

CORRESPONDENCE

Open Access

# Non-canonical protein-DNA interactions identified by ChIP are not artifacts

Richard P Bonocora<sup>1</sup>, Devon M Fitzgerald<sup>2</sup>, Anne M Stringer<sup>1</sup> and Joseph T Wade<sup>1,2\*</sup>

## Abstract

**Background:** ChIP-chip and ChIP-seq are widely used methods to map protein-DNA interactions on a genomic scale *in vivo*. Waldminghaus and Skarstad recently reported, in this journal, a modified method for ChIP-chip. Based on a comparison of our previously-published ChIP-chip data for *Escherichia coli*  $\sigma^{32}$  with their own data, Waldminghaus and Skarstad concluded that many of the  $\sigma^{32}$  targets identified in our earlier work are false positives. In particular, we identified many non-canonical  $\sigma^{32}$  targets that are located inside genes or are associated with genes that show no detectable regulation by  $\sigma^{32}$ . Waldminghaus and Skarstad propose that such non-canonical sites are artifacts, identified due to flaws in the standard ChIP methodology. Waldminghaus and Skarstad suggest specific changes to the standard ChIP procedure that reportedly eliminate the claimed artifacts.

**Results:** We reanalyzed our published ChIP-chip datasets for  $\sigma^{32}$  and the datasets generated by Waldminghaus and Skarstad to assess data quality and reproducibility. We also performed targeted ChIP/qPCR for  $\sigma^{32}$  and an unrelated transcription factor, AraC, using the standard ChIP method and the modified ChIP method proposed by Waldminghaus and Skarstad. Furthermore, we determined the association of core RNA polymerase with disputed  $\sigma^{32}$  promoters, with and without overexpression of  $\sigma^{32}$ . We show that (i) our published  $\sigma^{32}$  ChIP-chip datasets have a consistently higher dynamic range than those of Waldminghaus and Skarstad, (ii) our published  $\sigma^{32}$  ChIP-chip datasets are highly reproducible, whereas those of Waldminghaus and Skarstad are not, (iii) non-canonical  $\sigma^{32}$  target regions are enriched in a  $\sigma^{32}$  ChIP in a heat shock-dependent manner, regardless of the ChIP method used, (iv) association of core RNA polymerase with some disputed  $\sigma^{32}$  target genes is induced by overexpression of  $\sigma^{32}$ , (v)  $\sigma^{32}$  targets disputed by Waldminghaus and Skarstad are predominantly those that are most weakly bound, and (vi) the modifications to the ChIP method proposed by Waldminghaus and Skarstad reduce enrichment of all protein-bound genomic regions.

**Conclusions:** The modifications to the ChIP-chip method suggested by Waldminghaus and Skarstad reduce rather than increase the quality of ChIP data. Hence, the non-canonical  $\sigma^{32}$  targets identified in our previous study are likely to be genuine. We propose that the failure of Waldminghaus and Skarstad to identify many of these  $\sigma^{32}$  targets is due predominantly to the lower data quality in their study. We conclude that surprising ChIP-chip results are not artifacts to be ignored, but rather indications that our understanding of DNA-binding proteins is incomplete.

**Keywords:** ChIP-chip, ChIP-seq,  $\sigma^{32}$

\* Correspondence: [jwade@wadsworth.org](mailto:jwade@wadsworth.org)

<sup>1</sup>Wadsworth Center, New York State Department of Health, Albany, NY 12208, USA

<sup>2</sup>Department of Biomedical Sciences, University at Albany, Albany, NY 12201, USA

## Background

ChIP-chip (sometimes referred to as ChIP-on-chip) and ChIP-seq are widely-used genomic methods that combine chromatin immunoprecipitation (ChIP) with microarrays and deep sequencing, respectively, to map protein-DNA interactions *in vivo* [1]. The genome-wide binding profiles of hundreds of proteins have been mapped using ChIP-chip and ChIP-seq in organisms ranging from bacteria to humans. ChIP-chip/ChIP-seq often identifies non-canonical target regions for DNA-associated proteins, i.e. target regions that are inconsistent with our current understanding of the protein being studied. In many cases, these discoveries have provided new insight into the function of those proteins. In bacteria, many transcription factor (TF) binding sites identified using ChIP-chip/ChIP-seq are located in “unexpected” genomic regions: (i) upstream of genes whose described function is seemingly unconnected to the described function of the TF [2-4], (ii) upstream of genes whose expression does not change detectably when the TF-encoding gene is mutated [2,4-8], (iii) inside genes [2-4,9-13], and (iv) far from any DNA sequences that are close matches to the known consensus binding site [2,3,8,14,15]. In most cases, the significance of these observations is unclear, although they suggest that (i) gene annotations are often incomplete, (ii) TFs often function redundantly, such that expression of the regulated gene does not change unless multiple TF-encoding genes are deleted, (iii) TFs often regulate the expression of non-coding RNAs that initiate within genes [16], and (iv) TFs often bind DNA cooperatively such that the DNA sequence requirements are altered or relaxed.

Our published ChIP-chip study of  $\sigma^{32}$ , an alternative  $\sigma$  factor in *E. coli*, led to the identification of 22 putative  $\sigma^{32}$  binding sites within genes [11]. These represent ~25% of all the  $\sigma^{32}$  binding sites we identified. All but 2 of the gene-internal promoters are >300 bp from an annotated translation start codon. We proposed that RNA polymerase (RNAP) associated with  $\sigma^{32}$  (RNAP: $\sigma^{32}$ ) often binds to promoter elements within genes and initiates transcription of non-coding RNAs in either the sense or antisense orientation. We confirmed this for three examples that we examined in more detail. Furthermore, five of the  $\sigma^{32}$  binding sites within genes are immediately adjacent to genes identified in previous studies as being upregulated by  $\sigma^{32}$ , but for which no promoter could be identified in the upstream region [17,18]. Our ChIP-chip data also permitted identification of 65  $\sigma^{32}$  binding sites in intergenic regions, 26 of which are not associated with genes identified in either of two transcriptomic studies of  $\sigma^{32}$  [17,18]. Thus, many of the sites of  $\sigma^{32}$  association we identified are non-canonical.

