# GENETICS OF SPONTANEOUS IDIOPATHIC PRETERM BIRTH: EXPLORATION OF MATERNAL AND FETAL GENOMES

By

Jude James McElroy

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements for

the degree of

DOCTOR OF PHILOSOPHY

in

Human Genetics

August, 2013

Nashville, Tennessee

Approved:

Dana C. Crawford, Ph.D.

Michael R. DeBaun, M.D., M.P.H.

Louis J. Muglia, M.D., Ph.D.

Dan M. Roden, M.D.

Scott M. Williams, Ph.D.

To all my friends and family without whom I would never have made it this far

and

A very special thank you to a charming person who saw me better than I saw myself and helped me believe…

## ACKNOWLEDGEMENTS

The work presented in this dissertation was supported by the Vanderbilt Medical Scientist Training Program (MSTP) grant T32 GM007347, as well as funding form the March of Dimes (21-FY2011-16 "Genetic Analysis of Human Preterm Birth"). A very heartfelt thank you to all the individuals who participated in these studies. All of the work presented here would have not been possible without you.

The work presented here was guided by the input from my thesis committee: Dana Crawford (committee chair), Michael DeBaun, Louis Muglia, Dan Roden and Scott Williams. Special acknowledgements are needed for my mentor, Louis Muglia. Lou has helped give me the skills required to complete my dissertation and allowed me grow to become an excellent scientist, but even more importantly was an excellent role model for how to be an outstanding person.

Thanks to the students of the Program in Human Genetics for their discussions and support. Special thanks to Will Bush and Stephen Turner for their excellent genetics blog, “Getting Genetics Done”, and especially for their R code which was used for the QQ and Manhattan plots presented here and used throughout my training.

I would also like to thank all of the staff of the DNA Resources Core especially Holli Dilks, Cara Sutcliffe and Paxton Baker. Additionally, Christian

Shafer was extremely helpful with our whole-exome sequencing and answering all of my computational questions.

Modern human genetics requires a team, and all of the people above have been instrumental in helping me complete my dissertation and made the journey more enjoyable.

## TABLE OF CONTENTS

Page

## LIST OF TABLES

## LIST OF FIGURES

## LIST OF APPENDICES

# CHAPTER I

## INTRODUCTION*

Preterm birth (PTB) is defined as live birth before 37 complete weeks gestation. Parturition in humans, typically occurs between 37-42 weeks gestation, with 40 weeks gestation being the most common (Figure 1.1). The rate of preterm birth has risen over the past two decades worldwide and in 2010 was 12.0% in the United States, which is a 15% increase from 1990.[1] PTB is responsible for 75% of perinatal mortality and greater than 50% of long-term morbidity.[2] The lungs[3] and brain[4] are particularly susceptible to insult. Studies have estimated that PTB is responsible for half of all pediatric neurodevelopmental disabilities.[5] The severity and incidence of these associated comorbidities increases with decreased gestational length.[6] When one considers the constellation of health conditions and disabilities prematurity creates, it is the leading cause of disability in the United States and worldwide.[7] In addition to the disabilities and medical comorbidities that PTB causes, its cost to society and families is devastating.[8,9] A 2006 report from the Institute of Medicine estimates the cost to care for preterm infants and the associated comorbidities at greater than $26 billion per year.[10] Clearly, PTB is a major public health concern.

*Adapted from: Jude J. McElroy's thesis proposal, Defining the Genetics of Spontaneous Idiopathic Preterm Birth

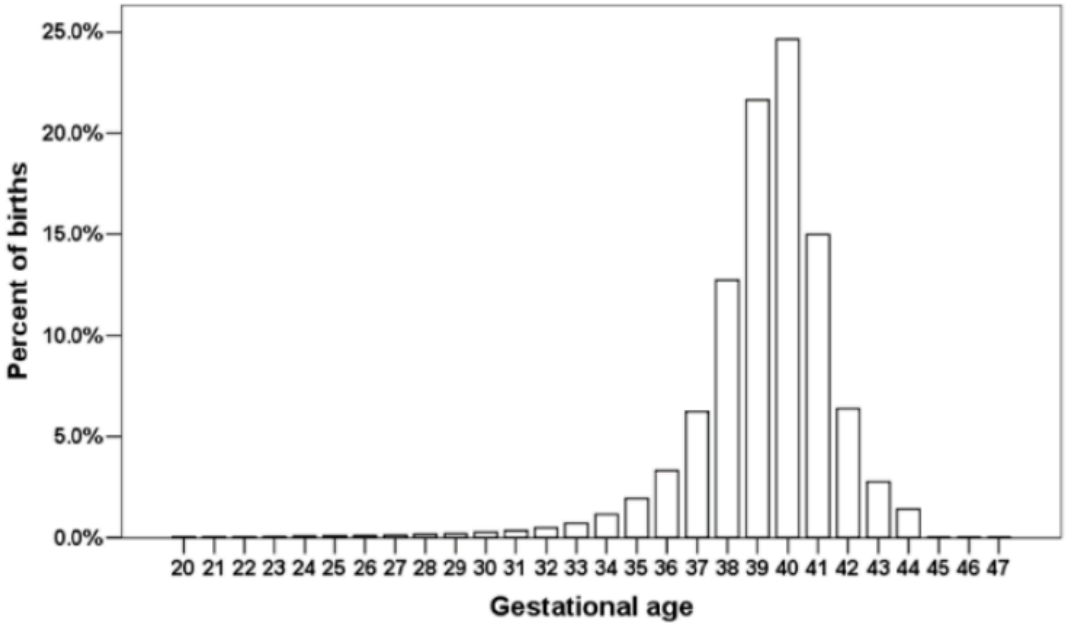

Figure 1.1: **Birth timing in Missouri (1978-1997).** From Plunkett et al. 2008

*Classification of Preterm birth*

There are two main categories of PTB: spontaneous and medically indicated. PTB can almost equally be divided between these two groups.[11] Medically indicated PTB can be grouped by indication, the most common of which include preeclampsia, placental abruption, and maternal or fetal distress. Similarly, spontaneous PTB can be divided into two groups: idiopathic and accompanied by preterm premature rupture of the membranes (PPROM). Racial disparity is seen in the most common causes and rates of spontaneous PTB. In Caucasian women, preterm labor is the most common cause while PPROM is responsible for the majority in African American women.[5] Spontaneous PTB is almost twice as common in African American women compared to Caucasian

women when all other risk factors are equal (Figure 1.2).[12] The cause of this large discrepancy remains unknown and represents one of the largest in medicine.

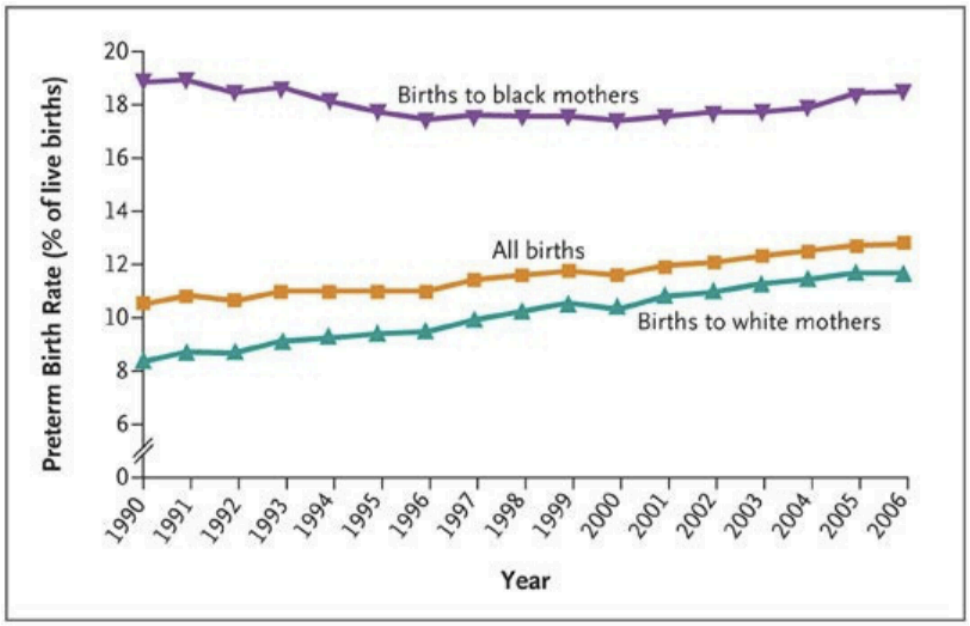


Figure 1.2: **Singleton preterm birth rates in the US 1990-2006.** Data from the National Center for Health Statistics by ethnicity. From Muglia & Katz 2010

*Pathways hypothesized to be involved in preterm birth*

Currently there are four main pathways hypothesized to be involved in the etiology of PTB: 1) activation of the maternal or fetal hypothalamic-pituitary-adrenal (HPA) axis, 2) infection and inflammation, 3) decidual hemorrhage, and 4) uterine distention (Figure 1.3).[13] These pathways are not mutually exclusive and there is much crosstalk between them. All of these pathways converge on a final common pathway that results in PTB.

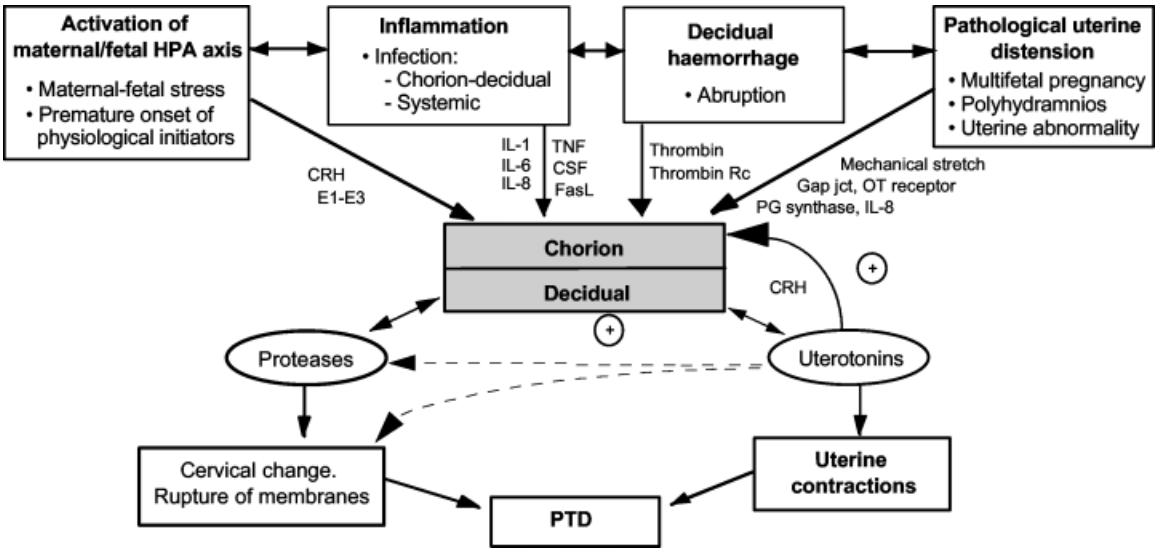


Figure 1.3: **Overview of pathways believed to be involved in preterm delivery.** From Lockwook & Kuczynski 2001

An important outstanding question is whether PTB represents inappropriate early activation of the normal parturition pathway or is a unique mechanism. This question is harder to answer than one might expect because human parturition is not well understood. Studies in sheep during the 1970s are the basis for the progesterone withdrawal theory of labor.[14,15] Briefly, levels of progesterone rise during pregnancy which is critical for maintenance of uterine quiescence, but abruptly decline prior to labor. The decline in progesterone is believed to be the precipitating event leading to labor. This abrupt decline is seen in mouse, rat, hamster, cow, goat, and sheep; however, it is absent in humans and great apes.[16,17] While there is not a systemic decline in progesterone concentration in humans prior to labor, there is some evidence that there may be a more localized decrease in progesterone or changes in the receptor isoform, co-

activators, and co-repressors resulting in a "functional withdrawal".[15,18,19] Additionally, there is evidence that the human progesterone receptor has been the target of adaptive evolution.[20] Because of the difference in labor initiation and evolution of the human progesterone receptor, studying the etiology and pathophysiology of human PTB is best performed in humans.

*Epidemiological risk factors*

Epidemiological studies have identified a number of risk factors that are associated with an increased risk of PTB. As previously mentioned race has been shown to play a significant role.[21-26] Other well recognized risk factors for PTB are low socioeconomic status and education, low and high maternal age, and single marital status.[27-29] Multiple gestations almost always deliver prior to term with nearly 60% of twins delivering preterm.[2] While only 2-3% of pregnancies are multiple-gestations they make up 15-20% of all PTBs.[2] Infection has long been recognized as a risk factor for PTB especially early (22-24 weeks) PTB.[30] Studies have shown that bacterial vaginosis,[31] trichomoniasis,[32] and periodontal infection[33] are more common in women who deliver preterm. Unfortunately, prophylactic treatment of gravid women with antibiotics has been unsuccessful at decreasing the rate of PTB even though it decreased the rate of positive cultures.[34,35] Both extremes of body mass index (BMI) have been shown to be risk factors for PTB. Women with low BMI[36] prior to pregnancy are more likely to have a spontaneous preterm delivery while obese[2] women are more likely to

develop complications that require a medically indicated PTB. However, some studies have seen a protective effect for obesity.[36] Low serum concentration of iron, folate, and zinc associate with delivering preterm.[37-39] Studies have also shown that high levels of psychological or social stress increase the risk of PTB even after correcting for sociodemographic, medical, and behavioral risk factors.[40,41] In the United States, 12-15% of pregnant women smoke throughout their pregnancy and smoking has been shown to increase the risk of PTB.[42-44] Heavy alcohol consumption and the use of cocaine and heroin have been associated with increased risk of PTB.[2] The interpregnancy interval has been shown to be a significant variable in PTB.[45] When this interval is less than six months, the rate of PTB is more than double after adjusting for other confounders.[46]

*Evidence of a genetic component*

A number of lines of evidence support the role of genetics in PTB, and birth timing in general. One of the most significant risk factors of having a preterm delivery is a positive family history.[47-49] The risk of preterm birth in females whose sister has had a preterm delivery is 80% higher themselves.[50] A number of twin and family studies investigating birth timing have been performed and report heritability ($h^2$) ranging between 30-40%.[51-53] These family studies have also illustrated that birth timing (both pre and post term) is stable across generations and sibships. The most common time or recurrence is during

the same week of gestation (Figure 1.4).[54-56] Segregation analysis has shown that the maternal genome plays the largest role in preterm birth.[53] Another line of evidence supporting a genetic component of PTB was a study by Ward et al. (2005), which found a significantly higher coefficient of kinship in PTB cases compared to controls in a Utah population.[57] However, even with this wealth of evidence that preterm birth has a large genetic component we are still currently unable to identify the variation responsible for the vast majority of cases.

Due to the known risk factors of PTB and four hypothesized pathways, the majority of previous candidate gene studies have focused on genes involved in the immune response, inflammation, drug metabolism, and connective tissue

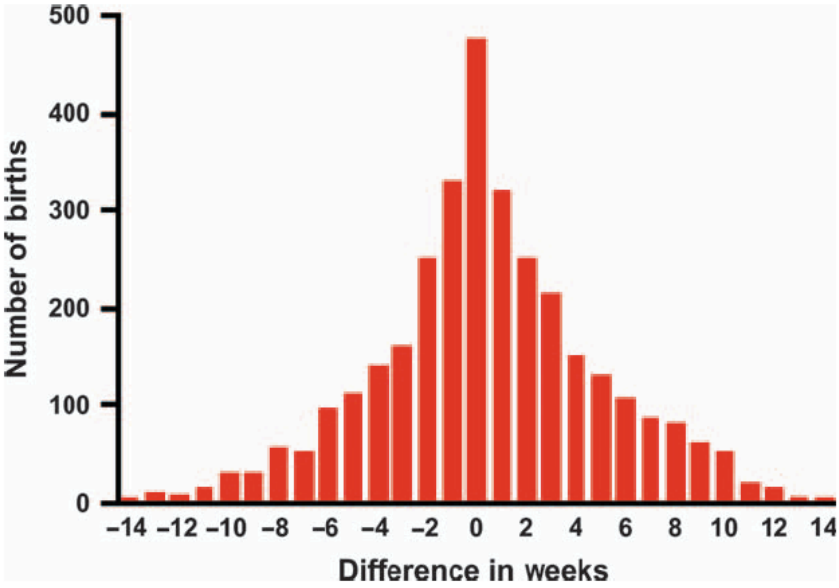


Figure 1.4: **A histogram illustrating the difference in gestational age between consecutive preterm pregnancies to an individual mother in Missouri, 1989-1997**. From Chaudhari et al. 2008

remodeling.[58-60] There have been a handful of positive associations between common variants and PTB many of which have not replicated due to a number of reasons one of which is lack of power due to small sample sizes. For a complete list of genes associated with PTB please see Appendix A.

While the genetic causes of the vast majority of PTBs are idiopathic, the Mendelian disorder Ehlers-Danlos Syndrome (EDS) serves as a proof-of-principle that a genetic disorder can increase the risk of PTB. EDS are a group of Mendelian disorders inherited in both an autosomal dominant and autosomal recessive manner in which the connective tissue is affected. Individuals with vascular (Type IV) EDS are at an increased risk of PTB primarily due to PPROM.[61]

Because PPROM involves rupture of the fetal membranes this raises the unresolved question of who should be considered the proband when investigating PTB, the infant or the mother (Figure 1.5)? Additionally, the placenta is fetal tissue and placental insufficiency and/or abruption are common causes of medically indicated PTB. While genetic modeling and segregation analysis have shown that the maternal genome plays the largest role in the genetic component of PTB risk, all children born to mothers who have had a preterm delivery are not born preterm. This raises the possibility that the fetus's genome individually or through an interaction with the maternal genome or the environment might be playing a partial role in etiology of PTB.

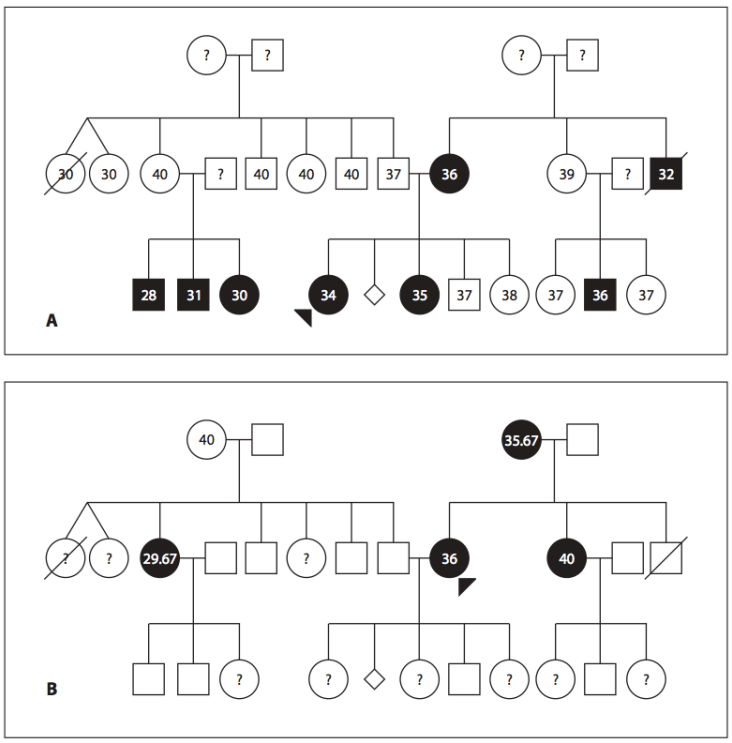


Figure 1.5: **The same pedigree illustrating the infant (A) and mother (B) as the affected proband.** The affected individuals (<37 weeks) are filled in with black and their gestational age is the number in the symbol. Individuals with a question mark have unknown gestational ages. Multiple gestation offspring were excluded due to the possibility of alternative mechanisms being involved compared to singleton preterm birth. From Plunkett et al. 2009

*Summary and objectives of dissertation*

Previous genetic research into PTB has had limited success at discovering novel insight into the mechanism of birth timing and parturition. The primary objective of this dissertation was to identify unidentified or unappreciated genetic variants responsible for the predisposition or pathogenesis of PTB. In order to investigate this hypothesis, we performed studies using a number of complementary approaches, which included genome-wide association studies

(GWAS), exome array association, whole-exome sequence and pathway analysis. In addition to using a group of complementary methods, we also interrogated both maternal and fetal genomes for their potential role in the etiology or pathogenesis of PTB. The ultimate goal of this research is to help prevent PTB and improve public health.

# CHAPTER II

## MATERNAL GENOME-WIDE ASSOCIATION FOR PRETERM BIRTH RISK AND GESTATION LENGTH[†]

### Introduction

Preterm birth (PTB) is defined by the World Health Organization (WHO) as live birth before 37 weeks' completed gestation and occurs in 12.0% of pregnancies in the United States.[62] Despite this major public health concern,[6,63] little is known about the pathogenesis of PTB. The limited insight into PTB is contributed to by the fact that the mechanism for normal parturition in general is not well understood in humans.[64]

A number of lines of evidence suggests that PTB has a genetic component such as PTB aggregating in families, segregation analysis and genetic modeling.[16,51,53] Primarily through candidate gene studies, there have been a number of SNPs associated with PTB; however, contradictory evidence from replication studies exists, and none of these have large effect sizes or have implicated new mechanisms in parturition control.[60]

Prior genetic model and segregation analysis have illustrated that the maternal genome plays an important role in the genetic contribution to PTB risk.[53] However, this does not mean that the fetal genome or paternally inherited

[†] Adapted from McElroy et al. In Preparation

genes in the fetus are not also potentially important, but only that the maternal genome plays the most significant role and models indicating a maternal genome contribution are the best fitting.

Due to this and the minimal success of candidate gene/pathway studies' ability to explain only a small percentage of cases of PTB, we decided to take a different approach and performed a genome-wide association study (GWAS) in a Finnish dataset of mothers form Helsinki and Oulu. This is conceptually different than a candidate/gene pathway study, which is limited by prior understanding or hypothesized mechanisms and is "hypothesis free".

The only "hypothesis" is that common alleles i.e. those with ≥ 5% minor allele frequency (MAF) are involved with the phenotype and have smaller effect sizes e.g. odds ratios (OR) generally ≤ 1.3.[65] GWAS are not without their own limitations, while candidate gene/pathway studies are limited by prior biology GWAS studies are limited by SNPs that are known at the time of the genotyping array design and also by the linkage disequilibrium (LD) in the population being studied. Additionally, "genome-wide" is somewhat of a misnomer and there are many genes/regions of the genome that are not covered at all or where the majority of the variation is not captured. However, even with these caveats GWAS have the ability to transform our understanding of the genetic underpinnings of disease and open up new avenues of research.[66]

Here we report the first GWAS investigating PTB and gestational age in a dataset of mothers from Helsinki and Oulu Finland that were genotyped using the Affymetrix Genome-Wide Human SNP Array 6.0. We chose to examine the more traditional dichotomous PTB phenotype as well as the transformed quantitative trait, gestational age, which increased our statistical power and the results should be somewhat complementary.[67] While we did not identify any associations that surpassed the genome-wide significance threshold of p-value = 5e-8, we were able to discover a number of strong associations that could help elucidate new biology and better understand the basic mechanism of parturition.

Due to the lack of an independent dataset to replicate/generalize our most significant associations from this GWAS should be considered tentative until replicated in a large independent dataset. We present these results with the hopes that others will attempt to generalize our strongest associations in their datasets.

## Results

*Study population characteristics*

The Finish mother dataset consisted of 539 total individuals of which 252 were cases and the remaining 287 samples were controls. These samples were a collection of mothers from Helsinki and Oulu Finland. Out of the 539 total samples, 152 were from Oulu and the remaining from Helsinki.

For the samples collected in Helsinki, we have access to dense demographic information some of which included maternal age, body mass index (BMI), gravidity, parity, smoking and alcohol use; however, unfortunately this information was not available for the Oulu mothers. If we had decided to adjust for variables, which were found to differ significantly between cases and controls in the Helsinki samples, we would be excluding all of the Oulu samples (~28% of the total number) from our analysis. Due to our small sample size, we could not afford the loss of that many samples; therefore all association analysis described below is unadjusted for any covariates.

*Finnish mothers preterm birth chi-square association*

The first analysis we performed was a standard allelic chi-squared genome-wide association study (GWAS) examining an association with PTB. In evaluating the quantile-quantile plot (QQ), we observed fewer than expected significant associations at the higher levels of significance (Figure 2.1). Our most significant SNP was rs871476 (p-value = 6.22e-6; OR = 0.484). This is an intergenic SNP which is ~114kb 5′ upstream of solute carrier family 34 (sodium phosphate), member 2 (*SLC34A2*) and ~123 kb 3′ downstream of anaphase promoting complex subunit 4 (*ANAPC4*). How this SNP is protective against PTB is currently unclear and will require further functional studies. For a complete list of SNPs with p-value < $10^{-4}$ please see Appendix B.

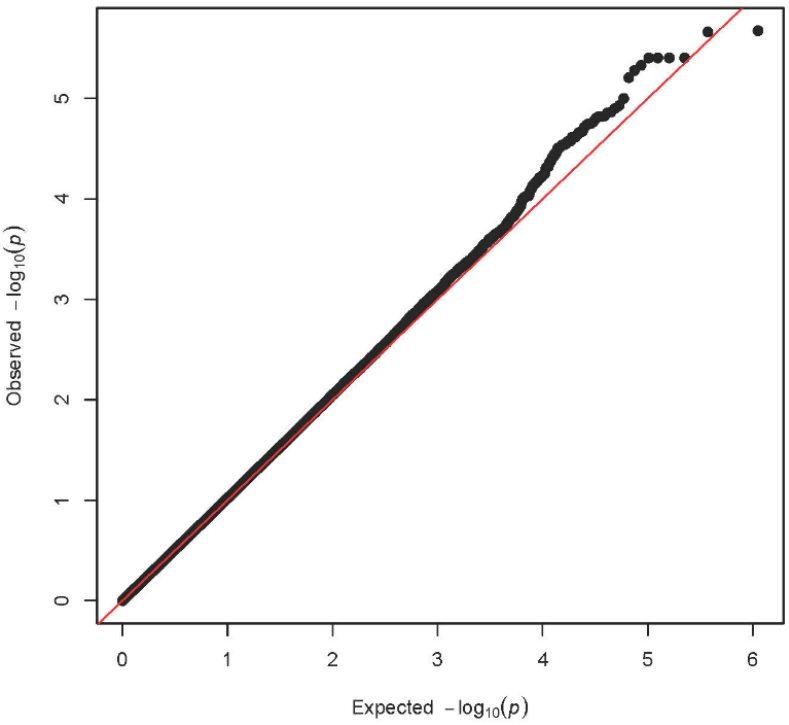


Figure 2.1: **QQ plot for chi-square allelic association analysis in our Finnish mothers genotyped on the Affymetrix 6.0 SNP array**.

*Box-Cox transformation of gestational ages*

Due to the smaller size of our Finnish mother dataset (539 total samples 252 born in the preterm range (< 37 weeks) and 287 term or (37-41 weeks)), we re-analyzed our Affymetrix 6.0 SNP arrays using the quantitative trait gestational age (in days), which increased our power to detect an association.[67] However, one of the main assumptions of linear regression is that the dependent variable is normally distributed. While linear regression is robust to deviations from normality we wanted to formally test whether our dependent variable i.e. gestational ages were normally distributed. To do so we performed a one-sample

Kolmogorov-Smirnov test which revealed that our gestational ages were very skewed, p-value = 1.49e-85 (Figure 2.2).

In order to normalize this distribution, we performed a Box-Cox power transformation. The lambda was determined using the Stata software package "boxcox" function with the rs6534679 genotypes coded recessively and was determined to be ~4.2. The gestational age was transformed using the original gestational age^4.2 and the distribution was graphed again and appeared more normally distributed (Figure 2.3). We chose to use rs6534679 coded recessively to determine the best lambda because this was the SNP and model that had the most significant p-value for the untransformed gestational age linear regression when we examined the additive, dominant and recessive models (data not shown).

*Box-Cox transformed gestational age unadjusted linear regression association analysis of Finnish mothers genotyped on the Affymetrix 6.0 SNP array*

The unadjusted linear regression was performed using the Box-Cox transformed gestational age as our dependent variable. When the additive model was performed we observed three SNPs with p-values at the $10^{-6}$ level of significance with the most significant SNP being rs10874644 which is ~185kb 3' downstream of collagen, type XI, alpha 1 (*COL11A1*), p-value = 5.26e-6 and is protective i.e. promotes longer gestational age (Table 2.1). For a complete

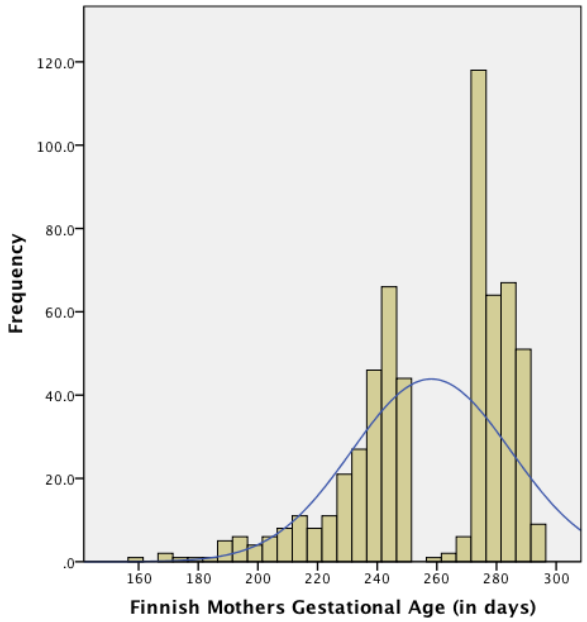


Figure 2.2: **Distribution of Finnish mother gestational ages in days with a normal curve overlaid.** The gestational ages are not normal (one-sample Kolmogorov-Smirnov test p-value = 1.49e-85). In order to try to normalize our gestational ages we performed a Box-Cox power transformation and used that variable in the linear regression

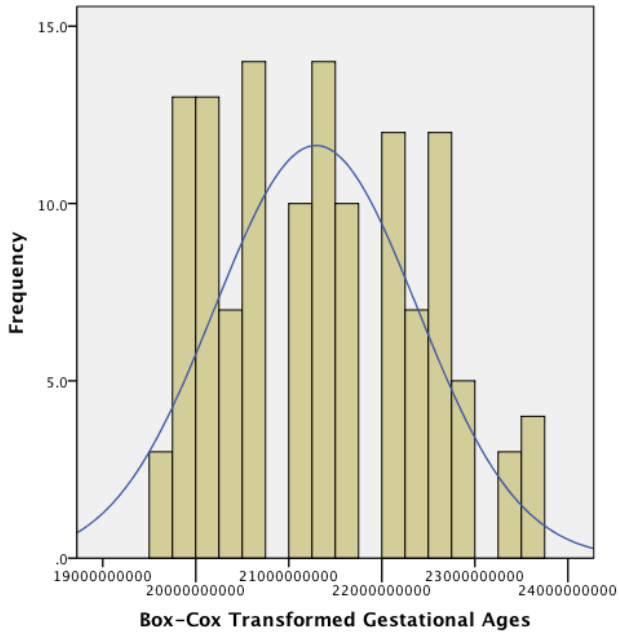


Figure 2.3: **Distribution of Finnish mother Box-Cox transformed gestational age with a normal curve overlaid.** Using the Stata software package the optimal lambda for the Box-Cox power transformation was determined to be ~4.2. The transformed variable was then graphed and a normal curve was overlaid**.**

list of SNPs with p-values < $10^{-4}$ please see Appendix C.

Because there is no a priori evidence to require gestational age to follow an allelic model, we also examined the genotypic 2-degree of freedom Box-Cox transformed gestational age linear regression for our Finnish mothers. We identified four SNPs with p-values ≤ $10^{-7}$ with the SNP 5′ upstream of *TACC1* rs10104530 (p-value = 7.09e-8) having the strongest genotypic association. When these four SNPs were tested using a dominant and recessive model rs10104530 had the most significant p-value, p-value = 9.45e-7, under the recessive model (Table 2.2).

## Discussion

To our knowledge, this is the first reported GWAS investigating preterm birth and gestational age in mothers. Due to this fact alone, we believe that this report is with merit.

When investigating PTB our most significant association was observed for rs871476, p-value = 6.22e-6, OR = 0.48, which is an intergenic SNP that is ~114kb 5′ upstream of *SLC34A2* and ~123 kb 3′ downstream of *ANAPC4*.

SLC34A2, or the solute carrier family 34 (sodium phosphate), member 2, is involved in the active transport of phosphate into cells and maybe involved in surfactant production in type II alveolar cells in the lungs.[68,69] While *SLC34A2* is

expressed in all tissues not surprisingly, it is most highly expressed in human lungs and trachea.[70,71] Interestingly, surfactant production and increased surfactant levels in amniotic fluid has been shown to be a trigger for human parturition.[72,73] While this is a possible mechanistic explanation how perturbation is *SLC34A2* could be involved in PTB, surfactant is produced by the fetus and the association was observed in the mothers. One could argue that this variant is being over-transmitted and/or preferentially transmitted to affected fetuses; however, this will require fetal genotyping and should be considered a future direction.

The gene most 5' to rs871476 is the anaphase promoting complex subunit 4, *ANAPC4*. ANAPC4 is involved in the control of progression through mitosis and the G1 phase of the cell cycle. How perturbation in this gene could be involved in PTB is unclear, but a GWAS from 2010 found an intronic SNP in *ANAPC4* to be the strongest associated with weight (p-value = 1.44e-6) in a population of Filipino women.[74] The Illumina Human Body Map 2.0 (http://www.ensembl.info/blog/2011/05/24/human-bodymap-2-0-data-from-illumina/), identified expression of *ANAPC4* in human cervix. While level of expression in the cervix was not higher than the other tissues investigated, perturbation of this protein in the cervix could be a potential mechanism for how this gene could be involved in PTB.

Table 2.1: **All SNPs with $10^{-6}$ level of significance for the unadjusted additive linear regression for the Box-Cox transformed gestational age in Finnish mothers genotyped on the Affymetrix 6.0 SNP array.** Base pair (BP) positions refer to NCBI36 (hg18, March 2006 assembly) build of the human genome.

| CHR | SNP | BP | GENE | LOCATION | MINOR ALLELE | BETA | L95 | U95 | P |
|---|---|---|---|---|---|---|---|---|---|
| 1 | rs10874644 | 102931572 | *COL11A1* | 3'downstream/intergenic | G | 2.34E+09 | 1.34E+09 | 3.34E+09 | 5.26E-06 |
| 2 | rs7583085 | 42876472 | *HAAO* | 5' upstream | G | 1.78E+09 | 1.02E+09 | 2.54E+09 | 5.77E-06 |
| 2 | rs7569325 | 42978493 | *HAAO* | 5' upstream | G | 1.81E+09 | 1.02E+09 | 2.59E+09 | 7.54E-06 |

Table 2.2**: All SNPs with a genotypic p-values ≤ $10^{-7}$ level of significance for Box-Cox gestational age and the subsequent dominant and recessive model p-values. The SNPs are sorted by GENO p-value.** Base pair (BP) positions refer to NCBI36 (hg18, March 2006 assembly) build of the human genome.

| CHR | SNP | BP | GENE | LOCATION | GENO P | DOM P | REC P |
|---|---|---|---|---|---|---|---|
| 8 | rs10104530 | 38682002 | *TACC1* | 5' upstream | 7.09E-08 | 0.2987 | 9.45E-07 |
| 10 | rs7094463 | 114701973 | *TCF7L2* | intron | 4.88E-07 | 6.37E-04 | 1.10E-03 |
| 15 | rs1568209 | 90382294 | *SLCO3A1* | intron | 4.90E-07 | 1.82E-04 | 0.0199 |
| 15 | rs11074035 | 90381355 | *SLCO3A1* | intron | 9.18E-07 | 2.49E-04 | 0.0208 |

Additionally, it is overly simplistic to believe that rs871476 is regulating genes that are over 100kb away or only these genes. To investigate this we interrogated the web expression quantitative trait loci (eQTLs) database, SCAN,[75] searching for rs871476 and found that this SNP is an eQTL controlling expression of *ACCS* (p-value = 7e-5) in the HapMap Northern and Western European ancestry (CEU) population. Unfortunately this gene, 1-aminocyclopropane-1-carboxylate synthase homolog (Arabidopsis), is believed to be non-functional and therefore is unlikely to be involved in the etiology of PTB.

In an attempt to increase statistical power, we performed linear regression using gestational age as the phenotype.[67] After we utilized a Box-Cox power transformation to transform our gestational ages and analyzed our data using an additive linear regression the most significant SNP was found to be rs10874644 (p-value = 5.26e-6), which is ~185kb 3' downstream of collagen, type XI, alpha 1 (*COL11A1*). This is one of the two alpha chains of type XI collagen, a minor collagen. This gene is ubiquitously expressed in all tissues examined via mRNA microarray.[70,71] While this gene has not previously been implicated in preterm birth or gestational age control, mutations in *COL11A1* have previously been associated with type II Stickler syndrome,[76-79] Marshall syndrome[79-82] and an increased risk for lumbar disc herniation.[83] None of these previously associated diseases have pregnancy/fertility phenotypes or at increased risk for PTB.

Once again additionally, this SNP is a far distance from the closest gene so it may very well exert its functional effect or be in LD with the causative variant

which could have its functional effect on a completely different gene possibly even on a different chromosome.

Examining the genotypic, dominant and recessive models after the Box-Cox transformation we observed that the same gene, *TACC1*, has the most significant genotypic, p-value = 7.09e-8. We also observe transforming, acidic coiled-coil containing protein 1 (*TACC1*) under the recessive model change from the third most significant to the most significant SNP with a p-value of 9.45e-7.

How *TACC1* is involved in the etiology of PTB or control of gestational age is currently unknown. *TACC1* is believed to be involved in the process that promotes cell division prior to tissue differentiation. Microarray expression analysis has found this gene to be expressed in all tissue types interrogated; however, one of the tissues with the highest expression was human uterus.[70,71] The fact that one of the tissues with the highest expression was the uterus could help bolster the argument that it is involved in PTB or gestational age control. *TACC1* has currently been shown to be dysregulated in breast,[84-86] ovarian[87] and gastric[88] carcinomas.

More interesting however, was the gene with the second most significant recessive model Box-Cox transformed gestational age, *TCF7L2* (p-value = 1.10e-3). Transcription factor 7-like 2 (*TCF7L2*), as the name implies is a transcription factor that plays a role in the Wnt signaling pathway.[89] Expression analysis has

found *TCF7L2* mRNA expression in all tissues examined, but one of the organ systems with higher expression is the female reproductive system.[70,71]

This gene has been previously been associated with type 2 diabetes risk,[90-106] glucose control,[107,108] metabolic syndrome,[109] coronary heart disease[110] and most importantly for us birth weight.[111] While SNPs in this gene have not yet been associated with gestational age if a carrier of a SNP which has been shown to alter birth weight that could indirectly control gestational age due to the uterine stretch response.[18,112]

The SNP that we observed the strongest association in *TCF7L2* in this study, rs7094463, which promoted longer gestational ages, was not the same SNP, which was associated with increased birth, weight (rs7903146) and in our dataset these two SNPs were not in LD ($r^2 = 0.038$).

In conclusion, we present our preliminary results for our Finnish mothers GWAS investigating preterm birth and gestational age. Due to our lack of an independent dataset to replicate/generalize our most significant result all associations presented should be considered tentative. Our study actually creates more questions than it answers, but we present it with the hope that it will motivate others to interrogate their datasets for associations with our top-hits with the ultimate goal of understanding preterm birth etiology and the control of gestational age.

## Materials and methods

*Sample collection*

Mothers from preterm or term deliveries were enrolled for genetic analysis by methods approved by the Institutional Review Boards/Ethics Committees of University Central Hospital, Helsinki and the University of Oulu. DNA was extracted from whole blood or Oragene® saliva kits. Standard manufacturer protocols were followed. All families filled out a detailed questionnaire, which included questions about their current pregnancy, past pregnancies and family histories.

*Sample inclusion and exclusion criteria*

For the Affymetrix Genome-Wide Human SNP Array 6.0 the inclusion and exclusion criteria are as described below. Case mothers were required to have been delivered spontaneously preterm between 22-36 weeks gestation and additionally either the child or mother had to have a first degree relative with a history of PTB, or a spontaneous idiopathic preterm delivery less than 35 weeks of gestation. Mothers were excluded if they had any medical indication for a preterm delivery, where part of a multiple gestation pregnancy, or other identified risk from preterm delivery such as recent trauma or clinical evidence of infection.

*Finnish mothers Affymetrix Genome-Wide Human SNP Array 6.0*

750ng of genomic DNA was sent to the Vanderbilt DNA microarray core or Washington University in St. Louis School of Medicine Genome Sciences Center where they performed QC and processed the Affymetrix arrays per the manufacture's protocol. Affymetrix® Genotyping Console™ software (GTC) version 4.0 was used to perform initial quality control (QC) and to make genotype calls. Briefly, all CEL files, both Washington University and Vanderbilt samples, were imported into GTC and initial QC determined which arrays are acceptable for genotype calling. Next, the Birdseed V2 genotype-calling algorithm that is embedded in GTC was used to call genotypes from the passing CEL files creating a Chp file containing all of the genotypes. The genotypes were exported as plain text files using a Perl script written by one of our collaborators at Washington University, Dr. Justin Fay, was used to create PLINK formatted MAP and PED files from the GTC output. PLINK v1.07 was used to perform QC using the following filters: 1.) Exclude SNPs with < 95% genotype efficiency, 2.) Remove samples with < 95% genotypes, 3.) Filter with minor allele frequency (MAF) < 0.05, 4.) Hardy Weinberg Equilibrium (HWE) < 0.0001 in controls. After data cleaning, our Finnish mother dataset consisted of 539 individuals of which 252 were cases and the remaining 287 were controls. The Affymetrix Genome-Wide Human SNP Array 6.0 consists of ~906,600 single nucleotide

polymorphisms (SNPs) and after standard quality control (QC) filters there were 561,598 SNPs remaining in our Finnish mothers dataset for interrogation.

*Statistical analysis*

All association analysis was performed using the software package PLINK v1.07. SPSS18 was used to perform one-sample Kolmogorov-Smirnov tests and plot histograms of gestational ages. Stata 11.1 and its "boxcox" function was used to determine the ideal lambda for the Box-Cox power transformation of gestational ages. The R software package was used to create QQ plots. All p-values are two-tailed unless noted otherwise.

# CHAPTER III

## FETAL VARIANTS IN THE *INF2* REGION ARE ASSOCIATED WITH GESTATIONAL AGE AND PRETERM BIRTH‡

### Introduction

While there is growing evidence that genetics are involved in the etiology of preterm birth (PTB), we can currently ascribe a role to the genome in a small percentage of cases.[16,60] Traditionally, PTB genetic studies focus on candidate gene and pathways studies, and while there have been some significant associations found many of these have failed to replicate.[60] Besides not replicating, another weakness of prior candidate gene/pathway studies is that they are limited by our current biological understanding.

Segregation analysis and genetic modeling found that models where PTB risk was conferred by the maternal genome to be the most parsimonious and best fitting.[16,51,53] However, all offspring of mothers' who have experienced a preterm delivery are not delivered preterm, though, so there is almost certainly a fetal genome effect.[53]

There have been a handful number of studies examining the role of fetal candidate genes and pathways for PTB risk;[58,113-120] in this study we utilized an agnostic approach and report to our knowledge the first fetal genome-wide

‡ Adapted from McElroy et al. In Preparation

association study (GWAS) in examining two phenotypes: dichotomous PTB (case/control) and the quantitative trait gestational age. Here we report a SNP located within an intron of the Ensembl gene, *ENSG0256050*, and 5' upstream to *INF2,* rs7153053, to be significantly and reproducibly associated with PTB and gestational age. This association may elucidate new biology and possibly medical interventions for prevention of PTB.

## Results

*Study population characteristics*

The Helsinki infant sample consisted of a total of 471 individuals (159 cases/312 controls) of which 236 were male and the remaining 235 were female. The software package SPSS Statistics V.20, was used to test for differences in demographic data between the case and control infants. The variables that were tested for differences include maternal age, body mass index (BMI), parity, gravidity, birth weight, birth length, alcohol use and smoking use. Besides the expected significant differences in birth weight and birth length between cases and controls, BMI (p-value = 0.015), gravidity (p-value = 0.016) and tobacco use (p-value = 0.003) differed significantly between groups and will be adjusted for in our analysis (Table 3.1). P-values were determined using a one-way ANOVA.

