X-Chromosomal Maternal and Fetal SNPs and the Risk of Spontaneous Preterm Delivery in a Danish/Norwegian Genome-Wide Association Study

Solveig Myking1,2, Heather A. Boyd3, Ronny Myhre1, Bjørke Feenstra2, Astanand Jugessur1,3, Aase S. Devold Pay1,4, Ingrid H. G. Østensen1, Nils-Halvdan Morken5,6, Tamara Busch8, Kelli K. Ryckman8,9, Frank Geller2, Per Magnus1, Håkon K. Gjessing1,5, Mads Melbye2,9, Bo Jacobsson1,7,*, Jeffrey C. Murray8,9,

1 Department of Genes and Environment, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway, 2 Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark, 3 Craniofacial Research, Murdoch Childrens Research Institute, Royal Children’s Hospital, Parkville, Australia, 4 Department of Obstetrics and Gynecology, Women and Children’s Division, Oslo University Hospital, Oslo, Norway, 5 Department of Public Health and Primary Health Care, University of Bergen, Bergen, Norway, 6 Department of Obstetrics and Gynecology, Haukeland University Hospital, Bergen, Norway, 7 Department of Obstetrics and Gynecology, Institute for the Health of Women and Children, Sahlgrenska University Hospital, Göteborg, Sweden, 8 Department of Pediatrics, Carver College of Medicine, University of Iowa, Iowa City, Iowa, United States of America, 9 Department of Epidemiology, College of Public Health, University of Iowa, Iowa City, Iowa, United States of America

Abstract

Background: Recent epidemiological studies suggest that the maternal genome is an important contributor to spontaneous preterm delivery (PTD). There is also a significant excess of males among preterm born infants, which may imply an X-linked mode of inheritance for a subset of cases. To explore this, we examined the effect of maternal and fetal X-chromosomal single nucleotide polymorphisms (SNPs) on the risk of PTD in two independent genome-wide association studies and one replication study.

Methods: Participants were recruited from the Danish National Birth Cohort and the Norwegian Mother and Child cohort studies. Data from these two populations were first analyzed independently, and then combined in a meta-analysis. Overall, we evaluated 12,211 SNPs in 1,535 case-mother dyads and 1,487 control-mother dyads. Analyses were done using a hybrid design that combines case-mother dyads and control-mother dyads, as implemented in the Haplin statistical software package. A sex-stratified analysis was performed for the fetal SNPs. In the replication study, 10 maternal and 16 fetal SNPs were analyzed using case-parent triads from independent studies of PTD in the United States, Argentina and Denmark.

Results: In the meta-analysis, the G allele at the maternal SNP rs2747022 in the FERM domain containing 7 gene (FRMD7) increased the risk of spontaneous PTD by 1.2 (95% confidence interval (CI): 1.1, 1.4). Although an association with this SNP was confirmed in the replication study, it was no longer statistically significant after a Bonferroni correction for multiple testing.

Conclusion: We did not find strong evidence in our data to implicate X-chromosomal SNPs in the etiology of spontaneous PTD. Although non-significant after correction for multiple testing, the mother’s G allele at rs2747022 in FRMD7 increased the risk of spontaneous PTD across all populations in this study, thus warranting further investigation in other populations.

Citation: Myking S, Boyd HA, Myhre R, Feenstra B, Jugessur A, et al. (2013) X-Chromosomal Maternal and Fetal SNPs and the Risk of Spontaneous Preterm Delivery in a Danish/Norwegian Genome-Wide Association Study. PLoS ONE 8(4): e61781. doi:10.1371/journal.pone.0061781

Editor: Wei Yan, University of Nevada School of Medicine, United States of America

Received October 19, 2012; Accepted March 13, 2013; Published April 16, 2013

Copyright: © 2013 Myking et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by the National Institutes of Health (NIH) [grant numbers HD52953, HD57192 to J.C.M.]. TheDNBC was established with the support of a major grant from the Danish National Research Foundation. Additional support for the DNBC has been provided by the Danish Pharmacists' Fund, the Egmont Foundation, the March of Dimes Birth Defects Foundation, the Augustinus Foundation and the Health Fund of the Danish Health Insurance Societies. The generation of GWAS genotype data for the DNBC samples was funded as part of the Gene Environment Association Studies (GENEVA) under GEI. Other support came from the Danish Medical Research Council and the Lundbeck Foundation. Assistance with phenotype harmonization and genotype cleaning, as well as with general study coordination, was provided by the GENEVA Coordinating Center (U01 HG004446). Assistance with data cleaning was provided by the National Center for Biotechnology Information (NCBI). Funding support for genotyping, which was performed at the Johns Hopkins University Center for Inherited Disease Research (CIDR) and the Broad Institute of MIT and Harvard, was provided by the NIH GEI (U01HG004438 for CIDR; U01HG04424 for Broad) and the NIH contract “High throughput genotyping for studying the genetic contributions to human disease” (HHSN26820072096C for CIDR only). The Norwegian Mother and Child Cohort Study (MoBa) is supported by the Norwegian Ministry of Health and the Ministry of Education and Research, NIH/NIEHS (contract no. NO-ES-75558), NIH/NINDS (grant no. 1 U01 NS 047537-01), and the Norwegian Research Council/FUGE (grant no. 151918/S10). Support also came from the Norwegian Research Council (FUGE 183220/S10, FRIMEDKLI-05 ES236011); Swedish government grants to researchers in the public health service (ALF) [ALFGBG-136431]; Sahlgrenska University Hospital; Sahlgrenska Academy; Gothenburg, Sweden; the Swedish Medical Society, Stockholm, Sweden (2008-21198); and the Jane and Dan Olsson Research Foundation, Gotthenburg, Sweden. The MoBa samples were genotyped at the genotyping core facility at Oslo University Hospital, Norway. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: bo.jacobsson@obgyn.gu.se
9 These authors contributed equally to this work.
Introduction

