THE UNIVERSITY OF CHICAGO

GENETIC AND ENVIRONMENTAL MODIFIERS OF IMMUNE PROFILES IN TWO US FOUNDER POPULATIONS

A DISSERTATION SUBMITTED TO

THE FACULTY OF THE DIVISION OF THE BIOLOGICAL SCIENCES

AND THE PRITZKER SCHOOL OF MEDICINE

IN CANDIDACY FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

DEPARTMENT OF HUMAN GENETICS

BY

MICHELLE MARIE STEIN

CHICAGO, ILLINOIS

JUNE 2017

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Carole Ober for her wonderful guidance and mentorship over my years as a graduate student at the University of Chicago. She has been a remarkable role model and I am so appreciative of the time and attention she has given to me for feedback on writing, presentations and development as a scientist. I feel extremely fortunate to have had her as an advisor and so grateful to be a part of her lab.

I would also like to thank my committee members, Yoav Gilad, Anne Sperling, and Matthew Stephens for their constructive comments and insightful feedback throughout the years on my thesis project. I felt privileged that three such outstanding scientists would spend time advising me and my work. I really appreciated having such a diverse array of scientific outlooks influence my thinking. I would also like to thank Donata Vercelli and Erika von Mutius for their advice and insight in the Amish Hutterite project. I learned so much and gained new perspective from both of them. I would also like to thank Cara Hrusch for being a great and generous collaborator through the last five years, covering several projects and countless meetings.

I would be remiss to not thank the past and present members of the Ober lab, who have created a wonderfully positive environment to work and learn. In particular, I would like to thank Minal Çalışkan, Jessie Nicodemus-Johnson, and Jessica Chong for their help and advice early in my graduate career, William Wentworth-Sheilds, Tiratat Patana-anake, and Oren Livne for technical advice and tips, and Chris Billstrand, Raluca Nicolae, and Katie Naughton for help with sample processing and good conversations. I would like to specially thank Sahar Mozaffari, Emma Thompson, and especially, and Katie Igartua for their advice and friendship over the years.

Without the consent and enthusiastic participation of the Hutterite and Amish in our studies, my thesis would end here. They deserve special acknowledgment for welcoming us into their homes (and barns), and being so willing to let us into their lives for a time. I feel so privileged to have had the conversations and experiences with these two wonderful groups of people.

I would also like to thank some of my earliest mentors in science, who cultivated my love of learning: Matthew Cole, Elizabeth Ramseyer, Howard Swider, Debbie Schmid, and Michelle Whaley. I would also like to thank all of my friends that I have met in graduate school. I never expected to meet such a wonderful group of people, and to maintain that friendship over nearly seven years.

I would especially like to thank James Fuller, who has been an incredible partner these last five years. I am so grateful for his intellectual and emotional support as we went through our graduate student journeys together. I count myself so lucky to have met him.

Finally, I would like to thank my family. My parents have always been there to support me, in matters both weighty and trivial, and my brother has always been ready with a joke. I have always been able to turn to them for encouragement and guidance, even if they only understood a portion of what I was talking about. I can't thank them enough for their encouragement, care, and love.

ABSTRACT

In common, complex diseases like asthma, the environment shapes how genotype affects phenotype. An understanding of how the environment, from the cellular to the population level, affects biological responses is critical in broadening our understanding of the genetic risks of disease, and may also help to explain the estimated heritabilities of complex diseases like asthma. In my dissertation, I investigated the genetic and environmental modifiers of immune profiles to gain greater understanding in inter-individual variation in immune response, and how population differences can affect asthma and allergic disease. These studies were completed in two US founder populations of European descent, the Amish and the Hutterites. By comparing the immune profiles of schoolchildren between these two unique populations, we were able to assess the effect of different levels of microbial exposures on asthma prevalence and innate immunity. I also examined the effects of variation on immune response in peripheral blood leukocytes (PBLs) in two separate studies. In the first, I conducted a close investigation of a clinically relevant polymorphism, rs1801274, in FcγIIA receptor on gene expression and cytokine responses to anti-CD3+anti-CD28 antibodies in whole blood and reported a unique response profile in heterozygotes for rs1801274 compared to the major and minor allele homozygotes. Finally, I conducted a quantitative trait loci (QTL) mapping study of cytokine levels in the supernatants of LPS-treated whole blood using a candidate gene approach. Only one association was significant after permutations, suggesting that cytokine QTL studies should be conducted in a more restricted set of cell types.

Chapter 1

Introduction

1.1 'Missing' heritability of complex disease

The last 15 years have seen an explosion in the number of identified genetic variants that contribute to risk of common, complex disease, ranging from cardiovascular (atherosclerosis[1]), to metabolic (type 2 diabetes[2]), to psychiatric diseases (schizophrenia[3]). These discoveries have been made possible due to the invention of cheap, high throughput genotyping arrays, which made it feasible to interrogate hundreds of thousands of single nucleotide polymorphisms (SNPs) in large numbers of individuals in a genome-wide association study (GWAS). Though GWAS for a wide range of diseases successfully identified variants that contribute to risk, these SNPs only explained a small portion of the estimated heritability of these complex diseases[4]. Perhaps unsurprisingly, identifying genetic variation in diseases with significant clinical, environmental, and genetic heterogeneity was not so simple. Instead, teasing apart the genetic causes of complex diseases necessitated more sophisticated investigation: of genetic regulation underlying specific disease-related phenotypes; of environmental exposures that may affect disease risk independently of or interaction with previously unobserved risk SNPs; and of the nature and molecular underpinnings of the disease itself.

1.2 Asthma as a classic example of complex disease

In every respect, asthma typifies the missing heritability, tangled environmental risks, and clinical heterogeneity observed in common, complex diseases. Indeed, asthma can be thought of as a set of related but heterogeneous diseases, all marked by chronic airway inflammation and periods of reversible airway obstruction[5,6]. In asthmatic individuals, there is also significant

airway remodeling and bronchial hyper-responsivness[5,7], which can be triggered by inhalation of allergens or other stimuli, like respiratory viral infections[8-10], air pollution[11-13], cigarette smoke[14-16], or exercise[17-20]. Asthma is estimated to affect as many as 300 million people worldwide, most of them children, and contributes to nearly 250,000 deaths per year[21,22]. Disease onset can occur at any age, but childhood-onset (before puberty) and adult-onset (after puberty) asthma are generally considered different subtypes[23,24]. Within the United States, the overall prevalence of childhood asthma is 8.4%, and varies among ethnicities, ranging between 7.4% in individuals of European descent and 13.9% in Puerto Ricans, with some variation due to differences in diagnostic criteria[25].

Studies of the pathogenesis of asthma have long demonstrated the importance of airway-infiltrating Th2 cells and the Th2-secreted cytokines IL-13, IL-4, and IL-5[26-30] in promoting allergic inflammation and, in mouse models, airway remodeling[7,31-33]. Airway remodeling may also be driven by the mechanical stress of airway constriction[34]. More recently, the role of the innate immune system in asthmatic inflammation has been better appreciated. In particular, inflammation-damaged airway epithelial cells release cytokines (IL-25, IL-33, TSLP) that activate innate cells like eosinophils, basophils, NKT cells, and the newly discovered ILC2 cells[28,35-37]. ILC2 cells are innate in origin, but release Th2 cytokines, providing a possible link between Th2 inflammation and innate immune mechanisms[35-38].

1.3 Risk factors for asthma

1.3.1 Genetic contributors to asthma risk

The causes of asthma are a complex constellation of genetic and early-life environmental factors, as well as the interactions between them[39]. Family and twin studies from the past four decades

have established a significant genetic component to asthma[40,41], including a heritability estimate of 54% from a recent large twin-study meta-analysis of common, complex disease[42]. Early candidate gene studies[43] and family-based linkage (Table 1.1) identified genes and genomic regions associated with asthma[43-46].

Table 1.1. Prominent genomic regions associated with asthma or atopy phenotypes identified from linkage studies [45-47.]

Chromosomal band	Gene	Chromosomal band	Gene
2q14.1	*DPP10*	12q21-24	*IFNG, STAT6*
4q		13q12-14	*PHF11, SETDB2*
5q23-31	*CD14,* cytokine genes, *SPINK5*	14q22.1	*PTGR*
6p21-24	*HLA-G,* other *HLA* genes	16q21-23	
7p14.3	*GPRA*	19q	
11q13-21	*FCER1G*	20p13	*ADAM33*

However, these linkage studies generally lacked the precision to identify the causal gene and/or polymorphisms. Following the advent of cheap, high throughput genotyping arrays, genome-wide association studies allowed for greater resolution to identify variants, and even genes, that contribute to risk of asthma. Recent large meta-analyses of Europeans and diverse populations of North Americans have consolidated a core list of approximately 10 asthma-associated loci (Table 1.2)[48,49]. Many of these variants, notably the 17q12-21 locus, had not been previously identified using linkage studies.

Table 1.2. Variants associated with asthma at $P<5\times10^{-8}$ in the North American meta-analysis [49] or European meta-analysis [48].

SNP	Chr:position	Chromosomal band	Alleles	Nearest Gene	Meta-analysis P
rs10173081	2:102323780	2q12.1	C/T	*IL1RL1*	1.4×10^{-8} *
rs3771166	2:102352654	2q12.1	G/A	*IL18R1*	3.4×10^{-9} §
rs1837253	5:110429771	5q22.1	C/T	*TSLP*	7.3×10^{-10} *
rs9273349	6:32733847	6p21	T/C	*HLA-DQ*	7.0×10^{-14} §
rs1342326	9:6180076	9p24.1	A/C	*IL33*	9.2×10^{-10} §
rs744910	15:65233839	15q22.33	G/A	*SMAD3*	3.9×10^{-9} §
rs11078927	17:35317931	17q21.1	C/T	*GSDMB*	1.2×10^{-14} *
rs3894194	17:35375519	17q21.1	G/A	*GSDMA*	4.6×10^{-9} §
rs2284033	22:35863980	22q12.3	G/A	*IL2RB*	1.2×10^{-8} §

* Identified in meta-analysis across all ethnicities in Torgerson *et al.*[49]
§ Identified in meta-analysis total sample in Moffat *et al.*[48]

When combined, these risk variants and genomic regions represent only a small portion of the estimated heritability of asthma[48], underscoring that most of the genetic risk factors for asthma have yet to be identified. In this regard, asthma is typical of the 'missing heritability' problem that has dogged human geneticists for nearly a decade[4]. To overcome this challenge, alternative approaches, such as assessing the impact of rare variants[50], or characterizing epigenetic variation[51] have been carried out, but have not contributed significantly to the missing heritability problem in asthma.

1.3.2 Epidemiological factors associated with asthma risk

In parallel with the ongoing development of genetics research, epidemiologic studies of environmental risks for asthma have made significant advances, despite the difficulty in assessing single environmental exposures in the context of many others. The development of allergic sensitization in early life has long been noted as a strong predictor to the development of childhood asthma[52], particularly if a child is sensitized to multiple allergens[53,54].

One of the largest risk factors for asthma is having a mother with asthma, for which odds ratios range from 3.17 to 5.0, depending on the study[55-57]. Because paternal asthma status does not confer the same increase in asthma risk, this finding highlights that the *in utero* environment may be especially important in shaping susceptibility to asthma[55]. Exposure to viral infection in early life also increases asthma risk. Infants hospitalized with respiratory syncytial virus (RSV) before one year were found to have a 10-fold (30% vs. 3%) higher prevalence of asthma at age 7 compared to controls[58], and children with lower respiratory tract RSV illness before age 3 had increased risk (OR=4.3) of frequent wheeze at age 6[59]. In a birth cohort selected with one or both parents having a previous asthma diagnosis, 87% of children (26 out of 30) who wheezed with rhinovirus infection before age 3 went on to have an asthma diagnosis by age six[8].

Other studies have focused on purely environmental exposures, and how they may increase asthma risk and/or trigger exacerbations. Prenatal exposure to maternal smoking shows robust associations with risk for asthma at age 6 (Odds Ratio=1.39)[60], and post-natal exposure to tobacco smoke similarly increases the risk of childhood asthma[61] wheeze[62,63], and plays a role in asthma exacerbations[64]. Exposure to visible mold is likewise associated with asthma (OR=1.49) and wheeze (OR=1.68)[65]. Air pollution, especially levels of gasses from automobile emissions, has also been correlated with a higher prevalence of asthma in children[66] and increases in the severity of asthma exacerbations[67]. Additionally, unconventional natural gas development ("fracking") in residential areas has also recently been associated with increased risk of asthma exacerbations[68].

1.3.3 Gene by environment interactions in asthma

The identification of gene by environment interactions in asthma have added a layer of complexity in characterizing risk factors for asthma – but have also been critical in understanding how genetic susceptibility is modified by the environment. A striking example of an asthma GEI is rhinovirus infection with wheeze in early life and genotype at the 17q12 asthma susceptibility locus. While both 17q12 genotype and rhinovirus wheezing illness independently increased risk for asthma, this increase in susceptibility is principally due to the enormous increase in risk in individuals with genotype TT at rs7216389 in the 17q12 locus who had rhinovirus wheezing illness in the first year of life (OR=26.1, compared to OR=5.2 for rhinovirus with wheeze and OR=2.3 for TT genotype alone)[69]. In the case of a polymorphism in CD14, individuals with the C allele and exposed to low levels of endotoxin (lipopolysaccharide, or LPS) increases risk for asthma, while individuals with the T allele exposed to high levels of endotoxin have increased risk for asthma, highlighting the complexity and interactions between genetic risk and environmental exposures in asthma[70].

1.4 Towards an organized framework of environmental protection from asthma: the biodiversity hypothesis

The critical role of environmental exposures on asthma risk is illustrated by the dramatic rise in the prevalence of asthma in the late 20th and early 21st century, too sudden to be explained by genetic drift or selection. That this increase in prevalence was observed in Westernized, highly developed nations led to the formation of the hygiene hypothesis[71,72]. The effect of this hypothesis is best demonstrated in epidemiologic studies of children growing up in the farms in central Europe compared to their non-farming neighbors[73,74]. In these populations, children who

lived on traditional dairy farms in early life had a 5-fold lower prevalence of asthma (1.4%) compared to their non-farming neighbors (11.8%)[73]. A larger questionnaire-based study found a similar farm protection against asthma (adjusted OR= 0.68), hay fever (aOR= 0.43), atopic dermatitis (aOR=0.80), and atopic sensitization (aOR=0.54)[75]. A recent birth cohort of children raised in rural areas across central Europe and Finland (PASTURE) revealed that exposure to a farm environment also protects against respiratory tract infections and fever[76]. Levels of specific IgE antibodies against common allergens is also decreased in farming children (12.4%) compared to their non-farming peers (32.9%)[73]. This protection from asthma and allergy has been attributed to sustained contact with livestock animals, as well as consumption of unpasteurized cows' milk[73,74,77-81]. Similarly, in non-farming homes, some studies have suggested that exposure to dogs and cats in early life can provide some protection against asthma and allergic disease[82-84], and day care before age one was associated with lower levels of asthma in later childhood and puberty[85]. Taken together, these protective exposures are likely a proxy for sustained, high levels of microbial exposure[86]. In particular, animal barns contain a diverse array of bacteria, fungi, and even plant pollen[87-89]. Supporting this idea are the observations that levels of various microbial products, like endotoxin, muramic acid, and extracellular polysaccharides of fungal origin, are inversely related to the prevalence of asthma in farming communities[74,78,90,91].

These findings fit into the framework of the emerging biodiversity hypothesis, developed in studies conducted in Karelia, a region straddling urbanized Finland and rural Russia. Like children growing up on farms in central Europe, the former Soviet region of Karelia continued to live on small, subsistence farms and have low prevalence of asthma and allergic disease, while people living in the Finnish region of Karelia underwent rapid economic development and have a much higher prevalence of asthma and allergic disease[92]. The loss of animal, plant, and microbial

diversity following urbanization, Haahtela *et al*.[93] argue, leads to immune system dysfunction in early life, which, in turn, contributes to increased risk of asthma and allergic disease in later life[93]. Consequently, characterization of the profiles and mechanisms of immune protection or dysregulation stemming from specific environmental exposures or stimuli has become a major focus in asthma research.

1.4.1 Farm exposures modulate immune system response

Studies investigating the mechanisms of protection conferred by high levels of microbial exposure have revealed numerous differences in immune profiles and responses. Early studies investigating the molecular underpinnings of the farm effect identified increased expression of *CD14* and the innate immune receptors *TLR2* and *TLR4* in whole blood from children whose mothers were exposed to barns during pregnancy, suggesting that higher microbial exposures modify immune cell responses to bacteria and fungi[94,95]. In mononuclear cord blood cells stimulated with phorbol 12-myristate 13-acetate/ionomycin (PMA), cells from farm infants had significantly higher levels of the proinflammatory cytokines TNFα and IFNγ, and further, IFNγ was negatively correlated with levels of specific IgE, suggesting that changes in cytokine production due to prenatal farm exposures may have downstream effects related to decreased specific IgE[96,97], a marker for allergic disease. Modified cytokine production was also observed in children at age 4.5 years in the PASTURE birth cohort study. Unstimulated peripheral blood mononuclear cells in farm children produced more IL-10, IL-12, and IFNγ compared to non-farm children. Notably, following stimulation with LPS, farm children produced less TNF than non-farming children, suggestive of a tolerogenic effect of farm exposures[98]. These studies highlight some of the diverse effects on the immune system in children exposed to farming

environments, and demonstrate the importance of assessing immune response to specific environmental stimuli (e.g., LPS) to understand how environmental exposures, such as living on a farm, affects risk for asthma.

1.5 Farming populations in the United States: the Amish

The Amish are a religious community that emerged in the late 17th century after a schism within the Swiss Brethren community, an Anabaptist group from Switzerland formed during the upheavals of the Protestant Reformation in the century before[99]. This split came out of disagreements arising between the stricter religiosity of people living in hill regions, the Oberländers, and the peoples living in the Emmental valley[100]. Both of these factors survive today as the Amish and the Mennonites, respectively. After years of religious persecution, the Amish immigrated to the North America, settling in Pennsylvania, and later, Ohio and Indiana. Today, nearly a quarter of a million Americans identify as Amish. Most still follow the traditional farming lifestyle of their ancestors, eschewing modern technology[101].

1.6 Farming populations in the United States: the Hutterites

Like the Amish, the Hutterites are an Anabaptist religious sect formed during the Protestant Reformation by their founder and namesake, Jakob Hutter, in Tyrol, Austria. Their faith and communal style of living meant the Hutterites encountered significant religious persecution, including the execution of Hutter in 1536, forcing the Hutterites to move throughout Europe during the 16th-18th century and resulting in several population bottlenecks[102]. Following a century of stability and population growth in Russia, 443 Hutterites emigrated to the Dakota territory in the U.S. between 1874-1877, settling on three communal farms, which eventually

gave rise to the three main branches of the Hutterite brethren: Schmiedeleut, Dariusleut, and Lehrerleut, each with their own specific traditions and cultural norms[102,103]. Each branch has remained reproductively isolated from the others since the first World War[103]. Today, there are more than 40,000 Hutterites living in the northern Plains states and central Canada in more than 400 communal farms (called colonies)[104]. Nearly all modern Hutterites are descendants of approximately 90 ancestors who lived in the 18th century on the Wischenka colony in Russia[103]. Due to their communal lifestyle, diet and environmental exposures are remarkably similar across and within colonies[104,105].

1.6.1 Asthma risk and prevalence in Amish and Hutterite children

While both the Amish and the Hutterites are farming populations, there is a striking difference in the prevalence of asthma and allergy between the two communities. Amish children (ages 6-10) are typical of European children who grew up on farms, with a low prevalence of asthma (5.2%) and allergic sensitization (7.2%)[106]. In marked contrast, 21.3% of Hutterite children have asthma, and 33.3% showed evidence of allergic sensitization by skin prick test[107]. However, both groups are similar with respect to lifestyle factors known to affect asthma risk – both groups have large sibship sizes, consume unpasteurized raw milk, have low rates of childhood obesity, long nursing durations, diets rich in fat and salt, high rates of childhood vaccinations, and no indoor pets. For these reasons, studies of Amish and Hutterite children circumvent the inherent challenges of studying the lifestyles of farming children and their non-farming peers and provide excellent subjects in which to isolate the protective effects of farm exposures.

1.7 Dissertation overview

The primary objective of my thesis was to examine and characterize how the environment – at the population and at the cellular level – can provide insight into genetic variation in immune response, immune function, and asthma risk. To do so, I utilized an approach leveraging the unique contrast between Amish and Hutterite farming populations, and investigated cytokine responses to both innate and adaptive immune stimulation. In Chapter 2, I report on our study characterizing and contrasting the genetics, immune profiles and environments of Amish and Hutterite schoolchildren and highlight how innate immunity is specifically shaped by the microbial exposures in airborne dust. The results of this study were published in *The New England Journal of Medicine* (Innate immunity and asthma risk in Amish and Hutterite farm children, 375(5), 411-421, 2016). In Chapter 3, I extend our understanding of a clinically important polymorphism in the FcIIγA receptor by measuring cytokine and gene expression responses to anti-CD3 and anti-CD28 IgG, with an emphasis on illustrating differences between heterozygotes and major allele homozygotes. In Chapter 4, I describe the results of a LPS-treated cytokine QTL association study using a candidate gene approach to identify regulatory and/or coding polymorphisms associated with cytokine responses following innate immune stimulation.

Chapter 2

Innate immunity and asthma risk in Amish and Hutterite farm children

2.1 Abstract[1]

The Amish and the Hutterites are U.S. agricultural populations whose lifestyles are remarkably similar in many respects but whose farming practices, in particular, are distinct; the former follow traditional farming practices whereas the latter use industrialized farming practices. The populations also show striking disparities in the prevalence of asthma, and little is known about the immune responses underlying these disparities. We studied environmental exposures, genetic ancestry, and immune profiles among 60 Amish and Hutterite children, measuring levels of allergens and endotoxins and assessing the microbiome composition of indoor dust samples. Whole blood was collected to measure serum IgE levels, cytokine responses, and gene expression, and peripheral blood leukocytes were phenotyped with flow cytometry. The effects of dust extracts obtained from Amish and Hutterite homes on immune and airway responses were assessed in a murine model of experimental allergic asthma. Despite the similar genetic ancestries and lifestyles of Amish and Hutterite children, the prevalence of asthma and allergic sensitization was 4 and 6 times as low in the Amish, whereas median endotoxin levels in Amish house dust was 6.8 times as high. Differences in microbial composition were also observed in dust samples from Amish and Hutterite homes. Profound differences in the proportions, phenotypes, and functions of innate immune cells were also found between the two groups of children. In a mouse model of experimental allergic asthma, the intranasal instillation of dust

---

[1] Reproduced with permission from: Stein MM, Hrusch CL, and Gozdz J., *et al.* Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* **375**, 411-421, doi:10.1056/NEJMoa1508749 (2016). Copyright Massachusetts Medical Society.

extracts from Amish but not Hutterite homes significantly inhibited airway hyperreactivity and eosinophilia. These protective effects were abrogated in mice that were deficient in MyD88 and Trif, molecules that are critical in innate immune signaling. The results of our studies in humans and mice indicate that the Amish environment provides protection against asthma by engaging and shaping the innate immune response.

2.2 Introduction

Many genetic risk factors have been reported to modify susceptibility to asthma and allergy[108,109], but the dramatic increase in the prevalence of these conditions in westernized countries in the past half-century suggests that the environment also plays a critical role[110]. The importance of environmental exposures in the development of asthma is most exquisitely illustrated by epidemiologic studies conducted in Central Europe that show significant protection from asthma and allergic disease in children raised on traditional dairy farms. In particular, children's contact with farm animals and the associated high microbial exposures[78,91] have been related to the reduced risk[74,90]. However, the effect of these traditional farming environment on immune responses is not well defined.

To address this gap in knowledge, we designed a study that compares two distinctive U.S. farming populations – the Amish of Indiana and the Hutterites of South Dakota – that recapitulate the differences in the prevalences of asthma and allergy observed in farmers and nonfarmers in Europe. These two particular groups of farmers originated in Europe – the Amish in Switzerland and the Hutterites in South Tyrol – during the Protestant Reformation and then emigrated to the United States in the 1700s and 1800s, respectively. Both groups have since remained reproductively isolated[101,105]. Their lifestyles are similar with respect to most of the factors know to influence the risk of asthma, including large sibship size, high

rates of childhood vaccination, diets rich in fat, salt, and raw milk, low rates of childhood obesity, long durations of breast-feeding, minimal exposure to tobacco smoke and air pollution, and taboos against indoor pets. However, whereas the Amish practice traditional farming, live on single-family dairy farms, and use horses for fieldwork and transportation, the Hutterites live on large, highly industrialized, communal farms. Strikingly, the prevalence of asthma in Amish versus Hutterite schoolchildren is 5.2% versus 21.3% and the prevalence of allergic sensitization is 7.2% versus 33.3%, as previously reported[106,107].

2.3 Methods

2.3.1 Overview

We characterized the immune profiles of Amish and Hutterite schoolchildren. Furthermore, we used mouse models of asthma to study the effect of the environment on airway responses and to create a mechanistic framework for the interpretation of our observations in humans.

2.3.2 Study participants and study oversight.

In November 2012, we studied 30 Amish children 7 to 14 years of age who lived in Indiana, and in December 2012 we studied 30 Hutterite children who lived in South Dakota and were matched with the Amish schoolchildren for sex and for age within 1 year. Written informed consent was obtained from the parents and written assent was obtained from the children. One parent of each child responded to a questionnaire on asthma symptoms and previous diagnoses. The study was approved by the institutional review boards at the University of Chicago and at St. Vincent Hospital in Indianapolis.

Table 2.1. Demographic and clinical characteristics of the study populations

	Amish (N=30)	Hutterite (N=30)
Median age yrs (range)	11 (8-14)	12 (7-14)
Girls (no.)	10	10
Sibships (no.)	15	14
Children with asthma (no.)	0	6
Positivity for allergen-specific IgE (no.)		
>0.7kUA/L	5	9
>3.5kUA/L	2	9
Serum IgE (kU/liter)		
Median	21	64
Interquartile range	10, 57	15, 288

UA denotes allergen-specific unit

2.3.3 Blood-sample collection and analysis

Whole blood was collected in tubes that contained culture medium alone, medium plus 0.1μg per milliliter of lipopolysaccharide, or medium plus 0.4 μg per milliliter of anti-CD3 plus 0.33 μg per milliliter of anti-CD28 monoclonal antibodies (TruCulture Blood Collection System, Myriad RBM). After incubation at 37ºC for 30 hours, supernatant and cells were frozen for use in gene-expression and cytokine studies. Levels of 26 cytokines were measured with the use of the Milliplex Map Human Th17 Magnetic Bead Panel (EMD Millipore) or enzyme-linked immune-absorbent assay (eBiosciences) in the supernatant, in accordance with standard protocols. Additional blood was collected to obtain peripheral blood leukocytes for flow cytometry and DNA isolation, and serum was collected for IgE studies (as described in Section 2.7.2).

Cyropreserved human peripheral-blood leukocytes were incubated for 10 minutes with pooled human IgG antibodies (FcX, Biolegend) to block nonspecific antibody binding before undergoing surface staining with fluorescently conjugated antibodies (Supplementary Table 2.1) and intracellular staining for FoxP3 (eBioscience). Flow cytometry data were acquired on an

LDRFortessa cell analyzer (BD Biosciences), and acquisition data were analyzed with FlowJo software (Tree Star).

2.3.4 Dust collection and extract preparation

Electrostatic dust collectors were placed in one bedroom and the living room in each of 10 Amish and 10 Hutterite homes to collect airborne house dust. All 10 Amish homes and 9 of 10 Hutterite homes housed children who participated in the study. After 1 month, dust was analyzed for endotoxin and allergen levels, and extracts were prepared for studies in mice. In addition, a vacuum was used to collect dust from the living-room floor in Amish homes and from mattresses in Amish and Hutterite homes for use in microbiome studies (as described in Section 2.7.5). Aqueous extracts of house dust from Amish and Hutterite homes were prepared as described in Section 2.7.10.

2.3.5 Genetic studies

RNA was extracted from thawed cells with the use of the AllPrep DNA/RNA Mini Kits (Qiagen). RNA underwent complementary DNA synthesis and was then hybridized to HumanHT-12 v4 Expression BeadChip arrays (Illumina). A common set of 118,789 single-nucleotide polymorphisms (SNPs) was genotyped or imputed in the 60 children in the study (see Supplementary Figure 2.1).

2.3.6 Mouse models

We instilled 50 µl of house-dust extract intranasally every 2 to 3 days (for a total of 14 times) into 7-week old BALB/c mice (Harlan Laboratories), beginning on day 0. The mice had been

sensitized intraperitoneally with 20 μg of ovalbumin (grade V, Sigma) plus alum (Pierce) on days 0 and 14 were challenged intranasally with 50 μg of ovalbumin on days 28 and 38. Beginning 5 days before day 0, we also instilled 50 μl of dust extract from Amish homes intranasally every 2 to 3 days (for a total of 14 times) in 7-week old, C57BL6 wild-type, MyD88-deficient mice[111] and in mice deficient in both MyD88 and Trif[112] (Jackson Laboratories). These mice were sensitized intraperitoneally with 20 μg of ovalbumin plus alum on days 0 and 14 and challenged intranasally with 75 μg of ovalbumin on days 26, 27, and 28.

2.3.7 Statistical analysis

Statistical analyses were performed with the use of R or Prism software (GraphPad). Differences in the distributions of median cytokine levels between Amish and Hutterite children were determined with use of a Wilcoxon signed-rank test. We used linear regression to identify differentially expressed genes in the Amish and Hutterite untreated samples of peripheral-blood leukocytes. The methods of Benjamini and Hochberg[113] were used to control the false discovery rate. For flow cytometric studies and the study in mice, differences in cell populations and airway resistance were assessed with the used of an unpaired Student's t-test. Additional details on sample processing, quality control, and statistical analysis for all methods described here are provided in Section 2.7.11.

2.4 Results

2.4.1 Asthma and allergic-sensitization rates and genetic ancestry

None of the Amish children and six (20%) of the Hutterite children had asthma, rates similar to hose reported in earlier studies[106,107]. Levels of total serum IgE and the number of children whose

levels of IgE against common allergens were high (defined as more than 3.5 kUA [allergen-specific unit] per liter) were lower in the Amish group than in the Hutterite group (Table 2.1, and Supplementary Table 2.2). No statistical differences were observed in levels of serum Ig isotypes other than IgE (Supplementary Figure 2.2).

