MethodsX 8 (2021) 101504

Contents lists available at ScienceDirect

MethodsX

journal homepage: www.elsevier.com/locate/mex

Method Article

A novel method for the normalization of ChIP-qPCR data

Elizabeth R Solomon, Kevin K Caldwell, Andrea M Allan*

Department of Neurosciences, University of New Mexico HSC, United States

A B S T R A C T

ChIP-qPCR permits the study of protein and chromatin interactions. The general technique can apply to the study of the interactions of protein with RNA, and the methylation state of genomic DNA. While the technique is vital to our understanding of epigenetic processes, there is much confusion around the proper normalization methods. Percent Input has recently emerged as a normalization standard, due to its reproducibility and accuracy. This method relies on the use of a constant volume of ChIP Isolate in each qPCR assay. Researchers may accidentally run qPCR assays with a constant amount of isolate, a common practice for RT-qPCR; however, the traditional Percent Input method cannot accurately normalize these data. We developed a novel method that can normalize these data to provide the same reproducible Percent Input value. Here, we present evidence that this novel method of normalizing ChIP-qPCR data works with real samples. Later, we present a mathematical proof which shows how a Percent Input value calculated from Cq (quantification cycle) values obtained from qPCR run with a constant amount (in nanograms of DNA in ChIP isolate) is equal to the traditional Percent Input calculated from quantification cycle (Cq) values obtained from running a constant volume of ChIP isolate.

- *Increases the number of possible data points per sample*
- *End values are the same % Input values as the traditional normalization method*

© 2021 Published by Elsevier B.V.
This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

A R T I C L E   I N F O

Method name: Constant Amount % Input (ChIP-qPCR Normalization)
Keywords: Percent input, Dilution factor, Chromatin, Immunoprecipitation
Article history: Received 30 June 2021; Accepted 29 August 2021; Available online 31 August 2021

DOI of original article: 10.1016/j.taap.2020.114920

* Corresponding author.
E-mail address: aallan@salud.unm.edu (A.M. Allan).

https://doi.org/10.1016/j.mex.2021.101504

2215-0161/© 2021 Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

Specifications table

Subject Area:	Biochemistry, Genetics and Molecular Biology
More specific subject area:	Epigenetics
Method name:	Constant Amount % Input (ChIP-qPCR Normalization)
Resource Availability:	NA
Name and reference of original method:	Common Name: % Input Original Name: Percentage of Immunoprecipitation Nagak Kiyotaka et al. Chromatin Immunoprecipitation Reveals That the 180-bp Satellite Repeat Is the Key Functional DNA Element of Arabidopsisthaliana Centromeres. Genetics163:1221–1225. March 2003. PMID: 12663558

Background

Chromatin Immunoprecipitation (ChIP) is a technique used to study the interaction between chromatin and a specific bio-molecular target. However, the chemistry behind this assay is not limited to the study of DNA-Histone complexes, any nucleic acid in a complex with an immuno-reactive molecule can be co-immunoprecipitated (Co-IP). Other nucleic acid Co-IP assays include RNA binding protein IP (RIP) and methylated DNA IP (meDIP or mDIP). Nucleic acid Co-Immunoprecipitation is a vital tool for understanding the epigenome. However, even with its importance in modern research, there is still much debate around the proper method to normalize the data produced by the qPCR of nucleic acids obtained by Co-IP. The goal of this manuscript is to present a novel method for analyzing a ChIP-qPCR, RIP-qPCR, or meDIP-qPCR assay, supported by real sample data and a logical proof.

The most common and cost effective way to measure specific nucleic acid sequences yielded in a ChIP, RIP, or meDIP isolate is by a quantitative polymerase chain reaction (qPCR) assay. In simplified terms, qPCR is the process of repeatedly heating and cooling DNA in the presence of DNA polymerase, free nucleotides, sequences specific primers, and an amplification reporting fluorophore. When operating at 100% efficiency the sequence targeted by the primers, known as the amplicon, doubles in concentration each heating and cooling cycle. The increase in the amplicon increases the fluorescence reported. Once the fluorescence reaches an arbitrary threshold value, the time (in cycles) it took to ‘break threshold’ is recorded as the quantification cycle (Cq) also known as cycle threshold (Ct). The Cq value for a given qPCR assay is inversely proportional to the amount of target sequence before amplification.

A more technical definition of ChIP is a nucleic acid isolation procedure that yields an enriched pool of nucleic acids in complex with immuno-reactive molecules (IRMs). [1]IRMs are any substance that binds to the antibody or beads used in the assay; background (noise) occurs when the intended bio-molecular target ([2]BMT) is not the only [1]IRM in the immunoprecipitation step. Any nucleic acid chain that is in a stable complex with immuno-reactive molecules is isolated from other components through an immunoprecipitation reaction using an antibody bound to magnetic or agarose beads [3–8]. Once this DNA is isolated, it can be analyzed by microarray, qPCR, or sequencing. Of these methods, qPCR provides high resolution quantitative data while also remaining affordable. The end goal of ChIP-qPCR is to allow the user to determine the degree of which a specific region is associated with a bio-molecular target, such as a histone with a unique post-translational modification (PTM), methylated cytosine or DNA/RNA binding proteins.

