ORIGINAL ARTICLE: GENETICS

The relationship between polycystic ovary syndrome and ancestry in European Americans

Andrew C. Bjonnes, M.S.,[a] Richa Saxena, Ph.D.,[a] and Corrine K. Welt, M.D.[b]

[a] Department of Anesthesia, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts; and [b] Division of Endocrinology, Metabolism, and Diabetes, University of Utah, Salt Lake City, Utah

Objective: To determine whether European Americans with polycystic ovary syndrome (PCOS) exhibit genetic differences associated with PCOS status and phenotypic features.
Design: Case-control association study in European Americans.
Setting: Academic center.
Subject(s): Women with PCOS diagnosed with the use of the National Institutes of Health criteria (n = 532) and control women with regular menstrual cycles and no evidence of hyperandrogenism (n = 432).
Intervention(s): Blood was drawn for measurement of sex steroids, metabolic parameters, and genotyping.
Main Outcome Measure(s): Associations among PCOS status, phenotype, and genetic background identified with the use of principal component analysis.
Result(s): Principal component analysis identified five principal components (PCs). PC1 captured northwest-to-southeast European genetic variation and was associated with PCOS status. Acanthosis was associated with southern European ancestry, and larger waist:hip ratio was associated with northern European ancestry. PC2 was associated with east-to-west European genetic variation and cholesterol levels.
Conclusion(s): These data provide evidence for genetic influence based on European ethnicity in women with PCOS. There is also evidence for a genetic component in the phenotypic features of PCOS within a mixed European population. The data point to the need to control for population stratification in genetic studies in women of mixed European ethnicity. They also emphasize the need for better studies of PCOS prevalence and phenotype as a function of genetic background. (Fertil Steril® 2016;■:■–■. ©2016 by American Society for Reproductive Medicine.)
Key Words: European, genetics, population stratification, waist:hip ratio

Discuss: You can discuss this article with its authors and with other ASRM members at

It is important to control for ancestry in genetic association studies to avoid false positive results confounding population differences between case and control subjects. Genetic association studies in Americans of European descent may be particularly prone to population stratification (1). Although it is often accepted that Europeans are genetically homogeneous, there is distinct genetic substructure in the European population (2–5). Three unique populations can be identified, with individuals of northwest European, southeast European, and Ashkenazi Jewish ancestry sharing genetic substructure (1, 2, 6). The genetic substructure can affect studies in European countries with multiple immigrant populations and countries such as the United States, in which ethnic groups from these European regions have distinct immigration patterns (2, 5). Therefore, population stratification may occur if the ethnic populations are not carefully matched. Genome-wide association studies use a predetermined set of variants that can be used to examine, and control for, population substructure and stratification. However, the European population markers may not be taken into account in candidate association and replication studies. They also fail to prevent spurious associations from occurring when using next-generation sequencing to study associations between disease and rare (<5%) exome variants (7).

Earlier studies have examined population stratification among ethnically

Received May 18, 2016; revised July 21, 2016; accepted August 16, 2016.
A.C.B. has nothing to disclose. R.S. has nothing to disclose. C.K.W. has received royalties from Uptodate unrelated to the current topic.
Supported by award no. R01HD065029 from the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, award no. 1 UL1 RR025758, Harvard Clinical and Translational Science Center, from the National Center for Research Resources, and award no. 1-10-CT-57 from the American Diabetes Association. The content is solely the responsibility of the authors and does not necessarily represent the official views of the *Eunice Kennedy Shriver* National Institute of Child Health and Human Development, National Center for Research Resources, National Institutes of Health, or American Diabetes Association.
Reprint requests: Corrine K. Welt, M.D., Division of Endocrinology, Metabolism, and Diabetes, University of Utah, 15 N 2030 E, Salt Lake City, Utah 84112 (E-mail: cwelt@genetics.utah.edu).

Fertility and Sterility® Vol. ■, No. ■, ■ 2016 0015-0282/$36.00

http://dx.doi.org/10.1016/j.fertnstert.2016.08.033

diverse women with polycystic ovary syndrome (PCOS). A multiethnic group of women with PCOS from the Netherlands (8) was examined with the use of a genome-wide panel of ancestry informative markers that distinguished women of African, southeast Asian, Hindustani, and European ancestry. Six distinct clusters were identified, representing the distinct subgroups of African, Surinam Creole, Asian, and Caribbean ethnicities, and, importantly, women of northern European and Turkish ethnicity clustered into two distinct groups. In addition, the genetic ancestral background accounted for a proportion of the phenotypic features of PCOS (9), with previous work by the same group demonstrating that the subset of women from Mediterranean Europe manifested greater obesity and hyperandrogenism compared with other groups in that study (10). In addition, a mixed European group from Boston manifested greater hyperandrogenism than women from Iceland (11). Therefore, the PCOS phenotype may include distinctive features depending on European ethnic origin (10, 11), and these may be partially determined by differences in genetics. Taken together, the genetic and phenotypic distinctions among Europeans may result in differences in ascertainment of PCOS or in expression of its features in distinct ethnic groups. These studies make it important to examine a broad population of European women and to compare genetic stratification in women with PCOS and control women.

