PeerJ

Bioinformatic analysis of cis-regulatory interactions between progesterone and estrogen receptors in breast cancer

Matloob Khushi*, Christine L. Clarke and J. Dinny Graham

Centre for Cancer Research, Westmead Millennium Institute, Sydney Medical School—Westmead, University of Sydney, Australia

* Current affiliation: Bioinformatics Unit, Children's Medical Research Institute, Westmead, NSW, Australia

ABSTRACT

Chromatin factors interact with each other in a cell and sequence-specific manner in order to regulate transcription and a wealth of publically available datasets exists describing the genomic locations of these interactions. Our recently published BiSA (Binding Sites Analyser) database contains transcription factor binding locations and epigenetic modifications collected from published studies and provides tools to analyse stored and imported data. Using BiSA we investigated the overlapping cis-regulatory role of estrogen receptor alpha (ERα) and progesterone receptor (PR) in the T-47D breast cancer cell line. We found that ERα binding sites overlap with a subset of PR binding sites. To investigate further, we re-analysed raw data to remove any biases introduced by the use of distinct tools in the original publications. We identified 22,152 PR and 18,560 ERα binding sites (<5% false discovery rate) with 4,358 overlapping regions among the two datasets. BiSA statistical analysis revealed a non-significant overall overlap correlation between the two factors, suggesting that ERα and PR are not partner factors and do not require each other for binding to occur. However, Monte Carlo simulation by Binary Interval Search (BITS), Relevant Distance, Absolute Distance, Jaccard and Projection tests by Genometricorr revealed a statistically significant spatial correlation of binding regions on chromosome between the two factors. Motif analysis revealed that the shared binding regions were enriched with binding motifs for ERα, PR and a number of other transcription and pioneer factors. Some of these factors are known to co-locate with ERα and PR binding. Therefore spatially close proximity of ERα binding sites with PR binding sites suggests that ERα and PR, in general function independently at the molecular level, but that their activities converge on a specific subset of transcriptional targets.

Subjects Bioinformatics, Computational Biology, Molecular Biology

Keywords Transcription factors, Estrogen receptor alpha, Progesterone receptor, ERα, ESR1, PR, Breast cancer, T47D, BiSA, Genomic region database

Submitted 21 May 2014
Accepted 15 October 2014
Published 18 November 2014

Corresponding author
Matloob Khushi,
mkhushi@uni.sydney.edu.au

Academic editor
Kenta Nakai

Additional Information and Declarations can be found on page 14

DOI **10.7717/peerj.654**

INTRODUCTION

The ovarian steroid hormones progesterone and estrogen play critical roles in the development and progression of breast cancer and endometriosis (*D'Abreo & Hindenburg, 2013*; *Salehnia & Zavareh, 2013*; *Shao et al., 2014*). These hormones exert their functions

by activating specific nuclear receptors, estrogen binds to estrogen receptor (ERα) and progesterone binds to progesterone receptor (PR) (*Tsai & O'Malley, 1994*).

Once activated these receptors bind to their DNA response elements and regulate transcription of target genes. ERα and PR, along with human epidermal growth factor receptor 2 (HER2), are used to classify phenotypes in breast cancers and to predict response to specific therapies (*Cadoo, Fornier & Morris, 2013*; *Kittler et al., 2013*). A high number of ERα positive breast cancers are also PR positive (*Cadoo, Fornier & Morris, 2013*; *Penault-Llorca & Viale, 2012*). Furthermore, studies from animal models and clinical trials have shown that progesterone via its receptor PR is a major player in development and growth of breast cancer and uterine fibroids, however, PR inhibits the development of estrogen-driven endometrial cancer (*Ishikawa et al., 2010*; *Kim, Kurita & Bulun, 2013*). Many recent reviews highlight the importance of the role that progesterone and estrogen play via their receptors in various types of breast cancers (*Abdel-Hafiz & Horwitz, 2014*; *Kalkman, Barentsz & van Diest, 2014*; *Obiorah et al., 2014*; *Wang & Di, 2014*; *Yadav et al., 2014*). Therefore it is important to understand how ERα and PR work together in regulating a number of cellular pathways, and clinical and molecular research on these factors continue to unveil new insights (*Bulun, 2014*).

It is acknowledged that ERα and PR binding, as well as that of other steroid hormone receptors, is assisted by binding of the pioneer transcription factor FOXA1 (*Ballare et al., 2013*; *Lam et al., 2013*) to condensed chromatin, therefore, the interactions of FOXA1 with other factors have been well studied (*Augello, Hickey & Knudsen, 2011*; *Bernardo & Keri, 2012*). There are a number of publications that have studied PR binding sites in progesterone-treated breast and other tissues (*Ballare et al., 2013*; *Clarke & Graham, 2012*; *Yin et al., 2012*). Many studies have also published ERα binding sites (*Joseph et al., 2010*; *Schmidt et al., 2010*; *Tsai et al., 2010*). However there is lack of investigation into the combined action of the two factors on DNA. Therefore in this report we investigated the interaction of these nuclear receptors on DNA. Our previously published BiSA database (*Khushi et al., 2014*) contains a number of datasets describing ERα and PR binding sites for various cell lines, therefore, we investigated the binding pattern of these factors in the T-47D breast cancer cell line. T-47D cells are derived from metastatic female human breast cancer and are known to be ERα and PR positive and their growth is simulated by the treatment of estrogen (*Chalbos et al., 1982*; *Ström et al., 2004*).

METHODS

PR data were taken from the study of *Clarke & Graham (2012)* and ERα data were obtained from the ENCODE project (*Gertz et al., 2012*). PR data were obtained by treating T47D cells with the progestin ORG2058 for 45 min, followed by PR-specific chromatin immunoprecipitation and deep sequencing (ChIP-Seq). Gertz et al. studied ERα binding sites by treating with estradiol (E2), GEN (Genistein) and BPA (Bisphenol A) and conclude that compared to E2, GEN and BPA treatment results in fewer ERα binding sites and less change in gene expression. We selected the E2-treated dataset for our study. Datasets from both studies were of 36 base pair lengths on the Illumina platform. The PR data were

generated using an Illumina Genome Analyzer IIx while ERα libraries were sequenced on Illumina HiSeq 2000. The data used in this study have been derived from peer-reviewed publications, suggesting that they are of an acceptable quality, in addition we also ensured standard quality control checks prior to our re-analysis of the raw data. The two studies used different genome assemblies and different tools to align the reads and to call the peaks. Therefore, to remove any biases we re-analysed the raw ERα and PR data. We mapped the raw data to the GRCh37/hg19 assembly using Bowtie version 2 (*Langmead & Salzberg, 2012*). The aligned replicates were merged using Picard tools (*Li et al., 2009*) and Model-based Analysis of ChIP-seq Algorithm (MACS) version 1.4.2 (*Zhang et al., 2008*) was employed, with default settings, to identify PR and ERα binding regions in the two datasets. Regions associated with greater than 5% false discovery rate (FDR) were removed (*Zhang et al., 2008*).

We performed motif analysis using HOMER software (*Heinz et al., 2010*). HOMER employs a differential motif discovery algorithm by comparing two sets of sequences and quantifying consensus motifs that are differentially enriched in a set. HOMER automatically generates an appropriate background sequence matched for the GC content to avoid bias from CpG Islands. The tool is exclusively written for analysing DNA regulatory elements in ChIP-Seq experiments and has been used in number of high impact publications (*Berman et al., 2012*; *Wang et al., 2011b*; *Xie et al., 2013*).

Overlapping features were studied in BiSA (*Khushi et al., 2014*). BiSA is a bioinformatics database resource that can be run on Windows as a personal resource or web-based under Galaxy (*Goecks et al., 2010*) as a collaborative tool. BiSA is pre-populated with published transcription factor and histone modification datasets and allows investigators to run a number of overlapping and non-overlapping genomic region analyses using their own datasets, or against the pre-loaded Knowledge Base. Overlapping features can be visualised as a Venn diagram and binding regions of interest can also be annotated with nearby genes. BiSA also provides an easy graphical interface to find the statistical significance of observed overlap between two genomic region datasets by implementing the IntervalStat tool (*Chikina & Troyanskaya, 2012*). The tool calculates a *p*-value for each peak region by comparing a region from the query dataset to all regions in a reference dataset. The tool restricts the analysis to regions that are within a domain dataset which can be a whole genome or can be possible interval locations such as promoter proximal regions. Based on IntervalStat calculated *p*-values BiSA calculates a summary statistic that we refer to as the Overlap Correlation Value (OCV). The OCV ranges from 0 to 1, the closer the value to 1 the stronger the significance of overlap of two datasets. The OCV represents the fraction of regions in the query dataset with a *p*-value less than a specified threshold. In BiSA, we have set the threshold *p*-value to 0.05 and used a number of domains such as whole genome and promoter proximal regions for this analysis.

