Pediatric RESEARCH

REVIEW ARTICLE

Spontaneous premature birth as a target of genomic research

Mikko Hallman[1], Antti Haapalainen[1], Johanna M. Huusko[2], Minna K. Karjalainen[1], Ge Zhang[2], Louis J. Muglia[2] and Mika Rämet[1]

Spontaneous preterm birth is a serious and common pregnancy complication associated with hormonal dysregulation, infection, inflammation, immunity, rupture of fetal membranes, stress, bleeding, and uterine distention. Heredity is 25–40% and mostly involves the maternal genome, with contribution of the fetal genome. Significant discoveries of candidate genes by genome-wide studies and confirmation in independent replicate populations serve as signposts for further research. The main task is to define the candidate genes, their roles, localization, regulation, and the associated pathways that influence the onset of human labor. Genomic research has identified some candidate genes that involve growth, differentiation, endocrine function, immunity, and other defense functions. For example, selenocysteine-specific elongation factor (EEFSEC) influences synthesis of selenoproteins. WNT4 regulates decidualization, while a heat-shock protein family A (HSP70) member 1 like, HSPAIL, influences expression of glucocorticoid receptor and *WNT4*. Programming of pregnancy duration starts before pregnancy and during placentation. Future goals are to understand the interactive regulation of the pathways in order to define the clocks that influence the risk of prematurity and the duration of pregnancy. Premature birth has a great impact on the duration and the quality of life. Intensification of focused research on causes, prediction and prevention of prematurity is justified.

Pediatric Research (2019) 85:422–431; https://doi.org/10.1038/s41390-018-0180-z

Approximately 15 million preterm live births (PTB) contribute to around 1 million infant deaths annually. According to estimates, each year the consequences of PTBs cause a loss of 80–90 million quality-adjusted life-years (QALY).[1,2] QALY ranges between 1 (good health) and 0 (death). Despite the increased survival of preterm infants, the lifelong adverse consequences of very preterm birth remain high. Preterm birth rates are 5.5–10% in Europe, 10–11% in the USA, and 11–13% globally. Spontaneous preterm labor and delivery (SPTD) accounts for ~70% of all cases of PTB. About 50% of spontaneous preterm live births (SPTB) occur in apparently low-risk pregnancies. A number of acquired factors may increase the risk of SPTB, and African-American ethnicity increases the adjusted risk of SPTB. The most significant risk factor in singleton births is previous preterm delivery, which increases the risk of preterm birth 3–6-fold in subsequent pregnancies.[3,4]

The synthetic nonsteroidal estrogen diethylstilbestrol (DES) was administered to prevent miscarriages and premature births during 1943–1972 but had no beneficial effects. However, it increased the risk of squamous genital neoplasia in female offspring as well as infertility and hypospadias in male offspring.[5] Other serious drug side effects raised concerns about the use of medication in early pregnancy.[6] Tocolytic therapies administered later in pregnancy also failed to prevent SPTD/SPTB.[7,8]

Reproduction is exceptionally species-specific. According to epidemiological family studies, the risk of prematurity depends mostly on genetic predisposition of the mother. Genome of the fetus has also influence on the risk of premature birth.[9,10] Both maternal and fetal genomes harbor variants that increase the risk of SPTD and SPTB. In live births, these two phenotypes converge. In present review we link the susceptibility of SPTD to maternal genetic factors and susceptibility of SPTB phenotype to fetal genetic factors. However, maternal and fetal genes likely interactively influence the risk of SPTD/SPTB. As the genetic predisposition in SPTD/SPTB ranges from 25 to 40%,[9,10] the Human Genome Project provides tools that are useful in identifying genes that influence the risk of SPTD/SPTB. This brief review summarizes the new genomic findings about susceptibility to SPTD/SPTB. We envision that these and further genomic and epigenetic discoveries will improve prediction of risk in early pregnancy and eventually provide new tools for decreasing the risk of SPTD/SPTB.

FACTORS ASSOCIATED WITH ONSET OF PRETERM LABOR AND DELIVERY

Prostaglandin E2 (PGE2), PGF2α, and oxytocin mediate the induction of cervical maturity and uterine contractions leading to delivery. Inflammation is involved in preterm, term, and post-term labor. The obstetric data that define the factors that trigger SPTD are presented below in brief.[11]

Uterine distention

The pressure induced by intrauterine mass in relation to uterus size leads to stretching and contraction of smooth muscle. This increases expression of contraction-associated proteins (e.g., prostaglandins, gap junction proteins, and oxytocin) and is associated with a 6–10-fold increase in the prematurity rate of twin pregnancies. The distending pressure and fetal movements generate shear forces and inflammation in fetal membranes,

[1]PEDEGO Research Unit, Medical Research Center Oulu, University of Oulu, and Department of Children and Adolescents, Oulu University Hospital, Oulu, Finland and [2]Division of Human Genetics, Center for Prevention of Preterm Birth, Perinatal Institute, Cincinnati Children's Hospital Medical Center, Department of Pediatrics, University of Cincinnati College of Medicine, March of Dimes Prematurity Research Center Ohio Collaborative, Cincinnati, OH, USA
Correspondence: Mikko Hallman (mikko.hallman@oulu.fi)

Received: 28 May 2018 Revised: 20 August 2018 Accepted: 23 August 2018
Published online: 18 September 2018

which may be transmitted to the uterus and cervix with consequent cervical maturation and labor.[12,13]

Preterm prelabor rupture of fetal membranes (PPROM)
PPROM is an initiating event in 20–30% of all SPTB cases. It is associated with increased risk of ascending infection. Not all PPROM cases lead to SPTD; many cases remain silent until term. PPROM is more common and often recurrent in African Americans.[14]

Infection and activated inflammation
Infection and activated inflammation are associated with 25–40% of SPTD cases. Ascending infection is the most common. Intrauterine (endometriosis), salpinx, and extrauterine infections from various sources are also risk factors for SPTB. Aside from clinical chorioamnionitis (CA), CA is mostly silent (i.e., histological CA). Cell wall components of microbes (lipopolysaccharides and other toxins) bind to specific receptors, inducing synthesis of inflammatory cytokines (e.g., IL-1 and TNF) and PGE2.[3,15]

Loss of immune tolerance
In early pregnancy, endometrial tissue transforms into the decidua. The decidua contains maternal Th2 and regulatory T (Treg) cells, CD56-positive lymphocytes, natural killer cells, antigen presenting cells (APC), fibroblasts, stromal cells, and fetal decidual trophoblasts. Syncytiotrophoblasts and a portion of the cytotrophoblasts are in direct contact with maternal blood cells. To suppress maternal immune activation, trophoblasts express tolerogenic HLA-G molecule. Aberrant activation of the silenced host immune system may induce SPTD.[16,17]

Stress
Reaction to stress may perturb production of the hormones necessary to maintain pregnancy. For example, stress is known to increase placental production of corticotropin-releasing hormone (CRH).[17]

Intrauterine bleeding
Intrauterine bleeding that occurs in the absence of evidence of placenta previa increases the risk of SPTB. The decidua is a common source of bleeding.[18,19]

Placental and fetal growth
Placenta and fetal membranes likely play roles in regulating the duration of pregnancy. Critical early events include implantation and vascularization. Fetuses that undergo SPTB tend to have slow intrauterine growth. However, the causes and consequences of growth delay remain poorly understood with respect to SPTD. Elective preterm birth is indicated for serious intrauterine growth restriction.[20]

Fetal maturity
Immune proteins, including surfactant proteins (SP) A and D, have been suggested to promote labor.[21] In human pregnancy, the appearance of SP-A in amniotic fluid and the onset of labor may not coincide.[22] Antenatal glucocorticoid has no effect on human pregnancy duration, either. Although inflammatory activation induces surfactant maturity and preterm labor, these two events may not be of consequence.[23]

Endocrine system
Human progesterone (P4) synthesis takes place in trophoblasts after early involution of the corpus luteum.[24] P4 is an anti-inflammatory hormone that promotes immune tolerance, and maintains myometrium quiescence. P4 levels increase throughout pregnancy and even during labor. However, toward the end of human pregnancy, the responsiveness of progesterone receptor (PR) decreases as levels of PR-A (the truncated form of PR) increase and levels of PR-B (the active, full-length form of PR) decrease.[25,26] Near the time of term delivery, synthesis of CRH increases in the placenta, boosting the synthesis of corticotrophin. This increases synthesis of cortisol and dehydroepiandrosterone (DHEA), a precursor of estradiol that antagonizes P4.[27–29]

BRIEF COMMENT ON CURRENT PHARMACOLOGICAL APPROACHES TO PREVENT OR DELAY SPTD/SPTB

Current practice recognizes primary, secondary and tertiary interventions in prevention of SPTD.[30] Primary interventions target to whole population before or during pregnancy. Secondary prevention targets women with established risk of preterm birth. Tertiary conditions include early detection of imminent preterm labor and treatments that delay or prevent early birth, or accelerate fetal maturity. The focus has been on tertiary prevention.

