Developing and analysing novel tools to study endogenous WNT signalling in mice

van de Moosdijk, A.A.A.

Publication date
2021

Chapter 2

Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland

Anoeska van de Moosdijk and Renée van Amerongen

Swammerdam Institute for Life Sciences, Section of Molecular Cytology, University of Amsterdam, The Netherlands

Adapted from

Scientific Reports (2016) 6:35595

DOI: 10.1038/srep35595

Abstract

Cell growth and differentiation are often driven by subtle changes in gene expression. Many challenges still exist in detecting these changes, particularly in the context of a complex, developing tissue. Quantitative reverse transcription polymerase chain reaction (qRT-PCR) allows relatively high throughput evaluation of multiple genes and developmental time points. Proper quantification of gene expression levels by qRT-PCR requires normalization to one or more reference genes. Traditionally, these genes have been selected based on their presumed "housekeeping" function, with the implicit assumption that they are stably expressed over the entire experimental set. However, this is rarely tested empirically. Here we describe the identification of novel reference genes for the mouse mammary gland based on their stable expression in published microarray datasets. We compared eight novel candidate reference genes (*Arpc3, Clock, Ctbp1, Phf7, Prdx1, Sugp2, Taf11* and *Usp7)* to eight traditional ones (*18S, Actb, Gapdh*, *Hmbs*, *Hprt*, *Rpl13a*, *Sdha* and *Tbp*) and analysed all genes for stable expression in the mouse mammary gland from pre-puberty to adulthood using four different algorithms (GeNorm, DeltaCt, BestKeeper and NormFinder). *Prdx1, Phf7* and *Ctbp1* were validated as novel and reliable, tissue-specific reference genes that outperform traditional reference genes in qRT-PCR studies of postnatal mammary gland development.

Introduction

One of the challenges in developmental biology is to understand how subtle changes in gene expression drive growth and differentiation of complex tissues. The mouse mammary gland serves as a prime example of a dynamic and complex tissue that harbours many different cell types. During embryonic development, the mammary placode develops into a rudimentary mammary epithelial tree (Richert et al., 2000; Watson and Khaled, 2008). After birth, this epithelial tree remains relatively quiescent until the onset of puberty, when the epithelium grows out via branching morphogenesis and invades the surrounding fat pad. By the time the mouse reaches adulthood, the ducts have reached the end of the fat pad. However, dynamic growth and differentiation do not stop at this stage. Adult mice undergo a complete estrus cycle every 4–5 days, and during this time the epithelium also expands and regresses. Furthermore, during pregnancy the gland forms large alveolar structures that will produce milk during lactation once fully differentiated. After lactation, the mammary gland involutes and is reshaped back to a virgin-like state. This cycle can

repeat itself multiple times throughout the reproductive life of an animal. As a result, during postnatal development the mammary gland does not only dramatically change in size, but also in cell type composition and differentiation status (Macias and Hinck, 2012; Visvader and Stingl, 2014).

Quantitative reverse transcription polymerase chain reaction (qRT-PCR) is a highly sensitive method for measuring changes in gene expression across multiple experimental conditions. However, even with current technologies it remains challenging to compare different developmental timepoints (Dhorne-Pollet et al., 2013; Li et al., 2016; Liu et al., 2014; Nakamura et al., 2016; Wu et al., 2016; Yang et al., 2015; Ye et al., 2015), due to changes in tissue composition. To control for technical errors, individual samples are typically normalized against the expression of one or more internal reference genes. The ideal reference gene shows stable expression across the entire experimental dataset and is not influenced by any of the experimental parameters. This ensures that changes in reference gene expression are only due to technical variation such as pipetting errors and differences in sample input. When analysing genes that show subtle expression differences, as is the case for many developmental regulators, even small changes in reference gene expression can lead to aberrant results. It is therefore important to use the most stably expressed reference possible. Including more than one reference gene for the analyses further minimizes the effect of individual reference gene expression variation. While not yet commonplace, the use of at least three *bona-fide* reference genes is advised for proper normalization (Vandesompele et al., 2002).

Reference genes have historically been picked based on their presumed "housekeeping" function, ensuring that they would be abundantly expressed and easily detectable. However, traditionally used reference genes such as *Gapdh, 18S* and *Actb* can change expression in response to experimental treatment and might therefore not be appropriate to use under all circumstances (Schmittgen and Zakrajsek, 2000). Attempts have been made to identify universal reference genes, which could be applied to any sample of interest irrespective of its developmental origin, by comparing published datasets of multiple different human tissues and cell lines (Jin et al., 2004; de Jonge et al., 2007; Kwon et al., 2009; Lee et al., 2007; Szabo et al., 2004). However, so far this has failed to yield a consistent list of candidate genes, raising the question whether such universal references exist at all (Zhang et al., 2015). Therefore, finding the best performing reference genes requires a dedicated effort focussing on the specific tissue or organism of interest (Cieslak et al., 2015; Dhorne-Pollet et al., 2013; Franzellitti et al., 2015; Petriccione et al., 2015; Schliep et al., 2015; Thomas et al., 2014; Yang et al., 2015).

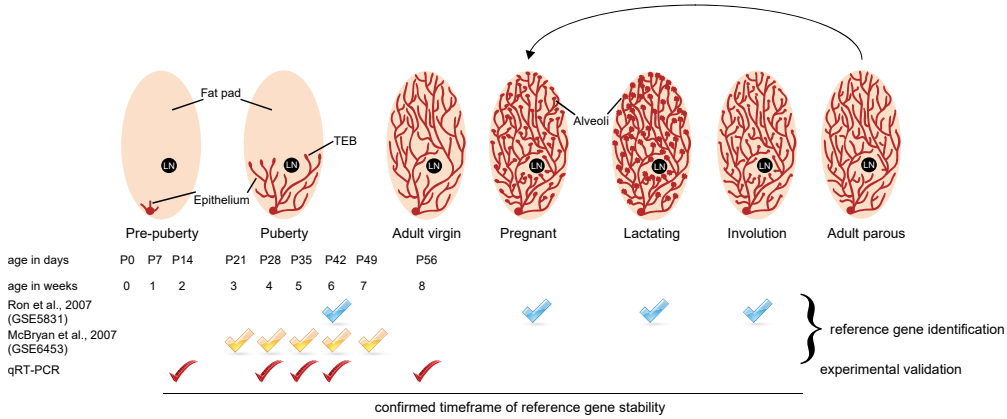


Figure 1. Postnatal development of the mouse mammary gland. Top: scheme illustrating developmental dynamics of the mammary gland. Before puberty, the fat pad only contains a rudimentary epithelial tree that starts from the nipple. During puberty, the epithelial tree invades the surrounding fat pad. This process is driven by highly proliferative structures called terminal end buds (TEBs). At the start of adulthood, the epithelial tree has fully invaded the fat pad and the TEBs regress. During each pregnancy, the gland forms alveoli for milk production during lactation. After weaning of the pups, the epithelial tree reverts back to a virgin-like state in a process called involution. LN= lymph node. Bottom: overview of the experimental time points used to generate the microarray expression data sets GSE5831 and GSE6453, as well as the time points used for qRT-PCR validation experiments.

Because of the dynamic growth and differentiation properties of the mammary gland, it is particularly challenging to find genes that can serve as reliable reference genes across different stages of postnatal development. In 2010, Han and colleagues ranked the stability of eight commonly used reference genes for qRT-PCR studies of the mouse mammary gland (Han et al., 2010). However, their study focussed on the pregnancy and lactation stages in the adult only. Moreover, a single algorithm was used to rank the genes, while several algorithms are currently available to provide a more comprehensive overview of reference gene stability and sensitivity (Andersen et al., 2004; Pfaffl et al., 2004; Silver et al., 2006; Vandesompele et al., 2002; Xie et al., 2012).

Here we describe the identification of novel candidate reference genes for the mouse mammary gland based on their stable expression in multiple microarray datasets. We validate *Ctbp1*, *Phf7* and *Prdx1* as tissue-specific reference genes. They show stable expression across multiple stages of postnatal mammary gland development and allow the detection of subtle changes in *Wnt* gene expression in whole mammary gland RNA preparations. As such, they outperform traditionally used reference genes and are exquisitely suited for comparative qRT-PCR analyses in this complex developing tissue.

