

Genetic variants associated with spontaneous preterm birth in women from India: a prospective cohort study



Esha Bhattacharjee,^{a,b} Ramachandran Thiruvengadam,^{c,d} Ayushi,^c Chitrapita Das,^a GARBH-Ini Team, Nitya Wadhwa,^c Uma Chandra Mouli Natchu,^{c,e} Pallavi Kshetrapal,^c Shinijini Bhatnagar,^{c,***} Partha Pratim Majumder,^{a,f,**} and Arindam Maitra^{a,*}



^aNational Institute of Biomedical Genomics, PO: NSS, Kalyani, India

^bRegional Centre for Biotechnology, 3rd Milestone, Faridabad-Gurugram Expressway, Faridabad, India

^cTranslational Health Science and Technology Institute, 3rd Milestone, Faridabad-Gurugram Expressway, Faridabad, India

^dPondicherry Institute of Medical Sciences, Ganapathichettikulam, Kalapet, Puducherry, India

^eDivision of Infectious Diseases, St. John's Research Institute, 100 Feet Road, John Nagar, Koramangala, Bengaluru, India

^fIndian Statistical Institute, Barrackpore Trunk Road, Kolkata, India

Summary

Background Despite having the highest number of preterm births globally, no genomic study on preterm birth was previously published from India or other South-Asian countries.

Methods We conducted a genome-wide association (GWA) study of spontaneous preterm birth (sPTB) on 6211 women from India. We used a novel resampling procedure to identify the associated single nucleotide polymorphisms (SNPs) followed by haplotype association analysis and imputation.

Findings We found that 512 maternal SNPs were associated with sPTB ($p < 2.51e-3$), of which minor allele at 19 SNPs (after Bonferroni correction) had increased genotype relative risk. Haplotypes containing six of the 19 SNPs (rs13011430, rs8179838, rs2327290, rs4798499, rs7629800, and rs13180906) were associated with sPTB ($p < 9.9e-4$; Bonferroni adjusted p -value < 0.05). After imputation in regions around the 19 SNPs, 15 imputed SNPs were found to be associated with sPTB (Bonferroni adjusted p -value < 0.05). One of these imputed SNPs, rs35760881, and three other SNPs (rs17307697, rs4308815, and rs10983507) were also reported to be associated with sPTB in women belonging to European ancestry. Moreover, we found that GG genotype at rs1152954, one of the associated SNPs, enhanced risk of sPTB and reduced telomere length.

Interpretation This is the first study from South Asia on the genome-wide identification of maternal SNPs associated with sPTB. These SNPs are known to alter the expression of genes associated with major pathways in sPTB viz. inflammation, apoptosis, cervical ripening, telomere maintenance, selenocysteine biosynthesis, myometrial contraction, and innate immunity. From a public health perspective, the trans-ethnic association of four SNPs identified in our study may help to stratify women with risk of sPTB in most populations.

Funding Department of Biotechnology (India), Grand Challenges India - All Children Thriving Program and Biotechnology Industry Research Assistance Council (BIRAC).

Copyright © 2023 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

Keywords: Spontaneous preterm birth; GWAS; India; Maternal genome; Cluster file

Introduction

Preterm birth (PTB), defined as birth before 37 completed weeks of gestation,¹ being the leading cause of neonatal and infant mortality, imposes enormous burden on the family, and public health infrastructure. Of all babies born annually in India, about 13% are born preterm; accounting for 23.4% of PTBs globally.¹

Preterm infants suffer from various medical complications including delayed development and are at increased risk of adult-onset diseases early in their lives.²

Spontaneous preterm birth (sPTB), the major subtype of PTB, occurs either due to spontaneous early onset of labour or due to preterm prelabour rupture of

*Corresponding author. National Institute of Biomedical Genomics, Kalyani, India.

**Corresponding author.

***Corresponding author. Translational Health Science and Technology Institute, Faridabad, India.

E-mail addresses: shinijini.bhatnagar@thsti.res.in (S. Bhatnagar), pmp1@nibmg.ac.in (P.P. Majumder), am1@nibmg.ac.in (A. Maitra).

Research in context

Evidence before this study

Previous candidate-gene association studies as well as limited genome-wide association study in European ancestral population have provided evidence on maternal genetic predisposition to spontaneous preterm birth (sPTB). The largest genome-wide association study (GWAS) of preterm birth, which used an additive model, found common variants in three loci (*EBF1*, *EEFSEC*, and *AGTR2*) to be significantly associated with spontaneous preterm birth. However, both the discovery and replication datasets belonged to European ancestral population. Information on other ancestral populations and use of non-additive models of inheritance are not available as on date. Despite having the highest number of preterm births globally, no genomic study on preterm birth was previously published from India or other South-Asian countries.

Added value of this study

To our knowledge, this is the first GWAS from South Asian ancestral population, which investigated the association of maternal genotypes with spontaneous preterm birth. In addition to the identification of 15 significantly associated single nucleotide polymorphisms (SNPs) with sPTB, we also identified four significantly associated transethnic SNPs (rs35760881-A, rs17307697-T, rs4308815-C and rs10983507-G) which showed association both in Indian women as well as in the women belonging to the European ancestral population. Interestingly, rs35760881 is a cis-eQTL (cis-expression quantitative trait loci) of *AKIP1* in blood, which is known to activate the pro-inflammatory state by enhancing

the NF- κ B signalling. We were also able to identify 16 other novel population-specific loci in our study, in which either only the minor allele homozygotes or both the minor allele homozygotes and heterozygotes showed enhanced risk for sPTB compared to major allele homozygotes. Additionally, we found that women with GG genotype at rs1152954, one of the population-specific loci, which is in the enhancer region of *CCT2*, was associated with increased risk of delivering preterm as well as with reduced relative telomere length, a genomic marker of cellular stress. Functional annotation and biological process enrichment analysis based on our GWAS results, revealed several previously known as well as novel yet relevant biological processes that might culminate in spontaneous preterm birth.

Implications of all the available evidence

Our study on Indian population adds to the existing limited GWAS findings on sPTB, by identifying both trans-ethnic and population specific maternal genomic variants, that might affect different biological processes and molecular mechanisms culminating in spontaneous preterm birth. The SNPs showing trans-ethnic association requires further investigation in other populations which will help us to understand its global clinical relevance. From public health perspective, the genomic markers either alone or in combination with other clinical or biological variables could be evaluated for their predictive abilities. This might help in the early triaging of women at high risk of sPTB to appropriate antenatal medical care.

membranes.³ Environment as well as maternal and foetal genomic factors influence the risk of PTB. The heritability estimate of PTB in a Swedish twin study was 36% while in an Australian twin study, the heritability varied from 17% in primigravidity to 27% in multi-gravidity.^{4,5} Moreover, the enhanced risk of recurrent sPTB in mothers suggests a genetic predisposition to sPTB.⁶ Most epidemiological studies have shown that sPTB is influenced more by the maternal genome than the foetal or paternal genomes.^{7–10} However, to our knowledge only two GWAS have investigated the association of maternal genomic variants with sPTB.^{11,12} Of these, the study by Zhang and colleagues on women of European ancestry, was able to discover and replicate genomic associations with sPTB.¹²

Most genomic studies on sPTB have been conducted on European and African-American ancestral populations.¹³ Although India contributes the highest number of PTBs globally, there are no previous genomic studies on sPTB from India or other South-Asian countries. We hypothesized that maternal genetic variants are associated with sPTB and conducted GWAS in a well-characterised cohort of pregnant women in

India¹⁴ using a nested matched preterm-term analysis to identify maternal genotypes that are significantly associated with sPTB.

