Identification of reference genes for qRT-PCR in granulosa cells of healthy women and polycystic ovarian syndrome patients

Yue Lv1,2,3, Shi Gang Zhao1,3, Gang Lu2,3, Chi Kwan Leung1,3,4, Zhi Qiang Xiong1,3,4, Xian Wei Su1,3,4, Jin Long Ma1,2,3,4, Wai Yee Chan2,3 & Hong Bin Liu1,2,3

Comparative gene expression analysis by qRT-PCR is commonly used to detect differentially expressed genes in studies of PCOS pathogenesis. Impaired GC function is strongly associated with PCOS pathogenesis, and a growing body of studies has been dedicated to identifying differentially expressed genes in GCs in PCOS patients and healthy women by qRT-PCR. It is necessary to validate the expression stability of the selected reference genes across the tested samples for target gene expression normalization. We examined the variability and stability of expression of the 15 commonly used reference genes in GCs from 44 PCOS patients and 45 healthy women using the GeNorm, BestKeeper, and NormFinder statistical algorithms. We combined the rankings of the three programs to produce a final ranking based on the geometric means of their stability scores. We found that HPRT1, RPLP0, and HMBS out of 15 examined commonly used reference genes are stably expressed in GCs in both controls and PCOS patients and can be used for normalization in gene expression profiling by qRT-PCR. Future gene-expression studies should consider using these reference genes in GCs in PCOS patients for more accurate quantitation of target gene expression and data interpretation.

Polycystic ovarian syndrome (PCOS) is characterized by ovulatory dysfunction, polycystic ovaries, and hyperandrogenism and is a common gynecological endocrinopathy and a leading cause of female reproductive failure1,2. Its prevalence is estimated to be 6% to 18% depending on the diagnostic criteria used3–5. PCOS patients are prone to multiple metabolic disorders, including insulin resistance, type 2 diabetes mellitus, obesity, dyslipidemia, chronic low-grade inflammation, cardiovascular diseases, and psychological disturbances1–3,6–11.

While the etiopathogenesis of PCOS remains elusive, it likely involves aberrant hormonal responses mediated by ovarian granulosa cells (GCs) during the progression of follicular development. Previous studies have reported that PCOS is associated with elevated luteinizing hormone (LH) and overexpression and overactivation of the luteinizing hormone/chorionic gonadotropin receptor (LHCGR) in GCs from PCOS patients and that this prevents follicle maturation and ovulation12–16. Impaired GC function is associated with disruption of folliculogenesis and with elevated intraovarian androgens and circulating anti-Müllerian hormone levels17–23.

Comparative gene expression profiling in normal and disease samples provides invaluable insights into the underlying mechanism of PCOS pathogenesis. Quantitative real-time PCR (qRT-PCR), microarray and RNA-seq have been widely used to measure differential gene expression of GCs from metabolically normal subjects and PCOS patients to identify the genetic regulators and their associated pathways as well as to identify biomarkers in PCOS24–35. Most studies have simply selected a reference gene, e.g. β-actin (ACTB)36,37, 18S ribosomal RNA (RNA18S)38, peptidylprolyl isomerase B (PPIB)39,40, or glyceraldehyde-3-phosphatedehydrogenase (GAPDH)41, as an endogenous internal control for target gene expression normalization without careful examination of their expression variability and stability in the tested samples. Accumulating evidence suggests that these reference genes...
Table 1. Anthropometric and clinical data of PCOS patients and healthy controls. BMI: body mass index; E2: estrogen; P: progesterone; PRL: prolactin; T: testosterone; FSH: follicle stimulating hormone LH: luteinizing hormone. *p < 0.05 indicates statistical significance.

|   | Control (n = 45) | PCOS (n = 44) | p-value |
|---|-----------------|---------------|---------|
| Age (years) | 29.64 ± 2.73 | 28.59 ± 3.20 | 0.09   |
| BMI (kg/m²) | 22.45 ± 2.99 | 24.20 ± 4.73 | 0.04*  |
| E2 (pg/ml) | 35.27 ± 10.66 | 39.09 ± 15.49 | 0.17 |
| P (ng/ml) | 0.62 ± 0.23 | 0.67 ± 0.39 | 0.48 |
| PRL (ng/ml) | 18.66 ± 9.63 | 17.69 ± 9.51 | 0.63 |
| T (ng/dl) | 23.87 ± 7.77 | 36.12 ± 17.6 | <0.001* |
| FSH (IU/L) | 6.54 ± 1.28 | 6.26 ± 1.51 | 0.35 |
| LH (IU/L) | 5.14 ± 1.84 | 8.10 ± 3.87 | <0.001* |
| LH/FSH | 0.80 ± 0.29 | 1.33 ± 0.74 | <0.001* |

Genes, previously thought to be stably expressed with minimal variability across samples, are in fact variably expressed under different physiological and pathological conditions41,49. The adoption of an inappropriate housekeeping gene could result in under- or overestimation of gene expression and subsequent misinterpretation of the data42,44. To the best of our knowledge, there are no published data that can be used to select reliable reference genes for gene expression normalization in the GCs of reproductively healthy women and PCOS patients.

Given the crucial role of GCs in PCOS pathology, we examined the expression stability and variability of the following 15 commonly used reference genes: RNA18SS, ACTB, GAPDH, ubiquitin C (UBC), β-2-microglobulin (B2M), RNA polymerase II subunit A (POLR2A), hypoxanthine phosphoribosyltransferase 1 (HPRT1), 3-tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein (YWHAZ), hydroxymethylbilane synthase (HMBS), β-glucuronidase (GUSB), importin 8 (IPO8), phosphoglcerate kinase 1 (PGK1), ribosomal protein lateral stalk subunit P0 (RPLP0), transferrin receptor (TFRC), and ribosomal protein L13a (RPL13A) in GCs from healthy control subjects and PCOS patients. This study thus provides valuable insights into the selection of appropriate reference genes for gene expression profiling in GCs from PCOS patients.

