

# Identification of Key Pathways and Genes in Polycystic Ovary Syndrome (PCOS)

Dilara AGACIK, Maxim VAN ASSEL

Supervisors : Prof. dr. ir. Tim De Meyer, Prof. dr. ir. Jo Vandesompele

Counselors: Dr. ir. Louis Coussement, Ir. Menno Van Damme

Faculty of Science  
Ghent University

Project Applied High Throughput Analysis

Academic year: 2023 – 2024

## Abstract

Numerous data from omics-related research are easily accessible to the public. It may be feasible to get novel insights from combining several omics data sources and technologies that are not achievable with a single omics research. This article presents an overview of the molecular pathways underlying PCOS by the execution of two microarray profiling investigations, an RNA-seq, and a methylation analysis

**Keywords:** Omics, Multiomics, PCOS, Polycystic Ovarian Syndrome, RNA-seq, Microarray Expression Profiling, Methylation Profiling

## 1. Introduction

Polycystic Ovarian Syndrome (PCOS) stands as one of the most prevalent endocrine disorders significantly impacting the female population. In the absence of other particular diagnoses, PCOS, which is heterogeneous by nature, is characterized by a combination of signs and symptoms of androgen excess and ovarian dysfunction.[1] Diabetes mellitus, obesity, dyslipidemia, hypertension, anxiety, and depression are all more common in women with PCOS. PCOS affects women's health at every stage of life and continuing into the postmenopausal years.[2]

Additionally, risk factors for cardiovascular disease are more common in adolescents and reproductive age women with PCOS.[3] On the subject of PCOS, numerous molecular studies have been conducted; yet, data source integration continues to be difficult. Integrating data from omics research may offer novel insights on the molecular factors that contribute to PCOS.

### 1.1. Datasets

Based on PCOS data, this article aims to investigate the characteristics of PCOS vs healthy tissue. Selected datasets are shown below with their GEO Accession code.

#### 1.1.1. Expression profiling by array (GSE124226):

The first selected dataset contains gene expression data of Adipose stem cells from normal-weight of 4 PCOS and 4 control women. The data was obtained from Affymetrix Human Genome U133 Plus 2.0 Array.

#### 1.1.2. Expression profiling by high throughput sequencing (GSE216609):

The dataset is RNA Sequencing data of 4 PCOS and 3 control cells sample, aim of this study to compare the transcriptome expression of cumulus granulosa cells between PCOS patients with insulin resistance and women undergoing assisted reproduction for fallopian tube reasons. The data sequenced by Illumina HiSeq 3000 (Homo sapiens)

#### 1.1.3. Expression profiling by array (GSE87435):

The dataset is Microarray expression data derived from ovaries of 12 women with Polycystic Ovary Syndrome and 24 Female to Male Transexual individuals after treatment with Testosterone.

#### 1.1.4. Methylation profiling by genome tiling array (GSE80468):

The last dataset is Genome-wide DNA methylation analysis on PCOS using Illumina HumanMethylation 450K BeadChips in 30 PCOS patients and 30 healthy controls.

The remainder of this paper is structured as follows. Section 2 provides an explanation of the various methods used to examine the datasets. Section 3 is a list of these analyses' findings. In section 4, these results are thoroughly discussed. The article is concluded in Section 5.

## 2. Methods

In this section, a comprehensive description of the step-by-step procedures employed for the analysis of the four datasets is provided. For a more in-depth exploration, including code snippets and visualizations, please consult the accompanying appendix.

### 2.1. Microarray based gene expression analysis

Data preprocessing of the first microarray dataset involved the download of CEL files from the NCBI GEO website, with subsequent manual addition of sample information due to the absence of metadata. The affy package facilitated background correction and quantile normalization, employing the robust multi-array average (RMA) method. Quality control checks, both pre- and post-processing, were conducted using the arrayQualityMetrics package to ensure data integrity.

