

Supplemental Material for Lerner, Hepperla, Keele *et al.*

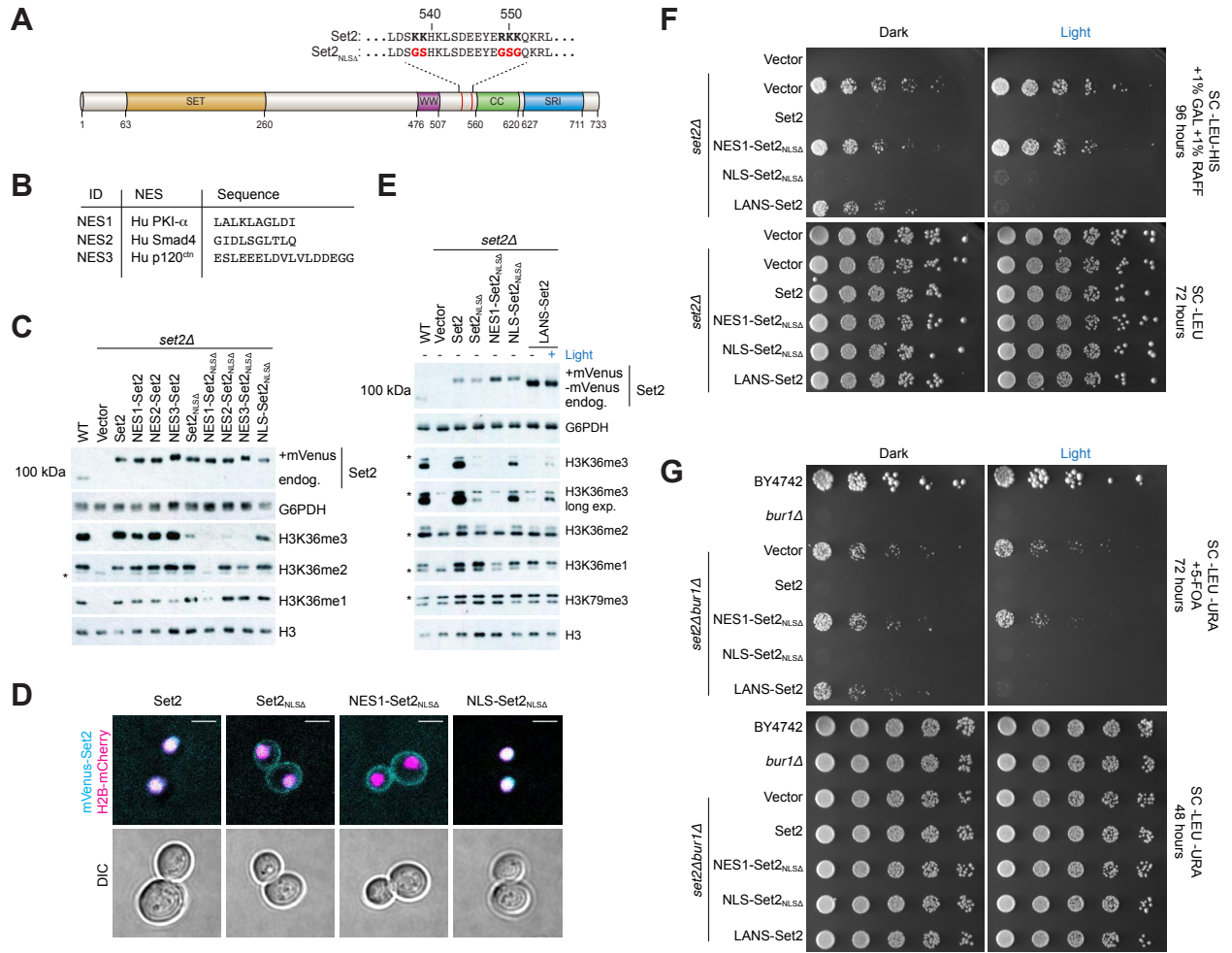
An optogenetic switch for the Set2 methyltransferase provides evidence for transcription-dependent and independent dynamics of H3K36 methylation

Contents

1	Supplemental Figures	1
1.1	Supplemental Fig S1: Related to Figures 1 and 2.	1
1.2	Supplemental Fig S2: Related to Figure 2.	3
1.3	Supplemental Fig S3: Related to Figure 3.	5
1.4	Supplemental Fig S4: Related to Figure 4.	7
1.5	Supplemental Fig S5: Related to Figure 5.	9
1.6	Supplemental Fig S6: Related to Figure 6.	11
2	Supplemental Tables	12
2.1	Supplemental Table S1: Yeast strains used in this study.	12
2.2	Supplemental Table S2: Plasmids used in this study and their Addgene deposition IDs.	12
3	Supplemental Data and Code	13
3.1	Supplemental Data S1	13
3.2	Supplemental Data S2	13
3.3	Supplemental Data S3	13
3.4	Supplemental Data S4	13
3.5	Supplemental Data S5	13
3.6	Supplemental Data S6	13
3.7	Supplemental Data S7	13
3.8	Supplemental Data S8	13
3.9	Supplemental Data S9	13
3.10	Supplemental Code S1	13
4	Additional Files	14
4.1	Supplemental Movie S1: LANS-Set2 light activation and reversion in <i>Saccharomyces cerevisiae</i>	14
4.2	Supplemental Note: Protein amino acid sequences for mVenus-FLAG-LANS-Set2 and FLAG-LANS-Set2.	14
5	Supplemental Statistical Methods	15
5.1	Description of the data	15
5.2	General statistical model	15
5.3	Zero-one-inflated beta regression model of H3K36me3 quantiles	15
5.4	Zero-inflated negative binomial regression model of H3K36me3 quasi-counts	16
5.5	Time models	16
5.5.1	Continuous time model	16

5.5.2 Testing for non-zero effect of condition with time	17
5.5.3 Categorical time model	17
5.5.4 Interpreting estimated effects	18
5.6 Model specification, sampling, and convergence	18