Results

Clinical and biochemical parameters of the study subjects. The clinical and biochemical parameters of the 44 PCOS patients and 45 healthy controls are presented in Table 1. There were no significant differences in the serum concentrations of estrogen, progesterone, prolactin, or FSH for the PCOS patients and the age-matched healthy controls. As expected, the BMI of the PCOS group was significantly higher than the control group by 7.8% (PCOS: 24.20 ± 4.73 kg/m², control: 22.45 ± 2.99 kg/m², 95% CI: −1.771 to 9.411, t = 1.358 df = 87, p < 0.05). The PCOS group also exhibited 33.9% higher levels of testosterone compared to the control group (PCOS: 36.12 ± 17.6 ng/dl, control: 23.87 ± 7.77, 95% CI: 6.540 to 17.96, t = 4.264 df = 87, p < 0.0001). Although the FSH levels were indistinguishable between the two groups, both LH and the LH/FSH ratio were strongly increased in the PCOS group by 36.4% and 39.8%, respectively (PCOS LH: 8.10 ± 3.87 IU/L, control LH: 5.14 ± 1.84 IU/L, 95% CI: 1.688 to 4.232, t = 4.624 df = 87, p < 0.0001; PCOS LH/FSH: 1.33 ± 0.74, control LH/FSH: 0.80 ± 0.29, 95% CI: 0.2942 to 0.7658, t = 4.467, df = 87, p < 0.0001).

Expression levels of the reference genes. The expression levels of the 15 candidate reference genes in all 44 PCOS patients and all 45 controls were evaluated based on their qRT-PCR threshold cycle value (Ct). As shown in Fig. 1, their Ct values exhibited a dynamic expression range from the lowest of 13.81 ± 0.54 for RNA18SS to the highest of 28.25 ± 1.13 for HMBS with a median Ct value of 21.35 ± 1.12. Among the 89 samples, RNA18SS and ACTB were the most abundant genes, followed by GAPDH, while the least abundant gene was HMBS. The assessed reference genes showed significant inter-individual variations across the samples, with SD values ranging from 0.54 for RNA18SS to 1.26 for RNA18SS with the lowest M-values of 0.27 were the most stably expressed gene pair, followed by ACTB (0.41), HPRT1 (0.45), YWHAZ (0.46), GAPDH (0.49), PGK1 (0.52), HMBS (0.53), UBC (0.55), IPO8 (0.56), B2M (0.58), POLR2A (0.60), GUSB (0.62), TRFC (0.64), and RNA18SS with the highest M-value of 0.67 making it the least stable gene across all samples.

Analysis of reference gene expression stability. We used three common algorithms – GeNorm, NormFinder, and BestKeeper – to individually examine the expression stability of the 15 candidate reference genes (Please refer to the Methods).

GeNorm analysis. GeNorm computed the M-values using a pairwise comparison approach to grade the 15 reference genes based on the similarity of these expression profiles rather than the minimal variation across the 89 samples. Lower M-values indicated greater expression stability. The computed M-values of all of the reference genes were below the defined cut-off value of 1.5, which are considered stably expressed by the algorithm (Fig. 2A). Following the step-wise exclusion of the least stably expressed gene with the highest M-value, the stability values were recalculated. In the pooled group, RPL13A and RLPP0 with the lowest M-values of 0.27 were the most stably expressed gene pair, followed by ACTB (0.41), HPRT1 (0.45), YWHAZ (0.46), GAPDH (0.49), PGK1 (0.52), HMBS (0.53), UBC (0.55), IPO8 (0.56), B2M (0.58), POLR2A (0.60), GUSB (0.62), TRFC (0.64), and RNA18SS with the highest M-value of 0.67 making it the least stable gene across all samples.
We also computed the optimal number of reference genes based on the pairwise variation value \( V_{n/n} \). GeNorm defines a pairwise variation of 0.15 as the cutoff value, below which \( V_{n/n} < 0.15 \) the inclusion of an additional reference gene for normalization is not needed. As shown in Fig. 2B and C, the V2/3 value is just over the cutoff threshold of 0.15 with an expression ratio of 0.1536 \( (r^2 = 0.978) \), while the V3/4 value was below the cutoff with an expression ratio of 0.1047 \( (r^2 = 0.986) \), indicating that the top three ranked reference genes RPL13A, RLPP0, and ACTB (with the lowest M-values) were suggested for gene expression normalization in GCs from PCOS patients.

**BestKeeper analysis.** The BestKeeper applet computes the gene expression variation for individual reference genes according to the SD and CV of their Ct values across all of the samples. Figure 3 shows that GUSB was ranked the most stably expressed reference gene with a SD \( (\pm \text{Ct}) \) and CV \( (%\text{Ct}) \) of 0.70 and 2.67, respectively, followed by R1A18S5 \( (\text{SD: 0.40, CV: 2.87}) \), HMBS \( (\text{SD: 0.90, CV: 3.19}) \), IP08 \( (\text{SD: 0.88, CV: 3.38}) \), TFRC \( (\text{SD: 0.85, CV: 3.45}) \), POLR2A \( (\text{SD: 0.88, CV: 3.58}) \), HPRT1 \( (\text{SD: 0.88, CV: 3.59}) \), RLPP0 \( (\text{SD: 0.75, CV: 3.77}) \), RPL13A \( (\text{SD: 0.75, CV: 3.97}) \), YWHAZ \( (\text{SD: 0.88, CV: 4.12}) \), PGK1 \( (\text{SD: 0.95, CV: 4.30}) \), ACTB \( (\text{SD: 0.76, CV: 4.36}) \), UBC \( (\text{SD: 1.00, CV: 4.72}) \), B2M \( (\text{SD: 0.93, CV: 4.73}) \), and GAPDH \( (\text{SD: 0.94, CV: 5.02}) \), which was ranked the least stable gene in GCs from PCOS.