Preterm delivery (PTD), defined as delivery before 37 weeks of gestation, affected 15 million births in 2010 [1]. It is associated with a substantially increased risk of mortality, as well as short- and long-term morbidity [1,2]. The PTD rate ranges from 6% in Scandinavian countries [3] to 18% in some African populations [1]. PTD is routinely divided into two main groups according to clinical presentation: i) spontaneous PTD, in which delivery starts with either uterine contractions (preterm labor) or membrane rupture (preterm prelabor rupture of membranes (PPROM)), and ii) iatrogenic PTD, which is induced by medical or surgical intervention.

Spontaneous PTD is etiologically heterogeneous, involving both genetic and environmental risk factors. Twin studies have estimated the heritability of PTD at 17–36% [4,5], and that of parturition timing at 34% [6]. Although there is compelling evidence for a genetic component to PTD, no common genetic variants or mode of inheritance have yet been established [7]. Recent generational epidemiological studies suggest that the maternal genome is a key genetic contributor to PTD inheritance [7–9], and a personal history of PTD is considered the most important risk factor for PTD in multiparous women [7]. Women who were born preterm themselves, or who have sisters or maternal half-sisters with a history of PTD, are also at increased risk [7,8].

A small but significant excess of males has been observed among preterm-born infants in most populations, especially for spontaneous PTD [10]. Different hypotheses have been proposed to explain this excess, including a shorter gestational length due to higher average fetal weight [11], increased vulnerability to certain pregnancy complications [10], and the fact that biochemical processes, such as increased estrogen production from androgen precursors or higher levels of interleukin-1 in the amniotic fluid of males, could lead to uterine contractions and PTD [12,13]. It is also possible that this male excess is due to fetal X-linked risk alleles contributing to a subset of the PTD cases, in which hemizygosity would increase the risk compared to heterozygosity [12].

Several complex disorders have been associated with genetic variants on the X chromosome, including prostate cancer [14], type 2 diabetes [15], X-linked dystonia Parkinsonism [16] and psychiatric disorders such as schizophrenia and autism spectrum disorders [17]. Hypospadias, a congenital malformation of male external genitalia, is strongly associated with genetic variants on the X chromosome [18]. As males inherit only one X chromosome, they are hemizygous for X-linked genes, and although females inherit two X chromosomes, one of them is inactivated in each cell. A recent meta-analysis found a significant association between skewed X-chromosome inactivation in females and idiopathic recurrent spontaneous abortion [19]. Recurrent second-trimester spontaneous abortion is a well-known risk factor for spontaneous PTD [20,21], and it is plausible that some of the same mechanisms may be involved in both outcomes.

Most candidate-gene studies of PTD have used a case-control study design and focused primarily on autosomal markers. This may be partly due to limited knowledge about plausible X-linked candidate genes for PTD and a lack of appropriate statistical methodology for family-based association studies of X chromosome markers [22]. Nearly all methods that have subsequently been developed for handling X-linked markers are based on the transmission/disequilibrium test (TDT) [23–27]. A family-based likelihood ratio test for the X chromosome (X-LRT) [28] and Haplin [22] are among the few existing methods that can estimate genetic relative risks, in contrast to other methods [23–25,27,29] that only generate a p-value for hypothesis-testing.

Taking into consideration the strong evidence of a maternally mediated genetic effect in PTD and the higher proportion of males among infants born preterm, we examined the effects of maternal and fetal X-linked gene variants on the risk of PTD, using data from two genome-wide association studies (GWAS) in Scandinavia (Norway and Denmark). To verify our findings, we conducted a replication study using case-parent triads from independent studies in the United States, Argentina and Denmark.

Materials and Methods

Ethics Statement

All the studies outlined in this paper were approved by the regional ethics committees or institutional review boards (IRB) at each site, and a written informed consent was obtained from each participant. The Danish National Birth Cohort (DNBC) study protocol was approved by the Danish Scientific Ethics Committee for the Copenhagen Capital City Region and by the Danish Data Protection Agency. The Norwegian Mother and Child Cohort Study (MoBa) was approved by the Regional Committee for Medical Research Ethics in South-Eastern Norway and the Norwegian Data Inspectorate. For the replication study, ethics approvals were obtained from the Human Subjects Office, University of Iowa (Iowa City, USA), the University of Pittsburgh Institutional Review Board (Pittsburgh Pennsylvania), the University of Rochester Research Subjects Review Board (Rochester, New York), the Wake Forest University Review Board (Wake Forest, North Carolina), and El Comite de Etica en Investigacion del Cemic (Argentina).

Study Participants

This study was conducted using data from DNBC [30] and MoBa [31]. A replication study of 10 of the most significant maternal SNPs and 16 of the most significant fetal SNPs from the combined analysis of DNBC and MoBa data was performed in independent family studies of PTD in the United States (US), Argentina and Denmark.

DNBC. The DNBC includes approximately 100,000 pregnancies from 1996 to 2002. Women were invited to participate at their first prenatal visit with their general practitioner at gestational weeks 6–12 [30]. Information on exposures not registered in medical records was obtained from national registers, telephone interviews, and a food frequency questionnaire. Blood samples from mothers were collected by the general practitioner during the routine visit at gestational weeks 6–12 and 24, and a cord blood sample was taken at delivery. Blood samples were stored in the Danish National Biobank at the Statens Serum Institut in Copenhagen, Denmark.

