



ΤΜΗΜΑ ΙΑΤΡΙΚΗΣ
Σχολή Επιστημών Υγείας
Πανεπιστήμιο Θεσσαλίας



Πρόγραμμα Μεταπτυχιακών Σπουδών (ΠΜΣ)

«Μεθοδολογία Βιοϊατρικής Έρευνας, Βιοστατιστική και Κλινική Βιοπληροφορική»

Διπλωματική Εργασία

Μελέτη γενετικής συσχέτισης γονιδιωματικής κλίμακας πολυμορφισμών που σχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών σε Ευρωπαϊκό Πληθυσμό.

Genome-wide association study for polymorphisms associated with PCOS in European Population.

Τοκαμάνη Μαρία, *PhD*

Μοριακός Βιολόγος & Γενετίστρια

ΑΜ:00202

ΑΓΜ: M060618015

Ακαδημαϊκό Έτος 2018-2019

Τριμελής Επιτροπή

Δοξάνη Χρυσούλα, *MSc, MD, PhD* (Επιβλέπουσα)
Ακαδημαϊκός Υπότροφος στο Εργαστήριο Βιομαθηματικών

Στεφανίδης Ιωάννης, *MD, PhD* (Μέλος)
Καθηγητής Παθολογίας-Νεφρολογίας
Διευθυντής Νεφρολογικής Κλινικής
Τμήμα Ιατρικής, Πανεπιστήμιο Θεσσαλίας

Ζιντζαράς Ηλίας, *MSc, PhD* (Μέλος)
Καθηγητής Βιομετρίας-Βιομαθηματικών
Διευθυντής Εργαστηρίου Βιομαθηματικών
Τμήμα Ιατρικής, Πανεπιστήμιο Θεσσαλίας

A. Περίληψη:

A. Περίληψη

Εισαγωγή: Το σύνδρομο πολυκυστικών ωοθηκών (PCOS, ΣΠΩ) είναι μια πολύπλοκη και ετερογενής ασθένεια με ενδοκρινολογικές και μεταβολικές διαταραχές. Προσβάλλει το 5-20% του γυναικείου πληθυσμού και είναι η κύρια αιτία υπογονιμότητας παγκοσμίως. Η μελέτη γενετικών πολυμορφισμών μεταξύ γυναικών με PCOS μπορεί να οδηγήσει στην καλύτερη κατανόηση των μοριακών μηχανισμών που το προκαλούν, στη βελτιστοποίηση των διαγνωστικών και των θεραπευτικών προσεγγίσεων μακροπρόθεσμα.

Στόχος: Αξιοποιώντας δεδομένα από τη βιοτράπεζα του Ηνωμένου Βασιλείου, γυναικών με Ευρωπαϊκή καταγωγή, σκοπός είναι να συσχετισθούν πιθανοί γενετικοί τόποι με το ΣΠΩ.

Μέθοδοι: Συλλέγοντας γενετικά δεδομένα γονιδιωματικής κλίμακας γυναικών ευρωπαϊκής καταγωγής που πάσχουν από PCOS βάσει ίδιας δήλωσης (συμμετέχουσες 194.174, πάσχουσες 687), αναλύθηκαν με λογιστική παλινδρόμηση 13.7 εκατομμύρια πολυμορφισμοί.

Αποτελέσματα και Συμπεράσματα: Από την ανάλυση γενετικής συσχέτισης προέκυψαν 2 SNPs με στατιστικά σημαντική συσχέτιση με το ΣΠΩ (όριο σημαντικότητας 5×10^{-8}) και 171 οριακά σημαντικοί (όριο σημαντικότητας 10^{-5}). Οι πολυμορφισμοί συγκεντρώνονται σε 12 γενετικούς τόπους, από τους οποίους οι 9 περιέχουν κωδικά και μη κωδικά γονίδια που συμμετέχουν σε διάφορες κυτταρικές διεργασίες, την απόκριση του ανοσοποιητικού και τον κυτταρικό μεταβολισμό. Οι γενετικοί αυτοί τόποι θα μπορούσαν να αποτελέσουν επίκεντρο έρευνας για την κατανόηση του ρόλου τους στο σύνδρομο.

Λέξεις Κλειδιά: Σύνδρομο Πολυκυστικών Ωοθηκών, Ανάλυση Γενετικής Συσχέτισης, Πολυμορφισμοί.

A. Abstract:

Introduction: Polycystic ovary syndrome (PCOS) is a complex disease, with multiple phenotypic traits. It brings forth a plethora of endocrinological and metabolic disorders. Polycystic ovary syndrome affects about 5-20% of the female population and it is the main reason of infertility worldwide. Genome wide association studies could provide better understanding of the molecular mechanism that lead to PCOS of different phenotypes. Also, they could offer improved diagnostic and therapeutic methods.

Aim: Using UK biobank data from women of European ancestry, we aim to find genetic loci associated with PCOS.

Method: Genetic high throughput data of 13.7 million polymorphisms from 194.174 females of European ancestry, of which 687 had PCOS, were analyzed using logistic regression analysis.

Results and Conclusion: From the GWAS emerged 2 SNPs with a statistically significant association to PCOS (significance threshold $<5 \times 10^{-8}$) and 171 marginally significant (significance threshold $<10^{-5}$). Polymorphisms are located in 12 genetic loci, of which 9 contain coding and non-coding genes. These genes are implicated in many cellular processes, in immunogenic response and in cellular metabolism. These genetic sites are promising as focal points for future research on PCOS.

Keywords: Polycystic ovary syndrome (PCOS), Genome Wide Association Study (GWAS), Single Nucleotide Polymorphisms (SNPs).

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τα μέλη της τριμελούς επιτροπής, την κα Δοξάνη Χρυσούλα, Ακαδημαϊκός Υπότροφος στο Εργαστήριο Βιομαθηματικών, τον καθηγητή κ. Στεφανίδη Ιωάννη και τον καθηγητή κ. Ζιντζαρά Ηλία για τις παρατηρήσεις τους. Επίσης θέλω να ευχαριστήσω θερμά τον καθηγητή του τμήματος Μοριακής Βιολογίας και Γενετικής (ΔΠΘ), κ. Ραφαήλ Σανδαλτζόπουλο για τη στήριξή του αυτήν τη χρονιά. Οφείλω να ευχαριστήσω τον μεταπτυχιακό φοιτητή κ. Μαργαρίτη Τσιφιντάρη για τις πολύτιμες συμβουλές του στα θέματα Πληθυσμιακής Γενετικής. Τέλος, το μεγαλύτερο ευχαριστώ αρμόζει στην οικογένειά μου, τους γονείς μου Βάσω και Παντελή και τον αδερφό μου Κωνσταντίνο, για τη συνεχή και ανιδιοτελή τους υποστήριξη και ενίσχυση.

Περιεχόμενα

A. Περίληψη:	2
B. Εισαγωγή	6
Σύνδρομο Πολυκυστικών Ωοθηκών	6
Κλινική Εικόνα και Επιδημιολογία	6
Παθοφυσιολογία και Μοριακοί Μηχανισμοί	6
Μελέτες Γενετικής Συσχέτισης γονιδιωματικής κλίμακας	8
Σκοπός	11
Γ. Μέθοδος	12
Συλλογή Δεδομένων	12
Ανάλυση Δεδομένων	12
Ανάλυση Γονιδιακής Οντολογίας	14
Δ. Αποτελέσματα	15
Ε. Συμπεράσματα	21
Στ. Αναφορές	25

B. Εισαγωγή

Σύνδρομο Πολυκυστικών Ωοθηκών

Κλινική Εικόνα και Επιδημιολογία

Το σύνδρομο πολυκυστικών ωοθηκών (Polycystic Ovary Syndrome -PCOS, ΣΠΩ) είναι συχνή ασθένεια σε γυναίκες της αναπαραγωγικής ηλικίας και προσβάλλει περίπου το 5-20% του πληθυσμού των γυναικών παγκοσμίως [1]. Για πρώτη φορά ο όρος χρησιμοποιήθηκε από τους Stein και Leventhal το 1936 για να περιγράψει μια παθολογική κατάσταση που συνοδευόταν από παχυσαρκία, υπερτρίχωση, ανωοθυλακιορρηξία, διαταραχές περιόδου και λευκωπές διογκωμένες ωοθήκες [2]. Το σύνδρομο πολυκυστικών ωοθηκών είναι η κύρια αιτία υπογονιμότητας παγκοσμίως και κύριο χαρακτηριστικό του είναι η ύπαρξη πολλών μη ώριμων ωοθυλακίων στην ωοθήκη ταυτόχρονα. Ωστόσο, δεν συνεπάγεται ότι η ύπαρξη πολλών ωοθυλακίων οδηγεί απαραίτητα στην εμφάνιση του συνδρόμου. Το σύνδρομο των πολυκυστικών ωοθηκών είναι μια ιδιαίτερα πολύπλοκη και ετερογενής ασθένεια με πολύπλοκες ενδοκρινολογικές (κυρίως στην παραγωγή ανδρογόνων) και μεταβολικές διαταραχές. [3]

Το σύνδρομο PCOS παρουσιάζει ποικίλα χαρακτηριστικά τα οποία διαχωρίζουν τις πάσχουσες γυναίκες σε 4 βασικούς υποφαινότυπους ανάλογα με τα συμπτώματα (κριτήρια 2003 Rotterdam) [4]. Ο πιο «βαρύς» φαινότυπος Α φέρει όλα τα συμπτώματα, υπερανδρογοναιμία, ωοθηκική δυσλειτουργία και πολυκυστικής μορφολογίας ωοθήκες. Ο φαινότυπος Β παρουσιάζει υπερανδρογοναιμία και ωοθηκική δυσλειτουργία, ο φαινότυπος Γ υπερανδρογοναιμία και πολυκυστικής μορφολογίας ωοθήκες, και τέλος ο φαινότυπος Δ ωοθηκική δυσλειτουργία και πολυκυστικής μορφολογίας ωοθήκες. Οι φαινότυποι Α και Β μαζί αφορούν περίπου το 40-45%, ο Γ το 35% και ο Δ το 20% των περιπτώσεων PCOS. [1,4]

Η κύρια συμπτωματολογία του συνδρόμου περιλαμβάνει:

- α) διαταραχές κατά τον εμμηνορρυσιακό κύκλο: αραιομηνόρροια, αμηνόρροια, δυσμηνόρροια
- β) δερματολογικές διαταραχές: δακτυλιδισμός, ακμή, λιπαρότητα [5]
- γ) μεταβολική δυσλειτουργία: υπερινσουλιναίμία ανεξάρτητη από τον δείκτη μάζας σώματος, αυξημένο κίνδυνο εμφάνισης διαβήτη κύησης και διαβήτη τύπου 2 [6-12]
- δ) παχυσαρκία [13,14]
- ε) υπογονιμότητα και επιπλοκές κατά την κύηση που δεν εξηγούνται από τη χρήση υποβοηθούμενης τεκνοποίησης [15,16]
- στ) ψυχολογικές διαταραχές: κατάθλιψη, που σχετίζεται με την αντίσταση στην ινσουλίνη [17], και άγχος [18-20]
- ζ) αυξημένο κίνδυνο δημιουργίας κακοηθειών (OR 2.7) [21]

Παθοφυσιολογία και Μοριακοί Μηχανισμοί

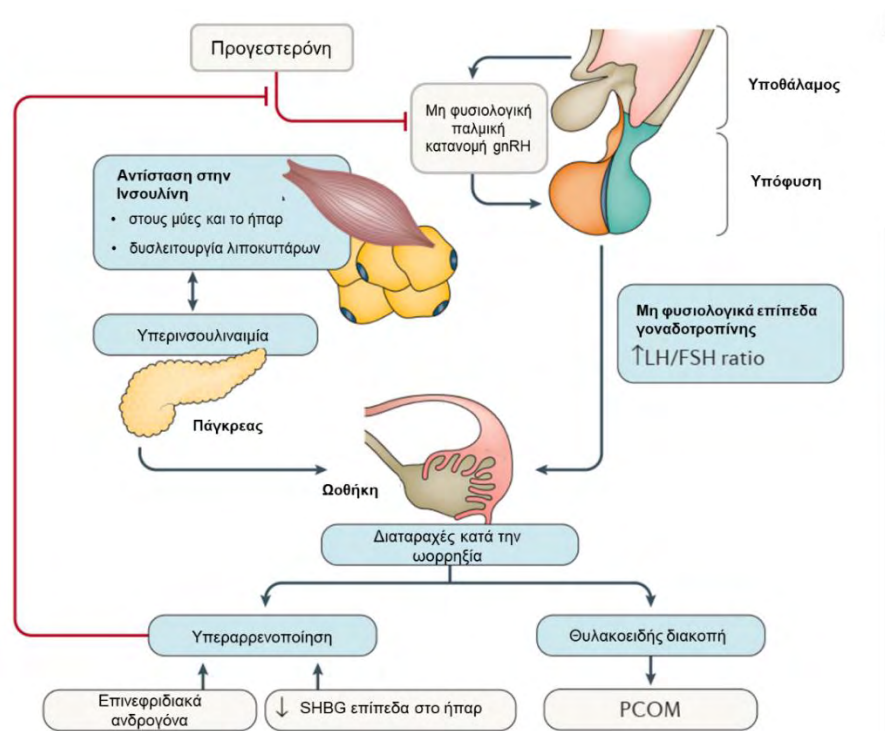
Η παθοφυσιολογία και ο μηχανισμός εμφάνισης του συνδρόμου είναι ιδιαίτερα πολύπλοκοι καθ' ότι οι αιτίες ποικίλουν και τα διαφορετικά συμπτώματα είναι αλληλένδετα μεταξύ τους (**εικόνα 1**). Είναι πλέον όμως

σίγουρο ότι, η εμφάνιση του συνδρόμου πολυκυστικών ωοθηκών έχει ισχυρό γενετικό υπόβαθρο και κληρονομικότητα που επηρεάζεται παράλληλα από περιβαλλοντικούς παράγοντες. Φαίνεται να δρα ως αυτοσωμικό επικρατές νόσημα, που επηρεάζει μόνο τις γυναίκες.