**NormFinder analysis.** We next used NormFinder, a model-based variance estimation algorithm based on the inter-group variation across samples and intra-group expression variation within the same condition to identify the optimal reference gene with minimal variability and coregulation across the samples. As shown in Fig. 4A, in contrast to the GeNorm and BestKeeper results, NormFinder ranked HPRT1 as the most stable reference gene with an M-value of 0.07, followed by YWHAZ \( (0.09) \), ACTB \( (0.42) \), HMBS \( (0.11) \), PGK1 \( (0.11) \), UB \( (0.12) \), GAPDH \( (0.19) \), GUSB \( (0.2) \), IP08 \( (0.21) \), POLR2A \( (0.22) \), RLPP0 \( (0.24) \), R1A18S5 \( (0.24) \), TFRC \( (0.25) \), RPL13A \( (0.25) \), and B2M; which, with the highest M-value of 0.28, was ranked as the least stable gene in GCs from PCOS patients.

Figure 4B shows the log difference between groups of the 15 reference genes ranked by the three different algorithms. The model-based NormFinder selected the most stable genes with minimal combined inter- and intra-group expression variation, whereas the pairwise comparison algorithms GeNorm and BestKeeper selected the genes with a low intragroup variation and roughly the same nonvanishing intergroup gene expression variation.

**Final ranking of the 15 reference genes.** Given the specific features of the three statistical algorithms, it is not surprising to see marked differences in ranking the 15 candidate reference genes. We used a previously reported method to take into consideration the three sets of results to produce a final ranking of the 15 reference genes47, 48. Briefly, the 15 reference genes were re-ranked according to their computed geometric mean from the three algorithms, and a smaller geometric mean indicated higher gene expression stability across all samples. As shown in Table 2, HPRT1 was the most stably expressed gene with the lowest geometric mean, followed by RLPP0, HMBS, YWHAZ, GUSB, ACTB, RPL13A, R1A18S5, IP08, PGK1, GAPDH, UBS, POLR2A, TFRC, and B2M, which was the least stable. Thus HPRT1, RLPP0, and HMBS were the top three ranked reference genes for use as internal controls for qRT-PCR to normalize gene expression in GCs from PCOS patients.

**Discussion**

PCOS is a heterogeneous gynecological endocrine disorder characterized by a broad spectrum of anomalies, including chronic anovulation, hyperandrogenism, and hyperinsulinism with insulin resistance, and a growing body of evidence has linked dysfunctional GCs to the pathogenicity of PCOS. To gain insight into the underlying...
mechanism of impaired GC functions in PCOS development, qRT-PCR has been used extensively to measure differential gene expression in normal and PCOS samples. Most researchers have chosen a reference gene for target gene expression normalization based on the assumption of its inherent expression stability across the tested samples without prior validation. The ideal reference gene requires a constant expression level across all treatments, physiological conditions, tissue/cell types, and experimental designs, but no such reference genes have been identified and they might not even exist.

The expressions of the commonly used reference genes in qRT-PCR analyses might vary considerably in specific biological contexts resulting in errors in gene expression estimations. Sadek et al. reported that YWHAZ, CYCI, and ACTB were the most stably expressed reference genes out of nine examined housekeeping genes.
in endometrial tissues collected from PCOS patients, and the commonly used GAPDH gene was not recommended\(^5\). Milutinović et al. showed that HPRT, in comparison to ACTB, BSM, and GAPDH, was the most stable
In this study, we evaluated 15 candidate reference genes commonly used for routine gene expression normalization in comparative gene expression profiling by qRT-PCR. Given their dynamic expression levels with significant expression variation across the samples (Fig. 1), we used three commonly used statistical programs – GeNorm, BestKeeper, and NormFinder – to analyze their Ct values and to rank them according to their M-values. GeNorm indicated that RPL13A and RPLP0 were the two most stably expressed genes with the same computed M-values, followed by ACTB (Fig. 2A). The best combination recommended by GeNorm was RPL13A, RPLP0, and ACTB (Fig. 2B,C). In contrast, BestKeeper ranked GUSB, RNA18S5, and HMBS as the top three most stable genes, while RPLP0, RPL13A, and ACTB were ranked at 8th, 9th, and 12th place, respectively. It is also notable that RNA18S5 was the least stable gene according to GeNorm (Fig. 3). NormFinder determined HPRT1 and YWHAZ to be the two most stable genes, which were ranked at 4th and 5th place, respectively, by GeNorm and at 7th and 12th, respectively, by BestKeeper (Fig. 4). It is worth noting that both GeNorm and NormFinder ranked ACTB and GAPDH, the two most commonly used reference genes, at the top position; however, they were ranked as the 12th and the least stable genes by BestKeeper (Table 2). The rankings by GeNorm and NormFinder were more consistent with each other than with the BestKeeper algorithm. Finally, we produced the final ranking by considering the results generated by the three programs. We identified HPRT1, RPLP0, and HMBS as the three most stably expressed genes for use as internal controls for comparative gene expression profiling by qRT-PCR (Table 2).

To the best of our knowledge, there is no gold standard for the optimal number of the candidate reference genes required for the expression stability study. In general, six to twelve candidate genes are commonly included for the analysis. The supply of cDNA samples and the lab resources dedicated for the characterization should also be considered to make the final decision on the number of candidate genes used for the analysis.