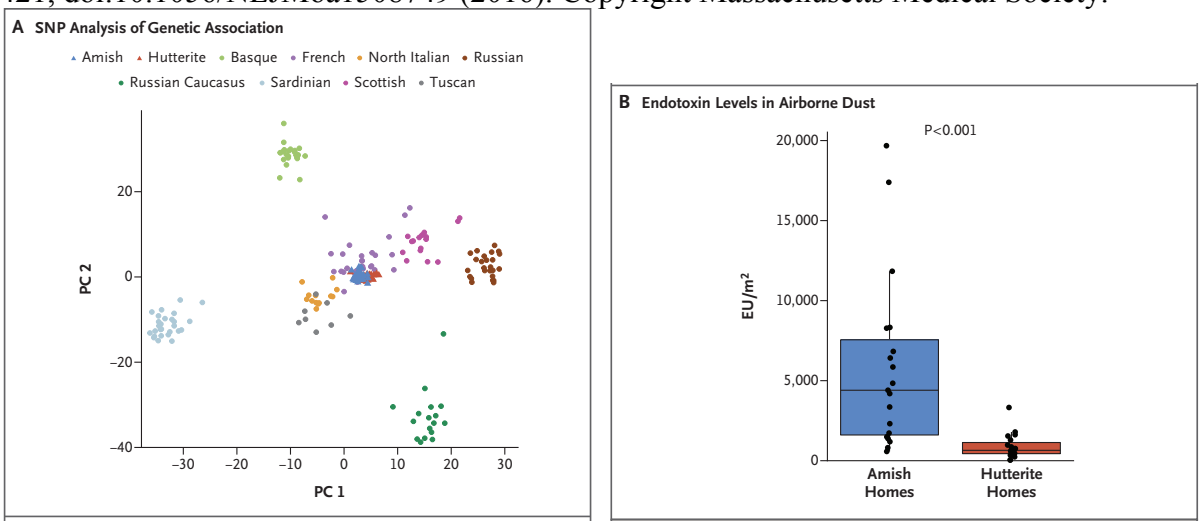
To evaluate whether the differences in the prevalence of asthma and of allergic sensitization and differences in IgE level could be attributed to population history, we assessed ancestry by conducting principal components analysis and compared allele frequencies using genomewide SNPs. These studies revealed remarkable genetic similarities between the Amish children and the Hutterite children, as compared with other European populations [114] (Panel A in Figure 2.1, and Supplementary Figure 2.3).

2.4.2 Exposures to allergens, microbes, and microbial products

Common allergens (from cats, dogs, house-dust mites, and cockroaches) were detectable in air-borne dust from 4 of 10 Amish and 1 of 10 Hutterite homes (Supplementary Table 2.3). In contrast, endotoxin levels were measurable in airborne dust from all 20 homes, and median levels were strikingly higher (6.8 times as high) in Amish homes than in Hutterite homes (4399 endotoxin units [EU] per square meter vs. 648 EU per square meter, $P<0.001$) (Figure 2.1, Panel B). Analysis of a single pooled sample of mattress dust from each population revealed different profiles of the relative abundance of bacteria at the family level (Supplementary Figure 2.4).

Figure 2.1. Ancestries and environments of Amish and Hutterite children.

Panel A shows a principal components plot of the first two principal components (PC 1 and PC 2) of the analysis of 72,034 single-nucleotide polymorphisms (SNPs). Amish and Hutterite genotypes were projected onto the sample space created by Human Genome Diversity Project (HGDP) for European populations (15). Panel B shows endotoxin levels in airborne dust from 10 Amish and 10 Hutterite homes. Box-and-whisker plots show a horizontal line indicating median value, a box representing the interquartile range, and whiskers showing the 95% confidence interval. The P value was calculated with the use of the Wilcoxon rank-sum test. EU denotes endotoxin units. Reproduced with permission from: Stein MM, Hrusch CL, and Gozdz J., *et al.* Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* **375**, 411-421, doi:10.1056/NEJMoa1508749 (2016). Copyright Massachusetts Medical Society.



2.4.3 Composition and phenotype of peripheral-blood leukocytes

Peripheral blood leukocytes from Amish children had increased proportions of neutrophils, decreased proportions of eosinophils, and similar proportions of monocytes as compared with samples from Hutterite children (Figure 2.2, Panel A). Neutrophils from Amish children expressed lower levels of the chemokine receptor CXCR4 and the adhesion molecules CD11b and CD11c than did neutrophils from Hutterite children, suggesting that these cells may have recently emigrated from the bone marrow (Figure 2.2, Panel B). Although proportions of monocytes were similar in Hutterite and Amish children, monocytes from Amish children, unlike

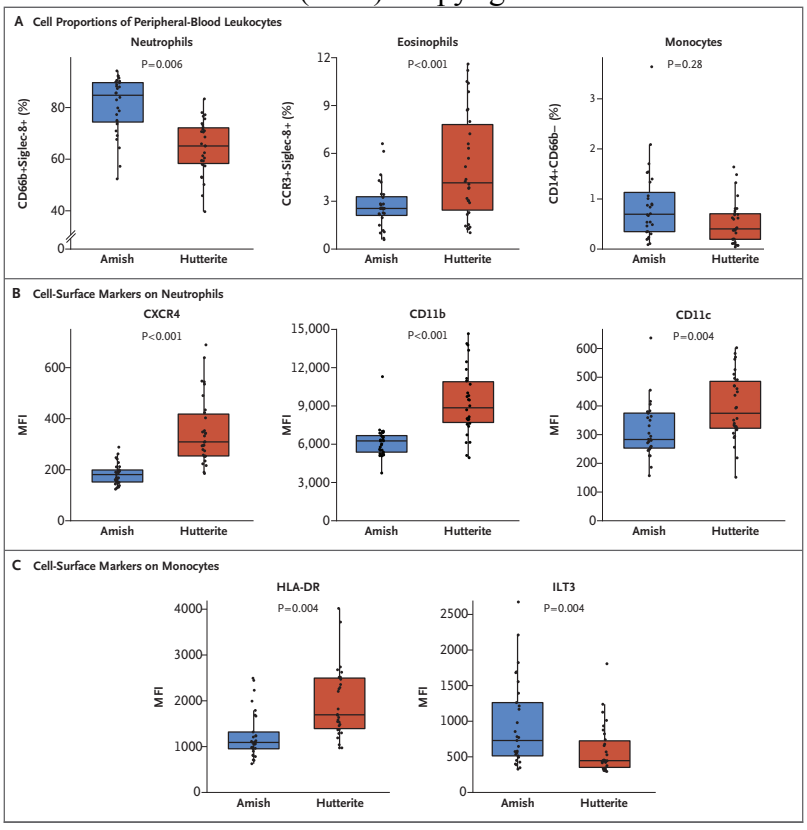
those from Hutterite children, exhibited a suppressive phenotype characterized by lower levels of human leukocyte antigen DR (HLA-DR) and higher levels of the inhibitory molecule immunoglobulin-like transcript 3 (ILT3) [115,116] (Figure 2.2, Panel C). In contrast with previous studies [117,118], no significant differences in percentages of T regulatory cells (defined as CD3+, CD4+, FoxP3+, and CD127- was observed in Amish and Hutterite children (0.056±0.054% vs. 0.079±0.081% of peripheral-blood leukocytes, P=0.29).

2.4.4 Cytokine responses to innate and adaptive stimulation

Cytokine levels were measured in supernatants from peripheral-blood leukocytes that were cultivated for 30 hours, with or without immune stimuli (lipopolysaccharide) or adaptive stimuli (combined anto-CD3 and anti-CD28 antibodies). Twenty-three cytokines were detectable in the supernatants from peripheral-blood leukocytes treated with lipopolysaccharide. Median levels of each of these 23 cytokines were lower in the Amish children than in the Hutterite children, and these distributions were significantly different (P<0.001 by Wilcoxon signed-rank test) (see Supplementary Table 2.5 and Supplementary Table 2.6). Results were similar after the exclusion of children who were known to have asthma or allergies (Supplementary Table 2.7). In contrast, after adaptive stimulation the overall distributions of median cytokine levels were not significantly different in peripheral blood leukocytes from Amish and Hutterite children (P=0.08 by Wilcoxon signed-rank test) (Supplementary Table 2.8).

Figure 2.2. Proportions of peripheral-blood leukocytes and cell-surface marker phenotypes in Amish and Hutterite children

The percentages of total peripheral-blood leukocytes (Panel A) were determined with flow cytometry for neutrophils (defined as CD66b+ Siglec-8+), eosinophils (defined as CCR3+Siglec-8+), and monocytes (defined as CD14+CD66b-). Box-and-whisker plots show a line indicating median value, with the box showing the interquartile range and whiskers representing the 95% confidence interval. Neutrophils (Panel B) were characterized according to the surface expression of CXCR4, CD11b, and CD11c (shown here), along with CXCR1 and CXCR2, expressed as mean fluorescence intensity (MFI). The expression of the interleukin-8 coreceptors CXCR1 and CXCR2 was not significantly different between groups (P=0.26 and P=0.91, respectively). Monocytes (Panel C) were characterized for the surface expression of HLA-DR and immunoglobin-like transcripts (ILTs), including ILT3 (shown here). There was no significant difference in the MFI of inhibitory receptors ILT2 and ILT4 between Amish and Hutterite children (P=0.69 and P-0.21, respectively; data not shown), whereas the surface expression of ILT5 was increased on Amish monocytes (P=0.001; data not shown). All P values were calculated with the use of an unpaired Student's t-test. Cell proportions and phenotypes after the exclusion of children with asthma or allergic sensitization are shown in Supplementary Table 2.4. Reproduced with permission from: Stein MM, Hrusch CL, and Gozdz J., *et al.* Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* **375**, 411-421, doi:10.1056/NEJMoa1508749 (2016). Copyright Massachusetts Medical Society.

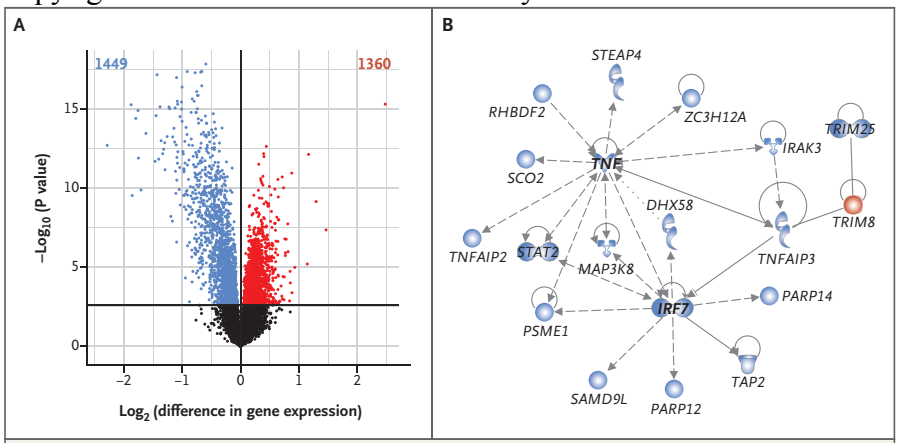


2.4.5 Gene-expression profiles

The striking differences in the proportions of peripheral-blood leukocytes observed in Amish and Hutterite children were reflected in the gene-expression profiles of these cells (Supplementary Figure 2.5). At a false discovery rate of 1%, 1449 genes were up-regulated in the peripheral-blood leukocytes of Amish children (blue points in Figure 2.3, Panel A) as compared with 1360 genes up-regulated in the cells of Hutterite children (red points in Figure 2.3, Panel A). These differentially expressed genes were organized into 15 coexpression modules with the use of the Whole Genome Co-Expression Network Analysis (Supplementary Table 2.9). To better understand the biologic relationships within each set of genes in each module, we used Ingenuity Pathway Analysis (Qiagen) to construct unsupervised networks on the basis of prior knowledge of the physical and functional connections between the molecules encoded by the genes. The most significant network ($P=1.0\times10^{-30}$ by Fisher's exact test) was in a module that contained 43 genes. The module was associated with both Amish and Hutterite status ($P=7.1\times10^{-9}$) and was therefore also associated with the proportions of neutrophils ($P=1.5\times10^{-6}$) and eosinophils ($P=1.0\times10^{-3}$). Eighteen of the genes in this module were over-expressed in the Amish peripheral-blood leukocytes, and all were clustered in a network that had as hubs tumor necrosis factor (TNF) and interferon regulatory factor 7 (IRF7), two key proteins in the innate immune response to microbial stimuli (Figure 2.3, Panel B).

Figure 2.3. Gene-expression profiles in peripheral-blood leukocytes from Amish and Hutterite children

In Panel A, a volcano plot shows differences in baseline gene expression in peripheral blood leukocytes from Amish and Hutterite children. The x axis indicates the $\log_2$ differences in gene-expression level between groups, with larger positive values representing genes with higher expression in the Hutterites relative to the Amish (1360 genes, shown in red points) and larger negative values representing genes with higher expression in the Amish relative to the Hutterites (1449 genes, shown in blue points). The y axis shows the $-\log_{10}$ of the P values for each gene, with larger values indicating greater statistical significance. The solid horizontal line indicates the 1% false discovery rate. Black points represent genes from Amish and Hutterite cells for which there was no significant difference in gene expression. Differences in gene expression remain after the data for children with asthma or allergic sensitization were excluded (Supplementary Figure 2.6 and 2.7). Changes in gene expression between the two groups after correcting for differences in cell proportion are shown in Supplementary Figure 2.4. In Panel B, a network of differentially expressed genes in untreated peripheral-blood leukocytes is shown. Genes shown in blue have increased expression in Amish children, and the gene shown in red has increased expression in Hutterite children. The gene shapes indicate the class of each gene's protein product (spirals denote enzymes, a v-shape denotes cytokines, conjoined circles denote a transcription regulators, hollow upside-down triangles denote kinases, cups denote transporters, and circles denote other products). Lines represent different biologic relationships (solid lines indicate direct interaction, arrows direction of activation, arrows with a horizontal line direction of activation and inhibition, and lines without arrows binding only. Reproduced with permission from: Stein MM, Hrusch CL, and Gozdz J., *et al.* Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* **375**, 411-421, doi:10.1056/NEJMoa1508749 (2016). Copyright Massachusetts Medical Society.

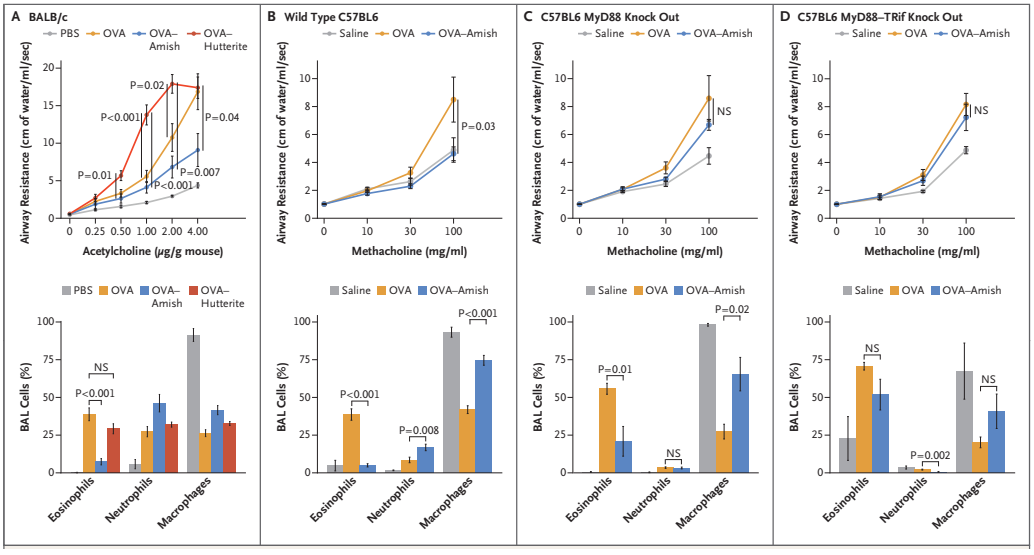


2.4.6 Effects of house-dust extracts on experimental asthma

To create a framework that would help us to interpret our observations, which were pointing toward a protective role of innate immunity, we used a classic ovalbumin mouse model of allergic asthma, comparing the effects of house dust obtained from Amish and Hutterite homes by administering extracts intranasally to mice over the course of 4 to 5 weeks. Eosinophilia was observed in bronchoalveolar-lavage samples, and airway hyperresponsiveness was exacerbated in mice treated with ovalbumin and Hutterite dust extracts as compared with mice treated with ovalbumin alone, findings that were consistent with the absence of protection from asthma observed in Hutterite children (Figure 2.4, Panel A). In contrast, inhalation of Amish dust extracts was sufficient to significantly inhibit ovalbumin-induced airway hyperresponsiveness, eosinophilia in the bronchoalveolar lavage, and the elevation of serum ovalbumin-specific IgE levels (Figure 2.4, Panel A, and Supplementary Table 2.10). Levels of lung T regulatory cells (defined as CD3+, CD4+, and FoxP3+) were not increased (Supplementary Table 2.11). The inhibitory effects of these extracts in wild-type mice probably required innate immunity, because protection was strongly reduced in mice deficient in MyD88 (Figure 2.4, Panel C) and completely abrogated in mice deficient in both MyD88 and Trif (Figure 2.4, Panel D), two molecules that are critical to the development of multiple innate immune-signaling pathways.

Figure 2.4. Effects of Amish and Hutterite house-dust extracts on airway responses in mouse models of allergic asthma

Panel A shows the effect of the intranasal instillation of 50 µl of Amish or Hutterite dust extract in 7-week-old mice (BALB/c strain) every 2 to 3 days for a total of 14 times beginning at day 0. The mice were sensitized with ovalbumin (OVA) intraperitoneally on days 0 and 14 and challenged with ovalbumin intranasally on days 28 and 38. Airway resistance (shown as centimeters of water per milliliter per second and stimulated in response to increasing doses of acetylcholine administered intravenously) and bronchoalveolar-lavage (BAL) cellularity were measured on day 39 (4 to 6 mice per group). The total amount of Amish and Hutterite dust extract administered over the course of the experiment represented the total load of airborne dust deposited on electrostatic dust collectors placed in Amish or Hutterite homes for 1 month. Statistical differences in experimental measures were assessed with the use of Student's t-test. Amish house-dust extracts (7.5 mg of dust equivalent in 50 µl) were instilled intranasally every 2 to 3 days for a total of 14 times beginning 5 days before day 0 into 7-week old wild-type mice (Panel B), mice deficient in MyD88 (Panel C), and mice deficient in MyD88 and Trif (Panel D) (all C57BL6 strains). These mice were sensitized intraperitoneally with 20 µg of ovalbumin on days 0 and 14 and were challenged intranasally with 75 µg of ovalbumin on days 26, 27, and 28. Airway resistance (shown as an increase from baseline in response to increasing doses of nebulized methacholine) and bronchoalveolar-lavage cellularity were measured on day 30 (12 mice per group for wild-type mice and 6 mice per group for those deficient in MyD88 or MyD88 and Trif). Statistical differences in experimental measures were assessed with the use of Student's t-test. I bars represent the standard errors of the data. NS denotes not significant and PBS phosphate-buffered saline. Reproduced with permission from: Stein MM, Hrusch CL, and Gozdz J., *et al.* Innate immunity and asthma risk in Amish and Hutterite farm children. *N Engl J Med* **375**, 411-421, doi:10.1056/NEJMoa1508749 (2016). Copyright Massachusetts Medical Society.



2.5 Discussion

Our studies in Amish and Hutterite schoolchildren revealed marked differences in the prevalence of asthma despite similar gene ancestries and lifestyles. As compared with the Hutterites, the Amish, who practice traditional farming and are exposed to an environment rich in microbes, showed exceedingly low rates of asthma and distinct immune profiles that suggest profound effects on innate immunity. Data generation in an experimental model of asthma support this notion by showing that the protective effect of the Amish environment requires the activation of innate immune signaling.

Analyses of the proportions and gene-expression profiles of peripheral-blood immune cells in Amish and Hutterite children revealed differences in the cells and genes involved in innate immune responses to microbes. Indeed, neutrophils, eosinophils, and monocytes appeared to be major targets of the distinct environments to which Amish and Hutterite children are exposed because these cell types differed between the two groups in terms of their relative abundance, their phenotypes, or both. Moreover, the network most associated with these differences consisted of innate immune genes. Notable among the genes that were more highly expressed in the Amish children was *TNFAIP3*, which encodes A20, a ubiquitin-editing enzyme that limits the activity of multiple inflammatory pathways that depend on nuclear factor κB (NF-κB)[119] and that has also been shown to mediate the protective effects of European farm-dust extracts in murine models of allergic asthma[120]. *IRF7*, a hub in this network, regulates type I interferon transcription and is therefore essential for innate airway responses against viruses[121] that are linked to susceptibility to asthma[69,122]. In turn, *TRIM8*, the one gene in the network that was more highly expressed in the Hutterites, acts as a positive regulator of TNF-α– and interleukin-1β–induced activation of NF-κB[123]. These findings suggest that in the Amish, intense

and presumably sustained exposure to microbes activates innate pathways that shape and calibrate downstream immune responses.

Sustained microbial exposure was also reflected in the phenotypes of peripheral innate immune cells in the Amish. Repeated microbial stimulation can lead to reduced expression of HLA-DR on monocytes[124,125] and drive immature neutrophils from the bone marrow[126-129]. Indeed, Amish children had immature neutrophils bearing markers suggestive of recent emigration from the bone marrow, and they had monocytes with reduced expression of HLA-DR and increased expression of ILT3, all of which are suggestive of anti-inflammatory function. Proportions of T regulatory cells and levels of interleukin-10, which typically mediate immune – balancing effects, were not increased in the Amish children. However, qualitative and functional differences in regulatory-cell populations remain to be defined.

Innate immunity has evolved to sense the environment and transduce signals that calibrate adaptive responses to exogenous antigens. The proteins MyD88 and Trif are located at the convergence of multiple innate signaling pathways,[130] and deletion of these molecules virtually disables innate immune responses, thereby also dysregulating adaptive immunity. The fact that the loss of protection was more marked in mice deficient in both MyD88 and Trif than in mice deficient only in MyD88 points to the involvement of multiple innate pathways. The concordance between findings from studies in humans and mice was remarkable: in both studies protection was accompanied by lower levels of eosinophils, higher levels of neutrophils, generally suppressed cytokine responses, and no increase in levels of T regulatory cells or interleukin-10. Thus, the finding that these features were largely dependent on innate immune signaling may also be the primary target of protection in the Amish children, in whom downstream adaptive immune responses may also be modulated.

Our study has several limitations. First, we were unable to include children younger than 6 years of age, we collected samples at a single time point, and the numbers of Amish and Hutterite children in our study were relatively small. As a result, we may have missed important windows of immune development or lacked the ability to detect early, subtle shifts in cell composition, response, or phenotype that are critical for immune maturation. Second, our microbiome assessments were limited, since only pooled dust samples from a limited number of homes were available for the studies in which were assessed bacterial composition. Therefore, we cannot further dissect microbial composition and identify potentially protective microbes to target. However, the striking differences found in endotoxin levels support the notion that the Amish indoor environment is much richer in microbial exposures than the Hutterite environment. Third, the strategy used for sampling the Hutterite children enriched selection for those with asthma, although the prevalence of asthma in our sample was similar to that reported in previous population-based studies[107]. Moreover, the exclusion of children with asthma or allergic sensitization from our analyses of gene expression, cell composition, and immune phenotypes did not affect the outcomes. Our study in a small number of children was sufficient to show significant differences in the prevalence of asthma and in immune profiles, suggesting that very strong environmental factors must account for these differences. Indeed, we showed that there are remarkable genetic similarities between Amish and Hutterite children. Although we interrogated only common variants, other variants that occur at very low frequency in these populations are unlikely to account for the observed large differences in the prevalence of asthma. In the end, the novelty of our work lies in the identification of innate immunity as the primary target of the protective Amish environment, a finding supported by results obtained in both humans and mice. Conversely, our work suggests that susceptibility to asthma may be

increased when innate immune stimulation is weak. A deeper understanding of the relevant stimuli and the innate immune pathways they engage may ultimately pave the way for the development of effective strategies for the prevention of asthma.

2.6 Acknowledgements

I would like to thank the tireless effort of the two other co-first authors on this study, Justyna Gozdz and Cara Hrusch. Without their work, this paper would not have been possible. We also thank the Hutterite and Amish volunteers and their families for participating in this study; Gorka Alkorta-Aranburu, Maitane Arrubarrena Orbegozo, Kathleen Bailey, Christine Billstrand, Kelly Blaine, Daniel Cook, Donna Decker, Mohammad Jaffery, Courtney Burrows, Katherine Naughton, Raluca Nicolae, Rob Stanaker, Meghan Sullivan, and Emma Thompson for assistance on field trips and with sample processing; Peace Ezeh, Amanda Herrell, Ashley Horner, Kenneth Addision, Dominik Schenten, and Shane Snyder for their contributions to the studies in mice; and Minal Çalışkan, Yoav Gilad, Jessie Nicodemus-Johnson, John Novembre, and Matthew Stephens for helpful comments and statistical advice.

2.7 Appendix A: Supplementary Methods

2.7.1 Study participants

Thirty Amish schoolchildren (7-14 years old) volunteers had blood sampled in Middlebury, Indiana in November 2012. Fifty-one Hutterite schoolchildren volunteers had blood drawn at two Hutterite colonies near Mitchell, South Dakota, in December 2012. Thirty Hutterite children were selected from this larger sample to be age- (±1 year) and sex-matched to the 30 Amish children, and to represent similar numbers of sibships (families) as for the Amish. When there

was more than one Hutterite child who could serve as a match to an Amish child, we selected a child with asthma, when possible. This occurred in two cases. The prevalence of asthma and atopy in individuals with known disease status was 14% (7/51) and 63% (26/41), respectively, among all schoolchildren in the two Hutterite colonies and 20% and 60%, respectively, among the 30 children included in these studies.

Written consent was obtained from the parents and written assent was obtained from the children. Parents of participants responded to a short questionnaire on asthma symptoms and previous diagnoses. The study was approved by the institutional review boards at the University of Chicago and St. Vincent Hospital, Indianapolis.

2.7.2 Collection of blood samples

One mL of whole blood was drawn into three TruCulture® (Myriad RBM) tubes: one containing TruCulture® media + 0.1 µg/mL LPS, one containing TruCulture® media + 0.4 µg/mL anti-CD3 + 0.33 µg/mL anti-CD28, and one containing TruCulture® media alone. Samples were incubated upright in a dry heat block at 37°C for 30 hours. Supernatant and cells were isolated and frozen for cytokine and gene expression studies, respectively. An additional three tubes of blood (5 mL each) were drawn and shipped to the University of Chicago: (1) heparinized vacutainer (BD Biosciences) shipped at room temperature for PBL isolation; (2) EDTA vacutainer (BD Biosciences) shipped on wet ice for DNA studies; and (3) silicon-coated serum tube (BD Biosciences) shipped on wet ice for IgE studies. The TruCulture® samples from Amish children were processed after 30 hours at the University of Chicago, and the samples from Hutterite children were processed after 30 hours on site in a makeshift lab in South Dakota. The Hutterite

and Amish samples were processed by the same individuals and using the same equipment at both sites.

2.7.3 Measurement of specific and total human serum IgE

Total serum IgE and serum IgE against six common airborne allergens were measured at the University of Chicago Hospital laboratory. The allergens were: *D. pteronyssinus* (house dust mite), *Alternaria alternata* (mold), cat dander, *Blattella germanica* (German cockroach), a tree mix (red maple, birch, hazelnut, white oak, and sycamore) and a grass mix (orchard grass, meadow fescue, rye grass, Timothy grass, and Kentucky bluegrass). Total and specific IgE were measured with separate Immunocap 250 assays (Thermo Scientific).

2.7.4 Genetic studies

Eighteen of the Hutterite children had been genotyped on the Affymetrix 6.0 chip as part of previous studies [131]; an additional nearly 7 million genotypes were imputed to those children from whole genome sequence data obtained in 98 Hutterites using PRIMAL, an in-house method which has 99.7% imputation accuracy [132]. DNA from the 30 Amish schoolchildren and three Hutterite schoolchildren were genotyped with the Illumina CytoSNP at the Genomics Core at the University of Chicago (Supplementary Figure 2.1). The remaining nine Hutterite schoolchildren were genotyped with the Illumina OmniExpress Bead Chip at the same core. SNPs from both Illumina chips were merged, and the overlapping Illumina SNPs were retained if they had a MAF $\geq$5% and genotypes were called in all 42 children. To avoid strand errors in matching CG and AT SNPs across platforms, only CT/GA SNPs were included, yielding genotypes at 123,060 SNPs in the 30 Amish and 12 Hutterite children. We then extracted genotypes at those 123,060

SNPs for the remaining 18 Hutterite children who were previously genotyped. This resulted in a final set of 118,789 SNPs for which genotypes were available in all 60 children for analyses of ancestry. We then found the overlap (73,590 SNPs) between this set of SNPs and genotypes from the European populations included in the Human Genome Diversity Project[114]. Removing SNPs with a MAF ≤5% and call rate ≤99.5% resulted in a final data set of 72,034 SNPs. We performed Principal Components Analysis (PCA) using the function *prcomp* in R on the HGDP European populations and projected the Hutterite and Amish genotypes onto the sample space.

2.7.5 Dust collections in Amish and Hutterite homes

At the time of whole blood sampling in 2012, electrostatic dust collectors (EDCs) were placed in two different locations (bedroom and living room) in each of 10 Amish and 10 Hutterite homes to collect airborne house dust. All 10 Amish and 9 of 10 Hutterite homes housed children who participated in the study. After one month, EDCs were shipped to the University of Iowa where dust was extracted for mouse studies and analyzed for endotoxin (Limulus Amoebocyte Lysate Assay) and common allergens (MARIA, MRA-P8): house dust mite (*Der p*1, *Der f*1, mite group 2), cat (*Fel d*1), dog (*Can f*1), and German cockroach (*Bla g*2) using a fluorescent multiplex array kit (MARIA, MRA-P8) and previously described methods[133].

In 2014, vacuumed mattress dust was collected from Amish and Hutterite homes. We vacuumed Amish and Hutterite beds in homes without pets and without rubber or plastic protective sheets covering the mattress. The bottom half of each mattress was vacuumed (3-5 min) in long parallel motions using the same battery powered hand vacuum (Black and Decker CHV1510L 15.6 Volt Cyclonic Action Cordless Dustbuster) in each population. DNA was extracted from an aliquot of each pooled dust sample using FastDNA Spin Kit for Soils (MP

Biomedicals) and concentrated with DNA Clean and Concentrator (Zymo Research). For each of the two pooled samples, libraries for the V4 16S rRNA V4 region were prepared and sequenced on the Illumina MiSeq using a 151bp paired end protocol at Argonne National Laboratory[134].

Data was pre-processed using CASAVA 1.8.1 and downstream analyses utilized QIIME v1.8.0 software[135]. Paired sequence reads were filtered for high quality using the following criteria. Reads were required to have 1) an exact match to an expected barcode, 2) zero ambiguous base calls, 3) less than three consecutive low quality base calls, and 4) a minimum Phred quality score of 20 along the entire read. All reads were clustered *de novo* using QIIME script pick_otus.py with 97% identity and each dust sample was randomly subsampled to 645,000 reads to reduce variability in sequence depth. To assign taxonomy, a representative sequence from each of the 79,323 clusters was classified with using QIIME script assign_taxonomy.py against the Greengenes 13.8 database. An operational taxonomic unit (OTU) table was constructed from these classified clusters and rarefied relative abundance was calculated at each taxonomic level and is presented at the family level.

Dust was also collected in 2015 by vacuuming the floor of 5 Amish homes. Individual samples were pooled, and dust was extracted for mouse studies.

