Pharmacogenomics of 17-alpha hydroxyprogesterone caproate for recurrent preterm birth: a case–control study

TA Manuck,[a,b,c] WS Watkins,[d] MS Esplin,[a,b] J Biggio,[e] R Bukowski,[f] S Parry,[g] H Zhan,[h] H Huang,[h] W Andrews,[e] G Saade,[f] Y Sadovsky,[i] UM Reddy,[j] J Ilekis,[j] M Yandell,[d] MW Varner,[a,b] LB Jorde,[d] for the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) Genomics and Proteomics Network for Preterm Birth Research (GPN-PBR)

[a] Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, University of Utah School of Medicine, Salt Lake City, UT, USA [b] Intermountain Healthcare Department of Maternal Fetal Medicine, Salt Lake City, UT, USA [c] Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, University of North Carolina-Chapel Hill, Chapel Hill, NC, USA [d] Department of Human Genetics, University of Utah, Salt Lake City, UT, USA [e] Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine and Center for Women's Reproductive Health, University of Alabama at Birmingham, Birmingham, AL, USA [f] Department of Obstetrics and Gynecology, Division of Maternal-Fetal Medicine, University of Texas Medical Branch, Galveston, TX, USA [g] Department of Obstetrics and Gynecology, University of Pennsylvania School of Medicine, Philadelphia, PA, USA [h] Collaborative Center for Statistics in Science, Yale University School of Public Health, New Haven, CT, USA [i] Magee-Womens Research Institute, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA [j] Pregnancy and Perinatology Branch, Center for Developmental Biology and Perinatal Medicine, Eunice Kennedy Shriver National Institute of Child Health and Human Development, Bethesda, MD, USA
Correspondence: TA Manuck, UNC Department of Obstetrics and Gynecology, Division of Maternal Fetal Medicine, 3010 Old Clinic Building, CB#7516, Chapel Hill, NC 27599-7516, USA. Email tmanuck@med.unc.edu

Objective To compare maternal genotypes between women with and without significant prolongation of pregnancy in the setting of 17-alpha hydroxyprogesterone caproate (17-P) administration for the prevention of recurrent preterm birth (PTB).

Design Case–control.

Setting Three tertiary-care centres across the USA.

Population Women ($n = 99$) with ≥ 1 prior singleton spontaneous PTB, receiving 17-P.

Methods Women were classified as having successful prolongation of pregnancy during the 17-P treated pregnancy, in two ways: (1) Definition A: success/non-success based on difference in gestational age at delivery between 17-P-treated and untreated pregnancies (success: delivered ≥ 3 weeks later with 17-P) and (2) Definition B: success/non-success based on reaching term (success: delivered at term with 17-P).

Main outcome measures To assess genetic variation, all women underwent whole exome sequencing. Between-group sequence variation was analysed with the Variant Annotation, Analysis, and Search Tool (VAAST). Genes scored by VAAST with $P < 0.05$ were then analysed with two online tools: (1) Protein ANalysis THrough Evolutionary Relationships (PANTHER) and (2) Database for Annotation, Visualization, and Integrated Discovery (DAVID).

Results Using Definition A, there were 70 women with successful prolongation and 29 without; 1375 genes scored by VAAST had $P < 0.05$. Using Definition B, 47 women had successful prolongation and 52 did not; 1039 genes scored by VAAST had $P < 0.05$. PANTHER revealed key differences in gene ontology pathways. Many genes from definition A were classified as prematurity genes ($P = 0.026$), and those from definition B as pharmacogenetic genes ($P = 0.0018$); (P, non-significant after Bonferroni correction).

Conclusion A novel analytic approach revealed several genetic differences among women delivering early *vs* later with 17-P.

Keywords 17-Alpha hydroxyprogesterone caproate, current preterm birth, pharmacogenomics, spontaneous prematurity.

Tweetable abstract Several key genetic differences are present in women with recurrent preterm birth despite 17-P treatment.

Introduction

Although the rate of preterm birth (PTB) has recently declined slightly, prematurity and its consequences remain major public health problems worldwide.[1] PTB prediction and prevention remain clinically challenging. One of the strongest predictors is a history of a prior PTB.[2,3] Weekly intramuscular injections of 17-alpha hydroxyprogesterone caproate (17-P) effectively reduces the risk of recurring PTB; offering this prophylaxis to women with a history of a prior spontaneous PTB is standard care in the United States.[4,5]

