|  |
| --- |
|  |
| **Supplementary Figure 1** |
| FLR-Exp results between different cell type pairs. |
| Correlation between Kolmogorov-Smirnov (KS) test statistics from FLR scores versus expression fold change for cell type pairs H1-hESC versus K562 (left), H1-hESC versus GM12878 (middle) and GM12878 versus K562 (right) for footprints predicted by: HINT-BC, DNase2TF, Neph and FLR (from top to bottom, respectively). We observe high FLR-Exp (Spearman correlation) values (*r* > 0.8) for all cases. Moreover, similar rankings of methods are obtained on the FLR-Exp for each cell pair: H1- hESC/K562 versus H1-hESC/GM12878 *r* = 0.99, H1-hESC/K562 versus GM12878/K562 *r* = 0.96 and H1-hESC/GM12878 versus GM12878/K562 *r* = 0.97. |
|  |
| **Supplementary Figure 2** |
| FLR-Exp results for different footprint quality metrics. |
| Correlation between Kolmogorov-Smirnov (KS) test statistics versus expression fold change for cell type pair H1- hESC versus K562 by evaluating either the FLR (left), FS (middle) and TC (right) as quality metric for the footprints. Footprints were predicted with HINT-BC, DNase2TF, Neph and FLR (from top to bottom, respectively). The use of FLR as quality metric presents the highest Spearman correlation values (FLR-Exp). On the other hand, TC exhibits small correlation values (*r* < 0.4) and presents several cases in which the signal of KS and fold change disagree (off diagonal points). Note that the use of FS also have a high average correlation with fold change expression on all evaluated data/methods (average *r* = 0.73) and indicates a ranking of footprint methods similar to FLR (*r* = 0.89). Therefore, FS can be used as an alternative to the FLR score as a footprint quality metric. |
|  |
| **Supplementary Figure 3** |
| Clustering of sequence bias estimates. |
| Ward's minimum variance clustering on pairwise Spearman correlation coefficient (*r*) of sequence bias estimates of all ENCODE's Tier 1 and 2 DNase-seq data sets and naked DNA DNase-seq data sets. DNase-seq experiments were based on single-hit (red), double-hit (blue) protocols or naked DNA (yellow). We observe a high average correlation between sequence biases estimated on DNase-seq data sets originated from the same protocol: single-hit = 0.89; double-hit = 0.84. Also, lower average correlation values are observed from experimental biases estimates from different protocols: single-hit versus double-hit = 0.39. The group of sequence bias estimates based on the three naked DNA data sets have an average correlation of 0.96. |
|  |
| **Supplementary Figure 4** |
| Correlation between the performance of methods and their OBS on *He Dataset*. |
| The *x*-axis represents the observed sequence bias. The *y*-axis represents the ratio between the AUC at 10% FPR for a particular method and the TC-Rank method. In accordance with He et al.1, we observe that FS-Rank method has a high negative correlation (*r* = −0.4144; adjusted *p*-value < 0.001) with the sequence bias score, while no significant correlation is found for all other evaluated methods HINT, HINT-BCN, HINT-BC and PWM-Rank. It is important to notice that the correlation value for FS-Rank method differs from He *et al.*1. This stems from a different strategy to find the DHSs and MPBSs used in the evaluation dataset. Nevertheless, we were able to observe a strong bias for the FS-Rank method as in He *et al.*1.  1. He, H.H. et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73−78 (2014). |
|  |
| **Supplementary Figure 5** |
| Evaluation of sequence bias correction strategies and CG content contribution. |