We also investigated the spatial correlation of regions of whole datasets being closer to each other by Binary Interval Search (BITS) (*Layer et al., 2013*) and Genometricorr (*Favorov et al., 2012*). BITS implements a Monte Carlo simulation by comparing actual overlapping regions to random observed overlap. Genometricorr considers one genomic

Table 1 Motif analysis of PR regions. Known motif analysis of PR top 1,000 regions using Homer software.

Motif	Name	*P*-value	% of targets sequences with motif
	PR(NR)/T47D	1e–123	59.40%
	FOXA1(Forkhead)/LNCAP-FOXA1	1e–28	27.10%
	AP-2gamma(AP2)/MCF7-TFAP2C	1e–10	13.70%

region set as a reference and other set as a query and provides four asymmetric pair-wise statistical tests (i) relative distance also called local correlation, (ii) absolute distance, (iii) Jaccard statistic and (iv) projection statistical tests. In local correlation the significance of relative distance between the genomic regions is measured by Kolmogorov–Smirnov test, in absolute distance test the significance of base pair distance among the regions is measured by permutation test, Jaccard statistic takes into account the ratio of intersecting bases to the union base pairs. A projection test calculates the overlapping centre points of query to reference regions and finds the significance of result outside of the null expectation by binomial test (*Favorov et al., 2012*). We performed 10,000 simulations for BITS and Genometricorr statistical tests.

We performed functional annotation of ERα-PR common cis-regulatory regions using GREAT (Genomic Regions Enrichment of Annotations Tool) (*McLean et al., 2010*). GREAT incorporates annotations from 20 ontologies covering gene ontology, phenotype data, human disease pathways, gene expression, regulatory motifs and gene families. We performed GREAT annotation using its default settings. A region was considered to have a proximal association with a gene if it was within 5 kb upstream or 1 kb downstream of the transcription start site (TSS). Regions outside this distance and up to 1,000 kb from the TSS to the next gene proximal region were considered to have a distal association.

RESULTS

Analysis of PR and ERα ChIP-seq data from T-47D breast cancer cells revealed 22,152 PR and 18,560 ERα binding regions with FDR <5%. HOMER motif analysis on the top ranked 1,000 regions by peak score revealed the strong presence of a PRE motif (59.40%) and ERE motif (48.80%) (Tables 1 and 2). These were the most statistically significant motifs identified, in agreement with other studies (*Kim, Kurita & Bulun, 2013*; *Lin et al., 2007*). In addition, in PR binding regions we found motifs for the transcriptional partners FOXA1 and AP-2 (TFAP2C) as other top ranked motifs. The transcription factor activator protein 2C (TFAP2C) is known to be involved in normal mammary development, differentiation, and oncogenesis (*Cyr et al., in press*; *Lal et al., 2013*; *Woodfield et al., 2010*). Interestingly PR motifs were present in 344 (34.4%) of the 1,000 top ranked ERα binding regions. Consensus FOXA1 motifs were also detected in 27% of PR binding regions and 24% of regions bound by ERα. FOXA1 is a member of the forkhead family of transcription factors, which are known to bind and reconfigure condensed chromatin to

Table 2 Motif analysis of ERα regions. Known motif analysis of ESR1 top 1,000 regions.

Motif	Name	*P*-value	% of targets sequences with motif
	ERE(NR/IR3)/MCF7-ERa	1e–474	48.80%
	FOXA1(Forkhead)/LNCAP-FOXA1	1e–22	24.30%
	PR(NR)/T47D-PR	1e–20	34.40%

enable the binding of other transcription factors (*Bernardo & Keri, 2012*) . The presence of high quality (p-value <1.00e–05) peaks and known conserved PR and ERα recognition sequences confirmed the success of the alignment and peak-calling process.

The size distribution of ERα (18,560 regions) and PR (22,152 regions) binding regions were visualised by drawing a histogram and box plot (Figs. 1 and 2). Mean PR binding region size was 1508 with a median of 1336. In contrast, ERα binding regions were on average half the size of PR binding regions, with a mean size of 601 and median 529. Most PR binding regions (∼94%) were greater than 1 kb, whereas most ERα binding regions (∼95%) were less than 1 kb. The longer PR regions may be due to longer input DNA fragment lengths in the original samples (*Kharchenko, Tolstorukov & Park, 2008*; *Landt et al., 2012*).

Limited overlap of ERα and PR regions

Using BiSA, we identified that almost one quarter (23.6%) of ERα binding regions (4,344) overlap with 3,870 unique PR binding regions. This revealed that some long PR binding regions spanned more than one ERα binding region and the reverse was also true for large ERα binding regions. In total, we found 4,358 sections that were common to the two datasets. The Venn diagram in Fig. 3A shows this overlap between the two ligand-activated transcription factors. The 4,358 overlapping sections of the regions common to the two datasets were extracted and plotted for their region lengths (Fig. 3B). Out of 4,358 overlapping sections 4,279 (98.2%) were more than 100 bases long, suggesting a strong binding overlap between the two transcription factor data sets. An example of a shared ERα and PR binding region is shown in Fig. 4. The 631 bp ERα binding region (red dotted lines) is completely contained within the 813 bp PR binding region (blue dotted lines) and the two regions share the peak centre location (Fig. 4).

Statistical analysis of ERα-PR overlap

To determine whether the overlap between ERα and PR binding was statistically significant, statistical analysis was performed in BiSA, BITS and Genometricorr. In BiSA, using a whole genome domain and selecting the ERα cistrome as query and PR as reference revealed an overlap correlation value of 0.33. The value decreased to 0.26 when PR was selected as query and ERα as reference. This showed that, although a considerable proportion of ERα binding regions are also bound by PR, the two receptors