Although many women have successful pregnancy outcomes when treated with 17-P, others experience recurrent PTB despite prophylaxis. The reasons for this variable response to therapy are poorly understood. Many experts have theorised that an individualised, or personalised medicine approach to PTB prevention may improve outcomes compared with a 'one size fits all' remedy. Personalised medicine is a rapidly expanding area of medicine, which seeks to provide the right drug, at the right dose, to the right patient at the right time.[6,7] Both clinical information and biological information can be used to inform personalised medicine approaches. For example, many experts have recently used clinical information to phenotype women with PTB to gain additional insight into the underlying pathophysiology and recurrent PTB despite the administration of 17-P.[8] One result of this research is the finding that women with a history of preterm premature rupture of membranes have lower rates of recurrent PTB when treated with 17-P than do those with a history of idiopathic PTB.[9,10] Thus, a woman's clinical risk factors combined to determine her PTB phenotype may impact on her clinical response to 17-P; however, this relationship remains poorly understood.

Genetic factors may also inform the personalised medicine approach and contribute to the variable pregnancy outcomes among women on 17-P. One preliminary pharmacogenomic investigation implicated genes in several broad biological pathways, including cell adhesion, cell communication, signal transduction, receptor activity and nitric oxide signalling as contributing to the variable response to 17-P.[11] Another study found that the progesterone receptor genotype may influence the risk of recurrent PTB among women on 17-P.[12] However, these studies were small, and it remains uncertain whether the results are widely applicable.

The objective of this study was to further characterise maternal genotype and pregnancy outcomes among women with a history of a prior spontaneous singleton PTB who received 17-P, utilising a personalised medicine framework. We hypothesised that maternal genotype influences pregnancy outcomes among women treated with 17-P for the prevention of recurrent spontaneous PTB.

Methods

This is a case–control genetic association study, and is a planned secondary analysis of a prospective multicentre cohort study of PTB conducted by the NICHD Genomics and Proteomics Network for Preterm Birth Research (GPN). The original cohort has previously been described.[13] Briefly, the main study was a prospective, observational cohort study of women enrolled at one of three main clinical sites (University of Alabama, Birmingham, University of Texas Medical Branch, Galveston, University of Utah/Intermountain Healthcare). All women were clinically managed by individual clinicians; there were no proscribed study algorithms with regards to pregnancy management, specific treatments (e.g., progesterone prophylaxis) or clinical surveillance (e.g., cervical length). Women ($n = 99$) with at least one documented singleton, non-anomalous spontaneous PTB between 16 and 37 weeks' gestation before the studied pregnancy, who received 17-P during the enrolled pregnancy and with DNA available, were included in this analysis. All women provided written, informed consent for participation in the parent study, which included consent for genetic analyses such as this secondary analysis. This specific secondary analysis was approved by the institutional review boards at Intermountain Healthcare and the University of Utah.

We studied two primary outcomes related to gestational age at delivery of the studied pregnancy treated with 17-P. The first method, subsequently referred to as pregnancy outcome 'Definition A', considered each woman's earliest historical delivery gestational age without 17-P to define pregnancy outcome with the 17-P-treated pregnancy.[11,14] Women who delivered at least 3 weeks later than the gestational age of the earliest prior PTB without 17-P were considered to have had a successful subsequent pregnancy prolongation with 17-P. Individuals who did not gain at least 3 weeks' gestation or delivered earlier when taking 17-P were considered to have had an unsuccessful pregnancy prolongation despite 17-P. Those who delivered 3 or more weeks earlier with 17-P were identified as having potential harm with 17-P. The second classification, subsequently referred to as pregnancy outcome 'Definition B', defined recurrent PTB in a traditional manner. Those women who delivered at term (≥ 37 weeks' gestation) with 17-P were considered to have had a successful pregnancy prolongation, whereas those delivering before 37 weeks had a recurrent PTB and were considered to have had an unsuccessful pregnancy prolongation despite 17-P. Demographic data and clinical characteristics were compared between responders and non-responders using Student's t-test, the chi-square test or Fisher's exact test as appropriate, using Stata version 13.1 (StataCorp LP, College Station, TX, USA).

DNA was extracted from stored buffy coats, and DNA quality assessed using NanoDrop spectrophotometer reading as well as evaluation on a 1% agarose gel prior to genomic library construction. Genomic libraries were constructed, and samples underwent additional quality-control measures including quantitative polymerase chain reaction quantitation of library concentration using Illumina® primers and library evaluation on an Agilent Bioanalyzer DNA 1000 chip. A PhiX control library (constructed by Illumina) was spiked into each lane at a concentration representing approximately 0.5% of the reads. This platform targets approximately 180 000 protein-coding exons, located in approximately 20 000 genes, for capture. Whole exome sequencing was then performed at The University of Utah Huntsman Cancer Institute's Microarray Core Facility (Salt Lake City, UT) using Illumina (San Diego, CA) HiSeq2000 technology. We indexed four samples per lane.