2.7.6 Flow cytometry

For flow cytometry studies, PBLs were isolated from whole blood collected in a heparinized vacutainer (BD Bioscience). Samples were kept at room temperature and processed approximately 30 hours after collection by lysing red blood cells in a solution of 0.15 M ammonium chloride, 10 mM potassium bicarbonate, and 0.1 mM EDTA. Isolated PBLs were cryopreserved in freeze media (Invitrogen) and stored in liquid nitrogen until analysis. Prior to

the flow studies, frozen PBLs were thawed, washed in RPMI containing Deoxyribonuclease I (0.02 mg/mL), and resuspended in FACS buffer (PBS containing 0.1% sodium azide and 1% BSA). Approximately $3x10^5$ cells in 100 µL per sample were incubated for 10 min with pooled human IgG (FcX, Biolegend) to block non-specific antibody binding before staining with fluorescently conjugated antibodies (Supplementary Table 2.1). For surface phenotyping, flow cytometry data were acquired immediately after staining on an LSRFortessa (BD Biosciences, San Jose, CA), and the data were analyzed with FlowJo software (Tree Star, Inc.). For cell composition studies, cell types were identified as follows: eosinophils (Siglec-8^+CCR3^+), neutrophils (CD16^+CD$66b^+$Siglec-8^-), monocytes (CD14^+CD$66b^-$), and T regulatory cells (CD3^+CD4^+FoxP3^+CD127^-). For FoxP3 staining, cells were surface stained as described above before performing the FoxP3 staining according to manufacturer's instructions (FoxP3 Fix/Perm Kit, eBioscience).

2.7.7 Gene expression profiling

RNA was extracted from each sample using AllPrep DNA/RNA Mini Kits (Qiagen). RNA concentration was assayed with a Nanodrop ND-100 Spectrophotometer (NanoDrop Technologies); RNA quality was assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies). RNA samples with RIN (RNA integrity) scores >6.5 were submitted to the Functional Genomics Core at the University of Chicago for gene expression profiling. Samples underwent cDNA synthesis and were then hybridized on the HumanHT-12 v4 Expression BeadChip arrays (Illumina, Inc). Each array included samples from 6 Hutterite and 6 Amish children (matched for age and sex), and samples were further randomized between arrays by sex, asthma status, and extraction batch to minimize confounding with chip effect. The arrays were

scanned and underwent image processing, and probe intensity data was returned for gene expression analysis. One Amish sample failed cDNA synthesis and hybridization and was excluded from further analyses.

Probe intensity was $\log_2$-transformed and quantile normalized using the *lumi* package in R[136]. Probes that overlapped >1 HapMap SNPs were removed, as were probes that did not have a detection *P* value <0.05 in at least 20% (n=12) of samples, resulting in a final set of 14,119 probes. To estimate gene expression, we used the most 3' probe in genes with multiple probes. We used PCA to identify and remove confounding technical covariates. RNA integrity number and RNA concentration were regressed out using linear regression, and chip effect was removed using ComBat[137]. To correct for differences in innate cell proportions, neutrophil, monocyte, and eosinophil cell proportion was regressed out of the linear gene expression model.

We used the R package *Weighted Gene Co-expression Network Analysis* (WGCNA) to identify co-expression modules within differentially expressed genes, using standard settings[138]. The genes in each module were analyzed further with Ingenuity Pathway Analysis® (IPA, QIAGEN) to identify networks enriched within each module using the IPA Knowledge Base. Within IPA, the significance of each network is determined using a Fisher's exact test, and a network score is calculated from the –log10(*P*-value). The 43 genes in module 1 were organized using the Connect tool in Qiagen's Ingenuity Pathway Analysis® (IPA, QIAGEN, Redwood City) Pathway Builder, to display a network that only includes genes that are a part of the module.

2.7.8 Measurement of cytokines

The Milliplex Map Human TH17 Magnetic Bead Panel (EMD Millipore) was used to measure levels of 25 cytokines in the supernatants of LPS-treated and anti-CD3 + anti-CD28-treated cells in the Hutterite and Amish children. Assays were performed at the University of Chicago Flow Cytometry Core facility, using standard protocols. Median fluorescent intensity (MFI) was measured using a Luminex Bio-Plex plate reader (Bio-Rad Laboratories, Inc.). Samples were run in duplicate and averaged to calculate the mean cytokine level for each individual. In cases where one of the duplicates failed, the remaining measurement was used as the cytokine level for that individual. Samples with concentrations above or below the range of detection were substituted with the highest or lowest recorded concentration for that cytokine, respectively. Only cytokines with >40% of all measurements falling within the standard curve were included in the analyses. As a result, one cytokine (TNFβ) was not included in the analyses. In addition, two cytokines (IL-21 and IL-17F) were not further studied due to technical failure.

Because measurements of two cytokines (IL-6, IL-1β) were higher than the maximum level of detection on the cytokine panel, and IL-8 was not included in the panel, we measured these cytokines in the supernatant of LPS-treated samples by ELISA, following manufacturer's instructions (eBioscience). Supernatants were diluted 1:50 (IL-1β) or 1:200 (IL-6, IL-8).

2.7.9 Preparation of dust extracts

To obtain aqueous extracts of Amish and Hutterite indoor dust collected in 2012, electrostatic cloths from EDCs were extracted with sterile pyrogen-free water (10 ml) by shaking for 1 hour at 22°C. Supernatants were centrifuged at 600x*g* for 15 min and decanted. Dust extracts from EDCs deployed in Amish and Hutterite homes (n=10 each) were pooled separately into one

sample. SpeedVac centrifugation was then used to reduce each sample to the same volume (5-6 ml). The same procedure was followed to prepare aqueous extracts of floor dust collected from Amish homes in 2015, but dust extracts were standardized by weight and resuspended at 100 mg/ml of dust equivalent. Aqueous dust extracts were aliquoted and stored in cryovials at -80°C until used.

2.7.10 Experimental asthma models

Amish or Hutterite house dust extracts (50 µl) were instilled intra-nasally (i.n.) every 2-3 days (14 times total) beginning at day 0 into adult (7 weeks old) Balb/c mice (Harlan Laboratories) sensitized intra-peritoneally (i.p.) with OVA (grade V, Sigma: 20 µg in 200 µL of PBS)-Alum (Pierce) at day 0 and 14 and challenged with OVA i.n. (50 µg in 50 µL of PBS) at day 28 and 38. The total amount of Amish and Hutterite dust extracts administered over the course of the experiment represented the total load of airborne dust collected on EDCs placed in Amish or Hutterite homes over a month's period. In other experiments, Amish house dust extracts (7.5 mg of dust equivalent in 50 µl) were instilled i.n. every 2-3 days for a total of 14 times beginning at day -5 into 7-week old wild-type, Myd88-deficient[111] (Jackson Laboratories) and Myd88/Trif-deficient[112] C57BL6 mice that were sensitized i.p. with 20 µg OVA-Alum at day 0 and 14, and challenged i.n. with 75 µg OVA at day 26, 27 and 28. Airway resistance and BAL cellularity were measured at day 30. Terminal assessments at day 39 (Balb/c mice) and day 30 (C57BL6 mice) included airway resistance and BAL cellularity. For Balb/c mice, to measure airway resistance in response to increasing concentrations of intravenous acetylcholine (Ach: 0-4 µg/g mouse), the animals were anesthetized with pentobarbital (90 µg/g body weight). For C57BL6 mice, to measure airway resistance in response to increasing concentrations of nebulized

methacholine (0-100 mg/ml), the animals were anesthetized with ketamine and xylazine (100 and 10 mg/kg, respectively) and paralyzed with pancuronium bromide (4 μg/g). The trachea was then dissected free and cannulated with a 20 gauge cannula (BD) which was kept in place with a single tie suture. Mice were then connected to a small ventilator (FlexiVent, SCIREQ, Inc.) and ventilated with a tidal volume of 10 mL/kg, inspiratory/expiratory ratio of 66.67%, respiratory rate of 150/min and maximum pressure of 30 cm H_2O as described in ref.[139] . BAL was performed by delivering cold 1% BSA in PBS (2 mL) into the airway via a tracheal cannula and gently aspirating the fluid. Cells were counted using a Countess II FL automated cell counter (Thermo Fisher Scientific) and differentials were determined by an operator blinded to mouse ID/grouping after staining with Hema 3 (Fischer) and examining at least 400 cells/slide. Cytokines in BAL were measured by ELISA (Quantikine, R&D). OVA-specific serum IgE levels were measured as described in ref[140], but using 3% BSA as a blocking agent.

To assess proportions of $CD4^+Foxp3^+$ T regulatory cells (among live $CD3^+$) cells in the lungs of mice treated i.n. with Amish or Hutterite dust extracts (4 treatments over 10 days, 5 mg of dust equivalent/treatment), single cell suspensions were prepared from mouse lungs using Liberase TM (Roche Life Science) digestion. Cells (1×10^6) were stained as described for human PBLs using anti-mouse CD3-FITC, Foxp3-PE, CD4-PECy5, CD8-AF700 (all from BD Biosciences, San Jose, CA). Data were acquired on an LSR II flow cytometer (BD Biosciences) and analyzed with FlowJo software gating on live $CD3^+$ cells. All animal procedures conform to the principles set forth by the Animal Welfare Act and the National Institutes of Health guidelines for the care and use of laboratory animals in biomedical research and were approved by the University of Arizona Institutional Animal Care and Use Committee.

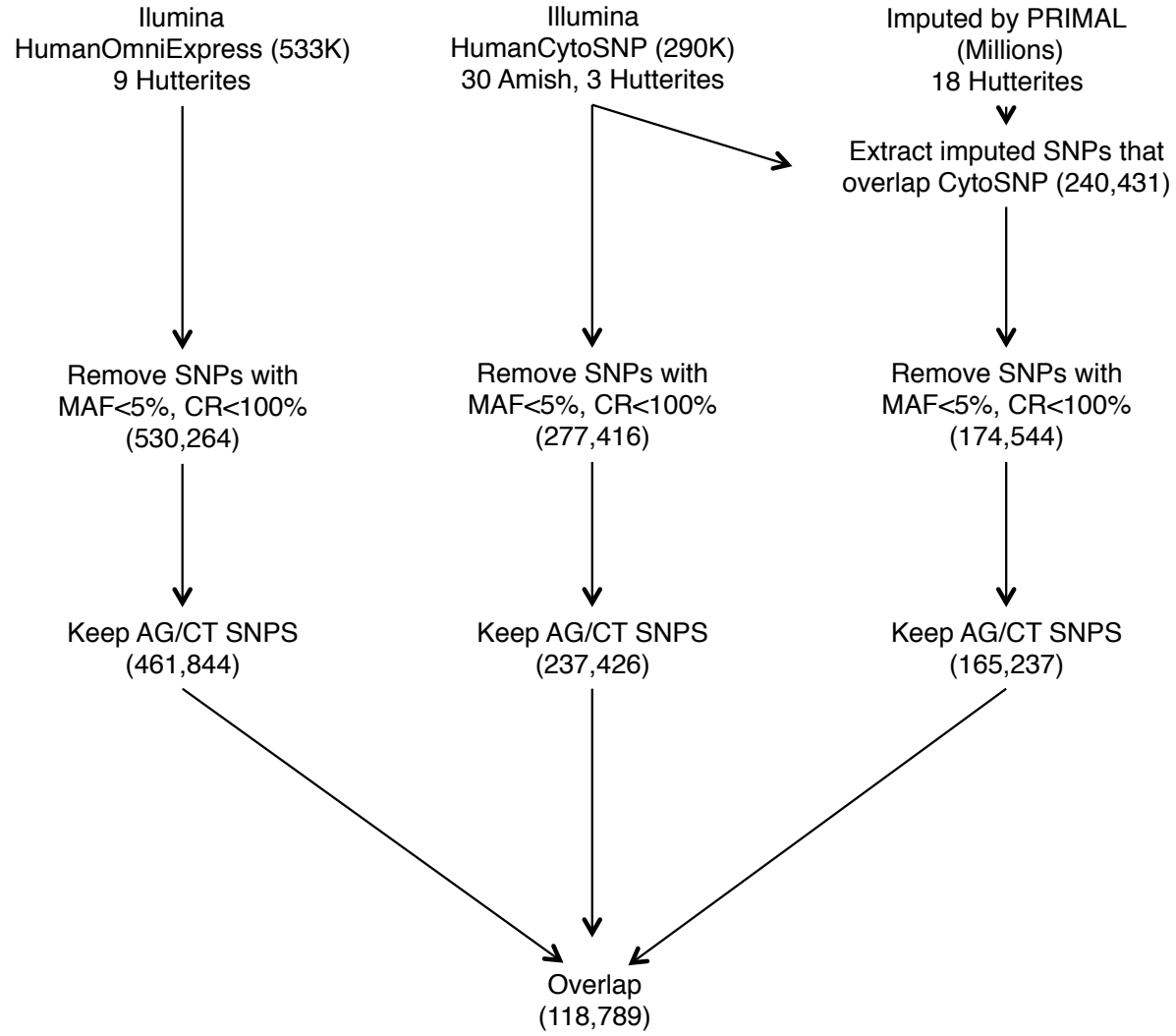
2.7.11 Statistical analyses

Statistical analyses were performed using R or Prism (GraphPad Software, Inc.). Differences in cytokine levels within LPS-treated or anti-CD3+anti-CD28-treated cells between Amish and Hutterite children were assessed using a Wilcoxon-rank sum test. Overall differences between Amish and Hutterite children across all cytokines was determined using a Wilcoxon signed rank test of median cytokine level. We used linear regression to identify differentially expressed genes between Amish and Hutterites in the untreated gene expression samples. Pearson correlation was used to correlate module eigengenes with cell proportions. The square root of the linear regression r^2 was used as an approximate correlation coefficient of module eigengene with Amish or Hutterite status. The false discovery rate methods of Benjamini and Hochberg[113,141] were used to control the false discovery rate. For flow cytometric and mouse studies, differences in cell populations, airway resistance, and OVA-specific IgE were assessed using an unpaired Student's t test. Differences in mouse T regulatory cells and cytokines were assessed with a Wilcoxon rank sum test.

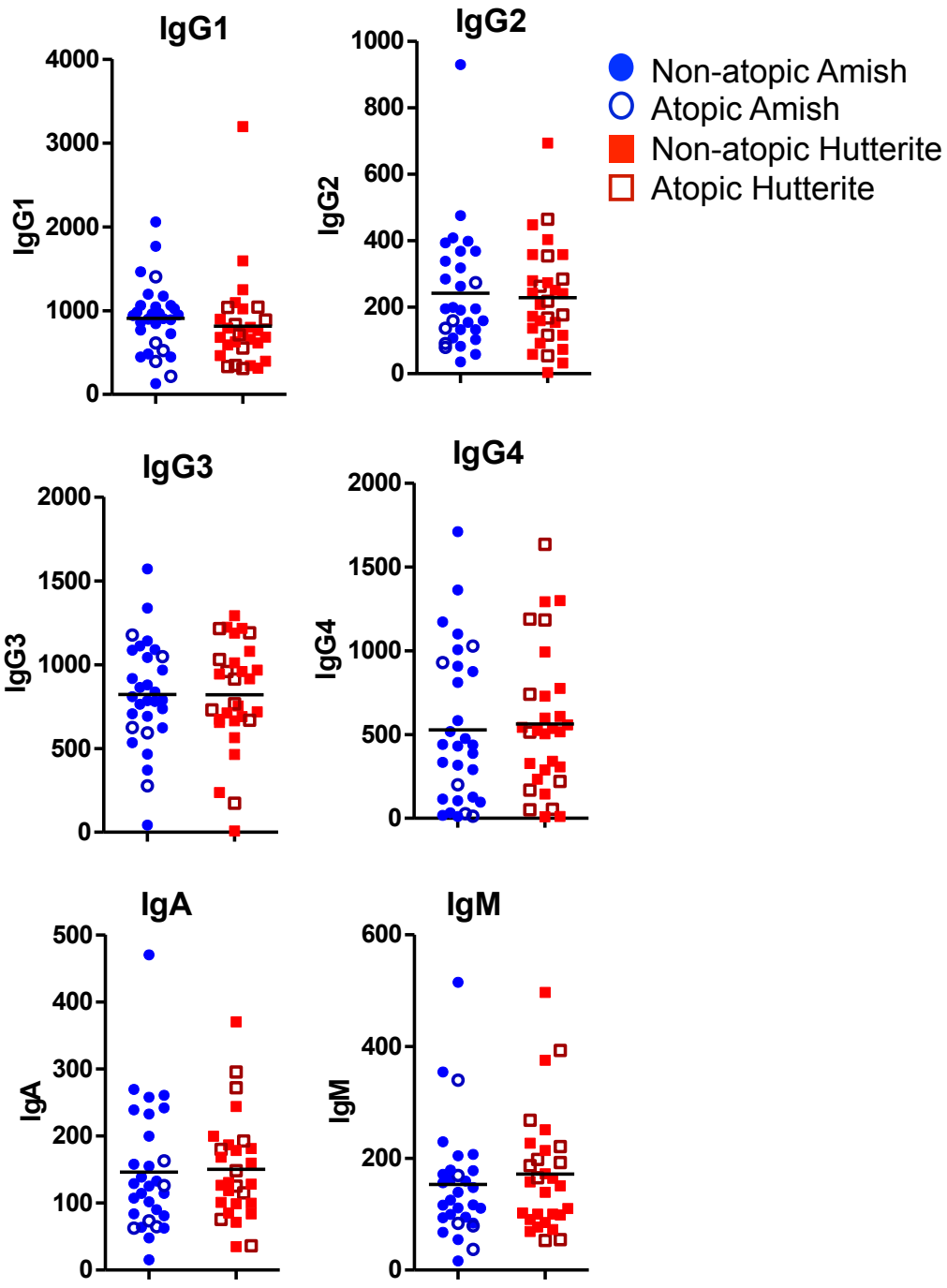
2.8 Appendix B: Supplementary Figures

Supplemental Figure 2.1 Schematic of genotype data QC.

Genotype data was generated from 3 different sources: (1) Illumina HumanOmniExpress chip for 9 Hutterites, (2) Illumina HumanCytoSNP chip for 30 Amish and 3 Hutterites, (3) 18 Hutterites with genotype data imputed from Hutterite whole genome sequence data (see Methods). MAF, minor allele frequency; CR; Call Rate. Numbers in parentheses indicate number of SNPs at each processing step.

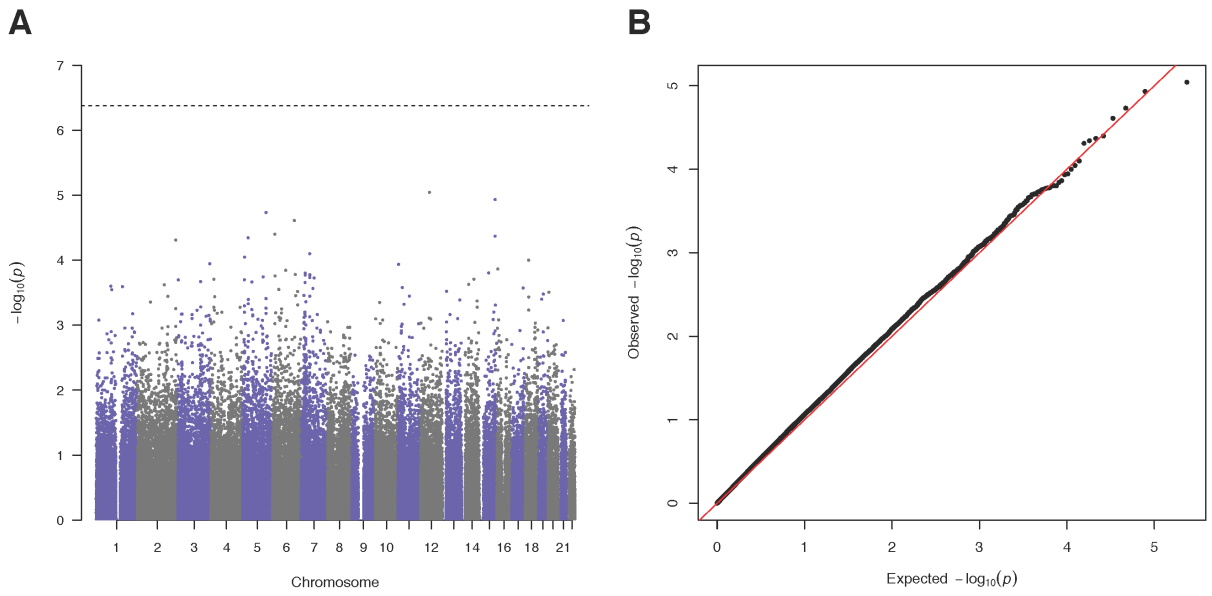


Supplemental Figure 2.2 Isotypes of non-IgE circulating antibody in Amish and Hutterite schoolchildren.



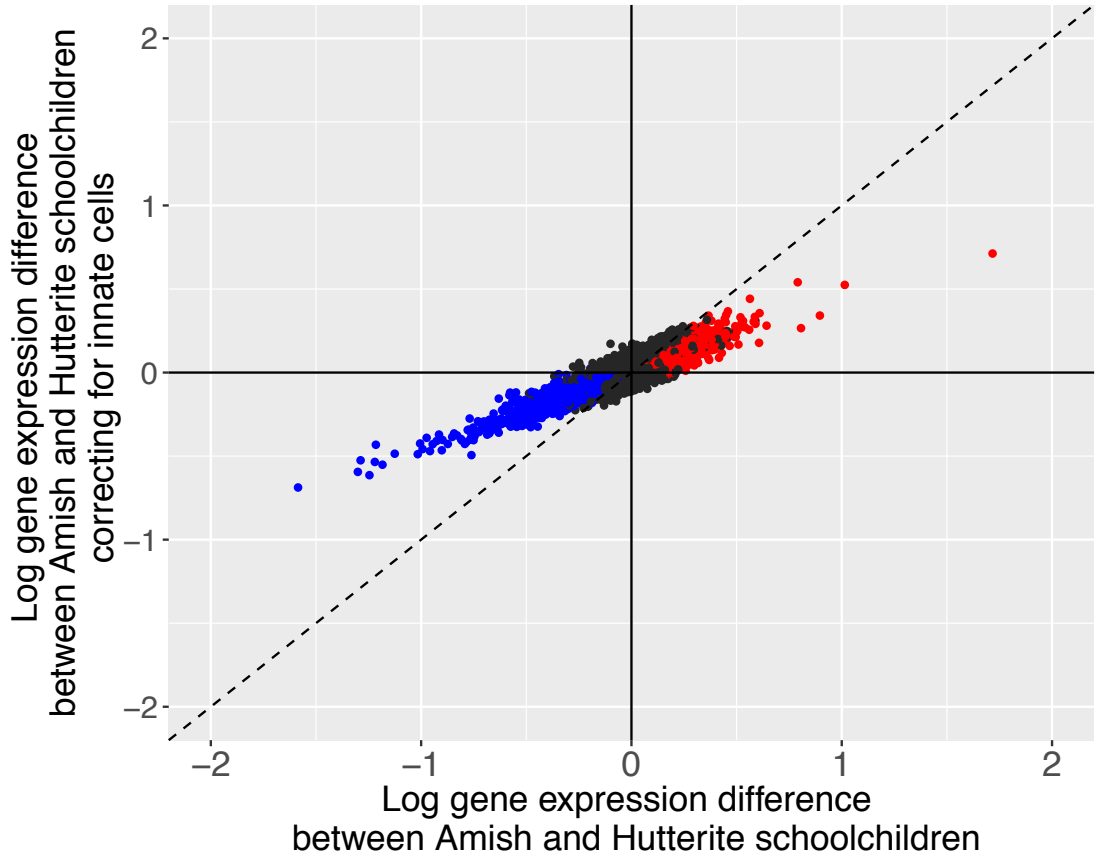
Supplemental Figure 2.3. Allele frequency differences between Amish and Hutterite children across 118,775 SNPs.

(A) Manhattan plot of log10 P values (y-axis) by chromosome location (x-axis). Dashed line indicates a Bonferroni-corrected significance threshold. (B) QQ plot of expected (x-axis) and observed (y-axis) log10 P values. A kinship matrix was included in the model as a random effect to correct for relatedness within each group.



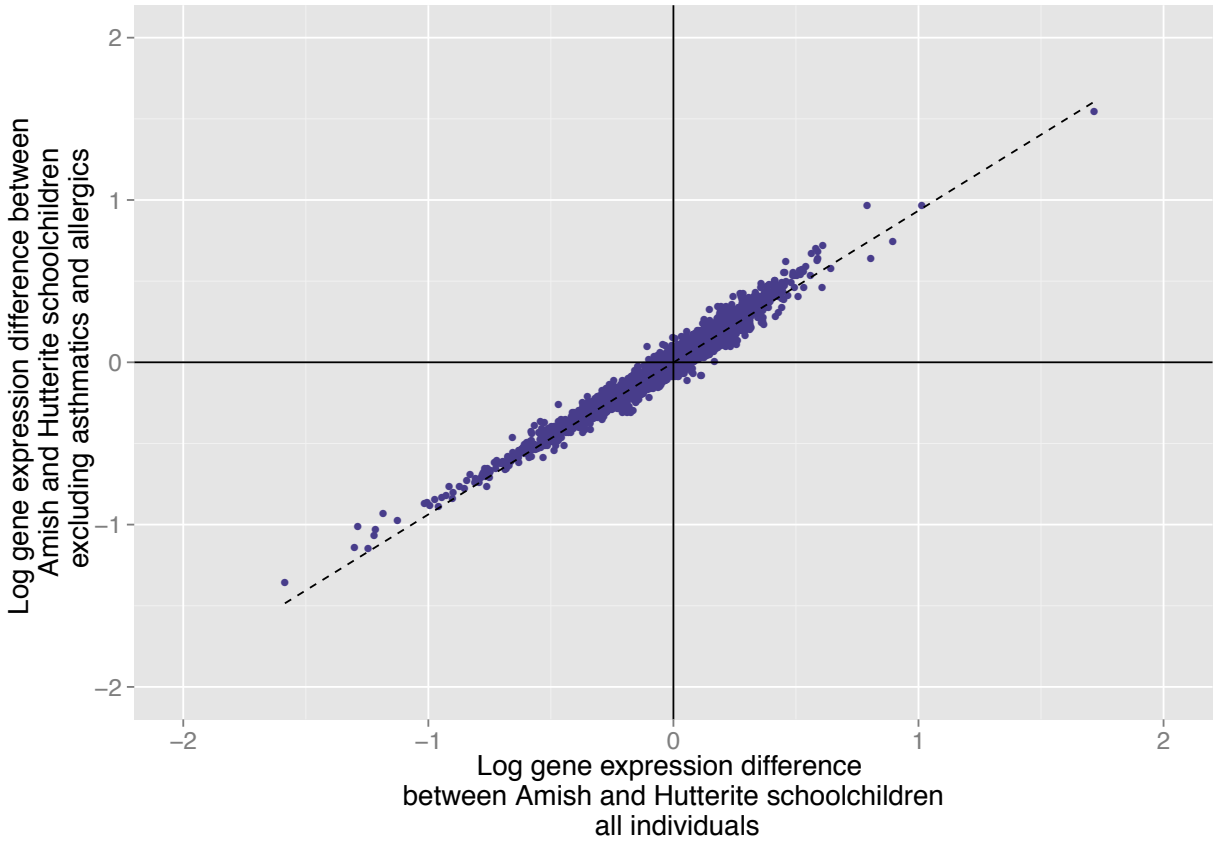
Supplemental Figure 2.4. Scatter plot of the log gene expression difference between Amish and Hutterite schoolchildren in untreated samples after correction for innate cell proportion differences.

The x-axis represents the log gene expression difference between Amish and Hutterite schoolchildren, and the y-axis represents the log gene expression difference between Amish and Hutterite schoolchildren after correcting for differences in innate cell proportion. Genes with red points indicate genes with increased expression (FDR 1%) in Hutterite PBLs (Figure 2.3, Panel A). Genes with blue points represent genes with significantly increased expression (FDR 1%) in Amish schoolchildren (Figure 2.3, Panel A). Dashed line indicates y=x. Overall, differences are reduced after correction for differences in innate immune cell proportions.



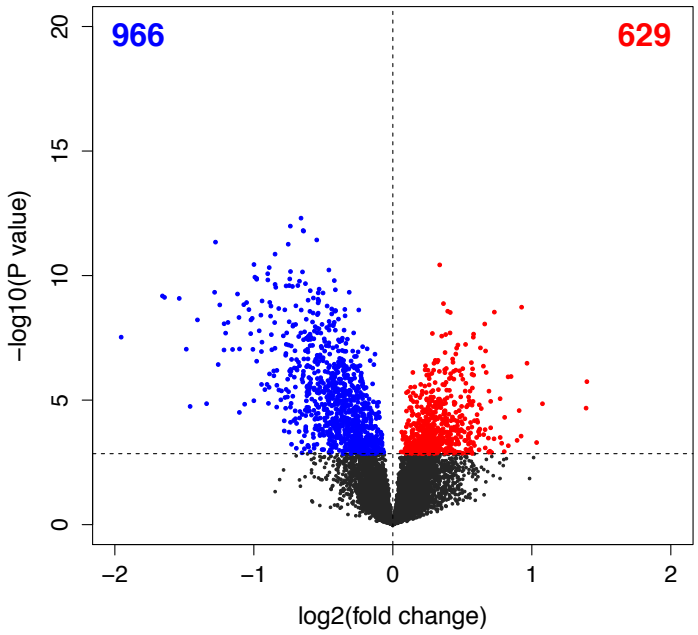
Supplemental Figure 2.5. Scatter plot of log gene expression difference between Hutterite and Amish schoolchildren in untreated samples.

The x-axis represents the log gene expression difference between all Hutterite and all Amish schoolchildren included in the study, and the y-axis represents the log gene expression difference between Hutterite and Amish schoolchildren after removal of individuals with asthma and/or allergy (Spearman Rank correlation ρ=0.96, P = < 2.2×10^{-16}). Positive values indicate genes with increased expression in Hutterite schoolchildren, while negative values indicate genes with increased expression in the Amish schoolchildren. Dashed line indicates y=x.



Supplemental Figure 2.6. Volcano plot of gene expression in untreated PBLs in Amish and Hutterite schoolchildren after excluding children with asthma and allergic sensitization.

After exclusion of these children and QC, 18 Hutterites and 23 Amish samples remained for analysis. The decreased number of differentially expressed genes (n=1,595, FDR 1%) relative to the full sample is due to this decreased sample size. Genes in blue are significantly increased (at FDR 1%) in Amish PBLs and genes in red are significantly increased (at FDR 1%) in Hutterite PBLs.



2.9 Appendix C: Supplemental Tables

Supplemental Table 2.1. Antibodies used in flow cytometric studies

Antibody	Clone	Fluorophore	Company
HLA-DR	L243	Brilliant Violet 510	Biolegend
CD3	OKT3	Qdot605NC	eBioscience
CD14	61D3	Qdot605NC	eBioscience
Siglec-8	837535	FITC	R&D Systems
CD4	OKT4	PerCP/Cy5.5	Biolegend
ILT3	ZM4.1	APC	eBioscience
FoxP3	206D	AlexaFluor647	Biolegend
CCR3	5E8	PE	Biolegend
CD66b	G10F5	PE/Cy7	Biolegend
CD127	A019D5	FITC	Biolegend

Supplemental Table 2.2. Mean level of specific IgE across six common allergens.

