also to Dr. Stephen Williams for his valuable advice, patience during crises, and teaching me new skills. To Dr. Keith Keene, countless thanks for his invaluable advice and unending support. I am extremely lucky that I have had these lab mates and great friends throughout my time at UVA.

I also thank my committee members Drs. David Wotton, Stefan Bekiranov, and Jason Papin. They helped to guide me and the direction of my studies to complete this thesis.

Many thanks to Dr. Birgit Winther who taught me so much about the clinical aspects of research. Dr. Winther involved me in many studies and collaborations that have advanced my skills, broadened my networking, and fine-tuned my career goals. I cannot thank her enough for her support and the opportunities she provided for me.

To the many collaborators at UVA, especially those in the Center for Public Health Genomics, thank you for the various contributions you have made to my scientific knowledge and to my project. Thank you Drs. Josyf Mychaleckyj and Ani Manichaikul for being wonderful mentors who helped to guide me and my project. Thank you to Dr. Wei-Min Chen for his guidance on statistical analyses for my project. Thank you Cheryl Bryan for teaching me the in's and out's of grant submission and also for her friendship and support during my time at UVA. I would also like to thank all of the faculty in the CPHG who taught me so much about how to approach a scientific question, especially during my presentations at the CPHG Research in Progress.

Thank you to Gordon Laurie and the Biotechnology Training program members. I have learned so much about multiple scientific disciplines. I also can't thank the BTP training grant enough for giving me the opportunity to work at Mobidiag as my externship. I would also like to extend a sincere thank you to the amazing scientists at Mobidiag, especially Drs. Minna Mäki and Sanna Laakso, and Dr. Anne Pitkäranta at the University of Helsinki for their guidance and everything they taught me during my externship.

Finally, I owe so many thanks to my family and friends. I owe an extreme amount of gratitude to my friends from home and the friends I made during my time at UVA. To my big brother Stu, for always letting me know he is proud of me and for making me

laugh when I needed it. To my phenomenal parents Barbara and Carl, who always encouraged me to stay motivated and reach for my goals, and on top of that, who endlessly supported me in my pursuits. They never let me lose sight of my purpose and always listened when I needed someone. They always make me feel like I can achieve anything. Thank you to my fiancé John for always calming me down and having my back. He is always being understanding during crazy times and letting me know that I will achieve anything I put my mind to. Finally, thank you to my sweet Bailie and Remley for their constant love.

## Table of Contents

<b>Abstract .....</b>	<b>i</b>
<b>Acknowledgements .....</b>	<b>iii</b>
<b>Table of Contents .....</b>	<b>1</b>
<b>List of Abbreviations.....</b>	<b>5</b>
<b>List of Figures .....</b>	<b>8</b>
<b>List of Tables.....</b>	<b>10</b>
<b>Chapter 1: Introduction.....</b>	<b>11</b>
1.1 Upper respiratory tract infections .....	11
1.2 Introduction to otitis media .....	11
1.3 Treatment options for children with otitis media .....	16
1.4 Heritability of otitis media .....	17
1.5 Mouse models of otitis media.....	17
1.6 Candidate gene studies of otitis media.....	19
1.7 Human Middle Ear Epithelial Cell line .....	20
1.8 Agnostic approaches to discover novel otitis media risk variants .....	21
1.9 Resources needed to discover novel risk variants of otitis media.....	28
1.10 Conclusions and goals .....	29

## **Chapter 2:** Survey of the nasopharyngeal microbiome in healthy adults

2.1 Chapter Introduction .....	32
2.2 Bacteria in the nose of young adults during wellness and rhinovirus colds - detection by culture and microarray methods in 100 nasal lavage specimens.....	33
2.2.1 Abstract .....	34
2.2.2 Introduction.....	35
2.2.3 Methods .....	36
2.2.4 Results.....	41
2.2.5 Discussion .....	53

## **Chapter 3:** Nasopharyngeal microbiome during both health and viral infection

3.1 Chapter Introduction .....	56
3.2 Characterization of the nasopharyngeal microbiome in health and during rhinovirus challenge .....	57
3.2.1 Abstract .....	57
3.2.2 Introduction.....	59
3.2.3 Methods .....	61
3.2.4 Results.....	67
3.2.5 Discussion .....	83



**Chapter 4:** Genome-wide association study of chronic otitis media with effusion and recurrent otitis media

4.1 Chapter Introduction .....	87
4.2 A genome-wide association study of chronic otitis media with effusion and recurrent otitis media identifies a novel susceptibility locus on chromosome 2.....	88
4.2.1 Abstract .....	89
4.2.2 Introduction.....	90
4.2.3 Methods .....	92
4.2.4 Results.....	98
4.2.5 Discussion .....	107

**Chapter 5:** Raine cohort GWAS variant replication study

5.1 Chapter Introduction.....	114
5.2 Evaluation of replication of variants associated with genetic risk of otitis media .....	115
5.2.1 Abstract .....	115
5.2.2 Introduction.....	116
5.2.3 Methods .....	119
5.2.4 Results and Discussion .....	120

**Chapter 6:** GWAS of COME/ROM follow up study to identify causal variants

6.1 Chapter Introduction .....	125
6.2 Honing in on genetic risk of chronic otitis media with effusion and/or recurrent otitis media (COME/ROM): a GWAS follow-up study.....	126
6.2.1 Abstract .....	127
6.2.2 Introduction.....	128
6.2.3 Methods .....	130
6.2.4 Results.....	141
6.2.5 Discussion .....	150
<b>Chapter 7:</b> Discussion.....	157
<b>References .....</b>	<b>169</b>
<b>Supplement .....</b>	<b>199</b>

## List of Abbreviations

<b>AOM</b>	acute otitis media
<b>ASO</b>	allele specific primer
<b>ASP</b>	affected sibling pairs
<b>CEU</b>	HapMap abbreviation for CEPH (Utah residents with ancestry from northern and western Europe)
<b>CNS</b>	Coagulase-negative <i>Staphylococcus</i>
<b>COM</b>	chronic otitis media
<b>COME</b>	chronic otitis media with effusion
<b>ENT</b>	ear, nose, and throat
<b>ENU</b>	<i>N</i> -ethyl- <i>N</i> -nitrosourea mutagenesis screening method
<b>eQTL</b>	expression quantitative trait loci
<b>ET</b>	Eustachian tube
<b>GWAS</b>	genome wide association study
<b>HMEEC-1</b>	Human middle ear epithelial cell line
<b>HRV</b>	Human Rhinovirus
<b>LD</b>	linkage disequilibrium
<b>LOD</b>	Logarithm of odds
<b>LSO</b>	locus specific primer

<b>MAF</b>	minor allele frequency
<b>ME</b>	middle ear
<b>MEE</b>	middle ear effusion
<b>NLF</b>	nasal lavage fluid
<b>NPL</b>	non-parametric linkage
<b>OM</b>	otitis media
<b>OME</b>	otitis media with effusion
<b>OR</b>	odds ratio
<b>OSA</b>	obstructive sleep apnea
<b>OTIGEN</b>	International Consortium of the genetics of otitis media
<b>PCA</b>	principal component analysis
<b>PCoA</b>	principal coordinates analysis
<b>PCR</b>	polymerase chain reaction
<b>Raine</b>	Western Australian Pregnancy cohort
<b>ROM</b>	recurrent otitis media
<b>RSV</b>	respiratory syncytial virus
<b>r<sup>2</sup></b>	measure of LD and its relationship to allele frequencies of the loci
<b>SNP</b>	single nucleotide polymorphism
<b>TCID<sub>50</sub></b>	50% Tissue Culture Infective Dose

<b>TM</b>	tympanic membrane
<b>UMN</b>	University of Minnesota
<b>UPitt</b>	University of Pittsburgh
<b>URT</b>	upper respiratory tract
<b>WAFSOM</b>	Western Australia family study of otitis media

<b>List of Figures</b>	<b>Page</b>
<b>Figure 1.1</b> Diagram of the ear .....	13
<b>Figure 1.2</b> Diagram linking chapters of this thesis in terms of OM pathogenesis .....	31
<b>Figure 2.1:</b> Study timeline including when information or samples were collected from participants. ....	38
<b>Figure 2.2:</b> Bacterial titer by semi-quantitative cultures in nasal wash samples of study participants. ....	45
<b>Figure 2.3:</b> Prevalence of bacteria identified by microarray in 100 samples from 10 volunteers inoculated with HRV39.....	49
<b>Figure 3.1:</b> Study timeline including when information or samples were collected from study participants. ....	63
<b>Figure 3.2:</b> Box and whisker plot of read counts at each time point during the study after quality control filtering.....	70
<b>Figure 3.3:</b> Taxonomic bar charts showing the relative abundances of phyla (A) in each sample, and (B) at each time point; (C) Box and whisker plot of dominant phyla in infected and not infected individuals by time period. ....	72
<b>Figure 3.4:</b> Rarefaction plots.....	75
<b>Figure 3.5:</b> Principal coordinates analysis (PCoA) of pre-inoculation NLF samples and samples from other body sites. ....	77
<b>Figure 3.6:</b> Unifrac distance comparisons between (A) different individuals (inter-individual) and the same individual (intra-individual); (B) infected and not infected subjects. ....	78

<b>Figure 3.7:</b> Comparisons of not infected vs. infected subjects: (A) stacked taxonomic bar chart of genera; (B): Box and whisker plot of Shannon diversity index. ....	80
<b>Figure 4.1</b> COME/ROM GWAS results using the QLS <sub>w</sub> method: (A) Quantile-quantile plot; (B) Manhattan plot. ....	99
<b>Figure 4.2</b> Genomic characteristics of the chromosome 2 region of high LD. ....	104
<b>Figure 4.3</b> Locus Zoom plots of the chromosome 2 region using UMN GWAS data and HapMap3 imputation. ....	106
<b>Figure 6.1:</b> Haplotype blocks of (A) the chromosome 2 region, and (B) the chromosome 15 region from the fine-mapping project. ....	145
<b>Figure 6.2:</b> ChIP-seq of the chromosome 2 region for enhancer activity (H3K27Ac, H3K4me1, and H3K4me3) in HMEEC-1 compared with ENCODE ChIP-seq data. ....	147
<b>Figure 6.3:</b> Relative <i>IQGAP1</i> expression levels in adenoids by rs1110060 risk allele status. ....	149
<b>Figure 6.4:</b> Relative <i>KIF7</i> expression levels in HMEEC-1 cells after treatment with TGF $\beta$ . ....	151
<b>Figure 6.5:</b> Relative <i>KIF7</i> expression levels in HMEEC-1 cells after treatment with <i>P. aeruginosa</i> LPS. ....	152

<b>List of Tables</b>	<b>Page</b>
<b>Table 2.1:</b> List of Species covered on the Mobidiag Prove-it™ Bone and Joint Assay...	42
<b>Table 2.2:</b> Infection and Illness in Volunteers following self-inoculation of HRV 39 ...	44
<b>Table 2.3:</b> Bacteria detected in all NLF samples from Volunteers.....	50
<b>Table 3.1:</b> Subject demographic information .....	68
<b>Table 3.2:</b> Comparisons of genus abundance between infected and uninfected samples .....	82
<b>Table 4.1:</b> Participant characteristics for the University of Minnesota (UMN) and University of Pittsburgh (UPitt) <i>family-based</i> study populations .....	93
<b>Table 4.2:</b> Replication results with $P < 0.10$ and Meta-analysis results of those SNPs. ....	102
<b>Table 4.3:</b> Significant results from the three eQTL browsers used for the in silico eQTL analysis for the four SNPs of interest. ....	108
<b>Table 5.1:</b> Top SNPs from the Raine cohort GWAS that were genotyped in our GWAS of COME/ROM .....	121
<b>Table 5.2:</b> Top SNPs from Raine cohort GWAS that were genotyped in our family population of COME/ROM.....	123
<b>Table 6.1:</b> Categories of SNPs used for fine-mapping of GWAS association signals ....	142
<b>Table 6.2:</b> Most significant associations from genotyping project using the QLSw association test.....	143



## Chapter 1

### Introduction

#### 1.1 Upper respiratory tract infections

The symptom complex of viral upper respiratory tract infections (URTIs) and bacterial complications (bronchitis, otitis media (OM) and sinusitis) are self-limiting, but they cause billions of dollars in healthcare costs and millions of days of school and work missed in the United States every year.<sup>1</sup> Rhinovirus is the major cause of URTIs and young preschool age children contract as many as six rhinovirus infections per year.<sup>2,3</sup> In the U.S., acute OM, a very common pediatric URTI, is the leading reason for prescription of antibiotics.<sup>4,5</sup> However, it remains unclear whether commensal bacteria may become pathogenic during the viral infection. Many pathogenic bacteria, including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis*, reside in the nasopharynx in 60% -70% of healthy adults and children<sup>6-10</sup>, but these bacteria may exacerbate symptoms during viral infections.<sup>11</sup>

#### 1.2 Introduction to otitis media

##### 1.2.1 Pathogenesis of otitis media

The cause of OM has been shown to be viral URTIs which elicit inflammation of the URT including the Eustachian tube (ET). Mucus from the middle ear (ME) is continually removed by cilia through the ETs to be swallowed, and without its removal due to ET blockage/dysfunction or excess mucins, bacteria may ascend the ETs and initiate inflammation.<sup>12</sup> The role of the ET is to provide air flow, regulate ME pressure, and allow for clearance of any buildup of effusion, or fluid, in the ME. When the ET is

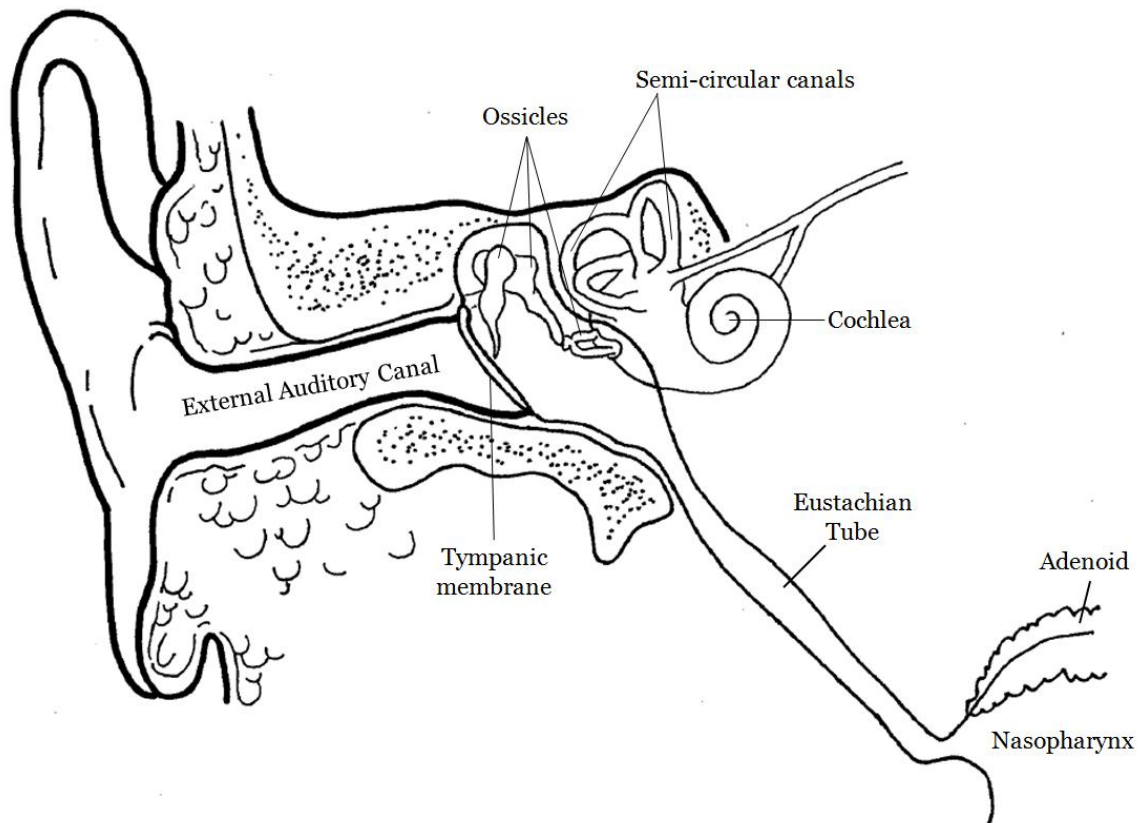
inflamed, airflow is blocked, negative pressure builds in the ME, and clearance of buildup in the ME is inhibited. Blocked airflow to the ME leads to hypoxia which has been shown to be a contributor to risk of OM.<sup>13</sup> Also, negative pressure in the ME causes tympanic membrane retraction.<sup>14,15</sup>

### 1.2.2 The adenoid as a pathogenic reservoir

The adenoid is a mass comprised of lymphatic tissue that resides in the roof of the nasopharynx where the nasal cavity bends to meet the throat. It is also situated between the openings of the two ET which lead to the ME cavities (**Figure 1.1**). In children, hypertrophic, or enlarged, adenoids are caused by bacterial, viral, or polymicrobial infections. Hypertrophic adenoids can block nasal passages, causing issues with breathing, and can block the openings of the ETs.<sup>16,17</sup> Several studies have examined the microbial composition of the surface of adenoids removed from children using culture-based methods, and have shown that adenoids harbor bacteria.<sup>18-20</sup> An adenoidectomy therefore not only removes the enlarged adenoids which obstruct ET openings, but it also removes a reservoir of bacteria that is positioned to allow for bacteria to ascend the ETs. It has been shown that adenoidectomy is an effective method to reduce recurrence of OM in children, regardless of the size of the adenoid, indicating the adenoid's role in OM pathogenesis as a source of pathogenic bacteria rather than a physical obstruction of the ETs in most cases.<sup>16,17,21</sup>

Two studies have investigated the microbial communities of the adenoid surface in children using culture-independent methods.<sup>16,22</sup> Nistico, *et al.* used the Ibis T5000 Biosensor which used broad-range PCR and mass spectrometry to determine the presence of 16 bacterial species on adenoids from children with chronic otitis media

**Figure 1.1** Diagram of the ear



Adjusted from <http://www.patient.co.uk/health/glue-ear>.

(COM) and with obstructive sleep apnea (OSA).<sup>16</sup> This study found that bacteria on adenoids from both COM and OSA patients were polymicrobial and included the known pathogens *Haemophilus influenza* (significantly more in the COM group,  $P=0.005$ ), *Streptococcus pneumoniae*, and *Staphylococcus aureus*. Additionally, results confirmed the presence of bacterial biofilms on the surface of adenoids. In our group's study of the adenoid microbiome, we used sequencing of the bacterial 16S rRNA gene to comprehensively survey the microbiome of the adenoid from children undergoing surgery for COM, OSA, and hypertrophic adenoids.<sup>22</sup> The 16S rRNA gene is characterized by highly conserved regions flanking hypervariable sequences allowing primers to be designed in conserved regions for sequencing across the hypervariable regions to determine what bacteria are present. We found the adenoid microbiome was diverse, but distinct from other body sites. Multiple pathogens were found on the adenoids, including those known to cause OM. Together, these studies demonstrate that the adenoid is a reservoir of pathogenic bacteria positioned at the base of the ETs.

### 1.2.3 OM classification and health burden

Each year in the United States, over \$5 billion is spent on healthcare for OM in children.<sup>1,23</sup> There is a spectrum of severities of OM, classified on the basis of stage of infection, the host's ability to clear infection, and the damage to the ME due to infection. Acute OM (AOM) is a single infection which causes an inflamed ME and is usually accompanied by symptoms like fever, irritability, and ME pain. Almost all children are estimated to have at least one episode of acute otitis media (AOM) by age 3.<sup>24</sup> OM with effusion (OME), or glue ear, is defined by persistent middle ear effusion (MEE) with no other symptoms. OME is known to be subsequent to an AOM episode, or it could be due to ET dysfunction after an URTI.<sup>25</sup> Some children do have the ability to clear AOM and

OME, but there is a subset of children highly susceptible to subsequent infections. Children who develop 3 or more infections in a period of 6 months, or 4 or more infections in a year, are considered to have recurrent OM (ROM). Chronic OME (COME) is characterized by the presence of MEE for at least 3 months.<sup>14</sup> COME and ROM lead to repeated cycles of antibiotic prescriptions and multiple doctors' visits.<sup>4,26</sup> High numbers of pediatrician visits each year dramatically increase the number of missed work days of parents.<sup>27,28</sup> The research in this dissertation focused on the severe end of the OM spectrum, *i.e.* COME and ROM, together referred to as COME/ROM.

#### 1.2.4 COME/ROM as a multifactorial disease

COME/ROM is a multifactorial disease arising from the interactions between viral upper respiratory infection, bacterial colonization, host genetic factors, and environmental risk factors. Like AOM, COME/ROM is polymicrobial, with three main otopathogens responsible for infection: *Moraxella catarrhalis*, non-typeable *Haemophilus influenzae*, and *Streptococcus pneumoniae*. It is known that these three otopathogens can form biofilms in the ME, which has broad implications for appropriate antibiotic selection to treat OM with antibiotics and may explain some of treatment failures in OM patients.<sup>29,30</sup> The role of viruses on predisposition to bacterial OM has been shown to be due to multiple mechanisms including effects on bacterial adherence and colonization<sup>31-33</sup>, ET function<sup>17,34</sup>, antibiotic efficacy<sup>35</sup>, and mucociliary transport<sup>36,37</sup>. Viruses most commonly detected in nasopharyngeal secretions of children with AOM include adenoviruses, influenza, respiratory syncytial virus (RSV), enterovirus, coronavirus, rhinovirus, and parainfluenza virus.<sup>38,39</sup>

Host factors found to contribute to COME/ROM include ET obstruction, naïve immune systems, gender, age, race, and host genetics. There is a peak incidence of AOM between the ages of 6 and 18 months in the United States.<sup>23</sup> Environmental risk factors have been extensively reviewed<sup>40-43</sup> and include day care attendance, seasonality, allergies, and shorter duration or lack of breastfeeding. ROM can be caused by an inability to generate an immune memory response, and therefore the child remains susceptible to re-infection.<sup>44-47</sup> Studies have not been able to resolve whether there is a sex bias in the heritability of OM<sup>48</sup>, because both male gender<sup>42,49-51</sup> and female gender<sup>52</sup> have been documented as risk factors for AOM, ME effusion, and ROM. It has, however, been shown that there is a higher prevalence of surgery for OM in males than females.<sup>53,54</sup>

### 1.3 Treatment options for children with otitis media

Despite clinical advancements, there are still few options to treat children with ROM and/or COME. OM is the leading cause for antibiotic treatment in children in the United States.<sup>55</sup> Since 1995, up to 80% of clinician visits for OM have been found to result in antibiotic prescriptions<sup>4</sup> despite evidence of broad spectrum antibiotic use causing increasing bacterial resistance and side effects like diarrhea. The other main treatment option for COME/ROM is tympanostomy tube surgery, the most common pediatric surgery with about 10,000 cases per week in the US.<sup>56</sup> Advantages of tympanostomy tube insertion include removal of MEE, air relieving negative ME pressure, and increased air flow. Additionally, tube insertion with adenoidectomy has been effective in treating OM and reducing the amount of further surgeries.<sup>21,57,58</sup> Disadvantages include risks associated with pediatric surgery, and a risk of disease progression (by changing ME pressure).

#### 1.4 Heritability of otitis media

There is considerable evidence for a genetic contribution to COME/ROM. Twin, sibling, and family studies show a heritable component for OM, with heritability referring to the proportion of the phenotypic variance due to inherited factors. Estimated heritability of OM in a Norwegian study of 2,750 twin pairs was 0.45 for males and 0.74 for females.<sup>52</sup> Another study completed at the University of Pittsburgh followed 175 same-sex twin or triplet sets and concluded that the heritability of the mean time with OME in the first 2 years of life was 0.73, with a slightly lower heritability of time with OME found in boys.<sup>59</sup> In a U.K. study of 715 monozygotic twins and 658 dizygotic twins, heritability of acute OM was 0.57, and heritability of chronic airway obstruction, a strong predictor of COME, was 0.72.<sup>60</sup> A study of the heritability of COME/ROM in Helsinki, Finland, using a cohort of 1279 children and their parents, found estimated heritability to be 0.385 for ROM, 0.221 for COME, and 0.478 for any OM severity.<sup>61</sup> A genetic component of OM was confirmed in these studies, giving a solid rationale for genetic studies.

#### 1.5 Mouse models of otitis media

Due to anatomical and functional similarities between the ears of humans and mice, mouse models have been used to study the biology of OM and the genetic component of OM.<sup>62-64</sup> Examples of knockout mouse models found to lead to spontaneous chronic otitis media include Eyes absent 4 (*Eya4*), characterized by ME cavity and ET malformation<sup>65</sup>, Sal-like protein 4 (*Sall4*) involved in early embryonic development<sup>66</sup>, Myeloid differentiation primary response gene 88 (*MyD88*)<sup>67</sup> and Toll like receptor 2 (*Tlr2*)<sup>68</sup> that both cause defects in the immune system, and Chibby (*Cby*) which lacks motile cilia and therefore is unable to clear effusion buildup<sup>69</sup>.

Currently, there are two spontaneous models of COME/ROM identified through phenotype-driven *N*-ethyl-*N*-nitrosourea (ENU) mutagenesis screens. These models are the *Jeff* (*Jf*)<sup>70,71</sup> mouse, which has a point mutation in the F-Box Only Protein 11 (*Fbxo11*), and the *Junbo*<sup>72</sup> mouse which has a loss of function mutation in the transcription factor Ectopic viral integration site 1 (*Evi1*). Studies of the *Jeff* and *Junbo* mouse models have revealed that both of these genes interact with the TGF $\beta$  pathway and can lead to loss of TGF $\beta$  signaling. These findings have led to the investigation of the TGF $\beta$  pathway in the pathobiology of OM, and led researchers at MRC Harwell to examine the effects of knocking out *Tgfr* in the mouse.<sup>73</sup> The *Tgfr* null mouse was found to develop otitis media with chronic effusion leading to conductive deafness, and is therefore considered a new model of COME.

Inflamed microenvironments have been characterized by hypoxia<sup>74</sup>, so the *Jeff* and *Junbo* mouse models have also been used to investigate the role of the HIF-VEGF pathways in the pathogenesis of chronic otitis media<sup>13</sup>. In this study, the authors showed that inflammatory cells in the ME cavity of mice with OM were hypoxic, and that inflammatory cells in middle ear fluid had elevated levels of *Hif- $\alpha$* . The *Jeff* and *Junbo* mice showed elevated levels of numerous VEGF signaling pathway target genes, including *Vegfa*. Treatment with inhibitors of HIF-VEGF pathway reduced hearing loss, angiogenesis, and lymphatic vessel formation in the *Junbo* mouse. Using these mouse models of OM, a new pathway involved in increased susceptibility to OM was discovered, and a potential therapeutic target was identified for further study.



Other animal models have been employed to study the pathogenesis of OM, including chinchillas, rats, and rabbits. Specifically, the chinchilla has become an important model of OM because it replicates most aspects of human disease and has the ability to induce bacterial infection by small intranasal inoculum doses.<sup>75,76</sup> The chinchilla has become the most commonly used OM model due to its anatomic advantages including a large ME space and a clearly visible tympanic membrane.<sup>77</sup> Chinchillas have been pivotal in the understanding of gene expression and mucins in the ME during OM<sup>78-82</sup> as well as the polymicrobial nature of the ME<sup>83</sup>. With the chinchilla genome recently sequenced (accession number AGCDO1000000), the OM community will be able to further study many of the current candidate genes from human studies in this model.

## 1.6 Candidate gene studies of otitis media

Several studies have focused on determining the role of specific genes in the pathogenesis of OM and increased susceptibility to OM. These candidate genes have been extensively reviewed previously.<sup>84</sup> The majority of genes investigated are involved in host defense and immune response, as they are biologically plausible, as well as genes identified in mouse studies (Section 1.6). Several OM candidate genes investigated in human studies have been significantly associated with OM, including Toll like receptor 4 (*TLR4*)<sup>85,86</sup>, interleukin 1 beta (*IL1β*)<sup>87</sup>, tumor necrosis factor alpha (*TNFα*)<sup>46,88,89</sup>, mannose-binding lectin 2 (*MBL2*)<sup>90</sup>, surfactant protein a (*SFTPA*)<sup>91,92</sup>, surfactant protein D (*SFTPD*)<sup>85</sup>, mucin genes (*MUC2* and *MUC5AC/MUC5B*<sup>85,93</sup>), the TGFβ signaling pathway genes *FBXO11*, *SMAD2*, and *SMAD4*, and sodium channel voltage-gated type 1-β (*SCN1β*)<sup>85,94</sup>. The *FBXO11* gene, originally identified in a mouse model of OM, has been successfully replicated in two human studies of COME/ROM.<sup>95,96</sup> The Solute carrier family 11a member 1 (*SLC11A1*) gene was investigated for its role in resistance to

infection in mice in the Western Australia Family Study of OM (WAFSOM).<sup>97</sup> Four SNPs were found to be associated with OM, especially in those children with intact adenoids.<sup>97</sup> A recent study tested association of SNPs in candidate OM genes in 100 OM cases and 100 controls using 192 tag-SNPs in eight genes including *TLR4*, *FBXO11*, *MUC2*, *MUC5A/B*, *SMAD2*, *SMAD4*, *SCN1B*, and *SFTPD*.<sup>98</sup> No SNPs reached significance when adjusted for multiple testing due to lack of power to detect association, though SNPs from *TLR4*, *MUC5B*, *SMAD2*, *SMAD4* require follow up in larger, independent populations as they had nominal *P*-values of <0.05 before adjusting for multiple comparisons.

Candidate gene studies of OM are largely unsuccessful because they require an understanding of the biological basis of the pathology of the disease, but, as with most complex diseases, we do not fully understand the pathology of OM. Additionally, a large sample size and replication is required for validation of association. The majority of genetic analyses of OM have been candidate gene studies which are unlikely to explain the heritability of OM, described in Section 1.5, emphasizing the need for genome-wide studies.<sup>99</sup>

### 1.7 Human Middle Ear Epithelial Cell line

To investigate the function of variants associated with COME/ROM in our studies, the cell line used was derived from Human Middle Ear Epithelial cells (HMEECs). Most *in vitro* OM studies have used animal models and animal cell lines including gerbil<sup>100</sup>, rat<sup>101</sup>, and chinchilla<sup>102</sup> middle ear epithelial cell lines. The first human middle ear epithelial cell line, HMEEC-1, was immortalized by Dr. David Lim at the House Ear

Institute in Los Angeles from primary cultures of normal human middle ear epithelial cells (NHMECs) using human papillomavirus E6/E7 genes which deregulate cell growth.<sup>103</sup> Dr Lim's group found that the HMEEC-1 cells have many morphological and phenotypic properties of the NHMECs. This cell line allowed us to investigate gene expression in a cell type relevant to OM, providing information in a human *in vitro* model about molecular mechanisms that increase risk of COME/ROM in children.

## 1.8 Agnostic approaches to discover novel otitis media risk variants

Active research of OM has been focused on animal models, epidemiology, and the microbiology and virology of OM, however genetic components of chronic and recurrent infections are not well-studied. Agnostic approaches to gene discovery provide a means of identifying new genes and pathways underlying susceptibility to OM by probing evidence for association in a comprehensive scan of the whole genome. To date, only two linkage studies and two genome wide association studies (GWAS) of OM have been completed, including one described in this dissertation (Chapter 4).

### 1.8.1 Introduction to linkage studies

A linkage study is a genetic analysis which utilizes pedigree information (*i.e.*, affected sibling pairs, nuclear families, extended families) to determine segments of the genome inherited more often in affected subjects than by chance. Likewise, unaffected family members would be expected to not share these same alleles with their affected relatives. In this way, it is possible to identify genomic regions that show evidence of linkage with the disease of interest. Linkage studies do not provide fine resolution of the identified linkage peaks because they rely on genetic recombination events within families to

narrow the genomic region of interest. The family-based design of linkage studies provides a genetic analysis that is robust to population stratification and environmental factors. Recent studies have demonstrated that large linkage regions may represent multiple rare (private) variants responsible for disease inheritance within different families.<sup>104,105</sup>

### 1.8.2 Linkage studies of otitis media

#### *1.8.2.1 University of Minnesota linkage scan*

The University of Minnesota (UMN) linkage scan was the first study to investigate genetic determinants of COME/ROM across the genome.<sup>24</sup> This study recruited COME/ROM affected children along with siblings and parents. The Center for Inherited Disease Research (CIDR) genotyped 404 microsatellite markers in the COME/ROM families. After removal of genotypes and samples in quality control analyses, a total of 248,219 genotypes were generated in 591 samples, from 133 families. Significant evidence of linkage was found on 10q (LOD 3.78;  $P=3.0 \times 10^{-5}$ ) and suggestive evidence of linkage was found on 19q (LOD 2.61;  $P=5.3 \times 10^{-4}$ ). Multi-point non-parametric linkage (NPL) analyses also provided evidence of linkage to COME/ROM on chromosome 19q (LOD 2.53). Analyses conditional for linkage at chromosomes 10q and 19q resulted in increased support for linkage at 3p (unconditional LOD 0.60; conditional [10q] LOD 2.43; conditional [19q] LOD 1.84). The chromosome 19q linkage region was a large region encompassing 90 genes with biologically relevant genes with known roles in inflammation and BMP and TGF $\beta$  signaling.

### 1.8.2.2 Fine mapping of Chromosome 19q linkage scan

In a study fine mapping the chromosome 19q region previously found in the UMN genome scan of COME/ROM,<sup>106</sup> a total of 607 individuals from 139 families were included, *i.e.* all families from the UMN linkage scan with the addition of six newly-recruited families. After removal of SNPs during quality control analyses, there were 1,091 SNPs available for analyses. The greatest evidence for linkage between COME/ROM and chromosome 19q was found at 11.5cM (LOD 3.75;  $P=1.6 \times 10^{-5}$ ) which is 63.4 Mb physical distance (hg18) when marker-to-marker LD was modeled at  $r^2 > 0.5$  or  $r^2 > 0.2$ . The recurrence risk of COME/ROM in a sibling of an affected individual due to the 19q locus is twice as high as that in the general population. The 1-LOD interval region of this locus (61.6 – 63.8 Mb, hg18) contains 93 known and hypothetical genes. Within this group, there are many potential COME/ROM susceptibility candidate genes, including zinc finger protein 8 (*ZNF8*), which binds with *SMAD1* to repress of BMP and TGF $\beta$  signaling during development in mice<sup>107</sup>; nucleotide-binding domain and leucine-rich repeat containing (NLR) family, pyrin domain containing 13 (*NLRP13*), NLR family, pyrin domain containing 5 (*NLRP5*), and NLR family, pyrin domain containing 8 (*NLRP8*), members of the inflammasome complex which is an important regulator of the innate immune response<sup>108-110</sup>; and Alpha-1-B glycoprotein (*A1BG*) which has been associated with severe inflammation.<sup>111</sup>

### 1.8.2.3 University of Pittsburgh linkage scan

The University of Pittsburgh (UPitt) linkage scan<sup>112</sup> study sample included at least two full siblings who had history of tympanostomy tube surgery due to a significant history of OM, along with their parents and other full siblings with no history of tube surgery.

Genotyping was carried out using two versions of the Affymetrix's GeneChip Human Mapping 10K Array in a total of 1506 individuals, including 1431 European American and 75 African American individuals. After removal of SNPs for quality control measures, 8,802 SNPs were tested for linkage to recurrent/persistent OM. Analyses of European American families only detected the most significant region of linkage at 17q12 (LOD 2.83;  $P=7 \times 10^{-5}$ ) and four other peaks with suggestive P-values: 10q22.3 ( $P=1.81 \times 10^{-3}$ ), 7q33 ( $P=1.05 \times 10^{-3}$ ), 6p25.1 (LOD 2.25;  $P=2.61 \times 10^{-3}$ ), and 4p15.2 ( $P=3.01 \times 10^{-3}$ ). Analyses using the combined European American and African American families strengthened evidence of linkage to recurrent/persistent OM at 10q22.3 ( $P=2.6 \times 10^{-4}$ ) but significance at the other peaks decreased. In the chromosome 17q12 linkage peak, OM susceptibility candidate genes include adaptor-related protein complex 2, beta 1 subunit (*AP2B1*), a gene associated with the down-regulation of CD8 and a cluster of chemokine C-C motif ligand (*CCL*) genes including *CCL5* which has previously been associated with OM. This study did not confirm the regions found to be significantly linked to COME/ROM in the UMN linkage study: 10q26.3 and 19q13.43. The findings of this study need further fine mapping, replication, and functional studies to determine causal SNPs increasing risk of recurrent/persistent OM.

#### *1.8.2.4 University of Western Australia investigation of regions previously found in linkage studies*

This study aimed to investigate replication of previously reported regions of linkage to COME/ROM, including 3p25.3, 10q22.3, 10q26.3, 17q12, and 19q13.43. A minimum of four microsatellites were genotyped across each of the five regions of interest in Western Australia family study on otitis media (WAFSOM), which included 468 individuals from 101 families, and 208 OM cases. Significant evidence of linkage to COME/ROM was

found only at 10q26.3 using both multi-point NPL analysis ( $P=0.0036$ ) and single-point NPL ( $P=0.012$ ). Borderline evidence of linkage to COME/ROM was found at 10q22.3 ( $P=0.05$ ), a linkage peak discovered in the UPitt linkage scan. There was no evidence of linkage at the three other previously reported linkage peaks.

In addition, association mapping was carried out across the 10q26.3 peak in a subset of individuals from the Western Australia Pregnancy Cohort (also known as the Raine Cohort) in 831 individuals, including 256 OM cases with available genotype data. Data from a total of 10,185 SNPs (2,270 genotyped SNPs and 7,915 imputed SNPs) across 10q26.3 were used. Significant associations robust to Bonferroni correction were detected at rs9418832 ( $P_{\text{adj-PC}}=7.48 \times 10^{-5}$ ), an intronic SNP in dedicator of cytokinesis (*DOCK1*), and rs7922424 ( $P_{\text{adj-PC}}=9.47 \times 10^{-6}$ ), an intergenic SNP found between transcription elongation regulator like protein (*TCERG1L*) and protein phosphatase 2, regulatory subunit B, delta (*PPP2R2D*). This study was able to replicate the linkage between chromosome 10q22.3 and COME/ROM in WAFSOM, and narrowed down the list of potential candidate genes in the region for future functional studies.

### 1.8.3 Introduction to genome wide association studies

While linkage studies examine patterns of inheritance within families, GWA studies use a population-based approach to examine evidence of allelic association with disease outcomes. In a GWAS, SNPs are genotyped across the entire genome, and each SNP is then individually tested for association with the trait or disease of interest. SNPs are selected using known patterns of linkage disequilibrium (LD) across the genome in a relevant population, providing finer resolution than linkage studies. This allows for the resulting association to be narrowed down to a relatively modest number of SNPs,

compared with the megabase-level resolution of linkage studies. GWA studies are able to find common SNPs associated with disease, but associated variants frequently have relatively small effect sizes. The ability to detect association for a particular genetic locus depends on SNP density, coverage, and patterns of LD within the genetic region harboring the causal variant(s) for the disease of interest. Associations with rare and infrequent variants typically cannot be identified through GWAS, as such variants may not be captured by the LD-based SNP selection, and analyses are often underpowered to detect them.

Risk alleles have been significantly associated in GWAS, but results do not explain a substantial portion of the disease/trait heritability. Researchers have expanded GWAS SNP panels from less than a hundred thousand to millions of SNPs and also have recruited thousands more individuals in hopes of finding variants that explain more heritability of diseases. Most of the variants identified in these association studies fall in noncoding regions of the genome, many of which are a large distance from protein coding regions.<sup>113</sup> These findings indicate that risk variants of complex diseases may affect gene regulation, instead of directly impacting gene function. Large-scale efforts have been made to generate public data for a variety of functional elements in noncoding regions, *i.e.* the Encyclopedia of DNA elements (ENCODE).<sup>114</sup> These resources provide evidence of functionality of noncoding variants, including many associated in GWAS, which can then be followed up in relevant tissues.



#### 1.8.4 Raine cohort genome-wide association study of acute and/or chronic otitis media

This was the first study to investigate genetic contributors to OM using a genome-wide association approach.<sup>115</sup> Analyses included all subjects with available GWAS data from the Raine cohort. OM status was determined using clinical examination and parental report from the first three years of life. Participants were classified as cases if they had presence of inflamed, retracted or scarred tympanic membrane, ME effusion or tympanostomy tubes. Cases were also defined by parental report of  $\geq 3$  episodes of AOM by the age of 3 yrs, with 35% of cases defined by this parental report alone. Genotype data was generated on the Illumina 660W Quad Beadchip at the Centre for Applied Genomics. There were 2,524,817 SNPs (genotyped and imputed) available for analysis in 416 cases and 1,075 controls. No SNP exceeded genome-wide significance ( $P < 5 \times 10^{-8}$ ). There were 46 variants that, when adjusted for two principal components (PCs) to account for population stratification, showed associations at  $P_{\text{adj-PCA}} < 10^{-5}$ . The top associations in this study were found in the region upstream of Calpain 14 (*CAPN14*) on chromosome 2. Association with this gene was further supported when the full imputed dataset was analyzed and the top association was rs6755194 ( $P_{\text{adj-PCA}} = 8.3 \times 10^{-7}$ ) which lies in an intron of *CAPN14*. The most highly-associated SNPs were located in the genes *CAPN14* and *GALNT14*, and these novel genes may have putative functional roles in OM susceptibility. Gene-based analysis was carried out using Versatile Gene-based Association Study (VEGAS)<sup>116</sup> to investigate gene-wide evidence of association. VEGAS incorporates genotype data from all SNPs within a gene to test for association with LD between SNPs taken into consideration. The most highly significant genes from this analysis included *CAPN14*, BPI fold containing family A, member 3 (*BPIFA3*), and BPI fold containing family A, member 1 (*BPIFA1*).

#### 1.8.4.1 Replication study of Raine cohort GWAS risk variants in Western Australia

##### *Family Study of Otitis Media (WAFSOM)*

To follow up the discovery GWAS, variants from the Raine cohort GWAS of acute and chronic OM were genotyped in an independent population of COME/ROM.<sup>115</sup> The most significant associations ( $P < 10^{-5}$ ) from the discovery GWAS were analyzed using the Ingenuity Pathway Analysis and were investigated for replication in WAFSOM. Genes which interact with the TGF $\beta$  pathway were focused on due to previous association of *Fbxo11*<sup>95</sup> in WAFSOM. Twenty SNPs were chosen from the top hits from the discovery GWAS plus genes identified from the pathway analysis, however none of the associations were replicated. Lack of replication is postulated to be due to phenotypic differences in the populations (acute and chronic OM versus COME/ROM), study design, and sample size. They calculated that at least 1700 families would be necessary to provide >80% power to confirm discovered loci with MAF 0.010 and  $OR \leq 1.4$ , assuming the observed GWAS effect sizes were over-inflated.

#### 1.9 Resources needed to discover novel risk variants of otitis media

Although candidate gene studies have contributed to the knowledge of the pathogenesis of OM, agnostic approaches can discover novel genes and pathways associated with OM. This knowledge will allow for the discovery of novel treatments that could reduce the amount of pediatric antibiotic prescriptions and surgeries. OM research in general lacks genome-wide studies which aim to uncover new mechanisms which increase susceptibility to OM and its various severities. Before the research described in this dissertation commenced, the two linkage scans of OM had been carried out in family-based populations. During the progress of this research, the most significant linkage

peak from the UMN linkage scan was fine mapped and confirmed, and the linkage peak at 10q22.3 from the UPitt linkage scan was replicated in the Western Australian family study of OM. Also, the GWAS of acute and chronic OM conducted in the Raine cohort at the University of Western Australia was published. Though there have been three genome-scale studies that have discovered novel loci, there are so few population studies of OM that it is difficult to replicate findings. It is critical that researchers recruit more OM populations using standardized phenotype definitions, to conduct genome-scale analyses and replication studies to advance our knowledge of genetic contributions to the pathogenesis of OM.

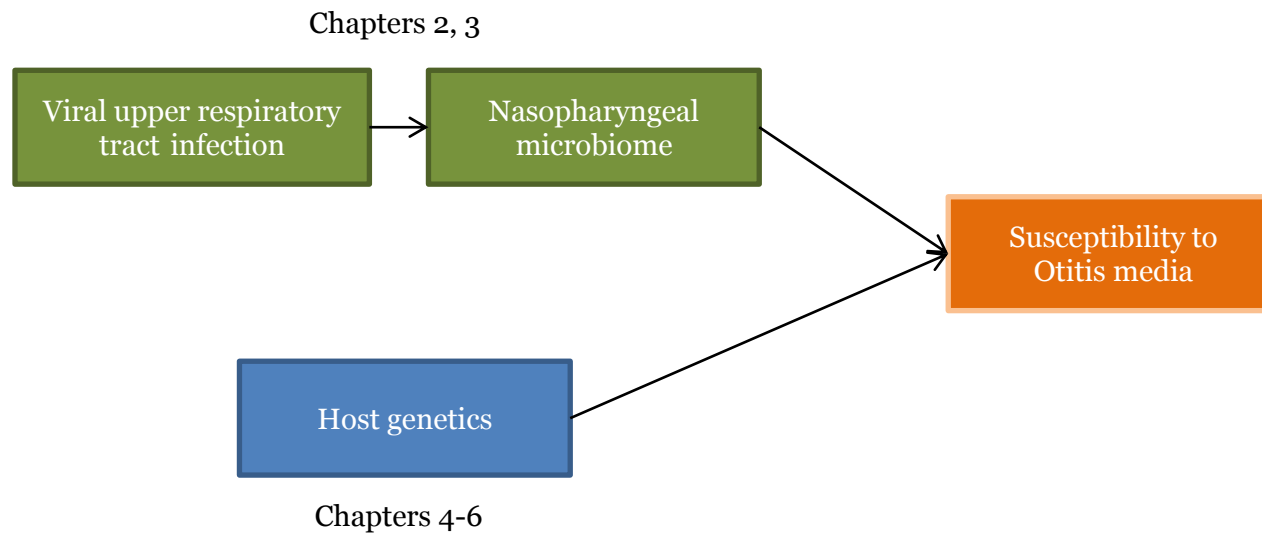
#### 1.10 Conclusion and goals

The aims of the research contained in this thesis were to identify and explore novel contributors to the pathogenesis of otitis media, from both bacterial and host perspectives. We investigated changes in the bacterial composition (microbiome) of the nasopharynx during Rhinovirus infection in human subjects. Understanding these changes provides important information on the impact of viral infection on bacterial community shifts in the nasopharynx, which is known to be a reservoir of pathogenic bacteria. This reservoir is located at the base of ETs, and pathogenic bacteria may ascend the ETs after viral infection. Shifts in the microbial community structure due to an URTI could provide insight into causes of complications of URTI, *i.e.* OM, and also increase specificity of antibiotic prescriptions and potentially reduce treatment failures. A second aim was to identify novel genetic contributors to COME/ROM through a GWAS approach, determine likely causal variants, and explore the potential roles of implicated genes and variants in the disease. We have conducted the studies described in the

following chapters, with the rationale linking these complementary approaches outlined in **Figure 1.2**.

Recurrent and chronic infections can lead to effusion build up behind the eardrum leading to deafness and developmental delays in speech and learning.<sup>117-119</sup> Due to the severity of current treatments and the detrimental nature of sequelae for children with COME/ROM, determining reasons for susceptibility are crucial. The successful completion of our research has uncovered novel host and bacterial aspects of OM pathogenesis that were previously unknown, providing insights into OM susceptibility.

**Figure 1.2** Diagram linking chapters of this thesis in terms of OM pathogenesis



## Chapter 2

### Survey of the nasopharyngeal microbiome in healthy adults

#### 2.1 Chapter Introduction

The paradigm of OM pathogenesis is that a viral infection of the URT initiates inflammation which extends into the ET, allowing pathogenic bacteria from the nasopharynx to ascend into the ME. It is known how viruses can aid in colonization and invasion of bacteria (Section 1.3.4), though it is not well understood how viral infection may alter the bacterial composition of the nasopharynx. Knowing what changes in bacterial composition occur in response to URTI could impact our understanding of URTIs, including OM. In our study, we sought to examine changes in the bacterial community of the nasopharynx due to viral infection in healthy adults. We used a self-inoculation method to better model naturally occurring viral infections. Using healthy individuals limits confounding factors such as natural viral infections and co-infections. Also, challenge with a known virus gives us exact timing of infection, which is helpful when investigating changes in the microbiome. Most studies which aim to determine effects of microbial infection on the host have looked at populations of individuals with a naturally occurring URTI or complication (*e.g.*, sinusitis) and sample each individual for specific pathogens using culture-based methods or PCR of a modest number of selected bacterial species. These studies are limited because there could be multiple viruses causing changes in the microbiome or host response, the timing of infection is unknown so a time course across various samples will not be comparable, and detection methods are limited. Our method of pathogen detection was a clinical diagnostic microarray that harnesses culture-independent PCR methods to detect a panel of known pathogenic bacterial species in the samples before, during, and after infection. We utilized the Prove-it™ Bone and Joint array as it contained URT pathogens similar to the Prove-it™

Sepsis array, but also included *Moraxella catarrhalis*, a known otopathogen. The long-term goal of our study is to use these novel strategies to prevent and even treat URTIs through modifying host response to infection by promoting a healthy, balanced microbiome.