The three identified stably expressed reference controls are specific to our experimental design. It would be of great interest to test the expression stability of these normalizers in the GCs sampled from the PCOS patients from different geographic populations or different species. Nonetheless, our statistical pipelines to identify stably expressed reference genes for comparative gene expression analysis should be broadly applied for general biomedical research.

### Method

**Patient recruitment.** Written informed consent was obtained from all participants or their next of kins, and the study was approved by the Institutional Review Board of Reproductive Medicine of Shandong University, and all methods were performed in accordance with the relevant guidelines and regulations. Eighty-nine Chinese Han women aged 29.12 ± 3.01 years were recruited from the Reproductive Hospital Affiliated to Shandong University between January 2015 and June 2016. This study included 44 PCOS patients and 45 women with normal ovarian function undergoing in vitro fertilization (IVF) for tubal and/or male infertility treatment. The study required no modification of our routine IVF protocol.

**Clinical and biochemical measurements.** PCOS was diagnosed according to the 2003 Rotterdam criteria, including any two of the following three clinical features: oligo/anovulation, clinical and/or biochemical hyperandrogenism, and polycystic ovaries on ultrasound. Women with other pathophysiological conditions associated with hyperandrogenemia, including adrenal congenital hyperplasia, Cushing’s syndrome, and androgen-secreting tumors, were excluded.

| Ranking | GeNorm       | NormFinder | BestKeeper | Final ranking |
|---------|--------------|------------|------------|---------------|
| 1       | RPL13A/RPLP0 | HPRT1      | GUSB       | HPRT1         |
| 2       | ACTB         | YWHAZ      | RNA18S5    | RPLP0         |
| 3       | HPRT1        | HMBS       | IPOP8      | YWHAZ         |
| 5       | YWHAZ        | PGK1       | TFRC       | GUSB          |
| 6       | GAPDH        | UBC        | POLR2A     | ACTB          |
| 7       | PGK1         | GAPDH      | HPRT1      | RPL13A        |
| 8       | HMBS         | GUSB       | RPLP0      | RNA18S5       |
| 9       | UBC          | IPOP8      | RPL13A     | IPOP8         |
| 10      | IPOP8        | POLR2A     | YWHAZ      | PGK1          |
| 11      | B2M          | RPLP0      | PGK1       | GAPDH         |
| 12      | POLR2A       | RNA18S5    | ACTB       | UBC           |
| 13      | GUSB         | TFRC       | UBC        | POLR2A        |
| 14      | TFRC         | RPL13A     | B2M        | TFRC          |
| 15      | RNA18S5      | B2M        | GAPDH      | B2M           |

Table 2. Final rankings of the 15 candidate reference genes.
| Symbol | Gene name | Accession number | Forward primer (5′→3′) | Reverse primer (5′→3′) |
|-------|-----------|-----------------|-----------------------|----------------------|
| RNA18S5 | 18S ribosomal RNA | NR_003286.2 | GGC GCC CCC CTT CGA GTG CTT TTA G | GCT CGG GGC CTT GCT TGA AC A C T CT |
| ACTB | Actin beta | NM_001101.3 | CGA CAG GAT GCC AGA GAGG | AC A CT GCT GGA AGG TGG GA |
| GAPDH | Glyceraldehyde-3-phosphate dehydrogenase | NM_001256799.2 | TGA TCA TCA A GA A GTG GTA GA | TCC TGT GAG GCA CTA GAG GCA T |
| UBC | Ubiquitin C | NM_021009.6 | AAA GGG CTC GCT TCT TTT TG | TGG CTC TGA ACT CTC GGT TGA |
| B2M | Beta-2-microglobulin | NM_000408.2 | TGA CAT TGT CAC GCA CAA GATA | GCG CAT TCA AAC ACCT TCA |
| POLR2A | RNA polymerase II subunit A | NM_000937.4 | GA CCA CAG TCT CAA TTA GTA C | GTT GC GGC GTC TTC CAA |
| HPRT1 | Hypoxanthine phosphoribosyltransferase 1 | NM_001194.2 | TG A CAC TGG CAA A A A GT C A C | GCG CAG GGA CAA A CCA TGA |
| YWHAZ | Tyrosine 3-monoxygenase/tryptophan 5-monoxygenase activation protein, zeta polypeptide | NM_001135699.1 | ACT TTT GTC AT A TTT GCG TTC TCAA | CGC CAC GGA CAA A CCA TGA |
| HMBS | Hydroxymethyl-bilane synthase | NM_001135699.1 | GGC AAC TCG GCT GAA | GGT TAC CAC CGG CAT A C |
| GUSB | Glucuronidase beta | NM_000181.3 | GTA TTT ACC TAA TGG CCA | CAC AGG CAA GGT A TGA |
| IPO8 | Importin 8 | NM_001190995.1 | TCC GCA ACT ATT TCG ACG GAC | GTT CAA A GAG GCC GAC TCA A |
| PGK1 | Phosphoglycerate kinase 1 | NM_000291.3 | GTA CCA A TGA TTA AGG GAC | GTT GC GAG TAC TCT CAA |
| RPLP0 | Ribosomal protein lateral stalk subunit P0 | NM_001002.3 | AG CCA GAA AC TGT GTC TCT | ACT CAG AT TTT CAC GGT GAA |
| TFR2 | Transferrin receptor | NM_001128148.2 | GGC TAC TTG GGC TAT TGG AAGG | CAG TTT CCT CCG A CAA CTT TCT |
| RPL13A | Ribosomal protein L13a | NM_001270491.1 | GCC ATC GAG TCT A A CAG GTA | GTT GGT GTT CAT CCG GTC TGC |

