Exploiting Molecular Genetic Diagnoses of Polycystic Ovary Syndrome to Achieve Better Patient Outcome

Silvia Sini-Laulu Borgen

BSc Forensic Biology and Toxicology (Major)
&
BSc Genetics and Molecular Biology (Major)

School of Veterinary and Life Sciences
Murdoch
August 2018

Supervisors:  Associate Professor Scott G. Wilson
              Clinical Professor Bronwyn Stuckey

Academic Supervisor:  Dr Garth Maker

This thesis was submitted in partial fulfilment of the requirements of the award of Honours degree of Bachelor of Science at Murdoch University, Western Australia.
Table of Contents

Declaration .................................................................................................................................................. I
Acknowledgements ......................................................................................................................................... II
Index of Figures ............................................................................................................................................... III
Index of Tables ................................................................................................................................................ IV
Abstract ...................................................................................................................................................... V-VI

Chapter 1: INTRODUCTION

1.1 General introduction ............................................................................................................................... 2
  1.1.1 History of Polycystic Ovary Syndrome ............................................................................................ 2

1.2 Diagnosis ................................................................................................................................................. 3

1.3 Clinical Signs and Symptoms .................................................................................................................. 3
  1.3.1 Hyperandrogenism ........................................................................................................................... 3
    1.3.1.1 Clinical Hyperandrogenism ........................................................................................................ 4
    1.3.1.2 Biochemical Hyperandrogenism .................................................................................................. 4
  1.3.2 Polycystic Ovary Morphology .......................................................................................................... 4
  1.3.3 Ovulatory Dysfunction .................................................................................................................... 5

1.4 Secondary Increased Health Risks ......................................................................................................... 6
  1.4.1 Weight Gain and Obesity .................................................................................................................. 6
  1.4.2 Insulin Resistance ............................................................................................................................ 7
  1.4.3 Metabolic Syndrome ......................................................................................................................... 7
  1.4.4 Gestational Diabetes ......................................................................................................................... 8
  1.4.5 Impaired Glucose Tolerance and Type II Diabetes ........................................................................ 8
  1.4.6 Dyslipidaemia ................................................................................................................................. 8
  1.4.7 Hypertension ..................................................................................................................................... 9
  1.4.8 Cardiovascular Disease .................................................................................................................. 9
  1.4.9 Sleep Disorders .............................................................................................................................. 10
  1.4.10 Mental Health Disorders ............................................................................................................. 10
1.5 Other Potential Risk Factors

1.5.1 Lifestyle Factors

1.5.2 Environmental

1.5.2.1 Exposure to Chemicals

1.5.2.2 Foetal Exposure to Androgens and Other Intrauterine Events

1.5.3 Genetic Factors

1.6 Genetics of PCOS

1.7 GDF9 as a Susceptibility Gene for PCOS

1.7.1 Functions and Characteristics of GDF9

1.7.2 Why GDF9 was Chosen for Functional Study

1.8 Statement of Hypothesis and Aims

Chapter 2: MATERIALS

2.1 General Reagents and Biochemicals

2.2 Instruments and Equipment

2.3 Commercial Kits

2.4 Bacterial Cells

2.5 Cell Lines - Human

2.6 Primers

2.7 Plasmid Constructs

2.8 GDF9 Plasmid Template DNA

2.9 Antibodies

Chapter 3: GENERAL METHODS

3.1 Bacterial Techniques

3.1.1 Bacterial Cultures

3.1.1.1 Agar Plates

3.1.2 Transformation

3.1.2.1 Heat Shock Transformation
3.1.2.1.1 Transformation of Mutagenesis Products......27
3.1.2.1.2 Transformation of Subcloned Constructs.......28
3.1.3 Preparation of E. coli XL2 Competent Cells.................................28

3.2 DNA Techniques..........................................................................................28
3.2.1 DNA Preparation..........................................................................................28
  3.2.1.1 Extraction of Plasmid DNA from Bacterial Cells............................28
      3.2.1.1.1 DNA Wizard Miniprep.........................................................28
      3.2.1.1.2 DNA PureYield Miniprep.......................................................29
  3.2.1.2 Extraction of DNA from Agarose Gel.............................................29
  3.2.1.3 DNA Purification.................................................................................29
      3.2.1.3.1 QIAquick Purification...............................................................29
  3.2.1.4 Ethanol Precipitation...........................................................................29
3.2.2 DNA Analysis..............................................................................................30
  3.2.2.1 Agarose Gel Electrophoresis..............................................................30
  3.2.2.2 DNA Quantification............................................................................30
      3.2.2.2.1 NanoDrop....................................................................................30
3.2.3 DNA Sequencing..........................................................................................30
  3.2.3.1 Australian Genome Research Facility.............................................30
      3.2.3.1.1 Sequencing of Plasmid DNA.....................................................30
3.2.4 PCR Reactions.............................................................................................31
  3.2.4.1 General PCR.......................................................................................31
  3.2.4.2 Expand High Fidelity PCR.................................................................31
      3.2.4.2.1 Addition of FLAG-tag to the N-Terminus of GDF9..................32
      3.2.4.2.2 Addition of Restriction Enzymes to N and C-Terminus of GDF9...32
3.2.5 PCR Colony Screening...............................................................................32
3.2.6 DNA Modifications.....................................................................................32
  3.2.6.1 Site-Directed Mutagenesis.................................................................32
      3.2.6.1.1 DpnI Digestion of Parental Strand............................................33
  3.2.6.2 Restriction Enzyme Digests..................................................................33
  3.2.6.3 Dephosphorylation of 5’-Ends............................................................34
3.2.7 Ligation Reactions

3.2.7.1 Ligation into pDrive

3.2.7.1.1 TA Cloning

3.2.7.2 Ligation into pcDNA3.1

3.3 Tissue Culture

3.3.1 Resuscitation of Mammalian Cell Lines

3.3.2 Maintenance of Mammalian Cell Lines

3.3.2.1 A2780

3.3.2.2 HEK293

3.3.3 Transfection

3.3.3.1 Counting/Seeding of Mammalian Cell Lines

3.3.3.2 Transfection of Wildtype and Mutant GDF9 Constructs into HEK293 Cells using Lipofectamine 2000

3.3.4 Cell Harvesting

3.3.5 Collection of Cell Lysates

3.4 Protein Analysis

3.4.1 Protein Quantitation Using the Bicinchoninic Acid Assay

3.4.2 SDS-PAGE

3.4.3 Membrane Transfer

3.4.4 Western Analysis

3.4.5 ECL Detection

3.4.6 Membrane Stripping

3.4.7 Beta Actin

3.4.7.1 Beta Actin Optimization

Chapter 4: CLINICAL RECRUITMENT AND EXOME SEQUENCING FOR THE IDENTIFICATION OF PUTATIVE CAUSAL MUTATIONS IN PCOS

4.1 Introduction

4.2 Materials and Methods

4.2.1 Clinical Recruitment

4.2.2 Extraction of Genomic DNA from Patient Blood Samples
4.2.3 Exome Sequencing by the Australian Genome Research Facility .................................................................45
4.2.4 Bioinformatics Analyses ................................................................................................................................46
4.3 Results ........................................................................................................................................................................47
  4.3.1 Clinical Recruitment and Exome Sequencing .................................................................................................47
  4.3.2 Mutations Identified ........................................................................................................................................50
4.4 Discussion ..................................................................................................................................................................53
4.5 Conclusion ..................................................................................................................................................................58

Chapter 5: CLONING AND OPTIMIZATION OF EXPRESSION OF GDF9 CONSTRUCTS

5.1 Introduction .................................................................................................................................................................60
5.2 Methods ..........................................................................................................................................................................60
  5.2.1 Cloning of the GDF9 ORF into the pcDNA3.1 Mammalian Expression Vector ...........................................60
    5.2.1.1 Primer Design ..............................................................................................................................................60
    5.2.1.2 Extraction of GDF9 DNA and Cloning into the pDrive Cloning Vector .................................................61
    5.2.1.3 Sanger Sequencing .......................................................................................................................................61
    5.2.1.4 Site-Directed Mutagenesis ..........................................................................................................................61
    5.2.1.5 Ligation of GDF9 Wildtype and Mutant into pcDNA3.1 .........................................................................61
  5.2.2 Plasmid Construction – FLAG-tagged Constructs ...........................................................................................62
  5.2.3 Selection of Cell Lines for Transfection .............................................................................................................62
  5.2.4 Transfection ............................................................................................................................................................63
    5.2.4.1 Cell Lysis and Protein Quantitation .............................................................................................................63
    5.2.4.2 Western Blot Analysis .....................................................................................................................................63
5.3 Results ...........................................................................................................................................................................65
  5.3.1 Isolation of GDF9 DNA .........................................................................................................................................65
  5.3.2 Sequence Confirmation of Amplified GDF9 ........................................................................................................65
  5.3.3 Cloning of GDF9 ....................................................................................................................................................66
    5.3.3.1 Incorporation of FLAG-tag ..........................................................................................................................67
  5.3.4 Transfection Optimisation of Wildtype and Mutant pcDNA3.1-
Declaration

I declare this thesis is my own account of my research and contains as its main content, work which has not been previously submitted for a degree at any tertiary institution.

Silvia Sini-Laulu Borgen

21/06/2019
Acknowledgements

Firstly, thanks must go to my supervisor Prof. Scott Wilson who gave me the opportunity of this project and who supported me in my late start of the honours project with a kick-start! Thank you for your guidance, knowledge and expertise, especially when I was in the deep water of recruiting patients for the study and analysing the massive amounts of data from the exome sequencing. And many thanks to my supervisor Prof. Bronwyn Stuckey and Dr Jennifer Ng for further aiding me in the patient recruitment process. I must also thank Dr Bryan Ward. You have taught me so much throughout this year with all my wet lab work, your secret techniques and troubleshooting for success. Your help has been invaluable, and I can never thank you enough. Bryan, thank you for being an absolute lab wiz! And thank you to all of the girls at the lab for helping me out when in doubt and sharing your knowledge, tips and tricks with me 😊 Finally, thanks to everyone else working at the Keogh Institute and Bone Density, your daily smiles and interest in my project and progress has been very supportive and important to me.

Many thanks must also go to my Murdoch supervisor Dr Garth Maker for all your help with important paperwork and your guidance in the process of writing this thesis. Go science!! And thanks to the Murdoch honours coordinator Jacqueline Dyer for explaining and aiding me through the many forms, approvals, plus other paperwork, and for overall being an amazing lady.

Special thanks also go to my wonderful partner for supporting and encouraging me throughout the entire process, from applying for the honours program until my completion. You have been my rock through all of the good and the hard times, always cheering me forward! Thank for all your love and care. I would not have excelled like this without you by my side.
<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>HiC chromatin interaction data heatmap showing the potential relationship between rs13164856 and GDF9</td>
<td>15</td>
</tr>
<tr>
<td>4.1</td>
<td>Pedigree diagram of the individuals for whom DNA was exome sequenced in this study</td>
<td>46</td>
</tr>
<tr>
<td>4.2</td>
<td>Variant character distribution observed across the DNA of three PCOS study subject samples analysed by exome sequencing</td>
<td>49</td>
</tr>
<tr>
<td>4.5</td>
<td>Analysis of rs767893675 by PolyPhen-2 predictor</td>
<td>51</td>
</tr>
<tr>
<td>5.1</td>
<td>Flow diagram for the cloning approach used for the creation of the mutant GDF9 into the pcDNA3.1 mammalian expression vector</td>
<td>63</td>
</tr>
<tr>
<td>5.2</td>
<td>Isolation of GDF9 DNA</td>
<td>64</td>
</tr>
<tr>
<td>5.3</td>
<td>GDF9 amplified from the pMD-GDF9 clone</td>
<td>64</td>
</tr>
<tr>
<td>5.4</td>
<td>Chromatograms of pDrive-GDF9 showing correctly incorporated A. stop signal (TGA) and XhoI sequence, and B. EcoRV sequence</td>
<td>65</td>
</tr>
<tr>
<td>5.5</td>
<td>Chromatogram showing sequence confirmation of the GDF9 c.783delC mutant in pDrive</td>
<td>66</td>
</tr>
<tr>
<td>5.6</td>
<td>Chromatogram showing sequence confirmation of the incorporated FLAG-tag of the FLAG-tagged pcDNA3.1-GDF9 mutant</td>
<td>67</td>
</tr>
<tr>
<td>5.7</td>
<td>Effect of varying the dilutions of rabbit polyclonal GDF9 antibody, and incubation time</td>
<td>68</td>
</tr>
<tr>
<td>5.8</td>
<td>Effect of different transfection reagents and ratio on transfection efficiency of pcDNA3.1-GDF9 wildtype into A2780 cells</td>
<td>70</td>
</tr>
<tr>
<td>5.9</td>
<td>Effect of three ratios of Fugene HD on the expression of pcDNA3.1-GDF9 WT and mutant into A2780 cells</td>
<td>71</td>
</tr>
<tr>
<td>5.10</td>
<td>Effect of varying the amount of DNA and the Fugene HD: DNA ratio for expression of wildtype and mutant pcDNA3.1-GDF9 plasmid into A2780 cells</td>
<td>73</td>
</tr>
<tr>
<td>5.11</td>
<td>Effect of increased the incubation time on blocking</td>
<td>74</td>
</tr>
<tr>
<td>5.12</td>
<td>Effect of variable amounts of DNA transfected of pcDNA3.1-GDF9 wildtype, mutant, and empty vector into HEK293 cells, using Lipofectamine 2000</td>
<td>76</td>
</tr>
<tr>
<td>5.13</td>
<td>Effect of variable amounts of DNA transfected of FLAG-tagged GDF9 mutant, EEF2K, and empty vector in HEK293 cells, using Lipofectamine 2000</td>
<td>77</td>
</tr>
</tbody>
</table>
Index of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1</td>
<td>BCA protein standards</td>
<td>37</td>
</tr>
<tr>
<td>4.1</td>
<td>Showing the mean and standard deviation for the three sisters</td>
<td>47</td>
</tr>
<tr>
<td>4.2</td>
<td>Summary of information from PCOS questionnaires</td>
<td>47</td>
</tr>
<tr>
<td>4.3</td>
<td>Analyst remarks from AGRF of the three samples submitted for exome sequencing</td>
<td>48</td>
</tr>
<tr>
<td>4.4</td>
<td>Potentially pathogenic variants identified in all three sisters diagnosed with PCOS through exome sequencing</td>
<td>50</td>
</tr>
<tr>
<td>4.5</td>
<td>Potentially pathogenic variants identified in all patients through exome sequencing</td>
<td>50</td>
</tr>
</tbody>
</table>
Abstract

Polycystic ovary syndrome (PCOS) is an endocrinopathy that affects 5% to 10% of women. It is the most common reproductive disorder in women and is characterised by menstrual irregularity, hyperandrogenism, and polycystic ovarian morphology. This disorder also involves a major risk factor for non-insulin dependent diabetes, hypertension, and cardiovascular disease.

While PCOS is believed to have a complex genetic component, the exact nature of this component remains largely unknown. However, Genome Wide Association Studies (GWAS) and exome sequencing have begun to generate statistically significant and replicable data from which more specific investigations can be launched. PCOS related variants highlighted by these methods can then be the subject of functional studies that seek to identify the functional role that these variants may be involved in. The efficient uncovering of variants, including exomic identification and functional characterisation of the genetic lesions, can aid genetic diagnosis of PCOS, which in turn will help achieve better patient outcomes and provide the basis for targeted therapies.

The focus of this study was that efficient genetic diagnosis of PCOS will be assisted by an increased understanding of the aetiology of PCOS gained via exome sequencing identification of related individuals and functional characterisation of variants. To this end, related individuals with clinically defined PCOS were recruited and exome sequencing carried out to uncover putative causal mutations.

In addition, a functional study of growth differentiation factor-9 (GDF9) was conducted due to its role in folliculogenesis and female fertility, as well as its proximity to a PCOS linked variant (rs13164856) identified via a GWAS. A single 1-bp deletion mutation (783celC) in GDF9 was targeted, as this deletion is known to cause primary ovarian insufficiency. The possible functional role of this GDF9 variant was then tested using an in vitro model. This was achieved by creating mammalian expression vectors containing the wildtype and mutant sequences.
The primary achievement of this thesis was the uncovering of causal putative mutations shared between first-degree relatives suffering from PCOS. This data will potentially aid in uncovering the aetiology of PCOS. Functional characterisation of the GDF9 mutation was unsuccessful due to failure of overexpression of the GDF9 mutant construct.
Chapter 1

Introduction
1 INTRODUCTION

1.1 General Introduction

Polycystic ovary syndrome (PCOS) is the most common endocrinopathy among women of reproductive age, affecting 5-10 per cent of the population (Morin-Papunen et al., 2012). PCOS is a syndrome and, as such, no single diagnostic criterion is sufficient for clinical diagnosis; it is also a diagnosis of exclusion. The principal features are hyperandrogenism, oligomenorrhea and polycystic ovarian morphology (PCO). The condition is also characterised by, hirsutism, acne, insulin resistance, obesity, impaired glucose tolerance, type II diabetes, and elevated serum luteinizing hormone (LH) levels. Moreover, women with PCOS are at an increased risk of cardiovascular incidents and may display features of cardiometabolic syndrome (CMS) (Fauser 2004; Broekmans et al., 2006).

1.1.1 History of Polycystic Ovary Syndrome

The first incidence of PCOS was officially described by the two American gynaecologists, Irving F. Stein and Michael L. Leventhal, in 1935 and they are regarded as the first investigators of PCOS. However, much earlier in 1721, an Italian medical scientist, Dr Vallisneri, reported a married, infertile woman, with shiny large ovaries (Insler and Lunenfeld 1990). Another report of abnormal ovaries descriptive of PCOS was made in 1844 by the French physician Dr Chereau, and in 1845 the Czech physician Dr Rokitansky described fibrous and sclerotic lesions in the ovaries with hydrops follicles. Soon after, hyperthecosis caused by luteinization producing androgens was described for the first time in 1897 by Bulius and Kretschmar. Later, in 1902, Kahlden published a review on the clinical implications and pathology of these abnormal ovaries. In 1935, Stein and Leventhal presented a group of women with common clinical characteristics of hirsutism, menstruation disturbances, and enlarged ovaries with multiple small follicles. With time, the then called Stein-Leventhal syndrome developed to what is now known as PCOS.
1.2 Diagnosis

There is a controversy regarding the diagnosis of PCOS and there are currently three sets of diagnostic criteria used: The National Institute of Health (NIH), Rotterdam, and the Androgen Excess-PCOS Society (AE-PCOS) criteria. The NIH criteria were developed during a conference held in 1990 and PCOS was defined as clinical or biochemical evidence of hyperandrogenism (excess male hormones) and oligo-ovulation (irregular ovulation) (Kauffman et al., 2008). The NIH criteria were modified by a consensus meeting held in 2003 in Rotterdam. The defining symptoms of PCOS were expanded in the Rotterdam criteria to include women with any two of these three criteria: polycystic ovaries on ultrasound, oligo- or anovulation, and clinical or biochemical hyperandrogenism. However, the flexibility of this new criteria allows for the inclusion of a larger number of women and hence this wide spectrum raises the question if all women diagnosed with PCOS are exposed to the same long-term health implications (Broekmans et al., 2006). The most recent criteria for diagnosing PCOS came out in 2006 and was formulated by the task force of the Androgen Excess Society. These criteria include hirsutism and/or hyperandrogenemia; oligo-ovulation and/or polycystic ovaries; and exclusion of other androgen excess or related disorders (Trivax and Azziz 2007). The diagnosis and prevalence of PCOS therefore depends to a great degree on the criteria used to define the syndrome.

1.3 Clinical Signs and Symptoms

1.3.1 Hyperandrogenism

PCOS may present with a variety of clinical manifestations. These clinical symptoms may vary according to racial, ethnic, environmental factors, and the presence of other factors such as insulin resistance and obesity, and none of the clinical features are pathognomonic on their own. However, the main clinical features of PCOS are hyperandrogenism, polycystic ovaries, and ovulatory dysfunction (Sirmans and Pate 2013).

Hyperandrogenism is present in more than 80% of women presenting with PCOS (Azziz et al., 2004). The presence of hyperandrogenism is determined either clinically or biochemically (Sirmans and Pate 2013).
1.3.1.1 Clinical Hyperandrogenism
Clinical hyperandrogenism often presents with hirsutism (male-pattern hair growth) and occurs in up to 70% of PCOS women (Fauser et al., 2012). However, the assessment of hirsutism is relatively subjective, and few physicians use standardized scoring methods (Futterweit et al., 2006). Moreover, hirsutism is often treated long before a diagnosis of PCOS is made. Normative data is still lacking in large populations, but hirsutism may be significantly less prevalent in adolescents and women of East Asian origin who have PCOS. The presence of acne is another potential marker for hyperandrogenism but is less prevalent and less specific than hirsutism for diagnosis of PCOS (Fauser 2004). The presence of androgenic alopecia may indicate hyperandrogenism, however, it is less well studied and appears to be a relatively poor marker (Futterweit et al., 2006).

1.3.1.2 Biochemical Hyperandrogenism
Biochemical hyperandrogenism is a measure of circulating androgen levels in blood made using various steroid assay techniques (e.g. radioimmunoassay, chemiluminometric immunoassay, gas chromatography/mass spectrometry). Limitations to the measurement of androgen levels are in part due to the inaccuracy and variable laboratory methods used (Boots et al., 1998; Rosner 1997; Vermeulen et al., 1999). Moreover, several androgens may not be considered in routine assessments, there is wide variability in the apparently healthy population, and androgens are rapidly altered by hormonal suppression and may remain low even after discontinuation of such treatment (Ruth 2016: Fauser 2004). As a result, normative ranges have not been well established.

