H3K27me3 natural variation selectively marks genes predicted to be important for differentiation in unicellular algae

Xue Zhao[1,3], Achal Rastogi[1†], Anne Flore Deton Cabanillas[1], Ouardia Ait Mohamed[1], Catherine Cantrel[1], Berangère Lombard[2], Omer Murik[1¥], Auguste Genovesio[1], Chris Bowler[1], Daniel Bouyer[1], Damarys Loew[2], Xin Lin[1 §], Alaguraj Veluchamy[4], Fabio Rocha Jimenez Vieira[1] and Leila Tirichine[1,3*]

[1]Institut de Biologie de l'ENS (IBENS), Département de biologie, École normale supérieure, CNRS, INSERM, Université PSL, 75005 Paris, France

[2]Institut Curie, PSL Research University, Centre de Recherche, Laboratoire de Spectrométrie de Masse Protéomique, 26 rue d'Ulm 75248 Cedex 05 Paris, France

[3]CNRS UMR6286, UFIP UFR Sciences et Techniques, Université de Nantes, 2 rue de la Houssinière 44322, Nantes Cedex 03

[4]Laboratory of Chromatin Biochemistry, BESE Division Building 2, Level 3, Office B2-3327, 4700 King Abdullah University of Science and Technology (KAUST), Thuwal 23955-6900, Kingdom of Saudi Arabia

[†]Present address: Corteva Agriscience, Madhapur, Hyderabad 500 081, Telangana, India

[¥]Present address: Medical Genetics Institute, Shaare Zedek Medical Center, Jerusalem, Israel

[§] Present address: State Key Laboratory of Marine Environmental Science, Xiamen University

Abstract

In multicellular organisms H3K27me3 has been shown to be deposited by Polycomb Repressive Complex 2 (PRC2) to establish and maintain gene silencing, critical for cell fate and developmentally regulated processes. PRC2 complex is absent in both yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*, which initially suggested that PRC2 arose with the emergence of multicellularity. However, its discovery in several unicellular species including microalgae questions its role in unicellular eukaryotes. Here, we show in the model diatom *Phaeodactylum tricornutum* (Pt), using mutants in the homologue of the catalytic subunit of PRC2, *enhancer of zeste E(z)*, that Pt *E(z)* is responsible for di and tri-methylation of lysine 27 of histone H3. H3K27me3 depletion abolishes cell morphology in Pt providing evidence for a role of H3K27me3 in cell differentiation in unicellular species. Genome wide profiling of H3K27me3 in fusiform and triradiate cells further revealed genes that may specify cell identity. These results suggest a role for PRC2 and its associated histone mark in cell differentiation in unicellular species and highlights their ancestral function in a broader evolutionary context than is currently appreciated.

Introduction

Tri-methylation of lysine 27 of histone H3 (H3K27me3) is a mark deposited by Polycomb Repressive Complex 2 (PRC2), which mediates silencing of gene expression during differentiation and development in both animals and plants [1-3]. PRC2 is comprised of four core proteins, highly conserved among multicellular organisms: the histone methyltransferase (HMTase) enhancer of zeste *E(z)*, the WD40 domain containing polypeptide Extra Sex Comb *Esc*, the C2H2 type zinc finger protein Suppressor of zeste 12 *Su(z)12* and the Nucleosome remodeling factor 55 kDa subunit *Nurf-55* [4,5]. The absence of PRC2 in the unicellular yeast models *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* initially led to suggestions that it arose to regulate cell differentiation in multicellular organisms [6]. This hypothesis has recently been questioned because components of PRC2 and the associated mark H3K27me3 are found in several unicellular species that belong to different lineages including, but not only, Stramenopiles, Alveolates and Rhizaria (SAR)[7-9] (Supplementary Fig. 1a-d), thus questioning the function of such a widespread complex in single celled organisms.

Attempts to understand the role of H3K27me3 in the unicellular green alga *Chlamydomonas reinhardtii* were not conclusive because tri-methylation of lysine 27 could not be assessed reliably due to its nominal mass which was found similar to acetylation of lysine 27 of the same histone[7]. However, we have previously identified H3K27me3 by mass spectrometry and mapped its localization genome-wide in the pennate diatom *Phaeodactylum tricornutum*[9] (Pt), which belongs to the stramenopile group of eukaryotes, only distantly related to the animal (Opisthokonta) and plant (Archaeplastida) eukaryotic crown groups. Pt has different morphotypes, fusiform (FM hereafter), which is the most prevailing morphology among the sampled accessions known so far, triradiate (TM hereafter), oval (OM hereafter) and cruciform (CM hereafter) [10-13] (Fig. 1a). Each morphotype can switch reversibly into a different morphology in response to several growth and environmental cues[10]. FM is the most stable morphotype while switching is more prominent in TM, CM and OM, which tend to convert to FM in the growth conditions used in this study[10,11].

Interestingly, western blot analysis using a monoclonal antibody against the mark in FM, TM, CM and OM cells revealed a strong correlation between the complexity of the morphology (branching of the cell) and the absolute quantity of H3K27me3, which is higher in both CM and TM cells compared to FM and OM cells (Supplementary Fig. 1e), suggesting that PRC2 activity controls cell differentiation in *P. tricornutum*. Cell differentiation is often orchestrated by H3K27me3-mediated silencing that underlies the establishment and maintenance of cellular

identity in multicellular model species [14]. These results prompted us to investigate the putative role of H3K27me3 in cell differentiation in a unicellular model. In silico annotation of polycomb complex members and identification of H3K27me3 by mass spectrometry[15] in a well-developed experimental model such as *P. tricornutum* (Supplementary Fig. 1a-c) might give new insights for the unsolved questions in the study of PRC2 complex in multicellular species, especially the availability of different morphotypes which is unique to *P. tricornutum* among stramenopiles, present an opportunity to decipher its role in single celled organisms with respect to its potential contribution to establish morphotype switches as well as its function in an evolutionary context.

To gain insights into the function of *E(z)* and its associated H3K27me3 mark in Pt, we generated two CRISPR/cas9 knockouts of the gene in each of the three morphotypes (FM, TM and CM) leading to putative loss of function mutations including premature stop codon and frameshifts (Fig. 1b). Light microscopy analysis of *E(z)* knockouts shows a change in cell morphology which becomes shorter in the FM background. Both triradiate and cruciform morphologies were abolished in TM and CM, respectively while transgenic lines carrying cas9 control vectors in each morphotype remain unchanged (Fig. 1c). Whereas *P. tricornutum* transgenic lines with the Cas9 control vector show similar H3K27me3 enrichment than the wild type (Supplementary Fig. 2a). knockout of *E(z)* led to an overall depletion of H3K27me3 shown by western blot using a monoclonal antibody against the mark (Fig. 1d), suggesting that the loss or diminution of *E(z)* activity, and hence H3K27me3, causes the observed changes in cell morphology (Fig. 1c; Supplementary Fig. 2a; Table S1). Overall, this suggests that *E(z)* and its associated mark are required for morphotype switch/to establish specific cell identity.

Mass spectrometry analysis of histones extracted from both wild type and *E(z)* knockout confirmed the loss of H3K27me3 and revealed a depletion of di-methylation of H3K27, corroborating the role of *E(z)* in di and tri-methylation of lysine 27 of histone H3 (Supplementary Fig. 2b,c,d). This is similar to fungi [16] and mammals[17] but different from *A. thaliana* where PRC2 loss of function leads to specific depletion of H3K27me3[18], although in vitro assays with reconstituted *A. thaliana* PRC2 components showed mono, di and tri-methylation of lysine 27 of histone H3[19,20]. Western blot analysis of *E(z)* knockout mutants using a monoclonal antibody against H3K27me2 (Supplementary Fig. 2e) confirmed the depletion of the mark from the mutants, supporting further the mass spectrometry analysis and the role of *E(z)* in di-methylation of lysine 27 of histone H3 in *P. tricornutum*.

Similar to animals, plants and *Neurospora crassa*, H3K27me3 appears not to be essential for cell survival in *P. tricornutum*, as indicated by the overall growth of *E(z)* knockout lines, which are only slightly retarded compared to wild type lines (Supplementary Fig. 2f).

To further investigate the role of H3K27me3 and its targets in different morphotypes, we carried out a genomic approach and performed Chromatin Immuno-Precipitation (ChIP) on two biological replicates of TM morphotypes using an antibody against H3K27me3 followed by DNA sequencing (ChIP-Seq) to generate a map of H3K27me3 distribution, which we compared to the one previously generated in FM[21]. ChIP-Seq data analysis revealed a similar H3K27me3 enrichment profile between TM and FM that localizes principally on transposable elements (TEs), with 58% and 60% of the reads overlapping with TE annotations for FM[21] and TM, respectively (Fig. 2a). The mark was found to occupy, on average, ~11.6% of the genome in FM cells, targeting approximately 15% of genes (consistent with [15]) and ~13.2% of the genome within TM, targeting 19% of genes in agreement with the absolute amount of H3K27me3 to be elevated in TM compared to FM as detected by western blot (Supplementary Fig. 1e). Indeed, more genes are marked by H3K27me3 in TM than in FM (Fig. 2a, Supplementary Fig. 3a), although most of the PRC2 targets are shared between both morphotypes (Fig. 2a) and exhibit globally broad coverage over the annotation (Supplementary Fig. 3b). Among the PRC2 target genes, 635 and 297 genes are found to be specifically marked by H3K27me3 in TM and FM, respectively (Fig. 2a). We used ChIP followed by quantitative PCR (ChIP-qPCR) to validate H3K27me3 enrichment over specifically marked loci in both backgrounds, which corroborated the genome wide data for most of the tested genes (Fig. 2b, c). Common marked loci and unmarked loci were tested as internal controls (Supplementary Fig. 3c). To test for the loss of H3K27me3 in *E(z)* knockouts, we performed ChIP-qPCR in FM and TM as well as the respective mutants, which confirmed the depletion of H3K27me3 in these mutants (Fig. 1e,f).

To gain insights into the functional categories enriched in H3K27me3 target genes that are shared between the two morphotype or specific, we applied to the updated Phatr3 annotation a gene ontology analysis using DAMA[22] and CLADE[23] which is a machine learning methodology that uses pHMMs, positive score matrix and support vector machines to infer the corresponding most probable GO category to genes. DAMA and Clade allow a more sensitive remote homology that permits to assign to genes with no or poor domain conservation the corresponding GO categories that would have been missed by other methods. Are considered statistically significant, only GO classes that are represented by at least 3 standard deviations above the average of observed entries.

Out of 1,640 H3K27me3 marked genes, 753 could not be assigned to a more specific GO category and are therefore marked as unknown. The genes that are marked in both morphotypes show top enrichment in RNA related biological processes such as RNA-dependent DNA biosynthetic process, RNA phosphodiester bond hydrolysis and RNA-DNA hybrid ribonuclease activity. Genes marked by H3K27me3 specifically in TM displayed top enrichment exclusively in (1) glycoprotein biosynthetic processes involved in the transfer of sugar moieties that might determine different sugar composition of the cell wall which is known to be sugar rich in *P. tricornutum*[24], (2) Peptidyl tyrosine dephosphorylation processes with Ankyrin repeats proteins known to act as scaffold for connecting molecular interactions, likely important for development of the numerous signaling pathways associated generally to more complex multicellular organisms[25] (Fig. 2d; Supplementary Table S2). Genes that are specifically marked in FM cells exhibit enrichment in categories such as peptidyl-tyrosine phosphorylation containing genes with central roles as modulators of cell differentiation and cell fate decisions[26] (Fig. 2d; Supplementary Table S2). Interestingly, additional genes specifically marked in each FM or TM share categories with predicted functions in positive regulation of (1) GTPase activity with a role in cell morphology changes, and neurite outgrowth and guidance[27] as well as the differentiation of many cell types, including neurons, T lymphocytes and myocytes[28]; (2) protein ubiquitination shown to play a role in the complex regulation of the levels and function of many proteins and signaling pathways involved in determining cell fate[29]. Overall, the genes that are specifically marked in the TM or FM morphotypes reflect processes related to cell growth, proliferation and differentiation.

