Mapping a New Spontaneous Preterm Birth Susceptibility Gene, *IGF1R*, Using Linkage, Haplotype Sharing, and Association Analysis

Ritva Haataja[1⚬], Minna K. Karjalainen[1⚬]*, Aino Luukkonen[1], Kari Teramo[2], Hilkka Puttonen[2], Marja Ojaniemi[1], Teppo Varilo[3,4,5], Bimal P. Chaudhari[6], Jevon Plunkett[7,8], Jeffrey C. Murray[9], Steven A. McCarroll[10], Leena Peltonen[3,4,5,11,12]†, Louis J. Muglia[7], Aarno Palotie[3,5,11,12], Mikko Hallman[1]

1 Department of Pediatrics, Institute of Clinical Medicine, University of Oulu, Oulu, Finland, 2 Department of Obstetrics and Gynecology, University Central Hospital, Helsinki, Finland, 3 Department of Medical Genetics, Haartman Institute, University of Helsinki and Helsinki University Central Hospital, Helsinki, Finland, 4 National Institute for Health and Welfare (THL), Helsinki, Finland, 5 Institute for Molecular Medicine Finland (FIMM), University of Helsinki, Helsinki, Finland, 6 Department of Pediatrics, Washington University School of Medicine, St. Louis, Missouri, United States of America, 7 Department of Pediatrics, Vanderbilt University School of Medicine, Nashville, Tennessee, United States of America, 8 Human and Statistics Genetics Program, Washington University School of Medicine, St. Louis, Missouri, United States of America, 9 Department of Pediatrics, University of Iowa, Iowa City, Iowa, United States of America, 10 Department of Genetics, Harvard Medical School, Boston, Massachusetts, United States of America, 11 The Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, Massachusetts, United States of America, 12 Wellcome Trust Sanger Institute, Cambridge, United Kingdom

Abstract

Preterm birth is the major cause of neonatal death and serious morbidity. Most preterm births are due to spontaneous onset of labor without a known cause or effective prevention. Both maternal and fetal genomes influence the predisposition to spontaneous preterm birth (SPTB), but the susceptibility loci remain to be defined. We utilized a combination of unique population structures, family-based linkage analysis, and subsequent case-control association to identify a susceptibility haplotype for SPTB. Clinically well-characterized SPTB families from northern Finland, a subisolate founded by a relatively small founder population that has subsequently experienced a number of bottlenecks, were selected for the initial discovery sample. Genome-wide linkage analysis using a high-density single-nucleotide polymorphism (SNP) array in seven large northern Finnish non-consanguinous families identified a locus on 15q26.3 (HLOD 4.68). This region contains the *IGF1R* gene, which encodes the type 1 insulin-like growth factor receptor IGF-1R. Haplotype segregation analysis revealed that a 55 kb 12-SNP core segment within the *IGF1R* gene was shared identical-by-state (IBS) in five families. A follow-up case-control study in an independent sample representing the more general Finnish population showed an association of a 6-SNP *IGF1R* haplotype with SPTB in the fetuses, providing further evidence for *IGF1R* as a SPTB predisposition gene (frequency in cases versus controls 0.11 versus 0.05, $P = 0.001$, odds ratio 2.3). This study demonstrates the identification of a predisposing, low-frequency haplotype in a multifactorial trait using a well-characterized population and a combination of family and case-control designs. Our findings support the identification of the novel susceptibility gene *IGF1R* for predisposition by the fetal genome to being born preterm.

Introduction

Preterm birth, defined as birth before 37 wk of gestation, accounts for an estimated 2 million annual deaths worldwide and is the major cause of serious morbidity in infants born preterm. Currently, approximately 12% of all births in the United States are premature. The serious acute diseases of prematurely born infants are principally caused by functional immaturity. Common life-long diseases that result in deteriorating quality of life among individuals born preterm include a chronic respiratory disease called bronchopulmonary dysplasia; retinopathy of prematurity, which is the most common cause of blindness in infants; cerebral palsy; and cognitive disorders [1]. The majority of preterm births (approximately 70%) occur after spontaneous onset of labor;

Author Summary

Preterm birth is the major cause of infant deaths and life-long neurologic and cardiopulmonary morbidity. More than 10% of births in the United States occur prematurely, and the rate is increasing without known effective prevention. Previous premature birth increases the risk 3-fold in subsequent pregnancies. We report here, for the first time to our knowledge, a genome-wide study on susceptibility to spontaneous preterm birth in singleton pregnancies. To detect novel regions of the genome associated with preterm birth, we performed linkage analysis on seven carefully selected large families with recurrent spontaneous premature births. When we studied the fetuses, evidence was found for linkage of a region on chromosome 15 with spontaneous preterm birth, with the highest linkage signals occurring within a single gene, *IGF1R*. Evidence of the involvement of this gene in the etiology of preterm birth was further strengthened by subsequent haplotype segregation analysis and case-control analysis of an independent patient population. The *IGF1R* gene encodes insulin-like growth factor receptor 1 (IGF-1R), an important protein that potentially regulates signaling cascades involved in the onset of labor. Our analyses are unique in providing evidence that fetal *IGF1R* influences the risk of spontaneous preterm labor, leading to preterm birth.

nearly 50% of these cases are preceded by rupture of fetal membranes. Apart from excessive uterine distension in multiple pregnancies or certain fetal malformations, and severe maternal diseases such as sepsis and abdominal trauma, no obvious environmental risk factors can be identified in most preterm births. Activation of spontaneous preterm labor and preterm birth is thought to result from the action of multiple pathways and mechanisms, including endocrine dysfunction or ascending intrauterine infection and inflammation that can lead to the induction of labor-producing mediators [2]. Despite ongoing research efforts, there is no effective medication for the prevention of spontaneous preterm birth (SPTB).

A history of SPTB of a single fetus is a strong predictor of its recurrence in families [3]. Approximately 20% of mothers with a preterm delivery have another baby born preterm [4], suggesting that factors that are stable over time, such as genetics, affect birth timing [5]. Mothers and daughters [6] and sisters [7] share the risk of delivering preterm. Twin studies suggest a heritability estimate of about 30% [8–10]. Both the fetal and maternal genome, as well as gene–gene and gene–environment interactions are likely to influence predisposition to SPTB. Several studies using fetal or maternal DNA have reported associations of individual gene polymorphisms [11]. These studies have focused on genes involved in infection, inflammation, and innate immunity; e.g. those encoding the cytokines tumor necrosis factor alpha and interleukins 4, 6, and 10, and mannose-binding lectin [12–21]. However, most of these associations were not replicated in subsequent studies and across populations. So far, only case-control candidate gene studies have been conducted for SPTB. Genome-wide methods of identifying genes *a priori* may reveal genes not considered to be obvious candidates, which are potentially important unexplored sources of variability in preterm birth. These studies may contribute to defining the risk of SPTB and developing potential preventive interventions.

The overall aim of the present approach was to define major genes that influence the susceptibility to SPTB. In this first report, we describe a SNP-based genome-wide linkage and haplotype

segregation analysis of recurrent familial SPTB using a strictly defined phenotype and carefully selected families, followed by case-control association analysis of a study population independent of the subjects used for the linkage scan. The linkage scan was performed with seven large families originating from northern Finland, where the population is characterized by genetic homogeneity, making it advantageous for gene-mapping studies [22]. With the phenotype defined as being born preterm, significant linkage signals ($HLOD_{max} = 4.68$) were obtained for chromosome locus 15q26.3 in a region harboring *IGF1R* (MIM *147370), the gene encoding insulin-like growth factor receptor 1 (IGF-1R). Haplotype segregation analysis performed for markers encompassing the *IGF1R* gene revealed prominent identical-by-descent (IBD) within-family and identical-by-state (IBS) between-family haplotype sharing among affected relatives. Evidence of the involvement of *IGF1R* in the etiology of SPTB was further strengthened by case-control analysis of an independent cohort located in northern and southern Finland, with a 6-SNP *IGF1R* haplotype overrepresented in SPTB infants. In summary, evidence from our linkage, haplotype sharing, and association analyses implicated *IGF1R* as a candidate gene for susceptibility to SPTB.

Results

The overall study design is illustrated in Figure 1.

