2016

GENETIC AND EPIGENETIC MECHANISMS OF COMPLEX REPRODUCTIVE DISORDERS

Bhavi P. Modi
Virginia Commonwealth University

GENETIC AND EPIGENETIC MECHANISMS OF COMPLEX REPRODUCTIVE DISORDERS

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

By

BHAVI P. MODI
Master of Science, Biotechnology, University of Mumbai, 2012

Advisor: Jerome F. Strauss III, M.D., Ph.D.
Dean of the School of Medicine
Professor, Dept. of Obstetrics and Gynecology and Dept. of Human and Molecular Genetics

Virginia Commonwealth University
Richmond, Virginia
December 2016

Acknowledgement

It is a pleasure to thank the many people who have made this dissertation possible and because of whom I will cherish my graduate experience forever.

First and foremost, I would like to express my sincere gratitude to my advisor and mentor, Dr. Jerome F. Strauss for giving me the opportunity to work under his guidance. I appreciate all of the time and effort he has put in training me as an independent research scientist. He made sure I learnt as much as I could and widened my expertise beyond the limitations of graduate requirements. He will always be a positive research role model for me.

I am sincerely grateful to the members of my graduate committee including Dr. John Nestler, Dr. Joyce Lloyd, Dr. Rita Shiang and Dr. Kellie Archer. Their individual expertise and willingness to invest time in my training have been a huge help. Together they ensured that I got a well-rounded graduate experience and will always be an inspiration to hold myself to high research standards.

I would also like to express my appreciation to our collaborators, Dr. Jan M. McAllister, Dr. Roberto Romero, Dr. Timothy P. York, Dr. Laurel Pearson, Dr. Colleen Jackson-Cook, Dr. Scott Walsh and Sonya Washington for providing samples, helping with data and statistical analysis, providing their intellectual insight and feedback on the manuscripts that have resulted from this work.

I am incredibly grateful for my work family – members of the Strauss Lab for being the best colleagues I could have asked for. Dr. Maria Teves and Dr. Eun Lee have taught me a lot that I know, have not only guided me with the science, but have also been amazing friends and confidants. It has been a pleasure working alongside all other past and current lab members and colleagues, especially DaShaunda Hillard.

Pursuing a Ph.D. alone in a new country is difficult in more ways than I can express and it would have been impossible without the love and support of my family. I dedicate this thesis to my parents, Pinesh and Jasmina Modi, my sister Rashi, and my boyfriend, Nitish Nayak for being my strength and support system from miles away. They have always stood by my side, encouraged me in my lowest times and cheered the loudest with each achievement along the way. Their love and relentless faith in my capabilities has been the strongest motivation in completing this work.

Table of Contents

List of Tables

List of Figures

Page

List of Abbreviations

AMH: anti-mullerian hormone

CYP17A1: cytochrome P450, family 17, subfamily A, polypeptide 1

C-21: Chromosome 21

DENND1A: differentially expressed in normal and neoplastic cells domain containing protein 1A.

DHEA: dehydroepiandrosterone

GWAS: genome-wide association studies

miRNA: microRNA

PCOS: polycystic ovary syndrome

PCR: polymerase chain reaction

PPROM: preterm premature rupture of membranes

PTB: preterm birth

sPTB: spontaneous preterm birth

SNP: single nucleotide polymorphism

T21: Trisomy 21

T2D(M): type II diabetes (mellitus)

TBP: TATA Binding Protein

WES: whole exome sequencing

Abstract

GENETIC AND EPIGENETIC MECHANISMS OF COMPLEX REPRODUCTIVE DISORDERS

By, Bhavi P. Modi, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2016

Major Director: Jerome F. Strauss III, M.D., Ph.D.
Dean of the School of Medicine
Professor, Dept. of Obstetrics and Gynecology; Dept. of Human and Molecular Genetics

Common, complex disorders are usually polygenic, heterogeneous and multifactorial traits representing interactions between environmental, genetic and epigenetic factors. Next generation genomic technologies have uncovered risk factors for several complex traits. More often than not, contributions of these risk factors have been studied individually. The application of these technologies to the study of complex female reproductive traits has been challenging. This thesis explores the potential of genetic and epigenetic components contributing to a better understanding of the

biological pathways underlying disease risk in two specific female complex reproductive traits - polycystic ovary syndrome (PCOS) and preterm premature rupture of membranes (PPROM). After introducing the clinical context and the genetic architecture for each disorder, I describe individual projects that together aim at understanding the genetic and epigenetic regulation of genes associated with disease susceptibility. The PCOS projects focus on characterization of a gene, *DENND1A*, whose association to PCOS was first established by Genome Wide Association Studies (GWAS) and one of its isoforms – DENND1A.V2, is known to functionally contribute to the PCOS pathology via increased androgen production. This study showed that decreased miR-125a-3p expression is correlated to increased DENND1A.V2 and increased DHEA accumulation in PCOS theca cells. A common network of differentially regulated miRNAs identified in this study and known PCOS candidate loci was also identified. The studies on PPROM utilize a Whole Exome Sequencing approach to identify rare variants in fetal genes contributing to extracellular matrix composition and synthesis contributing to PPROM risk. The results suggest that fetal contribution to PPROM is polygenic and is driven by a significant genetic burden of potentially damaging rare variants in genes contributing to fetal membrane strength and integrity. Tissue and location specific expression patterns of the Chromosome 21 miRNA cluster (miR-99a, miR-125b, let-7c) in fetal membranes from term pregnancies with spontaneous rupture were investigated. Unique expression profiles for all three miRNAs between the rupture and non-rupture sites were identified. The miRNAs were also differentially expressed between trisomy 21 (T21) and euploid control amniocyte cultures. The results suggest that these miRNAs play

potential roles in fetal membrane rupture and fetal membrane defects associated with T21.

Keywords: Polycystic Ovary Syndrome, DENND1A, Preterm Premature Rupture of Membranes, microRNA.

Chapter 1: Introduction and Background

1.1. Introduction

The genetic architecture of complex genetic disorders is represented by interactions between environmental, genetic and epigenetic factors. In contrast with single gene Mendelian disorders, single genetic variants are insufficient to cause complex genetic disorders. Instead, multiple variants of low penetrance at multiple loci act synergistically to contribute to complex disease risk. While environmental factors can be modified, the genetic factors predisposing to disease risk remain stable and need to be determined. The approaches used to study complex disease genetics have evolved rapidly with the development of new genomic technologies – from family-based studies and linkage analysis, twin studies, candidate-gene studies, to population-based techniques using microarray and next-generation sequencing approaches such as Genome Wide Association Studies (GWAS) and Whole Exome Sequencing (WES) (1,2).

Genetic contributions to female reproductive disorders and pregnancy complications have been established but the precise genetic etiology contributing to disease mechanism is not yet completely understood. These disorders typically manifest as heterogeneous and complex traits resulting from multiple genetic and environmental factors (3). The progress of genetic studies in these disorders has been slow as compared to other complex genetic disorders such as schizophrenia (4) and coronary artery disease (5). This thesis describes the studies designed and performed

to understand the genetic and epigenetic mechanisms of complex reproductive disorders, with specific focus on polycystic ovary syndrome (PCOS) and preterm premature rupture of membranes (PPROM).

Demonstrating direct causation in any complex genetic trait is almost always impossible, and as discussed in the sections below, it is even more challenging for disorders of female reproduction. The research approaches utilized to identify genetic variants should be aimed at gaining evidence that variants at least show a consistent correlation with gene expression or gene regulation that is highly correlated with the human trait originally assayed (6). Together, the two models present opportunities to study complex disease genetics using three broad approaches. The studies on PCOS (Chapter 2) essentially focus on the characterization of a trait (PCOS) - associated locus originally identified by GWAS, but whose association to PCOS has been securely implicated. The studies on PPROM (Chapter 4) utilize a WES approach to identify rare variants contributing to PPROM risk, focusing on variant prioritization for replication and follow up in genes with biological relevance to PPROM. In addition the role of miRNAs in epigenetic mechanisms of gene regulation in both PCOS (Chapter 3) and PPROM (Chapter 5) is investigated.

The following sections describe the clinical overview, genetic architecture and the ongoing research and its limitations in both PCOS and PPROM followed by a brief chapter-wise thesis outline.

1.2. PCOS background

1.2.1. Clinical Overview of PCOS

PCOS is a common reproductive endocrinological disorder that affects 5% - 15% women of reproductive age and is the leading cause of female infertility worldwide (7,8). It is a multisystem, heterogeneous disorder characterized by a broad spectrum of reproductive, metabolic and endocrinological features. The typical ovarian morphology, which gives the disorder its name, consists of multiple, small, subcortical follicular cysts in bilaterally enlarged ovaries. (Figure 1.1) Other features include irregular periods (oligo/ amenorrhea), oligo/ anovulation, infertility, hyperandrogenism, hirsutism, acne, insulin resistance and weight-gain (9). PCOS women have higher prevalence rates of cardiovascular disease, Type 2 Diabetes Mellitus (T2DM) and metabolic syndrome, irrespective of their age and body mass index (10,11).

PCOS is one of the most ill-defined endocrinological conditions with considerably debated diagnosis schemes and varying prevalence rates based on the diagnostic criteria applied (7). Three different classification systems have been used for PCOS diagnosis over the years – the National Institutes of Health (NIH) criteria (12), the Rotterdam criteria (13), and the Androgen Excess and PCOS Society criteria (14) (Figure 1.2). This shows that the existing diagnostic criteria are based on expert consensus of clinical findings as opposed to strong molecular evidence and these are limited to reproductive-aged women. Several symptoms, especially reproductive and some endocrine features improve over time, however associated metabolic dysfunction persists throughout a woman's lifetime further adding to risk for cardiovascular disease. Recently, the use of serum Anti-Mullerian hormone (AMH) levels as a biomarker for

PCOS diagnosis has been considered (15,16). In addition, none of these criteria include other endocrinopathies and metabolic features that are commonly observed in PCOS patients such as abnormal glucose metabolism, insulin resistance, dyslipidemia, weight gain and obesity. Development of molecular diagnostic techniques for PCOS will aid in early diagnosis in adolescents and pre-pubertal girls as opposed to late symptomatic diagnosis or more invasive procedures such as trans-vaginal ultrasound, and this can help prevent complications through lifestyle choices when possible for e.g. by controlling diet, exercise etc. Thus, accurate and early diagnosis of PCOS is essential not only to prevent potentially life-long associated comorbidities, but to also reduce the healthcare related economic burden of PCOS. The US health care system spends around 4 billion dollars annually in diagnosis and treatment of the primary PCOS symptoms as well as associated morbidities in pre-menopausal women (17). The hidden costs of life-long sequelae resulting from a PCOS background are difficult to estimate. It is also important to note that male and female first-degree relatives of women with PCOS have increased risk of developing associated metabolic and endocrine conditions including hyperandrogenism, obesity, insulin-resistance and T2DM (18,19). This underlines the genetic complexity of PCOS and further stresses the need to identify candidate genes.

Most consensus diagnostic schemes recommend the presence of hyperandrogenism (clinical or biochemical) a must for concluding a PCOS diagnosis. A cross-sectional case-control study suggested that hyperandrogenism is implicated in both the metabolic and reproductive morbidities of PCOS and maybe a common link between the two (9,20). Ovarian androgen production is increased in PCOS and studies show that ovarian theca cells from PCOS women secrete more androgen than theca from

regularly ovulating women (21,22). Thus, hyperandrogenism is considered to be the hallmark of PCOS, which is directly or indirectly, linked to several other PCOS symptoms and is central to understanding PCOS etiology.

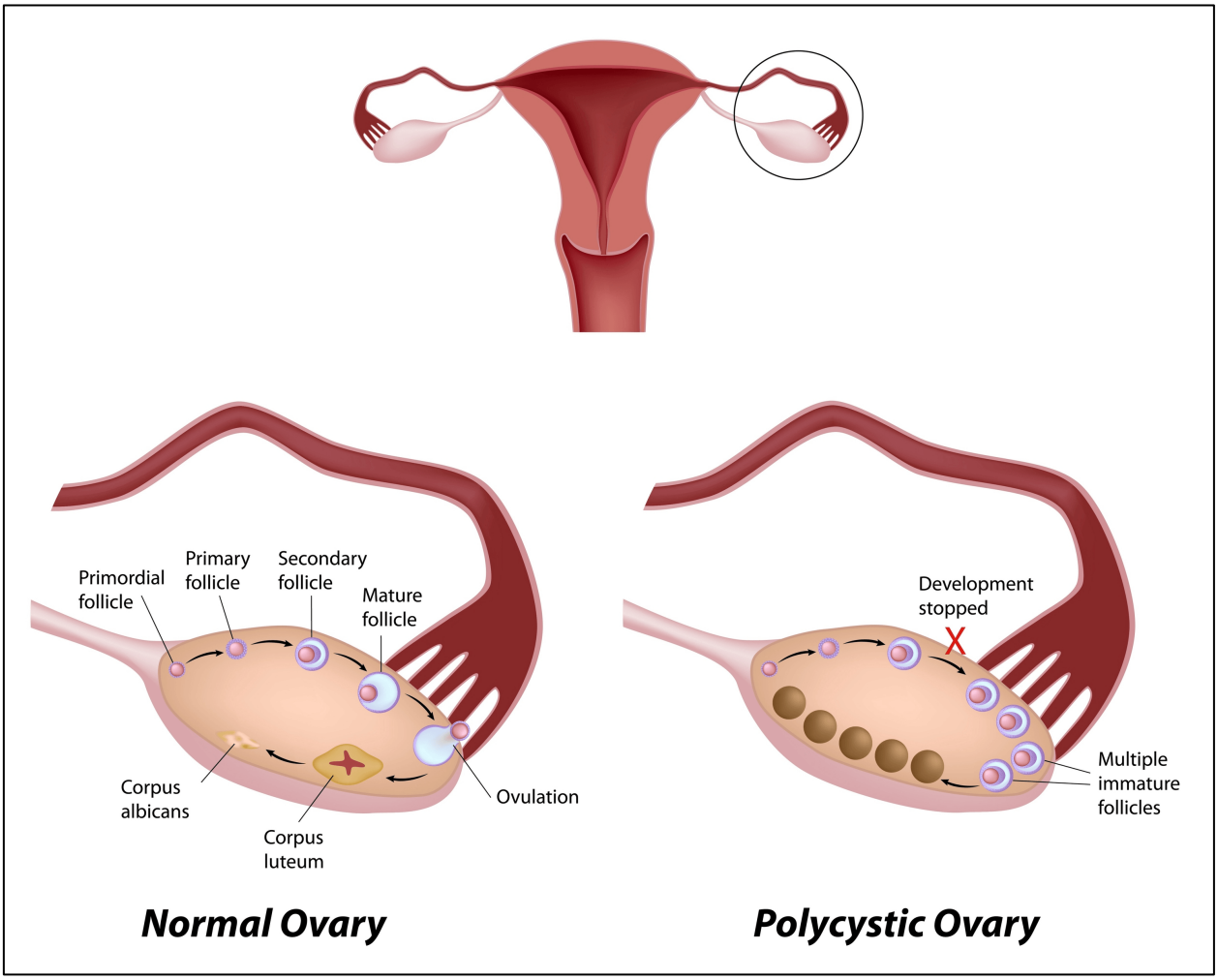


Figure 1.1. Normal vs. Polycystic Ovary: Normal ovulation follows a sequential series of developmental events in the follicles resulting in release of mature egg (ovulation). In a polycystic ovary, follicular development is arrested, resulting in accumulation of multiple immature follicles (< 10 mm), which overtime, appear as multiple small, subcortical follicular cysts.

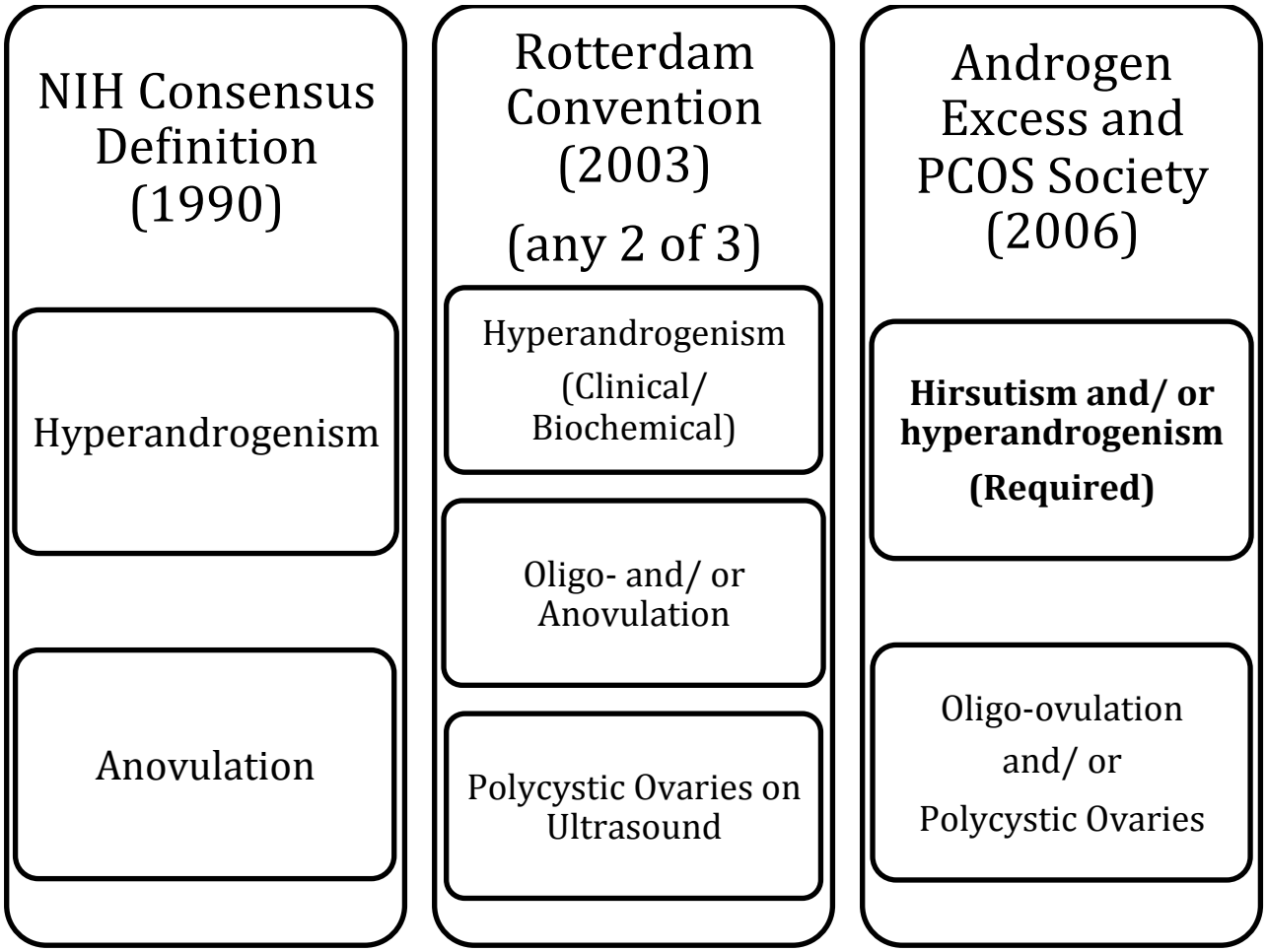


Figure 1.2: PCOS diagnostic criteria. Three commonly used PCOS diagnostic criteria are described. A PCOS diagnosis requires the exclusion of other diseases that may contribute to these clinical criteria used including 21-hydroxylase deficiency, Cushing's syndrome, and hyperprolactinemia.

1.2.2. Genetic architecture of PCOS

Initial evidence for genetic predisposition in PCOS was based on familial clustering of cases (23). The disorder was first thought to follow an autosomal dominant mode of inheritance with variable penetrance (8). Association and linkage studies and numerous candidate gene studies, though largely unsuccessful in yielding statistically significant and replicable results, provided evidence that a number of genes are differentially expressed in PCOS, which lend support to a polygenic mode of inheritance as opposed to a single-gene Mendelian model. Now, it is well established that PCOS is a complex genetic disorder with multiple genetic and environmental factors interacting and contributing to disease pathology (24,25). Twin studies and family-based association studies carried out on a Dutch population including more than 1300 monozygotic twins and 1800 dizygotic twins/ singleton sisters of twins showed that genetic factors account for ~70% of heritability in PCOS (26). The clinical features used for PCOS diagnosis, including polycystic ovaries, hyperandrogenism as well as oligo-amenorrhea have hereditary components with characteristics suggestive of an autosomal dominant trait (26–29). Hyperandrogenism is shown to be the strongest heritable characteristic in families with multiple PCOS patients (24,30). Other features commonly found in association with PCOS including metabolic syndrome, T2DM and obesity are also heritable characteristics (10,31,32). Environmental factors that may contribute include intra-uterine factors such as prenatal nutrition and/ or androgen exposure, while major postnatal factors are related to a person's lifestyle such as diet, exercise, abdominal fat and body mass index (8).

The genetic architecture of PCOS has been summarized by Rosenfield *et al* (8) – *"At its simplest, this is a "two-hit" hypothesis that can be thought of because of a congenitally programmed predisposition ("first hit") that becomes manifest upon exposure to a provocative environmental factor ("second hit")".* The congenital "hit" consists mostly of genetic variants affecting hyperandrogenism or ovarian function but can also be acquired factors such as intra-uterine nutrition or androgen exposure *in utero*. The provocative "hits" are postnatal factors including obesity and insulin resistance that may be caused by T2DM related gene variants.

1.2.3. DENND1A: From GWAS to molecular mechanisms

Candidate gene association studies in PCOS were primarily focused on genes involved in folliculogenesis, steroidogenesis, insulin production and function. Owing to small sample sizes and resulting limited statistical power, most of them failed to yield statistically significant results that could be replicated (24,25,33). The first successful GWAS in PCOS was published in 2011 on a Han-Chinese population with a follow-up GWAS published in 2012 (34,35). Subsequent GWAS on European populations have confirmed associations of some (as shown in Table 1.1) of the variants identified in the Han-Chinese GWAS to PCOS, and also identified additional candidate loci (36,37). Together, the four GWA studies have identified 22 candidate loci for PCOS (Table 1.1). Interestingly, association of the *DENND1A* gene locus (Chromosome 9: 9p33.3) to PCOS has been consistently replicated in several studies (38–40). In addition, association to related phenotypes including elevated serum glucose levels, insulin levels and body weight has been documented (41).

As described in greater detail in chapter 2 studies in theca cells from normal cycling women and women with PCOS provide evidence for functional association of increased expression of an alternatively spliced isoform of the *DENND1A* gene - DENND1A.V2 (V2) to PCOS. While showing potential for V2 to be utilized as a diagnostic and therapeutic target, this study does not determine the mechanisms of V2 overexpression in PCOS theca cells (42). The *DENND1A* SNPs (single nucleotide polymorphisms) - identified in Han Chinese GWAS – rs10818854, rs2479106, rs10986105 - are located in the intronic regions of the gene (34,35). The *DENND1A* signal reported (suggestive locus – rs10760321) in European population is intergenic and was not confirmed in follow-up samples (37).

The genetic mechanisms contributing to V2 overexpression need to be studied. Moreover, the reported odds ratios of association of the gene with PCOS are low to modest, which suggests variations in *DENND1A* and the resultant functional consequences of V2 overexpression alone cannot explain the heritability of PCOS. Considering PCOS is a polygenic disorder and the known functional significance of *DENND1A* in PCOS etiology, it is more likely that other genes, possibly in a common network with *DENND1A* contribute to PCOS pathophysiology.

Table 1.1. PCOS risk loci reported to date. Genetic variants associated with PCOS risk identified in Han Chinese (CHN) and European (EUR) populations are described.

Locus	Gene	GWAS Index SNP	Risk/ Other allele	Discovery p-value	Discovery population	Ref
2p16.3	FSHR	rs2268361	T/C	9.89 x 10-13	CHN	35
		rs2349415	T/C	2.35 x 10-12	CHN	35
2p16.3	LHCGR	rs13405728	G/A	7.55 x 10-21	CHN	34
2p21	THADA	rs13429458	C/A	1.73 x 10-23	CHN	34
		rs12478601	T/C	3.48 x 10-23	CHN	34
		rs12468394	A/C	1.59 x 10-20	CHN	34
		rs7563201	G/A	3.3 x 10-10	EUR	36
2q34	ERBB4	rs1351592	G/C	1.2 x 10-12	EUR	36
5q31.1	RAD50	rs13164856	T/C	3.5 x 10-9	EUR	36
8p23.1	GATA4/NEIL2	rs804279	A/T	8 x 10-10	EUR	37
9q22.32	C9orf3	rs4385527	A/G	5.87 x 10-9	CHN	35
		rs3802457	A/G	5.28 x 10-14	CHN	35
		rs10993397	C/T	4.6 x 10-13	EUR	37
9q33.3	DENND1A	rs2479106	G/A	8.12 x 10-19	CHN	34
		rs10818854	A/G	9.4 x 10-18	CHN	34
		rs10986105	C/A	6.9 x 10-15	CHN	34
		rs10760321	A/G	1.4 x 10-6	EUR	36
11p14.1	KCNA4/FSHB	rs11031006	G/A	1.9 x 10-8	EUR	37
11q22.1	YAP1	rs1894116	G/A	1.08 x 10-22	CHN	35
		rs11225154	A/G	7.6 x 10-11	EUR	36
12q13.2	RAB5B/SUOX	rs705702	G/A	8.64 x 10-26	CHN	35
12q13.2	ERBB3	rs7312770	C/T	2.1 x 10-7	EUR	36
12q14.3	HMGA2	rs2272046	C/A	1.95 x 10-21	CHN	35
12q21.2	KRR1	rs1275468	C/T	1.9 x 10-8	EUR	24
16q12.1	TOX3	rs4784165	G/ T	3.64 x 10-11	CHN	35
17q12	ERBB2	rs7218361	A/G	9.6 x 10-7	EUR	36
19p13.2	INSR	rs2059807	G/A	1.09 x 10-8	CHN	35
20q13.2	SUMO1P1/ ZNF217	rs6022786	A/G	1.83 x 10-9	CHN	35

1.3. PPROM Background

1.3.1. Clinical Overview of PPROM

In humans, the normal term gestational period for singleton pregnancies is 37 – 41 weeks, with gestational age determined from the first day of the last menstrual period or early second trimester estimation for confirming gestational age (43). Maternal and fetal physiological changes at term promote uterine contractions and breakdown of fetal membranes (amnion and chorion) to facilitate parturition. Variety of factors can contribute to early initiation or changes in these pathways resulting in preterm birth (PTB), which is defined as birth prior to 37 weeks of gestation. PTB is the leading cause of neonatal mortality and infant morbidity occurring in ~12% of all pregnancies in the United States (44,45). Children who are born preterm have increased risks of several short and long-term health sequelae owing to systemic immaturity. Immediate health complications include respiratory distress (due to poorly developed lungs), intraventricular hemorrhage, jaundice, anemia and infection. Long-term health impacts can include intellectual and developmental disabilities, cerebral palsy, lung problems, vision and hearing loss. In addition, children who are born preterm also have an increased risk of adult-onset diseases such as obesity, diabetes and hypertension (www.marchofdimes.com/prematurebabies). Despite the advances in research, healthcare and increasing knowledge of risk factors leading to PTB, the rate of PTB in US has not decreased in the last few decades (46,47).

There are two major proximate causes for PTB – spontaneous PTB (sPTB) with intact membranes and preterm premature rupture of membranes (PPROM). Preterm premature rupture of membranes (PPROM) is characterized by rupture of the fetal

membranes prior to the onset of labor before 37 weeks of gestation. It occurs in 2% - 3% of all pregnancies and accounts for approximately 40% of all PTBs (48). PPROM occurs at higher frequencies among African-American women compared to other populations in the US and these population-specific disparities cannot be explained by socio-economic factors. Genetic variation and therefore gene-environment interactions are thought to be involved (49,50).

Diagnosis of PPROM is made based on visual pooling of clear amniotic fluid in the cervix, confirmation of alkaline pH of the cervicovaginal discharge (amniotic fluid) which is typically done using a nitrazine (pH indicator) test and/ or a fern test, which is microscopic observation of ferning patterns of amniotic fluid on drying. Using these clinical signs have limitations in terms of accuracy as most of these tests become less accurate more than 1 hour after membrane rupture (51).

Understanding the genetic factors contributing to PPROM risk can help us identify early on in pregnancy, which women have an increased risk for PPROM. This can help identify the healthcare management options to prevent PPROM (by controlling other risk factors) and to assure best possible perinatal outcomes.

1.3.2. Genetic Architecture of PTB/ PPROM

Genetic contribution to PTB was initially widely debated with suggestions that variations in PTB risk including population-specific differences can be attributed to socioeconomic (education, access to healthcare), behavioral (smoking, alcohol and drug use), and environmental factors (stress). However, more studies demonstrated that these factors do not completely account for the risk of PTB in families and even after controlling for

these factors, PTB risk (specifically PPROM) is greater among African-Americans (49,50). This indicated that genetic factors likely contribute to PTB risk. Supporting this were reports indicating increased frequency of PTB in mothers with prior history of PTB and correlations observed between mother-daughter, sister-sister relationships within the same families. This is true even for PPROM. (52,53). These correlation patterns can be explained by both genetic as well as environmental factors shared within families and indicate that PTB and its specific sub-phenotype PPROM are complex genetic disorders. However, determining the genetic architecture is difficult because in addition to the heterogeneous, multifactorial, complex genetic nature of the disorder, the involvement of not one, but two genomes has to be considered – maternal and fetal. The contribution of fetal genetic factors have been estimated to range from 11 – 35%, whereas the contribution of maternal genetic factors is in the range of 13 – 20% (54).

Gene x environment interactions are considerably more challenging to understand in PTB as compared to other complex genetic disorders due to the interaction between the *in utero* environmental effects of the maternal genotype and the developmental effect of the fetal genotype.

Environmental factors contributing to PTB risk include familial sources of variation that would, in theory, affect all births of the same mother such as socioeconomic status, stress, smoking, congenital maternal health problems, and also those environmental factors that are unique to individual pregnancies such as infections (Figure 1.3).

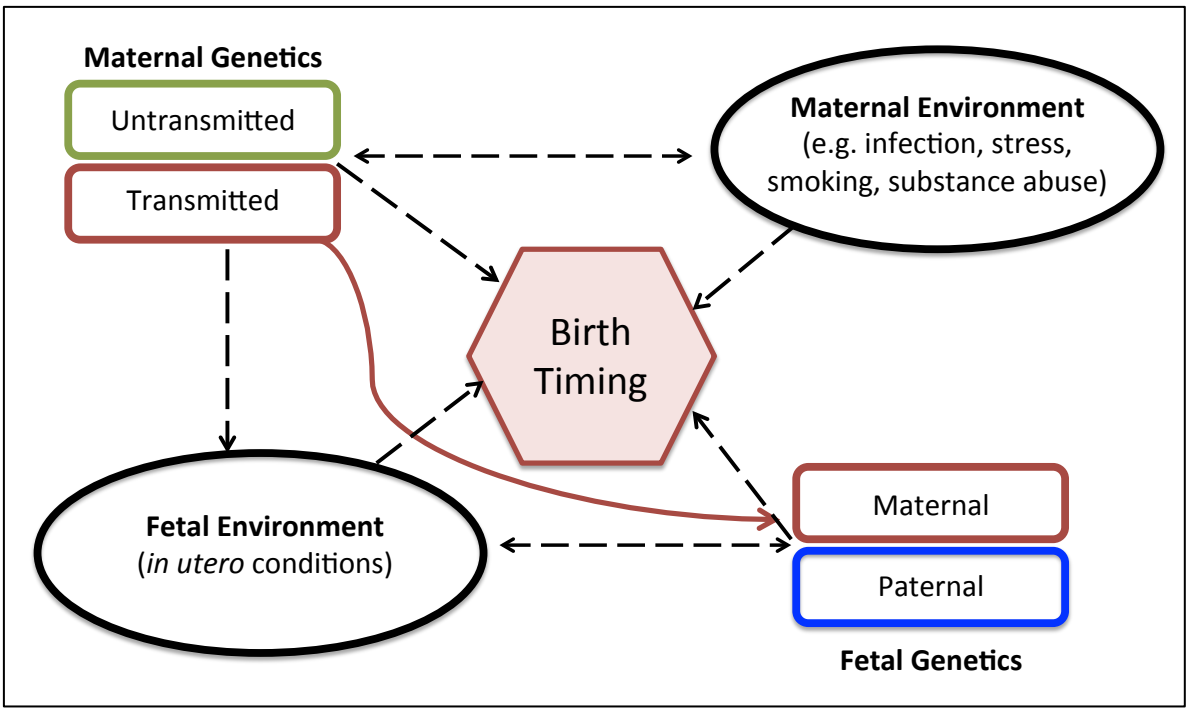


Figure 1.3. Genetic Architecture of Birth Timing. The complex genetic architecture of PPROM is influenced by both maternal and genetic factors. Maternal genetic factors influence the fetal genetic factors as well as the fetal environment *in utero*. In addition, gene x environment interactions exist between maternal genetic factors and general maternal environment such as nutrition status, smoking and substance abuse and pregnancy specific maternal environment such as acquired infections during pregnancy.