We have reported previously that H3K27me3 marked genes in FM are characterized by low expression[15], consistent with the role of H3K27me3 as a repressive mark. Interestingly, when genes are marked by H3K27me3 in both FM and TM, their expression is lower compared to the genes that are uniquely marked in FM and to a lesser extent in TM (Fig. 3a,b). This suggests that specifically marked genes are kept under less stringent and tight repression which might be due to their putative role in morphotype switch, which is known to be a dynamic process[10].

Considering the conserved role of H3K27me3 in repression, we tested the effect of *E(z)* knockout on gene expression. Therefore, RNA sequencing (RNA-seq) of two biological replicates of the *E(z)* mutant (Del6) was carried out and compared to previously generated RNA-seq in the wild type (FM). Around 1/4 of all genes are (23%, 2795 out of 12152 annotations) differentially expressed in the *E(z)* mutant (P-value. < 0.05), (Supplementary Fig. 3d; Supplementary Table S2), indicating an essential role in gene regulation by PRC2 in Pt.

We further monitored by RT-qPCR the expression of 27 specifically marked genes in the *E(z)* knockout of the TM and found that 18 genes out of 27 showed a gain of expression in the mutant compared to the TM background, demonstrating further that depletion of H3K27me3 likely releases the repression of target genes and correlates with the loss of the triradiate morphology (Fig. 3c). Although the remaining genes showed no change or even a gain in expression, these genes can be targets of other repressive or active marks as shown previously with co-occurrence of several repressive marks over genic regions[9]. The analysis of the R value[30] which reflects the entropy, and therefore the variability in expression of genes between FM, TM and loci marked in both morphotypes, showed a higher value in specifically marked genes compared to commonly marked ones (Fig. 3d). This supports further the finding that specific enrichment in each of the morphotypes are less silenced and potentially more dynamic compared to genes marked by H3K27me3 in both TM and FM cells whereas commonly H3K27me3 marked genes are globally silenced.

To substantiate the assumption that phenotypic plasticity and morphotype switch are regulated by PRC2 in *P. tricornutum*, we took advantage of the lack of stability of the TM phenotype and its tendency to switch to FM. Specifically, we used clonal cell samples with FM and TM morphologies from the same genetic background (TM), which switches habitually to fusiform and therefore contains a mixture of FM and TM cells. We reasoned that the activity of *E(z)* should correlate with H3K27me3 levels in the following way: (1) a pure triradiate population isolated from TM-N (named here TM-T1): highest level of H3K27me3), (2) a population of cells from TM after N generations (N is 60 ± 5) of cell division containing a mixture of triradiate and fusiform morphotypes (TM-N): medium level of H3K27me3) and (3) fusiform cells isolated from the triradiate background TM (TM-Fusi): lowest level of H3K27me3 (Fig. 4a). *E(z)* transcript levels show a clear decrease in TM-N and TM-Fusi compared to TM-T1 (Fig. 4b) which correlates with the switch from TM to FM, reflecting a lower activity of E(z) and H3K27me3 levels. We then asked whether specifically H3K27me3-marked loci in TM lose the mark upon cell switching to FM after multiple generations of sub-culturing leading to TM-N and in transformed fusiform cells (TM-Fusi). As expected, ChIP-qPCR showed clearly a loss of the mark in a population containing a mixture of fusiform and triradiate cells (TM-N) as well as in TM-Fusi compared to TM-T1 (Fig. 4c,d), which contains only triradiate cells, thus correlating the morphology with the level of enrichment in H3K27me3 over specific genes.

In summary, we have demonstrated in this study the role in *P. tricornutum* of *E(z)* as a histone methyltransferase responsible for di and tri-methylation of lysine 27 of histone H3. Knockout

of *E(z)* causes H3K27me3 depletion and loss of triradiate cell shape maintenance, providing evidence for the involvement of *E(z)* and its associated mark in establishing and/or maintaining cell morphology in unicellular species. We showed the dynamic nature of the mark, depending on the specific morphology between and within *P. tricornutum* accessions that correlate with the level of H3K27me3 enrichment. We showed differential marking in two different accessions of *P. tricornutum*, FM versus TM, which identified genes related to cell fate decisions compared to commonly marked genes. This is the first evidence of the involvement of H3K27me3 in cell differentiation in unicellular eukaryotes only distantly related to animals and plants. Our study points to the emerging function of PRC2 and its H3K27me3 associated mark as a determinant of the establishment and maintenance of cell morphology in single celled species such as *P. tricornutum* that shows signs of differentiation of the cell into diverse morphologies. This same function likely diversified with the emergence of multicellularity with PRC2 orchestrating development in plants and animals.

Methods

Strains and growth conditions

Phaeodactylum tricornutum Bohlin Clone Pt1 8.6 (CCMP2561) (referred as FM) and Clone Pt8Tc (referred as TM) cells were grown as described previously [31].

Isolation and immunoprecipitation of chromatin

Chromatin isolation and immunoprecipitation were performed as described previously [32]. The following antibodies were used for immunoprecipitation: H3K27me3 (07-449) from Millipore and H3K27me3 from cell signaling technology. qPCR on recovered DNA was performed as described previously [32]

CRISPR/Cas9 plasmid construction

hCAS9n (Cas9 from *Streptococcus pyogenes*, adapted to human codon usage, fused to SV40 nuclear localization sequence, and containing a D10A mutation) was amplified from pcDNA3.3-TOPO-hCAS9n (kindly received from Dr. Yonatan B. Tzur, Hebrew University of Jerusalem), using the primers 5'-CAC CAT GGA CAA GAA GTA CTC-3' and 5'- TCA CAC CTT CCT CTT CTT CTT-3'. The PCR product was first cloned into pENTR using pENTR/D-TOPO cloning kit (ThermoFisher Scientific), and then sub-cloned into a PT pDest, containing

an N-terminal HA-tag [31], following the manufacturer's protocol, which was named pDest-HA-hCAS9n.

For the sgRNA vector we first cloned the snRNA U6 promoter[33] from *P. tricornutum* genomic DNA using the primers 5'- AAA CGA CGG CCA GTG AAT TCT CGT TTC TGC TGT CAT CAC C-3' and 5'- TCT TTA ATT TCA GAA AAT TCC GAC TTT GAA GGT GTT TTT TG-3'. PU6::unc-119_sgRNA (kindly received from Dr. Yonatan B. Tzur) backbone was amplified using the primers 5'-CAA AAA ACA CCT TCA AAG TCG GAA TTT TCT GAA ATT AAA GA-3' and 5'-GGT GAT GAC AGC AGA AAC GAG AAT TCA CTG GCC GTC GTT T-3'. The two PCR products were used as template for a second round fusion PCR reaction as described in [34]. We further transformed the resulting product into *E. coli*, and extracted the ligated plasmid. The terminator sequence of the P. tricornutum U6 was amplified using the primers 5'-CATTCTAGAAGAACCGCTCACCCATGC-3' and 5'-GTTAAGCTTGAAAAGTTCGTCGAGACCATG-3', digested by XbaI/HindIII and ligated into XbaI/HindIII digested pU6::unc-119. The resulting vector, ptU6::unc-119-sgRNA, was used as template to replace the target sequence to *E(Z)* target by PCR using primers 32817TS12fwd GTG TCG GAG CCC GCC ATA CCG TTT TAG AGC TAG AAA TAG C and 32817TS12rev GGT ATG GCG GGC TCC GAC ACC GAC TTT GAA GGT GTT TTT TG. Target sequences were picked using PhytoCRISP-Ex [35].

Transformation of *P. tricornutum* cells and screening for mutants

Wild type cells of the reference strain FM and the TM were transformed with three plasmids (pPhat1, Cas9 and guide RNA with the target sequence) as described previously [36]. Positive transformants were validated by triple PCR screen for pPhaT1 shble primers (ACT GCG TGCACTTCGTGGC/TCGGTCAGTCCTGCTCCTC), sgRNA (GAGCTGGAAATTGGTTGTC/GACTCGGTGCCACTTTTTCAAGTT) and CAS9n (GGGAGCAGGCAGAAAACATT/TCACACCTTCCTCTTCTTCTT). For each colony, a rapid DNA preparation was performed as described previously and fragment of 400 bp was amplified with primers flanking the target sequence in the *E(z)* gene. The forward primer used is 5'-TAAGATGGAGTATGCCGAAATTC-3' and reverse primer is 5'-AGGCATTTATTCGTGTCTGTTCG-3' PCR product was run in 1% agarose gel and a single band was extracted using Machery Nagle kit and according to the manual manufacturer. PCR product was sequenced using the primer 5'-AGCCACCCTGCGTTAACTGAAAAT-3'.

To make sure that the fusiform cells originating from the switch of TM are not contaminants from the FM, each of the TM-T1, TM-Fusi and TM-N were checked for their genetic background whether it is FM or TM using a molecular marker designed around a 400 bp insertion in the FM background (Supplementary Fig. 4g), identified from genome sequencing of FM and TM strains of *P. tricornutum*[37]. The PCR check confirmed that all the cell samples described above are in the TM genetic background.

Validation of enrichment and expression of target genes

qPCR: Total RNA was extracted from TM and FM cells as described previously [31] and cDNA was synthesized with cDNA high yield synthesis kit according to the manufacturer user manual. Quantitative PCR was performed as described previously [31] using the primer list in Supplementary Table S3

Proteomics and PRM Measurements

Three independent histone purifications recovered from FM wild type cells as well as *E(z)* knockout mutant Del6 in FM genetic background were simultaneously separated by SDS-PAGE and stained with colloidal blue (LabSafe Gel Blue GBiosciences). Three gel slices were excised for each purification and in-gel digested by using trypsin/LysC (Promega). Peptide extracted from each set were pooled and analyzed by nanoLC-MS/MS using an Ultimate 3000 system (Thermo Scientific) coupled to a TripleTOFTM 6600 mass spectrometer (AB Sciex). Peptides were first trapped onto a C18 column (75 µm inner diameter × 2 cm; nanoViper Acclaim PepMapTM 100, Thermo Scientific) with buffer A (2/98 MeCN/H2O in 0.1% formic acid) at a flow rate of 2.5 µL/min over 4 min. Separation was performed on a 50 cm x 75 µm C18 column (nanoViper C18, 3 µm, 100Å, Acclaim PepMapTM RSLC, Thermo Scientific) regulated to 50°C and with a linear gradient from 1% to 30% buffet B (100 MeCN in 0.085% formic acid) at a flow rate of 400 nL/min over 90 min. The mass spectrometer was operated in PRM top30 high sensitivity mode with 100 ms acquisition time for MS1 and MS2 scans respectively with included precursor mass list for 600 sec (see Supplementary Table S4)

PRM Data Analysis

The PRM data were analyzed using Skyline version 3.7.0.11317 MacCoss Lab Software, Seattle, WA; https://skyline.ms/project/home/software/Skyline/begin.view, fragment ions for each targeted mass were extracted and peak areas were integrated. The peptide areas were log2

transformed and the mean log2- area was normalized by the mean area of peptide STDLLIR using software R version 3.1.0. On each peptide a linear model was used to estimate the mean fold change between the conditions, its 97.5% confidence interval and the p-value of the two sided associated t-test. The p-values were adjusted with the Benjamini-Hochberg procedure[38]. The mass spectrometry proteomics data have been deposited to the ProteomeXchange Consortium via the PRIDE [1] partner repository with the dataset identifier PXD012347.

Western blot analysis: Chromatin was extracted from wild type as well as mutants of both TM and FM cells and western blot performed as described previously [32].