Differential expression analysis was carried out using the limma package, involving the creation of a design matrix without intercept, linear model fitting, and identification of top genes through the eBayes method. Adjusted p-values were calculated for the results of the differential expression analysis using the Benjamini-Hochberg procedure. For gene set analysis, Goana was employed to identify over-represented Pathways (KEGG) and Gene Ontology: Molecular Function terms among the top 300 genes, selected based on p-value and logFC. P-values were used considering the absence of differently expressed genes when considering adjusted p-values. Additionally, a more lenient logFC cutoff of 0.8 was chosen due to the lack of substantial logFC values in the dataset. Finally, the adjusted p-values for the results of gene set analysis were calculated using the Benjamini-Hochberg procedure.

## 2.2. RNA-sequencing based gene expression analysis

The analysis of the dataset initiated with the retrieval of raw data from the Sequence Read Archive (SRA) using sample IDs. Subsequently, preprocessing of the data commenced by downloading the raw sequencing reads in FASTQ format. Quality control was conducted using the FastQC toolkit to assess data quality and identify potential overrepresented adapter sequences. Following this, a trimming step utilizing TRIMOMATIC removed adapter sequences and low-quality regions from the reads. A secondary FastQC analysis verified the effectiveness of the trimming process in enhancing data quality and reducing noise. Once preprocessed, the GTF file for Homo sapiens were loaded, and the GTF file was converted to RNA sequences in FASTA format. This sequence information was employed for mapping the RNA-seq data using the pseudo-aligner KALLISTO. The resulting pseudo-count data was adjusted for transcript length and summarized to the gene level using genome annotation, as we are not interested in differential transcript analysis but differential gene analysis. A library size normalization (TMM) and filtering step were then applied.

To address the notable differences in library size between samples, voom with quality weights was used to analyze the preprocessed data. After this, differential gene expression analysis was performed by fitting a model with no intercept using limma. Differential expression analysis was carried out using the limma package, involving the creation of a design matrix without intercept, linear model fitting, and identification of top genes through the eBayes method. Following this, differential gene expression analysis was performed by fitting a model with no intercept using limma. Adjusted p-values for the differential expression analysis were calculated using the Benjamini-Hochberg procedure.

For gene set analysis, an identical methodology was applied to ensure consistency across datasets. Goana was employed to identify over-represented Pathways (KEGG) and Gene Ontology: Molecular Function terms among the top 300 genes, selected based on p-value and logFC, akin to the approach used in the previous dataset. Goana's consideration of gene length is important for sequencing data, where genes with higher lengths were more likely to have elevated counts, potentially leading to more differentially expressed results.

## 2.3. 2nd Microarray based gene expression analysis

A gene expression microarray is utilized to determine differential expression analysis between PCOS samples and Female to Male Transexual individuals after treatment with Testosterone samples. Dataset includes two batches which are Affymetrix

Human Genome U133A Array and Affymetrix Human Genome U133B Array, only first batch was used. The expression levels from the Affymetrix microarray data were normalized and summarized using the Robust Multi-array Average (RMA) technique. RMA is a preferred preprocessing technique for microarray data that includes summarizing probe intensities, quantile normalization, and background correction to provide expression values with the arrayQualityMetrics package, quality control metrics are calculated before and after preprocessing. Subsequently, differential expression is tested with the limma package. Finally, Gene Set Enrichment Analysis is performed by goana function, but with a 0.05 cut-off. More details on this subject can be found in the results sections.

## 2.4. Methylation profiling gene expression analysis

In the exploration of Polycystic Ovary Syndrome (PCOS), the Infinium HumanMethylation450k BeadChip platform was employed to analyze methylation profiles from 60 patients, including both PCOS and healthy tissues. Probes with insufficient statistical significance were filtered out using the pfilter function.

Data preprocessing involved reading raw intensity data from IDAT files using the readEPIC function from the wateR-melon package. Quality control, preprocessing, and normalization were then conducted with the minfi package. Dye color adjustment addressed technical biases, and Figure 1 illustrates density and color bias adjustments using methylumi objects.

For differential methylation analysis, the limma package was employed to identify positions with distinct methylation patterns between PCOS and control groups.. Finally, gene set analysis was performed on both differentially methylated positions and regions, following the same methodology as described earlier. This comprehensive approach provides insights into the epigenetic differences associated with PCOS.