Linkage Analysis

We selected the families for the linkage analysis from a total 120,000 births that took place in a single regional hospital in northern Finland; this region is characterized by a homogeneous and stable population with a low prevalence of prematurity (approximately 5.5–6.5% of all births). We chose mothers with recurrent SPTB using very stringent criteria, with known risk
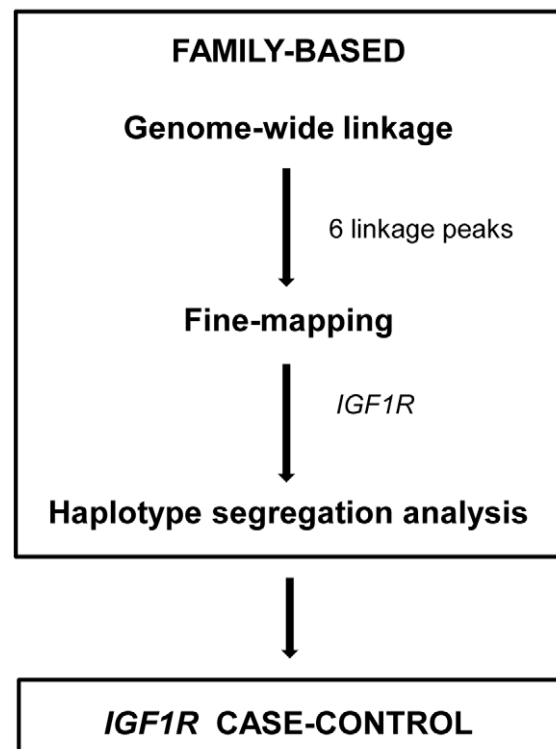

Figure 1. Study design.
doi:10.1371/journal.pgen.1001293.g001

factors for SPTB and elective preterm births without labor among the exclusion criteria. We identified 120 mothers with at least two spontaneous singleton preterm deliveries. Family interviews revealed 20 large families with multiple relatives affected by SPTB. According to a genealogical study, these families were non-consanguinous. Finally, families with apparent maternal inheritance of SPTB were chosen for the analysis.

We conducted parametric linkage analysis using seven large northern Finnish families with recurrent SPTB. Because genetic factors acting either on the fetus or the mother may influence SPTB, we used two settings in this study: affected fetus or infant phenotype (being spontaneously born preterm as the phenotype, $n = 41$) and affected mother phenotype (giving spontaneous preterm birth as the phenotype, $n = 21$). The pedigrees of the families are shown in Figure S1.

Before the analysis, markers were linkage-disequilibrium (LD) pruned to exclude high-LD SNPs, leaving 6377 markers with an average distance of 0.43 Mb between consecutive markers. We considered a heterogeneity logarithm of odds (HLOD) score of >2 as an initial signal of linkage. When we studied the affected fetus phenotype, analysis of the pruned marker set revealed HLOD scores of >2 on six autosomes, as depicted in Figure 2. The maximum HLOD score ($HLOD_{max}$) of 2.59 was detected for SNP rs2715416 on chromosome locus 15q26.3 ($\theta = 0.04$, $\alpha = 1.00$). Figure 2 also shows the linkage signals for mother-based analysis, with an $HLOD_{max}$ of 1.53 at rs11167102 on chromosome locus 8q24.3 ($\theta = 0.00$, $\alpha = 1.00$).

For the affected fetus phenotype, we performed fine mapping using the unpruned marker set on the regions flanking (approximately 5 Mb) the six initial linkage signals. We considered an HLOD of >3 as a further sign of linkage. Table 1 reveals that the three markers with an HLOD of >3 were all on the same chromosome locus 15q26.3 and shows the HLOD scores in the fine-scale analysis for each of the seven families with recurrent SPTB. The $HLOD_{max}$ of 4.68 was obtained at rs2684811. Because one preterm infant not fulfilling the stringent criteria of SPTB (preterm infant from a twin pregnancy) was included among the premature births, we repeated the linkage analyses while excluding this infant from the affected individuals, with little effect on the results. All three markers with the highest HLOD scores are located within a single gene *IGF1R*. Interestingly, the markers that yielded the second- and third-highest HLOD scores in the mother-based analysis (SNPs rs329292 and rs11247268 with HLOD scores of 1.51 and 1.48 respectively; Figure 2) were located on the same chromosomal region (15q26.3) as the marker with $HLOD_{max}$ in the infants, with a distance of approximately 2.2 Mb between the markers. Because of the colocalized linkage signals in both the fetal- and mother-based analysis in the region including *IGF1R*, we chose to explore this gene in greater detail.

IGF1R Haplotype Segregation and *Post Hoc* Linkage Analysis

We performed haplotype segregation analysis using the SNPs flanking the highest linkage peak within *IGF1R* in the six linked families 24, 70, 126, 150, 185 and 253. Family 210 did not undergo haplotype segregation analysis due to an absence of linkage in this family (Table 1). The analysis was performed for 30 SNPs covering a 330 kb region encompassing the entire *IGF1R* gene. We aimed to resolve whether one or more distinguishable haplotype segments inferred from these SNPs cosegregated with SPTB. While not suitable as such for statistical evaluation, segregation analysis is a useful tool for an empirical hypothesis-generating approach [23]. The model used for the segregation analysis was initially designed to be best-fit to the mode of

inheritance (MOI) used in the linkage analysis (dominant MOI, allowing for the presence of healthy carriers as transmitters of a disease-cosegregating haplotype). However, haplotype-sharing analysis offers the advantage over parametric linkage analysis of dissecting the genetic data regardless of true MOI.

IGF1R haplotypes for family 70 with the highest linkage signal are represented in Figure 3, and Figure S2 shows these haplotypes for all of the linked families, including family 70. In all six families, and considering the haplotypes comprising 30 SNPs spanning the entire *IGF1R* gene, we observed prominent IBD within-family haplotype sharing among the affected relatives (Figure 3 and Figure S2). A single disease-cosegregating haplotype was shared IBD within family in families 70, 126, 253 and 185. In the remaining two families, 24 and 150, an IBD-shared haplotype was identified in a subset of members of each family, and a second haplotype derived from another carrier was identified in a different subset of members. On the whole, the segregation analysis supported the assumed model very precisely, because it predicted the carriership of a disease-cosegregating haplotype in 34 out of all 38 affected individuals (absent from individuals 70-2, 150-15, 150-19 and 150-24), whereas only six nonaffected family members were deemed to be carriers of a disease-cosegregating haplotype (present in individuals 24-1, 24-5, 24-11, 24-9, 150-8, and 150-26) (Figure 3 and Figure S2). The recombinations that we observed, particularly that in the unaffected male 253-5, also fit the segregation model well. Three of the families (126, 253, and 185) showed a complete haplotype-disease cosegregation (Figure S2). However, unexpected male carriers were identified in four families (spouses of females 24-2 and 70-1, and males 253-1 and 150-3). Families 24 and 150 were consistent with a mixed maternal-paternal bilineal transmission pattern, and family 70 exhibited nearly complete penetrance with mixed unilineal maternal–paternal transmission. The segregation patterns were consistent with complete penetrance in families 126, 253, and 185; with 100% maternal transmission in families 126 and 185; and unexpected 100% paternal transmission in family 253.

An attempt to identify a similar IBD- or IBS-shared chromosomal segment among families led us to discover two core haplotypes with overlapping locations (illustrated in Figure S2). A 55-kb 12-SNP haplotype, CAGACGATACTC (core I, comprising the interval between SNPs rs1879612–rs2715416), was shared IBS among five families: all of families 24, 70, 126 and 253 and part of 150. Another 79-kb 11-SNP core haplotype, ATGTGTAATGT (core II, SNPs rs2684761–rs3743259), was shared IBS between part of family 150 and the whole of family 185. The disease-segregating chromosomes with the core II haplotype were maternal in 100% of preterm-born individuals carrying these chromosomes, while haplotypes with core I were maternal in only about half (47%) of the cases.

