Sex bias evaluation of classic and novel Housekeeping Genes in adipose tissue through the massive analysis of transcriptomics data

Maria Guaita-Cespedes[1,+], Rubén Grillo-Risco[1,+], Marta R. Hidalgo[1], Sonia Fernández-Veledo[2], Deborah Burks[3], María de la Iglesia-Vayá[4], Amparo Galán[3,*], and Francisco Garcia-Garcia[1,*]

[1]Bioinformatics and Biostatistics Unit, Principe Felipe Research Center (CIPF), Valencia, 46012, Spain
[2]Department of Endocrinology and Nutrition and Research Unit, University Hospital of Tarragona Joan XXIII, Institut d'Investigaciò Sanitària Pere Virgili (IISPV), Tarragona, Spain; CIBER de Diabetes y Enfermedades Metabólicas Asociadas (CIBERDEM), Instituto de Salud Carlos III, Madrid, Spain
[3]Molecular Neuroendocrinology Unit, Principe Felipe Research Center (CIPF), Valencia, 46012, Spain
[4]Biomedical Imaging Unit FISABIO-CIPF, Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunidad Valenciana, 46012, Valencia, Spain
[+]these authors contributed equally to this work
[*]corresponding.author: agalan@cipf.es, fgarcia@cipf.es

ABSTRACT

Housekeeping genes (HKG), those involved in the maintenance of basic cell functions, are considered to have constant expression levels in all cell types, and are therefore commonly used as internal controls in gene expression studies. Nevertheless, multiple studies have shown that not all of them have stable expression levels across different cells, tissues, and conditions, introducing a systematic error in the experimental results. The proper selection and validation of control housekeeping genes in the specific studied conditions is crucial for the validity of the obtained results, although, up to date, sex has never been taken into account as a biological variable.

In this work, we evaluate the expression profiles of six classical housekeeping genes, (four metabolic: *HPRT*, *GAPDH*, *PPIA* and *UBC*, and two ribosomal: *18S* and *RPL19*) used as controls in several tissues, to determine the stability of their expression in adipose tissue of *Homo sapiens* and *Mus musculus* and asses sex bias and control suitability. We also evaluated gene expression stability of the genes included in different whole transcriptome microarrays available at the Gene Expression Omnibus database (GEO), to identify new genes suitable to be used as sex-unbiased controls. We perform a sex-based analysis to test for/reveal sexual dimorphism of mRNA expression stability.

We use a novel computational strategy based on meta-analysis techniques which evidence that some classical housekeeping genes do not fit to analyze human adipose tissue when sex variable is included. For instance, the extensively used *18S* has shown to be variable in this tissue, while *PPIA* and *RPL19* have shown to be good HKG targets. Further, we propose new sex-unbiased human and mouse housekeeping genes, derived from sex-specific expression profiles, including, *RPS8* or *UBB*. All the results generated in this work are available in an open web resource (https://bioinfo.cipf.es/metafun-HKG), so that they can be consulted and used in further studies.

Introduction

Housekeeping genes (HKG) are a large class of genes that are constitutively expressed and subjected to low levels of regulation in different conditions, which perform biological actions that are fundamental for the basic functions of the cell including cell cycle, translation, metabolism of RNA and cell transport[1,2]. Thus, they are expected to be expressed in all cells of an organism independently of the tissue, the developmental stage, the cell cycle state, or the external signals[3,4].

When performing quantitative gene expression analysis, such as microarrays, RNA-Seq and quantitative reverse transcriptase-polymerase chain reaction (qRT-PCR), the use of internal controls is the most common strategy to normalise gene expression, correcting intrinsic errors related to sample manipulation and technical protocol. The gene expression profiles obtained highly depend on the reference gene, leading to inaccurate results when using inappropriate internal controls.

Due to their fundamental roles, HKG tend to be expressed in medium-high levels, which makes them specially suitable to be used as internal controls/reference genes to normalise gene expression data in quantitative gene expression analysis[2,5,6]. Ideally, internal controls should exhibit stable gene expression across most sample types and experimental conditions, so that no differences in their expression may lead to undesired experimental variation. However, the literature suggests that the expression of the most commonly used HKG can vary depending on the experimental conditions and chosen set up, and also the analyzed tissue[6–13]. These limitations do not invalidate the use of HKG as normalization strategy, in any case, they evidence the need for increasing the knowledge about how they behave under different conditions or tissues. Reference HKG need to be previously selected and their expression stability validated under the particular conditions of interest of each study as a mandatory step[5], taking into account all experimental, biological, or clinical variables[7,14–16], including sex.

The role of sex in biomedical studies has often been overlooked, despite evidence of sexually dimorphic effects in some biological studies[17]. In this recent study, a large proportion of mammalian traits, both in wildtype and mutants have been found to be phenotypically influenced by sex. Further, another recent work has reported the impact of sex on gene expression across human tissues through metadata analysis by the GTEx platform, generating a catalogue of sex differences in gene expression and its genetic regulation[18]. This study revealed that sex effects on gene expression were found ubiquitously, and one of the tissues most affected was the human adipose tissue, both visceral and subcutaneous[18]. This intrinsic factor has not historically been considered. In a recent review work, from more than 600 papers analyzed for animal research, 22% of the publications did not specifiy the sex of the animals[19]. Of the reports that did specifiy the sex, 80% of publications included only males, 17% only females, and 3% both sexes[20]. A greater disparity was found in the number of animals studied (16,152 males vs 3,173 females). Only 7 (1%) studies reported sex-based results. Thus, across ages, the number of male-only studies and usage of male animals has become more disparate over time[20,21]. Unfortunately, human counterpart studies are not more encouraging, and although sex variable is being introduced boosted by international institutions[22,23], it has not been done in the past and a male perspective has always been predominant. The consequence of not taking the sex variable into consideration may either accentuate, or attenuate gene expression analysis in several conditions with further implications on interpretation at a biological or biomedical purpose.

The analysis of quantitative gene expression data of RNA levels has proven to be of utility to assess the genome expression levels at different tissues and conditions, and can be used to identify stable expression profiles/patternn.[1,9,12,24–28]. In the last decade, different public repositories of gene expression data have appeared. Gene Expression Omnibus (GEO[29]) is one of the most known international public repositories that stores and allows the access to gene expression data generated

by different high throughput technologies, such as microarrays or next-generation sequencing. Exploiting and reusing the vast amount of data in these repositories has become a powerful tool to find patterns across many different types of tissues and conditions.

According to the HKG generally studied, a survey of 40 studies published since 2001 shows that, in 70% of the papers, *ACTB*, *GAPDH*, and *18S* rRNA were used as reference genes in human adipose tissue (HAT)[14]. Other studies have reflected the use of other common HKG generally used in several cells and tissues, like *PPIA*, *HPRT, RPS18* or RPL19 to be also useful in HAT[16, 22]. However, none of these studies have included sex as a biological variable, thus, using this markers may affect the final outcome. In short, there is an important limitation in biomedical studies due to the lack of inclusion of the sex perspective.

In this work we determine the gene expression variability levels of six housekeeping genes commonly used in human and mouse adipose tissue and of the genes included in different whole transcriptome microarrays available at GEO, taking into account the sex covariable. We also identify novel candidates for reference genes unbiased for sex in human adipose tissue. Besides, we have extended to mice experimental analyses deposited in GEO. Our studies have revealed that either no sex specificity or only males have been studied in mice, and that some classical HKG do not accomplish male and female requisites of what is defined as a constitutive gene. Also, we have established new putative HKG to be used for gene expression analysis in male and female human adipose tissue and putative orthologs for mouse adipose tissue. A general framework is presented for choosing reference genes that may be useful in gene expression studies on normal tissues and organs. Further, we have developed an open web tool to select the proper HKG according to each customized experimental design.

Results

Classic HKG selection

An extensive bibliographic review has revealed that most reference genes for qRT-PCR measurements of gene expression in human adipose tissue (HAT) or adipocytes of several types include metabolic genes *GAPDH*[7,14–16,30,31], *HPRT*[7,16], *PPIA*[14,30,31], UBC and ribosomal genes *18S*[7,14,16,31–34] and *RPL19*[35]. All these genes, have been employed to analyze gene expression as reference genes in several experimental conditions, although sex variable has never been taken into account. We select these six classic HAT HKG genes to be evaluated in the context of a sex-perspective, to assess their suitability as sex-unbiased HKG (suHKG).

Systematic review and data collection

We have conducted an advanced research at GEO defining the sample tissue, the type of study and the organism of interest, and obtained a total of 187 candidate studies for *Homo sapiens* (Hsa) and 214 studies for *Mus musculus* (Mmu). For each species we selected the main microarray platforms that contain the greatest number of studies, 4 platforms for Hsa (**Table 1**) and 5 platforms for Mmu (**Table 2**). Up to 138 studies of Hsa and 171 studies of Mmu were excluded for not meeting the inclusion criteria. Finally, we selected 49 Hsa studies and 43 Mmu studies (**Fig 1**), including 2,724 Hsa samples and 1,072 Mmu samples.

In Hsa, just twenty-four (49%) of the forty-nine selected studies included information about their samples' sex. Ten studies covered both sexes in their analysis, eleven included exclusively females

and three contained only men samples (**Fig 2A**). In total, 681 Hsa male samples and 875 Hsa female samples were properly identified (Supplementary **Table S1, Supplementary Fig S1**). In Mmu, twenty-two (51%) of the forty-three selected studies informed about their samples' sex. Only one study covered both sexes, two included exclusively female samples, and nineteen contained only male samples (**Fig 2B**). In total, 559 Mmu male samples and 34 Mmu female samples were properly identified (Supplementary **Table S2, Supplementary Fig S2**). Due to the lack of female samples, Mmu studies were excluded from the sex-based analysis.

Stability data meta-analysis

After downloading and annotating the normalized expression data of the selected studies, we calculated three estimators of variability: the coefficient of variation (CV), the interquartile range divided by the median value (IQR/median) and the mean absolute deviation divided by the median value (MAD/median). **Fig. 3** and **Fig. 4** summarize the obtained levels of variability of the six selected HAT HKG (UBC, RPL19, RNA18S5, PPIA, HPRT1 and GAPDH) for Hsa and Mmu respectively. The specific values for male and female samples are shown at the Supp. Material (**Fig. S3 and Fig. S4**).

To integrate the statistical results from all the different platforms we conducted a meta-analysis based on the Rank Product (RP) method. This approach combines the ranks of the genes rather than their variability scores, creating platform-independence, and identifies the elements that systematically occupy higher positions in ranked lists, giving to each element of the ranking its RP score. We calculated the RP score of 41,975 Hsa genes and 47,203 Mmu genes and sorted them. In this ranking, lower positions indicate higher expression stability. We found that, from the six selected HAT HKG, *RPL19*, *PPIA* and *UBC* are the most stable genes (**Fig. 5**) in human samples. On the contrary, 18S shows an important variability in Hsa, although surprisingly it is the second most stable gene in Mmu. Probably, the high presence of mouse male samples and the fact that this gene shows an important sex bias, especially in mouse and rat, may explain this behaviour. Figure 5 and Figure 6 show the positions occupied by the six selected HAT HKG in Hsa and Mmu, and in Hsa males and Hsa females, respectively. The whole rankings with the positions and RP scores of all evaluated genes in each experimental condition are available at the Metafun-HKG webtool.

To select sex-unbiased, highly expressed and stable human HAT HKG candidates, we combined the scores of the three statistics in a unique list of positions for each experimental condition (metaRanking) and filtered out of them genes with low expression (TPM < 20) in the GTEx database, obtaining 5,315 genes. Then, we intersected the Top 10% (532) most stable genes in the Hsa male and Hsa female metaRankings separately, which resulted in a list of 195 candidate suHKG . This analysis revealed relative stability and expression values high enough to be detected by the different gene expression analysis technologies in Hsa male and female samples **(Table 3, Figure S6)** of some of the selected HAT HKG, as *PPIA*, *UBC*, *RPL19* and *RPS18* and of other novel candidate genes, as *RPS8* and *UBB*. We also found genes which were stable and highly expressed in one sex but not in the other, as *ANXA2*, *DDX39B* and *PLIN4* for males, and *DNASE2*, *NDUFB11* or *RARA* for females (**Table S3. Figure S6**), which may be used as sex-specific reference genes. We failed to find the expression of gene 18S in GTEx, although different aliases were searched *(RNA18S5, RNA18S1, RNA18SN1, RNA18SN5, RN18S1)*.