In a recent study published in this journal, Waldminghaus and Skarstad describe modifications to the standard ChIP-chip procedure [19]. The key modifications are

avoiding the use of Spin-X filter columns during immunoprecipitation (IP) wash steps, including an RNase treatment following the IP, and collecting reference material after the IP rather than the traditional “input” starting chromatin. Waldminghaus and Skarstad propose that the standard ChIP-chip method results in identification of false positives that are eliminated when using the modified method. Waldminghaus and Skarstad demonstrated their modified ChIP-chip procedure by performing ChIP-chip of *E. coli*  $\sigma^{32}$ . They identified many fewer target regions for  $\sigma^{32}$  than our earlier study. We will refer to the 46  $\sigma^{32}$  target regions identified in our previous study but not by Waldminghaus and Skarstad as “Disputed  $\sigma^{32}$  targets” (DSTs). DSTs are enriched for non-canonical  $\sigma^{32}$  binding sites. Specifically, 16 of the 46 DSTs are located inside genes or between convergently transcribed genes, and 21 DSTs are located in intergenic regions but are not associated with genes identified in transcriptomic studies of  $\sigma^{32}$  [17,18]. We have reanalyzed our published ChIP-chip datasets and those of Waldminghaus and Skarstad. This reanalysis demonstrates low reproducibility in the datasets of Waldminghaus and Skarstad. We also used targeted ChIP/qPCR to directly compare the standard and modified ChIP methods. We demonstrate that non-canonical targets of  $\sigma^{32}$  are real and that the lower data quality and deficiencies in the modified ChIP method are sufficient to explain the absence of DSTs in the list of  $\sigma^{32}$  targets generated by Waldminghaus and Skarstad.

## Results and discussion

### Existing evidence that DSTs are genuine sites of $\sigma^{32}$ association

Waldminghaus and Skarstad suggest that DSTs are artifacts that result from non-specific IP of RNA that is then amplified by Klenow DNA polymerase during sample preparation for ChIP-chip [19]. However, there are several features of DSTs that are consistent with them being genuine sites of  $\sigma^{32}$  association and inconsistent with them being artifacts resulting from amplification from RNA:

- (i) Nine of the DSTs (*mfd*, *phoP*, *ldhA*, *recF*, *narP*, *holC*, *glnS*, *ileS*, and *yffN*) are  $\sigma^{32}$  targets identified in independent studies that did not involve ChIP [17,18]. With the exception of the DSTs inside *yffN* and *recF*, these would all be considered canonical  $\sigma^{32}$  binding sites, i.e. located in an intergenic region upstream of a gene known from previous studies to be transcribed by  $\sigma^{32}$  [17,18].
- (ii) Our previous study included validation of three non-canonical DSTs (between *tdk* and *yehG*, within *dhaM*, and within *ydeP*) using ChIP/qPCR [11]. This method does not involve amplification of ChIP DNA

using Klenow DNA polymerase. Furthermore, we demonstrated heat shock-dependent increases of  $\sigma^{32}$  association with all three regions [11].

- (iii) Although many DSTs are located inside genes, there are significantly more DSTs located in intergenic regions than expected by chance (Binomial Test  $p = 0.00033$ ).

Note that, for all the analyses described herein, we have excluded the two DSTs that are located in repetitive sequence (*yibA* and *yrdA*; see Conclusions).

#### Comparison of data quality between our data and those of Waldminghaus and Skarstad

The disparity between the  $\sigma^{32}$  targets identified in the two studies led us to compare the quality of the ChIP-chip data. For each dataset we used an established method to estimate the null distribution of ChIP-chip signals [20,21]. Specifically, we determined the modal value and used the probes with scores at or below this value to fit a normal distribution. Using this fitted normal distribution we determined the mean and standard deviation of the null distribution. This allowed us to calculate z-scores (number of standard deviations from the mean) for each microarray probe, thus providing a measure of dynamic range that is independent of the absolute ChIP-chip signals, which have arbitrary units. Scatter plots of z-scores for the duplicate datasets from each study are shown in Figure 1A-B. These scatter plots demonstrate several key features of the datasets from each study:

- (i) The two replicate datasets for our study correlate very well (Spearman Correlation Coefficient of 0.93) whereas those of Waldminghaus and Skarstad correlate less well (Spearman Correlation Coefficient of 0.64).
- (ii) One of the datasets of Waldminghaus and Skarstad has a substantially lower dynamic range than the other. Several of the targets identified in both studies have z-scores within the noise for this replicate.
- (iii) Although the dynamic range of one Waldminghaus and Skarstad dataset is high, the vast majority (~98.5%) of the probes have z-scores lower than 3, suggesting that these datasets are effective at identifying strong protein-DNA interactions but not weaker interactions.
- (iv) Although they were not called as targets, DSTs have significantly higher z-scores for the datasets of Waldminghaus and Skarstad than expected by chance (Mann Whitney U Test  $p < 1e^{-30}$  for each replicate dataset).

We conclude that our ChIP-chip data are of substantially higher quality with respect to both dynamic range and reproducibility. Figure 1C-H shows normalized ChIP-chip