Differences in mean specific IgE were assessed with a Wilcoxon rank sum test.

Allergen	Amish mean specific IgE (number children with specific IgE>0)	Hutterite mean specific IgE (number children with specific IgE>0)	*P* value
D. pteronyssinus	0.11 (3)	2.97 (7)	0.09
Cat dander	0.01 (1)	0.06 (3)	0.36
German cockroach	0.35 (10)	0.14 (8)	0.49
Alternaria alternata	0.01 (2)	6.53 (6)	0.06
Mixed grass	3.62 (7)	0.29 (9)	0.56
Mixed trees	0.29 (7)	0.34 (4)	0.48

Supplemental Table 2.3. Measurement of allergens in dust collected from electrostatic dust collectors (EDCs) deployed in 10 Amish and 10 Hutterite homes.

HDM,house dust mite. Overall, allergens were detected in 4 Amish homes and 1 Hutterite home.

Allergen	Mean allergen (ng/m^2) Amish homes	Number Amish homes with detectable allergen	Mean allergen (ng/m^2) Hutterite homes	Number Hutterite homes with detectable allergen
*Der p*1 (HDM)	97.6	2	NA	0
*Der f*1 (HDM)	131.76	1	NA	0
Mite gp2 (HDM)	29.28	1	NA	0
*Fe*l *d*1 (Cat)	22.77	2	9.76	1
*Can f*1 (Dog)	48.80	1	31.72	1
*Bla g*2 (German cockroach)	795.44	1	NA	0

Supplemental Table 2.4. PBL phenotypic differences between Amish and Hutterites remained after excluding children with asthma or allergic sensitization.

Cell Type	Phenotype marker	*P* value all children (t test)	*P* value asthmatic/atopic excluded
Neutrophil	Percent of PBLs	0.0055	0.0102
Neutrophil	CXCR4	<0.0001	<0.0001
Neutrophil	CD11c	0.0039	0.0026
Neutrophil	CD11b	<0.0001	<0.0001
Eosinophils	Percent of PBLs	0.0005	0.0088
Monocyte	Percent of PBLs	0.28	0.86
Monocyte	HLA-DR	<0.0001	0.0007
Monocyte	ILT3	0.0041	0.0041
Treg	Percent of PBLs	0.29	0.13

Supplemental Table 2.5. Cytokine levels in supernatants from LPS-treated PBLs.

22 cytokines measured on the Milliplex Map Human TH17 Magnetic Bead Panel were measured at detectable levels in supernatant of LPS-treated PBLs. IL-6 and IL-1β were also measured in LPS-treated supernatant of PBLs using an ELISA.

Cytokine	Amish	Hutterite	*P* value
IL-17	6.19 (4.85, 7.15)	9.24 (7.56, 12.23)	2.97×10^{-6}*
IL-33	12.16 (6.2, 19.1)	25.8 (20.3, 33.1)	4.47×10^{-6}*
IL-25	0.10 (0.07, 0.15)	0.20 (0.14, 0.24)	1.16×10^{-5}*
IL-31	0.01 (0, 0.03)	0.05 (0.03, 0.07)	6.16×10^{-5}*
IL-27	0.25 (0.22, 0.29)	0.34 (0.26, 0.4)	8.38×10^{-5}*
IL-4	0 (0, 0.019)	0.03 (0.011, 0.05)	1.25×10^{-3}*
IL-5	1.96 (1.75, 2.31)	2.79 (2.12, 3.94)	2.37×10^{-3}
IL-22	0.34 (0.29, 0.44)	0.49 (0.39, 0.56)	2.49×10^{-3}
IL-2	0.77 (0.44, 3.58)	5.89 (1.15, 9.38)	3.86×10^{-3}
IL-15	5.52 (2.8, 8.03)	8.74 (6.31, 12.92)	0.01
IL-10	58.81 (36.94, 79.36)	85.86 (56.93,112.2)	0.04
IL-9	2.99 (1.72, 4.29)	3.90 (2.57, 5.81)	0.08
IL-13	141.2 (112.5, 181.4)	161.5 (131.7, 214.7)	0.12
IL-12p70	51.4 (33.0, 81.9)	77.0 (41.5, 119.3)	0.13
GMCSF	0.143 (0.096, 0.179)	0.108 (0.060, 0.164)	0.15
MIP3α	3039 (2160, 4137)	3379 (2618, 4679)	0.27
IL-1β	40810 (21270, 66250)	49690 (34900, 66250)	0.28
IL-6	10000 (10000, 10000)	10000 (8097, 10000)	0.42
TNFα	5866 (2300, 9646)	8020 (2536, 10120)	0.56
IL-28A	0.23 (0.14, 0.43)	0.31 (0.15, 0.57)	0.59
IL-23	3.18 (0.42, 4.29)	3.21 (0.56, 5.12)	0.66
IFN-γ	2434 (1575, 6043)	2432 (1426, 5822)	0.67

Median cytokine levels (pg/μL) are shown with first and third interquartile range presented in parentheses
P value determined by Wilcoxon Rank Sum Test
* Bonferroni significant ($P< 2.3\times10^{-3}$)

Supplemental Table 2.6. Cytokine levels in supernatants from LPS-treated PBLs (measured by ELISA).

Cytokine	Amish	Hutterite	*P* value
IL-1β	3,441 (2908, 3966)	3828 (2852, 4587)	0.29
IL-8	11210 (6251, 13120)	11240 (8948, 15110)	0.24
IL-6	887400 (782200, 964300)	857400 (808200, 994900)	0.76

Median cytokine levels (pg/μL) are shown with first and third interquartile range presented in parentheses
P value determined by Wilcoxon Rank Sum Test

Supplemental Table 2.7. Cytokines from LPS-treated supernatant after excluding children with asthma or allergic sensitization.

Cytokine	Amish	Hutterite	*P* value
IL-17	6.19	9.43	3.04×10^{-4}*
IL-33	12.67	24.79	3.66×10^{-4}*
IL-25	0.1	0.19	1.04×10^{-4} *
IL-31	0.02	0.05	5.23×10^{-4} *
IL-27	0.25	0.34	1.16×10^{-3}*
IL-4	0	0.04	3.02×10^{-4}*
IL-5	2.02	2.77	0.06
IL-22	0.42	0.48	0.06
IL-2	0.44	7.88	1.64×10^{-3}*
IL-15	5.71	9.51	0.02
IL-10	60.41	89.58	0.02
IL-9	2.95	3.97	0.06
IL-13	137.0	160.4	0.21
IL-12p70	69.67	62.92	0.50
GMCSF	0.14	0.12	0.25
MIP3α	2993	3660	0.06
IL-1β	37014	49689	0.19
IL-6	10001	10001	0.39
TNFα	6432	5943	0.64
IL-28A	0.22	0.23	0.76
IL-23	3.20	2.72	0.68
IFN-γ	2218	1804	0.30

Median cytokine levels (pg/μL) are shown with first and third interquartile range presented in parentheses
P value determined by Wilcoxon Rank Sum Test
* Bonferroni significant ($P< 2.3\times10^{-3}$)

Supplemental Table 2.8. Cytokine levels in supernatants from PBLs treated with anti-CD3 and anti-CD28 antibodies.

Cytokine	Amish	Hutterite	*P* value
IL-33	10.32 (6.49, 13.40)	17.83 (12.44, 23.58)	1.01×10^{-4}*
IL-31	0.005 (0, 0.019)	0.032 (0.011, 0.045)	5.86×10^{-4}*
IL-27	0.12 (0.08, 0.15)	0.16 (0.11, 0.21)	2.61×10^{-3}
IL-12p70	8.37 (6.19, 13.78)	13.30 (11.26, 20.66)	2.88×10^{-3}
IL-25	0.085 (0.07, 0.12)	0.13 (0.09, 0.20)	7.53×10^{-3}
IL-23	0.42 (0.04, 0.61)	0.60 (0.11, 1.03)	0.04
IFN-γ	461.2 (96.66, 907.1)	865.3 (254.7, 1598)	0.07
IL-15	3.62 (0.28, 6.41)	5.11 (3.74, 7.35)	0.08
MIP3α	312.5 (110.8, 470.9)	445.0 (260.7, 643.6)	0.11
IL-10	68.54 (12.61, 170.5)	149.1 (56.86, 182.8)	0.12
IL-28A	0.20 (0.15, 0.54)	0.35 (0.18, 0.71)	0.14
IL-2	101.7 (22.66, 214.4)	144.8 (83.53, 332.7)	0.16
IL-17	53.25 (23.28, 84.96)	66.75 (40.12, 115.9)	0.22
IL-4	0.36 (0.015, 0.62)	0.25 (0.09, 0.32)	0.25
IL-22	0.36 (0.26, 0.48)	0.39 (0.30, 0.48)	0.29
IL-5	114.8 (13.09, 227.7)	91.95 (42.55, 140.8)	0.52
GMCSF	0.48 (0.096, 0.74)	0.51 (0.19, 0.66)	0.73
IL-13	411.7 (149.9, 668.8)	353.0 (207.2, 495.3)	0.81
TNFα	430.6 (142.4, 710.5)	486.3 (167.9, 633.4)	0.85
IL-1b	44.66 (23.55, 101.3)	41.19 (25.99, 81.89)	0.93
IL-6	58.68 (13.78, 122.5)	58.23 (38.99, 101.2)	0.97
IL-9	21.29 (5.47, 34.18)	19.38 (7.35, 37.06)	0.97

Median cytokine levels (pg/μL) are shown with first and third interquartile range presented in parentheses
P value determined by Wilcoxon Rank Sum Test
* Bonferroni significant ($P< 2.3\times10^{-3}$)

Supplemental Table 2.9. Correlation of Weighted gene co-expression network analysis (WGCNA) module eigengenes to cell proportions.

Nine of the 2,809 genes that had expression patterns that did not fit into any module. Correlation with cell proportion phenotypes determined with Pearson's correlation. The square root of the linear regression r^2 was used as an approximate correlation coefficient of module eigengene with Amish or Hutterite status.

Module number	Number genes in module	Correlation with Neutrophils (*P* value)	Correlation with Eosinophils (*P* value)	Correlation with Monocytes (*P* value)	Correlation with Amish/ Hutterite (*P* value)	Number IPA networks with *P* value $>1.0\times10^{-20}$ (*P* value)
1	43	0.60 (1.5×10^{-6})	-0.43 (1.0×10^{-3})	0.15 (0.26)	0.69 (1.5×10^{-6})	1 (1.0×10^{-30})
2	27	-0.58 (4.3×10^{-3})	0.73 (2.5×10^{-10})	-0.20 (0.14)	0.69 (4.3×10^{-6})	0
3	321	0.58 (4.2×10^{-6})	-0.49 (1.7×10^{-4})	0.20 (0.14)	0.78 (1.6×10^{-12})	1 (1.0×10^{-20})
4	296	-0.56 (8.7×10^{-6})	0.46 (4.6×10^{-4})	-0.18 (0.19)	0.66 (5.1×10^{-8})	0
5	839	0.54 (2.0×10^{-5})	-0.48 (2.0×10^{-4})	0.19 (0.16)	0.85 (3.8×10^{-16})	0
6	39	0.49 (1.6×10^{-4})	-0.52 (4.3×10^{-5})	0.18 (0.18)	0.65 (8.7×10^{-9})	0
7	105	-0.39 (3.0×10^{-3})	0.40 (2.4×10^{-3})	-0.18 (0.20)	0.58 (3.0×10^{-6})	0
8	142	-0.35 (0.01)	0.28 (0.04)	-0.21 (0.12)	0.58 (2.9×10^{-6})	1 (1.0×10^{-20})
9	340	0.36 (0.01)	-0.35 (0.01)	0.13 (0.33)	0.71 (1.1×10^{-9})	0
10	48	0.33 (0.01)	-0.24 (0.07)	0.27 (0.05)	0.63 (2.1×10^{-7})	0
11	123	-0.32 (0.02)	0.10 (0.47)	-0.11 (0.42)	0.70 (3.4×10^{-9})	2 (1.0×10^{-23} 1.0×10^{-21})
12	127	0.33 (0.02)	-0.28 (0.04)	0.23 (0.09)	0.71 (9.4×10^{-10})	1 (1.0×10^{-20})
13	108	-0.27 (0.04)	0.26 (0.05)	-0.18 (0.20)	0.55 (1.2×10^{-5})	0
14	143	0.26 (0.06)	-0.29 (0.03)	0.10 (0.45)	0.59 (2.0×10^{-6})	1 (1.0×10^{-29})
15	99	-0.21 (0.13)	0.08 (0.58)	-0.12 (0.40)	0.61 (6.1×10^{-7})	1 (1.0×10^{-21})

Supplemental Table 2.10. Effect of Amish dust extracts on total BAL cellularity and serum OVA-specific IgE levels in OVA-treated mice.

	Saline	OVA	OVA/Amish	P value
BAL Total Cells	409,425 ± 142,358	2,356,250 ± 647,484	1,144,458 ± 209,521	0.08
Total Eosinophils	7,284 ± 5,574	1,076,755 ± 340,283	61,812 ± 21,864	0.007*
Total Neutrophils	2,376 ± 193	130,872 ± 57,921	153,266 ± 20,156	0.72
Total Macrophages	137,677 ± 4,800	974,787 ± 259,360	917,374 ± 191,636	0.86
Serum OVA-specific IgE	496.6 ± 10.5	1457.7 ± 254.7	858.9 ± 56.6	0.031*

C57BL6 mice (n = 12 mice/group) were treated with Amish house dust extracts (7.5 mg of dust equivalent in 50 µl) i.n. every 2-3 days for a total of 14 times beginning at day -5 and sensitized i.p. with 20 µg OVA-Alum at day 0 and 14. Animals were challenged i.n. with 75 µg OVA at day 26, 27 and 28 and sacrificed at day 30. Cells were counted using a hemocytometer and differentials were determined by an operator blinded to mouse ID/grouping after staining with Hema 3 and examining at least 400 cells/slide. OVA-specific IgE in serum obtained at day 30 were measured by ELISA. Shown are means ± SE.
P value for OVA vs. OVA/Amish as determined by Student's t test, * = P<0.05

Supplemental Table 2.11. Percentages of CD3+ CD4+ Foxp3+ regulatory T cells in the lungs of mice treated intranasally with Amish or Hutterite dust extracts.

PBS	Amish	Hutterite	*P* value
4.84 ± 1.93	8.97 ± 1.36	8.78 ± 2.54	0.69

Balb/c mice (n = 4 mice/group) received dust extracts (5 mg of dust equivalent/treatment) i.n. 4 times over 10 days. Following digestion with Liberase, single cell lung suspensions ($1x10^6$ cells) were stained with anti-mouse CD3-FITC, Foxp3-PE, CD4-PECy5 and CD8-AF700. Data were acquired on an LSR II flow cytometer and analyzed with FlowJo software gating on CD3+ cells. Shown are means ± SE.
P value for Amish vs. Hutterite determined by the Wilcoxon two-sample test, * = P<0.05

Supplemental Table 2.12. Effects of Amish and Hutterite dust extracts on cytokine levels in BAL collected from OVA-treated Balb/c mice at day 39.

All cytokines were measured by ELISA.

Cytokine	PBS	OVA	OVA/Amish	OVA/Hutterite	P value
IL-13	3.9 (3.9-3.9)	116.4 (47.6-182)	3.9 (3.9-3.9)	47.02 (33.9-107.6)	0.007*
IL-5	7.8 (7.8-7.8)	142.7 (107.8-210.5)	40.56 (30.8-46.7)	101.80 (68.7-131.4)	0.014*
IL-4	3.9 (3.9-3.9)	42.4 (21.8-213.0)	13.12 (3.9-152.7)	35.8 (13.4-70.2)	0.22
IL-10	7.8 (7.8-7.8)	33.4 (19.9-49)	7.8 (7.8-7.8)	15.7 (7.8-35.4)	0.029*
MIP-1α	7.9 (3.9-12.1)	19.9 (11.4-31.2)	7.22 (3.9-9.9)	19.42 (13.1-33.9)	0.014*

Median cytokine levels (pg/ml) are shown, with range in parentheses. Undetectable samples were assigned a value corresponding to half of the lower limit of sensitivity of the relevant standard curve.
**P* value for OVA/Amish vs. OVA/Hutterite determined by the Wilcoxon rank-sum test (Mann-Whitney; corrected for ties, * = $P<0.05$)
n=4-5 mice/group

Chapter 3

Profound effects of FcγIIA polymorphism (Arg131His) on blood leukocyte global gene expression and cytokine responses to treatment with anti-CD3 and anti-CD28 IgG1

3.1 Abstract

The low affinity Fcγ receptor, FcγIIA, harbors a common variant, rs1801274 that results in a missense mutation (G>A, Arg131His) that modifies binding affinity to human IgG2 and mouse IgG1 antibodies. Previous studies have reported associations between this polymorphisms and autoimmune and inflammatory diseases, as well as with efficacy of cancer immunotherapies. Despite the potentially important role of the Arg131His variant, little is known about genotype effects on global gene expression and cytokine responses between individuals. To address this gap in knowledge, we treated whole blood samples from 130 individuals with mouse IgG1 anti-CD3 and anti-CD28 antibodies, and characterized the genome-wide gene expression profiles and cytokines among individuals stratified by genotype at this variant. A genome-wide association study of anti-CD3 and anti-CD28-treated cytokines revealed significantly decreased expression of IL-2, GM-CSF, IFNγ, and TNFα, and nominally significant associations in an additional 13 cytokines among homozygotes for the low affinity FcγIIA His allele. Moreover, large differences in gene expression patterns were observed between all three genotype classes, including many genes that differed between heterozygous individuals and both homozygous genotype classes. Overall, heterozygotes showed suboptimal T cell receptor activation compared to samples from individuals homozygous for the higher affinity FcγIIA Arg allele. The results of this study provide the novel observation that variation in response to IgG treatment may vary by genotype

and result in profound differences in both cytokine responses and global gene expression patterns in whole blood leukocytes, and yields potential insight into the heterogeneity of clinical outcomes from therapeutic antibody treatment.

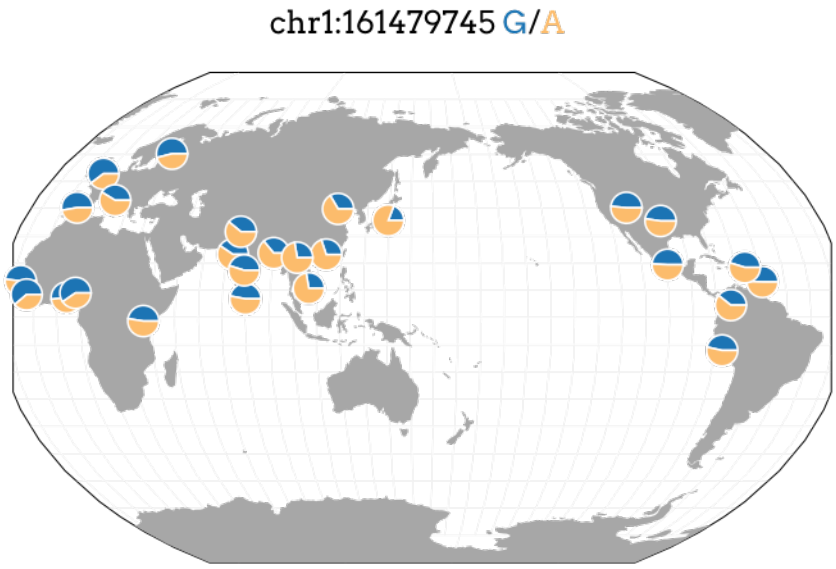
3.2 Introduction

Fc receptors (FcRs) bind to the Fc region of immunoglobins (Ig) G, IgM, IgE, and IgA secreted by B cells. FcRs play a critical role in activating neutrophils and NK cells to kill infected cells during the process of antibody-dependent cell-mediated cytotoxicity, and in inducing basophil and mast cell degranulation[142]. The class of FcRs that bind to IgG, FcγR, can be activating (FcγRI, FcγRIIA, FcγRIIC, FcγIIIA) or inhibitory (FcγRIIB), and are further classified by their affinity to different classes of IgG antibodies, such that FcγRI is a high affinity receptor and FcγRIIA is a low affinity receptor for IgG1, for example[142]. FcγIIA is broadly expressed on monocytes, macrophages, dendritic cells, eosinophils, and neutrophils, and is altered by a missense mutation (G>A, rs1801274) in the fourth exon of the *FCGR2A* gene encoding FcγIIA. This polymorphism results in an arginine to histidine (Arg131His) amino acid substitution within the binding interface, resulting in increased affinity of the His (A) allele for the monomeric human IgG2 isotype of IgG[143,144] and an increase in phagocytosis of opsonized bacteria in granulocytes from individuals with AA (His/His) genotype at rs1801274[145]. The prevalence of this common mutation varies widely around the world (Figure 3.1), with frequency of the His (A) allele ranging from 0.81 in Japanese to 0.39 in East Africans[146]. Genome-wide association studies (GWAS) have identified associations between rs1801274 genotype and autoimmune and inflammatory diseases, such as ulcerative colitis[147], Kawasaki disease[148], systemic lupus

erythematosus[149], periodontitis[150], and severe sepsis[151]. In all conditions except severe sepsis, the G (Arg) allele was associated with increased risk.

Figure 3.1. Distribution of rs1801274 allele frequency in 1000 genomes populations.

Relative frequency of the G allele (Arg) colored in blue, and relative frequency of the A allele (His) colored in yellow. Different populations are indicated by different pie charts.



In addition to modifying affinity to human IgG2, the rs1801274 polymorphism changes binding of human cells to mouse IgG1, an antibody frequently used in *in vitro* studies[152]. In contrast to binding of human IgG2, the Arg (G) allele has *higher* affinity to the mouse IgG1 antibody compared to the His (A) allele. The effects of this polymorphism on peripheral blood mononuclear cell responses to soluble monomeric mouse IgG1 anti-CD3 antibody was reported over 30 years ago when OKT3 (anti-CD3 antibody) treatment failed to induce a mitogenic response in blood cells from five of eight members of a single family[153,154]. Despite these

observations, mouse IgG1 antibodies, and specifically anti-CD3 and/or anti-CD28 antibodies, continue to be used in studies of cytokine responses in stimulated blood cells[155,156].

The effect of the rs1801274 genotype on response to antibodies is not limited to *in vitro* studies – the efficacy of several therapeutic antibodies is also modified by this polymorphism. For example, the proapoptotic receptor agonist, drozitumab, an antibody against death receptor 5 (DR5) has a two- to three-fold higher affinity for 131His FcγIIA compared to 131Arg FcγIIA[157]. Follicular lymphoma patients with the His/His (AA) genotype had higher response rates to a chimeric monoclonal antibody with a human Fc region, rituximab, after 6, 9, and 12 months of treatment[158]. The response rate of Arg/Arg (GG) patients with metastatic colorectal cancer improved with therapy that included the addition of the monoclonal chimeric antibody cetuximab, which contains a human IgG1 Fc region[159], and Arg/Arg rheumatoid arthritis patients had better responses to a human IgG1 antibody, adalimumab, a TNF inhibitor[160,161]. In two retrospective studies of the anti-HER2 breast cancer treatment, trastuzumab, cells from patients with the His/His genotype had increased trastuzumab-mediated cytotoxicity[162] and these patients had a significantly better response rate compared to those with the Arg/Arg genotype[163]. Other studies, however, reported that this polymorphism in the *FCGR2A* gene did not affect the outcome in follicular lymphoma patients treated with rituximab[164] or tumor response after treatment with cetuximab in patients with advanced colorectal cancer[165], and was not predictive of response to trastuzumab in breast cancer patients[166].

In spite of the potential clinical and experimental importance of the Arg131His genotype on human immune responses to different classes of mouse and human antibodies, relatively little is known of the effect of this genotype on broad cytokine or genome-wide gene expression responses. Moreover, heterozygous (AG) individuals are typically grouped with the homozygote

class not of interest to the study (e.g., see ref.[161]), assuming that only the homozygous genotype is functionally affected by this variant. As a result, the effect of heterozygosity at this variant site on response is largely unknown.

During the course of our GWAS of cytokine and global gene expression responses of whole blood leukocytes to anti-CD3+anti-CD28 antibodies, we observed a dramatic impact of this polymorphism on a broad panel of cytokine responses as well as on global gene expression responses. Moreover, our studies clearly show that individuals with the His/Arg (A/G) heterozygous genotype respond differently to treatment with mouse anti-CD3+anti-CD28 antibodies compared to either the His/His (AA) or Arg/Arg (GG) homozygotes at both the gene expression and cytokine levels. These observations yield insight into the gene regulation of anti-CD3+anti-CD28 response, and challenge the assumption of equivalent responses in individuals with the AA (His/His) and AG (His/Arg) genotypes, assumed in previous studies[155,161].

3.3 Methods

3.3.1 Study population

This study was conducted in 153 Hutterites (7-76 years old) who are a subset of the >1,400 Hutterites who have participated in our population-based studies of complex phenotypes [131,167,168]. The Hutterites are an Anabaptist sect of central European ancestry. The Hutterites in our studies live on communal farms in South Dakota and are related to each other through multiple lines of descent in a single 3,671-member pedigree with 64 ancestors. Written consent for these studies was obtained from the adult participants and parents of children under 18; written assent was obtained from all children.

3.3.2 Collection of Whole blood samples

One milliliter of whole blood was drawn into each of three TruCulture (Myriad RBM) tubes containing either proprietary TruCulture media alone, media + 0.1 μg/ml lipopolysaccharide (LPS), or media + 0.4 μl/ml mouse IgG1 anti-CD3 + 0.33 μl/ml mouse IgG1 anti-CD28 antibodies. Samples were incubated upright in a dry heat block at 37C for 30 hours. Following incubation, tubes were processed in a makeshift lab in South Dakota. Briefly, supernatants were separated from the cell pellets, and then aliquoted and flash-frozen on dry ice. The remaining samples were washed twice with Buffer EL (Qiagen) and the cell pellets were resuspended in 350 μl RLT Buffer (Qiagen) and frozen on dry ice. Frozen samples were shipped on dry ice overnight to Chicago and stored at -80^{o}C. DNA and RNA were extracted from thawed cell pellets using AllPrep DNA/RNA Mini Kits (Qiagen).

3.3.3 Cytokine measurement and processing

Supernatants from 130 LPS-treated and 130 anti-CD3+anti-CD28 treated samples were submitted to the Human Flow Cytometry Core at the University of Chicago for measurement of 25 cytokines using the Milliplex Th17 Magnetic Bead Panel (Millipore). Each sample was run in duplicate. Results from duplicate samples were averaged; in 10 instances, each one duplicate did not yield a detectable result and only a single measurement was retained. After removing cytokines with measurements outside the range of detection in >30% of individuals, 17 cytokines from LPS-treated samples and 21 cytokines from anti-CD3+anti-CD28-treated samples were included in downstream analysis. Cytokine measurements were quantile normalized; and principal components analysis identified bead panel "plate" as a confounding technical covariate, which was adjusted using ComBat[137].

3.3.4 Genotyping and trans cytokine QTL mapping

Of the 130 individuals with measured cytokine levels, 107 were previously genotyped on Affymetrix platforms 6.0 and 5.0; the QC of those data was described previously[131,167,168]. The remaining 23 individuals were genotyped on Illumina CytoSNP Bead Chip and the Illumina Infinium HumanCore Exome Bead Chip. Variants with call rates <90%, Hardy-Weinberg equilibrium $P<10^{-5}$, or MAF <5% were removed; all individuals had SNP call rates of >95%. Using these SNPs as framework markers, all variants present in the genome sequences of 121 Hutterites were imputed to all 130 individuals using PRIMAL, an in-house pedigree-based imputation method with >99% accuracy in the Hutterites[132]. This yielded a final set of 4,998,317 imputed variants with a call rate >80% and MAF >5% that were retained for downstream analyses. Associations between each variant and each cytokine level were calculated using GEMMA[169], including a kinship matrix based on identity by descent (IBD) to correct for the relatedness between Hutterites in our sample. In all analyses, age and sex were included as covariates; P-values of $<5\times10^{-8}$ was used as the threshold for significance in genome-wide analyses. A Wilcoxon rank-sum test was used to determine if cytokines differed between individuals AG or GG at rs1801274, in which we used a Bonferroni-corrected P-value threshold ($P=2.3\times10^{-3}$) to correct for the 21 tests.

3.3.5 Analysis of Gene expression

RNA samples from all three TruCulture tubes was used to create RNA-seq libraries using the TruSeq Library kit (Illumina); quality and concentration of libraries were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantitative PCR using the Kapa library

quantification kit (Kapa Biosystems). Samples were sequenced in pools of 16-18 samples across three flow cells of an Illumina HiSeq 2500; 119 samples with low read count were re-sequenced on two flow cells on the same machine. Adaptor sequences were trimmed, and then reads were mapped to hg19 using Tophat2[170]. Reads were sorted and indexed with samtools[171], and genes counted with HTseq[172]. Read counts for the same sample sequenced multiple times were summed together, and then Trimmed Means of M-values normalization (TMM) and a voom transformation was used to correct for differences in library sizes[173]. Confounding technical effects were assessed in the normalized expression data using principal components analysis (PCA), and sequencing pool was adjusted using the function RemoveBatchEffect() from the R package limma[174]. PCA was also used to visualize differences in overall patterns of gene expression variation according to rs1801274 genotype in three analyses: (1) untreated vs. anti-CD3+anti-CD28 treated samples; (2) within anti-CD3+anti-CD28-treated samples; and (3) within anti-CD3+anti-CD28-treated samples after removing individuals homozygous AA at rs1801274.

Differential gene expression in anti-CD3+anti-CD28 treated samples by rs1801274 genotypes AG vs. GG was examined using a linear model, with age and sex included as covariates. Significance was assessed using the methods of Benjamini and Hochberg[113]. Genes differentially expressed at a false discovery rate of 5% were analyzed with Ingenuity Pathway Analysis (Qiagen, IPA) to identify networks and upstream regulators enriched within the gene set.

3.4 Results

3.4.1 GWAS of cytokine responses to anti-CD3 + anti-CD28 antibodies.

We used mouse IgG1 anti-CD3 and IgG1 anti-CD28 antibodies, as prepared in the TruCulture tubes, to stimulate T cells for studies of cytokine and gene expression responses. We conducted a GWAS for each of the 21 cytokines detected in our sample. This GWAS identified highly significant associations between rs1801274 genotype and four cytokines (Table 3.1; Figure 3.2, Figure 3.3). The responses of an additional two cytokines (Table 3.1; Figure 3.2, Figure 3.3) were also associated with rs1801274 genotype, but at a lower level of significance ($P<5\times10^{-6}$).

Table 3.1. Summary table of results of association studies of rs1801274 (ch1:161479745) at the *FCGR2A* locus and cytokines.

Cytokines with significant ($P<5\times10^{-8}$) or suggestive ($P<5\times10^{-6}$) associations are denoted with * or §, respectively. The beta values are shown for the minor allele (A, MAF=0.326). rs1801274 was genotyped in 118/130 individuals.

