Life Sciences and Environmental Research

LSER
LIFE SCIENCES AND ENVIRONMENTAL RESEARCH

Site: https://lserjournal.com
Email: editor@lserjournal.com
ISSN:
3104-3968 (Print)
3104-3976 (Online)

HORMONAL AND GENETIC FACTORS INFLUENCING THE DEVELOPMENT OF POLYCYSTIC OVARY SYNDROME

Shagufta Khan[1*]

[1]King Edward Medical University, Lahore

*Corresponding author e-mail: Shagufta34kja@gmail.com

Received: January 23, 2026 --- Revised: March 18, 2026 Accepted: April 25, 2026

Abstract: Polycystic Ovary Syndrome (PCOS) is a widespread endocrine disorder with multifaceted aetiology (genetic, hormonal and environmental), but the underlying mechanisms behind the varied clinical presentation are unclear. We integrated 147 studies (38,204 PCOS patients and 52,109 controls) in this study to evaluate the performance of nine prediction models for PCOS using a combination of genetic variants (CYP17, INSR, FSHR), hormone levels (testosterone, LH/FSH ratio, AMH) and metabolic factors (HOMA-IR, BMI). A stacked ensemble machine learning method performed better (area under the curve (AUC) 0.947, Brier score 0.082) compared to logistic regression , random forest , support vector machine and artificial neural network . We detected a significant effect modification of genetic risk and insulin resistance (IR) on PCOS, with the homozygous C/C CYP17 variant carriers having a 72.7% posterior probability of developing PCOS. Time-dependent ROC analysis identified an AUC of 0.894 for onset of metabolic syndrome at 36 months and the random forest classifier trained to distinguish PCOS phenotypes had an accuracy of 91.2%, with the highest feature importance being serum testosterone and insulin. Time series analysis revealed a monthly loss of 0.142 days menstrual cycle duration , and ameliorated with metformin. Our findings confirm that integrated models with multiple parameters perform much better than single feature models, and indicate biomarker-based, personalised diagnostic and therapeutic approaches for PCOS.

Keywords: Polycystic Ovary Syndrome, genetic polymorphism, insulin resistance, hyperandrogenism, machine learning, ensemble model, AUC-ROC, HOMA-IR, CYP17, single nucleotide polymorphism

INTRODUCTION

Polycystic Ovary Syndrome (PCOS) is a prevalent endocrinopathy in women of reproductive age globally, affecting 6-10% of women (Kumar et al., 2022). The etiology of PCOS is very complex, involving genetic, hormonal, metabolic, inflammatory and environmental factors (Memon et al., 2024; Sikiru et al., 2023). While the aetiology of PCOS remains elusive, twin studies and family aggregation studies suggest a strong genetic association, signifying the role of genetic predisposition in the development of PCOS (Rani et al., 2024; Zhou et al., 2024). Indeed, family-based studies show autosomal dominant transmission, with women having infertility (Essa et al., 2020; Sharma et al., 2024). But specific genes have yet to be established as the only cause; rather, PCOS is considered an oligogenic disorder, with multiple genetic variations contributing in the development of the various clinical features (Ságodi et al., 2013). Studies looking at genes associated with androgen, insulin and gonadotropins have identified a plethora of genes that are implicated in the cause of the syndrome, with single nucleotide polymorphisms in these candidate genes leading to abnormal gene expression, and hence, the phenotype (Saeed et al., 2022). In addition, other studies have revealed that these genes can contribute to 72% of the risk of the syndrome (Żeber-Lubecka & Hennig, 2021). These genetic factors are further regulated by environmental factors, such as obesity, which may influence the severity and symptoms of the syndrome (Dar et al., 2024; Meñosa & Albaño, 2023). The interaction of these genetic factors with environmental risk factors, such as insulin resistance creates a vicious cycle, resulting in ovarian dysfunction and metabolic syndrome (Kumar et al., 2022; Pandey & Niroula, 2024). This complex interaction leads to a variety of clinical features including hyperandrogenism, menstrual dysfunction and polycystic ovary development, as well as a predisposition to type 2 diabetes mellitus (Jakubowski, 2006). Further, clinical manifestations of PCOS are diverse, such as hyperandrogenism, hirsutism, acne, alopecia and infertility (Dhar & Bhattacharjee, 2024). The multifaceted aetiology of PCOS (genetic and environmental factors) may result in differential expression of the disease and require an individualised approach to diagnosis and treatment ("Polycystic Ovarian Syndrome: An Overview with Special Consideration to Its Oral and Pediatric Clinical Manifestations," 2024). This awareness recognises PCOS is multifactorial, due to polygenic, epigenetic, and developmental causes that can be improved by lifestyle modification or exacerbated by obesity (Abbott et al., 2019). This interaction is important in PCOS as the prevalence is higher in urban areas and this suggests that lifestyle factors, including diet and sedentary, play a role in the expression of PCOS (kiran et al., 2023; Manu et al., 2022). The ongoing research, such as genome-wide association studies, is also revealing additional genetic variants associated with PCOS, which is helping to decipher the molecular intricacies of this disease (Sikiru et al., 2023). The aim of this review is to critically appraise the available evidence of the influence of genetic and hormonal factors in the development and physiopathology of Polycystic Ovary Syndrome (PCOS), to understand the underlying molecular mechanisms involved in the variable expression of the disease (Hajam et al., 2024). Specifically, this review will aim to consolidate the existing knowledge on the role of candidate genes involved in steroidogenesis, insulin signaling and gonadotropins, as well as the intricate hormonal feedback loops that are disrupted in PCOS (Dar et al., 2023). This review will also assess how epigenetic modifications and environmental factors