Experimental Validation

According to the computational assessment of variability performed, we selected genes *RPL19*, *PPIA* and *18S* for experimental validation. HAT mRNA from lean, obese, and diabetic, male and female individuals were analysed using these genes as reference (**Table 3; Fig. 7**). Raw crossing points (Cps) values coefficient variation (CV) analysis revealed significant differences in *18S* expression values between male and female samples in the three analyzed conditions (**Fig. 7A**), thus affecting relative gene expression analysis of other experimental targets when used as internal control (**Fig. 7B**).

However, *RPL19* and *PPIA* presented similar Cps values between male and female samples with low CV, thus evidencing their suitability as suHKG (**Fig. 7**). We conclude that experimental procedures validate computational metadata analysis, discarding *18S* and selecting *PPIA* and *RPL19* as suHKG for HAT analysis.

To fulfill the lack of sex-based Mmu data with which to compute a Mmu metaRanking, we tested experimentally the mouse orthologs (Ppia, Gapdh, Rpl19 and 18s) of the validated human suHKG. Relative gene expression analysis shows that, also in mice, relative expression of different experimental targets may be affected by the internal control used. In particular, relative gene expression of Irs2 changes dramatically when normalized with respect to 18s in contrast to Ppia, Gadph or Rpl19, for which it remains comparable (**Fig. S5**). These results confirm that our suHKG candidates' orthologs may be used as suHKG themselves in mice.

Metafun-HKG Web Tool

The Metafun-HKG web tool (https://bioinfo.cipf.es/metafun-HKG) contains information related to the 49 Hsa studies (681 male and 875 female samples) and Mmu 43 studies (559 male and 34 female samples). The portal includes stability indicators for each of the genes in the studies evaluated by platform, species and sex, which can be explored by users to identify profiles of interest.

Discussion

Assessment of HKG Expression. The two main objectives of this work have been first, the evaluation of the suitability of a group of six classical HKGs, generally used as reference in gene expression analysis of several tissues, to act as controls in adipose tissue, and second, the identification of genes with a stable and high expression profile, which allow them to be potential candidates for the normalisation of expression data in human and mouse adipose tissue, taking into account a sex perspective. We propose a novel strategy that has made it possible to review the role of HKGs by considering several variables of interest in the studies evaluated.

We performed the analysis pipeline on four different experimental conditions based on sex and species: male Mmu, female Hsa, male Hsa, and both male and female Hsa samples. We failed to perform the female and the both male and female Mmu samples analysis due to the lack of female Mmu reported samples. In all conditions HKG have shown platform-dependent variability, given that each microarray platform has its particular probe design and technical protocol. Previous studies comment on the existence of technology-dependent variability and conclude that it is not as determinant as the existing differences of transcript expression levels related to cell conditions[24].

Proposed candidate genes. We propose a list of 193 suHKG candidates with high relative levels of stability and high and comparable expression levels (TPM>20) in male and female samples. These genes would be suitable to be used as internal controls in HAT studies including male and female samples, as they show high expression and stability levels, and, also, minimal influece of sex in their expression patterns. We were not able to reproduce the pipeline followed with human samples in mouse studies due to the lack of mouse female samples, but we suggest as mouse suHKG the orthologs of the human proposed suHKG.

Experimental validation. We validated experimentally a selection of the proposed suHKG candidates, together with *18S* to assess computational findings. Gene expression analysis validated in silico results (**Table 3**). Specifically, *PPIA,* a generally extended HAT HKG, and *RPL19*, currently used as HKG in several cell types[39–41] and ocasionally in HAT studies[35] , have been validated as HAT suHKG. On the contrary, experimental validation shows that *18S*, which is widely used as HAT HKG[7,14,16,31–34], displays not only important levels of variability in both male and female samples separately, but also different

expression patterns between sexes. These results have also been evidenced in mouse adipose tissue (**Fig. 7 y S3**), where 18s shows clear differences in male and female samples (Fig S5) despite the high stability found in the metaRanking, which could be due to the important presence of male samples in this group.. These results could explain why *18s* has been systematically reported as a stable gene to be used as a reference when most of the studies do not incorporate females. In this context, *HPRT1* also presents high levels of variability, but its expression pattern is similar in males and females introducing similar errors in both sexes, thus it is not as robust and adequate as *PPIA* or *RPL19,* but does not introduce such an error as *18S* does.

Further, several genes have also been computationally suggested, such as *RPS18* *RPS8*, or *UBB* (**Table 3**). Although no experimental validation has been carried out, these genes present suitable stability and expression levels which make them promising suHKG to be taken into account. A web tool has been designed to customize the best suHKG fit in each experimental design of human or mouse adipose tissue. We also suggest the orthologs of the human proposed suHKG as mouse suHKG, and experimental validation corroborates the suitability of the tested genes.

With this strategy we have revealed that thanks to proper selection of HKG, such as *PPIA*, *RPL19*,or other potential markers (*RPS18*, *UBB)*, a proper experimental design can be performed. We will be able to finally avoid the common practice of pooling males and females or even discard the extended male effect.

Strengths and limitations. Massive data analysis of gene expression has been implemented through the years as a key tool to deepen and further understand different biological scenarios which may eventually help to elucidate mechanisms affecting all levels from basic research to biomedical implications. All these data analyses must be assessed in the laboratory practice by relative gene expression analysis referred to an adequate HKG. Selection of the ideal HKG is tricky and essential to ensure an accurate result, and that has to be done taking into consideration experimental conditions and biological variables[6,16]. Incorporating sex analysis into research can improve reproducibility and experimental efficiency, as it can influence the outcome of experiments and must be accounted for as a critical biological variable, and it is important to take it into account to monitor sex differences and similarities for all diseases and biological processes that affect both sexes. It also may help to reduce bias, enable social equality in scientific outcomes and encourage new opportunities for discovery and innovation as evidenced by several studies analyzing this new issue[20,22].

Numerous lines of evidence indicate that the current status quo is not addressing fundamental issues of sex differences that are evident in gene expression. On the one hand, many of the most used HKG have not been tested including sex as a biological variable. It is the case of ACTB, GAPDH and 18S, commonly used in HAT, and also of other HKG such as *PPIA*, *HPRT*, *RPS18* or *RPL19*, which have proven to be also useful in HAT. To use a HKG to normalize samples without assessing its behaviour in the particular experimental conditions of the study, including sex, may lead to different biases in the outcome. In particular, the HKG may be stable in one sex but not in the other, as in the case of *DDX39B* and *PLIN4* (stables just in males) or *NDUFB11* and *RARA* (stables just in females), or, although stable, it may have different levels of expression in both sexes, as *18S*. This would lead to confounded variables and results in which it is not possible to assess whether the differences in the data come from the experimental design or the normalization process. At the same time, this would be a potential source of variability in the data which would reduce statistical power, making it more difficult to find significant results. In this study, we have analyzed the role of six conventional HAT HKG taking into consideration the sex biological variable for the first time.

On the other hand, published studies do not include a sex perspective, omiting the sex of the animals or performing studies with animals of only one sex, tipically males. In consonance, we found in our systematic review that 51% of Hsa studies and 49% of Mmu studies did not include sex sample

information, and just a 19% of Hsa and a strikingly 2% of Mmu studies included samples from both sexes. Also, Mmu studies including only female samples represented just 5% of the total. The low number of Mmu studies including female sample information has been a great limitation of the work, which has prevented from creating a Mmu meta-ranking to select highly expressed stable Mmu suHKG candidates, as in the case of Hsa. To overcome this limitation, we tested experimentally the Mmu orthologs of the selected Hsa suHKG candidates, confirming their suitability as Mmu suHKG.

Despite the general use of *18S* RNA as HKG, its annotation has been another limiting factor, as we could not find this gene in the GTEx platform under any of its proposed aliases in GeneCards. We also notice that the identifiers for this gene are not stable or not included in the reference assemblies. In addition, the DNA sequence of the *RNA18SN5* gene (accession number NR_003286.4) has 99-100% identity with other ribosomal RNAs such as *RNA18SN1*, *RNA18SN2*, *RNA18SN3*, *RNA18SN4*, or *RNA18SP3* (accession numbers NR_145820.1, NR_146146.1, NR_146152.1, NR_146119.1, NG_054871.1, respectively). It is known that 18S rRNA has different copy number among individuals and also varies with age[36], so taking all this into account makes the *18S* gene less suitable as suHKG in human adipose tissue than other genes proposed in this work. Our experimental validations also support these findings.

Other limitations of the study have been the filtering and pre-processing of the biological information located at GEO to identify the published studies with transcriptomic data of adipose tissue, and the classification of the samples depending on the sex. The main limiting factor has been the absence of standardized vocabulary to tag the sex at the sample records of the studies. Even though the gene expression data in GEO is presented as a standardized expression matrix, the metadata (including sample source, tissue type or sample's sex) is reported through free-text fields written by the researcher that submits the study. The absence of standardized vocabulary and structured information constrains the data mining power on large scale data, and their improvement could aid the processing of data in public repositories[38].

This work presents, for the first time, a computational strategy that includes a massive data analysis capable of assessing the sex bias in expression levels of classical and novel HGK, over a large volume of studies and samples.

With this strategy we have revealed that thanks to proper selection of suHKG, such as *PPIA*, *RPL19*, or other potential markers (*RPS18*, *UBB)*, a proper experimental design can be performed. We will be able to finally avoid the common practice of pooling males and females or even discard the extended male effect. In this work we present not only the relative expression stability of six commonly used housekeeping genes, but also the variability levels of other genes as long as they are covered by the analysed microarray platforms. This same workflow/methodology is translatable to assess the expression stability in other tissues modifying the sample source at the advanced search step to collect data from GEO and the SQL queries of GEOmetadb to obtain the sample information. This strategy is aligned with the FAIR principles[37] (Findability, Accessibility, Interoperability and Reusability) to ensure the further utility and reproducibility of the generated information.

Our findings, although limited to HAT, suggest that the same sex-bias in commonly used HKG could be appearing in other tissues, affecting the normalization process of gene expression analysis of any kind. Incorrect normalization may alter significantly gene expression data, as we have shown in the case of 18S, and lead to erroneous conclusions. This work shows the importance of taking into account the sex-prespective in biomedical studies, and hints that a thorough analysis of the different HKG used as internal controls in all tissues should be promptly adressed.

Methods

The bioinformatics analysis strategy was carried out using the programming languages R 3.5.0[42] and Python 3.0. This approach consists of several phases described below and summarised in **Fig. 8**:

Systematic review and data collection

A comprehensive systematic review was carried out to identify all the available studies with adipose tissue samples processed with array platforms at GEO. First, we searched in GEO defining the fields sample source - adipose, type of study - expression profiling by array, and organism of interest - *Homo sapiens* or *Mus musculus*. The search was conducted during the first quarter of 2020, and the review period covered the years 2000-2019. We extracted from the returned records: the studies GSE ID, the platform GPL ID and the study type using the Python 3.0 library Beautiful Soup. After that, we used the R package GEOmetadb[43] to identify microarray platforms and selected the top four and top five most used platforms in Hsa (**Table 1)** and Mmu (**Table 2)** respectively to work with.

We then used the GEOmetadb package to identify samples from adipose tissue. Given the complex nature of some of the studies, we manually determined the studies informing about their samples' sex and homogenized the keywords used to annotate them.

Finally, we filtered out studies not meeting the following predefined inclusion criteria: i) include at least 10 adipose tissue samples, ii) use one of the selected microarray platforms to analyze gene expression data, iii) present data in a standardized way, iv) not include duplicate sample records (as superseries).