Cytokine	P-value	Beta	Standard error (Beta)
IL-2	6.59×10^{-13} *	-0.815	0.102
GM-CSF	2.56×10^{-9} *	0.604	0.094
IFNγ	4.81×10^{-9} *	-0.723	0.115
TNFα	7.43×10^{-9} *	-0.658	0.106
IL-1β	1.57×10^{-6} §	-0.597	0.118
IL-10	3.13×10^{-6} §	-0.600	0.123
IL-13	8.09×10^{-6}	-0.493	0.106
IL-5	1.28×10^{-5}	-0.544	0.120
MIP-3α	3.80×10^{-5}	-0.507	0.119
IL-9	3.87×10^{-5}	-0.491	0.115
IL-17	5.74×10^{-5}	-0.487	0.117
IL-6	1.60×10^{-5}	-0.510	0.131
IL-4	1.97×10^{-4}	-0.448	0.117
IL-22	1.66×10^{-3}	-0.199	0.062
TNFβ	2.26×10^{-3}	-0.326	0.105
IL-12p70	8.95×10^{-3}	-0.262	0.099
IL-17F	9.88×10^{-3}	-0.333	0.127
IL-25	0.096	-0.131	0.078
IL-21	0.101	-0.169	0.103
IL-23	0.129	-0.104	0.068
IL-15	0.179	-0.134	0.099

Figure 3.2. Manhattan plots showing association with rs1801274 in anti-CD3+anti-CD28-treated samples.

(A) IFNγ response and (B) IL-2 response. The x-axis represents location along the chromosome, y-axis indicates the –log10 P-value of each SNP tested. In each plot, the dashed line indicates genome-wide significance ($P=5\times10^{-8}$).

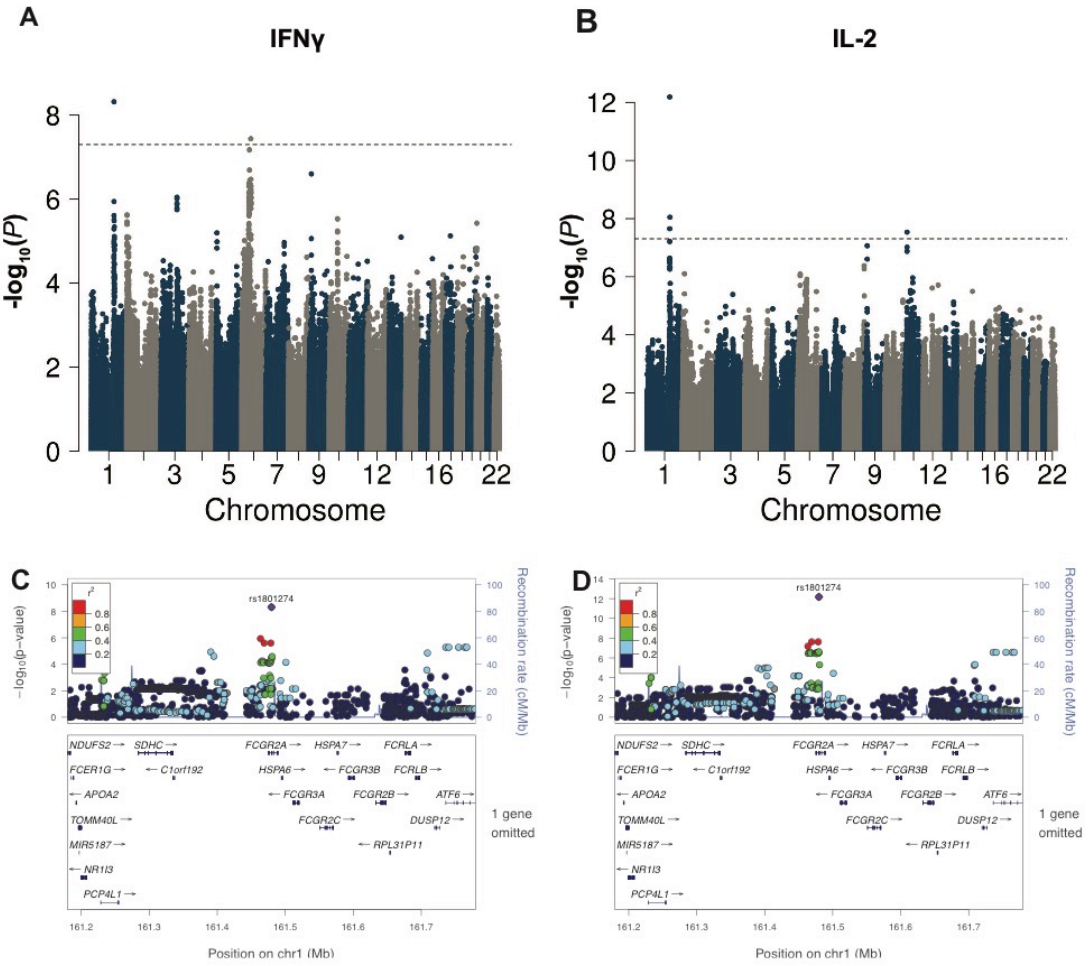
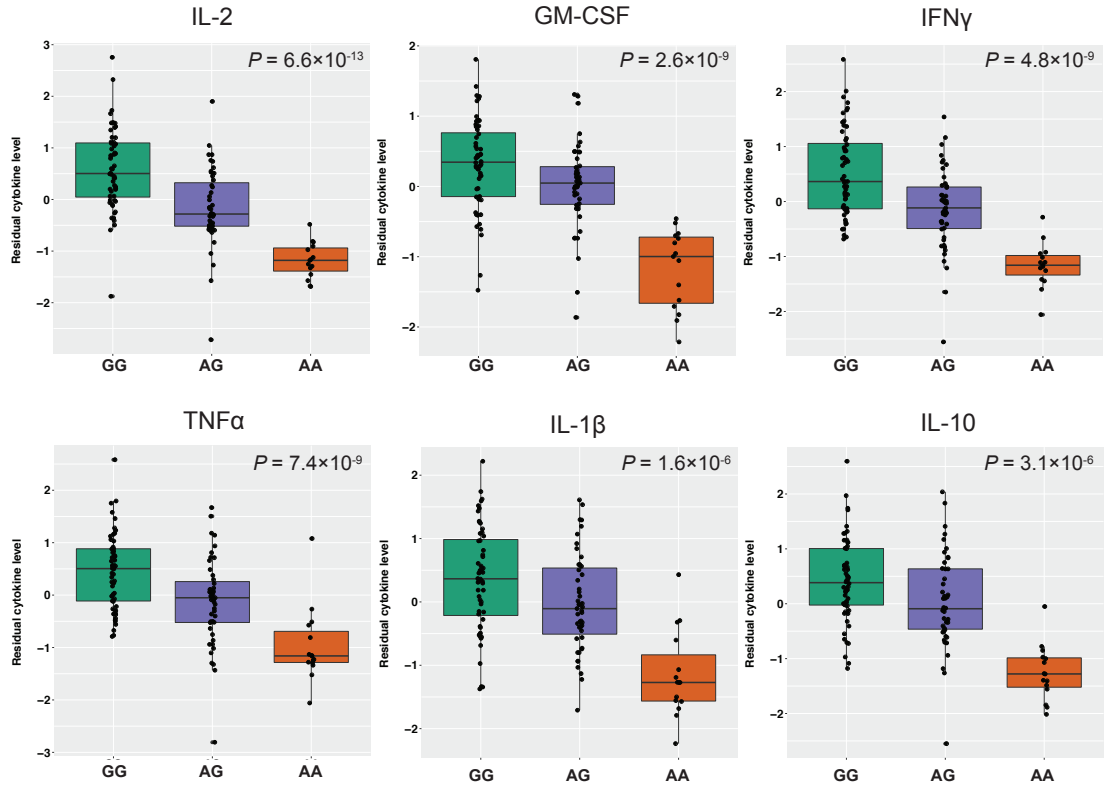


Figure 3.3. Cytokine levels following treatment with anti-CD3+anti-CD28 antibodies in whole blood by rs1801274 genotype in the *FCGR2A* gene.

Boxes indicate the interquartile ranges, whiskers represent the 95% confidence intervals. The six cytokines shown are associated with genotype at the *FCGR2A* locus at a $P<5\times10^{-6}$. For all other cytokines, see Supplementary Figure 3.1.



Overall, levels of 17 of the 21 cytokines were reduced in AA homozygotes at nominal levels of significance (P<0.01) (Table 3.1, Supplementary Figure 3.1, Panel A). Among the remaining four cytokines (IL-15, IL-21, IL-23, and IL-25) no difference were observed by rs1801274 genotype, likely due to the low levels of cytokine produced across all genotype classes in these cytokines (Table 3.1, Supplementary Figure 3.1, Panel B). Four of the 17 cytokines were also significantly reduced in the AG heterozygotes compared to GG homozygotes (IFNγ, IL-12, IL-2, and TNFα at $P=9.6\times10^{-5}$, 9.1×10^{-4}, 1.8×10^{-6}, and 8.3×10^{-4}

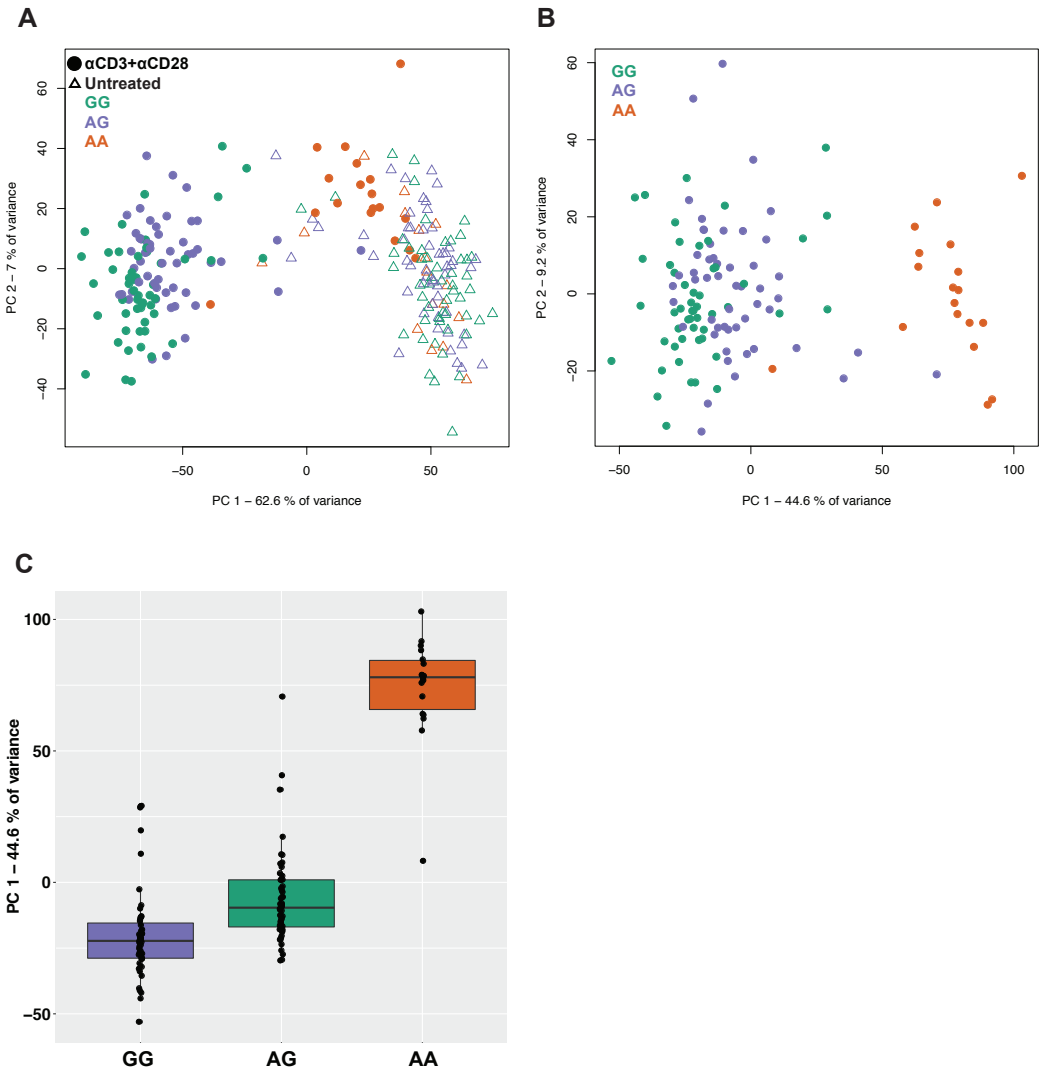
respectively; Wilcoxon rank sum test), whereas the AG heterozygotes had levels similar to the GG homozygotes for the remaining 13 cytokines, suggestive of an additive effect of the G allele in the former but more of a recessive effect in the latter group of cytokines.

3.4.2 Gene expression profiles of whole blood cells and rs1801274 genotype

The known effects of rs1801274 genotype on immunoglobulin binding and the ability to cross-link T cell receptors[143] are consistent with the cytokine response patterns described above. However, the downstream effects of this variant on global gene expression have not previously been investigated. To directly assess this, we measured gene expression in RNA from the same anti-CD3+anti-CD28 antibody-treated samples used in the cytokine studies along with the untreated samples from the same individuals. A principal component analysis (PCA) of gene expression showed clear separation of treated and untreated samples along the first PC (PC1), which explained 62.6% of the variance (Figure 3.4, Panel A), although subsets of the treated samples clustered intermediate to or even within the untreated sample cluster (Figure 3.4, Panel A). These samples were those with the AG (His/Arg) or AA (His/His) genotype at rs1801274, respectively. In a PCA of anti-CD3+anti-CD28 treated samples alone, the non-responding AA homozygotes cluster far from samples with AG (His/Arg) and GG (His/His) genotypes along PC1, accounting for 44.6% of the variance in global gene expression (Figure 3.4, Panel B). The distributions of PC1 values by genotype additionally reveals a subtle separation between samples with GG and AG genotype (Figure 3.4, Panel C, Supplementary Table 3.1). Overall, these results reveal global alterations in gene expression response to anti-CD3+anti-CD28 antibodies in whole blood cells from individuals carrying one or two copies of the A (His) allele at this SNP.

Figure 3.4. Principal components plots of gene expression profiles in whole blood.

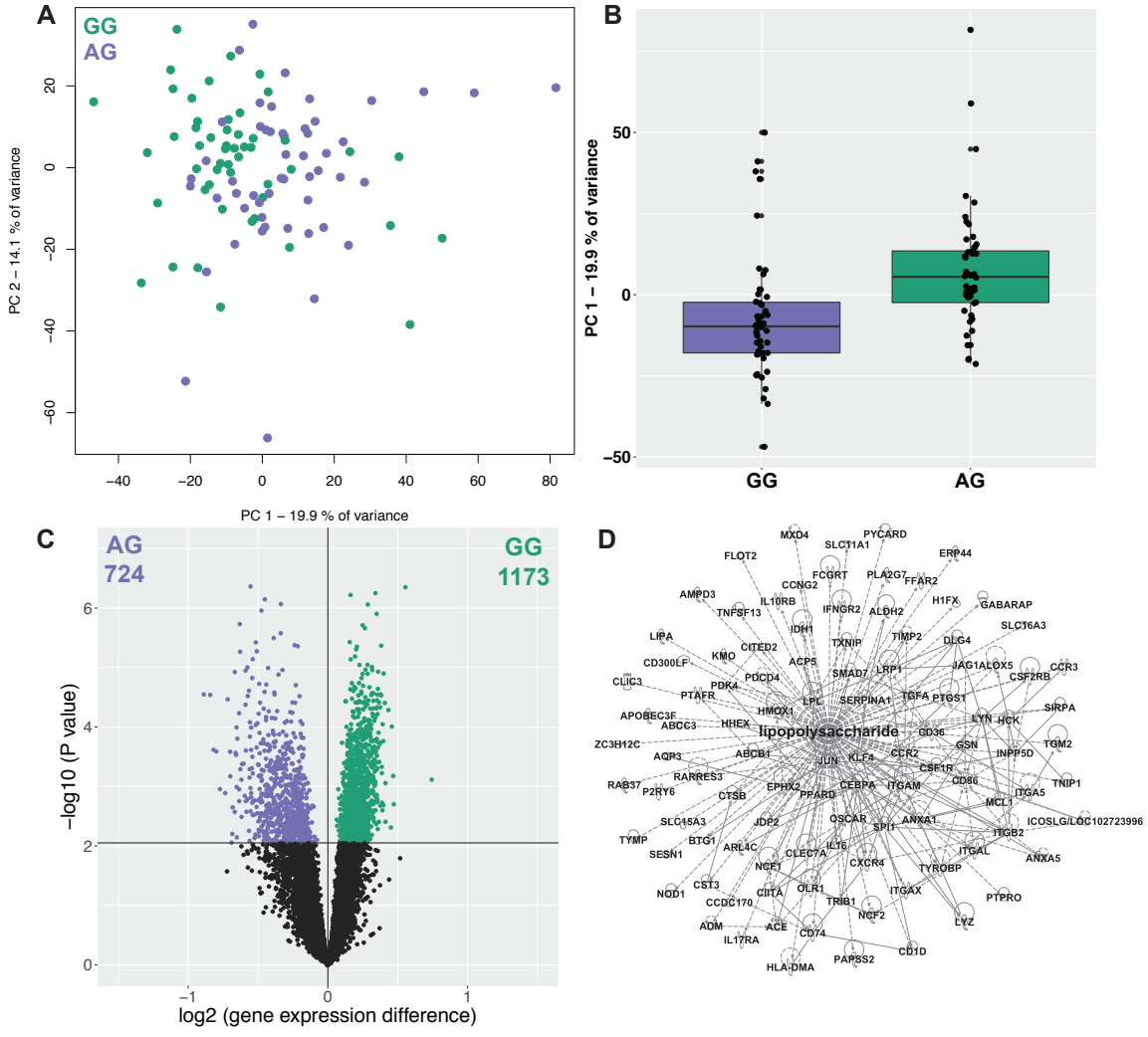
(A) PC plot of the PCs 1 and 2 (x- and y-axis, respectively) of untreated and anti-CD3+anti-CD28 treated samples colored by rs1801274 genotype. Shape indicates untreated (open triangles) or anti-CD3+anti-CD28-treated samples (filled circles). (B) PC plot of anti-CD3+anti-CD28-treated samples only, colored by rs1801274 genotype. (C) Boxplot of PC1 by rs1801274 genotype.



To better visualize gene expression differences between samples with the GG and AG genotypes, we removed individuals with the AA genotype and reran the PCA. In this analysis, genotype at rs1801274 was still significantly correlated with PC1, which now accounted for 19.9% of the variance in global gene expression response to anti-CD3+anti-CD28 treatment (Figure 3.5, Panels A-B). To identify the set of genes whose response to anti-CD3+anti-CD28 treatment differed between individuals with AG and GG genotypes, we tested for differential expression between these genotype groups. Because genotype at rs1801274 should not affect transcriptional response to LPS, we also performed the same analysis of differential expression in the LPS-treated samples as a negative control. As predicted, there were no differentially expressed genes between AG and GG individuals after treatment with LPS or with media alone, at a false discovery rate of 5%. In contrast, 1,897 genes were differentially expressed (DE) between anti-CD3+anti-CD28-treated samples with AG genotypes compared to anti-CD3+anti-CD28-treated samples with genotype GG at rs1801274 (Figure 3.5, Panel C). Ingenuity pathway analysis (IPA) of the 1,173 genes that were upregulated in GG individuals unsurprisingly identified CD3 and CD28 as among the most significant upstream regulators of these genes ($P=8.59\times10^{-10}$ and $P=2.04\times10^{-7}$, respectively; Supplementary Table 3.2), reflecting a robust T cell response among GG individuals and a significantly suppressed response among the AG heterozygotes. In contrast, the significant upstream regulators of the 503 genes with increased expression in AG heterozygotes reflected more of a generalized immune response, including LPS ($P=7.32\times10^{-13}$), TNF ($P=5.80\times10^{-13}$), and IFNγ ($P=7.55\times10^{-16}$) (Supplementary Table 3.3, Figure 3.5, Panel D), which were upstream of 44 to 81 upregulated genes in AG individuals.

Figure 3.5. Differences in anti-CD3+anti-CD28-treated gene expression responses between individuals with the AG (Arg/His) and GG (His/His) genotype at rs1801274 in the *FCGR2A* gene.

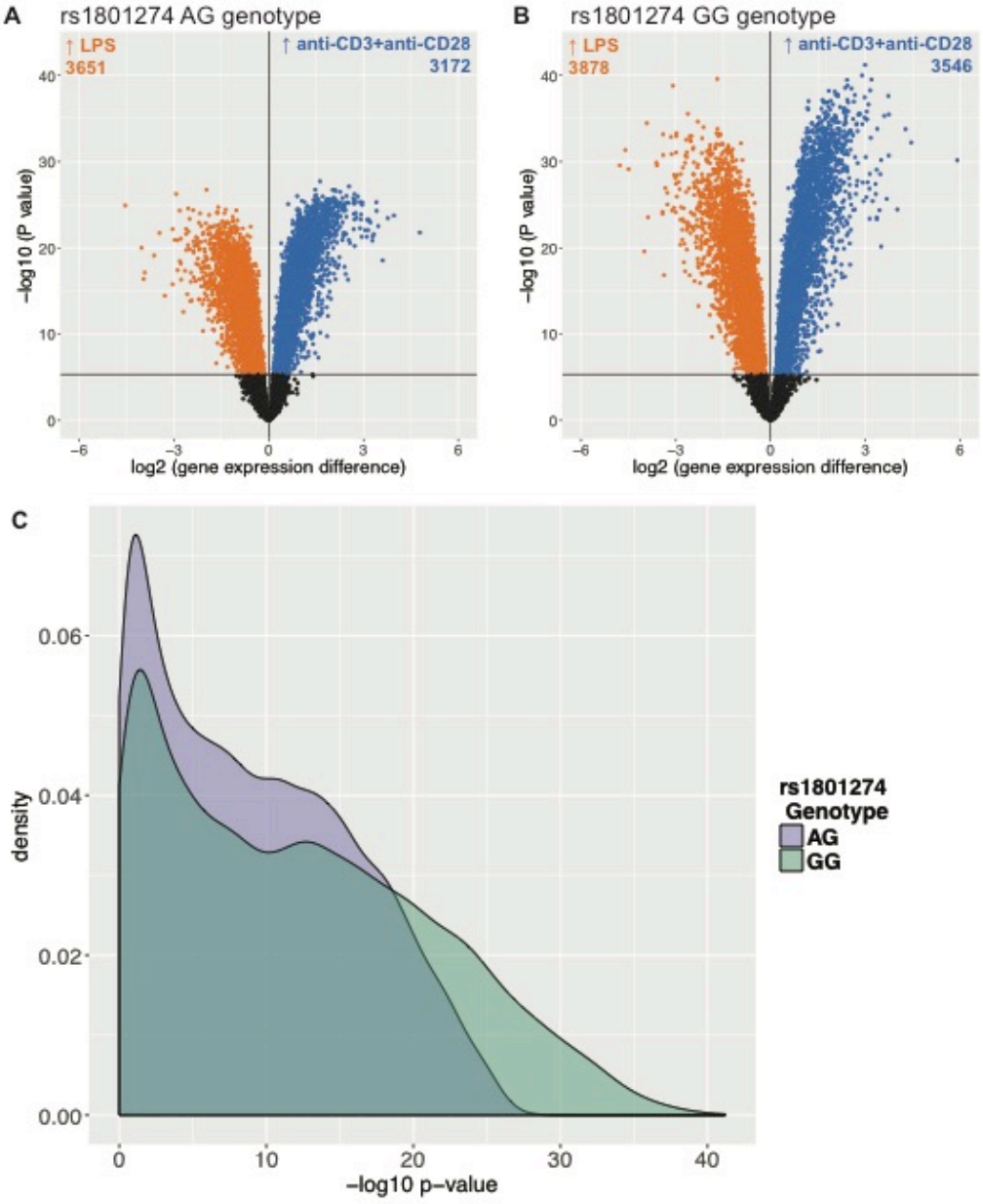
(A) PC plot of PCs 1 and 2 (x- and y-axis, respectively) of anti-CD3+anti-CD28 treated samples with AG or GG rs1801274 genotype. (B) Boxplot of PC1 by rs1801274 genotype. (C) Volcano plot of gene expression differences between treated samples from individuals with genotype AG compared to anti-CD3+anti-CD28-treated samples from individuals with genotype GG. Each point represents a gene, horizontal black line indicates the 5% FDR threshold. Genes colored blue have increased expression in GG individuals, genes colored purple have increased expression in AG individuals. Number of genes differentially expressed is listed in corresponding quadrant. (D) Network of 106 genes upregulated in AG individuals after treatment that share LPS (lipopolysaccharide) as an upstream regulator.



To investigate this finding further, we asked whether anti-CD3+anti-CD28 gene expression profiles of AG individuals are more similar to LPS-treated gene expression profiles of the same individuals compared to responses of individuals with a GG genotype to the same two treatments. Although the majority of genes were differentially expressed between LPS-treated and anti-CD3+anti-CD28-treated samples in both AG (n=47, Figure 3.6, Panel A) and GG (n=45, Figure 3.6, Panel B) individuals (94.5% of DE genes in AG individuals, 86.8% of DE genes in GG individuals), P-values were smaller and associated beta values larger in GG individuals (K-S test for P-value distributions $P<2.2\times10^{-16}$, K-S test for beta distributions $P=6.65\times10^{-11}$). These results suggest that AG individuals show less of a difference between LPS-treated and anti-CD3+anti-CD28-treated gene expression profiles compared to GG individuals. Individuals who are heterozygous AG at rs1801274 have both decreased T cell activation by anti-CD3+anti-CD28 antibodies, and relatively greater activation of innate immune pathways compared to individuals who are GG at rs1801274 (Figure 3.6, Panel C).

Figure 3.6. Gene expression profiles in LPS- and anti-CD3+anti-CD28-treated samples with the AG and GG genotypes at rs1801274 in the *FCGR2A* gene.

(A) Volcano plot of gene expression differences between LPS-treated and anti-CD3+anti-CD28-treated whole blood samples in individuals with the AG (n=47) or (B) GG (n=45) genotype at rs1801274. Genes in green are significantly increased after treatment with LPS at a Bonferroni-corrected significance threshold, genes in blue are significantly increased after anti-CD3+anti-CD28 treatment. Number of genes differentially expressed is listed in corresponding quadrant. (C) Density plot of –log10 P-values (x-axis) of differential expression by treatment and genotype; Kolmogorov-Smirnov $P<2.2\times10^{-16}$. Individuals with GG genotype at rs1801274 are skewed toward more significant P-values, and therefore, larger $-\log_{10}$ P-values.



3.5 Discussion

In this study of the effect of rs1801274 on whole blood leukocytes treated with mouse IgG1 anti-CD3 and anti-CD28 antibodies, we observed an overall striking lack of activation of the immune response, as measured by cytokine and gene expression profiles, in individuals homozygous for the A (131His) allele. In nearly every cytokine measured, AA (131His/131His) individuals had significantly lower levels after treatment with anti-CD3+anti-CD28 antibodies. Consistent with the low levels of cytokine activation, gene expression responses in anti-CD3+anti-CD28-treated whole blood leukocytes from AA homozygotes were more similar to untreated samples than treated samples from individuals with AG (His/Arg) or GG (Arg/Arg) genotypes.

On the one hand, the absence of cytokine responses to anti-CD3+anti-CD28 antibodies is in agreement with previous studies of rs1801274 in the context of differential therapeutic antibody response[160,161] or *in vitro* experiments[155]. On the other hand, differences in cytokine responses between AG heterozygotes (His/Arg) and GG homozygotes (Arg/Arg) have not been previously appreciated. In particular, in our study IFNγ, IL-12, TNFα, and IL-2 cytokine responses to anti-CD3+anti-CD28-treatment were significantly different between all three genotype classes. Such differences are likely to result in alterations of downstream immune responses. Genotype difference were further observed in gene expression response profiles, in which the expression of 18% of all expressed genes were affected, including many that are critical to immune activation processes. Importantly, in most cases, differences in expression were due to an overall decreased magnitude of response in heterozygotes compared to GG (Arg/Arg) homozygotes. Moreover, the subset of genes that were upregulated in heterozygous AG cells shared innate immune mediators as upstream regulators, suggesting both greater involvement of innate immune pathways and suboptimal activation of T cell receptors in these

individuals. That these differences may be relevant beyond laboratory studies using these antibodies is evidenced by GWAS implicating genotype at rs1801274 in susceptibility to autoimmune and inflammatory diseases[147-151]. Thus, *in vivo* responses to stimuli may be skewed more toward innate or T cell activation depending on genotype, and confer risk or protection from a variety of immune-mediated conditions.

Beyond the anti-CD3 and anti-CD28 antibodies used in this study, the broad differences in gene expression by rs1801274 genotype suggest that measurement of gene expression before and during treatment with therapeutic antibodies could be an important tool for identifying differential response to treatment by rs1801274 genotype. Study of a more proximal phenotype (i.e., gene expression) may help to dissect the inconsistent associations of the His131Arg polymorphism in the *FCGR2A* gene and tumor response and/or progression free survival reported in earlier studies[157-166].

Generalizing the results of this study comes with several caveats. Most importantly, we studied the effect of rs1801274 in response to mouse IgG1 anti-CD3 and mouse IgG1 anti-CD28 antibodies, which bind to the His allele in the FcγIIA receptors with lower affinity than to the Arg allele. In contrast, human IgG2 antibodies bind to Arg allele in the FcγIIA receptors with lower affinity than to the His allele. While we can infer that treatment of whole blood leukocytes with human IgG2 antibodies would reveal similar genotype differences in immune response, this has not been confirmed experimentally in our study or by others.

Regardless, the results of this study provide clear evidence that individual responses to antibodies differ according to rs1801274 genotype. While differences between individuals with the low-affinity homozygous genotype class (AA [His/His] for mouse IgG1, or GG [Arg/Arg] for human IgG2[143]) and the higher-affinity heterozygote and homozygous individuals have long

been known, differences between the high-affinity FcγIIA homozygote and heterozygote have not previously been explored, but have potentially important implications for future studies of response to these stimuli as well as to studies of therapeutic antibodies. *In vitro* studies can side-step this issue with the use of antibody beads, which act as a steric hindrance to the FcγIIA-Fc binding interface. However, in circumstances where the use of soluble immunoglobins cannot be avoided, designs and interpretation of results should consider the rs1801274 genotype of the study subjects. Moving towards a more nuanced understanding of the spectrum of rs1801274 effects on gene expression and immune responses may contribute to a greater understanding of variable therapeutic treatment outcomes, to improved efficacy of treatment among specific genotype classes, and potentially to risk for immune-mediated conditions. Assessment of genotype effects for this polymorphism in more general clinical contexts may even broaden further the relevance of this functional variant to precision medicine.

