



Article

Single-Cell RNA-Seq Identifies Pathways and Genes Contributing to the Hyperandrogenemia Associated with Polycystic Ovary Syndrome

R. Alan Harris ^{1,2,*} , Jan M. McAllister ³ and Jerome F. Strauss III ^{4,*}

¹ Human Genome Sequencing Center, Baylor College of Medicine, Houston, TX 77030, USA

² Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA

³ Department of Pathology, Penn State Hershey College of Medicine, Hershey, PA 17033, USA; jxm63@psu.edu

⁴ Department of Obstetrics and Gynecology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA

* Correspondence: rharris1@bcm.edu (R.A.H.); jerome.strauss@penmedicine.upenn.edu (J.F.S.III)

Abstract: Polycystic ovary syndrome (PCOS) is a common endocrine disorder characterized by hyperandrogenemia of ovarian thecal cell origin, resulting in anovulation/oligo-ovulation and infertility. Our previous studies established that ovarian theca cells isolated and propagated from ovaries of normal ovulatory women and women with PCOS have distinctive molecular and cellular signatures that underlie the increased androgen biosynthesis in PCOS. To evaluate differences between gene expression in single-cells from passaged cultures of theca cells from ovaries of normal ovulatory women and women with PCOS, we performed single-cell RNA sequencing (scRNA-seq). Results from these studies revealed differentially expressed pathways and genes involved in the acquisition of cholesterol, the precursor of steroid hormones, and steroidogenesis. Bulk RNA-seq and microarray studies confirmed the theca cell differential gene expression profiles. The expression profiles appear to be directed largely by increased levels or activity of the transcription factors SREBF1, which regulates genes involved in cholesterol acquisition (*LDLR*, *LIPA*, *NPC1*, *CYP11A1*, *FDX1*, and *FDXR*), and GATA6, which regulates expression of genes encoding steroidogenic enzymes (*CYP17A1*) in concert with other differentially expressed transcription factors (*SP1*, *NR5A2*). This study provides insights into the molecular mechanisms underlying the hyperandrogenemia associated with PCOS and highlights potential targets for molecular diagnosis and therapeutic intervention.

Keywords: polycystic ovary syndrome; single-cell RNA sequencing; human ovarian theca cells; androgens; steroids; cholesterol



Citation: Harris, R.A.; McAllister, J.M.; Strauss, J.F., III. Single-Cell RNA-Seq Identifies Pathways and Genes Contributing to the Hyperandrogenemia Associated with Polycystic Ovary Syndrome. *Int. J. Mol. Sci.* **2023**, *24*, 10611. <https://doi.org/10.3390/ijms241310611>

Academic Editor: Lalit Kumar Parameswaran Grace

Received: 31 May 2023
Revised: 14 June 2023
Accepted: 20 June 2023
Published: 25 June 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Polycystic ovary syndrome (PCOS) is a complex endocrine disorder that affects up to 10% of women of reproductive age worldwide [1–7]. PCOS is characterized by various clinical features, including hyperandrogenism, irregular menstrual cycles, infertility, obesity, and insulin resistance [6,8–10]. Despite extensive research, the molecular mechanisms that contribute to the development of PCOS are not fully understood.

Theca cells, a component of ovarian follicles, are thought to play a crucial role in the pathogenesis of PCOS. They are responsible for producing androgens, which are converted into estrogens by granulosa cells [2,11,12]. Several studies have shown that theca cells from women with PCOS exhibit altered gene expression profiles and increased androgen production compared to theca cells from women without PCOS [7,11–20]. However, the molecular mechanisms that underlie these changes in theca cell function have not been elucidated.

Single-cell RNA sequencing (scRNA-seq) [21] is a powerful tool for identifying molecular pathways and genes that are affected by the disease. This technology allows for identifying cell subpopulations, gene expression patterns, and gene regulatory networks

at high resolution. In the context of PCOS, scRNA-seq can be used to identify molecular changes in theca cells that are associated with disease pathogenesis, as well as to identify potential targets for therapeutic intervention.

Here, we describe using scRNA-seq to identify pathways and genes affected by PCOS in theca cells. Our findings provide new insights into the molecular mechanisms underlying PCOS, which may have diagnostic and therapeutic implications for this common endocrine disorder.

2. Results

2.1. scRNA-Seq Reveals That PCOS Theca Cells Have Distinctive Gene Expression Profiles

The 10× Cell Ranger [21] processing of the raw scRNA-seq data identified 70,223 cells across all 20 samples from women with and without PCOS that were untreated or forskolin-treated. After standard filtering for mitochondrial read counts < 5% and a number of gene features > 200 using Seurat [22], a total of 20,775 cells (29.58% of total cells) were retained. After integration with sub-type anchoring correction for alignment in Seurat (STACAS) [23] and additional filtering consisting of removing outlier cells based on UMAP visualization, a total of 15,565 (22.17%) theca cells were included in the downstream analyses. After all filtering, the number of cells per group was: unaffected/untreated 4672; unaffected/forskolin-treated 4793; PCOS/untreated 3446; PCOS/forskolin-treated 2654. UMAP visualizations of the filtered cells show the clustering of groups by affection status/treatment (Figure 1a), which is also consistent with the clustering of the samples in each group (Figure 1b).

Gene set enrichment analysis (GSEA) of molecular signatures database (MSigDB) C2: curated gene sets [24] using singleseqset [25] identified pathways significantly (FDR corrected $p < 0.05$) differentially expressed in comparisons across the four groups. Among these significant pathways were ones involved in cholesterol acquisition (six with CHOLESTEROL in the pathway terms) and steroidogenesis (10 with ANDROGEN or STEROID in the pathway terms) (Figure 2 and Table S1).

Working from the significant differentially expressed pathways, we identified individual genes in the pathways that were themselves significantly (Bonferroni corrected $p < 0.05$) differentially expressed genes (DEGs). We identified 168 genes differentially expressed in at least one group, including 18 genes in cholesterol acquisition and steroidogenesis pathways. There were 59 DEGs involved in cholesterol acquisition (Figure S1 and Table S2). These included the gene encoding the rate-limiting enzyme in de novo cholesterol biosynthesis (*HMGCR*) and the transcription factor *SREBF1*, which regulates genes involved in cholesterol acquisition.

We identified 127 DEGs involved in steroidogenesis (Figure S2 and Table S3). These included *STAR*, which catalyzes the rate-limiting step in steroid synthesis, the translocation of cholesterol to the inner mitochondrial membrane where the cholesterol side-chain cleavage enzyme, *CYP11A1*, another DEG, resides. The genes encoding the *CYP11A1* electron transport proteins (*FDX1*, *FDXR*), the enzyme involved in the metabolism of pregnenolone to DHEA (*CYP17A1*), and proteins that enhance the *CYP17A1* 17/20-lyase reaction (*POR*, *CYB5A*) [26–28], were also differentially expressed, with highest expression in PCOS theca cells stimulated with forskolin.

Based on transcription factors identified by Lambert et al. [29], the cholesterol acquisition pathways had six differentially expressed transcription factors (*NCOA1*, *NR1H2*, *NR1H3*, *RXRA*, *RXRB*, and *SREBF1*). The steroidogenesis pathways had 15 differentially expressed transcription factors (*AR*, *CREB1*, *FOXO1*, *JUN*, *KAT7*, *MAF*, *MYC*, *NCOA1*, *NCOA3*, *NR2C2*, *NR3C1*, *PATZ1*, *SP1*, *ZBTB10*, and *ZNF318*).