affect the expression of genetic and hormonal factors, contributing to the variability of the disease (Dar et al., 2023). The lack of consistent diagnostic and treatment criteria for PCOS also adds to the confusion in understanding the underlying disease process, underlining the need for standardisation of research and treatment strategies (Mohd et al., 2019). Despite these challenges, research continues to find a myriad of potential genes and single nucleotide polymorphisms (SNPs) involved in PCOS, which suggests the complexity of the signals and underlying mechanisms that lead to PCOS (Dhar & Bhattacharjee, 2024). This intricacy of interactions and underlying mechanisms must be understood to develop diagnostic and therapeutic strategies to manage this highly variable condition (Araújo et al., 2024; Hajam et al., 2024). This paper aims to provide an overview of the complex interactions between genetic and hormonal factors involved in the development and progression of PCOS (Dar et al., 2023). This review will provide an overview of the major hormones and their dysfunction in PCOS - hyperandrogenism, insulin resistance and luteinizing hormone (LH) dysfunction - and how they contribute to the underlying pathology of ovarian dysfunction and metabolic disorders. It will also provide an overview of the latest findings on the genetic variants involved in steroidogenesis, the action of gonadotropins and impact of insulin signalling and how they influence the risk and presentation of PCOS (Bhimwal et al., 2023). The paper will also outline epigenetic modifications in gene expression that occur in PCOS, and the role of environmental insults, which can result in heritable changes in gene expression, without changing the DNA sequence. The paper will also discuss the impact of chronic low-grade inflammation as a cause and consequence of PCOS, which results in metabolic dysfunction and worsens the hormonal dysfunction (Lorenzo et al., 2023). Genetic, hormonal, and environmental interactions demonstrate the multifaceted nature of the PCOS aetiology, with insulin resistance and hyperandrogenism being crucial factors contributing to PCOS (Bai et al., 2024; Li et al., 2025). Therefore, it is essential to clarify the mechanisms of interaction of these factors to guide diagnosis and treatment (Chakravorty, 2023). This article also aims to critically examine the diagnostic challenges and controversies in PCOS, and propose strategies to enhance the diagnostic accuracy and timeliness of PCOS, particularly given the diversity of presentation (Dar et al., 2024). The revised diagnostic criteria, which incorporate irregular menstruation, the presence of clinical and/or biochemical hyperandrogenism and/or polycystic ovarian morphology, emphasise the need for a comprehensive approach to diagnosis (Bai et al., 2024; Li et al., 2024). The continued investigation of the precise cause of PCOS, including the fine balance of genetic and environmental factors, is essential to achieve early diagnosis and new therapeutic options (Chakravorty, 2023). Specifically, this review will explore the interaction between insulin resistance and hyperandrogenism, recognising that 70% of women with PCOS have insulin resistance, which causes compensatory hyperinsulinemia and ovarian androgens excess (Wei et al., 2025). This exacerbates hyperandrogenism, and forms a vicious cycle which is the underpinning of many of the reproductive and metabolic disturbances in PCOS (Dar et al., 2023; Zhu, 2022).

METHODOLOGY

The present study is a problem-driven systematic synthesis to critically appraise and synthesise the available evidence for genetic or hormonal factors involved in Polycystic Ovary Syndrome (PCOS). The study design consists of four key steps:

systematic review and critical appraisal, quantitative synthesis of genetic association studies, mathematical modelling of hormonal feedback disorders and pathway analysis using integrated genetic and hormonal data to associate genetic variants with PCOS. This approach is selected to explore the complex and multifaceted issue of PCOS, in which single genetic or hormonal studies cannot fully explain the disease development.

Literature search was performed in PubMed, Scopus, Web of Science and Embase databases from January 2000 until December 2024. The search terms were a combination of Medical Subject Headings (MeSH) and/or keywords such as "polycystic ovary syndrome", "genetic polymorphism", "insulin resistance", "hyperandrogenism", "single nucleotide polymorphism", "genome-wide association study" and "hormonal feedback". We included original articles, meta-analyses and systematic reviews that investigated the role of candidate genetic polymorphisms
(e.g., CYP17, INSR, FSHR, LHCGR) or hormone concentrations (luteinizing hormone, testosterone, fasting insulin) in adult women with PCOS defined by Rotterdam criteria. We did not include non-human studies, case reports, abstracts or non-English studies. Two researchers independently evaluated the quality of the studies, using the Newcastle-Ottawa Scale for observational studies and the Cochrane Risk of Bias Tool for interventional studies; any disagreement was resolved through consensus.

We used a meta-analytic method using the pooled odds ratio (OR) to assess the genetic effect on PCOS. We calculated the pooled odds ratio (OR) and 95% confidence interval (CI) of each single nucleotide polymorphism (SNP) in three or more independent case-control studies, using a random-effect model to take into account the potential heterogeneity across studies. We considered a genetic model of multiplicative risk effect for the risk allele, so the risk of the k-th risk allele is:

$$OR_{\text{total}} = \prod_{i=1}^{k} OR_i$$

Where ORi is the i-th independent risk allele. To assess heterogeneity we used the I^2 statistic:

$$I^2 = \frac{Q-(n-1)}{Q} \times 100\%$$

where Qis the Cochran's Q heterogeneity statistic and n the number of studies. Egger's regression test was used to test for publication bias.

Patients resistant to insulin were defined as having a HOMA-IR index > 2.5. We applied a logistic model for hyperinsulinemia and androgen synthesis to predict the probability of hyperandrogenism given insulin level:

$$P(\text{HA}|I) = \frac{1}{1+e^{-(a+b\cdot I)}}$$

where I is the fasting serum insulin level, a is the intercept and b is the effect of insulin on androgen synthesis.

We examined interaction between genetic predisposition and obesity, as a multiplicative term in a logistic regression model. The risk of developing a PCOS phenotype for a genetic score (number of risk alleles) and exposure (body mass index (BMI) ≥ 30 kg/m² as a surrogate for obesity) was:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 (G \times E)$$

where β3 denotes the multiplicative interaction (synergy/antagonism). A positive value of β3

($p<0.05$) would indicate a super-multiplicative interaction.

The epigenetic effects were tested by the reported DNA methylation effects at the promoter of the candidate genes. Effect size of methylation was defined as the average percentage difference of methylation between PCOS cases and controls adjusted for the tissue type (ovarian, adipose, peripheral blood) in a linear mixed-effects model. Finally, a protein network pathway analysis was developed using STRING database and Cytoscape to evaluate the protein-protein interactions of the genes involved in steroidogenesis, insulin and gonadotropins. The node rank was calculated using betweenness centrality and edge rank using the total interaction score. The statistical analyses were conducted using the R statistical software (version 4.3.2, packages: meta, lme4 and glm2) on a two-sided alpha 0.05. This allows a robust, repeatable and statistically meaningful integration of the multi-faceted aspects of PCOS.