1.3.3. Rare Variant Analysis in fetal collagen genes contributing to PPROM risk

GWAS has had very limited success in identifying loci associated to PTB. Identified loci lose significance after testing for multiple corrections or are not reproducible in follow up studies. This can be attributed to several limitations in study design such as limited sample size, inadequate power, population heterogeneity and variations in definitions of PTB including the proximate cause of PTB used for each study, which makes replication of results more difficult. (46,55,56). It is important to recognize that different sub-phenotypes of PTB (sPTB or PPROM) may result from different pathways, and can even have varying contributions of participating genomes (maternal vs. fetal) towards the genetic risk. The poor outcome from standard genetic approaches used to study PTB suggests that other methods need to be utilized to identify the genetic contribution to PTB. There have been no PTB GWAS focusing specifically on PPROM as the etiology leading to PTB. The majority of studies have focused on candidate genes in the maternal genome involved in inflammatory pathways such as Tumor Necrosis Factor-alpha (*TNFA),* Toll-like receptor 4 *(TLR4)* and Caspase Recruiting Domain protein - 15 (*CARD15)* (57,58). Despite evidence of fetal contribution to PPROM established by twin studies, not many studies have investigated variations in fetal genes contributing to PPROM (54,59). As suggested earlier, the pathways and genomes contributing to PPROM can be vastly different to those contributing to sPTB. This also has been suggested in a recent study investigating the pathway analysis of genetic factors associated with the two PTB subtypes (60). Moreover, all studies on PTB in general have focused on identifying common variants contributing to preterm birth risk. A study of rare variants contributing to birth timing has not been performed. As described in

Chapter 4, this thesis describes a unique approach towards identification of rare variants in the fetal genome in select genes that are known to contribute to fetal membrane structure and integrity.

1.4. MicroRNA regulation in PCOS and PPROM

GWAS and WES are approaches used to investigate alterations within the DNA sequence (common and/ or rare variants) contributing to disease association or disease risk. In contrast, epigenetic approaches are used in the study of complex genetic diseases to investigate heritable gene expression changes that occur without alterations of the underlying DNA sequences and contribute to disease mechanisms. Commonly studied epigenetic changes include DNA modifications such as methylation and histone modifications.

MiRNAs are small (20 – 24 nucleotides), single-stranded, non-coding RNA molecules involved in post-transcriptional regulation of gene expression via translation inhibition or mRNA degradation (61,62). Since miRNAs induce gene expression changes without altering underlying DNA sequences, they contribute to epigenetic mechanisms of gene expression regulation. Normal ovarian function and normal parturition require a series of closely coordinated events that require tightly controlled gene regulation. Several investigators have studied and identified the role of miRNAs in regulating the normal functioning of these pathways including folliculogenesis, androgen regulation, contributing to preterm birth and preeclampsia (63–65). PCOS as well as PPROM are perturbations of the normal ovarian and parturition or birth timing pathways and this led

us to the hypothesis that differential miRNA expression patterns could contribute to the epigenetic landscape of disease mechanisms for these two disorders.

In addition, the study of miRNAs in female reproductive disorders and adverse pregnancy outcomes is especially exciting because of accessibility of miRNAs in the serum, plasma, urine and saliva either as free circulating forms or encapsulated in vesicles, making them excellent non-invasive diagnostic and prognostic biomarkers of disease. This is particularly helpful in PCOS and PPROM where non-invasive accessibility of disease relevant organs and tissues is difficult or impossible, making diagnosis and following of disease progression extremely challenging.

1.5. Thesis Outline

Chapter 2 describes the genetic mechanisms for V2 overexpression in PCOS theca cells via WES analysis, copy number variation assessment and analysis of genetic variants in the intronic region where alternative splicing of DENND1A gene is thought to occur to produce V2. None of these mechanisms provide concrete evidence of contributing to V2 overexpression in PCOS, suggesting the potential role of miRNA regulation.

In Chapter 3 the epigenetic mechanisms of miRNA contribution to PCOS etiology via miRNA regulation are studied. Using miRNA deep sequencing, the study shows evidence for differential microRNA expression in PCOS theca cells as compared to unaffected control theca cells. It describes the potential role of miR-125a-3p in regulating DENND1A.V2 gene expression and contributing to hyperandrogenism in the PCOS theca.

Chapter 4 describes a whole exome sequencing approach in African Americans to identify rare genetic variants in fetal genes encoding fibrillar collagen and enzymes and chaperone proteins involved in collagen synthesis contributing to increased risk for PPROM. The findings suggest that the fetal contribution to PPROM is polygenic, and driven by an increased burden of rare variants that may also contribute to the disparities in rates of preterm birth among African Americans.

Chapter 5 investigates tissue and location specific expression patterns of three miRNAs (miR-99a, miR-125b and let-7c) derived from Chromosome 21 in chorioamniotic membranes from term pregnancies with spontaneous rupture. Results suggest the Chr – 21 derived miRNAs play a potential role in fetal membrane rupture and fetal membrane defects associated with Trisomy 21.

Chapter 6 discusses the connecting themes underlying the dissertation work in Chapters 2 – 5 and comments on future directions.

Chapter 2: Genetic mechanisms of DENND1A.V2 overexpression in PCOS theca cells.

Please note: Part of the material in this chapter has been published in

i) McAllister JM, **Modi B**, Miller BA, Biegler J, Bruggeman R, Legro RS, Strauss JF 3rd. *Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype*. Proc Natl Acad Sci USA. 2014; 111(15): E1519-27

ii) Tee MK Speek M, Legeza B, **Modi B**, Teves ME, McAllister JM, Strauss JF 3rd, Miller WL. *Alternative splicing of DENND1A, a PCOS candidate gene, generates variant 2*. Mol Cell Endocrinol. 2016; 434: 25 – 35

Abstract

Identification of DENND1A as a PCOS candidate gene in Han Chinese and European Genome Wide Association Studies, replication in follow-up association studies and evidence for overexpression of one of the isoforms – DENND1A.V2 (V2) - in PCOS theca cells responsible for their steroidogenic phenotype, makes the gene the most potential key factor in the molecular pathogenesis of PCOS. In order to completely understand the functional significance of this gene and its contribution to PCOS, it is first important to elucidate the exact mechanism by which V2 is produced and

overexpressed in PCOS theca cells. The present study indicates that V2 is produced by the exonization of intronic regions between exons 20 and 21 within the DENND1A gene. In addition, genetic mechanisms including coding/ splice site variations, copy number changes as well as intronic region variations were investigated and no evidence for their contribution to overexpression of V2 in PCOS theca cells was found.

2.1. Introduction

Polycystic Ovary Syndrome (PCOS) affects 5% - 15% women of reproductive age across different ethnicities (7,8). The disorder is associated with a typical ovarian morphology of multiple, small, subcortical follicular cysts in bilaterally enlarged ovaries, irregular periods (oligo/ amenorrhea), oligo/ anovulation, infertility, hyperandrogenism, insulin resistance and weight-gain (9). It is a heterogeneous disorder with a strong evidence for a genetic component, either autosomal dominant or oligogenic/ polygenic (25,30). Incomplete penetrance, epigenetic modification, and environmental contributions have hindered attempts to clarify the underlying mode of inheritance. Numerous candidate gene association studies have been conducted, but few have yielded statistically significant associations that have been replicated. Genome wide association studies (GWAS) with large populations of Han Chinese individuals identified eleven PCOS candidate loci, including *DENND1A* (34,35). Additional studies in European populations confirmed the association of *DENND1A* with PCOS strongly supporting *DENND1A* as a PCOS candidate gene (Chapter 1, Table 1.1) (36–40). In addition, association to related phenotypes including elevated serum glucose levels, insulin levels and body weight has been documented (41).

The 18 human genes in the DENND family encode proteins that function as Rab-specific GTPases involved in membrane trafficking. The *DENND1A* gene consists of 22 exons and encodes the protein connecdenn 1, which has a clathrin-binding domain thought to facilitate endocytosis (66). *DENND1A* yields two principal transcripts: variant 1 (V1), encodes a 1009 AA protein with three DENN (differentially expressed in normal and neoplastic cells) domains and a C-terminal proline-rich domain; variant 2 (V2), encodes a 559 AA protein that contains the three DENN domains, but lacks the proline-rich domain, and includes a C-terminal 33 AA sequence that differs from V1 (Figure 2.1). Recent work has directly implicated V2 in the pathophysiology of PCOS: V2 mRNA and protein were overexpressed in PCOS theca cells compared to unaffected control theca cells; over-expression of V2 in unaffected control theca cells increased the abundance of cytochrome P450, family 17, subfamily A, polypeptide 1 (*CYP17A1)* mRNA, augmented dehydroepiandrosterone (DHEA) production and increased the activity of transfected *CYP11A1* and *CYP17A1* promoter-reporter constructs; and siRNA knockdown of V2 in PCOS theca cells reversed the PCOS phenotype. In addition, V2 mRNA was abundant in urinary exosomes of PCOS women but not normally-cycling women (42). These results suggest the diagnostic potential of V2. Especially, a diagnostic method based on urine exosomal RNA would have a great potential as non-invasive and early diagnosis tool, which will be helpful for early detection in prepubertal adolescent females (67). However, in order to completely understand the functional significance of this gene and its contribution to PCOS and to better exploit its use as potential diagnostic tool, it is important to elucidate the exact mechanism by which V2 is produced and overexpressed in PCOS theca cells. The SNPs in *DENND1A* identified by

GWAS are located in introns, and lack apparent functions that would favor splicing of V2 over V1.

As suggested previously overexpression of V2 in PCOS suggests the contribution of gain-of-function variation (42). Thus, we sought to determine the mechanisms leading to production of V2 and its overexpression in PCOS by testing several possible mechanisms including examination of coding or splicing variants in the *DENND1A* gene, investigation of copy number variations in the *DENND1A* gene and sequencing of the intronic region that may explain alternative splicing of V1 and V2.

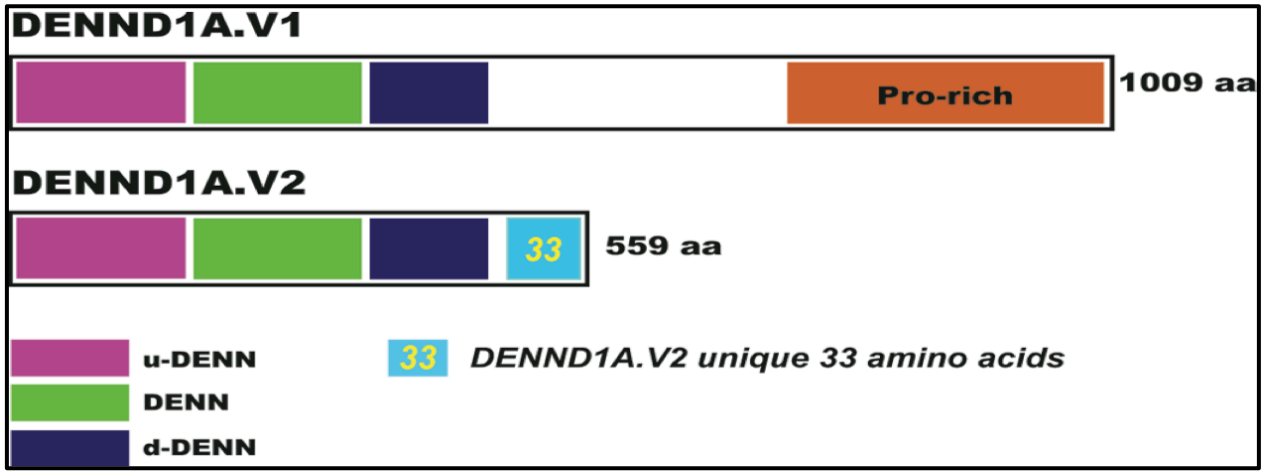


Figure 2.1. The DENND1A gene.

The DENND1A gene is a member of a family of 18 genes, which encode for proteins containing DENN domains: domains that are differentially expressed in normal and neoplastic cells. *DENND1A* has a tripartite DENN domain consisting of upstream (u-DENN), central / core DENN and a downstream (d-DENN) elements which are separated by linker sequences. V1 encodes a full-length protein with 1009 amino acids consisting of all three DENN domains, a clathrin-binding domain and a C-terminal proline rich domain. V2 encodes a protein with 559 amino acids containing the three DENN domains and the clathrin-binding domain (not shown in the figure but is present downstream of the three DENN domains), but lacks the proline rich domain and has a unique 33 AA sequence that is not present in the full-length protein.

2.2. Materials and Methods

2.2.1. Unaffected and PCOS theca cells: Human theca interna tissue was obtained from age-matched (38 – 40 years old), normal cycling women and women with PCOS undergoing hysterectomy, following informed consent under a protocol approved by the Institutional Review Board of The Pennsylvania State University College of Medicine, as previously described (68). All oophorectomies were performed during the luteal phase. The diagnosis of PCOS was made according to NIH consensus guidelines (9), which include hyperandrogenemia, oligoovulation, polycystic ovaries, and the exclusion of 21-hydroxylase deficiency, Cushing's syndrome, and hyperprolactinemia. All of the PCOS theca cell preparations studied came from ovaries of women with fewer than six menses per year and elevated serum total testosterone or bioavailable testosterone levels (42). Each of the PCOS ovaries contained multiple subcortical follicles of less than 10 mm in diameter. The control (unaffected) theca cell preparations came from ovaries of fertile women with normal menstrual histories, menstrual cycles of 21-35 days, and no clinical signs of hyperandrogenism. Neither PCOS nor unaffected control subjects were receiving hormonal medications at the time of surgery. Indications for surgery were dysfunctional uterine bleeding, endometrial cancer, and/or pelvic pain. Theca cells were isolated and cultured as previously described (68). All experiments comparing PCOS and unaffected control theca were performed utilizing 4th-passage theca cells propagated from frozen stocks of second passage cells in media as described previously (68). The passage conditions and split ratios for all control and PCOS cells were identical.

2.2.2. WES Analysis of Control and PCOS Theca Cell DNA. Unaffected control (n = 6) and PCOS (n = 6) genomic DNA samples were subjected to whole exome sequencing at 100 millions reads, at 100X coverage using the Agilent SureSelect 51M capture kit with Illumina HiSEq 2000 sequencing, in conjunction with BGI Americas. Sequence reads obtained were aligned to the reference genome (Human hg19) by BGI Americas. Variant calling and summary of data production was performed by BGI. Visualization of data and identification of called variants was done using the IGV (Integrative Genomic Viewer) software. Only those variants with IGV support numbers >15 (count of the uniquely mapped base) for the alternative allele were selected from the control unaffected and PCOS theca cell samples for further analysis. SNPs in *DENND1A* were identified and their alternative allele frequencies were calculated.

2.2.3. CNV Analysis

From WES data: The reads obtained from WES were used to call for CNV using two different platforms – Conifer and ExomeCNV, in conjunction with BGI Americas.

CNV Assessment using qPCR: Genomic DNA samples from unaffected (n = 6) and PCOS (n = 8) theca cell cultures were all adjusted to a concentration of 5 ng/μl for CNV qPCR assays. Quantitative PCR analysis was performed using commercially available, predesigned TaqMan Copy Number Assays from Life Technologies targeting two different regions in the *DENND1A* gene – i) overlapping exon 2 – intron 2 boundary and ii) overlapping exon 20 – intron 20 boundary. RNAse P Copy Number Reference Assay was used as the endogenous control. QPCR assay and analysis was performed in quadruplicates following manufacturer's protocols on an ABI 7900 HT Fast PCR

System. The raw data obtained from the qPCR assay was analyzed by the maximum likelihood method available in the Copy Caller v1.0 software (Applied Biosystems).

2.2.4. Genomic DNA Sequencing

Genomic DNA isolated from leukocytes (20 unaffected control and 20 PCOS) and theca cell cultures (8 unaffected control and 8 PCOS) was prepared using the "Wizard" genomic DNA purification kit (Promega), and initial direct sequencing was initiated with primers S1280 and AS1670. The region encompassing exons 20 to 20A was amplified using primers S1626 and AS3776 and the resulting 2.15 kb fragment was cloned into pSC-A-amp/kan vector. Between 7 and 9 individual clones were selected from each sample and sequenced with primers S1280 and AS1670. Original chromatograms were visually inspected to identify double peaks.

2.2.5. Minigene construction and splicing analysis

Please note: minigene construction and splicing analysis were performed by Dr. Walter Miller's laboratory at the University of California, San Francisco. Detailed description of the methods and results can be found in the published version of the paper (69). Briefly, in order to determine what forms of RNA were produced from each variant identified in Table 2.2, the 2.15 Kb fragment (of the region encompassing exons 20 – 20A including the intronic region in between) from each sequence variant (from Table 2.2) was digested with *PciI* and *BsrFI* to produce a 1731 bp fragment, which was used to create minigene constructs in pcDNA3.1 vector using CMV promoter and a bGH polyA tail. The design retains sequences from exons 20, 20A and 21, and parts of intronic region

between 20 and 20A and 20A and 21. These constructs were individually transfected into steroidogenic human adrenal NCI-H295A cells. After transfection, RNA splicing was assessed by RT-PCR followed by electrophoresis in 2% agarose gels. To distinguish the splicing products arising from the minigene constructs vs. those endogenously expressed in NCI-H295A cells, we used the vector-specific primer for RT-PCR.

2.2.6. Statistical Analysis

The significance of the association of intron sequence variation with PCOS was assessed using Pearson's Chi Square Test. ($P < 0.05$, was considered significant).

2.3. Results

2.3.1. Genetic variants in the DENND1A gene

Data from WES performed on DNA obtained from unaffected control (n = 6) and PCOS (n = 6) theca cell cultures were investigated to identify functional genetic variants in the *DENND1A* gene. As shown in Table 2.1, only synonymous coding-sequence and intronic variants were identified. No missense, splice-site or nonsense variants that could contribute to alternative splicing of V2 were identified (42).

2.3.2. Copy Number Variations in the DENND1A gene

No copy number differences were found between unaffected control and PCOS theca cell DNA by examining copy number changes in the WES data (6 control, 6 PCOS) as well as by *in vitro* TaqMan Copy Number Assays (6 control, 8 PCOS). The qPCR results from the *in vitro* Copy Number Assays targeting two regions (exon 2 – intron 2; exon 20 – intron 20) in the *DENND1A* gene are shown in Figure 2.2.

2.3.3. DENND1A.V2 arises by exonization of sequences in intron 20

The NCBI database lists multiple *DENND1A* transcripts; V1 and V2 being the major ones (42,66). Examining the coding regions of the V1 and V2 mRNAs showed that they are identical through the end of exon 20 and then diverge. V1 continues with canonical exons 21 and 22 to yield a 1009 AA open reading frame, while V2 lacks the sequences encoded by exons 21 and 22 and has a different coding sequence that is responsible for a 559 AA open reading frame with a unique C-terminal 33 AA. Using this sequence *in silico* to probe the *DENND1A* gene, we found that the 33 AA C-terminus of V2 was generated by exonization of sequences located ~1.5 kb downstream from canonical exon 20 and 19.5 kb upstream from canonical exon 21 (Figure 2.3 A) Because this "exonic" or coded sequence was flanked by intronic region on 5'end but not the 3'end, (coding for a stop codon causing production of truncated isoform V2 with 559 AA), we named this exon, cryptic exon 20A.

2.3.4. Intron 20 Sequence Variation

To determine if there is genetic variation in intron 20 that might give rise to alternative splicing that might determine whether the *DENND1A* gene expresses mRNA encoding DENND1A V1 vs. V2, genomic DNA from human leukocytes obtained from blood collected from unaffected control (n = 20) women and women with PCOS (n = 19) as well as from theca cell cultures obtained from unaffected control women (n = 8) and women with PCOS (n = 9) was sequenced. Using primers S1626 and AS3776, we amplified a 2.15 kb region and then used primers S1280 and AS1670 (Figure 2.3 B) to sequence the PCR products. Sequencing identified a complex variable region 330 bp upstream from exon 20A with an extended dinucleotide TG repeats. Direct sequencing of this region in some genomic DNA samples yielded garbled sequence data,

suggesting the presence of multiple sequence variants. Therefore, we cloned the 2.15 kb product and then sequenced multiple clones and identified different sequence variants (Figure 2.3 C). Among the 10 variants (variants 2 - 11) identified in genomic DNA from unaffected and PCOS women, none appeared to be overrepresented in PCOS women. Variants 2,4,9 and 11 were found in a limited number of samples from theca cells, and variants 2 and 4 were only found in samples from PCOS patients. We detected known SNPs including rs10739631, rs45480399, rs71490807, and rs10986007 (Figure 2.3 C) (69). The sequence variants were assessed and results are represented in Table 2.2. None of the variants were significantly associated to PCOS (*P* >0.05)

2.3.5. Minigene Analysis of alternative splicing.

Please note: minigene construction and splicing analysis were performed by Dr. Walter Miller's laboratory at the University of California, San Francisco. Detailed description of the results can be found in the published version of the paper (69). Briefly, five splicing products were obtained arising from the minigene constructs created with each of the eleven variants shown in Figure 2.3 and Table 2.2. However, in two independent experiments with all constructs, none of these consistently favored splicing to favor the mRNA encoding the V2 isoform (69).

Table 2.1. WES analysis of the *DENND1A* gene. DENND1A variants identified by WES are described along with the calculated alternate allele frequencies in the control (n = 6) and PCOS (n = 6) theca cell cultures. Alternate allele frequency in European American population (obtained from Exome 6500) was reported for comparison. (42)

SNP	Location (Chr 9)	Function	Ref Allele	Alt allele Freq_Control	Alt allele Freq_PCOS	Alt allele Freq (EU)
rs3739837	126143662	utr-3	G	A (0.083)	A (0.083)	A (0.1109)
rs2491348	126144746	Synonymous	T	C (1.0)	C (1.0)	T (0.0003)
rs2808409	126144758	Synonymous	T	C (1.0)	C(1.0)	ABSENT
rs2808411	126146197	Intron	C	G (1.0)	G (1.0)	C (0.082)
rs10739631	126164562	Intron	T	C (1.0)	C (1.0)	ABSENT
rs7872778	126201672	Intron	G	A (0.083)	Absent	ABSENT
rs62581072	126201742	Intron	G	Absent	A (0.083)	ABSENT
rs10739633	126202551	Intron	A	C (0.75)	C (0.75)	ABSENT
rs3829851	126219706	Synonymous	A	G (0.083)	Absent	G (0.055)
snp4115	126219742	Intron	C	Absent	T (0.083)	ABSENT
snp3842	126220063	Intron	G	Absent	T (0.083)	ABSENT
rs61736953	126220114	Synonymous	C	T (0.083)	T (0.166)	T(0.0884)
rs12377595	126439100	Intron	T	G (0.166)	G (0.33)	G (0.2490)
rs9785285	126520068	Synonymous	T	C (0.33)	C (0.166)	C (0.2123)
rs41274356	126531724	Intron	T	C (0.083)	Absent	C (0.0239)
rs1778890	126531755	Intron	T	C (0.33)	C (0.166)	C (0.2148)
snp3895	126554844	Intron	A	Absent	G (0.083)	ABSENT
rs670028	126641091	Intron	A	G (0.25)	Absent	ABSENT

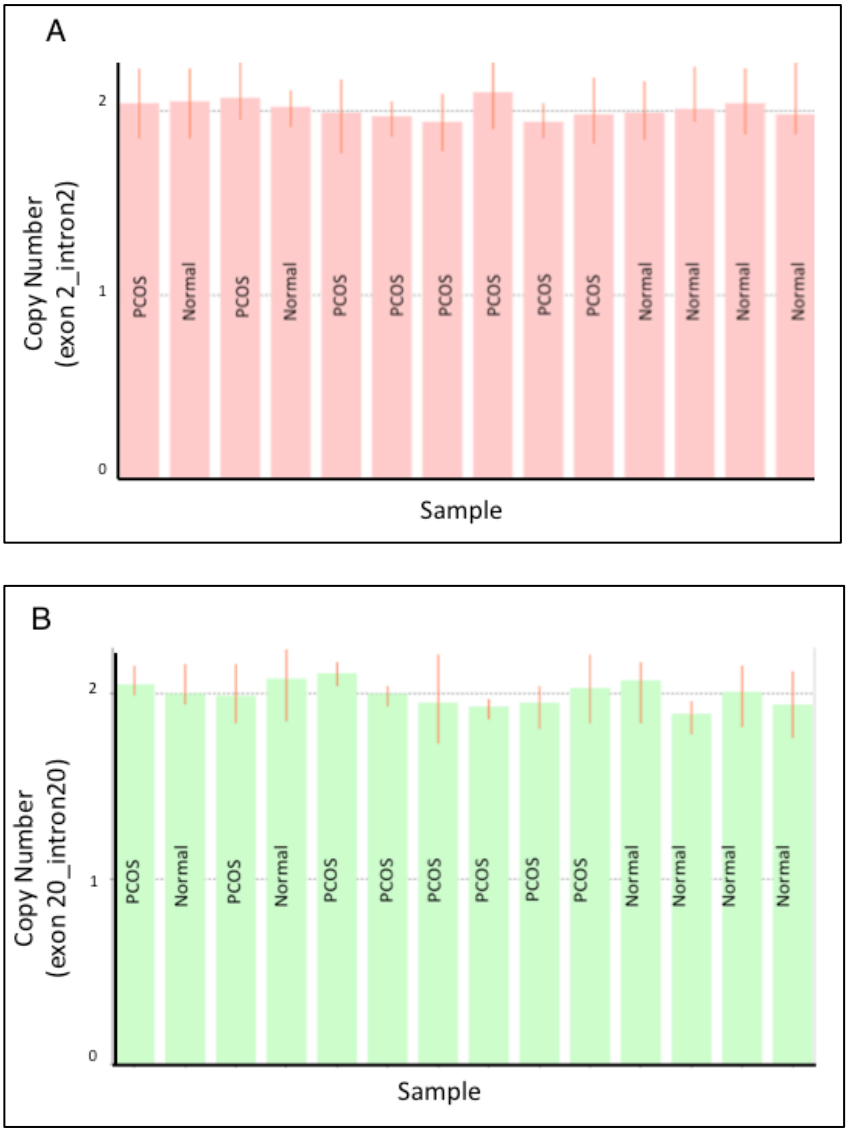


Figure 2.2. Copy Number Analysis in the DENND1A gene. Results from TaqMan Copy Number qPCR assays for regions overlapping exon 2 – intron 2 (2.2 A) and exon 20 – intron 20 (2.2 B) were analyzed in CopyCaller software and no copy number differences between control (n = 6) and PCOS (n = 8) theca cells were observed.

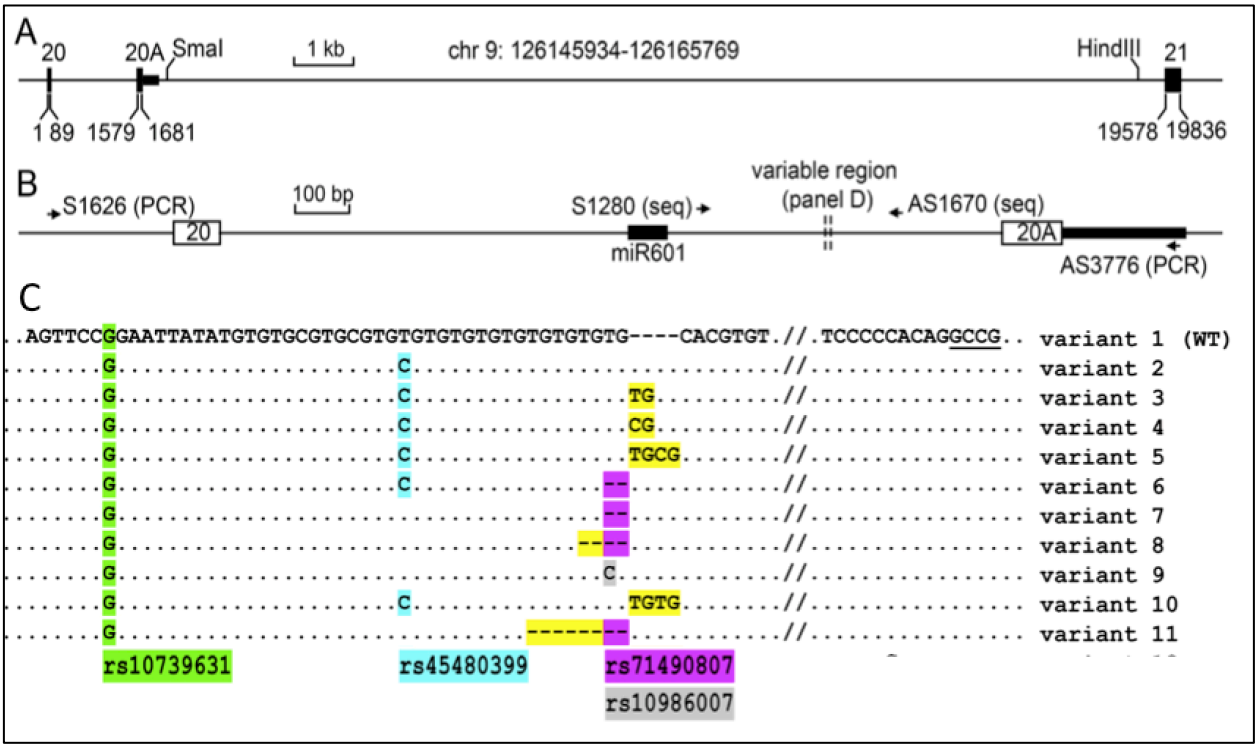


Figure 2.3. Diagram of the 3' end of the DENND1A gene and intron 20 sequence variants identified. A. Scale diagram of ~20 kb spanning DENND1A exons 20 and 21; bases are numbered with the first base of exon 20 (Chr 9: 126145934), designated as base 1; exons are shown as boxes. B. Magnified diagram of the region between exons 20 and 20A; the hypervariable region between the vertical dotted lines was identified by sequencing PCR products generated with primers S1626 and AS3776. C. Sequence variants in the hypervariable region with variant 1 depicting the WT sequence. Known SNPs in this region are highlighted: green, rs10739631; cyan, rs45480399; purple, rs71490807; grey, rs10986007. Newly identified variants at chr 9: 126164525 - 126164521 are highlighted in yellow. (69)

Table 2.2. Intron 20 Sequence Variation. The intronic region between exons 20 and 20A was assessed in the genomic DNA from human leukocytes obtained from blood collected from unaffected control (n = 20) women and women with PCOS (n = 19) as well as from theca cell cultures obtained from unaffected control women (n = 8) and women with PCOS (n = 9). The number of clones carrying each variant sequence (shown in Figure 2.3 C) in normal and PCOS samples and the average number of clones identified per sample for each variant are described.

Variant	No. of Clones with Variant		Avg. No. of Clones with Variant per sample	
	Control	PCOS	Control	PCOS
1	65	42	3	2
2	0	2	NA	1
3	4	4	1	1
4	0	2	NA	1
5	18	20	3	2
6	9	23	3	2
7	18	28	2	2
8	6	2	1	1
9	2	2	2	2
10	3	0	2	NA
11	1	0	1	NA

2.4. Discussion

Considering the established association between *DENND1A* and PCOS based on genetic association studies (34–39), genotype-phenotype correlation studies (40,41), *in vitro* functional studies (70) and its potential diagnostic and therapeutic applications (67), it was first important to understand the mechanism of V2 production and overexpression in PCOS theca cells. We report that the V2 isoform of the *DENND1A* gene is produced by exonization of sequences in intron 20, which generates a unique exon 20A, encoding the C-terminus of V2 (69).

This is a comprehensive study investigating several possible genetic mechanisms for V2 overexpression in PCOS theca cells including sequencing analysis for coding region and splice-site variants, copy number analysis, analysis of the intronic region between exon 20 (present in both V1 and V2) and cryptic exon 20A (present in only V2). As shown in the results none of these mechanisms appear to contribute to increased V2 expression in PCOS theca. The WES results are consistent with another study reporting no changes in *DENND1A* exons or splice junctions in PCOS (71).

Genetic variation in the intronic region between exons 20 and 20A was identified. However, when each of these sequence variants was tested in a mini gene system, none of them consistently favored mRNA splicing to produce V2 instead of V1. Thus increased V2 expression in PCOS theca cells does not appear to result from genomic sequence variation in intron 20.

It is important to note that though none of the genetic mechanisms we investigated, seemed to contribute to the observed V2 overexpression in PCOS theca cells, there maybe other potential genetic mechanisms including promoter region variants, variants

affecting mRNA stability (positively or negatively) in both *DENND1A* isoforms and epigenetic mechanisms including DNA methylations, histone modifications and miRNA regulation. Further investigation of these mechanisms contributing to V2 overexpression in PCOS theca cells is warranted.

Chapter 3: MiR-125a-3p targets DENND1A.V2 and contributes to the V2 mediated steroidogenic phenotype in PCOS theca cells

Abstract

Polycystic ovary syndrome (PCOS) is a multisystem disorder with a complex pathophysiology. Several microRNAs (miRNAs), small, non-coding post-transcriptional gene regulators, are known to be differentially expressed in granulosa cells, follicular fluid, ovarian stroma, and in the circulation of women with PCOS. However, there are no reports on miRNA expression and target gene analysis in control and PCOS theca cells. Understanding gene expression and its regulation in theca cells may represent a key to understanding the hyperandrogenemia characteristic of PCOS. The previously reported altered transcriptome profiles in PCOS theca cells could be explained, at least in part, by miRNA regulation. In the present study, we identified 18 miRNAs, which are differentially expressed in PCOS theca cells, some of which are predicted to target PCOS candidate genes identified by Genome Wide Association Studies. Using Ingenuity pathway analysis, we identified a common network with ~45% of the differentially expressed miRNAs identified in this study, and 73% of the PCOS candidate genes. In addition, miR-125a-3p, which was predicted to target DENND1A, one of the PCOS candidate genes, was shown to have specific interaction for the DENND1A.V2 isoform known to have a functional role in PCOS theca cell

steroidogenesis via correlation expression and mimic studies. In addition, miR-125a-3p expression was also negatively correlated to DHEA accumulation and overexpression of miR-125a-3p mimic also downregulated CYP17 mRNA expression. Our findings suggest an epigenetic mechanism of V2 overexpression in PCOS theca cells via miR-125a-3p and suggest that the PCOS candidate genes derived from GWAS are part of a complex network regulated by miRNAs.