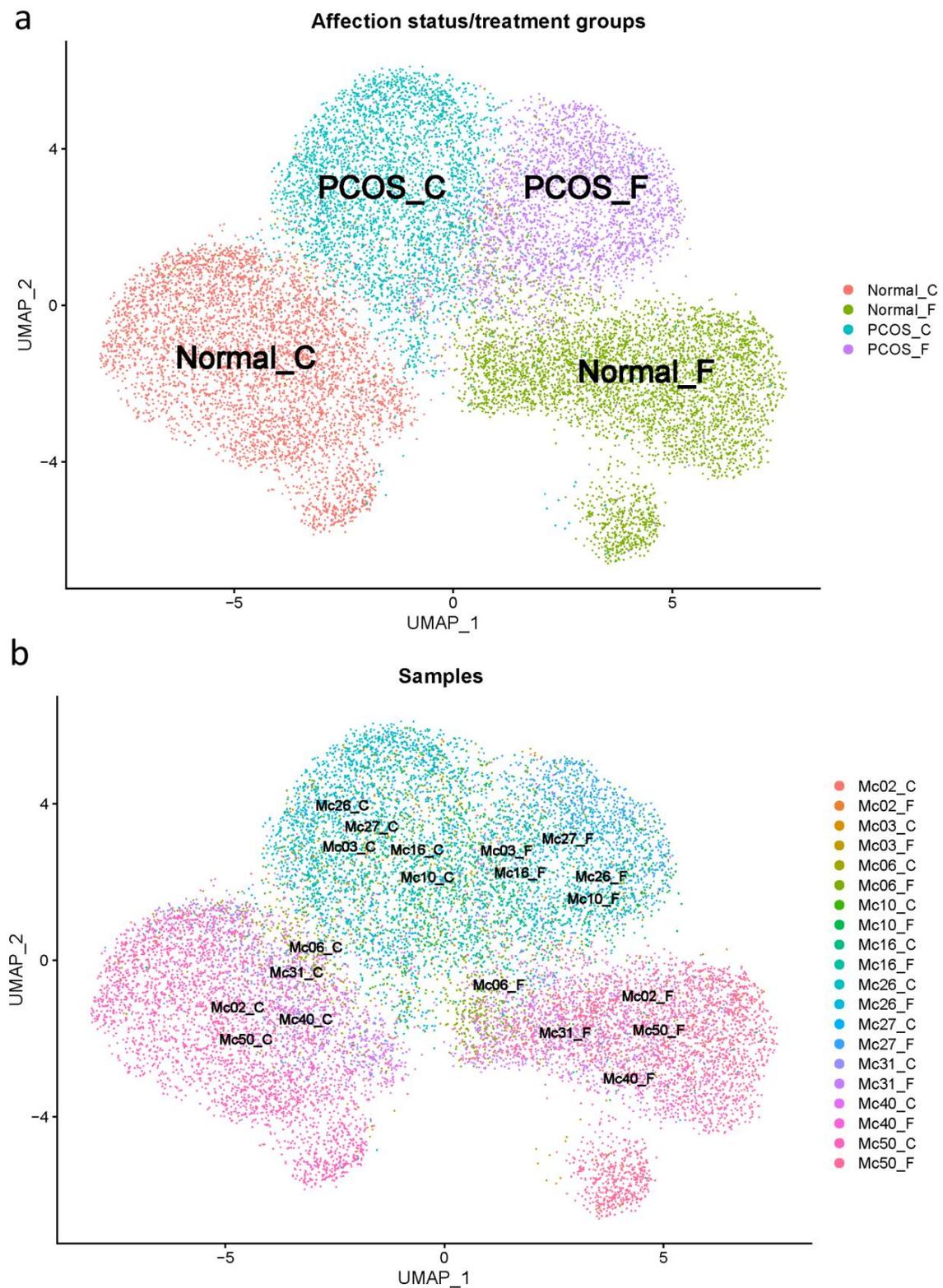


Figure 1. UMAP of affection status/treatment groups (a) and samples (b). Values from normal (control) theca cells were generated using the five theca preparations from PCOS women (MC03, Mc10, MC16, MC26, and MC27) and five theca cell preparations from normal cycling women (MC02, MC06, MC31, MC40, and MC50), that were grown until subconfluent, and treated with (F) and without (C), 20 mM forskolin for 24 h in a serum-free medium.

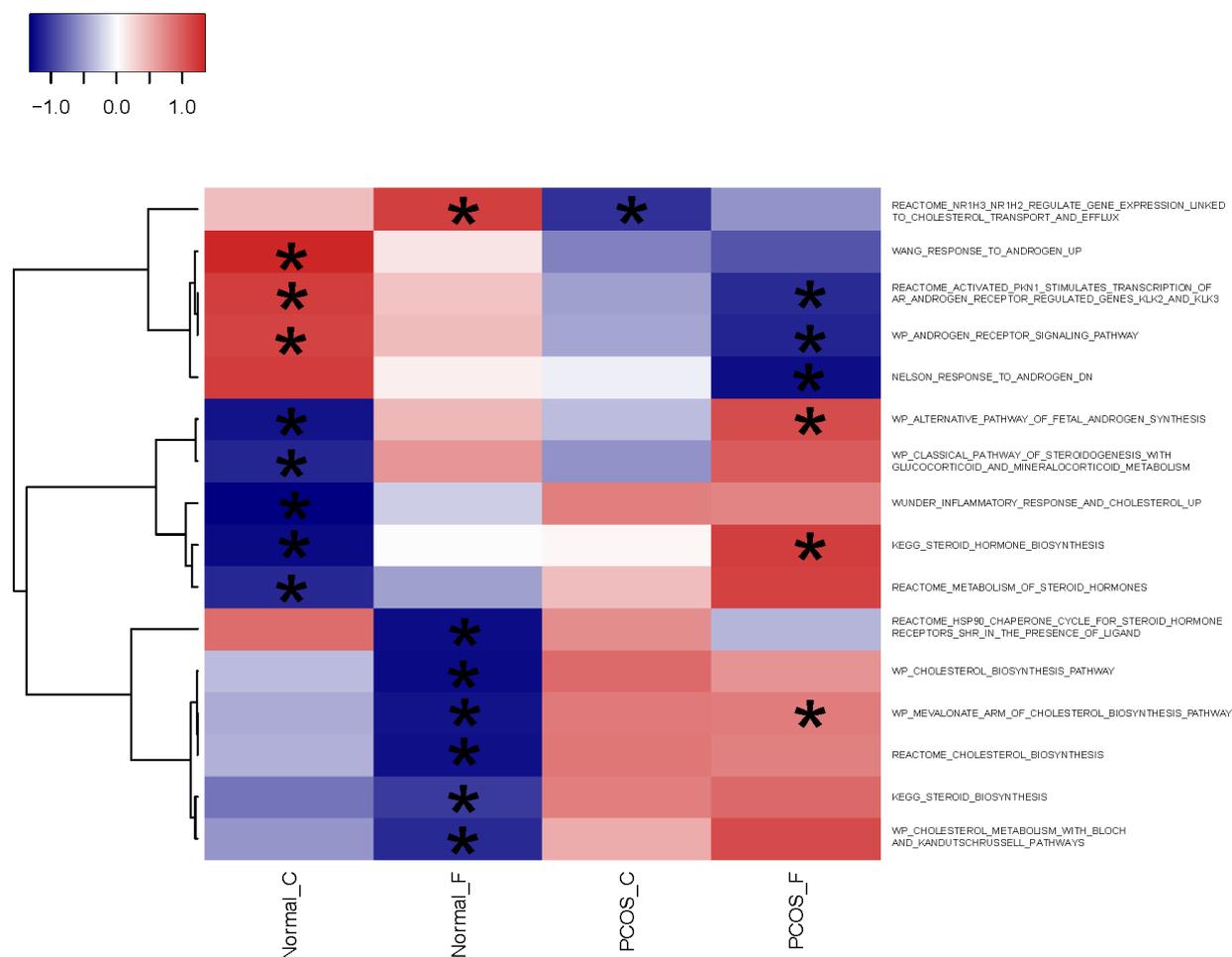


Figure 2. Heatmap of the Z scores for cholesterol acquisition and steroidogenesis pathways from molecular signatures database (MSigDB) C2: curated gene sets significantly enriched in at least one group. Asterisks show the significant differentially expressed group(s). scRNA-seq data were generated using theca cell preparations from normal cycling and PCOS women grown until subconfluent and either untreated (C) or treated (F) with 20 μ M forskolin for 24 h in a serum-free medium.

Among the individual DEGs in significant differentially expressed pathways, a number are genes that have been previously suggested as having a role in PCOS. There are 18 significant genes in cholesterol acquisition pathways (Table S2) and 31 significant genes in steroidogenesis pathways that are potential drivers of the thecal PCOS phenotype (Table S3).