We hypothesized that European population stratification would be present in association studies of PCOS in European Americans. Based on data from the Netherlands, we also hypothesized that phenotypic features in women of European ancestry with PCOS would exhibit differences based on the southeast-to-northwest population substructure of European Americans (1, 2, 6). To test these hypotheses, we analyzed PCOS status and phenotype as a function of the principal components of population structure, with the use of markers informative for European ancestry, in our cohort from a genome-wide association study of women with PCOS and control women of European ancestry (12, 13).

RESEARCH DESIGN AND METHODS

Subjects

All subjects were U.S. women of reported European ancestry, aged 18–45 years, and recruited at Massachusetts General Hospital in Boston, Massachusetts. Subjects with PCOS (n = 532) had oligomenorrhea (<9 menstrual periods per year) and clinical and/or biochemical evidence of hyperandrogenism, fulfilling the National Institutes of Health criteria (11). Clinical hyperandrogenism was defined by: 1) Ferriman-Gallwey score >9 (14); or 2) acne on the face or back. Biochemical hyperandrogenism was defined as T >63 ng/dL (2.8 nmol/L), DHEAS >430 μg/dL (1.16 μmoL/L) or A >3.8 ng/mL (13.3 nmol/L) (11). Control subjects (n = 432) had regular menstrual cycles of 21–35 days and no physical or biochemical evidence of hyperandrogenism.

Subjects were excluded for a personal history or biochemical evidence of late onset congenital adrenal hyperplasia (11). All subjects had normal thyroid function and PRL levels and a follicular-phase FSH level in the premenopausal range. Subjects were on no hormonal medication, except for stable thyroid hormone replacement.

Protocol

The study was approved by the Institutional Review Board of the Massachusetts General Hospital, and all subjects gave written informed consent. All PCOS subjects were studied ≥10 days after their last menstrual period and after a 12-hour fast (11). Subjects underwent a detailed history, a physical exam, pelvic ultrasound (ATL HDI 1500; 5-MHz convex array transducer), and blood samples for lipids, glucose, insulin, gonadotropin, and sex steroid levels. An oral glucose tolerance test was performed, with blood sampling 2 hours after a 75-gram glucose load.

Genotyping

Patient DNA was isolated from whole blood and genotyped with the use of the Omnihumanexpress Bead Chip (Illumina) with 951,117 single-nucleotide polymorphisms (SNPs). Subjects were removed for inbreeding (n = 16) and for population stratification after analysis with the use of Eigenstrat for subjects failing to cluster with European cohorts (n = 60), with some samples excluded for both (n = 15). SNPs with >5% missing genotype were excluded.

Statistical Analysis

PCOS status. A subset of 240,000 markers informative for European, African-American, and Latin-American ancestry was used for analysis (2, 15, 16). These variants were used to determine the genetic variability mathematically by structuring the data into principal components (PCs). PC analysis was performed with the use of Eigensoft for case subjects, control subjects, and the combined groups with age and body mass index (BMI) as a covariates (17, 18).

Phenotype. Quantitative traits were log transformed for analysis. Logistic or linear regression analysis was used to examine associations between the five PCs identified and PCOS status and 17 log-transformed quantitative traits in the combined sample of PCOS case subjects and control subjects and the case and control subjects as separate groups, adjusting for PCOS status, age, and BMI. A P value of $< .007$ was considered to be significant after Benjamini and Hochberg false discovery rate correction for five PCs and ten independent traits, with other variables highly correlated (trait family [correlated measurements], gonadotropins [LH, FSH], 17OH-P, T [A, DHEAS, SHBG], cholesterol [low-density lipoprotein (LDL), high-density lipoprotein (HDL)], acanthosis nigricans, blood pressure [systolic blood pressure, diastolic blood pressure], body mass index [waist:hip ratio], fasting glucose, fasting insulin, and ovarian volume). Data were plotted against those obtained from European-based Human Genome Diversity Project datasets, including Italian from Bergamo, Tuscan (central Italy), Russian, Orcadian (Orkney Islands, Scotland), French, Basque (Northern Spain and Southern France), Sardinian (autonomous Italian island), and Adygei (Republic of Russia, Caucasian) (19–21).