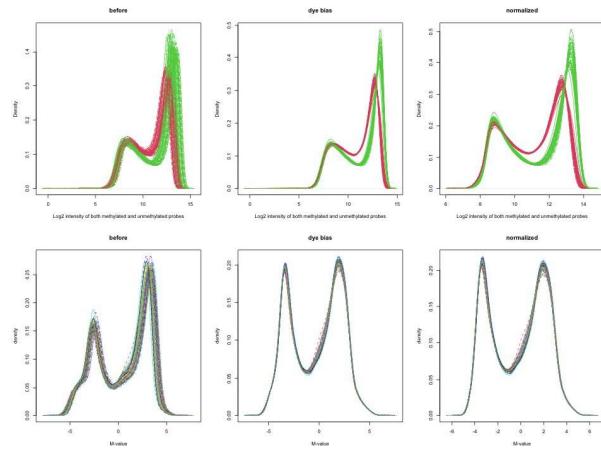


Figure 1: Density and color bias adjustment using methylumi objects.

## 2.5. Integrating the results

To integrate findings across the four datasets, a comprehensive approach was employed to identify similarities in differently expressed genes. To ensure a robust comparison between differently expressed genes, we employed a 'threshold' approach based on the dataset with the highest number of differently ex-

pressed genes. The number of differently expressed genes from this dataset served as the threshold, and an equivalent subset was selected from each dataset. These subsets were then overlapped and compared using Gene Symbol to identify shared elements and assess the degree of similarity across the datasets. This approach allows us to identify consistently differentially expressed genes enhancing the reliability and coherence of our integrated results.

### 3. Results

#### 3.1. Microarray Based Expression Analysis

The successful execution of background correction and quantile normalization using the RMA method is evident from the box-plots, displaying nearly identical means and quantiles across conditions (see Fig. 2). This positive transformation was further confirmed by the post-RMA arrayQualityMetrics report. However, visual inspection of the PCA plot did not reveal distinct clustering of control or PCOS samples, indicating potential challenges in identifying clear differences in the data.

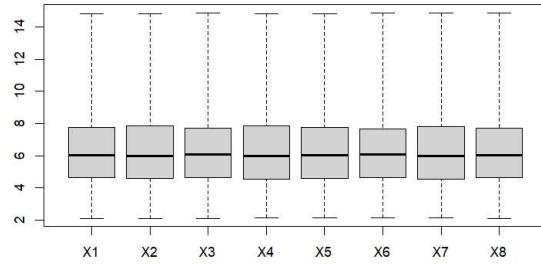


Figure 2: Boxplot showing the data distribution of intensity values of each sample after normalization for dataset 1.

The differential expression analysis using the limma package did not reveal any significantly differentially expressed genes, evident from consistently high adjusted p-values (0.9999858). This outcome implies a potential limitation in statistical power, primarily coming from the experimental design, wherein only data from four control and four PCOS-positive women were utilized. Additionally, the volcano plot highlighted a scarcity of substantial changes in logFC (See Fig. 3).

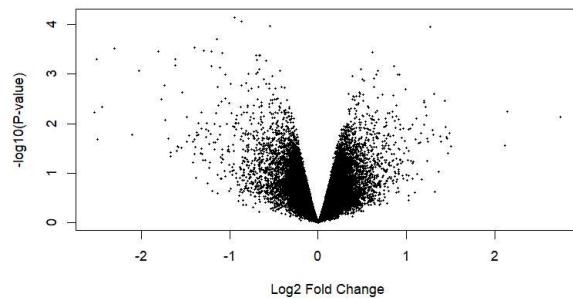


Figure 3: Volcano plot illustrating the results of differential gene expression analysis for dataset 1 using the limma package.