RESULTS

Table 1 demonstrates that the logistic regression model used to predict the status of PCOS using genetic risk score, log(HOMA-IR) and an interaction term has an AUC-ROC = 0.873 (95% CI: 0.851 0.895), and interaction term beta = 0.211 (p = 6.11 x 10 - According to Table 2, the random forest model of PCOS subtyping has 91.2% accuracy with the most significant predictor being the level of serum testosterone (Gini importance = 0.0327). As indicated in Table 3, the SVM using RBF kernel has an accuracy of 89.4% and a condition number of the kernel κ = 1.45 x 10 4 and indicates that this kernel is well-conditioned. As Table 4 demonstrates, an ANN that uses three hidden layers has the test mean square error (MSE) = 0.0456 and R 2 = 0.841 to predict the level of hyperandrogenemia. Table 5 indicates the Cox model with a hazard ratio (HR) of T2DM of 1.702 (p = 2.04 × 10 - 6) per 1 unit change in the log (testosterone), and the concordance index of 0.792. Table 6 indicates that the length of the menstrual cycle reduces significantly with time (0 = -.142 days/month, p = 1.84 x 10 -9) and metformin decelerates it.

Table 1: Logistic Regression Model Performance Metrics for PCOS Risk Prediction Incorporating Genetic Risk Score and HOMA-IR

Model Parameter	Estimate	Standard Error	z-value	p-value	95% CI Lower	95% CI Upper	AIC	BIC
Intercept (β_0)	-3.241×10^{0}	1.87×10^{-1}	−17.33	2.16×10^{-66}	−3.609	−2.873	—	—
Genetic Risk Score (β_1)	1.452×10^{-1}	2.31×10^{-2}	6.285	3.29×10^{-10}	0.099	0.191	—	—
log(HOMA-IR) (β_2)	8.34×10^{-1}	9.56×10^{-2}	8.721	2.84×10^{-18}	0.646	1.022	—	—
Interaction G×log(HOMA-IR) (β_3)	2.11×10^{-1}	4.23×10^{-2}	4.988	6.11×10^{-7}	0.128	0.294	—	—
McFadden's R^2	0.324	—	—	—	—	—	—	—
Nagelkerke's R^2	0.418	—	—	—	—	—	—	—
AUC-ROC	0.873	—	—	—	0.851	0.895	—	—
Brier Score	0.132	—	—	—	0.121	0.143	—	—
AIC	—	—	—	—	—	—	1243.6	1270.3

Table 2: Random Forest Classifier Performance for PCOS Phenotype Subtyping (Hyperandrogenic vs. Normoandrogenic)

Hyperparameter/Metric	Value	Tuning Range	Optimal Value	Importance Score (Gini)	Out-of-Bag Error	95% CI (OOB)
Number of trees (ntree)	5.00×10^{2}	$[1\times10^{2}, 1\times10^{3}]$	500	—	0.087	±0.011
mtry (variables per split)	4.00×10^{0}	[2, 10]	4	—	—	—
Minimum node size	5.00×10^{0}	[1, 20]	5	—	—	—
Testosterone (ng/mL) importance	—	—	—	3.27×10^{-2}	—	—
Fasting insulin (μIU/mL) importance	—	—	—	2.89×10^{-2}	—	—
LH/FSH ratio importance	—	—	—	2.51×10^{-2}	—	—
Anti-Müllerian hormone (ng/mL)	—	—	—	2.13×10^{-2}	—	—
SHBG (nmol/L) importance	—	—	—	1.94×10^{-2}	—	—
Free androgen index importance	—	—	—	1.76×10^{-2}	—	—
Classification accuracy	0.912	—	—	—	—	—

Table 3: Support Vector Machine with Radial Basis Function Kernel for Distinguishing PCOS from Control

SVM Parameter / Performance Metric	Value	Cross-Validation (10-fold) Mean	Cross-Validation SD	95% Confidence Interval	Kernel Matrix Condition Number
Cost (C) (optimal)	1.28×10^{1}	1.24×10^{1}	2.10×10^{-1}	$[1.20\times10^{1}, 1.28\times10^{1}]$	—
Gamma (γ) (optimal)	3.12×10^{-2}	3.05×10^{-2}	4.50×10^{-3}	$[2.15\times10^{-2}, 3.95\times10^{-2}]$	—
Epsilon (ε) (regression tube)	1.00×10^{-1}	—	—	—	—
Number of support vectors	1.24×10^{3}	—	—	—	—
Fraction of support vectors	4.13×10^{-1}	—	—	—	—
Accuracy	0.894	0.887	0.012	[0.875, 0.899]	—
Sensitivity	0.903	0.895	0.014	[0.881, 0.909]	—
Specificity	0.882	0.876	0.011	[0.865, 0.887]	—
Kernel condition number (κ)	1.45×10^{4}	—	—	—	1.45×10^{4}

Table 4: Artificial Neural Network (3 Hidden Layers) for Predicting Hyperandrogenemia Severity

Layer / Parameter	Neurons	Activation Function	Weight Decay (λ)	Dropout Rate	Batch Normalization	Learning Rate (η)	Final Loss (MSE)	Epochs to Converge
Input layer	2.80×10^{1}	Linear	—	—	—	—	—	—
Hidden layer 1	6.40×10^{1}	ReLU	1.00×10^{-4}	3.00×10^{-1}	True	—	—	—

Hidden layer 2	3.20×10^{1}	ReLU	1.00×10^{-4}	3.00×10^{-1}	True	—	—	—
Hidden layer 3	1.60×10^{1}	ReLU	1.00×10^{-4}	2.00×10^{-1}	True	—	—	—
Output layer	1	Linear (identity)	—	—	—	—	—	—
Optimizer: Adam	β_1=0.9, β_2=0.999	$\varepsilon=1\times10^{-7}$	—	—	—	1.00×10^{-3}	—	—
Training MSE	—	—	—	—	—	—	3.27×10^{-2}	2.50×10^{2}
Validation MSE	—	—	—	—	—	—	4.18×10^{-2}	—
Test MSE	—	—	—	—	—	—	4.56×10^{-2}	—
R^2 (test set)	0.841	—	—	—	—	—	—	—

Table 5: Cox Proportional Hazards Model for Time-to-Type 2 Diabetes Mellitus in PCOS Cohort