Methods

Study participants

Pregnant women enrolled between May 2015 to April 2021 in a large single-centre prospective cohort¹⁴ (GARBH-Ini: interdisciplinary Group for Advanced Research on Birth Outcomes - DBT India Initiative) in Haryana, India, were followed up with written informed consent till six months after delivery (cohort description provided in [Appendix](#)). In this cohort, we identified the subset of women who delivered preterm (<37 weeks) and those who delivered at term (37⁺⁰ weeks–41⁺⁶ weeks). To minimise misclassification of preterm and term delivering individuals, (i) we included those women who were enrolled in the cohort before 14 weeks of gestational age as per ultrasonography dating, and (ii) from the individuals who were enrolled between 14⁺⁰ weeks and 19⁺⁶ weeks of gestational age, we included those women whose gestational age at enrolment

concurrent within seven days between last menstrual date-based dating and ultrasonography-based dating. Women with age <18 years, having multiple pregnancies, stillbirths, and care-giver initiated preterm deliveries were excluded from this study.

This study was approved by the institutional review boards and ethics committees of Gurugram Civil Hospital where the pregnant women were enrolled, Translational Health Science and Technology Institute where the biospecimens were stored and DNA was extracted, and National Institute of Biomedical Genomics where the genotyping and analyses were conducted. Personal identifiers of the participants were replaced with pseudonymised unique identifiers.

Genome-wide genotyping

Genome-wide genotyping assay of all 6211 mothers in the cohort was carried out using Infinium Global Screening Array (GSA) version 3.0 with multi-disease drop-in panel which contains 700,604 markers (Illumina Inc, USA). We performed genotype clustering using custom cluster file, developed by us in house (Appendix), in GenomeStudio 2.0 software (Genotyping module 2.0.4, Illumina Inc, USA).

Genotype data quality control and imputation

PLINK¹⁵ version 1.9 was used for quality control of the genome-wide genotype data. We excluded loci with call rate <95% (numbers excluded, $n_e = 4551$), minor allele frequency <5% ($n_e = 348,443$), loci that significantly deviated from Hardy–Weinberg equilibrium [False discovery rate Benjamini-Hochberg (FDR-BH) q-value <0.05, $n_e = 2840$], and loci for which the number of individuals of any genotype was <10 ($n_e = 93,766$), from the autosomal and X-chromosomal biallelic single nucleotide polymorphisms (SNPs; $n = 677,695$). We removed individuals showing sex discrepancy ($n_e = 1$), having call rate <95% ($n_e = 47$), cryptic relatedness (Pihat >0.2) ($n_e = 117$), high or low heterozygosity rates (± 3.84 SD from the mean) ($n_e = 78$), and those who did not meet our inclusion criteria described in the “Study participants” section ($n_e = 1286$).

Since samples were genotyped randomly in an anonymised manner without knowing the delivery status (preterm or term), we checked for the presence of batch effect in the data. In PLINK, we tested for the frequency of non-random missingness of genotype data between individuals delivering preterm and term using Fisher’s exact test and excluded 18,802 SNPs which showed batch effect (unadjusted $p < 0.1$; strict cut-off to not allow any SNP with batch-effect in the data). After quality control, 209,293 biallelic SNPs and 4682 individuals were retained for analysis.

We imputed untyped SNPs ± 500 Kbp around associated SNPs, with IMPUTE2¹⁶ using reference haplotypes from both 1000 Genome Phase 3¹⁷ and Genome Asia Pilot project¹⁸ after phasing with SHAPEIT2.¹⁹

Apart from performing quality control of the imputed loci as undertaken in genotyped data, SNPs with INFO score <0.9 were excluded to enhance the accuracy of imputed data.

Statistical analysis

Among 4682 individuals that passed quality control and sample selection criteria, 521 women delivered preterm while 4161 women delivered at term. For each preterm delivering women, we randomly selected two term delivering women matched for the age group (18–24 years, 25–30 years, 31–35 years, 36–40 years, >40 years), BMI status (underweight, normal weight, overweight, obese), parity (0, 1, 2, 3, ≥ 4), and occupational status (unemployed, unskilled or semi-skilled, and skilled), to avoid possible confounding effects of these variables on birth outcomes. Thus, this study is a 1:2 matched nested preterm-term association study with samples sizes of 521 preterm and 1042 term delivering women (Fig. 1).

To test for association with a SNP locus, we calculated the association statistic for the observed counts in the 3×2 table of three genotypes and the two attributes (preterm and term) as:

$$X = \frac{\sum_i \sum_j (O_{ij} - E_{ij})^2}{E_{ij}}$$

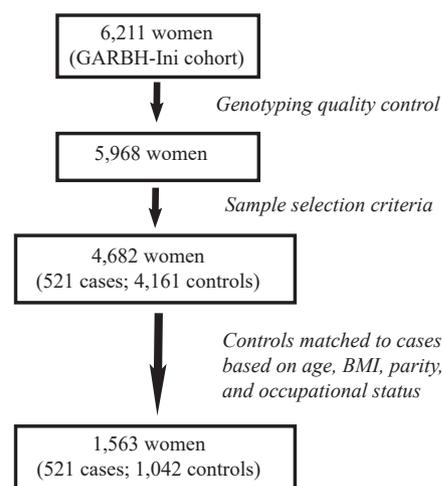


Fig. 1: Overflow of sample selection criteria. 6211 women from the cohort were genotyped and 5968 women passed the genotyping quality control. We then applied the selection criteria (gestational age at enrolment <14 weeks or difference between LMP dating and USG dating <7 days, participant’s age ≥ 18 years, only spontaneous singleton live birth without any complications) on these 5968 women and 4682 women were selected among which 521 had spontaneous preterm delivery (<37 weeks) while 4161 had term delivery (≥ 37 weeks – ≤ 41 weeks). Next, we matched preterm and term delivering women in 1:2 ratio based on age, BMI, parity and occupational status for the association analysis.

where O_{ij} and E_{ij} are the observed and expected (under the hypothesis of independence of the genotypes and the two attributes) number of individuals in the (i,j) -th cell of the 3×2 table ($i = 1,2,3; j = 1,2$).

To test the significance of an observed association value, we adopted a resampling technique. We did not employ the standard method of deriving the p -value of the association statistic using the chi-squared distribution with appropriate degrees of freedom and then correcting for multiple tests because (a) the assayed loci are in linkage disequilibrium with variable strengths of disequilibrium that result in the association statistics being correlated, (b) there is no community-accepted statistical method of taking such correlations into account, and (c) correcting for multiple tests assuming that the tests are independent leads to overcorrection, resulting in declaring significant associations as non-significant. Some further reasons are provided in the [Appendix](#). To implement our resampling (bootstrapping) strategy, we randomly sampled, with replacement, 521 preterm delivering women from our dataset. For each preterm delivering woman sampled, we included the corresponding pair of mothers who delivered at term. In the bootstrap replicate, we computed the association statistic for each SNP. Thus, for one bootstrap replication we calculated 209,293 association statistics, equalling the total number of SNPs after curation. We replicated the bootstrap procedure 500 times, thereby obtaining $209,293 \times 500 = 104,646,500$ values of the association statistic. We sorted these values in increasing order of magnitude and obtained an empirical bootstrap frequency distribution of the values of association statistic, from which the value of the statistic corresponding to the upper 5% tail of the distribution was derived. We then compared our observed values of the association statistics against this 5% threshold value and declared those SNPs to be significantly associated at 5% level whose association statistics exceeded this threshold.

Relative risks (RRs) for both heterozygous and minor allele homozygous genotypes compared to major allele homozygous genotype were calculated with 95% CI per internally validated SNP in R (version 4.0.1).²⁰ We identified significantly associated SNPs with increased relative risk in either minor allele homozygotes, or both minor allele homozygotes and heterozygotes compared to major allele homozygotes using Bonferroni's correction.

The SNPs passing Bonferroni corrected p -value significance threshold were retested for validation using a logistic regression model to test for genotype effects after adjusting for the covariates—age, BMI, occupational status, and parity.

A subgroup analysis was performed to identify whether any of the significant SNPs was associated with early sPTB, i.e., birth before 33 weeks of

pregnancy. 62 women delivered before 33 weeks of gestational age. For each early preterm delivering women, two women who delivered at term were randomly selected after matching for age group, BMI group, parity and occupational status alike the previous analysis. Thus, in this subgroup analysis, there were 62 early preterm delivering women and 124 term delivering women. After the association analysis, RR was calculated for the SNPs.