Data processing and statistical analysis

We downloaded the normalized microarray expression data of the selected studies from GEO using the GEOQuery R package. All the probe sets of each platform were converted to gene symbol, averaging expression values of multiple probe sets targeting the same gene to the median value.

To determine the relative expression variability, we calculated three statistical stability indicators for each gene in each individual study: the coefficient of variation (CV), the IQR/median and the MAD/median. The CV, computed as the standard deviation divided by the mean, is used to compare variation between genes with expression levels at different orders of magnitude but can be greatly affected by extreme values. Therefore, we also considered other two statistics based on the median, known to be more robust in skewed distributions[44]: the interquartile range (IQR) divided by the median, and the median absolute deviation (MAD) divided by the median. Both statistics were multiplied by a correction factor of 0.75 and 1.4826 respectively to make them comparable to the CV in normal distributions.

Lastly, we resumed the variability scores of each gene in each platform as the median of the statistics of the studies analyzed with each platform.. The calculated median values were ranked, lower ranks corresponding to higher stability levels.

We performed the described analysis pipeline on four different experimental conditions based on sex and species: male Mmu, female Hsa, male Hsa, and both male and female Hsa samples. We failed to perform the female and the both male and female Mmu samples analysis due to the lack of female

Mmu reported samples.

Meta-analysis

We integrate the individual results obtained with all the platforms for each condition, using the Rank Product (RP) method[45,46]. The Rank Product is a non-parametrical statistic that identifies the elements that systematically occupy higher positions in ranked lists. It is used in transcriptomics studies to identify differentially expressed genes across biological replicas[45]. This approach combines the ranks of the genes rather than their variability scores, creating platform-independence. We use the RankProd package[47,48], to calculate the RP score (equation 4), which equals the geometric mean of the gene position in each ranking, and generate a global ranked list. The final list is obtained by sorting the genes in increasing order of RP.

$$(1) \quad RP_i = \left(\prod_{j=1}^{K} rank_{ij} \right)^{1/K}$$

Selection of housekeeping candidate genes

The expression stability is the main characteristic for a gene to be considered as an internal control to normalise quantitative gene expression data, meaning that any gene could be potentially used as a reference gene whilst its expression is not affected by the particular experimental conditions. We generated a unique ranking ("MetaRanking") for each experimental condition (male Mmu, female Hsa, male Hsa, and both male and female Hsa samples) combining the results of the meta-analysis as the mean of the three statistics rankings (equation 5).

$$(2) \quad MetaRanking\ position = \frac{positionCV + positionIQR/median + positionMAD/median}{3}$$

We also calculated the difference in the ranking positions occupied in males and females to reveal sex-based stability differences at a gene level.

To select stable suHKG with high levels of expression, we first i) downloaded the "GTEx_Analysis_2017-06-05_v8_RNASeQCv1.1.9_gene_median_tpm.gct.gz" file from GTEx, ii) selected the adipose tissue samples and iii) took the gene median Transcript Per Million (TPM) value in Visceral Adipose tissue. We then iv) filtered out from our sex-specific rankings genes with median TPM < 20, and v) selected the genes in the Top 10% positions of each sex-specific ranking. Finally, we vi) intersected the two top lists to find stable and highly expressed genes common to both sexes.

Experimental validation

Study selection and sample processing. Subjects were recruited by the endocrinology and surgery departments at the University Hospital Joan XXIII (Tarragona, Spain) in accordance of Helsinki declaration. Visceral adipose tissue (VAT) and subcutaneous adipose tissue (SAT) samples were obtained during surgery. Total RNA was extracted from adipose tissue using the RNeasy lipid tissue midi kit (Qiagen Science). One microgram of RNA was reverse transcribed with random primers using

the reverse transcription system (Applied Biosystems)[31].

Mouse adipose tissue was obtained from Wild-type and Irs2-/- C57BL/6 littermates. All animals received humane care according to the criteria outlined in the "Guide for the Care and Use of Laboratory Animals"[22]. Total RNA was extracted from abdominal fat using a combined protocol including Trizol (Sigma) and RNeasy Mini Kit (Qiagen) with DNase Digestion. First-strand synthesis was performed using EcoDry Premix (Takara).

Gene Expression Analysis. Quantitative gene expression analysis was performed on 50 ng cDNA template. Real Time-PCR was carried out in LightCycler 480 Instrument IIR (Roche) using SYBR PreMix ExTaqTM (mi RNaseH Plus, Takara). Primers used in this study are specified in **Table S4**. Crossing points (Cp) values were analysed for stability between samples and for relative quantification following 2^-ΔCt.

Statistical Analyses were performed with GraphPad Prism 8 (Graphpad Software V 8.0). The results are expressed as arithmetic mean ± the standard error of the mean (SEM). When two data sets were compared, Student's t-test was used. The differences observed were considered significant when: p-value<0.05 (*), p-value<0.01 (**) and p-value<0.001 (***).

Web Tool

The large volume of data and results generated in this work is freely available in the metafun-HK web tool (https://bioinfo.cipf.es/metafun-HK), which will allow users to review the results described in the manuscript and any other results of interest to researchers. The front-end was developed using the Bootstrap library. All graphics used in this tool were implemented with Plot.ly, except for the exploratory analysis cluster plot, which was generated with the ggplot2[49] package.

This easy-to-use resource is organized into four sections: (1) a quick summary of the results obtained with the analysis pipeline in each of the phases. Then, for each of the studies, the detailed results of the 2) exploratory analysis and 3) variability assessment. Finally, all results will be integrated and summarized in 4) gene stability meta-analysis by sex and organism. The user can interact with the web tool through graphics and tables and search information for specific genes.

References

1. Chang CW, Cheng WC, Chen CR, et al. Identification of Human Housekeeping Genes and Tissue-Selective Genes by Microarray Meta-Analysis. *PLOS ONE*. 2011;6(7):e22859. doi:10.1371/journal.pone.0022859
2. Caracausi M, Piovesan A, Antonaros F, Strippoli P, Vitale L, Pelleri MC. Systematic identification of human housekeeping genes possibly useful as references in gene expression studies. *Mol Med Rep*. 2017;16(3):2397-2410. doi:10.3892/mmr.2017.6944
3. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013;29(10):569-574. doi:10.1016/j.tig.2013.05.010
4. Butte AJ, Dzau VJ, Glueck SB. Further defining housekeeping, or "maintenance," genes Focus on "A compendium of gene expression in normal human tissues." *Physiol Genomics*. 2001;7(2):95-96. doi:10.1152/physiolgenomics.2001.7.2.95
5. Dheda K, Huggett JF, Bustin SA, Johnson MA, Rook G, Zumla A. Validation of housekeeping genes for normalizing RNA expression in real-time PCR. *BioTechniques*. 2004;37(1):112-119. doi:10.2144/04371RR03

6. Bustin SA, Benes V, Garson J, et al. The need for transparency and good practices in the qPCR literature. *Nat Methods*. 2013;10(11):1063-1067. doi:10.1038/nmeth.2697
7. Chechi K, Gelinas Y, Mathieu P, Deshaies Y, Richard D. Validation of Reference Genes for the Relative Quantification of Gene Expression in Human Epicardial Adipose Tissue. *PLOS ONE*. 2012;7(4):e32265. doi:10.1371/journal.pone.0032265
8. Stürzenbaum SR, Kille P. Control genes in quantitative molecular biological techniques: the variability of invariance. *Comp Biochem Physiol B Biochem Mol Biol*. 2001;130(3):281-289. doi:10.1016/S1096-4959(01)00440-7
9. She X, Rohl CA, Castle JC, Kulkarni AV, Johnson JM, Chen R. Definition, conservation and epigenetics of housekeeping and tissue-enriched genes. *BMC Genomics*. 2009;10(1):269. doi:10.1186/1471-2164-10-269
10. Zhu J, He F, Song S, Wang J, Yu J. How many human genes can be defined as housekeeping with current expression data? *BMC Genomics*. 2008;9(1):172. doi:10.1186/1471-2164-9-172
11. Suzuki T, Higgins PJ, Crawford DR. Control Selection for RNA Quantitation. *BioTechniques*. 2000;29(2):332-337. doi:10.2144/00292rv02
12. Hsiao LL, Dangond F, Yoshida T, et al. A compendium of gene expression in normal human tissues. *Physiol Genomics*. 2001;7(2):97-104. doi:10.1152/physiolgenomics.00040.2001
13. Jonge HJM de, Fehrmann RSN, Bont ESJM de, et al. Evidence Based Selection of Housekeeping Genes. *PLOS ONE*. 2007;2(9):e898. doi:10.1371/journal.pone.0000898
14. Gabrielsson BG, Olofsson LE, Sjögren A, et al. Evaluation of reference genes for studies of gene expression in human adipose tissue. *Obes Res*. 2005;13(4):649-652. doi:10.1038/oby.2005.72
15. Heo JS, Choi Y, Kim HS, Kim HO. Comparison of molecular profiles of human mesenchymal stem cells derived from bone marrow, umbilical cord blood, placenta and adipose tissue. *Int J Mol Med*. 2016;37(1):115-125. doi:10.3892/ijmm.2015.2413
16. White JM, Piron MJ, Rangaraj VR, Hanlon EC, Cohen RN, Brady MJ. Reference Gene Optimization for Circadian Gene Expression Analysis in Human Adipose Tissue. *J Biol Rhythms*. 2020;35(1):84-97. doi:10.1177/0748730419883043
17. Karp NA, Mason J, Beaudet AL, et al. Prevalence of sexual dimorphism in mammalian phenotypic traits. *Nat Commun*. 2017;8(1):15475. doi:10.1038/ncomms15475
18. Oliva M, Muñoz-Aguirre M, Kim-Hellmuth S, et al. The impact of sex on gene expression across human tissues. *Science*. 2020;369(6509). doi:10.1126/science.aba3066
19. Yoon DY, Mansukhani NA, Stubbs VC, Helenowski IB, Woodruff TK, Kibbe MR. Sex bias exists in basic science and translational surgical research. *Surgery*. 2014;156(3):508-516. doi:10.1016/j.surg.2014.07.001
20. Tannenbaum C, Ellis RP, Eyssel F, Zou J, Schiebinger L. Sex and gender analysis improves science and engineering. *Nature*. 2019;575(7781):137-146. doi:10.1038/s41586-019-1657-6
21. Woitowich NC, Beery A, Woodruff T. A 10-year follow-up study of sex inclusion in the biological sciences. Sugimoto C, Rodgers P, Shansky R, Schiebinger L, eds. *eLife*. 2020;9:e56344. doi:10.7554/eLife.56344
22. McCullough LD, de Vries GJ, Miller VM, Becker JB, Sandberg K, McCarthy MM. NIH initiative to balance sex of animals in preclinical studies: generative questions to guide policy, implementation, and metrics. *Biol Sex Differ*. 2014;5:15. doi:10.1186/s13293-014-0015-5
23. Accounting for sex and gender makes for better science. *Nature*. 2020;588(7837):196-196. doi:10.1038/d41586-020-03459-y
24. Lee PD, Sladek R, Greenwood CMT, Hudson TJ. Control Genes and Variability: Absence of Ubiquitous Reference Transcripts in Diverse Mammalian Expression Studies. *Genome Res*. 2002;12(2):292-297. doi:10.1101/gr.217802
25. Lee SR, Jo MJ, Lee JE, Koh SS, Kim SY. Identification of Novel Universal Housekeeping Genes by Statistical Analysis of Microarray Data. *BMB Rep*. 2007;40(2):226-231. doi:10.5483/BMBRep.2007.40.2.226
26. Popovici V, Goldstein DR, Antonov J, Jaggi R, Delorenzi M, Wirapati P. Selecting control genes