Results

Selection of novel candidate reference genes from published microarray data

The mammary gland is composed of multiple different cell types, including basal and luminal epithelial cells, stromal fibroblasts and adipocytes, as well as cells contributing to other structures in the gland such as nerves and blood vessels. In addition, the gland shows dynamic growth and differentiation during puberty, pregnancy and lactation, as well as extensive tissue remodelling during involution (Figure 1). To date, no dedicated attempt has been made to determine whether suitable reference genes exist for use in qRT-PCR studies of whole mammary gland preparations across all stages of postnatal development. Therefore, we set out to identify novel, tissue-specific candidate reference genes that are stably expressed from pre-puberty through puberty and adulthood by analysing published expression datasets.

We identified two gene expression profiling studies in the NCBI gene expression omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) that compared different stages of mammary gland development (Figure 1). Both datasets (GSE5831 (Ron et al., 2007) and GSE6453 (McBryan et al., 2007)) used microarrays to analyse the whole mammary gland transcriptome, but the studies were performed by different labs, on different mouse strains (C57BL6/J and CD-1, respectively) and using different array platforms. Furthermore, one experiment (GSE5831) compared adult stages (6 week old virgins, pregnancy day 14, lactation day 10 and involution day 4), whereas the other (GSE6453) compared pubertal time points (3, 4, 5, 6 and 7 week old females). We hypothesized that any gene that would be stably expressed across all of these arrays could qualify as a suitable reference gene. To identify these genes, we compared physiologically divergent stages using GEO2R, an interactive webtool that allows gene expression analysis of published microarray datasets using the GEOquery and limma R packages (Sean and Meltzer, 2007).

First, we performed differential gene expression analysis of GSE5831 for the virgin and lactating, terminally differentiated gland. We selected genes with a log fold change (logFC) between -0.1 and 0.1, thus reflecting stable expression between the two conditions (Figure 2a). Next, we performed a similar analysis for the pregnant (actively growing) and involuting (regressing) gland (Figure 2b). Of the 852 and 1754 genes that showed stable expression in the virgin/lactating and pregnant/involuting comparisons, respectively, only a subset (194 genes) approached a logFC of 0 in both.

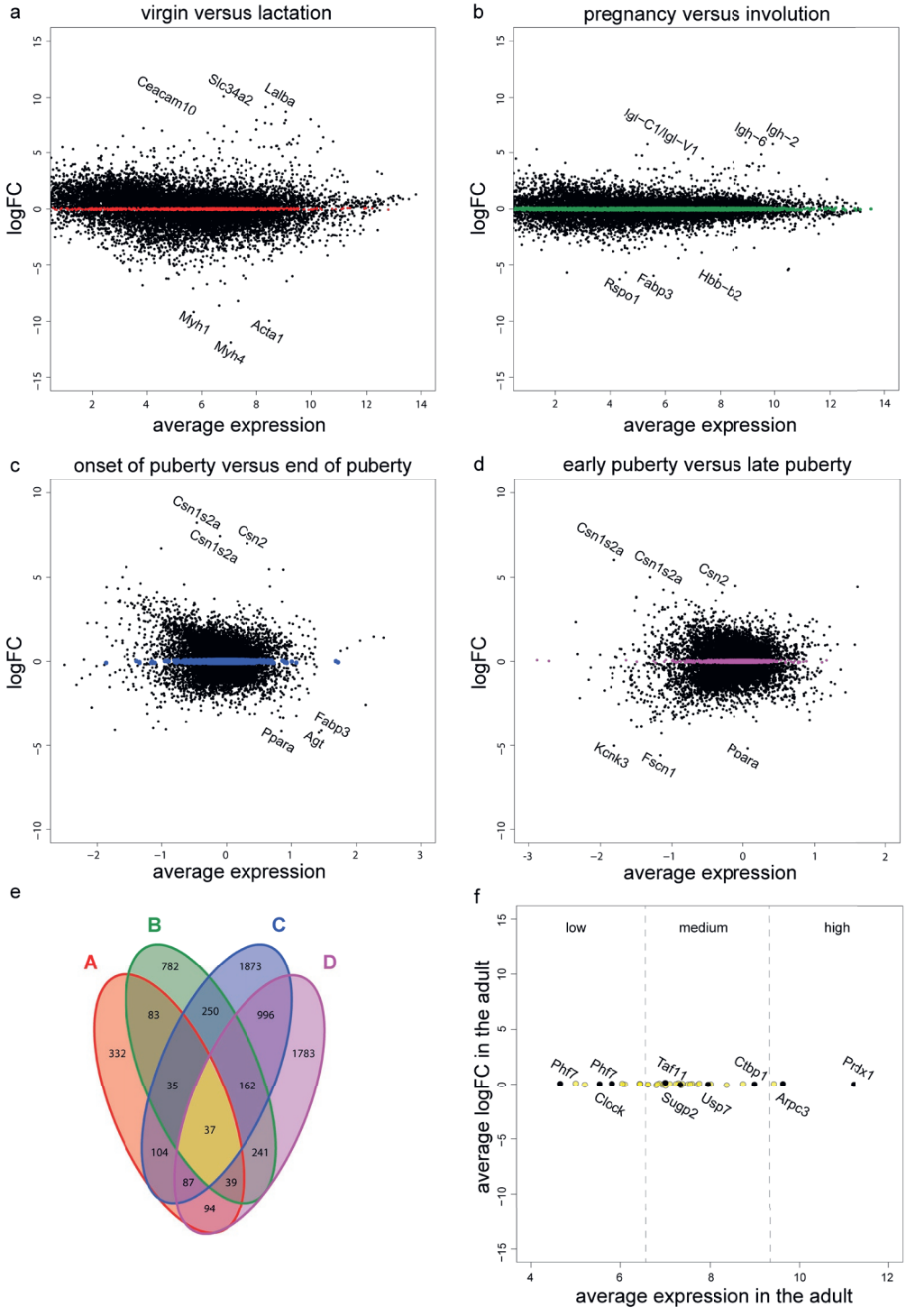

a
virgin versus lactation
Ceacam10
Slc34a2
Lalba
Myh1
Myh4
Acta1
logFC
average expression
b
pregnancy versus involution
Igl~C1/Igl~V1
Igh~6
Igh~2
Rspo1
Fabp3
Hbb~b2
logFC
average expression
c
onset of puberty versus end of puberty
Csn1s2a
Csn2
Csn1s2a
Fabp3
Ppara
Agt
logFC
average expression
d
early puberty versus late puberty
Csn1s2a
Csn1s2a
Csn2
Kcnk3
Fscn1
Ppara
logFC
average expression
e
A
B
C
D
782
1873
83
250
996
332
1783
35
162
37
104
241
87
39
94
f
low
medium
high
Phf7
Phf7
Clock
Taf11
Sugp2
Usp7
Ctbp1
Arpc3
Prdx1
average logFC in the adult
average expression in the adult

Figure 2. Identification of novel candidate reference genes. (a-d) MA plots for GSE5831(a-b) and GSE6453 (c-d), comparing the virgin versus the lactating gland (a), the pregnant versus the involuting gland (b), the onset versus the end of puberty (3 versus 7 weeks, c) and early versus late puberty (4 weeks versus 6 weeks, d). The top 3 up- and downregulated genes for each of the four comparisons are indicated. Genes with a log fold change (logFC) between -0.1 and 0.1 are highlighted. (e) Venn diagram showing the overlap of genes with a logFC between -0.1 and 0.1 identified in a-d. Only 37 genes are shared between all four comparisons. (f) Ranking of the 37 genes with a logFC between -0.1 and 0.1 according to their average expression level in the GSE5831 dataset. The names of the candidate reference genes that were picked for further testing are highlighted.

To narrow down the number of potential candidate reference genes, we then performed differential gene expression analysis of GSE6453 for different pubertal stages, comparing the start and end of puberty (3-week versus 7-week old mice, Figure 2c) as well as early and late puberty (4-week versus 6-week old mice, Fig. 2d). This resulted in a list of 37 genes that showed stable expression across all of the tested conditions (Figure 2e and Supplementary Table S1). From this list, we selected a final set of 8 putative reference genes (*Arpc3*, *Clock*, *Ctbp1*, *Phf7*, *Prdx1*, *Sugp2*, *Taf11* and *Usp7*, Figure 2f).

We made certain that the resulting candidates represented a variety of expression levels (Figure 2f) and divergent functions as predicted by gene ontology analysis (Supplementary Table S1), and ensured that gene-specific qRT-PCR primers with unique, intron-spanning sequences and sufficient amplification efficiency could be designed (Supplementary Table S2).