Sequences were then called simultaneously on all samples using the University of Utah Department of Human Genetics (Salt Lake City, UT) variant-calling software pipeline. Paired-end 101 bp fastq reads were aligned to the reference genome (b37) using the Burrows–Wheeler aligner.[15] Additional processing, including sorting, mate-fixing and duplicate read removal, was performed using Samtools and Picard Tools.[16] Indel realignment and base recalibration were performed using the Genome Analysis Tool Kit (GATK, Broad Institute, Cambridge, MA).[17,18] Processed call-ready BAM files were called jointly using the Unified Genotyper (GATK package). Raw genotypes were evaluated and filtered using the Variant Quality Score Recalibrator provided in the GATK package.

Individual exomes were analysed for evidence of population stratification using Eigenstrat software, and the proportion of European and African ancestry for each individual was estimated using the Admixture software package.[19] Single nucleotide polymorphisms (SNPs) with a minor allele frequency of < 0.05 and/or a strong deviation from Hardy–Weinberg equilibrium ($P < 0.00001$) were removed. SNPs were also filtered to remove all SNPs with pairwise linkage disequilibrium of $r^2 > 0.2$. Population stratification was then analysed among the remaining subjects, and individual outliers were excluded from further analysis.

The remaining exomes were compared between women with successful and those with unsuccessful pregnancy prolongation during their 17-P-treated pregnancy using the Variant Annotation, Analysis and Search Tool (VAAST Omicia, Inc., Oakland, CA and University of Utah, Salt Lake City, UT).[20–22] VAAST is a publicly available probabilistic search tool for identifying disease-causing variants. VAAST evaluates coding variants based on the allele and amino acid substitution frequency differences between case and control genomes, and has been shown to be effective in identifying causative disease alleles both in cases of rare variants in rare diseases and in combinations of rare and common variants in common diseases.[20–22] The VAAST analysis produced a 'raw' list of genes, prioritised by the likelihood of allelic differences between 17-P responders and non-responders. We utilised genes with raw P-values of < 0.05 (reflecting differences in genotypes between 17-P responders and non-responders) (uncorrected for multiple comparisons) for our VAAST analysis gene list.

In the next step of our pathway analysis, we again utilised the raw VAAST list of genes, and selected those genes with VAAST P-values of <0.05 from that list. Each gene on this raw list was classified by the online Protein ANalysis THrough Evolutionary Relationships (PANTHER) classification system (annotation version 10.0, released 5/15/2015) into known gene ontology molecular functions and biologic processes.[23–25] The percentage of genes within each molecular gene ontology group was compared with an online referent population to search for categories of over- and under-representation using the binomial test available through PANTHER tools online. A Bonferroni-corrected P-value of <0.05 for the binomial test was considered significant for the PANTHER analysis. Genes classified into these gene ontology groups were searched for possible biologic plausibility and/or known prior association with PTB, prematurity risk factors and other major pregnancy and perinatal complications using the online Database for Preterm Birth (dbPTB), the UCSC genome browser, and PubMed.[26]

Finally, the raw VAAST gene list was also analysed using the online Database for Annotation, Visualization, and Integrated Discovery (DAVID) version 6.7, sponsored by the National Institutes of Health National Institute of Allergy and Infectious Diseases.[27–29] In addition to assessing gene groupings for over- and under-representation, the DAVID database also queries the gene list for known disease associations. A Bonferroni-corrected P-value of < 0.05 was considered significant for the DAVID portion of the analysis. The overall study analysis schematic is summarised in Figure 1.