Γενετικά αίτια όπως πολυμορφισμοί ή μεταλλάξεις προκαλούν απορρύθμιση του μονοπατιού βιοσύνθεσης στεροειδών ορμονών. Σε αυτό κύριο ρόλο φαίνεται να παίζει η ωοθήκη και σε δευτερεύοντα ρόλο η δυσλειτουργία του νευράξονα υποθαλάμου-υπόφυσης. Η πολυκυστική ωοθήκη έχει αυξημένη ικανότητα να εκκρίνει ανδρογόνα μέσω της αυξημένης κυτταρικής διαίρεσης των κυττάρων της θήκης και αύξηση της βιοσύνθεσης των στεροειδών. Η απορρύθμιση των παραπάνω διεργασιών προκαλείται από μεταλλάξεις σε αυξητικούς παράγοντες και βιοχημικά ένζυμα ή λόγω απορρύθμιση της έκφρασής τους.

Η βιοσύνθεση των στεροειδών από την ωοθήκη και τα επινεφρίδια δεν επιτρέπουν την ολοκλήρωση της ωρίμανσης των ωοθυλακίων. Κατά συνέπεια, δεν αυξάνεται η συγκέντρωση της προγεστερόνης, που μειώνει την αρνητική επανατροφοδότηση στον υποθάλαμο, οδηγώντας έτσι σε συχνότερες παλμικές εκκρίσεις της ελκτικής ορμόνης γοναδοτροπινών GnRH. Υπό φυσιολογικές συνθήκες για την ωρίμανση ενός ωαρίου κύριο ρόλο παίζει η περιοδική αυξομείωση των γοναδοτροπινών LH και FSH και κυρίως ο λόγος τους. Ο λόγος αυτός ρυθμίζεται από τις παλμικές εκκρίσεις της GnRH που οδηγεί σε αύξηση της LH δρώντας στην υπόφυση. Επομένως, η LH διεγείρει τη σύνθεση και έκκριση ανδρογόνων. Επίσης, τα αυξημένα ανδρογόνα προκαλούν μείωση της SHBG που φυσιολογικά μειώνει τα ελεύθερα δραστικά ανδρογόνα. Δημιουργείται, συνεπώς, ένα κύκλος ενίσχυσης της σύνθεσης ανδρογόνων οδηγώντας έτσι σε συχνότερες παλμικές εκκρίσεις της GnRH. Η συχνότερη παλμική έκκριση της GnRH απορρυθμίζει το σύστημα προκαλώντας υπερανδρογοναιμία και συσσώρευση μικρών ωοθυλακίων που δεν ολοκληρώνουν τη διαδικασία της ωρίμανσής τους (follicular arrest.) [^{3,22-24}]

Εξίσου σημαντικό ρόλο παίζουν περιβαλλοντικοί παράγοντες, που σχετίζονται άμεσα με την ανάπτυξη παχυσαρκίας, όπως η διαίτα και η άσκηση, μέσω της επίδρασης της ινσουλίνης και του ινσουλινόμορφου αυξητικού παράγοντα I (IGF-I). Φαίνεται ότι γυναίκες με σύνδρομο πολυκυστικών ωοθηκών εμφανίζουν αντίσταση στην ινσουλίνη και υπερινσουλιναιμία ταυτόχρονα με μη σωστή λειτουργία των β-παγκρεατικών κυττάρων. [²⁵] Σε αντίθεση με τους πάσχοντες από διαβήτη τύπου 2, οι γυναίκες με σύνδρομο πολυκυστικών ωοθηκών παρουσιάζουν αντίσταση στην ινσουλίνη λόγω της υπερενεργοποίησης (μέσω φωσφορυλίωσης) του υποδοχέα της ινσουλίνης IGR και του υποστρώματός του IRS1, στους σκελετικούς μύες, με αποτέλεσμα την ιδιοστατική ενεργοποίηση καθοδικών μορίων (MEK1 και MEK2) του μονοπατιού. [^{26,27}] Δυσλειτουργία του λιπώδη ιστού οδηγεί και αυτή με τη σειρά της σε αντίσταση στην ινσουλίνη και φλεγμονώδη αντίδραση. [^{28,29}] Κατ' επέκταση η αντίσταση στην ινσουλίνη και η αντισταθμιστική υπερινσουλιναιμία συμβάλουν στην υπερανδρογοναιμία διότι προκαλούν την παραγωγή ανδρογόνων, τόσο άμεσα επηρεάζοντας την έκφραση ενζύμων της βιοσύνθεσής τους, όσο και έμμεσα αναστέλλοντας τη σύνθεση της SHBG. Παράλληλα, μειώνουν την έκφραση μιας πρωτεΐνης της IGF-BP, που δεσμεύει τον παράγοντα IGF-1 στο ωοθυλάκιο, συνεπώς αυξάνεται η απελευθέρωση του IGF-1 για να δράσει στο ορμονοπαραγωγικό σύστημα των κυττάρων της θήκης. [³]



Εικόνα 1 Η παθοφυσιολογία του συνδρόμου πολυκυστικών ωοθηκών [Τροποποιημένη από ³].

Εν κατακλείδι, μέσα από μελέτες έχουν απομονωθεί περίπου 100 γονίδια που σχετίζονται με το σύνδρομο και συμμετέχουν σε μονοπάτια που προκαλούν τους διάφορους φαινοτύπους. Τα πιο σημαντικά γονίδια και η βιολογική διεργασία που συμμετέχουν είναι [^{30,31}]:

- Βιοσύνθεση στεροειδών και ανδρογόνων ορμονών: *cyp11a1, cyp11a, cyp17a1, cyp19, hsd17b6, gata4, fdft1*
- Δράση ανδρογόνων: *ar, shbg, srd5a1, srd5a2*
- Σηματοδότηση Ινσουλίνης: *insr, irs1, irs2, pparg, capn10, thada, dennd1a, rab5b/suox*
- Μεταβολισμός και αποθήκευση γλυκογόνου: *adipoq, fto, gys2*
- Δράση Γοναδοτροπίνης και Γένεση Θυλακίων: *fshr, fshb, lhcg, amhr2*
- Φλεγμονώδης αντίδραση: *il1a, il1b, il6, il18, pall, fbn1, tnfr, mepl1a*
- Κυτταρική ανάπτυξη, επιδιόρθωση γενετικού υλικού, δράση τελομερασών: *c9orf3, yap1, neil2, rad50, krr1, khdrbs3, tox3, sumo1p1, znf217, hmga2*
 - Υποδοχείς αυξητικών παραγόντων: *errb3/her3, errb4/her4, errb2/her2*
- Ανάπτυξη αναπαραγωγικού συστήματος: *gtf2a1l*

Μελέτες Γενετικής Συσχέτισης γονιδιωματικής κλίμακας

Έχουν πραγματοποιηθεί τέσσερις κύριες μελέτες γενετικής συσχέτισης για να διερευνηθούν πιθανοί γενετικοί τόποι που είτε συνδέονται με το σύνδρομο είτε είναι η αιτία εμφάνισης του [^{25,32–34}], καθώς και δευτερεύουσες μελέτες που είτε επικεντρώνονται σε συγκεκριμένους γενετικούς τόπους, μεταξύ ενός πληθυσμού κοινής καταγωγής, είτε μελέτες διδύμων ή οικογένειας ή κοορτής. Παράλληλα έχει πραγματοποιηθεί μία μεγάλης κλίμακας μετανάλυση από την ερευνητική ομάδα του F. Day σε μία προσπάθεια

να απομονωθούν οι κυριότεροι γενετικοί τόποι. Η μετανάλυση βασίστηκε σε 7 μελέτες κοορτής ατόμων Ευρωπαϊκής καταγωγής (10,074 πάσχουσες και 103,164 μάρτυρες), από την οποία προέκυψαν 14 πολυμορφισμοί που οι 3 αναφέρθηκαν για πρώτη φορά.^[35] Οι πολυμορφισμοί που αφορούν το σύνδρομο πολυκυστικών ωοθηκών έως σήμερα έχουν συνοψισθεί στον **πίνακα 1**.

Πίνακας 1 Μονονουκλεοτιδικοί πολυμορφισμοί που σχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών, ταξινομημένοι βάση της θέσης του σε χρωμοσώματα, σύμφωνα με την έκδοση hg19 του γονιδιώματος.

Χρ.	Θέση (hg19)	rsID	Ref. Allele	Minor Allele	Πληθυσμός	Γονίδια	Αναφ.
1	887801	rs3828047	A	G, T	Korean	<i>smc3, noc2l</i>	36
1	3109151	rs7539775	G	A	Korean	<i>gnaq, prdm16</i>	36
1	3538501	rs4364818	C	T	Korean	<i>ccne2, al513320.1, tprgl1</i>	36
1	4312205	rs1908490	C	T	Korean	<i>plcb3</i>	36
1	4355593	rs2029490	A	G	Korean	<i>plcb2</i>	36
1	5072295	rs1870516	C	T	Korean	<i>stxbp1</i>	36
1	7999602	rs161811	C	T	Korean	<i>spdy, tnfrsf9, al009183.1</i>	36
1	158945708	rs16841398	A	G	Korean	<i>pyhin1</i>	37
1	233278612	rs2477858	A	G	Korean	<i>pcnx2</i>	38
1	239840107	rs1867265	C	T	Korean	<i>chrn3</i>	37
2	38280021	rs7557493	A	C	Korean	<i>rmdn2, rmdn2-as1</i>	37
2	43532302	rs11891936	C	T	European	<i>thada</i>	39
2	43561161	rs12468394	C	A	European, Chinese	<i>thada</i>	39,40
2	43561780	rs7563201	A	G	European	<i>thada</i>	32
2	43638838	rs13429458	A	C	Han Chinese, European	<i>rnu6-958p, thada</i>	33,39,40
2	43711925	rs6544661	A	G	European	<i>thada</i>	39
2	43721508	rs12478601	C	T	Han Chinese	<i>thada, rn7sl531p</i>	33,40
2	48921375	rs2293275	T	C	Sardinian	<i>ac073082.1, ston1-gtf2a1l, lhgr, gtf2a1l</i>	41
2	48921438	rs12470652	T	C		<i>ac073082.1, ston1-gtf2a1l, lhgr, gtf2a1l</i>	42
2	48965603	rs6732721	T	C	European	<i>lhgr, ston1-gtf2a1l, ac073082.1</i>	39
2	48978159	rs13405728	A	G	Han Chinese	<i>ac073082.1, ston1-gtf2a1l, lhgr</i>	33,40
2	49189921	rs6166	C	T	Dutch, Japanese	<i>ac009975.1, fshr</i>	43-48
2	49191041	rs6165	C	G, T	Italian, Chinese	<i>ac009975.1, fshr</i>	43-48
2	49201612	rs2268361	C	T	Han Chinese, European (Greek)	<i>fshr, ac009975.1</i>	34,40,49,50
2	49226271	rs12994034	A	G	European	<i>fshr, ac009975.1</i>	39
2	49247832	rs2349415	T	C	Han Chinese, European	<i>fshr</i>	34,39,40,49
2	213391766	rs2178575	G	A	European	<i>erbb4</i>	32
2	213394712	rs1351592	C	G	European	<i>erbb4</i>	32
2	228481419	rs11889798	C	T	Korean	<i>c2orf83, ac064853.1</i>	37
3	99477182	rs792837	C	T	Saudi Arabs	<i>col8a1</i>	51
3	145518144	rs11709200	T	C	Korean		38
3	179831733	rs7652876	C	A	Korean		38
4	47003076	rs1159315	T	C	Korean	<i>gabrbl, gabra4</i>	37,38
4	157586380	rs1521716	A	C	Korean		38
4	189477801	rs11930115	G	A	Korean	<i>linc01060</i>	38
4	189498064	rs7666129	C	T	Korean	<i>linc01060</i>	38
5	131813204	rs13164856	T	C	European	<i>irfl/rad50</i>	32
5	167341594	rs11134480	T	C	Korean	<i>tenm2</i>	38
6	68908391	rs1112082	T	A	Korean	<i>ppp2r5a</i>	36
6	112500650	rs1158747	T	G	Korean	<i>ppp1cb, lama4</i>	36

