Novel Quantitative ChIP-seq Methods Measure Absolute Fold-Change in ER Binding Upon Fulvestrant Treatment

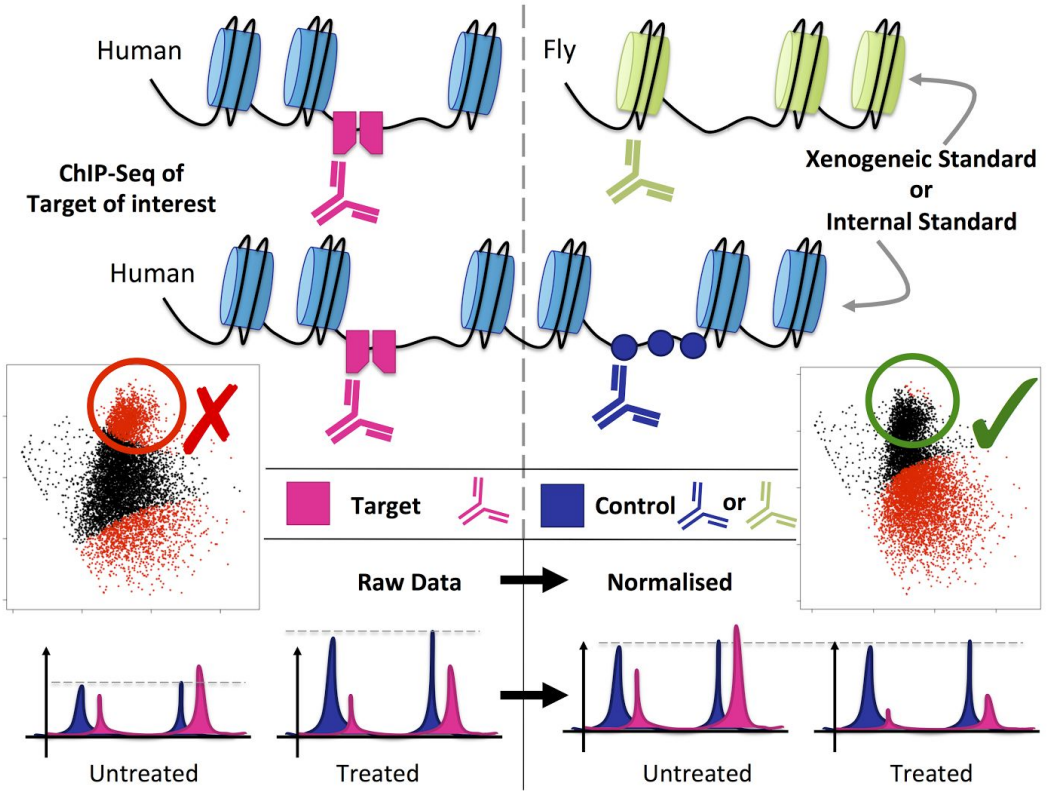
Michael J Guertin[1], Florian Markowetz[2] & Andrew N Holding[2]*

[1] Department of Biochemistry and Molecular Genetics,
School of Medicine, University of Virginia, Charlottesville, Virginia, USA

[2] Cancer Research UK Cambridge Institute,
University of Cambridge, Cambridge, Cambridgeshire, CB2 0RE, UK

* To whom correspondence should be addressed.
Tel: +44 (0) 1223 769500;
Fax: +44 (0) 1223 769510;
Email: andrew.holding@cruk.cam.ac.uk

Graphical Abstract



Abstract

ChIP-seq (Chromatin ImmunoPrecipitation combined with massively parallel DNA Sequencing) maps protein/DNA interactions genome-wide with high resolution.

A key challenge in quantitative ChIP-seq is the normalisation of data in the presence of genome-wide changes in occupancy. Analysis-based normalisation methods were initially developed for transcriptomic data and these methods are dependent on the underlying assumption that total transcription does not change between conditions. For genome-wide changes in transcription factor binding, these assumptions do not hold true. Misapplication of these methods to ChIP-seq data results in the suppression of the biological signal or erroneous measurement of differential occupancy. The challenges in normalisation are confounded by experimental variability introduced during sample preparation and processing. Current experimental methodologies do not fully control for important variables, such as the efficiency of DNA recovery by immunoprecipitation.

We present a novel normalisation strategy utilising an internal standard of unchanged peaks for reference. We compare our approach to normalisation by total read depth and two alternative methods that utilise external controls to study transcription factor binding. We successfully resolve the key challenges in quantitative ChIP-seq analysis and demonstrate its application by monitoring the changes in Estrogen Receptor-alpha (ER) binding upon fulvestrant-mediated degradation of ER, which compromises ER binding genome-wide. Additionally, we developed an adaptable pipeline to normalise and quantify differential transcription factor binding genome-wide and generate metrics for differential binding at individual sites.

Key Results

- ChIP-seq data can be normalised by utilising a second control antibody.
- Our implementation avoids downsampling of the normalised data.
- We developed a customisable pipeline to undertake the analysis, implementing both DESeq2 and DiffBind.
- We provide a robust, normalised dataset that quantifies the effect of fulvestrant on ER binding.

Introduction

ChIP combined with high throughput sequencing (ChIP-seq) quantifies the relative binding intensity for all interactions of a protein to genomic DNA for a single

condition. However, comparing relative intensities of binding between samples and between conditions is an ongoing challenge[1-4]. Conventionally, correcting for sample-to-sample variability occurs at the analysis stage, but these methods assume that experimental variables remain constant between datasets. In practice, different efficiencies in nuclear extraction, DNA shearing, and immunoprecipitation present potential points within a typical ChIP-seq protocol[5] to introduce experimental variation and error. Controlling for experimental variables at the point of data analysis is limited because differential binding is indistinguishable from differential ChIP efficiency and DNA recovery. Therefore, unless a substantial number of known binding events remain constant to provide an internal control, these methods can lead to the suppression of the biological signal[1-3]. Previously, studies have aimed to resolve these challenges when analysing genome-wide changes in histone modifications[1,2]. Similarly, total genomic occupancy of a Transcription Factor (TF) binding can vary drastically between conditions. An extreme case is the binding of nuclear receptors to DNA, which are nearly absent from DNA when unliganded and can be induced by their ligand to bind DNA within minutes. To overcome this limitation, the field has exploited a deficiency in ChIP-seq to approximate normalisation. In short, the vast majority of reads resulting from ChIP-seq are outside of true TF binding sites, so the total read depth is used to approximate normalisation. Nonetheless, these do not control for any of the aforementioned causes of experimental variability which can be exacerbated by changes in the nuclear concentration of the target protein. External spike-in controls are proposed to resolve these challenges in the analysis of histone modifications[1,2], but these methods rely on xenogeneic chromatin (i.e. from a second organism) and the cross-reactivity of the antibody to the factor of interest[1] in both organisms or the use of a second species-specific antibody[2].

Here we present a method, termed *parallel-factor ChIP*, which utilises a second antibody to provide an internal control. The process of utilising a second antibody against the target chromatin avoids the need of a xenogeneic spike-in and controls for more experimental variables than previous methods. We present this method alongside the application of two xenogeneic methods for the analysis of the nuclear receptor Estrogen Receptor-alpha (ER) binding after treatment with fulvestrant. Further, we have developed an adaptable pipeline to apply these strategies and provide a highly reliable quantitative analysis of differential binding sites utilising established statistical software packages.

Accurate analysis of ER binding is of key interest as 70% of all breast cancer tumours are classified at ER+[6]. Fulvestrant is a targeted therapeutic to prevent the growth of these tumours. The mode of action for fulvestrant is to bind to the ER as an antagonist and to recruit a different set of cofactors when compared to the native ligand estra-2-diol. This promotes the degradation of the ER *via* the ubiquitination pathway and the proteasome[7]. The family of compounds to which Fulvestrant belongs is called Selective Estrogen Receptor Degraders or Downregulators

(SERDs). Cellular loss of ER protein results in compromised ER binding genome-wide and presents a model for which to develop novel controls and analysis methods for quantitative ChIP-seq.

Materials and Methods

Cell Culture

All experimental conditions were conducted in the MCF7 (Human, ATCC) cell line. Spike-in standards were generated using HC11 (Mouse, ATCC) and S2 (Drosophila, ATCC) cells. MCF7 were authenticated using STR DNA profiling.

For each individual ChIP pull-down, 4 x 10^7 MCF7 cells were cultured across two 15 cm diameter plates in DMEM (Dulbecco's Modified Eagle's Medium, Glibco) with 10% FBS, Glutamine and Penicillin/Streptomycin (Glibco). Incubators were set to 37°C and to provide a humidified atmosphere with 5% CO_2. The cells were treated with fulvestrant (final concentration 100 nM, Sigma-Aldrich). After 48 hours, the cells were washed with ice cold PBS twice and then fixed by incubating with 15 mL per plate of 1% formaldehyde in unsupplemented clear media for 10 minutes. The reaction was stopped by the addition of 1.5 mL of 2.5 M glycine and the plates were washed twice with ice cold PBS. Cells were released from each plate using a cell lifter and 1 mL of PBS with protease inhibitors (PI) into a 1.5 mL microcentrifuge tube. The cells were centrifuged at 8000 rpm in for 3 minutes at 4 ^{o}C and the supernatant removed. The process was repeated for a second wash in 1 mL PBS+PI and the PBS removed before storing at -80 ^{o}C.

S2 Cells were grown in T175 flask with Schneider's Drosophila Medium + 10% FBS at 27 ^{o}C. Cells were released by agitation and transferred to a 50 mL falcon tube. The cells were then pelleted at 1300 rpm for 3 minutes. The media was removed and the cells resuspended in 7.5 mL PBS. In a fume hood, cells were crosslinked by the addition of 7.5mL 2% formaldehyde in unsupplemented clear media. The reaction was stopped with 3 mL of 1M glycine at 10 minutes. The suspension of cells was then centrifuged at 2000 x g for 5 minutes. The cells were then washed twice with 1.5 mL PBS+PI before the PBS+PI was removed and the cells stored at -80 ^{o}C.

