

# A Potential Novel Spontaneous Preterm Birth Gene, *AR*, Identified by Linkage and Association Analysis of X Chromosomal Markers

Minna K. Karjalainen<sup>1\*</sup>, Johanna M. Huusko<sup>1</sup>, Johanna Ulvila<sup>1</sup>, Jenni Sotkasiira<sup>1</sup>, Aino Luukkonen<sup>1</sup>, Kari Teramo<sup>2</sup>, Jevon Plunkett<sup>3</sup>, Verner Anttila<sup>4,5</sup>, Aarno Palotie<sup>4,5,6,7</sup>, Ritva Haataja<sup>1,8</sup>, Louis J. Muglia<sup>9</sup>, Mikko Hallman<sup>1</sup>

**1** Department of Pediatrics, Institute of Clinical Medicine, University of Oulu, Oulu, Finland, **2** Department of Obstetrics and Gynecology, Helsinki University Central Hospital, Helsinki, Finland, **3** Stanford University School of Medicine, Palo Alto, California, United States of America, **4** Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, **5** Wellcome Trust Sanger Institute, Cambridge, United Kingdom, **6** Department of Medical Genetics, Haartman Institute, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, **7** The Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America, **8** Biocenter Oulu, University of Oulu, Oulu, Finland, **9** Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Childrens Hospital Medical Center, Cincinnati, Ohio, United States of America

## Abstract

Preterm birth is the major cause of neonatal mortality and morbidity. In many cases, it has severe life-long consequences for the health and neurological development of the newborn child. More than 50% of all preterm births are spontaneous, and currently there is no effective prevention. Several studies suggest that genetic factors play a role in spontaneous preterm birth (SPTB). However, its genetic background is insufficiently characterized. The aim of the present study was to perform a linkage analysis of X chromosomal markers in SPTB in large northern Finnish families with recurrent SPTBs. We found a significant linkage signal (HLOD = 3.72) on chromosome locus Xq13.1 when the studied phenotype was being born preterm. There were no significant linkage signals when the studied phenotype was giving preterm deliveries. Two functional candidate genes, those encoding the androgen receptor (*AR*) and the interleukin-2 receptor gamma subunit (*IL2RG*), located near this locus were analyzed as candidates for SPTB in subsequent case-control association analyses. Nine single-nucleotide polymorphisms (SNPs) within these genes and an *AR* exon-1 CAG repeat, which was previously demonstrated to be functionally significant, were analyzed in mothers with preterm delivery ( $n = 272$ ) and their offspring ( $n = 269$ ), and in mothers with exclusively term deliveries ( $n = 201$ ) and their offspring ( $n = 199$ ), all originating from northern Finland. A replication study population consisting of individuals born preterm ( $n = 111$ ) and term ( $n = 197$ ) from southern Finland was also analyzed. Long *AR* CAG repeats ( $\geq 26$ ) were overrepresented and short repeats ( $\leq 19$ ) underrepresented in individuals born preterm compared to those born at term. Thus, our linkage and association results emphasize the role of the fetal genome in genetic predisposition to SPTB and implicate *AR* as a potential novel fetal susceptibility gene for SPTB.

**Citation:** Karjalainen MK, Huusko JM, Ulvila J, Sotkasiira J, Luukkonen A, et al. (2012) A Potential Novel Spontaneous Preterm Birth Gene, *AR*, Identified by Linkage and Association Analysis of X Chromosomal Markers. PLoS ONE 7(12): e51378. doi:10.1371/journal.pone.0051378

**Editor:** Balraj Mittal, Sanjay Gandhi Medical Institute, India

**Received:** August 8, 2012; **Accepted:** November 7, 2012; **Published:** December 5, 2012

**Copyright:** © 2012 Karjalainen 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 research was supported by the Academy of Finland (grant number 105734 to RH, 126662 and 125911 to MH), the Sigrid Juselius Foundation (MH), the Emil Aaltonen Foundation (MKK), the Alma and K.A. Snellman Foundation (MKK), the Foundation of Pediatric Research in Finland (RH), and University of Oulu Faculty of Medicine (RH). Oulu University Hospital was supported by the Ministry of Social Affairs and Health (MH), the Wellcome Trust (grant number WTO89062 to AP), the Academy of Finland Center of Excellence for Complex Disease Genetics 200923 (AP) and by a grant from the March of Dimes, the Children's Discovery Institute at Washington University School of Medicine and St. Louis Children's Hospital awarded to LJM. The Broad Institute Center for Genotyping and Analysis was supported by grant U54 RR020278 from the National Center for Research Resources. 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: minna.k.karjalainen@oulu.fi

## Introduction

Preterm birth (birth before 37 completed weeks of gestation) is a major global healthcare problem, affecting approximately 13 million people every year [1]. Approximately 12% of all infants are born preterm in the United States; this figure is 6% in Europe [2]. Seventy percent of mortality and 75% morbidity in the neonatal period are caused by preterm birth, and preterm infants also face a substantial risk of several serious morbidities that may have life-long consequences [3]. Multiple gestation, maternal chronic diseases, fetal malformations, and infections are among the factors

known to increase the risk of preterm birth [4]. Preterm birth tends to recur in families, with mothers with previous preterm deliveries being at a significantly increased risk of preterm delivery in subsequent pregnancies [5–9]. Preterm birth also occurs across generations and sibships, with the risk of preterm delivery shared between mothers and daughters, and sisters [7,8,10–13]. Aggregation of preterm deliveries in families suggests a significant role for genetic factors in preterm birth. Approximately 70% of all preterm births are spontaneous, and no environmental risk factor can be identified for most of these births [4]. Thus, it is likely that

several spontaneous preterm births (SPTBs) are explained by genetic factors.

Because initiation of parturition is preceded by physiological changes occurring in both the mother and fetus, it is possible that both maternal and fetal genes may predispose to preterm birth [14]. Most large population-based studies suggest that maternal genetic factors play the most significant role with little or a substantially smaller contribution of the fetal genes [8,12,13]. Some genetic modeling studies have further estimated that, in addition to environmental factors, the variation in gestational age is explained by both maternal and fetal genetic factors [15–17], while other studies have failed to find a significant contribution of the fetal genes [12,13]. Overall, these studies suggest that the genetic influences of SPTB are most likely to be either direct effects of maternal genes acting in the mother, of maternally-inherited mitochondrial genes acting in the fetus, or of maternally-inherited fetal genes. Knowledge about the actual genes predisposing to SPTB is insufficient. So far, the significance of specific candidate genes in SPTB has been assessed in numerous case-control studies [18], focusing mostly on genes involved in the infection and inflammation pathway, such as those encoding tumor necrosis factor  $\alpha$  (TNF- $\alpha$ ) and several interleukins [19–23]. Although significant associations of both maternal and fetal genetic factors have been reported in several studies, most of the studies have been limited in several respects, including small sample sizes, inconsistent phenotype definitions, and failure to replicate in subsequent studies and across populations [18]. Recent more comprehensive association studies have implied a role for genes involved in multiple pathways in SPTB emphasizing the heterogeneity in the SPTB phenotype; these associating genes include the maternal *FSHR* and *TIMP2* genes, encoding follicle-stimulating hormone receptor and tissue inhibitor of metalloproteinase 2, respectively [24,25], and the fetal *IL6R* and *COL5A2* genes, encoding interleukin-6 receptor and collagen type V, alpha 2, respectively [25,26]. To date, genome-wide case-control association analyses of SPTB have not been reported.

Recently, we reported the results of the first linkage analysis of SPTB in seven large northern Finnish families with recurrent SPTBs using a high-density array of autosomal markers [27]. The gene encoding insulin-like growth factor receptor 1 (IGF-1R) was identified as a potential SPTB-predisposing gene. In the present study, linkage analysis was performed for these same families using X chromosomal markers. After linkage analysis, two genes, encoding the androgen receptor (AR) and the interleukin-2 receptor gamma subunit (IL2RG), located near a significant linkage signal were further analyzed as candidates for SPTB in case-control populations originating from northern and southern Finland.

## Materials and Methods

### Ethics Statement

The study was approved by the Ethics Committees of Oulu University Hospital and Helsinki University Central Hospital. Written informed consent was obtained from the subjects participating in the study.