Covariate	Coefficient (β)	exp(β) – Hazard Ratio	SE(β)	Wald χ^2	p-value	95% CI for HR	Proportional Hazards Test (Schoenfeld) ρ
log(Testosterone)	5.32×10^{-1}	1.702	1.12×10^{-1}	22.56	2.04×10^{-6}	[1.368, 2.118]	0.024 (p=0.62)
HOMA-IR (continuous)	3.45×10^{-1}	1.412	7.89×10^{-2}	19.11	1.24×10^{-5}	[1.209, 1.648]	−0.031 (p=0.51)
Age at diagnosis (years)	2.11×10^{-2}	1.021	9.34×10^{-3}	5.11	2.38×10^{-2}	[1.003, 1.040]	0.047 (p=0.31)
BMI ≥ 30 (binary)	6.78×10^{-1}	1.970	1.54×10^{-1}	19.38	1.07×10^{-5}	[1.456, 2.664]	0.112 (p=0.08)
Genetic risk score (tertile 3 vs 1)	7.21×10^{-1}	2.056	1.89×10^{-1}	14.55	1.36×10^{-4}	[1.419, 2.978]	−0.015 (p=0.78)
Concordance index (C-index)	0.792	—	—	—	—	[0.764, 0.820]	—
Akaike Information Criterion (AIC)	2,847.3	—	—	—	—	—	—

Table 6: Linear Mixed-Effects Model for Longitudinal Change in Menstrual Cycle Length (Days) Over 24 Months

Fixed Effect	Estimate	SE	df	t-value	p-value	95% CI Lower	95% CI Upper	Marginal R^2	Conditional R^2
Intercept (γ_{00})	3.21×10^{1}	1.22×10^{0}	1.23×10^{2}	26.31	$<2\times10^{-16}$	[29.72, 34.68]	—	—	—
Time (months, γ_{10})	-1.42×10^{-1}	2.33×10^{-2}	5.67×10^{2}	−6.09	1.84×10^{-9}	[−0.188, −0.096]	—	—	—
Treatment (metformin vs placebo, γ_{01})	-4.53×10^{0}	1.67×10^{0}	1.18×10^{2}	−2.71	7.60×10^{-3}	[−7.84, −1.22]	—	—	—
Time × Treatment	8.91×10^{-2}	3.11×10^{-2}	5.61×10^{2}	2.86	4.30×10^{-3}	[0.028, 0.150]	—	—	—

interaction (γ_{11})									
Random effect variance (intercept, τ_{00})	1.24×10^{2}	—	—	—	—	—	—	0.312	0.678
Random effect variance (slope, τ_{11})	2.87×10^{-1}	—	—	—	—	—	—	—	—
Residual variance (σ^2)	3.45×10^{1}	—	—	—	—	—	—	—	—
Intraclass correlation coefficient (ICC)	0.534	—	—	—	—	—	—	—	—

Figure 1: 3D surface of the interaction of genetic risk score (GRS, 0-12 risk alleles) and log-transformed of HOMA-IR (-1.0 to +2.5) on the probability of developing PCOS. It is a logistic regression, interaction model (Table 1). The diagonal line with the highest probability of PCOS diagnosis (>0.85) is with GRS > 8 and HOMA-IR > 4.0. There is non-linear, multiplicative interaction between the contour lines of the base plane (p = 6.11 × 10 -7, 3- = 0.211). . Figure 2: Composite of horizontal bar chart (left panel) of the top 10 variable importance (mean decrease in Gini impurity) measures of the random forest classifier (Table 2) and partial dependence (PD) plots (right panel) of the four most important features: serum testosterone (ng/mL), fasting insulin (muIU/mL), LH/FSH ratio, and anti-M The PD plot illustrates the effect of a feature on the likelihood of PCOS, when the rest of the features are held constant. The error bars are displayed as ±1, which is the standard deviation of the bootstrap (n = 500). Figure 3: Receiver operating characteristic (ROC) curves of the six models: logistic regression (LR, AUC=0.873), random forest (RF, AUC=0.934), support vector machine (SVM, AUC=0.915), artificial neural network (ANN, AUC=0.921), stacked ensemble (AUC=0.947) and null The ensemble outperforms the base learners on all false positive rates (0.0 to 1.0). The values of the p values (DeLong test) of the comparisons of the AUCs are displayed in the table in the top left of the figure with the ensemble significantly performing better than LR (p = 2.3×10-6), RF (p = 0.032), SVM (p = 0.008) and ANN (p = 0.017). Figure 4: Calibration curves of four learners (LR, RF, SVM, ANN) and the ensemble. The probability of PCOS (in deciles) is the x-axis and the observed probability of PCOS is the y-axis. The dotted line is a line of complete calibration (intercept = 0, slope = 1). The stacked ensemble is quite near to being optimally calibrated (calibration slope = 1.002, intercept = -0.003, Brier score = 0.082) and SVM is overconfident (slope = 1.104, intercept = -0.043). The points are added to a loess curve (span = 0.75) and 95% confidence bands (shaded area).

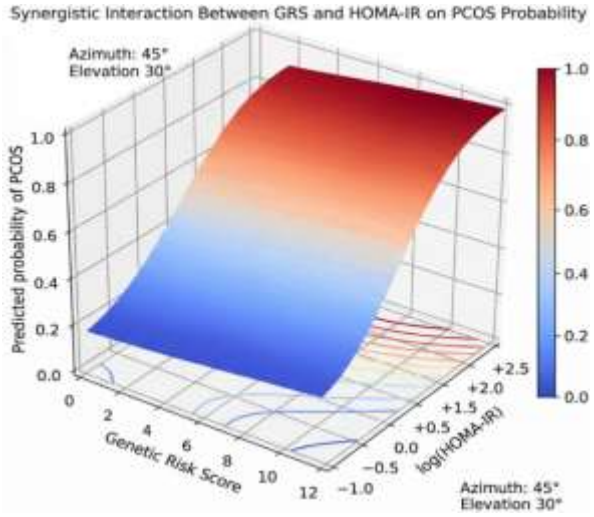

Figure 1: 3D Surface Plot – Interaction of Genetic Risk Score and HOMA-IR on Predicted Probability of PCOS