Antibiotics and probiotics
Antibiotics may prolong the duration of pregnancy after PPROM.[31] In general, however, antibiotics do not decrease the risk of SPTD/SPTB. Further investigations of the types of microbes present in maternal and intrauterine tissues, as well as prophylactic trials with probiotics, are required to substantiate the potential effects of microbial colonization on the rate of SPTD.[32]

Tocolytic agents
Tocolytics silence uterine contractions and may delay SPTB for several days, allowing time for antenatal steroids to take effect.[33] Nonsteroidal anti-inflammatory agents are effective but may cause serious adverse effects. Oxytocin antagonist and calcium channel blockers are preferred for their safety. According to a meta-analysis, calcium channel blockers delay preterm labor.[33]

Progesterone
According to meta-analysis, in rare pregnancies with short cervix syndrome (cervix <0.25 cm before mid-gestation), administration of vaginal P4 increases the length of pregnancy and decreases prematurity rate and morbidity of the newborn.[34] Recent meta-analysis confirmed the beneficial effect of P4 in short-cervix syndrome,[35] despite in a large recent trial P4 had no beneficial effect on preterm births or postnatal outcomes.[36,37] Uncertainty remains concerning the indications, dosage, route, onset, and duration of administration of P4.

GENETIC PREDISPOSITION TO SPTD/SPTB

Studies of the duration of pregnancy in twin sisters and in families with recurrent SPTB/SPTD demonstrate that both maternal and fetal genetic factors contribute to the variance in gestation length.[38,39] Hereditary factors contribute an estimated 25–40% toward the risk of premature birth. In one study, the maternal and fetal genetic contributions were ~60% and 40%, respectively, toward the variance in pregnancy duration.[9] According to Wilcox[10] pregnancy duration is transmitted as a matrilineal trait. However, there is no consistent evidence to indicate whether either mitochondrial genome[40–42] or parental imprinting[43] influence the SPTD/SPTB rate or duration of pregnancy.

Ethnicity influences the predisposition to preterm birth. Ethnic differences in the risk of SPTB may be due to genetic or environmental factors or a combination of both.[44] In African American populations, the prematurity rate is high and the influence of environmental factors is prominent.[45] Acquired risk factors include very young, and also old age of the mother; nutrition; extremes in body mass index (BMI); heavy exercise or severe stress; and exposure to infection, drugs, smoking, and toxins.[38]

Expressivity is likely low for common alleles associated with complex traits like SPTD that occur at an early age. Rare variants that are dominant and have a high expressivity influencing the pregnancy duration are likely damaging, disease-causing variants

identifiable in families with repeated SPTD/SPTBs. Pathways that influence the duration of pregnancy may contain a large number of variants and multiple SPTD-predisposing candidate genes.

Large population sizes based on power analysis, an accurate definition of phenotypes, and minimization of ethnic and acquired differences between SPTD/SPTB and control populations will influence the success of genomic analysis. Accuracy of diagnostics is important with exclusions of multiple pregnancies, elective preterm deliveries due to maternal/fetal complications, and cases of severe clinical CA.

Populations that underwent a recent bottleneck followed by population growth without significant population admixture are attractive for genetic studies. They may have unique single-gene diseases and lower allelic variation, which decrease the population size requirements for genetic discoveries. Northeastern Finland, which has a typical population history and a unique pattern of single gene diseases, is one such example.[46]

Candidate gene studies

Conventional candidate gene studies have focused on genes presumed to influence SPTD/SPTB-triggering events. These genes influence the inflammatory response or resistance against infection. Another set of candidate genes influence uterine muscle contractions, tensile strength of fetal membranes, or quantity of cervical connective tissue. Still others influence placental function, differentiation, and stress; response to environmental toxins; and activity of hormones necessary to maintain pregnancy. Currently, variants representing several hundred genes have been evaluated and may show significant trends.[47–50] However, these findings have either not been replicated or they have had a nominal association in genomic studies. These types of approaches, however, are still justified especially when addressing the significance of suggestive associations from the genome-wide studies in more detail.

Genome-wide approaches

The whole genome consists of ~3.1 billion base pairs, while the whole exome comprises ~180,000 exons, 35 million base pairs and 22,000 genes (about 1.2% of the whole-genome sequence). Thus far, the most common approach taken by genomic studies to investigate complex diseases has been genome-wide association study (GWAS) analyses. Typically, chips contain 0.2–2 million single nucleotide variants (SNV). Less than 0.2% of all SNVs are located in exons, and these SNVs are enriched in many chips. The expanding approaches to genomic studies include whole exome sequence (WES) and whole-genome sequence (WGS) analyses. In

Table 1. List of major genomic approaches towards understanding the genetics of SPTD and SPTB

	POPULATION-BASED APPROACH	FAMILY-BASED APPROACH
Definition of discovery and replication populations	Phenotype of SPTD and/or SPTB defining phenotypes length of pregnancy another phenotype consider subcategories of SPTD/SPTB Defining controls limits of pregnancy length prospective recruitment preferable Other population characteristics Homogeneous ethnicity preferable Similar ethnicity in cases and controls Mother-child dyads or family triads analyzed: transmission of SNVs Defining replication populations consider several studies Genetic power analysis increase the size of control population Ethics permissions required	Families with densely affected SPTD/SPTB phenotypes Several generations, if possible mode of inheritance; maternal vs paternal transmission Population homogeneity critical (replication of rare SNVs) Consider subcategories of SPTD/SPTB (families with PPROM-, bleeding-, infection-associated SPTD, or with similar duration of pregnancy) Consider acquired risk factors Replication populations Family-based replications Families with different ethnic background may have similar candidate genes despite different SNVs Utilization of large GWAS data for replication
Genetic analysis	Availability of genetic expertise in genetics and bioinformatics DNA sample preparation, genotyping, criteria of SNV selection Quality control and statistical imputation Statistical analyses of SNVs (Manhattan blot, etc) Study of population stratification (heterogeneity) Analysis of phenotype subgroups, gene environment interaction	Linkage analysis: genome-wide haplotype markers followed by sequencing. No longer used. Whole exome sequencing (WES) currently utilized. Whole genome sequencing (WGS) additionally evaluates regulatory regions of genes and large intergenic regions. Variant annotation, filtering, prioritizing Quality control tests Prior to association analyses, selection criteria may include data quality, SNV frequency (mostly < 0.01) and predicted pathogenicity
Interpretation of results	Gene–gene interaction Inheritance pattern/transmission of associating SNVs Similar approaches in population- and in family-studies Bioinformatics Functional annotation and potential pathogenicity of candidate genes and nearby SNVs Comparison of new SNV data with available genetic data on SPTD/SPTB and on other traits Pathway-associations of gene discoveries Correlation of SNV with mRNA levels of nearby genes, with putative regulatory regions Identification of SNV that is literature-informed quantitative trait locus (QTL); i.e. it has established tissue expression, correlating with a quantitative trait	Functional analyses of gene discoveries Analysis of biologic material recovered from pregnancies (placenta, specimens from cervix and uterus); primary cell cultures or continuous cell lines; animal models. Omics studies (DNA methylomics, genomics of microRNA, proteomics, transcriptomics, etc) Transgenic experiments in vitro (gene silencing or upregulation, followed by transcriptomics/ proteomics/ pathway analyses, etc). Transgenic animals; disease models Many lucrative possibilities with advancing knowledge, collaboration, technology and resources

multifactorial phenotypes, like SPTB, disease associating genetic alterations are often outside the coding areas and thus WGS approaches have increasing importance. Table 1 shows some practical aspects of genomic studies on SPTD/SPTB.