A case was defined as a live, singleton spontaneous PTD occurring before 259 days (37 weeks) of gestation; a control was defined as a live, singleton full-term delivery (i.e. occurring at 280–286 days (40 weeks) of gestation). The exclusion criteria were: fetal malformations, preeclampsia/eclampsia, placenta previa, placental abruption, polyhydramnios, isoimmunization and placental insufficiency. In addition, the infant’s parents and grandparents had to be of Nordic ancestry (i.e. born in Denmark or one of the other Nordic countries).

MoBa. MoBa is a nationwide Norwegian pregnancy cohort study administered by the Norwegian Institute of Public Health (NIHP). The study includes more than 107,000 pregnancies.
recruited from 1999 through 2008. Women were invited by postal invitation in connection with a routine ultrasound screening offered to all pregnant women in Norway at gestational weeks 17–19. Most of the pregnant women in Norway were invited and the participation rate was 42.7%. Participation rates for the first three questionnaires were 92–95% [31]. For the current study, cases and controls were selected from Version 4 of the MoBa cohort, which included a total of 71,669 pregnancies. This version was released in 2008 for research use.

Blood samples were drawn from the pregnant woman and the fetus’ father during the ultrasound appointment. A new blood sample from the woman and a cord-blood sample from the infant were collected at delivery. All biological specimens were sent to the MoBa Biobank where DNA was extracted, processed and stored until retrieval [32].

A case was defined as a live, singleton spontaneous PTD occurring between 154 and 258 days of gestation (220/7–364/7 weeks); a control was defined as a live, singleton full-term delivery, i.e., occurring at 273–286 days of gestation (390/7 and 404/7 weeks). Gestational age was estimated by ultrasound at gestational weeks 17–19. In the few cases without ultrasound dating, gestational age was estimated using the date of the last menstrual period. Strict selection criteria were applied to both cases and controls in order to yield the clearest possible phenotype. Only women in the age group 20–34 years were selected. As women aged <20 years and >35 years have an increased risk of spontaneous PTD [21], only women in the age group 20–34 years were selected in order to prevent the increased risk from affecting the results. Pregnancies involving pre-existing medical conditions, such as diabetes, hypertension, specific autoimmune diseases (inflammatory bowel disease, systemic lupus erythematosus, rheumatoid arthritis and scleroderma) and immune-compromised conditions, were excluded from the study. Lastly, pregnancies with complications such as preeclampsia, hypertension, gestational diabetes, placental abruption, placenta previa, cervical cerclage, small for gestational age and fetal malformation were also excluded, as were pregnancies conceived by in vitro fertilization.

The US prematurity study. This study focuses on PTD cases and their families. Study participants (cases and both parents and grandparents when available) were enrolled at various locations in the US, including Iowa City (Iowa), Wake Forest (North Carolina), Pittsburgh (Pennsylvania) and Rochester (New York). DNA from each participant was extracted from saliva or blood samples. Cases were defined as PTD if gestational age was <37 weeks. In order to harmonize the phenotypes of the replication cohort with those of the original study populations, we excluded indicated deliveries without PPROM, multiple gestations, fetal malformations, preeclampsia or hypertension in pregnancy, placental abruption, placenta previa, conception via assisted reproductive technology, and maternal age <20 or >39 years. Only white, non-Hispanic individuals were included in the study. For the maternal replication, all ethnicities were included in one of the analyses, but after other inclusion criteria were applied, only white, non-Hispanic individuals remained for analysis.

The Argentina prematurity study. The Argentina Prematurity Study had similar enrollment criteria as the US Prematurity study. Cases were defined as PTD if gestational age was <37 weeks. We excluded deliveries with multiple gestations, preeclampsia, maternal age <20 or >39 years, and indicated deliveries without PPROM. For the current study, maternal SNPs were replicated using triads consisting of the mother of a preterm-born infant and her parents.

The Denmark family study of PTD. This study includes mothers of preterm-born infants and their parents (the infant’s maternal grandparents). Cases were defined as PTD if gestational age was <37 weeks. Indicated deliveries and congenital malformations were excluded.

Genotyping

The Illumina Human660W-Quad BeadChip platform (Illumina, San Diego, CA, USA) was used for genotyping in both study populations. The DNBC samples were genotyped by the Center for Inherited Disease Research (CIDR) at the Johns Hopkins University (Baltimore, MD, USA). The MoBa samples were genotyped at the genotyping core facility at Oslo University Hospital (Oslo, Norway). For replication, the 28 selected SNPs were genotyped at the University of Iowa using the TaqMan® chemistry genotyping system (Applied Biosystems, Foster City, CA, USA). All reactions were performed under standard conditions supplied by Applied Biosystems. Following thermocycling, fluorescence levels of the FAM and VIC dyes were measured and genotypes were scored using the proprietary Sequence Detection Systems 2.2 software (Applied Biosystems) and reviewed manually by at least two independent observers.

Quality Control

DNBC. In the Danish data, 2,035 mother-infant pairs were available for analysis (1,061 case pairs and 974 control pairs). Of these, 20 dyads were excluded because of a call rate <97%, in either the mother or the infant. One dyad was excluded because of unknown gender in the infant, and a further nine dyads were excluded because either the mother or the infant had a sibling or half-sibling in the cohort. After quality control, 2,005 mother-infant dyads (1,046 case dyads and 959 control dyads) were available for the current analysis.