All of the linked families shared one of the core haplotype segments, I or II. To demonstrate further the relevance of these haplotypes, we compared the haplotype frequencies in the affected and unaffected members of the linked families with those of an independent Finnish reference population from the Nordic Database Lund-Malmö dataset [24]. These Finnish reference samples (referred to as Nordic–Finn, $n = 955$) are derived from a region in Bothnia in which there is evidence of western late settlement. The people in this region are genetically close to those in the region of western early settlement representing the origin of the linkage families [22]. Thus, these samples represented a good control and were also better matched with our study population than HapMap CEU individuals (CEPH; Utah residents with ancestry from northern and western Europe). Frequencies of haplotype core segments I and II were 0.30 and 0.33, respectively,
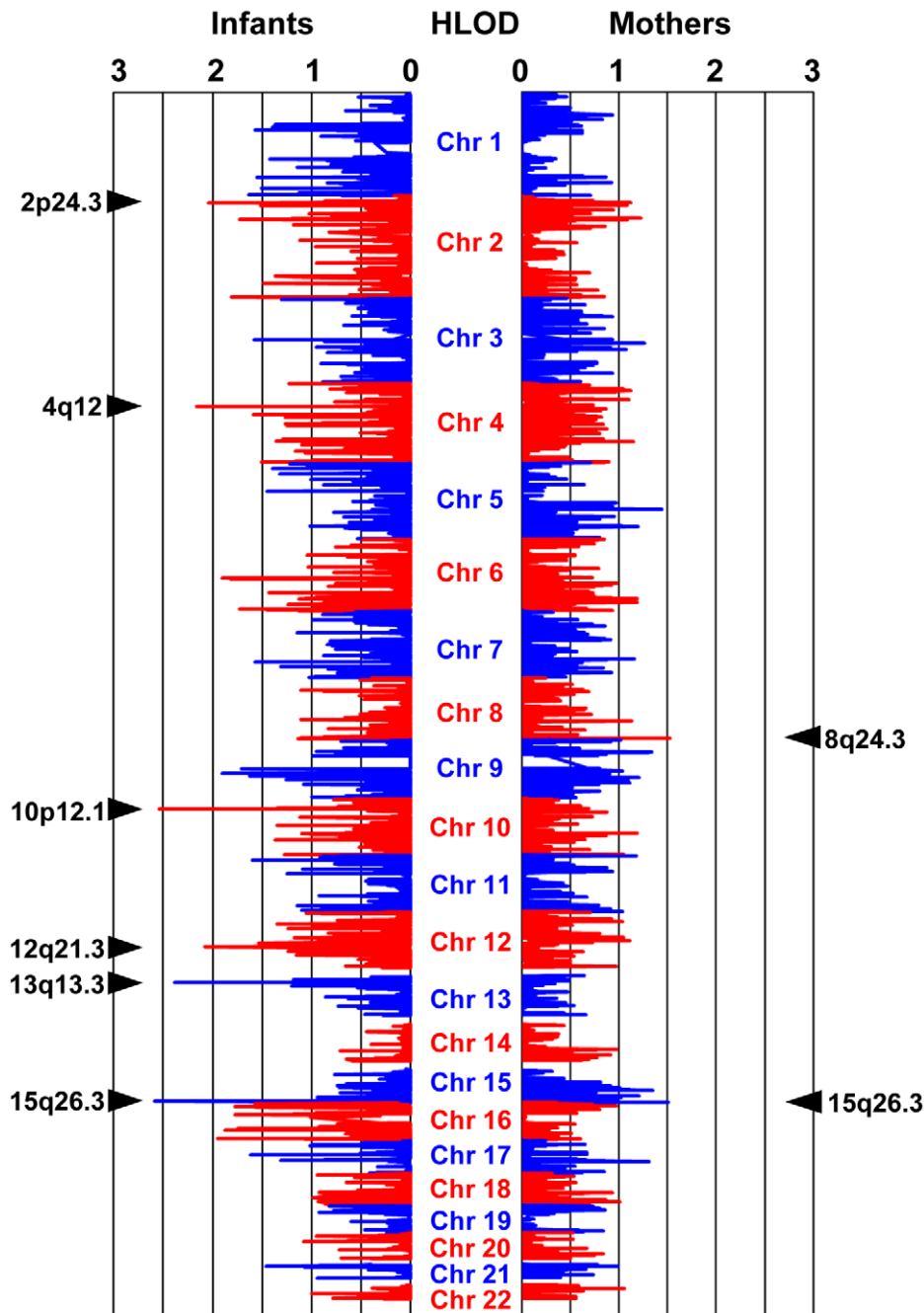
Figure 2. Heterogeneity LOD (HLOD) scores in parametric linkage analysis of spontaneous preterm birth. For the affected fetus phenotype (left panel), initial linkage signals (HLOD>2) were detected on chromosomes 2p24.3 (SNP rs6531074, HLOD = 2.05), 4q12 (SNP rs2037046, HLOD = 2.16), 10p12.1 (SNP SNP_A-419261, HLOD = 2.54), 12q21.3 (SNP rs7138288, HLOD = 2.08), 13q13.3 (SNP rs4245377, HLOD = 2.39), and 15q26.3 (SNP rs2715416, HLOD = 2.59) as indicated by arrows. For the mothers (right panel), the three maximum HLOD scores were detected on chromosomes 8q24.3 (SNP rs11167102, HLOD = 1.53) and 15q26.3 (SNPs rs11247268 and rs329292, HLODs of 1.48 and 1.51, respectively). doi:10.1371/journal.pgen.1001293.g002

in the affected members of the linkage families. Frequencies of core I and II haplotypes were 0.17 and 0.12, respectively, in the unaffected members of the linkage families, and 0.18 and 0.06 in the Nordic–Finn population. Thus, frequencies of the core haplotypes in the reference population were close to those of the unaffected members of the linkage families, while these frequencies were increased in the affected members of the linkage families. The higher frequency of haplotype core II in the unaffected members of the linkage families compared to Nordic–Finn

individuals is explained by the high overall incidence of this haplotype in the largest linked family (Family 150; Figure S2). In terms of the LD patterns (Figure S3) in the unaffected individuals, the genetic profile of our study population was representative of the Nordic–Finn reference population. Furthermore, there was a short two-SNP major-allele haplotype AT at rs11630259– rs1357112 shared IBS between core segments I and II in 100% of the affected inviduals ($n = 38$) and predicted carriers ($n = 18$) (Figure S2), whereas the same haplotype was present in only 62%

Table 1. HLOD values for the SNPs with significant signals in fine-scale linkage analysis of SPTB with the affected fetus phenotype.

	rs1521480	rs4966036	rs2684811
Chromosome region	15q26.3	15q26.3	15q26.3
Physical position	97,231,926	97,247,272	97,271,086
HLOD: All families	3.63[a]	3.63[a]	4.68[a]
HLOD: Per family			
Family 24	0.56	0.56	0.47
Family 70	1.37	1.37	1.31
Family 126	0.60	0.60	0.60
Family 150	0.38	0.38	0.70
Family 185	0.06	0.06	0.92
Family 210	0.02	0.02	0.03
Family 253	0.65	0.65	0.65

[a]Values of theta and alpha were 0.00 and 1.00, respectively, for all the markers.
doi:10.1371/journal.pgen.1001293.t001

of the unaffected family members ($n = 37$) and 72% of the Nordic-Finn individuals. Whether this particular segment sharing reflected the critical location of disease predisposition or arose by chance, could not be statistically evaluated in the current setting. Unfortunately, these two SNPs, which are not in LD with the surrounding region SNPs (Figure S3) and thus could not be imputed from other SNPs, failed to settle in the Sequenom iPLEX association platform and therefore could not be included in the following case-control association study.

We additionally performed a *post hoc* linkage analysis in which we used the transmission information obtained from the segregation analysis model, with both unaffected carriers and individuals born preterm defined as affected ($n_{affected} = 55$). An overall increase in the HLOD scores was observed within *IGF1R*, with an $HLOD_{max}$ of 3.81 at rs2715416 after pruning and an $HLOD_{max}$ of 5.15 at rs2684811 without pruning, further supporting the view that this genomic region may harbor a true susceptibility gene.

IGF1R Case-Control Association Analysis

To validate the potential linkage and association between *IGF1R* and SPTB, we enrolled a new Finnish population comprising cases originating from the northern (Oulu) and southern (Helsinki) regions of the country; these cases were independent of the families used for the linkage analysis. Tagging SNPs covering the entire *IGF1R* gene were determined using HapMap data from the CEU population. Among the 20 SNPs studied, two (rs7165181 and rs4966038) showed a weak association ($P<0.05$) in the infants but not in the mothers (Table 2). Even so, a 55-kb region of six SNPs in LD (Figure 4) spanning these two SNPs and extending to rs2715416 (which yielded the original linkage signal) showed statistically significant haplotype association with SPTB in the infants exclusively (Table 3). Similar associations were evident in both recurrent and sporadic SPTB. The signals obtained from independent sets of study populations using linkage, segregation and association approaches were colocalized within the same region of *IGF1R* (Figure 5 and Table S1) and were consistently observed in the infants but not the mothers; i.e., when the affected phenotype was being born preterm instead of giving birth to preterm infant(s). The birthweights and gestational ages of preterm infants carrying the associating haplotype ($n = 71$; $2,025\pm626$ g; 32.6 ± 3.0 wk; mean $\pm$ standard deviation) did not

differ significantly from those of preterm infants without this haplotype ($n = 263$; $2,105\pm616$ g; 32.3 ± 2.9 wk; P values of 0.64 and 0.35, respectively). Similar to our control population, the predicted frequency of the associating haplotype was 0.052 in the HapMap CEU population. In HapMap CHB (Han Chinese in Beijing, China) and JPT (Japanese in Tokyo, Japan) populations, allele and haplotype distibutions were completely different from our controls, as well as from the HapMap CEU population. The CHB and JPT populations completely lacked the associating haplotype. We were not able to estimate the frequency of this haplotype in the rest of the HapMap populations, including YRI (Yoruba in Ibadan, Nigeria), due to unavailable genotype data for part of the SNPs in these populations.

Discussion

According to epidemiological studies, there is evidence for a heritable predisposition to preterm birth with both maternal and fetal contribution [11]. To our knowledge, the present report describes the first genome-wide investigation and the first linkage study to identify genomic regions associated with SPTB. We detected significant parametric HLOD scores in the infants for three intronic markers (rs1521480, HLOD = 3.63; rs4966936, HLOD = 3.63; and rs2684811, HLOD = 4.68) on chromosome 15q26.3 within a single candidate gene, *IGF1R*, encoding the type 1 insulin-like growth factor receptor IGF-1R (Table 1).