Results

From the original GPN cohort, 99 women met the inclusion criteria. Using Definition A (women with successful pregnancy prolongation were those who delivered at least 3 weeks later than they did in the earliest prior PTB), there were 70 with favourable and 29 with non-favourable outcomes. Women with successful prolongation delivered an average of 7.7 ± 5.4 weeks later (95% CI, 6.4–9.0) with 17-P compared with 0.4 ± 2.3 weeks later (95% CI, −0.5 to 1.2) for unsuccessful prolongation ($P < 0.001$). Eight women delivered earlier in their 17-P-treated pregnancy; four delivered less than one week earlier and the others delivered 13, 18, 33 and 54 days earlier. Thus, only two women delivered 3 or more weeks earlier with 17-P and
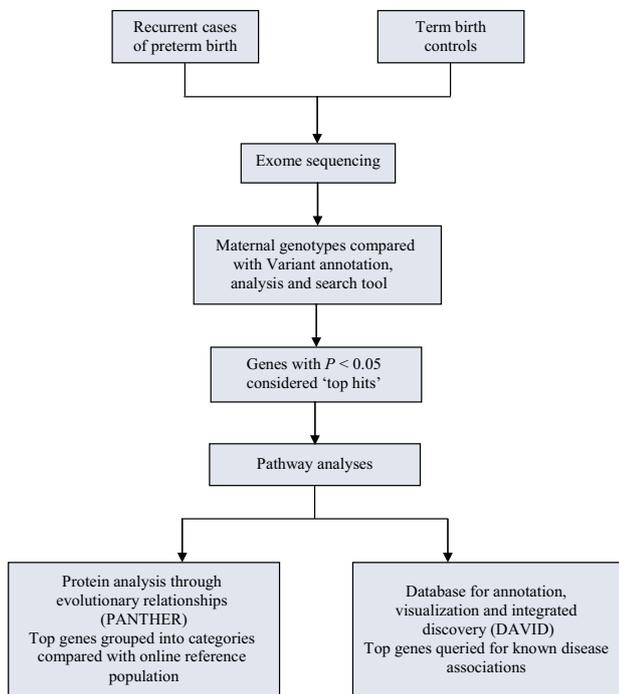
Figure 1. Overall study schematic and analysis summary.

were identified as having potential harm from 17-P. Owing to the small numbers, these two individuals were combined with women classified as having unsuccessful pregnancy prolongation for the purposes of the analysis. Using Definition B (women with successful pregnancy prolongation delivered at term), there were 47 with favourable and 52 with unfavourable outcomes.

Demographic and prior pregnancy characteristics were similar between those with favourable and those with unfavourable subsequent pregnancy gestational ages (Table 1), although using Definition A, women with favourable outcomes were more likely to have a history of cervical insufficiency in one or more prior pregnancies. The distribution of genetic ancestry was also similar (Table 1), regardless of pregnancy outcome definition. Most women were healthy without major medical co-morbidities; two (2%) had pre-existing type I or type II diabetes mellitus, three (3%) had a history of chronic hypertension and none had a history of chronic renal insufficiency. Six developed pre-eclampsia, but four were diagnosed at term, and none required PTB due to maternal or fetal indications (e.g., delivery due to worsening pre-eclampsia). Other pregnancy characteristics did not vary by degree of pregnancy prolongation (Table 2).

Table 1. Demographic characteristics, stratified by 17P outcome definition

	Outcome Definition A			Outcome Definition B		
	Significant pregnancy prolongation with 17P n =70	Delivered at similar gestational age with 17P n = 29	P-value	Delivered at term with 17P n = 47	Delivered preterm with 17P n = 52	P-value
Genetic Ancestry Markers						
>90% European	33 (49.3)	17 (65.4)	0.161	22 (46.8)	28 (53.8)	0.361
>70% African	21 (31.3)	7 (26.9)	0.677	15 (31.9)	13 (25.0)	0.511
Prepregnancy body mass index, kg/m², mean ± SD	27.2 ± 7.7	25.0 ± 5.6	0.177	27.9 ± 7.7	25.4 ± 6.5	0.081
Married, n(%)	41 (58.6)	20 (69.0)	0.333	28 (59.6)	33 (63.5)	0.691
Tobacco use, n(%)	11 (15.7)	4 (13.8)	>0.99	8 (17.0)	7 (13.5)	0.622
Number of prior pregnancies, mean ± SD	2.4 ± 1.3	2.4 ± 1.4	0.963	2.4 ± 1.2	2.4 ± 1.5	0.883
One or more prior term deliveries, n(%)	22 (31.4)	7 (24.1)	0.468	13 (25.0)	16 (34.0)	0.324
Total number of prior preterm (20.0 - 36.9 weeks) deliveries, median (IQR)	1 (1, 2)	1 (1, 2)	0.790	1 (1, 2)	1 (1, 2)	0.107
History of PPROM	4 (5.7)	3 (10.3)	0.415	5 (9.6)	2 (4.3)	0.440
History of cervical insufficiency	12 (17.1)	0 (0)	0.016	6 (11.5)	6 (12.8)	>0.99
Delivery gestational age (weeks) of earliest prior preterm birth, median (IQR)	32 (24, 34)	34 (32, 35)	0.0016	33 (28, 35)	32 (27, 34)	0.643