Sequencing and computational data analysis

ChIP-Seq: Chromatin Immunoprecipitation (ChIP) was done with monoclonal cell cultures grown using single triradiate cell from Pt8 population (referred as TM). CHIP-Seq was performed as described previously [15,39]. Two replicates were performed and showed a good Pearson correlation (Supplementary Fig. 5). Raw reads were filtered and low quality read-pairs were discarded using FASTQC with a read quality (Phred score) cutoff of 30. Using the genome assembly published in 2008 as reference (Pt1 8.6), we performed reference-assisted mapping of filtered reads using BOWTIE. We then performed the processing and filtering of the alignments using SAMTOOLS and BEDTOOLS. SICER [40] was then used to identify significant enriched H3K27me3 peaks by comparing it with the INPUT. Differential H3K27me3 peak enrichment analysis between FM and TM backgrounds was also done using SICER-df plugin. Peaks with Padj < 0.05 differential enrichment or depletion were considered significant. Functional inferences were obtained by overlapping the differentially enriched peaks over structural annotations from Phatr3 genome annotation[41].

RNA sequencing (RNA-Seq): Total RNA was extracted from FM, TM, and FM Ez-KO (Del6) cell lines. RNA expression and differential gene expression analysis was performed using Eoulsan version 1.2.2 with default parameters [42]. Genes having at least 2 folds expression change with P-value < 0.05 were considered as significant different expressed genes (DEGs).

Whole genome sequencing (WGS): Whole genome sequencing was performed using DNA extracted from monoclonal cell cultures grown using single triradiate cell taken from Pt8 and Pt1 accession, referred to as Pt8tc and Pt1 8.6, respectively. At least 6 µg of genomic DNA from each accession was used to construct a sequencing library following the manufacturer's

instructions (Illumina Inc.). Paired-end sequencing libraries with a read size of 100 bp and an insert size of approximately 400 bp were sequenced on an Illumina HiSeq 2000 sequencer at Berry Genomics Company (China) and Fasteris for Pt1 8.6 and Pt8tc, respectively. Low quality read-pairs were discarded using FASTQC with a read quality (Phred score) cutoff of 30. Using the genome assembly published previously[43], we performed reference-assisted assembly of all the accessions. We used BOWTIE (-n 2 –X 400) for mapping the high quality NGS reads to the reference genome followed by the processing and filtering of the alignments using SAMTOOLS and BEDTOOLS. For estimating the genetic diversity between Pt1 8.6 and Pt8Tc genome, GATK [44] configured for diploid genomes, was used for variant calling, which included single nucleotide polymorphisms (SNPs), small insertions and deletions ranging between 1 and 300 base pairs (bp). The genotyping mode was kept default (genotyping mode = DISCOVERY), Emission confidence threshold (-stand_emit_conf) was kept 10 and calling confidence threshold (-stand_call_conf) was kept at 30. The minimum number of reads per base to be called as a high quality SNP was kept at 4 (i.e., read-depth >=4x). SNPEFF was used to annotate the functional nature of the polymorphisms.

GO enrichment analysis: GO categories were grouped by 3 different levels of expression according to a simple density clustering algorithm (also confirmed by iterative k-means clustering).

Data Availability:

All data are available through NCBI Sequence Read Archive with accession number PRJNA565539.

References

1 Aldiri, I. & Vetter, M. L. PRC2 during vertebrate organogenesis: a complex in transition. *Dev Biol* **367**, 91-99, doi:10.1016/j.ydbio.2012.04.030 (2012).

2 Fragola, G. *et al.* Cell reprogramming requires silencing of a core subset of polycomb targets. *PLoS Genet* **9**, e1003292, doi:10.1371/journal.pgen.1003292 (2013).

3 Surface, L. E., Thornton, S. R. & Boyer, L. A. Polycomb group proteins set the stage for early lineage commitment. *Cell Stem Cell* **7**, 288-298, doi:10.1016/j.stem.2010.08.004 (2010).

4 Schwartz, Y. B. & Pirrotta, V. A new world of Polycombs: unexpected partnerships and emerging functions. *Nature reviews. Genetics* **14**, 853-864, doi:10.1038/nrg3603 (2013).

5 Martinez-Balbas, M. A., Tsukiyama, T., Gdula, D. & Wu, C. Drosophila NURF-55, a WD repeat protein involved in histone metabolism. *Proceedings of the National Academy of Sciences of the United States of America* **95**, 132-137 (1998).

366 6 Kohler, C. & Villar, C. B. Programming of gene expression by Polycomb group proteins. *Trends*
367 *Cell Biol* **18**, 236-243, doi:10.1016/j.tcb.2008.02.005 (2008).
368 7 Shaver, S., Casas-Mollano, J. A., Cerny, R. L. & Cerutti, H. Origin of the polycomb repressive
369 complex 2 and gene silencing by an E(z) homolog in the unicellular alga Chlamydomonas.
370 *Epigenetics* **5**, 301-312 (2010).
371 8 Mikulski, P., Komarynets, O., Fachinelli, F., Weber, A. P. M. & Schubert, D. Characterization of
372 the Polycomb-Group Mark H3K27me3 in Unicellular Algae. *Front Plant Sci* **8**, 607,
373 doi:10.3389/fpls.2017.00607 (2017).
374 9 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the
375 marine diatom Phaeodactylum tricornutum. *Genome biology* **16**, 102, doi:10.1186/s13059-
376 015-0671-8 (2015).
377 10 De Martino, A. *et al.* Physiological and Molecular Evidence that Environmental Changes Elicit
378 Morphological Interconversion in the Model Diatom Phaeodactylum tricornutum. *Protist*
379 **162**, 462-481, doi:S1434-4610(11)00006-X [pii]

380 10.1016/j.protis.2011.02.002 (2011).
381 11 De Martino, A. M., A. Juan Shi, K.P. Bowler, C. Genetic and phenotypic characterization of
382 Phaeodactylum tricornutum (Bacillariophyceae) accessions. *J. Phycol.* **43**, 992–1009 (2007).
383 12 He, L., Han, X. & Yu, Z. A rare Phaeodactylum tricornutum cruciform morphotype: culture
384 conditions, transformation and unique fatty acid characteristics. *PloS one* **9**, e93922,
385 doi:10.1371/journal.pone.0093922 (2014).
386 13 Borowitzka, M. A., Volcani, B.E. The polymorphic diatom Phaeodactylum tricornutum:
387 Ultrastructure of its morphotypes. *J. Phycolo.* **14**, 10-21 (1978).
388 14 Margueron, R. & Reinberg, D. The Polycomb complex PRC2 and its mark in life. *Nature* **469**,
389 343-349, doi:10.1038/nature09784 (2011).
390 15 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the
391 marine diatom Phaeodactylum tricornutum. *Genome biology* **16**, 102, doi:10.1186/s13059-
392 015-0671-8 (2015).
393 16 Jamieson, K., Rountree, M. R., Lewis, Z. A., Stajich, J. E. & Selker, E. U. Regional control of
394 histone H3 lysine 27 methylation in Neurospora. *Proceedings of the National Academy of*
395 *Sciences of the United States of America* **110**, 6027-6032, doi:10.1073/pnas.1303750110
396 (2013).
397 17 Ferrari, K. J. *et al.* Polycomb-dependent H3K27me1 and H3K27me2 regulate active
398 transcription and enhancer fidelity. *Molecular cell* **53**, 49-62,
399 doi:10.1016/j.molcel.2013.10.030 (2014).
400 18 Lafos, M. *et al.* Dynamic regulation of H3K27 trimethylation during Arabidopsis
401 differentiation. *PLoS Genet* **7**, e1002040, doi:10.1371/journal.pgen.1002040 (2011).
402 19 Jacob, Y. *et al.* ATXR5 and ATXR6 are H3K27 monomethyltransferases required for chromatin
403 structure and gene silencing. *Nat Struct Mol Biol* **16**, 763-768, doi:10.1038/nsmb.1611
404 (2009).
405 20 Schmitges, F. W. *et al.* Histone methylation by PRC2 is inhibited by active chromatin marks.
406 *Molecular cell* **42**, 330-341, doi:10.1016/j.molcel.2011.03.025 (2011).
407 21 Veluchamy, A. *et al.* An integrative analysis of post-translational histone modifications in the
408 marine diatom Phaeodactylum tricornutum. *Genome Biol* **16**, 102 (2015).
409 22 Bernardes, J. S., Vieira, F. R., Zaverucha, G. & Carbone, A. A multi-objective optimization
410 approach accurately resolves protein domain architectures. *Bioinformatics* **32**, 345-353,
411 doi:10.1093/bioinformatics/btv582 (2016).
412 23 Bernardes, J., Zaverucha, G., Vaquero, C. & Carbone, A. Improvement in Protein Domain
413 Identification Is Reached by Breaking Consensus, with the Agreement of Many Profiles and
414 Domain Co-occurrence. *PLoS Comput Biol* **12**, e1005038, doi:10.1371/journal.pcbi.1005038
415 (2016).

416 24 Le Costaouëc, T., Unamunzaga, C., Mantecon, L. and Helbert, W. New structural insights into
417 the cell-wall polysaccharide of the diatom Phaeodactylum tricornutum. *Algal research* **26**
418 172-179 (2017).
419 25 Marcotte, E. M., Pellegrini, M., Yeates, T. O. & Eisenberg, D. A census of protein repeats. *J*
420 *Mol Biol* **293**, 151-160, doi:10.1006/jmbi.1999.3136 (1999).
421 26 Yu, J. S. & Cui, W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR
422 signalling in pluripotency and cell fate determination. *Development* **143**, 3050-3060,
423 doi:10.1242/dev.137075 (2016).
424 27 Etienne-Manneville, S. & Hall, A. Rho GTPases in cell biology. *Nature* **420**, 629-635,
425 doi:10.1038/nature01148 (2002).
426 28 Bryan, B. A., Li, D., Wu, X. & Liu, M. The Rho family of small GTPases: crucial regulators of
427 skeletal myogenesis. *Cell Mol Life Sci* **62**, 1547-1555, doi:10.1007/s00018-005-5029-z (2005).
428 29 Thompson, S. J., Loftus, L. T., Ashley, M. D. & Meller, R. Ubiquitin-proteasome system as a
429 modulator of cell fate. *Curr Opin Pharmacol* **8**, 90-95, doi:10.1016/j.coph.2007.09.010 (2008).
430 30 Maheswari, U. *et al.* Digital expression profiling of novel diatom transcripts provides insight
431 into their biological functions. *Genome biology* **11**, R85, doi:gb-2010-11-8-r85 [pii]

432 10.1186/gb-2010-11-8-r85 (2010).
433 31 Siaut, M. *et al.* Molecular toolbox for studying diatom biology in Phaeodactylum tricornutum.
434 *Gene* **406**, 23-35, doi:S0378-1119(07)00275-2 [pii]

435 10.1016/j.gene.2007.05.022 (2007).
436 32 Lin, X., Tirichine, L. & Bowler, C. Protocol: Chromatin immunoprecipitation (ChIP)
437 methodology to investigate histone modifications in two model diatom species. *Plant*
438 *methods* **8**, 48, doi:10.1186/1746-4811-8-48 (2012).
439 33 Rogato, A. *et al.* The diversity of small non-coding RNAs in the diatom Phaeodactylum
440 tricornutum. *BMC genomics* **15**, 698, doi:10.1186/1471-2164-15-698 (2014).
441 34 Hobert, O. PCR fusion-based approach to create reporter gene constructs for expression
442 analysis in transgenic C. elegans. *Biotechniques* **32**, 728-730 (2002).
443 35 Rastogi, A., Murik, O., Bowler, C. & Tirichine, L. PhytoCRISP-Ex: a web-based and stand-alone
444 application to find specific target sequences for CRISPR/CAS editing. *BMC bioinformatics* **17**,
445 261, doi:10.1186/s12859-016-1143-1 (2016).
446 36 Falciatore, A., Casotti, R., Leblanc, C., Abrescia, C. & Bowler, C. Transformation of
447 nonselectable reporter genes in marine diatoms. *Mar Biotechnol (NY)* **1**, 239-251, doi:MBT30
448 [pii] (1999).
449 37 Rastogi, A. V. F., Deton-Cabanillas AF, Veluchamy A., Cantrel, C., Wang, G., Vanormelingen,
450 P., Bowler, C., Piganeau, G., Hu, H. and Leila Tirichine. A genomics approach reveals the
451 global genetic polymorphism, structure and functional diversity of ten accessions of the
452 marine model diatom Phaeodactylum tricornutum

453

454 *ISME J* (2019).
455 38 Benjamini, Y., and Yekutieli, D. . The control of the false discovery rate in multiple testing
456 under dependency. *Annals of Statistics* **29**, 1165–1188 (2001).
457 39 Veluchamy, A. *et al.* Insights into the role of DNA methylation in diatoms by genome-wide
458 profiling in Phaeodactylum tricornutum. *Nat Commun* **4**, doi:10.1038/ncomms3091 (2013).
459 40 Zang, C. *et al.* A clustering approach for identification of enriched domains from histone
460 modification ChIP-Seq data. *Bioinformatics* **25**, 1952-1958,
461 doi:10.1093/bioinformatics/btp340 (2009).
462 41 Rastogi, A. *et al.* Integrative analysis of large scale transcriptome data draws a
463 comprehensive landscape of Phaeodactylum tricornutum genome and evolutionary origin of
464 diatoms. *Sci Rep* **8**, 4834, doi:10.1038/s41598-018-23106-x (2018).