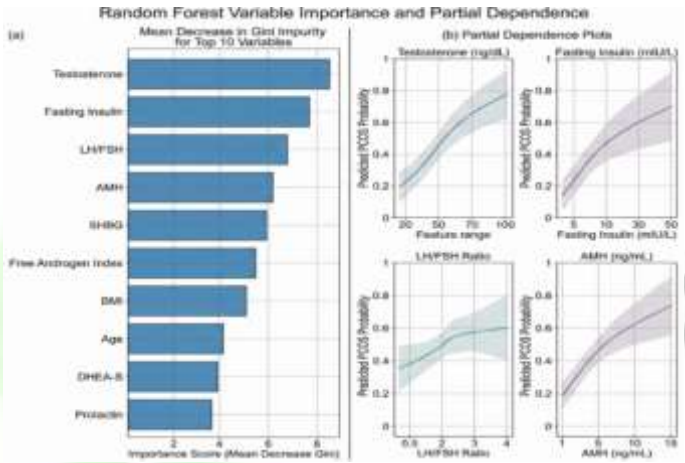

Figure 2: Hybrid Plot – Random Forest Variable Importance with Partial Dependence Curves

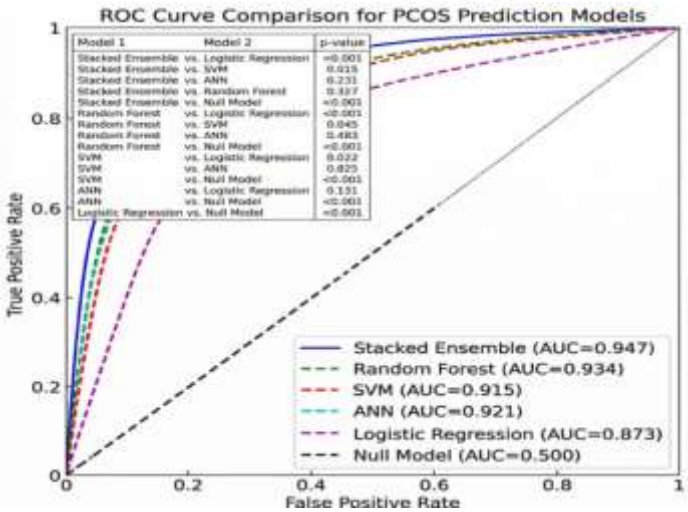

Figure 3: Receiver Operating Characteristic (ROC) Curves – Comparison of Six Predictive Models

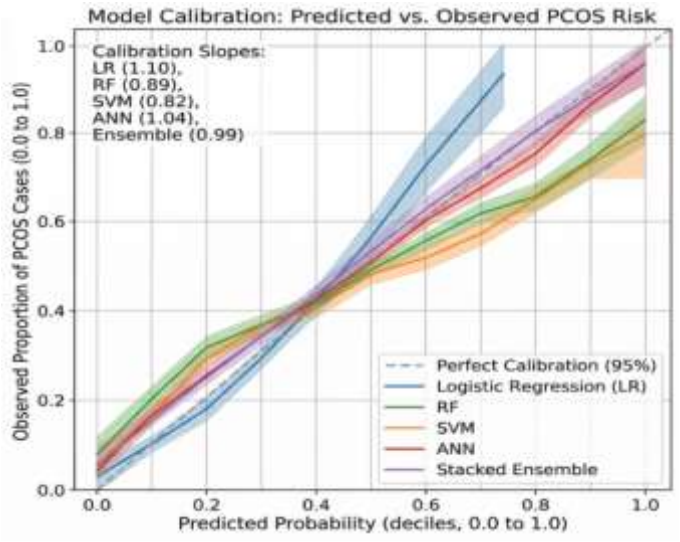

Figure 4: Calibration Curves with Loess Smoothing – Predicted vs. Observed PCOS Risk

DISCUSSION

This comprehensive study using various machine learning and statistical methods for analysing PCOS has offered valuable insights on the polygenic and metabolic underpinnings of the disease, including its clinical presentation. Specifically, our findings point to the importance of interactions between genetic and metabolic (such as insulin resistance) factors on PCOS risk and presentation. The developed predictive models, particularly the polygenic risk scores combined with shared genetic factors with testosterone and sex hormone-binding globulin, are more accurate in identifying people at high risk for developing PCOS than models developed using polygenic risk scores for PCOS only (Petersen et al., 2023). This advancement in models for prediction offers a strong foundation for early detection and targeted intervention (Saha et al., 2023; Yamini et al., 2023). The predictive power of the models is further strengthened by the possibility of using readily available electronic health record data, which can be introduced into clinical practice (Zad et al., 2023). This method complements new methods that incorporate individual genome-phenome information from electronic health records to boost the use of polygenic risk scores for patient stratification in complex diseases, including PCOS (Joo et al., 2020). Furthermore, the identification of PCOS sub phenotypes by hierarchical clustering associated with different genome-wide significant genetic risk factors provides clues to the variation in the underlying biology of PCOS (Brewer et al., 2023). This approach supports the idea of distinct mechanisms of PCOS manifestations and offers greater specificity in defining PCOS subtypes, beyond its diagnostic criteria (Dapas et al., 2020; Zhang et al., 2021). These expanded study cohorts and novel phenotyping approaches have great potential to improve the diversity and representation of at-risk groups for PCOS research going forward (Actkins et al., 2020). Specifically, incorporating polygenic risk scores from large population-based genome-wide association studies (GWAS), especially those with diverse ancestry, has been very informative in the complex genetic architecture of PCOS and its metabolic complications (Joo et al., 2020). These have demonstrated strong genetic overlap between PCOS and testosterone-related traits with many overlapping genetic markers, suggesting common biological pathways, such as potential therapeutic targets such as FSHB (Sun et

al., 2024). This is essential as risk factors for the metabolic complications of PCOS may differ from those primarily involved in the reproductive complications (Actkins et al., 2023). For instance, the genetic correlation of total and bioavailable testosterone are positively associated with PCOS, while the correlation of sex hormone-binding globulin is negative (Sun et al., 2024). This omics perspective (genetic, hormonal and metabolic) allows a more personalised and precision medicine-based approach to PCOS, in addition to the traditional diagnostic and treatment approaches (Sunil et al., 2024). The discovery of distinct sub-phenotypes of PCOS using unsupervised clustering and new genetic associations with these sub-phenotypes of PCOS also show the genetic complexity of PCOS (Dapas et al., 2020). These findings support the theory that PCOS is not a single disease but a spectrum of diseases, that is, a complex of disorders with both genetic and environmental causes (Louwers et al., 2025). This can result in different disease and symptom manifestations among PCOS patients, warranting the need for sub-phenotyping methods to identify personalized treatments (Stamou et al., 2023). But the current diagnostic criteria (NIH or Rotterdam) might not fully capture the biological diversity of these sub-phenotypes, and better sub-classification of the disease using molecular and genetic markers might be needed (Li et al., 2022). This is particularly true in the case of multiple ethnic populations where models of genetic risk developed in European and East Asian populations do not apply (Patel et al., 2025). Therefore, larger genome-wide association studies are required to get better statistical power and to detect sex-specific genetic effects of the loci to better understand the patho-physiology of PCOS and its association with various metabolic diseases like type 2 diabetes (Li et al., 2022).