Table 2. Pregnancy management and outcomes during the studied pregnancy

	Outcome Definition A			Outcome Definition B		
	Significant pregnancy prolongation with 17P n = 70	Delivered at similar gestational age with 17P n = 29	P-value	Delivered at term with 17P n = 47	Delivered preterm with 17P n = 52	P-value
Unexplained vaginal bleeding during pregnancy, n (%)	8 (11.4)	1 (3.4)	0.276	4 (8.5)	5 (9.6)	>0.99
Vaginal cervical length assessed at least once during pregnancy, n (%)	43 (61.4)	13 (44.8)	0.129	30 (63.8)	26 (50.0)	0.166
Short cervix <2.50 cm detected during pregnancy*	6/43 (14.0)	1/13 (7.7)	>0.99	2/30 (6.7)	5/26 (19.2)	0.231
Cervical Cerclage Placed, n (%)	6 (8.6)	0 (0)	0.176	3 (6.4)	3 (5.8)	>0.99
Preterm premature rupture of membranes, n (%)	9 (12.9)	7 (24.1)	0.165	16 (30.8)	0 (0)	<0.001
Delivery gestational age, weeks, mean ± SD	37.0 ± 3.4	33.4 ± 3.6	<0.001	38.9 ± 0.9	33.3 ± 3.5	<0.001
Birthweight, grams, mean ± SD	2850 ± 737	2202 ± 694	<0.001	3222 ± 381	2153 ± 697	<0.001

*Among women with one or more transvaginal cervical length assessments during pregnancy.

Sequencing depth of coverage averaged 43.2 (interquartile range, 38.0–49.1), and did not vary by outcome definition (data not shown). In our VAAST analysis, we allowed for recessive inheritance and locus heterogeneity, and tested our genotypes using one million permutations. The top 100 genes with the greatest differences between successful and unsuccessful pregnancy prolongation are listed in Table S1 and represent the 'raw' VAAST list of genes. The P-values displayed in Table S1 are unadjusted; none meet genome-wide significance ($P < 2.5 \times 10^{-6}$).

Recently, Uzun and colleagues[30] applied a pathway-based genetic analysis approach to examine a large custom database for PTB with a refined set of genes curated from the literature and biological databases. These investigators identified two main pathways (breast cancer estrogen signalling and oxidoreductase activity), each with 10 different genes contributing significantly to PTB.[30] One gene (CYP19A1) was shared by both pathways; thus, they identified a total of 19 distinct genes across these two pathways. Using Definition A, 2/19 genes had a raw VAAST P of <0.05 (TP53, ESR1) and 6/19 (TP53, ESR1, IL6, ADH6, PTGS1, P4HA2) had a raw VAAST P of <0.10. Using Definition B, 3/19 genes had a raw VAAST P of <0.05 (TP53, PTGS2, ESR1) and 4/19 (TP53, PTGS2, ESR1, PTGS1) had a raw VAAST P of <0.10.

Using outcome Definition A, 1375 of 11 277 genes (12%) scored by VAAST had $P < 0.05$; 1039 of 11 422 (9%) genes scored by VAAST had $P < 0.05$ using outcome Definition B; top genes for each definition were entered into PANTHER for pathway analysis. In the PANTHER analysis, genes generated from the VAAST analysis using Definition A were over-represented in three broad regulatory gene ontology pathways – system process, biological regulation and metabolic process (Table S2). Genes from the raw VAAST analysis that fell into each of these categories and that could serve as potential candidate genes for response to 17-P are shown in Table S2. In contrast, genes generated from the VAAST analysis using Definition B were over-represented in two different gene ontology pathways – receptor activity and structural component of cytoskeleton (Table S3). Representative genes from each of these pathways are also listed in both Tables S2 and S3; a complete list is available online (PANTHER website, www.pantherdb.org and others).

Finally, analysis was performed using DAVID. Again, the raw list of genes for each outcome definition as generated by VAAST was utilised. Overall, 21.0% of raw VAAST genes from Definition A and 17.6% from Definition B could be grouped into known prior disease associations (Table S4). Notably, using Definition A, genes were grouped into a 'preterm birth' category, and using Definition B, genes were grouped into a 'pharmacogenomics' category. However, no P-values for any of the DAVID analyses remained significant after controlling for multiple comparisons.

Discussion

Main findings

A novel analytic approach revealed several key genetic differences among women with recurrent PTB or recurrent

PTB at similar gestational ages despite 17-P prophylaxis, and highlighted genes and pathways suspected in the pathogenesis of PTB. Genotypes varied by outcome definition; these findings emphasise the importance of refining definitions of successful pregnancy outcomes following one or more prior spontaneous PTB(s).

