Supplementary material for "Joint modelling of ChIP-seq data via a Markov random field model"

YANCHUN BAO\textsuperscript{1}, VERONICA VINCIOTTI\textsuperscript{1,*}, ERNST WIT\textsuperscript{2} and PETER 'T HOEN\textsuperscript{3,4}

\textsuperscript{1}School of Information Systems, Computing and Mathematics, Brunel University, UK
\textsuperscript{2}Institute of Mathematics and Computing Science, University of Groningen, Groningen, The Netherlands
\textsuperscript{3}Department of Human Genetics, Leiden University Medical Center, Leiden, The Netherlands
\textsuperscript{4}Netherlands Bioinformatics Centre, The Netherlands

veronica.vinciotti@brunel.ac.uk

In Section 1 we provide more details about the posterior distributions and the MCMC implementation of the Markov random field model (MRF). In Section 2, we present a comparison between MRF, a classical hidden Markov model and an Ising model. In Section 3, we provide additional results.

1. Posterior distributions for the parameters

For simplicity, we assume only one condition and the same antibody for all replicates. Therefore, in the following sections we drop the subscripts $c$ and $a$. A similar derivation applies to the more general case. We present distributions for the case of a mixture of zero-inflated negative Binomial (background) and negative Binomial (signal).
1.1 Posterior distributions for a single experiment

We start with the case when only one replicate is available. Following the same notation of the manuscript, we choose Gamma and Beta conjugate priors for the parameters as follows:

\[ \tilde{q}_0 \sim \text{Beta}(A_{\tilde{q}_0}, B_{\tilde{q}_0}), \tilde{q}_1 \sim \text{Beta}(A_{\tilde{q}_1}, B_{\tilde{q}_1}) \]

\[ \pi \sim \text{Beta}(A_{\pi}, B_{\pi}) \]

\[ \lambda_0 \sim \Gamma(A_{\lambda_0}, B_{\lambda_0}), \lambda_1 \sim \Gamma(A_{\lambda_1}, B_{\lambda_1}) \]

\[ \mu_0 \sim \Gamma(A_{\mu_0}, B_{\mu_0}), \mu_1 \sim \Gamma(A_{\mu_1}, B_{\mu_1}) \]

\[ \phi_0 \sim \Gamma(A_{\phi_0}, B_{\phi_0}), \phi_1 \sim \Gamma(A_{\phi_1}, B_{\phi_1}), \]

where \( \Gamma(\alpha, \beta) \) represents the Gamma distribution with density \( f(y) = \frac{y^{\alpha-1} \exp(-\beta y)}{\Gamma(\alpha)} \), for \( y > 0 \). These prior distributions lead to the posterior distributions

\[ \tilde{q}_0 \sim \text{Beta}(A_{\tilde{q}_0} + n_{0,1}, B_{\tilde{q}_0} + n_{0,0}) \] (1.1)

\[ \tilde{q}_1 \sim \text{Beta}(A_{\tilde{q}_1} + n_{1,1}, B_{\tilde{q}_1} + n_{1,0}) \]

\[ \pi \sim \text{Beta}(A_{\pi} + \sum_i I(X_m = 0, Z_m = 1), B_{\pi} + \sum_i I(X_m = 0, Z_m = 0)) \]

\[ \lambda_0 \sim \Gamma(A_{\lambda_0} + \sum_m y_m I(X_m = 0, Z_m = 1), B_{\lambda_0} + \sum_m I(X_m = 0, Z_m = 1)) \]

\[ \lambda_1 \sim \Gamma(A_{\lambda_1} + \sum_m I(X_m = 1), B_{\lambda_1} + \sum_m I(X_m = 1)) \]

\[ \mu_0 \sim f(\mu_0|\phi_0, Y, X, Z) \]

\[ \mu_0 = \frac{\sum_m y_m I[X_m = 0, Z_m = 1]}{A_{\mu_0} - 1} \exp(-B_{\mu_0} \mu_0) \]

