Annals of
human genetics WILEY

Recent advances in the genetics of preterm birth

Megan Wadon[1] (iD) | Neena Modi[2] | Hilary S. Wong[3] | Anita Thapar[1]* |
Michael C. O'Donovan[1]*

[1]MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Cardiff, Wales

[2]Section of Neonatal Medicine, Department of Medicine, Chelsea and Westminster Hospital Campus, Imperial College, London, United Kingdom

[3]Department of Paediatrics, University of Cambridge, Cambridge, United Kingdom

Correspondence
Michael C. O'Donovan, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff, Wales, CF24 4HQ.
Email: odonovanmc@cardiff.ac.uk

*These authors contributed equally to this work.

Abstract

Preterm birth is associated with short- and long-term impairments affecting physical, cognitive, and neuropsychiatric health. These sequelae, together with a rising preterm birth rate and increased survival, make prematurity a growing public health issue because of the increased number of individuals with impaired health throughout the life span. Although a major contribution to preterm birth comes from environmental factors, it is also modestly heritable. Little is known about the architecture of this genetic contribution. Studies of common and of rare genetic variation have had limited power, but recent findings implicate variation in both the maternal and fetal genome. There is some evidence risk alleles in mothers may be enriched for processes related to immunity and inflammation, and in the preterm infant, processes related to brain development. Overall genomic discoveries for preterm birth lag behind progress for many other multifactorial diseases and traits. Investigations focusing on gene–environment interactions may also provide insights, but these studies still have a number of limitations. Adequately sized genetic studies of preterm birth are a priority for the future especially given the breadth of its negative health impacts across the life span and the current interest in newborn genome sequencing.

KEYWORDS

genetics, gestational age, preterm birth

1 | INTRODUCTION

Birth is defined as preterm if it occurs before 37 weeks of gestation, and accounts for ∼60,000 births in the United Kingdom every year (Blencowe et al., 2012). It is further categorized as very preterm (birth before 32 weeks gestation) and extreme preterm (birth before 28 weeks gestation). Preterm birth can also be categorized into medically indicated and spontaneous. The former is where labor is induced or a Caesarean section is performed because of medical risks or complications, the latter where onset of labor occurs spontaneously prior to 37 weeks gestation. However, this distinction fails to take into

consideration the possibility that medical conditions necessitating delivery by medical intervention has factors in common with "spontaneous" preterm birth.

Globally, in 2010, preterm births were estimated to account for ∼11.1% of all live births and evidence suggests this rate is increasing (Blencowe et al., 2012). The survival of very preterm infants has improved steadily over recent years and is now over 90% in high-income settings (Blencowe et al., 2012; Santhakumaran et al., 2018). However, globally, complications related to preterm birth remain a major cause of neonatal and under-5 mortality (Liu et al., 2015). Preterm individuals also are at substantially higher risk of

impairments affecting multiple organ systems (Haraldsdottir et al., 2018; Huang, Lin, Wang, Su, & Lin, 2018; Parkinson, Hyde, Gale, Santhakumaran, & Modi, 2013), including potentially disabling neurological and neuropsychiatric disorders (Agrawal, Rao, Bulsara, & Patole, 2018; Franz et al., 2018; Heuvelman et al., 2018; Nosarti et al., 2012; Sutton & Darmstadt, 2013). This highlights the importance of determining the mechanisms behind preterm birth and how these mechanisms might relate to the difficulties faced later in life. Although environmental factors are known to contribute to preterm birth risk, there is growing evidence that highlight the contribution of genetic factors. In this article, we aim to provide a brief review of genetic findings for preterm birth and highlight issues that need to be addressed in future research.

2 | TWIN AND FAMILY STUDIES

Preterm birth runs in families. The risk of preterm birth for women who have a mother or a sister with a history of having delivered preterm is higher than that observed in the general population (Sherf et al., 2017). There have been a number of twin studies, all of which find that there is a significant, genetic contribution to preterm birth. As is typical of this type of study, observed heritability estimates have varied (Table 1), ranging from approximately 15% (Lunde, Melve, Gjessing, Skjaerven, & Irgens, 2007; Wu et al., 2015) up to 30%–40% (Clausson, Lichtenstein, & Cnattingius, 2000; Kistka et al., 2008; Plunkett et al., 2009) for the maternal genetic contribution. These estimates should be interpreted with caution, as they may not account for potential confounding; for example, an observed maternal genetic effect may also include the effects of the fetal genome (York, Eaves, Neale, & Strauss, 2014) as well as lifestyle factors that are socially transmitted from mothers to daughters (Svensson et al., 2009). A large Swedish registry-based ($n = 244,00$) extended twin-sibling design study showed that fetal genetic factors contributed 13.1% of the genetic variation to gestational age at delivery (York et al., 2013). This study was important in also showing that the largest contribution came from environmental factors, specifically pregnancy-specific factors (56%). However, another large study of 630,000 families observed minimal fetal contribution, while maternal genetic factors explained 25% of the variation in preterm birth (Svensson et al., 2009). Paternal effects have been less commonly investigated, with either negligible effects (Kistka et al., 2008) or a small paternal contribution (6%) (Wu et al., 2015) being found. Consistent with a paternal contribution, it has been reported that maternally reported ancestry, including paternal ancestry, modifies risk for preterm birth (Palomar, DeFranco, Lee, Allsworth, & Muglia, 2007).