Haplotype estimation

We estimated haploblocks ± 500 kb around the validated SNPs using Gabriel's method of 95% CI in Haploview 4.1²¹ and tested the association of haplotype frequency with sPTB.

Evaluation of trans-ethnic association

SNPs which were common between our data and the summary statistics of top 10,000 variants ($p < 1e-3$) from a previous GWAS conducted on women of European ancestry,¹² were identified. Since this study¹² had assumed an additive model, we investigated whether these SNPs were significantly associated under additive model in our population using PLINK.

Functional annotation

We at first annotated the most significantly associated SNPs from each imputed region based on their position—exonic, intronic, or intergenic using ANNOVAR.²² We used CADD²³ to predict the deleteriousness of these SNPs under GRCh37-v1.6. Next, we used FUMA²⁴ to identify whether these SNPs overlap with any cis-expression quantitative trait loci (cis-eQTL) or enhancer region of any gene of utmost relevance to sPTB.

Functional impact of minor allele at rs1152954 on telomere length

We included relative telomere length [Telomere-to-Single Copy Gene ratio (T/S ratio)] data²⁵ which was available for 220 individuals among 4682 individuals that passed genotyping quality control and selection criteria. A logistic regression model was used to test for the association of T/S ratio at both the time points with sPTB adjusting for age (years), BMI at enrolment (kg/m^2), parity, occupational status, and sex of the infant.

We also tested whether mean T/S ratio was significantly smaller in individuals with minor allele homozygous genotype than heterozygous genotype at rs1152954 using t -test.

Biological process enrichment analysis

Variants having $p < 0.001$ were positionally mapped to genes. Enrichment analysis of biological processes was performed at Gene Ontology Resources (<http://geneontology.org/>; release date: 16/05/2022) using Fisher's exact test.

Role of funding source

Funder had no role in study design, data collection, data analysis, data interpretation, or writing of the report.

Results

Association of maternal genomic variants with spontaneous preterm birth

Demographic characteristics of 521 preterm and 1042 term delivering women are provided in [Appendix Table S1](#). From the genotype-based genome-wide association analysis with 209,293 maternal biallelic SNPs, we found that the value of the association statistic corresponding to the upper 5% tail of the bootstrap frequency distribution to be 11.98. Of the 209,293 observed values of the association statistic, 512 exceeded the threshold of 11.98 ([Fig. 2](#)); indicating that there were 512 SNPs associated with statistical significance ($p < 2.51e-3$). Genomic inflation factor further indicated that this result is not confounded by significant population sub-structuring ($\lambda_{0.5} = 1.004$).

Genotype relative risk

We assessed the risk of delivering preterm due to the presence of minor allele homozygous genotype (BB) and heterozygous genotype (AB) compared to the major allele homozygous genotype (AA) for each of the 512 statistically significant SNPs. We identified 149 SNPs, of which, at 37 SNPs both BB and AB genotypes enhanced sPTB risk, while at 112 SNPs only BB genotype enhanced sPTB risk with statistical significance.

Among these 149 SNPs, where either BB, or both AB and BB genotypes showed increased relative risk as compared to AA, after applying Bonferroni correction ($p < 0.05/149$, or $3.36e-4$), 19 SNPs were prioritised

([Appendix Table S2](#)). We used logistic regression model to test genotype effects at a locus on the same data for validation and found all these 19 SNPs to be associated with sPTB with statistical significance. The four covariates— age, parity, BMI at enrolment, and occupational status— were also used as independent predictors of sPTB; none of these turned out to be significant.

To ensure the quality of genotyping data of these 19 validated SNPs (rs4362135, rs2327290, rs13011430, rs10026052, rs2070235, rs12208914, rs8179838, rs10983328, rs4798499, rs7629800, rs13180906, rs2689089, rs11696299, rs710702, rs10485983, rs11727167, rs1799041, rs7645913, and rs16238), we manually inspected their SNP graphs, generated by joint-calling of the genotype data of all the 6211 individuals using GenomeStudio. We found that all these 19 SNPs had three well-separated genotype clusters ([Appendix Fig. S1](#)). These 19 validated SNPs were considered for further downstream analysis.

On evaluation of the association of 19 validated SNPs with early sPTB, we found that six SNPs were associated with early sPTB at statistical significance ($p < 0.05$), however, five of them (rs4798499, rs2689089, rs7645913, rs10026052 and rs16238) significantly conferred increased risk on early sPTB ([Appendix Table S3](#)).

Haplotype association with sPTB

Among the 19 validated SNPs, ten were present in haploblocks, and rs13011430 and rs8179838 were in the same haploblock ([Table 1](#)). Upon evaluation of the association of haplotype frequency with sPTB, we identified ten haplotypes from five haploblocks that showed statistically significant association with sPTB after Bonferroni correction ($p < 9.9e-4$).

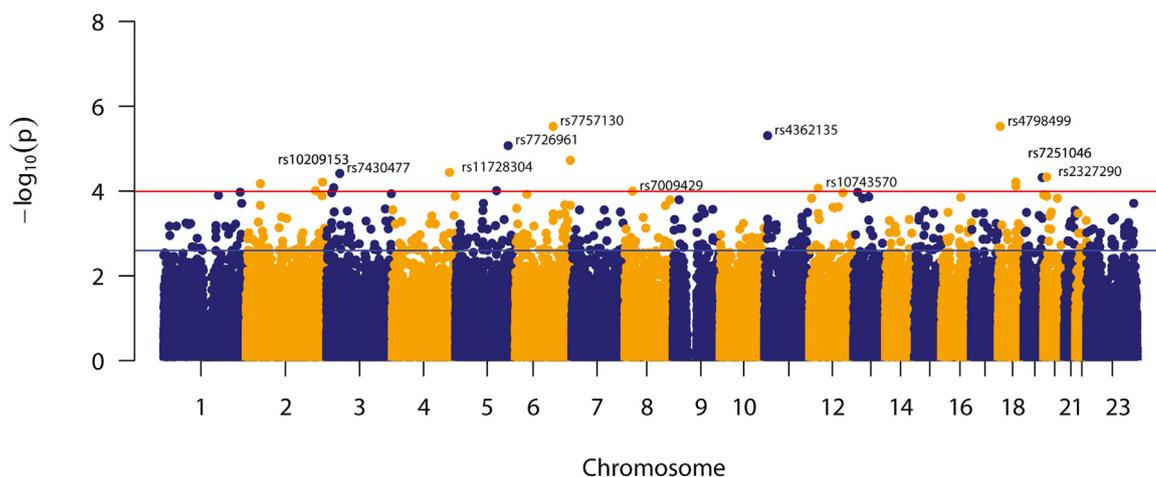


Fig. 2: Manhattan plot showing strength of association (negative logarithm of p-value of association statistic) of the tested SNPs with sPTB. Horizontal lines indicate strength of associations i.e. $p = 1.02e-4$ and $p = 2.51e-3$ corresponding to the 1% (red) and 5% (blue) upper tail association statistic thresholds obtained from the empirical bootstrap frequency distribution, respectively.