The linkage disequilibrium of individual variants within haplo-blocks enables fairly accurate estimation of a very large number of SNVs. This imputation method yields an accurate estimate of 2–15 million SNVs. Variants that remain unaccounted for are mostly rare SNVs (frequency < 0.01), particularly those located in intergenic regions. Copy number variations (CMV), insertion or deletion (InDel) variants and mitochondrial DNA (mtDNA) are often included in genome analysis. Other studies have a focus on specific metabolic pathways or pharmacogenomics. Thus far, genomic studies on SPTD/SPTB have been limited.

Genomic discovery requires replication. Further molecular and translational studies are essential to understand the specific function of a gene, its variants, and the functional pathways that contribute to its expression. Bioinformatics tools have become increasingly accurate as the data on molecular cellular processes in various species, tissues, cells, and diseases continue to accumulate. These data are helpful in planning and interpretation of experiments that address the molecular aspects of specific genes and their influences on individual targets[51] (Fig. 1).

RNA molecules include messenger RNA (mRNA), ribosomal RNA (rRNA), transfer RNA (tRNA), signal recognition particle RNA (SRP RNA), and others. Several epigenetic systems participate in gene expression. In addition to DNA methylation (known as the methylome; i.e., the pattern of genome-wide DNA methylation), a number of proteins (particularly histones) and regulatory RNA species, e.g., microRNA (miRNA), long noncoding RNA (lncRNA), and antisense RNA (asRNA) are involved in regulation of DNA replication as well as transcription or translation. Indeed, binding of transcription factors to their target sites as well as interactions among transcription factors play critical roles in gene expression. The proteome consists of ~170,000 individual proteins with multiple splice variants, intracellular processing, distribution, structure, function, and pathway involvement. Post-transcriptional and post-translational genomic entities include the proteome, lipidome, and metabolome, to mention a few. However, these areas are outside of the scope of the current review.

The current studies on association of microbiome and SPTD/SPTB have been excluded from the present review.[52–54]

Genome-wide linkage analysis

In the past, linkage analysis was the principal tool used to define single-gene diseases. In the first genomic analysis of SPTB/SPTD, seven large families from Northern Finland were analyzed using autosomal markers for either SPTB or SPTD outcomes.[55] The study found that the gene encoding type 1 insulin-like growth factor receptor (*IGFR1*) was associated with SPTB. The study also identified joint haplotypes shared by several SPTB-affected families, and these results were replicated in a case–control study (Table 2). IGF1R influences growth, differentiation, inflammation, immunity, and susceptibility to infections. *IGF1* may influence susceptibility to preterm birth.[56] IGFbp1 binds IGF1 and IGF2, thereby altering their half-life and function. Cervical-phosphorylated IGFbp1 has value in predicting SPTD.[57,58]

X-chromosome linkage analysis revealed an association between SPTB and an exonic variant of androgen receptor (*AR*). This SPTB-associating variant comprises long CAG repeats on the exon that decrease AR activity and are associated with susceptibility to SPTB.[59] There is evidence that androgen activity throughout pregnancy and androgen withdrawal in late pregnancy are of importance.[60,61] An intron variant of *CXCR3* in X-chromosome was also associated with SPTB.[62] This chemokine receptor is abundant in trophoblasts. In cord blood in SPTB cases, the SPTB-predisposing *CXCR3*-SNV on rs220964 was highly expressed, and this increased expression is associated with high cord blood levels of CXCL9, a CXCR3 ligand. *Cxcr3* deletion modified the maternal LPS-induced cytokine response in the fetal compartments[62] (Table 2).

Whole-exome sequencing

WES analyses involve families with high prevalence of SPTB and SPTD. The analysis provides a very high number of variants in comparison to mostly a limited study population. According to feasible assumption, many WES studies focus on rare (frequency < 0.01), possibly damaging SNVs (Table 1).

A WES analysis followed by pathway analysis of ten mothers, including two mother–daughter pairs from families highly affected by SPTD/SPTB, revealed the complement and coagulation cascade among the top pathways. This feasible finding was further tested in case–control setting of 565 Finnish mothers and involving 67 promising coding region SNVs of *CFH*, *CR1*, *F13B*, *F5*, *CR2* and *C4BPA*. A missense variant of *CR1* was most significantly associated with SPTD[63] (Table 2).

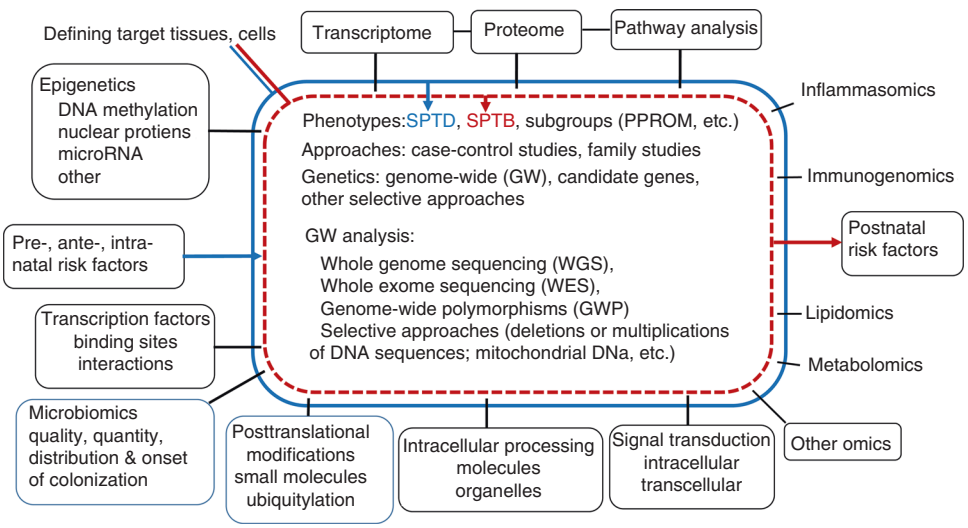

Fig. 1 Schematic presentation of projects defining the function and roles of the candidate genes. Significant genome-wide gene discovery starts a series of molecular and translational studies

Table 2. Genomic or selective genomic candidate gene studies on the risk of SPTD and SPTB. When indicated, association to duration of pregnancy is reported[a]