Variation in expression levels of candidate reference genes

Previously, Han and colleagues investigated the stability of eight commonly used reference genes in the pregnant and lactating mouse mammary gland (Han et al., 2010). To determine whether any of these genes are also stably expressed over a larger developmental timeframe, we selected four of the most stable genes identified in this study (*Rpl13a, Hprt, Gapdh* and *Actb*) as well as four additional commonly used reference genes (Kosir et al., 2010; Silberstein et al., 2006; Thellin et al., 1999) (*18S, Hmbs, Tbp* and *Sdha*) as a "traditional reference gene set".

To test the variation in expression levels of the traditional and novel candidate reference genes over a large developmental time course, we performed qRT-PCR analyses on cDNA synthesized from whole mouse mammary gland RNA preparations ranging from pre-puberty (P14) to early, mid- and late puberty (P28, P35 and P42) and adult virgin FVB mice (P56). When the expression levels (Ct values) of all sixteen (candidate) reference genes are compared, variable expression patterns across the different stages of development emerge for both the traditional (Figure 3a) and the newly identified (Figure 3b) genes. Some genes show a relatively

stable pattern throughout, with all Ct values clustering closely together irrespective of the developmental time point. For instance, across pre-puberty and adulthood the Ct values for *Prdx1* and *Ctbp1* show a total spread of only 2.0 and 1.9 cycles, respectively. Other genes show more variation and are therefore probably less suitable as reference genes for our purpose. As an example, the expression of *Hmbs* and *Actb* shows a spread of 8.0 and 7.7 cycles, respectively. The 3rd thoracic gland was analysed as a biological replicate and shows a similar pattern (Supplementary Figure S1).

GeNorm stability analysis of candidate reference genes

Reference genes can be ranked based on their expression stability over all samples using the GeNorm algorithm (Vandesompele et al., 2002). The underlying principle of this algorithm is that ideal reference genes show stable expression patterns regardless of the experimental parameters (in our case: different developmental time points). When multiple *bona-fide* reference genes are analysed, their average expression should best represent the "true" normalisation factor. Therefore, GeNorm not only scores for stable expression per candidate, but also for gene expression variation of each individual reference compared to the average reference gene expression. By performing a pairwise variation of any candidate reference gene compared to the average expression of the entire pool, a so-called M-value is calculated for each gene, which serves as a measure for gene stability: the lower the value, the more stably a gene is expressed. Next, the genes are ranked based on M-value, by performing stepwise exclusion of the gene with the highest M-value (i.e. poorest stability). M-value calculation and stepwise exclusion are repeated until only two genes remain. These final two genes are ranked based on their individual stability value.

M-values should be as low as possible. However, they strongly depend on the complexity of the experimental samples and fixed criteria or cut offs do not exist (Hellemans et al., 2007). For homologous samples, such as cell lines, the M-value can drop below 0.2. In contrast, for complex tissues M-values of 0.5 or even >1.0 have been reported (Liu et al., 2014; Martins et al., 2016; Schliep et al., 2015; Yang et al., 2015; Ye et al., 2015). When the traditional gene set was analysed, *Rpl13a, Hprt* and *Gapdh* came out as the best references, with an M-value of 0.62 for *Rpl13a* (Figure 4a). GeNorm analysis for the novel candidates revealed *Phf7, Prdx1* and *Ctbp1* as best performing references in this set (Figure 4b). The M-values for these genes (0.49, 0.51 and 0.53, respectively) are lower than the best M-value for the traditional gene set. Combined analysis of both novel (*Arpc3*, *Clock*, *Ctbp1*, *Phf7*, *Prdx1*, *Sugp2*, *Taf11* and *Usp7*) and traditional (*18S*, *Actb*, *Gapdh*, *Hbms*, *Hprt*, *Rpl13a*, *Sdha* and *Tbp*) reference genes confirms that *Phf7, Prdx1* and *Ctbp1* rank as the most stable genes across all developmental time points analysed, outperforming *Rpl13a, Hprt* and *Gapdh* (Figure 4c).

Comprehensive ranking of *Prdx1*, *Phf7* and *Ctbp1* as the most stable reference genes in the postnatal mammary gland

For a more comprehensive overview of reference gene suitability, we used the web-based tool RefFinder (Xie et al., 2012) to rank the candidate reference genes according to three additional algorithms: DeltaCt (Silver et al., 2006) , BestKeeper (Pfaffl et al., 2004) and NormFinder (Andersen et al., 2004). DeltaCt ranks genes based only on the spread in Ct values (similar as depicted in Figure 3), considering the gene with the least variation to be the best ranking reference. BestKeeper performs pairwise variations for all possible pairs of reference genes and calculates a geometric mean to determine which genes have highly similar expression profiles, creating a ranking based on the fit of the geometric means. NormFinder analyses the variance in expression data of all given candidate reference genes, with the assumption that an ideal reference gene should only differ mildly from the overall estimate.

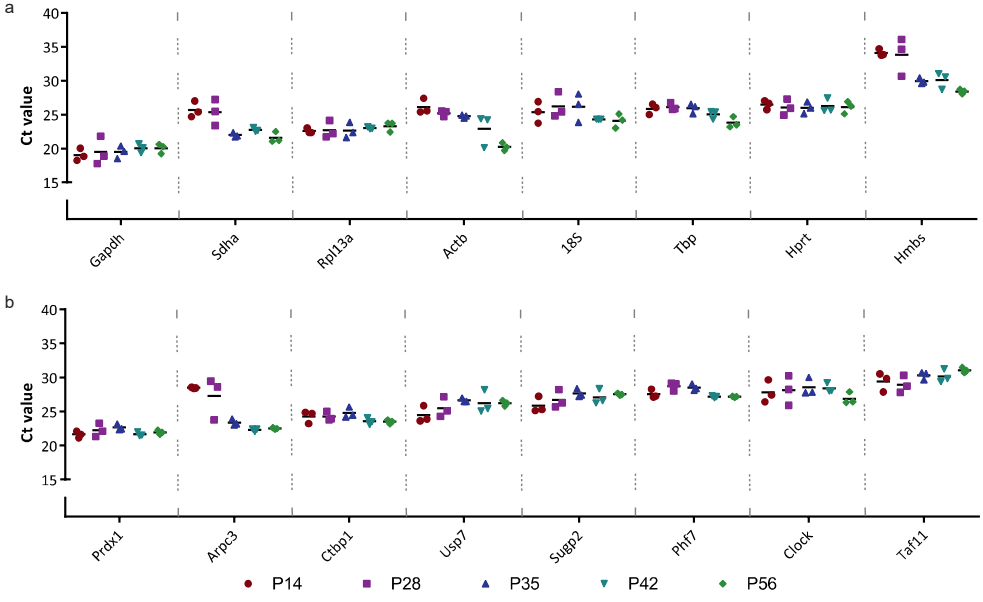


Figure 3. Ct values of reference genes for biological replicas of #4 mammary gland. Scatter plot depicting Ct values for all sixteen reference genes in the 4th inguinal mammary gland over a developmental time series from pre-puberty (P14) to puberty (P28, P35 and P42) and adult (P56). Each data point represents a single biological replicate. a) Traditionally used reference genes. b) Novel candidate reference genes. Genes were plotted in order of their absolute expression levels, from high expression (lowest Ct value) to low expression (n=3 biological replicates for each developmental time point).