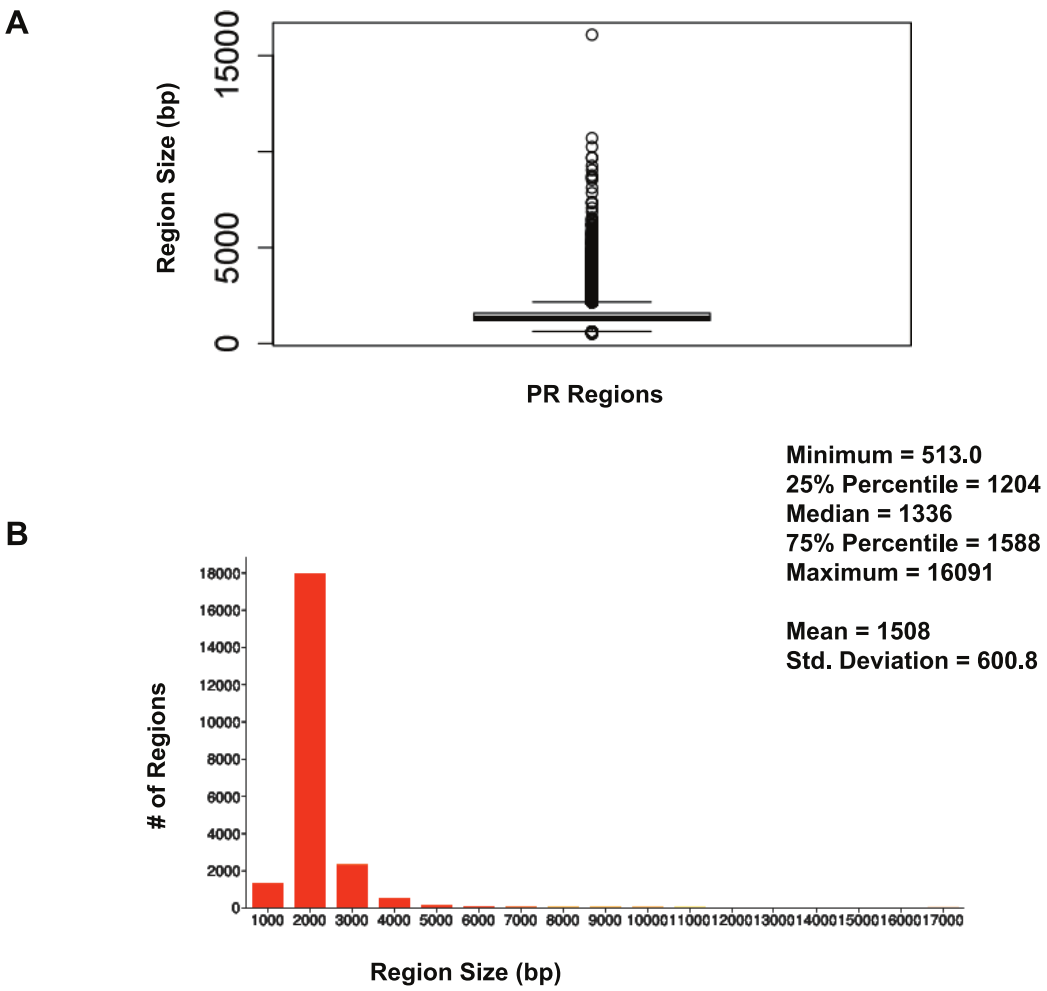

Figure 1 Distribution of PR binding region sizes. (A) Box plot with mean and median information. (B) Histogram of region sizes with bin size 1,000.

do not cooperate for binding at all sites. To determine whether the significance of ERα-PR binding overlap was greater in functionally relevant genomic regions, we compared the level of binding overlap over a range of genomic domains from promoter proximal (within 500 b of a TSS) to more distal regions (Table 3). We found a low though consistent overlap correlation value ($\sim$0.3) whether promoter proximal or distal sites were included in the analysis (Table 3). To confirm that the OCV result is independent of the mean region sizes of the two datasets, we fixed the PR region sizes to 300 bases from each side of peak summits to match mean ERα region length (mean = 601) and performed the OCV test again. This did not change the OCV (0.33) for the whole genome dataset, and there was negligible change in OCV observed for other domains (Table 3).

Using BITS and Genometricorr, we further investigated whether the spatial proximity correlation between PR and ERα binding was more significant than expected by chance. BITS Monte Carlo simulation reported that the spatial correlation of ERα and PR was statistically significant, with a p-value of 0.0001. Similarly Genometricorr's Relative

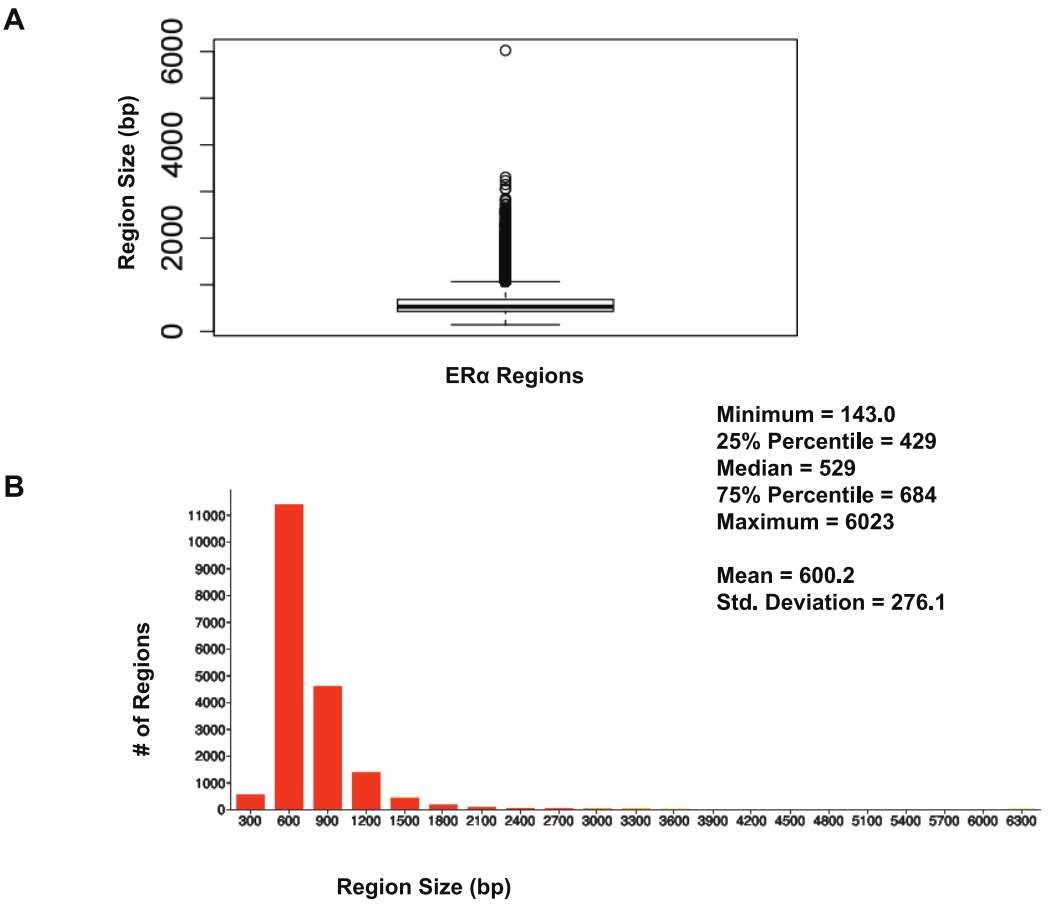

Figure 2 Distribution of ERα binding region sizes. (A) Box plot with mean and median information. (B) Histogram of ERα region sizes with bin 200.

Correlation test, Absolute Distance test, Jaccard test and Projection tests also reported the spatial correlation between the two factors as statistically significant (p-value =<1e–04) (Fig. 5). We repeated the tests for the 600bp fixed-width PR dataset and found no change in reported p-values from BITS or Genometricorr. This confirmed that a change in average region size between the two datasets does not affect the statistical analysis and demonstrated that the tendency for binding events for the two factors to be close to each other is statistically significant. Therefore we conclude that, although there are a number of statistically significant shared binding sites in the ERα and PR datasets, and that ERα and PR often bind in proximity to each other, the observed overlap of the two factors is not strong enough for them to be considered as co-factors that consistently co-operate on shared binding regions. However, the close proximity of the binding regions for the two factors shows a spatial convergence and is statistically significant.

Motif analysis

The 4,358 common sections of ERα-PR were searched for known motifs. Known motif analysis in these common sections revealed a strong presence of ERE, forkhead protein