Strengths and limitations of the study

A strength of our study is our ability to examine the effects of genotype on subsequent pregnancy outcome in the setting of 17-P administration. Pharmacogenomics studies in obstetrics lag significantly behind those in other specialities. This may be attributed partly to difficulties with defining outcomes in obstetric populations, e.g. response to medication for recurrent PTB is a prime example. Though a prior PTB is a leading risk factor for a subsequent PTB, even without treatment, 30–70% of women may deliver at term; this number is widely variable based on patient population and other risk factors, including the number of prior PTB (s) and the gestational age(s) of the prior PTB(s). However, those with a recurrent PTB tend to experience this recurrence at a similar gestational age (70% deliver within 2 weeks of the gestational age of their previous delivery), and this is the basis for outcome Definition A.[31,32]

We defined our outcomes in two ways, and directly compared the results. Although outcome Definition B is more commonly used, the use of Definition A identifies women as having favourable outcomes if they gain significant gestational age with 17-P, even if they do not reach 37 weeks' gestation.[11,14] It is interesting that, regardless of pregnancy outcome definition, both analyses yielded biologically plausible results, and highlighted both pathways and individual genes previously associated with PTB. These results are somewhat surprising, though, because we expected the fact that all women had a prior PTB would have decreased the likelihood of detecting 'prematurity genes '. Given the different definitions, however, it is not unexpected that there are differences in the specific genes and pathways. There is significant overlap between genes within each pathway, as some genes are part of multiple pathways. CYP2E1 is only one example – it forms a part of the pathways that were highlighted for both responder definitions (Table S4).

Using Definition A, we noted that women with significant pregnancy prolongation were more likely to have had a history of cervical insufficiency in a prior pregnancy and were also more likely to have had cervical cerclage placed (although this difference did not reach statistical significance). Overall, however, this is of unclear significance given that the prevalence of a short cervix did not differ between the groups (Table 2), and may be attributed to the small numbers of women with cervical insufficiency. Nonetheless, PANTHER results for responder Definition A were notable for several collagen genes (COL4A1, COL9A2, COL18A1,

COL15A1) and interleukin signalling/ inflammatory genes (IL16, ELK1, IRAK4, PLCB1, MKNK1) (Table S2). Although these specific collagen and inflammatory genes have not previously been associated with cervical insufficiency per se, other similar genes (e.g, COL1A1, IL6) have.[33,34] Women with significant mean prolongation of pregnancy (Definition A) also tended to have a slightly earlier median gestational age of their earliest prior PTB (32 vs 34 weeks), which may be attributed to the definition itself. In contrast, there were no apparent differences in clinical phenotype between responders and non-responders using Definition B.

The differences in results based on outcome definition again highlight the challenges involved with pharmacogenomics research in prematurity and the difficulties in defining 'successful' outcomes. Our approach identifies broader, larger-scale genetic pathways that may be involved in the pathogenesis of response to 17-P, unlikely to be related to a single genetic variant or even a single pathway. This approach may provide the key first step for identifying the general area of genetic investigation that should be pursued to provide further insight into the pharmacogenomics of PTB prevention.

Our study should be interpreted with several limitations in mind. We did not have fetal DNA available to test the association between fetal genotype and progesterone response. Ultimately, we were limited by the small sample size, and our sample size limitation makes this an exploratory study. We made an *a priori* decision to include genes with a raw VAAST P-value of < 0.05 in our pathway analysis to balance inclusion of potentially causative variants (which might be missed by a more stringent P-value cutoff owing to sample size limitations) and including genes unlikely to be involved in the pharmacogenomics of 17-P for the prevention of recurrent PTB. It is possible that some of the results are associated with severe PTB phenotypes (recurrent, or early PTB) and not 17-P treatment. Ideally we would limit the study to women of one ancestry, and focus on those with a severe prior PTB history (e.g., very early PTB < 28 weeks). Our sample size also prevented us from directly examining the influence of gene–environment interactions on recurrent PTB and from performing subgroup analysis of women with specific PTB phenotypes.

Interpretation

A previous report using similar methods found an association between nitric oxide signalling pathways and response to 17-P for the prevention of recurrent prematurity.[11] In contrast, we did not specifically isolate the nitric oxide signalling pathway nor the NOS1 gene as a top candidate in our VAAST analysis. However, our analyses did identify several genes within broader gene pathways with a known role in nitric oxide response, metabolism or signalling (including OPRM1, CDH2, ATP2B4, CCR7, CCR8, MICAL3; see Tables S3 and S4). Nitric oxide is a ubiquitous gas

present in the body, and has crucial roles in cell signalling, regulation of inflammation and myometrial relaxation. It provides a biologically plausible alternative pathway for other therapeutic options to treat women who are identified as most likely to experience recurrent spontaneous PTB despite the administration of 17-P.