3.1. Introduction

Polycystic ovary syndrome (PCOS) is a common endocrine disorder that affects 5% - 7% women worldwide. The disorder is heterogeneous, characterized by a broad spectrum of reproductive, metabolic and endocrinological features including hyperandrogenism, anovulation, infertility, and presence of multiple small subcortical follicular cysts embedded in bilaterally enlarged ovaries (20,28,72,73). Other endocrinopathies and metabolic features are commonly observed in PCOS patients including abnormal glucose metabolism, insulin resistance, dyslipidemia and obesity. PCOS women are known to have higher prevalence rates of type 2 diabetes (T2D), metabolic syndrome and cardiovascular disease (10,72). Most consensus diagnostic schemes recommend the presence of hyperandrogenism (clinical or biochemical) as an essential characteristic for reaching a PCOS diagnosis, and it is considered to be the hallmark of PCOS. A cross-sectional case-control study suggested that hyperandrogenism is implicated in both the metabolic and reproductive morbidities of PCOS, and maybe a common link between the two (9,20). Ovarian androgen production is increased in PCOS and studies show that ovarian theca cells from PCOS

women secrete more androgen than theca cells from regularly ovulating women (21,22,74–76). Thus, understanding gene expression and its regulation in theca cells is central to understanding PCOS pathophysiology, or at least hyperandrogenism associated with PCOS. Wood *et al*. (77,78) have shown that PCOS theca cells have a characteristic molecular signature as compared to normal theca cells. Although these results revealed altered gene expression in PCOS theca as compared to normal theca cells, it is not yet known whether the altered expression profile is the result of transcriptional or post-transcriptional regulatory mechanisms.

MicroRNAs (miRNAs) are small (20-24 nucleotides) single-stranded, noncoding, regulatory RNA molecules. They are involved in post-transcriptional regulation of gene expression either by complimentary binding to the 3' untranslated region of their target mRNA, thereby inhibiting translation or inducing mRNA degradation (61,62). The post-transcriptional mechanism operationalized for down regulation of gene expression depends upon the binding target sequence. Messenger RNA degradation occurs in case of complete or near complete complementarity between miRNA and the target, while repression or inhibition of translation occurs if there is not sufficient complementarity for cleavage (62). Although not much is known about the roles of miRNAs in PCOS pathophysiology, there have been studies providing evidence of differential miRNA expression in the ovarian stroma, follicular fluid and granulosa cells of women with PCOS (63,79–81). Metformin, an insulin-sensitizing drug, which is used as a treatment to lower insulin and androgen levels and initiate ovulation in women with PCOS, changes global miRNA expression patterns (82). MicroRNAs are also reported to be stably expressed in encapsulated vesicles or free-circulating in the serum, plasma,

urine and saliva, making them potential non-invasive biomarkers for PCOS diagnosis (63).

Several genome-wide association studies (GWAS) have identified loci significantly associated with PCOS. In a two-part study, GWAS on Han-Chinese populations identified 11 candidate loci (34,35). Subsequent GWAS on European populations confirmed associations of some of the Han-Chinese GWAS to PCOS, and also identified additional candidate loci (36,37). Together, the 4 GWA studies have identified 22 candidate loci for PCOS (Table 1.1, chapter 1), including *DENND1A*. Although some of these loci consist of plausible PCOS candidate genes, the molecular mechanisms by which most of the identified loci contribute to PCOS is not clearly understood. We demonstrated that the total mRNA transcript as well as a splice variant of *DENND1A*, DENND1A.V2, is increased in PCOS theca cells, with no change in the DENND1A.V1 (full-length isoform) transcript expression (70). Forced expression of DENND1A.V2 in normal theca cells increased androgen production, whereas knockdown of DENND1A.V2 expression in PCOS theca cells reduced androgen synthesis, implicating DENND1A.V2 in the regulation of steroidogenesis and the PCOS phenotype (70). The mechanisms underlying the increased DENND1A.V2 expression in PCOS theca cells, as well as their altered transciptome signature in PCOS, remain to be elucidated. One possibility is that these alterations result, in part, from differential expression of miRNAs. However, a detailed study of miRNA expression and target gene analysis has not been performed in theca cells.

In the present study, miRNA expression profiles of human theca cell cultures, established from women with PCOS and without the disease were determined using

next generation sequencing. Target gene analysis of the differentially expressed miRNAs was focused on PCOS candidate genes identified by GWAS (34–37). Pathway and Network analyses (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) were performed on the differentially expressed miRNAs between normal and PCOS theca cells identified in this study and the established GWAS candidate genes to identify a common network that may be involved in PCOS pathophysiology. Expression patterns of select miRNAs and target genes in the common network were validated by quantitative real-time PCR and associations of differential expression patterns (of both miRNAs and target genes) with androgen biosynthesis were investigated.

3.2. Materials and Methods

3.2.1 Cell culture: Human theca interna tissue was obtained from age-matched, normal cycling women (n = 6) and women with PCOS (n = 5) undergoing hysterectomy, following informed consent under a protocol approved by the Institutional Review Board of The Pennsylvania State University College of Medicine as previously described (68). Isolation of theca cells and theca cell cultures was done as described in Chapter 2. All theca cell cultures used for the following experiments were treated with and without 20mM forskolin for 16 hours. Refer to Appendix, Table A.3.1 for details about samples used for different assays performed.

H295R cells were obtained from ATCC and were cultured according to ATCC recommended culture conditions.

3.2.2. MicroRNA Discovery Sequencing: MicroRNA discovery sequencing was performed using the Illumina technology (by LC Sciences, www.lcsciences.com) on

theca cell RNA from normal cycling women (n = 4) and women with PCOS (n = 4) each treated with (F) and without (C) forskolin. Total RNA was extracted using TRIzol reagents. Small RNA library prep, cDNA deep sequencing and initial data analysis including alignment of raw reads to miRbase (v17) and normalization of count data was performed by LC Sciences. The $\log_2$ normalized expression levels for each miRNA were used for statistical analyses after filtering for human miRNAs.

3.2.3. Target identification and Pathway analysis: The miRNA target prediction filter in Qiagen's Ingenuity® Pathway Analysis tool (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) was used to identify potential target genes for the 18 miRNAs that were differentially expressed between PCOS and normal theca cells (FDR < 0.05). MicroRNA-target gene relationships having a confidence level of "Experimentally Observed" or "High-Predicted" were considered. Pathway and network analyses were also performed in IPA.

3.2.4. Quantitative RT-PCR based analysis of miRNA expression: Expression patterns of select miRNAs were validated *in vitro* using the TaqMan MicroRNA Assays (Life Technologies) following the manufacturer's instructions. For the miRNA-qPCR assays, total RNA was extracted from the established theca cell cultures (Normal: n = 5 and PCOS: n= 4; treated with and without forskolin) using miRVana miRNA isolation kit (Ambion, Life Technologies) following manufacturer's protocol. Briefly, 10 ng RNA was reverse transcribed using the target (miRNA) specific stem-loop RT primer and the TaqMan MicroRNA reverse transcription kit. The cDNA was then amplified by real-time qRT-PCR using target specific TaqMan primer-probe mix. The qRT-PCR was performed in triplicates per sample. Different small RNAs (RNU48, RNU43 and miR-16)

were tested for consistent expression between PCOS and unaffected control theca cell samples and RNU48 was selected for normalization of the qRT-PCR expression data. The mean expression value for each miRNA was divided by the mean RNU48 expression value to normalize each sample.

3.2.5. Quantitative RT-PCR based analysis of DENND1A.V1, DENND1A.V2 and CYP17 expression. Quantitation of DENND1A.V1, DENND1A.V2 and CYP17 mRNA abundance was determined using the Single Step Brilliant III Ultra Fast SYBR qRT-PCR Reagents (Agilent) using 50-100 ng total RNA/tube, and 200nM final concentration of each forward and reverse primers, and 100nM probe. The gene specific one step PCR was carried out in duplicate for each mRNA sample and for a series of dilutions in an Mx3000p Thermocycler system (Stratagene) according to manufacturers instructions for this instrument as previously described (70). The primer and probe sets utilized were previously described (70). TATA Binding Protein (TBP) mRNA was used for normalization based on our previous studies (70). The mean expression value for each mRNA was divided by the mean TBP expression value to normalize each sample.

3.2.6. Quantitation of Dehydroepiandrosterone (DHEA): ELISA for DHEA was performed on cell culture media without organic solvent extraction using kits from DRG International, Inc. (Springfield, NJ) as described by the manufacturer's protocol, and normalized by cell count.

3.2.7. Mimic studies: H295R cells were transfected with miRVana hsa-miR-125a-3p mimic and negative control-1 mimic (Life Technologies) at a concentration of 75 nM using RNAiMAX transfection reagents and the manufacturer's protocol. Cells were harvested 48 hours post transfection and total RNA was extracted using miRVana

miRNA isolation kit (Ambion Life Technologies) following the manufacturer's protocol. V1, V2 and CYP17 mRNA qPCRs were performed on control and miR-125a mimic transfected RNA as described in section 3.2.5.

3.2.8. Statistical analysis: For miRNA discovery sequencing analysis, for each miRNA, a one-way analysis of variance (ANOVA) model was fit to the $\log_2$ normalized expression levels to identify miRNAs differentially expressed among the four groups - Normal Untreated Control (Normal-C), Normal Forskolin Treated (Normal-F), PCOS Untreated Control (PCOS-C) and PCOS Forskolin Treated (PCOS-F). After fitting the ANOVA models for each miRNA, linear contrasts were used to test for differences between the following group comparisons: Normal-C vs. Normal-F; PCOS-C vs. PCOS-F; Normal-C vs. PCOS-C; Normal-F vs. PCOS-F and Untreated (Normal-C and PCOS-C) vs. Treated (Normal-F and PCOS-F). Meta-analysis was performed on miRNAs common between and having significant $P < 0.05$ in both group comparisons; Normal-C vs. PCOS-C, and Normal-F vs. PCOS-F. Their individual *P* values were combined using Fisher's method (83) for combining *P* values to get an overall comparison of miRNA expression between PCOS and normal theca cells. Subsequently, the combined *P* values were used to obtain the false discovery rate (FDR) using the Benjamini and Hochberg method (84). MicroRNAs having a statistically significant expression difference between both pair-wise comparisons (Normal-C vs. PCOS-C and PCOS-F vs. PCOS-F), a meta-analysis combined $P < 0.05$ and FDR < 0.05 were considered significant. All statistical analyses were performed using R (85).

For the miRNA qRT-PCR assays, two sample t-tests were performed to test for differences between groups Normal-C vs. PCOS-C and Normal-F vs. PCOS-F. The *P*

values thus obtained were combined using Fisher's method (83) and combined $P < 0.05$ were considered statistically significant. Each miRNA specific cDNA was run in triplicates for the qRT-PCR assays.

The results from mRNA qRT-PCR (DENND1A.V1 and DENND1A.V2) and steroid (DHEA) were collected from individual patients and *ANOVA* was performed to test for differences. *P* values were determined by the Bonferroni method for multiple comparisons. Data are presented and described in the text as the mean ± SEM performed triplicate and were prepared using Prism 5.0c (GraphPad Software, San Diego, CA).

Spearman's nonparametric correlation coefficient (ρ) was calculated to estimate the degree of correlation between all pairwise comparisons. (ρ was considered significant when $P < 0.05$). Please note that all samples treated with (F) and without (C) forskolin were considered together for the statistical analysis, giving a total of 10 unaffected control and 8 PCOS samples for Spearman's correlation testing. Performing statistical tests for treated and untreated samples individually gave insignificant results owing to small sample size (n = 5 control unaffected; n = 4 PCOS samples)

3.3. Results

3.3.1. PCOS theca cells have differential miRNA expression as compared to Normal theca cells: A total of 1,823 miRNAs were identified from deep sequencing. After meta-analysis of the Normal-C vs. PCOS-C and Normal-F vs. PCOS-F pairwise comparisons, 18 miRNAs were found to be significantly differentially expressed between normal and PCOS theca cells (Fisher's combined $P < 0.05$) with a FDR < 0.05

(Table 3.1). Of these 18 miRNAs, 13 miRNAs had increased expression and 5 miRNAs had decreased expression in PCOS theca cells as compared to unaffected control theca cells.

3.3.2. Differentially expressed miRNAs in PCOS theca are predicted to target known PCOS candidate genes: The 18 miRNAs listed in Table 3.1 were used for bioinformatics analysis including target prediction and pathway analysis in IPA as described below. The miRNA target prediction filter in IPA was used to identify the validated and predicted target genes of the differentially expressed miRNAs (the miRNA-target gene relationships with IPA described confidence level of "Experimentally Observed" and "High-Predicted") and a total of 6443 such interactions were identified. Interestingly 6 of the 18 miRNAs described in Table 3.1, were predicted to target (confidence level = High Predicted) 12 of 22 PCOS GWAS candidate genes identified in Han-Chinese (34,35) and European populations (36,37) and described in Table 1.1 (Chapter 1). The predicted miRNA-GWAS gene interactions are shown in Table 3.2.

Table 3.1. Differentially expressed miRNAs in unaffected control and PCOS theca cells. Using MicroRNA Discovery Sequencing, 18 miRNAs were differentially expressed between unaffected control vs. PCOS theca cells, 13 of which were upregulated and 5 were down-regulated in PCOS theca cells compared to normal. (Fisher's combined *P* value < 0.05 and FDR < 0.05 considered significant)

ID (hsa)	PCOS/ Control (C)	*P* value PCOS vs. Control (C)	PCOS/ Control (F)	*P* value PCOS vs. Control (F)	meta *P* value	meta FDR
miR-501-3p	6.31	2.13E-06	5.56	5.20E-06	2.90E-10	5.29E-07
miR-100-5p	4.89	4.74E-06	3.90	2.83E-05	3.19E-09	2.43E-06
miR-409-5p	9.27	4.66E-06	6.51	3.65E-05	4.00E-09	2.43E-06
miR-125a-3p	0.69	9.19E-06	0.76	1.17E-04	2.32E-08	1.06E-05
miR-1271-5p	7.44	8.83E-05	5.61	3.70E-04	5.95E-07	2.17E-04
miR-1301-3p	2.70	1.04E-03	3.08	2.87E-04	4.79E-06	1.45E-03
miR-130b-3p	0.61	3.63E-04	0.72	6.70E-03	3.39E-05	8.23E-03
miR-99b-5p	2.79	7.42E-04	2.37	3.51E-03	3.61E-05	8.23E-03
miR-127-3p	2.15	1.72E-04	1.42	2.86E-02	6.45E-05	1.31E-02
miR-148b-5p	0.69	1.55E-03	0.75	6.33E-03	1.23E-04	2.18E-02
miR-654-5p	2.41	1.72E-04	1.37	6.35E-02	1.35E-04	2.18E-02
miR-195-5p	0.46	8.74E-04	0.62	1.33E-02	1.44E-04	2.18E-02
miR-744	1.48	2.10E-03	1.36	1.00E-02	2.48E-04	3.29E-02
mir-1293	2.05	8.56E-03	2.31	2.80E-03	2.78E-04	3.29E-02
miR-410-3p	1.69	1.30E-03	1.41	1.92E-02	2.90E-04	3.29E-02
miR-4524a-5p	0.31	9.42E-04	0.50	2.67E-02	2.91E-04	3.29E-02
hsa-miR-502-3p	1.77	1.52E-02	2.22	1.75E-03	3.07E-04	3.29E-02
hsa-miR-494-3p	2.53	5.50E-03	2.60	5.95E-03	3.71E-04	3.76E-02

Table 3.2. MiRNAs targeting GWAS genes. MiRNA target identification (IPA) filter identified 6 of the differentially expressed miRNAs with predicted interactions with 12 of the 22 GWAS genes previously identified (listed in Table 1).

ID	Targeted Gene	Source	IPA Confidence
hsa-miR-125a-3p	DENND1A	TargetScan Human	High (predicted)
hsa-miR-1271-5p	TOX3	TargetScan Human	High (predicted)
hsa-miR-1271-5p	ERBB4	TargetScan Human	High (predicted)
hsa-miR-130b-3p	DENND1A	TargetScan Human	High (predicted)
hsa-miR-130b-3p	ZNF217	TargetScan Human	High (predicted)
hsa-miR-130b-3p	RAB5B	TargetScan Human	High (predicted)
hsa-miR-130b-3p	ERBB4	TargetScan Human	High (predicted)
hsa-miR-130b-3p	ERBB3	TargetScan Human	High (predicted)
hsa-miR-130b-3p	KCNA4	TargetScan Human	High (predicted)
hsa-miR-195-5p	INSR	TargetScan Human	High (predicted)
hsa-miR-195-5p	YAP1	TargetScan Human	High (predicted)
hsa-miR-195-5p	HMGA2	TargetScan Human	High (predicted)
hsa-miR-195-5p	RAD50	TargetScan Human	High (predicted)
hsa-miR-410-3p	HMGA2	TargetScan Human	High (predicted)
hsa-miR-654-5p	GATA4	TargetScan Human	High (predicted)

3.3.3. A common network of differentially expressed miRNAs and PCOS candidate genes may be at work in the PCOS theca: Target prediction revealed that ~33 % (6/18) of the miRNAs that are differentially expressed between unaffected control and PCOS theca cells are predicted to target 55% (12/22) of the GWAS candidate genes including *DENND1A,* which has been previously shown to have a functional role in PCOS theca cell steroidogenesis. This led us to hypothesize that like *DENND1A*, other PCOS GWAS candidates may also have significant functional and biological relevance to the disorder, possibly in a common network. To test this *in silico*, IPA core pathway analysis was performed on all of the 22 GWAS genes identified thus far to examine pathways, networks, diseases and bio-functions that are "over-represented" in the GWAS candidates. Several PCOS relevant features were identified in the canonical pathways, diseases and biofunctions, physiological system development and functions that were over-represented in the GWAS genes dataset and are listed in Table 3.3. These results suggested that at least some of the 22 GWAS candidate genes and the differentially expressed miRNAs might be involved in a common pathway/ network with functional relevance to PCOS pathophysiology central to the theca cells. To investigate this, we utilized the "Build Network" application available in IPA and created a network with the 22 GWAS genes and overlayed the 18 differentially expressed miRNAs and using the "grow" option added any known or predicted connections present in the Ingenuity Knowledge Base. The resulting network is represented in Figure 3.1. As shown, ~73% (16/22) of PCOS GWAS candidate genes and ~44% (8/18) of differentially expressed miRNAs in PCOS theca are a part of a common network.

Table 3.3. Ingenuity Pathway Analysis of the PCOS risk loci reported to date. Key findings from the pathway analysis (performed in IPA) of the 22 loci (Table 1.1) associated with PCOS risk are described. The Top Canonical Pathways (3.3.A), Top Diseases and Bio-functions (3.3.B) and Physiological System Development and Functions (3.3.C) over-represented in the PCOS risk loci are shown.

3.3.A. Top Canonical Pathways

Name	*P* value	Overlap
Agrin interactions at Neuromuscular Junction	3.80E-05	4.30%
ErbB Signaling	7.35E-05	3.50%
Neuregulin Signaling	7.87E-05	3.40%
Ovarian Cancer Signaling	2.56E-04	2.30%
ErbB2-ErbB3 Signaling	1.35E-03	3.50%

3.3.B. Top Diseases and Bio-functions

Name	*P* value	# Molecules
Cancer	9.61E-04 - 4.31E-11	20
Organismal Injury and Abnormalities	9.61E-04 -4.31E-11	20
Respiratory Disease	4.34E-04 - 4.38E-10	6
Endocrine System Disorders	9.60E-04 - 6.83E-09	12
Reproductive System Disease	9.60E-04 - 6.83E-09	17

3.3.C. Physiological System Development and Function

Name	p-value	# Molecules
Organ Development	9.60E-04 - 4.46E-09	9
Reproductive System Development and Function	9.60E-04 - 4.46E-09	10
Organ Morphology	9.6E-04 - 2.81E-08	10
Renal and Urological System Development and Function	8.80E-04 - 2.81E-08	7
Organismal Development	9.60E-04 - 1.44E-07	10

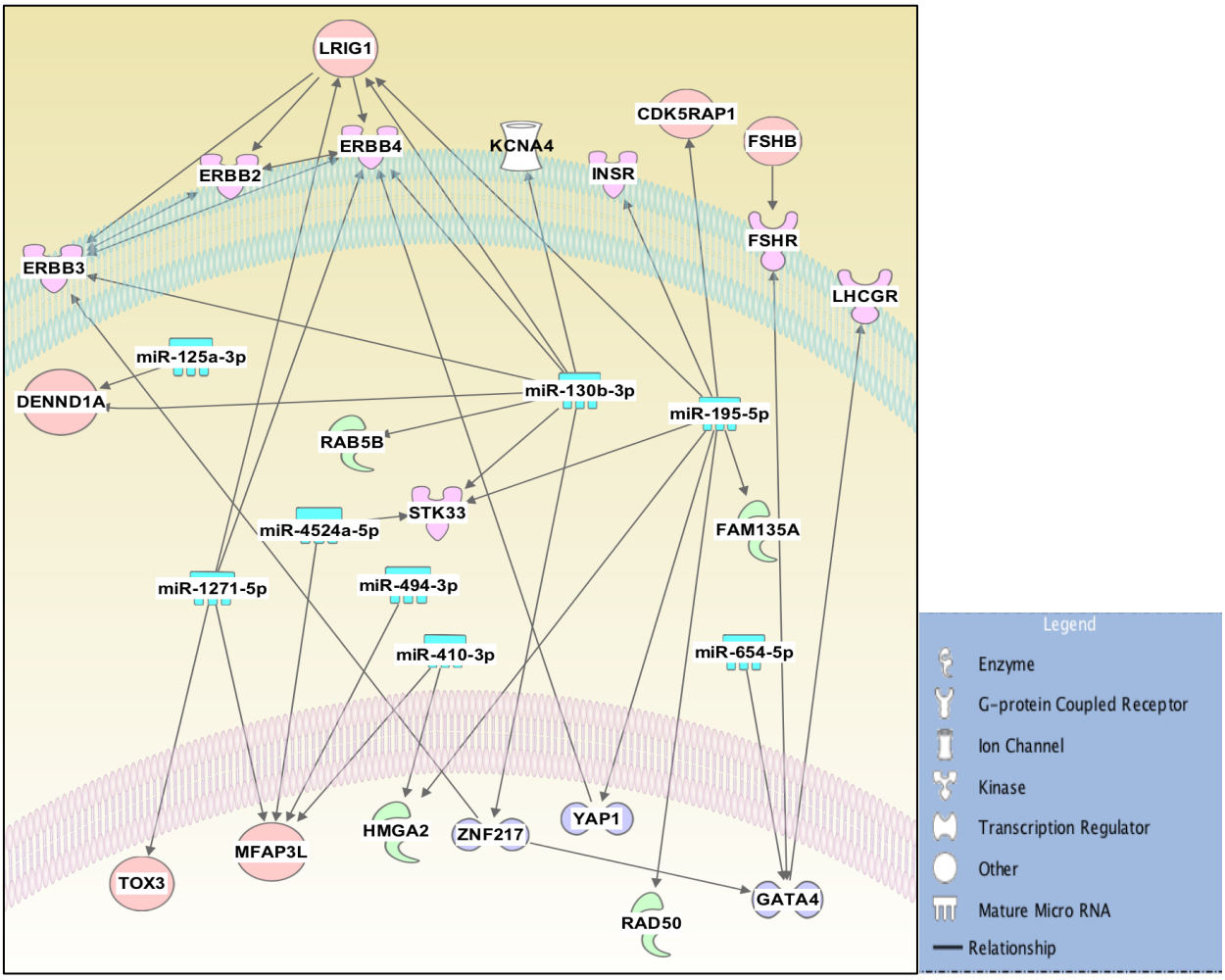


Figure 3.1. IPA network of miRNAs and PCOS candidate genes. Common network of the 16/22 PCOS candidate genes and 8/18 of the differentially expressed miRNAs is shown. IPA pathway figure legend is described as well.

3.3.4. qRT-PCR validation of miRNA expression: Next, we determined the expression patterns of all the 8 miRNAs in Figure 3.1. *in vitro* in 5 unaffected control and 4 PCOS theca cell cultures treated with and without 20 mM forskolin by TaqMan-based miRNA qPCR assays (Life Technologies). Of the 8 miRNAs, statistically significant expression differences ($P < 0.05$) between unaffected control and PCOS theca cells were only observed for hsa-miR-125a-3p (Figure 3.2.A) and hsa-miR-130b-3p (Figure 3.2.B). miR-125a-3p expression was significantly downregulated in PCOS theca cells under both basal and forskolin stimulated conditions. MiR-130b-3p expression was reduced in PCOS theca cells only under basal conditions. There were no significant changes in miR125a-3p and miR-130b-3p expression in response to forskolin treatment in unaffected control and PCOS theca cells.

3.3.5. Decreased miR-125a-3p expression is correlated to increased DENND1A.V2 expression in PCOS theca cells. Since both miR-125a-3p and miR-130b-3p are predicted to target *DENND1A*, we sought to determine the relation between their expression patterns in theca cells. As shown in Figure 3.3, expression of both miRNAs was compared against expression of both isoforms of *DENND1A*, DENND1A.V1 and DENND1A.V2, in the same unaffected control (n = 5) and PCOS (n = 4) samples treated with and without forskolin. A significant negative correlation was observed between expression of miR-125a-3p and DENND1A.V2 (Figure 3.3.A) but no significant relation with DENND1A.V1 expression was observed (Figure 3.3.B). No significant correlations ($P > 0.05$) were observed between miR-130b-3p and DENND1A.V2 or DENND1A.V1

In silico miRNA-mRNA target prediction strength was assessed for miR-125a-3p and

both isoforms of *DENND1A* in microRNA.org, which utilizes predicted miR-target interactions from miRanda and predicts the probability of miRNA downregulating gene expression by calculating mirSVR scores. Positive or miRSVR scores = 0 indicate a null probablility of interaction and higher negative scores indicate increasing probability of interaction. miR-125a-3p alignment with 3'UTR of DENND1A.V1 gave a mirSVR score = -0.0004, whereas miR-125a-3p alignment with 3'UTR of DENND1A.V2 (not present in V1) gave a mirSVR score = -1.0634. This preliminary *in silico* analysis supported our observed expression correlation results described above suggesting that the predicted interaction of miR-125a-3p with *DENND1A* is likely more specific for DENND1A.V2 than for DENND1A.V1 in theca cells.

3.3.6. Decreased miR-125a-3p expression in PCOS theca cells is correlated with increased DHEA production. Because a relationship between reduced miR-125a-3p expression and increased DENND1A.V2 expression was established and as increased DENND1A.V2 expression in PCOS is known to be associated with increased androgen production (70), we sought to determine whether miR125a-3p expression was associated with increased androgen biosynthesis. We examined DHEA accumulation in the matched 5 unaffected control and 4 PCOS theca cell preparations under the same conditions. As shown in Figure 3.4.A, miR-125a-3p expression has a significant negative correlation with DHEA accumulation in theca cells. Interestingly, miR-130b-3p (Figure 3.4.B) expression also showed a negative correlation with DHEA accumulation but not as significant as with miR-125a-3p and DHEA.

3.3.7. MiR-125a-3p overexpression in H295R cells downregulates V2 and CYP17 mRNA expression, but has no effect on V1 expression.

In order to confirm specific interaction of miR-125a-3p with V2 and its impact on androgen production (as shown by DHEA accumulation), we overexpressed miR-125a-3p mimic in H295R cells (adrenal corticocarcinoma cell line) and analyzed V1, V2 and CYP17 mRNA expression by qRT-PCR in negative-control and mimic treated cells. As shown in Figure 3.5.A, mimic treated H295R cells showed a significant decrease in V2 expression (P = 0.042) when compared to control treated cells. No change in V1 expression was observed (Figure 3.5.B). In addition, CYP17 expression was also significantly downregulated upon mimic overexpression as compared to control, which was in agreement to the miRNA-DHEA correlation data shown in Figure 3.4.

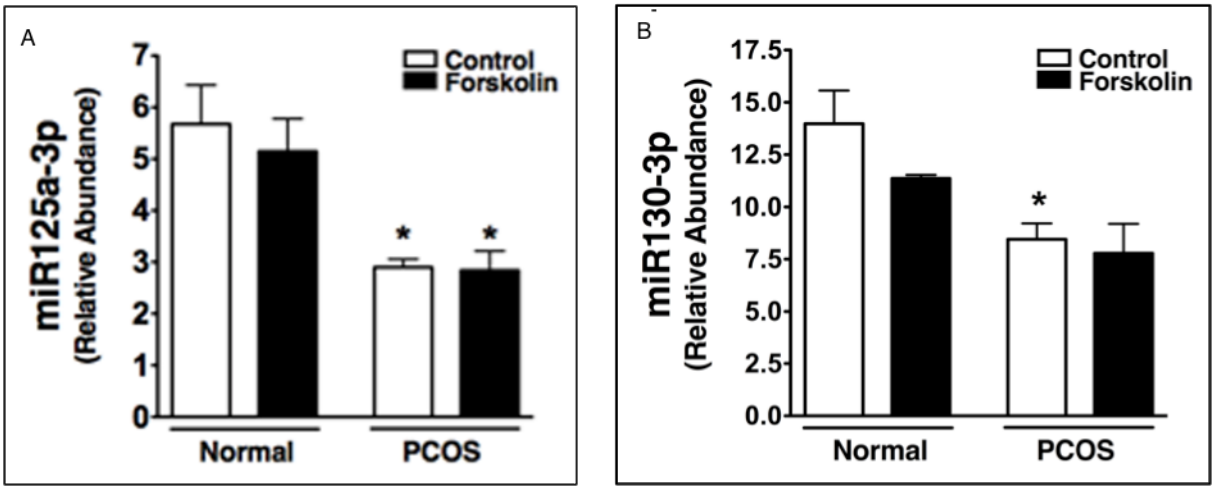


Figure. 3.2. Decreased miR-125a-3p and miR-130b-3p expression in PCOS theca cells. miR-125a-3p and miR-130b-3p expression was compared, by qRT-PCR, in theca cells propagated from 5 individual unaffected control and 4 individual PCOS women that were treated in the absence (C) and presence of 20 mM forskolin (F) for 16 h. As shown, both, miR-125a-3p (A) and miR-130b-3p (B) expression was decreased in PCOS theca cells under basal conditions as compared to unaffected control cells. However, only miR-125a expression was significantly reduced under forskolin stimulated conditions *(*P<0.05*)

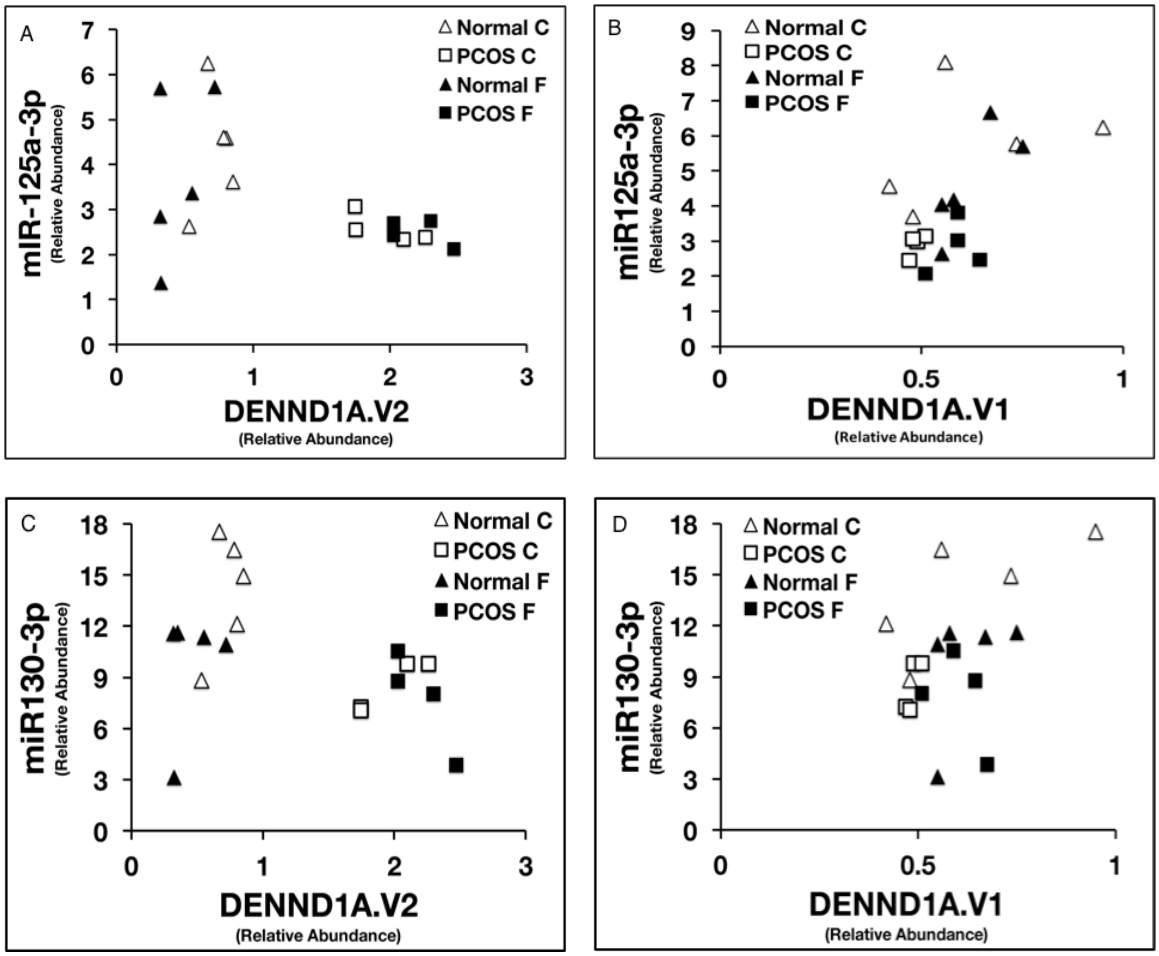


Figure. 3.3. Decreased miR-125a-3p expression is correlated with increased DENND1A.V2 expression. qRT-PCR miRNA expression of miR-125a-3p and miR-130b-3p was compared with qRT-PCR mRNA expression of DENND1A.V1 and DENND1A.V2 in same normal (n = 5) and PCOS (n = 4) theca cells. MiR-125a-3p expression showed significant negative correlation (Spearman's ρ = -0.52, *P* = 0.0279) with DENND1A.V2 expression (3.3.A) but showed no significant relation with expression of DENND1A.V1 (3.3.B) (Spearman's ρ = 0.41, *P* = 0.0874). MiR-130b-3p expression showed no significant relationships (*P* > 0.05) with both DENND1A.V2 (3.3.C) as well as DENND1A.V1 (3.3.D), but showed a trend of negative correlation with DENND1A.V2 (3.3.C, Spearman's ρ = -0.36) but not with DENND1A.V1 (3.3.D, Spearman's ρ = 0.39).