Gene Set Analysis was performed on the top 300 genes, selected based on p-values and logFC due to the lack of significant adjusted p-values. A more lenient logFC cutoff of 0.8

was chosen considering the absence of substantial logFC values in the dataset. However, due to these lax cut-offs, it is more likely that many of the selected genes are not differentially expressed in PCOS, a notion supported by the results of the over-represented pathway (KEGG) analysis. Only two significantly upregulated pathways were detected: Viral protein interaction with cytokine and cytokine receptor (adj.P.val = 0.001391039) and Chemokine signaling pathway (adj.P.val. = 0.029937282). In the gene set analysis using gene ontology: molecular function, three were significantly upregulated, and 14 were significantly downregulated.

#### 3.2. RNA-seq Based Expression Analysis

The RNA-seq preprocessing was conducted on the HPC of UGent, leveraging its computational power for resource-intensive tasks such as multiple FastQC analyses, trimming, and pseudo-alignment with KALLISTO. While the FastQC report reflected high read quality, a substantial variation in library sizes was observed. To address this discrepancy, voom with quality weights was implemented (See Fig. 4) to mitigate the impact of library size differences on downstream analyses.

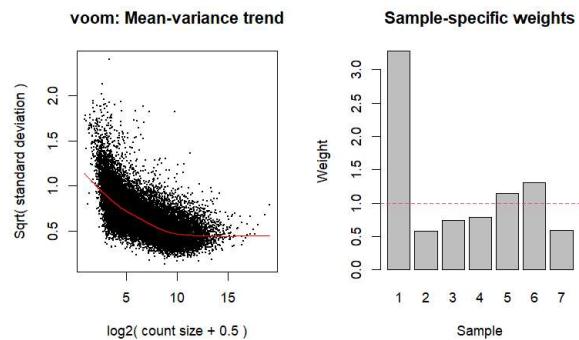


Figure 4: The Voom transformation's Mean-Variance Trend Plot unveils insights into the correlation between estimated mean expression and biological coefficient of variation, while the Sample-Specific Weights Plot visualizes how individual weights optimize precision by down-weighting high-variability observations in the pursuit of refined differential expression analysis.

The differential expression analysis was performed utilizing the limma package. Remarkably, only six genes exhibited significant differential expression, as evidenced by the adj.p.value in Table 1. Notably, two of these genes demonstrated an absolute log fold change exceeding 1.5, suggesting a substantial potential difference between cases and controls with the potential for biological impact in the context of PCOS.

Table 1: Results of Differential Expression Analysis on Dataset 2, 6 genes were found to be significantly differentially expressed.

logFC	adj.P.Value	Gene
1.077184	0.03324	SPON2
-1.503025	0.03324	SCN3B
-1.794698	0.03324	NaN
1.5413606	0.03324	AADAC
1.2211944	0.03324	APOA1
-3.21892115	0.03324	MUC6

Gene set analysis was executed on a curated selection of 300 genes, chosen for their substantial absolute log fold change (1.5) and the smallest p-values. When it comes to KEGG enrichment analysis, no pathways exhibited significant up or downregulation based on adjusted p-values. Noteworthy findings encompassed the Complement and coagulation cascades pathway and the Fluid shear stress and atherosclerosis pathway. When it comes to Gene Ontology: Molecular Function terms, once more, no significant outcomes emerged based on adjusted p-values. Notable top hits included terms related to extracellular matrix structural constituent and signaling receptor regulator activity.

### 3.3. 2nd Microarray based gene expression analysis

Boxplots confirm the successful implementation of RMA for background correction and quantile normalization. Consistency in means and quantiles across diverse conditions indicates the effectiveness of these techniques. Array quality metrics validate data robustness. (see Fig. 5)

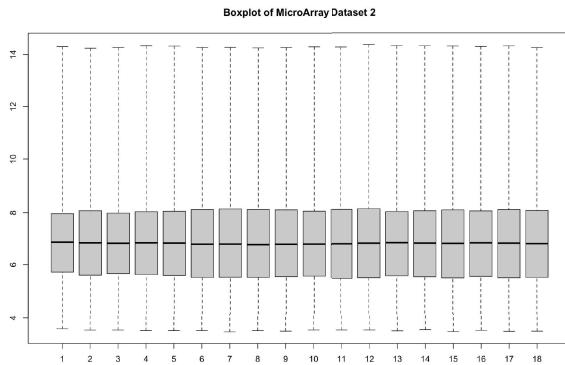


Figure 5: Boxplot showing the data distribution of intensity values of each sample after normalization for 2nd MicroArray Dataset.