**Table 3.** qPCR primer set of the 15 reference genes.

Anthropometric parameters of all subjects, including body height, and weight were measured during the first visit to the outpatient department. The body mass index (BMI) was calculated as weight (kg) divided by the square of the height (m²). Venous blood samples were collected between 8:00 a.m. and 10:00 a.m. after a 12 h overnight fast. All blood samples from PCOS patients were obtained during the early follicular phase of their menstrual cycle on days 3–5. The serum levels of follicle-stimulating hormone (FSH), luteinizing hormone (LH), total testosterone, estrogen, progesterone, and prolactin were assayed enzymatically using an automated biochemistry analyzer (Olympus 600, Clinical Chemistry Analyzer, Olympus Diagnostica GmbH, Co. Clare, Ireland).

**Isolation of GCs.** Cumulus GCs were isolated from all 44 PCOS patients undergoing IVF and the 45 women in the control group undergoing regular treatment for tubal and/or male factor infertility. Volunteers with a history of other gynecological or medical disorders were excluded. All women were injected with gonadotropin-releasing hormone agonist at the onset of their midluteal phase, and an ultrasound scan and serum estradiol assays were performed to monitor follicle size. When three or more follicles with a mean diameter of R1.8 cm were detected, 8,000–10,000 IU human chorionic gonadotropin (Profasi, Serono) was administered 36 hours prior to the ultrasound-guided immature oocyte retrieval procedure. After anesthetization, the oocyte retrieval was performed through a 17-gauge double-lumen aspiration needle (K-OPS-WOOD-1235, Cook Australia). The cumulus GCs around the oocytes were collected and washed twice with Dulbecco’s modified Eagle medium (DMEM) for subsequent analysis after oocyte removal in follicular aspirates using a Pasteur pipette. Red blood cells were removed with lysis buffer.

**RNA extraction and cDNA synthesis.** Total RNA from isolated GCs was extracted using the Trizol Plus RNA Purification kit (Life Technologies) according to the manufacturer’s instructions. The RNA purity was confirmed using a NanoDrop 2000 (Thermo Scientific) and an A260:A280 ratio of 1.9–2.1. Total RNA (1 µg) was reverse transcribed to cDNA using the PrimeScript RT reagent Kit (Takara) and diluted with nuclease-free water to a final volume of 20 µl. The cDNAs were further diluted 1:20 with nuclease-free water for use as the DNA template for qRT-PCR.

**qRT-PCR.** We used an epMotion 5075 LH (Eppendorf, Germany) automated liquid handling workstation for precise and accurate pipetting of qPCR reagents. The qPCR mix was prepared in a 10 µl final volume containing 5 µl SYBR Green Master Mix, 0.5 µl of primer pair solution (10 µM), 3 µl of nuclease-free water, and 1 µl of the diluted cDNA. The amplification reactions were performed in 384 well plates using PowerUp™ SYBR® Green Master Mix (ThermoFisher) on a QuantStudio™ 7 Flex Real-Time PCR System (Applied Biosystem, USA) according to the manufacturer’s instructions. The thermal cycling conditions were set as follows: uracil-DNA glycosylase (UDG) activation at 50 °C for 2 minutes, DNA polymerase activation at 95 °C for 2 minutes, 40 cycles of denaturation at 95 °C for 15 seconds and annealing/extension at 55 °C for 30 seconds. Specificity of amplification was confirmed by melting curve analysis. The qPCR primers sequences are shown in Table 3.

**Statistical Analyses.** The expression stability of the 15 reference genes was analyzed using the three R-based or Microsoft Excel-based algorithms – GeNorm 46, NormFinder 79, and BestKeeper 80. GeNorm selects an ideal reference gene pair whose expression ratios exhibit minimal influence by external factors across different experimental conditions and samples. The first step in the algorithm is to determine the computed stability value (M-value) based on the average pair-wise variation between a gene and all other assessed reference genes as an indication of the stability of the reference gene expression. The gene with the highest M-value is then eliminated, and the selection process will be repeated to select the two most stably expressed genes with the lowest M-values.
GeNorm was also used to determine the minimal number of reference genes for gene expression normalization. According to the pairwise variation calculation, for any gene pair with a cutoff ($V_{n+1}$) lower than 0.15, an additional (n + 1) reference gene will not significantly improve the normalization and thus is not required for normalization.

NormFinder calculates the variance within and between groups to determine the gene expression stability. Based on the estimate of the intragroup and intergroup stability, a stability value will be assigned for each reference gene program ranks all assessed genes based on the Stability values and the gene with the lowest stability is considered the most stably expressed gene.

BestKeeper determines the expression stability according to the coefficient of correlation to the BestKeeper index, which consists of the geometric means of the Ct values of the tested gene set. The coefficient of variance (CV) and standard derivation (SD) are determined, and the reference gene with the lowest CV and SD is considered the most stably expressed gene. Genes with $SD > 1$ are discarded. This program considers the relationship between the BestKeeper index and the reference genes by calculating the Pearson correlation coefficient, the coefficient of determination ($r^2$), and the $p$-value.

Numerical values of the clinical characteristics of PCOS cases and controls were expressed as the mean ± SD. Independent-samples t-tests were performed using SPSS v.19.0 (SPSS Inc., Chicago, IL, USA). A value of $p < 0.05$ was regarded as statistically significant.

