Supplementary material
Chainy, an universal tool for standardized relative quantification in real-time PCR

Izaskun Mallona, Anna Díez-Villanueva, Berta Martín and Miguel A. Peinado

Contents

List of Figures

List of Tables

1 Supplemental methods

1.1 Experimental procedures

Cell lines SW480 (colon cancer) and PTN2 (immortalized normal prostate epithelium) were purchased from the ATCC (Manassas, Virginia) and growth in standard conditions.

Chromatin immunoprecipitation was performed as described in (Rodriguez et al., 2008) with ChIP-grade antibodies specific for H3 (Abcam, ab1791), H3K4me3 (Millipore, 17-614) and Rb IgG (Jackson Immunoresearch 011-000-120).

Total RNA was extracted with the PureLink RNA mini kit (Ambion) following manufacturer's instructions, including the DNase I digestion step. cDNA was synthesized using SuperScript III (Invitrogen) and random hexamers.

Cells were treated with 0.3 μM Trichostatin A (Sigma T8552) for 24 h or 5 μM and 10 μM 5-Azacytidine (Sigma A2385) for 96, 120 and 144 h.

Real-time PCRs were conducted in triplicates (gene expression) or duplicates (ChIP) on a Light Cycler 480 with LightCycler DNA Master SYBR Green I chemistry. Primers and targets are described in table S1 and figure S1. PCR amplification data were extracted from the LC software version 1.5.

1.2 Data load

Chainy home menu offers two entry levels for the uploading of either raw or pre processed data. Raw fluorescence data files directly generated by most popular thermocyclers, including the Light Cycler 480 II (Roche), 7900 HT (Applied Biosystems) and Mx3000p (Agilent Technologies), as well as plain text files with structured data and RDML files are accepted.

1.3 Kinetic fitting

In Real time PCR, the kinetics of DNA amplification is visualized as the amount of fluorescence detected. The most informative descriptor of the amplification is the Cq (cycle of quantification), when the fluorescence takes off above the background. Another important parameter is the efficiency of the reaction, defined as the fold increase of the DNA per cycle, hence ranging from 1 (no amplification) to 2 (complete doubling) (Schlereth et al., 1998; Liu and Saint, 2002; Tichopad et al., 2002).

Chainy implements the qpcR (Ritz and Spiess, 2008) kinetic fitting methods to extract Cq and efficiency values at the maximum of the first and second derivative curves; at the 20% value of the fluorescence at second derivative maximum, as defined by Corbett Research; and at the intersection of a tangent on the first derivative maximum with the abscissa (Cy0 method) (Guescini et al., 2008). Kinetic outliers are detected at two tiers: first, as amplifications with unsuccessful model fitting (name flanked by a single asterisk, *name*), second, as Cq and efficiency calling failures (tagging by two asterisks, **name**). Note that Chainy will not remove these samples from the analysis. The Akaike's goodness of fit criterion AIC is reported for each model fit. By default, no baseline correction is performed (recommended for SYBR Green), but users can opt

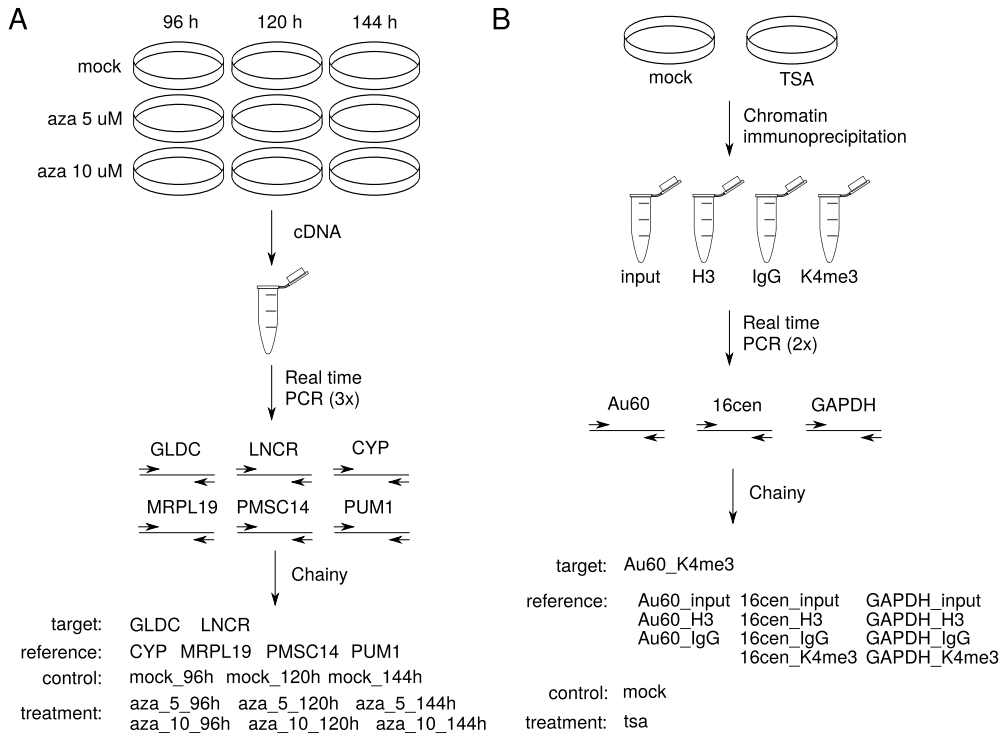


Figure S1: **Experimental design.** **A**, PTN2 cells (human immortalized normal prostate epithelium) was subjected to two different concentrations of 5-azacytidine and cells were collected at three time points. GLDC and LNCR were measured as targets to be quantified and four different candidate reference genes were measured. **B**, SW480 cells (human colon cancer cell line) was treated with Trichostatin A (TSA) and chromatin was subjected to three different immunoprecipitations. The amount of H3K4me3 at Au60 locus was the target of quantification; other pairs of loci and immunoprecipitation were evaluated as references.

to subtract the background signal detected from the fitted amplification curve (best suited for Taqman).

Individual curve for each sample is plotted using any of the kinetic methods selectable from a pull-down list. Navigation through the different samples is achieved using previous/next buttons. Results are displayed on screen and may be downloaded as a text file.

1.4 Experiment annotation and inspection

To ease the differential quantification and references stability assessment Chainy will ask for a simple text file reporting the basic design of the experiment (figure S1). An appropriate template prefilled with the user's uploaded data may be downloaded and easily edited using any word processor or spreadsheet software. In this step, each well or tube must be labeled according to the amplicon (i.e.: gene or locus name, primers name, etc.) and the sample or group name (i.e.: control vs. treated, normal, vs. tumor, time points, etc.). For raw fluorescence RDML inputs, the template is already filled with the amplicon and treatment names; the user might nonetheless edit the template to exclude some samples (i.e. failing replicates). Samples with matching labels will be treated as technical replicates and averaged afterwards. To achieve accurate quantification technical and biological replicates must be considered distinct entities. Therefore, biological replicates should contain different names (i.e. treated1, treated2 and so on). Once the replicates are collapsed, the scaled quantities (q) are calculated by comparison to the sample:amplicon combination with the highest levels (minCq) as follows:

$$q = E^{minCq - sampleCq} = E^{\Delta Cq} \tag{1}$$

In this way, the sample:amplicon combination with lower Cq is set to 1. The standard deviation of the scaled quantity (sd_q) is calculated according to the formula:

$$sd_q = E^{\Delta Cq} \cdot log(E) \cdot s_{sampleCq} \tag{2}$$

Being $log(E)$ the natural logarithm of the sample replicates mean efficiency value and $s_{sampleCq}$ the standard deviation of the sample replicates Cq values.