To achieve this goal, careful planning, recording, and pipetting must be used throughout the assay, as differences in volume or dilutions can have a direct impact on the final calculation. Each unknown sample lysate is split into two processing paths, the “Input” ([4]IN) path and the “Immunoprecipitated” ([3]IP) path. The IN processing path simply involves the nucleic acid isolation from the sample lysate. This path produces the IN isolate. In the IP path, an aliquot of the same sample lysate is immunoprecipitated and nucleic acids are, subsequently, isolated from the precipitate. The IP path produces the IP isolate. Due to the often-low yields of the IP path, the volume of the sample lysate used in the IP path is 5 to 100 times more than the volume of sample lysate used in the IN path. An accurate calculation of the ratio of starting material between both paths, commonly referred to as a dilution factor, is essential for ChIP-qPCR data normalization.

A variety of negative controls can be used in a nucleic acid Co-IP assay, the two most common being an isotype control and a no antibody control. The isotype control utilizes heterogeneous antibody from the same host species and isotype as the target antibody in order to pull down anything that may bind non-specifically to the antibody or the immunoprecipitation beads used. The no antibody control only pulls down material that binds to the IP beads used, of which Protein A/G conjugated magnetic beads are the most popular choice.

Additional controls require a thorough understanding of the target that is not always available to the researcher when designing a ChIP, RIP, or meDIP protocol [3,9]. These involve designing primers that target a region known to associate or not associate with the immuno-reactive molecule of interest. Such positive and negative controls can validate the accuracy of an assay to a much higher degree than the isotype and no antibody negative controls; which only serve to confirm the antibody is more specific than background noise.

Current ChIP-qPCR normalization methods

The two most used methods for ChIP-qPCR data normalization are fold enrichment (Eq. (1)) and percent input (Eq. (2)). Fold enrichment is a signal-to-noise ratio comparing the amount of target sequence measured in the IP isolate to the amount measured in a negative control isolate. This method is typically capable of minimal precision, because its results are dependent on the type and quality of negative control, it is a challenging value to reproduce. The benefit of calculating fold enrichment is to estimate the signal-to-noise ratio. The percent input method compares the amount of target sequence measured in the IP isolate to the total amount of the target sequence in the IN isolate [12]. An ideal percent input value is the ratio of the amount of target sequence measured in the IP isolate to the amount of target sequence found in the original sample lysate.

Eq. (1): Fold Enrichment

$$Fold\ Enrichment = 2^{(Cq(Negative\ Control) - Cq(IP))} \tag{1}$$

The dependence of normalizing to an assay negative control makes fold enrichment a difficult value to reproduce. A difference in manufacturer or even lot-to-lot variability can change how Protein A/G magnetic beads or antibody interact in an immunoprecipitation reaction. The negative control from a high-quality immunoprecipitation often will not have a quantification cycle (Cq) value within the allotted cycles of a qPCR assay. In the scenario where a negative control produces no Cq value, Cq(Negative Control) in Eq. (1) is often substituted by the number of allotted cycles. Another pitfall of fold enrichment is that subtle variations in background between samples can lead to artificial differences and misinterpretation of the data [1,9].

Eq. (2): Percent Input

$$\%\ Input = 2^{((Cq(IN) - Log_2(DF)) - Cq(IP))} * 100 \tag{2}$$

Unlike fold enrichment, which is a representation of signal over background noise, percent input normalization is an expression of the ratio of the number of target sequences measured in the IP isolate to the number of target sequences measured in the IN isolate. This expression allows scientists to normalize ChIP-qPCR data for a value that is representative of the *in vivo* percentage of target sequences in complex with a given molecule. The current method to calculate percent input relies heavily on precise pipetting and using a constant volume of isolate in each qPCR reaction. The IN path usually receives 5–100 times less lysate than the amount put into the IP path. This saves on material required for the assay, however this must be accounted for when normalizing the data. The ratio of starting amounts is referred to as the dilution factor (DF) and it is vital for accurate percent input calculations. Eq. (3) describes how to calculate the dilution factor for the traditional percent input normalization method when starting with a constant volume of IP and IN isolate. Eqs. (2) and (3) also help identify the three key steps where precise pipetting is required. These are, the addition of Lysate into the IN or IP paths, the volume of elution buffer to dissolve the final IN and IP isolates, and the volume of isolate added to each qPCR reaction.

Eq. (3): [7]DFv is the initial volume of sample lysate used in the IP path (IP iV) divided by the initial volume of sample lysate used in the IN path (IN iV).

$$Dilution\ Factor\ (Constant\ Volume) = DFv = \frac{iV(IP)}{iV(IN)} \tag{3}$$

In their 2007 paper, Haring and colleagues detail two other, less common normalization methods [9]. The first normalizes the IP ChIP-qPCR signal relative to a positive or negative control gene associated with the same biomolecular target (BMT). However, it can be difficult to find a gene sequence that does not change its association with the BMT after exposure to the experiment's treatment. The other method was developed for measuring histone modifications on specific genes, this method normalizes the modified histone IP ChIP-qPCR signal relative to a total histone IP signal, e.g., H3K9me3 relative to total H3 on the same gene sequence. This method is difficult to optimize, and like fold enrichment, it can produce artifacts due to subtle changes in total histone levels between samples.

