



OPEN

Selection of Endogenous Control Reference Genes for Studies on Type 1 or Type 2 Endometrial Cancer

Thangesweran Ayakannu^{1,2,3}✉, Anthony H. Taylor^{1,4}  & Justin C. Konje^{1,5}

A panel of 32 candidate reference genes was used to identify the most stable genes for gene normalisation in quantitative RT-PCR studies using endometrial biopsies obtained from women with endometrial cancer (type 1 or type 2) and without cancer (controls). RNA from the biopsies was isolated, examined for purity and quality, and then reverse transcribed into cDNA before being subjected to real-time qRT-PCR analysis in triplicate within the TaqMan gene Expression Assay kit. The most 'stable' endogenous control genes were then identified using the geNorm qbase + 2 and NormFinder software packages. PSMC4, PUM1 and IPO8 were identified as the best reference genes combination for type 1 endometrial cancer (grades 1, 2 and 3), whereas for type 2 endometrial cancer (serous and carcinosarcoma), UBC, MRPL19, PGK1 and PPIA were the best reference genes combination. We conclude that the use of these normaliser combinations should provide accurate interpretation of gene expression at the transcript level in endometrial cancer studies especially for types 1 and 2 cancers.

The use of biomarkers is the cornerstone of effective precision medicine^{1,2}. RNA forms an excellent source of biomarkers to enable early disease detection, assessment of prognosis, monitoring patient response to therapy or selecting the right treatment for the right patient (personalised medicine). The gold standard method for studying biomarkers at the RNA level in the past decade has been measurement of microRNA in plasma³ and in the past two decades measurement of messenger RNA (mRNA) in tissue biopsies^{4,5}. The measurement of these moieties has occurred primarily through quantitative real time PCR (qRT-PCR) with microRNA- and gene-specific primers, respectively. Application of qRT-PCR to the study of transcript levels in many disease processes has increasingly replaced northern blotting because it is easy to use, fast, reproducible, highly sensitive, specific, and provides high sample throughput^{6–9}. In particular, it has been used to identify and assess several molecular markers associated with the staging¹⁰, initiation¹¹, progression¹², and metastatic potential¹³ of endometrial cancer and recently in patient prognosis^{14,15}.

Elimination of sources of error in the qRT-PCR technique, such as differences in the quantity and quality of extracted mRNA, the presence of contaminating genomic or operator DNA, divergences in reverse transcription and PCR efficiencies must occur for the qRT-PCR to be valid^{16–18}. To ensure uniformity and reproducibility of published data, the “Minimum Information for publication of Quantitative real time PCR Experiments” (MIQE) guidelines suggest that the choice of and number of reference genes should be an essential part of all qRT-PCR studies. Validation of this important step guarantees normalisation of resulting data¹⁹. This normalisation step, where an endogenous reference (housekeeping) gene compensates for any variations in sample or experimental conditions, is essential because all genes under test (both the genes of interest and the reference gene) are assessed under the same experimental conditions. A unified, single “best” reference gene is therefore unlikely to be found routinely, because almost all genes are modified under some conditions²⁰.

¹Faculty of Health and Life Sciences, University of Liverpool, Thompson Yates Brownlow Hill, Liverpool, L69 3GB, UK. ²Reproductive Sciences Section, Department of Cancer Studies and Molecular Medicine, University of Leicester, Leicester, UK. ³Gynaecology Oncology Cancer Centre, Liverpool Women's NHS Foundation Trust, Liverpool Women's Hospital, Liverpool, UK. ⁴Department of Molecular and Cell Biology, University of Leicester, Leicester, Leicestershire, UK. ⁵Department of Obstetrics and Gynaecology, Sidra Medicine, Women's Wellness and Research Center, HMC, Doha, Qatar. ✉e-mail: t.ayakannu@nhs.net

In earlier qRT-PCR studies of endometrial cancer, normalisation was done using reference genes, such as β -actin (ACTB)^{21,22}, glyceraldehyde-3-phosphate dehydrogenase (GAPDH)^{23,24}, or 18S RNA ribosomal unit 1 (18S rRNA)²⁵, chosen without rigorous testing. Previously, we demonstrated that all of the aforementioned genes are poor reference genes for the study of gene expression changes that occur in the two main types of endometrial cancer⁵. In our previous studies, we showed that a combination of 3 genes (IPO8, MRPL19 and PPIA) provided the best combination of normalisation factors in qRT-PCR studies of endometrial cancer and that geNorm qbase+²⁶ is more robust than the other software packages in its statistical corrections for all the possible sources of experimental error listed above. Although this conclusion remains applicable and supportable when comparing mRNA levels in type 1 endometrial cancer with type 2 endometrial cancer in the same gene expression study, recent studies have identified additional targeted biomarkers for individualised endometrial cancer. In those studies, the authors advocate the linking of transcript studies directly to personalised treatments, without indicating what the correcting reference genes for such studies should be¹³. This is especially important because endometrial cancer is becoming more prevalent in the reproductive age woman¹⁰, a point that has been missed in those recommendations^{1,12}.

The aim of this study was therefore to identify the best reference gene combinations for the normalisation of studies, when using qRT-PCR, to quantify the number of gene transcripts in endometrial tissues from normal women with either endometrial cancer type 1 or type 2. By doing so, we hope to provide investigators with the tools to effectively investigate their chosen RNA biomarkers in future studies of women with type 1 or type 2 endometrial cancer so as to provide meaningful clinical data, especially when only one of these patient groups is under consideration. If both groups are under consideration, then our previously published data remains the correct choice⁴.

Results

GeNorm analyses for type 1 EC and for type 2 EC. Using geNorm, the least stable to most stable reference genes evaluated for type 1 EC was found to be: RPL30 > MT-ATP6 > 18S > ACTB > TBP > RPLP0 > PES1 > POLR2A > TFRC > HPRT1 > ABL1 > GADD45A > HMBS > CDKN1A > RPL37A > UBC > GAPDH > CDKN1B > CASC3 > POP4 > PGK1 > GUSB > YWHAZ > PPIA > RPS17 > MRPL19 > B2M > EIF2B1 > ELF1 > PSMC4 > PUM1 > IPO8 (Fig. 1); (gene names, identities, accession numbers and amplicon sizes of the PCR products can be found in Supplemental Table 1). Analysis of this order indicated that PUM1 and IPO8 were the two most stable genes (defined as $M \leq 1.0$) in type 1 EC samples, with the two least stable genes being RPL30 and MT-ATP6. The commonly used reference genes β -actin (ACTB), GAPDH and 18S were outside the least stable ($M \geq 1.0$) category with M-values of 1.355, 0.950 and 1.300, respectively.