In the next step, Chainy renders a table and a plot to visually inspect and quality check non normalized data. The table summarizes the scaled relative quantities with their standard deviation. Error-propagated quantities take into account replicates' variability, therefore high values will indicate poor repeatability. It should be noted that in these data have not been yet normalized. Hence, the scaled quantities are highly dependent on the starting amount of template and cannot be directly compared without further normalization. A barplot to visually check the scaled quantities is included at the ZIP-compressed final report. Independently, Chainy draws a scatter plot which displays the individual efficiency and Cq readouts using a combination of color/shape codes according to the experimental design (blue for targets, red for references; a circle for controls, a cross for treatments). In this case, dispersed replicates indicate high technical variability, whilst tightly grouped replicates indicate robust technical repeatability.

1.5 Differential quantification

In the next step Chainy computes fold changes between experimental groups following the qpcR (Ritz and Spiess, 2008) strategy, in which Cqs and efficiencies are arithmetically averaged. Whilst

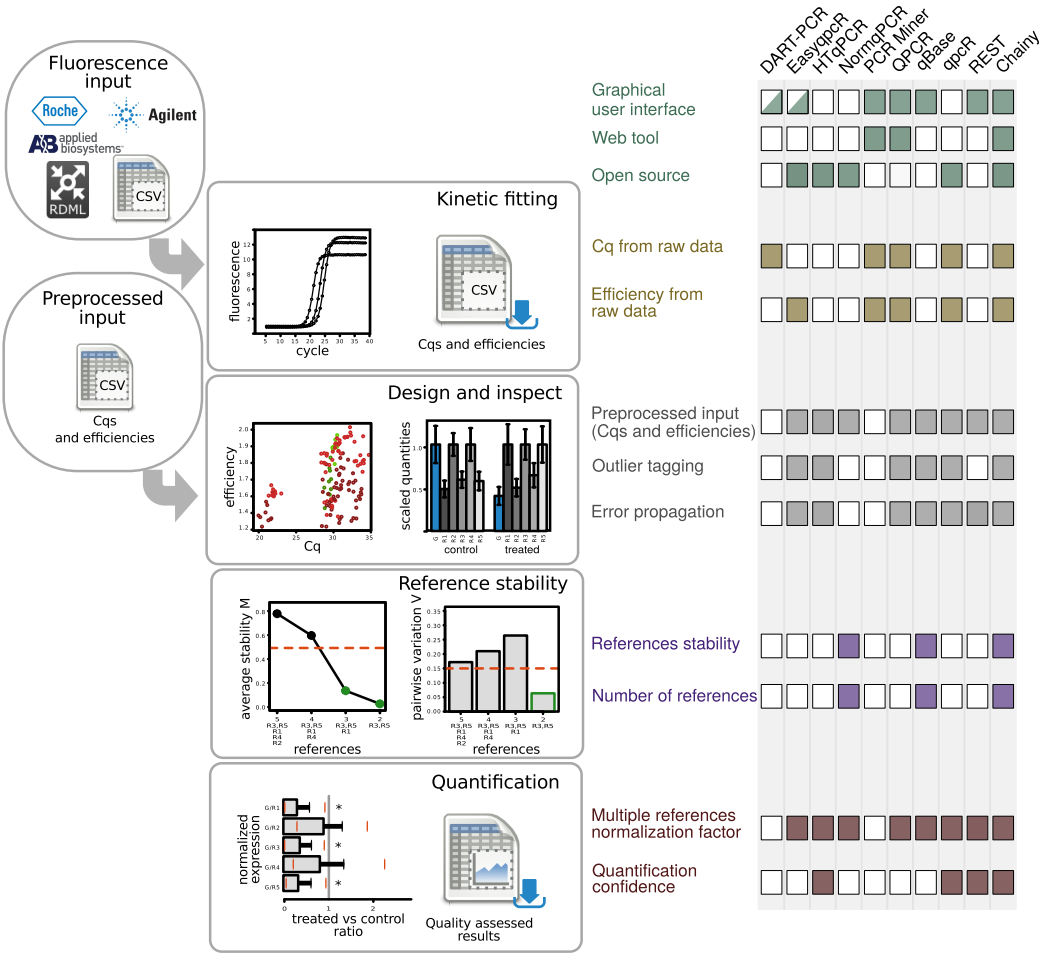


Figure S2: **Data flow**. Chainy offers an integrative workflow with a graphical user interface. Chainy capabilities include efficiency and Cq calling from raw fluorescence, quality check, reference stability evaluation and statistically assessed relative quantification. The right panel presents a comparison of available tools able to perform relative quantification or used to estimate efficiencies by kinetic methods (software listing and capabilities adapted from (Pabinger et al., 2014); half filled squares indicate that DART-PCR is an Excel spreadsheet and EasyqpcR an R package with an R inbuilt graphical interface).

quantitatively equivalent to the geometric averaging of the relative quantities as described by (Vandesompele et al., 2002), the former allows introducing propagated errors into the gene expression ratios (Nordgård et al., 2006).

According to the user's choice, Chainy is able to calculate the fold changes to either individual references or normalization factors composed by multiple references.

Chainy assesses the fold change between treatments by permutation tests. Briefly, scaled quantities are taken from both groups (i.e. treatment and control) and randomly reallocated 100 times. For each randomization, a treatment vs control ratio is calculated. The number of tests in which the shuffled ratio is below or above the original ratio can be read as a p-value of a test against the null hypothesis that the ratios calculated by permutation are just by chance (Pfaffl et al., 2002).

1.6 Reference stability assessment

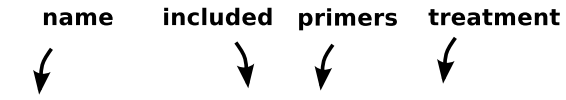
Chainy evaluates the suitability of references by the standard geNorm method (Vandesompele et al., 2002) providing a ranking based on their stability. The optimal normalization factor with the minimum number of genes (lower V value) and the highest stability (lower M value) is also reported. Chainy is able to normalize ChIP-qPCR results by leveraging the cross-loci and/or intra-locus controls included in ChIP-qPCR settings, thus enabling the normalization process to be customized to the user's choice. It should be noted that Chainy performs an agnostic evaluation of references and does not take into account the biological rationale of the experiment. Therefore the choice of the most appropriate reference should be ultimately decided by the investigator.