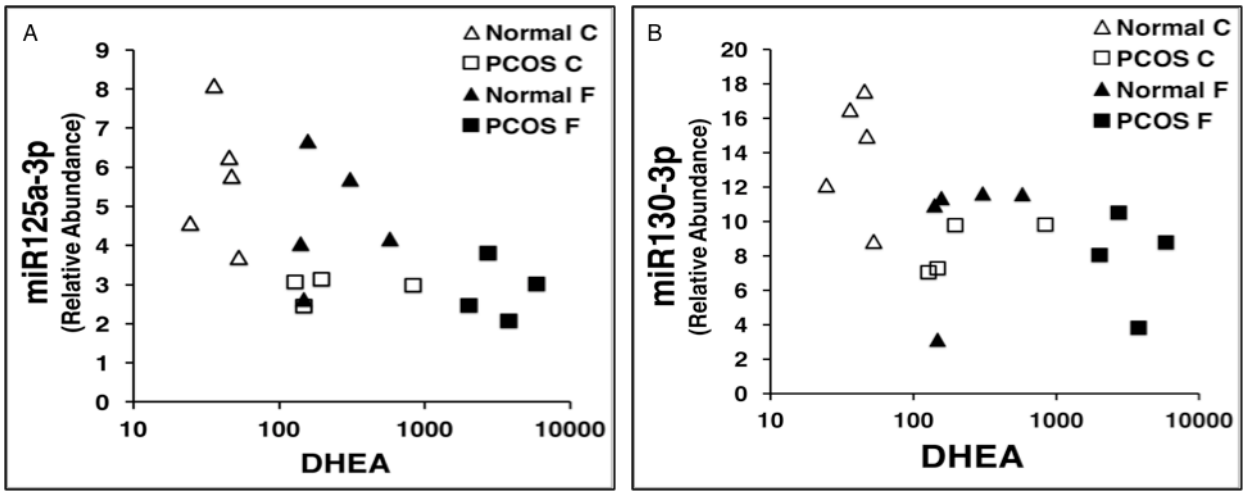


Figure. 3.4. Decreased miRNA expression is correlated with increased DHEA accumulation. qRT-PCR miRNA expression of miR-125a-3p and miR-130b-3p was compared with DHEA accumulation (estimated by ELISA) from same unaffected control (n = 5) and PCOS (n = 4) theca cell cultures. Significant negative correlation with DHEA accumulation was observed with miRNA expression of both miR-125a-3p (3.4.A, Spearman's ρ = -0.56, *P* = 0.0156) and miR-130b-3p (3.4.B, Spearman's ρ = -0.48, *P* = 0.0408)

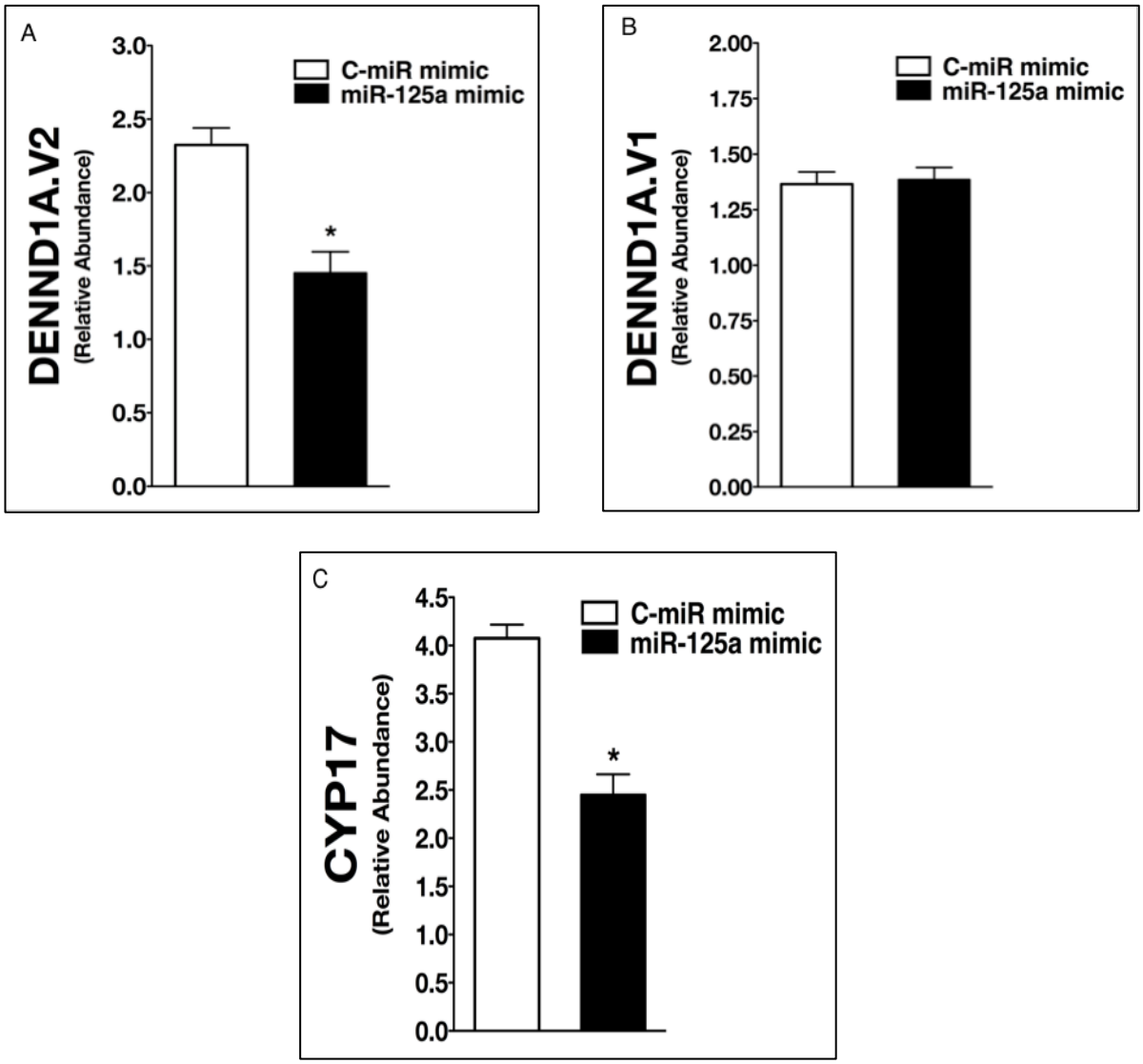


Figure. 3.5. MiR-125a-3p overexpression in H295R cells downregulates V2 and CYP17 mRNA expression, but has no effect on V1 expression. qRT-PCR mRNA expression for V1, V2 and CYP17 genes was analyzed in H295R cells treated with miR-125a-3p mimic (75nM) and negative control mimic (75nM) for 48 hours. MiR-125a-3p overexpression causes a down-regulation of V2 (P = 0.042) and CYP17 (P = 0.0237) mRNA expression when compared to control. No significant differences in V1 expression were observed (*P* > 0.05).

3.4. Discussion

PCOS is a complex disorder. Recent GWAS have identified candidate loci, some of which have been replicated by multiple laboratories, providing a basis for the molecular dissection of the pathophysiology of PCOS (34–37). Theca cells, the androgen producing cells of the ovary and central to PCOS pathophysiology, also have evidence of an altered transcriptome profile in PCOS, which could be contributed by, at least in part, by miRNA regulation. (77,78). However, a detailed study of miRNA expression and target gene analysis has not been performed in theca cells. Altered miRNA levels have been associated with several phenotypes in the PCOS spectrum including diabetes, insulin resistance and inflammation. An increasing number of studies are showing differential miRNA expression in the ovarian stroma, follicular fluid, cumulus cells and granulosa cells of women with PCOS (63,64,79–81,86,87). Our study is the first to identify differential miRNA expression in theca cell cultures from PCOS women as compared to normal cycling women.

The first GWAS conducted on Han Chinese populations identified 11 loci with significant associations with PCOS (34,35). The association of some of the loci identified in these studies including *DENND1A, RAB5B* and *THADA* to PCOS has also been replicated in European populations (38–40). Subsequent GWAS conducted on women of European descent identified other loci including *ERBB2, ERBB3, ERBB4, RAD50, KRR1, FSHB, GATA4, NEIL2* and *KCNA4* and also replicated association of some of the Han Chinese GWAS including *YAP1, THADA, DENND1A* and *C9orf3* (36,37). Of the total 22 loci identified in the 4 GWAS, association with some candidate genes including *INSR, FSHR, LHCGR* and GATA4 is plausible because of their known roles in reproductive

biology /ovarian function and their involvement in PCOS candidate gene studies. For e.g. *FSHR* and *LHCGR* are gonadotropins involved in regulation of folliculogenesis and ovulation. The *INSR* gene is involved in insulin metabolism and can play a crucial role in PCOS pathophysiology via insulin resistance. *GATA4* is a transcription factor known to be differentially expressed in PCOS granulosa and theca cells and is speculated to be involved in transcriptional activation of PCOS relevant genes including CYP17, which is responsible for androgen production. (8,24). However, the functional relevance of the remaining loci with PCOS is not clear and the exact molecular mechanism by which any of these candidate genes may contribute to the disorder is not known. Genotype-phenotype correlation studies of the susceptibility SNPs identified by GWAS have yielded low to modest relative risk ratios (41,88). This is not unexpected for a complex disorder like PCOS and establishing a causal relationship between the genetic variants identified in GWAS and pathophysiological phenotype is challenging. In a previous study, we demonstrated that a splice variant (DENND1A.V2) of *DENND1A,* one of the candidate genes identified in the Han Chinese GWAS and replicated in studies of women of European ancestry is involved in PCOS theca cell steroidogenesis (69,70). In the present study, we show that 6 of the 18 differentially expressed miRNAs in PCOS theca cells are predicted to target 12 of the 22 PCOS candidate genes identified in GWAS including *DENND1A.*

IPA pathway analysis of the 22 GWAS genes revealed an over-representation of several PCOS relevant features. Three of the top 5 canonical pathways are central to ovarian function and PCOS pathophysiology including ovarian cancer signaling, ErbB signaling and ErbB2-ErbB3 signaling. Although there is no firm evidence for

associations between PCOS and ovarian cancers because of limited sample sizes and confounding risk factors, perturbation of pathways in ovarian cancer are important for normal ovarian function (89). Ovarian morphology is altered in PCOS, which can be explained by changes in cellular growth and proliferation pathways, which overlap with ovarian cancer signaling. ErbB2, ErbB3 and ErbB4 are epidermal growth factors receptors, and EGFR signaling plays a crucial role in ovarian steroidogenesis and oocyte maturation (90,91). GWAS genes also show an over-representation of diseases and disorders that are highly concordant with PCOS including organismal injury and abnormalities, endocrine system disorders and reproductive system disorders, with 20, 12 and 17 molecules respectively, from the GWAS genes dataset being significantly overrepresented in these diseases/ disorders. The physiological system development and functions also had similar trends with organ development, reproductive system development and function as well as organ morphology with 9, 10 and 10 molecules respectively. The top IPA network which had ~45% (10/22) of the GWAS genes had a network score of 24 and associated network functions including cancer, organismal injury and abnormalities and respiratory disease. These results suggested that more than one PCOS candidate gene are involved in PCOS pathophysiology, which warrants detailed study of other GWAS candidate genes in a PCOS context for e.g. compare gene expression patterns of the GWAS candidate genes in control unaffected and PCOS theca cell cultures.

Overlaying the GWAS genes network with the 18 differentially expressed miRNAs identified in this study revealed ~73% (16/22) PCOS GWAS candidate genes and ~44% (8/18) differentially expressed miRNAs in PCOS theca to be a part of a common

network. Our qPCR assays validated the expression patterns of some of these miRNA-target gene pairs *in vitro* in theca cells, including hsa-miR-125a-3p and hsa-miR-130b-3p down-regulation and increased expression of *DENND1A* isoform DENND1A.V2 in PCOS theca cells.

There are studies demonstrating that hsa-miR-93 is differentially expressed in PCOS, with evidence for decreased expression in human PCOS blastocysts (92) and increased expression in human PCOS adipocytes (93), granulosa cells (94) as well as in circulation (95). Target identification in IPA revealed that miR-93 is predicted to target ("High" confidence level) 3 of the 11 GWAS genes: *RAB5B*, *ZNF217* and *THADA*. However, we did not find any differences in miR-93 expression levels in PCOS theca cells in our sequencing as well as TaqMan based qPCR assays

It is important to note that this study provides further evidence for the functional role played by *DENND1A* in PCOS as both the miRNAs, hsa-miR-125a-3p and hsa-miR-130b-3p, whose differential expression in PCOS theca has been validated *in vitro*, are predicted to target the *DENND1A.* The gene is also a part of the common network shown in Figure 3.1. In addition, decreased miR-125a-3p expression is correlated with increased DENND1A.V2 expression and increased DHEA accumulation, which is characteristic of PCOS cells. The miR-125a-3p mimic studies in addition confirmed that the predicted interaction of miR-125a-3p & DENND1A is specific for V2 and not V1. Decreased miR-125a-3p expression can possibly explain, at least in part, the V2 overexpression in PCOS theca cells as documented previously, which was not explained by any of the other genetic mechanisms investigated by Tee et al and described in Chapter 2 in this thesis. (42,69). MiRNAs are usually involved in complex

regulatory networks and their expression patterns are themselves regulated by several factors. Even though a significant association between miR-130b-3p and DENND1A.V2 was not observed, a trend for inverse correlation with DENND1A.V2 but not V1 was found and miR130-3p was inversely correlated with DHEA production. Thus, further investigation of miR130-3p with respect to PCOS pathophysiology is warranted.

In addition to *DENND1A* and miR-125a-3p and miR-130b-3p, there are other genes in the network that deserve further discussion including *YAP1* and *ZNF217*. Though not a lot is not known about these genes in the context of reproduction, we have evidence showing that an isoform of *YAP1* – YAP1.V2 and ZNF217 are significantly downregulated in PCOS theca cells as compared to unaffected control theca cells and reduced expression is correlated with increased DHEA accumulation. (Data not shown) MiRNA - target gene interactions are complex and as mentioned above, correlated inverse expression patterns are not always possible to observe as each gene can be regulated by multiple miRNAs. Performing luciferase-reporter assays is crucial to confirm the miRNA-target gene interaction *in vitro*. Even though, expression patterns are not directionally correlated with the expression patterns of the miRNAs predicted to target these genes, it is important to note that two other PCOS GWAS candidate genes from the same network show evidence for differential expression associated with functional relevance of increased androgen production that is central to PCOS.

These results suggest a potential epigenetic mechanism (in lieu of miRNA regulation) of V2 overexpression in PCOS theca cells via miR-125a-3p regulation. These results also support our hypothesis that miRNAs and *DENND1A* are involved in a complex network, perhaps in combination with other GWAS genes and can explain the ovarian

hyperandrogenemia associated with and central to PCOS.

Acknowledgement

We thank Bruce Miller, Jessica Biegler, Lauren Sauders-Miller and Kira Oshaben for their technical support. This research was funded by the National Institutes of Health grants U54HD3449 (to J.F.S. and J.M.M), R01HD033852 (to J.M.M), R01HD058300 (to J.M.M) and R01HD083323 (to J.M.M and J.F.S.)

Chapter 4: Rare Mutations and Potentially Damaging Missense Variants in Genes Encoding Fibrillar Collagens and Proteins Involved in Their Production Are Candidates for Risk for Preterm Premature Rupture of Membranes

Please note: This chapter is under review at PLoS One as
Bhavi Modi, Maria Teves, Laurel Pearson, Hardik Parikh, Piya Chaemsaithong, Nihar Sheth, Tim York, Roberto Romero, Jerome Strauss. *Rare Mutations and Potentially Damaging Missense Variants in Genes Encoding Fibrillar Collagens and Proteins Involved in their Production are Candidates for Risk for PPROM*.

Abstract

Preterm premature rupture of membranes (PPROM) is the leading identifiable cause of preterm birth with ~ 40% of preterm births being associated with PPROM and occurs in 1% - 2% of all pregnancies. We hypothesized that multiple rare variants in fetal genes involved in extracellular matrix synthesis would associate with PPROM, based on the assumption that impaired elaboration of matrix proteins would reduce fetal membrane tensile strength, predisposing to unscheduled rupture. We performed whole exome sequencing (WES) on neonatal DNA derived from pregnancies complicated by PPROM (49 cases) and healthy term deliveries (20 controls) to identify candidate mutations/variants. Genotyping for selected variants from the WES study was carried

out on an additional 188 PPROM cases and 175 controls. All mothers were self-reported African Americans, and a panel of ancestry informative markers was used to control for genetic ancestry in all genetic association tests. In support of the primary hypothesis, a statistically significant genetic burden (all samples combined, SKAT-O p-value = 0.0225) of damaging/potentially damaging rare variants was identified in the genes of interest - fibrillar collagen genes, which contribute to fetal membrane strength and integrity. These findings suggest that the fetal contribution to PPROM is polygenic, and driven by an increased burden of rare variants that may also contribute to the disparities in rates of preterm birth among African Americans.

4.1. Introduction

Although there is strong evidence from twin-based studies that both maternal and fetal genetic factors contribute to preterm birth, attempts to identify specific loci contributing to prematurity in genome-wide association studies (GWAS) have largely failed to yield robust and reproducible findings (54,59,96–98). A number of candidate gene association studies have found significant relationships, but meta-analyses indicate that these associations are at best weak or population-specific (50,99,100). The disappointing output from genetic studies may relate to the subject inclusion and exclusion criteria, including the proximate cause of preterm birth - spontaneous preterm birth (sPTB) or preterm premature rupture of membranes (PPROM), population heterogeneity (partially attributable to genetic admixture), and environmental exposures, including viral and bacterial infections. Further, identification of genetic variants contributing to preterm birth risk is mostly based on the “common disease-common

variant" hypothesis, which assumes that a large number of common allelic variants can explain the genetic variance in complex diseases. This approach has been employed for several complex traits such as human height and schizophrenia and despite having large sample sizes required for success, there still exists an issue of "missing heritability" with identified variants being able to explain only a small fraction of the genetic variance observed (101). Considering the complex genetic nature of preterm birth, a systematic identification of even a small number of rare variants with moderate-large effect sizes in genes of known/putative biological significance can be helpful (54,102). None of the genetic studies on preterm birth so far have taken this approach.

PPROM is the leading identifiable cause of preterm birth, with ~40% of preterm deliveries being associated with PPROM (48). Previous studies have focused on selected candidate genes involved in infection/ inflammation pathways, with the majority of these studies investigating only maternal genomes (47,50,60,103,104) despite evidence of a fetal contribution established by twin studies (59,97).

Human fetal membranes consist of an inner layer, the amnion, and an adherent outer layer, the chorion. The amnion is the load-bearing component of the fetal membranes and major contributor to their structural integrity. The strength of the fetal membranes is influenced by both synthesis and degradation of the extracellular matrix (ECM) components. Fibrillar collagens and associated proteins are major components of the fetal membranes contributing to their tensile strength (105,106). Thus, defects in the fibrillar collagen synthesis and/or altered ECM metabolism can adversely affect fetal membrane integrity, and may result in preterm birth as a result of preterm rupture. Epidemiological studies show that fetuses/neonates with Ehlers-Danlos syndrome,

osteogenesis imperfecta, and restrictive dermopathy, disorders of ECM synthesis, are at increased risk of adverse pregnancy outcomes including PPROM (106–109). In addition, a functional promoter SNP in the *SERPINH1* gene, which encodes a chaperone protein (HSP47) necessary for fibrillar collagen synthesis, was previously shown to be associated with PPROM (110).

The primary hypothesis explored in this study is that PPROM is the result of rare mutations or variants in the fetal genes involved in the elaboration of the amnion ECM. An initial Whole Exome Sequencing (WES) was performed followed up by genotyping of select variants in additional samples of neonatal DNA from normal term pregnancies (controls) and pregnancies complicated by PPROM (cases). All of the neonates were delivered from mothers of self-reported African-American ancestry. In an effort to identify functional variants with large effect, the analysis was selectively focused on damaging mutations: either nonsense mutations or frameshift mutations that precluded production of a functional protein, and missense mutations, which were predicted to be damaging or potentially damaging. Furthermore, the investigation was restricted to genes whose disruption could theoretically promote PPROM by weakening the fetal membranes by disruption of the ECM fibrillar collagens and genes encoding proteins involved in their production.

We discovered that rare heterozygous, nonsense, frameshift and damaging missense mutations were more prevalent in the genomes of neonates born of pregnancies complicated by PPROM as compared to normal term controls. The combined burden of the rare damaging variants identified in the fetal genome yielded a statistically significant genetic association with PPROM. These results suggest that PPROM may

be caused by infrequent genetic variants that modulate fetal membrane strength leading to weakening of the membranes and ultimately ending in premature rupture.

4.2. Materials and Methods

4.2.1. Study Population: The initial WES was performed on 49 case and 20 healthy term control neonatal DNA samples. Additional genotyping of select variants was performed on an independent cohort of 188 case and 175 control neonatal DNA samples. Subjects were self-reported African-American women and their neonates receiving obstetrical care at MCV Hospitals, Richmond, VA (all samples in the initial WES) and Hutzel Hospital in Detroit, MI. The study was approved by the respective Institutional Review Boards and written informed consent was obtained from mothers before sample collection. Demographic and clinical data were obtained from surveys and medical records. Control DNA samples (n = 20 + 175) were obtained from neonates of singleton pregnancies delivered at term (> 37 weeks of gestation) of mothers with no prior history of PPROM or preterm labor. Cases of PPROM (n = 49 + 188) were defined as neonates from pregnancies complicated by spontaneous rupture of membranes prior to 37 weeks of gestation. The diagnosis of membrane rupture was based on pooling of amniotic fluid in the vagina, amniotic fluid ferning patterns and a positive nitrazine test. Women with multiple gestations, fetal anomalies, trauma, connective tissue diseases and medical complications of pregnancy requiring induction of labor were excluded.

4.2.2. Ancestry Estimates: Genetic ancestry was estimated to control for the presence of population structure in all genetic association tests. Genetic ancestry estimates were generated in a two-way model of admixture, European and West African, for the

neonates of each self-reported African-American study subject using 102 ancestry informative markers (AIMs), single nucleotide polymorphisms with large allele frequency differences between ancestral populations (Appendix, Table A.4.1). Prior allele frequencies derived from the HapMap West Africans (YRI, Yoruba in Ibadan, Nigeria) and Europeans (CEU, CEPH Utah residents with ancestry from northern and western Europe) were used to estimate individual genetic ancestry estimates following a maximum likelihood approach with mean allele frequency difference between ancestral populations ($\delta = 0.733$) (111–113).

4.2.3. Whole Exome Sequencing: Whole Exome Capture and Sequencing was performed on the initial set of samples at BGI (BGI, Cambridge, MA) using the SureSelect Target Enrichment System Capture Process and high-throughput sequencing on an Illumina HiSeq2000 platform. Raw image files are processed by Illumina base calling Software 1.7 for base calling with default parameters and the sequences of each individual are generated as 90bp paired-end reads. The raw sequence data generated from the Illumina pipeline were used for bioinformatic analysis.

4.2.4. Read Mapping and pre-processing: Raw sequence data for each individual were mapped to the human reference genome (build hg19) using the BWA-MEM algorithm of Burrows-Wheeler Aligner (v 0.7.12) (114). This was followed by a series of pre-processing steps – marking duplicates, realignment around indels and base quality recalibration. PCR duplicates were marked within the aligned reads using Picard tools. (http://picard.sourceforge.net) Next, mapping artifacts around indels were cleaned up using the RealignerTargetCreator, the IndelRealigner and the LeftAlignIndels walkers of

the Genome Analysis ToolKit (GATK) (115,116). Inaccurate / biased base quality scores were recalibrated using the BaseRecalibrator, the AnalyzeCovariates and the PrintReads walkers of GATK, which use machine learning to model these errors empirically and adjust the quality scores accordingly. Alignment statistics for each sample were calculated on the "clean" sample BAM files.

4.2.5. Variant Discovery and Quality Filtering: The pre-processing steps were followed by variant calling using the HaplotypeCaller walker of GATK on each sample BAM file. Variant sites are identified by taking into account the haplotype likelihood predicted by building Dr. Brujin-like graphs in regions where the data displays variation relative to the hg19 reference genome. This step is also guided using the dbSNP, and Mills and 1000G gold standard SNP and indel databases. The output is a set of unfiltered/raw SNP and indel calls in the Genomic Variant Call Format (gVCF) file. Sample-specific gVCFs were merged into a single VCF file and a cohort-wide joint genotyping was performed using the CombineGVCFs walker of GATK. Finally, Variant Quality Score Recalibration (VQSR) was performed to assign the statistical probability to each variant call and produce a call-set distilled to a desired level of truth sensitivity. The raw SNP call-set was filtered using the GATK VariantFilter module, with variants required to pass the following criteria – "QUAL >= 30" AND "DP >= 25". The raw indel call-set was filtered with variants required to pass the following criteria – "QD > 2.0" AND "FS < 200.0" AND "InbreedingCoeff > -0.8" AND "ReadPosRankSum > -20.0".

4.2.6. Annotation and Filtering for Genes and Variants of Interest: SnpEff was used for annotating the functional effects of *high-quality* SNPs and INDELs on genes, transcripts and protein sequences including: a) their genomic location (*i.e.,* intron, 5' or

3' untranslated region, upstream/downstream of a transcript, or intergenic region); b) their consequence on protein sequence (*i.e.*, stop-gained, missense, frameshift); c) known variants from dbSNP (117), ClinVar (118), and the 1000 Genomes Project (119).

A total of sixteen candidate genes were selected for investigation of rare variants based on their involvement in the extra-cellular matrix (ECM) composition and synthesis and previously linked to connective tissue disorders such as classical types of Ehlers-Danlos syndrome (types I and II) as well as Ehler's-Danlos Syndrome types VIIA and VIIB, osteogenesis imperfecta type II and restrictive dermopathy (106). These genes encode for major ECM components including fibrillar collagens (*COL1A1, COL1A2, COL2A1, COL3A1, COL5A1, COL5A2)* and associated proteins *(CRTAP, ELN*) as well as enzymes involved in collagen processing and ECM production (*ADAMTS2, BMP1, LEPRE1, LOX, LOXL1, SERPINH1, ZMPSTE24, FKBP10*). Further analyses were focused on variants affecting only coding regions of the selected genes to best identify functional variation and this included nonsense, frameshift, splice site and damaging missense variants. Damaging missense variants were selected on the basis of most deleterious predictions in both Polyphen2 (HumDiv - probably damaging) as well as SIFT (damaging) platforms. When multiple isoforms were identified, only variants with the most deleterious damaging predictions for all isoforms were included.

4.2.7. Custom Genotyping: The variants identified and selected for further analysis from Whole Exome Sequencing (Table 4.2) were validated and additional samples (an independent cohort of additional 188 cases and 175 controls) were genotyped for the selected variants. Genotyping was performed on the Agena (previously Sequenom)

MassArray iPLEX platform following manufacturer's instructions (120) at the University of Minnesota Genomics Center.

4.2.8. Testing for Genetic Association: The combined set of variants identified in the initial WES and by additional genotyping were tested for genetic association using the combined Optimized Sequence Kernel Association Test (SKAT-O) software package in R version 3.2.3 (121–123) with default parameters adjusting for small sample size and ancestry estimates as covariates. A burden test was selected because the variants under study were rare (variant frequency in sample set < the calculated fixed frequency threshold T = 0.034), in the coding region and known/ expected to alter amino acid sequence and thus assumed to contribute to risk with same direction if effect (124).

4.2.9. Statistical Analysis: Mean levels of demographic variables were tested using a 2-tailed Student's t-test. Count data (for gravidity and parity) was square-root transformed before performing tests. P-values < 0.05 were considered statistically significant.

4.3. Results

4.3.1. Study Population: The characteristics of the subjects used in the study are presented in Tables 4.1.A and 4.1.B. There were no significant differences in maternal age, gravidity and parity between the cases and controls. As expected, the pregnancies complicated by PPROM had a significantly shorter gestational age at delivery than the term pregnancy control group ($p < 0.001$) and the PPROM neonatal birth weights were also significantly lower ($p < 0.001$) for both sample sets. Individual neonatal genetic

ancestry estimates were calculated and used to compare the ancestry proportions between cases and controls and no differences were found. (Appendix, Table A.4.1)

4.3.2. Variant Discovery: The analysis of WES was focused on damaging mutations (non-sense and frameshift mutations) and predicted damaging missense variants as identified by SIFT and Polyphen2 in the preselected genes of interest in neonatal DNA derived from of 49 pregnancies complicated by PPROM and 20 normal term pregnancies.

4.3.3. Variants Identified in Genes involved in ECM Components and ECM Synthesis: The variants identified in the selected genes of interest in the initial WES are described in Table 4.2. The position-specific annotation of the variant in the protein ("within feature" column) describes the molecular processing outcome of that particular region in the final protein product. An additional 188 cases and 175 controls were genotyped for these variants.

In the discovery WES, all variants identified were unique to cases. A heterozygous nonsense mutation in *BMP1* (rs116360985), an enzyme involved in procollagen processing, was discovered in 1 PPROM case. This mutation truncates the protein at amino acid residue 721 in a protein that has an isoform of 730 amino acids, so the functional significance is not clear. A frameshift mutation (rs137853883) in the *FKBP10* gene, which encodes for a chaperone protein involved in ECM metabolism, was identified in 1 case. Heterozygous predicted damaging (Polyphen2 and SIFT) missense variants were found in *COL1A2*; rs139528613 in 3 cases and rs145693444 in 2 PPROM cases. Predicted damaging heterozygous missense variants were found in *COL5A1* (rs2229817, rs116003670 and rs61739195) in 3 different PPROM cases. Although

rs61739195 and rs2229817 are predicted to be damaging by both Polyphen 2 and SIFT, these variants are listed to be benign in ClinVar. All other mutations and missense variants are listed as having unknown clinical significance.

Interestingly, two missense variants, rs201234519 and rs78690642, both predicted to be damaging by Polyphen2 and SIFT were identified in *COL2A1*, which encodes a fibrillar collagen found in cartilage and tendons, and not previously thought to be expressed in amnion. Rs78690642 was present in 5 cases and rs201234519 was present in one different case with none of the controls having either of the two variants. We performed an RT-PCR on RNA extracted from fetal membranes from normal term pregnancies (n = 6) using *COL2A1* specific primers, and obtained a robust amplicon that was sequence-verified to be *COL2A1* mRNA (Appendix, Figure A.4.1). Thus, it is possible that *COL2A1* mRNA, and possibly protein, is expressed in fetal membranes, and this has been overlooked in previous studies.

None of the heterozygous variants in the *COL1A2*, *COL2A1 and COL5A1* genes appeared in the same subject. However, one case had a predicted damaging mutation in *COL1A2* (rs139528613) as well as a nonsense mutation in *BMP1* (rs116360985).

Missense variants that were not predicted to be damaging using our stringent criteria were found in the fibrillar collagen genes as well as genes involved in their synthesis (Appendix, Table A.4.2). A number of these variants were novel and only detected in cases. In some instances one of the predictive algorithms suggested that the variant was potentially damaging (e.g., rs201944190), but their significance with respect to PPROM remains to be evaluated.