Untreated HC11 were prepared following the same procedure as MCF7.

Chromatin Immunoprecipitation (ChIP)

ChIP was performed as previously reported[5] with the modifications listed below. It should be noted that in both cases, the levels of xenogeneic spike-in chromatin was

used at higher than the minimum necessary for normalisation to facilitate additional analysis in this study (Figure 17). For future experiments, the results of this study should be reviewed in context of the following discussion section to establish a suitable level of spike-in for the experiment undertaken.

For the *D. melanogaster* chromatin spike-in experiment (sequencing data: SLX-8047), *D. melanogaster* and *H. sapiens* samples were prepared separately following the reported protocol until after sonication. At this point, the MCF7 (experimental) chromatin was combined with the S2 derived chromatin (control) in a ratio of 10:1. Magnetic protein A beads were prepared as normal for both the target antibody (100µg, ER, SC-543, lot K0113, Santa Cruz) and the control antibody (10µL, H2Av, 39715, lot 1341001). The washed beads were then combined in a ratio of 1:4 for pulldown. The protocol was then continued as cited.

For the *M. musculus* chromatin spike-in experiment (sequencing data: SLX-12998), *M. musculus* and *H. sapiens* cells were prepared separately following the aforementioned protocol until after sonication. At this point, for each replicate the chromatin from the experimental samples (4×10^7 MCF7 cells) was combined with that from a single plate of HC11 cells (2×10^6 cells). The protocol was continued unmodified using only the ER antibody and protein A beads.

For experiments containing the CTCF antibody control (sequencing data: SLX-14229 & SLX-14438): 100µL magnetic protein G beads were prepared separately for both antibodies, CTCF (10µL, 3418 XP, Cell Signalling) and ER (100µg, SC-543, lot F1716, Santa Cruz). The beads were then combined 1:1 giving 200µL of beads. The only exception was the two CTCF controls (one with and one without Fulvestrant) where no ER beads were added. These were used to generate a CTCF consensus peak set. The protocol was continued as previously reported.

Library Prep

ChIP and input DNA was processed using the Thruplex Library DNA-seq Kit (Rubicon) according to the manufacturer's protocol.

Sequencing

Sequencing was carried out by the CRUK Cambridge Institute using a HiSeq 4000, 50 bp single end reads.

Alignment

Previously Egan *et al.*[2] aligned the reads to the genomes of the two species separately for the generation of correction factors. Their method also required the downsampling to the lowest number of reads. We developed our protocol around the

alignment to a single combined reference genome, either Drosophila-Human (DmHs) or Mouse-Human (MmHs). The reference genomes were generated from Hg19 and Mm9 or Dm3. The alignment of the FASTQ format reads was undertaken with BowTie2 (version 2.3.2). This resolves and simplifies the challenge of ambiguous alignments between the two genomes and does not require the removal of reads. Reads were removed from blacklisted regions using a combined blacklist before peak calling. Normalisation was carried out at the point of differential binding analysis.

Peak Calling

We used MACS2 (version 2.1.1, default parameters) for peak calling in the treatment sample against an input control on the combined genomes. The reads and peaks file were then split into human or spike-in control. For CTCF spike-in, the peaks were split using the CTCF controls to generate a consensus of CTCF peaks. By finding peaks that overlapped with these consensus peaks (bedtools, version 2.25.0), CTCF peak lists for each sample were generated. Likewise, for each sample, an ER peaks list was generated by subtracting the CTCF consensus peaks from the combined list. An example with input data is provided at https://github.com/andrewholding/Brundle_Example/tree/master/preprocessing/ctcfExample.

Bioinformatic Analysis

The bioinformatic analysis was implemented using R (version 3.3.2) with a modified version of DiffBind (version 2.5.6, https://github.com/andrewholding/BrundleDevelopment/raw/master/Diffbind/DiffBind_2.5.6.tar.gz) and DESeq2 (version 1.14.1). Motif analysis was performed using Homer (v4.9).

For full reproducibility of our results, the source code and output of the analysis is available from https://github.com/andrewholding/BrundleDevelopment. Files greater than 100Mb are not included because of the GitHub file size limit . All sequence data utilised for this study is available from the Gene Expression Omnibus (GSE102882).

Summaries of the analysis applied in this paper can be found at https://github.com/andrewholding/BrundleDevelopment/tree/master/examples with read counts provided as either a DiffBind object in Rdata format or as CSV. Example 001 provides documentation on the normalisation of ER ChIP-seq data using the binding of CTCF as an internal control. Example 002 provides documentation on the normalisation of ER ChIP-seq data using only DESeq2 size factors and no-linear fit. Example 003 provides documentation on the normalisation of ER ChIP-seq data using *D. melanogaster* derived chromatin and H2Av specific antibody.

Pipeline and R packages

A R package containing the functions used for the analysis can be installed from https://github.com/andrewholding/Brundle using the *install_github* found in the *Devtools* package.

A R package containing two sets (one internal and one spike-in control) of test data provided as aligned reads, peak files and samples sheets can be installed from https://github.com/andrewholding/BrundleData.

Two complete documented examples of how to undertake analysis of data using these two packages can be found at https://github.com/andrewholding/Brundle_Example/

The complete set of scripts for the preprocessing pipeline, including generation of, indexing and the alignment of reads to the merged genome, is provided to support the implmentation of future analysis with Brundle at https://github.com/andrewholding/Brundle_Example/tree/master/preprocessing/dmExample. This data set includes example input data in FastQ format so the complete preprocessng and analysis pipeline can be tested before use.

All the contents of the Brundle_Example repository are also packaged in a Docker container for easy use. Instructions on downloading and running the container are available at https://github.com/andrewholding/Brundle_Example/#using-docker-container.

Results

Commonly applied analytical normalisation methods highlight the need for experimental quantitative ChIP-seq controls

Four data-based normalisation strategies, raw counts, RPM reads in peaks, RPM total reads, and RPM aligned reads, are commonly used to normalise ChIP-seq binding between conditions. We apply these methods to each of the datasets we generated to highlight their deficiencies. Although the experimental controls had reads associated with xenogeneic DNA and non-ER peaks, we exclusively analysed reads that align to the *H. sapiens* genome or, in the case of the parallel-factor ChIP, using data only from ER binding sites. Therefore, these analyses are representative of a normal, uncontrolled, ChIP-seq experiment. All datasets gave equivalent results

and conclusions for each of the normalisation strategies, showing a strong decrease in ER binding upon addition of fulvestrant. Therefore, only the analysis of the human aligned reads from our xenogeneic spike-in and IP of ER using the same ER antibody for control and sample chromatin (SLX-12998) is discussed below.

We first plotted the average ER peak intensity, as determined by raw counts and three counts-based normalisation methods, by the change in ER intensity upon fulvestrant treatment (Figure 1). These MA plots showed similar results, with the main difference being a dampened decrease in ER binding when only reads within peaks are used to normalise. When MA plots are properly normalised, the unchanged peaks between conditions should be distributed with a log fold difference around zero, with increasing variance as the peak intensity decreases. However, the distribution of data points in the raw counts MA plot show that this distribution is shifted up to a y value of ~1 (Figure 1A). We hypothesised that these are true ER binding sites that do not change upon fulvestrant treatment or false-positive peaks. In both cases, the apparent increase in binding is an artefact of the data processing. As expected, the apparent fold-change for the increase in ER binding was most pronounced when the data was normalised with respect to total number of reads in peaks (Figure 1B) because this method is reliant on the majority of binding events between the two experimental conditions remaining constant. For our system, that assumption is incorrect as the majority of ER binding events are substantially decreased. Other common normalisation methods that have been applied to ChIP-seq data, such as quantile normalisation[8,9], would result in a similar systematic error in the final data. Nonetheless, other more appropriate methods – e.g. reads per million (RPM) total reads, which correct for total library size – showed little improvement for our datasets over the raw number of reads counts in peaks. The same result was seen if RPM peak counts were calculated using either aligned (Figure 1C) or total read counts (Figure 1D). All three strategies for all three datasets displayed a similar grouping of points above the axis, incorrectly implying an increase in ER binding to the chromatin at these sites 48 hours after treatment.