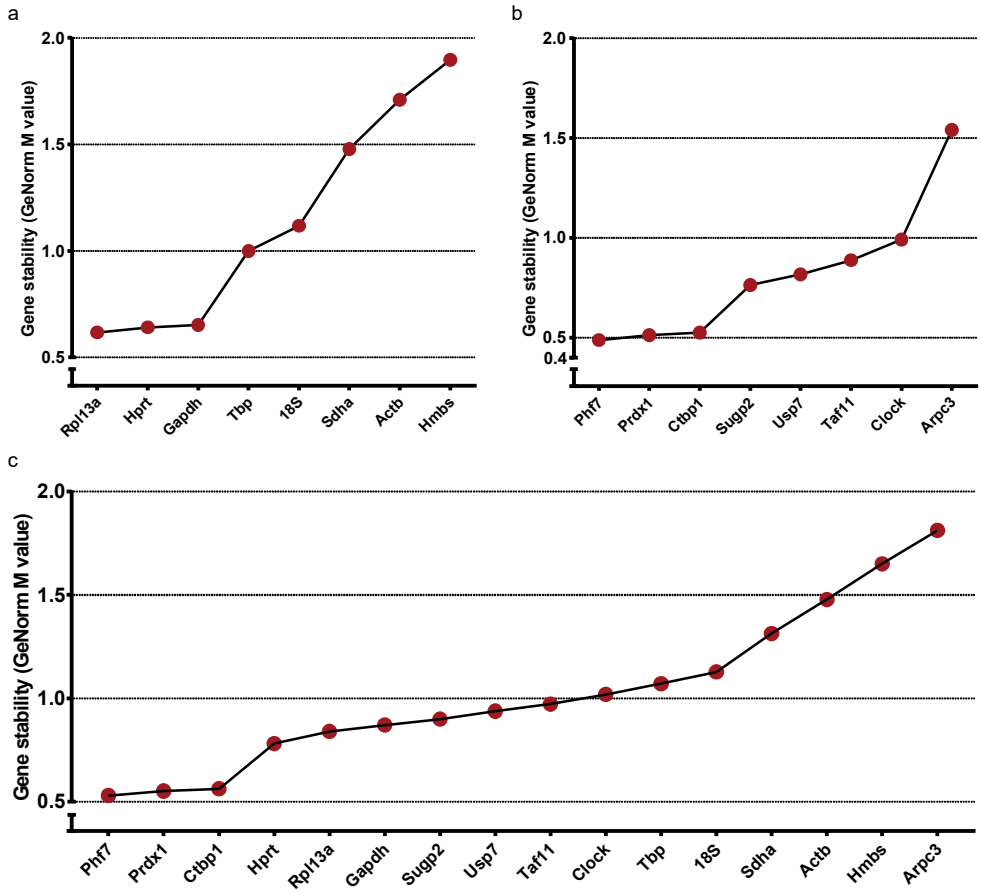


Figure 4. Reference gene ranking according to GeNorm. Graph showing candidate reference gene ranking according to the M-value for gene expression stability in the developing mammary gland. a) Ranking of eight traditional reference genes. b) Ranking of eight novel candidate reference genes. c) Comprehensive ranking of the traditional and candidate reference gene sets. *Ctbp1*, *Phf7* and *Prdx1* have a lower M-value, indicating higher stability.

The different approaches from the above-mentioned algorithms led to small differences in their respective reference gene rankings, but ideally the overall trend should be the same. Table 1 shows the comprehensive ranking of all novel candidates and traditional reference genes for the different algorithms. Indeed, the results confirm the GeNorm analysis: *Arpc3, Hmbs, Actb* and *Sdha* rank poorly for all algorithms and are thus considered to be less suitable reference genes in all tests. In contrast, *Prdx1 Phf7* and *Ctbp1* are ranked in the top four in all tests, complemented with either *Tbp*, *Hprt* or *Rpl13a* depending on the algorithm. A comprehensive

ranking of the results from all four algorithms thus confirms *Prdx1*, *Phf7* and *Ctbp1* as the top three reference genes (Table 1 and Supplementary Figure S2).

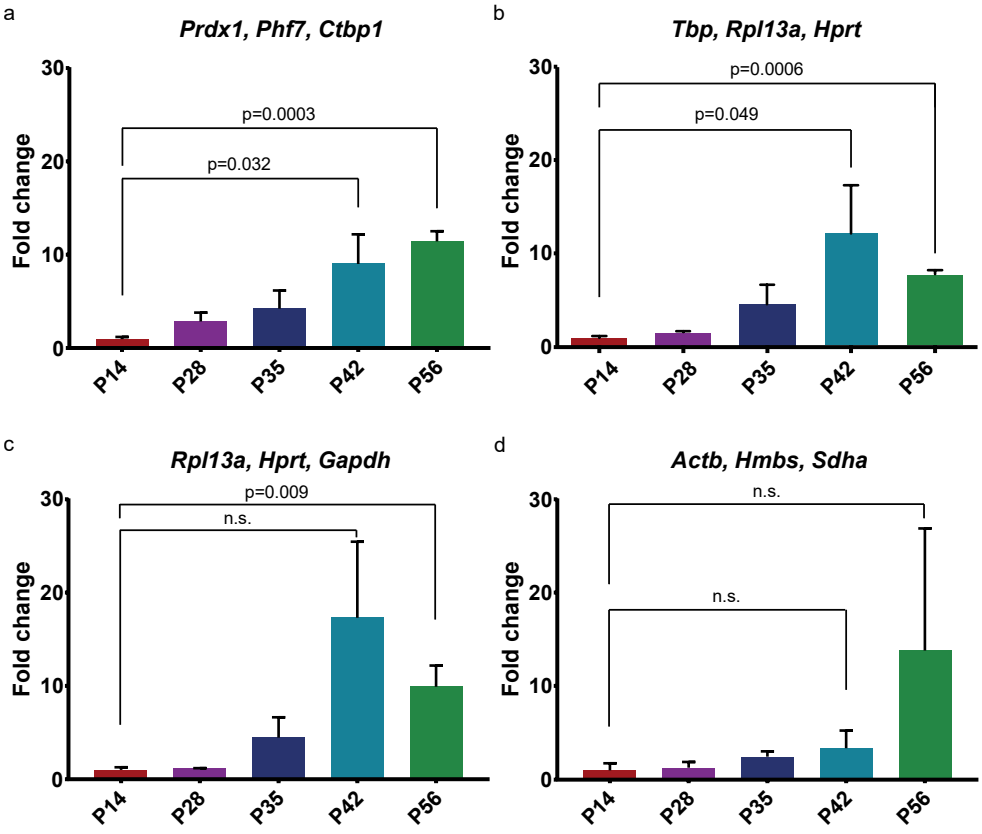


Figure 5. Reference gene validation. Bar graphs showing qRT-PCR analysis *Wnt4* expression in the developing mouse mammary gland between P14 and P56. The lowest value (P14) was set to 1. a) Relative *Wnt4* expression normalised to *Prdx1*, *Phf7* and *Ctbp1*, which were identified as novel reference genes in this study. b) Relative *Wnt4* expression normalised to *Rpl13a*, *Hprt* and *Tbp*, which were identified as the best traditional reference genes in this study. c) Relative *Wnt4* expression normalised to *Rpl13a*, *Hprt* and *Gapdh*, which were previously identified as the best reference genes for the mouse mammary gland (Han et al., 2010). c) Relative *Wnt4* expression normalised to *Actb*, *Hmbs* and *Sdha*. Although frequently used as references, these genes rank poorly in our stability analyses. Statistically significant differences are indicated with their respective p-values. Error bars represent SEM (n=3 biological replicates for each developmental time point). n.s. = not statistically significant.

Validation of novel reference genes by qRT-PCR analysis of *Wnt* gene expression

To validate that *Prdx1, Phf7* and *Ctbp1* can be used as reliable reference genes for the postnatal mouse mammary gland, we tested their ability to reveal subtle changes in gene expression across mammary gland development. For this purpose, we focused on *Wnt* gene expression. WNT signalling is crucial for mammary gland development and function, and WNT-responsive stem cells reside in the mammary gland from an early time point on (van Amerongen et al., 2012; De Visser et al., 2012; Wang et al., 2015b). Multiple *Wnt* genes have previously been shown to be expressed in the mammary gland, but so far, *Wnt* gene expression has mainly been studied using semi-quantitative methods (Silberstein et al., 2006; Weber-Hall et al., 1994). Furthermore, *Wnt* genes are typically expressed at relatively low levels, making it challenging to detect expression level changes in RNA preparations from the whole mammary gland. *Wnt4* is known to be expressed in the adult mammary gland, where it plays a role in the estrus cycle (Joshi et al., 2010) and in side-branching during pregnancy (Brisken et al., 2000). A recent study also revealed a role for *Wnt4* in branching morphogenesis during puberty (Rajaram et al., 2015), but if and how *Wnt4* expression levels change during postnatal mammary gland development remains incompletely understood. To this end, we quantified *Wnt4* expression across pre-puberty, puberty and adulthood. Only when *Prdx1, Phf7* and *Ctbp1* (our set of novel and preferred reference genes) are used for normalization, a steady increase in *Wnt4* levels is revealed between P14 and P56 (Figure 5a). Subtle changes in *Wnt4* expression are also apparent when *Tbp*, *Rpl13a* and *Hprt* (the most reliable traditional reference genes) are used as a reference geneset (Figure 5b). Using a set of even slightly less stable reference genes (*Rpl13a, Hprt* and *Gapdh*) already obscures this trend (Figure 5c). Finally, when we normalized *Wnt4* expression to the three commonly used reference genes that were least stable in our analysis (*Actb, Hmbs* and *Sdha*), no statistically significant changes can be detected between the different developmental stages at all (Figure 5d). This underscores the importance of using properly validated reference genes for the tissue of choice and confirms that the new reference gene set comprised of *Prdx1*, *Phf7* and *Ctbp1* allows small changes in developmental gene expression to be detected with high precision in a complex mammary tissue preparation.