Observed allele frequencies for both the initial WES as well as the follow up study are shown in Table 4.3. In total, WES identified 7 predicted to be damaging missense variants in the candidate genes involved in ECM formation in 14 cases, one frameshift variant in *FKBP10* and one nonsense mutation in *BMP1* in one case each and none in the 20 controls. In addition, the follow-up study revealed an increased burden of the damaging missense variants in 10 additional cases and 6 controls. The nonsense mutation in *BMP1* was identified in an additional 4 cases and 5 controls and the frameshift variant in *FKBP10,* was not found in any of the additional samples genotyped. Combining the two sample sets revealed an increased frequency of the risk allele in cases as compared to controls.

4.3.4. Genetic Association Analysis: To determine if the rare variants collectively contributed to PPROM risk, we performed a genetic burden test using the combined initial WES and follow up genotypes including adjustment for West African ancestry. The omnibus SKAT-O test yielded a significant association for the rare variants with PPROM (p-value = 0.0225).

Table 4.1.A. Characteristics of the study population in the initial WES set

Cases (n = 49) and Controls (n = 20) were compared for key patient characteristics. Values represent means with SDs reported in parentheses for each group.

Characteristic	Cases Mean (SD)	Controls Mean (SD)	p-value
Maternal Age (years)	27.57 (5.67)	25.8 (4.86)	0.198
Gestational Age at Delivery (weeks)	29.59 (4.17)	38.55 (1.19)	**< 0.001**
Neonatal Weight (kgs)	1.47 (0.65)	3.21 (0.53)	**< 0.001**
Gravidity	1.73 (0.54)	1.78 (0.48)	0.685
Parity	0.85 (0.72)	1.04 (0.62)	0.280

Table 4.1.B. Characteristics of the study population used for follow-up genotyping.

Cases (n = 188) and Controls (n = 175) were compared for key patient characteristics. Values represent means with SDs reported in parentheses for each group.

Characteristic	Cases Mean (SD)	Controls Mean (SD)	p-value
Maternal Age (years)	26.76 (5.91)	26.6 (6.02)	0.798
Gestational Age at Delivery (weeks)	31.05 (3.63)	39.47 (0.10)	**< 0.001**
Neonatal Weight (kgs)	1.7 (1.07)	3.35 (0.31)	**< 0.001**
Gravidity	1.91 (0.59)	1.89 (0.58)	0.09
Parity	1.06 (0.77)	1.10 (0.69)	0.62

Table 4.2. Variants identified in genes involved in ECM composition and synthesis The positional and putative functional impact of the variants identified in the initial WES (unique to cases in the WES) in the selected genes of interest and selected for additional genotyping are shown. (PPD: Propeptide = part of a protein that is cleaved during maturation or activation, chain = extent of polypeptide chain in the mature protein).

Gene	SNP ID	Location	Allele Change	Type	AA Position (Residue Change)	Within Feature
BMP1	rs116360985	Chr8: 22058684	C > T	Nonsense	721 (R > Ter*)	Chain
FKBP10	rs137853883	Chr17:39975558	- > C	Frameshift	278 (G > R)	Chain
COL1A2	rs139528613	Chr7: 94028386	G > A	Missense	41 (R > H)	PPD
COL1A2	rs145693444	Chr7: 94038721	G > T	Missense	294 (V > F)	Chain
COL2A1	rs201234519	Chr12:48391685	G > T	Missense	64 (P > H)	PPD
COL2A1	rs78690642	Chr12:48367327	C > T	Missense	1374 (G > S)	Chain
COL5A1	rs116003670	Chr9:137582842	G > A	Missense	65 (R > Q)	Chain
COL5A1	rs2229817	Chr9:137726950	C > T	Missense	1757 (T > M)	PPD
COL5A1	rs61739195	Chr9:137708884	C > T	Missense	1379 (P > S)	Chain

Table 4.3. Allele frequencies of the variants identified

The table shows the allele frequencies of the putative risk allele (RAF) of variants listed in Table 4.2 in the general populations of CEU - Northern Europeans from Utah (European-American), AFR –African (combined African populations) and ASW – Americans of African ancestry in Southwest USA (admixed African Americans) ancestries as reported in the 1000 Genomes Project [50] and their observed risk allele frequencies in the initial WES and in the follow up study on an independent sample cohort used for custom genotyping, separate and combined. Please note that the AFR allele frequencies constitute a super population, which includes the allele frequencies from all African populations in the 1000 Genomes Project including the ASW.

SNP ID	RAF in CEU/AFR/ASW (1000 genomes)	RAF in Initial WES Case/ Control	RAF in Follow-Up Case/ Control	RAF in Combined Case/ Control
rs116360985	0.000/0.005/ 0.000	0.01/ 0	0.011/ 0.014	0.011/ 0.013
rs137853883	NA	0.01/ 0	0/ 0	0.002 / 0.00
rs139528613	0.000/0.011/ 0.025	0.031/ 0	0.006/ 0.009	0.011/0.008
rs145693444	0.000/0.004/ 0.008	0.02/ 0	0/ 0	0.004/ 0
rs201234519	0.000/0.002/ 0.000	0.01/ 0	0.006/ 0	0.006/ 0
rs78690642	0.000/0.014/ 0.025	0.05/ 0	0.009/ 0.003	0.018/ 0.003
rs116003670	0.000/0.007/ 0.000	0.01/ 0	0.006/ 0.006	0.006/ 0.005
rs2229817	0.000 /0.001/ 0.008	0.01/ 0	0.003/ 0	0.004/ 0
rs61739195	0.005 /0.000/ 0.000	0.01/ 0	0/ 0	0.002/ 0

4.4. Discussion

The present study using WES to identify mutations and damaging variants in the neonatal genome associated with PPROM supported the working hypotheses that rare variants could be significant contributors to preterm birth, particularly mutations or variants that would alter ECM integrity, resulting in weaker fetal membranes or increased susceptibility of ECM proteins to proteolytic enzymes.

Crosslinked networks of several collagen types constitute the major components of the ECM of the fetal membranes. Altered expression patterns or altered metabolism of any of the ECM or collagen components could lead to a loss of integrity of the fetal membranes (105,106). Some studies have shown that PPROM membranes have an altered amnion collagen content as compared to normal term pregnancies (125–127). We anticipated identifying variants that would disrupt the ECM, and looked for mutations and variants in genes encoding proteins involved in the production of the major ECM proteins in fetal membranes, especially variants that would alter the collagen content or structure of the amnion. With the exception of *BMP1* and *FKBP10*, we discovered no damaging mutations (nonsense, frameshift, splice junctions) in the selected genes of interest. *BMP1* is a metalloproteinase involved in collagen processing. *FKBP10* codes for a chaperone protein, FKBP65, which is known to be associated with the extracellular matrix protein - tropoelastin. Mutations in *FKBP10* have been found in several family members with osteogenesis imperfecta and clinical consequences of these mutations, one of which is the frameshift mutation identified in our study (rs137853883), included loss of FKBP65 protein function leading to delayed type I procollagen secretion and improper crosslinking of collagen (128,129).

The tensile strength of fetal membranes is determined by the assembly of fibrillar collagens (I, III, V) into fibrils. The size of the fibrils is mainly determined by type V collagen in association with types I and III collagens and proteoglycans (105,130). Both *COL5A1* and *COL5A2* are involved in the production of type V collagen, which participates in early fibril initiation, determination of fibril structure and matrix organization (131,132). Type I collagen is involved in fibril formation and consists of two alpha-1 chains (*COL1A1*) and one alpha-2 chain (*COL1A2*). Mutations in the *COL1A1* and *COL1A2* genes are known to cause rare forms of Ehler's Danlos Syndromes, types VIIA and B and osteogenesis imperfecta types I and II (106,133,134). Mutations in *COL5A1* and *COL5A2* genes resulting in haploinsufficiency or structural modifications of type V collagen are common causes for classical Ehler's Danlos Syndrome (types I and II) (135). These disorders have been associated with increased risk of PPROM when the fetus/neonate is affected (106,107,109). A case-control study with case-parent triads and case-mother dyads suggested a significant association of *COL5A1* (combined fetal-maternal association) and *COL5A2* (fetal association) with spontaneous preterm birth (131). Our study identified several predicted to be damaging missense variants in the *COL1A2* and *COL5A1* genes that were unique to only cases in the initial WES. Missense variants were identified in the *COL1A1*, *COL3A1* and *COL5A2* genes but their predicted impact on protein function was benign.

Potentially damaging missense variants in the *COL2A1* gene were also discovered with increased frequency in PPROM cases. *COL2A1* codes for cartilage collagen and until now there has been no evidence of *COL2A1* expression in the amnion making the significance of this finding uncertain (136). However, human amniotic membranes

(HAM) have been used as a source of stem cells for chondrocyte culture, where chondrocytes grown on the chorionic side of the HAM express type II Collagen (137,138). RT-PCR using mRNA from fetal membranes (amnion) obtained from normal term pregnancies and *COL2A1* specific primers, revealed detectable *COL2A1* mRNA expression, raising the possibility of low levels of expression of *COL2A1* protein that might play a significant role in fetal membrane integrity. Alternatively, the detected RNA could be generated from illegitimate transcription of the *COL2A1* gene. The putative functional significance of the variants could be due to variants in genes that are in linkage disequilibrium with the *COL2A1* SNPs or they could have a disrupting impact on overlapping coding sequences on the DNA strand opposite the *COL2A1* gene. A noncoding RNA (LOC105369752) of unknown function does reside there. There was no linkage disequilibrium (LD) information available for rs201234519. Rs78690642 is in LD with three other SNPs in the *COL2A1* gene (rs1455684563, rs76519927 and rs2071358).

It is important to note that there are significant disparities in the prevalence rates of PPROM with African-American women experiencing a 2-fold increased risk of PPROM as compared to European-American women. This disparity cannot be explained by socio-economic factors alone and genetic variation and gene-environment interactions are involved (49,50). Most of the rare variants described in this study are more prevalent in individuals of West African descent than European descent in the general population. Interestingly, this is also true for the *SERPINH1* promoter SNP that has previously been associated with PPROM (110). Admixed populations such as the African Americans have varying proportions of West African and European genetic

ancestry contributions across individuals and also differences in groups across different regions within the US (139,140). Even though the sample set in our initial WES and the independent sample set on which custom genotyping was performed consist of individuals from two different regions, Richmond and Detroit respectively, the fact that their combined SKAT-O gives a significant association is promising, suggesting a higher functional impact of the rare variants identified. Moreover, the fact that the combined burden of rare variants, which are of African origin (all except rs61739195), is significantly associated with increased PPROM risk suggests that the increased prevalence of PPROM in African-American populations is partly attributed to these rare population-specific alleles.

There are potentially other rare variants inactivating transcription factors or disrupting transcription factor binding sites that might result in reduced production of fibrillar collagens in the amnion. These were not explored in our study, nor were epigenetic factors that could influence collagen gene expression. These should be explored in future research. Conversely, variants that affect expression of matrix degrading enzymes, particularly the matrix metalloproteinases, could contribute to PPROM risk. This has been suggested by association studies with promoter variants in the MMP1, MMP8, and MMP9 genes (141–143).

In summary, using a screen to detect deleterious genetic variants that could promote PPROM in pregnancies hosted by women of African-American descent, we discovered evidence that rare damaging non-sense and frameshift mutations and predicted to be damaging missense variants in a variety of genes involved in negatively modulating the ECM metabolism-related genes are more prevalent in neonates born from pregnancies

complicated by PPROM than normal term pregnancies. Despite sample size being a limitation in our study, the variants identified strongly suggest that the fetal contribution to PPROM is polygenic and driven by multiple rare rather than common genetic variants

ACKNOWLEDGEMENTS

We thank Ms. Sonya Washington for enrolling subjects for the study at MCV Hospitals, Richmond VA. This research was funded by National Institutes of Health Grants R01 HD073555 and P60 MD002256. This research was also supported, in part, by the Perinatology Research Branch, Division of Intramural Research, *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Institutes of Health, Department of Health and Human Services (NICHD/ NIH); and, in part, with Federal funds from NICHD, NIH under Contract No. HSN275201300006C

Chapter 5: Expression Patterns of the Chromosome 21 MicroRNA Cluster (miR-99a, miR-125b and let-7c) in Chorioamniotic Membranes.

Please note: This chapter is accepted for publication in Placenta as

Bhavi Modi, Sonya Washington, Scott W. Walsh, Colleen Jackson-Cook, Kellie J. Archer, Jerome F. Strauss, III. *Expression Patterns of the Chromosome 21 MicroRNA Cluster (miR-99a, miR-125b and let-7c) in Chorioamniotic Membranes.*

Abstract

Trisomy 21 (T21) is the most common chromosome abnormality in humans and is associated with a spectrum of phenotypes, including cognitive impairment, congenital heart defects and immune system defects. In addition, T21 is also associated with abnormalities of fetal membranes including chorioamniotic separation, delayed fusion of the chorioamniotic membranes, defects in syncytiotrophoblast formation, as well as amniocyte senescence. There is evidence indicating miRNAs encoded by sequences on chromosome 21 (Chr-21) are involved in several of the cognitive and neurological phenotypes of T21, but the role of Chr-21 derived miRNAs in fetal membrane abnormalities associated with T21 has not been investigated. In the current study, we determined the expression patterns of three miRNAs derived from a cluster on Chr-21 –

hsa-miR-99a, hsa-miR-125b and hsa-let-7c in chorioamniotic membranes obtained from term pregnancies with spontaneous rupture (n = 20). Tissue and location specific expression patterns within the chorioamniotic membranes were identified. The rupture zone in the choriodecidua had distinct expression patterns compared to other fetal membrane locations. Despite the increased gene dosage associated with T21, the expression of all three miRNAs was reduced in cultured T21 amniocytes as compared to cultured euploid amniocytes. *In silico* analysis of experimentally validated targets of the three miRNAs suggest these Chr-21 derived miRNAs play a potential role in fetal membrane rupture and the fetal membrane defects associated with T21.

5.1. Introduction

Trisomy 21 (T21) is associated with abnormalities of the fetal membranes, including chorioamniotic separation and delayed fusion of the amnion and chorion (144–146). Cytotrophoblast abnormalities and early amniocyte senescence have also been reported (147,148). Early senescence of fetal extraembryonic tissues has been proposed to contribute to preterm birth (149). The underlying pathophysiological explanations for these abnormalities have not been elucidated. Epigenetic alterations in the chorion and amnion (differential DNA methylation) have been reported, raising the possibility that both chromosome dose, as well as epigenetic factors, contribute to dysfunction associated with cells in the fetal membranes (150).

MicroRNAs (miRNAs) are small (20-24 nucleotides) single-stranded, noncoding, regulatory RNA molecules. They are involved in post-transcriptional regulation of gene expression either by complimentary binding to the 3' untranslated region of their target

mRNA, thereby inhibiting translation or inducing mRNA degradation (61,62). Altered expression of microRNAs, including those encoded on chromosome 21 (Chr-21), are thought to contribute to the phenotypes associated with T21 (151,152). However, the role of these miRNAs in the fetal membrane abnormalities associated with T21 has not been studied in detail. We examined the pattern of expression of a cluster of Chr-21 microRNAs that are potential candidates for altering fetal membrane function in T21. We determined the relative levels of expression of the 3 microRNAs in this cluster in the amnion and choriodecidua, as well as differences in expression in different locations within the fetal membranes. Examination of *in vitro* cultured amniocytes from normal pregnancies and pregnancies hosting a T21 fetus revealed that T21 is associated with reduced expression levels of these three microRNAs. Based on the known roles of these microRNAs in the epigenetic control of gene expression, some of the fetal membrane structural defects and cellular abnormalities associated with T21 could be explained by altered expression of this microRNA cluster.

5.2. Materials and Methods

5.2.1. Study Population: Self-reported African-American women admitted at MCV Hospitals, Richmond, VA, with spontaneous labor at term (defined as more than 37 weeks of gestation) with rupture of membranes (n = 20) were recruited for the study. Women with multiple gestations, fetal anomalies, connective tissue diseases and pregnancy complications requiring induction of labor were excluded. Fetal membrane specimens were collected as described below from 12 male and 8 female fetal subjects delivered from 20 term pregnancies.

In order to test the effect of gene dosage on the expression of the Chr-21 localized miRNA cluster that were assessed in this study, cultured amniocyte specimens from local cell repository of cryopreserved specimens were evaluated from fetuses (race unknown) with a T21 complement (n = 7; T21 cases) or a disomic chromosomal complement (n = 6; controls). Fetal sex of the amniocyte cultures were as follows: T21 cases (2 males, 5 females); euploid controls (4 males, 2 females).

The study was approved by the Institutional Review Board of MCV Hospitals, Richmond, VA. Written informed consent was obtained from the patients before collection of samples.

5.2.2. Collection of samples: Human placentas and fetal membranes were obtained at the time of vaginal delivery or when C-section was performed at term. Fetal membranes specimens were collected within 1-2 hours of delivery. After visual identification of the rupture zone (Figure 5.1) as done previously by Nhan-Chang et al (153), the two adherent fetal membranes (amnion and choriodecidua) were separated/ teased apart from the rupture tear line. A piece of membrane tissue (~ $2cm^2$) spanning the rupture site was dissected sharply from both the amnion as well as the choriodecidua, constituting the rupture zone amnion (RZA) and rupture zone choriodecidua (RZC) samples, respectively. The two membranes were continuously teased apart and pieces of both the amnion and choriodecidua were dissected at the mid region, which was halfway between the rupture site and the placenta, followed by at the peri-placental region, which is adjacent to the placenta. This method yielded a set of 6 specimens from each patient: 1) amnion from the rupture zone (RZA), 2) choriodecidua from the rupture zone (RZC) (3) amnion from the mid-region (MRA), 4) choriodecidua from the

mid-region (MRC), 5) amnion from the peri-placental region (PPA), 6) choriodecidua from the peri-placental region (PPC).

The amniocytes from T21 cases and euploid controls (defined above) were established in culture (monolayer cultures in flasks), propagated *in vitro* (1 – 3 passages), and cryopreserved using standard techniques (154,155). The frozen cell aliquots were retrieved and maintained on ice prior to RNA isolation, which was accomplished as described below. The amniocyte cultures constitute a mixed population of fibroblast cells, amniotic fluid cells (derived from fetal membranes and trophoblasts) and epithelial cells (derived from epithelial fetal skin, bladder and other epithelia). The tissue culture procedure selects for growth of fibroblast-like cells, followed by amniotic fluid cells and epithelial cells with decreasing proliferative capacity in culture respectively (155).

5.2.3. RNA isolation and miRNA qPCR: For the miRNA-qPCR assays, total RNA was extracted using miRVana miRNA isolation kit (Ambion Life Technologies) following the manufacturer's protocol. MicroRNA qPCRs were performed using the TaqMan MicroRNA Assays (Life Technologies) following the manufacturer's instructions. Briefly, 10ng RNA was reverse transcribed using the target (miRNA) specific stem-loop RT primer and the TaqMan MicroRNA reverse transcription kit. The cDNA was then amplified by real-time qRT-PCR using target specific TaqMan primer-probe mix. The qRT-PCR was performed in triplicates per sample. Different small RNAs (RNU48, RNU43 and miR-16) were tested for consistent expression across all samples and RNU48 was selected for normalization of the qRT-PCR expression data. The mean expression value for each miRNA was divided by the mean RNU48 expression value to normalize each sample.

5.2.4. Statistical analysis: For miRNA qPCR assays in amniocyte samples from T21 cases and controls, normalized qPCR expression values for the three miRNAs (hsa-miR-99a, hsa-miR-125b and hsa-let-7c) were log2 transformed and used for student's t-test for determining statistical significance ($P < 0.05$).

For miRNA expression analysis in the fetal membranes normalized expression values from qPCR for the three miRNAs (hsa-miR-99a, hsa-miR-125b and hsa-let-7c) from the 6 specimens collected for each of the 20 participating subjects were log2 transformed prior to statistical analysis. For each miRNA, a random coefficient mixed effects model predicting log2 expression was fit, including the location/tissue factor variable as a fixed effect and subject as a random effect (156). Linear contrasts from the mixed effects model were constructed to test the following: MRA versus MRC; PPA versus PPC; RZA versus RZC; MRA versus PPA; MRA versus RZA; PPA versus RZA; MRC versus PPC; MRC versus RZC; PPC versus RZC; and Amnion versus Choriodecidua. Adjusted p-values are reported for the linear contrasts using the single-step method ($p < 0.05$ was considered significant) (157).

To compare male and female fetal subjects with respect to miRNA $\log_2$ qPCR expression, Welch's t-test was performed for each miRNA (let-7c, miR-125b and miR-99a), location (MR, PP, RZ and VP) and tissue (Amnion, Choriodecidua), adjusting for possible unequal variances between the groups. A Bonferroni correction was used to control for multiple comparisons and an adjusted alpha level of 0.0071 or below was considered significant (157).

All statistical analyses were performed in the R programming environment (121).

5.2.5. MicroRNA target identification and pathway analysis. The miRNA target prediction filter in Qiagen's Ingenuity® Pathway Analysis tool (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity) was used to identify potential target genes for hsa-miR-125b, hsa-miR-99a and hsa-let-7c. Pathway analysis was performed in IPA on target genes with experimentally observed interactions with the miRNAs.

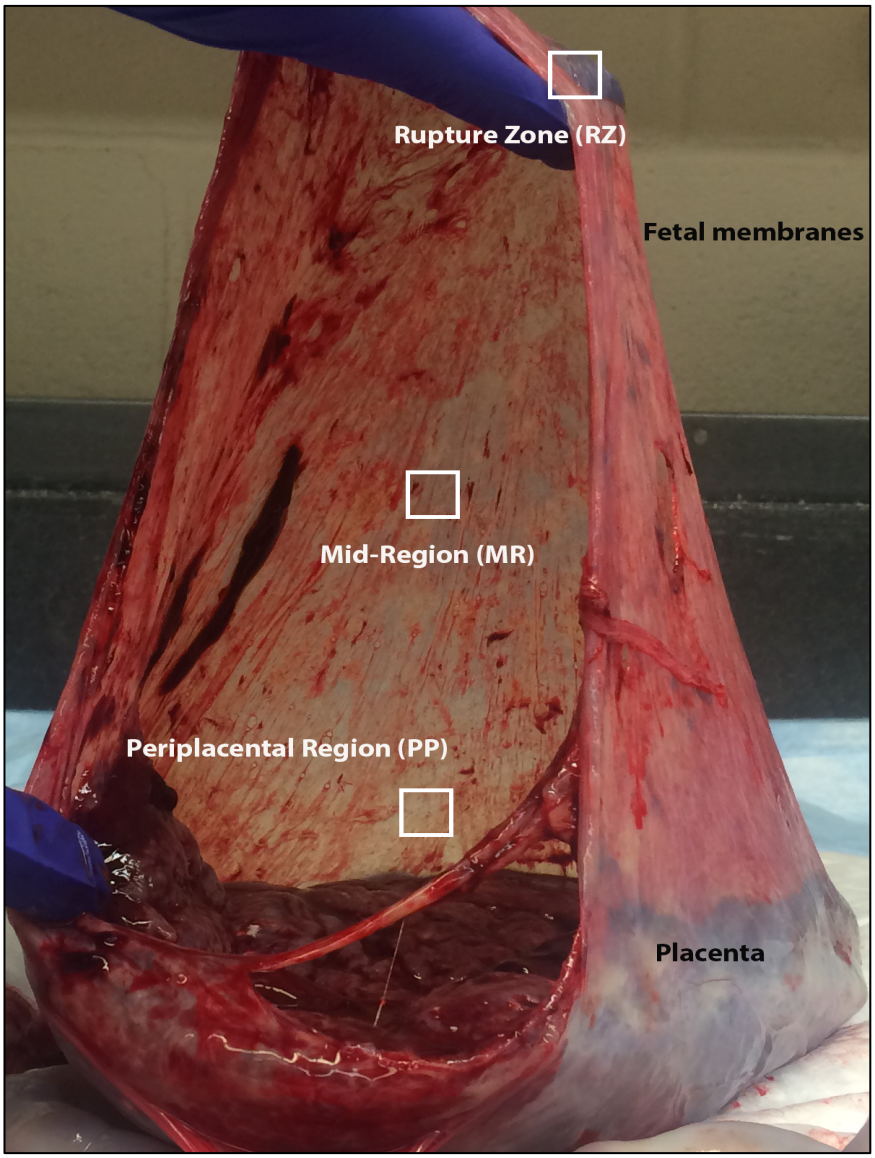


Figure 5.1. Sites of tissue collection

The placenta with the fetal membranes showing the representative sites of tissue collection by region i.e. gross sites for the membrane rupture called the Rupture Zone (RZ), Mid-Region (MR) and the Peri-placental (PP) regions. For RZ, MR and PP regions tissues were collected from both the amnion (A) and the choriodecidua (C) separately.

5.3. Results

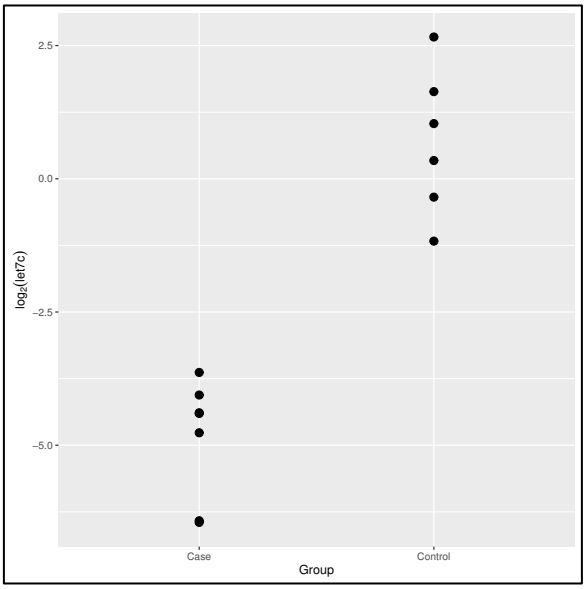
5.3.1. Trisomy impacts expression of the Chr21Orf34 microRNA cluster: The qPCR miRNA expression assays that were completed on cultured amniocytes revealed significant differences ($p < 0.05$) in the expression of all three miRNAs between the T21 cases ($n = 7$) compared to disomic controls ($n = 6$), with reduced expression levels in T21 amniocytes. For each miRNA, dot-plots of $\log_2$ expression (normalized by RNU48) by case and control are presented in Figure 5.2.

5.3.2. Differential miRNA expression patterns are observed between the amnion and the choriodecidua and at different locations within the choriodecidua: For each miRNA, normalized qPCR expression values for all samples ($n = 20$) were $\log_2$ transformed for each of the location/ tissue types and used for statistical analysis. Q-PCR results are shown in Figure 5.3. The differences and adjusted p-values for linear contrasts from the random coefficient mixed effects model for all pairwise comparisons are reported in Table 5.1 and visually displayed in Figure 5.4. Comparisons between the amnion and the choriodecidua at different locations revealed that all three miRNAs had significant expression differences at the rupture zone. For the non-rupture sites, for miR99a and let-7c there were significant expression differences at the mid-region, whereas miR-99a and miR-125b were differentially expressed between the amnion and the choriodecidua at the peri-placental region. None of the three miRNAs had significant differences at any location within the amnion. Comparatively, the choriodecidua showed significant differences (lower abundance) between the rupture zone and non-rupture sites, with all three miRNAs showing significant lower abundances between the rupture zone and the peri-placental region, and miR-99a and let-7c showing significant

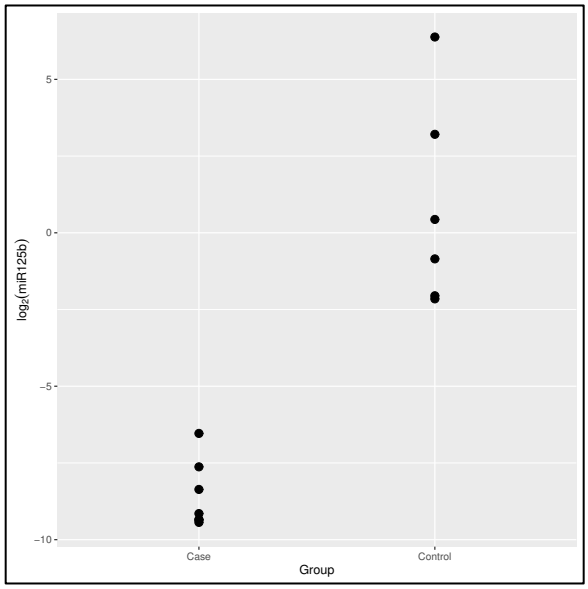
differences (lower abundance) between the rupture zone and the mid-region choriodecidua as well.

5.3.3. MiRNA Correlation in amnion and the choriodecidua by location: Pearson's correlation was used to examine the strength of linear relationship (pairwise) between the $\log_2$ expression of the three miRNAs by location for amnion (Table 5.2) and for the choriodecidua (Table 5.3). All pairwise miRNA relationships in all 3 locations in the amnion as well as the choriodecidua were positively correlated, though the strength of correlation (determined by ρ) varied with specimen tissue/ site. For the periplacental region, all pairwise miRNA comparisons showed significant correlations ($P < 0.05$) for both the amnion and the choriodecidua. Other significant pair-wise correlations in expression between miRNAs were observed in the choriodecidua at the mid-region (let-7c and miR-125b) and at the rupture zone (let-7c and miR-125b, as well as miR-99a and miR-125b) with miR-99a and miR-125b showing a strong correlation ($\rho > 0.5$) at this site.

(Fig. 5.2.A)



(Fig. 5.2.B)



(Fig. 5.2.C)

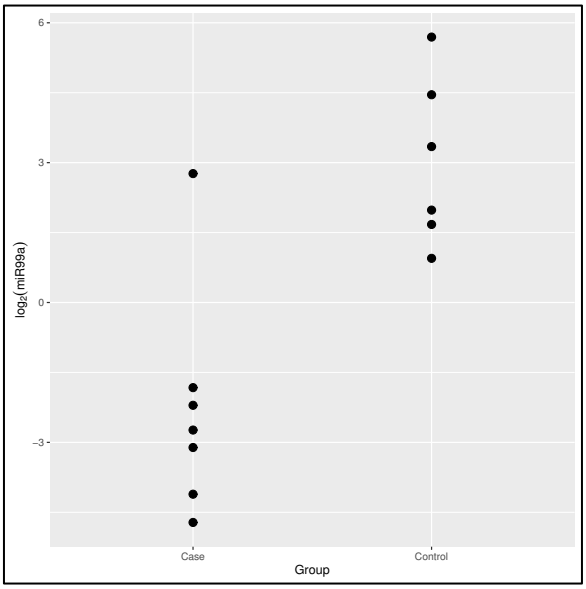
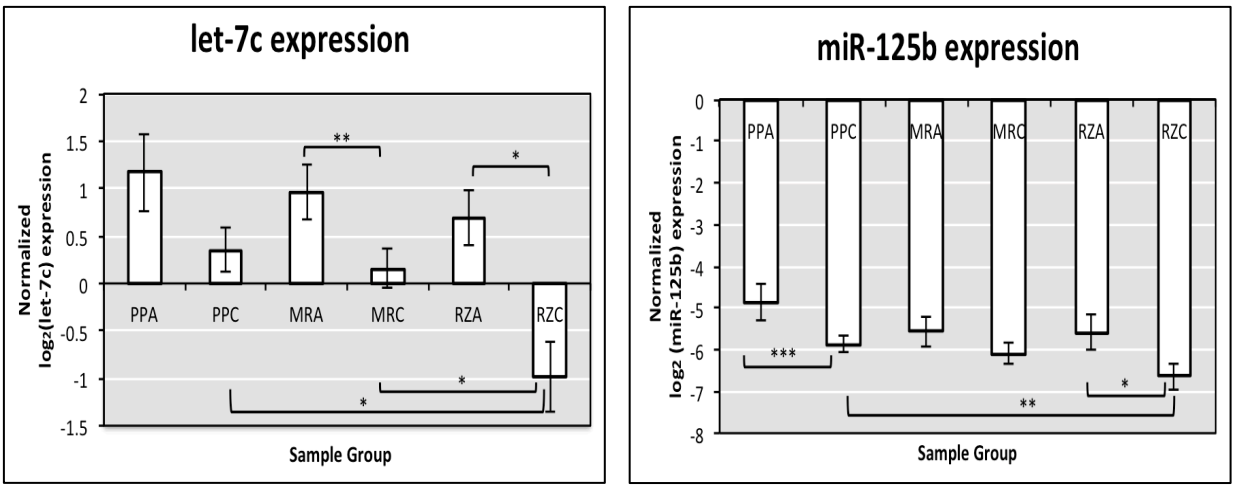


Figure 5.2 MiRNA Expression in Amniocytes

Significant differences ($p < 0.001$) in miRNA expression were observed between T21 case (n = 7) and disomic control (n = 6) amniocyte samples for hsa-miR-let7c (A), hsa-miR-125b (B) and hsa-miR-99a (C).

