Statistical methods for Lerner & Hepperla *et al.*

Description of the data

We sought to characterize the methylation dynamics of histone H3 at lysine 36 (H3K36) at a majority of yeast genes in three biological conditions: after light activation of Set2 (writer add:WA), and after dark inactivation of Set2 with and without the presence of Rpd3, a histone deacetylase (writer loss:WL and writer eraser loss:WEL, respectively). H3K36 trimethyl (H3K36me3) levels were measured through ChIP-seq across four timepoints for each condition (0/20/40/60 min for WA and 0/30/60/90 min for WL and WEL) for 5,368 genes. Seq data were collected and quantified from three sample replicates for each condition across the timepoints. We used two different approaches to process the seq data for analysis. First, within each replicate, H3K36me3 levels for a gene and condition were scaled to the proportion of maximum H3K36me3 level observed (for said replicate, gene, and condition), resulting in a common quantile scale across all replicates, genes, and conditions. Second, we took the measurements, previously standardized based on spike-in samples and normalized to adjust for gene length, multiplied them by 1000, and rounded to an integer, resulting in count data (quasi-counts) with a distribution proportional to the starting data.

General statistical model

Though the H3K36me3 data were large and well-balanced–the vast majority of genes having complete measurements of three replicates from each condition across four timepoints–they possess challenging features from a statistical modeling perspective. The statistical model should accommodate the non-normal distribution of the H3K36me3 data, as well as the time course, *i.e.* longitudinal, nature of observations within a replicate. This broad range of features can be flexibly handled using Bayesian inference [1, 2].

The Stan statistical platform [3] is one such computational tool for fitting complex Bayesian hierarchical models. We used the brms software package [4, 5], which acts as a wrapper of Stan in the R statistical programming language [6]. Let y_{ijkl} be the H3K36me3 measurement for the i^{th} replicate sample ($i = 1, 2, 3$) at the j^{th} timepoint ($j = 0, 1, 2, 3$) for the k^{th} gene ($k = 1, 2, \ldots, 5368$) for l^{th} condition ($l = 1, 2, 3$). Briefly, we model y with a GLM by defining $E(y) = g^{-1}(\eta)$, where $E(.)$ is the expected value of a random variable, $g^{-1}(.)$ is the inverse link function, and η is the linear predictor that relates the outcome to factors of interest.

A Bayesian analysis could in principle simultaneously model all genes and conditions, though in practice, such an approach would likely be computationally infeasible, particularly when considering the need for sufficient sampling in order for the parameter estimates to converge. To avoid these challenges, we instead model more manageable subsets of H3K36me3 data. First, we fit a model of H3K36me3 data for a specific gene and condition. Second, to make more direct comparisons between the writer loss and writer eraser loss conditions, we modeled their data jointly within a gene.

Zero-one-inflated beta regression model of H3K36me3 quantiles

For the H3K36me3 data transformed to the quantile scale, we initially considered dose response (DR) models [7, 8]. Because the DR model implementations did not easily accommodate covariates or the replicate observations and also fit relatively complex models (≤ 5 parameters), we instead modeled the data with a zero-one-inflated beta (ZOIB) distribution with the following parameterization:

$$p(y_i) = \begin{cases} \alpha(1-\gamma) & y_i = 0 \\ \alpha\gamma & y_i = 1 \\ \frac{y_i^{\mu\phi-1}(1-y_i)^{(1-\mu)\phi-1}}{B(\mu\phi,(1-\mu)\phi)} & y_i \notin \{0,1\} \end{cases} \tag{1}$$

α is the zero-one-inflation probability (probability that a zero or one occurs), γ is the conditional one-inflation probability (probability that one occurs rather than zerio), $B(.)$ is the beta function [9], and ϕ is a positive precision parameter. For the link function, we used the logit: $g(\eta) = \log(\frac{\eta}{1-\eta})$. Zero-one-inflation was necessary because standard beta regression expects $y \in (0, 1)$, meaning it cannot handle values at the boundaries.