Table 1. Reference gene ranking according to four algorithms. DeltaCt, BestKeeper, NormFinder and GeNorm were used to rank reference genes for suitability. Ranking was performed in RefFinder (DeltaCt, BestKeeper and NormFinder) or qBase+ (GeNorm). Individual rankings are shown in the top four rows. The overall comprehensive ranking is shown in the bottom row. A similar analysis performed separately on the traditional and the novel reference gene sets is presented in Supplementary Figure S2.

Ranking	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Delta CT	Prdx1	Ctbp1	Phf7	Tbp	Hprt	Rpl13a	Sugp2	Gapdh	Clock	18S	Usp7	Taf11	Sdha	Actb	Hmbs	Arpc3
BestKeeper	Prdx1	Ctbp1	Rpl13a	Phf7	Tbp	Hprt	Sugp2	Gapdh	Taf11	Usp7	Clock	18S	Sdha	Actb	Hmbs	Arpc3
NormFinder	Phf7	Tbp	Prdx1	Ctbp1	Hprt	Rpl13a	Clock	Sugp2	Gapdh	18S	Usp7	Taf11	Sdha	Actb	Hmbs	Arpc3
GeNorm	Phf7	Prdx1	Ctbp1	Hprt	Rpl13a	Gapdh	Sugp2	Usp7	Taf11	Clock	Tbp	18S	Sdha	Actb	Hmbs	Arpc3
Overall ranking	Prdx1	Phf7	Ctbp1	Tbp	Rpl13a	Hprt	Sugp2	Gapdh	Clock	Usp7	Taf11	18S	Sdha	Actb	Hmbs	Arpc3

Discussion

Traditionally, reference genes used for normalization purposes in qRT-PCR experiments have been chosen based on their supposed role in cellular housekeeping processes. However, these genes do not perform well as reference genes in all settings, especially when complex tissues or larger developmental time series are concerned (Thellin et al., 1999). It is therefore important to critically select and evaluate candidate reference genes for each specific experimental condition.

In this study we report the identification and validation of novel reference genes for qRT-PCR analysis of a wide range of postnatal stages of mouse mammary gland development based on published mammary gland microarray expression data.

Efforts to detect novel reference genes for a diverse set of human samples by comparing large published datasets have been made before (Jin et al., 2004; de Jonge et al., 2007; Kwon et al., 2009; Lee et al., 2007; Szabo et al., 2004). A study by de Jonge *et al* (de Jonge et al., 2007) analysed an astonishing number of 13,629 human microarray sets to determine which genes are stably expressed. The resulting genes, mostly encoding ribosomal proteins, showed improved stability values over traditionally used reference genes such as *GAPDH, ACTB* and *HPRT*. The top ranking genes for human samples also showed high stability values in mouse microarray data (de Jonge et al., 2007). However, the resulting genes were not

analysed for tissue-specificity and were not experimentally validated for the mouse mammary gland. Moreover, since almost all of the identified candidate reference genes were involved in very similar biological processes, they may be co-regulated, making it more difficult to test their suitability as true reference genes (Vandesompele et al., 2002).

Since reference gene stability can be particularly variable depending on the biological data set, reference genes should be validated on any organism- and tissue-specific experimental setting. We employed two separate microarray expression sets to obtain gene expression data for the mouse mammary gland at multiple developmental time points (puberty, pregnancy, lactation and involution) and for two different mouse strains (McBryan et al., 2007; Ron et al., 2007). We hypothesized that any gene showing stable expression across all these different conditions, would serve as a suitable reference gene for the postnatal mouse mammary gland. A total of 37 genes met our selection criteria (i.e. a logFC between -0.1 and 0.1 in four different cross comparisons). Of note, our approach allows the identification of both low, medium and highly expressed genes (Figure 2f), unlike traditional 'housekeeping' genes, which are often expressed at significantly higher levels than the experimental genes of interest. A final set of eight novel candidate reference genes was selected based on average expression levels, amplification efficiency and gene ontology analysis (Supplementary table S1). The latter was performed to reduce the chance of selecting co-regulated genes with similar functions.

The resulting eight candidate reference genes were *Arpc3, Clock, Ctbp1, Phf7, Prdx1, Sugp2, Taf11* and *Usp7*. We compared the performance of these genes to eight traditional reference genes (*Actb, Gapdh, Hprt, Rpl13a, 18S, Hmbs, Sdha* and *Tbp*) using the GeNorm, DeltaCt, BestKeeper and NormFinder algorithms to determine the most stably expressed reference genes (Figure 4 & Table 1) (Andersen et al., 2004; Pfaffl et al., 2004; Silver et al., 2006; Vandesompele et al., 2002). Out of the traditional reference gene set, *Tbp*, *Rpl13a* and *Hprt* perform better than the previously suggested combination of *Gapdh*, *Rpl13a* and *Hprt* (Han et al., 2010). However, our newly identified genes *Prdx1, Phf7* and *Ctbp1* rank as most stable overall. In addition, they are also in the top four of most stable reference genes in terms of their individual rankings (Table 1 and Supplementary Figure S2) and show a small spread in Ct values (Figure 3), further supporting their suitability as reference genes in the mouse mammary gland. Interestingly, *CTBP1* and *PRDX1* have previously been found to be stably expressed in a variety of human samples, suggesting they may be more widely applicable (Kwon et al., 2009; Lee et al., 2007; Zhang et al., 2015). In fact, *PRDX1* was recently shown to be the only reference gene commonly identified in fifteen published human reference gene studies (Zhang et al., 2015).

By using *bona-fide* and highly stable reference genes, even subtle changes in developmental gene expression can be picked up by qRT-PCR on a complex tissue sample (Figure 5). Furthermore, in concordance with published guidelines for performing qRT-PCR experiments (Bustin et al., 2009; Hellemans et al., 2007) the use of three reference genes minimizes errors. Most qRT-PCR software programs accept the use of multiple reference genes for normalization. Alternatively, the geometric mean of three (or more) reference genes can be used for normalization, making this a feasible approach for most researchers today (Dean et al., 2016). However, our validation experiment stresses the importance of selecting the most stable reference gene set (Figure 5). In this respect, *Prdx1, Phf7* and *Ctbp1* perform better over a large developmental time series than previously identified reference genes for the mouse mammary gland (Han et al., 2010).

The fact that *Prdx1, Phf7* and *Ctbp1* outperform traditionally used reference genes, demonstrates the feasibility of our strategy to select tissue-specific, stably expressed genes based on existing microarray data. However, it should be noted that not all candidates showed stable expression upon further testing. In fact, *Arpc3* is one of the less well performing reference genes. One explanation is that our experimental samples contain an earlier developmental time point than the microarray datasets used to identify candidate genes. *Arpc3* expression in 2-week old mice (P14) deviates from the expression values at the other time points (Figure 3b). Another explanation is that strain-specific differences could play a role. Candidates were identified based on microarray experiments on C57BL/6 and CD-1 tissues, but our experimental validation was performed on FVB tissues. This further underscores the importance of testing and validating candidate reference genes, even if they are tissue-specific.

In conclusion, large datasets such as microarray or RNAseq analyses allow the selection of tissue-specific reference genes, although these should always be tested experimentally for stability in the biological samples of interest. Using this approach, we have identified *Prdx1*, *Phf7* and *Ctbp1* as novel, tissue-specific reference genes for qRT-PCR studies of the postnatal mouse mammary gland, and validated that this reference gene set is suitable for detecting subtle gene expression changes that occur during the pre-pubertal, pubertal and adult stages.

Materials and Methods

Animals

FVB/N mice were purchased from Envigo. Breeding for the appropriate developmental time points (P14, P28, P35, P42 and P56) was performed in-house. Mice were housed in open conventional cages on a 12h light/dark cycle and received food and water *ad libitum*. All experiments were performed in accordance with institutional and national guidelines and regulations. All experimental protocols were approved by the Animal Welfare Committee of the University of Amsterdam. As biological replicates, n=3 animals were used for each developmental time point.