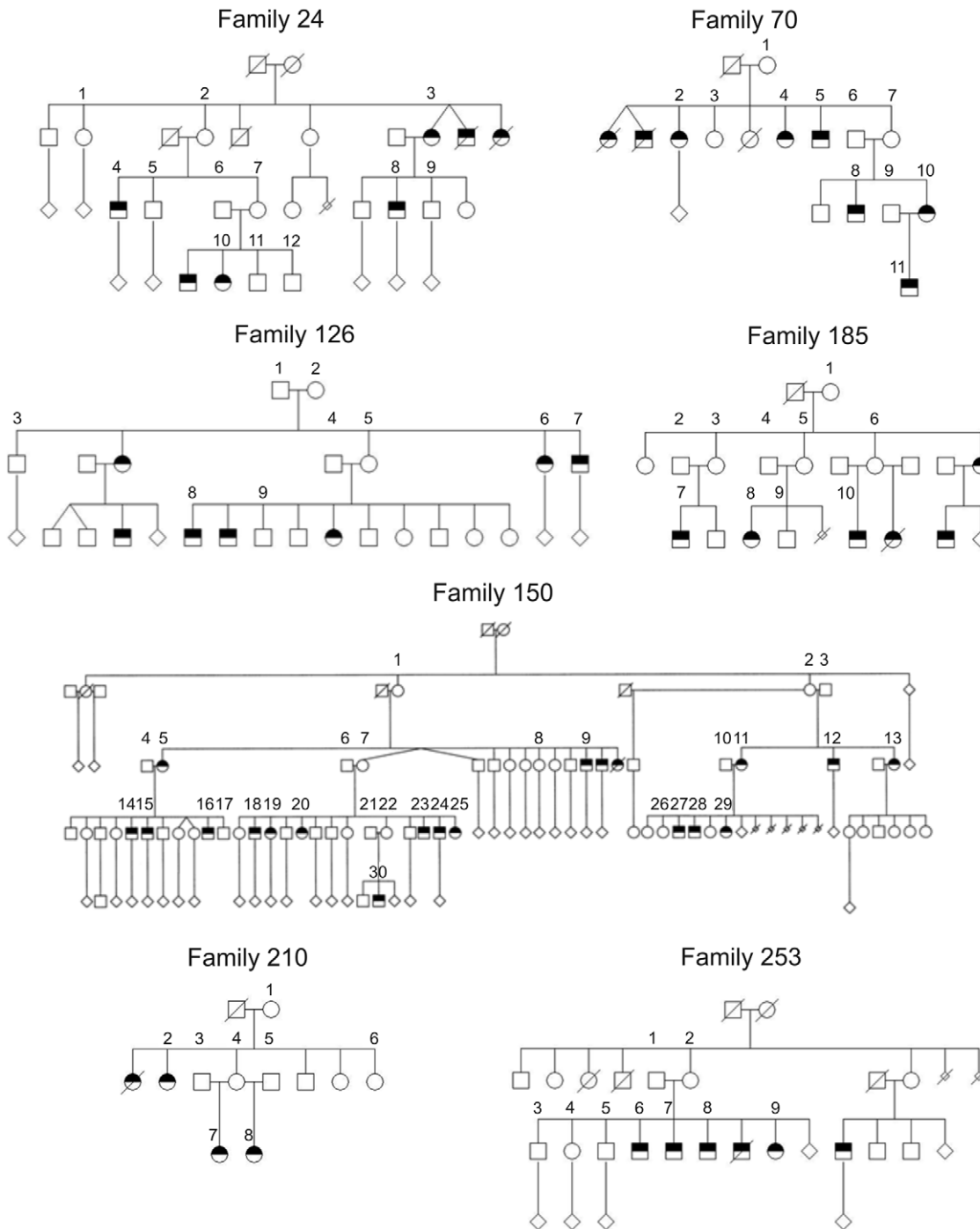
### Study Populations

**Definition of SPTB and inclusion criteria for the SPTB cases.** SPTB was defined as spontaneous preterm birth occurring before 36 completed weeks of gestation. Preterm deliveries occurring both with intact membranes and those following preterm premature rupture of fetal membranes (PPROM, defined as leakage of amniotic fluid as the presenting symptom before the

onset of contractions) were included among the SPTBs. Deliveries involving known risk factors for preterm birth (multiple gestation, polyhydramnios, septic or acute systemic infection of the mother, diseases of the mother that possibly affect the pregnancy outcome including those influencing cardio-pulmonary, hepato-renal, or endocrine functions, and chronic inflammatory disease, alcohol or narcotic use, accidents, and fetuses with congenital anomalies) and all medically indicated preterm births without labor (preeclampsia, intrauterine growth restriction, and placental abruption) were excluded. All analyzed subjects were of Finnish origin. The overall rate of preterm birth is approximately 5.5% in Finland [28]. Considering that approximately 70% of all preterm deliveries are spontaneous and that some spontaneous deliveries were further excluded from our study due to border-line gestational ages (birth at gestational age ranging from 36 weeks + 0 days to 36 weeks + 6 days; approximately 35% of all preterm births), multiple gestation (approximately 30% of all preterm births and 15% of all deliveries with a spontaneous onset) or any of the other above-mentioned exclusion criteria (leading to exclusion of approximately 15% of all deliveries), the proportion of all preterm deliveries fulfilling our selection criteria is estimated to be approximately 20%, which corresponds to approximately 1.1% of all births in Finland.

**Families in Linkage Analysis.** The families used in the current study were the same as those used in our recently published linkage study performed with autosomal markers [27]. Families with recurrent SPTB were selected retrospectively from birth diaries of Oulu University Hospital during 1973–2003, and prospectively during 2003–2005. Family interviews revealed 20 large families with SPTBs in two or more generations or in first cousins. To ensure that the families were nonconsanguineous, a genealogical study was performed according to published criteria [29]. An examination of the microfiche copies available in the provincial and national archive of Finland did not reveal close consanguinity or common residence dating from the 17<sup>th</sup> century among the families. Finally, the families with apparent maternal inheritance of the SPTB phenotype were selected for the study because most previous family studies suggest that maternal genes or maternal genes acting in the fetus (rather than paternal) influence the duration of pregnancy and the inheritance of preterm delivery [8,12,13,16]. Whole-blood DNA samples ( $n = 89$ ) were collected from as many family members as possible. The pedigrees of the seven families in the linkage analysis are depicted in Figure 1; gestational ages and years of birth for the family members born preterm are shown in Table S1.

**Study Sample in the Initial Case-Control Association Analysis.** Mothers with one or more spontaneous preterm deliveries and their preterm offspring (born at <36 gestational weeks) born in the Oulu University Hospital region (northern Finland) were selected for the study during the same time period as the linkage analysis families. Cases included mothers and their offspring from both families with multiple (two or more) SPTBs as well as those with a single SPTB in the family. For 39 families with multiple SPTBs, DNA samples of two or more individuals born preterm were available. To include deliveries involving as few as possible other than genetic risk factors for SPTB, a single individual born preterm was selected for analysis from these families according to the following low-risk criteria: 1) age of the mother between 20 and 35 years at the time of the birth of the individual born preterm (criterion applied because low and high maternal ages are risk factors for preterm birth [4]); 9 individuals selected using this criterion), and 2) no deliveries within the preceding 2 years, or the longest time interval from the previous delivery (criterion applied because short inter-pregnancy interval is



**Figure 1. Pedigrees of the seven families with recurrent spontaneous preterm births analyzed in the linkage analysis of X chromosomal markers.** Individuals born preterm are shown by half-black symbols. Squares represent males and circles females. Individuals analyzed in the linkage analysis are indicated by a number ( $n = 89$ ). The diamonds denote an unspecified number of term infants, the smaller diamonds with a line through the symbol represent a miscarriage, and the lines through the other symbols indicate that the person is deceased. Gestational ages and years of birth for the family members born preterm are shown in Table S1. Linkage analysis was performed in two settings: 1) being spontaneously born preterm as the phenotype (affected offspring phenotype,  $n_{affected} = 41$ ), and 2) giving spontaneous preterm births as the phenotype (affected mother phenotype,  $n_{affected} = 21$ ). These pedigrees have been previously shown in the supplementary material of our linkage study of SPTB performed using autosomal markers [27]. doi:10.1371/journal.pone.0051378.g001

a risk factor for preterm birth [4]; 22 individuals selected using this criterion). If these criteria were not applicable, a female individual was selected over a male individual (criterion applied because a somewhat higher proportion of male fetuses are born preterm

compared to female fetuses [30]; 8 individuals selected using this criterion). None of the SPTB cases were members of any of the linkage analysis families. The clinical characteristics of the case study population are shown in Table 1. Whole-blood or buccal

samples were obtained from cases (141 whole-blood and 131 buccal samples of SPTB mothers; 104 whole-blood and 165 buccal samples of SPTB offspring), and when available, also from fathers ( $n = 226$ ; 117 whole-blood and 109 buccal samples). The control population comprised mothers with exclusively at least three term (gestational age  $\geq 37$  wk) deliveries ( $n = 201$ ) without any pregnancy- or labor-associated complications (intrauterine growth restriction, placental abruption, polyhydramnios, preeclampsia, fetuses with congenital anomalies, requirement of special care of the newborn) and their singleton offspring ( $n = 199$ ; 110 males and 89 females) collected prospectively following delivery in the Oulu University Hospital during 2004–2007. The selection criterion “exclusively at least three term deliveries” was applied because women who have had several uncomplicated term deliveries without any preterm deliveries are likely to be at low risk to give birth preterm in subsequent pregnancies and thus most unlikely to carry genetic factors that predispose to SPTB. Whole-blood or buccal samples were obtained from controls (143 whole-blood and 58 buccal samples of control mothers; 199 whole-blood samples of control offspring).