In addition to genes previously suggested to be involved in producing the hyperandrogenemia phenotype, we identified several other DEGs that could potentially be involved in the pathogenesis of PCOS. There were 41 potential novel candidates in cholesterol acquisition pathways (Table S2) and 96 potential novel candidates in steroidogenesis pathways (Table S3).

We also examined significant DEGs not annotated in these pathways but likely to play a role in those functions (Table S4). The cholesterol acquisition genes *LDLR*, *LIPA*, *NPC1*, and *SCARB1* were significantly (Bonferroni corrected $p < 0.05$) differentially expressed. They are involved in the uptake of and processing of lipoprotein-carried cholesterol. Low-density lipoprotein cholesterol esters, taken into cells by LDLR, are hydrolyzed by LIPA in lysosomes, and free cholesterol is released from lysosomes by NPC1 and then trafficked to various subcellular compartments, including the mitochondria where the first committed step in steroid hormone synthesis takes place, cholesterol side-chain cleavage [30,31]. The steroidogenesis-associated genes encoding transcription factors *GATA4*, *GATA6*, *NR5A1*,

and *NR5A2* were significantly (Bonferroni corrected $p < 0.05$) differentially expressed, as was *CITED2*.

In Figure 3, we highlight a subset of significant DEGs either in differentially expressed pathways or among the additional genes we added that are likely to play a major role in cholesterol acquisition and steroidogenesis. These genes are regulated by two transcription factors that control genes involved in cholesterol acquisition and metabolism (*SREBF1*) and steroidogenesis (*GATA6*), both of which were found to be elevated in PCOS theca cells.

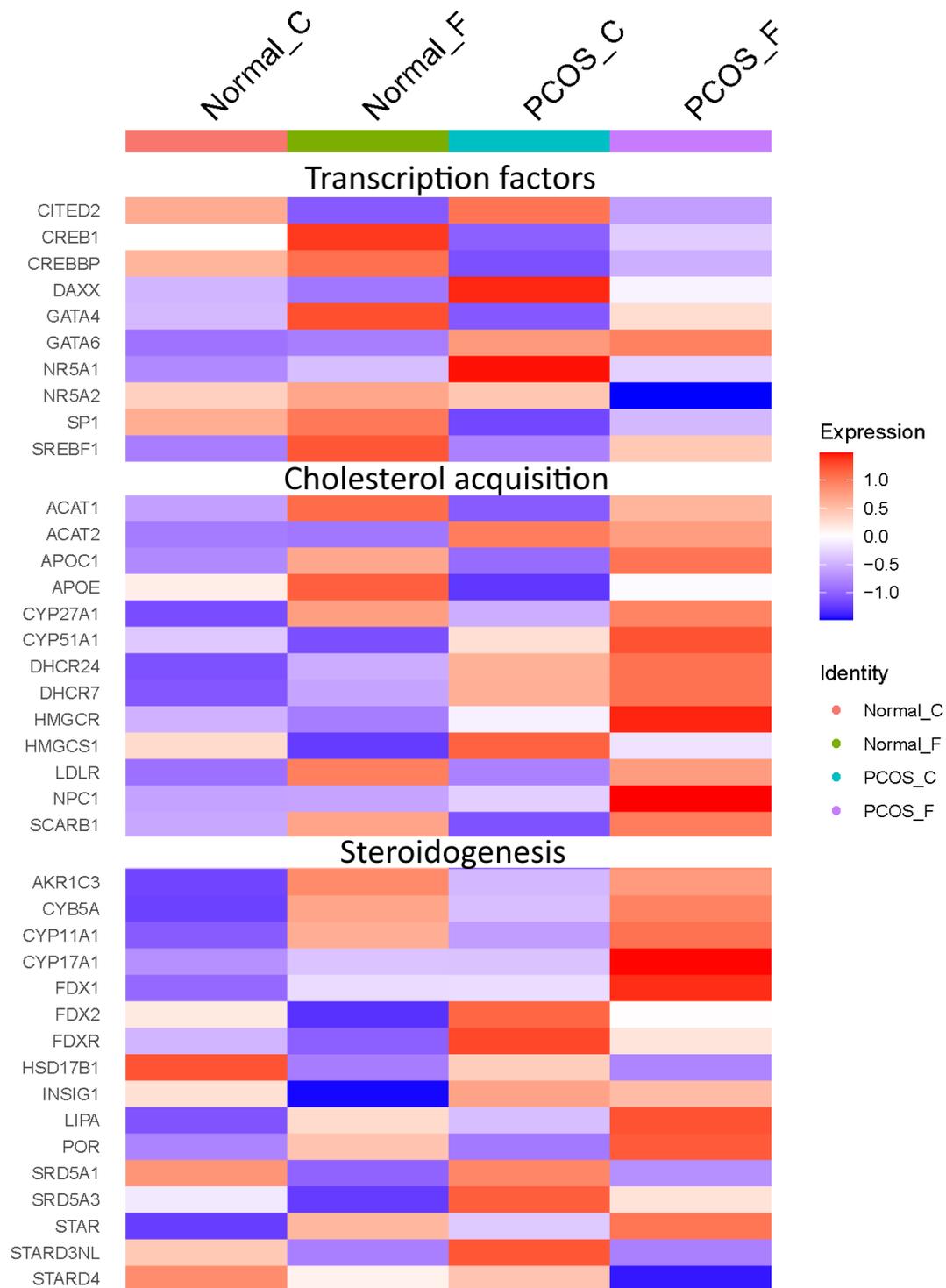


Figure 3. Heatmap of significant differentially expressed genes likely to play a major role in cholesterol acquisition and steroidogenesis.

In order to examine associations among proteins translated from DEGs identified by scRNA-seq analysis, we use the STRING [32] protein–protein interaction database. Figure 4 is a STRING interaction network visualization of the 13 proteins translated from DEGs under the header “Cholesterol acquisition” in the Figure 3 heatmap and the transcription factor SREBF1. SREBF1 is directly associated with 10 of the cholesterol acquisition proteins, which excludes APOC1, ACAT1, and ACAT2. HMGCR, the rate-limiting enzyme in de novo cholesterol biosynthesis, is directly associated with all the proteins in the network except for APOC1.