Selection of candidate reference genes

A total of eight reference genes commonly used in qRT-PCR analyses for mouse tissues for which primers were already available were selected as 'traditional' reference genes. Another set of eight novel candidate reference genes was selected from published mammary gland microarray datasets (GEO accession numbers GSE5831 (Ron et al., 2007) and GSE6453 (McBryan et al., 2007)) as described in the text. Briefly, GEO2R analysis was performed in RStudio using the GEOquery (Sean and Meltzer, 2007) and limma (Ritchie et al., 2015) R packages from the Bioconductor project to determine the log fold change (logFC) between virgin and lactating mice (GSE5831), pregnant and involuting mice (GSE5831), 3 versus 7 week old mice (GSE6453) and 4 versus 6 week old mice (GSE6453).

Genes with a logFC >-0.1 | <0.1 in each of these comparisons were selected as putative tissue-specific reference genes. These gene lists are provided in Supplementary Table S1. Gene ontology term analysis was performed using the "biological process" and "molecular function" GO terms from the GEO2R analysis to ensure genes with different cellular functions were used to build the final set of eight candidate reference genes to prevent the selection of co-regulated genes.

Primer design and validation

Primer sequences for the eight traditional reference genes were already available in the lab. Novel primers were selected from Primerbank (https://pga.mgh.harvard.edu/primerbank/) when available or designed with the Roche Universal Probe Library Assay Design Centre (https://lifescience.roche.com/). Only intron-spanning primer pairs were used. Specific amplification was confirmed by a single peak after melting curve analysis, a single band of appropriate size on an agarose gel and sequence verification. The amplification efficiency of each primer pair was determined by generating standard curves, using different dilutions of mixed mammary gland cDNA samples as input (see below). The amplification

efficiency of the primers was calculated from the slope of the standard curve using Biogazelle qBase+ software. All primer sequences, amplicon lengths and amplification efficiencies are available in Supp. Table 1.

RNA isolation and cDNA synthesis

Total RNA was isolated from the pooled left and right #4 (inguinal) mammary glands of each biological replicate (n=3 for each time point) using TRIzol reagent (Fisher Scientific) according to the manufacturer's guidelines. RNA from the pooled left and right #3 mammary glands was isolated separately. Residual genomic DNA was digested by RQ1 RNAse-free DNAse treatment (Promega) according to the manufacturer's instructions. The RNA concentration was determined using a Nanodrop spectrophotometer. The purity of all samples was assessed by the absorbance ratios of OD260/280 and OD260/230. cDNA was synthesized from 2 µg RNA using SuperScript® II Reverse Transcriptase (Invitrogen) and Random Hexamers (Fisher Scientific), according to the manufacturer's instructions. The cDNA was diluted 10-fold for subsequent qRT-PCR analysis. Standard curves were generated by diluting the cDNA 2–, 5–, 10–, 100– and 1000-fold.

Quantitative Real-time PCR (qRT-PCR)

qRT-PCR was performed using an Applied Biosystems 7500 Real-Time PCR System. PCR reactions (total 20 µl) were set up containing 13 µl RNAse-free H_2O, 4 µl 5x HOT FIREPol® EvaGreen® qRT-PCR Mix Plus ROX (Solis Biodyne), 0.5 µl of each specific forward and reverse primer (10 µM stock) and 2 µl of diluted cDNA template. The reactions were set up in technical triplicates in 96-well PCR plates. One negative control (no-RT) reaction was included for each sample/primer combination. Thermal cycling was performed, starting with an initial step at 95°C for 10 min, followed by 40 cycles of denaturation at 95°C for 15s and annealing at 60°C for 60s. Each run was completed with a melting curve analysis.

Testing expression stability of candidate reference genes

Raw Ct values were used for GeNorm analysis using the Biogazelle qBase+ software (Vandesompele et al., 2002). All technical triplicates were included to calculate a single Ct value for each gene and biological sample. To minimize the effect of spread within technical replicates, Ct values were calculated based on the median (Vandesompele et al., 2007). Expression stability values (M) were calculated as described in the text.

Median Ct values of the technical triplicates were also used as import for the web-based tool RefFinder (Xie et al., 2012) (http://fulxie.ofees.us). The resulting rankings for the DeltaCt, BestKeeper and NormFinder algorithms were used to assign an appropriate weight to each of the individual reference genes. To this end, all reference genes were given a value (1 – 16) for their position. The same was done

for the GeNorm ranking derived from the qBase+ software. An overall final ranking was calculated based on the geometric mean of their four individual positions.

***Wnt4* expression analysis**
Relative expression levels of *Wnt4* were calculated by the comparative Delta-Ct method (Livak and Schmittgen, 2001; Schmittgen and Livak, 2008), taking the different amplification efficiencies into account, and were normalized to three reference genes (either *Prdx1*, *Phf7* and *Ctbp1; or Rpl13a, Hprt* and *Tbp;* or *Rpl13a, Hprt* and *Gapdh;* or *Actb, Hmbs* and *Sdha*) using the Biogazelle qBase+ software. Statistical analyses were done using qBase+ Statistics Wizard and the Analysis Toolpack Plugin for Microsoft Excel 2016. An F-test for variances was performed to test whether the data distribution was equal. Subsequently, a t-test was performed to test whether there was a difference between respective samples.

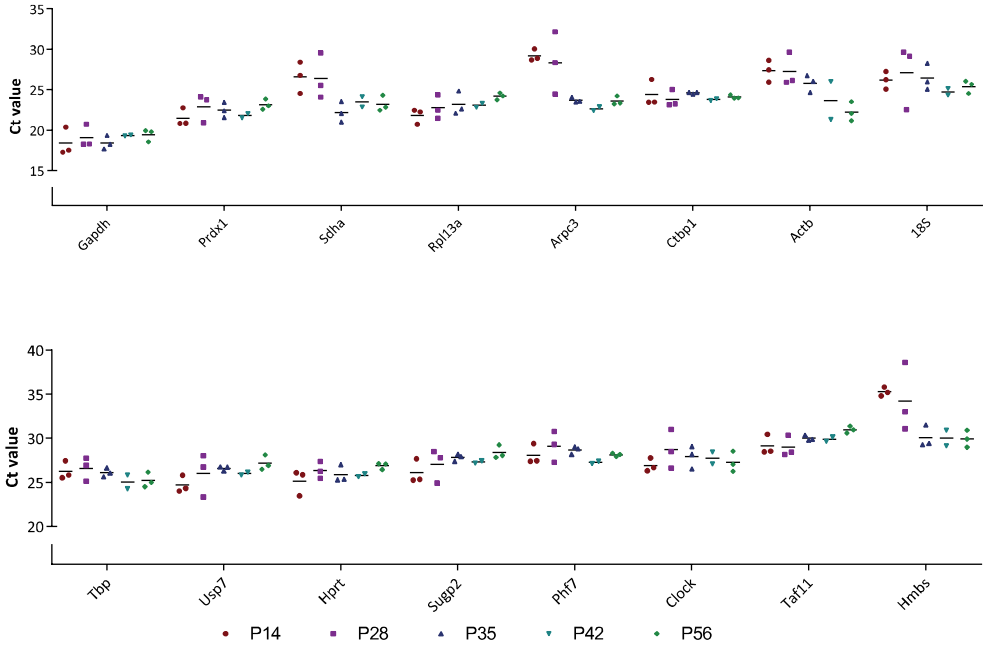
Acknowledgements

We thank Eva Naninck for providing primer sequences for some of the traditional reference genes and for initial help with the qBase+ software, and Amber Zeeman for help with the animal breeding. We thank all lab members for fruitful discussions and Joachim Goedhart and Katrin Wiese for critical feedback on the manuscript. This research was funded by the University of Amsterdam (MacGillavry fellowship to R.vA), the Netherlands Organization for Scientific Research (NWO ALW VIDI 864.13.002 to R.vA) and the Dutch Cancer Society (KWF grants ANW 2013-6057 and 2015-8014 to R.vA).

Author contributions

A.vdM. and R.vA. designed the study. A.vdM. performed the laboratory experiments. A.vdM. and R.vA. analysed the data. A.vdM. and R.vA. wrote the manuscript.

Supplementary Figures



Supplementary figure S1. Ct values and reference gene ranking for #3 mammary gland. Scatter plot depicting spread in expression of candidate reference genes in different stages of the #3 mouse mammary gland. a) Traditional reference genes b) Novel candidate reference genes.