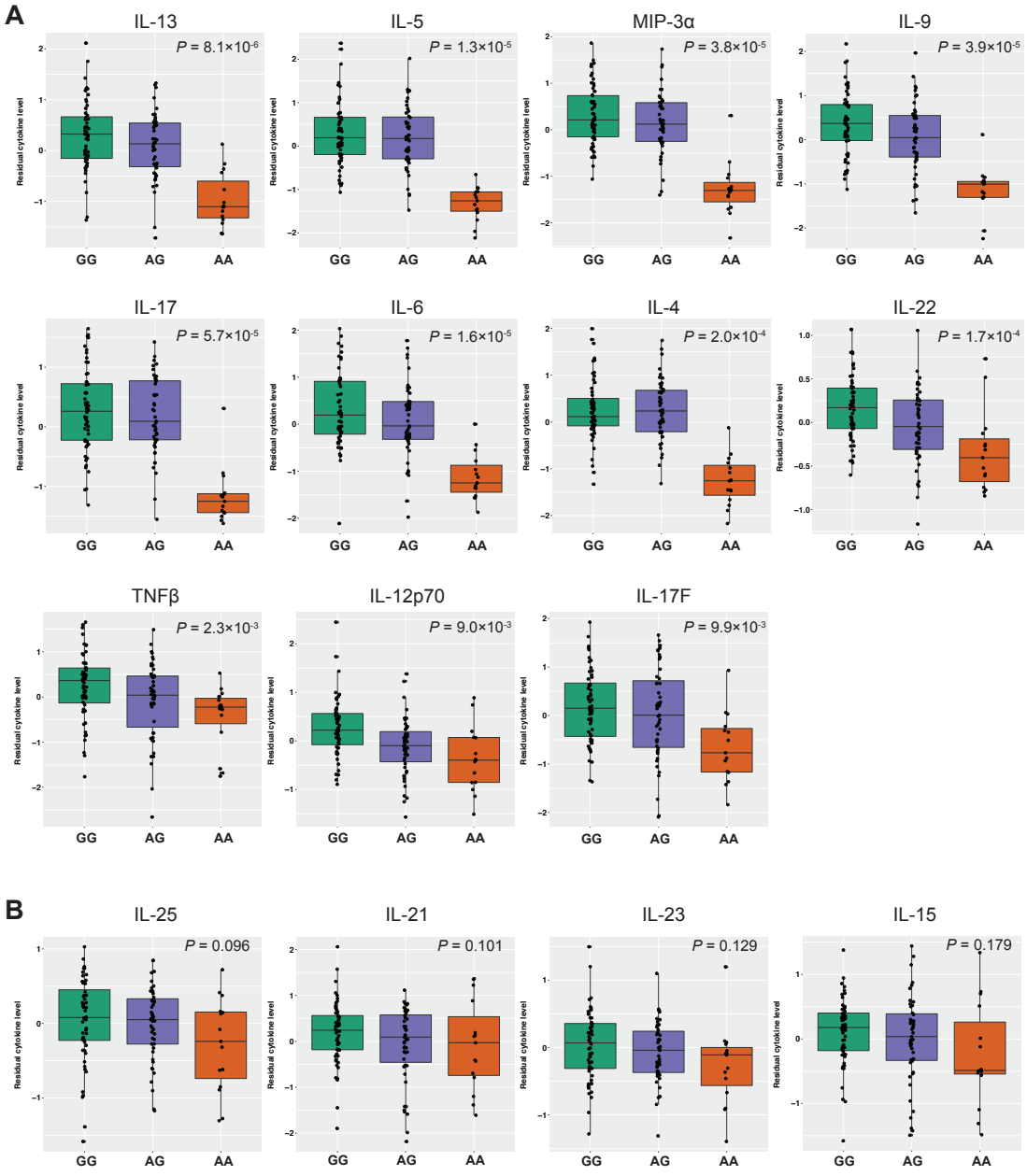
3.6 Acknowledgments

The authors would like to thank Christine Billstrand and Raluca Nicolae for making RNAseq sample libraries; Tiratat Patana-anake for generating the Hutterite imputation genotype data; and Catherine Igartua and Mark Abney for helpful comments and statistical advice.

3.7 Appendix D: Supplementary Figure and Tables

Supplemental Figure 3.1. Cytokine levels after treatment with anti-CD3+anti-CD28 antibodies of whole blood cells by rs1801274 genotype at the *FCGR2A* locus.

Box indicates the interquartile range, whiskers represent the 95% confidence interval. (A) Cytokines nominally associated with rs1801274 at a $P<10^{-3}$, (B) Cytokines not associated with the rs1801274.



Supplemental Table 3.1. Correlation P-values between rs1801274 genotype and PCs from anti-CD3+anti-CD28-treated gene expression profiles.

PCs calculated from 10,654 expressed genes in data set.

PC (% variance)	rs1801274 genotype
PC1 (44.6)	2.68×10^{-37}
PC2 (9.2)	0.610
PC3 (6.3)	0.291
PC4 (4.0)	0.119
PC5 (3.6)	0.414
PC6 (2.6)	0.512
PC7 (2.3)	0.888
PC8 (1.7)	0.016
PC9 (1.4)	0.238
PC10 (1.3)	0.534

Supplemental Table 3.2. Significant upstream regulators of genes with increased expression in individuals who are GG at rs1801274 compared to individuals who are AG following anti-CD3+anti-CD28 treatment.

Upstream regulators determined using Ingenuity Pathway Analysis (IPA).

Upstream regulator	P-value of overlap	Number target genes in data set
HNF4A	4.71E-25	220
MMP3	2.97E-15	31
MYC	6.95E-11	101
CD3	8.59E-10	74
1,2-dithiol-3-thione	3.57E-09	35
E2F1	1.57E-08	45
let-7a-5p (and other miRNAs w/seed GAGGUAG)	1.29E-07	26
5-fluorouracil	1.39E-07	29
CD28	2.04E-07	43
MAX	4.28E-07	16

Supplemental Table 3.3. Significant upstream regulators of gene with increased expression in individuals who are AG at rs1801274 compared to individuals who are GG following anti-CD3+anti-CD28 treatment.

Upstream regulators determined using Ingenuity Pathway Analysis (IPA).

Upstream regulator	P-value of overlap	Number target genes in data set
Genistein	2.27E-18	52
TGFB1	7.42E-16	109
IFNG	7.55E-16	94
Fluticasone	2.98E-14	31
TNF	5.80E-13	109
LPS	7.32E-13	105
Dexamethasone	1.49E-12	101
TP53	2.37E-12	85
Immunoglobulin	7.61E-12	32
IL13	2.51E-08	36

Chapter 4

A cytokine QTL study in LPS-treated whole blood in a founder population

4.1 Abstract

Expression quantitative trait loci (eQTL) studies have identified numerous examples of genetic variation regulating gene expression following treatment, particularly in the case of peripheral blood cells and immune stimuli. However, studies of protein QTLs remain relatively unexplored, likely due to the difficulty and cost of measuring the entire proteome. To identify genetic regulation underlying protein response to an innate immune stimulus, LPS, we narrowed the scope of our analysis to cytokines, which play pivotal roles in immune activation. Using a relatively small sample size from a founder population of European descent, we measured 17 cytokines and from LPS-treated whole blood supernatant and tested for associations with DNA variation ± 250kb of cytokine and cytokine receptor genes. We identified one significant (permuted $P<0.05$) and one suggestive (permuted $P<0.1$) association with IL-12 and IL-27, respectively, and found evidence of non-zero heritability in 7 of the 17 cytokines tested. Based on the paucity of significant results, we recommend that future cytokine QTL studies be conducted using a more constricted set of peripheral blood leukocytes instead of whole blood.

4.2 Introduction

Although genome-wide association studies (GWAS) of immune-mediated diseases have led to the discovery of many risk alleles, these disease-associated variants account for little of their heritabilities. It is likely that some of this "missing heritability" for immune phenotypes is due to gene by environment interactions (GEIs), most of which would not be discovered in

GWAS[175,176]. Nonetheless, the collectively large experience of GWAS has revealed important insights into the genetic architecture of complex phenotypes, including many immune-mediated diseases. In particular, that variants that are associated with disease in GWAS are enriched for expression quantitative trait loci (eQTLs)[177-181]. This indicates that a significant portion of the genetic risk for common, complex diseases is due to changes in transcript abundance and not to changes in protein coding genes. Therefore, identification of eQTLs in different tissues and in response to different treatments (or "environments") may identify otherwise difficult-to-identify GEIs, which contribute substantially to the risk of many immune-mediate diseases, such as asthma[69,182] and rheumatoid arthritis[183]. Today, dozens of studies have revealed thousands of cis-eQTLs and hundreds of response eQTLs at steady state or after treatment (e.g., lipopolysaccharide, IFNβ, bacteria, viruses) in various components of blood (immune) cells (e.g., lymphoblastoid cells lines [LCLs], peripheral blood mononuclear cells [PBMCs], dendritic cells, CD4+ T cells, Th17 cells)[184-187].

The relationship between complex diseases and eQTLs or response eQTLs in immune cells largely assumes a correlation between transcript abundance and protein abundance, which in turn leads to changes in cellular networks and pathways that results in disease phenotypes. However, correlations between mRNA and protein level is not perfect[188-190], presumably due to additional processes such as post-translational modification, protein folding, and variation in protein stability. Consequently, studies have been attempted in humans to identify protein QTLs (pQTLs) across the proteome using HapMap lymphoblastoid cell lines[191-194], which may have a more direct effect on disease risk. These studies identified a few hundred cis-pQTLs, but only approximately half of the cis-pQTLs were also eQTLs, indicating indeed that significant post-translational modifications perturb the relationship between transcript level and protein level.

These studies used different methods to measure protein levels, underscoring the present difficulty in measuring thousands of proteins in a cell with accuracy and precision at a reasonable cost, and hampering the ability to make comparisons across studies.

Despite the practical and financial limitations on proteome-wide QTL studies, they represent an important step forward in knitting together genotype-phenotype relationships. For specific diseases or phenotypes, equivalent insight might be gained by studying a smaller number of proteins with more direct biological relevance to the trait of interest. In the case of immune-mediated diseases, cytokines are a critical class of proteins with enormous effects on immune response and disease pathology, and secreted cytokines have reproducible and well established methods of measurement. For example, the Human Functional Genomics Project (HFGP), mapped variation in six cytokines (TNFα, IL-1β, IL-6, IFNγ, IL-17, IL-22) in response to 18 different stimuli in PBMCs, macrophages, and whole blood treated for 1, 2 or 7 days[195,196]. In a cohort of 500 individuals, Li *et al.* identified 17 cytokine QTLs (cQTLs) that were genome-wide significant ($P<5\times10^{-8}$), most of which were QTLs for IL-6 (7/17) or IL-1β (4/17) in PBMCs [196]. The 17 QTLs were significantly enriched in ENCODE monocyte-specific enhancers[197], and were detected in bacterial-, fungal-, and TLR ligand-stimulated cells. Additionally, cQTLs associated with levels of monocyte-derived cytokines (TNFα, IL-6, IL-1β) were enriched for significant associations with infections disease, while cQTLs associated with IL-22 and IFNγ were enriched for SNPs associated with autoimmune disease[196]. Outside of this initiative, identification of both cis and trans cytokine QTLs is still largely unexplored.

In an orthogonal approach to Li *et al*., we measured both gene expression and 17 cytokines in whole blood treated with a single innate immune stimulus, LPS, in 130 Hutterite individuals, a founder population of European descent. The communal practices of the Hutterites

leads to a relatively homogeneous lifestyle with regards to environmental exposure, and therefore may have greater power at a given study size than a randomly selected sample. To relieve multiple testing burden, we used a candidate gene approach to assess the effect of DNA variation near genes that code for these cytokines or their receptors. Our study revealed evidence of significant heritabilities for seven of the 17 cytokines investigated, and identified loci affecting levels of two cytokines, IL-12 and IL-27, after correcting for multiple testing by permutation. These data further support a role for genetic regulation of cytokine response.

4.3 Methods

4.3.1 Study population

This study was conducted in 153 Hutterites (7-76 years old) who are a subset of the >1,400 Hutterites who have participated in our population-based studies of complex phenotypes [131,167,168]. The Hutterites are an Anabaptist sect of central European ancestry. The Hutterites in our studies live on communal farms in South Dakota and are related to each other through multiple lines of descent in a single 3,671-member pedigree with 64 ancestors. Written consent for these studies was obtained from the adult participants and parents of children under 18; written assent was obtained from all children.

4.3.2 Collection of whole blood samples

One milliliter of whole blood was drawn into each of three TruCulture (Myriad RBM) tubes containing either proprietary TruCulture media alone or media + 0.1 μg/ml lipopolysaccharide (LPS). Samples were incubated upright in a dry heat block at 37C for 30 hours. Following incubation, tubes were processed in a makeshift lab in South Dakota. Briefly, supernatants were

separated from the cell pellets, which contained peripheral blood leukocytes (PBLs), and then aliquoted and flash-frozen on dry ice. The remaining samples were washed twice with Buffer EL (Qiagen) and the cell pellets were resuspended in 350 µl RLT Buffer (Qiagen) and frozen on dry ice. Frozen samples were shipped on dry ice overnight to Chicago and stored at -80^{o}C. DNA and RNA were extracted from thawed cell pellets using AllPrep DNA/RNA Mini Kits (Qiagen).

4.3.3 Cytokine measurement and processing

Supernatants from 130 LPS-treated samples were submitted to the Human Immunology Core at the University of Chicago for measurement of 25 cytokines using the Milliplex Th17 Magnetic Bead Panel (Millipore). Each sample was run in duplicate. Results from duplicate samples were averaged; in 10 cases each, one duplicate did not yield a detectable result and only a single measurement was retained. After removing cytokines with measurements outside the range of detection in >30% of individuals, 17 cytokines from LPS-treated samples were included in downstream analysis. Cytokine measurements were quantile normalized; and principal components analysis identified bead panel "plate" as a confounding technical covariate, which was adjusted using ComBat[137].

4.3.4 Genotyping and chip heritability estimate

Of the 130 individuals with measured cytokine levels, 107 were previously genotyped on Affymetrix platforms 6.0 and 5.0; the QC of those data was described previously[131,167,168]. The remaining 23 individuals were genotyped on Illumina CytoSNP Bead Chip and the Illumina Infinium HumanCore Exome Bead Chip. Variants with call rates <90%, Hardy-Weinberg equilibrium $P<10^{-5}$, or MAF <5% were removed; all individuals had SNP call rates of >95%.

Using these SNPs as framework markers, 9,914,265 variants (single nucleotide polymorphisms [SNPs] and short insertions/deletions [indels]) present in the genome sequences of 121 Hutterites were imputed to all 130 individuals using PRIMAL, an in-house pedigree-based imputation method with >99% accuracy in the Hutterites[132].

4.3.5 Cis-cytokine QTL analysis

Variants ± 250kb of the transcription start site of genes that encode the 17 cytokines, or their receptors were included in the association analyses. The full list of genes for each cytokine, and the number of variants tested, is shown in Supplementary Table 4.1. Associations between each variant and each cytokine level were calculated using GEMMA[169], including a kinship matrix based on identity by descent (IBD) to correct for the relatedness between Hutterites in our sample. In all analyses, age and sex were included as covariates. Chip heritability was estimated using GEMMA[169], from the proportion of variance in phenotypes explained by all 4,998,317 genotyped and imputed variants with a call rate >80% and MAF >5%. Significance was assessed using permuted phenotypes using the R package MVNpermute[198], which permutes phenotypes while retaining the covariance structure of the phenotype. Permutations are calculated on a linear transformation of residuals that is exchangeable with the original phenotypes when the data is multivariate normally distributed. Each cytokine was permuted 1000 times and association P-values were calculated using GEMMA. The most significant P-value from each permuted phenotype was retained to determine the empirical P-value for the most significant association for each cytokine.

4.3.6 Analysis of Gene expression

RNA from the TruCulture tubes was used to create RNA-seq libraries using the TruSeq Library kit (Illumina); quality and concentration of libraries were assessed with an Agilent 2100 Bioanalyzer (Agilent Technologies) and quantitative PCR using the Kapa library quantification kit (Kapa Biosystems). Samples were sequenced in pools of 16-18 samples across three flow cells of an Illumina HiSeq 2500; 119 samples with low read count were re-sequenced on two flow cells on the same machine. Adaptor sequences were trimmed, and then reads were mapped to hg19 using Tophat2[170]. Reads were sorted and indexed with samtools[171], and genes counted with HTseq [172]. Read counts for the same sample sequenced multiple times were summed together, and then Trimmed Means of M-values normalization (TMM) and a voom transformation was used to correct for differences in library sizes[173]. Confounding technical effects were assessed in the normalized expression data using principal components analysis (PCA), and sequencing pool was adjusted using the function RemoveBatchEffect() from the R package limma[174].

4.3.7 Trans cytokine QTL analysis

For the seven cytokines with an estimated non-zero chip based heritability (IL-12p70, IL-23, IL-22, MIP-3α, IFNγ, IL-33, and IL-10; Supplementary Table 4.2), we performed an exploratory genome-wide association study to identify candidate variants acting in trans, using the same methods as for cis-cQTL mapping described above. A P-value of $<5\times10^{-8}$ was used as the threshold for genome-wide significance.

4.4 Results

4.4.1 Results of local cytokine QTL associations

We conducted association studies of 17 cytokines in LPS-treated whole blood and genetic variants ± 250kb of the transcription start site (TSS) of each cytokine or cytokine receptor gene (Supplementary Table 4.1). Using a permutation approach to assess significance, we identified variants associated with IL-12p70 and IL-27 at a permuted genome-wide P-value <0.1 (Table 4.1). The most significant association was with a variant (rs6441271) and IL-12p70 levels (permuted P=0.021). This variant is located 230kb upstream of *IL12A*, a gene that encodes the 35kD subunit of IL-12 (Figure 4.1, Panel A). The number of copies (0, 1, 2) of the minor allele (MAF=0.398) was associated with increasing levels of IL-12p70 (Figure 4.1, Panel B). The *IL12A* transcript was not detected as expressed in the untreated or LPS-treated PBLs. We were therefore unable to test if rs6441271 is also an eQTL for this gene. While this locus resides in an area near bivalent enhancers and a bivalent transcription start site in primary PBMCs[199]; rs6441271 or any SNPs in strong LD with it (r^2>0.8) have not been previously associated with any traits or diseases at genome-wide significance ($P<5\times10^{-8}$,[200]), or as eQTLs in other tissues[201].

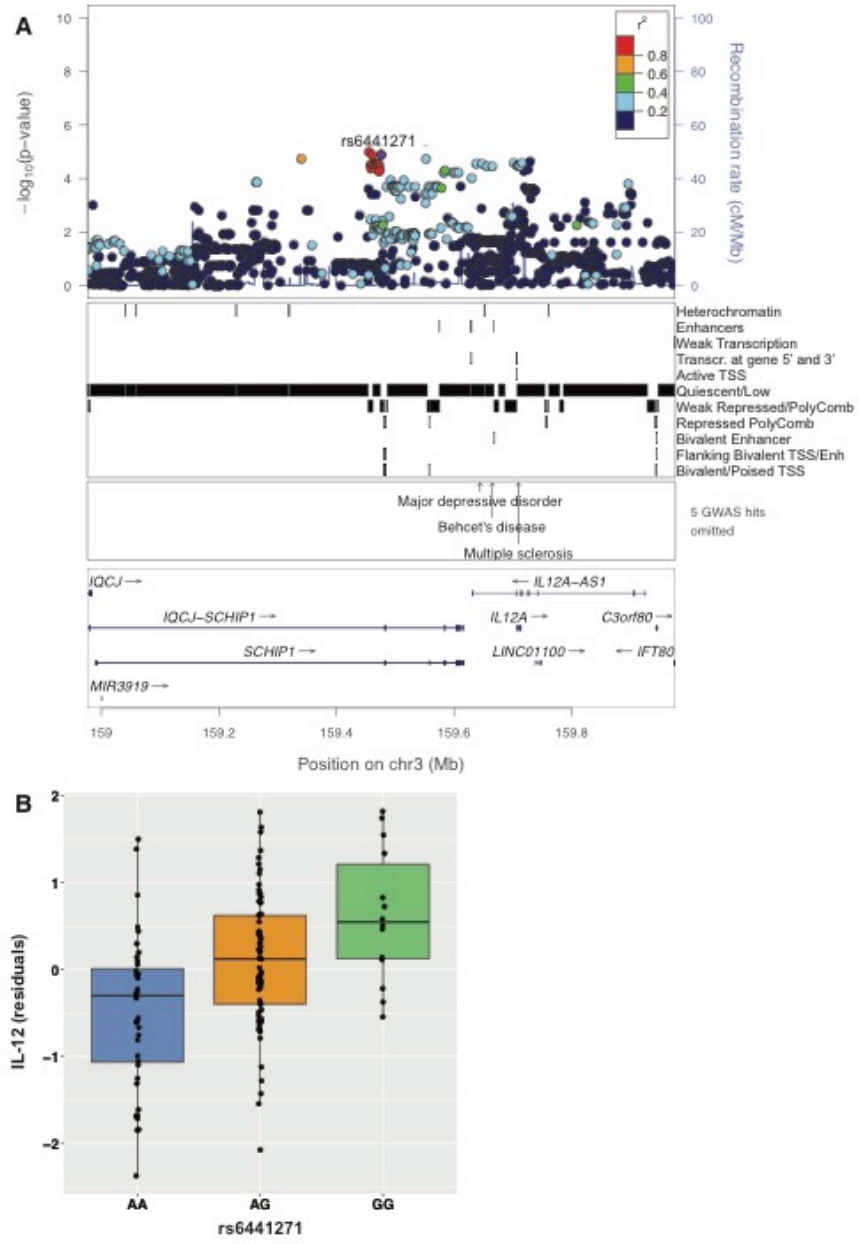
Table 4.1. Summary of association results for each cytokine.

The most significant association for each cytokine is shown by rsID, chromosome position, and nearest gene. The beta is presented for the minor allele.

Cyto-kine	rsID	Chr:loc	Nearest gene	Alleles	MAF	Beta	P-value	Genome-wide permuted P-value
IL-12 p70	rs6441271	3:159476003	*IQCJ-SCHIP1, SCHIP1*	A/G	0.398	0.561	1.28×10^{-5}	0.021
IL-27	rs56116604	19:4193497	*ANKRD4*	-/AAG	0.465	-0.279	3.17×10^{-4}	0.076
IL-23	rs2815376	1:67508531	*SLC35D1*	A/G	0.480	-0.297	2.30×10^{-4}	0.162
IL-15	rs10014610	4:142724101	*IL15*	G/A	0.209	-0.290	5.84×10^{-4}	0.195
IL-22	rs11177183	12:68753921	*MDM1*	G/A	0.223	0.250	4.61×10^{-4}	0.270
MIP-3α	rs73994542	2:228691267	*CCL20*	C/T	0.200	-0.330	8.41×10^{-4}	0.278
IFNγ	rs276556	6:137417649	*IL22RA2*	T/G	0.392	0.436	7.04×10^{-4}	0.309
TNFα	rs9266329	6:31330788	*HLA-B*	G/A	0.357	0.353	7.99×10^{-4}	0.309
IL-25	rs9606606	22:17568213	*IL17RA*	C/T	0.134	0.351	8.60×10^{-4}	0.362
IL-4	rs35630199	16:27344199	*IL4R*	T/G	0.289	0.245	5.16×10^{-3}	0.539
IL-13	rs60549201	16:27186970	*KDM8*	G/A	0.479	0.222	7.45×10^{-3}	0.623
IL-33	rs10188022	2:102874866	*IL1RL2*	A/T	0.404	0.142	4.44×10^{-3}	0.691
IL-21	rs10518402	4:123554790	*IL21-AS1*	T/C	0.285	-0.271	7.69×10^{-3}	0.693
IL-17F	rs2241044	22:9779068	*IL17RA*	A/C	0.434	0.273	6.11×10^{-3}	0.768
IL-5	rs2022068	22:37199806	*PVALB*	G/A	0.305	0.192	6.67×10^{-3}	0.856
IL-10	rs7276112	21:34816479	*TMEM50B*	G/A	0.426	0.317	5.20×10^{-3}	0.874
IL-9	rs367950266	5:135340288	*TGFBI*	ACTCAGCTTAAAGTGT/-	0.054	-0.295	0.088	0.959

Figure 4.1. Association of rs6441271 with levels of LPS-treated IL-12p70.

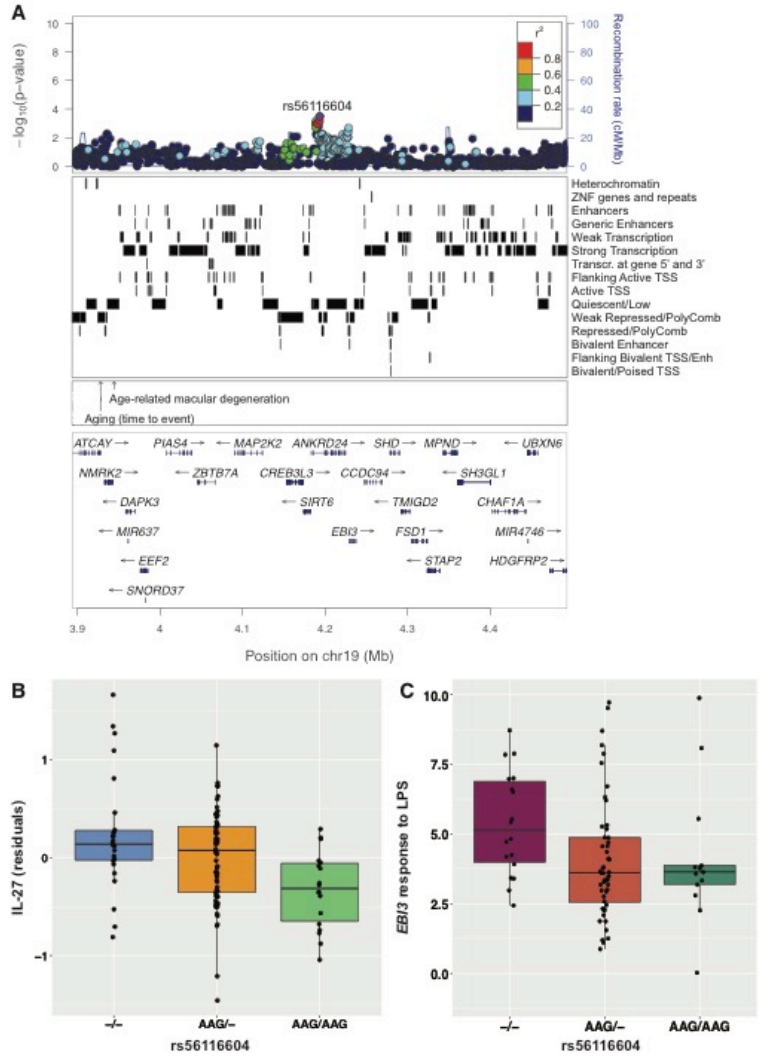
(A) Locus zoom plot of rs6441271 ($P=1.28\times10^{-5}$, permuted P=0.021), located upstream of the *IL12A* gene, a subunit of the IL-12 cytokine. rs6441271 indicated by a purple diamond, different colors represent the level of linkage disequilibrium between each SNP and rs6441271. The blue line represents the recombination rate in the Hutterites in this genomic location. The second track represents chromatin states from primary peripheral mononuclear blood cells from the 15-model Epigenome Roadmap, and the third track represents genome-wide significant associations included in the NHGRI catalog. Omitted hits shown in Supplementary Table 4.2. (B) Boxplot of rs6441271 genotype by IL-12 cytokine level.



We also observed a suggestive association (P=0.076) on chromosome 19 with LPS-treated IL-27 (Figure 4.2, Panel A). Located 36kb upstream of the IL-27 subunit *EBI3*, the minor allele of rs56116604 is a three-base pair insertion that is correlated with decreased levels of LPS-treated IL-27 (Figure 4.2, Panel B, MAF=0.465). While rs56116604 and SNPs in high LD (r^2>0.8) with this indel are not significantly associated with traits in the GWAS catalog, *EBI3* gene expression response to LPS treatment is modestly correlated with rs56116604 (P=0.0401, Figure 4.2, Panel C) in the same direction as the LPS-treated IL-27 cytokine levels. This suggests that this indel is both a cQTL and response eQTL, and that the effect of this SNP on LPS-treated IL-27 may act through changes in gene expression.

Figure 4.2 Association of rs56116604 with levels of LPS-treated IL-27.

(A) Locus zoom plot of rs56116604 ($P=3.17\times10^{-4}$, permuted P=0.076), located upstream of the *EBI3* gene, a subunit of IL-27. rs56116604 indicated by a purple diamond, different colors represent the level of linkage disequilibrium between each SNP and rs56116604. The blue line represents the recombination rate in the Hutterites in this genomic location. The second track represents chromatin states from primary peripheral mononuclear blood cells from the 15-model Epigenome Roadmap, and the third track represents genome-wide significant associations included in the NHGRI catalog. Omitted hits shown in Supplementary Table 4.2. (B) Boxplot of rs56116604 genotype by IL-27 cytokine level. (C) Gene expression of *EBI3* in response to LPS treatment (LPS/untreated, P=0.04).



4.4.2 Estimate of chip-based heritability and genome-wide cytokine QTL associations

Using the SNP-based kinship coefficients in our sample, we were able to estimate the proportion of variance in phenotype explained by genotypes ('chip' heritability, or h^2) for all 17 cytokines tested. Seven of the 17 cytokines (IL-12p70, IL-23, IL-22, MIP-3α, IFNγ, IL-33, and IL-10) had evidence of non-zero h^2 after accounting for the standard error of the estimate (Supplementary Table 4.3). Although these estimates were generated from small samples sizes, a non-zero h^2 suggests that additional genetic variation outside of cytokine and cytokine receptor genes contribute to variance in LPS-treated cytokine level. To address this question, we performed a GWAS to identify variants acting on cytokine production in trans for the 7 cytokines with evidence of non-zero h^2. For each cytokine, the most significant result is listed in Supplementary Table 4.4. No variants reached genome-wide significance ($P<5\times10^{-8}$) for any cytokine (Supplementary Figure 4.1).

4.5 Discussion

In this study, we assessed the effect of common genetic variation on cytokine response to LPS treatment in and around cytokine and cytokine receptor genes. We hypothesized that DNA variation proximal to the cytokine genes or to their receptors, which frequently play a role in cytokine regulation, would have the largest and most easily observable effect on cytokine variation. Across 17 LPS-treated cytokines in 130 Hutterite individuals, we identified loci suggestively associated with LPS-treated IL-12 and IL-27. IL-12 cytokine levels were associated with a SNP (rs6441271), located upstream of *IL12A*, which encodes a cytokine subunit unique to IL-12 and therefore represents an excellent target to specifically regulate IL-12. Because *IL12A* was not detected as expressed in PBLs after 30 hours of treatment, it is unknown whether this

locus is also acting at the transcriptional level as an eQTL. In contrast, LPS-treated IL-27 levels were associated with an indel (rs56116604), which was also a modest eQTL for *EBI3* in LPS-treated PBLs, suggesting that this indel affects IL-27 cytokine levels through a transcriptional mechanism. However, further study is required to identify the causal SNP and possible mechanism at each locus. Of the 17 cytokines studied, seven had evidence of non-zero heritability across genome-wide genotyped SNPs. An exploratory genome-wide association study for these seven LPS-treated cytokines did not reveal any significant associations; however, a better-powered study might be able to find evidence that reach genome-wide significance.