## **2.2 Bacteria in the nose of young adults during wellness and rhinovirus colds - detection by culture and microarray methods in 100 nasal lavage specimens**

*International Forum of Allergy and Rhinology, 2013 Sep;3(9):731-9.*

E. Kaitlynn Allen, BA<sup>1,2</sup>, Anne Pitkäranta, MD<sup>3</sup>, Minna Mäki, PhD<sup>4</sup>, J. Owen Hendley, MD<sup>5</sup>, Sanna Laakso, MSc<sup>4</sup>, Michèle M. Sale, PhD<sup>1,2,5</sup>, Birgit Winther, MD<sup>6</sup>

<sup>1</sup>Center for Public Health Genomics, <sup>2</sup>Department of Biochemistry and Molecular Genetics, University of Virginia, USA. <sup>3</sup>Department of Otolaryngology, University of Helsinki, Finland. <sup>4</sup>Mobidiag, Ltd. Helsinki, Finland. <sup>5</sup>Department of Pediatrics, <sup>5</sup>Department of Medicine, <sup>6</sup>Department of Otolaryngology, University of Virginia, USA.

E. Kaitlynn Allen conducted all experiments except bacterial culturing, and drafted the manuscript with Dr. Birgit Winther. Drs. Birgit Winther and J. Owen Hendley conceived and designed the challenge study, and Dr. Birgit Winther coordinated study recruitment, and carried out sample processing and storage. Kathleen Ashe conducted bacterial titer by culture methods. Dr. Minna Mäki and Ms. Sanna Laakso acted in an advisory capacity over technical work completed at Mobidiag, Ltd. Drs. Anne Pitkäranta, Michèle M. Sale, Minna Mäki and Birgit Winther contributed to data interpretation and reviewed the manuscript.

### 2.2.1 Abstract

**Background:** Patients with viral respiratory infections/viral rhinitis/common colds are often treated with antibiotic although there is little information on whether or how bacterial microbiota in the nose and nasopharynx might influence the course of viral illnesses.

**Methods:** To initiate investigation of possible interaction between viral respiratory illness and microbiota of the nose/nasopharynx, we utilized microarray technology to examine 100 nasal lavage fluid samples (NLF) for bacterial species and recorded the bacterial titer of culturable bacteria. Rhinovirus illnesses were induced by self-inoculation using the “finger to nose or eye natural transmission route” in ten otherwise healthy young adults. NLF samples were collected during wellness and at specific time points following experimental rhinovirus inoculation.

**Results:** The rhinovirus infection rate was 70%. There were no consistent changes in the prevalence of different bacterial species determined by microarray and bacterial titer by culture methods during rhinovirus infection. The bacterial profile in NLF samples showed high variability between volunteers but low variability in multiple NLF's obtained before and following infection from the same volunteer. *S. epidermidis*/coagulase negative *staphylococcus* (CNS) were identified in all ten subjects. One or more bacterial sinus/otitis pathogens were identified by microarray in six of the ten volunteers. The microarray identified a few bacteria not included in traditional bacterial cultures.

**Conclusion:** Our pilot study showed that each of ten volunteers had a unique bacterial profile in the nose by microarray analysis and that bacterial load did not change during experimental rhinovirus colds. Larger scale studies are warranted.



### 2.2.2 Introduction

The symptom complex of viral upper respiratory tract infections (URTIs, common colds) and bacterial complications (bronchitis, otitis media and sinusitis) are self-limiting, but they cause billions of dollars in healthcare costs and millions of days of school and work missed in the U.S. every year. Rhinovirus is the major cause of URTIs and young preschool age children contract as many as six rhinovirus infections per year.<sup>2,3</sup> In the U.S., the majority of antibiotics are used for URTIs, otitis, sinusitis and bronchitis. However, it remains unclear whether commensal bacteria may become pathogenic during the viral infection. Many pathogenic bacteria including *Streptococcus pneumoniae*, *Haemophilus influenzae*, and *Moraxella catarrhalis* reside in the nasopharynx in 60% -70% of healthy adults and children<sup>6-10</sup>, but these bacteria may exacerbate symptoms during viral infections.<sup>11</sup> Further understanding of possible viral-bacterial interactions during URTIs are warranted to provide evidence for more appropriate use of antibiotics in colds.

Studies of nasopharyngeal bacteria during URTIs have been done without knowledge of the timing of the viral infection. Traditionally, bacterial identification was based on culture methods, but new technologies allow detection of bacterial genomes by microarray and sequencing (e.g. 454 pyrosequencing).<sup>120</sup> In the present study, we utilized a new clinical model to study bacteria during colds by evaluating patients with experimentally-induced rhinovirus colds and the application of a novel method of bacterial detection using microarray technology to determine the bacteria in the nasal cavity of otherwise healthy adults prior to and during rhinovirus infections. Experimental cold inoculation of volunteers with rhinovirus allows examination of

changes in the commensal and pathogenic bacterial flora in the upper airways at precise time points during rhinovirus infections.

### **2.2.3 Methods**

#### *2.2.3.1 Participants*

Volunteers between the ages of 18 and 65 were recruited by advertisement during October-November 2010. Subjects were eligible if they had a screening serum neutralizing antibody of 1:4 or less to the challenge rhinovirus type 39. Subjects were not eligible if they had a history of chronic sinus disease, history of sinus surgery, allergic rhinitis or if they had received topical or oral antibiotics within the four weeks prior to study participation or had upper respiratory tract symptoms during the two weeks prior to rhinovirus challenge. The study was approved by the Institutional Review Board for Health Sciences Research at the University of Virginia. Informed consents were obtained from all participants prior to enrollment.

#### *2.2.3.2 Rhinovirus self- inoculation*

Rhinovirus immunotype 39 was used as the challenge inoculum. This inoculum pool has been safety tested and approved for use by the FDA (IND12934). All subjects were exposed to a total of 100-300 TCID<sub>50</sub> of rhinovirus in 250µl by self-inoculation. Following hand washing, the volunteer used the tip of the pointer finger on the dominant hand to dip in the inoculum in a sterile Petri dish. The subject then inoculated her/himself by touching the medial cantus and conjunctiva of one eye or the septum in

the nasal vestibulum of one side of the nose. The procedure was repeated once after a 5-15 minute interval.

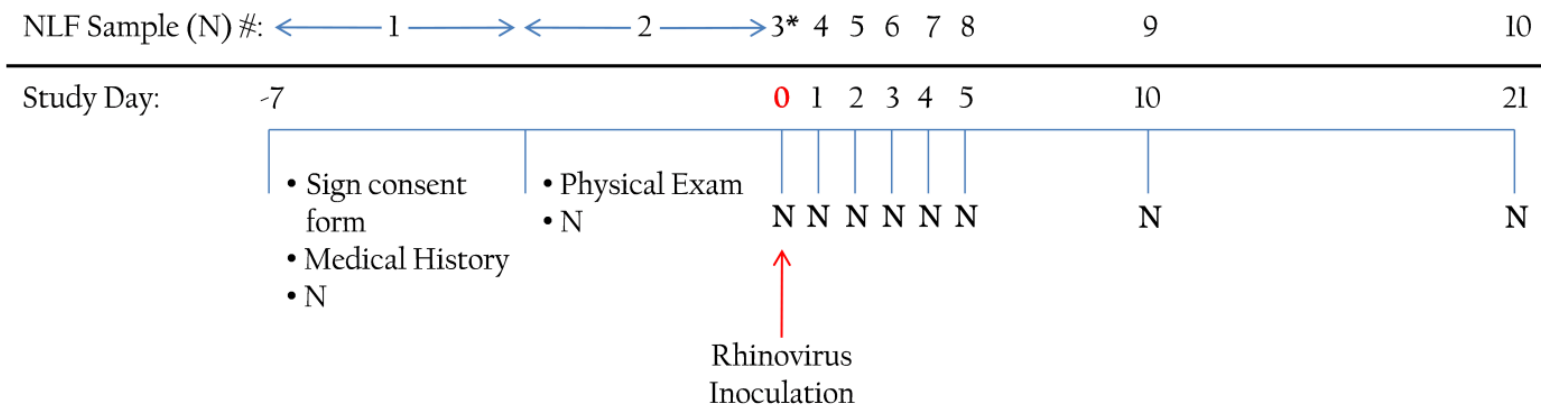
#### *2.2.3.3 Specimen collection*

Nasal lavage fluid (NLF) was obtained by installation of 5mL of 0.9% saline into each nasal cavity while the subject had the head tilted backwards, repeating the “k-k-k” sound multiple times to close the soft palate to minimize fluid running into the throat. After a few seconds, the subject moved the head forward and 5-8 mL of nasal lavage fluid was recovered into a waxed paper cup. Ten nasal washes were obtained from each volunteer: three during the week prior to HRV inoculation, one on each of five days immediately following inoculation, and on days 10 and 21 following inoculation (**Figure 2.1**). An aliquot of 1mL of each NLF samples was placed in tubes containing an equal volume of viral collection broth, kept on ice, and transported to the laboratory within one hour for rhinovirus isolation in tissue cultures. The remaining NLF was transferred to 2mL cryotubes and stored frozen at -80°C.

#### *2.2.3.4 Assessment of Rhinovirus infection*

The presence of rhinovirus in nasal and nasopharyngeal mucus was examined in NLF during each of five days following inoculation. 0.2 mL of NLF from each subject was inoculated into two fibroblast monolayer tubes (embryonic lung fibroblast cell/MRC-5 or WI 38) and incubated at 33°C in roller drums for 14 days. Tubes were examined every other day for rhinovirus cytopathic effect. One isolate from each subject was confirmed as the challenge virus by neutralization testing with antibody to rhinovirus 39. Subjects with a rhinovirus isolated from the nasal lavage specimens were considered to be

**Figure 2.1:** Study timeline including when information or samples were collected from participants. \*NLF samples were obtained prior to Rhinovirus inoculation on this day.



infected. The serum antibody response to the challenge rhinovirus type 39 was examined on sera obtained prior to inoculation and three weeks following inoculation by standard methods.<sup>121</sup> Subjects with a 4-fold increase in antibody titer to the challenge virus were considered to be infected.

#### *2.2.3.5 Assessment of illness*

The severity of symptoms was recorded each morning by the study nurse when the subjects returned to the study site, prior to any study procedures. Upper respiratory tract symptoms/nasal symptoms including sneezing, runny nose, nasal obstruction and sore throat, and non-nasal symptoms including malaise, chilliness, cough and headache were recorded at baseline and each morning following rhinovirus inoculation by asking the subject to judge the severity of their symptoms on a scale from 0 to 4. The scale indicated 0 as absent (symptom not noticeable); 1 as mild (symptom by itself causes no limitation of usual activities); 2 as moderate (symptom by itself causes some limitation of usual activities); 3 as severe (symptom by itself causes severe limitation or inability to carry out usual activities); and 4 as very severe (symptom by itself leaves you unable to carry out usual activities). The score for each cold symptom present before the virus challenge was subtracted from each daily score for that symptom. The daily symptom scores were obtained at day zero and for 5 days following virus inoculation. The diagnosis of a cold illness (modified Jackson cold<sup>122</sup>) required a total symptom score of  $\geq 6$  for 5 days and either the presence of rhinorrhea on 3 or more days or the subjective impression of having a cold.

#### *2.2.3.6 Assessment of bacterial load by culture*

Nasal lavage fluid was transported to the laboratory on ice and cultured within one hour after collection. One tenth of 1 mL of NLF was inoculated onto both a blood and chocolate agar plate and incubated at 35° C for 48 hours. Number of bacterial colonies was examined at 24 and 48 hours. If there were less than 200 colonies, they were counted; 25-75 colonies was recorded as  $10^{2.5}$  colonies per ml; 200 colonies was recorded as  $10^{3.5}$  per ml; confluent growth on plates was recorded as  $10^{4.0}$ , and more than 200 but less than confluent was recorded as  $10^{3.5}$  colonies per mL of NLF.

#### *2.2.3.7 DNA extraction from nasal lavage fluid samples*

From each NLF sample, 1.5 mL of NLF was centrifuged at 2350 g (5000 rpm) for 10 minutes, and the remainder of the sample was stored at -80° C. After centrifugation, 1mL of supernatant was discarded and 20µL Proteinase K, at 20mg/mL (Roche Proteinase K, recombinant, PCR grade, Roche, Indianapolis, IN), was added to the 0.5 mL of sample and incubated at 60°C with 300 rpm agitation for 2 hours. Samples were then heated to 95°C to block Proteinase K activity. Extraction with a starting volume of 520µL and an elution volume of 60µL were done with Nordiag Arrow VIRAL NA® (CE/IVD) kit and Viral NA v1.0 program (NorDiag (Autogen), Holliston, MA). Seven NLF samples did not have sufficient extracted DNA and were re-extracted using the remainder of the stored supernatant and a different extracting method suitable for smaller volumes (Nuclisens® easyMAG®, BioMerieux, Inc., Marcy-l'Étoile, France). DNA was quantified using 1µL of extracted DNA by NanoDrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA). Concentrations and purity measures (A260/280) were recorded.

#### *2.2.3.8 DNA Amplification and hybridization to the microarray*

Broad range bacterial PCR was carried out according to the instructions of the Prove-it™ Bone and Joint assay (Mobidiag, Helsinki, Finland, <http://www.mobidiag.com>) which includes primers for bacteria tested (**Table 2.1**). Amplicons were hybridized onto the Prove-it™ Bone and Joint Strip Array using a modified version of the hybridization assay.<sup>123,124</sup> The modified version of the assay used 10µL PCR product instead of using the standard volume of 3µL PCR product for hybridization, though all other steps were followed as stated in the protocol.

#### *2.2.3.9 Microarray scanning and analysis*

Microarray images were scanned using the Prove-it™ StripArray Reader and analyzed by Prove-it™ Advisor software using Mobidiag default parameters.<sup>123</sup> Analysis by Prove-it™ Advisor software includes simultaneous identification of bacterial targets and evaluation of the control probes included in the array. Results were detected by signals from the hybridization of PCR

products with target specific oligonucleotides on the array. The software interpreted the array image using built-in rules and parameters which are specific for the assay type.

### **2.2.4 Results**

#### *2.2.4.1 Rhinovirus infection and illness rates*

Ten eligible subjects with serum neutralizing antibodies <4 were enrolled. Four were females (mean age 18 1/2 years) and six were males (mean age 20 years). None of the subjects withdrew from the study or had serious adverse events.

**Table 2.1** List of Species covered on the Mobidiag Prove-it™ Bone and Joint Assay

(<http://www.mobidiag.com>)

Gram+ Bacteria	Gram - Bacteria
Coagulase negative <i>staphylococcus</i> *	<i>Haemophilus influenzae</i>
<i>Staphylococcus aureus</i>	<i>Moraxella catarrhalis</i>
<i>Staphylococcus epidermidis</i>	<i>Acinetobacter baumannii</i>
<i>Streptococcus agalactiae</i>	<i>Bacteriodes fragilis</i> group**
<i>Streptococcus dysgalactiae</i> subspecies <i>equisimilis</i>	<i>Enterobacteriaceae</i> ***
<i>Streptococcus pneumoniae</i>	<i>Enterobacter aerogenes</i>
<i>Streptococcus pyogenes</i>	<i>Enterobacter cloacae</i>
<i>Clostridium perfringens</i>	<i>Escherichia coli</i>
<i>Enterococcus faecalis</i>	<i>Fusobacterium necrophorum</i>
<i>Enterococcus faecium</i>	<i>Kingella kingae</i>
<i>Listeria monocytogenes</i>	<i>Klebsiella oxytoca</i>
	<i>Klebsiella pnueumoniae</i>
*Detects <i>S. haemolyticus</i> , <i>S. hominis</i> , <i>S. lugdunensis</i> , <i>S. saprophyticus</i> , <i>S. warneri</i> , <i>S. xylosus</i>	<i>Neisseria meningitidis</i>
	<i>Neisseria</i> species non-meningitidis****
	<i>Proteus mirabilis</i>
<i>mecA</i> meticillin resistance marker	<i>Proteus vulgaris</i>
	<i>Pseudomonas aeruginosa</i>
	<i>Salmonella enterica</i> subspecies <i>enterica</i> *****
	<i>Serratia marcescens</i>
	<i>Stenotrophomonas maltophilia</i>
	<i>Campylobacter jejuni/coli</i>
	**Detects <i>B. fragilis</i> , <i>B. vulgatus</i> , <i>B. thetaiotaomicron</i>
	***Detects <i>C. amalonaticus</i> , <i>C. braakii</i> , <i>C. freundii</i> , <i>C. koseri</i> , <i>E. hormaechei</i> , <i>E. sakazakii</i> , <i>K. intermedia</i> , <i>M. morganii</i> , <i>P. agglomerans</i> , <i>P. rettgeri</i> , <i>P. stuartii</i> , <i>Y. enterocolitica</i> , <i>Y. pseudotuberculosis</i>
	****Detects <i>N. gonorrhoeae</i> , <i>N. subflava</i> , <i>N. sicca</i> , <i>N. cinerea</i> , <i>N. elongata</i> subspecies <i>nitroreducens</i> , <i>N. flavescens</i> , <i>N. lactamica</i> .
	*****Detects the following serovars: <i>Enteritidis</i> , <i>Oranienburg</i> , <i>Othmarschen</i> , <i>Panama</i> , <i>Paratyphi</i> , <i>Stanley</i> , <i>Typhimurium</i> , <i>Virchow</i> , group A, B, C, D.



Six subjects were inoculated into the nose and four subjects into one eye. Overall, 70% of rhinovirus challenged subjects became infected. Four of six subjects inoculated into the nose got infected; three of the four inoculated into one eye got infected (**Table 2.2**).

Five subjects shed rhinovirus into nasal secretion sometime during days 1-5 after inoculation, and six subjects had a 4-fold rise in serum neutralizing antibodies to the challenge virus. Three of the seven HRV infected subjects had clinical illness based on the modified Jackson colds method. The severity of illness in the rhinovirus infected volunteers appeared similar regardless of the nasal or eye inoculation method. One of three subjects who did not get infected with the challenge rhinovirus had clinical illness based on the modified Jackson method.

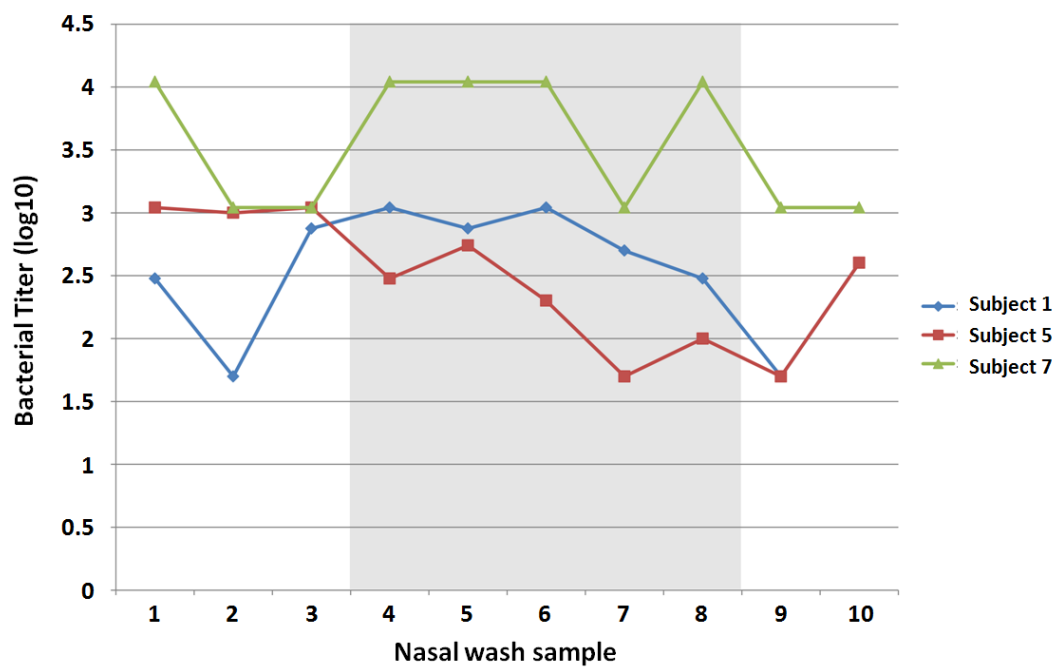
#### *2.2.4.2 Bacterial culture*

The titer of bacteria in NLF from each subject was surprisingly stable, varying only one log in 10 different samples, whereas the inter-individual variation was up to 4 fold. There were no apparent changes in the culturable bacterial load during rhinovirus infection compared to during wellness (**Figure 2.2a-2.2c**). Two of four HRV infected subjects (volunteers 4 and 7) had a high bacterial load ( $>10^3$  cfu/mL) prior to inoculation and it remained high; the two other subjects (volunteers 1 and 5) dropped the bacterial load by one log from day 3 to 5 following rhinovirus inoculation. Three subjects (volunteers 3, 6 and 8) had a low bacterial titer ( $\leq 10^3$  cfu/mL) throughout the study. The bacterial titer was not different in three symptomatic rhinovirus infected subjects (**Figure 2.2a**) compared to four rhinovirus infected subjects without symptoms (**Figure 2.2b**).

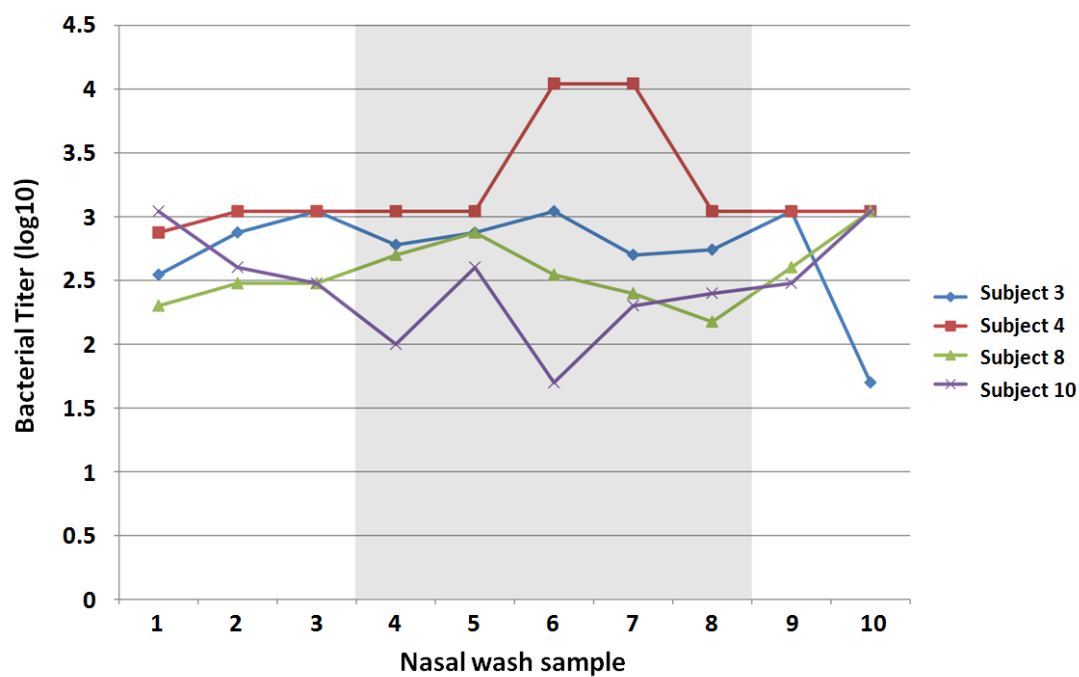
**Table 2.2** Infection and Illness in Volunteers following self-inoculation of HRV 39

<b>Subject</b>	1	6	7	8	2	3	4	5	9	10
<b>Route of inoculation</b>	Eye	Eye	Eye	Eye	Nose	Nose	Nose	Nose	Nose	Nose
<b>HRV infection</b>	yes	no	yes	yes	no	yes	yes	yes	no	yes
<b>Clinical illness</b>	yes	no	yes	no	no	no	no	yes	yes	no
<b>Total symptoms score</b>	7	0	29	6	3	4	4	20	12	0

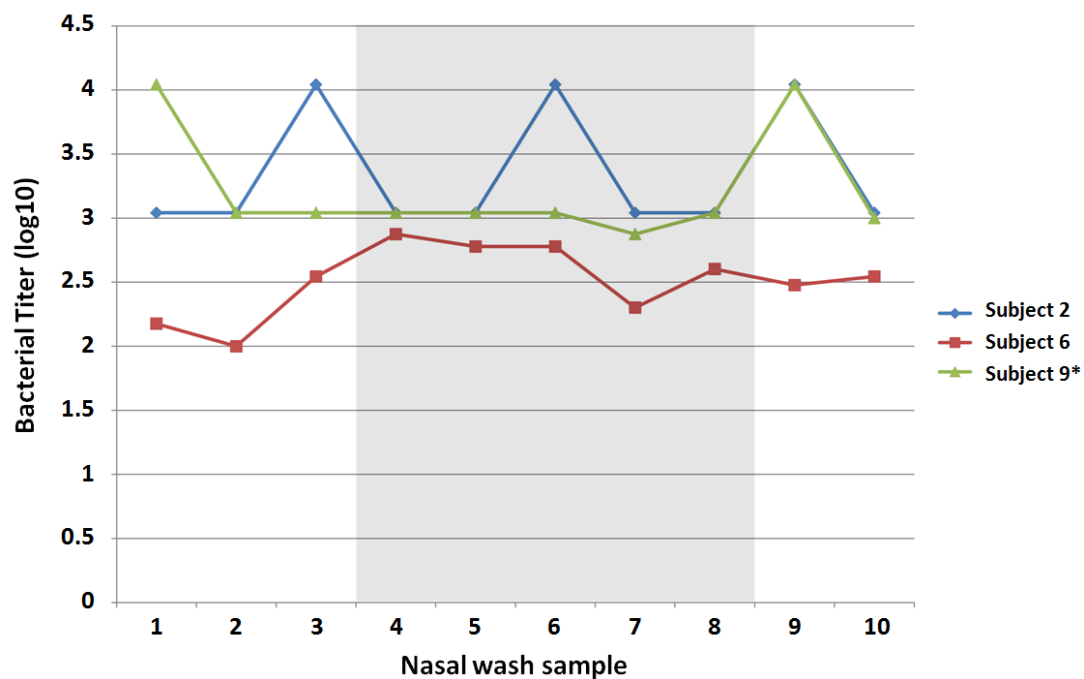
**Figure 2.2 A:** Bacterial titer by semi-quantitative cultures in nasal wash samples Rhinovirus infected volunteers with illness. Shaded region indicates nasal wash samples obtained during the first five days following HRV 39.



**Figure 2.2 B:** Bacterial titer by semi-quantitative cultures in nasal wash samples in Rhinovirus infected volunteers without illness. Shaded region indicates nasal wash samples obtained during the first five days following HRV 39.



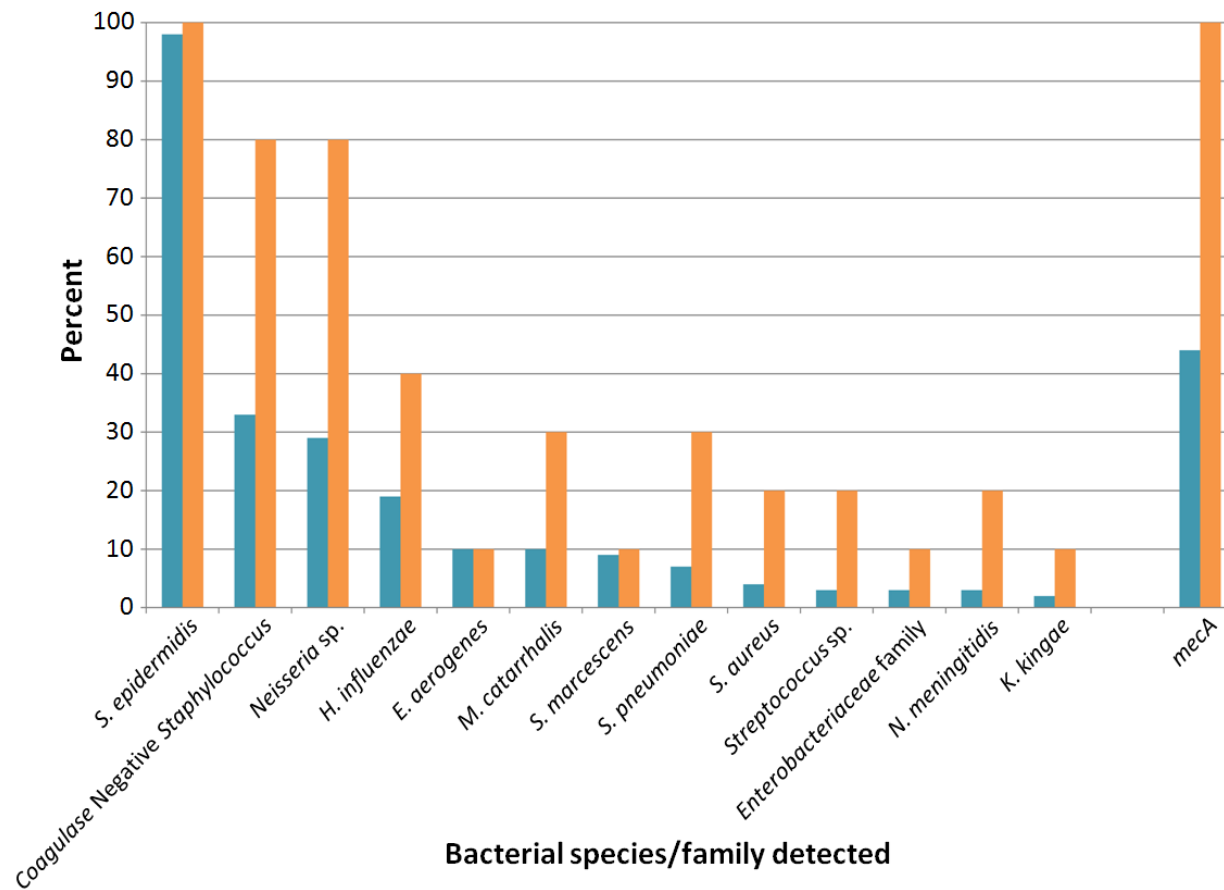
**Figure 2.2 C:** Bacterial titer by semi-quantitative cultures in nasal wash samples in volunteers who were not infected with Rhinovirus. Shaded region indicates nasal wash samples obtained during the first five days following HRV 39. \*Indicates symptomatic subject



### 2.2.4.3 Microarray

Amplification of bacterial DNA and hybridization to the microarray was successful in all 100 NLF samples from ten volunteers. Prove-it™ Advisor software identified 13 bacterial species out of the bacterial species tested (**Table 2.1**). Detailed microarray results for each subject are listed in Table S2.1. Samples analyzed for quality assurance were identical with original bacterial results. *S. epidermidis*/Coagulase negative *Staphylococcus* (CNS) was present in all volunteers, and *mecA* was present sporadically in samples from all ten volunteers. One or more of the potential otitis or sinus pathogens (*H. influenzae*, *S. pneumoniae* and *M. catarrhalis*) were detected in at least one NLF sample from 60% of volunteers. *H. influenzae*, *S. pneumoniae* and *M. catarrhalis* were present in 40%, 30% and 30% of subjects, respectively (**Figure 2.3**). *H. influenzae* was present in 90% of samples from one volunteer (**Table 2.3c**, Volunteer 9). In each of the ten individuals, the prevalence of different bacterial species was similar across the ten samples obtained prior to and following rhinovirus inoculation (**Table 2.3a-2.3c**). A few bacterial species, such as *S. marcescens*, *E. aerogenes* and *K. kingae*, are not usually assayed with traditional culturing technique but were detected in the NLF samples by microarray (**Table 2.3a-2.3c**).

**Figure 2.3:** Prevalence of bacteria identified by microarray in 100 samples from 10 volunteers inoculated with HRV39. Blue bars represent the percentage of **samples** (out of 100) which the species/gene was detected. The orange bars represent the percentage of **subjects** (out of 10) in whom the species/gene was detected.



**Table 2.3 A:** Bacteria detected by microarray in nasal lavage fluid samples obtained during wellness (well) and on days following rhinovirus challenge (days 1-21). *S. epidermidis*/CNS was positive in 99 samples/all volunteers and *mecA* was detected in all volunteers sporadically and therefore are not included in the table.

*Rhinovirus infected volunteers with illness*

<b>Volunteer 1</b>	Well	Well	Well	Day1	Day2	Day3	Day4	Day5	Day10	Day21
<i>S. aureus</i>				pos						
<i>S. pneumoniae</i>			pos		pos				pos	
<i>H. influenzae</i>			pos		pos	pos	pos			
<i>Enterobacteriaceae</i> family			pos	pos	pos					
<i>Neisseria</i> sp.					pos			pos		
<b>Volunteer 5</b>										
<i>Neisseria</i> sp.						pos	pos	pos		
<b>Volunteer 7</b>										
<i>Neisseria</i> sp.			pos	pos	pos	pos			pos	pos



[illegible]

**Table 2.3 C:** Bacteria detected by microarray in nasal lavage fluid samples obtained during wellness (well) and on days following rhinovirus challenge (days 1-21). *S. epidermidis*/CNS was positive in 99 samples/all volunteers and *mecA* was detected in all volunteers sporadically and therefore are not included in the table.

*Not rhinovirus infected volunteers*

<b>Volunteer 2</b>	well	well	well	Day1	Day2	Day3	Day4	Day5	Day10	Day21
<i>E. aerogenes</i>	pos	pos	pos	pos	pos	pos	pos	pos	pos	pos
<i>Neisseria</i> sp.	pos									
<i>S. marcescens</i>	pos	pos	pos	pos	pos	pos	pos	pos		pos
<b>Volunteer 6</b>										
<i>S. pneumoniae</i>		pos								
<i>N. meningitidis</i>			pos							
<i>Neisseria</i> sp.	pos	pos		pos	pos	pos	pos	pos	pos	
<b>Volunteer 9*</b>										
<i>S. pneumoniae</i>									pos	pos
<i>Streptococcus</i> sp.	pos			pos						
<i>H. influenzae</i>	pos	pos	pos		pos	pos	pos	pos	pos	pos
<i>M. catarrhalis</i>	pos	pos								
<i>N. meningitidis</i>					pos			pos		
<i>Neisseria</i> sp.	pos			pos			pos			pos
<i>K. kingae</i>	pos			pos						

\*Had symptomatic illness not associated with rhinovirus infection with challenge virus.

### 2.2.5 Discussion

This study examined the bacteria present in the nasal cavity of ten young healthy adults prior to and during rhinovirus infection. Rhinovirus infection was successfully induced using self-inoculation with inoculum applied with the finger into either the nose or eye. Inoculation with larger volume ( $\frac{1}{2}$  mL) dropped into each nasal cavity was avoided in order to decrease the risk of spreading bacteria in the nasal cavity at time of inoculation.

An astonishing number of studies in the past have evaluated bacteria in the upper airways during colds by culturing bacteria on agar plates. Overgrowth of some bacteria may make less abundant pathogenic bacteria difficult to identify.<sup>125</sup> In this study, we have used microarray technology based on isolation of bacterial DNA to identify bacteria in the NLF samples. This approach allows for rapid detection of clinically relevant pathogens at the species level and the resistance marker *mecA*.<sup>123</sup> Bacterial species that are known to either be common in the upper airways or to have a pathogenic potential were included in the platform. Not surprisingly, the most common species found in this study was *Staphylococcus epidermidis*/CNS with at least one of these bacteria detected in 99 NLF samples out of the 100 total. The bacterial profile by microarray showed that samples from the same participant had a low variability, whereas inter-individual variability was high. The detection rate of bacterial sinusitis and otitis pathogens identified in nasal lavage fluid by microarray was similar to the detection rate by selective agar culture methods.<sup>6</sup>

Nasal lavages were chosen as the sampling method (rather than swabbing specific areas) in order to obtain bacteria from the entire the nose/nasopharynx. The nasal wash

sampling method has been shown to detect more potential pathogens than swabbing the nasopharynx in healthy adults.<sup>126</sup> The microarray did identify a few bacterial species that are not included in traditional culture methods. The importance of those bacterial species remains to be explored with a larger number of subjects.

During wellness, the nasopharynx and nasal vestibulum are colonized with bacteria, whereas the nasal turbinates and especially the osteomeatal complex are not colonized with pathogenic bacteria.<sup>9,10</sup> In this study, the bacterial profile in the NFL by microarray did not show any major changes following rhinovirus infection. This finding is in agreement with a prior study in adults with natural colds.<sup>127</sup> The overall bacterial titer by culture on blood agar plates also remained within one log difference in the ten nasal lavage samples obtained from each volunteer. Following rhinovirus inoculation into the eye, the virus enters the nasal cavity via the lacrimal duct and is transported back to the nasopharynx where it initially can be detected in the majority of subjects. Later, the rhinovirus can also be detected on the nasal turbinates, suggesting that the rhinovirus spread.<sup>128</sup> Similarly, potential pathogenic bacteria from the nasopharynx have been detected in the osteomeatal complex of the nasal cavity during natural colds but not during wellness.<sup>9</sup> Bacteria transient in the nose are removed by the mucociliary clearance mechanism during wellness but during colds the mucociliary clearance function is decreased for 3-4 weeks.<sup>129-131</sup> A longer transit time of the bacteria in the nose may increase the risk of further displacement into the paranasal sinuses and/or the middle ear cavity by nose-blowing.<sup>132-134</sup> Kaiser *et al.* showed a benefit of antibiotic treatment in the subgroup of patients with early stage common colds with pathogenic bacteria in the nose/nasopharynx by culture method.<sup>135</sup> It would be of great interest to know whether displacement of potentially pathogenic bacterial to the nose during colds

would increase the overall morbidity of colds and/or affect the severity of nasal symptoms. If so, treatment of colds should attempt to decrease spread of virus and bacteria into the nasal cavity by a combination of several different approaches such as 1) an antiviral and antibacterial nasal spray, 2) decrease of mucus production, and 3) enhancement of nasopharyngeal clearance by sniffing rather than nose blowing.

In conclusion, we present an experimental clinical model to study possible effects of bacteria during viral respiratory tract infection. We successfully infected healthy adults with rhinovirus by self-inoculation into the eye or nose and demonstrated that microarray is an excellent way to detect more bacterial species than traditional bacterial culture methods. We used a commercially available microarray platform for rapid detection of clinically relevant bacteria. In the future, a microarray platform designed using 16S rDNA sequence data or RNASeq data results of a cross-section of samples from normal children and adults and patients with upper respiratory diseases should be used. In this pilot study, each of 10 volunteers had a unique bacterial profile as evidenced by microarray and the bacterial load did not change during viral infection. Larger scale studies are needed to assess definitively whether bacterial load or species change during rhinovirus colds.

## **Chapter 3**

### **Nasopharyngeal microbiome during both health and viral infection**

#### **3.1 Chapter Introduction**

In our previous study of the bacterial species in the nasopharynx of healthy adults during health and viral infection, we did not find changes in the species detected following the presence of a viral infection.<sup>136</sup> The clinical diagnostic microarray tested 27 individual species-level probes, along with four group probes that detect six, three, 13, and seven species respectively, and one additional group probe that detects eight serovars. These results may be due to the fact that we only tested for a panel of approx. 60 species, or that we were unable to measure expansion or reduction of species because detection of species was not quantitative. In order to determine relative abundance of all bacteria present, we sequenced the hypervariable 1 and 2 regions of the 16S rRNA gene in each sample from the previous study described in Chapter 2. This study combines the innovative HRV challenge model and sample collection time course before, during, and after inoculation with the ability to investigate the microbiome using high-throughput sequencing methods.

Our goal was to better understand microbial changes in the nasopharynx due to a viral infection in otherwise healthy individuals. Future studies can harness the same methodology to sample the nasopharyngeal microbiome in a larger population, providing greater power to determine which bacteria respond, expand, and disappear in response to HRV infection. These findings may lead to the development of more specific clinical diagnostics, enabling a better, more personalized approach to treatment of these diseases. This, in turn, could reduce the use of broad spectrum antibiotics prescribed for URTI's, potentially also decreasing antibiotic resistance.

### 3.2 Characterization of the nasopharyngeal microbiome in health and during rhinovirus challenge

*Manuscript under review at Microbiome (originally submitted November 11, 2013, revision submitted February 28, 2014)*

E. Kaitlynn Allen<sup>1,2</sup>, Alex Koeppel<sup>3</sup>, J. Owen Hendley<sup>4</sup>, Stephen D. Turner<sup>3</sup>, Birgit Winther<sup>5</sup>, Michèle M. Sale<sup>1,2,6</sup>

<sup>1</sup>Center for Public Health Genomics, <sup>2</sup>Department of Biochemistry and Molecular Genetics, <sup>3</sup>Bioinformatics Core Facility, <sup>4</sup>Department of Pediatrics, <sup>5</sup> Department of Otolaryngology, <sup>6</sup>Department of Medicine, University of Virginia, USA.

E. Kaitlynn Allen carried out molecular genetic studies (DNA extraction, PCR amplification, sequencing preparation), participated in data analysis, and drafted the manuscript. Dr. Alex Koeppel carried out data analysis and statistical analyses, and helped in drafting the manuscript. Drs. Birgit Winther and J. Owen Hendley conceived and designed the challenge study, and Dr. Birgit Winther coordinated study recruitment, and carried out sample processing and storage. Dr. Stephen D. Turner participated in data analysis and statistical analyses. Dr. Michèle M. Sale conceived of the microbiome study and coordinated molecular genetic studies. All authors contributed to data interpretation, and read and approved the final manuscript.

#### 3.2.1 Abstract

**Background:** The human microbiome might play an important role in host immune response to various microbes. The microbiome of the nasopharynx is a bacterial

community responsible for numerous upper respiratory tract infections (URTIs).

Though this microbiome has been surveyed before, it has not yet been surveyed during a known, controlled viral challenge. Understanding the microbial composition of the nasopharynx during viral infection will help in understanding better treatments of URTIs including the common cold.

**Methods:** Rhinovirus illnesses were induced by self-inoculation using the finger to nose or eye natural transmission route in ten otherwise healthy young adults. NLF samples were collected at specific time points before, during, and following experimental rhinovirus inoculation. Bacterial DNA from each sample (N=97 from 10 subjects) was subjected to 16S rRNA sequencing by amplifying the V1-V2 hypervariable region followed by sequencing using the 454-FLX platform.

**Results:** This survey of the nasopharyngeal microbiome revealed a highly complex microbial ecosystem. Taxonomic composition varied widely between subjects and between time points of the same subject. We also observed significantly higher diversity in not infected individuals compared to infected individuals. Principal coordinates analysis (PCoA) demonstrated that the microbiome of the nasopharynx clusters with similar environments including the nostril, external auditory canal, adenoid, and hair.

**Conclusions:** Our results demonstrate a highly complex microbiome of the healthy and viral-challenged adult nasopharynx. Each subject has a unique microbiome, though there are phyla that occur in most samples including Firmicutes, Actinobacteria, and Proteobacteria. Deeper metagenomic sequencing of the microbiome of more individuals at the species level could result in detection of significant community changes due to viral infection.



### 3.2.2 Introduction

Upper respiratory tract (URT) infections are polymicrobial, with complex interactions of bacterial pathogens and viruses in the URT, both synergistic and competitive associations, playing a large role in the pathogenesis of the disease.<sup>137-140</sup> Like other areas of the body, when in a balanced state, this microbiome is beneficial for the health of the host via interactions with the immune system and protection against invading pathogenic bacteria.<sup>141-143</sup> Bacteria in the URT must colonize the nasopharynx before being able to cause disease.<sup>144</sup> The nasopharynx is normally colonized by both commensal and pathogenic bacteria, with the latter being able to cause different infections including bacterial sinusitis and otitis media.<sup>6</sup> Many pathogenic species, including *Streptococcus pneumoniae*, *Haemophilus influenza*, *Moraxella catarrhalis*, *Staphylococcus aureus*, and *Neisseria meningitidis*, exist in the nasopharynx of apparently healthy individuals.<sup>6,7,11,145</sup> Prior studies have examined the nasopharyngeal microbiome in healthy and diseased individuals. Ling *et al.* demonstrated that the nasopharyngeal microbiome was distinct from other body sites surveyed (saliva, dominant hands, and feces) in healthy Chinese adults.<sup>146</sup> Studies of the nasopharyngeal microbiome in children show that there are changes due to season (winter/fall versus spring)<sup>145</sup>, and treatment with antimicrobials or the heptavalent conjugated pneumococcal polysaccharide vaccine.<sup>147,148</sup> Also, a study of the nasopharyngeal microbiome of children with severe bronchitis shows that the microbiome shifts depending on the viral infection (HRV only, RSV only, or co-infection)<sup>149</sup>, demonstrating the importance of examining changes in both the bacterial and viral communities during disease. Previously, we surveyed the microbiome of the adenoid in children undergoing adenoidectomies for various diseases, including chronic otitis media, hypertrophic adenoids, chronic adenoiditis, and obstructive sleep apnea.<sup>22</sup> Results of our study showed a complex, highly diverse microbiome including known pathogenic bacteria.

Given the high occurrence of viral infection preceding various diseases including otitis media and sinusitis, it is important to understand how the interaction of the microbiome and virus changes the host's risk of disease.<sup>150</sup> For example, viral infection of the URT is a predisposing risk that is thought to allow for pathogenic bacteria in the nasopharynx to induce inflammation and permit microbial invasion of the mucosa in the middle ear causing otitis media.<sup>37</sup> Like pathogenic bacteria, viruses including rhinoviruses, enteroviruses, coronaviruses, and adenoviruses have been found in asymptomatic, healthy individuals.<sup>151,152</sup> Mechanisms by which viruses in the URT can alter bacterial adherence<sup>153</sup>, bacterial colonization<sup>154</sup>, host immune response<sup>155-157</sup> have been reviewed.<sup>144,158</sup> The disease model of a virus predisposing the host to a secondary bacterial infection is thought to be an oversimplification, with a much more complex dynamic being more likely.<sup>139</sup> Understanding the relationships between microbes of the URT, especially during perturbations, would give us insight into the pathogenesis of URT infections. Mice with a perturbed microbiome due to antibiotic treatment develop an allergic reaction to airway allergens, while mice with an intact endogenous microbiome do not.<sup>142,159</sup> The results of this study provide evidence that the gut microbiome is important in the host's response to upper respiratory disease. Another study identified specific commensal taxa that were negatively associated with colonization of known pathogenic bacteria and with acute otitis media, and these relationships changed depending on antibiotic usage of the children.<sup>55</sup> Also, a study on the lung microbiome in COPD patients and healthy individuals during a rhinovirus challenge showed a significant change in the microbiome in COPD patients after infection<sup>160</sup>, suggesting rhinovirus infection may give rise to secondary bacterial infections. Understanding the baseline diversity and any potential changes in the microbial community in the nasopharynx during viral infection can elucidate the role of bacteria in the pathogenesis

of URT infections. This knowledge could provide novel insight into URT infections and the common cold.

The goal of our study was to characterize the bacterial community found in the adult nasopharynx prior to, during, and after rhinovirus infections. In a previous study, we used the same samples to detect bacteria in the nasopharynx using microarray technology.<sup>136</sup> The microarray allowed us to detect bacteria at the species level accurately, but we were only able to test presence or absence of about 60 bacteria clinically informative for a range of diseases. In order to characterize the relative abundance of microbial communities of the nasopharynx at the phylum and genus levels, we carried out sequencing of the 16S rRNA gene of bacteria in nasal lavage fluid (NLF) samples from these adults during rhinovirus challenge.