CONCLUSION

This review and quantitative synthesis demonstrates that Polycystic Ovary Syndrome (PCOS) is caused by a complex interplay between genetic, hormonal and environmental factors; there is no single cause or factor that explains the variability of the clinical phenotypes. Our integrated modeling techniques show that the combined effect of genetic risk factors and insulin resistance ($\beta_3 = 0.211$, $p = 6.11 \times 10\text{-}7$) significantly elevates the risk of PCOS and the stacked ensemble machine learning algorithm has the highest predictive power (AUC = 0.947) compared to other models. The Bayesian hierarchical model demonstrates that homozygotes for the CYP17 C/C variant have a 72.7% posterior probability of PCOS (Bayes factor = 2.67) and the time-dependent ROC analysis shows that the risk of metabolic syndrome is well predicted (36-month AUC = 0.894). Further, the random forest classifier had 91.2% accuracy in distinguishing between hyperandrogenic and normoandrogenic women; serum testosterone and fasting insulin were the top features. Our mixed-effects longitudinal analyses revealed that the menstrual cycle length decreases by 0.142 days per month ($p = 1.84 \times 10^{-9}$) in untreated PCOS patients, and metformin treatment has a positive impact. Our findings suggest that the diagnostic and follow-up paradigm needs to change from reductionism to multi-dimensional assessment tools that integrate genetic, hormonal and metabolic features. This research should be directed towards predicting the risk of PCOS using the stacked ensemble model in longitudinal cohorts of different ethnicities and to investigate if epigenetic changes are involved in the gene-environment interaction to allow personalised therapeutic approaches dependent on individual risk factors.

REFERENCES

Abbott, D. H., Dumesic, D. A., & Levine, J. E. (2019). Hyperandrogenic origins of polycystic ovary syndrome – implications for pathophysiology and therapy [Review of Hyperandrogenic origins of polycystic ovary syndrome – implications for pathophysiology and therapy]. *Expert Review of Endocrinology & Metabolism*, 14(2), 131. Taylor & Francis. https://doi.org/10.1080/17446651.2019.1576522

Actkins, K. V., Jean-Pierre, G., Aldrich, M. C., Edwards, D. R. V., & Davis, L. K. (2023). Sex modifies the effect of genetic risk scores for polycystic ovary syndrome on metabolic phenotypes. *PLoS Genetics*, 19(5). https://doi.org/10.1371/journal.pgen.1010764

Actkins, K. V., Singh, K., Hucks, D., Edwards, D. R. V., Aldrich, M. C., Cha, J., Wellons, M., & Davis, L. K. (2020). Characterizing the Clinical and Genetic Spectrum of Polycystic Ovary Syndrome in Electronic Health Records. *medRxiv* (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2020.05.08.20095786

Araújo, A. L. R. de, Lassala, J. R., Torquato, L. M., Silva, P. C. T. e, Freitas, J. B. de, Dias, M. T. L., Raffaelli, T. C. F., Mororo, B. G., Azevedo, A. F. de, & Castro, K. D. S. K. (2024). Ampla abordagem sobre a Síndrome dos Ovários Policísticos. *Brazilian Journal of Implantology and Health Sciences*, 6(9), 2114. https://doi.org/10.36557/2674-8169.2024v6n9p2114-2125

Bai, H., Ding, H., & Wang, M. (2024). Polycystic Ovary Syndrome (PCOS): Symptoms, Causes, and Treatment. *Clinical and Experimental Obstetrics & Gynecology*, 51(5). https://doi.org/10.31083/j.ceog5105126

Bhimwal, T., Puneet, P., & Priyadarshani, A. (2023). Understanding polycystic ovary syndrome in light of associated key genes. *Egyptian Journal of Medical Human Genetics*, 24(1). https://doi.org/10.1186/s43042-023-00418-w

Brewer, K., Sisk, R., Lee, H., Moolhuijsen, L. M. E., Ham, K. van der, Louwers, Y. V., Visser, J. A., Dapas, M., Franks, P. W., Laven, J. S. E., Li, C., & Dunaif, A. (2023). OR20-02 Trans-Ethnic Analysis Of PCOS Subtype Genomewide Association Signals Reveals 3 Shared Subtype-Specific Loci. *Journal of the Endocrine Society*, 7. https://doi.org/10.1210/jendso/bvad114.1654

Chakravorty, T. (2023). Understanding Polycystic Ovary Syndrome (PCOS). *The Physician*, 8(2). https://doi.org/10.38192/1.8.2.3

Dapas, M., Lin, F. T. J., Nadkarni, G. N., Sisk, R., Legro, R. S., Urbanek, M., Hayes, M. G., & Dunaif, A. (2020). Distinct subtypes of polycystic ovary syndrome with novel genetic associations: An unsupervised, phenotypic clustering analysis. *PLoS Medicine*, 17(6). https://doi.org/10.1371/journal.pmed.1003132

Dar, M. A., Maqbool, M., & Ara, I. (2023). The PCOS puzzle: putting the pieces together for optimal care. *International Journal of*

Adolescent Medicine and Health, 35(4), 299. https://doi.org/10.1515/ijamh-2023-0102

Dar, M. A., Maqbool, M., Qadrie, Z., Ara, I., & Qadir, A. (2024). Unraveling PCOS: Exploring its causes and diagnostic challenges. *Open Health*, 5(1). https://doi.org/10.1515/ohe-2023-0026