We identified one major disease-cosegregating haplotype (core I) in all but one of the six linked families (Figure 3 and Figure S2). Rather than IBD, this sharing is likely to be IBS, because of the high frequency of this haplotype in the population, a view consistent with the genealogical survey, which suggested no common ancestry among the linked families. The occurrence of an alternative shared haplotype segment (core II) in all of family 185 and part of family 150 may reflect allelic heterogeneity or absence of true linkage in this subgroup, which would be typical of complex phenotypes even in an isolated population [25]. Taken together, our data on linkage and disease segregation of *IGF1R* SNPs are consistent with a role for this genomic region in SPTB under dominant MOI with incomplete penetrance allowing for healthy carriers or disease transmitters under etiological hetero-geneity (the existence of phenocopies). However, because the segregation pattern supports the model of disease-cosegregating *IGF1R* haplotype transmission *via* both healthy male and female carriers, our initial MOI based on pure unilineal maternal transmission is likely to be oversimplified.

We examined the gene encoding IGF-1R in a separate investigation with a Finnish population originating from two regions of the country, which revealed an association of a fetal *IGF1R* haplotype with SPTB (Table 3 and Figure 5). Both the pattern of maximal haplotype sharing in the linked families and the region of association observed in the case-control study independently placed SPTB susceptibility on the same segment within *IGF1R*. As a whole, these analyses provide evidence that the fetal *IGF1R* influences the risk of spontaneous preterm labor leading to preterm birth.

IGF-1R is a heterotetramer composed of two extracellular alpha subunits containing a ligand-binding site for IGFs and two membrane-spanning beta subunits harboring intracellular tyrosine kinase activity involved in a variety of cellular functions [26,27]. Upon activation by IGF-1 or IGF-2, IGF-1R participates in regulation of the cell cycle. Accordingly, certain *IGF1R* mutations result in intrauterine growth restriction, whereas polymorphisms of this ubiquitously expressed gene may not influence fetal growth [28–32]. The roles of IGF-1R in normal and pathological growth
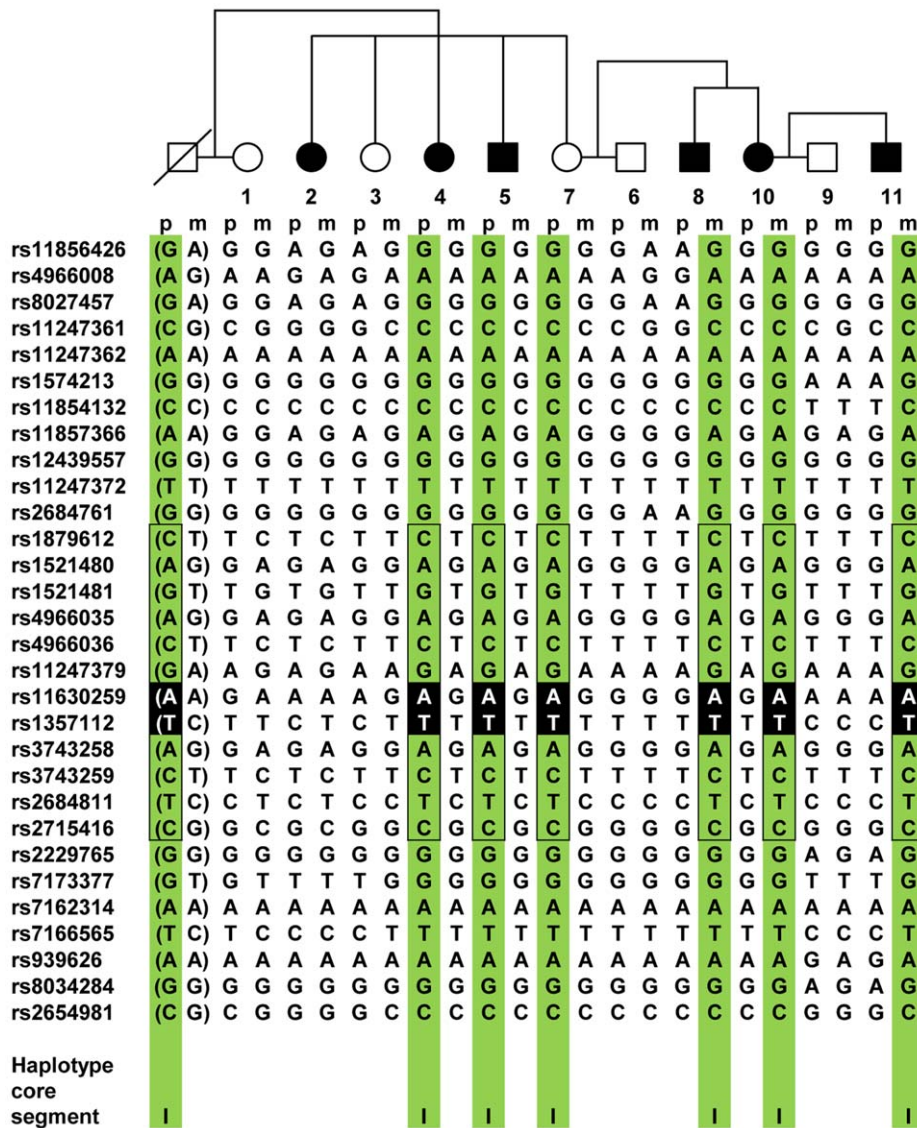
Figure 3. *IGF1R* haplotype segregation in family 70. Each successive column represents a single chromosome, with the paternal (p) and maternal (m) origin indicated above the haplotype. The individuals genotyped are indicated by the same numbers as in Figure S1. The bracketed () haplotypes were imputed for untyped individuals. Green-shaded areas indicate within-family haplotype sharing. Haplotype core segment I (55 kb), shared IBS between families 24, 70, 126, 253, and 150, is indicated by a solid box. The location of maximal haplotype sharing (rs11630259–rs1357112 haplotype AT) in all disease-cosegregating chromosomes in all families is shown by white letters on a black background. See text for details and Figure S2 for *IGF1R* haplotypes in all of the linked families.
doi:10.1371/journal.pgen.1001293.g003

and differentiation and aging involve interactions with the ubiquitous growth factors, hormones, and proinflammatory cytokines that are considered to be mediators of the labor process [33–35]. Several IGF-binding proteins that regulate IGF-1R–dependent signaling cascades have been studied in the context of SPTB, and some of them have been implicated in preterm labor [36–38]. However, these studies did not involve *IGF1R* and thus the extension of studies involving IGF-1R (particularly the ligand-binding alpha subunit encoded by exons 1–10) to the process of labor, with special consideration of fetal involvement in the endocrine and paracrine control of the preterm labor process, is clearly indicated. The range of regulatory roles performed by the IGF system is consistent with our view that *IGF1R* influences susceptibility to SPTB. Furthermore, a recent study revealed a parent-of-origin-specific methylated site within intron 2 of *IGF1R*

[39], which localizes to the same region that was identified in our haplotype segregation and case-control association analyses. This site was predominantly methylated on the maternally derived chromosome and may be involved in the imprinting process. This finding suggests a possible epigenetic mechanism through which *IGF1R* may be involved in the preterm labor process. In our linkage families, the core II *IGF1R* haplotype was maternally transmitted to the individuals born preterm, whereas the core I haplotype had a mixed transmission. This is an important aspect to consider in future studies involving IGF1R and preterm birth.

In the present study, we obtained no clear linkage or association signals when the outcome phenotype was giving preterm birth (affected mother phenotype). Because the number of affected mothers was small ($n = 21$), the power to detect regions of linkage was limited. Although the maternal $HLOD_{max}$ was <2, which we

Table 2. Case-control association analysis of *IGF1R* SNPs in SPTB.

Marker	SNP	Allele	Mothers (*n* = 348/143)		Infants (*n* = 334/197)	
			Allele frequency case/control	Allelic test *P* value	Allele frequency case/control	Allelic test *P* value
1	rs874305	A	0.474/0.496	0.519	0.467/0.454	0.677
2	rs11247361	C	0.641/0.651	0.570	0.640/0.640	1.000
3	rs4966013	G	0.407/0.364	0.211	0.415/0.391	0.449
4	rs10794486	A	0.747/0.734	0.676	0.733/0.727	0.827
5	rs907807	A	0.919/0.934	0.425	0.916/0.912	0.833
6	rs2684774	G	0.082/0.076	0.751	0.090/0.090	0.970
7	rs1879612	T	0.630/0.613	0.609	0.658/0.619	0.216
8	rs7165181	G	0.295/0.280	0.650	0.315/0.255	0.041[a]
9	rs4966038	C	0.375/0.359	0.635	0.408/0.343	0.037[a]
10	rs883149	A	0.789/0.825	0.202	0.824/0.814	0.712
11	rs8039666	A	0.881/0.906	0.264	0.887/0.878	0.688
12	rs2715416	G	0.389/0.402	0.708	0.412/0.369	0.166
13	rs1546714	G	0.984/0.983	0.859	0.988/0.979	0.282
14	rs12438493	G	0.571/0.563	0.808	0.605/0.584	0.508
15	rs2229765	G	0.573/0.570	0.937	0.609/0.588	0.513
16	rs2684799	A	0.429/0.392	0.287	0.437/0.405	0.318
17	rs2684792	T	0.628/0.629	0.995	0.673/0.629	0.153
18	rs7166565	G	0.603/0.608	0.880	0.640/0.611	0.340
19	rs2684789	C	0.445/0.465	0.560	0.521/0.531	0.739
20	rs2016347	C	0.481/0.444	0.291	0.478/0.469	0.771

[a]$P > 0.05$ after permutation correction.
doi:10.1371/journal.pgen.1001293.t002

considered to be below an initial threshold suggesting linkage in this study, we identified the regions yielding the highest maternal linkage signals (HLODs of 1.53 and 1.51 for markers rs11167102 and rs329292 on chromosomes 8q24.3 and 15q26.3 respectively). An interesting finding was that the SNP with the maternal HLOD score of 1.51 was located on the same region as the one with the HLOD_{max} obtained from the infants, separated by 2 Mb (Figure 5). To conclude, the lack of association in the mothers and the observation of both maternal and paternal transmission in the segregation analysis unexpectedly did not support a major maternal contribution in *IGF1R*-mediated genetic susceptibility to SPTB. Our discovery of a fetal gene was unforeseen, but does not exclude the role of maternal effects *via* other susceptibility genes. Interestingly, a recent study suggested that the fetal genome plays a more significant role than the maternal genome in individuals of European ancestry [40]. However, other studies have pointed out the importance of the maternal genome in genetic predisposition to SPTB [41,42].