1.3.2 Polycystic Ovary Morphology
Polycystic ovaries (PCO) are defined as the presence of 12 or more follicles in each ovary with a size range of 2-9 mm in diameter, and/or by the measurement of increased ovarian volume of more than 10 ml (Balen et al., 2003). The presence of either in just one of the ovaries is sufficient for the diagnosis of PCO (Fauser 2004). However, since this definition was established there has been significant advancements in ultrasound image technologies, resulting in improved resolution enabling reliable detection of smaller follicles (Lujan et al., 2013). The presence of PCO alone is not sufficient for PCOS diagnosis; largely because 20-
30% of otherwise normal women of reproductive age have evidence of multiple cysts on their ovaries (Sirmans and Pate 2013)

1.3.3 Ovulatory Dysfunction

Ovulatory dysfunction creates a menstrual disturbance and is described as abnormal if menstrual periods are i) irregular (9 or fewer menses per year) or ii) absent (no release of the oocyte at ovulation). Menstrual disturbances commonly observed in PCOS women include the absence of menstruation for three or more menstrual cycles (amenorrhea), infrequent menstrual cycles of no less than 35-day intervals or no less than 10 bleeds per year (oligomenorrhea), and prolonged erratic menstrual bleeding (Futterweit et al., 2006; Sirmans and Pate 2013).

The World Health Organization (WHO) categorizes ovulatory dysfunction into three groups. Group one ovulation disorders are caused by hypothalamic-pituitary failure and includes, among others, hypothalamic amenorrhea and hypogonadotropic hypogonadism. This group is present in about 10% of women with ovulatory disorder. Group two ovulatory disorders include conditions like PCOS and hyperprolactinemic amenorrhea and are defined as dysfunctions of the hypothalamic-pituitary-ovarian axis. About 85% of women suffering from ovulation dysfunction belong to group two. Finally, group three ovarian dysfunction is caused by ovarian failure, and approximately 5% of women with an ovulation disorder are classified as group three (Treasure et al., 2013).

Ovulation disorders are the cause of infertility in approximately 25% of women that experience difficulty conceiving and PCOS accounts for 75% of patients with ovulatory infertility (Futterweit et al., 2006; Treasure et al., 2013; O’Flynn 2014). Moreover, ovulatory dysfunction affects 80-100% of PCOS women, depending on the diagnostic criteria used (Azziz et al., 2004; Futterweit et al., 2006). However, not all PCOS patients display a clinically apparent abnormality in their cyclic vaginal bleeding pattern and 30% of women with PCOS will have normal menses. Approximately 85-90% of women presenting with oligomenorrhea have PCOS whereas 30-40% of women with amenorrhea have PCOS (Balen et al., 1995; Hart 2004). It should also be noted that more subtle changes in menstrual cycling may occur in PCOS, including dysfunctional uterine bleeding and intermittent anovulation, these may
occur in association. Additionally, menstrual function in PCOS women often becomes more regular at the age of 30-40 years old. Irregular or absent ovulation may also be caused by obesity, low BMI, extreme exercise, advanced maternal age, premature ovarian failure, thyroid dysfunction and stress (Futterweit et al., 2006). Consequently, ovulatory dysfunction cannot alone be taken as proof of PCOS.

1.4 Secondary Increased Health Risks

1.4.1 Weight Gain and Obesity

Weight gain after puberty and the early development of obesity can be associated with PCOS. It is common for symptoms of PCOS to develop after onset of menstruation, however, the development of menstrual dysfunction and cutaneous signs of hyperandrogenism, like acne, hirsutism, skin tags and acanthosis nigricans (patches of dark discolouration), are commonly incited by a significantly increased weight gain (Hooff et al., 2004). Women suffering from PCOS often have a history of sudden weight gain over a period of 6-12 months, and most commonly a persistent gain of 4-7 kg per year (Futterweit et al., 2006). Body mass index (BMI) has also been noted as a predictor for the persistence of oligomenorrhea into adulthood and the likely development of PCOS (Hooff et al., 2004). Obesity in PCOS women is highly prevalent in the United States and here obesity is observed in up to 60% of the patients (Azziz et al., 2004). However, obesity in PCOS women is less prevalent in other populations. In addition, the prevalence of obesity in PCOS women is higher among adults than adolescents within the same population (Guleki et al., 1993). Hence, adolescents should not be expected to have the same degree of obesity as older PCOS patients. The buildup of visceral adiposity (fat stored within the abdominal cavity) is frequent and is associated with many metabolic abnormalities in PCOS, including, but not limited to, insulin resistance, dyslipidemia, and glucose intolerance (Ehrmann 2005). Weight gain may be associated with carbohydrate craving and serve as evidence of reactive hypoglycemia. Hypoglycemia is the pre-diabetic condition for insulin resistance (Holte et al., 1995). Type two diabetes and impaired glucose tolerance are lower in non-obese compared to obese PCOS women and the conversion of impaired glucose tolerance to type 2 diabetes is closely correlated with increased BMI. Overall, weight gain and obesity are quite modest predictors of PCOS development (Futterweit et al., 2006).
1.4.2 Insulin Resistance

Insulin resistance is not required for the diagnosis of PCOS, conversely, it is clear that hyperinsulinemic insulin resistance has an important role in the syndrome. 50-70% of PCOS women are diagnosed with insulin resistance, and this prevalence is independent of obesity (Carmina and Lobo 2004; Dunaif et al., 1989). More specifically, the effect that obesity has on insulin resistance is additive to that of PCOS. The insulin resistance observed in PCOS appears to be due to a specific but abnormal pattern of insulin receptor phosphorylation. Hyperinsulinemic insulin resistance produces an increase of free androgens in the blood resulting in a hyperandrogenic state; thought to be due to a combination of increased androgen production in theca cells (endocrine cells associated with ovarian follicles) and reduced liver production of sex hormone-binding globulin (SHBG). This increased prevalence of insulin resistance is also consistent with the metabolic syndrome being more common in PCOS women (Sirmans and Pate 2013).

1.4.3 Metabolic Syndrome

Metabolic syndrome is a clustering of both lipid and non-lipid risk factors that identify individuals at a heightened risk of developing type two diabetes mellitus and cardiovascular disease. Numerous metabolic abnormalities of PCOS overlap with components of metabolic syndrome (Ehrmann et al., 2006). One such component is insulin resistance which appears to play a pathogenic role in the metabolic syndrome (Ehrmann et al., 2006). The National Cholesterol Education Program Adult Treatment Panel guidelines define metabolic syndrome as having three or more of the following: fasting serum glucose at least 110 mg/dl; fasting serum triglycerides at least 150 mg/dl; Serum HDL-Cholesterol less than 50 mg/dl; waist circumference in females greater than 88 cm; and blood pressure at least 130/85 mm Hg (Apridonidze et al., 2005). The presence of acanthosis nigrican (dark, velvety patches in body folds and creases) and/or acrochordons (small soft skin growths) are suggestive of metabolic syndrome (Futterweit et al., 2006). Metabolic syndrome is a common disorder and it is particularly prevalent in PCOS. According to the Adult Treatment Panel III criteria, 43-46% of American women with PCOS have metabolic syndrome (Carmina et al., 2006). Moreover, a study by Ehrmann et al. (2006) concluded that the prevalence of metabolic syndrome in PCOS does not vary significantly between ethnic groups.
1.4.4 Gestational Diabetes

Gestational diabetes mellitus (GDM) is the first recognition, or onset, of any degree of glucose intolerance during pregnancy (Mpondo et al., 2015). While several studies have shown an increased risk for GDM in women with PCOS (Rees et al., 2016; Mumm et al., 2015), this has been found to be related to obesity and maternal age, rather than to PCOS itself (Mustaniemi et al., 2018). This suggests that routine early screening for GDM in women with PCOS may not be necessary unless other risk factors are present.

1.4.5 Impaired Glucose Tolerance and Type II Diabetes

PCOS is associated with a higher prevalence of impaired glucose tolerance and type two diabetes, independently of obesity. Impaired glucose tolerance is very common in PCOS women with a 30-40% prevalence and 7-10% have type two diabetes. These rates are significantly higher than the normative ranges observed. The Second National Health and Nutrition Survey reported a prevalence for women in the general population of similar age as 7.8% for impaired glucose tolerance and a 2.5% rate for diabetes. However, most women with PCOS maintain adequate beta-cell function for insulin production to prevent a decrease in glucose tolerance, although a significant proportion demonstrates an abnormal beta-cell response to glucose challenge, especially when having a first degree relative with type two diabetes (Sirmans and Pate 2013).

1.4.6 Dyslipidemia

An increased risk of dyslipidemia has been demonstrated in PCOS and these lipid abnormalities include reduced high-density lipoprotein-cholesterol, increased low-density lipoprotein-cholesterol, and increased triglycerides. Moreover, studies have revealed that women with PCOS have highly atherogenic small lipoprotein that promotes the formation of fatty acids in the arteries and increases the risk of cardiac events. Obesity and insulin resistance further promote the formation of very-low-density lipoprotein particles. Moreover, insulin resistance is also associated with reduced clearance of the lipoprotein particles and triglycerides, and when triglycerides are in high abundance larger lipoprotein particles are formed. Metabolism of these large lipoprotein particles leads to the formation of small and dense lipoproteins that are more atherogenic and not easily removed from the body (Sirmans and Pate 2013).
1.4.7 Hypertension
Several studies have found that PCOS women have a higher prevalence of hypertension (Sirmans and Pate 2013; Lo et al., 2006). A potential mechanism for this is increased aldosterone concentrations related to insulin resistance. A study by Lo et al. (2006) found that PCOS women were more likely to have hypertension than controls after adjusting for age, BMI, diabetes, and dyslipidemia. Hypertension was found to be present in 22% of the PCOS subjects compared to 2.1% of the controls (Sirmans and Pate 2013).

1.4.8 Cardiovascular Disease
Cardiovascular disease risk factors have an increased prevalence and appear to be clustering in PCOS, yet there is scarce evidence that PCOS is associated with increased cardiovascular events (Sirmans and Pate 2013). There are limited longitudinal studies on the risk of cardiovascular disease in PCOS women, and those studies are too small to distinguish any differences in event rates (Legro et al., 2013). However, a greater prevalence of subclinical atherosclerosis has been demonstrated in PCOS by the noninvasive assessments of coronary artery calcium (CAC) scores and carotid intima-media thickness (CIMT). CAC scores predict myocardial infarction and sudden cardiac death, and CIMT is positively correlated with myocardial infarction and stroke and serves as a reliable measurement for atherosclerosis. The increase of CAC scores and CIMT in PCOS is independent of BMI and age (Sirmans and Pate 2013). The revised 2003 consensus on diagnostic criteria and long-term health risks related to PCOS highlights two factors to keep in mind when evaluating the increase in cardiovascular events of PCOS women: the relatively young age of cohorts studied so far and the possibility that unknown factors present in PCOS could potentially protect from cardiovascular disease (Fauser 2004). Furthermore, a study by Stuckey et al. (2014) concluded that the clustering of cardiovascular and metabolic risk factors is consistent with the different pathogenetic pathways affected in PCOS and/or the different cardiometabolic outcomes.
1.4.9 Sleep Disorders
The prevalence of sleep disorders is high in PCOS, especially sleep apnea. The most common symptom of sleep apnea is snoring and is caused by intermittent airflow blockage during sleep, resulting in decreased oxygen supply. Sleep apnea increases the risk of developing hypertension, heart attack, and stroke and it is positively correlated with obesity and may be linked to hyperandrogenism and insulin resistance. Laboratory-proven sleep apnea is found at a 4- to 30-fold higher incidence in women with PCOS compared to controls. The resulting consequences of sleep apnea are sleepiness during daytime, reduced mental performance, and overall reduced quality of life (Futterweit et al., 2006).

1.4.10 Mental Health Disorders
Women with PCOS are also more likely to have or develop several mental health disorders including, bipolar disorder, anxiety, depression, and various eating disorders. A study screening attendees of the Polycystic Ovarian Association regional meeting found that approximately 27% of the participants had bipolar disorder. In another study, the prevalence of bipolar disorder in PCOS women was reported to be 11%. Those reported levels of occurrence of bipolar disorder in PCOS is high in relation to the estimated prevalence of 0.5-2% in the general population. The prevalence of anxiety disorders in PCOS have been reported by several studies and range from 13-64% (Sirmans and Pate 2013). A recent meta-analysis indicates that anxiety is more common among PCOS women than controls (Dokras et al., 2012). Depression has been demonstrated to be significantly higher in PCOS women with a 35% prevalence, compared to 11% in controls. Various eating disorders are also more common in PCOS with binge eating disorders affecting 13% of PCOS women compared to 2% of controls (Hollinrake et al., 2007).
1.5  Other Potential Risk Factors

1.5.1  Lifestyle Factors
Lifestyle factors that cause weight gain, such as high fat and carbohydrate diets and physical inactivity, can worsen the hormonal and metabolic symptoms of PCOS. Conversely, lifestyle changes of improved diet, increased exercise, and behavioural modification, appears to improve the abnormal metabolic symptoms and reproductive difficulties of overweight and obese PCOS women (Panidis et al., 2013). The relationship between PCOS and obesity is complex and not well understood, and most likely involves the interaction of genetic and environmental factors (Sam 2007).

1.5.2  Environmental
The complete aetiology of PCOS continues to be unclear, however, environmental and genetic components are involved in the pathogenesis of the syndrome. Environmental insults may affect susceptible women throughout life to eventually develop PCOS. Exposure to environmental triggers in the early stages of development, in particular, could change a predisposed genotype to a phenotype of PCOS. Moreover, the clinical severity of PCOS may be moderated by environmental triggers such as chemical exposures (Diamanti-Kandarakis et al., 2012).

1.5.2.1 Exposure to Chemicals
Environmental exposure to chemicals, such as industrial products, may also aggravate the clinical presentation of PCOS. Bisphenol A (BPA), which is an exogenously derived substance, in particular, may act as an endocrine disruptor. Endocrine disruptors are environmental chemicals which have the potential to interfere with metabolism, hormone biosynthesis, or processes resulting in abnormal reproduction or homeostasis (Diamanti-Kandarakis et al., 2012).

Recent research data suggest an association of increased tissue and fluid levels of BPA with the metabolic, endocrine and reproductive features of PCOS. Exposure to BPA appears to have detrimental reproductive effects by acting directly on the ovary and may impair ovarian steroidogenesis and folliculogenesis. This is supported by a reported impairment of ovarian follicular growth and decreased corpus luteum formation in rodents exposed to BPA.
at early development. Interestingly, the level of BPA environmental exposure to humans in daily life is equivalent to doses used in these experiments (Diamanti-Kandarakis et al., 2012). It could be speculated that BPA may contribute to the development of anovulation seen in PCOS. Also, exposure to BPA has been linked to changes in ovarian steroidogenesis. The administration of PBA in rats results in enhanced testosterone synthesis. Furthermore, granulosa cell function appears to be affected by BPA. Porcine granulosa cells that had BPA administered at physiological dose were shown to increase FSH-stimulated progesterone synthesis and suppress FSH-induced estradiol production (Mlynarčíková et al., 2005). Animals exposed to BPA during developmental periods present with abnormal ovarian morphology, with a resemblance to PCOS (Diamanti-Kandarakis et al., 2012).

BPA exposure also has potential adverse metabolic effects, especially on diabetes and cardiovascular disease. Development of metabolic syndrome may be favoured by BPA through inhibition of the adiponectin hormone which is involved in regulating glucose levels. Adiponectin increases insulin sensitivity and the inhibition by BPA could be suggested to exacerbate insulin resistance. Moreover, findings by Do et al. (2017) suggest that BPA has an adverse interaction with lipid homeostasis and regulation of body weight and this conveys a potential link between obesity and BPA. Advanced glyced end products (AGEs) are another chemical that may contribute to the development of PCOS. Exogenous AGEs appear to act as endocrine disruptors and could have major implications in the reproductive and metabolic dysfunctions characteristic of PCOS (Diamanti-Kandarakis et al., 2012).

### 1.5.2.2 Foetal Exposure to Androgens and Other Intrauterine Events

In a 2015 study by Gur et al. (2015) it was found that prenatally androgenized monkeys had high levels of androstenedione at birth and increasing levels of adrenal origin androgens for a period of 4 to 25 months after birth. This suggests that prenatal androgen exposure may change adrenal androgen production. Morphological changes seen in these prenatally androgenized foetuses, such as increased numbers of follicles and accelerated proliferation of granulosa cells, are similar to the increased follicular development that is seen in PCOS patients (Gur et al., 2015).
It is not clear if this animal model is useful when looking at human foetuses. While significantly elevated serum androgen levels have been reported in pregnant women with PCOS during mid-gestation (Sir-Petermann et al., 2002), these increased maternal serum androgens may not result in increased foetal androgen exposure, as placental aromatase is known to form a barrier to maternal androgens by converting androgens into estrogens (Kaňová and Bičíková, 2011). However, this conversion may not fully take place when excessive amounts of androgens are present, and placental aromatase is unable to aromatise dihydrotestosterone (DHT), which is an androgen often found at higher levels in hyperandrogenic states of ovarian aetiologies, such as PCOS (Kaňová and Bičíková, 2011). Further research is needed to elucidate whether foetal exposure to androgens plays a role in the development of PCOS.

1.5.3 Genetic Factors
PCOS is of unknown aetiology but there is strong evidence of a genetic component to PCOS and it has been proven to be a familial condition through studies of the clustering of PCOS in families (Urbanek 2007; Prapas et al., 2009). Family history is an important tool for predicting PCOS in women, because the risk of developing the syndrome when a first-degree relative is affected is greater than 30% (Futterweit et al., 2006). The mode of inheritance is complex, due to the great genetic heterogeneity of PCOS and diverse clinical manifestations, and PCOS is generally considered a polygenetic disorder (Urbanek 2007; McAllister et al., 2015). The roles of more than 70 genes in the aetiology of PCOS have been evaluated, however, findings in these studies have often been ambiguous and lacking replication most likely due to an absence of consistent phenotypic criteria across studies, incomplete characterization of candidate genes, and limited sample sizes, resulting in the slow progress of genetic study in PCOS (Urbanek 2007). This slow progress may also be due to the syndrome being perceived as a relatively benign disorder when compared to many debilitating genetic disorders. This results in funding difficulties for PCOS genetics research. Genetic factors very likely contribute to PCOS, and more research is needed. The genetic factors involved in PCOS and recent developments in this area are discussed in detail in section 1.6.
1.6 Genetics of PCOS

The involvement of genetics in the aetiology of PCOS was established due to observations of familial clustering of PCOS consistent with a genetic component (Legro et al., 1998). Research into the genes involved in this component began with the use of candidate gene studies. These studies investigate genes that may be involved with PCOS due to their function in known biochemical pathways that are relevant to the clinical symptoms of the syndrome. Once identified, variants near or in these genes are targeted if it is believed they are in linkage disequilibrium (LD) with a variant that could alter the encoded protein or could alter the encoded protein themselves. Differences in allele frequency are then analysed between cases and controls or between variants and disorder phenotype variation (Tabor et al., 2002). Many genes have been investigated in this way (Jones et al., 2006; Jones et al., 2007; Jones et al., 2008). However, these studies have mostly been carried out on small populations, with limited statistical power, and poor replicability, which has made them difficult to meaningfully interpret (Strauss et al., 2012).

Genome Wide Association Studies (GWAS) are a powerful method of identifying variants that are associated with a genetic disease that overcomes the flaws of candidate gene studies and provides evidence that PCOS is indeed a complex polygenic disorder (Strauss et al., 2012). These studies involve searching the genome of cases and controls for single nucleotide polymorphisms (SNPs) associated with a specific trait. These study designs are hypothesis-free and typically show reliable replication for genuinely associated variants. Multiple studies have managed to identify gene variants with enough statistical power and replicability to significantly associate them with PCOS (Day et al., 2018; Day et al., 2015; Shi et al., 2012; Hayes et al., 2015; Lee et al., 2015; Goodarzi et al., 2012; Chen et al., 2011). The genes identified in these studies include THADA, ERBB4, IRF1/RAD50, GATA4/NEIL2, FANCC, DENND1A, ARL14EP/FSHB, YAP1, ERBB3/RAB5B, KRR1, TOX3, and the most recently identified PLGRKT, ZBTB16, and MAPRE1 (Day et al., 2018). Such studies are a promising source of information on the genetic components of PCOS, but much research remains to be conducted, as many of these variants are located in non-coding DNA regions, are in strong linkage disequilibrium (LD) with other variants, or are not the nucleotides involved in the regulation of transcription.
1.7 GDF9 as a Susceptibility Gene for PCOS

1.7.1 Function and Characteristics of GDF9

Growth differentiation factor-9 (GDF9) is a member of the transforming growth factor-beta (TGFβ) superfamily first discovered in 1993 (McPherron and Lee, 1993). GDF9 plays a key role in folliculogenesis and female fertility through selective expression in developing oocytes (Otsuka et al., 2011; Filho et al., 2002; Erickson and Shimasaki, 2000; McGrath et al., 1995). Reproductive abnormalities such as arrested follicle growth at the primary stage, disturbed theca development, reduced granulosa proliferation, cyst formation, and infertility have all been shown to occur in GDF9 deficient female mice (Elvin et al., 1999; Carabatsos et al., 1998; Dong et al., 1996). In humans, ovarian insufficiency has been reported as the result of a homozygous 1 bp deletion in GDF9 (França et al., 2018).

1.7.2 Why GDF9 was Chosen for Functional Study

In a 2015 GWAS paper, Day et al. identified a novel variant (rs13164856) associated with PCOS near the genes RAD50 and IRF1 (Day et al., 2015). While neither of these genes are likely to be involved in PCOS, Hi C Chromatin interaction maps, as displayed in Figure 1.1, show that GDF9 is within the DNA looping vicinity of the PCOS linked variant. This variant could therefore be highlighting a regulatory element for GDF9. Given the physiological role of GDF9 and this potential variant link to PCOS, it was decided that GDF9 is a strong candidate to be involved in PCOS and an excellent opportunity to develop a protocol for molecular genetic functional characterization of variants in PCOS studies.
Figure 1.1: HiC chromatin interaction data heatmap showing the potential relationship between rs13164856 and GDF9
1.8 Statement of Hypothesis and Aims

Current diagnostic criteria for PCOS often result in heterogeneous groups of patients with very different genetic lesions and limited treatment options. Exome sequencing and functional genomics are capable and cost-effective methods of investigating pathological genetic variants and their cellular effects. Therefore, the hypothesis for this study is that efficient genetic diagnosis of PCOS, including exomic identification and functional characterisation of the genetic lesions, will help achieve better patient outcomes and provide the basis for targeted therapies.