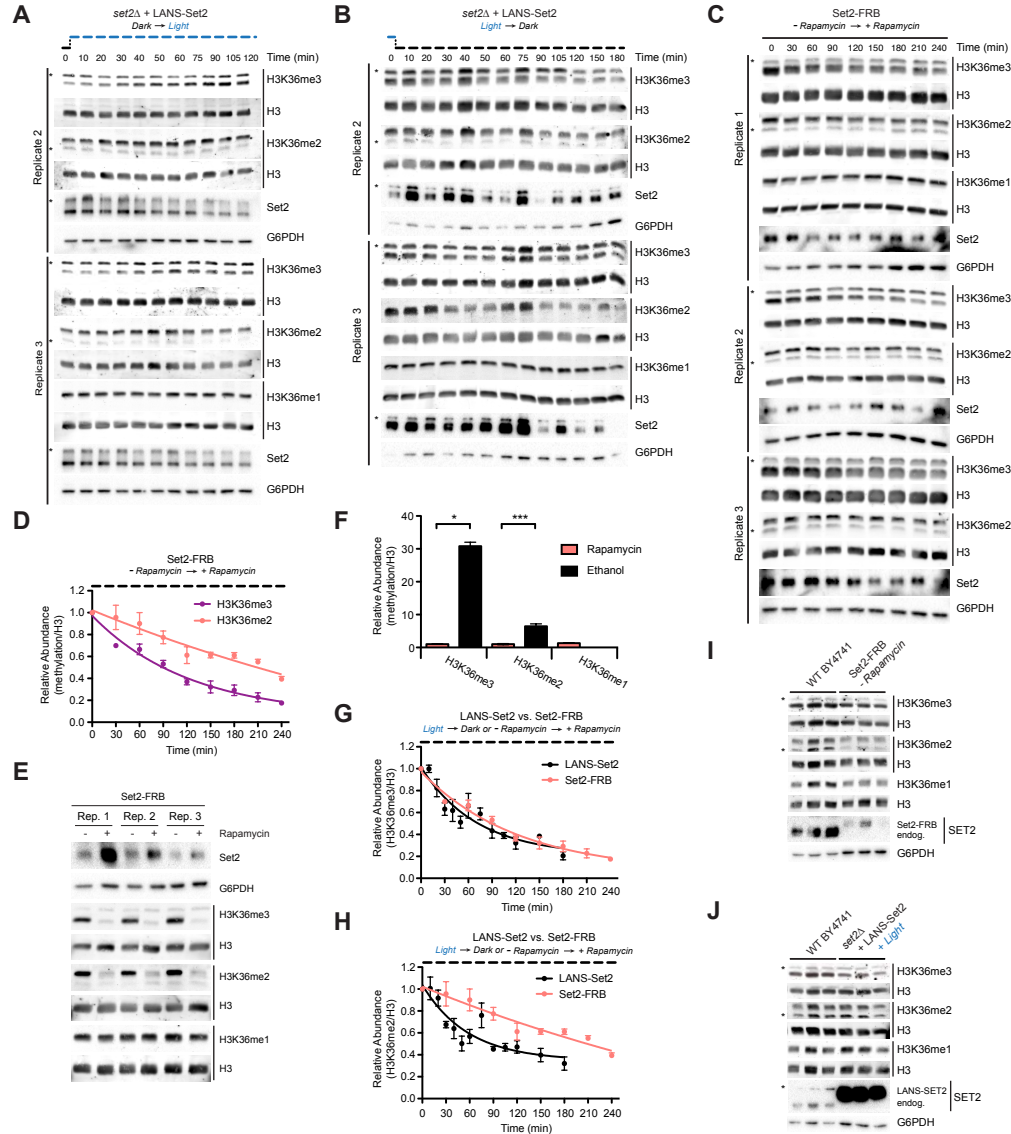
1 Supplemental Figures



1.1 Supplemental Fig S1: Related to Figures 1 and 2.

(A) Sequence alignment and domain map depicting mutations (red) to a putative NLS (bold) and their locations relative to other domains. Set2 is characterized by the SET (Su(var)3-9, Enhancer-of-zeste and Trithorax), WW, CC (coiled coil), and SRI (Set2-Rpb1 interacting) domains. (B) List of the nuclear export signals tested, including their origins and sequences. (C) Immunoblots showing testing in a *SET2* deletion strain (*set2Δ*) of nuclear export signals fused to Set2 with and without inactivating mutations of the Set2 NLS (Set2_{NLSΔ}), as well as testing of the designed NLS used in LANS. Constructs are mVenus-tagged and migrate around 115 kDa, whereas the endogenous Set2 is not mVenus-tagged and migrates around 85 kDa. The asterisk indicates a nonspecific band. (D) Confocal images demonstrating localization of several mVenus-tagged Set2 constructs (cyan) in yeast cells with histone H2B endogenously tagged with mCherry (magenta); (scale bar, 3 μm). (E) Light-mediated control of H3K36 tri- and dimethylation in *set2Δ* cells as evidenced by immunoblotting with antibodies specific to various histone modifications. Several constructs are mVenus-tagged whereas endogenous Set2 (endog.) and LANS-Set2 are not. Asterisks indicate nonspecific bands. (F) *FLO8-HIS3* cryptic transcription initiation spotting assay testing LANS-Set2 and constructs toward its development. Fourfold serial dilutions of overnight cultures expressing one of several constructs

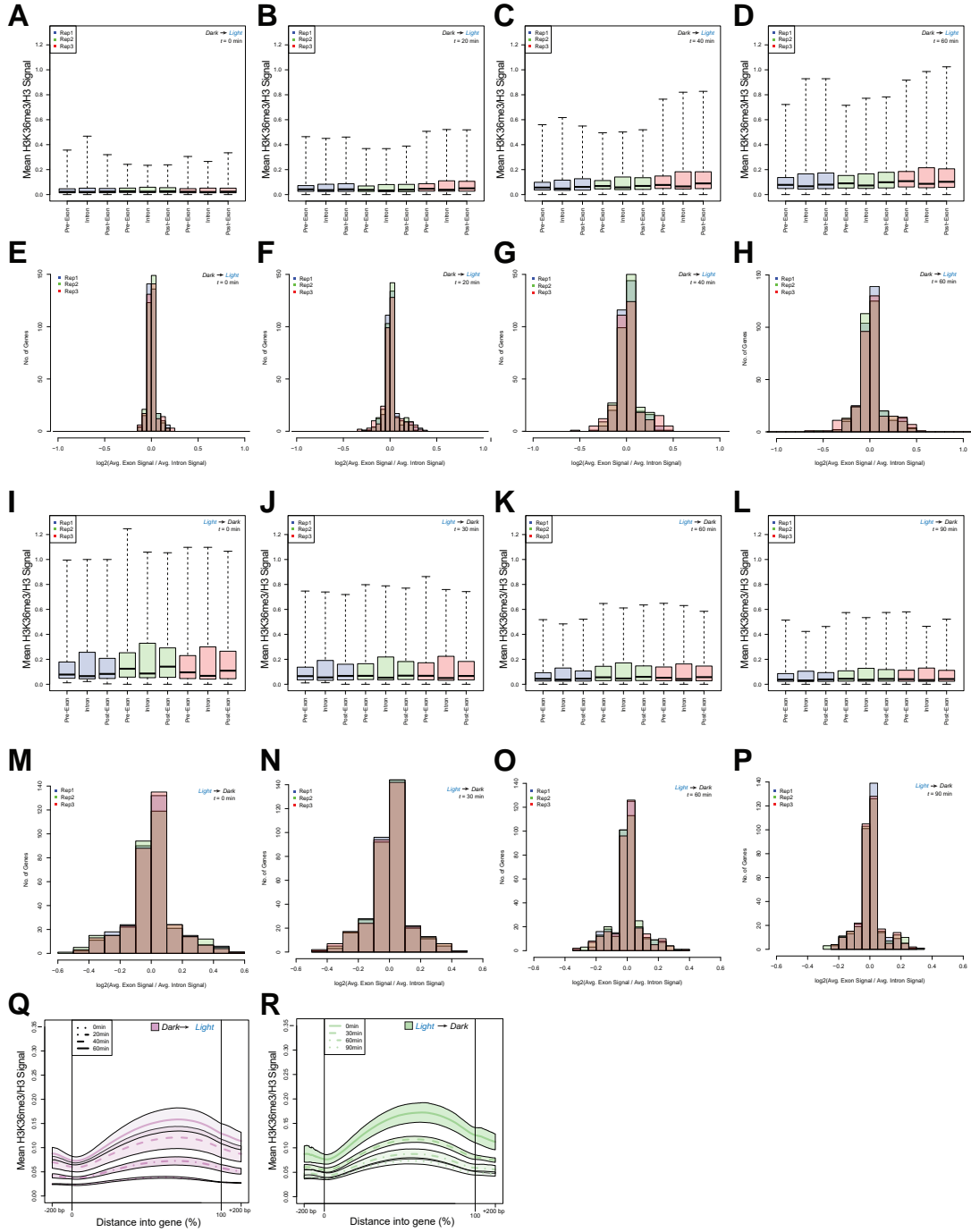
were spotted on solid media with or without histidine as per the *FLO8-HIS3* reporter assay depicted in Figure 2C, and plates were incubated in the dark or blue light. Images were collected 3-4 days later. All constructs are mVenus-tagged. (G) *bur1* Δ bypass spotting assay testing LANS-Set2 and constructs toward its development. Fivefold serial dilutions of overnight cultures of wild-type BY4742 and *BUR1* plasmid shuffling strains were spotted on solid media with or without 5-FOA to select against the *BUR1/URA3* plasmid and plates were incubated in the dark or blue light. Images were collected 2-3 days later. All constructs are mVenus-tagged.