For researchers lacking information on a BMT's association with specific genes, there is currently only one option for ChIP-qPCR data normalization, that is a Percent Input normalization. Fold Enrichment is highly susceptible to background noise variation, and the more advanced methods of normalization rely on specific knowledge of a BMT's association with target genes. Epigenetics is at the forefront of modern research, with many novel pathways to study. Percent Input normalization is the most robust method for repeatable research of less studied or unknown epigenetic associations.

Two ways to calculate percent input

In the following sections, the traditional percent input method to normalize ChIP-qPCR data will be referred to as the constant volume ([5]CV) method and the new method described in this manuscript will be labeled as the constant amount ([6]CA) method. This highlights the major difference between the two methods: the constant volume method relies on the use of an equal volume of elution buffer, as well as an equal volume of chromatin for each qPCR reaction from both the IP and IN chromatin isolates. While the constant amount method relies on knowing the yielded amount (in nanograms) of chromatin from the IP and IN chromatin isolates in order to calculate a sample specific dilution factor and then using an equal amount (in nanograms) of chromatin for each qPCR reaction.

Dilution factors

In the traditional constant volume method, a dilution factor must be used to account for the difference of initial tissue lysate used in the IP and IN sample processing paths. In our example, the volume *"iV(IP)"* of initial tissue lysate used in the IP path is 10 times the volume *"iV(IN)"* used in the IN path, giving a dilution factor of 10 as described by Eq. (3). This dilution factor can be used to normalize any sample whose initial lysate used in the IP path is a factor of 10 greater than the initial lysate used in the IN path. In this paper the symbol for the constant volume dilution factor is "[7]DFv".

The constant amount method requires a new dilution factor. In our example, we are working with the same IP and IN samples for both normalization methods. Again, the initial lysate volume used for the IP path is a factor of 10 greater than the initial lysate volume used for the IN path. However, we will also need to know the factor of enrichment, by dividing the final yield of chromatin from the IN isolate *"fA(IN)"* by the final yield of chromatin from the IP isolate *"fA(IP)"*. The final yield is calculated by multiplying the elution volume by the concentration. The product of *DFv* and the factor of enrichment gives us the dilution factor for constant amount (*DFa*) as described by Eq. (4). The constant amount dilution factor *"DFa"* is unique for each IP/IN pair. Unlike [7]*DFv* the same [8]*DFa* cannot be used for all samples in an assay.

Eq. (4): *DFa* is the product of *DFv* and the final amount recovered from the IN path *fA(IN)* divided by the final amount recovered from the IP path *fA(IP)*.

$$DFa = \frac{iV(IP)}{iV(IN)} * \frac{fA(IN)}{fA(IP)} = DFv * \frac{fA(IN)}{fA(IP)} \tag{4}$$

The reliance on knowing the final yield of chromatin in the IP and IN isolates is the major drawback to the CA normalization method. IP isolate concentrations can oftentimes be too low to be reliably quantified by Qubit or Nano-Drop. In such a case, where to measure the sample' would use most or all of the sample, then the CV method should be used.

Example of CV and CA percent input calculations with real data

The constant amount method was put to the test using ten samples of mouse frontal cortex lysate. For each sample, *iV(IP) = 2160 μL* of lysate went to the IP path and was immunoprecipitated with an anti-H3K9me3 antibody then the co-immunoprecipitated DNA was isolated and eluted with a final volume of *fV(IP)* = 50 μL. For the Input (IN) path, *iV(IN)* =216 μL of lysate was treated to the same DNA isolation as the IP path and eluted with a final volume of *fV(IN)* =50 μL. Final DNA concentrations *(M)* of IP and IN isolates were quantified using a Qubit 2.0 Fluorometer (Thermo Fisher). Final DNA yields *(fA)* were calculated by multiplying the elution volume *(fV)* by the DNA concentration *(M)*. Please refer to Solomon et al. to review more details of the ChIP protocol used to prepare these samples [13].

Following Eq. (3), a *DFv* value of 10 was calculated by dividing *iV(IP)* by *iV(IN)*. Since all IP or IN paths used the same volume of starting material, the constant volume dilution factor *(DFv)* is 10 for all samples. To use the CA normalization method a unique constant amount dilution factor (*DFa)* must be calculated for each individual sample. Following Eq. (4), a *DFa* value was calculated by dividing *fA(IN)* by *fA(IP),* for each sample, then multiplying by *DFv*. Fig. 1a illustrates the steps and variables in the *DFv* and *DFa* calculations for one of the ten frontal cortex samples used.

These ten sets of IP and IN isolates were then analyzed via qPCR assay. All reactions were run on a LightCycler96 (Roche) with FastStart Essential DNA Green Master (Roche) following manufacturer's instructions for cycling parameters. Forward and reverse primers targeting the promoter region of the Crhr1 gene were previously validated and were used at a final concentration of 1 μM for all reactions [13].

Fig. 1b details the steps and variables used to calculate percent input from the CV and CA methods. For the constant volume qPCR assays, $qV(IP_{CV}) = 2.00$ μL of IP isolate and $qV(IN_{CV}) = 2.00$ μL of IN isolate are all run in triplicate qPCR assays. The mean quantification cycle (Cq) of each triplicate is calculated; labeled *"Cq"* in Fig. 1b. Following Eq. (2) to calculate the percent input with the constant volume method using *DFv* calculated with Eq. (3).