Table 3.1: **Demographic information for Helsinki infants genotyped on the Illumina Omni2.5 BeadChip**. Numbers in the table are mean (standard deviation) except for dichotomous variables where percentages were used. P-values were determined using a one-way analysis of variance (ANOVA).

| Variable | Preterm (n=159) | Term (n=312) | P-value |
|---|---|---|---|
| Maternal age (yr) | 31.3 (5.0) | 31.4 (4.1) | 0.767 |
| Body Mass Index (kg/m$^2$) | 23.6 (4.6) | 22.7 (3.3) | 0.015 |
| Parity (n) | 1.6 (0.093) | 1.5 (0.71) | 0.080 |
| Gravidity (n) | 2.1 (1.4) | 1.9 (1.0) | 0.016 |
| Birth weight (g) | 2389.9 (480.9) | 3577.1 (417.6) | <0.0001 |
| Birth length (cm) | 45.3 (2.6) | 50.3 (1.9) | <0.0001 |
| Alcohol use (%) | 3.9% | 1.6% | 0.123 |
| Tobacco use (%) | 9.2% | 2.9% | 0.003 |

*Preterm birth additive logistic regression adjusted for BMI, gravidity and smoking status association analysis in Helsinki infants on the Illumina Omni2.5 BeadChip*

After quality control (QC) of the Illumina Omni2.5 BeadChips, there were 1,695,052 single nucleotide polymorphism (SNPs) remaining for interrogation.

The initial analysis we preformed was an adjusted additive logistic regression for PTB adjusting for three factors, which were shown to differ significantly. A quantile-quantile plot (QQ) showed almost no deviation from the expected distribution (Figure 3.1). There is one SNP with an adjusted additive logistic regression p-value at the $10^{-7}$ level of significance. The strongest associated SNP, rs7153053, was located in an intron of an Ensembl gene,

*ENSG0256050,* and 5′ upstream of inverted formin, FH2 and WH2 domain containing (*INF2*) (p-value = 5.72e-7; OR = 2.11). For a complete list of all SNPs with p-values < $10^{-4}$ please see Appendix D. For comparison purposes, please see Appendix E for all SNPs with p-values < $10^{-4}$ for the unadjusted additive logistic regression.

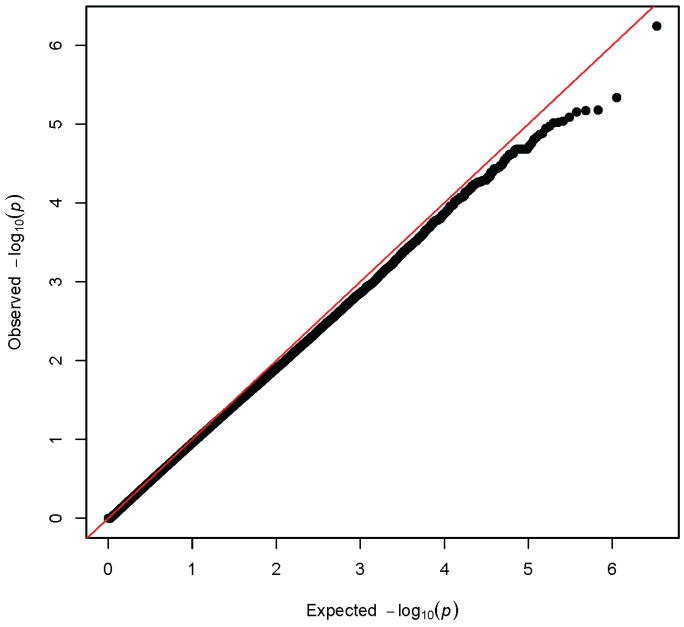


Figure 3.1: **QQ plot of additive logistic regression adjusted for body mass index, gravidity and smoking status association analysis for dichotomous preterm birth phenotype in our Helsinki infants Omni2.5 dataset**

*Preterm birth genotypic logistic regression adjusted for BMI, gravidity and smoking status association analysis in Helsinki infants on the Illumina Omni2.5 BeadChip*

Because there is no a priori evidence to conclude that PTB follows an allelic model, we re-analyzed our Helsinki infant dataset under a genotypic

model. We once again adjusted for the three covariates; BMI, gravidity and smoking that were identified to differ significantly between cases and controls.

Our most significantly associated SNP, and only one with a single SNP at the $10^{-7}$ level of significance is rs7153053, genotypic p-value = 7.40e-7. This is the same SNP as observed under the additive model, but p-value is slightly less significant, 5.72e-7 vs. 7.40e-7, almost certainly not statistically different. However, because the p-values were so similar we concluded that the additive adjusted logistic regression was the model to use because it had very similar p-values and also saved us a degree of freedom. For a complete list of all SNPs with p-values < $10^{-4}$ please see Appendix F. For comparison purposes, please see Appendix G for all SNPs with p-values < $10^{-4}$ for the unadjusted genotypic logistic regression.

*Box-Cox transformation of gestational age in Helsinki infants*

Since our dataset is considered small for a GWAS, to gain additional statistical power,[67] we also analyzed the Helsinki infant dataset, 471 total samples 159 born in the preterm range (<37 weeks) and 312 term or (37-41 weeks) using the quantitative trait gestational age (in days).

One of the main assumptions of linear regression is that the dependent variable, in this case gestational age, is normally distributed. To test this we performed a one-sample Kolmogorov-Smirnov test on the gestational ages and

discovered that our gestational ages were very skewed, p-value = 1.07e-67 (Figure 3.2).

In order to normalize this distribution, we performed a Box-Cox power transformation.[121] The lambda was determined using the Stata software package "boxcox" function with the rs7153053 genotypes coded additively and was determined to be 7.639 The gestational age was transformed using the original gestational age^7.639 and the distribution was graphed again and appeared more normally distributed (Figure 3.3).

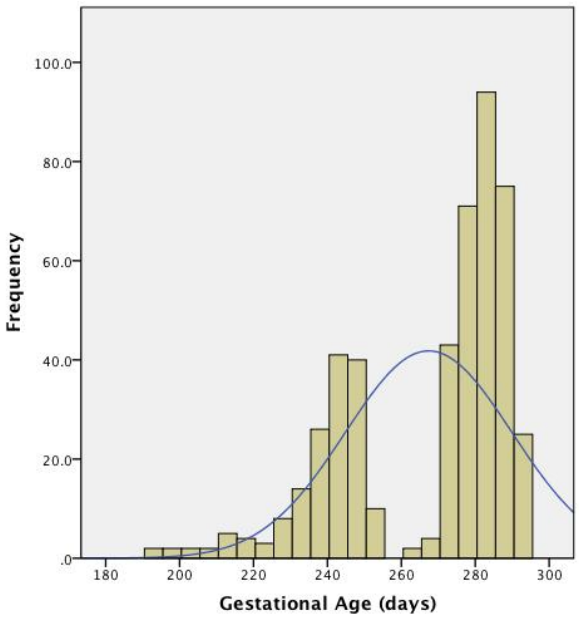


Figure 3.2: **Distribution of Helsinki infant gestational age in days with a normal curve overlaid**. The gestational ages are not normal which (one-sample Kolmogorov-Smirnov Test p-value = 1.07e-67). In order to try to normalize our gestational ages we performed a Box-Cox power transformation and used that variable in the linear regression.

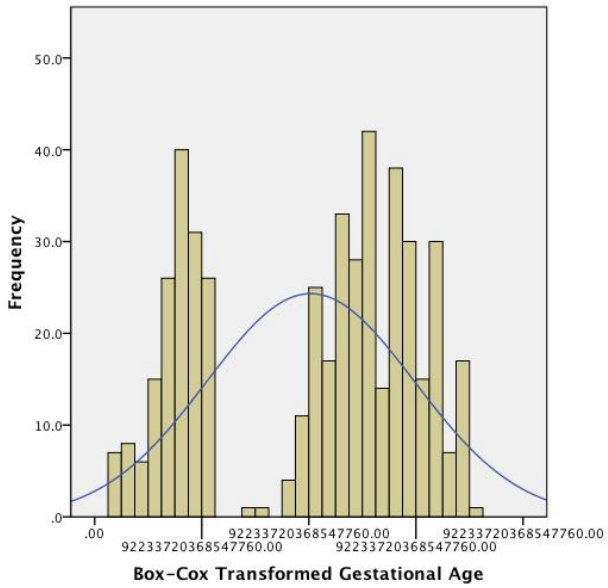


Figure 3.3: **Distribution of Helsinki infant Box-Cox transformed gestational age with a normal curve overlaid.** Using the Stata 12.1 software package the optimal lambda for the Box-Cox power transformation was determined to be 7.639. The transformed variable was then graphed and a normal curve was overlaid

*Box-Cox transformed gestational age adjusted for BMI, gravidity and smoking status linear regression in Helsinki infants*

The first regression that we performed was an adjusted additive model linear regression in which we adjusted for the three variables found to differ significantly between cases and controls: BMI, gravidity and smoking. We observed only a single SNP at the $10^{-8}$ level of significance. The strongest association was once again observed for rs7153053, which was also the SNP with the strongest association observed for the dichotomous PTB analysis, p-value = 6.28e-8, and was risk promoting i.e. earlier gestational age. For the complete list of SNPs with p-values < $10^{-4}$ please see Appendix H. The QQ plot illustrates that

the results do not deviate from the expected until we observe a slight underrepresentation in more highly significant associations and then there is the single SNP, rs7153053, which lies above the expected line (Figure 3.4).

Next, we ran an adjusted genotypic model linear regression for the Box-Cox transformed gestational age in which we were interested in SNPs in which the genotypic 2-degree of freedom have highly significant p-values. Just like for the PTB analysis, there is no evidence that gestational age needs to follow an additive allelic model, which is why we performed this investigation. This analysis was adjusted for the same variables as the additive linear regression. We utilized a genotypic p-value $\leq 10^{-7}$ as highly significant, we observed only a single SNP, rs7153053, p-value = 5.96e-8. This SNP, rs7153053, was then tested under the more defined dominant and recessive models. The most significant association was observed under the recessive model, p-value = 1.56e-7.

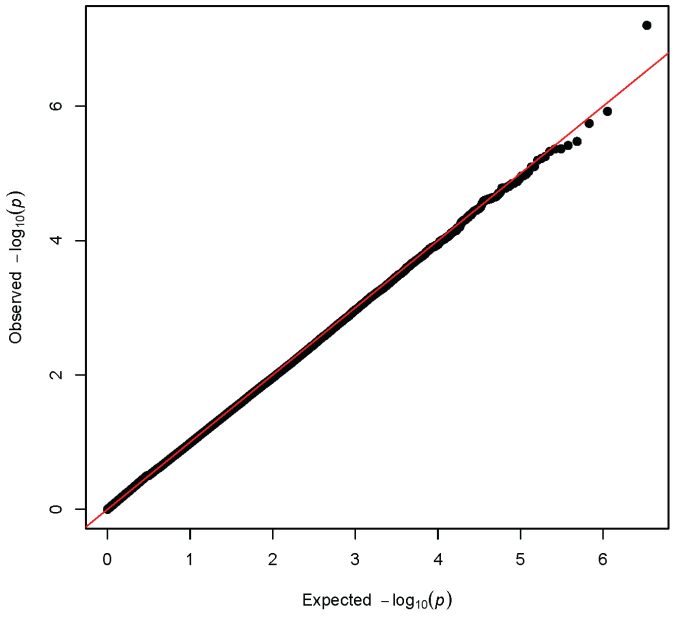


Figure 3.4: **QQ plot of the results of the adjusted additive linear regression using Box-Cox transformed gestational age in our Helsinki infants Omni2.5 dataset.** The covariates, which we adjusted for included, body mass index, gravidity and smoking.

*TaqMan genotyping of rs7153053 in an Oulu infant cohort*

In an attempt to replicate our most significant result, we used Applied Biosystems TaqMan to genotype our most significant SNP, rs7153053, in an Oulu infant cohort, which consisted of 310 cases (22.7-35.9 weeks) and 180 controls (37.3-41.9 weeks) of which 271 were male and the remaining 216 female. Unfortunately, unlike for our Helsinki infant cohort we did not have access to as extensive demographic information for these samples; therefore, these analyses discussed below were not adjusted for any covariates. After discussion, we felt as

if it was more appropriate to perform unadjusted analysis than to lose power by having to remove samples for which we were missing demographic data and not being able to test for significant differences in variables known to be/suggested to be involved in PTB or gestational age.

We investigated both the quantitative trait gestational age (in days) and dichotomous preterm birth phenotype. Before running an unadjusted linear regression for gestational age, we tested the normality of the gestational ages using the one-sample Kolmogorov-Smirnov test and the data was found to be skewed with a p-value = 1.43e-8 (Figure 3.5). Using the Stata software packing "boxcox" function and rs7153053 coded additively the ideal lambda was

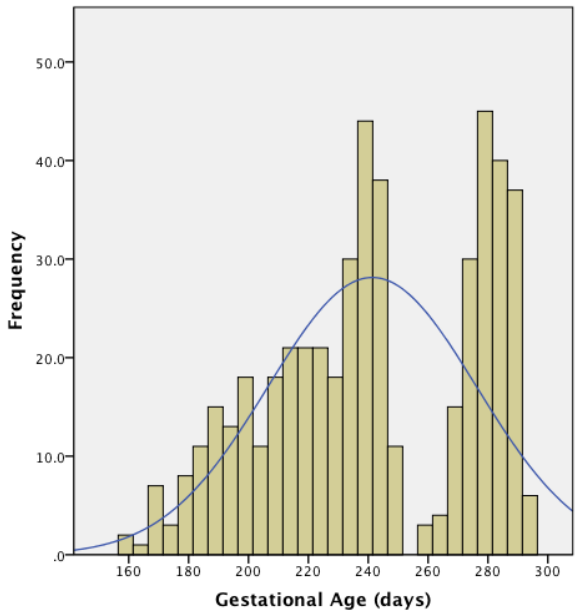


Figure 3.5: **Distribution of Oulu infant gestational ages in days with a normal curve overlaid**. The gestational ages are not normal (one-sample Kolmogorov-Smirnov Test p-value = 1.43e-8). In order to try to normalize our gestational ages we performed a Box-Cox power transformation and used that variable in the linear regression

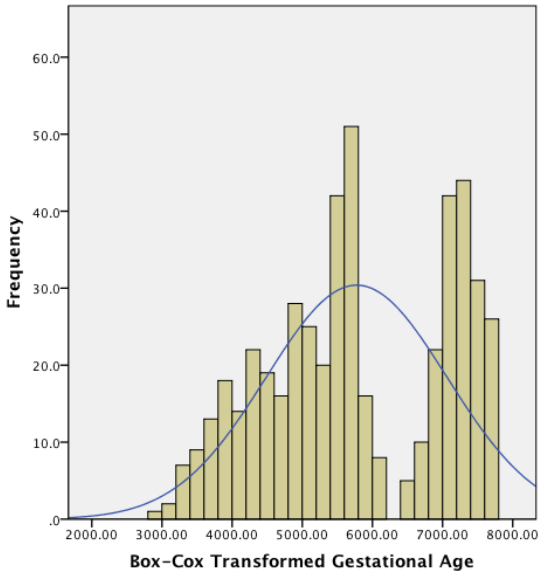


Figure 3.6: **Distribution of Oulu infant Box-Cox transformed gestational age with a normal curve overlaid.** Using the Stata 12.1 software package the optimal lambda for the Box-Cox power transformation was determined to be 1.577. The transformed variable was then graphed and a normal curve was overlaid.

determined to be 1.577 and the transformed variable was computed by raising the original gestational age to the 1.577 power. The new Box-Cox gestational age was graphed and appeared slightly more normally distributed (Figure 3.6). The one-tailed unadjusted additive linear regression was significant with a p-value = 0.040 with the effect in the same direction as observed in the Helsinki infants. We believed that the one-tailed test was appropriate due to the fact that we had a priori knowledge that we expected this SNP to be risk promoting i.e. minor allele more common in the cases.

For PTB, a simple chi-squared association analysis was performed and the one-tailed p-value = 0.059; OR = 1.23. While this does not meet the traditional p-value threshold of ≤ 0.05, the effect is trending in the same direction, risk promoting. Taken together, we believe this should be considered a positive replication of rs7153053 identified in our Helsinki infant GWAS.

## Discussion

To our knowledge, this is the fist reported GWAS in preterm birth infants. Additionally, it is the first study to observe an association between rs7153053 and PTB and gestational age.

The one SNP, rs7153053, that was consistently the strongest associated is located within an intron of the Ensembl gene, *ENSG0256050,* and ~5.4kb 5' upstream of *INF2*. While *ENSG0256050* is not a traditional protein-coding gene, it is predicted to be a long intergenic non-coding RNAs (lincRNA). This class of genes is emerging to be involved in a number of key cellular processes and have the potential to regulate a number of genes.[122] While the exact role of *ENSG0256050* remains to be elucidated, other lincRNAs have been discovered to be involved in X chromosome inactivation,[123] imprinting[124,125] and development.[126,127] Most importantly for this investigation, one can make the argument *ENSG0256050* is important for PTB risk and control of gestational age length due to its potential role controlling imprinting and/or fetal development. Unfortunately, tissue specific expression of *ENSG0256050* remains to be

performed therefore it is unknown if this gene is expressed in tissues which we believe would be important for PTB risk and/or gestational age control such as the fetal membranes or placenta.

While *ENSG0256050* remains somewhat of an enigma, there is more known about inverted formin, FH2 and WH2 domain containing (*INF2*). This protein, like all formins, is involved in the regulation of proteins that accelerate actin polymerization; however, INF2 is unique it its ability to additionally accelerate actin depolymerization *in vitro*.[128,129] Actin is ubiquitously expressed in almost all eukaryotic cells including the fetal membranes (amnion and chorion) and placenta; therefore any depletion of actin integrity could be important for the pathogenesis of PTB and/or controlling gestational age timing. Further strengthening its potential role in PTB, *INF2* is expressed in the human placenta.[130] Any perturbation of placental integrity or implantation can be devastating.[131]

While this is to the best of our knowledge the first report of a SNP near *INF2* being associated with preterm birth and gestational age, it is not the first time this gene has been implicated in human disease pathogenesis. A study by Brown et al. in 2009, identified missense variants in *INF2* as the cause of autosomal dominant focal segmental glomerulosclerosis (FSGS).[130] Whether or not families with the autosomal dominant form of FSGS caused by *INF2* variants are at an increased risk of preterm delivery is not known. Because women with FSGS are advised not to become pregnant because of increased health risks, it

would require women harboring *INF2* missense variants who had incomplete penetrance to help answer whether the variant in *INF2* was playing a role in PTB age gestational age determination.

Our study is not without limitations. Our replication cohort is small and to further validate this report our association needs to be replicated/generalized in an independent dataset. An additional limitation of our replication cohort is the lack of dense demographic data. If these were available we would be able to test for significant differences between potential covariates and then adjust for any covariate illustrated to differ significantly in our association models. Furthermore, our most significantly associated SNP, rs7153053, is almost certainly not the causal variant and is acting as a surrogate and "tagging" the true causative variant. In order to find the true "causative" variant, DNA sequencing of the region around rs7153053 will be required and should be a future direction.

In conclusion, we report the first genome-wide association study in preterm birth infants and report an association between a SNP within an intron of the Ensembl gene, *ENSG0256050,* and ~5.4kb 5' upstream of *INF2,* rs7153053, to be reproducibly associated with PTB and Box-Cox transformed gestational age. We believe this SNP may elucidate new biology and possibly medical interventions for prevention of PTB; however, further functional studies will be required to make this a reality.

## Materials and methods

*Sample collection*

Infants from preterm or term delivery were enrolled for genetic analysis by methods approved by the Institutional Review Boards/Ethics Committees of University Central Hospital, Helsinki and the University of Oulu. DNA was extracted from whole blood or Oragene® saliva kits. Standard manufacturer protocols were followed.

*Sample inclusion and exclusion criteria*

For the Illumina Omni2.5 BeadChips and Applied Biosystems TaqMan SNP genotyping the inclusion and exclusion criteria are as described below. Case infants were required to have been delivered spontaneously preterm between 22-36 weeks gestation and additionally either the child or mother had to have a first degree relative with a history of PTB, or a spontaneous idiopathic preterm delivery less than 35 weeks of gestation. Infants were excluded if their mother had any medical indication for a preterm delivery, where part of a multiple gestation pregnancy, or other identified risk from preterm delivery such as recent trauma or clinical evidence of infection.

*Helsinki infant Illumina Omni2.5 BeadChips*

750ng of genomic DNA was sent to the Vanderbilt DNA microarray core where they performed QC and processed the Illumina arrays per the manufacture's protocol. The GTC created PLINK ready BED files from the processed array genotype calls. QC was performed in PLINK. Monomorphic SNPs were removed by using a minor allele frequency filter (MAF) of 0.000001%. We chose to not filter by Hardy-Weinberg equilibrium (HWE), but HWE checks were performed in order to see if our most significant SNPs were out of HWE in controls. SNPs which were found to have a HWE $p < 0.0001$ in controls were excluded from further analysis.

*Applied Biosystems TaqMan genotyping*

Genomic DNA from Oulu was sent to the Cincinnati Children's Hospital Medical Center Genetic Variation and Gene Discovery Core Facility. Standard manufacturers protocols were followed and for each assay 10ng of genomic DNA was used.

*Statistical analysis*

All association analysis, chi-squared, logistic regression and linear regression were performed using the software package PLINK v1.07. SPSS18 was

used to perform one-sample Kolmogorov-Smirnov tests and plot histograms of gestational age and Box-Cox transformed gestational ages. The software package SPSS Statistics V.20, was used to test for differences in demographic data between the case and control infants and p-values were determined using a one-way ANOVA. Stata 12.1 and its "boxcox" function were used to determine the ideal lambda for the Box-Cox power transformation of gestational ages. The R software package was used to create all of the QQ plots. All p-values are two-tailed unless noted otherwise.

# CHAPTER IV

## FETAL CODING REGION VARIANTS IN ADAM METALLOPEPTIDASE DOMAIN 29 (*ADAM29*) ARE SIGNIFICANTLY ASSOCIATED WITH INCREASED BIRTH WEIGHT Z-SCORES[§]

### Introduction

Birth weight is complex trait that is controlled by a number of factors, the strongest, which is gestational age. Other factors known to influence birth weight include maternal age, ethnicity, parity, body mass index (BMI), education and smoking.[132,133] Both extremes of birth weight are at increased risk for perinatal morbidity and mortality.[134-136] Additionally, low birth weight infants are at an increased risk of type 2 diabetes, cardiovascular disease and hypertension in adulthood.[137,138]

A number of previous twin and family studies have illustrated that birth weight has a genetic component.[139-142] While the intrauterine environment, controlled by the maternal genome and environmental factors, almost certainly plays an important role in birth weight, the correlation between paternal height or weight and offspring birth weight also illustrates the importance of fetal genetics on birth weight.[132,143-145] Heritability estimates have been shown to vary depending on the gestational age and decreases with increasing gestational age, 38% at 25 weeks gestation to only 15% at 42 weeks.[146]

[§] Adapted from McElroy et al. In Preparation

Prior studies have identified a handful of genes associated with birth weight, most of which are also associated with type 2 diabetes, including: insulin-like growth factor binding protein 3 (*IGFGP3*),[147] peroxisome proliferator-activated receptor-γ (*PPARG*)[148] and transcription factor 7-like 2 (*TCF7L2*).[111] While the overlap between genes associated with type 2 diabetes and birth weight may signify important overlap in the underlying etiology/pathophysiology and pleiotropic effect of variants in these genes, most of these associations have not been replicated in independent datasets.[149-152]

In 2010, a genome-wide association study (GWAS) meta-analysis by Freathy et al. identified two variants that were associated with decreased fetal birth weights.[153] One of these variants, rs9883204, was in an intron of adenylate cyclase type 5 (*ADCY5*) and the other, rs900400, located between cyclin-L1 (*CCNL1*) and leucine, glutamate and lysine rich 1 (*LEKR1*). Further strengthening the link between birth weight and type 2 diabetes correlated SNPs in *ADCY5* have been associated with glucose control and risk for type 2 diabetes.[107] A recent study, found a modest replication of rs900400 in a meta-analysis of preterm infants.[154]

Here we report a GWAS in which we identified a robust association between birth weight z-scores and SNPs in the coding region of ADAM metallopeptidase domain 29 (*ADAM29*) in a Finnish infant dataset from Helsinki that includes both preterm and term infants. Our study is somewhat different from the earlier investigations of birth weight because we used normalized birth

weight z-scores instead of the absolute birth weight. Using birth weight z-scores allowed us to include infants that were born both term and preterm. To our knowledge, this is the first report to show this link between coding region *ADAM29* SNPs and increased birth weight z-scores. We believe that these associations may elucidate new biology.

## Results

*Study population characteristics*

The Helsinki infant sample consisted of a total of 471 individuals (159 cases/312 controls) of which 236 were male and the remaining 235 were female. The software package SPSS Statistics V.20, was used to test for differences in demographic data between the case and control infants. The variables that were tested for differences include maternal age, body mass index (BMI), parity, gravidity, birth weight, birth length, alcohol use and smoking use. Besides the expected significant differences in birth weight and birth weight between cases and controls, BMI (p-value = 0.015), gravidity (p-value = 0.016) and tobacco use (p-value = 0.003) differed significantly between groups and will be adjusted for in our analysis (Table 4.1). P-values were determined using a one-way ANOVA.

Table 4.1: **Demographic information for Helsinki infants genotyped on the Illumina Omni2.5 BeadChip**. Numbers in the table are mean (standard deviation) except for dichotomous variables where percentages were used. P-values were determined using a one-way analysis of variance (ANOVA).

| Variable | Preterm (n=159) | Term (n=312) | P-value |
|---|---|---|---|
| Maternal age (yr) | 31.3 (5.0) | 31.4 (4.1) | 0.767 |
| Body Mass Index (kg/m$^2$) | 23.6 (4.6) | 22.7 (3.3) | 0.015 |
| Parity (n) | 1.6 (0.093) | 1.5 (0.71) | 0.080 |
| Gravidity (n) | 2.1 (1.4) | 1.9 (1.0) | 0.016 |
| Birth weight (g) | 2389.9 (480.9) | 3577.1 (417.6) | <0.0001 |
| Birth length (cm) | 45.3 (2.6) | 50.3 (1.9) | <0.0001 |
| Alcohol use (%) | 3.9% | 1.6% | 0.123 |
| Tobacco use (%) | 9.2% | 2.9% | 0.003 |

*Birth weight z-score linear regression adjusted for BMI, gravidity and smoking status in Helsinki infants genotyped on Illumina Omni2.5 BeadChips*

Due to the smaller size of our dataset, 471 infants of whom 312 were born in the period considered term gestation (37-41 weeks) and 159 preterm infants ranging in gestational age from ~28 to ~36 weeks gestation, we chose to use the quantitative phenotype birth weight z-scores to increase our power to detect an association.[67] Linear regression was adjusted for the variables shown to differ significantly between infants born in the term and preterm period: body mass index (BMI), gravidity and smoking.

We used birth weight z-scores instead of the absolute birth weights as our phenotype/dependent variable because when they are converted to z-scores they are normally distributed, a requirement for linear regression. We calculated the

birth weight z-scores by comparing the sample birth weights to a population sample of around 7,000 Finish births.[155] A one-sample Kolmogorov-Smirnov test was performed and the normality assumptions of the z-scores was valid, p-value = 0.171.

Under the adjusted additive linear regression, there were 5 SNPs with p-values at the level of $10^{-7}$ (Table 4.2). However, all of these SNPs are in very high linkage disequilibrium (LD) ($r^2$ = 0.98-1.0) therefore should conceptually be considered a single robust association in *ADAM29* which increases birth weight z-score, beta = 0.30, most significant p-value = 4.25e-7 (Figure 4.1). Examining the quantile-quantile plot (QQ) showed almost no deviation from the expected distribution except for the five SNPs discussed above (Figure 4.2). This is also illustrated when we examine a "Manhattan" plot and the *ADAM29* SNPs are the only SNPs, which are highly significant and rise above the background (Figure 4.3). For a complete list of all SNPs with p-values < $10^{-4}$ please see Appendix I.

In addition to an additive model, we also tested for an association under a general 2-degree of freedom genotypic model and the more defined dominant and recessive models. The SNPs, which we carried forward and tested under more specific genetic models, were the SNPs that had genotypic p-values = $10^{-7}$ (Table 4.3). All of these SNPs with the exception of two, rs1864180 and KGP10390109, are the same as under the additive model. The first additional SNP (rs1864180) is in a long intergenic non-coding RNA (lincRNA), *CTC-454M9.1* while the later is intergenic. For all of the SNPs that were significant

under both the additive and genotypic model the additive model had the more significant p-values.

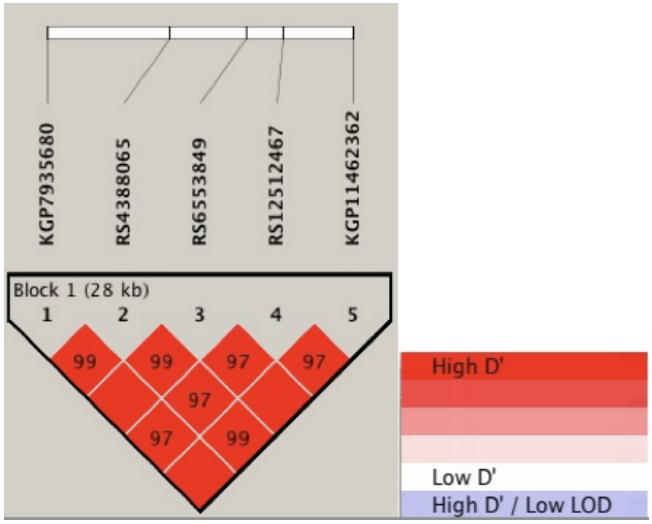


Figure 4.1: **A Haploview diagram illustrating the LD ($r^2$) values for the five SNPs in *ADAM29* for our Helsinki infants genotyped on the Illumina Omni2.5**. All five of the SNPs reside in the same 28kb LD block on chromosome 4

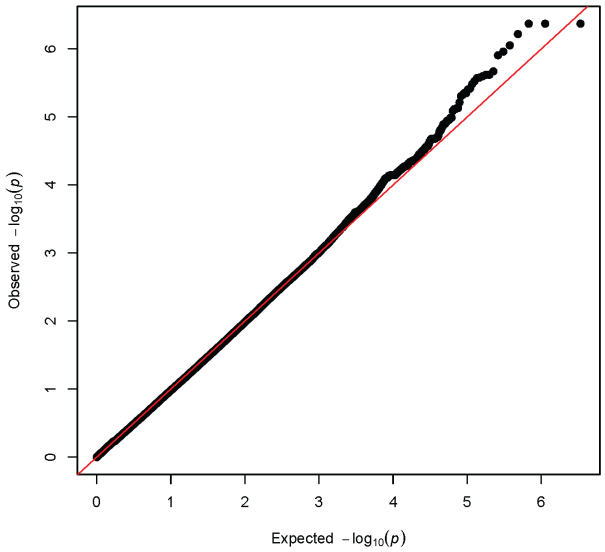


Figure 4.2: **QQ plot of the results of adjusted additive linear regression using birth weight z-scores in our Helsinki infants Omni2.5 dataset**. The variables adjusted for include body mass index (BMI), gravidity and smoking

*TaqMan genotyping of ADAM29 SNPs in Oulu infants*

Because all genetic association is tentative until replicated, in an attempt to replicate the results we observed in our Helsinki infants we used Applied Biosystems (ABI) TaqMan to genotype three of these SNPs, rs6553849, rs12512467 and rs4388065, in a cohort of infants from Oulu Finland. This cohort consisted of 310 cases (22.7-35.9 weeks) and 180 controls (37.3-41.9 weeks) of which 271 were male and the remaining 216 female. We chose to only run the unadjusted additive linear regression for birth weight z-scores. The reason why we were unable to perform adjusted linear regression is because of lack of

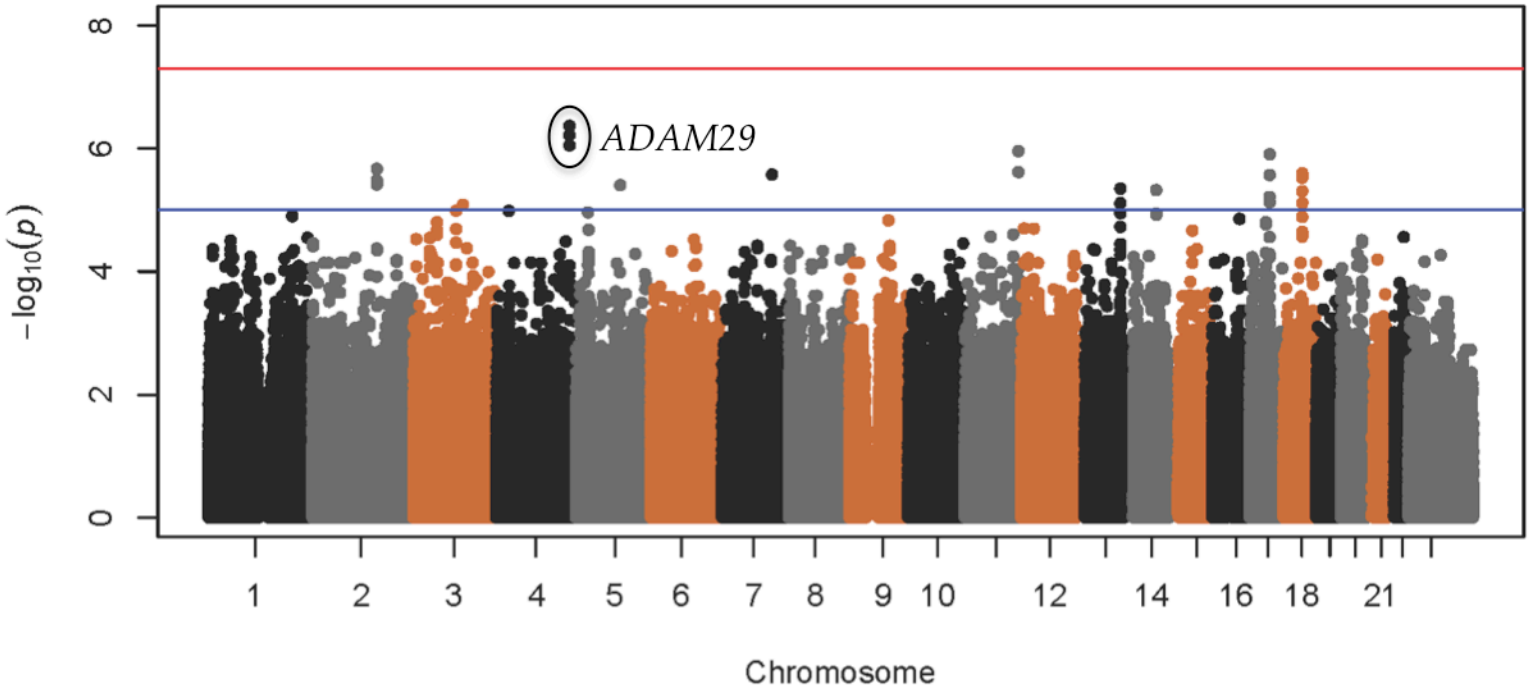


Figure 4.3**: Manhattan plot of Helsinki infants on Omni2.5 BeadChips adjusted additive linear regression for birth weight z-scores.** The variables adjusted for include body mass index (BMI), gravidity and smoking. The blue line is for p-value = 1e-5 and the red line is for p-value = 5e-8 which is considered genome-wide significant. The SNPs, which are circled, are the five most significant SNPs in *ADAM29*: KGP7935680, rs6553849, KGP11462362, rs12512467 and rs4388065.

demographic data for this Oulu infant dataset. Before regression was performed, genotype efficiency was tested using PLINK v1.07 and while both rs4388065 and rs6553849 were >97% unfortunately, rs12512467 had only an ~82% genotyping efficiency and was excluded from further analysis. Once again, the normality assumption of the birth weight z-scores was tested using a one-sample

Table 4.2**: Adjusted additive model linear regression for birth weight z-scores in our Helsinki infant dataset**. The variables adjusted for include body mass index (BMI), gravidity and smoking.

| CHR | SNP | GENE | LOCATION | MINOR ALLELE | BETA | L95 | U95 | ADJ P |
|---|---|---|---|---|---|---|---|---|
| 4 | KGP7935680 | *ADAM29* | Intron/5' upstream | T | 0.30 | 0.18 | 0.41 | 4.25E-07 |
| 4 | rs6553849 | *ADAM29* | Intron/5' upstream | G | 0.30 | 0.18 | 0.41 | 4.25E-07 |
| 4 | KGP11462362 | *ADAM29* | Intron | C | 0.30 | 0.18 | 0.41 | 4.25E-07 |
| 4 | rs12512467 | *ADAM29* | Intron | G | 0.29 | 0.18 | 0.41 | 6.04E-07 |
| 4 | rs4388065 | *ADAM29* | Intron/5' upstream | C | 0.29 | 0.18 | 0.40 | 8.85E-07 |

Table 4.3: **Adjusted linear regression for birth weight z-scores in Helsinki infants under a number of different models.** All SNPs that had a genotypic 2DF (GENO) p-value ≤ e-7 were tested under a dominant (DOM) and recessive (REC) model. The variables adjusted for include body mass index (BMI), gravidity and smoking. The base pair positions (BP) refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | GENO P-VALUE | DOM P-VALUE | REC P-VALUE |
|---|---|---|---|---|---|
| 5 | rs1864180 | 88421363 | 2.28E-07 | 0.01 | 2.85E-04 |
| 7 | rs17156420 | 102940582 | 6.46E-07 | 0.42 | 9.85E-07 |
| 7 | KGP3967285 | 103011509 | 6.54E-07 | 0.40 | 1.08E-06 |
| 7 | KGP4607969 | 102980732 | 6.54E-07 | 0.40 | 1.08E-06 |
| 7 | KGP5238014 | 103008646 | 7.51E-07 | 0.43 | 1.09E-06 |
| 2 | KGP10390109 | 155375235 | 8.54E-07 | 5.32E-05 | 4.18E-05 |

Kolmogorov-Smirnov test and found to be valid, p-value = 0.096.

For the unadjusted linear regression additive model for birth weight z-scores the most significant SNP was rs6553849 which had a one-tailed p-value = 0.096, beta = 0.093 (Table 4.4). Due to the a priori knowledge that we expected the minor allele of rs6553849 to increase birth weight z-scores we felt justified using one-tailed p-values. While this association does not meet the standard level of significance, p-value ≤ 0.05, it does meet the more liberal cutoff of p-value < 0.1. Additionally, the beta is in the same direction as in the initial Helsinki infant dataset i.e. the minor allele is weight promoting. We would consider this a tentative replication of our original GWAS association.

Table 4.4: **Unadjusted additive model linear regression for birth weight z-scores in our Oulu infant cohort.** Please note that p-values in this table are one-tailed.

| CHR | SNP | BP | GENE | LOCATION | MINOR ALLELE | BETA | L95 | U95 | UNADJ P |
|---|---|---|---|---|---|---|---|---|---|
| 4 | rs6553849 | 175839066 | *ADAM29* | Intron/5' upstream | G | 0.093 | -0.047 | 0.233 | 0.096 |
| 4 | rs4388065 | 175832109 | *ADAM29* | Intron/5' upstream | C | 0.087 | -0.051 | 0.224 | 0.109 |

## Discussion

There have been a number of studies investigating birth weight z-scores. However, until recently most of these datasets have only included infants born during the gestational age range considered term. While there maybe important biological rationale to do so such as different mechanisms of weight gain in

preterm vs. term infants; here we report the first genome-wide association study including both term and preterm infants investigating birth weight z-scores.

All of our strongest associations were observed in ADAM metallopeptidase domain 29 (*ADAM29*). ADAM29 is thought to play an important role in spermatogenesis and fertilization. Not surprisingly, the gene is strongly expressed in the testicles almost exclusively.[156] Explaining how *ADAM29* is involved in birth weight z-score control is still a mystery and will require a number of functional studies to try to illuminate a possible mechanism.

There have been a number of studies recently investigating the association of an intronic *ADCY5* SNP, rs9883204, and an intergenic SNP, rs900400, between *LEKR1* and *CCNL1* with birth weight. As in most of human genetics, the associations have been inconsistent in their replication/generalizability. Unlike the associations we observed with *ADAM29* both of these SNPs conferred the risk of lower birth weight z-scores. However, much like with *ADAM29* how the variant between *LEKR1* (~35kb) and *CCNL1* (~67kb), rs900400, is exactly playing a role in birth weight control is unknown. Interestingly however, this same variant and specific allele has been associated with decreased birth weight and placental weight in two prospective cohorts from The Netherlands and Australia. Each copy of the C allele for rs900400 beginning in the second trimester was associated with decreased fetal head circumference and femur length and in the third trimester in addition to the previously mentioned anthropometric measures smaller abdominal circumference and estimated fetal weight.[157] Unfortunately,

both of these SNPs were not on the Illumina Omni2.5 BeadChip therefore we were unable to test for an association in our Helsinki infants.

On the other hand, the *ADCY5* intronic variant, rs9883204, has been associated with insulin control in addition to birth weight and joins a number of genes (*IGFGP3*, *PPARG* and *TCF7L2*), which have been, associated with birth weight and the have pleiotropic effects controlling glucose control and type 2 diabetes risk. With the modicum of knowledge known about *ADAM29*, it does not seem to be involved in this mechanism/pathophysiology controlling birth weight. Thus new opportunity to elucidate a new potential mechanism controlling birth weight exists.

As is with all investigations, this study is not without important limitations. Even though our most significantly associated SNPs are located in the coding region of *ADAM29* it is simplistic to believe that the SNPs are only affecting *ADAM29*. When we interrogate the web expression quantitative trait loci (eQTLs) database, SCAN,[75] with our three most significant *ADAM29* SNPs with rs numbers two of the SNPs (rs6553849 and rs4388065) have no known effect on gene expression while rs12512467 has been shown to effect expression of two genes, *WDR5B* (p-value = 7e-5) and *ANAPC4* (p-value = 0.0001), in the HapMap Yorubian (YRI) population. Further complicating the functional consequences of our proposed associated SNPs, *WDR5B* is not located on the same chromosome as *ADAM29 (ADAM29* and *ANAPC4* chromosome 4 vs.

*WDR5B* chromosome 3) illustrating that SNPs can have functional effects in *trans*..

Anaphase promoting complex subunit 4 (*ANAPC4*) was the gene most 5' to the most significantly associated SNP, rs871476, identified in our chi-square association analysis between maternal DNA and PTB. For a more extensive discussion about *ANAPC4* please see Chapter II. The other gene identified as an eQTL, WD repeat domain 5B (*WDR5B*) as the name suggests contains a number of WD40 repeats, is ubiquitously expressed,[70,71] and hypothesized to be involved in protein-protein interactions. These two genes, much like *ADAM29*, also have no obvious mechanistic role controlling birth weight z-scores.

Additionally, considering the eQTL was only observed in the HapMap YRI samples it is unlikely that the eQTL effect of rs12512467 is generalizable to the Finnish. Another limitation with this study is that our replication cohort was somewhat small and to further validate this report our association needs to be replicated/generalized in an independent dataset. Finally, the lack of demographic data to adjust for in our Oulu infant dataset was another weakness of this study.

In conclusion, we report the first genome-wide association study investigating birth weight z-scores that includes both infants born during the gestational age considered preterm and term. Here we found a reproducible association with SNPs in *ADAM29*. Functional work will be required to

understand how these SNPs in *ADAM29* could be controlling birth weight z-scores, nevertheless this report will help elucidate new biology and offer potential opportunities for intervention.

## Materials and methods

*Sample collection*

Infants from preterm or term delivery were enrolled for genetic analysis by methods approved by the Institutional Review Boards/Ethics Committees of University Central Hospital, Helsinki and the University of Oulu. DNA was extracted from whole blood or Oragene® saliva kits. Standard manufacturer protocols were followed. Demographic information was collected by a trained study nurse using an approved questionnaire and all variables with the exceptions of alcohol and tobacco use were quantitative.

*Sample inclusion and exclusion criteria*

For the Illumina Omni2.5 BeadChips and Applied Biosystems TaqMan SNP genotyping the inclusion and exclusion criteria are as described below. Case infants were required to have been delivered spontaneously preterm between 22-36 weeks gestation and additionally either the child or mother had to have a first degree relative with a history of PTB, or a spontaneous idiopathic preterm delivery less than 35 weeks of gestation. Infants were excluded if their mother

had any medical indication for a preterm delivery, where part of a multiple gestation pregnancy, or other identified risk from preterm delivery such as recent trauma or clinical evidence of infection.