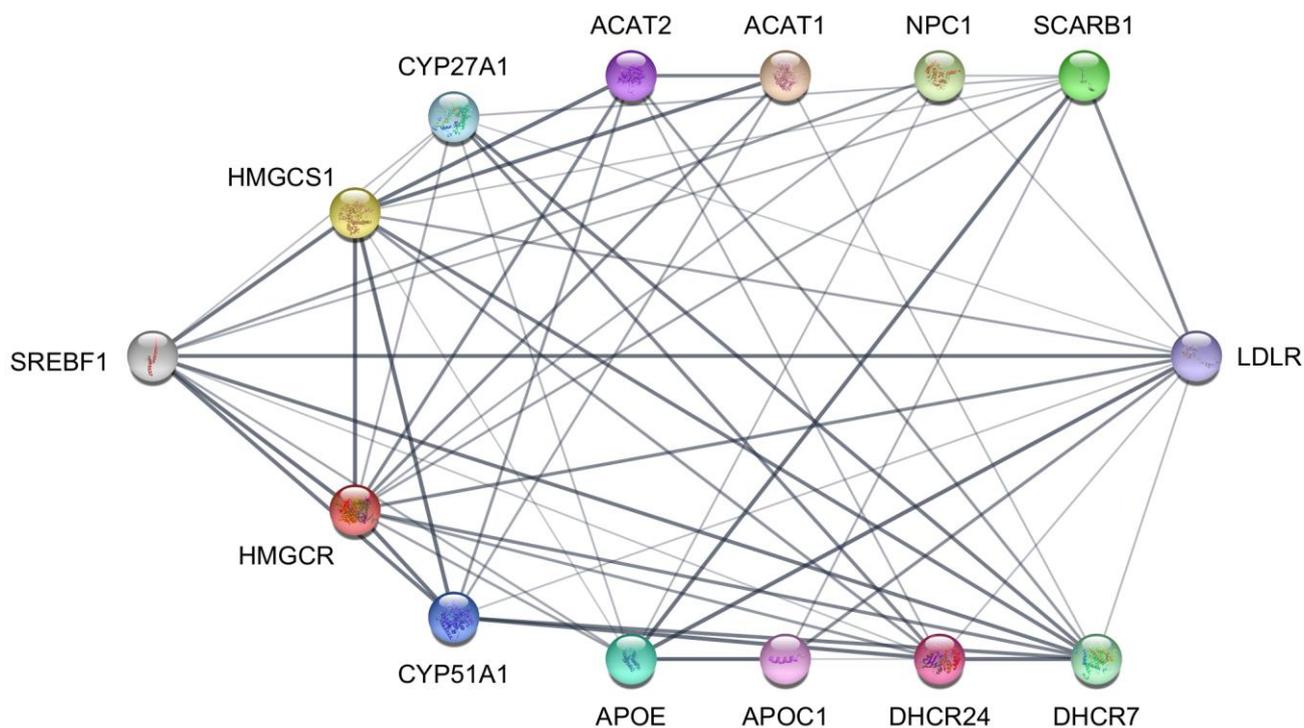


Figure 4. STRING protein–protein interaction network of cholesterol acquisition proteins translated from differentially expressed genes identified by scRNA-seq of normal and PCOS theca cells. The nodes represent the proteins, and the width of the edges shows the confidence of the association between connected proteins.

Figure 5 is a STRING interaction network visualization of the 16 proteins translated from scRNA-seq identified DEGs under the header “Steroidogenesis” in the Figure 3 heatmap together with the transcription factor GATA6. GATA6 shows direct associations with CYP17A1 and HSD17B1. CYP17A1 is directly associated with all the proteins in the network except for LIPA, INSIG1, FDX1L, STARD4, and STARD3NL. FDX1L is indirectly associated with CYP17A1 through FDXR and FDX1. STARD4 and STARD3NL are indirectly associated with CYP17A1 through STAR.

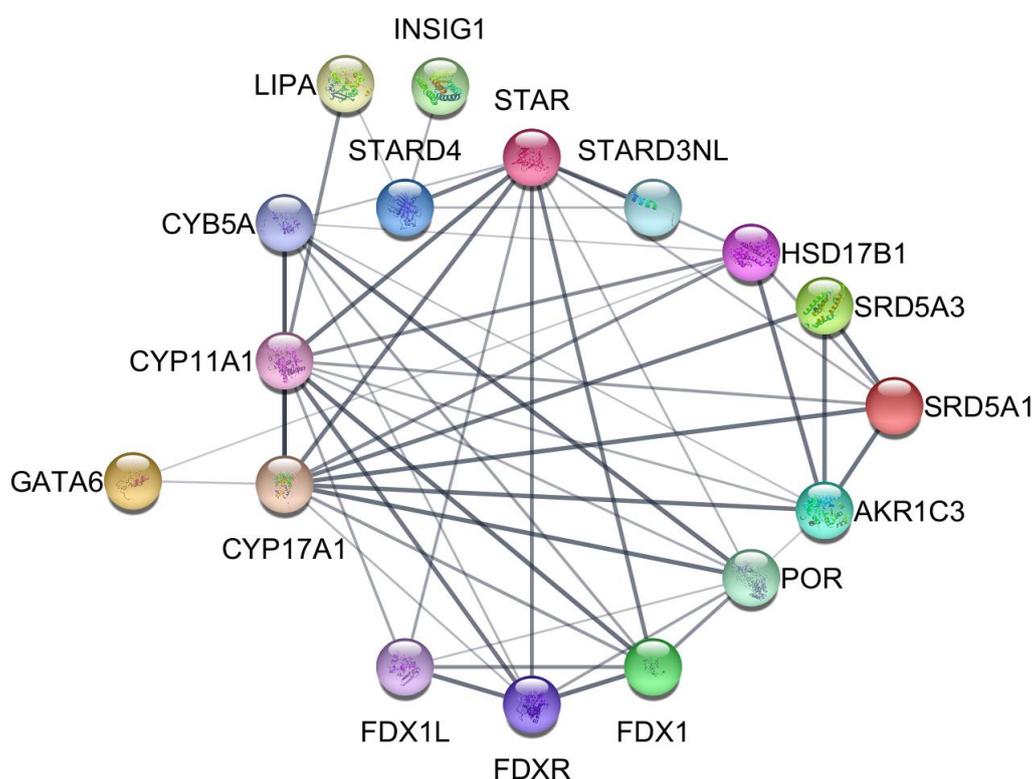


Figure 5. STRING protein–protein interaction network of steroidogenesis proteins translated from significant differentially expressed genes identified by scRNA-seq of normal and PCOS theca cells. The nodes represent the proteins, and the width of the edges shows the confidence of the association between connected proteins.

2.2. Bulk RNA-Seq Confirms a Distinctive Repertoire of Gene Expression in PCOS Theca Cells

We performed bulk RNA-seq on theca cells from normal and PCOS women treated with forskolin. In contrast to scRNA-seq, which characterizes the transcriptomes of individual cells, bulk RNA-seq assesses average gene expression across a large population of cells. We then used Cufflinks [33] to identify significant (q value < 0.05) DEGs and compared the bulk RNA-seq DEGs to scRNA-seq DEGs from pairwise comparisons of normal and PCOS theca cells treated with forskolin (Table S5). In this comparison, 86 significant DEGs were shared between scRNA-seq and bulk RNA-seq and showed fold changes in the same direction, including *GATA6* and *CYP17A1*. The findings confirmed the differential expression of genes involved in cholesterol acquisition and androgen biosynthesis.

2.3. Differential Gene Expression Identified by scRNA-Seq Is Consistent with the PCOS Molecular Signature Detected by Microarrays

We compared DEGs discovered in our scRNA-seq dataset to our previously published microarray data obtained from the analysis of normal and PCOS theca cells cultured without and with forskolin [34]. In forskolin-treated theca cells, 143 DEGs were shared between scRNA-seq and microarray datasets and showed fold changes in the same direction (Table S5). A total of 15 DEGs were shared across scRNA-seq, microarray, and bulk RNA-seq datasets with fold changes in the same direction, including *GATA6* and *CYP17A1*. There were 128 PCOS candidates shared only between scRNA-seq and microarray from forskolin-treated theca cells, including *HMGCR*, *FDX1*, *CYB5A*, and *CITED2*. In untreated theca cells, 100 DEGs were shared between scRNA-seq and microarray datasets and showed fold changes in the same direction, including the PCOS candidate *STAR* (Table S6).

3. Discussion

Our study provides new insights into the molecular phenotype of PCOS theca cells. Our findings suggest that theca cells from women with PCOS have an enhanced capacity to acquire cholesterol and metabolize it into androgens, which disrupt the function of the ovarian–hypothalamic–pituitary axis, leading to anovulation and infertility, as well as promoting metabolic dysfunction. In addition, our study identified several novel candidate genes that may be involved in the pathogenesis of PCOS. Future studies are needed to validate these findings and to elucidate the functional roles of these genes in PCOS. Of note, genes involved in the acquisition of cholesterol, the precursor of ovarian steroid hormones, including those involved in de novo biosynthesis or uptake of lipoprotein-carried cholesterol, were up-regulated in PCOS theca cells, especially when they were stimulated with forskolin. Many of these genes are transcribed in response to the activation of the transcription factor SREBF1. Additionally, the expression of genes encoding steroidogenic proteins catalyzing the metabolism of cholesterol into pregnenolone and subsequently into androgens was elevated in PCOS theca cells [12,13,35]. Of note, genes encoding proteins that increase the 17/20-lyase activity of CYP17A1, which converts pregnenolone into DHEA (*POR*, *CYB5A*), were significantly up-regulated.

Because cholesterol is the precursor of steroid hormones, there is an expected overlap among the proteins depicted in the STRING interaction networks presented in Figure 4 (Cholesterol acquisition) and Figure 5 (Steroidogenesis). For example, *LIPA* encodes the lysosomal acid lipase involved in the hydrolysis of LDL-carried cholesteryl esters, thus generating free cholesterol that can be trafficked to the mitochondria, where the first committed step in steroidogenesis takes place. Mutations inactivating *LIPA* result in lysosomal accumulation of cholesteryl esters (Wolman disease), which damages steroidogenic cells, reducing steroidogenesis [36,37].