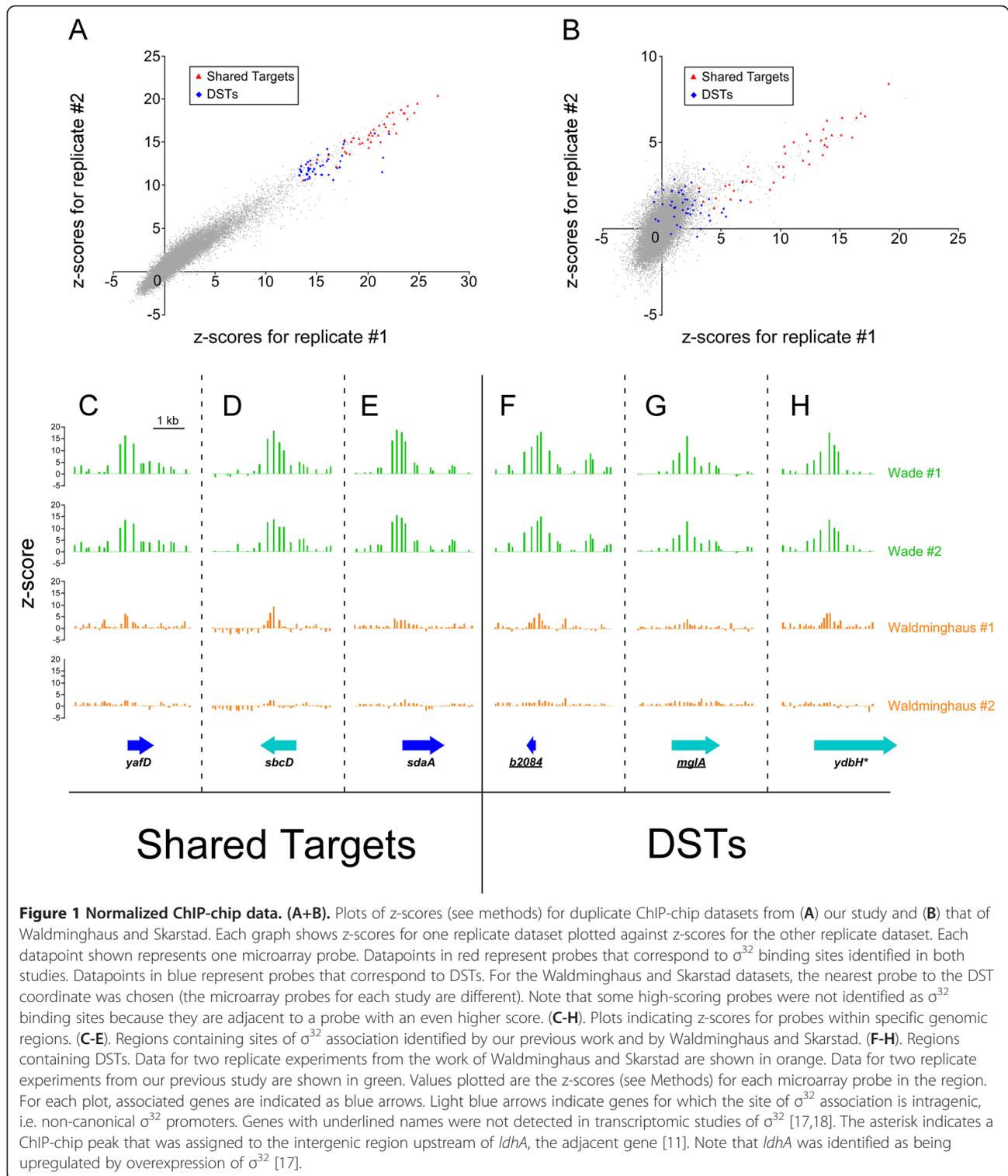
data for replicate datasets from both studies for six selected genomic regions. These data further demonstrate the differences in reproducibility and dynamic range between the two studies. The genomic regions shown include DSTs and non-canonical targets (inside genes and/or no detectable regulation in transcriptomic studies).

Several factors likely contribute to the difference in data quality between the two studies. First, we used a TAP-tagged derivative of  $\sigma^{32}$  whereas Waldminghaus and Skarstad used an antibody raised against the native protein. Second, our heat shock conditions (50°C for 10 minutes) were different to those of Waldminghaus and Skarstad (43°C for 5 minutes). Third, as described below, the modifications to the ChIP method reduce the sensitivity of the assay.

#### ChIP/qPCR validation of DSTs

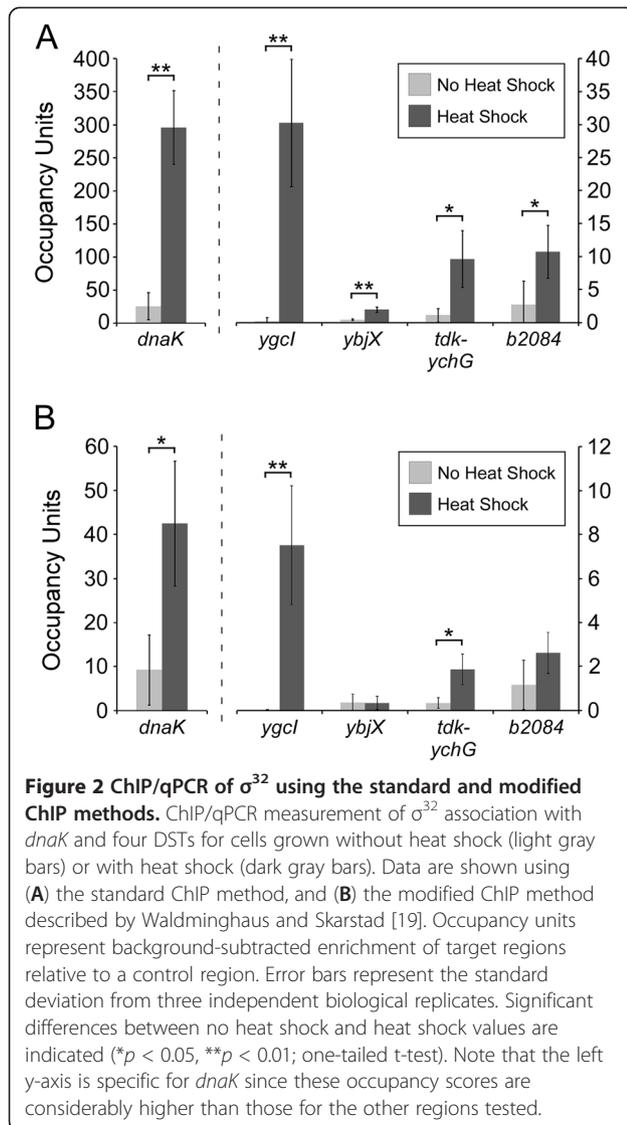
We used ChIP/qPCR with the standard and modified ChIP methods to measure association of  $\sigma^{32}$  with four DSTs in cells before and after heat shock. As a positive control, we measured association of  $\sigma^{32}$  with the region upstream of *dnaK*, a well-established  $\sigma^{32}$  target [17,18] identified both in our study and that of Waldminghaus and Skarstad. We used cells expressing an N-terminally FLAG-tagged copy of  $\sigma^{32}$  expressed from its native locus (our earlier study used a C-terminally TAP-tagged copy of  $\sigma^{32}$ ). Using the standard ChIP method, we observed significant association of  $\sigma^{32}$  with all regions tested and a significant increase in  $\sigma^{32}$  association with all regions tested following heat shock (Figure 2A). Previous ChIP-seq studies have revealed biases in the level of some genomic regions in input DNA, the most common control sample for ChIP experiments [22-24]. In the case of ChIP-chip, this bias is likely to be due to nucleosomes, and is hence specific to eukaryotes [23,24]. Nevertheless, we wished to rule out the possibility that DSTs were identified as a result of input biases. Therefore, we repeated the ChIP/qPCR using an untagged strain. We observed no significant ChIP/qPCR signal for any region tested (Additional file 1: Supplementary Data). We conclude that all four DSTs tested are genuine sites of  $\sigma^{32}$  binding.