**Replication Study Sample for the Case-Control Association Analysis.** Due to a significant linkage signal with the affected offspring phenotype and association of a fetal polymorphism in the initial case-control population, a replication study sample consisting of individuals born preterm (at  $< 36$  gestational weeks;  $n = 111$ ; 56 males and 55 females) and term ( $n = 197$ ; 94 males and 103 females) collected in the Helsinki University Central Hospital (southern Finland) during 2005–2008 was analyzed. The cases (SPTB offspring) were selected using the same criteria as described for the initial study sample. Similarly to the initial offspring control population, the replication control population was comprised of individuals born at gestational age  $\geq 37$  wk for mothers who had experienced no pregnancy- or labor-associated complications. However, it should be noted that the replication control study sample did not meet one of the criteria for the selection of the initial controls, that is, the occurrence of exclusively at least three term deliveries in the family. Whole-blood samples were obtained from cases and controls.

## DNA Sample Preparation

Genomic DNA was extracted from whole blood or buccal cells using standard methods as described [27]. Buccal DNA was whole-genome amplified (WGA) using the Illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare Sciences, Cardiff, UK) and purified using Illustra Microspin G-50 columns (GE Healthcare Sciences). Nontemplate reactions were included in the WGA reactions to control for contamination. The quality of the WGA samples was controlled in the set-up stage by genotyping amplified DNA samples in duplicate reactions simultaneously with the corresponding unamplified samples and whole-blood DNA samples obtained from the same individuals using a test set of single-nucleotide polymorphisms (SNPs); this yielded  $> 99\%$  consistent genotypes. Samples with low genotyping success rate were further excluded from analyses.

## Genotyping and Data Analysis in the Linkage Study

DNA samples for the linkage analysis were genotyped in the Broad Institute Center for Genotyping and Analysis (CGA) using the Affymetrix Genome-Wide Human SNP Array 5.0Kb.

Before linkage analysis, the Affymetrix SNP array data was processed using PLINK, v. 1.02 [31]. A 1-Mb-to-1 cM map was used as justified before [32]. SNPs occurring at a minor allele frequency (MAF)  $< 0.08$  in the studied subjects, genotyping failure  $> 0.1$ , Mendelian errors, or deviation from HWE ( $p < 0.001$ ) were excluded. The absence of Mendelian errors was confirmed using Pedcheck [33]. After these pruning steps, 4,870 X chromosomal SNPs with an average spacing of 0.031 Mb remained for linkage analysis.

Linkage analysis of X chromosomal markers was performed in two settings: 1) being spontaneously born preterm as the phenotype (affected offspring phenotype,  $n_{affected} = 41$ ), and 2) giving spontaneous preterm births as the phenotype (affected mother phenotype,  $n_{affected} = 21$ ). Two-point linkage analysis was applied to avoid bias potentially arising in multipoint analyses due to markers displaying linkage disequilibrium (LD) or missing parental genotype data [34]. We used a dominant, low-penetrance model with a disease allele frequency set to 0.001 and penetrances of 0.001 for the homozygotes, 0.001 for the heterozygotes and 0

**Table 1.** Clinical characteristics of the case study populations of spontaneous preterm birth.

Characteristic	Initial population from northern Finland		Replication population from southern Finland
	Offspring	Mothers	Offspring
Total $n$	269 <sup>a</sup>	272 <sup>a</sup>	111
Single spontaneous preterm delivery in the family, $n$	194	199	95
Two spontaneous preterm deliveries in the family, $n$	60	57	16
3–5 spontaneous preterm deliveries in the family, $n$	15	16	0
Gestational age, wk	31.8 $\pm$ 2.99 (24.1–35.9) <sup>b</sup>		34.0 $\pm$ 1.92 (27.6–35.9) <sup>b</sup>
Gestational age $< 32$ wk, $n$	109		15
Gestational age $\geq 32$ wk, $n$	160		96
Birth weight, g	1,864 $\pm$ 618 (538–3,070) <sup>b</sup>		2,361 $\pm$ 492 (1,130–3,598) <sup>b</sup>
PPROM, yes/no	111/158		67/44
Male/female	149/120		56/55
Maternal smoking during pregnancy, yes/no/unknown	34/170/65		33/175/64
			6/105/0

<sup>a</sup>Mismatch in the numbers of mothers and offspring is due to occasional low quality of DNA samples.

<sup>b</sup>Mean  $\pm$  standard deviation (range).

doi:10.1371/journal.pone.0051378.t001

for the wild-type homozygotes. The effect of misspecifying the actual mode of inheritance (MOI) is minimized in this model. Therefore, this model is nearly equivalent to a model-free analysis while the high power of parametric analysis to detect linkage is retained [35,36]. Analyze, v. 1.9.3. BETA [37], which uses FASTLINK 4.1P for calculation of pedigree likelihoods, was used in parametric two-point linkage analysis of SPTB. Heterogeneity LOD (HLOD) scores and the estimates of proportion of linked families ( $\alpha$ ) and recombination fraction ( $\theta$ ) were used to assess parametric linkage in the presence of heterogeneity. An HLOD score  $>3.3$  was considered as a significant signal of linkage.

### Selection of Genes For the Case-Control Study

Due to a significant linkage signal for marker rs6525299 on chromosome locus Xq13.1, we screened the region ( $\sim 5$  Mb) surrounding the linkage peak for the presence of plausible candidate genes for SPTB and chose to analyze two genes near this locus (*AR* and *IL2RG*, located upstream and downstream from rs6525299, respectively, each at a distance of  $\sim 2$  Mb) as candidates for SPTB. The *AR* and *IL2RG* genes were chosen for analysis based on their function: *IL2RG*, encoding for a signaling component (the common gamma chain) of several interleukin receptors with essential roles in the immunity [38], emerged as an interesting candidate because infections are estimated to be involved in approximately 30–40% of preterm births [39]; *AR* emerged as an interesting candidate due to the role of androgen receptor in the development of several tissues and due to involvement of androgens in the biosynthesis of estrogen, which is essentially involved in the regulation of parturition initiation [40,41]. Altogether 65 and 41 SNPs (at an average SNP spacing of 0.03 Mb) analyzed in the Affymetrix linkage array were located within the interval spanning from the most significant linkage peak (rs6525299) to the *AR* and *IL2RG* genes, respectively. In linkage analysis, suggestive or nearly suggestive linkage signals (HLODs 1.5–2.7) were detected for several SNPs located between the linkage peak and these two genes, while some of the SNPs located in this region showed no signals (data not shown). According to 1000 Genomes Project sequence data, SNP rs6525299 (the marker yielding the significant linkage signal) occurs at a MAF of 0.16 in the Finnish population, while many of the SNPs of the Affymetrix linkage array located between this SNP and the *AR* and *IL2RG* genes occur at a MAF of 0.25–0.35. Ability to detect LD between a SNP and a given genetic variant is influenced by allele frequencies of these variants [42–44]. Thus, a SNP may display linkage with a given genetic variant even though other markers located close to this variant show weaker linkage signals.

### Selection of Polymorphisms and Genotyping in the Case-Control Study

Tagging SNPs (tSNPs) were selected for the *AR* and *IL2RG* genes using HapMap data (release23a/phaseII) from the CEU (CEPH; Utah residents with ancestry from northern and western Europe) population. Using pairwise tagging with MAF and  $r^2$  cutoffs of 0.05 and 1.0, respectively, six tSNPs (rs5918757, rs5919393, rs2361634, rs5918762, rs12014709, and rs5031002) were identified for the *AR* gene. In addition, two exonic *AR* polymorphisms were further selected for analysis due to previously reported associations of these polymorphisms to pregnancy-related conditions and/or due to reported functional consequences: the *AR* exon-1 SNP rs6152 (Glu213Glu) was selected because it was previously shown to associate with recurrent spontaneous abortions [45], and the exon-1 CAG repeat polymorphism ( $CAG_n$ ) was selected because previous studies have shown that this polymorphism affects the AR-mediated transactivation [46], correlates

with the concentration of androgens [47], and associates with recurrent spontaneous abortions [48]. For the *IL2RG* gene, no tSNPs were identified even with a MAF cutoff of 0. Therefore, we analyzed all common SNPs (MAF  $>0.05$ ) indicated for this gene (rs4612655, rs4073907, rs11574625, rs7880291, and rs28743771) in the dbSNP database ([www.ncbi.nlm.nih.gov/SNP](http://www.ncbi.nlm.nih.gov/SNP)). After validation of the genotyping procedure, one SNP in *AR* (rs2361634) and two SNPs in *IL2RG* (rs4073907 and rs11574625) were excluded from the analysis.