Transcription factors controlling theca cell cholesterol acquisition (SREBF1) [38] and androgen biosynthesis (GATA6) were significantly increased in PCOS theca cells [39,40], especially when they were stimulated with forskolin. Notably, statin drugs, which inhibit the HMGCR protein product 3-hydroxy-3-methyl-glutaryl-CoA (HMG-CoA) reductase, the rate-limiting enzyme in cholesterol biosynthesis, impact the growth of human theca cells and reduce androgen levels in women with PCOS, consistent with a key role for cholesterol metabolism in the theca cell dysfunction in PCOS [41]. In addition to GATA6 and SREBF1, other transcription factors are involved in differential gene expression. For example, GATA6 activates *CYP17A1* transcription by binding to another differentially expressed transcription factor, SP1, that interacts with the *CYP17A1* proximal promoter [42,43].

It should be noted that differential gene expression detected by RNA-seq or microarrays can reflect differences in gene transcription as well as differences in mRNA half-life. For example, we have previously shown that in addition to increased transcription, *GATA6* [40], *CYP11A1* [44], and *CYP17A1* [45] gene expression is augmented in PCOS theca cells by increased mRNA stability. Thus, transcriptional and post-transcriptional mechanisms contribute to the increased transcript levels encoded by these genes in PCOS theca cells.

We have employed three different methods (scRNA-seq, bulk RNA-seq, and microarray methods) to analyze the molecular phenotype of normal and PCOS theca cells. Overall, the findings are congruent with increased expression of genes encoding proteins involved in cholesterol biosynthesis or uptake and metabolism to androgens. Each of the methods we have employed has strengths and limitations. For example, a limitation of the scRNA-seq approach to profiling gene expression is that most scRNA-seq methods identify DEGs at the gene level but not at the isoform level due to the limitations of current single-cell sequencing technologies and software analysis tools [46]. Consequently, the scRNA-seq method we employed could fail to detect variation in the expression of some isoforms that play important pathophysiologic roles in PCOS. Microarray studies can fail to detect expressed genes not represented in the nucleotide arrays, in addition to failing to detect splice variants if the arrayed nucleotide sequences are not complementary to sequences in

the variant transcripts. Bulk RNA-seq, while capable of identifying splice variants, is less sensitive and may fail to detect variants present in low abundance or not represented in genomic databases [47].

As an example, we previously demonstrated that DENND1A variant transcript 2 (DENND1A.V2) is elevated in PCOS theca cells (by qPCR and western blotting). That forced expression of DENND1A.V2 in normal theca cells augments androgen production, and conversely, that DENND1A.V2 mRNA knock-down in PCOS theca cells diminishes androgen production [7]. However, DENND1A.V2 was not identified as significantly differentially expressed in our scRNA-seq or microarray studies but was increased in bulk RNA-seq experiments, although the increase was not statistically significant. We attribute the failure to detect differential expression of this variant transcript, which encodes a major PCOS candidate gene, to the limitations of the abovementioned methods.

DENND1A.V2 encodes a protein associated with coated pits, where gonadotropin receptors cluster to initiate signal transduction followed by endocytosis and vesicular trafficking through the cytoplasm [7]. The DENND1A.V2 traffic pattern includes translocation into the nucleus, where we have proposed that it plays a role in initiating the transcription of genes involved in androgen biosynthesis [7,48]. Our scRNA-seq findings suggest that DENND1A.V2 could trigger the expression of SREBF1 and GATA6, which in turn, could account for the DENND1A.V2-driven increase in theca cell androgen production through transcriptional activation of genes encoding steroidogenic enzymes.

4. Materials and Methods

4.1. Theca Cell Cultures

Human theca interna tissue was obtained from follicles of women undergoing hysterectomy, following informed consent under a protocol approved by the Institutional Review Board of The Pennsylvania State University College of Medicine. As a standard of care, oophorectomies were performed during the luteal phase of the cycle. Theca cells from normal cycling and PCOS follicles were isolated and grown as previously reported in detail [7,20,40,49,50]. PCOS and normal ovarian tissue came from age-matched women, 38–41 years old. The diagnosis of PCOS was made according to the National Institutes of Health (NIH) consensus guidelines [9,10], which include hyperandrogenemia/hyperandrogenism and oligo-ovulation and the exclusion of other causes of hyperandrogenemia (e.g., 21-hydroxylase deficiency, Cushing's syndrome, and adrenal or ovarian tumors). All PCOS theca cell preparations studied came from the ovaries of women with fewer than six menses per year and elevated serum total testosterone or bioavailable testosterone levels [11,13,14]. Each PCOS ovary contained multiple subcortical follicles less than 10 mm in diameter. The control (normal) theca cell preparations came from ovaries of fertile women with normal menstrual histories, menstrual cycles of 21–35 days, and no clinical signs of hyperandrogenism. Neither PCOS nor normal subjects were receiving hormonal medications at the time of surgery. Indications for surgery were dysfunctional uterine bleeding, endometrial cancer, and pelvic pain. Experiments comparing PCOS and normal theca were performed using fourth-passage (31–38 population doublings) theca cells isolated from individual size-matched follicles obtained from age-matched subjects in the absence of *in vivo* stimulation. The use of fourth-passage cells allowed us to perform multiple experiments from the same patient population that were propagated from frozen stocks of second-passage cells in the media described above. The passage conditions and split ratios for all normal and PCOS cells were identical. These studies were approved by the Human Subjects Protection Offices of Virginia Commonwealth University and Penn State College of Medicine.

4.2. Single-Cell RNA Sequencing (scRNA-Seq)

For scRNA-seq studies, fourth passage theca cells were grown and propagated from five theca cell preparations obtained from women with PCOS (MC03, MC10, MC16, MC26, and MC27), and five theca cell preparations from normal cycling women (MC02, MC06,

MC31, MC40, and MC50), in a serum-containing medium, in a 5% O₂/5% CO₂ incubator, as we have previously described [7,20,40,49,50]. All of the subjects we recruited were unrelated and of European ancestry. Once all of theca cell preparations were approximately 85–95% confluent, the cells were rinsed and transferred to serum-free defined media and treated under basal control conditions or treated with forskolin (20 μM), an activator of adenylate cyclase for 24 h. Following treatment, the cells were rinsed with 5mL of PBS and trypsinized and harvested in a cold 1:1 DME/F12 medium containing 10% FBS, as we have previously described [7,20,40,49,50]. The cells were then centrifuged at 600× g, and the cells were resuspended according to Active Motif's (Carlsbad, CA, USA) protocol in 20% DMSO and 50% FBS in 1:1 DME/F12 and frozen in 2 mL cryovials using a Corning isopropanol –80 degree cryofreezing apparatus for 24 h. Afterward, the cells were transferred and stored in a ThermoFisher liquid N₂ cell storage unit (Waltham, MA, USA) until all of the theca cell preparations were processed and frozen in lq N₂ stable cryovials, prior to sending the whole frozen cell preparation to Active Motif (Carlsbad, CA, USA) for subsequent cell dilution, RNA isolation, and single-cell library preparation, using Active Motifs proprietary conditions. Following single-cell library preparation, 10× single-cell RNA sequencing (scRNA-seq) was performed using an Illumina NextSeq 500 (San Diego, CA, USA) sequencing apparatus to generate 91 bp sequencing reads.