1.2 Supplemental Fig S2: Related to Figure 2.

(A) Western blots probing gain of H3K36 methylation over time using LANS-Set2 in *set2Δ* after the transition from dark to light. Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantification of histone modifications as a function of time in Figure 2F. Asterisks indicate nonspecific bands. (B) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2Δ* after the transition from light to dark. Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantification of histone modifications as a function of time in Figure 2H. Asterisks indicate nonspecific bands. (C) Western blots probing loss of H3K36 methylation over time using Set2-FRB after addition of rapamycin. Triplicate immunoblots of whole-cell lysates prepared from log phase cultures. (D) Quantification of histone modifications as a function of time from immunoblots in (C). ($n = 3$, mean \pm SEM). (E) Immunoblots comparing levels of H3K36 methylation in Set2-FRB cultures grown with and without rapamycin exposure. (F) Quantification of histone modifications from immunoblots in (E). Data represent mean values \pm SD ($n = 3$ independent experiments). (*) $P < 0.05$; (***) $P < 0.0001$. (G) Quantification of H3K36me3 loss using LANS-Set2 in *set2Δ* after transition from light to dark (from Figure 2H) and using

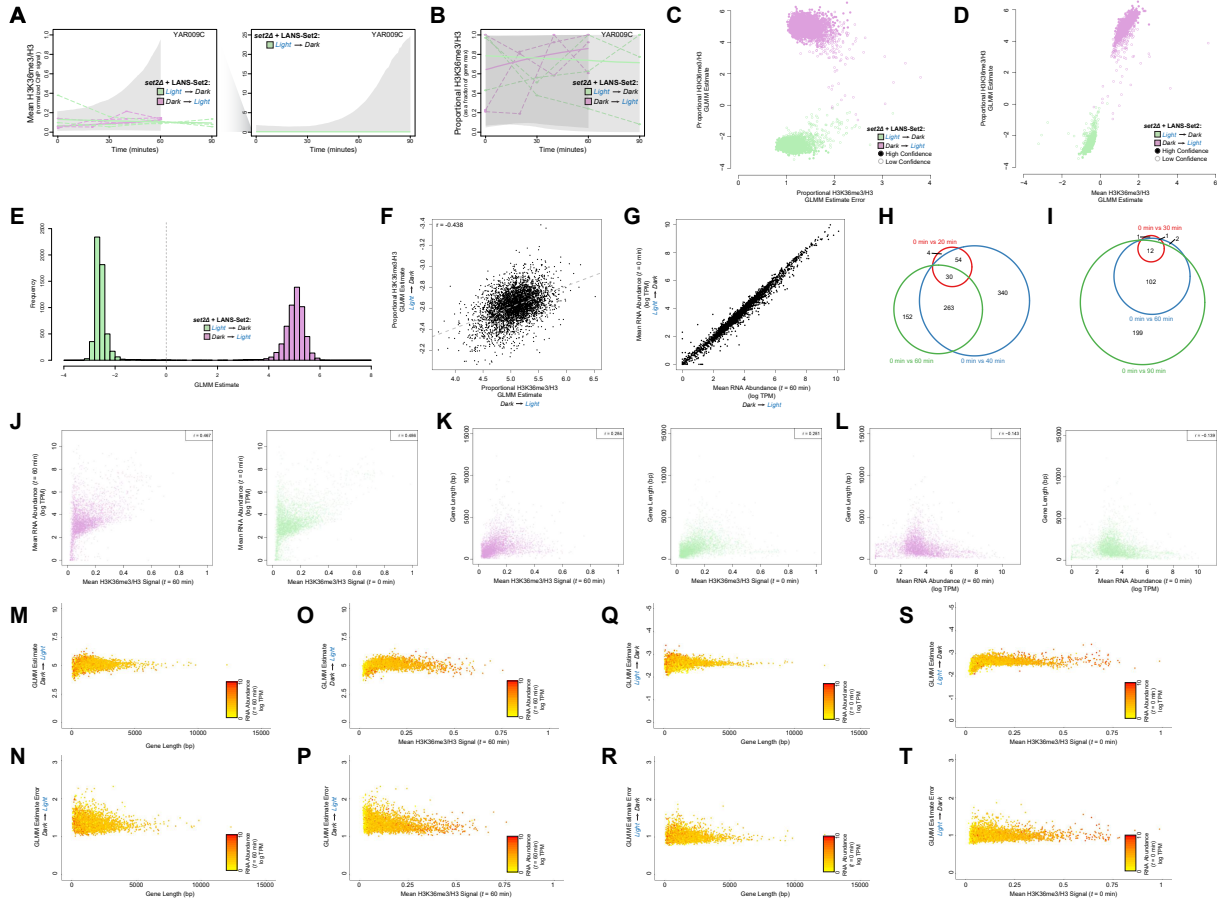
Set2-FRB after addition of rapamycin (from Supplemental Fig S2D). (H) Quantification of H3K36me2 loss using LANS-Set2 in *set2* Δ after transition from light to dark (from Figure 2H) and using Set2-FRB after addition of rapamycin (from Supplemental Fig S2D). (I) Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for wild-type BY4741 yeast and Set2-FRB without exposure to rapamycin. Asterisks indicate nonspecific bands. (J) Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for wild-type BY4741 yeast and LANS-Set2 in *set2* Δ grown continuously in the light. Asterisks indicate nonspecific bands.