It is interesting that several of our top VAAST hits are among the leading PTB genetic candidates identified through the pathway analysis published by Uzun et al.[30] We would anticipate some concordance with 'prematurity genes' in this pharmacogenetic study, but because all women in this cohort had a prior PTB we thought that this would reduce the likelihood that we would detect 'prematurity genes'. We also note that several of the genes identified through the PANTHER pathway analysis (MICAL3, GSTZ1, AKR1C2) are involved in oxidoreductase activity, one of the two main pathways identified by Uzun et al.[30]

Our approach is unique in that we allow for the possibility of multiple variants within a single gene due to the scoring algorithms within the VAAST analysis programme. Concerns regarding which SNP best represents a given gene are irrelevant, because scoring occurs at the level of the gene. Our pathway approach considers the multifactorial nature of spontaneous PTB and decreases the burden of discovery of individual genes or haplotypes.

Conclusion

These exploratory results should be confirmed in larger studies, nevertheless they provide additional evidence implicating maternal genotype in pregnancy outcomes among women given 17-P for the prevention of recurrent prematurity. Some scientists have recently questioned the use of progestogens for PTB prevention. The clinician should be careful to note both the indication for its use and its formulation, as vaginal progesterone and 17-P do not appear to be interchangeable. Work such as this is important in identifying reasons for individual differences in outcomes between progestogen treatment and within progestogen treatment, a necessary step that must be taken in order to realise the ultimate goal of pharmacogenomics and individualised health care.

Disclosure of interest
Full disclosure of interests available to view online as supporting information.

Contributions to authorship
TAM conceived, planned and designed the study. TAM and WSW analysed the data. TAM wrote the first draft of the manuscript. TAM, WSW, MY, MWV, and LBJ contributed to data interpretation. All authors (TAM, WSW, MSE, JB, RB, SP, HZ, HH, WA, GS, YS, UMR, JI, MY, MWV, LBJ) contributed to data analysis, interpretation of the results and manuscript revisions. All authors (TAM, WSW, MSE, JB, RB, SP, HZ, HH, WA, GS, YS, UMR, JI, MY, MWV, LBJ) reviewed and approved the final manuscript.

Details of ethics approval
This study was approved by the institutional review board (IRB) at each centre, and written informed consent was obtained from all participants. IRB approval numbers as well as appropriate documentation are available upon request. This secondary analysis was also reviewed locally by the University of Utah IRB (IRB #44342, initially approved 9/17/2010).

Funding

■

Acknowledgements

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 Top 100 genes identified by VAAST.

Table S2 PANTHER results for outcome Definition A (women delivered later with 17P than without it).

Table S3 PANTHER results for outcome Definition B (women delivered at term with 17P).

Table S4 DAVID Results. All p-values were non-significant after Bonferroni correction.

References

1 Martin JA, Hamilton BE, Osterman MJ. Births in the United States, 2014. *NCHS Data Brief* 2015;216:1–8.

2 Simonsen SE, Lyon JL, Stanford JB, Porucznik CA, Esplin MS, Varner MW. Risk factors for recurrent preterm birth in multiparous Utah women: a historical cohort study. *BJOG* 2013;120:863–72.

3 Ananth CV, Getahun D, Peltier MR, Salihu HM, Vintzileos AM. Recurrence of spontaneous versus medically indicated preterm birth. *Am J Obstet Gynecol* 2006;195:643–50.

4 Meis PJ, Klebanoff M, Thom E, Dombrowski MP, Sibai B, Moawad AH, et al. Prevention of recurrent preterm delivery by 17 alpha-hydroxyprogesterone caproate. *New Engl J Med* 2003;348:2379–85.

5 Society for Maternal Fetal Medicine Publications Committee, et al. ACOG Committee Opinion number 419 October 2008 (replaces no. 291, November 2003). Use of progesterone to reduce preterm birth. *Obstet Gynecol* 2008;112:963–5.

6 Manuck TA. Pharmacogenomics of preterm birth prevention and treatment. *BJOG* 2016;123:368–75.

7 Blumenfeld YJ, Reynolds-May MF, Altman RB, El-Sayed YY. Maternal–fetal and neonatal pharmacogenomics: a review of current literature. *J Perinatol* 2010;30:571–9.

8 Manuck TA, Esplin MS, Biggio J, Bukowski R, Parry S, Zhang H, et al. The phenotype of spontaneous preterm birth: application of a clinical phenotyping tool. *Am J Obstet Gynecol* 2015;212:487.e1–11.

9 Gonzalez-Quintero VH, Cordova YC, Istwan NB, Tudela F, Rhea DJ, Romary LM, et al. Influence of gestational age and reason for prior preterm birth on rates of recurrent preterm delivery. *Am J Obstet Gynecol* 2011;205:275.e1–5.

10 Coleman S, Wallace L, Alexander J, Istwan N. Recurrent preterm birth in women treated with 17 alpha-hydroxyprogesterone caproate: the contribution of risk factors in the penultimate pregnancy. *J Matern Fetal Neonatal Med* 2012;25:1034–8.