Gene stability analysis using geNorm (Fig. 2) indicated that the optimal number of reference gene targets was 5 in the EC1 analyses (geNorm $V < 0.15$ when comparing a normalisation factor based on the 5 or 6 most stable targets). Thus, geNorm qbase + 2 predicts that the optimal normalisation factor would be the geometric mean of the reference targets PSMC4, PUM1 and IPO8 (based on 3 most stable genes), ELF1, PSMC4, PUM1 AND IPO8 (based on 4 most stable genes), or EIF2B1, ELF1, PSMC4, PUM1 AND IPO8 (based on 5 most stable genes).

Similar analyses of samples from patients with type 2 EC demonstrated the least stable to most stable genes to be: 18S > RPL30 > ACTB > MT-ATP6 > TBP > CDKN1A > PES1 > POLR2A > HMBS > GADD45A > B2M > GAPDH > RPLP0 > TFRC > HPRT1 > RPL37A > GUSB > CDKN1B > ABL1 > RPS17 > PSMC4 > CASC3 > EIF2B1 > POP4 > ELF1 > IPO8 > PUM1 > YWHAZ > UBC > MRPL19 > PGK1 > PPIA (Fig. 1). In studies using normal and type 2 cancer endometria, PGK1 and PPIA would be the two most stable genes with 18S and RPL30 being the two least stable. The commonly used reference genes (β -actin (ACTB), GAPDH and 18S) were also outside the least stable category with M-values of 1.450, 1.205 and 1.600, respectively.

The optimal number of reference targets for type 2 EC, was also 5 genes (Fig. 2). Thus, geNorm^{PLUS} with qbase+2 predicts that optimal normalisation factor would be achieved by using the geometric mean of the reference targets MRPL19, PGK1 and PPIA (based on 3 most stable genes), UBC, MRPL19, PGK1 and PPIA (based on 4 most stable genes), or YWAZ, UBC, MRPL19, PGK1 and PPIA (based on 5 most stable genes).

NormFinder analyses for type 1 EC and for type 2 EC. NormFinder software, which uses a mathematical model that considers both intergroup and intragroup expression variations (stability), and ranks them in order from the lowest to highest stability value²⁷, identified PSMC4 (proteasome (prosome, Macropain) 26S subunit, ATPase, 4) as the single most stable gene (stability value = 0.268). The software specified that the best two gene combination was IPO8 (importin 8) and MRPL19 (mitochondrial ribosomal protein L19) with a stability value of 0.224 for patients with type 1 EC (Fig. 3). By contrast, the single most stable gene for type 2 EC patients was MRPL19 (mitochondrial ribosomal protein L19; stability value of 0.379) and the best combination of two genes was ELF1 (E74-like factor 1) and PUM1 (Pumilio homolog 1) with a stability value of 0.259 (Fig. 3).

Discussion

In this first systematic study, we show which combination of reference genes from a panel of 32 endogenous control reference genes should be used for normalisation in quantitative real-time PCR (qRT-PCR) studies of types 1 and 2 EC. We have identified 2 sets (one for type 1 and another set for type 2) of five 'housekeeping' genes that provide the best combination. When the TaqMan gene Expression Assay kit was used, none of the genes in the two identified combinations was common to both types of EC. When the stability list was increased to six normalising genes, one gene (PUM1) was found to be common to both sets of samples and interestingly one gene (IPO8) in the type 1 EC group and two genes (PPIA and MRPL19) in the type 2 EC group were identical to those we identified in our previous report⁵ where only 3 normalising genes were used and in other cancers where 4 normalising genes were used²⁸. Several genes (B2M, EIF2B1, ELF1 and PSMC4; type 1 EC, and YWAZ, UBC, and PGK1; type 2 EC) were not identified in the combinations of five best genes for the two cancer sub-types (this