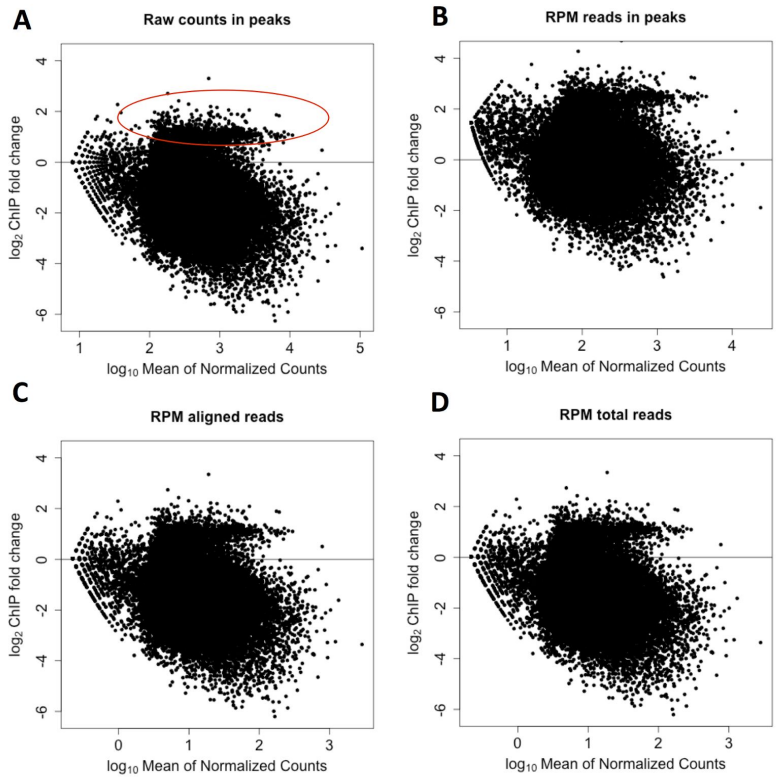


Figure 1 - Comparison of simple normalisation strategies employed. *MA plots showing the changes in ER binding after 48 hours treatment with 100 nM Fulvestrant. Three simple normalisation methods were applied to this data (A) Raw Counts (B) Reads Per Million (RPM) reads in Peaks. (C) RPM aligned reads (D) RPM total reads. Note that the highlighted peaks remain above zero under all three standard normalisations.*

Internal and spike-in normalisation controls

Normalisation using *D. melanogaster* chromatin and species specific antibody for H2Av

To overcome the challenges of normalising ChIP-seq data, *Egan et al.*[2] combined the extract with xenogeneic chromatin and a second antibody that is specific to the spike-in organism's chromatin. This controls for the efficiency of the immunoprecipitation, if the same ratio of target to control chromatin is achieved between samples. This work reported that a reduction in H3K27me3 in response to inhibition of the EZH2 methyltransferase cannot be detected by standard normalisation techniques. Instead, the study was able to demonstrate genomic H3K27me3 reduction by including *D. melanogaster* (Dm) derived chromatin and a Dm-specific histone variant H2Av antibody as spike-in control for normalisation. However, this method fails to control for variation in sonication, which could alter fragment length distributions, or errors in quantifying chromatin concentration.

The challenge in analysing the genome-wide reduction in H3K27 methylation by ChIP-seq shares many similarities to quantifying changes in ER binding after fulvestrant treatment. In particular, both result in a global unidirectional change in chromatin occupancy due to the specific loss of the target molecule, which is not compatible with the underlying assumptions of purely analytical normalisation approaches.

We applied this method of normalisation to fulvestrant-depleted ER samples using xenogeneic *D. melanogaster* chromatin and an H2Av antibody. We aligned our reads to a combined *H. sapiens/D. Melanogaster* (HsDm) genome. We separated peaks by their genomic identity, resulting in both Hs peaks (experimental) and Dm peaks (control). Figure 2A shows a similar distribution to Figure 1C, including the off-centre putative unchanged ER binding events (Figure 2A, within red triangle) as highlighted in Figure 1A. Overlaying the peak information from the *D. Melanogaster* peaks indicated that they overlapped along the same y-axis value (Figure 2B) as the ER binding events (Figure 2A) that were concluded previously as being unchanged, or possibly false positive peaks. We then applied a linear fit to Dm $\log_2$(fold-change) values for each binding site. The coefficients generated from the linear regression were then used to adjust the $\log_2$(fold-change) of all data points (Figure 2C). The normalisation of the data resulted in a reduced number of increased ER binding events at 48 hours. The remaining loci of increased binding resulted from the higher variation at lower intensities.

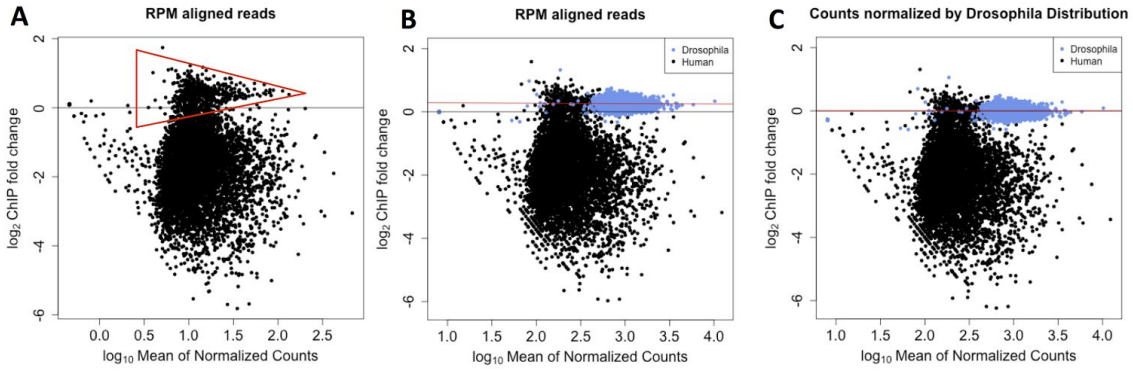


Figure 2 MA plots showing ER binding before and after treatment with fulvestrant including matched Dm H2av spike-in control. *(A) Reads corrected to total aligned reads showed the same off-centre peak density as observed in Figure 1. Putative unchanged ER binding sites are within the red triangle. (B) Overlaying the MA plot combining the changes in chromatin binding of Hs ER (black) and Dm H2Av (blue). Dm peaks overlay the off-centre peak density. (C) Utilising the Dm H2Av binding events as a ground truth for 0-fold change, a linear fit to the log-fold change is generated and the fit is applied to adjust the Hs ER binding events.*

Normalisation utilising the cross-reactivity of an ER antibody and mouse chromatin as a spike-in control

A challenge with the use of spike-in Dm chromatin to provide a standard for Hs ChIP-seq experiments is that tight control is required on both the amounts of antibody and the amount of chromatin used. This is complicated as the amount of Hs chromatin may not be constant between experimental conditions. In an attempt to reduce the number of variables that can result in experimental error, we developed a similar method to that of *Bonhoure et al*[3]. Their study utilised the cross-reactivity of a Pol II antibody against Hs control chromatin and sample chromatin from Mus musculus (Mm). The ER antibody utilised in this study is known to have cross-reactivity between the Hs and Mm homologues of the ER. We therefore expected that the inclusion of Mm chromatin would provide a series of control data points that would not alter between conditions. To visualise and provide a preliminary analysis of the data, we applied the same strategy as we developed for our Dm spike-in experiment.

We found that Mm genomic ER peaks were greatly increased after treatment with Fulvestrant (Figure 3A).

We compared the level of Hs and Mm reads between samples and found the ratio consistent between samples (Figure 4). Therefore, poor sample balancing could not cause the results presented in Figure 3.

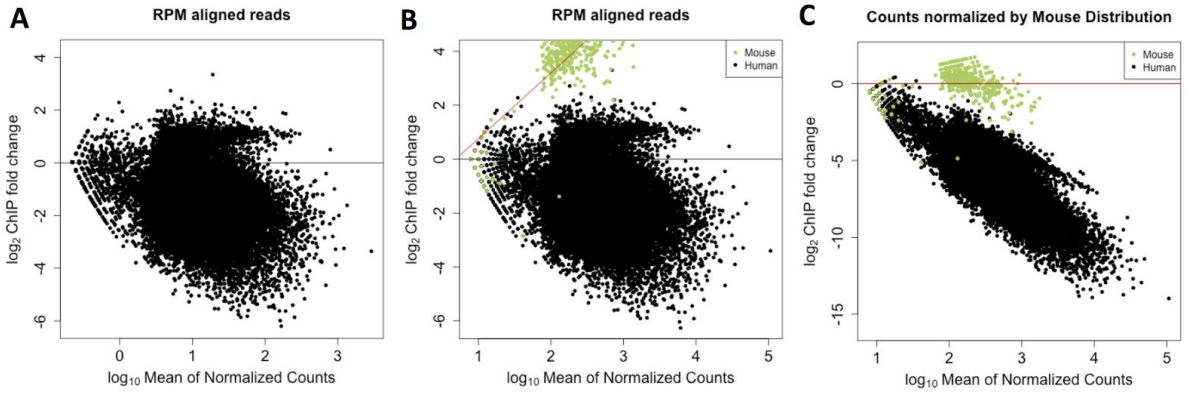


Figure 3 MA plots showing the addition of Mm derived chromatin spike-in to the ChIP-seq analysis of MCF7 before and after treatment with fulvestrant. *(A) MA plot after scaling factor based normalisation shows same characteristic grouping of peaks off axis. (B) ER binding in Mm samples shows considerable increase in binding after treatment of the MCF7 cell line with fulvestrant. (C) Attempting to fit a correction factor to the data results in a significant distortion.*

These results highlight a problem with using a constant antibody and a xenogeneic source of chromatin for normalisation. We propose that the ER antibody has lower affinity for mouse ER, compared to human ER. Therefore, we conclude that the increase in Mm reads from ER binding sites results from a reduction in competition with human ER for the same antibody, because fulvestrant is degrading human ER. These challenges are likely to be less of a concern when applying this method to a more conserved target and this explains why there has been previous success in applying this strategy to the analysis of histones[2] and RNA Polymerase[3].

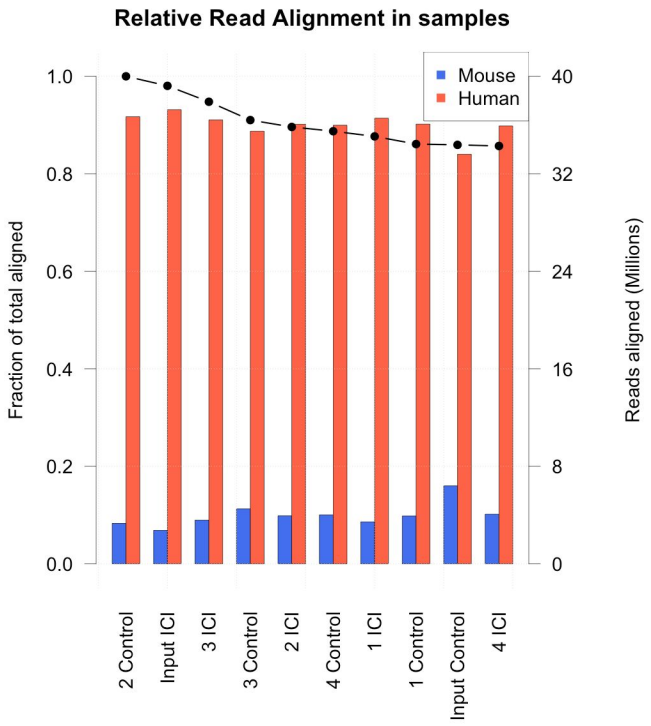


Figure 4 Distribution of reads for Mm chromatin spike-in normalisation strategy. Comparison of mouse chromatin between samples showed no systematic bias in the sample preparation. Bar plots (left axis) represent the fraction of total aligned reads. The dot plot represents the total aligned reads (right axis) for each sample.