Article	Discovery population	Diagnoses	Replication population	Genome analyses	Result
Haataja et al.[55]	89 Finnish individuals from 7 unrelated families, with recurrent prematurity 21 SPTD, 41 SPTB	<37 wks, term. Replicates: <36 wks, term	348 SPTD, 143 term 334 SPTB, 197 term	Autosomal linkage analysis Affymetrix NSP array, 6377 autosomal SNPs for linkage	SPTB. large common haploblock within *IGF1R* rs2684811, $p = 2.0 \times 10^{-5}$; rs4966036, $p = 2.1 \times 10^{-4}$. Replication: $p = 0.037$
Plunkett et al.[70]	165 SPTD, 163 controls Finnish and African-American	<36 w vs. term. Replicates: <37 wks, term	79 SPTD 171 control African-Americans 73 SPTD 292 control Hispanic	Selective QWAS: 150 selected genes with accelerated evolution Affymetrix HSNP 6.0 8490	SPTD. *FSHR* rs6741370; $p = 8.2 \times 10^{-5}$. Replication rs11686474 $p = 0.004$ in African-Americans
Karjalainen et al.[59]	Same as Haataja 2011[55]	<37 wks, term.	272 < 36 wks SPTD, 201 term 269 < 36 wks SPTB, 199 term	X-chromosome linkage analysis Affymetrix. 4870 chrx SNPs	SPTB. Androgen receptor (*AR*) exon-1 CAG repeat: HLOD 3,72 Replication in Finnish cohort: AR; $p = 0.0007$.
McElroy et al.[63]	10 Caucasian or Finnish mothers with multiple SPTD, including two SPTD/SPTB pairs.	<37 wks, term: both in WES and replication studies	237 SPTD, 328 term control	WES analysis followed by pathway analysis.	SPTD. WES and pathway analysis: complement and coagulation cascade/ genes among the top. Replication: association with missense, rs 6691117 in *CR1*.
Karjalainen et al.[62]	Same as Haataja 2011[55]	<37 wks, term.	251 < 36 wks SPTD, 192 term 68 < 36 wks SPTB, 192 term	Affymetrix. 4870 chrx SNPs	SPTB. *CXCR3:* a. intron rs2280964; $p = 0.009$; replication in Finnish cohort with repeat SPTB. b. *CXCR3* mRNA in basal placenta; $p = 0.01$. c. Rs2280964 associates with CXCR3-ligand CXCL9 in cord blood; $p < 0.05$. *Cxcr3*-deficient mice: altered maternal–fetal cytokine trafficking.
Zhang et al.[67]	959 SPTD vs. 985 controls Multiethnic. Cases and controls matched for ethnicity, maternal age and parity	20–33.9 wks, 39.0–41.9 wks	293 SPTD, 200 control population. 243 SPTD, 149 control population.	Affymetrix HSNP Array 6.0 Validation: 96 best SNPs	SPTD. rs17053026 in Chr3; $p = 10^{-6}$. SPTB: rs17527054 in Chr 6p22; $p = 2.7 \times 10^{-12}$. rs3777722 Chr 6p22; $p = 1.4 \times 10^{-10}$. Associations not replicated.
Bacelis et al.[68]	1921 mothers, 1199 children. 3-model evaluations: SPTD/SPTB; PPROM-SPTD/SPTB; no-PPROM-SPTD/SPTB	<37 wks, term Norwegian MoBa cohort linked to birth register	None	GWAS analysis involving 525 577 SNPs Three outcomes analysis of fetal and maternal DNA in each case	SPTD. Top association with PPROM mothers: rs6977716 in *DPP6*; $p = 5.1 \times 10^{-7}$. SPTD/SPTB. Enrichment ($p = 0.005$–0.001) in infection/inflammation-related genes ($n = 13$)
Modi et al.[65]	PPROM-SPTB: 49; 20 controls African-American	<37 wks; vs. Term	African-American PPROM-SPTB 188; 175 term controls	WES analysis of rare alleles on fibrillary collagens Case-control replication: 9 SNPs Involvement in extracellular matrix (ECM)	SPTB. Rare variants in *FKBP10, COL1A2, COL2A1, COL5A1* associated with PPROM-SPTB Case–control replication suggest genetic burden of rare damaging variants
Modi et al.[66]	PPROM-SPTB: 76; 42 controls African-American	<37 wks; vs. Term	188 African-American PPROM-SPTB; 175 term	WES discovery of rare alleles Case-control replication: 14 SNPs	SPTB. Rare alleles of *CARD6, DEFB1, FUT2, MBL2, NLRP10, NOD2*. associated with PPROM-SPTB. Case–control replication
Rappoport et al.[69]	1349 SPTB and 12595 controls Controls and cases were ancestry-matched	<37 week cases; control population: African, the Americas, European, South- or East-Asian	Sage II cohort, GALA II cohort and ITMI cohort and known, ancestry specific cohorts that include both SPTB and term populations.	GWAS HumanOmni2.5–4 v1 BeadChip (Illumina). Altogether 2 015 750 SNPs.	SPTB. Intergenic rs217591250 on Chr1 in African ancestry; $p = 4.55 \times 10^{-9}$. Intergenic rs1979081 on Chr8 in African ancestry; $p = 3.72 \times 10^{-8}$. Associations not replicated.
Zhang et al.[43]		Self-reported gestation of 1st delivery. <37 wks; vs. 37	SPTD 2565, term 6034. SPTB 1918, term 2172	GWAS Four genotypic platforms, range	SPTD. *EBF1, EEFSEC, AGTR2* Duration of gestation: EBF1, *EEFSEC AGTR2,*

Table 2 continued

Article	Discovery population	Diagnoses	Replication population	Genome analyses	Result
	SPTD: 3 331, Ctr: 40 236 controls Caucasian	+ 0–42 + 0 weeks. Post-term >42 weeks	SPTD, SPTB cases and controls of Norwegian, Danish and Finnish origin	560 00–950 000 15 635 593 imputed SNPs 3184 mother-infant pairs for transmittance of significant maternal associations.	*WNT4* All above replicated *ADCY5, RAP2C* borderline; became significant in replication and combination sets. Mother infant dyad analyses confirm maternal genome.
Huusko et al.[89]	SPTD: 17 Finnish mothers from seven complex families	<37 wks	93 Danish sister pairs and two sister triads, all with history of preterm delivery	WES analysis of rare alleles (fr < 0.01) Pathway analysis of 406 candidate genes.	Glucocorticoid Receptor (*GR*) pathway enrichment, $p < 1.7e{-}8$). Replication. Among its affected genes, *HSPA1* rs34620296(Ala268Thr) replicates in GWAS ($p < 0.002$). In silico analysis and decidualization experiment: Ala268Thr generates phosphorylation site; Ala268Thr mutant unable to activate *WNT4* (cf. ref. [39]).

[a]The listed studies do not include focused pathway analyses,[56, 64] primary proteomics analysis,[71] Mendelian randomization analysis,[90] or microbiome analyses[52–54]

A genome-wide targeted screening identified *IGF1* as candidate of preterm birth.[56] Subsequent study involved sequencing of haplotype blocks and selected exomes with the aim to define pathways and candidate genes. Comparison of preterm birth-affected families and final analysis identified *IGF1*, *ATM*, and *IQGAP2* as the most promising and connected.[64]

In African Americans prone to PPROM-SPTB, selective WES analysis identified potentially damaging missense variants in genes that encode fibrillar collagens and proteins involved in collagen synthesis. The SMV data were further studied in a SPTB-PPROM population and controls born at term.[65] Another WES analysis identified genes involved in innate immunity. A targeted case–control replication study confirmed associations between SPTB and several rare alleles. Of the likely damaging rare alleles in the innate immunity and host defense genes, *CARD6, DEFB1, FUT2, MBL2, NLRP10*, and *NOD2* were only detected in the PPROM-SPTB cases[66] (Table 2).

GWAS studies

In GWAS analyses, because there are about 10^6 binomial comparisons to be made, the p value must be $<5 \times 10^{-8}$ to reach statistical significance. Since a consistent large difference in the risk ratio of the common allele distribution between cases and controls of identical ethnicity is practically unattainable, a large population of cases and controls is required. A lack of homogeneity, particularly in multiethnic populations, is a limitation that causes population stratification, which complicates analyses. Independent replications are required to confirm the results. Large and accurate GWASs and meta-analyses will likely identify the most robust candidate genes for future research. Rare SNVs identified among families with recurrent phenotype by WES/GWS studies may nominally replicate in large GWAS studies (Table 1).

The first GWAS of SPTB included 2072 pregnant singleton mothers and their newborns in the USA. The racial distribution was mixed in this study, with 59% of Caucasian ethnicity, and the ethnic distribution was similar between cases and controls. The team validated the 96 best SNVs in 493 mothers and their children: variants associated with *CCDC25* and *DCP1A* yielded the lowest p values (1×10^{-6}). However, these results were not replicated. The authors stated that the heterogeneity of the SPTB/SPTD phenotype may be a serious limitation[67] (Table 2).

Another GWAS included 1921 mothers and 1199 children from a Norwegian birth register cohort.[68] The outcomes included labor-initiated and PPROM-initiated SPTD and SPTB. These phenotypes were analyzed separately and together. The top candidate gene, *DPP6*, was associated with PPROM-SPTD ($p = 5.1 \times 10^{-7}$). Several replicated loci from the maternal GWAS and a literature-informed top candidate gene set showed a nominal association with the phenotypes.

A recent GWAS study of SPTB included 1349 SPTB cases and 12,595 ancestry-matched controls classified into population cohorts by ethnicity (Table 2).[69] Two intergenic loci, each in one specific ancestry, were associated with SPTB at a genome-wide level of significance. The investigators were unable to identify a replication cohort for these findings. Despite the negative result of their large study, the authors stressed the possibility that ancestry-based environmental selective pressure may contribute to genetic susceptibility to SPTD.