### **3.2.3 Methods**

#### *3.2.3.1 Sample Collection*

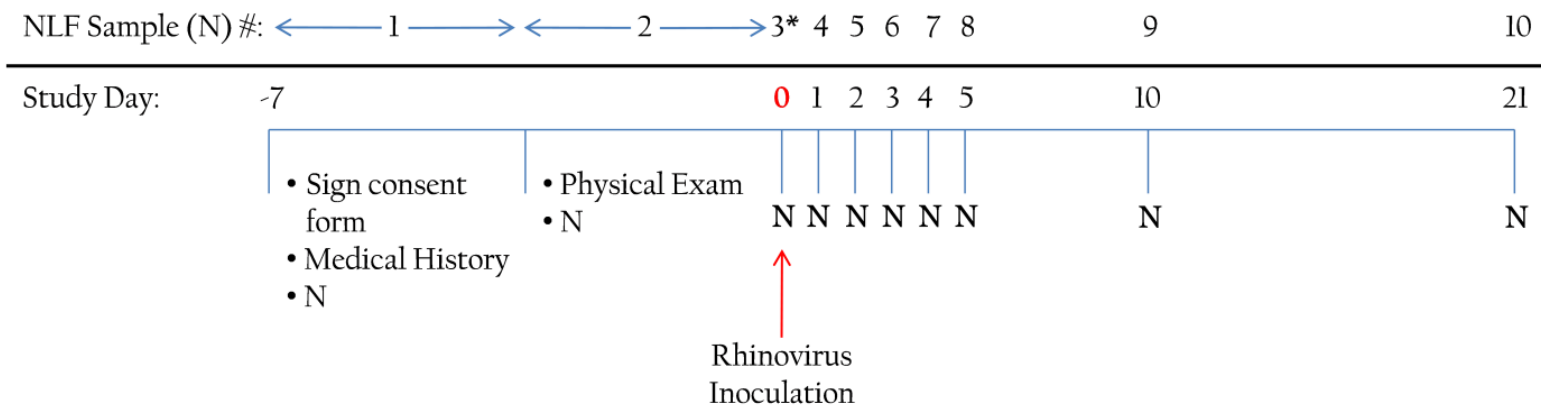
This protocol has been previously described (also see Chapter 2, section 2.2.3).<sup>136</sup> Briefly, participants were volunteers between the ages of 18 and 65 that responded to advertisements in 2010. Subjects were eligible if they had a screening serum neutralizing antibody of 1:4 or less to the challenge Rhinovirus type 39. The study was approved by the Institutional Review Board for Health Sciences Research at the University of Virginia. Informed consents were obtained from all ten participants prior to enrollment. The Rhinovirus immunotype 39 inoculum pool has been safety tested and approved for use by the FDA (IND12934). This study was conducted with all sample collection during Fall 2010. Subjects were exposed to a total of 100-300 TCID<sub>50</sub> of Rhinovirus in 250µl by self-inoculation by touching the medial cantus and conjunctiva of

one eye or the septum in the nasal vestibulum on one side of the nose, and repeated once after a 5-15 minute period. NLF was obtained by installation of 5mL of 0.9% saline into each nasal cavity which was recovered into a waxed paper cup. The saline is able to reach throughout the nasopharynx due to the position of the participant's head (tilted backwards). Ten nasal washes were obtained from each volunteer: three during the week prior to HRV inoculation, one on each of five days immediately following inoculation, and on days 10 and 21 following inoculation (**Figure 3.1**). An aliquot of 1mL of each NLF sample was placed in tubes containing an equal volume of viral collection broth, kept on ice, and transported to the laboratory within one hour for rhinovirus isolation in tissue cultures. The remaining NLF was transferred to 2mL cryo-tubes and stored frozen at -80°C.

### *3.2.3.2 Assessment of infection*

This protocol has previously been described (see Chapter 2.2.3).<sup>136</sup> Briefly, a subject was considered *infected* if the challenge rhinovirus was detected in the NLF at least once during the five days following inoculation. Additionally, the serum antibody response to the challenge rhinovirus type 39 was examined in sera obtained prior to inoculation and three weeks following inoculation by standard methods.<sup>121</sup> Subjects with a 4-fold increase in antibody titer to the challenge virus were considered to be infected. If either of these tests were positive, the subject was considered infected. Assessment of illness of the participants is described in Figshare (<http://dx.doi.org/10.6084/m9.figshare.832471>) **NOTE:** Figshare documents have been updated after journal reviews and contain the latest data) Additional file 1 (see Chapter 3 Supplement) and Chapter 2, section 2.2.3.4.

**Figure 3.1:** Study timeline including when information or samples were collected from study participants. \*Day which NLF samples were obtained prior to Rhinovirus inoculation.



### 3.2.3.3 DNA extraction

DNA was isolated from NLF using bead beating and phenol-chloroform extraction methods. This protocol<sup>22</sup> was optimized from Ren *et al.* to increase DNA output from a more difficult sample type (NLF). Briefly, 1.8 mL of NLF was transferred to a 2 mL BeadBeater tube (BioSpec) and centrifuged at 2348 *g* (5000 rpm) for 10 minutes. Supernatant (1.3 mL) was aspirated and 20 µL Proteinase K (20 mg/mL) (Roche) was added. The sample was incubated at 60°C for two hours, then 95°C for 10 minutes. A total of 200 µL 10% SDS (Fischer Scientific), 400 µL of 0.1 mm zirconia/silica beads (BioSpec), 100 µL of buffer (10 mM Tris-HCl pH 7.5, 5 mM EDTA, 100 mM NaCl), and 500 µL of phenol:chloroform:isoamyl alcohol (25:24:1) (Fischer Scientific) were added to the tube. The sample was homogenized on a Biospec Mini-BeadBeater-8 for three minutes and centrifuged for 10 minutes at 10,000 *g*. The top aqueous layer was transferred to a new Phase-Lock-Gel tube (VWR) with an equal volume of chloroform:isoamyl alcohol (24:1) (Acros Organics) and centrifuged at 20,000 *g* for 10 minutes. This step was repeated twice to ensure removal of phenol. The sample was ethanol precipitated, and the DNA pellet resuspended in 50 µL 10 mM Tris-HCl pH 8.0. Each set of extractions was accompanied by a negative control to ensure no contaminants in the reagents.

### 3.2.3.4 Tag-PCR amplification of the V1-V2 regions of the bacterial 16S rRNA gene and pyrosequencing

This protocol has been previously described.<sup>22</sup> Briefly, the V1-V2 hypervariable regions of the 16S rRNA gene were amplified from extracted DNA samples using two primers containing the universal sequences 27F (5'-AGRGTTTGATCMTGGCTCAG-3') and 534R

(5'-TTACCGCGGCTGCTGGCAC-3'). Each sample was tagged with a unique barcode (10 bp) added to the 5' end of the forward primer. Conditions for amplification were 94 °C for three minutes, then 30 cycles of 94 °C for 30 seconds, 57 °C for 45 seconds, and 72 °C for 60 seconds with a final extension step of 72 °C for five minutes. Each set of samples was amplified with a negative control to ensure no contaminants from reagents. PCR products were run on a gel to verify amplification of the correct size. Amplicons from all samples were quantified using a Qubit 2.0 Fluorometer and pooled in equal molar ratios, with 49 samples in Pool 1 and 48 samples in Pool 2. The pooled samples were gel purified, and each pooled sample was sequenced using Titanium chemistry on a 454 Life Science Genome Sequencer FLX platform at the University of Virginia Department of Biology Genome Core Facility using the standard protocol for sequencing. Before sequencing, the samples were run on an Agilent 2100 Bioanalyzer High-Sensitivity chip to determine the size distribution of the library samples. Then, a KAPA Biosystems qPCR assay was run to determine the effective concentration of the samples.

### *3.2.3.5 Quality assessment and filtering of sequences*

The quality of the read sequences was assessed using FastQC v0.10.1.<sup>161</sup> This assessment revealed that a significant fraction of our reads were identical short (<100bp) reads (homopolymers). Based on this assessment, and the distribution of read lengths, we filtered the sequences by length, removing any reads shorter than 100bp in length, or greater than 600bp. This filter alone reduced our initial dataset from 4.32 Mbp, to 3.32 Mbp (~23.1% reduction). In addition, reads with an average quality score of less than 25 (phred scale) were removed, as were any reads that contained 50-base windows with average quality scores below that threshold. Qiime's default filters also removed any

reads with mismatches in the primer or barcode sequences, and any reads with more than 6 ambiguous bases.

### *3.2.3.6 Clustering & Filtering of OTUs (Operational Taxonomic Units)*

Sequences were clustered into operational taxonomic units (OTUs) (97% identity cutoff) using the uclust algorithm v1.2.22 (*de novo* clustering)<sup>162</sup> in QIIME v1.8<sup>163</sup>, and assigned taxonomy using RDP classifier v2.2<sup>151</sup>, trained with the Greengenes (gg\_13\_8)<sup>164</sup> dataset. OTUs were removed from the dataset if they could not be classified at the domain level, or were classified as archaea, or were classified as chloroplast sequences. OTUs estimated to be putatively chimeric by the uchime algorithm v4.2.40<sup>165</sup> were also filtered out, as were singleton OTUs. After filtering, we were left with a mean of 1021.6 sequences per sample (std. dev: 630.84, median: 913, min: 134, max: 4501). Samples containing fewer sequences than three standard deviations below the mean count (146 sequences/sample) (n=1) were eliminated from downstream analyses.

### *3.2.3.7 Sequence Analysis*

Analyses of taxonomic composition and diversity were performed using QIIME, version 1.8.<sup>163</sup> Default parameters were used for the core analyses. OTU representatives were aligned with pyNAST v1.2.2<sup>166</sup>, and phylogenies were generated using FastTree v2.1.3<sup>167</sup>. Alpha diversity (within-population diversity) of samples was measured using counts of observed species (OTUs), the chao1 estimator for species richness, and the Shannon diversity index, which estimates total diversity taking into account both species richness and evenness. Beta diversity (diversity between populations) was calculated using UniFrac distances (unweighted)<sup>168</sup> between samples (based on the relative abundance of



OTUs), and visualized using Principal Coordinates Analysis (PCoA)<sup>169</sup>. In order to compare diversity of OTUs in the nasopharynx to that of other body sites, we also performed a combined beta diversity analysis of our nasopharyngeal samples with samples from several different body sites (gut, oral cavity, external auditory canal, nostril, and hair) from Costello *et al.*<sup>170</sup>, as well as adenoid samples from Ren *et al.*<sup>22</sup>.

### 3.2.3.8 Statistical Analysis

Analyses of phylum abundance between inoculation states (before, during, and after) and between subjects, was performed using the Tukey Kramer test. The Tukey Kramer test was also used for the analyses of class, order, and family abundances between inoculation states (before, during, and after). Comparisons of alpha diversity between the infected and non-infected sample groups were done using Mann-Whitney Wilcoxon tests, as were the abundances of selected genera between the infected and non-infected samples, and the comparison of inter-individual and intra-individual Unifrac distances.

## 3.2.4 Results

### 3.2.4.1 Subject population

Ten volunteers were eligible and enrolled into the study. Six of the subjects self-inoculated into the nose and four self-inoculated into the eye. Seven out of the ten subjects were infected by the Rhinovirus challenge. Out of the infected subjects, three were considered ill based on the modified Jackson Cold method (Figshare Additional file 1). All subjects were sampled at all ten time points. Demographic information and information on infection status is included in **Table 3.1**.

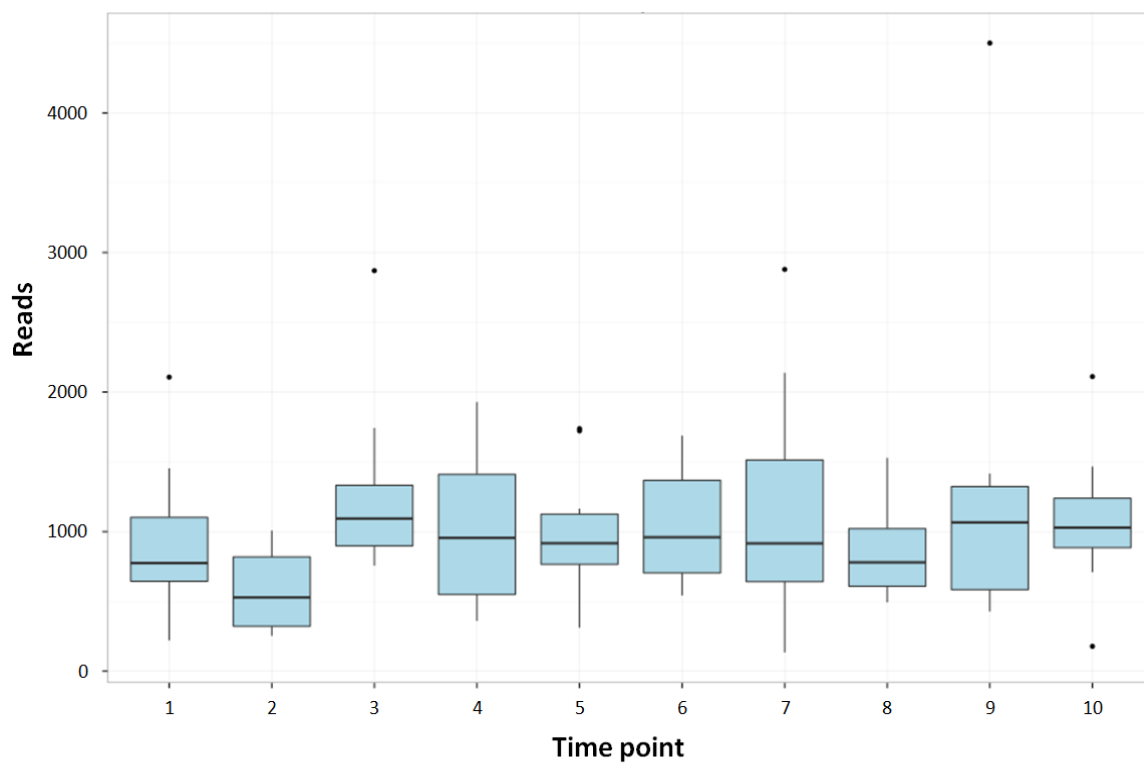
**Table 3.1** Subject information

<b>Subject</b>	1	2	3	4	5	6	7	8	9	10
<b>Age</b>	19	19	18	18	20	21	21	20	19	21
<b>Sex</b>	F	F	F	M	F	M	M	M	M	M
<b>Race</b>	White*	White	White	White	White, Asian	White	White	White	African American	White
<b>Route of inoculation</b>	Eye	Nose	Nose	Nose	Nose	Eye	Eye	Eye	Nose	Nose
<b>HRV infection</b>	yes	no	yes	yes	yes	no	yes	yes	no	yes
<b>Criteria for infection status</b>	Ab	-	HRV/Ab	HRV	HRV/Ab	-	HRV/Ab	Ab	-	HRV/Ab

For Race, “\*” indicates the subject did *not* specify “Non-hispanic.” For Criteria for infection status, “HRV” indicates HRV shedding in at least one sample during timepoints 4-8 (during infection period), and “Ab” indicates that the subject had a 4-fold rise in serum-neutralizing antibody in response to the challenge virus.

DNA was successfully extracted from 99/100 NLF samples, with concentration values ranging from 5.86ng/ $\mu$ L-33.8ng/ $\mu$ L which is above the recommended minimum threshold for microbiome studies.<sup>171</sup> One sample (Subject 6, Day 3) did not have enough NLF for DNA extraction. PCR amplification and pyrosequencing of the V1-V2 hypervariable regions of the bacterial 16S rRNA gene was successful in 97 out of the 99 NLF samples. PCR amplification was unsuccessful in two samples (Subject 3, Day 2 and Subject 10, Day 3). Sequencing of the two pools of DNA extracted from the NLF samples resulted in 679,135 reads. These reads were quality filtered to select the most reliable reads for analysis (see Methods section 3.2.3.5), resulting in 383,969 reads with a median of 913 reads per sample (**Figure 3.2**). After filtering, reads had a mean length of 376 bp (median 414 bp), and one sample (Subject 10, Day 7 (134 reads)) was excluded for containing fewer sequences than three standard deviations below the mean count (146 sequences/sample).

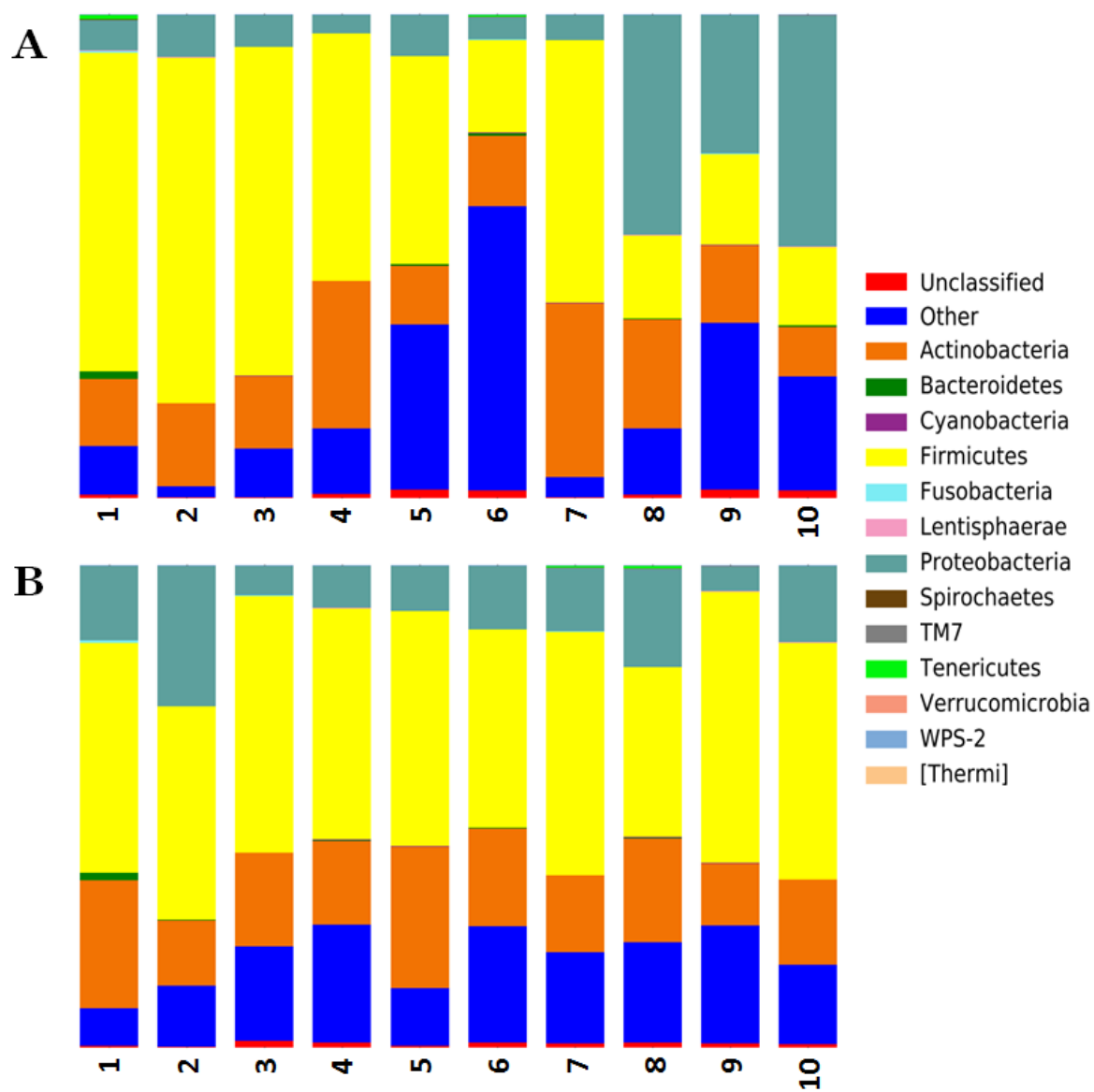
**Figure 3.2:** Box and whisker plot of read counts at each time point during the study after quality control filtering.



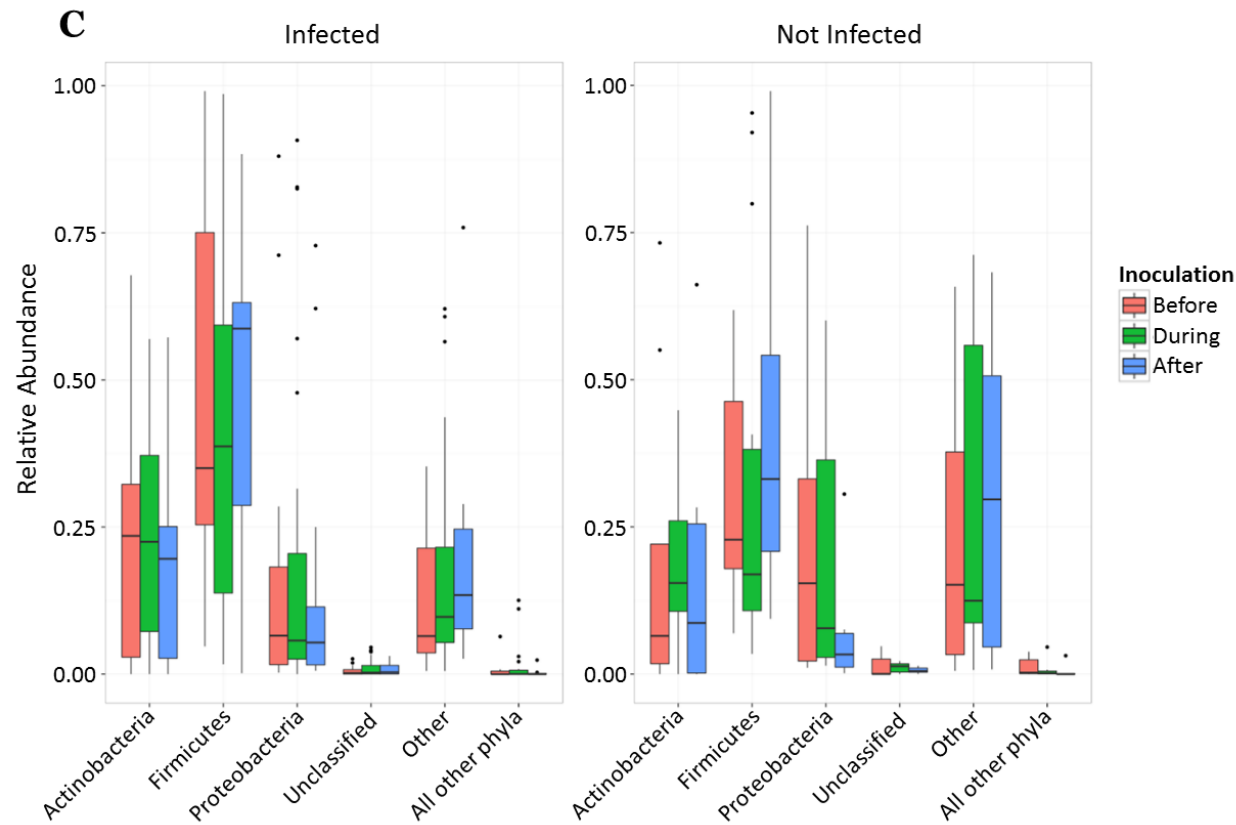
### 3.2.4.2 Analysis of nasopharyngeal bacterial richness and composition by sequencing

The sequences were clustered into 3229 distinct OTUs (97% identity cutoff). Singleton OTUs were filtered out, leaving 1608 OTUs for downstream analysis. Across all ten subjects, the dominant phyla were Firmicutes (40.6%), Actinobacteria (20.9%), Proteobacteria (17.1%), and unclassified (20.6%), with the relative abundance of each phylum being highly variable across the subjects (**Figure 3.3A**; Figshare Additional Files 2 & 3, see Chapter 3 Supplement) and time points (**Figure 3.3B**; Figshare Additional Files 2 & 4). To determine if bacterial composition changes occur due to viral infection, we compared the relative abundance of the dominant phyla before (time points 1-3), during (time points 4-8), and after (time points 9-10) Rhinovirus infection in infected and not infection subjects (**Figure 3.3C**). No phylum showed a significant change in relative abundance between time points. Additionally, we have conducted a Tukey-Kramer test which revealed no significant differences in the abundance of the three dominant phyla before, during, or after inoculation, in either the infected or uninfected categories (Figshare Additional File 5). No significant differences were found in the abundances of dominant classes, orders, or families before, during, or after inoculation, in either infected or not infected individuals (all  $P$ -values  $>0.11$ ). There are some large changes in relative abundances of phyla within individuals at close time points. It will be important to investigate whether these reflect true biological fluctuations in this dynamic ecological niche or low microbial density via greater sequencing depth. A Tukey-Kramer test was performed to test the null hypothesis that the relative abundances of the dominant phyla were the same between all pairs of subjects (Figshare Additional File 6). Significant differences in relative abundances of Proteobacteria were found in Subjects 8 and 10 compared to Subjects 1, 3, 4, 5, 6, and 7. Subject 8 had significantly different relative abundances of Firmicutes than Subjects 1 and 3, and Subject 9 also differed from Subjects 1, 2, and 3 in Firmicutes relative

**Figure 3.3:** Taxonomic bar chart (A) showing the relative abundance of phyla in each sample and (B) showing the relative abundance of phyla in each time point. Time points 1-3 represent before viral inoculation, time points 4-8 represent during infection, and time points 9-10 represent after infection.



**Figure 3.3C:** Box and whisker plot of the relative abundance of dominant phyla in infected and not infected individuals by time period. Pink = before inoculation, green = during inoculation, and blue = after inoculation. Sample counts for Infected samples (left panel): before N=19, during N=34, after N=14. Sample counts for Non-Infected samples (right panel): before N=8, during N=15, after N=6. Phyla that were classified but whose abundances were <1% are represented by the classification “All other phyla”.



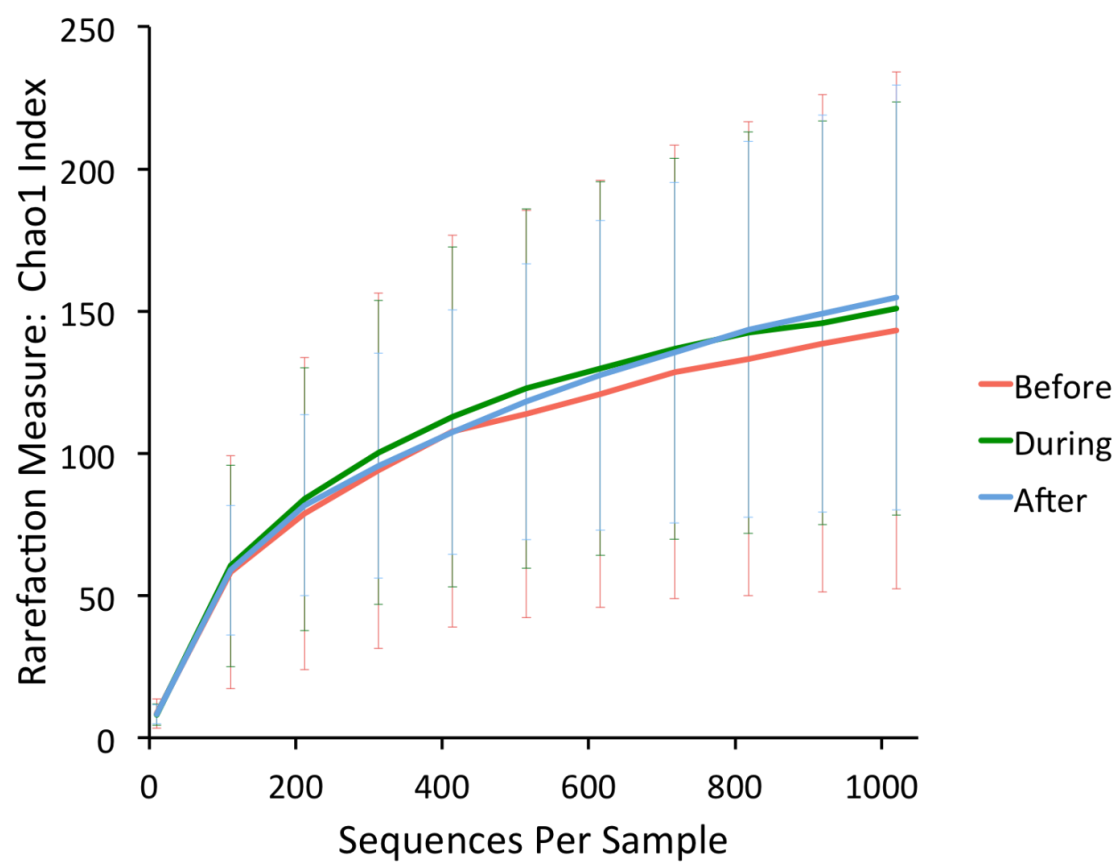
abundance. Subject 10 showed a significantly different relative abundance of Actinobacteria than Subject 7. These differences can be visualized using **Figures 3.3A** and **3.3B**. However, for the majority of subjects, the abundances of the dominant phyla were not significantly different from one another.

We also carried out a rarefaction analysis to compare bacterial species richness between the samples before, during, and after infection by plotting the rarefaction measure (the Chao1 Diversity index for OTUs at 97% identity threshold) by the number of sequences per sample (**Figure 3.4**; Figshare Additional Files 7-8). The rarefaction plot curves indicate that the nasopharynx, sampled by NLF, is a very complex microbial environment. Deeper sequencing could reveal even more diversity in this very complex microbiome. Additionally, there is no significant difference between the species richness between the three time points, though there is a trend for more bacterial richness in the time points during and after inoculation compared with the time points before inoculation.

The most abundant genus in the nasopharyngeal microbiome across samples was *Alloiococcus* (21.2% of total abundance), followed by *Corynebacterium* (14.7%), *Staphylococcus* (8.9%), *Haemophilus* (5.3%), *Propionibacterium* (4.6%), and *Streptococcus* (3.2%). All other classified OTUs belonged to genera comprising less than 2% of the total abundance. It should be noted that these abundances reflect total abundance across all samples. Taxonomic composition varied widely between samples (Figshare Additional File 2).

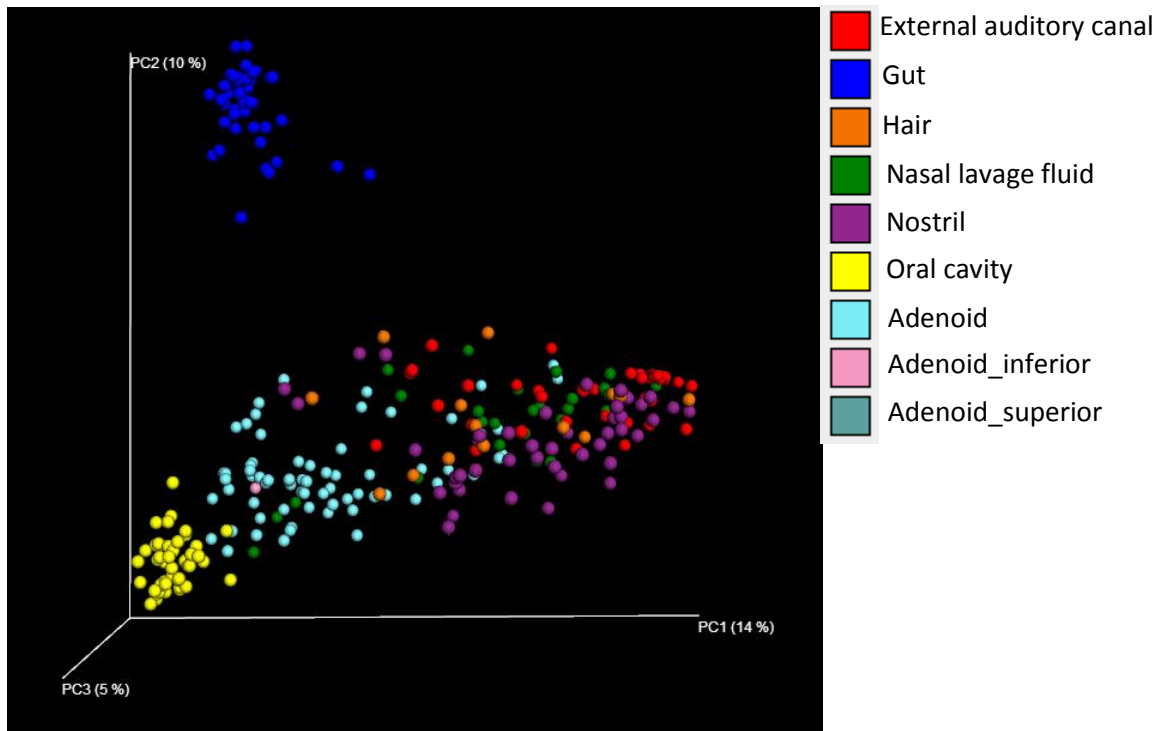


**Figure 3.4:** Rarefaction plots displaying the relationship between the sample size and the number of observed species.

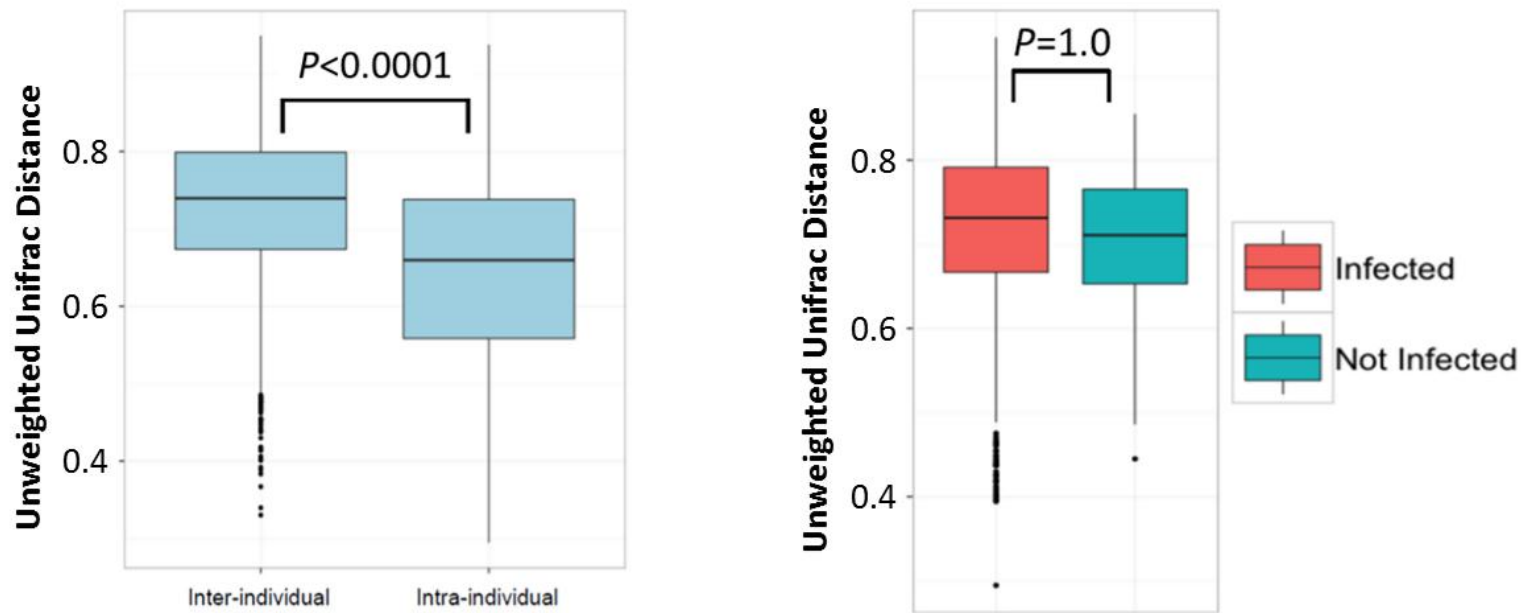


We compared the microbiome of the nasopharynx (NLF samples) to those of other body sites including the gut, oral cavity, external auditory canal, nostril, hair, and adenoid using PCoA (**Figure 3.5**; Figshare Additional Files 9-10). For the purpose of comparison, only the NLF samples taken prior to inoculation were included for this analysis. The microbiome of the nasopharynx is most similar to the external auditory canal, nostril, hair, and adenoid. The gut and oral cavity are distinct from the rest of the samples. When we used PCoA to compare the NLF samples before, during, and after infection; infected versus not infected; ill versus not ill; and by route of infection, we saw no distinct differences in the samples (Figshare Additional File 11-12). When looking at the PCoA to compare NLF samples between each subject (Figshare Additional File 11), there is no distinct clustering by subject. When comparing NLF samples within each subject (Figshare Additional File 12), there is also no distinct clustering. Unifrac distances between samples from different individuals is significantly greater than the distances between samples from the same individual (**Figure 3.6A**). Inter-individual Unifrac distances were not significantly different between infected and not infected subjects (**Figure 3.6B**).

**Figure 3.5:** Nasopharyngeal microbiome is similar to neighboring body parts, but distinct from gut and oral cavity. Principal coordinates analysis (PCoA) was performed using the unweighted UniFrac distance matrix using nasal lavage fluid (NLF) samples from pre-inoculation time points. Each sample is represented by a point with NLF (N=27) in green, gut (N=45) in dark blue, oral cavity (N=46) in yellow, external auditory canal (N=44) in red, nostril (N=46) in purple, hair (N=14) in orange, and adenoids (N=69) in light blue. “Adenoid\_inferior” and “Adenoid\_superior” refer to a sample from the inferior and superior areas of one adenoid for comparison.



**Figure 3.6:** Unifrac distance comparisons: **(A)** Unweighted Unifrac distances between samples from different individuals (inter-individual) and the distances between samples from the same individual (intra-individual); **(B)** Inter-individual Unweighted Unifrac distances compared between infected and not infected subjects. *P*-values displayed are the results of Mann-Whitney Wilcoxon tests.



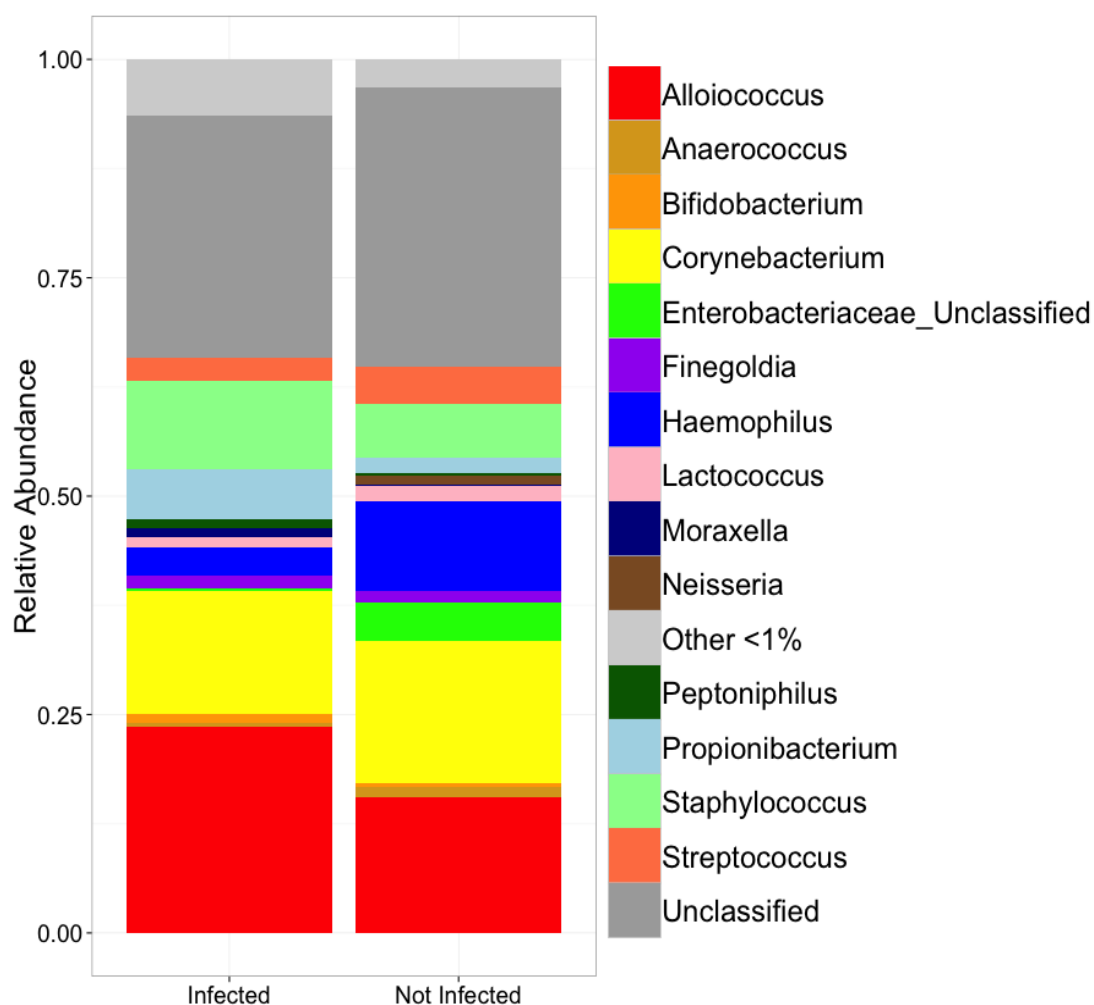
### 3.2.4.3 Bacterial community structure of infected and not infected subjects

The community structure of bacteria in the NLF samples of infected (N=7) and not infected (N=3) subjects were compared at the genus level. Both groups showed similar patterns of genera distribution (**Figure 3.7A**), but there were significant differences ( $P=0.031$ ) in the overall community diversity (Chao1 Index) between infected and not infected subjects (**Figure 3.7B**). A Mann-Whitney Wilcoxon test was used to compare the relative abundances between infected and not infected samples of all genera comprising 1% or more of the total abundance across samples. Significant differences were detected in two genera (*Neisseria* and *Propionibacterium*) after using the Benjamini & Hochberg False Discovery Rate correction for multiple comparisons (**Table 3.2**). Prior to correction for multiple comparisons, the genus *Haemophilus* was significantly less abundant in the infected samples, suggesting a potential association which could be further investigated with additional samples.

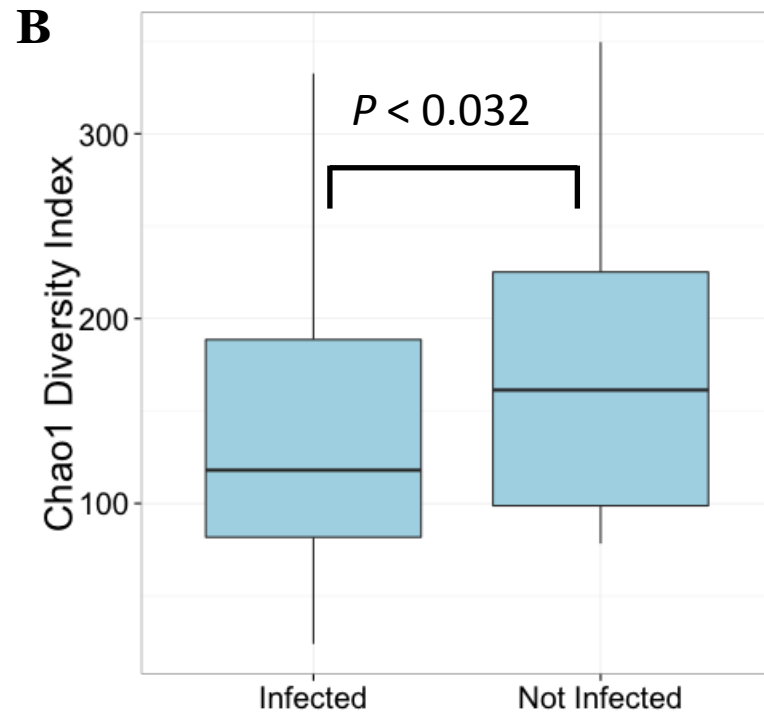
**Figure 3.7A:** Stacked taxonomic bar chart of Not infected (N=29) vs. Infected (n=67).

OTUs were categorized as “Unclassified” if RDP classifier could not classify them to at least the family level. Genera that could be classified to the family level, but not the genus level were categorized by the family name, followed by “\_Unclassified” to indicate that the genus is not known. All genera comprising less than 1% of the total abundance (whether classified or not) were combined into the “Other <1%” category.

**A**



**Figure 3.7B:** Box and whisker plot- Shannon diversity index of Not infected (N=29) vs. Infected (N=67).



**Table 3.2** Comparisons of genus abundance between infected and uninfected samples, using the Mann-Whitney Wilcoxon test.

*P*-values <0.05 (with & without FDR correction) are highlighted.

Genus	Mean Relative Abundance			P-value	P-value adj
	Infected	Not Infected	W		
Alloiococcus	0.237	0.155	1198	0.071	0.227
Anaerococcus	0.004	0.013	801	0.137	0.274
Bifidobacterium	0.01	0.004	975.5	0.969	0.97
Corynebacterium	0.141	0.162	946	0.842	0.97
Enterobacteriaceae_Unclassified	0.003	0.043	864	0.343	0.534
Fingoldia	0.015	0.013	1166	0.113	0.263
Haemophilus	0.032	0.104	709	0.02	0.095
Lactococcus	0.013	0.016	898.5	0.227	0.398
Moraxella	0.009	0.002	976	0.97	0.97
Neisseria	0.001	0.011	667	0.001	0.02
Peptoniphilus	0.009	0.003	1176.5	0.081	0.227
Propionibacterium	0.058	0.018	1315.5	0.006	0.042
Staphylococcus	0.101	0.061	964.5	0.959	0.97
Streptococcaceae_Unclassified	0.027	0.043	871.5	0.417	0.583



### 3.2.5 Discussion

This study is the first to survey and describe the microbiome of the nasopharynx in healthy adults during a Rhinovirus challenge. Strengths of our study include a controlled rhinovirus study, collection of clinical information, and use of high-throughput sequencing technology. Our study offers the advantage of using a known, controlled rhinovirus challenge to examine changes in the bacterial community before, during, and after inoculation in the same individual. This is important as inter-individual variability has been shown to be very high<sup>172</sup>; therefore, comparing healthy individuals to individuals with disease is not as conclusive as within-individual comparisons. Additionally, results from a known, controlled viral challenge are more interpretable than surveys in populations with naturally occurring viral infections. We discovered that the microbiome of the nasopharynx is highly complex and observed trends in the differences in the microbial community during viral infection.

The results from this survey have been compared to the results of the microarray study<sup>136</sup> using the same samples. Due to the design of the microarray study, it can only detect the presence or absence of bacterial species on the panel included, while the 16S method is able to determine relative abundance of all bacteria in the sample. We can only compared the number of mismatches in those genera that the microarray was able to detect (N=8). Out of the 96 samples total, the percent of samples which were detected in one study but not the other ranged from 4%-53% in each genus. These results could be due to the use of two aliquots from the same NLF sample, extracted using two different methods. It might also be due to the microarray detecting species in very low abundance because of enrichment by PCR.

Due to the level of complexity of the nasopharyngeal microbiome, it is possible that significant changes in the bacterial communities before, during, and after infection could likely be detected with deeper sequencing and/or recruiting more subjects. Additionally, using true shotgun metagenomics and taxonomic classification<sup>173</sup> to distinguish bacteria at the species level may allow significant changes in pathogenic and opportunistic bacteria during infection to be discovered. Information at the species level is especially important because, in numerous cases, there are pathogenic and commensal bacteria within the same genus, e.g. *Haemophilus influenzae*, a known pathogen, and its closest phylogenetic relative *Haemophilus haemolyticus*, a commensal bacterium.<sup>174,175</sup>

PCoA of the nasopharyngeal microbiome samples along with samples from various other body habitats showed logical clustering with similar locations throughout the URT. The nasopharynx, adenoid, ear, hair, and nostril samples all formed a single large cluster. While this cluster was distinct from the separate clusters formed by the gut and oral samples, there was not a clear differentiation between the other body sites. Bacterial niches in close proximity have some sharing of bacteria due to physiological events like sneezing and nose blowing<sup>132-134</sup>, which could help to explain the overlap between the microbiome of the nasopharynx and other proximal body habitats. Differences in age and health status of subjects, along with season of sample collection, may impact results. Participants in the Costello *et al.* study included healthy adults, with two samples collected during June and two during the following September<sup>170</sup>. Adenoids in the Ren *et al.* study were from children undergoing adenoidectomy, collected year-round.<sup>22</sup> Demographic data was not accessible because these were discarded surgical specimens. NLF samples in our study were collected during the Fall season in healthy adults. Future

studies will need to sample complementary niches in the URT to understand coordinated changes of bacterial communities in response to viral infection.

In our study, we surveyed the microbiome of the healthy nasopharynx during viral infection and cataloged a highly diverse population of microbes. The nine dominant phyla of the adenoid microbiome<sup>22</sup> contained the dominant phyla found in our study, including Firmicutes, Proteobacteria, and Actinobacteria. We found the same dominant phyla as a study of nasopharyngeal microbiome of healthy Chinese adults.<sup>146</sup> We did not observe an effect of viral infection on microbial composition, which is likely due to a small population, insufficient sequencing depth to detect the full complexity of the microbiome, and the inability to classify OTUs to the species level. In addition to deeper sequencing to achieve species level information, shotgun metatranscriptomics could be used as a complementary approach, to characterize the changes in gene expression that the invasion of a virus can elicit in a microbial community.<sup>176</sup> This would help to address the open question of the functional role that the microbiome plays in viral infections. It is uncertain, for instance, whether the microbiome can help to protect the host from viral infection by providing a barrier or by initiating an immune response. Alternatively, it is also possible that the microbiome can aid a virus in invading the host mucosa. One metagenomic study of the airway DNA virome of cystic fibrosis (CF) showed that the metabolic profiles of CF and non-CF individuals were distinctly different.<sup>177</sup> This study sought to find target pathways that have altered expression levels in the diseased state, rather than taxonomic composition differences of the microbiome to discover novel therapeutic strategies.