(Fig. 5.3.A) (Fig. 5.3.B)



(Fig. 5.3.C)

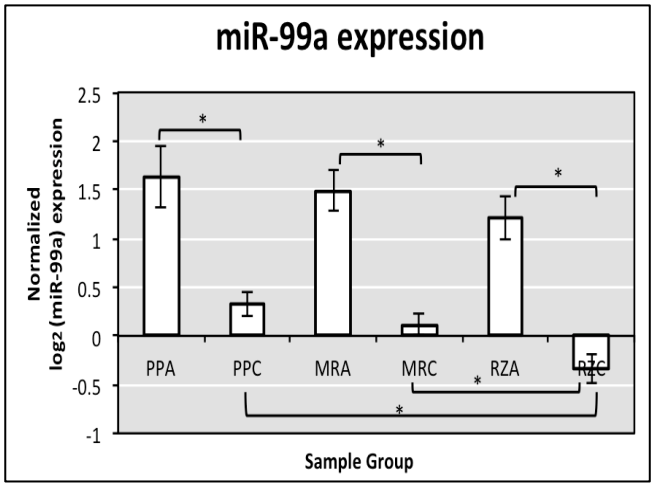
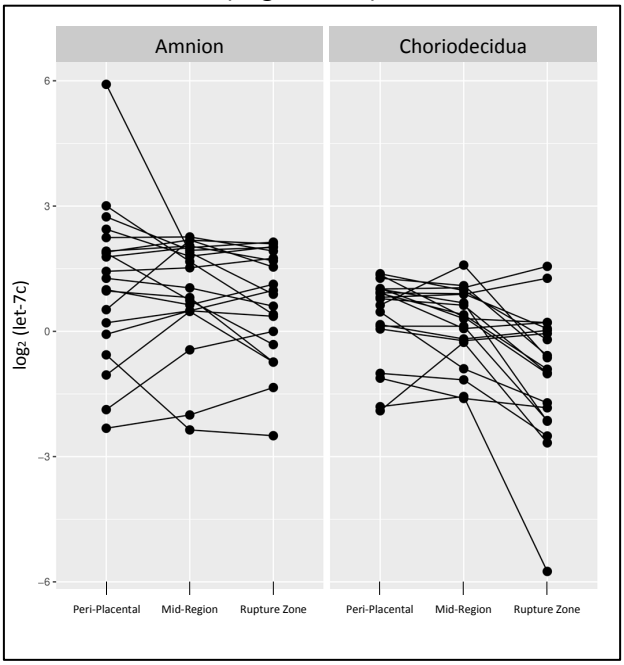


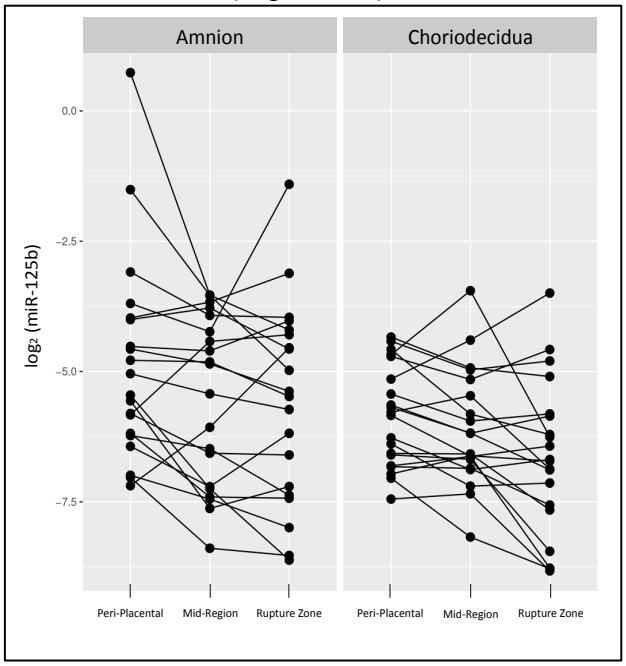
Figure 5.3. MiRNA Expression in Fetal Membranes

qPCR results for log2 expression of hsa-let-7c (A), hsa-miR-125b (B) and hsa-miR-99a (C) for each group/site of collection are shown. (PP=Periplacental, MR=MidRegion, RZ=Rupture Zone; A = Amnion/C = Choriodecidua). (Sample N = 20). Data are represented as Mean ± Standard Error (*P < 0.001, **P = 0.002, ***P<0.05)

(Fig. 5.4.A)



(Fig. 5.4.B)



(Fig. 5.4.C)

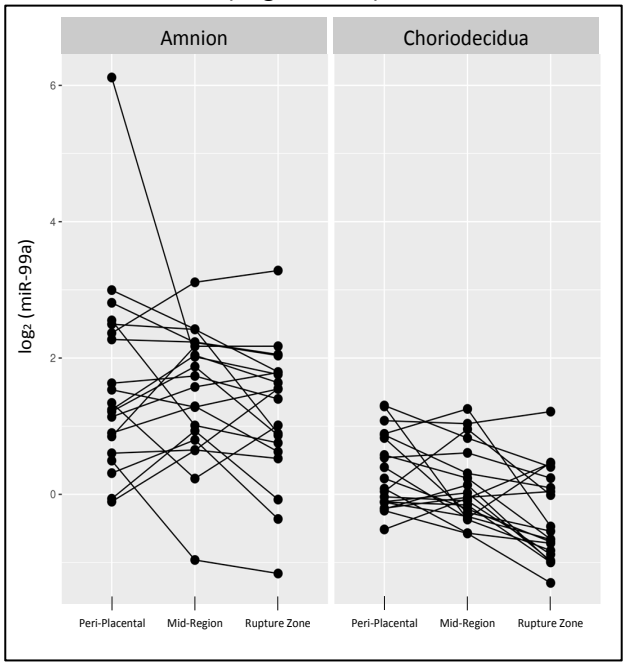


Figure 5.4. Comparison of miRNA expression by tissue and location. Dot-plot of log2 expression for hsa-let7c (A), hsa-miR-125b (B) and hsa-miR-99a (C) are shown for each specimen collection location, separated by amnion (left) versus choriodecidua (right). Each subject's observations are connected with a line.

Table 5.1. Pairwise comparisons of miRNA expression by tissue and location

The differences (log_2 contrast estimates) and adjusted p-values for linear contrasts from the random coefficient mixed effects model for all pairwise comparisons are shown for hsa-let7c, hsa-miR-125b and hsa-miR-99a. ($P < 0.05$, considered significant, shown in bold)

Comparison	Let-7c Estimate (Adjusted p-value)	miR-125b Estimate (Adjusted p-value)	miR-99a Estimate (Adjusted p-value)
MRA vs. MRC	0.8127 **(<0.001)**	0.5421 (0.1222)	1.3674 **(<0.001)**
PPA vs. PPC	0.8161 (0.0873)	1.0005 **(0.0331)**	1.3056 **(<0.001)**
RZA vs. RZC	1.6851 **(<0.001)**	1.0628 **(<0.001)**	1.5455 **(<0.001)**
A vs. C	1.1046 **(<0.001)**	0.8685 **(<0.001)**	1.4062 **(<0.001)**
MRA vs. PPA	-0.1980 (0.961)	-0.70407 (0.0643)	-0.1492 (0.9879)
MRA vs. RZA	0.2743 (0.287)	0.01895 (1.0000)	0.2807 (0.2066)
PPA vs. RZA	0.4722 (0.648)	0.72302 (0.3538)	0.4299 (0.5728)
MRC vs. PPC	-0.1945 (0.638)	-0.24567 (0.3453)	-0.2110 (0.4481)
MRC vs. RZC	1.1466 **(<0.001)**	0.53970 (0.0757)	0.4588 **(0.0027)**
PPC vs. RZC	1.3412 **(<0.001)**	0.78537 **(0.0024)**	0.6698 **(<0.001)**

Table 5.2. Pairwise Pearson's correlation for log2 miRNA expression by location for amnion. The strength of linear relationship (pairwise) between the $\log_2$ expression of the three miRNAs in the amnion by location is shown. Values represent Pearson's coefficient of correlation (ρ) with estimated p-values in parentheses. ($P < 0.05$ considered significant, shown in bold)

Comparison	Periplacental ρ (p-value)	MidRegion ρ (p-value)	Rupture Zone ρ (p-value)
let-7c & miR-99a	0.627 **(0.003)**	0.255 (0.28)	0.407 (0.07)
let-7c & miR-125b	0.560 **(0.01)**	0.209 (0.38)	0.029 (0.90)
miR-99a & miR-125b	0.541 **(0.01)**	0.291 (0.21)	0.336 (0.15)

Table 5.3. Pairwise Pearson's correlation for log2 miRNA expression by location for the choriodecidua. The strength of linear relationship (pairwise) between the $\log_2$ expression of the three miRNAs in the choriodecidua by location is shown. Values represent Pearson's coefficient of correlation (ρ) with estimated p-values in parentheses. ($p < 0.05$ considered significant, shown in bold)

Comparison	Periplacental ρ (p-value)	MidRegion ρ (p-value)	Rupture Zone ρ (p-value)
let-7c & miR-99a	0.469 **(0.04)**	0.445 (0.05)	0.383 (0.10)
let-7c & miR-125b	0.514 **(0.02)**	0.495 **(0.03)**	0.453 **(0.04)**
miR-99a & miR-125b	0.665 **(0.001)**	0.362 (0.12)	0.646 **(0.002)**

5.3.4. Pathway analysis of genes targeted by miRNAs: MicroRNA target prediction filter in IPA identified a total of 217 experimentally observed miRNA-target gene interactions for all three miRNAs. The 212 target genes from these interactions were used for core pathway analysis in IPA to identify pathways, diseases & bio-functions and networks that are "over-represented" in target genes of the Chr-21 microRNA cluster. The key findings from the pathway analysis are shown in Table 5.4. Of note, several of the functional annotations suggest involvement of cell survival, cell death and cellular development pathways (molecular mechanisms of cancer), which play a role in fetal membrane rupture. In addition, a significant number of targeted genes overlap with the categories of reproductive system disease (101 genes) and developmental disorder (45 genes), which also have a strong association with preterm birth, as well as T21. The top network (Figure 5.5), based on the functional annotations of the target genes, had 26 focus molecules (26 of the 212 target genes) and IPA Network Score = 48. The associated network functions involving the cell cycle, connective tissue development and function, and cancer. All of these functions have been shown to have strong correlations with membrane rupture and/ or phenotypic traits associated with Down's syndrome.

5.3.5. Fetal gender and miRNA expression: Fetal gender is known to be associated with adverse pregnancy outcomes (158,159). Male (n = 12) and female (n = 8) fetal subjects in the study were compared with respect to miRNA expression (log_2 qPCR estimates) for each miRNA for each of the 6 groups or sampling sites. Only let-7c in

mid-region choriodecidua samples was significantly different ($p < 0.0071$) when comparing fetal gender. (Appendix, Tables A.5.1 – A.5.3)

Table 5.4. Ingenuity pathway analysis of the genes targeted by hsa-miR-99a, hsa-miR-125b and hsa-let-7c. Key findings from the pathway analysis (performed in IPA) of the 212 genes targeted by the three miRNAs in the Chr21 cluster are described. The Top Canonical Pathways (5.4.A), Top Diseases and Disorders (5.4.B), Top Molecular and Cellular Functions (5.4.C) and Physiological System Development and Functions (5.4.D) over-represented in the target genes are shown.

5.4.A: Top Canonical Pathways

Name	p-value	Overlap
Bladder Cancer Signaling	1.04E-09	12.60%
Chronic Myeloid Leukemia Signaling	5.26E-10	11.50%
Glioblastoma Multiforme signaling	4.18E-13	10.70%
Colorectal Cancer Metastasis Signaling	3.80E-09	6.50%
Molecular Mechanisms of Cancer	3.61E-12	6.10%

5.4.B: Top Diseases and Disorders

Name	p-value	No. of gene
Organismal Injury and Abnormalities	3.7E-07 - 3.82E-18	153
Cancer	3.7E-07 - 3.82E-18	151
Reproductive System Disease	3.7E-07 - 5.73E-15	101
Tumor Morphology	2.02E-07 - 3.85E-17	51
Developmental Disorder	3.68E-07 - 2.88E-14	45

5.4.C: Top Molecular and Cellular Functions

Name	p-value	No. of genes
Cellular Growth and Proliferation	3.70E-07 - 1.34E-20	125
Cellular Development	3.70E-07 - 1.34E-20	112
Cell Death and Survival	3.32E-07 - 5.40E-18	110
Cellular Movement	3.10E-07 - 3.43E-13	70
Cell Cycle	3.70E-07 - 2.93E-16	64

5.4.D: Physiological System Development and Function

Name	p-value	No. of genes
Organismal Development	3.31E-07 - 1.03E-17	115
Organismal Survival	5.59E-15 - 3.97E-18	90
Tissue Morphology	1.22E-07 - 4.65E-14	78
Hematological System Development and Function	3.21E-07 - 6.85E-14	67
Lymphoid Tissue Structure and Development	3.21E-07 - 2.97E-14	62

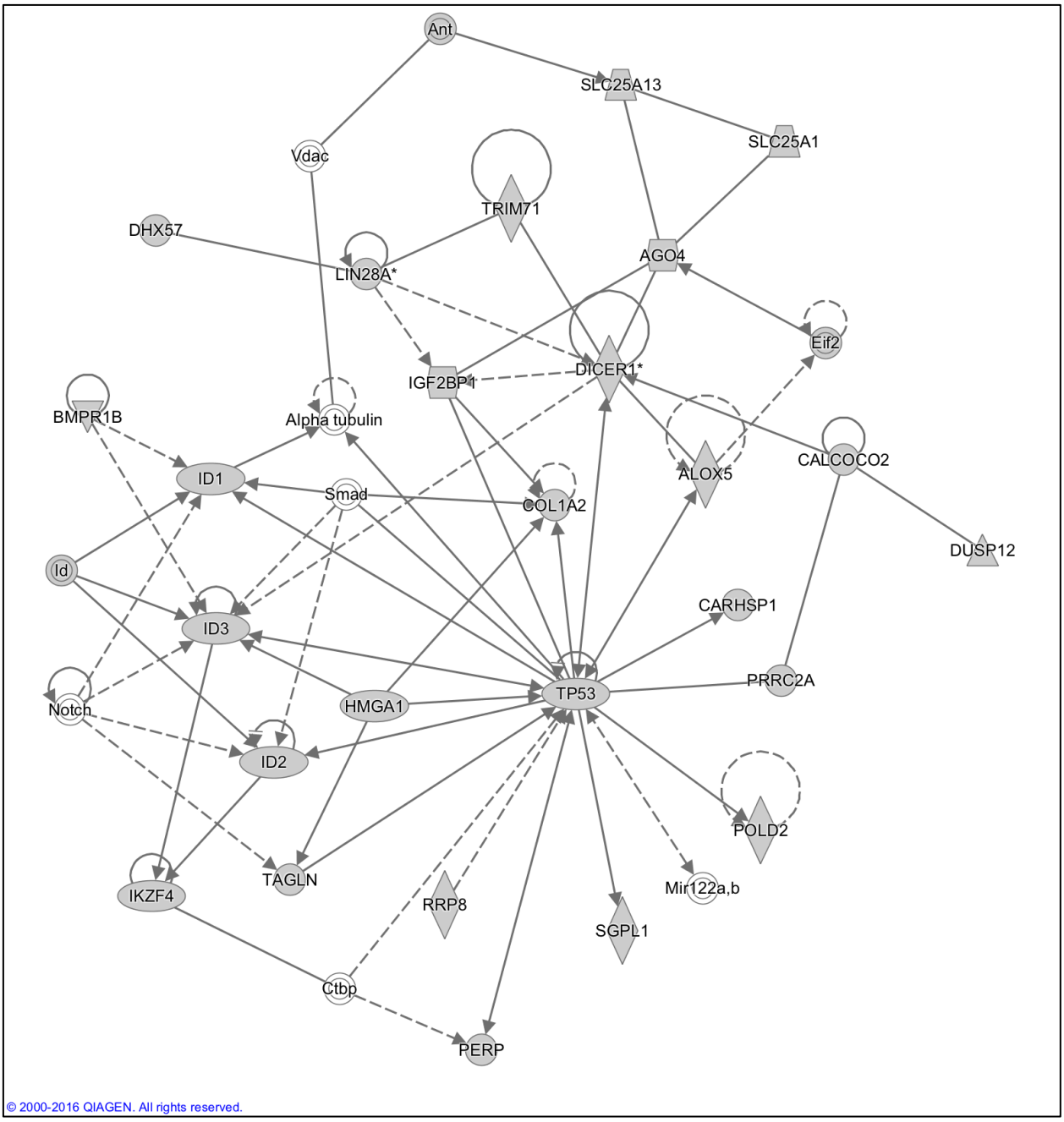


Figure 5.5. IPA network. The top IPA network (Score = 48) with 26 of the 212 experimentally observed target genes of the 3 miRNAs. The associated network functions are cell cycle, connective tissue development and function and cancer.

5.4. Discussion

Increasing numbers of investigators are identifying miRNAs to be involved in pathologies of several diseases and disorders, including the features of T21 (151,152,160,161). Expression patterns of miRNAs have been studied in the placenta as well as the chorioamniotic membranes and associations with human pregnancy, gestational age, and parturition and pregnancy complications (such as preeclampsia and preterm labor) have been found (65,162–164). MicroRNAs are key regulators of extracellular matrix (ECM) gene expression and homeostasis, which is perturbed in preterm premature rupture of membranes (PPROM). Interestingly, the relationship between miRNAs and ECM is thought to be bi-directional with studies suggesting that not only can miRNAs regulate ECM composition, but the ECM can also affect miRNA expression and function (165).

Despite the presence of three chromosomes 21 (a trisomic imbalance), the T21 amniocyte specimens showed reduced expression of the microRNA cluster. This has been observed by others who quantitated expression of the same microRNAs in other tissue types and found reduced expression in T21 (152,160). However, other studies have documented increased expression of these miRNAs in T21 as compared to euploid controls (151,161,166). These apparent discrepancies could be attributed to differences in the tissue specimens investigated or differences in normalization techniques. Reduced expression of the microRNAs in the face of increased gene dosage may reflect epigenetic changes such as DNA methylation that dampen expression (167). The reduced expression of the three microRNAs would presumably increase the expression of the proteins encoded by the target mRNAs. Although all

three miRNAs are known to be expressed in the placenta and chorioamniotic membranes, an in-depth expression analysis has not been performed. Prior studies investigating the roles of these miRNAs in context of other diseases reveal findings suggesting a true potential for their involvement in fetal membrane pathology. For example, miR-99a expression has been studied in several cancers and plays crucial roles in the PI3K/AKT and NF-κB signaling pathways (168,169). In fact, the miR-99a/let-7c/miR-125b locus on Chr-21 is phylogenetically conserved, is transcribed as a polycistronic message, and regulates TGFβ/Wnt signaling pathways in hematopoietic stem and progenitor cell homeostasis (170). Members of the miR-125 family of miRNAs are involved in diverse cell functions and have been associated with several diseases. MiR-125 miRNAs target a range of genes, including transcription factors and matrix metalloproteinases (MMPs), which are involved in ECM metabolism and also have important functions in immunological pathways. Another study found that miR-125b is upregulated in third trimester placentas compared to first trimester placentas, suggesting it has a role in gestational age (162,171). Let-7 miRNAs are negatively regulated by the RNA-binding protein, Lin28, and both act downstream of the NF-κB signaling pathways. Like miR-125b, let-7 miRNAs, including let-7c are upregulated in third trimester human placentas compared to the first trimester (162). In support of our results, another study found tissue-specific let-7c expression differences at term, with significantly reduced expression in choriodecidua as compared to amnion and placenta (172).

In the current study, in addition to the tissue-specific expression patterns, we also show differential expression patterns at different locations within the choriodecidua.

Differences between the rupture zone and the non-rupture sites are especially biologically relevant. The rupture site of fetal membranes is characterized by unique histologic and biochemical changes. These changes are observed even in the portion of fetal membranes overlying the cervix (thought to be eventual site of spontaneous rupture) in patients at term without labor. This area is termed as the "zone of altered morphology" (ZAM) and is characterized by increased thickness of connective tissue and decreased thickness of cytotrophoblast and decidua (130,173–175). These changes would require large-scale transcriptional reprogramming, which can be carried out, at least in part, by miRNA-regulation. Nhan-Chang *et al* (153) showed gene expression differences between the amnion and the choriodecidua at both the rupture as well as non-rupture sites, whereas within each fetal membrane individually, only the choriodecidua showed gene expression differences between the rupture and non-rupture sites. These results are in accordance with the miRNA expression patterns in our study, and support the concept that the altered gene expression in fetal membranes in the rupture zone could be due to altered miRNA expression patterns. It is important to note that the decreased thickness in the cytotrophoblast and maternal decidual cell layers at the rupture site in the chorion compared to non-rupture sites (130,173–175) can influence the observed miRNA expression levels at these sites. Furthermore, the genes identified by Nhang-Chang et al (such as *IL-6, IL1RN, CXCL12, ADAMTS5*) were involved in inflammatory response, immune tolerance pathways as well as ECM-receptor interactions. This suggests that miRNAs from the Chr21orf34 locus could be involved in regulation of these pathways in the fetal membranes as well, which is quite plausible considering the evidence for these functional roles is already established in

other tissues and disease states (165,168–171). *In silico* pathway analysis of the experimentally observed gene targets of the three miRNAs revealed an over-representation of genes related to cellular growth and proliferation, cell death and cell survival pathways, all of which are crucial in ECM function during membrane rupture (176). Notably, the top network based on functional annotation of the genes targeted by the three miRNAs included *COL1A2*, which is one of the major fibrillar collagens present in fetal membranes contributing to the tensile strength of the amnion and associated with preterm birth (105). The network also included *SMAD* which suggests involvement of the TGFβ-Smad signaling pathway which plays a role in ECM metabolism and fetal membrane rupture via transcription regulation of various collagens, MMPs and tissue inhibitors of metalloproteinases (TIMPs) (48,177).

In conclusion, we show evidence for distinct tissue and region specific miRNA expression patterns in fetal membranes from subjects with spontaneous term labor. These miRNAs are expressed at a reduced level in T21 and this alteration may promote abnormalities in fetal membranes associated with T21, including chorio-amnion separation or delayed fusion of the chorion and amnion.

Acknowledgements

This research was funded by National Institutes of Health Grants R01 HD073555 and P60 MD002256.

Chapter 6: Perspectives and Future Directions

Please note:

The section “6.2.1: A hypothetical model for DENND1A and GWAS candidate loci in PCOS” is adapted from a published journal article

(Jan M. McAllister, Richard S. Legro, Bhavi P. Modi, and Jerome F. Strauss III. *Functional genomics of PCOS: from GWAS to molecular mechanisms.* **Trends in Endocrinol Metab, 2015; 26** (3))

The section “6.2.3: Pharmacogenomics in PCOS” is part of a published journal article

(Cindy T. Pau*, Kai I. Cheang*, **Bhavi P. Modi**, Thushiga Kasippillai, Candace C. Keefe, Maria Shulleeta, William S. Evans, Lubna Pal, Jerome F. Strauss III, John E. Nestler, Corrine K Welt. *The Role of Variants Regulating Metformin Transport and Action in Women with Polycystic Ovary Syndrome.* (*Authors contributed equally); **Pharmacogenomics 2016; 17 (16)**)

6.1. Underlying themes

While all of the projects described in this thesis are more or less distinct, they all are focused on understanding the genetic and epigenetic differences leading to disease susceptibility. Each chapter employs a unique data set with the most relevant approaches used to answer specific questions. Beyond the specific research questions addressed in each chapter, certain wider themes were apparent across multiple projects, which pertain to the study of complex genetic studies as a whole. Central to the study of complex genetic traits is the understanding of gene expression regulation and function. This is a multistep process that includes defining the genetic architecture, identifying genes and variants within these genes associated with disease risk and translating these associations into disease mechanisms. As explained in the introduction, the chapters in this thesis take different approaches towards reaching this goal in the context of two complex reproductive disorders – PCOS and PPROM. Chapter 2 describes the investigation of different possible genetic mechanisms for V2 overexpression in PCOS and the negative results obtained indicate either we do not have sufficient data to clearly elucidate a genetic mechanism (owing to small sample size) or that epigenetic mechanisms may play a role and the results in Chapter 3 support our hypothesis by showing evidence for potential miRNA regulation of the *DENND1A* gene. In addition, it is important to note other epigenetic mechanisms including DNA methylation and histone modifications may be involved and need to be studied. In addition, Chapter 4 describes an approach to identify low frequency rare variants that contribute to PPROM risk susceptibility in select candidate genes by integrating bioinformatics and known disease pathology (discussed below).

One additional theme is the role of miRNAs contributing to the epigenetic mechanisms of complex disorders as discussed in Chapters 3 and 5. While DNA methylation and histone modifications are usually investigated in lieu of epigenetic regulation of gene expression, miRNAs also contribute to this landscape of gene expression regulation by post-transcriptional translational inhibition or mRNA degradation with no changes to the underlying DNA sequence. Differential miRNA expression patterns have contributed to several disease mechanisms including cancer, neurological disorders, cardiovascular diseases, diabetes etc. Chapter 3 shows evidence for differential expression for two different miRNAs in PCOS theca cells. In addition, differential expression was correlated to increased androgen (DHEA) production – which is a measure of the hyperandrogenism associated with PCOS. In addition, Chapter 5 describes tissue and location specific expression patterns of Chr-21 derived miRNAs in the chorioamniotic membranes in term pregnancies with spontaneous rupture. The results suggest a unique expression pattern in rupture site of the chorion, supporting our hypothesis that gene expression changes at the rupture site leading to membrane rupture at term could be regulated by miRNAs. This pilot study provides support for a detailed miRNA expression analysis in fetal membranes from term pregnancies with and without membrane rupture and pregnancies complicated by PPROM.

In addition, utilization of *in silico* network analysis in Chapters 3 and 5 really underlines the overall scope of this thesis by giving preliminary evidence for a common network of miRNAs (epigenetic) and genes (genetic) implicated either *a priori* (by association studies) or via validated and/ or predicted miRNA-gene interactions being involved in complex disease susceptibility.

Another common theme is integration of biological theory and standard next-generation genetic analysis techniques to make the studies more functionally informative. Identification of risk loci will have applicable value only when risk loci can improve our understanding of disease pathophysiology and in the long-term impact healthcare management and a “hypothesis-driven” approach can aid in designing studies that are better equipped in this direction. This was particularly helpful in Chapter 4 where knowledge of biological function of genes and association to biological etiology of disease allowed us to focus on a manageable number of candidate genes and variants there in from a large data set obtained from Whole Exome Sequencing. In addition, prior studies on biological relevance of *DENND1A* in PCOS directed the mechanistic and functional aspects of Chapters 2 and 3. Moreover, Chapter 5 integrates epidemiological evidence of fetal membrane abnormalities in Trisomy 21 and contribution of Chr-21 miRNAs to T21 phenotypes leading to the study of location and tissue specific expression patterns of Chr-21 miRNAs in chorioamniotic membranes.

The results obtained in each separate project reinforce the notion that both genetic and epigenetic mechanisms and their complex interactions need to be studied in detail to understand the complex traits. As discussed in the introduction, several of these mechanisms are even more challenging to study for complex reproductive traits and adverse pregnancy outcomes such as PCOS and PPROM respectively. This thesis is an example of how a better understanding of these concepts can be achieved by defining specific research questions and identifying the best methods to address those questions effectively. In addition, several of these studies have led to potential future directions that will add more clarity towards the genetic and epigenetic mechanisms in

complex reproductive disorders. The rest of the chapter entails brief descriptions of these future projects some of which are already in progress.

6.2 Future Directions

6.2.1. A hypothetical model for *DENND1A* and other GWAS candidate loci in PCOS

A previous study has described the role of increased expression of a *DENND1A* isoform - V2 contributing to PCOS pathophysiology via steroidogenic regulation in PCOS theca cells (70), however the precise etiology by which V2 alters theca cell function is not known. However, the structural features of the V2 protein and its cytoplasmic and nuclear location in theca cells described previously (70) provide important clues for speculating a possible molecular model for its function, perhaps in conjunction with other GWAS candidate loci (Figure 6.1.). V2 could be involved in modification of gonadotropins (*FSHR, LHCGR*) and *INSR* (all three are known GWAS candidate loci) signal transduction cascades via receptor-mediated recycling/ endocytosis through clathrin-coated vesicles by interactions via the clathrin-binding domain in the V2 structure. In addition, *DENND1A* functions as a Rab-specific guanine nucleotide exchange factor (GEF) that interacts with Rab-specific GTPases such as *RAB5B* (also a GWAS candidate gene) and are involved in endocytic processes. V2 could influence gene transcription of androgen producing genes by transport of ligand (*FSH, LH, INS)* and/ or receptor (*FHSR, LHCGR, INSR*) into the nucleus via these processes. Alternatively, V2 localized in the nucleus could be a scaffold for transcription factors some of which are GWAS genes including *TOX3,* YAP1, HMGA2 and ZNF217 or others

such as GATA family of transcription factors which ultimately regulate steroidogenic enzyme gene expression. These transcription factors can act independently of V2 to influence gene expression as well. Preliminary evidence (not shown) indicates V2-mediated activation and regulation of PI3K/MAPK pathways, which could then regulate several other downstream signaling cascades leading to transcriptional regulation of gene expression contributing to the PCOS phenotype. Detailed molecular studies designed to clearly define and outline these hypothesized interactions need to be performed.

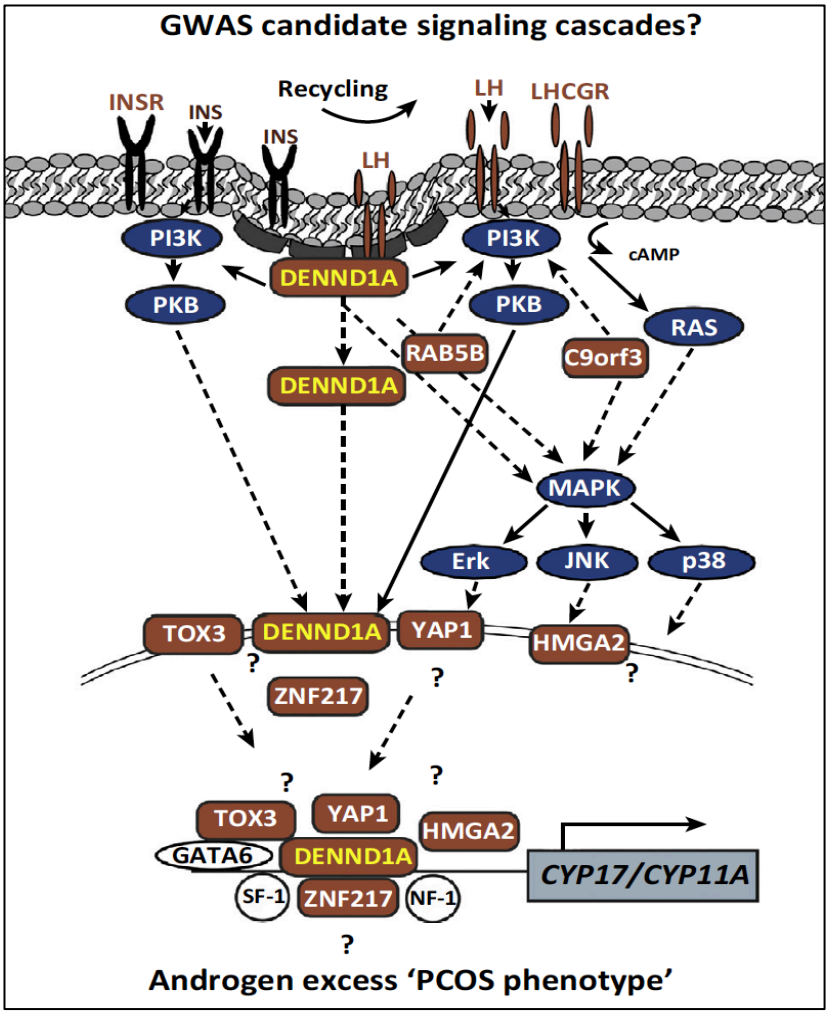


Figure. 6.1. A hypothetical model for DENND1A and other GWAS candidate loci in PCOS. The different possible mechanisms by which DENND1A and other GWAS candidate loci (shown in brown boxes) can contribute to the steroidogenic phenotype of PCOS theca cells is shown. The different mechanisms (described in detail in section 6.2.1) include gene transcription regulation via DENND1A mediated endocytosis of ligand/ receptor via clathrin-coated vesicles without or in conjunction with RAB5B and gene expression regulation via transcription factors dependent or independent of nuclear V2.

6.2.2. A transgenic mouse model for PCOS

The existing PCOS animal models are based on treatment of the animals with androgens, or prenatal exposure to androgens *in utero* (178,179). However, a mouse model created following a gene-based approach is lacking. Based on our studies with *DENND1A* in theca cells (70), it was hypothesized that global overexpression of human DENND1A.V2 in a transgenic mouse should reproduce the PCOS phenotype (at least the steroidogenic phenotype). In addition, this provides an opportunity to study the pathophysiological role of V2 in PCOS *in vivo*. A C57BL6 background was used to create three transgenic lines overexpressing V2 using the construct shown in Figure 6.2.

While transgene expression at the mRNA level was detectable in two of the three transgenic lines, establishing transgene protein expression was inconsistent and difficult. Fertility and fecundity did not appear to be different between wildtype (WT) and transgenic (Tg) animals. Steroidogenic phenotypes of increased plasma DHEA, androstenedione, progesterone and testorone levels were observed in Tg animals compared to WT. Unlike humans, mouse adrenal glands do not synthesize androgen (180). However, Cyp17 gene expression (mRNA and protein) was detected in the adrenal glands of Tg animals overexpressing V2, whereas, no expression was detected in the WT animals. Adrenal androgen production needs to be further confirmed by cortisol assays.