Advanced genomic studies

Midway between genomic and candidate gene studies is the selective genomic approach. Plunkett[70] conducted a representative study of this type based on the hypothesis that human gestation time has been shortened relative to other primates. Accordingly, many genes involved in parturition would display accelerated evolution in human lineages. This study tested the association between SPTD and 150 genes and around 8400 SNVs

with rapid phylogeny. Of these, *FSHR* contained eight of the ten SNVs associated with SPTD in Caucasian (including Finnish) populations. An African American population also had a haplotype of the *FSHR* gene that nominally associated with preterm delivery.[70] FSH/FSHR influences contractility of the uterus and expression of progesterone receptor (*PR*).

In another approach genetic analysis is combined with placental proteome analysis. A recent analysis investigated SNVs ($n = 77$) in ten fetal genes encoding placental proteins associated with the duration of pregnancy. Of these, only a SNV within *CPPED1* associated with pregnancy duration.[71] Both CPPED1 protein and *CPPED1* mRNA in the basal plate of the placenta were associated with labor and length of pregnancy. CPPED1 is a phosphatase that inactivates AKT1 in the PI3K/AKT1 signaling pathway in bladder cancer cells.[72] A pathway analysis in which siRNA was used to inhibit CPPED1 expression in trophoblast-derived cells revealed that CPPED1 affects expression of genes involved in angiogenesis and proximal PI3K/AKT1 signaling pathways.[71] Potential phosphorylation targets of AKT1 include inactivation of hypo phosphorylated FOXO1, which participates in the progesterone receptor B (PR-B) transcriptional complex.[73]

Thus far, only one GWAS study has identified a significant association with the risk of SPTD. This study also found significant associations with duration of pregnancy.[43] This landmark study described the GWAS findings from a total of 43,568 mothers of European ancestry. The discovery population represents self-reported data regarding the first pregnancy, including length of gestation, and apparently excludes indicated preterm births. The replication studies included three Nordic data sets with a total of 8643 mothers. Overall, 14 genomic loci achieved significant ($p < 5 \times 10^{-8}$, four genes) or suggestive ($p < 1 \times 10^{-6}$, ten genes) associations (Table 1). *EBF1*, *EEFSEC*, *AGTR2*, and *WNT4* were significantly associated with at least one of the outcomes in the discovery stage. Remarkably, all of these findings were replicated in the Nordic data sets. In addition, *ADCY5* and *RAP2C*, which had suggestive significance in the discovery stage, achieved genome-wide significance in the joint analysis with the replicates. All associations were of maternal origin, as suggested by the analysis of mother–infant dyads.

The top SNVs in both the discovery and replication stages for SPTD and for gestational length were associated with early B-cell factor 1 (*EBF1*), which is essential for normal B cell development. Previous GWAS studies have identified associations of *EBF1* with cardiovascular, metabolic, and endocrine risk, as well as with low birth weight.[74] Likewise, SPTD-associating variants near angiotensin II receptor, type 2 (*AGTR2*) on the X chromosome may modulate uteroplacental circulation and contribute to the risk of preeclampsia.[75] The renin–angiotensin system and oxidative stress are potential mechanisms for both preeclampsia and SPTD. Identification of an association between SPTD and *AGTR2* is consistent with the possibility that insufficient oxygen and nutrient uptake by the placenta are not exclusively associated with fetal growth restriction and elective preterm delivery. The combination of placental perfusion defect and CA is associated with SPTD and poor neurological outcome;[76] a modest decrease in fetal growth is also associated with SPTB.[77,78]

Zhang et al. also found that variants of eukaryotic elongation factor selenocysteine-tRNA-specific (*EEFSEC*) are associated with SPTD and length of pregnancy.[43] The SPTD-associated SNV rs9998765 is in linkage disequilibrium with hypospadias- and menarche age-associated SNVs. EEFSEC participates in the incorporation of selenocysteine (Sec) into Sec-proteins. Phosphorylated selenium serves as a substrate when phosphoserine-tRNA is converted to Sec-tRNA (Fig. 2a–c). Transfer of Sec-tRNA to Sec-protein requires both a cis-acting Sec insertion sequence (*SECIS*) element in the 3′ untranslated region (3′ UTR) of each Sec-protein and several specific trans-acting factors. These may include EEFSEC, SECISBP2, L30, nucleolin, and EIF4A3 (Fig. 2d).

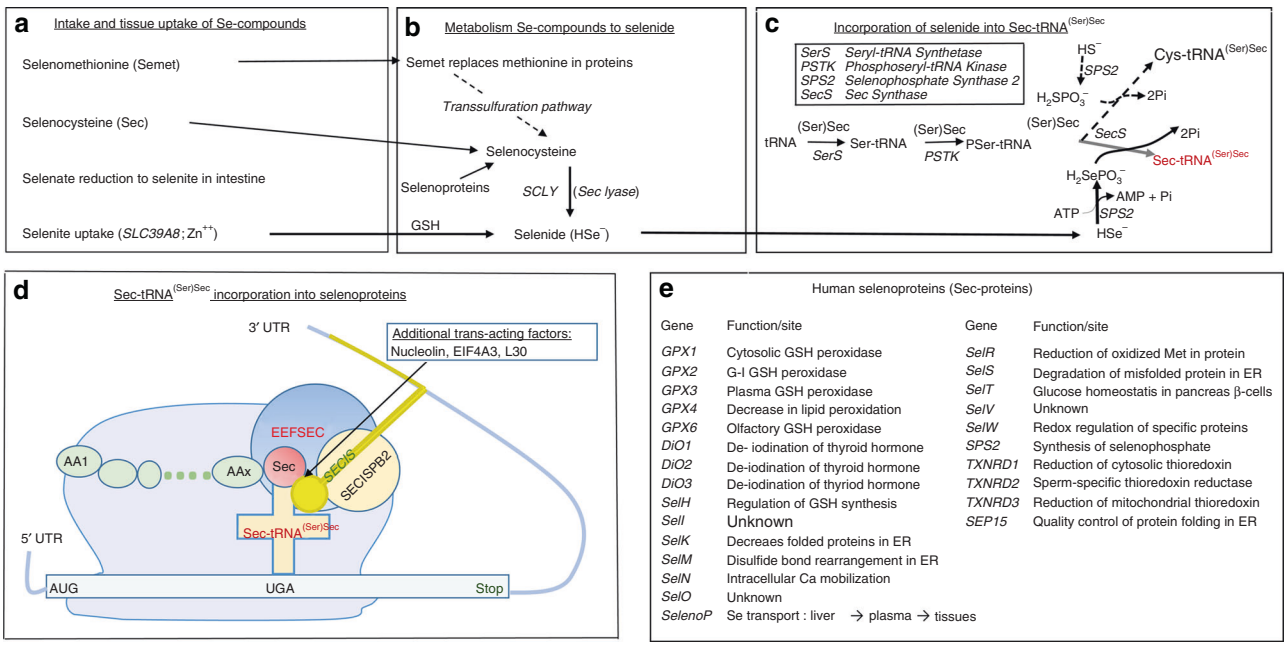

Fig. 2 Intake, metabolism, and incorporation of selenium into selenoproteins. **a** List of selenium-containing compounds in the diet. **b** Metabolism of Se-compounds. Selenomethionine may replace methionine in proteins. According to what is currently known, it does not alter protein function and therefore is likely part of selenium storage. **c** Incorporation of selenide into selenocysteine (Sec) tRNA$^{(Sec)Ser}$. The pathway consists of transformations catalyzed by the enzymes shown in the figure. SPS2 is a Sec-protein. **d** Synthesis of Sec-proteins. Sec-tRNA$^{(Sec)Ser}$ uniquely has a UGA in the open reading frame. In the 3′ UTR of the genes encoding Sec-proteins, the stem-loop structure known as Sec insertion sequence (*SECIS*) designates the UGA triplet that otherwise is a termination codon. Gene-specific differences in the *SECIS* elements affect how well trans-acting factors influence translation efficiency. EEFSEC and SECISPB2 are major trans-acting factors; additional factors may include nucleolin, EIF4A3, and L30. **e** List of 25 Sec-proteins identified in the human genome and transcriptome and brief statement about the function. Alone among Sec-proteins, SelenoP contains more than one Sec amino acid

The human genome contains 25 Sec-proteins, including glutathione peroxidases (GPX) ($n = 5$), thioredoxin reductases ($n = 3$), thyroid hormone deiodinases ($n = 3$), proteins involved in endoplasmic reticulum functions ($n = 3$), and others (Fig. 2e). SelenoP contains 10 Sec amino acids and it transports selenium to tissues.[79] Apolipoprotein E receptor-2 (apoER2) facilitates uptake of SEPP1 in the nervous system, testis, and placenta. In addition, protein-bound methionine binds selenium to selenomethionine (Semet).