We compared the standard ChIP method with the modified method proposed by Waldminghaus and Skarstad. Importantly, ChIP with the modified method used the same sonicated, cross-linked cell extracts as the standard method. Using the modified method, we detected significant  $\sigma^{32}$  association with the region upstream of *dnaK* (Figure 2B), and association increased significantly following heat shock (Figure 2B). However, the absolute ChIP signal was substantially lower than that observed using the standard ChIP method (Figure 2A). Thus, the modified ChIP method has a decreased sensitivity relative to the standard method. Using the modified ChIP method we detected significant association of  $\sigma^{32}$  following heat



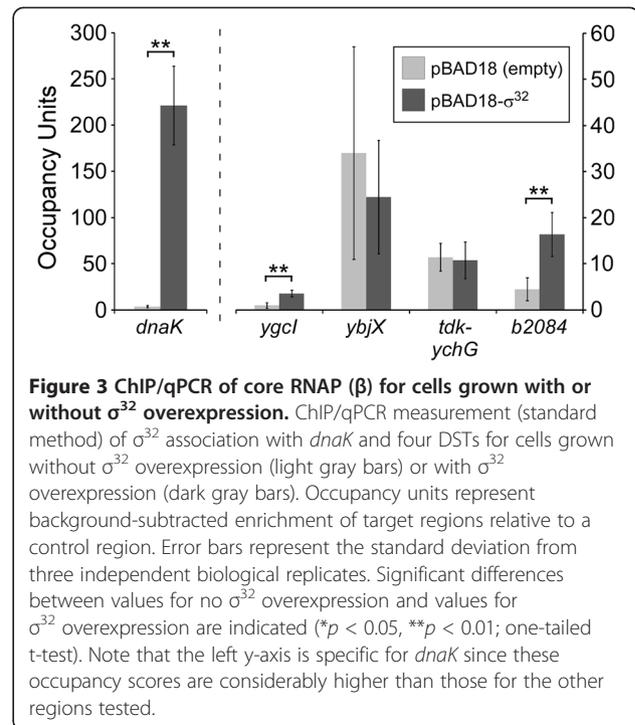
shock with three of the four DSTs tested (Figure 2B). We also observed a significant reduction in  $\sigma^{32}$  association in the absence of heat shock at two of these DSTs (Figure 2B). Thus, even with the decreased sensitivity of the modified ChIP method, three of the four DSTs tested were validated

as genuine sites of  $\sigma^{32}$  association. We believe that we were unable to detect significant association of  $\sigma^{32}$  with the fourth DST, *ybjX*, due to the substantial decrease in sensitivity relative to the standard ChIP method. We note that the ChIP signal for *ybjX* was the lowest of all the



regions tested using the standard method (Figure 2A). We conclude that the reduced sensitivity of the modified ChIP method prevented Waldminghaus and Skarstad from identifying DSTs as sites of  $\sigma^{32}$  association. This is consistent with the observation that DSTs have above average ChIP-chip scores in the Waldminghaus and Skarstad datasets (Figure 1B).

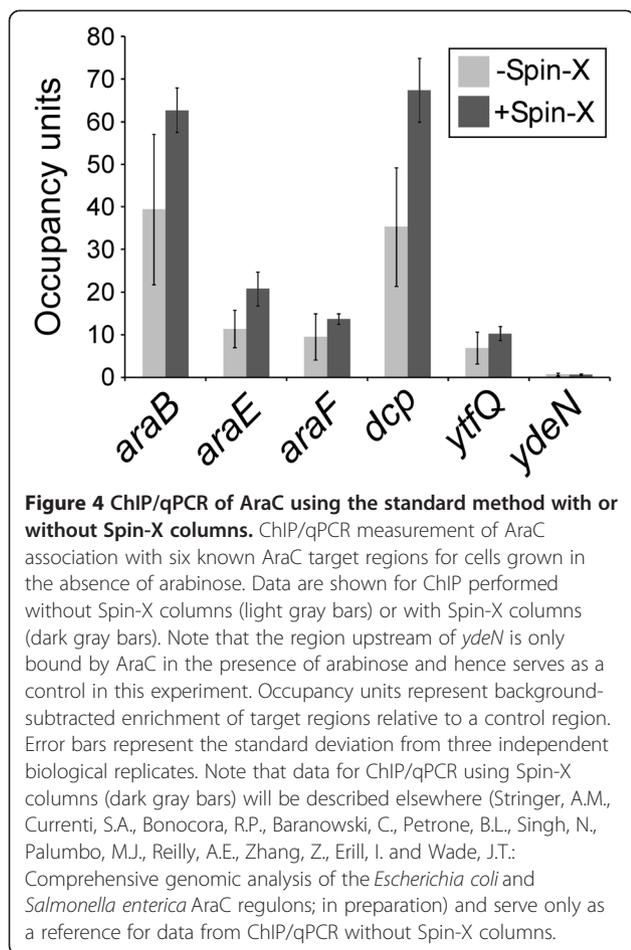
As an independent assessment of  $\sigma^{32}$  association with DSTs, we measured association of core RNAP ( $\beta$  subunit) with *dnaK* and the four DSTs described above, with and without overexpression of  $\sigma^{32}$  from a plasmid. Association of  $\beta$  with *dnaK* and two DSTs was significantly higher in cells overexpressing  $\sigma^{32}$  as compared to those with empty vector (Figure 3). This provides independent validation of the association of  $\sigma^{32}$  with these regions. Two of the DSTs tested showed no significant difference in the association of  $\beta$  between cells overexpressing  $\sigma^{32}$  and those with empty vector. In the case of *ybjX*, we propose that the lack



of increase in RNAP levels is due to the relatively low association of  $\sigma^{32}$  (Figure 2A). Thus, association of RNAP: $\sigma^{32}$  may not significantly increase the overall association of RNAP in the presence of a relatively high level of RNAP that is independent of  $\sigma^{32}$  (presumably RNAP: $\sigma^{70}$ ). Consistent with our ChIP/qPCR data, *ybjX* expression was not detectably increased by  $\sigma^{32}$  overexpression in two transcriptomic studies [17,18]. In the case of *tdk/ychG*, we propose that RNAP: $\sigma^{32}$  binds this region specifically during heat shock but not following  $\sigma^{32}$  over-expression without heat shock, perhaps due to the requirement for other heat shock-induced/activated proteins.