\[ \mu_1 \sim f(\mu_1|\phi_1, Y, X) \]

\[ \phi_0 \sim f(\phi_0|\mu_0, Y, X, Z) \]

\[ \phi_0 = \frac{\Gamma(y_0 + \phi_0)}{\Gamma(\phi_0) \Gamma(y_0 + 1)} \left( \frac{\mu_0}{\mu_0 + \phi_0} \right)^{y_0} \left( \frac{\phi_0}{\mu_0 + \phi_0} \right)^{\phi_0} \exp(-B_{\phi_0} \phi_0) \] (1.2)
\[
\phi_1 \sim f(\phi_1 | \mu_1, Y, X) = \prod_{m=1}^{M} \left[ \frac{\Gamma(y_m + \phi_1)}{\Gamma(\phi_1) \Gamma(y_m + 1)} \left( \frac{\mu_1}{\mu_1 + \phi_1} \right)^{y_m} \left( \frac{\phi_1}{\mu_1 + \phi_1} \right)^{\phi_1} \right]^{I[X_m=1]} \phi_1^{A_{\phi_1}-1} \exp(-B_{\phi_1} \phi_1). 
\]

Since the posterior densities of \( \mu_0, \mu_1, \phi_0 \) and \( \phi_1 \) have no closed form, we sample from their posterior distributions using a Metropolis-Hastings algorithm. New parameter values \( \mu_0', \mu_1', \phi_0' \) and \( \phi_1' \) are proposed given the current values \( \mu_0, \mu_1, \phi_0, \phi_1 \), using symmetric random walk updates. To ensure that the parameters \( \mu_i \) and \( \phi_i, i = 0, 1 \) are non-negative, we use the following Normal proposal densities,

\[
\mu_i' \sim N(\mu_i, \sigma_{\mu_i}^2), \phi_i' \sim N(\phi_i, \sigma_{\phi_i}^2)
\]

truncated at zero, where \( \sigma_{\mu_i} \) and \( \sigma_{\phi_i} \) can be adjusted to improve the efficiency of the algorithm. The new values of \( \mu_i \) are accepted with probabilities \( \min(1, D_{\mu_i}) \) and \( \phi_i \) are accepted with probabilities \( \min(1, D_{\phi_i}) \) for \( i = 0, 1 \), where

\[
D_{\mu_0} = \frac{p(\mu_0'|\phi_0)q(\mu_0'|\mu_0)}{p(\mu_0'|\phi_0)q(\mu_0'|\mu_0')} \cdot \frac{\exp(-B_{\mu_0}(\mu_0' - \mu_0))}{\exp(-B_{\mu_0}(\mu_0' - \mu_0))} 
\] \[
= \left( \frac{\mu_0'}{\mu_0 + \phi_0} / \frac{\mu_0'}{\mu_0 + \phi_0} \right)^{\sum y_m I[X_m=0,Z_m=1]} \left( \frac{\mu_0' + \phi_0}{\mu_0' + \phi_0} \right)^{\phi_0 \sum I[X_m=0,Z_m=1]} \left( \frac{\mu_0'^2}{\mu_0'^2} \right)^{A_{\mu_0}-1} \exp(-B_{\mu_0}(\mu_0' - \mu_0)) E_{\mu_0} 
\]

\[
D_{\mu_1} = \frac{p(\mu_1'|\phi_1)q(\mu_1'|\mu_1)}{p(\mu_1'|\phi_1)q(\mu_1'|\mu_1')} \cdot \frac{\exp(-B_{\mu_1}(\mu_1' - \mu_1))}{\exp(-B_{\mu_1}(\mu_1' - \mu_1))} 
\] \[
= \left( \frac{\mu_1'}{\mu_1 + \phi_1} / \frac{\mu_1'}{\mu_1 + \phi_1} \right)^{\sum y_m I[X_m=1]} \left( \frac{\mu_1' + \phi_1}{\mu_1' + \phi_1} \right)^{\phi_1 \sum I[X_m=1]} \left( \frac{\mu_1'^2}{\mu_1'^2} \right)^{A_{\mu_1}-1} \exp(-B_{\mu_1}(\mu_1' - \mu_1)) E_{\mu_1} 
\]