The H3K36me3 data scaled to the proportion of the maximum H3K36me3 measured within a replicate (three replicates per gene per condition) results in three values of 1 for each gene and condition pairing. Additionally, zeros may be observed. We used this scale of the data to better standardize ChIP-seq dynamics across cells and better detect consistent patterns compared to the normalized ChIP-seq data. However, it does represent a challenging formulation in the context of a GLMM, and possibly even an abuse of the underlying assumptions of a beta regression model, specifically. The zeros and ones are informative, but would strongly violate the expectations of standard beta regression, resulting in anti-conservative, extreme parameter estimates. To avoid this issue, ZOIB conservatively drops out zeros and ones in terms of parameter estimation, resulting in less extreme estimates for the time effect. Alternatively, we modeled the non-quantile, normalized ChIP-seq data (quasi-counts).

Zero-inflated negative binomial regression model of H3K36me3 quasi-counts

The normalized ChIP-seq data after being processed and normalized, ranged from 0 to 1.374346. The distribution was consistent in shape with common count distributions, *e.g.* Poisson and negative binomial, exemplified by being non-negative with a right skew. We calculated quasi-counts as $y_i^{\text{quasi}} = \text{round}(y_i \times 1000)$. Though this transformation is artificial, it produces a new distribution that is proportional to the original and consistent with count distributions.

For the quasi-count scale, we modeled the data with a zero-inflated negative binomial (ZINB) with the following parameterization:

$$p(y_i) = \begin{cases} \binom{y_i+\phi-1}{y_i}(\frac{\mu}{\mu+\phi})^{y_i}(\frac{\phi}{\mu+\phi})^{\phi} & y_i > 0 \\ \xi + (1-\xi)\text{NB}(y_i = 0) & y_i = 0 \end{cases} \tag{2}$$

ξ is the zero-inflation probability, NB(.) is the non-zero-inflated negative binomial probability mass function, and ϕ is a positive precision parameter. As $\phi \to \infty$, the negative binomial distribution converges to the Poisson distribution. The ZINB distribution estimates the zero observations as a mixture of true zeros, expected by the NB distribution, with an additional component from drop-outs, resulting in the zeros having less influence on the parameter estimates.

The inference on the H3K36me3 time rate dynamics based on either ZOIB and ZINB should be consistent with each other. The quantile data scale is more standardized across genes and conditions, but also overly conservative when modeled by ZOIB regression, due to essentially removing the effect of zeros

and ones on parameter inference. By contrast, the quasi-count scale is more artificial and disparate across genes and conditions, but more closely represented the raw data. It also makes more complete use of the data–by not excluding the maximum values which was transformed to one in the quantile data–for parameter estimation.

Time models

Continuous

We modeled the gain or loss of H3K36me3 with a continuous time variable in the linear predictor of the model,

$$\eta_{ij} = \mu + u_i + (\beta_{\text{time}} + v_i)x_{ij}, \tag{3}$$

where μ is a shared intercept term, x_{ij} is the j^{th} timepoint (in minutes) for the i^{th} sample, and β_{time} is the change rate with time. u_i and v_i are random, or group-level [1], effects that account for the longitudinal nature of the data, and modeled with the following priors: $\text{N}(0, \tau_u^2)$ and $\text{N}(0, \tau_v^2)$, respectively. For all genes and conditions, we recorded the posterior mean ($\widehat{\beta}_{\text{time},kl}$) as a point estimate for the change in H3K36me3 (proportion or quasi-count) with time for gene k and condition l, the standard error on the estimate, and the 95% Credible Interval (CrI), which we used to define "confident" genes (CrI that do not cover 0) for a given condition. $\widehat{\beta}_{\text{time},kl}$ were plotted and correlated with additional covariates of interests in order to identify potential relationships with other factors of interest.