for RT-QPCR using public microarray data. *BMC Bioinformatics*. 2009;10(1):42. doi:10.1186/1471-2105-10-42
27. Pilbrow AP, Ellmers LJ, Black MA, et al. Genomic selection of reference genes for real-time PCR in human myocardium. *BMC Med Genomics*. 2008;1(1):64. doi:10.1186/1755-8794-1-64
28. Zhang Y, Li D, Sun B. Do Housekeeping Genes Exist? *PLoS ONE*. 2015;10(5). doi:10.1371/journal.pone.0123691
29. Home - GEO - NCBI. Accessed December 1, 2021. https://www.ncbi.nlm.nih.gov/geo/
30. Amisten S. Quantification of the mRNA expression of G protein-coupled receptors in human adipose tissue. *Methods Cell Biol*. 2016;132:73-105. doi:10.1016/bs.mcb.2015.10.004
31. Ceperuelo-Mallafré V, Duran X, Pachón G, et al. Disruption of GIP/GIPR axis in human adipose tissue is linked to obesity and insulin resistance. *J Clin Endocrinol Metab*. 2014;99(5):E908-919. doi:10.1210/jc.2013-3350
32. Ebrahimi R, Bahiraee A, Alipour NJ, Toolabi K, Emamgholipour S. Evaluation of the Housekeeping Genes; β-Actin, Glyceraldehyde-3-Phosphate-Dehydrogenase, and 18S rRNA for Normalization in Real-Time Polymerase Chain Reaction Analysis of Gene Expression in Human Adipose Tissue. *Arch Med Lab Sci*. 2018;4(3). doi:10.22037/amls.v4i3.26269
33. Gómez-Abellán P, Díez-Noguera A, Madrid JA, Luján JA, Ordovás JM, Garaulet M. Glucocorticoids Affect 24 h Clock Genes Expression in Human Adipose Tissue Explant Cultures. *PLOS ONE*. 2012;7(12):e50435. doi:10.1371/journal.pone.0050435
34. Petrus P, Mejhert N, Corrales P, et al. Transforming Growth Factor-β3 Regulates Adipocyte Number in Subcutaneous White Adipose Tissue. *Cell Rep*. 2018;25(3):551-560.e5. doi:10.1016/j.celrep.2018.09.069
35. Catalano-Iniesta L, Sánchez Robledo V, Iglesias-Osma MC, et al. Evidences for Expression and Location of ANGPTL8 in Human Adipose Tissue. *J Clin Med*. 2020;9(2):512. doi:10.3390/jcm9020512
36. Hall AN, Turner TN, Queitsch C. Thousands of high-quality sequencing samples fail to show meaningful correlation between 5S and 45S ribosomal DNA arrays in humans. *Sci Rep*. 2021;11(1):449. doi:10.1038/s41598-020-80049-y
37. The FAIR Guiding Principles for scientific data management and stewardship | Scientific Data. Accessed December 1, 2021. https://www.nature.com/articles/sdata201618
38. Wang Z, Lachmann A, Ma'ayan A. Mining data and metadata from the gene expression omnibus. *Biophys Rev*. 2019;11(1):103-110. doi:10.1007/s12551-018-0490-8
39. Manzano-Núñez F, Arámbul-Anthony MJ, Albiñana AG, et al. Insulin resistance disrupts epithelial repair and niche-progenitor Fgf signaling during chronic liver injury. *PLOS Biol*. 2019;17(1):e2006972. doi:10.1371/journal.pbio.2006972
40. Galan A, Diaz-Gimeno P, Poo ME, et al. Defining the Genomic Signature of Totipotency and Pluripotency during Early Human Development. *PLOS ONE*. 2013;8(4):e62135. doi:10.1371/journal.pone.0062135
41. Galán A, Simón C. Monitoring Stemness in Long-Term hESC Cultures by Real-Time PCR. In: Turksen K, ed. *Human Embryonic Stem Cell Protocols*. Methods in Molecular Biology. Humana Press; 2010:135-150. doi:10.1007/978-1-60761-369-5_8
42. Team R. Core. R: A language and environment for statistical computing. *2013*.:16.
43. Zhu Y, Davis S, Stephens R, Meltzer PS, Chen Y. GEOmetadb: powerful alternative search engine for the Gene Expression Omnibus. *Bioinformatics*. 2008;24(23):2798-2800. doi:10.1093/bioinformatics/btn520
44. Arachchige C, Prendergast L, Staudte R. Robust analogs to the coefficient of variation. *J Appl Stat*. Published online August 20, 2020:1-23. doi:10.1080/02664763.2020.1808599
45. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. *FEBS Lett*. 2004;573(1-3):83-92. doi:https://doi.org/10.1016/j.febslet.2004.07.055
46. Mitchell L. A parallel implementation of the Rank Product method for R. :12.

47. Hong F, Breitling R, McEntee CW, Wittner BS, Nemhauser JL, Chory J. RankProd: a bioconductor package for detecting differentially expressed genes in meta-analysis. *Bioinformatics*. 2006;22(22):2825-2827. doi:10.1093/bioinformatics/btl476
48. Del Carratore F, Jankevics A, Eisinga R, Heskes T, Hong F, Breitling R. RankProd 2.0: a refactored bioconductor package for detecting differentially expressed features in molecular profiling datasets. *Bioinformatics*. 2017;33(17):2774-2775. doi:10.1093/bioinformatics/btx292
49. Wickham H. Elegant Graphics for Data Analysis. In: Wickham H, ed. *Ggplot2: Elegant Graphics for Data Analysis*. Use R. Springer; 2009:1-7. doi:10.1007/978-0-387-98141-3_1
50. Burks DJ, White MF. IRS proteins and beta-cell function. *Diabetes*. 2001;50(suppl 1):S140. doi:10.2337/diabetes.50.2007.S140
51. Garcia-Barrado MJ, Iglesias-Osma MC, Moreno-Viedma V, et al. Differential sensitivity to adrenergic stimulation underlies the sexual dimorphism in the development of diabetes caused by Irs-2 deficiency. *Biochem Pharmacol*. 2011;81(2):279-288.
52. Besse-Patin A, Jeromson S, Levesque-Damphousse P, Secco B, Laplante M, Estall JL. PGC1A regulates the IRS1:IRS2 ratio during fasting to influence hepatic metabolism downstream of insulin. *Proc Natl Acad Sci*. 2019;116(10):4285-4290. doi:10.1073/pnas.1815150116

Acknowledgements

The authors thank the Principe Felipe Research Center (CIPF) for providing access to the cluster, co-funded by European Regional Development Funds (FEDER) in Valencian Community 2014–2020.

Funding

Research supported by Principe Felipe Research Center, GV/2020/186 and SAF2017-84708-R grants.

Competing interests

The authors declare no competing interests.

Author contributions statement

Conceptualization, F.G.G., and A.G.; methodology, M.G., R.G.R., M.R.H., A.G., and F.G.G.; software, M.R.H.; formal analysis, M.G., and R.G.R.; investigation, M.G., R.G.R., M.R.H., A.G., and F.G.G.; data curation, M.G., and R.G.R.; experiment conduction: A.G. and S.F.V.; writing—original draft preparation, M.G., R.G.R., M.R.H., D.B., S.F.V., A.G., and F.G.G.; writing—review and editing, M.G., R.G.R., M.R.H., A.G., and F.G.G.; visualization, M.G., R.G.R., M.R.H.; supervision, A.G., M.R.H., and F.G.G.; funding acquisition, F.G.G., and D.B.; project administration, F.G.G., M.R.H., and A.G. All authors have read and agreed to the published version of the manuscript.

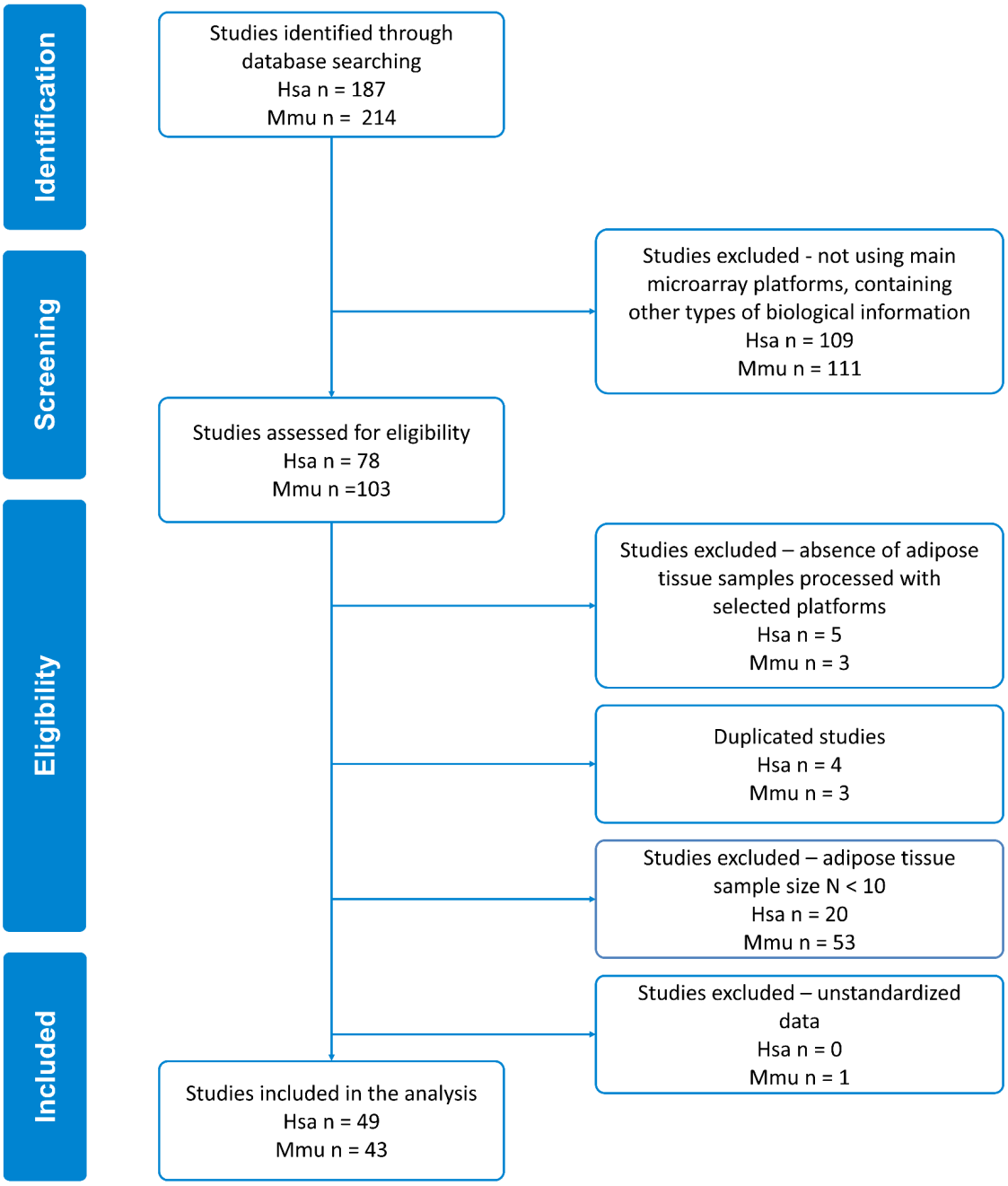

Figure 1. Flow diagram of the systematic review and selection of studies for meta-analysis according to PRISMA statement guidelines for searches of databases.