References
1. Wild, S., Pierpoint, T., Jacobs, H. & McKeigue, P. Long-term consequences of polycystic ovary syndrome: results of a 31 year follow-up study. Hum Fertil (Camb) 3, 101–105 (2000).
2. Kovacs, G. & Wood, C. The current status of polycystic ovary syndrome. Aust N Z J Obstet Gynaecol 41, 65–68 (2001).
3. Azizi, F. et al. The prevalence and features of the polycystic ovary syndrome in an unselected population. J Clin Endocrinol Metab 89, 2745–2749, doi:10.1210/jc.2004-032046 (2004).
4. Christian, R. C. et al. Prevalence of polycystic ovary syndrome: a meta-analysis. J Clin Endocrinol Metab 88, 2562–2568, doi:10.1210/jc.2003-030334 (2003).
5. March, W. et al. The prevalence of polycystic ovary syndrome in a community sample assessed under contrasting diagnostic criteria. Hum Reprod 25, 544–551, doi:10.1093/humrep/dep399 (2010).
6. Ehrmann, D. A. Polycystic ovary syndrome. The New England journal of medicine 352, 1223–1236, doi:10.1056/NEJMra041536 (2005).
7. Jayasena, C. N. & Franks, S. The management of patients with polycystic ovary syndrome. Nat Rev Endocrinol 10, 624–636, doi:10.1038/nrendo.2014.102 (2014).
8. Orio, F. J. et al. The cardiovascular risk of young women with polycystic ovary syndrome: an observational, analytical, prospective case-control study. J Clin Endocrinol Metab 89, 3696–3701, doi:10.1210/jc.2004-032049 (2004).
9. Spritzer, P. M., Lecke, S. R., Satler, F. & Morsch, D. M. Adipose tissue dysfunction, adipokines, and low-grade chronic inflammation in polycystic ovary syndrome. Reproduction 149, R219–227, doi:10.1530/REP-14-0435 (2015).
10. Stepot, N. K. et al. Women with polycystic ovary syndrome have intrinsic insulin resistance on euglycaemic-hyperinsulinaemic clamp. Hum Reprod 28, 777–784, doi:10.1093/humrep/deo463 (2013).
11. Ozcan Dag, Z., Oguzturk, O., Isik, Y., Turkel, Y. & Bulcun, E. Personality profile in patients with polycystic ovary syndrome. Gynecol Endocrinol 31, 540–542, doi:10.3109/09513590.2015.1018162 (2015).
12. Ehrmann, D. A., Barnes, R. B. & Rosenfield, R. L. Polycystic-Ovary-Syndrome as a Form of Functional Ovarian Hyperandrogenism Due to Dysregulation of Androgen Secretion. Endocrine Reviews 16, 322–330, doi:10.1210/er.16.3.322 (1995).
13. Yong, E. L., Baird, D. T., Yates, R., Reichert, I. E. & Hillier, S. G. Hormonal-Regulation of the Growth and Steroidogenic Function of Human Granulosa-Cells. J Clin Endocrinol Metab 74, 842–849, doi:10.1210/jc.74.4.842 (1992).
14. Jakimiuk, A. J., Weitsman, S. R., Navah, A. & Magoffin, D. A. Luteinizing hormone receptor, steroidogenesis acute regulatory protein, and steroidogenic enzyme messenger ribonucleic acids are overexpressed in theca and granulosa cells from polycystic ovaries. J Clin Endocrinol Metab 86, 1318–1323, doi:10.1210/jc.86.3.1318 (2001).
15. Willis, D. S. et al. Premature response to luteinizing hormone granulosa cells from anovulatory women with polycystic ovary syndrome: Relevance to mechanism of anovulation. J Clin Endocrinol Metab 83, 3984–3991, doi:10.1210/jc.83.11.3984 (1998).
16. Kanamarlapudi, V., Gordon, U. D. & Berrall, A. L. Luteinizing hormone receptor, choric gonadotropin receptor, and adipsin in polycystic ovary syndrome (PCOS) patients are functionally active. Reproductive Biomedicine Online 32, 635–641, doi:10.1016/j.rbmo.2016.03.003 (2016).
17. Das, M. et al. Granulosa cell survival and proliferation are altered in polycystic ovary syndrome. J Clin Endocrinol Metab 93, 881–887, doi:10.1210/jc.2007-1650 (2008).
18. Magoffin, D. A. & Weitsman, S. R. Insulin-Like Growth Factor-I Regulation of Luteinizing Hormone (LH) Receptor Messenger-Ribonucleic Acid Expression and LH-Stimulated Signal-Transduction in Rat Ovarian Theca-Interstitial Cells. Biol Reprod 51, 766–775, doi:10.1095/biolreprod.199.4.766 (1994).
19. Udoff, L. & Adashi, E. Y. Polycystic Ovarian Disease - a New Look at an Old Subject. Curr Opin Obstet Gynecol 7, 340–343 (1995).
20. Pellat, J. et al. Granulosa cell production of anti-Mullerian hormone is increased in polycystic ovaries. J Clin Endocrinol Metab 92, 240–245, doi:10.1210/jc.2006-1582 (2007).
21. Homburg, R. et al. The relationship of serum anti-Mullerian hormone with polycystic ovarian morphology and polycystic ovary syndrome: a prospective cohort study. Hum Reprod 28, 1077–1083, doi:10.1093/humrep/det015 (2013).
22. Homburg, R. & Crawford, G. The role of AMH in anovulation associated with PCOS: a hypothesis. Hum Reprod 29, 1117–1121, doi:10.1093/humrep/deu076 (2014).
23. Lan, C. et al. Functional microarray analysis of differentially expressed genes in granulosa cells from women with polycystic ovary syndrome related to MAPK/ERK signaling. Sci Rep 5, doi:10.1038/srep14994 (2015).
24. Schmidt, J. et al. Differential expression of inflammation-related genes in the ovarian stroma and granulosa cells of PCOS women. Mol Hum Reprod 20, 49–58, doi:10.1093/molehr/gat051 (2014).
25. Chronowska, E. High-throughput analysis of ovarian granulosa cell transcriptome. Biomed Res Int 2014, 213570, doi:10.1155/2014/213570 (2014).
26. Nivet, A. L., Vigneault, C., Blondin, P. & Sirard, M. A. Changes in granulosa cells’ gene expression associated with increased oocyte competence in bovine. Reproduction 145, 555–565, doi:10.1530/REP-13-0032 (2013).
27. Martorini, A. & Gerard, N. Interleukin-1 (II-1) system gene expression in granulosa cells: kinetics during terminal preovulatory follicle maturation in the mare. Reprod Biol Endocrinol 1, 42, doi:10.1186/1477-8277-1-42 (2003).
28. Wang, F. et al. Alternative splicing of the androgen receptor in polycystic ovary syndrome. Proc Natl Acad Sci USA 112, 4743–4748, doi:10.1073/pnas.1418216112 (2015).
29. Chazenbalk, G. et al. Abnormal expression of genes involved in inflammation, lipid metabolism, and Wnt signaling in the adipose tissue of polycystic ovary syndrome. J Clin Endocrinol Metab 97, E765–770, doi:10.1210/jc.2011-2377 (2012).
30. de Resende, L. O. et al. Single-cell expression analysis of BMP15 and GDF9 in mature oocytes and BMP2R in cumulus cells of women with polycystic ovarian syndrome undergoing controlled ovarian hyperstimulation. J Assist Reprod Genet 29, 1057–1065, doi:10.1007/s10515-012-9825-8 (2012).