SNP genotyping was performed using the SNaPshot Multiplex kit (Applied Biosystems, Carlsbad, CA, USA). The *AR* CAG repeat was genotyped using fragment analysis. A DNA fragment containing the CAG repeat (at physical position 66,763,874–66,944,119 in chromosome X) in the *AR* gene was amplified using the polymerase chain reaction (PCR) with the forward and reverse primers 5'-ACC GAG GAG CTT TCC AGA AT-3' and 5'-CTG GGA CGC AAC CTC TCT C-3' respectively. The primers were purchased from Oligomer (Helsinki, Finland) and the forward primer was labeled with the fluorescent 6-FAM dye. The sizes of the PCR fragments were determined using the ABI3130xl Genetic Analyzer (Applied Biosystems), and the GeneScan-500 LIZ Size Standard (Applied Biosystems) was used as a size standard. The fragment lengths were analyzed using the GeneMapper software, v. 3.7 (Applied Biosystems). PCR resulted in fragments with sizes ranging from 300 to 393 bp. To determine the exact number of CAG repeats, a part of the samples were sequenced. Sequencing revealed that fragment analysis had resulted in a 12-bp downwards shift in the actual size of the PCR products. To correct for this bias, the fragment sizes were adjusted for the shift. The shortest and longest PCR fragments (adjusted sizes of 312 and 405 bp) corresponded to 11 and 42 CAG repeats, respectively.

### Statistical Analysis in the Case-Control Study

**Quality control.** Before association analysis, the case-control data was checked for Mendelian inconsistencies using PedCheck [33]. Any inconsistent genotypes were recoded as missing.

**Analysis of SNP frequencies.** SNP allele distributions were analyzed with logistic regression (additive model) with sex as a covariate using PLINK, v. 1.07 [31]. Genotypes of female individuals were also analyzed under the dominant and recessive models. Case-control comparisons of haplotypes were performed using Haploview v. 4.2 [49].

**Analysis of  $AR$   $CAG_n$  distribution.** Differences in *AR*  $CAG_n$  distributions were analyzed using the nonparametric Mann-Whitney  $U$  test using Predictive Analytics Software (PASW) Statistics, v. 17.0.3 (IBM, SPSS, Inc.). To equalize the contributions of male and female alleles, the tests were performed with the hemizygous males treated as homozygotes. The  $CAG_n$  distributions were divided into quintiles and further analyzed using the Pearson  $\chi^2$  test with PASW Statistics. The  $CAG_n$  analyses were replicated using biallelic means for the females. Family-based association tests of  $CAG_n$  were performed using Pseudomarker, v. 1.0.6b [50], and FBAT, v. 2.0.2c [51], under the dominant model. Pseudomarker was used to test for linkage in the presence of LD (linkage given LD test) in mother-father-offspring trios or mother-offspring pairs; this test is analogous to the model-free transmission disequilibrium test. Pseudomarker can handle missing parental data and different data structures such as mixtures of singletons and families jointly; therefore both the mother-father-affected offspring trios and mother-control offspring pairs are informative in this analysis. FBAT was further used to determine significances in preferential transmission of individual  $CAG_n$  alleles in mother-father-affected offspring trios.

**Significance threshold and power consideration.** Using the Bonferroni correction to account for testing of ten polymorphisms (nine SNPs and the *AR* CAG<sub>n</sub>), a  $p$  value  $\leq 0.005$  was considered significant in case-control comparisons. Power consideration for the initial case-control study (considering multiple testing of ten polymorphisms,  $\alpha = 0.005$ ; prevalence of 0.05; MAFs ranging from 0.15 down to 0.05; allelic test under additive model): the sample size was estimated to provide 80% power to detect relative risks of 1.92–2.55 and 2.84–4.10 for the heterozygous and homozygous genotypes, respectively, both in the offspring and mothers.

### Genome Coordinates

Chromosomal positions refer to NCBI Build 37.2 of the human genome.

### Web Resources

1000 Genomes Project: <http://www.1000genomes.org>.

Genetic Power Calculator: <http://pngu.mgh.harvard.edu/~purcell/gpc>.

International HapMap project: [www.hapmap.org](http://www.hapmap.org).

### Results

Some of these results were described in a preliminary form in the Ph.D. thesis of the first author [52].

#### Linkage Analysis of X chromosomal Markers in SPTB

Seven northern Finnish families with recurrent SPTB were selected for linkage analysis using stringent phenotypic criteria, with deliveries involving known risk factors of SPTB excluded. Because family studies suggest that maternal factors may play an important role in the inheritance of SPTB [8,12,13,16], only families with apparent maternal inheritance were included. Genealogical surveys ensured that the selected families were not consanguineous.

Parametric two-point linkage analysis of SPTB was performed using a high-density panel of X chromosomal markers. HLOD scores are depicted in Figure 2. For the affected offspring phenotype, a significant linkage signal was obtained for SNP rs6525299 ( $\theta = 0.00$ ,  $\alpha = 1.00$ ) on chromosome locus Xq13.1 (HLOD = 3.72). HLOD scores for the three markers that yielded the highest signals in linkage analysis are shown separately for each of the families in Table 2. Two families (126 and 253) displayed HLOD scores of 0.00 for these markers, indicating that there was no evidence of linkage in these families. For the affected mother phenotype, no significant linkage signals were detected (Figure 2); HLOD<sub>max</sub> of 1.83 was detected for SNP rs4240068 on Xq21.2.

#### Case-Control Association Analysis of *AR* and *IL2RG* Polymorphisms in the Initial SPTB Study Population Originating from Northern Finland

The genes investigated as candidates for SPTB were selected based on the location of the marker giving the significant signal in linkage analysis. Based on their function, two genes located near the marker yielding the maximum HLOD in linkage analysis (rs6525299) emerged as interesting candidates for SPTB: those encoding androgen receptor (*AR*, MIM \*313700, located on Xq12) and interleukin-2 receptor gamma subunit (*IL2RG*, MIM \*308380, located on Xq13.1). These genes are located at a distance of approximately 2 Mb from SNP rs6525299.

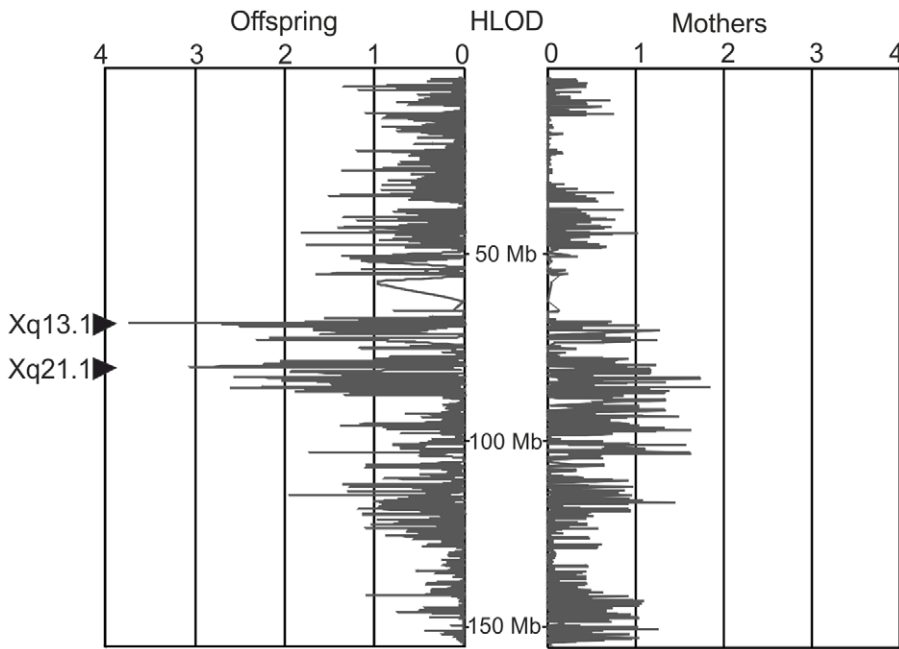
Case-control association analysis of the selected genes in SPTB was performed in a population consisting of mothers with preterm

and term deliveries and their preterm and term-born offspring, respectively. Tagging SNPs representative of common variation within each gene were selected using the HapMap CEU population. In addition, other polymorphisms with potential biological significance were included. For the offspring, the SNP analyses were first performed in mixed-sex samples with sex as a covariate. *IL2RG* SNP rs7880291 was not polymorphic in the study population. Results of logistic regression analysis performed for the *AR* and *IL2RG* SNPs are shown in Table 3; genotype distributions of these polymorphisms did not differ between SPTB and term mothers or offspring. Analysis of female individuals under the dominant and recessive models did not reveal any significant associations (data not shown). Five *AR* SNPs (rs6152, rs5918757, rs5919393, rs5918762, and rs12014709) and two *IL2RG* SNPs (rs4612544 and rs28743771) resided in haplotype blocks. Altogether three 5-SNP *AR* haplotypes and two 2-SNP *IL2RG* haplotypes with a frequency  $> 0.01$  were predicted (Table 4). Analysis of these haplotypes did not reveal any significant differences in haplotype frequencies between SPTB offspring or mothers and their controls (Table 4).