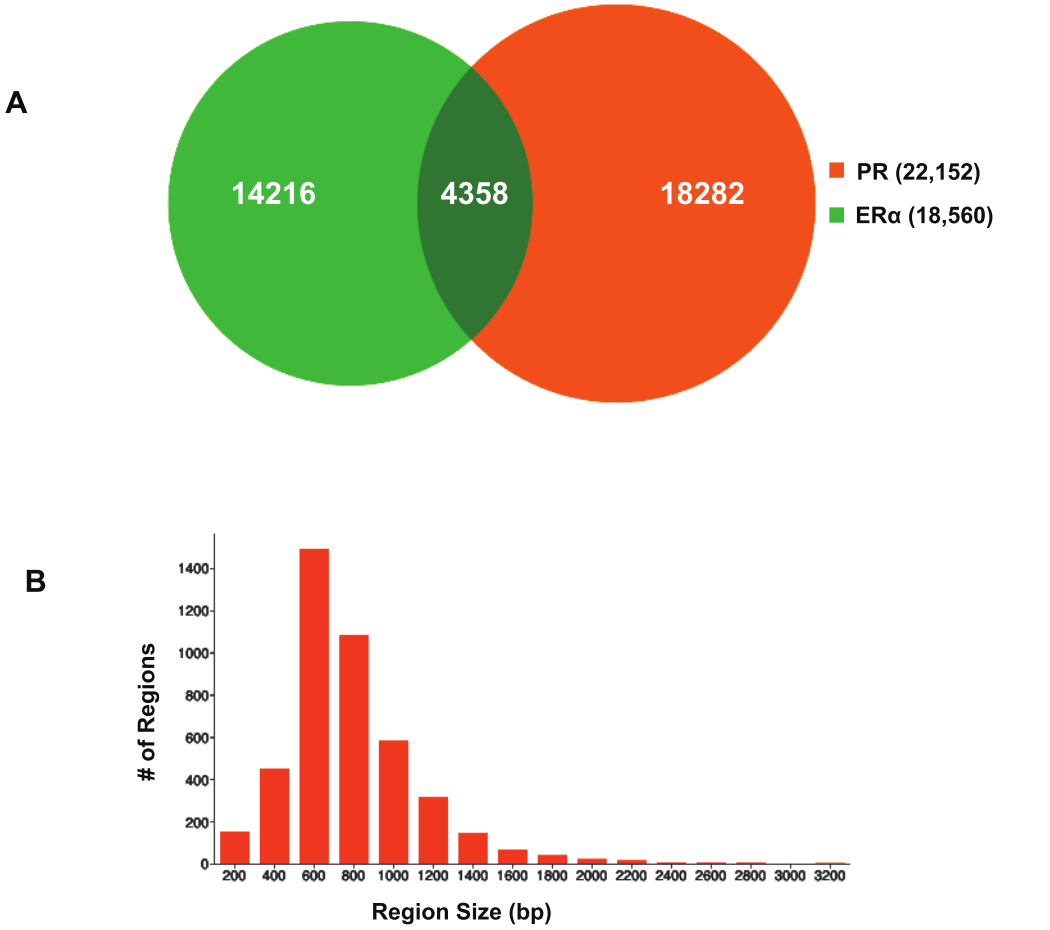

Figure 3 Visualisation of ERα and PR binding region overlap. (A) Venn diagram showing overlap between ERα and PR data. The 4,344 ERα binding regions overlap with 3,870 unique PR binding regions making up 4,358 overlapping sections. (B) Region sizes of 4,358 regions common to the ERα and PR datasets.

and PRE motifs. In Table 4, we listed the top ranked motifs, ordered by *p*-value. A PRE motif was found in 41.88% (1,825) of the total 4,358 regions, which was much higher than the number of ERE motifs detected 14.3% (623) of the sequences. However, this may reflect the higher stringency of the position specific scoring matrix used to identify ERE motif occurrence than the matrix used to find PRE motifs since the *p*-value for ERE motif detection (1e–291) was much stronger than the *p*-value for PRE motif occurrence in the dataset (1e–179). The presence of FOXA1 motifs in these regions confirms that the factor facilitates the binding of ERα and PR on these regions as previously reported (*Augello, Hickey & Knudsen, 2011*; *Bernardo & Keri, 2012*; *Nakshatri & Badve, 2009*). In addition AP-2 and TEAD4 (TEA) motifs were also identified in these regions and in the 1,000 top scoring PR binding regions. AP-2 has a known role in normal mammary development and breast cancer (*Cyr et al., in press*; *Lal et al., 2013*; *Woodfield et al., 2010*). TEAD4 has also been shown to be co-expressed with other oncogenes and is correlated with poor prognosis (*Xia et al., 2014*; *Mesrouze et al., 2014*; *Lim et al., 2014*). The presence of the related motifs in

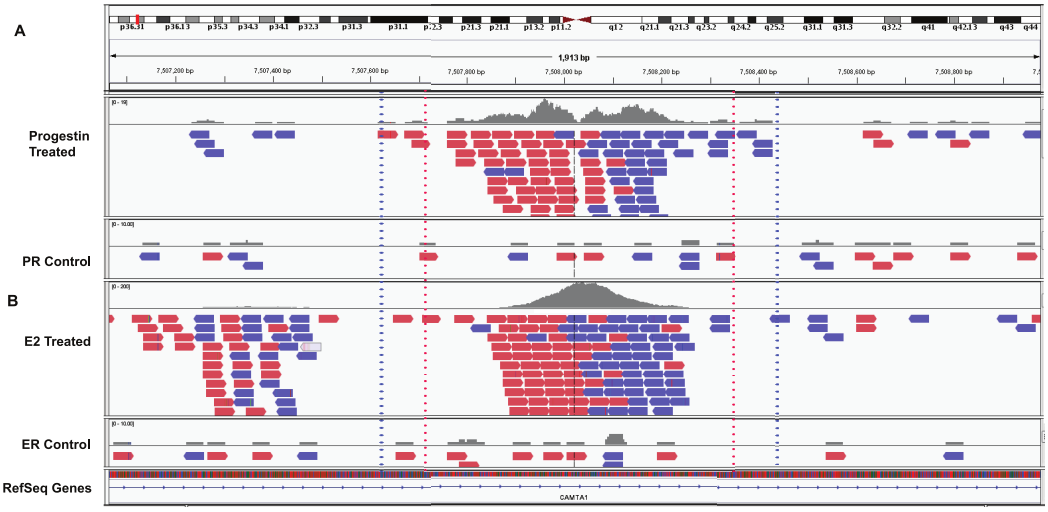

Figure 4 Example overlapping region. IGV snapshot of PR binding region at chr1:7507615–7508428 (marked by blue dotted lines) and ERα binding region (marked by red dotted lines). (A) Progestin treated and control samples. (B) Estradiol (E2) treated and control sample. The red boxes are reads that mapped to the forward strand and blue boxes are reads that mapped to the reverse strand of the human genome (build hg19).

Table 3 BiSA Overlap Correlation Value (OCV) testing. BiSA Statistical analysis of overlap between ERα and PR datasets using different domain datasets.

Domain	Overlap Correlation Value (OCV)			# of overlaps[b]/total ERα regions in domain
	Query = ERα Reference = PR	Query = PR Reference = ERα	Query = ERα Reference = PR (600 bp long)[a]	
Whole Genome	0.33	0.26	0.33	4,344/18,560
500 bp upstream, downstream of TSS	0.3	0.17	0.22	112/419
1 kb upstream, downstream of TSS	0.28	0.18	0.25	157/647
5 kb upstream of TSS	0.3	0.21	0.28	304/1,224
5 kb upstream, downstream of TSS	0.31	0.22	0.3	522/2,147
10 kb upstream, downstream of TSS	0.31	0.22	0.3	929/3,666
45 kb–55 kb upstream of TSS	0.29	0.21	0.28	449/1,929
95 kb–105 kb upstream of TSS	0.31	0.24	0.3	514/2,017
90 kb–110 kb upstream of TSS	0.31	0.23	0.3	878/3,495

Notes.

[a] PR regions are fixed to 600 bp long by cutting off 300 bp on both sides of peak summits.

[b] Number of overlaps in this column is reported by selecting ERα as the query and PR as the reference dataset.

the ERα-PR shared regions as well as in regions that bind uniquely ERα or PR suggests that AP-2 and/or TEAD play a key role for both receptors and could be important in facilitating cooperation between the two nuclear receptors.

Using Homer, we also looked at relative position distributions of these motifs (Fig. 6). We found that the motifs converge around the centres of the peaks, supporting their biological significance as primary binding events.

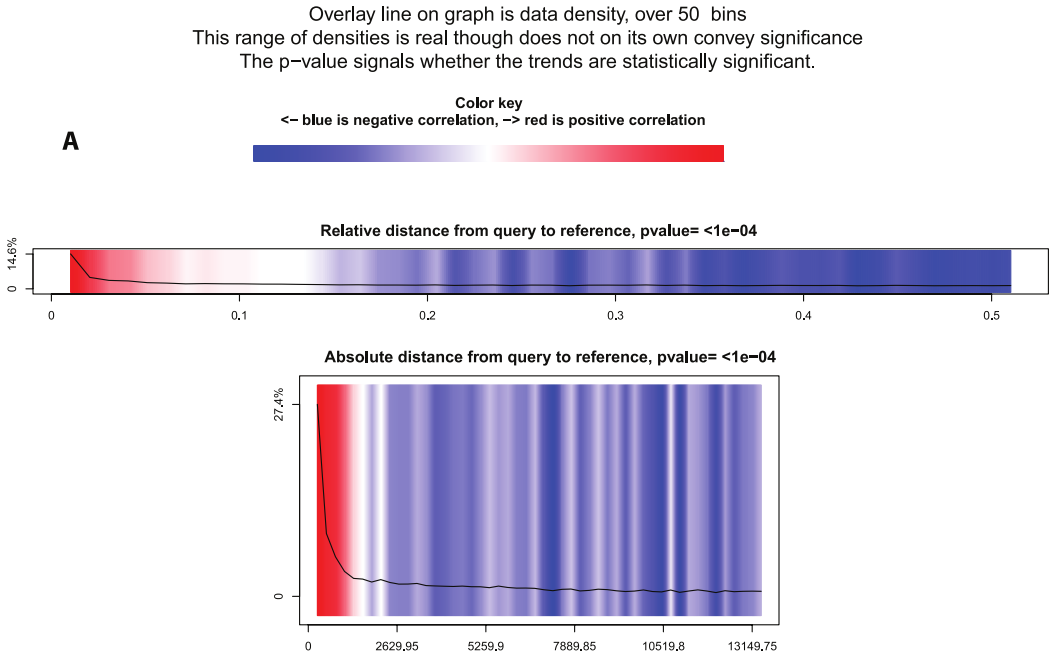

Figure 5 Statistical significance test using Genometricorr. Genometricorr statistical significance analysis of ERα (query)-PR (reference). (A) Relative and Absolute Distance Correlation tests are shown graphically. Overlay line (data density) when in the blue section shows negative correlation while the high density in the red section shows positive correlation. (B) Results from Jaccard and Projection tests are shown in text.

