

S1 Table. Properties of the phenotype and genotype data and data handling of the meta-analysis cohorts.

	Phenotype data			Genotype data				
	Source	Definition of preterm birth	Exclusion criteria	Quality control prior to GWAS	Data transformation	GWAS Method	GWAS covariates	Quality control after GWAS
<b>FinnGenR6</b>	Register-based	Based on ICD-codes and gestational duration. ICD8, 63497 (Partus praematurus imminens, alibi non classificabilis, at 29-37 weeks of gestation); ICD9, 644A (partus praematurus/imminens), individuals with term births (>37wk) excluded; ICD9, 644B (partus praematurus, at 29-37 weeks of gestation); ICD10, O60 (preterm birth, delivery started <37 weeks of gestation)	Multiple gestation or birth, pre- or eclampsia, polyhydroamnios	Sample exclusion criteria: ambiguous sex, non-Finnish ancestry, missingness >5%, excess heterozygosity ( $\chi^2/4SD$ ); Variant exclusion criteria: missingness >2%, minor allele count (MAC)<3, deviation from HWE ( $p < 1 \times 10^{-8}$ ), imputation INFO <0.7	Gestational duration were inverse-normalized	Linear / logistic regression with Scalable and Accurate Implementation of Generalized mixed model, SAIGE, MAC=5	Maternal age, genotyping batch, the 10 leading principal components	Variant exclusions: MAF<0.01%
<b>Northern/Central Finland</b>	Register-based	Based on gestational duration: Preterm, <36 wk + 1d of gestation; term > 38-41 wk + 6 d of gestation	Multiple gestation or birth, pre- or eclampsia, polyhydroamnios, intrauterine growth restriction, placental abruption, anomalies of the fetus, clinical chorioamnionitis or acute septic infection in the mother, alcohol or narcotic use, accidents	Sample exclusion criteria: high missingness (>10%), uncertain sex; Variant exclusion criteria: minor allele frequency (MAF)<1%, deviation from HWE ( $p < 1 \times 10^{-8}$ ), genotyping rate <90%, imputation INFO <0.7	Gestational duration were mean centered, scaled, and quantile normalized	Frequentist association test / quantitative trait test with SNPTTESTV2	The top three MDS (multidimensional scaling) components	Variant exclusions: MAF<1%, SNPTTEST info<0.7
<b>23andMe</b>	Self-reported	Based on gestational duration: Preterm, <37 wk of gestation; term >37wk of gestation	Medical indication for preterm delivery	Sample exclusion criteria: <97% of European ancestry; Variant exclusion criteria: deviation from HWE ( $p < 1 \times 10^{-8}$ ), call rate <95%, allele frequency discrepancy with 1000Genomes Europeans	None	Linear / logistic regression	Maternal age and the top five principal components	Variant exclusions: MAF<0.01%