Normalisation using a second control antibody to provide an internal control

A key reason for utilising the cross-reactivity of antibodies between organisms to provide a normalisation standard was the potential to reduce the number of sources for experimental variation. For the same reason, we developed the use of a second antibody as an experimental control, *parallel-factor ChIP*, to normalise signal. The advantages of using a second antibody over a spike-in control is that the target:control antibody ratio can be maintained for all samples by producing a single

stock solution. For concurrent experiments, a single stock of antibody bound beads can be prepared and used for all samples with minimal variation. It is critical to identify a DNA-binding protein for the control whose genomic distribution and intensities are not affected by the treatment. For the analysis of ER binding, we chose CTCF as our control antibody. While CTCF is affected by compounds that target ER, the effects of these changes have been documented at only a small fraction of the total number of sites[10], a result that was subsequently replicated in our own analysis (Figure 5 and Figure 10b).

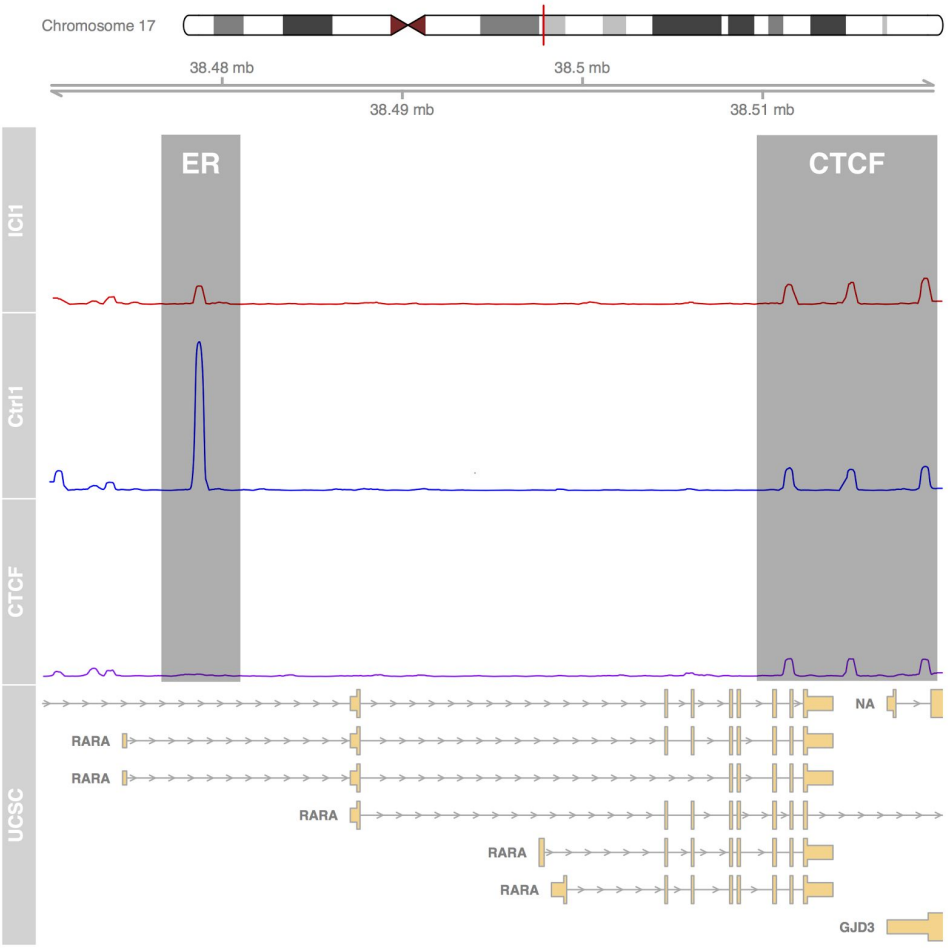


Figure 5. CTCF peak height remains constant while ER peaks change upon treatment with fulvestrant. As the binding of CTCF at the 3 control peaks (right) will remain constant in all 3 conditions (see also Figure 7B), the data is scaled to CTCF peak height. After 100 nM fulvestrant treatment for 48 hours, ER binding (left) shows a reduction in binding at the RARA gene (red) when compared to control (blue). The CTCF peaks can be confirmed against a CTCF only ChIP-seq experiment (purple).

To confirm that the effects seen in Figure 5 were consistent across the genome, we compared the clustering of the CTCF and the ER peaks with respect to the treatment with Fulvestrant. Initial clustering was weakly correlated with that of the treatment condition (Figure 6A). Clustering specifically to CTCF derived peak data (Figure 6B) resulted in a loss of grouping by treatment, while clustering specifically ER-derived peak data (Figure 6C) led to a clearer separation by treatment.

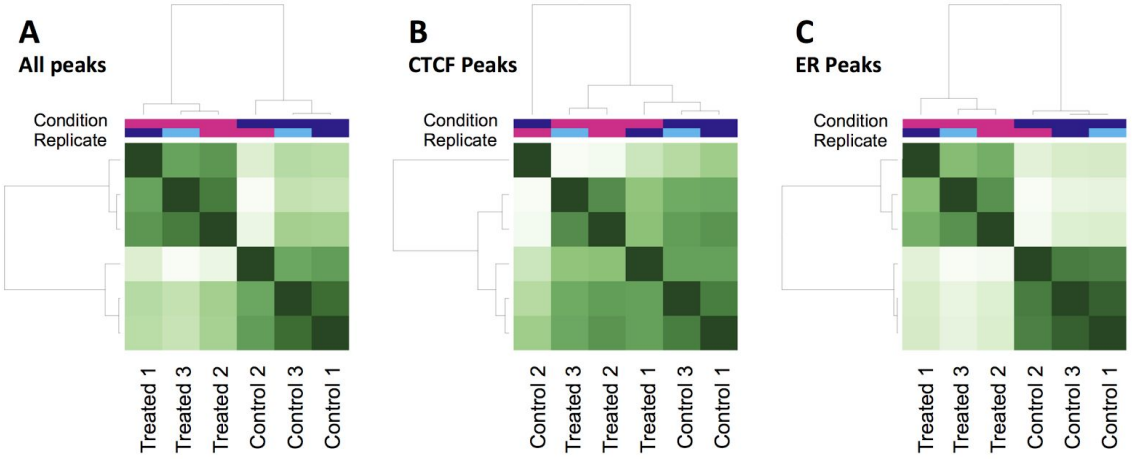


Figure 6. Cluster of samples before and after ER and CTCF peak extractions shows the effect of fulvestrant on ER peaks drive clustering of the raw data. *(A) Clustering of the ChIP-seq peak sets before peak extraction shows that the fulvestrant treatment is significant enough to result in sample grouping. (B) Extraction of only the CTCF alters the clusters and the heatmap shows no strong pattern. (C) Extraction of only the ER peaks greatly increases the clustering by treatment demonstrating a clear separation of signal.*

Having separated the ER and CTCF binding events, we plotted them separately on an MA plot (Figure 7A and B). As previously shown for Dm spike-in control, we applied a global fit to the $\log_2$-fold change between the two conditions correcting the bias in fold-change between conditions in ER binding (Figure 7C). Taken together, we show that performing a parallel ChIP-seq experiment with an unrelated and unchanged factor is an alternative and complementary method to account for extreme genomic changes in factor occupancy.

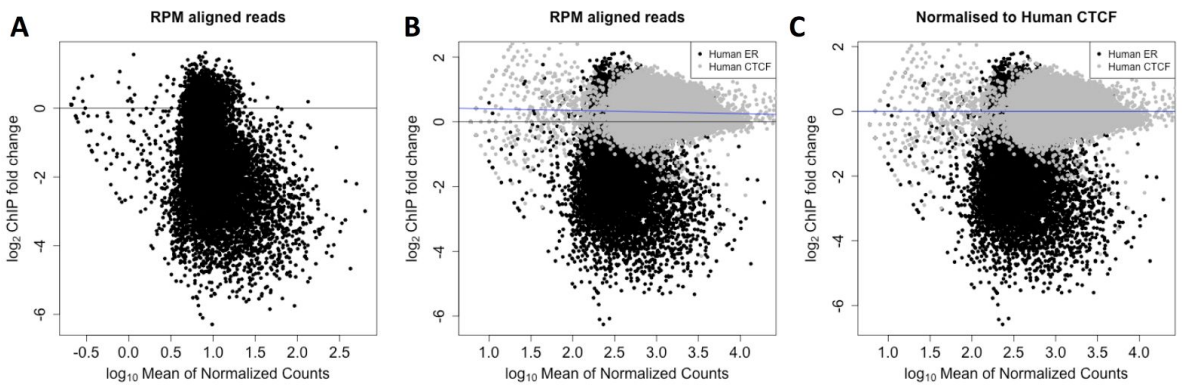


Figure 7. MA plots showing ER binding before and after treatment with fulvestrant including matched CTCF control. *(A) Reads corrected to total aligned reads showed the same off-centre peak density as observed with all that was not-normalised with an internal spike-in control. (B) Overlaying the MA plot combining the changes in chromatin binding of ER (black) and CTCF (grey). CTCF peaks overlay the off-centre peak density. (C) Utilising the CTCF binding events as a ground truth for 0-fold change, a linear fit to the log-fold change is generated (blue line). The fit is then also applied to the ER binding events.*

Pipeline and Quantitative Analysis

Normalisation implementation using DESeq2 and Size Factors

DESeq2 was initially developed for the analysis of RNA-seq data[11]. It provides a method to establish significant differential gene expression between two samples by modelling the counts with a negative binomial distribution. Given the similarities in ChIP-seq and RNA-seq, primarily that they are both based on the same high throughput sequencing technologies, DESeq2 has been successfully adapted to ChIP-seq analysis to establish differential intensity analysis of histone modifications.