*Helsinki infant Illumina Omni2.5 BeadChips*

750ng of genomic DNA was sent to the Vanderbilt DNA microarray core where they performed QC and processed the Illumina arrays per the manufacture's protocol. The GTC created PLINK ready BED files from the processed array genotype calls. QC was performed in PLINK. Monomorphic SNPs were removed by using a minor allele frequency filter (MAF) of 0.000001%. We chose to not filter by Hardy-Weinberg equilibrium (HWE), but HWE checks were performed in order to see if our most significant SNPs were out of HWE in controls. SNPs which were found to have a HWE $p < 0.0001$ in controls were excluded from further analysis.

*Applied Biosystems TaqMan genotyping*

Genomic DNA from Oulu was sent to the Cincinnati Children's Hospital Medical Center Genetic Variation and Gene Discovery Core Facility. Standard manufacturers protocols were followed and for each assay 10 ng of genomic DNA was used.

*Statistical analysis*

All association analysis was performed using the software package PLINK v1.07. For association tests Fisher's exact allelic tests were used due to SNPs with low MAF. SPSS18 was used to perform one-sample Kolmogorov-Smirnov tests and plot histograms of birth weight z-scores. The software package SPSS Statistics V.20, was used to test for differences in demographic data between the case and control infants and p-values were determined using a one-way ANOVA. The R software package was used to create QQ and Manhattan plots. All p-values are two-tailed unless noted otherwise.

# CHAPTER V

## MATERNAL CODING REGION VARIANTS IN COMPLEMENT RECEPTOR 1 INCREASE RISK FOR SPONTANEOUS IDIOPATHIC PRETERM BIRTH [**]

### Introduction

Preterm birth (PTB), defined as live birth before 37 weeks' completed gestation, is the leading cause of infant mortality worldwide.[1] Despite this major public health concern, little is known about the pathogenesis of PTB. The limited insight into PTB is contributed to by the fact that the mechanism for normal parturition in general is not known in humans. There have been a number of suggested pathways believed to play a role in PTB pathogenesis, but direct evidence for any of these is modest at best.[13]

A number of lines of evidence suggest that PTB has a genetic component such as PTB aggregating in families, segregation analysis and genetic modeling.[16,53] Primarily through candidate gene studies, there have been a number of SNPs associated with PTB; however, contradictory evidence from replication studies exists, and none of these have large effect sizes or have implicated new mechanisms in parturition control.[60]

With the advent of next-generation sequencing (NGS) and exon capture technology the ability to sequence a patient's exome provides an important new approach to disease gene discovery. Whole-exome sequencing has been used to

[**] Adapted from McElroy et al. In Press at *Human Genetics*

identify the casual variant/gene for a number of Mendelian diseases.[158,159] While the potential of whole-exome sequencing to identify the cause of complex diseases has been discussed,[160] this approach has only been used sparingly such as in autism spectrum disorders.[161]

In this study, we test the hypothesis that rare variants aggregate in specific genes and pathways that contribute to PTB risk. In order to test our hypothesis, we performed whole-exome sequencing in multiplex families with a history of spontaneous idiopathic PTB. We identified predicted deleterious variants aggregating in complement/coagulation pathway genes, and extended this observation to more common coding region variants to demonstrate a significant association of the complement receptor 1 (*CR1*) gene with spontaneous idiopathic preterm birth.

## Results

*Analysis of shared variants in mother-daughter pairs*

To identify potential rare variants that contribute to the risk for preterm birth, we analyzed two Finnish mother-daughter pairs (families 1168 and 1281) each of whom experienced preterm delivery (Figure 5.1). These families were selected from more than 100 pedigrees due to high penetrance of preterm birth to each mother, more than one affected generation with spontaneous idiopathic preterm birth phenotype, and exhibiting a maternal pattern of transmission. Variants were called using standard best-practice quality control (QC) thresholds

in Genome Analysis Toolkit (GATK).[162] Only variants that passed QC filters were then input into the Variant Annotation, Analysis and Search Tool (VAAST) pipeline.[163] Each set of these variants was then individually compared to a 1000 Genomes background VAAST file that contained data from 1093 individuals.

In family 1168 there were 202 genes/features, which had the most significant VAAST genome-wide permutated p-value (1.67e-06) (Appendix J). For family 1281 there were 275 genes with the most significant VAAST genome-wide permutated p-value (1.67e-06) (Appendix K). Examining the overlap of these gene lists reveals 163 genes that are common between the two families. The top genome-wide permutation p-value for both families are considered genome-wide significant due to the fact that it surpasses the Bonferroni corrected p-value for 20,000 genes, 2.5e-06.

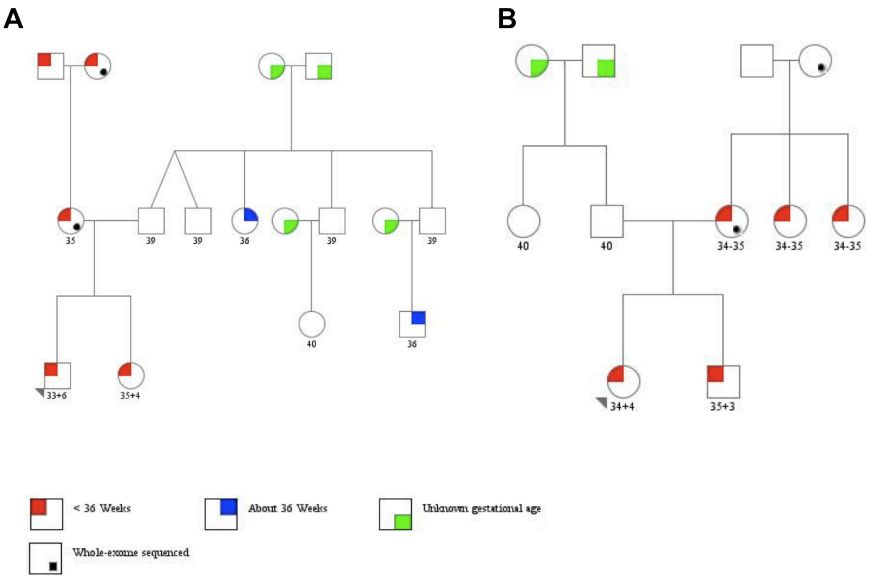


Figure 5.1: **Pedigrees for the two mother-daughter pairs that were whole-exome sequenced.** Family 1168 (A) and family 1281 (B)

*Pathway analysis of the most significant genes in mother-daughter pairs*

In order to glean insight into PTB pathophysiology we used a pathway analysis approach. The most significant genes for our two families were tested for pathway enrichment using the Kyoto Encyclopedia of Genes and Genomes (KEGG).[164]

Overall, for family 1168 genes from the most significant list played a role in 102 pathways (Table 5.1). The top three pathways were olfactory transduction, metabolic pathways and complement and coagulation cascades with 25, eight, and five genes respectively. In family 1281, there were 86 KEGG pathways with at least one gene, which had the most significant VAAST p-value (Table 5.2). The top three pathways were olfactory transduction, focal adhesion and metabolic pathways with 20, six, and five genes respectively. The complement and coagulation cascades is in the next group of three pathways which each have four genes. Table 5.3 lists the genes in each family that were involved in the complement and coagulation cascades.

Table 5.1: **KEGG pathways with more than three genes from the list of most significant p-value genes for family 1168.** Only pathways with three or greater genes are listed. The number of genes for each pathway is the number inside the parentheses. For a complete list of KEGG pathways for family 1168 please see Appendix L.

| KEGG Pathways Top P-value Genes Family 1168 |
|---|
| hsa04740 Olfactory transduction (25) |
| hsa01100 Metabolic pathways (8) |
| hsa04610 Complement and coagulation cascades (5) |
| hsa04640 Hematopoietic cell lineage (4) |
| hsa04510 Focal adhesion (4) |
| hsa04972 Pancreatic secretion (4) |
| hsa05164 Influenza A (4) |
| hsa04974 Protein digestion and absorption (4) |
| hsa04080 Neuroactive ligand-receptor interaction (3) |
| hsa04141 Protein processing in endoplasmic reticulum (3) |
| hsa05144 Malaria (3) |
| hsa00520 Amino sugar and nucleotide sugar metabolism (3) |
| hsa04666 Fc gamma R-mediated phagocytosis (3) |
| hsa04360 Axon guidance (3) |

Table 5.2: **KEGG pathways with more than three genes from the list of most significant p-value genes for family 1281.** Only pathways with three or greater genes are listed. The number of genes for each pathway is the number inside the parentheses. For a complete list of KEGG pathways for family 1168 please see Appendix M.

| KEGG Pathways Top P-value Genes Family 1281 |
|---|
| hsa04740 Olfactory transduction (20) |
| hsa04510 Focal adhesion (6) |
| hsa01100 Metabolic pathways (5) |
| hsa04141 Protein processing in endoplasmic reticulum (4) |
| hsa04610 Complement and coagulation cascades (4) |
| hsa04512 ECM-receptor interaction (4) |
| hsa04972 Pancreatic secretion (3) |
| hsa04650 Natural killer cell mediated cytotoxicity (3) |
| hsa05164 Influenza A (3) |
| hsa04612 Antigen processing and presentation (3) |
| hsa04974 Protein digestion and absorption (3) |

Table 5.3: **The complement and coagulation cascade genes from the KEGG analysis of the most significant p-value VAAST genes for families 1168 and 1281.** The higher the VAAST score the more likely the gene is to be disease causing.

| Gene | Gene Name | VAAST Score Family 1168 | VAAST Score Family 1281 | VAAST Rank Family 1168 | VAAST Rank Family 1281 | Total Missense SNPs for the other 6 exomes |
|---|---|---|---|---|---|---|
| *CR1* | Complement component (3b/4b) receptor 1 (Knops blood group) | 174.31 | 71.81 | 12 | 74 | 60*Φ |
| *F5* | Coagulation factor V (proaccelerin, labile factor) | 91.08 | 53.38 | 45 | 130 | 44* |
| *F13B* | Coagulation factor XIII, B polypeptide | 57.60 | 57.60 | 64 | 86 | 6 |
| *CR2* | Complement component (3d/Epstein Barr virus) receptor 2 | 56.76 | 39.38 | 70 | 162 | 30* |
| *C4BPA* | Complement component 4 binding protein, alpha | 23.64 | 38.25 | 182 | 170 | 7 |
| *CFH* | Complement factor H | n/a | 76.52 | n/a | 65 | 10* |

* Contains variants predicted to be probably damaging

Φ Contains variants predicted to be possibly damaging

*Examination of complement and coagulation cascades in our other PTB exomes*

We chose to focus our analysis on the KEGG complement and coagulation cascade instead of the two others that they shared, olfactory transduction and metabolic pathways, for three main reasons. First, whole-genome sequencing has shown that olfactory transduction genes harbor more predicted loss-of-function variants than expected when interrogating 1000 Genomes Project data;[165] therefore we believed that these shared variants in the mother-daughter pairs were unlikely to be involved in PTB. Second, coagulation and immune activation (complement system) are two of the proposed pathways previously hypothesized to contribute to PTB.[13,166] Third, prior modest associations exist between coagulation pathway genes, *F5*,[167,168] *F7*,[168,169] *F13A1*,[169] and *PLAT*[168] and PTB.

Using our two mother-daughter pairs as a "discovery" cohort we examined the six genes from the KEGG complement and coagulation cascades identified by VAAST in six other PTB exomes (5 Finnish, 1 European American). Because of the higher probability of being damaging, we focused on novel variants. Of the six exomes, three harbored novel variants. There were 19 total novel variants: 14 were unique, and half were missense SNPs. Using the *in silico* tool PolyPhen-2, we assessed the novel missense variants for potential to be deleterious using the HumDiv algorithm.[170] The only variant predicted to be "probably damaging" was a complement factor H (*CFH*) Thr956Met variant seen

in a single family. All of the other novel missense variants were predicted to be "benign" (Table 5.4).

We also analyzed all missense variants in these genes. All of the six genes harbored between 6 and 60 missense variants for a total of 157, and 36 were unique. We once again tested the potential for these variants to be deleterious using the PolyPhen-2 *in silico* tool's HumDiv algorithm. Only half of the genes, *CR1, CR2, and F5,* contained missense variants that were predicted to be "probably damaging" and *CR1* was the sole gene that contained missense variants predicted to be "possibly damaging" by PolyPhen-2 (Table 5.5).

Table 5.4: **All novel missense variants in our six other exomes in the six complement and coagulation cascade genes identified by VAAST.** For the PolyPhen-2 prediction the HumDiv algorithm was used.

| Individual | Gene | Variant | PolyPhen-2 |
|---|---|---|---|
| Family 150 | *F5* | GLU1390GLN | **Benign** |
| Family 150 | *F5* | LEU1370PHE | **Benign** |
| Family 150 | *F5* | PRO1361LEU | **Benign** |
| Family 150 | *F5* | LEU1357ILE | **Benign** |
| Family 1165 | *F5* | LEU1370PHE | **Benign** |
| Family 1165 | *F5* | PRO1361LEU | **Benign** |
| Family 1165 | *F5* | LEU1357ILE | **Benign** |
| Family 1165 | *F5* | PHE1334LEU | **Benign** |
| Family 1165 | *F5* | ARG1220THR | **Benign** |
| Family 14w | *CFH* | THR956MET | **Probably Damaging** |

Table 5.5**: Counts of missense SNPs in our six proband mother exomes.** The numbers are for counts of missense variants in the six exomes, and not unique variants, except when in parentheses or if there is a single variant.

| GENE | TOTAL MISSENSE | NOVEL MISSENSE | POLYPHEN-2 PROBABLY DAMAGING | POLYPHEN-2 POSSIBLY DAMAGING |
|---|---|---|---|---|
| *CR1* | 60 | 4 | 9 (2) | 3 (2) |
| *F5* | 44 | 14 | 2 (1) | 0 |
| *CR2* | 30 | 0 | 12 (2) | 0 |
| *CFH* | 10 | 1 | 1 | 0 |
| *C4BPA* | 7 | 0 | 0 | 0 |
| *F13B* | 6 | 0 | 0 | 0 |

*Interrogation of the complement and coagulation cascade in nuclear PTB mothers*

Based upon the exome sequencing findings, we next tested the hypothesis that coding-region variants in the complement/coagulation cascade genes identified in the Finnish families contributed more broadly to the pathogenesis of preterm birth. We conducted an association study in 237 case and 328 control Finnish mothers. We performed additive logistic regression adjusting for the variables shown to differ significantly between the preterm and term mothers: body mass index (BMI), gravidity, ethanol use and smoking use (Table 5.6). Our examination was focused on the six complement and coagulation cascade genes identified by VAAST in our two whole-exome families. In total, 67 coding region SNPs from the six gene regions were analyzed (Appendix N). The most significantly associated SNP was an exonic missense SNP, rs6691117, unadjusted

p-value = 6.93e-5, OR = 1.74 (1.33, 2.29 95% CI); adjusted additive logistic regression p-value = 1.07e-4, OR = 1.73 (1.31, 2.29 95% CI) in *CR1*. This association withstands a conservative Bonferroni corrected p-value of 7.64e-4. Depending on the transcript, this SNP changes an isoleucine to a valine at amino acid position 1615 or 2065. Both of these substitutions are predicted to be "benign" by the HumDiv algorithm in PolyPhen-2.

Table 5.6: **Demographic information for Helsinki mothers genotyped on the Illumina exome BeadChip**. Numbers in the table are mean (standard deviation) except for dichotomous variables where percentages were used. P-values were determined using a one-way analysis of variance (ANOVA).

| Variable | Preterm (n=237) | Term (n=328) | P-value |
|---|---|---|---|
| Maternal age (yr) | 31.1 (5.0) | 31.5 (4.2) | 0.334 |
| Body Mass Index (kg/m$^2$) | 23.5 (4.5) | 22.7 (3.1) | 0.012 |
| Parity (n) | 1.6 (0.89) | 1.5 (0.73) | 0.062 |
| Gravidity (n) | 2.1 (1.3) | 1.9 (1.0) | 0.050 |
| Birth weight (g) | 2337.0 (500.1) | 3578.7 (421.3) | <0.0001 |
| Birth length (cm) | 45.0 (2.8) | 50.3 (1.9) | <0.0001 |
| Alcohol use (%) | 5.2% | 1.5% | 0.012 |
| Tobacco use (%) | 8.7% | 2.7% | 0.002 |

Due to the robust association of the *CR1* coding region SNP we applied a similar analysis to our Finnish mothers (252 cases/287 controls), which were previously genotyped on the Affymetrix 6.0 SNP arrays.[171] We provide this additional analysis to explore whether our coding SNP, rs6691117, on the exome

array may be tagging another variant elsewhere in the *CR1* gene or its regulatory regions which would be detected with this more densely sampled, largely noncoding variant array. We identified 103 SNPs in the region spanning 10 kb 5' through 10 kb 3' of the *CR1* gene boundary (Appendix O). The most significantly associated SNP, rs10429953, was located in an intron of *CR1* unadjusted additive logistic regression p-value = 1.31e-4, OR = 1.93. This p-value surpasses the Bonferroni corrected p-value of 4.85e-4 for 103 SNPs tested. In addition to the most significantly associated SNPs surpassing Bonferroni, the second highest associated SNP, rs10429943, also in a *CR1* intron p-value = 3.74e-4 OR = 1.84 clears the threshold. However, these two SNPs are in strong linkage disequilibrium ($r^2 = 0.96$) so this should be considered a single strong association in *CR1* (Figure 5.2).

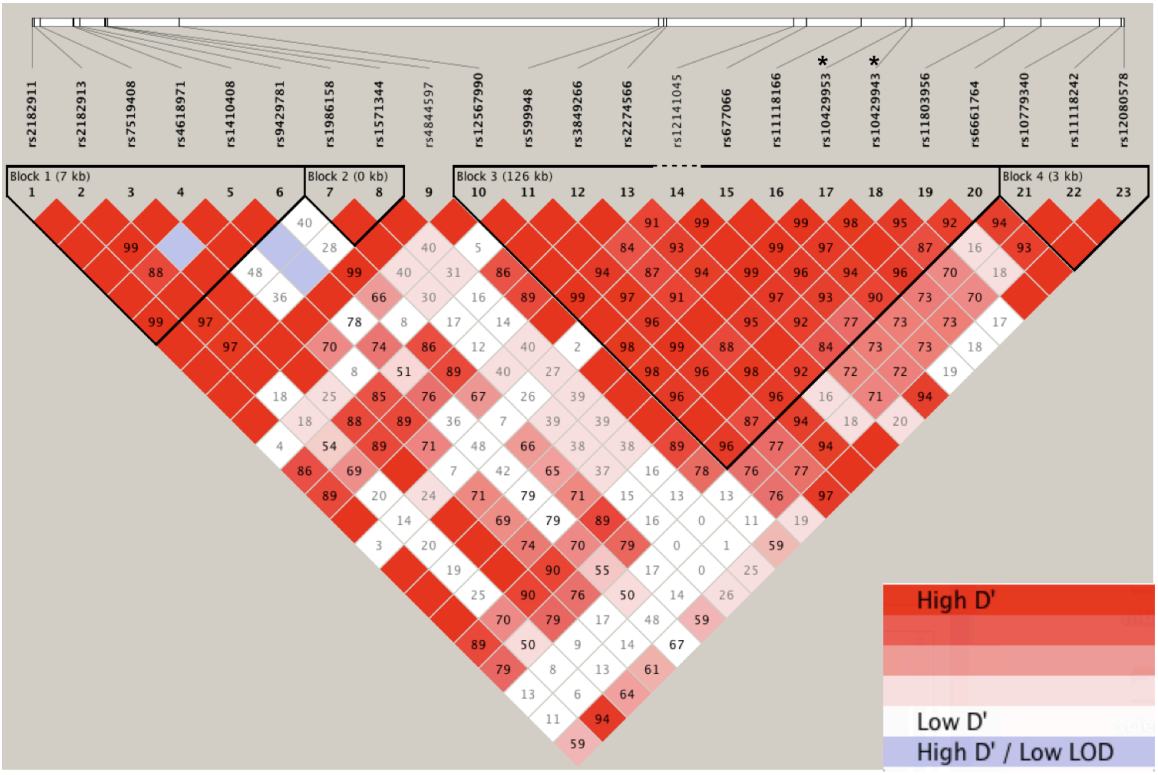


Figure 5.2: **A Haploview view LD diagram showing D′ values for our Affymetrix 6.0 SNP array samples for the *CR1* gene region with the addition of 10Kb 5′ and 3′.** The two intronic SNPs on the Affymetrix array are marked with an asterisk. The intronic SNP on the Illumina exome beadchip resides in the same 126Kb LD block.

## Discussion

To our knowledge, this is the first report using whole-exome sequencing to interrogate for rare variants that aggregate in specific genes and pathways in the complex disease of PTB. We used our results from the exome sequencing of mother-daughter pairs in highly affected families to expand to candidate gene/pathway association studies using genotyping arrays. This more focused

analysis enhances the power to detect significant variants in smaller nuclear datasets, and should accelerate gene discovery. We believe that this approach can be effectively applied to other complex disorders.

CR1, and the complement/coagulation factor pathway, as revealed by our exome sequencing and follow-up analysis, provides a biologically plausible pathway related to adverse pregnancy outcomes. There is a growing body of literature describing activation of the complement system and adverse pregnancy outcomes, including PTB.[172,173] There have been a number of reports illustrating that an increased level of fragment Bb (FBb) early in pregnancy is associated with an increased risk of PTB < 34 weeks.[174] FBb is a marker of alternative pathway complement activation. In addition to increased FBb, increased maternal plasma levels of complement C3a during the first trimester have been associated with an increased risk of a number of adverse pregnancy outcomes including PTB.[175] It has also been shown that erythrocyte membrane complement receptor 1 (CR1) levels are reduced during pregnancy and reach their nadir during the third trimester.[176] In this study, we found a significant association between three SNPs in *CR1* and an increased risk of PTB. However, when we examine linkage disequilibrium all three of the SNPs are in a large 126 kb linkage block (Figure 5.2). The strongest association residing in the coding region variant suggests that the other non-coding SNPs are detected based on tagging this variant.

While predicted to be "benign" the isoleucine to valine exonic SNP in our exome arrays may be functional; such a subtle substitution would be misclassified by *in silico* tools such as PolyPhen-2[170] and SIFT.[177] Evidence for functional consequences of the SNP leading to the isoleucine to valine change in *CR1* we identified in our exome arrays is its association with alteration in erythrocyte sedimentation rate (ESR), as detected in a large genome-wide association study.[178] Our SNP (rs6691117) is associated with a decreased ESR. As CR1 on erythrocytes leads to increased clearance of immune complexes to limit their deposition in vessel walls, we would predict greater systemic inflammation or coagulability resulting from our risk-promoting allele. ESR normally increases in pregnancy due to increased fibrinogen levels and the need for clearance of immune complexes;[179] attenuating this process by less functional CR1 variants would be predicted to increase risk for adverse pregnancy outcomes such as preterm birth. Alternatively, this SNP may be tagging a different causative variant in the gene. It should be noted that the predicted damaging SNP rs2274567, resulting in a histidine to arginine change, is in linkage disequilibrium with rs6691117, and is associated with similar changes in ESR.[178]

In conclusion, this is the first report of using whole-exome sequencing to interrogate for rare variants that aggregate in specific genes and pathways in the complex disease of PTB. We believe that our results strengthen the argument that the complement and coagulation cascade are involved in the pathophysiology of PTB, and suggest potential screening and intervention approaches to prevent

prematurity, which target this pathway.[180] Possible interventions include the use of soluble CR1[181] or monoclonal antibodies to C5[182] to limit complement activation and inflammation or enhance clearance, which our risk allele may compromise. While there are FDA approved anti-C5 antibodies, eculizumab,[183] soluble CR1 is still in the clinical trial stage. Future functional studies will be essential to determine the specific mechanisms by which these pathways increase prematurity risk.

## Materials and Methods

*Sample collection*

Mothers of preterm or term infants were enrolled for genetic analysis by methods approved by Institutional Review Boards/Ethics Committees of University Central Hospital, Helsinki, University of Oulu, Cincinnati Children's Hospital Medical Center and Vanderbilt University. DNA was extracted from whole blood or Oragene® saliva kits. Standard manufacturer protocols were followed.

*Sample inclusion and exclusion criteria*

Samples were selected for whole-exome sequencing based on a number of features, which we termed, "preterm birth load". What we took into account

was the shortest gestation in a pedigree, number of preterm children and whether the mother herself was born preterm. For this study we only sequenced case mothers.

For the Illumina and Affymetrix SNP genotyping arrays the inclusion and exclusion criteria are as described below. Case mothers were required to have a spontaneous preterm birth between 22-36 weeks gestation and additionally either the child or mother had to have a first degree relative with a history of PTB, or spontaneous idiopathic preterm birth less than 35 weeks of gestation. Families for whole exome sequencing were selected by more than one generation of the pedigree affected with an apparent maternal mode of transmission. Mothers were excluded if they had any medical indication for a preterm delivery, a multiple gestation pregnancy, or other identified risk from preterm birth such as recent trauma or clinical evidence of infection.

*Exome capture*

A total of 3μg of genomic DNA was submitted to the Vanderbilt Genome Technology Core (GTC) for whole-exome capture. The Agilent Technologies 50Mb SureSelect Human All Exome Kit was used to capture and amplify the submitted samples. Per the manufacturer's website, the kit interrogates 1.22% of the human genomic regions corresponding to the NCBI Consensus CDS database (CCDS).

*Whole-exome sequencing*

The GTC performed the sequencing using an Illumina HiSeq and 100bp paired-end reads. This resulted in an average of 76x read-depth for all of our exome with a range between ~60-104x.

*Sequencing quality control*

The Genome Analysis Toolkit (GATK) using current best practices was used to trim, align and call variants in our exome sequence data against the Hg19 genome build.

*VAAST analysis*

Only variants, which passed GATK QC were annotated against the Hg19 genome build using the Variant Annotation Tool (VAT). Once annotated the Variant Selection Tool (VST) was used to select variants that were shared between the mother-daughter pairs in family 1168 and 1281. Each set of these shared variants were then individually compared to a 1000 Genomes background VAAST background file that contained data from 1093 individuals using the Variant Analysis Tool (VAT). When VAT was ran the codon-bias

option and default additive genetic model were used. VAAST p-values were calculated using the fast genome-permutation option with 1e5 permutations.

*KEGG pathway analysis of VAAST most significant p-value genes*

All of the genes with the top p-value for each mother-daughter pair were searched against the KEGG Homo Sapiens pathways to test for pathway enrichment in the top p-value genes.

*Finnish mothers Illumina HumanExome beadchips*

750ng of genomic DNA was sent to the Vanderbilt DNA microarray core where they performed QC and processed the Illumina arrays per the manufacture's protocol. The GTC created PLINK ready BED files form the processed array genotype calls. QC was performed in PLINK. Monomorphic SNPs were removed by using a minor allele frequency filter (MAF) of 0.000001%. We chose to not filter by Hardy-Weinberg equilibrium (HWE), but HWE checks were performed in order to see if our most significant SNPs were out of HWE in controls.

*Finnish mothers Affymetrix Genome-Wide Human SNP Array 6.0*

750ng of genomic DNA were genotyped in the Vanderbilt microarray core. All samples, which passed the core's QC thresholds, were processed using the manufacturer's protocol for the Affymetrix 6.0 arrays. The PLINK ready BED files were processed for QC using PLINK. The QC steps were as follows and performed in the following order: remove SNPs with genotype frequency <95%, remove samples with <95% SNP calls, remove SNPs with MAF <5%, and remove SNPs with HWE p-values in controls <0.0001.

*Statistical analysis*

All SNP statistical analysis was performed using the software package PLINK v1.07. Demographic data was analyzed using SPSS Statistics V.20 and means were tested for a significant difference using a one-way analysis of variance (ANOVA). For association tests of complement/coagulation factor cascade SNPs, we used additive logistic regression adjusting for the factors shown to differ significantly between cases and controls (Table 5.6). Associations were considered statistically significant if they survived a Bonferroni correction for the number of SNPs tested.

# CHAPTER VI

# CONCLUSION

## Summary

Preterm birth (PTB) and parturition in general are controlled by a complex interaction of genetic and environmental factors. While we have known that gestational age is a heritable trait, we can currently only explain a small proportion of this heritability.[16,51-53,60] While this so-called "missing heritability"[184] is not unique to PTB, it is accentuated because compared to most complex disease or phenotypes studied, PTB genetics is very verdant. In Chapter I, I discuss some of the genetic variants associated with PTB and also some of the reasons why the majority of associations have not replicated and/or generalized.[60] Additionally, since PTB genetics is still in its infancy, access to appropriate datasets to replicate associations is often lacking.

Genome-wide association studies (GWAS) have been used successfully as an agnostic approach to identify SNPs associated with other complex diseases such as type 2 diabetes and cardiovascular disease reviewed by Manolio.[66] As of August 16, 2012, the National Human Genome Research Institute (NHGRI) GWAS catalog of published GWAS includes 1353 publications and 7039 SNPs. While this catalog includes more than 700 phenotypes/traits there has yet to be a published GWAS investigating PTB, gestational age or birth weight z-scores. To complete this dissertation, I have interrogated both the maternal and fetal

genome in an attempt to explain the etiology of PTB or the associated traits, gestational age or birth weight z-scores.

In Chapter II, I utilized a GWAS in Finnish mothers to investigate PTB and the quantitative trait gestational age. When we used PTB as the phenotype, we identified only a single SNP at the $10^{-6}$ level of significance, rs871476, p-value = 6.22e-6, odds ratio (OR) = 0.484. This is an intergenic SNP that is more than 100kb away from the two closest genes *SLC34A2* and *ANAPC4*. Unfortunately, this SNP is unlikely to be a true causative SNP for PTB. Due to our smaller sample size of 539 mothers, especially for a GWAS, the "real" SNP which is tagging the causative variant is probably still somewhere lower down in significance, in the noise. We will only be able to answer this question more definitively when we increase our dataset by adding new samples, which our colleagues in Finland send to us every 6-8 months. Larger sample sizes in the future will allow us to re-investigate these and other genetic associations with PTB.

After transforming the gestational ages using Box-Cox power transformation in order to make the dependent variable more Gaussian our results were more encouraging. We observed three SNPs with p-values at the $10^{-6}$ level of significance. The strongest association, p-value = 5.26e-6, for a SNP 3' downstream of *COL11A1*. While *COL11A1* has not been associated with preterm birth or gestational age previously, it has been associated with Marshall

syndrome,[79-82] type II Stickler syndrome[76-79] and an increased risk for lumbar disc herniation.[83]

For the Box-Cox transformed gestational age linear regression using genotypic, dominant and recessive linear regression our strongest association is observed for rs10104530, which is 5' upstream of *TACC1,* for both the genotypic and recessive model, 7.09e-8 and 9.45e-7 respectively. *TACC1* has not been previously associated with PTB or gestational age control, but has previously been shown to be dysregulated in breast,[84-86] ovarian[87] and gastric[88] carcinoma. While this association with Box-Cox transformed gestational ages could be a false positive, alternatively it could be illustrating pleiotropy, an area of human genetics that is currently garnering substantial interest.[185,186]

Also of note, and further strengthening the element of pleiotropy the second most significant SNP under an additive model is in *TCF7L2* (p-value 1.10e-3). Variants within this gene have been associated with a number of phenotypes such as type 2 diabetes risk, glucose control,[90-102,187] coronary heart disease, [110] metabolic syndrome[109] and absolute birth weight;[109] however, this is the first report of a variant within this gene being associated with gestational age.

While GWAS are an ideal agnostic approach to discover common variants with minor allele frequency (MAF) > 5% and smaller effect sizes, they are unable to discover/test rare variants with larger effect sizes.[65,66] To remedy this, we used complementary approaches such as whole-exome sequencing, Illumina HumanExome BeadChip association and pathway analysis to investigate rare and

lower (MAF < 5%) variants in the maternal genome. These results are presented and discussed in Chapter V of this dissertation. These experiments added evidence to the growing theory of involvement of the complement and coagulation system systems in the pathogenesis of PTB.[172-175] By selecting mothers to sequence form highly effected families and using two mother-daughter pairs as a "discovery" cohort then following up with exome beadchips in a larger dataset of mothers, we discovered a robust association with coding SNPs in complement receptor 1 (*CR1*). We hope that this finding will allow for screening of gravid women and targets for potential therapeutic intervention.

Chapters III and IV of this dissertation, take a different tact and instead of investigating the maternal genome we switched our focus to the fetal genome. We first utilized a GWAS testing for associations with preterm birth risk and Box-Cox transformed gestational age. The strongest association is observed for rs7153053 when we explore both phenotypes. This SNP is located within an intron of the Ensembl gene, *ENSG0256050,* and 5' upstream of *INF2*. Neither of these two genes has previously been associated with PTB or gestation length control; however, missense variants in *INF2* have previously been discovered to be a cause of autosomal dominant focal segmental glomerulosclerosis (FSGS).[130] Additionally, *ENSG0256050*, is predicted to be a long intergenic non-coding RNAs (lincRNA) which can control a number of genes and been shown to be involved in development.[126,127]

We once again alter our analysis and in Chapter IV interrogate the fetal genome for SNPs that associate with birth weight z-scores. Unlike earlier GWAS of birth weight, we chose to use the normalized birth weight z-scores instead of the absolute birth weight because this allowed us to have the largest dataset possible and include both infants born in the preterm and term range. This investigation discovered a number of highly significant associations with SNPs in the coding region of *ADAM29*. When we examine the associations closer for the adjusted additive linear regression we observed five SNPs in the coding region of *ADAM29*, but they are all in very high linkage disequilibrium (LD); therefore it is best to consider this one very robust association with three SNP (rs6553849, KGP11462362 and KGP7935680) having the identical p-value and beta, p-value = 4.25e-7 and beta = 0.33. These three SNPs as well as the next most significant, rs12512467, p-value = 6.04e-7 and beta = 0.29, all while not reaching the threshold of "genome-wide" significance p-value < 5e-8 were close. Like all of the other GWAS analysis, we also tested birth weight z-scores under a general genotypic and dominant and recessive models. The most significant SNPs for each analysis were the same five SNPs from the additive model and for each SNP the additive model produced the most significant p-values.

Following up the most significant birth weight z-score associations in an independent Oulu infant cohort we observed a one-tailed p-value = 0.096 for rs6553849. While this does not surpass the traditional p-value ≤ 0.05 level of significance I would still consider it a positive replication especially since the

beta is in the same direction. Unfortunately, we do not have access to demographic data for the infants born during the term period (37-41weeks) therefore we were unable to adjust our linear association in the Oulu infant birth weight z-score analysis.

Overall, in this dissertation I present a comprehensive interrogation of the maternal and fetal genome investigating a number of phenotypes: preterm birth, gestational age and birth weight z-scores. Not unexpectedly, the top "hits" for all of these analyses were not always the same SNP or within the same gene/region. Also based on current biological understanding the most significantly associated SNPs would not have been identified using the more traditional candidate gene or candidate pathway approach because the majority of the top SNPs do not have an obvious connection to any of the hypothesized pathways involved in preterm birth and parturition.[13,64] If these associations replicate, they add to the growing evidence showing the importance of pleiotropy.[185,186] This is the beauty of the agnostic GWAS and whole-exome sequencing approaches and these results should not be considered the end of the investigation, but only the beginning.

## Future Directions

While in this dissertation, I present a number of interesting results there is still much more work to be done in order to achieve the ultimate goal of this work: to transform pubic health by helping prevent or delay preterm birth.

A logical first step for all of the association results discussed in this dissertation, and human genetics in general, is to attempt to replicate and/or generalize all of the SNPs with the most significant associations. The terms "replicate" and "generalize" are often incorrectly used interchangeably. Replication is an experiment in which one attempts to find a significant association and affect in the same direction, when the alleles are coded the same way, for variants of interest in an independent dataset from the same ancestral background. On the other hand, generalization is when the most highly associated SNPs from an association study are interrogated an independent datasets from different ancestral backgrounds i.e. a Caucasian GWAS identified SNP is tested for an association in an African American cohort. It is possible for a SNP to replicate and not generalize.

The primary datasets used for this dissertation are Finnish, which due to the small population size (~5.4 million) will unfortunately make finding an appropriate replication cohort virtually impossible. Due to this limitation, I believe that finding an appropriate generalization cohort to be a key future direction and should be undertaken as soon as possible.

Because of the lack of appropriate replication cohort availability and the fact that Finland is a genetic isolate, I believe that the strongest associations discovered in this dissertation may not be generalizable. The two primary reasons I believe this is due to possible different genetic architecture of PTB and gestational age control in the Finnish due to private mutations or more likely due to different patterns of LD. However, while the same actual variant may not generalize other variants within a gene/region or SNPs in other genes within a particular pathway may be discovered to have robust associations and should be considered evidence in support of the results presented here. Due to the low cost of ABI TaqMan or Sequenom genotyping I recommend that in addition to the most highly significant SNPs additional SNPs in that gene/region and additional genes within a pathway (if one is known for a gene) are interrogated when completing this future direction.

In addition to genotyping, with the decreasing costs and throughput of next-generation sequencing (NGS) always increasing another future direction is to perform deep sequencing of genes and genes within pathways identified in the GWAS presented. Because I expect the majority of the discovered putative functional variants to be very rare and/or private I believe the best use of funds would be to sequence only affected individuals. When interesting variants are discovered, ABI TaqMan assays can always be created to explore if any term mothers/infants harbor said variant.

Finally, computational and statistical analysis can only teach us so much about the pathogenesis and/or etiology of preterm birth, gestation length control and birth weight z-scores. Functional analysis will be required in order to discover if particular SNPs have measurable effects *in vitro*. While the specific functional experiment will be guided by the gene or region a SNP of interest is located in mRNA or protein level variation will be examined. Allele expression imbalance is another potential consequence one may want to explore.

As discussed above, preterm birth research is a young field overall and this is true for the sphere of human genetics. Before we are able to make large advancements in understanding the etiology or pathogenesis with the ultimate goal of being able to prevent or delay PTB a number of things must happen. First and foremost, what we really require are larger datasets which will have more power to identify variants with smaller, more realistic effect size. Currently, the two largest PTB datasets are the Danish National Birth Cohort (DNBC) and Norwegian Mother and Child Cohort (MoBa) both of which include ~1000 mother-child pair cases and ~1000 mother-child pair controls.[188]

These two larger datasets also illustrate another important aspect, which we need to standardize in order to move this field ahead; a more standardized phenotype of preterm birth. While PTB is inherently a heterogeneous phenotype, if we have a standardized phenotypic definition, which must include demographic factors known, to be associated with PTB or speculated to be associated some of which include: maternal age, body mass index (BMI),

gravidity, parity, preterm premature rupture of the membranes (PPROM), infection status (bacterial vaginosis (BV), Group Beta Streptococcus (GBS), etc), drug use during and prior to pregnancy, prior pregnancy complications, smoking and alcohol use. A more uniform set of variables will allow for more efficient and powerful meta-analyses in the future.

Other areas of research that will be important to interrogate in the future to help understand PTB are maternal-fetal interactions, gene-environment (GxE) and gene-gene (GxG) in both maternal and fetal samples. These type of analyses in genome-wide SNP and whole-exome/whole-genome sequence data are still in their infancy and will need to mature in order to be performed in all PTB studies, but I believe will be a fruitful area of discovery.

In conclusion, while the results discussed in this dissertation are interesting and have potentially implicated new pathways in the etiology of preterm birth, gestational age control and birth weight z-score this should only be considered a starting off point. There is much more analysis and functional studies needed to untangle how these SNPs and/or genes are involved. However, I believe the results presented in this dissertation to be an important step in the process of improving public health by understanding preterm birth.