Before data processing, there were 14,441 SNPs on the X chromosome. After filtering out SNPs that had a call rate <95%, more than five Mendelian inconsistencies, or a minor allele frequency (MAF) of <1%, we were left with 12,345 SNPs for further analysis.

MoBa. In the Norwegian data, there were 1,086 complete mother-infant dyads eligible for analysis (529 case dyads and 557 control dyads). Sixty-two dyads were excluded because they had a genotype call rate <97% in either the mother or the infant. In addition, 7 mother-infant pairs were excluded because of inconsistencies in parenthood. This left 1,017 mother-infant dyads (489 case-dyads and 528 control-dyads) for the current analysis.

Overall, 14,441 SNPs on the X chromosome were available before data processing. SNPs that had a call rate <95%, more than five Mendelian inconsistencies, or a MAF <1% were excluded, leaving 12,361 SNPs for the current analysis.

The US prematurity study and the Denmark family study of PTD. The fetal triads consisted of a preterm infant and his/her parents. For the fetal triads, 286 case families (787 individuals) were eligible for analysis. Of these, three families (15 individuals) were removed because of Mendelian inconsistencies, 25 individuals were removed because of inconsistencies in parenthood. This left 267 families (727 individuals) for further analysis. A total of 18 SNPs were chosen for replication. One SNP (rs5918890) failed assay design and another SNP (rs6524611) had a call rate <95%, leaving 16 SNPs for analysis.

The maternal triads consisted of the mother of a preterm infant and the infant’s maternal grandparents. For analysis of the maternal triads, 98 case families (293 individuals) were eligible for analysis. Of these, 116 individuals were excluded because they had
different analyses (Figure S1). In addition, regional association plots were generated for the combined p-values for each of the combined using Fisher’s method [35], and quantile-quantile (QQ) plots for the SNPs with the lowest uncorrected p-values were made using a modified version of the R script available at http://www.broadinstitute.org/files/shared/diabetes/scandinavs/assocplot.R (Figure 1). With the meta-analysis approach, relative risks with opposing directions in the two populations may cancel each other out, whereas the Fisher method combines p-values irrespective of the direction of the effect.

The replication data from the US prematurity study were analyzed using the case-parent triad module in Haplin. In this setting, the unit of analysis is a preterm infant and his/her parents/siblings. The analyzed maternal triads consisted of the mother of a preterm infant and her parents (the infant’s maternal grandparents).

Haplin was originally designed to analyze genetic and environmental risk factors using a case-parent triad approach, a case-control approach or a combination of the two [34]. In addition to fetal effects, Haplin can also estimate maternal effects unambiguously. Haplin is based on log-linear modeling and uses a full maximum likelihood (ML) model for estimation of relative risks. In addition, missing genotypes are imputed using the expectation maximization (EM) algorithm.

The relationship between male and female allele effects may be influenced by X inactivation in females. The analyses were therefore carried out using a parameterization model in which boys and girls were assigned different baseline risks ($B_B$ and $B_G$) and a shared relative risk ($R_R$). By assuming separate baseline risks, confounding due to other effects that can influence sex differences can be avoided. The baseline risk applies when no risk alleles are present (i.e., $A_1$ in boys and $A_1A_1$ in girls, with $A_2$ representing the risk allele). The risk increases similarly in boys and girls ($B_B*RR$ and $B_G*RR$) when they are hemizygous for the risk allele ($A_1$ in boys and $A_2A_2$ in girls). When girls are heterozygous ($A_1A_2$), the risk is the average of $B_G$ and $B_G*RR$. 

Table 1. Demographic and pregnancy characteristics for the cases and controls in the Danish and Norwegian GWAS.

|                | DNBC (n = 1,046) | Controls (n = 959) | p     | MoBa (n = 489) | Controls (n = 528) | p     |
|----------------|------------------|--------------------|-------|----------------|--------------------|-------|
| Maternal age (years) | 26 (17–44)       | 30 (18–42)         | <0.001| 28.7 (20–34)   | 29.4 (20–34)       | 0.003 |
| Gestational age (days) | 245 (142–258)   | 283 (273–293)      | <0.001| 251.8 (172–256) | 280.5 (273–286)    | <0.001|
| Birth weight (g)* | 2550 (600–4040)  | 3700 (2165–5250)   | <0.001| 2760 (747–3840) | 3670 (2610–4950)   | <0.001|
| Male            | 573 (54.8%)      | 502 (52.4%)        | 0.28  | 265 (54.2%)    | 247 (46.8%)        | 0.018 |
| Previous PTD**  | 62 (19.9%)       | 20 (3.2%)          | <0.001| 49 (25.5%)     | 12 (4.0%)          | <0.001|
| Primiparous***  | 725 (70.2%)      | 341 (35.6%)        | <0.001| 297 (60.7%)    | 226 (42.8%)        | <0.001|

DNBC – Danish National Birth Cohort; MoBa – Norwegian Mother and Child Cohort Study. For continuous variables, median is reported with the range in brackets.

**Previous PTD based on data for multipara only, 4 missing for controls in DNBC.

***13 missing for cases and 1 missing for controls in DNBC.

doi:10.1371/journal.pone.0061781.t001
addition to joint analyses of males and females, Haplin has an option for running sex-specific analyses [22].

**Results**

Demographic and pregnancy characteristics for cases and controls in DNBC and MoBa are outlined in Table 1. Table 2 shows the number of families in the replication study.