One problem is that studies have used inconsistent definitions of preterm birth. For example, Lunde et al. (2007) (Table 1) excluded births before 35 weeks gestation and investigated gestational age as a quantitative trait, and others have used a categorical variable of preterm birth (<37 weeks) (Treloar, Macones, Mitchell, & Martin, 2000). Also, some studies have not differentiated between spontaneous and medically indicated preterm birth, and others have focused on spontaneous preterm birth. This may be relevant as in the large study of Svensson et al. (2009), for the medically indicated preterm birth group, the maternal genetic contribution was half that (13%) of the spontaneous group (27%). The fetal genetic contribution was 14% in the medically indicated group compared with none in the spontaneous group, although it should be noted the confidence around this estimate of 14% included 0%. Overall, despite the differences in studies, findings suggest estimates of heritability that are modest for the maternal genome, and considerably weaker, possibly negligable, from the fetal genome.

3 | GENOME-WIDE ASSOCIATION STUDIES (GWAS): SEARCHING FOR COMMON GENE VARIANTS

3.1 | Maternal GWAS

Effect sizes are expected to be particularly weak for common alleles that influence liability to a phenotype, which is likely to be subject to negative selection, probable for preterm birth given its associated infant mortality and morbidity. Unsurprisingly then, given their small sample sizes, the first few GWASs based on the maternal genome ($n = 1,881$ [Zhang et al., 2015]; $n = 3022$ [Myking et al., 2013]) did not yield significant associations.

The largest maternal GWAS to date included 43,568 mothers of European ancestry of whom 3,331 reported a gestational duration <37 weeks for their first live singleton birth where onset of labor was spontaneous (Zhang et al., 2017). Promising findings were further tested in a sample of 8,643 individuals (2,565 preterm). In the combined sample, six genetic loci were significantly associated with gestational duration, of which three were associated with preterm birth. The same loci were also found in the fetal genome; however, the association effect sizes were around half that of those conferred by the maternal genotypes. This suggests the associations are driven by the maternal genome (the allele frequencies of the fetus being passively elevated as a consequence of the elevated allele frequencies in the mothers). These findings suggest that, at least in part, from the perspective of genetic liability, preterm birth represents the tail of a continuous trait of gestational age. It is the case that complex phenotypes in general behave as quantitative traits, but it cannot

TABLE 1 Twin studies of gestational age and preterm birth

Heritability estimate	Authors	Study type	Variable measured
36% (maternal)	Clausson et al., 2000	Twin study investigating concordance and correlation between twin mothers of differing zygosity (potential confounding from fetal genome)	Gestational age (no distinction between spontaneous and medically indicated onset of labor)
14% (maternal) 11% (fetal)	Lunde et al., 2007	Parent–offspring study using path analysis and maximum likelihood principles and Pearson's correlation coefficients for different familial relationships	Gestational age (births by Caesarean section were excluded as these births are more likely to be medically indicated)
34% (maternal) Negligible (paternal)	Kistka et al., 2008	Extended twin design investigating the correlations between monozygotic twins and first-degree relatives reporting on the gestational age of their first-born child (potential confounding from fetal genome)	Gestational age (no distinction between spontaneous or medically indicated onset of labor)
44% (maternal) 33% (fetal)	Plunkett et al., 2009	Family study with mother–infant pairs using segregation analysis and estimating the parameters using maximum likelihood method (potential confounding between maternal and fetal estimates)	Gestational age (spontaneous onset of labor)
27% (maternal)	Treloar et al., 2000	Twin study (potential confounding from fetal genome)	Any preterm birth (no distinction between spontaneous or medically indicated onset of labor)
20.6% (maternal) 13.1% (fetal)	York et al., 2013	Extended twin-sibling design investigating the offspring of twins, full siblings, and half-siblings using structural equation modeling. (excluded multifetal pregnancies)	Gestational age (exclusions for some medical reasons including preeclampsia and placental abruption)
25% (maternal) (29% for spontaneous; 13% for indicated) 5% (fetal) (negligible for spontaneous; 14% for indicated)	Svensson et al., 2009	Family study investigating children of siblings using an alternating logistic regression model	Preterm birth (live birth at <37 weeks gestation) (data were analyzed as a combined group including spontaneous and medically indicated onset of labor and also as two separate groups)
14.21% (maternal) 6.77% (paternal) 13.33% (narrow-sense) 24.45% (broad-sense)	Wu et al., 2015	Family study assessing the correlation of gestational age between offspring and full/half siblings and a regression analysis in parent offspring pairs; narrow-sense heritability was calculated as 4 times the paternal half-sib correlation and broad-sense heritability was calculated as the narrow sense heritability plus the full sibling correlation (representing the dominance variance) (not able to control for environmental risk factors)	Gestational age (excluded births that were unlikely to have a spontaneous onset of labor)

be taken as certain that all alleles that influence those traits are efficiently observed in general populations reflecting the full range of trait outcomes. In the neurodevelopmental field, for example, there are mutations of large effects but whose frequencies are so rare that they are essentially not observed in unselected samples of the size that it is currently feasible to study. Similar considerations may hold for very preterm births where it is possible that there may be some variants, for example, highly penetrant deleterious de novo mutations, which may have very large effects on fetal development and the age