We successfully integrated our datasets with DESeq2 by providing two matrices: one for the reads counts from peaks detected in our experimental sample, and a second for our control. The initial analysis using DESeq2 *estimateSizeFactors()* on the Hs aligned reads showed a large distortion of the signal (Figure 8A). The distortion is expected as the default analysis of DESeq2 is designed for an RNA-seq library where total transcription is assumed to not change between conditions and ~100% of counts are signal (in contrast, the ChIP-seq signal is often contributed by less than 5% of reads). Since the H2Av peaks will not change intensity, we used the read counts in these H2Av control peaks as a size factors parameter estimate for the experimental data. As expected, the spike-in control size factors corrected the majority of the bias in the sample analysis (Figure 8B). Overlaying the ER treatment

analysis onto the H2Av Drosophila control showed that the counts in H2Av peaks showed minimal changes. To aid with the analysis, we have provided the series of functions to count peaks and normalise using DESeq2 as an R package (https://github.com/andrewholding/Brundle).

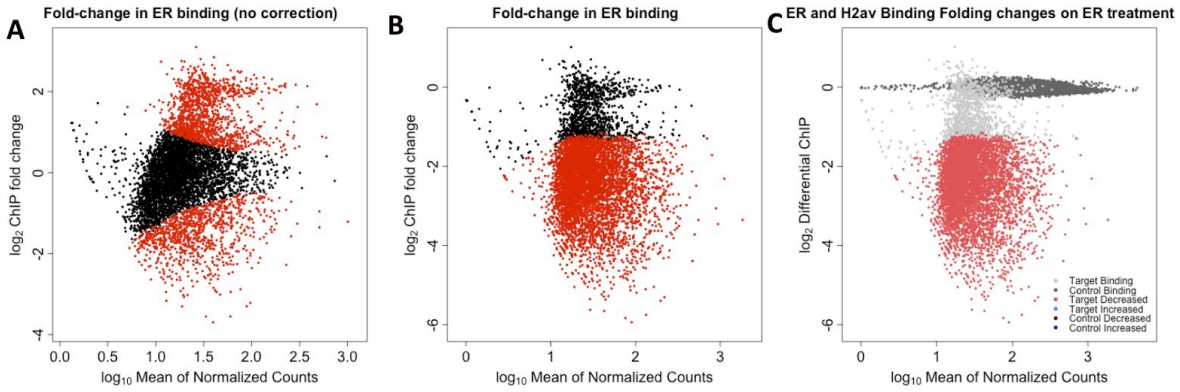


Figure 8 Normalisation of ER binding external spike implemented using DESeq2. *Highlighted data points are considered significant fold-changes with a FDR = 0.01. (A) Initial analysis of the ER binding with default parameters shows an equal increase and decrease in ER binding. The distribution seen is not reflective of the documented response of ER on treatment of fulvestrant. (B) Estimating the DESeq2 size factors from the sample spike-in corrects the distortion in the results. (C) Overlaying the two datasets shows that the read counts in H2Av peaks from the D. melanogaster spike-in (dark grey) remain constant.*

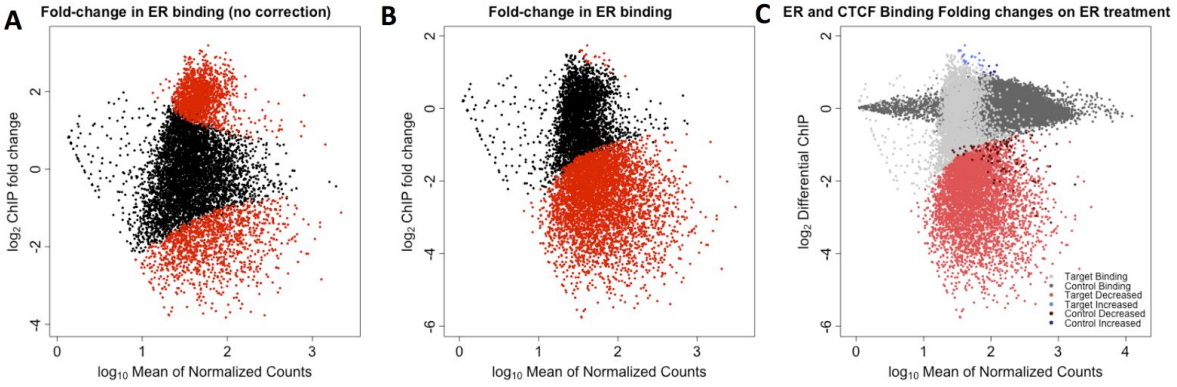


Figure 9 Normalisation of ER binding internal CTCF control. *Highlighted data points are considered significant fold-changes with a FDR = 0.01. (A) Initial analysis with default DESeq2 parameters gives similar distortion as seen in Figure 8A. (B) Correction using the CTCF peaks to provide an internal control allows for the data to be corrected. (C) Overlaying the ER and CTCF datapoints shows that, similar to the H2Av/Drosophila spike-in, CTCF (dark grey) remains mostly constant. Only a small percentage of CTCF sites are differentially bound (<0.2%) as previously noted in the literature.*

We processed the CTCF internal control data in the same manner, using the counts with CTCF peaks to adjust the size factors parameter. The analysis of the raw data displayed the same initial bias as seen with the Dm/H2Av spike in data (Figure 9A). We normalised the data using the counts within CTCF peaks to estimate the DESeq2 size factors (Figure 9B). Overlaying the control and target peaks (Figure 9C) showed that of the CTCF binding events, only a low percentage of sites change between samples, which is consistent with previous reports in the literature[10].

H2Av and CTCF provide a set of unchanged reference peaks for normalisation

For CTCF to work as an internal control, the binding of the factor must show minimal changes between the two conditions. Using the CTCF locations established from two CTCF replicates that did not include any ER antibody, it is possible to implement a differential binding analysis of only the control regions, as we did for H2Av binding for the Dm ChIP control. Comparison of the two control datasets (Figure 10) displayed a lower variance and a lower maximum fold-change for H2Av compared to the CTCF control binding regions. In contrast, the CTCF dataset provides a much greater number of data points for normalisation as a result of relative size of the human and Drosophila genomes. None of the H2Av sites in the Drosophila genome showed a significant change in occupancy. Of the CTCF peaks, 0.01% were considered significantly upregulated, and 0.18% downregulated at an FDR of 0.01 (Figure 10B). We confirmed the differential bound CTCF sites were consistent with the previously characterised proximal binding of ER and CTCF[10] by motif analysis. Sequences of the differentially bound CTCF sites gave strong enrichment for known ERE and CTCF binding motifs (Figure 10C), both with q-values much smaller than 0.0001. Inspection of these sites (Figure 10D) confirmed proximal binding at these locations. These only account for a small percentage (<0.2%) of the CTCF binding locations, therefore CTCF sites as a whole meet the assumption that the average fold-change in CTCF binding between the two conditions equals 1 (i.e. an average log-fold change of 0 in the MA plots). As long as the vast majority of control peaks do not change, they can be used to normalise the peaks that change with the experimental conditions.