**Maternal Effects**

Of 12,211 analyzed maternal SNPs, 29 had a combined p-value \(10^{-2}\) before Bonferroni correction for multiple testing and 14 of these had an uncorrected p-value <0.05 in both the Norwegian and Danish population (Table S1). The best result was for rs7892483, with a combined p-value of \(9.8 \times 10^{-6}\) and a relative risk (RR) of 1.7 (95% CI: 1.3, 2.1; Table 3). However, none of the SNPs remained significant after correcting for multiple testing.

**Fetal Effects**

Of the analyzed fetal SNPs, 19 had a combined p-value \(10^{-3}\) before Bonferroni correction and 9 of these had an uncorrected p-value <0.05 in both populations in the discovery phase (Table S2). None of the SNPs reached the chromosome-wide significance level threshold of \(4.1 \times 10^{-6}\), but rs2961403 was closest to significance, with a combined p-value of \(1.2 \times 10^{-5}\) and a RR of 1.3 (95% CI: 1.1, 1.4) (Table 4). However, this SNP was only borderline significant in the Norwegian sample (p-value = \(1.8 \times 10^{-6}\)). In the Danish study, rs6528251 in the “DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 26B” (DDX26B) gene was the SNP closest to significance (p-value = \(1.1 \times 10^{-4}\)).

**Sex-stratified Analyses**

When girls and boys were analyzed separately, 10 SNPs had a combined, uncorrected p-value <0.05 for boys and five of these SNPs had a p-value <0.05 in both studies (Table S3). The SNP closest to significance was rs6652393 in the interleukin-1 receptor-associated protein-like 2 gene (IL1RAPL2) (Table 5). In boys, the A-allele of this SNP had an overall p-value of \(3.1 \times 10^{-5}\) and a RR of 1.2 (95% CI: 1.1, 1.3). Moreover, three of the other top SNPs were located in IL1RAPL2 (Table S3). For girls, there were 17 SNPs with a combined p-value <0.05, only one of which had a p-value <0.05 in both studies (Table 6, Table S4).

**Replication Analyses**

In the maternal replication study, SNP rs2747022 located in FRMD7 had an uncorrected p-value of 0.01 (Table 7). After including the Argentinean families, this SNP had an uncorrected p-value of 0.03 (Table 8, Figure 1A). In the fetal replication, four SNPs showed a positive association with preterm delivery, although none reached genome-wide significance.

---

**Table 2. Case families in the replication study.**

|                      | Maternal US | Maternal Denmark | Maternal Argentina | Fetal |
|----------------------|-------------|------------------|--------------------|-------|
| Full triads          | 22          | 4                | 46                 | 182   |
| Mother-infant dyads  | 23          | 2                | 38                 | 66    |
| Father-infant dyads  | 2           | 0                | 3                  | 14    |
| Father, father, two siblings | 0          | 0                | 0                  | 4     |
| Father, mother, two siblings | 0        | 0                | 0                  | 1     |
| Total number of families | 47        | 6                | 87                 | 267   |
| Total number of individuals | 116       | 16               | 220                | 727   |

---

**Table 3. Maternal results from the two independent GWAS and the combined analysis.**

| Gene     | SNP      | Alleles | DNBC    | MoBa    | Combined analysis |
|----------|----------|---------|---------|---------|-------------------|
|          |          |         | RR 95% CI | MAF      | RR 95% CI | p RR       |
| rs7892483| A/g      | 0.03    | 1.76 (1.29, 2.37) | 0.05 | 1.58 (1.08, 2.30) | 1.69 (1.34, 2.13) | 9.81E-06 |
| rs5972070| a/G      | 0.03    | 1.65 (1.21, 2.23) | 0.05 | 1.62 (1.10, 2.36) | 1.64 (1.30, 2.08) | 3.67E-05 |
| rs6619677*| a/C      | 0.03    | 1.59 (1.16, 2.17) | 0.03 | 1.81 (1.17, 2.77) | 1.67 (1.30, 2.14) | 5.85E-05 |
| REPS2    | rs12557633| a/G     | 0.08    | 1.42 (1.14, 1.75) | 0.09 | 1.34 (1.00, 1.78) | 1.39 (1.17, 1.64) | 1.54E-04 |
| FRMD7    | rs4562494| a/C     | 0.28    | 0.82 (0.71, 0.94) | 0.28 | 0.78 (0.63, 0.96) | 0.80 (0.72, 0.90) | 2.14E-04 |
| FGDI     | rs2747022| A/g     | 0.31    | 1.17 (1.02, 1.33) | 0.31 | 1.33 (1.10, 1.61) | 1.22 (1.10, 1.36) | 2.62E-04 |
| IL1RAPL1 | rs5927786**| a/G     | 0.47    | 0.79 (0.69, 0.89) | 0.47 | 0.94 (0.78, 1.13) | 0.84 (0.75, 0.93) | 5.77E-04 |
| IL1RAPL1 | rs4829104| A/g***  | 0.52    | 1.25 (1.42, 1.10) | 0.49 | 1.08 (0.90, 1.28) | 1.19 (1.07, 1.32) | 8.39E-04 |
| COL4A6   | rs2295912| a/G     | 0.02    | 0.50 (0.30, 0.84) | 0.02 | 0.48 (0.22, 1.04) | 0.50 (0.33, 0.76) | 1.30E-03 |

*deviates from HWE in the Norwegian population; failed replication analysis in Haplin.
**deviates from HWE in the Danish population.
***a/G in the Danish population.

Abbreviations: MAF, minor allele frequency; RR, relative risk; 95% CI, 95% confidence interval; HWE, Hardy Weinberg Equilibrium.