#### ChIP method comparison for AraC

The comparison of the ChIP methodologies described above demonstrates that the modified ChIP method is less sensitive. There are multiple changes to the standard method, so it is unclear which specific change(s) results in the decreased sensitivity. One significant change in the method described by Waldminghaus and Skarstad is the omission of Spin-X columns during the IP wash steps. We directly assessed the importance of Spin-X columns by measuring association of AraC (C-terminally FLAG-tagged) with target regions in *E. coli* using ChIP/qPCR performed either with or without Spin-X columns. The use of Spin-X columns increased the ChIP/qPCR signal for all regions tested but qualitatively the data are the same for both methods (Figure 4). Importantly, we detected association of AraC with a non-canonical target



within the *dcp* gene using both methods (Figure 4). This site of AraC association is hundreds of base pairs from either end of the gene and there is no detectable change in transcription of *dcp* or association of RNAP at this region following deletion of *araC* and/or addition of arabinose (Stringer, A.M., Currenti, S.A., Bonocora, R.P., Baranowski, C., Petrone, B.L., Singh, N., Palumbo, M.J., Reilly, A.E., Zhang, Z., Erill, I. and Wade, J.T.: Comprehensive genomic analysis of the *Escherichia coli* and *Salmonella enterica* AraC regulons; in preparation). Thus, the Spin-X column-free ChIP method detects association with non-canonical target regions, although association with all target regions is reduced relative to the standard ChIP method. In a control experiment using an untagged strain, we observed no significant ChIP/qPCR signal (using the standard ChIP method) for any region tested (Additional file 1: Supplementary Data).

## Conclusions

We conclude that Waldminghaus and Skarstad failed to identify DSTs not because of an improvement in the ChIP methodology, but because of lower data quality. Consistent with this, the majority of DSTs showed

relatively low association of  $\sigma^{32}$  in our study: when ranked by the level of  $\sigma^{32}$  association, 36 of the bottom 43 targets are DSTs (Figure 1A) [11]. Furthermore, DSTs have significantly higher signal in the Waldminghaus and Skarstad datasets than expected by chance ( $p < 1e^{-30}$ ; Figure 1B), consistent with the idea that these regions represent true binding sites for  $\sigma^{32}$  but fall below the detection threshold of this analysis. We note that Waldminghaus and Skarstad did not present any  $\sigma^{32}$  ChIP data generated using the standard methodology, precluding direct comparison of our work, nor did they use ChIP/qPCR with their modified method to measure association of  $\sigma^{32}$  with specific target regions [19]. Furthermore, Waldminghaus and Skarstad demonstrated a dramatic improvement in ChIP-chip data for SeqA using the modified ChIP method [19], but their data is very similar to that generated using the standard ChIP method by another group [25].

Our comparison of ChIP-chip datasets highlights the importance of data quality for correct identification of protein-DNA interactions. Guidelines for ChIP-chip and ChIP-seq experimental and analytical approaches have been described previously [26,27]. Key components of these methods that are especially relevant to our own study are the comparison of replicates, the choice of control, and the importance of repetitive sequence. Current guidelines for ChIP-seq recommend the use of only two independent biological replicates [27], but also stress the importance of reproducibility. As shown in Figure 1B, the poor reproducibility of the Waldminghaus and Skarstad datasets is likely to be a major cause of their failure to identify DSTs as regions truly bound by  $\sigma^{32}$ . Recommended controls are either input DNA or ChIP-enriched DNA from an untagged strain (when using an epitope-tagged protein). Waldminghaus and Skarstad instead used DNA left in the supernatant after the initial IP, acknowledging that this DNA would be de-enriched for target regions. While this may increase the apparent signal, we caution against this approach as the ChIP-chip or ChIP-seq signals may not accurately reflect the actual level of binding. Finally, Waldminghaus and Skarstad highlighted the importance of treating repetitive DNA sequences with caution when interpreting ChIP-chip (or ChIP-seq) datasets. In the case of  $\sigma^{32}$ , two of the ChIP peaks identified in our earlier study overlap repetitive regions. It is impossible to determine from ChIP-chip data alone whether  $\sigma^{32}$  associates with one or all of the repetitive regions. Since this caveat applies to repetitive sequences in any ChIP-chip or ChIP-seq experiment, we echo the sentiment expressed by Waldminghaus and Skarstad and caution against analysis of sequences in these regions.

Many ChIP-chip studies have revealed the existence of unexpected protein-DNA interactions. For example,

ChIP-chip studies in bacteria have demonstrated that transcription factors often bind to sites within genes, sites without a recognizable motif, and sites that are not associated with described regulation by the transcription factor [15]. This is one of the great strengths of ChIP-chip and ChIP-seq, since these non-canonical binding sites often cannot be identified using other genomic approaches such as transcription profiling. In the case of  $\sigma^{32}$ , our data provide strong evidence that RNAP: $\sigma^{32}$  initiates transcription of many RNAs from within genes, and our original study described three such examples in greater detail [11]. The function of intragenic transcripts in bacteria is poorly understood, although several antisense transcripts have been shown previously to regulate expression of the overlapping mRNA [28]. Our own studies have revealed pervasive antisense transcription in *E. coli* [16], and this has since been observed in several other bacterial species [28]. Intriguingly, many ChIP-chip studies of bacterial DNA-binding TFs have revealed sites of association inside genes [10,15], suggesting regulation of intragenic transcripts. Similar phenomena have been observed in eukaryotes, including human cells [29,30]. Other types of non-canonical transcription factor binding sites, i.e. sites without a recognizable motif and sites that are not associated with described regulation by the transcription factor, are also poorly understood. However, sites without a recognizable motif could be explained by indirect association with DNA (detectable using ChIP) [15,31] or cooperative interactions with other DNA-binding proteins [32]. Sites that are not associated with described regulation by the transcription factor could be explained by combinatorial regulation by multiple, redundant transcription factors. In the case of  $\sigma^{32}$ , our data suggest that many  $\sigma^{32}$  promoters are not associated with detectable regulation using transcriptomic approaches due to a high basal level of transcription, or a specific requirement for heat shock conditions.

It is important to note that Waldminghaus and Skarstad identified many non-canonical  $\sigma^{32}$ -target regions in their study. Specifically, Waldminghaus and Skarstad detected  $\sigma^{32}$  association upstream of four genes whose expression was not detectably upregulated by overexpression of  $\sigma^{32}$  in either of two transcriptomic studies (*yafU*, *rpsL*, *yjhI*, and *fimB*) [17,18], and six sites of  $\sigma^{32}$  association within genes or between convergently transcribed genes (*yfbM/yfbN*, *yffU*, *ypjA*, *sbcD*, *cycA*, and *macB*) [19]. Waldminghaus and Skarstad suggest that “surprising”, non-canonical protein-DNA interactions are often artifacts. We caution against this dogmatic approach. Artifacts can arise from ChIP-chip and ChIP-seq experiments; however, with the appropriate experimental and analytical methods, and with the appropriate controls, it is possible to identify protein-DNA interactions with high confidence. Atypical binding

sites identified using these methods may indicate novel functions for well-studied proteins. These binding sites should not be dismissed, but rather should be the focus of additional studies.