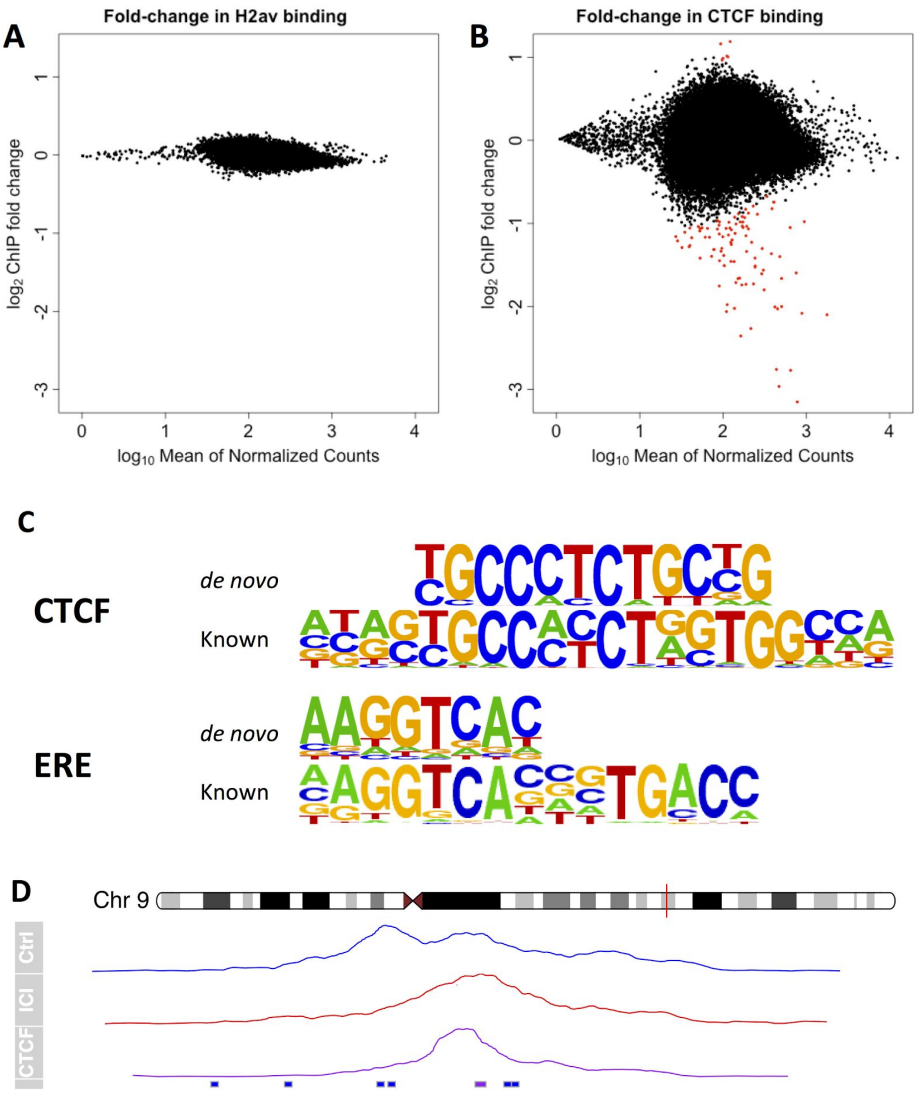


Figure 10 Comparison of the control regions used to normalise ER analysis before and after treatment. Dots highlighted in red are significant (FDR = 0.01). *(A) H2Av occupancy of the Drosophila genome shows no significant changes before and after treatment. (B) The majority of CTCF peaks show no significant change in the number reads (>99.8% of sites) before and after treatment. (C) Motif Analysis of differentially bound CTCF loci. Motifs matching binding sites for CTCF and ER (ERE) were found by* de novo *analysis. (D) Example site of ER and CTCF proximal binding site (Chr9:97,545,000-97,546,000). Due to the DNA fragment length, the peaks overlap. Top/Ctrl is the combined ER & CTCF ChIP before treatment. Middle/ICI is the combined ER & CTCF ChIP after fulvestrant treatment. Bottom/CTCF is CTCF ChIP only. Blue/Purple blocks under tracks represent the presence of ER/CTCF motifs respectively.*

Integration with DiffBind using Corrected Size Factors

DiffBind[12] is an established R package to provide a pipeline to quantitatively measure differential binding from ChIP-seq data. Similar to our strategy, DiffBind is

implemented using RNA-seq analysis tools, including DESeq2, but has many advantages that make it convenient for ChIP-seq analysis. A key feature of DiffBind is that, to calculate size factors, it utilises the total library size from the sequence data provided in a sample sheet (e.g. BAM files) rather than the *estimateSizeFactors* function provided by DESeq2. Nonetheless, while improved, the analysis of the raw data by DiffBind is incomplete with the putative unchanging peaks showing a greater than 0 log-fold change (Figure 11).

Utilising a modified version of DiffBind package (available from https://github.com/andrewholding/BrundleDevelopment/raw/master/Diffbind/DiffBind_2.5.6.tar.gz), we were able to extract the count matrix, apply the size factors calculated by our DESeq2 pipeline, and then reinsert these into the DiffBind object. To prevent DiffBind from applying further normalisation data, the relative read counts in the DiffBind object were set to 1. The final output was similar to that achieved by our implementation (Figure 11B).

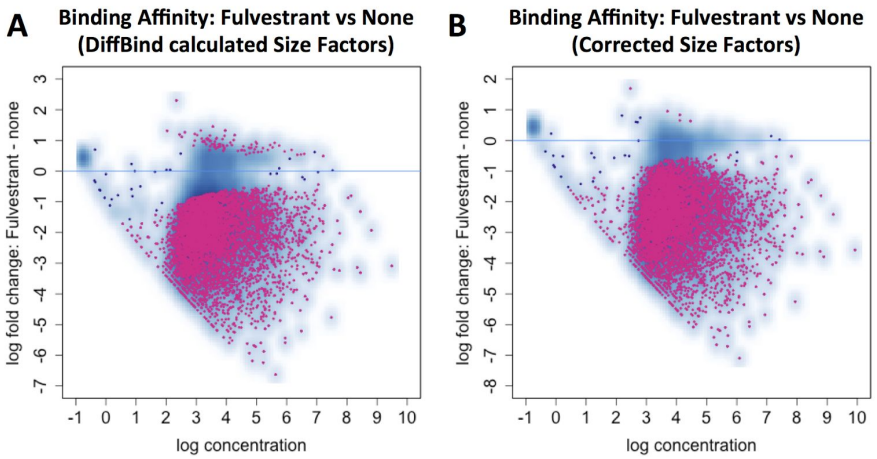


Figure 11 Comparison of DiffBind output before and after applying the corrected size factors from our pipeline generated from Drosophila spike-in control. *(A) Analysis of ER binding before and after treatment with fulvestrant demonstrates that DiffBind's default normalisation strategy is more effective than the DESeq2 default (Fig 8A), but demonstrates a bias between samples. (B) Applying the correct size factors from our DESeq2 pipeline reduces the bias in the analysis (Data: SLX-8047).*

Establishing a normalisation coefficient by linear regression of control peak counts

In our implementation, DESeq2 generates the size factor estimates through the summation of all reads within the peaks. Inherently, a summation has a bias to the peaks with the largest read count. We therefore hypothesised that the normalisation process could be improved by calculating the sample bias through the application of

linear regression. We plot the read count in each CTCF peak of one condition against the other (Figure 12) and then apply a linear model to the data. The advantage over summation is that this step gives equal weight to each point. Any systematic bias will be represented by the gradient of resultant fit deviating from a value of 1. Our normalisation coefficient is defined as the constant by which we need to scale the count data for each CTCF peak from the treated samples to correct this systematic bias (and thereby setting the gradient of the linear fit equal to 1). This normalisation coefficient is then applied in the same manner to ER count data. The correct data is then re-inserted into the DiffBind object for analysis. The only caveat to this method is because DiffBind is programmed to correct by total library size, our pipeline then includes a second factor, named the correction factor in our code, to ensure that part of the normalisation is not applied twice.

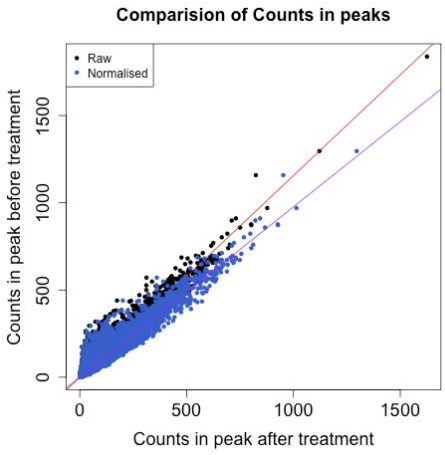


Figure 12 Comparison of Mean Counts in CTCF peaks before and after treatment. *If the samples have no systematic bias before and after treatment then the linear fit would be expected to have a gradient of 1. Here, we establish that the gradient is < 1, implying a systematic bias between samples. The read counts in the treated samples peaks are corrected (blue), removing the bias, and resulting in a new gradient of 1.*

Comparison of size factors and linear modelling to establish normalisation

To compare methods, we applied normalisation by total library size, normalisation using size factors generated by DESeq2 from the CTCF binding events, and then used linear regression to establish the normalisation coefficient, defined and generated as described above. These methods were all applied to the same initial dataset (Figure 13). The use of the linear regression based method was found to be most successful (Figure 13C) showing a 10.2% increase in the number of sites found as differentially bound (FDR < 0.05) compared to normalisation by library size alone (Figure 13A). Our alternative method based on DESeq2 size factors (Figure 13B)

gave an increase of 1.4% when compared to normalisation by library size alone (Figure 13A).

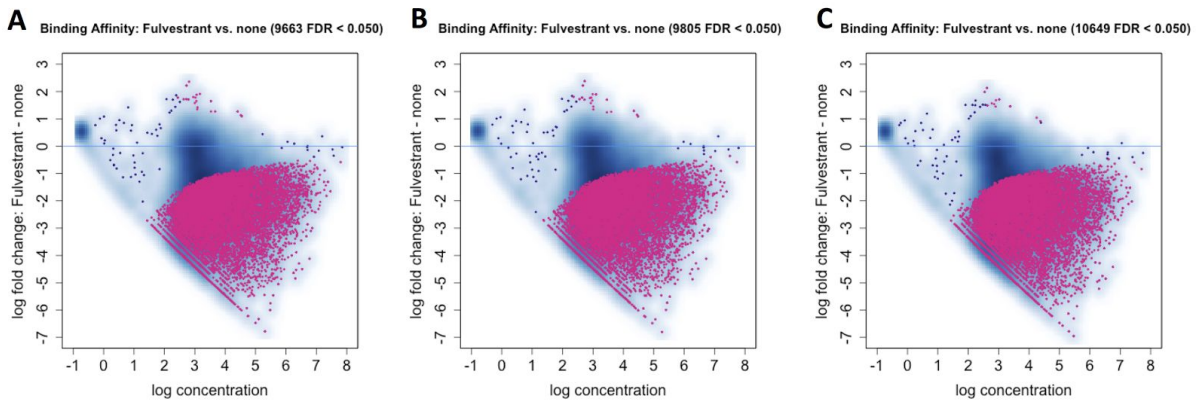


Figure 13 Comparison of DiffBind results before and after our two methods of normalisation. (*A) Normalisation to Library Size, (B) Applying the corrected size factors from our DESeq2 pipeline generated from CTCF internal control. (C) Applying correction using linear regression of CTCF peaks between conditions to normalise data. The result is a 10.2% increase in the number of loci detected as significantly changed ER binding.*

Comparison of absolute fold-change from parallel-factor ChIP and xenogeneic spike-in

In principle, if both the internal control using CTCF binding events and the use of the spike-in Dm/H2Av control are accurate, the normalised fold-change for each genomic loci should be equal. We tested this by matching peaks between the two experiments (taking the nearest peak by genomic location) and plotting the fold-change of one normalised experimental result against the other (Figure 14). As predicted, the datasets gave a result near parity (linear fit of gradient = 0.96) and strong statistical evidence from the correlation (Pearson correlation tending to 0).