(A)

Ranking order (Better—Good—Average)								
Method	1	2	3	4	5	6	7	8
Delta CT	Tbp	Rpl13a	Hprt	18S	Gapdh	Sdha	Actb	Hmbs
BestKeeper	Rpl13a	Tbp	Hprt	Gapdh	18S	Sdha	Actb	Hmbs
Normfinder	Tbp	Rpl13a	Hprt	18S	Sdha	Gapdh	Actb	Hmbs
GeNorm qBase	Rpl13a	Hprt	Gapdh	Tbp	18S	Sdha	Actb	Hmbs
Recommended comprehensive ranking	**Rpl13a**	**Tbp**	**Hprt**	**Gapdh**	**18S**	**Sdha**	**Actb**	**Hmbs**

(B)

Ranking order (Better—Good—Average)								
Method	1	2	3	4	5	6	7	8
Delta CT	Prdx1	Sugp2	Ctbp1	Phf7	Usp7	Taf11	Clock	Arpc3
BestKeeper	Prdx1	Ctbp1	Phf7	Sugp2	Taf11	Usp7	Clock	Arpc3
Normfinder	Prdx1	Phf7	Ctbp1	Sugp2	Clock	Usp7	Taf11	Arpc3
GeNorm qBase	Phf7	Prdx1	Ctbp1	Sugp2	Usp7	Taf11	Clock	Arpc3
Recommended comprehensive ranking	**Prdx1**	**Phf7**	**Ctbp1**	**Sugp2**	**Usp7**	**Taf11**	**Clock**	**Arpc3**

Supplementary Figure S2. Comprehensive ranking of the old/traditional (A) and novel candidate reference genes (B) using RefFinder (available at http://fulxie.0fees.us/?type=reference). A combined comprehensive ranking of both the traditional and novel candidate genes is presented in Table 1.

Supplementary Table S1. GO terms candidate reference genes; .xslx file available on https://osf.io/sjfr4/?view_only=1259dbaf2de447fb8f95969fb97f2e67. Data for genes showing LogFC<0.1 for all performed microarray data comparisons: LogFC and average expression values. Selected candidate reference genes sorted on average expression levels. Gene ontology terms for candidate reference genes.

Supplementary table S2. Overview of all genes and primer sequences used in this study. The length of the amplicon and the efficiency of the primers is also listed.

Gene name	Accession Number	Primer sequence (5' – 3')	Amplicon length	PCR Efficiency %
Reference genes				
18S	NR_003278	F: CCGCCGCCATGTCTCTAGT R: CTTTCCTCAACACCACATGAGC	150 bp	90
Arpc3	NM_019824	F: TGAAGCGGACAGGACATTGAT R: AGTTGGTGATTCCTAGCGTGT	122 bp	92
Actb	NM_007393	F: GAGCTGCCTGACGGCCAGGT R: TACTCCTGCTTGCTGATCCA	365 bp	72
Clock	NM_001289826	F: TTTACAGGCGTTGTTGATTGGA R: ACGCAAGGCCGTCTTCTG	92 bp	97
Ctbp1	NM_001198859	F: GTGCCCTGATGTACCATACCA R: GCCAATTCGGACGATGATTCTA	83 bp	92
Gapdh	NM_001289726	F: CTGGTGCTGCCAAGGCT R:CTGCTTCACCACCTTCTTGATGTCATCATA	161 bp	99
Hmbs	NM_013551	F: GCCCCATGTGCCTTCAGTC R: TCTTTGAGCCGTTTTCTTCCGC	151 bp	101
Hprt	NM_013556	F: TGTTGTTGGATATGCCCTTG R: TTGCGCTCATCTTAGGCTTT	108 bp	97
Phf7	NM_027949	F: TGTTTGCCTTCTATGCCTTCAA R: GGCAGTCTGCTAGACAGGATAAG	114 bp	88
Prdx1	NM_011034	F: AATGCAAAAATTGGGTATCCTGC R: CGTGGGACACACAAAAGTAAAGT	150 bp	96
Rpl13a	NM_009438	F: CCCTCCACCCTATGACAAGA R: GCCCCAGGTAAGCAAACTT	95 bp	101
Sdha	NM_023281	F: GTTGCTGTGTGGCTGATCG R: GCACAGTGCAATGACACCAC	149 bp	89
Sugp2	NM_001168290	F: TGACCAGAGGCCGATCTCTAA R: GGAAGTAATACGACTGGCAACAG	158 bp	105
Tbp	NM_013684	F: GTCATTTTCTCCGCAGTGCC R: GCTGTTGTTCTGGTCCATGAT	151 bp	94
Taf11	NM_026836	F: GATGGAGACGCGGACTTGAAA R: GCAGGAGTAAGTAAGGACGAGTT	113 bp	99
Usp7	NM_001003918	F: CCACAAGGAAAACGACTGGG R: GTAACACGTTGCTCCCTGATT	196 bp	88
Target genes				
Wnt2	NM_023653	F: CAGAGATCACAGCCTCTTTGG R: GCGTAAACAAAGGCCGATT	65 bp	102
Wnt4	NM_009523	F: ACTGGACTCCCTCCCTGTCT R: TGCCCTTGTCACTGCAAA	109 bp	97

References

- van Amerongen, R., Bowman, A.N., and Nusse, R. (2012). Developmental Stage and Time Dictate the Fate of Wnt/β-Catenin-Responsive Stem Cells in the Mammary Gland. Cell Stem Cell 11, 387–400.
- Andersen, C.L., Jensen, J.L., and Ørntoft, T.F. (2004). 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.
- Brisken, C., Heineman, A., Chavarria, T., Elenbaas, B., Tan, J., Dey, S.K., Mcmahon, J.A., McMahon, A.P., and Weinberg, R.A. (2000). Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. Genes Dev. 14, 650–654.
- Bustin, S.A., Benes, V., Garson, J.A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M.W., Shipley, G.L., et al. (2009). The MIQE guidelines:Minimum Information for publication of quantitative real-time PCR experiments. Clin. Chem. 55, 611–622.
- Cieslak, J., Mackowski, M., Czyzak-Runowska, G., Wojtowski, J., Puppel, K., Kuczynska, B., and Pawlak, P. (2015). Screening for the Most Suitable Reference Genes for Gene Expression Studies in Equine Milk Somatic Cells. PLoS One 10, e0139688.
- Dean, B., Udawela, M., and Scarr, E. (2016). Validating reference genes using minimally transformed qpcr data: findings in human cortex and outcomes in schizophrenia. BMC Psychiatry 16, 154.
- Dhorne-Pollet, S., Thélie, A., and Pollet, N. (2013). Validation of novel reference genes for RT-qPCR studies of gene expression in Xenopus tropicalis during embryonic and post-embryonic development. Dev. Dyn. 242, 709–717.
- Franzellitti, S., Kiwan, A., Valbonesi, P., and Fabbri, E. (2015). Selection of best-performing reference gene products for investigating transcriptional regulation across silvering in the European eel (Anguilla anguilla). Sci. Rep. 5, 16966.
- Han, L.Q., Yang, G.Y., Zhu, H.S., Wang, Y.Y., Wang, L.F., Guo, Y.J., Lu, W.F., Li, H.J., and Wang, Y.L. (2010). Selection and use of reference genes in mouse mammary glands. Genet. Mol. Res. 9, 449–456.
- Hellemans, J., Mortier, G., De Paepe, A., Speleman, F., and Vandesompele, J. (2007). qBase relative quantification framework and software for management and automated analysis of real-time quantitative PCR data. Genome Biol. 8, R19.
- Jin, P., Zhao, Y., Ngalame, Y., Panelli, M.C., Nagorsen, D., Monsurró, V., Smith, K., Hu, N., Su, H., Taylor, P.R., et al. (2004). Selection and validation of endogenous reference genes using a high throughput approach. BMC Genomics 5, 55.
- de Jonge, H.J.M., Fehrmann, R.S.N., de Bont, E.S.J.M., Hofstra, R.M.W., Gerbens, F., Kamps, W.A., de Vries, E.G.E., van der Zee, A.G.J., te Meerman, G.J., and ter Elst, A. (2007). Evidence based selection of housekeeping genes. PLoS One 2, 1–5.