**Appendix A. Summary of candidate gene association studies' findings as of August 2012. Modified from Jevon et al.**

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *ABCA1* | ATP-binding cassette, sub-family A, member 1 | 1 | 0 |
| *ACE* | angiotensin I converting enzyme 1 | 8 | 1 |
| *ADD1* | adducin 1 (alpha) | 2 | 0 |
| *ADH1B* | alcohol dehydrogenase 1B | 4 | 0 |
| *ADH1C* | alcohol dehydrogenase 1c | 5 | 0 |
| *ADRB2* | adrenergic, beta-2, receptor, surface | 9 | 3 |
| *AGT* | angiotensinogen | 4 | 2 |
| *AGTR1* | angiotensin II receptor, type 1 | 3 | 0 |
| *ALOX5* | arachidonate 5-lipoxygenase | 1 | 0 |
| *ALOX5AP* | arachidonate 5-lipoxygenase-activating protein | 2 | 0 |
| *ANG* | angiogenin, ribonuclease, RNase A family, 5 | 1 | 0 |
| *ANGPT2* | angiopoietin 2 | 1 | 0 |
| *ANXA5* | annexin A5 | 1 | 0 |
| *APOA1* | apolipoprotein A-I | 2 | 0 |
| *APOA4* | apolipoprotein A-4 | 1 | 0 |
| *APOA5* | apolipoprotein A-5 | 1 | 0 |
| *APOB* | apolipoprotein B | 2 | 0 |
| *APOC2* | apolipoprotein C2 | 1 | 0 |
| *APOC3* | apolipoprotein C3 | 2 | 0 |
| *APOE* | apolipoprotein E | 2 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *AQP2* | aquaporin 2 (collecting duct) | 1 | 0 |
| *BHMT* | betaine-homocysteine methyltransferase | 1 | 0 |
| *CBS* | cystathionine-beta-synthase | 5 | 1 |
| *CCL2* | chemokine (C-C motif) ligand 2 | 5 | 0 |
| *CCL3* | chemokine (C-C motif) ligand 3 | 4 | 0 |
| *CCL8* | chemokine (C-C motif) ligand 8 | 4 | 0 |
| *CCR2* | chemokine (C-C motif) receptor 2 | 1 | 0 |
| *CD14* | monocyte differentiation antigen CD14 | 6 | 1 |
| *CD55* | CD55 molecule, decay accelerating factor for complement | 1 | 0 |
| *CETP* | cholesteryl ester transfer protein, plasma | 2 | 0 |
| *COL1A1* | collagen, type I, alpha 1 | 4 | 0 |
| *COL1A2* | collagen, type I, alpha 2 | 4 | 0 |
| *COL3A1* | collagen, type 3, alpha 1 | 4 | 0 |
| *COL4A1* | collagen, type IV, alpha 1 | 1 | 0 |
| *COL4A2* | collagen, type IV, alpha 2 | 1 | 0 |
| *COL4A3* | collagen, type IV, alpha 3 | 1 | 1 |
| *COL4A4* | collagen, type IV, alpha 4 | 1 | 0 |
| *COL4A5* | collagen, type IV, alpha 5 | 1 | 0 |
| *COL4A6* | collagen, type IV, alpha 6 | 1 | 0 |
| *COL5A1* | collagen, type V, alpha 1 | 5 | 0 |
| *COL5A2* | collagen, type V, alpha 2 | 5 | 0 |
| *CRH* | corticotropin releasing hormone | 4 | 0 |
| *CRHBP* | corticotropin releasing hormone binding | 5 | 1 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| | protein | | |
| *CRHR1* | corticotropin releasing hormone receptor 1 | 5 | 0 |
| *CRHR2* | corticotropin releasing hormone receptor 2 | 5 | 0 |
| *CRP* | C-reactive protein, pentraxin-related | 4 | 0 |
| *CSF1* | colony stimulating factor 1 | 1 | 0 |
| *CSF2* | colony stimulating factor 2 | 1 | 0 |
| *CSF3* | colony stimulating factor 3 | 2 | 0 |
| *CSPG2* | Chondroitin sulfate proteoglycan core protein 2 | 1 | 0 |
| *CTGF* | connective tissue growth factor | 1 | 0 |
| *CTLA4* | cytotoxic T-lymphocyte-associated protein 4 | 4 | 1 |
| *CYP19A1* | cytochrome P450, family 19, subfamily A, polypeptide 1 | 4 | 0 |
| *CYP1A1* | cytochrome P450, family 1, subfamily A, polypeptide 1 | 7 | 2 |
| *CYP2C19* | cytochrome P450, family 2, subfamily C, polypeptide 19 | 1 | 0 |
| *CYP2D6* | cytochrome P450, family 2, subfamily D, polypeptide 6 | 4 | 0 |
| *CYP2E1* | cytochrome P450, family 2, subfamily E, polypeptide 1 | 2 | 0 |
| *CYP3A4* | cytochrome P450, family 3, subfamily A, polypeptide 4 | 1 | 0 |
| *DEFA5* | defensin, alpha 5, Paneth cell-specific | 1 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *DEFB1* | defensin, beta 1 | 1 | 0 |
| *DHCR24* | 24-dehydrocholesterol reductase | 1 | 0 |
| *DHCR7* | 7-dehydrocholesterol reductase | 1 | 0 |
| *DHFR* | dihydrofolate reductase | 4 | 0 |
| *DLAT* | dihydrolipoamide S-acetyltransferase | 1 | 0 |
| *DRD2* | dopamine receptor D2 | 1 | 0 |
| *EDN1* | endothelin 1 | 1 | 0 |
| *EDN2* | endothelin 2 | 4 | 0 |
| *ELN* | elastin | 1 | 0 |
| *PROCR* | protein C receptor, endothelial (EPCR) | 1 | 0 |
| *EPHX1* | epoxide hydrolase 1, microsomal (xenobiotic) | 5 | 0 |
| *EPHX2* | epoxide hydrolase 2, microsomal (xenobiotic) | 4 | 0 |
| *ESR1* | estrogen receptor 1 | 1 | 0 |
| *ESR2* | estrogen receptor 2 | 1 | 0 |
| *F12* | coagulation factor XII (Hageman factor) | 1 | 0 |
| *F13A1* | coagulation factor XIII, A1 polypeptide | 2 | 0 |
| *F13B* | coagulation factor XIII, B polypeptide | 1 | 0 |
| *F2* | coagulation factor II (thrombin) | 10 | 0 |
| *F3* | coagulation factor III (thromboplastin, tissue factor) | 1 | 0 |
| *F5* | coagulation factor V | 13 | 4 |
| *F7* | coagulation factor VII | 8 | 1 |
| *FABP2* | fatty acid binding protein 2, intestina | 1 | 0 |
| *FAS* | Fas | 4 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *FADS1* | Fatty acid desaturase 1 | 1 | 0 |
| *FADS2* | Fatty acid desaturase 2 | 1 | 0 |
| *FASLG* | Fas ligand | 4 | 0 |
| *FGB* | fibrinogen beta chain | 3 | 0 |
| *FGF1* | fibroblast growth factor 1 (acidic) | 1 | 0 |
| *FGF2* | fibroblast growth factor 2 (basic) | 1 | 0 |
| *FGF4* | fibroblast growth factor 4 | 1 | 0 |
| *FIGF* | c-fos induced growth factor (vascular endothelial growth factor D) | 1 | 0 |
| *FLT1* | fms-related tyrosine kinase 1 | 2 | 0 |
| *FLT4* | fms-related tyrosine kinase 4 | 1 | 0 |
| *FN1* | fibronectin 1 | 1 | 0 |
| *FSHR* | follicle stimulating hormone receptor | 1 | 1 |
| *GJA4* | gap junction protein, alpha 4 | 1 | 0 |
| *GJB2* | gap junction protein, beta 2, | 1 | 0 |
| *GNB3* | guanine nucleotide binding protein, beta polypeptide 3 | 2 | 0 |
| *GP1BA* | glycoprotein Ib (platelet), alpha polypeptide | 1 | 0 |
| *GSTM1* | glutathione S-transferase mu 1 | 4 | 2 |
| *GSTP1* | glutathione S-transferase pi 1 | 5 | 0 |
| *GSTT1* | glutathione S-transferase theta 1 | 5 | 2 |
| *GSTT2* | glutathione S-transferase theta 2 | 4 | 0 |
| *HLA-E* | major histocompatibility complex, class I, E | 1 | 0 |
| *HLA-G* | major histocompatibility complex, class I, G | 1 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *HMGCR* | 3-hydroxy-3-methylglutaryl-Coenzyme A reductase | 1 | 1 |
| *HPGD* | hydroxyprostaglandin dehydrogenase 15-(NAD) | 1 | 0 |
| *HSD11B1* | hydroxysteroid (11-beta) dehydrogenase 1 | 4 | 0 |
| *HSD17B7* | hydroxysteroid (17-beta) dehydrogenase 7 | 4 | 0 |
| *HSPA14* | heat shock 70kDa protein 14 | 4 | 0 |
| *HSPA1A* | heat shock 70kDa protein 1A | 4 | 0 |
| *HSPA1B* | heat shock 70kDa protein 1B | 4 | 0 |
| *HSPA1L* | heat shock 70kDa protein 1-like | 4 | 1 |
| *HSPA4* | heat shock 70kDa protein 4 | 4 | 0 |
| *HSPA6* | heat shock 70kDa protein 6 | 4 | 0 |
| *HSPG2* | heparan sulfate proteoglycan 2 | 1 | 0 |
| *HTR2A* | serotonin receptor 2A | 1 | 0 |
| *ICAM1* | intercellular adhesion molecule 1 | 2 | 1 |
| *ICAM3* | intercellular adhesion molecule 3 | 1 | 0 |
| *IFNG* | interferon, gamma | 6 | 2 |
| *IFNGR1* | interferon, gamma receptor 1 | 2 | 0 |
| *IFNGR2* | interferon, gamma receptor 2 | 1 | 0 |
| *IGF1* | insulin-like growth factor 1 | 5 | 0 |
| *IGF1R* | insulin-like growth factor 1 receptor | 1 | 0 |
| *IGF2* | insulin-like growth factor 2 | 1 | 1 |
| *IGF2R* | insulin-like growth factor 2 receptor | 1 | 0 |
| *IGFBP3* | insulin-like growth factor binding protein 3 | 4 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *IL10* | interleukin 10 | 11 | 2 |
| *IL10RA* | interleukin 10 receptor, alpha | 6 | 1 |
| *IL10RB* | interleukin 10 receptor, beta | 4 | 0 |
| *IL11* | interleukin 11 | 1 | 0 |
| *IL12A* | interleukin 12A | 2 | 0 |
| *IL12B* | interleukin 12B | 1 | 0 |
| *IL12RB1* | interleukin 12 receptor, beta 1 | 1 | 0 |
| *IL12RB2* | interleukin 12 receptor, beta 2 | 1 | 0 |
| *IL13* | interleukin 13 | 5 | 1 |
| *IL13RA2* | interleukin 13 receptor, alpha 2 | 1 | 0 |
| *IL15* | interleukin 15 | 4 | 1 |
| *IL18* | interleukin 18 | 6 | 0 |
| *IL18BP* | interleukin 18 binding protein | 1 | 0 |
| *IL1A* | interleukin 1, alpha | 10 | 2 |
| *IL1B* | interleukin 1, beta | 11 | 2 |
| *IL1R1* | interleukin 1 receptor, type I | 6 | 0 |
| *IL1R2* | interleukin 1 receptor, type 2 | 6 | 1 |
| *IL1RAP* | interleukin 1 receptor accessory protein | 4 | 0 |
| *IL1RAPL1* | interleukin 1 receptor accessory protein-like 1 | 1 | 0 |
| *IL1RN* | interleukin 1 receptor antagonist | 11 | 4 |
| *IL2* | interleukin 2 | 7 | 1 |
| *IL2RA* | interleukin 2 receptor, alpha | 5 | 1 |
| *IL2RB* | interleukin 2 receptor, beta | 4 | 1 |
| *IL3* | interleukin 3 | 1 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *IL3RA* | interleukin 3 receptor, alpha | 1 | 0 |
| *IL4* | interleukin 4 | 9 | 3 |
| *IL4R* | interleukin 4 receptor | 5 | 0 |
| *IL5* | interleukin 5 | 5 | 1 |
| *IL5RA* | interleukin 5 receptor, alpha | 1 | 0 |
| *IL6* | interleukin 6 | 16 | 5 |
| *IL6R* | interleukin 6 receptor | 8 | 4 |
| *IL8* | interleukin 8 | 8 | 0 |
| *IL8RA* | interleukin 8 receptor alpha | 5 | 0 |
| *IL8RB* | interleukin 8 receptor beta | 1 | 0 |
| *IL9* | interleukin 9 | 1 | 0 |
| *IL9R* | interleukin 9 receptor | 1 | 0 |
| *IRS1* | insulin receptor substrate 1 | 1 | 0 |
| *ITGA2* | integrin, alpha 2 | 1 | 0 |
| *ITGB3* | integrin, beta 3 | 2 | 0 |
| *KL* | Klotho | 4 | 1 |
| *LCAT* | lecithin-cholesterol acyltransferase | 1 | 0 |
| *LDLR* | LDL receptor | 1 | 0 |
| *LEP* | leptin | 1 | 0 |
| *LIPC* | lipase, hepatic | 2 | 0 |
| *LNPEP* | leucyl/cystinyl aminopeptidase | 1 | 0 |
| *LOXL1* | lysyl oxidase-like 1 | 1 | 0 |
| *LPA* | lipoprotein | 1 | 0 |
| *LPL* | lipoprotein lipase | 2 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *LST1* | leukocyte specific transcript 1 | 4 | 0 |
| *LTA* | lymphotoxin alpha | 6 | 1 |
| *LTF* | lactotransferrin | 1 | 0 |
| *LYZ* | lysozyme | 1 | 0 |
| *MASP2* | mannan-binding lectin serine peptidase 2 | 1 | 0 |
| *MBL2* | mannose-binding lectin 2, soluble | 11 | 6 |
| *MGP* | matrix Gla protein | 1 | 0 |
| *MIF* | macrophage migration inhibitory factor | 1 | 0 |
| *MMP1* | matrix metallopeptidase 1 | 6 | 1 |
| *MMP10* | matrix metallopeptidase 10 | 1 | 0 |
| *MMP11* | matrix metallopeptidase 11 | 1 | 0 |
| *MMP12* | matrix metallopeptidase 12 | 1 | 0 |
| *MMP13* | matrix metallopeptidase 13 | 1 | 0 |
| *MMP14* | matrix metallopeptidase 14 | 1 | 0 |
| *MMP15* | matrix metallopeptidase 15 | 1 | 0 |
| *MMP16* | matrix metallopeptidase 16 | 1 | 0 |
| *MMP17* | matrix metallopeptidase 17 | 1 | 0 |
| *MMP19* | matrix metallopeptidase 19 | 1 | 0 |
| *MMP2* | matrix metallopeptidase 2 | 5 | 1 |
| *MMP3* | matrix metallopeptidase 3 | 6 | 0 |
| *MMP7* | matrix metallopeptidase 7 | 1 | 0 |
| *MMP8* | matrix metallopeptidase 8 | 5 | 0 |
| *MMP9* | matrix metallopeptidase 9 | 6 | 2 |
| *MTHFD1* | methylenetetrahydrofolate dehydrogenase | 4 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *MTHFR* | 5,10-methylenetetrahydrofolate reductase | 12 | 0 |
| *MTR* | 5-methyltetrahydrofolate-homocysteine methyltransferase | 1 | 0 |
| *MTRR* | 5-methyltetrahydrofolate-homocysteine methyltransferase reductase | 2 | 1 |
| *NAT1* | N-acetyltransferase 1 | 5 | 0 |
| *NAT2* | N-acetyltransferase 2 | 6 | 0 |
| *NFKB1* | nuclear factor of kappa light polypeptide gene enhancer in B-cells 1 | 5 | 0 |
| *NFKB2* | nuclear factor of kappa light polypeptide gene enhancer in B-cells 2 | 4 | 0 |
| *NFKBIA* | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, alpha | 4 | 0 |
| *NFKBIB* | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, beta | 4 | 0 |
| *NFKBIE* | nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, epsilon | 4 | 0 |
| *NOD2* | nucleotide-binding oligomerization domain containing 2 | 2 | 1 |
| *NOD2/CARD 15* | nucleotide-binding oligomerization domain containing 2 | 5 | 0 |
| *NOS2A* | nitric oxide synthase 2, inducible | 3 | 2 |
| *NOS3* | nucleotide-binding oligomerization domain containing 3 | 8 | 1 |
| *NPPA* | natriuretic peptide precursor A | 2 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *NPR1* | natriuretic peptide receptor A | 1 | 0 |
| *NPY* | neuropeptide Y | 1 | 0 |
| *NQO1* | NAD(P)H dehydrogenase, quinone 1 | 1 | 0 |
| *NR3C1* | glucocorticoid receptor | 4 | 0 |
| *OPRM1* | opioid receptor, mu 1 | 1 | 1 |
| *OXT* | oxytocin | 1 | 0 |
| *OXTR* | oxytocin receptor | 2 | 1 |
| *PAFAH1B1* | platelet-activating factor acetylhydrolase, isoform Ib, subunit 1 | 5 | 0 |
| *PAFAH1B2* | platelet-activating factor acetylhydrolase, isoform Ib, subunit 2 | 4 | 0 |
| *PDE4D* | phosphodiesterase 4D, cAMP-specific | 1 | 0 |
| *PDGFB* | platelet-derived growth factor beta polypeptide | 1 | 0 |
| *PDGFC* | platelet derived growth factor C | 1 | 0 |
| *PECAM1* | platelet/endothelial cell adhesion molecule 1 | 1 | 0 |
| *PGEA1* | chibby homolog 1 | 4 | 0 |
| *PGF* | placental growth factor | 1 | 0 |
| *PIGF* | phosphatidylinositol glycan anchor biosynthesis, class F | 1 | 0 |
| *PLA2G4C* | phospholipase A2, group IVC (cytosolic, calcium-independent) | 1 | 1 |
| *PGR* | progesterone receptor | 9 | 1 |
| *PGRMC1* | progesterone receptor membrane component | 4 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| | 1 | | |
| *PGRMC2* | progesterone receptor membrane component 2 | 4 | 0 |
| *PLA2G4A* | phospholipase A2, group IVA | 4 | 0 |
| *PLAT* | plasminogen activator, tissue | 7 | 2 |
| *PLAU* | plasminogen activator, urokinase | 1 | 0 |
| *PLAUR* | plasminogen activator, urokinase receptor | 1 | 0 |
| *POMC* | proopiomelanocortin | 4 | 0 |
| *PON1* | paraoxonase 1 | 8 | 3 |
| *PON2* | paraoxonase 2 | 6 | 2 |
| *PPARA* | peroxisome proliferator-activated receptor alpha | 1 | 0 |
| *PPARG* | peroxisome proliferator-activated receptor gamma | 2 | 1 |
| *PRKCA* | protein kinase C, alpha | 1 | 1 |
| *PROC* | protein C | 2 | 0 |
| *PROS1* | protein S (alpha) | 1 | 0 |
| *PTCRA* | pre T-cell antigen receptor alpha | 4 | 0 |
| *PTGER1* | prostaglandin E receptor 1 | 1 | 0 |
| *PTGER2* | prostaglandin E receptor 2 | 5 | 1 |
| *PTGER3* | prostaglandin E receptor 3 | 4 | 2 |
| *PTGER4* | prostaglandin E receptor 4 | 1 | 0 |
| *PTGES* | prostaglandin E synthase | 6 | 0 |
| *PTGES2* | prostaglandin E synthase 2 | 2 | 1 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *PTGES3* | prostaglandin E synthase 3 | 2 | 0 |
| *PTGFR* | prostaglandin F receptor | 5 | 0 |
| *PTGS1* | prostaglandin G/H synthase (cyclooxygenase) | 6 | 0 |
| *PTGS2* | prostaglandin-endoperoxide synthase 2 | 6 | 0 |
| *PTPN22* | protein tyrosine phosphatase, non-receptor type 22 | 4 | 0 |
| *REN* | renin | 2 | 0 |
| *RFC1* | replication factor C (activator 1) 1 | 1 | 0 |
| *RLN1* | Relaxin 1 | 1 | 0 |
| *RLN2* | Relaxin 2 | 1 | 0 |
| *RLN3* | Relaxin 3 | 1 | 0 |
| *SCGB1A1* | secretoglobin, family 1A, member 1 (uteroglobin) | 4 | 0 |
| *SCNN1A* | sodium channel, nonvoltage-gated 1 alpha | 1 | 0 |
| *SELE* | selectin E | 3 | 0 |
| *SELP* | selectin P | 1 | 0 |
| *SERPINB2* | serpin peptidase inhibitor, clade B (ovalbumin), member 2 | 1 | 0 |
| *SERPINC1* | serpin peptidase inhibitor, clade C, member 1 | 1 | 0 |
| *SERPINE1* | serpin peptidase inhibitor, clade E, member 1 | 7 | 1 |
| *SERPINH1* | serpin peptidase inhibitor, clade H, member 1 (collagen binding protein 1) | 6 | 2 |
| *SFTPA1* | surfactant protein A1 | 1 | 0 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *SFTPA2* | surfactant protein A2 | 1 | 0 |
| *SFTPC* | Surfactant protein C | 2 | 0 |
| *SFTPD* | Surfactant protein D | 2 | 1 |
| *SHMT1* | serine hydroxymethyltransferase 1 | 1 | 1 |
| *SLC23A1* | solute carrier family 23 (nucleobase transporters), member 1 | 5 | 0 |
| *SLC23A2* | solute carrier family 23, member 2 | 1 | 1 |
| *SLC6A4* | solute carrier family 6, member 4 | 4 | 0 |
| *SOD3* | superoxide dismutase 3, extracellular | 1 | 0 |
| *SPARC* | secreted protein, acidic, cysteine-rich | 1 | 0 |
| *TAP1* | transporter 1, ATP-binding cassette, sub-family B | 1 | 0 |
| *TBXAS1* | thromboxane A synthase 1 (platelet) | 1 | 0 |
| *TCN2* | transcobalamin II | 4 | 0 |
| *TFPI* | tissue factor pathway inhibitor | 1 | 0 |
| *TGFA* | transforming growth factor, alpha | 1 | 0 |
| *TGFB* | transforming growth factor, beta | 1 | 0 |
| *TGFB1* | transforming growth factor, beta 1 | 7 | 0 |
| *THBD* | thrombomodulin | 2 | 1 |
| *THBS1* | thrombospondin 1 | 1 | 0 |
| *THBS2* | thrombospondin 2 | 1 | 0 |
| *THPO* | thrombopoietin | 1 | 0 |
| *TIMP1* | TIMP metallopeptidase inhibitor 1 | 1 | 0 |
| *TIMP2* | TIMP metallopeptidase inhibitor 2 | 1 | 1 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| *TIMP3* | TIMP metallopeptidase inhibitor 3 | 4 | 0 |
| *TIMP4* | TIMP metallopeptidase inhibitor 4 | 4 | 0 |
| *TLR1* | toll-like receptor 1 | 1 | 0 |
| *TLR10* | toll-like receptor 10 | 1 | 1 |
| *TLR2* | toll-like receptor 2 | 7 | 1 |
| *TLR3* | toll-like receptor 3 | 5 | 0 |
| *TLR4* | toll-like receptor 4 | 11 | 2 |
| *TLR5* | toll-like receptor 5 | 1 | 0 |
| *TLR6* | toll-like receptor 6 | 1 | 0 |
| *TLR7* | toll-like receptor 7 | 4 | 0 |
| *TLR8* | toll-like receptor 8 | 4 | 0 |
| *TLR9* | toll-like receptor 9 | 5 | 0 |
| *TNF* | tumor necrosis factor | 27 | 12 |
| *TNFR1* | tumor necrosis factor receptor 1 | 7 | 4 |
| *TNFR2* | tumor necrosis factor receptor 2 | 3 | 2 |
| *TNFRSF1A* | tumor necrosis factor receptor superfamily, member 1A | 5 | 0 |
| *TNFRSF1B* | tumor necrosis factor receptor superfamily, member 1B | 5 | 0 |
| *TNFRSF6* | tumor necrosis factor receptor superfamily, member 6b, decoy | 2 | 1 |
| *TNR* | tenascin R | 1 | 0 |
| *TRAF2* | TNF receptor-associated factor 2 | 4 | 0 |
| *TREM1* | triggering receptor expressed on myeloid cells | 4 | 1 |

| Gene Symbol | Gene Name | Number of Studies | Studies reporting positive findings |
|---|---|---|---|
| | 1 | | |
| *TSHR* | thyroid stimulating hormone receptor | 4 | 0 |
| *UGT1A1* | UDP glucuronosyltransferase 1 family, polypeptide A1 | 4 | 0 |
| *VEGF* | vascular endothelial growth factor A | 7 | 1 |
| *VEGFB* | vascular endothelial growth factor B | 1 | 0 |
| *VEGFC* | vascular endothelial growth factor C | 1 | 0 |
| *VWF* | von Willebrand factor | 1 | 0 |

**Appendix B. All SNPs with p-value < $10^{-4}$ for chi-square PTB association in Finnish mothers genotyped on the Affymetrix 6.0 array.** Base pair (BP) positions refer to NCBI36 (hg18, March 2006 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | FREQ CASE | FREQ CONTROL | MAJOR ALLELE | P | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 4 | rs871476 | 25152250 | G | 0.137 | 0.247 | A | 6.22E-06 | 0.48 | 0.35 | 0.67 |
| 6 | rs732496 | 152995144 | C | 0.362 | 0.496 | A | 1.15E-05 | 0.58 | 0.45 | 0.74 |
| 17 | rs222745 | 3435620 | T | 0.155 | 0.071 | C | 1.17E-05 | 2.41 | 1.61 | 3.60 |
| 4 | rs904132 | 55496865 | G | 0.093 | 0.030 | A | 1.72E-05 | 3.30 | 1.86 | 5.84 |
| 13 | rs17337271 | 23250220 | C | 0.036 | 0.103 | G | 2.00E-05 | 0.32 | 0.19 | 0.56 |
| 14 | rs767757 | 60944127 | G | 0.318 | 0.204 | T | 2.04E-05 | 1.82 | 1.38 | 2.40 |
| 10 | rs11193681 | 84068972 | T | 0.155 | 0.073 | C | 2.14E-05 | 2.32 | 1.56 | 3.45 |
| 12 | rs2302728 | 2644929 | C | 0.205 | 0.320 | A | 2.29E-05 | 0.55 | 0.41 | 0.73 |
| 6 | rs1110227 | 153003973 | A | 0.365 | 0.493 | G | 2.51E-05 | 0.59 | 0.46 | 0.76 |
| 4 | rs12648485 | 55499294 | C | 0.089 | 0.029 | T | 2.52E-05 | 3.31 | 1.84 | 5.95 |
| 10 | rs3908834 | 84068197 | T | 0.153 | 0.073 | C | 2.85E-05 | 2.30 | 1.54 | 3.41 |
| 2 | rs332867 | 117602473 | G | 0.273 | 0.167 | A | 2.85E-05 | 1.87 | 1.39 | 2.51 |
| 6 | rs9384026 | 153035435 | C | 0.370 | 0.497 | G | 3.27E-05 | 0.60 | 0.47 | 0.76 |
| 3 | rs4682892 | 43558769 | G | 0.262 | 0.159 | T | 3.30E-05 | 1.88 | 1.39 | 2.53 |
| 22 | rs9605923 | 15445079 | T | 0.208 | 0.320 | A | 3.52E-05 | 0.56 | 0.42 | 0.74 |
| 3 | rs7428158 | 43558323 | C | 0.262 | 0.160 | T | 4.07E-05 | 1.86 | 1.38 | 2.51 |
| 1 | rs1890844 | 208973862 | T | 0.516 | 0.391 | G | 4.48E-05 | 1.66 | 1.30 | 2.12 |
| 12 | rs4765701 | 2641152 | T | 0.153 | 0.255 | C | 4.64E-05 | 0.53 | 0.39 | 0.72 |
| 5 | rs17463165 | 39140675 | A | 0.337 | 0.459 | G | 4.80E-05 | 0.60 | 0.47 | 0.77 |
| 2 | rs12464127 | 19949783 | C | 0.193 | 0.300 | T | 4.95E-05 | 0.56 | 0.42 | 0.74 |
| 18 | rs4798834 | 9714549 | A | 0.242 | 0.356 | G | 5.19E-05 | 0.58 | 0.44 | 0.75 |
| 3 | rs17407870 | 43471145 | G | 0.262 | 0.161 | A | 5.22E-05 | 1.84 | 1.37 | 2.49 |

| CHR | SNP | BP | MINOR ALLELE | FREQ CASE | FREQ CONTROL | MAJOR ALLELE | P | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 10 | rs3904726 | 84081058 | G | 0.153 | 0.075 | A | 5.45E-05 | 2.22 | 1.50 | 3.29 |
| 16 | rs344357 | 1776256 | G | 0.223 | 0.335 | C | 5.59E-05 | 0.57 | 0.43 | 0.75 |
| 3 | rs3755602 | 43594314 | A | 0.262 | 0.162 | G | 5.67E-05 | 1.84 | 1.36 | 2.47 |
| 3 | rs6809134 | 43559114 | T | 0.263 | 0.163 | C | 6.24E-05 | 1.83 | 1.36 | 2.47 |
| 6 | rs9479367 | 153014823 | T | 0.363 | 0.484 | C | 6.36E-05 | 0.61 | 0.48 | 0.78 |
| 6 | rs12200492 | 47646089 | C | 0.472 | 0.352 | T | 6.48E-05 | 1.65 | 1.29 | 2.10 |
| 9 | rs7045593 | 1530499 | T | 0.145 | 0.070 | C | 6.50E-05 | 2.26 | 1.50 | 3.41 |
| 2 | rs11676603 | 209969736 | G | 0.300 | 0.193 | C | 6.60E-05 | 1.79 | 1.34 | 2.38 |
| 4 | rs2051428 | 100342209 | C | 0.239 | 0.144 | T | 6.99E-05 | 1.87 | 1.37 | 2.55 |
| 2 | rs7602876 | 117576259 | T | 0.276 | 0.175 | C | 7.00E-05 | 1.80 | 1.34 | 2.41 |
| 2 | rs13396426 | 42965448 | C | 0.407 | 0.528 | T | 7.03E-05 | 0.61 | 0.48 | 0.78 |
| 14 | rs9944098 | 90557351 | T | 0.460 | 0.342 | C | 7.31E-05 | 1.64 | 1.29 | 2.10 |
| 15 | rs1868243 | 55667810 | C | 0.344 | 0.234 | T | 7.42E-05 | 1.72 | 1.31 | 2.25 |
| 2 | rs12986437 | 20068171 | G | 0.213 | 0.323 | T | 7.54E-05 | 0.57 | 0.43 | 0.75 |
| 16 | rs169844 | 16162267 | C | 0.468 | 0.350 | T | 7.67E-05 | 1.64 | 1.28 | 2.09 |
| 3 | rs2372433 | 43473252 | C | 0.260 | 0.162 | T | 7.72E-05 | 1.82 | 1.35 | 2.45 |
| 3 | rs17407912 | 43471255 | G | 0.262 | 0.164 | A | 7.83E-05 | 1.81 | 1.35 | 2.44 |
| 6 | rs1744397 | 152962545 | G | 0.387 | 0.509 | C | 8.23E-05 | 0.61 | 0.48 | 0.78 |
| 8 | rs9918898 | 94373271 | A | 0.028 | 0.084 | G | 8.52E-05 | 0.31 | 0.17 | 0.58 |
| 12 | rs961445 | 73790333 | C | 0.175 | 0.093 | T | 8.67E-05 | 2.06 | 1.43 | 2.96 |
| 3 | rs13063227 | 172429742 | A | 0.184 | 0.100 | G | 9.29E-05 | 2.01 | 1.41 | 2.88 |
| 6 | rs9395272 | 47603093 | G | 0.480 | 0.362 | A | 9.66E-05 | 1.63 | 1.27 | 2.08 |
| 21 | rs243693 | 28127529 | T | 0.366 | 0.484 | C | 9.70E-05 | 0.61 | 0.48 | 0.79 |
| 5 | rs28050 | 96171180 | G | 0.280 | 0.180 | A | 9.74E-05 | 1.77 | 1.33 | 2.36 |

| CHR | SNP | BP | MINOR ALLELE | FREQ CASE | FREQ CONTROL | MAJOR ALLELE | P | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|---|---|---|
| 6 | rs873889 | 153033302 | G | 0.364 | 0.482 | A | 9.76E-05 | 0.61 | 0.48 | 0.78 |
| 4 | rs7663862 | 55454959 | G | 0.089 | 0.033 | A | 9.83E-05 | 2.86 | 1.65 | 4.97 |
| 2 | rs4665765 | 25216019 | T | 0.355 | 0.473 | C | 9.91E-05 | 0.61 | 0.48 | 0.78 |
| 16 | rs2247696 | 1792776 | T | 0.209 | 0.315 | C | 9.92E-05 | 0.58 | 0.44 | 0.76 |
| 2 | rs332868 | 117602387 | C | 0.286 | 0.185 | T | 9.97E-05 | 1.76 | 1.32 | 2.34 |

**Appendix C. A complete list of all SNPs with p-value < $10^{-4}$ from the unadjusted additive linear regression for the Box-Cox transformed gestational ages for Finnish mothers genotyped on the Affymetrix 6.0 SNP array.** Base pair (BP) positions refer to NCBI36 (hg18, March 2006 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | P |
|---|---|---|---|---|---|---|---|
| 1 | rs10874644 | 102931572 | G | 2.34E+09 | 1.34E+09 | 3.34E+09 | 5.26E-06 |
| 2 | rs7583085 | 42876472 | G | 1.78E+09 | 1.02E+09 | 2.54E+09 | 5.77E-06 |
| 2 | rs7569325 | 42978493 | G | 1.81E+09 | 1.02E+09 | 2.59E+09 | 7.54E-06 |
| 2 | rs12464127 | 19949783 | C | 1.77E+09 | 9.92E+08 | 2.55E+09 | 1.04E-05 |
| 6 | rs4714551 | 42098919 | A | -1.65E+09 | -2.37E+09 | -9.18E+08 | 1.13E-05 |
| 5 | rs10070878 | 28220077 | T | -2.42E+09 | -3.49E+09 | -1.35E+09 | 1.14E-05 |
| 6 | rs12208368 | 76474506 | T | -2.12E+09 | -3.06E+09 | -1.18E+09 | 1.21E-05 |
| 12 | rs12827976 | 65560052 | T | 1.70E+09 | 9.46E+08 | 2.46E+09 | 1.22E-05 |
| 4 | rs41389750 | 11370357 | C | 2.40E+09 | 1.33E+09 | 3.46E+09 | 1.30E-05 |
| 2 | rs920391 | 42872851 | C | 1.73E+09 | 9.56E+08 | 2.49E+09 | 1.34E-05 |
| 2 | rs7582883 | 42876508 | C | 1.70E+09 | 9.37E+08 | 2.46E+09 | 1.42E-05 |
| 6 | rs4711698 | 42095429 | G | -1.63E+09 | -2.36E+09 | -8.99E+08 | 1.51E-05 |
| 13 | rs17337271 | 23250220 | C | 2.78E+09 | 1.53E+09 | 4.03E+09 | 1.55E-05 |
| 4 | rs871476 | 25152250 | G | 1.80E+09 | 9.90E+08 | 2.62E+09 | 1.65E-05 |
| 6 | rs732496 | 152995144 | C | 1.51E+09 | 8.30E+08 | 2.19E+09 | 1.66E-05 |
| 2 | rs6736894 | 42995273 | A | 1.67E+09 | 9.14E+08 | 2.43E+09 | 1.82E-05 |
| 23 | rs1172046 | 103629907 | T | 1.60E+09 | 8.75E+08 | 2.33E+09 | 1.89E-05 |
| 10 | rs1904693 | 52585849 | G | 1.49E+09 | 8.10E+08 | 2.16E+09 | 1.92E-05 |
| 6 | rs12210386 | 76610556 | A | -2.17E+09 | -3.16E+09 | -1.18E+09 | 2.05E-05 |
| 6 | rs6901077 | 42097696 | A | -1.61E+09 | -2.34E+09 | -8.70E+08 | 2.24E-05 |
| 14 | rs767757 | 60944127 | G | -1.61E+09 | -2.35E+09 | -8.70E+08 | 2.34E-05 |
| 12 | rs2302728 | 2644929 | C | 1.64E+09 | 8.87E+08 | 2.39E+09 | 2.35E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | P |
|---|---|---|---|---|---|---|---|
| 2 | rs332867 | 117602473 | G | -1.74E+09 | -2.54E+09 | -9.40E+08 | 2.38E-05 |
| 12 | rs12820694 | 65550268 | T | 1.59E+09 | 8.60E+08 | 2.32E+09 | 2.39E-05 |
| 14 | rs17127816 | 54264954 | C | -1.41E+09 | -2.06E+09 | -7.60E+08 | 2.43E-05 |
| 12 | rs7310441 | 65540438 | T | 1.63E+09 | 8.80E+08 | 2.38E+09 | 2.46E-05 |
| 7 | rs10225158 | 76939468 | C | -1.53E+09 | -2.24E+09 | -8.27E+08 | 2.49E-05 |
| 2 | rs4350800 | 42966710 | G | 1.72E+09 | 9.30E+08 | 2.52E+09 | 2.49E-05 |
| 12 | rs17182681 | 65560266 | A | 1.63E+09 | 8.76E+08 | 2.38E+09 | 2.59E-05 |
| 23 | rs5916847 | 104101048 | A | 1.44E+09 | 7.74E+08 | 2.11E+09 | 2.82E-05 |
| 6 | rs9352244 | 76628350 | A | -2.12E+09 | -3.10E+09 | -1.14E+09 | 2.82E-05 |
| 6 | rs9360946 | 76629692 | A | -2.12E+09 | -3.10E+09 | -1.14E+09 | 2.82E-05 |
| 6 | rs2208798 | 76640363 | T | -2.12E+09 | -3.10E+09 | -1.14E+09 | 2.82E-05 |
| 6 | rs12196105 | 76575424 | T | -2.12E+09 | -3.10E+09 | -1.14E+09 | 2.82E-05 |
| 16 | rs344357 | 1776256 | G | 1.57E+09 | 8.43E+08 | 2.30E+09 | 2.85E-05 |
| 13 | rs7999686 | 109785738 | A | 1.44E+09 | 7.72E+08 | 2.11E+09 | 2.90E-05 |
| 6 | rs12200169 | 76293841 | T | -1.98E+09 | -2.91E+09 | -1.06E+09 | 2.94E-05 |
| 16 | rs4781675 | 9708081 | T | 1.48E+09 | 7.91E+08 | 2.17E+09 | 3.07E-05 |
| 2 | rs4953664 | 42878180 | C | 1.67E+09 | 8.93E+08 | 2.45E+09 | 3.09E-05 |
| 5 | rs52252 | 141944700 | G | 1.85E+09 | 9.88E+08 | 2.72E+09 | 3.13E-05 |
| 12 | rs2098414 | 95798431 | C | 1.77E+09 | 9.39E+08 | 2.60E+09 | 3.35E-05 |
| 11 | rs7945752 | 33916012 | A | -1.58E+09 | -2.32E+09 | -8.39E+08 | 3.37E-05 |
| 11 | rs16908162 | 10655488 | C | -1.53E+09 | -2.25E+09 | -8.12E+08 | 3.42E-05 |
| 6 | rs13216921 | 76640907 | A | -2.12E+09 | -3.11E+09 | -1.12E+09 | 3.62E-05 |
| 18 | rs1945148 | 20387524 | G | -1.50E+09 | -2.20E+09 | -7.93E+08 | 3.67E-05 |
| 2 | rs4594497 | 42959781 | A | 1.65E+09 | 8.71E+08 | 2.42E+09 | 3.73E-05 |
| 7 | rs16219 | 24224133 | C | 1.48E+09 | 7.80E+08 | 2.18E+09 | 4.11E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | P |
|---|---|---|---|---|---|---|---|
| 9 | rs7045593 | 1530499 | T | -2.24E+09 | -3.30E+09 | -1.17E+09 | 4.43E-05 |
| 10 | rs1904692 | 52585942 | T | 1.45E+09 | 7.58E+08 | 2.14E+09 | 4.67E-05 |
| 5 | rs4702798 | 11377554 | C | 1.57E+09 | 8.18E+08 | 2.31E+09 | 4.71E-05 |
| 12 | rs4765701 | 2641152 | T | 1.73E+09 | 9.01E+08 | 2.55E+09 | 4.72E-05 |
| 8 | rs4633077 | 6059848 | G | -1.48E+09 | -2.18E+09 | -7.70E+08 | 4.77E-05 |
| 8 | rs2320562 | 24782519 | T | 1.37E+09 | 7.15E+08 | 2.03E+09 | 4.82E-05 |
| 12 | rs7967486 | 3780172 | T | -1.59E+09 | -2.35E+09 | -8.26E+08 | 5.00E-05 |
| 16 | rs1420050 | 9719224 | T | 1.38E+09 | 7.17E+08 | 2.04E+09 | 5.00E-05 |
| 2 | rs12986437 | 20068171 | G | 1.62E+09 | 8.40E+08 | 2.39E+09 | 5.15E-05 |
| 2 | rs3816184 | 42869261 | G | 1.58E+09 | 8.20E+08 | 2.34E+09 | 5.23E-05 |
| 8 | rs2632839 | 18718981 | G | -2.21E+09 | -3.27E+09 | -1.15E+09 | 5.23E-05 |
| 2 | rs7602876 | 117576259 | T | -1.62E+09 | -2.40E+09 | -8.40E+08 | 5.27E-05 |
| 6 | rs1110227 | 153003973 | A | 1.38E+09 | 7.13E+08 | 2.04E+09 | 5.66E-05 |
| 13 | rs10507333 | 23267776 | A | 2.70E+09 | 1.40E+09 | 4.00E+09 | 5.68E-05 |
| 6 | rs9350591 | 76298247 | T | -1.89E+09 | -2.80E+09 | -9.77E+08 | 5.69E-05 |
| 4 | rs16841283 | 7844571 | A | -2.17E+09 | -3.22E+09 | -1.12E+09 | 5.74E-05 |
| 11 | rs1120306 | 21719669 | C | 2.18E+09 | 1.12E+09 | 3.23E+09 | 6.10E-05 |
| 22 | rs5768864 | 45284762 | C | 1.74E+09 | 8.93E+08 | 2.58E+09 | 6.15E-05 |
| 3 | rs4682892 | 43558769 | G | -1.67E+09 | -2.48E+09 | -8.59E+08 | 6.21E-05 |
| 8 | rs2638610 | 18719115 | G | -2.20E+09 | -3.26E+09 | -1.13E+09 | 6.30E-05 |
| 6 | rs12207159 | 76609496 | A | -2.41E+09 | -3.59E+09 | -1.24E+09 | 6.35E-05 |
| 23 | rs3135207 | 122776922 | C | -1.46E+09 | -2.17E+09 | -7.47E+08 | 6.50E-05 |
| 20 | rs4813182 | 15424232 | G | -2.66E+09 | -3.96E+09 | -1.37E+09 | 6.58E-05 |
| 6 | rs3798430 | 76660572 | C | -2.05E+09 | -3.05E+09 | -1.05E+09 | 6.60E-05 |
| 2 | rs9917172 | 42963980 | C | 1.57E+09 | 8.06E+08 | 2.34E+09 | 6.63E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | P |
|---|---|---|---|---|---|---|---|
| 2 | rs11883440 | 42976693 | A | 1.62E+09 | 8.29E+08 | 2.41E+09 | 6.89E-05 |
| 3 | rs7428158 | 43558323 | C | -1.66E+09 | -2.47E+09 | -8.46E+08 | 6.97E-05 |
| 10 | rs11193681 | 84068972 | T | -2.15E+09 | -3.20E+09 | -1.10E+09 | 7.05E-05 |
| 8 | rs11786728 | 24801098 | G | 1.35E+09 | 6.87E+08 | 2.01E+09 | 7.34E-05 |
| 6 | rs12202443 | 76295334 | T | -1.89E+09 | -2.81E+09 | -9.61E+08 | 7.35E-05 |
| 3 | rs978485 | 64298841 | T | 1.66E+09 | 8.45E+08 | 2.47E+09 | 7.37E-05 |
| 22 | rs9605923 | 15445079 | T | 1.46E+09 | 7.43E+08 | 2.18E+09 | 7.73E-05 |
| 6 | rs9384026 | 153035435 | C | 1.38E+09 | 7.00E+08 | 2.06E+09 | 7.84E-05 |
| 2 | rs1861459 | 51547229 | C | 1.59E+09 | 8.06E+08 | 2.37E+09 | 7.93E-05 |
| 5 | rs258833 | 11377531 | A | 1.55E+09 | 7.85E+08 | 2.31E+09 | 8.08E-05 |
| 2 | rs437866 | 126864929 | A | -1.35E+09 | -2.01E+09 | -6.81E+08 | 8.28E-05 |
| 10 | rs3908834 | 84068197 | T | -2.13E+09 | -3.18E+09 | -1.07E+09 | 8.61E-05 |
| 16 | rs11645147 | 9709958 | A | 1.33E+09 | 6.69E+08 | 1.99E+09 | 8.69E-05 |
| 2 | rs4953687 | 43002989 | A | 1.51E+09 | 7.60E+08 | 2.26E+09 | 8.88E-05 |
| 9 | rs12683498 | 84671083 | G | -1.37E+09 | -2.05E+09 | -6.90E+08 | 8.93E-05 |
| 20 | rs1412977 | 56142696 | T | -1.38E+09 | -2.07E+09 | -6.95E+08 | 8.97E-05 |
| 3 | rs17407870 | 43471145 | G | -1.62E+09 | -2.43E+09 | -8.15E+08 | 9.19E-05 |
| 6 | rs9447540 | 76512109 | G | -1.92E+09 | -2.87E+09 | -9.64E+08 | 9.23E-05 |
| 3 | rs3755602 | 43594314 | A | -1.62E+09 | -2.42E+09 | -8.12E+08 | 9.30E-05 |
| 14 | rs17094971 | 58252951 | A | 2.27E+09 | 1.14E+09 | 3.40E+09 | 9.34E-05 |
| 20 | rs16980283 | 54765104 | A | 1.66E+09 | 8.31E+08 | 2.48E+09 | 9.35E-05 |
| 11 | rs10767349 | 3597975 | G | -2.02E+09 | -3.03E+09 | -1.01E+09 | 9.55E-05 |
| 18 | rs4798834 | 9714549 | A | 1.41E+09 | 7.05E+08 | 2.11E+09 | 9.62E-05 |