42 Jourdren, L., Bernard, M., Dillies, M. A. & Le Crom, S. Eoulsan: a cloud computing-based framework facilitating high throughput sequencing analyses. *Bioinformatics* **28**, 1542-1543, doi:10.1093/bioinformatics/bts165 (2012).

43 Bowler, C. *et al.* The Phaeodactylum genome reveals the evolutionary history of diatom genomes. *Nature* **456**, 239-244, doi:nature07410 [pii]

10.1038/nature07410 (2008).

44 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**, 1297-1303, doi:10.1101/gr.107524.110 (2010).

Acknowledgements

Hanhua Hu from the Chinese Academy of Science is acknowledged for the gift of CM morphotype.). LT acknowledges funds from the CNRS and the region of Pays de la Loire (ConnecTalent EPIALG project). CB acknowledges funding from the ERC Advanced Award ‘Diatomite. XZ was supported by a PhD fellowship from the Chinese Scholarship Council (CSC-201604910722). AR was supported by an International PhD fellowship from the MEMO LIFE Program.

Competing interests

The authors declare no competing interests.

Figure legends

Figure 1. *Phaeodactylum tricornutum* morphotypes and enhancer of zeste knockout mutants. (a) *P. tricornutum* morphotypes (top left fusiform, scale bar = 3µm, top right oval, bottom left triradiate, scale bar = 2µm, bottom right cruciform). Scale bars correspond to 1 µm in OM and CM. **(b)** Sequence chromatograms of PCR product from WT cells and CRISPR cas9 mutants of enhancer of zeste showing the different indels in FM, TM and CM. (**c)** Light microscopy images of WT, E(z) KO and Cas9 control cells. Empty vector controls containing Cas9 and Shble antibiotic resistance gene show no loss of H3K27me3 and retain the wild type morphology suggesting that morphology distortion is not due to the transformation but to the absence of H3K27me3. (**d**) Western blots of WT and two E(z)KOs from each morphotype using a monoclonal antibody against H3K27me3. Histone H4 was used as a loading control. ChIP-QPCR enrichment levels of H3K27me3 on genes in WT and E(z) KO in FM (**e**) and TM (**f**) backgrounds.

Figure 2. Genomic features of H3K27me3 targets in FM and TM cells. (**a)** Venn diagrams showing the number of common and specific genomic features [Genes, Transposable elements (TEs), and Intergenic Regions (IGRs)] targeted by H3K27me3 in TM (orange circles) and FM (blue circles). ChIP-QPCR validation of H3K27me3 specifically marked genes in **(b)** TM and **(c)** FM morphotypes. (**d**) Distribution of the most frequent GO terms on genes marked with H3K27me3. The distribution was sub-divided into different categories, where TM and FM represent the GOs observed exclusively on triradiate and fusiform genes, respectively. A third category (Both) also presents a GO distribution for genes observed on both morphotypes.

Figure 3. Gene expression profiles of 3HK27me3 specific and common targets in FM and TM. (**a**) The box plot represents mean enrichment of H3K27me3 (Y-axis), with log10 scaling, over genes marked specifically in FM (blue), TM (orange), and also on genes marked in both (commonly marked) morphotypes. The enrichment profile is generated using number of genes marked by H3K27me3 specifically in each morphotype and also in both. The significant H3K27me3 enrichment difference between specifically marked and commonly marked is estimated using two-tailed t-test with P value < 0.0001, denoted by "****". (**b**) Expression of

genes marked specifically in FM (blue), TM (orange), and in both phenotypic backgrounds with same principle aesthetics and categorical genes, used in 2f. The significance/non-significance of the variability of expression between specifically and commonly marked genes is estimated using two-tailed t-test with P value = 0.0433, as denoted by "*". "ns" denote non-significant. (**c**) Relative expression level of H3K27me3 targets genes in the TM and enhancer of zeste knock out M2-11. In the plot, fold change log2(PtM2-11/Pt8Tc) values are shown (**d**) Boxplot showing the entropy value distribution of H3K27me3 marked genes. Entropy values measures the differential expression of genes under different experimental conditions. Entropy values are derived from expression data (fragments per kilobase of exon per million fragments mapped) under five different experimental conditions. Genes marked specifically by H3K27me3 in TM (Triradiate) shows higher variation in expression followed by FM (Fusiform) specific H3K27me3 marked genes. Horizontal lines represent the median entropy values. The significant of FM/TM-specific H3K27me3 marked genes and genes marked on both conditions are estimated using two-tailed t-test with P value < 0.0001.

Figure 4. H3K27me3 enrichment levels and morphotype switch. (a) Schematic diagram showing how TM-N, TM-T1 and TM-Fusi were generated. After generations of culture in lab growth condition (ASW media, 19°C, 12h/12h light dark period), some triradiate cells switch to fusiform forming a mixture of FM and TM cells, named TM-N. A single fusiform cell from TM-N was picked and propagated clonally giving rise to TM-Fusi. Similarly, single triradiate cell was isolated from TM-N and its clonal propagation gave a population of pure triradiate cells named TM-T1. (**b)** Relative expression level of Enhancer of Zest in FM, TM-N, TM-T1 and TM-Fusi respectively, two pairs of primer were designed at N-terminal of Enhancer of Zeste gene, and at the CXC domain respectively. (**c)** ChIP-QPCR enrichment levels of H3K27me3 in TM-T1 and TM-Fusi. (**d)** ChIP-QPCR showing enrichment levels of H3K27me3 in TM-T1 and TM-N.

Supplementary Table 1. Cell counts of different morphologies and cell size measurements in each wild type morphotype and knock out of E(z)

Supplementary Table 2. List of genes marked by H3K27me3 with their annotation and GOs.

Supplementary Table 3. List of the primers used in this study

Supplementary Table 4. Mass spectrometry quantification of Di- and Tri-methylation of H3K27me3

Supplementary Figure 1. Phylogeny **(a-c)**. **(d)** Western blots using a monoclonal antibody against H3K27me3 on protein or chromatin extracts of different species representative of the super SAR lineage (1: *Bigelowiella_natans,* 2: *Gymnophora dimorpha,* 3: *Skeletonema marinoi,* 4: *Thalassiosira pseudonana,* 5: *Raphoneis sp,* 6: *Synedra sp,* 7: *Asterionellopsisglacialis,* 8: *Thalassiosira rotula,* 9: *Phaeodactylum tricornutum,* 10: *Isocrhrysis lutea,* 11: *Amhedinium klebselii,* 12: *Amhedinium carteri).* **(e)** Western blot on chromatin extracts from OM, FM, TM and CM with a monoclonal antibody against H3K27me3 showing a significant difference in enrichment levels of H3K27me3 in TM and CM which are higher compared to FM and OM.

Supplementary Figure 2. **(a)** Western blot of chromatin extracts of Cas9 control lines from each of FM, TM and CM. H4 histone antibody was used as a loading control. **(b, c)** Mass spectrometry quantification of di and tri-methylation of lysine 27 of histone H3 in both wild type and enhancer of zeste knockout mutant. MS/MS spectrum of the $[M + 2H]^{2+}$ precursor ion of histone H3 (27–36 residue peptide) tri-methylated or di-methylated on K27. Broken bonds above and below sequence denote b and y ions, respectively, that were annotated from the spectrum. **(d)** Abundance of H3 K27 di- and tri-methylated KSAPATGGVK peptide. Y axis shows normalized log2 (WT/Mutant) of the di-methylated and tri-methylated peptides. All measurements have been performed in triplicate, and error bars indicated the 97.5% confidence interval (see supplementary Table 1). **(e)** Western blot of chromatin extracts from wild type FM and TM as well as E(z) knockouts in both backgrounds with a monoclonal antibody against H3K27me2. H4 histone antibody was used as loading control. **(f)** Growth curves of wild type, enhancer of zeste mutants and cas9 control line in each of FM, TM and CM. **(g)** Gel picture of a molecular marker distinguishing FM and TM and amplifying an insertion in one allele present in FM but absent in TM.

Supplementary Figure 3.

(a) Total genome coverage of H3K27me3 within TM (orange) and FM (blue) showing a higher mapping of H3K27me3 in TM compared to FM. **(b)** mean distribution of H3K27me3 over 500 bp upstream, gene body, and 500 bp downstream region of all the gene targets in TM (orange line) and FM (blue line). **(c)** Unmarked genes (J01910, J31617), commonly marked genes (J34600, J44413) were chosen as internal controls. FM (EG00164) and TM specifically marked (J49062) genes were used as controls for the reproducibility of independent ChIP-QPCR results. (**d)** violin plot represents the mean fold change of gene expression in Ez(KO) lines in TM compared to the wild-type (WT). The significant expression difference between Ez(KO) and WT is estimated using two-tailed t-test with P value < 0.0001

Supplementary Figure 4. Scatter plots with Pearson correlation coefficient displaying the relationship between TM and E(z) knock out RNA Seq replicates.

Supplementary Figure 5. Scatter plots with Pearson correlation coefficient displaying the relationship between TM ChIP-Seq replicates

Supplementary Table 1. Cell counts of different morphologies and cell size measurements in each wild type morphotype and knock out of E(z)

Supplementary Table 2. List of genes marked by H3K27me3 with their annotation and GOs.

Supplementary Table 3. List of the primers used in this study

Supplementary Table 4. Mass spectrometry quantification of Di- and Tri-methylation of H3K27me3

Supplementary Figure 1. Phylogeny **(a-c)**. **(d)** Western blots using a monoclonal antibody against H3K27me3 on protein or chromatin extracts of different species representative of the super SAR lineage (1: *Bigelowiella_natans,* 2: *Gymnophora dimorpha,* 3: *Skeletonema marinoi,* 4: *Thalassiosira pseudonana,* 5: *Raphoneis sp,* 6: *Synedra sp,* 7: *Asterionellopsisglacialis,* 8: *Thalassiosira rotula,* 9: *Phaeodactylum tricornutum,* 10: *Isocrhrysis lutea,* 11: *Amhedinium klebselii,* 12: *Amhedinium carteri).* **(e)** Western blot on chromatin extracts from OM, FM, TM and CM with a monoclonal antibody against H3K27me3

showing a significant difference in enrichment levels of H3K27me3 in TM and CM which are higher compared to FM and OM.