1.3 Supplemental Fig S3: Related to Figure 3.

(A,B,C,D) Boxplots representing the distribution of all mean, normalized H3K36me3 ChIP-seq signal for the pre-exons, introns, and post-exons at (A) $t = 0$ minutes, (B) $t = 20$ minutes, (C) $t = 40$ minutes, and (D) $t = 60$ minutes after LANS-Set2 activation. A pre-exon is the exon preceding a given intron while the post-exon is the succeeding exon to a given intron. Blue boxplots are replicate 1, green boxplots are replicate 2, and blue boxplots are replicate 3. Boxplot borders represent the 2nd and 3rd quartile, while the midline reflects the median. (E,F,G,H) Histograms representing the \log_2 fold change of mean, normalized

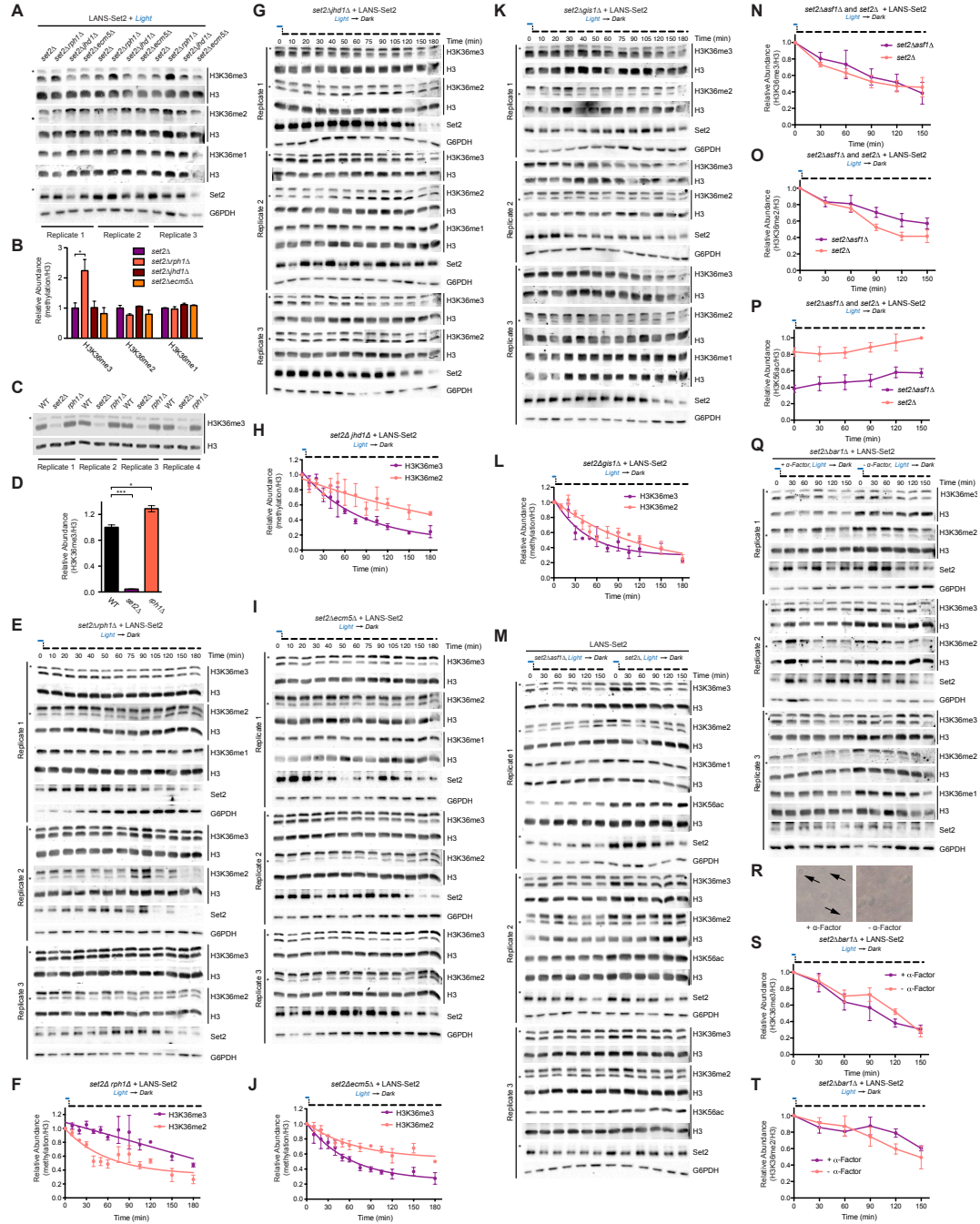
H3K36me3 ChIP-seq signal between a pre-exon and the corresponding intron at (E) $t = 0$ minutes, (F) $t = 20$ minutes, (G) $t = 40$ minutes, and (H) $t = 60$ minutes after LANS-Set2 activation. Blue histograms represent replicate 1, green histograms represent replicate 2, and blue histograms represent replicate 3. (I,J,K,L) Boxplots representing the distribution of all mean, normalized H3K36me3 ChIP-seq signal for the pre-exons, introns, and post-exons (I) $t = 0$ minutes, (J) $t = 30$ minutes, (K) $t = 60$ minutes, and (L) $t = 90$ minutes after LANS-Set2 inactivation. (M,N,O,P) Histograms representing the \log_2 fold change of mean, normalized H3K36me3 ChIP-seq signal between a pre-exon and the corresponding intron at (M) $t = 0$ minutes, (N) $t = 30$ minutes (O) $t = 60$ minutes, and (P) $t = 90$ minutes after LANS-Set2 inactivation. (Q,R) Mean, normalized H3K36me3 ChIP-seq signal metagene for the time course after (Q) LANS-Set2 activation or (R) LANS-Set2 inactivation, scaled between 0-100% of the gene length. Metagenes contain 200 base pairs up- and downstream of the genes. Bold lines represent the mean across all three replicates for each time point, while the shaded bands represent the standard deviation about the mean. The different line types indicate the time points.