PCA identified outliers in samples (Patients PCOS 149A and PCOS 152A). Despite detection, these outliers were retained due to limited sample availability, aiming to avoid potential information loss.(see Fig. 6)

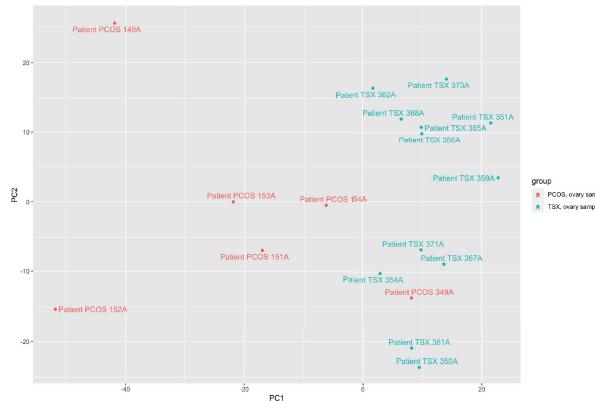


Figure 6: PCA applied to the samples.

Following the preprocessing steps, the limma package was utilized for differential expression analysis. As there were no cofounders, the intercept was set to zero. A predefined threshold was applied, with a cutoff for a p-value of 0.05 and a fold change of 1.5. Consequently, a total of 5338 genes were identified as significantly differentially expressed based solely on p-value criteria.

Table 2: Following a Differential Expression Analysis with pre-defined criteria for both fold change and p-value, only three genes were deemed statistically significant.

logFC	adj.P.Value	Gene
1.504750	0.00589	DDX17
-1.512543	0.0337	CYP11A1
-1.973567	0.0502	CYP17A1

We found in our research that for several genes, the p-values suggested statistical significance, but the corresponding fold changes fell short of statistical significance. This implies that even when expression levels could vary somewhat, these variations might only have a little impact.(see Table 2)

It is crucial to use caution when interpreting these results, taking into account both the statistical significance and the biological importance of the effects that were seen.

In the final step, Gene Set Analysis (GSA) was conducted using both KEGG and goana functions. In the goana analysis, 186 molecular function terms were identified as significant, with the most prominent terms including protein binding, organic cyclic and heterocyclic compound binding. Similar pathways observed in other omics analyses, such as cell adhesion molecule binding and cytoskeletal protein binding, were also significant.

The over-representation pathway analysis using KEGG revealed that the most significant terms were associated with Focal Adhesion and the P13K-Akt signaling pathway. Other pathways, including proteoglycans in cancer and regulation of actin cytoskeleton, were also identified and exhibited connections with findings from other omics analyses

### 3.4. Methylation Profiling Analysis

Methylation profiles of both PCOS (Polycystic Ovary Syndrome) and healthy tissues were analyzed using three packages: minfi, limma, and wateRmelon. The normalized data was visually represented in boxplots. (see Fig. 7)

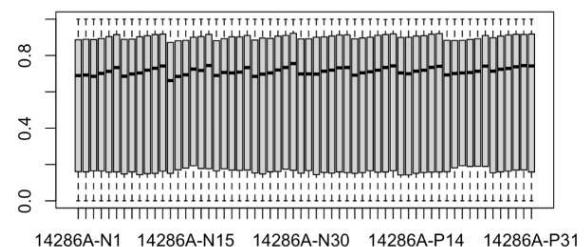


Figure 7: Boxplot showing the data distribution of intensity values of each sample after normalization.

To compare methylation patterns between PCOS and healthy tissues, average beta values were calculated for each group. A statistical test called a t-test was used to check for differences. The p-value is reported as 0.09934. This value is close to 0.05, suggesting some evidence against the null hypothesis. However, the result is not statistically significant at the 0.05 level.