4.3. Bulk RNA Sequencing (Bulk RNA-Seq)

For bulk RNA-seq experiments, fourth passage theca cells were gown and propagated from four theca cell preparations obtained from women with PCOS (MC03, MC10, MC26, and MC27) and four theca cell preparations from normal cycling women (MC02, MC31, MC40, and MC50), in a serum-containing medium, in a 5% O₂ and 5% N₂ incubator, as we have previously described [7,20,40,49,50]. As described above for scRNA-seq, once the cells were 85–95% confluent, all of the cells were rinsed with PBS, and transferred into a serum-free medium in the presence of 20 mM forskolin for 16 h. Following treatment, the cells were rinsed in PBS, and after complete aspiration, the cell dishes were tightly wrapped in foil, labeled, and flash frozen in Liq N₂ prior to storage in a –80 degree ThermoFisher freezer (Waltham, MA, USA). Once all the cells were treated and collected, we harvested RNA using Sigma Tri-reagent, using Sigma's standard protocol for cells in monolayer culture, as we have previously described [7,20,40,49,50], following isopropanol precipitation and centrifugation at 10× g in a 4 degrees C microfuge for 30 min. The resulting total RNA cell pellet was rinsed in 70% ethanol and re-centrifuged at 10× g in a Beckman 4 degrees C microfuge for 20 min. The resulting total RNA pellet was resuspended in 65 mL RNAase-free ultrapure water, OD'd, and aliquoted into 5–10 mL aliquots, then frozen in a –80 freezer prior to transfer to ThermoScientific (previously Life Sciences) (Waltham, MA, USA) for library preparation for bulk RNA-seq. The bulk RNA-seq was performed by sequencing paired-end 100 bp reads at a depth of 100 million reads using Illumina HiSeq (San Diego, CA, USA).

4.4. Cell Ranger Data Processing

Cell Ranger [21] v.6.1.1 was used to process the raw scRNA-Seq data. The Cell Ranger mkfastq function was run with default parameters to demultiplex raw base call (BCL) files generated by Illumina sequencers into FASTQ files. The Cell Ranger count function was run with default parameters to perform genome alignment, filtering, barcode counting, and unique molecular identifier (UMI) counting that assigns each read to its originating cell and transcript. The transcript counts were then aggregated to generate a matrix of gene expression values for each cell.

4.5. Seurat Pre-Processing and Quality Control

The Seurat [22] v.4.1.1 package loaded into R v.4.1.3 performed standard pre-processing and quality control on the scRNA-Seq data. For each sample, Cell Ranger output files were read using Read10×. The percent of mitochondrial genes was calculated, and cells with

>5% mitochondrial genes were removed. Cells with <200 RNA features were also removed. The Seurat functions `NormalizeData` and `FindVariableFeatures` were then run using default parameters.

4.6. STACAS Integration

The Sub-Type Anchoring Correction for Alignment in Seurat (STACAS v.2.0.1) [23] package loaded into R v.4.1.3 was used to perform semi-supervised integration of the individual theca cell scRNA-seq datasets. Highly variable genes (HVG) were identified, excluding blacklist genes consisting of mitochondrial, ribosomal, heat-shock, and cell cycle genes. PCOS candidate genes and other genes known to be expressed in the ovary were selected from the HVG genes as features for integration. Integration was performed using the `Run.STACAS` function with default parameters except for `anchor.features` set to the ovarian genes, `cell.labels` set to the known affection status and treatment of the cell, and `label.confidence` set to 0.99. Setting the `cell.labels` parameter causes STACAS to perform semi-supervised integration by using this information to penalize anchors where cell types are inconsistent.

4.7. UMAP Generation

Uniform Manifold Approximation and Projection (UMAP) visualizations were generated using the Seurat `RunUMAP` function with default parameters except for `n.neighbors` = 50, `min.dist` = 0.5 to optimize for global structure. The initial UMAP visualization was used to identify 5210 outlier cells that were removed.

4.8. Identification of Differentially Expressed Genes

The Seurat `FindMarkers` function was used to identify differentially expressed genes across the groups based on known affection status and treatment of the cell. `FindMarkers` was run with default parameters except for `min.pct` = 0.01 and `logfc.threshold` = 0.01 to allow for more sensitive detection of differential expression. This function calculates a Wilcoxon rank-sum test for each gene to compare its expression level in each individual group to all the other groups combined. The test results were then adjusted for multiple testing using Bonferroni correction.

4.9. Gene Set Enrichment Analysis

The `singleseqset` [25] v.0.1.2.9000 package loaded into R v.4.1.3 was used to perform Gene Set Enrichment Analysis (GSEA). The `logFC` function was used to calculate the log fold change between each group and all the other groups combined for all genes using log-normalized data to correct for library size. The `gse.res` function was used to calculate enrichment scores and perform Wilcoxon rank sum tests, followed by Benjamini–Hochberg FDR correction, on the gene sets. The gene sets used were the Molecular Signatures Database (MSigDB) C2: curated gene sets [24], which are curated from sources, including online pathway databases, biomedical literature, and contributions by individual domain experts.

4.10. STRING Protein–Protein Interaction Database and Network Visualization

The STRING [32] protein–protein interaction database was used to identify associations between proteins translated from differentially expressed genes. Protein–protein interaction networks were generated and visualized using the Cytoscape `stringApp` plugin [51].

4.11. Bulk RNA Sequencing Analysis

Tophat [33] v.1.4.1 was used to align RNA-seq reads to the human (GRCh37/hg19) genome assembly. Cufflinks [33] v.2.1.1 was used to perform differential expression analysis.

5. Conclusions

Our study demonstrates the feasibility and utility of performing scRNA-seq on theca cells to identify molecular pathways and genes affected by PCOS. Differentially expressed genes in PCOS theca cells included those involved in the acquisition of cholesterol, the precursor of steroid hormones, and genes involved in the biosynthesis of androgens. The transcription factors SREBF1, which regulates genes involved in cholesterol acquisition, and GATA6, which regulates the expression of genes encoding steroidogenic enzymes (CYP17A1, CYP11A1) in concert with other differentially expressed transcription factors, appear to play a major role in driving the PCOS theca cell molecular signature. Our findings provide new insights into the molecular mechanisms underlying theca cell dysfunction in PCOS and identify potential targets for therapeutic intervention. scRNA-seq of theca cells may also have diagnostic and prognostic applications for PCOS.

Supplementary Materials: The supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms241310611/s1/>.

Author Contributions: Conceptualization, J.M.M. and J.F.S.III; methodology, R.A.H., J.M.M. and J.F.S.III; software, R.A.H.; formal analysis, R.A.H.; investigation, R.A.H., J.M.M. and J.F.S.III; resources, J.M.M. and J.F.S.III; data curation, R.A.H.; writing—original draft preparation, R.A.H., J.M.M. and J.F.S.III; writing—review and editing, R.A.H., J.M.M. and J.F.S.III; visualization, R.A.H. and J.F.S.III; supervision, J.M.M. and J.F.S.III; project administration, J.M.M. and J.F.S.III; funding acquisition, J.M.M. and J.F.S.III. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Institutes of Health, grant numbers R01HD083323 (J.M.M. and J.F.S.) and R01HD033852 (J.M.M.).

Institutional Review Board Statement: The study was conducted following the Declaration of Helsinki and was approved by the Institutional Review Board of The Pennsylvania State University College of Medicine (STUDY00007086, Title: Acquisition of Human Specimens to be Used in Studying the Molecular, Genetic, and Epigenetic Basis for Increased Androgen Biosynthesis and Other Associated Sequelae in Polycystic Ovary Syndrome (PCOS). This approval is effective through 16 November 2023 inclusive).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The 10× Cell Ranger filtered feature-barcode matrix data files and the scRNA-seq data are publicly available for download at <https://zenodo.org/record/7942968> (accessed on 21 May 2023). These files are the read counts by gene by cell that can be used to analyze single cell data.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Azziz, R.; Dumesic, D.A.; Goodarzi, M.O. Polycystic Ovary Syndrome: An Ancient Disorder? *Fertil. Steril.* **2011**, *95*, 1544–1548. [[CrossRef](#)]
2. Balen, A.; Homburg, R.; Franks, S. Defining Polycystic Ovary Syndrome. *BMJ* **2009**, *338*, 426. [[CrossRef](#)] [[PubMed](#)]
3. Franks, S.; Webber, L.J.; Goh, M.; Valentine, A.; White, D.M.; Conway, G.S.; Wiltshire, S.; McCarthy, M.I. Ovarian Morphology Is a Marker of Heritable Biochemical Traits in Sisters with Polycystic Ovaries. *J. Clin. Endocrinol. Metab.* **2008**, *93*, 3396–3402. [[CrossRef](#)]
4. Franks, S.; Stark, J.; Hardy, K. Follicle Dynamics and Anovulation in Polycystic Ovary Syndrome. *Hum. Reprod. Update* **2008**, *14*, 367–378. [[CrossRef](#)]
5. McAllister, J.M.; Legro, R.S.; Modi, B.P.; Strauss, J.F. Functional Genomics of PCOS: From GWAS to Molecular Mechanisms. *Trends Endocrinol. Metab.* **2015**, *26*, 118–124. [[CrossRef](#)] [[PubMed](#)]
6. Dapas, M.; Dunaif, A. Deconstructing a Syndrome: Genomic Insights into PCOS Causal Mechanisms and Classification. *Endocr. Rev.* **2022**, *43*, 927–965. [[CrossRef](#)]
7. McAllister, J.M.; Modi, B.; Miller, B.A.; Biegler, J.; Bruggeman, R.; Legro, R.S.; Strauss, J.F. Overexpression of a DENND1A Isoform Produces a Polycystic Ovary Syndrome Theca Phenotype. *Proc. Natl. Acad. Sci. USA* **2014**, *111*. [[CrossRef](#)] [[PubMed](#)]
8. Azziz, R. PCOS in 2015: New Insights into the Genetics of Polycystic Ovary Syndrome. *Nat. Rev. Endocrinol.* **2016**, *12*, 74. [[CrossRef](#)]