Testing for non-zero effect of condition with time

To formally test for a non-zero effect of condition with time is infeasible because it would require the joint estimation across all genes and conditions in a Bayesian context. Further, GLMMs are challenging models to fit, and not amenable to reliable likelihood-based inference, given the number of genes observed here. As an alternative, we fit a second model from the posterior mean $\widehat{\beta}_{\text{time},kl}$, estimated for gene k and condition l, described above. These parameters were modeled as normally distributed, which can be viewed as a latent variable that we model in a second regression as:

$$\beta_{\text{time},kl} = u_k + \delta_{\text{WA}} I\{l = \text{WA}\} + \delta_{\text{WL}} I\{l = \text{WL}\} + \delta_{\text{WEL}} I\{l = \text{WEL}\} + \varepsilon_{kl}, \tag{4}$$

where u_k is a gene-specific random effect, modeled as $\text{N}(0, \tau_u^2)$, δ_{WA}, δ_{WL}, and δ_{WEL} are the condition-specific effects on the previously measured continuous time effect, modeled as fixed effects, and ε_{kl} is a noise term, distributed according to $\text{N}(0, \frac{\sigma^2}{w_{ij}})$ with σ^2 representing the noise variance and w_{kl} is a weight specific to gene k and condition l. For the weight, we used $1/\text{SE}(\beta_{\text{time},kl})$ from the GLMM model, effectively down weighting the influence of effect estimates with large standard error. $I\{A\}$ represents the indicator function, returning 1 if the conditional statement A is satisfied and 0 if not.

Given that the effects can be modeled with a normal distribution, representing a linear mixed effect model (LMM) with weights, we used the R package lme4 [10]. Through ANOVA [11] with maximum likelihood estimates, we compared the model in Equation 4 to a null model with an intercept and no condition fixed effects, resulting in an ANOVA p-value $< 2.2 \times 10^{-16}$. Using the R package emmeans [12], we performed Tukey *post hoc* tests of pairwise differences [11] between the conditions, which were all found to be significant (Tukey p-values < 0.0001). Notably, the rate of H3K36me3 loss was greater for WEL compared to WL.

Categorical

WL and WEL had similar H3K36me3 dynamics with time. To more directly compare their dynamics, we simultaneously analyzed WL and WEL data per gene within a categorical time model, allowing for greater

flexibility and non-linearity with respect to time:

$$\begin{aligned}\eta_{ijl} = &\underbrace{\mu + u_i}_{\text{WL time0}} + \underbrace{(\omega_{\text{WEL}} + w_i) I\{l = \text{WEL}\}}_{\text{WEL time0}} \\ &+ \underbrace{\sum_{p=1}^{3} (\beta^p_{\text{timepoint}} + v_i^p) I\{j = p\}}_{\text{WL timepoint } p} \\ &+ \underbrace{\sum_{p=1}^{3} (\delta^p_{\text{timepoint}} + z_i^p) I\{j = p, l = \text{WEL}\}}_{\text{WEL timepoint } p},\end{aligned} \tag{5}$$

where μ is the intercept representing, representing WL at time0, ω_{WL} is the deviation of WEL from WL, $\beta^p_{\text{timepoint}}$ is the effect comparing timepoint p to 0 for WL, and $\delta^q_{\text{timepoint}}$ represents effect for timepoint p for WEL compared to WL. u_i, w_i, v_i^p, z_i^p are group-level effects that model the correlations within replicates, and have the following prior distributions: $\text{N}(0, \tau_u^2)$, $\text{N}(0, \tau_w^2)$, $\text{N}(0, \tau_{v,p}^2)$, and $\text{N}(0, \tau_{z,p}^2)$, respectively. Similar to the continuous time model, we recorded posterior means and 95% CrI for $\delta^p_{\text{timepoint}}$. We fit the categorial time model with the quantile data and ZOIB, though it could be used with the quasi-counts and ZINB as well.