The distributions of *AR* CAG repeat numbers in each group studied are shown in Figure S1. The observed distributions are consistent with those previously reported for individuals of European descent [53]. Nearly all of the CAG<sub>n</sub> alleles in our population were in the range of 13–30, which falls in the normal range of repeat lengths (10–36). Very long alleles (38 or 42 repeats) were detected in four individuals (two SPTB mothers and their SPTB offspring). The median number of *AR* CAG repeats was 22 in all groups studied, except in female SPTB offspring, who had median of 23. The distribution of CAG repeats differed significantly between individuals born preterm and those born at term (Mann Whitney  $U$  test,  $p = 0.0006$ ), with carriers of the longest repeats ( $\geq 26$ ) being at a more than 2-fold risk (OR = 2.45, 95% CI 1.46–4.13) of being born preterm compared to the carriers of the shortest repeats ( $\leq 19$ ) (Table 5), suggesting that longer alleles may predispose to SPTB. The difference was clearer in females than in males (Table S2). Differences in CAG<sub>n</sub> distributions between SPTB and term offspring were also significant when biallelic means were used in the comparisons (data not shown). There was no significant difference in the *AR* CAG<sub>n</sub> distribution between SPTB mothers and mothers with term deliveries (Mann-Whitney  $U$  test,  $p = 0.13$ ; mean  $\pm$  standard deviation, SD,  $22.41 \pm 3.02$  vs.  $22.06 \pm 2.81$ , respectively). The fathers of the preterm offspring had a CAG<sub>n</sub> mean of  $22.15 \pm 2.75$ .

To evaluate whether the predisposing *AR* CAG<sub>n</sub> alleles had been inherited from the mother or the father, distributions of maternally and paternally inherited alleles were compared between SPTB and term offspring. This revealed that distributions of both maternally and paternally inherited CAG<sub>n</sub> alleles differed between cases and controls; the difference was more significant for the maternal alleles (Mann Whitney  $U$  test,  $p = 0.0045$ , mean  $\pm$  SD,  $22.59 \pm 3.19$  vs.  $21.74 \pm 2.63$ , respectively, for maternal alleles; Mann Whitney  $U$  test,  $p = 0.040$ ,  $22.77 \pm 2.82$  vs.  $21.84 \pm 2.72$ , respectively, for paternal alleles). When the repeat lengths were divided into quintiles, under- and overrepresentation of the lowest and highest groups, respectively, could be seen in both maternally and paternally inherited alleles (data not shown).

Family-based tests were performed to further investigate whether specific CAG<sub>n</sub> alleles had been preferentially transmitted to affected individuals, i.e. offspring born preterm. The linkage given LD test showed evidence for disequilibrium in the transmission of *AR* CAG<sub>n</sub> alleles ( $p = 0.004$ ), indicating that this polymorphism is linked to the affected offspring phenotype in the families of the case-control population. FBAT analysis further



**Figure 2. HLOD scores in parametric linkage analysis of X chromosomal markers in spontaneous preterm birth.** HLODs for the affected offspring phenotype are shown on the left and those for the affected mother phenotype on the right. The position of the marker yielding a significant linkage signal (SNP rs6525299 on Xq13.1, HLOD of 3.72) for the affected offspring phenotype is shown. The position of the markers yielding the second highest HLODs (SNPs rs11266953 and rs7056400 on Xq21.1, HLODs of 3.05) is also indicated. These three SNPs are intergenic. doi:10.1371/journal.pone.0051378.g002

suggested that the allele with 19 repeats had been under-transmitted and the allele with 24 repeats overtransmitted to preterm-born offspring (Table 6), again suggesting that shorter alleles may be protective and longer alleles predisposing to the affected offspring phenotype.

Allele or haplotype frequencies of any of the studied polymorphisms, including *AR* CAG<sub>n</sub>, did not differ between SPTB mothers or offspring with or without PPRM, according to gender, or between individuals born very preterm (gestational age <32 wk) and moderately preterm (gestational age ≥32 wk) (data not

shown). The frequencies and *AR* CAG<sub>n</sub> distributions were similar in offspring from families with multiple and single SPTBs (data not shown).

### Case-Control Association Analysis of *AR* and *IL2RG* Polymorphisms in the Replication SPTB Study Population Originating from Southern Finland

Because association analyses performed in the initial case-control study population revealed an association of a fetal polymorphism with SPTB, the case-control analyses were repeated in a study population consisting of preterm and term offspring originating from the southern Finland, where the population is known to be genetically more diverse compared to northern Finland [54]. Similarly to the initial population, none of the analyzed *AR* and *IL2RG* SNPs associated with SPTB in the replication population (data not shown). The range of the *AR* CAG<sub>n</sub> repeats was 11–40. The allele with 40 repeats was detected in a single individual born preterm while rest of the alleles were ≤32 repeats in length. Similarly to the population of northern Finland, the longer alleles tended to be more frequent in SPTB than in term offspring but the difference in the distribution was not significant (Mann Whitney *U* test, *p* = 0.67; Table 5). However, similarly to the northern Finnish population, carriers of the longest CAG repeats (≥26) were at a more than 2-fold risk (OR 2.25, 95% CI 1.14–4.43) of being born preterm compared to the carriers of the shortest repeats (≤19) (Table 5), further suggesting that the long fetal *AR* CAG<sub>n</sub> alleles may be involved in giving predisposition to SPTB. Finally, when the populations of northern and southern Finland were combined, the difference in CAG<sub>n</sub> distribution between SPTB and term offspring was significant (Mann Whitney *U* test, *p* = 0.0027; Table 5).

**Table 2. SNPs with highest HLOD scores in parametric linkage analysis of spontaneous preterm birth with the affected fetus phenotype.**

	rs6525299	rs11266593	rs7056400
<b>Locus</b>	Xq13.1	Xq21.1	Xq21.1
<b>Physical position</b>	68,623,050	80,124,948	80,173,940
<b>Overall HLOD</b>	3.72 <sup>a</sup>	3.05 <sup>a</sup>	3.05 <sup>a</sup>
<b>HLOD for each family:</b>			
Family 24	0.42	0.01	0.01
Family 70	1.38	1.72	1.72
Family 126	0.00	0.00	0.00
Family 150	0.69	0.69	0.69
Family 185	0.80	0.20	0.20
Family 210	0.42	0.42	0.42
Family 253	0.00	0.00	0.00

<sup>a</sup>*θ* = 0.00, *α* = 1.00.  
doi:10.1371/journal.pone.0051378.t002



**Table 3.** Logistic regression analysis of *AR* and *IL2RG* polymorphisms in SPTB in the initial case-control population.

Gene and polymorphism	Position	Minor allele frequency		Logistic regression (additive model)	
		SPTB	Term	<i>p</i>	OR (95% CI)
<b>Mothers</b>		<i>n</i> = 272	<i>n</i> = 201		
<i>AR</i> , rs6152	exon 1, Glu213Glu	0.146	0.150	0.88	0.97 (0.67–1.42)
<i>AR</i> , rs5918757	intron 1	0.139	0.154	0.55	0.89 (0.60–1.31)
<i>AR</i> , rs5919393	intron 1	0.148	0.154	0.80	0.95 (0.66–1.38)
<i>AR</i> , rs5918762	intron 3	0.139	0.145	0.80	0.95 (0.64–1.41)
<i>AR</i> , rs12014709	intron 5	0.047	0.067	0.22	0.70 (0.39–1.23)
<i>AR</i> , rs5031002	intron 6	0.006	0.025	0.03 <sup>a</sup>	0.23 (0.06–0.86)
<i>IL2RG</i> , rs4612544	3'	0.090	0.122	0.11	0.69 (0.45–1.09)
<i>IL2RG</i> , rs7880291	intron 1	0	0	-	-
<i>IL2RG</i> , rs28743771	intron 1	0.095	0.124	0.17	0.73 (0.47–1.15)
<b>Offspring</b>		<i>n</i> = 269	<i>n</i> = 199		
<i>AR</i> , rs6152	exon 1, Glu213Glu	0.126	0.126	1.00	1.00 (0.63–1.60)
<i>AR</i> , rs5918757	intron 1	0.118	0.122	0.87	0.96 (0.58–1.58)
<i>AR</i> , rs5919393	intron 1	0.128	0.126	0.94	1.02 (0.65–1.61)
<i>AR</i> , rs5918762	intron 3	0.106	0.113	0.79	0.93 (0.56–1.56)
<i>AR</i> , rs12014709	intron 5	0.053	0.042	0.55	1.23 (0.62–2.44)
<i>AR</i> , rs5031002	intron 6	0.006	0.020	0.15	0.29 (0.05–1.53)
<i>IL2RG</i> , rs4612544	3'	0.102	0.102	0.99	1.00 (0.60–1.67)
<i>IL2RG</i> , rs7880291	intron 1	0	0	-	-
<i>IL2RG</i> , rs28743771	intron 1	0.103	0.108	0.85	0.95 (0.57–1.59)

<sup>a</sup>Not significant at the Bonferroni-corrected significance level ( $p \leq 0.005$  considered significant).  
doi:10.1371/journal.pone.0051378.t003

## Discussion

We recently reported the first linkage analysis of SPTB, which was performed for autosomal markers in seven families with

**Table 4.** Association analysis of *AR* and *IL2RG* haplotypes in SPTB in the initial case-control population.

Gene	Haplotype	Frequency		<i>p</i>
		SPTB	Term	
<b>Mothers</b>		<i>n</i> = 272	<i>n</i> = 201	
<i>AR</i> <sup>a</sup>	GAAC	0.851	0.847	0.85
	AGGT	0.098	0.087	0.57
	AGTG	0.049	0.064	0.33
<i>IL2RG</i> <sup>b</sup>	AA	0.907	0.876	0.13
	GG	0.091	0.122	0.14
<b>Offspring</b>		<i>n</i> = 269	<i>n</i> = 199	
<i>AR</i> <sup>a</sup>	GAAC	0.880	0.874	0.83
	AGGT	0.067	0.082	0.47
	AGTG	0.047	0.043	0.83
<i>IL2RG</i> <sup>b</sup>	AA	0.899	0.891	0.73
	GG	0.098	0.102	0.85

<sup>a</sup>SNPs in the haplotype: rs6152, rs5918757, rs5919393, rs5918762 and rs12014709.