Our study has several strengths compared to previous genetic studies of preterm birth. These candidate gene studies have been limited in several respects, such as small or mixed populations that may represent only the mother or the fetus, lack of replication, inconsistencies in phenotype definitions and incomplete coverage of variation within a gene [11]. In contrast, our study first focused on detecting novel genes involved in SPTB using a nonhypothesis-driven genome-wide approach. We evaluated both the maternal and fetal contribution. Additionally, careful attention was paid to selection of the families for the linkage analysis to ensure a precise phenotype definition. Lastly, we replicated the finding of a linked and disease-cosegregating region in an independent case-control cohort covering a large part of the entire country. The effect size we observed in the infants fell within the range of power calculations described in Materials and Methods. Although the Finnish population represents a higher genetic similarity than most populations, a clear substructure caused by population history is known to exist within the country [22]. However, the linkage and segregation analyses were performed using a relatively homogenous study population, which is amenable to using a search for shared genomic segments [25]. Our case-control cohort comprised two regions (northern Finland and southern Finland/Helsinki) known to represent slightly distinct patterns of genetic variation; the Helsinki region is known to be representative of a genetically more diverse population [22,43]. The fact that the association was observed across these two subpopulations gives further credence to our findings.

Our results constitute proof of principle for using a genome-wide SNP linkage scan to elucidate a complex heterogenic trait *via* stringent initial selection of a limited set of individuals. Despite the important strengths mentioned above, our study has some limitations. The result remains yet to be confirmed and extended by independent case-control analyses in populations with larger sample sizes than in the current study, and by studying additional SNPs covering the large gene more extensively. It is possible that while the ethnically homogenous nature of our Finnish linkage families can facilitate detection of genetic factors, the relatively high degree of genetic similarity among the Finnish population compared to others may mean that this finding cannot be generalized to more outbred populations. The frequency of the associating *IGF1R* haplotype in the HapMap CEU population was similar (0.052) to the frequency of controls in our case-control
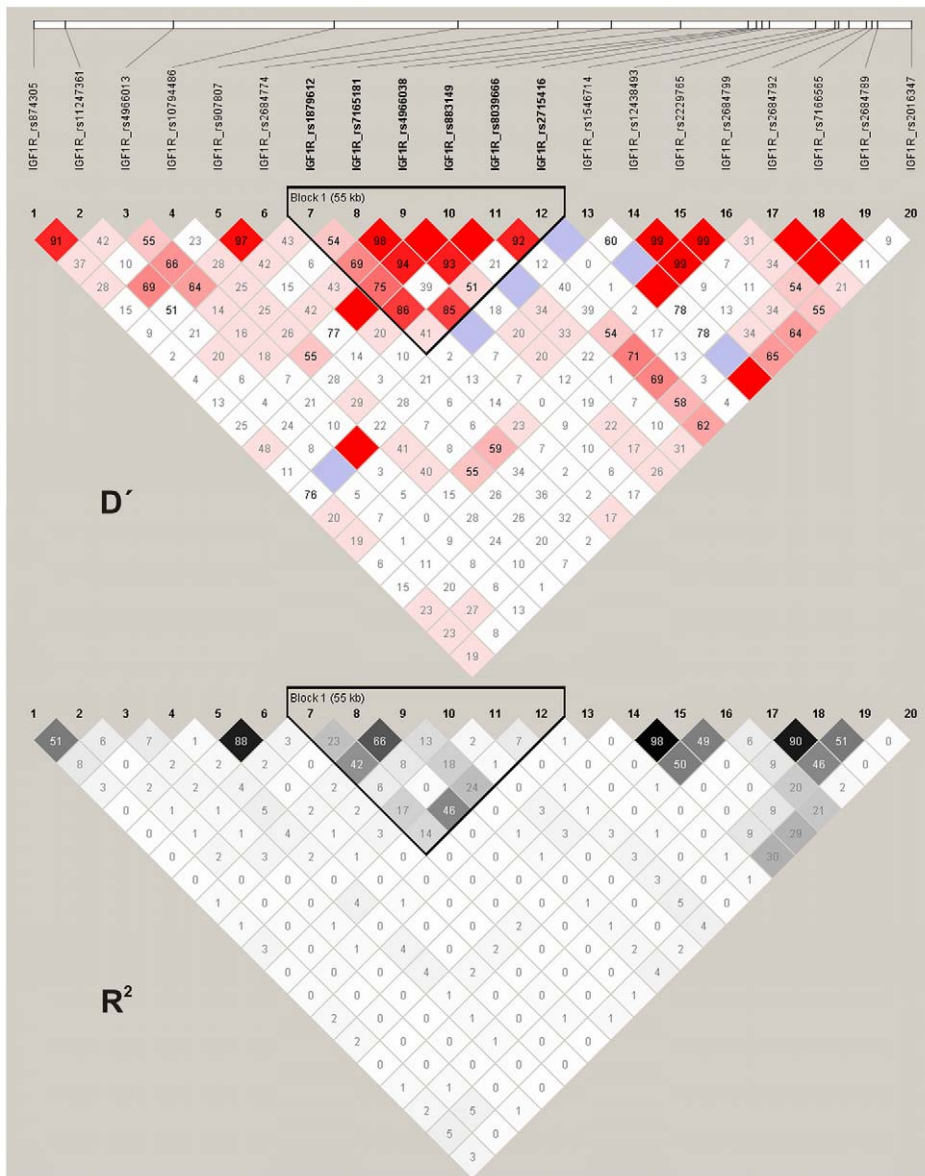
Figure 4. Pairwise D prime and r^2 LD plots of *IGF1R* SNPs in the association analysis. The largest 55 kb region of LD spans the two associating SNPs rs7165181 (marker 8) and rs4966038 (marker 9), extending to the SNPs rs2684811 (between markers 11 and 12; the highest linkage signal) and rs2715416 (marker 12; the original linkage signal).
doi:10.1371/journal.pgen.1001293.g004

analysis (Table 3), while this haplotype was completely lacking from HapMap CHB and JPT populations, reflecting population-specific variation within this genomic segment. In the current setting, it cannot be evaluated whether the observed association was due to a haplotypic effect or another linked polymorphic site that was not analyzed in the current study. Therefore, this region needs to be more thoroughly investigated to find the causative sequence variant(s) and also to determine whether other populations may have an *IGF1R*-mediated risk of SPTB. Furthermore, although we discussed here only regions with an HLOD of >3, other regions with initial linkage signals on chromosomes 2, 4, 10, 12, and 13 should not be ignored as candidates for the risk of spontaneous preterm labor leading to preterm birth.

In conclusion, our genetic linkage and haplotype segregation analysis mapped the novel fetal SPTB susceptibility gene *IGF1R* on chromosome 15q26.3; this was further confirmed by association in an independent case-control population replicate. This result was unexpected, because the studies on the etiology of SPTB have been focused on cytokine-mediated inflammatory signaling as a possible route of predisposition by the maternal genome. Future clarification of the molecular mechanism of a growth factor pathway with considerable influence on the heritable proportion of the risk of spontaneous premature labor and birth is likely to open up a new potential avenue for preventive therapies.

Materials and Methods

Ethics Statement

Written informed consent was obtained from the participants and the study was approved separately by the Ethics Committee of Oulu University Hospital and that of Helsinki University Central Hospital.

Table 3. *IGF1R* haplotype case-control association analysis of SPTB in the infants.