\[
D_{\phi_0} = \frac{p(\phi_0'|\mu_0)q(\phi_0'|\phi_0)}{p(\phi_0'|\mu_0)q(\phi_0'|\phi_0')} \cdot \frac{\exp(-B_{\phi_0}(\phi_0' - \phi_0))}{\exp(-B_{\phi_0}(\phi_0' - \phi_0))} 
\] \[
= \prod_{m=1}^{M} \left[ \frac{\Gamma(y_m + \phi_0)}{\Gamma(\phi_0) \Gamma(y_m + \phi_0)} \left( \frac{\mu_0 + \phi_0}{\mu_0 + \phi_0} \right)^{y_m} \left( \frac{\phi_0}{\mu_0 + \phi_0} \right)^{\phi_0} \left( \frac{\phi_0}{\mu_0 + \phi_0} \right)^{-\phi_0} \right]^{I[X_m=0,Z_m=1]} \left( \frac{\phi_0'}{\phi_0} \right)^{A_{\phi_0}-1} \exp(-B_{\phi_0}(\phi_0' - \phi_0)) E_{\phi_0} 
\]
Y. Bao and others

\[ D_{\phi} = \frac{p(\phi_1 | \mu_1)q(\phi_1 | \phi_1^t)}{p(\phi_1 | \mu_1)q(\phi_1 | \phi_1^t)} \]

\[ = \prod_{m=1}^{M} \left[ \frac{\Gamma(y_m + \phi_1^t)\Gamma(\phi_1^t)}{\Gamma(y_m + \phi_1)\Gamma(\phi_1)} \right] \left( \frac{\mu_1 + \phi_1}{\mu_1 + \phi_1^t} \right)^{y_m} \left( \frac{\phi_1^t}{\mu_1 + \phi_1^t} \right)^{-\phi_1^t} \right]_{[X_m=1]} \]

\[ \left( \frac{\phi_1^t}{\phi_1} \right)^{A_{\theta_1} - 1} \exp(-B_{\phi_1}(\phi_1^t - \phi_1^t))E_{\phi_1} \]

and \( E_{\mu_i} = \Phi(\mu_i^t/\sigma_{\mu_i})/\Phi(\mu_i^t/\sigma_{\mu_i}) \) and \( E_{\phi_i} = \Phi(\phi_i^t/\sigma_{\phi_i})/\Phi(\phi_i^t/\sigma_{\phi_i}) \) are the ratio of normalising constants due to the truncated proposals for \( \mu \) and \( \phi \) respectively.

1.2 Posterior distributions in the presence of replicates

Similarly to before, the posterior distributions of the parameters for \( r \) replicates, with \( r = 1, \ldots, R \), are

\[ q_0 \sim \text{Beta}(A_{\theta_0} + n_{0,1}, B_{\theta_0} + n_{0,0}) \]

\[ q_1 \sim \text{Beta}(A_{\theta_1} + n_{1,1}, B_{\theta_1} + n_{1,0}) \]

\[ \pi_r \sim \text{Beta}(A_{\pi_r} + \Sigma_m I(X_m = 0, Z_{mr} = 1), B_{\pi_r} + \Sigma_m I(X_m = 0, Z_{mr} = 0)) \]

\[ \lambda_{0r} \sim \Gamma(A_{\lambda_{0r}} + \Sigma_m y_{mr} I(X_m = 0, Z_{mr} = 1), B_{\lambda_{0r}} + \Sigma_m I(X_m = 0, Z_{mr} = 0)) \]

\[ \lambda_{1r} \sim \Gamma(A_{\lambda_{1r}} + \Sigma_m y_{mr} I(X_m = 1), B_{\lambda_{1r}} + \Sigma_m I(X_m = 1)) \]