**Appendix D. All SNPs with p-value < $10^{-4}$ from PTB additive logistic regression adjusted for body mass index (BMI), gravidity and smoking status association analysis for the Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | ADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 14 | rs7153053 | 105150273 | T | 5.72E-07 | 2.11 | 1.58 | 2.83 |
| 2 | rs651418 | 4495174 | T | 4.61E-06 | 0.47 | 0.34 | 0.65 |
| 13 | KGP1703195 | 23467854 | C | 6.65E-06 | 0.44 | 0.31 | 0.63 |
| 4 | KGP8375539 | 56993480 | G | 6.75E-06 | 0.49 | 0.36 | 0.67 |
| 12 | KGP2468001 | 123572794 | G | 7.01E-06 | 4.51 | 2.34 | 8.70 |
| 13 | rs9510404 | 23462530 | G | 8.20E-06 | 0.44 | 0.31 | 0.63 |
| 22 | KGP11155947 | 19226260 | A | 9.16E-06 | 4.05 | 2.18 | 7.52 |
| 22 | rs712950 | 19204210 | C | 9.55E-06 | 4.04 | 2.18 | 7.49 |
| 18 | KGP10756992 | 55899135 | C | 9.68E-06 | 2.56 | 1.69 | 3.87 |
| 12 | KGP6141129 | 31496481 | A | 1.07E-05 | 0.47 | 0.34 | 0.66 |
| 11 | rs4757417 | 16615280 | C | 1.13E-05 | 2.00 | 1.47 | 2.73 |
| 4 | KGP4968386 | 27091299 | C | 1.31E-05 | 0.37 | 0.23 | 0.58 |
| 11 | KGP9466771 | 16637583 | T | 1.36E-05 | 1.99 | 1.46 | 2.71 |
| 3 | KGP2829197 | 59893355 | A | 1.46E-05 | 2.12 | 1.51 | 2.98 |
| 10 | KGP5967610 | 117393184 | C | 1.56E-05 | 2.14 | 1.52 | 3.03 |
| 10 | rs17724010 | 117273552 | C | 1.76E-05 | 2.12 | 1.51 | 2.99 |
| 23 | rs17215777 | 14045928 | A | 1.90E-05 | 2.82 | 1.75 | 4.54 |
| 19 | rs3745751 | 48519241 | C | 2.09E-05 | 2.84 | 1.76 | 4.59 |
| 22 | rs1633399 | 19183787 | G | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 22 | KGP3688261 | 19195680 | C | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 22 | rs712952 | 19197949 | A | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 22 | rs807459 | 19223352 | C | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 22 | KGP4024090 | 19280168 | G | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 22 | KGP2366391 | 19283478 | G | 2.09E-05 | 3.92 | 2.09 | 7.34 |
| 9 | KGP535977 | 73076516 | A | 2.14E-05 | 0.45 | 0.31 | 0.65 |
| 19 | rs16982100 | 48511888 | G | 2.36E-05 | 3.38 | 1.92 | 5.94 |
| 9 | rs10780947 | 73165078 | T | 2.40E-05 | 0.51 | 0.37 | 0.70 |
| 13 | rs9571702 | 67532237 | A | 2.41E-05 | 0.46 | 0.32 | 0.66 |
| 10 | KGP10886825 | 117230790 | C | 2.43E-05 | 2.10 | 1.49 | 2.97 |
| 8 | rs900213 | 3454880 | T | 2.57E-05 | 1.89 | 1.41 | 2.55 |
| 10 | KGP349528 | 131918109 | G | 2.65E-05 | 2.17 | 1.51 | 3.11 |
| 10 | KGP22784786 | 116847333 | T | 2.81E-05 | 1.93 | 1.42 | 2.63 |

| CHR | SNP | BP | MINOR ALLELE | ADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 13 | rs1028730 | 67527244 | C | 2.86E-05 | 0.48 | 0.34 | 0.68 |
| 10 | rs1171727 | 131908016 | G | 3.04E-05 | 2.14 | 1.50 | 3.05 |
| 10 | rs1264794 | 116902943 | C | 3.23E-05 | 1.94 | 1.42 | 2.65 |
| 2 | rs662510 | 4491654 | T | 3.32E-05 | 0.50 | 0.36 | 0.69 |
| 10 | KGP6938912 | 116941272 | G | 3.36E-05 | 1.93 | 1.41 | 2.63 |
| 18 | rs1529905 | 73077313 | T | 3.52E-05 | 0.33 | 0.20 | 0.56 |
| 14 | rs11160817 | 105168414 | T | 3.57E-05 | 1.81 | 1.37 | 2.40 |
| 5 | rs2270822 | 60459040 | A | 3.65E-05 | 0.27 | 0.15 | 0.50 |
| 5 | KGP10062212 | 60367377 | C | 3.67E-05 | 0.27 | 0.15 | 0.50 |
| 5 | KGP9219331 | 60457627 | T | 3.67E-05 | 0.27 | 0.15 | 0.50 |
| 3 | rs1864427 | 5626851 | A | 3.69E-05 | 0.52 | 0.38 | 0.71 |
| 9 | KGP3932024 | 127536726 | A | 3.71E-05 | 1.88 | 1.39 | 2.53 |
| 9 | KGP1766722 | 73095674 | T | 3.98E-05 | 0.46 | 0.32 | 0.67 |
| 13 | KGP7715737 | 67553342 | T | 4.10E-05 | 0.47 | 0.33 | 0.68 |
| 15 | rs11855354 | 78443631 | G | 4.12E-05 | 0.55 | 0.41 | 0.73 |
| 15 | rs2304824 | 78466127 | A | 4.12E-05 | 0.55 | 0.41 | 0.73 |
| 2 | KGP11014586 | 187487160 | G | 4.62E-05 | 1.90 | 1.40 | 2.60 |
| 2 | rs2197138 | 187493841 | A | 4.62E-05 | 1.90 | 1.40 | 2.60 |
| 2 | rs7596996 | 187494427 | G | 4.62E-05 | 1.90 | 1.40 | 2.60 |
| 10 | rs1899721 | 116965037 | T | 4.86E-05 | 1.90 | 1.39 | 2.59 |
| 13 | KGP9920102 | 67545692 | G | 4.96E-05 | 0.48 | 0.34 | 0.69 |
| 7 | KGP11556052 | 22141292 | A | 5.13E-05 | 2.00 | 1.43 | 2.79 |
| 19 | KGP11012186 | 48517676 | T | 5.15E-05 | 3.22 | 1.83 | 5.66 |
| 19 | KGP3429713 | 48522480 | A | 5.15E-05 | 3.22 | 1.83 | 5.66 |
| 19 | KGP12010912 | 48522553 | C | 5.15E-05 | 3.22 | 1.83 | 5.66 |
| 23 | rs5965496 | 67319094 | C | 5.15E-05 | 0.45 | 0.30 | 0.66 |
| 3 | KGP5507304 | 47649998 | C | 5.20E-05 | 1.95 | 1.41 | 2.70 |
| 3 | rs1014228 | 47652639 | C | 5.20E-05 | 1.95 | 1.41 | 2.70 |
| 1 | KGP4299054 | 243602390 | A | 5.24E-05 | 1.90 | 1.39 | 2.60 |
| 5 | rs2071239 | 149755421 | G | 5.36E-05 | 2.31 | 1.54 | 3.47 |
| 9 | KGP1996626 | 434222 | A | 5.38E-05 | 2.00 | 1.43 | 2.80 |
| 19 | KGP1072697 | 48522746 | A | 5.42E-05 | 3.21 | 1.82 | 5.64 |
| 1 | rs3842895 | 180596196 | C | 5.44E-05 | 0.53 | 0.39 | 0.72 |
| 5 | KGP12461110 | 60199201 | C | 5.44E-05 | 0.21 | 0.10 | 0.44 |
| 5 | rs976080 | 60217651 | T | 5.44E-05 | 0.21 | 0.10 | 0.44 |
| 2 | KGP22789102 | 187483472 | C | 5.45E-05 | 1.89 | 1.39 | 2.58 |
| 5 | rs7713638 | 149759096 | C | 5.48E-05 | 2.30 | 1.53 | 3.44 |
| 19 | KGP1220612 | 48517387 | A | 5.60E-05 | 3.26 | 1.83 | 5.79 |
| 4 | KGP11607606 | 91912899 | A | 5.67E-05 | 2.26 | 1.52 | 3.37 |

| CHR | SNP | BP | MINOR ALLELE | ADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 10 | rs1414632 | 117244968 | G | 5.70E-05 | 2.02 | 1.43 | 2.84 |
| 17 | KGP3405370 | 9719103 | C | 5.77E-05 | 1.79 | 1.35 | 2.38 |
| 12 | rs12300476 | 31514886 | C | 5.79E-05 | 2.01 | 1.43 | 2.82 |
| 2 | rs2370693 | 187475104 | G | 5.81E-05 | 1.89 | 1.39 | 2.57 |
| 3 | rs3732530 | 47618953 | C | 5.94E-05 | 1.95 | 1.41 | 2.70 |
| 5 | rs1559050 | 134470030 | T | 5.95E-05 | 1.92 | 1.40 | 2.65 |
| 2 | rs4233788 | 187462581 | A | 6.12E-05 | 1.89 | 1.38 | 2.58 |
| 5 | KGP4818545 | 149763305 | G | 6.14E-05 | 2.28 | 1.52 | 3.41 |
| 10 | rs589508 | 117171076 | C | 6.14E-05 | 1.89 | 1.38 | 2.58 |
| 10 | rs10885649 | 116843474 | A | 6.40E-05 | 1.88 | 1.38 | 2.56 |
| 10 | KGP6011600 | 117148182 | A | 6.56E-05 | 1.88 | 1.38 | 2.57 |
| 10 | rs1170884 | 117173622 | T | 6.56E-05 | 1.88 | 1.38 | 2.57 |
| 2 | rs3911084 | 187475357 | G | 6.61E-05 | 1.88 | 1.38 | 2.56 |
| 23 | KGP22809830 | 67338660 | T | 6.65E-05 | 0.45 | 0.31 | 0.67 |
| 9 | KGP11168890 | 8402449 | A | 6.67E-05 | 0.52 | 0.37 | 0.72 |
| 7 | KGP3934027 | 36087216 | C | 6.76E-05 | 4.05 | 2.04 | 8.06 |
| 5 | KGP12517799 | 60180474 | T | 7.01E-05 | 0.21 | 0.10 | 0.45 |
| 17 | KGP1296647 | 12534475 | T | 7.08E-05 | 2.15 | 1.48 | 3.14 |
| 1 | rs28456011 | 1567206 | A | 7.09E-05 | 3.02 | 1.75 | 5.20 |
| 23 | KGP22806909 | 67347391 | G | 7.17E-05 | 0.46 | 0.31 | 0.67 |
| 13 | rs9562406 | 43097559 | A | 7.24E-05 | 2.51 | 1.59 | 3.94 |
| 5 | rs13173226 | 134786259 | C | 7.25E-05 | 1.95 | 1.40 | 2.71 |
| 16 | KGP46650 | 6028700 | T | 7.26E-05 | 0.54 | 0.39 | 0.73 |
| 9 | KGP22833554 | 127410515 | A | 7.31E-05 | 1.86 | 1.37 | 2.52 |
| 19 | rs1423056 | 33780444 | T | 7.33E-05 | 0.53 | 0.38 | 0.72 |
| 2 | rs6758285 | 19029816 | G | 7.50E-05 | 2.04 | 1.43 | 2.90 |
| 5 | KGP12328891 | 134486618 | A | 8.05E-05 | 1.91 | 1.38 | 2.63 |
| 9 | KGP5987914 | 433708 | C | 8.12E-05 | 1.95 | 1.40 | 2.72 |
| 9 | KGP4579224 | 433978 | A | 8.12E-05 | 1.95 | 1.40 | 2.72 |
| 3 | KGP5265567 | 147371594 | T | 8.32E-05 | 2.37 | 1.54 | 3.65 |
| 6 | KGP12324497 | 11292289 | A | 8.36E-05 | 0.52 | 0.37 | 0.72 |
| 1 | rs13375369 | 197007229 | A | 8.42E-05 | 6.81 | 2.62 | 17.73 |
| 1 | KGP15746110 | 197020510 | T | 8.42E-05 | 6.81 | 2.62 | 17.73 |
| 13 | rs1323903 | 67591078 | C | 8.50E-05 | 0.55 | 0.41 | 0.74 |
| 13 | KGP10648476 | 67608580 | T | 8.50E-05 | 0.55 | 0.41 | 0.74 |
| 13 | rs260156 | 67613729 | C | 8.50E-05 | 0.55 | 0.41 | 0.74 |
| 5 | KGP11421325 | 134780494 | A | 8.52E-05 | 2.10 | 1.45 | 3.03 |
| 5 | rs744247 | 134781400 | T | 8.52E-05 | 2.10 | 1.45 | 3.03 |
| 5 | rs12519481 | 134784086 | A | 8.52E-05 | 2.10 | 1.45 | 3.03 |

| CHR | SNP | BP | MINOR ALLELE | ADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 3 | rs1973780 | 59904539 | T | 8.73E-05 | 1.82 | 1.35 | 2.46 |
| 13 | KGP9582727 | 23502042 | T | 8.80E-05 | 0.53 | 0.38 | 0.72 |
| 5 | KGP11882290 | 60219480 | A | 8.85E-05 | 0.22 | 0.10 | 0.47 |
| 10 | rs11595630 | 117277009 | C | 8.91E-05 | 1.99 | 1.41 | 2.80 |
| 10 | rs12781613 | 117302717 | C | 8.91E-05 | 1.99 | 1.41 | 2.80 |
| 3 | KGP1917465 | 47642726 | G | 8.94E-05 | 1.87 | 1.37 | 2.56 |
| 11 | rs1950008 | 8058910 | T | 8.96E-05 | 3.10 | 1.76 | 5.46 |
| 1 | KGP4449990 | 1567206 | A | 9.08E-05 | 2.93 | 1.71 | 5.02 |
| 3 | rs2888432 | 45292244 | G | 9.15E-05 | 0.30 | 0.16 | 0.55 |
| 1 | rs6660568 | 48082491 | G | 9.21E-05 | 2.22 | 1.49 | 3.30 |
| 2 | KGP5053590 | 187467750 | A | 9.34E-05 | 1.87 | 1.37 | 2.56 |
| 9 | KGP2727660 | 73053196 | G | 9.42E-05 | 0.48 | 0.33 | 0.69 |
| 10 | KGP4162559 | 116939803 | C | 9.43E-05 | 1.82 | 1.35 | 2.46 |
| 12 | KGP5357931 | 123519112 | A | 9.44E-05 | 5.02 | 2.23 | 11.29 |
| 9 | rs7040842 | 433347 | C | 9.51E-05 | 1.90 | 1.38 | 2.61 |
| 11 | KGP12037102 | 8031000 | T | 9.52E-05 | 2.31 | 1.52 | 3.52 |
| 16 | KGP6524413 | 80280761 | T | 9.66E-05 | 1.98 | 1.41 | 2.80 |
| 16 | rs16977041 | 23203690 | T | 9.69E-05 | 5.54 | 2.34 | 13.12 |
| 11 | KGP208171 | 8030988 | T | 9.82E-05 | 2.33 | 1.52 | 3.55 |

**Appendix E. All SNPs with p-value < $10^{-4}$ from PTB unadjusted additive logistic regression association analysis for the Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | UNADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 14 | rs7153053 | 105150273 | T | 9.36E-07 | 2.05 | 1.54 | 2.73 |
| 4 | KGP8375539 | 56993480 | G | 5.04E-06 | 0.50 | 0.37 | 0.67 |
| 9 | KGP535977 | 73076516 | A | 7.37E-06 | 0.44 | 0.31 | 0.63 |
| 1 | rs28456011 | 1567206 | A | 8.68E-06 | 3.36 | 1.97 | 5.72 |
| 13 | KGP1703195 | 23467854 | C | 9.27E-06 | 0.45 | 0.32 | 0.64 |
| 13 | rs9510404 | 23462530 | G | 1.13E-05 | 0.45 | 0.32 | 0.65 |
| 1 | KGP4449990 | 1567206 | A | 1.20E-05 | 3.25 | 1.92 | 5.50 |
| 10 | KGP349528 | 131918109 | G | 1.22E-05 | 2.20 | 1.55 | 3.13 |
| 2 | rs651418 | 4495174 | T | 1.24E-05 | 0.50 | 0.37 | 0.68 |
| 12 | KGP2468001 | 123572794 | G | 1.26E-05 | 4.21 | 2.21 | 8.04 |
| 15 | rs11855354 | 78443631 | G | 1.43E-05 | 0.54 | 0.41 | 0.71 |
| 15 | rs2304824 | 78466127 | A | 1.43E-05 | 0.54 | 0.41 | 0.71 |
| 9 | KGP1766722 | 73095674 | T | 1.50E-05 | 0.45 | 0.31 | 0.65 |
| 3 | KGP2829197 | 59893355 | A | 1.51E-05 | 2.09 | 1.50 | 2.91 |
| 3 | rs1864427 | 5626851 | A | 1.52E-05 | 0.51 | 0.38 | 0.69 |
| 9 | rs10780947 | 73165078 | T | 1.57E-05 | 0.51 | 0.38 | 0.69 |
| 22 | KGP11155947 | 19226260 | A | 1.67E-05 | 3.77 | 2.06 | 6.91 |
| 10 | rs1171727 | 131908016 | G | 1.70E-05 | 2.15 | 1.52 | 3.05 |
| 22 | rs712950 | 19204210 | C | 1.76E-05 | 3.76 | 2.05 | 6.88 |
| 4 | KGP4968386 | 27091299 | C | 1.93E-05 | 0.38 | 0.25 | 0.60 |
| 11 | rs4757417 | 16615280 | C | 2.55E-05 | 1.92 | 1.42 | 2.60 |
| 5 | rs2270822 | 60459040 | A | 2.66E-05 | 0.27 | 0.15 | 0.50 |
| 12 | rs12300476 | 31514886 | C | 2.68E-05 | 2.05 | 1.47 | 2.86 |
| 5 | KGP10062212 | 60367377 | C | 2.70E-05 | 0.27 | 0.15 | 0.50 |
| 5 | KGP9219331 | 60457627 | T | 2.70E-05 | 0.27 | 0.15 | 0.50 |
| 19 | rs1423056 | 33780444 | T | 2.77E-05 | 0.51 | 0.38 | 0.70 |
| 16 | KGP6524413 | 80280761 | T | 3.03E-05 | 2.06 | 1.47 | 2.89 |
| 11 | KGP9466771 | 16637583 | T | 3.14E-05 | 1.90 | 1.41 | 2.58 |
| 9 | KGP1996626 | 434222 | A | 3.24E-05 | 2.02 | 1.45 | 2.81 |
| 9 | KGP2727660 | 73053196 | G | 3.27E-05 | 0.46 | 0.32 | 0.67 |
| 18 | rs1529905 | 73077313 | T | 3.40E-05 | 0.33 | 0.20 | 0.56 |
| 22 | rs1633399 | 19183787 | G | 3.72E-05 | 3.65 | 1.97 | 6.75 |
| 22 | KGP3688261 | 19195680 | C | 3.72E-05 | 3.65 | 1.97 | 6.75 |
| 22 | rs712952 | 19197949 | A | 3.72E-05 | 3.65 | 1.97 | 6.75 |

| CHR | SNP | BP | MINOR ALLELE | UNADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 22 | rs807459 | 19223352 | C | 3.72E-05 | 3.65 | 1.97 | 6.75 |
| 22 | KGP4024090 | 19280168 | G | 3.72E-05 | 3.65 | 1.97 | 6.75 |
| 22 | KGP2366391 | 19283478 | G | 3.72E-05 | 3.65 | 1.97 | 6.75 |
| 17 | KGP3405370 | 9719103 | C | 3.74E-05 | 1.79 | 1.36 | 2.37 |
| 23 | rs17215777 | 14045928 | A | 3.75E-05 | 2.67 | 1.67 | 4.25 |
| 10 | KGP5967610 | 117393184 | C | 3.76E-05 | 2.03 | 1.45 | 2.84 |
| 13 | rs9562406 | 43097559 | A | 3.96E-05 | 2.56 | 1.63 | 4.00 |
| 5 | KGP12461110 | 60199201 | C | 4.15E-05 | 0.20 | 0.10 | 0.44 |
| 5 | rs976080 | 60217651 | T | 4.15E-05 | 0.20 | 0.10 | 0.44 |
| 7 | KGP11556052 | 22141292 | A | 4.28E-05 | 1.99 | 1.43 | 2.76 |
| 17 | KGP1296647 | 12534475 | T | 4.29E-05 | 2.16 | 1.50 | 3.13 |
| 2 | KGP11014586 | 187487160 | G | 4.44E-05 | 1.88 | 1.39 | 2.55 |
| 2 | rs2197138 | 187493841 | A | 4.44E-05 | 1.88 | 1.39 | 2.55 |
| 2 | rs7596996 | 187494427 | G | 4.44E-05 | 1.88 | 1.39 | 2.55 |
| 19 | KGP1577871 | 41550317 | T | 4.47E-05 | 1.80 | 1.36 | 2.38 |
| 1 | KGP4299054 | 243602390 | A | 4.53E-05 | 1.89 | 1.39 | 2.56 |
| 4 | KGP4441723 | 92890696 | A | 4.66E-05 | 3.13 | 1.81 | 5.43 |
| 10 | KGP22784786 | 116847333 | T | 4.74E-05 | 1.88 | 1.39 | 2.54 |
| 15 | KGP11866130 | 78470310 | A | 4.87E-05 | 0.56 | 0.42 | 0.74 |
| 9 | KGP11168890 | 8402449 | A | 4.91E-05 | 0.52 | 0.38 | 0.71 |
| 10 | rs1264794 | 116902943 | C | 4.94E-05 | 1.89 | 1.39 | 2.56 |
| 5 | KGP11882290 | 60219480 | A | 4.98E-05 | 0.21 | 0.10 | 0.44 |
| 4 | rs2343115 | 109111726 | C | 5.02E-05 | 0.55 | 0.41 | 0.73 |
| 5 | KGP12517799 | 60180474 | T | 5.20E-05 | 0.21 | 0.10 | 0.45 |
| 2 | KGP22789102 | 187483472 | C | 5.22E-05 | 1.87 | 1.38 | 2.54 |
| 15 | KGP824359 | 39012965 | G | 5.25E-05 | 6.95 | 2.72 | 17.78 |
| 9 | rs7040842 | 433347 | C | 5.27E-05 | 1.92 | 1.40 | 2.63 |
| 4 | KGP3024940 | 109104841 | T | 5.27E-05 | 0.55 | 0.41 | 0.73 |
| 9 | KGP5987914 | 433708 | C | 5.33E-05 | 1.97 | 1.42 | 2.73 |
| 9 | KGP4579224 | 433978 | A | 5.33E-05 | 1.97 | 1.42 | 2.73 |
| 12 | KGP6141129 | 31496481 | A | 5.37E-05 | 0.52 | 0.38 | 0.71 |
| 5 | rs13173226 | 134786259 | C | 5.42E-05 | 1.95 | 1.41 | 2.70 |
| 2 | rs2370693 | 187475104 | G | 5.45E-05 | 1.87 | 1.38 | 2.53 |
| 11 | rs1950008 | 8058910 | T | 5.58E-05 | 3.15 | 1.80 | 5.50 |
| 15 | rs3816253 | 78458485 | G | 5.58E-05 | 1.79 | 1.35 | 2.38 |
| 10 | rs17724010 | 117273552 | C | 5.59E-05 | 1.98 | 1.42 | 2.77 |
| 4 | KGP11607606 | 91912899 | A | 5.75E-05 | 2.22 | 1.50 | 3.27 |
| 2 | rs4233788 | 187462581 | A | 5.78E-05 | 1.87 | 1.38 | 2.53 |
| 1 | rs4839478 | 116295007 | G | 5.92E-05 | 3.93 | 2.02 | 7.66 |

| CHR | SNP | BP | MINOR ALLELE | UNADJ P-value | OR | L95 | U95 |
|---|---|---|---|---|---|---|---|
| 10 | KGP6938912 | 116941272 | G | 6.04E-05 | 1.87 | 1.38 | 2.53 |
| 11 | rs10832869 | 2910621 | C | 6.07E-05 | 1.89 | 1.38 | 2.58 |
| 2 | rs3911084 | 187475357 | G | 6.10E-05 | 1.86 | 1.37 | 2.52 |
| 17 | KGP1449113 | 9720357 | C | 6.10E-05 | 1.76 | 1.33 | 2.31 |
| 17 | rs4791876 | 9722150 | C | 6.10E-05 | 1.76 | 1.33 | 2.31 |
| 4 | KGP1392454 | 109105287 | A | 6.37E-05 | 0.55 | 0.41 | 0.74 |
| 2 | KGP9614636 | 33440045 | G | 6.40E-05 | 1.81 | 1.35 | 2.41 |
| 1 | rs7554304 | 116284997 | G | 6.48E-05 | 3.90 | 2.00 | 7.60 |
| 1 | KGP6147061 | 116291094 | A | 6.48E-05 | 3.90 | 2.00 | 7.60 |
| 16 | KGP46650 | 6028700 | T | 6.59E-05 | 0.54 | 0.40 | 0.73 |
| 13 | rs1028730 | 67527244 | C | 6.81E-05 | 0.51 | 0.36 | 0.71 |
| 10 | KGP10886825 | 117230790 | C | 7.12E-05 | 1.97 | 1.41 | 2.76 |
| 2 | KGP5053590 | 187467750 | A | 7.18E-05 | 1.87 | 1.37 | 2.54 |
| 18 | KGP10756992 | 55899135 | C | 7.20E-05 | 2.27 | 1.52 | 3.40 |
| 20 | rs458332 | 53425217 | A | 7.26E-05 | 1.72 | 1.31 | 2.24 |
| 19 | KGP7273430 | 48524294 | G | 7.46E-05 | 1.77 | 1.33 | 2.34 |
| 15 | KGP10073062 | 78435527 | T | 7.95E-05 | 1.74 | 1.32 | 2.30 |
| 7 | rs7357254 | 135466740 | C | 8.08E-05 | 0.46 | 0.32 | 0.68 |
| 10 | rs1899721 | 116965037 | T | 8.18E-05 | 1.84 | 1.36 | 2.50 |
| 9 | rs1329376 | 431860 | T | 8.48E-05 | 1.93 | 1.39 | 2.67 |
| 17 | KGP3023594 | 9722104 | C | 8.51E-05 | 1.74 | 1.32 | 2.29 |
| 7 | KGP6023091 | 135466550 | A | 8.76E-05 | 0.46 | 0.31 | 0.68 |
| 9 | KGP3932024 | 127536726 | A | 8.76E-05 | 1.79 | 1.34 | 2.40 |
| 14 | rs11160817 | 105168414 | T | 8.80E-05 | 1.73 | 1.32 | 2.28 |
| 16 | rs7204357 | 80252871 | T | 8.90E-05 | 1.97 | 1.40 | 2.76 |
| 16 | KGP383502 | 57112865 | A | 8.91E-05 | 0.33 | 0.19 | 0.57 |
| 13 | rs9571702 | 67532237 | A | 8.97E-05 | 0.50 | 0.35 | 0.71 |
| 19 | rs3745751 | 48519241 | C | 8.99E-05 | 2.56 | 1.60 | 4.09 |
| 8 | rs900213 | 3454880 | T | 9.45E-05 | 1.78 | 1.33 | 2.37 |
| 11 | KGP12037102 | 8031000 | T | 9.52E-05 | 2.27 | 1.50 | 3.43 |
| 1 | rs13375369 | 197007229 | A | 9.66E-05 | 6.56 | 2.55 | 16.88 |
| 1 | KGP15746110 | 197020510 | T | 9.66E-05 | 6.56 | 2.55 | 16.88 |
| 5 | KGP11421325 | 134780494 | A | 9.75E-05 | 2.05 | 1.43 | 2.95 |
| 5 | rs744247 | 134781400 | T | 9.75E-05 | 2.05 | 1.43 | 2.95 |
| 5 | rs12519481 | 134784086 | A | 9.75E-05 | 2.05 | 1.43 | 2.95 |
| 5 | rs11742346 | 45003293 | T | 9.88E-05 | 2.14 | 1.46 | 3.13 |
| 2 | rs6758285 | 19029816 | G | 9.96E-05 | 1.99 | 1.41 | 2.80 |
| 10 | rs10885649 | 116843474 | A | 9.99E-05 | 1.83 | 1.35 | 2.48 |

**Appendix F. All SNPs with p-value < $10^{-4}$ from PTB genotypic logistic regression adjusted for body mass index (BMI), gravidity and smoking status association analysis for the Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome

| CHR | SNP | BP | Minor Allele | Adjusted Genotypic P-value |
|---|---|---|---|---|
| 14 | rs7153053 | 105150273 | T | 7.40E-07 |
| 20 | rs3865536 | 2362961 | A | 3.90E-06 |
| 8 | rs900213 | 3454880 | T | 8.44E-06 |
| 16 | rs6497441 | 9829848 | T | 9.93E-06 |
| 3 | rs6800348 | 67010311 | A | 1.30E-05 |
| 10 | KGP349528 | 131918109 | G | 1.49E-05 |
| 18 | KGP10756992 | 55899135 | C | 1.71E-05 |
| 15 | KGP2297008 | 37151356 | A | 2.20E-05 |
| 3 | KGP7628141 | 15701184 | G | 2.37E-05 |
| 11 | rs12364666 | 81395174 | A | 2.41E-05 |
| 2 | KGP9077006 | 50268228 | T | 2.49E-05 |
| 2 | rs651418 | 4495174 | T | 2.84E-05 |
| 2 | rs1819972 | 50268228 | T | 2.90E-05 |
| 4 | KGP8375539 | 56993480 | G | 3.14E-05 |
| 22 | KGP10885912 | 46017699 | G | 3.73E-05 |
| 5 | rs10072136 | 174470775 | T | 3.95E-05 |
| 5 | KGP11777967 | 174472781 | C | 4.17E-05 |
| 5 | rs10053511 | 174472998 | A | 4.17E-05 |
| 10 | KGP22784786 | 116847333 | T | 4.20E-05 |
| 10 | rs1264794 | 116902943 | C | 4.29E-05 |
| 15 | rs8025854 | 37165249 | G | 4.33E-05 |
| 1 | KGP4299054 | 243602390 | A | 4.59E-05 |
| 3 | KGP9937064 | 15699347 | G | 4.86E-05 |
| 11 | rs4757417 | 16615280 | C | 5.30E-05 |
| 5 | KGP5909768 | 174473277 | A | 5.48E-05 |
| 15 | rs11855354 | 78443631 | G | 5.61E-05 |
| 15 | rs2304824 | 78466127 | A | 5.61E-05 |
| 12 | rs12300476 | 31514886 | C | 5.70E-05 |
| 11 | KGP9466771 | 16637583 | T | 6.12E-05 |
| 15 | KGP10073062 | 78435527 | T | 6.13E-05 |
| 2 | rs716335 | 169402127 | T | 6.17E-05 |
| 10 | rs1171727 | 131908016 | G | 6.29E-05 |
| 10 | KGP6938912 | 116941272 | G | 6.82E-05 |

| CHR | SNP | BP | Minor Allele | Adjusted Genotypic P-value |
|---|---|---|---|---|
| 22 | KGP9592633 | 46016136 | G | 6.99E-05 |
| 15 | rs1450418 | 37166878 | C | 7.04E-05 |
| 20 | rs458332 | 53425217 | A | 7.35E-05 |
| 11 | rs10832869 | 2910621 | C | 7.47E-05 |
| 4 | KGP1592798 | 127266480 | G | 7.51E-05 |
| 14 | KGP6461089 | 22103779 | C | 7.57E-05 |
| 10 | rs10885649 | 116843474 | A | 7.57E-05 |
| 1 | rs1129590 | 161953015 | T | 7.61E-05 |
| 14 | rs4982475 | 22101945 | A | 7.70E-05 |
| 10 | KGP5967610 | 117393184 | C | 7.81E-05 |
| 9 | KGP535977 | 73076516 | A | 7.86E-05 |
| 19 | rs16982100 | 48511888 | G | 7.90E-05 |
| 10 | rs17093770 | 92830260 | A | 7.96E-05 |
| 3 | KGP2829197 | 59893355 | A | 8.33E-05 |
| 10 | rs11253689 | 6668830 | G | 8.43E-05 |
| 10 | rs7915166 | 6409502 | C | 8.74E-05 |
| 4 | rs17384008 | 122580875 | A | 9.15E-05 |
| 14 | KGP3600601 | 92582472 | A | 9.32E-05 |
| 14 | rs3814833 | 92583579 | C | 9.32E-05 |
| 14 | rs3818263 | 92588002 | C | 9.32E-05 |
| 11 | KGP208171 | 8030988 | T | 9.44E-05 |
| 6 | KGP1255072 | 130351313 | T | 9.44E-05 |
| 1 | rs12726227 | 161954448 | C | 9.74E-05 |
| 5 | KGP2263756 | 140570974 | G | 9.77E-05 |
| 5 | rs4912742 | 140571526 | C | 9.77E-05 |
| 10 | rs17724010 | 117273552 | C | 9.82E-05 |
| 11 | KGP12037102 | 8031000 | T | 9.85E-05 |
| 4 | KGP4142511 | 16136484 | G | 9.90E-05 |

**Appendix G. All SNPs with p-value < $10^{-4}$ from PTB genotypic logistic regression unadjusted association analysis for the Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome

| CHR | SNP | BP | Minor allele | Unadjusted Genotypic P-value |
|---|---|---|---|---|
| 14 | rs7153053 | 105150273 | T | 7.88E-07 |
| 20 | rs3865536 | 2362961 | A | 2.25E-06 |
| 16 | rs6497441 | 9829848 | T | 7.56E-06 |
| 10 | KGP349528 | 131918109 | G | 9.40E-06 |
| 2 | KGP9077006 | 50268228 | T | 1.21E-05 |
| 2 | rs1819972 | 50268228 | T | 1.41E-05 |
| 12 | rs12300476 | 31514886 | C | 2.04E-05 |
| 20 | rs458332 | 53425217 | A | 2.21E-05 |
| 15 | rs11855354 | 78443631 | G | 2.59E-05 |
| 15 | rs2304824 | 78466127 | A | 2.59E-05 |
| 8 | rs900213 | 3454880 | T | 2.60E-05 |
| 4 | KGP8375539 | 56993480 | G | 2.64E-05 |
| 11 | rs12364666 | 81395174 | A | 2.67E-05 |
| 4 | KGP1592798 | 127266480 | G | 2.83E-05 |
| 9 | KGP535977 | 73076516 | A | 2.95E-05 |
| 5 | rs10072136 | 174470775 | T | 3.35E-05 |
| 10 | rs1264794 | 116902943 | C | 3.79E-05 |
| 10 | KGP22784786 | 116847333 | T | 3.81E-05 |
| 22 | KGP10885912 | 46017699 | G | 3.83E-05 |
| 1 | KGP4299054 | 243602390 | A | 3.96E-05 |
| 10 | rs1171727 | 131908016 | G | 4.01E-05 |
| 3 | rs6800348 | 67010311 | A | 4.03E-05 |
| 2 | rs4281904 | 107151134 | G | 4.09E-05 |
| 23 | rs9699237 | 123461240 | C | 4.11E-05 |
| 15 | KGP2297008 | 37151356 | A | 4.13E-05 |
| 9 | KGP720068 | 16012195 | T | 4.46E-05 |
| 11 | rs10832869 | 2910621 | C | 4.61E-05 |
| 15 | KGP10073062 | 78435527 | T | 4.83E-05 |
| 9 | KGP1766722 | 73095674 | T | 5.43E-05 |
| 9 | rs10962251 | 16012225 | A | 5.56E-05 |
| 8 | rs7818882 | 88803575 | G | 5.92E-05 |
| 14 | KGP6461089 | 22103779 | C | 6.22E-05 |
| 14 | rs4982475 | 22101945 | A | 6.36E-05 |
| 6 | rs3828886 | 31440552 | G | 6.41E-05 |

| CHR | SNP | BP | Minor allele | Unadjusted Genotypic P-value |
|---|---|---|---|---|
| 10 | rs10885649 | 116843474 | A | 6.42E-05 |
| 2 | rs651418 | 4495174 | T | 6.69E-05 |
| 18 | rs7228791 | 71341426 | A | 6.77E-05 |
| 22 | KGP9592633 | 46016136 | G | 7.04E-05 |
| 2 | rs6729848 | 107154331 | G | 7.29E-05 |
| 14 | KGP3600601 | 92582472 | A | 7.34E-05 |
| 14 | rs3814833 | 92583579 | C | 7.34E-05 |
| 14 | rs3818263 | 92588002 | C | 7.34E-05 |
| 10 | KGP6938912 | 116941272 | G | 7.49E-05 |
| 15 | rs8025854 | 37165249 | G | 7.57E-05 |
| 5 | KGP11777967 | 174472781 | C | 7.71E-05 |
| 5 | rs10053511 | 174472998 | A | 7.71E-05 |
| 10 | rs17093770 | 92830260 | A | 7.78E-05 |
| 18 | rs1529905 | 73077313 | T | 8.13E-05 |
| 14 | KGP9989748 | 92584025 | T | 8.25E-05 |
| 3 | KGP7628141 | 15701184 | G | 8.66E-05 |
| 23 | KGP22772552 | 78565662 | G | 9.08E-05 |
| 1 | KGP12448491 | 243558785 | G | 9.08E-05 |
| 3 | KGP2829197 | 59893355 | A | 9.09E-05 |
| 18 | KGP10576230 | 71347120 | T | 9.09E-05 |
| 16 | KGP9902020 | 9866186 | A | 9.48E-05 |
| 6 | KGP1999769 | 88421485 | C | 9.73E-05 |
| 2 | rs716335 | 169402127 | T | 9.95E-05 |

**Appendix H. A complete list of all SNPs with p-value < $10^{-4}$ from the additive linear regression for Box-Cox transformed gestational age adjusted for BMI, gravidity and smoking status for the Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | ADJ Box Cox P-value |
|---|---|---|---|---|
| 14 | rs7153053 | 105150273 | T | 6.28E-08 |
| 18 | KGP10756992 | 55899135 | C | 1.20E-06 |
| 4 | KGP8375539 | 56993480 | G | 1.80E-06 |
| 23 | rs17215777 | 14045928 | A | 3.35E-06 |
| 20 | rs6079395 | 14327899 | G | 3.85E-06 |
| 18 | KGP1841629 | 24195715 | A | 4.32E-06 |
| 22 | KGP1327745 | 27571605 | T | 4.35E-06 |
| 19 | rs3745751 | 48519241 | C | 4.73E-06 |
| 18 | rs1529905 | 73077313 | T | 5.60E-06 |
| 18 | KGP12553043 | 24194562 | G | 6.02E-06 |
| 1 | KGP2324307 | 161955804 | C | 6.39E-06 |
| 4 | KGP10363154 | 92027017 | A | 7.91E-06 |
| 14 | rs11160817 | 105168414 | T | 8.02E-06 |
| 1 | rs12726227 | 161954448 | C | 9.38E-06 |
| 1 | rs1129590 | 161953015 | T | 1.04E-05 |
| 12 | KGP2468001 | 123572794 | G | 1.08E-05 |
| 19 | rs16982100 | 48511888 | G | 1.10E-05 |
| 13 | KGP1703195 | 23467854 | C | 1.21E-05 |
| 10 | KGP1429545 | 116400883 | C | 1.32E-05 |
| 5 | rs11242497 | 102926108 | A | 1.35E-05 |
| 23 | KGP22783547 | 66980923 | T | 1.41E-05 |
| 23 | KGP22820383 | 66339984 | C | 1.42E-05 |
| 13 | rs9510404 | 23462530 | G | 1.50E-05 |
| 13 | KGP2028096 | 100549885 | G | 1.57E-05 |
| 3 | rs1864427 | 5626851 | A | 1.59E-05 |
| 6 | KGP17157059 | 118340886 | T | 1.66E-05 |
| 6 | KGP2973353 | 118378899 | A | 1.66E-05 |
| 6 | KGP8877966 | 118384078 | G | 1.66E-05 |
| 6 | KGP9462937 | 118397112 | A | 1.66E-05 |
| 1 | rs4839478 | 116295007 | G | 1.88E-05 |
| 1 | rs7554304 | 116284997 | G | 2.00E-05 |
| 1 | KGP6147061 | 116291094 | A | 2.00E-05 |
| 23 | rs11095194 | 30975402 | C | 2.15E-05 |

| CHR | SNP | BP | MINOR ALLELE | ADJ Box Cox P-value |
|---|---|---|---|---|
| 20 | rs6079391 | 14306953 | G | 2.21E-05 |
| 9 | KGP11955310 | 36213849 | A | 2.25E-05 |
| 10 | rs10903443 | 1454864 | T | 2.27E-05 |
| 16 | rs7187030 | 65922323 | A | 2.30E-05 |
| 23 | KGP22820496 | 30973949 | C | 2.30E-05 |
| 19 | KGP11012186 | 48517676 | T | 2.39E-05 |
| 19 | KGP3429713 | 48522480 | A | 2.39E-05 |
| 19 | KGP12010912 | 48522553 | C | 2.39E-05 |
| 19 | KGP1072697 | 48522746 | A | 2.39E-05 |
| 6 | rs11756591 | 14088091 | G | 2.47E-05 |
| 23 | KGP22789476 | 66783343 | T | 2.48E-05 |
| 20 | rs4640454 | 14337539 | G | 2.50E-05 |
| 14 | KGP2938981 | 24910076 | T | 2.50E-05 |
| 18 | KGP4741739 | 24181001 | A | 2.55E-05 |
| 11 | rs4757417 | 16615280 | C | 2.59E-05 |
| 14 | KGP4248600 | 85507035 | G | 2.71E-05 |
| 19 | KGP1220612 | 48517387 | A | 2.80E-05 |
| 9 | KGP2866004 | 78159780 | A | 3.02E-05 |
| 11 | KGP9466771 | 16637583 | T | 3.19E-05 |
| 23 | KGP22802116 | 30958392 | T | 3.22E-05 |
| 16 | rs17494421 | 64819806 | T | 3.35E-05 |
| 5 | rs1559050 | 134470030 | T | 3.38E-05 |
| 6 | rs2013807 | 14088302 | A | 3.41E-05 |
| 6 | rs12660382 | 31443323 | T | 3.43E-05 |
| 7 | KGP3934027 | 36087216 | C | 3.54E-05 |
| 5 | KGP10407767 | 134493981 | G | 3.56E-05 |
| 1 | rs6660568 | 48082491 | G | 3.57E-05 |
| 5 | KGP12328891 | 134486618 | A | 3.59E-05 |
| 12 | KGP9155397 | 86114882 | T | 3.66E-05 |
| 5 | rs6874368 | 102921968 | C | 3.67E-05 |
| 3 | KGP2829197 | 59893355 | A | 3.79E-05 |
| 16 | KGP6524413 | 80280761 | T | 3.88E-05 |
| 7 | KGP6023091 | 135466550 | A | 3.94E-05 |
| 14 | KGP1568287 | 85884768 | A | 3.99E-05 |
| 14 | KGP22760935 | 85892075 | C | 4.18E-05 |
| 20 | rs6074716 | 14293522 | T | 4.18E-05 |
| 4 | KGP4968386 | 27091299 | C | 4.20E-05 |
| 2 | rs4270372 | 71717062 | C | 4.20E-05 |
| 12 | KGP5098575 | 71380074 | C | 4.37E-05 |

| CHR | SNP | BP | MINOR ALLELE | ADJ Box Cox P-value |
|---|---|---|---|---|
| 10 | KGP22784786 | 116847333 | T | 4.50E-05 |
| 4 | KGP12486975 | 91937972 | G | 4.52E-05 |
| 4 | KGP9305680 | 92113737 | C | 4.52E-05 |
| 3 | KGP11570798 | 41493591 | A | 4.60E-05 |
| 2 | KGP9171376 | 27611454 | T | 4.72E-05 |
| 19 | KGP12296823 | 35670808 | A | 4.81E-05 |
| 23 | rs7885415 | 30975885 | A | 4.84E-05 |
| 23 | rs12838662 | 30976242 | G | 4.95E-05 |
| 23 | KGP22742897 | 30977301 | T | 4.95E-05 |
| 23 | KGP22729818 | 30978353 | C | 4.95E-05 |
| 1 | KGP480502 | 169623594 | T | 5.05E-05 |
| 9 | rs4879961 | 36231483 | T | 5.10E-05 |
| 1 | KGP1517463 | 94569447 | G | 5.12E-05 |
| 10 | KGP10902145 | 116567094 | C | 5.13E-05 |
| 9 | KGP398693 | 36298820 | T | 5.35E-05 |
| 1 | rs13375369 | 197007229 | A | 5.39E-05 |
| 1 | KGP15746110 | 197020510 | T | 5.39E-05 |
| 23 | KGP22754915 | 66198870 | G | 5.67E-05 |
| 1 | rs17524161 | 169622927 | C | 5.78E-05 |
| 10 | KGP349528 | 131918109 | G | 6.02E-05 |
| 10 | rs1264794 | 116902943 | C | 6.29E-05 |
| 10 | rs4749049 | 25864482 | A | 6.33E-05 |
| 14 | rs11621863 | 85918380 | A | 6.36E-05 |
| 20 | rs3747927 | 14230505 | C | 6.38E-05 |
| 13 | rs1028730 | 67527244 | C | 6.40E-05 |
| 16 | KGP10540547 | 6024123 | G | 6.49E-05 |
| 2 | KGP437725 | 213991770 | C | 6.49E-05 |
| 11 | rs7128152 | 40859924 | C | 6.51E-05 |
| 2 | KGP11659142 | 38925610 | G | 6.96E-05 |
| 10 | rs11259736 | 15595361 | A | 7.10E-05 |
| 19 | KGP12256054 | 35673896 | T | 7.10E-05 |
| 22 | KGP11155947 | 19226260 | A | 7.11E-05 |
| 10 | KGP6938912 | 116941272 | G | 7.14E-05 |
| 22 | rs712950 | 19204210 | C | 7.15E-05 |
| 10 | KGP1874043 | 134540672 | A | 7.21E-05 |
| 10 | rs1171727 | 131908016 | G | 7.24E-05 |
| 14 | rs4900406 | 99056323 | A | 7.24E-05 |
| 5 | rs2071239 | 149755421 | G | 7.31E-05 |
| 1 | rs6691548 | 161967287 | C | 7.41E-05 |

| CHR | SNP | BP | MINOR ALLELE | ADJ Box Cox P-value |
|---|---|---|---|---|
| 4 | KGP12445244 | 55094467 | A | 7.44E-05 |
| 16 | KGP46650 | 6028700 | T | 7.55E-05 |
| 13 | rs9571702 | 67532237 | A | 7.61E-05 |
| 10 | rs11196905 | 116528294 | C | 7.62E-05 |
| 2 | KGP8418824 | 115621917 | A | 7.64E-05 |
| 4 | KGP11607606 | 91912899 | A | 7.79E-05 |
| 19 | rs9807905 | 35685456 | C | 7.80E-05 |
| 17 | rs2343202 | 70222600 | A | 7.89E-05 |
| 11 | rs12804793 | 122643283 | T | 8.11E-05 |
| 19 | KGP4180994 | 35658188 | A | 8.21E-05 |
| 19 | rs12110 | 35660508 | G | 8.21E-05 |
| 10 | KGP10799699 | 25867396 | T | 8.32E-05 |
| 16 | KGP383502 | 57112865 | A | 8.44E-05 |
| 17 | KGP9084790 | 21296416 | A | 8.45E-05 |
| 13 | KGP7715737 | 67553342 | T | 8.60E-05 |
| 13 | KGP9920102 | 67545692 | G | 8.62E-05 |
| 1 | rs3842895 | 180596196 | C | 8.70E-05 |
| 2 | rs4536672 | 71714223 | A | 8.77E-05 |
| 5 | rs7713638 | 149759096 | C | 8.82E-05 |
| 1 | rs12756866 | 161968297 | T | 8.83E-05 |
| 5 | rs11745068 | 134475494 | T | 8.84E-05 |
| 10 | rs10885649 | 116843474 | A | 8.91E-05 |
| 15 | rs8034908 | 98124026 | A | 8.98E-05 |
| 5 | rs2270822 | 60459040 | A | 9.08E-05 |
| 1 | KGP5269992 | 169580510 | T | 9.14E-05 |
| 1 | KGP10559043 | 169580717 | C | 9.14E-05 |
| 16 | KGP5557476 | 65932588 | A | 9.25E-05 |
| 2 | rs3739028 | 135907846 | G | 9.39E-05 |
| 2 | KGP3867438 | 135988235 | T | 9.39E-05 |
| 15 | rs12591876 | 98115684 | C | 9.45E-05 |
| 20 | rs6071381 | 59434798 | G | 9.48E-05 |
| 5 | KGP1480373 | 174609264 | C | 9.56E-05 |
| 5 | KGP4818545 | 149763305 | G | 9.56E-05 |
| 10 | KGP7663783 | 9028946 | T | 9.61E-05 |
| 19 | rs11084724 | 33781008 | T | 9.65E-05 |
| 9 | rs1043313 | 36214971 | C | 9.72E-05 |
| 4 | KGP7789951 | 38370504 | A | 9.72E-05 |
| 10 | rs1899721 | 116965037 | T | 9.73E-05 |
| 1 | KGP11853718 | 169643176 | A | 9.79E-05 |

| CHR | SNP | BP | MINOR ALLELE | ADJ Box Cox P-value |
|---|---|---|---|---|
| 4 | KGP4062013 | 186578529 | A | 9.81E-05 |
| 17 | KGP1290608 | 72047939 | T | 9.90E-05 |
| 3 | rs1500005 | 46102620 | A | 9.97E-05 |