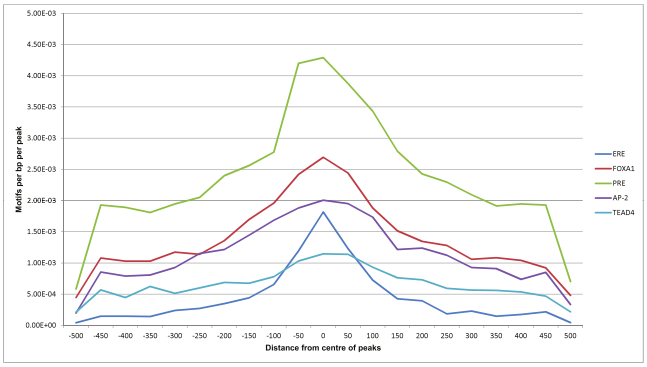

Figure 6 Motif position distributions in ERα-PR overlapping regions. Frequency distribution of ERE, FOXA1, PRE, AP-2 and TEAD4 motifs around centres of peaks using a 50 bp bin size.

Table 4 Known motif analysis of ERα and PR overlapping common regions. Top ranked known motif analysis of ERα-PR common sections (4,358 regions).

Motif	Name	*P*-value	% of targets sequences with motif
	ERE(NR/IR3)/MCF7-ERa	1e–291	14.30%
	FOXA1(Forkhead)/LNCAP-FOXA1	1e–249	35.11%
	PR(NR)/T47D-PR	1e–179	41.88%
	AP-2gamma(AP2)/MCF7-TFAP2C	1e–122	20.38%
	TEAD4(TEA)/Tropoblast-Tead4	1e–86	17.97%

Enrichment analysis of ERα-PR common regions

We used GREAT (Genomic Regions Enrichment of Annotations Tool) (*McLean et al., 2010*) to interpret the functional role of 4,358 ERα-PR common regions. GREAT revealed that only 34 regions (∼0.8%) are not associated with any gene and 3,687 (∼85%) regions are associated with 2 genes (Fig. 7). Most of the regions were found to be distal binding events while 405 (∼9%) regions are within 5 kb of transcription start sites (TSS). Region to gene association revealed MYC has the maximum number of regions linked to this gene (26 regions). The known role of estrogen-induced MYC oncogene in breast cancer (*Orr et al., 2012*; *Wang et al., 2011a*) confirms a biological relevant regions-to-gene association. PGR was also among the top 10 genes identified with the largest number of associated regions (File S1). Gene ontology enrichment analysis of the common regions revealed epithelial cell development as the most significant biological process (File S1). Epithelial cell development was linked to 30 genes associated with 120 regions out of which 4 regions were within 5 kb of a TSS. Pathway Commons, a meta-database of public biological pathway information (*Cerami et al., 2006*), revealed the ERα signalling network as the most significant term (p-value = 5.7e–37) where 137 regions were found regulating 24 genes associated with this pathway. The FOXA1 transcription factor network and IL6-dedicated signalling events were also significant terms (p-value 1.6e–19 and 2.6e–17). Mouse phenotype analysis revealed two breast cancer related ontologies (abnormal mammary gland epithelium physiology and abnormal mammary gland development) as the most significant terms. There were 32 regions associated with 5 genes linked to abnormal mammary gland epithelium physiology and 189 regions associated with 52 genes linked to mammary gland development. The File S1 also lists regions and associated genes with the ontologies.

DISCUSSION

The BiSA database provides a good starting point for studying overlapping binding by a range of transcription factors from a comprehensive collection of published studies (*Khushi et al., 2014*). The datasets available in BiSA represent the original genomic

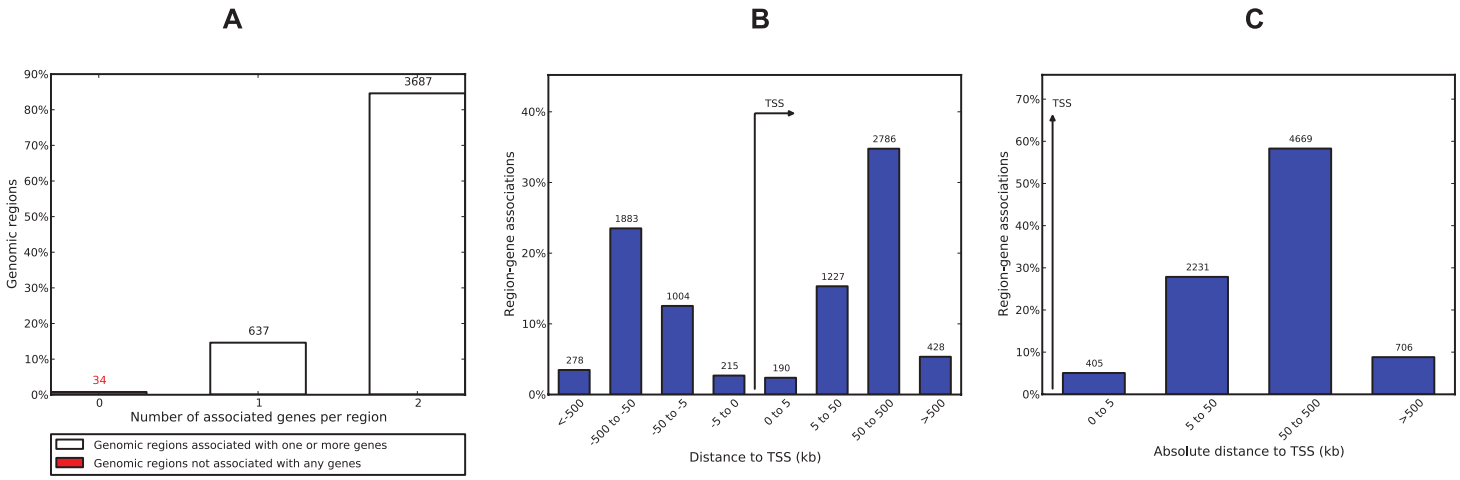

Figure 7 ERα-PR common region-gene association. (A) Number of associated genes per region. (B) Region-gene association binned by orientation and distance to TSS. (C) Region-gene association binned by absolute distance to TSS.

locations identified in the published studies from which they are sourced. Although the same standard pipeline has often been applied, it must be acknowledged that differences in read alignment algorithms (*Kerpedjiev et al., 2014*; *Lunter & Goodson, 2011*) and the use of a variety of peak-caller programmes (*Ladunga, 2010*; *Pepke, Wold & Mortazavi, 2009*; *Wilbanks & Facciotti, 2010*) has an impact on downstream analysis, largely due to differences in stringency that affects the number of genomic regions identified. Our initial investigation of the overlap in ERα and PR binding in T-47D cells, utilizing the published binding regions, revealed an overlap of ∼27% of ERα binding regions with the published PR cistrome (data not shown). This suggested an interesting functional relationship between the receptors, which justified further study. To perform a more rigorous exploration of their overlapping binding patterns, we reanalysed the raw ERα and PR ChIP-seq data using a standardized pipeline. This illustrates the great value of BiSA as an easy to implement first pass tool to investigate potential functional relationships in transcription factor binding and epigenomic datasets.