\[ \mu_{0r} \sim f(\mu_{0r} | \phi_{0r}, Y_r, X, Z_r) \]

\[ \mu_{1r} \sim f(\mu_{1r} | \phi_{1r}, Y_r, X) \]

\[ \phi_{0r} \sim f(\phi_{0r} | \mu_{0r}, Y_r, X, Z_r) \]

\[ \phi_{1r} \sim f(\phi_{1r} | \mu_{1r}, Y_r, X) \]

where \( f(\mu_{0r} | \phi_{0r}, Y_r, X, Z_r), f(\mu_{1r} | \phi_{1r}, Y_r, X), f(\phi_{0r} | \mu_{0r}, Y_r, X, Z_r) \) and \( f(\phi_{1r} | \mu_{1r}, Y_r, X) \) are defined in a similar way to (1.1).
1.3 Posterior distributions assuming the same proportion of binding sites

Using the same notation of the manuscript, if we assume a Uniform prior distribution for \( s \), then the posterior distributions for the parameters for conditions \( c, c = 1, \ldots, C \), and replicates \( r = 1, \ldots, R_c \) are

\[
s \sim \prod_{c=1}^{C} \left( \frac{1}{1+s} \right)^{I(X_{1c}=1)} \left( 1 - \frac{1}{1+s} \right)^{I(X_{1c}=0)} \left( 1 - s\tilde{q}_{0c} \right)^{n_{1,1}^c} \left( s\tilde{q}_{0c} \right)^{n_{1,0}^c}
\]

\[
\tilde{q}_{0c} \sim (1 - s\tilde{q}_{0c})^{n_{1,1}^c} \left( s\tilde{q}_{0c} \right)^{n_{1,0}^c + A \tilde{q}_{0c}^{-1} (1 - \tilde{q}_{0c})^{n_{0,0}^c + B \tilde{q}_{0c}^{-1}}
\]

\[
\pi_{cr} \sim B(A_{\pi_{cr}} + \Sigma_m I(X_{mc} = 0, Z_{mcr} = 1), B_{\pi_{cr}} + \Sigma_m I(X_{mc} = 0, Z_{mcr} = 0))
\]

\[
\lambda_{0cr} \sim \Gamma(A_{\lambda_{0cr}} + \Sigma_m y_{mcr} I(X_{mc} = 0, Z_{mcr} = 1), B_{\lambda_{0cr}} + \Sigma_m I(X_{mc} = 0, Z_{mcr} = 1))
\]

\[
\lambda_{1cr} \sim \Gamma(A_{\lambda_{1cr}} + \Sigma_m y_{mcr} I(X_{mc} = 1), B_{\lambda_{1cr}} + \Sigma_m I(X_{mc} = 1))
\]

\[
\mu_{0cr} \sim f(\mu_{0cr} | \phi_{0cr}, Y_{cr}, X_c, Z_{cr})
\]

\[
\mu_{1cr} \sim f(\mu_{1cr} | \phi_{1cr}, Y_{cr}, X_c)
\]

\[
\phi_{0cr} \sim f(\phi_{0cr} | \mu_{0cr}, Y_{cr}, X_c, Z_{cr})
\]

\[
\phi_{1cr} \sim f(\phi_{1cr} | \mu_{1cr}, Y_{cr}, X_c)
\]

where \( f(\mu_{0cr} | \phi_{0cr}, Y_{cr}, X_c, Z_{cr}), f(\mu_{1cr} | \phi_{1cr}, Y_{cr}, X_c), f(\phi_{0cr} | \mu_{0cr}, Y_{cr}, X_c, Z_{cr}), f(\phi_{1cr} | \mu_{1cr}, Y_{cr}, X_c) \) and \( f(\phi_{1cr} | \mu_{1cr}, Y_{cr}, X_c) \) are defined in a similar way to (1.1).