Our findings that each individual had a unique nasopharyngeal microbiome broaden our views on how to study and treat URTIs. Most prior studies have looked at a population of individuals with a specific URTI or complication (*i.e.*, sinusitis) and sampled each individual for specific pathogens. From the findings of such studies, conclusions may be drawn as to which pathogens are likely involved and targeted treatments may be chosen at the population level. However, with a new understanding of the complexity of the NP microbiome at the individual level, we can appreciate that future studies need to sample patients with URTIs or complications at baseline comprehensively, and then during acute disease to determine which bacteria respond, expand, disappear. This approach is anticipated to lead to improved understanding of disease and more tailored approaches to treatments of these diseases.

In conclusion, we have conducted a longitudinal survey of the nasopharyngeal microbiome of healthy adults during rhinovirus challenge. The microbiome of each individual differs between time points throughout the viral challenge, though variability within individuals was less than between individuals. No significant changes in bacterial presence or relative abundance between time points were found due to viral infection at various taxonomic levels. Comparison of the relative abundance of genera between infected and not infected individuals resulted in significant differences in two genera. Deeper sequencing and metagenomic sequencing will further characterize this microbiome and the response of the microbiome to viral infection. With this knowledge, we will better understand the pathogenesis of URT infections.

## Chapter 4

### Genome-wide association study of chronic otitis media with effusion and recurrent otitis media

#### 4.1 Chapter Introduction

In addition to studying the microbial contributors to OM pathogenesis (Chapters 2-3), we sought to determine novel genetic loci and pathways associated with COME/ROM in a human population. It has been shown that OM risk is heritable, but at the commencement of this project there had been only two genome-scale (linkage) studies<sup>24,112</sup> and no GWAS of OM. Candidate gene studies, selected on the basis of biological plausibility in the pathogenesis and progression of OM, have confirmed that genes involved in inflammatory pathways, mucin production, and bacterial recognition may confer risk of susceptibility to OM, but the field had not gained knowledge of novel genes or pathways related to OM progression.

The two published linkage studies of OM provided large regions of the genome to follow up. Late in the progression of this research, a GWAS of acute and chronic OM was published from a group at the University of Western Australia.<sup>115</sup> In contrast to our GWAS of COME/ROM, this first GWAS of OM included cases with various severities of OM from a general population. With agnostic approaches such as ENU mutagenesis screens in mice<sup>178</sup> and the first GWAS of acute and chronic OM, additional pathways have been found to be involved in OM risk, including the TGF $\beta$  pathway<sup>72,115,179</sup> and HIF-VEGF pathway<sup>13</sup>. Table S4.1 describes published agnostic approaches to identify OM susceptibility loci. To better understand risk of COME/ROM, we aimed to conduct a discovery GWAS in the family-based population recruited at the University of Minnesota

and evaluate replication in an independent population of OM. The novel regions of the genome associated with OM is anticipated to lead to the development of new methods of screening and treatment of children diagnosed with COME/ROM.

#### **4.2 A genome-wide association study of chronic otitis media with effusion and recurrent otitis media identifies a novel susceptibility locus on chromosome 2**

*Journal of the Association for Research in Otolaryngology*. 2013 Dec;14(6):791-800.

E. Kaitlynn Allen<sup>1,2</sup>, Wei-Min Chen<sup>1,3</sup>, Daniel E. Weeks<sup>5,6</sup>, Fang Chen<sup>1</sup>, Xuanlin Hou<sup>1</sup>, José L. Mattos<sup>1</sup>, Josyf C. Mychaleckyj<sup>1,3</sup>, Fernando Segade<sup>7</sup>, Margaretha L. Casselbrant<sup>8</sup>, Ellen M. Mandel<sup>8</sup>, Robert E. Ferrell<sup>5</sup>, Stephen S. Rich<sup>1,3,4</sup>, Kathleen A. Daly<sup>9</sup>, Michèle M. Sale<sup>1,2,4</sup>

<sup>1</sup>Center for Public Health Genomics, <sup>2</sup> Department of Biochemistry and Molecular Genetics, <sup>3</sup>Department of Public Health Sciences, <sup>4</sup>Department of Medicine , University of Virginia, Charlottesville, VA 22908; <sup>5</sup>Department of Human Genetics, <sup>6</sup>Department of Biostatistics, Graduate School of Public Health, University of Pittsburgh, Pittsburgh, PA 15261; <sup>7</sup>Department of Anatomy and Cell Biology, University of Pennsylvania, Philadelphia, PA 19104; <sup>8</sup>Children's Hospital of Pittsburgh of UPMC, Department of Otolaryngology, Pittsburgh, PA 15224; <sup>9</sup>Department of Otolaryngology, University of Minnesota, Minneapolis MN 55455.

E. Kaitlynn Allen prioritized UMN GWAS SNPs for replication in UPitt, carried out meta-analysis of UMN and UPitt genotype data and *in silico* eQTL analysis. Drs. Josyf C. Mychaleckyj and Xuanlin Hou carried out QC of GWAS data. Drs. Wei-Min Chen and Fang Chen carried out statistical analyses and imputation analysis of GWAS data. Drs. Stephen S. Rich, Kathleen Daly and Michèle M. Sale conceived of the GWAS study and Dr. Sale coordinated GWAS genotyping. Drs. Kathleen Daly and Stephen Rich conceived and designed the UMN family-based study, and Dr. Daly coordinated recruitment. Drs. Margaretha L. Casselbrant and Ellen M. Mandel coordinated study recruitment of the UPitt population. Dr. Robert E. Ferrell coordinated replication genotyping, and Dr. Weeks conducted UPitt statistical analyses. E. Kaitlynn Allen, Dr. Wei-Min Chen, and Dr. Michèle M. Sale drafted the manuscript, and all authors edited and approved of the manuscript before submission.

#### **4.2.1 Abstract**

**Objective:** Chronic otitis media with effusion (COME) and recurrent otitis media (ROM) have been shown to be heritable, but candidate gene and linkage studies to date have been equivocal. Our aim was to identify genetic susceptibility factors using a genome-wide association study (GWAS).

**Methods:** We genotyped 602 subjects from 143 families with 373 COME/ROM subjects using the Illumina HumanCNV370-Duo DNA BeadChip (324,748 SNPs). We carried out the GWAS scan and imputed SNPs at the regions with the most significant associations. Replication genotyping in an independent family-based sample was conducted for 53 SNPs: the 41 most significant SNPs with  $P < 10^{-4}$ , and 12 imputed SNPs with  $P < 10^{-4}$  on chromosome 15 (near the strongest signal).

**Results:** We replicated the association of rs10497394 (GWAS discovery  $P=1.30 \times 10^{-5}$ ) on chromosome 2 in the independent otitis media population ( $P=4.7 \times 10^{-5}$ ; meta-analysis  $P=1.52 \times 10^{-8}$ ). Three additional SNPs had replication  $P$ -values  $< 0.10$ . Two were on chromosome 15q26.1 including rs1110060, the strongest association with COME/ROM in the primary GWAS ( $P=3.4 \times 10^{-7}$ ) in *KIF7* intron 7 ( $P=0.072$ ), and rs10775247, a non-synonymous SNP in *TICRR* exon 2 ( $P=0.075$ ). The third SNP, rs386057, was on chromosome 5 in *TPPP* intron 1 ( $P=0.045$ ).

**Conclusion:** We have performed the first GWAS of COME/ROM and have identified a SNP rs10497394 on chromosome 2 significantly associated with COME/ROM susceptibility. This SNP is within a 537 kb intergenic region, bordered by *CDCA7* and *SP3*. The genomic and functional significance of this newly-identified locus in COME/ROM pathogenesis requires additional investigation.

#### 4.2.2 Introduction

Every year, over 10 million children in the United States are treated for otitis media (OM), inflammation of the middle ear.<sup>180,181</sup> OM is a leading reason for physician visits during childhood and represents a major component of the pediatric healthcare burden.<sup>182</sup> Antibiotic prescriptions for OM contribute significantly to increasing multi-drug resistance<sup>182</sup>, and insertion of ventilation tubes into the tympanic membrane in the middle ear to assist fluid drainage remains one of most common childhood surgeries performed in the U.S.<sup>183</sup>

Acute OM is characterized by middle ear inflammation with signs of middle ear effusion and acute symptoms of infection including middle ear pain and fever.<sup>184,185</sup> A minority of



children who suffer from acute OM will develop either chronic otitis media with effusion (COME) or recurrent otitis media (ROM). Estimates suggest that by two years of age, nearly a third of all children will have had more than three episodes of OM within six months or four episodes within one year.<sup>186</sup> Recurrent and chronic forms of OM are associated with high economic and societal costs such as medical care costs, parental days lost from work, and loss of earnings.<sup>187</sup> Complications of COME/ROM include permanent hearing loss, tympanic membrane (TM) abnormalities (e.g. chronic ear drainage), other infections such as meningitis, and other sequelae.<sup>182,188,189</sup> The annual economic burden of otitis media in the United States is estimated to be at least \$5 billion for direct and indirect costs.<sup>1</sup>

COME/ROM has long been recognized to aggregate in families.<sup>190-193</sup> The first family study of COME/ROM<sup>194</sup> demonstrated that first-degree relatives (parents and siblings) of probands with COME/ROM had greater than expected rates of OM based on population rates. Subsequent twin and triplet studies confirmed the strong familial aggregation of COME and ROM, with heritability estimates of 0.64-0.74 in monozygotic twins and 0.20-0.53 in dizygotic twins.<sup>48,195-198</sup> Two linkage studies for COME/ROM have been published to date<sup>106,199,200</sup>; however, the loci identified in the two studies did not overlap. Recently, the first GWAS for acute and chronic OM in the first three years of life discovered novel associations with OM, including *CAPN14*, *GALNT14*, and the *BPIFA* gene cluster, though the study lacked replication in an independent population of COME/ROM.<sup>115</sup> In an effort to detect common variants contributing to COME/ROM susceptibility, we have conducted the first genome-wide association study (GWAS) of COME/ROM.

### 4.2.3 Methods

This study was conducted with Institutional Review Board approval at the University of Minnesota, Wake Forest University, the University of Virginia, and the University of Pittsburgh, and adhered to the tenets of the Declaration of Helsinki.

#### *4.2.3.1 University of Minnesota (UMN) Family Study*

Details of recruitment and examination of the study participants have been described previously.<sup>194,200,201</sup> Index cases (probands) who had tympanostomy tube surgery for COME/ROM and their family members were recruited for the study. Four data sources used to define affected status included findings from ENT exam, tympanometry, reported OM history, and medical record, described in detail in Table S4.2. At least two positive data sources were required for an individual to be considered affected. An otolaryngologist performed an ear examination to determine presence of OM sequelae without knowledge of the subject's prior OM history. Tympanometric testing was performed in subjects at three frequencies (226, 630 or 710, and 1400) to detect abnormal middle ear mechanics, and hearing was screened at 20dB for speech frequencies. Individuals from 143 families with phenotypic data and DNA available were enrolled in genetic studies. The sample includes 44 families with 5-10 members, 55 families with four members, 36 trios, and 8 families with less than three members (**Table 4.1**).

**Table 4.1** Participant characteristics for the University of Minnesota (UMN) and University of Pittsburgh (UPitt) *family-based* study populations

Trait	UMN Study	UPitt Study
	Value	Value
Number of families	143	441
Number of genotyped subjects	602	1584
Female (%)	53.3%	51.4%
Affected subjects	62.0% (373/602)	59.1% (932/1584)
Caucasian	95.7%	100%
Non-Hispanic	98.7%	-

#### *4.2.3.2 University of Pittsburgh (UPitt) Study*

We carried out a replication analysis in an independent study of OM<sup>199</sup> that consisted of 1,584 genotyped individuals from 441 Caucasian families. In order to assure a history of significant ear disease, two or more full siblings who both or all had undergone tympanostomy tube insertion were enrolled. The need for tympanostomy tube insertion established a subject's history of significant middle ear disease. A subject was only considered affected if he/she had undergone tympanostomy tube insertion at least once for recurrent/persistent OM, while a subject was considered unaffected if he/she had never had tympanostomy tubes and had no known history of recurrent/persistent OM. The remaining subjects were considered as having unknown disease status. Otoscopic examinations and tympanometry were conducted at entry when possible, but the condition of the ears at entry did not determine eligibility and the tubes may have been inserted many years prior to study entry.<sup>199</sup> The UPitt sample includes 87 families with 5 to 8 members, 330 families with four members, and 12 trios.

#### *4.2.3.3 GWAS genotyping and data cleaning*

The Illumina HumanCNV370-Duo DNA BeadChip was used for genotyping DNA family members of the UMN Study. Removal of SNPs was based upon filtering for poor genotype clusters, low minor allele frequency ( $MAF < 0.01$ ), and genotypes inconsistent with Hardy Weinberg proportions ( $P < 10^{-5}$ ). Samples and SNPs with excessive Mendelian errors ( $> 0.6\%$ ) and/or low genotype call rates ( $< 95\%$ ) were removed. A total of 324,748 SNPs were genotyped in 602 subjects and available for analyses.

Kinship coefficients were estimated between all pairs of individuals based on the GWAS SNP data using software KING<sup>202</sup> and these data were used to correct both within- and between-family pedigree errors. The Minnesota study samples are from a predominantly Caucasian population (95.7%), with the remaining 4.3% of the population from Asian, Native American, and Mixed populations. Two families, consisting of 2 and 4 individuals respectively, had non-Caucasian ethnicity, and 19 individuals self-identified as belonging to more than one race. These individuals did not substantially change the family-based association results and so were retained for analyses.

#### *4.2.3.4 Imputation*

To improve genomic coverage, we used the imputation method implemented in the software package MACH using HapMap3 with the CEU reference population.<sup>203,204</sup> This method uses Markov models to identify stretches of shared chromosomes between individuals, and then to infer intervening genotypes by contrasting study samples with densely typed HapMap samples. Although imputation is accurate<sup>203</sup>, improvement in accuracy can be achieved by tuning the quality metric threshold. The “most likely” genotypes were used in the QLSw test; therefore, imputed SNPs were treated the same as genotyped SNPs. Imputed SNPs that passed a quality threshold of  $r^2 > 0.3$  and  $MAF > 0.5$  were included in analysis. Imputed genotypes that were not consistent with Mendelian inheritance in a family were set to missing. To reduce the multiple comparisons penalty of evaluating all imputed SNPs, our investigation of imputed data was focused only on GWAS regions with prior evidence of association in the primary (genotyped) analysis ( $P < 10^{-4}$ ). In our region of interest on chromosome 15, there were 2,818 SNPs available for imputation.

After evidence of replication on chromosome 2, we investigated imputed data in the chromosome 2 region (chr2:174286992-174307309) using the same imputation methods already described. Eighteen imputed SNPs were available in this region for association analysis.

#### *4.2.3.5 Statistical analyses*

We applied the family-based association test QLS<sub>w</sub> to the genome scan data. This method was recently developed<sup>205</sup> as an extension of the powerful method MQLS<sup>206</sup>, and the implementation is publicly available through software package GDT.<sup>207</sup> The QLS<sub>w</sub> method has been shown more powerful than the MQLS method in the presence of population stratification, and/or in the presence of family ascertainment such as the affected sibling pairs (ASP) design. Population substructure does not affect analysis using the QLS<sub>w</sub> family-based association test as this approach is robust to population stratification. The QLS<sub>w</sub> family-based association test (January 2011) was used for GWAS analysis and UPitt replication analyses.

Association of SNPs on the X chromosome with COME/ROM was analyzed using the Generalized Disequilibrium Test (GDT).<sup>207</sup>

The GDT method<sup>207</sup> was used for the conditional analysis of the chromosome 2 locus. The genotype at SNP rs10497394 was considered as a covariate and was adjusted in the GDT test statistic.

#### 4.2.3.6 UPitt Replication study

A total of 53 SNPs were selected for replication genotyping in the UPitt families. The 53 SNPs included the 41 most significant SNPs with  $P < 10^{-4}$  from our discovery GWAS (N=41), plus 12 imputed SNPs with  $P < 10^{-4}$  located on chromosome 15 near rs1110060, the only genotyped SNP with marginal significance. Genotyping was conducted on the Sequenom platform using the iPlex assay.<sup>208</sup>

Meta-analysis was carried out on the data from the UMN and UPitt studies. Briefly, a weighted z score–based fixed effects meta-analysis method was used to combine the results of the UMN and UPitt studies using the program METAL.<sup>209</sup> Weighting was proportional to the sample size of each study.<sup>210</sup>

#### 4.2.3.7 In silico eQTL analysis

An eQTL analysis of the chromosomes 2, 5, and 15 regions was conducted using three different databases including “eQTL resources @ the pritchard lab” (<http://eqtl.uchicago.edu>), Wellcome Trust Sanger Institute’s Genevar<sup>211</sup>, and SCAN: SNP and CNV annotation database<sup>212</sup>. The eQTL analysis using “eQTL resources @ the pritchard lab” was conducted using RNAseq data from a study using total RNA from lymphoblastoid cell lines in 63 HapMap individuals of European ancestry.<sup>213</sup> The eQTL analysis using Genevar was conducted using expression data from the total RNA of 109 lymphoblastoid cell lines of European ancestry.<sup>214</sup> The eQTL analysis using SCAN: SNP and CNV annotation database was conducted using expression data from total RNA from lymphoblastoid cell lines of 30 trios of European ancestry and 30 trios of African

ancestry<sup>215</sup>, though based on our population, only significant results from the population of European ancestry are included.

#### 4.2.4 Results

##### 4.2.4.1 Association results of GWAS

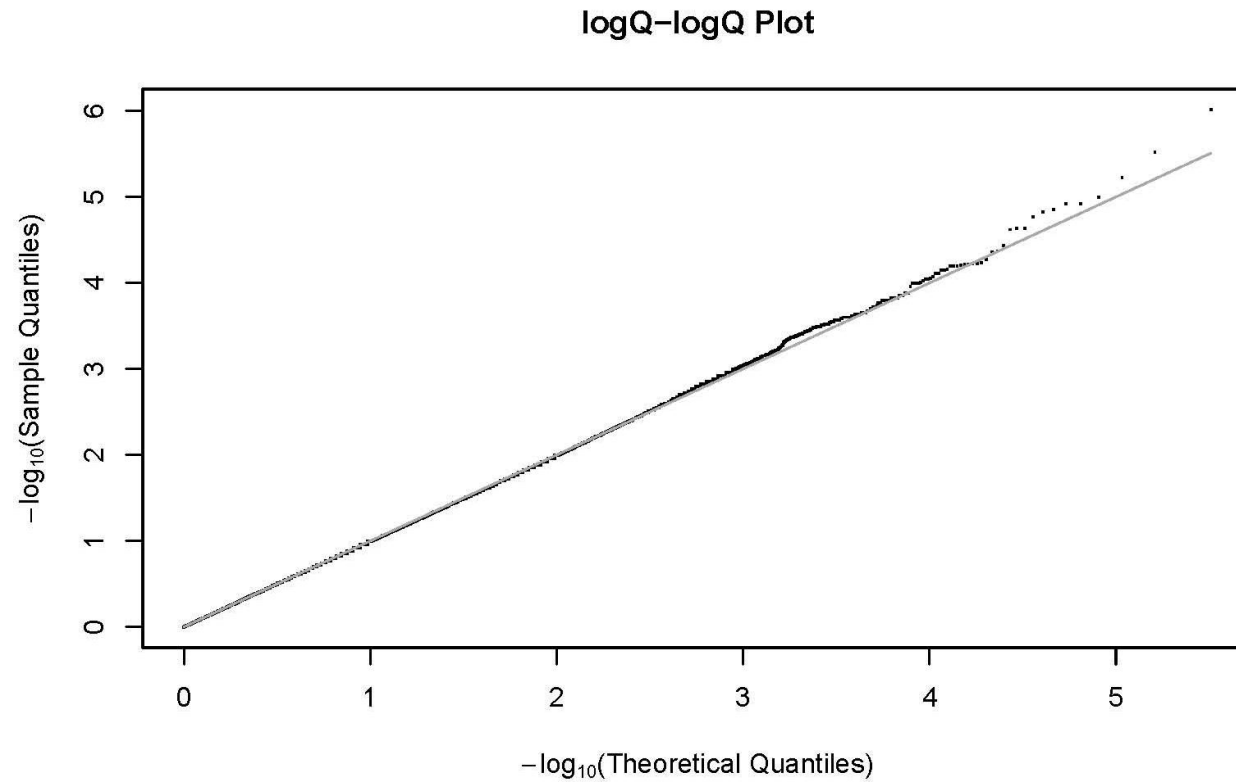
We performed the first GWAS for COME/ROM by analyzing 324,748 SNPs on the Illumina HumanCNV370-Duo DNA BeadChip (genomic control lambda value (GC) = 0.993) in the UMN family population (**Figure 4.1**). The strongest association with COME/ROM in the UMN families was rs1110060 ( $P=9.1 \times 10^{-7}$  (**Figure 4.1**)) located on chromosome 15 within the gene encoding *KIF7*. One region had a “suggestive”<sup>216</sup>  $P<10^{-5}$  located on chromosome 6 (rs10499006,  $P=2.1 \times 10^{-6}$ ), and one region approached the “suggestive”  $P$  value threshold on chromosome 3 (rs6438779,  $P=2.9 \times 10^{-6}$ ), both of which were in intergenic regions. CNV association analyses conducted with PennCNV<sup>217</sup> did not reveal any evidence for significant association (data not shown).

Analyses of the X chromosome using the GDT<sup>207</sup> did not result in any significant findings. The most significant result on the X chromosome was with rs2215100 ( $P=3.3 \times 10^{-5}$ ). This was the only SNP that exceeded  $P=10^{-4}$ . SNP rs2215100 lies within a large intergenic region with the closest genic feature the VENT homeobox 1 pseudogene (*VENTXP1*), 115.8 kb distal to the SNP, and closest gene Melanoma antigen family B, 6 (*MAGEB6*), 481.3 kb distal to the SNP.



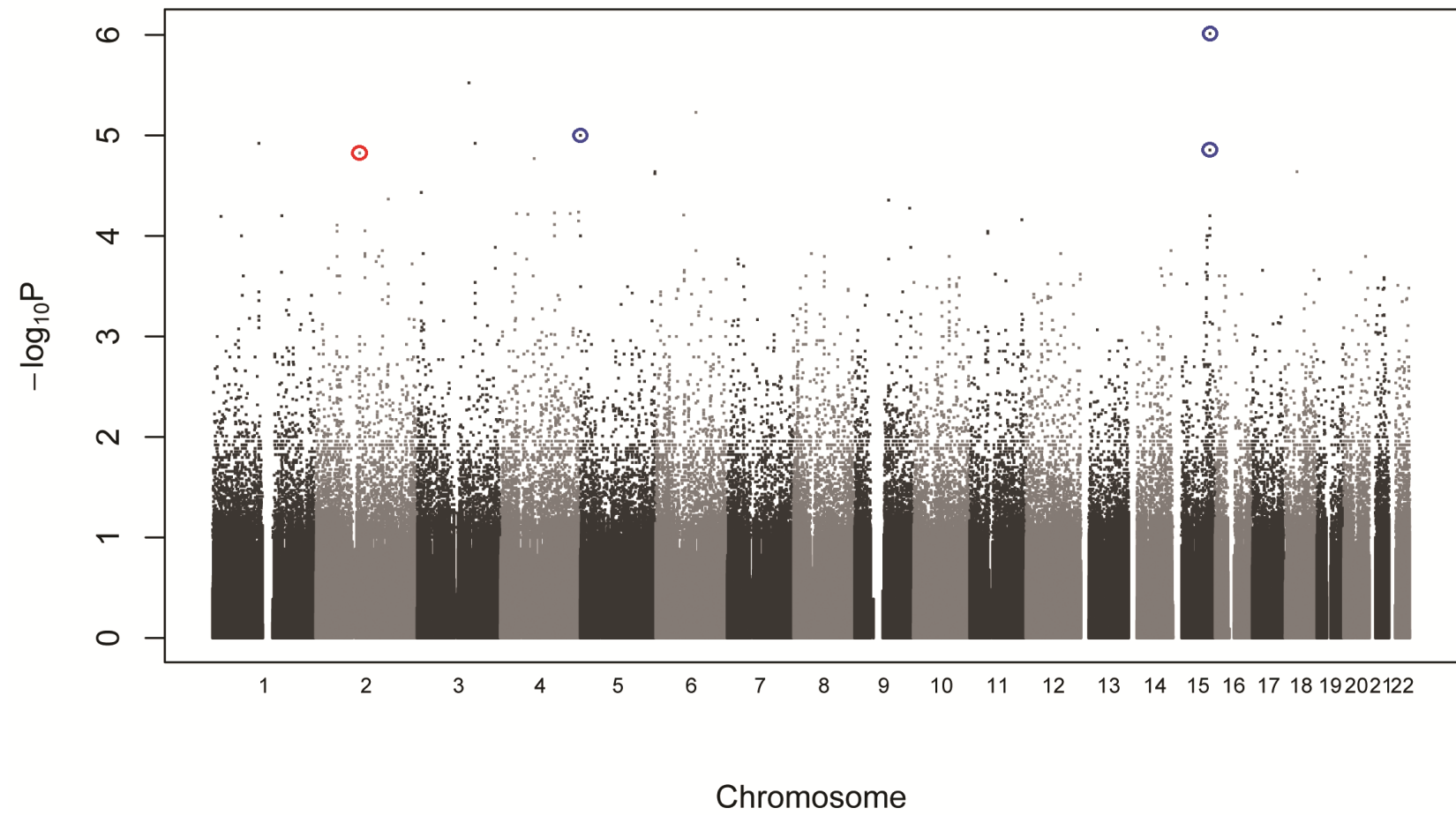
**Figure 4.1** COME/ROM genome-wide scan association results using the  $QLS_w$  method: **A.** Quantile-quantile (Q-Q) plot (Genomic control value (GC) = 0.993). **B.** Manhattan plot of GWAS data. The circled SNPs represent the SNPs in Table 2 (replication  $P < 0.10$ .) The SNP circled in red is rs10497394, the significantly replicated SNP. X chromosome analyses are not presented since they were analyzed using a different method (GDT<sup>218</sup>).

**A.**



B.

### Otitis Media GWAS Scan ( GC = 0.993 )



#### 4.2.4.2 Results of UPitt Replication study

All replication results with  $P < 0.10$  are shown in **Table 4.2**. All four SNPs listed in **Table 4.2** had genotyping success rates over 99.35% in the UMN GWAS and over 98.0% in the UPitt replication study. Five of the 53 SNPs failed quality control assessment (rs3773770, rs11639117, rs17360083, rs6808013 and rs11259905). Three SNPs (rs1361049, rs4701136, and rs11852287) were unable to be successfully genotyped on the Sequenom platform. The significance threshold after performing a conservative Bonferroni correction for multiple testing (*i.e.*, 45 SNPs) was  $P = 1.1 \times 10^{-3}$ . The locus identified on chromosome 2 (rs10497394,  $P = 2.9 \times 10^{-5}$ ) was successfully replicated in the UPitt Study ( $P = 4.7 \times 10^{-5}$ ) with a meta-analysis  $P = 1.52 \times 10^{-8}$ . Meta-analysis results are listed in Table 4.2 and confirmed directionality of these four SNPs in the two studies. Meta-analysis results for all 45 SNPs are listed in Table S4.3. Note a naïve Bonferroni corrected  $P$  value threshold that assumes both datasets are typed at 324,748 SNPs is  $0.05/324,748 = 1.54 \times 10^{-7}$ , and the significance threshold for a more appropriate two-stage joint analysis is  $1.80 \times 10^{-6}$ .<sup>219</sup>

#### 4.2.4.3 Characterization of top associations

The replicated SNP on chromosome 2 (rs10497394) maps within a 537 kb intergenic region bordered by two genes (*CDCA7* and *SP3*) coding for nuclear proteins with transcription factor activities. Subsequent imputation analyses of the chromosome 2 region using HapMap3 data found 14 SNPs with  $P < 10^{-5}$  in this region, with the lowest  $P = 1.1 \times 10^{-7}$ .  $P$ -values for all 14 of these SNPs exceeded our top genotyped SNP, rs10497394. Haplotype analysis using Haploview (Version 3; Release 27; CEU analysis panel)<sup>220</sup> identified two haplotype blocks of high LD ( $r^2 > 0.9$ ), one block spanning approximately 16kb (chr2:174286992-174303858) and the other less than 1kb

**Table 4.2** Replication results with  $P < 0.10$  and meta-analysis results of those SNPs. Information provided in this table relates to the reference allele.

				UMN Study				UPitt Study				Meta-Analysis		
Chr	SNP	Position (hg19)	Ref Allele	Allele Freq	OR	Z	P-value	Allele Freq	OR	Z	P-Value	OR (95% CI)	Z	P-Value
2	rs10497394	174,297,659	G	0.629	1.754	4.18	$2.9 \times 10^{-5}$	0.637	1.422	4.07	$4.7 \times 10^{-5}$	1.51 (1.3-1.7)	5.66	$1.52 \times 10^{-8}$
5	rs386057	685,748	G	0.447	1.733	4.33	$1.5 \times 10^{-5}$	0.585	1.193	2.01	0.045	1.35 (1.2-1.6)	3.98	$6.94 \times 10^{-5}$
15	rs10775247	90,126,121	C	0.570	0.573	-4.00	$6.3 \times 10^{-5}$	0.460	0.861	-1.78	0.075	0.77 (0.67-0.89)	-3.62	$3.00 \times 10^{-4}$
15	rs1110060	90,190,048	A	0.551	0.510	-4.91	$9.1 \times 10^{-7*}$	0.409	0.856	-1.80	0.072	0.74 (0.64-0.85)	-4.10	$4.09 \times 10^{-5}$

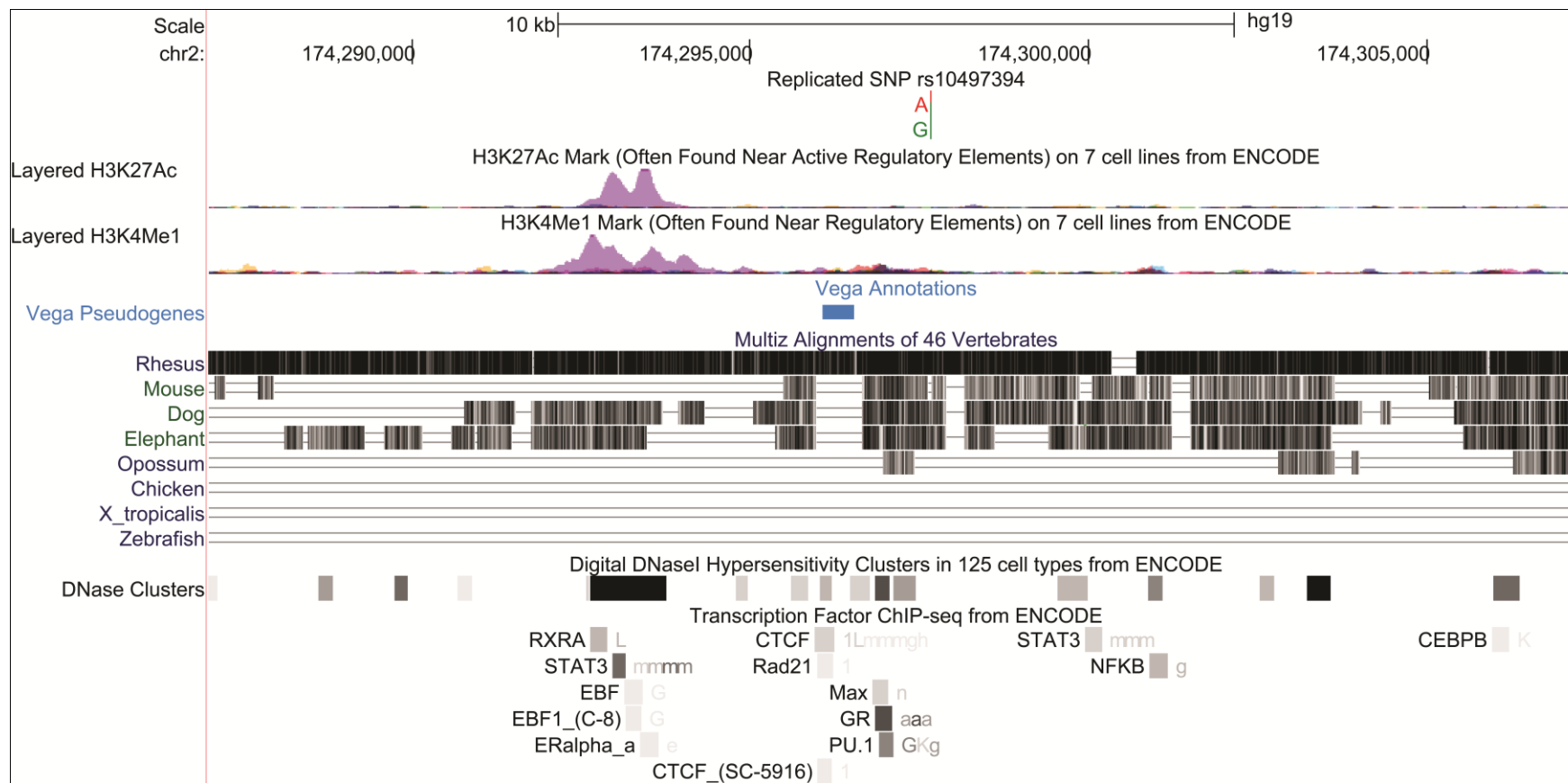
\*Most significant result from UMN Study GWAS; Z score listed for minor allele.

(chr2:174306514-174307309), and the total region containing both haplotype blocks spanning chr2:174286992-174307309 (UCSC Genome Browser, build hg19; Figure S4.1).

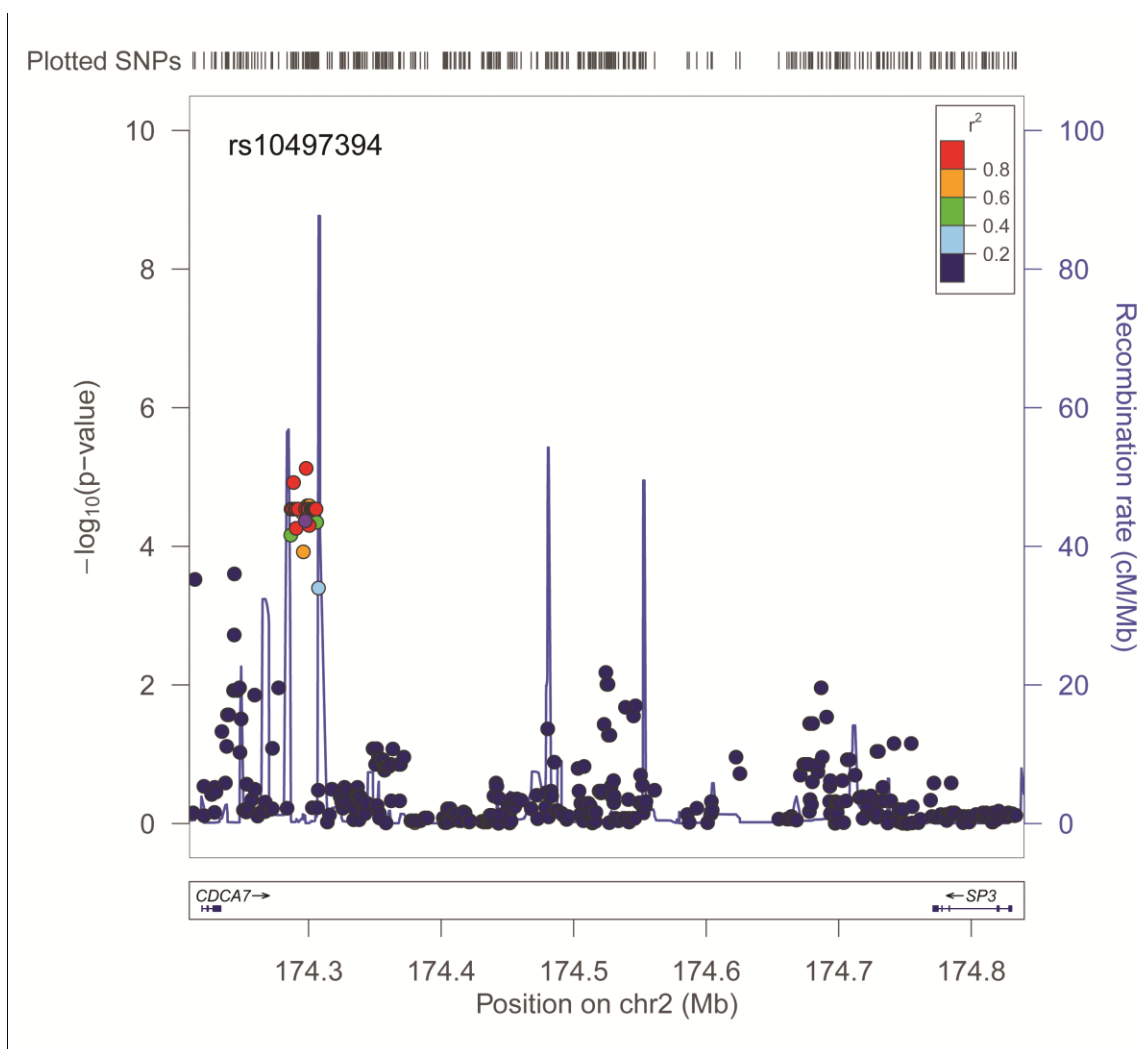
This region of high LD contains many regulatory and genomic elements that have putative functional effects (**Figure 4.2**). Locus Zoom<sup>221</sup> plots using our primary GWAS data with HapMap3 imputation results (**Figure 4.3**) confirmed this region is in high LD and lies between two areas of high recombination. HaploReg<sup>222</sup> identified two SNPs (rs1435769 and rs10185704) in high LD ( $r^2=1.0$ ) with rs10497394 that lie within transcription factor binding sites for STAT3 and NFκB, respectively. In order to investigate whether other SNPs in this region represent independent associations with COME/ROM in this population, we used a conditional analysis using the genotype at SNP rs10497394 as a covariate. There were 14 SNPs in the chromosome 2 region of interest with GDT  $P<0.001$  (the smallest  $P=8.6\times 10^{-5}$ ) before the conditional analysis, and none of them have  $P<0.001$  after the conditional analysis. Note that although the existing method GDT is slightly less powerful than the method QLSw that we are currently using, it has the ability to adjust for covariates. This analysis suggests that the other SNPs in this region do not represent independent associations with COME/ROM.

The strongest locus from our discovery sample (rs1110060) produced a replication  $P=0.072$  and a meta-analysis  $P=4.09\times 10^{-5}$  (**Table 4.2**.) The two chromosome 15 SNPs (rs10775247 and rs1110060) lie about 64 kB from each other in adjacent genes *TICRR* and *KIF7* ( $r^2=0.47$ ;  $D'=0.77$ ). SNP rs386057 (replication  $P=0.045$ ; meta-analysis  $P=6.94\times 10^{-5}$ , **Table 4.2**) is located on chromosome 5p15.33 within intron 1 of the gene tubulin polymerization promoting protein (*TPPP*).

**Figure 4.2** Genomic characteristics of the chromosome 2 region of high LD (chr2:174286992-174307309) from UCSC Genome Browser build hg19 which includes a processed pseudogene with 89% sequence identity to the coding sequence of the hematological and neurological expressed 1 (HN1) gene; an H3K27Ac mark and an H3K4Me1 mark from ENCODE; DNaseI hypersensitive regions including those discovered in relevant cell lines (lung, non-pigment ciliary epithelial, and small airway epithelial cell lines); a conserved CTCF insulator binding site; and numerous predicted transcription factor binding sites.<sup>223,224</sup> Conservation in this region is measured using Multiz Alignments of eight vertebrates.



**Figure 4.3** Locus Zoom<sup>221</sup> plots of the chromosome 2 region using UMN GWAS data and HapMap3 imputation. SNP shown in purple is our replicated SNP rs10497394, and linkage disequilibrium (LD) is in relation to this SNP.





#### 4.2.4.4 *In silico* eQTL analysis

From our *in silico* eQTL analysis, the SNPs on chromosomes 2 and 15 were found to regulate genes in the CEU population (**Table 4.3**; results from all populations in Table S4.4). Analysis of the chromosome 2 region of interest using the eQTL browser of the Pritchard Lab and Genevar produced no significant results. Using the SCAN database, the chromosome 2 SNP (rs10497394) is associated with the expression of the LDLR gene on chromosome 19 in the CEU population ( $P=6 \times 10^{-5}$ ). eQTL analysis of the chromosome 15 region of interest with the “eQTL resources @ the pritchard lab” found our most significantly associated SNP in the discovery population (rs1110060) acts in *cis* as an exon-QTL for IQGAP1.<sup>213</sup> Using the SCAN database, SNP rs10775247 regulates expression of Resistin (RETN) in the CEU population ( $P=4 \times 10^{-5}$ ). Due to differences in sample sets and methods for measuring and analyzing expression data, results varied between the databases.

#### 4.2.5 Discussion

We have conducted the first GWAS of COME/ROM and identified and investigated replication in an independent collection of families with otitis media. In our discovery GWAS, one SNP on chromosome 15q26.1 (rs1110060) approached genome-wide significance ( $P=9.1 \times 10^{-7}$ ). A panel of 53 SNPs with P-values less than  $10^{-4}$  from our GWAS was tested for replication in an independent otitis OM family population at the University of Pittsburgh. We successfully replicated a novel locus on chromosome 2q31.1 (rs10497394), with replication  $P=4.7 \times 10^{-5}$  in the UPitt family population of OM and meta-analysis  $P=1.52 \times 10^{-8}$ .

**Table 4.3** Significant results from the three eQTL browsers used for the *in silico* eQTL analysis for the four SNPs of interest.

Chromosome	SNP	Database	Gene	Population	P value	Cis/Trans
2	rs10497394	eQTL browser @ pritchard lab	<i>LDLR</i>	CEU	6.00x10 <sup>-05</sup>	Trans (Chr19)
5	rs386057	--	--	--	--	--
15	rs10775247	SNP SNP and CNV annotation database	<i>RETN</i>	CEU	4.00x10 <sup>-05</sup>	Trans (Chr19)
15	rs1110060	eQTL browser @ pritchard lab	<i>IQGAP1</i>	CEU	3.07x10 <sup>-05</sup>	Cis (exonQTL)

The replicated SNP rs10497394 is in a large intergenic region on chromosome 2q31.1 between the genes *CDCA7* and *SP3*. Although predicting functional consequences of non-protein coding SNPs proves challenging, loci have been identified in a GWAS in gene deserts and have been found to have functional significance in complex diseases.<sup>225</sup> SNP rs10497394 was found to regulate expression of Low density lipoprotein receptor (LDLR) on chromosome 19 using SCAN: SNP and CNV annotation database. LDLR is expressed on ciliated airway epithelial cells, and LDLR has been found to have a role in the pathogenesis of asthma in mice.<sup>226</sup> Additionally, LDLR is a binding site for human rhinovirus species C (HRV-C) which have been implicated in upper and lower respiratory infections in children and adults with chronic respiratory disease.<sup>227</sup> Replication and significant meta-analysis results for rs10497394, when combined with evidence for potential functional consequences, support a role in COME/ROM pathogenesis that warrants further investigation.

Association results on chromosomes 5 and 15 did not meet Bonferroni-corrected significance for replication, but are of interest for further evaluation in additional populations. SNP rs1110060, with strongest evidence for association in the primary GWAS, is located on 15q26.1 in the proximity of the 5' end of intron 7 of *KIF7*, a location that may conceivably affect the splicing of the primary *KIF7* transcript. The likelihood of this SNP affecting splicing was confirmed by Spliceman Web Server, an online tool used to predict how likely mutations around annotated splice sites are to disrupt splicing.<sup>228</sup> *KIF7* is involved in regulating mammalian *Sonic Hedgehog* (*Shh*)<sup>229</sup> and Indian hedgehog (*IHH*)<sup>230</sup> via protein trafficking within the primary cilium<sup>229</sup>, resulting in skeletal abnormalities, possibly providing a link between craniofacial development such as Eustachian tube formation and COME/ROM susceptibility. Hedgehog interacting

protein (*HHIP*), a regulator of the Hedgehog (HH) pathway, and Patched (*PTCH1*), a Shh protein upstream of *KIF7*, have been associated with abnormal lung function and chronic obstructive pulmonary disease (COPD) in two large meta-analyses of GWAS studies of lung function.<sup>231</sup> Polymorphisms in a related kinesin gene, *KIF3A*, were recently shown to be associated with aspirin-intolerance in asthma and the decline of forced expiratory volume at 1 s (FEV<sub>1</sub>) % by aspirin provocation<sup>232</sup>, suggesting a possible role in immune response.

The second SNP on chromosome 15 in Table 4.3, rs10775247, is a missense mutation (Arg287Cys) in *C15orf42* (*TICRR*). *TICRR* encodes the protein Treslin which has been shown to play a role in initiation of DNA replication.<sup>233</sup> Although the link with COME/ROM is unclear, Treslin interacts with TopBP1, a member of a complex impacted by inactivation of host DNA replication by certain viruses, including adenovirus and parainfluenza virus type 2<sup>234</sup> which are both commonly coincident with OM.<sup>38</sup> This suggests that *TICRR* may be worth evaluating in a larger sample to determine a role in COME/ROM susceptibility. A SNP in *TICRR* (rs8032553, not genotyped in our study) was found to be associated with cholesterol/LDL in the Framingham Heart Study.<sup>235</sup> As far as we are aware, none of the other regions reported in our study have been previously associated in any GWAS.

The eQTL analysis of the chromosome 15 region using the “eQTL resources @ the pritchard lab” showed regulation of *IQGAP1* by rs1110060 as an exon-QTL. *IQGAP1* is located about 741 kB distal to this SNP (chr15:90931473-91045475; hg19). *IQGAP1* is a widely conserved and ubiquitously expressed scaffold protein with numerous protein-

interaction domains allowing for binding to diverse proteins.<sup>236</sup> *IQGAP1* has been implicated as a target of several bacteria as a player in their mechanisms of infection. *Salmonella typhimurium* has been shown to bind to *IQGAP1* which allows *S. typhimurium* to escape the host immune response allowing for the establishment of chronic infection.<sup>237</sup> Additionally, *IQGAP1* was found to be needed to organize ROS-dependent *VEGF* signaling via *VEGFR2* in endothelial cells which may contribute to the repair and maintenance of blood vessels and angiogenesis.<sup>238</sup> *VEGF* signaling is known to induce angiogenesis, increase vascular permeability, and influence inflammation in several ways, all which could increase risk of chronic otitis media by effusion build up and increased inflammation.<sup>239</sup> Inhibitors of *VEGF* signaling have been found to reduce hearing loss and fluid buildup in the middle ear of *Junbo* mice, a spontaneous mouse model of OM.<sup>13</sup> These roles of *IQGAP1* indicate that *IQGAP1* may be involved in COME/ROM pathogenesis though further investigation is necessary.

SNP rs386057 is located on chromosome 5p15.33 in the intronic region of the gene tubulin polymerization promoting protein (*TPPP*). *TPPP* is important in microtubule (MT) assembly, MT bundling and the stabilization of existing MTs, and also may play a role in mitotic spindle assembly and nuclear envelope breakdown.<sup>240</sup> It has been identified as a brain-specific protein, but its orthologs are ciliary proteins, suggesting a role of *TPPP* in the basic function of cilia.<sup>241</sup> This potential role ciliary function is interesting due to our findings of the association of rs1110060 in *KIF7*, an important intraflagellar transport protein.

Many epidemiological studies have shown that male children are more likely to be affected with COME/ROM than females.<sup>49,180,242</sup> Despite evidence of a gender bias, analysis of the X chromosome from our GWAS data did not result in any significant loci associated with COME/ROM.

The primary strength of our study is the family-based study design for both the GWAS and UPitt replication phases, permitting an analytical approach robust to population stratification, and the application of a powerful new test of association. Though the QLSw family-based association test is powerful and robust, it is unable to model covariates. The main limitations to this study include SNP density and modest population size. We carried out a power analysis for a two-stage genome-wide association study using the software tool CaTS.<sup>219</sup> In a study with 1309 cases and 877 controls, with 27.5% of the samples genotyped at all 324,748 SNPs in the first stage, and 0.015% of SNPs followed up by genotyping the remaining samples in Stage 2, assuming the prevalence of OM is 0.15, and the genotype relative risk is 1.5 at a causal SNP with the disease allele frequency 0.5, we have 59% power to identify a variant at significance level  $1.5 \times 10^{-7}$ . In common diseases, statistical power has been shown to be slightly lower in family-based studies than in case-control studies<sup>243</sup>; therefore, the statistical power calculated is slightly inflated. Another limitation is that the UMN Study included only a small sample of non-European families, and the UPitt Study excluded all non-European families. Consequently, we are unable to draw conclusions about whether the observed associations are also present in populations of non-European ancestry. Due to the nominal significance of the chromosome 5 and 15 SNPs in the replication study, we cannot exclude the possibility of Type I error in the discovery GWAS or Type II error in the replication study.