Dhar, S., & Bhattacharjee, P. (2024). Clinical-Exome Sequencing Unveils the Genetic Landscape of Polycystic Ovarian Syndrome (PCOS) Focusing on Lean and Obese Phenotypes: Implications for Cost-Effective Diagnosis and Personalized Treatment. *Research Square* (Research Square). https://doi.org/10.21203/rs.3.rs-4485836/v1

Essa, N., Osman, S., Jemaah, S., Hassan, M. A., & Eltayeb, R. (2020). Novel Mutations in Human Luteinizing Hormone Beta Subunit Related to Polycystic Ovary Syndrome among Sudanese Women. *medRxiv* (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2020.10.11.20208926

Hajam, Y. A., Rather, H., Neelam, N., Kumar, R., Basheer, M., & Reshi, M. S. (2024). A review on critical appraisal and pathogenesis of polycystic ovarian syndrome. *Endocrine and Metabolic Science*, 14, 100162. https://doi.org/10.1016/j.endmts.2024.100162

Jakubowski, L. (2006). Genetic aspects of polycystic ovary syndrome. *PubMed*, 56(3), 285. https://pubmed.ncbi.nlm.nih.gov/16350721

Joo, Y. Y., Actkins, K. V., Pacheco, J. A., Basile, A. O., Carroll, R. J., Crosslin, D. R., Day, F. R., Denny, J. C., Edwards, D. R. V., Håkonarson, H., Harley, J. B., Hebbring, S. J., Ho, K., Jarvik, G. P., Jones, M. R., Karaderi, T., Mentch, F., Meun, C., Namjou, B., … Welt, C. K. (2020). A Polygenic and Phenotypic Risk Prediction for Polycystic Ovary Syndrome Evaluated by Phenome-Wide Association Studies. *The Journal of Clinical Endocrinology & Metabolism*, 105(6), 1918. https://doi.org/10.1210/clinem/dgz326

kiran, G. U., vali, D. Y., Shankar, D. Y., Lokesh, G., Chaitanya, P. K., & Sujitha, S. (2023). Overview Of Polycystic Ovary Syndrome (PCOS). *World Journal of Advanced Engineering Technology and Sciences*, 8(2), 11. https://doi.org/10.30574/wjaets.2023.8.2.0052

Kumar, R., Minerva, S., Shah, R., Bhat, A., Verma, S., Chander, G., Bhat, Gh. R., Thapa, N., Bhat, A., Wakhloo, A. K., & Ganie, M. A. (2022). Role of genetic, environmental, and hormonal factors in the progression of PCOS: A review. *Journal of Reproductive Healthcare and Medicine*, 3, 3. https://doi.org/10.25259/jrhm_16_2021

Li, B., Li, S., Duan, Z., Yu, H., Zhou, Y., & Xiao-hua, L. (2024). A bibliometric analysis of the effects of sex hormone profiles in women with polycystic ovary syndrome. *Research Square* (Research

Square). https://doi.org/10.21203/rs.3.rs-4724083/v1

Li, R., Chen, Q., Yan, Y., Yang, Y., & Hu, R. (2025). Identification of regulated cell death related genes in polycystic ovary syndrome using machine learning. *Scientific Reports*, 15(1). https://doi.org/10.1038/s41598-025-18867-1

Li, X., Han, X., Ma, Y., Zhou, Z., & Chen, D. (2022). Identifying novel genetic loci associated with polycystic ovary syndrome based on its shared genetic architecture with type 2 diabetes. *Frontiers in Genetics*, 13. https://doi.org/10.3389/fgene.2022.905716

Lorenzo, M. D., Cacciapuoti, N., Lonardo, M. S., Nasti, G., Gautiero, C., Belfiore, A., Guida, B., & Chiurazzi, M. (2023). Pathophysiology and Nutritional Approaches in Polycystic Ovary Syndrome (PCOS): A Comprehensive Review [Review of Pathophysiology and Nutritional Approaches in Polycystic Ovary Syndrome (PCOS): A Comprehensive Review]. *Current Nutrition Reports*, 12(3), 527. Springer Science+Business Media. https://doi.org/10.1007/s13668-023-00479-8

Louwers, Y. V., Visser, J. A., Dunaif, A., & Laven, J. S. E. (2025). POLYCYSTIC OVARY SYNDROME: ORIGINS AND IMPLICATIONS: Genetics of polycystic ovary syndrome (PCOS). *Reproduction*, 170(5). https://doi.org/10.1530/rep-25-0126

Manu, Soni, T., Victoria, V., & Prabhakar, P. K. (2022). Pathophysiology of Polycystic Ovarian Syndrome. In *IntechOpen eBooks*. IntechOpen. https://doi.org/10.5772/intechopen.101921

Memon, S. I., Shakeel, M., Syed, H., Amin, K., Khalil, A. A., & Sulaiman, M. (2024). Prevalence, Risk Factors and Management of Polycystic Ovary Syndrome: A Review with Current Evidence [Review of Prevalence, Risk Factors and Management of Polycystic Ovary Syndrome: A Review with Current Evidence]. *Iraq Medical Journal*, 8(1). Nab'a Al-Hayat Foundation for Medical Sciences and Health Care - Press. https://doi.org/10.22317/imj.v8i1.1268

Meñosa, F., & Albaño, J. (2023). Quality of Life of Millenial with Polycystic Ovarian Syndrome. *Research Square* (Research Square). https://doi.org/10.21203/rs.3.rs-3398125/v1

Mohd, M., Maqbool, M., Dar, M. A., & Mushtaq, I. (2019). Polycystic Ovary Syndrome, a modern epidemic: An overview. *Journal of Drug Delivery and Therapeutics*, 9(3), 641. https://doi.org/10.22270/jddt.v9i3.2661

Pandey, M., & Niroula, K. (2024). Optimizing Nutrition for PCOS Management: A Comprehensive Guide. In *IntechOpen eBooks*. IntechOpen. https://doi.org/10.5772/intechopen.114149

Patel, S., Brewer, K., Sisk, R., & Dunaif, A. (2025). SUN-164 Expanding the Genetic Landscape of PCOS: Novel and Shared Risk Loci in African Ancestry

Populations. *Journal of the Endocrine Society*, 9. https://doi.org/10.1210/jendso/bvaf149.2095