Since the posterior distributions of \( s \) and \( \tilde{q}_{0c}, c = 1, \ldots, C \), have no closed form, we use a Metropolis-Hastings algorithm for sampling. To ensure that the parameter \( s \) is non-negative and parameters \( \tilde{q}_{0c} \) are in the interval \([0, 1]\), we use a Normal proposal density \( N(s, \sigma_s^2) \) truncated at zero and \( N(\tilde{q}_{0c}, \sigma_{\tilde{q}_{0c}}^2) \) truncated at \([0, 1]\) for \( s \) and \( \tilde{q}_{0c} \) respectively, where \( \sigma_s \) and \( \sigma_{\tilde{q}_{0c}} \) are adjusted to improve the efficiency of the algorithm.
2. Model comparison between MRF and iSeq

In this section, we compare our proposed Markov random field model (MRF) with the existing Ising model of Mo (2012), which is implemented in iSeq for single experiments, and with a classical hidden Markov model, which is implemented in BayesPeak for single experiments. For this reason, we drop the subscripts \( c, a \) and \( r \) for MRF and consider the case of one replicate.

2.1 One dimensional Ising model (iSeq)

For comparison purposes, we follow the notation of Mo (2012). Let \( X_m \in \{-1, 1\} \) be the binary latent variable, where \( X_m = -1 \) (1) denotes that the bin belongs to a non-enriched (enriched) region, respectively. Mo (2012) uses a one-dimensional Ising model without exterior field, defined as following:

\[
P(X_1, \ldots, X_M | \kappa) = \frac{1}{Z_1(\kappa)} \exp(\kappa \sum_{m=1}^{M-1} X_m X_{m+1}),
\]

where \( \kappa \) is the interaction parameter and \( Z_1(\kappa) \) is the normalizing constant of the distribution given by:

\[
Z_1(\kappa) = 2^n (\cosh(\kappa))^{n-1}, \quad \cosh(\kappa) = \frac{e^\kappa + e^{-\kappa}}{2}.
\]

2.2 Hidden Markov model

In this section, we describe a classical one-dimensional Markov random field model, often referred to as hidden Markov model (HMM). This is implemented in BayesPeak (Spyrou and others, 2009) for the detection of peak binding sites. We use the same notation of Mo (2012).

A classical two states Markov chain is defined by

\[
P(X_1, \ldots, X_M) = \pi_0(X_1) \prod_{m} q_{X_m, X_{m+1}}
= f_0(X_1) q_{1,1}^{n_{1,1}} q_{1,-1}^{n_{1,-1}} q_{-1,1}^{n_{-1,1}} q_{-1,-1}^{n_{-1,-1}}
\]

(2.5)
where $f_0$ is the distribution of the initial state and $q_{i,j} = P(X_{m+1} = j | X_m = i), i, j \in \{-1, 1\}$ are the transition probabilities. This is a one dimensional random field Markov model since it follows the Markov property,

$$P(X_m = i | X_{m-1}) = P(X_m = i | X_{m-1}, X_{m+1}).$$ \hspace{1cm} (2.6)

### 2.3 One dimensional Markov random field model (MRF)

In the manuscript, we define a one dimensional random field Markov model with the following representation for the joint density of the latent states:

$$P(X_1, \ldots, X_M) = \frac{\prod_{m=1}^{M-1} P(X_m, X_{m+1})}{\prod_{m=2}^{M-1} P(X_m)},$$ \hspace{1cm} (2.7)

where $P(X_m, X_{m+1})$ is the joint probability of $X_m$ and $X_{m+1}$ and $P(X_m)$ is the marginal probability of $X_m$. We have $P(X_m) = \sum_{x_{m+1}} P(X_m; X_{m+1} = x_{m+1})$. When $X$ is a binary variable, as in our case, we can further re-write model (2.7) as