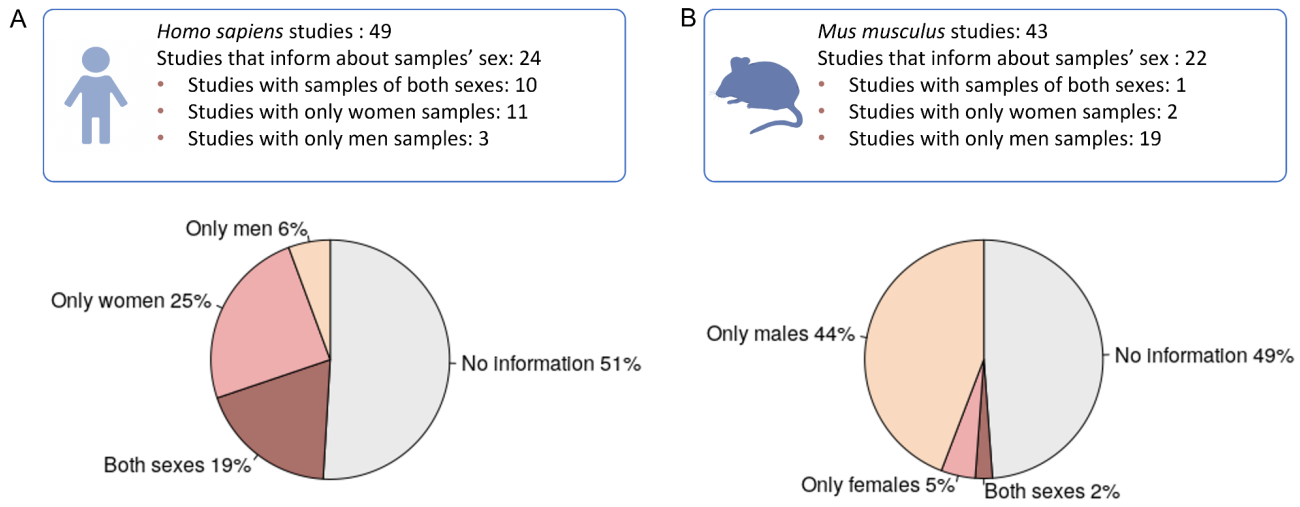

Figure 2. Summary of the sex variable inclusion found during the review of studies on each organism. A. Out of 49 Hsa studies 24 studies(49%) properly specified the sex of the samples, 51% of the studies lacked this information, only 19% of the studies used samples of both sexes in the experimental procedure. B. In Mmu, 51% of the studies presented information about the sex, but focused mainly in male samples, almost no female samples were found in these studies. Only one study included samples of both sexes.

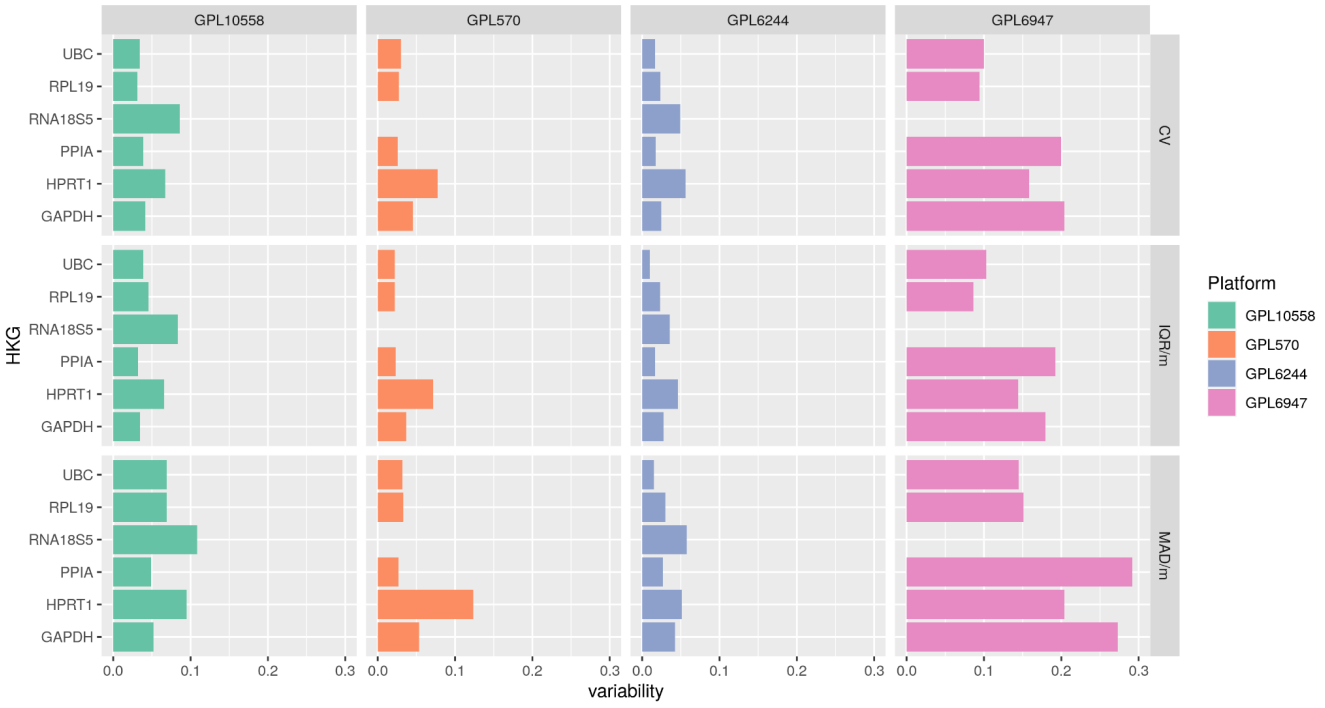

Figure 3. Variability levels for the classic housekeeping group evaluated in Hsa. For each of the housekeeping genes, the variability level found in the selected microarray platforms with the three statistical approaches (C.V., IQR/median and MAD/median) is described on the X-axis.

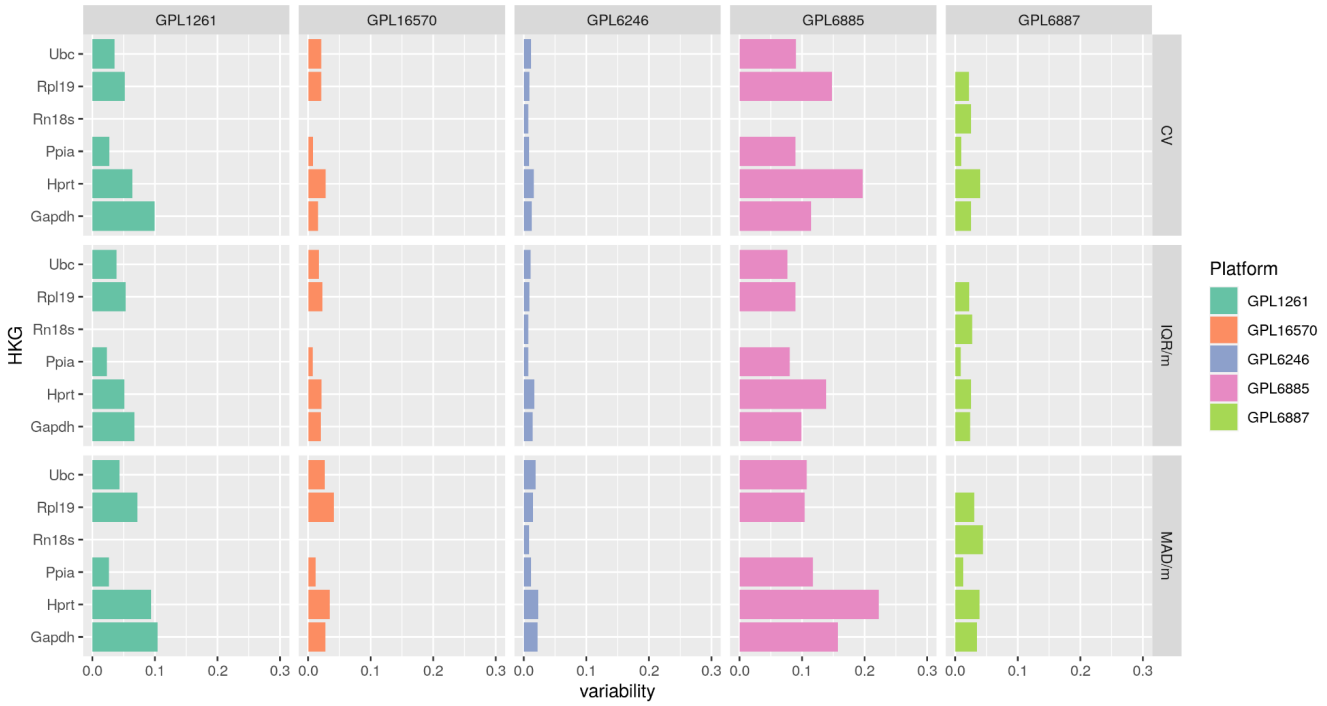

Figure 4. Variability levels for the classic housekeeping group evaluated in Mmu. For each of the housekeeping genes, the variability level found in the selected microarray platforms with the three statistical approaches (C.V., IQR/median and MAD/median) is described on the X-axis.

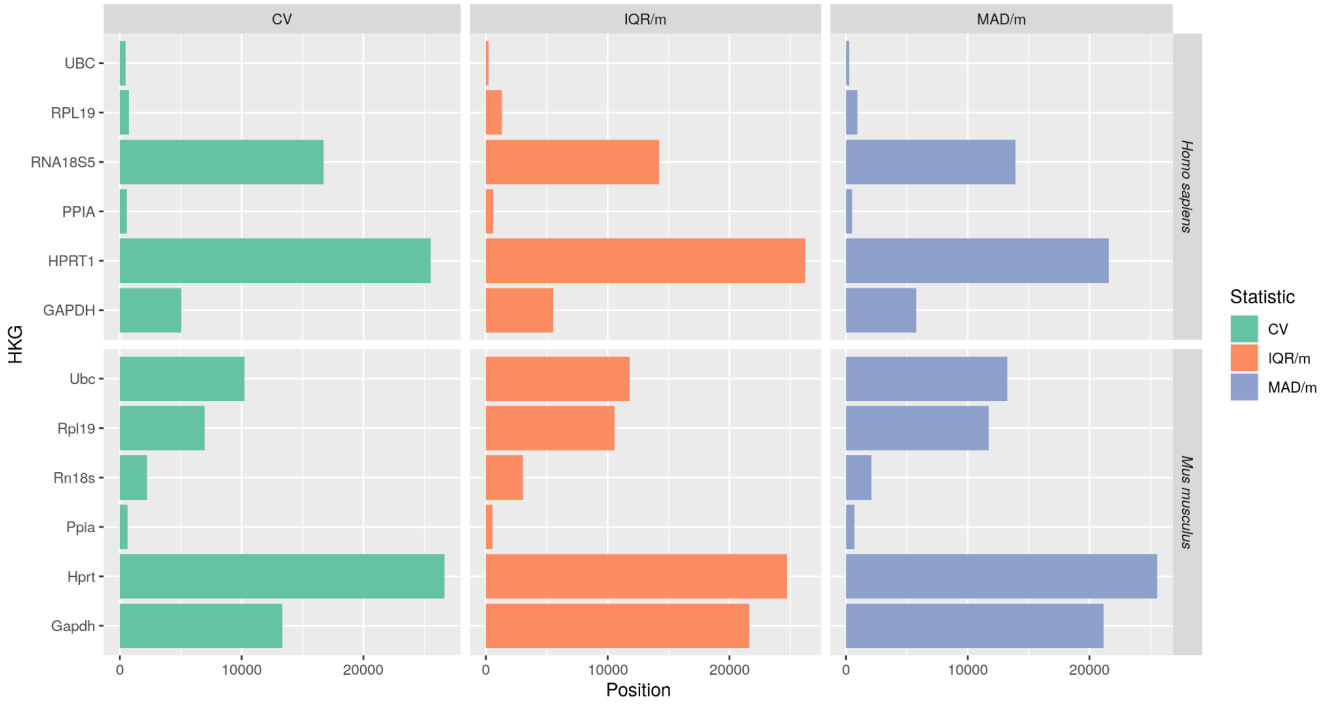

Figure 5. MetaRanking of stability levels for the classic housekeeping group evaluated in Hsa and Mmu. For each of the selected genes, its position in the ranking is described on the X-axis. This MetaRanking has been generated resuming by the mean the obtained RP values of the three statistics (C.V., IQR/median and MAD/median) alter filtering the non-coding genes.