1.4 Supplemental Fig S4: Related to Figure 4.

The (A) mean, normalized H3K36me3 ChIP-seq signal and (B) relative H3K36me3 ChIP-seq signal for the gene YAR009C throughout the time course after LANS-Set2 activation (green) and LANS-Set2 inactivation (purple). The gene YAR009C was selected at random from the gene set not included in the high confidence gene set from Figure 4B in order to show that the model discerns between high and low confidence gene. The gene YAR009C serves as an example of a gene lacking a clear trend in the data and exhibiting wide credible intervals for either normalized or relative H3K36me3. Dashed lines represent individual replicates, while bold lines represent the Bayesian generalized linear mixed effect model (GLMM) posterior mean H3K36me3. The shaded regions represent the posterior credible interval for H3K36me3 from the GLMM. The credible interval for LANS-Set2 activation was plotted on a separate plot (right) to accommodate the large difference in the y-axis scale. (C) Per-gene GLMM posterior rates for relative H3K36me3 ChIP-seq signal for LANS-Set2 activation (green) and LANS-Set2 inactivation (purple) and their respective errors. Solid circles represent high confidence genes for LANS-Set2 activation or LANS-Set2 inactivation, while hollow circles represent genes with low confidence. (D) Per-gene GLMM posterior rates for normalized H3K36me3 ChIP-seq signal and relative H3K36me3 ChIP-seq signal for both LANS-Set2 activation (green) and LANS-Set2 inactivation (purple). All genes were included, whereas some were excluded in Figure 4C for clarity. (E) Histogram of the GLMM rates for of all genes for LANS-Set2 activation (green) and LANS-Set2 inactivation (purple). (F) GLMM rate comparison between LANS-Set2 inactivation and LANS-Set2 activation for the high confidence gene set. Pearson correlation coefficient is included. (G) Comparison of mean RNA abundance levels (log TPM) between LANS-Set2 inactivation and LANS-Set2 activation at

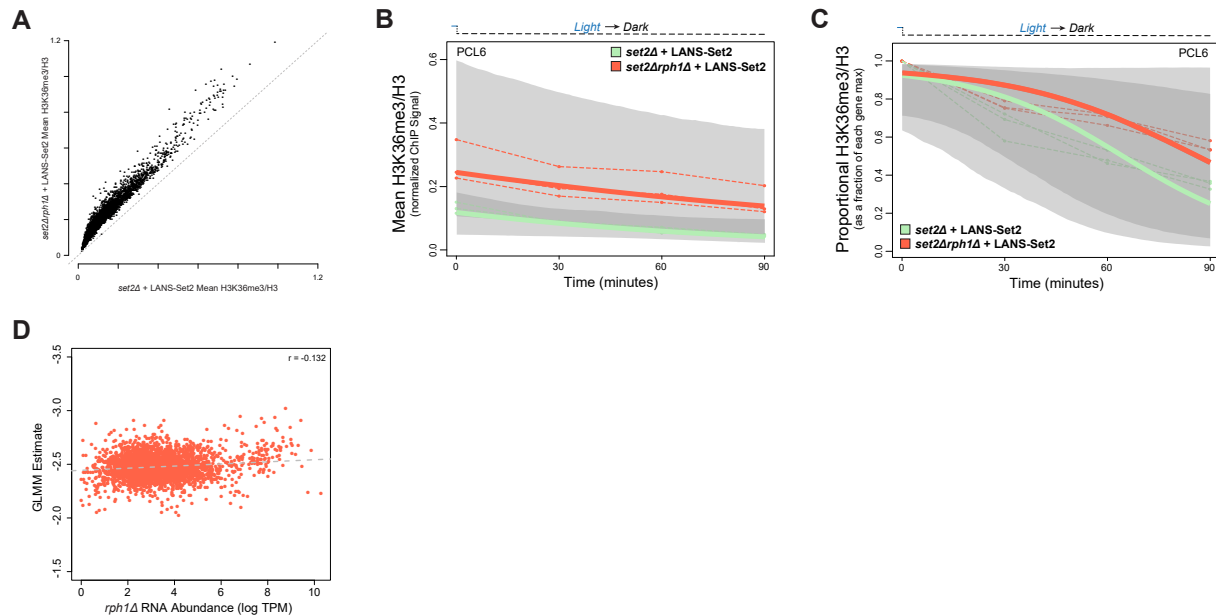
$t = 0$ minutes and $t = 60$ minutes, respectively, for high confidence gene set. (H) Venn diagram of genes with significantly different RNA abundances between $t = 0$ minutes and $t = 30$ minutes (red), $t = 60$ minutes (blue), and $t = 90$ minutes (green) after LANS-Set2 activation. (I) Venn diagram of differential genes between $t = 0$ minutes and $t = 20$ minutes (red), $t = 40$ minutes (blue), and $t = 60$ minutes (green) after LANS-Set2 inactivation. (J) Mean RNA abundance levels (log TPM) by mean, normalized H3K36me3 levels for the high confidence gene set for both LANS-Set2 activation at $t = 60$ minutes (left) and LANS-Set2 inactivation at $t = 0$ minutes (right). These time points were chosen as they are the ones with the highest H3K36me3 level, and thus most like wild-type conditions. (K) Gene length by mean, normalized H3K36me3 levels for the high confidence gene set for both LANS-Set2 activation at $t = 60$ minutes (left) and LANS-Set2 inactivation at $t = 0$ minutes (right). (L) Gene length by mean RNA abundance levels (log TPM) for the high confidence gene set for both LANS-Set2 activation at $t = 60$ minutes (left) and LANS-Set2 inactivation at $t = 0$ minutes (right). (M) The GLMM rate by gene length for the high confidence gene set during LANS-Set2 activation. (N) GLMM rate errors by gene length for the high confidence gene set during LANS-Set2 activation. (O) The GLMM rate by mean, normalized H3K36me3 levels for the high confidence gene set during LANS-Set2 activation. (P) The GLMM rate errors by mean, normalized H3K36me3 levels for the high confidence gene set during LANS-Set2 activation. Genes are colored based on their RNA abundance at $t = 60$ minutes (log TPM) (M,N,O,P). (Q) The GLMM rate by gene length for the high confidence gene set during LANS-Set2 inactivation. (R) GLMM rate errors by gene length for the high confidence gene set during LANS-Set2 inactivation. (S) The GLMM rate by mean, normalized H3K36me3 levels for the high confidence gene set during LANS-Set2 inactivation. (T) The GLMM rate errors by mean, normalized H3K36me3 levels for the high confidence gene set during LANS-Set2 inactivation. Genes are colored based on their RNA abundance at $t = 0$ minutes (log TPM) (Q,R,S,T).