| (**a**) Distribution of AUC (at 10% FPR) differences between HINT-BC and HINT; HINT-BCN and HINT; HINT-BC and HINT-BCN for all 233 TFs of the *comprehensive dataset*. TFs are ranked by the difference between HINT-BC and HINT-BCN. There is a clear increase in AUC values between sequence bias-corrected methods (HINT-BC and HINT-BCN) and the uncorrected method HINT (*p*-value < 10−30; Mann- Whitney-Wilcoxon test). Moreover, HINT-BC has higher AUC values for all but seven TFs in the comparison with HINT-BCN. (**b**) CG content of TF motifs. We observe no correlation between CG content of the motifs and the individual AUC of each method: HINT *r* = 0.0144, HINT-BC *r* = 0.0254 and HINT-BCN *r* = 0.0108 (*p*-value > 0.05; Spearman correlation test). Furthermore, we observe no correlation between CG content of motifs and differences in AUC: HINT-BC − HINT-BCN *r* = 0.0188, HINT-BC − HINT *r* = 0.0724 and HINT-BCN − HINT *r* = 0.0644 (*p*-value > 0.05; Spearman correlation test). |
|  |
| **Supplementary Figure 6** |
| Average DNase-seq signals around selected TFs with ChIP-seq evidence in H1-hESC (DU) cell type. |
| These TFs had the higher AUC gain between HINT-BC and HINT: (**a**) ATF3, (**b**) EGR1, (**c**) NRF1, (**d**) RAD21, (**e**) SP1 and (**f**) SP4. In the top panel of each graph, we show the strand-specific average DNase-seq signal on naked DNA DNase-seq experiments (MCF-7 cell type); the middle panel shows the strand-specific estimated DHS sequence bias signal; and the bottom panels shows the (1) uncorrected – observed DNase-seq I cleavage signal and (2) corrected – DNase-seq signal after the bias correction. Signals in the bottom graph were standardized to be in the interval [0,1]. The motif logo represents all underlying DNA sequences centered on the TFBSs. The bias correction led to a substantial change in the average DNase-seq sequence bias patterns surrounding several TFs. On EGR1, for instance, we observed that the bias-corrected DNase-seq signal presents three clear depletions, which fit the high affinity regions of EGR1 motif (two CC and one C). In contrast, EGR1 uncorrected DNase-seq signal presents a single peak in the center of the motif. The same observations can be made for other TFs, such as NRF1 (with affinity regions (C/G)(C/G)(G/C)C and G(G/C)(C/G)(C/G)C) and SP4 (with affinity region CGCCC). Such patterns reflect bias corrections which are clearly beneficial to footprinting method accuracy. |
|  |
| **Supplementary Figure 7** |
| Association between *6*-mer CG content and DNase-seq sequence bias. |
| We sorted *6*-mers by their bias estimates and grouped similar ranked *6*-mers. We show scatter plots with CG content versus average sequence bias for *6*-mer groups on DNase-seq data generated with the (**a**) single-hit (DU), (**b**) double-hit (UW) protocols and (**c**) naked DNA experiments. There is a strong positive correlation between DNase-seq sequence bias and CG content for all DHS sequence bias estimates from both single-hit and double-hit protocols (*p*-value < 0.01). Interestingly, we observe a negative correlation for two naked DNA experiments: K562 and IMR90 (*p*-value < 10−5). |
|  |
| **Supplementary Figure 8** |
| Analysis of footprint ranking strategies. |
| Distribution of AUC values (at 10% FPR) by using distinct ranking strategies for site centric methods: (**a**) BinDNase, (**b**) Centipede, (**c**) Cuellar, (**d**) FLR, (**e**) PIQ and (**f**) segmentation methods DNase2TF and Wellington. Ranking strategies (*x*-axis) are ranked by decreasing median AUC. The site-centric methods are tested based on probability (P) cutoffs of 0.8, 0.85, 0.9, 0.95, 0.99 and their own ranking strategy (Own rank). Segmentation methods are tested based on the TC metric ranking and their own ranking strategy (Own rank). Methods not shown in this figure do not contain an intrinsic ranking methodology. In all cases, using TC-based strategies/cutoff was significantly better than the original ranking of the methods (*p*-value < 10−12; Mann-Whitney-Wilcoxon test). Concerning site-centric methods, the use of a probability threshold (P) of 0.9 was best for all methods, with the exception of BinDNase, where 0.8 was best. The box plot depicts the distribution median value (middle dot) and first and third quartiles (box extremities). The whiskers represent the 1.5 IQR and external dots represent outliers (data greater than or smaller than 1.5 IQR). |