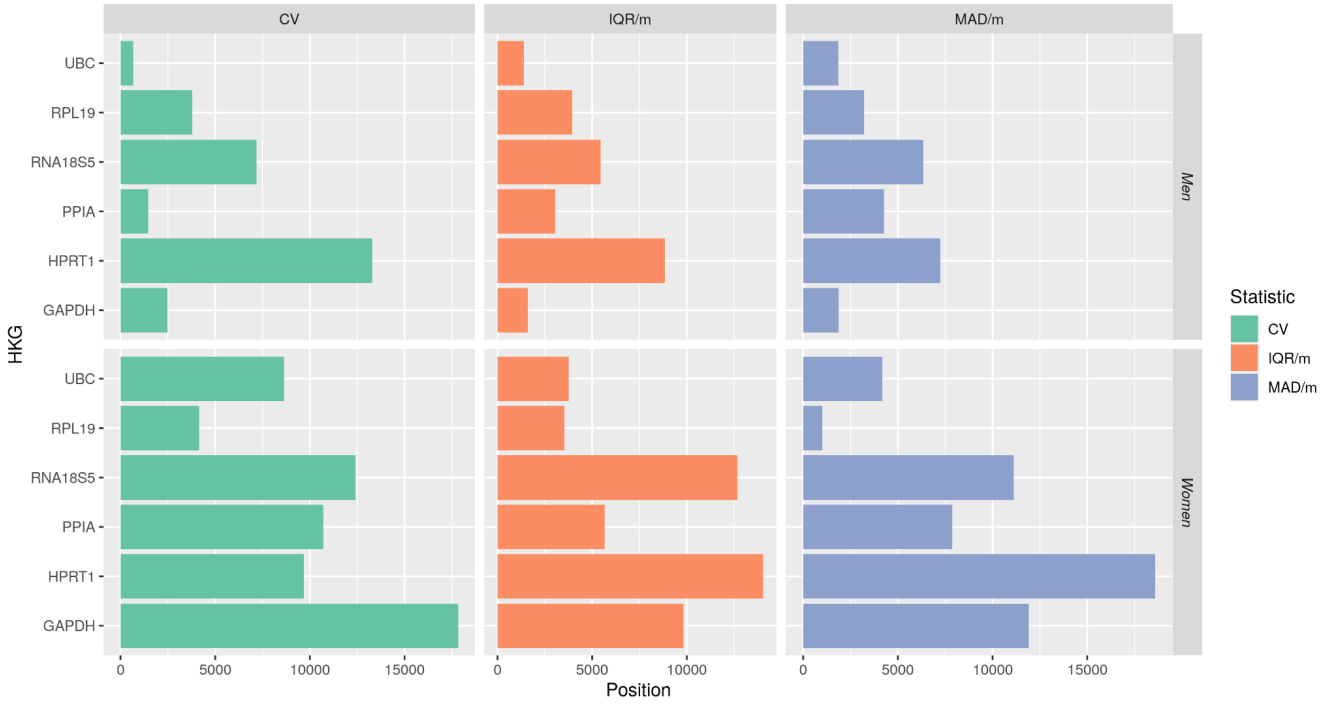

Figure 6. Ranking of stability levels for the classic housekeeping group evaluated in male and females of *Homo sapiens*. For each of the selected genes, its position in the ranking is described on the X-axis. This MetaRanking has been generated resuming by the mean the obtained RP values of the three statistics (C.V., IQR/median and MAD/median) alter filtering the non-coding genes.

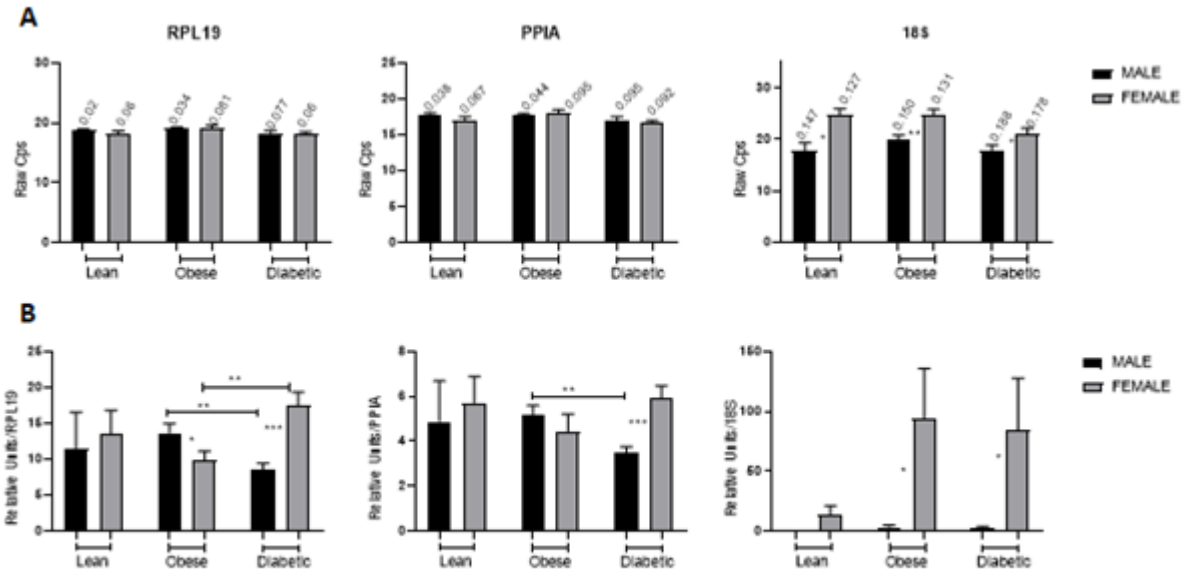

Figure 7. Gene expression analysis in Human adipose tissue from male and female samples using different Housekeeping genes (HKG). (A) Coefficient of variation (CV) in the Cp values of each candidate gene calculated in male and female for lean, obese, and diabetic samples.(B) *IGF1* expression analysis using *RPL19*, *PPIA*, and *18S* as reference genes. Male Lean n=3; Female Lean n=7; Male Obese n=10; Female Obese n=10; Male Diabetic n=12; Female Diabetic n=9. t test student's applied for significance.

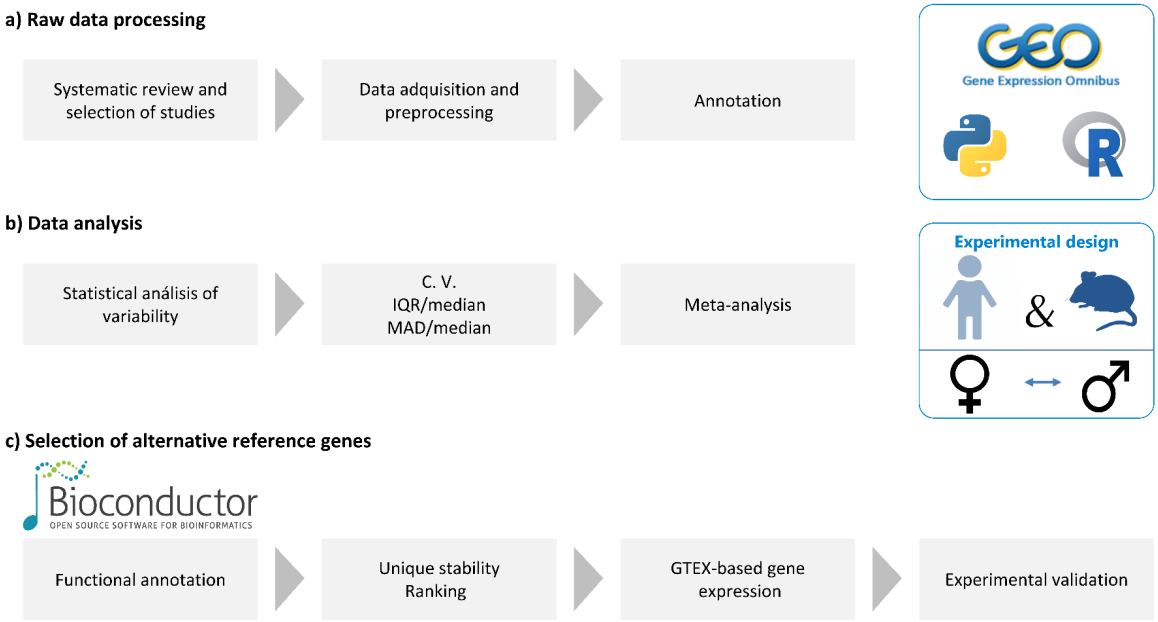

Figure 8. Data-analysis workflow. This study consisted of three main block-steps: A. The collection and pre-processing of public microarray information located at GEO (Gene Expression Omnibus) database with Python and R, B. The statistical data analysis with three different statistics to get the gene expression variability of adipose tissue samples of Hsa and Mmu, considering the biological sex as a variable, and a meta-analysis, and C. The selection of potential

reference genes suitable to compare both sexes in gene expression analyses.

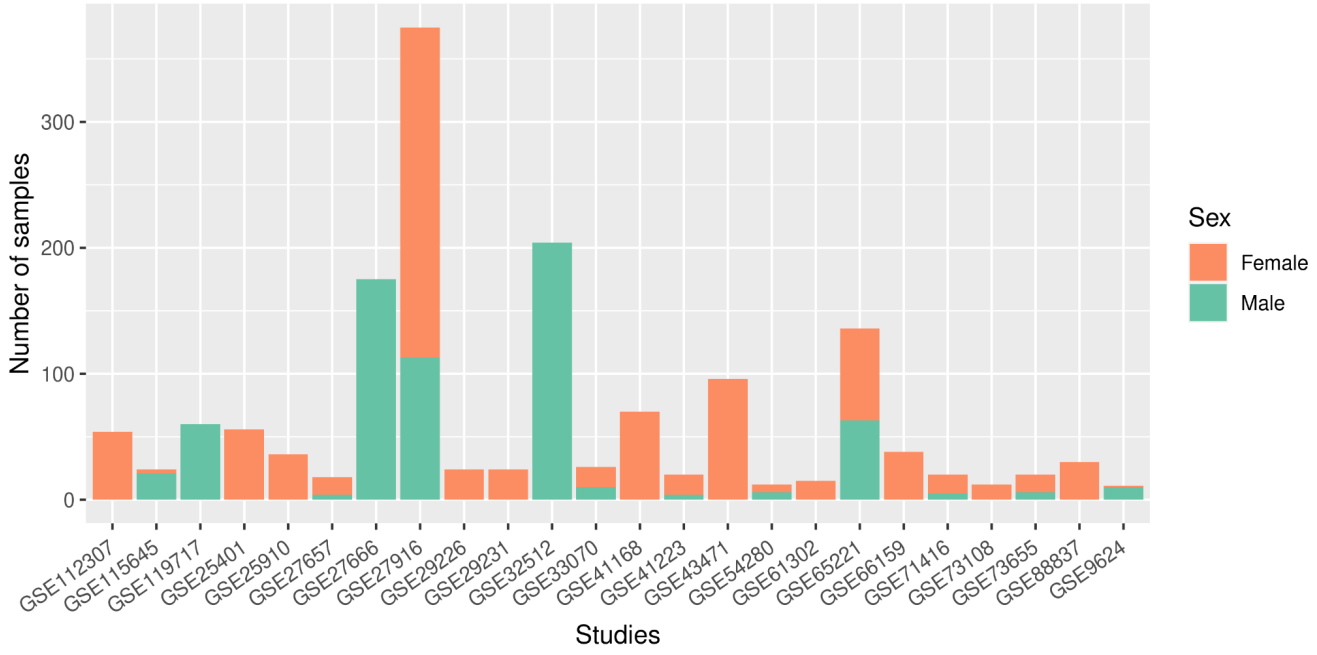

Figure S1. Summary of the number of female and male samples found in each study of Homo sapiens.

