SNP ANALYSIS OF lncRNA GENES ASSOCIATED WITH POLYCYSTIC OVARY SYNDROME USING GWAS CATALOG AND THE 1000 GENOMES PROJECT

Authors:

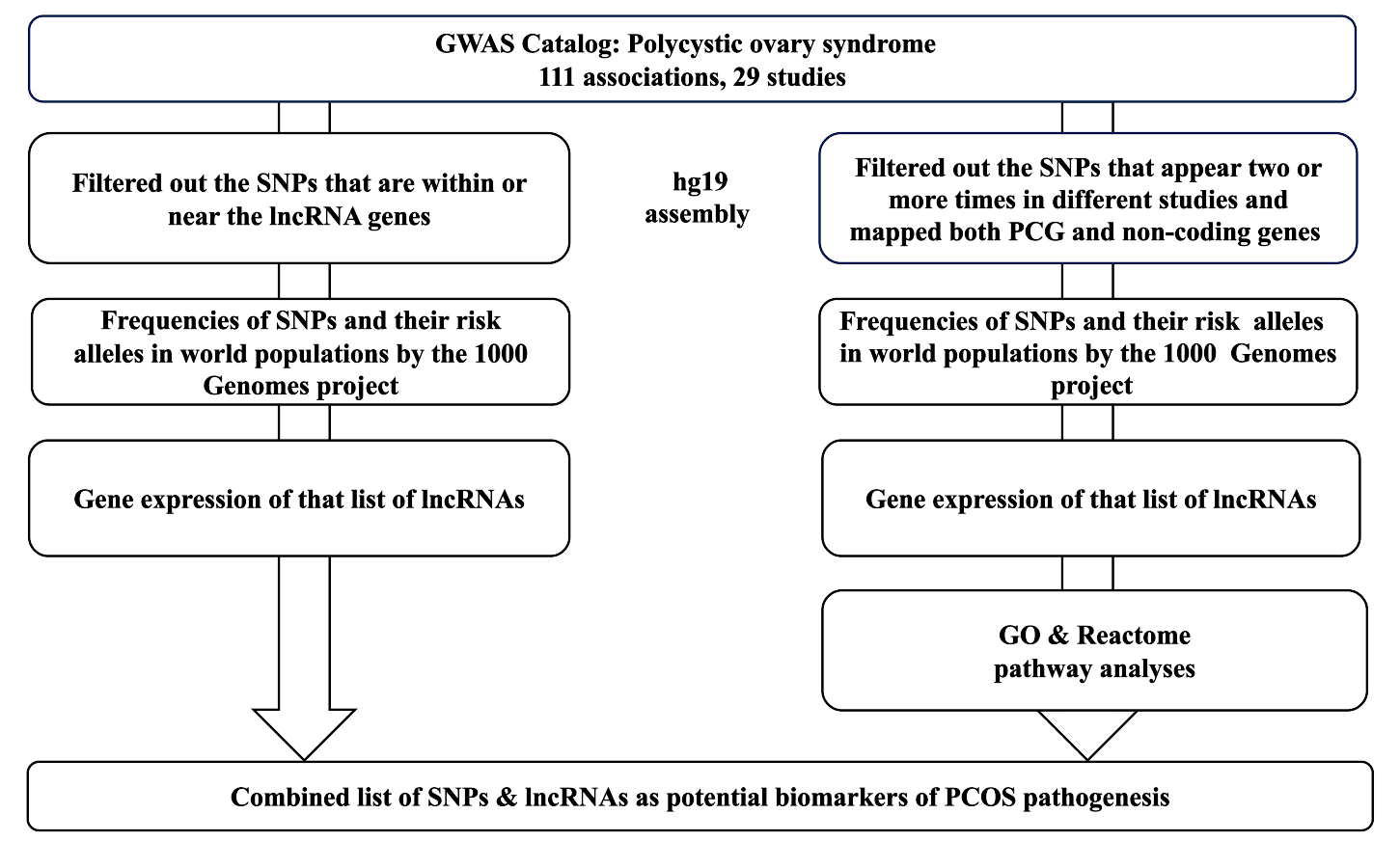
I. Bulanov1, D. Romanov1, L. Lipovich2,3

1. Southern Federal University, Rostov-on-Don, Russia
2. Shenzhen Huayuan Biological Science Research Institute, Shenzhen Huayuan Biotechnology Co., Ltd., 601 Building C1, Guangming Science Park, Fenghuang Street, Shenzhen 518000, Guangdong Province, People's Republic of China.
3. Center for Molecular Medicine and Genetics, Wayne State University, Detroit, Michigan 48201, USA.