|  |
| **Supplementary Figure 9** |
| Accuracy of methods based on TF ChIP-seq evaluation strategy. |
| Accuracy distribution for all 15 footprinting methods regarding all TF ChIP-seq validation sets (ordered by Friedman Ranking). Accuracies are shown for the statistics: (**a**) AUC at 100% FPR (**b**) AUC at 10% FPR (**c**) AUC at 1% FPR and (**d**) AUPR. We used the Friedman-Nemenyi hypothesis test for statistical evaluation (see **Supplementary Tables 3-6**). The box plot depicts the distribution median value (middle dot) and first and third quartiles (box extremities). The whiskers represent the 1.5 IQR and external dots represent outliers (data greater than or smaller than 1.5 IQR). |
|  |
| **Supplementary Figure 10** |
| Average sequence bias and DNase-seq signals around nuclear receptors. |
| Results are shown for the TFs: (**a**) AR (R1881), (**b**) GR (with DEX), (**c**) ER (40 min) and (**d**) ER (160 min). In the top panel, we show the strand-specific average DNase-seq signal on naked DNA DNase-seq experiments (MCF-7 (DU) for data sets from single-hit and IMR90 (UW) for data sets with double-hit protocol); the middle panel shows the strand-specific estimated DHS sequence bias signal; and the bottom panels shows the (1) uncorrected – observed DNase-seq signal and (2) corrected – DNase-seq signal after the bias correction with the DHS sequence bias estimates. Signals in the bottom graph were standardized to be in the interval [0,1]. The motif logo represents all underlying DNA sequences centered on the TFBSs. While corrected DNase-seq profiles from ER have a better match with the underlying motif, this is not the case for AR and GR. However, we observed a small gain in the AUC score comparing HINT- BC and HINT. This difference is in the upper quartile range for all 233 TFs analyzed. These results indicate that cleavage bias correction also brings improvements to footprint prediction of nuclear receptors. However, all these TFs have low AUC scores in all footprinting methods, i.e. lower quartiles for HINT-BC or TC-Rank AUC scores. This indicates that short binding time indeed poses a challenge in footprint prediction. |
|  |
| **Supplementary Figure 11** |
| Average sequence bias and DNase-seq signals around binding sites of *de novo* motifs found using Neph footprints. |
| Results are shown for de novo motifs: (**a**) #0458 and (**b**) #0500 binding on cell type H7-hESC (UW). In the top panel, we show the strand-specific average DNase-seq signal on naked DNA DNase-seq experiments (MCF-7 cell type); the middle panel shows the strand-specific estimated DHS sequence bias signal; and the bottom panels shows the (1) uncorrected – observed DNase-seq signal and (2) corrected – DNase-seq signal after the bias correction using DHS sequence bias estimates. Signals in the bottom graph were standardized to be in the interval [0,1]. The motif logo represents all underlying DNA sequences centered on the TFBSs. These motifs were discovered in the footprint analysis of Neph et al.1 and indicated in He et al.2 to be artifacts of sequence bias. Bias-corrected DNase-seq profiles reveal no clear footprint shape. Furthermore, we compared the overlap between footprints generated by HINT-BC and Neph in H7-hESC (UW) cells. We considered only the MPBSs that overlapped DHSs in H7-hESC. We observed that 24.99% (motif #0458) and 28.58% (motif #0500) of MPBSs were associated with a Neph footprint. In contrast, only 0.73% (motif #0458) and 1.71% (motif #0500) of MPBSs overlapped with a HINT-BC footprint. Altogether, this indicates that these motifs are indeed potential artifacts of sequence bias and reinforces the importance of bias correction prior to any DNase-seq analysis.  1. Neph, S. et al. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012). 2. He, H.H. et al. Refined DNase-seq protocol and data analysis reveals intrinsic bias in transcription factor footprint identification. *Nat. Methods* **11**, 73−78 (2014). |