1.5 Supplemental Fig S5: Related to Figure 5.

(A) Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for the indicated strains transformed with LANS-Set2 and grown continuously in the light. Asterisks indicate nonspecific bands. (B) Quantification of Western blots in (A). Data represent mean values \pm SD ($n = 3$). (*) $P < 0.05$. (C) Replicate immunoblots of whole-cell lysates prepared from log phase cultures for the indicated strains. (D) Quantification of Western blots in (C). Data represent mean values \pm SD ($n = 4$). (*) $P < 0.05$; (***) $P < 0.0001$. (E) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2Δrph1Δ*. Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantifica-

tion of histone modifications as a function of time after the transition from the light to dark. Asterisks indicate nonspecific bands. (F) Quantification of immunoblots in (E); ($n = 3$, mean \pm SEM). (G) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2 Δ jhd1 Δ* . Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantification of histone modifications as a function of time after the transition from the light to dark. Asterisks indicate nonspecific bands. (H) Quantification of immunoblots in (G); ($n = 3$, mean \pm SEM). (I) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2 Δ ecm5 Δ* . Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantification of histone modifications as a function of time after the transition from the light to dark. Asterisks indicate nonspecific bands. (J) Quantification of immunoblots in (I); ($n = 3$, mean \pm SEM). (K) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2 Δ gis1 Δ* . Triplicate immunoblots of whole-cell lysates prepared from log phase cultures for quantification of histone modifications as a function of time after the transition from the light to dark. Asterisks indicate nonspecific bands. (L) Quantification of immunoblots in (K); ($n = 3$, mean \pm SEM). (M) Western blots probing loss of H3K36 methylation over time using LANS-Set2 in *set2 Δ* and *set2 Δ asf1 Δ* . Triplicate immunoblots of whole-cell lysates prepared from log phase cultures after the transition from the light to dark. Asterisks indicate nonspecific bands. (N) Quantification of H3K36me3 from immunoblots in (M); ($n = 3$, mean \pm SEM). (O) Quantification of H3K36me2 from immunoblots in (M); ($n = 3$, mean \pm SEM). (P) Quantification of H3K56ac from immunoblots in (M); ($n = 3$, mean \pm SEM). (Q) Western blots probing loss of H3K36 methylation over time with and without α -factor treatment using LANS-Set2 in *set2 Δ bar1 Δ* . Triplicate immunoblots of whole-cell lysates prepared from log phase cultures grown with and without pre-treatment with α -factor to arrest cells in G1 prior to transition from the light to dark. Asterisks indicate nonspecific bands. (R) Representative widefield microscopy images of *set2 Δ bar1 Δ* cells with and without α -factor pre-treatment. Arrows point to the aberrant morphology (“shmoo” shape) that develops after α -factor treatment of MATa cells. (S) Quantification of H3K36me3 from immunoblots in (Q); ($n = 3$, mean \pm SEM). (T) Quantification of H3K36me2 from immunoblots in (Q); ($n = 3$, mean \pm SEM).



1.6 Supplemental Fig S6: Related to Figure 6.

(A) Comparison of mean, normalized H3K36me3 ChIP-seq signal for high confidence gene set of LANS-Set2 inactivation between *set2Δ* and *set2Δrph1Δ* backgrounds. The (B) normalized H3K36me3 ChIP-seq signal and (C) relative H3K36me3 ChIP-seq signal for the gene *PCL6* (YAL012W) throughout the time courses of LANS-Set2 inactivation in the backgrounds of *set2Δ* (green) or *set2Δrph1Δ* (red). Dashed lines represent individual replicates, while bold lines represent the posterior mean H3K36me3 from the Bayesian generalized linear mixed effect model (GLMM). The shaded regions represent the posterior credible interval for H3K36me3 from the GLMM. (D) Comparison of the LANS-Set2 inactivation GLMM rate in *set2Δrph1Δ* cells and RNA abundance levels (log TPM) in *rph1Δ* cells across the high confidence gene set shared between *set2Δ* and *set2Δrph1Δ* backgrounds. Dashed line represents the line of best fit. Pearson correlation coefficient is included ($r = -0.132$).

2 Supplemental Tables

2.1 Supplemental Table S1: Yeast strains used in this study.

Name	Genotype	Source
BY4741	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0</i>	Open Biosystems, Huntsville, Alabama
<i>set2Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH</i>	this study
H2B-mCherry	MATa <i>trp1Δ63 leu2Δ1 ura3-52 his3Δ200 lys2-801</i> H2B-mCherry::NatMX	Gift of Kerry Bloom
KLY78	MATx, <i>ura3-52 leu2Δ1 or leu1- trp1Δ63 or trp1- his3Δ200</i> KanMX6::GAL1pr-flo8::HIS3 <i>lys2-128Δ</i>	Silva et al, J. Biol. Chem., 2012
KLY78 <i>set2Δ</i> (YSM138)	MATx, <i>ura3-52 leu2Δ1 or leu1- trp1Δ63 or trp1- his3Δ200</i> KanMX6::GAL1pr-flo8::HIS3 <i>lys2-128Δ</i> , <i>set2Δ::NatMX</i>	Stephen McDaniel, unpublished
BY4742	MATa, <i>his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0</i>	Open Biosystems, Huntsville, Alabama
<i>bur1Δ BUR1</i> shuffle (YSB788)	MATa, <i>ura3-52 leu2Δ trp1Δ63 his3Δ200 lys2Δ202 bur1Δ::HIS3 (pRS316-Bur1)</i>	Keogh et al., 2003
<i>set2Δbur1Δ BUR1</i> shuffle (YSB1003)	MATa, <i>ura3-52 or 3Δ0 leu2Δ1 or 2Δ0 trp1Δ63 his3Δ200 or 3Δ1 lys2Δ202 bur1Δ::HIS3 set2Δ::KanMX (pRS316-Bur1)</i>	Keogh et al., 2005
Set2-FRB	MATa <i>ade2-1 trp1-1 can1-100 leu2-3, 112 his3-11, 15 ura3 tor1-1 frp1::NAT RPL13A-2XFKBP12::TRP1 Set2-FRB::KanMX6</i>	this study
<i>set2Δrph1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH rph1Δ::KanMX</i>	this study
<i>set2Δjhd1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH jhd1Δ::NatMX</i>	this study
<i>set2Δecm5Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH ecm5Δ::KanMX</i>	this study
<i>set2Δgis1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH gis1Δ::KanMX</i>	this study
<i>rph1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 rph1Δ::KanMX</i>	this study
<i>set2Δbar1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::HPH bar1Δ::KanMX</i>	this study
<i>set2Δasf1Δ</i>	MATa, <i>his3Δ0 leu2Δ0 met15Δ0 ura3Δ0 set2Δ::NatMX bar1Δ::HPH</i>	Stephen McDaniel, unpublished

2.2 Supplemental Table S2: Plasmids used in this study and their Addgene deposition IDs.

Name	Insert	Addgene ID
mVenus-FLAG-LANS-Set2	LANS-Set2	122002
FLAG-LANS-Set2	LANS-Set2	122003

3 Supplemental Data and Code

3.1 Supplemental Data S1

Relative H3K36 methylation data across genes, replicates, and four time points for LANS-Set2 activation (“writer add”) and LANS-Set2 inactivation, in both *set2*Δ (“writer loss”) and *set2*Δ*rph1*Δ (“writer eraser loss”) backgrounds. For a given gene, replicate, and LANS-Set2 status, the maximum signal value over all timepoints was set to 1, and the other time points as fractions of that maximum. (Supplemental_Data_S1.txt)

3.2 Supplemental Data S2

H3K36 methylation data across genes, replicates, and four time points for LANS-Set2 activation (“writer add”) and LANS-Set2 inactivation, in both *set2*Δ (“writer loss”) and *set2*Δ*rph1*Δ (“writer eraser loss”) backgrounds. (Supplemental_Data_S2.txt)