The Raine cohort GWAS of OM discovered novel associations with combined acute/chronic OM, including *CAPN14*, *GALTN14*, and the *BPIFA* gene cluster, which are all found within or connected to the TGF $\beta$  pathway.<sup>115</sup> The TGF $\beta$  pathway has recently been identified as important in COME/ROM in multiple studies.<sup>71,73,95</sup> These loci were not able to be replicated in the Western Australia Family Study of Otitis Media (WAFSOM). In our family-based population, we also found no evidence of association with the reported SNPs at the loci that had genotype data in our study (N=21). The lack of replication could be due to the differences in inclusion criteria for case classification (acute/chronic OM in first three years versus COME/ROM), or due to lack of power in the UMN population to detect associations.

This investigation has demonstrated that a family-based GWAS approach can successfully identify susceptibility loci for COME/ROM, despite the high prevalence of acute OM in the general population and modest sample size of our study. Larger genome-scale investigations of COME/ROM would be anticipated to lead to identification of additional susceptibility loci. Functional investigation of the chromosome 2 susceptibility locus identified by our study is warranted and may provide insights into COME/ROM initiation and pathophysiology.

## **Chapter 5**

### **Raine cohort GWAS variant replication study**

#### **5.1 Chapter Introduction**

We aimed to investigate replication of the most significant variants from the Raine cohort GWAS<sup>115</sup> of acute and chronic OM. The Raine cohort is a longitudinal birth cohort from Western Australia that has been followed since enrollment began in 1989-1991. Clinical examinations and parental questionnaires from the first three years of life were used to define OM status. Cases from this population presented with various severities of OM, from AOM to COM. Controls were children with no clinical indication or parental report of OM history in the first three years of life. This GWAS was published late in the progress of my research (October 25, 2012) and has allowed us to investigate new variants of OM. Replication of significant findings in an independent population is critical for genetic studies to confirm association of a novel locus. In the Raine cohort GWAS paper, they attempted to replicate a subset of the most significant associations in an independent population of OM, the Western Australia Family Study of OM. They chose to genotype SNPs that were most significantly associated in their discovery GWAS and those located in genes identified in the Ingenuity Pathway Analysis of their GWAS data. They were unsuccessful in finding significant replication. Due to the small number of OM populations and to the variable nature of clinical diagnosis of the various OM severities and recruitment strategies, it is difficult to find statistically significant replication. Our goal was to evaluate the most significant associations from the Raine discovery GWAS for replication in our family population of COME/ROM.



## 5.2 Evaluation of replication of variants associated with genetic risk of otitis media

*Under Review at PLoS ONE (Submitted January 7, 2013)*

E. Kaitlynn Allen<sup>1,2</sup>, Ani Manichaikul<sup>1</sup>, Stephen S. Rich<sup>1,3</sup>, Kathleen A. Daly<sup>4</sup>, Michèle M. Sale<sup>1,2,5</sup>

<sup>1</sup>Center for Public Health Genomics, <sup>2</sup>Dept of Biochemistry, Molecular Biology, and Genetics, <sup>3</sup>Dept of Public Health Sciences, University of Virginia, Charlottesville, VA 22908; <sup>4</sup>Dept of Otolaryngology, University of Minnesota, Minneapolis MN 55455; <sup>5</sup>Dept of Medicine, University of Virginia, Charlottesville, VA 22908

E. Kaitlynn Allen, Dr. Michèle M. Sale, and Dr. Ani Manichaikul conceived of the study design. E. Kaitlynn Allen carried out genotyping, participated in data analysis, and drafted the manuscript. Dr. Ani Manichaikul carried out QC and statistical analysis of genotyped SNPs. Dr. Wei-Min Chen carried out statistical analyses of GWAS data and advised on statistical approaches. Drs. Kathleen Daly and Stephen Rich conceived and designed the UMN family-based study, and Dr. Daly coordinated recruitment. All authors edited and approved of the manuscript before submission.

### 5.2.1 Abstract

The first Genome Wide Association Study (GWAS) of otitis media (OM) found evidence of association in the Western Australian Pregnancy Cohort (Raine) study, but lacked replication in an independent OM population. The aim of this study was to investigate association at these loci in our family-based sample of chronic otitis media with effusion

and recurrent otitis media (COME/ROM). Autosomal SNPs were selected from the Raine OM GWAS results. SNPs from the Raine cohort GWAS genotyped in our GWAS of COME/ROM had P-values ranging from  $P=0.11-0.92$ . After removal of SNPs previously genotyped in our GWAS of COME/ROM ( $N=21$ ) and those that failed Fluidigm assay design ( $N=1$ ), 26 SNPs were available for genotyping in 716 individuals from our COME/ROM family population. None of the SNPs were replicated in our family-based population ( $P=0.11-0.88$ ). Replication in an independent sample would confirm that these represent novel OM loci, and that further investigation is warranted.

### 5.2.2 Introduction

Inflammation of the middle ear, known as otitis media (OM), is a highly prevalent disease in the pediatric population worldwide. Children at higher risk of OM may develop recurrent or chronic otitis media (COME/ROM), a condition that may lead to multiple antibiotic treatments and at least one tympanostomy tube insertion surgery. Environmental factors are known to play a role in risk of OM; however, it is also known that there is a significant genetic component to OM risk.

The first family study of COME/ROM showed a higher rate of OM than general population rates<sup>244</sup>, indicating familial aggregation of OM and potential genetic contribution to risk. Estimated heritability of OM (the proportion of risk attributable to additive genetic factors) has been shown to vary from 0.2 to 0.73 in various populations.<sup>52,59-61,245,246</sup> Although the estimated heritability of OM is high, there have been few studies to discover new putative loci or genes. Identification of OM candidate genes may provide new understanding of the etiology of OM, increase precision for

disease prediction, as well as provide novel targets for prevention and treatment. This impact could reduce the quantity of antibiotics prescribed and the number of pediatric surgeries.

Currently, there are few studies that have the ability to study the genetics of OM. Prior genome-scale studies of OM include a linkage study from the University of Pittsburgh (UPitt)<sup>112</sup>, a GWAS from the University of Western Australia (UWA)<sup>115</sup>, and both linkage and GWAS from the University of Minnesota (UMN).<sup>24,247</sup> The first GWAS of OM was conducted in Western Australia using the Western Australian Pregnancy (Raine) cohort to determine variants associated with acute and chronic OM. The Raine cohort was derived from a longitudinal birth cohort, and cases were defined if clinical exam indicated the presence of inflamed, retracted, or scarred tympanic membrane; middle ear effusion; or tympanostomy tube surgery in the first three years of life. Parental self-report of at least three episodes of acute otitis media (AOM) by three years could also determine affected status. The Raine cohort analysis found no significant genome-wide association, but did identify two genes (*CAPN14* and *GALNT14*) as strong candidates from the GWAS, and another cluster of genes (*BPIFA*) in a gene-based analysis. The most significantly-associated SNPs were followed up in UWA's family study of OM (WAFSOM). A total of 20 SNPs within seven genes (*CAPN14*, *GALNT14*, *GALNT13*, *BMP5*, *NELL1*, *TGFB3*, and *BPIFA1*) were genotyped. None of these were found to be significantly associated with OM in the WAFSOM population. Though they did not find replication in this study, many of the candidates have biological plausibility including *CAPN14*, *GALNT14*, and the *BPIFA* gene cluster and may warrant further investigation in an independent population of OM.

The second GWAS of OM was conducted at the University of Virginia using the UMN family population to determine variants associated with COME/ROM.<sup>247</sup> Children were recruited into this study if they had tympanostomy tube surgery for COME/ROM, and their family members were also recruited. Participants were considered affected if they had at least two pieces of positive evidence of COME/ROM from an ear examination from an ENT, a tympanometric test, medical record, and self-reported history.<sup>24</sup> As in the case of the Raine cohort, no SNP exceeded genome-wide significance for association with COME/ROM. The most significantly associated SNP in this GWAS was rs1110060 in Kinesin family member 7 (*KIF7*). Significant replication of rs10497394 on chromosome 2 was found in the UPitt family population of OM.

A hallmark of gene discovery is replication of results. Fortunately, there is strong collaboration between groups in the genetics of OM. Critical to replication is the coordinated clinical phenotyping of study populations, as a failure to replication could be due to inconsistent phenotypes used for recruitment and lack of power due to sample size. This ongoing collaboration in the genetics of OM has achieved replication of UPitt's linkage at chromosome 10q26.3 in the UWA's family population of COME/ROM<sup>248</sup> and a SNP on chromosome 2 from UMN's GWAS in UPitt's family population of OM. With evidence of association in the Raine cohort GWAS of OM, it was logical to investigate replication of the most significant SNPs from the Raine GWAS in the UMN family population.

### 5.2.3 Methods

This study was conducted with Institutional Review Board approval at the University of Minnesota, Wake Forest University, the University of Virginia, and the University of Pittsburgh, and adhered to the tenets of the Declaration of Helsinki.

#### *5.2.3.1 SNP Identification for Replication*

We identified 46 autosomal SNPs to follow up in this replication study. The Raine cohort GWAS reported a list of their top 25 statistically significant SNPs from their GWAS and a list of the top 25 statistically significant SNPs from a subset of participants with full covariate data. Three SNPs overlapped these two lists and one was located on the X chromosome, resulting in a total of 46 autosomal SNPs.

#### *5.2.3.2 Investigation of Replication: Data mining and genotyping*

From these 46 SNPs, we searched our GWAS of COME/ROM data to determine if any associations had already tested. To investigate the remaining SNPs from the Raine GWAS results, we genotyped the SNPs in the UMN family population using the Fluidigm SNPtype assay platform. Briefly, 100ng/uL of each DNA sample was combined with Biotium 2X Fast Probe Master Mix, SNPtype Sample Loading Reagent (Fluidigm), and the reference dye ROX (Invitrogen Inc.). Each SNP type assay was mixed with 2X Assay Loading Reagent (Fluidigm). Both sample mixes and assay mixes were then loaded onto Fluidigm 96.96 Dynamic Genotyping Arrays, and nanofluidic circuitry loads and mixes the 96 loci with 96 samples in 9216 reaction chambers. Thermocycling was performed

on Fluidigm's Stand-alone Thermal Cycler and fluorescence detection performed on the EP1 genotyping system (Fluidigm).

SNP rs2704219, which did not pass quality control measures, was re-genotyped after a pre-amplification step performed with the following thermocycling conditions: 95°C for 15 minutes, then 14 cycles of 5 seconds at 95°C and 4 minutes at 60°C. Pre-amplified DNA was diluted 1:100 in suspension buffer and genotyped as previously described.

#### *5.2.3.3 Quality control measures and Association Test*

Quality control of SNPs after genotyping was based upon filtering for poor genotype clusters, low minor allele frequency ( $MAF < 0.01$ ), and genotypes inconsistent with Hardy Weinberg proportions ( $P < 10^{-5}$ ). Samples and SNPs with excessive Mendelian errors ( $> 1\%$ ) and/or low genotype call rates ( $< 95\%$ ) were removed. Genotype data from these 26 SNPs was analyzed using the family based association test QLSw.<sup>205</sup>

### **5.2.4 Results and Discussion**

Initially, we searched our GWAS data for any SNPs that were already genotyped. We found no significant associations from these 21 SNPs (**Table 5.1**) with P-values ranging from  $P = 0.11$  (rs4627412 in *RPTOR*) to  $P = 0.92$  (rs4709819, intergenic near *QK1*). To investigate the remaining SNPs from the Raine GWAS results, we aimed to genotype the SNPs ( $N = 27$ ) in the UMN family population using the Fluidigm SNPtype assay platform, including one SNP (rs10242197) which had been genotyped in our GWAS for quality

**Table 5.1:** Top SNPs from the Raine cohort GWAS that were genotyped in our GWAS of COME/ROM

Chr	SNP	Position (hg18)	Nearest RefSeq Gene	Risk Allele	Other Allele	Risk Allele Freq	OR	P-value	Raine OR (95% CI)	Raine Risk Allele Freq	Raine P-value <sub>adj-PCA</sub>
1	rs1175549	3.681587	<i>SMIM1</i>	A	C	0.84	0.79	0.23	1.56 (1.27-1.91)	0.76	2.65x10 <sup>-5</sup>
2	rs13386850	31.2992	<i>CAPN14</i> *	C	A	0.09	1.38	0.19	1.86 (1.44-2.38)	0.10	1.32x10 <sup>-6</sup>
2	rs2098787^	31.15015	<i>GALNT14</i>	G	A	0.51	1.08	0.56	1.58 (1.28-1.97)	0.53	3.17x10 <sup>-5</sup>
2	rs1862981^	31.15103	<i>GALNT14</i>	T	G	0.49	0.91	0.45	1.60 (1.29-1.99)	0.54	2.20x10 <sup>-5</sup>
2	rs2377445^	106.0271	<i>C2orf40</i> *	C	T	0.71	1.02	0.88	1.57 (1.26-1.96)	0.70	5.20x10 <sup>-5</sup>
4	rs10008015	106.2247	<i>TET2</i> *	C	T	0.09	1.06	0.8	1.66 (1.30-2.11)	0.11	3.89x10 <sup>-6</sup>
4	rs6826919^	113.0233	<i>C4orf32</i> *	A	G	0.42	1.18	0.24	1.59 (1.27-1.96)	0.43	4.05x10 <sup>-5</sup>
6	rs1457955^	77.46879	<i>IMPG1</i> *	G	T	0.85	0.96	0.84	1.85 (1.37-2.49)	0.87	4.98x10 <sup>-5</sup>
6	rs4709819^	164.3833	<i>QKI</i> *	A	G	0.45	0.98	0.92	1.59 (1.28-2.0)	0.40	2.67x10 <sup>-6</sup>
7	rs10242197	90.09785	<i>CDK14</i> *	C	T	0.82	0.83	0.3	1.59 (1.27-1.98)	0.19	4.38x10 <sup>-5</sup>
7	rs10488001	90.49398	<i>CDK14</i>	T	C	0.09	0.89	0.6	1.72 (1.33-2.24)	0.09	4.53x10 <sup>-5</sup>
8	rs2882460^	62.68845	<i>ASPH</i>	C	A	0.81	0.81	0.22	1.66 (1.30-2.12)	0.79	4.96x10 <sup>-5</sup>
8	rs6471969^	62.72939	<i>ASPH</i>	G	T	0.84	0.81	0.28	1.83 (1.42-2.37)	0.82	3.85x10 <sup>-6</sup>
8	rs11990408^	62.74755	<i>ASPH</i>	A	G	0.83	0.84	0.35	1.76 (1.27-2.27)	0.81	1.10x10 <sup>-5</sup>
8	rs11787089^	62.78339	<i>ASPH</i>	C	T	0.87	0.80	0.29	1.88 (1.43-2.47)	0.85	6.83x10 <sup>-6</sup>
10	rs4575213	53.09533	<i>PRKG1</i>	C	A	0.62	0.96	0.8	1.41 (1.20-1.67)	0.60	4.65x10 <sup>-5</sup>
13	rs1336708^	101.763	<i>FGF14</i>	G	A	0.21	0.90	0.51	1.75 (1.33-2.27)	0.23	5.17x10 <sup>-5</sup>
13	rs4512966^	109.8801	<i>COL4A2</i>	C	T	0.53	0.88	0.34	1.55 (1.25-1.91)	0.50	5.01x10 <sup>-5</sup>
17	rs4627412^	76.13884	<i>RPTOR</i>	A	G	0.65	1.25	0.11	1.58 (1.27-1.96)	0.69	4.92x10 <sup>-5</sup>
21	rs2839520^	42.74273	<i>UBASH3A</i> *	G	A	0.39	1.18	0.25	1.59 (1.30-1.96)	0.46	1.63x10 <sup>-5</sup>

We report our OR and AF using the risk allele corresponding to the Risk Allele listed in the Raine cohort GWAS.<sup>115</sup>

\*Indicates SNP is intergenic and therefore reports the nearest gene. ^Indicates the SNP was a top genotyped SNP from a subset of Raine study participants with full covariate data

control purposes. One SNP (rs2704219 in *ALDH1A2*), did not pass quality control measures; however, rs2704219 was successfully re-genotyped using a different Fluidigm protocol to pre-amplify samples before genotyping. A total of 26 SNPs were genotyped in 596 subjects and available for analyses.

Data was analyzed using the family based association test QLSw.<sup>205</sup> Again, no significant associations with any SNP with OM status were found after Bonferroni correction (**Table 5.2**), with nominal P-values ranging from  $P=0.11$  (rs7846684, intergenic between *GSDMC* and *PVT1*) to  $P=0.88$  (rs330787 in *FBXO11* and rs11658297, intergenic between *FAM20A* and *ABCA8*). Comparing the OR observed in our GWAS versus the Raine cohort GWAS, we observed 17/46 SNPs (37%) with the same direction of effect. Additionally, in the Raine GWAS, a top SNP from imputation analysis (rs6755194: OR=1.90; 95% CI 1.47-2.45) resulted in higher significance ( $P\text{-value}_{\text{adj-PCA}}=8.3\times 10^{-7}$ ) than their most associated genotyped SNP located in the *CAPN14* gene. This SNP was included in our analysis. This SNP was genotyped in our GWAS (OR=1.26; 95% CI 1.21-1.31;  $P\text{-value}=0.36$ ) and was not found to be associated with COME/ROM. The SNP genotyped both in our GWAS and this replication project (rs10242197) had consistent results (**Table 5.1, Table 5.2**). Lack of replication in our population is likely due to similar reasons as were listed in the Raine GWAS paper including phenotypic heterogeneity and sample size.<sup>115</sup> Inclusion criteria for case classification were less stringent in the Raine cohort than in classification of cases in the UMN family population.



**Table 5.2:** Top SNPs from Raine cohort GWAS that were genotyped in our family population of COME/ROM.

Chr	SNP	Position	Nearest	Risk	Other	Risk	OR	P-value	Raine OR	Raine Risk	Raine
1	rs12728900	26.746807	<i>LIN28A</i>	A	G	0.20	0.84	0.27	1.47 (1.23-1.75)	0.25	2.28x10 <sup>-5</sup>
2	rs13408922	31.444826	<i>CAPN14</i> *	A	C	0.12	1.31	0.26	1.86 (1.44-2.38)	0.10	1.32x10 <sup>-6</sup>
2	rs13386745	31.445615	<i>CAPN14</i> *	G	A	0.12	1.26	0.33	1.84 (1.44-2.37)	0.10	1.63x10 <sup>-6</sup>
2	rs330787	48.041377	<i>FBXO11</i>	C	T	0.63	1.02	0.88	1.43 (1.21-1.70)	0.64	4.64x10 <sup>-5</sup>
3	rs17624623^	61.620466	<i>PTPRG</i>	T	C	0.58	0.93	0.58	1.60 (1.29-1.98)	0.59	2.02x10 <sup>-5</sup>
4	rs11097383	94.58384	<i>GRID2</i>	C	T	0.17	0.77	0.12	1.55 (1.25-1.92)	0.16	5.97x10 <sup>-5</sup>
4	rs1859161	106.042692	<i>TET2</i> *	G	A	0.12	1.16	0.46	1.78 (1.37-2.32)	0.09	1.90x10 <sup>-5</sup>
4	rs11940126	186.790909	<i>SORBS2</i>	A	G	0.05	0.94	0.85	2.09 (1.46-3.00)	0.05	5.70x10 <sup>-5</sup>
7	rs10242197	90.259912	<i>CDK14</i> *	C	T	0.81	0.86	0.42	1.59 (1.27-1.98)	0.81	4.38x10 <sup>-5</sup>
8	rs1496306	50.540708	<i>SNUG1</i> *	A	G	0.08	0.73	0.16	1.76 (1.36-2.28)	0.10	1.63x10 <sup>-5</sup>
8	rs2132528	50.605655	<i>SNUG1</i> *	G	A	0.08	0.72	0.14	1.75 (1.35-2.26)	0.09	2.41x10 <sup>-5</sup>
8	rs13438948	87.858778	<i>CNBD1</i> *	A	G	0.07	1.40	0.22	1.82 (1.39-2.39)	0.09	1.19x10 <sup>-5</sup>
8	rs7846284	130.370235	<i>GSDMC</i> *	A	G	0.42	0.82	0.12	1.40 (1.19-1.65)	0.43	4.06x10 <sup>-5</sup>
8	rs7846684	130.370414	<i>GSDMC</i> *	C	T	0.42	0.81	0.11	1.41 (1.20-1.65)	0.43	3.60x10 <sup>-5</sup>
9	rs11790808	94.588498	<i>ROR2</i>	C	T	0.04	0.93	0.79	2.09 (1.50-2.91)	0.05	1.15x10 <sup>-5</sup>
10	rs16919668^	20.279158	<i>PLXDC2</i>	G	A	0.15	1.16	0.47	2.0 (1.43-2.78)	0.16	3.28x10 <sup>-5</sup>
10	rs10884043	106.531703	<i>SORCS3</i>	A	G	0.77	1.19	0.25	1.53 (1.25-1.86)	0.74	2.74x10 <sup>-5</sup>
10	rs4556466	106.558025	<i>SORCS3</i>	A	G	0.77	1.21	0.22	1.53 (1.26-1.87)	0.74	2.74x10 <sup>-5</sup>
13	rs9564897	72.912338	<i>MZT1</i> *	T	G	0.34	0.90	0.46	1.43 (1.22-1.69)	0.34	1.98x10 <sup>-5</sup>
15	rs6493973^	58.292636	<i>ALDH1A2</i>	C	T	0.97	0.82	0.59	3.43 (1.90-6.20)	0.97	4.40x10 <sup>-5</sup>
15	rs2218261^	58.294022	<i>ALDH1A2</i>	G	A	0.97	0.82	0.59	3.43 (1.90-6.20)	0.97	4.40x10 <sup>-5</sup>
15	rs2704219^	58.328448	<i>ALDH1A2</i>	T	C	0.97	0.74	0.45	3.39 (1.90-6.05)	0.97	3.72x10 <sup>-5</sup>
17	rs11658127^	66.714282	<i>FAM20A</i> *	A	G	0.88	0.91	0.65	1.93 (1.42-2.61)	0.87	2.29x10 <sup>-5</sup>
17	rs11658297^	66.728983	<i>FAM20A</i> *	G	A	0.88	0.97	0.88	1.94 (1.43-2.63)	0.87	1.89x10 <sup>-5</sup>
17	rs9911978^	78.524407	<i>RPTOR</i>	A	G	0.68	1.21	0.18	1.58 (1.27-1.96)	0.69	4.92x10 <sup>-5</sup>
20	rs17396317	31.790377	<i>BPIFA4P</i> **	A	G	0.12	1.22	0.32	1.62 (1.28-2.05)	0.12	5.15x10 <sup>-5</sup>

We report our OR and AF using the risk allele corresponding to the Risk Allele listed in the Raine cohort GWAS.<sup>115</sup>

\*Indicates SNP is intergenic and therefore reports the nearest gene. \*\*Indicates SNP is coding. ^Indicates the SNP was a top genotyped SNP from a subset of Raine study participants with full covariate data.

Due to the complexity of this disease, it is exceedingly important that study sites use the same definitions of affected status so that replication studies and meta-analyses can be successfully conducted. Though this is a significant task to achieve, the International Consortium of the Genetics of Otitis Media (OTIGEN) is striving to standardize inclusion criteria for OM studies to better understand the genetic components increasing risk for OM. Additionally, for these population studies to be more informative, recruitment of diverse populations is necessary. The majority of OM population studies thus far have recruited participants with European ancestry, but diverse populations will enable researchers to narrow in on regions of linkage disequilibrium (LD). Narrowing the regions of high LD will get us closer to the underlying causal variants. Another issue is that OM controls are very difficult to find since most children will have at least one episode of OM in the first three years of life. Trio and family-based approaches could provide a solution to the problem of OM controls.

In this study, we attempted to replicate associations from the Raine cohort GWAS of OM in our family population of COME/ROM recruited at the University of Minnesota. Though we did not find any of these SNPs to be significantly associated in our population, this study illustrates the need for larger, diverse populations recruited with standardized inclusion criteria for cases.

## Chapter 6

### GWAS of COME/ROM follow up study to identify causal variants

#### 6.1 Chapter Introduction

Genome-wide association studies have been utilized to detect variants associated with a specific trait or disease in populations of interest.<sup>249</sup> These studies have been successful in identifying common SNPs with modest to small effects that impact risk of complex diseases, but are underpowered to detect rare variants. The causal variant is also rarely associated in GWA studies but is in LD with the associated SNP.<sup>250</sup> In many cases, the functional role of the genes and variants identified in GWA studies is unknown, especially in context of a specific disease. Finally, most variants associated in GWA studies are non-coding variants therefore requiring follow-up studies to translate findings into biological knowledge which can lead to therapeutic advances.<sup>251</sup> In our study, we investigated those regions which either approached significance in our discovery GWAS or were replicated in an independent population.<sup>247</sup>

Chapter 4 describes our discovery GWAS of COME/ROM. Though no SNP exceeded genome-wide significance, our strongest signal on chromosome 15 was supported by multiple SNPs and fell in a biologically plausible gene, *KIF7*, and replication was achieved at SNP rs10497394 in a large intergenic region on chromosome 2. Our follow-up study to determine the causal nature of these regions used a multi-tiered approach. Initially, we utilized targeted resequencing of regions of interest to discover SNPs in a subset of family members from our COME/ROM family-based population. This was followed-up by genotyping variants detected to localize the signals. Analysis using a within-family association test resulted in six significantly associated SNPs. We investigated the regions and genes containing these SNPs using relevant functional assays.

## **6.2 Honing in on genetic risk of chronic otitis media with effusion and/or recurrent otitis media (COME/ROM): a GWAS follow-up study**

E. Kaitlynn Allen<sup>1,2</sup>, Ani Manichaikul<sup>3</sup>, Stephen R. Williams<sup>1,4</sup>, Keith L. Keene<sup>5</sup>, Josyf C. Mychaleckyj<sup>1,3</sup>, Aaron Quinlan<sup>1,2,3</sup>, Wei-Min Chen<sup>1,3</sup>, David Wotton<sup>2</sup>, Jeremy Thorpe<sup>2</sup>, Mazhar Adli<sup>2</sup>, Stephen S. Rich<sup>1,3,6</sup>, Kathleen A. Daly<sup>7</sup>, Michèle M. Sale<sup>1,2,6</sup>

<sup>1</sup>Center for Public Health Genomics, <sup>2</sup> Department of Biochemistry and Molecular Genetics, <sup>3</sup>Department of Public Health Sciences, <sup>4</sup>Cardiovascular Research Center, University of Virginia, Charlottesville, VA 22908; <sup>5</sup>Department of Biology, East Carolina University, Greenville, NC 27858; <sup>6</sup>Department of Medicine, University of Virginia, Charlottesville, VA 22908; <sup>7</sup>Department of Otolaryngology, University of Minnesota, Minneapolis MN 55455.

E. Kaitlynn Allen carried out molecular genetic studies: capture, targeted resequencing, Fluidigm genotyping, and functional assays, analyzed sequence data for variants using Unified Genotyper and freebayes, aided in ChIP experiments, conducted statistical analyses for all functional assays, and drafted the manuscript. E. K. Allen and Dr. Michèle M. Sale conceived of study design for fine-mapping and functional assays. Drs. Kathleen Daly and Stephen Rich conceived and designed the UMN family-based study, and Dr. Daly coordinated recruitment. Dr. Ani Manichaikul conducted statistical analyses of genotype data. Dr. Josyf C. Mychaleckyj carried out sequence data QC analysis, alignment, and variant calling with mpileup. Dr. Keith L. Keene advised on targeted resequencing. Dr. Aaron Quinlan advised on sequence analysis and variant calling. Dr. Stephen R. Williams advised on functional assays. Dr. Wei-Min Chen carried out statistical analyses of prior GWAS data. Dr. David Wotton advised on functional assays, especially TGF $\beta$  treatment.

Dr. Mazhar Adli advised on ChIP-seq experiments. Mr. Jeremy Thorpe carried out ChIP of HMEEC-1 cells, initial QC and alignment. Dr. Sale coordinated genetic studies and analyses.

### 6.2.1 Abstract

**Background:** Few genome-wide studies of otitis media (OM) have been conducted, and little is known about genetic susceptibility to OM. We previously carried out a GWAS of chronic OM with effusion and recurrent OM (COME/ROM) and successfully replicated a locus on chromosome 2 in an independent OM population. The most significantly associated SNP lies in kinesin family member 7 (*KIF7*), a member of the Hedgehog (HH) signaling pathway located on chromosome 15. This study aimed to fine-map and functionally explore these two high priority loci from our GWAS.

**Methods:** We used targeted resequencing in the regions of association on chromosomes 2 and 15 in a subset of family members (N=100). SNPs (N=240) identified through this approach, supplemented with those from a region with low coverage, were genotyped in the full family-based population of COME/ROM (N=596). The chromosome 2 region was evaluated for enhancer activity using ChIP-seq in a human middle ear epithelial cell line, HMEEC-1. We tested the most associated SNP for its potential to affect *IQGAP1* gene expression in adenoids. *KIF7* was investigated for response to TGF $\beta$  and lipopolysaccharide (LPS).

**Results:** We detected 303 SNPs in the chromosome 2 and 15 regions, and 240 were genotyped in all COME/ROM families. Association tests using QLSw found two associated SNPs on chromosome 2 and four on chromosome 15. Investigation of chromatin marks revealed an enhancer region on chromosome 2 in HMEEC-1. The “G” risk allele of the most significantly-associated SNP, rs1110060, was found to impact expression of *IQGAP1*

in adenoids. Also, both LPS and TGF $\beta$  treatment were found to regulate expression of *KIF7* in HMEEC-1.

**Conclusions:** Our fine-mapping honed the two regions of association, with the most associated GWAS SNP – rs1110060 – remaining the strongest signal. Although the associated chromosome 2 SNPs lie in a large intergenic region, confirmation of enhancer activity via ChIP-seq suggests this region may impact regulation of distant genes.

Influences on *IQGAP1* expression by rs1110060 genotype establishes a plausible role of this variant. Our study links *KIF7* and the HH signaling pathway with COME/ROM for the first time, and points to roles in response to infection and TGF $\beta$  signaling. Further study is required to elucidate the details of the mechanisms by which the chromosome 2 region and *KIF7* and *IQGAP1* genes increase susceptibility to COME/ROM.

### 6.2.2 Introduction

Otitis media (OM) – inflammation of the middle ear (ME) – is one of the most common diseases in children and is responsible for over \$5 billion in direct and indirect costs.<sup>1</sup> In a subset of children with OM, persistent and/or subsequent infections can occur, which are referred to as chronic otitis media with effusion (COME) and recurrent otitis media (ROM), or COME/ROM indicating the presence of either. According to the World Health Organization, deafness in children due to chronic OM often leads to long term effects on early communication, language, auditory processing, psychosocial and cognitive development, and education.<sup>118</sup> OM is the leading cause for antibiotic treatment in children in the United States.<sup>55</sup> The other main treatment option for COME/ROM is tympanostomy tube surgery, the most common pediatric surgery.<sup>56</sup> Understanding why some children are susceptible to chronic ME infection is important for clinical knowledge and treatment.

Little is known about genetic susceptibility to OM. To date, only two linkage studies<sup>24,112</sup> and two genome-wide association studies (GWAS)<sup>115,247</sup> have been conducted to discover risk variants for OM. Replication of results from these studies have been minimal, with the only successful replication evidence to date from the 10q22.3 linkage peak<sup>248</sup> and association at rs10497394 on chromosome 2<sup>247</sup>. Few opportunities to evaluate replication exist due to a lack of available genetic data and limited sample sizes.

Our prior GWAS of COME/ROM revealed novel regions associated with susceptibility to OM.<sup>247</sup> The only SNP replicated in an independent population (rs10497394) lies in a large intergenic region on chromosome 2.<sup>247</sup> ENCODE data support regulatory elements throughout this region. The most significantly associated SNP in our discovery GWAS, rs1110060, lies on chromosome 15 and is located in an intron of kinesin family member 7 (*KIF7*), a member of the hedgehog (HH) signaling pathway. Prior studies have shown that treatment with bacterial toxins LPS<sup>252</sup> or ExoS<sup>253</sup>, as models of Gram negative bacterial infection, downregulate HH signaling pathway genes, including *KIF7*, and induces cytotoxicity<sup>253</sup>. In eQTL studies of lymphoblastoid cell lines, SNP rs1110060 has also been found to regulate expression of Ras GTPase-activating-like protein (*IQGAP1*), a scaffolding protein known to be involved in bacterial adherence and invasion. OM mouse models *Jeff*<sup>179</sup>, *Junbo*<sup>72</sup> and the *Tgfr1* knockout<sup>73</sup> indicate the TGF $\beta$  signaling pathway is an important player in OM pathogenesis. This has been supported by human candidate gene<sup>95,96</sup> and GWAS<sup>115</sup>.

The goals of our current study were to fine map the two highest priority regions from our GWAS, and investigate potential functional contributions to OM risk. For *in vitro* assays, we used the human middle ear epithelium cell line (HMEEC-1) because of its relevance to

ME infection and response. Expression was also explored in adenoid tissue, used previously by others to evaluate candidate genes from the 10q22.3 linkage peak.<sup>248</sup> These approaches were applied to provide support for novel yet plausible roles of these loci in COME/ROM susceptibility.

### **6.2.3 Methods**

#### **6.2.3.1 University of Minnesota (UMN) family-based study**

##### *6.2.3.1.1 Ethics Statement*

This study was conducted with Institutional Review Board approval at the University of Minnesota, Wake Forest University, the University of Virginia, and the University of Pittsburgh, and adhered to the tenets of the Declaration of Helsinki.

##### *6.2.3.1.2 Study design*

The UMN family-based population of COME/ROM has been described in detail previously.<sup>24,244</sup> Briefly, probands, children who had tympanostomy tube surgery for COME/ROM, and their family members were recruited. Affected status of family members was determined using four data sources including ear examination, tympanogram, medical records, and self-reported COME/ROM history.

#### **6.2.3.2 Targeted Sequencing of Regions of Interest**

##### *6.2.3.2.1 Regions sequenced*

The two regions sequenced included chromosome 2 and chromosome 15 COME/ROM-associated loci from our prior GWAS<sup>247</sup>. An additional 3Mb region from a separate investigation of COME/ROM was sequenced concurrently and is not discussed further.



The regions sequenced from the GWAS were determined using a haplotype analysis of HapMap CEU samples using Haploview<sup>220</sup> (Gabriel block method). The haplotype blocks containing the associated SNPs from the GWAS were sequenced and included chr2: 174286992-174307309 and chr15: 90098150-90200735 (Build 37.) The chromosome 2 region is intergenic, while the chromosome 15 region contains two genes: *KIF7* and TOPBP1-interacting checkpoint and replication regulator (*TTCRR*).

#### *6.2.3.2.2 Sample selection and pooling for sequencing*

For sequencing, at least two members from 45 families from the UMN family-based study were included in this analysis. Initial concentration of each sample was measured using Quant-iT Picogreen (Invitrogen, Carlsbad, CA). Sample quality was measured using a NanoDrop 8000 Spectrometer (Thermo Fisher Scientific, Waltham, MA). Samples were pooled into ten pools, with ten individuals per pool. To pool the ten samples, 3ng DNA from each sample was combined. Members of each family were contained within the same pool to enrich a single pool for private and/or rare variants. Each of the ten pools was then combined in equimolar ratios into one sample for sequencing in a single lane.

#### *6.2.3.2.3 Library preparation and indexing*

Library construction of each pool (of 10 samples) was carried out individually, then each pool was indexed prior to sequencing using Agilent's SureSelect<sup>XT</sup> Target Enrichment System for Illumina Paired-End Sequencing Library, SureSelect<sup>XT</sup> Target Enrichment for Illumina Multiplexed Sequencing, Version 1.0, November 2010, according to manufacturer's instruction (Agilent, San Francisco, CA), and resulted in a captured region from chromosomes 2 and 15 of 1.3Mb. Briefly, DNA from each pool was sheared using a

Covaris S-series Single Tube Sample Preparation system (Covaris, Woburn, MA). The target peak for base pair size was 150-200bp, which was assessed after purification with AMPure XP beads using an Agilent 2100 Bioanalyzer and a Bioanalyzer DNA 1000 chip. The prepared, adapter-ligated libraries were then purified with AMPure XP beads and assessed with the Bioanalyzer DNA 1000 chip for peak size approximately 250-275 bp. DNA was hybridized with the Custom RNA baits, and the library was captured using 50µL Dynal Dynabeads® M-280 Streptavidin (Invitrogen, Carlsbad, CA), washed and resuspended in 200µL SureSelect Binding Buffer with *at least* 20µL hybridized library. Each captured library was then purified with AMPure XP beads. Index tags were added to each of the pools using post-hybridization amplification. Samples were purified with AMPure XP beads and assessed for quality and size range (approximately 300-400 nucleotides) with a Bioanalyzer High Sensitivity DNA Assay. Concentration of each pool was quantified using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA), and then pooled to a final concentration of 10nM and final volume of 20µL.

#### 6.2.3.2.4 Sequencing

Sequencing of the 10 pooled libraries (100 samples) was carried out at the Genomics Services Lab at HudsonAlpha Institute for Biotechnology (Huntsville, AL) using the Illumina HiSeq 2000 platform. Quality control measures carried out at Hudson Alpha prior to sequencing included DNA quantification using a Qubit® 2.0 Fluorometer (Invitrogen, Carlsbad, CA), bioanalysis using Agilent's 2100-High Sensitivity assay to determine the library peak size and to determine if there were multiple peaks, and Kapa RT-PCR to validate concentration of the sample. The pooled sample (15ng/µL) was sequenced using 100bp, paired-end reads on a single lane of the HiSeq 2000.

#### 6.2.3.2.5 *Primary analysis*

Alignment of the paired-end sequence data to the reference genome (hg19) was carried out using Novocraft's Novoalign (v2.07.10; <http://www.novocraft.com/main/index.php>).

Bam files were then sorted, and reads filtered based on whether they were properly paired and using a minimum MapQ value of 30. Duplicate reads were removed using Picard's MarkDuplicates algorithm. Coverage statistics were calculated for each of the sequenced regions.

#### 6.2.3.2.6 *Variant calling*

In order to improve and compare SNP detection from our sample of multiple pools, we evaluated results from three complementary SNP discovery approaches: SamTools<sup>254</sup> mpileup, GATK Unified Genotyper<sup>255</sup>, and Freebayes<sup>256</sup>. We ran mpileup with its default parameters, limiting the analysis to the genomic coordinates of the sequenced regions. Parameters used in the Unified Genotyper analysis included a minimum base quality score of 30 (`--min_base_quality_score 30`), a minimum phred-scaled confidence threshold at which variants can be called (`--stand_call_conf 30`), and a minimum phred-scaled confidence threshold at which variants will be emitted (`--stand_emit_conf 30`). Importantly, Freebayes was written specifically for datasets with pooled data and accounts for ploidy, enabling identification of rare SNPs not identified by other tools. Freebayes version 0.5.1 was used, and parameters included a minimum of 100x coverage (`--min-coverage 100`), a minimum base quality of 30 (`--min-base-quality 30`), at least 5 reads had to have the SNP to avoid spurious variation caused by sequencing artifacts (`--min-alternate-total 5`), we did not calculate the marginal probability of genotypes (`--no-marginals`), and ploidy was set to 10 individuals per pool (`--ploidy 20`). Analyses for both

Unified Genotyper and Freebayes were limited to the genomic coordinates of the RNA baits created for capture.

### **6.2.3.3 Fine-mapping of chromosome 2 and chromosome 15 regions**

#### *6.2.3.3.1 SNP prioritization for genotyping*

All SNPs detected in the chromosome 2 and chromosome 15 regions by one or more of the three SNP detection programs were collated. Indels and tri-allelic SNPs were not investigated due to technical challenges of obtaining reliable genotype data. SNPs that had been genotyped in our GWAS of COME/ROM were not regenotyped, except for two SNPs from the chromosome 2 region (rs10497394 and rs16861690) and two SNPs from the chromosome 15 region (rs10775247 and rs1110060), included as QC measures. One segment of the chromosome 15 region (chr15:90166854-90172852) was not well-covered by sequence data but contained reported, non-synonymous variants; therefore, we added seven SNPs from this region to our genotyping panel. Additionally, we included SNPs from a different OM GWAS performed in Western Australia<sup>115</sup> (N=27) to evaluate replication, reported in Chapter 5. In total, 298 SNPs were submitted to the Fluidigm SNPtype Assay Design service. Two hundred sixty-seven SNPs were considered designable and included in the panel.

#### *6.2.3.3.2 SNP Genotyping*

Using Fluidigm's SNPtype Assays<sup>257</sup>, 240 SNPs were genotyped in 716 individuals from our family population of COME/ROM (<http://www.fluidigm.com/snp-type-assays.html>) using Fluidigm's SNP Genotyping Analysis Software with manual clustering and calling.

Samples were not pre-amplified, but pre-amplification was used subsequently on samples which had high missing rates (>5%) in the initial genotyping project.

#### *6.2.3.3 Data QC and Association Analysis*

Samples were removed if they were monomorphic, had missing genotypes >5%, Mendelian error >3, or no genotyped/phenotyped relative pairs. Association analysis of the genotype data was conducted using the family-based association test QLSw.<sup>205,247</sup> Haplotype analysis was conducted with our data using Haploview<sup>220</sup> (Version 4.2) to determine and visualize linkage disequilibrium between genotyped SNPs.

### **6.2.3.4 Assays to investigate potential functions of associated variants**

#### *6.2.3.4.1 Relevant cell line and tissue used for molecular assays*

The human middle ear epithelial cell line (HMEEC-1)<sup>103</sup>, kindly provided by Dr. David Lim (House Ear Institute, Los Angeles, CA), is the most relevant cell line for *in vitro* assays related to OM. HMEEC-1 cells were maintained in 1:1 mixture of Dulbecco's modified Eagle's medium (DMEM) (Invitrogen, Carlsbad, CA) and bronchial epithelial basal medium (BEBM) (Lonza, Basel, Switzerland). The mixture was supplemented with BEGM™ BulletKit™ SingleQuots™ (Lonza, Basel, Switzerland) which includes bovine pituitary extract, hydrocortisone, hEGF, epinephrine, transferring, insulin, triiodothyronine, retinoic acid, gentamicin, and amphotericin B. Cells were maintained at 37°C in a carbon dioxide-enriched (95% air, 5% CO<sub>2</sub>) incubator.

Adenoids were collected from 43 children undergoing adenoidectomies for hypertrophic adenoids, chronic otitis media, and obstructive sleep apnea. Since these were obtained as discarded surgical tissues, no clinical information was able to be accessed for these patients. Adenoids were immediately transported to the lab on ice in a sterile container. After being cut into small cubes approximately 8mm<sup>3</sup>, they were frozen at -80°C in RNAlater (Qiagen, Valencia, CA). Before DNA and RNA extraction, adenoid cells were lysed by incubating at 60°C in RLT buffer with  $\beta$ ME for 15 minutes, and homogenized using a tissue homogenizer (Omni International, Tulsa, OK).

#### *6.2.3.4.2 Enhancer activity of HMEEC-1: Chromatin immunoprecipitation sequencing (ChIP-seq) analysis*

Because our associated SNPs in the chromosome 2 region lie in a large intergenic region which encompasses ENCODE enhancer peaks<sup>258</sup>, we investigated enhancer activity in the HMEEC-1 cell line. ChIP experiments for H3K27Ac, H3K4me1, and H3K4me3 were carried out on wild type HMEEC-1 cells, and biological replicate experiments were conducted using experimental samples from different cell preparations. HMEEC-1 cell samples were subject to ChIP-seq, as described.<sup>259</sup> In brief, cells were crosslinked in 1% formaldehyde, lysed at room temperature for 12 minutes, and sonicated in a Branson 250 Sonifier to obtain chromatin fragments between 200 and 700bp. Sonicated samples were diluted 1:10 in ChIP dilution buffer and incubated with antibody at 4°C overnight. Dynabeads Protein A and Protein G (Life Technologies, Carlsbad, CA) were mixed in equal volumes and incubated with the chromatin-antibody complex for 2 hours at 4°C. The chromatin-antibody-Dynabead complex was precipitated using DynaMag-2 Magnet (Life Technologies, Carlsbad, CA) and washed twice with each of low salt, LiCl, and TE (pH 8.0) buffers. Enriched fragments were eluted at 65°C for 15 minutes in elution buffer, reverse

crosslinked at 65°C for 5 hours, treated with Proteinase K (1mg/ml; New England Biolabs, Ipswich, MA) for 2 hours at 37°C, and phenol-chloroform-isoamyl alcohol extracted and ethanol precipitated overnight at -20°C. The precipitant was reconstituted with 1:50 RNase A (10mg/ml; Thermo Fisher Scientific, Waltham, MA) and incubated at 37°C for 30 minutes. ChIP DNA was quantified with a Quantifluor dsDNA kit (Promega, Madison, WI) and Qubit (Invitrogen, Carlsbad, CA). ChIP was carried out in duplicate.

#### *6.2.3.4.3 Library Preparation and Sequencing*

Library preparation was carried out using the TruSeq ChIP Sample Preparation kit (Illumina, San Diego, CA) with minor modifications. An input of 1-10ng ChIP DNA was used for each library. End repair, 3' adenylation, and adapter ligation were performed using the standard protocol in accord with the TruSeq ChIP Sample Preparation Guide (August 2012). Adapter-ligated ChIP DNA fragments between 300 and 650bp were gel-purified and 18 cycles of PCR performed. Completed libraries were quantified with Quantifluor dsDNA kit and Qubit. Sequencing was carried out using Illumina's MiSeq Benchtop Sequencer and the MiSeq Reagent Kit v3 (150 cycle kit, single end) according to the manufacturer's instructions.

#### *6.2.3.4.4 ChIP-seq Data Analysis*

Sequence reads from each of the ChIP experiments were aligned to the human reference genome (hg19) using the BOWTIE alignment tool.<sup>260</sup> Duplicate reads were removed using SAMTOOLS<sup>254</sup> *rmdup* command. Processed reads were then converted into BigWig format<sup>261</sup> for viewing of the data in the UCSC Genome Browser (<http://genome.ucsc.edu>).

#### 6.2.3.4.5 Real time qPCR

RNA was extracted from adenoid samples and HMEEC-1 cells using the Qiagen RNeasy kit according to the manufacturer's instructions (Qiagen, Valencia, CA). Following extraction, RNA was converted to cDNA using qScript™ cDNA Synthesis Kit (Quanta Biosciences, Gaithersburg, MD). For cDNA conversion, 1µg RNA was mixed with 4µL 5X qScript Reaction Mix and was then cycled using the following parameters: 22°C for 5 minutes, 42°C for 30 minutes, 85°C for 5 minutes. Real-time qPCR was conducted using the ABI TaqMan Gene Expression Assay (Applied Biosystems, Life Technologies, Carlsbad, CA) for human mRNA expression. Briefly, the qPCR reaction mix was 5µL 2X TaqMan Universal PCR Master Mix, No AmpErase UNG, 0.5µL probe, and 2.5µL H<sub>2</sub>O with 2µL 50ng/µL cDNA. Expression of genes of interest (described below) was measured on the Applied Biosystems Prism 7900 HT using 95°C for 10 minutes and 40 cycles of 95°C for 15 seconds then 60°C for 1 minute. All samples were run in triplicate with *GAPDH* as an endogenous control. The cycle threshold (CT) values were determined for each reaction during PCR amplification. Relative expression levels were quantified with the  $\Delta\Delta CT$  method.

#### 6.2.3.4.6 *IQGAP1* eQTL assay

The genotypes of the 43 adenoid samples at chromosome 15 SNP rs1110060, previously shown to be an eQTL for *IQGAP1*<sup>213</sup>, were determined using Sanger sequencing. Primers used were 19F (5'-ACAGCTGAAAGCAGCCTCTC-3') and 20R (5'-TGTCCTCGACCTCATAACAG-3'). The PCR reaction mix included 12.5µL 2X Mango Mix (Bioline, Taunton, MA), 1µL F primer, 1µL R primer, and 8.5µL H<sub>2</sub>O with 2µL 20ng/µL DNA. The samples were amplified using the following conditions: 95°C for 5 minutes, 35



cycles of 95°C for 30 seconds, 53°C for 40 seconds, and 72°C for 15 seconds, then 72°C for 7 minutes. Amplified products were run on a 1% agarose gel. PCR products were purified using HT ExoSAP-IT® High-Throughput PCR Product Cleanup (Affymetrix, Inc., Santa Clara, CA); 5µL of product was mixed with 2µL ExoSAP-IT reagent and incubated at 37°C for 15 minutes, then 80°C for 15 minutes. Purified products were diluted to 4ng/µL and bi-directionally sequenced at ACGT, Inc (Wheeling, IL). Bioedit Sequence Alignment Editor (Version 7.2.3)<sup>262</sup> was used to determine the genotypes of rs1110060 for each adenoid sample.

Expression of *IQGAP1* in adenoid samples was measured using RT-PCR as described above. *IQGAP1* expression-level differences due to genotype were tested using a one-way ANOVA test. Because we observed a trend towards a reduction in *IQGAP1* expression in A/G and G/G genotypes compared to A/A, correlation of expression level by risk allele (G) carrier status was tested using a paired, one-tailed Student's T-test in GraphPad Prism.