$$P(X_1, \ldots, X_M) = \frac{\delta_{1,1}^{n_{1,1}} \delta_{1,-1}^{n_{1,-1}} \delta_{-1,1}^{n_{-1,1}} \delta_{-1,-1}^{n_{-1,-1}}}{\delta_1^{n_{1,1} + n_{1,-1} - I[X_1 = 1]} \delta_{-1}^{n_{-1,1} + n_{-1,-1} - I[X_1 = -1]}}$$

where

$$\delta_1 = P(X_m = 1) = \delta_{1,1} + \delta_{1,-1} \quad \delta_{-1} = P(X_m = -1) = 1 - \delta_1.$$

$$n_{i,j} = \# \{X_m = i, X_{m+1} = j, m = 1, \ldots, M - 1\}, i, j \in \{-1, 1\}$$

$$\delta_{i,j} = P(X_m = i, X_{m+1} = j), i, j \in \{-1, 1\}; m = 1, \ldots, M - 1.$$

It is easy to prove that our model follows the Markov property (2.6). From the definition of $P(X_m)$, one can show that $\delta_{-1,1} = \delta_{1,-1}$ and since $\sum_{i,j \in \{-1,1\}} \delta_{i,j} = 1$ we have

$$\delta_{1,-1} = \delta_{-1,1} = (1 - \delta_{1,1} - \delta_{-1,-1})/2.$$
We can further write the parameters $\delta_{i,j}$ in terms of the transition probabilities $q_{i,j}$ as following

$$
\delta_{1,1} = \frac{q_{-1,1}q_{1,1}}{q_{-1,1}q_{1,-1}}, \quad \delta_{-1,-1} = \frac{q_{1,-1}q_{-1,-1}}{q_{-1,1}q_{1,-1}}
$$

$$
\delta_{1,-1} = \delta_{-1,1} = \frac{q_{-1,1}q_{1,-1}}{q_{-1,1}q_{1,-1}}.
$$

This allows to write the joint density of the latent states in terms of the transition probabilities:

$$
P(X_1, \ldots, X_M) = \delta_1 I(X_1=1) \delta_{-1} I(X_1=-1) \left( \frac{\delta_{1,1}}{\delta_1} \right)^{n_{1,1}} \left( \frac{\delta_{-1,-1}}{\delta_{-1}} \right)^{n_{-1,-1}} \left( \frac{\delta_{1,-1}}{\delta_1} \right)^{n_{1,-1}} \left( \frac{\delta_{-1,1}}{\delta_{-1}} \right)^{n_{-1,1}}
$$

$$
= \left( \frac{q_{-1,1}}{q_{-1,1}q_{1,-1}} \right)^{I(X_1=1)} \left( \frac{q_{1,-1}}{q_{-1,1}q_{1,-1}} \right)^{I(X_1=-1)} \delta_1 I(X_1=1) \delta_{-1} I(X_1=-1) \cdot \left( \frac{q_{1,-1}}{q_{1,1}q_{1,-1}} \right)^{n_{1,1}} \left( \frac{q_{-1,1}}{q_{-1,1}q_{1,-1}} \right)^{n_{1,-1}} \left( \frac{q_{1,-1}}{q_{1,1}q_{1,-1}} \right)^{n_{-1,1}} \left( \frac{q_{-1,-1}}{q_{-1,1}q_{1,-1}} \right)^{n_{-1,-1}}. \quad (2.8)
$$

### 2.4 Link between HMM and MRF

Comparing (2.5) with (2.8), we can see that the only difference between HMM and MRF is that, in a classical Markov chain model the initial distribution can be any distribution of the initial state, whereas in our model the initial distribution is the stationary distribution and therefore can be expressed in terms of the transition probabilities. When the classical Markov chain converges to a stationary state, the two models are equivalent.

### 2.5 Link between iSeq and MRF

Now we prove that under the following condition, a one-dimensional Ising model is the same as a HMM and a MRF model.