This study aimed to:

i) recruit related individuals with clinically defined PCOS in families that may harbour private mutations or rare variants of large effect size.

ii) perform exome sequencing followed by analysis with Exomiser software on the affected individuals from a selected family to assess the efficiency of this approach as a molecular diagnostic.

iii) develop a protocol for efficient functional characterisation of PCOS variants to inform on the relationship between genotype and PCOS phenotype.

Achievement of these aims will aid PCOS patients to better outcomes, particularly in the case of mutations with a large effect size that is segregating within a family (as opposed to other cases where there is likely to be a relatively large number of alleles contributing to the disease susceptibility), provide a basis for further PCOS functional characterisations, and gather valuable data on the complex genetic aetiology of PCOS.
Chapter 2

Materials
## 2 MATERIALS

### 2.1 General Reagents and Biochemicals

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose, Analytical Grade</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Alkaline Phosphatase, Calf Intestinal</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Alkaline Phosphatase, Buffer 10X</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Ampicillin (50 mg/mL)</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>β-mercaptoethanol</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Bacto™ Agar</td>
<td>Becton Dickinson, USA</td>
</tr>
<tr>
<td>Bacto™ Tryptone</td>
<td>Becton Dickinson, USA</td>
</tr>
<tr>
<td>Bacto™ Yeast Extract</td>
<td>Becton Dickinson, USA</td>
</tr>
<tr>
<td>Bovine Serum Albumin</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Bromophenol Blue</td>
<td>Hopkin and Williams, USA</td>
</tr>
<tr>
<td>Calcium Chloride (CaCl₂)</td>
<td>Ajax FineChem, AU</td>
</tr>
<tr>
<td>ddH₂O Ultra-Pure</td>
<td>Invitrogen (Thermo Fisher), USA</td>
</tr>
<tr>
<td>Disodium Phosphate (Na₂HPO₄)</td>
<td>BDH Chemicals, USA</td>
</tr>
<tr>
<td>DNA Ladder, 1 kb Plus</td>
<td>Life Technologies (Thermo Fisher), USA</td>
</tr>
<tr>
<td>Dulbecco’s Modification of Eagles Medium (DMEM)</td>
<td>Gibco (Thermo Fisher), AU</td>
</tr>
<tr>
<td>dNTPs, 5 mM</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Ethanol, 70%</td>
<td>Rowe Scientific, AU</td>
</tr>
<tr>
<td>Ethanol, Absolute</td>
<td>Rowe Scientific, AU</td>
</tr>
<tr>
<td>Ethidium Bromide (EtBr)</td>
<td>Roche Diagnostics, Switzerland</td>
</tr>
<tr>
<td>Ethylenediaminetetra-acetic acid (EDTA)</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Foetal Bovine Serum (FBS)</td>
<td>Moregate Biotech, AU</td>
</tr>
<tr>
<td>Foetal Calf Serum (FCS)</td>
<td>Moregate Biotech, AU</td>
</tr>
<tr>
<td>Fugene 6</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Fugene HD</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>GelRed NA Stain (10,000X)</td>
<td>Fisher Biotech, AU</td>
</tr>
<tr>
<td>Compound</td>
<td>Supplier</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>-------------------------------</td>
</tr>
<tr>
<td>Glucose ($C_6H_{12}O_6$)</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Glycerol ($C_3H_{10}O_3$)</td>
<td>Merck, USA</td>
</tr>
<tr>
<td>Glycine ($C_2H_5NO_2$)</td>
<td>MP Biomedicals, USA</td>
</tr>
<tr>
<td>HEPES</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Isopropanol</td>
<td>Rowe Scientific, AU</td>
</tr>
<tr>
<td>Lipofectamine 2000</td>
<td>Life Technologies (Invitrogen), USA</td>
</tr>
<tr>
<td>Liquid Nitrogen</td>
<td>BOC Limited, UK</td>
</tr>
<tr>
<td>Magnesium Chloride ($MgCl_2$): 25 mM</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>1 M, Filter Sterilised</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Magnesium Sulphate ($MgSO_4$), 1 M, Filter Sterilised</td>
<td>BDH Chemicals, USA</td>
</tr>
<tr>
<td>Methanol</td>
<td>Rowe Scientific, AU</td>
</tr>
<tr>
<td>OPTIMEM, X1</td>
<td>Gibco (Thermo Fisher), AU</td>
</tr>
<tr>
<td>Pen/Strep; Penicillin – 10,000 U/mL</td>
<td>Harry Perkins Institute of Medical Research, AU</td>
</tr>
<tr>
<td>Streptomycin – 10 mg/mL</td>
<td></td>
</tr>
<tr>
<td>Potassium Chloride ($KCl$)</td>
<td>David Brown Scientific, AU</td>
</tr>
<tr>
<td>Potassium Phosphate Monobasic ($KH_2PO_4$)</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Protease Inhibitor (tablets)</td>
<td>Roche Diagnostics, Switzerland</td>
</tr>
<tr>
<td>Precision Plus Protein™ Dual Colour Standards</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>Restriction Enzymes: EcoRI</td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td></td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td></td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td>Restriction Enzyme Buffer: NEBuffer3.1</td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td></td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td>Salmon Sperm, 10 mg/mL</td>
<td>Provided by Dr Bryan Ward, Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, WA</td>
</tr>
<tr>
<td>Sodium Acetate ($C_2H_3NaO_2$)</td>
<td>BDH Chemicals, USA</td>
</tr>
<tr>
<td>Sodium Chloride (NaCl)</td>
<td>BDH Chemicals, USA</td>
</tr>
<tr>
<td>Sodium Deoxycholate</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Item</td>
<td>Supplier</td>
</tr>
<tr>
<td>-------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Sodium Dodecyl Sulfate</td>
<td>ICN Biomedicals inc., USA</td>
</tr>
<tr>
<td>Sodium Hydroxide (NaOH)</td>
<td>Ajax FineChem, AU</td>
</tr>
<tr>
<td>T4 DNA Ligase</td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td>T4 DNA Ligase, Buffer 10X</td>
<td>New England Biolabs Inc., USA</td>
</tr>
<tr>
<td>Taq DNA Polymerase, GoTaq® Flexi</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Taq DNA 5X Buffer, GoTaq® Flexi</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Tetracycline (50 µg/mL)</td>
<td>Unknown Source</td>
</tr>
<tr>
<td>Trizma® Base</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Triton X-100</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Trypsin-EDTA, 0.25%</td>
<td>Harry Perkins Institute of Medical Research, AU</td>
</tr>
<tr>
<td>Tween® 20</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Viafect</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Xylene Cyanole</td>
<td>Sigma-Aldrich, USA</td>
</tr>
</tbody>
</table>

### 2.2 Instruments and Equipment

<table>
<thead>
<tr>
<th>Item</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>25 cm Cell Scrapers</td>
<td>Sarstedt, Inc., Germany</td>
</tr>
<tr>
<td>8 cm Petri Dishes</td>
<td>Sarstedt, Inc., Germany</td>
</tr>
<tr>
<td>25 cm² Tissue Culture Flasks, Cellstar®</td>
<td>Greiner Bio-One, Austria</td>
</tr>
<tr>
<td>75 cm² Tissue Culture Flasks, Cellstar®</td>
<td>Greiner Bio-One, Austria</td>
</tr>
<tr>
<td>500 mL Centrifuge Tubes</td>
<td>Beckman Coulter, USA</td>
</tr>
<tr>
<td>0.5 mL Microcentrifuge Tubes</td>
<td>SSI Bio (Interpath), AU</td>
</tr>
<tr>
<td>1.5 mL Microcentrifuge Tubes</td>
<td>Eppendorf®, Germany</td>
</tr>
<tr>
<td>15 mL Round Bottom Tubes, Cellstar®</td>
<td>Greiner Bio-One, Austria</td>
</tr>
<tr>
<td>50 mL Round Bottom Tubes, Cellstar®</td>
<td>Greiner Bio-One, Austria</td>
</tr>
<tr>
<td>0.2 mL PCR Tubes: Flat Top</td>
<td>Axygen Inc., USA</td>
</tr>
<tr>
<td>Rounded Top</td>
<td>Axygen Inc., USA</td>
</tr>
<tr>
<td>1 mL Syringes</td>
<td>Becton Dickinson, USA</td>
</tr>
<tr>
<td>Item</td>
<td>Supplier</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>6 Well Tissue Culture Plates</td>
<td>Greiner Bio-One, Austria</td>
</tr>
<tr>
<td>Autoclave</td>
<td>Thackray Graham Hart, AU</td>
</tr>
<tr>
<td>Centrifuge: Tabletop, IEC Micromax</td>
<td>Thermo Electron, USA</td>
</tr>
<tr>
<td>Large, Kontron TGA-6</td>
<td>Measuring and Scientific Equipment, UK</td>
</tr>
<tr>
<td>Tissue Culture, Orbital 900</td>
<td>Clements Medical Equipment, AU</td>
</tr>
<tr>
<td>ChemiDoc™ XRS+ with Image Lab™ Software</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>Cuvette</td>
<td>Sarstedt, Inc., Germany</td>
</tr>
<tr>
<td>Electrophoresis Mini-Sub DNA tank</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>Heating Plate</td>
<td>IEC, AU</td>
</tr>
<tr>
<td>Incubator (Water-Jacketed), Temperor</td>
<td>Techne, USA</td>
</tr>
<tr>
<td>Latex Gloves, Micro-Touch® (Size Small)</td>
<td>Ansell Healthcare, AU</td>
</tr>
<tr>
<td>Microscope, IMT-2</td>
<td>Olympus, AU</td>
</tr>
<tr>
<td>Mini-PROTEAN Tetra Cell Casting Module</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>Mini Trans-Blot Electrophoretic Transfer Cell</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>Motorised Pipetman: P1000</td>
<td>Gilson, France</td>
</tr>
<tr>
<td>P200</td>
<td>Gilson, France</td>
</tr>
<tr>
<td>P20</td>
<td>Gilson, France</td>
</tr>
<tr>
<td>P2</td>
<td>Gilson, France</td>
</tr>
<tr>
<td>Nalgene Filter, Rapid-Flow 50 mm</td>
<td>Thermo Fisher Scientific, AU</td>
</tr>
<tr>
<td>NanoDrop One</td>
<td>Thermo Scientific, USA</td>
</tr>
<tr>
<td>Neubauer Counting Chamber</td>
<td>Sigma-Aldrich, USA</td>
</tr>
<tr>
<td>Nitrocellulose Blotting Membrane, Amersham™ Protran™</td>
<td>GE Healthcare Life Sciences, USA</td>
</tr>
<tr>
<td>Parafilm</td>
<td>Bemis Company, Inc., USA</td>
</tr>
<tr>
<td>pH and Temperature Meter, lab CHEM-pH</td>
<td>TPS Pty Ltd, AU</td>
</tr>
<tr>
<td>Pipette Tips, Aerosol Barrier: P1000</td>
<td>Interpath Services, AU</td>
</tr>
<tr>
<td>P200</td>
<td>Interpath Services, AU</td>
</tr>
<tr>
<td>P20</td>
<td>Interpath Services, AU</td>
</tr>
<tr>
<td>Pipette Tips, Blue, P1000</td>
<td>Sarstedt, Inc., Germany</td>
</tr>
</tbody>
</table>
### 2.3 Commercial Kits

<table>
<thead>
<tr>
<th>Kit</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Expand High Fidelity PCR System</td>
<td>Roche Holding, Switzerland</td>
</tr>
<tr>
<td>Pierce™ BCA Protein Assay Kit</td>
<td>Thermo Fisher Scientific, USA</td>
</tr>
<tr>
<td>PureYield™ Plasmid Miniprep System</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>TGX Stain-Free™ FastCast™ Acrylamide, 12%</td>
<td>Bio-Rad, USA</td>
</tr>
<tr>
<td>QIAEX II® Gel Extraction Kit (500)</td>
<td>QIAGEN, Germany</td>
</tr>
<tr>
<td>QIAGEN PCR Cloning Kit (10)</td>
<td>QIAGEN, Germany</td>
</tr>
<tr>
<td>QIAquick® PCR Purification Kit (250)</td>
<td>QIAGEN, Germany</td>
</tr>
<tr>
<td>QuickChange® Lightning Site-Directed Mutagenesis Kit</td>
<td>Agilent Technologies, USA</td>
</tr>
<tr>
<td>Western Lightning® Plus-ECL Reagent</td>
<td>PerkinElmer, USA</td>
</tr>
<tr>
<td>Wizard® Plus SV Minipreps DNA Purification System</td>
<td>Promega, USA</td>
</tr>
</tbody>
</table>
### 2.4 Bacterial Cells

<table>
<thead>
<tr>
<th>Cell Strain</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Escherichia coli</em> (E. coli) XL-2</td>
<td>Dr B. Ward, Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital</td>
</tr>
<tr>
<td><em>Escherichia coli</em> (E. coli) XL-2 Supercompetent Cells</td>
<td>Stratagene, USA</td>
</tr>
</tbody>
</table>

### 2.5 Cell Lines – Human

<table>
<thead>
<tr>
<th>Cell Line</th>
<th>Culturing Media</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>A2780</td>
<td>DMEM</td>
<td>Prof. AM Dharmarajan, Department of Anatomy and Human Biology, University of Western Australia</td>
</tr>
<tr>
<td>HEK293</td>
<td>DMEM</td>
<td>Prof K Eidne, Harry Perkins Institute of Medical Research, University of Western Australia</td>
</tr>
</tbody>
</table>

### 2.6 Primers

<table>
<thead>
<tr>
<th>Primers</th>
<th>Type</th>
<th>Sequence</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>FLAGGDF9-F</td>
<td>PCR</td>
<td>5’ GATATCGACTACAAAGACGATGACGACAAGATGGCACGTCCCAAC 3’</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td>FLAGGDF9-F2</td>
<td>PCR</td>
<td>5’GATATCATGGACTACAAAGACGATGACGACAAGGCACGTCCCAAC 3’</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td>GDF9-R</td>
<td>PCR</td>
<td>5’CTCGAGTCAACGACAGGTGACCTTTG 3’</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td>GDF9Int-F</td>
<td>Sequencing/</td>
<td>5’ CAGATTGATGTGACCAG 3’</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td>GDF9Int-R</td>
<td>Sequencing/</td>
<td>5’ GAGCACTTGTGTCATTC 3’</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td></td>
<td>Plasmid</td>
<td>Supplier</td>
<td></td>
</tr>
<tr>
<td>----------------</td>
<td>---------</td>
<td>------------------------------------------------------------------------------------------------------------------------------------------</td>
<td></td>
</tr>
<tr>
<td>GDF9-783delC-F</td>
<td>SDM</td>
<td>5’ GACTCTGGTGTCCTCCCCTCAC TGATCTTATATTTG 3’</td>
<td></td>
</tr>
<tr>
<td>GDF9-783delC-R</td>
<td>SDM</td>
<td>5’ CAAATATAAGATCAGTGAG GGGACACCAGAGTC 3’</td>
<td></td>
</tr>
</tbody>
</table>

### 2.7 Plasmid Constructs

<table>
<thead>
<tr>
<th>Plasmid</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>pcDNA3.1</td>
<td>Provided by Dr Bryan Ward, Department of Endocrinology and Diabetes, Sir Charles Gairdner Hospital, WA</td>
</tr>
<tr>
<td>pDrive</td>
<td>QIAGEN, Germany</td>
</tr>
</tbody>
</table>

### 2.8 GDF9 Plasmid Template DNA

<table>
<thead>
<tr>
<th>GDF9</th>
<th>Plasmid</th>
<th>Vector</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Human GDF-9 Gene cDNA clone plasmid (10 μg)</td>
<td>pMD-GDF9 (NM_005260.3)</td>
<td>pMD18-T Simple</td>
<td>Sino Biological Inc. Beijing</td>
</tr>
</tbody>
</table>

### 2.9 Antibody

<table>
<thead>
<tr>
<th>Type</th>
<th>Antibody</th>
<th>Dilution</th>
<th>Supplier</th>
</tr>
</thead>
<tbody>
<tr>
<td>Primary</td>
<td>Anti-GDF9 antibody, rabbit polyclonal to GDF9 (N terminal)</td>
<td>1:500</td>
<td>Abcam, UK</td>
</tr>
<tr>
<td>Primary</td>
<td>Anti-FLAG monoclonal, mouse</td>
<td>1:5,000</td>
<td>Sigma-Aldrich, US</td>
</tr>
<tr>
<td>Secondary</td>
<td>Goat anti rabbit-HRP conjugate</td>
<td>1:5,000</td>
<td>Promega, USA</td>
</tr>
<tr>
<td>Secondary</td>
<td>Goat anti-Mouse</td>
<td>1:10,000</td>
<td>Sigma, USA</td>
</tr>
<tr>
<td>Primary</td>
<td>β-actin, Mouse mAb</td>
<td>1:5,000</td>
<td>Cell Signaling Technology, USA</td>
</tr>
</tbody>
</table>
Chapter 3

General Methods
3 METHODS

3.1 Bacterial Techniques

3.1.1 Bacterial Cultures
Bacterial cultures were prepared by inoculating one colony of transformed cells (E. coli XL-2 or E. coli XL-2 Supercompetent Cells) in 5 mL of 2xYT medium containing 50 µg/mL ampicillin and left on a shaker (225 rpm) overnight at 37°C.

3.1.1.1 Agar Plates
Agar plates for the propagation of single bacterial colonies were prepared by either spreading an appropriate volume of bacterial suspension or by streaking out of bacteria onto 2xYT agar plates containing ampicillin (50 µg/mL).

3.1.2 Transformation

3.1.2.1 Heat Shock Transformation
3.1.2.1.1 Transformation of Mutagenesis Products
E. coli XL-2 supercompetent cells stored at -70°C were thawed on wet ice for ~5 minutes. 15 mL round-bottom tubes were pre-chilled on ice and 200 µL aliquots of E. coli XL-2 supercompetent cells added to each tube. DpnI digested PCR products (5 µL) were aliquoted into the appropriate tubes containing the cells. The transformation reactions were gently mixed and incubated on ice for 30 min. SOC medium was prepared from supplied and sterile SOB by adding 1 M glucose, 1 M magnesium chloride, and 1 M of magnesium sulphide to final concentrations of 0.02 M for glucose and 0.02 M for Mg²⁺ ions, and then placed in a 42°C water bath. Following the 30 min incubation on ice, the tubes were heat-shocked in the 42°C water bath for 90 seconds and then placed back on ice for a further 2 min. 800 µL of the pre-heated SOC medium was added to each reaction and the tubes incubated at 37°C with shaking at 225 rpm for 1 hour. 200 µL, 20 µL, and 2 µL of each transformation mix were spread onto each of three 2xYT ampicillin agar plates that were left to absorb the bacterial suspension for 30 min and then inverted and incubated at 37°C overnight.
3.1.2.1.2 Transformation of Subcloned Constructs

*E. coli* XL-2 cells were thawed on ice and 5 µL ligation-reaction mix or 10 µL plasmid DNA (5ng/µL) was added to 200 µL cells in 15 mL round bottom tubes. The reactions were incubated on ice for 30 min before being heat-shocked at 42°C for 90 seconds and then returned to incubate on ice for 2 min. 800 µL warm 2xYT or SOC medium was added to each tube and then the reactions were incubated with shaking at 225 rpm for 1 hour at 37°C. 200 µL, 20 µL, and 2 µL of each transformation mix was spread onto each of three 2xYT ampicillin agar plates and incubated at 37°C overnight.

For transformations with dephosphorylated vector 200 µL, 100 µL, and 50 µL of each transformation was spread instead.

3.1.3 Preparation of *E. coli* XL2 Competent Cells

An overnight culture of *E. coli* XL2 cells was grown in 2xYT broth containing 50 µg/mL tetracycline and then 5 mL inoculated in 500 mL 2xYT. The optical density (OD) was measured at regular intervals at a wavelength of 600 nm until an OD of 0.5-0.7 was reached. The cells were then pelleted in 500 mL centrifuge tubes (4000, 10 min, 4°C), resuspended in 50 mL ice-cold 100 mM MgCl₂, incubated on ice for 20 min, and pelleted in 50 mL tubes with the above settings and the supernatant decanted. The same procedure was then carried out using 50 mL ice-cold 100 mM CaCl₂. Both pellets were resuspended in a final total volume of 22 mL ice-cold 14% glycerol / 100 mM CaCl₂, and aliquoted out in 1.5 mL Eppendorf tubes before being snap-frozen in liquid nitrogen and stored at -70°C for use.

3.2 DNA Techniques

3.2.1 DNA Preparation

3.2.1.1 Extraction of Plasmid DNA from Bacterial Cells

3.2.1.1.1 DNA Wizard Miniprep

Plasmid DNA from bacteria was routinely extracted from overnight cell cultures using the Wizard® Plus SV Minipreps DNA Purification System following their Quick Protocol; Centrifugation Protocol, using ~5 mL culture pelleted per column. DNA was eluted from the column in 30 µL ultra-pure ddH₂O.
3.2.1.2 DNA PureYield Miniprep

The PureYield™ Miniprep System was used when high-quality DNA was needed for sequencing and/or transfection. Extraction was performed following their Quick Protocol; Centrifugation Protocol and using their Alternative Protocol for Larger Culture Volumes.

3.2.1.2 Extraction of DNA from Agarose Gel

The appropriate bands on the agarose gel were excised under UV light using a sterile scalpel blade and ~200-250 mg of gel allocated to each Eppendorf tube. The QIAEX II® Gel Extraction Kit (500) was used together with the QIAEX® II Handbook for DNA extraction from agarose and polyacrylamide gels and desalting and concentrating DNA from solutions. The eluates were combined into one single tube and stored at -20°C until future use.