## Methods

### Strains and plasmids

*E. coli* MG1655 *rpoH*-NFLAG containing the *rpoH* gene at its native chromosomal location fused to three FLAG tags was constructed using FRUIT [33]. Primer sequences are available on request. Construction of MG1655 with C-terminally FLAG-tagged AraC (AMD187) will be described elsewhere (Stringer, A.M., Currenti, S.A., Bonocora, R.P., Baranowski, C., Petrone, B.L., Singh, N., Palumbo, M.J., Reilly, A.E., Zhang, Z., Erill, I. and Wade, J.T.: Comprehensive genomic analysis of the *Escherichia coli* and *Salmonella enterica* AraC regulons; in preparation).

pRB1 for expression of the *rpoH* gene ( $\sigma^{32}$ ) was constructed by PCR amplification from chromosomal DNA with primers JW2199 and JW2200 (Table 1). The

**Table 1 List of oligonucleotides used in this work**

Name	Sequence
JW071	ACCAAAGCCATGACAAAA
JW072	TGGCATAGCAAAGTGTGA
JW073	CGCAGCAATTTAATCCAT
JW074	CCTGCCAGCAGAGAGTAA
JW075	TGCGATGTGATATTGCTC
JW076	TAGGGCAAAAACGAATGA
JW125	AAGCGAAAAATCGGCAATA
JW126	CATGGCCTGCAACATATC
JW389	CGCGAACATCTTTAACC
JW390	CAACGCCATAGACGACAT
JW393	GTCAACGCTTTATGGACTG
JW394	ACGAAGAGAAAATAAGTGATGT
JW1312	TTCGCTGTACCTCTGGAA
JW1313	TTGGCCGATGATGGTTAT
JW1610	ACGTTTCGCCCTATTAC
JW1611	ACCCAGGTGCATACCAAT
JW1612	GGTAGGTCGCGTTCCTT
JW1613	GAGGCGTGAACATGAGAT
JW1614	GCGTCGATTTCCACATT
JW1615	GGTTTTCCGCTTTTTTCAT
JW1616	ATACCCTTCTCGGCAGTT
JW1617	CGCCTAAACACAGGGATA
JW1622	ACACCTTCTGTCTTCAGCTC
JW1623	GCCAGGGGTAGAATATCTG
JW2199	CTAGGCTAGCGAGAGGATTTGAATGACTGAC
JW2200	CTAGGCATGCTTACGCTTCAATGGCAGCAC

PCR product was digested with *NheI* and *SphI* and ligated into similarly digested pBAD18-Cm [34].

#### Cell growth

For heat shock ChIP experiments, 100 ml LB was inoculated with 1 ml of fresh overnight culture of MG1655 *rpoH*-NFLAG and cells were grown at 30°C at 225 rpm to an OD<sub>600</sub> of 0.5-0.6. Cultures were split (40 ml each) for further incubation at either 30°C or 50°C for 10 - minutes. For ChIP experiments involving overexpression of  $\sigma^{32}$ , 40 ml LB supplemented with 30 µg/ml chloramphenicol was inoculated with 0.4 ml of a fresh overnight culture of MG1655 containing either pRB1 or pBAD18-Cm. Cells were grown at 37°C at 225 rpm to an OD<sub>600</sub> of 0.7-0.8. Expression of *rpoH* from pRB1 was induced by the addition of 0.2% arabinose and further incubation at 37°C for 10 minutes. For ChIP of AraC, AMD187 was grown in LB at 37°C at 225 rpm to an OD<sub>600</sub> of 0.6-0.8.

#### Standard ChIP method

Cells were crosslinked by the addition of formaldehyde to a final concentration of 1% for 20 minutes. Formaldehyde was quenched with glycine (0.5 M final concentration) and cultures were pelleted by centrifugation. Pellets were washed twice with Tris-buffered saline (TBS; pH 7.5) and resuspended in 1 ml FA lysis buffer (50 mM Hepes-KOH, pH 7, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 0.1% SDS) supplemented with 4 mg/ml lysozyme. After a 30 minute incubation at 37°C, cells were chilled on ice and sonicated in 30 second on/off pulses for 30 minutes at 100% output using a BioRuptor Sonicator. Lysates were centrifuged for five minutes to pellet cell debris. The supernatant was transferred to a new tube, brought up to a final volume of approximately 2 ml, and frozen in 0.5 ml aliquots. 0.5 ml crosslinked, sonicated cell lysate was brought up to a final volume of 0.8 ml with FA lysis buffer. A 20 µl aliquot was removed for "input" DNA control sample. 25 µl of protein A-Sepharose beads (50% slurry in TBS) and either 1 µl anti-RNA polymerase beta subunit (Neoclone) or 2 µl anti-FLAG (M2 monoclonal; Sigma) was added to the lysate and incubated for 90 minutes at room temperature with gentle rotation. Beads were pelleted at 4000 rpm in a microcentrifuge for one minute and the supernatant was removed. Beads were resuspended in 700 µl FA lysis buffer, transferred to a Spin-X column (Corning) and washed for three minutes by rotation, centrifuged for 1 - minute at 4,000 rpm in a microcentrifuge and the flow through discarded. The beads were washed in a similar fashion with 750 µl of each of the following: FA lysis buffer, FA lysis buffer 500 mM NaCl, ChIP wash buffer (10 mM Tris-HCl, pH 8.0, 250 mM LiCl, 1 mM EDTA, 0.5% Nonidet-P40, 0.5% sodium deoxycholate) and TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA). The Spin-X

column was transferred to a fresh tube and the chromatin was eluted from the beads by addition of 100 µl ChIP elution buffer (50 mM Tris-HCl, pH 7.5, 10 mM EDTA, 1% SDS) and incubation at 65°C for 10 minutes. The eluate was collected by centrifugation for 1 min at 4,000 rpm in a microcentrifuge. Crosslinks were reversed for both the eluate and the input samples by incubation for 10 - minutes at 100°C. DNA was purified using QIAGEN PCR purification kit followed by elution in either 50 µl or 200 µl for the IP samples or 200 µl for the input samples. For AraC ChIP, Spin-X columns were omitted from this procedure when indicated in the figure. Note that data shown for AraC ChIP/qPCR with Spin-X columns will be presented elsewhere (Stringer, A.M., Currenti, S.A., Bonocora, R.P., Baranowski, C., Petrone, B.L., Singh, N., Palumbo, M.J., Reilly, A.E., Zhang, Z., Erill, I. and Wade, J.T.: Comprehensive genomic analysis of the *Escherichia coli* and *Salmonella enterica* AraC regulons; in preparation).