**Condition 1**: Equal transition probabilities $q_{1,1} = q_{-1,-1} = q$.

Under the above condition we have $q_{-1,1} = q_{1,-1} = 1 - q$ and the stationary distribution
Joint modelling of ChIP-seq data via a RFMM

\(P(X_m = 1) = 1/2\). Therefore equation (2.5) or (2.8) could be further written as

\[
P(X_1, \ldots, X_M) = \frac{1}{2} q^{n_1,1+n_{-1,-1}(1-q)}
\]

\[
= \frac{1}{2} q^{n_{-1,-1}+n_{1,1}(1-q)}
\]

\[
= \tilde{Z}_2(q)\left(\frac{q}{1-q}\right)^{-n_{-1,-1}+n_{1,1}}
\]

\[
= \tilde{Z}_2(q)\exp(-\log\left(\frac{q}{1-q}\right)(n_{1,-1} + n_{-1,1}))
\]

(2.9)

where \(\tilde{Z}_2(q) = \frac{1}{2} q^{n-1}\).

The Ising model (2.3) can be expressed as

\[
p(X_1, \ldots, X_M|\kappa) = \frac{1}{Z_1(\kappa)} \exp(\kappa(n_{1,1} + n_{-1,-1}) - \kappa(n_{1,-1} + n_{-1,1}))
\]

\[
= \frac{1}{Z_1(\kappa)} \exp(\kappa(n-1 - [n_{1,-1} + n_{-1,1}]) - \kappa(n_{1,-1} + n_{-1,1}))
\]

\[
= \frac{1}{Z_1(\kappa)} \exp(-2\kappa(n_{1,-1} + n_{-1,1}))
\]

(2.10)

where \(\tilde{Z}_1(\kappa) = Z_1(\kappa) / \exp(k(n-1))\).

If we let \(\kappa = \frac{1}{2} \log\left(\frac{q}{1-q}\right) = \log\left(\frac{\sqrt{q}}{1-q}\right)^{1/2}\) and notice that

\[
\cosh(\kappa) = \frac{e^{-\kappa} + e^{\kappa}}{2} = \frac{e^{-\kappa}(1+e^{2\kappa})}{2} = \frac{[q/(1-q)]^{1/2}(1+(1-q)/q)}{2} = \frac{1}{2}(q)^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}},
\]

we have

\[
\frac{1}{Z_1(\kappa)} = \frac{\exp(\kappa(n-1))}{Z_1(\kappa)} = \frac{\exp(\kappa(n-1))}{2^n \cosh^{n-1}(\kappa)}
\]

\[
= \frac{2^n \cosh^{n-1}(\kappa)}{\cosh^{(n-1)/2}(q)^{1/2}}
\]

\[
= \frac{2^n (q^{1/2} + (1-q)^{1/2})^{n-1}}{2(q)^{-\frac{1}{2}}(1-q)^{-\frac{1}{2}}} = \frac{1}{2} q^{n-1}.
\]

(2.11)

Therefore, the models (2.9) and (2.10) are equal when condition 1 is satisfied.

We can further see that condition 1 can also be expressed as \(q_{1,1} + q_{1,-1} = 1\) if we notice that \(q_{-1,-1} = 1 - q_{1,-1}\). Condition 1 requires the same transition probabilities when two neighboring bins have the same states. It is a strong condition which is not imposed by the HMM and the MRF models.
3. Additional Results

In Figure 1, we give the BIC values for the eight experiments, using a (non-Markovian) latent mixture model. In general, we find that the BIC values are lower for the ZINB-NB mixture than for the NB mixture, suggesting a better fit for the zero-inflated mixture.

Table 1 complements Figure 1 in the manuscript, where we show evidence of spatial dependencies in the data. After dividing the genome into 200bp fixed windows, we have detected the enriched regions using the latent (non-Markovian) mixture model at a 5% FDR and then calculated the conditional frequencies for each region, given that the previous region is enriched or non enriched. We denote these by $f_{1|1}$ and $f_{1|0}$, respectively. The table shows that these two frequencies are generally not equal. The conditional frequency of a current bin being enriched given that the previous bin is enriched is generally larger than the conditional frequency of a bin being enriched given that the previous bin is not enriched.