3.2.1.3 DNA Purification

3.2.1.3.1 QIAquick Purification

The QIAquick® PCR Purification Kit was mainly used for purification of restriction digests in between and after double digests and occasionally for the clean-up of other DNA products. The QIAquick PCR Purification Kit Protocol using a microcentrifuge was followed with final elution in 30 µL and the eluates combined. Storage was at -20°C until future use.

3.2.1.4 Ethanol Precipitation

Ethanol precipitation with salt was used when a high concentration of DNA was needed. The volume of DNA solution was measured using a single channel pipette and then the appropriate quantities of 3 M sodium acetate and 100% ethanol was added for a final ethanol concentration of 70% and a final salt concentration of 0.3 M. 2 µg salmon sperm was added to act as a carrier for the DNA, and the solution briefly vortexed and centrifuged at 15000 for 10 min. The supernatant was discarded carefully, and the pellet washed gently in 1 mL 70% ethanol to not dislodge the pellet. If the pellet became loose then the tube was re-centrifuged for 10 min. The supernatant was discarded as before, and every trace of ethanol removed by drying for 30 min in a fume hood. The pellet was finally resuspended in 4 µL ddH₂O.
3.2.2 DNA Analysis

3.2.2.1 Agarose Gel Electrophoresis
DNA was typically run on a 1% (w/v) agarose gel with 0.5-1 µg of DNA was usually loaded per well alongside a 1 Kb Plus DNA ladder used for size determination. The agarose gels were run in a DNA electrophoresis mini-sub DNA tank in 1x TAE buffer at 110V. The DNA was visualised on a ChemiDoc™ XRS+ Imager (Bio-Rad) and a picture taken using the Image Lab™ Software (Bio-Rad). For the gel visualisation of gel bands, either ethidium bromide (0.002%) or GelRed (0.0017%) was added to the gel before pouring.

3.2.2.2 DNA Quantification
3.2.2.2.1 NanoDrop
DNA and protein were quantitated using a NanoDrop One spectrophotometer. Calibration was generally performed using 1 µL ddH₂O or the appropriate elution buffer of the sample, and each sample was analysed in duplicate and the average used to determine the concentration of DNA in each sample. DNA was analysed at wavelengths A260/280nm and A260/230nm for purity and contamination, and the concentration given as ng/µL.

3.2.3 DNA Sequencing
3.2.3.1 Australian Genome Research Facility
3.2.3.1.1 Sequencing of Plasmid DNA
Sequencing of plasmid DNA was carried out at the Australian Genome Research Facility (AGRF) located at the Harry Perkins Institute of Medical Research using Sanger sequencing and the option of ‘force all base-calls’ was used as the sequence read-out. The AGRF Quick Reference Guide for Purified DNA was used and each reaction contained 1.0-1.5 µg DNA and 9.6 pmol sequencing primer in a final volume of 12 µL.
3.2.4 PCR Reactions

3.2.4.1 General PCR

The following settings were used on the thermocycler for routine PCR reactions. Generally, a reaction volume of 20-25 µL and heated lids were used.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Activation Step</td>
<td>95</td>
<td>5min</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>30sec</td>
</tr>
<tr>
<td>Annealing</td>
<td>60</td>
<td>30sec</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1.5min</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72</td>
<td>5min</td>
</tr>
</tbody>
</table>

Each reaction typically contained 5 µL 5X Taq polymerase buffer, 2 µL of 25 mM MgCl₂, 1 µL of 5 mM dNTPs, 2.5 µL of 5 µM each of forward and reverse primer, and 0.2 µL of Taq polymerase, made up to a final volume of 25 µL with ddH₂O. Amount of template DNA used was dependent on the purpose of the specific PCR and generally ranged from 5-100 ng.

3.2.4.2 Expand High Fidelity PCR

The Expand High Fidelity PCR System from Roche was used for the initial PCR reaction to increase the copy number of GDF9 plasmid template DNA. The following settings were used with a reaction volume of 50 µl and heated lids.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Activation Step</td>
<td>94</td>
<td>2min</td>
</tr>
<tr>
<td>Denaturation</td>
<td>94</td>
<td>15sec</td>
</tr>
<tr>
<td>Annealing</td>
<td>60</td>
<td>30sec</td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>1.5min</td>
</tr>
<tr>
<td>Final Extension</td>
<td>72</td>
<td>7min</td>
</tr>
</tbody>
</table>

The first reaction mix contained 2 µL of 5 mM dNTPs, 3 µL of 5 µM each of forward and reverse primer, and 15 ng template DNA, made up to 25 µL with ddH₂O. The second reaction mix contained 5 µL of 10X Expand High Fidelity Buffer, 4 µL of 25 mM MgCl₂, and 0.75 µL Expand High Fidelity Enzyme, made up to 25 µL with ddH₂O. The two reaction mixes were combined immediately prior to PCR cycling for a final total volume of 50 µL.
3.2.4.2.1 Addition of FLAG-tag to the N-Terminus of GDF9

The FLAG-tag (DYKDDDDK) was added to the N-terminus of the GDF9 gene (see primer table in Materials section) by incorporating the sequence encoding the FLAG-tag into my forward PCR primer. The FLAG-tag was to be used for detection of FLAG-tagged GDF9 in mammalian cell expression studies.

3.2.4.2.2 Addition of Restriction Enzymes to N and C-Terminus of GDF9

In addition to the FLAG tag the restriction enzyme site for EcoRV was added to the forward PCR primer and the restriction enzyme XhoI was added to the reverse PCR primer for subsequent subcloning from pDrive into the pcDNA3.1 vector.

3.2.5 PCR Colony Screening

PCR colony screening was routinely performed with bacteria colonies growing on plates after transformation reactions to screen for successful incorporation of my DNA into the appropriate vector. Each reaction typically contained 5 μL 5X Taq polymerase buffer, 2 μL of 25 mM MgCl₂, 1 μL of 5 mM dNTPs, 2.5 μL of 5 μM each of forward and reverse primer, and 0.2 μL of Taq polymerase, made up to a final volume of 25 μL with ddH₂O. Single bacterial colonies, enough to cover the opening of a yellow pipette tip, were added to the respective thin-walled 0.2 mL PCR tubes containing the PCR reaction. Mixing by vigorous pipetting was required to lyse the bacterial cells. A positive control from a streak-out plate known to contain the target sequence was also prepared, as well as a negative control using a colony from the negative control transformation plate. The general PCR settings (see section 3.2.4.1.) was used for all PCR colony screenings with a reaction volume of 25 μL. Products were analysed by agarose gel electrophoresis (see section 3.2.2.1).

3.2.6 DNA Modifications

3.2.6.1 Site-Directed Mutagenesis

Site-directed mutagenesis (SDM) was performed to generate the mutant DNA (GFD9 c.783delC) using the QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies). Oligonucleotide primers were designed according to their instruction manual and synthesised to contain the specific nucleotide deletion.
The mutant strand synthesis reaction was set up according to the manufacturer’s protocol and contained 5 µL 10X reaction buffer, 85 ng template DNA, 1 µL of dNTP mix, 1.5 µL Quick Solution Reagent, and 125 ng each of forward and reverse primers, made up to 50 µL with ultra-pure ddH2O. 1 µL of Quick Change Lightning Enzyme was then added to the mix. The cycling conditions included an initial denaturation cycle at 95°C for 2 minutes, followed by 18 cycles of 95°C for 20 seconds, 60°C for 10 seconds, and 68°C for 30 seconds per Kb of plasmid length (2 minutes and 36 seconds, pDrive-GDF9), and a final extension cycle of 68°C for 5 minutes. The PCR products were then cooled to 4°C and stored at -20°C.

3.2.6.1 DpnI Digestion of Parental Strand
Following SDM the parental (methylated) DNA was digested with DpnI enzyme. 2 µL of DpnI enzyme was added to the 50 µL sample and digested at 37°C for 1.5 hours with mixing and spin down of reaction multiple times to ensure complete digestion.

3.2.6.2 Restriction Enzyme Digests
Restriction enzyme digests were routinely performed as part of subcloning and when analysing DNA products on an agarose gel. Both the pDrive-GDF9 and pcDNA3.1 vector DNA were cut with the restriction enzymes EcoRV and XhoI using NEBuffer3.1 for subsequent subcloning of GDF9 insert into the pcDNA3.1 vector. The restriction enzyme digests were generally set up as 20 µL reactions containing a minimum of 500 ng DNA sample, 1x NEBuffer3.1, 5 U/µg of each enzyme, and the appropriate volume of ddH2O. At times, plasmid DNA was simply linearized for analysis on agarose gel with either EcoRI (in 1x NEBuffer2.1) or EcoRV (in 1x NE3uffer3.1).

The reactions were incubated at 37°C for 1-2 hours and the samples mixed by pipetting and spun down every 15 minutes. Storage was at -20°C.
3.2.6.3 Dephosphorylation of 5’-Ends

pcDNA3.1 vector DNA was dephosphorylated with alkaline phosphatase to prevent religation when subcloning. Each sample for dephosphorylation was made up to 40.5 µL with ddH₂O, then 4.5 µL of 10x calf intestinal alkaline phosphatase (CIAP) buffer and 2.5 µL of 1:5 CIAP (in 1x CIAP buffer) was added, and incubated at 37°C for 30 min. Following the incubation, another 2.5 µL of 1:5 CIAP was added and the reaction incubated for a further 37°C for 30 min. To stop the reaction 2 µL 50mM EDTA (pH 8.0) was added to the sample and incubated at 75°C for 15 min. The reactions were then cleaned up using the QIAquick® PCR Purification Kit with a final elution in 25 µL ddH₂O.

3.2.7 Ligation Reactions

3.2.7.1 Ligation into pDrive

3.2.7.1.1 TA Cloning

TA cloning, also commonly known as UA cloning or PCR cloning, was used for the expression of GDF9 PCR product in pDrive cloning vector. The TA cloning was performed using the QIAGEN PCR Cloning Kit and following the QIAGEN PCR Cloning Handbook, and the following equation was used for calculating the amount of PCR product required:

\[
\text{ng PCR product required} = \frac{\text{Cloning Vector (ng) x PCR Product Size (bp) x Molar Ratio}}{\text{Vector Size (bp)}}
\]

Usually, 50 ng of the pDrive vector was used with a molar ratio of 5. The PCR product size was 1362 bp for the WT and 1361 bp for the mutant, the pDrive vector size was 3851 bp. Controls containing no insert DNA were also prepared for the WT and mutant.

3.2.7.2 Ligation into pcDNA3.1

Ligation of GDF9 insert into pcDNA3.1 was achieved by engineering the restriction endonucleases (enzymes) EcoRV onto the forward PCR primer and XhoI into the reverse primer. pcDNA3.1 contains these restriction endonucleases cutting sites in its multiple cloning site (MCS) which allowed for the directional ligation of insert (GDF9) into pcDNA3.1. The same equation used for the TA cloning was also used for the ligation into pcDNA3.1:

\[
\text{ng PCR product required} = \frac{\text{Cloning Vector (ng) x PCR Product Size (bp) x Molar Ratio}}{\text{Vector Size (bp)}}
\]
Generally, 50 ng of the pcDNA.3.1 vector was used with a molar ratio of 5 or 8. The PCR product size was 1362/1361 bp, and the pcDNA3.1 vector size was 5428 pb. Controls containing no DNA insert were also prepared.

3.3 Tissue Culture

3.3.1 Resuscitation of Mammalian Cell lines

Cells were resuscitated from liquid nitrogen using a quick thaw method. Aliquots of cells that were stored in liquid nitrogen were quickly thawed in a 37°C water bath. The thawed cells were added to 4 mL warmed DMEM media (+10% FCS, 1% Pen/Strep) in a 15 mL round bottom tube and centrifuged at 9000 for 2 min at room temperature. Supernatant containing DMSO (dimethyl sulfoxide) cell culture reagent for cryopreservation and DMEM media was removed via aspiration, and the pellet resuspended in 1 mL fresh DMEM medium (+10% FCS, 1% Pen/Strep) and added to a 25 cm² tissue culture flask containing 4 mL warm DMEM media (+10% FCS, 1% Pen/Strep). The cells were gently mixed with the medium and incubated at 37°C with 5% CO₂.

The media was changed after 2-3 days and the cells trypsinised and transferred to a 75 cm² tissue culture flask when the confluency reached approximately 90%.

3.3.2 Maintenance of Mammalian Cell Lines

Human cell lines were generally maintained in DMEM medium with 10% foetal bovine serum and 1% penicillin/streptomycin (10,000 U/mL / 10 mg/mL) in a water-jacketed incubator at 37°C with 5% CO₂. DMEM medium was generally replaced every 2 or 3 days and re-passaged once a week.

Cells were generally grown to 80-90% confluency and then trypsinised, either to be further passaged or to be seeded out for experimental work. The old medium was aspirated, and the cells washed with 3 mL of pre-warmed 1X PBS. The PBS was then aspirated and 3 mL of 2-5X trypsin (pre-warmed to 37°C) added and the flask incubated at 37°C for 3 minutes. To fully dislodge the cells, the bottom of the flask was firmly tapped against a flat surface.
To inactivate the trypsin, 4 mL of pre-warmed DMEM medium was added and vigorously pipetted to individualise the cells. The cell suspension was collected and transferred into a 15 mL round bottom tube and centrifuged at 900 for 2 minutes at room temperature. The supernatant was aspirated, and the cells typically resuspended in 4 mL of fresh DMEM medium (37°C). The cells were then either counted in a counting chamber and seeded out into plates for experiments, or a small amount, typically 50 µL, transferred to a new flask for cell maintenance.

3.3.2.1  A2780
The human ovarian epithelial cell line, A2780, was maintained in DMEM medium with 10% foetal bovine serum and 1% penicillin / streptomycin (10,000 U/mL / 10 mg/mL).

3.3.2.2  HEK293
The human embryonic kidney cell line, HEK293, was maintained in DMEM medium with 10% foetal bovine serum, 1% penicillin/streptomycin (10,000 U/mL / 10 mg/mL).

3.3.3  Transfection
3.3.3.1  Counting/Seeding of Mammalian Cell Lines
The cells were trypsinised as above and typically resuspended in 7 mL pre-warmed DMEM media and 50 µL of cell suspension added to a Neubauer counting chamber (25 µL for each chamber). Cells were counted for 16 squares in each chamber and the average number of cells per mL determined. The total number of cells required for the seeding was calculated and from this, the amount of cell suspension to be used was determined.

For the A2780 cell line, 1x10⁶ cells per well were typically seeded into 6-well plates and incubated for 4 hours before transfection. Following transfection, the 6-well plates were incubated overnight. For the HEK293 cell line, 5x10⁶ cells per flask were seeded into either 25cm² or 75cm² flasks and incubated overnight to grow to a confluency of 50-60% and then transfected with Lipofectamine 2000 at a ratio of 2:1. Following transfection of HEK293 cells, the flasks were incubated for 7 hours, then pooled and seeded into 25cm² flasks at 2x10⁶ cells per flask and incubated overnight. DMEM media was added to a total volume of 2 mL for each well of the 6-well plate for A2780 and each 25cm² flask for the HEK293 cell
line, for this overnight incubation with DMEM containing 1% pen/strep included in the DMEM for the HEK293 cells but not in that for the A2780 cells.

Media for all plates/flasks were changed after this first overnight incubation, and fresh DMEM media (10% FBS, 1% p/s) added to each well (2 mL) and flask (5 mL). All incubations were at 37°C with 5% CO₂.

3.3.3.2 Transfection of Wildtype and Mutant GDF9 Constructs into HEK293 Cells using Lipofectamine 2000

HEK293 cells were seeded into 25cm² tissue culture flasks at 2x10⁶ cells per flask, each flask was used for only one condition, and the transfection of cells occurred on the following day. The ratios of Lipofectamine to DNA were 2:1 for all conditions and the amount of DNA was 2 µg, 4 µg, and 6 µg each for pcDNA3.1-GDF9 wildtype, mutant, FLAG-tagged mutant, and empty vector. For transfection, the DNA for wild type, mutant, and empty vector were each pooled for a total of 12 µg of DNA each, this was diluted in 600 µL OPTIMEM. 96 µL of Lipofectamine was diluted in 2.5 mL OPTIMEM and 600 µL of this diluted Lipofectamine was added to each tube containing 12 µg of DNA and left to incubate for 15 minutes at room temperature. Media from the tissue flasks were aspirated and the cells washed twice with 2 mL 1X PBS. Following the incubation, 200 µL of the appropriate DNA/Lipofectamine mix was added to the flasks for transfection of 2 µg DNA, 400 µL DNA/Lipofectamine mix to the flask for 4 µg DNA, and 600 µL DNA/Lipofectamine mix for transfection of 6 µg DNA. OPTIMEM was added to each flask for a total volume of 1 mL, and finally, 1 mL of DMEM (no antibiotics) was added to each flask. The cells were incubated for 7 hours at 37°C with 5% CO₂, 2 mL of fresh DMEM was then added and the cells further incubated at 37°C with 5% CO₂ for approximately 40 hours before being harvested. The rest of the procedure for transfection into the HEK293 cell line was as per section 3.3.3, with the only change of using 150 µL cell lysate buffer instead of the typical 100 µL.
3.3.4 Cell Harvesting

Cells were typically harvested 48 hours after transfection. Fresh DMEM media containing 1% p/s and 10% FCS was added 24 hours after transfection. After 48 hours the media was aspirated, and cells gently washed twice with pre-warmed 1X PBS, 1 mL for cells in 6-well plates, and 2 mL for 25cm² flasks. Following the final washing, the PBS was thoroughly aspirated, and the cells immediately placed on ice.

3.3.5 Collection of Cell Lysates

10 mL RIPA buffer with one protease inhibitor tablet dissolved was used for the cell lysis. For highly confluent cells (~100%) 100 µL of RIPA buffer was used per well in a 6-well plate, 80 µL for 80% confluency and so on. 150 µL was used for per 25cm² flask. The RIPA buffer was added dropwise with a pipette and the plates/flasks incubated for 5 minutes on a shaker at room temperature. Following the incubation, the cells were collected in 1.5 mL Eppendorfs kept on ice using a cell scraper and a yellow tip pipette. The cell scraper was thoroughly cleaned with 70% ethanol in between scraping cells from each well/flask, and a separate cell scraper was used for each transfection reagent. After collection of cells, all tubes were methodically vortexed, and a precision 25-gauge glide needle and 1 mL syringe were used to suck up and down the contents of each tube 10 times to further lyse the cells. The tubes were then centrifuged for 35 minutes at 15,000 at 4°C and the cell lysate supernatant carefully collected in new Eppendorfs on ice.

3.4 Protein Analysis

3.4.1 Protein Quantitation Using the Bicinchoninic Acid Assay

Protein quantitation was determined using the BCA (Bicinchoninic Acid Assay) Protein Assay Kit from Thermo Scientific and standards for all protein quantitation were made up according to table 3.1.
Table 3.1. BCA protein standards.

<table>
<thead>
<tr>
<th>Standard</th>
<th>Concentration (µg/µL)</th>
<th>Diluent (µL)</th>
<th>BSA (µL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2,000</td>
<td>-</td>
<td>10</td>
</tr>
<tr>
<td>B</td>
<td>1,500</td>
<td>7.5</td>
<td>22.5</td>
</tr>
<tr>
<td>C</td>
<td>1,000</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>D</td>
<td>750</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>E</td>
<td>500</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>F</td>
<td>250</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>G</td>
<td>125</td>
<td>15</td>
<td>15</td>
</tr>
<tr>
<td>H</td>
<td>25</td>
<td>24</td>
<td>6</td>
</tr>
<tr>
<td>I</td>
<td>0</td>
<td>15</td>
<td>-</td>
</tr>
</tbody>
</table>

RIPA buffer (with protease inhibitor) was used as the diluent and added to all standards according to table 3.1. Bovine Serum Albumin (2 mg/mL, diluent; RIPA) was added to standards A, B, and C, and these were then used to make up standards D to H. From the BCA kit, the appropriate amounts of reagent A and reagent B were mixed according to the manufacturer, generally 7 mL of reagent A and 2% of reagent B was used, and 200 µL of this solution added to new tubes labelled A to I, and one additional tube. 10 µL of each standard was then transferred to the corresponding new tubes.

The one additional tube containing 200 µL (reagent A and B mix) was used as a test to ensure the appropriate dilution of the lysates was made so that the concentration would fall within the standards. Generally, a 1 in 5 dilution of lysate, using RIPA as the diluent, was appropriate and 10 µL of diluted lysate added to the tube.

Contents of all tubes were thoroughly mixed by vortexing and incubated at 37°C for 30 minutes. After 10 minutes of incubation, the colour change of the standards were compared to the additional test tube and any adjustments to the dilution ratio made if necessary. All protein dilutions were made in duplicates. Following incubation, the tubes were stored at 4°C until use.

Standards were analysed, and protein quantitated at 562 nm using the BCA software on the NanoDrop. RIPA buffer was used as the blank and the standard ‘I’ was used as the reference.
sample. Results for the protein quantitation was analysed in Microsoft Excel and calculations made for the loading of 100 µg protein of each sample in an SDS-PAGE gel.

3.4.2 SDS-PAGE

Protein samples were separated using Sodium Dodecyl Sulfate Polyacrylamide Gel Electrophoresis (SDS-PAGE). The separating gel was prepared using the TGX Stain-Free™ FastCast™ Acrylamide kit and a concentration of 12% was made. The gel was loaded using the Mini-PROTEAN Tetra Cell Casting Module (Bio-Rad), typically using the 1.5 mm cast. Immediately following loading of the separating gel, the stacking gel was gently loaded on top in an even layer using a transfer pipette and the comb inserted carefully to avoid bubbles, and the gel left to set for 30 minutes. The comb was then carefully removed, any excess gel washed off, and the gel placed into the Mini-PROTEAN Tetra Cell and Tank apparatus, with the appropriate amount of 1X SDS-PAGE running buffer added.

The protein samples were typically prepared and run as 100 µg with 6X SDS-PAGE loading buffer in a final volume of 20-30 µL, made up with RIPA buffer. Before loading of samples, a 10-minute incubation in boiling water was performed to fully denature the protein. For size determination, the Precision Plus Protein Dual Colour Standards was loaded neat (5 µL) and run alongside the protein samples. The gel was first run at 80V until past the stacking buffer and then at 180V until the 25 kDa marker of the protein dual colour standard was at the bottom of the gel.