Selenium protects against acute pro-oxidant injury. An in vitro study revealed that knockdown of *SECISBP2* reduces levels of selenoproteins *GPX1*, *SelK*, and *DiO2* and increases oxidative stress in trophoblast cells. When *SECISBP2* expression is inhibited by siRNA, the synthesis of human chorion gonadotropin (hGC) and progesterone decreased, and the functions of trophoblasts deteriorated.[80] The recommended daily intake of selenium ranges from 55 to 90 µg. Bioavailability of selenium may depend on the quality of Se-nutrient and on intracellular uptake of SelenoP[79] or selenite[81] (Fig. 2a). Both severe Se-excess and Se-deficiency cause serious symptoms.[82,83] Inborn errors of Sec-proteins cause protein-specific traits.[84] Male genitalia, thyroid, brain, and placenta have high Se-content and their Se-contents are resistant to acute Se-deficiency.[79] Serum selenium concentration in early pregnancy does not consistently associate with duration of pregnancy.[85] In Malawi, which has a high prematurity rate, the selenium content of the soil is low.[86] In Finland, selenium supplementation of the fertilizers started in 1984, because of the low Se-content in soil. In Northern Finland, the singleton SPTB-rate was ~50% lower in 1987 than in 1966.[87] Multiple factors could influence on such a large difference, starting from the difference in the calculation of the pregnancy length.

Zhang et al. showed that *WNT4* is associated with pregnancy duration, and they discussed one feasible mechanism behind this phenotype.[43] The *WNT* gene family consists of structurally related genes that encode secreted signaling proteins. These proteins have been implicated in several developmental processes. During placentation, WNT4 is a downstream target of bone morphogenetic protein 2 (BMP2), a member of the TGFβ superfamily; activation of WNT4 by BMP2 induces endometrial stroma cell proliferation and differentiation into decidua stroma cells.[88] Zhang et al. showed that *WNT4* variants influence the binding affinity of estrogen receptor-1 (ESR1) during decidualization, which includes differentiation of endothelial fibroblasts into secretory decidua cells. Estrogen plays a major role in vascularization during formation of the decidua. Analysis of expression of human endothelial stromal cells revealed a striking difference in the *WNT4* allele-dependent binding affinity of electrophoretic mobility shift assay(EMSA)-validated ESR1 to WNT4. These findings imply that differentiation of the endometrium into the decidua in early pregnancy sets the stage for placentation and may influence the duration of pregnancy.

The most recent study concerns the WES analysis of families with predisposition to SPTD.[89] The aim was to identify rare frequency (<0.01), possibly damaging nucleotide variants in complex families with recurrent SPTD. WES analyses included 17 mothers of the northern Finnish origin. The replication population included 93 Danish sister pairs and two triads, all with history of preterm delivery (Table 1). The glucocorticoid receptor (*GR*) signaling pathway was significantly affected by these rare variants ($p < 1.7e^{-8}$). This pathway was replicated in the Danish sister pairs. A gene in this pathway, heat shock protein family A (*Hsp70*) member 1 like (*HSPA1L*), contained two likely damaging missense alleles, identified in four different Finnish families. One of them (rs34620296) was also more frequent in cases than in controls (0.0025 vs. 0.0010, $p = 0.002$) of the large GWAS study.[43] In silico analysis predicted an additional phosphorylation site generated by the rs34620296 variant. Finally, decidualization experiment revealed that *HSPA1L* Ala268Thr (rs34620296) variants influenced the protein level of GR and expression of *WNT4*; the latter was previously shown to associate with SPTD.[43]

COMMENT

Currently, in about 50% of preterm births, there are no identified risk factors. Mendelian randomization analysis based on genomic data suggest that the association between maternal height and length of pregnancy is causal, which indicates that anthropometry influences the duration of pregnancy.[90] Human evolution, characterized by prone position, a narrow birth canal, and a large head, may have favored a decrease in the duration of pregnancy that may even associate with SPTD, leading to fetal demise, which was less harmful than a too large fetus that additionally threatened the mother.[70] Animal models are useful for investigating the induction of SPTD but understanding the variance in the duration of pregnancy requires species-specific approaches.

Genomic candidate genes for SPTD/SPTB identified thus far indicate that development, growth and immunity are important for pregnancy duration, particularly formation of the decidua and vascularization of the placenta. The cervix and uterine corpus respond to signals from both the mother and fetus. Further genomic studies are required. The generally low-expressivity alleles have a higher impact on the prematurity rate than the rare high-penetrance alleles associated with recurrent SPTD. However, the latter likely reveal SPTD-predisposing pathways that contain additional disease genes and variants.

Investigation of the candidate genes, pathways, gene–environment interactions and the underlying epigenetic regulation will further explain why some individuals are more sensitive to SPTD-triggering events than the other. Prevention of SPTD/SPTB requires multiple approaches and individualized prevention strategies.

CONCLUSION

Current research on prevention of SPTB has been concentrating on prevention of the imminent preterm birth. Shifting the target towards early detection of the risk, may be a more promising approach that eventually can lead to effective prophylaxis or early treatment of pregnancies with established risk of SPTD. Identifying mothers that are at the highest genetic risk for SPTD may allow individualized strategies for prevention.

Optimal duration of pregnancy that involves appropriate early development, growth, and differentiation, is among the most critical events influencing both the duration and the quality of postnatal life and therefore require considerable intensification of research.

ACKNOWLEDGEMENTS

Supported by the Jane and Aatos Erkko Foundation (M.H., M.R.), and the Sigrid Jusélius Foundation (M.H.).

ADDITIONAL INFORMATION

Competing interests: The authors declare no competing interests.

REFERENCES

1. Blencowe, H. et al. Preterm birth-associated neurodevelopmental impairment estimates at regional and global levels for 2010. *Pediatr. Res.* **74**(Suppl 1), 17–34 (2013).
2. Yoshida, S. et al. Setting research priorities to improve global newborn health and prevent stillbirths by 2025. *J. Glob. Health* **6**, 010508 (2016).