**References**

1. Haplin KL, et al. [22]
SNPs had p-values \( < 0.05 \): rs4239992 in MIR-505 (microRNA 505), rs17328647 in ATP11C (ATPase, class 6 type, 11C), and rs5953790 and rs2485729 located close to the ATP11C gene region (Table 9, Figure 1B). All of the SNPs were in strong linkage disequilibrium (LD) with each other and the G allele at rs4239992 increased the relative risk of spontaneous PTD by 2.3 (95% CI: 1.3, 4.1). In the sex-stratified analysis, there was suggestive evidence of association with these SNPs in boys, but not in girls.

### Discussion

In this meta-analysis of two independent GWAS from two ancestrally and geographically closely related populations, many SNPs had p-values \( < 0.05 \) but none remained significant after a Bonferroni adjustment for multiple testing. Replication was attempted for the SNPs closest to significance, but the results were inconclusive.

### Maternal Effects

For the maternal SNPs, the most promising finding was rs2747022 in FRMD7. Before Bonferroni correction, the G allele at this SNP was associated with an increased risk of spontaneous PTD in the Norwegian and Danish data, the combined analysis and the replication study, also when the Argentinean families were included. Another SNP in this gene, rs7880476, was also associated in the combined analyses of the Norwegian and Danish data. But like rs2747022 above, the association was no longer significant after Bonferroni correction.

FRMD7 maps to Xq26.2 and is associated with X-linked idiopathic congenital nystagmus [36]. This condition is fully penetrant in males, but has incomplete penetrance in females [36]. Expression analyses show that mRNA is present at low levels in most human adult tissues. In embryos, there is expression in various parts of the brain [36], and it has been postulated that the FRMD7 protein is important for neurite development and neuronal differentiation [37]. It is unclear how this gene might

### Table 4. Fetal results from the two independent GWAS and the combined analysis.

| Gene   | SNP       | Alleles | MAF   | RR (95% CI) | MAF   | RR (95% CI) | RR (95% CI) | p     | p overall |
|--------|-----------|---------|-------|-------------|-------|-------------|-------------|-------|-----------|
| DNBC   | rs5962953 | A/g     | 0.42  | 1.11 (1.01, 1.22) | 0.44  | 1.27 (1.10, 1.45) | 1.16 (1.07, 1.25) | 2.53E-04 | 4.37E-04 |
| MoBa   | rs1313191 | A/g*    | 0.31  | 1.18 (1.07, 1.31) | 0.29  | 1.20 (1.03, 1.39) | 1.19 (1.10, 1.29) | 3.16E-05 | 2.23E-04 |
| Combined |         |         |       |             |       |             |             |       |           |
|        |           |         |       |             |       |             |             |       |           |

1 The five most significant SNPs.

| doi:10.1371/journal.pone.0061781.t005

### Table 5. Sex-stratified analysis, males.

| Gene   | SNP       | Alleles | MAF   | RR (95% CI) | MAF   | RR (95% CI) | RR (95% CI) | p     | p overall |
|--------|-----------|---------|-------|-------------|-------|-------------|-------------|-------|-----------|
| DNBC   | rs6652393 | a/G     | 0.38  | 1.14 (1.04, 1.26) | 0.40  | 1.32 (1.15, 1.52) | 1.20 (1.11, 1.29) | 5.47E-06 | 1.15E-05 |
| MoBa   | rs3008952 | a/G     | 0.07  | 1.10 (0.92, 1.30) | 0.05  | 1.68 (1.33, 2.12) | 1.28 (1.11, 1.46) | 4.60E-04 | 7.50E-05 |
| Combined |         |         |       |             |       |             |             |       |           |
|        |           |         |       |             |       |             |             |       |           |
|        |           |         |       |             |       |             |             |       |           |

| doi:10.1371/journal.pone.0061781.t004

PLOS ONE | www.plosone.org 7 April 2013 | Volume 8 | Issue 4 | e61781
be involved in the pathogenesis of spontaneous PTD. An associated SNP may be a surrogate for etiologic variants adjacent to the gene in which the most associated SNP is located. However, the genes immediately flanking FRMD7 (MST4, RAP2C, MBNL3) do not appear to be obvious candidates for PTD, although MBNL3 is involved in alternative splicing, a mechanism that would be consistent with altered developmental expression patterns.

The SNP closest to significance in the combined analysis was rs7892483. This finding was consistent across the Norwegian and Danish samples, but not in the replication study involving the US and Argentinean samples. Of the other most associated SNPs, four (rs5972070, rs5972071, rs5973734, and rs5973741; Table S1) were in strong LD with rs7892483. These SNPs lie in a gene desert and have not previously been associated with any disease. It is possible that they are in LD with a causal variant of unknown location. Several loci associated with disease have been localized to gene deserts, and it has been shown that these regions may harbor regulatory elements that can modulate gene expression over large distances on the chromosome [38].

**Fetal Effects**

Before Bonferroni correction, the closely linked SNPs (rs4239992, rs17328647, rs5953790 and rs2485729) do not appear to be obvious candidates for PTD, although MBNL3 is involved in alternative splicing, a mechanism that would be consistent with altered developmental expression patterns.