Our sample was approximately one-fourth of the 500 person discovery panel employed in the first major study of cQTLs following immune stimulation[196]. This small sample size may have limited our power to identify variants associated with cytokine levels, especially at a genome-wide levels of significance. However, because the Hutterites have a relatively homogenous lifestyle and environmental exposures compared to the general population, we assumed that a smaller sample of Hutterites could provide equivalent statistical power to find associations compared to a larger, randomly-selected sample. While we still can not rule out this possibility, the lack of associations in our study likely reflects some limitation of our design.

Notably however, Li *et al.* identified only one cytokine QTL in whole blood – a locus on chromosome 19 that was associated with IL-6 after a 48-hour incubation with *C. albicans*[196]. The other 16 associations were identified in PBMCs (n=500). These results, together with the findings of our study, suggest that whole blood may be a poor choice of tissue in which to identify genetic factors influencing cytokine responses. The reasons behind this observation may be due, in part, to the diversity of cell types found in whole blood, and the inter-individual variation in cellular composition and activation state of peripheral blood leukocytes. Importantly,

whole blood cells are composed largely (>50%) of neutrophils, whereas these innate immune cells are not included in the PBMC fraction of blood cells.

The cytokine QTL study presented here is the first to study genetic variation underlying LPS-treated cytokine levels for 13 cytokines, including IL-12, IL-27, and MIP-3α. With a relatively small sample size, we were able to identify three loci that are suggestively associated with levels of IL-12 and IL-27, and potentially identifying novel candidates of cytokine regulation. Due to the broad array and functional importance of cytokines in immune response, cytokine QTL studies may still be a powerful new avenue for understanding the genetic architecture of immune-mediated disease in specific cell types and in large samples of individuals.

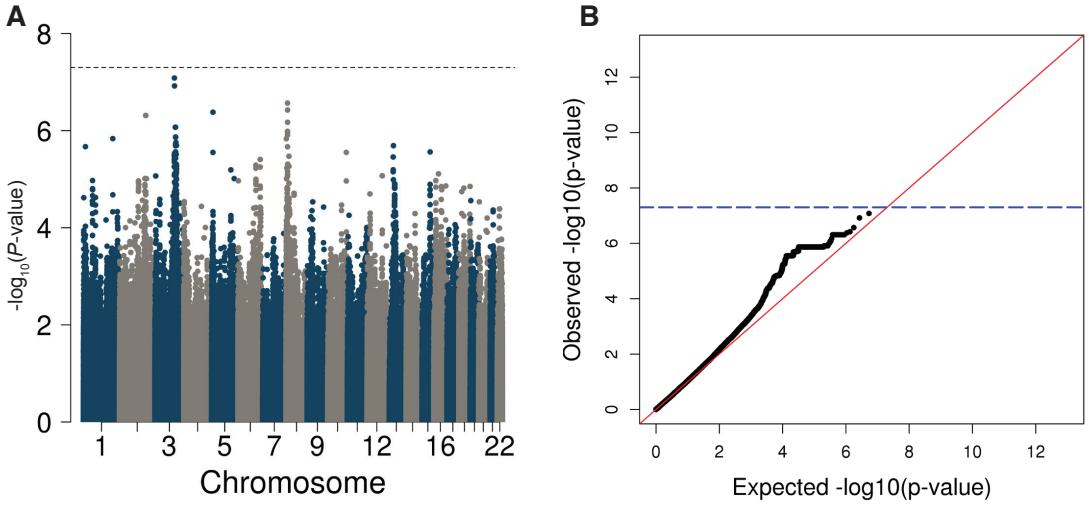
4.6 Acknowledgements

The authors would like to acknowledge the efforts of Christine Billstrand and Raluca Nicolae in generating RNAseq sample libraries, Tiratat Patana-anake for the Hutterite imputation genotype data; Catherine Igartua and Mark Abney for statistical advice; the entire South Dakota field studies team (Gorka Alkorta-Aranburu, Rebecca Anderson, Maitane Arrubarrena Orbegozo, Kathleen Bailey, Courtney Burrows, Dan Cook, Catherine Igartua, Rob Stanaker, and Emma Thompson), and the Hutterite volunteers.

4.7 Appendix E: Supplementary Figure and Tables

Supplemental Figure 4.1. Results of genome-wide association with LPS-treated MIP-3α.

(A) Manhattan plot for LPS-treated MIP-3α. Chromosomal position of each variant indicated along the x-axis, significance of association represented along the y-axis as -$\log_{10}$Pvalue. Dashed line shows genome wide significant P=5×10^{-8}. (B) QQ plot for LPS-treated MIP-3α associations. Dashed blue line shows genome wide significant P=5×10^{-8}.



Supplemental Table 4.1 Genes and number of variants tested for each LPS-treated cytokine.

To get an estimate of how many independent loci are tested in each cytokine, genotypes were pruned for LD (r^2<0.5). Genes located on the X chromosome were not included in the study.

Cytokine	Genes	Number variants	Number of independent regions (with r^2<0.05)
IL-12p70	*IL12A, IL12B, IL12RB1, IL12RB2*	4000	262
IL-27	*IL27, EBI3*	1352	110
IL-23	*IL23A, IL12B, IL23R, IL12RB1*	3733	235
IL-15	*IL15, IL15RA, IL2RB*	3062	236
IL-22	*IL22, IL22RA1, IL10RB*	3354	224
MIP-3α	*CCL20, CCR6*	2803	152
IFNγ	*IFNG, IFNGR1, IFNGR2*	3170	200
TNFα	*TNF, TNFRSF1A, TNFRSF1B*	4114	203
IL-25	*IL25, IL17RA, IL17RB*	3015	274
IL-4	*IL4, IL4R*	1767	132
IL-13	*IL13, IL4R*	1759	134
IL-33	*IL33, IL1RL1, IL1RAP*	3504	134
IL-21	*IL21, IL21R*	1400	94
IL-17F	*IL17F, IL17RA, IL17RC*	3334	241
IL-5	*IL5, IL5RA, CSF2RB*	3767	288
IL-10	*IL10, IL10RA, IL10RB*	3196	259
IL-9	*IL9*	1219	53

Supplemental Table 4.2 Genome-wide significant associations included in the NHGRI catalog from the locus zoom plot track in Figures 4.1 and 4.2.

Table includes both omitted and included associated traits.

Chr	Pos (Mb)	Trait	SNP
IL-12			
3	159.2953	Quantitative traits	rs2222328
3	159.6443	Major depressive disorder	rs6799788
3	159.6651	Behet's disease	rs17810546
3	159.6551	Celiac disease	rs17810546
3	159.6989	Multiple sclerosis	rs4680534
3	159.7097	Multiple sclerosis	rs2243123
3	159.7289	Primary biliary cirrhosis	rs6441286
3	159.7459	Primary biliary cirrhosis	rs485499
IL-27			
19	3.797100	Obesity−related traits	rs12104221
19	3.927771	Aging (time to event)	rs10412199
19	3.944240	Age−related macular degeneration	rs10406174

Supplemental Table 4.3. Estimates of chip-based heritability of cytokine response to LPS.

Heritability calculated using a pedigree based on Affymetrix and Illumina genotyping of the Hutterites (see methods), with age and sex included as covariates. Heritability was calculated in GEMMA, denoted as the proportion of variance (pve) explained.

Cytokine	pve estimate	pve standard error
IL-12p70	0.371	0.214
IL-27	0.040	0.138
IL-23	0.259	0.199
IL-15	0.038	0.118
IL-22	0.440	0.189
MIP-3α	0.313	0.176
IFNγ	0.284	0.160
TNFα	0.157	0.166
IL-25	0.108	0.125
IL-4	9.18e-06	0.136
IL-13	0.007	0.152
IL-33	0.308	0.180
IL-21	0.169	0.169
IL-17F	9.18e-06	0.159
IL-5	0.065	0.109
IL-10	0.351	0.176
IL-9	0.029	0.137

Supplemental Table 4.4. Summary of the most significant genome-wide association results.

Results shown for the 7 cytokines with a non-zero chip-based heritability (Supplementary Table 4.2). The beta is presented for the minor allele. No associations reach genome-wide significance ($P=5\times10^{-8}$).

Cytokine	rsID	Chr:loc	Nearest gene	Alleles	MAF	Beta	P-value
IL-12 p70	rs6034497	20:16464676	*KIF16B*	C/T	0.196	0.677	3.03×10^{-7}
IL-23	rs568198249	4:175566524-43	*GLRA3*	CTGCACTCCAGCCCTCCAGC/-	0.092	-0.750	1.38×10^{-7}
IL-22	rs71540205	6:48285624-5	*PTCHD4*	CA/TG	0.386	0.367	1.62×10^{-7}
MIP-3α	rs558667825	3:133681303-5	*SLCO2A1*	AGG/-	0.034	-0.634	8.25×10^{-8}
IFNγ	rs77351266	4:7540278	*SORCS2*	C/A	0.067	-1.28	5.49×10^{-7}
IL-33	rs9406745	9:18365057	*ADAMTSL1*	A/G	0.488	0.235	1.80×10^{-7}
IL-10	rs613718	22:26800409	*SEZ6L*	G/A	0.265	0.577	1.66×10^{-7}

Chapter 5

Conclusions

5.1 Summary and Significance

In common, complex diseases like asthma, genetic risk does not exist in a vacuum independent of environmental exposures. Instead, these exposures may act to exacerbate or ameliorate overall risk, or interact with genetic risk variants to contribute to the overall likelihood of developing disease. The studies included in this dissertation were motivated by a desire to understand how environmental contexts, from the cellular to population level, affect asthma or immune responses. Chapter 2 details how differences in the farming environment in Amish and Hutterite children contribute to differences in the prevalence of childhood asthma and allergic sensitization. Chapters 3 and 4 characterize and identify genetic variation that mediates host response to environmental stimuli.

The work described in Chapter 2 encompasses the results of a collaboration with the Sperling lab at the University of Chicago and the Vercelli lab at the University of Arizona to characterize the immune profiles and environments of Amish and Hutterite schoolchildren. This work began with an appreciation that the Amish and Hutterites, two US founder populations of European descent, had very similar lifestyles with respect to asthma risk, but only the Amish had a low prevalence of asthma and allergic sensitization. In contrast, the Hutterites were a farming population with relatively high rates of asthma and allergic sensitization. A comparison of these two populations therefore affords a unique opportunity to dissect the protections of the farm effect. I was able to determine that across common DNA variants, the Amish and Hutterite schoolchildren in our study shared remarkable genetic similarity, underscoring that the observed differences in asthma prevalence were likely driven by environmental exposures. The

population-specific differences in exposures could be recapitulated using Amish and Hutterite home dust extract in a mouse model of allergic asthma. Mice treated with Amish house dust extract were protected against airway hyperresponsiveness and eosinophilia, while mice treated with Hutterite house dust extract were not. Using MyD88/Trif knockout mice, we were able to show that innate immunity is required for protection against airway hyperresponsiveness. This farm protection in the Amish and lack thereof in the Hutterites translates to stark differences in the peripheral immune system by school age: vast differences peripheral blood leukocyte composition drove large differences in baseline gene expression profiles; increased regulatory and immature cell surface markers in Amish monocytes and neutrophils, respectively; and a dampened cytokine response to stimulation with LPS in Amish schoolchildren. The results of this study were published in *The New England Journal of Medicine*.

In Chapter 3, I reported on the effects of a single clinically relevant polymorphism, rs1801274, on cytokine response and gene expression profiles following treatment with anti-CD3 and anti-CD28 mouse IgG1 antibodies. This variant, located in the gene that encodes the FcγIIA receptor, results in a missense mutation that modifies binding affinity to human IgG2 and mouse IgG1 antibodies. While rs1801274 previously has been associated as a risk factor for autoimmune disease, and studied in the context of anti-cancer therapeutic antibody efficacy, little is known about rs1801274 genotype effects on global gene expression and variation in cytokine responses across all three genotype classes. Following stimulation with anti-CD3 and anti-CD28 antibodies, I observed a striking lack of immune activation in low-binding affinity homozygotes (131His/131His) in global gene expression and cytokine levels. More notably, I recorded hitherto unappreciated differences between heterozygotes and high affinity (131Arg/131Arg)

homozygotes in cytokine levels and gene expression profiles, suggesting that downstream immune responses to IgG vary by each genotype class.

In Chapter 4, I describe the results of an LPS-treated cytokine QTL study using a candidate gene approach. Due to the relatively small sample size of 130 Hutterite individuals with measured cytokine levels after 30 hours of treatment with LPS, I employed a candidate gene approach, testing variation in and around cytokine genes and cytokine receptor genes. I identified one significant ($P<0.05$) and one suggestive loci ($P<0.1$) that correlated with IL-12 and IL-27 levels, respectively. The lack of significant results points not only to low statistical power due to a small sample size, but also the use of whole blood as a tissue choice. A previous study of cytokine QTLs uncovered over a dozen genome-wide significant QTLs in peripheral blood mononuclear cells (PBMCs), but only one in whole blood[196], supporting the conclusion that whole blood may be too heterogeneous to reliably identify DNA variation that affects cytokine response.

5.2 Future Directions

5.2.1 Innate immunity and asthma risk in Amish and Hutterite children

The results of this study open an array of new research avenues to pursue, in order to better understand the mechanisms of the development of protection and/or risk of asthma. One logical next step is a more rigorous study of the composition of Amish house dust. Recent experiments conducted by the Vercelli lab have shown that treatment with either Amish or Hutterite barn dust confers protection against airway hyperresponsivnes and eosinophilia[202], suggesting that high exposures to animal or microbial products is behind this protection. Assessment of the dust composition by 16S rRNA sequencing and characterization of metabolites using mass

spectrometry is required in order to begin testing specific components in a mouse model, to move towards the long-term goal of identifying a reproducible "environmental cocktail" that could induce the same protection from asthma in a mouse model, and later, in a controlled clinical trial in children.

One of the major limitations of our study was that gene expression profiles were measured in peripheral blood leukocytes, a heterogeneous cell population. While we could observe gene expression differences between individuals, we could not observe expression differences on a per cell basis. Since we were able to observe striking differences in cell surface markers like CXCR4 and HLA-DR on neutrophils and monocytes, respectively, between Amish and Hutterite schoolchildren, it is likely the underlying gene expression profiles within each cell differ considerably. Single-cell RNAseq (scRNAseq) would allow for better characterization of peripheral blood leukocytes and help to pinpoint specific cell types that may be potential targets in innate immunity-mediated protection from asthma and allergic disease. Additionally, scRNAseq of airway immune and epithelial cells in mice treated with Amish house dust extract will be extremely informative in generating functional hypotheses to unravel the mechanisms of protection.

ScRNAseq of Amish and Hutterite PBLs could also provide greater understanding of a previously observed gene by environment interaction of 17q12-21 asthma-associated SNPs and barn exposure[122]. In that study, individuals with the asthma risk allele in the 17q21 region (rs8076131, A allele) and who spent more than two hours/week in animal barns had the lowest asthma risk— suggesting there is a flip-flop effect of the 17q21 region depending on exposure to animal barns. Consequently, Amish children, who are exposed to barns, and Hutterite children, who are not, are an excellent comparison in which to further investigate this finding. Though the

mechanism of risk or protection is not yet known, the 17q12-21 variants, including rs8076131 and rs7216389, are eQTLs for the nearby gene *ORMDL3*, a gene highly expressed in B cells, eosinophils, and CD8+ T cells[203-205]. Testing for rs7216389 effect on *ORMDL3* expression in specific cells from Amish and Hutterite children might provide additional insight into this 17q12-21 by barn exposure interaction effect.

5.2.2 Profound effects of FcγIIA polymorphism (Arg131His) on blood leukocyte global gene expression and cytokine responses to treatment with anti-CD3 and anti-CD28 IgG1

A critical next step in characterizing the effect of this variant is to confirm that the same effect on rs1801274 heterozygotes is observed after treatment with human IgG2 antibodies. In this case, the A allele (131His) has higher affinity with human IgG2, a reverse of the binding pattern with mouse IgG1. While we can infer that there would be a similar effect following treatment with IgG2, differences in the binding affinities due to variation in the structure of human IgG2 versus mouse IgG1 Fc may result in an altered response. Quantification of response to human IgG2, in addition to the response to mouse IgG1 detailed in Chapter 3, will be crucial for design of future experiments or drug trials that use soluble antibodies in humans.

5.2.3 A cytokine QTL study in LPS-treated whole blood in a founder population

I began this study to identify variants near cytokine subunit coding genes and cytokine receptor genes that contribute to variation in cytokine response to innate immune stimulation with LPS. In the future, the results of this study should be replicated in a more restricted set of cell types, such as PBMCs, or macrophages, where previous studies have identified genome-wide associations with treated cytokines [196]. Future studies should also designate an earlier time point, such as 8

hours of treatment, to sample gene expression, and a later time point, such as 24 or 48 hours, to assay cytokine production. Assignment of each time point may depend on the type of stimulus used.

5.3 Concluding remarks

The studies presented in this dissertation contribute to the rapidly growing literature investigating the role of the environment, both as the macro- and micro-scale, in immune profiles in asthma. In Chapter 2, we were able to leverage a unique comparison of two very similar populations to gain insight into the protective effect of the immune response. In Chapter 3, a focused study of how one polymorphism modifies gene expression and cytokine response to anti-CD3+anti-CD28 antibodies in peripheral blood leukocytes revealed previously unappreciated differences in immune response by genotype class. In Chapter 4, I looked for associations between genetic variation around cytokine and cytokine receptor genes and cytokines from LPS-treated PBLs, and gained an appreciation for the pitfalls of working with heterogeneous cell populations, like PBLs. Overall, these studies illustrate how investigation into the different levels of environment can yield biological insight.

BIBLIOGRAPHY

1 Natarajan, P. *et al.* Multiethnic Exome-Wide Association Study of Subclinical Atherosclerosis. *Circulation. Cardiovascular genetics* **9**, 511-520, doi:10.1161/circgenetics.116.001572 (2016).

2 Mahajan, A. *et al.* Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility. *Nat Genet* **46**, 234-244, doi:10.1038/ng.2897 (2014).

3 Ripke, S. Genome-wide Association Analysis Identifies 14 New Risk Loci for. **45**, 1150-1159, doi:10.1038/ng.2742 (2013).

4 Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747-753, doi:10.1038/nature08494 (2009).

5 Lemanske, R. F., Jr. & Busse, W. W. Asthma: clinical expression and molecular mechanisms. *J Allergy Clin Immunol* **125**, S95-102, doi:10.1016/j.jaci.2009.10.047 (2010).

6 Expert Panel Report 3 (EPR-3): Guidelines for the Diagnosis and Management of Asthma-Summary Report 2007. *J Allergy Clin Immunol* **120**, S94-138, doi:10.1016/j.jaci.2007.09.043 (2007).

7 Martinez, F. D. & Vercelli, D. Asthma. *Lancet* **382**, 1360-1372, doi:10.1016/s0140-6736(13)61536-6 (2013).

8 Jackson, D. J. *et al.* Wheezing rhinovirus illnesses in early life predict asthma development in high-risk children. *American journal of respiratory and critical care medicine* **178**, 667-672, doi:10.1164/rccm.200802-309OC (2008).

9 Kusel, M. M. *et al.* Early-life respiratory viral infections, atopic sensitization, and risk of subsequent development of persistent asthma. *J Allergy Clin Immunol* **119**, 1105-1110, doi:10.1016/j.jaci.2006.12.669 (2007).

10 Johnston, S. L. *et al.* Community study of role of viral infections in exacerbations of asthma in 9-11 year old children. *BMJ (Clinical research ed.)* **310**, 1225-1229 (1995).

11 Li, N., Hao, M., Phalen, R. F., Hinds, W. C. & Nel, A. E. Particulate air pollutants and asthma. A paradigm for the role of oxidative stress in PM-induced adverse health effects. *Clinical immunology (Orlando, Fla.)* **109**, 250-265 (2003).

12 Bernstein, J. A. *et al.* Health effects of air pollution. *J Allergy Clin Immunol* **114**, 1116-1123, doi:10.1016/j.jaci.2004.08.030 (2004).

13 Samoli, E., Nastos, P. T., Paliatsos, A. G., Katsouyanni, K. & Priftis, K. N. Acute effects of air pollution on pediatric asthma exacerbation: evidence of association and effect modification. *Environmental research* **111**, 418-424, doi:10.1016/j.envres.2011.01.014 (2011).

14 Robays, L. J., Maes, T., Joos, G. F. & Vermaelen, K. Y. Between a cough and a wheeze: dendritic cells at the nexus of tobacco smoke-induced allergic airway sensitization. *Mucosal immunology* **2**, 206-219, doi:10.1038/mi.2009.7 (2009).

15 Thacher, J. D. *et al.* Pre- and postnatal exposure to parental smoking and allergic disease through adolescence. *Pediatrics* **134**, 428-434, doi:10.1542/peds.2014-0427 (2014).

16 Mitchell, E. A., Beasley, R., Keil, U., Montefort, S. & Odhiambo, J. The association between tobacco and the risk of asthma, rhinoconjunctivitis and eczema in children and adolescents: analyses from Phase Three of the ISAAC programme. *Thorax* **67**, 941-949, doi:10.1136/thoraxjnl-2011-200901 (2012).

17 Carlsen, K. H. & Carlsen, K. C. Exercise-induced asthma. *Paediatric respiratory reviews* **3**, 154-160 (2002).

18 Jones, R. S., Buston, M. H. & Wharton, M. J. The effect of exercise on ventilatory function in the child with asthma. *British journal of diseases of the chest* **56**, 78-86 (1962).

19 Randolph, C. Exercise-induced bronchospasm in children. *Clinical reviews in allergy & immunology* **34**, 205-216, doi:10.1007/s12016-007-8035-0 (2008).

20 Bonini, M. & Palange, P. Exercise-induced bronchoconstriction: new evidence in pathogenesis, diagnosis and treatment. *Asthma research and practice* **1**, 2, doi:10.1186/s40733-015-0004-4 (2015).

21 Masoli, M., Fabian, D., Holt, S. & Beasley, R. The global burden of asthma: executive summary of the GINA Dissemination Committee report. *Allergy* **59**, 469-478, doi:10.1111/j.1398-9995.2004.00526.x
ALL526 [pii] (2004).

22 World Health Organization. *Global surveillance, prevention and control of chronic respiratory diseases*, <http://www.who.int/gard/publications/GARD Book 2007.pdf?ua=1> (2007).

23 Lemanske, R. F., Jr. Asthma therapies revisited: what have we learned? *Proceedings of the American Thoracic Society* **6**, 312-315, doi:10.1513/pats.200806-055RM (2009).

24 Wenzel, S. E. Asthma: defining of the persistent adult phenotypes. *Lancet* **368**, 804-813, doi:10.1016/s0140-6736(06)69290-8 (2006).

25 Centers for Disease Control and Prevention. *National Current Asthma Prevalence (2015)*, <https://www.cdc.gov/asthma/most_recent_data.htm> (2017).

26 Bentley, A. M. *et al.* Activated T-lymphocytes and eosinophils in the bronchial mucosa in isocyanate-induced asthma. *J Allergy Clin Immunol* **89**, 821-829 (1992).

27 Bentley, A. M. *et al.* Increases in activated T lymphocytes, eosinophils, and cytokine mRNA expression for interleukin-5 and granulocyte/macrophage colony-stimulating factor in bronchial biopsies after allergen inhalation challenge in atopic asthmatics. *American journal of respiratory cell and molecular biology* **8**, 35-42, doi:10.1165/ajrcmb/8.1.35 (1993).

28 Bradley, B. L. *et al.* Eosinophils, T-lymphocytes, mast cells, neutrophils, and macrophages in bronchial biopsy specimens from atopic subjects with asthma: comparison with biopsy specimens from atopic subjects without asthma and normal control subjects and relationship to bronchial hyperresponsiveness. *J Allergy Clin Immunol* **88**, 661-674 (1991).

29 Huang, S. K. *et al.* IL-13 expression at the sites of allergen challenge in patients with asthma. *Journal of immunology* **155**, 2688-2694 (1995).

30 Robinson, D. S. *et al.* Predominant TH2-like bronchoalveolar T-lymphocyte population in atopic asthma. *N Engl J Med* **326**, 298-304, doi:10.1056/nejm199201303260504 (1992).

31 Lee, J. J. *et al.* Interleukin-5 expression in the lung epithelium of transgenic mice leads to pulmonary changes pathognomonic of asthma. *The Journal of experimental medicine* **185**, 2143-2156 (1997).

32 Zhu, Z. *et al.* Pulmonary expression of interleukin-13 causes inflammation, mucus hypersecretion, subepithelial fibrosis, physiologic abnormalities, and eotaxin production. *The Journal of clinical investigation* **103**, 779-788, doi:10.1172/jci5909 (1999).

33 Saglani, S. & Lloyd, C. M. Novel concepts in airway inflammation and remodelling in asthma. *The European respiratory journal* **46**, 1796-1804, doi:10.1183/13993003.01196-2014 (2015).

34 Tschumperlin, D. J. *et al.* Mechanotransduction through growth-factor shedding into the extracellular space. *Nature* **429**, 83-86, doi:10.1038/nature02543 (2004).

35 Locksley, R. M. Asthma and allergic inflammation. *Cell* **140**, 777-783, doi:10.1016/j.cell.2010.03.004 (2010).

36 Moro, K. *et al.* Innate production of T(H)2 cytokines by adipose tissue-associated c-Kit(+)Sca-1(+) lymphoid cells. *Nature* **463**, 540-544, doi:10.1038/nature08636 (2010).

37 Licona-Limon, P., Kim, L. K., Palm, N. W. & Flavell, R. A. TH2, allergy and group 2 innate lymphoid cells. *Nature immunology* **14**, 536-542, doi:10.1038/ni.2617 (2013).

38 Kim, H. Y., DeKruyff, R. H. & Umetsu, D. T. The many paths to asthma: phenotype shaped by innate and adaptive immunity. *Nature immunology* **11**, 577-584, doi:10.1038/ni.1892 (2010).

39 Ober, C. & Vercelli, D. Gene-environment interactions in human disease: nuisance or opportunity? *Trends in genetics : TIG* **27**, 107-115, doi:10.1016/j.tig.2010.12.004 (2011).

40 Duffy, D. L., Martin, N. G., Battistutta, D., Hopper, J. L. & Mathews, J. D. Genetics of asthma and hay fever in Australian twins. *The American review of respiratory disease* **142**, 1351-1358, doi:10.1164/ajrccm/142.6_Pt_1.1351 (1990).

41 Nieminen, M. M., Kaprio, J. & Koskenvuo, M. A population-based study of bronchial asthma in adult twin pairs. *Chest* **100**, 70-75 (1991).

42 Polderman, T. J. *et al.* Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet* **47**, 702-709, doi:10.1038/ng.3285 (2015).

43 Ober, C. & Hoffjan, S. Asthma genetics 2006: the long and winding road to gene discovery. *Genes and immunity* **7**, 95-100, doi:10.1038/sj.gene.6364284 (2006).

44 Vercelli, D. Discovering susceptibility genes for asthma and allergy. *Nat Rev Immunol* **8**, 169-182, doi:10.1038/nri2257 (2008).

45 Nicolae, D. *et al.* Fine mapping and positional candidate studies identify HLA-G as an asthma susceptibility gene on chromosome 6p21. *American Journal of Human Genetics* **76**, 349-357, doi:10.1086/427763 (2005).

46 Cookson, W. & Moffatt, M. Making sense of asthma genes. *N Engl J Med* **351**, 1794-1796, doi:10.1056/NEJMe048232 (2004).

47 Wills-Karp, M. & Ewart, S. L. Time to draw breath: asthma-susceptibility genes are identified. *Nature reviews. Genetics* **5**, 376-387, doi:10.1038/nrg1326 (2004).

48 Moffatt, M. F. *et al.* A large-scale, consortium-based genomewide association study of asthma. *N Engl J Med* **363**, 1211-1221, doi:10.1056/NEJMoa0906312 (2010).

49 Torgerson, D. G. *et al.* Meta-analysis of genome-wide association studies of asthma in ethnically diverse North American populations. *Nat Genet* **43**, 887-892, doi:10.1038/ng.888 (2011).

50 Igartua, C. *et al.* Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. *Nature communications* **6**, 5965, doi:10.1038/ncomms6965 (2015).

51 Liang, L. *et al.* An epigenome-wide association study of total serum immunoglobulin E concentration. *Nature* **520**, 670-674, doi:10.1038/nature14125 (2015).

52 Jackson, D. J., Gern, J. E. & Lemanske, R. F., Jr. The contributions of allergic sensitization and respiratory pathogens to asthma inception. *J Allergy Clin Immunol* **137**, 659-665; quiz 666, doi:10.1016/j.jaci.2016.01.002 (2016).

53 Simpson, A. *et al.* Beyond atopy: multiple patterns of sensitization in relation to asthma in a birth cohort study. *American journal of respiratory and critical care medicine* **181**, 1200-1206, doi:10.1164/rccm.200907-1101OC (2010).

54 Stoltz, D. J. *et al.* Specific patterns of allergic sensitization in early childhood and asthma & rhinitis risk. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **43**, 233-241, doi:10.1111/cea.12050 (2013).

55 Litonjua, A. A., Carey, V. J., Burge, H. A., Weiss, S. T. & Gold, D. R. Parental history and the risk for childhood asthma. Does mother confer more risk than father? *American journal of respiratory and critical care medicine* **158**, 176-181, doi:10.1164/ajrccm.158.1.9710014 (1998).

56 Xu, R., DeMauro, S. B. & Feng, R. The impact of parental history on children's risk of asthma: a study based on the National Health and Nutrition Examination Survey-III. *Journal of asthma and allergy* **8**, 51-61, doi:10.2147/jaa.s80245 (2015).

57 Martinez, F. D. *et al.* Asthma and wheezing in the first six years of life. The Group Health Medical Associates. *N Engl J Med* **332**, 133-138, doi:10.1056/nejm199501193320301 (1995).

58 Sigurs, N., Bjarnason, R., Sigurbergsson, F. & Kjellman, B. Respiratory syncytial virus bronchiolitis in infancy is an important risk factor for asthma and allergy at age 7. *American journal of respiratory and critical care medicine* **161**, 1501-1507, doi:10.1164/ajrccm.161.5.9906076 (2000).

59 Stein, R. T. *et al.* Respiratory syncytial virus in early life and risk of wheeze and allergy by age 13 years. *Lancet* **354**, 541-545, doi:10.1016/s0140-6736(98)10321-5 (1999).

60 Neuman, A. *et al.* Maternal smoking in pregnancy and asthma in preschool children: a pooled analysis of eight birth cohorts. *American journal of respiratory and critical care medicine* **186**, 1037-1043, doi:10.1164/rccm.201203-0501OC (2012).

61 Vork, K. L., Broadwin, R. L. & Blaisdell, R. J. Developing asthma in childhood from exposure to secondhand tobacco smoke: insights from a meta-regression. *Environmental health perspectives* **115**, 1394-1400, doi:10.1289/ehp.10155 (2007).