TABLE 2   Genome-wide association studies of preterm birth

Authors	Category of preterm birth	Sample size	Limitation	Concerns
Zhang et al., 2015	<34 weeks gestation (spontaneous onset of labor)	1,851 (916 cases)	Small sample size	Reduces power
			Inclusion of highly heterogeneous groups (in terms of ancestry)	Further reduces power
			Lack of stringent control for background ancestry	Increases false-positive rate
			Lack of support from markers in the same chromosomal regions	Increases probability of genotyping errors
			Failure of replication in independent samples	Suggests unreliable results
Rapport et al. 2018	Birth between 25 and 30 weeks gestation (spontaneous onset of labor)	13,944 (1,349 cases) split into five ancestral groups EUR 9,890 (260 cases) AFR 1874 (190 cases) AMR 1847 (745 cases) EAS 249 (131 cases) SAS 59 (23 cases)	Each association was only observed in one of the five samples	
			Unable to obtain replication in any external samples	Suggests unreliable results
			Samples from which the associations were seen were small (AFR N preterm = 190, AMR N preterm = 745)	Reduces power
			Each subsample was genetically heterogeneous, controlling for ancestry challenging	Further reduces power
			Cases and controls were taken from different populations	Possible genotyping batch effects

at birth, such that are also only observed in the very preterm range.

As is typical for complex phenotypes that are under negative selection, the effect sizes of the common GWAS associations are small and explain only a trivial fraction of the heritability of preterm birth. Nevertheless, the ability to find preterm-associated alleles in what is still a relatively small (in GWAS terms) study suggests that the GWAS approach is likely to yield many more findings as sample sizes increase.

3.2 | Fetal/offspring GWAS

The first GWAS based on the fetal genome of spontaneous preterm birth (<37 weeks gestation) failed to identify any associated loci (Myking et al., 2013). The second GWAS of the fetal genome investigated spontaneous preterm birth <34 weeks gestation (Zhang et al., 2015); two variants achieved genome-wide significance. However, with a sample of only 916 cases (total neonatal sample size 1,851), as well as other limitations (Table 2), the initial associations are at best highly preliminary (as recognized by the authors) and, we suggest, more likely to be spurious rather than true associations.

A larger GWAS investigated spontaneous preterm birth between 25 and 30 weeks gestation in a total sample size of 13,944, of whom 1,349 met the preterm criterion, split into five ancestry groups (Rappoport et al., 2018). The authors reported two significant associations. Again, these findings have to be viewed with caution for many of the reasons outlined above (Table 2), including each association only being seen in one of the five small samples (African ancestry sample N preterm birth = 190, American sample 745). One interpretation of the findings is that it suggests different alleles exist in people of different ancestries, and this may partly explain differences in preterm birth rates seen in different populations. However, this should also be viewed with caution. First, the individual findings in each population are not yet robust. Second, it is well known that in underpowered studies such as the subgroups here represented, even in the face of genetic homogeneity, sampling variance makes replication across studies unlikely. Nevertheless, should more secure findings of differences in the architectures across ancestral (or indeed ethnic) groups be identified, it will be important to investigate interactions with variable environmental exposures that also differ if the findings are to translate to reducing the disparity in preterm births (Burris et al., 2019).

The largest published GWAS to date included 84,689 children; 4,775 were preterm of whom 1,139 of these were considered early preterm infants (<34 weeks) (Liu et al., 2019). Infants were excluded if their birth was medically indicated.

The study reported association between markers in the fetal genome at 2q13 and gestational duration, and, moreover, the allele that was associated with increased gestation was also associated with postterm birth greater than 42 weeks. Overall, 7.3% of the variance in gestation duration was explained by common fetal genetic variants. There were no replicated associations with either preterm birth or early preterm birth and, interestingly, the allele associated with gestational age was reported to show no association with preterm or early preterm birth. This suggests the genetic variants that influence the early extreme of gestational age may differ from those that influence gestational age in general, although the lack of association may also reflect the incomplete power of the study.

To summarize the GWAS findings for preterm birth, thus far studies have, in general, been underpowered. Nevertheless, they strongly suggest that common alleles influencing preterm birth with even modest (odds ratio [OR] > 1.5) effect sizes do not exist, but as shown by Zhang et al. (2017), alleles with smaller effects do exist and they can be captured by larger GWASs. It is likely that large-scale population studies of gestational age can uncover some of the heritability attributable to common variation, although it is not yet clear whether at the extremes of preterm birth, the trait shows some qualitative as well as quantitative differences from gestational age.

Sample sizes lag behind GWASs of many complex diseases and traits, for example, increasing the sample size to include 36,989 cases and 113,075 controls when using GWAS to investigate schizophrenia resulted in the discovery of 83 novel associated loci (out of 108 associated loci found) (Schizophrenia Working Group of the Psychiatric Genomics Consortium, 2014), a pattern that also holds true for other traits (Warren et al., 2017). These common gene variants only partially explain observed twin heritability and it is also important to realize that extensive correlations between genotypes at associated loci means the index single-nucleotide polymorphisms identified by GWAS may not necessarily be those that directly contribute to risk. However, the hope is that common gene variant discovery ultimately will provide valuable insights into biological mechanisms, and thereby identify potential novel therapeutic targets, and inform risk prediction and disease stratification. For example, there is growing interest in whether polygenic risk scores could add value to current risk-prediction models used for disease onset prediction (e.g., coronary artery disease) and be used in combination with family history and clinical variables for predicting prognosis as well as disease onset (Lambert, Abraham, & Inouye, 2019).

Meta-analysis of large population-based or case-register cohorts of hundreds of thousands (or more) will likely provide sufficient power to identify hundreds of GWAS loci associated with gestational age. However, for assuring the validity of the findings for clinically relevant phenotypes, particularly very preterm birth (as well as for identifying the contribution of rare alleles of large effect), large-scale (tens of thousands) targeted recruitment from clinics is likely to be needed.