1.7 Availability and runtime

Chainy is written in R/shiny and freely accessible at `http://maplab.cat/chainy` without registration. As the calculations are performed at the server-side, the application can be accessed from commodity computers or mobile devices, as far as they have a recent Web browser installed. Chainy can be run locally by downloading its source code from `http://bitbucket.com/imallona/chainy` (under the GNU Public License terms); an R script to check and install dependencies as well to launch the application is included. Chainy imports the qpcR (Ritz and Spiess, 2008), NormqPCR (Perkins et al., 2012) and RDML R packages. The user manual is available at the main page and covers results interpretation. A zip-compressed verbose report conveying the outputs and the user parameters can be downloaded at the final step. The final report includes barplots for both the scaled and the differential quantifications (for the sake of simplicity, the dynamic interface only renders tables to summarize these results).

2 Supplemental results

Chainy's graphical user interface is built to maximize usage versatility (figure S2). The simplest application estimates PCR efficiency from raw fluorescence inputs. However, Chainy's full capabilities are better depicted when used as a comprehensive workflow from raw data to differential quantification. To exemplify the whole process we have applied Chainy in two classic experimental settings. The first one corresponds to the quantification of a histone modification



name	included	primers	treatment
I1..96H.C_ciclof	TRUE	cyp	control_96
I2..96H.C_ciclof	TRUE	cyp	control_96
I3..96H.C_ciclof	TRUE	cyp	control_96
I16..120H.C_ciclof	TRUE	cyp	control_120
I17..120H.C_ciclof	TRUE	cyp	control_120
I18..120H.C_ciclof	TRUE	cyp	control_120
J7..144H.C_ciclof	TRUE	cyp	control_144
J8..144H.C_ciclof	TRUE	cyp	control_144
J9..144H.C_ciclof	TRUE	cyp	control_144
I10..96H._AZA.5_ciclof	TRUE	cyp	aza_5_96
I11..96H._AZA.5_ciclof	TRUE	cyp	aza_5_96
I12..96H._AZA.5_ciclof	TRUE	cyp	aza_5_96
I13..96H._AZA.10_ciclof	TRUE	cyp	aza_10_96

Figure S3: **Design file format**. Chainy requires a structured text file (tab-separated) with 4 columns and as many rows as PCR wells/tubes. Each row must indicate the well/tube name, whether the datum is included or not, the primers used and the experimental condition.

by chromatin immunoprecipitation (ChIP) and the second to the analysis of differential gene expression in response to a drug treatment (figure S1).

2.1 ChIP enrichment analysis

The histone deacetylase inhibitor trichostatin A (TSA) disrupts the epigenetic landscape affecting the distribution of histone modifications. In this example we analyze the differential enrichment of the repressive H3K4me3 mark at an specific Alu sequence identified as Au60 upon TSA treatment of the SW480 colon cancer cells. Using standard nomenclature implemented in Chainy, the H3K4me3 measurement at the Au60 locus represents the target of our analysis. Two other loci, the active GAPDH and the inactive 16cen pericentromeric region (table S1), were also evaluated. Three controls were included: the chromatin input diluted 1:5, immunoprecipitated histone H3 and a blank control of immunoprecipitated IgG (background signal). Thus, the experimental groups to be compared are treated vs. control (figure S1).

Raw fluorescence data were exported from the thermocycler (supplemental data 1) and directly uploaded to Chainy using the *Go chainy raw* button. Cy0 Cq and efficiencies were calculated using default parameters. No kinetic outliers were detected; even though some samples amplified very late and did not reach the plateau (figure S4A-C). The next step will ask for a file with the experimental design (analyzed loci names and sample grouping). To ease the process, Chainy generates a template file that the user may download and fill-out using any text processor or spreadsheet software. The template must include four columns: the well/tube *name*, whether the datum should be *included*; the *primers* used; and the *treatment* name (figure S3). The template file before and after being annotated is provided as supplemental files 2 and 3. After uploading the annotated file, clickable lists allow the selection of the H3K4me3 data at Au60 as target (leaving the rest of loci and ChIP combinations as references) and the treated-control groups. Afterwards, the *Overview* screen presents a table displaying the scaled relative quantities of all loci. The downloadable final report also includes a barplot depicting these data. In this example, the efficiency calling was consistent among the replicates (figure S4D).

In the current setting, the most conventional normalization option is using either the input

or the H3 signal at the target locus. Nevertheless, signal at other loci or a combination of determinations may be more suited depending on the rationale of the experiment and providing more robust results. Chainy allows the agnostic evaluation of multiple references providing available alternatives to obtain relative measurements of the target loci. The investigator should ultimately decide the most appropriate normalization factor. In our example setting, we set the amplicon and immunoprecipitation pair Au60 H3K4me3 as target to be quantified and the rest (16cen H3K4me3, 16cen H3, 16cen input, Au60 H3, Au60 input, GAPDH H3K4me3, GAPDH H3, GAPDH input) as putative references whose normalization suitability was being checked. IgG-related data were discarded due to their unspecificity. A normalization factor built upon all the references was found to be stable (geNorm's $M = 0.48 < 0.5$); as the stepwise exclusion of the least stable gene of a normalization factor only can increase stability, subset normalization factors were also under the stability cut-off (figure S4E and table S2). Regarding the optimal number of references, the smallest normalization factor containing 16cen H3 and GAPDH H3 satisfied the stability criterion (geNorm's $V = 0.02 < 0.15$, figure S4F and table S2).

TSA treated cells showed statistically significant enrichment of H3K4me3 at the target Au60 using the multiple normalization factor; most of the individual references led to similar fold change estimates (figure S4G-H and table S3).

2.2 Differential expression analysis

In some genomic contexts, DNA demethylating agents induce local DNA hypomethylations and gene re-expression. We wondered whether this was the case for the GLDC locus (figure S1) and its nearby lncRNA LNCR target in PTN2 epithelium cells, which show high levels of regional DNA methylation (data not shown). Total RNA was extracted after 96, 120 and 144h of being treated with 5 μm and 10 μM of 5-Azacytidine (see supplemental methods). CYPA, MRPL19, PSMC4 and PUM1 were evaluated as a reference panel and GLDC and the long non coding LNCR as targets. To describe the expression ratios, the cells treated with the two concentrations of aza-cytidine were compared to the mock-treated at the same time course data point. Kinetic data analysis including outlier detection was conducted as in the previous ChIP analysis (figure S5A to S5C).

The scaled quantities to the gene with highest expression (CYPA at the mock-treated after 96 h) showed a gradual decrease of overall transcript levels at both 5-Azacytidine concentrations. The visual inspection of the scaled quantities' standard deviation, which collapse technical replicates, showed no obvious failing replicates. The Cq-efficiency space pointed to a consistent biological grouping among targets, references, treatments and controls (figure S5D).

Regarding the reference genes stability, the highest ranked reference genes PUM1 and PSMC4 (M=0.12) were below the proposed cut-off of 0.5 for homogeneous datasets (Vandesompele et al., 2002), thus fulfilling the stability requisite. Nonetheless, their V value $V = 0.18 > 0.15$ recommend the addition of a third gene, CYPA. The normalization factor for the three genes passed both M and V thresholds ($M = 0.4$, $V = 0.14$). The inclusion of the least stably expressed gene, MRPL19 made the normalization factor unreliable ($M = 0.52$, figure S5E and F and table S4). To avoid the noise inherent of including MRPL19 as reference, we updated the experimental design to tag MRPL19 as a target, therefore excluding it from the reference normalization factor.