Supplementary Figure 2. **(a)** Western blot of chromatin extracts of Cas9 control lines from each of FM, TM and CM. H4 histone antibody was used as a loading control. **(b, c)** Mass spectrometry quantification of di and tri-methylation of lysine 27 of histone H3 in both wild type and enhancer of zeste knockout mutant. MS/MS spectrum of the $[M + 2H]^{2+}$ precursor ion of histone H3 (27–36 residue peptide) tri-methylated or di-methylated on K27. Broken bonds above and below sequence denote b and y ions, respectively, that were annotated from the spectrum. **(d)** Abundance of H3 K27 di- and tri-methylated KSAPATGGVK peptide. Y axis shows normalized log2 (WT/Mutant) of the di-methylated and tri-methylated peptides. All measurements have been performed in triplicate, and error bars indicated the 97.5% confidence interval (see supplementary Table 1). **(e)** Western blot of chromatin extracts from wild type FM and TM as well as E(z) knockouts in both backgrounds with a monoclonal antibody against H3K27me2. H4 histone antibody was used as loading control. **(f)** Growth curves of wild type, enhancer of zeste mutants and cas9 control line in each of FM, TM and CM. **(g)** Gel picture of a molecular marker distinguishing FM and TM and amplifying an insertion in one allele present in FM but absent in TM.

Supplementary Figure 3.

(a) Total genome coverage of H3K27me3 within TM (orange) and FM (blue) showing a higher mapping of H3K27me3 in TM compared to FM. **(b)** mean distribution of H3K27me3 over 500 bp upstream, gene body, and 500 bp downstream region of all the gene targets in TM (orange line) and FM (blue line). **(c)** Unmarked genes (J01910, J31617), commonly marked genes (J34600, J44413) were chosen as internal controls. FM (EG00164) and TM specifically marked (J49062) genes were used as controls for the reproducibility of independent ChIP-QPCR results. (**d)** violin plot represents the mean fold change of gene expression in Ez(KO) lines in TM compared to the wild-type (WT). The significant expression difference between Ez(KO) and WT is estimated using two-tailed t-test with P value < 0.0001

Supplementary Figure 4. Scatter plots with Pearson correlation coefficient displaying the relationship between TM and E(z) knock out RNA Seq replicates.

Supplementary Figure 5. Scatter plots with Pearson correlation coefficient displaying the relationship between TM ChIP-Seq replicates

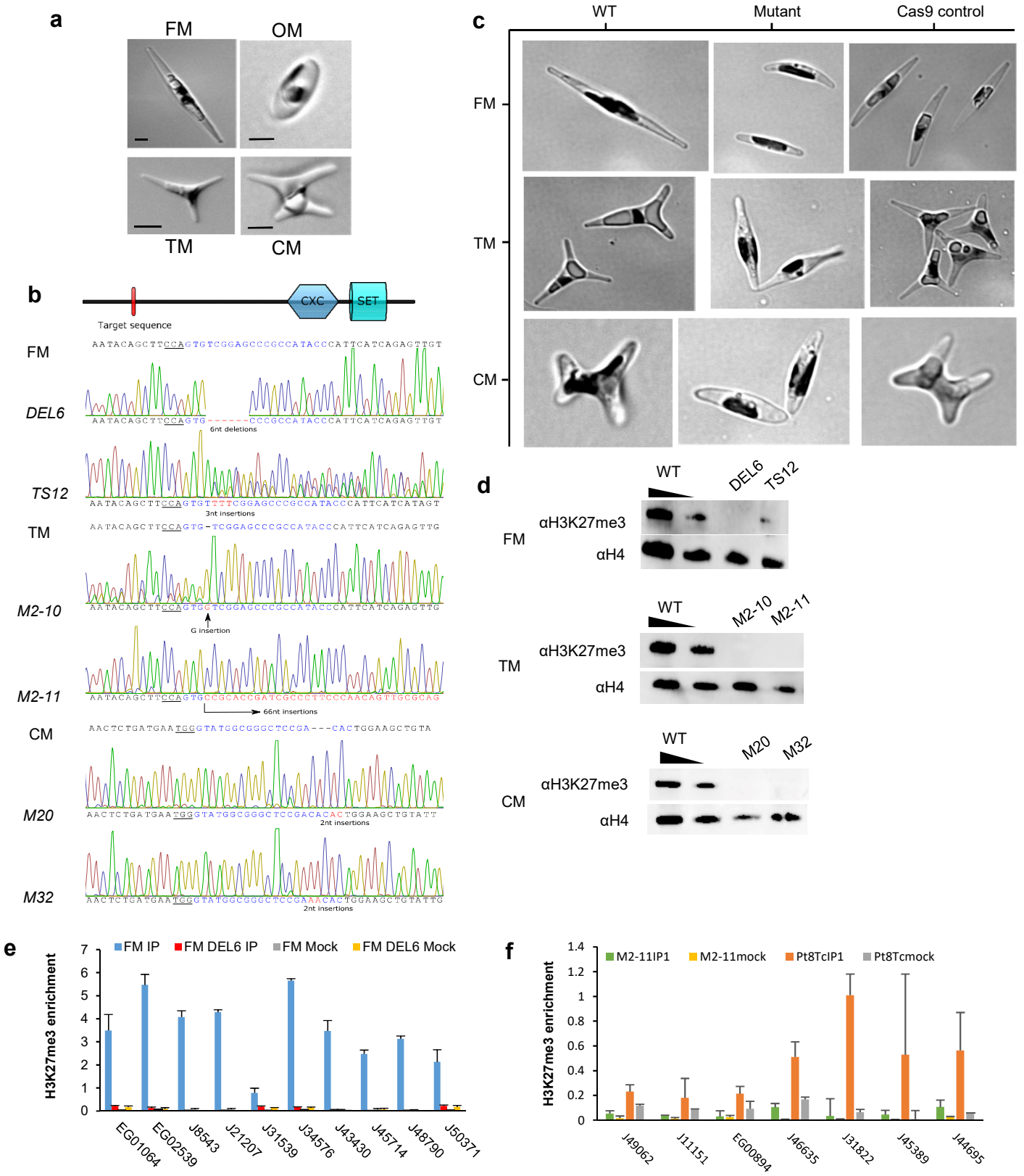

a
FM
OM
TM
CM
b
CXC
SET
Target sequence
FM
DEL6
6nt deletions
TS12
3nt insertions
TM
M2-10
G insertion
M2-11
66nt insertions
CM
M20
2nt insertions
M32
2nt insertions
c
WT
Mutant
Cas9 control
FM
TM
CM
d
WT
DEL6
TS12
FM
αH3K27me3
αH4
WT
M2-10
M2-11
TM
αH3K27me3
αH4
WT
M20
M32
CM
αH3K27me3
αH4
e
FM IP
FM DEL6 IP
FM Mock
FM DEL6 Mock
H3K27me3 enrichment
EG01064
EG02539
J8543
J21207
J31539
J34576
J43430
J45714
J48790
J50371
f
M2-11IP1
M2-11mock
Pt8TcIP1
Pt8Tcmock
H3K27me3 enrichment
J49062
J11151
EG00894
J46635
J31822
J45389
J44695

Figure 1

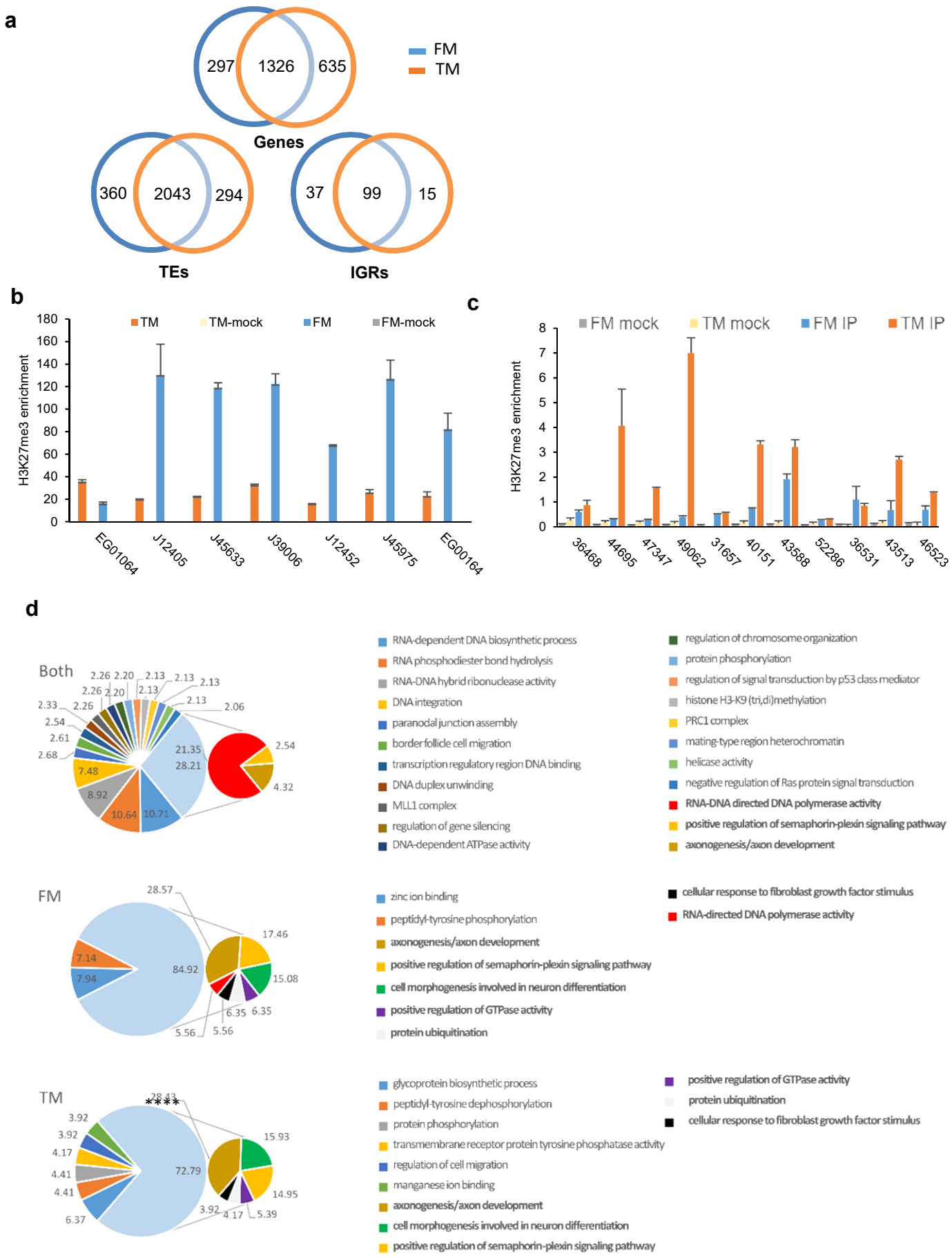

a
297
1326
635
FM
TM
Genes
360
2043
294
TEs
37
99
15
IGRs
b
TM
TM-mock
FM
FM-mock
H3K27me3 enrichment
EG01064
J12405
J45633
J39006
J12452
J45975
EG00164
c
FM mock
TM mock
FM IP
TM IP
H3K27me3 enrichment
36468
44695
47347
49062
31657
40151
43588
52286
36531
43513
46523
d
Both
RNA-dependent DNA biosynthetic process
RNA phosphodiester bond hydrolysis
RNA-DNA hybrid ribonuclease activity
DNA integration
paranodal junction assembly
border follicle cell migration
transcription regulatory region DNA binding
DNA duplex unwinding
MLL1 complex
regulation of gene silencing
DNA-dependent ATPase activity
regulation of chromosome organization
protein phosphorylation
regulation of signal transduction by p53 class mediator
histone H3-K9 (tri,di)methylation
PRC1 complex
mating-type region heterochromatin
helicase activity
negative regulation of Ras protein signal transduction
RNA-DNA directed DNA polymerase activity
positive regulation of semaphorin-plexin signaling pathway
axonogenesis/axon development
FM
zinc ion binding
peptidyl-tyrosine phosphorylation
axonogenesis/axon development
positive regulation of semaphorin-plexin signaling pathway
cell morphogenesis involved in neuron differentiation
positive regulation of GTPase activity
protein ubiquitination
cellular response to fibroblast growth factor stimulus
RNA-directed DNA polymerase activity
TM
glycoprotein biosynthetic process
peptidyl-tyrosine dephosphorylation
protein phosphorylation
transmembrane receptor protein tyrosine phosphatase activity
regulation of cell migration
manganese ion binding
axonogenesis/axon development
cell morphogenesis involved in neuron differentiation
positive regulation of semaphorin-plexin signaling pathway
positive regulation of GTPase activity
protein ubiquitination
cellular response to fibroblast growth factor stimulus