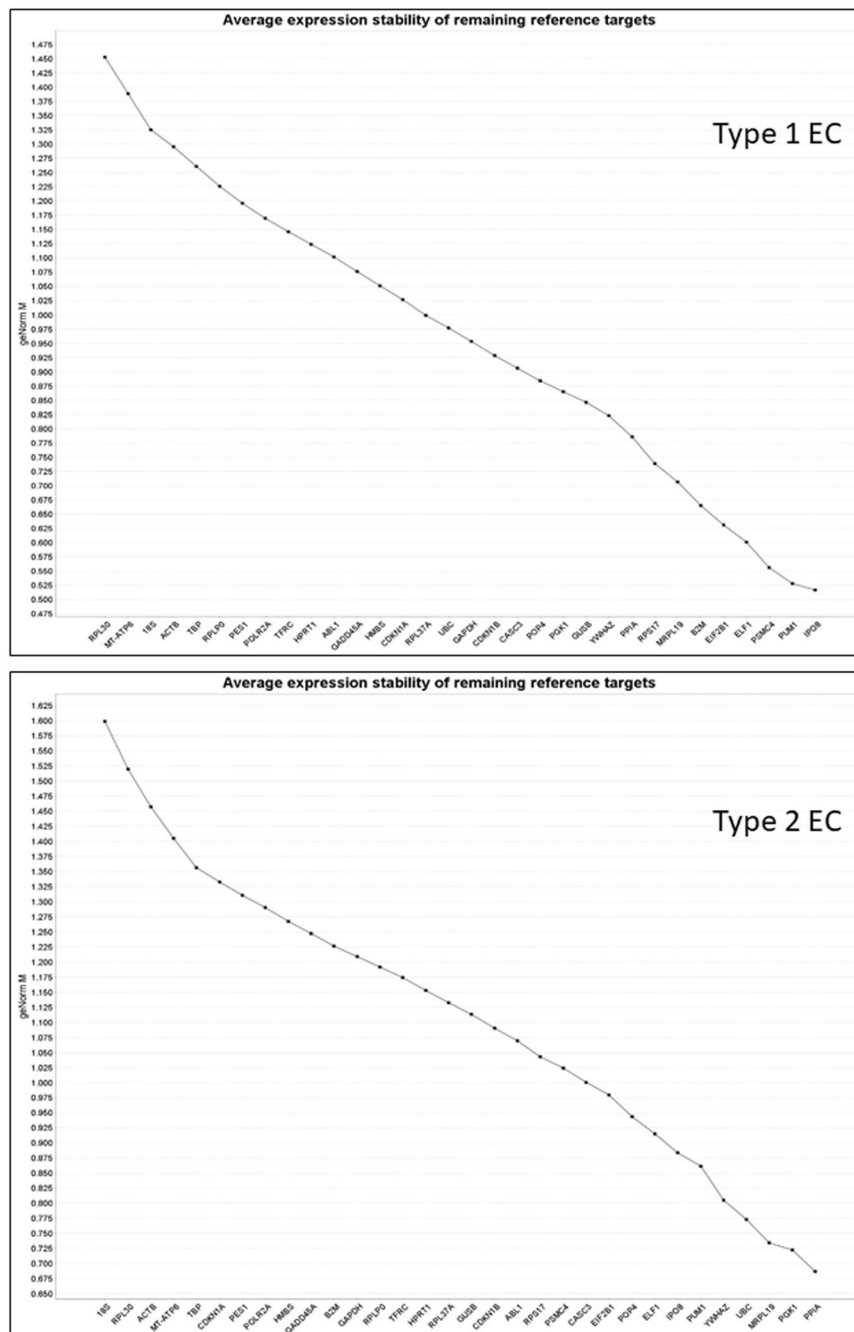


Figure 1. GeNorm analyses of the reference genes. Average expression stability values (geNorm M) of the 32 reference genes are ranked from least to most stable expression (left to right) for type 1 EC data (upper panel) and type 2 EC data (lower panel). Data are presented according to the output file obtained from qbase + 2 software. An M value of ≤ 1.0 indicates more stable gene expression.

study) suggesting that the transcriptome of these two EC sub-types may in fact be very different. Indeed, type 1 EC is phenotypically distinct from type 2 EC both in its biology and response to treatment, further supporting our hypothesis that the normalising genes for these two subtypes of EC would also be different; qRT-PCR studies must therefore take this into consideration in order to generate meaningful and clinically relevant data. The data in this new analysis suggest that the best normalisers for qRT-PCR studies that are confined to either type 1 EC or type 2 EC alone are different hence combining these together when studying either in isolation is likely to generate erroneous data.

These findings have important implications for future study design, because qRT-PCR studies that only focus on type 1 EC (with a control group) require a different combination of normalising genes to a study that focusses on only type 2 EC (also with a control group). If the study design utilises type 1 and type 2 EC samples (along with a control of normal tissues), then an additional and different set of normalising genes will be required to provide

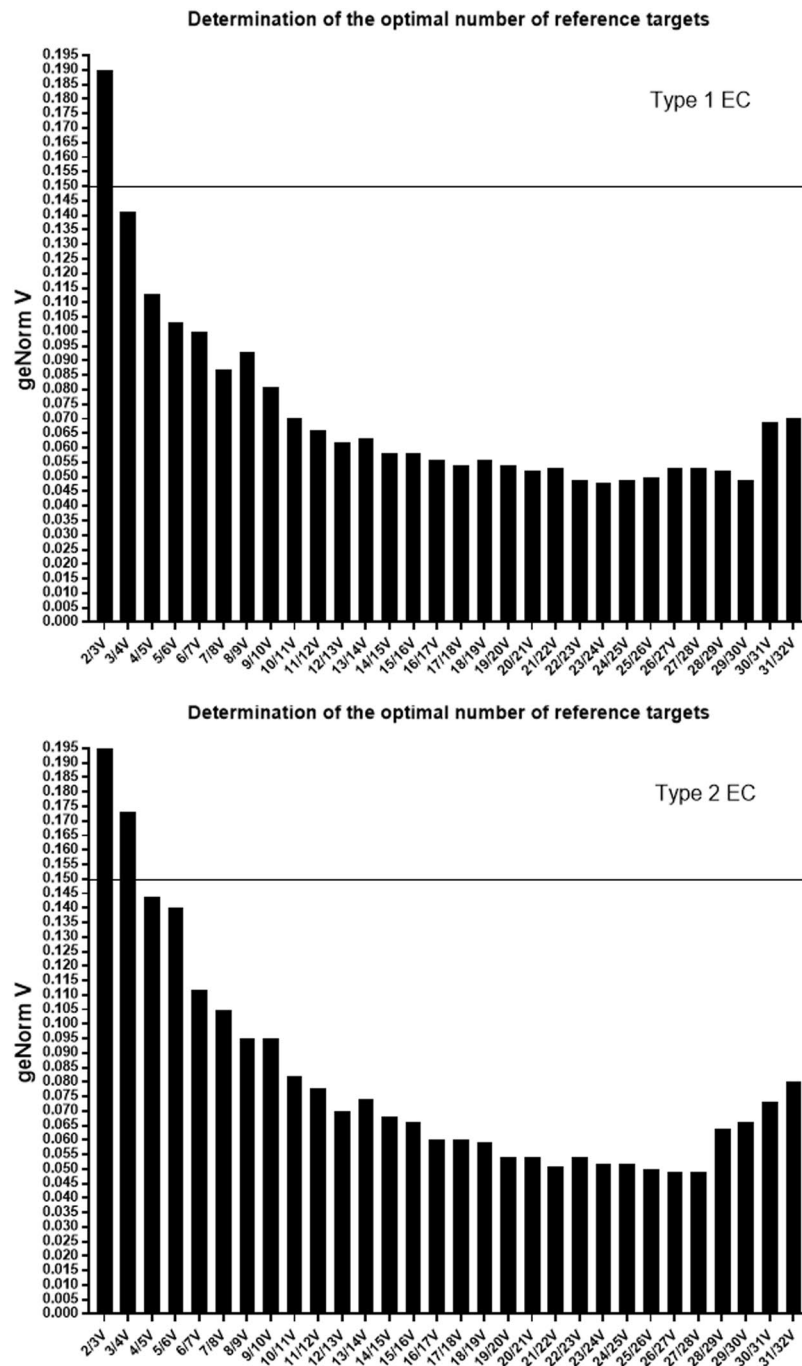


Figure 2. Determination of the optimal number of reference genes for normalisation. geNorm V scores, using the pair-wise variation (V) analyses within the geNorm^{PLUS} analysis software for the 32 reference genes for the incremental combination of V2/3 to V31/32 genes, are shown. The upper panel shows the data output file obtained for control and type 1 EC samples and the lower panel equivalent output data for control and type 2 EC samples. The horizontal line indicates the position of the geNorm threshold for stability ($V = 0.15$).