For the constant amount qPCR assays, $qA(IP_{CA}) = 1.25$ ng of IP isolate chromatin and $qA(IN_{CA}) = 1.25$ ng of IN isolate chromatin are each run in triplicate qPCR reactions. Again, the mean quantification cycle (Cq) of each triplicate is calculated; labeled *"Cq"* in Fig. 1b. Following Eq. (2) to calculate the percent input with the constant amount method by replacing *DFv* with *DFa* calculated from Eq. (4). The calculated percent input (CV) and percent input (CA) values of all ten frontal cortex samples, which highlight the variation between both methods, can be seen on Fig. 2.

Variation in percent input from both methods correlated with variation in qPCR efficiency

Fig. 2 presents the results of the direct comparison of the CV and CA percent input normalization methods across ten individual samples. The difference between CV and CA methods ranged from 0.36 to 9.34 percentage points, with an average variation of 2.62% points. Fig. 4a presents this data in a graph. Later in the manuscript is a mathematical proof deriving Eq. (4) from the expression: Percent Input (CV) is equal to Percent Input (CA). However, the proof assumes that the efficiency *"E"* is the same for all reactions, yet concentrations of dNTPs, Mg+ ions, Taq polymerase, primers, and the initial amount (qA) of DNA all influence the efficiency of the polymerase chain reaction.

The experiment maintained a constant *qA* = 1.25 ng of DNA in initial isolate for all CA qPCR assays, and used a constant *qV* = 2.00 μL of initial isolate for all CV qPCR assays. Due to the differences in initial concentrations of IP and IN isolates, there was a large variation in the initial amount of DNA *(qA)* used in each CV assay, with this amount ranging from 5.56 ng to 13.12 ng for the IP(CV) assays and from 1.92 ng to 8.64 ng for the IN(CV) assays. Every CV assay had more DNA than its corresponding CA assay. All other components in the qPCR assays, including primers, were the same.

a						
Sample Path	**Initial Sample Lysate Volume (μL)**	**Eq. #3 Dilution Factor Constant Volume**	**Elution Buffer Volume (μL)**	**Isolate Conc. [ng/μL]**	**Isolate Yield (ng)**	**Eq. # 4 Dilution Factor Constant Amount**
Symbol	iV	DFv	fV	M	fA	DFa
IP	2160 μL	$= \frac{iV(IP)}{iV(IN)}$	50 μL	2.78 ng/μL	$= fV * M$ 139 ng	$= DFv * \frac{fA(IN)}{fA(IP)}$
IN	216 μL	10	50 μL	1.62 ng/μL	$= fV * M$ 81 ng	5.83

b						
Method	**qPCR Reaction Name**	**qPCR Initial Template Volume**	**qPCR Initial Template Amount**	**Cq Mean**	**Dilution Factor**	**Eq. #2 Percent Input**
CA, CV		qV	qA	Cq	DFa, DFv	*% Input*
Constant Amount method	IP_{CA}	0.45 μL Stock	1.25 ng	25.25	DFa	$= 2^{Cq(IN_{CA}) - Log2(DFa) - Cq(IP_{CA})}$
	IN_{CA}	0.77 μL Stock	1.25 ng	25.33	5.83	**18.13%**
Constant Volume method	IP_{CV}	2 μL Stock	5.56 ng	23.2	DFv	$= 2^{Cq(IN_{CV}) - Log2(DFv) - Cq(IP_{CV})}$
	IN_{CV}	2 μL Stock	3.24 ng	24.03	10	**17.78%**

Fig. 1. (a). Example dilution factor calculation, two methods
Example calculation of the dilution factors required for the traditional constant volume normalization method (*DFv*) and the new constant amount method (*DFa*). Rows "IP" and "IN" detail the parameters of the two sample paths in a ChIP assay. Dilution Factor Constant Volume (*DFv*) is calculated with Eq. (3) using the Initial Sample Lysate Volumes (*iV*) of IP and IN paths. Dilution Factor Constant Amount (*DFa*) is calculated with Eq. (4) using *DFv* and the final Isolate Yield (*fA*) of IP and IN paths. *fA* is calculated from the product of final Isolate Volume (*fV*) and the Isolate Concentration (*M*). (b). Example percent input calculation, two methods
Example normalization of real ChIP-qPCR data, comparing the traditional constant volume normalization method and the new constant amount method. For this example the same IP and IN isolates were analyzed by the constant volume and constant amount methods. The constant volume qPCR reactions received the same volume (*qV*) of IP and IN isolate solution, while the constant amount qPCR reactions received the same amount (*qA*) of IP and IN isolate chromatin. Cq Mean is the actual quantification cycle values obtained from this example experiment. Using the appropriate dilution factor with Eq. (2) an accurate Percent Input can be calculated from Cq values obtained by either constant volume or constant amount methods.

From our qPCR data we can calculate the efficiency offset between CV and CA reactions of the same isolate with Eq. (5).