Figure 2

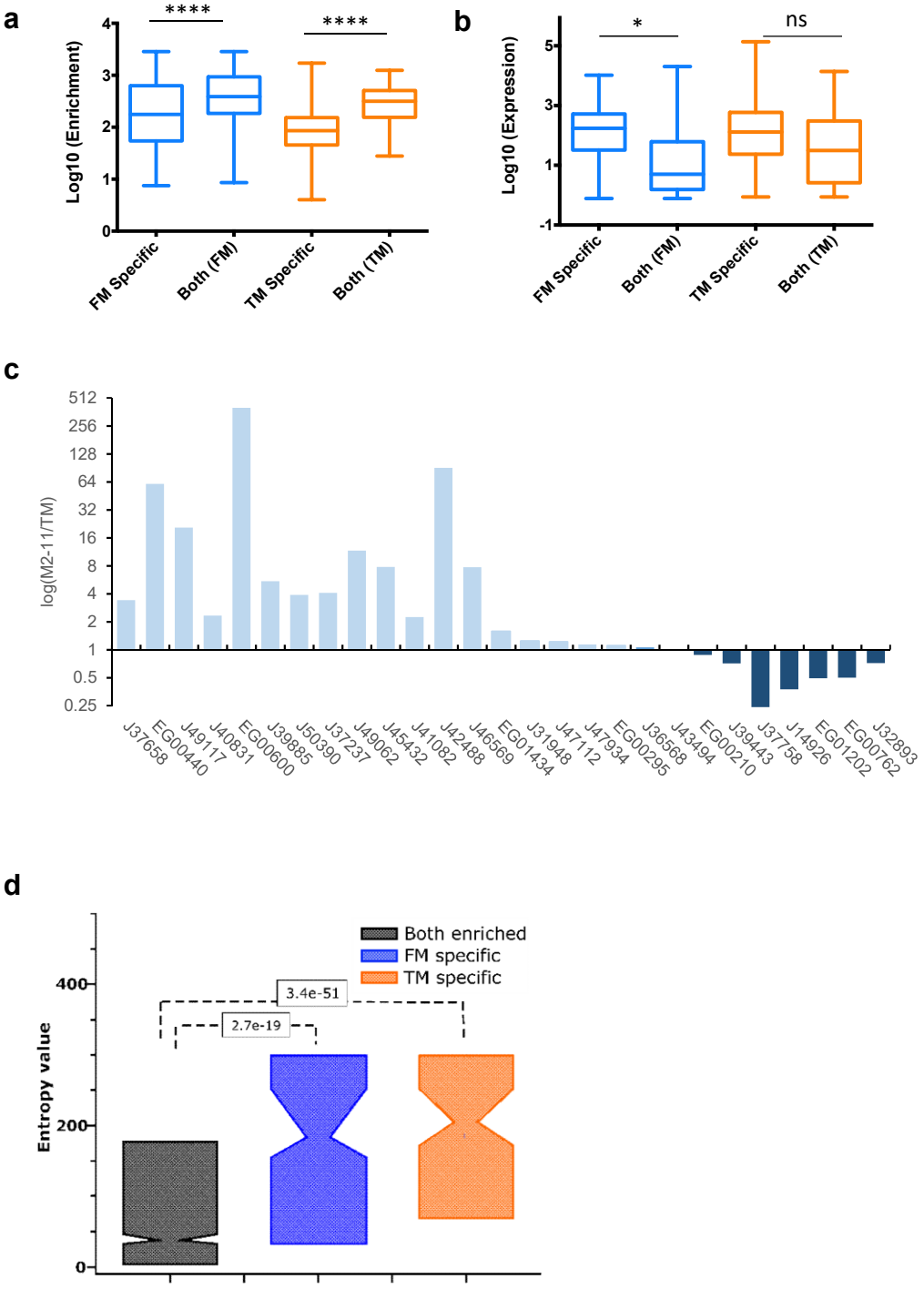

a
Log10 (Enrichment)
****
****
FM Specific
Both (FM)
TM Specific
Both (TM)
b
Log10 (Expression)
*
ns
FM Specific
Both (FM)
TM Specific
Both (TM)
c
log(M2-11/TM)
J37658
EG00440
J49117
J40831
EG00600
J39885
J50390
J37237
J49062
J45432
J41082
J42488
J46569
EG01434
J31948
J47112
J47934
EG00295
J36568
J43494
EG00210
J39443
J37758
J14926
EG01202
EG00762
J32893
d
Entropy value
Both enriched
FM specific
TM specific
3.4e-51
2.7e-19

Figure 3

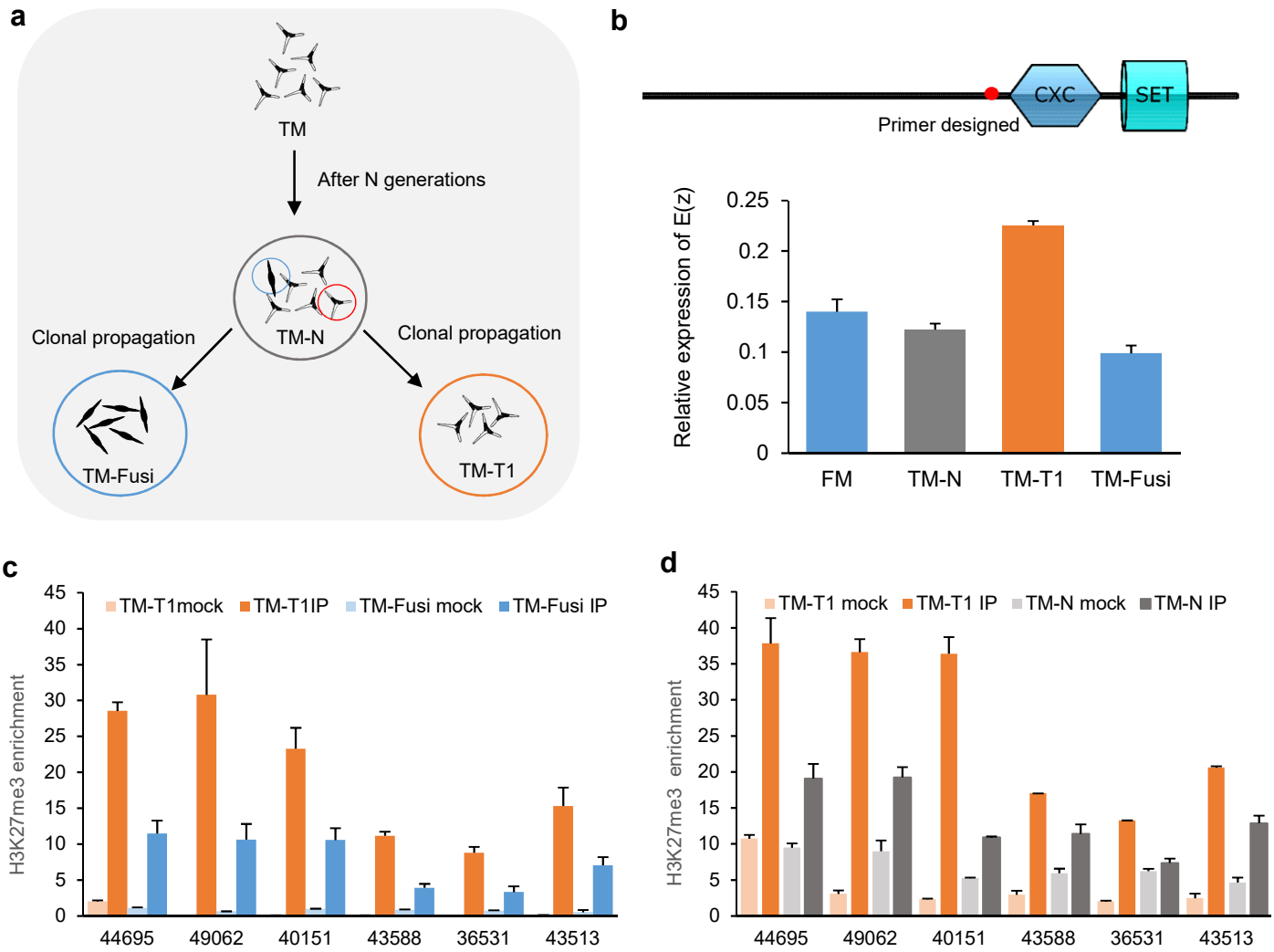

a
TM
After N generations
TM-N
Clonal propagation
Clonal propagation
TM-Fusi
TM-T1
b
CXC
SET
Primer designed
Relative expression of E(z)
0.25
0.2
0.15
0.1
0.05
0
FM
TM-N
TM-T1
TM-Fusi
c
TM-T1mock
TM-T1IP
TM-Fusi mock
TM-Fusi IP
H3K27me3 enrichment
45
40
35
30
25
20
15
10
5
0
44695
49062
40151
43588
36531
43513
d
TM-T1 mock
TM-T1 IP
TM-N mock
TM-N IP
H3K27me3 enrichment
45
40
35
30
25
20
15
10
5
0
44695
49062
40151
43588
36531
43513