4 $\mid$ RARE VARIANTS

4.1 $\mid$ Copy number variants

Some investigators have studied the role of copy number variants (CNVs) in preterm birth, but few have taken a whole genome perspective, focusing instead on specific candidate CNVs. The single study we are aware of that has taken a genome-wide approach found no specific CNV associations in the genomes of 454 mothers giving birth before 34 weeks (most of whom had spontaneous onset of labor) compared with 1,018 mothers who delivered at term, nor was there any difference in overall CNV burden between the groups (Uzun et al., 2016a).

The most intensively studied candidate region for CNVs has spanned *glutathione-S-transferase (GST)* genes, involved in the metabolism of toxins, including a maternal deletion CNV of approximately 28 kb at the *glutathione-S-transferase theta 1 (GSTT1)* in spontaneous preterm birth at <37 weeks gestation (Zheng et al., 2013). Meta-analysis of nine studies (including 2,526 cases and 4,565 controls) found evidence for association (Liu, Tang, Chen, & Huang, 2014), but the deletion CNV at *GSTT1* is common (15.8%) and if genuinely associated, confers a weak effect size in the typical common variant range (OR ~1.18). Another study of 231 mothers of infants born preterm and 378 control mothers (188 infants born preterm and 391 infants born at term) observed a trend for association with maternal insertion at *glutathione-S-transferase Mu 1 (GSTM1)* and protection against preterm birth, but the primary driver of the association appeared to be acting via the fetal genome (Bustamante et al., 2012). These results support previous findings that suggest maternal *GSTM1-/GSTT1-*(null) genotype increases the risk for preterm birth (Mustafa et al., 2010), but with major caveats. It is unclear whether the statistical evidence (association statistics were not reported in the meta-analysis) meets the standard of genome-wide significance that is now widely required in the field of human genetics, even for candidate genes, but with the reported sample and effect sizes, this seems unlikely. Further (larger) stringently conducted studies are required to definitively answer whether CNVs at these loci confer liability to preterm birth, and if so, whether this is a maternal or fetal effect.

4.2 $\mid$ Whole-exome and -genome sequencing

To date, sequencing studies of preterm births have been small, typically focusing on multiplex families in which more than one woman has had a child born prematurely, the premise

being whether such families might be segregating a rare allele with large effect.

4.2.1 | Maternal rare coding and noncoding variants

McElroy and colleagues performed exome sequencing on a small sample (10 cases including two mother–daughter pairs) to select variants to genotype in a larger sample (237 cases of spontaneous preterm birth, 328 controls) (McElroy et al., 2013). No genome-wide significant findings emerged. Uzun et al. (2016b) in a study based on 32 women who had given birth preterm (with a family history of preterm birth) and 16 controls sequenced 329 candidate genes but again observed no significant associations with preterm birth. While the focus on candidate genes in these studies may be considered pragmatic given the sample sizes, the failure of the candidate gene approach for other complex phenotypes suggests that in most cases, the choice of candidates is not much better than random for identifying associated genes. It is therefore widely considered that regardless of the number of genes sequenced, genome-wide significant thresholds are still appropriate.

A recent whole-exome study was conducted in 17 mothers from seven Finnish multiplex families, 13 of whom had experienced spontaneous preterm birth, and 10 of them had experienced recurrent spontaneous preterm birth. A further 93 sister pairs and two sister triads were used as a "replication population." A number of biological pathways were identified that were enriched for genes containing rare variants predicted to be damaging (Huusko et al., 2018). One of the main drivers of pathway enrichment was the gene *HSPA1L* (encoding *heat shock protein family A member 1 like*). The authors next examined imputed alleles at this locus in a GWAS dataset of over 40,000 mothers (approximately 3,300 cases); of two variants that were testable in that dataset, one was nominally associated ($P = 0.0022$). Although some evidence was provided that the associated allele might confer differential endometrial fibroblast responsiveness to glucocorticoids, the associations did not replicate in two other (small) datasets and use of imputed data for alleles of very low frequency is not necessarily reliable. It will be necessary for this gene to obtain much stronger support before it can be decisively considered to be involved in preterm birth.

4.2.2 | Fetal rare coding and noncoding variants

Modi and colleagues recently presented a series of studies utilizing exome sequencing of the fetal genome to identify rare variants in a number of candidate genes in a small (but variable) number of cases of preterm premature rupture of membranes (PPROM), a subcategory of spontaneous preterm birth, as well as a small (variable) number of normal term control mothers. They reported variants in genes involved in the structural integrity of fetal membranes (Modi et al., 2017a), innate immunity (Modi et al., 2017b), and other pathways (Modi et al., 2018) but in none of the studies was a clear relationship established between PPROM and the genetic variants observed.

The final study we consider here adopted a whole-genome sequencing approach (Li, Oehlert, Snyder, Stevenson, & Shaw, 2017). Rather than focus on cases and controls, the authors used 816 parent-proband trios (295 with preterm birth) to identify de novo genetic variants. While this study was underpowered to detect specifically associated genes and did not distinguish between medically indicated and spontaneous preterm birth, it did provide evidence that those born preterm have an increased rate of de novo variants at a genome-wide level. Also when damaging de novo variants were observed in those born preterm, they were more likely to occur in genes that do not tolerate mutations well, and that are expressed in early fetal brain development (Li et al., 2017). While not definitive, this study does provide evidence that suggests further investigation of rare de novo genetic variants in preterm birth will be important.