**Appendix I. A complete list of all SNPs with p-value ≤ $10^{-4}$ from the additive birth weight z-score linear regression adjusted for BMI, gravidity and smoking status for Helsinki infants genotyped on the Illumina Omni2.5 BeadChip.** Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 4 | KGP7935680 | 175820536 | T | 0.296 | 0.183 | 0.409 | 4.25E-07 |
| 4 | rs6553849 | 175838816 | G | 0.296 | 0.183 | 0.409 | 4.25E-07 |
| 4 | KGP11462362 | 175848583 | C | 0.296 | 0.183 | 0.409 | 4.25E-07 |
| 4 | rs12512467 | 175842230 | G | 0.294 | 0.180 | 0.408 | 6.04E-07 |
| 4 | rs4388065 | 175831859 | C | 0.289 | 0.175 | 0.402 | 8.85E-07 |
| 11 | KGP5677965 | 128764571 | A | 0.486 | 0.293 | 0.678 | 1.09E-06 |
| 17 | KGP10605353 | 47459646 | C | -0.367 | -0.514 | -0.221 | 1.24E-06 |
| 2 | KGP10390109 | 155375235 | G | 0.446 | 0.264 | 0.628 | 2.13E-06 |
| 11 | KGP12295813 | 128769010 | T | 0.472 | 0.278 | 0.666 | 2.40E-06 |
| 11 | KGP10855022 | 128769387 | A | 0.472 | 0.278 | 0.666 | 2.40E-06 |
| 18 | KGP216041 | 44526798 | A | 0.276 | 0.163 | 0.390 | 2.52E-06 |
| 7 | KGP7565682 | 119519894 | T | 2.412 | 1.418 | 3.405 | 2.63E-06 |
| 17 | rs6917 | 47481543 | T | -0.329 | -0.465 | -0.193 | 2.68E-06 |
| 18 | rs2010834 | 44560875 | T | 0.273 | 0.160 | 0.386 | 2.97E-06 |
| 2 | KGP411313 | 155370926 | A | 0.444 | 0.259 | 0.629 | 3.29E-06 |
| 2 | rs16837788 | 155371557 | C | 0.433 | 0.252 | 0.615 | 3.82E-06 |
| 5 | rs967489 | 107325759 | T | -0.277 | -0.393 | -0.161 | 3.93E-06 |
| 13 | KGP5600928 | 104714446 | C | -0.362 | -0.515 | -0.209 | 4.45E-06 |
| 13 | KGP6718572 | 104716555 | C | -0.362 | -0.515 | -0.209 | 4.45E-06 |
| 14 | KGP4386312 | 75701221 | T | 0.374 | 0.216 | 0.532 | 4.71E-06 |
| 18 | rs7244778 | 44541915 | T | 0.266 | 0.153 | 0.379 | 4.92E-06 |
| 17 | KGP12240597 | 47244412 | T | -0.287 | -0.410 | -0.164 | 6.09E-06 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 17 | KGP6975802 | 47293848 | T | -0.280 | -0.402 | -0.159 | 7.37E-06 |
| 17 | rs2233362 | 47287067 | G | -0.282 | -0.404 | -0.160 | 7.43E-06 |
| 18 | KGP12164454 | 44537102 | T | 0.261 | 0.148 | 0.374 | 7.64E-06 |
| 13 | rs7998621 | 104647290 | T | -0.348 | -0.499 | -0.197 | 7.69E-06 |
| 3 | KGP9945238 | 118281147 | A | -2.892 | -4.149 | -1.636 | 8.13E-06 |
| 3 | rs12486534 | 101816344 | C | -0.254 | -0.366 | -0.143 | 1.02E-05 |
| 4 | KGP5822115 | 30338739 | T | 0.762 | 0.427 | 1.096 | 1.03E-05 |
| 5 | KGP1886224 | 28998113 | G | 0.411 | 0.230 | 0.592 | 1.08E-05 |
| 13 | KGP3791399 | 104652517 | A | -0.343 | -0.494 | -0.192 | 1.12E-05 |
| 14 | KGP10544388 | 75705656 | A | 0.377 | 0.211 | 0.544 | 1.13E-05 |
| 14 | KGP6657936 | 75705863 | T | 0.376 | 0.210 | 0.542 | 1.19E-05 |
| 1 | KGP10162962 | 201260642 | A | 0.353 | 0.197 | 0.510 | 1.25E-05 |
| 18 | rs509647 | 44383675 | C | 0.257 | 0.143 | 0.371 | 1.28E-05 |
| 18 | rs642897 | 44396917 | T | 0.257 | 0.143 | 0.371 | 1.28E-05 |
| 16 | KGP11936906 | 65193896 | C | -1.953 | -2.825 | -1.082 | 1.39E-05 |
| 9 | KGP643692 | 93597588 | A | 0.404 | 0.223 | 0.584 | 1.46E-05 |
| 3 | rs12488237 | 56114861 | C | -0.674 | -0.976 | -0.371 | 1.57E-05 |
| 17 | KGP11999849 | 38133922 | G | 0.306 | 0.169 | 0.444 | 1.57E-05 |
| 17 | KGP4921252 | 38133545 | C | 0.307 | 0.169 | 0.446 | 1.69E-05 |
| 13 | KGP16741843 | 104646537 | A | -0.340 | -0.494 | -0.186 | 1.87E-05 |
| 12 | KGP2332927 | 7596747 | T | -2.700 | -3.928 | -1.473 | 1.98E-05 |
| 12 | rs16906797 | 30955572 | C | -0.586 | -0.853 | -0.320 | 1.99E-05 |
| 3 | rs6774331 | 101823824 | G | -0.247 | -0.359 | -0.134 | 2.02E-05 |
| 5 | rs1436969 | 30689900 | G | 0.245 | 0.133 | 0.356 | 2.08E-05 |
| 3 | rs6804221 | 56207448 | C | -0.618 | -0.899 | -0.336 | 2.09E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 3 | rs6764295 | 56207634 | A | -0.618 | -0.899 | -0.336 | 2.09E-05 |
| 3 | rs4974120 | 56220993 | C | -0.618 | -0.899 | -0.336 | 2.09E-05 |
| 3 | KGP3087267 | 56230867 | T | -0.618 | -0.899 | -0.336 | 2.09E-05 |
| 3 | KGP9981812 | 56288321 | T | -0.618 | -0.899 | -0.336 | 2.09E-05 |
| 3 | KGP10188926 | 56303321 | C | -0.618 | -0.899 | -0.336 | 2.09E-05 |
| 15 | rs12916632 | 55118942 | T | -0.290 | -0.423 | -0.158 | 2.14E-05 |
| 18 | rs9304337 | 44493094 | G | 0.249 | 0.135 | 0.362 | 2.26E-05 |
| 18 | KGP2783710 | 44488514 | A | 0.247 | 0.134 | 0.360 | 2.29E-05 |
| 11 | rs7110240 | 116343848 | C | -0.258 | -0.377 | -0.139 | 2.48E-05 |
| 3 | rs4974196 | 56357553 | A | -0.599 | -0.875 | -0.323 | 2.51E-05 |
| 18 | rs10853545 | 44475220 | A | 0.249 | 0.134 | 0.364 | 2.69E-05 |
| 11 | KGP12477693 | 62434740 | A | -2.648 | -3.872 | -1.424 | 2.69E-05 |
| 17 | rs2177309 | 47262672 | C | -0.266 | -0.389 | -0.143 | 2.73E-05 |
| 22 | KGP11418130 | 39551628 | T | -0.243 | -0.355 | -0.131 | 2.74E-05 |
| 1 | KGP12121704 | 237836880 | C | -0.245 | -0.358 | -0.131 | 2.81E-05 |
| 3 | rs12053903 | 38593393 | C | 0.246 | 0.132 | 0.361 | 2.82E-05 |
| 3 | KGP5143211 | 6363787 | T | 0.943 | 0.505 | 1.381 | 2.95E-05 |
| 6 | rs7768059 | 103775306 | G | -0.532 | -0.779 | -0.284 | 3.01E-05 |
| 6 | rs9404359 | 103788127 | T | -0.532 | -0.779 | -0.284 | 3.01E-05 |
| 20 | KGP19186526 | 50555325 | C | 0.873 | 0.467 | 1.279 | 3.06E-05 |
| 1 | rs11811484 | 53775436 | G | 0.974 | 0.520 | 1.428 | 3.12E-05 |
| 20 | KGP19243097 | 50578768 | G | 0.871 | 0.465 | 1.277 | 3.17E-05 |
| 4 | rs2611209 | 166578236 | A | -0.240 | -0.352 | -0.128 | 3.23E-05 |
| 20 | KGP4604730 | 50573234 | G | 0.892 | 0.475 | 1.310 | 3.34E-05 |
| 3 | KGP2675136 | 101813131 | C | -0.239 | -0.351 | -0.127 | 3.35E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 2 | KGP6370333 | 1982376 | C | 0.269 | 0.143 | 0.394 | 3.47E-05 |
| 10 | KGP10748275 | 131699951 | A | 0.295 | 0.157 | 0.434 | 3.47E-05 |
| 13 | rs2806768 | 104724443 | T | -0.316 | -0.464 | -0.168 | 3.54E-05 |
| 13 | KGP12344613 | 104724623 | A | -0.316 | -0.464 | -0.168 | 3.54E-05 |
| 7 | KGP6839477 | 84338112 | A | -0.379 | -0.558 | -0.201 | 3.77E-05 |
| 8 | KGP20187005 | 4100582 | A | -1.667 | -2.453 | -0.882 | 3.79E-05 |
| 9 | rs10821325 | 97018657 | T | -0.266 | -0.391 | -0.141 | 3.80E-05 |
| 2 | rs7561518 | 1983619 | C | 0.268 | 0.142 | 0.395 | 3.86E-05 |
| 6 | KGP3656064 | 107650309 | G | -0.308 | -0.454 | -0.163 | 3.97E-05 |
| 6 | rs2430464 | 107639414 | T | -0.308 | -0.454 | -0.162 | 4.11E-05 |
| 3 | KGP3425928 | 127913954 | A | 0.576 | 0.303 | 0.848 | 4.15E-05 |
| 7 | KGP2678353 | 84594661 | A | -0.369 | -0.544 | -0.194 | 4.18E-05 |
| 1 | KGP9894193 | 53902833 | A | 0.900 | 0.474 | 1.327 | 4.19E-05 |
| 2 | KGP11300174 | 155370862 | A | 0.431 | 0.227 | 0.635 | 4.19E-05 |
| 8 | rs7017102 | 145118648 | G | 0.434 | 0.228 | 0.640 | 4.24E-05 |
| 15 | KGP9865938 | 65809642 | A | 1.495 | 0.785 | 2.204 | 4.29E-05 |
| 15 | KGP2935410 | 65871017 | T | 1.495 | 0.785 | 2.204 | 4.29E-05 |
| 1 | KGP5615287 | 10807369 | A | 0.287 | 0.151 | 0.423 | 4.32E-05 |
| 13 | KGP3146424 | 39647492 | T | -0.476 | -0.702 | -0.250 | 4.32E-05 |
| 13 | KGP11565338 | 104731202 | G | -0.307 | -0.453 | -0.161 | 4.35E-05 |
| 13 | KGP16832172 | 43520205 | C | 1.112 | 0.584 | 1.641 | 4.38E-05 |
| 1 | KGP7415512 | 53917385 | A | 0.899 | 0.472 | 1.326 | 4.42E-05 |
| 1 | KGP22758040 | 205390079 | A | 2.576 | 1.352 | 3.801 | 4.43E-05 |
| 2 | rs1896831 | 155381331 | A | 0.429 | 0.225 | 0.633 | 4.48E-05 |
| 9 | KGP3331281 | 97021624 | T | -0.262 | -0.386 | -0.137 | 4.49E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 13 | KGP1596046 | 43531085 | A | 1.106 | 0.580 | 1.633 | 4.52E-05 |
| 3 | KGP9532024 | 38599886 | C | 0.240 | 0.126 | 0.355 | 4.55E-05 |
| 8 | KGP2375841 | 81922820 | T | -2.579 | -3.808 | -1.351 | 4.56E-05 |
| 6 | rs495335 | 49701523 | A | -0.251 | -0.371 | -0.131 | 4.63E-05 |
| 6 | rs509699 | 49704546 | A | -0.251 | -0.371 | -0.131 | 4.63E-05 |
| 15 | KGP3967604 | 55134440 | G | -0.342 | -0.505 | -0.179 | 4.75E-05 |
| 5 | KGP2332271 | 30802243 | C | 0.269 | 0.140 | 0.397 | 4.77E-05 |
| 7 | rs11532826 | 53967836 | A | -0.244 | -0.360 | -0.127 | 4.77E-05 |
| 8 | KGP6731082 | 20985530 | A | -1.813 | -2.680 | -0.946 | 4.90E-05 |
| 1 | rs2268147 | 201252866 | T | 0.371 | 0.194 | 0.548 | 4.91E-05 |
| 17 | rs9972882 | 37807698 | A | -0.273 | -0.404 | -0.143 | 4.94E-05 |
| 20 | KGP22832251 | 38839231 | G | -0.305 | -0.451 | -0.159 | 5.01E-05 |
| 5 | KGP4721260 | 143219584 | C | -0.280 | -0.415 | -0.146 | 5.13E-05 |
| 17 | rs12948798 | 47275360 | A | -0.255 | -0.377 | -0.133 | 5.17E-05 |
| 4 | rs17028431 | 152949686 | A | 0.336 | 0.175 | 0.497 | 5.21E-05 |
| 8 | rs7004867 | 145114844 | T | 0.430 | 0.224 | 0.637 | 5.27E-05 |
| 10 | KGP4567182 | 99497671 | T | 0.380 | 0.197 | 0.562 | 5.28E-05 |
| 10 | rs17108375 | 99498123 | G | 0.380 | 0.197 | 0.562 | 5.28E-05 |
| 1 | KGP1786634 | 53847964 | T | 0.890 | 0.463 | 1.317 | 5.28E-05 |
| 4 | KGP8409301 | 152966262 | T | 0.336 | 0.175 | 0.498 | 5.29E-05 |
| 13 | rs6491817 | 104736365 | G | -0.305 | -0.452 | -0.159 | 5.34E-05 |
| 4 | rs12511908 | 152957388 | A | 0.334 | 0.174 | 0.495 | 5.34E-05 |
| 4 | KGP8058190 | 152969838 | C | 0.334 | 0.174 | 0.495 | 5.34E-05 |
| 4 | KGP10209584 | 152971850 | A | 0.334 | 0.174 | 0.495 | 5.34E-05 |
| 23 | KGP22766765 | 77086257 | T | 1.282 | 0.665 | 1.898 | 5.38E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 1 | rs205478 | 10811672 | C | 0.281 | 0.146 | 0.416 | 5.51E-05 |
| 3 | KGP7144741 | 43385905 | T | -2.083 | -3.086 | -1.080 | 5.54E-05 |
| 12 | KGP6908040 | 128021163 | T | -0.235 | -0.349 | -0.122 | 5.54E-05 |
| 14 | KGP10753036 | 57781048 | C | 1.149 | 0.595 | 1.702 | 5.55E-05 |
| 11 | rs1508098 | 116338899 | C | -0.239 | -0.354 | -0.124 | 5.66E-05 |
| 8 | rs6558293 | 145118501 | G | 0.419 | 0.217 | 0.621 | 5.72E-05 |
| 1 | rs620281 | 100359333 | A | -0.275 | -0.408 | -0.142 | 5.74E-05 |
| 14 | rs17097726 | 21338824 | C | 0.336 | 0.174 | 0.499 | 5.76E-05 |
| 17 | KGP2193194 | 5864270 | T | -0.411 | -0.609 | -0.212 | 5.82E-05 |
| 5 | KGP2886748 | 30797764 | C | 0.264 | 0.137 | 0.392 | 5.87E-05 |
| 17 | KGP4356413 | 47322264 | G | -0.238 | -0.352 | -0.123 | 5.90E-05 |
| 3 | rs6793245 | 38599037 | A | 0.239 | 0.123 | 0.354 | 5.91E-05 |
| 2 | rs2215944 | 103308109 | A | 0.283 | 0.146 | 0.420 | 5.93E-05 |
| 1 | rs3015315 | 59271489 | G | -0.259 | -0.385 | -0.134 | 5.97E-05 |
| 12 | KGP10449250 | 16997890 | C | 0.243 | 0.125 | 0.361 | 6.01E-05 |
| 1 | rs185580 | 182519861 | T | 0.433 | 0.224 | 0.643 | 6.05E-05 |
| 11 | KGP11936758 | 46695365 | T | 3.523 | 1.816 | 5.230 | 6.17E-05 |
| 5 | rs1553238 | 30805404 | T | 0.268 | 0.138 | 0.397 | 6.22E-05 |
| 9 | KGP6347718 | 80933424 | A | 0.242 | 0.124 | 0.359 | 6.23E-05 |
| 16 | KGP22775411 | 27303201 | T | 3.516 | 1.810 | 5.222 | 6.27E-05 |
| 8 | KGP7605529 | 124688861 | A | -0.332 | -0.493 | -0.171 | 6.29E-05 |
| 21 | KGP6049103 | 24879624 | T | 0.703 | 0.362 | 1.044 | 6.35E-05 |
| 7 | KGP4058397 | 117484407 | T | 3.513 | 1.807 | 5.220 | 6.38E-05 |
| 2 | KGP10541057 | 202578735 | C | -0.398 | -0.591 | -0.204 | 6.46E-05 |
| 1 | KGP7574149 | 53757935 | A | 0.933 | 0.479 | 1.387 | 6.51E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 12 | KGP11305355 | 126878310 | C | -0.282 | -0.419 | -0.145 | 6.52E-05 |
| 12 | rs7972835 | 126881318 | T | -0.282 | -0.419 | -0.145 | 6.52E-05 |
| 7 | KGP22777208 | 67384864 | C | 3.537 | 1.816 | 5.259 | 6.61E-05 |
| 20 | rs16996025 | 50061314 | T | 0.552 | 0.283 | 0.820 | 6.67E-05 |
| 17 | rs9913468 | 8987850 | A | 3.537 | 1.814 | 5.260 | 6.68E-05 |
| 8 | KGP4424636 | 4263108 | C | -0.246 | -0.365 | -0.126 | 6.81E-05 |
| 7 | KGP5625284 | 116680062 | A | 3.539 | 1.812 | 5.266 | 6.88E-05 |
| 23 | KGP22754315 | 38921384 | A | 0.287 | 0.147 | 0.427 | 6.88E-05 |
| 2 | KGP724647 | 202593280 | A | -0.386 | -0.574 | -0.197 | 6.88E-05 |
| 12 | KGP19017736 | 7981180 | G | 3.534 | 1.808 | 5.259 | 6.99E-05 |
| 4 | KGP10278399 | 83506328 | T | -0.378 | -0.563 | -0.193 | 7.04E-05 |
| 3 | KGP17690561 | 141642493 | G | 3.539 | 1.809 | 5.270 | 7.11E-05 |
| 15 | rs7182589 | 55121068 | A | -0.274 | -0.407 | -0.140 | 7.17E-05 |
| 2 | KGP14352045 | 28078741 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 2 | KGP14630788 | 39417280 | C | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 2 | KGP4441825 | 67206868 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 2 | rs9678580 | 67911392 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 2 | KGP5719345 | 77595289 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 3 | KGP17790171 | 58260764 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 4 | KGP20914043 | 43210788 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 7 | KGP8187902 | 68466002 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 7 | KGP10217381 | 68469575 | C | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 8 | KGP5758399 | 80102186 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 9 | KGP11502707 | 9878200 | G | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 9 | KGP18685884 | 23017950 | C | 3.535 | 1.806 | 5.265 | 7.18E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 9 | KGP18500611 | 87043507 | G | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 10 | KGP22030309 | 122576439 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 11 | KGP12773346 | 46784448 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 11 | KGP12659342 | 63342699 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 12 | KGP18905928 | 8188744 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 12 | KGP18726703 | 8202273 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 12 | KGP18911582 | 8284164 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 12 | KGP19124898 | 8290881 | C | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 12 | KGP18807926 | 40940381 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 16 | KGP1402542 | 6498791 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 16 | rs9282774 | 11647532 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 17 | KGP3621758 | 8978285 | G | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 18 | KGP16029087 | 40550514 | T | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 18 | KGP7261933 | 76950069 | A | 3.535 | 1.806 | 5.265 | 7.18E-05 |
| 5 | KGP7070362 | 30789521 | A | 0.261 | 0.133 | 0.389 | 7.18E-05 |
| 16 | KGP1320860 | 61998278 | T | 0.240 | 0.122 | 0.357 | 7.19E-05 |
| 11 | rs6485688 | 46762142 | C | 3.538 | 1.806 | 5.269 | 7.22E-05 |
| 13 | KGP1341706 | 82856926 | C | 3.533 | 1.804 | 5.263 | 7.25E-05 |
| 11 | KGP2548532 | 46926246 | A | 3.536 | 1.806 | 5.267 | 7.25E-05 |
| 7 | KGP22829342 | 71404269 | A | 3.536 | 1.805 | 5.267 | 7.26E-05 |
| 5 | rs2616314 | 99243418 | T | 3.534 | 1.804 | 5.265 | 7.28E-05 |
| 4 | KGP9086245 | 107715244 | T | 3.534 | 1.804 | 5.265 | 7.29E-05 |
| 12 | rs10848031 | 130676091 | A | -0.228 | -0.340 | -0.116 | 7.41E-05 |
| 16 | KGP4010190 | 61998405 | A | 0.238 | 0.121 | 0.355 | 7.50E-05 |
| 8 | KGP8492828 | 55199771 | A | 1.260 | 0.642 | 1.878 | 7.54E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 8 | KGP22764564 | 55211121 | A | 1.260 | 0.642 | 1.878 | 7.54E-05 |
| 6 | rs2500585 | 107637582 | C | -0.280 | -0.417 | -0.142 | 7.59E-05 |
| 3 | KGP2510828 | 101821068 | C | -0.229 | -0.341 | -0.116 | 7.60E-05 |
| 9 | KGP9181314 | 97050935 | A | 0.293 | 0.149 | 0.437 | 7.77E-05 |
| 9 | rs1536690 | 97055310 | A | 0.293 | 0.149 | 0.437 | 7.77E-05 |
| 5 | rs17623890 | 29151261 | A | 0.406 | 0.207 | 0.606 | 7.77E-05 |
| 9 | rs17373603 | 98493505 | T | 0.408 | 0.208 | 0.609 | 7.78E-05 |
| 12 | KGP2171465 | 128015666 | T | 0.230 | 0.117 | 0.343 | 7.82E-05 |
| 12 | rs16912235 | 16963281 | T | 0.241 | 0.123 | 0.360 | 7.83E-05 |
| 20 | KGP5722281 | 38827848 | T | -0.300 | -0.447 | -0.152 | 7.86E-05 |
| 3 | rs1399272 | 101811776 | G | -0.227 | -0.339 | -0.115 | 7.88E-05 |
| 3 | KGP4822429 | 101811776 | C | -0.227 | -0.339 | -0.115 | 7.88E-05 |
| 14 | KGP19526628 | 33086732 | A | -0.825 | -1.231 | -0.419 | 7.90E-05 |
| 4 | KGP3751996 | 175827177 | A | -0.224 | -0.335 | -0.114 | 7.90E-05 |
| 4 | rs6553847 | 175836048 | G | -0.224 | -0.335 | -0.114 | 7.90E-05 |
| 8 | rs7838113 | 4264676 | G | -0.289 | -0.431 | -0.147 | 7.98E-05 |
| 14 | rs2281677 | 23284572 | T | 0.224 | 0.114 | 0.335 | 7.99E-05 |
| 1 | KGP38486 | 216899635 | G | 0.349 | 0.177 | 0.521 | 8.02E-05 |
| 6 | KGP4132992 | 103866352 | A | -0.492 | -0.735 | -0.250 | 8.04E-05 |
| 4 | rs4691212 | 166585445 | G | -0.232 | -0.346 | -0.118 | 8.16E-05 |
| 1 | KGP2685985 | 102204932 | G | 1.775 | 0.900 | 2.651 | 8.16E-05 |
| 17 | KGP11116724 | 38125891 | C | 0.261 | 0.132 | 0.390 | 8.20E-05 |
| 3 | rs17216007 | 6475608 | C | 0.431 | 0.218 | 0.643 | 8.20E-05 |
| 1 | rs6694800 | 40551745 | T | 0.521 | 0.264 | 0.777 | 8.25E-05 |
| 20 | rs3787207 | 50056419 | A | 0.691 | 0.350 | 1.033 | 8.38E-05 |

| CHR | SNP | BP | MINOR ALLELE | BETA | L95 | U95 | ADJ P-value |
|---|---|---|---|---|---|---|---|
| 2 | KGP1852138 | 2004564 | G | 0.256 | 0.130 | 0.383 | 8.54E-05 |
| 7 | KGP11111926 | 57343139 | A | -0.672 | -1.004 | -0.340 | 8.55E-05 |
| 4 | rs17028403 | 152943626 | A | 0.300 | 0.152 | 0.449 | 8.58E-05 |
| 3 | KGP7730892 | 38598606 | T | 0.276 | 0.140 | 0.413 | 8.65E-05 |
| 20 | KGP11474206 | 2376920 | T | 0.242 | 0.122 | 0.361 | 8.69E-05 |
| 8 | rs1430449 | 4258195 | T | -0.298 | -0.445 | -0.150 | 8.76E-05 |
| 17 | KGP22820022 | 77463529 | C | 0.554 | 0.279 | 0.828 | 8.80E-05 |
| 14 | KGP3068459 | 29845225 | A | 0.237 | 0.119 | 0.354 | 8.91E-05 |
| 8 | KGP20314184 | 55208974 | T | 1.175 | 0.592 | 1.758 | 8.98E-05 |
| 1 | rs10864742 | 230463537 | A | 0.316 | 0.159 | 0.472 | 9.00E-05 |
| 11 | rs17119743 | 116344404 | G | -0.248 | -0.371 | -0.125 | 9.07E-05 |
| 10 | KGP4897218 | 99503021 | T | 0.383 | 0.193 | 0.573 | 9.14E-05 |
| 11 | rs10893871 | 128325507 | T | 0.247 | 0.124 | 0.370 | 9.22E-05 |
| 20 | KGP7935258 | 798547 | G | 0.534 | 0.269 | 0.800 | 9.31E-05 |
| 13 | rs12100138 | 23770808 | T | -0.575 | -0.860 | -0.289 | 9.38E-05 |
| 14 | rs11161052 | 29836773 | T | 0.234 | 0.118 | 0.351 | 9.38E-05 |
| 13 | KGP6651515 | 85865655 | G | -0.343 | -0.513 | -0.172 | 9.40E-05 |
| 1 | rs266549 | 182528269 | A | 0.420 | 0.211 | 0.629 | 9.50E-05 |
| 1 | rs1257394 | 201254116 | C | 0.295 | 0.148 | 0.442 | 9.52E-05 |
| 7 | KGP1446355 | 57363286 | T | -0.667 | -0.999 | -0.335 | 9.63E-05 |
| 1 | rs2984908 | 59304719 | T | -0.249 | -0.373 | -0.125 | 9.70E-05 |
| 1 | KGP6439718 | 59306823 | A | -0.249 | -0.373 | -0.125 | 9.70E-05 |

**Appendix J. VAAST genes with most significant p-values in family 1168 sorted by rank.** P-values were calculated using the fast genome-permutation option with 1e5 permutations.

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 1 | *PDE4DIP* | 1.67E-06 | 512.63 |
| 2 | *HEATR1* | 1.67E-06 | 337.39 |
| 3 | *HRNR* | 1.67E-06 | 310.50 |
| 4 | *OR6N1* | 1.67E-06 | 247.08 |
| 5 | *USH2A* | 1.67E-06 | 235.23 |
| 6 | *OR2L8* | 1.67E-06 | 234.92 |
| 7 | *OR11L1* | 1.67E-06 | 231.18 |
| 8 | *OR2M7* | 1.67E-06 | 192.86 |
| 9 | *EXO1* | 1.67E-06 | 189.52 |
| 10 | *OR2W3* | 1.67E-06 | 184.78 |
| 11 | *KIR3DL1* | 1.67E-06 | 178.51 |
| 12 | *CR1* | 1.67E-06 | 174.31 |
| 13 | *LGALS8* | 1.67E-06 | 173.87 |
| 14 | *SPTA1* | 1.67E-06 | 170.29 |
| 15 | *NES* | 1.67E-06 | 169.47 |
| 16 | *OR14I1* | 1.67E-06 | 167.38 |
| 17 | *OR4C3* | 1.67E-06 | 158.42 |
| 18 | *KIAA1614* | 1.67E-06 | 151.24 |
| 19 | *KIAA1324* | 1.67E-06 | 148.60 |
| 20 | *PEAR1* | 1.67E-06 | 145.22 |
| 21 | *TCHH* | 1.67E-06 | 135.75 |
| 22 | *CENPF* | 1.67E-06 | 121.82 |
| 23 | *OR2G2* | 1.67E-06 | 119.26 |
| 24 | *FCRL5* | 1.67E-06 | 119.03 |
| 25 | *IFI16* | 1.67E-06 | 118.95 |
| 26 | *CHIA* | 1.67E-06 | 117.83 |
| 27 | *PGLYRP4* | 1.67E-06 | 115.09 |
| 28 | *TOR1AIP1* | 1.67E-06 | 113.73 |
| 29 | *TLR5* | 1.67E-06 | 113.51 |
| 30 | *DTL* | 1.67E-06 | 112.78 |
| 31 | *CCDC76* | 1.67E-06 | 112.56 |
| 32 | *FAM71A* | 1.67E-06 | 112.36 |
| 33 | *OR2T11* | 1.67E-06 | 111.33 |
| 34 | *LAMC1* | 1.67E-06 | 111.24 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 35 | *OR2C3* | 1.67E-06 | 109.80 |
| 36 | *CDC27* | 1.67E-06 | 105.25 |
| 37 | *OR9G9* | 1.67E-06 | 104.26 |
| 38 | *OR2T6* | 1.67E-06 | 98.89 |
| 39 | *OR2B11* | 1.67E-06 | 97.80 |
| 40 | *THEM5* | 1.67E-06 | 95.04 |
| 41 | *IQGAP3* | 1.67E-06 | 93.68 |
| 42 | *OR10J1* | 1.67E-06 | 93.46 |
| 43 | *ADAR* | 1.67E-06 | 92.91 |
| 44 | *MRPL9* | 1.67E-06 | 91.19 |
| 45 | *F5* | 1.67E-06 | 91.08 |
| 46 | *PTGFRN* | 1.67E-06 | 87.94 |
| 47 | *ZNF805* | 1.67E-06 | 83.06 |
| 48 | *OR2T12* | 1.67E-06 | 82.08 |
| 49 | *IGSF3* | 1.67E-06 | 81.48 |
| 50 | *OR14C36* | 1.67E-06 | 79.06 |
| 51 | *ADAM15* | 1.67E-06 | 77.87 |
| 52 | *AKNAD1* | 1.67E-06 | 77.61 |
| 53 | *ASTN1* | 1.67E-06 | 76.82 |
| 54 | *KIR2DL4_DUP_01* | 1.67E-06 | 75.51 |
| 55 | *CR1L* | 1.67E-06 | 75.22 |
| 56 | *ARHGEF11* | 1.67E-06 | 74.39 |
| 57 | *MUC6* | 1.67E-06 | 74.29 |
| 58 | *ZNF695* | 1.67E-06 | 73.25 |
| 59 | *FAM104B_DUP_01* | 1.67E-06 | 72.30 |
| 60 | *HYDIN* | 1.67E-06 | 68.67 |
| 61 | *EPS8L3* | 1.67E-06 | 65.99 |
| 62 | *CTBP2* | 1.67E-06 | 60.41 |
| 63 | *GPR37L1* | 1.67E-06 | 58.49 |
| 64 | *F13B* | 1.67E-06 | 57.60 |
| 65 | *RRNAD1* | 1.67E-06 | 57.52 |
| 66 | *MYBPHL* | 1.67E-06 | 57.30 |
| 67 | *BCL2L15* | 1.67E-06 | 57.30 |
| 68 | *TMEM81* | 1.67E-06 | 57.09 |
| 69 | *PRODH* | 1.67E-06 | 57.09 |
| 70 | *CR2* | 1.67E-06 | 56.76 |
| 71 | *CSF1* | 1.67E-06 | 56.54 |
| 72 | *PRG4* | 1.67E-06 | 56.54 |
| 73 | *ASPM* | 1.67E-06 | 56.32 |
| 74 | *ACP6* | 1.67E-06 | 56.07 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 75 | *DSTYK* | 1.67E-06 | 55.95 |
| 76 | *ITPKB* | 1.67E-06 | 55.70 |
| 77 | *XKR3* | 1.67E-06 | 55.66 |
| 78 | *PTPN22* | 1.67E-06 | 55.65 |
| 79 | *TRIM33* | 1.67E-06 | 55.63 |
| 80 | *OR14A16* | 1.67E-06 | 55.63 |
| 81 | *MIA3* | 1.67E-06 | 55.61 |
| 82 | *C1orf111* | 1.67E-06 | 55.43 |
| 83 | *CDC42BPA* | 1.67E-06 | 55.43 |
| 84 | *CEP350* | 1.67E-06 | 55.38 |
| 85 | *ANKRD35* | 1.67E-06 | 55.38 |
| 86 | *PPP1R15B* | 1.67E-06 | 55.36 |
| 87 | *AHCTF1* | 1.67E-06 | 55.36 |
| 88 | *FAM63A* | 1.67E-06 | 55.25 |
| 89 | *C1orf116* | 1.67E-06 | 54.92 |
| 90 | *C1orf204* | 1.67E-06 | 54.76 |
| 91 | *RGS16* | 1.67E-06 | 54.66 |
| 92 | *LYPLAL1* | 1.67E-06 | 54.64 |
| 93 | *OR2T27* | 1.67E-06 | 54.61 |
| 94 | *CD101* | 1.67E-06 | 54.46 |
| 95 | *PRSS38* | 1.67E-06 | 54.46 |
| 96 | *SYT11* | 1.67E-06 | 54.42 |
| 97 | *S100A7* | 1.67E-06 | 54.32 |
| 98 | *RXFP4* | 1.67E-06 | 54.10 |
| 99 | *CNST* | 1.67E-06 | 54.10 |
| 100 | *EPRS* | 1.67E-06 | 53.81 |
| 101 | *FNDC7* | 1.67E-06 | 52.92 |
| 102 | *TSHB* | 1.67E-06 | 52.52 |
| 103 | *ASH1L* | 1.67E-06 | 52.52 |
| 104 | *PRSS3* | 1.67E-06 | 52.52 |
| 105 | *DBT* | 1.67E-06 | 52.49 |
| 106 | *PPM1J* | 1.67E-06 | 52.36 |
| 107 | *PM20D1* | 1.67E-06 | 52.30 |
| 108 | *APOBEC4* | 1.67E-06 | 50.18 |
| 109 | *SETD8* | 1.67E-06 | 50.13 |
| 110 | *HMCN1* | 1.67E-06 | 49.47 |
| 111 | *COL11A1* | 1.67E-06 | 49.22 |
| 112 | *RNASEL* | 1.67E-06 | 48.93 |
| 113 | *BGLAP* | 1.67E-06 | 47.00 |
| 114 | *LRRC71* | 1.67E-06 | 46.59 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 115 | *OR10T2* | 1.67E-06 | 45.59 |
| 116 | *ITLN1* | 1.67E-06 | 41.03 |
| 117 | *OR6K3* | 1.67E-06 | 40.72 |
| 118 | *FMN2* | 1.67E-06 | 40.22 |
| 119 | *ITLN2* | 1.67E-06 | 40.22 |
| 120 | *SMYD2* | 1.67E-06 | 39.95 |
| 121 | *URB2* | 1.67E-06 | 39.75 |
| 122 | *ST7L* | 1.67E-06 | 39.71 |
| 123 | *EDEM3* | 1.67E-06 | 39.62 |
| 124 | *LEFTY1* | 1.67E-06 | 38.83 |
| 125 | *OR10R2* | 1.67E-06 | 38.44 |
| 126 | *CHI3L1* | 1.67E-06 | 38.30 |
| 127 | *GPRIN2* | 1.67E-06 | 38.30 |
| 128 | *TTF2* | 1.67E-06 | 38.23 |
| 129 | *ZNF845* | 1.67E-06 | 38.00 |
| 130 | *CHD1L* | 1.67E-06 | 37.84 |
| 131 | *SLC6A17* | 1.67E-06 | 37.76 |
| 132 | *OR10Z1* | 1.67E-06 | 37.74 |
| 133 | *POTEF* | 1.67E-06 | 37.56 |
| 134 | *SEMA6C* | 1.67E-06 | 37.55 |
| 135 | *HSD3B1* | 1.67E-06 | 37.51 |
| 136 | *AGT* | 1.67E-06 | 37.48 |
| 137 | *FCGR2A* | 1.67E-06 | 37.28 |
| 138 | *SEC16B* | 1.67E-06 | 37.28 |
| 139 | *SHISA4_DUP_01* | 1.67E-06 | 37.26 |
| 140 | *ARHGAP30* | 1.67E-06 | 37.23 |
| 141 | *AXDND1* | 1.67E-06 | 37.09 |
| 142 | *SLAMF9* | 1.67E-06 | 37.09 |
| 143 | *NSL1* | 1.67E-06 | 37.09 |
| 144 | *MTMR11* | 1.67E-06 | 37.09 |
| 145 | *ATF6* | 1.67E-06 | 37.09 |
| 146 | *ACBD3* | 1.67E-06 | 36.94 |
| 147 | *SLC16A1* | 1.67E-06 | 36.44 |
| 148 | *SCYL3* | 1.67E-06 | 36.00 |
| 149 | *C1orf105* | 1.67E-06 | 35.79 |
| 150 | *LBR* | 1.67E-06 | 35.79 |
| 151 | *HBXIP* | 1.67E-06 | 35.79 |
| 152 | *PLA2G4A* | 1.67E-06 | 35.71 |
| 153 | *LGR6* | 1.67E-06 | 35.54 |
| 154 | *MOSC1* | 1.67E-06 | 35.14 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 155 | *GSTM3* | 1.67E-06 | 34.98 |
| 156 | *PROK1* | 1.67E-06 | 34.98 |
| 157 | *PLEKHA6* | 1.67E-06 | 34.98 |
| 158 | *OR13G1* | 1.67E-06 | 34.92 |
| 159 | *OR2G3* | 1.67E-06 | 34.92 |
| 160 | *VAV3* | 1.67E-06 | 34.67 |
| 161 | *MTX1* | 1.67E-06 | 34.67 |
| 162 | *EFNA1* | 1.67E-06 | 27.65 |
| 163 | *ATP1A4* | 1.67E-06 | 26.50 |
| 164 | *DARC* | 1.67E-06 | 26.50 |
| 165 | *PBXIP1* | 1.67E-06 | 26.50 |
| 166 | *SCCPDH* | 1.67E-06 | 26.11 |
| 167 | *SELP* | 1.67E-06 | 25.31 |
| 168 | *DENND2C* | 1.67E-06 | 25.21 |
| 169 | *LIPK* | 1.67E-06 | 24.94 |
| 170 | *IBA57* | 1.67E-06 | 24.79 |
| 171 | *FLG2* | 1.67E-06 | 24.55 |
| 172 | *CELSR2* | 1.67E-06 | 24.34 |
| 173 | *UAP1* | 1.67E-06 | 24.34 |
| 174 | *HDGF* | 1.67E-06 | 24.29 |
| 175 | *C1orf124* | 1.67E-06 | 24.29 |
| 176 | *ADAMTSL4* | 1.67E-06 | 24.23 |
| 177 | *CLDN16* | 1.67E-06 | 24.23 |
| 178 | *CLCC1* | 1.67E-06 | 24.00 |
| 179 | *TCEB3CL_DUP_02* | 1.67E-06 | 24.00 |
| 180 | *RPTN* | 1.67E-06 | 23.83 |
| 181 | *PLXNA2* | 1.67E-06 | 23.83 |
| 182 | *C4BPA* | 1.67E-06 | 23.64 |
| 183 | *CAPN2* | 1.67E-06 | 23.62 |
| 184 | *EDARADD* | 1.67E-06 | 23.61 |
| 185 | *CEP170* | 1.67E-06 | 23.54 |
| 186 | *C1orf129* | 1.67E-06 | 23.37 |
| 187 | *LHX4* | 1.67E-06 | 23.37 |
| 188 | *OR2T4* | 1.67E-06 | 23.29 |
| 189 | *PIGC* | 1.67E-06 | 23.29 |
| 190 | *BCL9* | 1.67E-06 | 23.29 |
| 191 | *PCMTD1* | 1.67E-06 | 23.13 |
| 192 | *OR2T1* | 1.67E-06 | 22.67 |
| 193 | *METTL13* | 1.67E-06 | 22.48 |
| 194 | *LY9* | 1.67E-06 | 22.48 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 195 | *KCNJ12* | 1.67E-06 | 22.44 |
| 196 | *SIPA1L2* | 1.67E-06 | 22.33 |
| 197 | *CBWD6* | 1.67E-06 | 21.83 |
| 198 | *CD1E* | 1.67E-06 | 21.40 |
| 199 | *ADAMTS4* | 1.67E-06 | 21.40 |
| 200 | *KDM4DL* | 1.67E-06 | 21.40 |
| 201 | *PRSS1* | 1.67E-06 | 21.19 |
| 202 | *ADCY10* | 1.67E-06 | 20.31 |

**Appendix K. VAAST genes with most significant p-values in family 1281 sorted by rank.** P-values were calculated using the fast genome-permutation option with 1e5 permutations.