valid and important data, as previously described⁵. For example, in a study of only type 1 EC samples, the list of normalising genes required for qRT-PCR studies are those listed in the results section of this paper, whilst a study of all types of EC requires the list of genes provided previously⁵, i.e. PPIA, MRPL19 and IPO8. The 3 normalising genes we reported previously was not significantly improved upon by increasing the list to 4, 5 or 6 normalising genes⁵ and so this extra expense is not required, nor warranted.

Using NormFinder to analyse the best combination of normalising genes and the best gene for each patient group revealed that two genes previously reported to be the best normalisers for studies that included all types of EC (i.e. IPO8 and MRPL19)⁵ were also identified in the present study. Furthermore, MRPL19 appears to be the only common gene in the combination of six that is most stable in studies separately investigating type 1 EC and

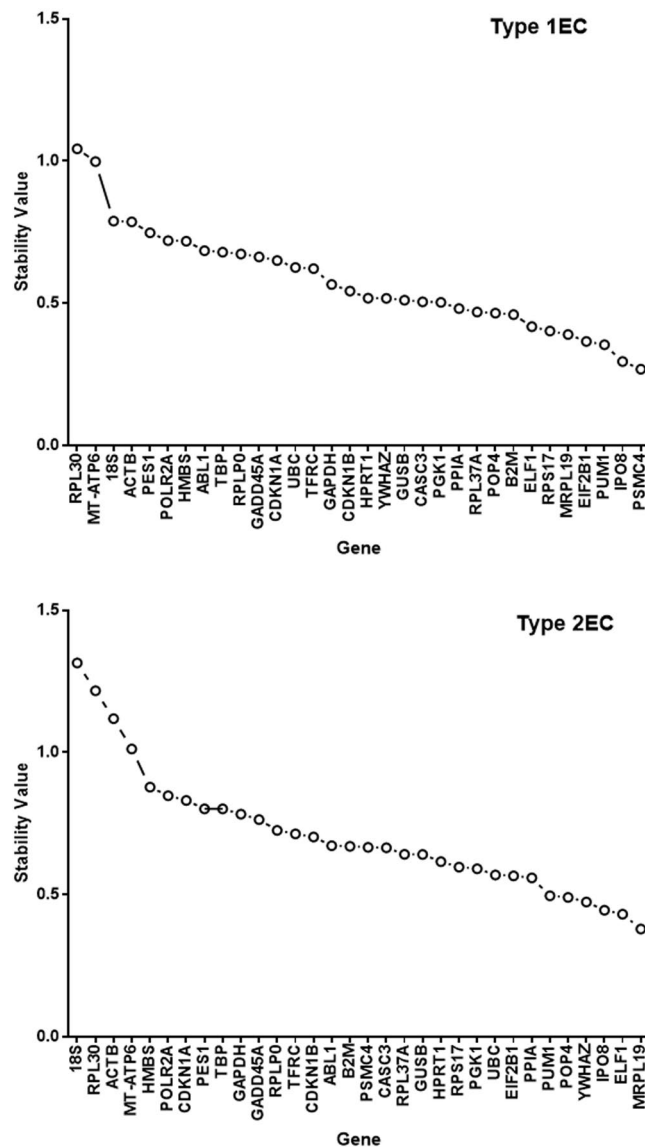


Figure 3. Normfinder analyses of the reference genes. The stability values for each of the 32 ‘housekeeping’ genes are shown, with the least stable on the left and the most stable on the right. A stability value < 0.5 is sufficient to predict either a stable gene in the type 1 EC cohort (upper panel) or the type 2 EC cohort (lower panel).

type 2 EC (as in this study), and indeed is one of the genes we have previously recommended when both types of EC were combined⁵.

Previous studies focussing on the best normalising genes for studies of patients with endometrial cancer have a limitation in that they evaluated only ten possible normalising genes^{29,30}. In those studies, the best normalising gene lists identified genes that were identified here and also previously (PPIA) in type 2 EC, but not those identified in type 1 EC. An additional issue in the stated studies was the large number of samples being examined that was not balanced by an equivalent number of control samples³⁰. Increasing the number of biological sample replicates minimises, any internal variations. This could result in erroneous conclusions based on those internal limitations. By limiting our studies to the minimum number of biological replicates, we have maximally increased the biological and experimental variations. This means that the software needs to perform in a robust manner to generate the best normalisers for the study populations. Additionally, by increasing the number of ‘housekeeping genes’ from ten to 32 genes, we have increased the probability of identifying extra most stable normalisers for qRT-PCR studies, not identified previously. Consequently, we also reported the minimum number of genes ($n = 3$) for each type of endometrial cancer and the maximum number of genes ($n = 6$) for absolute robustness. The decision on whether to choose 3, 4, 5 or 6 genes in combination as normalisers for qRT-PCR studies, is dependent on two factors. These factors are: (1) the magnitude of change in transcript levels (more subtle changes need a larger number of genes to robustly identify them) and (2) cost (the cost of six primer pairs is obviously

more than that of 3, 4 or 5). These considerations are important because an unwise choice may invalidate the study undertaken.