Eq. (5). Based on the qPCR efficiency equation [2], Efficiency Offset is used to compare the amplification efficiency between two qPCR reactions with different amounts of the same starting material. Efficiency offset (^{9}EO) between two qPCR assays from the same isolate sample, where qA is the initial amount of DNA in nanograms for a qPCR reaction and Cq is the resulting quantification

Mouse Frontal Cortex Sample	Assay ID	qPCR Initial IP Isolate Amount	Mean IP Cq Values	IP PCR Efficiency Offset CV/CA	qPCR Initial IN Isolate Amount	Mean IN Cq Values	IN PCR Efficiency Offset CV/CA	Dilution Factor	% Input CA, CV	Difference between % Input Methods	Efficiency Offset Difference IP-IN
#	*CA#, CV#*	*qA(IP)*	*Cq(IP)*	*EO(IP)*	*qA(IN)*	*Cq(IN)*	*EO(IN)*	*Dfa, DFv*	*% Input*	*%Input(CA)-%Input(CV)*	*EO(IP)-EO(IN)*
1	CA-1	1.25 ng	26.01	1.17	1.25 ng	26.33	0.99	8.31	15.02%	2.27	0.18
	CV-1	9.24 ng	23.35		7.68 ng	23.70		10.00	12.75%		
2	CA-2	1.25 ng	25.71	1.18	1.25 ng	25.23	1.02	3.82	18.75%	2.51	0.16
	CV-2	8.4 ng	23.20		3.22 ng	23.90		10.00	16.25%		
3	CA-3	1.25 ng	26.13	1.12	1.25 ng	25.84	1.02	5.45	14.99%	1.24	0.09
	CV-3	11.00 ng	23.15		6.00 ng	23.61		10.00	13.76%		
4	CA-4	1.25 ng	25.88	1.09	1.25 ng	25.79	1.02	5.65	16.63%	1.04	0.07
	CV-4	9.84 ng	23.03		5.56 ng	23.67		10.00	15.58%		
5	CA-5	1.25 ng	25.92	1.24	1.25 ng	26.18	1.16	7.91	15.13%	0.99	0.08
	CV-5	10.92 ng	23.10		8.64 ng	23.60		10.00	14.14%		
6	CA-6	1.25 ng	26.18	1.24	1.25 ng	24.39	1.09	1.48	19.48%	2.31	0.15
	CV-6	12.96 ng	23.12		1.92 ng	23.90		10.00	17.17%		
7	CA-7	1.25 ng	25.99	1.37	1.25 ng	26.59	1.02	4.18	36.29%	9.34	0.35
	CV-7	13.12 ng	23.05		5.48 ng	24.48		10.00	26.94%		
8	CA-8	1.25 ng	26.25	1.21	1.25 ng	26.44	1.00	5.11	22.33%	3.79	0.21
	CV-8	12.76 ng	23.17		6.52 ng	24.06		10.00	18.53%		
9	CA-9	1.25 ng	25.68	1.14	1.25 ng	24.88	1.02	2.32	24.73%	2.38	0.11
	CV-9	8.56 ng	23.09		1.98 ng	24.25		10.00	22.35%		
10	CA-10	1.25 ng	25.25	1.07	1.25 ng	25.33	1.05	5.83	18.14%	0.36	0.02
	CV-10	5.56 ng	23.20		3.24 ng	24.03		10.00	17.78%		
										Average	Average
										2.62	0.14
										Pearson's Correlation	
										r = 0.957	

Fig. 2. Comparing two normalization methods to calculate percent input with real ChIP-qPCR Data from 10 individual tissue samples
Real ChIP-qPCR data from mouse frontal cortex. The Assay ID column indicates the assay and normalization method used for that row. The IP and IN isolate analysis are separated in 3 column groups. The Efficiency Offset columns show the difference in PCR amplification efficiency between the CV and CA assays of the IP or IN isolates. The Dilution Factor column shows the unique *DFa* value for each sample as well as the unchanging *DFv* value. The % Input column shows the calculated percent input from the data in this table, CA or CV depending on row. The final two columns emphasize the variance between calculated % Input from the CA and CV methods and the correlation to difference in Efficiency Offset between IP and IN paths within normalization methods. The novel constant amount normalization method varies from the traditional constant volume method only when the amplification efficiency between reactions also varies. When optimized for efficiency the constant amount method has the potential to be more accurate than the constant volume method.

Eq. #5 Efficiency Offset CV/CA	Efficiency Offset Difference Between IP and IN
EO	***ED***
$= \frac{2^{Cq(IP_{CV})} * qA(IP_{CV})}{2^{Cq(IP_{CA})} * qA(IP_{CA})}$ IP 1.07	$= EO(IP_{CV/CA}) - EO(IN_{CV/CA})$
$= \frac{2^{Cq(IN_{CV})} * qA(IN_{CV})}{2^{Cq(IN_{CA})} * qA(IN_{CA})}$ IN 1.05	.0021

Fig. 3. Efficiency comparison
Example calculation of efficiency offset (Eq. (5)) to compare the amplification efficiency between two qPCR reactions with different amounts of the same starting material. Using the data in Fig. 1b, the *EO* is calculated for the IP and IN isolates. The IP isolate has an *EO* of 1.07, meaning the numerator reaction (Constant Volume) amplifies at a rate of 1.07 times that of the denominator reaction (Constant Amount). The IN isolate has an *EO* of 1.05. From this the difference in EO can be calculated, the efficiency offset difference between this pair of IP and IN isolates is 0.0021. The smaller the efficiency offset difference between IP and IN isolates from the same sample, the closer the calculated percent input will be between the constant volume and constant amount normalization methods.