#### 6.2.3.4.7 Evaluation of TGFβ signaling pathway regulation of *KIF7*

Prior studies have found the TGFβ pathway to be important for OM risk in humans<sup>95,96,115,98</sup> and mice<sup>71,72,179</sup>. *KIF7* lies within the Sonic hedgehog (SHH) pathway, which has been shown to interact with the TGFβ pathway.<sup>263,264</sup> Because of these connections, we sought to investigate whether the TGFβ pathway could regulate *KIF7*. HMEEC-1 cells were grown in HMEEC-1 media in 6-well plates to 80-90% confluency. Cells were treated with 100pM recombinant TGFβ (R&D Systems, Minneapolis, MN) for 3 hours, 24 hours, and 48 hours. Control cells were untreated. After treatment, the media was aspirated, cells were washed, then scraped using RLT lysis buffer (Qiagen, Valencia,

CA) and stored at -80 °C until RNA extraction. Each experiment was carried out in triplicate. RNA was extracted and converted to cDNA as described above. Relative expression of selected SHH pathway members (*KIF7*, GLI family zinc finger 1 (*GLI1*), Suppressor of fused homolog (*SUFU*), and Patched 1 (*PTCH1*)), TGFβ pathway member mothers against decapentaplegic homolog 2 (*SMAD2*), also associated with OM<sup>95,98</sup>, and positive controls (known TGFβ responsiveness genes) *PAI1* and *SMAD7* were measured as described above. Expression levels of each gene across the treatment time course were compared using a one-way ANOVA in GraphPad Prism.

#### 6.2.3.4.8 Evaluation of LPS treatment in the regulation of *KIF7*

Administration of LPS has been used extensively as a model of infection.<sup>252,265-268</sup> HMEEC-1 cells were grown in HMEEC-1 media in 6-well plates to 80-90% confluency. Cells were serum-starved for 24 hours, then treated with 100μg/mL Lipopolysaccharide (LPS) derived from *Pseudomonas aeruginosa* (Sigma-Aldrich, St. Louis, MO) for two hours and four hours, with untreated cells as controls. After treatment, RNA was extracted and converted to cDNA as described above. Relative expression of *KIF7*, *GLI1*, *SUFU*, and *PTCH1* were measured to determine whether LPS treatment impacted expression of HH signaling pathway members. *GAPDH* expression was measured as an endogenous control. Expression levels were compared using Student's T-test in GraphPad Prism.

## 6.2.4 Results

### 6.2.4.1 Fine mapping of GWAS signals

Sequencing on the HiSeq 2000 yielded 36.95 Gb of high-quality sequence data with an average read length of 216 bp  $\pm$  62. Average coverage was 224x per individual for the chromosome 2 region and 128x per individual for the chromosome 15 region. After alignment of sequences to the reference genome (Human; hg19) with Novoalign v2.07.10, quality control measures were performed, including removing reads that did not properly map, reads that did not properly pair (mapQ $\geq$ 30), and excluding duplicates to remove potential PCR bias. Mapping sequence reads resulted in 98.64% of reads properly mapped and 96.86% of reads properly paired (89.21% reads with mapQ $\geq$ 30). After these filters, there were 51.9 million reads (average reads per index: 5.19 million reads; median reads per index: 4.63 million reads) aligned to the genome for variant detection.

After indels were removed, Mpileup, Unified Genotyper, and Freebayes detected a total of 110 unique SNPs in the chromosome 2 region and 193 SNPs in the chromosome 15 region (303 total). Two hundred forty SNPs were designable on the Fluidigm platform, 26 of which were novel variants from our sequence data (**Table 6.1**). After initial genotyping, 96 SNPs with high missing data were pre-amplified and re-genotyped in 96 samples. From these genotyped SNPs (N=240), QC filters excluded 17 SNPs that did not cluster, 12 SNPs that exceeded missing data, HWE ( $P$ -value  $< 10^{-5}$ ) and/or Mendelian inconsistency ( $>3$ ) thresholds, and 21 SNPs with MAF $<0.01$ . After QC procedures, there were 190 SNPs available for analysis in 596 samples. Six SNPs were found to be significantly associated with COME/ROM after Bonferroni correction for 190 tests ( $P$ -value  $< 2.63 \times 10^{-4}$ ). These results are shown in **Table 6.2**. Two associated SNPs were in the chromosome 2 GWAS

**Table 6.1:** Categories of SNPs used for fine-mapping of GWAS association signals

<b>SNP category</b>	<b>Number of SNPs genotyped</b>
Chromosome 2 Sequence data	85
Chromosome 15 Sequence data	148
Chromosome 15 Annotation	7

**Table 6.2:** Most significant associations ( $P < 2.63 \times 10^{-4}$ ) from genotyping project using the within-family QLSw association test

Chr	Position (hg19)	SNP	Effect Allele	Other Allele	Allele Freq	OR	Z	QLSw P-Value
2	174,289,993	rs6433410	C	T	0.28	0.597	-3.81	$1.39 \times 10^{-4}$
2	174,302,727	rs10179575	G	C	0.23	0.598	-3.84	$1.22 \times 10^{-4}$
15	90,098,280	rs12442646	A	G	0.58	1.788	4.29	$1.82 \times 10^{-5}$
15	90,124,446	rs62023109	C	T	0.07	6.300	3.70	$2.13 \times 10^{-4}$
15	90,174,955	rs9672296	T	C	0.56	1.686	3.80	$1.43 \times 10^{-4}$
15	90,190,048	rs1110060	G	A	0.6	1.894	4.60	$4.32 \times 10^{-6}$

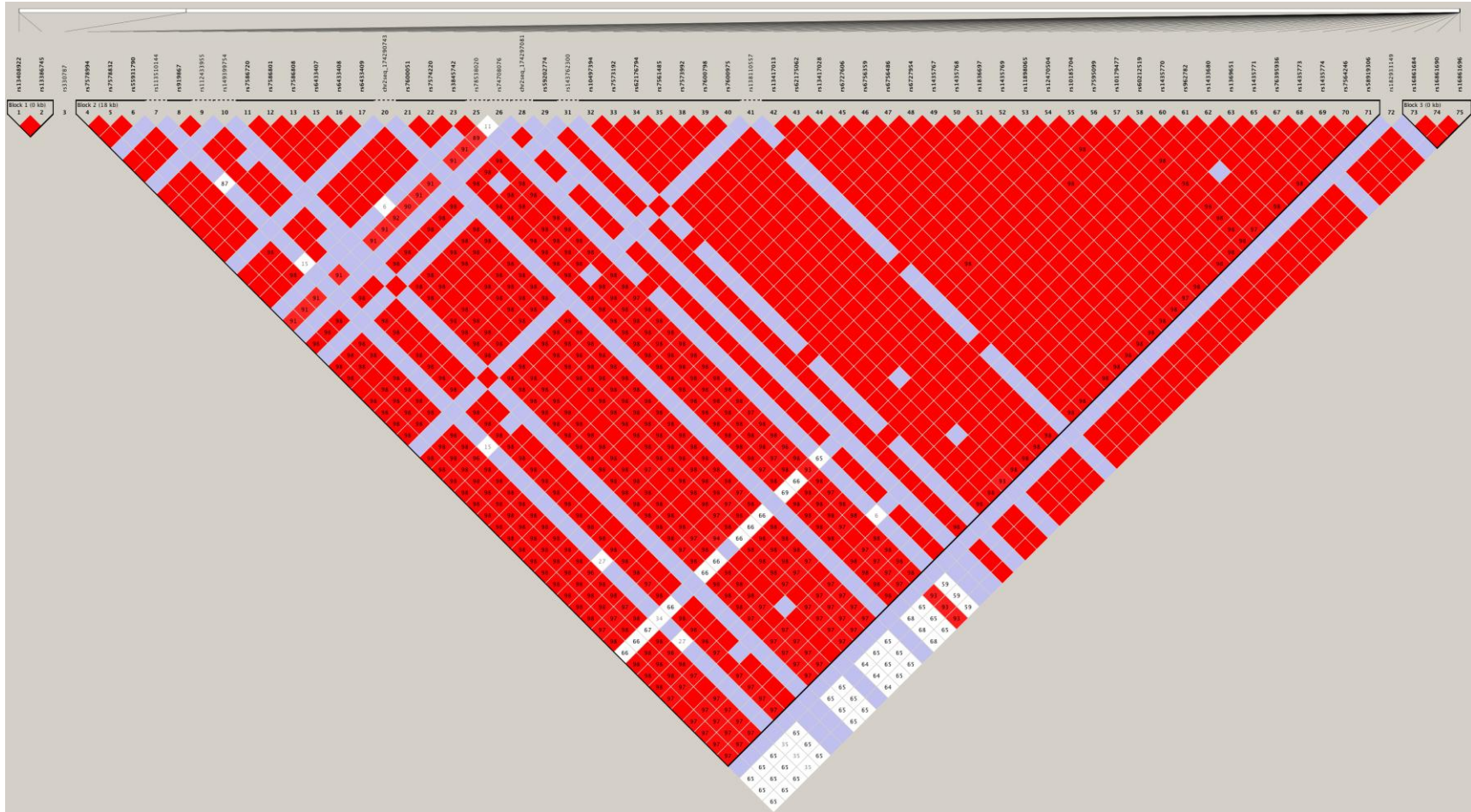
Allele frequency, OR, and Z listed for effect allele.

region, flanking the ENCODE enhancer region (characterized by H3K27Ac, H3K4me1, and H3K4me3 histone marks). Additionally, these SNPs flank 11 DNaseI hypersensitive sites (DHS) detected in 125 cell types from ENCODE, and 13 transcription factor binding sites from ENCODE (<http://genome.ucsc.edu/cgi-bin/hgTrackUi?hgsid=365409379&g=wgEncodeReg>). Haplotype analysis using Haploview<sup>220</sup> (Version 4.2) found very high linkage disequilibrium (LD) between SNPs genotyped in this region (**Figure 6.1A**). Specifically, the two significantly associated SNPs were in high LD ( $r^2=0.86$ ,  $D'=0.99$ ). Four associated SNPs were located in the chromosome 15 GWAS region, two in introns of *KIF7*, one in an intron of *TICRR*, and one intergenic SNP. The most significantly associated SNP, rs1110060, was also the most significant result in our GWAS of COME/ROM.<sup>247</sup> LD across the chromosome 15 region is shown in **Figure 6.1B**. SNP rs1110060 was in moderate LD with the two other common (MAF>0.10) associated chromosome 15 SNPs (rs12442646-  $r^2=0.413$ ,  $D'=0.678$ ; rs9672296-  $r^2=0.507$ ,  $D'=0.780$ ), but LD was complete between rs1110060 and the rare SNP rs62023109 ( $r^2=0.0063$ ,  $D'=1.000$ ).

#### 6.2.4.2 Evidence of chromosome 2 region enhancer activity in HMEEC-1 cells

Since chromosome 2-associated SNPs flanked an enhancer reported in other cell lines, interactions between the DNA in this region and enhancer marks H3K27Ac, H3K4me1, and H3K4me3 were measured to confirm enhancer activity in a relevant cell line, HMEEC-1. After processing the ChIP-seq data, the number of sequence reads per histone mark ranged from 10.3 to 18.3 million reads, exceeding our pre-specified threshold of 10 million reads. Enhancer activity was detected in the region between the two associated SNPs (**Figure 6.2**).

**Figure 6.1A:** Haplotype blocks of the chromosome 2 region. The Gabriel block method was used along with default Haploview parameters (minimum MAF=0.001 and  $D'$  used for measure of LD).



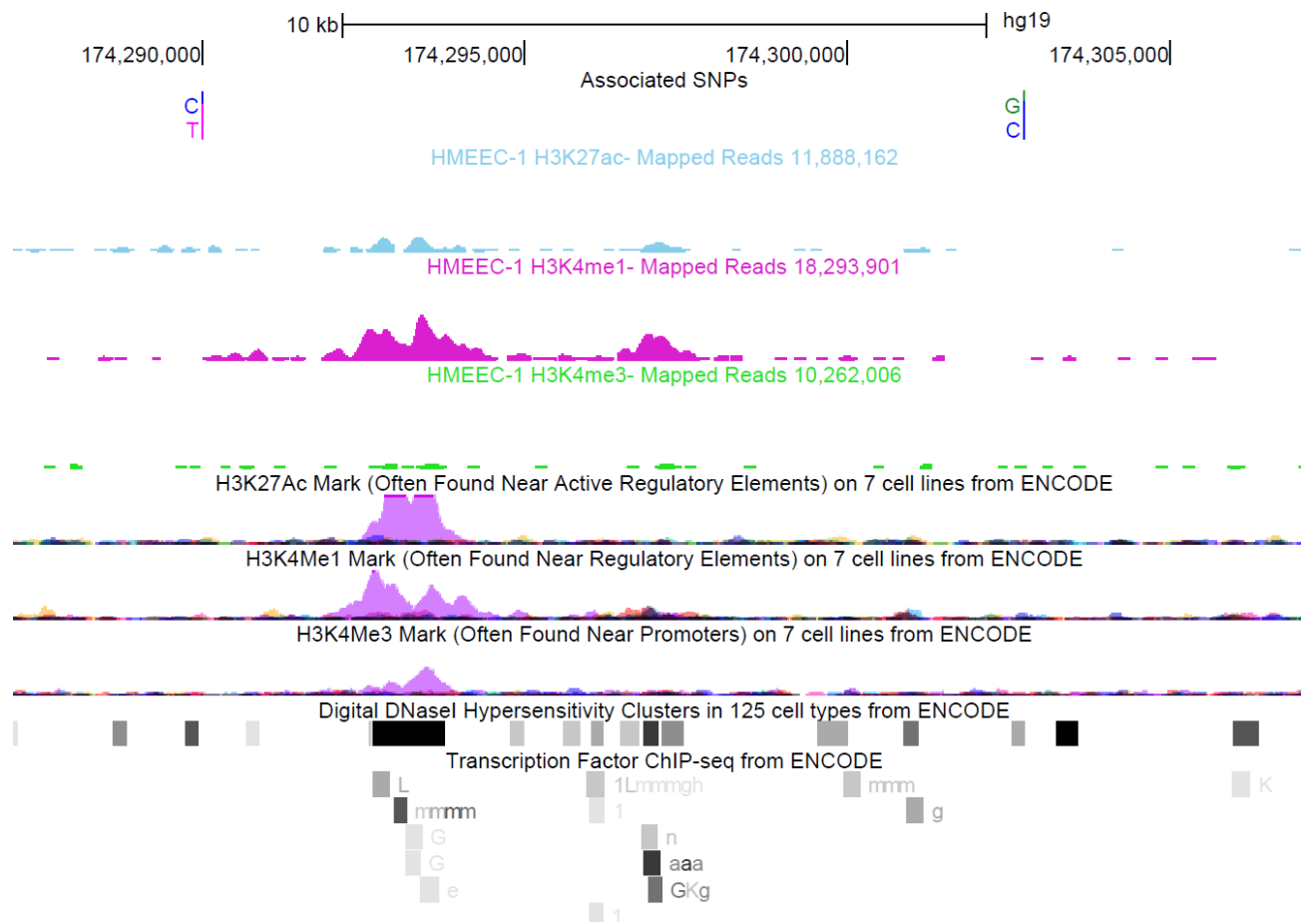


**Figure 6.1B:** Haplotype block of the chromosome 15 region from finemapping project. The Gabriel block method was used along with default Haploview parameters (minimum MAF=0.001 and  $D'$  used for measure of LD).





**Figure 6.2:** ChIP-seq of the chromosome 2 region (chr2:174,286,992-174,307,309) for enhancer activity (H3K27Ac, H3K4me1, and H3K4me3) in human middle ear epithelial cells (HMEEC-1) compared with ENCODE ChIP-seq data. Significantly associated SNPs from our fine-mapping project are also listed as a custom track.



In HMEEC-1 cells, this region contains a robust H3K4me1 peak (chr2:174,292,431-174,294,318), two nominal H3K27Ac peaks at the same location, and a second smaller H3K4me1 peak (seen only in HMEEC-1 cells) approximately 2.1kb downstream of the robust H3K4me1 peak. There was no detectable H3K4me3 peak in this genomic region. Monomethylation of lysine 4 on histone H3 in the absence of trimethylation indicates enhancer activity.<sup>269</sup>

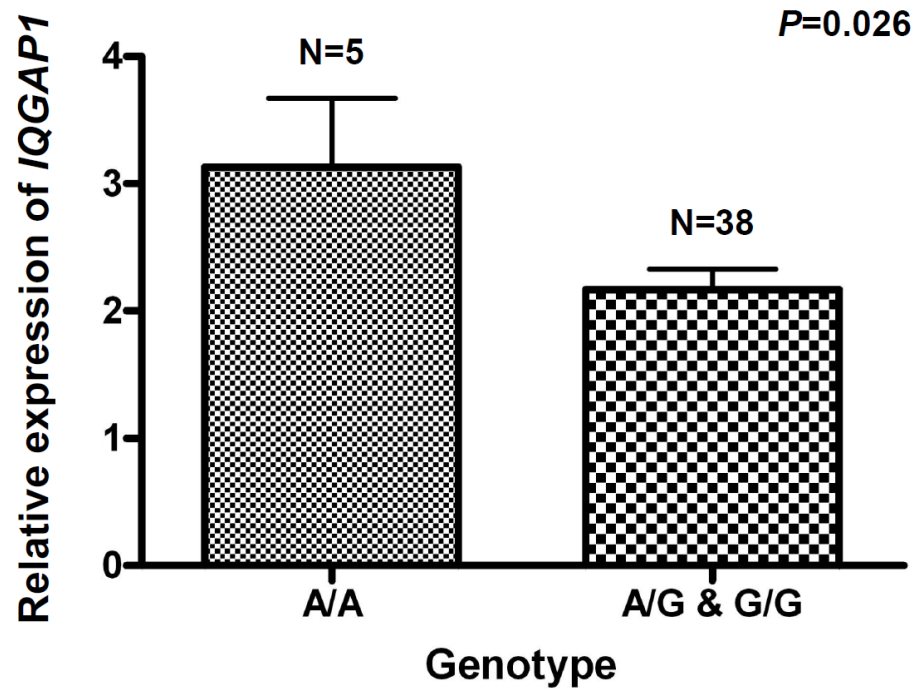
#### *6.2.4.3 Associated SNP rs1110060 impacts IQGAP1 expression in adenoids*

SNP rs1110060, the SNP most significantly associated with COME/ROM, was reported to be an eQTL in an RNA-seq study of 63 CEU lymphoblastoid cell lines<sup>213</sup>, acting to regulate expression of *IQGAP1* located 741.4 kb downstream . It is well known that adenoids play an important role in OM pathogenesis, with prior studies showing reduction of OM episodes after adenoidectomy.<sup>58,270</sup> In order to study if this regulatory impact could also be detected in a tissue relevant to OM pathology, we conducted a pilot eQTL study using available human adenoid samples (N=43). We detected reduced expression of *IQGAP1* ( $P=0.026$ ) in adenoids from individuals carrying the rs1110060 “G” risk allele (**Figure 6.3**).

#### *6.2.4.4 TGF signaling pathway upregulates KIF7 and other Hh signaling pathway genes*

SNP rs1110060 lies within the *KIF7* gene. Expression level changes of *PAI1* and *SMAD7* showed significant increases in expression levels after treatment with TFG $\beta$  (Figure S6.1), demonstrating the HMEEC-1 cell line is responsive to TGF $\beta$  signaling.

**Figure 6.3:** Relative *IQGAP1* expression levels in adenoids that carry the risk “G” allele of the SNP rs1110060 (N=38) and adenoids with the A/A genotype (N=5).



Treatment with TGF $\beta$  upregulated expression of *KIF7* (**Figure 6.4**), other HH signaling pathway genes *GLI1*, *SUFU*, and *PTCH1*, and *SMAD2* (Figure S6.2).

#### *6.2.4.5 LPS treatment downregulates expression of HH signaling pathway genes*

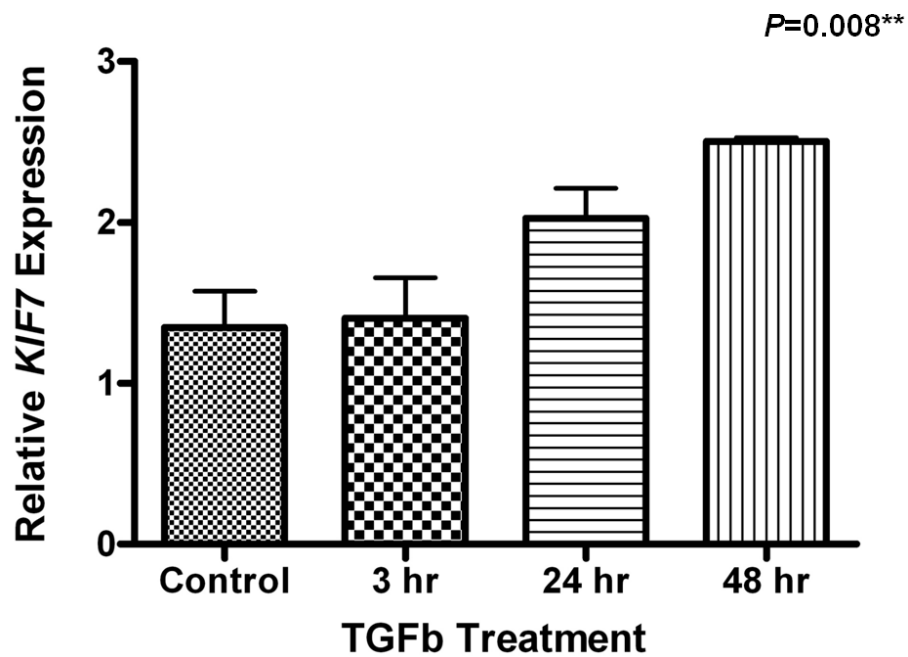
Based on prior studies showing an impact of bacterial LPS on gene expression of HH signaling pathway members, including *KIF7*, we evaluated regulation of HH signaling in response to LPS treatment. After treatment of HMEEC-1 cells with LPS for 4 hours, we detected significant reduction ( $P=0.0026$ ) in the expression of *KIF7* (**Figure 6.5**) and other HH signaling pathway genes *PTCH1*, *GLI1*, and *SUFU* (Figure S6.3) compared to controls.

### **6.2.5 Discussion**

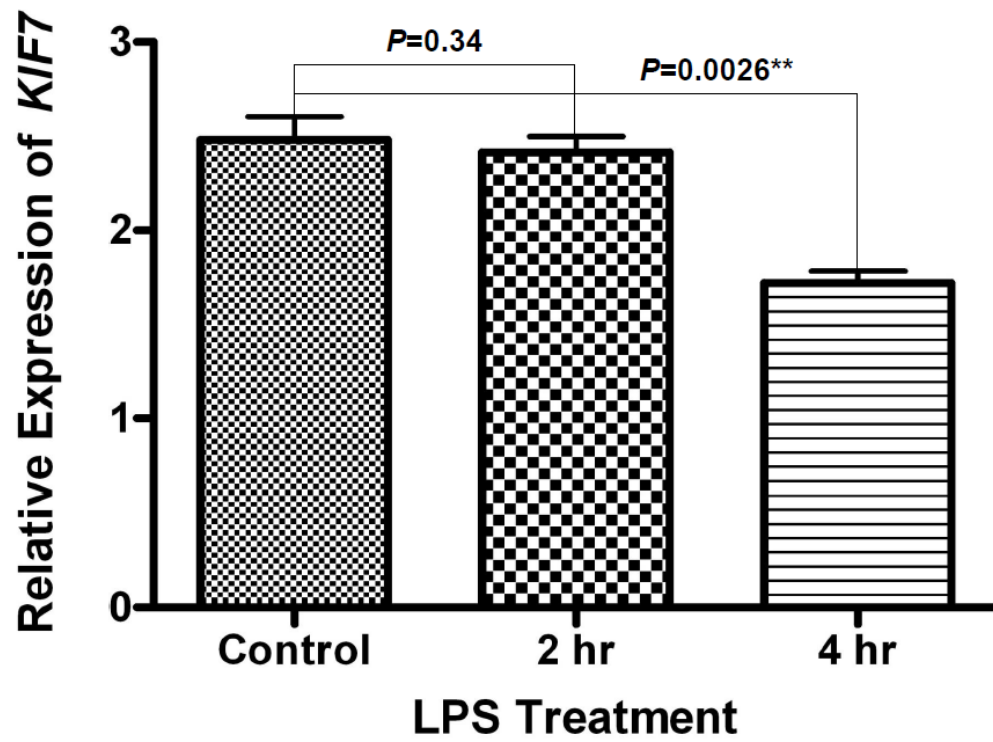
We successfully performed targeted resequencing of two GWAS-associated regions on chromosomes 2 and 15, permitting identification of known and novel genetic variants in these regions. Genotyping variants from these regions able to be typed on the Fluidigm platform effectively localized these association signals, providing solid evidence for association between COME/ROM and two SNPs in the chromosome 2 region, and four SNPs on chromosome 15.

The associated variants in the chromosome 2 region lie in a large intergenic region which, in our population, is characterized by high LD. This region encompasses numerous regulatory elements including enhancer peaks, transcription factor binding sites, and DNaseI hypersensitive regions.

**Figure 6.4:** Relative *KIF7* expression levels in HMEEC-1 cells after treatment with TGF $\beta$



**Figure 6.5:** Relative *KIF7* expression levels in HMEEC-1 cells after treatment with *P. aeruginosa* LPS



Enhancer activity characterized by H3K27Ac and H3K4me1 peaks was found in the ENCODE normal human epidermal keratinocytes (NHEK) cell line between our two associated SNPs (chr2:174,291,989-174,294,667), and overlapped DHS and multiple transcription factor binding sites (RXRA, STAT3, EBF, EBF1\_(C-8), ER $\alpha$ \_a). The other 6 ENCODE cell types lacked enhancer activity at this region. Enhancer activity (H3K4me1 in the absence of H3K4me3) at the same location as the NHEK enhancer peak (within our chromosome 2 region of interest) was detected in an OM-relevant cell line – HMEEC-1 – using ChIP-Seq. An additional H3K4me1 peak not found in ENCODE cell lines was also detected in HMEEC-1 cells in this region. Future genome-scale investigations will be required to determine which genes are impacted by enhancer activity in this region.

The most significantly associated SNP from our study, rs1110060, is a *KIF7* intronic SNP on chromosome 15, previously found to regulate expression of *IQGAP1* in a study of the transcriptome of lymphoblastoid cell lines from European ancestry.<sup>213</sup> *IQGAP1* is a scaffolding protein with two known activities conceivably related to OM, including pathogen adherence and infection and ROS-dependent VEGF signaling. *IQGAP1* is a key modulator of cytoskeletal function by regulating activity of the GTPases *rac1* and *cdc42*. Bacteria interact with *IQGAP1* at the epithelial barrier, causing changes in phosphorylation of *rac1* and *cdc42*, leading to altered cell shape and motility<sup>271</sup>, allowing for bacteria to internalize into the host.<sup>272</sup> *IQGAP* also functions as a VEGFR2-associated scaffolding protein in ROS-dependent VEGF signaling, known to contribute to the repair and maintenance of blood vessels.<sup>273</sup> Investigation using the *Junbo* mouse model of OM has implicated the HIF-VEGF pathways in OM, with upregulation of hypoxia and VEGF pathway members in the middle ear epithelium of mice with chronic

OM and not in control mice.<sup>13</sup> During OM, gene expression of FGF and VEGF family members was altered, prompting neovascularization to nourish a thickening of the ME mucosa in mice.<sup>274</sup> We observed reduced *IQGAP* expression in adenoids from surgical patients carrying the rs1110060 risk allele (G), although the number of A/A homozygotes in our pilot exploration of available adenoid samples was small (N=5), thus further examination in a larger population will be necessary to confirm this result. However, these results suggest further examination of *IQGAP* regulation in OM models may provide a productive avenue of enquiry in terms of explaining differential COME/ROM susceptibility in humans.

After fine mapping, the most significantly associated SNP from the GWAS, rs1110060, remained our top signal. This SNP lies in an intron of *KIF7*, a member of the HH signaling pathway which is important in developmental processes, including cleft palate and respiratory functions.<sup>275</sup> OM mouse models have provided evidence of defects in the regulation of TGF $\beta$  signaling in the middle ear<sup>71,73</sup>, and prior studies have identified a link between HH signaling and the TGF $\beta$  pathway in holoprocencephaly and cancer.<sup>263,264</sup> Because of the connections between TGF $\beta$  signaling pathway and OM<sup>71,73,95,96</sup> and TGF $\beta$  and HH signaling pathways<sup>263,264</sup>, we investigated whether TGF $\beta$  regulated *KIF7* expression. We also measured expression of other HH signaling pathway members, both upstream (*PTCH1* and *SUFU*) and downstream (*GLI1*) of *KIF7* activity. Expression of *SMAD2*, a signal transducer and transcriptional modulator in the TGF $\beta$  signaling pathway that has also been found associated with OM<sup>95</sup> was also measured. Treating HMEEC-1 cells with TGF $\beta$  resulted in an increase in expression of *KIF7* and other HH signaling pathway members, providing a link between *KIF7* and a key pathway in OM pathogenesis.



Concurrently, we investigated the effect of LPS treatment as a model of infection on the expression levels of *KIF7* in HMEEC-1 cells. *KIF7* binds to the bacterial protein ExoS, and an increase in ExoS expression or a decrease in *KIF7* expression causes a reduction in HH signaling pathway activity, leading to cytotoxicity.<sup>253</sup> Further, HH signaling pathway members *Shh*, *Ptc-1*, and *Gli1* are downregulated in cells treated with LPS causing cellular injury.<sup>252</sup> The authors speculate that because reduction of HH signaling activity was correlated with injury in cells, this study demonstrated the potential protective role of HH signaling during infection. Consistent with studies in other cell types, HH signaling pathway members *KIF7*, *GLI1*, *PTCH1*, and *SUFU* were downregulated following a 4 hour LPS challenge in HMEEC-1 cells. These findings suggest that Gram negative bacterial infection may reduce expression of HH signaling in ME epithelia. Since this pathway has not previously been implicated in OM, these results suggest an entirely new potential mechanism may be relevant to OM pathogenesis and susceptibility.

Our fine-mapping study is based on findings from one of only two GWA studies of OM. We investigated two regions, the region on chromosome 15 with most significantly associated SNP rs1110060 and the region on chromosome 2 encompassing the significantly replicated SNP rs10497394. We comprehensively sampled variants (MAF >0.01) across these two regions to determine likely causal variants. The primary strength of this study is the family-based study design utilized in our fine-mapping of these two regions. Based on the study design, we utilized a more powerful analytical tool to test association that is robust to population stratification. Additionally, *in vitro* assays were carried out in a human middle ear epithelial cell line and adenoid tissue, both relevant to OM. Limitations to our study include a modest sample size for both fine-

mapping and adenoid studies. We were also unable to design genotyping assays for indels detected in these regions, and not all SNPs detected in our sequence data were designable on the Fluidigm platform.

This study represents the first functional follow up to GWAS findings in COME/ROM. Our results indicate potential novel mechanisms may be involved in OM pathogenesis. We uncovered plausible candidate genes including *KIF7* and *IQGAP1*, and a novel role for the HH pathway. Additionally, the enhancer activity found in the associated chromosome 2 region appears likely to regulate expression of genes involved in OM pathogenesis. Further study is necessary to understand the details of the mechanisms of each of these findings in OM pathogenesis. This investigation has demonstrated that fine mapping followed by a functional assessment of GWAS findings illuminates novel mechanisms of OM pathology and susceptibility.

## Chapter 7

### Discussion

Otitis media is the most common pediatric disease in the United States and can lead to long-term hearing and developmental issues. Viral infection of the URT induces inflammation of the nasopharynx, an environment that is highly connected to the ME. The adenoid, known to be a pathogenic bacterial reservoir in the nasopharynx, is positioned between the openings of the two ETs. The bacteria on the surface of the adenoid ascend the ETs into the ME during URTI, causing OM. Other known contributors to OM risk and progression include ET obstruction and reduced mucociliary clearance. We have sought to discover novel risk factors that increase OM susceptibility using complementary genomic approaches to investigate two key aspects of the development of OM: bacterial changes in response to URTI and host genetics.

First, knowing that URTIs predispose the host to bacterial invasion of the ME via the ET<sup>12</sup>, we examined changes in the microbiome of the nasopharynx following controlled rhinovirus infection in healthy adults. We used two different methods to study these changes, including a clinical diagnostics microarray (Chapter 2) and high-throughput sequencing (Chapter 3). In Chapter 2, we successfully carried out a rhinovirus challenge study using a natural route of self-inoculation in healthy adults. NLF samples were used to survey the nasopharyngeal microbiome comprehensively. Though we did not see significant differences in bacteria between time points of inoculation, we found evidence to support prior studies showing that pathogenic bacterial species reside in otherwise healthy adults. In Chapter 3, we sought to determine the relative abundance of all bacteria present in the nasopharynx. Using high-throughput sequencing, we were able

to detect significant differences in two genera – *Neisseria* and *Propionibacterium* – between infected and not infected individuals. Differences between infected and not infected groups, but not between the different time-points (before, during, and after inoculation), may indicate the importance of host response during infection. We speculate that a higher predominance of *Neisseria* and/or greater bacterial diversity at baseline may have conferred a protective advantage in individuals that did not become infected, despite inoculation with an active virus. We also discovered that inter-individual differences were higher than intra-individual differences using Unifrac distances. Furthermore, “Other” bacteria were one of the dominant phylum-level classifications detected in these samples, suggesting that bacteria not yet sequenced may play an important role in URTIs and need further study. Our results have highlighted that individual microbiomes and responses to perturbations (*i.e.* viral infection) are variable and complex. This validates the potential advantages of a more personalized approach to treatment, instead of current standard therapies of broad-spectrum antibiotics for any URTI, including OM.

Results of our nasopharyngeal microbiome study (see Chapter 3) provide insights into suitable future study questions and methods. Not only would it be informative to collect deeper sequencing data on more subjects, but sequencing of other genes (*parE* and *gyrB*) would provide species-level classification of bacteria present in the sample.<sup>276-278</sup> Species-level information is imperative for these studies, as pathogenic and non-pathogenic species belong to the same genus. In our microbiome study, we found statistically different relative abundances of two genera – *Neisseria* and *Propionibacterium* – between infected and not-infected individuals. Interestingly, *Neisseria* and *Haemophilus* relative abundance was higher, and *Propionibacterium*

lower in not-infected than infected individuals. We do not know if there are species within these genera that contribute a protective effect on the host.

The bacterial transcriptome has proven to be a valuable tool in understanding the relationship between a microbial community and pathology of disease.<sup>279</sup> Shotgun metatranscriptomics could be used as a complementary approach to high-throughput sequencing of bacterial genes to characterize the changes in gene expression that viral infection can elicit in a microbial community.<sup>176</sup> This information would help us to address the question of the functional role that the microbiome plays during URTI. It is uncertain, for instance, whether the microbiome can help to protect the host from viral infection by providing a barrier, or by initiating an immune response. Alternatively, it is possible that the microbiome can aid a virus in invading the host mucosa. Knowing the genetic and metabolic profiles of bacteria present in the nasopharyngeal microbiome will elucidate novel mechanisms that could be targeted for therapeutic strategies.

Employing a GWAS<sup>247</sup>, we sought to discover novel OM risk variants in a family-based population of COME/ROM. This agnostic approach led to the discovery of two novel regions of the genome associated with COME/ROM (Chapter 4). The most significantly associated SNP was rs1110060 that lies in an intron of kinesin family member 7 (*KIF7*), a gene not previously known to be important in OM. We sought replication of the most significantly associated SNPs in an independent OM population, and validated association with an intergenic SNP on chromosome 2. We employed eQTL databases and splice prediction software to predict functional activity of our most significant SNPs.

These results provide novel insights into putative contributors to OM pathogenesis that can be investigated in other populations.

Chapter 5 describes investigation of replication of the most significantly associated variants from an independent GWAS of acute and chronic OM from an Australian group.<sup>115</sup> Replication of results in independent populations is important to validate genetic findings. This study represents one of only four attempts to replicate genome-wide results in OM.<sup>115,247,248</sup> Unfortunately, no associations were replicated in our family-based population of COME/ROM. Lack of replication in this study brings light to issues in genetic studies of OM including lack of resources (*i.e.*, few population studies, small sample sizes of current OM studies, and sparse GWAS data), as well as inconsistent recruitment strategies and affected status definitions.

The discovery of causal variants of COME/ROM will increase our knowledge of genes and pathways involved in COME/ROM pathogenesis, furthering our understanding of this very common, multifactorial disease. Understanding mechanisms that make some children more susceptible to chronic infections is anticipated to impact clinical treatment of COME/ROM, ideally precipitating a shift from prescribing antibiotics and surgery to treat illness to the prevention of chronic infection. This would not only decrease the disease incidence, but also the economic burden from COME/ROM in the U.S.

Although candidate gene and animal model approaches have highlighted important pathways in the pathogenesis of OM, additional genetic contributors remain to be

identified to explain the heritability of this disease. Genome-scale agnostic approaches have the advantage that they do not presume prior knowledge of disease pathogenesis, and thus hold the promise of novel insights into mechanism. Unfortunately, in OM research, there have been insufficient genomic scans and even fewer follow-up investigations of these results. Promisingly, the association at rs10497394 discovered in our GWAS was replicated in the UPitt family population<sup>247</sup>, and the linkage peak at 10q22.3 discovered in the UPitt family population was replicated in WAFSOM<sup>248</sup>. Results of genome-scale studies have revealed evidence for novel mechanisms of OM pathogenesis not discovered using candidate gene approaches that could provide new screening and treatment strategies. The two GWA studies discovered novel putative genes contributing to OM (*CAPN14* and *GALNT14*<sup>115</sup>) and COME/ROM (*KIF7* and *TICRR*<sup>247</sup>).

Most linkage studies and GWAS have yet to confirm the identity of the underlying causal variant. SNPs genotyped in GWAS are selected on the basis of genomic coverage and patterns of LD. The SNP associated with the trait of interest could be in high or even complete LD with other SNPs inherited from a common ancestor, and therefore cannot be distinguished from each other by association methods. One way to hone in on the causal variants is to examine more diverse populations which have different patterns of LD and can narrow down regions to be functionally examined. The 1000 Genomes Project<sup>280</sup> has generated whole-genome sequence data from multiple populations, including those from European, East Asian, West African, and South Asian ancestry, and populations from the Americas, as a publicly available resource to the genomics community. The HapMap project<sup>281</sup> characterizes the patterns of LD in a subset of these

populations. These resources, along with recruitment of diverse populations with OM, may help us narrow down linkage and GWAS signals.

Lack of replication between genetic studies of OM has been attributed to phenotypic heterogeneity due to inconsistent application of clinical criteria and lack of power due to sample size. There is a critical need for more studies of larger sample size and more diverse populations to discover novel mechanisms of OM. Additional independent studies will enable the investigation of replication suspected loci and genes from genomic studies. In order to avoid lack of replication due to heterogeneous definitions of OM and its various severities, there is a need for consistency in phenotype definitions and recruitment criteria spanning both the clinic and research studies. The International Consortium of the Genetics of OM (OTIGEN; <http://www.otigen.org/>) was initiated by OM investigators with the common intention to resolve the genetic contributions to OM. OTIGEN investigators strive to bring consistency to OM phenotypic definitions and strategize larger studies and collaborations to replicate findings of individual studies.

In summary, genome-scale studies provide new insights into OM pathobiology which may provide the necessary understanding for the development of novel screening or treatment tools. Candidate gene studies are able to verify biological knowledge, but agnostic approaches aren't limited and are able to discover new targets. Efforts need to be placed on consistency in diagnosis, recruitment, and phenotyping to allow for more efficient population studies.



Following our GWAS, functional validation of associated regions using targeted re-sequencing, fine-mapping, and functional assays was carried out (Chapter 6). This approach honed our GWAS regions to six associated SNPs across the two highest priority regions. Our most significantly associated SNP from our GWAS, rs1110060, remained the top signal in our fine-mapping study. ChIP-seq of HMEEC-1 found enhancer activity in a large intergenic region on chromosome 2, confirming a regulatory role in OM, putatively affecting gene expression of an unknown gene or set of genes. The chromosome 15 region contained four associated variants across a large region containing two genes, *KIF7* and *TICRR*. Based on association significance and biological plausibility, we focused our functional assays on impacts of rs1110060 and *KIF7*. We measured the effect of rs1110060 on expression of *IQGAP1*, and we observed effects of bacterial infection (LPS challenge) and TGF $\beta$  signaling on levels of *KIF7* and its HH signaling pathway members. We have determined that rs1110060 appears to affect expression of *IQGAP1* in adenoids, though a larger population is needed to validate this observation. TGF $\beta$  signaling was found to upregulate expression of *KIF7* which is important because the TGF $\beta$  pathway has been associated with OM in both mouse<sup>71,73</sup> and human studies<sup>95,96,98</sup>. Finally, expression of *KIF7* and other HH signaling pathway members decreased after treatment with bacterial LPS. This has also been shown in studies of acute lung injury, demonstrating the role of HH signaling in preventing cell injury.<sup>252</sup> Further functional characterization is necessary to understand the details of the mechanisms by which these regions affect OM pathogenesis.

In the future, we can ascertain HMEEC-1 enhancers across the genome using our existing ChIP-seq data. We also have generated ChIP-seq data for the H3K27Ac enhancer mark in HMEEC-1 cells treated with bacterial LPS for two hours, however our

subsequent LPS treatments (described in Chapter 6) showed that four hours of LPS treatment was necessary to induce a robust *NFκβ* and *KIF7* response to LPS challenge. Although some genes may respond at the earlier time point, we feel it will be informative to treat cells with LPS for four hours to compare enhancer activity. Changes in enhancer activity may indicate those genes which are important for host response to bacterial infection. Further studies would be to ChIP for repressive marks in untreated and LPS-treated cells. Characterization of epigenetic marks in the HMEEC-1 cell line will facilitate further investigation of regulatory elements which contribute to the pathobiology of OM. Blocking specific chromosome 2 enhancer regions using gene editing tools like CRISPR<sup>282</sup> and measuring differential expression with RNA-seq would identify genes regulated by this enhancer. Comparison of RNA-seq data from normal and LPS-treated cells could also highlight pathways active in the ME epithelium during infection.

We generated evidence of potential links between *KIF7* regulation in response to infection or TGFβ. We plan to knock down *KIF7* in HMEEC-1 cells using a reported miRNA methodology.<sup>253</sup> After successful knockdown of *KIF7*, we can carry out multiple molecular assays to explore its role. Is host response to infection enhanced or impaired by the absence of *KIF7*? Will reduction of HH signaling by knockdown of *KIF7* induce cell injury, as seen in a prior study<sup>252</sup>? HH signaling is activated by HH ligands, including Sonic (SHH), Desert (DHH), and Indian (IHH) Hedgehog, binding to transmembrane protein Patched (PTCH1). When bound to the ligand, PTCH1 no longer represses Smoothened (SMO), which results in *KIF7* trafficking Gli transcription factors to the tip of the primary cilium where they become activated. Activated Gli transcription factors then translocate to the nucleus and regulate transcription of target genes. It is

known that blocking the SMO/PTCH1 axis does not inhibit Gli activation because SMAD3-dependent TGF $\beta$  signaling is able to activate the Gli transcription factors.<sup>264</sup> In HMEEC-1 cells, can TGF $\beta$  signaling compensate for absence of *KIF7*, or is *KIF7* essential for the activation of Gli transcription factors? Answers to these questions will illustrate how *KIF7* may impact the pathobiology of OM.

We propose to employ assays to extend our studies of the function of the most significantly associated SNP, rs1110060. Spliceman<sup>228</sup> predicted that rs1110060 is likely to affect splicing. We predict that this SNP might cause inclusion of intronic sequence in the transcribed mRNA and create a downstream premature stop codon, truncating the full-length protein from 1343 amino acids (AA) to 601 AA (calculated using Bioedit Sequence Alignment Editor, Version 7.2.3<sup>262</sup>). This truncation would remove a KIF7-KIF27 homologous (KIF727H) domain and coiled coil domains.<sup>283,284</sup> Other *KIF7* truncating mutations are responsible for primary ciliopathies fetal hydroletharus syndrome<sup>285</sup>, acrocallosal syndrome<sup>285,286</sup>, and Joubert syndrome<sup>287</sup>, suggesting a critical role for KIF7 in ciliogenesis and development. A highly penetrant truncation of KIF7 might therefore seem unlikely to cause a non-syndromic common disease like OM. We speculate that rs1110060 could be a “wobbly” splice site, creating a pool of truncated protein products with impacts on responses to infection or TGF $\beta$ , or on ciliogenesis in a subset of ME epithelial cells. Although primary cilia have signaling and sensory functions, while motile cilia move in coordinated sweeps and are responsible for mucociliary clearance, temporal studies of the respiratory epithelium suggest the primary cilium may play a role in ciliogenesis of motile cilia.<sup>288</sup>

Additionally, rs1110060 “G” risk allele creates a CpG site, permitting the adjacent cytosine to be methylated. DNA methylation is a mechanism that controls gene expression, in general resulting in repression. As we saw a reduction in *IQGAP1* expression in adenoid samples that carried the risk allele, it would be logical to measure DNA methylation using pyrosequencing at this region in affected and unaffected samples. Differential methylation may explain how our causal variant affects OM pathogenesis, whether by reduction of *IQGAP1* or *KIF7* expression, or by regulating expression of another gene or pathway.

In Chapter 6, chromosome 15 SNP rs1110060 was found to act as an eQTL, decreasing expression of *IQGAP1* in adenoid samples that carried the “G” risk allele. Because the reference AA homozygous group was modest, it is important to validate this result in a larger sample. Additionally, the significantly replicated chromosome 2 SNP from our GWAS, rs10497394, was previously found by others to be an eQTL regulating expression of *LDLR* in a population of European ancestry<sup>213</sup> (Chapter 4). Our fine-mapping project found more strongly associated SNPs in the chromosome 2 region, yet one of these – rs6433410 – has also been reported to regulate *LDLR* in the same population<sup>213</sup>. We could initially test whether this SNP regulates expression of *LDLR* in our adenoid specimens. These results might provide insight into another novel pathway associated with COME/ROM risk.

We will seek to replicate variants associated with COME/ROM from our family-based fine-mapping study (Chapter 6) in independent populations of COME/ROM. WAFSOM and a U.K. family-based COME/ROM population, currently being recruited by OTIGEN

member Dr. Mahmood Bhutta, would be the most appropriate samples, however do not have existing genotype data for these loci. Dr. Bhutta's is recruiting trios, using high diagnostic stringency for COME/ROM. Additionally, our collaborator Dr. Steve Brown, has added *KIF7* to the International Mouse Phenotyping Consortium (IMPC) (<http://www.mousephenotype.org/>) queue for production of the knockout and phenotyping at the Mammalian Research Center (MRC) at Harwell. Researchers at MRC Harwell are experts at phenotyping mice for hearing-related diseases, including OM, illustrated by their experience discovering and phenotyping OM mouse models *Junbo*, *Jeff*, and the *Tgif1* knockout. With the *KIF7* knockout mouse at MRC Harwell, we would be able to investigate the ME for any defects in hearing, anatomy, mucociliary clearance, and ET function. ME epithelium and other tissues from these mice could be used to measure longitudinal gene expression. There is currently a *KIF7* null mouse, generated at Genentech, that displays polydactyly, microphthalmia, and exencephaly by 18.5 days of gestation.<sup>289</sup> Using this *Kif7* knockout mouse, it was shown that *Kif7* plays a major role in transducing the Hh signal in the primary cilium.

Although not discussed in this thesis, we generated high coverage sequence data of the chromosome 19 COME/ROM linkage region<sup>24,106</sup> concurrently with the sequencing described in Chapter 6. Applying the same variant-calling pipeline (Chapter 6), we identified 10,602 SNPs across this region. A subset of these variants, based on gene enrichment, biological plausibility, and predicted functional impacts, needs to be selected for genotyping in our full family-based population of COME/ROM to discover variants and/or genes responsible for linkage.

To move the field of OM genetic research forward and discover additional genetic variants that increase risk of OM susceptibility, we need diverse population studies with consistently applied phenotypic criteria and large sample sizes. Both our research (Chapters 4-6) and the Australian GWAS<sup>115</sup> suggest that existing family-based and population studies lack power to compensate for differences in study designs. Family-based approaches have the added advantages of being robust to population stratification and avoiding issues of control misclassification. Recruiting probands with Cholesteatoma –a non-neoplastic yet highly damaging cystic lesion that contains layers of keratinized epithelium induced by chronic inflammation and is a severe middle ear pathology<sup>290</sup>– would provide an objective clinical measure consistently applied across the globe. Additionally, children with Cholesteatoma have CT scans prior to<sup>291</sup> and after<sup>292</sup> surgery<sup>290</sup>, providing image data that could be used to generate novel measures that have not been previously explored in genetic studies of OM. This thesis has highlighted the importance of bacterial communities and host response to infection; therefore, collection of adenoids swabs during surgery or NLF samples for microbiome and virome in the context of host susceptibility would provide an additional unique aspect. In my opinion, this study design provides a definitive affected status definition, novel imaging and microbial data, and host molecular resources for genomic studies of OM susceptibility.