11 Manuck TA, Watkins WS, Moore B, Esplin MS, Varner MW, Jackson GM, et al. Pharmacogenomics of 17-alpha hydroxyprogesterone caproate for recurrent preterm birth prevention. *Am J Obstet Gynecol* 2014;210:321.e1–21.

12 Manuck TA, Lai Y, Meis PJ, Dombrowski MP, Sibai B, Spong CY, et al. Progesterone receptor polymorphisms and clinical response to 17-alpha-hydroxyprogesterone caproate. *Am J Obstet Gynecol* 2011;205:135.e1–9.

13 Parry S, Zhang H, Biggio J, Bukowski R, Varner M, Xu Y, et al. Maternal serum serpin B7 is associated with early spontaneous preterm birth. *Am J Obstet Gynecol* 2014;211:678.e1–12.

14 Manuck TA, Esplin MS, Biggio J, Bukowski R, Parry S, Zhang H, et al. Predictors of response to 17-alpha hydroxyprogesterone caproate for prevention of recurrent spontaneous preterm birth. *Am J Obstet Gynecol* 2016;214:376.e1–8.

15 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009;25:1754–60.

16 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009;25:2078–9.

17 DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011;43:491–8.

18 McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* 2010;20:1297–303.

19 Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res* 2009;19:1655–64.

20 Hu H, Huff CD, Moore B, Flygare S, Reese MG, Yandell M. VAAST 2.0: improved variant classification and disease-gene identification using a conservation-controlled amino acid substitution matrix. *Genet Epidemiol* 2013;37:622–34.

21 Yandell M, Huff C, Hu H, Singleton M, Moore B, Xing J, et al. A probabilistic disease-gene finder for personal genomes. *Genome Res* 2011;21:1529–42.

22 Kennedy B, Kronenberg Z, Hu H, Moore B, Flygare S, Reese MG, et al. Using VAAST to Identify Disease-Associated Variants in Next-Generation Sequencing Data. *Curr Protoc Hum Genet* 2014;81:6.14.1–25.

23 Mi H, Muruganujan A, Thomas PD. PANTHER in 2013: modeling the evolution of gene function, and other gene attributes, in the context of phylogenetic trees. *Nucleic Acids Res* 2013;41(Database issue):D377–86.

24 Mi H, Thomas P. PANTHER pathway: an ontology-based pathway database coupled with data analysis tools. *Methods Mol Biol* 2009;563:123–40.

25 Thomas PD, Kejariwal A, Campbell MJ, Mi H, Diemer K, Guo N, et al. PANTHER: a browsable database of gene products organized by biological function, using curated protein family and subfamily classification. *Nucleic Acids Res* 2003;31:334–41.

26 Uzun A, Laliberte A, Parker J, Andrew C, Winterrowd E, Sharma S, et al. DBPTB: a database for preterm birth. *Database (Oxford)* 2012;2012:bar069.

27 da Huang W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* 2009;4:44–57.

28 Huang dW, Sherman BT, Zheng X, Yang J, Imamichi T, Stephens R, et al. Extracting biological meaning from large gene lists with DAVID. *Curr Protoc Bioinformatics* 2009; Chapter 13: Unit 13.11.

29 da Huang W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res* 2009;37:1–13.

30 Uzun A, Dewan AT, Istrail S, Padbury JF. Pathway-based genetic analysis of preterm birth. *Genomics* 2013;101:163–70.

31 Mercer BM, Goldenberg RL, Moawad AH, Meis PJ, Iams JD, Das AF, et al. The preterm prediction study: effect of gestational age and cause of preterm birth on subsequent obstetric outcome. National Institute of Child Health and Human Development Maternal-Fetal Medicine Units Network. *Am J Obstet Gynecol* 1999;181:1216–21.

32 Bloom SL, Yost NP, McIntire DD, Leveno KJ. Recurrence of preterm birth in singleton and twin pregnancies. *Obstet Gynecol* 2001;98:379–85.

33 Warren JE, Silver RM, Dalton J, Nelson LT, Branch DW, Porter TF. Collagen 1Alpha1 and transforming growth factor-beta polymorphisms in women with cervical insufficiency. *Obstet Gynecol* 2007;110:619–24.

34 Sundtoft I, Uldbjerg N, Steffensen R, Sommer S, Christiansen OB. Polymorphisms in Genes Coding for Cytokines, Mannose-Binding Lectin, Collagen Metabolism and Thrombophilia in Women with Cervical Insufficiency. *Gynecol Obstet Invest* 2016;81:15–22.