3.4.3 Membrane Transfer

All membrane transfers were conducted using the Mini Trans-Blot Electrophoretic Transfer Cell (Bio-Rad). Following SDS-PAGE, the gels were disassembled and trimmed by removal of the stacking gel. The separating gel was then carefully sandwiched into the cassette with the outermost layer being a fluffy pad on either side, then chromatography paper (Whatman) on the inside of each fluffy pad. The gel was placed in the middle of this sandwich towards the cathode side of the cassette and the nitrous cellulose blotting membrane (Amersham) placed on top of the gel (towards the anode side of the cassette). During the construction of the sandwich, all components were soaked in 1X membrane transfer buffer, and the
sandwich finally placed into the tank containing the same buffer. The transfer was typically run at 30V at 4°C overnight, or at 70V for 1 hour at 4°C.

### 3.4.4 Western Analysis

For the western analysis several optimization experiments were performed. However, the initial western analysis was based on the lab's own in-house protocol for FLAG antibody detection. The membranes were blocked with 1X TBS-T + 5% Bovine Serum Albumin (BSA) for 2 hours at room temperature. The primary antibody (GDF9 polyclonal antibody) was then added at 1:500 dilution in 1X TBS-T + 10% BSA and incubated overnight at 4°C. Following the overnight incubation, the membrane was washed three times with 5 mL TBS-T for 10 minutes each wash and then incubated with secondary antibody (goat anti-rabbit-HRP conjugate) at 1:5,000 dilution in 1X TBS-T + 10% BSA for 1 hour at room temperature. The membrane was again washed three times with 5 mL TBS-T for 10 minutes each wash and then incubated with 1 mL ECL reagent (PerkinElmer) for 1 minute before being imaged using the ChemiDoc Imager (Bio-Rad). All membranes were analysed using the above approach unless otherwise stated.

### 3.4.5 ECL Detection

Western blot protein-membrane detection was performed using the ECL detection kit (PerkinElmer). 500 µL of each reagent was mixed, and the membrane incubated in the solution for 1 minute at room temperature. Detection of chemiluminescence was achieved using the ChemiDoc imager.

### 3.4.6 Membrane stripping

Mild membrane stripping was performed before reprobing with beta-actin. The stripping of membranes was necessary due to the relatively high degree of non-specific bands on the blots after initial western analysis antibody probing. The mild stripping buffer was made with 1.5% glycine, 0.1% SDS, 1% Tween-20 made up to a total volume of 1 litre with ddH₂O, and the pH adjusted to 2.2. The membrane was incubated at room temperature on a shaker with volumes sufficient to cover the membrane.
The membrane was incubated two times with the mild stripping buffer for 10 minutes, washed twice with 1X PBS for 10 minutes, and then washed twice with 1X TBS-T for 5 minutes. The membrane was then blocked as per procedure below (3.4.6).

3.4.7 Beta Actin

Beta-actin (β-actin), a housekeeping protein, was used as a loading control for all Western blot membranes for proper interpretation of the blots. It was used to normalize the levels of protein detected by determining the evenness of loading of protein across the gel. All membranes were stripped as described above prior to probing with beta-actin.

The membrane was blocked with 1X TBS-T + 3% SMP for 1 hour at room temperature. The beta-actin primary antibody was then added at 1:5,000 dilution in 1X TBS-T + 3% SMP and incubated for 1 hour at room temperature. Following incubation, the membrane was washed three times with 15 mL 1X TBS-T for 5 minutes each wash, and then incubated with goat anti-mouse, secondary antibody, at 1:10,000 dilution in 1X TBS-T + 3% SMP for 1 hour at room temperature. The membrane was washed again three times with 15 mL 1X TBS-T for 5 minutes each wash and detected using the ECL reagent and ChemiDoc imaging.

3.4.7.1 Beta Actin Optimization

The initial probing with beta-actin was conducted following the manufacturers recommended dilution of 1:1,000 and using the above-described method. The dilution was then adjusted to 1:5,000.
Chapter 4

Clinical Recruitment

&

Exome Sequencing for the Identification of Putative Causal Mutations in PCOS
4 CLINICAL RECRUITMENT AND EXOME SEQUENCING FOR THE IDENTIFICATION OF PUTATIVE CAUSAL MUTATIONS IN PCOS

4.1 Introduction
The Keogh Institute at Sir Charles Gairdner Hospital is the centre of the Western Australian Database study of PCOS women. The database was put together in 2004 by Prof. Bronwyn Stuckey and Prof. Scott Wilson. One of the aims of this study was to recruit at least one family for samples to be processed and added to this database.

Whole exome sequencing is an approach to determine the sequence of the gene coding regions of the genome, which constitutes 2% of the genome. Pathogenic mutations are present within these parts of the genome and with the aid of exome sequencing mutations like these can be efficiently uncovered on a large scale and for a significantly cheaper cost than is the case for whole genome sequencing, sanger sequencing and functional approaches. Exome sequencing is therefore ideal for screening for potential disease-causing functional variants in PCOS.

4.2 Materials and Methods
4.2.1 Clinical Recruitment
Ethics approval was granted and patient recruitment for this study was undertaken through the Keogh Institute of Medical Research, WA, with the aid of my supervisor Professor Bronwyn Stuckey and Dr Jennifer Ng. Suitable patients were identified during normal medical consultations. Individuals eligible for enrolment in the study were females diagnosed with PCOS who had at least one other first-degree relative (e.g. mother, sister, or daughter) diagnosed with PCOS and willing to participate. After obtaining an initial expression of interest from a patient for participation in the research, contact details were made available to the Honours student to attempt to progress recruitment into the study. The recruitment process involved calling the patients directly and asking if they were willing to participate in the PCOS study at the Keogh Institute by donating a small sample of their blood and receive, complete and submit a few simple forms and a short medical questionnaire. These papers included a participant information sheet and consent form (consent for blood taking and DNA banking for clinical research), a PCOS questionnaire, and
a PathWest blood collection form to bring when donating their blood either at any of the PathWest centres or at the Keogh Institute. All paperwork for the recruitment process can be found in Appendix II.

4.2.2 Extraction of Genomic DNA from Patient Blood Samples

The patient blood samples were collected through the PathWest state-wide specimen collection network. Specimens received at PathWest were transferred through the pneumatic tube system to the Department of Endocrinology and Diabetes laboratory, Sir Charles Gairdner Hospital, laboratory for analysis. Each sample contained 1x 6 mL clotted whole blood tube and 2x 9 mL K3-EDTA whole blood. The clotted whole blood sample was prepared by centrifugation at 2400 rpm (10 min, 4°C) and the plasma supernatant transferred to 1x 5 mL sterile vials and stored at -20°C for potential future use. The 2x 9 mL samples were aliquoted into 5x5 mL vials and stored at -20°C for DNA extraction and analysis.

DNA was extracted using 200 µL patient whole blood (K3-EDTA) using the QIAamp® DNA Blood Mini Kit (50) and following the manufacturer’s instructions. DNA was eluted in 100 µL ultra-pure ddH₂O for increased DNA concentration. The quantity and quality of the DNA in the sample was determined using a NanoDrop Spectrophotometer. The ratio of absorbance at 260 nm and 230 nm was used to assess the purity of the DNA and the ratio measurement at 260 nm and 280 nm used to determine the quantity of DNA.

4.2.3 Exome Sequencing by the Australian Genome Research Facility

Exome sequencing was contacted to the Australian Genome Research Facility (AGRF) and was performed on the patient blood samples (three sisters diagnosed with PCOS in a single-family). AGRF exome sequencing requires a minimum of 20 ng/µL purified DNA for standard library preparation, the samples submitted for sequencing each contained ~5 µg DNA for high-quality exome sequencing (~60 ng/µL). Exome sequencing at AGRF incorporates sample quality assessment, Illumina library preparation, the exome capture, and sequencing on the HiSeq® 2500 Illumina Sequencing System using a paired-end (2x100 bp) protocol. The Human reference genome used was GRCh37/hg19, and the Burrows-Wheeler Aligner
Tool was used for mapping to the reference genome. Variant calls were performed with HaplotypeCaller by GATK (version 4.0.4.0).

### 4.2.4 Bioinformatics Analyses

Data generated from all samples were returned as .FASTQ, .BAM, .VCF, and .TXT file formats for analysis. Prior to analysis, a checksum function was performed to test the integrity of the data files by verifying that the data was complete and error-free.

The Java program Exomiser detects potential disease-causing variants from exome sequencing data. Filtered VCF files (supplied by AGRF) for each study subject were analysed in conjunction with a set of phenotypes specific to PCOS using the Exomiser program, which was run using the root directory and text editor for Linux. Phenotypes were selected based on the phenotypes expressed in and/or used for PCOS diagnosis, questionnaires from the patients, and confidential patient medical notes that were made available for the study. The phenotypes were encoded using the Human Phenotype Ontology (HPO) for use in Exomiser. Exomiser uses the VCF files and HPO terms to annotate, filter, and prioritise likely causative variants, and create a summary output generated in a .HTML file format. In addition to the HPO terms an exclusion criterion for the minor allele frequency (MAF) of <2% was applied, which functioned to filter all variants more frequent than 2% out of the analysis. The .HTML files for each sample submitted were investigated and the top 50 gene mutations (in prioritised order) were analysed in a spreadsheet to find predicted pathogenic mutations present in all three sisters. The genes containing these mutations were then analysed for a role in PCOS, by investigating these genes using the Mouse Genome Informatics (IGM), PubMed online databases, UCSC genome browser, Ensembl genome browser, and The Database of Short Genetic Variation (dbSNP) (Smedley et al., 2015).

The specific HPO terms used were: HP:0000140, abnormality of the menstrual cycle; HP:0001061, acne; HP:0000141, amenorrhea; HP:0000868, decreased fertility in females; HP:0100879, enlarged ovaries; HP:0008675, enlarged polycystic ovaries; HP:0030348, increased circulating androgen level; HP:0000858, menstrual irregularities; HP:0000876, oligomenorrhea; HP:0000147 polycystic ovaries.
4.3 Results

4.3.1 Clinical Recruitment and Exome Sequencing

Numerous individuals who expressed interest in the research were followed-up during this project. One family consisting of three sisters, and two daughters from one of the sisters, were successfully recruited for the study. Physician diagnosis of PCOS was confirmed for the three sisters, however diagnosis of PCOS was not possible for the two daughters who are both of a very young age and showing only early symptoms of PCOS; these children could not be recruited for the study because the human research ethics approval provides only for adults to be included in the investigation. Therefore, blood from the two daughters was not able to be collected. Genomic DNA samples from the three sisters were submitted to AGRF for exome sequencing.

Figure 4.1. Pedigree diagram of the individuals for whom DNA was exome sequenced in this study. Proband is highlighted by an arrow, probands daughters are provisionally diagnosed with PCOS, but are yet to be classified as physician-diagnosed.

The mean age for the three sisters was 37.3 years with a +/- 2.5 standard deviation (SD), the mean weight in kg was 54.8 +/- 4 SD, and the mean height was 165.7 cm +/- 2.5 SD, as seen in table 4.1. The age ranged from 35 to 40 years, the weight from 51 to 59 kg, and the height from 163 to 168 cm.
Table 4.1. Showing the mean and standard deviation for the three sisters.

<table>
<thead>
<tr>
<th></th>
<th>AGE</th>
<th>WEIGHT (KG)</th>
<th>HEIGHT (CM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>MEAN</td>
<td>37.3</td>
<td>54.8</td>
<td>165.7</td>
</tr>
<tr>
<td>STANDARD DEVIATION</td>
<td>2.5</td>
<td>4.0</td>
<td>2.5</td>
</tr>
</tbody>
</table>

The PCOS questionnaires were investigated and the three sisters share several common features. Age at first period for each of the sisters were 13, 14, and 16 years of age and the contraceptive pill was used to control acne for all sisters, and two of the sisters further used the pill for painful and heavy periods. In the absence of the pill they all experienced irregular periods, and each of the sisters also experienced more than 6 months in between periods. Two out of the three sisters took longer than 12 months to conceive and required assisted reproductive technology. Moreover, none of the sisters has had gestational diabetes or gestational hypertension (pre-eclampsia) (table 4.2).

Table 4.2. Summary of information from PCOS questionnaires.

<table>
<thead>
<tr>
<th>Q2358</th>
<th>Q2359</th>
<th>Q2361</th>
</tr>
</thead>
<tbody>
<tr>
<td>MENARCHE</td>
<td>14</td>
<td>13</td>
</tr>
<tr>
<td>PERIOD (ABSENCE OF PILL)</td>
<td>Irregular then regular</td>
<td>Always irregular</td>
</tr>
<tr>
<td>TIME BETWEEN PERIODS</td>
<td>&gt;6 months</td>
<td>&gt;6 months</td>
</tr>
<tr>
<td>MENOPAUSE HIRSUTISM</td>
<td>Yes</td>
<td>Yes (a little bit)</td>
</tr>
<tr>
<td>GESTATIONAL DIABETES</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>GESTATIONAL HYPERTENSION</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>ETHNICITY</td>
<td>Caucasian</td>
<td>Caucasian</td>
</tr>
<tr>
<td>REASON FOR THE PILL</td>
<td>Acne, painful/heavy periods</td>
<td>Acne, regulate periods, contraception</td>
</tr>
<tr>
<td>LONGER THAN 12 MONTHS TO CONCEIVE</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>REQUIRED ASSISTED REPRODUCTIVE TECHNOLOGY</td>
<td>Yes, in vitro fertilisation</td>
<td>No, but used metformin medication</td>
</tr>
</tbody>
</table>
Prior to analysis of the exome sequencing, data for the sequencing quality control metrics of the three samples were analysed which showed that the per-base sequence quality for the three samples looked excellent with more than 86% bases above Q30 (Phred quality score; quality bases) across all samples. The samples were also screened by the AGRF for the presence of any contamination before exome sequencing and passed this assessment. Exome sequencing for each of the three samples was successful and the bioinformatics analysis conducted by the AGRF involved demultiplexing, realignment and variant calls of indels (insertion and deletions of bases) and single-nucleotide polymorphisms (SNV). The primary data was generated using the Illumina (bcl2fastq 2.20.0.422) pipeline and the sequence files were produced in a standard FASTQ file format, ready for bioinformatic analysis.

The analyst remarks from AGRF for total reads, total reads mapped, aligned proportion, read pairs mapped, the proportion of read pairs mapped and the target coverage breadth can be seen in table 4.3. These remarks are excellent for all samples analysed.

Table 4.3. Analyst remarks from AGRF of the three samples submitted for exome sequencing.

<table>
<thead>
<tr>
<th>SAMPLE ID</th>
<th>TOTAL READS</th>
<th>TOTAL READS MAPPED</th>
<th>ALIGNED PROPORTION</th>
<th>READ PAIRS MAPPED</th>
<th>PROPORTION PAIRS MAPPED</th>
<th>TARGETS COVERAGE BREADTH</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q2358</td>
<td>62,293,467</td>
<td>62,118,344</td>
<td>99.72%</td>
<td>30,476,326</td>
<td>97.85%</td>
<td>99.53%</td>
</tr>
<tr>
<td>Q2359</td>
<td>57,273,060</td>
<td>62,118,344</td>
<td>99.75%</td>
<td>28,055,909</td>
<td>97.97%</td>
<td>99.53%</td>
</tr>
<tr>
<td>Q2361</td>
<td>60,501,228</td>
<td>60,350,612</td>
<td>99.75%</td>
<td>29,657,691</td>
<td>98.04%</td>
<td>99.52%</td>
</tr>
</tbody>
</table>

The variant character distribution was observed across the DNA of the three PCOS study subject samples analysed by exome sequencing. The variant classes and the variants by most severe cases were analysed for each sample, with a very high degree of consistency among the three sisters. A summary of these results can be seen in the pie charts of figure 4.2. Single nucleotide polymorphisms (SNP) was the single most common variant within the variant classes with 91.2%. Of the most severe variant cases, the top three were intron variants with 44.2%, and missense variants and synonymous variants, both with 18.5%.
Figure 4.2. Variant character distribution observed across the DNA of three PCOS study subject samples analysed by exome sequencing showing A. variant classes and B. variants by most severe consequences.

4.3.2 Mutations Identified

The .HTML output files for each sample analysed by Exomiser were investigated manually and the top 50 gene mutations (in prioritised order) were selected for further analysis. Potentially pathogenic variants in 7 genes were found to be shared between the three sisters. The name of these 7 genes can be seen in table 4.4 alongside the corresponding rs numbers for the SNPs, chromosome location, mutation type, and the minor allele frequency.
Table 4.4: Potentially pathogenic variants identified in all three sisters diagnosed with PCOS through exome sequencing. Showing the gene name, rs ID number, chromosome location (GRCh37/hg19), human genome variation society (HGVS) name, and minor allele frequency.

<table>
<thead>
<tr>
<th>GENE</th>
<th>RS ID</th>
<th>CHROMOSOME LOCATION</th>
<th>HGVS NAME</th>
<th>MINOR ALLELE FREQUENCY*</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTTP</td>
<td>rs767893675</td>
<td>4: 100512944</td>
<td>ENST00000265517:c.755C&gt;T:p.(Ser252Leu)</td>
<td>0.00002</td>
</tr>
<tr>
<td>IFT57</td>
<td>rs574081689</td>
<td>3:107937439</td>
<td>ENST00000264538:c.437A&gt;G:p.(Tyr146Cys)</td>
<td>0.00001</td>
</tr>
<tr>
<td>MDK</td>
<td>No rs#</td>
<td>11: 46404337</td>
<td>ENST000000489525:c.447_451del:p.(Ala150Profs*15)</td>
<td>ND</td>
</tr>
<tr>
<td>HAS1</td>
<td>rs1474834071</td>
<td>19:52217309</td>
<td>ENST000000594621.1:c.596C&gt;T:p.(Pro199Leu)</td>
<td>ND</td>
</tr>
<tr>
<td>SIPA1L1</td>
<td>No rs#</td>
<td>14:72085603</td>
<td>ENST00000358550:c.1628A&gt;T:p.(Glu543Val)</td>
<td>ND</td>
</tr>
<tr>
<td>MUC16</td>
<td>rs200575681*</td>
<td>19:9049665</td>
<td>ENST00000397910:c.31966G&gt;A:p.(Val10656lle)</td>
<td>ND</td>
</tr>
<tr>
<td>ZFYVE19</td>
<td>rs1351648333</td>
<td>15:41102928</td>
<td>ENST00000336455:c.769G&gt;A:p.(Asp257Asn)</td>
<td>0.00002</td>
</tr>
</tbody>
</table>

* Minor allele frequency derived from gnomAD, * new allelic variant of existing rs#, ND – no data.

Some further statistical scores were needed for interpretation of the potential consequences that the shared variants may have, these can be seen in table 4.5.

Table 4.5: Potentially pathogenic variants identified in all patients through exome sequencing. Showing the PolyPhen-2, SIFT and CADD scores for the mutations identified which are in silico predictions of likely pathogenicity.

<table>
<thead>
<tr>
<th>GENE</th>
<th>MUTATION TYPE</th>
<th>ALLELE</th>
<th>RS ID</th>
<th>POLYPHEN-2 SCORE</th>
<th>SIFT SCORE</th>
<th>CADD SCORE</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTTP</td>
<td>Missense</td>
<td>C&gt;T</td>
<td>rs767893675</td>
<td>0.976</td>
<td>0.005</td>
<td>0.9996</td>
</tr>
<tr>
<td>IFT57</td>
<td>Missense</td>
<td>T&gt;C</td>
<td>rs574081689</td>
<td>0.974</td>
<td>0.009</td>
<td>0.9953</td>
</tr>
<tr>
<td>MDK</td>
<td>Frameshift Truncation</td>
<td>GCTGCC&gt;G</td>
<td>No rs#</td>
<td>ND</td>
<td>ND</td>
<td>ND</td>
</tr>
<tr>
<td>HAS1</td>
<td>Missense</td>
<td>G&gt;A</td>
<td>rs1474834071</td>
<td>0.912</td>
<td>0.0</td>
<td>0.1496</td>
</tr>
<tr>
<td>SIPA1L1</td>
<td>Missense</td>
<td>A&gt;T</td>
<td>No rs#</td>
<td>0.996</td>
<td>0.0</td>
<td>0.9964</td>
</tr>
<tr>
<td>MUC16</td>
<td>Missense</td>
<td>C&gt;T</td>
<td>rs200575681*</td>
<td>0.351</td>
<td>0.0</td>
<td>0.6991</td>
</tr>
<tr>
<td>ZFYVE19</td>
<td>Missense</td>
<td>G&gt;A</td>
<td>rs1351648333</td>
<td>1.0</td>
<td>0.0</td>
<td>0.9981</td>
</tr>
</tbody>
</table>

* new allelic variant of existing rs#, ‡ Truncations early in a transcript is generally considered pathogenic, ND - no data.
SIFT and PolyPhen-2 are free in silico pathogenicity software that are available online and the scores predict the possible impact of an amino acid substitution on the structure and function of a protein. These tools model the likelihood of the variant having the deleterious effect, based on the degree of conservation of the affected allele throughout evolution (Flanagan et al., 2010). PolyPhen-2 scores range from 0 being a benign variant to 1 being probably damaging. An example of the output for rs767893675 observed in this study is illustrated in Figure 4.5. In contrast, SIFT scores are scaled in the opposite direction, having the scores range from 0 (pathogenic) to 1 (benign). CADD is a combined score integrating pathogenicity modelling data from a variety of sources and the interpretation.

![Figure 4.5](image)

**Figure 4.5. Analysis of rs767893675 by PolyPhen-2 predictor** (note scale ranges from 0 being benign and 1.0 being probably damaging (Adzhubei et al., 2010). rs767893675 achieves a score of 0.998 which is reported as probably damaging.