The five most significant SNPs.

| Gene | SNP | Alleles | MAF | RR (95% CI) | p RR | p overall |
|------|-----|---------|-----|-------------|------|-----------|
| FRMD7 | rs2747022 | A/g | 0.23 | 2.59 (1.18, 5.57) | 0.018 | 0.014 |
| IL1RAPL1 | rs5927786 | a/G | 0.32 | 1.83 (0.88, 3.74) | 0.105 | 0.097 |
| FGDI | rs3213533 | A/g | 0.06 | 2.18 (0.58, 7.82) | 0.251 | 0.216 |
| rs619677 | A/C | 0.04 | 0.23 (0.22, 2.49) | 0.232 | 0.221 |
| IL1RAPL1 | rs4829104 | A/G | 0.33 | 1.42 (0.88, 2.31) | 0.346 | 0.343 |
| rs4562494 | A/C | 0.04 | 0.29 (0.32, 1.55) | 0.394 | 0.410 |
| REPS2 | rs12557633 | A/G | 0.04 | 1.63 (0.32, 7.88) | 0.555 | 0.539 |
| rs7892483 | A/G | 0.06 | 0.57 (0.09, 3.37) | 0.538 | 0.549 |
| rs5972070 | A/G | 0.04 | 0.70 (0.11, 4.20) | 0.703 | 0.704 |

*Deviates from HWE.

doi:10.1371/journal.pone.0061781.t007

Sex-stratified Analyses

In the combined study, the most promising SNP among males was rs6652393 in *IL1RAPL2*. Before Bonferroni correction this SNP was significant at p<0.05 in both populations, but not in the sex-stratified analysis of females, indicating a possible sex-specific effect. *IL1RAPL2* maps to Xq22 and is a member of the IL-1 receptor family. In mice, *IL1RAPL2* is specifically expressed in the central nervous system (CNS) from embryonic day 12.5 onwards [41]. A study by Born and co-workers [42] detected expression of *IL1RAPL2* in skin, liver, placenta, and fetal brain tissues. Only

| Gene | SNP | Alleles | MAF | RR (95% CI) | p RR | p overall |
|------|-----|---------|-----|-------------|------|-----------|
| FRMD7 | rs2747022 | A/g | 0.16 | 1.80 (1.03, 3.08) | 0.040 | 0.032 |
| rs7892483 | A/g | 0.04 | 0.43 (0.12, 1.48) | 0.181 | 0.193 |
| rs4562494 | A/C | 0.46 | 0.84 (0.54, 1.37) | 0.402 | 0.416 |
| rs619677 | A/C | 0.10 | 0.73 (0.36, 1.51) | 0.409 | 0.424 |
| FGDI | rs3213533 | A/G | 0.24 | 1.14 (0.69, 1.85) | 0.614 | 0.622 |
| REPS2 | rs12557633 | A/G | 0.07 | 1.19 (0.53, 2.61) | 0.671 | 0.670 |
| IL1RAPL1 | rs5927786 | A/G | 0.43 | 1.10 (0.71, 1.69) | 0.674 | 0.684 |
| rs4829104 | A/G | 0.45 | 1.09 (0.71, 1.66) | 0.683 | 0.691 |
| rs5972070 | A/G | 0.04 | 0.85 (0.27, 2.52) | 0.766 | 0.766 |
| COLA46 | rs2299512 | A/G | 0.04 | 1.04 (0.38, 2.74) | 0.940 | 0.946 |

*Deviates from HWE.

doi:10.1371/journal.pone.0061781.t008

Table 6. Sex-stratified analyses, females.

Table 7. Replication of maternal SNPs in the US and Denmark study.

Table 8. Replication of maternal SNPs in the US, Denmark and Argentina studies.
SNPs closest to significance were in the glutamate receptor 3 gene (GRIA3) close to significance in females, whereas in the Danish study the SNP was applicable to other perinatal disorders and has several attractive features.

The hybrid design used in this study is widely applicable to other perinatal disorders and has several attractive features. However, the respective relative risks were in opposite directions of association with one SNP across both study samples in females. It has also been shown that IL-1 plays a role in the response to IL-1 [42].

In the sex-stratified analysis, there was only suggestive evidence for this. First of all, the associations did not remain significant after Bonferroni correction for multiple testing, increasing the likelihood that they were false positives. Whereas the primary studies were based on a relatively homogeneous Scandinavian population, the replication study was based on a more heterogeneous condition, we applied strict inclusion criteria in order to obtain a clearly defined phenotype.

In conclusion, our data did not provide any strong evidence for population stratification than the case-control design, and since it involves more controls than the case-parent triad design alone, it not only provides more statistical power for detecting an effect, it also allows the main effects of an exposure (genetic or environmental) to be estimated. Furthermore, our study subjects come from two relatively homogeneous populations that share a common ancestry and geography, further protecting against population stratification. Because spontaneous PTD is a heterogeneous condition, we applied strict inclusion criteria in order to obtain a clearly defined phenotype.

Although this study is based on one of the largest collections of PTD samples to date, the power to detect small genetic effects in particular may still be limited. Several associations that were significant in the two independent discovery populations from Denmark and Norway failed to replicate in a third study population from the US. There are several plausible explanations for this. First of all, the associations did not remain significant after a Bonferroni correction for multiple testing, increasing the likelihood that they were false positives. Whereas the primary studies were based on a relatively homogeneous Scandinavian population, the replication study was based on a more heterogeneous/admixed US and Argentinean population. Also, some of the variables that were used as exclusion criteria in the primary studies (for example diabetes and cerclage) were not available for the replication population. Furthermore, the replication study had a smaller sample size than the combined study, in particular for the evaluation of maternal SNPs. Finally, the case-parent triad study design in the replication study offers better protection against population stratification than the hybrid design in the primary study, but the statistical power in the former is lower because fewer control alleles are available for comparison.

In conclusion, our data did not provide any strong evidence for the involvement of X-chromosomal SNPs in the risk of spontaneous PTD. However, there were several interesting findings, such as the maternal SNP rs7473479 in FRMD7, which represents a particularly attractive candidate for further investigations in other populations.
large studies. The hybrid study design described here and the analytic opportunities provided by Haplin should prove valuable not only for the replication but also for exploring other perinatal disorders.