6	142506290	rs225673	T	C		<i>vta1</i>	52
8	11278146	rs11250135	T	G	Korean	<i>fam167a-as1, fam167a</i>	38
8	11623889	rs804279	A	T	European	<i>gata4, neil2, c8orf49</i>	25
8	11623889	rs804279	A	T	European	<i>gata4/neil2</i>	25
8	118184783	rs13266634	C	T	Saudi Arabs	<i>slc30a8</i>	51
8	137156450	rs10505648	A	G	Korean	<i>ac079098.1</i>	38
9	5440589	rs10739076	C	A	European	<i>plgrkt</i>	35
9	78268853	rs4745427	C	T	Korean		38
9	85704575	rs11536913	C	T	Korean		38
9	97505608	rs7866848	G	A	European	<i>c9orf3</i>	39
9	97648587	rs4385527	G	A	Han Chinese	<i>c9orf3</i>	34,40,49
9	97679771	rs10993397	C	T	European	<i>c9orf3</i>	25
9	97723266	rs7864171	G	A	European	<i>fancc</i>	25
9	97741336	rs3802457	G	A	Han Chinese	<i>c9orf4</i>	34,40,49
9	120480812	rs1554973	T	C	Korean	<i>tlr4</i>	37
9	126446778	rs10818854	G	A	Chinese	<i>dennd1a</i>	33,40
9	126525212	rs2479106	A	G	Han Chinese, European	<i>dennd1a, al390774.1</i>	33,39,40
9	126549955	rs10986105	T	G	Chinese	<i>dennd1a</i>	40
9	126566945	rs12337273	A	G	European	<i>dennd1a</i>	39
9	126619233	rs9696009	G	A	European	<i>dennd1a</i>	
9	126864196	rs10760321	A	G	European		32
11	2491883	rs61873498	C	G	Saudi Arabs	<i>kcnq1</i>	51
11	30226356	rs11031005	T	C	European	<i>arl14ep/fshb</i>	25,32
11	30226528	rs11031006	G	A	European	<i>kcna4, fshb</i>	25,32
11	33376095	rs4755571	A	G	Korean	<i>hipk3</i>	38
11	76648121	rs10751267	T	G	Korean	<i>acer3</i>	38
11	102043240	rs11225154	G	A	European	<i>yap1</i>	32
11	102070639	rs1894116	A	G	Han Chinese, European (Greek)	<i>yap1</i>	34,40,49,50
11	113949232	rs1784692	A	G	European	<i>zbtb16</i>	35
11	116409093	rs10502220	T	C	Korean		38
12	21691891	rs10841843	T	C	Korean	<i>gys2</i>	53
12	21699508	rs1904121	T	C	Korean	<i>gys4</i>	53
12	21699928	rs6487237	C	A	Korean	<i>gys3</i>	53
12	56390636	rs705702	A	G	Han Chinese	<i>rab5b, suox</i>	34,40,49,54
12	56467587	rs7312770	C	T	European		32
12	56477694	rs2271194	A	T	European	<i>erbb3/rab5b</i>	32
12	66224461	rs2272046	A	C	Han Chinese, European	<i>hmga2, rpsap52, ac107308.1</i>	34,39,40,49
12	75935157	rs1275468	T	C	European		32
12	75941042	rs1795379	C	T	European	<i>krr1</i>	32
13	54770376	rs9563201	A	T	European		32
14	75734471	rs8004059	G	A	Korean	<i>af111167.1, al691403.1</i>	38
16	1050877	rs500492	T	C	Korean	<i>ac009041.1</i>	38
16	52347819	rs4784165	T	G	Han Chinese, European (Greek)	<i>tox3</i>	34,40,49,50
16	52375777	rs8043701	A	T	European	<i>tox3</i>	
17	31373402	rs16589	T	C	Korean	<i>asic2</i>	37
17	31380311	rs12450465	C	T	Korean	<i>ac003687.1, asic2</i>	37
17	31380860	rs1988598	A	G	Korean	<i>ac003687.1, asic2</i>	37
17	31391470	rs8071961	T	A, G	Korean	<i>asic2</i>	37
17	31399596	rs223031	G	A	Korean	<i>asic2</i>	37
17	37725404	rs7218361	A	G	European	<i>ac009283.1, cdk12</i>	32
17	37982883	rs2060941	G	T	Korean	<i>ikzf3</i>	37
17	38082831	rs17608925	T	C	Korean	<i>ac090844.2, ormdl3</i>	37
17	52959686	rs2958927	C	T	Korean		38
17	59485555	rs1057987	C	G, T	Korean	<i>c17orf82, tbx2-as1, tbx2</i>	38
17	78396004	rs41298712	A	G	Korean	<i>endov, mir4730</i>	38
19	7125297	rs1799817	G	A		<i>insr</i>	55
19	7163065	rs2962	G	A	European	<i>insr</i>	56
19	7163140	rs2245655	T	G	European	<i>insr</i>	56

19	7163154	rs2963	G	A	European	<i>insr</i>	56
19	7163214	rs2245649	T	C	European	<i>insr</i>	56
19	7163230	rs2245648	T	C	European	<i>insr</i>	56
19	7165539	rs8107575	C	T	European	<i>insr</i>	56
19	7166109	rs2059807	A	G	Han Chinese, European	<i>insr</i>	34,39,40,49
19	17417565	rs6903	C	T	Korean	<i>mrpl34, abhd8, dda1</i>	37
19	17422402	rs8100448	C	A	Korean	<i>mrpl34, abhd8, dda1</i>	37
19	17426501	rs3826700	T	G	Korean	<i>dda1</i>	37
19	17430757	rs9234	G	A	Korean	<i>ano8, dda1</i>	37
20	15522635	rs7262810	T	C	Korean	<i>macrod2</i>	37
20	31420757	rs853854	A	T	European	<i>mapre1</i>	35
20	52447303	rs6022786	A	G	Han Chinese	<i>sumo1p1</i>	34,40,49
22	20906367	rs1075339	G	A	Korean	<i>plcz1, med15</i>	36

Σκοπός

Σκοπός της παρούσας μελέτης είναι η διερεύνηση πιθανών μονονουκλεοτιδικών πολυμορφισμών που σχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών από δεδομένα που προέκυψαν από τη μελέτη κοορτής ατόμων ευρωπαϊκής καταγωγής της βιοτράπεζας του Ηνωμένου Βασιλείου (UKBB) και ανασκόπηση στη υπάρχουσα βιβλιογραφία και σε δεδομένα βάσεων της πιθανής βιολογικής λειτουργίας των γενετικών τόπων που σχετίζεται με το σύνδρομο.

Γ. Μέθοδος

Για την πραγματοποίηση της ανάλυσης γενετικής συσχέτισης γονιδιωματικής κλίμακας, τα βήματα ελέγχου ποιότητας και η στατιστική ανάλυση, βασίστηκαν στις οδηγίες των Marees et al.^[57] και στις οδηγίες της ιστοσελίδας UK_Biobank_GWAS στο Github ^[58,59].

Συλλογή Δεδομένων

Για την πραγματοποίηση της παρούσας μελέτης, δεδομένα λήφθηκαν από τη βιοτράπεζα του Ηνωμένου Βασιλείου (UK Biobank- UKBB). Η UKBB αποτελεί μία μελέτη κοορτής με σκοπό τη συλλογή γενετικών δεδομένων και συσχέτισης τους με συνήθειες, δημογραφικά στοιχεία αλλά κυρίως με ασθένειες και χαρακτηριστικά που σχετίζονται με αυτές. ^[60] Η λήψη των πληροφοριών αυτών βασίζεται κυρίως στη συμπλήρωση ενός ειδικά σχεδιασμένου για τη μελέτη ερωτηματολογίου από τον συμμετέχοντα ^[61], ενώ για συγκεκριμένες νόσους η κατηγοριοποίηση βασίστηκε σε ιατρική γνωμάτευση από εξειδικευμένο προσωπικό ή επιβεβαίωση από γενετικούς βιομάρτυρες.

Η επιστράτευση των συμμετεχόντων πραγματοποιήθηκε στο χρονικό διάστημα 2006 – 2017 σε διάφορα κέντρα συλλογής της Αγγλίας (Wexham, Swansea, Stoke, Stockport, Sheffield, Reading, Oxford, Nottingham, Newcastle, Middleborough, Manchester, Liverpool, Leeds, Hounslow, Glasgow, Edinburg, Croydon, Cardiff, Bury, Bristol, Birmingham και Bards) και έως τώρα απαρτίζεται από 487,409 συμμετέχοντες. Σύμφωνα με στοιχεία βάσει ιδίας δήλωσης οι συμμετέχοντες ανήκουν σε διάφορες εθνικότητες (Καυκάσια φυλή [94.06%], Μαύρη Φυλή [1.61%], Ασιατική φυλή [2.28%] και μιγάδες [0.59%]) και κοινωνικοοικονομικές τάξεις, ενώ παράλληλα έγινε μια προσπάθεια να γνωστοποιείται στα δεδομένα άτομα που φέρουν έως και τρίτου βαθμού συγγενείας. ^[62]

Όσον αφορά τη γονοτύπηση, αρχικά για τους 50,000 πρώτους συμμετέχοντες πραγματοποιήθηκε με την πλατφόρμα Affymetrix UK BiLEVE Axiom array ενώ για τους υπόλοιπους με την Affymetrix UK Biobank Axiom® array, το οποίο γονοτυπεί περίπου 850,000 πολυμορφισμούς. Στα δεδομένα της γονοτύπησης πραγματοποιήθηκε ένας αρχικός έλεγχος ποιότητας των δεδομένων και ακολούθησε η διαδικασία του πρόβλεψης γονότυπων (imputation), όπου αύξησε τον αριθμό των μελετημένων πολυμορφισμών στα 90 εκατομμύρια από το κέντρο Γενετικής Ανθρώπου Wellcome Trust. ^[62,63]

Ανάλυση Δεδομένων

Αποσκοπώντας στη μελέτη γενετικών τόπων που σχετίζονται με το Σύνδρομο Πολυκυστικών Ωοθηκών, από το σύνολο των δεδομένων που προσφέρει η βιοτράπεζα επιλέχθηκαν μόνο οι συμμετέχοντες γένους θηλυκού με Καυκάσια Ευρωπαϊκή Καταγωγή (194,153 άτομα). Ο φυλετικός διαχωρισμός ήταν απαραίτητος καθώς διαφορές στις συχνότητες αλληλόμορφων ανά πληθυσμό που μετρούνται σε ένα δείγμα μπορούν συχνά να οδηγήσουν σε ψεύτικες συσχετίσεις όταν συμπεριλαμβάνουν άτομα με αρκετά διαφορετικό προγονικό υπόβαθρο.

Λόγω έλλειψης γενετικών βιοδεικτών για το συγκεκριμένο νόσημα η κατάταξη σε πάσχοντες βασίστηκε στην ίδια δήλωση του συμμετέχοντα. Σύμφωνα με την UKBB προέκυψαν 687 γυναίκες που νοσούν από

σύνδρομο πολυκυστικών ωοθηκών [64]. Τα δεδομένα που αφορούσαν το σύνδρομο πολυκυστικών ωοθηκών λήφθηκαν από την πιο πρόσφατη έκδοση του αρχείου UKBB GWAS Imputed v3 – File Manifest Release 20180731 [65].