Combined Annotation Dependent Depletion (CADD) is another in silico tool for scoring the deleteriousness/likely pathogenicity of single nucleotide variants and small insertion/deletions variants in the human genome. Like Polyphen-2, a CADD score of 0 represents a benign variant versus a score of 1 being pathogenic. While many variant annotation software available, most annotations tend to exploit a single information type (e.g. conservation) and/or are restricted in scope (e.g. to missense changes). CADD integrates multiple annotations into one metric by contrasting variants that survived natural selection with simulated mutations. C-scores strongly correlate with allelic diversity, the pathogenicity of both coding and non-coding variants, and experimentally measured regulatory effects, and also highly rank causal variants within individual genome sequences. (Rentzsch et al. 2018).

A variant with the PolyPhen-2 score of 0.85 – 1.0 are predicted to be damaging, and scores in the range of 0.15 to 1.0 are possibly damaging. According to the PolyPhen-2 scores the
mutations found in the genes MTTP, ZFYVE19, SIPA1L1, and IFT57 are all predicted to be damaging and the mutations found in genes MUC16 and HAS1 are possibly damaging. For the SIFT score, variants with scores between 0.0 to 0.05 are considered deleterious and all of the mutations discovered falls within this category. However, no PolyPhen-2 or SIFT score could be obtained for the mutation discovered in MDK.

The CADD score is especially powerful as it integrates multiple annotations into one metric. And like Polyphen-2, a CADD score of 0 represents a benign variant versus a score of 1 being pathogenic. The mutations observed in the MTTP, IFT57, SIPA1L1, and ZFYVE19 all have CADD scores close to 1, indicating that these are all pathogenic variants. A CADD score could not be obtained for the MDK variant, however, a frameshift causing a truncation of the gene early in a transcript is generally considered to be pathogenic.

When combining the scores for the PolyPhen-2, SIFT, and CADD, the most likely pathogenic variants in relative ranked order are found in the genes; MTTP, ZFYVE19, SIPA1L1, IFT57, MUC16, HAS1, and MDK.

4.4 Discussion

All of the women included in this study were diagnosed with PCOS in their adulthood (2007-2019), likely while experiencing trouble conceiving, as this is when women often seek medical help. It is also common to seek diagnosis due to amenorrhea, which all of the sisters experienced. A family history of PCOS, diabetes, obesity and masculinisation in women are also common motives to seek a diagnosis. However, none of these factors are part of the symptoms of the women included in this study, they are all of a lean build, elevated androgen levels were not detected, and hirsutism was minor or non-present.

Exome sequencing is a very efficient technique of discovering novel mutations shared by individuals diagnosed with complex syndromes, such as PCOS, compared to selecting genes and sanger sequencing each of these to look for mutations. Moreover, an approach of having to try and guess which gene is involved in a candidate gene study design or to Sanger sequence a whole lot of genes are not time-efficient and hence not labour efficient. Whereas exome sequencing is relatively inexpensive, very time efficient, labour efficient,
and saves the cost of laboratory products (only one DNA extraction is required). Exome sequencing is a powerful technique where you get to survey coding changes in pretty much every gene in the genome.

The economic efficiency of exome sequencing is very good as it is relatively inexpensive at $750, compared to most diagnostic procedures. It is performed as an outpatient from a single blood sample and is very different to for example a colonoscopy which is a complex day surgery or inpatient exploration commonly costing $3,000, a computed tomography (CT) scan costing $5,000 on the high end, or a magnetic resonance imaging (MRI) scan ($3,500) which are similarly complex medical investigations.

Moreover, exome sequencing has a higher diagnostic yield compared to many simulated disease-specific panels for monogenic disorders (Dillon et al., 2018). Monogenic disorders are the result of a single defective gene on the autosomes. The mutation implicated in a monogenic disorder can be spontaneous and without any linkage to previous family history. Whole exome sequencing (WES) has much broader coverage than the disease-specific panels and together with the cost decline of WES this is increasingly being applied for clinical diagnosis, and it is now the primary alternative for diagnosis of suspected genetic disorders (Dillon et al. 2018). The research paper by Daga, Majmundar et al. found that whole exome sequencing frequently detects a monogenic cause in the early onset of nephrolithiasis (kidney stones) and nephrocalcinosis (deposition of calcium salts in the renal parenchyma), and concludes that the approach of using exome sequencing for this purpose represents a major advance in providing specific etiologic diagnosis and aids in enabling personalised treatment plans for patients with a detectable monogenic cause of disease. However, exome sequencing is still much less used for complex genetic disease. This could be due to the factor that it is still somewhat impractical for smaller laboratories to perform whole-exome sequencing for large numbers of subjects at sufficiently high coverage, which is required to complete valid large-scale genetic association studies of complex traits, such as PCOS. Exome sequencing for variants in rare Mendelian diseases has in the last few years become widespread, with more than 100 genes being characterised, however, application of exome sequencing for non-Mendelian phenotypes are to date much less widespread (Wang et al., 2013).
In this study, the most compelling finding was the detection of rs767893675 in the *Homo sapiens* microsomal triglyceride transfer protein (MTTP) gene in all three of the PCOS affected sisters. This variant is known as NP_000244.2:p.Ser252Leu and is predicted to cause a serine to leucine amino acid change at amino acid number 252. Despite the in-silico prediction software characterising the variant as likely pathogenic, there are no reports in ClinVar of this being confirmed through functional studies as a pathogenic variant.

MTTP encodes the large subunit of the heterodimeric microsomal triglyceride transfer protein. Protein disulfide isomerase (PDI) works to complete the heterodimeric microsomal triglyceride transfer protein, and this process has been shown to play a central role in lipoprotein assembly. Moreover, mutations in MTTP can cause abetalipoproteinemia (Kent et al. 2002). Abetalipoproteinemia is a disorder that interferes with the normal absorption of fat and fat-soluble vitamins. The protein product of this gene is reported to play a role in the pathways of metabolism of lipids and lipoproteins, the Statin Pathway, fat digestion and absorption, and in the circadian rhythm. To date, there are no previous reports of an association specifically with PCOS. GWAS have highlighted variants in the vicinity of this gene for the following traits: social interaction measurement, high-density lipoprotein cholesterol measurement, red blood cell distribution, lung vital capacity, high altitude adaptation malaria susceptibility, alcohol dependence, celiac disease, and hypertension.

Variants of the MTTP may cause dysregulation of metabolism of lipids and lipoproteins, and fat digestion and absorption. It is therefore possible that the variant discovered in this gene could be contributing to the metabolic syndrome often observed in women with PCOS, together with the increased risk factor of hypertension (also related to MTTP).

The second most compelling finding was the discovery of rs1351648333 in the Homo Sapiens Zinc Finger FYVE Domain-Containing Protein 19 (ZFYVE19) gene found in all three of the PCOS affected sisters. This variant, known as NP_001070736.1:p.Asp267Asn is predicted to cause an aspartic acid to asparagine amino acid change at amino acid number 267. Like for the MTTP variant, despite the in-silico prediction software characterising the
rs1351648333 variant as likely pathogenic, there are no reports in ClinVar of this being a pathogenic variant confirmed via functional studies.

ZFYVE19 is a key regulator of the abscission step in cytokinesis and aberrations in this gene are associated with acute myeloblastic leukaemia (Kent et al. 2002). ZFYVE19 is not reported to play a role in any pathways or diseases listed in GeneCards. To this date, there are no reports of an association specifically with PCOS. Variants in the vicinity of this gene have been reported by GWAS for the following traits: blood protein measurement, cardiovascular disease, eosinophil percentage of granulocytes, and triglyceride measurement. Overall, the variant observed in ZFYVE19 seems less likely to be implicated in the aetiology of PCOS.

The third most compelling finding was the finding of a missense variant in the Homo Sapiens Signal Induced Proliferation Associated 1 Like 1, also known as the High-Risk Human Papilloma Viruses E6 Oncoproteins Targeted Protein 1 (SIPA1L1) gene discovered in all three sisters diagnosed with PCOS. This variant does not have an rs number, but according to the HGVS name in table 4.4 this variant is predicted to cause a glutamic acid to valine amino acid change at amino acid number 543. Despite the in-silico predictions software characterising this variant as likely pathogenic there is no report in ClinVar of this being confirmed through functional studies as a pathogenic variant, however, the Fusion Gene Annotation Database relates mutations in the SIPA1L1 gene to weight gain.

SIPA1L1 acts to stimulate GTPase activity of the RAP2A gene of the RAS oncogene family. It promotes reorganization of the actin cytoskeleton, contributes to the regulation of dendritic spine morphogenesis, and controls neuronal morphology by interaction with other genes and domains (Kent et al. 2002). Among its related pathways are Protein-protein interactions at synapses and Transmission across Chemical Synapses, however, no diseases for the SIPA1L1 gene was found in GeneCards. To date, there are no previous reports of an association specifically with PCOS. GWAS have highlighted variants nearby this gene for traits of a variation of heart conditions and hip bone mineral density, and none of the PCOS GWAS has so far reported any associations in the neighbourhood of this gene. According to these findings, the variant discovered in SIPA1L1 does not seem a likely contributor to the aetiology of PCOS.
The fourth next persuasive finding was the detection of rs574081689 in the Homo Sapiens Intraflagellar Transport 57 protein (IFT57) gene exposed in all three sisters included in this study. This variant is known as NP_060480.1:p.Tyr146Cys and is predicted to cause a tyrosine to cysteine amino acid change at amino acid number 146. Although the in-silico prediction software characterised the variant as a likely pathogenic variant there are no reports in ClinVar confirming this.

IFT57 is required for the formation of cilia, plays a role in apoptosis, and may act as a transcription regulator (Kent et al. 2002). Among its related pathways are Signalling by GPCR and the fMLP Pathway. There are no reports of an association specifically with PCOS and GWAS have highlighted one variant in the vicinity of this gene with the trait of social interaction measurement. This variant does not appear to be involved in PCOS specifically.

The fifth most convincing discovery was the finding of a new allelic variant for the existing rs200575681 in the Homo Sapiens Mucin 16, Cell Surface Associated (MUC16) gene. As there is no rs number for this variant there is also no available data in ClinVar regarding the pathogenicity of this variant. The MUC16 gene encodes a protein that is a member of the mucin family and play an important role in forming a protective mucous barrier and are found of the apical surfaces of the epithelia (Kent et al. 2002). To date, there are no previous reports of an association with PCOS. GWAS have highlighted variants in the vicinity of this gene, however, none of these traits appears relevant to PCOS. Among the gene cards disease association (not GWAS) are ovarian cysts, papillary adenocarcinoma, and ovarian disease. The association with ovarian cysts may be of interest, however, it does not lend much credibility as the gene is associated with many forms of cancers, including ovarian cancer.

Investigation of the Hyaluronan Synthase 1 (HAS1) gene and the rs1474834071 missense variant does not lend much credibility towards involvement in the development of PCOS, and there is no ClinVar data available for this variant. However, a GWAS have highlighted two variants in the vicinity of this gene with the following traits: high-density lipoprotein cholesterol measurement and alopecia areata.
Little information could be found on the variant for HAS1, however, a GWAS associated it with high-density lipoprotein cholesterol measurement and alopecia areata, so it may be associated with PCOS.

The last variant discovered in all PCOS affected sisters through exome sequencing was the variant discovered in the Midkine (MDK) gene. No rs number, PolyPhen-2/SIFT/CADD scores, or minor allele frequency was available for this variant, however, a frameshift truncation early in a transcript is generally considered pathogenic and hence investigation of this gene too was important.

The MDK gene is developmentally regulated and is a secreted growth factor. It binds to and induces activation of anaplastic lymphoma kinase which causes phosphorylation of the insulin receptor substrate, followed by activation of some other kinases leading to induction of cell proliferation (Kent et al. 2002). MDK is involved in, amongst others, the ERK signalling pathway. There is no association of MDK specifically with PCOS, and GWAS has only highlighted this gene for the involvement in schizophrenia.

The variant discovered in the MDK gene did not initially appear to be of pathogenic significance, however, a frameshift truncation is likely to cause pathogenicity (as the sequence would be read out of frame, yielding a nonsense protein) and this gene is interestingly a secreted growth factor. It is possible that dysregulation of this gene could be involved in the formation of cysts observed in PCOS.

4.5 Conclusion
In this study exome sequencing has proven to be a highly efficient diagnostic approach to study short genetic variations in the coding sequence of the genome that are relevant to PCOS. The technique is relatively inexpensive, and the tissue sample required for analysis is a simple venous blood sample. Out of the 7 variants discovered, 2 appear to be of special relevance; the variant in MTTP and the variant in MDK. The next step would be to annotate these variants in ANNOVAR and to do Sanger sequencing for variant confirmation.
Chapter 5

Cloning and Optimization of Expression of GDF9 Constructs
5 CLONING AND OPTIMIZATION OF EXPRESSION OF GDF9 CONSTRUCTS

5.1 Introduction

Growth differentiation factor-9 (GDF), as stated previously, growth differentiation factor-9 (GDF9) is a member of the transforming growth factor-beta (TGFβ) superfamily first discovered in 1993 (McPherron and Lee, 1993). GDF9 plays a key role in folliculogenesis and female fertility through selective expression in developing oocytes (Otsuka et al., 2011; Filho et al., 2002; Erickson and Shimasaki, 2000; McGrath et al., 1995). GDF9 was chosen as the candidate gene for the functional work of this thesis as a novel variant (rs13164856) associated with PCOS was discovered to be in the DNA looping vicinity of GDF9 (via Hi C chromatin interaction), and this variant could therefore be highlighting a regulatory element for GDF9. Given the physiological role of GDF9 and this potential variant link to PCOS, it was decided that GDF9 is a strong candidate to be involved in PCOS and an excellent choice for functional study.

In order to investigate this gene, a mutation observed in the GDF9 gene of a woman suffering from PCOS was selected and the sequence introduced into the pcDNA3.1 mammalian expression vector.

5.2 Methods

5.2.1 Cloning of the GDF9 ORF into the pcDNA3.1 Mammalian Expression Vector

5.2.1.1 Primer design

Primers were designed to clone DNA from the open reading frame (ORF) GDF9 (section 2.6). This was completed by using the first 12 bases of the ORF in the forward primer and the reverse complement of the last 19 bases in the reverse primer. The primers also needed to include restriction endonuclease sites for ligation into the pcDNA3.1 mammalian cloning vector. Suitable restriction endonucleases needed to be present once in the multiple cloning site of the pcDNA3.1 vector and not present within the GDF9 ORF. The restriction endonucleases EcoRV and XhoI met these criteria, and also have the added benefit of digesting equally well in the same buffer (NEBuffer3.1). The sequence for EcoRV was added 5’ to the ORF sequence in the forward primer and the XhoI restriction endonuclease sequence 5’ to the ORF sequence in the reverse primer (see figure 5.1, flow diagram).
Translation of the XhoI site was prevented by the addition of a stop codon between the ORF and XhoI site (see section 2.6).

5.2.1.2 Extraction of GDF9 DNA and Cloning into the pDrive Cloning vector
PCR of GDF9 plasmid, gel extraction, and ligation into the pDrive cloning vector as per figure 5.1.

The initial amplification and purification of GDF9 DNA from the pMD-GDF9 clone was of a too low concentration for TA cloning into the pDrive cloning vector. Following an investigation of the concentration of pMD-GDF9, it was found to be of a much lower concentration than stated. This obstacle was overcome by ethanol precipitation to concentrate the isolated GDF9 DNA.

5.2.1.3 Sanger Sequencing
Internal primers were designed for Sanger sequencing of the entire GDF9 DNA (see section 2.6). These primers overlapped and revealed the vector sequence on either side of the DNA insert which allowed for confirmation of the correct insert sequence. Reaction prepared and analysed by AGRF.

5.2.1.4 Site-Directed Mutagenesis
The c.783delC mutation was introduced into the GDF9 DNA within the pDrive cloning vector using the QuickChange® Lightning Site-Directed Mutagenesis Kit (Agilent Technologies, USA), with mutagenesis primers designed as per section 2.6 (GDF9-783delC-F/R).

5.2.1.5 Ligation of GDF9 Wildtype and Mutant into pcDNA3.1
The GDF9 wildtype and mutant inserts needed to be subcloned from the pDrive cloning vector into the pcDNA3.1 mammalian expression vector, as the pDrive vector was simply used as a cloning tool to clone the PCR products. pDrive-GDF9 containing the wildtype and mutant inserts underwent restriction enzyme digestion with EcoRV and XhoI restriction enzymes. Empty pcDNA3.1 vector was also restriction digested with the same restriction enzymes and then dephosphorylated as described in section 3.2.6.3. These products were analysed by gel electrophoresis and the band size corresponding to the GDF9 insert and
pcDNA3.1 vector were excised using a surgical blade. DNA was extracted from the agarose gel using the QIAEX II® Gel Extraction Kit (Qiagen), and the GDF9 wildtype and mutant inserts were ligated into the pcDNA3.1 vector using T4 DNA ligase. The ligation mixes containing potential pcDNA3.1-GDF9 wild type or mutant recombinants was transformed into XL-2 (E. coli) cells and single colonies were analysed by PCR colony screening and agarose gel electrophoresis for the GFD9 wildtype or mutant insert sequence. Single colonies showing the presence of insert were inoculated into 2X TY culture containing 50 µg/mL ampicillin and DNA was extracted from the overnight bacterial culture using the Wizard® Plus SV Minipreps DNA Purification System (Promega). Plasmid DNA containing the GDF9 inserts was sent for Sanger Sequencing, using the internal primers, to check the sequence and confirm that ligation into pcDNA3.1 was successful. Glycerol stocks were made for the recombinants containing wildtype and mutant GDF9, stored at -70°C, and colonies were inoculated as needed and DNA was extracted using the PureYield™ Plasmid Miniprep System (Promega).

The wild type GDF9 released from pDrive was also cloned into pcDNA3.1 by the EcoRV and XhoI endonuclease restriction sites as described for the mutant in figure 5.1.

5.2.2 Plasmid Construction – FLAG-tagged Constructs
The same approach as described above was used when introducing the FLAG-tag into the constructs. The FLAG-tag sequence was placed 5’ to the ORF sequence in the forward primer, and the reverse primer was the same as that described in section 5.2.2.1.

5.2.3 Selection of Cell Lines for Transfection
Three cell lines were available for functional genomics: A2780 and HEK297 which are both human cell lines, and COS1 derived from monkey. COS1 was the least appropriate cell line as it is non-human. The A2780 cell line was determined the most relevant as it is a human ovarian epithelial cell line, whereas the HEK293 cells are derived from the kidney.
5.2.4 Transfection

Cells were grown in 75cm² flasks and seeded into 6-well plates at 1x10⁶ cells per well for A2780 cells or into 25cm² flasks at 2x10⁶ cells per flask for HEK293 cells for transfection. Transfections were generally carried out using Fugene HD for the A2780 cell line at varying ratios and Lipofectamine 2000 for the HEK293 cell line at ratio 2:1.

All transfections reagents were brought to room temperature, then OPTIMEM, DNA, and transfection reagent were added to respective tubes to obtain the desired ratio in a total volume of 200 µL, and the tubes incubated for 10 minutes at room temperature. Media of the 6-well plates containing the previously seeded out cells were changed and exactly 1.8 mL of DMEM (without antibiotics) added, and the full content of each tube added to the respective well of the 6-well plate for a final volume of 2 mL. Incubation and media change were as described in section 3.3.3.1.

5.2.4.1 Cell Lysis and Protein Quantitation

Transfected cells were lysed in RIPA buffer and protein quantitated using the Pierce™ BCA Protein Assay Kit by Thermo Fisher Scientific and NanoDrop One spectrophotometer (Thermo Scientific, USA). 100 µg of cell lysate protein was analysed on a 12% SDS-PAGE gel.

5.2.4.2 Western Blot Analysis

For western blot analysis, the gel was blotted onto a nitrocellulose membrane (Amersham™ Protran™ by Sigma-Aldrich, USA) and probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate. For flag detection, the membrane was probed with mouse monoclonal anti-FLAG antibody followed by goat anti-mouse antibody. And for β-actin loading control, the membrane was stripped and then probed with β-actin mouse antibody followed by goat anti-mouse antibody. ECL detection of the membrane was performed using the Bio-Rad ChemiDoc imager.
GDF9 was amplified by PCR from the pMD-GDF9 clone and the PCR product gel-purified

TA/PCR cloning of GDF9 into pDrive cloning vector

Transformation of pDrive ligatants into E. coli XL-2 cells, and selection of single colonies subcultured overnight

Extraction of plasmid DNA from overnight cultures and analysis of pDrive clones to check for GDF9 insert

Site-directed mutagenesis to introduce c.783delC mutant using the pDrive construct as template

Transformation into E. coli XL-2 cells, and selection of single colonies, subcultured overnight

Extraction of plasmid DNA from overnight cultures

Sanger sequencing to confirm mutant GDF9 sequence in pDrive followed by release of GDF9 insert by EcoRV and XhoI enzymes

GDF9 mutant DNA was gel purified and ligated into pcDNA3.1

Transformation of pcDNA3.1 recombinants into E. coli XL-2 cells

PCR colony screening of single colonies, single colony selection and subculture overnight

Extraction of plasmid DNA from overnight cultures using the PureYield™ Plasmid Miniprep System

Plasmid DNA ready for functional study

Figure 5.1. Flow diagram for the cloning approach used for the creation of the mutant GDF9 into the pcDNA3.1 mammalian expression vector.
5.3 Results
GDF9 was cloned from the pMD-GDF9 clone into the pcDNA3.1 mammalian expression vector as described in the flowchart see figure 5.1.

5.3.1 Isolation of GDF9 DNA
Amplification of GDF9 DNA, from the pMD-GDF9 clone, resulted in strong bands at 1.36kb (Figure 5.2). These bands were excised, and then gel purified.

![Agarose gel showing PCR amplified product of GDF9 from the pMD-GDF9 clone](image)

**Figure 5.2. Isolation of GDF9 DNA.** Agarose gel (1%, EtBr) showing PCR amplified product of GDF9 from the pMD-GDF9 clone (Sino Biological Inc.) with bands at 1.36kb.

5.3.2 Sequence Confirmation of Amplified GDF9
GDF9 amplified product was cloned into the pDrive vector and sent for Sanger sequencing to ensure the template (pMD-GDF9) contained the correct GDF9 sequence (figure 5.3).