Figure 4

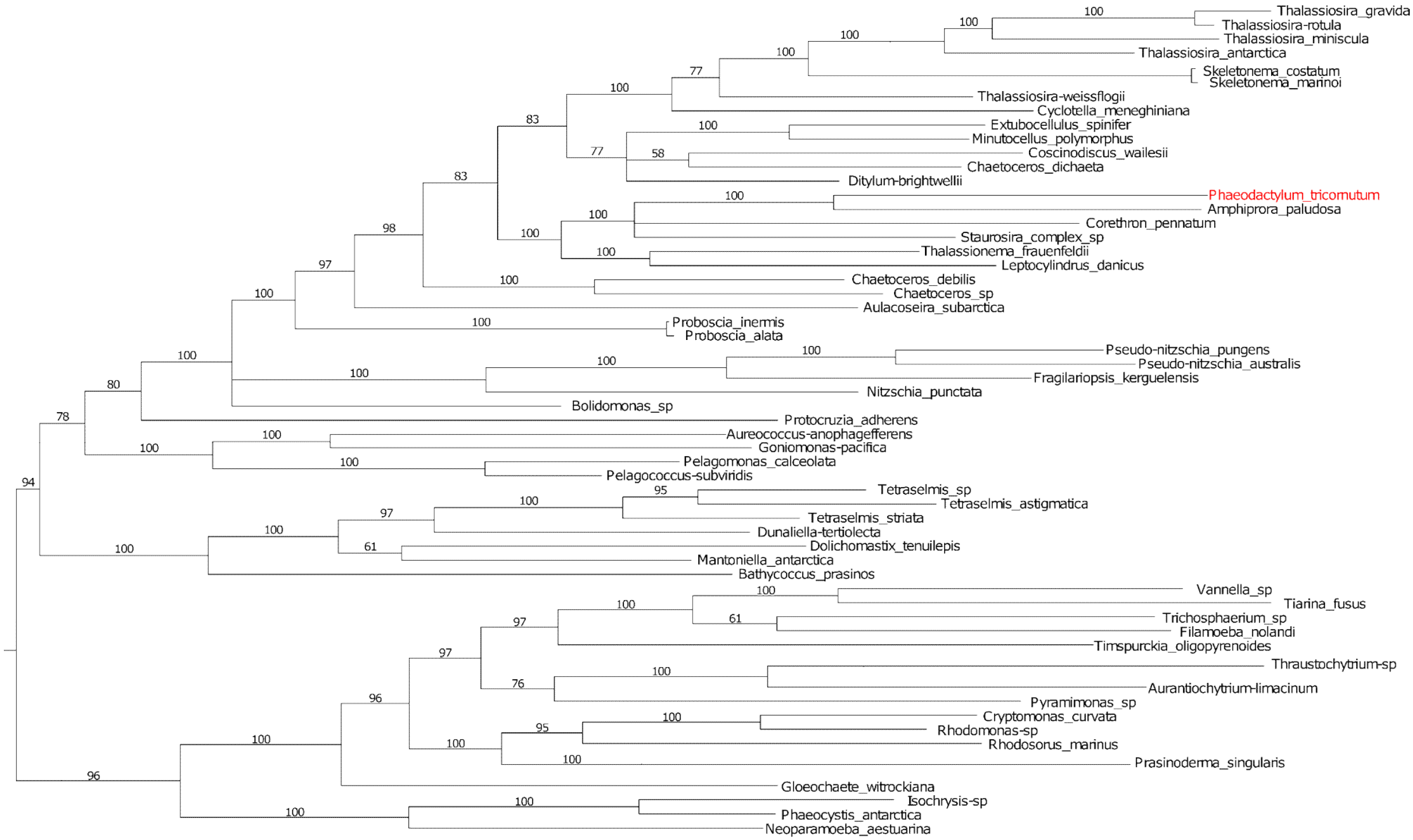

a
Thalassiosira_gravida
Thalassiosira-rotula
Thalassiosira_miniscula
Thalassiosira_antarctica
Skeletonema_costatum
Skeletonema_marinoi
Thalassiosira-weissflogii
Cyclotella_meneghiniana
Extubocellulus_spinifer
Minutocellus_polymorphus
Coscinodiscus_wailesii
Chaetoceros_dichaeta
Ditylum-brightwellii
Phaeodactylum_tricornutum
Amphiprora_paludosa
Corethron_pennatum
Staurosira_complex_sp
Thalassionema_frauenfeldii
Leptocylindrus_danicus
Chaetoceros_debilis
Chaetoceros_sp
Aulacoseira_subarctica
Proboscia_inermis
Proboscia_alata
Pseudo-nitzschia_pungens
Pseudo-nitzschia_australis
Fragilariopsis_kerguelensis
Nitzschia_punctata
Bolidomonas_sp
Protocruzia_adherens
Aureococcus-anophagefferens
Goniomonas-pacifica
Pelagomonas_calceolata
Pelagococcus-subviridis
Tetraselmis_sp
Tetraselmis_astigmatica
Tetraselmis_striata
Dunaliella-tertiolecta
Dolichomastix_tenuilepis
Mantoniella_antarctica
Bathycoccus_prasinos
Vannella_sp
Tiarina_fusus
Trichosphaerium_sp
Filamoeba_nolandi
Timspurckia_oligopyrenoides
Thraustochytrium-sp
Aurantiochytrium-limacinum
Pyramimonas_sp
Cryptomonas_curvata
Rhodomonas-sp
Rhodosorus_marinus
Prasinoderma_singularis
Gloeochaete_witrockiana
Isochrysis-sp
Phaeocystis_antarctica
Neoparamoeba_aestuarina

Figure S1

b

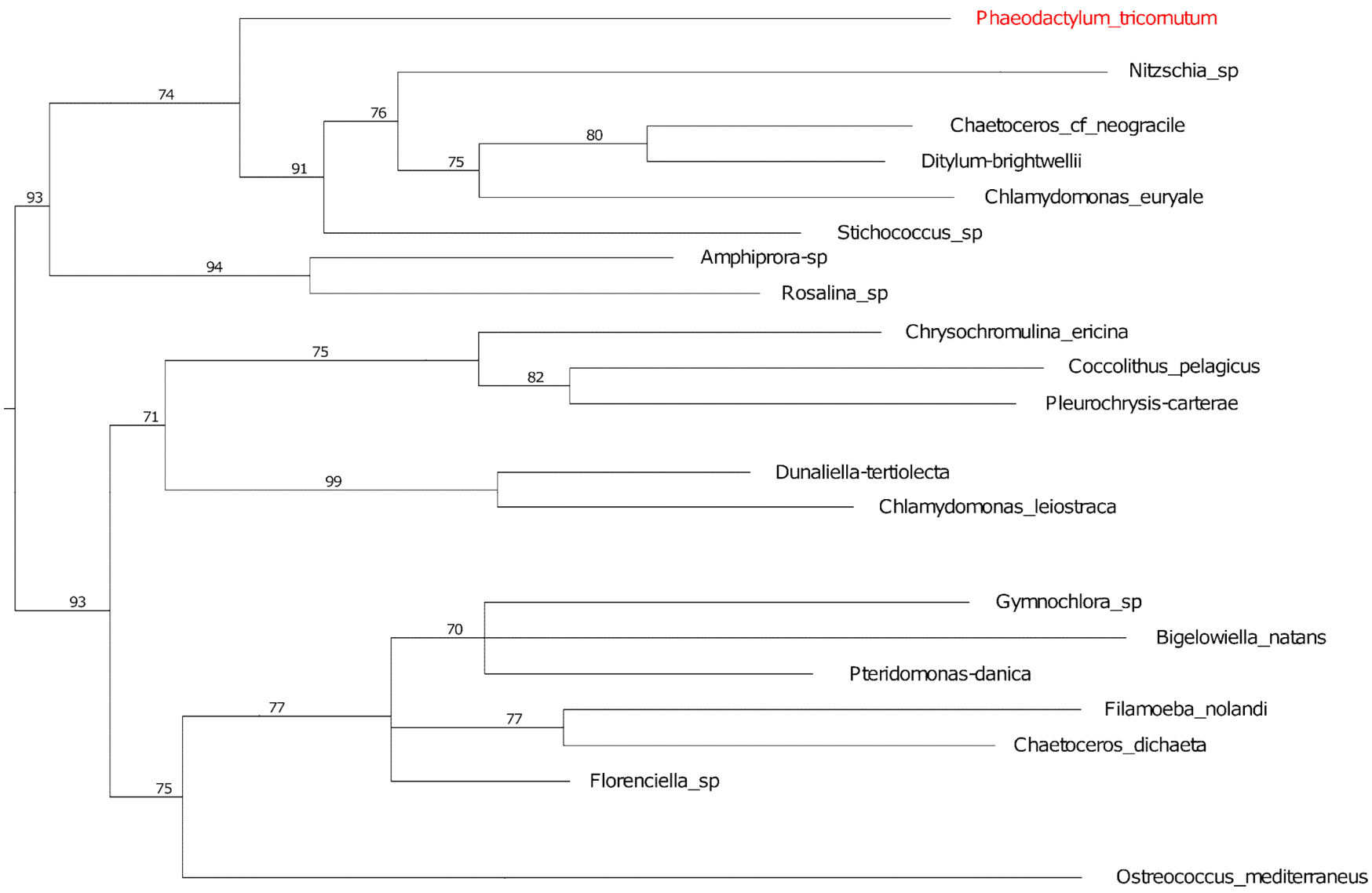

Phaeodactylum_tricornutum
Nitzschia_sp
Chaetoceros_cf_neogracile
Ditylum-brightwellii
Chlamydomonas_euryale
Stichococcus_sp
Amphiprora-sp
Rosalina_sp
Chrysochromulina_ericina
Coccolithus_pelagicus
Pleurochrysis-carterae
Dunaliella-tertiolecta
Chlamydomonas_leiostraca
Gymnochlora_sp
Bigelowiella_natans
Pteridomonas-danica
Filamoeba_nolandi
Chaetoceros_dichaeta
Florenciella_sp
Ostreococcus_mediterraneus
74
76
80
91
75
93
94
75
82
71
99
93
70
77
77
75

Figure S1

c

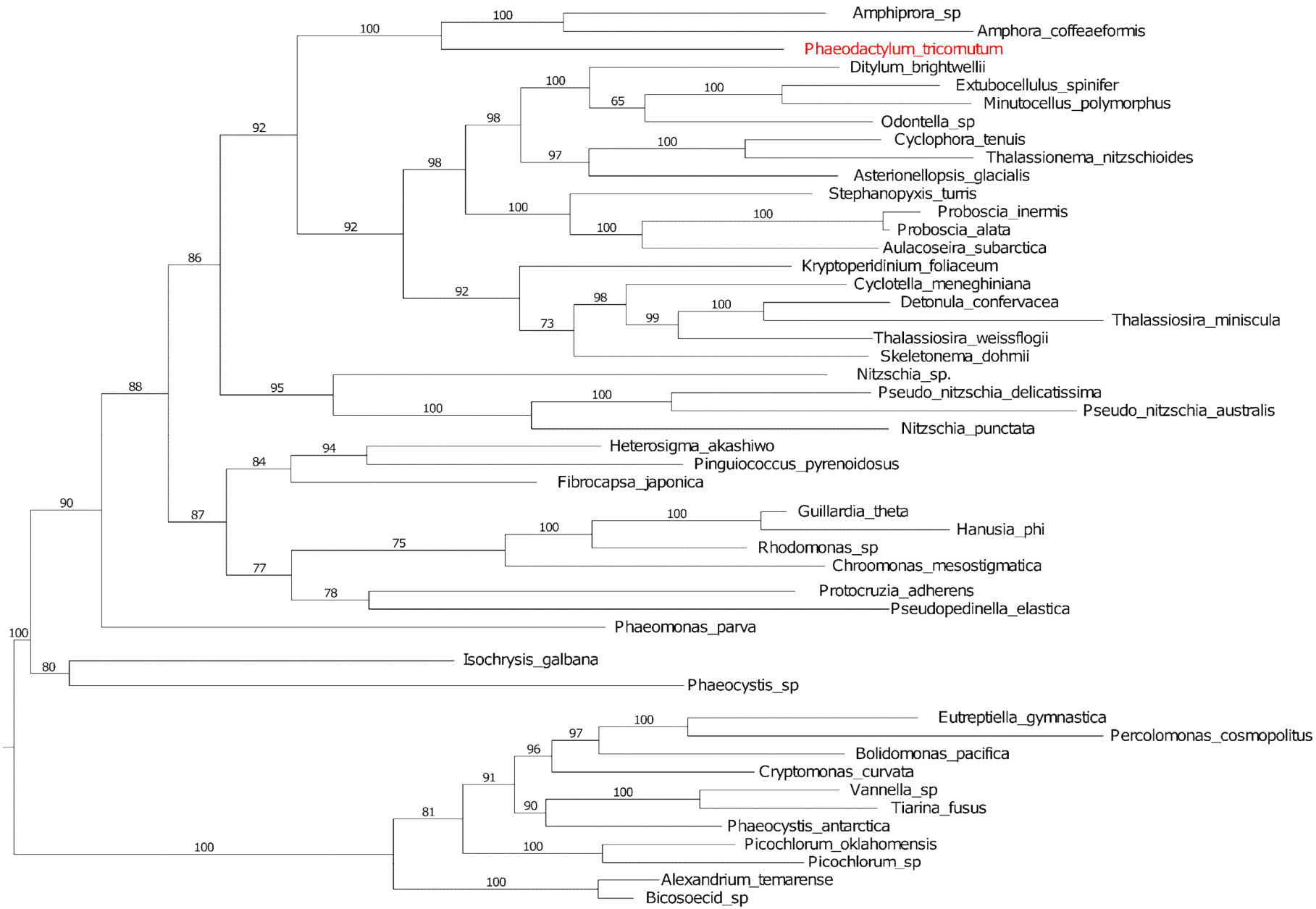
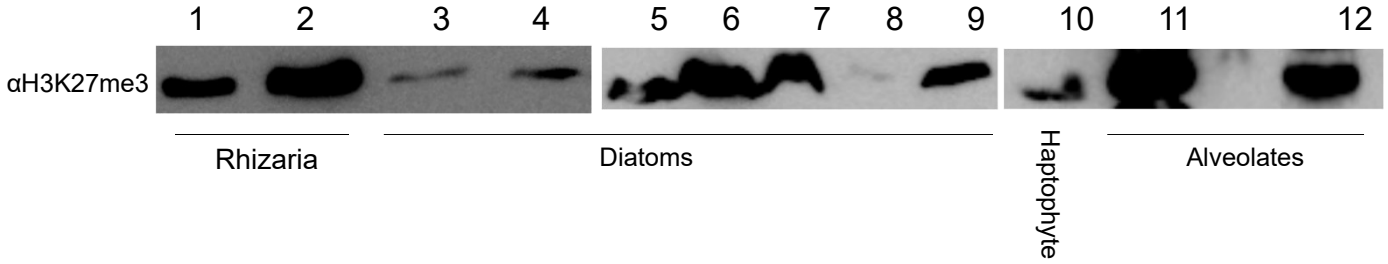