The use of the correct endogenous control reference gene(s) for normalisation in qRT-PCR experiments has been the subject of strong debate³¹; with advantages and disadvantages¹⁷, even with studies that follow MIQE guidelines¹⁹ being highlighted. Nevertheless, studies have shown disparities in the stability of reference genes in many different tissues³² and in the same tissues under different conditions³³. It is thus imperative that the appropriate selection of stable reference genes relative to the experiment undertaken is made, as has been discussed previously^{5,30}. In this regard, improved normalisation is possible when changing from one to multiple endogenous control reference genes, a single reference is unlikely to provide an ideal endogenous control, as is stated in the MIQE guidelines. It is for this reason that we chose to undertake this study, especially as advocated personalised medicine studies do not define the subject samples or the experimental conditions when suggesting new biomarkers for EC. We also deliberately chose samples from a cross section of the different groups that truly represent the patient population. By using both pre- and post-menopausal samples and also limiting our study to 3 patient samples from each group to provide nine control samples, nine type 1 EC samples and six type 2 EC samples that were clearly defined by an experienced gynaecology histopathologist, we maximised the variation within and between the sample groups. This manipulation results in the most robust study where variation is being minimised, whereas a much larger number of samples (e.g. in the tens or hundreds) would have hidden the variation that is needed by the software being used.

It is for this reason that we recommended geNorm qbase + 2 software for the analysis of normalisation studies where a given combination of reference genes are used to generate a normalisation factor and that the data are further complemented with analysis of the same data using NormFinder^{34,35}. By providing the best possible combination of normaliser genes, a platform whereby the biology of type 1 EC or type 2 EC can effectively be studied at the transcript level is provided. Additionally, in patient-specific prognosis studies where treatment outcomes are assessed, the observations reported herein are instantly applicable, since it provides a good starting point for normalising gene identification when treatments are applied. Thus, in gene expression studies using normal and type 1 malignant endometria, and where limited amounts of material or resources are available, more reliable normalisation is achieved when using the geometric mean of the Ct values obtained from the combination of three genes of PSMC4, PUM1 and IPO8 and is thus recommended. Similarly, for type 2 EC studies the geometric mean of the Ct values derived from the combination of MRPL19, PGK1 and PPIA provides a reliable normalisation factor. For absolute robustness, we recommend the use of geometric means from the 5 genes EIF2B1, ELF1, PSMC4, PUM1 AND IPO8 for type 1 EC studies and the use of geometric means from the 5 genes YWAZ, UBC, MRPL19, PGK1 and PPIA for type 2 EC studies, especially if the changes in gene expression when compared to controls are relatively subtle in contrast to other highly expressed gene targets. To compare gene expression patterns between type 1 EC and type 2 EC samples (without any control tissues as a reference), researchers are advised to follow the protocols described in this and our previous publication⁵, to generate the best normalising genes for their own patient cohorts.

In summary, by using a panel of 32 optimised and validated endogenous control reference genes in a Taqman gene expression assay format, we identified the most robust endogenous control reference genes for the study of either type 1 EC or type 2 EC. In doing so, we can categorically remove traditionally used normalising genes for the study of either type 1 EC or type 2 EC by qRT-PCR from the database, since the array used included representative genes from different gene families and functional classes. By reporting these data, we hoped to have provided a valuable tool for use in future studies of RNA biomarkers in the biology of type 1 and type 2 EC.

Materials and Methods

All volunteers provided signed, written informed consent to partake in the study (see Ethics Statement section). The Leicestershire, Northamptonshire and Rutland Research Ethics Committee (ref 06/Q2501/49) approved the study. Women undergoing hysterectomy for endometrial carcinoma or a benign gynaecological condition at the University Hospitals of Leicester National Health Service Trust were recruited (Table 1).

Histological diagnosis of the cancer was based on the FIGO system³⁶ and 15 endometrial carcinoma samples were studied: type 1 grade 1 endometrioid adenocarcinoma (n = 3), type 1 grade 2 endometrioid adenocarcinoma (n = 3) and type 1 grade 3 endometrioid adenocarcinoma (n = 3), type 2 serous (n = 3) and carcinosarcoma (n = 3). All cancer tissues was classified as being FIGO Stage 1. Normal endometrial tissue samples were obtained from volunteers who were undergoing a hysterectomy for benign indications (prolapse, dysfunctional uterine bleeding, fibroids) and were classified into pre-menopausal (secretory (n = 3) or proliferative (n = 3) phase) and postmenopausal demonstrating histological atrophic endometria (n = 3) by the criteria of Noyes *et al.*³⁷. Patient characteristics are shown in Table 1.

The type 1 EC patients (72.8 ± 11.4 years; Mean ± SD) were significantly older (p < 0.001; one way ANOVA with Dunnett's multiple comparison test) than either the controls (51.9 ± 8.5 years) or those with type 2 EC (54.5 ± 6.2 years). Both type 1 (32.4 ± 5.1 kg/m²) and type 2 EC (37.2 ± 4.4 kg/m²) patients had significantly higher BMIs (p < 0.01; one way ANOVA with Dunnett's multiple comparison test) when compared to the controls (26.2 ± 3.4 kg/m²), as has been reported previously³⁸ (Table 1).

Preparation of total cellular RNA and cDNA synthesis. From this point forward, the methodologies are essentially the same as described^{4,5}. Fresh uteri were transported on ice to the histopathology department and two adjacent tissue biopsy samples dissected out by an experienced consultant gynaecology histopathologist using a dissecting microscope; one sample was fixed in 10% formal saline, stained with haematoxylin and eosin and used for histological confirmation of the diagnosis. All cancer samples were stage 1 and all control tissues were devoid of myometrial contamination. The second sample used for this normalising gene study was washed