5 | GENE–ENVIRONMENT INTERACTION

Environmental as well as genetic factors contribute to preterm birth, and this has led some to look for gene-environment interaction. Examples include possible relationships between candidate genes involved in metabolic detoxification and exposure to pollution (Suh et al., 2008), organochlorine pesticides (Mustafa, Banerjee, Ahmed, Tripathi, & Guleria, 2013), benzene, a chemical found in cigarette smoke (Wang et al., 2000), and maternal smoking (Tsai et al., 2008). Nominally significant reports of interactions have been reported in these studies (some distinguishing between spontaneous and indicated preterm birth and some not) but the sample sizes are too small to generate robust results. The studies have not controlled for genome-wide testing, the underpinning candidate genes have often had larger effect sizes than seem feasible based on the larger studies discussed above, and the findings are as yet unsubstantiated by replication.

There is only one study we are aware of that has taken a genome-wide approach to investigate gene–environment interactions in preterm birth investigating spontaneous preterm birth separately and combined with medically indicated preterm birth. This analysis ($n = 1,733$, including 698 cases) based on African-American women reported genome-wide significant interactions between variants in *COL24A1* and pre-pregnancy body mass index on risk of preterm birth; this finding was subsequently replicated in 300 African-American women (Hong et al., 2017). It is important to note that studying gene–environment interactions poses many

challenges above and beyond straightforward GWAS (Hunter, 2005), including seemingly simple matters such as how to define and model interaction, practical issues such as larger sample size requirements than simple gene discovery, and design issues including the possibility of introducing confounding. As has been the case in complex disorders more widely, the potential for false-positive findings is high and therefore even the most promising of the current findings should be viewed with caution (Boffetta et al., 2012).

6 | CONCLUSIONS

Preterm birth is a major public health issue, with individuals at substantially higher risk of a wide range of adverse health outcomes that extend throughout the life-course and into the next generation. It is a multifactorial outcome that is the result of multiple environmental and genetic (maternal and fetal) risk factors. Gene discovery for preterm birth is lagging behind many other multifactorial disorders, most likely because GWASs of gestational age and preterm birth have been relatively small. However, emerging evidence suggests there are a large number of alleles involved, each contributing to a small amount of population liability. Sample sizes also still need to increase to detect likely causal de novo and inherited rare genetic variants. Preliminary investigations suggest rare, deleterious genetic variants could be enriched in both mothers and offspring, although no robustly associated variants have emerged. Gene–environment studies may provide further insight, but the current findings need rigorous testing. Advancements in statistical methods, such as polygenic risk scoring may also help with prediction of preterm birth, enabling women and prospective parents to make informed lifestyle choices (Torkamani, Wineinger, & Topol, 2018). However, so far, polygenic risk scores are only weakly predictive for many diseases, and not at all predictive for preterm birth. Given the growing global burden of preterm birth, much larger genetic studies are of relevance and importance, and collaborations across different groups are likely to be important for generating adequate sample sizes. This is likely to become a pressing scientific question with growing interest and debates about newborn genome sequencing (Berg & Powell, 2015; Berg et al., 2017).

CONFLICTS OF INTEREST

The authors state that they have no conflict of interest.

ORCID

Megan Wadon https://orcid.org/0000-0002-7694-6228

REFERENCES

Agrawal, S., Rao, S. C., Bulsara, M. K., & Patole, S. K. (2018). Prevalence of autism spectrum disorder in preterm infants: A meta-analysis. *Pediatrics*, *142*(3), e20180134. Retrieved from https://doi.org/10.1542/peds.2018-0134

Berg, J. S., & Powell, C. M. (2015). Potential uses and inherent challenges of using genome-scale sequencing to augment current newborn screening. *Cold Spring Harbor Perspectives in Medicine*, *5*(12), a023150. Retrieved from https://doi.org/10.1101/cshperspect.a023150

Berg, J. S., Agrawal, P. B., Bailey, D. B., Beggs, A. H., Brenner, S. E., Brower, A. M., … Wise, A. L. (2017). Newborn sequencing in genomic medicine and public health. *Pediatrics*, *139*(2), e20162252. Retrieved from https://doi.org/10.1542/peds.2016-2252

Blencowe, H., Cousens, S., Oestergaard, M. Z., Chou, D., Moller, A.-B., Narwal, R., … Lawn, J. E. (2012). National, regional, and worldwide estimates of preterm birth rates in the year 2010 with time trends since 1990 for selected countries: A systematic analysis and implications. *The Lancet*, *379*(9832), 2162–2172. Retrieved from https://doi.org/10.1016/S0140-6736(12)60820-4

Boffetta, P., Winn, D. M., Ioannidis, J. P., Thomas, D. C., Little, J., Smith, G. D., … Khoury, M. J. (2012). Recommendations and proposed guidelines for assessing the cumulative evidence on joint effects of genes and environments on cancer occurrence in humans. *International Journal of Epidemiology*, *41*(3), 686–704. Retrieved from https://doi.org/10.1093/ije/dys010

Burris, H. H., Lorch, S. A., Kirpalani, H., Pursley, D. M., Elovitz, M. A., & Clougherty, J. E. (2019). Racial disparities in preterm birth in USA: A biosensor of physical and social environmental exposures. *Archives of Disease in Childhood*, *104*(10), 931–935. Retrieved from https://doi.org/10.1136/archdischild-2018-316486