9. Azziz, R. Introduction: Determinants of Polycystic Ovary Syndrome. *Fertil. Steril.* **2016**, *106*, 4–5. [[CrossRef](#)]
10. Legro, R.S.; Arslanian, S.A.; Ehrmann, D.A.; Hoeger, K.M.; Murad, M.H.; Pasquali, R.; Welt, C.K. Diagnosis and Treatment of Polycystic Ovary Syndrome: An Endocrine Society Clinical Practice Guideline. *J. Clin. Endocrinol. Metab.* **2013**, *98*, 4565–4592. [[CrossRef](#)]
11. McAllister, J.M.; Kerin, J.F.P.; Trant, J.M.; Estabrook, R.W.; Mason, J.I.; Waterman, M.R.; Simpson, E.R. Regulation of Cholesterol Side-Chain Cleavage and 17 Alpha-Hydroxylase/Lyase Activities in Proliferating Human Theca Interna Cells in Long Term Monolayer Culture. *Endocrinology* **1989**, *125*, 1959–1966. [[CrossRef](#)]
12. Wickenheisser, J.K.; Nelson-DeGrave, V.L.; McAllister, J.M. Human Ovarian Theca Cells in Culture. *Trends Endocrinol. Metab.* **2006**, *17*, 65–71. [[CrossRef](#)] [[PubMed](#)]
13. Nelson, V.L.; Legro, R.S.; Strauss, J.F., 3rd; McAllister, J.M. Augmented Androgen Production Is a Stable Steroidogenic Phenotype of Propagated Theca Cells from Polycystic Ovaries. *Mol. Endocrinol.* **1999**, *13*, 946–957. [[CrossRef](#)]
14. McAllister, J.M. Functional, Long-Term Human Theca and Granulosa Cell Cultures from Polycystic Ovaries. *Endocrine* **1995**, *3*, 143–149. [[CrossRef](#)]
15. Comim, F.V.; Teerds, K.; Hardy, K.; Franks, S. Increased Protein Expression of LHCG Receptor and 17 α -Hydroxylase/17-20-Lyase in Human Polycystic Ovaries. *Hum. Reprod.* **2013**, *28*, 3086–3092. [[CrossRef](#)] [[PubMed](#)]
16. Gilling-Smith, C.; Willis, D.S.; Beard, R.W.; Franks, S. Hypersecretion of Androstenedione by Isolated Thecal Cells from Polycystic Ovaries. *J. Clin. Endocrinol. Metab.* **1994**, *79*, 1158–1165. [[CrossRef](#)]
17. Gilling-Smith, C.; Story, H.; Rogers, V.; Franks, S. Evidence for a Primary Abnormality of Thecal Cell Steroidogenesis in the Polycystic Ovary Syndrome. *Clin. Endocrinol.* **1997**, *47*, 93–99. [[CrossRef](#)] [[PubMed](#)]
18. Mcallister, J.M.; Han, A.X.; Modi, B.P.; Teves, M.E.; Mavodza, G.R.; Anderson, Z.L.; Shen, T.; Christenson, L.K.; Archer, K.J.; Strauss, J.F. MiRNA Profiling Reveals MiRNA-130b-3p Mediates DENND1A Variant 2 Expression and Androgen Biosynthesis. *Endocrinology* **2019**, *160*, 1964–1981. [[CrossRef](#)]
19. Alan Harris, R.; Archer, K.J.; Goodarzi, M.O.; York, T.P.; Rogers, J.; Dunaif, A.; McAllister, J.M.; Strauss, J.F. Loci on Chromosome 12q13.2 Encompassing ERBB3, PA2G4 and RAB5B Are Associated with Polycystic Ovary Syndrome. *Gene* **2023**, *852*, 147062. [[CrossRef](#)] [[PubMed](#)]
20. Waterbury, J.S.; Teves, M.E.; Gaynor, A.; Han, A.X.; Mavodza, G.; Newell, J.; Strauss, J.F.; McAllister, J.M. The PCOS GWAS Candidate Gene ZNF217 Influences Theca Cell Expression of DENND1A.V2, CYP17A1, and Androgen Production. *J. Endocr. Soc.* **2022**, *6*, bvac078. [[CrossRef](#)]
21. Zheng, G.X.Y.; Terry, J.M.; Belgrader, P.; Ryvkin, P.; Bent, Z.W.; Wilson, R.; Ziraldo, S.B.; Wheeler, T.D.; McDermott, G.P.; Zhu, J.; et al. Massively Parallel Digital Transcriptional Profiling of Single Cells. *Nat. Commun.* **2017**, *8*, 14049. [[CrossRef](#)]
22. Hao, Y.; Hao, S.; Andersen-Nissen, E.; Mauck, W.M.; Zheng, S.; Butler, A.; Lee, M.J.; Wilk, A.J.; Darby, C.; Zager, M.; et al. Integrated Analysis of Multimodal Single-Cell Data. *Cell* **2021**, *184*, 3573–3587. [[CrossRef](#)]
23. Andreatta, M.; Carmona, S.J. STACAS: Sub-Type Anchor Correction for Alignment in Seurat to Integrate Single-Cell RNA-Seq Data. *Bioinformatics* **2021**, *37*, 882–884. [[CrossRef](#)] [[PubMed](#)]
24. Liberzon, A.; Subramanian, A.; Pinchback, R.; Thorvaldsdóttir, H.; Tamayo, P.; Mesirov, J.P. Molecular Signatures Database (MSigDB) 3.0. *Bioinformatics* **2011**, *27*, 1739–1740. [[CrossRef](#)] [[PubMed](#)]
25. Cillo, A. Singleseqgset 2019. Available online: <https://github.com/arc85/singleseqgset> (accessed on 19 June 2023).
26. Auchus, R.J. Steroid 17-Hydroxylase and 17,20-Lyase Deficiencies, Genetic and Pharmacologic. *J. Steroid Biochem. Mol. Biol.* **2017**, *165*, 71–78. [[CrossRef](#)]
27. Xu, Y.; Jiang, S.; Yan, Z.; Niu, Y.; Du, W.; Liu, B.; Han, B.; Liu, X.; Zhao, S.; Song, H.; et al. Phenotypic Heterogeneity and Fertility Potential of Patients with 17-Hydroxylase/17,20-Lyase Deficiency. *J. Clin. Endocrinol. Metab.* **2022**, *107*, E2610–E2618. [[CrossRef](#)]
28. Miller, W.L.; Tee, M.K. The Post-Translational Regulation of 17,20 Lyase Activity. *Mol. Cell. Endocrinol.* **2015**, *408*, 99–106. [[CrossRef](#)] [[PubMed](#)]
29. Lambert, S.A.; Jolma, A.; Campitelli, L.F.; Das, P.K.; Yin, Y.; Albu, M.; Chen, X.; Taipale, J.; Hughes, T.R.; Weirauch, M.T. The Human Transcription Factors. *Cell* **2018**, *172*, 650–665. [[CrossRef](#)]
30. Watari, H.; Blanchette-Mackie, E.J.; Dwyer, N.K.; Sun, G.; Glick, J.M.; Patel, S.; Neufeld, E.B.; Pentchev, P.G.; Strauss, J.F. NPC1-Containing Compartment of Human Granulosa-Lutein Cells: A Role in the Intracellular Trafficking of Cholesterol Supporting Steroidogenesis. *Exp. Cell Res.* **2000**, *255*, 56–66. [[CrossRef](#)]
31. Miller, W.L. Steroid Hormone Synthesis in Mitochondria. *Mol. Cell. Endocrinol.* **2013**, *379*, 62–73. [[CrossRef](#)] [[PubMed](#)]
32. Szklarczyk, D.; Kirsch, R.; Koutrouli, M.; Nastou, K.; Mehryary, F.; Hachilif, R.; Gable, A.L.; Fang, T.; Doncheva, N.T.; Pyysalo, S.; et al. The STRING Database in 2023: Protein-Protein Association Networks and Functional Enrichment Analyses for Any Sequenced Genome of Interest. *Nucleic Acids Res.* **2023**, *51*, D638–D646. [[CrossRef](#)] [[PubMed](#)]
33. Trapnell, C.; Roberts, A.; Goff, L.; Pertea, G.; Kim, D.; Kelley, D.R.; Pimentel, H.; Salzberg, S.L.; Rinn, J.L.; Pachter, L. Differential Gene and Transcript Expression Analysis of RNA-Seq Experiments with TopHat and Cufflinks. *Nat. Protoc.* **2012**, *7*, 562–578. [[CrossRef](#)] [[PubMed](#)]
34. Wood, J.R.; Nelson, V.L.; Ho, C.; Jansen, E.; Wang, C.Y.; Urbanek, M.; McAllister, J.M.; Mosselman, S.; Strauss, J.F. The Molecular Phenotype of Polycystic Ovary Syndrome (PCOS) Theca Cells and New Candidate PCOS Genes Defined by Microarray Analysis. *J. Biol. Chem.* **2003**, *278*, 26380–26390. [[CrossRef](#)] [[PubMed](#)]