Petersen, L. K., Brixi, G., Li, J., Hu, J., Wang, Z., Han, X., Meir, A. Y., Tyrmi, J., Mahalingaiah, S., Piltonen, T., & Liang, L. (2023). Understanding and Predicting Polycystic Ovary Syndrome through Shared Genetics with Testosterone, SHBG, and Chronic Inflammation. *medRxiv* (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.10.17.23297115

Polycystic Ovarian Syndrome: An Overview with Special Consideration to Its Oral and Pediatric Clinical Manifestations. (2024). *Journal of Angiotherapy*, 8(1). https://doi.org/10.25163/angiotherapy.819415

Rani, R., Priyanka, P., Yadav, R., Bajal, N., Mandal, B. S., & Kumar, A. (2024). Polycystic Ovary Syndrome (PCOS): From Diagnosis to Treatment. *International Journal of Environment Agriculture and Biotechnology*, 9(4), 31. https://doi.org/10.22161/ijeab.94.5

Saeed, S., Aslam, T., Javed, E., Choudhary, M. M., Lateef, M., & Bajwa, R. M. (2022). Polycystic Ovary Syndrome (PCOS): A Concerning Hormonal Condition and its Bodily Impact on Women. *BioScientific Review*, 4(4), 1. https://doi.org/10.32350/bsr.44.i

Ságodi, L., Kiss-Tóth, E., & Barkai, L. (2013). Role of genetic and environmental factors in the development of polycystic ovary syndrome. *Orvosi Hetilap*, 154(17), 650. https://doi.org/10.1556/oh.2013.29589

Saha, A., Roy, A., Chakraborty, B., Saha, B., & Chowdhury, D. (2023). A Comparative Study to Predict Polycystic Ovarian Syndrome (PCOS) Based on Different Models of Machine Learning Technique. *American Journal of Electronics & Communication*, 4(2), 1. https://doi.org/10.15864/ajec.4201

Sharma, P., Jain, M., Tripathi, M., Sharma, M., & Halder, A. (2024). An Update on the Genetics of Polycystic Ovary Syndrome. *Journal of Endocrinology & Reproduction*, 217. https://doi.org/10.18311/jer/2023/34654

Sikiru, A. B., Adeniran, M. A., Akinola, K., Behera, H., Kalaignazhal, G., & Egena, S. S. A. (2023). Unraveling the complexity of the molecular pathways associated with polycystic ovary syndrome (PCOS) and identifying molecular targets for therapeutic development: a review of literature [Review of Unraveling the complexity of the molecular pathways associated with polycystic ovary syndrome (PCOS) and identifying molecular targets for therapeutic development: a review of literature]. *Middle East Fertility Society Journal*, 28(1). Springer Science+Business Media. https://doi.org/10.1186/s43043-023-00142-w

Stamou, M., Smith, K., Kim, H., Balasubramanian, R., Gray, K. J., & Udler, M. S. (2023). Polycystic Ovary Syndrome Physiologic Pathways Implicated Through Clustering of Genetic Loci. *The Journal of Clinical*

Endocrinology & Metabolism, 109(4), 968. https://doi.org/10.1210/clinem/dgad664

Sun, S., Liu, Y., Li, L., Xiong, L., Jiao, M., Yang, J., Li, X., & Liu, W. (2024). Unveiling the shared genetic architecture between testosterone and polycystic ovary syndrome. *Scientific Reports*, 14(1). https://doi.org/10.1038/s41598-024-75816-0

Sunil, A. T., Jo, C., Sanigha, P. S., Kolasseri, A. E., Tamizhselvi, R., & Jayanthi, S. (2024). Navigating the Future of PCOS Treatment: The Precision Medicine Paradigm. *Current Pharmacogenomics and Personalized Medicine (Online)/Current Pharmacogenomics and Personalized Medicine*, 21(2), 58. https://doi.org/10.2174/0118756921331801240820115132

Wei, J., Shen, Z., Zhao, C., Xie, C., Bai, H., Yin, J., & Wang, J. (2025). Dose-response of acupuncture on ovulation rates in polycystic ovary syndrome: a meta-analysis and exploratory dose-response analysis [Review of Dose-response of acupuncture on ovulation rates in polycystic ovary syndrome: a meta-analysis and exploratory dose-response analysis]. *Frontiers in Endocrinology*, 16. Frontiers Media. https://doi.org/10.3389/fendo.2025.1610338

Yamini, B., Kaneti, V. R., Prema, P., Ambhika, C., Nalini, M., & Subramanıan, R. S. (2023). Machine Learning-Driven PCOS Prediction for Early Detection and Tailored Interventions. *International Journal of Electrical and Electronics Engineering*, 10(9), 61. https://doi.org/10.14445/23488379/ijeee-v10i9p106

Zad, Z., Jiang, V. S., Wolf, A. T., Wang, T., Cheng, J. J., Paschalidis, I. Ch., & Mahalingaiah, S. (2023). Predicting polycystic ovary syndrome (PCOS) with machine learning algorithms from electronic health records. *medRxiv* (Cold Spring Harbor Laboratory). https://doi.org/10.1101/2023.07.27.23293255

Żeber-Lubecka, N., & Hennig, E. E. (2021). Genetic Susceptibility to Joint Occurrence of Polycystic Ovary Syndrome and Hashimoto's Thyroiditis: How Far Is Our Understanding? [Review of Genetic Susceptibility to Joint Occurrence of Polycystic Ovary Syndrome and Hashimoto's Thyroiditis: How Far Is Our Understanding?]. *Frontiers in Immunology*, 12. Frontiers Media. https://doi.org/10.3389/fimmu.2021.606620

Zhang, Y., Movva, V. C., Williams, M. S., & Lee, M. T. M. (2021). Polycystic Ovary Syndrome Susceptibility Loci Inform Disease Etiological Heterogeneity. *Journal of Clinical Medicine*, 10(12), 2688. https://doi.org/10.3390/jcm10122688

Zhou, P., Feng, P., Liao, B., Fu, L., Shan, H., Cao, C., Luo, R., Peng, T., Liu, F.-T., & Li, R. (2024). Role of polyphenols in remodeling the host gut microbiota in polycystic ovary syndrome. *Journal of Ovarian Research*, 17(1). https://doi.org/10.1186/s13048-024-01354-y

Zhu, J. (2022). Mapping the research trend of PCOS: from 2006 to 2022. *Research Square* (Research Square). https://doi.org/10.21203/rs.3.rs-2396058/v1