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 1 | *HRNR* | 1.67E-06 | 537.33 |
| 2 | *PDE4DIP* | 1.67E-06 | 529.39 |
| 3 | *USH2A* | 1.67E-06 | 297.29 |
| 4 | *OR2L8* | 1.67E-06 | 234.92 |
| 5 | *SPTA1* | 1.67E-06 | 234.17 |
| 6 | *OR11L1* | 1.67E-06 | 231.18 |
| 7 | *OR6N1* | 1.67E-06 | 228.36 |
| 8 | *OR2W3* | 1.67E-06 | 215.57 |
| 9 | *IFI16* | 1.67E-06 | 190.83 |
| 10 | *EXO1* | 1.67E-06 | 177.37 |
| 11 | *FCRL5* | 1.67E-06 | 171.17 |
| 12 | *HEATR1* | 1.67E-06 | 171.12 |
| 13 | *KIAA1324* | 1.67E-06 | 168.17 |
| 14 | *OVGP1* | 1.67E-06 | 157.97 |
| 15 | *KIR3DL1* | 1.67E-06 | 146.41 |
| 16 | *IGSF3* | 1.67E-06 | 134.53 |
| 17 | *OR10R2* | 1.67E-06 | 134.26 |
| 18 | *LGR6* | 1.67E-06 | 133.68 |
| 19 | *OR14I1* | 1.67E-06 | 131.01 |
| 20 | *ASPM* | 1.67E-06 | 130.25 |
| 21 | *OR2C3* | 1.67E-06 | 130.03 |
| 22 | *CHIA* | 1.67E-06 | 126.56 |
| 23 | *OR13G1* | 1.67E-06 | 123.55 |
| 24 | *OR10X1* | 1.67E-06 | 120.47 |
| 25 | *PGLYRP4* | 1.67E-06 | 115.09 |
| 26 | *ECM1* | 1.67E-06 | 114.95 |
| 27 | *TCHH* | 1.67E-06 | 114.52 |
| 28 | *DUSP27* | 1.67E-06 | 114.13 |
| 29 | *THEM5* | 1.67E-06 | 113.80 |
| 30 | *TLR5* | 1.67E-06 | 113.51 |
| 31 | *COL11A1* | 1.67E-06 | 113.19 |
| 32 | *KIAA1614* | 1.67E-06 | 113.13 |
| 33 | *DTL* | 1.67E-06 | 112.78 |
| 34 | *FAM71A* | 1.67E-06 | 112.36 |
| 35 | *ASTN1* | 1.67E-06 | 111.57 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 36 | *OR2T11* | 1.67E-06 | 111.33 |
| 37 | *TNR* | 1.67E-06 | 107.39 |
| 38 | *IQGAP3* | 1.67E-06 | 105.92 |
| 39 | *CDC27* | 1.67E-06 | 105.25 |
| 40 | *OR2G2* | 1.67E-06 | 104.18 |
| 41 | *CENPF* | 1.67E-06 | 103.83 |
| 42 | *OR4C3* | 1.67E-06 | 102.50 |
| 43 | *OR2M7* | 1.67E-06 | 102.45 |
| 44 | *HMCN1* | 1.67E-06 | 101.61 |
| 45 | *DNAH14* | 1.67E-06 | 100.87 |
| 46 | *MUC6* | 1.67E-06 | 100.82 |
| 47 | *VAV3* | 1.67E-06 | 98.54 |
| 48 | *CR1L* | 1.67E-06 | 97.06 |
| 49 | *CAPN9* | 1.67E-06 | 94.83 |
| 50 | *CHD1L* | 1.67E-06 | 94.43 |
| 51 | *OR10J1* | 1.67E-06 | 93.46 |
| 52 | *ITPKB* | 1.67E-06 | 89.90 |
| 53 | *KCNJ12* | 1.67E-06 | 87.50 |
| 54 | *CSF1* | 1.67E-06 | 80.73 |
| 55 | *TOR1AIP1* | 1.67E-06 | 78.97 |
| 56 | *KIR3DL2* | 1.67E-06 | 78.30 |
| 57 | *OR9G9* | 1.67E-06 | 78.17 |
| 58 | *RNASEL* | 1.67E-06 | 78.15 |
| 59 | *LGALS8* | 1.67E-06 | 77.91 |
| 60 | *GPRIN2* | 1.67E-06 | 77.85 |
| 61 | *CD101* | 1.67E-06 | 77.70 |
| 62 | *FAM177B* | 1.67E-06 | 77.63 |
| 63 | *ADAR* | 1.67E-06 | 76.92 |
| 64 | *CAPN2* | 1.67E-06 | 76.67 |
| 65 | *CFH* | 1.67E-06 | 76.52 |
| 66 | *LAMC1* | 1.67E-06 | 76.49 |
| 67 | *CDC42BPA* | 1.67E-06 | 75.03 |
| 68 | *OR10T2* | 1.67E-06 | 74.80 |
| 69 | *C1orf68* | 1.67E-06 | 74.52 |
| 70 | *NSL1* | 1.67E-06 | 74.00 |
| 71 | *MAGEC1* | 1.67E-06 | 72.63 |
| 72 | *FAM104B_DUP_01* | 1.67E-06 | 72.30 |
| 73 | *PTGFRN* | 1.67E-06 | 71.94 |
| 74 | *CR1* | 1.67E-06 | 71.81 |
| 75 | *SWT1* | 1.67E-06 | 69.18 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 76 | *HYDIN* | 1.67E-06 | 68.67 |
| 77 | *TNN* | 1.67E-06 | 63.40 |
| 78 | *FNDC7* | 1.67E-06 | 63.34 |
| 79 | *ATF6* | 1.67E-06 | 62.19 |
| 80 | *KIR3DL3* | 1.67E-06 | 61.58 |
| 81 | *PM20D1* | 1.67E-06 | 60.22 |
| 82 | *EFNA1* | 1.67E-06 | 59.63 |
| 83 | *ATP1A4* | 1.67E-06 | 58.49 |
| 84 | *GPR37L1* | 1.67E-06 | 58.49 |
| 85 | *LCE5A* | 1.67E-06 | 57.65 |
| 86 | *F13B* | 1.67E-06 | 57.60 |
| 87 | *RRNAD1* | 1.67E-06 | 57.52 |
| 88 | *SMYD2* | 1.67E-06 | 57.33 |
| 89 | *MYBPHL* | 1.67E-06 | 57.30 |
| 90 | *BCL2L15* | 1.67E-06 | 57.30 |
| 91 | *PRODH* | 1.67E-06 | 57.09 |
| 92 | *EPS8L3* | 1.67E-06 | 56.93 |
| 93 | *PRG4* | 1.67E-06 | 56.54 |
| 94 | *ADAMTSL4* | 1.67E-06 | 56.22 |
| 95 | *FLVCR1* | 1.67E-06 | 56.18 |
| 96 | *ACP6* | 1.67E-06 | 56.07 |
| 97 | *OR2B11* | 1.67E-06 | 56.07 |
| 98 | *OR14C36* | 1.67E-06 | 55.66 |
| 99 | *C1orf227* | 1.67E-06 | 55.63 |
| 100 | *TRIM33* | 1.67E-06 | 55.63 |
| 101 | *OR14A16* | 1.67E-06 | 55.63 |
| 102 | *TTF2* | 1.67E-06 | 55.61 |
| 103 | *MIA3* | 1.67E-06 | 55.61 |
| 104 | *C1orf111* | 1.67E-06 | 55.43 |
| 105 | *JMJD4* | 1.67E-06 | 55.43 |
| 106 | *SH2D2A* | 1.67E-06 | 55.36 |
| 107 | *PPP1R15B* | 1.67E-06 | 55.36 |
| 108 | *ADAM15* | 1.67E-06 | 55.35 |
| 109 | *FAM63A* | 1.67E-06 | 55.25 |
| 110 | *SEMA6C* | 1.67E-06 | 54.93 |
| 111 | *PEAR1* | 1.67E-06 | 54.92 |
| 112 | *C1orf116* | 1.67E-06 | 54.92 |
| 113 | *APOA1BP* | 1.67E-06 | 54.77 |
| 114 | *DIEXF* | 1.67E-06 | 54.76 |
| 115 | *C1orf204* | 1.67E-06 | 54.76 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 116 | *SEC16B* | 1.67E-06 | 54.66 |
| 117 | *UBQLN4* | 1.67E-06 | 54.64 |
| 118 | *LYPLAL1* | 1.67E-06 | 54.64 |
| 119 | *OR2T27* | 1.67E-06 | 54.61 |
| 120 | *SLAMF9* | 1.67E-06 | 54.47 |
| 121 | *PRSS38* | 1.67E-06 | 54.46 |
| 122 | *SYT11* | 1.67E-06 | 54.42 |
| 123 | *S100A7* | 1.67E-06 | 54.32 |
| 124 | *RXFP4* | 1.67E-06 | 54.10 |
| 125 | *CNST* | 1.67E-06 | 54.10 |
| 126 | *ZNF805* | 1.67E-06 | 53.85 |
| 127 | *SLC16A1* | 1.67E-06 | 53.81 |
| 128 | *EPRS* | 1.67E-06 | 53.81 |
| 129 | *SCYL3* | 1.67E-06 | 53.38 |
| 130 | *F5* | 1.67E-06 | 53.38 |
| 131 | *LBR* | 1.67E-06 | 53.17 |
| 132 | *PLA2G4A* | 1.67E-06 | 53.09 |
| 133 | *ERO1LB* | 1.67E-06 | 52.58 |
| 134 | *TSHB* | 1.67E-06 | 52.52 |
| 135 | *ASH1L* | 1.67E-06 | 52.52 |
| 136 | *PRSS3* | 1.67E-06 | 52.52 |
| 137 | *DBT* | 1.67E-06 | 52.49 |
| 138 | *PPM1J* | 1.67E-06 | 52.36 |
| 139 | *AQP10* | 1.67E-06 | 52.05 |
| 140 | *BCLAF1* | 1.67E-06 | 50.52 |
| 141 | *SELP* | 1.67E-06 | 50.25 |
| 142 | *APOBEC4* | 1.67E-06 | 50.18 |
| 143 | *SETD8* | 1.67E-06 | 50.13 |
| 144 | *POTED* | 1.67E-06 | 49.26 |
| 145 | *CEP89* | 1.67E-06 | 47.61 |
| 146 | *NES* | 1.67E-06 | 46.66 |
| 147 | *SAA2-SAA4* | 1.67E-06 | 46.38 |
| 148 | *C1orf85* | 1.67E-06 | 45.60 |
| 149 | *SPRR1A* | 1.67E-06 | 45.30 |
| 150 | *ARHGEF11* | 1.67E-06 | 45.17 |
| 151 | *PDE10A* | 1.67E-06 | 44.20 |
| 152 | *ZNF695* | 1.67E-06 | 44.04 |
| 153 | *FCRLB* | 1.67E-06 | 41.11 |
| 154 | *OR6K3* | 1.67E-06 | 40.72 |
| 155 | *KIR2DL1* | 1.67E-06 | 40.14 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 156 | *DENND2C* | 1.67E-06 | 39.82 |
| 157 | *URB2* | 1.67E-06 | 39.75 |
| 158 | *TMEM81* | 1.67E-06 | 39.71 |
| 159 | *NUF2* | 1.67E-06 | 39.69 |
| 160 | *C1orf162* | 1.67E-06 | 39.39 |
| 161 | *PIGR* | 1.67E-06 | 39.39 |
| 162 | *CR2* | 1.67E-06 | 39.38 |
| 163 | *IL19* | 1.67E-06 | 39.16 |
| 164 | *KPRP* | 1.67E-06 | 39.16 |
| 165 | *CLDN16* | 1.67E-06 | 38.84 |
| 166 | *OR2AK2* | 1.67E-06 | 38.69 |
| 167 | *DSTYK* | 1.67E-06 | 38.57 |
| 168 | *TBX15* | 1.67E-06 | 38.28 |
| 169 | *PTPN22* | 1.67E-06 | 38.27 |
| 170 | *C4BPA* | 1.67E-06 | 38.25 |
| 171 | *DDX20* | 1.67E-06 | 38.25 |
| 172 | *CEP350* | 1.67E-06 | 38.00 |
| 173 | *ANKRD35* | 1.67E-06 | 38.00 |
| 174 | *LHX4* | 1.67E-06 | 37.98 |
| 175 | *FAM46C* | 1.67E-06 | 37.54 |
| 176 | *TRAF3IP3* | 1.67E-06 | 37.38 |
| 177 | *AXDND1* | 1.67E-06 | 37.38 |
| 178 | *RGS16* | 1.67E-06 | 37.28 |
| 179 | *SHISA4_DUP_01* | 1.67E-06 | 37.26 |
| 180 | *METTL13* | 1.67E-06 | 37.09 |
| 181 | *LY9* | 1.67E-06 | 37.09 |
| 182 | *SDCCAG8* | 1.67E-06 | 36.94 |
| 183 | *CD1E* | 1.67E-06 | 36.00 |
| 184 | *BGLAP* | 1.67E-06 | 36.00 |
| 185 | *HBXIP* | 1.67E-06 | 35.79 |
| 186 | *MOSC1* | 1.67E-06 | 35.14 |
| 187 | *LCE1E* | 1.67E-06 | 35.11 |
| 188 | *PLEKHA6* | 1.67E-06 | 34.98 |
| 189 | *TARBP1* | 1.67E-06 | 34.97 |
| 190 | *ADCY10* | 1.67E-06 | 34.92 |
| 191 | *BBS9* | 1.67E-06 | 32.71 |
| 192 | *CHI3L2* | 1.67E-06 | 27.07 |
| 193 | *PAPPA2* | 1.67E-06 | 27.07 |
| 194 | *CACNA1S* | 1.67E-06 | 26.70 |
| 195 | *KMO* | 1.67E-06 | 26.70 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 196 | *DCST1* | 1.67E-06 | 26.70 |
| 197 | *PIP* | 1.67E-06 | 26.67 |
| 198 | *FMO3* | 1.67E-06 | 26.18 |
| 199 | *BCAN* | 1.67E-06 | 26.18 |
| 200 | *OR2T12* | 1.67E-06 | 26.13 |
| 201 | *KIAA1671* | 1.67E-06 | 26.13 |
| 202 | *VSIG4* | 1.67E-06 | 26.11 |
| 203 | *FMN2* | 1.67E-06 | 25.61 |
| 204 | *FRG1* | 1.67E-06 | 25.35 |
| 205 | *NUDT22* | 1.67E-06 | 25.22 |
| 206 | *FCRL4* | 1.67E-06 | 25.11 |
| 207 | *SHMT2_DUP_01* | 1.67E-06 | 25.11 |
| 208 | *YIF1B* | 1.67E-06 | 25.05 |
| 209 | *TPTE* | 1.67E-06 | 25.05 |
| 210 | *C1orf88* | 1.67E-06 | 24.73 |
| 211 | *ZNF337* | 1.67E-06 | 24.69 |
| 212 | *TDRD5* | 1.67E-06 | 24.53 |
| 213 | *C1orf110* | 1.67E-06 | 24.53 |
| 214 | *IL24* | 1.67E-06 | 24.34 |
| 215 | *C1orf129* | 1.67E-06 | 24.34 |
| 216 | *TEDDM1* | 1.67E-06 | 24.34 |
| 217 | *COL4A5* | 1.67E-06 | 24.29 |
| 218 | *INSRR* | 1.67E-06 | 24.29 |
| 219 | *C1orf124* | 1.67E-06 | 24.29 |
| 220 | *KIR2DL4_DUP_01* | 1.67E-06 | 24.29 |
| 221 | *TMEM79_DUP_01* | 1.67E-06 | 24.09 |
| 222 | *EFNA3* | 1.67E-06 | 24.09 |
| 223 | *HHIPL2* | 1.67E-06 | 24.03 |
| 224 | *KIF14* | 1.67E-06 | 24.00 |
| 225 | *CLCC1* | 1.67E-06 | 24.00 |
| 226 | *TCEB3CL_DUP_02* | 1.67E-06 | 24.00 |
| 227 | *DISC1* | 1.67E-06 | 23.70 |
| 228 | *NAIF1* | 1.67E-06 | 23.70 |
| 229 | *OR2M2* | 1.67E-06 | 23.69 |
| 230 | *CGB1* | 1.67E-06 | 23.69 |
| 231 | *EDARADD* | 1.67E-06 | 23.61 |
| 232 | *SLAMF1* | 1.67E-06 | 23.58 |
| 233 | *NOTCH2NL* | 1.67E-06 | 23.54 |
| 234 | *AHCTF1* | 1.67E-06 | 23.37 |
| 235 | *CREB3L4* | 1.67E-06 | 23.29 |

| Rank | Gene | P-value | Score |
|---|---|---|---|
| 236 | *NLRP3* | 1.67E-06 | 23.19 |
| 237 | *OR10Z1* | 1.67E-06 | 23.13 |
| 238 | *PCMTD1* | 1.67E-06 | 23.13 |
| 239 | *METTL11B* | 1.67E-06 | 22.93 |
| 240 | *HSD3B1* | 1.67E-06 | 22.91 |
| 241 | *AGT* | 1.67E-06 | 22.88 |
| 242 | *CD244* | 1.67E-06 | 22.80 |
| 243 | *MFSD4* | 1.67E-06 | 22.80 |
| 244 | *MR1* | 1.67E-06 | 22.67 |
| 245 | *FCGR2A* | 1.67E-06 | 22.67 |
| 246 | *FCRL2* | 1.67E-06 | 22.63 |
| 247 | *ARHGAP30* | 1.67E-06 | 22.62 |
| 248 | *SPRR3* | 1.67E-06 | 22.62 |
| 249 | *LAMC2* | 1.67E-06 | 22.59 |
| 250 | *RASA2* | 1.67E-06 | 22.59 |
| 251 | *CFHR4* | 1.67E-06 | 22.49 |
| 252 | *MTMR11* | 1.67E-06 | 22.48 |
| 253 | *CRTC2* | 1.67E-06 | 22.48 |
| 254 | *RTBDN* | 1.67E-06 | 22.36 |
| 255 | *ACBD3* | 1.67E-06 | 22.34 |
| 256 | *METTL18* | 1.67E-06 | 22.34 |
| 257 | *ATP8B2* | 1.67E-06 | 22.02 |
| 258 | *NCF2* | 1.67E-06 | 22.02 |
| 259 | *CBWD6* | 1.67E-06 | 21.83 |
| 260 | *NAP1L2* | 1.67E-06 | 21.83 |
| 261 | *KISS1* | 1.67E-06 | 21.40 |
| 262 | *AKNAD1* | 1.67E-06 | 21.19 |
| 263 | *AMY2B* | 1.67E-06 | 20.93 |
| 264 | *PARP1* | 1.67E-06 | 20.93 |
| 265 | *BEND3* | 1.67E-06 | 20.93 |
| 266 | *IVL* | 1.67E-06 | 20.53 |
| 267 | *CASQ2* | 1.67E-06 | 20.53 |
| 268 | *OR6Y1* | 1.67E-06 | 20.37 |
| 269 | *GSTM3* | 1.67E-06 | 20.37 |
| 270 | *NUP210L* | 1.67E-06 | 20.37 |
| 271 | *ILDR2* | 1.67E-06 | 20.37 |
| 272 | *XPR1* | 1.67E-06 | 20.31 |
| 273 | *ABL2* | 1.67E-06 | 20.07 |
| 274 | *NUP98* | 1.67E-06 | 19.56 |
| 275 | *NKRF* | 1.67E-06 | 18.20 |

**Appendix L. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of most significant p-value VAAST genes for family 1168.** The table includes KEGG pathway ID, pathway name and number of genes in parentheses.

| KEGG Pathways Top P-value Genes Family 1168 |
|---|
| hsa04740 Olfactory transduction (25) |
| hsa01100 Metabolic pathways (8) |
| hsa04610 Complement and coagulation cascades (5) |
| hsa04640 Hematopoietic cell lineage (4) |
| hsa04510 Focal adhesion (4) |
| hsa04972 Pancreatic secretion (4) |
| hsa05164 Influenza A (4) |
| hsa04974 Protein digestion and absorption (4) |
| hsa04080 Neuroactive ligand-receptor interaction (3) |
| hsa04141 Protein processing in endoplasmic reticulum (3) |
| hsa05144 Malaria (3) |
| hsa00520 Amino sugar and nucleotide sugar metabolism (3) |
| hsa04666 Fc gamma R-mediated phagocytosis (3) |
| hsa04360 Axon guidance (3) |
| hsa04810 Regulation of actin cytoskeleton (2) |
| hsa05145 Toxoplasmosis (2) |
| hsa05010 Alzheimer's disease (2) |
| hsa04530 Tight junction (2) |
| hsa04662 B cell receptor signaling pathway (2) |
| hsa04664 Fc epsilon RI signaling pathway (2) |
| hsa05140 Leishmaniasis (2) |
| hsa04514 Cell adhesion molecules (CAMs) (2) |
| hsa05134 Legionellosis (2) |
| hsa04380 Osteoclast differentiation (2) |
| hsa00310 Lysine degradation (2) |
| hsa04512 ECM-receptor interaction (2) |
| hsa05152 Tuberculosis (2) |
| hsa05200 Pathways in cancer (2) |
| hsa04650 Natural killer cell mediated cytotoxicity (2) |
| hsa05150 Staphylococcus aureus infection (2) |
| hsa04270 Vascular smooth muscle contraction (2) |
| hsa05146 Amoebiasis (2) |
| hsa05160 Hepatitis C (2) |
| hsa04670 Leukocyte transendothelial migration (2) |
| hsa04614 Renin-angiotensin system (1) |
| hsa00592 alpha-Linolenic acid metabolism (1) |

| KEGG Pathways Top P-value Genes Family 1168 |
| --- |
| hsa04060 Cytokine-cytokine receptor interaction (1) |
| hsa03430 Mismatch repair (1) |
| hsa04912 GnRH signaling pathway (1) |
| hsa04726 Serotonergic synapse (1) |
| hsa04062 Chemokine signaling pathway (1) |
| hsa03008 Ribosome biogenesis in eukaryotes (1) |
| hsa00590 Arachidonic acid metabolism (1) |
| hsa04976 Bile secretion (1) |
| hsa04120 Ubiquitin mediated proteolysis (1) |
| hsa00740 Riboflavin metabolism (1) |
| hsa05162 Measles (1) |
| hsa04914 Progesterone-mediated oocyte maturation (1) |
| hsa05222 Small cell lung cancer (1) |
| hsa04725 Cholinergic synapse (1) |
| hsa04970 Salivary secretion (1) |
| hsa04070 Phosphatidylinositol signaling system (1) |
| hsa04370 VEGF signaling pathway (1) |
| hsa05130 Pathogenic Escherichia coli infection (1) |
| hsa04964 Proximal tubule bicarbonate reclamation (1) |
| hsa00480 Glutathione metabolism (1) |
| hsa04975 Fat digestion and absorption (1) |
| hsa00591 Linoleic acid metabolism (1) |
| hsa04114 Oocyte meiosis (1) |
| hsa04310 Wnt signaling pathway (1) |
| hsa05132 Salmonella infection (1) |
| hsa05322 Systemic lupus erythematosus (1) |
| hsa05133 Pertussis (1) |
| hsa04973 Carbohydrate digestion and absorption (1) |
| hsa00982 Drug metabolism cytochrome P450 (1) |
| hsa00562 Inositol phosphate metabolism (1) |
| hsa05166 HTLV-I infection (1) |
| hsa00280 Valine, leucine and isoleucine degradation (1) |
| hsa05020 Prion diseases (1) |
| hsa04730 Long-term depression (1) |
| hsa00330 Arginine and proline metabolism (1) |
| hsa04350 TGF-beta signaling pathway (1) |
| hsa05332 Graft-versus-host disease (1) |
| hsa04660 T cell receptor signaling pathway (1) |
| hsa04620 Toll-like receptor signaling pathway (1) |
| hsa04971 Gastric acid secretion (1) |

| KEGG Pathways Top P-value Genes Family 1168 |
| --- |
| hsa00565 Ether lipid metabolism (1) |
| hsa04978 Mineral absorption (1) |
| hsa04110 Cell cycle (1) |
| hsa05320 Autoimmune thyroid disease (1) |
| hsa04623 Cytosolic DNA-sensing pathway (1) |
| hsa00563 Glycosylphosphatidylinositol(GPI)-anchor biosynthesis (1) |
| hsa04260 Cardiac muscle contraction (1) |
| hsa00140 Steroid hormone biosynthesis (1) |
| hsa04145 Phagosome (1) |
| hsa05220 Chronic myeloid leukemia (1) |
| hsa04210 Apoptosis (1) |
| hsa00970 Aminoacyl-tRNA biosynthesis (1) |
| hsa04010 MAPK signaling pathway (1) |
| hsa00564 Glycerophospholipid metabolism (1) |
| hsa04330 Notch signaling pathway (1) |
| hsa04612 Antigen processing and presentation (1) |
| hsa04320 Dorso-ventral axis formation (1) |
| hsa04724 Glutamatergic synapse (1) |
| hsa00230 Purine metabolism (1) |
| hsa05168 Herpes simplex infection (1) |
| hsa04020 Calcium signaling pathway (1) |
| hsa00980 Metabolism of xenobiotics by cytochrome P450 (1) |
| hsa04960 Aldosterone-regulated sodium reabsorption (1) |
| hsa05323 Rheumatoid arthritis (1) |
| hsa04961 Endocrine and other factor-regulated calcium reabsorption (1) |
| hsa00860 Porphyrin and chlorophyll metabolism (1) |

**Appendix M. Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis of most significant p-value VAAST genes for family 1281.** The table includes KEGG pathway ID, pathway name and number of genes in parentheses.

| KEGG Pathways Top P-value Genes Family 1281 |
|---|
| hsa04740 Olfactory transduction (20) |
| hsa04510 Focal adhesion (6) |
| hsa01100 Metabolic pathways (5) |
| hsa04141 Protein processing in endoplasmic reticulum (4) |
| hsa04610 Complement and coagulation cascades (4) |
| hsa04512 ECM-receptor interaction (4) |
| hsa04972 Pancreatic secretion (3) |
| hsa04650 Natural killer cell mediated cytotoxicity (3) |
| hsa05164 Influenza A (3) |
| hsa04612 Antigen processing and presentation (3) |
| hsa04974 Protein digestion and absorption (3) |
| hsa04810 Regulation of actin cytoskeleton (2) |
| hsa05145 Toxoplasmosis (2) |
| hsa04080 Neuroactive ligand-receptor interaction (2) |
| hsa04640 Hematopoietic cell lineage (2) |
| hsa05010 Alzheimer's disease (2) |
| hsa04370 VEGF signaling pathway (2) |
| hsa04664 Fc epsilon RI signaling pathway (2) |
| hsa05134 Legionellosis (2) |
| hsa05332 Graft-versus-host disease (2) |
| hsa04666 Fc gamma R-mediated phagocytosis (2) |
| hsa04360 Axon guidance (2) |
| hsa05146 Amoebiasis (2) |
| hsa00592 alpha-Linolenic acid metabolism (1) |
| hsa04060 Cytokine-cytokine receptor interaction (1) |
| hsa03430 Mismatch repair (1) |
| hsa04912 GnRH signaling pathway (1) |
| hsa04726 Serotonergic synapse (1) |
| hsa04062 Chemokine signaling pathway (1) |
| hsa03008 Ribosome biogenesis in eukaryotes (1) |
| hsa00590 Arachidonic acid metabolism (1) |
| hsa04976 Bile secretion (1) |
| hsa04120 Ubiquitin mediated proteolysis (1) |
| hsa00740 Riboflavin metabolism (1) |
| hsa05162 Measles (1) |

| KEGG Pathways Top P-value Genes Family 1281 |
| --- |
| hsa04530 Tight junction (1) |
| hsa04914 Progesterone-mediated oocyte maturation (1) |
| hsa05222 Small cell lung cancer (1) |
| hsa04725 Cholinergic synapse (1) |
| hsa04970 Salivary secretion (1) |
| hsa04070 Phosphatidylinositol signaling system (1) |
| hsa04662 B cell receptor signaling pathway (1) |
| hsa05130 Pathogenic Escherichia coli infection (1) |
| hsa04964 Proximal tubule bicarbonate reclamation (1) |
| hsa04975 Fat digestion and absorption (1) |
| hsa05140 Leishmaniasis (1) |
| hsa00591 Linoleic acid metabolism (1) |
| hsa04114 Oocyte meiosis (1) |
| hsa04380 Osteoclast differentiation (1) |
| hsa05132 Salmonella infection (1) |
| hsa05144 Malaria (1) |
| hsa04973 Carbohydrate digestion and absorption (1) |
| hsa00562 Inositol phosphate metabolism (1) |
| hsa05166 HTLV-I infection (1) |
| hsa00280 Valine, leucine and isoleucine degradation (1) |
| hsa05020 Prion diseases (1) |
| hsa04730 Long-term depression (1) |
| hsa00330 Arginine and proline metabolism (1) |
| hsa00310 Lysine degradation (1) |
| hsa04660 T cell receptor signaling pathway (1) |
| hsa04620 Toll-like receptor signaling pathway (1) |
| hsa00520 Amino sugar and nucleotide sugar metabolism (1) |
| hsa04971 Gastric acid secretion (1) |
| hsa04978 Mineral absorption (1) |
| hsa00565 Ether lipid metabolism (1) |
| hsa05152 Tuberculosis (1) |
| hsa05200 Pathways in cancer (1) |
| hsa04110 Cell cycle (1) |
| hsa05320 Autoimmune thyroid disease (1) |
| hsa04623 Cytosolic DNA-sensing pathway (1) |
| hsa04260 Cardiac muscle contraction (1) |
| hsa04210 Apoptosis (1) |
| hsa00970 Aminoacyl-tRNA biosynthesis (1) |
| hsa04010 MAPK signaling pathway (1) |
| hsa00564 Glycerophospholipid metabolism (1) |

| KEGG Pathways Top P-value Genes Family 1281 |
| --- |
| hsa04270 Vascular smooth muscle contraction (1) |
| hsa05150 Staphylococcus aureus infection (1) |
| hsa04724 Glutamatergic synapse (1) |
| hsa05168 Herpes simplex infection (1) |
| hsa04020 Calcium signaling pathway (1) |
| hsa05160 Hepatitis C (1) |
| hsa04670 Leukocyte transendothelial migration (1) |
| hsa04960 Aldosterone-regulated sodium reabsorption (1) |
| hsa05323 Rheumatoid arthritis (1) |
| hsa04961 Endocrine and other factor-regulated calcium reabsorption (1) |
| hsa00860 Porphyrin and chlorophyll metabolism (1) |

**Appendix N. Complete unadjusted and adjusted additive logistic regression association results for Finnish mothers' complement and coagulation factor exome SNPs on Illumina Omni2.5 BeadChips sorted by adjusted p-values**. The gene listed is that in which the SNP is located and in parentheses is the VAAST gene the SNP was selected to interrogate within the 10 kb 5′ and 3′ buffer. Base pair (BP) positions refer to GRCh37 (hg19, February 2009 assembly) build of the human genome.

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | ADJ P-VALUE | ADJ OR | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | rs6691117 | 207782931 | *CR1* | G | A | 1.07E-04 | 1.732 | 6.93E-05 | 1.741 |
| 1 | EXM133911 | 197031021 | *F13B* | C | T | 0.0213 | 0.5327 | 0.04428 | 0.5864 |
| 1 | EXM-RS12034383 | 207803595 | *CR1* | G | A | 0.0307 | 1.31 | 0.02107 | 1.326 |
| 1 | EXM-RS6656401 | 207692049 | *CR1* | A | G | 0.0504 | 0.7326 | 0.05833 | 0.7454 |
| 1 | EXM144910 | 207782769 | *CR1* | A | G | 0.0793 | 1.477 | 0.1053 | 1.423 |
| 1 | EXM-RS3818361 | 207784968 | *CR1* | T | C | 0.0927 | 0.7726 | 0.09383 | 0.7776 |
| 1 | EXM-RS6701713 | 207786289 | *CR1* | A | G | 0.0927 | 0.7726 | 0.09383 | 0.7776 |
| 1 | EXM144976 | 207795320 | *CR1* | A | G | 0.1011 | 0.7782 | 0.104 | 0.784 |
| 1 | EXM-RS3813948 | 207269858 | *C4BPA* | G | A | 0.1553 | 0.673 | 0.165 | 0.6842 |
| 1 | EXM-RS6677604 | 196686918 | *CFH* | A | G | 0.1601 | 0.752 | 0.1991 | 0.7769 |
| 1 | EXM144922 | 207782916 | *CR1* | A | T | 0.1796 | 0.493 | 0.127 | 0.4497 |
| 1 | EXM144743 | 207653395 | *CR2* | C | A | 0.1905 | 0.7905 | 0.1542 | 0.7775 |
| 1 | EXM-RS1061170 | 196659237 | *CFH* | C | T | 0.1964 | 0.8492 | 0.1775 | 0.8459 |
| 1 | EXM144486 | 207304980 | *C4BPA* | C | G | 0.2161 | 2.305 | 0.3081 | 1.987 |
| 1 | EXM121798 | 169510139 | *F5* | A | G | 0.2205 | 0.7313 | 0.1913 | 0.7219 |
| 1 | EXM121818 | 169510475 | *F5* | T | G | 0.2523 | 0.8353 | 0.3945 | 0.878 |
| 1 | EXM121977 | 169521853 | *F5* | G | A | 0.2732 | 0.7969 | 0.5509 | 0.8878 |
| 1 | EXM121943 | 169513583 | *F5* | T | G | 0.3048 | 0.8082 | 0.6067 | 0.9022 |
| 1 | EXM2250216 | 169513583 | *F5* | T | G | 0.3048 | 0.8082 | 0.6067 | 0.9022 |
| 1 | EXM-RS1329424 | 196646176 | *CFH* | A | C | 0.3288 | 0.8869 | 0.2994 | 0.8824 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | ADJ P-VALUE | ADJ OR | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | EXM133342 | 196642233 | *CFH* | A | G | 0.3511 | 1.135 | 0.2702 | 1.158 |
| 1 | EXM133469 | 196709816 | *CFH* | T | G | 0.3593 | 3.091 | 0.4045 | 2.783 |
| 1 | EXM144433 | 207286381 | *C4BPA* | A | C | 0.4025 | 1.58 | 0.5381 | 1.396 |
| 1 | EXM133467 | 196709774 | *CFH* | T | G | 0.4383 | 1.142 | 0.6694 | 1.074 |
| 1 | EXM122025 | 169541513 | *F5* | G | C | 0.4563 | 1.328 | 0.5909 | 1.224 |
| 1 | EXM121797 | 169510118 | *F5* | A | G | 0.4583 | 0.7704 | 0.4232 | 0.7584 |
| 1 | EXM122022 | 169529974 | *F5* | A | G | 0.4843 | 0.5536 | 0.4769 | 0.5498 |
| 1 | EXM-RS380390 | 196701051 | *CFH* | C | G | 0.4857 | 0.9182 | 0.4585 | 0.9149 |
| 1 | EXM144908 | 207782707 | *CR1* | G | A | 0.5277 | 1.139 | 0.4767 | 1.155 |
| 1 | EXM144887 | 207760772 | *CR1* | G | A | 0.5394 | 1.304 | 0.5605 | 1.282 |
| 1 | EXM121681 | 169483561 | *F5* | C | T | 0.5635 | 0.877 | 0.9626 | 1.01 |
| 1 | EXM122058 | 169563951 | *F5* | G | T | 0.5974 | 0.9059 | 0.5445 | 0.8951 |
| 1 | EXM121903 | 169511878 | *F5* | G | T | 0.5975 | 0.8857 | 0.9166 | 1.023 |
| 1 | EXM-RS17045328 | 207652176 | *CR2* | G | A | 0.6030 | 1.183 | 0.5552 | 1.206 |
| 1 | EXM144740 | 207653364 | *CR2* | G | A | 0.6062 | 0.8743 | 0.4428 | 0.8213 |
| 1 | EXM144473 | 207300070 | *C4BPA* | A | G | 0.6458 | 0.8845 | 0.5855 | 0.8667 |
| 1 | EXM144961 | 207791434 | *CR1* | G | A | 0.7044 | 1.218 | 0.887 | 1.075 |
| 1 | EXM121964 | 169519112 | *F5* | T | C | 0.7311 | 1.115 | 0.9299 | 1.027 |
| 1 | EXM144679 | 207646898 | *CR2* | C | T | 0.7341 | 0.95 | 0.5762 | 0.9208 |
| 1 | EXM144874 | 207755285 | *CR1* | A | G | 0.7434 | 0.7922 | 0.8784 | 1.109 |
| 1 | EXM144675 | 207646462 | *CR2* | A | G | 0.7572 | 0.9545 | 0.5927 | 0.9241 |
| 1 | EXM144681 | 207646923 | *CR2* | A | G | 0.7579 | 0.9545 | 0.5992 | 0.9254 |
| 1 | EXM133877 | 197026289 | *F13B* | G | A | 0.8023 | 0.7329 | 0.7631 | 0.6907 |
| 1 | EXM-RS10737680 | 196679455 | *CFH* | C | A | 0.8453 | 1.025 | 0.6321 | 1.062 |
| 1 | EXM-RS1329428 | 196702810 | *CFH* | A | G | 0.8453 | 1.025 | 0.6321 | 1.062 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | ADJ P-VALUE | ADJ OR | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|---|---|
| 1 | EXM-RS1410996 | 196696933 | *CFH* | T | C | 0.8453 | 1.025 | 0.6321 | 1.062 |
| 1 | EXM122075 | 169565346 | *F5* | A | C | 0.8731 | 0.9642 | 0.9244 | 0.9793 |
| 1 | EXM121879 | 169511555 | *F5* | C | T | 0.8767 | 1.023 | 0.7438 | 0.9545 |
| 1 | EXM121894 | 169511734 | *F5* | C | T | 0.8792 | 1.022 | 0.7491 | 0.9556 |
| 1 | EXM121896 | 169511755 | *F5* | C | T | 0.9022 | 1.018 | 0.7251 | 0.9513 |
| 1 | EXM144655 | 207644786 | *CR2* | A | T | 0.9062 | 0.8651 | 0.7631 | 0.6907 |
| 1 | EXM121743 | 169498975 | *F5* | C | T | 0.9359 | 0.9892 | 0.8742 | 0.9793 |
| 1 | EXM133490 | 196712596 | *CFH* | T | A | 0.9514 | 0.9559 | 0.7977 | 0.8282 |
| 1 | EXM121985 | 169524517 | *F5* | A | G | 0.9666 | 0.9682 | 0.6435 | 1.391 |
| 1 | EXM144888 | 207760773 | *CR1* | T | C | 0.9814 | 1.003 | 0.996 | 0.9993 |
| 1 | EXM2259849 | 169527856 | *F5* | C | T | 0.9866 | 1.002 | 0.8072 | 1.033 |
| 1 | EXM144899 | 207762095 | *CR1* | T | C | 0.9879 | 0.9815 | 0.7631 | 0.6907 |
| 1 | EXM121732 | 169497306 | *F5* | A | G | 0.9993 | 9.495E-10 | 0.9993 | 8.541E-10 |
| 1 | EXM121746 | 169499020 | *F5* | C | G | 0.9993 | 1.006E-09 | 0.9993 | 8.541E-10 |
| 1 | EXM121969 | 169519883 | *F5* | A | G | 0.9993 | 2.89E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM122036 | 169555582 | *F5* | T | C | 0.9993 | 7.867E-10 | 0.9993 | 8.515E-10 |
| 1 | EXM133460 | 196706677 | *CFH* | T | G | 0.9993 | 1.58E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM133479 | 196711067 | *CFH* | T | G | 0.9993 | 1.58E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM133865 | 197024914 | *F13B* | G | A | 0.9993 | 2.48E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM144467 | 207297622 | *C4BPA* | A | G | 0.9993 | 1.67E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM144499 | 207307932 | *C4BPA* | A | G | 0.9993 | 2.22E+09 | 0.9993 | 2.25E+09 |
| 1 | EXM144639 | 207643432 | *CR2* | G | C | 0.9993 | 9.624E-10 | 0.9993 | 8.541E-10 |

**Appendix O. Complete unadjusted additive logistic association results for Finnish mothers' Affymetrix 6.0 SNP arrays sorted by p-value.** The gene listed in the gene column is the gene where the SNP is located. The gene listed is the gene in which the SNP is located and in parentheses is the VAAST gene the SNP was selected to interrogate within the 10kb 5' and 3' buffer. Base pair (BP) positions refer to NCBI36 (hg18, March 2006 assembly) build of the human genome.

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|
| 1 | rs10429953 | 205855201 | *CR1* | G | A | 1.31E-04 | 1.931 |
| 1 | rs10429943 | 205856094 | *CR1* | T | C | 3.74E-04 | 1.841 |
| 1 | rs11118166 | 205848618 | *CR1* | G | A | 5.78E-04 | 1.776 |
| 1 | rs12567990 | 205748308 | *CR1* | T | C | 5.82E-04 | 1.791 |
| 1 | rs12141045 | 205838777 | *CR1* | G | A | 6.22E-04 | 1.776 |
| 1 | rs599948 | 205818862 | *CR1* | C | T | 7.04E-04 | 1.664 |
| 1 | rs677066 | 205840614 | *CR1* | C | T | 1.70E-03 | 1.591 |
| 1 | rs3753305 | 167820682 | *F5* | G | C | 5.52E-03 | 1.422 |
| 1 | rs9943077 | 205361921 | *C4BPA* | T | C | 0.02323 | 0.7266 |
| 1 | rs2491393 | 205366882 | *C4BPA* | G | A | 0.03275 | 1.335 |
| 1 | rs10489185 | 167815516 | *F5* | A | C | 0.03766 | 0.757 |
| 1 | rs12080578 | 205887235 | *CR1L (CR1)* | G | A | 0.04018 | 1.525 |
| 1 | rs17522707 | 167829686 | *SELP (F5)* | A | G | 0.04466 | 0.5847 |
| 1 | rs4844573 | 205371523 | *C4BPA* | C | T | 0.05165 | 1.28 |
| 1 | rs2213873 | 167810401 | *F5* | A | G | 0.05722 | 0.7711 |
| 1 | rs10733086 | 194943558 | *CFH* | A | T | 0.06023 | 0.7854 |
| 1 | rs2227245 | 167806704 | *F5* | T | C | 0.06292 | 0.7785 |
| 1 | rs572515 | 194912884 | *CFH* | A | G | 0.06589 | 0.7914 |
| 1 | rs6691048 | 167808759 | *F5* | T | C | 0.0678 | 0.7857 |
| 1 | rs4403634 | 205334714 | *C4BPB* | A | C | 0.06793 | 0.797 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|
| 1 | rs6427195 | 167747800 | *F5* | T | A | 0.07488 | 1.609 |
| 1 | rs12406092 | 167809755 | *F5* | T | C | 0.07674 | 0.7934 |
| 1 | rs7551623 | 167740978 | *F5* | A | G | 0.08331 | 0.6452 |
| 1 | rs2491395 | 205368346 | *C4BPA* | T | A | 0.08607 | 1.264 |
| 1 | rs12032512 | 205719801 | *CR2* | C | G | 0.09646 | 0.8065 |
| 1 | rs6695321 | 194942484 | *CFH* | G | A | 0.1016 | 1.239 |
| 1 | rs9332647 | 167758724 | *F5* | G | A | 0.1105 | 0.6706 |
| 1 | rs11118242 | 205886936 | *CR1L (CR1)* | T | C | 0.1129 | 0.8181 |
| 1 | rs2274566 | 205819968 | *CR1* | G | A | 0.1181 | 1.227 |
| 1 | rs1018827 | 167780630 | *F5* | T | C | 0.1195 | 1.501 |
| 1 | rs4618971 | 205732684 | *CR1* | C | T | 0.1263 | 0.8011 |
| 1 | rs6662176 | 167810667 | *F5* | T | A | 0.1271 | 0.8195 |
| 1 | rs6427197 | 167767214 | *F5* | G | T | 0.1272 | 1.484 |
| 1 | rs9429781 | 205733845 | *CR1* | G | T | 0.1343 | 0.7974 |
| 1 | rs7542088 | 167802864 | *F5* | A | C | 0.1538 | 0.8256 |
| 1 | rs1329423 | 194913010 | *CFH* | G | A | 0.1547 | 1.258 |
| 1 | rs17044576 | 205701293 | *CR2* | G | A | 0.1567 | 1.422 |
| 1 | rs10801555 | 194926884 | *CFH* | A | G | 0.1568 | 0.8273 |
| 1 | rs514943 | 194930536 | *CFH* | G | A | 0.1569 | 0.836 |
| 1 | rs6428357 | 194942194 | *CFH* | T | C | 0.1654 | 0.8365 |
| 1 | rs11803956 | 205869644 | *CR1* | T | C | 0.1693 | 1.187 |
| 1 | rs10779340 | 205883648 | *CR1* | C | T | 0.1704 | 1.187 |
| 1 | rs1831282 | 194940616 | *CFH* | A | C | 0.1724 | 0.8445 |
| 1 | rs395544 | 194964895 | *CFH* | A | G | 0.1894 | 0.8456 |
| 1 | rs2182911 | 205726694 | *CR2* | G | A | 0.203 | 0.8308 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|
| 1 | rs1759007 | 195294229 | *F13B* | T | C | 0.2242 | 0.7253 |
| 1 | rs10801556 | 194927087 | *CFH* | T | C | 0.2281 | 0.8539 |
| 1 | rs17258982 | 205719987 | *CR2* | C | T | 0.2337 | 0.7363 |
| 1 | rs6690215 | 205722673 | *CR2* | C | T | 0.2468 | 0.86 |
| 1 | rs4844597 | 205737892 | *CR1* | C | T | 0.2732 | 0.8503 |
| 1 | rs742855 | 194972143 | *CFH* | G | A | 0.2843 | 1.212 |
| 1 | rs7518773 | 195278259 | *F13B* | A | G | 0.2971 | 1.137 |
| 1 | rs10737680 | 194946078 | *CFH* | G | T | 0.2977 | 1.146 |
| 1 | rs974793 | 167745278 | *F5* | T | C | 0.2998 | 1.172 |
| 1 | rs203687 | 194940893 | *CFH* | C | T | 0.3151 | 0.8774 |
| 1 | rs1759008 | 195293518 | *F13B* | T | G | 0.3161 | 0.7802 |
| 1 | rs3917843 | 167826881 | *SELP (F5)* | T | C | 0.3344 | 0.8381 |
| 1 | rs9332627 | 167764444 | *F5* | T | C | 0.3483 | 1.153 |
| 1 | rs3820060 | 167751176 | *F5* | G | T | 0.3557 | 1.134 |
| 1 | rs403846 | 194963360 | *CFH* | A | G | 0.3581 | 0.8845 |
| 1 | rs1627765 | 195299666 | *F13B* | C | T | 0.371 | 0.8008 |
| 1 | rs424535 | 194975846 | *CFH* | T | A | 0.3888 | 1.124 |
| 1 | rs10754209 | 195278200 | *F13B* | A | T | 0.3949 | 1.111 |
| 1 | rs4915148 | 195302161 | *F13B* | T | C | 0.437 | 0.8258 |
| 1 | rs857021 | 195274102 | *F13B* | A | G | 0.437 | 0.8258 |
| 1 | rs2182913 | 205727001 | *CR2* | T | C | 0.4498 | 0.9044 |
| 1 | rs3849266 | 205819613 | *CR1* | T | C | 0.4527 | 1.112 |
| 1 | rs1571344 | 205737551 | *CR1* | G | A | 0.4535 | 1.185 |
| 1 | rs3917820 | 167831369 | *SELP (F5)* | T | C | 0.4566 | 1.172 |
| 1 | rs4656685 | 167750468 | *F5* | A | G | 0.4634 | 1.123 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|
| 1 | rs1759009 | 195293295 | *F13B* | T | C | 0.5075 | 0.847 |
| 1 | rs3766104 | 167755647 | *F5* | T | A | 0.5089 | 0.8546 |
| 1 | rs17020993 | 205355015 | *C4BPA* | G | A | 0.5116 | 1.145 |
| 1 | rs7519408 | 205727912 | *CR2* | C | G | 0.5476 | 0.9042 |
| 1 | rs3917824 | 167831206 | *SELP (F5)* | G | C | 0.5669 | 1.13 |
| 1 | rs3917819 | 167831513 | *SELP (F5)* | G | A | 0.598 | 1.13 |
| 1 | rs9332600 | 167779537 | *F5* | T | C | 0.6052 | 1.077 |
| 1 | rs8942 | 205336542 | *C4BPB* | A | G | 0.6097 | 1.113 |
| 1 | rs6128 | 167829528 | *SELP (F5)* | A | G | 0.6377 | 0.9326 |
| 1 | rs9332619 | 167766972 | *F5* | T | C | 0.6544 | 1.069 |
| 1 | rs9332661 | 167755827 | *F5* | C | T | 0.666 | 1.101 |
| 1 | rs6661764 | 205875008 | *CR1* | C | G | 0.7046 | 1.05 |
| 1 | rs4524 | 167778379 | *F5* | G | A | 0.7397 | 1.049 |
| 1 | rs7527218 | 205690452 | *CR2* | C | T | 0.7508 | 1.041 |
| 1 | rs6662593 | 167779218 | *F5* | T | C | 0.7594 | 1.045 |
| 1 | rs3766110 | 167781807 | *F5* | C | A | 0.7746 | 0.9597 |
| 1 | rs1986158 | 205737367 | *CR1* | G | A | 0.7833 | 1.065 |
| 1 | rs6029 | 167796597 | *F5* | A | G | 0.7891 | 0.9636 |
| 1 | rs12025910 | 167745405 | *F5* | C | T | 0.797 | 1.059 |
| 1 | rs17615 | 205713085 | *CR2* | A | G | 0.8337 | 0.9697 |
| 1 | rs9332618 | 167767105 | *F5* | T | C | 0.8377 | 0.9661 |
| 1 | rs11120211 | 205343002 | *C4BPA* | A | G | 0.8424 | 1.054 |
| 1 | rs2019727 | 194941337 | *CFH* | T | A | 0.8471 | 1.037 |
| 1 | rs2940253 | 205712366 | *CR2* | C | G | 0.8611 | 0.9746 |
| 1 | rs9332595 | 167780979 | *F5* | C | G | 0.8669 | 0.9767 |

| CHR | SNP | BP | GENE | MINOR ALLELE | MAJOR ALLELE | UNADJ P-VALUE | UNADJ OR |
|---|---|---|---|---|---|---|---|
| 1 | rs11120218 | 205345074 | *C4BPA* | T | C | 0.8718 | 1.032 |
| 1 | rs9332640 | 167760350 | *F5* | G | C | 0.8719 | 0.9804 |
| 1 | rs511678 | 205711905 | *CR2* | C | G | 0.8976 | 0.981 |
| 1 | rs1410408 | 205732931 | *CR1* | T | C | 0.9113 | 0.9738 |
| 1 | rs311311 | 205706318 | *CR2* | C | G | 0.9124 | 0.984 |
| 1 | rs6015 | 167786518 | *F5* | T | C | 0.9447 | 1.014 |
| 1 | rs9429774 | 205712518 | *CR2* | T | C | 0.9519 | 0.9912 |
| 1 | rs916438 | 167766283 | *F5* | A | T | 0.9561 | 0.993 |

## REFERENCES


1 Eds CP Howson, M. K., JE Lawn. March of Dimes, PMNCH, Save the Children, WHO. Born Too Soon: The Global Action Report on Preterm Birth. (World Health Organization, 2012).