## References

1. Gates, G.A. Cost-effectiveness considerations in otitis media treatment. *Otolaryngology - Head and Neck Surgery* **114**, 525-530 (1996).
2. Winther, B., Hayden, F.G. & Hendley, J.O. Picornavirus infections in children diagnosed by RT-PCR during longitudinal surveillance with weekly sampling: Association with symptomatic illness and effect of season. *Journal of medical virology* **78**, 644-650 (2006).
3. Pappas, D.E., Hendley, J.O., Hayden, F.G. & Winther, B. Symptom Profile of Common Colds in School-Aged Children. *The Pediatric infectious disease journal* **27**, 8-11 10.1097/INF.ob013e31814847d9 (2008).
4. Grijalva, C.G., Nuorti, J.P. & Griffin, M.R. Antibiotic prescription rates for acute respiratory tract infections in US ambulatory settings. *JAMA* **302**, 758-66 (2009).
5. Vergison, A. *et al.* Otitis media and its consequences: beyond the earache. *The Lancet infectious diseases* **10**, 195-203 (2010).
6. Chi, D.H. *et al.* Nasopharyngeal Reservoir of Bacterial Otitis Media and Sinusitis Pathogens in Adults during Wellness and Viral Respiratory Illness. *American Journal of Rhinology* **17**, 209-214 (2003).
7. Hendley, J.O., Hayden, F.G. & Winther, B. Weekly point prevalence of *Streptococcus pneumoniae*, *Hemophilus influenzae* and *Moraxella catarrhalis* in the upper airways of normal young children: effect of respiratory illness and season. *APMIS* **113**, 213-220 (2005).
8. Winther, B., Gross, B.C., Hendley, J.O. & Early, S.V. Location of bacterial biofilm in the mucus overlying the adenoid by light microscopy. *Archives of Otolaryngology—Head & Neck Surgery* **135**, 1239-1245 (2009).

9. Han, J.K., Hendley, J.O. & Winther, B. Bacterial pathogens of acute sinusitis in the osteomeatal complex during common colds and wellness. *International Forum of Allergy & Rhinology* **1**, 356-360 (2011).
10. Rawlings, B.A., Higgins, T.S. & Han, J.K. Bacterial pathogens in the nasopharynx, nasal cavity, and osteomeatal complex during wellness and viral infection. *Am J Rhinol Allergy* **27**, 39-42 (2013).
11. Puhakka, T. *et al.* The common cold: Effects of intranasal fluticasone propionate treatment. *Journal of Allergy and Clinical Immunology* **101**, 726-731 (1998).
12. Kubba, H., Pearson, J.P. & Birchall, J.P. The aetiology of otitis media with effusion: a review. *Clinical otolaryngology and allied sciences* **25**, 181-194 (2000).
13. Cheeseman, M.T. *et al.* HIF–VEGF Pathways Are Critical for Chronic Otitis Media in *Junbo* and *Jeff* Mouse Mutants. *PLoS Genet* **7**, e1002336 (2011).
14. Daly, K.A., Hunter, L.L. & Giebink, G.S. Chronic otitis media with effusion. *Pediatr Rev* **20**, 85-93; quiz 94 (1999).
15. Pichichero, M.E. Otitis media. *Pediatr Clin North Am* **60**, 391-407 (2013).
16. Nistico, L. *et al.* An Adenoid Reservoir for Pathogenic Biofilm Bacteria. *Journal of clinical microbiology* (2011).
17. Bluestone, C.D. Pathogenesis of otitis media: role of eustachian tube. *Pediatr Infect Dis J* **15**, 281-91 (1996).
18. Brook, I., Shah, K. & Jackson, W. Microbiology of healthy and diseased adenoids. *Laryngoscope* **110**, 994-9 (2000).
19. Brook, I. & Shah, K. Bacteriology of adenoids and tonsils in children with recurrent adenotonsillitis. *Ann Otol Rhinol Laryngol* **110**, 844-8 (2001).
20. Faden, H. *et al.* Nasopharyngeal flora in the first three years of life in normal and otitis-prone children. *Ann Otol Rhinol Laryngol* **100**, 612-5 (1991).



21. Gates, G.A. Otitis media--the pharyngeal connection. *JAMA* **282**, 987-9 (1999).
22. Ren, T. *et al.* 16S rRNA survey revealed complex bacterial communities and evidence of bacterial interference on human adenoids. *Environ Microbiol* **15**, 535-47 (2013).
23. Gould, J.M. & Matz, P.S. Otitis media. *Pediatrics in review / American Academy of Pediatrics* **31**, 102-116 (2010).
24. Daly, K.A. *et al.* Chronic and recurrent otitis media: a genome scan for susceptibility loci. *American Journal of Human Genetics* **75**, 988-997 (2004).
25. Lieberthal, A.S. *et al.* The diagnosis and management of acute otitis media. *Pediatrics* **131**, e964-99 (2013).
26. Grevers, G. & First International Roundtable, E.N.T.M.G. Challenges in reducing the burden of otitis media disease: an ENT perspective on improving management and prospects for prevention. *International journal of pediatric otorhinolaryngology* **74**, 572-577 (2010).
27. Meropol, S.B., Glick, H.A. & Asch, D.A. Age inconsistency in the American Academy of Pediatrics guidelines for acute otitis media. *Pediatrics* **121**, 657-68 (2008).
28. Meropol, S.B. Valuing reduced antibiotic use for pediatric acute otitis media. *Pediatrics* **121**, 669-73 (2008).
29. Bakaletz, L.O. Bacterial biofilms in otitis media: evidence and relevance. *Pediatr Infect Dis J* **26**, S17-9 (2007).
30. Post, J.C., Hiller, N.L., Nistico, L., Stoodley, P. & Ehrlich, G.D. The role of biofilms in otolaryngologic infections: update 2007. *Curr Opin Otolaryngol Head Neck Surg* **15**, 347-51 (2007).

31. Sanford, B.A., Shelokov, A. & Ramsay, M.A. Bacterial adherence to virus-infected cells: a cell culture model of bacterial superinfection. *J Infect Dis* **137**, 176-81 (1978).
32. Hakansson, A., Kidd, A., Wadell, G., Sabharwal, H. & Svanborg, C. Adenovirus infection enhances in vitro adherence of *Streptococcus pneumoniae*. *Infect Immun* **62**, 2707-14 (1994).
33. Faden, H. *et al.* Relationship between nasopharyngeal colonization and the development of otitis media in children. Tonawanda/Williamsville Pediatrics. *J Infect Dis* **175**, 1440-5 (1997).
34. Takahashi, H., Fujita, A. & Honjo, I. The influence of upper respiratory tract inflammation upon tubal function. *Clin Otolaryngol Allied Sci* **20**, 216-8 (1995).
35. Chonmaitree, T., Owen, M.J. & Howie, V.M. Respiratory viruses interfere with bacteriologic response to antibiotic in children with acute otitis media. *J Infect Dis* **162**, 546-9 (1990).
36. Carson, J.L., Collier, A.M. & Hu, S.S. Acquired ciliary defects in nasal epithelium of children with acute viral upper respiratory infections. *N Engl J Med* **312**, 463-8 (1985).
37. Bakaletz, L.O. Immunopathogenesis of polymicrobial otitis media. *J Leukoc Biol* **87**, 213-22 (2010).
38. Massa, H.M., Cripps, A.W. & Lehmann, D. Otitis media: viruses, bacteria, biofilms and vaccines. *The Medical journal of Australia* **191**, S44-9 (2009).
39. Chonmaitree, T. *et al.* Viral upper respiratory tract infection and otitis media complication in young children. *Clin Infect Dis* **46**, 815-23 (2008).
40. Hoffman, H.J. *et al.* Panel 1: Epidemiology, natural history, and risk factors. *Otolaryngol Head Neck Surg* **148**, E1-E25 (2013).

41. Daly, K.A. *et al.* Epidemiology, natural history, and risk factors: panel report from the Ninth International Research Conference on Otitis Media. *International journal of pediatric otorhinolaryngology* **74**, 231-240 (2010).
42. Kong, K. & Coates, H.L. Natural history, definitions, risk factors and burden of otitis media. *The Medical journal of Australia* **191**, S39-43 (2009).
43. Rovers, M.M. The burden of otitis media. *Vaccine* **26 Suppl 7**, G2-4 (2008).
44. Sharma, S.K., Casey, J.R. & Pichichero, M.E. Reduced serum IgG responses to pneumococcal antigens in otitis-prone children may be due to poor memory B-cell generation. *J Infect Dis* **205**, 1225-9 (2012).
45. Avanzini, A.M. *et al.* Children with recurrent otitis show defective IFN gamma-producing cells in adenoids. *Pediatric allergy and immunology : official publication of the European Society of Pediatric Allergy and Immunology* **19**, 523-526 (2008).
46. Emonts, M. *et al.* Genetic polymorphisms in immunoresponse genes TNFA, IL6, IL10, and TLR4 are associated with recurrent acute otitis media. *Pediatrics* **120**, 814-823 (2007).
47. Sharma, S.K., Casey, J.R. & Pichichero, M.E. Reduced memory CD4+ T-cell generation in the circulation of young children may contribute to the otitis-prone condition. *J Infect Dis* **204**, 645-53 (2011).
48. Kvestad, E. *et al.* Otitis media: genetic factors and sex differences. *Twin Res* **7**, 239-44 (2004).
49. Auinger, P., Lanphear, B.P., Kalkwarf, H.J. & Mansour, M.E. Trends in otitis media among children in the United States. *Pediatrics* **112**, 514-520 (2003).
50. Lanphear, B.P., Byrd, R.S., Auinger, P. & Hall, C.B. Increasing prevalence of recurrent otitis media among children in the United States. *Pediatrics* **99**, E1 (1997).

51. Teele, D.W., Klein, J.O. & Rosner, B. Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study. *The Journal of infectious diseases* **160**, 83-94 (1989).
52. Kvaerner, K.J., Tambs, K., Harris, J.R. & Magnus, P. Distribution and heritability of recurrent ear infections. *The Annals of Otology, Rhinology, and Laryngology* **106**, 624-632 (1997).
53. Karevold, G., Haapkyla, J., Pitkaranta, A., Nafstad, P. & Kvaerner, K.J. Paediatric otitis media surgery in Norway. *Acta Otolaryngol* **127**, 29-33 (2007).
54. Karevold, G., Haapkyla, J., Pitkaranta, A. & Kvaerner, K.J. Otitis media surgery: large variability between Finland and Norway. *Int J Pediatr Otorhinolaryngol* **71**, 1035-9 (2007).
55. Pettigrew, M.M. *et al.* Upper respiratory tract microbial communities, acute otitis media pathogens, and antibiotic use in healthy and sick children. *Appl Environ Microbiol* **78**, 6262-70 (2012).
56. Rosenfeld, R.M., Jang, D.W. & Tarashansky, K. Tympanostomy tube outcomes in children at-risk and not at-risk for developmental delays. *International journal of pediatric otorhinolaryngology* **75**, 190-195 (2011).
57. Coyte, P.C., Croxford, R., McIsaac, W., Feldman, W. & Friedberg, J. The role of adjuvant adenoidectomy and tonsillectomy in the outcome of the insertion of tympanostomy tubes. *N Engl J Med* **344**, 1188-95 (2001).
58. Gates, G.A., Avery, C.A., Prihoda, T.J. & Cooper, J.C., Jr. Effectiveness of adenoidectomy and tympanostomy tubes in the treatment of chronic otitis media with effusion. *N Engl J Med* **317**, 1444-51 (1987).
59. Casselbrant, M.L. *et al.* The heritability of otitis media: a twin and triplet study. *JAMA : the journal of the American Medical Association* **282**, 2125-2130 (1999).

60. Rovers, M., Haggard, M., Gannon, M., Koeppen-Schomerus, G. & Plomin, R. Heritability of symptom domains in otitis media: a longitudinal study of 1,373 twin pairs. *American Journal of Epidemiology* **155**, 958-964 (2002).
61. Hafren, L. *et al.* Genetic background and the risk of otitis media. *Int J Pediatr Otorhinolaryngol* **76**, 41-4 (2012).
62. Rye, M.S. *et al.* Unraveling the genetics of otitis media: from mouse to human and back again. *Mammalian genome : official journal of the International Mammalian Genome Society* **22**, 66-82 (2011).
63. Bhutta, M.F. Mouse models of otitis media: strengths and limitations. *Otolaryngol Head Neck Surg* **147**, 611-4 (2012).
64. Brown, S.D., Hardisty-Hughes, R.E. & Mburu, P. Quiet as a mouse: dissecting the molecular and genetic basis of hearing. *Nature reviews.Genetics* **9**, 277-290 (2008).
65. Depreux, F.F. *et al.* Eya4-deficient mice are a model for heritable otitis media. *The Journal of clinical investigation* **118**, 651-658 (2008).
66. Warren, M. *et al.* A Sall4 mutant mouse model useful for studying the role of Sall4 in early embryonic development and organogenesis. *Genesis* **45**, 51-8 (2007).
67. Hernandez, M. *et al.* Myeloid differentiation primary response gene 88 is required for the resolution of otitis media. *J Infect Dis* **198**, 1862-9 (2008).
68. Leichtle, A. *et al.* TLR4-mediated induction of TLR2 signaling is critical in the pathogenesis and resolution of otitis media. *Innate Immun* **15**, 205-15 (2009).
69. Voronina, V.A. *et al.* Inactivation of Chibby affects function of motile airway cilia. *J Cell Biol* **185**, 225-33 (2009).
70. Hardisty-Hughes, R.E. *et al.* A mutation in the F-box gene, Fbxo11, causes otitis media in the Jeff mouse. *Hum Mol Genet* **15**, 3273-9 (2006).

71. Tateossian, H. *et al.* Regulation of TGF-beta signalling by Fbxo11, the gene mutated in the Jeff otitis media mouse mutant. *PathoGenetics* **2**, 5 (2009).
72. Parkinson, N. *et al.* Mutation at the Evi1 locus in Junbo mice causes susceptibility to otitis media. *PLoS genetics* **2**, e149 (2006).
73. Tateossian, H. *et al.* Otitis media in the Tgif knockout mouse implicates TGFbeta signalling in chronic middle ear inflammatory disease. *Hum Mol Genet* (2013).
74. Frede, S., Berchner-Pfannschmidt, U. & Fandrey, J. Regulation of hypoxia-inducible factors during inflammation. *Methods Enzymol* **435**, 405-19 (2007).
75. Giebink, G.S. Otitis media: the chinchilla model. *Microb Drug Resist* **5**, 57-72 (1999).
76. Apicella, M.A. Bacterial otitis media, the chinchilla middle ear, and biofilms. *J Infect Dis* **199**, 774-5 (2009).
77. Kerschner, J.E. *et al.* Partial characterization of normal and Haemophilus influenzae-infected mucosal complementary DNA libraries in chinchilla middle ear mucosa. *Ann Otol Rhinol Laryngol* **119**, 270-8 (2010).
78. Kerschner, J.E. Mucin gene expression in human middle ear epithelium. *The Laryngoscope* **117**, 1666-1676 (2007).
79. Kerschner, J.E. *et al.* Gene expression differences in infected and noninfected middle ear complementary DNA libraries. *Arch Otolaryngol Head Neck Surg* **135**, 33-9 (2009).
80. Kerschner, J.E., Khamphang, P. & Samuels, T. Extending the chinchilla middle ear epithelial model for mucin gene investigation. *Int J Pediatr Otorhinolaryngol* **74**, 980-5 (2010).
81. Kerschner, J.E., Meyer, T.K., Burrows, A., Ehrlich, G. & Post, J.C. Mucin gene cDNA sequence characterization in chinchilla middle ear epithelium. *Int J Pediatr Otorhinolaryngol* **70**, 1449-56 (2006).

82. Kerschner, J.E., Meyer, T.K. & Burrows, A. Chinchilla middle ear epithelial mucin gene expression in response to inflammatory cytokines. *Arch Otolaryngol Head Neck Surg* **130**, 1163-7 (2004).
83. Bakaletz, L.O. Chinchilla as a robust, reproducible and polymicrobial model of otitis media and its prevention. *Expert Rev Vaccines* **8**, 1063-82 (2009).
84. Rye, M.S., Blackwell, J.M. & Jamieson, S.E. Genetic susceptibility to otitis media in childhood. *Laryngoscope* **122**, 665-75 (2012).
85. Sale, M.M. *et al.* Evaluation of 15 functional candidate genes for association with chronic otitis media with effusion and/or recurrent otitis media (COME/ROM). *PloS one* **6**, e22297 (2011).
86. Alpay, H.C. *et al.* Evaluation of the polymorphism in the Toll-like receptor 4 (TLR4) genes of tympanosclerosis patients. *Auris, Nasus, Larynx* (2009).
87. McCormick, D.P. *et al.* Acute otitis media severity: association with cytokine gene polymorphisms and other risk factors. *Int J Pediatr Otorhinolaryngol* **75**, 708-12 (2011).
88. Patel, J.A. *et al.* Association of proinflammatory cytokine gene polymorphisms with susceptibility to otitis media. *Pediatrics* **118**, 2273-2279 (2006).
89. Revai, K. *et al.* Association between cytokine gene polymorphisms and risk for upper respiratory tract infection and acute otitis media. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America* **49**, 257-261 (2009).
90. Nuytinck, L., De Meester, E., Van Thielen, M. & Govaerts, P. Role of mannose-binding lectin (MBL2) genotyping in predicting the risk of recurrent otitis media (rOM). *Adv Exp Med Biol* **586**, 281-90 (2006).
91. Ramet, M., Lofgren, J., Alho, O.P. & Hallman, M. Surfactant protein-A gene locus associated with recurrent otitis media. *J Pediatr* **138**, 266-8 (2001).

92. Pettigrew, M.M. *et al.* Association of surfactant protein A polymorphisms with otitis media in infants at risk for asthma. *BMC medical genetics* **7**, 68 (2006).
93. Ubell, M.L., Khampang, P. & Kerschner, J.E. Mucin gene polymorphisms in otitis media patients. *The Laryngoscope* **120**, 132-138 (2010).
94. Lin, J. *et al.* Analysis by cDNA microarrays of altered gene expression in middle ears of rats following pneumococcal infection. *Int J Pediatr Otorhinolaryngol* **65**, 203-11 (2002).
95. Rye, M.S. *et al.* FBXO11, a regulator of the TGFbeta pathway, is associated with severe otitis media in Western Australian children. *Genes and immunity* **12**, 352-359 (2011).
96. Segade, F. *et al.* Association of the FBXO11 gene with chronic otitis media with effusion and recurrent otitis media: the Minnesota COME/ROM Family Study. *Archives of Otolaryngology--Head & Neck Surgery* **132**, 729-733 (2006).
97. Rye, M.S. *et al.* Genetic and functional evidence for a role for SLC11A1 in susceptibility to otitis media in early childhood in a Western Australian population. *Infect Genet Evol* **16**, 411-8 (2013).
98. Macarthur, C.J. *et al.* Genetic susceptibility to chronic otitis media with effusion: Candidate gene SNPs. *Laryngoscope* (2013).
99. Casselbrant, M.L. & Mandel, E.M. Genetic susceptibility to otitis media. *Current opinion in allergy and clinical immunology* **5**, 1-4 (2005).
100. Herman, P. *et al.* Middle ear cell line that maintains vectorial electrolyte transport. *J Cell Physiol* **154**, 615-22 (1993).
101. Ueyama, S., Jin, S., Rhim, J.S., Ueyama, T. & Lim, D.J. Immortalization of rat middle ear epithelial cells by adeno 12-SV40 hybrid virus. *Ann Otol Rhinol Laryngol* **110**, 132-41 (2001).



102. Jin, S., Gu, X.X., Rhim, J.S. & Lim, D.J. Immortalization of chinchilla middle ear epithelial cells by adenovirus 12-simian virus 40 hybrid virus. *Ann Otol Rhinol Laryngol* **108**, 934-43 (1999).
103. Chun, Y.M. *et al.* Immortalization of normal adult human middle ear epithelial cells using a retrovirus containing the E6/E7 genes of human papillomavirus type 16. *Ann Otol Rhinol Laryngol* **111**, 507-17 (2002).
104. Bowden, D.W. *et al.* Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. *Hum Mol Genet* **19**, 4112-20 (2010).
105. Fearnhead, N.S. *et al.* Multiple rare variants in different genes account for multifactorial inherited susceptibility to colorectal adenomas. *Proc Natl Acad Sci U S A* **101**, 15992-7 (2004).
106. Chen, W.M. *et al.* Significant linkage at chromosome 19q for otitis media with effusion and/or recurrent otitis media (COME/ROM). *BMC medical genetics* **12**, 124 (2011).
107. Jiao, K., Zhou, Y. & Hogan, B.L. Identification of mZnf8, a mouse Kruppel-like transcriptional repressor, as a novel nuclear interaction partner of Smad1. *Molecular and cellular biology* **22**, 7633-7644 (2002).
108. Schroder, K. & Tschopp, J. The inflammasomes. *Cell* **140**, 821-832 (2010).
109. Taxman, D.J., Huang, M.T. & Ting, J.P. Inflammasome inhibition as a pathogenic stealth mechanism. *Cell host & microbe* **8**, 7-11 (2010).
110. Kanneganti, T.D. Central roles of NLRs and inflammasomes in viral infection. *Nature reviews.Immunology* (2010).
111. Valletta, E.A., Rigo, A., Bonazzi, L., Zanolli, L. & Mastella, G. Modification of some markers of inflammation during treatment for acute respiratory

- exacerbation in cystic fibrosis. *Acta Paediatrica (Oslo, Norway : 1992)* **81**, 227-230 (1992).
112. Casselbrant, M.L. *et al.* Otitis media: a genome-wide linkage scan with evidence of susceptibility loci within the 17q12 and 10q22.3 regions. *BMC medical genetics* **10**, 85 (2009).
  113. Hindorff, L.A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci U S A* **106**, 9362-7 (2009).
  114. Consortium, E.P. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science (New York, N.Y.)* **306**, 636-640 (2004).
  115. Rye, M.S. *et al.* Genome-Wide Association Study to Identify the Genetic Determinants of Otitis Media Susceptibility in Childhood. *PloS one* **7**, e48215 (2012).
  116. Liu, J.Z. *et al.* A versatile gene-based test for genome-wide association studies. *Am J Hum Genet* **87**, 139-45 (2010).
  117. Teele, D.W. *et al.* Otitis Media in Infancy and Intellectual Ability, School Achievement, Speech, and Language at Age 7 Years. *The Journal of infectious diseases* **162**, 685-694 (1990).
  118. Organization, W.H. Prevention of hearing impairment from chronic otitis media. in *Strategies for Prevention of Deafness and Hearing Impairment* (WHO/CIBA Foundation Workshop, London, UK, 1996).
  119. Lous, J. *et al.* Grommets (ventilation tubes) for hearing loss associated with otitis media with effusion in children. *Cochrane Database Syst Rev*, CD001801 (2005).
  120. Clarridge, J.E. Impact of 16S rRNA Gene Sequence Analysis for Identification of Bacteria on Clinical Microbiology and Infectious Diseases. *Clinical Microbiology Reviews* **17**, 840-862 (2004).

121. Hamparian, V.V. Rhinoviruses. in *Diagnostic procedures for viral, rickettsial and chlamydial infections* (eds. Lennette, E.H. & Schmidt, N.J.) 562 (American Public Health Association, Washington, D.C., 1979).
122. Jackson, G.G. Transmission of the common cold to volunteers under controlled conditions. I. The common cold as a clinical entity. *Archives of Internal Medicine* **101**, 267 (1958).
123. Jarvinen, A.K. *et al.* Rapid identification of bacterial pathogens using a PCR- and microarray-based assay. *BMC microbiology* **9**, 161 (2009).
124. Tissari, P. *et al.* Accurate and rapid identification of bacterial species from positive blood cultures with a DNA-based microarray platform: an observational study. *Lancet* **375**, 224-230 (2010).
125. Dudley, S., Ashe, K., Winther, B. & Hendley, J.O. Bacterial pathogens of otitis media and sinusitis: Detection in the nasopharynx with selective agar media. *Journal of Laboratory and Clinical Medicine* **138**, 338-342 (2001).
126. Gritzfeld, J., Roberts, P., Roche, L., El Batrawy, S. & Gordon, S. Comparison between nasopharyngeal swab and nasal wash, using culture and PCR, in the detection of potential respiratory pathogens. *BMC Research Notes* **4**, 122 (2011).
127. Winther, B. *et al.* Study of bacteria in the nasal cavity and nasopharynx during naturally acquired common colds. *Acta Oto-Laryngologica* **98**, 315-320 (1984).
128. Winther, B., Gwaltney, J.M., Jr, Mygind, N., Turner, R.B. & Hendley, J.O. Sites of rhinovirus recovery after point inoculation of the upper airway. *JAMA: The Journal of the American Medical Association* **256**, 1763-1767 (1986).
129. Sakakura, Y. *et al.* Mucociliary function during experimentally induced rhinovirus infection in man. *Annals of Otology, Rhinology and Laryngology* **82**, 203-211 (1973).

130. Pedersen, M., Sakakura, Y., Winther, B., Brofeldt, S. & Mygind, N. Nasal mucociliary transport, number of ciliated cells, and beating pattern in naturally acquired common colds. *European journal of respiratory diseases. Supplement* **128**, 355-65 (1983).
131. Sakakura, Y. Changes of mucociliary function during colds. *Eur J Respir Dis Suppl* **128 (Pt 1)**, 348-54 (1983).
132. Gwaltney, J.M. *et al.* Nose Blowing Propels Nasal Fluid into the Paranasal Sinuses. *Clinical Infectious Diseases* **30**, 387-391 (2000).
133. Schwartz, R.H., Pitkaranta, A. & Winther, B. Computed Tomography Imaging of the Maxillary and Ethmoid Sinuses in Children with Short-duration Purulent Rhinorrhea. *Otolaryngology -- Head and Neck Surgery* **124**, 160-163 (2001).
134. Winther, B., Gwaltney, J.M., Phillips, C.D. & Hendley, J.O. Radiopaque contrast dye in nasopharynx reaches the middle ear during swallowing and/or yawning. *Acta Oto-Laryngologica* **125**, 625-628 (2005).
135. Kaiser, L. *et al.* Effects of antibiotic treatment in the subset of common-cold patients who have bacteria in nasopharyngeal secretions. *The Lancet* **347**, 1507-1510 (1996).
136. Allen, E.K. *et al.* Bacteria in the nose of young adults during wellness and rhinovirus colds: detection by culture and microarray methods in 100 nasal lavage specimens. *Int Forum Allergy Rhinol* (2013).
137. Murphy, T.F., Bakaletz, L.O. & Smeesters, P.R. Microbial interactions in the respiratory tract. *Pediatr Infect Dis J* **28**, S121-6 (2009).
138. Peltola, V.T. & McCullers, J.A. Respiratory viruses predisposing to bacterial infections: role of neuraminidase. *Pediatr Infect Dis J* **23**, S87-97 (2004).

139. van den Bergh, M.R. *et al.* Associations between pathogens in the upper respiratory tract of young children: interplay between viruses and bacteria. *PloS one* **7**, e47711 (2012).
140. Brogden, K.A., Guthmiller, J.M. & Taylor, C.E. Human polymicrobial infections. *Lancet* **365**, 253-5 (2005).
141. Blaser, M.J. & Falkow, S. What are the consequences of the disappearing human microbiota? *Nat Rev Microbiol* **7**, 887-94 (2009).
142. Noverr, M.C. & Huffnagle, G.B. The 'microflora hypothesis' of allergic diseases. *Clin Exp Allergy* **35**, 1511-20 (2005).
143. Crowe, C.C., Sanders, W.E., Jr. & Longley, S. Bacterial interference. II. Role of the normal throat flora in prevention of colonization by group A Streptococcus. *J Infect Dis* **128**, 527-32 (1973).
144. Bosch, A.A., Biesbroek, G., Trzcinski, K., Sanders, E.A. & Bogaert, D. Viral and bacterial interactions in the upper respiratory tract. *PLoS Pathog* **9**, e1003057 (2013).
145. Bogaert, D. *et al.* Variability and diversity of nasopharyngeal microbiota in children: a metagenomic analysis. *PloS one* **6**, e17035 (2011).
146. Ling, Z. *et al.* Pyrosequencing analysis of the human microbiota of healthy Chinese undergraduates. *BMC genomics* **14**, 390 (2013).
147. Hilty, M. *et al.* Nasopharyngeal microbiota in infants with acute otitis media. *J Infect Dis* **205**, 1048-55 (2012).
148. Biesbroek, G. *et al.* Seven-valent pneumococcal conjugate vaccine and nasopharyngeal microbiota in healthy children. *Emerg Infect Dis* **20**, 201-10 (2014).

149. Hyde, E.R. *et al.* Nasopharyngeal Proteobacteria are associated with viral etiology and acute wheezing in children with severe bronchiolitis. *Journal of Allergy and Clinical Immunology*.
150. Winther, B., Alper, C.M., Mandel, E.M., Doyle, W.J. & Hendley, J.O. Temporal relationships between colds, upper respiratory viruses detected by polymerase chain reaction, and otitis media in young children followed through a typical cold season. *Pediatrics* **119**, 1069-75 (2007).
151. Wang, Q., Garrity, G.M., Tiedje, J.M. & Cole, J.R. Naive Bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Appl Environ Microbiol* **73**, 5261-7 (2007).
152. van der Zalm, M.M. *et al.* Respiratory pathogens in children with and without respiratory symptoms. *J Pediatr* **154**, 396-400, 400 e1 (2009).
153. Avadhanula, V. *et al.* Respiratory viruses augment the adhesion of bacterial pathogens to respiratory epithelium in a viral species- and cell type-dependent manner. *J Virol* **80**, 1629-36 (2006).
154. Tong, H.H., Weiser, J.N., James, M.A. & DeMaria, T.F. Effect of influenza A virus infection on nasopharyngeal colonization and otitis media induced by transparent or opaque phenotype variants of *Streptococcus pneumoniae* in the chinchilla model. *Infect Immun* **69**, 602-6 (2001).
155. McNamee, L.A. & Harmsen, A.G. Both influenza-induced neutrophil dysfunction and neutrophil-independent mechanisms contribute to increased susceptibility to a secondary *Streptococcus pneumoniae* infection. *Infect Immun* **74**, 6707-21 (2006).
156. Colamussi, M.L., White, M.R., Crouch, E. & Hartshorn, K.L. Influenza A virus accelerates neutrophil apoptosis and markedly potentiates apoptotic effects of bacteria. *Blood* **93**, 2395-403 (1999).

157. Abramson, J.S. & Hudnor, H.R. Role of the sialophorin (CD43) receptor in mediating influenza A virus-induced polymorphonuclear leukocyte dysfunction. *Blood* **85**, 1615-9 (1995).
158. Marom, T., Nokso-Koivisto, J. & Chonmaitree, T. Viral-bacterial interactions in acute otitis media. *Curr Allergy Asthma Rep* **12**, 551-8 (2012).
159. Noverr, M.C., Noggle, R.M., Toews, G.B. & Huffnagle, G.B. Role of antibiotics and fungal microbiota in driving pulmonary allergic responses. *Infect Immun* **72**, 4996-5003 (2004).
160. Molyneaux, P.L. *et al.* Outgrowth of the bacterial airway microbiome after rhinovirus exacerbation of chronic obstructive pulmonary disease. *Am J Respir Crit Care Med* **188**, 1224-31 (2013).
161. Andrews, S. FastQC A quality control tool for high throughput sequence data.
162. Edgar, R.C. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460-1 (2010).
163. Caporaso, J.G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**, 335-6 (2010).
164. DeSantis, T.Z. *et al.* Greengenes, a chimera-checked 16S rRNA gene database and workbench compatible with ARB. *Appl Environ Microbiol* **72**, 5069-72 (2006).
165. Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C. & Knight, R. UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**, 2194-200 (2011).
166. Caporaso, J.G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266-7 (2010).
167. Price, M.N., Dehal, P.S. & Arkin, A.P. FastTree 2--approximately maximum-likelihood trees for large alignments. *PloS one* **5**, e9490 (2010).

168. Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl Environ Microbiol* **71**, 8228-35 (2005).
169. Ramette, A. Multivariate analyses in microbial ecology. *FEMS Microbiol Ecol* **62**, 142-60 (2007).
170. Costello, E.K. *et al.* Bacterial community variation in human body habitats across space and time. *Science* **326**, 1694-7 (2009).
171. Biesbroek, G. *et al.* Deep sequencing analyses of low density microbial communities: working at the boundary of accurate microbiota detection. *PloS one* **7**, e32942 (2012).
172. Consortium, H.M.P. Structure, function and diversity of the healthy human microbiome. *Nature* **486**, 207-14 (2012).
173. Segata, N. *et al.* Metagenomic microbial community profiling using unique clade-specific marker genes. *Nat Methods* **9**, 811-4 (2012).
174. Murphy, T.F. *et al.* Haemophilus haemolyticus: a human respiratory tract commensal to be distinguished from Haemophilus influenzae. *J Infect Dis* **195**, 81-9 (2007).
175. McCrea, K.W. *et al.* Relationships of nontypeable Haemophilus influenzae strains to hemolytic and nonhemolytic Haemophilus haemolyticus strains. *J Clin Microbiol* **46**, 406-16 (2008).
176. Segata, N. *et al.* Computational meta'omics for microbial community studies. *Mol Syst Biol* **9**, 666 (2013).
177. Willner, D. *et al.* Metagenomic analysis of respiratory tract DNA viral communities in cystic fibrosis and non-cystic fibrosis individuals. *PloS one* **4**, e7370 (2009).
178. Nolan, P.M. *et al.* A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* **25**, 440-3 (2000).



179. Hardisty-Hughes, R.E. *et al.* A mutation in the F-box gene, Fbxo11, causes otitis media in the Jeff mouse. *Human molecular genetics* **15**, 3273-3279 (2006).
180. Daly, K.A. *et al.* Epidemiology, natural history, and risk factors: panel report from the Ninth International Research Conference on Otitis Media. *Int J Pediatr Otorhinolaryngol* **74**, 231-40 (2010).
181. Gunasekera, H., Hayson, L., Morris, P. & Craig, J. The global burden of childhood otitis media and hearing impairment: a systematic review. in *9<sup>th</sup> International Symposium on Recent Advances in Otitis* (St. Petersburg, FL, 2007).
182. Grevers, G. Challenges in reducing the burden of otitis media disease: an ENT perspective on improving management and prospects for prevention. *Int J Pediatr Otorhinolaryngol* **74**, 572-7 (2010).
183. Schraff, S.A. Contemporary indications for ventilation tube placement. *Curr Opin Otolaryngol Head Neck Surg* **16**, 406-11 (2008).
184. Toll, E.C. & Nunez, D.A. Diagnosis and treatment of acute otitis media: review. *The Journal of Laryngology & Otology* **126**, 976-983 (2012).
185. Coker Tr, C.L.S.N.S.J. & *et al.* Diagnosis, microbial epidemiology, and antibiotic treatment of acute otitis media in children: A systematic review. *JAMA: The Journal of the American Medical Association* **304**, 2161-2169 (2010).
186. Poehling, K.A. *et al.* Reduction of frequent otitis media and pressure-equalizing tube insertions in children after introduction of pneumococcal conjugate vaccine. *Pediatrics* **119**, 707-715 (2007).
187. Greenberg, D. *et al.* The burden of acute otitis media on the patient and the family. *European Journal of Pediatrics* **162**, 576-581 (2003).
188. Johnston, L.C. *et al.* Tympanic membrane abnormalities and hearing levels at the ages of 5 and 6 years in relation to persistent otitis media and tympanostomy

- tube insertion in the first 3 years of life: a prospective study incorporating a randomized clinical trial. *Pediatrics* **114**, e58-67 (2004).
189. Daly, K.A., Hunter, L.L., Levine, S.C., Lindgren, B.R. & Giebink, G.S. Relationships between otitis media sequelae and age. *Laryngoscope* **108**, 1306-10 (1998).
  190. Harsten, G., Prellner, K., Heldrup, J., Kalm, O. & Kornfalt, R. Recurrent acute otitis media. A prospective study of children during the first three years of life. *Acta Otolaryngol* **107**, 111-9 (1989).
  191. Teele, D.W., Klein, J.O. & Rosner, B. Epidemiology of otitis media during the first seven years of life in children in greater Boston: a prospective, cohort study. *J Infect Dis* **160**, 83-94 (1989).
  192. Teele, D.W., Klein, J.O. & Rosner, B.A. Epidemiology of otitis media in children. *Ann Otol Rhinol Laryngol Suppl* **89**, 5-6 (1980).
  193. De Melker, R.A. & Burke, P.D. Epidemiology of otitis media and the role of the general practitioner in management. *Fam Pract* **5**, 307-13 (1988).
  194. Daly, K.A. *et al.* The family study of otitis media: design and disease and risk factor profiles. *Genet Epidemiol* **13**, 451-68 (1996).
  195. Kvaerner, K.J., Tambs, K., Harris, J.R. & Magnus, P. Distribution and heritability of recurrent ear infections. *Ann Otol Rhinol Laryngol* **106**, 624-32 (1997).
  196. Casselbrant, M.L. *et al.* The heritability of otitis media: a twin and triplet study. *JAMA* **282**, 2125-30 (1999).
  197. Rovers, M., Haggard, M., Gannon, M., Koeppen-Schomerus, G. & Plomin, R. Heritability of symptom domains in otitis media: a longitudinal study of 1,373 twin pairs. *Am J Epidemiol* **155**, 958-64 (2002).
  198. Kvestad, E. *et al.* Recurrent otitis media and tonsillitis: common disease predisposition. *Int J Pediatr Otorhinolaryngol* **70**, 1561-8 (2006).

199. Casselbrant, M.L. *et al.* Otitis media: a genome-wide linkage scan with evidence of susceptibility loci within the 17q12 and 10q22.3 regions. *BMC Med Genet* **10**, 85 (2009).
200. Daly, K.A. *et al.* Chronic and recurrent otitis media: a genome scan for susceptibility loci. *Am J Hum Genet* **75**, 988-97 (2004).
201. Segade, F. *et al.* Association of the FBXO11 gene with chronic otitis media with effusion and recurrent otitis media: the Minnesota COME/ROM Family Study. *Arch Otolaryngol Head Neck Surg* **132**, 729-33 (2006).
202. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867-73 (2010).
203. Li, Y., Willer, C.J., Ding, J., Scheet, P. & Abecasis, G.R. MaCH: Using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet Epidemiol* (2010).
204. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. *Annu Rev Genomics Hum Genet* **10**, 387-406 (2009).
205. Chen, W.M., Manichaikul, A. & Rich, S.S. Testing and Estimating Within-family Association for Dichotomous Traits. *Genetic epidemiology* **34**, 959-959 (2010).
206. Thornton, T. & McPeck, M.S. Case-control association testing with related individuals: a more powerful quasi-likelihood score test. *American Journal of Human Genetics* **81**, 321-337 (2007).
207. Chen, W.M., Manichaikul, A. & Rich, S.S. A generalized family-based association test for dichotomous traits. *Am J Hum Genet* **85**, 364-76 (2009).
208. Jurinke, C., van den Boom, D., Cantor, C.R. & Koster, H. Automated genotyping using the DNA MassArray technology. *Methods Mol Biol* **187**, 179-92 (2002).
209. Willer, C.J., Li, Y. & Abecasis, G.R. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* **26**, 2190-2191 (2010).

210. Chen, W.-M. *et al.* Variations in the G6PC2/ABCB11 genomic region are associated with fasting glucose levels. *The Journal of clinical investigation* **118**, 2620-2628 (2008).
211. Yang, T.-P. *et al.* Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. *Bioinformatics* **26**, 2474-2476 (2010).
212. Gamazon, E.R. *et al.* SCAN: SNP and copy number annotation. *Bioinformatics* **26**, 259-262 (2010).
213. Montgomery, S.B. *et al.* Transcriptome genetics using second generation sequencing in a Caucasian population. *Nature* **464**, 773-777 (2010).
214. Stranger, B.E. *et al.* Patterns of cis regulatory variation in diverse human populations. *PLoS genetics* **8**, e1002639 (2012).
215. Duan, S. *et al.* Genetic architecture of transcript-level variation in humans. *Am J Hum Genet* **82**, 1101-13 (2008).
216. Duggal, P., Gillanders, E.M., Holmes, T.N. & Bailey-Wilson, J.E. Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. *BMC Genomics* **9**, 516 (2008).
217. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**, 1665-74 (2007).
218. Chen, W.M., Manichaikul, A. & Rich, S.S. A generalized family-based association test for dichotomous traits. *American Journal of Human Genetics* **85**, 364-376 (2009).
219. Skol, A.D., Scott, L.J., Abecasis, G.R. & Boehnke, M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet* **38**, 209-13 (2006).

220. Barrett, J.C. Haploview: Visualization and analysis of SNP genotype data. *Cold Spring Harbor protocols* **2009**, pdb.ip71 (2009).
221. Pruim, R.J. *et al.* LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* **26**, 2336-2337 (2010).
222. Ward, L.D. & Kellis, M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* **40**, D930-4 (2012).
223. The ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636-640 (2004).
224. Rosenbloom, K.R. *et al.* ENCODE whole-genome data in the UCSC Genome Browser: update 2012. *Nucleic Acids Res* **40**, D912-D917 (2012).
225. Jia, L. *et al.* Functional Enhancers at the Gene-Poor 8q24 Cancer-Linked Locus. *PLoS Genet* **5**, e1000597 (2009).
226. Yao, X., Remaley, A.T. & Levine, S.J. New Kids on the Block Apolipoprotein Mimetics for Asthma Treatment The Emerging Role of Apolipoproteins in the Pathogenesis and Treatment of Asthma. *Chest* **140**, 1048-1054 (2011).
227. Bochkov, Y.A. *et al.* Molecular modeling, organ culture and reverse genetics for a newly identified human rhinovirus C. *Nat Med* **17**, 627-632 (2011).
228. Lim, K.H. & Fairbrother, W.G. Spliceman--a computational web server that predicts sequence variations in pre-mRNA splicing. *Bioinformatics* **28**, 1031-2 (2012).
229. Ingham, P.W. & McMahon, A.P. Hedgehog signalling: Kif7 is not that fishy after all. *Curr Biol* **19**, R729-31 (2009).
230. Hsu, S.H. *et al.* Kif7 promotes hedgehog signaling in growth plate chondrocytes by restricting the inhibitory function of Sufu. *Development* **138**, 3791-3801 (2011).

231. Li, X. *et al.* Importance of hedgehog interacting protein and other lung function genes in asthma. *The Journal of allergy and clinical immunology* **127**, 1457-1465 (2011).
232. Kim, J.H. *et al.* KIF3A, a Cilia Structural Gene on Chromosome 5q31, and Its Polymorphisms Show an Association with Aspirin Hypersensitivity in Asthma. *J Clin Immunol* (2010).
233. Kumagai, A., Shevchenko, A. & Dunphy, W.G. Treslin collaborates with TopBP1 in triggering the initiation of DNA replication. *Cell* **140**, 349-59 (2010).
234. Lilley, C.E., Schwartz, R.A. & Weitzman, M.D. Using or abusing: viruses and the cellular DNA damage response. *Trends Microbiol* **15**, 119-26 (2007).
235. Kathiresan, S. *et al.* A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. *BMC Med Genet* **8 Suppl 1**, S17 (2007).
236. Brown, M.D. & Sacks, D.B. IQGAP1 in cellular signaling: bridging the GAP. *Trends in Cell Biology* **16**, 242-249 (2006).
237. Kim, H., White, C.D. & Sacks, D.B. IQGAP1 in microbial pathogenesis: Targeting the actin cytoskeleton. *FEBS Letters* **585**, 723-729 (2011).
238. Yamaoka-Tojo, M. *et al.* IQGAP1, a Novel Vascular Endothelial Growth Factor Receptor Binding Protein, Is Involved in Reactive Oxygen Species—Dependent Endothelial Migration and Proliferation. *Circulation Research* **95**, 276-283 (2004).
239. Angelo, L.S. & Kurzrock, R. Vascular Endothelial Growth Factor and Its Relationship to Inflammatory Mediators. *Clinical Cancer Research* **13**, 2825-2830 (2007).
240. Ovádi, J. & Orosz, F. An unstructured protein with destructive potential: TPPP/p25 in neurodegeneration. *BioEssays* **31**, 676-686 (2009).

241. Orosz, F. & Ovádi, J. TPPP orthologs are ciliary proteins. *FEBS Letters* **582**, 3757-3764 (2008).
242. Damoiseaux, R.A., Rovers, M.M., Van Balen, F.A., Hoes, A.W. & de Melker, R.A. Long-term prognosis of acute otitis media in infancy: determinants of recurrent acute otitis media and persistent middle ear effusion. *Family Practice* **23**, 40-45 (2006).
243. Laird, N.M. & Lange, C. Family-based designs in the age of large-scale gene-association studies. *Nat Rev Genet* **7**, 385-94 (2006).
244. Daly, K.A. *et al.* The family study of otitis media: design and disease and risk factor profiles. *Genetic epidemiology* **13**, 451-468 (1996).
245. Kvestad, E. *et al.* Otitis media: genetic factors and sex differences. *Twin research : the official journal of the International Society for Twin Studies* **7**, 239-244 (2004).
246. Kvestad, E. *et al.* Recurrent otitis media and tonsillitis: common disease predisposition. *International journal of pediatric otorhinolaryngology* **70**, 1561-1568 (2006).
247. Allen, E.K. *et al.* A Genome-Wide Association Study of Chronic Otitis Media with Effusion and Recurrent Otitis Media Identifies a Novel Susceptibility Locus on Chromosome 2. *J Assoc Res Otolaryngol* (2013).
248. Rye, M.S. *et al.* Genetic and functional evidence for a locus controlling otitis media at chromosome 10q26.3. *BMC Med Genet* **15**, 18 (2014).
249. Manolio, T.A. Genomewide Association Studies and Assessment of the Risk of Disease. *New England Journal of Medicine* **363**, 166-176 (2010).
250. McCarthy, M.I. & Hirschhorn, J.N. Genome-wide association studies: potential next steps on a genetic journey. *Hum Mol Genet* **17**, R156-65 (2008).

251. Maurano, M.T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190-5 (2012).
252. Yang, Y. *et al.* Protection from lipopolysaccharide-induced pulmonary microvascular endothelial cell injury by activation of hedgehog signaling pathway. *Mol Biol Rep* **38**, 3615-22 (2011).
253. Okuda, J., Hanabusa, A. & Gotoh, N. ExoS of *Pseudomonas aeruginosa* binds to a human KIF7 to induce cytotoxicity in cultured human bronchial epithelial cells. *J Infect Chemother* (2013).
254. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-9 (2009).
255. DePristo, M.A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491-8 (2011).
256. Garrison, E. & Marth, G. Haplotype-based variant detection from short-read sequencing. *eprint arXiv:1207.3907*, 1-9 (2012).
257. Schunter, C., Garza, J.C., Macpherson, E. & Pascual, M. SNP development from RNA-seq data in a nonmodel fish: how many individuals are needed for accurate allele frequency prediction? *Mol Ecol Resour* **14**, 157-65 (2014).
258. ENCODE. Integrated Regulation from ENCODE Tracks.
259. Ku, M. *et al.* Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains. *PLoS Genet* **4**, e1000242 (2008).
260. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S.L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10**, R25 (2009).
261. Bioinformatics, U.G. bigWig Track Format.



262. Hall, T.A. Bioedit: a user friendly biological sequence alignment editory and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**, 95-98 (1999).
263. Taniguchi, K., Anderson, A.E., Sutherland, A.E. & Wotton, D. Loss of Tgif function causes holoprosencephaly by disrupting the SHH signaling pathway. *PLoS Genet* **8**, e1002524 (2012).
264. Yoo, Y.A., Kang, M.H., Kim, J.S. & Oh, S.C. Sonic hedgehog signaling promotes motility and invasiveness of gastric cancer cells through TGF-beta-mediated activation of the ALK5-Smad 3 pathway. *Carcinogenesis* **29**, 480-90 (2008).
265. Nemzek, J.A., Hugunin, K.M. & Opp, M.R. Modeling sepsis in the laboratory: merging sound science with animal well-being. *Comp Med* **58**, 120-8 (2008).
266. Sweet, M.J. & Hume, D.A. Endotoxin signal transduction in macrophages. *J Leukoc Biol* **60**, 8-26 (1996).
267. Hodgson, J.C. Endotoxin and mammalian host responses during experimental disease. *J Comp Pathol* **135**, 157-75 (2006).
268. Song, J.J., Kwon, S.K., Cho, C.G., Park, S.W. & Chae, S.W. Microarray analysis of microRNA expression in LPS induced inflammation of human middle ear epithelial cells (HMEECs). *International journal of pediatric otorhinolaryngology* (2011).
269. Heintzman, N.D. *et al.* Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat Genet* **39**, 311-8 (2007).
270. van den Aardweg, M.T., Schilder, A.G., Herkert, E., Boonacker, C.W. & Rovers, M.M. Adenoidectomy for otitis media in children. *Cochrane database of systematic reviews (Online)* **(1)**, CD007810 (2010).