3. Goldenberg, R. L., Culhane, J. F., Iams, J. D. & Romero, R. Epidemiology and causes of preterm birth. *Lancet* **371**, 75–84 (2008).
4. Muglia, L. J. & Katz, M. The enigma of spontaneous preterm birth. *N. Engl. J. Med.* **362**, 529–535 (2010).
5. Al Jishi, T. & Sergi, C. Current perspective of diethylstilbestrol (DES) exposure in mothers and offspring. *Reprod. Toxicol.* **71**, 71–77 (2017).
6. Troisi, R., Hatch, E. E. & Titus, L. The diethylstilbestrol legacy: a powerful case against intervention in uncomplicated pregnancy. *Pediatrics* **138**(Suppl 1), S42–S44 (2016).
7. Stewart, L. A. et al. Evaluating progestogens for prevention of preterm birth international collaborative (EPPPIC) individual participant data (IPD) meta-analysis: protocol. *Syst. Rev.* **6**, 235 (2017).
8. Smid, M. C., Stringer, E. M. & Stringer, J. S. A worldwide epidemic: the problem and challenges of preterm birth in low- and middle-income countries. *Am. J. Perinatol.* **33**, 276–289 (2016).
9. Lunde, A., Melve, K. K., Gjessing, H. K., Skjaerven, R. & Irgens, L. M. 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**, 734–774 (2007).
10. Wilcox, A. J., Skjaerven, R. & Lie, R. T. Familial patterns of preterm delivery: maternal and fetal contributions. *Am. J. Epidemiol.* **167**, 474–479 (2008).
11. Romero, R., Dey, S. K. & Fisher, S. J. Preterm labor: one syndrome, many causes. *Science* **345**, 760–765 (2014).
12. Adams Waldorf, K. M. et al. Uterine overdistention induces preterm labor mediated by inflammation: observations in pregnant women and nonhuman primates. *Am. J. Obstet. Gynecol.* **213**, 830.e1–830.e19 (2015).
13. Renthal, N. E., Williams, K. C. & Mendelson, C. R. MicroRNAs--mediators of myometrial contractility during pregnancy and labour. *Nat. Rev. Endocrinol.* **9**, 391–401 (2013).
14. Hadley, C. B. 1, Main, D. M. & Gabbe, S. G. Risk factors for preterm premature rupture of the fetal membranes. *Am. J. Perinatol.* **7**, 374–379 (1990).
15. Kim, C. J. et al. Chronic inflammation of the placenta: definition, classification, pathogenesis, and clinical significance. *Am. J. Obstet. Gynecol.* **213**, S53–S69 (2015).
16. Bonney, E. A. Alternative theories: Pregnancy and immune tolerance. *J. Reprod. Immunol.* **123**, 65–71 (2017).
17. Hillhouse, E. W. & Grammatopoulos, D. K. Role of stress peptides during human pregnancy and labour. *Reproduction* **124**, 323–329 (2002).
18. Velez Edwards, D. R., Baird, D. D., Hasan, R., Savitz, D. A. & Hartmann, K. E. First-trimester bleeding characteristics associate with increased risk of preterm birth: data from a prospective pregnancy cohort. *Hum. Reprod.* **27**, 54–60 (2012).
19. Schatz, F., Guzeloglu-Kayisli, O., Arlier, S., Kayisli, U. A. & Lockwood, C. J. The role of decidual cells in uterine hemostasis, menstruation, inflammation, adverse pregnancy outcomes and abnormal uterine bleeding. *Hum. Reprod. Update* **22**, 497–515 (2016).
20. Morgan, T. K. Role of the placenta in preterm birth: a review. *Am. J. Perinatol.* **33**, 258–266 (2016).
21. Mendelson, C. R., Montalbano, A. P. & Gao, L. Fetal-to-maternal signaling in the timing of birth. *J. Steroid Biochem. Mol. Biol.* **170**, 19–27 (2017).
22. Hallman, M., Arjomaa, P., Mizumoto, M. & Akino, T. Surfactant proteins in the diagnosis of fetal lung maturity. I. Predictive accuracy of the 35 kD protein, the lecithin/sphingomyelin ratio, and phosphatidylglycerol. *Am. J. Obstet. Gynecol.* **158**, 531–535 (1988).
23. Hallman, M. The surfactant system protects both fetus and newborn. *Neonatology* **103**, 320–326 (2013).
24. Byrns, M. C. Regulation of progesterone signaling during pregnancy: implications for the use of progestins for the prevention of preterm birth. *J. Steroid Biochem. Mol. Biol.* **139**, 173–181 (2014).
25. Mesiano, S. et al. Progesterone withdrawal and estrogen activation in human parturition are coordinated by progesterone receptor A expression in the myometrium. *J. Clin. Endocrinol. Metab.* **87**, 2924–2930 (2002).
26. Patel, B. et al. Role of nuclear progesterone receptor isoforms in uterine pathophysiology. *Hum. Reprod. Update* **21**, 155–173 (2015).
27. Petraglia, F., Imperatore, A. & Challis, J. R. Neuroendocrine mechanisms in pregnancy and parturition. *Endocr. Rev.* **31**, 783–816 (2010).
28. Renthal N. E., et al. Molecular regulation of parturition: a myometrial perspective. *Cold Spring Harb. Perspect. Med.* **5**, pii: a023069 (2015).
29. Menon, R., Bonney, E. A., Condon, J., Mesiano, S. & Taylor, R. N. Novel concepts on pregnancy clocks and alarms: redundancy and synergy in human parturition. *Hum. Reprod. Update* **22**, 535–560 (2016).
30. Iams, J. D., Romero, R., Culhane, J. F. & Goldenberg, R. L. Primary, secondary, and tertiary interventions to reduce the morbidity and mortality of preterm birth. *Lancet* **371**, 164–175 (2008).
31. Kenyon, S., Boulvain, M. & Neilson, J. P. Antibiotics for preterm rupture of membranes. *Cochrane Database Syst. Rev.* **12**, CD001058 (2013).
32. Ruiz, L., Moles, L., Gueimonde, M. & Rodriguez, J. M. Perinatal microbiomes' influence on preterm birth and preterms' health: influencing factors and modulation strategies. *J. Pediatr. Gastroenterol. Nutr.* **63**, e193–e203 (2016).
33. Haas, D. M., Caldwell, D. M., Kirkpatrick, P., McIntosh, J. J. & Welton, N. J. Tocolytic therapy for preterm delivery: systematic review and network meta-analysis. *Brit. Med. J.* **345**, e6226 (2012).
34. Romero, R. et al. Vaginal progesterone decreases preterm birth and neonatal morbidity and mortality in women with a twin gestation and a short cervix: an updated meta-analysis of individual patient data. *Ultrasound Obstet. Gynecol.* **49**, 303–314 (2017).
35. Romero, R. et al. Vaginal progesterone for preventing preterm birth and adverse perinatal outcomes in singleton gestations with a short cervix: meta-analysis of individual patient data. *Am. J. Obstet. Gynecol.* **218**, 161–180 (2018).
36. Norman, J. E. et al. Vaginal progesterone prophylaxis for preterm birth (the OPPTIMUM study): a multicentre, randomised, double-blind trial. *Lancet* **387**, 2106–2116 (2016).
37. Norman, J. E. et al. Does progesterone prophylaxis to prevent preterm labour improve outcome? A randomized double-blind placebo-controlled trial (OPPTIMUM). *Health Technol. Assess.* **22**, 1–304 (2018).
38. Bezold, K. Y., Karjalainen, M. K., Hallman, M., Teramo, K. & Muglia, L. J. The genomics of preterm birth: from animal models to human studies. *Genome Med.* **5**, 34 (2013).
39. York, T. P., Strauss, J. F. 3rd, Neale, M. C. & Eaves, L. J. Estimating fetal and maternal genetic contributions to premature birth from multiparous pregnancy histories of twins using MCMC and maximum-likelihood approaches. *Twin Res Hum. Genet* **12**, 333–342 (2009).
40. Alleman, B. W. et al. No observed association for mitochondrial SNPs with preterm delivery and related outcomes. *Pediatr. Res.* **72**, 539–544 (2012).
41. Manuck, T. A. et al. Absence of mitochondrial progesterone receptor polymorphisms in women with spontaneous preterm birth. *Reprod. Sci.* **17**, 913–916 (2010).
42. Velez, D. R. et al. Mitochondrial DNA variant A4917G, smoking and spontaneous preterm birth. *Mitochondrion* **8**, 130–135 (2008).
43. Zhang, G. et al. Genetic associations with gestational duration and spontaneous preterm birth. *N. Engl. J. Med.* **377**, 1156–1167 (2017).
44. Rosenberg, N. A. et al. Genome-wide association studies in diverse populations. *Nat. Rev. Genet.* **11**, 356–366 (2010).
45. Parets, S. E., Conneely, K. N., Kilaru, V., Menon, R. & Smith, A. K. DNA methylation provides insight into intergenerational risk for preterm birth in African Americans. *Epigenetics* **10**, 784–792 (2015).
46. Palo, J. U., Ulmanen, I., Lukka, M., Ellonen, P. & Sajantila, A. Genetic markers and population history: Finland revisited. *Eur. J. Hum. Genet.* **17**, 1336–1346 (2009).
47. Sheikh, I. A. et al. Spontaneous preterm birth and single nucleotide gene polymorphisms: a recent update. *BMC Genom.* **17**(Suppl 9), 759 (2016).
48. Plunkett, J. & Muglia, L. J. Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. *Ann. Med.* **40**, 167–195 (2008).
49. Rood, K. M. & Buhimschi, C. S. Genetics, hormonal influences, and preterm birth. *Semin. Perinatol.* **41**, 401–408 (2017).
50. Strauss, J. F. et al. Spontaneous preterm birth: advances toward the discovery of genetic predisposition. *Am. J. Obstet. Gynecol.* **218**, 294–314 (2018).
51. Singh, A. J., Ramsey, S. A., Filtz, T. M. & Kioussi, C. Differential gene regulatory networks in development and disease. *Cell. Mol. Life Sci.* **75**, 1013–1025 (2017).
52. Prince, A. L. et al. The placental membrane microbiome is altered among subjects with spontaneous preterm birth with and without chorioamnionitis. *Am. J. Obstet. Gynecol.* **214**, 627.e16 (2016).
53. Nelson, D. B., Shin, H., Wu, J. & Dominguez-Bello, M. G. The gestational vaginal microbiome and spontaneous preterm birth among Nulliparous African American women. *Am. J. Perinatol.* **33**, 887–893 (2016).
54. Dahl, C. et al. Gut microbiome of mothers delivering prematurely shows reduced diversity and lower relative abundance of Bifidobacterium and Streptococcus. *PLoS ONE* **12**, e0184336 (2017).
55. Haataja, R. et al. Mapping a new spontaneous preterm birth susceptibility gene, IGF1R, using linkage, haplotype sharing, and association analysis. *PLoS Genet.* **7**, e1001293 (2011).
56. Uzun, A., Dewan, A. T., Istrail, S. & Padbury, J. F. Pathway-based genetic analysis of preterm birth. *Genomics* **101**, 163–170 (2013).
57. Rahkonen, L. et al. 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**, 701–710 (2010).
58. Conde-Agudelo, A. & Romero, R. Cervical phosphorylated insulin-like growth factor binding protein-1 test for the prediction of preterm birth: a systematic review and meta-analysis. *Am. J. Obstet. Gynecol.* **214**, 57–73 (2016).