3.3 Supplemental Data S3

ZOI beta GLMM H3K36 methylation rates and fit statistics per gene per LANS-Set2 status. Models fit for relative H3K36 methylation data, described in detail in Supplemental Statistical Methods. (Supplemental_Data_S3.txt)

3.4 Supplemental Data S4

ZI negative binomial GLMM H3K36 methylation rates and fit statistics per gene per LANS-Set2 status. Models fit for H3K36 methylation data, described in detail in Supplemental Statistical Methods. (Supplemental_Data_S4.txt)

3.5 Supplemental Data S5

RNA abundance data across genes, replicates, and four time points for LANS-Set2 activation (“writer add”). (Supplemental_Data_S5.txt)

3.6 Supplemental Data S6

RNA abundance data across genes, replicates, and four time points for LANS-Set2 inactivation (“writer loss”). (Supplemental_Data_S6.txt)

3.7 Supplemental Data S7

Difference in H3K36 methylation at four time points between LANS-Set2 inactivation *set2*Δ (“writer loss”) and *set2*Δ*rph1*Δ (“writer eraser loss”) backgrounds, using a ZOI beta GLMM, and fit statistics per gene. Models fit for relative H3K36 methylation data, described in detail in Supplemental Statistical Methods. (Supplemental_Data_S7.txt)

3.8 Supplemental Data S8

Annotations for genes analyzed in study, used to determine gene length. (Supplemental_Data_S8.txt)

3.9 Supplemental Data S9

RNA abundance data across genes and replicates from a *rph1*Δ background. (Supplemental_Data_S9.txt)

3.10 Supplemental Code S1

R code to generate figures included in manuscript and Supplemental Figures. (Supplemental_Code_S1.txt)

4 Additional Files

4.1 Supplemental Movie S1: LANS-Set2 light activation and reversion in *Saccharomyces cerevisiae*.

Still images from this video are shown in Figure 1B and quantification of nuclear/cytoplasmic fluorescence intensity change before, during, and after light activation is shown in Figure 1C. Activation is performed on the entire field of view and the appearance and disappearance of a blue circle indicate the time of blue light activation. Scale bar is 3 μm . (Supplemental_Movie_S1.mov)

4.2 Supplemental Note: Protein amino acid sequences for mVenus-FLAG-LANS-Set2 and FLAG-LANS-Set2.

mVenus-FLAG-LANS-Set2 sequence (NLSΔ mutations in red; size: 132.6 kDa):

```
MKALT VSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTCLKICTTGKLPVPWPT
LVTTLGYGLQCFARYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVN
RIELKGIDFKEDGNILGHKLEYNNSHNVYITADKQKNGIKANFKIRHNIEDGGVQLADHYQQNT
PIGDGPVLLPDNHYLSYQSKLSKDPNEKRDHMLLEFVTAAGITLGMDELYKSGSGSGGRDYKD
DDDKSR LALKLAGLDIGLQSGSGSLATTLEIEKNFVITDPRLPDNPIIFASDSFLQLTEYSREEI
LGRNCRFLQGPETDRATVRKIRDAIDNQTEVTVQLINYTKSGKKFWNLFHLQPMRDQKGDVQY
FIGVQLDGT EHV RDA AEREGVMLIKKTAENIDMAAKRVKLDGSGSGSGSRGMSKNQSVSASE
DEKEILNNAEGHKPQRLFDQEPDLTEEALTKFENLDDCIYANKRIGTFKNDFMECDCYEEFS
DGVNHACDESDCINRLTIECVNDLCSSCGNDCQNQRQKQYAPIAIFKTKHKGYGVRAEQ
DIEANQFIYEYKGEVIEEMEFRDRLIDYDQRHFHFKHYFMMLQNGEFIDATIKGSLARFCNHSCSP
NAYVNKWWVKDKLRMGIFAQRKILKGEEITFDYNVDYGAQAQKCYCEEPNCIGFLGGKTQTD
AASLLPQNIADALGVTVSMEKKWLKLLKLSGEPPIKNENENINIEFLQSLEVQPIDSPVDVTKIMSV
LLQQDNKIIASKLLKRLFTIDDDSLRHQAHLHYTCFSKMLKLFITEQPQVDGKGNETEEDDIKFI
KGILDFLLELPKTTNRNGIESSQIDNVVKTLPKFPFLKPNCDLLEKWSKFETYKRITKKDINVAAL
SKMIDLRRVRLPPGWEIHHENGRPLYNAEQKTKLHYPPSGSSKVFSSRSNTQVNSPSSSGIPK
TPGALDS SSHKLSDEEYE SSGQKRLEYERIALERAKQEELES LKQKLKLENERKSVLEDIIAEAN
KQKELQKEEAKKLVEAKEAKRLKRKTVSQSRLEHNWNKFFASFVPNLIKKNPQSKQFDHENI
KQCAKDIVKILTTKELKKDSSRAPDDLTKGKRHKVKEFINSYMDKIILKKKQKKALALSSASTRM
SSPPPSTSS
```

FLAG-LANS-Set2 sequence (NLSΔ mutations in red; size: 105.7 kDa):

```
MKALTSGGRDYKDDDDKSR LALKLAGLDIGLQSGSGSLATTLEIEKNFVITDPRLPDNPIIFA
SDSFLQLTEYSREEILGRNCRFLQGPETDRATVRKIRDAIDNQTEVTVQLINYTKSGKKFWNLFH
LQPMRDQKGDVQYFIGVQLDGT EHV RDA AEREGVMLIKKTAENIDMAAKRVKLDGSGSGSGS
RGMSKNQSVSASEDEKEILNNAEGHKPQRLFDQEPDLTEEALTKFENLDDCIYANKRIGTFKN
NDFMECDCYEEFS DGVNHACDESDCINRLTIECVNDLCSSCGNDCQNQRQKQYAPIAIF
KTKHKGYGVRAEQDIEANQFIYEYKGEVIEEMEFRDRLIDYDQRHFHFKHYFMMLQNGEFIDATIK
GSLARFCNHSCSPNAYVNKWWVKDKLRMGIFAQRKILKGEEITFDYNVDYGAQAQKCYCEEP
NCIGFLGGKTQTD AASLLPQNIADALGVTVSMEKKWLKLLKLSGEPPIKNENENINIEFLQSLEVQ
PIDSPVDVTKIMSVLLQQDNKIIASKLLKRLFTIDDDSLRHQAHLHYTCFSKMLKLFITEQPQVD
GKGNETEEDDIKFIKGILDFLLELPKTTNRNGIESSQIDNVVKTLPKFPFLKPNCDLLEKWSKFE
TYKRITKKDINVAASKMIDLRRVRLPPGWEIHHENGRPLYNAEQKTKLHYPPSGSSKVFSSRSN
TQVNSPSSSGIPKTPGALDS SSHKLSDEEYE SSGQKRLEYERIALERAKQEELES LKQKLKLENERKSVLEDIIAEAN
KQKELQKEEAKKLVEAKEAKRLKRKTVSQSRLEHNWNKFFASFVPNLIKKNPQSKQFDHENI
KQCAKDIVKILTTKELKKDSSRAPDDLTKGKRHKVKEFINSYMDKIILKKKQKKALALSSASTRM
SSPPPSTSS
```


5 Supplemental Statistical Methods

5.1 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 LANS-Set2 (writer add:WA), and after dark inactivation of LANS-Set2 with and without the presence of Rph1, 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 5,355 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, which we refer to as “relative H3K36me3 signal” in the text. Second, we took the measurements, previously standardized based on spike-in samples and normalized to adjust for gene length, multiplied them by 1,000, and rounded to an integer, resulting in count data (quasi-counts) with a distribution proportional to the starting data.