The 5-Azacytidine treatment produced an overall upregulation of the GLDC and LNCR genes

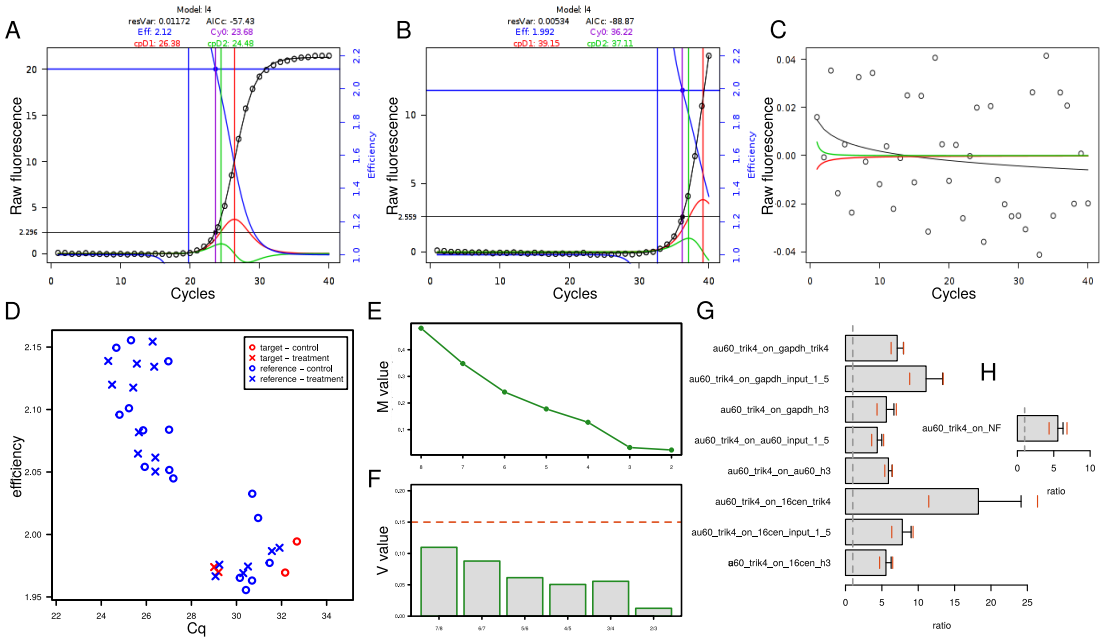


Figure S4: **ChIP quantification.** Relative enrichment of H3K4me3 at the Alu element Au60 after TSA treatment. The GAPDH and pericentromeric 16cen loci were also evaluated as cross-loci positive and negative controls, respectively. Input, H3 and IgG ChIPs were also measured and used as intra-loci controls. Data were processed in Chainy as a TSA-treated vs mock-treated cells, with H3K4me3 at the Au60 locus as target and the rest of determinations as putative references (see supplemental methods). **A**, kinetic fitting result of H3K4me3 at the GAPDH locus (early amplification). **B**, fitting of IgG at the GAPDH locus: despite the late take-off and lack of plateau, the efficiency value is successfully retrieved. **C**, failing amplification of input at the Au60 locus. **D**, efficiency vs Cq landscape of data points represented according to the experimental design. **E**, geNorm's M plot showing the references suitability (dotted red line, proposed cut-off of 0.5). **F**, geNorm's V plot (the lowest V value points to the normalization factor with optimal number of control genes). **G**, differential quantification of H3K4me3 at the target Au60 locus as compared to each of the references separately (dotted gray line, no changes); **H**, to a normalization factor leveraging all the references.

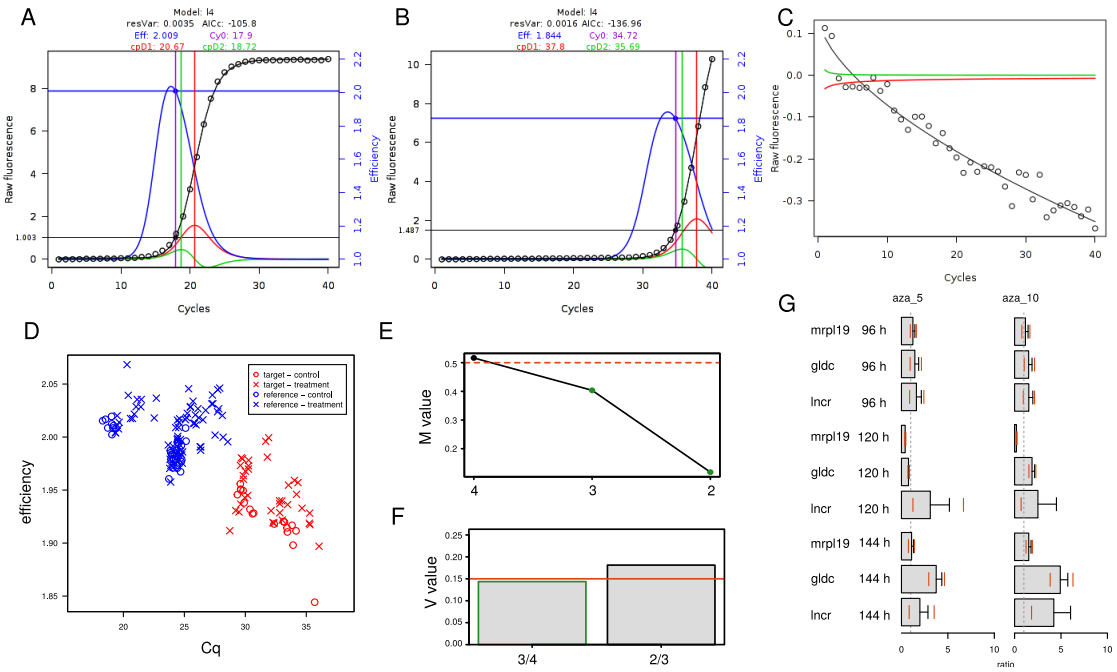


Figure S5: **Differential expression analysis.** GLDC and LNCR differential expression analysis after treatment with two doses of 5-azacytidine (see supplemental methods). CYPA, MRPL19, PSMC4 and PUM1 are evaluated as a putative reference panel. **A**, kinetic fitting result of CYPA in a mock-treated sample at 96 h (early amplification). **B**, fitting of LNCR in a mock-trated sample at 120 h: despite the late take-off and lack of plateau, the efficiency value is successfully retrieved. C, non-amplifying non template control unable to be fitted. **D**, efficiency vs Cq space plot: the tight grouping of treatments and primers indicate consistent technical repeatability. **E**, the geNorm's M highest value for the fourth gene, MRPL19, indicates unsuitability as reference gene. **F**, geNorm's V plot points to an optimal normalization factor with three genes (PUM1, PSMC4 and CYPA). **G**, expression ratios against an averaged normalization factor (note that MRPL19 is evaluated as a target as it does not fulfill the references stability criteria).

during the late points of the time course (figure S5G and table S5). GLDC and LNCR did not change synchronously, being GLDC the most upregulated gene at 144 h at both concentrations whereas the summit in LNCR expression occurred at 120 h for 5-Azacytidine 5 μM but at 144 for 5-Azacytidine 10 μM. The unstable reference gene MRPL19 showed an erratic expression behaviour, ranging from down- to upregulation, including non detectable changes. Thus, the variability in MRPL19 expression was consistent with the previous result of reference stability inappropriateness.