The BiSA statistical overlap correlation value (OCV) represents a statistical summary value of the set of *p*-values calculated by the IntervalStat tool and reflects the overall correlation of two binding site datasets. IntervalStat calculates a *p*-value for each query region against the closest reference region within the given domain. It is designed to identify factors that target the same genomic locations. As described in examples in our previous study (*Khushi et al., 2014*) the OCV should be greater than 0.5 for partner factors, reflecting a statistically significant correlation between two binding patterns. For example the OCV for known partners, FOXA3 (query) to FOXA1 (reference) was 0.72 (*Motallebipour et al., 2009*). Similarly the OCV for CTCF (query) and SA1 (reference), which are known to co-locate on DNA, was 0.82 (*Schmidt et al., 2010*). Therefore the lower OCV for ERα-PR suggests that the majority of ERα and PR binding events are independent of each other, however, the OCV test does not challenge the biological

co-occurrence of binding of the two factors on the reported regions where IntervalStat reports a statistically significant *p*-value.

A consistent overlap was found both proximal and distal to gene promoters (Table 3). It is acknowledged that gene expression is regulated through interaction at a number of cis-regulatory elements, which includes promoters and enhancers. Moreover, enhancers can spread over a range of distances from the TSS. Therefore, the detection of binding sites over a range of distances and locations is to be expected (*Bulger & Groudine, 2011*; *Calo & Wysocka, 2013*). This spatial correlation between the two factors is identified as statistically significant by Monte Carlo simulation using BITS, Relevant Distance, Absolute Distance, Jaccard and Projection tests using Genometricorr. Therefore, the regions from the two factors are found in close proximity more often than expected by chance although they do not exactly overlap. Therefore the consistent OCV observed using various domains and statistically significant spatial convergence suggest that the consistent overlap may have biological significance. Although not all sites overlapped, many of the shared ERα and PR binding regions were highly statistically significant binding sites for both receptors, as determined by a strong *p*-value and low FDR value in MACS, suggesting that these are biologically valid binding regions for these receptors and that their overlap reflects converging function on a subset of gene targets.

In recent years a number of studies have published ERα binding regions in the MCF-7 cell line (*Grober et al., 2011*; *Gu et al., 2010*; *Hu et al., 2010*; *Hurtado et al., 2008*; *Joseph et al., 2010*; *Schmidt et al., 2010*; *Tsai et al., 2010*; *Welboren et al., 2009*). However only two studies have published ERα data in T47D cells (*Gertz et al., 2012*; *Joseph et al., 2010*). We chose to study the *Gertz et al. (2012)* dataset because using data from the *Joseph et al. (2010)* study we called only 1,817 peaks with FDR <5%, which can be an indication of low quality ChIP (*Landt et al., 2012*). On the other hand for the PR dataset, we did not employ the datasets published by *Yin et al. (2012)* because the experiment was performed with an antiprogestin (RU486) treatment, which would not be expected to elicit the same binding pattern as PR agonist, and lacked any control sample. MACS distributes read tags from the control sample along the genome to model Poisson distribution, and false discovery rate (FDR) is calculated by swapping control and ChIP samples. Therefore it is recommended for ChIP-seq studies to have an appropriate input control sample (*Wilbanks & Facciotti, 2010*). ENCODE guidelines also emphasise the importance of using a suitable control dataset to adjust for variable DNA fragment lengths (*Landt et al., 2012*).

There is a slight difference in the reported low-significance motifs for PR data between this report and the Clarke and Graham study (*Clarke & Graham, 2012*). The two most significant motifs (PRE are FOXA1) are the same in the two studies, however, Clarke and Graham found an NF1 half-site as one of the significant motifs and AP-1 sites as non-significant while in this study we found an AP-2 motif higher in significance than the NF1 motif (not shown). This minor difference is due to the difference in binding regions as Clarke and Graham published 6,312 PR bound regions in T47D cells by aligning to hg18 and using the ERANGE peak caller, however, in this study we reported 22,152 PR regions by aligning to hg19 assembly and using MACS as our peak caller.

The ERα-PR data was collected from two separate publications where the binding of each factor was studied by stimulation of T-47D cells with estrogen or progesterone independently. Therefore the focus of this study was to examine the correlation of ERα-PR binding patterns which revealed an interesting convergence on specific loci. We studied the association between common regions and nearby genes and found biologically relevant gene pathways. The Myc oncogene, which was most highly associated with binding sites common to ERα and PR, is a known target of both estrogen and progesterone and plays a key role in the normal breast and breast cancer (*Curtis et al., 2012*; *Hynes & Stoelzle, 2009*). PR itself is also regulated by both hormones and the PGR gene was highly associated with shared ERα and PR binding regions. Transcriptional regulation by estrogen and progesterone co-treatment in this cell model was not available, however it would be interesting to study the binding of the two factors under the influence of both stimuli (estrogen and progesterone) to observe the impact of converging ERα and PR regulation in comparison to individual stimulation.

CONCLUSION

In summary, we have evidence for a biologically relevant interplay between PR and ERα in a subset of binding sites in breast cancer cells. Our analysis demonstrated the utility of our previously published software BiSA (*Khushi et al., 2014*), which has a comprehensive knowledge base, consisting of transcription factor binding sites and histone modifications collected from previously published studies. Using BiSA we identified that ERα and PR co-locate on a subset of binding sites. The BiSA statistical testing of overlap revealed a low overlap correlation value (OCV) suggesting that the two factors are not obligate cofactors. However, spatial correlation testing using Monte Carlo simulation by BITS, Relevant Distance, Absolute Distance, Jaccard and Projection tests by Genometricorr revealed a statistically significant correlation between the two factors. In addition, the discovery that ERα, FOXA1, PR, AP-2 and TEAD4 binding motifs are significantly enriched in regions that are bound by both ERα and PR suggests that their overlap is biologically relevant.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

MK was previously supported by Australian Postgraduate Award (APA) and Westmead Medical Research Foundation (WMRF) Top-Up scholarship. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
Australian Postgraduate Award (APA).
Westmead Medical Research Foundation (WMRF).

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Matloob Khushi conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Christine L. Clarke and J. Dinny Graham conceived and designed the experiments, reviewed drafts of the paper.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.654#supplemental-information.

REFERENCES

Abdel-Hafiz HA, Horwitz KB. 2014. Post-translational modifications of the progesterone receptors. *Journal of Steroid Biochemistry and Molecular Biology* **140**:80–89 DOI 10.1016/j.jsbmb.2013.12.008.

Augello MA, Hickey TE, Knudsen KE. 2011. FOXA1: master of steroid receptor function in cancer. *EMBO Journal* **30**:3885–3894 DOI 10.1038/emboj.2011.340.

Ballare C, Castellano G, Gaveglia L, Althammer S, Gonzalez-Vallinas J, Eyras E, Le Dily F, Zaurin R, Soronellas D, Vicent GP, Beato M. 2013. Nucleosome-driven transcription factor binding and gene regulation. *Molecular Cell* **49**:67–79 DOI 10.1016/j.molcel.2012.10.019.

Berman BP, Weisenberger DJ, Aman JF, Hinoue T, Ramjan Z, Liu Y, Noushmehr H, Lange CPE, Van Dijk CM, Tollenaar RAEM, Van Den Berg D, Laird PW. 2012. Regions of focal DNA hypermethylation and long-range hypomethylation in colorectal cancer coincide with nuclear lamina-associated domains. *Nature Genetics* **44**:40–46 DOI 10.1038/ng.969.

Bernardo GM, Keri RA. 2012. FOXA1: a transcription factor with parallel functions in development and cancer. *Bioscience Reports* **32**:113–130 DOI 10.1042/BSR20110046.

Bulger M, Groudine M. 2011. Functional and mechanistic diversity of distal transcription enhancers. *Cell* **144**:327–339 DOI 10.1016/j.cell.2011.01.024.

Bulun SE. 2014. Aromatase and estrogen receptor alpha deficiency. *Fertility and Sterility* **101**:323–329 DOI 10.1016/j.fertnstert.2013.12.022.

Cadoo KA, Fornier MN, Morris PG. 2013. Biological subtypes of breast cancer: current concepts and implications for recurrence patterns. *The Quarterly Journal of Nuclear Medicine and Molecular Imaging* **57**:312–321.

Calo E, Wysocka J. 2013. Modification of enhancer chromatin: what, how, and why? *Molecular Cell* **49**:825–837 DOI 10.1016/j.molcel.2013.01.038.