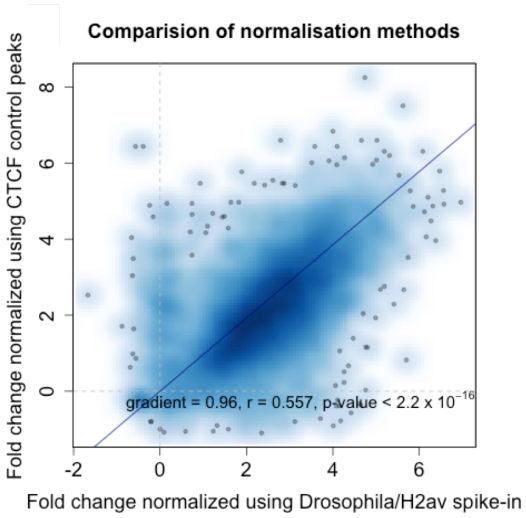


Figure 14 *Comparison of fold-change values for ER binding sites between two datasets. Y-axis is after internal normalisation to CTCF binding sites, x-axis is the use of H2Av binding on spike-in D. melanogaster chromatin. Binding sites were equated between datasets by taking the nearest peak. Ratio of fold-changes is approximately 1 and the datasets correlate with a high degree of confidence.*

Distribution of high-intensity low-fold change peaks

A subset of high-intensity low-fold change peaks, i.e. those at the narrow end of the triangle in Figure 2A, were absent in the MA plots of samples generated with the parallel pulldown of CTCF and ER (Figure 2A and 7A). To address potential concerns arising from the differences in these putative unchanged ER binding sites, we repeated the above comparison using a consensus set of 10,000 high-confidence ER binding sites (as established by MACS2[13]) in place of the nearest neighbour algorithm previously employed. On overlaying the data analysed with the consensus peak set, we established the same high-intensity low-fold change sites in both datasets (Figure 16A). Repeating our comparison of locus-specific fold-change data between the two normalisation methods, but this time using the consensus peak set, supported our previous analysis (Figure 16B). As before, we found near parity between the methods (linear fit of gradient = 0.94) and a correlation of r = 0.77, with a similar p-value (tending to 0).

On further analysis, the genomic locations of these high-intensity low-fold change peaks were found to have been removed as CTCF binding sites in the filtering step of the internal normalisation preprocessing. Nonetheless, their inclusion in the consensus data showed good correlation between the two methods, implying that the CTCF binding event at these locations did not interfere with analysis of the ER sites.

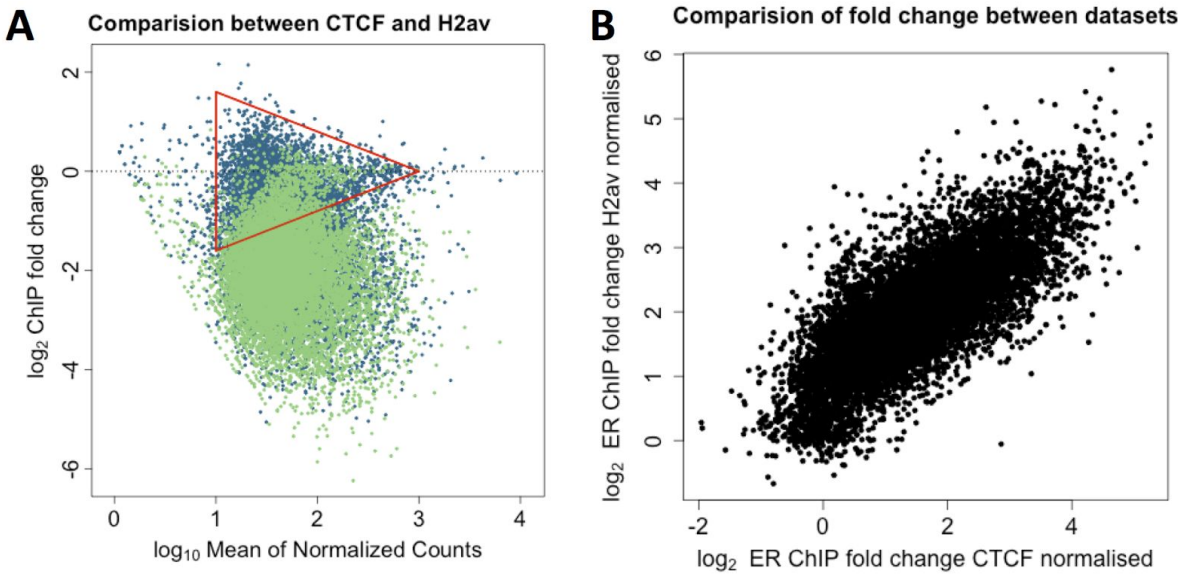


Figure 16 Comparison of normalisation methods using consensus peak set. *(A) The analysis for the CTCF normalised (blue) and H2Av normalised (green) dataset using an ER consensus peak set of 10,000 peaks were formatted as an MA plot and overlaid. This recovered the low-fold change higher-intensity peaks that were not visible in Figure 7A and both datasets showed a similar distribution. (B) Comparison of fold-change values for individual ER binding sites between two datasets showed that the inclusion of these sites did not appear to affect the correlation* (r = 0.77)*.*

Simulating the level of xenogeneic chromatin spike-in for normalisation

To establish the most efficient normalisation strategy, spike-in samples were prepared with higher levels of Dm chromatin than previously described[2]. This provided an opportunity to computationally titrate the number of reads against the Hs chromatin.

To systematically compute different levels of spike-in reads, we calculated the percentage of total number of reads aligned to the Dm genome across all samples and downsampled Dm reads equally across all samples. All MA plots showed considerable improvement (Figures 17A, B and C, top) over correction based on total number of reads (Figure 7A). We found that, while the information used to generate the normalisation coefficient was significantly reduced in quality from 1% (Figure 17A, bottom) compared to 5% (Figure 17B, bottom), the final corrected data was similar. A caveat of using only a very low level of reads (e.g. 1%) from the Dm chromatin is that peak calling will not be successful; however, in these cases, the peak locations established in our study, or other studies, can be used in place of peak calling.

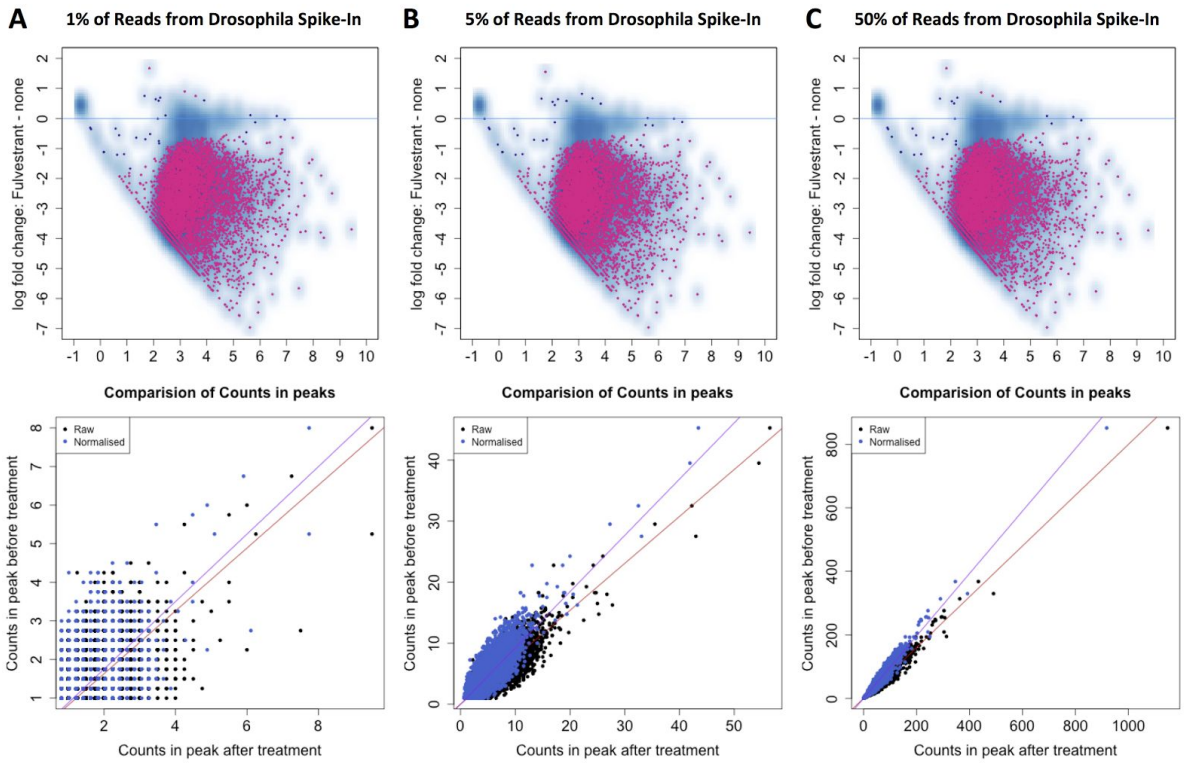


Figure 17 Comparison of normalisation efficiency at 1%, 5% and 50% control sample reads. *Control reads from D. melanogaster chromatin were calculated as a fraction across all samples. Reads were then randomly removed from each sample to give a simulated spike-in of (A) 1%, (B) 5% and (C) 50% reads for normalisation. The overall effect of reducing the level of spike-in on normalisation appears minimal down to 1%.*

Discussion

We have described and benchmarked a reproducible quantitative normalisation strategy using internal ChIP-seq controls. We applied this technique to normalise TF binding and applied statistical analysis at the level of individual binding sites, which was lacking from previous spike-in methodologies. We demonstrate that a parallel-factor control antibody is a reliable alternative to the use of cross-reactivity[1] against a second species or a second species-specific antibody[2] as a method of normalising ChIP-seq data. Further, we have validated both methods in the context to ER binding, and more broadly supported their application to transcription factors in general.