<sup>b</sup>SNPs in the haplotype: rs4612544 and rs28743771.

doi:10.1371/journal.pone.0051378.t004

recurrent SPTB [27]. In the present study, linkage analysis of X chromosomal markers in these same families revealed a significant linkage signal on chromosome locus Xq13.1 (HLOD of 3.72 for SNP rs6525299) when the phenotype was being born preterm (affected offspring phenotype). Two genes near this locus, *AR* and *IL2RG*, were investigated in subsequent case-control association analysis of mothers with preterm and term deliveries and their offspring. Finally, our results indicated fetal *AR* as a novel gene potentially involved in genetic predisposition to SPTB.

The length of the *AR* exon-1 CAG repeat was associated with SPTB in the offspring, with longer repeat lengths overrepresented in individuals born preterm compared to those born at term (Table 5). *AR* is a nuclear receptor for the androgens testosterone or dihydrosterone. It functions as a transcription factor regulating expression of its target genes involved in the development of several tissue, in maintenance of male sexual characteristics, and in female reproductive physiology [40]. *AR* also has non-genomic actions through interactions with signal transduction proteins in the cytoplasm [55,56]. The *AR* CAG repeat encodes a polyglutamine tract present in the transactivation domain of *AR* [46]. This domain is involved in ligand-induced transcriptional activity. Longer polyglutamine chains have been shown to hinder the transactivation activity of *AR* *in vitro*, while short polyglutamine chains lead to enhanced *AR* activation resulting in hyperandrogenism [46,57–59]. Thus, it is evident that the CAG repeat length polymorphism has biologically significant consequences for *AR* function.

*AR* has important roles during pregnancy and may be involved in some pregnancy-related conditions. *AR* is widely expressed in reproductive tissues of both males and females [60]. In developing first-trimester human embryos, expression of *AR* has been



**Table 5.** AR CAG repeat distributions in individuals born spontaneously preterm and in those born at term.

	<i>n</i>	Mean ± SD <sup>a</sup>	Mann-Whitney <i>U</i> test, <i>p</i>	Frequency of repeats in quintiles <sup>b</sup>	$\chi^2$ test for quintiles, <i>p</i>	OR (95% CI) <sup>c</sup>
<b>Oulu:</b>						
SPTB	252	22.59±3.11	0.0006	0.109/0.282/0.260/0.208/0.141	0.009	2.45 (1.46–4.13)
Term	189	21.87±2.64		0.201/0.267/0.275/0.151/0.106		
<b>Helsinki:</b>						
SPTB	111	22.45±3.35	0.67	0.122/0.338/0.216/0.180/0.144	0.047	2.25 (1.14–4.43)
Term	197	22.09±2.54		0.145/0.305/0.269/0.206/0.076		
<b>Combined Oulu-Helsinki:</b>						
SPTB	363	22.55±3.18	0.0027	0.113/0.299/0.247/0.200/0.142	0.0006	2.39 (1.58–3.60)
Term	386	21.98±2.59		0.172/0.286/0.272/0.179/0.091		

<sup>a</sup>Mean ± standard deviation of CAG<sub>n</sub> repeat length.

<sup>b</sup>1<sup>st</sup> quintile 10–19 repeats, 2<sup>nd</sup> quintile 20–21 repeats, 3<sup>rd</sup> quintile 22–23 repeats, 4<sup>th</sup> quintile 24–25 repeats, 5<sup>th</sup> quintile 26–42 repeats; frequencies shown separately for male and female individuals in Table S2.

<sup>c</sup>ORs for the highest quintile relative to the lowest quintile.

doi:10.1371/journal.pone.0051378.t005

detected in several tissues, including several extragenital sites, such as thymus, bronchial epithelium of the lung, spinal cord, and cardiac valves [61]. AR levels in the fetal lung change significantly with gestational age during mid- and late gestation [62]. Longer AR CAG repeats (>19) have been shown to be overrepresented in women with recurrent spontaneous abortions [48], and increased AR expression has been detected in placentas of preeclamptic women [63], suggesting that AR plays a role in these pregnancy-related complications. Furthermore, successful induction of labor by local prostaglandin treatment leads to decreased AR and progesterone receptor levels in human uterine cervix [64]. This suggests that besides the decrease in progesterone activity, androgen withdrawal may be involved in the onset of parturition. However, complete androgen insensitivity syndrome of the fetus does not seem to lead to preterm birth [65]. Further studies on the roles of AR exon-1 CAG repeat and the genes regulated by AR

could lead to further understanding of the function of fetal AR in regulation of the onset of labor.

In our previous linkage study performed in the same families as the present analysis, we found a significant linkage signal within the gene encoding *IGF1R* and identified a 6-SNP haplotype within this gene as a fetal susceptibility factor for SPTB [27]. Interestingly, recent studies have identified *IGF1R* as a downstream target for AR [66,67]. Furthermore, androgens have been shown to be involved in activation of *IGF1R* through non-genomic pathways involving AR [67–69]. In addition, hypermethylation of the AR promoter may be involved in the progression of prostate cancer and downregulate expression of *IGF1R* [67]. Because these two genes were identified as fetal susceptibility factors for SPTB in our previous and current studies, it is tempting to postulate that interactions of these genes may affect the onset of spontaneous preterm labor. However, further studies are needed to test this hypothesis.

In the present study, we did not find significant linkage in the mothers, i.e. when the phenotype was giving preterm deliveries. This may be due to the small sample numbers in this analysis ( $n_{affected} = 21$ ) leading to low power to detect linkage. Previous genetic modeling studies have estimated that both maternal and fetal genetic effects contribute to the timing of parturition [15–17]. This is consistent with our finding of a fetal gene associated with SPTB. However, several large population-based studies suggest a significant contribution of maternal genes to preterm delivery with little effect of fetal genes [8,12,13]. In a Dutch twin study, there was no detectable evidence for paternal heritability of parturition timing [70]. These studies suggest that paternal genes acting in the fetus have little effect on the risk of preterm birth compared to maternal genes acting either in the pregnant mother or in the fetus. It remains to be investigated whether both maternal and paternal alleles of the fetus could contribute to the SPTB phenotype.

The strengths of our study include 1) the use of the family-based linkage analysis for localization of the associating genomic region, 2) the use of the northern Finnish population, which is known to be genetically relatively homogeneous [54], 3) further analysis in the genetically more diverse southern Finnish population, 4) analysis of both the fetal and maternal contributions, and 5) stringent selection criteria for the SPTB phenotype with deliveries involving major risk

**Table 6.** FBAT analysis of the AR CAG<sub>n</sub> polymorphism under the dominant model in mother-father-preterm offspring trios of the initial SPTB case population.

CAG <sub>n</sub> allele (number of repeats) <sup>a</sup>	Number of families	Z <sup>b</sup>	<i>p</i>
18	31	−0.180	0.857
19	20	−2.236	0.025
20	42	1.234	0.217
21	56	−1.336	0.181
22	42	1.234	0.217
23	41	−0.469	0.639
24	28	1.890	0.059
25	46	0.000	1.000
26	25	−0.200	0.841
27	11	−0.905	0.366

<sup>a</sup>Analysis was not performed for alleles with less than 10 transmissions.

<sup>b</sup>Positive and negative values of the FBAT test statistic Z are indicative of over- and undertransmission, respectively.

doi:10.1371/journal.pone.0051378.t006

factors for preterm birth excluded. In addition, our initial control population consisted of mothers with exclusively at least three term deliveries, and their offspring. These individuals therefore represented the so-called “super controls” for the SPTB phenotype. To avoid false positives in the case-control study, the significances were considered on the stringent Bonferroni-corrected level.