SNPs in haploblock	Haplotypes (>1%) ^a	Frequency in preterm delivering women	Frequency in term delivering women	p	Adjusted p ^b
rs13011430 rs8179838	CT	0.67	0.74	2.66e-5	7.98e-5
	TC	0.29	0.22	2.62e-5	7.86e-5
	CC	0.03	0.03	0.91	ns
rs10775437 rs4798497 rs2889405 rs4798499	GACC	0.4	0.39	0.89	ns
	GAAT	0.31	0.26	0.0015	0.0075
	AAAC	0.23	0.27	0.008	0.04
	GGCC	0.04	0.06	0.09	ns
	GAAC	0.02	0.01	0.68	ns
rs11727167 rs114981738	AT	0.49	0.51	0.21	ns
	GT	0.35	0.33	0.35	ns
	AC	0.16	0.16	0.62	ns
rs2689089 rs182107462	AA	0.73	0.76	0.14	ns
	CA	0.19	0.18	0.35	ns
	CC	0.08	0.07	0.28	ns
rs12669058 rs1799041 rs740209	AGG	0.29	0.30	0.34	ns
	CGT	0.28	0.28	0.82	ns
	AGT	0.24	0.25	0.53	ns
	AAT	0.19	0.16	0.06	ns
rs16237 rs16238 rs16968617	TCG	0.33	0.36	0.15	ns
	GAG	0.36	0.32	0.06	ns
	GCA	0.23	0.22	0.57	ns
	GCG	0.08	0.09	0.10	ns
rs7629800 rs11129012	TC	0.42	0.36	0.003	0.009
	CC	0.31	0.34	0.09	ns
	CT	0.27	0.29	0.15	ns
rs75603134 rs13180906 rs10078386 rs77414036	ACTT	0.48	0.41	8.44e-5	3.38e-4
	ATTT	0.30	0.35	0.0074	0.03
	ATGT	0.14	0.15	0.74	ns
	GTTC	0.07	0.09	0.07	ns
rs2327290 rs6039882 rs6131083 rs6074125	CTTG	0.34	0.37	0.04	ns
	CCTA	0.22	0.26	0.0099	0.04
	ACTG	0.26	0.22	0.0094	0.04
	ACGG	0.18	0.14	0.0069	0.03

^aSNPs of interest" are marked in bold. ns means not significant. ^bHaplotypes with more than 1% frequency among the study participants were considered. ^b Bonferroni-adjusted p-values have been enlisted. P-values have been adjusted for the number of tests within each block.

Table 1: Association of haplotypes of the SNPs of interest with spontaneous preterm birth.

Imputation to identify significant SNPs in proximity to the SNPs of interest

Genotype imputation ±500 kb of the 19 validated SNPs led to the detection of 15 SNPs (twelve SNPs in chromosome 20, and three in chromosome 11) that showed statistically significant association with sPTB, after Bonferroni correction ($p < 2e-6$). In chromosome 20, rs6032870 [16.18 kbp downstream of the genotyped SNP rs2327290 ($r^2 = 0.6$ and $D' = 1$)] and in chromosome 11, rs35760881 [153 bp downstream of the genotyped SNP rs4362135 ($r^2 = 0.8$ and $D' = 1$)] showed the most statistically significant associations ($p = 3.058e-8$ and $1.475e-6$, respectively).

Regional plots showed that, in most of the imputed regions, the imputed SNPs showed higher association signals than the genotyped SNPs of interest (Appendix

Fig. S2). Among 19 validated SNPs, rs13011430 and rs8179838 in chromosome 2, as well as rs7629800 and rs7645913 in chromosome 3 were located in the same regional plots respectively, and hence, there were 17 imputed regions.

Trans-ethnic association analysis

Among the 10,000 strongly associated variants from a previously published GWAS of sPTB, conducted on the women of European ancestry,¹² we found 218 SNPs in our data.

Out of these, eight SNPs (rs35760881, rs2999572, rs12293813, rs16949470, rs17307967, rs4308815, rs10983508, and rs10983507) were associated with sPTB with statistical significance of $p < 0.01$; the ORs of seven SNPs except for rs12293813, were in the same direction

of effect as in the European population. Interestingly, four SNPs (rs35760881, rs17307967, rs4308815, and rs10983507) showed statistically significant trans-ethnic association with sPTB after Bonferroni correction ($p < 0.05/218$ or, $2.3e-4$) (Appendix Table S4).

Further analysis of rs35760881, which showed statistically significant association with sPTB in GARBH-Ini cohort, as well as in the European dataset, revealed that both BB and AB genotypes enhanced risk of sPTB [RR (95% CI) for BB: 2.05 (1.49–2.81) and for AB: 1.39 (1.2–1.61)]. After assaying individuals with TaqMan assay, we found a genotype concordance of 98.05% between the imputed and assayed genotypes at rs35760881 (Appendix) which was within expectations.

Functional annotation of the associated SNPs

Of the index SNPs from the 17 imputed regions, 13 were intronic while four were intergenic. Several of these index SNPs showed regulatory features. Interestingly, one of the associated SNPs, rs2070235 (CADD score: 17.80) is a missense variant on Myb-related protein B (*MYBL2*) gene that leads to Ser427Gly amino acid change²⁶ (Table 2).

Minor allele homozygotes at rs1152954 have shorter telomere length

Since rs1152954 lies within the enhancer region of Chaperonin Containing T-Complex Polypeptide 1, Subunit 2 (*CCT2*) gene, which is involved in telomere maintenance, we evaluated the effect of genotypes at rs1152954 on telomere length. First, we assessed whether telomere length measured as T/S ratio at 18–20 weeks and 26–28 weeks was associated with sPTB after controlling for age, BMI at enrolment, parity, occupational status, and sex of the infant. We found that decrease in the T/S ratio at 26–28 weeks showed statistically significant association with sPTB ($p = 7.74e-7$, $\beta = -2.96$) (Appendix Table S5). As the minor allele homozygotes at rs1152954 had enhanced risk of sPTB (RR: 1.96 [1.51–2.55]), we postulated that they might have shortened telomere. Interestingly, we found that one copy of major allele (A) in rs1152954 is sufficient for telomere maintenance while two copies of minor allele (G) results in shortened telomere length (Fig. 3).

Discussion

This is the first GWAS report from South Asia on the identification of maternal genotypes associated with sPTB. In addition to performing a traditional logistic regression analysis, we also performed a preterm-term birth analysis. For both analyses, we considered a 1:2 matched set of preterm-to term-delivering women, matched by age, BMI, and occupation. If we considered the entire data set, then the number of mothers delivering at term would be almost an order of magnitude

greater than the number of preterm delivering mothers and the results of the analyses would likely be adversely affected by such unbalanced data. We found that either the minor allele homozygotes, or both the minor allele homozygotes and heterozygotes conferred enhanced risk of sPTB compared to major allele homozygotes at these loci. This not only showed the importance of considering genotypes in the association test, but also indicated that different SNPs affect phenotype in different manners.

On investigation of the association of these validated SNPs with early sPTB, we found that six of 19 SNPs were associated, of which five SNPs (rs4798499, rs2689089, rs7645913, rs10026052, and rs16238) showed increased relative risk of early sPTB as compared to sPTB. This suggests that these five SNPs influence both early sPTB and sPTB.

After imputation and Bonferroni correction, 15 SNPs in two chromosomes (chromosome 20, chromosome 11) were found to be associated with sPTB. One of the associated SNP in our cohort, rs35760881 in chromosome 11, exhibited trans-ethnic association in European population along with three other intronic SNPs (rs17307967, rs4308815, and rs10983507) in *Astrotactin-2* (*ASTN2*) gene in chromosome 9. We also found population-specific association of 16 unique genomic loci with sPTB in Indian women. Functional annotation of most of these risk loci revealed that they are either located on or regulate genes which belong to biological pathways with considerable relevance to parturition. Thus, our findings indicate the existence of both trans-ethnic as well as population specific maternal loci associated with preterm birth outcome.