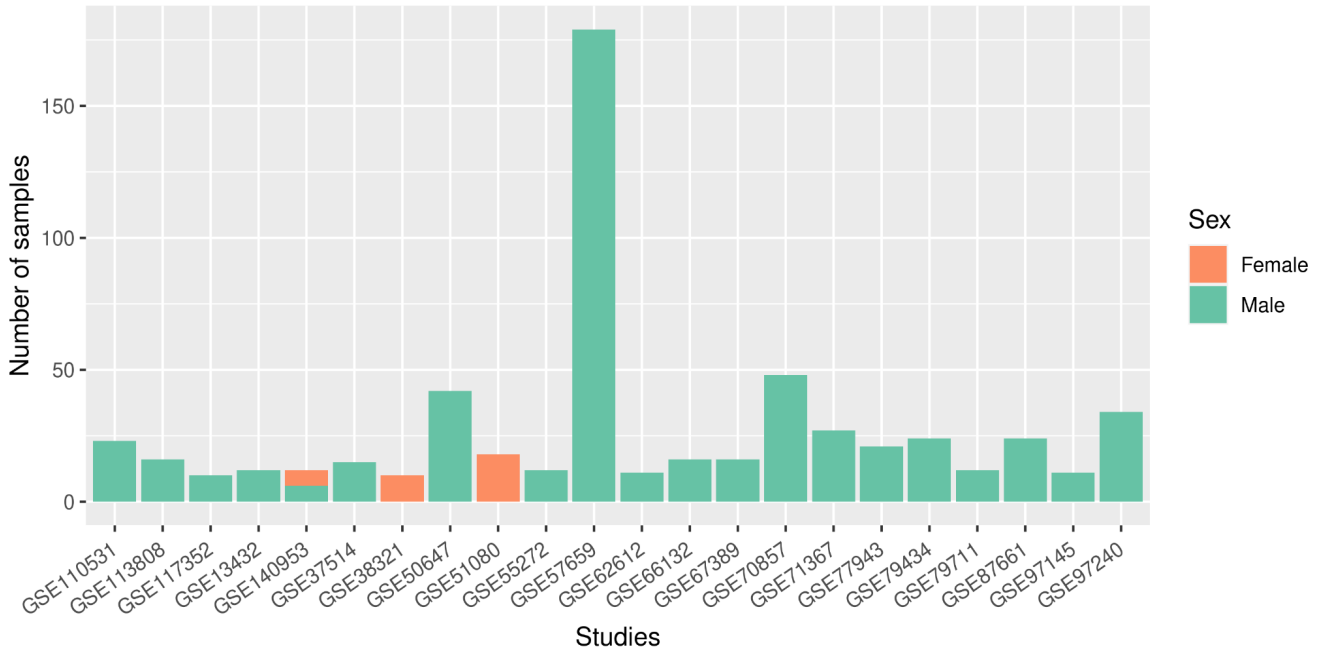

Figure S2. Summary of the number of female and male samples found in each study of Mus musculus. Just one study included samples of both sexes. The majority of the collected samples corresponded to males, evidencing the striking absence of females.

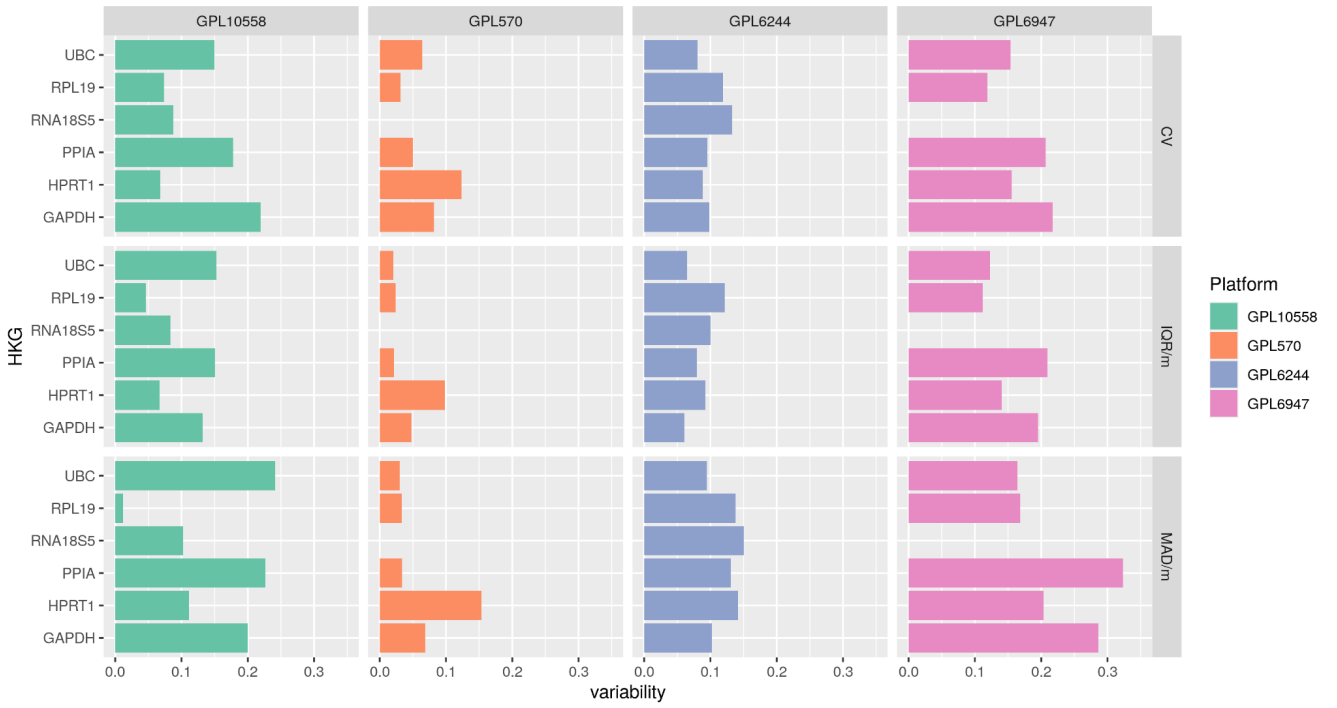

Figure S3. Variability levels for the classic housekeeping group evaluated in females (Human). For each of the housekeeping genes, the variability level found in the selected microarray platforms with the three statistical approaches (C.V., IQR/median and MAD/median) is described on the X-axis.

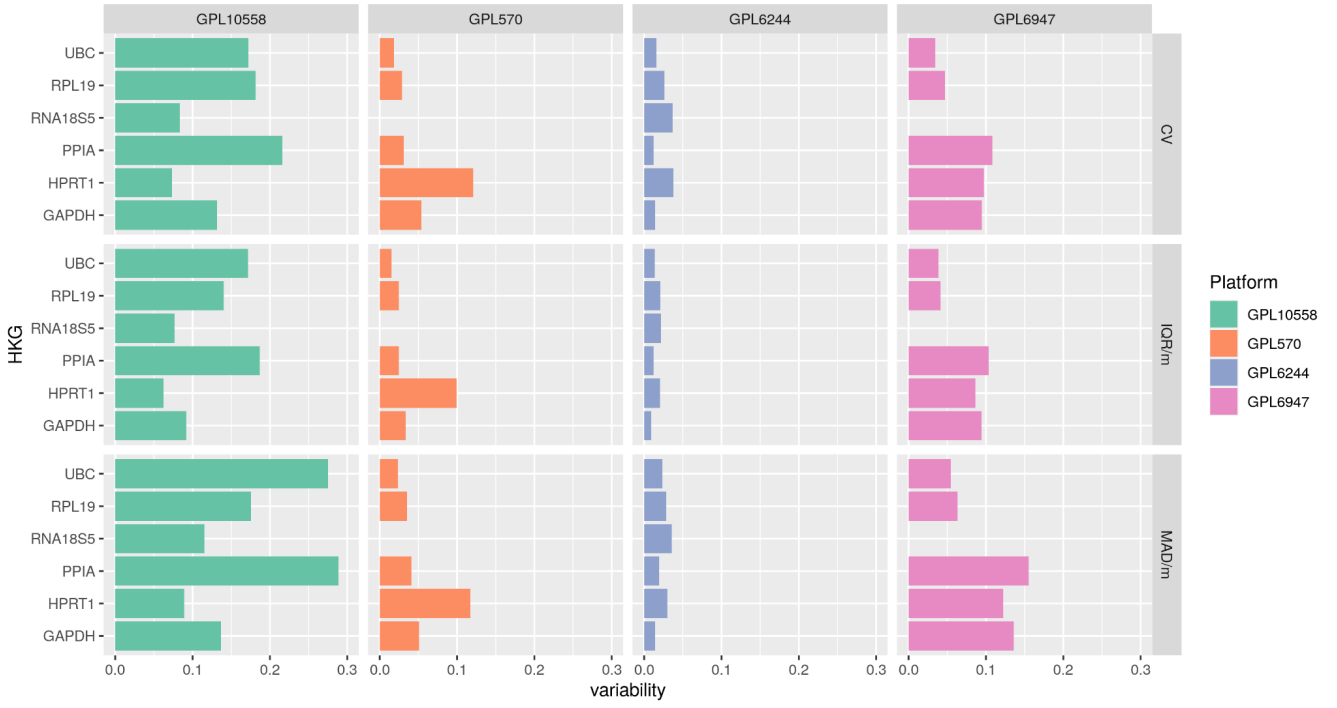

Figure S4. Variability levels for the classic housekeeping group evaluated in males (Human). For each of the housekeeping genes, the variability level found in the selected microarray platforms with the three statistical approaches (C.V., IQR/median and MAD/median) is described on the X-axis.

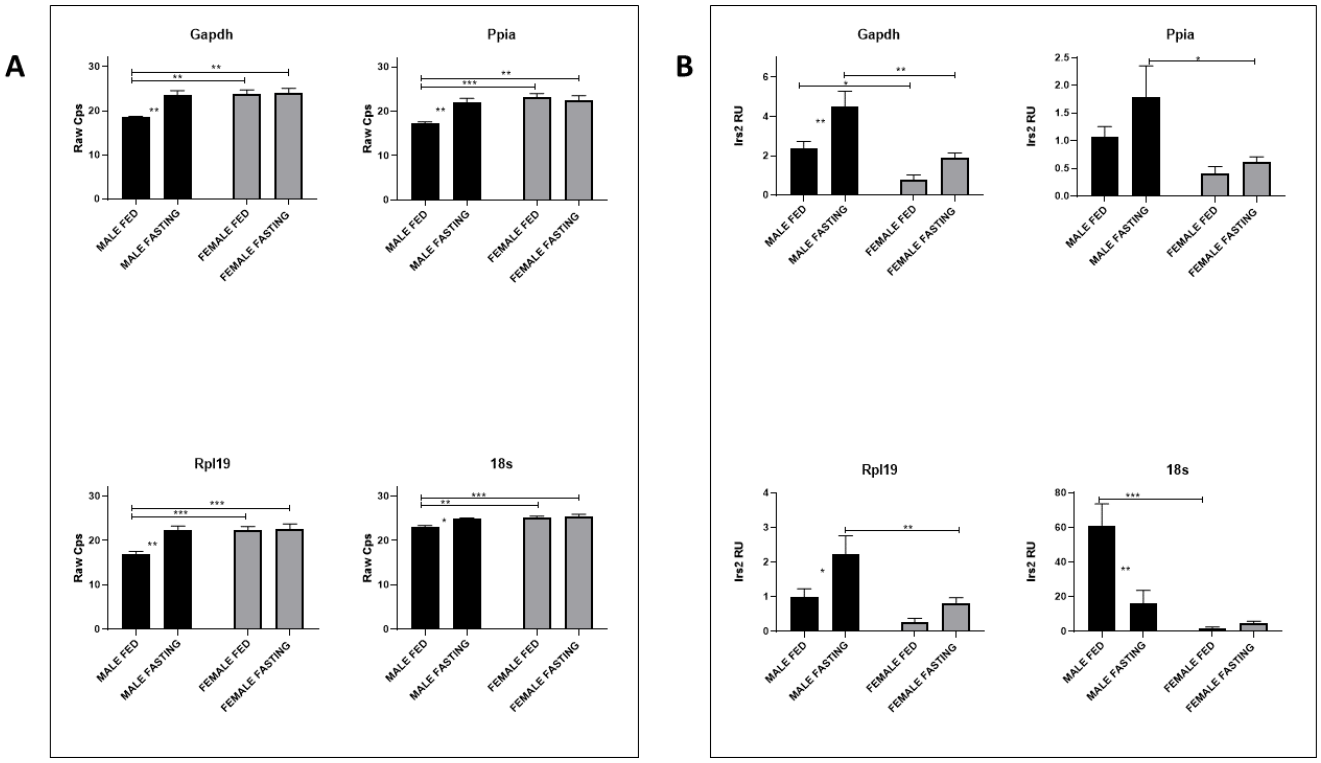

Figure S5. Housekeeping gene candidates analysis in mouse adipose tissue. (A) *Gapdh*, *Ppia*, *Rpl19* and *18s* Raw Cps were analysed using Male and Females in Fed and Fasting conditions. (B) Irs2 gene expression analysis using the same HKG. Relative gene expression may vary depending on the HKG used as reference. Male fed n=6; Male fasting n=4; Female fed=6; Female fasting n=6. 1way ANOVA was performed for statistical analysis. The differences observed were considered significant when: $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).