Using the limma package, a differential expression analysis was performed. Since there were no confounding factors, the intercept was set to zero, and the Benjamini-Hochberg (BH) method was applied. There were initially 484,175 Probe IDs in the data, but in the annotation file,

IlluminaHumanMethylation450kanno.ilmn12.hg19, only 484,110 Probe IDs had matching gene names. This mismatch was resolved by assigning NA (Not Available) to the gene names of the unmatched probes. After applying a lax cutoff of 0.1, 146 genes were identified as significantly differentially expressed. Top three genes are shown in Table 3

Table 3: Top Hits for DE Between PCOS and Healthy Tissue in the Methylation Analysis.

logFC	adj.P.Value	Gene
-0.3007335	0.00102	RBL2
-0.7155935	0.0025	C3orf35
-0.5485005	0.0025	FNDC3B

Due to an insufficient number of significant genes, gene set analysis (GSA) was performed on the entire gene set instead of just the significant ones, resulting in 100 significant Molecular Function terms. In the KEGG enrichment analysis for Infinium data, the most significant terms included "neuroactive ligand-receptor interaction." When compared with microarray data, the matching pathways were found to be MAPK and PI3K signaling pathways, as well as the regulation of the actin cytoskeleton.

### 3.5. Integrating the results

The datasets, designated as Dataset 1, Dataset 2, Dataset 3, and Dataset 4, underwent individual analyses using the limma package. Subsequently, the results were combined to yield a comprehensive overview. The integration process was performed on a subset comprising the top 5112 genes from each dataset, presenting both upregulated and downregulated genes for detailed comparison. This method facilitated a robust comparative analysis, indicating consistently differentially expressed genes and contributing to a more dependable and coherent understanding of the genetic landscape associated with Polycystic Ovary Syndrome (PCOS).

An exploration of upregulated genes across datasets unveiled eight common genes: TGOLN2, CRYZ, SQSTM1, ALG5, BLVRA, ST3GAL1, CAMLG, and SNCA. Examination of the adjusted p-value and logFC for each gene emphasized the importance of evaluating the consistency across datasets. Notably, none of the logFC values exceeded 1 in any dataset, indicating a lack of substantial changes between control and PCOS patients. Furthermore, the logFC exhibited variability, ranging from 0.1 to 0.6 across different datasets for each gene.

For downregulated genes, two common genes, DEPDC1 and NTRK3, were identified. Similar to the analysis of upregulated genes, the adjusted p-value and logFC underscored the need for assessing consistency across datasets. In this context,

the logFC values again displayed variability, ranging from -0.3 to -1.1 for DEPDC1 and from -0.1 to -1.1 for NTRK3.

The collective observation of high p-values, low logFC, and variable logFC across genes should serve as a cautionary note, urging a critical interpretation of the results.

## 4. Discussion

### 4.1. Dissecting Gene Associations in PCOS

In our integrated results, three genes—SQSTM1, ALG5, and SNCA—emerged as potentially linked to PCOS as they were upregulated in all datasets and had a relationship with PCOS in published research papers. When compared with 3 independent research, SQSTM1 exhibited differential expression in PCOS, with studies reporting a significant decrease in its expression, contrary to our findings [5,6,7]. The overexpression of ALG5, a glycosyltransferase, was found in our research and is associated with the NLRP3 inflammasome-dependent pathway in PCOS, impacting glycan synthesis, sex hormone synthesis, autophagy, and apoptosis [8]. The upregulation of SNCA, implicated in neurodegenerative diseases, showed potential relevance to PCOS in our results, contrasting with a study reporting its significant downregulation in PCOS patients [9]. This dysregulation in AD-related protein expression within the PCOS context is likely exacerbated by factors like obesity and inflammation related to insulin resistance. Protective heat shock protein (HSP) mechanisms may play a role in modulating these intricate molecular interactions. These comparisons highlight the intricate molecular landscape of PCOS, emphasizing the need for further research to unravel underlying mechanisms.