The trans-ethnic SNP, rs35760881, is a cis-eQTL of *A-kinase-interacting protein 1* (*AKIP1*) gene in blood. *AKIP1* encoded protein enhances nuclear factor kappaB (NF- κ B) dependent transcription²⁶ of matrix metalloproteinases, cyclooxygenase-2 (COX-2), and pro-inflammatory cytokines which results in cervical ripening, myometrial contraction, and rupture of foetal membranes.^{27,28} Our finding thus suggests a fundamental role of early activation of these phenomena which may result in sPTB in population independent manner. However, the biological role of the transethnic SNPs located in the intronic region of *ASTN2* gene is not yet clear. Recent findings indicate the existence of population specific enrichment of disease-associated genes. Further, a significant variation of disease predisposition background across the worldwide populations for common diseases, have been found.²⁹ The identification of SNPs associated with sPTB, and the predicted biological processes which might be modulated by them, in our cohort, but not reported in other populations, might indicate the preponderance of these in sPTB outcome in Indian population. Targeted studies are required to be undertaken to illuminate such

SNPs	Variant type	CADD score	Index SNP ^a	Relevant Gene ^b	Gene mapping
rs35760881	Intergenic	1.31	rs35760881	AKIP1	Cis-eQTL
rs4362135	Intergenic	5.79			
rs6032870	Intergenic	2.30	rs6032870	TMX4	Enhancer
rs2327290	Intronic (SNAP25-AS1)	0.29			
rs10211311	Intronic (MTA3)	1.34	rs10211311	MTA3	Cis-eQTL; Positional
rs13011430	Intronic (MTA3)	0.03			
rs8179838	Intronic (MTA3)	7.22			
rs2135743	Intronic (LINCO1060)	1.68	rs2135743	—	NA
rs10026052	Intronic (LINCO1060)	4.52			
rs3818215	Intronic (MYBL2)	0.50	rs3818215	MYBL2	Cis-eQTL; Positional
rs2070235	Missense (MYBL2)	17.80			
rs12208914	Intronic (UST)	0.09	rs12208914	UST	Positional
rs10983328	Intronic (ASTN2)	0.43	rs10983328	—	NA
rs4798499	Intronic (ARHGAP28)	7.05	rs4798499	ARHGAP28	Positional
rs1379490	Intronic (ZNF385D)	1.27	rs1379490	KCNH8	Enhancer
rs7629800	Intronic (ZNF385D)	9.47			
rs7645913	Intronic (ZNF385D)	3.78			
rs11743963	Intergenic	0.20	rs11743963	ZNF608	Cis-eQTL
rs13180906	Intergenic	9.13			
rs2689089	Intronic (ANKS1A)	10.40	rs2689089	RPL10A	Cis-eQTL
rs11696299	Intronic (TRIB3)	5.08	rs11696299	RBCK1	Cis-eQTL
rs1152954;	Intronic (MYRFL)	15.55	rs1152954	CCT2	Enhancer
rs710702	Intronic (MYRFL)	4.80			
rs9969211	Intronic (DNAH11)	2.17	rs9969211	CDCA7L	Cis-eQTL
rs10485983	Intronic (DNAH11)	6.35			
rs7670046	Intronic (MAEA)	2.11	rs7670046	SPON2	Cis-eQTL
rs11727167	Intronic (MAEA)	2.70			
rs61045241	Intergenic	0.23	rs61045241	SDHAF3	Enhancer
rs1799041	Intergenic	1.89			
rs9908136	Intronic (PLXDC1)	2.66	rs9908136	GSDMB	Cis-eQTL
rs16238	Intronic (PLXDC1)	2.67			

Genotyped array SNPs are marked in bold. If the index SNP in the imputed region is the imputed one, then both imputed SNP and genotyped SNP of interest are mentioned otherwise, genotyped SNP of interests are mentioned only. CADD: Combined Annotation Dependent Depletion — No relevant genes. NA: not applicable. ^aSNP with the lowest p-value in each of the 17 imputed regions. ^bGene playing major role in pathways reported to be involved in spontaneous preterm birth.

Table 2: Functional annotation of the significant SNPs from 17 imputed regions.

population specific predisposition and the role of nutritional, environmental, and other lifestyle related factors which might be unique to this region and population. Some of the biological processes which might be modulated by these SNPs are discussed below.

The most significantly associated SNP in this study, rs6032870, is an enhancer of Thioredoxin-Related Transmembrane Protein 4 (TMX4) gene in trophoblast-like cells. Although TMX4 possesses disulphide-reductase activity,²⁶ however, investigations are required to be undertaken to understand its biological pathways.

60S ribosomal protein L10a (RPL10A) gene takes part in selenocysteine synthesis by interacting with a complex formed by Selenocysteine-specific elongation factor (EEFSEC), Selenocysteine insertion sequence-binding protein 2 (SECISBP2), Sec-tRNA^{Sec}, and GTP.³⁰ Selenoproteins not only maintain the redox and

anti-inflammatory state but also play important role in reproduction.³¹ A previous study identified maternal EEFSEC variants to be associated with sPTB.¹² Interestingly, Selenium is an essential trace mineral and has been implicated in PTB by previous epidemiological studies. A recent study found statistically significant associations between maternal Selenium concentration and PTB in some, but not all, geographical sites.³² Whether population-specific genetic predisposition explains such heterogeneity and whether use of Selenium as a micronutrient in pregnant women reduces the risk of preterm birth, warrant further investigation.

T-complex protein 1 subunit beta, encoded by CCT2, is a member of T-Complex protein 1 ring complex (TRiC) which regulates telomere maintenance by promoting the folding of telomerase cofactor Telomerase Cajal body protein 1 (TCAB1).²⁶ Shortened telomere length promotes cellular senescence which triggers pro-

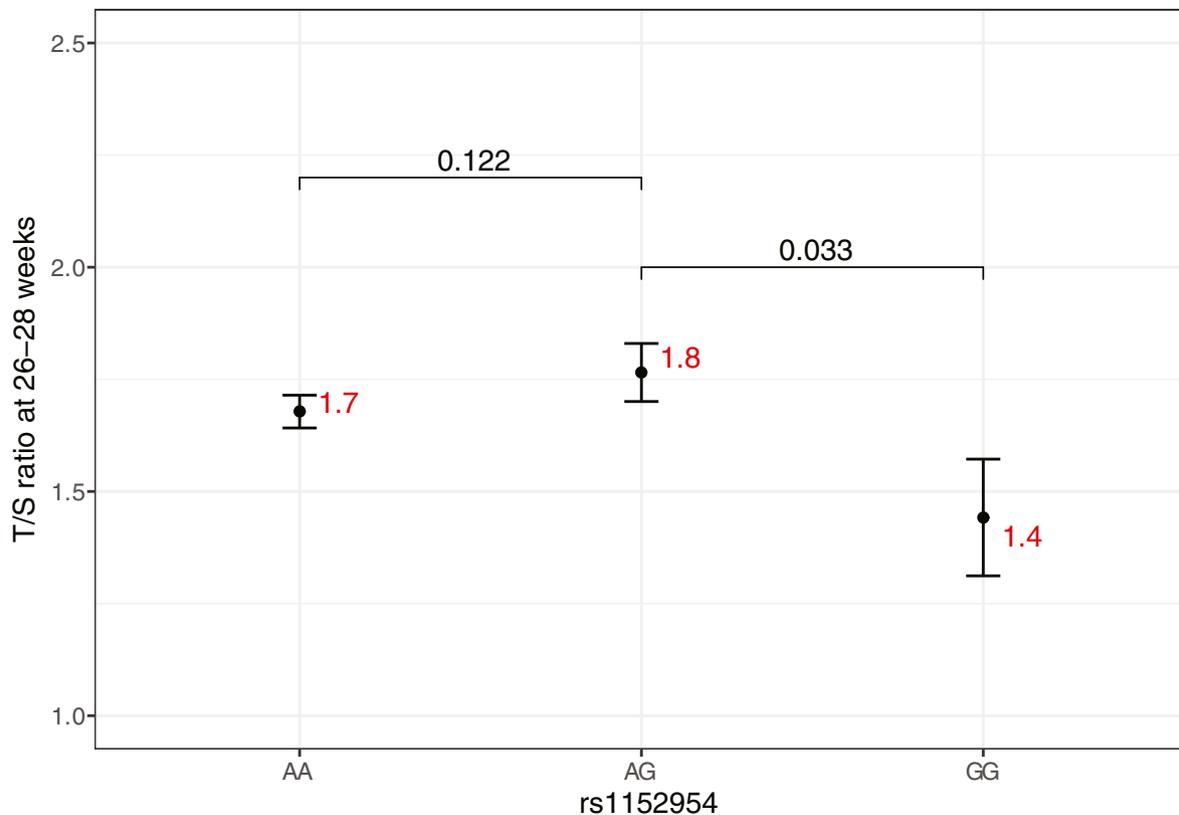


Fig. 3: Mean and standard error of Telomere-to-Single Copy Gene ratio (T/S ratio) at 26–28 weeks of each genotype at rs1152954. A is the major allele and G is the minor allele. Mean value for each genotype is indicated in red. p values for testing the null hypothesis of equality of the mean T/S ratio between genotypes are also provided. Number of individuals with AA genotype: 164, AG genotype: 51, and GG genotype: 5.

inflammatory responses.³³ Alike previous studies,^{25,34} we found that women with shorter telomere length in peripheral blood at 26–28 weeks have enhanced risk of sPTB. Moreover, women with GG genotype compared to AA and AG genotypes at rs1152954, an enhancer of *CCT2* gene, have shorter telomeres as well as higher risk of sPTB. Our results might indicate that women who are genetically predisposed to reduced telomere length, ultimately leading to enhanced cellular senescence, might be at enhanced risk of preterm delivery. This also raises the interesting prospect of whether such at risk women can be triaged early on in pregnancy and provided interventions which can reduce such risk.