62 Tanaka, K., Miyake, Y., Sasaki, S., Ohya, Y. & Hirota, Y. Maternal smoking and environmental tobacco smoke exposure and the risk of allergic diseases in Japanese infants: the Osaka Maternal and Child Health Study. *The Journal of asthma : official journal of the Association for the Care of Asthma* **45**, 833-838, doi:10.1080/02770900802339742 (2008).

63 Haberg, S. E., Stigum, H., Nystad, W. & Nafstad, P. Effects of pre- and postnatal exposure to parental smoking on early childhood respiratory health. *American journal of epidemiology* **166**, 679-686, doi:10.1093/aje/kwm134 (2007).

64 Jackson, D. J., Sykes, A., Mallia, P. & Johnston, S. L. Asthma exacerbations: origin, effect, and prevention. *J Allergy Clin Immunol* **128**, 1165-1174, doi:10.1016/j.jaci.2011.10.024 (2011).

65 Tischer, C., Chen, C. M. & Heinrich, J. Association between domestic mould and mould components, and asthma and allergy in children: a systematic review. *The European respiratory journal* **38**, 812-824, doi:10.1183/09031936.00184010 (2011).

66 Gasana, J., Dillikar, D., Mendy, A., Forno, E. & Ramos Vieira, E. Motor vehicle air pollution and asthma in children: a meta-analysis. *Environmental research* **117**, 36-45, doi:10.1016/j.envres.2012.05.001 (2012).

67 Trasande, L. & Thurston, G. D. The role of air pollution in asthma and other pediatric morbidities. *J Allergy Clin Immunol* **115**, 689-699, doi:10.1016/j.jaci.2005.01.056 (2005).

68 Rasmussen, S. G. *et al.* Association Between Unconventional Natural Gas Development in the Marcellus Shale and Asthma Exacerbations. *JAMA internal medicine* **176**, 1334-1343, doi:10.1001/jamainternmed.2016.2436 (2016).

69 Caliskan, M. *et al.* Rhinovirus wheezing illness and genetic risk of childhood-onset asthma. *N Engl J Med* **368**, 1398-1407, doi:10.1056/NEJMoa1211592 (2013).

70 Martinez, F. D. CD14, Endotoxin, and Asthma Risk: Actions and Interactions. *Proceedings of the American Thoracic Society* **4**, 221-225, doi:10.1513/pats.200702-035AW (2007).

71 Schaub, B., Lauener, R. & von Mutius, E. The many faces of the hygiene hypothesis. *J Allergy Clin Immunol* **117**, 969-977; quiz 978, doi:10.1016/j.jaci.2006.03.003 (2006).

72 Strachan, D. P. Hay fever, hygiene, and household size. *BMJ (Clinical research ed.)* **299**, 1259-1260 (1989).

73 Riedler, J. *et al.* Exposure to farming in early life and development of asthma and allergy: a cross-sectional survey. *Lancet* **358**, 1129-1133 (2001).

74 Ege, M. J. *et al.* Not all farming environments protect against the development of asthma and wheeze in children. *J Allergy Clin Immunol* **119**, 1140-1147, doi:10.1016/j.jaci.2007.01.037 (2007).

75 Illi, S. *et al.* Protection from childhood asthma and allergy in Alpine farm environments-the GABRIEL Advanced Studies. *J Allergy Clin Immunol* **129**, 1470-1477.e1476, doi:10.1016/j.jaci.2012.03.013 (2012).

76 Loss, G. *et al.* Consumption of unprocessed cow's milk protects infants from common respiratory infections. *J Allergy Clin Immunol* **135**, 56-62, doi:10.1016/j.jaci.2014.08.044 (2015).

77 von Mutius, E. & Vercelli, D. Farm living: effects on childhood asthma and allergy. *Nat Rev Immunol* **10**, 861-868, doi:nri2871 [pii]
10.1038/nri2871 (2010).

78 Ege, M. J. *et al.* Exposure to environmental microorganisms and childhood asthma. *N Engl J Med* **364**, 701-709, doi:10.1056/NEJMoa1007302 (2011).

79 Riedler, J., Eder, W., Oberfeld, G. & Schreuer, M. Austrian children living on a farm have less hay fever, asthma and allergic sensitization. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **30**, 194-200 (2000).

80 Von Ehrenstein, O. S. *et al.* Reduced risk of hay fever and asthma among children of farmers. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **30**, 187-193 (2000).

81 Douwes, J. *et al.* Farm exposure in utero may protect against asthma, hay fever and eczema. *The European respiratory journal* **32**, 603-611, doi:10.1183/09031936.00033707 (2008).

82 Ownby, D. R., Johnson, C. C. & Peterson, E. L. Exposure to dogs and cats in the first year of life and risk of allergic sensitization at 6 to 7 years of age. *Jama* **288**, 963-972 (2002).

83 Wegienka, G. *et al.* Lifetime dog and cat exposure and dog- and cat-specific sensitization at age 18 years. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **41**, 979-986, doi:10.1111/j.1365-2222.2011.03747.x (2011).

84 Chen, C. M. *et al.* Longitudinal study on cat allergen exposure and the development of allergy in young children. *J Allergy Clin Immunol* **119**, 1148-1155, doi:10.1016/j.jaci.2007.02.017 (2007).

85 Ball, T. M. *et al.* Siblings, day-care attendance, and the risk of asthma and wheezing during childhood. *N Engl J Med* **343**, 538-543, doi:10.1056/nejm200008243430803 (2000).

86 von Mutius, E. The microbial environment and its influence on asthma prevention in early life. *J Allergy Clin Immunol* **137**, 680-689, doi:10.1016/j.jaci.2015.12.1301 (2016).

87 Perkin, M. R. Unpasteurized milk: health or hazard? *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **37**, 627-630, doi:10.1111/j.1365-2222.2007.02715.x (2007).

88 Seedorf, J. *et al.* Concentrations and emissions of airborne endotoxins and microorganisms in livestock buildings in Northern Europe. *Journal of Agricultural Engineering Research* **70**, 97-109 (1998).

89 Peters, M. *et al.* Arabinogalactan isolated from cowshed dust extract protects mice from allergic airway inflammation and sensitization. *J Allergy Clin Immunol* **126**, 648-656.e641-644, doi:10.1016/j.jaci.2010.05.011 (2010).

90 Ege, M. J. *et al.* Gene-environment interaction for childhood asthma and exposure to farming in Central Europe. *J Allergy Clin Immunol* **127**, 138-144, 144 e131-134, doi:S0091-6749(10)01630-1 [pii]
10.1016/j.jaci.2010.09.041 (2011).

91 Braun-Fahrlander, C. *et al.* Environmental exposure to endotoxin and its relation to asthma in school-age children. *N Engl J Med* **347**, 869-877 (2002).

92 Vartiainen, E., Petays, T., Haahtela, T., Jousilahti, P. & Pekkanen, J. Allergic diseases, skin prick test responses, and IgE levels in North Karelia, Finland, and the Republic of Karelia, Russia. *J Allergy Clin Immunol* **109**, 643-648 (2002).

93 Haahtela, T. *et al.* Hunt for the origin of allergy - comparing the Finnish and Russian Karelia. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **45**, 891-901, doi:10.1111/cea.12527 (2015).

94 Lauener, R. P. *et al.* Expression of CD14 and Toll-like receptor 2 in farmers' and non-farmers' children. *Lancet* **360**, 465-466, doi:10.1016/s0140-6736(02)09641-1 (2002).

95 Ege, M. J. *et al.* Prenatal farm exposure is related to the expression of receptors of the innate immunity and to atopic sensitization in school-age children. *J Allergy Clin Immunol* **117**, 817-823, doi:10.1016/j.jaci.2005.12.1307 (2006).

96 Pfefferle, P. I. *et al.* Cord blood allergen-specific IgE is associated with reduced IFN-gamma production by cord blood cells: the Protection against Allergy-Study in Rural Environments (PASTURE) Study. *J Allergy Clin Immunol* **122**, 711-716, doi:10.1016/j.jaci.2008.06.035 (2008).

97 Pfefferle, P. I. *et al.* Cord blood cytokines are modulated by maternal farming activities and consumption of farm dairy products during pregnancy: the PASTURE Study. *J Allergy Clin Immunol* **125**, 108-115.e101-103, doi:10.1016/j.jaci.2009.09.019 (2010).

98 Kaario, H. *et al.* Exposure to a farm environment is associated with T helper 1 and regulatory cytokines at age 4.5 years. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **46**, 71-77, doi:10.1111/cea.12636 (2016).

99 Nolt, S. M. *A history of the Amish*. (Good Books, 1992).

100 Eshleman, H. F. *Historic background and annals of the Swiss and German pioneer settlers of southeastern Pennsylvania*. (1917).

101 Hostetler, J. A. *Amish society*. 4th edn, (Johns Hopkins University Press, 1993).

102 Triggs-Raine, B. *et al.* Development of a diagnostic DNA chip to screen for 30 autosomal recessive disorders in the Hutterite population. *Molecular Genetics & Genomic Medicine* **4**, 312-321, doi:10.1002/mgg3.206 (2016).

103 Martin, A. O. The founder effect in a human isolate: Evolutionary implications. *American Journal of Physical Anthropology* **32**, 351-367, doi:10.1002/ajpa.1330320305 (1970).

104 Ober, C., Abney, M. & McPeek, M. S. The Genetic Dissection of Complex Traits in a Founder Population. *American Journal of Human Genetics* **69**, 1068-1079 (2001).

105 Hostetler, J. A. *Hutterite Society*. (The Johns Hopkins University Press, 1974).

106 Holbreich, M. *et al.* Amish children living in Northern Indiana have a very low prevalence of allergic sensitization. *J. Allergy Clin. Immunol.* **129**, 1671-1673, doi:S0091-6749(12)00519-2 [pii]
10.1016/j.jaci.2012.03.016 (2012).

107 Motika, C. A., Papachristou, C., Abney, M., Lester, L. A. & Ober, C. Rising prevalence of asthma is sex-specific in a US farming population. *J Allergy Clin Immunol* **128**, 774-779, doi:10.1016/j.jaci.2011.06.046 (2011).

108 Ober, C. & Yao, T. C. The genetics of asthma and allergic disease: a 21st century perspective. *Immunological reviews* **242**, 10-30, doi:10.1111/j.1600-065X.2011.01029.x (2011).

109 Meyers, D. A., Bleecker, E. R., Holloway, J. W. & Holgate, S. T. Asthma genetics and personalised medicine. *The Lancet. Respiratory medicine* **2**, 405-415, doi:10.1016/s2213-2600(14)70012-8 (2014).

110 Bach, J. F. The effect of infections on susceptibility to autoimmune and allergic diseases. *N Engl J Med* **347**, 911-920 (2002).

111 Adachi, O. *et al.* Targeted disruption of the MyD88 gene results in loss of IL-1- and IL-18-mediated function. *Immunity* **9**, 143-150 (1998).

112 Yamamoto, M. *et al.* Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science (New York, N.Y.)* **301**, 640-643, doi:10.1126/science.1087262 (2003).

113 Benjamini, Y., Hochberg, Yosef. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society* **57**, 11 (1995).

114 Cann, H. M. *et al.* A human genome diversity cell line panel. *Science (New York, N.Y.)* **296**, 261-262 (2002).

115 Kim-Schulze, S. *et al.* Recombinant Ig-like transcript 3-Fc modulates T cell responses via induction of Th anergy and differentiation of CD8+ T suppressor cells. *Journal of immunology* **176**, 2790-2798 (2006).

116 Velten, F. W., Duperrier, K., Bohlender, J., Metharom, P. & Goerdt, S. A gene signature of inhibitory MHC receptors identifies a BDCA3(+) subset of IL-10-induced dendritic cells with reduced allostimulatory capacity in vitro. *European journal of immunology* **34**, 2800-2811, doi:10.1002/eji.200324732 (2004).

117 Lluis, A. *et al.* Increased regulatory T-cell numbers are associated with farm milk exposure and lower atopic sensitization and asthma in childhood. *J Allergy Clin Immunol* **133**, 551-559, doi:10.1016/j.jaci.2013.06.034 (2014).

118 Schaub, B. *et al.* Maternal farm exposure modulates neonatal immune mechanisms through regulatory T cells. *J Allergy Clin Immunol* **123**, 774-782.e775, doi:10.1016/j.jaci.2009.01.056 (2009).

119 Ma, A. & Malynn, B. A. A20: linking a complex regulator of ubiquitylation to immunity and human disease. *Nat Rev Immunol* **12**, 774-785, doi:10.1038/nri3313 (2012).

120 Schuijs, M. J. *et al.* Farm dust and endotoxin protect against allergy through A20 induction in lung epithelial cells. *Science (New York, N.Y.)* **349**, 1106-1110, doi:10.1126/science.aac6623 (2015).

121 Bosco, A., Ehteshami, S., Panyala, S. & Martinez, F. D. Interferon regulatory factor 7 is a major hub connecting interferon-mediated responses in virus-induced asthma exacerbations in vivo. *J Allergy Clin Immunol* **129**, 88-94, doi:10.1016/j.jaci.2011.10.038 (2012).

122 Loss, G. J. *et al.* The Early Development of Wheeze. Environmental Determinants and Genetic Susceptibility at 17q21. *American journal of respiratory and critical care medicine* **193**, 889-897, doi:10.1164/rccm.201507-1493OC (2016).

123 Li, Q. *et al.* Tripartite motif 8 (TRIM8) modulates TNFalpha- and IL-1beta-triggered NF-kappaB activation by targeting TAK1 for K63-linked polyubiquitination. *Proceedings of the National Academy of Sciences of the United States of America* **108**, 19341-19346, doi:10.1073/pnas.1110946108 (2011).

124 Meisel, C. *et al.* Granulocyte-macrophage colony-stimulating factor to reverse sepsis-associated immunosuppression: a double-blind, randomized, placebo-controlled multicenter trial. *American journal of respiratory and critical care medicine* **180**, 640-648, doi:10.1164/rccm.200903-0363OC (2009).

125 Gomez, H. G. *et al.* Immunological characterization of compensatory anti-inflammatory response syndrome in patients with severe sepsis: a longitudinal study*. *Critical care medicine* **42**, 771-780, doi:10.1097/ccm.0000000000000100 (2014).

126 Fortunati, E., Kazemier, K. M., Grutters, J. C., Koenderman, L. & Van den Bosch v, J. Human neutrophils switch to an activated phenotype after homing to the lung irrespective of inflammatory disease. *Clinical and experimental immunology* **155**, 559-566, doi:10.1111/j.1365-2249.2008.03791.x (2009).

127 Orden, S. *et al.* Efavirenz induces interactions between leucocytes and endothelium through the activation of Mac-1 and gp150,95. *The Journal of antimicrobial chemotherapy* **69**, 995-1004, doi:10.1093/jac/dkt468 (2014).

128 Devi, S. *et al.* Neutrophil mobilization via plerixafor-mediated CXCR4 inhibition arises from lung demargination and blockade of neutrophil homing to the bone marrow. *The Journal of experimental medicine* **210**, 2321-2336, doi:10.1084/jem.20130056 (2013).

129 Yamada, M. *et al.* The increase in surface CXCR4 expression on lung extravascular neutrophils and its effects on neutrophils during endotoxin-induced lung injury. *Cellular & molecular immunology* **8**, 305-314, doi:10.1038/cmi.2011.8 (2011).

130 Netea, M. G., Wijmenga, C. & O'Neill, L. A. Genetic variation in Toll-like receptors and disease susceptibility. *Nature immunology* **13**, 535-542, doi:10.1038/ni.2284 (2012).

131 Yao, T. C. *et al.* Genome-wide association study of lung function phenotypes in a founder population. *J Allergy Clin Immunol* **133**, 248-255 e241-210, doi:10.1016/j.jaci.2013.06.018 (2014).

132 Livne, O. E. *et al.* PRIMAL: Fast and Accurate Pedigree-based Imputation from Sequence Data in a Founder Population. *PLoS computational biology* **11**, e1004139, doi:10.1371/journal.pcbi.1004139 (2015).

133 Kilburg-Basnyat, B., Metwali, N. & Thorne, P. S. Effect of deployment time on endotoxin and allergen exposure assessment using electrostatic dust collectors. *The Annals of occupational hygiene* **59**, 104-115, doi:10.1093/annhyg/meu063 (2015).

134 Caporaso, J. G. *et al.* Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. *The ISME journal* **6**, 1621-1624, doi:10.1038/ismej.2012.8 (2012).

135 Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature methods* **7**, 335-336, doi:10.1038/nmeth.f.303 (2010).

136 Du, P., Kibbe, W. A. & Lin, S. M. lumi: a pipeline for processing Illumina microarray. *Bioinformatics* **24**, 1547-1548, doi:10.1093/bioinformatics/btn224 (2008).

137 Johnson, W. E., Li, C. & Rabinovic, A. Adjusting batch effects in microarray expression data using empirical Bayes methods. *biostatistics.oxfordjournals.org*.

138 Langfelder, P. & Horvath, S. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* **9**, 559, doi:10.1186/1471-2105-9-559 (2008).

139 McGovern, T. K., Robichaud, A., Fereydoonzad, L., Schuessler, T. F. & Martin, J. G. Evaluation of respiratory system mechanics in mice using the forced oscillation technique. *Journal of visualized experiments : JoVE*, e50172, doi:10.3791/50172 (2013).

140 Hamada, K. *et al.* Allergen-independent maternal transmission of asthma susceptibility. *Journal of immunology (Baltimore, Md. : 1950)* **170**, 1683-1689 (2003).

141 Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)* **57**, 289-300 (1995).

142 Guilliams, M., Bruhns, P., Saeys, Y., Hammad, H. & Lambrecht, B. N. The function of Fcgamma receptors in dendritic cells and macrophages. *Nat Rev Immunol* **14**, 94-108, doi:10.1038/nri3582 (2014).

143 Bruhns, P. *et al.* Specificity and affinity of human Fcgamma receptors and their polymorphic variants for human IgG subclasses. *Blood* **113**, 3716-3725, doi:10.1182/blood-2008-09-179754 (2009).

144 Ivan, E. & Colovai, A. I. Human Fc receptors: critical targets in the treatment of autoimmune diseases and transplant rejections. *Human immunology* **67**, 479-491, doi:10.1016/j.humimm.2005.12.001 (2006).

145 Bredius, R. G. *et al.* Role of neutrophil Fc gamma RIIa (CD32) and Fc gamma RIIIb (CD16) polymorphic forms in phagocytosis of human IgG1- and IgG3-opsonized bacteria and erythrocytes. *Immunology* **83**, 624-630 (1994).

146 Marcus, J. H. & Novembre, J. Visualizing the geography of genetic variants. *Bioinformatics*, doi:10.1093/bioinformatics/btw643 (2016).

147 Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* **47**, 979-986, doi:10.1038/ng.3359 (2015).

148 Khor, C. C. *et al.* Genome-wide association study identifies FCGR2A as a susceptibility locus for Kawasaki disease. *Nat Genet* **43**, 1241-1246, doi:10.1038/ng.981 (2011).

149 Bentham, J. *et al.* Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. *Nat Genet* **47**, 1457-1464, doi:10.1038/ng.3434 (2015).

150 Song, G. G. & Lee, Y. H. Associations between FCGR2A rs1801274, FCGR3A rs396991, FCGR3B NA1/NA2 polymorphisms and periodontitis: a meta-analysis. *Molecular biology reports* **40**, 4985-4993, doi:10.1007/s11033-013-2599-y (2013).

151 Sole-Violan, J. *et al.* The Fcgamma receptor IIA-H/H131 genotype is associated with bacteremia in pneumococcal community-acquired pneumonia. *Critical care medicine* **39**, 1388-1393, doi:10.1097/CCM.0b013e31820eda74 (2011).

152 Stewart, R., Hammond, S. A., Oberst, M. & Wilkinson, R. W. The role of Fc gamma receptors in the activity of immunomodulatory antibodies for cancer. *Journal for ImmunoTherapy of Cancer* **2**, 29, doi:10.1186/s40425-014-0029-x (2014).

153 Ceuppens, J. L., Bloemmen, F. J. & Van Wauwe, J. P. T cell unresponsiveness to the mitogenic activity of OKT3 antibody results from a deficiency of monocyte Fc gamma receptors for murine IgG2a and inability to cross-link the T3-Ti complex. *Journal of immunology* **135**, 3882-3886 (1985).

154 Ceuppens, J. L., Meurs, L. & Van Wauwe, J. P. Failure of OKT3 monoclonal antibody to induce lymphocyte mitogenesis: a familial defect in monocyte helper function. *Journal of immunology* **134**, 1498-1502 (1985).

155 Duffy, D. *et al.* Functional analysis via standardized whole-blood stimulation systems defines the boundaries of a healthy immune response to complex stimuli. *Immunity* **40**, 436-450, doi:10.1016/j.immuni.2014.03.002 (2014).

156 Mueller, S. C., Marz, R., Schmolz, M. & Drewelow, B. Intraindividual long term stability and response corridors of cytokines in healthy volunteers detected by a standardized whole-blood culture system for bed-side application. *BMC medical research methodology* **12**, 112, doi:10.1186/1471-2288-12-112 (2012).

157 Wilson, N. S. *et al.* An Fcgamma receptor-dependent mechanism drives antibody-mediated target-receptor signaling in cancer cells. *Cancer cell* **19**, 101-113, doi:10.1016/j.ccr.2010.11.012 (2011).

158 Weng, W. K. & Levy, R. Two immunoglobulin G fragment C receptor polymorphisms independently predict response to rituximab in patients with follicular lymphoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **21**, 3940-3947, doi:10.1200/jco.2003.05.013 (2003).

159 Kjersem, J. B. *et al.* FCGR2A and FCGR3A polymorphisms and clinical outcome in metastatic colorectal cancer patients treated with first-line 5-fluorouracil/folinic acid and oxaliplatin +/- cetuximab. *BMC cancer* **14**, 340, doi:10.1186/1471-2407-14-340 (2014).

160 Avila-Pedretti, G. *et al.* Variation at FCGR2A and functionally related genes is associated with the response to anti-TNF therapy in rheumatoid arthritis. *PloS one* **10**, e0122088, doi:10.1371/journal.pone.0122088 (2015).

161 Davila-Fajardo, C. L. *et al.* FcGR genetic polymorphisms and the response to adalimumab in patients with rheumatoid arthritis. *Pharmacogenomics* **16**, 373-381, doi:10.2217/pgs.14.178 (2015).

162 Musolino, A. *et al.* Immunoglobulin G fragment C receptor polymorphisms and clinical efficacy of trastuzumab-based therapy in patients with HER-2/neu-positive metastatic breast cancer. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* **26**, 1789-1796, doi:10.1200/jco.2007.14.8957 (2008).

163 Tamura, K. *et al.* FcgammaR2A and 3A polymorphisms predict clinical outcome of trastuzumab in both neoadjuvant and metastatic settings in patients with HER2-positive breast cancer. *Annals of oncology : official journal of the European Society for Medical Oncology* **22**, 1302-1307, doi:10.1093/annonc/mdq585 (2011).

164 Ghesquieres, H. *et al.* Clinical outcome of patients with follicular lymphoma receiving chemoimmunotherapy in the PRIMA study is not affected by FCGR3A and FCGR2A polymorphisms. *Blood* **120**, 2650-2657, doi:10.1182/blood-2012-05-431825 (2012).

165 Paez, D. *et al.* Immunoglobulin G fragment C receptor polymorphisms and KRAS mutations: are they useful biomarkers of clinical outcome in advanced colorectal cancer treated with anti-EGFR-based therapy? *Cancer science* **101**, 2048-2053, doi:10.1111/j.1349-7006.2010.01621.x (2010).

166 Hurvitz, S. A. *et al.* Analysis of Fcgamma receptor IIIa and IIa polymorphisms: lack of correlation with outcome in trastuzumab-treated breast cancer patients. *Clinical cancer research : an official journal of the American Association for Cancer Research* **18**, 3478-3486, doi:10.1158/1078-0432.ccr-11-2294 (2012).

167 Ober, C. *et al.* Effect of variation in CHI3L1 on serum YKL-40 level, risk of asthma, and lung function. *N Engl J Med* **358**, 1682-1691, doi:10.1056/NEJMoa0708801 (2008).

168 Ober, C. *et al.* Genome-wide association study of plasma lipoprotein(a) levels identifies multiple genes on chromosome 6q. *Journal of lipid research* **50**, 798-806, doi:10.1194/jlr.M800515-JLR200 (2009).

169 Zhou, X. & Stephens, M. Genome-wide efficient mixed-model analysis for association studies. *Nat Genet* **44**, 821-824, doi:10.1038/ng.2310 (2012).

170 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. *Genome biology* **14**, R36, doi:10.1186/gb-2013-14-4-r36 (2013).

171 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078-2079, doi:10.1093/bioinformatics/btp352 (2009).

172 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* **31**, 166-169, doi:10.1093/bioinformatics/btu638 (2015).

173 Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome biology* **15**, R29, doi:10.1186/gb-2014-15-2-r29 (2014).

174 Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **43**, e47, doi:10.1093/nar/gkv007 (2015).

175 Bonnelykke, K. & Ober, C. Leveraging gene-environment interactions and endotypes for asthma gene discovery. *J Allergy Clin Immunol* **137**, 667-679, doi:10.1016/j.jaci.2016.01.006 (2016).

176 Sparks, J. A. & Costenbader, K. H. Genetics, environment, and gene-environment interactions in the development of systemic rheumatic diseases. *Rheumatic diseases clinics of North America* **40**, 637-657, doi:10.1016/j.rdc.2014.07.005 (2014).

177 Nicolae, D. L. *et al.* Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS genetics* **6**, e1000888, doi:10.1371/journal.pgen.1000888 (2010).

178 Gusev, A. *et al.* Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. *American Journal of Human Genetics* **95**, 535-552, doi:10.1016/j.ajhg.2014.10.004 (2014).

179 Torres, J. M. *et al.* Cross-tissue and tissue-specific eQTLs: partitioning the heritability of a complex trait. *American Journal of Human Genetics* **95**, 521-534, doi:10.1016/j.ajhg.2014.10.001 (2014).

180 Croteau-Chonka, D. C. *et al.* Expression Quantitative Trait Loci Information Improves Predictive Modeling of Disease Relevance of Non-Coding Genetic Variation. *PloS one* **10**, e0140758, doi:10.1371/journal.pone.0140758 (2015).

181 Murphy, A. *et al.* Mapping of numerous disease-associated expression polymorphisms in primary peripheral blood CD4+ lymphocytes. *Human molecular genetics* **19**, 4745-4757, doi:10.1093/hmg/ddq392 (2010).

182 Bieli, C. *et al.* A polymorphism in CD14 modifies the effect of farm milk consumption on allergic diseases and CD14 gene expression. *J Allergy Clin Immunol* **120**, 1308-1315, doi:10.1016/j.jaci.2007.07.034 (2007).

183 Mahdi, H. *et al.* Specific interaction between genotype, smoking and autoimmunity to citrullinated alpha-enolase in the etiology of rheumatoid arthritis. *Nat Genet* **41**, 1319-1324, doi:10.1038/ng.480 (2009).

184 Lee, M. N. *et al.* Common genetic variants modulate pathogen-sensing responses in human dendritic cells. *Science (New York, N.Y.)* **343**, 1246980, doi:10.1126/science.1246980 (2014).

185 Ye, C. J. *et al.* Intersection of population variation and autoimmunity genetics in human T cell activation. *Science (New York, N.Y.)* **345**, 1254665, doi:10.1126/science.1254665 (2014).

186 Caliskan, M., Baker, S. W., Gilad, Y. & Ober, C. Host Genetic Variation Influences Gene Expression Response to Rhinovirus Infection. *PLoS genetics* **11**, e1005111, doi:10.1371/journal.pgen.1005111 (2015).

187 Moyerbrailean, G. A. *et al.* High-throughput allele-specific expression across 250 environmental conditions. *Genome research* **26**, 1627-1638, doi:10.1101/gr.209759.116 (2016).

188 Schwanhausser, B. *et al.* Global quantification of mammalian gene expression control. *Nature* **473**, 337-342, doi:10.1038/nature10098 (2011).

189 de Sousa Abreu, R., Penalva, L. O., Marcotte, E. M. & Vogel, C. Global signatures of protein and mRNA expression levels. *Molecular bioSystems* **5**, 1512-1526, doi:10.1039/b908315d (2009).

190 Gygi, S. P., Rochon, Y., Franza, B. R. & Aebersold, R. Correlation between protein and mRNA abundance in yeast. *Molecular and cellular biology* **19**, 1720-1730 (1999).

191 Garge, N. *et al.* Identification of quantitative trait loci underlying proteome variation in human lymphoblastoid cells. *Molecular & cellular proteomics : MCP* **9**, 1383-1399, doi:10.1074/mcp.M900378-MCP200 (2010).

192 Wu, L. *et al.* Variation and genetic control of protein abundance in humans. *Nature* **499**, 79-82, doi:10.1038/nature12223 (2013).

193 Hause, R. J. *et al.* Identification and validation of genetic variants that influence transcription factor and cell signaling protein levels. *American Journal of Human Genetics* **95**, 194-208, doi:10.1016/j.ajhg.2014.07.005 (2014).

194 Battle, A. *et al.* Genomic variation. Impact of regulatory variation from RNA to protein. *Science (New York, N.Y.)* **347**, 664-667, doi:10.1126/science.1260793 (2015).

195 Li, Y. *et al.* Inter-individual variability and genetic influences on cytokine responses to bacteria and fungi. *Nat Med* **22**, 952-960, doi:10.1038/nm.4139 (2016).

196 Li, Y. *et al.* A Functional Genomics Approach to Understand Variation in Cytokine Production in Humans. *Cell* **167**, 1099-1110.e1014, doi:10.1016/j.cell.2016.10.017 (2016).

197 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57-74, doi:10.1038/nature11247 (2012).

198 Abney, M. Permutation testing in the presence of polygenic variation. *Genetic epidemiology* **39**, 249-258, doi:10.1002/gepi.21893 (2015).

199 Kundaje, A. *et al.* Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317-330, doi:10.1038/nature14248 (2015).

200 Welter, D. *et al.* The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic acids research* **42**, D1001-1006, doi:10.1093/nar/gkt1229 (2014).

201 The Genotype-Tissue Expression (GTEx) project. *Nat Genet* **45**, 580-585, doi:10.1038/ng.2653 (2013).

202 Gozdz, J., Ober, C. & Vercelli, D. Innate Immunity and Asthma Risk. *N Engl J Med* **375**, 1898-1899, doi:10.1056/NEJMc1611699 (2016).

203 Liu, Y. P. *et al.* Association of ORMDL3 with rhinovirus-induced endoplasmic reticulum stress and type I Interferon responses in human leucocytes. *Clinical and experimental allergy : journal of the British Society for Allergy and Clinical Immunology* **47**, 371-382, doi:10.1111/cea.12903 (2017).

204 Cantero-Recasens, G., Fandos, C., Rubio-Moscardo, F., Valverde, M. A. & Vicente, R. The asthma-associated ORMDL3 gene product regulates endoplasmic reticulum-mediated calcium signaling and cellular stress. *Human molecular genetics* **19**, 111-121, doi:10.1093/hmg/ddp471 (2010).

205 Carreras-Sureda, A. *et al.* ORMDL3 modulates store-operated calcium entry and lymphocyte activation. *Human molecular genetics* **22**, 519-530, doi:10.1093/hmg/dds450 (2013).