Στα δεδομένα έχει πραγματοποιηθεί ανάλυση λογιστικής παλινδρόμησης για να διερευνηθεί η συσχέτιση των πολυμορφισμών που σχετίζονται με το σύνδρομο. Η λογιστική παλινδρόμηση είναι μία στατιστική ανάλυση που εξετάζει τη συσχέτιση μεταξύ δύο ποιοτικών μεταβλητών λαμβάνοντας υπόψιν πιθανές συμμεταβλητές, όπως η ηλικία, το είδος της πλατφόρμας και η παρτίδα που χρησιμοποιήθηκε για την αλληλούχηση. Λεπτομέρειες σχετικά με τη στρατηγική της ανάλυσης παρουσιάζονται στη δημοσίευση των Millard et al [66]. Από το PHESANT προκύπτει η τιμή b (beta) που χαρακτηρίζει τη συσχέτιση. Αν το beta φέρει θετική τιμή τότε οι μεταβλητές είναι ανάλογες, δηλαδή η ύπαρξη του αλληλομόρφου θα μπορούσε να προκαλεί την ασθένεια/ φαινότυπο και το αντίθετο υποδηλώνει ότι ο πολυμορφισμός προστατεύει από την ασθένεια/ φαινότυπο. Επίσης όσο η τιμή beta απομακρύνεται από το 0 τόσο πιο έντονη είναι η επίδραση της μεταβλητής του γονότυπου στη ασθένεια/ φαινότυπο. [66,67]

Όσον αφορά τα στάδια ελέγχου ποιότητας έχει προηγηθεί έλεγχος ποιότητας σε επίπεδο δείγματος (μόνο Ευρωπαϊκής καταγωγής γυναίκες, φίλτρο για ανευπλοειδία στα φυλετικά χρωμοσώματα, σύνολο 194.174 δείγματα) και σε επίπεδο πολυμορφισμών ύστερα από πρόβλεψη γονότυπου με τη χρήση αλγορίθμων (imputation) (ισορροπίας HWE $p\text{-value} > 1 \times 10^{-10}$, INFO score > 0.8 , MAF > 0.0001 , σύνολο 13.7 εκατομμύρια πολυμορφισμοί) απευθείας από την UKBB [62,66,68]. Συνεπώς, μένει ο καθορισμός της ελάχιστης επιτρεπτής συχνότητας σπανιότερου αλληλομόρφου. Συνηθίζεται στις έρευνες να υπολογίζεται το μέγεθος του δείγματος που απαιτείται για τη μελέτη ενός χαρακτηριστικού ανάλογα με τον επιπολασμό του στο πληθυσμό. Στις μελέτες γενετικής συσχέτισης γονιδιωματικής κλίμακας μελετούνται ταυτόχρονα πολλά χαρακτηριστικά (πολυμορφισμοί) που ο καθένας φέρει διαφορετική συχνότητα και συνήθως είναι σχετικά χαμηλή, ιδιαίτερα αν σχετίζεται με νόσο όπου ασκείται έντονη φυσική επιλογή. Από την άλλη, λόγω του μεγάλου κόστους των μελετών γενετικής συσχέτισης ο αριθμός των συμμετεχόντων αρχικά είναι μικρός και αυξάνεται με την πάροδο του χρόνου. Συνεπώς στις μελέτες συσχέτισης γίνεται το αντίστροφο, δεν υπολογίζεται το μέγεθος του δείγματος, αλλά σε μια μελέτη λαμβάνονται υπόψιν οι πολυμορφισμοί με συχνότητα άνω του κατώτατου ορίου. Ο τύπος που καθορίζει τη χαμηλότερη τιμή συχνότητας σπανιότερου αλληλομόρφου είναι:

$$MAF = \frac{25}{2 * \text{minor sample size}}$$

Minor sample size είναι η ομάδα (με ή χωρίς το χαρακτηριστικό) που φέρει τα λιγότερα άτομα, και αυτή συνήθως είναι η ομάδα των πασχόντων. Πιο συγκεκριμένα εδώ η ομάδα των πασχουσών γυναικών αποτελείται από 687 άτομα σύμφωνα με τα στοιχεία της UKBB. Συνεπώς από την ανάλυση μας απορρίφθηκαν όλοι οι πολυμορφισμοί με συχνότητα μικρότερη του 0.018.

Σύμφωνα με την τιμή της συχνότητας σπανιότερου αλληλομόρφου (MAF) και την τιμή της στάθμης σημαντικότητας απομονώθηκαν μονονουκλεοτιδικοί πολυμορφισμοί. Τέλος, ορίστηκε το επίπεδο στατιστικής σημαντικότητας κατόπιν διόρθωσης Bonferroni στο $p = 5 \times 10^{-8}$ και ως οριακής σημαντικότητας το $p = 1 \times 10^{-5}$.

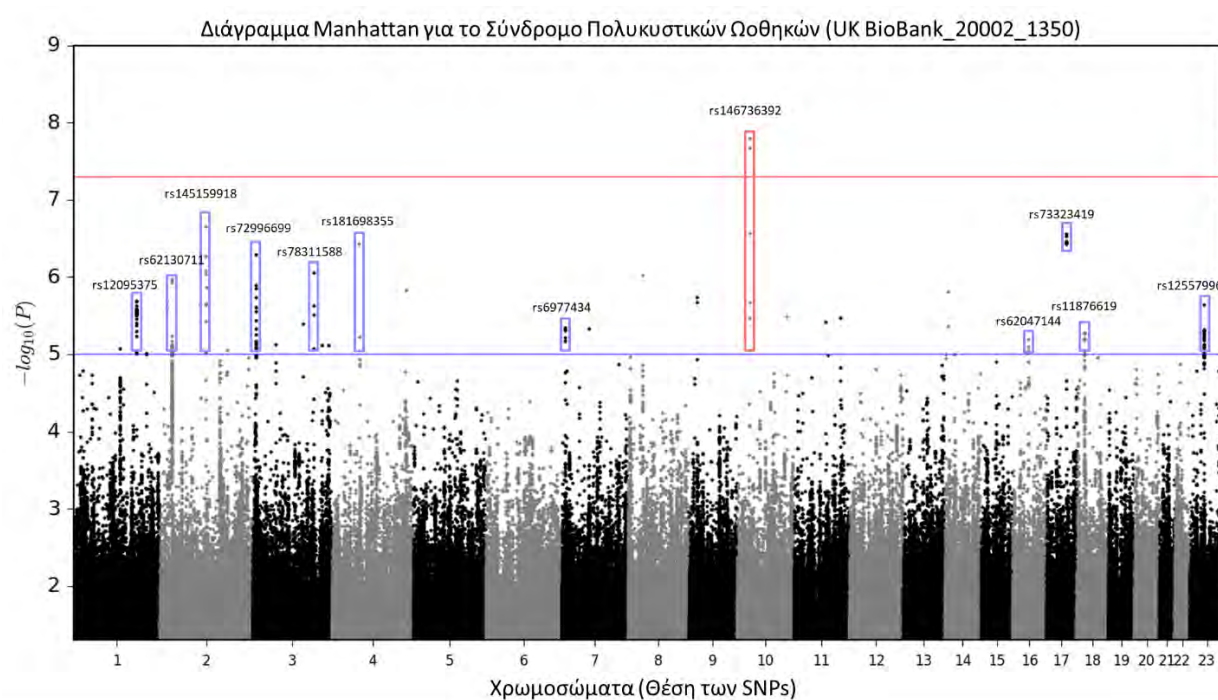
Ανάλυση Γονιδιακής Οντολογίας

Στη συνέχεια διερευνήθηκε ο πιθανός βιολογικός ρόλος των μονονουκλεοτιδικών πολυμορφισμών που προέκυψαν από το προηγούμενο βήμα. Αυτό πραγματοποιήθηκε με τον εντοπισμό της θέσης του κάθε πολυμορφισμού σε κωδικές και μη κωδικές περιοχές του γονιδιώματος μέσω του εργαλείου SNPense [69]. Για τα γονίδια που προέκυψαν πραγματοποιήθηκε ανάλυση εμπλουτισμού για τη διερεύνηση της λειτουργίας τους. Τέλος για κάθε γονίδιο πραγματοποιήθηκε διερεύνηση πιθανού βιολογικού ρόλου μέσω βιβλιογραφικής ανασκόπησης και αναζήτησης από βάσεις δεδομένων [70–72].

Δ. Αποτελέσματα

Προκειμένου να διερευνηθούν πιθανοί γενετικοί τόποι που να σχετίζονται με την εμφάνιση του συνδρόμου πολυκυστικών ωοθηκών αναλύθηκαν δεδομένα της βιοτράπεζας UKBB. Από τα δεδομένα αυτά επιλέχθηκαν μονονουκλεοτιδικοί πολυμορφισμοί που πληρούν τα κριτήρια των ποιοτικών ελέγχων (γυναίκες ευρωπαϊκής καταγωγής με σχέση συγγένειας άνω του 3ου βαθμού, ισορροπίας HWE $p\text{-value} > 1 \cdot 10^{-10}$, INFO score > 0.8 , MAF > 0.018).

Συνολικά υπάρχουν δύο στατιστικά σημαντικοί πολυμορφισμοί ($p < 5 \cdot 10^{-8}$) που βρίσκονται στο χρωμόσωμα 10 (rs146736392 και rs140283997) σε γειτονική γενετική περιοχή. Οι πολυμορφισμοί που έφεραν όριο στατιστικής σημαντικότητας μικρότερο του $1 \cdot 10^{-5}$ (οριακή σημαντικότητα) ανέρχονται στους 171. Όλοι οι γενετικοί πολυμορφισμοί παρουσιάζονται στο διαγραμμα διασποράς Manhattan (εικόνα 2).



Εικόνα 2 Διαγραμμα διασποράς Manhattan. Οι πολυμορφισμοί που έφεραν όριο στατιστικής σημαντικότητας μικρότερο του $1 \cdot 10^{-5}$ (οριακή σημαντικότητα) ανέρχονται στους 171. Όλοι οι γενετικοί πολυμορφισμοί που είναι οριακά σημαντικοί σημαίνονται με γαλάζιο πλαίσιο και όσοι είναι σημαντικοί με κόκκινο, και αναφέρεται το rsID του πολυμορφισμού που φέρει το χαμηλότερο επίπεδο σημαντικότητας για κάθε γενετικό τόπο.

Από την ανάλυση, διακρίνονται 12 γενετικοί τόποι που φαίνεται να σχετίζονται με το σύνδρομο με οριακή σημαντικότητα. Οι πολυμορφισμοί με τη χαμηλότερη τιμή $p\text{-value}$ από κάθε γενετικό τόπο παρουσιάζονται στον **πίνακα 2**. Βλέπουμε ότι τη μεγαλύτερη απόλυτη τιμή β με θετικό πρόσημο έχουν οι rs12095375 (χρωμ. 1, $\beta=0.002670420$ (0.000562721), sig. $2.08 \cdot 10^{-06}$) και rs146736392 (χρωμ. 10, $\beta=0.002789$ (0.000493517), sig. $1.59 \cdot 10^{-08}$) άρα ενδεχομένως να προκαλούν το σύνδρομο, ενώ ο rs62130711 (χρωμ. 2, $\beta=-0.000894$ (0.000183222), sig. $1.07 \cdot 10^{-06}$) δρά πιθανώς προστατευτικά.