- Joshi, P. a, Jackson, H.W., Beristain, A.G., Di Grappa, M. a, Mote, P. a, Clarke, C.L., Stingl, J., Waterhouse, P.D., and Khokha, R. (2010). Progesterone induces adult mammary stem cell expansion. Nature 465, 803–807.
- Kosir, R., Acimovic, J., Golicnik, M., Perse, M., Majdic, G., Fink, M., and Rozman, D. (2010). Determination of reference genes for circadian studies in different tissues and mouse strains. BMC Mol. Biol. 11, 60.
- Kwon, M.J., Oh, E., Lee, S., Roh, M.R., Kim, S.E., Lee, Y., Choi, Y. La, In, Y.H., Park, T., Koh, S.S., et al. (2009). Identification of novel reference genes using multiplatform expression data and their validation for quantitative gene expression analysis. PLoS One 4.
- Lee, S., Jo, M., Lee, J., Koh, S.S., and Kim, S. (2007). Identification of novel universal housekeeping genes by statistical analysis of microarray data. J. Biochem. Mol. Biol. 40, 226–231.
- Li, M.-Y., Wang, F., Jiang, Q., Wang, G.-L., Tian, C., and Xiong, A.-S. (2016). Validation and Comparison of Reference Genes for qPCR Normalization of Celery (Apium graveolens) at Different Development Stages. Front. Plant Sci. 7, 1–12.
- Liu, C., Xin, N., Zhai, Y., Jiang, L., Zhai, J., Zhang, Q., and Qi, J. (2014). Reference gene selection for quantitative real-time RT-PCR normalization in the half-smooth tongue sole (Cynoglossus semilaevis) at different developmental stages, in various tissue types and on exposure to chemicals. PLoS One 9.
- Livak, K.J., and Schmittgen, T.D. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2^- ΔΔCT Method. Methods 25, 402–408.
- Macias, H., and Hinck, L. (2012). Mammary gland development. Wiley Interdiscip. Rev. Dev. Biol. 1, 533–557.
- Martins, P.K., Mafra, V., de Souza, W.R., Ribeiro, A.P., Vinecky, F., Basso, M.F., da Cunha, B.A.D.B., Kobayashi, A.K., and Molinari, H.B.C. (2016). Selection of reliable reference genes for RT-qPCR analysis during developmental stages and abiotic stress in Setaria viridis. Sci. Rep. 6, 28348.
- McBryan, J., Howlin, J., Kenny, P. a, Shioda, T., and Martin, F. (2007). ERalpha-CITED1 co-regulated genes expressed during pubertal mammary gland development: implications for breast cancer prognosis. Oncogene 26, 6406–6419.
- Nakamura, A.M., Chahad-Ehlers, S., Lima, A.L.A., Taniguti, C.H., Sobrinho Jr., I., Torres, F.R., and de Brito, R.A. (2016). Reference genes for accessing differential expression among developmental stages and analysis of differential expression of OBP genes in Anastrepha obliqua. Sci. Rep. 6, 17480.
- Petriccione, M., Mastrobuoni, F., Zampella, L., and Scortichini, M. (2015). Reference gene selection for normalization of RT-qPCR gene expression data from Actinidia deliciosa leaves infected with Pseudomonas syringae pv. actinidiae. Sci. Rep. 5, 16961.
- Pfaffl, M.W., Tichopad, A., Prgomet, C., and Neuvians, T.P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper--Excel-based tool using pair-wise correlations. Biotechnol. Lett. 26, 509–515.
- Rajaram, R.D., Buric, D., Caikovski, M., Ayyanan, A., Rougemont, J., Shan, J., Vainio, S.J., Yalcin-Ozuysal, O., and Brisken, C. (2015). Progesterone and Wnt4 control mammary stem cells via myoepithelial crosstalk. EMBO J. 34, 641–652.

- Richert, M.M., Schwertfeger, K.L., Ryder, J.W., and Anderson, S.M. (2000). An atlas of mouse mammary gland development. J. Mammary Gland Biol. Neoplasia 5, 227–241.
- Ritchie, M.E., Phipson, B., Wu, D., Hu, Y., Law, C.W., Shi, W., and Smyth, G.K. (2015). limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43.
- Ron, M., Israeli, G., Seroussi, E., Weller, J.I., Gregg, J.P., Shani, M., and Medrano, J.F. (2007). Combining mouse mammary gland gene expression and comparative mapping for the identification of candidate genes for QTL of milk production traits in cattle. BMC Genomics 8, 183.
- Schliep, M., Pernice, M., Sinutok, S., Bryant, C. V, York, P.H., Rasheed, M.A., and Ralph, P.J. (2015). Evaluation of Reference Genes for RT-qPCR Studies in the Seagrass Zostera muelleri Exposed to Light Limitation. Sci. Rep. 5, 17051.
- Schmittgen, T.D., and Livak, K.J. (2008). Analyzing real-time PCR data by the comparative CT method. Nat. Protoc. 3, 1101–1108.
- Schmittgen, T.D., and Zakrajsek, B.A. (2000). Effect of experimental treatment on housekeeping gene expression: Validation by real-time, quantitative RT-PCR. J. Biochem. Biophys. Methods 46, 69–81.
- Sean, D., and Meltzer, P.S. (2007). GEOquery: A bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics 23, 1846–1847.
- Silberstein, G.B., Van Horn, K., Hrabeta-Robinson, E., and Compton, J. (2006). Estrogen-triggered delays in mammary gland gene expression during the estrous cycle: Evidence for a novel timing system. J. Endocrinol. 190, 225–239.
- Silver, N., Best, S., Jiang, J., and Thein, S.L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. BMC Mol. Biol. 7, 33.
- Szabo, A., Perou, C.M., Karaca, M., Perreard, L., Palais, R., Quackenbush, J.F., and Bernard, P.S. (2004). Statistical modeling for selecting housekeeper genes. Genome Biol. 5, R59.
- Thellin, O., Zorzi, W., Lakaye, B., De Borman, B., Coumans, B., Hennen, G., Grisar, T., Igout, A., and Heinen, E. (1999). Housekeeping genes as internal standards: Use and limits. J. Biotechnol. 75, 291–295.
- Thomas, K.C., Zheng, X.F., Suarez, F.G., Raftery, J.M., Quinlan, K.G.R., Yang, N., North, K.N., and Houweling, P.J. (2014). Evidence based selection of commonly used RT-qPCR reference genes for the analysis of mouse skeletal muscle. PLoS One 9.
- Vandesompele, J., De Preter, K., Pattyn, F., Poppe, B., Van Roy, N., De Paepe, A., and Speleman, F. (2002). Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol. 3, RESEARCH0034.
- Vandesompele, J., Preter, K. De, Pattyn, F., and Poppe, B. (2007). GeNorm software manual. Last Updat. March 2014 1–16.
- De Visser, K.E., Ciampricotti, M., Michalak, E.M., Tan, D.W.M., Speksnijder, E.N., Hau, C.S., Clevers, H., Barker, N., and Jonkers, J. (2012). Developmental stage-specific contribution of LGR5+ cells to basal and luminal epithelial lineages in the postnatal mammary gland. J. Pathol. 228, 300–309.
- Visvader, J.E., and Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy: current status and perspectives. Genes Dev. 28, 1143–1158.

- Wang, D., Cai, C., Dong, X., Yu, Q.C., Zhang, X.-O., Yang, L., and Zeng, Y.A. (2015). Identification of multipotent mammary stem cells by protein C receptor expression. Nature 517, 81–84.
- Watson, C.J., and Khaled, W.T. (2008). Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development 135, 995–1003.
- Weber-Hall, S.J., Phippard, D.J., Niemeyer, C.C., and Dale, T.C. (1994). Developmental and hormonal regulation of Wnt gene expression in the mouse mammary gland. Differentiation 57, 205–214.
- Wu, Z.-J., Tian, C., Jiang, Q., Li, X.-H., and Zhuang, J. (2016). Selection of suitable reference genes for qRT-PCR normalization during leaf development and hormonal stimuli in tea plant (Camellia sinensis). Sci. Rep. 6, 19748.
- Xie, F., Xiao, P., Chen, D., Xu, L., and Zhang, B. (2012). miRDeepFinder: A miRNA analysis tool for deep sequencing of plant small RNAs. Plant Mol. Biol. 80, 75–84.
- Yang, C., Pan, H., Noland, J.E., Zhang, D., Zhang, Z., Liu, Y., and Zhou, X. (2015). Selection of reference genes for RT-qPCR analysis in a predatory biological control agent, Coleomegilla maculata (Coleoptera: Coccinellidae). Sci. Rep. 5, 18201.
- Ye, X., Zhang, F., Tao, Y., Song, S., and Fang, J. (2015). Reference gene selection for quantitative real-time PCR normalization in different cherry genotypes, developmental stages and organs. Sci. Hortic. (Amsterdam). 181, 182–188.
- Zhang, Y., Li, D., and Sun, B. (2015). Do housekeeping genes exist? PLoS One 60, 512-521.