![Sequence around the mutation site](image)

**Figure 5.3. GDF9 amplified from the pMD-GDF9 clone.** Figure showing the sequence around the mutation site. No errors were introduced during the PCR amplification and the correct sequence was confirmed (not shown).
5.3.3 Cloning of GDF9

Sanger sequencing also revealed the presence of GDF9 containing the EcoRV and XhoI endonuclease restriction sequences (Figure 5.4).

Figure 5.4. Chromatograms of pDrive-GDF9 showing correctly incorporated A. stop signal (TGA) and XhoI sequence, and B. EcoRV sequence.
Site-directed mutagenesis resulted in the deletion of a cytosine at base pair position 783 in the ORF of GDF9, leading to a premature stop codon (Figure 5.5).

**Figure 5.5. Chromatogram showing sequence confirmation of the GDF9 c.783delC mutant in pDrive.** The CCC codon for proline in wild type GDF9 is mutated to CCT, also coding for proline in the GDF9 mutant. The cytosine (C) deletion (marked with an asterisk) at base-pair position 783 changes the reading frame of the GDF9 mutant leading to a premature stop codon (TGA) and changing the TCA codon for serine to CAC for histidine in the GDF9 mutant.

pDrive-GDF9 wildtype and mutant were restriction enzyme digested with EcoRV and XhoI restriction endonucleases, analysed on 1% agarose gel, the correct bands excised, and gel purified for ligation into the pcDNA3.1 mammalian cloning vector. Sanger sequencing of the wildtype and mutant inserts in pcDNA3.1 confirmed the presence of GDF9 within the vectors, together with the restriction endonuclease sites (results not shown).

### 5.3.3.1 Incorporation of FLAG-tag

The same approach (as described in section 5.3.3 and shown in figure 5.1) was used when introducing the FLAG-tag into the constructs, except that the FLAG-tag sequence was placed 5’ to the ORF sequence in the forward primer with the initial ATG sequence of the GDF9 ORF placed 5’ to the sequence encoding the FLAG peptide (DYKDDDDK). The reverse primer was kept as described in section 5.2.2.1.
Figure 5.6. Chromatogram showing sequence confirmation of the incorporated FLAG-tag of the FLAG-tagged pcDNA3.1-GDF9 mutant. The restriction enzyme EcoRV cutting site and the ATG start codon precede the sequence encoding the FLAG-tag (DYKDDDDK) (highlighted in the red box).

It was necessary to reconstruct the clones with the incorporation of a FLAG-tag as expression of the mutant construct, using the GDF9 primary antibody, was unsuccessful. The FLAG-tag WT and mutant GDF9 constructs were successfully created in the pDrive cloning vector, however, only the mutant GDF9 was successfully subcloned into the pcDNA3.1 vector; the wild type GDF9 still awaits subcloning into pcDNA3.1.

5.3.4 Transfection Optimisation of Wildtype and Mutant pcDNA3.1-GDF9 into the A2780 Mammalian Cell Line

5.3.4.1 Optimisation of Primary Antibody in SMP

As an attempt to reduce non-specific binding following ECL detection previously prepared protein lysates were used and the ratios of Fugene HD to WT DNA were 1.5:1 and 2:1 for 2µg of DNA (lysate from the experiment described in section 5.3.4.3). The membrane was blocked with 1X TBS-T + 5% Skim Milk Powder (SMP) for 2 hours at room temperature. Following the blocking, the membrane was cut in four parts with each part including the Bio-Rad Precision Plus Protein Dual Standards for size determination and the two WT protein samples. The first three membranes were incubated with the primary antibody
(GDF9) at 1:5000 dilution in 1X TBS-T + 10% SMP for 2 hours, 5 hours, and overnight. The fourth membrane was incubated with a 1:500 dilution of the primary antibody (GDF9) overnight also in 1X TBS-T + 10% SMP. All membranes were incubated at 4°C.

Following incubation, all the membranes were washed as described in section 3.4.4. and then incubated with secondary antibody (goat anti-rabbit-HRP conjugate) at 1:5000 dilution in 1X TBS-T + 10% SMP for 1 hour at room temperature. Finally, the membranes were washed again as per section 3.4.4 and detected using the ECL reagent.

![Blot Image](image)

**Figure 5.7. Effect of varying the dilutions of rabbit polyclonal GDF9 antibody, and incubation time.** Shown in the lower panel is the β-Actin loading controls. The position of GDF9 wildtype can be seen from the protein marker on the left-hand side. 100 µg cell lysate protein was loaded and separated on a 12% SDS-PAGE gel before blotting onto a nitrocellulose membrane. The membrane was blocked with 5% SMP for 2 hours and then cut into four parts; A, B, C, and D. A, B, and C, were incubated with rabbit polyclonal GDF9 antibody at 1:5000 dilution. A; incubated for 2 hours, B; 5 hours, C; overnight. D was incubated with rabbit polyclonal GDF9 antibody at 1:500 dilution overnight. All four membranes were incubated with goat anti-rabbit-HRP conjugate at 1:5000 dilution for 1 hour. Membrane detection with ECL reagent and the Bio-Rad ChemiDoc imager.

The conditions used for blot D (Figure 5.7) was determined to be the most effective. However, for the next blots the following conditions were adopted; blocking with 5% BSA for 2 hours, overnight incubation with rabbit polyclonal GDF9 antibody at 1:500 dilution (10% BSA), and incubation with goat anti-rabbit-HRP conjugate at 1:5000 dilution for 1 hour (10% BSA). Blocking with BSA appears more efficient for the reduction of non-specific bands.
Determination of Optimal Transfection Reagent for A2780s

The optimal transfection reagent for my cell line needed to be established. The optimal transfection reagent to use may depend on the vector used and the size of the gene of interest. The transfection reagents tested were Viafect, Fugene 6, an old aliquot of Fugene HD, and newly ordered Fugene HD, the ratios of transfection reagent to DNA were 1.5:1, 2:1, and 2.5:1, and 2 µg of DNA, WT pcDNA3.1-GDF9 vector, was used.

Figure 5.8. Effect of different transfection reagents and ratio on transfection efficiency of pcDNA3.1-GDF9 wildtype into A2780 cells. The β-actin loading control can be seen in the lower panel. The position of GDF9 wildtype is indicated from the protein markers on the left-hand side of the upper panel (ladder). The first 4 conditions and the last 8 conditions were run on separate gels. 2 µg of DNA was transfected for each condition. The type of transfection reagent and the ratio of transfection reagent to DNA is indicated above each lane. 100 µg cell lysate protein was separated on a 12% SDS-PAGE gel and blotted onto a nitrocellulose membrane. The membrane was probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate. ECL detection was performed using the Bio-Rad ChemiDoc imager.
When 2 µg of pcDNA.3.1-GDF9 wildtype was transfected with each of the four transfection reagents, at varying ratios, no difference in transfection efficiency was observed (Figure 5.8). The intensity of the bands for ratio 2.5:1 Viafect and 2.5:1 Fugene HD (old) were lower but corresponds to the lower intensity of the β-actin loading control for these samples. The expression appears to be slightly higher for ratio 2.5:1 Fugene 6 and 2.5:1 Fugene HD, however, the intensity of the β-actin loading control for these samples are also slightly stronger. Because the β-actin loading control and all samples are of proportional intensity across the blots, there is no clear evidence that any transfection reagent is better than any other.

Fugene HD was chosen for further transfection optimisation as this transfection reagent was observed as successful in the A2780 cell line by work conducted by research assistant Shelby Mullin on the EEF2K gene, Department of Endocrinology and Diabetes lab, Sir Charles Gairdner Hospital.

5.3.4.3 Assessment of Transfection Efficiency of Fugene HD in both WT and Mutant constructs into A2780s

Following the experiment for determination of the optimal transfection reagent in A2780s, an experiment to assess the transfection efficiency of Fugene HD in both my WT and mutant constructs were conducted based on the results from Shelby Mullin’s experiments that indicated that Fugene HD might be the most effective transfection reagent for pcDNA3.1 constructs. The ratios of Fugene HD to DNA were 1.5:1, 2:1, and 2.5:1, and 2 µg of DNA with either WT or mutant pcDNA3.1-GDF9, empty vector (EV), or control (no DNA). The procedure for transfection was performed as per section 3.3.3.2.
Figure 5.9. Effect of three ratios of Fugene HD on the expression of pcDNA3.1-GDF9 WT and mutant into A2780 cells. The β-actin loading control for each lane can be seen in the lower panel. The positions of GDF9 wildtype, GDF9 mutant, and β-actin is indicated by the protein markers and the supporting text. The first 6 conditions and the last 6 conditions were run on separate SDS-PAGE gels, and 2 µg of DNA was transfected for each condition. The ratio of Fugene HD to DNA is shown above each lane. 100 µg cell lysate protein was loaded on a 12% SDS-PAGE gel and blotted onto a nitrocellulose membrane. The membrane was probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate. ECL detection was performed using the Bio-Rad ChemiDoc imager.

Minor overexpression can be seen with the three ratios of Fugene HD tested for the pcDNA3.1-GDF9 wildtype constructs (Figure 5.9, left panel, compare GDF9 and GDF9 mutant lanes). This was determined by comparison to the lower intensity of the bands observed for the mutant at the same ratios of Fugene HD. Moreover, the β-actin loading control lends strength to this as the strongest β-actin band can be seen with the pcDNA3.1-GDF9 mutant at ratio 1.5:1, whereas the β-actin bands for the rest of the lanes are all relatively even or (in the case of the 2.5:1 GDF9 mutant) proportionate to the low intensity of the GDF9 protein band. There is no evidence of expression for the mutant. It should be noted that the two blots came from separate SDS-PAGE gels and hence cannot be compared with great confidence.
5.3.4.4 Determination of Optimal Fugene HD to DNA ratio in A2780s

To achieve overexpression of WT (compared to endogenous levels), and expression of mutant GDF9, an optimization experiment of Fugene HD to DNA ratio was conducted. The ratios of Fugene HD to DNA were; 3:1, 4:1, and 5:1 for 4 µg WT DNA; 3:1, 4:1, 5:1 and 6:1 for 4 µg mutant DNA; 3:1, 4:1, and 6:1 for 5 µg mutant DNA; and 4:1 for 4 µg EV. The procedure for transfection was completed as per section 3.3.3.2. To reduce non-specific bands seen in previous experiments, three hours of blocking was performed in place of the two hours and the membrane washed five times instead of three (3.4.4).

Figure 5.10. Effect of varying the amount of DNA and the Fugene HD: DNA ratio for expression of wildtype and mutant pcDNA3.1-GDF9 plasmid into A2780 cells. The β-actin loading control is shown in the lower panel. The positions of GDF9 wildtype and mutant can be seen from the protein markers on the left-hand side of the upper panel. 100 µg cell lysate protein was separated on a 12% SDS-PAGE gel, then blotted onto a nitrocellulose membrane. The membrane was probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate. ECL detection was performed using the Bio-Rad ChemiDoc imager.
Some overexpression is observed for the pcDNA3.1-GDF9 wildtype constructs at ratio 3:1, 4 µg DNA transfected, compared to empty vector (Figure 5.10). This is supported by reference to the bands for the β-actin loading control. Moreover, this indicates that increasing the ratio of Fugene HD to DNA for the wildtype construct does not increase the degree of overexpression. There is no detectable expression of the mutant pcDNA3.1-GDF9 construct observed from this blot. The band seen at 29 kDa appears to be a non-specific band (not the pcDNA3.1-GDF9 mutant) as it occurs in the WT GDF9 lanes where mutant GDF9 constructs were not transfected so cannot be expressed. The “missing” band at ~29 kDa for the ‘Empty Vector’ is likely the result of a bubble being present during the blotting (protein transfer from SDS-PAGE gel to the nitrocellulose membrane) process or another aberration during the blotting process.

5.3.4.5 Optimization of Membrane Blocking

As a second attempt to reduce non-specific binding following ECL detection, the blocking stage was altered (Figure 5.11). Previously prepared protein lysate from the experiment at section 5.3.4.3 was used and protein lysate from WT, M, C (non-transfected control), and EV was used at a ratio of 2:1 of Fugene HD to DNA. Two membranes were prepared each containing the Bio-Rad Precision Plus Protein Dual Standards for size determination and WT, M, C (non-transfected), and EV protein. Both membranes were blocked in 1X TBS-T + 10% BSA, one for 2 hours at room temperature and the other overnight at 4°C. Western analysis was then followed as per section 3.4.4.
**Figure 5.11. Effect of increased the incubation time on blocking.** **A.** 2 hour blocking with 10% BSA at room temperature. **B.** Overnight blocking with 10% BSA at 4°C. The β-actin loading control is shown in the lower panel for each blot. The positions of GDF9 wildtype and mutant is indicated from the protein markers on the left-hand side of the upper panel. 100 µg cell lysate protein was separated on a 12% SDS-PAGE gel, then blotted onto a nitrocellulose membrane. The membrane was probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate and ECL detection was conducted using the Bio-Rad ChemiDoc imager.

Both figure A and B (of figure 5.11) are difficult to interpret as the protein bands are quite faint and the β-actin loading control is not even across the samples for either blot. Overall, the blocking optimization experiment shows that increasing time of the blocking stage does not appreciably change the non-specific binding pattern following ECL detection. The strength of the bands of interest observed are very faint and difficult to interpret. Hence, these optimisation approaches were not adopted.
5.3.5 Transfection of pcDNA3.1-GDF9 Constructs into the HEK293 Mammalian Cell Line

As an attempt to achieve sufficient overexpression of pcDNA3.1-GDF9 wildtype and expression of the mutant pcDNA3.1-GDF9 construct, HEK293 was used. The mammalian cell line HEK293 was used for expression of other pcDNA3.1 constructs and has proved to be efficient multiple times in the Department of Endocrinology and Diabetes Laboratory, Sir Charles Gairdner Hospital, WA.

The FLAG-tagged pcDNA3.1-GDF9 mutant construct was also transfected into the HEK293 cell line for FLAG detection.

The ratios of Lipofectamine 2000 to DNA were 2:1 for all conditions and the amount of DNA was 2 µg, 4 µg, and 6 µg each for pcDNA3.1-GDF9 wildtype, mutant, FLAG-tagged mutant, and empty vector. The procedure for transfection into the HEK293 cell line was as per section 3.3.3.2.
Figure 5.12. Effect of variable amounts of DNA transfected of pcDNA3.1-GDF9 wildtype, mutant, and empty vector into HEK293 cells, using Lipofectamine 2000. The β-actin loading control is presented in the lower panel and the positions of GDF9 wildtype and mutant can be seen from the protein markers on the left-hand side of the upper panel. 100 µg cell lysate protein was separated on a 12% SDS-PAGE gel, then blotted onto a nitrocellulose membrane. The membrane was probed with rabbit polyclonal GDF9 antibody followed by goat anti-rabbit-HRP conjugate and ECL detection using the Bio-Rad ChemiDoc imager.

From the blot in figure 5.12, it appears that there is no overexpression of pcDNA3.1-GDF9 wildtype or expression of GDF9 mutant, only endogenous GDF9 seems to be present.
Figure 5.13. Effect of variable amounts of DNA transfected of FLAG-tagged GDF9 mutant, EEF2K, and empty vector in HEK293 cells, using Lipofectamine 2000. The β-actin loading control is presented just below the 50 kDa marker. The positions of FLAG-tagged pcDNA3.1-GDF9 mutant and the EEF2K SDS-PAGE and blotting controls can be seen from the protein markers on the left-hand side and right-hand side, respectively, of the upper panel. 100 µg cell lysate protein was separated on a 12% SDS-PAGE gel, then blotted onto a nitrocellulose membrane. The membrane was probed with mouse monoclonal anti-FLAG antibody (1:5,000) followed by goat anti-mouse antibody (1:10,000). ECL detection was performed using the Bio-Rad ChemiDoc Imager.

Transfection and expression of FLAG-tagged pcDNA3.1-GDF9 mutant in the HEK293 cell line were unsuccessful (Figure 5.13). The control FLAG-tagged pcDNA3.1-EEFK2 was successfully detected, indicating that the SDS-PAGE, blotting, and ECL detection were performed correctly.
5.4 Discussion

5.4.1 Isolation of GDF9 DNA
GDF9 DNA was successfully amplified from the pMD-GDF9 clone and isolated by agarose gel extraction.

5.4.2 Cloning
Ligation of GDF9 DNA containing the EcoRV and XhoI restriction endonuclease sites was successful. Site-directed mutagenesis was successful for the introduction of the GDF9 deletion mutation. Ligation of both wildtype and mutant GDF9 sequences into the pcDNA3.1 mammalian expression vector was successful, and the sequence within the vectors was confirmed to have the desired restriction nuclease sequences. However, the FLAG-tag was only successfully incorporated into the mutant GFD9 construct due to time constraints.

5.4.3 Selection of Cell Line
The most relevant cell line for this study would have been a human ovarian granulosa tumour cell line, as GDF9 is most readily expressed in this type of tissue. However, an ovarian epithelial cell line (A2780) was chosen and used in this study, as there has been no previous success in this laboratory at growing various granulosa cell lines, and the faster growth rate of the epithelial cells was desirable due to time constraints. However, expression was minimal or not observable in the A2780 cell line following transfection of plasmid constructs containing wildtype or mutant GDF9. The HEK293 human kidney cells were also investigated for expression of the pcDNA3.1-GDF9 constructs. However, there was no observed expression in this cell line. The next step would be to keep optimizing expression in the A2780 cell line and also attempt transfection and expression in the COS-1 monkey cell line (also available at the Diabetes and Endocrinology Laboratory) as the expression of pcDNA3.1 constructs in this cell line is generally observed to be very high. At least this would show that the GDF9 constructs could be expressed before pursuing further optimisation studies in human cell lines.
5.4.4 Presence of Endogenous GDF9

Endogenous wildtype GDF9 was observed for transfections into the A2780 cell line, however, no endogenous levels of GDF9 mutant were observed. Moreover, some overexpression of GDF9 wildtype was achieved in the A2780 cell line, but of insufficient levels for further functional studies.

5.5 Conclusion

This experiment aimed to produce GDF9 wildtype and mutant (c.783delC) sequences in mammalian cell lines and observing its effects on cell cycle signalling pathways in which GDF9 is involved. The construction of mammalian expression vectors containing the wildtype and mutant GDF9 sequences was successfully conducted. However, the overexpression of GDF9 wildtype and mutant constructs were unsuccessful, and further functional analysis could not be achieved within the set timeframe. However, the transfection and western blot optimisation did provide valuable information about how to proceed with the study of these constructs. Further research could focus on the signalling pathways which GDF9 is known to play an important role (i.e. the TGF-β BMP signalling pathway).
Chapter 6

General Discussion

&

Conclusion
6 GENERAL DISCUSSION AND CONCLUSION

6.1 General Discussion

The hypothesis that thorough and efficient molecular genetic diagnoses of PCOS, including identification and full characterisation of the genetic lesions will achieve better patient outcomes and facilitate targeted therapies was indeed proven correct. Exome sequencing was used as a molecular diagnostic approach to efficiently highlight the aetiological variants and metabolic pathways affected, and when successfully executed, could direct which therapy be the most appropriate. This is further illustrated by the functional molecular biology work were working towards the demonstration of the functional effect of a different variant (GDF9 c.783delC) that may be relevant to PCOS and reported in the literature to be pathogenic with no functional proof provided (Franca et al., 2018). GDF9 is in linkage disequilibrium with the PCOS GWAS variant rs13164856 reported in the literature and it is also within the DNA looping (by HiC analysis/long-range control element range, Figure 1.1).

Growth differentiation factor 9 (GDF9) is required during early ovarian folliculogenesis and directly affect oocyte growth and function (Otsuka et al. 2011). Large numbers of mutations in the GDF9 gene have been identified in women with premature ovarian failure which also highlights the crucial role this gene plays in ovarian function and female fertility. GDF9 is not required for male fertility, which was demonstrated by a complete loss of GDF9 in homozygous male mice, whereas in female mice this caused a block in follicle growth and a complete absence of normal follicles beyond the primary stage (Dong et al. 1996). Moreover, oocytes deficient in GDF9 are demonstrated to grow more rapidly than controls and achieve a larger size, despite the block in follicle growth at the primary stage (Carabatsos et al. 1998).

GDF9 is most homologous to BMP-15 in primary structure, as well as in expression distribution in the ovary, and hence play an important role in the TGFβ/BMP signalling pathway. The GDF9 c.783delC mutation is a truncation mutation and is likely to be pathogenic with severe implication to the normal functions of GDF9 and therefore also possibly disrupt several parts of the TGFβ/BMP signalling pathway (i.e. binding sites might
be missing due to the truncation). One of the aims of this thesis was to functionally characterise the effects of the GDF9 mutant using the Qiagen TGFβ/BMP signalling pathway RT² profiler PCR Array system and to investigate how the mutant alters the regulation (i.e. upregulation or downregulation) of important genes in this cell cycle pathway, such as the SMAD genes. The SMAD genes are of particular interest as GDF9 and BMP-15 act synergistically to target the SMAD3 pathway.

Another future aspect to investigate would be the stability of the truncation mutant, as it might be rapidly degrading and could explain why there is no expression of the mutant following transfection. To assess the stability of the GDF9 mutant protein the cycloheximide assay could be conducted.

Once the expression of mutant and wildtype GDF9 is achieved, a knockdown of wildtype GDF9 by siRNA technology before transfecting the mutant plasmid would be desirable. This would provide a clearer picture of what effects the mutant might be having on the regulation of genes in the RT² profiler PCR Array system since there would be much less endogenous wildtype GDF9 present. Silencer Select pre-designed and validated siRNA sequences (including for GDF9) are available from companies such Ambion. Another possibility is to use CRISPR/Cas9 technology to introduce the GDF9 mutant in the homozygous state into a haploid cell line.