Πίνακας 2 Οι πολυμορφισμοί με τη χαμηλότερη τιμή *p-value* από κάθε γενετικό τόπο όπως προέκυψαν από την ανάλυση των δεδομένων της UKBB. Για κάθε πολυμορφισμό στον πίνακα αναφέρεται αναλυτικά η γενετική του θέση στο χρωμόσωμα σύμφωνα με το γονιδίωμα αναφοράς hg19 (GRCh37), ο ξεχωριστός rsID κωδικός, οι δύο μορφές του αλληλομόρφου, η συχνότητα του σπανιότερου αλληλομόρφου στο δείγμα, το σύνολο των ατόμων, η τιμή *beta* από τη λογιστική παλινδρόμηση που χαρακτηρίζει την επίδραση του πολυμορφισμού στο σύνδρομο μαζί με το σφάλμα και τέλος το επίπεδο στατιστικής σημαντικότητας. Με έντονα γράμματα τονίζονται οι πολυμορφισμοί που φέρουν τιμή *beta* με τη μέγιστη απόσταση από το 0 για αυτά τα δεδομένα, με είτε θετική είτε αρνητική τιμή, κατά συνέπεια είναι πιθανά οι πολυμορφισμοί με τη μεγαλύτερη επίδραση στο σύνδρομο πολυκυστικών ωοθηκών.

chr	bp	rsID	a1	a2	minor_allele	minor_AF	beta	se	p-value
1	194120736	rs12095375	G	T	T	0.0186456	0.002670420	0.000562721	2.08E-06
2	27517683	rs62130711	A	G	G	0.2198770	-0.000893722	0.000183222	1.07E-06
2	120562499	rs145159918	C	T	T	0.0279054	0.002412510	0.000465438	2.18E-07
3	5541436	rs72996699	A	G	G	0.0928598	0.001313320	0.0002614	5.06E-07
3	154028751	rs78311588	A	G	G	0.0829344	0.001374120	0.00027946	8.79E-07
4	62634513	rs181698355	G	A	A	0.0303653	0.002332420	0.000458644	3.67E-07
7	5836304	rs6977434	C	A	A	0.0218316	0.002359180	0.000523932	6.71E-06
10	25939389	rs146736392	C	T	T	0.0255750	0.002789020	0.000493517	1.59E-08
16	52999630	rs62047144	C	T	T	0.0282634	0.002077700	0.000460395	6.40E-06
17	55338047	rs73323419	T	C	C	0.0267394	0.002435240	0.000473855	2.76E-07
18	20198700	rs11876619	G	A	A	0.3186430	0.000744142	0.000163321	5.21E-06
X	70166471	rs12557996	C	T	T	0.0369567	0.001908120	0.000403704	2.29E-06

Στη συνέχεια για όλους τους οριακά σημαντικούς πολυμορφισμούς διερευνήθηκε η πιθανή βιολογική λειτουργία τους με το σύνδρομο πολυκυστικών ωοθηκών. Αυτό επιτεύχθηκε, αναζητώντας την τοποθεσία και κατ' επέκταση το ρόλο των SNP σε περιοχές του γονιδιώματος που εκφράζονται. Τα αποτελέσματα παρουσιάζονται στον **πίνακα 3** και την **εικόνα 3**.

Πίνακας 3 Γενετικοί τόποι που συσχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών και τα γονίδια που περιέχουν.

Γεν. Τόπ	Χρωμ.	Έναρξη	Λήξη	Top SNP rsID	Κίνδυνος	Γονίδια
A	1	194120736	194250935	rs12095375	+	EEF1A1P14
B	2	27475318	27653336	rs62130711	-	TRIM54, SLC30A3, DNAJC5G, UCN, MPV17, GTF3C2, NRBPI
Γ	2	120489240	120849018	rs145159918	+	PTPN4, EPB41L5
Δ	3	5541436	5571785	rs72996699	+	-
E	3	153831864	154028751	rs78311588	+	ARHGEF26-AS1, ARHGEF26, DHX36
ΣΤ	4	62634513	63735125	rs181698355	+	ADGRL3
Z	7	5836153	5836304	rs6977434	+	-
H	10	25823261	25957511	rs146736392	+	GPR158, GPN3P1, LINC00836
Θ	16	52998399	53013272	rs62047144	+	-
I	17	55328725	55342824	rs73323419	+	MSI2
ΙΑ	18	20176086	20245248	rs11876619	+	RNU6-1032P
ΙΒ	X	70149188	70210271	rs12557996	+	SLC7A3

Φαίνεται ότι, από τους 12 γενετικούς τόπους οι 9 βρίσκονται σε περιοχές που κωδικοποιούν γονίδια (*TRIM54, SLC30A3, DNAJC5G, UCN, MPV17, GTF3C2, NRBPI, PTPN4, EPB41L5, ARHGEF26, DHX36, ADGRL3, GPR158, MSI2, SLC7A3*) με διάφορες λειτουργίες. Πραγματοποιήθηκε λοιπόν μία ανάλυση εμπλουτισμού μέσω της βάσης Gene Ontology- Panther [72] για να διερευνηθεί ο ρόλος των γονιδίων που φέρουν τους πολυμορφισμούς. Στην **εικόνα 4** παρουσιάζονται τα αποτελέσματα της ανάλυσης εμπλουτισμού των γονιδίων συνολικά και όχι μεμονομένα για κάθε γονίδιο ξεχωριστά. Αναλυτικότερα, βλέπουμε ότι τα γονίδια, που προέκυψαν από το προηγούμενο βήμα, δρουν κυρίως καταλύοντας αντιδράσεις, είναι μεταγραφικοί παράγοντες, συμμετέχουν ως σηματοδοτικά μορια ή φέρουν δραστικότητα μεταφορέα μορίων. Κατ' επέκταση, τα βλέπουμε να συμμετέχουν σε κυτταρικές και μεταβολικές διεργασίες, ενώ συμμετέχουν

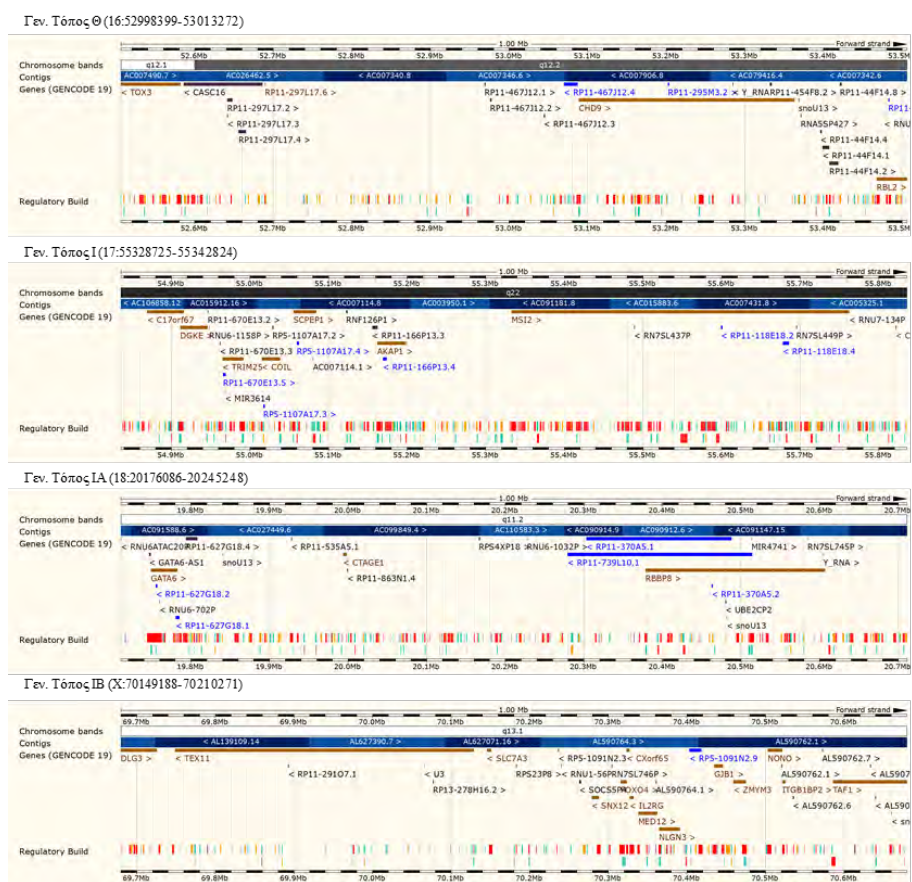
και στην ανταπόκριση σε εξωκυτταρικά ερεθίσματα. Αναλυτικά τα αποτελέσματα της ανάλυσης για κάθε γονίδιο παρουσιάζονται στον **πίνακα 4** όπου αναφέρετε η οικογένεια που ανήκει κάθε γονίδιο, η βιολογική διαδικασία στην οποία συμμετέχει, η μοριακή λειτουργία και το συστατικό του κυττάρου είναι, καθώς και τα μονοπάτια με τα οποία αλληλεπιδρά.

Από τα αποτελέσματα προκύπτει ότι δύο γονίδια εμπλέκονται σε ανοσοαποκρίσεις σε μοριακό επίπεδο (*PTPN4*, *DHX36*) και σε διεργασίες στις συνάψεις του νευρικού συστήματος (*EPB4115*). Τέσσερα γονίδια παίζουν ρόλο στον μεταβολισμό όπου συμμετέχουν σε μεταφορές στοιχείων (*SLC7A3* και *SLC30A3*), τον μεταβολισμό των πρωτεϊνών (*MPV17*, *SLC7A3*) και τον κύκλο του cAMP (*ADGRLE*). Επίσης, το γονίδιο *ARHGEF26* συμμετέχει στη σηματοδότηση RHO που ρυθμίζει την κυττατική προσκόλληση, τον πολλαπλασιασμό και τη διαίρεση σε διάφορα ερεθίσματα όπως στη μηχανική καταπόνηση. Το γονίδιο *UCN* εμπλέκεται στο μεταβολισμό πεπτιδίων-ορμονών μέσω σηματοδότησης με υποδοχείς συζευγμένους με πρωτεΐνη G και τη ρύθμιση της όρεξης. Τέλος τα γονίδια *MSI2* και *TRIM54* δεν εμπεριέχονται στη βάση δεδομένων PANTHER. Ωστόσο στο NCBI αναφέρετε ότι το γονίδιο *MSI2* είναι μεταγραφικός παράγοντας που στοχεύει γονίδια σχετικά με την ανάπτυξη και τη ρύθμιση του κυτταρικού κύκλου και σχετίζεται με διάφορους τύπους καρκίνου. Το *TRIM54* κωδικοποιεί για μια πρωτεΐνη με μοτίβα δακτυλίου που δημιουργεί ετεροδιμερή με τις όμοιες πρωτεΐνες RNF28/MURF1 και RNF29/MURF2 και εμπλέκονται στη ρύθμιση της κίνησης τιτίνης και σε μονοπάτια που εξαρτώνται από τους μικροσωληνίσκους του κυτταροσκελετού στους μύες.

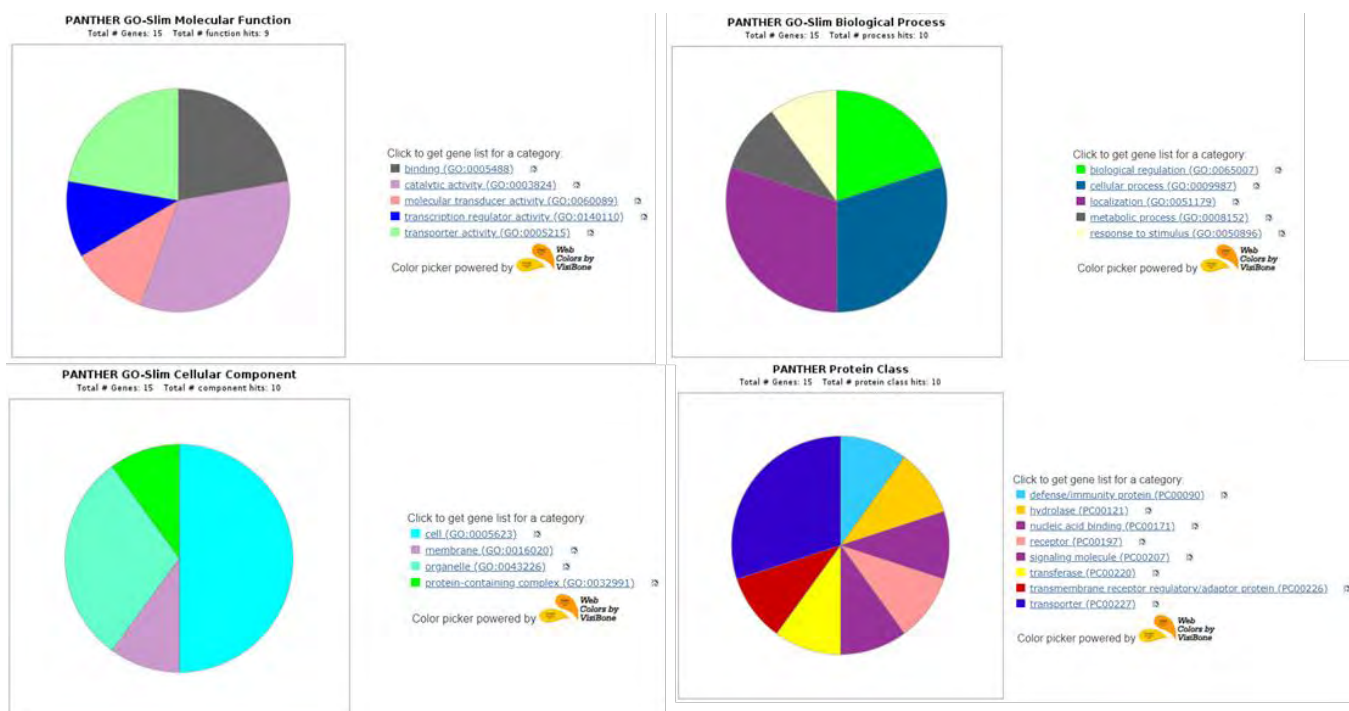
Παράλληλα, πραγματοποιήθηκε ανάλυση για την ακριβή τοποθεσία των SNPs στο γονιδίωμα και τι μεταλλαγές μπορεί να δημιουργεί και σε ποια περιοχή των γονιδίων, χρησιμοποιώντας το SNPENSE [69]. Τα αποτελέσματα της ανάλυσης για όλους τους μονονουκλεοτιδικούς πολυμορφισμούς παρουσιάζονται στην **εικόνα 5**. Φαίνεται ότι οι μεταλλαγές που δημιουργούνται από τους οριακά σημαντικούς πολυμορφισμούς δεν προκαλούν δομικές αλλαγές σε κωδικά προϊόντα ή μεταβολή στην έκφραση τους.