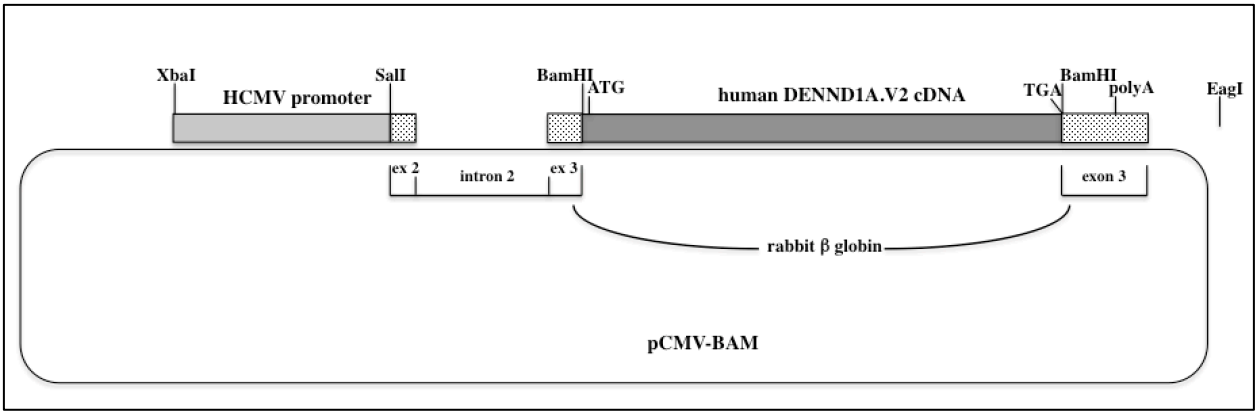


Figure 6.2. DENND1A.V2 construct. pCMV-BAM vector containing the human DENND1A.V2 cDNA under control of the CMV promoter, as well as rabbit β-globin sequences that provide an intron for efficient transgene expression and a polyadynelation site. This construct was used to create three transgenic lines on a C57BL/6 background.

6.2.2. Rare variant analysis in negative regulators of the immune response contributing to PPROM.

As described in Chapter 4, identification of low-frequency rare variants in the fetal genome is an approach towards identification of genetic variants contributing to disease susceptibility in PPROM that has not been used. As Chapter 4 focused on damaging variants in fibrillar collagen genes contributing to ECM structure and metabolism, using a similar approach, future studies will focus on identification of rare variants in fetal genes involved in the suppression of the inflammatory response triggering matrix degradation.

Chorioamnionitis (inflammation of fetal membranes) and/ or unchecked innate immune response resulting from bacterial infection are major risk factors associated to PPROM (181–183). Gene expression differences of key mediators of inflammation and innate immune response pathways in fetal membranes of normal term pregnancies and PPROM patients have been identified. Negative regulators on innate immune signaling help to keep inflammation in check. Host genetic factors significantly contribute to inter-individual variation in susceptibility to infection and severity on innate immune response. Identification of damaging variants that potentially disrupt expression and/ or function of genes involved in negative regulation of inflammation and the innate immune response may lead to identification of genetic predictors of PPROM. Using this rational and the same methodology described in Chapter 4, the candidate genes interrogated included anti-inflammatory cytokines such (*IL10, IL10RA, IL10RB*), suppressors of cytokine signaling (*SOCS1, SOCS2, SOCS3*), intracellular mediators (*CARD6, CARD8, NOD2, NFKBIE*) and pattern recognition receptors (*NLRP3, NLRP10, NLRP12, TLR10*).

Of particular note were the mutations/potentially damaging variants in *NFKBIE*, an inhibitor of the NFkB activation (Appendix Table A.6.1). In addition, predicted to be damaging missense variants were found only in PPROM cases in *NLRP10*, a negative regulator of the innate immune inflammasome pathway and *NOD2,* an intracellular protein in the NFkB pathway. A frameshift mutation rs2066847 was identified in the *NOD2* gene with allele frequency of 0.0042 in all PPROM cases combined (49 cases used in WES and 188 cases used for the follow-up genotyping). A nonsense mutation, rs765522475, was identified in the *NLRP10* gene with an allele frequency of 0.0021 in all PPROM cases combined. Detailed analysis of these variants in association to PPROM is warranted.

6.2.3. Pharmacogenomics in PCOS

As mentioned earlier, true value of identification of genetic variants impacting disease susceptibility lies in their applications towards understanding disease pathophysiology and contribution towards healthcare. Pharmacogenomics, which is the study of how genes and genetic variants affect a person's response to drugs, is now being increasingly applied to the study of complex genetic disorders as well. In an attempt to study the application of pharmacogenomics in PCOS, a small project to understand *the role of variants regulating metformin transport and action in women with Polycystic Ovary Syndrome* was performed. The study was a collaborative effort and has been published in *Pharmacogenomics* (184). A brief description of study design and key findings for the work performed at Virginia Commonwealth University are described below.

Metformin is widely used to treat women with PCOS based on the concept that it improves insulin resistance. However, reports have demonstrated that metformin does not improve insulin resistance, rather it improves fasting glucose levels and glucose effectiveness in women with PCOS. In addition, other reports indicate that metformin lowers testosterone levels and increases ovulation rates in a subset of women with PCOS independent of insulin action (185–189). Thus, the mechanism of metformin action in PCOS remains unclear and identifying the subset of patients who will benefit from metformin remains a challenge. Metformin is transported into hepatocytes by organic cation transporter 1 (*OCT1* or *SLC22A1*), out of hepatocytes and renal tubular epithelium by multidrug and toxin extrusion 1 (*MATE1* or *SLC47A1*), and eliminated through the renal tubule cells by organic cation transporter 2 (*OCT2* or *SLC22A2*). All three transporters are expressed in the ovaries, hepatocytes and skeletal muscle and genetic variants in these genes have been demonstrated to affect metformin transport response in patients with type 2 diabetes. In addition to metformin transport protein heterogeneity, a variant in linkage disequilibrium with the ataxia-telangiectasia mutated gene (ATM; rs11212617) has been associated with reduced HbA1C levels after metformin treatment through phosphorylation and activation of the AMPK pathway (190–194). This suggested that the genetic variability in metformin transporters and mediators may affect metformin levels and action in relevant tissues in women with PCOS and have clinical significance.

Genotyping for loss of function variants (associated with protein change) in *SLC22A1*, *SLC22A2, SLC47A1* and *ATM* was performed on genomic DNA isolated from women

with PCOS (n = 26). (Details regarding the other cohorts, clinical criteria and metformin dosage used can be found in the full version of the paper (184)).

Briefly, the study failed to demonstrate a relationship between metformin transporter variants and glucose, insulin and testosterone levels, a response to an oral glucose tolerance test, and ovulation in patients with PCOS. Thus, the data from the current study does not support an important effect of these variants in the response to metformin in PCOS.

However, it is important to note the heterogeneity in the different sample cohorts used in the study – including the strength and duration of metformin treatment, which can mask true effects. This highlights the need for replication studies with increased sample sizes and uniform study design.

Bibliography

1. Iyengar SK, Elston RC. The genetic basis of complex traits: rare variants or "common gene, common disease"? Methods Mol Biol [Internet]. 2007;376:71–84. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=17984539
2. Koboldt DC, Steinberg KM, Larson DE, Wilson RK, Mardis ER. XThe next-generation sequencing revolution and its impact on genomics. Cell [Internet]. 2013;155(1):27–38. Available from: http://dx.doi.org/10.1016/j.cell.2013.09.006
3. Zhao H, Chen ZJ. Genetic association studies in female reproduction: From candidate-gene approaches to genome-wide mapping. Mol Hum Reprod. 2013;19(10):644–54.
4. Kavanagh DH, Tansey KE, O'Donovan MC, Owen MJ. Schizophrenia genetics: emerging themes for a complex disorder. Mol Psychiatry [Internet]. 2015;20(1):72–6. Available from: http://dx.doi.org/10.1038/mp.2014.148
5. McPherson R, Tybjaerg-Hansen A. Genetics of Coronary Artery Disease. Circ Res. 2016;118(4):564–78.
6. Editorial. On beyond GWAS. Nat Genet [Internet]. 2010;42(7):551–551.

7. Lauritsen MP, Bentzen JG, Pinborg A, Loft A, Forman JL, Thuesen LL, et al. The prevalence of polycystic ovary syndrome in a normal population according to the Rotterdam criteria versus revised criteria including anti-Müllerian hormone. Hum Reprod [Internet]. 2014 Apr 1;29(4):791–801. Available from: http://humrep.oxfordjournals.org/content/29/4/791.abstract

8. Rosenfield RL, Ehrmann DA. The Pathogenesis of Polycystic Ovary Syndrome (PCOS). Endocr Rev [Internet]. 2016;(July):er.2015-1104. Available from: http://press.endocrine.org/doi/10.1210/er.2015-1104

9. Legro RS, Arslanian SA, Ehrmann DA, Hoeger KM, Murad MH, Pasquali R, et al. Diagnosis and treatment of polycystic ovary syndrome: an Endocrine Society clinical practice guideline. J Clin Endocrinol Metab [Internet]. 2013;98(12):4565–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24151290

10. Moran LJ, Misso ML, Wild RA, Norman RJ. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: A systematic review and meta-analysis. Vol. 16, Human Reproduction Update. 2010. p. 347–63.

11. Scicchitano P, Dentamaro I, Carbonara R, Bulzis G, Dachille A, Caputo P, et al. Cardiovascular risk in women with PCOS. Vol. 10, International Journal of Endocrinology and Metabolism. 2012. p. 611–8.

12. Zawadski JK. Diagnostic criteria for polycystic ovary syndrome, towards a rational approach, Dunaif A., Givens JR, Haseltine F., Polycystic ovary syndrome, 1992, 377-384. Blackwell Scientific, Boston;

13. The Rotterdam ESHRE/ASRM-sponsored PCOS consensus workshop group. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). Hum Reprod [Internet]. 2004;19(1):41–7. Available from: http://humrep.oxfordjournals.org/content/19/1/41.long

14. Azziz R, Carmina E, Dewailly D, Diamanti-Kandarakis E, Escobar-Morreale HF, Futterweit W, et al. Position statement: Criteria for defining polycystic ovary syndrome as a predominantly hyperandrogenic syndrome: An androgen excess society guideline. Vol. 91, Journal of Clinical Endocrinology and Metabolism. 2006. p. 4237–45.

15. Eilertsen TB, Vanky E, Carlsen SM. Anti-Mullerian hormone in the diagnosis of polycystic ovary syndrome: Can morphologic description be replaced. Hum Reprod. 2012;27(8):2494–502.

16. Casadei L, Madrigale A, Puca F, Manicuti C, Emidi E, Piccione E, et al. The role of serum anti-Müllerian hormone (AMH) in the hormonal diagnosis of polycystic ovary syndrome. Gynecol Endocrinol [Internet]. 2013;29(6):545–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23506275

17. El Hayek S, Bitar L, Hamdar LH, Mirza FG, Daoud G. Poly Cystic Ovarian Syndrome: An Updated Overview. Front Physiol [Internet]. 2016;7:124. Available from: http://journal.frontiersin.org/article/10.3389/fphys.2016.00124/abstract

18. Legro RS, Bentley-Lewis R, Driscoll D, Wang SC, Dunaif A. Insulin resistance in the sisters of women with polycystic ovary syndrome: Association with hyperandrogenemia rather than menstrual irregularity. J Clin Endocrinol Metab. 2002;87(5):2128–33.

19. Recabarren SE, Smith R, Rios R, Maliqueo M, Echiburú B, Codner E, et al. Metabolic profile in sons of women with polycystic ovary syndrome. J Clin Endocrinol Metab. 2008;93(5):1820–6.

20. Sung YA, Oh JY, Chung H, Lee H. Hyperandrogenemia is implicated in both the metabolic and reproductive morbidities of polycystic ovary syndrome. Fertil Steril. 2014;101(3):840–5.

21. Ehrmann DA, Rosenfield RL, Barnes RB, Brigell DF, Sheikh Z. Detection of functional ovarian hyperandrogenism in women with androgen excess. N Engl J Med [Internet]. 1992;327(3):157–62. Available from: http://www.ncbi.nlm.nih.gov/pubmed/1319000

22. Nelson VL, Qin KN, Rosenfield RL, Wood JR, Penning TM, Legro RS, et al. The biochemical basis for increased Testosterone production in theca cells propagated from patients with polycystic ovary syndrome. J Clin Endocrinol Metab. 2001;86(12):5925–33.

23. Givens JR. Familial polycystic ovarian disease. Endocrinol Metab Clin North Am. 1988;17(4):771–83.

24. Amato P, Simpson JL. The genetics of polycystic ovary syndrome. Best Pract Res Clin Obstet Gynaecol [Internet]. 2004;18(5):707–18. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15380142

25. Strauss JF, McAllister JM, Urbanek M. Persistence Pays Off for PCOS Gene Prospectors. J Clin Endocrinol Metab [Internet]. 2012 Jul 1;97(7):2286–8. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3387390/

26. Vink JM, Sadrzadeh S, Lambalk CB, Boomsma DI. Heritability of polycystic ovary

syndrome in a Dutch twin-family study. J Clin Endocrinol Metab. 2006;91(6):2100–4.

27. Legro RS, Driscoll D, Strauss JF, Fox J, Dunaif A. Evidence for a genetic basis for hyperandrogenemia in polycystic ovary syndrome. Proc Natl Acad Sci U S A [Internet]. 1998;95(25):14956–60. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=24557&tool=pmcentrez&rendertype=abstract

28. Franks S, Webber LJ, Goh M, Valentine A, White DM, Conway GS, et al. Ovarian morphology is a marker of heritable biochemical traits in sisters with polycystic ovaries. J Clin Endocrinol Metab. 2008;93(9):3396–402.

29. Govind A, Obhrai MS, Clayton RN. Polycystic ovaries are inherited as an autosomal dominant trait: Analysis of 29 polycystic ovary syndrome and 10 control families. J Clin Endocrinol Metab. 1999;84(1):38–43.

30. Legro RS, Spielman R, Urbanek M, Driscoll D, Strauss JF, Dunaif A. Phenotype and genotype in polycystic ovary syndrome. Recent Prog Horm Res. 1998;53:217–56.

31. Povel CM, Boer JMA, Feskens EJM. Shared genetic variance between the features of the metabolic syndrome: Heritability studies. Mol Genet Metab. 2011;104(4):666–9.

32. Waalen J. The genetics of human obesity. Vol. 164, Translational Research. 2014. p. 293–301.

33. McAllister JM, Legro RS, Modi BP, Strauss JF. Functional genomics of PCOS: From GWAS to molecular mechanisms. Obstet Gynecol Surv. 2015;70(9).

34. Chen ZJ, Zhao H, He L, Shi Y, Qin Y, Li Z, et al. Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome 2p16.3, 2p21 and 9q33.3. Nat Genet [Internet]. 2011;43(1):55–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21151128

35. Shi Y, Zhao H, Shi Y, Cao Y, Yang D, Li Z, et al. Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. Nat Genet. 2012;44:1020–5.

36. Hayes MG, Urbanek M, Ehrmann DA, Armstrong LL, Lee JY, Sisk R, et al. Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. Nat Commun [Internet]. 2015;6(February 2016):7502. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4557132&tool=pmcentrez&rendertype=abstract

37. Day FR, Hinds DA, Tung JY, Stolk L, Styrkarsdottir U, Saxena R, et al. Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. Nat Commun [Internet]. 2015;6:8464. Available from: http://dx.doi.org/10.1038/ncomms9464\nhttp://www.nature.com/doifinder/10.1038/ncomms9464

38. Goodarzi MO, Jones MR, Li X, Chua AK, Garcia OA, Chen Y-DI, et al. Replication of association of DENND1A and THADA variants with polycystic ovary syndrome in European cohorts. J Med Genet [Internet]. 2012;49(2):90–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3536488&tool=pmcentrez&rendertype=abstract

39. Welt CK, Styrkarsdottir U, Ehrmann DA, Thorleifsson G, Arason G, Gudmundsson JA, et al. Variants in DENND1A are associated with polycystic ovary syndrome in women of European ancestry. J Clin Endocrinol Metab. 2012;97(7).

40. Louwers Y V., Stolk L, Uitterlinden AG, Laven JSE. Cross-Ethnic meta-Analysis of genetic variants for polycystic ovary syndrome. J Clin Endocrinol Metab. 2013;98(12).

41. Cui L, Zhao H, Zhang B, Qu Z, Liu J, Liang X, et al. Genotype-phenotype correlations of PCOS susceptibility SNPs identified by GWAS in a large cohort of Han Chinese women. Hum Reprod. 2013;28(2):538–44.

42. McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS, et al. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. Proc Natl Acad Sci U S A [Internet]. 2014;111(15):E1519-27. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3992676&tool=pmcentrez&rendertype=abstract

43. Kramer MS, McLean FH, Boyd ME, Usher RH. The validity of gestational age estimation by menstrual dating in term, preterm, and postterm gestations. JAMA. 1988;260(22):3306–8.

44. MacDorman MF, Munson ML, Kirmeyer S. Fetal and perinatal mortality, United States, 2004. Natl Vital Stat Rep [Internet]. 2007;56(3):1–19. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17983023

45. Chaudhari BP, Plunkett J, Ratajczak CK, Shen TT, Defranco EA, Muglia LJ. The genetics of birth timing: Insights into a fundamental component of human

development. Clin Genet. 2008;74(6):493–501.

46. Knight AK, Smith AK. Insights into genetic susceptibility in the etiology of spontaneous preterm birth. 2015;283–90.

47. Goldenberg RL, Culhane JF, Iams JD, Romero R. Epidemiology and causes of preterm birth. Vol. 371, The Lancet. 2008. p. 75–84.

48. Parry S, Strauss JF. Premature rupture of the fetal membranes. N Engl J Med [Internet]. 1998;338(10):663–70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9486996

49. Shen TT, DeFranco EA, Stamilio DM, Chang JJ, Muglia LJ. A population-based study of race-specific risk for preterm premature rupture of membranes. Am J Obstet Gynecol. 2008;199(4).

50. Anum EA, Springel EH, Shriver MD, Strauss JF. Genetic contributions to disparities in preterm birth. Vol. 65, Pediatric Research. 2009. p. 1–9.

51. Caughey AB, Robinson JN, Norwitz ER. Contemporary diagnosis and management of preterm premature rupture of membranes. Rev Obstet Gynecol [Internet]. 2008;1(1):11–22. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2492588&tool=pmcentrez&rendertype=abstract

52. Magnus P, Bakketeig LS, Skjaerven R. Correlations of birth weight and gestational age across generations. Ann Hum Biol [Internet]. 1993;20(3):231–8. Available from: http://www.ncbi.nlm.nih.gov/htbin-post/Entrez/query?db=m&form=6&dopt=r&uid=0008489198

53. Winkvist A, Mogren I, Högberg U. Familial patterns in birth characteristics: Impact

on individual and population risks. Int J Epidemiol. 1998;27(2):248–54.

54. York TP, Eaves LJ, Neale MC, Strauss JF. The contribution of genetic and environmental factors to the duration of pregnancy. Am J Obstet Gynecol. 2014;210(5):398–405.

55. Zhang H, Baldwin DA, Bukowski RK, Parry S, Xu Y, Song C, et al. A genome-wide association study of early spontaneous preterm delivery. Genet Epidemiol [Internet]. 2015;39(3):217–26. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25599974\nhttp://onlinelibrary.wiley.com/store/10.1002/gepi.21887/asset/gepi21887.pdf?v=1&t=i8u9hn8v&s=fcf9d22c80396f3b0e155a5afc32ed605bd4980c

56. Wu W, Clark EAS, Manuck TA, Esplin MS, Varner MW, Jorde LB. A Genome-Wide Association Study of spontaneous preterm birth in a European population [version 1 ; referees : 2 approved with reservations] Referee Status : 2015;(0):1–12.

57. Ferrand PE, Fujimoto T, Chennathukuzhi V, Parry S, Macones GA, Sammel M, et al. The CARD15 2936insC mutation and TLR4 896 A>G polymorphism in African Americans and risk of preterm premature rupture of membranes (PPROM). Mol Hum Reprod [Internet]. 2002 Nov 1;8(11):1031–4. Available from: http://molehr.oxfordjournals.org/content/8/11/1031.abstract

58. A.K. R, F. M-B, P.G. VD, J. H, G.A. M, M.A. M, et al. Association of polymorphism within the promoter of the tumor necrosis factor α gene with increased risk of preterm premature rupture of the fetal membranes. Am J Obstet Gynecol [Internet]. 1999;180(5):1297–302. Available from:

http://www.embase.com/search/results?subaction=viewrecord&from=export&id=L29259032\nhttp://dx.doi.org/10.1016/S0002-9378(99)70632-0

59. York TP, Strauss JF, Neale MC, Eaves LJ. Estimating fetal and maternal genetic contributions to premature birth from multiparous pregnancy histories of twins using MCMC and maximum-likelihood approaches. Twin Res Hum Genet. 2009;12(4):333–42.

60. Capece A, Vasieva O, Meher S, Alfirevic Z, Alfirevic A. Pathway analysis of genetic factors associated with spontaneous preterm birth and pre-labor preterm rupture of membranes. PLoS One. 2014;9(9).

61. Ambros V. microRNAs: Tiny regulators with great potential. Vol. 107, Cell. 2001. p. 823–6.

62. Bartel DP. MicroRNAs: Genomics, Biogenesis, Mechanism, and Function. Vol. 116, Cell. 2004. p. 281–97.

63. Sørensen A, Wissing M, Salö S, Englund A, Dalgaard L. MicroRNAs Related to Polycystic Ovary Syndrome (PCOS). Genes (Basel) [Internet]. 2014;5(3):684–708. Available from: http://www.mdpi.com/2073-4425/5/3/684/

64. Imbar T, Eisenberg I. Regulatory role of microRNAs in ovarian function. Vol. 101, Fertility and Sterility. 2014. p. 1524–30.

65. Mayor-Lynn K, Toloubeydokhti T, Cruz AC, Chegini N. Expression Profile of MicroRNAs and mRNAs in Human Placentas From Pregnancies Complicated by Preeclampsia and Preterm Labor. Reprod Sci [Internet]. 2011;18(1):46–56. Available from: http://rsx.sagepub.com/content/18/1/46.full\nhttp://www.ncbi.nlm.nih.gov/pmc/artic

les/PMC3343068/\nhttp://www.ncbi.nlm.nih.gov/pmc/articles/PMC3343068/pdf/10.1177_1933719110374115.pdf

66. Marat AL, Dokainish H, McPherson PS. DENN domain proteins: Regulators of Rab GTPases. Vol. 286, Journal of Biological Chemistry. 2011. p. 13791–800.

67. McAllister JM, Legro RS, Modi BP, Strauss JF. Functional genomics of PCOS: From GWAS to molecular mechanisms. Trends Endocrinol Metab. 2015;26(3).

68. McAllister JM. Functional, long-term human theca and granulosa cell cultures from polycystic ovaries. Endocrine [Internet]. 1995;3(2):143–9. Available from: http://dx.doi.org/10.1007/BF02990066

69. Tee MK, Speek M, Legeza B, Modi B, Teves ME, McAllister JM, et al. Alternative splicing of DENND1A, a PCOS candidate gene, generates variant 2. Mol Cell Endocrinol. 2016;434.

70. McAllister JM, Modi B, Miller BA, Biegler J, Bruggeman R, Legro RS, et al. Overexpression of a DENND1A isoform produces a polycystic ovary syndrome theca phenotype. Proc Natl Acad Sci U S A. 2014;111(15).

71. Eriksen MB, Nielsen MFB, Brusgaard K, Tan Q, Andersen MS, Glintborg D, et al. Genetic Alterations within the DENND1A Gene in Patients with Polycystic Ovary Syndrome (PCOS). PLoS One. 2013;8(9).

72. Azziz R, Dumesic DA, Goodarzi MO. Polycystic ovary syndrome: An ancient disorder? Fertil Steril. 2011;95(5):1544–8.

73. Balen A, Homburg R, Franks S. Defining polycystic ovary syndrome. BMJ [Internet]. 2009;338:a2968. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19218318

74. Gilling-Smith C, Willis DS, Beard RW, Franks S. Hypersecretion of androstenedione by isolated thecal cells from polycystic ovaries. J Clin Endocrinol Metab. 1994;79(4):1158–65.

75. Gilling-Smith C, Story H, Rogers V, Franks S. Evidence for a primary abnormality of thecal cell steroidogenesis in the polycystic ovary syndrome. Clin Endocrinol (Oxf) [Internet]. 1997;47(1):93–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9302378

76. Nelson VL, Legro RS, Strauss JF, McAllister JM. Augmented androgen production is a stable steroidogenic phenotype of propagated theca cells from polycystic ovaries. Mol Endocrinol. 1999;13(6):946–57.

77. Wood JR, Nelson VL, Ho C, Jansen E, Wang CY, Urbanek M, et al. The molecular phenotype of polycystic ovary syndrome (PCOS) theca cells and new candidate PCOS genes defined by microarray analysis. J Biol Chem. 2003;278(29):26380–90.

78. Wood JR, Ho CKM, Nelson-Degrave VL, McAllister JM, Strauss JF. The molecular signature of polycystic ovary syndrome (PCOS) theca cells defined by gene expression profiling. In: Journal of Reproductive Immunology. 2004. p. 51–60.

79. Sang Q, Yao Z, Wang H, Feng R, Wang H, Zhao X, et al. Identification of MicroRNAs in human follicular fluid: Characterization of MicroRNAs that govern steroidogenesis in vitro and are associated with polycystic ovary syndrome in vivo. J Clin Endocrinol Metab. 2013;98(7):3068–79.

80. Roth L, McCallie M, Alvero R, Schoolcraft W, Minjarez D, Katz-Jaffe M. Altered

microRNA and gene expression in the follicular fluid of women with polycystic ovary syndrome. J Assist Reprod Genet. 2014;31:355–62.

81. Schmidt J, Weijdeg??rd B, Mikkelsen AL, Lindenberg S, Nilsson L, Br??nnstr??m M. Differential expression of inflammationrelated genes in the ovarian stroma and granulosa cells of PCOS women. Mol Hum Reprod. 2014;20(1):49–58.

82. Blandino G, Valerio M, Cioce M, Mori F, Casadei L, Pulito C, et al. Metformin elicits anticancer effects through the sequential modulation of DICER and c-MYC. Nat Commun [Internet]. 2012;3(May):865. Available from: http://dx.doi.org/10.1038/ncomms1859

83. Fisher R. Statistical methods for research workers [Internet]. Biological monographs and manuals. 1925. xv, 356 . Available from: http://psychclassics.yorku.ca/Fisher/Methods

84. Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. Vol. 57, Journal of the Royal Statistical Society. 1995. p. 289–300.

85. R Core Team. R Core Team. R A Lang Environ Stat Comput R Found Stat Comput Vienna, Austria [Internet]. 2013;ISBN 3-900051-07-0, URL http://www.R-project.org/. Available from: http://www.mendeley.com/research/r-language-environment-statistical-computing-96/\npapers2://publication/uuid/A1207DAB-22D3-4A04-82FB-D4DD5AD57C28

86. Huang X, Liu C, Hao C, Tang Q, Liu R, Lin S, et al. Identification of altered microRNAs and mRNAs in the cumulus cells of PCOS patients. Reproduction [Internet]. 2016; Available from: http://www.ncbi.nlm.nih.gov/pubmed/27001999

87. Sørensen AE, Wissing ML, Lis A, Englund M, Dalgaard LT. MicroRNA species in follicular fluid associating with polycystic ovary syndrome and related intermediary phenotypes. J Clin Endocrinol Metab [Internet]. 2016;jc20153588. Available from: http://press.endocrine.org/doi/pdf/10.1210/jc.2015-3588\nhttp://www.ncbi.nlm.nih.gov/pubmed/26771704

88. Xu Y, Li Z, Ai F, Chen J, Xing Q, Zhou P, et al. Systematic Evaluation of Genetic Variants for Polycystic Ovary Syndrome in a Chinese Population. PLoS One [Internet]. 2015;10(10):1–10. Available from: http://dx.doi.org/10.1371%2Fjournal.pone.0140695

89. Amsterdam T, Pcos EA. Consensus on women's health aspects of polycystic ovary syndrome (PCOS). Hum Reprod [Internet]. 2012;27(1):14–24. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22147920

90. Jamnongjit M, Gill A, Hammes SR. Epidermal growth factor receptor signaling is required for normal ovarian steroidogenesis and oocyte maturation. Proc Natl Acad Sci [Internet]. 2005;102(45):16257–62. Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0508521102

91. Evaul K, Jamnongjit M, Hammes S. EPIDERMAL GROWTH FACTOR RECEPTOR SIGNALING IS REQUIRED FOR STEROIDOGENESIS IN THE OVARY, TESTES AND ADRENAL GLANDS. Biol Reprod [Internet]. 2007 Jul 1;77(Suppl 1):137. Available from: http://www.biolreprod.org/content/77/Suppl_1/137.2.abstract

92. McCallie B, Schoolcraft WB, Katz-Jaffe MG. Aberration of blastocyst microRNA expression is associated with human infertility. Fertil Steril. 2010;93(7):2374–82.

93. Chen Y-H, Heneidi S, Lee J-M, Layman LC, Stepp DW, Gamboa GM, et al. miRNA-93 Inhibits GLUT4 and Is Overexpressed in Adipose Tissue of Polycystic Ovary Syndrome Patients and Women With Insulin Resistance. Diabetes [Internet]. 2013 Jun 25;62(7):2278 LP-2286. Available from: http://diabetes.diabetesjournals.org/content/62/7/2278.abstract

94. Jiang L, Huang J, Li L, Chen Y, Chen X, Zhao X, et al. MicroRNA-93 promotes ovarian granulosa cells proliferation through targeting CDKN1A in polycystic ovarian syndrome. J Clin Endocrinol Metab. 2015;100(5):E729–38.

95. Sathyapalan T, David R, Gooderham NJ, Atkin SL. Increased expression of circulating miRNA-93 in women with polycystic ovary syndrome may represent a novel, non-invasive biomarker for diagnosis. Sci Rep [Internet]. 2015;5:16890. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-84947800144&partnerID=tZOtx3y1

96. Wilcox AJ, Skjærven R, Lie RT. Familial patterns of preterm delivery: Maternal and fetal contributions. Am J Epidemiol. 2008;167(4):474–9.

97. York TP, Eaves LJ, Lichtenstein P, Neale MC, Svensson A, Latendresse S, et al. Fetal and Maternal Genes' Influence on Gestational Age in a Quantitative Genetic Analysis of 244,000 Swedish Births. Am J Epidemiol [Internet]. 2013 Aug 15;178(4):543–50. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3736752/

98. Myking S, Boyd HA, Myhre R, Feenstra B, Jugessur A, Devold Pay AS, et al. X-Chromosomal Maternal and Fetal SNPs and the Risk of Spontaneous Preterm Delivery in a Danish/Norwegian Genome-Wide Association Study. Yan W, editor.

PLoS One [Internet]. 2013 Apr 16;8(4):e61781. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3628886/

99. Plunkett J, Muglia LJ. Genetic contributions to preterm birth: implications from epidemiological and genetic association studies. Ann Med [Internet]. 2008;40(3):167–95. Available from: http://www.tandfonline.com/doi/full/10.1080/07853890701806181\nhttp://www.ncbi.nlm.nih.gov/pubmed/18382883

100. Dolan SM, Hollegaard M V, Merialdi M, Betran AP, Allen T, Abelow C, et al. Synopsis of Preterm Birth Genetic Association Studies: The Preterm Birth Genetics Knowledge Base (PTBGene). Public Health Genomics [Internet]. 2010;13(7–8):514–23. Available from: http://www.karger.com/DOI/10.1159/000294202

101. Hindorff L a, Sethupathy P, Junkins H a, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A [Internet]. 2009;106(23):9362–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19474294

102. Cirulli ET, Goldstein DB. Uncovering the roles of rare variants in common disease through whole-genome sequencing. Nat Rev Genet [Internet]. 2010 Jun;11(6):415–25. Available from: http://dx.doi.org/10.1038/nrg2779

103. Romero R, Velez DR, Kusanovic JP, Hassan SS, Mazaki-Tovi S, Vaisbuch E, et al. Identification of Fetal and Maternal Single Nucleotide Polymorphisms in Candidate Genes That Predispose to Spontaneous Preterm Labor with Intact

Membranes. Am J Obstet Gynecol [Internet]. 2010 May;202(5):431.e1-431.34. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3604889/

104. Weinberg CR, Shi M. The genetics of preterm birth: Using what we know to design better association studies. Am J Epidemiol. 2009;170(11):1373–81.

105. Strauss JF. Extracellular Matrix Dynamics and Fetal Membrane Rupture. Reprod Sci. 2012;20(2):140–53.

106. Anum EA, Hill LD, Pandya A, Strauss JF. Connective Tissue and Related Disorders and Preterm Birth: Clues to Genes Contributing to Prematurity. Placenta. 2009;30(3):207–15.

107. Lind J, Wallenburg HC. Pregnancy and the Ehlers-Danlos syndrome: a retrospective study in a Dutch population. Acta Obs Gynecol Scand. 2002;81(4):293–300.