Haplotype[a]	Haplotype frequency case/control	*P* value	Corrected *P* value	OR (95% CI)[b]
TAGAAC	0.367/0.410	0.176	NS	
CGCAAG	0.179/0.172	0.770	NS	
TGCAAG	0.110/0.052	0.001	0.021	2.3 (1.4–3.8)
CACACC	0.085/0.076	0.611	NS	
TAGTAC	0.085/0.075	0.556	NS	
TAGTAG	0.072/0.068	0.800	NS	
CAGAAG	0.025/0.034	0.392	NS	
CAGTAG	0.013/0.039	0.007	NS	
CGCACC	0.019/0.028	0.326	NS	
CAGAAC	0.015/0.023	0.359	NS	

[a]Markers in the haplotype: rs1879612, rs7165181, rs4966038, rs883149, rs8039666, and rs2715416; see Table 2 and Figure 4.
[b]OR = odds ratio, CI = confidence interval.
doi:10.1371/journal.pgen.1001293.t003

Study Population

Families for linkage study and haplotype-sharing analysis. We selected families for the linkage analysis from among the population born in the Oulu University Hospital region. We selected mothers with recurrent SPTB from approximately 120,000 birth diaries between 1975–2005 (prospectively from 2002). To avoid misclassification bias at borderline gestational ages, we defined SPTB as spontaneous onset of labor and birth before 36 (rather than ≤37) completed weeks of gestation. Both labors initiated with intact membranes and those following premature rupture of fetal membranes (PROM, defined as leakage of amniotic fluid as the presenting symptom before the onset of contractions) were included among the SPTBs. Additional exclusion criteria were known risk factors for SPTB (multiple gestation, polyhydramnios, septic infection or chronic disease of the mother, narcotic or alcohol abuse, accidents, and fetuses with congenital anomalies) and all elective preterm births without labor (preeclampsia, intrauterine growth restriction, and placental abruption). Only families of Finnish origin were included. We identified 120 families altogether, with at least two spontaneous singleton preterm births. The family interview revealed that 20 of them were actually large families with multiple relatives affected by SPTB. To search for potential consanguinity, we performed a genealogical study in accordance with the published criteria [44], tracing ancestors from the Finnish Population Registries and scrutinizing microfiche copies available in the provincial and national archives of Finland. The genealogical survey showed that most of the large families originated from Northern Ostrobothnia. However, neither close consanguinity nor the common residence of ancestors dating from the 17th century was apparent. Because prior evidence suggests that a maternal family history of preterm birth may have a greater influence on this disorder, we selected families with apparent maternal inheritance of predisposition to preterm birth, excluding families with seemingly paternal or mixed transmission of the preterm birth phenotype. Using these criteria, seven families were selected. Whole-blood DNA samples (*n* = 89) were taken from both affected and unaffected family members. The pedigrees of the families appear in Figure S1. Six of the mothers who gave preterm birth were themselves born preterm.

Subjects for the *IGF1R* case-control association study. We selected mothers with a history of at least one SPTB and their preterm infants from singleton pregnancies from the regions of Oulu (northern Finland) and Helsinki (southern Finland) University Hospitals for an association study using the same phenotypic criteria as for the linkage analysis, resulting in a case population of 348 mothers (244 from Oulu and 104 from Helsinki) and 334 infants (238 from Oulu and 96 from Helsinki). One infant *per* family was included using the criteria described [45]. One hundred and forty-six (42.6%) of the mothers gave birth following PROM and 197 (57.4%) had labor initiated with intact membranes, while 128 (39.5%) of the infants were born following PROM and 196 (60.5%) without PROM. For five mothers and ten infants, there was no information concerning the occurrence of PROM. The control population consisted of 143 mothers and 197 newborn singleton infants who were prospectively recruited from Oulu University Hospital in 2004–2005 among mothers who had at least three exclusively term deliveries with no pregnancy- or labor-associated complications. It is of note that this part of the study was not restricted to familial SPTB, because it included mothers with both recurrent (*n* = 79) and sporadic (*n* = 265) SPTB and their preterm infants (recurrent *n* = 76 and sporadic *n* = 258). All of the individuals from the families included in the linkage analysis and their relatives were excluded from the association study.

DNA Sample Preparation and Genotyping

Linkage study. DNA was prepared from whole blood specimens (*n* = 89) by standard methods. DNA samples were genotyped in the Broad Institute Center for Genotyping and Analysis (CGA) using the Affymetrix Genome-Wide Human SNP Array 5.0Kb consisting of 500,568 SNP markers.

Association study. Whole blood (*n* = 797) and buccal cell samples (*n* = 305) were used for DNA extraction. Buccal DNA was extracted using Chelex 100 (Bio-Rad, Hercules, CA, USA), after which it was were whole-genome amplified (WGA) in duplicate reactions using the Illustra GenomiPhi V2 DNA Amplification kit (GE Healthcare Sciences, Cardiff, UK), followed by pooling of the duplicates and purification with Illustra Microspin G-50 columns (GE Healthcare Sciences). The quality and quantity of the WGA samples was confirmed with ethidium bromide–stained agarose gels and UV absorbance measurements, and by including nontemplate reactions to control for DNA contamination. In the set-up stage, we genotyped the WGA products in parallel with the corresponding unamplified DNA for a set of test SNPs, yielding >99% consistent genotypes. *IGF1R* SNP genotyping was performed at the Institute for Molecular Medicine Finland FIMM Technology Centre using the Sequenom iPLEX Gold assay on the MassARRAY Platform for user-defined SNP sets.
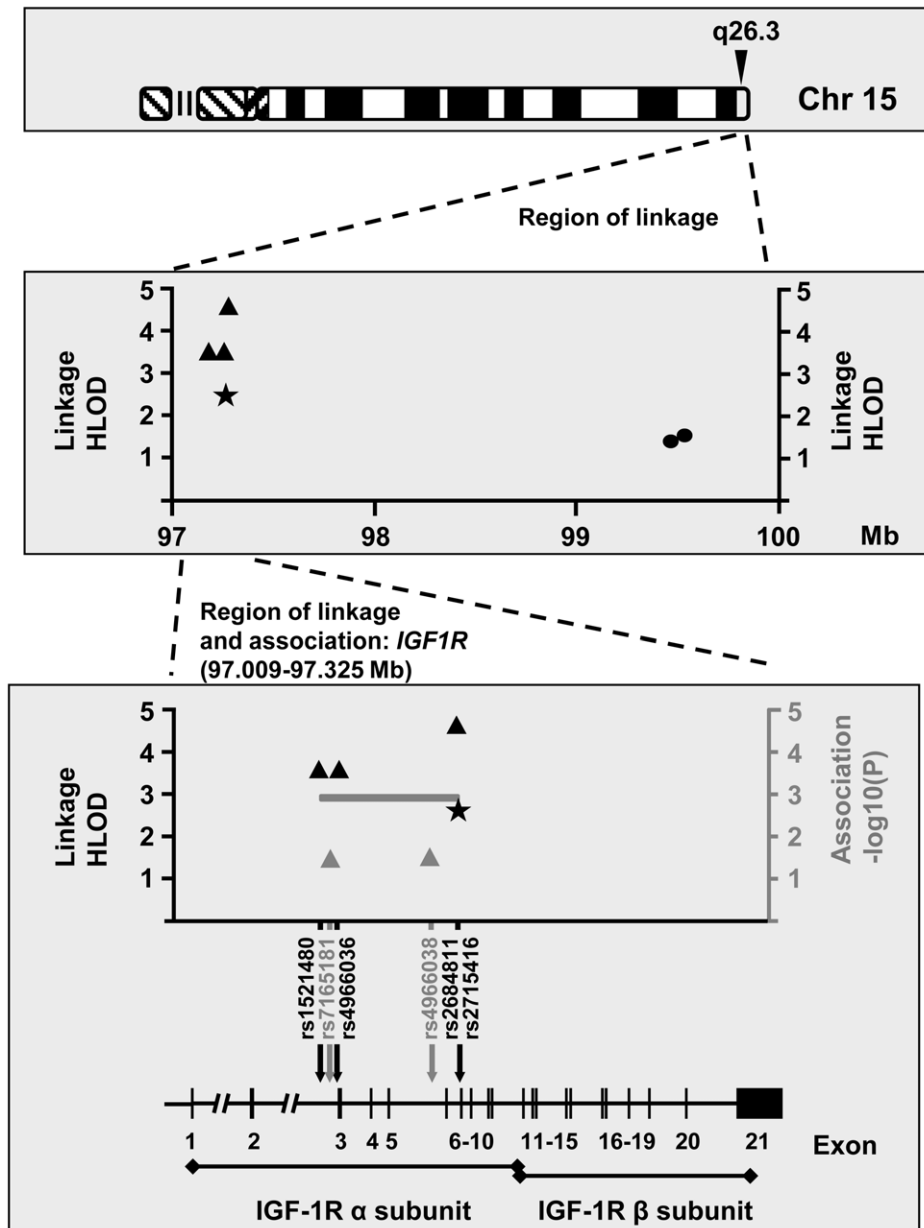
Figure 5. Illustration of SNPs yielding the most significant linkage and association signals on chromosome 15q26.3. The most significant linkage (black symbols) and association (gray symbols) signals in the infants (triangles) and mothers (rounded dots) are shown. *IGF1R* contains 21 exons at 97.0–97.3 Mb with the IGF-1R alpha and beta subunits encoded by exons 1–11 and 11–21, respectively. Black asterisk represents the initial linkage peak obtained from the infants using the 6,377 LD-pruned SNPs (rs2715416 at position 97,271,554). The region of linkage and association localizes at a 55-kb LD interval extending from intron 2 to intron 6 (gray line). A statistically significant association was detected for haplotype TGCAAG (unadjusted $P = 0.001$, log10(P) = 2.89, and permutation-corrected $P = 0.021$).
doi:10.1371/journal.pgen.1001293.g005

Affymetrix Data Processing, Error Checking, and SNP–Marker Pruning for Linkage Analysis

We processed the Affymetrix array data files using PLINK, v. 1.02 [46]. We used a 1 Mb-to-1 cM converted map, as justified in [47]. Before linkage analysis, the markers were pruned using a linkage disequilibrium (LD) r^2 threshold of 0.35 with PLINK. Markers with either minor allele frequency (MAF)<0.08, genotyping failure >0.1 or Mendelian errors were excluded, as well as those violating the Hardy–Weinberg equilibrium (HWE, P<0.001). Data processing after removal of the high-LD SNPs yielded a selection of 6,377 autosomal SNPs for the genome-wide

linkage analysis. We rechecked the genotype data for any Mendelian inconsistencies using PedCheck [48] before proceeding to linkage analysis.