2 Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75-84, doi:10.1016/S0140-6736(08)60074-4 (2008).

3 Jobe, A. H. & Bancalari, E. Bronchopulmonary dysplasia. *American Journal of Respiratory and Critical Care Medicine* **163**, 1723-1729 (2001).

4 Rees, S. & Inder, T. Fetal and neonatal origins of altered brain development. *Early Hum Dev* **81**, 753-761, doi:10.1016/j.earlhumdev.2005.07.004 (2005).

5 Ananth, C. V. & Vintzileos, A. M. Epidemiology of preterm birth and its clinical subtypes. *J Matern Fetal Neonatal Med* **19**, 773-782, doi:10.1080/14767050600965882 (2006).

6 Green, N. S. *et al.* Research agenda for preterm birth: recommendations from the March of Dimes. *Am J Obstet Gynecol* **193**, 626-635, doi:10.1016/j.ajog.2005.02.106 (2005).

7 Murray, C. J. & Lopez, A. D. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* **349**, 1269-1276, doi:10.1016/S0140-6736(96)07493-4 (1997).

8 Gilbert, W. M., Nesbitt, T. S. & Danielsen, B. The cost of prematurity: quantification by gestational age and birth weight. *Obstet Gynecol* **102**, 488-492 (2003).

9 Petrou, S. The economic consequences of preterm birth during the first 10 years of life. *BJOG* **112 Suppl 1**, 10-15, doi:10.1111/j.1471-0528.2005.00577.x (2005).

10 Preterm birth: causes, consequences and prevention. (Institute of Medicine, 2006).

11 Muglia, L. J. & Katz, M. The enigma of spontaneous preterm birth. *N Engl J Med* **362**, 529-535, doi:10.1056/NEJMra0904308 (2010).

12 Martin, J. *et al.* Births: Final data for 2006. *Public Health Resources*, 65 (2009).

13 Lockwood, C. J. & Kuczynski, E. Risk stratification and pathological mechanisms in preterm delivery. *Paediatr Perinat Epidemiol* **15 Suppl 2**, 78-89 (2001).

14 Liggins, G. C., Fairclough, R. J., Grieves, S. A., Forster, C. S. & Knox, B. S. Parturition in the sheep. *Ciba Found Symp*, 5-30 (1977).

15 Sfakianaki, A. K. & Norwitz, E. R. Mechanisms of progesterone action in inhibiting prematurity. *J Matern Fetal Neonatal Med* **19**, 763-772, doi:10.1080/14767050600949829 (2006).

16 Chaudhari, B. P. *et al.* The genetics of birth timing: insights into a fundamental component of human development. *Clin Genet* **74**, 493-501, doi:10.1111/j.1399-0004.2008.01124.x (2008).

17 JRG, C., Matthews, S. G., Gibb, W. & Lye, S. J. Endocrine and paracrine regulation of birth at term and preterm. *Endocr Rev* **21**, 514-550 (2000).

18 Mendelson, C. R. Minireview: fetal-maternal hormonal signaling in pregnancy and labor. *Mol Endocrinol* **23**, 947-954, doi:10.1210/me.2009-0016 (2009).

19 Merlino, A. A. *et al.* Nuclear progesterone receptors in the human pregnancy myometrium: evidence that parturition involves functional progesterone withdrawal mediated by increased expression of progesterone receptor-A. *J Clin Endocrinol Metab* **92**, 1927-1933, doi:10.1210/jc.2007-0077 (2007).

20 Chen, C. *et al.* The human progesterone receptor shows evidence of adaptive evolution associated with its ability to act as a transcription factor. *Mol Phylogenet Evol* **47**, 637-649, doi:10.1016/j.ympev.2007.12.026 (2008).

21 Velez, D. R. *et al.* Ethnic differences in interleukin 6 (IL-6) and IL6 receptor genes in spontaneous preterm birth and effects on amniotic fluid protein levels. *Ann Hum Genet* **71**, 586-600, doi:10.1111/j.1469-1809.2007.00352.x (2007).

22 Burris, H. H., Collins, J. W. & Wright, R. O. Racial/ethnic disparities in preterm birth: clues from environmental exposures. *Current Opinion in Pediatrics*, doi:10.1097/MOP.0b013e328344568f (2011).

23 Burris, H. H. & Collins, J. W. Race and preterm birth--the case for epigenetic inquiry. *Ethnicity & disease* **20**, 296-299 (2010).

24 Dunlop, A. L., KRAMER, M., HOGUE, C. J., Menon, R. & RAMAKRISHAN, U. Racial disparities in preterm birth: an overview of the potential role of nutrient deficiencies. *Acta obstetricia et gynecologica Scandinavica*, doi:10.1111/j.1600-0412.2011.01274.x (2011).

25 Menon, R., Dunlop, A. L., Kramer, M. R., Fortunato, S. J. & Hogue, C. J. An overview of racial disparities in preterm birth rates: caused by infection or inflammatory response? *Acta obstetricia et gynecologica Scandinavica*, no-no, doi:10.1111/j.1600-0412.2011.01135.x (2011).

26 Menon, R. *et al.* Racial disparity in pathophysiologic pathways of preterm birth based on genetic variants. *Reproductive biology and endocrinology : RB&E* **7**, 62, doi:10.1186/1477-7827-7-62 (2009).

27 Brett, K. M., Strogatz, D. S. & Savitz, D. A. Employment, job strain, and preterm delivery among women in North Carolina. *American Journal of Public Health* **87**, 199-204 (1997).

28 Smith, L. K., Draper, E. S., Manktelow, B. N., Dorling, J. S. & Field, D. J. Socioeconomic inequalities in very preterm birth rates. *Arch Dis Child Fetal Neonatal Ed* **92**, F11-14, doi:10.1136/adc.2005.090308 (2007).

29 Thompson, J. M. D., Irgens, L. M., Rasmussen, S. & Daltveit, A. K. Secular trends in socio-economic status and the implications for preterm birth. *Paediatr Perinat Epidemiol* **20**, 182-187, doi:10.1111/j.1365-3016.2006.00711.x (2006).

30 Mueller-Heubach, E., Rubinstein, D. N. & Schwarz, S. S. Histologic chorioamnionitis and preterm delivery in different patient populations. *Obstet Gynecol* **75**, 622-626 (1990).

31 Nygren, P. *et al.* Evidence on the benefits and harms of screening and treating pregnant women who are asymptomatic for bacterial vaginosis: an update review for the U.S. Preventive Services Task Force. *Ann Intern Med* **148**, 220-233 (2008).

32 Okun, N., Gronau, K. A. & Hannah, M. E. Antibiotics for bacterial vaginosis or Trichomonas vaginalis in pregnancy: a systematic review. *Obstet Gynecol* **105**, 857-868, doi:10.1097/01.AOG.0000157108.32059.8f (2005).

33 Michalowicz, B. S. *et al.* Treatment of periodontal disease and the risk of preterm birth. *N Engl J Med* **355**, 1885-1894, doi:10.1056/NEJMoa062249 (2006).

34 Carey, J. C. *et al.* Metronidazole to prevent preterm delivery in pregnant women with asymptomatic bacterial vaginosis. National Institute of Child Health and Human Development Network of Maternal-Fetal Medicine Units. *N Engl J Med* **342**, 534-540, doi:10.1056/NEJM200002243420802 (2000).

35 Klebanoff, M. A. *et al.* Failure of metronidazole to prevent preterm delivery among pregnant women with asymptomatic Trichomonas vaginalis infection. *N Engl J Med* **345**, 487-493, doi:10.1056/NEJMoa003329 (2001).

36 Hendler, I. *et al.* The Preterm Prediction Study: association between maternal body mass index and spontaneous and indicated preterm birth. *Am J Obstet Gynecol* **192**, 882-886, doi:10.1016/j.ajog.2004.09.021 (2005).

37 Scholl, T. O. Iron status during pregnancy: setting the stage for mother and infant. *Am J Clin Nutr* **81**, 1218S-1222S (2005).

38 Scholl, T. O., Hediger, M. L., Schall, J. I., Fischer, R. L. & Khoo, C. S. Low zinc intake during pregnancy: its association with preterm and very preterm delivery. *Am J Epidemiol* **137**, 1115-1124 (1993).

39 Tamura, T. *et al.* Maternal serum folate and zinc concentrations and their relationships to pregnancy outcome. *Am J Clin Nutr* **56**, 365-370 (1992).

40 Lobel, M., Dunkel-Schetter, C. & Scrimshaw, S. C. Prenatal maternal stress and prematurity: a prospective study of socioeconomically disadvantaged women. *Health Psychol* **11**, 32-40 (1992).

41 Copper, R. L. *et al.* The preterm prediction study: maternal stress is associated with spontaneous preterm birth at less than thirty-five weeks' gestation. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. *Am J Obstet Gynecol* **175**, 1286-1292 (1996).

42 Andres, R. L. & Day, M. C. Perinatal complications associated with maternal tobacco use. *Semin Neonatol* **5**, 231-241, doi:10.1053/siny.2000.0025 (2000).

43 Ebrahim, S. H., Floyd, R. L., Merritt, R. K., Decoufle, P. & Holtzman, D. Trends in pregnancy-related smoking rates in the United States, 1987-1996. *JAMA* **283**, 361-366 (2000).

44 Cnattingius, S. The epidemiology of smoking during pregnancy: smoking prevalence, maternal characteristics, and pregnancy outcomes. *Nicotine Tob Res* **6 Suppl 2**, S125-140, doi:10.1080/14622200410001669187 (2004).

45 Conde-Agudelo, A., Rosas-Bermúdez, A. & Kafury-Goeta, A. C. Birth spacing and risk of adverse perinatal outcomes: a meta-analysis. *JAMA* **295**, 1809-1823, doi:10.1001/jama.295.15.1809 (2006).

46 Smith, G. C. S., Pell, J. P. & Dobbie, R. Interpregnancy interval and risk of preterm birth and neonatal death: retrospective cohort study. *BMJ* **327**, 313, doi:10.1136/bmj.327.7410.313 (2003).

47 Bakketeig, L. S., Hoffman, H. J. & Harley, E. E. The tendency to repeat gestational age and birth weight in successive births. *Am J Obstet Gynecol* **135**, 1086-1103 (1979).

48 Carr-Hill, R. A. & Hall, M. H. The repetition of spontaneous preterm labour. *Br J Obstet Gynaecol* **92**, 921-928 (1985).

49 Porter, T. F., Fraser, A. M., Hunter, C. Y., Ward, R. H. & Varner, M. W. The risk of preterm birth across generations. *Obstet Gynecol* **90**, 63-67, doi:10.1016/S0029-7844(97)00215-9 (1997).

50 Winkvist, A., Mogren, I. & Högberg, U. Familial patterns in birth characteristics: impact on individual and population risks. *Int J Epidemiol* **27**, 248-254 (1998).

51 Kistka, Z. A.-F. *et al.* Heritability of parturition timing: an extended twin design analysis. *Am J Obstet Gynecol* **199**, 43.e41-45, doi:10.1016/j.ajog.2007.12.014 (2008).

52 Boyd, H. A. *et al.* Maternal contributions to preterm delivery. *Am J Epidemiol* **170**, 1358-1364, doi:10.1093/aje/kwp324 (2009).

53 Plunkett, J. *et al.* Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. *Hum Hered* **68**, 209-219, doi:10.1159/000224641 (2009).

54 Bloom, S. L., Yost, N. P., McIntire, D. D. & Leveno, K. J. Recurrence of preterm birth in singleton and twin pregnancies. *Obstet Gynecol* **98**, 379-385 (2001).

55 Kistka, Z. A.-F. *et al.* Risk for postterm delivery after previous postterm delivery. *Am J Obstet Gynecol* **196**, 241.e241-246, doi:10.1016/j.ajog.2006.10.873 (2007).

56 Melve, K. K., Skjaerven, R., Gjessing, H. K. & Oyen, N. Recurrence of gestational age in sibships: implications for perinatal mortality. *Am J Epidemiol* **150**, 756-762 (1999).

57 Ward, K., Argyle, V., Meade, M. & Nelson, L. The heritability of preterm delivery. *Obstet Gynecol* **106**, 1235-1239, doi:10.1097/01.AOG.0000189091.35982.85 (2005).

58 Romero, R. *et al.* Identification of fetal and maternal single nucleotide polymorphisms in candidate genes that predispose to spontaneous preterm labor with intact membranes. *Am J Obstet Gynecol* **202**, 431.e431-434, doi:10.1016/j.ajog.2010.03.026 (2010).

59 Crider, K. S., Whitehead, N. & Buus, R. M. Genetic variation associated with preterm birth: a HuGE review. *Genet Med* **7**, 593-604 (2005).

60 Plunkett, J. & Muglia, L. J. Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. *Ann Med* **40**, 167-195, doi:10.1080/07853890701806181 (2008).

61 Volkov, N., Nisenblat, V., Ohel, G. & Gonen, R. Ehlers-Danlos syndrome: insights on obstetric aspects. *Obstetrical & gynecological survey* **62**, 51-57, doi:10.1097/01.ogx.0000251027.32142.63 (2007).

62 Ventura, S. Births: Preliminary Data for 2010. *NCHS data brief* (2012).

63 Esplin, M. S. Preterm birth: a review of genetic factors and future directions for genetic study. *Obstetrical & gynecological survey* **61**, 800-806, doi:10.1097/01.ogx.0000248747.52343.5f (2006).

64 Smith, R. Parturition. *N Engl J Med* **356**, 271-283, doi:10.1056/NEJMra061360 (2007).

65 Manolio, T. A., Brooks, L. D. & Collins, F. S. A HapMap harvest of insights into the genetics of common disease. *J Clin Invest* **118**, 1590-1605, doi:10.1172/JCI34772 (2008).

66 Manolio, T. A. Genomewide association studies and assessment of the risk of disease. *The New England journal of medicine* **363**, 166-176, doi:10.1056/NEJMra0905980 (2010).

67 Yang, J., Wray, N. R. & Visscher, P. M. Comparing apples and oranges: equating the power of case-control and quantitative trait association studies. *Genetic Epidemiology* **34**, 254-257, doi:10.1002/gepi.20456 (2010).

68 Hashimoto, M. *et al.* Isolation and localization of type IIb Na/Pi cotransporter in the developing rat lung. *The American journal of pathology* **157**, 21-27, doi:10.1016/S0002-9440(10)64512-9 (2000).

69 Traebert, M., Hattenhauer, O., Murer, H., Kaissling, B. & Biber, J. Expression of type II Na-P(i) cotransporter in alveolar type II cells. *The American journal of physiology* **277**, L868-873 (1999).

70 Wu, C. *et al.* BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biology* **10**, R130, doi:10.1186/gb-2009-10-11-r130 (2009).

71 Su, A. I. *et al.* A gene atlas of the mouse and human protein-encoding transcriptomes. *Proceedings of the National Academy of Sciences of the United States of America* **101**, 6062-6067, doi:10.1073/pnas.0400782101 (2004).

72 López Bernal, A., Newman, G. E., Phizackerley, P. J. & Turnbull, A. C. Surfactant stimulates prostaglandin E production in human amnion. *British journal of obstetrics and gynaecology* **95**, 1013-1017 (1988).

73 Mendelson, C. R. & Condon, J. C. New insights into the molecular endocrinology of parturition. *The Journal of steroid biochemistry and molecular biology* **93**, 113-119, doi:10.1016/j.jsbmb.2004.12.027 (2005).

74 Croteau-Chonka, D. C. *et al.* Genome-Wide Association Study of Anthropometric Traits and Evidence of Interactions With Age and Study Year in Filipino Women. *Obesity* **19**, 1019-1027, doi:10.1038/oby.2010.256 (2009).

75 Gamazon, E. R. *et al.* SCAN: SNP and copy number annotation. *Bioinformatics (Oxford, England)* **26**, 259-262, doi:10.1093/bioinformatics/btp644 (2010).

76 Martin, S. *et al.* Stickler syndrome: further mutations in COL11A1 and evidence for additional locus heterogeneity. *European Journal of Human Genetics* **7**, 807-814, doi:10.1038/sj.ejhg.5200377 (1999).

77 Richards, A. J. *et al.* Stickler syndrome and the vitreous phenotype: mutations in COL2A1 and COL11A1. *Human mutation* **31**, E1461-1471, doi:10.1002/humu.21257 (2010).

78 Richards, A. J. *et al.* A family with Stickler syndrome type 2 has a mutation in the COL11A1 gene resulting in the substitution of glycine 97 by valine in alpha 1 (XI) collagen. *Human Molecular Genetics* **5**, 1339-1343 (1996).

79 Annunen, S. *et al.* Splicing mutations of 54-bp exons in the COL11A1 gene cause Marshall syndrome, but other mutations cause overlapping Marshall/Stickler

phenotypes. *American journal of human genetics* **65**, 974-983, doi:10.1086/302585 (1999).

80 Griffith, A. J., Gebarski, S. S., Shepard, N. T. & Kileny, P. R. Audiovestibular phenotype associated with a COL11A1 mutation in Marshall syndrome. *Archives of otolaryngology--head & neck surgery* **126**, 891-894 (2000).

81 Griffith, A. J. *et al.* Marshall syndrome associated with a splicing defect at the COL11A1 locus. *American journal of human genetics* **62**, 816-823, doi:10.1086/301789 (1998).

82 Shanske, A., Bogdanow, A., Shprintzen, R. J. & Marion, R. W. Marshall syndrome and a defect at the COL11A1 locus. *American journal of human genetics* **63**, 1558-1561, doi:10.1086/302110 (1998).

83 Mio, F. *et al.* A functional polymorphism in COL11A1, which encodes the alpha 1 chain of type XI collagen, is associated with susceptibility to lumbar disc herniation. *American journal of human genetics* **81**, 1271-1277, doi:10.1086/522377 (2007).

84 Conte, N. *et al.* TACC1-chTOG-Aurora A protein complex in breast cancer. *Oncogene* **22**, 8102-8116, doi:10.1038/sj.onc.1206972 (2003).

85 Ghayad, S. E. *et al.* Identification of TACC1, NOV, and PTTG1 as new candidate genes associated with endocrine therapy resistance in breast cancer. *Journal of molecular endocrinology* **42**, 87-103, doi:10.1677/JME-08-0076 (2009).

86 Still, I. H., Hamilton, M., Vince, P., Wolfman, A. & Cowell, J. K. Cloning of TACC1, an embryonically expressed, potentially transforming coiled coil containing gene, from the 8p11 breast cancer amplicon. *Oncogene* **18**, 4032-4038, doi:10.1038/sj.onc.1202801 (1999).

87 Lauffart, B. *et al.* Aberrations of TACC1 and TACC3 are associated with ovarian cancer. *BMC women's health* **5**, 8, doi:10.1186/1472-6874-5-8 (2005).

88 Line, A., Slucka, Z., Stengrevics, A., Li, G. & Rees, R. C. Altered splicing pattern of TACC1 mRNA in gastric cancer. *Cancer genetics and cytogenetics* **139**, 78-83 (2002).

89 Jin, T. & Liu, L. The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. *Molecular endocrinology (Baltimore, Md)* **22**, 2383-2392, doi:10.1210/me.2008-0135 (2008).

90 Burton, P. R. *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678, doi:10.1038/nature05911 (2007).

91 Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, L. U., and Novartis Institutes of BioMedical Research *et al.* Genome-Wide Association

Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels. *Science (New York, NY)* **316**, 1331-1336, doi:10.1126/science.1142358 (2007).

92 Perry, J. R. B. *et al.* Stratifying Type 2 Diabetes Cases by BMI Identifies Genetic Risk Variants in LAMA1 and Enrichment for Risk Variants in Lean Compared to Obese Cases. *PLoS Genetics* **8**, e1002741, doi:10.1371/journal.pgen.1002741.t003 (2012).

93 Rung, J. *et al.* Genetic variant near IRS1 is associated with type 2 diabetes, insulin resistance and hyperinsulinemia. *Nature Genetics* **41**, 1110-1115, doi:10.1038/ng.443 (2009).

94 Salonen, J. T. *et al.* Type 2 Diabetes Whole-Genome Association Study in Four Populations: The DiaGen Consortium. *The American Journal of Human Genetics* **81**, 338-345, doi:10.1086/520599 (2007).

95 Scott, L. J. *et al.* A Genome-Wide Association Study of Type 2 Diabetes in Finns Detects Multiple Susceptibility Variants. *Science (New York, NY)* **316**, 1341-1345, doi:10.1126/science.1142382 (2007).

96 Sladek, R. *et al.* A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature* **445**, 881-885, doi:10.1038/nature05616 (2007).

97 Steinthorsdottir, V. *et al.* A variant in CDKAL1 influences insulin response and risk of type 2 diabetes. *Nature Genetics* **39**, 770-775, doi:10.1038/ng2043 (2007).

98 Takeuchi, F. *et al.* Confirmation of Multiple Risk Loci and Genetic Impacts by a Genome-Wide Association Study of Type 2 Diabetes in the Japanese Population. *Diabetes* **58**, 1690-1699, doi:10.2337/db08-1494 (2009).

99 Timpson, N. J. *et al.* Adiposity-Related Heterogeneity in Patterns of Type 2 Diabetes Susceptibility Observed in Genome-Wide Association Data. *Diabetes* **58**, 505-510, doi:10.2337/db08-0906 (2008).

100 Voight, B. F. *et al.* Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. *Nature Genetics* **42**, 579-589, doi:10.1038/ng.609 (2010).

101 Zeggini, E. *et al.* Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. *Nature Genetics* **40**, 638-645, doi:10.1038/ng.120 (2008).

102 Zeggini, E. *et al.* Replication of Genome-Wide Association Signals in UK Samples Reveals Risk Loci for Type 2 Diabetes. *Science (New York, NY)* **316**, 1336-1341, doi:10.1126/science.1142364 (2007).

103 Palizban, A., Nikpour, M., Salehi, R. & Maracy, M.-R. Association of a common variant in TCF7L2 gene with type 2 diabetes mellitus in a Persian population. *Clinical and experimental medicine* **12**, 115-119, doi:10.1007/s10238-011-0144-7 (2012).

104 Greenawalt, D. M. *et al.* Integrating genetic association, genetics of gene expression, and single nucleotide polymorphism set analysis to identify susceptibility Loci for type 2 diabetes mellitus. *American journal of epidemiology* **176**, 423-430, doi:10.1093/aje/kws123 (2012).

105 Gamboa-Meléndez, M. A. *et al.* Contribution of Common Genetic Variation to the Risk of Type 2 Diabetes in the Mexican Mestizo Population. *Diabetes*, doi:10.2337/db11-0550 (2012).

106 Mtiraoui, N. *et al.* Contribution of common variants of ENPP1, IGF2BP2, KCNJ11, MLXIPL, PPARγ, SLC30A8 and TCF7L2 to the risk of type 2 diabetes in Lebanese and Tunisian Arabs. *Diabetes & metabolism*, doi:10.1016/j.diabet.2012.05.002 (2012).

107 Dupuis, J. *et al.* New genetic loci implicated in fasting glucose homeostasis and their impact on type 2 diabetes risk. *Nature Genetics* **42**, 105-116, doi:10.1038/ng.520 (2010).

108 Saxena, R. *et al.* Genetic variation in GIPR influences the glucose and insulin responses to an oral glucose challenge. *Nature Genetics* **42**, 142-148, doi:10.1038/ng.521 (2010).

109 Zabaneh, D. & Balding, D. J. A genome-wide association study of the metabolic syndrome in Indian Asian men. *PLoS ONE* **5**, e11961, doi:10.1371/journal.pone.0011961 (2010).

110 Lettre, G. *et al.* Genome-Wide Association Study of Coronary Heart Disease and Its Risk Factors in 8,090 African Americans: The NHLBI CARe Project. *PLoS Genetics* **7**, e1001300, doi:10.1371/journal.pgen.1001300.t004 (2011).

111 Freathy, R. M. *et al.* Type 2 diabetes TCF7L2 risk genotypes alter birth weight: a study of 24,053 individuals. *American journal of human genetics* **80**, 1150-1161, doi:10.1086/518517 (2007).

112 Gibb, W. & Challis, J. R. G. Mechanisms of term and preterm birth. *J Obstet Gynaecol Can* **24**, 874-883 (2002).

113 Ryckman, K. K. *et al.* Maternal and fetal genetic associations of PTGER3 and PON1 with preterm birth. *PLoS ONE* **5**, e9040, doi:10.1371/journal.pone.0009040 (2010).

114 Kalish, R. B., Vardhana, S., Gupta, M., Perni, S. C. & Witkin, S. S. Interleukin-4 and -10 gene polymorphisms and spontaneous preterm birth in multifetal gestations. *American journal of obstetrics and gynecology* **190**, 702-706, doi:10.1016/j.ajog.2003.09.066 (2004).

115 Lorenz, E., Hallman, M., Marttila, R., Haataja, R. & Schwartz, D. A. Association between the Asp299Gly polymorphisms in the Toll-like receptor 4 and premature births in the Finnish population. *Pediatric research* **52**, 373-376 (2002).

116 Genç, M. R., Gerber, S., Nesin, M. & Witkin, S. S. Polymorphism in the interleukin-1 gene complex and spontaneous preterm delivery. *American journal of obstetrics and gynecology* **187**, 157-163 (2002).

117 Fujimoto, T. *et al.* A single nucleotide polymorphism in the matrix metalloproteinase-1 (MMP-1) promoter influences amnion cell MMP-1 expression and risk for preterm premature rupture of the fetal membranes. *The Journal of biological chemistry* **277**, 6296-6302, doi:10.1074/jbc.M107865200 (2002).

118 Ferrand, P. E. *et al.* A polymorphism in the matrix metalloproteinase-9 promoter is associated with increased risk of preterm premature rupture of membranes in African Americans. *Molecular human reproduction* **8**, 494-501 (2002).

119 Aidoo, M. *et al.* Tumor necrosis factor-alpha promoter variant 2 (TNF2) is associated with pre-term delivery, infant mortality, and malaria morbidity in western Kenya: Asembo Bay Cohort Project IX. *Genetic Epidemiology* **21**, 201-211, doi:10.1002/gepi.1029 (2001).

120 Dizon-Townson, D. S., Major, H., Varner, M. & Ward, K. A promoter mutation that increases transcription of the tumor necrosis factor-alpha gene is not associated with preterm delivery. *American journal of obstetrics and gynecology* **177**, 810-813 (1997).

121 Box, G. E. P. & Cox, D. R. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252 (1964).

122 Cabili, M. N. *et al.* Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes & Development* **25**, 1915-1927, doi:10.1101/gad.17446611 (2011).

123 Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J. & Lee, J. T. Polycomb Proteins Targeted by a Short Repeat RNA to the Mouse X Chromosome. *Science (New York, NY)* **322**, 750-756, doi:10.1126/science.1163045 (2008).

124 Pandey, R. R. *et al.* Kcnq1ot1 Antisense Noncoding RNA Mediates Lineage-Specific Transcriptional Silencing through Chromatin-Level Regulation. *Molecular cell* **32**, 232-246, doi:10.1016/j.molcel.2008.08.022 (2008).

125 Leighton, P. A., Ingram, R. S., Eggenschwiler, J., Efstratiadis, A. & Tilghman, S. M. Disruption of imprinting caused by deletion of the H19 gene region in mice. *Nature* **375**, 34-39, doi:10.1038/375034a0 (1995).

126 Rinn, J. L. *et al.* Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* **129**, 1311-1323, doi:10.1016/j.cell.2007.05.022 (2007).

127 Heo, J. B. & Sung, S. Vernalization-mediated epigenetic silencing by a long intronic noncoding RNA. *Science (New York, NY)* **331**, 76-79, doi:10.1126/science.1197349 (2011).

128 Faix, J. & Grosse, R. Staying in shape with formins. *Developmental Cell* **10**, 693-706, doi:10.1016/j.devcel.2006.05.001 (2006).

129 Higgs, H. N. Formin proteins: a domain-based approach. *Trends in biochemical sciences* **30**, 342-353, doi:10.1016/j.tibs.2005.04.014 (2005).

130 Brown, E. J. *et al.* Mutations in the formin gene INF2 cause focal segmental glomerulosclerosis. *Nature Genetics* **42**, 72-76, doi:10.1038/ng.505 (2009).

131 Cross, J. C. Placental function in development and disease. *Reproduction, fertility, and development* **18**, 71-76, doi:10.1071/RD05121 (2006).

132 Kramer, M. Determinants of low birth weight: methodological assessment and meta-analysis. *Bulletin of the World Health Organization* (1987).

133 Mortensen, L. H., Diderichsen, F., Smith, G. D. & Andersen, A. M. N. The social gradient in birthweight at term: quantification of the mediating role of maternal smoking and body mass index. *Human reproduction (Oxford, England)* **24**, 2629-2635, doi:10.1093/humrep/dep211 (2009).

134 Acker, D. B., Sachs, B. P. & Friedman, E. A. Risk factors for shoulder dystocia. *Obstetrics and gynecology* **66**, 762-768 (1985).

135 Battaglia, F. C. & Lubchenco, L. O. A practical classification of newborn infants by weight and gestational age. *The Journal of pediatrics* **71**, 159-163 (1967).

136 Fanaroff, A. A. *et al.* Trends in neonatal morbidity and mortality for very low birthweight infants. *American journal of obstetrics and gynecology* **196**, 147.e141-148, doi:10.1016/j.ajog.2006.09.014 (2007).

137 Barker, D. J. *et al.* Type 2 (non-insulin-dependent) diabetes mellitus, hypertension and hyperlipidaemia (syndrome X): relation to reduced fetal growth. *Diabetologia* **36**, 62-67 (1993).

138 Jarvelin, M.-R. *et al.* Early Life Factors and Blood Pressure at Age 31 Years in the 1966 Northern Finland Birth Cohort. (2004).

139 Klebanoff, M. A., Meirik, O. & Berendes, H. W. Second-generation consequences of small-for-dates birth. *PEDIATRICS* **84**, 343-347 (1989).

140 Clausson, B., Lichtenstein, P. & Cnattingius, S. Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG : an international journal of obstetrics and gynaecology* **107**, 375-381 (2000).

141 Emanuel, I., Filakti, H., Alberman, E. & Evans, S. J. Intergenerational studies of human birthweight from the 1958 birth cohort. 1. Evidence for a multigenerational effect. *British journal of obstetrics and gynaecology* **99**, 67-74 (1992).

142 Magnus, P. Causes of variation in birth weight: A study of offspring of twins - Magnus - 2008 - Clinical Genetics - Wiley Online Library. *Clinical genetics* (1984).

143 Knight, B. *et al.* Evidence of genetic regulation of fetal longitudinal growth. *Early human development* **81**, 823-831, doi:10.1016/j.earlhumdev.2005.06.003 (2005).

144 Klebanoff, M. A., Mednick, B. R., Schulsinger, C., Secher, N. J. & Shiono, P. H. Father's effect on infant birth weight. *American journal of obstetrics and gynecology* **178**, 1022-1026 (1998).

145 Magnus, P., Gjessing, H. K., Skrondal, A. & Skjaerven, R. Paternal contribution to birth weight. *Journal of epidemiology and community health* **55**, 873-877 (2001).

146 Gielen, M. *et al.* Modeling Genetic and Environmental Factors to Increase Heritability and Ease the Identification of Candidate Genes for Birth Weight: A Twin Study. *Behavior genetics* **38**, 44-54, doi:10.1007/s10519-007-9170-3 (2007).

147 Sweeney, C. *et al.* Insulin-like growth factor pathway polymorphisms associated with body size in Hispanic and non-Hispanic white women. *Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology* **14**, 1802-1809, doi:10.1158/1055-9965.EPI-05-0149 (2005).

148 Meirhaeghe, A. *et al.* A possible role for the PPARG Pro12Ala polymorphism in preterm birth. *Diabetes* **56**, 494-498, doi:10.2337/db06-0915 (2007).

149 Hattersley, A. T. & Tooke, J. E. The fetal insulin hypothesis: an alternative explanation of the association of low birthweight with diabetes and vascular disease. *Lancet* **353**, 1789-1792, doi:10.1016/S0140-6736(98)07546-1 (1999).

150 Freathy, R. M. *et al.* Type 2 diabetes risk alleles are associated with reduced size at birth. *Diabetes* **58**, 1428-1433, doi:10.2337/db08-1739 (2009).

151 Zhao, J. *et al.* Examination of type 2 diabetes loci implicates CDKAL1 as a birth weight gene. *Diabetes* **58**, 2414-2418, doi:10.2337/db09-0506 (2009).

152 Pulizzi, N. *et al.* Interaction between prenatal growth and high-risk genotypes in the development of type 2 diabetes. *Diabetologia* **52**, 825-829, doi:10.1007/s00125-009-1291-1 (2009).

153 Freathy, R. M. *et al.* Variants in ADCY5 and near CCNL1 are associated with fetal growth and birth weight. *Nat Genet* **42**, 430-435, doi:10.1038/ng.567 (2010).

154 Ryckman, K. K., Dagle, J. M., Kelsey, K., Momany, A. M. & Murray, J. C. Replication of genetic associations in the inflammation, complement, and coagulation pathways with intraventricular hemorrhage in LBW preterm neonates. *Pediatric research* **70**, 90-95, doi:10.1203/PDR.0b013e31821ceb63 (2011).

155 Pihkala, J., Hakala, T., Voutilainen, P. & Raivio, K. [Characteristic of recent fetal growth curves in Finland]. *Duodecim; lääketieteellinen aikakauskirja* **105**, 1540-1546 (1989).

156 Kouadjo, K. E., Nishida, Y., Cadrin-Girard, J. F., Yoshioka, M. & St-Amand, J. Housekeeping and tissue-specific genes in mouse tissues. *BMC genomics* **8**, 127, doi:10.1186/1471-2164-8-127 (2007).

157 Mook-Kanamori, D. O. *et al.* Variants near CCNL1/LEKR1 and in ADCY5 and fetal growth characteristics in different trimesters. *The Journal of clinical endocrinology and metabolism* **96**, E810-815, doi:10.1210/jc.2010-2316 (2011).

158 Züchner, S. *et al.* Whole-Exome Sequencing Links a Variant in DHDDS to Retinitis Pigmentosa. *American journal of human genetics* **88**, 201-206, doi:10.1016/j.ajhg.2011.01.001 (2011).

159 Ng, S. B. *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* **42**, 30-35, doi:10.1038/ng.499 (2010).

160 Kiezun, A. *et al.* Exome sequencing and the genetic basis of complex traits. *Nature Genetics* **44**, 623-630, doi:10.1038/ng.2303 (2012).

161 O'Roak, B. J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe de novo mutations. *Nature Genetics* **43**, 585-589, doi:10.1038/ng.835 (2011).

162 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

163 Yandell, M. *et al.* A probabilistic disease-gene finder for personal genomes. *Genome Research* **21**, 1529-1542, doi:10.1101/gr.123158.111 (2011).

164 Kanehisa, M. & Goto, S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28**, 27-30 (2000).

165 MacArthur, D. G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science (New York, NY)* **335**, 823-828, doi:10.1126/science.1215040 (2012).

166 Wang, X. *et al.* Molecular epidemiology of preterm delivery: methodology and challenges. *Paediatr Perinat Epidemiol* **15 Suppl 2**, 63-77 (2001).

167 Hao, K. *et al.* A candidate gene association study on preterm delivery: application of high-throughput genotyping technology and advanced statistical methods. *Hum Mol Genet* **13**, 683-691, doi:10.1093/hmg/ddh091 (2004).

168 Velez, D. R. *et al.* Preterm birth in Caucasians is associated with coagulation and inflammation pathway gene variants. *PLoS ONE* **3**, e3283, doi:10.1371/journal.pone.0003283 (2008).

169 Härtel, C. *et al.* Polymorphisms of haemostasis genes as risk factors for preterm delivery. *Thrombosis and haemostasis* **94**, 88-92, doi:10.1267/THRO05010088 (2005).

170 Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248-249, doi:10.1038/nmeth0410-248 (2010).

171 Plunkett, J. *et al.* An Evolutionary Genomic Approach to Identify Genes Involved in Human Birth Timing. *PLoS Genetics* **7**, e1001365, doi:10.1371/journal.pgen.1001365.t002 (2011).

172 Vaisbuch, E. *et al.* Activation of the alternative pathway of complement is a feature of pre-term parturition but not of spontaneous labor at term. *American journal of reproductive immunology (New York, N.Y. : 1989)* **63**, 318-330, doi:10.1111/j.1600-0897.2009.00800.x (2010).

173 Soto, E. *et al.* Anaphylatoxins in preterm and term labor. *Journal of perinatal medicine* **33**, 306-313, doi:10.1515/JPM.2005.051 (2005).

174 Lynch, A., Gibbs, R., Murphy, J. & Byers, T. ScienceDirect.com - American Journal of Obstetrics and Gynecology - Complement activation fragment Bb in early pregnancy and spontaneous preterm birth. *American journal of …* (2008).

175 Lynch, A. M. *et al.* Early elevations of the complement activation fragment C3a and adverse pregnancy outcomes. *Obstetrics and gynecology* **117**, 75-83, doi:10.1097/AOG.0b013e3181fc3afa (2011).

176 Imrie, H., McGonigle, T., Liu, D. & Jones, D. Reduction in erythrocyte complement receptor 1 (CR1, CD35) and decay accelerating factor (DAF, CD55) during normal pregnancy. *Journal of reproductive immunology* **31**, 221-227 (1996).

177 Ng, P. C. & Henikoff, S. SIFT: Predicting amino acid changes that affect protein function. *Nucleic Acids Research* **31**, 3812-3814 (2003).

178 Kullo, I. J. *et al.* Complement receptor 1 gene variants are associated with erythrocyte sedimentation rate. *American journal of human genetics* **89**, 131-138, doi:10.1016/j.ajhg.2011.05.019 (2011).

179 van den Broe, N. R. & Letsky, E. A. Pregnancy and the erythrocyte sedimentation rate. *BJOG : an international journal of obstetrics and gynaecology* **108**, 1164-1167 (2001).

180 Ricklin, D. & Lambris, J. D. Complement-targeted therapeutics. *Nature Biotechnology* **25**, 1265-1275, doi:10.1038/nbt1342 (2007).

181 Weisman, H. F. *et al.* Soluble human complement receptor type 1: in vivo inhibitor of complement suppressing post-ischemic myocardial inflammation and necrosis. *Science (New York, NY)* **249**, 146-151 (1990).

182 Wang, Y., Rollins, S. A., Madri, J. A. & Matis, L. A. Anti-C5 monoclonal antibody therapy prevents collagen-induced arthritis and ameliorates established disease.

*Proceedings of the National Academy of Sciences of the United States of America* **92**, 8955-8959 (1995).

183 Zareba, K. Eculizumab: A novel therapy for paroxysmal nocturnal hemoglobinuria. *Drugs Today* (2007).

184 Makowsky, R. *et al.* Beyond Missing Heritability: Prediction of Complex Traits. *PLoS Genetics* **7**, e1002051, doi:10.1371/journal.pgen.1002051.t002 (2011).

185 Pavlicev, M. & Wagner, G. P. A model of developmental evolution: selection, pleiotropy and compensation. *Trends in ecology & evolution* **27**, 316-322, doi:10.1016/j.tree.2012.01.016 (2012).

186 Sivakumaran, S. *et al.* Abundant pleiotropy in human complex diseases and traits. *American journal of human genetics* **89**, 607-618, doi:10.1016/j.ajhg.2011.10.004 (2011).

187 Kho, A. N. *et al.* Use of diverse electronic medical record systems to identify genetic risk for type 2 diabetes within a genome-wide association study. *Journal of the American Medical Informatics Association* **19**, 212-218, doi:10.1136/amiajnl-2011-000439 (2012).

188 Alleman, B. W. *et al.* No observed association for mitochondrial SNPs with preterm delivery and related outcomes. *Pediatric research* **72**, 539-544, doi:10.1038/pr.2012.112 (2012).