108. Wesche WA, Cutlan RT, Khare V, Chesney T, Shanklin D. Restrictive dermopathy: report of a case and review of the literature. J Cutan Pathol [Internet]. 2001;28(4):211–8. Available from: http://dx.doi.org/10.1034/j.1600-0560.2001.028004211.x

109. Krakow D. Osteogenesis Imperfecta and Pregnancy. In: Osteogenesis Imperfecta: A Translational Approach to Brittle Bone Disease. 2013. p. 243–51.

110. Wang H, Parry S, Macones G, Sammel MD, Kuivaniemi H, Tromp G, et al. A functional SNP in the promoter of the SERPINH1 gene increases risk of preterm premature rupture of membranes in African Americans. Proc Natl Acad Sci U S A [Internet]. 2006;103(36):13463–7. Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-

33748631588&partnerID=tZOtx3y1

111. Chakraborty R, Weiss KM. Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. Proc Natl Acad Sci [Internet]. 1988;85(23):9119–23. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=282675&tool=pmcentrez&rendertype=abstract

112. International T, Consortium H. The International HapMap Project. Nature. 2003;426(6968):789–96.

113. Hanis CL, Chakraborty R, Ferrell RE, Schull WJ. Individual admixture estimates: disease associations and individual risk of diabetes and gallbladder disease among Mexican-Americans in Starr County, Texas. Am J Phys Anthropol [Internet]. 1986;70(4):433–41. Available from: http://www.ncbi.nlm.nih.gov/pubmed/3766713

114. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010;26(5):589–95.

115. 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(9):1297–303.

116. DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011;43(5):491–8.

117. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res. 2001;29(1):308–11.

118. Landrum MJ, Lee JM, Benson M, Brown G, Chao C, Chitipiralla S, et al. ClinVar: Public archive of interpretations of clinically relevant variants. Nucleic Acids Res. 2016;44(D1):D862–8.

119. The 1000 Genomes Project Consortium. A global reference for human genetic variation [Internet]. Vol. 526, Nature. 2015. p. 68–74. Available from: http://www.nature.com/nature/journal/v526/n7571/full/nature15393.html%5Cnhttp://www.nature.com/nature/journal/v526/n7571/pdf/nature15393.pdf

120. Gabriel S, Ziaugra L, Tabbaa D. SNP genotyping using the sequenom massARRAY iPLEX Platform. Current Protocols in Human Genetics. 2009.

121. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vol. 0, R Foundation for Statistical Computing Vienna Austria. 2015. p. {ISBN} 3-900051-07-0. Available from: http://www.r-project.org/

122. Wu MC, Lee S, Cai T, Li Y, Boehnke M, Lin X. Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet. 2011;89(1):82–93.

123. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet. 2012;91(2):224–37.

124. Ionita-Laza I, Lee S, Makarov V, Buxbaum JD, Lin X. Sequence kernel association tests for the combined effect of rare and common variants. Am J Hum Genet. 2013;92(6):841–53.

125. Hampson V, Liu D, Billett E, Kirk S. Amniotic membrane collagen content and

type distribution in women with preterm premature rupture of the membranes in pregnancy. Br J Obstet Gynaecol. 1997;104(9):1087–91.

126. Kanayama N, Terao T, Kawashima Y, Horiuchi K, Fujimoto D. Collagen types in normal and prematurely ruptured amniotic membranes. Am J Obstet Gynecol. 1985;153(8):899–903.

127. Skinner S, Campos G, Liggins G. Collagen content of human amniotic membranes: effect of gestation length and premature rupture. Obstet Gynecol. 1981;57(4):487–9.

128. Davis EC, Broekelmann TJ, Ozawa Y, Mecham RP. Identification of tropoelastin as a ligand for the 65-kD FK506-binding protein, FKBP65, in the secretory pathway. J Cell Biol. 1998;140(2):295–303.

129. Schwarze U, Cundy T, Pyott SM, Christiansen HE, Hegde MR, Bank RA, et al. Mutations in FKBp10, which result in bruck syndrome and recessive forms of osteogenesis imperfecta, inhibit the hydroxylation of telopeptide lysines in bone collagen. Hum Mol Genet. 2013;22(1):1–17.

130. Moore RM, Mansour JM, Redline RW, Mercer BM, Moore JJ. The Physiology of Fetal Membrane Rupture: Insight Gained from the Determination of Physical Properties. Placenta. 2006;27(11–12):1037–51.

131. Myking S, Myhre R, Gjessing HK, Morken N-H, Sengpiel V, Williams SM, et al. Candidate gene analysis of spontaneous preterm delivery: New insights from re-analysis of a case-control study using case-parent triads and control-mother dyads. BMC Med Genet [Internet]. 2011;12(1):174. Available from: http://www.biomedcentral.com/1471-2350/12/174

132. Wenstrup RJ, Florer JB, Brunskill EW, Bell SM, Chervoneva I, Birk DE. Type V collagen controls the initiation of collagen fibril assembly. J Biol Chem. 2004;279(51):53331–7.

133. JH B. Osteogenesis imperfecta. Radiol Technol [Internet]. 2008;79(6):535–551 17p. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=rzh&AN=105813760&site=ehost-live

134. Taksande A, Vilhekar K, Khangare S. Osteogenesis Imperfecta Type II with Congenital Heart Disease. Iran J Pediatr [Internet]. 2008;18(2):175–8. Available from: papers3://publication/uuid/64E88C5C-1409-441C-92FC-6296D10D06F4

135. Malfait F, De Paepe A. Molecular genetics in classic Ehlers-Danlos syndrome. Vol. 139 C, American Journal of Medical Genetics - Seminars in Medical Genetics. 2005. p. 17–23.

136. Barat-Houari M, Sarrabay G, Gatinois V, Fabre A, Dumont B, Genevieve D, et al. Mutation Update for COL2A1 Gene Variants Associated with Type II Collagenopathies. Hum Mutat. 2016;37(1):7–15.

137. Díaz-Prado S, Rendal-Vázquez ME, Muiños-López E, Hermida-Gómez T, Rodríguez-Cabarcos M, Fuentes-Boquete I, et al. Potential use of the human amniotic membrane as a scaffold in human articular cartilage repair. Cell Tissue Bank [Internet]. 2010;11(2):183–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20386989

138. Díaz-Prado S, Muiños-López E, Hermida-Gómez T, Cicione C, Rendal-Vázquez ME, Fuentes-Boquete I, et al. Human amniotic membrane as an alternative

source of stem cells for regenerative medicine. Vol. 81, Differentiation. 2011. p. 162–71.

139. Parra EJ, Marcini A, Akey J, Martinson J, Batzer MA, Cooper R, et al. Estimating African American admixture proportions by use of population-specific alleles. Am J Hum Genet [Internet]. 1998;63(6):1839–51. Available from: http://www.sciencedirect.com/science/article/pii/S0002929707616280

140. Bryc K, Durand EY, Macpherson JM, Reich D, Mountain JL. The genetic ancestry of African, Latino, and European Americans across the United States. bioRxiv [Internet]. 2014;9340. Available from: http://biorxiv.org/content/early/2014/09/26/009340.abstract\nhttp://biorxiv.org/lookup/doi/10.1101/009340

141. Ferrand PE, Parry S, Sammel M, Macones GA, Kuivaniemi H, Romero R, et al. A polymorphism in the matrix metalloproteinase-9 promoter is associated with increased risk of preterm premature rupture of membranes in African Americans. Mol Hum Reprod [Internet]. 2002;8(5):494–501. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?cmd=Retrieve&db=PubMed&dopt=Citation&list_uids=11994547\nhttp://molehr.oxfordjournals.org/content/8/5/494.full.pdf

142. Fujimoto T, Parry S, Urbanek M, Sammel M, Macones G, Kuivaniemi H, et al. A single nucleotide polymorphism in the matrix metalloproteinase-1 (MMP- 1) promoter influences amnion cell MMP-1 expression and risk for preterm premature rupture of the fetal membranes. J Biol Chem [Internet]. 2002;277(8):6296–302. Available from: http://www.ncbi.nlm.nih.gov/htbin-

post/Entrez/query?db=m&form=6&dopt=r&uid=11741975\nhttp://www.jbc.org/cgi/content/full/277/8/6296\nhttp://www.jbc.org/cgi/content/abstract/277/8/6296

143. Wang H, Parry S, Macones G, Sammel MD, Ferrand PE, Kuivaniemi H, et al. Functionally significant SNP MMP8 promoter haplotypes and preterm premature rupture of membranes (PPROM). Hum Mol Genet. 2004;13(21):2659–69.

144. Bromley B, Shipp TD, Benacerraf BR. Amnion-Chorion Separation After 17 Weeks ' Gestation. 7844(99):1024–6.

145. Abboud P, Mansour G, Zejli A, Gondry J. Chorioamniotic separation after 14 weeks' gestation associated with trisomy 21 [1]. Ultrasound Obstet Gynecol. 2003;22(1):94–5.

146. Campbell A. Letters To the Editor Letters To the Editor. Harv Bus Rev. 1992;15(October):1106–7.

147. Amiel A, Fejgin MD, Liberman M, Sharon Y, Kidron D, Biron-Shental T. Senescence in amniocytes and placentas from trisomy 21 pregnancies. J Matern Fetal Neonatal Med [Internet]. 2013;7058(11):1–4. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23339291

148. Pidoux G, Guibourdenche J, Frendo JL, Gerbaud P, Conti M, Luton D, et al. Impact of trisomy 21 on human trophoblast behaviour and hormonal function. Placenta. 2004;25(SUPPL. A):79–84.

149. Polettini J, Dutta EH, Behnia F, Saade GR, Torloni MR, Menon R. Aging of intrauterine tissues in spontaneous preterm birth and preterm premature rupture of the membranes: A systematic review of the literature. Placenta [Internet]. 2015;36(9):969–73. Available from:

http://dx.doi.org/10.1016/j.placenta.2015.05.003

150. Eckmann-Scholz C, Bens S, Kolarova J, Schneppenheim S, Caliebe A, Heidemann S, et al. DNA-methylation profiling of fetal tissues reveals marked epigenetic differences between Chorionic and Amniotic samples. PLoS One. 2012;7(6).

151. Elton TS, Sansom SE, Martin MM. Trisomy-21 gene dosage over-expression of miRNAs results in the haploinsufficiency of specific target proteins. RNA Biol. 2010;7(5):540–7.

152. Lim JH, Kim DJ, Lee DE, Han JY, Chung JH, Ahn HK, et al. Genome-wide microRNA expression profiling in placentas of fetuses with Down syndrome. Placenta. 2015;36(3):322–8.

153. Nhan-Chang CL, Romero R, Tarca AL, Mittal P, Kusanovic JP, Erez O, et al. Characterization of the transcriptome of chorioamniotic membranes at the site of rupture in spontaneous labor at term. Am J Obstet Gynecol. 2010;202(5):1–31.

154. Verma RS, Babu A. Tissue culture techniques and chromosome preparation. Hum Chromosom Princ Tech McGraw-Hill, Inc, Heal Prof Div New York. 1995;10–3.

155. DE R, BH C. Human Cytogenetics: a practical approach. In: DE R, BH C, editors. Washington, DC: IRL Press; 1986. p. 35–7.

156. Hedeker D, Gibbons RD. Longitudinal data analysis. Hoboken. 2006;(February):337.

157. Kutner MH, Nachtsheim CJJ, Neter J, Li W. Applied Linear Statistical Models [Internet]. Vol. Fifth, Journal Of The Royal Statistical Society Series A General.

2004. 1408 p. Available from: http://www.jhs14.business.msstate.edu/bqa9333/tips/15.70_TukeyMC_NeterKutner.pdf

158. Melamed N, Yogev Y, Glezerman M. Fetal gender and pregnancy outcome. J Matern Neonatal Med [Internet]. 2010;23(4):338–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20222831

159. Weng Y-H, Yang C-Y, Chiu Y-W. Neonatal outcomes in relation to sex differences: a national cohort survey in Taiwan. Biol Sex Differ [Internet]. 2015;6:30. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4675056&tool=pmcentrez&rendertype=abstract

160. Xu Y, Li W, Liu X, Ma H, Tu Z, Dai Y. Analysis of microRNA expression profile by small RNA sequencing in Down syndrome fetuses. Int J Mol Med. 2013;32(5):1115–25.

161. Erturk B, Karaca E, Aykut A, Durmaz B, Guler A, Buke B, et al. Prenatal Evaluation of MicroRNA Expressions in Pregnancies with Down Syndrome. 2016;2016.

162. Gu Y, Sun J, Groome LJ, Wang Y. Differential miRNA expression profiles between the first and third trimester human placentas. Am J Physiol Endocrinol Metab [Internet]. 2013;304:E836-43. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3625781&tool=pmcentrez&rendertype=abstract

163. Montenegro D, Romero R, Kim SS, Tarca AL, SDraghici, Kusanovic JP, et al.

Expression patterns of microRNAs in the chorioamniotic membranes: A role for microRNAs in human pregnancy and parturition. J Pathol. 2009;217(1):113–21.

164. Montenegro D, Romero R, Pineles BL, Tarca AL, Kim YM, Draghici S, et al. Differential expression of microRNAs with progression of gestation and inflammation in the human chorioamniotic membranes. Am J Obstet Gynecol. 2007;197(3).

165. Piccinini AM, Midwood KS. Illustrating the interplay between the extracellular matrix and microRNAs. Int J Exp Pathol. 2014;95(3):158–80.

166. Svobodová I, Korabečná M, Calda P, Břešťák M, Pazourková E, Pospíšilová Š, et al. Differentially expressed miRNAs in trisomy-21 placentas. Prenat Diagn [Internet]. 2016; Available from: http://doi.wiley.com/10.1002/pd.4861

167. Lim JH, Kim SY, Han JY, Kim MY, Park SY, Ryu HM. Comprehensive investigation of DNA methylation and gene expression in trisomy 21 placenta. Placenta [Internet]. 2016;42:17–24. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0143400416300480

168. Yu S, Zhang C, Dong F, Zhang Y. miR-99a suppresses the metastasis of human non-small cell lung cancer cells by targeting AKT1 signaling pathway. J Cell Biochem. 2015;116(2):268–76.

169. Wu D, Zhou Y, Pan H, Zhou J, Fan Y, Qu P. MicroRNA-99a inhibiting cell proliferation, migration and invasion by targeting fibroblast growth factor receptor 3 in bladder cancer. Oncol Lett. 2014;7(4):1219–24.

170. Emmrich S, Rasche M, Schöning J, Reimer C, Keihani S, Maroz A, et al. [Supplemental]miR-99a/100~125b tricistrons regulate hematopoietic stem and

progenitor cell homeostasis by shifting the balance between TGFβ and Wnt signaling. Genes Dev. 2014;1–15.

171. Sun Y-M, Lin K-Y, Chen Y-Q. Diverse functions of miR-125 family in different cell contexts. J Hematol Oncol [Internet]. 2013;6(1):6. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3566921&tool=pmcentrez&rendertype=abstract

172. Chan HW, Lappas M, Yee SWY, Vaswani K, Mitchell MD, Rice GE. The expression of the let-7 miRNAs and Lin28 signalling pathway in human term gestational tissues. Placenta. 2013;34(5):443–8.

173. Malak TM, Bell SC. Structural characteristics of term human fetal membranes: a novel zone of extreme morphological alteration within the rupture site. Br J Obstet Gynaecol [Internet]. 1994;101(5):375–86. Available from: http://www.ncbi.nlm.nih.gov/pubmed/8018607

174. McLaren J, Malak TM, Bell SC. Structural characteristics of term human fetal membranes prior to labour: Identification of an area of altered morphology overlying the cervix. Hum Reprod. 1999;14(1):237–41.

175. McParland PC, Taylor DJ, Bell SC. Mapping of zones of altered morphology and chorionic connective tissue cellular phenotype in human fetal membranes (amniochorion and decidua) overlying the lower uterine pole and cervix before labor at term. Am J Obstet Gynecol. 2003;189(5):1481–8.

176. Fortunato SJ, Menon R, Bryant C, Lombardi SJ. Programmed cell death (apoptosis) as a possible pathway to metalloproteinase activation and fetal membrane degradation in premature rupture of membranes. In: American Journal

of Obstetrics and Gynecology. 2000. p. 1468–76.

177. Menon R, Fortunato SJ. The role of matrix degrading enzymes and apoptosis in rupture of membranes. Vol. 11, Journal of the Society for Gynecologic Investigation. 2004. p. 427–37.

178. Elia E, Sander V, Luchetti CG, Solano ME, Di Girolamo G, Gonzalez C, et al. The mechanisms involved in the action of metformin in regulating ovarian function in hyperandrogenized mice. Mol Hum Reprod. 2006;12(8):475–81.

179. Padmanabhan V, Veiga-Lopez A. Animal models of the polycystic ovary syndrome phenotype. Vol. 78, Steroids. 2013. p. 734–40.

180. van Weerden WM, Bierings HG, Van Steenbrugge GJ, De Jong FH, Schröder FH. Adrenal glands of mouse and rat do not synthesize androgens. Life Sci. 1992;50(12):857–61.

181. Abrahams VM, Potter JA, Bhat G, Peltier MR, Saade G, Menon R. Bacterial modulation of human fetal membrane Toll-like receptor expression. Am J Reprod Immunol [Internet]. 2013;69(1):33–40. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3535068&tool=pmcentrez&rendertype=abstract

182. Abu-Maziad A, Schaa K, Bell EF, Dagle JM, Cooper M, Marazita ML, et al. Role of polymorphic variants as genetic modulators of infection in neonatal sepsis. Pediatr Res. 2010;68(4):323–9.

183. Tambor V, Kacerovsky M, Lenco J, Bhat G, Menon R. Proteomics and bioinformatics analysis reveal underlying pathways of infection associated histologic chorioamnionitis in pPROM. Placenta [Internet]. 2013;34(2):155–61.

Available from: http://dx.doi.org/10.1016/j.placenta.2012.11.028

184. Candace C, Evans WS, Pal L. Pharmacogenomics The role of variants regulating metformin. 2016;2:1–9.

185. Essah PA, Apridonidze T, Iuorno MJ, Nestler JE. Effects of short-term and long-term metformin treatment on menstrual cyclicity in women with polycystic ovary syndrome. Fertil Steril. 2006;86(1):230–2.

186. Diamanti-Kandarakis E, Kouli C, Tsianateli T, Bergiele a. Therapeutic effects of metformin on insulin resistance and hyperandrogenism in polycystic ovary syndrome. Eur J Endocrinol [Internet]. 1998;138(3):269–74. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9539300

187. Ehrmann DA, Cavaghan MK, Imperial J, Sturis J, Rosenfield RL, Polonsky KS. Effects of metformin on insulin secretion, insulin action, and ovarian steroidogenesis in women with polycystic ovary syndrome. J Clin Endocrinol Metab. 1997;82(2):524–30.

188. Pau CT, Keefe C, Duran J, Welt CK. Metformin improves glucose effectiveness, not insulin sensitivity: Predicting treatment response in women with polycystic ovary syndrome in an open-label, interventional study. J Clin Endocrinol Metab. 2014;99(5):1870–8.

189. Acbay O, Gundogdu S. Can metformin reduce insulin resistance in polycystic ovary syndrome? Fertil Sterility. 1996;65(5):949.

190. Becker ML, Visser LE, van Schaik RHN, Hofman A, Uitterlinden AG, Stricker BHC. Genetic variation in the organic cation transporter 1 is associated with metformin response in patients with diabetes mellitus. Pharmacogenomics J.

2009;9:242–7.

191. Kang H-J, Song I-S, Shin HJ, Kim W-Y, Lee C-H, Shim J-C, et al. Identification and functional characterization of genetic variants of human organic cation transporters in a Korean population. Drug Metab Dispos. 2007;35(4):667–75.

192. Tkáč I, Klimčáková L, Javorský M, Fabianová M, Schroner Z, Hermanová H, et al. Pharmacogenomic association between a variant in SLC47A1 gene and therapeutic response to metformin in type 2 diabetes. Diabetes, Obes Metab. 2013;15(2):189–91.

193. Wang Z-J, Yin OQP, Tomlinson B, Chow MSS. OCT2 polymorphisms and in-vivo renal functional consequence: studies with metformin and cimetidine. Pharmacogenet Genomics. 2008;18(7):637–45.

194. Zhou K, Donnelly LA, Kimber CH, Donnan PT, Doney ASF, Leese G, et al. Reduced-function SLC22A1 polymorphisms encoding organic cation transporter 1 and glycemic response to metformin: A GoDARTS study. Diabetes. 2009;58(6):1434–9.

Appendix

Table of Contents

Table A.3.1. Specific theca cell cultures used for miRNA sequencing and qPCR assays in Chapter 3 (miRNA and gene qPCRs). C and F indicate control (untreated) and forskolin treated samples

Sample ID	Category	Included in MiRNA Sequencing	Included in QPCR assays
MC02_C/F	Normal	Yes	No
MC06_C/F	Normal	Yes	Yes
MC31_C/F	Normal	Yes	Yes
MC38_C/F	Normal	No	Yes
MC40_C/F	Normal	Yes	Yes
MC50_C/F	Normal	Yes	Yes
MC10_C/F	PCOS	Yes	Yes
MC16_C/F	PCOS	Yes	Yes
MC27_C/F	PCOS	Yes	Yes
MC190_C/F	PCOS	Yes	Yes

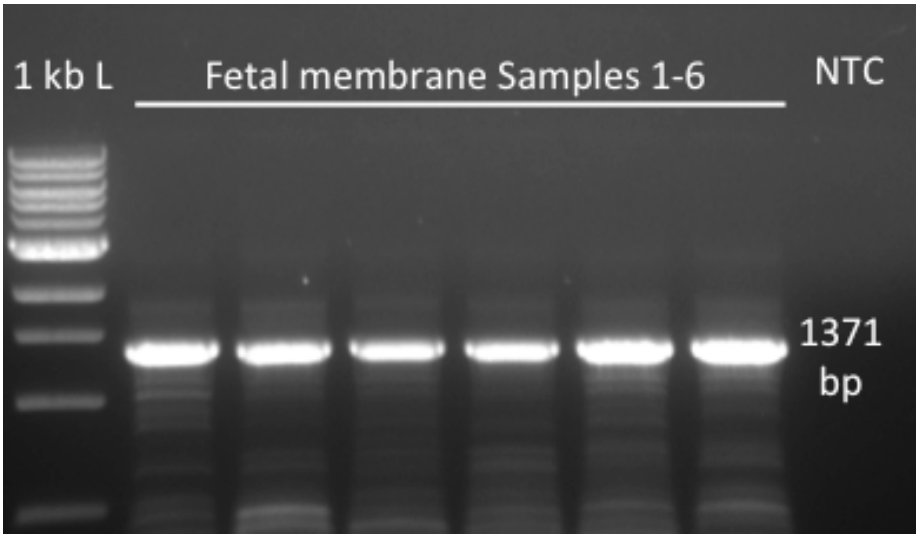


Figure. A.4.1. COL2A1 mRNA expression in amnion samples from normal term pregnancies.

RT-PCR gel picture for *COL2A1* (1371 bp) using 1 μg RNA from amnion tissue samples obtained from normal term pregnancies (lanes 1-6) showing mRNA expression of the *COL2A1* gene in the amnion.

(COL2A1 Primers: 5-CAAGGCCCCAGAGGTGACA-3 and

5-AGGCGGACATGTCGATGCCA-3)

Table A.4.1.A. Comparison of neonatal ancestry between cases and controls in the initial WES:

Neonatal case (n = 49) and control (n = 20) samples were compared for European and West-African ancestry proportions. Values represent mean genetic ancestry estimates generated using two-way model of admixture following maximum likelihood method with SD in parentheses.

Ancestry	Cases Mean (SD)	Control Mean (SD)	p-value
European	0.295 (0.070)	0.314 (0.107)	0.463
West-African	0.705 (0.070)	0.685 (0.107)	0.463

Table A.4.1.B Comparison of neonatal ancestry between cases and controls in the follow-up genotyping study.

Neonatal case (n = 188) and control (n = 175) samples were compared for European and West-African ancestry proportions. Values represent mean genetic ancestry estimates generated using two-way model of admixture following maximum likelihood method with SD in parentheses.

Ancestry	Cases Mean (SD)	Control Mean (SD)	p-value
European	0.1823 (0.083)	0.1658 (0.119)	0.134
West-African	0.8176 (0.083)	0.8341 (0.119)	0.134

Table A.4.2. Missense variants not selected.

Missense variants identified by WES in the genes of interest but not selected for analysis in this study are listed along with their SIFT and Polyphen2 predictions and the observed putative risk allele frequencies (RAF) for cases (n = 49) and controls (n = 20). Novel variants were submitted to dbSNP and ss ids are provided. In cases of multiple SIFT and PolyPhen2 predictions for variants that were annotated to multiple isoforms, only distinct predictions are listed (i.e. does not reflect the total number of isoforms that were actually annotated). (SIFT predictions: T = Tolerated, D = Damaging; Polyphen2 predictions: B = Benign, P = Possibly Damaging, D = Probably Damaging).

Gene	SNP ID	Location	Allele Change	SIFT prediction	PolyPhen 2_HDIV prediction	RAF Cases	RAF Controls
COL1A1	rs142570406	chr17: 48264244	G > C	T	B	0.01	0
COL1A1	rs116794104	chr17: 48272827	C > T	T	B	0.01	0.025
COL1A1	rs1800215	chr17: 48265495	T > C	T	B	0.93	1
COL1A1	rs145446512	chr17: 48266851	C > T	T	B	0	0.025
COL1A2	ss2019492554	chr7: 94029523	C > T	T	P,P	0	0.025
COL1A2	rs42524	chr7:	C > G	T	B	0.91	0.85

		94043239					
COL1A2	rs150124840	chr7: 94057691	C > T	T	P	0.01	0
COL1A2	rs35820023	chr7: 94056353	G > A	T	P	0	0.025
COL2A1	rs140368756	chr12: 48372444	G > A	T	B,B	0	0.025
COL2A1	rs3803183	chr12: 48398080	T > A	T	B,B	0.56	0.575
COL2A1	rs34392760	chr12: 48391657	T > A	T	B,B	0.01	0.025
COL2A1	rs12721427	chr12: 48368541	C > T	T	B,B	0.03	0
COL2A1	rs2070739	chr12: 48367976	C > T	T	D,D	0.05	0.075
COL2A1	rs201823490	chr12: 48380213	G > A	T	B,B	0.03	0
COL3A1	rs201220788	chr2: 189870163	G > A	T	D	0	0.025
COL3A1	rs373838193	chr2: 189862102	C > T	T	D	1	0
COL3A1	rs1800255	chr2: 189864080	G > A	T	B	0.1	0.1

COL3A1	rs1801183	chr2: 189863424	C > T	T	P	0.01	0
COL3A1	rs41263775	chr2: 189864044	C > G	T	B	0.01	0
COL3A1	rs1516446	chr2: 189875421	T > G	T	B	1	1
COL5A1	rs147008954	chr9: 137593122	C > G	D	P	0.01	0
COL5A1	ss2019492556	chr9: 137698068	G > A	T	D	0	0.025
COL5A1	rs61729495	chr9: 137734039	G > A	T	B	0.01	0
COL5A1	rs61735045	chr9: 137642654	G > A	T	D	0.01	0
COL5A1	ss2019492555	chr9: 137620639	A > T	T	B	0	0.025
COL5A2	rs35852101	chr2: 189931144	A > G	N/A	B,P	0.01	0
COL5A2	rs6434313	chr2: 189916110	C > G	N/A	B	0.01	0.025
CRTAP	rs145623565	chr3: 33161975	A > C	T,T	P	0.01	0
CRTAP	ss20194925	chr3:	G > C	T	D	0.01	0

	50	33155942					
ELN	ss2019492553	chr7: 73474329	G > T	T	B	0.01	0
ELN	rs2071307	chr7: 73470714	G > A	T	B	0.21	0.25
ELN	rs61734581	chr7: 73459575	G > T	T	D	0.02	0.05
ELN	rs144026807	chr7: 73474485	C > T	T	B	0.01	0
ELN	rs137987089	chr7: 73474484	G > T	T	P	0.01	0
ELN	rs140425210	chr7: 73477524	G > A	T	D	0.01	0.025
ELN	ss2019492552	chr7: 73457326	G > A	T	D,P	0.01	0
ELN	rs145612009	chr7: 73462008	G > T	T	D	0.01	0
ELN	rs34945509	chr7: 73472000	A > G	T	P	0	0.05
ELN	rs144341345	chr7: 73470729	C > G	T	P	0.03	0.025
ADAMTS2	rs35445112	chr5: 178555097	C > T	T	B	0	0.05

ADAMTS 2	rs398829	chr5: 178634672	C > T	T	B	0.43	0.525
ADAMTS 2	rs59567206	chr5: 178634704	T > C	T	B	0.08	0.075
ADAMTS 2	rs11750821	chr5: 178634683	C > T	T	B	0.01	0.075
ADAMTS 2	rs143764421	chr5: 178634657	C > T	T	B	0	0.025
ADAMTS 2	rs35372714	chr5: 178563002	C > T	T	B	0	0.025
ADAMTS 2	rs1054480	chr5: 178540975	G > A	T	P	0.07	0.1
ADAMTS 2	rs79330641	chr5: 178549719	G > A	T	B	0.02	0.05
ADAMTS 2	ss20194925 51	chr5: 178771124	C > T	T	D,P	0.01	0
SERPIN H1	ss20194925 57	chr11: 75277973	G > C	T	P	0	0.025
SERPIN H1	rs141721173	chr11: 75277974	C > A	T	B	0	0.025
SERPIN H1	rs138784081	chr11: 75277959	A > G	D	B,P	0.01	0
FKBP10	rs34764749	chr17:	A > G	T	B	0.12	0.1

		39974642					
FKBP10	rs76022961	chr17: 39978052	G > A	T	P,B	0.02	0.05
FKBP10	rs201944190	chr17: 39978604	G > A	D	B	0.01	0
FKBP10	rs145424241	chr17:39969495	A > G	T	D	0.01	0

Table A.5.1. Fetal gender comparison for hsa-let-7c.

Comparisons of $\log_2$ expression of hsa-let-7c at the 6 different sampling sites with respect to male vs. female fetal gender are shown. (P < 0.0071 was considered significant (bold)).

Location/Tissue	mean Female	mean Male	p-value
MRA	1.490	0.622	0.12032
MRC	0.857	-0.310	**0.00095**
PPA	1.443	0.983	0.57719
PPC	0.614	0.176	0.36579
RZA	1.178	0.373	0.15320
RZC	-0.273	-1.468	0.09397

Table A.5.2. Fetal gender comparison for hsa-miR-125b.

Comparisons of $\log_2$ expression of hsa-let-7c at the 6 different sampling sites with respect to male vs. female fetal gender are shown. (P < 0.0071 was considered significant (bold)).

Location/Tissue	mean Female	mean Male	p-value
MRA	-6.212	-6.202	0.991
MRC	-6.137	-6.624	0.483
PPA	-5.401	-5.667	0.777
PPC	-6.408	-6.184	0.624
RZA	-6.583	-5.859	0.502
RZC	-7.179	-7.056	0.870

Table A.5.3. Fetal gender comparison for hsa-miR-99a.

Comparisons of log_2 expression of hsa-let-7c at the 6 different sampling sites with respect to male vs. female fetal gender are shown. ($P < 0.0071$ was considered significant (bold)).

Location/Tissue	mean Female	mean Male	p-value
MRA	1.689	1.323	0.579
MRC	0.351	-0.015	0.291
PPA	1.712	1.514	0.706
PPC	0.165	0.154	0.970
RZA	1.221	1.369	0.824
RZC	-0.397	-0.483	0.792

Table A.6.1. NFKBIE variants identified in WES and unique to cases.

SNP	Chromosome 6 Location	Type	Reference Allele	Alternate Allele	Alternate allele frequency
rs10081001	44227330	Intron	G	A	0.09
rs12211448	44227751	Intron	C	A	0.04
rs2233429	44233488	missense	T	A	0.01
rs2233437	44228162	Intron	G	A	0.17
rs28362860	44227224	Intron	C	T	0.04
rs369182955	44232670	Intron	G	A	0.01
rs372668192	44232707	Intron	C	T	0.01
rs79985077	44230663	Intron	C	T	0.01

Vita

Bhavi Pinesh Modi was born on January 6, 1989 in Mumbai, India. She graduated from the University of Mumbai with a Bachelor's (2010) and a Master's (2012) degree in Biotechnology. She gained experience in reproductive biology while working on her Master's dissertation thesis entitled "Receptor for Advanced Glycation End Products (RAGE): Effect of AGEs, Glucose and Hormone Stimulation in Endometrial Cells" at the National Institute for Research in Reproductive Health. She then moved to the United States to pursue a Ph.D. in Human Genetics and continued her passion for reproductive research in Dr. Jerome F. Strauss's laboratory. Under the guidance of Dr. Strauss, she completed her Ph.D. dissertation entitled "Genetic and epigenetic mechanisms of complex reproductive disorders." During her graduate training, she became well versed in several genomic and bioinformatics analysis techniques, miRNA regulation and application of these techniques towards understanding and identification of risk factors contributing to disease susceptibility. She received the C. C. Clayton Award from Virginia Commonwealth University in March, 2016 and was inducted into the Phi Kappa Phi Honor Society in November, 2016. She has presented her research at international conferences for the Society of Reproductive Investigation in 2015 and 2016 and has published several peer-reviewed journal articles as well as a book chapter.