Our method showed considerably better performance than utilising a xenogeneic spike-in combined with cross reactivity of the target antibody in the context of ER binding. Utilising the cross-reactivity of the antibody displayed significant distortion within our data set. This method resulted in skewed murine-ER peak intensities due

to variable competition between the ER from the respective species upon ER-degradation. Intriguingly, a similar challenge was previously reported when normalising the changes of H3K27me3 occupancy when originally applying the Drosophila and H2Av spike-in control.[2]

Using an internal parallel-factor control showed comparable performance to using a second antibody and xenogeneic chromatin as a spike-in control, but there are many advantages to using a second antibody (CTCF) that IPs a protein within the same extract. Primarily, the parallel-factor ChIP controls for the greatest number of steps in the process and gives fewer opportunities for variation being introduced into the sample preparation. In particular, stocks of the antibody mixture can be much more tightly controlled between experiments than xenogeneic chromatin spike-ins. Stocks can be kept over periods of time between experiments. In contrast, the addition of xenogeneic chromatin relies on the precision that the concentration of the chromatin of both the experimental samples and the spike-in can be established reliably and must be added to each sample individually. As chromatin is routinely crosslinked for ChIP-seq, the resultant mixture of protein and DNA makes accurate quantification of DNA challenging without purification, which presents another challenge for the use of xenogeneic spike-in methods. The potential concerns for parallel-factor normalisation is that it may lead to the masking of a subset of sites (Figure 2A and 7A); however, these are only a minority of the ER binding sites and we have shown that their inclusion using a ER consensus peak set does not negatively impact the analysis (Figure 16). Although we cannot accurately quantify changes at this minority of ER binding proximal to CTCF sites upon fulvestrant treatment, the inclusion of an ER-alone ChIP-seq +/- fulvestrant would resolve this issue. Since relative binding intensities are comparable within a ChIP-seq experiment, the known absolute changes in ER binding measured by parallel-factor ChIP at non-CTCF proximal peaks can be used as a reference to quantify changes in CTCF-proximal ER peaks in the ER-alone ChIP-seq[14]. Finally, in terms of more general adoption of the method, the challenges of using CTCF binding to provide control data disappear for the large number of TFs that do not interact with CTCF. Depending on the particular needs of a study, we understand the choice of strategy may change, which is why we have provided a complete pipeline for both methods. Nonetheless, we believe the evidence we have presented demonstrates that the use of parallel-factor ChIP provides the most easily and reliably deployed solution for quantitative ChIP-seq normalisation.

Our strategy also provides a distinct advantage over ChIP-qPCR. While qPCR provides a very accurate measure of the amount of DNA present and methods exist to control against other regions of the genome, it does not account for variability in the efficiency of the immunoprecipitation step. Normalisation of ChIP-qPCR data also presents its own challenges[15]. Our method – by utilising two antibodies against

the same species – means that the reads from CTCF binding regions can control for both biases in sample loading and, more importantly, for the overall efficiency of the pulldown. Interestingly, while our method could be adapted to ChIP-qPCR, the number of controls would be limited by the number of qPCR primer sets, thereby reducing the power of the method.

The ability to confirm if the majority CTCF binding is unchanged (Figure 7B) without the need to generate further data, provides a reliable QC step for the method. Moreover, it is trivial to blacklist a set of ER-proximal CTCF sites, if there is a concern about their impact on normalisation.

For the experiments within our study, we aimed for a higher level of spike-in than previously reported to measure the efficiency of the spike-in. By randomly removing the xenogeneic reads in the Dm/H2Av controlled sample, we were able to show that with only 1% of reads coming from the spike-in, we achieved similar results in normalising the ER binding as at higher levels of spike-in. While the number of reads at 1% spike-in will not be enough to undertake reliable peak calling, standard peak locations can be used. Our analysis suggests only a very small number of spike-in reads are needed to produce robust datasets. Due to the nature of the experiment, it was not possible to apply the same simulation to the CTCF antibody control. We expect, though, that the only constraint on the level of CTCF antibody is that it must be in excess of CTCF.

Most importantly, we have developed the preprocessing and analysis tools to integrate both normalisation strategies, xenogeneic dual antibody ChIP and parallel-factor ChIP, into well-established quantitative ChIP-seq analysis methods[11]. By providing an open and reproducible pipeline, we permit others the ability to measure absolute fold-change of transcription factor binding and further develop and refine both the in-parallel control ChIP and xenogeneic spike-in normalisation methods we describe. We expect future studies of ER and similar transcription factors that undergo rapid and genome-wide changes will find the methods we present essential to accurately characterise biological effects. Our analysis tools, combined with the benefits and relative simplicity of parallel-factor ChIP to normalise ChIP-seq data, have provided a fundamental resource for quantitative transcription factor analysis.

Funding

We would like to acknowledge the support of The University of Cambridge, Cancer Research UK and Hutchison Whampoa Limited.

Parts of this work were funded by CRUK core grant [grant numbers C14303/A17197, A19274] to FM; Breast Cancer Now Award [grant number 2012NovPR042] to FM; CRUK Travel Award [grant number C60571/A24631] to AH; and a Thomas Jefferson Fellowship to AH.

Acknowledgements

MJG and ANH conceived the normalisation strategies, designed the experiments, undertook the data analysis, and wrote the manuscript. ANH undertook the ChIP-seq experiments and developed the R data and analysis packages. FM reviewed and contributed to the initial proposal and provided feedback to the project.

We would like to acknowledge the contribution from the CRUK Genomics and Bioinformatic core facilities in supporting this work.

We are grateful to Rory Stark for supplying a modified version of his DiffBind Package and to Ashley Sawle and Federico Giorgi for their ideas and support in strategies for alignments to multiple genomes.

References

1. Orlando, D.A., Chen, M.W., Brown, V.E., Solanki, S., Choi, Y.J., Olson, E.R., Fritz, C.C., Bradner, J.E. and Guenther, M.G. (2014) Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. *Cell reports* 9, **3**, 1163-1170.

2. Egan, B., Yuan, C.C., Craske, M.L., Labhart, P., Guler, G.D., Arnott, D., Maile, T.M., Busby, J., Henry, C., Kelly, T.K. and Tindell, C.A. (2016) An alternative approach to ChIP-Seq normalization enables detection of genome-wide changes in histone H3 lysine 27 trimethylation upon EZH2 inhibition. *PloS one* 11, **11**, e0166438.

3. Bonhoure, N., Bounova, G., Bernasconi, D., Praz, V., Lammers, F., Canella, D., Willis, I.M., Herr, W., Hernandez, N., Delorenzi, M. and Deplancke, B. Quantifying ChIP-seq data: a spiking method providing an internal reference for sample-to-sample normalization. *Genome research* 24, **7**,1157-1168.

4. Chen, K., Hu, Z., Xia, Z., Zhao, D., Li, W. and Tyler, J.K. (2016) The overlooked fact: fundamental need for spike-in control for virtually all genome-wide analyses. *Molecular and cellular biology* 36, **5**, 662-667.

5. Holmes, K.A., Brown, G.D. and Carroll, J.S. (2016) Chromatin Immunoprecipitation-Sequencing (ChIP-seq) for Mapping of Estrogen Receptor-Chromatin Interactions in Breast Cancer. *Estrogen Receptors: Methods and Protocols*, 79-98.

6. Early Breast Cancer Trialists' Collaborative Group. (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. *The Lancet* 351, **9114**. 1451-1467.

7. Howell, A. (2006) Pure oestrogen antagonists for the treatment of advanced breast cancer. *Endocrine-related cancer* 13, **3**, 689-706.

8. Bolstad, B.M., Irizarry, R.A., Åstrand, M. and Speed, T.P. (2003) A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* 19, **2**, 185-193.

9. Nair, N.U., Sahu, A.D., Bucher, P. and Moret, B.M. (2012) ChIPnorm: a statistical method for normalizing and identifying differential regions in histone modification ChIP-seq libraries. *PLoS one* 7, **8**, e39573.

10. Ross-Innes, C.S., Brown, G.D. and Carroll, J.S. (2011) A co-ordinated interaction between CTCF and ER in breast cancer cells. *BMC genomics* 12, **1**, 593.

11. Love, M.I., Huber, W. and Anders, S. (2014) Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* 15, **12**, 550.

12. Ross-Innes, C.S., Stark, R., Teschendorff, A.E., Holmes, K.A., Ali, H.R., Dunning, M.J., Brown, G.D., Gojis, O., Ellis, I.O., Green, A.R. and Ali, S. (2012) Differential oestrogen receptor binding is associated with clinical outcome in breast cancer. *Nature* 481, **7381**, 389.

13. Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W. and Liu, X.S. (2008) Model-based analysis of ChIP-Seq (MACS). *Genome biology* 9, **9**, R137.

14. Guertin, M.J., Martins, A.L., Siepel, A. and Lis, J.T. (2012) Accurate prediction of inducible transcription factor binding intensities in vivo. PLoS genetics, 8, **3**, p.e1002610.

15. Haring, M., Offermann, S., Danker, T., Horst, I., Peterhansel, C. and Stam, M. (2007) Chromatin immunoprecipitation: optimization, quantitative analysis and data normalization. *Plant methods* 3, **1**, 11.