Εικόνα 3 Γενετικοί τόποι που συσχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών όπως αναπαρίστανται στη βάση δεδομένων Ensembl [71]. Στην εικόνα φαίνονται γονίδια και μη κωδικά γονίδια τις κάθε περιοχής.



Εικόνα 3 (συνέχεια) Γενετικοί τόποι που συσχετίζονται με το σύνδρομο πολυκυστικών ωοθηκών όπως αναπαρίστανται στη βάση δεδομένων Ensembl [71]. Στην εικόνα φαίνονται γονίδια και μη κωδικά γονίδια τις κάθε περιοχής.



Εικόνα 4 Συνολικά αποτελέσματα της ανάλυσης εμπλουτισμού των γονιδίων που προέκυψαν από την ανάλυση γενετικής συσχέτισης με το σύνδρομο πολυκυστικών ωοθηκών. Παρουσιάζονται συνολικά οι βιολογικές διεργασίες, η μοριακή λειτουργία, τι μέρος/ συστατικό του κυττάρου και η τάξη των πρωτεϊνών σύμφωνα με τα δεδομένα της βάσης PANTHER [72].

Πίνακας 4 Αποτελέσματα ανάλυσης εμπλουτισμού για τα γονίδια που βρίσκονται στους 12 γενετικούς τόπου που προέκυψαν από την ανάλυση γενετικής συσχέτισης για το σύνδρομο πολυκυστικών ωοθηκών [72].

ID Γονιδίου	Γονίδιο	Πλήρες Όνομα	Οικογένεια/Υποοικογένεια κατά PANTHER	Τάξη Πρωτεϊνών	Βιολογική Διαδικασία	Μοριακή Λειτουργία	Κυτταρικό Συστατικό	Μονοπάτι	Μονοπάτια Reactome
HUMAN HGNC=9656 UniProtKB=P29074	PTPN4	Tyrosine-protein phosphatase non-receptor type 4	TYROSINE-PROTEIN PHOSPHATASE NON-RECEPTOR TYPE 4 (PTHR45706:SF7)					-	Immune System;Interleukin-1 family signaling;Toll-Like Receptors Cascades;Interleukin-37 signaling;Signaling by Interleukins;Innate Immune System;Toll Like Receptor 4 (TLR4) Cascade;Cytokine Signaling in Immune system
HUMAN HGNC=20974 UniProtKB=Q9HAR2	ADGRL3	Adhesion G protein-coupled receptor L3	ADHESION G PROTEIN-COUPLED RECEPTOR L3 (PTHR12011:SF60)	G-protein coupled receptor(PC00021);antibacterial response rotein (PC00051);protease(PC00190)	activation of adenylate cyclase activity(GO:0007190);adenylate cyclase-activating G-protein coupled receptor signaling pathway(GO:0007189);regulation of adenylate cyclase activity(GO:0045761);regulation of cAMP-mediated signaling(GO:0043949)	G-protein coupled receptor activity(GO:0004930);adenylate cyclase activity(GO:0004016)	integral component of plasma membrane(GO:0005887);intracellular(GO:0005622)	-	
HUMAN HGNC=7224 UniProtKB=P39210	MPV17	Protein Mpv17	PROTEIN MPV17 (PTHR11266:SF17)	transmembrane receptor regulatory/adaptor protein(PC00226)			cytoplasm(GO:0005737)	-	Peroxisomal protein import;Metabolism of proteins
HUMAN HGNC=23689 UniProtKB=Q5T848	GPR158	Probable G-protein coupled receptor 158	G-PROTEIN COUPLED RECEPTOR 158-RELATED (PTHR32546:SF11)	transporter(PC00227)				-	
HUMAN HGNC=12516 UniProtKB=P55089	UCN	Urocortin	UROCORTIN (PTHR15035:SF11)	peptide hormone(PC00179)				-	Signaling by GPCR;Signal Transduction;Synthesis, secretion, and deacylation of Ghrelin;Class B/2 (Secretin family receptors);Peptide hormone metabolism;GPCR ligand binding;Metabolism of proteins
HUMAN HGNC=11014 UniProtKB=Q99726	SLC30A3	Zinc transporter 3	ZINC TRANSPORTER 3 (PTHR11562:SF30)	transporter(PC00227)	cellular divalent inorganic cation homeostasis(GO:0072503);cellular metal ion homeostasis(GO:0006875);divalent metal ion transport(GO:0070838);inorganic cation transmembrane transport(GO:0098662);maintenance of location(GO:0051235);regulation of cellular process(GO:0050794);regulation of localization(GO:0032879);response to metal ion(GO:0010038);transition metal ion homeostasis(GO:0055076);transition metal ion transport(GO:0000041)	metal ion transmembrane transporter activity(GO:0046873)	plasma membrane(GO:0005886)	-	Transport of bile salts and organic acids, metal ions and amine compounds;Metal ion SLC transporters;Zinc efflux and compartmentalization by the SLC30 family;SLC-mediated transmembrane transport;Zinc transporters;Transport of small molecules
HUMAN HGNC=7993 UniProtKB=Q9UHY1	NRBP1	Nuclear receptor-binding protein	NUCLEAR RECEPTOR-BINDING PROTEIN (PTHR13902:SF48)	non-receptor serine/threonine protein kinase(PC00167)	ER to Golgi vesicle-mediated transport(GO:0006888);intracellular signal transduction(GO:0035556)	protein serine/threonine kinase activity(GO:0004674)	endomembrane system(GO:0012505);plasma membrane(GO:0005886);vacuole(GO:0005779)	-	Nuclear Receptor transcription pathway;RNA Polymerase II Transcription;Generic Transcription Pathway;Gene expression (Transcription)
HUMAN HGNC=19819 UniProtKB=Q9HGM4	EPB41L5	Band 4.1-like protein 5	BAND 4.1-LIKE PROTEIN 5 (PTHR23280:SF15)		actomyosin structure organization(GO:0031032)			-	Protein-protein interactions at synapses;Neuronal System;Neurexins and neuroligins
HUMAN HGNC=16008 UniProtKB=Q9BYV2	TRIM54	Tripartite motif-containing protein 54	TRIPARTITE MOTIF-CONTAINING PROTEIN 54 (PTHR24103:SF596)					-	
HUMAN HGNC=11061 UniProtKB=Q8WY07	SLC7A3	Cationic amino acid transporter 3	CATIONIC AMINO ACID TRANSPORTER 3 (PTHR43243:SF20)	amino acid transporter(PC00046)	import into cell(GO:0098657)	L-lysine transmembrane transporter activity(GO:0015189);arginine transmembrane transporter activity(GO:0015181)		-	Transport of inorganic cations/anions and amino acids/oligopeptides;Amino acid transport across the plasma membrane;SLC-mediated transmembrane transport;Transport of small molecules
HUMAN HGNC=4665 UniProtKB=Q8WUA4	GTF3C2	General transcription factor 3C polypeptide 2	GENERAL TRANSCRIPTION FACTOR 3C POLYPEPTIDE 2 (PTHR15052:SF2)		transcription by RNA polymerase III(GO:0006383)	transcription regulator activity(GO:0140110)	transcription factor TFIIC complex(GO:0000127)	-	RNA Polymerase III Transcription Initiation From Type 2 Promoter;RNA Polymerase III Transcription Initiation From Type 1 Promoter;RNA Polymerase III Transcription Initiation;RNA Polymerase III Abortive And Retraction Initiation;Gene expression (Transcription);RNA Polymerase III Transcription
HUMAN HGNC=24844 UniProtKB=Q8N7S2	DNAJC5G	DnaJ homolog subfamily C member 5G	DNAJ HOMOLOG SUBFAMILY C MEMBER 5G (PTHR44027:SF2)					-	
HUMAN HGNC=24490 UniProtKB=Q96DR7	ARHGEF26	Rho guanine nucleotide exchange factor 26	RHO GUANINE NUCLEOTIDE EXCHANGE FACTOR 26 (PTHR12845:SF4)					-	Signal Transduction;G alpha (12/13) signalling events;Signaling by Rho GTPases;Rho GTPase cycle;Death Receptor Signalling;Cell death signalling via NRAGE, NRIF and NADK;p75 NTR receptor-mediated signalling;Signaling by GPCR;NRAGE signals death through JNK;GPCR downstream signalling
HUMAN HGNC=14410 UniProtKB=Q9H2U1	DHX36	ATP-dependent RNA helicase DHX36	ATP-DEPENDENT RNA HELICASE DHX36 (PTHR18934:SF213)	RNA helicase(PC00032)		ATP-dependent helicase activity(GO:0008026);RNA binding(GO:0003723);catalytic activity, acting on RNA(GO:0140098)	nucleus(GO:0005634)	-	Cytosolic sensors of pathogen-associated DNA ;Immune System;DEX/H-box helicases activate type I IFN and inflammatory cytokines production ;Innate Immune System
HUMAN HGNC=18585 UniProtKB=Q96DH6	MSI2	RNA-binding protein Musashi homolog 2	RNA-BINDING PROTEIN MUSASHI HOMOLOG 2 (PTHR15241:SF82)			RNA binding(GO:0003723)	cytoplasm(GO:0005737);nucleolus(GO:0005730);polyosome(GO:0005844)	-	

Ε. Συμπεράσματα

Το σύνδρομο πολυκυστικών ωοθηκών είναι μια ενδοκρινολογική διαταραχή ανδρογόνων ορμονών που πλήττει γυναίκες προκαλώντας υπερανδρογοναιμία, διαταραχές στον εμμηνορροϊκό κύκλο, αντίσταση στην ινσουλίνη αυξάνοντας την πιθανότητα εμφάνιση διαβήτη τύπου II και τον κίνδυνο εμφάνισης κακοηθειών. Έχει απασχολήσει όμως ιδιαίτερα την επιστημονική κοινότητα καθώς ευθύνεται σε μεγάλο βαθμό για την υπογονιμότητα των γυναικών. Φαίνεται ότι το σύνδρομο φέρει γενετικά και περιβαλλοντικά αίτια. Έχουν πραγματοποιηθεί πολλές έρευνες στοχευμένες και γονιδιωματικής κλίμακας για να συσχετισθούν γενετικές και επιγενετικές τροποποιήσεις με τον κίνδυνο εμφάνισης του συνδρόμου.

Στη παρούσα εργασία πραγματοποιήθηκε μια ανάλυση γενετικής συσχέτισης γονιδιωματικής κλίμακας σε δεδομένα της βιοτράπεζας του Ηνωμένου Βασιλείου για να βρεθούν γενετικοί τόποι που σχετίζονται με την εμφάνιση του συνδρόμου πολυκυστικών ωοθηκών σε γυναίκες με ευρωπαϊκή καταγωγή. Η μελέτη πραγματοποιήθηκε σε δείγμα 194,000 περίπου γυναικών με 678 να πάσχουν από το σύνδρομο σύμφωνα με δική τους δήλωση, και ανίχνευε 850,000 πολυμορφισμούς όπου αυξάνονταν στα 13.7 εκατομμύρια ύστερα από πρόβλεψη πιθανών γονότυπων (imputation) και ποιοτικό έλεγχο των δεδομένων.