31. Lemieux, S., Lewis, G. F., Ben-Chetrit, A., Steiner, G. & Greenblatt, E. M. Correction of hyperandrogenemia by laparoscopic ovarian cauter in women with polycystic ovarian syndrome is not accompanied by improved insulin sensitivity or lipid-lipoprotein levels. J Clin Endocrinol Metab 84, 4278–4282, doi:10.1210/jc.84.11.4440 (1999).

32. Long, W. et al. Characterization of serum microRNAs profile of PCOS and identification of novel non-invasive biomarkers. Cell Physiol Biochem 33, 1304–1315, doi:10.1159/000358698 (2014).

33. Salilew-Wondim, D. et al. Polycystic ovarian syndrome is accompanied by repression of gene signatures associated with biosynthesis and metabolism of steroids, cholesterol and lipids. J Ovarian Res 8, 24, doi:10.1186/s13052-015-0151-5 (2015).

34. Wissing, M. L. et al. The transcriptome of corona radiata cells from individual McaU1, UkrainianCap1, Ukrainian oocytes that after ICSI developed to embryos selected for transfer: PCOS women compared to healthy women. J Ovarian Res 7, 110, doi:10.1186/1305-2775-7-110 (2014).

35. Zhang, W. X. et al. Selection of reference genes from Shiraia bambusicola for RT-qPCR analysis under different culturing conditions. BMC Genomics 13, 309, doi:10.1186/1471-2164-13-309 (2012).

36. Vandesompele, J. et al. Comparative abundance of quantification (qPCR) control genes. J Mol Endocrinol 27, 365–371, doi:10.1677/jme.0.02617 (2001).

37. Schmittgen, T. D. & Zakrajsek, B. A. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative PCR. Anal Biochem 341, 297–308, doi:10.1016/j.ab.2005.02.022 (2005).

38. Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 34, 549–558, doi:10.1677/jme.0.031849 (2010).

39. Patel, P., Boyd, C. A., Johnston, D. G. & Williamson, C. Analysis of GAPDH as a standard for gene expression quantification in human placenta. Placenta 23, 497–498 (2002).

40. Meller, M., Vadachkoria, S., Luthy, D. A. & Williams, M. A. Evaluation of housekeeping genes in placental comparative expression studies. Placenta 26, 601–607, doi:10.1016/j.placenta.2004.09.009 (2005).

41. Nestorov, J., Matic, G., Elakovic, I. & Tanic, N. Gene Expression Studies: How to Obtain Accurate and Reliable Data by Quantitative Real-Time RT PCR. J Med Biochem 34, 325–338, doi:10.2478/jm-2014-0001 (2013).

42. Dheda, K. et al. The implications of using an inappropriate reference gene for real-time reverse transcription PCR data normalization. Anal Biochem 344, 141–143, doi:10.1016/j.ab.2005.02.022 (2005).

43. Murphy, J. & Bustin, S. A. Reliability of real-time reverse-transcription PCR in clinical diagnostics: gold standard or substandard? Expert Rev Mol Diagn 9, 187–197, doi:10.1586/14737159.9.2.187 (2009).

44. Chen, D., Pan, X., Xiao, P., Farwell, M. A. & Zhang, B. Evaluation and identification of reliable reference genes for pharmacogenomics, toxicogenomics, and small RNA expression analysis. J Cell Physiol 226, 2469–2477, doi:10.1002/jcp.22723 (2011).

45. Zhang, X., Chen, D., Smith, M. A., Zhang, B. & Pan, X. Selection of reliable reference genes in Caenorhabditis elegans for analysis of nanotoxicity. PLoS One 7, e31849, doi:10.1371/journal.pone.0031849 (2012).

46. Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. J Mol Endocrinol 25, 169–193 (2000).

47. Nolan, T., Hands, R. E. & Bustin, S. A. Quantification of mRNA using real-time RT-PCR. Nat Protoc 1, 1559–1582, doi:10.1038/nprot.2008.236 (2006).