2.3 Technique resolution validation

We evaluated the resolution of the workflow by using a series of known quantities of template, two independent targets and a normalization factor with three references. To do so, we generated a series of templates carrying a fixed amount of human genomic DNA (0.9 ng/μL) spiked in with a dilution series of the luciferase plasmid pGL3basic (2 fg/μL to 2 pg/μL; figure S6A). We quantified two plasmid-specific sequences placed at the pGL3basic backbone, LUC and LUC2, against a normalization factor composed by three single copy human targets (THOC3, GAPDH and 16CEN; tables S1) whose readout is expected to be invariant.

This design reproduces a realistic complex template (human genomic DNA). As evaluated for two target sequences placed at the same template (i.e. being equimolecular), it provides a parallel scrutiny of both targets. Finally, the dynamic range of quantification is easily evaluated by using serial dilutions of the plasmid.

The dilution series resulted in a linear decay of Cq as the plasmid quantity diminished (figure S6B; linear regression adjusted R squared of 0.986 (LUC) and 0.998 (LUC2)) whilst the efficiency remained unaffected (figure S6B). Taking the 20 fg/μL as yardstick (in Chainy nomenclature the *control*, whose fold change is 1), the three dilutions tested (0.1x, 10x, and 100x) were quantified with confidence intervals of 0.39–0.46, 11.19–14.68 and 65.96–73.48 (LUC amplicon); and 0.13–0.21, 7.88–10.04 and 63.27–94.73 (LUC2), indicating the relative quantification provides a consistent readout of the absolute quantities spanning four orders of magnitude (table S6 and figure S6C) for the two targets tested independently.

3 Supplemental discussion

Real-time PCR has become the technique of choice in a wide range of quantitative applications. In most settings the quantification of the target is referred to one or more references, the so-called normalization factor. The selection of a stable normalization factor is the key point to deliver reliable and accurate results (Bustin et al., 2013). Surprisingly, it is not uncommon the arbitrary choice of one or more references without an appropriate appraisal of the normalization process, which has important drawbacks in the final results (reviewed in (Wong and Medrano, 2005; Guénin et al., 2009; Kozera and Rapacz, 2013)). References validation has become a publication trend, with over 10,000 papers citing the most used algorithm (Vandesompele et al., 2002). Most studies evaluate a shortlist of reference genes and provide a ranked final panel. However, even though reference validation studies clearly provide good normalization factors for the assayed experimental conditions, it is impossible to say whether these results can be extrapolated to a different setting, specifically when conducted in a different lab. Despite the obvious advantages of evaluating the references stability as a part of the quantification workflow, its inclusion is still marginal, partly because of the lack of easy-to-use software. This drawback implies that the vast

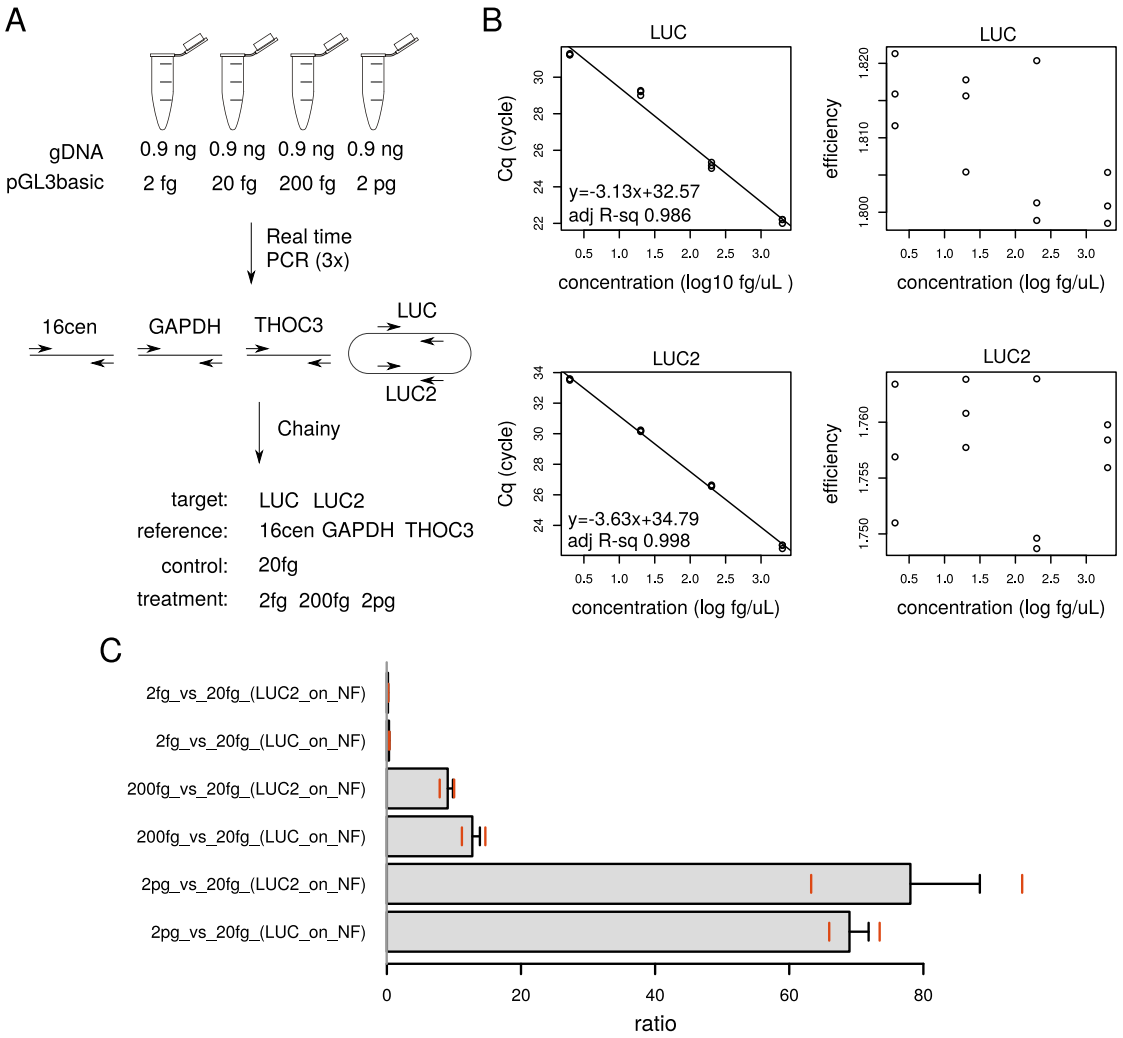


Figure S6: **Experimental resolution validation**. Fold change of two amplicons, LUC and LUC2, using plasmid templates at 2pg/μL, 200 fg/ul and 2 fg/μL as compared to 20 fg/μL. Expected fold changes are 0.1, 10 and 100 for both LUC and LUC2, respectively. **A**, experimental design; template amounts refer to DNA mass per μL of final volume. **B**, Cq vs concentration plots indicating the dilution linearity; and lack of dilution effects on efficiency values. **C**, fold changes (mean and standard deviations; confidence intervals are depicted in orange). adj R-sq, adjusted R squared (linear regression); NF, normalization factor.

majority of applications of real time PCR rely on references without internal validation.