Bustamante, M., Danileviciute, A., Espinosa, A., Gonzalez, J., Subirana, I., Cordier, S., … Kogevinas, M. (2012). Influence of fetal glutathione *S*-transferase copy number variants on adverse reproductive outcomes. *BJOG: An International Journal of Obstetrics & Gynaecology*, *119*(9), 1141–1146. Retrieved from https://doi.org/10.1111/j.1471-0528.2012.03400.x

Clausson, B., Lichtenstein, P., & Cnattingius, S. (2000). Genetic influence on birthweight and gestational length determined by studies in offspring of twins. *BJOG: An International Journal of Obstetrics and Gynaecology*, *107*(3), 375–381. Retrieved from https://doi.org/10.1111/j.1471-0528.2000.tb13234.x

Franz, A. P., Bolat, G. U., Bolat, H., Matijasevich, A., Santos, I. S., Silveira, R. C., … Moreira-Maia, C. R. (2018). Attention-deficit/hyperactivity disorder and very preterm/very low birth weight: A meta-analysis. *Pediatrics*, *141*(1), e20171645. Retrieved from https://doi.org/10.1542/peds.2017-1645

Haraldsdottir, K., Watson, A. M., Goss, K. N., Beshish, A. G., Pegelow, D. F., Palta, M., … Eldridge, M. W. (2018). Impaired autonomic function in adolescents born preterm. *Physiological Reports*, *6*(6), e13620. Retrieved from https://doi.org/10.14814/phy2.13620

Heuvelman, H., Abel, K., Wicks, S., Gardner, R., Johnstone, E., Lee, B., … Rai, D. (2018). Gestational age at birth and risk of intellectual disability without a common genetic cause. *European Journal*

of Epidemiology, *33*(7), 667–678. Retrieved from https://doi.org/10.1007/s10654-017-0340-1

Hong, X., Hao, K., Ji, H., Peng, S., Sherwood, B., Di Narzo, A., … Wang, X. (2017). Genome-wide approach identifies a novel gene-maternal pre-pregnancy BMI interaction on preterm birth. *Nature Communications*, *8*, 15608. Retrieved from https://doi.org/10.1038/ncomms15608

Huang, Y.-T., Lin, H.-Y., Wang, C.-H., Su, B.-H., & Lin, C.-C. (2018). Association of preterm birth and small for gestational age with metabolic outcomes in children and adolescents: A population-based cohort study from Taiwan. *Pediatrics & Neonatology*, *59*(2), 147–153. Retrieved from https://doi.org/10.1016/j.pedneo.2017.07.007

Hunter, D. J. (2005). Gene–environment interactions in human diseases. *Nature Reviews. Genetics*, *6*(4), 287–298. Retrieved from https://doi.org/10.1038/nrg1578

Huusko, J. M., Karjalainen, M. K., Graham, B. E., Zhang, G., Farrow, E. G., Miller, N. A., … Muglia, L. J. (2018). Whole exome sequencing reveals HSPA1L as a genetic risk factor for spontaneous preterm birth. *PLOS Genetics*, *14*(7), e1007394. Retrieved from https://doi.org/10.1371/journal.pgen.1007394

Kistka, Z. A.-F., DeFranco, E. A., Ligthart, L., Willemsen, G., Plunkett, J., Muglia, L. J., & Boomsma, D. I. (2008). Heritability of parturition timing: An extended twin design analysis. *American Journal of Obstetrics and Gynecology*, *199*(1), 43.e1–43.e5. Retrieved from https://doi.org/10.1016/J.AJOG.2007.12.014

Lambert, S. A., Abraham, G., & Inouye, M. (2019). Towards clinical utility of polygenic risk scores. *Human Molecular Genetics*, *28*(R2), R133–R142. Retrieved from https://doi.org/10.1093/hmg/ddz187

Li, J., Oehlert, J., Snyder, M., Stevenson, D. K., & Shaw, G. M. (2017). Fetal de novo mutations and preterm birth. *PLOS Genetics*, *13*(4), e1006689. Retrieved from https://doi.org/10.1371/journal.pgen.1006689

Liu, Y., Tang, Y.-B., Chen, J., & Huang, Z.-X. (2014). Meta-analysis of GSTT1 null genotype and preterm delivery risk. *International Journal of Clinical and Experimental Medicine*, *7*(6), 1537–1541. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/25035777

Liu, L., Oza, S., Hogan, D., Perin, J., Rudan, I., Lawn, J. E., … Black, R. E. (2015). Global, regional, and national causes of child mortality in 2000–13, with projections to inform post-2015 priorities: An updated systematic analysis. *The Lancet*, *385*(9966), 430–440. Retrieved from https://doi.org/10.1016/S0140-6736(14)61698-6

Liu, X., Helenius, D., Skotte, L., Beaumont, R. N., Wielscher, M., Geller, F., … Feenstra, B. (2019). Variants in the fetal genome near pro-inflammatory cytokine genes on 2q13 associate with gestational duration. *Nature Communications*, *10*(1), 3927. Retrieved from https://doi.org/10.1038/s41467-019-11881-8

Lunde, A., Melve, K. K., Gjessing, H. K., Skjaerven, R., & Irgens, L. M. (2007). Genetic and environmental influences on birth weight, birth length, head circumference, and gestational age by use of population-based parent-offspring data. *American Journal of Epidemiology*, *165*(7), 734–741. Retrieved from https://doi.org/10.1093/aje/kwk107