Interpreting estimated effects

The interpretation of the regression coefficients will depend on whether ZOIB or ZINB was used. For ZOIB, effects represent log odds ratio (OR) for change in proportion H3K36me3 (relative to the sample maximum) per minute. For ZOIB, the effects are log change in quasi-counts per minute. We emphasize that these data were highly derived, resulting in some reduction in the interpretability or tangibility of their estimates. Regardless of data scale, positive effects represent increasing levels of trimethyl marks and negative effects represent decreasing marks. We view large scale trends across genes and/or conditions as meaningful.

Model specification, sampling, and convergence

A Bayesian model requires the specification of prior distributions for the various parameters. We used the default settings from brms. Briefly, for fixed, or population-level, effects, an improper flat prior over the reals was used. Group-level effects are modeled as normal variables with standard deviation parameter. These are modeled with half Student-t distributions with 3 degrees of freedom and a minimal scale parameter of 10 [1], which brms potentially increases based on the data to insure that the prior is minimally informative. The LKJ-correlation prior is used to model the correlations between the group-level effects on the same grouping factor [13].

Bayesian inferences involves random sampling from the joint posterior of the model parameters. For each Stan model, we ran four Monte Carlo Markov Chains (MCMC), each with 2,000 iterations of warm-up and sampling. Initial values for parameters for each chain were randomly generated within Stan. The the adaptive delta parameter, necessary for the No-U-turn Sampler [14] used by Stan, was set to 0.8, the default used by brms.

There are various diagnostics for the MCMC samples that can be used to determining whether the model is converging and performing well. We used the split-$\widehat{R}$ statistic [2], a ratio of the average variance within chains to the variance with the pooled chains. A split-$\widehat{R}$ value close to 1 means the variances within each chain are similar, and thus the model is likely mixing well and converging. Guidelines from the Stan

development team state that models with split-$\widehat{R} > 1.1$ have not converged and should not be used for inference. To ensure replicable results and declare convergence and ensure reproducible results, we set the seed and ran all models in Stan (through brms) for a given gene. We then checked that split-$\widehat{R} < 1.1$ for all parameters. If this was satisfied, we declared convergence and recorded the seed as well as posterior means and CrIs for the parameters. If any split-$\widehat{R} > 1.1$, convergence failed, and we set a new seed and repeated the process. We capped that number of repeats at 10. If convergence is not met after 10 iterations, we declare modeling to have failed for those genes, likely due to oddly distributed and noisy data.

References

[1] Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006.

[2] A. Gelman, J.B. Carlin, H.S. Stern, D.B. Dunson, A. Vehtari, and D.B. Rubin. *Bayesian Data Analysis, Third Edition*. Chapman & Hall/CRC Texts in Statistical Science. Taylor & Francis, 2013.

[3] Bob Carpenter, Andrew Gelman, Matthew D. Hoffman, Daniel Lee, Ben Goodrich, Michael Betancourt, Marcus Brubaker, Jiqiang Guo, Peter Li, and Allen Riddell. Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, 76(1), 2017.

[4] Paul-Christian Bürkner. Advanced Bayesian Multilevel Modeling with the R Package brms. 10(July):395–411, 2017.

[5] Paul-Christian Bürkner. brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 2017.

[6] R Core Team. RSoftware2019, 2019.

[7] Wout Slob. Dose-Response Modeling of Continuous Endpoints. *Toxicological Sciences*, 66(2):298–312, apr 2002.

[8] Ander Wilson, David M. Reif, and Brian J. Reich. Hierarchical dose-response modeling for high-throughput toxicity screening of environmental chemicals. *Biometrics*, 70(1):237–246, mar 2014.

[9] G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002.

[10] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.

[11] W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer Publishing Company, Incorporated, 2010.

[12] Russell Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2019. R package version 1.3.4.

[13] Daniel Lewandowski, Dorota Kurowicka, and Harry Joe. Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001, oct 2009.

[14] Matthew D. Hoffman and Andrew Gelman. The no-u-turn sampler: Adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, 15(1):1593–1623, January 2014.