Cerami EG, Bader GD, Gross BE, Sander C. 2006. cPath: open source software for collecting, storing, and querying biological pathways. *BMC Bioinformatics* **7**:497 DOI 10.1186/1471-2105-7-497.

Chalbos D, Vignon F, Keydar I, Rochefort H. 1982. Estrogens stimulate cell proliferation and induce secretory proteins in a human breast cancer cell line (T47D). *Journal of Clinical Endocrinology and Metabolism* **55**:276–283 DOI 10.1210/jcem-55-2-276.

Chikina MD, Troyanskaya OG. 2012. An effective statistical evaluation of ChIPseq dataset similarity. *Bioinformatics* **28**:607–613 DOI 10.1093/bioinformatics/bts009.

Clarke CL, Graham JD. 2012. Non-overlapping progesterone receptor cistromes contribute to cell-specific transcriptional outcomes. *PLoS ONE* **7**:e35859 DOI 10.1371/journal.pone.0035859.

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, Speed D, Lynch AG, Samarajiwa S, Yuan Y, Graf S, Ha G, Haffari G, Bashashati A, Russell R, McKinney S, Group M, Langerod A, Green A, Provenzano E, Wishart G, Pinder S, Watson P, Markowetz F, Murphy L, Ellis I, Purushotham A, Borresen-Dale AL, Brenton JD, Tavare S, Caldas C, Aparicio S. 2012. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**:346–352 DOI 10.1038/nature10983.

Cyr AR, Kulak MV, Park JM, Bogachek MV, Spanheimer PM, Woodfield GW, White-Baer LS, O'Malley YQ, Sugg SL, Olivier AK, Zhang W, Domann FE, Weigel RJ. 2014. TFAP2C governs the luminal epithelial phenotype in mammary development and carcinogenesis. *Oncogene* In Press.

D'Abreo N, Hindenburg AA. 2013. Sex hormone receptors in breast cancer. *Vitamins and Hormones* **93**:99–133 DOI 10.1016/B978-0-12-416673-8.00001-0.

Favorov A, Mularoni L, Cope LM, Medvedeva Y, Mironov AA, Makeev VJ, Wheelan SJ. 2012. Exploring massive, genome scale datasets with the GenometriCorr package. *PLoS Computational Biology* **8**:e1002529 DOI 10.1371/journal.pcbi.1002529.

Gertz J, Reddy TE, Varley KE, Garabedian MJ, Myers RM. 2012. Genistein and bisphenol a exposure cause estrogen receptor 1 to bind thousands of sites in a cell type-specific manner. *Genome Research* **22**:2153–2162 DOI 10.1101/gr.135681.111.

Goecks J, Nekrutenko A, Taylor J, Galaxy T. 2010. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biology* **11**:R86 DOI 10.1186/gb-2010-11-8-r86.

Grober OM, Mutarelli M, Giurato G, Ravo M, Cicatiello L, De Filippo MR, Ferraro L, Nassa G, Papa MF, Paris O, Tarallo R, Luo S, Schroth GP, Benes V, Weisz A. 2011. Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC Genomics* **12**:36 DOI 10.1186/1471-2164-12-36.

Gu F, Hsu HK, Hsu PY, Wu J, Ma Y, Parvin J, Huang TH, Jin VX. 2010. Inference of hierarchical regulatory network of estrogen-dependent breast cancer through ChIP-based data. *BMC Systems Biology* **4**:170 DOI 10.1186/1752-0509-4-170.

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589 DOI 10.1016/j.molcel.2010.05.004.

Hu M, Yu J, Taylor JM, Chinnaiyan AM, Qin ZS. 2010. On the detection and refinement of transcription factor binding sites using ChIP-Seq data. *Nucleic Acids Research* **38**:2154–2167 DOI 10.1093/nar/gkp1180.

Hurtado A, Holmes KA, Geistlinger TR, Hutcheson IR, Nicholson RI, Brown M, Jiang J, Howat WJ, Ali S, Carroll JS. 2008. Regulation of ERBB2 by oestrogen receptor-PAX2 determines response to tamoxifen. *Nature* **456**:663–666 DOI 10.1038/nature07483.

Hynes NE, Stoelzle T. 2009. Key signalling nodes in mammary gland development and cancer: Myc. *Breast Cancer Research* **11**:210 DOI 10.1186/bcr2406.

Ishikawa H, Ishi K, Serna VA, Kakazu R, Bulun SE, Kurita T. 2010. Progesterone is essential for maintenance and growth of uterine leiomyoma. *Endocrinology* **151**:2433–2442 DOI 10.1210/en.2009-1225.

Joseph R, Orlov YL, Huss M, Sun W, Kong SL, Ukil L, Pan YF, Li G, Lim M, Thomsen JS, Ruan Y, Clarke ND, Prabhakar S, Cheung E, Liu ET. 2010. Integrative model of genomic factors for determining binding site selection by estrogen receptor-alpha. *Molecular Systems Biology* **6**:456 DOI 10.1038/msb.2010.109.

Kalkman S, Barentsz MW, Van Diest PJ. 2014. The effects of under 6 hours of formalin fixation on hormone receptor and HER2 expression in invasive breast cancer: a systematic review. *American Journal of Clinical Pathology* **142**:16–22 DOI 10.1309/AJCP96YDQSTYBXWU.

Kerpedjiev P, Frellsen J, Lindgreen S, Krogh A. 2014. Adaptable probabilistic mapping of short reads using position specific scoring matrices. *BMC Bioinformatics* **15**:100 DOI 10.1186/1471-2105-15-100.

Kharchenko PV, Tolstorukov MY, Park PJ. 2008. Design and analysis of ChIP-seq experiments for DNA-binding proteins. *Nature Biotechnology* **26**:1351–1359 DOI 10.1038/nbt.1508.

Khushi M, Liddle C, Clarke CL, Graham JD. 2014. Binding sites analyser (BiSA): software for genomic binding sites archiving and overlap analysis. *PLoS ONE* **9**:e87301 DOI 10.1371/journal.pone.0087301.

Kim JJ, Kurita T, Bulun SE. 2013. Progesterone action in endometrial cancer, endometriosis, uterine fibroids, and breast cancer. *Endocrine Reviews* **34**:130–162 DOI 10.1210/er.2012-1043.

Kittler R, Zhou J, Hua S, Ma L, Liu Y, Pendleton E, Cheng C, Gerstein M, White KP. 2013. A comprehensive nuclear receptor network for breast cancer cells. *Cell Reports* **3**:538–551 DOI 10.1016/j.celrep.2013.01.004.

Ladunga I. 2010. An overview of the computational analyses and discovery of transcription factor binding sites. *Methods in Molecular Biology* **674**:1–22 DOI 10.1007/978-1-60761-854-6_1.

Lal G, Contreras PG, Kulak M, Woodfield G, Bair T, Domann FE, Weigel RJ. 2013. Human Melanoma cells over-express extracellular matrix 1 (ECM1) which is regulated by TFAP2C. *PLoS ONE* **8**:e73953 DOI 10.1371/journal.pone.0073953.

Lam EW, Brosens JJ, Gomes AR, Koo CY. 2013. Forkhead box proteins: tuning forks for transcriptional harmony. *Nature Reviews Cancer* **13**:482–495 DOI 10.1038/nrc3539.

Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shoresh N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. 2012. ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome Research* **22**:1813–1831 DOI 10.1101/gr.136184.111.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nature Methods* **9**:357–359 DOI 10.1038/nmeth.1923.

Layer RM, Skadron K, Robins G, Hall IM, Quinlan AR. 2013. Binary Interval Search: a scalable algorithm for counting interval intersections. *Bioinformatics* **29**:1–7 DOI 10.1093/bioinformatics/bts652.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S. 2009. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**:2078–2079 DOI 10.1093/bioinformatics/btp352.

Lim B, Park JL, Kim HJ, Park YK, Kim JH, Sohn HA, Noh SM, Song KS, Kim WH, Kim YS, Kim SY. 2014. Integrative genomics analysis reveals the multilevel dysregulation and oncogenic characteristics of TEAD4 in gastric cancer. *Carcinogenesis* **35**:1020–1027 DOI 10.1093/carcin/bgt409.