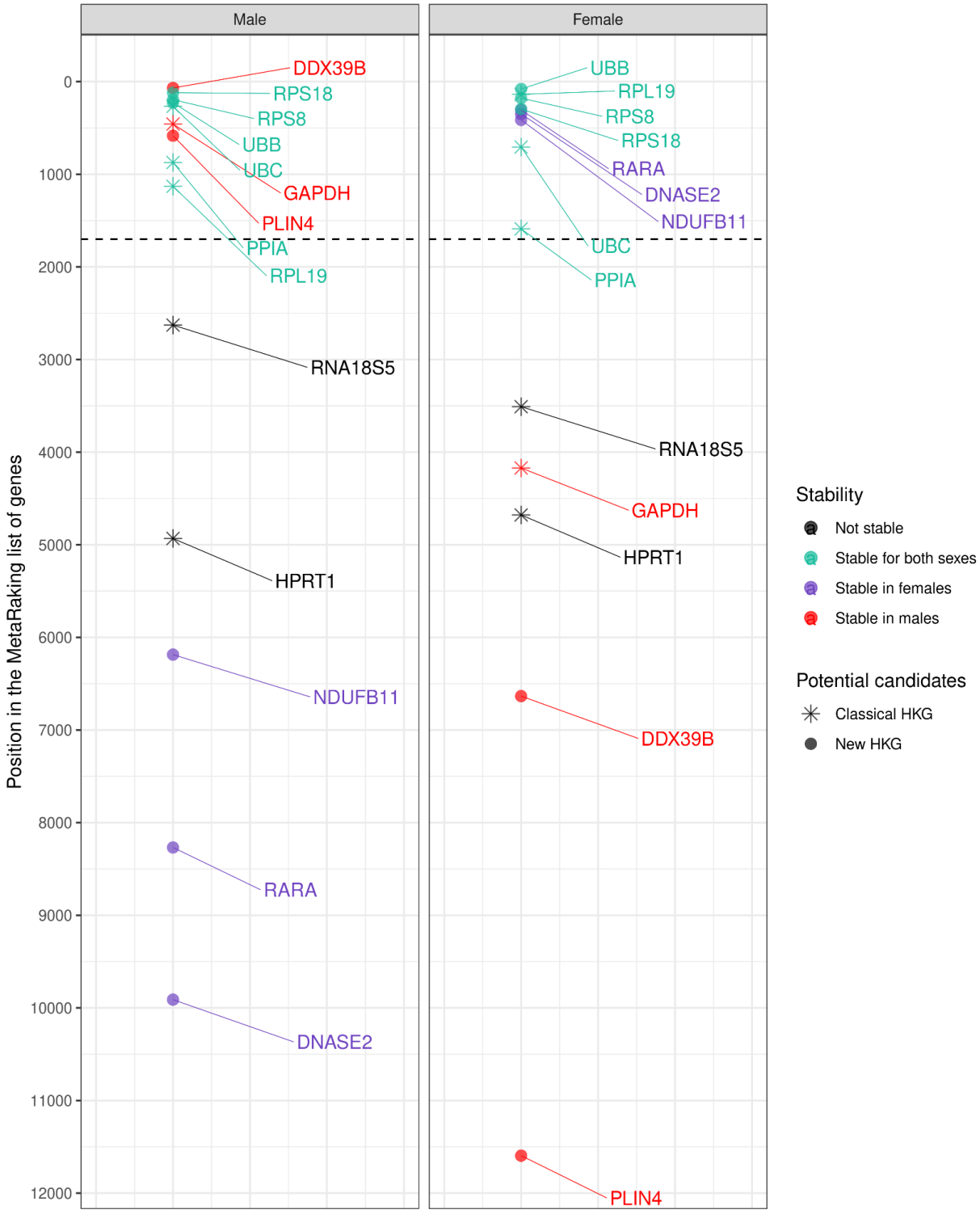

Figure S6. MetaRanking of stability levels housekeeping genes mentioned in this work for human males and females. The dot shape indicates whether they are classical HKG (star) or new putative HKGs proposed (circle). Color indicates if a gene is stable for both sexes (green), stable only in females (violet), stable only in males (red) or not stable (black). Dashed line indicates the limit position of the top 10% most stable genes and with an expression of at least 20 TPM.

Platform	Description	Eligible studies	Included studies	Analyzed samples	Identified genes
GPL570	Affymetrix Human Genome U133 Plus 2.0 Array	37	20	1058	22881
GPL6244	Affymetrix Human Gene 1.0 ST Array transcript (gene) version	15	13	343	23307
GPL10558	Illumina HumanHT-12 V4.0 expression BeadChip	14	7	498	31426
GPL6947	Illumina HumanHT-12 V3.0 expression BeadChip	12	9	825	25159

Table 1. Processed data sets for selected studies of Hsa. For each selected platform, the number of studies that used the platform (eligible studies) are shown, including the number of studies that made the cut (refer exclusion criteria), the number of adipose tissue samples and the maximum number of genes that were able to be identified. A total of 49 studies and 2724 samples have been included in the statistical analysis.

Platform	Description	Eligible studies	Included studies	Analyzed samples	Identified genes
GPL1261	Affymetrix Mouse Genome430 2.0 Array	34	16	280	21495
GPL6246	Affymetrix Mouse Gene 1.0 ST Array transcript (gene) version	24	6	133	24213
GPL6887	Illumina MouseWG-6 v2.0 expression BeadChip	20	8	183	30886
GPL6885	Illumina MouseRef-8 v2.0 expression BeadChip	15	8	375	18120
GPL16570	Affymetrix Mouse Gene 2.0 ST Array transcript (gene) version	10	5	101	24647

Table 2. Processed data sets for selected studies of Mmu. For each selected platform, the number of studies that used the platform (eligible studies) are shown, including the number of studies that made the cut (refer exclusion criteria), the number of adipose tissue samples and the maximum number of genes that were able to be identified. A total of 43

studies and 1072 samples have been included in the statistical analysis.

Gene	**Relative stability Male**	**Relative stability Female**	**Expression level (TPM)**	**Expression level (TPM) in female**	**Expression level (TPM) in male**
PPIA	873	1,589	234.597	236.1	233.6
RPL19	1,129.67	137.33	1,707.61	1707	1708
RPS8	194.67	178.33	952.191	944.5	957.9
RPS18	119.33	296.33	3,173.82	3180	3168
UBB	228	79	252.293	249.8	254.1
UBC	267.33	706.33	432.547	396.9	447.7

Table 3. Selection of housekeeping candidate genes proposed to be used as a reference to compare both sexes in gene expression analysis. PPIA and RPL19 have been experimentally validated, RPS8, RPS18, UBB and UBC are computationally suggested. These genes are proposed based on their sex-specific values of relative expression stability, obtained from the final MetaRanking positions, and the expression levels have been extracted from GTEx, given in TPM (Transcripts Per Million) , which are high enough to be detected by different technologies.

GSE ID	GPL ID	N samples	Male	Female
GSE27657	GPL570	18	4	14
GSE27916	GPL570	375	113	262
GSE41168	GPL570	70	0	70
GSE61302	GPL570	15	0	15
GSE66159	GPL570	38	0	38
GSE71416	GPL570	20	5	15
GSE88837	GPL570	30	0	30
GSE9624	GPL570	11	10	1

GSE25401	GPL6244	56	0	56
GSE25910	GPL6244	36	0	36
GSE33070	GPL6244	26	10	16
GSE73655	GPL6244	20	6	14
GSE41223	GPL6244	20	4	16
GSE54280	GPL6244	12	6	6
GSE73108	GPL10558	12	0	12
GSE65221	GPL10558	136	63	73
GSE119717	GPL10558	60	60	0
GSE115645	GPL10558	24	21	3
GSE43471	GPL6947	96	0	96
GSE32512	GPL6947	204	204	0
GSE29231	GPL6947	24	0	24
GSE29226	GPL6947	24	0	24
GSE27666	GPL6947	175	175	0
GSE112307	GPL6947	54	0	54

Table S1. Distribution of the number of samples by study (GSE ID), platform (GPL ID) and sex for Hsa, in those studies that included the information of the sex variable in the GEO entry.

GSE ID	GPL ID	N samples	Male	Female
GSE117352	GPL1261	10	10	0
GSE140953	GPL1261	12	6	6
GSE110531	GPL1261	23	23	0
GSE66132	GPL1261	16	16	0
GSE77943	GPL1261	21	21	0
GSE97240	GPL1261	34	34	0
GSE71367	GPL1261	27	27	0
GSE67389	GPL1261	16	16	0
GSE51080	GPL1261	18	0	18
GSE13432	GPL1261	12	12	0
GSE38321	GPL6246	10	0	10
GSE79434	GPL6246	24	24	0
GSE55272	GPL6246	12	12	0
GSE37514	GPL6246	15	15	0
GSE113808	GPL6885	16	16	0
GSE70857	GPL6885	48	48	0
GSE57659	GPL6885	179	179	0
GSE97145	GPL6887	11	11	0
GSE62612	GPL6887	11	11	0
GSE50647	GPL6887	42	42	0
GSE87661	GPL16570	24	24	0
GSE79711	GPL16570	12	12	0

Table S2. Distribution of the number of samples by study (GSE ID), platform (GPL ID) and sex for Mmu, *in* those studies that included the information of the sex variable in the GEO entry.

Target	Species	Primer Forward	Primer Reverse
18s	Human	GCAATTATTCCCCATGAACG	GGCCTCACTAAACCATCCAA
	Mouse	AGAAACGGCTACCACATCCA	CATTCCAATTACAGGGCCCG
GAPDH	Human	GTCAGTGGTGGACCTGACCT	CACCACCCTGTTGCTGTAGC
	Mouse	CGTCCCGTAGACAAAATGGT	TCGTTGATGGCAACAATCTC
IGF1	Human	CAGCAGTCTTCCAACCCAAT	ACAGCGCCAGGTAGAAGAGA
IRS2	Mouse	GCCGCCCGAGCAAAGTGACT	ACTACCGCTGGACGGACGCT
PPIA	Human	CCTAAAGCATACGGGTCCTG	TTTCACTTTGCCAAACACCA
	Mouse	AGCATACAGGTCCTGGCATC	TTCACCTTCCCAAAGACCAC
RPL19	Human	CGAATGCCAGAGAAGGTCAC	CCATGAGAATCCGCTTGTTT
	Mouse	GGTGACCTGGATGAGAAGGA	TTCAGCTTGTGGATGTGCTC

Table S4. List of Primers used in the experimental validation. Housekeeping gene candidates analysis in mouse adipose tissue. (A) Gapdh, Ppia, Rpl19 and 18s Raw Cps were analysed using Male and Females in Fed and Fasting conditions. (B) Irs2 gene expression analysis using the same HKG. Relative gene expression may vary depending on the HKG used as reference. Male fed n=6; Male fasting n=4; Female fed=6; Female fasting n=6. 1way ANOVA was performed for statistical analysis. The differences observed were considered significant when: $p<0.05$ (*), $p<0.01$ (**) and $p<0.001$ (***).