Linkage Analysis

We studied two main outcome phenotypes for SPTB: (1) being spontaneously born preterm (affected fetus/infant, $n_{affected} = 41$), and (2) giving spontaneous preterm birth (affected mother, $n_{affected} = 21$) (pedigrees in Figure S1). A parametric two-point linkage analysis of SPTB as a dichotomous trait was performed with ANALYZE, v. 1.9.3 BETA, which is a linkage and LD

analysis package that uses FASTLINK 4.1P to calculate the pedigree likelihoods and is capable of managing both extended pedigrees and large numbers of markers [49]. We used a dominant low-penetrance model, assuming a disease allele frequency of 0.001 and penetrances of 0.001, 0.001, and 0 for the homozygotes, heterozygotes and wild-type homozygotes, respectively. This kind of model is nearly convergent with a model-free analysis, because it minimizes the effect of misspecifying the true MOI while retaining the highest power of a parametric analysis to detect linkage [50,51]. Parametric linkage in the presence of heterogeneity was assessed using heterogeneity LOD (HLOD) scores and their accompanying estimates of the proportion of linked families (α) and recombination fraction (θ).

We applied two-point linkage analysis to avoid bias that could result in multipoint analyses due to missing parental genotypes and markers residing in LD [52]. Our linkage analysis strategy was to perform an analysis with an LD-pruned marker set and then to screen regions with initial linkage signals (HLOD>2) with a denser marker map including all markers with MAF>0.08 on the region. Using high-density SNPs in linkage analysis has the advantages of a low error rate, high per-marker call rates, and higher information content when compared to microsatellites [53]. In two-point analysis, LD does not increase type 1 error [54].

When a marker showed HLOD>2 (six positions), we selected the genomic region for further analysis. This fine-scale linkage analysis was performed using the unpruned marker set on the flanking region (~5 Mb) of each of the initial linkage signals with the same parametric model as the original scan. We considered an HLOD score of >3 as a signal of linkage.

Haplotype Segregation Analysis

We used PedPhase version 2.0 utilizing integer linear programming (ILP) to find the minimum-recombinant haplotype configuration in the linked families 24, 70, 126, 150, 185 and 253 for the SNPs flanking the best linkage peak [55]. Deceased or untyped individuals were also included if their haplotypes could be reconstructed from their relatives. The potential presence of disease-cosegregating haplotypes ("affected haplotypes", affected fetus phenotype) was evaluated using a model compatible with the assumption of maternal unilineal inheritance with incomplete penetrance, utilizing the scheme of the most likely segregation pattern to search for a minimum number of affected haplotypes. The model assumed dominant MOI, allowing for the presence of healthy carriers as transmitters of an affected haplotype.

We used an independent Finnish reference population of the Nordic Database Lund-Malmö dataset [24]. These samples (referred to as Nordic–Finn, $n = 955$) are derived from a region in Bothnia in which there is evidence of western late settlement. We used Beagle 3.1 [56] to infer the haplotype phases for unrelated individuals in the Nordic–Finn samples.

IGF1R Case-Control Study: SNP Selection and Association Analysis

IGF1R is a large, >300 kb gene with nearly 2,000 SNPs (NCBI B36 assembly dbSNP b126), but little frequent exonic variation; it has relatively weak intragenic LD and is not readily divided into discrete blocks of limited haplotype diversity. Using HapMap data (release 23a/phaseII) from the CEU population to determine tagging SNPs (tSNPs) covering the entire *IGF1R* gene (chr15: 97,004,502–97,329,396), we obtained a list of 100 tSNPs (pairwise tagging, r^2 cutoff 0.8, MAF>0.1). Given the number of individuals available for the study and the power to detect associations, it was reasonable to set up a single iPLEX set, which can include up to

approximately 30 compatible SNPs. To select SNPs, we visualized the gene region's LD pattern in the HapMap CEU population in the Haploview program v. 4.1 [57] and selected one SNP with the highest MAF from each haplotype block designated by this program. We also included the coding SNP rs2229765 (Glu1043-Glu, also referred to as 3174 G>A) because individuals carrying at least one A allele are reported to have lower levels of free plasma IGF-1 than GG homozygotes [58], suggesting that this polymorphism may have a functional effect. After testing and validation, a selection of 20 SNPs (listed in Table 2 and details in Table S1) remained for the case-control association study.

We performed statistical tests using Haploview v. 4.1 [57]. We took multiple testing into account by performing permutations with 10,000 replicates and considered a corrected *P* value of <0.05 to be statistically significant. For analysis of haplotypes with birthweight and gestational age, we inferred phased *IGF1R* haplotypes from unphased data using Beagle 3.2 [56]. To analyze potential associations between haplotypes and birthweight or gestational age, the nonparametric Mann-Whitney U-test was used with Predictive Analytics SoftWare (PASW) statistics, version 17.0.3 (IBM SPSS, Inc.). Power consideration of the case-control study: the sample size provides 80% power (alpha = 0.0025, allowing for multiple testing of 20 SNPs) to detect the risk allele carrier relative risks of approximately 2.1–2.7 in the infants and 2.4–3.1 in the mothers for allele frequencies ranging from 0.2 down to 0.05, considering SPTB as a discrete trait, assuming causal SNP, a prevalence of 0.05, and using the allelic 1 df test [59].

Genome Coordinates

All of the chromosomal positions refer to NCBI Build 36 of the human genome.

Web Resources

International HapMap Project: http://www.hapmap.org
National Center for Biotechnology: http://www.ncbi.nlm.nih.gov
Nordic Database: http://www.nordicdb.org
Genetic Power Calculator: http://pngu.mgh.harvard.edu/~purcell/gpc

Supporting Information

Figure S1 Pedigrees of the seven families with recurrent SPTB included in the linkage analysis.
Found at: doi:10.1371/journal.pgen.1001293.s001 (0.46 MB PDF)

Figure S2 *IGF1R* haplotype segragation in the linked families.
Found at: doi:10.1371/journal.pgen.1001293.s002 (2.21 MB PDF)

Figure S3 Pairwise marker linkage disequilibrium (LD) in the unaffected members of the linked families and the reference population.
Found at: doi:10.1371/journal.pgen.1001293.s003 (1.92 MB PDF)

Table S1 Combined results of *IGF1R* SNPs in fine-scale linkage analysis and association analysis with the affected fetus phenotype.
Found at: doi:10.1371/journal.pgen.1001293.s004 (0.08 MB PDF)

Author Contributions

Conceived and designed the experiments: RH MKK AL KT HP MO TV JCM SAM LP LJM AP MH. Performed the experiments: RH MKK BPC JP. Analyzed the data: RH MKK AL TV BPC JP SAM LJM AP. Contributed reagents/materials/analysis tools: LJM MH. Wrote the paper: RH MKK. Contributed in recruiting patients: KT HP. Contributed in obtaining patient data and DNA: HP. Contributed in obtaining the DNA array: LP. Contributed in writing the paper: AP MH. Served as a liaison between the investigators: MH.