48. Schmittgen, T. D. & Zakrzewska, B. A. Effect of experimental treatment on housekeeping gene expression: validation by real-time, quantitative RT-PCR. J Biochem Biophys Methods 46, 69–81 (2000).

49. Suzuki, T., Higgins, P. J. & Crawford, D. R. Control selection for RNA quantitation. Biotechniques 29, 332–337 (2000).

50. Thellin, O. et al. Housekeeping genes as internal controls: use and limits. J Biotechnol 75, 291–295 (1999).

51. Tricario, G. et al. Quantitative real-time reverse transcription polymerase chain reaction: normalization to rRNA or single housekeeping genes is inappropriate for human tissue biopsies. Anal Biochem 309, 293–300 (2002).

52. Warrington, J. A., Nair, A., Mahadevappa, M. & Tygstrup, K. Comparison of human adult and fetal expression and identification of 535 housekeeping/maintenance genes. Physiol Genomics 2, 143–147 (2000).

53. Vandesompele, J. et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome biology 3, RESEARCH0034 (2002).

54. Sadek, K. H. et al. Variation in stability of housekeeping genes in endometrium of healthy and polycystic ovarian syndrome women. Hum Reprod 27, 251–256, doi:10.1093/humrep/der363 (2012).

55. Nestorov, D. V. M. D. M. I. B. A. J. B. M. M. N. G. M. J. Hypoxanthine Guanine Phosphoribosyltransferase Is the Most Stable Reference Gene for Gene Expression Analysis by Quantitative PCR in Peripheral Blood Mononuclear Cells from Women with the Polycystic Ovary Syndrome. J Med Biochem 33, 356–363 (2014).

56. Zhang, C., Li, T., Hou, C. L. & Shen, X. Y. Selection of reference genes from Shiraia bambusicola for RT-qPCR analysis under different culturing conditions. AMB Express 7, 14, doi:10.1186/s13568-016-0314-9 (2017).

57. Chen, Y. et al. Selection of reference genes for quantitative real-time PCR normalization in creeping bentgrass involved in four abiotic stresses. Plant Cell Rep 34, 1823–1834, doi:10.1007/s00299-015-1830-9 (2015).

58. Xu, L. et al. Validation of Reference Genes for Quantitative Real-Time PCR during Bicolor Tepal Development in Asiatic Hybrid Lilium (Lilium spp.). Front Plant Sci 8, 609, doi:10.3389/fpls.2017.00669 (2017).

59. Zhang, W. X. et al. Selection of Suitable Reference Genes for Quantitative Real-Time PCR Normalization in Three Types of Rat Adipose Tissue. Int J Mol Sci 17, doi:10.3390/ijms17060968 (2016).

60. Wang, H., Zhang, X., Liu, Q., Liu, X. & Ding, S. Selection and evaluation of new reference genes for RT-qPCR analysis in Epinephelus akaara based on transcriptome data. PLoS One 12, e0171646, doi:10.1371/journal.pone.0171646 (2017).

61. Robledo, D. et al. Analysis of qPCR reference gene stability determination methods and a practical approach for efficiency calculation on a turbob (Scophthalmus maximus) gonad dataset. BMC Genomics 15, 648, doi:10.1186/1471-2164-15-648 (2014).

62. van de Moosdijk, A. A. & van Amerongen, R. Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland. Sci Rep 6, 35595, doi:10.1038/srep35595 (2016).

63. Rydbirk, R. et al. Assessment of brain reference genes for RT-qPCR studies in neurodegenerative diseases. Sci Rep 6, 37116, doi:10.1038/srep37116 (2016).

64. Gong, H. et al. Evaluation of candidate reference genes for RT-qPCR studies in three metabolism related tissues of mice after caloric restriction. Sci Rep 6, 38513, doi:10.1038/srep38513 (2016).
76. Mo, F., Zhao, J., Liu, N., Cao, L. H. & Jiang, S. X. Validation of reference genes for RT-qPCR analysis of CYP4T expression in crucian carp. *Genet Mol Biol* **37**, 500–507 (2014).

77. Chang, J. et al. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome. *Fertility and Sterility* **81**, 19–25, doi:10.1016/j.fertnstert.2003.10.004 (2004).

78. Fauser, B. C. J. M. et al. Revised 2003 consensus on diagnostic criteria and long-term health risks related to polycystic ovary syndrome (PCOS). *Human Reproduction* **19**, 41–47, doi:10.1093/humrep/deh098 (2004).

79. Andersen, C. L., Jensen, J. L. & Orntoft, T. F. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res* **64**, 5245–5250, doi:10.1158/0008-5472.CAN-04-0496 (2004).

Acknowledgements
This work was supported in part by funds from the One-off Funding of CUHK for Joint Lab/Research Collaboration provided to CUHK-Shandong University Joint Laboratory on Reproductive Genetics (Project 3132970) and to the Hong Kong Branch of the Chinese National Center for Assisted Reproductive Technology and Reproductive Genetics (Project 3132987), from the Hospital for Reproductive Medicine, Shandong University provided to CUHK–Shandong University Joint Laboratory on Reproductive Genetics (Project 7104693), and from the Young Scholars Program of Shandong University (2016WLJH50).

Author Contributions
L.V.Y., Z.H.A.O., S.G., L.U.G., L.E.U.N.G., C.K., X.I.O.N., Z.Q., S.U., X.W., M.A., J.L., C.H.A.N., W.Y., and L.I.U., H.B. participated in the literature search, experimental work, data mining, data analysis and interpretation, writing the manuscript, and creating the figures and tables. L.U.G. supervised the study design and critically read and edited the manuscript. All authors approved the final version for publication.

Additional Information

Competing Interests: The authors declare that they have no competing interests.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.