Metastasis-associated protein encoded by Metastasis Associated 1 Family Member 3 (*MTA3*) gene, maintains epithelial architecture by positively regulating E-cadherin expression³⁵ and negatively regulating Wingless-Type MMTV Integration Site Family, Member 4 (*WNT4*) signalling.³⁶ Altered E-cadherin expression and *WNT4* signalling in cervix³⁷ and endometrium³⁸ respectively have been reported to culminate in sPTB.

MYBL2 codes for MYB-related protein B which transactivates the expression of the anti-apoptotic gene

clusterin (*CLU*).²⁶ Altered expression of *CLU* might result in apoptosis of placenta, leading to sPTB.

Uronyl-2-sulfotransferase (*UST*) gene participates in dermatan sulfate biosynthesis.²⁶ Loss of cervical dermatan sulphate makes the cervix soft and swollen resulting in cervical ripening³⁹ which is an important feature that triggers parturition. Early loss of dermatan sulfate might culminate in sPTB.

Rho GTPase activating protein 28 (*ARHGAP28*) gene inhibits Ras Homolog Family Member A (*RHOA*).⁴⁰ A previous study found significantly increased GTP-bound *RHOA* in myometrium in women having spontaneous preterm labour,⁴¹ which might be a consequence of altered expression of *ARHGAP28*.

The protein encoded by Ubiquitin Conjugating Enzyme 7 Interacting Protein 3 (*RBCK1*) gene is a component of LUBAC (linear ubiquitin chain assembly complex) which ubiquitinates inhibitor of kappaB kinase gamma (*IKKgamma*), thus activating NF- κ B signalling pathway.²⁶ Altered expression of *RBCK1* might promote sPTB due to upregulation of pro-inflammatory state.

Spondin-2 (*SPON2*) gene, plays a key role in initiating innate immune response.²⁶ Innate immune cells (neutrophils, macrophages, and mast cells) initiates labour by releasing pro-inflammatory factors like cytokines, chemokines, and matrix metalloproteinases.⁴² Interestingly, a previous study on gene expression revealed that *SPON2* was highly expressed in idiopathic preterm placenta.⁴³

The protein encoded by Cell Division Cycle Associated 7 Like (*CDCA7L*) gene represses the gene expression and activity of monoamine oxidase A (*MAOA*)²⁶ which was previously reported to be downregulated in preterm placentas and to have a negative correlation with placental Interleukin-6 (*IL6*) gene expression.⁴⁴ Increased expression of *CDCA7L* in the presence of rs9969211-T, thus suggests an enhanced pro-inflammatory state which might result in sPTB.

Succinate Dehydrogenase Complex Assembly Factor 3 (*SDHAF3*) gene encodes for the protein which is involved in the assembly of mitochondrial respiratory complex II, Krebs cycle, and electron transport chain.²⁶ Altered expression of *SDHAF3* might affect these biological processes resulting in increased reactive oxygen species (ROS) generation⁴⁵ which might culminate in sPTB.⁴⁶

Potassium Voltage-Gated Channel Subfamily H Member 8 (*KCNH8*) gene participates in voltage-gated potassium channel activity. Altered expression of this gene in myometrial smooth muscle cells may result in aberrant uterine activity resulting in sPTB.⁴⁷

Gasdermin-B (*GSDMB*) gene triggers pyroptosis, an inflammatory type of apoptosis.²⁶ Pyroptosis has been previously reported to occur in women who experienced spontaneous preterm labour.⁴⁸

The encoded protein of Zinc Finger Protein 608 (*ZNF608*) gene is a transcription factor which represses the transcription of *ZNF609*, and thus represses transcription of Recombination Activating Gene 1 (*RAG1*).²⁶ In the presence of rs13180906-C, as the expression of *ZNF608* increases in blood, the expression of *RAG1* might be decreased. Previous study has reported that *Rag1* knockout mice were deficient in both T cells and B cells and had enhanced risk of preterm delivery.⁴⁹

Moreover, enrichment analysis also revealed similar biological processes that were identified by dissecting the biological relevance of the above-mentioned individual genes, which added further value to this study. Additionally, several other relevant biological processes were identified viz. calcium-release channel activity, negative regulation of circadian sleep/wake cycle, chaperone-mediated autophagy, cellular responses to molecule of fungal origin, magnesium ion, prostaglandin E stimulus, and caffeine (Appendix Table S6) which merits further attention.

Our haplotype-based association analysis revealed that haplotypes containing six validated SNPs (rs4798499, rs13011430, rs8179838, rs7629800,

rs13180906, and rs2327290) were significantly associated with sPTB. This suggests that the association of these SNPs is strong enough to influence the association of their respective haplotypes with sPTB.

The main strength of our study is a well-phenotyped cohort. Being a prospective single-centre cohort with ultrasound-based gestational age estimation done in early pregnancy, the enrolled women were stringently phenotyped into preterm and term delivering women. Moreover, we ensured the quality of genotype data by using strict thresholds. Regarding limitations of the study, although the Infinium Global Screening Array with multi-disease drop-in panel is the most relevant array for genome-wide genotyping of South Asian population (as it contains polymorphic markers from 26 populations of 1000 Genomes Phase III), 12.38% of the markers present in the array were found to be monomorphic in our study population and 40.31% of the markers were found to be rare variants, i.e. have minor allele frequency <5%. This resulted in a reduction of the total number of available loci for association testing. Moreover, because of lack of samples from the same ethnic background, we were unable to replicate our results. However, we have undertaken a comparative evaluation of the associated SNPs with a previously published study,¹² in which the participants were of a different ancestry.

Overall, our findings provided evidence of the genetic heterogeneity of sPTB in Indian women. Identification of both trans-ethnic and novel genetic loci in the cohort might reflect the fundamental or predisposing nature of the underlying biological processes. Further functional investigations on the modulation of gene expression under the presence of different genotypes might provide new insights into biological mechanism of sPTB. From public health perspective, it would be interesting to evaluate the ability of these genomic markers to predict sPTB. If they are found predictive either by themselves, or in combination with other clinical and biological predictors, they will add value to a risk stratification algorithm to be used in clinical settings. Such a tool will enable early triaging of at-risk women to appropriate level of antenatal medical care. The identification of significant trans-ethnic association of four SNPs (rs35760881-A, rs17307967-T, rs4308815-C, and rs10983507-G) with sPTB in women belonging to India and European ancestral population is interesting, and further investigation is warranted to understand its clinical relevance globally.

Contributors

EB analysed and interpreted the data and was a major contributor in writing the manuscript. Ayushi and RT collated the clinical data of the study participants. CD prepared the bootstrapped datasets for identification of significance threshold. SB, PPM, AM, RT and UCN conceptualized the study. AM, PPM, SB, and RT substantively revised the manuscript. PK, RT, and NW assisted in biospecimen storage, identification, and shipment. GARBH-Ini Team developed and implemented

the clinical data collection methods, and data management procedures in GARBH-Ini cohort. PPM and AM have accessed the data and verified the underlying data reported in the manuscript. All authors confirm that they have full access to all the data in the study and accept responsibility to submit for publication. All authors read and approved the final manuscript.