5.2 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 (Gelman et al. 2013; Gelman and Hill 2006).

The Stan statistical platform (Carpenter et al. 2017) is one such computational tool for fitting complex Bayesian hierarchical models. We used the brms software package (Bürkner 2017, 2018), which acts as a wrapper of Stan in the R statistical programming language (R Core Team 2019). 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 ($k = 1, 2, \dots, 5355$) for l^{th} condition ($l = 1, 2, 3$). Briefly, we model y with a GLM by defining $E(y) = g^{-1}(\eta)$, where $E(\cdot)$ is the expected value of a random variable, $g^{-1}(\cdot)$ 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.

5.3 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 (Slob 2002; Wilson et al. 2014). Because the DR model implementations did not easily accommodate covariates or the replicate observations and also fit relatively complex models (\leq five parameters), we instead modeled the data with a zero-one-inflated beta (ZOIB) distribution with the following parameterization:

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

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 zero), $B(\cdot)$ is the beta function (Casella and Berger 2002), 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 $y \in (0, 1)$, *i.e.*, 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 one 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).

5.4 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:

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

ξ is the zero-inflation probability, $\text{NB}(\cdot)$ is the non-zero-inflated negative binomial probability mass function, and ϕ is a positive precision parameter. As $\phi \rightarrow \infty$, 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.

5.5 Time models

5.5.1 Continuous time model

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

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

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 (Gelman and Hill 2006), effects that account for the longitudinal nature of the data, and modeled with the following priors: $N(0, \tau_u^2)$ and $N(0, \tau_v^2)$, respectively. For all genes and conditions, we recorded the posterior mean ($\hat{\beta}_{\text{time},kl}$) 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 zero) for a given condition. $\hat{\beta}_{\text{time},kl}$ were plotted and correlated with additional covariates of interests in order to identify potential relationships with other factors of interest.

5.5.2 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 $\hat{\beta}_{\text{time},kl}$, 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:

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

where u_k is a gene-specific random effect, modeled as $N(0, \tau_u^2)$, 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 $N(0, \frac{\sigma^2}{w_{ij}})$ 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 (Bates et al. 2015). Through ANOVA (Venables and Ripley 2010) 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 (Lenth 2019), we performed Tukey *post hoc* tests of pairwise differences (Venables and Ripley 2010) 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.

5.5.3 Categorical time model

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:

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

where μ is the intercept representing, representing WL at time0, ω_{WL} is the deviation of WEL from WL, $\beta_{\text{timepoint}}^p$ is the effect comparing timepoint p to zero for WL, and $\delta_{\text{timepoint}}^q$ 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: $N(0, \tau_u^2)$, $N(0, \tau_w^2)$, $N(0, \tau_{v,p}^2)$, and $N(0, \tau_{z,p}^2)$, respectively. Similar to the continuous time model, we recorded posterior means and 95% CrI for $\delta_{\text{timepoint}}^p$. We fit the categorical time model with the quantile data and ZOIB, though it could be used with the quasi-counts and ZINB as well.

5.5.4 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.

5.6 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 (*i.e.*, 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 three degrees of freedom and a minimal scale parameter of 10 (Gelman and Hill 2006), 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 (Lewandowski et al. 2009).

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 adaptive delta parameter, necessary for the No-U-turn Sampler (Hoffman and Gelman 2014) 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- \hat{R} statistic (Gelman et al. 2013), a ratio of the average variance within chains to the variance with the pooled chains. A split- \hat{R} value close to one 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- $\hat{R} > 1.1$ 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- $\hat{R} < 1.1$ 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- $\hat{R} > 1.1$, 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

- 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. doi: 10.18637/jss.v067.i01.
- Paul-Christian Bürkner. brms : An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1), 2017. ISSN 1548-7660. doi: 10.18637/jss.v080.i01. URL <http://www.jstatsoft.org/v80/i01/>.

- Paul-Christian Bürkner. Advanced Bayesian Multilevel Modeling with the R Package brms. *The R Journal*, 10(1):395–411, 2018. doi: 10.32614/RJ-2018-017. URL <https://doi.org/10.32614/RJ-2018-017>.
- 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. ISSN 1548-7660. doi: 10.18637/jss.v076.i01. URL <http://www.jstatsoft.org/v76/i01/>.
- G. Casella and R.L. Berger. *Statistical Inference*. Duxbury advanced series in statistics and decision sciences. Thomson Learning, 2002. ISBN 9780534243128. URL https://books.google.com/books?id=0x_vAAAAAAJ.
- 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. ISBN 9781439840955. URL <https://books.google.com/books?id=ZXL6AQAAQBAJ>.
- Andrew Gelman and Jennifer Hill. *Data Analysis Using Regression and Multilevel/Hierarchical Models*. Cambridge University Press, 2006. ISBN <https://doi.org/10.1017/CBO9780511790942>.
- 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. ISSN 1532-4435. URL <http://dl.acm.org/citation.cfm?id=2627435.2638586>.
- Russell Lenth. *emmeans: Estimated Marginal Means, aka Least-Squares Means*, 2019. URL <https://CRAN.R-project.org/package=emmeans>. R package version 1.3.4.
- 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. ISSN 0047259X. doi: 10.1016/j.jmva.2009.04.008. URL <http://dx.doi.org/10.1016/j.jmva.2009.04.008><https://linkinghub.elsevier.com/retrieve/pii/S0047259X09000876>.
- R Core Team. RSoftware2019, 2019.
- Wout Slob. Dose-Response Modeling of Continuous Endpoints. *Toxicological Sciences*, 66(2):298–312, apr 2002. ISSN 10960929. doi: 10.1093/toxsci/66.2.298. URL <https://academic.oup.com/toxsci/article-lookup/doi/10.1093/toxsci/66.2.298>.
- W. N. Venables and B. D. Ripley. *Modern Applied Statistics with S*. Springer Publishing Company, Incorporated, 2010. ISBN 1441930086, 9781441930088.
- 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. ISSN 0006341X. doi: 10.1111/biom.12114. URL <http://doi.wiley.com/10.1111/biom.12114>.