### 4.2. Challenges in Statistical Interpretation and the Impact of Experimental Design

In our research, we encountered instances where the p-values associated with certain genes indicated statistical significance, implying differences in gene expression between experimental conditions. However, a closer examination revealed that the corresponding fold changes were relatively small, suggesting that while statistically significant, the observed variations had a modest impact. This underscores the importance of a nuanced interpretation, as statistical significance alone does not necessarily translate into significant biological effects.

In gene expression studies, where assessing changes in expression levels is common, it is essential to go beyond statistical metrics and consider the practical significance of observed alterations to accurately gauge their biological relevance. The fact that expression levels vary statistically doesn't automatically imply that these variations have a substantial impact on the biological system under investigation. Small changes in gene expression may have limited biological consequences or may not lead to noticeable effects at the organismal level. This is particularly relevant in cases where the observed differences, although statistically significant, fall short of representing a "big" or substantial change in terms of biological function.

Three of our datasets showed little to no significant results based on adjusted p-values. This can be attributed, in part, to the experimental design, where smaller sample sizes result in reduced statistical power—making it more challenging to detect true effects even when they exist. However, it's also plausible that the Benjamini-Hochberg procedure, employed to control the false discovery rate, may be too stringent. This procedure, while effective in minimizing false positives or Type I errors,

involves a delicate balance between avoiding false positives and potentially increasing false negatives or Type II errors.

The conservative nature of the Benjamini-Hochberg procedure means that it might cut out too many true positives for the sake of removing false positives. In other words, it could be overly cautious in declaring a result as statistically significant, potentially leading to a lack of significant findings even in cases where they could be valid.

When designing the linear models for fitting the data, intentional efforts were made to exclude confounding variables. All the models were intentionally constructed without an intercept, emphasizing the exploration of the relationship between cases and controls. This choice was justified by the matched experimental design used in the source studies of the datasets. In matched studies, researchers aim to mitigate the impact of specific confounding variables by carefully pairing participants in the treatment group with those in the control group, ensuring similarity in terms of potential confounders. The selection criteria for cases and controls were meticulous, encompassing factors such as equal weight, age, length, etc., thereby creating a controlled environment where the only differing factor impacting gene expression is the presence or absence of PCOS.

However, it's crucial to acknowledge the inherent imperfections in any experimental design. Despite the efforts to minimize confounders in matched studies, they persist, albeit in reduced proportions. Access to additional information on these confounding factors could potentially enhance the interpretability of our results, offering a more nuanced understanding that aligns closely with findings in existing research on PCOS.

Revisiting the integration of results, we observed consistent patterns of upregulation or downregulation for a small number of genes across all datasets. However, a noteworthy observation was the lack of a clear trend in the fold changes for these genes. One would expect similar log-fold changes across all datasets when genes are consistently regulated, however, this was not the case. The comparison of fold changes across different datasets introduces several challenges.

Primarily, variations in experimental setups, encompassing differences in laboratory protocols, sample handling, and the technology platforms used for data generation, can introduce substantial variability. Additionally, the utilization of different technologies and targets, such as microarrays, RNA-seq, and methylation data, further contributes to divergent results. These technical disparities pose challenges to achieving consistent fold change measurements, even when the underlying biological regulation remains similar.

Secondarily, the influence of biological heterogeneity among study populations or conditions cannot be overlooked. Datasets may encompass diverse people with distinct genetic backgrounds, environmental exposures, or other factors influencing gene expression. Consequently, the inherent diversity among these datasets may lead to disparities in observed fold changes, complicating direct comparisons.

Lastly, the intrinsic statistical variability and noise in biological data contribute to fluctuations in fold change measurements. Differences in statistical power, sample sizes, and data preprocessing methods across datasets, in conjunction with the considerations mentioned earlier, further complicate the challenge of establishing a uniform trend in fold changes.

## 5. Summary & Conclusion

An summary of the -omics landscapes in PCOS is given in the report's conclusion. We demonstrate the considerable overlap that exists between the findings of various data sources and provide evidence for the rationale—both statistical and biological—of utilizing information from disparate data formats. This paper, however, offers an illustration of the fundamentals of utilizing several omics data sources to generate novel biological findings. We think that integrating data from many sources can produce better outcomes with more data and better experimental designs.