cycle from that reaction. If the efficiency is the same between two assays $EO = 1.00$, if the numerator has the higher efficiency $EO > 1.00$, if the numerator has the lower efficiency $EO < 1.00$.

$$Efficiency\ Offset = \frac{2^{Cq(CV)} * qA(CV)}{2^{Cq(CA)} * qA(CA)} \tag{5}$$

Eq. (5) is used to compare the efficiencies of two qPCR assays with the same isolate. The next step is to see if the Efficiency Offset is the same between IP and IN isolates from the same sample. Illustrated in Fig. 3, the difference between $EO(IP_{CV/CA})$ and $EO(IN_{CV/CA})$ gives the efficiency offset difference *(ED)* between the isolates which varies with respect to the initial amount of DNA *(qA)* used in each qPCR assay. It is possible for the variation in percent input to be minimal and to have EO values for IP and IN assays to not equal 1.00, if the EO values for the IP and IN assays are the same. That is, if the IP and IN assays have equal efficiency offset then the percent input will be the same for the CV and CA methods.

The two far right columns in Fig. 2 highlight the variation of CV vs CA percent input as well as the efficiency offset difference. These two data sets have a Pearson's Correlation value of $r = 0.957$. This data, graphed in Fig. 4b, shows the strong correlation between the variation in CV vs CA normalization methods and the variation in PCR efficiency between CV and CA assays.

Mathematical proof showing the relationship between dilution factors DFv and DFa

Here we present the proof behind the novel method which allows a ChIP-qPCR researcher to calculate a percent input from Cq (quantification cycle) values generated by using a constant amount (ng) of DNA in each qPCR reaction. This method relies on a different dilution factor *(DFa)* that is unique for each sample and is based on the yielded amounts from each IP and IN isolate. To calculate an accurate percent input when a qPCR assay is run using a constant amount of DNA, *DFa* is the dilution factor *(DF)* used in the percent input equation (Eq. (2)).

We developed a direct proof to logically show how percent input, calculated with Cq generated from a constant amount and adjusted with *DFa,* is equal to a percent input calculated with Cq generated from a constant volume and adjusted with *DFv*.

The key axiom of our proof is explained by Tellinghuisen and Spiess's paper on qPCR analysis methods [2], which states that data from quantitative polymerase chain reaction can be represented

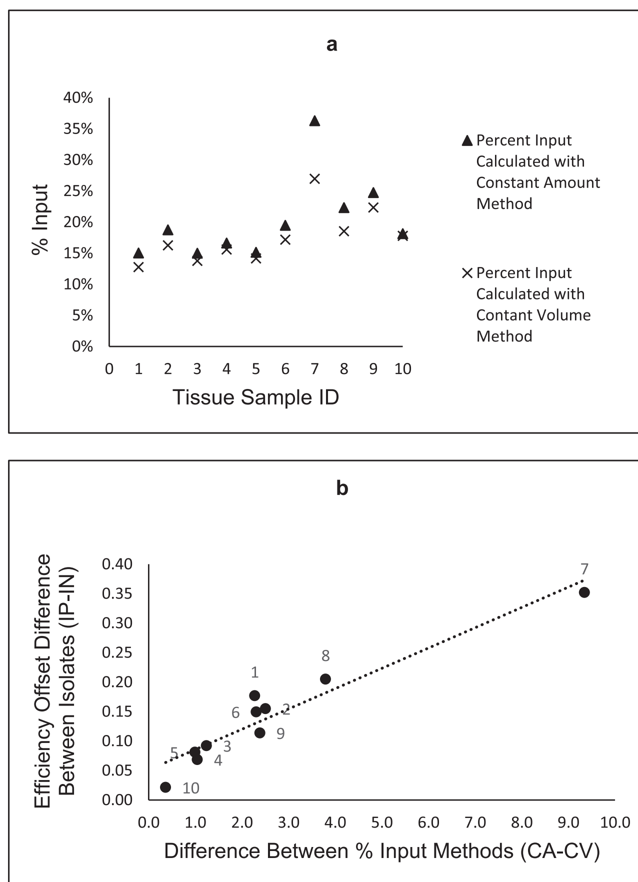

Fig. 4. (A). Graph comparing percent input normalization methods of ten individual ChIP-qPCR samples, Triangles represent percent input calculated with constant amount, X's represent percent input calculated with constant volume. Most data points are on top of or near their counterparts with the trend between data points being retained in both normalization methods. (B). The correlation of the difference between percent input normalization methods and the difference in amplification efficiency offset between IP and Input pairs. This shows how the perceived lack in precision of the constant amount normalization method is strongly correlated with the variance of PCR efficiency between reactions.

by an exponential growth equation. From this, we can solve for *Cq*, which gives us Eq. (6); where the *Threshold (ng)* is the number of nanograms of double stranded DNA required to break the fluorescent threshold and produce a Quantification Cycle (*Cq*) value. The *Initial Template (ng)* is the initial number of nanograms of the, to be amplified, template strand of DNA. *E* is the amplification efficiency i.e. when *E* equals 2.00, the amplicon is doubled each cycle.