McElroy, J. J., Gutman, C. E., Shaffer, C. M., Busch, T. D., Puttonen, H., Teramo, K., … Muglia, L. J. (2013). Maternal coding variants in complement receptor 1 and spontaneous idiopathic preterm birth. *Human Genetics*, *132*(8), 935–942. Retrieved from https://doi.org/10.1007/s00439-013-1304-5

Modi, B. P., Teves, M. E., Pearson, L. N., Parikh, H. I., Chaem-saithong, P., Sheth, N. U., … Strauss, J. F. (2017a). Rare mutations and potentially damaging missense variants in genes encod-ing fibrillar collagens and proteins involved in their production are candidates for risk for preterm premature rupture of membranes. *PLOS ONE*, *12*(3), e0174356. Retrieved from https://doi.org/10.1371/journal.pone.0174356

Modi, B. P., Teves, M. E., Pearson, L. N., Parikh, H. I., Haymond-Thornburg, H., Tucker, J. L., … Strauss, J. F., III (2017b). Mutations in fetal genes involved in innate immunity and host defense against microbes increase risk of preterm premature rupture of membranes (PPROM). *Molecular Genetics & Genomic Medicine*, *5*(6), 720–729. Retrieved from https://doi.org/10.1002/mgg3.330

Modi, B. P., Parikh, H. I., Teves, M. E., Kulkarni, R., Liyu, J., Romero, R., … Strauss, J. F., III (2018). Discovery of rare ancestry-specific variants in the fetal genome that confer risk of preterm premature rupture of membranes (PPROM) and preterm birth. *BMC Medical Genetics*, *19*(1), 181. Retrieved from https://doi.org/10.1186/s12881-018-0696-4

Mustafa, M. D., Pathak, R., Ahmed, T., Ahmed, R. S., Tripathi, A. K., Guleria, K., & Banerjee, B. D. (2010). Association of glutathione *S*-transferase M1 and T1 gene polymorphisms and oxidative stress markers in preterm labor. *Clinical Biochemistry*, *43*(13–14), 1124–1128. Retrieved from https://doi.org/10.1016/J.CLINBIOCHEM.2010.06.018

Mustafa, M. D., Banerjee, B. D., Ahmed, R. S., Tripathi, A. K., & Guleria, K. (2013). Gene–environment interaction in preterm delivery with special reference to organochlorine pesticides. *Molecular Human Reproduction*, *19*(1), 35–42. Retrieved from https://doi.org/10.1093/molehr/gas039

Myking, S., Boyd, H. A., Myhre, R., Feenstra, B., Jugessur, A., Devold Pay, A. S., … Murray, J. C. (2013). X-chromosomal maternal and fetal SNPs and the risk of spontaneous preterm delivery in a Danish/Norwegian genome-wide association study. *PloS One*, *8*(4), e61781. Retrieved from https://doi.org/10.1371/journal.pone.0061781

Nosarti, C., Reichenberg, A., Murray, R. M., Cnattingius, S., Lambe, M. P., Yin, L., … Hultman, C. M. (2012). Preterm birth and psychiatric disorders in young adult life. *Archives of General Psychiatry*, *69*(6), 610–617. Retrieved from https://doi.org/10.1001/archgenpsychiatry.2011.1374

Palomar, L., DeFranco, E. A., Lee, K. A., Allsworth, J. E., & Muglia, L. J. (2007). Paternal race is a risk factor for preterm birth. *American Journal of Obstetrics and Gynecology*, *197*(2), 152.e1–152.e7. Retrieved from https://doi.org/10.1016/j.ajog.2007.03.035

Parkinson, J. R. C., Hyde, M. J., Gale, C., Santhakumaran, S., & Modi, N. (2013). Preterm birth and the metabolic syndrome in adult life: A systematic review and meta-analysis. *Pediatrics*, *131*(4), e1240–e1263. Retrieved from https://doi.org/10.1542/peds.2012-2177

Plunkett, J., Feitosa, M. F., Trusgnich, M., Wangler, M. F., Palomar, L., Kistka, Z. A.-F., … Muglia, L. J. (2009). Mother's genome or maternally-inherited genes acting in the fetus influence gestational age in familial preterm birth. *Human Heredity*, *68*(3), 209–219. Retrieved from https://doi.org/10.1159/000224641

Rappoport, N., Toung, J., Hadley, D., Wong, R. J., Fujioka, K., Reuter, J., … Sirota, M. (2018). A genome-wide association study identifies only two ancestry specific variants associated with spontaneous preterm birth. *Scientific Reports*, *8*(1), 226. Retrieved from https://doi.org/10.1038/s41598-017-18246-5

Santhakumaran, S., Statnikov, Y., Gray, D., Battersby, C., Ashby, D., Modi, N. & Medicines for Neonates Investigator Group. (2018). Survival of very preterm infants admitted to neonatal care in England

2008-2014: Time trends and regional variation. *Archives of Disease in Childhood. Fetal and Neonatal Edition*, *103*(3), F208–F215. Retrieved from https://doi.org/10.1136/archdischild-2017-312748

Schizophrenia Working Group of the Psychiatric Genomics Consortium. (2014). Biological insights from 108 schizophrenia-associated genetic loci. *Nature*, *511*(7510), 421–427. Retrieved from https://doi.org/10.1038/nature13595

Sherf, Y., Sheiner, E., Vardi, I., Sergienko, R., Klein, J., & Bilenko, N. (2017). Recurrence of preterm delivery in women with a family history of preterm delivery. *American Journal of Perinatology*, *34*(4), 397–402. Retrieved from https://doi.org/10.1055/s-0036-1592131