Table 2 complements the Venn diagrams of Figure 2 in the manuscript by providing the number of enriched regions for all four single experiments.

Figure 2 shows the ChromHMM validation on replicated experiments, using a 4-state hidden Markov model on the enrichment profile given by MRF and the mixture model of Bao and others (2013), each at a 5% FDR. The emission probabilities (left) show how, for all analyses, two of the three states explain most of the enrichment pattern in the identified lists. The relative fold enrichment plots (right) show how these two states seem to be mostly enriched with TSS and CpGIsland features for both methods.

Table 3 reports the number of enriched and differentially bound regions for chromosome 21 under the assumption that the two proteins, CBP and p300, have the same number of binding sites under the same condition (here the time point). In (Bao and others, 2013), we show how this assumption is justified for the T0 and Wang experiments, so we report the results only for these two cases.
REFERENCES


Fig. 1. BIC values for mixture of two NB distributions (dashed grey line) and mixture of zero-inflated NB distribution for the background and NB distribution for the signal (dashed black line).

Table 1. Conditional frequencies of enrichment given that the previous bin is enriched or not, denoted by $f_{1|1}$ and $f_{1|0}$ respectively. A region is called enriched or not using a latent mixture model at a 5% FDR.

| Experiment  | $f_{1|1}$ | $f_{1|0}$ |
|-------------|-----------|-----------|
| CBPT0       | 0.0909    | 0.0002    |
| CBPT301     | 0.3004    | 0.0008    |
| CBPT302     | 0.4924    | 0.0017    |
| p300T0      | 0.3342    | 0.0021    |
| p300T301    | 0.4015    | 0.0020    |
| p300T302    | 0.5653    | 0.0027    |
| Wang CBP    | 0.4129    | 0.0014    |
| Wang p300   | 0.2190    | 0.0005    |

Table 2. Number of enriched regions identified by MRF, iSeq and BayesPeak for four single experiments, using the data for chromosome 21 and at the 5% FDR.

<table>
<thead>
<tr>
<th>Method</th>
<th>CBPT0</th>
<th>p300T0</th>
<th>WangCBP</th>
<th>Wangp300</th>
</tr>
</thead>
<tbody>
<tr>
<td>MRF</td>
<td>2073</td>
<td>4393</td>
<td>1443</td>
<td>639</td>
</tr>
<tr>
<td>iSeq</td>
<td>488</td>
<td>1115</td>
<td>1126</td>
<td>326</td>
</tr>
<tr>
<td>BayesPeak</td>
<td>1102</td>
<td>1834</td>
<td>603</td>
<td>576</td>
</tr>
<tr>
<td>MACS</td>
<td>265</td>
<td>893</td>
<td>895</td>
<td>169</td>
</tr>
<tr>
<td>CisGenome</td>
<td>14</td>
<td>407</td>
<td>860</td>
<td>46</td>
</tr>
</tbody>
</table>
Fig. 2. Validation of the enriched bins detected by mixture model (MIX, top) and MRF (bottom) for technical replicates of CBP and p300 at time T30. We use a 4-state ChromHMM. The left plots show heatmaps of the probabilities (in %) that the detected bins are enriched given each identified chromatin-state. The right plots show the relative percentage of the genome represented by each chromatin state (column 1) and the relative fold enrichment for several types of annotation (columns 2-8).

Table 3. Number of enriched and differentially regions identified by MRF under the assumption that the two proteins, CBP and p300, have the same number of binding sites at the same time point.

<table>
<thead>
<tr>
<th></th>
<th>Enriched regions</th>
<th>Differentially bound regions</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CBP</td>
<td>p300</td>
</tr>
<tr>
<td>CBPT0 vs p300T0</td>
<td>1606</td>
<td>3842</td>
</tr>
<tr>
<td>WangCBP vs Wangp300</td>
<td>1426</td>
<td>643</td>
</tr>
</tbody>
</table>