Data sharing statement

GWAS summary statistics generated for 233,748 SNPs (both genotyped and imputed) has been made available at GWAS Catalog (Accession ID: GCST90179179). Custom cluster file (.egt file) generated for Infinium Global Screening Array version 3.0 with multi-disease drop-in panel would be shared upon request to the corresponding authors.

Declaration of interests

None.

Acknowledgments

This work was funded by the Department of Biotechnology, Ministry of Science and Technology, Govt. of India (BT/PR9983/MED/97/194/2013) and for some components of the biorepository by the Grand Challenges India – All Children Thriving Program (supported by the Programme Management Unit), Biotechnology Industry Research Assistance Council (BIRAC/GCI/0114/03/14–ACT). PPM acknowledges support of his National Science Chair fellowship from SERB, Government of India. EB is supported by the Senior Research Fellowship (NET) of the Council of Scientific and Industrial Research (CSIR-HRDG), Govt. of India.

Professor M. K. Bhan will always be remembered reverently for his critical scientific and technical feedback. The authors recognize the efforts of the personnel of the GARBH-Ini Cohort. The authors are thankful to Mr. Sukdev Sinha, Dr. Anamika Gambhir and Dr. Shirshendu Mukherjee for their technical feedback and support. We also thank the assistance of National Genomics Core facility in NIBMG, India, specially Mr. Indranil Bagchi, Mr Arup Ghosh, Mr. Shekhar Ghosh and Dr. Subrata Patra, for conducting genotyping and sequencing. We thank the members of NIBMG for providing valuable suggestions in the generation and validation of custom cluster file (Dr. Chandrika Bhattacharyya, Ms. Vijay Laxmi Roy, Dr. Analabha Basu) and providing the sequence information of the individuals who were used for the validation of custom cluster file (Mr. Arnab Ghosh, Dr. Nidhan K. Biswas). EB acknowledges the scientific advises from her scientific advisory committee members (Dr. Samsiddhi Bhattacharjee and Professor Sanghamitra Sengupta). We are grateful to Dr. Louis J. Muglia for providing us the list of top 10,000 variants obtained from the GWAS of preterm birth and gestational age, conducted in women belonging to European ancestry.

GARBH-Ini Team members:

Vineeta Bal^a, Shinjini Bhatnagar^a, Bhabatosh Das^a, Babu Koundinya Desiraju^a, Pallavi Kshetrapal^a, Sumit Misra^a, Uma Chandra Mouli Natchu^a, Satyajit Rath^a, Kanika Sachdeva^a, Dharmendra Sharma^a, Amanpreet Singh^a, Shailaja Sopory^a, Ramachandran Thiruvengadam^a, Nitya Wadhwa^a, Arindam Maitra^b, Partha P Majumder^b, Tushar K Maiti^c, Monika Bahl^d, Shubra Bansal^d, Umesh Mehta^e, Sunita Sharma^e, Brahmdeep Sindhu^f, Sugandha Arya^f, Rekha Bharti^f, Harish Chellani^f, Pratima Mittal^g, Anju Garg^g, Siddharth Ramji^g, Ashok Khurana^h, Reva Tripathi^o, Yashdeep Gupta^p, Smriti Hari^p, Nikhil Tandon^p, Rakesh Gupta^q, Dinakar M Salunke^r, Balakrish G Nair^r, and Gagandeep Kang^s.

^aTranslational Health Science and Technology Institute, NCR Biotech Cluster, Faridabad, Delhi NCR, India-Coordinating Institute, India

^bNational Institute of Biomedical Genomics, Kalyani, India

^cRegional Centre for Biotechnology, NCR Biotech Cluster, Faridabad, Delhi NCR, India

^dClinical Development Services Agency, Translational Health Science and Technology Institute, NCR-Biotech Cluster, Faridabad, Delhi NCR, India

^eGurugram Civil Hospital, Haryana, India

^fSafdarjung Hospital, New Delhi, India

^mMaulana Azad Medical College, New Delhi, India

ⁿThe Ultrasound Lab, Defence Colony, New Delhi, India

^oHamdard Institute of Medical Sciences and Research, Jamia Hamdard University, New Delhi, India

^pIndia Institute of Medical Sciences, New Delhi, India

^qGovernment of Haryana, India

^rInternational Centre for Genetic Engineering and Biotechnology, New Delhi, India

^sRajiv Gandhi Centre for Biotechnology, Trivandrum, India

^tChristian Medical College, Vellore, India

Appendix A. Supplementary data

Supplementary data related to this article can be found at <https://doi.org/10.1016/j.lansea.2023.100190>.

References

- World Health Organization. Preterm birth. <https://www.who.int/news-room/fact-sheets/detail/preterm-birth>; 2018. Accessed November 18, 2022.
- Raju TNK, Buist AS, Blaisdell CJ, Moxey-Mims M, Saigal S. Adults born preterm: a review of general health and system-specific outcomes. *Acta Paediatr.* 2017;106:1409–1437. <https://doi.org/10.1111/apa.13880>.
- Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. *Lancet.* 2008;371:75–84. [https://doi.org/10.1016/S0140-6736\(08\)60074-4](https://doi.org/10.1016/S0140-6736(08)60074-4).
- Clausson B, Lichtenstein P, Cnattingius S. Genetic influences on birthweight and gestational length determined by studies in offspring of twins. *BJOG.* 2000;107:375–381. <https://doi.org/10.1111/j.1471-0528.2000.tb13234.x>.
- Treloar SA, Macones GA, Mitchell LE, Martin NG. Genetic influences on premature parturition in an Australian twin sample. *Twin Res.* 2000;3:80–82. <https://doi.org/10.1375/136905200320565526>.
- Phillips C, Velji Z, Hanly C, Metcalfe A. Risk of recurrent spontaneous preterm birth: a systematic review and meta-analysis. *BMJ Open.* 2017;7:e015402. <https://doi.org/10.1136/bmjopen-2016-015402>.
- Svensson AC, Sandin S, Cnattingius S, et al. Maternal effects for preterm birth: a genetic epidemiologic study of 630,000 families. *Am J Epidemiol.* 2009;170:1365–1372. <https://doi.org/10.1093/aje/kwp328>.
- Plunkett J, Feitosa MF, Trusgnich M, et al. Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. *Hum Hered.* 2009;68:209–219. <https://doi.org/10.1159/000224641>.
- Porter TF, Fraser AM, Hunter CY, Ward RH, Varner MW. The risk of preterm birth across generations. *Obstet Gynecol.* 1997;90:63–67. [https://doi.org/10.1016/S0029-7844\(97\)00215-9](https://doi.org/10.1016/S0029-7844(97)00215-9).
- Boyd HA, Poulsen G, Wohlfahrt J, Murray JC, Feenstra B, Melbye M. Maternal contributions to preterm delivery. *Am J Epidemiol.* 2009;170:1358–1364. <https://doi.org/10.1093/aje/kwp324>.
- Zhang H, Baldwin DA, Bukowski RK, et al. A genome-wide association study of early spontaneous preterm delivery. *Genet Epidemiol.* 2015;39:217–226. <https://doi.org/10.1002/gepi.21887>.
- Zhang G, Feenstra B, Bacelis J, et al. Genetic associations with gestational duration and spontaneous preterm birth. *N Engl J Med.* 2017;377:1156–1167. <https://doi.org/10.1056/NEJMoa1612665>.
- Bhattacharjee E, Maitra A. Spontaneous preterm birth: the underpinnings in the maternal and fetal genomes. *NPJ Genom Med.* 2021;6:43. <https://doi.org/10.1038/s41525-021-00209-5>.
- Bhatnagar S, Majumder PP, Salunke DM. A pregnancy cohort to study multidimensional correlates of preterm birth in India: study design, implementation and baseline characteristics of the participants. *Am J Epidemiol.* 2019;188:621–663. <https://doi.org/10.1093/aje/kwy284>.
- Chang CC, Chow CC, Tellier LCAM, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience.* 2015;4. <https://doi.org/10.1186/s13742-015-0047-8>.
- Howie B, Marchini J, Stephens M. Genotype imputation with thousands of genomes. *G3: GenesGenom Genet.* 2011;1(6):457–470. <https://doi.org/10.1534/g3.111.001198>.
- The 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature.* 2015;526:68–74. <https://doi.org/10.1038/nature15393>.