The rs767893675 mutation within the MTTP gene is predicted to be pathogenic, segregates with family members, is rare in the general population (MAF = 0.00002), and the PolyPhen-2 and SIFT scores suggest that this variant may be pathogenic, thereby increasing its chances of being disease-causing. rs767893675 has the potential of being involved in three common symptoms of PCOS: metabolism of lipids and lipoproteins (metabolic syndrome), fat digestion and absorption (obesity), and hypertension.
6.2 General Conclusion

The current diagnostic criteria for PCOS often result in heterogeneous groups of patients with very different genetic lesions and limited treatment options. Exome sequencing and functional genomics are capable and cost-effective methods of investigating pathological genetic variants and their cellular effects. Therefore, the efficient uncovering of genetic diagnosis of PCOS, including exomic identification and functional characterisation of the genetic lesions, will help achieve better patient outcomes and provide the basis for targeted therapies. The work of this thesis intended to identify variants that are likely to play a role in the development of PCOS, and to characterise the phenotypic effects of a different mutation discovered in an individual with PCOS by the following aims:

1. Recruitment of one or more families with multiple individuals with clinically defined PCOS and identify functional variants implicated in the disease aetiology
2. Developing a protocol for characterisation of putative pathogenic PCOS variants
3. If/when the variant is fully characterised that too may inform on the best course of treatment for the patient

Moreover, this work also aimed to analyse the pathogenicity of the GDF9 c.783delC variant by observing its involvement in and effect on cell cycle pathways in which GDF9 is involved, and hence its possible implication in the development of PCOS. Functional studies are important to generate proof for in-silico predictions of what detrimental effects a variant has based on how a specific mutation affects the function of a gene. Mammalian clones for expression studies were successfully created containing the wildtype and mutant GDF9 sequences. However, optimization for sufficient overexpression of GDF9 wildtype and the expression of mutant GDF9 proved unsuccessful. As overexpression was not achieved, the experiments to assess the stability of these proteins and the effect of mutant GDF9 on relevant cell cycle pathways could not be conducted. However, generated clones for this study and the optimizations conducted provided valuable information on how to proceed with further experiments and will aid any further functional study of this gene. The future research pathways for this functional work would be to fully analyse the pathogenicity of the mutant variant by observing its effect on relevant cell signalling pathways. If successful, the achievement of these aims will substantially aid PCOS patients and particularity those
where there is evidence of a mutation with a large effect size that is segregating within a family.

Overall, therefore, the work of this thesis has addressed important aspects of how exploiting genetic molecular diagnoses of PCOS may lead to achieving better patient outcomes.
Chapter 7

References
References:


Urbanek, M. (2007). The genetics of the polycystic ovary syndrome. *Nature Clinical Practice Endocrinology &amp; Metabolism*, 3, 103. doi:10.1038/ncpendmet0400


Appendix

I

Recipes
RECIPES

1% Agarose Gel

<table>
<thead>
<tr>
<th></th>
<th>Small Gel</th>
<th>Large Gel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agarose</td>
<td>0.3 g</td>
<td>1 g</td>
</tr>
<tr>
<td>DdH₂O</td>
<td>30 mL</td>
<td>100 mL</td>
</tr>
<tr>
<td>Ethidium Bromide OR</td>
<td>0.6 µL</td>
<td>2 µL</td>
</tr>
<tr>
<td>Gel Red</td>
<td>0.5 µL</td>
<td>1.7 µL</td>
</tr>
</tbody>
</table>

Microwave agarose and ddH₂O on high for 2 min, swirl, and heat for a further minute. Leave to cool to approximately 50°C, and then add the appropriate volume of EtBr or Gel Red, pour gel and leave to set for minimum 30 min before use.

10% Ammonium Persulphate (APS)

APS 0.5 g

Make up to 5 mL with ddH₂O, aliquot into 1.5 mL Eppendorfs and store at -20°C.

50 mg / mL Ampicillin

Ampicillin 1 g

Dissolve the ampicillin in 20 mL ddH₂O. Filter sterilize and aliquot into 1.5 mL Eppendorfs, store at -20°C.

10 x Blue Juice Gel Loading Buffer (For 1 Kb DNA Ladder)

<table>
<thead>
<tr>
<th></th>
<th>0.325g</th>
<th>65%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sucrose</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Tris-HCl pH 7.5</td>
<td>0.5 mL</td>
<td>10 mM</td>
</tr>
<tr>
<td>EDTA</td>
<td>0.1 mL</td>
<td>10 mM</td>
</tr>
<tr>
<td>Bromophenol Blue</td>
<td>1.5 mL</td>
<td>0.3%</td>
</tr>
</tbody>
</table>

Make up to 50 mL with ddH₂O. Store at 4°C.

DMEM (Dulbecco’s Modification of Eagles Medium)

DMEM 1 Packet

HEPES 4.77 g

NaHCO₃ 3.7 g

Make up to 1 L with ddH₂O, adjust pH to 7.4. Filter sterilize and store at 4°C.
5x DNA Loading Buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Xylene Cyanole</td>
<td>50 mg</td>
<td>0.25% w/v</td>
</tr>
<tr>
<td>Bromophenol Blue</td>
<td>50 mg</td>
<td>0.25% w/v</td>
</tr>
<tr>
<td>Glycerol</td>
<td>6 mL</td>
<td>30% w/v</td>
</tr>
</tbody>
</table>

Make up to 20 mL with ddH₂O. Store at room temperature.

100 mg / mL Ethidium Bromide (EtBr)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
</tr>
</thead>
<tbody>
<tr>
<td>EtBr</td>
<td>50 mg</td>
</tr>
</tbody>
</table>

Make up to 5 mL with ddH₂O. Store at 4°C away from light.

1 Kb DNA Ladder

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 Kb DNA Ladder</td>
<td>250 µL</td>
<td>10 X Blue Juice Buffer</td>
</tr>
</tbody>
</table>

Make up to 1250 µL with ddH₂O. Aliquot to 1.5 mL Eppendorfs and store at -20°C. Final concentration of ladder is 1 µg / 5 µL.

1 X Phosphate Buffered Saline (PBS)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>8 g</td>
<td>137 mM</td>
</tr>
<tr>
<td>KCl</td>
<td>0.2 g</td>
<td>2.7 mM</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>1.44 g</td>
<td>10.14 mM</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>0.24 g</td>
<td>1.8 mM</td>
</tr>
</tbody>
</table>

Make up to 1 L with ddH₂O, adjust pH to 7.7. Aliquot into 3 x 500 mL Schott bottles, autoclave. Store at room temperature.

10 X Phosphate Buffered Saline (PBD)

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>NaCl</td>
<td>80 g</td>
<td>137 mM</td>
</tr>
<tr>
<td>KCl</td>
<td>2 g</td>
<td>2.7 mM</td>
</tr>
<tr>
<td>Na₂HPO₄</td>
<td>14.4 g</td>
<td>10.14 mM</td>
</tr>
<tr>
<td>KH₂PO₄</td>
<td>2.4 g</td>
<td>1.8 mM</td>
</tr>
</tbody>
</table>

Make up to 1 L with ddH₂O. Store at room temperature. Dilute to 1 X PBS with ddH₂O, adjust pH to 7.4, autoclave before use.

RadioImmunoPrecipitation Assay (RIPA) Buffer

<table>
<thead>
<tr>
<th>Component</th>
<th>Amount</th>
<th>Volume</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5 M Tris Cl pH 8.0</td>
<td>20 mL</td>
<td>50 mM</td>
<td></td>
</tr>
<tr>
<td>NaCl</td>
<td>1.752 g</td>
<td>150 mM</td>
<td></td>
</tr>
<tr>
<td>Triton X-100</td>
<td>2 mL</td>
<td>1%</td>
<td></td>
</tr>
<tr>
<td>20% SDS</td>
<td>1 mL</td>
<td>0.1%</td>
<td></td>
</tr>
</tbody>
</table>
Sodium Deoxycholate  1 g  0.5%

Make up to 200 mL with ddH$_2$O. Store at 4°C.

Transfer Buffer (For SDS-PAGE / Western)
- Trizma  3.016 g
- Glycine  14.4 g
- Methanol  200 mL

Add trizma and glycine to 750 mL ddH$_2$O and dissolve. Add methanol and then make up to 1 L with ddH$_2$O. Store at 4°C.

1 X Tris Buffered Saline (TBS)
- 10 X TBS  50 mL

Make up to 500 mL with ddH$_2$O, adjust pH to 7.4. Store at 4°C.

10 X Tris Buffered Saline (TBS)
- Trizma  60.5 g  0.5 M
- NaCl  85.0 g  1.75 M

Make up to 1 L with ddH$_2$O. Store at room temperature.

1 X Tris Buffered Saline + 0.1% Tween (TBS-T)
- 10X TBS  50 mL
- Tween 20  500 µL

Make up to 500 mL with ddH$_2$O, adjust pH to 7.4. Store at 4°C.

1 X Trypsin
- 10X Trypsin  3 mL
- 1X PBS  27 mL

Store at 4°C, heat to 37°C before use.

2 X YT Agar
- Bacto Tryptone  16g  1.6% w/v
- Bacto Yeast Extract  10g  1% w/v
- NaCl  5g  85.5 mM

Make up to 1 L with ddH$_2$O, adjust to pH 7.0, autoclave, store at room temperature.
**2 X YT Media/Broth**

- Bacto Tryptone: 16g (1.6% w/v)
- Bacto Yeast Extract: 10g (1% w/v)
- NaCl: 5g (85.5 mM)

Make up to 1 L with ddH$_2$O, adjust to pH 7.0, add 15g Bacto Agar, autoclave, add antibiotics (1mL of 50mg/mL ampicillin) at 50°C mix gently and pour plates. Store at 4°C.
Appendix

II

Clinical Recruitment
<table>
<thead>
<tr>
<th>Time of Last Dose</th>
<th>Time of Last Dose</th>
<th>Pain or Disease</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

ENDO

**Clinical Recruitment Form**

PathWest Blood Collection Form:

**Clinical Study**

Polypsichogical Wary Syndrome Study

Clinical Study

RET HE1D: CST 218

**PathWest**
Participant Information Sheet and Consent Form

Study Title: Genetic variation in genes regulating steroid biosynthesis and insulin metabolism, and their association with polycystic ovary syndrome.

Introduction

You are being invited to take part in a research study involving the analyses of genes (DNA) associated with polycystic ovary syndrome. Before you decide if you want to participate in this research study, it is important for you to understand why the research is being done and what it will involve.

Please take time to read the following information carefully and ask us if there is anything that is not clear or if you would like more information.

Purpose of the study

Polycystic ovarian syndrome (PCOS) is the most common form of endocrine (gland and hormone) disorder in premenopausal women. This common clinical syndrome starts soon after beginning of menstrual function and features consist of irregular periods, hirsutism (hairiness) and obesity often in association with enlarged ovaries.

The aim is to study the genetic basis of PCOS. We will examine whether gene variations predispose to developing PCOS. We will examine whether particular genetic code (DNA) variations are associated with a particular form of PCOS.

We are requesting your consent to a sample of your blood, as well as your permission to obtain related health information about you for the purposes of medical research. This information sheet provides information relating to this consent.

Why have I been chosen?

You have been chosen to take part because you have PCOS. We will be recruiting 400 women with this condition to take part in the study, which is being conducted at the Keogh Institute.

If you do not have PCOS and have normal menstrual cycles, you will be selected as controls (your clinical details will be matched with women who have PCOS). Equal numbers of controls will be age- matched with equal number of PCOS.

Study description

You will need to sign the consent form before you can take part. If you are under 18 years of age, your parent or guardian will also need to sign.

Why am I being asked to provide this sample?

You are being asked to provide a sample of blood (approximately 30 mLs) for the analysis of your blood biochemistry and you genetic code or DNA. Information about you will be used for research on PCOS.
Does the decision to participate affect my care in any way?

The choice to take part in this genetic research is entirely up to you and your medical care will not be affected by your decision not to take part.

What will be done with my blood sample?

Medical researchers for genetic studies of polycystic ovarian syndrome will use your stored DNA for this study. The study has been approved by the Human Research Ethics Committee of Sir Charles Gairdner Hospital. It will be conducted according to the principles set out in the National Health and Medical Research Council of Australia’s National Statement on Ethical Conduct in Human Research and the guidelines on the Privacy Act.

Researchers working on this study are required to demonstrate that they meet the appropriate Australian standards of ethics and privacy as detailed above.

Your DNA sample will be stored under strict security at Sir Charles Gairdner Hospital until it used up or you contact the Medical Director at the Keogh Institute for Medical Research to request its destruction (contact details at the end of this leaflet). We will store your sample securely and carefully, but cannot guarantee against inadvertent loss or damage that is beyond the Institute’s control. In all cases you have an over-riding right to revoke your consent at any time.

The aim of our research is to increase scientific and medical knowledge in the management of women with PCOS. However sometimes research on your DNA may lead to findings that result in the development of a commercial test or treatment. There will be no financial reward or remuneration to you in this event.

Access to medical records: protecting your privacy

Your DNA sample made available to researchers will be identified by a code only and it will not be possible for them to link it in any way to your personal information. However, the discovery of factors important in understanding PCOS may also require the knowledge of other relevant information about you. This information may come from hospital notes, specialists and GP records or may be information kept about you by the Department of Health as part of its regular function.

You are asked to give your consent to the researchers to access medical information kept about you relevant to medical research. Your details will be held in strict confidence at all times. The appropriate guidelines and statues required with Australia abide access to your health information.

Your consent is for ongoing and continued access to your health information and you may withdraw this consent at any time by writing to the Medical Director at the Keogh Institute (name and telephone number at the end of this document).

What will be done with my DNA sample?

Your sample will be used solely for genetic analyses looking at gene types and gene variations associated with PCOS. Your sample will expressly not be used for research that involves reproductive cloning. No researchers will be permitted to derive a genetic profile of you as an individual. No sample or health information will be released to any 3rd party.

Future implications of donating my DNA

Whilst your donated samples are not intended to be used in your diagnoses or treatment, it is possible that future DNA testing may result in new information about diseases or potential diseases that you carry. Such information will require extensive testing and validation before it can
be determined to be useful but some of this information may have health implications for you or your family and through approved medical channels to ensure proper care is available to you. At that time you may be approached by a medical practitioner associated with the study. You may choose what further information or investigations you would like to proceed with.

Findings from this research will be presented at scientific meetings and published in scientific journals but any publications will not contain any specific details of your identity. If you have any concern you may contact the medical director at the Keogh Institute at any time.

**What bad effects can happen to me by donating this sample?**

A needle will be used to draw blood from a vein in your arm. Most people experience slight pain at the site of the needle insertion and some may develop a bruise or very rarely an infection of the puncture site.

**Informed consent**

We ask that you give careful consideration to the information set out in this brochure before providing your consent. If you decide to take part in this study we suggest that you consider:

- Advising your family members, or the Executor of your will, of the existence of your DNA sample; perhaps even provide them with a copy of this Information Sheet and your signed Consent Form
- Advising your family members of the purpose for which you have provided your DNA sample.

**What if I change my mind?**

You maintain the right to withdraw your DNA sample at any time. If you wish to have your sample withdrawn from the DNA bank or stop access to your health information, please notify the Medical Director in writing. A letter confirming removal of your DNA sample or health information will be sent to you.

**Further information**

For independent advice or for any complaints you can contact the Human Research Ethics Committee at Sir Charles Gairdner Hospital. Their telephone number is written on the bottom of this participant information and consent from that you be will be given a copy to keep.

**Do you have further questions?**

If you have any concerns or questions about the genetic study or storage of your DNA sample, please contact the Medical Director, Dr Bronwyn Stuckey of the Keogh Institute on (08) 6457 2008 or Dr Scott Wilson on (08) 6457 2089.
Consent for blood taking and DNA banking for clinical research

I, _______________________________ of _______________________________

have read the Information sheet entitled:

"Genetic variation in genes regulating steroid biosynthesis and insulin metabolism and their association with polycystic ovary syndrome."

I have had explained to me and I understand the consequences involved in my voluntary donation of my DNA.

I have had an opportunity to discuss my participation with my family members, ask questions and am satisfied with the answers given.

In making my donation, I understand and agree that:

1) The blood (which in this consent form, includes its constituents and any cell lines derived from the blood) will be used in relation to genetic research study associated with polycystic ovarian syndrome

2) A sample of my DNA or derived cell line/s held in a bank will be discarded upon my written request to the Medical Directory of the Keogh Institute for Medical Research

3) I will be approached by a medical professional in regard to potentially important health information that arises from this research that may impact on me or my family. At that time I may choose what further information or investigations I would like to proceed with

4) I can request to know more specific details of any studies that used my samples at any time by contacting the Medical Director of the Keogh Institute for Medical Research

5) Research results and the fact that I have made this donation will not be revealed to any 3rd party not directly part of medical research without my written consent, except under subpoena

6) The Keogh Institute for Medical Research will not be liable for any loss of or damage to DNA used in accordance with this form

7) I will not benefit financially if this research leads to development of a new treatment or medical test

8) Storage of and access to my DNA sample will be managed by The Keogh Institute for Medical Research and only released where the research proposal has been approved by the Human Research Ethics Committee and with consent

9) I understand that international research collaboration using my DNA will only take place where researchers abide by equal or more stringent regulations of privacy and ethics as those in Australia, as assessed by the Human Research Ethics Committee

10) I give permission to access health information about me related to the research area defined in point 1 above, such as is kept in a medical record or by the WA Department of Health, to assist medical research only where the research proposal has been approved by a Human Research Ethics Committee.
I do □  I do not □

Consent to the storage and use of my blood or biochemical and genetic based medical research.

…………………………………………………………………………………………………………
… Name of Patient                     Signature of Patient                     Date

…………………………………………………………………………………………………………
… Name of Parent/Guardian             Signature of Parent/Guardian           Date

…………………………………………………………………………………………………………
… Name of Investigator (or Delegate)  Signature of Investigator (or Delegate)

Date

The Sir Charles Gairdner Group Human Research Ethics Committee has given ethics approval for the conduct of this project. If you have any ethical concerns regarding the study you can contact the SCGG HREC office on telephone number (08) 6457 2999.

All study participants will be provided with a copy of the Information Sheet and Consent Form for their personal records.
**PCOS Questionnaire:**

<table>
<thead>
<tr>
<th>Given name</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Family name</td>
<td></td>
</tr>
<tr>
<td>Date of birth</td>
<td></td>
</tr>
<tr>
<td>Weight</td>
<td></td>
</tr>
<tr>
<td>Height</td>
<td></td>
</tr>
<tr>
<td>Date of clinic visit</td>
<td></td>
</tr>
<tr>
<td>Diagnosis (PCOS or other?)</td>
<td></td>
</tr>
<tr>
<td>Blood pressure at visit (if known)</td>
<td></td>
</tr>
</tbody>
</table>

**How did you hear about the study? Please circle your response.**

- a) Social Media
- b) Pamphlets
- c) University
- d) Keogh Institute

**PCOS history:**

<table>
<thead>
<tr>
<th>Have you ever been diagnosed with PCOS?</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>At what age did you get your first period?</td>
<td></td>
</tr>
<tr>
<td>In the absence of hormonal contraceptive use, are/were your periods: Please circle only one response</td>
<td></td>
</tr>
<tr>
<td>Always regular</td>
<td></td>
</tr>
<tr>
<td>Regular then irregular</td>
<td></td>
</tr>
<tr>
<td>Always irregular</td>
<td></td>
</tr>
<tr>
<td>If your periods were ever irregular, what was the longest time between periods? Please circle only one response</td>
<td></td>
</tr>
<tr>
<td>Less than 6 months</td>
<td></td>
</tr>
<tr>
<td>More than 6 months</td>
<td></td>
</tr>
<tr>
<td>Between which ages did you use the contraceptive pill?</td>
<td></td>
</tr>
<tr>
<td>What were your main reasons for oral contraceptive use? Circle all that apply</td>
<td></td>
</tr>
<tr>
<td>Contraception</td>
<td></td>
</tr>
<tr>
<td>Acne</td>
<td></td>
</tr>
<tr>
<td>Hirsutism</td>
<td></td>
</tr>
<tr>
<td>Regulate periods</td>
<td></td>
</tr>
<tr>
<td>Other (please specify)</td>
<td></td>
</tr>
<tr>
<td>Have you ever used any other form of hormonal contraception? (Contraceptive patch, NuvaRing, intrauterine device, Depo Provera or Noristerat injection, Jadelle or implanon implants)</td>
<td></td>
</tr>
<tr>
<td>Yes</td>
<td></td>
</tr>
<tr>
<td>No</td>
<td></td>
</tr>
</tbody>
</table>

<p>| In your premenopausal years, did you experience hirsutism (excess body hair)? Please circle the pictures which best apply to you on the next page. |  |
| As an adult, what was your lowest weight? |  |
| As an adult what was your top weight? |  |</p>
<table>
<thead>
<tr>
<th><strong>Fertility history:</strong></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>How many children have you had?</td>
<td></td>
</tr>
<tr>
<td>Did it ever take you longer than 12 months to conceive (from when you first started trying)?</td>
<td></td>
</tr>
<tr>
<td>Did you ever require assisted reproductive technology?</td>
<td></td>
</tr>
<tr>
<td>Was male infertility a contributing factor?</td>
<td></td>
</tr>
<tr>
<td>Did you ever have gestational diabetes?</td>
<td></td>
</tr>
<tr>
<td>Did you ever have gestational hypertension (pre-eclampsia)?</td>
<td></td>
</tr>
<tr>
<td>How many of your first degree relatives have PCOS? Tick the boxes</td>
<td></td>
</tr>
<tr>
<td>Mother</td>
<td></td>
</tr>
<tr>
<td>Sisters (how many?)</td>
<td></td>
</tr>
<tr>
<td>Ethnicity: Please circle most appropriate response.</td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td></td>
</tr>
<tr>
<td>Asian</td>
<td></td>
</tr>
<tr>
<td>Aboriginal</td>
<td></td>
</tr>
<tr>
<td>Maori</td>
<td></td>
</tr>
<tr>
<td>Pacific Islander</td>
<td></td>
</tr>
<tr>
<td>Hispanic</td>
<td></td>
</tr>
<tr>
<td>Other. Please specify</td>
<td></td>
</tr>
</tbody>
</table>