Από τη μελέτη προέκυψαν δύο στατιστικά σημαντικοί πολυμορφισμοί στον γενετικό τόπο 10p12.1, και 171 οριακά σημαντικοί μονονουκλεοτιδικοί πολυμορφισμοί σε 12 γενετικούς τόπους (πίνακας 3). Οι έντεκα γενετικοί τόποι φαίνεται να σχετίζονται με τον κίνδυνο εμφάνισης του συνδρόμου ενώ ο ένας στο χρωμόσωμα 2 (2p23.3) φαίνεται να σχετίζεται με προστασία έναντι του συνδρόμου. Οι δώδεκα τόποι αναφέρονται για πρώτη φορά και δεν γειτονεύουν με γενετικούς τόπους που έχουν αναφερθεί σε προηγούμενες αναλύσεις γενετικής συσχέτισης (πίνακας 1).

Κατ'επέκταση, διερευνήθηκαν τα γενετικά στοιχεία των τόπων αυτών. Από την αναζήτηση προέκυψε ότι οι 9 γενετικοί τόποι φέρουν τα γονίδια *trim54*, *slc30a3*, *dnajc5g*, *ucn*, *mpv17*, *gtf3c2*, *nrbp1*, *ptpn4*, *epb41l5*, *arhgef26*, *dhx36*, *adgrl3*, *gpr158*, *msi2* και *slc7a3*, που εμπλέκονται με ανοσοαποκρίσεις, με διεργασίες του μεταβολισμού (μεταφορές στοιχείων, μεταβολισμός των πρωτεϊνών, μεταβολισμός πεπτιδίων ορμονών και κύκλος του cAMP) με ρύθμιση της κυτταρικής προσκόλλησης, του πολλαπλασιασμού και της διαίρεση σε διάφορα ερεθίσματα. Παράλληλα, φαίνεται ότι οι πολυμορφισμοί που προέκυψαν από τη γενετική συσχέτιση δεν προκαλούν αλλαγές που να επηρεάζουν άμεσα το εκφραζόμενο προϊόν, είτε σε δομικό επίπεδο είτε σε επίπεδο έκφρασης (εικόνα 5). Ύστερα από μια σύντομη βιβλιογραφική ανασκόπηση στο NCBI χρησιμοποιώντας ως λέξεις κλειδιά τα γονίδια και το σύνδρομο πολυκυστικών ωοθηκών βρέθηκε μία δημοσίευση για την ουροκορτίνη (UCN), που σχετίζεται με την ελεύθερη τεστροστερόνη στο αίμα γυναικών με σύνδρομο πολυκυστικών ωοθηκών [73].

Τα δεδομένα από τη παρούσα μελέτη και από άλλες μελέτες γενετικής συσχέτισης θα μπορούσαν να κατευθύνουν τους επιστήμονες σε γενετικούς τόπους που ευθύνονται για την εμφάνιση του συνδρόμου και παράλληλα να συμβάλλουν στην ανάπτυξη ενός μοντέλου κινδύνου, διότι το σύνδρομο φαίνεται να επιρρεάζεται έντονα από περιβαλλοντικούς παράγοντες [74], προκειμένου να υπάρξουν στοχευμένες

στρατηγικές παρέμβασης σε γυναίκες που διατρέχουν αυξημένο κίνδυνο. Ωστόσο, για να είναι πιο αξιόπιστα τα αποτελέσματα, οι μελέτες θα πρέπει να διευρυνθούν, χρησιμοποιώντας περισσότερους πληθυσμούς με συμμετοχή περισσότερων πάσχουσων γυναικών και κυρίως κατηγοριοποίηση των συμμετεχόντων ανάλογα τον φαινότυπο σύμφωνα με τα κριτήρια του Rotterdam 2003. Επίσης σύμφωνα με μία νεα έρευνα σε μοντέλο μυών φαίνεται ότι και η κατάσταση της μητέρας κατά την περίοδο της κύησης παίζει καθοριστικό ρόλο στην ανάπτυξη του αναπαραγωγικού συστήματος του θηλυκού απόγονου [75] συνεπώς ενδιαφέρον θα αποτελούσε να πραγματοποιηθεί μία μελέτη γενετικής συσχέτισης που θα εστιάζει στη γενετική και επιγενετική σχέση γονέων με απόγονους.

							3' prime UTR variant	5' prime UTR variant	NMD transcript variant	intron variant	non coding transcript exon variant	non coding transcript variant	splice region variant
id	chr	start	end	strand	gene ids	gene names							
rs12564097	1	151566817	151566817	+	ENSG00000143367	TUFT1					1	2	4
rs12095375	1	194151606	194151606	+	ENSG00000227240	AL356272.1					1	1	1
rs12093379	1	194153035	194153035	+	ENSG00000227240	AL356272.1					1	1	1
rs12095985	1	194154209	194154209	+	ENSG00000227240	AL356272.1					1	1	1
rs10921504	1	194157623	194157623	+	ENSG00000227240	AL356272.1					2	2	2
rs755945584	1	194158628	194158628	+	ENSG00000227240	AL356272.1					1	1	1
rs59157004	1	194161033	194161033	+	ENSG00000227240	AL356272.1					2	2	2
rs12057425	1	194161518	194161518	+	ENSG00000227240	AL356272.1					1	1	1
rs12066411	1	194170002	194170002	+	ENSG00000227240	AL356272.1					1	1	1
rs10921506	1	194170839	194170839	+	ENSG00000227240	AL356272.1					1	1	1
rs76675176	1	194171825	194171825	+	ENSG00000227240	AL356272.1					1	1	1
rs12082464	1	194184254	194184254	+	ENSG00000227240	AL356272.1					1	1	1
rs16836486	1	194188895	194188895	+	ENSG00000227240	AL356272.1					1	1	1
rs139622816	1	194190703	194190703	+	ENSG00000227240	AL356272.1					1	1	1
rs149734501	1	194190741	194190741	+	ENSG00000227240	AL356272.1					2	2	2
rs12090496	1	194194919	194194919	+	ENSG00000227240	AL356272.1					1	1	1
rs77088952	1	194198538	194198538	+	ENSG00000227240	AL356272.1					1	1	1
rs12075017													
rs10921513													
rs60096291													
rs79055547													
rs148002941													
rs9662046													
rs78503800													
rs74471373													
rs75495973													
rs12057943													
rs12066981													
rs12065894													
rs9661433													
rs12072567													
rs57353113													
rs10921523													
rs7515513													
rs79981534													
rs113228185													
rs79513293													
rs12080723													
rs16836610													
rs10921526													
rs74738470													
rs76005189													
rs1397176													
rs189303289													
rs2083363													
rs11126932													
rs11126934	2	27253745	27253745	+	ENSG00000115194	SLC30A3					2	2	2
rs11126936	2	27260233	27260233	+	ENSG00000115194	SLC30A3					2	2	2
rs60631624	2	27274947	27274947	+	ENSG00000115194	SLC30A3					2	2	2
rs62130707	2	27276169	27276169	+	ENSG00000163793	DNAJC5G	2				2	2	2
rs4665961	2	27277513	27277513	+	ENSG00000163793	DNAJC5G					2	2	2

id	chr	start	end	strand	gene ids	gene names	<div> <div> <div>5 prime UTR variant</div> <div>3 prime UTR variant</div> <div>intron variant</div> <div>non coding transcript variant</div> <div>non coding transcript variant</div> </div> <div> <div>5 prime UTR variant</div> <div>3 prime UTR variant</div> <div>intron variant</div> <div>non coding transcript variant</div> <div>non coding transcript variant</div> </div> </div>			
rs78668399										
rs606211	11	78931241	78931241	+	ENSG00000149256	TENM4				
rs45627831	11	113327963	113327963	+	ENSG00000149292	TTC12				
rs74549758	14	28826043	28826043	+	ENSG00000257748	LINC02281				
rs75105335										
rs62047143										
rs62047144										
rs9926732										
rs9944327										
rs9944333										
rs9916061										
rs7214294										
rs9890932										
rs7215360										
rs9905322										
rs3744086	17	57257290	57257290	+	ENSG00000153944	MSI2				
rs73323419	17	57260686	57260686	+	ENSG00000153944	MSI2				
rs12452040	17	57264741	57264741	+	ENSG00000153944	MSI2				
rs9902023	17	57265463	57265463	+	ENSG00000153944	MSI2				
rs8087945										
rs8087799										
rs11876619										
rs35561761										
rs199780392										
rs7236935										
rs12955536										
rs112307349										
rs35945038										
rs72628883	X	70929338	70929338	+	ENSG00000165349	SLC7A3				
rs5936608										
rs4844280										
rs12556893										
rs5981032										
rs5937027										
rs11094185										
rs4436293										
rs5936609										
rs4272540										
rs5007414										
rs112420976										
rs200379683										
rs370927748										
rs201782252										
rs12560098										
rs12560144										
rs5937030										
rs12557996										
rs72628888										
rs150325574										

Εικόνα 5 Μεταλλάξεις που προκαλούνται από τα SNPs που σχετίζονται με το PCOS, σε κωδικές και μη κωδικές περιοχές του γονιδιώματος, χρησιμοποιώντας το SNPense. Οι μεταλλάξεις που δημιουργούνται από τους πολυμορφισμούς δεν φαίνεται να προκαλούν δομικές αλλαγές σε κωδικά προϊόντα ή μεταβολή στην έκφραση τους [69].