$$Cq = Log_E\left(\frac{Threshold\ (ng)}{Intial\ Template\ (ng)}\right) \tag{6}$$

We can represent the *Initial Template (ng)* as the product of the total nanograms of chromatin *(qA)* and the fraction of chromatin *(r)* amplified by the Polymerase Chain Reaction, as described in Eq. (7).

$$Intitial\ Template\ (ng) = \ r * qA \tag{7}$$

Proof

If percent input (constant volume) is equal to percent input (constant amount), then it should be possible to derive Eq. (4) from this expression. Percent input, calculated from Cq values generated from using a constant volume (Cqv) and *DFv*, is equal to percent input calculated from Cq values generated from using a constant amount (Cqa) and *DFa*.

$$Percent\ Input = 2^{(Cqv(Input) - Log_2(DFv)) - Cqv(IP)} * 100 = 2^{(Cqa(Input) - Log_2(DFa)) - Cqa(IP)} * 100$$

Divide each side by 100 and take the Log base 2 of each side.

$$(Cqv(Input) - Log_2(DFv)) - Cqv(IP) = (Cqa(Input) - Log_2(DFa)) - Cqa(IP)$$

Replace *Cq* values with Eq. (6), our key axiom, derived from the qPCR exponential growth equation described in Tellinghuisen and Spiess's paper [2].

$$\left(Log_2\left(\frac{Threshold\ (ng)}{iT\,(INv)}\right) - Log_2(DFv)\right) - Log_2\left(\frac{Threshold\ (ng)}{iT\,(IPv)}\right) = \left(Log_2\left(\frac{Threshold\ (ng)}{iT\,(INa)}\right) - Log_2(DFa)\right) - Log_2\left(\frac{Threshold\ (ng)}{iT\,(IPa)}\right)$$

Simplify Log functions.

$$Log_2\left(\frac{Threshold\ (ng) * iT\,(IPv)}{iT\,(INv) * DFv * Threshold\ (ng)}\right) = Log_2\left(\frac{Threshold\ (ng) * iT\,(IPa)}{iT\,(INa) * DFa * Threshold\ (ng)}\right)$$

Take each side to the power of 2.

$$\frac{\cancel{Threshold\ (ng)} * iT(IPv)}{iT(INv) * DFv * \cancel{Threshold\ (ng)}} = \frac{\cancel{Threshold\ (ng)} * iT(IPa)}{iT(INa) * DFa * \cancel{Threshold\ (ng)}}$$

Threshold (ng) is the nanograms of double stranded DNA required to bind to sybrgreen for a qPCR reaction to pass the fluorescent threshold which is a constant value and cancels out on either side.

$$\frac{iT\,(IPv)}{iT\,(INv) * DFv} = \frac{iT\,(IPa)}{iT\,(INa) * DFa}$$

Replace Initial Template *(iT)* with Eq. (7). *qA* on the left side is replaced with *qV*M* since a constant volume is used, the amount is dependent on the concentration *M*.

$$\frac{r(IP) * qV * M(IP)}{r(IN) * qV * M(IN) * DFv} = \frac{r(IP) * qA}{r(IN) * qA * DFa}$$

Divide both sides by $\frac{r(IP)}{r(IN)}$.

$$\frac{\cancel{qV} * M(IP)}{\cancel{qV} * M(IN) * DFv} = \frac{\cancel{qA}}{\cancel{qA} * DFa}$$

qV and *qA* cancel out on their respective sides.

$$\frac{M(IP)}{M(IN) * DFv} = \frac{1}{DFa}$$

Concentration *M* can be written as Final Amount (*fA*) divided by Final Volume (*fV*).

$$\frac{fA(IP)/\cancel{fV\,(IP)}}{fA(IN)/\cancel{fV\,(IN)} * DFv} = \frac{1}{DFa}$$

Since the IP and IN samples are always eluted with the same volume, $fV(IP) = fV(IN)$, and these cancel each other out.

$$\frac{fA(IP)}{fA(IN)} * \frac{1}{DFv} = \frac{1}{DFa}$$

Take each side to the power of -1, and here is Eq. (4).

$$\frac{fA(IN)}{fA(IP)} * DFv = DFa$$

Conclusion

This new method for calculating a percent input from Cq values obtained from a qPCR assay using a constant amount of IP and IN Isolate, provides epigenetic researchers with more flexibility when designing a ChIP-qPCR, RIP-qPCR, and MeDIP-qPCR assays. This method allows for the use of the minimum template required for optimal efficiency in each qPCR reaction, thus extending the number of assays a researcher can run from each isolate. Using a constant amount of isolate material also allows the researcher to minimize the impact of variation in amplification efficiency between samples. It is so common for researchers to use a constant amount of material in a qPCR assay, that they may make the mistake of not using a constant volume of ChIP isolate, if not specified by their protocol. This method corrects for this oversight and provides percent input values that are just as reproducible as the constant volume method.

The major drawback to the constant amount percent input method is the requirement of quantifying the concentration of nucleic acids in each Isolate. If yields are low enough that they cannot be accurately measured by Qubit or Nano-drop or if measuring requires a substantial amount of isolate, then a constant amount method will waste more sample than it will save, in which case, it is best to use the original constant volume percent input method. For experiments attempting to measure the association of several different genes with one bio-molecular target (BMT), it is best to optimize the ChIP assay to ensure a quantifiable yield and use the constant amount percent input method to increase the number of possible qPCR assays per sample.