Suh, Y.-J., Ha, E.-H., Park, H., Kim, Y.-J., Kim, H., & Hong, Y.-C. (2008). GSTM1 polymorphism along with PM10 exposure contributes to the risk of preterm delivery. *Mutation Research/Genetic Toxicology and Environmental Mutagenesis*, *656*(1–2), 62–67. Retrieved from https://doi.org/10.1016/J.MRGENTOX.2008.07.006

Sutton, P. S., & Darmstadt, G. L. (2013). Preterm birth and neurodevelopment: A review of outcomes and recommendations for early identification and cost-effective interventions. *Journal of Tropical Pediatrics*, *59*(4), 258–265. Retrieved from https://doi.org/10.1093/tropej/fmt012

Svensson, A. C., Sandin, S., Cnattingius, S., Reilly, M., Pawitan, Y., Hultman, C. M., & Lichtenstein, P. (2009). Maternal effects for preterm birth: A genetic epidemiologic study of 630,000 families. *American Journal of Epidemiology*, *170*(11), 1365–1372. Retrieved from https://doi.org/10.1093/aje/kwp328

Torkamani, A., Wineinger, N. E., & Topol, E. J. (2018). The personal and clinical utility of polygenic risk scores. *Nature Reviews Genetics*, *19*, 581–590. Retrieved from https://doi.org/10.1038/s41576-018-0018-x

Treloar, S. A., Macones, G. A., Mitchell, L. E., & Martin, N. G. (2000). Genetic influences on premature parturition in an Australian twin sample. *Twin Research*, *3*(2), 80–82. Retrieved from https://doi.org/10.1375/twin.3.2.80

Tsai, H.-J., Liu, X., Mestan, K., Yu, Y., Zhang, S., Fang, Y., … Wang, X. (2008). Maternal cigarette smoking, metabolic gene polymorphisms, and preterm delivery: New insights on G×E interactions and pathogenic pathways. *Human Genetics*, *123*(4), 359–369. Retrieved from https://doi.org/10.1007/s00439-008-0485-9

Uzun, A., Sahin, Y., Schuster, J. S., Zheng, X., Ryckman, K., Feingold, E., & Padbury, J. (2016a). Structural and genomic variation in preterm birth. *Pediatric Research*, *80*(6), 829–836. Retrieved from https://doi.org/10.1038/pr.2016.152

Uzun, A., Schuster, J., McGonnigal, B., Schorl, C., Dewan, A., & Padbury, J. (2016b). Targeted sequencing and meta-analysis of preterm birth. *PLoS One*, *11*(5), e0155021. Retrieved from https://doi.org/10.1371/journal.pone.0155021

Wang, X., Chen, D., Niu, T., Wang, Z., Wang, L., Ryan, L., … Xu, X. (2000). Genetic susceptibility to benzene and shortened gestation: Evidence of gene–environment interaction. *American Journal of Epidemiology*, *152*(8), 693–700. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11052546

Warren, H. R., Evangelou, E., Cabrera, C. P., Gao, H., Ren, M., & Mifsud, B., … UK Biobank CardioMetabolic Consortium BP working group. (2017). Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk. *Nature Genetics*, *49*(3), 403–415. Retrieved from https://doi.org/10.1038/ng.3768

Wu, W., Witherspoon, D. J., Fraser, A., Clark, E. A. S., Rogers, A., Stoddard, G. J., … Jorde, L. B. (2015). The heritability of gestational age in a two-million member cohort: Implications for spontaneous preterm birth. *Human Genetics*, *134*(7), 803–808. Retrieved from https://doi.org/10.1007/s00439-015-1558-1

York, T. P., Eaves, L. J., Lichtenstein, P., Neale, M. C., Svensson, A., Latendresse, S., … Strauss, J. F., III. (2013). Fetal and maternal genes' influence on gestational age in a quantitative genetic analysis of 244,000 Swedish births. *American Journal of Epidemiology*, *178*(4), 543–550. Retrieved from https://doi.org/10.1093/aje/kwt005

York, T. P., Eaves, L. J., Neale, M. C., & Strauss, J. F. (2014). The contribution of genetic and environmental factors to the duration of pregnancy. *American Journal of Obstetrics and Gynecology*, *210*, 398–405. Retrieved from https://doi.org/10.1016/j.ajog.2013.10.001

Zhang, H., Baldwin, D. A., Bukowski, R. K., Parry, S., Xu, Y., & Song, C., … Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Genomic and Proteomic Network for Preterm Birth Research (GPN-PBR). (2015). A genome-wide association study of early spontaneous preterm delivery. *Genetic Epidemiology*, *39*(3), 217–226. Retrieved from https://doi.org/10.1002/gepi.21887

Zhang, G., Feenstra, B., Bacelis, J., Liu, X., Muglia, L. M., Juodakis, J., … Muglia, L. J. (2017). Genetic associations with gestational duration and spontaneous preterm birth. *The New England Journal of Medicine*, *377*(12), 1156–1167. Retrieved from https://doi.org/10.1056/NEJMoa1612665

Zheng, X., Feingold, E., Ryckman, K. K., Shaffer, J. R., Boyd, H. A., Feenstra, B., … Cuenco, K. T. (2013). Association of maternal CNVs in GSTT1/GSTT2 with smoking, preterm delivery, and low birth weight. *Frontiers in Genetics*, *4*, 196. Retrieved from https://doi.org/10.3389/fgene.2013.00196