Control Patients	Age (yrs)	Mean \pm SD	p-value [†]	Mean \pm SD for Control Patients	p-value ^{††}	BMI (kg/m ²)	Mean \pm SD	p-value [†]	Mean \pm SD for Control Patients	p-value ^{††}
Atrophic (1)	65					34				
Atrophic (2)	65	62.33 \pm 4.62	n.a.			22	26.67 \pm 6.43	n.a.		
Atrophic (3)	57					24				
Secretory Phase (1)	44					25				
Secretory Phase (2)	51	46.00 \pm 4.36	0.025	51.89 \pm 5.477	n.a.	27	26.00 \pm 1.00	n.s.	26.22 \pm 3.38	n.a.
Secretory Phase (3)	43					26				
Proliferative Phase (1)	47					27				
Proliferative Phase (2)	48	47.33 \pm 0.58	0.037			27	26.00 \pm 1.73	n.s.		
Proliferative Phase (3)	47					24				
Cancer Patients (all Stage 1)	Age (yrs)	Mean \pm SD	p-value [†]	Mean \pm SD for Cancer Patients	p-value ^{††}	BMI (kg/m ²)	Mean \pm SD	p-value [†]	Mean \pm SD for Cancer Patients	p-value ^{††}
EC1 Grade 1 (1)	63					33				
EC1 Grade 1 (2)	88	78.00 \pm 13.23	0.030			24	31.33 \pm 6.66	n.s.		
EC1 Grade 1 (3)	83					37				
EC1 Grade 2 (1)	78					29				
EC1 Grade 2 (2)	69	67.67 \pm 11.06	0.936	72.78 \pm 11.43	0.0002	32	30.67 \pm 1.53	n.s.	32.44 \pm 5.08	0.012
EC1 Grade 2 (3)	56					31				
EC1 Grade 3 (1)	74					42				
EC1 Grade 3 (2)	60	72.67 \pm 12.06	0.486			30	35.33 \pm 6.11	0.038		
EC1 Grade 3 (3)	84					34				
EC2 Serous (1)	61					40				
EC2 Serous (2)	55	59.00 \pm 3.46	0.994			35	37.67 \pm 2.52	0.011		
EC2 Serous (3)	61					38				
EC2 Carcinosarcoma (1)	50			54.50 \pm 6.25	0.817	34			37.17 \pm 4.40	0.0002
EC2 Carcinosarcoma (2)	45	50.00 \pm 5.00	0.317			44	36.67 \pm 6.43	0.019		
EC2 Carcinosarcoma (3)	55					32				

Table 1. Patient characteristics. [†]P-value compared to atrophic patient group; ^{††}P-value compared to control group; One-way ANOVA with Dunnett's multiple comparisons t-test. Significant differences are in bold font. EC1 = endometrial carcinoma type 1; EC2 = endometrial carcinoma type 2; n.a. = not applicable; n.s. = not significantly different; S.D. = standard deviation.

with phosphate buffered saline (PBS) to remove excess blood and stored in RNAlater (Life Technologies, Paisley, UK) for 24 hours at room temperature before transfer to -80°C for storage and further processing.

Frozen, endometrial tissues (~ 100 mg) in lysis/binding buffer (1 ml lysis/binding buffer solution per 100 mg of tissues (miRNA Isolation Kit) were homogenised using a TissueRuptor (Qiagen Crawley, UK) homogeniser at medium speed for 60 seconds on ice until all visible 'clumps' were dispersed. Total RNA was extracted using the mirVana[™] miRNA Isolation Kit (Life Technologies, Paisley, UK) according to the manufacturer's protocol. Total RNA was then quantified and its purity determined using a NanoDrop 2000c spectrophotometer (Thermo Scientific, Detroit USA). At this point, the RNA concentration was standardised to $10\mu\text{g}/100\mu\text{l}$, and contaminating genomic DNA digested by treating with a TURBO-DNase free kit (Life Technologies, Paisley, UK) at 37°C for 30 minutes. The reaction was inactivated with $10\mu\text{l}$ of inactivation buffer and the solution centrifuged for 90 seconds at $10\,000 \times g$. The purity of the extracted total RNA (supernatant), as measured with the Nanodrop spectrophotometer, indicated good quality RNA with a A_{260}/A_{280} ratio of 2.10 ± 0.31 (OD ratio \pm SD) and a A_{260}/A_{230} ratio of 2.19 ± 0.43 . The average nucleic acid yield of RNA after extraction was 1.17 ± 0.61 ($\mu\text{g}/\mu\text{l} \pm$ SD). The treated supernatants were subjected to first strand synthesis using the high capacity cDNA MultiScribe[™] Reverse Transcriptase Kit ((Life Technologies, Paisley, UK) according to the manufacturer's protocol with $1\mu\text{g}$ of RNA, oligodT16 primers and recombinant MMLV-RT enzyme; incubation at 25°C for 10 minutes, 37°C for 120 minutes, 85°C for 5 minutes and then cooled to 4°C . The resultant cDNA was stored at -20°C .

qRT-PCR studies. Quantitative Real-Time PCR experiments were performed in triplicate using optimised, validated human endogenous control assay TaqMan Array 96-Well Plates (Applied Biosystems by Life Technologies, Paisley, UK) that contain 32 reference genes (see Supplemental Table 1). TaqMan Gene Expression Assays were used (Applied Biosystems) and each consisted of a fluorogenic FAM dye-labelled MGB probe (final concentration 250 nM) and two amplification primers (forward and reverse; final concentration 900 nM) provided in a pre-formulated 20X mix. Each assay had an amplification efficiency of $100 \pm 10\%$, calculated by the system software. RT-minus and no template controls (NTC) containing DNase-free water instead of template mRNA were included in each run and no product was synthesised in the NTC and RT-minus reactions confirming the absence of contamination with exogenous DNA. The final reaction volume was $20\mu\text{l}$ and consisted of $2\mu\text{l}$ of cDNA, $8\mu\text{l}$ of DNase-free water and $10\mu\text{l}$ of TaqMan universal PCR Master Mix. A StepOne Plus instrument

(Applied Biosystems by Life Technologies, Paisley, UK) was used for the PCR with their proprietary software (StepOnePlus software version 2.3) automatically determining the Ct as being 0.5 standard deviations above baseline fluorescence. The cycle profile was 2 minutes at 50 °C, 10 minutes at 95 °C, 40 cycles of 15 seconds at 95 °C and 1 minute at 60 °C.

Determination of the most stable reference genes. Gene stability was evaluated using geNorm^{PLUS} software v2.2 (Biogazelle, Zwijnaarde, Belgium), which incorporates the updated version of qbase + 2²⁶ (available from <https://www.biogazelle.com/qbaseplus>) and NormFinder version 0.953²⁷ (available from Aarhus University, Denmark; <http://moma.dk/normfinder-software>). Accumulated standard deviation data obtained from the NormFinder algorithm were analysed using GenEx software version 5.3.6.170 (MultiD Analyses AB, Goteborg, Sweden) (<http://www.multid.se/contact.php>).