Another frequently overlooked aspect is the potential variability in amplification efficiency. The classical quantification methods assume perfect doubling of the template amount each cycle, this is efficiency = 2. Experimental data show, however, that efficiencies vary (Ramakers et al., 2003) depending on primer and template properties, among other factors (Mallona et al., 2011). Importantly, differences in efficiency have huge impacts on quantification (Bar et al., 2003). One strategy to circumvent this problem is to estimate the efficiency by using calibration curves based in dilution series, which are laborious and sample consuming (Ståhlberg et al., 2003). Efficiencies, however, can be directly evaluated by kinetic methods (Schlereth et al., 1998), providing also a powerful method for outlier detection (Bar et al., 2003). Remarkably, different kinetic methods run on the same dataset produce different Cqs and efficiency estimates. In some cases, a noticeable number of efficiency values above 2 are called (Ruijter et al., 2013). Importantly, quantification results are similar regardless of the Cq and efficiency algorithm used (Ruijter et al., 2013); keeping the procedure fixed in all samples is strongly recommended to detect real differences in efficiency between samples rather than differences between methods (Bar et al., 2003).

Another important aspect that is usually overlooked in this kind of analyses is the statistical treatment of the data. The most common and easy-to-interpret output of relative quantification in different samples is fold change. Fold changes are ratios and are likely to show no-normal distributions and high variances; therefore, the use of parametric tests is not straightforward. Permutation or randomization tests are a handy alternative, as they do not have data distribution assumptions and present a statistical power as good as parametric tests, although being slightly conservative (Pfaffl et al., 2002). Permutation testing relies on the probability of the observed ratio to occur under the null hypothesis of no treatment effect. Therefore, the statistical test is performed by taking the quantities from both groups (i.e. treatment and control), randomly reallocating them regardless of the group, recalculating the ratios and checking counting the number of tests in which the shuffled ratio is below or above the original ratio. This proportion can be read as a p-value of a test against the null hypothesis that the ratios calculated by permutation are just by chance (Pfaffl et al., 2002).

One of the most common applications of real time PCR aims to retrieve the gene expression variation between two conditions (Andersen et al., 2004). Variations in the amount of starting material, RNA integrity and cDNA synthesis efficiency, among others, may have an important impact in the final result (Dowling et al., 2016). Expression stability is crucial and has to be tested for the experimental conditions, as no universally valid references with constant expression exist (Vandesompele et al., 2002). A simple strategy that stabilizes most of variability is to normalize the target by an internal reference gene. This internal control is processed jointly with the gene of interest, being subjected to almost the same variability sources. If the control gene is stably expressed, the technical noise will be low and robust comparisons between conditions, including different samples, cells or treatments, may be performed. Importantly, Chainy evaluates the adequacy of the reference genes selected by the user, but it can also perform an agnostic analysis of both candidates and non candidates. The evaluated reference genes are ranked according to their stability, facilitating the decision making of how to optimize and standardize the relative gene expression analysis.

To illustrate the utility of Chainy we have also selected another common application or real time PCR: the enrichment analysis of chromatin immunoprecipitates (ChIP). ChIP-qPCR settings may differ in many ways and frequently present several positive and negative controls. Com-

monly used controls include a negative or background control (IgG immunoprecipitate obtained without specific antibody) and the so called input representing the total load of chromatin. Additional positive and negative controls include the analysis of loci with known chromatin states, for instance highly transcribed regions enriched in active chromatin marks (i.e. the GAPDH gene) or non-transcribed regions enriched in repressive marks (i.e. pericentromeric regions). The normalization strategy with such a complex setting is often underestimated in ChIP-qPCR data analysis. We consider two main scenarios to normalize ChIP-qPCR results: the cross-loci and the intra-locus settings. The cross-loci normalization, much alike as gene expression analysis, compares the target quantities to an external control locus which is expected to present an invariant level of the qPCR readout. On the other hand, the intra-locus method compares the target binding to both the chromatin quantity (using the input as reference) or to the nucleosome abundance (using an antibody specific to an invariant part of histone H3) of the same locus. These two normalization methods may or may be not comparable for two reasons: first, technically, they are subjected to different experimental procedures; and second, biologically, the results interpretation heavily relies on the yardstick used (i.e. comparing to the H3 abundance of the same locus takes into account nucleosome density). As well as for gene expression analysis, Chainy provides an integrated qPCR framework aware of the intrinsic technical challenges associated with the chromatin enrichment measurement. Chainy allows the appraisal of the effect of different normalization criteria on the final result, that may be essential to correctly interpret the data.

Chainy tool is specially suited to analyze treated vs control experimental designs, although the flexible workflow is easily adaptable to other designs. Following the MIQE guidelines, it assesses the efficiency in each step and allows normalization against multiple reference genes. The tool implements sophisticated options for the advanced user allowing reliable, traceable and reproducible real-time PCR quantifications. In the current version of Chainy some advanced options have been limited (e.g.: inter-run calibrators, stability assessment in more than 10 references) to empower a straightforward workflow, easing its adoption by sporadic and non expert users.

References

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 research*, 64(15):5245–5250.

Bar, T., Ståhlberg, A., Muszta, A., and Kubista, M. (2003). Kinetic outlier detection (kod) in real-time pcr. *Nucleic acids research*, 31(17):e105–e105.

Bustin, S. A., Benes, V., Garson, J., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G., Wittwer, C. T., Schjerling, P., Day, P. J., Abreu, M., Aguado, B., Beaulieu, J.-F., Beckers, A., Bogaert, S., Browne, J. A., Carrasco-Ramiro, F., Ceelen, L., Ciborowski, K., Cornillie, P., Coulon, S., Cuypers, A., De Brouwer, S., De Ceuninck, L., De Craene, J., De Naeyer, H., De Spiegelaere, W., Deckers, K., Dheedene, A., Durinck, K., Ferreira-Teixeira, M., Fieuw, A., Gallup, J. M., Gonzalo-Flores, S., Goossens, K., Heindryckx, F., Herring, E., Hoenicka, H., Icardi, L., Jaggi, R., Javad, F., Karampelias, M., Kibenge, F., Kibenge, M., Kumps, C., Lambertz, I., Lammens, T., Markey, A., Messiaen, P., Mets, E., Morais, S., Mudarra-Rubio, A., Nakiwala, J., Nelis, H., Olsvik, P. A., Perez-

Novo, C., Plusquin, M., Remans, T., Rihani, A., Rodrigues-Santos, P., Rondou, P., Sanders, R., Schmidt-Bleek, K., Skovgaard, K., Smeets, K., Tabera, L., Toegel, S., Van Acker, T., Van den Broeck, W., Van der Meulen, J., Van Gele, M., Van Peer, G., Van Poucke, M., Van Roy, N., Vergult, S., Wauman, J., Tshuikina-Wiklander, M., Willems, E., Zaccara, S., Zeka, F., and Vandesompele, J. (2013). The need for transparency and good practices in the qpcr literature. *Nature Methods*, 10(11):1063–1067. Commentary.