59. Karjalainen, M. K. et al. A potential novel spontaneous preterm birth gene, AR, identified by linkage and association analysis of X chromosomal markers. *PLoS ONE* **7**, e51378 (2012).
60. Bethin, K. E. et al. Microarray analysis of uterine gene expression in mouse and human pregnancy. *Mol. Endocrinol.* **17**, 1454–1469 (2003).
61. Makieva, S., Saunders, P. T. & Norman, J. E. Androgens in pregnancy: roles in parturition. *Hum. Reprod. Update* **20**, 542–559 (2014).
62. Karjalainen, M. K. et al. CXCR3 polymorphism and expression associate with spontaneous preterm birth. *J. Immunol.* **195**, 2187–2198 (2015).
63. McElroy, J. J. et al. Maternal coding variants in complement receptor 1 and spontaneous idiopathic preterm birth. *Hum. Genet.* **132**, 935–942 (2013).
64. Uzun, A. et al. Targeted sequencing and meta-analysis of preterm birth. *PLoS ONE* **11**, e0155021 (2016).
65. Modi, B. P. et al. Rare mutations and potentially damaging missense variants in genes encoding fibrillar collagens and proteins involved in their production are candidates for risk for preterm premature rupture of membranes. *PLoS ONE* **12**, e0174356 (2017).
66. Modi, B. P. et al. Mutations in fetal genes involved in innate immunity and host defense against microbes increase risk of preterm premature rupture of membranes (PPROM). *Mol. Genet. Genom. Med.* **5**, 720–729 (2017).
67. Zhang, H. et al. A genome-wide association study of early spontaneous preterm delivery. *Genet. Epidemiol.* **39**, 217–226 (2015).
68. Bacelis, J. et al. Literature-informed analysis of a genome-wide association study of gestational age in Norwegian women and children suggests involvement of inflammatory pathways. *PLoS ONE* **11**, e0160335 (2016).
69. Rappoport, N. et al. A genome-wide association study identifies only two ancestry specific variants associated with spontaneous preterm birth. *Sci. Rep.* **8**, 226 (2018).
70. Plunkett, J. et al. An evolutionary genomic approach to identify genes involved in human birth timing. *PLoS Genet.* **7**, e1001365 (2011).
71. Haapalainen, A. M. et al. Expression of CPPED1 in human trophoblasts is associated with timing of term birth. *J. Cell. Mol. Med.* **22**, 968–981 (2018).
72. Zhuo, D. X. et al. CSTP1, a novel protein phosphatase, blocks cell cycle, promotes cell apoptosis, and suppresses tumor growth of bladder cancer by directly dephosphorylating Akt at Ser473 site. *PLoS ONE* **8**, e65679 (2013).
73. Diep, C. H. et al. Progesterone receptors induce FOXO1-dependent senescence in ovarian cancer cells. *Cell Cycle* **12**, 1433–1449 (2013).
74. Beaumont, R. N. et al. Genome-wide association study of offspring birth weight in 86577 women identifies five novel loci and highlights maternal genetic effects that are independent of fetal genetics. *Hum. Mol. Genet.* **27**, 742–756 (2018).
75. GOPEC Consortium. Disentangling fetal and maternal susceptibility for pre-eclampsia: a British multicenter candidate-gene study. *Am. J. Hum. Genet.* **77**, 127–131 (2005).
76. Kaukola, T. et al. Population cohort associating chorioamnionitis, cord inflammatory cytokines and neurologic outcome in very preterm, extremely low birth weight infants. *Pediatr. Res.* **59**, 478–483 (2006).
77. Marsál, K. Intrauterine growth restriction. *Curr. Opin. Obstet. Gynecol.* **14**, 127–135 (2002).
78. Zeitlin, J., Ancel, P. Y., Saurel-Cubizolles, M. J. & Papiernik, E. The relationship between intrauterine growth restriction and preterm delivery: an empirical approach using data from a European case-control study. *Br. J. Obstet. Gynecol.* **107**, 750–758 (2000).
79. Burk, R. F. & Hill, K. E. Regulation of Selenium metabolism and transport. *Annu. Rev. Nutr.* **35**, 109–134 (2015).
80. Li, M. et al. Loss of selenocysteine insertion sequence binding protein 2 suppresses the proliferation, migration/invasion and hormone secretion of human trophoblast cells via the PI3K/Akt and ERK signaling pathway. *Placenta* **55**, 81–89 (2017).
81. McDermott, J. R. et al. Zinc- and bicarbonate-dependent ZIP8 trasporter mediates selenite uptake. *Oncotarget* **7**, 35327–35340 (2016).
82. MacFarquhar, J. K. et al. Acute selenium toxicity associated with a dietary supplement. *Arch. Intern. Med.* **170**, 256–261 (2010).
83. Zhou, H., Wang, T., Li, Q. & Li, D. Prevention of Keshan disease by selenium supplementation: a systematic review and meta-analysis. Biol. Trace Elem. Res. https://doi.org/10.1007/s12011-018-1302-5 (2018).
84. Schweizer, U. & Fradejas-Villar, N. Why 21? The significance of selenoproteins for human health revealed by inborn errors of metabolism. *FASEB J.* **30**, 3669–3681 (2016).
85. Rayman, M. P., Wijnen, H., Vader, H., Kooistra, L. & Pop, V. Maternal selenium status during early gestation and risk for preterm birth. *Can. Med. Assoc. J.* **183**, 549–555 (2011).
86. Hurst, R. et al. Soil-type influences human selenium status and underlies widespread selenium deficiency risks in Malawi. *Sci. Rep.* **3**, 1425 (2013).
87. Olsen, P. et al. Epidemiology of preterm delivery in two birth cohorts with an interval of 20 years. *Am. J. Epidemiol.* **142**, 1184–1193 (1995).
88. Knöfler, M. & Pollheimer, J. Human placental trophoblast invasion and differentiation: a particular focus on Wnt signaling. *Front. Genet.* **4**, 190 (2013).
89. Huusko, J. M. et al. Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLoS Genet.* **14**, e1007394 (2018).
90. Zhang, G. et al. Assessing the causal relationship of maternal height on birth size and gestational age at birth: a mendelian randomization analysis. *PLoS Med.* **12**, e1001865 (2015).