35. Chaudhary, H.; Patel, J.; Jain, N.K.; Joshi, R. The Role of Polymorphism in Various Potential Genes on Polycystic Ovary Syndrome Susceptibility and Pathogenesis. *J. Ovarian Res.* **2021**, *14*, 125. [[CrossRef](#)]
36. Pericleous, M.; Kelly, C.; Wang, T.; Livingstone, C.; Ala, A. Wolman's Disease and Cholesteryl Ester Storage Disorder: The Phenotypic Spectrum of Lysosomal Acid Lipase Deficiency. *Lancet Gastroenterol. Hepatol.* **2017**, *2*, 670–679. [[CrossRef](#)]
37. Miller, W.L. Disorders in the Initial Steps of Steroid Hormone Synthesis. *J. Steroid Biochem. Mol. Biol.* **2017**, *165*, 18–37. [[CrossRef](#)] [[PubMed](#)]
38. Christenson, L.K.; Osborne, T.F.; McAllister, J.M.; Strauss, J.F. Conditional Response of the Human Steroidogenic Acute Regulatory Protein Gene Promoter to Sterol Regulatory Element Binding Protein-1a. *Endocrinology* **2001**, *142*, 28–36. [[CrossRef](#)]
39. Wood, J.R.; Ho, C.K.M.; Nelson-Degrave, V.L.; McAllister, J.M.; Strauss, J.F. The Molecular Signature of Polycystic Ovary Syndrome (PCOS) Theca Cells Defined by Gene Expression Profiling. *J. Reprod. Immunol.* **2004**, *63*, 51–60. [[CrossRef](#)]
40. Ho, C.K.M.; Wood, J.R.; Stewart, D.R.; Ewens, K.; Ankener, W.; Wickenheisser, J.; Nelson-Degrave, V.; Zhang, Z.; Legro, R.S.; Dunaif, A.; et al. Increased Transcription and Increased Messenger Ribonucleic Acid (mRNA) Stability Contribute to Increased GATA6 mRNA Abundance in Polycystic Ovary Syndrome Theca Cells. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 6596–6602. [[CrossRef](#)]
41. Chen, J.; Huang, C.; Zhang, T.; Gong, W.; Deng, X.; Liu, H.; Liu, J.; Guo, Y. The Effects of Statins on Hyperandrogenism in Women with Polycystic Ovary Syndrome: A Systematic Review and Meta-Analysis of Randomized Controlled Trials. *Reprod. Biol. Endocrinol.* **2021**, *19*, 189. [[CrossRef](#)]
42. Flück, C.E.; Miller, W.L. GATA-4 and GATA-6 Modulate Tissue-Specific Transcription of the Human Gene for P450c17 by Direct Interaction with Sp1. *Mol. Endocrinol.* **2004**, *18*, 1144–1157. [[CrossRef](#)]
43. Sewer, M.B.; Waterman, M.R. ACTH Modulation of Transcription Factors Responsible for Steroid Hydroxylase Gene Expression in the Adrenal Cortex. *Microsc. Res. Tech.* **2003**, *61*, 300–307. [[CrossRef](#)] [[PubMed](#)]
44. Wickenheisser, J.K.; Biegler, J.M.; Nelson-DeGrave, V.L.; Legro, R.S.; Strauss, J.F.; McAllister, J.M. Cholesterol Side-Chain Cleavage Gene Expression in Theca Cells: Augmented Transcriptional Regulation and mRNA Stability in Polycystic Ovary Syndrome. *PLoS ONE* **2012**, *7*, e48963. [[CrossRef](#)] [[PubMed](#)]
45. Wickenheisser, J.K.; Nelson-Degrave, V.L.; McAllister, J.M. Dysregulation of Cytochrome P450 17 α -Hydroxylase Messenger Ribonucleic Acid Stability in Theca Cells Isolated from Women with Polycystic Ovary Syndrome. *J. Clin. Endocrinol. Metab.* **2005**, *90*, 1720–1727. [[CrossRef](#)]
46. Arzalluz-Luqueángeles, Á; Conesa, A. Single-Cell RNAseq for the Study of Isoforms-How Is That Possible? *Genome Biol.* **2018**, *19*, 110. [[CrossRef](#)]
47. Zhang, C.; Zhang, B.; Lin, L.L.; Zhao, S. Evaluation and Comparison of Computational Tools for RNA-Seq Isoform Quantification. *BMC Genom.* **2017**, *18*, 583. [[CrossRef](#)] [[PubMed](#)]
48. Tee, M.K.; Speek, M.; Legeza, B.; Modi, B.; Teves, M.E.; McAllister, J.M.; Strauss, J.F.; Miller, W.L. Alternative Splicing of DENND1A, a PCOS Candidate Gene, Generates Variant 2. *Mol. Cell. Endocrinol.* **2016**, *434*, 25–35. [[CrossRef](#)]
49. Kulkarni, R.; Teves, M.E.; Han, A.X.; McAllister, J.M.; Strauss, J.F., 3rd. Colocalization of Polycystic Ovary Syndrome Candidate Gene Products in Theca Cells Suggests Novel Signaling Pathways. *J. Endocr. Soc.* **2019**, *3*, 2204–2223. [[CrossRef](#)] [[PubMed](#)]
50. Nelson, V.L.; Qin, K.N.; Rosenfield, R.L.; Wood, J.R.; Penning, T.M.; Legro, R.S.; Strauss, J.F.; McAllister, J.M. The Biochemical Basis for Increased Testosterone Production in Theca Cells Propagated from Patients with Polycystic Ovary Syndrome. *J. Clin. Endocrinol. Metab.* **2001**, *86*, 5925–5933. [[CrossRef](#)]
51. Doncheva, N.T.; Morris, J.H.; Gorodkin, J.; Jensen, L.J. Cytoscape StringApp: Network Analysis and Visualization of Proteomics Data. *J. Proteome Res.* **2019**, *18*, 623–632. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.