Dowling, C. M., Walsh, D., Coffey, J. C., and Kiely, P. A. (2016). The importance of selecting the appropriate reference genes for quantitative real time pcr as illustrated using colon cancer cells and tissue. *F1000Research*, 5.

Guénin, S., Mauriat, M., Pelloux, J., Van Wuytswinkel, O., Bellini, C., and Gutierrez, L. (2009). Normalization of qrt-pcr data: the necessity of adopting a systematic, experimental conditions-specific, validation of references. *Journal of experimental botany*, 60(2):487–493.

Guescini, M., Sisti, D., Rocchi, M. B., Stocchi, L., and Stocchi, V. (2008). A new real-time pcr method to overcome significant quantitative inaccuracy due to slight amplification inhibition. *BMC bioinformatics*, 9(1):1.

Kozera, B. and Rapacz, M. (2013). Reference genes in real-time pcr. *Journal of applied genetics*, 54(4):391–406.

Liu, W. and Saint, D. A. (2002). Validation of a quantitative method for real time pcr kinetics. *Biochemical and biophysical research communications*, 294(2):347–353.

Mallona, I., Weiss, J., and Egea-Cortines, M. (2011). pcrefficiency: a web tool for pcr amplification efficiency prediction. *BMC bioinformatics*, 12(1):404.

Nordgård, O., Kvaløy, J. T., Farmen, R. K., and Heikkilä, R. (2006). Error propagation in relative real-time reverse transcription polymerase chain reaction quantification models: the balance between accuracy and precision. *Analytical biochemistry*, 356(2):182–193.

Pabinger, S., Rödiger, S., Kriegner, A., Vierlinger, K., and Weinhäusel, A. (2014). A survey of tools for the analysis of quantitative pcr (qpcr) data. *Biomolecular Detection and Quantification*, 1(1):23–33.

Perkins, J. R., Dawes, J. M., McMahon, S. B., Bennett, D. L., Orengo, C., and Kohl, M. (2012). Readqpcr and normqpcr: R packages for the reading, quality checking and normalisation of rt-qpcr quantification cycle (cq) data. *BMC genomics*, 13(1):296.

Pfaffl, M. W., Horgan, G. W., and Dempfle, L. (2002). Relative expression software tool (rest©) for group-wise comparison and statistical analysis of relative expression results in real-time pcr. *Nucleic acids research*, 30(9):e36–e36.

Ramakers, C., Ruijter, J. M., Deprez, R. H. L., and Moorman, A. F. (2003). Assumption-free analysis of quantitative real-time polymerase chain reaction (pcr) data. *Neuroscience letters*, 339(1):62–66.

Ritz, C. and Spiess, A. (2008). qpcR: an R package for sigmoidal model selection in quantitative real-time polymerase chain reaction analysis. *Bioinformatics*, 24(13):1549.

Rodriguez, J., Vives, L., Jordà, M., Morales, C., Muñoz, M., Vendrell, E., and Peinado, M. A. (2008). Genome-wide tracking of unmethylated dna alu repeats in normal and cancer cells. *Nucleic acids research*, 36(3):770–784.

Ruijter, J. M., Pfaffl, M. W., Zhao, S., Spiess, A. N., Boggy, G., Blom, J., Rutledge, R. G., Sisti, D., Lievens, A., Preter, K. D., Derveaux, S., Hellemans, J., and Vandesompele, J. (2013). Evaluation of qpcr curve analysis methods for reliable biomarker discovery: Bias, resolution, precision, and implications. *Methods*, 59(1):32–46. Transcriptional Biomarkers.

Schlereth, W., Bassukas, I. D., Deubel, W., Lorenz, R., and Hempel, K. (1998). Use of the recursion formula of the gompertz function for the quantitation of pcr-amplified templates. *International journal of molecular medicine*, 1(2):463–470.

Ståhlberg, A., Åman, P., Ridell, B., Mostad, P., and Kubista, M. (2003). Quantitative real-time pcr method for detection of b-lymphocyte monoclonality by comparison of κ and λ immunoglobulin light chain expression. *Clinical Chemistry*, 49(1):51–59.

Tichopad, A., Dzidic, A., and Pfaffl, M. W. (2002). Improving quantitative real-time rt-pcr reproducibility by boosting primer-linked amplification efficiency. *Biotechnology Letters*, 24(24):2053–2056.

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 biology*, 3(7):1–12.

Wong, M. L. and Medrano, J. F. (2005). Real-time pcr for mrna quantitation. *Biotechniques*, 39(1):75.

4 Supplemental tables

name	sequence (forward, reverse)	genomic location/accession
GAPDH	TTGCAACCGGGAAGGAAA, TAGCCTCGCTCCACCTGACTT	chr12:6535301-6535420
Au60	CTTTGTTGCAACATGGTCCTA, CCTTAACTCAAGCACTGATGA	chr16:31181983-31182328
16cen	CAGTTTGCCAGATAGTCTCTTT, GAGACATTTGGGAAGGTCACTGAAT	chr7:61891350-61891560
THOC3	TTCTTCTGGGGTTGTTCCGTAATC, CACACCCGCTAGCCCTTTTCAT	chr5:175968561-175968678
LUC	CGAACTGTGTGTGAGAGG, TCCAGAATGTAGCCATCCAT	No match
LUC2	TGGACGAAGTACCGAAAGGT, TGGTTTGTCCAAACTCATCAA	No match
GLDC	TGACCTGCGTTACATCTTC, ACGCCCTCTTTTGTTCAGA	NM_000170.2
LNCR	TGCTCTGATCCTGACTAATACT, TGTTCTCACTCCCTCTAG	ENST00000413145.1
CYPA (PP1A)	CTCCTTTGAGCTGTTTGCAG, CACCACATGCTTGCCATCC	NM_021130.4
MRPL19	CAGTTTCTGGGGATTTGCAT, TATTCAGGAAGGGCATCTCG	NM_014763.3
PSMC4	TGTTGGCAAAGGCGGTGGCA, TCTCTTGGTGGCGATGGCAT	NM_153001.2
PUM1	CGGTCGTCCTGAGGATAAAA, CGTACGTGAGGCGTGAGTAA	NM_014676.2

Table S1: **Primer and amplicon details.** The upper panel depicts the primers used for ChIP-QPCR and technique validation (hg38 genomic coordinates). The lower panel summarizes the primers used for expression analysis (spanning intron/exon boundaries; RefSeq or Ensembl accessions).

references	# references	NF stability (M value)	NF size optimality (V value)
h3_gapdh, h3_16cen	2	stable (0.02)	enough (0.01)
+ h3_au60	3	stable (0.03)	enough (0.06)
+ trik4_gapdh	4	stable (0.13)	enough (0.05)
+ Input_1_5_16cen	5	stable (0.18)	enough (0.05)
+ igg_au60	6	stable (0.21)	enough (0.06)
+ Input_1_5_au60	7	stable (0.27)	enough (0.07)
+ Input_1_5_gapdh	8	stable (0.34)	enough (0.08)
+ igg_16cen	9	stable (0.42)	enough (0.08)
+ trik4_16cen	10	stable (0.49)	enough (0.14)
+ igg_gapdh	11	unstable (0.67)	–