#### Modified ChIP method described by Waldminghaus and Skarstad

ChIP was performed as above but with the following modifications: (i) 100 µl of post-immunoprecipitation supernatant was substituted for the "input" control DNA sample, (ii) no Spin-X columns were used, (iii) 1 µl RNase A (30 mg/ml) was added after elution and incubated for 2 hours at 42°C for both the input and immunoprecipitated DNA samples, (iv) 80 µl TE and 20 µl proteinase K (20 mg/ml) was added incubated for 2 hours at 42°C, (v) crosslinks were reversed by incubation overnight at 65°C, and (vi) DNA was purified by phenol/chloroform/isoamyl alcohol and chloroform/isoamyl alcohol extraction followed by ethanol precipitation. Note that aliquots from the same sonicated, crosslinked cell extract were used for both the standard and modified ChIP methods.

#### qPCR

ChIP and input samples were analyzed by quantitative real time PCR using an ABI 7500 Fast real time PCR machine, as described previously [2]. Enrichment of ChIP samples was calculated relative to a control region within the transcriptionally silent *bglB* gene, and normalized to input DNA. Occupancy units represent background-subtracted fold-enrichment. Oligonucleotides used for real time PCR were JW125/JW126 (*bglB*), JW1610/JW1611 (*dnaK*), JW1612/JW1613 (*ygcl*), JW1614/JW1615 (*ybjX*), JW1616/JW1617 (*tdk-ychG*), JW1622/JW1623 (*b2084*), JW071/JW072 (*araB*), JW073/JW074 (*araE*), JW075/JW076 (*araF*), JW389/JW390 (*ytfQ*), JW1312/JW1313 (*dcp*), and JW393/JW394 (*ydeN*; Table 1). Note that primers for *ytfQ* produced primer dimers in qPCR for

ChIP with an untagged strain (Additional file 1: Supplementary Data), so we were not able to assess enrichment of this region.

### Estimating null distributions for ChIP-chip datasets to calculate z-scores

Previous studies have analyzed ChIP-chip datasets based on the assumption that the distribution of actual ChIP-chip signals below the modal value closely matches the null distribution, and fits a normal distribution [20,21]. We determined the modal value for each ChIP-chip dataset and used all probes scoring below the mode to estimate the standard deviation of a null distribution, treating the mode as the mean. We used these mean and standard deviation estimates to calculate z-scores (i.e. number of standard deviations from the mean) for each probe.

**Assessment of the number of DSTs in intergenic regions**  
88% of the *E. coli* genome is genic. Of the 46 DSTs, 15 have peak probe coordinates that fall in intergenic regions. Note that some additional DSTs were classified as being “intergenic” due to the stringent criterion used in our earlier work [11] to account for incomplete probe coverage on the microarray. We used a Binomial Test to determine the probability that 15 of 46 DSTs would be located in intergenic regions if their genomic position was unbiased with respect to genes.

### Comparison of DST z-scores to those of all z-scores for waldminghaus and skarstad datasets

For each replicate dataset, we determined the z-score for each DST peak probe. We then determined z-scores for 1,000 randomly-selected probes from the complete dataset. We used a Mann–Whitney U Test to determine the probability that the z-scores for DST peak probes are not larger than those of randomly-selected probes.

### Additional file

**Additional file 1: Control ChIP/qPCR data using an untagged strain.**

#### Abbreviations

ChIP: Chromatin Immunoprecipitation; TF: Transcription factor; RNAP: RNA Polymerase; IP: Immunoprecipitation; DSTs: Disputed  $\sigma^{32}$  Targets; TBS: Tris-buffered saline.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

RPB performed the experiments described in Figure 2 and the Additional file 1: Supplementary Data. DMF performed the experiment described in Figure 3. AMS performed the experiment described in Figure 4. JTW performed all other analyses. JTW wrote the paper with input from RPB and DMF. JTW conceived the study. All authors read and approved the final manuscript.

### Acknowledgements

We thank Todd Gray and David Grainger for comments on the manuscript. We thank Todd Gray, David Grainger, Stephen Busby, Kevin Struhl and Evgeny Nudler for helpful discussions. This work was supported by National Institutes of Health (NIH) Grant 1DP2OD007188. DMF was supported by NIH training grant T32AI055429.