- 271. Karlsson, T., Turkina, M.V., Yakymenko, O., Magnusson, K.E. & Vikstrom, E. The *Pseudomonas aeruginosa* N-acylhomoserine lactone quorum sensing molecules target IQGAP1 and modulate epithelial cell migration. *PLoS Pathog* **8**, e1002953 (2012).
- 272. Kim, H., White, C.D. & Sacks, D.B. IQGAP1 in microbial pathogenesis: Targeting the actin cytoskeleton. *FEBS Lett* **585**, 723-9 (2011).
- 273. Yamaoka-Tojo, M. *et al.* IQGAP1, a novel vascular endothelial growth factor receptor binding protein, is involved in reactive oxygen species--dependent endothelial migration and proliferation. *Circ Res* **95**, 276-83 (2004).
- 274. Husseman, J., Palacios, S.D., Rivkin, A.Z., Oehl, H. & Ryan, A.F. The role of vascular endothelial growth factors and fibroblast growth factors in angiogenesis during otitis media. *Audiol Neurotol* **17**, 148-54 (2012).
- 275. Prime, S.S., Pring, M., Davies, M. & Paterson, I.C. TGF-beta signal transduction in oro-facial health and non-malignant disease (part I). *Crit Rev Oral Biol Med* **15**, 324-36 (2004).
- 276. Tayeb, L.A. *et al.* Comparative phylogenies of Burkholderia, Ralstonia, Comamonas, Brevundimonas and related organisms derived from rpoB, gyrB and rrs gene sequences. *Res Microbiol* **159**, 169-77 (2008).
- 277. Dauga, C. Evolution of the gyrB gene and the molecular phylogeny of Enterobacteriaceae: a model molecule for molecular systematic studies. *Int J Syst Evol Microbiol* **52**, 531-47 (2002).
- 278. Roth, S.B., Jalava, J., Ruuskanen, O., Ruohola, A. & Nikkari, S. Use of an oligonucleotide array for laboratory diagnosis of bacteria responsible for acute upper respiratory infections. *J Clin Microbiol* **42**, 4268-74 (2004).
- 279. Maccaferri, S., Biagi, E. & Brigidi, P. Metagenomics: key to human gut microbiota. *Dig Dis* **29**, 525-30 (2011).

280. Abecasis, G.R. *et al.* An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56-65 (2012).
281. Altshuler, D.M. *et al.* Integrating common and rare genetic variation in diverse human populations. *Nature* **467**, 52-8 (2010).
282. Qi, L.S. *et al.* Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. *Cell* **152**, 1173-83 (2013).
283. Katoh, Y. & Katoh, M. Characterization of KIF7 gene in silico. *Int J Oncol* **25**, 1881-6 (2004).
284. Cheung, H.O. *et al.* The kinesin protein Kif7 is a critical regulator of Gli transcription factors in mammalian hedgehog signaling. *Science signaling* **2**, ra29 (2009).
285. Putoux, A. *et al.* KIF7 mutations cause fetal hydroletharus and acrocallosal syndromes. *Nat Genet* **43**, 601-6 (2011).
286. Putoux, A. *et al.* Novel KIF7 mutations extend the phenotypic spectrum of acrocallosal syndrome. *J Med Genet* **49**, 713-20 (2012).
287. Dafinger, C. *et al.* Mutations in KIF7 link Joubert syndrome with Sonic Hedgehog signaling and microtubule dynamics. *J Clin Invest* **121**, 2662-7 (2011).
288. Jain, R. *et al.* Temporal relationship between primary and motile ciliogenesis in airway epithelial cells. *Am J Respir Cell Mol Biol* **43**, 731-9 (2010).
289. Endoh-Yamagami, S. *et al.* The mammalian Cos2 homolog Kif7 plays an essential role in modulating Hh signal transduction during development. *Current biology : CB* **19**, 1320-1326 (2009).
290. Nevoux, J. *et al.* Childhood cholesteatoma. *Eur Ann Otorhinolaryngol Head Neck Dis* **127**, 143-50 (2010).
291. Ayache, D. *et al.* Imaging of non-operated cholesteatoma: clinical practice guidelines. *Eur Ann Otorhinolaryngol Head Neck Dis* **129**, 148-52 (2012).

292. Khemani, S., Singh, A., Lingam, R.K. & Kalan, A. Imaging of postoperative middle ear cholesteatoma. *Clin Radiol* **66**, 760-7 (2011).

## Supplement

<b>Table S2.1:</b> Complete data for each NLF sample collected separated by Subject.....	200
<b>Chapter 3 Supplement:</b> Figshare additional files .....	210
<b>Table S4.1:</b> Summary of the design and results of all published genome-wide studies of OM .....	213
<b>Figure S4.1:</b> Haploview haplotype block of the chromosome 2 region .....	215
<b>Table S4.2:</b> Criteria for Classifying Family Members as Affected with COME/ROM for the UMN Study .....	216
<b>Table S4.3:</b> Meta-analysis results for all 53 SNPs tested for replication. ....	217
<b>Table S4.4:</b> Significant results from the three eQTL browsers used for the in silico eQTL analysis for the four SNPs of interest.....	219
<b>Figure S6.1:</b> Relative expression levels of <i>PAI1</i> and <i>SMAD7</i> in HMEEC-1 cells after treatment with TGF $\beta$ .....	220
<b>Figure S6.2:</b> Relative expression levels of <i>GLI1</i> , <i>PTCH1</i> , <i>SUFU</i> , in HMEEC-1 cells and <i>SMAD2</i> after treatment with TGF $\beta$ .....	221
<b>Figure S6.3:</b> Relative expression levels of <i>GLI1</i> , <i>PTCH1</i> , and <i>SUFU</i> in HMEEC-1 cells after treatment with <i>P. aeruginosa</i> LPS.....	222

**Table S2.1-** Data for each NLF sample collected separated by Subject, including Illness status, HRV infection, bacterial load, bacteria detected by microarray, symptom scores, and subjective cold.

Subject # 1	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day 10	Normal
Illness (TSS 7)	ND	0	0	2	0	0	4	1	3	0
HRV infection*	nd	nd	nd	neg	neg	neg	neg	neg	nd	nd
Bacterial load by culture on blood agar	300	50	750	>1000	750	>1000	500	300	50	400
<i>S. epidermidis</i>	+	+	+	+	+	+	+	+	+	+
CNS										
<i>mecA</i>			+	+			+			
<i>S. aureus</i>				+						
<i>S. pneumoniae</i>			+		+				+	
<i>Streptococcus</i> sp.										
<i>H. influenzae</i>			+		+	+	+			
<i>Moraxella catarrhalis</i>										
Enterobacter family			+	+	+					
<i>E. aerogenes</i>										
<i>Neisseria meningitidis</i>										
<i>Neisseria</i> sp.					+			+		
<i>Serratia marcescens</i>										
<i>Kingella kingae</i>										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	0
Runny Nose		0	0	1	0	0	0	0	0	0
Nasal Obstruction		0	0	0	0	0	0	0	0	0
Sore/Scratchy Throat		0	0	0	0	0	2	1	2	0
Cough		0	0	0	0	0	2	0	1	0
Headache		0	0	0	0	0	0	0	0	0
Malaise		0	0	1	0	0	0	0	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	2	0	0	4	1	3	0
Subjective Cold		No	No	No	No	No	No	No	No	No

Subject # 2	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No Illness (TSS 3)	0	0	0	0	1	1	1	0	2	0
No HRV infection	nd	nd	nd	neg	neg	neg	neg	neg	nd	nd
Bacterial load by culture on blood agar	>1000	>1000	>10,000	>1000	>1000	>10,000	>1000	>1000	>10,000	>1000
<i>S. epidermidis</i>	+	+	+	+	+	+	+	+	+	+
CNS										
<i>mecA</i>	+		+	+	+	+	+	+		+
<i>S. aureus</i>										
<i>S. pneumoniae</i>										
<i>Streptococcus</i> sp.										
<i>H. influenzae</i>										
<i>Moraxella catarrhalis</i>										
Enterobacter family										
<i>E. aerogenes</i>	+	+	+	+	+	+	+	+	+	+
<i>Neisseria meningitidis</i>										
<i>Neisseria</i> sp.	+									
<i>Serratia marcescens</i>	+	+	+	+	+	+	+	+		+
<i>Kingella kingae</i>										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	0
Runny Nose		0	0	0	0	0	0	0	0	0
Nasal Obstruction		0	0	0	0	0	0	0	0	0
Sore/Scratchy Throat		0	0	0	1	1	1	0	2	0
Cough		0	0	0	0	0	0	0	0	0
Headache		0	0	0	0	0	0	0	0	0
Malaise		0	0	0	0	0	0	0	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	0	1	1	1	0	2	0
Subjective Cold		No	No	No	Yes	Yes	Yes	No	Yes	No

Subject # 3	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No Illness(TSS 4)	nd	0	0	0	1	0	1	2	0	9
HRV infection*	nd	nd	neg	neg	neg	neg	pos	pos	nd	nd
Bacterial load by culture on blood agar	350	750	>1000	600	750	>1000	500	550	>1000	50
S. epidermidis	+	+	+	+	+	+	+	+	+	+
CNS							+	+		
<i>mecA</i>							+			
S. aureus										
S. pneumoniae										
Streptococcus sp.										
H. influenzae										
Moraxella catarrhalis										
Enterobacter family										
E. aerogenes										
Neisseria meningitidis										
Neisseria sp.										
Serratia marcescens										
Kingella kingae										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	1
Runny Nose		0	0	0	0	0	0	0	0	3
Nasal Obstruction		0	0	0	0	0	0	0	0	2
Sore/Scratchy Throat		0	0	0	0	0	0	1	0	0
Cough		0	0	0	0	0	0	0	0	1
Headache		0	0	0	0	0	0	0	0	1
Malaise		0	0	0	1	0	1	1	0	1
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	0	1	0	1	2	0	9
Subjective Cold		No	No	No	No	No	No	No	No	Yes



Subject # 4	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No illness (TSS 4)	nd	0	0	0	0	2	1	1	1	0
HRV infection	nd	nd	neg	neg	neg	neg	neg	pos	nd	nd
Bacterial load by culture on blood agar	750	>1000	>1000	>1000	>1000	>10,000	>10,000	>1000	>1000	>1000
S. epidermidis	+	+	+	+	+	+	+	+	+	+
CNS	+								+	
<i>mecA</i>			+	+						
S. aureus										
S. pneumoniae										
Streptococcus sp.										
H. influenzae										
Moraxella catarrhalis		+	+			+	+	+	+	+
Enterobacter family										
E. aerogenes										
Neisseria meningitidis										
Neisseria sp.									+	
Serratia marcescens										
Kingella kingae										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	0
Runny Nose		0	0	0	0	1	0	0	0	0
Nasal Obstruction		0	0	0	0	0	0	0	0	0
Sore/Scratchy Throat		0	0	0	0	1	1	1	1	0
Cough		0	0	0	0	0	0	0	0	0
Headache		0	0	0	0	0	0	0	0	0
Malaise		0	0	0	0	0	0	0	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	0	0	2	1	1	1	0
Subjective Cold		No	No	No	No	Yes	No	No	No	No

Subject # 5	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
Illness (TSS 22)	nd	0	0	1	3	4	6	8	5	0
HRV infection*	nd	nd	neg	pos	pos	pos	pos	pos	nd	nd
Bacterial load by culture on blood agar	>1000	1000	>1000	300	550	200	50	100	50	400
S. epidermidis	+	+	+	+	+	+	+	+	+	+
CNS		+			+		+			+
<i>mecA</i>					+	+				
S. aureus										
S. pneumoniae										
Streptococcus sp.										
H. influenzae										
Moraxella catarrhalis										
Enterobacter Family										
E. aerogenes										
Neisseria meningitidis										
Neisseria sp.						+	+	+		
Serratia marcescens										
Kingella kingae										
<b>Symptoms</b>										
Sneezing		0	0	0	0	1	1	0	0	0
Runny Nose		0	0	0	0	1	2	2	1	0
Nasal Obstruction		0	0	1	1	1	1	2	2	0
Sore/Scratchy Throat		0	0	0	2	1	0	1	0	0
Cough		0	0	0	0	0	0	0	2	0
Headache		0	0	0	0	0	1	2	0	0
Malaise		0	0	0	0	0	1	1	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	1	3	4	6	8	5	0
Subjective Cold		No	No	No	Yes	Yes	Yes	Yes	Yes	No

Subject # 6	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No Illness (TSS o)	nd	o	o	o	o	o	o	o	o	o
No HRV infection	nd	nd	neg	neg	neg	neg	neg	neg	nd	nd
Bacterial load by culture on blood agar	150	100	350	750	600	600	200	400	300	350
<i>S. epidermidis</i>	+	+	+	+	+	+	+	+	+	+
CNS	+	+	+	+	+	+	+	+	+	+
<i>mecA</i>	+	+	+	+	+	+	+	+	+	+
<i>S. aureus</i>										
<i>S. pneumoniae</i>		+								
<i>Streptococcus</i> sp.										
<i>H. influenzae</i>										
<i>Moraxella catarrhalis</i>										
Enterobacter family										
<i>E. aerogenes</i>										
<i>Nesseria meningitidis</i>			+							
<i>Neisseria</i> sp.	+	+		+	+	+	+	+	+	
<i>Serratia marcescens</i>										
<i>Kingella kingae</i>										
<b>Symptoms</b>										
Sneezing		o	o	o	o	o	o	o	o	o
Runny Nose		o	o	o	o	o	o	o	o	o
Nasal Obstruction		o	o	o	o	o	o	o	o	o
Sore/Scratchy Throat		o	o	o	o	o	o	o	o	o
Cough		o	o	o	o	o	o	o	o	o
Headache		o	o	o	o	o	o	o	o	o
Malaise		o	o	o	o	o	o	o	o	o
Chilliness		o	o	o	o	o	o	o	o	o
Total		o	o	o	o	o	o	o	o	o
Subjective Cold		No	No	No	No	No	No	No	No	No

Subject # 7	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
Illness (TSS 30)	nd	0	0	0	6	5	10	9	2	0
HRV infection*	nd	nd	neg	neg	neg	neg	pos	pos	nd	nd
Bacterial load by culture on blood agar	>10,000	>1000	>1000	>10,000	>10,000	>10,000	>1000	>10,000	>1000	>1000
S. epidermidis	+	+	+	+	+	+	+	+	+	+
CNS										+
<i>mecA</i>	+		+	+	+	+	+			+
S. aureus										
S. pneumoniae										
Streptococcus sp.										
H. influenzae										
Moraxella catarrhalis										
Enterobacter family										
E. aerogenes										
Nisseria meningitidis										
Neisseria sp.			+	+	+	+			+	+
Serratia marcescens										
Kingella kingae										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	0
Runny Nose		0	0	0	0	1	1	2	0	0
Nasal Obstruction		0	0	0	0	0	1	2	0	0
Sore/Scratchy Throat		0	0	0	0	1	2	2	1	0
Cough		0	0	0	0	0	2	1	1	0
Headache		0	0	0	3	3	3	2	0	0
Malaise		0	0	0	3	0	1	0	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	0	6	5	10	9	2	0
Subjective Cold		No	No	No	Yes	Yes	Yes	Yes	Yes	No

<b>Subject # 8</b>	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No Illness (TSS 6)	nd	0	0	1	1	1	1	2	4	0
HRV infection*	nd	nd	neg	neg	neg	neg	neg	neg	nd	nd
Bacterial load by culture on blood agar	200	300	300	500	750	350	250	150	400	>1000
<i>S. epidermidis</i>	+	+	+	+	+	+	+	+	+	+
CNS	+	+	+		+		+		+	+
<i>mecA</i>						+				+
<i>S. aureus</i>						+			+	+
<i>S. pneumoniae</i>					+		+			
<i>Streptococcus</i> sp.										
<i>H. influenzae</i>							+			
<i>Moraxella catarrhalis</i>										
Enterobacter family										
<i>E. aerogenes</i>										
<i>Nisseria meningitidis</i>										
<i>Neisseria</i> sp.					+			+		
<i>Serratia marcescens</i>										
<i>Kingella Kingae</i>										
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	0	0
Runny Nose		0	0	1	0	0	0	0	0	0
Nasal Obstruction		0	0	0	0	0	0	0	0	0
Sore/Scratchy Throat		0	0	0	1	1	1	2	2	0
Cough		0	0	0	0	0	0	0	2	0
Headache		0	0	0	0	0	0	0	0	0
Malaise		0	0	0	0	0	0	0	0	0
Chilliness		0	0	0	0	0	0	0	0	0
Total		0	0	1	1	1	1	2	4	0
Subjective Cold		No	No	No	No	No	No	No	No	No

Subject # 9	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
Illness (TSS 12)	nd	0	1	1	2	4	0	4	18	1
No HRV infection	nd	nd	neg	neg	neg	neg	neg	neg	nd	nd
Bacterial load by culture on blood agar	>10,000	>1000	>1000	>1000	>1000	>1000	750	>1000	>10,000	1000
<i>S. epidermidis</i>		+	+	+	+	+	+	+		+
CNS										+
<i>mecA</i>			+		+	+		+	+	+
<i>S. aureus</i>										
<i>S. pneumoniae</i>									+	+
<i>Streptococcus</i> Sp.	+			+						
<i>H. influenzae</i>	+	+	+		+	+	+	+	+	+
<i>Moraxella catarrhalis</i>	+	+								
<i>Enterobacter</i> family										
<i>E. aerogenes</i>										
<i>Neisseria meningitidis</i>					+			+		
<i>Neisseria</i> sp.	+						+			+
<i>Serratia marcescens</i>										
<i>Kingella kingae</i>	+			+						
<b>Symptoms</b>										
Sneezing		0	0	0	0	0	0	0	2	0
Runny Nose		0	1	0	0	0	0	0	1	0
Nasal Obstruction		0	0	0	0	1	0	0	2	0
Sore/Scratchy Throat		0	0	1	1	1	0	2	3	0
Cough		0	0	0	0	2	0	0	3	1
Headache		0	0	0	0	0	0	1	2	0
Malaise		0	0	0	1	0	0	1	3	0
Chilliness		0	0	0	0	0	0	0	2	0
Total		0	1	1	2	4	0	4	18	1
Subjective Cold		No	No	No	No	Yes	Yes	Yes	Yes	No

Subject # 10	NW 1	NW 2	NW 3	NW 4	NW 5	NW 6	NW 7	NW 8	NW 9	NW 10
	Normal	Normal	Inoc	Day 1	Day 2	Day 3	Day 4	Day 5	Day10	Normal
No Illness (TSS o)	nd	o	o	o	o	o	o	o	o	o
HRV infection*	nd	nd	nd	neg	neg	pos	pos	neg	nd	nd
Bacterial load by culture on blood agar	>1000	400	300	100	400	50	200	250	300	>1000
S. epidermidis	+	+	+	+	+	+	+	+	+	+
CNS	+	+	+		+				+	+
<i>mecA</i>	+		+							
S. aureus										
S. pneumoniae										
Streptococcus sp.						+				
H. influenzae	+	+	+							+
Moraxella catarrhalis										+
Enterobacter family										
E. aerogenes										
Neisseria meningitidis										
Neisseria sp.									+	+
Serratia marcescens										
Kingella kingae										
<b>Symptoms</b>										
Sneezing		o	o	o	o	o	o	o	o	o
Runny Nose		o	o	o	o	o	o	o	o	o
Nasal Obstruction		o	o	o	o	o	o	o	o	o
Sore/Scratchy Throat		o	o	o	o	o	o	o	o	o
Cough		o	o	o	o	o	o	o	o	o
Headache		o	o	o	o	o	o	o	o	o
Malaise		o	o	o	o	o	o	o	o	o
Chilliness		o	o	o	o	o	o	o	o	o
Total		o	o	o	o	o	o	o	o	o
Subjective Cold		No	No	No	No	No	No	No	No	No

## Chapter 3 Supplement

The additional files and data set supporting the results of this article are available in Figshare (<http://dx.doi.org/10.6084/m9.figshare.832471>). Because of the size of these files, they are not included in this thesis, but will remain publicly available.

### Additional file 1

- Word document (.docx)
- Assessment of illness
- This document describes how subjects were assessed for clinical illness. This information was only used for Additional File 6 (2D PCoA- Ill vs. Not Ill.png).

### Additional file 2

- Qiime taxonomic composition results (compressed) (.tar.gz)
- Taxa Summary Plots for all samples
- Bar and area charts describing the taxonomic composition of all samples analyzed at all taxonomic ranks from genus to phylum.

### Additional file 3

- Qiime taxonomic composition results (compressed) (.tar.gz)
- Taxa Summary Plots organized by subject
- Bar and area charts describing the taxonomic composition of all samples analyzed at all taxonomic ranks from genus to phylum, organized by subject.

### Additional file 4

- Qiime taxonomic composition results (compressed) (.tar.gz)
- Taxa Summary Plots organized by time point



- Bar and area charts describing the taxonomic composition of all samples analyzed at all taxonomic ranks from genus to phylum, organized by time point.

#### **Additional file 5**

- Excel file (.xlsx)
- Tukey-Kramer test of dominant phyla between infected vs. not infected subjects
- Results for each test of the relative abundance of each dominant phyla (Actinobacteria, Firmicutes, and Proteobacteria) between infected and not infected subjects. Includes testing for each comparison between inoculation time point.

#### **Additional file 6**

- Excel file (.xlsx)
- Tukey-Kramer test of dominant phyla between all pairs of subjects
- Results for each test of the relative abundance of the dominant phyla (Actinobacteria, Firmicutes, and Proteobacteria) between every pair of subjects.

#### **Additional file 7**

- Compressed file (.tar.gz)
- Alpha diversity\_150
- Alpha diversity analyses including rarefaction plots. Plots go to minimum sequences/sample (~150)

#### **Additional file 8**

- Compressed file (.tar.gz)
- Alpha diversity\_1000
- Alpha diversity analyses including rarefaction plots. Plots exclude samples with fewer than the mean sequences/sample (~1000)

#### **Additional file 9**

- Compressed file (.tar.gz)
- Beta Diversity\_All body sites\_pre-inoculation NLF
- Beta diversity (both weighted and unweighted Unifrac) analyses/emperor plots for all body sites, including pre-inoculation nasal wash samples.

**Additional file 10**

- Compressed file (.tar.gz)
- Beta Diversity\_All body sites\_all NLF
- Beta diversity (both weighted and unweighted Unifrac) analyses/emperor plots for all body sites, including all nasal wash samples.

**Additional file 11**

- Compressed file (.tar.gz)
- Beta Diversity\_NLF samples only
- Beta diversity (both weighted and unweighted Unifrac) analyses/emperor plots for the nasal wash data only.

**Additional file 12**

- Compressed file (.tar.gz)
- Beta Diversity\_NLF samples only\_by subject
- Beta diversity (both weighted and unweighted Unifrac) analyses/emperor plots for the nasal wash data only, split up by subject.

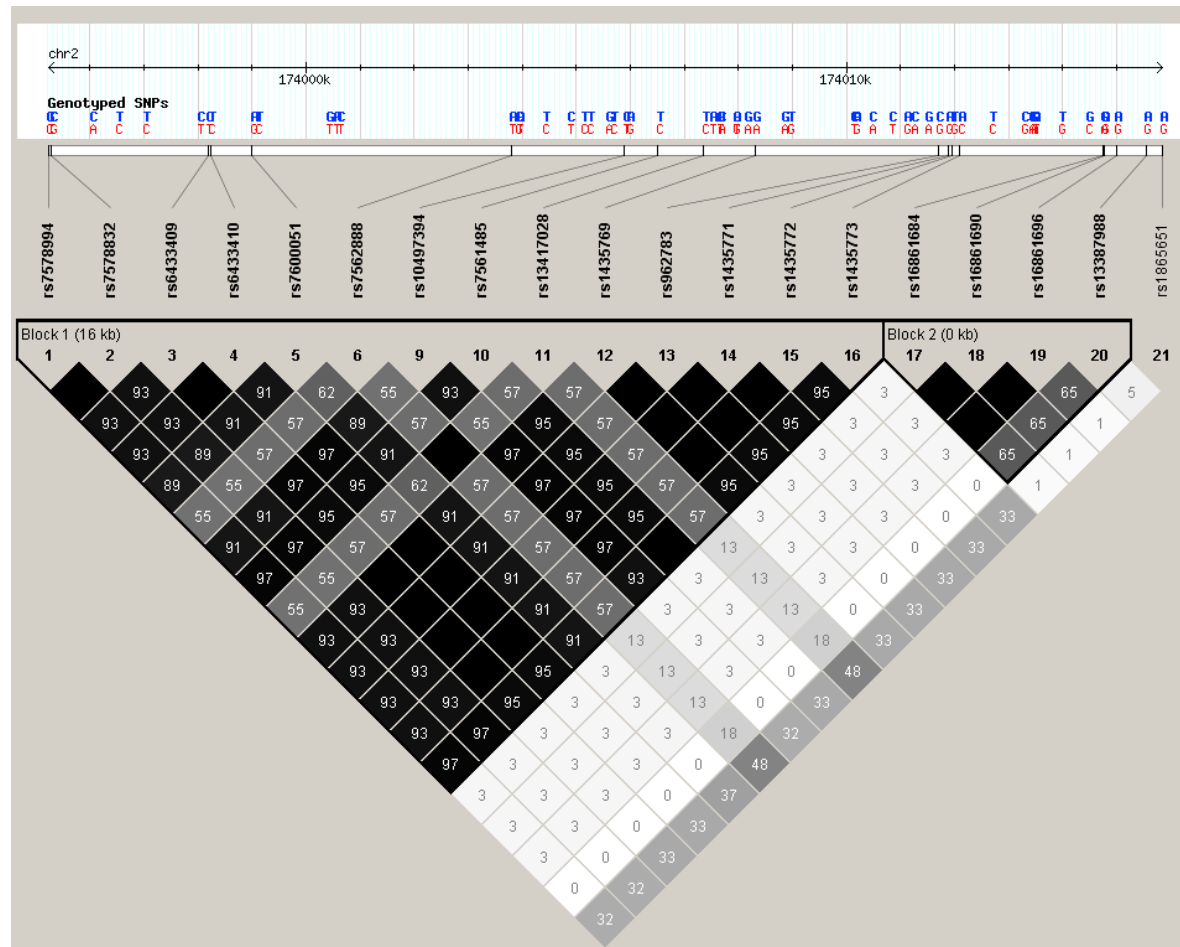
**Table S4.1** Summary of the design and results of all published genome-wide studies of OM<sup>1</sup>

Study	University of Pittsburgh	Western Australia Family Study of Otitis Media	Western Australia Pregnancy Cohort (Raine) Study	University of Minnesota
<b>Phenotype Definition</b>	<ul style="list-style-type: none"> <li>• Affected if he/she had tympanostomy tube surgery at least once for recurrent/persistent OM</li> <li>• Unaffected if he/she had never had tympanostomy tube surgery and had no known history of recurrent/persistent OM</li> <li>• Required one visit to the ENT research Center at the Children's Hospital of Pittsburgh</li> <li>• Detailed history regarding recurrent/persistent OM and risk factors was collected for each family member</li> <li>• Medical records were obtained when possible</li> <li>• Tympanograms were obtained</li> </ul>	<ul style="list-style-type: none"> <li>• Probands were recruited if they had tympanostomy tube surgery for RAOM/COME prior to 3 years of age.</li> <li>• Parents and siblings were recruited.</li> <li>• Affected siblings were considered as affected if defined as having <math>\geq 3</math> episodes of AOM by 3 years of age and/or COME with tubes inserted or recommended.</li> <li>• Parents completed a study-specific questionnaire. Information regarding risk factors was also collected.</li> <li>• Family members with no history of RAOM were classified as unaffected.</li> </ul>	<ul style="list-style-type: none"> <li>• Affected status was assessed using questionnaires and clinical examinations at ages one, two, and three.</li> <li>• Cases defined by examining clinical records from the first 3 years of life for record of the presence of: <ul style="list-style-type: none"> <li>○ Inflamed, retracted, or scarred tympanic membrane</li> <li>○ Middle ear effusion</li> <li>○ Tympanostomy tubes</li> </ul> </li> <li>• Parental reports were also used to define cases if they had <math>\geq 3</math> episodes of AOM by 3 years of age.</li> <li>• Controls were children with no evidence of AOM by 3 years of age.</li> </ul>	<ul style="list-style-type: none"> <li>• Affected if he/she had tympanostomy tube surgery for COME/ROM <b>or</b> if there were at least 2 positive findings from the following data sources: <ol style="list-style-type: none"> <li>1. Ear examination to determine presence of OM sequelae</li> <li>2. Tympanometric testing was performed in subjects at three frequencies to detect abnormal middle ear mechanics, and hearing was screened at 20dB for speech frequencies</li> <li>3. Questionnaire/interview about OM history, risk factors, and related conditions</li> <li>4. Medical record review</li> </ol> </li> </ul>
<b>Population</b>	<ul style="list-style-type: none"> <li>• <math>\geq 2</math> full siblings with history of tubes due to a significant history of OM</li> <li>• Parents and other full siblings with no history of tubes</li> <li>• Exclusion criteria: major congenital malformations, medical conditions with a predisposition for OM, cared for in the Intensive Care Unit as a neonate, been assisted on ventilation, or having known sensorineural hearing loss.</li> </ul>	<ul style="list-style-type: none"> <li>• Family-based cohort with probands, parents and any affected siblings recruited.</li> <li>• Exclusion criteria: predisposing medical conditions including immune deficiencies, craniofacial abnormalities, and specific genetic syndromes.</li> </ul>	<ul style="list-style-type: none"> <li>• Longitudinal cohort of children whose mothers were recruited from antenatal clinics.</li> <li>• Exclusion criteria: diagnosed with cleft palate</li> </ul>	<ul style="list-style-type: none"> <li>• Probands (children treated with tubes for COME/ROM) and their families (parents, siblings, and other relatives) were recruited</li> <li>• Exclusion criteria: adopted or if diagnosed with Down Syndrome, craniofacial anomaly, another genetic disorder with otologic complications, or tubes for a diagnosis other than OM</li> </ul>

<p><b>Results</b></p>	<ul style="list-style-type: none"> <li>• Evidence of linkage at 17q12-LOD 2.83</li> <li>• Suggestive linkage at: <ul style="list-style-type: none"> <li>◦ 10q22.3</li> <li>◦ 7q33</li> <li>◦ 6p25.1</li> <li>◦ 4p15.2</li> </ul> </li> <li>• Combined Caucasian and African American dataset improved linkage at 10q22.3 from <math>P=1.81 \times 10^{-3}</math> to <math>P=2.6 \times 10^{-4}</math></li> </ul> <hr/> <ul style="list-style-type: none"> <li>• Replication study of UMN top GWAS hits found significant association at rs10497394 (<math>P=4.7 \times 10^{-5}</math>)</li> </ul>	<ul style="list-style-type: none"> <li>• Replication of UPitt's 10q22.3 linkage peak at two SNPs: <ul style="list-style-type: none"> <li>◦ rs9418832 (<math>P_{\text{adj-PC}}=7.48 \times 10^{-5}</math>) an intronic SNP in <i>DOCK1</i></li> <li>◦ rs7922424 (<math>P_{\text{adj-PC}}=9.47 \times 10^{-6}</math>) an intergenic SNP</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Top associations: <ul style="list-style-type: none"> <li>◦ rs13386850 (<math>P=1.32 \times 10^{-6}</math>)</li> <li>◦ rs13408922 (<math>P=1.32 \times 10^{-6}</math>)</li> </ul> </li> <li>• Top hit from PC adjusted analysis of full imputed dataset: <ul style="list-style-type: none"> <li>◦ rs6755194 (<math>P=8.3 \times 10^{-7}</math>)</li> </ul> </li> <li>• New candidate genes for OM <ul style="list-style-type: none"> <li>◦ <i>CAPN14</i></li> <li>◦ <i>GALNT14</i></li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>• Evidence of linkage at 10q-LOD 3.78, <math>P=3.0 \times 10^{-5}</math></li> <li>• Evidence of linkage at 19q-LOD 2.61, <math>P=5.3 \times 10^{-4}</math></li> <li>• Conditional linkage at 3p: <ul style="list-style-type: none"> <li>◦ [10q] LOD 2.43</li> <li>◦ [19q] LOD 1.84</li> </ul> </li> <li>• Most significant association in discovery GWAS= rs1110060 (<math>P=9.1 \times 10^{-7}</math>)</li> <li>• Most significant result= rs10497394 <ul style="list-style-type: none"> <li>◦ <b>GWAS</b> <math>P=2.9 \times 10^{-5}</math></li> <li>◦ <b>Replication</b> <math>P=4.7 \times 10^{-5}</math></li> <li>◦ <b>Meta-analysis</b> <math>P=1.52 \times 10^{-8}</math></li> </ul> </li> <li>• New candidate genes for OM <ul style="list-style-type: none"> <li>◦ <i>KIF7</i></li> <li>◦ <i>TICRR</i></li> </ul> </li> </ul>
-----------------------	---	--	---	---

<sup>1</sup>Allen EK, Manichaikul A, Sale MM. Genetic contributors to otitis media: agnostic discovery approaches. Curr Allergy Asthma Rep. 2014 Feb;14(2):411. doi: 10.1007/s11882-013-0411-7.

**Figure S4.1** Haploview haplotype block of the chromosome 2 region including replicated SNP rs10497394 (chr2:173995238-174015555, hg18) using the Gabriel Block method and using the  $r^2$  color scheme. Numbers represent  $r^2 \times 100$  and black blocks represent  $r^2=1$ .



**Table S4.2** Criteria for Classifying Family Members as Affected with COME/ROM for the UMN Study

Data source	Abnormal (evidence of ROM/COME history)*	
Ear exam	Retraction Atrophy Tympanosclerosis Cholesteatoma	Perforation Drainage Tympanostomy tube Monomer
Tympanogram	<b>Age &gt;10 yrs (a)</b> $Y_1 \leq 0.3$ or $>1.7$ mmho (b) $TW_2 \geq 180$ daPa (b) Resonant frequency $<630$ Hz or $>1400$ Hz Abnormal van Huysse pattern at 630 or 1400 Hz  <b>Age <math>\leq 10</math> yrs (c)</b> $Y_1 \leq 0.3$ or $\geq 1.1$ mmho $TW_2 \geq 150$ daPa Resonant frequency $<710$ Hz or $>1400$ Hz Abnormal van Huysse pattern at 710 or 1400 Hz	
Self-reported history	<b>Adult</b> Had tympanostomy tubes $\geq 3$ OM episodes in a year or 6 episodes before age 6 History of perforation, otorrhea, tympanosclerosis, mastoidectomy, cholesteatoma, tympanoplasty.  <b>Child</b> Had tympanostomy tubes $\geq 3$ OM episodes in a year or 6 episodes before age 6 Tympanostomy tubes      Otorrhea OME $\Rightarrow$ 2 months      Perforation Tympanosclerosis      Mastoidectomy Cholesteatoma      Tympanoplasty $\Rightarrow$ 3 OM episodes in a year or 6 episodes by age 6	
Medical record		

\* A data source was considered abnormal if  $\geq 1$  finding was present.

(a):  $<5$ th or  $95$ th percentile ranges for adults [1]

(b):  $Y_1$  = static admittance,  $TW_2$  = tympanometric width.

(c):  $<5$ th or  $>95$ th percentile ranges for children [2,3,4]

#### Reference

1. Margolis RH, Goycoolea HG (1993) Multifrequency tympanometry in normal adults. *Ear Hear* 14: 408-413.
2. Koebsell KA, Margolis RH (1986) Tympanometric gradient measured from normal preschool children. *Audiology* 25: 149-157.
3. Margolis RH, Heller JW (1987) Screening tympanometry: criteria for medical referral. *Audiology* 26: 197-208.
4. Margolis RH, Hunter LL, Rykken JR, Giebink GS (1993) Effects of otitis media on extended high-frequency hearing in children. *Ann Otol Rhinol Laryngol* 102: 1-5.

**Table S4.3** Meta-analysis results for all 53 SNPs tested for replication.

Chr	SNP	Position	Allele 1	Allele 2	Z	P-Value	Direction
2	rs10497394	174,005,905	A	G	-5.659	1.52E-08	--
15	rs1110060	87,991,052	A	G	-4.102	4.09E-05	--
5	rs386057	738,748	A	G	-3.978	6.94E-05	--
15	rs10775247	87,927,125	T	C	3.616	0.0002995	++
15	rs4932232*	87,915,313	A	G	3.603	0.000314	++
15	rs9672286*	87,975,793	T	C	3.435	0.0005935	++
15	rs4932231	87,914,936	T	C	3.421	0.0006233	++
2	rs13025485	105,669,606	A	G	3.316	0.0009129	++
15	rs11855211*	87,920,618	C	G	3.268	0.001082	++
3	rs6438779	124,381,419	T	C	3.142	0.001679	++
15	rs11073878*	87,944,388	T	G	3.118	0.001823	++
15	rs3803530*	87,977,077	A	C	3.034	0.002412	++
11	rs12271647	125,129,266	A	C	3.008	0.002626	++
1	rs2268550	166,615,795	A	G	-2.984	0.002843	--
15	rs2350479	87,948,591	A	G	2.925	0.003449	++
15	rs12914042*	87,973,137	C	G	-2.906	0.003657	--
15	rs11856299	81,331,475	T	C	2.685	0.00725	++
4	rs10006892	65,991,076	A	G	-2.471	0.01348	--
18	rs17666735	28,717,818	T	C	-2.46	0.0139	--
4	rs2608816	39,103,204	T	C	-2.447	0.01441	--
15	rs2621228*	81,350,958	A	G	2.398	0.01648	++
14	rs11160664	101,201,298	T	C	2.324	0.02013	++
6	rs1513498	67,228,643	A	G	2.285	0.02231	++
1	rs7511842	165,909,808	T	C	-2.285	0.02234	--
15	rs1972873	87,875,160	T	G	2.204	0.02756	++
4	rs7667228	8,117,895	T	C	2.179	0.02933	++
15	rs1256453*	81,340,902	T	G	2.159	0.03085	+-
6	rs10499006	96,400,203	A	G	-2.068	0.03866	-+
4	rs6816903	187,337,378	T	C	1.957	0.05036	+-
4	rs10027488	187,344,843	A	G	-1.903	0.05709	-+
4	rs2036912	187,336,037	A	G	1.85	0.06424	+-
4	rs1364968	129,873,577	T	C	1.826	0.06786	+-

9	rs10780347	81,792,544	T	C	1.685	0.09199	+-
11	rs4354680	43,345,907	A	G	1.593	0.1111	+-
4	rs7692380	129,886,861	A	G	1.525	0.1273	+-
15	rs12324875	80,310,460	A	G	-1.521	0.1282	-+
3	rs4686753	187,434,342	T	C	1.512	0.1304	+-
5	rs4701137	178,958,798	A	C	-1.278	0.2012	-+
8	rs7459512	74,172,491	A	G	1.278	0.2012	+-
8	rs7821093	74,177,456	T	C	1.25	0.2114	+-
8	rs11992993	74,165,638	A	G	1.242	0.2143	+-
8	rs7006404	74,159,701	T	C	-1.121	0.2621	-+
16	rs7204986	24,220,269	T	C	1.058	0.29	+-
3	rs1629816	10,311,291	A	G	0.842	0.3998	+-
15	rs920563*	87,953,690	A	T	0.739	0.4599	+-
1	rs1361049	111,710,236	G	T		Failed Assay Design	
5	rs4701136	178,958,156	G	A		Failed Assay Design	
15	rs11852287*	81,732,738	G	C		Failed Assay Design	
3	rs3773770	139,221,620	T	C		Failed QC	
15	rs17360083*	81,361,085	C	A		Failed QC	
15	rs11639117*	87,911,954	T	C		Failed QC	
3	rs6808013	187,432,923	A	G		Failed Genotyping	
15	rs11259905*	81,740,755	A	C		Failed Genotyping	

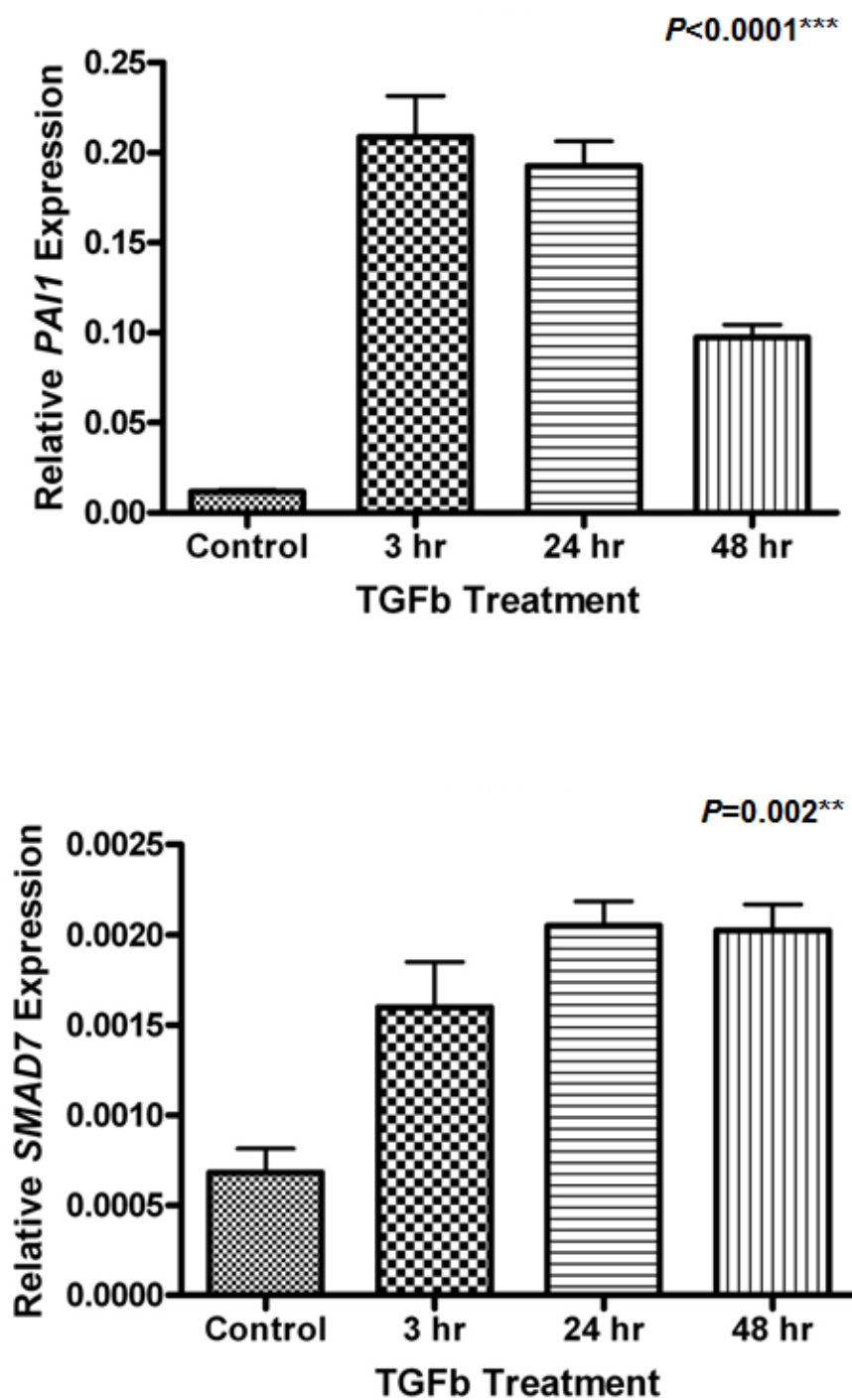
\*Indicates SNPs imputed from discovery GWAS



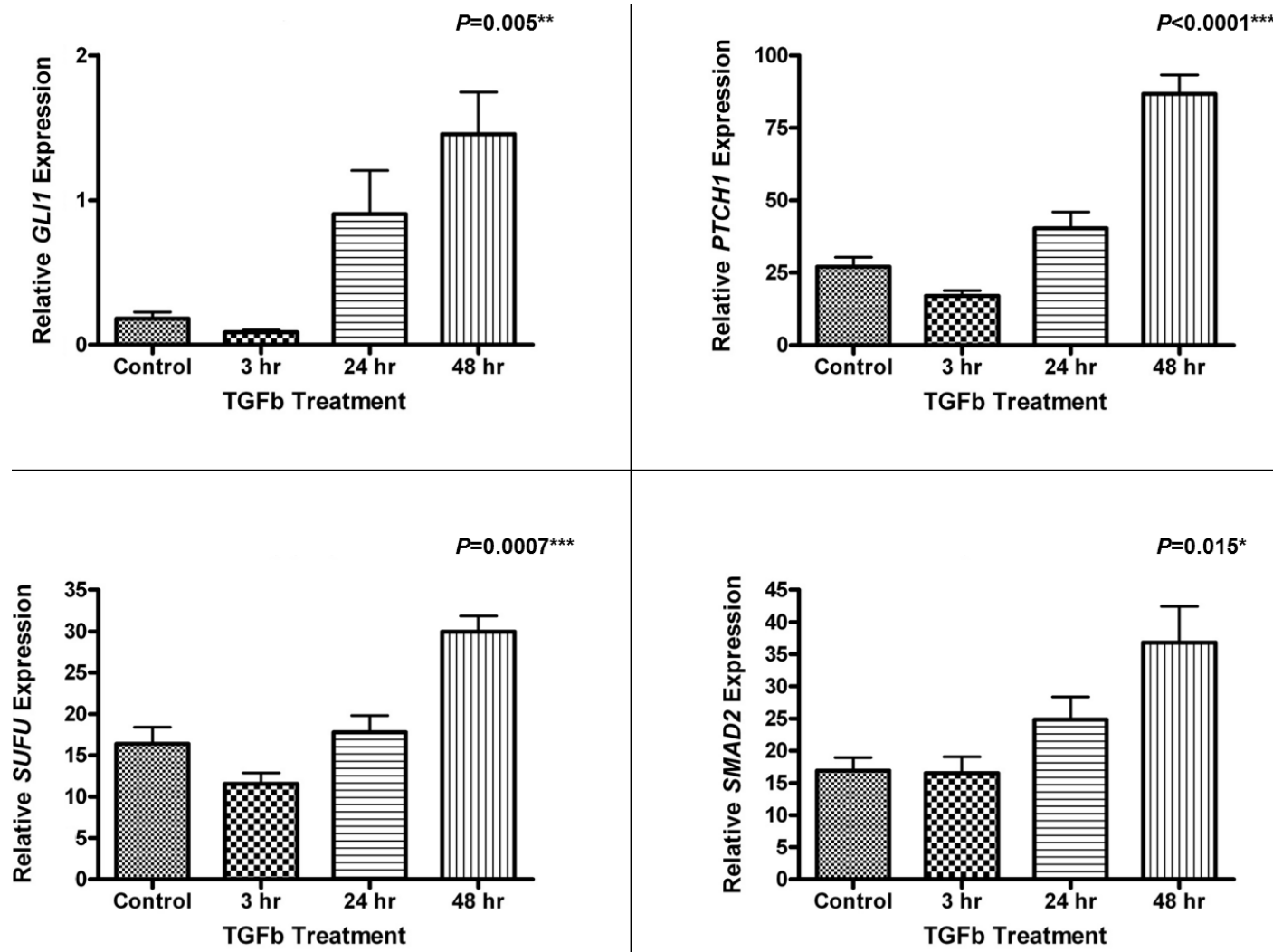
**Table S4.4** Significant results from the three eQTL browsers used for the in silico eQTL analysis for the four SNPs of interest.

Chr	SNP	Database	Gene	Population	P value	Cis/Trans
2	rs10497394	eQTL browser @ pritchard lab	<i>LDLR</i>	CEU	6.00x10 <sup>-5</sup>	Trans (Chr19)
5	rs386057	Wellcome Trust Sanger Institute's Genevar	<i>CEP72</i>	JPT	3.00x10 <sup>-4</sup>	Cis
15	rs10775247	SNP SNP and CNV annotation database	<i>RETN</i>	CEU	4.00 x10 <sup>-5</sup>	Trans (Chr19)
		SNP SNP and CNV annotation database	<i>LDLRAP1</i>	YRI	1.00 x10 <sup>-5</sup>	Trans (Chr1)
		SNP SNP and CNV annotation database	<i>NNT</i>	YRI	6.00 x10 <sup>-5</sup>	Trans (Chr5)
		SNP SNP and CNV annotation database	<i>TTC13</i>	YRI	9.00 x10 <sup>-5</sup>	Trans (Chr1)
15	rs1110060	eQTL browser @ pritchard lab	<i>IQGAP1</i>	CEU	3.07 x10 <sup>-5</sup>	Cis (exonQTL)

**Figure S6.1:** Relative expression levels of PAI1 and SMAD7 in HMEEC-1 cells after treatment with TGF $\beta$



**Figure S6.2:** Relative expression levels of *GLI1*, *PTCH1*, *SUFU*, and *SMAD2* in HMEEC-1 cells after treatment with TGF $\beta$



**Figure S6.3:** Relative expression levels of *GLI1*, *PTCH1*, and *SUFU* in HMEEC-1 cells after treatment with *P. aeruginosa* LPS