## 6. Contributions

The project's initial phase involved selecting and analyzing datasets on the topic of interest. Dilara identified and chose four datasets related to Polycystic Ovary Syndrome (PCOS). Subsequently, Maxim took charge of the analysis of Dataset 1 (microarray-based gene expression) and Dataset 2 (RNA-sequencing-based gene expression), and integrating the results from all four datasets. Dilara conducted the analysis for Dataset 3 (microarray-based gene expression) and Dataset 4 (methylation profiling gene expression).

As the project transitioned to report writing, Dilara took the lead in writing the Introduction and Conclusion sections, along with handling Methods sections 2.3 and 2.4, as well as Results sections 3.3 and 3.4. Meanwhile, Maxim contributed to Methods sections 2.1, 2.2, and 2.5, to Results sections 3.1, 3.2, and 3.5, and to the Discussion part of the report.

## 7. References

- [1] Selma Feldman Witchel, Helena J Teede, and Alexia S Peña, "Curtailing pcos," .
- [2] Anuja Dokras, "Cardiovascular disease risk in women with pcos," *Steroids*, vol. 78, pp. 773–776, 2013.
- [3] Daniel A. Dumescic, Julia D. Phan, Karen L. Leung, Tristan R. Grogan, Xiangmian Ding, Xinmin Li, Luis R. Hoyos, David H. Abbott, and Gregorio D. Chazenbalk, "Adipose insulin resistance in normal-weight women with polycystic ovary syndrome," *The Journal of Clinical Endocrinology and Metabolism*, vol. 104, pp. 2171, 6 2019.
- [4] Shuxia Li, Dongyi Zhu, Hongmei Duan, Anran Ren, Dorte Glintborg, Marianne Andersen, Vibe Skov, Mads Thomassen, Torben Kruse, and Qihua Tan, "Differential dna methylation patterns of polycystic ovarian syndrome in whole blood of chinese women," *Oncotarget*, vol. 8, pp. 20656–20666, 2017.
- [5] Paramita Saha, Sudhir Kumar, Kasturi Datta, and Rakesh K. Tyagi, "Upsurge in autophagy, associated with mifepristone-treated polycystic ovarian condition, is reversed upon thymoquinone treatment," *The Journal of Steroid Biochemistry and Molecular Biology*, vol. 208, pp. 105823, 4 2021.
- [6] Xiaoxue Li, Jia Qi, Qinling Zhu, Yaqiong He, Yuan Wang, Yao Lu, Hasiximuke Wu, and Yun Sun, "The role of androgen in autophagy of granulosa cells from pcos," *Gynecological Endocrinology*, vol. 35, pp. 669–672, 8 2019.
- [7] Da Li, Yue You, Fang Fang Bi, Tie Ning Zhang, Jiao Jiao, Tian Ren Wang, Yi Ming Zhou, Zi Qi Shen, Xiu Xia Wang,

and Qing Yang, “Autophagy is activated in the ovarian tissue of polycystic ovary syndrome,” *Reproduction*, vol. 155, pp. 85–92, 1 2018.

- [8] Bo Wang, Minfeng Shi, Chuanjin Yu, Hong Pan, Haiqing Shen, Yatao Du, Yi Zhang, Bin Liu, Di Xi, Jianzhong Sheng, Hefeng Huang, and Guolian Ding, “Nlrp3 inflammasome-dependent pathway is involved in the pathogenesis of polycystic ovary syndrome,” *Reproductive Sciences*, vol. 1, pp. 1–11, 10 2023.
- [9] Alexandra E Butler, Abu Saleh, Md Moin, Thozhukat Sathyapalan, and Stephen Lawrence Atkin, “Sat366 expression of amyloid-related proteins associated with dementia in polycystic ovary syndrome compared to a control population,” *Journal of the Endocrine Society*, vol. 7, 10 2023.