Table S2: **References selection for chromatin enrichment.** To quantify K4me3 at the Alu element Au60 after TSA treatment, the Alu Au60, GAPDH and pericentromeric 16cen loci were evaluated for the K4me3, H3 and IgG ChIPs and the input. Data were processed as a TSA-treated vs mock-treated experiment, with au60_trik4 as target 16cen_igg, 16cen_h3, 16cen_input, au60_h3, au60_input, gapdh_igg, gapdh_h3, gapdh_input as putative references (see supplemental methods). Normalization factors (NF) stability is assessed by their M value; M values below 0.5 are tagged as *stable* (*unstable* otherwise). The determination of the optimal number of references is addressed by the NF's V value, being the lowest the optimal. V values below 0.15 indicate that the inclusion of a further reference is not needed and hence tagged as *enough* (and *add* otherwise). Note that the minimal NF is built upon two references which cannot be ranked. The NF with maximum number of references cannot be assessed for size optimality.

contrast	FC	CI	status
trik4_au60_on_NF	8.45	6.56–8.5	up
trik4_au60_on_trik4_gapdh	8.36	6.52–8.37	up
trik4_au60_on_h3_gapdh	6.18	4.5–7.32	up
trik4_au60_on_Input_1_5_gapdh	9.23	9.16–14.07	up
trik4_au60_on_h3_au60	5.59	5.52–6.65	up
trik4_au60_on_Input_1_5_au60	4.3	3.67–5.43	up
trik4_au60_on_trik4_16cen	19.63	11.88–27.66	up
trik4_au60_on_h3_16cen	6.03	4.86–6.8	up
trik4_au60_on_Input_1_5_16cen	8.87	6.6–9.73	up

Table S3: **ChIP enrichment analysis.** Relative quantification of H3K4me3 at the Alu element Au60 after TSA treatment. As the candidate references stability assessment reported an overall suitability (table S2), we only discarded IgG immunoprecipitations due to their unspecificity. The resulting panel of 8 references was found to be stable ($M = 0.48$). The fold-change of K4me3 at Au60 was found to be 8.45 times higher in TSA-treated vs non-treated cells (significantly upregulated, 100 permutations test) when using the multiple normalization factor. The enrichment was also found taking the references separately. NF, normalization factor. FC, median fold change. CI, confidence of interval. up, significantly upregulated.

references	# references	NF stability (M value)	NF size optimality (V value)
psmc4, pum1	2	stable (0.12)	add (0.18)
+ cypa	3	stable (0.4)	enough (0.14)
+ mrpl19	4	unstable (0.52)	–

Table S4: **References selection for differential expression.** All controls and treatments were included in the analysis (mock and two 5-azacytidine treatments at three time points, see supplemental methods).

contrast	FC	CI	status
aza_5_96_vs_control_96_(gldc_on_NF)	1.87	1.49–2.49	up
aza_5_96_vs_control_96_(lncr_on_NF)	2.35	1.32–3.08	up
aza_5_96_vs_control_96_(mrpl19_on_NF)	1.71	1.53–1.8	up
aza_10_96_vs_control_96_(gldc_on_NF)	2.54	2.02–3.73	up
aza_10_96_vs_control_96_(lncr_on_NF)	2.89	1.78–4.25	up
aza_10_96_vs_control_96_(mrpl19_on_NF)	2.09	1.52–3.15	up
aza_5_120_vs_control_120_(gldc_on_NF)	1.7	1.56–2.21	up
aza_5_120_vs_control_120_(lncr_on_NF)	4.4	2.67–17.65	up
aza_5_120_vs_control_120_(mrpl19_on_NF)	0.95	0.73–1.24	ns
aza_10_120_vs_control_120_(gldc_on_NF)	3.94	3.43–4.59	up
aza_10_120_vs_control_120_(lncr_on_NF)	3.57	1.73–12.08	up
aza_10_120_vs_control_120_(mrpl19_on_NF)	0.41	0.32–0.5	down
aza_5_144_vs_control_144_(gldc_on_NF)	4.08	3.58–4.99	up
aza_5_144_vs_control_144_(lncr_on_NF)	2.47	0.99–4.09	up
aza_5_144_vs_control_144_(mrpl19_on_NF)	1.37	0.83–1.8	ns
aza_10_144_vs_control_144_(gldc_on_NF)	6	4.4–306.47	up
aza_10_144_vs_control_144_(lncr_on_NF)	7.26	3.36–164.54	up
aza_10_144_vs_control_144_(mrpl19_on_NF)	1.64	1.26–2.03	up

Table S5: **Differential expression analysis.** GLDC, LNCR and MRPL19 differential expression analysis after being treated with either 5 μM or 10 μM of 5-azacytidine (see supplemental methods). Expression fold change is assessed comparing the aza-treated sample to the mock-treated of the same time point (96, 120 and 144 h). The normalization factor leverages CYPA, PSMC4 and PUM1 references. NF, normalization factor. FC, median fold change. CI, confidence interval. up, upregulated; down, downregulated; ns, non significant.

contrast	FC	CI	status
2fg_vs_20fg_(LUC_on_NF)	0.32	0.39–0.46	down
2fg_vs_20fg_(LUC2_on_NF)	0.16	0.13–0.21	down
200fg_vs_20fg_(LUC_on_NF)	12.74	11.19–14.68	up
200fg_vs_20fg_(LUC2_on_NF)	9.07	7.88–10.04	up
2pg_vs_20fg_(LUC_on_NF)	68.98	65.96–73.48	up
2pg_vs_20fg_(LUC2_on_NF)	78.06	63.27–94.73	up

Table S6: **Experimental resolution validation.**. Fold change of two amplicons, LUC and LUC2, using plasmid templates at 2pg/μL, 200 fg/ul and 2 fg/μL as compared to 20 fg/μL. Expected fold changes are 0.1, 10 and 100 for both LUC and LUC2, respectively. The normalization factor leverages GAPDH, THOC3 and 16cen references as present in a fixed amount genomic spike-in, which are detected to be adequate (M=0.15). NF, normalization factor. FC, median fold change. CI, confidence interval. up, significant positive fold change; down, significant negative fold change.

5 Example files

- Supplemental data. Includes six text files.
  - 1_chip_data_raw_fluo_delim.tsv: ChIP experiment raw fluorescence (structured data).
  - 2_chip_data_unannotated_experimental_design.tsv: ChIP experiment unannotated design template.
  - 3_chip_data_annotated_experimental_design.tsv: ChIP experiment annotated design template to scrutinize H3K4me3 on Alu Au60.
  - 4_expression_data_raw_fluo_delim.tsv: Gene expression experiment raw fluorescence (structured data).
  - 5_expression_data_unannotated_experimental_design.tsv: Gene expression unannotated design template.
  - 6_expression_data_annotated_experimental_design.tsv: Gene expression annotated design template to look for changes in GLDC and LNCR.