Briefly, mRNA gene expression stability analysis for the 32 endogenous reference control genes was obtained using the mean qRT-PCR threshold cycle (Ct) value, defined as the number of cycles required for the fluorescent signal to cross the threshold (i.e. 0.5 standard deviations above background levels). The geNorm qbase+2 algorithm determined the medium reference target stability measure (M), as the average pair-wise variation of each reference gene in relation to all the other reference genes thus enabling the elimination of the least stable gene. This was followed by recalculation of the M values resulting in a ranking of the most stable genes, i.e. the lower the M value, the higher the gene stability. The software indicates that a stable reference gene should have an average geNorm M value ≤ 1.0 in a heterogeneous set of samples²⁶.

The in-built geNorm qbase + 2 ranking algorithm was used to calculate the minimum number of genes required for a reliable normalisation factor, using the pair-wise variation V_n/V_{n+1} between two sequential normalisation factors (NF_n and NF_{n+1}). In addition, using the in-built optimal cut off value ($V = 0.15$) for pairwise variation, below which the inclusion of an additional reference genes is unnecessary, the minimum number of reference gene for type 1 EC studies or type 2 EC studies was calculated²⁶.

Details of the statistical methods used by both qbase + 2 and NormFinder are described elsewhere^{39–41}.

Ethics statement. All subjects gave written informed consent in accordance with the Declaration of Helsinki. The study and the protocol adopted was approved by the Leicestershire, Northamptonshire and Rutland Research Ethics Committee (LREC Reference Number: 06/Q2501/49) and performed using guidelines provided within the National Health Service Research Governance Framework.

Received: 12 March 2019; Accepted: 2 April 2020;

Published online: 21 May 2020

References

- Jha, A. *et al.* Towards precision medicine: discovering novel gynecological cancer biomarkers and pathways using linked data. *J Biomed Semantics* **8**, 40, <https://doi.org/10.1186/s13326-017-0146-9> (2017).
- Xi, X. *et al.* RNA Biomarkers: Frontiers of precision medicine for cancer. *Noncoding RNA* **3**, <https://doi.org/10.3390/ncrna3010009> (2017).
- Torres, A., Torres, K., Wdowiak, P., Paszkowski, T. & Maciejewski, R. Selection and validation of endogenous controls for microRNA expression studies in endometrioid endometrial cancer tissues. *Gynecol Oncol* **130**, 588–594, <https://doi.org/10.1016/j.ygyno.2013.06.026> (2013).
- Ayakannu, T., Taylor, A. H. & Konje, J. C. Cannabinoid receptor expression in estrogen-dependent and estrogen-independent endometrial cancer. *J Recept Signal Transduct Res*, 1–8, <https://doi.org/10.1080/10799893.2018.1531890> (2018).
- Ayakannu, T. *et al.* Validation of endogenous control reference genes for normalizing gene expression studies in endometrial carcinoma. *Mol Hum Reprod* **21**, 723–735, <https://doi.org/10.1093/molehr/gav033> (2015).
- Blazquez, C. *et al.* Cannabinoids inhibit glioma cell invasion by down-regulating matrix metalloproteinase-2 expression. *Cancer Res* **68**, 1945–1952, <https://doi.org/10.1158/0008-5472.CAN-07-5176> (2008).
- Huggett, J., Dheda, K., Bustin, S. & Zumla, A. Real-time RT-PCR normalisation; strategies and considerations. *Genes Immun* **6**, 279–284, <https://doi.org/10.1038/sj.gene.6364190> (2005).
- Mishima, K. *et al.* Increased expression of podoplanin in malignant astrocytic tumors as a novel molecular marker of malignant progression. *Acta Neuropathol* **111**, 483–488, <https://doi.org/10.1007/s00401-006-0063-y> (2006).
- Radonic, A. *et al.* Guideline to reference gene selection for quantitative real-time PCR. *Biochem Biophys Res Commun* **313**, 856–862 (2004).
- Farias-Eisner, G. *et al.* Validation of serum biomarkers for detection of early- and late-stage endometrial cancer. *Am J Obstet Gynecol* **202**, 73 e71–75, <https://doi.org/10.1016/j.ajog.2009.07.049> (2010).
- Burns, J. M. *et al.* A novel chemokine receptor for SDF-1 and I-TAC involved in cell survival, cell adhesion, and tumor development. *J Exp Med* **203**, 2201–2213, <https://doi.org/10.1084/jem.20052144> (2006).
- Edlinger, M. *et al.* Lifestyle-related biomarkers and endometrial cancer survival: elevated gamma-glutamyltransferase as an important risk factor. *Cancer Epidemiol* **37**, 156–161, <https://doi.org/10.1016/j.canep.2012.12.003> (2013).
- Shaarawy, M. & El-Sharkawy, S. A. Biomarkers of intrinsic angiogenic and anti-angiogenic activity in patients with endometrial hyperplasia and endometrial cancer. *Acta Oncol* **40**, 513–518 (2001).
- Porichi, O. *et al.* BCL-2, BAX and P53 expression profiles in endometrial carcinoma as studied by real-time PCR and immunohistochemistry. *Anticancer Res* **29**, 3977–3982, <http://ar.iiarjournals.org/content/29/10/3977> (2009).
- Wong, J. Y., Huggins, G. S., Debidia, M., Munshi, N. C. & De Vivo, I. Dichloroacetate induces apoptosis in endometrial cancer cells. *Gynecol Oncol* **109**, 394–402, <https://doi.org/10.1016/j.ygyno.2008.01.038> (2008).
- Bustin, S. A. Absolute quantification of mRNA using real-time reverse transcription polymerase chain reaction assays. *J Mol Endocrinol* **25**, 169–193, <https://doi.org/10.1677/jme.0.0250169> (2000).
- Bustin, S. A. Quantification of mRNA using real-time reverse transcription PCR (RT-PCR): trends and problems. *J Mol Endocrinol* **29**, 23–39, <https://doi.org/10.1677/jme.0.0290023> (2002).
- Romanowski, T., Markiewicz, A., Bednarz, N. & Bielawski, K. P. [Housekeeping genes as a reference in quantitative real-time RT-PCR]. *Postępy higieny i medycyny doświadczalnej* **61**, 500–510, http://www.phmd.pl/pub/phmd/vol_61/11284.pdf (2007).
- Bustin, S. A. *et al.* The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments. *Clin Chem* **55**, 611–622, <https://doi.org/10.1373/clinchem.2008.112797> (2009).