Received: 24 April 2012 Accepted: 1 April 2013

Published: 15 April 2013

### References

1. Park PJ: ChIP-seq: advantages and challenges of a maturing technology. *Nat Rev Genet* 2009, **10**:669–680.
2. Wade JT, Reppas NB, Church GM, Struhl K: Genomic analysis of LexA binding reveals the permissive nature of the *Escherichia coli* genome and identifies unconventional target sites. *Genes Dev* 2005, **19**:2619–2630.
3. Partridge JD, Bodenmiller DM, Humphrys MS, Spiro S: NsrR targets in the *Escherichia coli* genome: new insights into DNA sequence requirements for binding and a role for NsrR in the regulation of motility. *Mol Microbiol* 2009, **73**:680–694.
4. Eichenberger P, Fujita M, Jensen ST, Conlon EM, Rudner DZ, Wang ST, Ferguson C, Haga K, Sato T, Liu JS, et al: The program of gene transcription for a single differentiating cell type during sporulation in *Bacillus subtilis*. *PLoS Biol* 2004, **2**:e328.
5. Danielli A, Roncarati D, Delany I, Chiarini V, Rappuoli R, Scarlato V: In vivo dissection of the *Helicobacter pylori* Fur regulatory circuit by genome-wide location analysis. *J Bacteriol* 2006, **188**:4654–4662.
6. Molle V, Nakaura Y, Shivers RP, Yamaguchi H, Losick R, Fujita Y, Sonenshein AL: Additional targets of the *Bacillus subtilis* global regulator CodY identified by chromatin immunoprecipitation and genome-wide transcript analysis. *J Bacteriol* 2003, **185**(6):1911–1922.
7. Laub MT, Chen SL, Shapiro L, McAdams HH: Genes directly controlled by CtrA, a master regulator of the *Caulobacter* cell cycle. *Proc Natl Acad Sci USA* 2002, **99**(7):4632–4637.
8. Grainger DC, Aiba H, Hurd D, Browning DF, Busby SJ: Transcription factor distribution in *Escherichia coli*: studies with FNR protein. *Nucleic Acids Res* 2007, **35**:269–278.
9. Grainger DC, Hurd D, Harrison M, Holdstock J, Busby SJ: Studies of the distribution of *Escherichia coli* cAMP-receptor protein and RNA polymerase along the *E. coli* chromosome. *Proc Natl Acad Sci USA* 2005, **102**:17693–17698.
10. Shimada T, Ishihama A, Busby SJ, Grainger DC: The *Escherichia coli* RutR transcription factor binds at targets within genes as well as intergenic regions. *Nucleic Acids Res* 2008, **36**:3950–3955.
11. Wade JT, Roa DC, Grainger DC, Hurd D, Busby SJ, Struhl K, Nudler E: Extensive functional overlap between  $\sigma$  factors in *Escherichia coli*. *Nat Struct Mol Biol* 2006, **13**:806–814.
12. Reppas NB, Wade JT, Church G, Struhl K: The transition between transcriptional initiation and elongation in *E. coli* is highly variable and often rate-limiting. *Mol Cell* 2006, **24**:747–757.
13. Tomljenovic-Berube AM, Mulder DT, Whiteside MD, Brinkman FS, Coombes BK: Identification of the regulatory logic controlling *Salmonella* pathoadaptation by the SsrA-SsrB Two-component system. *PLoS Genet* 2010, **6**:e1000875.
14. Molle V, Fujita M, Jensen ST, Eichenberger P, Gonzalez-Pastor JE, Liu JS, Losick R: The Spo0A regulon of *Bacillus subtilis*. *Mol Microbiol* 2003, **50**(5):1683–1701.
15. Wade JT, Struhl K, Busby SJ, Grainger DC: Genomic analysis of protein-DNA interactions in bacteria: insights into transcription and chromosome organization. *Mol Microbiol* 2007, **65**:21–26.
16. Dornenburg JE, DeVita AM, Palumbo MJ, Wade JT: Widespread antisense transcription in *Escherichia coli*. *mbio* 2010, **1**:e00024–00010.
17. Nonaka G, Blankschien M, Herman C, Gross CA, Rhodius VA: Regulon and promoter analysis of the *E. coli* heat shock factor, sigma 32, reveals a multifaceted cellular response to heat stress. *Genes Dev* 2006, **20**:1776–1789.
18. Zhao K, Liu M, Burgess RR: The global transcriptional response of *Escherichia coli* to induced sigma 32 protein involves sigma 32 regulon activation followed by inactivation and degradation of sigma 32 *in vivo*. *J Biol Chem* 2005, **280**:17758–17768.
19. Waldminghaus T, Skarstad K: ChIP on chip: surprising results are often artifacts. *BMC Genomics* 2010, **11**:414.

20. Li XY, MacArthur S, Bourgon R, Nix D, Pollard DA, Iyer VN, Hechmer A, Simirenko L, Stapleton M, Luengo Hendriks CL, *et al*: **Transcription factors bind thousands of active and inactive regions in the *drosophila* blastoderm.** *PLoS Biol* 2008, **6**:e27.
21. Gibbons FD, Proft M, Struhl K, Roth RP: **Chipper: discovering transcription factor targets from chromatin immunoprecipitation microarrays using variance-stabilization.** *Genome Biol* 2005, **6**:R96.
22. Lefrançois P, Euskirchen GM, Auerbach RK, Rozowsky J, Gibson T, Yellman CM, Gerstein M, Snyder M: **Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing.** *BMC Genomics* 2009, **10**:37.
23. Vega VB, Cheung E, Palanisamy N, Sung WK: **Inherent signals in sequencing-based chromatin-ImmunoPrecipitation control libraries.** *PLoS One* 2009, **4**:e5241.
24. Auerbach RK, Euskirchen G, Rozowsky J, Lamarre-Vincent N, Moqtaderi Z, Lefrançois P, Struhl K, Gerstein M, Snyder M: **Mapping accessible chromatin regions using sono-Seq.** *Proc Natl Acad Sci USA* 2009, **106**:14926–14931.
25. Sánchez-Romero MA, Busby SJ, Dyer NP, Ott S, Millard AD, Grainger DC: **Dynamic distribution of SeqA protein across the chromosome of *Escherichia coli* K-12.** *mBio* 2010, **1**:e00012-00010.
26. Buck MJ, Lieb JD: **ChIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments.** *Genomics* 2004, **83**(3):349–360.
27. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, *et al*: **ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia.** *Genome Res* 2012, **22**:1813–1831.
28. Georg J, Hess WR: ***cis*-antisense RNA, another level of gene regulation in bacteria.** *Microbiol Mol Biol Rev* 2011, **75**:286–300.
29. ENCODE pc: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**:57–74.
30. Cawley S, Bekiranov S, Ng HH, Kapranov P, Sekinger EA, Kampa D, Piccolboni A, Smentchenko V, Cheng J, Williams AJ, *et al*: **Unbiased mapping of transcription factor binding sites along human chromosomes 21 and 22 points to widespread regulation of non-coding RNAs.** *Cell* 2004, **116**:499–509.
31. Neph S, Vierstra J, Stergachis AB, Reynolds AP, Haugen E, Vernot B, Thurman RE, John S, Sandstrom R, Johnson A, *et al*: **An expansive human regulatory lexicon encoded in transcription factor footprints.** *Nature* 2012, **489**:83–90.
32. Belyaeva TA, Wade JT, Webster CL, Howard VJ, Thomas MS, Hyde EI, Busby SJ: **Transcription activation at the *Escherichia coli* melAB promoter: the role of MelR and the cyclic AMP receptor protein.** *Mol Microbiol* 2000, **36**:211–222.
33. Stringer AM, Singh N, Yermakova A, Petrone BL, Amarasinghe JJ, Reyes-Diaz L, Mantis NJ, Wade JT: **FRUIT, a scar-free system for targeted chromosomal mutagenesis, epitope tagging, and promoter replacement in *Escherichia coli* and *salmonella enterica*.** *PLoS One* 2012, **7**:e44841.
34. Guzman L-M, Belin D, Carson MJ, Beckwith JR: **Tight regulation, modulation, and high-level expression by vectors containing the arabinose P<sub>BAD</sub> promoter.** *J Bacteriol* 1995, **177**:4121–4130.

doi:10.1186/1471-2164-14-254

**Cite this article as:** Bonocora *et al*: Non-canonical protein-DNA interactions identified by ChIP are not artifacts. *BMC Genomics* 2013 **14**:254.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