Lin CY, Vega VB, Thomsen JS, Zhang T, Kong SL, Xie M, Chiu KP, Lipovich L, Barnett DH, Stossi F, Yeo A, George J, Kuznetsov VA, Lee YK, Charn TH, Palanisamy N, Miller LD, Cheung E, Katzenellenbogen BS, Ruan Y, Bourque G, Wei CL, Liu ET. 2007. Whole-genome cartography of estrogen receptor alpha binding sites. *PLoS Genetics* **3**:e87 DOI 10.1371/journal.pgen.0030087.

Lunter G, Goodson M. 2011. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. *Genome Research* **21**:936–939 DOI 10.1101/gr.111120.110.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nature Biotechnology* **28**:495–501 DOI 10.1038/nbt.1630.

Mesrouze Y, Hau JC, Erdmann D, Zimmermann C, Fontana P, Schmelzle T, Chene P. 2014. The surprising features of the TEAD4-Vgll1 protein–protein interaction. *ChemBioChem* **15**:537–542 DOI 10.1002/cbic.201300715.

Motallebipour M, Ameur A, Reddy Bysani MS, Patra K, Wallerman O, Mangion J, Barker MA, McKernan KJ, Komorowski J, Wadelius C. 2009. Differential binding and co-binding pattern of FOXA1 and FOXA3 and their relation to H3K4me3 in HepG2 cells revealed by ChIP-seq. *Genome Biology* **10**:R129 DOI 10.1186/gb-2009-10-11-r129.

Nakshatri H, Badve S. 2009. FOXA1 in breast cancer. *Expert Reviews in Molecular Medicine* **11**:e8 DOI 10.1017/S1462399409001008.

Obiorah IE, Fan P, Sengupta S, Jordan VC. 2014. Selective estrogen-induced apoptosis in breast cancer. *Steroids* **90**:60–70 DOI 10.1016/j.steroids.2014.06.003.

Orr N, Lemnrau A, Cooke R, Fletcher O, Tomczyk K, Jones M, Johnson N, Lord CJ, Mitsopoulos C, Zvelebil M, McDade SS, Buck G, Blancher C, Consortium KC, Trainer AH, James PA, Bojesen SE, Bokmand S, Nevanlinna H, Mattson J, Friedman E, Laitman Y, Palli D, Masala G, Zanna I, Ottini L, Giannini G, Hollestelle A, Ouweland AM, Novakovic S, Krajc M, Gago-Dominguez M, Castelao JE, Olsson H, Hedenfalk I, Easton DF, Pharoah PD, Dunning AM, Bishop DT, Neuhausen SL, Steele L, Houlston RS, Garcia-Closas M, Ashworth A, Swerdlow AJ. 2012. Genome-wide association study identifies a common variant in RAD51B associated with male breast cancer risk. *Nature Genetics* **44**:1182–1184 DOI 10.1038/ng.2417.

Penault-Llorca F, Viale G. 2012. Pathological and molecular diagnosis of triple-negative breast cancer: a clinical perspective. *Annals of Oncology* **23(Suppl 6)**:vi19–vi22 DOI 10.1093/annonc/mds190.

Pepke S, Wold B, Mortazavi A. 2009. Computation for ChIP-seq and RNA-seq studies. *Nature Methods* **6**:S22–S32 DOI 10.1038/nmeth.1371.

Salehnia M, Zavareh S. 2013. The effects of progesterone on oocyte maturation and embryo development. *International Journal of Fertility & Sterility* **7**:74–81.

Schmidt D, Schwalie PC, Ross-Innes CS, Hurtado A, Brown GD, Carroll JS, Flicek P, Odom DT. 2010. A CTCF-independent role for cohesin in tissue-specific transcription. *Genome Research* **20**:578–588 DOI 10.1101/gr.100479.109.

Shao R, Cao S, Wang X, Feng Y, Billig H. 2014. The elusive and controversial roles of estrogen and progesterone receptors in human endometriosis. *American Journal of Translational Research* **6**:104–113.

Ström A, Hartman J, Foster JS, Kietz S, Wimalasena J, Gustafsson J-Å. 2004. Estrogen receptor β inhibits 17β-estradiol-stimulated proliferation of the breast cancer cell line T47D. *Proceedings of the National Academy of Sciences of the United States of America* **101**:1566–1571 DOI 10.1073/pnas.0308319100.

Tsai MJ, O'Malley BW. 1994. Molecular mechanisms of action of steroid/thyroid receptor superfamily members. *Annual Review of Biochemistry* **63**:451–486 DOI 10.1146/annurev.bi.63.070194.002315.

Tsai WW, Wang Z, Yiu TT, Akdemir KC, Xia W, Winter S, Tsai CY, Shi X, Schwarzer D, Plunkett W, Aronow B, Gozani O, Fischle W, Hung MC, Patel DJ, Barton MC. 2010. TRIM24 links a non-canonical histone signature to breast cancer. *Nature* **468**:927–932 DOI 10.1038/nature09542.

Wang L, Di LJ. 2014. BRCA1 and estrogen/estrogen receptor in breast cancer: where they interact? *International Journal of Biological Sciences* **10**:566–575 DOI 10.7150/ijbs.8579.

Wang D, Garcia-Bassets I, Benner C, Li W, Su X, Zhou Y, Qiu J, Liu W, Kaikkonen MU, Ohgi KA, Glass CK, Rosenfeld MG, Fu XD. 2011b. Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. *Nature* **474**:390–394 DOI 10.1038/nature10006.

Wang C, Mayer JA, Mazumdar A, Fertuck K, Kim H, Brown M, Brown PH. 2011a. Estrogen induces c-myc gene expression via an upstream enhancer activated by the estrogen receptor and the AP-1 transcription factor. *Molecular Endocrinology* **25**:1527–1538 DOI 10.1210/me.2011-1037.

Welboren WJ, Van Driel MA, Janssen-Megens EM, Van Heeringen SJ, Sweep FC, Span PN, Stunnenberg HG. 2009. ChIP-Seq of ERalpha and RNA polymerase II defines genes differentially responding to ligands. *EMBO Journal* **28**:1418–1428 DOI 10.1038/emboj.2009.88.

Wilbanks EG, Facciotti MT. 2010. Evaluation of algorithm performance in ChIP-seq peak detection. *PLoS ONE* **5**:e11471 DOI 10.1371/journal.pone.0011471.

Woodfield GW, Chen Y, Bair TB, Domann FE, Weigel RJ. 2010. Identification of primary gene targets of TFAP2C in hormone responsive breast carcinoma cells. *Genes Chromosomes Cancer* **49**:948–962 DOI 10.1002/gcc.20807.

Xia Y, Chang T, Wang Y, Liu Y, Li W, Li M, Fan HY. 2014. YAP promotes ovarian cancer cell tumorigenesis and is indicative of a poor prognosis for ovarian cancer patients. *PLoS ONE* **9**:e91770 DOI 10.1371/journal.pone.0091770.

Xie W, Schultz MD, Lister R, Hou Z, Rajagopal N, Ray P, Whitaker JW, Tian S, Hawkins RD, Leung D, Yang H, Wang T, Lee AY, Swanson SA, Zhang J, Zhu Y, Kim A, Nery JR, Urich MA, Kuan S, Yen CA, Klugman S, Yu P, Suknuntha K, Propson NE, Chen H, Edsall LE, Wagner U, Li Y, Ye Z, Kulkarni A, Xuan Z, Chung WY, Chi NC, Antosiewicz-Bourget JE, Slukvin I, Stewart R, Zhang MQ, Wang W, Thomson JA, Ecker JR, Ren B. 2013. Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**:1134–1148 DOI 10.1016/j.cell.2013.04.022.

Yadav BS, Sharma SC, Chanana P, Jhamb S. 2014. Systemic treatment strategies for triple-negative breast cancer. *World Journal of Clinical Oncology* **5**:125–133 DOI 10.5306/wjco.v5.i2.125.

Yin P, Roqueiro D, Huang L, Owen JK, Xie A, Navarro A, Monsivais D, Coon JSt, Kim JJ, Dai Y, Bulun SE. 2012. Genome-wide progesterone receptor binding: cell type-specific and shared mechanisms in T47D breast cancer cells and primary leiomyoma cells. *PLoS ONE* **7**:e29021 DOI 10.1371/journal.pone.0029021.

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, Liu XS. 2008. Model-based analysis of ChIP-Seq (MACS). *Genome Biology* **9**:R137 DOI 10.1186/gb-2008-9-9-r137.