References

1. Damus K (2008) Prevention of preterm birth: A renewed national priority. Curr Opin Obstet Gynecol 20(6): 590–596.
2. Goldenberg RL, Culhane JF, Iams JD, Romero R (2008) Epidemiology and causes of preterm birth. Lancet 371(9606): 75–84.
3. Esplin MS, O'Brien E, Fraser A, Kerber RA, Clark E, et al. (2008) Estimating recurrence of spontaneous preterm delivery. Obstet Gynecol 112(3): 516–523.
4. Bakketeig LS, Hoffman HJ, Harley EE (1979) The tendency to repeat gestational age and birth weight in successive births. Am J Obstet Gynecol 135(8): 1086–1103.
5. Muglia LJ, Katz M (2010) The enigma of spontaneous preterm birth. N Engl J Med 362(6): 529–535.
6. Porter TF, Fraser AM, Hunter CY, Ward RH, Varner MW (1997) The risk of preterm birth across generations. Obstet Gynecol 90(1): 63–67.
7. 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.
8. Kistka ZA, DeFranco EA, Ligthart 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.
9. Clausson B, Lichtenstein P, Cnattingius S (2000) Genetic influence on birthweight and gestational length determined by studies in offspring of twins. BJOG 107(3): 375–381.
10. Treloar SA, Macones GA, Mitchell LE, Martin NG (2000) Genetic influences on premature parturition in an Australian twin sample. Twin Res 3(2): 80–82.
11. Plunkett J, Muglia LJ (2008) Genetic contributions to preterm birth: Implications from epidemiological and genetic association studies. Ann Med 40(3): 167–195.
12. 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: Asembo Bay Cohort project IX. Genet Epidemiol 21(3): 201–211.
13. 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.
14. Amory JH, Adams KM, Lin MT, Hansen JA, Eschenbach DA, et al. (2004) Adverse outcomes after preterm labor are associated with tumor necrosis factor-alpha polymorphism -863, but not -308, in mother-infant pairs. Am J Obstet Gynecol 191(4): 1362–1367.
15. Simhan HN, Krohn MA, Roberts JM, Zeevi A, Caritis SN (2003) Interleukin-6 promoter -174 polymorphism and spontaneous preterm birth. Am J Obstet Gynecol 189(4): 915–918.
16. 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.
17. 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.
18. Kalish RB, Vardhana S, Gupta M, Perni SC, Witkin SS (2004) Interleukin-4 and -10 gene polymorphisms and spontaneous preterm birth in multifetal gestations. Am J Obstet Gynecol 190(3): 702–706.
19. Macones GA, Parry S, Elkousy M, Clothier B, Ural SH, et al. (2004) A polymorphism in the promoter region of TNF and bacterial vaginosis: Preliminary evidence of gene-environment interaction in the etiology of spontaneous preterm birth. Am J Obstet Gynecol 190(6): 1504–8; discussion 3A.
20. Moore S, Ide M, Randhawa M, Walker JJ, Reid JG, et al. (2004) An investigation into the association among interleukin, cytokine gene polymorphisms and periodontal disease. BJOG 111(2): 125–132.
21. 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.
22. 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.
23. Onkamo P, Toivonen H (2006) A survey of data mining methods for linkage disequilibrium mapping. Hum Genomics 2(5): 336–340.
24. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical Research, Saxena R, Voight BF, Lyssenko V, Burtt NP, et al. (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. Science 316(5829): 1331–1336.
25. Houwen RH, Baharloo S, Blankenship K, Raeymaekers P, Juyn J, et al. (1994) Genome screening by searching for shared segments: Mapping a gene for benign recurrent intrahepatic cholestasis. Nat Genet 8(4): 380–386.
26. Abbott AM, Bueno R, Pedrini MT, Murray JM, Smith RJ (1992) Insulin-like growth factor I receptor gene structure. J Biol Chem 267(15): 10759–10763.
27. Vitale L, Lenzi L, Huntsman SA, Canaider S, Frabetti F, et al. (2006) Differential expression of alternatively spliced mRNA forms of the insulin-like growth factor 1 receptor in human neuroendocrine tumors. Oncol Rep 15(5): 1249–1256.
28. Abuzzahab MJ, Schneider A, Goddard A, Grigorescu F, Lautier C, et al. (2003) IGF-I receptor mutations resulting in intrauterine and postnatal growth retardation. N Engl J Med 349(23): 2211–2222.
29. Kawashima Y, Kanzaki S, Yang F, Kinoshita T, Hanaki K, et al. (2005) Mutation at cleavage site of insulin-like growth factor receptor in a short-stature child born with intrauterine growth retardation. J Clin Endocrinol Metab 90(8): 4679–4687.
30. Walenkamp MJ, van der Kamp HJ, Pereira AM, Kant SG, van Duyvenvoorde HA, et al. (2006) A variable degree of intrauterine and postnatal growth retardation in a family with a missense mutation in the insulin-like growth factor I receptor. J Clin Endocrinol Metab 91(8): 3062–3070.
31. Inagaki K, Tiulpakov A, Rubtsov P, Sverdlova P, Peterkova V, et al. (2007) A familial insulin-like growth factor-I receptor mutant leads to short stature: Clinical and biochemical characterization. J Clin Endocrinol Metab 92(4): 1542–1548.
32. Ester WA, Hokken-Koelega AC (2008) Polymorphisms in the IGF1 and IGF1R genes and children born small for gestational age: Results of large population studies. Best Pract Res Clin Endocrinol Metab 22(3): 415–431.
33. Clemmons DR (2007) Modifying IGF1 activity: An approach to treat endocrine disorders, atherosclerosis and cancer. Nat Rev Drug Discov 6(10): 821–833.
34. Laviola L, Natalicchio A, Giorgino F (2007) The IGF-I signaling pathway. Curr Pharm Des 13(7): 663–669.
35. Himpe E, Kooijman R (2009) Insulin-like growth factor-I receptor signal transduction and the janus Kinase/Signal transducer and activator of transcription (JAK-STAT) pathway. Biofactors 35(1): 76–81.
36. Lo HC, Tsao LY, Hsu WY, Chen HN, Yu WK, et al. (2002) Relation of cord serum levels of growth hormone, insulin-like growth factors, insulin-like growth factor binding proteins, leptin, and interleukin-6 with birth weight, birth length, and head circumference in term and preterm neonates. Nutrition 18(7–8): 604–608.
37. Cooley SM, Donnelly JC, Collins C, Geary MP, Rodeck CH, et al. (2010) The relationship between maternal insulin-like growth factors 1 and 2 (IGF-1, IGF-2) and IGFBP-3 to gestational age and preterm delivery. J Perinat Med 38(3): 255–259.
38. Rahkonen L, Rutanen EM, Nuutila M, Sainio S, Saisto T, et al. (2010) Elevated levels of decidual insulin-like growth factor binding protein-1 in cervical fluid in early and mid-pregnancy are associated with an increased risk of spontaneous preterm delivery. BJOG 117(6): 701–710.
39. Sharp AJ, Migliavacca E, Dupre Y, Stathaki E, Sailani MR, et al. (2010) Methylation profiling in individuals with uniparental disomy identifies novel differentially methylated regions on chromosome 15. Genome Res 20(9): 1271–1278.
40. York TP, Strauss JF, 3rd, Neale MC, Eaves LJ (2010) Racial differences in genetic and environmental risk to preterm birth. PLoS ONE 5: e12391. doi:10.1371/journal.pone.0012391.
41. 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.
42. 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.
43. Nelis M, Esko T, Magi R, Zimprich F, Zimprich A, et al. (2009) Genetic structure of Europeans: A view from the North-East. PLoS ONE 4: e5472. doi:10.1371/journal.pone.0005472.
44. Varilo T (1999) The age of the mutations in the Finnish disease heritage; a genealogical and linkage disequilibrium study. PhD thesis, University of Helsinki, Department of Medical Genetics, Faculty of Medicine and Department of Human Molecular Genetics, National Public Health Institute, Helsinki. 98 p.
45. Salminen A, Paananen R, Karjalainen MK, Tuohimaa A, Luukkonen A, et al. (2009) Genetic association of SP-C with duration of preterm premature rupture of fetal membranes and expression in gestational tissues. Ann Med 41(8): 629–642.
46. 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.
47. Ulgen A, Li W (2005) Comparing single-nucleotide polymorphism marker-based and microsatellite marker-based linkage analyses. BMC Genet 6 Suppl 1: S13.

48. 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.

49. 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.

50. 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.

51. Strauch K, Fimmers R, Baur MP, Wienker TF (2003) How to model a complex trait. 1. general considerations and suggestions. Hum Hered 55(4): 202–210.

52. 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: S72–6.

53. Wang S, Huang S, Liu N, Chen L, Oh C, et al. (2005) Whole-genome linkage analysis in mapping alcoholism genes using single-nucleotide polymorphisms and microsatellites. BMC Genet 6 Suppl 1: S28.

54. Huang Q, Shete S, Swartz M, Amos CI (2005) Examining the effect of linkage disequilibrium on multipoint linkage analysis. BMC Genet 6 Suppl 1: S83.

55. Li J, Jiang T (2005) Computing the minimum recombinant haplotype configuration from incomplete genotype data on a pedigree by integer linear programming. J Comput Biol 12(6): 719–739.

56. Browning SR, Browning BL (2007) Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. Am J Hum Genet 81(5): 1084–1097.

57. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics 21(2): 263–265.

58. Bonafe M, Barbieri M, Marchegiani F, Olivieri F, Ragno E, et al. (2003) Polymorphic variants of insulin-like growth factor I (IGF-I) receptor and phosphoinositide 3-kinase genes affect IGF-I plasma levels and human longevity: Cues for an evolutionarily conserved mechanism of life span control. J Clin Endocrinol Metab 88(7): 3299–3304.

59. Purcell S, Cherny SS, Sham PC (2003) Genetic power calculator: Design of linkage and association genetic mapping studies of complex traits. Bioinformatics 19(1): 149–150.