**Introduction:** According to WHO, around 8-13% of women of reproductive age suffer from polycystic ovary syndrome (PCOS), while up to 70% of women worldwide with PCOS remain undiagnosed. PCOS is the most common cause of anovulation and the main cause of infertility. Previously, scientists have observed a large number of ncRNAs with altered levels in the plasma, serum or other specimens from patients with PCOS (Han, J., e.a., 2019; Liangshan Mu, e.a., 2021; ElMonier, A. A., 2023). LncRNAs have been identified that regulate the response of androgen, estrogen and progesterone receptors, suggesting that lncRNAs play a role in the hormone regulatory system in PCOS (Zhao X, e.a., 2004; Nabi, M.,e.a.,2023). One of the reasons for impaired expression of reporter genes could be SNP alleles of lncRNA genes (Ali, R. M., e.a. 2022; Gonzalez-Moro, I., 2023, Li, Y. K., 2023).

**Methods:** For testing this hypothesis, we examined 29 Genome-Wide Association Studies comprising 111 associations (SNPs) and a total of 100170 cases and 2088636 controls from four populations: European, Chinese, African, American (https://www.ebi.ac.uk/gwas/efotraits/EFO\_0000660). Then we followed the two approaches of SNP examination: in the first one, for each association we determine the mapped gene (based on hg19 reference genome) and filter out lincRNA genes which contain or close to studied SNPs (within 170 kb in both sides around SNP). In the second approach, we obtained the list of SNPs that appear two or more times in different studies (mapped to both protein-coding and non-coding genes). For two lists of studied SNPs we examined the allele frequency in world populations using the 1000 Genomes project. Then we explored the gene expression of mapped genes using GTEx portal (https://www.gtexportal.org/home/). We also performed Gene Ontology analysis for mapped genes from the second approach using Panther database (<https://www.pantherdb.org/>) and Reactome pathway analysis. At last, based on SNP analysis, the combined list of lncRNAs that may be further examined to define their impact on pathogenesis of PCOS was determined.

**Results:**



Picture 1.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SNP ID | Overlapped Gene (hg19 assembly) | Gene Type | If SNP is intergenic, Nearest Gene | Type of Gene | Distance to Gene, bases | Upstream / Downstream |
| rs78025940 | KAZN-AS1 | lncRNA |  |  |  |  |
| rs181791462 | intergenic | none | RP11-192P3.4 | lincRNA | 99249 | downstream |
| rs1275468 | intergenic | None | RP11-114H23.1 | lincRNA | 21825 | downstream |
| rs1795379 | intergenic | None | RP11-114H23.1 | lincRNA | 15940 | downstream |
| rs7151257 | LINC00871 | lincRNA |  |  |  |  |
| rs500492 | RP11-161M6.3 | lincRNA |  |  |  |  |
| rs4784165 | intergenic | None | CASC22 | lincRNA | 33608 | upstream |
| rs8043701 | intergenic | None | CASC22 | lincRNA | 61566 | upstream |
| rs3813583 | LINC01229, MAFFTR | lincRNAs |  |  |  |  |
| rs9675376 | intergenic | None | RP11-396N11.1 | lincRNA | 893 | upstream |
| rs7574059 | intergenic | None | AC073409.1 | lincRNA | 168983 | upstream |
| rs6022786 | intergenic | None | AC006076.1 | lincRNA | 29446 | upstream |
| rs10938398 | intergenic | None | RP11-362I1.1 | lincRNA | 132470 | upstream |
| rs7666129 | LINC01060 | lincRNA |  |  |  |  |
| rs17186366 | intergenic | None | LINC02529 | lincRNA | 80786 | upstream |
| rs10505648 | LINC02055 | lincRNA |  |  |  |  |