When publishing data obtained by nucleic acid co-immunoprecipitation analyzed by qPCR, specific details must be included for the experiment to be reproducible. In addition to following the minimum information for publication of qRT-PCR experiments (MIQE) guidelines [10] for publishing qPCR data, researchers should also specify the key parameters of the nucleic acid co-IP assay. These include, but are not limited to, specifying the volume of elution buffer used, how a dilution factor was calculated, the amount or volume of isolate used in a qPCR assay, and the equation used to normalize the data.

Since the first nucleic acid co-IP assay described in 1984 by Gilmour and Lis [11] and the invention of quantitative PCR, this technology has developed into a powerful analytical tool for understanding the epigenome. In the past decade, normalization by percent input has emerged as the standard method of ChIP-qPCR, RIP-qPCR, and meDIP-qPCR data analysis. Our work here, has subtlety, yet significantly improved upon this standard method. By providing researchers more options to normalize their data in comparable and reproducible ways, we have increased the potential for the collaboration of research across the globe.

Acknowledgments

This work was supported with funding from the National Institute of Environmental Health Sciences grant 2RO1 ES-019583 (AMA). Normalization method and proof developed by ERS. Paper written by all authors.

Declaration of Competing Interest

The authors declare all work described in this paper was conducted in the absence of any conflicts of interest.

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] ThermoFisher Scientific. ChIP analysis. (n.d.). Retrieved February 25, 2020, from https://www.thermofisher.com/us/en/home/life-science/epigenetics-noncoding-rna-research/chromatin-remodeling/chromatin-immunoprecipitation-chip/chip-analysis.html.

[2] J. Tellinghuisen, A.N. Spiess, Comparing real-time quantitative polymerase chain reaction analysis methods for precision, linearity, and accuracy of estimating amplification efficiency, Anal. Biochem. 449 (2014) 76–82 15 March, doi:10.1016/j.ab.2013.12.020.

[3] F.M. Mendez, F.J. Núñez, R.I. Zorrilla-Veloz, P.R. Lowenstein, M.G. Castro, Native chromatin immunoprecipitation using murine brain tumor neurospheres, J. Vis. Exp. 131 (2018) e57016 10.3791%2F57016.

[4] R.C. Akkers, U.G. Jacobi, G.J.C. Veenstra, Chromatin immunoprecipitation analysis of xenopus embryos, in: S. Hoppler, P. Vize (Eds.), Xenopus Protocols. Methods in Molecular Biology (Methods and Protocols), Humana Press, Totowa, NJ, 2012 vol 917https://link.springer.com/protocol/10.1007%2F978-1-61779-992-1_17#Sec16.

[5] P. Asp, How to combine ChIP with qPCR, in: N. Visa, A. Jordán-Pla (Eds.), Chromatin Immunoprecipitation. Methods in Molecular Biology, Humana Press, New York, NY, 2018 vol 1689 https://link.springer.com/protocol/10.1007/978-1-4939-7380-4_3 .

[6] S. Flanagin, J.D. Nelson, D.G. Castner, O. Denisenko, K. Bomsztyk, Microplate-based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signaling of complex genomic events, Nucleic Acids Res. 36 (3) (2008) e17 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2241906/, doi:10.1093/nar/gkn001.

[7] J.L. Cotney, JP. Noonan, Chromatin immunoprecipitation with fixed animal tissues and preparation for high-throughput sequencing, Cold Spring Harb. Protoc. 2015 (4) (2015) 419 Published 2015 Apr 1 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5956515/ , doi:10.1101/pdb.err087585.

[8] S. Bhatia, J. Matthews, P.G. Wells, Characterization of epigenetic histone activation/repression marks in sequences of genes by chromatin immunoprecipitation-quantitative polymerase chain reaction (ChIP-qPCR), in: J. Hansen, L. Winn (Eds.), Developmental Toxicology. Methods in Molecular Biology, Humana, New York, NY, 2019 vol 1965, doi:10.1007/978-1-4939-9182-2_25.

[9] M. Haring, S. Offermann, T. Danker, et al., Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization, Plant Methods 3 (2007) 11, doi:10.1186/1746-4811-3-11.

[10] S.A. Bustin, et al., The MIQE guidelines: minimum information for publication of quantitative real-time PCR experiments, Clin. Chem. 55 (4) (2009) 611–622 1 April, doi:10.1373/clinchem.2008.112797.

[11] D.S. Gilmour, J.T. Lis, Detecting protein-DNA interactions *in vivo*: distribution of RNA polymerase on specific bacterial genes, in: Proceedings of the National Academy of Sciences, 81, 1984, pp. 4275–4279, doi:10.1073/pnas.81.14.4275.

[12] N. Kiyotaka, et al., Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres, Genetics 163 (2003) 1221–1225 MarchPMID:12663558 .

[13] E.R. Solomon, K.K. Caldwell, A.M. Allan, Developmental arsenic exposure is associated with sex differences in the epigenetic regulation of stress genes in the adult mouse frontal cortex, Toxicol. Appl. Pharmacol. 391 (2020) 114920, doi:10.1016/j.taap.2020.114920.