Supporting Information

Table S1 Maternal results for SNPs with \(p<1.00 \times 10^{-3}\).

Table S2 Fetal results for SNPs with \(p<1.00 \times 10^{-3}\).

Table S3 Sex-stratified analysis, males, \(p<1.00 \times 10^{-3}\).

Table S4 Sex-stratified analysis, females, \(p<1.00 \times 10^{-3}\).

Acknowledgments

We thank Kaare Christensen for contributing the Danish samples for the replication study. Also thanks to Hakon Boas and German Tapia for assisting in filtering of cases in the replication study. Finally, we want to thank all the participants in DNBC, MoBa, the US Prematurity Study, the Danish Prematurity Study and the Argentina Prematurity Study for making this study possible.

Author Contributions

Conceived and designed the experiments: HB BF FG HG AJ BJ MM NM PM SM AP KR IØ. Wrote the paper: SM HB RM AJ BJ MM.

References

1. Blencowe H, Cousens S, Oestergaard MZ, Chou D, Moller AB, et al. (2012) National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: a systematic analysis and implications. Lancet 379: 2162-2172.
2. Moster D, Lie RT, Marksted T (2008) Long-term medical and social consequences of preterm birth. N Engl J Med 359: 262-273.
3. Morken NV, Irving I, Kallen K, Skjaerven R, Langholz-Bloos J, et al. (2008) Reference population for international comparisons and time trend surveillance of preterm delivery proportions in three countries. BMC Womens Health 8: 16.
4. Treloar SA, Macouces GA, Mitchell EE, Martin NG (2006) Genetic influences on premature parturition in an Australian twin sample. Twin Res 3: 80-92.
5. Clausen B, Lichtenstein P, Castringsius S (2000) Genetic influence on birthweight and gestational length determined by studies in offspring of twins. Bpl 107: 373-391.
6. Kiiski ZA, De Francesco EA, Liegart L, Willemsen G, Planke J, et al. (2000) Heritability of parturition timing: an extended twin design analysis. Am J Obstet Gynecol 199: 43-e14.
7. Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B, et al. (2009) Maternal contributions to preterm delivery. Am J Epidemiol 170: 1358-1364.
8. Svensson AC, Sandin S, Cnattingius S, Reilly M, Pawitan Y, et al. (2006) Maternal effects for preterm birth: a genetic epidemiologic study of 630,000 families. Am J Epidemiol 170: 1365-1372.
9. Lundqvist M, Melve KK, Gesing HK, Skjaerven R, Ingem LS (2007) Genetic and environmental influences on birthweight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. Am J Epidemiol 165: 734-741.
10. Zeytlin J, Sael-Cubillas MJ, De Mouzon J, Rivera L, Ancel PY, et al. (2002) Fetal sex and preterm birth: are males at greater risk? Hum Reprod 27: 2762-2768.
11. McGreger JA, Leif M, Orlews M, Baron A (1992) Fetal Gender Differences in Preterm Birth: Findings in a North American Cohort. Amer J Perinatal 9: 43-48.
12. Cooperstock M, Campbell J (1996) Excess males in preterm birth: interactions with gestational age, race, and multiple birth. Obstetrics & Gynecology 88: 189-193.
13. Bry K, Teramo K, Lappalainen U, Waffarn F, Hallman M (1995) Interleukin-1 receptor antagonist in the fraternal compartment. Acta Paediatr 84: 223-236.
14. Gudmundsson J, Sulem P, Palfrad J, Berta J, Jonsdottir JT, et al. (2006) Common sequence variants on 2p15 and Xp11.22 confer susceptibility to prostate cancer. Nat Genet 40: 281-283.
15. Voight BF, Scott LJ, Steinthorhottir V, Morris AP, Dina C, et al. (2010) Twelve type 2 diabetes susceptibility loci identified through large-scale association analysis. Nat Genet 42: 579-589.
16. Nrennt AH, Nolte D, Dunne E, Nienman S, Kristensen M, et al. (1999) Refined Linkage Disequilibrium and Physical Mapping of the Gene Locus for X-Linked Dystonia-Parkinsonism (DYT3). Genomics 60: 320-329.
17. Plton A, Gauthier J, Hamdan FF, Lafreniere RG, Yang Y, et al. (2011) Systematic resequencing of X-chromosome synaptic genes in autism spectrum disorder and schizophrenia. Mol Psychiatry 16: 867-880.
18. van der Zanden LFM, van Rooij IALM, Feitz WFJ, Knight J, Donders ART, et al. (2011) Common variants in DGKK are strongly associated with risk of hypospadias. Nat Genet 43: 48-50.
19. Su MT, Lin SH, Chen YC (2011) Association of sex hormone receptor gene polymorphisms with recurrent pregnancy loss: a systematic review and meta-analysis. Fertil Steril 96: 1435-1444 e1431.
42. Born TL, Smith DE, Garla KE, Renshaw BR, Bertles JS, et al. (2000) Identification and characterization of two members of a novel class of the interleukin-1 receptor (IL-1R) family. Delineation of a new class of IL-1R-related proteins based on signaling. J Biol Chem 275: 41528.

43. Piton A, Michaud JL, Peng H, Aradhya S, Gauthier J, et al. (2008) Mutations in the calcium-related gene IL1RAPL1 are associated with autism. Human Molecular Genetics 17: 3965–3974.

44. Leviton A (1993) Preterm birth and cerebral palsy: is tumor necrosis factor the missing link? Dev Med Child Neurol 35: 533–538.