- 18 GenomeAsia100K Consortium. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature*. 2019;576:106–111. <https://doi.org/10.1038/s41586-019-1793-z>.
- 19 Delaneau O, Zagury JF, Marchini J. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Methods*. 2013;10:5–6. <https://doi.org/10.1038/nmeth.2307>.
- 20 R Core Team. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing; 2020. URL <https://www.R-project.org/>. Accessed December 12, 2021.
- 21 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics*. 2005;21:263–265. <https://doi.org/10.1093/bioinformatics/bth457>.
- 22 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38:e164. <https://doi.org/10.1093/nar/gkq603>.
- 23 Kircher M, Witten DM, Jain P, O’Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46:310–315. <https://doi.org/10.1038/ng.2892>.
- 24 Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017;8:1826. <https://doi.org/10.1038/s41467-017-01261-5>.
- 25 Sarkar S, Godara K, Sarkar S, et al. *Preterm birth is associated with reduction of maternal telomere length during pregnancy*. PREPRINT (Version 1). Research Square; 2021. <https://doi.org/10.21203/rs.3.rs-197790/v1>.
- 26 The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res*. 2021;49:D480–D489. <https://doi.org/10.1093/nar/gkaa1100>.
- 27 Ng PY, Ireland DJ, Keelan JA. Drugs to block cytokine signaling for the prevention and treatment of inflammation-induced preterm birth. *Front Immunol*. 2015;6:166. <https://doi.org/10.3389/fimmu.2015.00166>.
- 28 Gómez-Chávez F, Correa D, Navarrete-Meneses P, Cancino-Díaz JC, Cancino-Díaz ME, Rodríguez-Martínez S. NF-κB and its regulators during pregnancy. *Front Immunol*. 2021;12:679106. <https://doi.org/10.3389/fimmu.2021.679106>.
- 29 Nikoghosyan M, Hakobyan S, Hovhannisyanyan A, Loeffler-Wirth H, Binder H, Arakelyan M. Population levels assessment of the distribution of disease-associated variants with emphasis on Armenians – a machine learning approach. *Front Genet*. 2019;10:394. <https://doi.org/10.3389/fgene.2019.00394>.
- 30 Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome diagram viewer: data structures and strategies to boost performance. *Bioinformatics*. 2018;34:1208–1214. <https://doi.org/10.1093/bioinformatics/btx752>.
- 31 Mistry HD, Broughton PF, Redman CW, Poston L. Selenium in reproductive health. *Am J Obstet Gynecol*. 2012;206:21–30. <https://doi.org/10.1016/j.ajog.2011.07.034>.
- 32 Monangi N, Xu H, Khanam R, et al. Association of maternal prenatal selenium concentration and preterm birth: a multicountry meta-analysis. *BMJ Glob Health*. 2021;6:e005856. <https://doi.org/10.1136/bmjgh-2021-005856>.
- 33 Cha JM, Aronoff DM. A role for cellular senescence in birth timing. *Cell Cycle*. 2017;16:2023–2031. <https://doi.org/10.1080/15384101.2017.1371888>.
- 34 Page RL, Han G, Akinlotan M, Patron MP, Gandhi H, Kochan KJ. Telomere length and preterm birth in pregnant Mexican-origin women. *Matern Child Health J*. 2021;25:1798–1805. <https://doi.org/10.1007/s10995-021-03209-0>.
- 35 Fujita N, Jaye DL, Kajita M, Geigerman C, Moreno CS, Wade PA. MTA3, a Mi-2/NuRD complex subunit, regulates an invasive growth pathway in breast cancer. *Cell*. 2003;113:207–219. [https://doi.org/10.1016/S0092-8674\(03\)00234-4](https://doi.org/10.1016/S0092-8674(03)00234-4).
- 36 Zhang H, Singh RR, Talukder AH, Kumar R. Metastatic tumor antigen 3 is a direct corepressor of the Wnt4 pathway. *Genes Dev*. 2006;20:2943–2948. <https://doi.org/10.1101/gad.1461706>.
- 37 Nold C, Jensen T, O’Hara K, Stone J, Yellon SM, Vella AT. Replens prevents preterm birth by decreasing type I interferon strengthening the cervical epithelial barrier. *Am J Reprod Immunol*. 2020;83:e13192. <https://doi.org/10.1111/aji.13192>.
- 38 Li Q, Kannan A, Das A, et al. WNT4 acts downstream of BMP2 and functions via β-catenin signaling pathway to regulate human endometrial stromal cell differentiation. *Endocrinology*. 2013;154:446–457. <https://doi.org/10.1210/en.2012-1585>.
- 39 Majovic JE, Olson DM. The physiology of human parturition. In: Bittar EE, Zakar T, eds. *Advances in organ biology*. Elsevier; 1996;1:89–119. [https://doi.org/10.1016/S1569-2590\(08\)60069-5](https://doi.org/10.1016/S1569-2590(08)60069-5).
- 40 Yeung CY, Taylor SH, Garva R, et al. Arhgap28 is a RhoGAP that inactivates RhoA and downregulates stress fibers. *PLoS One*. 2014;9:e107036. <https://doi.org/10.1371/journal.pone.0107036>.
- 41 Lartey J, Smith M, Pawade J, Strachan B, Mellor H, López Bernal A. Up-regulation of myometrial RHO effector proteins (PKN1 and DIAPH1) and CPI-17 (PIR14A) phosphorylation in human pregnancy is associated with increased GTP-RHOA in spontaneous preterm labor. *Biol Reprod*. 2007;76:971–982. <https://doi.org/10.1095/biolreprod.106.058982>.
- 42 Gomez-Lopez N, StLouis D, Lehr MA, Sanchez-Rodriguez EN, Arenas-Hernandez M. Immune cells in term and preterm labor. *Cell Mol Immunol*. 2014;11:571–581. <https://doi.org/10.1038/cmi.2014.46>.
- 43 Brockway HM, Kallapur SG, Buhimschi IA, et al. Unique transcriptomic landscapes identified in idiopathic spontaneous and infection related preterm births compared to normal term births. *PLoS One*. 2019;14:e0225062. <https://doi.org/10.1371/journal.pone.0225062>.
- 44 Karahoda R, Robles M, Marushka J, et al. Prenatal inflammation as a link between placental expression signature of tryptophan metabolism and preterm birth. *Hum Mol Genet*. 2021;30:2053–2067. <https://doi.org/10.1093/hmg/ddab169>.
- 45 Vanova KH, Kraus M, Neuzil J, Rohlena J. Mitochondrial complex II and reactive oxygen species in disease and therapy. *Redox Rep*. 2020;25:26–32. <https://doi.org/10.1080/13510002.2020.1752002>.
- 46 Menon R. Oxidative stress damage as a detrimental factor in preterm birth pathology. *Front Immunol*. 2014;5:567. <https://doi.org/10.3389/fimmu.2014.00567>.
- 47 Brainard AM, Korovkina VP, England SK. Potassium channels and uterine function. *Semin Cell Dev Biol*. 2007;18:332–339. <https://doi.org/10.1016/j.semdev.2007.05.008>.
- 48 Gomez-Lopez N, Romero R, Tarca AL, et al. Gasdermin D: Evidence of pyroptosis in spontaneous preterm labor with sterile intra-amniotic inflammation or intra-amniotic infection. *Am J Reprod Immunol*. 2019;82:e13184. <https://doi.org/10.1111/aji.13184>.
- 49 Bizargity P, Del Rio R, Philippe M, Teuscher C, Bonney EA. Resistance to lipopolysaccharide-induced preterm delivery mediated by regulatory T cell function in mice. *Biol Reprod*. 2009;80:874–878. <https://doi.org/10.1095/biolreprod.108.074294>.