The limitations of our study include the relatively small number of families in linkage analysis and the relatively small sample sizes of both the initial case-control study population and particularly that of the replication study population. Although association between the fetal *AR* CAG repeat and SPTB was not directly replicated in the southern Finnish population, the quintile comparison suggested that, similarly to the northern Finnish population, the longest alleles may be overrepresented in individuals born preterm (Table 5). This analysis was apparently limited by the small sample sizes in each of the quintile subgroups in the southern Finnish population. However, because the differences in the *AR* CAG distribution were significant when the northern and southern Finnish populations were combined, we feel that results of the southern Finnish population strengthened the evidence for the role of fetal *AR* CAG repeat in genetic predisposition to SPTB. Our study was also limited in that we only analyzed two genes located near the linkage signal. Thus, we cannot exclude the possibility that other genes located in the linked region could play a role in SPTB. We also cannot exclude a possible role for genes located near the suggestive linkage signals on chromosome locus Xq21.1 (Figure 2) in SPTB. In the current setting, we cannot evaluate whether the linkage signal actually was due to linkage to *AR* CAG<sub>n</sub> polymorphism. Short tandem repeat polymorphisms (STRPs), such as *AR* CAG<sub>n</sub>, are known to display significant LD with SNPs even on long distances (up to 2 cM), and STRP-SNP LD is greater on the X chromosome compared to autosomes [44]. Therefore, it is possible that *AR* CAG<sub>n</sub> polymorphism indeed is linked to SPTB. This conclusion is further supported by the significant result of the family-based linkage given LD test performed in the northern Finnish case-control families in the present study. Further studies are needed to evaluate the association between the fetal *AR* CAG repeat polymorphism and SPTB; these should include replication in other, more outbred populations and larger sample sizes.

In conclusion, we found the highest linkage signal (HLOD = 3.72) on chromosome Xq13.1 when the phenotype was being

born preterm. Two genes located near this locus, encoding the androgen receptor (*AR*) and the interleukin-2 receptor gamma subunit (*IL2RG*), were considered functional candidates for SPTB and were thus analyzed in a subsequent case-control association analysis of mothers and their offspring originating from northern Finland, and in an independent case-control population of offspring from southern Finland. Finally, we demonstrated that a CAG repeat polymorphism present in exon 1 of the *AR* gene associates with SPTB in the offspring. Therefore, our results implicate *AR* as a potential fetal susceptibility gene for SPTB. The role of *AR* in parturition timing remains to be investigated in subsequent studies.

## Supporting Information

**Figure S1 *AR* CAG repeat distributions in the initial case-control population.**  
(PDF)

**Table S1 Gestational ages and years of birth for the preterm-born members of the seven families with recurrent SPTB analyzed in the linkage analysis of X chromosomal markers.**  
(PDF)

**Table S2 *AR* CAG repeat distributions in male and female SPTB and term offspring.**  
(PDF)

## Acknowledgments

The authors thank Maarit Haarala, Outi Kajula, Mirka Ovaska, Hilka Puttonen, and Riitta Vikeväinen for sample collection and technical assistance; Tero Hiekkalinna and Joseph Terwilliger for advice in linkage analysis; and Teppo Varilo for performing the genealogical analysis.

## Author Contributions

Conceived and designed the experiments: MKK AL JP AP RH LJM MH. Performed the experiments: MKK JMH JU JS RH. Analyzed the data: MKK JMH RH. Contributed reagents/materials/analysis tools: KT VA LJM MH. Wrote the paper: MKK. Contributed in obtaining patient data and DNA: AL KT. Served as a liaison between the investigators: MH.

## References

1. Simmons LE, Rubens CE, Darmstadt GL, Gravett MG (2010) Preventing preterm birth and neonatal mortality: Exploring the epidemiology, causes, and interventions. *Semin Perinatol* 34(6): 408–415.
2. Beck S, Wojdyla D, Say L, Betran AP, Merialdi M, et al. (2010) The worldwide incidence of preterm birth: A systematic review of maternal mortality and morbidity. *Bull World Health Organ* 88(1): 31–38.
3. Iacovidou N, Varsami M, Syggellou A (2010) Neonatal outcome of preterm delivery. *Ann N Y Acad Sci* 1205: 130–134.
4. Goldenberg RL, Culhane JF, Iams JD, Romero R (2008) Epidemiology and causes of preterm birth. *Lancet* 371(9606): 75–84.
5. Adams MM, Elam-Evans LD, Wilson HG, Gilbert DA (2000) Rates of and factors associated with recurrence of preterm delivery. *JAMA* 283(12): 1591–1596.
6. Basso O, Olsen J, Christensen K (1999) Study of environmental, social, and paternal factors in preterm delivery using sibs and half sibs. A population-based study in denmark. *J Epidemiol Community Health* 53(1): 20–23.
7. Bhattacharya S, Raja EA, Mirazo ER, Campbell DM, Lee AJ, et al. (2010) Inherited predisposition to spontaneous preterm delivery. *Obstet Gynecol* 115(6): 1125–1133.
8. Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B, et al. (2009) Maternal contributions to preterm delivery. *Am J Epidemiol* 170(11): 1358–1364.
9. Winkvist A, Mogren I, Hogberg U (1998) Familial patterns in birth characteristics: Impact on individual and population risks. *Int J Epidemiol* 27(2): 248–254.
10. Porter TF, Fraser AM, Hunter CY, Ward RH, Varner MW (1997) The risk of preterm birth across generations. *Obstet Gynecol* 90(1): 63–67.
11. Di Renzo GC, Giardina I, Rosati A, Clerici G, Torricelli M, et al. (2011) Maternal risk factors for preterm birth: A country-based population analysis. *Eur J Obstet Gynecol Reprod Biol* 159(2): 342–346.
12. Svensson AC, Sandin S, Cnattingius S, Reilly M, Pawitan Y, et al. (2009) Maternal effects for preterm birth: A genetic epidemiologic study of 630,000 families. *Am J Epidemiol* 170(11): 1365–1372.
13. Wilcox AJ, Skaerven R, Lie RT (2008) Familial patterns of preterm delivery: Maternal and fetal contributions. *Am J Epidemiol* 167(4): 474–479.
14. Chaudhari BP, Plunkett J, Ratajczak CK, Shen TT, DeFranco EA, et al. (2008) The genetics of birth timing: Insights into a fundamental component of human development. *Clin Genet* 74(6): 493–501.
15. Lunde A, Melve KK, Gjessing HK, Skjaerven R, Irgens LM (2007) Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *Am J Epidemiol* 165(7): 734–741.
16. Plunkett J, Feitosa MF, Trusgnich M, Wangler MF, Palomar L, et al. (2009) Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. *Hum Hered* 68(3): 209–219.
17. York TP, Strauss JF 3rd, Neale MC, Eaves LJ (2010) Racial differences in genetic and environmental risk to preterm birth. *PLoS One* 5(8): e12391.
18. Plunkett J, Muglia LJ (2008) Genetic contributions to preterm birth: Implications from epidemiological and genetic association studies. *Ann Med* 40(3): 167–195.