Στ. Αναφορές

1. Lizneva, D. *et al.* Criteria, prevalence, and phenotypes of polycystic ovary syndrome. *Fertil. Steril.* **106**, 6–15 (2016).
2. Stein, I. F. Bilateral polycystic ovaries: Significance in sterility. *Am. J. Obstet. Gynecol.* **50**, 385–398 (1945).
3. Azziz, R. *et al.* Polycystic ovary syndrome. *Nat. Rev. Dis. Prim.* **2**, 16057 (2016).
4. ESHRE, T. R., Group, A.-S. P. C. W. & others. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertil. Steril.* **81**, 19–25 (2004).
5. Azziz, R. *et al.* The Androgen Excess and PCOS Society criteria for the polycystic ovary syndrome: the complete task force report. *Fertil. Steril.* **91**, 456–488 (2009).
6. Dunaif, A., Segal, K. R., Futterweit, W. & Dobrjansky, A. Profound Peripheral Insulin Resistance, Independent of Obesity, in Polycystic Ovary Syndrome. *Diabetes* **38**, 1165–1174 (1989).
7. Moran, L. J., Misso, M. L., Wild, R. A. & Norman, R. J. Impaired glucose tolerance, type 2 diabetes and metabolic syndrome in polycystic ovary syndrome: a systematic review and meta-analysis. *Hum. Reprod. Update* **16**, 347–363 (2010).
8. Legro, R. S., Gnatuk, C. L., Kunselman, A. R. & Dunaif, A. Changes in Glucose Tolerance over Time in Women with Polycystic Ovary Syndrome: A Controlled Study. *J. Clin. Endocrinol. Metab.* **90**, 3236–3242 (2005).
9. Norman, R. J., Masters, L., Milner, C. R., Wang, J. X. & Davies, M. J. Relative risk of conversion from normoglycaemia to impaired glucose tolerance or non-insulin dependent diabetes mellitus in polycystic ovarian syndrome. *Hum. Reprod.* **16**, 1995–1998 (2001).
10. Hudecova, M. *et al.* Diabetes and impaired glucose tolerance in patients with polycystic ovary syndrome—a long term follow-up. *Hum. Reprod.* **26**, 1462–1468 (2011).
11. Gambineri, A. *et al.* Polycystic Ovary Syndrome Is a Risk Factor for Type 2 Diabetes. *Diabetes* **61**, 2369–2374 (2012).
12. Schmidt, J., Landin-Wilhelmsen, K., Brännström, M. & Dahlgren, E. Cardiovascular Disease and Risk Factors in PCOS Women of Postmenopausal Age: A 21-Year Controlled Follow-Up Study. *J. Clin. Endocrinol. Metab.* **96**, 3794–3803 (2011).
13. Ezech, U., Yildiz, B. O. & Azziz, R. Referral Bias in Defining the Phenotype and Prevalence of Obesity in Polycystic Ovary Syndrome. *J. Clin. Endocrinol. Metab.* **98**, E1088–E1096 (2013).
14. Luque-Ramírez, M. *et al.* Referral bias in female functional hyperandrogenism and polycystic ovary syndrome. *Eur. J. Endocrinol.* **173**, (2015).
15. Joham, A. E., Teede, H. J., Ranasinha, S., Zoungas, S. & Boyle, J. Prevalence of infertility and use of fertility treatment in women with polycystic ovary syndrome: data from a large community-based cohort study. *J. women's Heal.* **24**, 299–307 (2015).
16. Roos, N. *et al.* Risk of adverse pregnancy outcomes in women with polycystic ovary syndrome: population based cohort study. *BMJ* **343**, (2011).
17. Cinar, N. *et al.* Depression, anxiety and cardiometabolic risk in polycystic ovary syndrome. *Hum. Reprod.* **26**, 3339–3345 (2011).
18. Teede, H. J. *et al.* Assessment and management of polycystic ovary syndrome: Summary of an evidence-based guideline. *Med. J. Aust.* **195**, S65–S112 (2011).
19. Dokras, A., Clifton, S., Futterweit, W. & Wild, R. Increased risk for abnormal depression scores in women with polycystic ovary syndrome: a systematic review and meta-analysis. *Obstet. Gynecol.* **117**, 145–152 (2011).
20. Veltman-Verhulst, S. M., Boivin, J., Eijkemans, M. J. C. & Fauser, B. J. C. M. Emotional distress is a common risk in women with polycystic ovary syndrome: a systematic review and meta-analysis of 28 studies. *Hum. Reprod. Update* **18**, 638–651 (2012).
21. Dumesic, D. A. & Lobo, R. A. Cancer risk and PCOS. *Steroids* **78**, 782–785 (2013).
22. Banaszewska, B., Spaczynski, R. Z., Pelesz, M. & Pawelczyk, L. Incidence of elevated LH/FSH ratio in polycystic ovary syndrome women with normo- and hyperinsulinemia. *Rocz Akad Med Białymst* **48**, 131–134 (2003).
23. Dale, P. O., Tanbo, T., Vaaler, S. & Åbyholm, T. Body weight, hyperinsulinemia, and gonadotropin levels in the polycystic ovarian syndrome: evidence of two distinct populations. *Fertil. Steril.* **58**, 487–491 (1992).
24. Burt Solorzano, C. M., McCartney, C. R., Blank, S. K., Knudsen, K. L. & Marshall, J. C. Hyperandrogenaemia in adolescent girls: origins of abnormal gonadotropin-releasing hormone secretion. *BJOG An Int. J. Obstet. Gynaecol.* **117**, 143–149 (2010).
25. Hayes, M. G. *et al.* Genome-wide association of polycystic ovary syndrome implicates alterations in gonadotropin secretion in European ancestry populations. *Nat. Commun.* **6**, 1–12 (2015).
26. Corbould, A., Zhao, H., Mirzoeva, S., Aird, F. & Dunaif, A. Enhanced Mitogenic Signaling in Skeletal Muscle of Women With Polycystic Ovary Syndrome. *Diabetes* **55**, 751–759 (2006).
27. Rajkhowa, M. *et al.* Insulin resistance in polycystic ovary syndrome is associated with defective regulation of ERK1/2 by insulin in skeletal muscle in vivo. *Biochem. J.* **418**, 665–671 (2009).
28. Chazenbalk, G. *et al.* Regulation of Adiponectin Secretion by Adipocytes in the Polycystic Ovary Syndrome: Role of Tumor Necrosis Factor- α . *J. Clin. Endocrinol. Metab.* **95**, 935–942 (2010).
29. Mannerås-Holm, L. *et al.* Adipose Tissue Has Aberrant Morphology and Function in PCOS: Enlarged Adipocytes and Low Serum Adiponectin, But Not Circulating Sex Steroids, Are Strongly Associated with Insulin Resistance. *J. Clin. Endocrinol. Metab.* **96**, E304–E311 (2011).
30. Kosova, G. & Urbanek, M. Genetics of the polycystic ovary syndrome. *Mol. Cell. Endocrinol.* **373**, 29–38 (2013).
31. Hardy, T. & Norman, R. J. Chapter 20 - Genetics of Polycystic Ovary Syndrome. in *Human Reproductive and Prenatal Genetics* (eds. Leung, P. C. K. & Qiao, J.) 447–461 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-813570-9.00020-6
32. Day, F. R. *et al.* Causal mechanisms and balancing selection inferred from genetic associations with polycystic ovary syndrome. *Nat. Commun.* **6**, 8464 (2015).
33. Chen, Z. J. *et al.* Genome-wide association study identifies susceptibility loci for polycystic ovary syndrome on chromosome

- 2p16.3, 2p21 and 9q33.3. *Nat. Genet.* **43**, 55–59 (2011).
34. Shi, Y. *et al.* Genome-wide association study identifies eight new risk loci for polycystic ovary syndrome. *Nat. Genet.* **44**, 1020–1025 (2012).
 35. Day, F. *et al.* Large-scale genome-wide meta-analysis of polycystic ovary syndrome suggests shared genetic architecture for different diagnosis criteria. *PLoS Genet.* **14**, 1–20 (2018).
 36. Shim, U. *et al.* Pathway Analysis Based on a Genome-Wide Association Study of Polycystic Ovary Syndrome. *PLoS One* **10**, 1–11 (2015).
 37. Kim, S. H., Liu, M., Jin, H. S. & Park, S. High genetic risk scores of ASIC2, MACROD2, CHRM3, and C2orf83 genetic variants associated with polycystic ovary syndrome impair insulin sensitivity and interact with energy intake in Korean Women. *Gynecol. Obstet. Invest.* **84**, 225–236 (2019).
 38. Lee, H. *et al.* Genome-wide association study identified new susceptibility loci for polycystic ovary syndrome. *Hum. Reprod.* **30**, 723–731 (2015).
 39. Brower, M. A. *et al.* Further investigation in europeans of susceptibility variants for polycystic ovary syndrome discovered in genomewide association studies of chinese individuals. *J. Clin. Endocrinol. Metab.* **100**, E182–E186 (2015).
 40. Xu, Y. *et al.* Systematic evaluation of genetic variants for polycystic ovary syndrome in a Chinese population. *PLoS One* **10**, 1–10 (2015).
 41. Capalbo, A. *et al.* The 312N variant of the luteinizing hormone/choriogonadotropin receptor gene (LHCGR) confers up to 2·7-fold increased risk of polycystic ovary syndrome in a Sardinian population. *Clin. Endocrinol. (Oxf)*. **77**, 113–119 (2012).
 42. Piersma, D. D., Verhoef-Post, M. M., Berns, E. P. & Themmen, A. A. P. N. LH receptor gene mutations and polymorphisms: an overview. *Mol. Cell. Endocrinol.* **260**, 282–286 (2007).
 43. Unsal, T. *et al.* Genetic polymorphisms of FSHR, CYP17, CYP11A1, CAPN10, INSR, SERPINE1 genes in adolescent girls with polycystic ovary syndrome. *J. Assist. Reprod. Genet.* **26**, 205–216 (2009).
 44. Dolfín, E. *et al.* FSH-receptor Ala307Thr polymorphism is associated to polycystic ovary syndrome and to a higher responsiveness to exogenous FSH in Italian women. *J. Assist. Reprod. Genet.* **28**, 925–930 (2011).
 45. Gu, B.-H., Park, J.-M. & Baek, K.-H. Genetic variations of follicle stimulating hormone receptor are associated with polycystic ovary syndrome. *Int. J. Mol. Med.* **26**, 107–112 (2010).
 46. Wu, X. *et al.* Association between FSHR polymorphisms and polycystic ovary syndrome among Chinese women in north China. *J. Assist. Reprod. Genet.* **31**, 371–377 (2014).
 47. Du, J. *et al.* Two FSHR variants, haplotypes and meta-analysis in Chinese women with premature ovarian failure and polycystic ovary syndrome. *Mol. Genet. Metab.* **100**, 292–295 (2010).
 48. Conway, G. S. *et al.* Mutation screening and isoform prevalence of the follicle stimulating hormone receptor gene in women with premature ovarian failure, resistant ovary syndrome and polycystic ovary syndrome. *Clin. Endocrinol. (Oxf)*. **51**, 97–99 (1999).
 49. Zhao, S. *et al.* Family-based analysis of eight susceptibility loci in polycystic ovary syndrome. *Sci. Rep.* **5**, 1–6 (2015).
 50. Saxena, R. *et al.* Han Chinese polycystic ovary syndrome risk variants in women of European ancestry: Relationship to FSH levels and glucose tolerance. *Hum. Reprod.* **30**, 1454–1459 (2015).
 51. Ezzidi, I., Mtiraoui, N., Mohammed Ali, M. E., Al Masoudi, A. & Abu Duhier, F. Impact of variants on type-2 diabetes risk genes identified through genomewide association studies in polycystic ovary syndrome: a case–control study. *J. Genet.* **97**, 1213–1223 (2018).
 52. Goodarzi, M. O. *et al.* Replication of association of a novel insulin receptor gene polymorphism with polycystic ovary syndrome. *Fertil. Steril.* **95**, 1736–1741.e11 (2011).
 53. Hwang, J. Y. *et al.* Genome-wide association study identifies GYS2 as a novel genetic factor for polycystic ovary syndrome through obesity-related condition. *J. Hum. Genet.* **57**, 660–664 (2012).
 54. Ying, S. *et al.* Association between Common genetic variants of PCOS risk in a Chinese Han population. *J Clin Res Pediatr Endocrinol* **8**, 405–410 (2016).
 55. Talbot, J. A. *et al.* Molecular scanning of the insulin receptor gene in women with polycystic ovarian syndrome. *J. Clin. Endocrinol. Metab.* **81**, 1979–1983 (1996).
 56. Hanzu, F. A. *et al.* Association of insulin receptor genetic variants with polycystic ovary syndrome in a population of women from Central Europe. *Fertil. Steril.* **94**, 2389–2392 (2010).
 57. Marees, A. T. *et al.* A tutorial on conducting genome-wide association studies: Quality control and statistical analysis. *Int. J. Methods Psychiatr. Res.* **27**, 1–10 (2018).
 58. Association Model Nealelab Github. Available at: https://github.com/Nealelab/UK_Biobank_GWAS/blob/m. (Accessed: 5th August 2019)
 59. Summary Statistic Output Nealelab Github. Available at: https://github.com/Nealelab/UK_Biobank_GWAS/blob/master/imputed-v2-gwas/README.md#summary-stat-output. (Accessed: 5th August 2019)
 60. Bycroft, C. *et al.* Genome-wide genetic data on ~500,000 UK Biobank participants. *bioRxiv* (2017). doi:10.1101/166298
 61. UKBB Interview. Available at: <http://biobank.ctsu.ox.ac.uk/crystal/crystal/docs/Interview.pdf>. (Accessed: 26th July 2019)
 62. UKBB Genotyping QC. Available at: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank_genotyping_QC_documentation-web-1.pdf. (Accessed: 7th August 2019)
 63. UKBB Imputation.
 64. UKBB PCOS Phenotype. Available at: <https://biobank.ctsu.ox.ac.uk/crystal/field.cgi?id=20002>.
 65. UKBB GWAS Imputed v3 File Manifest Release 20180731- PCOS v3 female.
 66. Millard, L. A. C., Davies, N. M., Gaunt, T. R., Smith, G. D. & Tilling, K. Software application profile: PHESANT: A tool for performing automated phenome scans in UK Biobank. *Int. J. Epidemiol.* **47**, 29–35 (2018).
 67. PHESANT Pipeline Github. Available at: https://github.com/Nealelab/UK_Biobank_GWAS/blob/master/imputed-v2-gwas/PHESANT_pipeline.pdf. (Accessed: 7th August 2019)

68. Phenotypes and Applications Github.
69. SNPense. Available at: <https://biit.cs.ut.ee/gprofiler/snpense>. (Accessed: 12th August 2019)
70. NCBI Data Base. Available at: <https://www.ncbi.nlm.nih.gov>. (Accessed: 17th August 2019)
71. Ensembl Data Base. Available at: <https://www.ensembl.org>. (Accessed: 17th August 2019)
72. GeneOntology. Available at: <http://geneontology.org>. (Accessed: 16th August 2019)
73. Temur, M. *et al.* The relationship of urocortin-2 with insulin resistance patients having PCOS. *Gynecol. Endocrinol.* **33**, 124–127 (2017).
74. Barber, T. M. & Franks, S. Chapter 27 - Genetic and Environmental Factors in the Etiology of Polycystic Ovary Syndrome. in *The Ovary (Third Edition)* (eds. Leung, P. C. K. & Adashi, E. Y.) 437–459 (Academic Press, 2019). doi:<https://doi.org/10.1016/B978-0-12-813209-8.00027-3>
75. Tata, B. *et al.* Elevated prenatal anti-Müllerian hormone reprograms the fetus and induces polycystic ovary syndrome in adulthood. *Nat. Med.* **24**, 834–846 (2018).