20. Kubista, M. *et al.* The real-time polymerase chain reaction. *Mol Aspects Med* **27**, 95–125, <https://doi.org/10.1016/j.mam.2005.12.007> (2006).
21. Guida, M. *et al.* The levels of the endocannabinoid receptor CB2 and its ligand 2-arachidonoylglycerol are elevated in endometrial carcinoma. *Endocrinology* **151**, 921–928, <https://doi.org/10.1210/en.2009-0883> (2010).
22. Papageorgiou, I. *et al.* Expression of nodal signalling components in cycling human endometrium and in endometrial cancer. *Reprod Biol Endocrinol* **7**, 122, <https://doi.org/10.1186/1477-7827-7-122> (2009).
23. Du, X. L., Jiang, T., Sheng, X. G., Gao, R. & Li, Q. S. Inhibition of osteopontin suppresses *in vitro* and *in vivo* angiogenesis in endometrial cancer. *Gynecol Oncol* **115**, 371–376, <https://doi.org/10.1016/j.ygyno.2009.08.029> (2009).
24. Kashima, H. *et al.* Autocrine stimulation of IGF1 in estrogen-induced growth of endometrial carcinoma cells: involvement of the mitogen-activated protein kinase pathway followed by up-regulation of cyclin D1 and cyclin E. *Endocr Relat Cancer* **16**, 113–122, <https://doi.org/10.1677/ERC-08-0117> (2009).
25. Balduin, P. *et al.* CASC2a gene is down-regulated in endometrial cancer. *Anticancer Res* **27**, 235–243, <http://ar.iarjournals.org/content/27/1A/235.long> (2007).
26. Hellemans, J., Mortier, G., De Paep, A., Speleman, F. & Vandesompele, J. qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. *Genome Biol* **8**, R19, <https://doi.org/10.1186/gb-2007-8-2-r19> (2007).
27. NormFinder. (Aarhus University Hospital: Aarhus, Denmark, 2005).
28. Krasnov, G. S. *et al.* Pan-cancer analysis of TCGA data revealed promising reference gene for qPCR normalization. *Front Genet* **10**, 97, <https://doi.org/10.3389/fgene.2019.00097> (2019).
29. Fung, J. N. *et al.* The genetic regulation of transcription in human endometrial tissue. *Hum Reprod* **32**, 893–904, <https://doi.org/10.1093/humrep/dex006> (2017).
30. Romani, C. *et al.* Identification of optimal reference genes for gene expression normalization in a wide cohort of endometrioid endometrial carcinoma. *PLoS One* **9**, <https://doi.org/10.1371/journal.pone.0113781> (2014).
31. Thellin, O. *et al.* Housekeeping genes as internal standards: use and limits. *J Biotech* **75**, 291–295, <https://www.gene-quantification.de/thellin-1999.pdf> (1999).
32. Hoogewijs, D., Houthoofd, K., Matthijssens, F., Vandesompele, J. & Vanfleteren, J. R. Selection and validation of a set of reliable reference genes for quantitative *sod* gene expression analysis in *C. elegans*. *Mol Biol* **9**, 9, <https://doi.org/10.1186/1471-2199-9-9> (2008).
33. Nelson, P. S. Predicting prostate cancer behavior using transcript profiles. *J Urol* **172**, S28–32; discussion S33, <https://doi.org/10.1097/01.ju.0000142067.17181.68> (2004).
34. Amoako, A. A. *et al.* Impact of reference gene selection for type 2 cannabinoid receptor gene expression studies in human spermatozoa. *Andrologia* **45**, 278–284, <https://doi.org/10.1111/and.12006> (2013).
35. Gebeh, A. K., Marczylo, E. L., Amoako, A. A., Willets, J. M. & Konje, J. C. Variation in stability of endogenous reference genes in fallopian tubes and endometrium from healthy and ectopic pregnant women. *Int J Mol Sci* **13**, 2810–2826, <https://doi.org/10.3390/ijms13032810> (2012).
36. Pecorelli, S. Revised FIGO staging for carcinoma of the vulva, cervix, and endometrium. *Int J Gynecol Obstet* **105**, 103–104, <https://doi.org/10.1016/j.ijgo.2009.02.012> (2009).
37. Noyes, R. W., Hertig, A. T. & Rock, J. Dating the endometrial biopsy. *Fertil Steril* **1**, 3–25, <https://doi.org/10.1097/00006254-195008000-00044> (1950).
38. McDonald, M. E. & Bender, D. P. Endometrial cancer: Obesity, genetics, and targeted agents. *Obstet Gynecol Clin North Am* **46**, 89–105, <https://doi.org/10.1016/j.ogc.2018.09.006> (2019).
39. 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, <https://doi.org/10.1158/0008-5472.CAN-04-0496> (2004).
40. Beekman, L., Tohver, T., Dardari, R. & Leguillet, R. Evaluation of suitable reference genes for gene expression studies in bronchoalveolar lavage cells from horses with inflammatory airway disease. *Mol Biol* **12**, 5, <https://doi.org/10.1186/1471-2199-12-5> (2011).
41. Pinto, F., Pacheco, C. C., Ferreira, D., Moradas-Ferreira, P. & Tamagnini, P. Selection of suitable reference genes for RT-qPCR analyses in cyanobacteria. *PLoS One* **7**, e34983, <https://doi.org/10.1371/journal.pone.0034983> (2012).

Acknowledgements

We thank Dr. J.H. Pringle (Pathology Department, University of Leicester) for allowing us to use his StepOnePlus qRT-PCR machine and Linda Jane Potter for teaching AT how to use it. We also thank Dr. John McDonald (Department of Anaesthesia, Leicester Royal Infirmary) for teaching AT on how to process and prepare the tissues for the qRT-PCR experiments.

Author contributions

T.A. designed the study, recruited the volunteers, collected the tissues, and undertook the quantification with supervision from A.H.T. T.A. and A.H.T. performed the data analysis with input from J.C.K. T.A. and A.H.T. produced the initial draft of the manuscript and all authors contributed to its revision. J.C.K. is the senior supervisor and guarantor of the study.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-64663-4>.

Correspondence and requests for materials should be addressed to T.A.

Reprints and permissions information is available at www.nature.com/reprints.

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/>.

© The Author(s) 2020