19. Annells MF, Hart PH, Mullighan CG, Heatley SL, Robinson JS, et al. (2004) Interleukins-1, -4, -6, -10, tumor necrosis factor, transforming growth factor-beta, FAS, and mannose-binding protein C gene polymorphisms in Australian women: Risk of preterm birth. *Am J Obstet Gynecol* 191(6): 2056–2067.
20. Hartel C, Finas D, Ahrens P, Kattner E, Schaible T, et al. (2004) Polymorphisms of genes involved in innate immunity: Association with preterm delivery. *Mol Hum Reprod* 10(12): 911–915.
21. Aidoo M, McElroy PD, Kolczak MS, Terlouw DJ, ter Kuile FO, et al. (2001) Tumor necrosis factor-alpha promoter variant 2 (TNF2) is associated with preterm delivery, infant mortality, and malaria morbidity in western Kenya: Asemo bay cohort project IX. *Genet Epidemiol* 21(3): 201–211.
22. Chen D, Hu Y, Wu B, Chen L, Fang Z, et al. (2003) Tumor necrosis factor-alpha gene G308A polymorphism is associated with the risk of preterm delivery. *Beijing Da Xue Xue Bao* 35(4): 377–381.
23. Engel SA, Erichsen HC, Savitz DA, Thorp J, Chanock SJ, et al. (2005) Risk of spontaneous preterm birth is associated with common proinflammatory cytokine polymorphisms. *Epidemiology* 16(4): 469–477.
24. Plunkett J, Doniger S, Orabona G, Morgan T, Haataja R, et al. (2011) An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet* 7(4): e1001365.
25. Romero R, Velez Edwards DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, et al. (2010) 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(5): 431.e1–431.34.
26. Myking S, Myhre R, Gjessing HK, Morken NH, Sengpiel V, et al. (2011) Candidate gene analysis of spontaneous preterm delivery: New insights from re-analysis of a case-control study using case-parent triads and control-mother dyads. *BMC Med Genet* 12: 174.
27. Haataja R, Karjalainen MK, Luukkonen A, Teramo K, Puttonen H, et al. (2011) Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis. *PLoS Genet* 7(2): e1001293.
28. THL (The Finnish National Institute for Health and Welfare) (2010) Synnytykset ja vastasyntyneet 2009. Tilastoraportti 26/2010.
29. Varilo T (1999) The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. Ph.D. thesis, University of Helsinki, Department of Medical Genetics, Faculty of Medicine and Department of Human Molecular Genetics, National Public Health Institute, Helsinki. 98 p.
30. Di Renzo GC, Rosati A, Sarti RD, Cruciani L, Cutuli AM (2007) Does fetal sex affect pregnancy outcome? *Genet Med* 4(1): 19–30.
31. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81(3): 559–575.
32. Ulgen A, Li W (2005) Comparing single-nucleotide polymorphism marker-based and microsatellite marker-based linkage analyses. *BMC Genet* 6 Suppl 1: S13.
33. O'Connell JR, Weeks DE (1998) PedCheck: A program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63(1): 259–266.
34. Goode EL, Jarvik GP (2005) Assessment and implications of linkage disequilibrium in genome-wide single-nucleotide polymorphism and microsatellite panels. *Genet Epidemiol* 29 Suppl 1: S72–76.
35. Goring HH, Terwilliger JD (2000) Linkage analysis in the presence of errors IV: Joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am J Hum Genet* 66(4): 1310–1327.
36. Strauch K, Fimmers R, Baur MP, Wienker TF (2003) How to model a complex trait. I. general considerations and suggestions. *Hum Hered* 55(4): 202–210.
37. Hiekkalinna T, Terwilliger JD, Sammalisto S, Peltonen L, Perola M (2005) AUTOGSCAN: Powerful tools for automated genome-wide linkage and linkage disequilibrium analysis. *Twin Res Hum Genet* 8(1): 16–21.
38. Meazza R, Azzarone B, Orengo AM, Ferrini S (2011) Role of common-gamma chain cytokines in NK cell development and function: Perspectives for immunotherapy. *J Biomed Biotechnol* 2011: 861920.
39. Haram K, Mortensen JH, Wollen AL (2003) Preterm delivery: An overview. *Acta Obstet Gynecol Scand* 82(8): 687–704.
40. Kumar R, Atamna H, Zakharov MN, Bhasin S, Khan SH, et al. (2011) Role of the androgen receptor CAG repeat polymorphism in prostate cancer, and spinal and bulbar muscular atrophy. *Life Sci* 88(13–14): 565–571.
41. Nelson LR, Bulun SE (2001) Estrogen production and action. *J Am Acad Dermatol* 45(3 Suppl): S116–24.
42. Eberle MA, Rieder MJ, Kruglyak L, Nickerson DA (2006) Allele frequency matching between SNPs reveals an excess of linkage disequilibrium in genic regions of the human genome. *PLoS Genet* 2(9): e142.
43. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, et al. (2005) Whole-genome patterns of common DNA variation in three human populations. *Science* 307(5712): 1072–1079.
44. Payseur BA, Place M, Weber JL (2008) Linkage disequilibrium between STRPs and SNPs across the human genome. *Am J Hum Genet* 82(5): 1039–1050.
45. Karvela M, Stefanakis N, Papadopoulou S, Tsitilou SG, Tsilivakos V, et al. (2008) Evidence for association of the G1733A polymorphism of the androgen receptor gene with recurrent spontaneous abortions. *Fertil Steril* 90(5): 2010.e9–2010.12.
46. Palazzolo I, Gliozzi A, Rusmini P, Sau D, Crippa V, et al. (2008) The role of the polyglutamine tract in androgen receptor. *J Steroid Biochem Mol Biol* 108(3–5): 245–253.
47. Van Nieuwerburgh F, Stoop D, Cabri P, Dhont M, Deforce D, et al. (2008) Shorter CAG repeats in the androgen receptor gene may enhance hyperandrogenicity in polycystic ovary syndrome. *Gynecol Endocrinol* 24(12): 669–673.
48. Aruna M, Dasgupta S, Sirisha PV, Andal Bhaskar S, Tarakeswari S, et al. (2011) Role of androgen receptor CAG repeat polymorphism and X-inactivation in the manifestation of recurrent spontaneous abortions in Indian women. *PLoS One* 6(3): e17718.
49. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. *Bioinformatics* 21(2): 263–265.
50. Hiekkalinna T, Schaffer AA, Lambert B, Norrgrann P, Goring HH, et al. (2011) PSEUDOMARKER: A powerful program for joint linkage and/or linkage disequilibrium analysis on mixtures of singletons and related individuals. *Hum Hered* 71(4): 256–266.
51. Horvath S, Xu X, Laird NM (2001) The family based association test method: Strategies for studying general genotype-phenotype associations. *Eur J Hum Genet* 9(4): 301–306.
52. Karjalainen MK (2011) Genetic predisposition to spontaneous preterm birth: approaches to identify susceptibility genes. Ph.D. thesis, University of Oulu, Faculty of Medicine, Institute of Clinical Medicine, Department of Paediatrics, Oulu. Acta Universitatis Ouluensis Medica D1143. 140 p.
53. Ackerman CM, Lowe LP, Lee H, Hayes MG, Dyer AR, et al. (2011) Ethnic variation in allele distribution of the androgen receptor (AR) (CAG)<sub>n</sub> repeat. *J Androl* 33(2): 210–215.
54. Jakkula E, Rehnstrom K, Varilo T, Pietilainen OP, Paunio T, et al. (2008) The genome-wide patterns of variation expose significant substructure in a founder population. *Am J Hum Genet* 83(6): 787–794.
55. Heinelein CA, Chang C (2002) The roles of androgen receptors and androgen-binding proteins in nongenomic androgen actions. *Mol Endocrinol* 16(10): 2181–2187.
56. Michels G, Hoppe UC (2008) Rapid actions of androgens. *Front Neuroendocrinol* 29(2): 182–198.
57. Chamberlain NL, Driver ED, Miesfeld RL (1994) The length and location of CAG trinucleotide repeats in the androgen receptor N-terminal domain affect transactivation function. *Nucleic Acids Res* 22(15): 3181–3186.
58. Kazemi-Esfarjani P, Trifiro MA, Pinsky L (1995) Evidence for a repressive function of the long polyglutamine tract in the human androgen receptor: Possible pathogenetic relevance for the (CAG)<sub>n</sub>-expanded neuropathies. *Hum Mol Genet* 4(4): 523–527.
59. Ibanez L, Ong KK, Mongan N, Jaaskelainen J, Marcos MV, et al. (2003) Androgen receptor gene CAG repeat polymorphism in the development of ovarian hyperandrogenism. *J Clin Endocrinol Metab* 88(7): 3333–3338.
60. Wilton CM, McPhaul MJ (1996) A and B forms of the androgen receptor are expressed in a variety of human tissues. *Mol Cell Endocrinol* 120(1): 51–57.
61. Sajjad Y, Quenby S, Nickson P, Lewis-Jones DI, Vince G (2007) Androgen receptors are expressed in a variety of human fetal extragenital tissues: An immunohistochemical study. *Asian J Androl* 9(6): 751–759.
62. Simard M, Plante J, Boucher M, Provost PR, Tremblay Y (2010) Type 2 and 5 17beta-hydroxysteroid dehydrogenases and androgen receptor in human fetal lungs. *Mol Cell Endocrinol* 319(1–2): 79–87.
63. Hsu TY, Lan KC, Tsai CC, Ou CY, Cheng BH, et al. (2009) Expression of androgen receptor in human placentas from normal and preeclamptic pregnancies. *Taiwan J Obstet Gynecol* 48(3): 262–267.
64. Vladic-Stjernholm Y, Vladic T, Blesson CS, Ekman-Ordeberg G, Sahlin L (2009) Prostaglandin treatment is associated with a withdrawal of progesterone and androgen at the receptor level in the uterine cervix. *Reprod Biol Endocrinol* 7: 116.
65. Miles HL, Gidlof S, Nordenstrom A, Ong KK, Hughes IA (2010) The role of androgens in fetal growth: Observational study in two genetic models of disordered androgen signalling. *Arch Dis Child Fetal Neonatal Ed* 95(6): F435–8.
66. Schayek H, Seti H, Greenberg NM, Sun S, Werner H, et al. (2010) Differential regulation of insulin-like growth factor-I receptor gene expression by wild type and mutant androgen receptor in prostate cancer cells. *Mol Cell Endocrinol* 323(2): 239–245.
67. Schayek H, Bentov I, Sun S, Plymate SR, Werner H (2010) Progression to metastatic stage in a cellular model of prostate cancer is associated with methylation of the androgen receptor gene and transcriptional suppression of the insulin-like growth factor-I receptor gene. *Exp Cell Res* 316(9): 1479–1488.
68. Pandini G, Mineo R, Frasca F, Roberts CT Jr, Marcelli M, et al. (2005) Androgens up-regulate the insulin-like growth factor-I receptor in prostate cancer cells. *Cancer Res* 65(5): 1849–1857.
69. Pandini G, Genua M, Frasca F, Vigneri R, Belfiore A (2009) Sex steroids upregulate the IGF-1R in prostate cancer cells through a nongenotropic pathway. *Ann N Y Acad Sci* 1155: 263–267.
70. Kistka ZA, DeFranco EA, Lighthart L, Willemsen G, Plunkett J, et al. (2008) Heritability of parturition timing: An extended twin design analysis. *Am J Obstet Gynecol* 199(1): 43.e1–43.e5.