Amphiprora_sp
Amphora_coffeaeformis
Phaeodactylum_tricornutum
Ditylum_brightwellii
Extubocellulus_spinifer
Minutocellus_polymorphus
Odontella_sp
Cyclophora_tenuis
Thalassionema_nitzschioides
Asterionellopsis_glacialis
Stephanopyxis_turris
Proboscia_inermis
Proboscia_alata
Aulacoseira_subarctica
Kryptoperidinium_foliaceum
Cyclotella_meneghiniana
Detonula_confervacea
Thalassiosira_miniscula
Thalassiosira_weissflogii
Skeletonema_dohrnii
Nitzschia_sp.
Pseudo_nitzschia_delicatissima
Pseudo_nitzschia_australis
Nitzschia_punctata
Heterosigma_akashiwo
Pinguiococcus_pyrenoidosus
Fibrocapsa_japonica
Guillardia_theta
Hanusia_phi
Rhodomonas_sp
Chroomonas_mesostigmatica
Protocruzia_adherens
Pseudopedinella_elastica
Phaeomonas_parva
Isochrysis_galbana
Phaeocystis_sp
Eutreptiella_gymnastica
Percolomonas_cosmopolitus
Bolidomonas_pacifica
Cryptomonas_curvata
Vannella_sp
Tiarina_fusus
Phaeocystis_antarctica
Picochlorum_oklahomensis
Picochlorum_sp
Alexandrium_temarense
Bicosoecid_sp

Figure S1

d


1
2
3
4
5
6
7
8
9
10
11
12
αH3K27me3
Rhizaria
Diatoms
Haptophyte
Alveolates

e

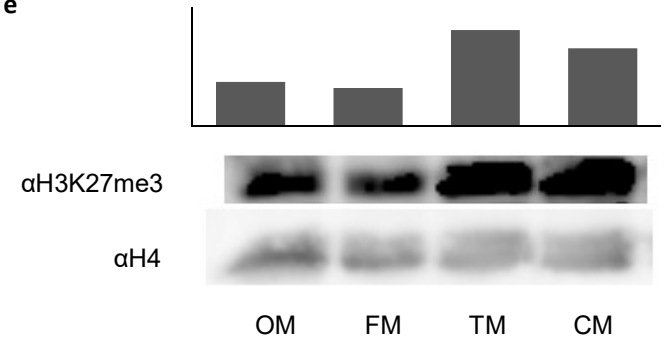

αH3K27me3
αH4
OM
FM
TM
CM

Figure S1

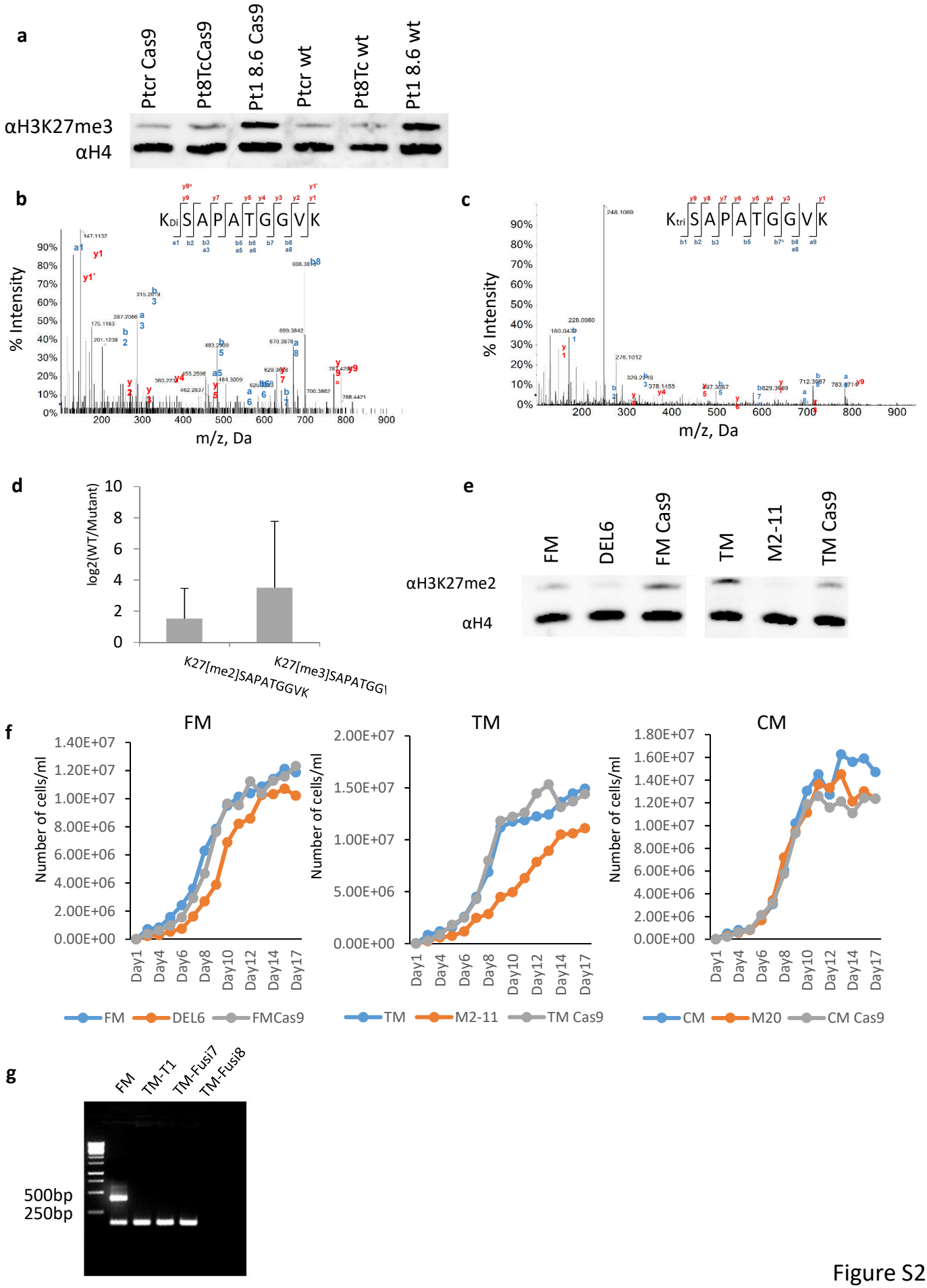

a
Ptcr Cas9
Pt8TcCas9
Pt1 8.6 Cas9
Ptcr wt
Pt8Tc wt
Pt1 8.6 wt
αH3K27me3
αH4
b
% Intensity
m/z, Da
c
% Intensity
m/z, Da
d
log2(WT/Mutant)
K27[me2]SAPATGGVK
K27[me3]SAPATGGV
e
FM
DEL6
FM Cas9
TM
M2-11
TM Cas9
αH3K27me2
αH4
f
FM
TM
CM
Number of cells/ml
Day1
Day4
Day6
Day8
Day10
Day12
Day14
Day17
FM
DEL6
FMCas9
TM
M2-11
TM Cas9
CM
M20
CM Cas9
g
FM
TM-T1
TM-Fusi7
TM-Fusi8
500bp
250bp

Figure S2

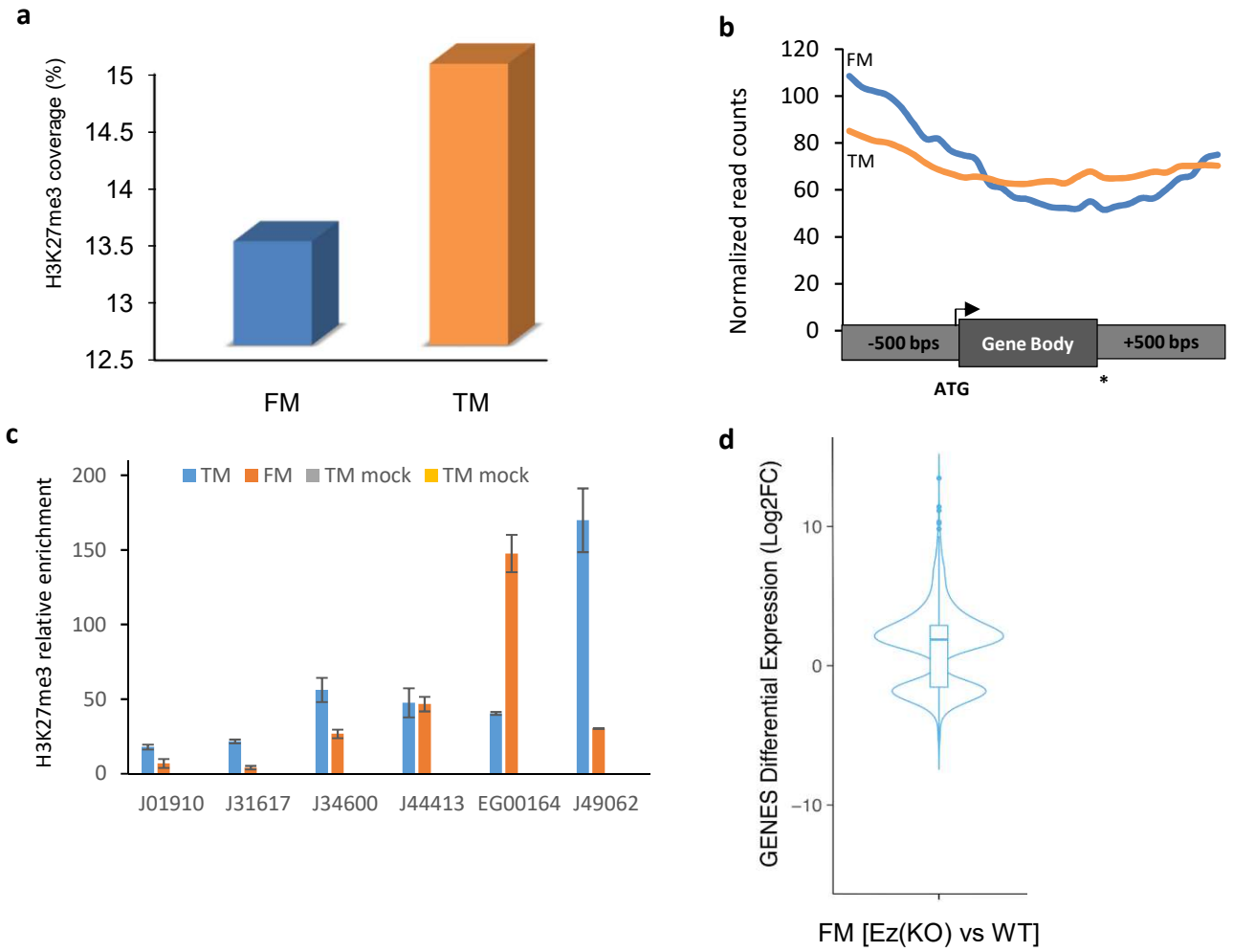
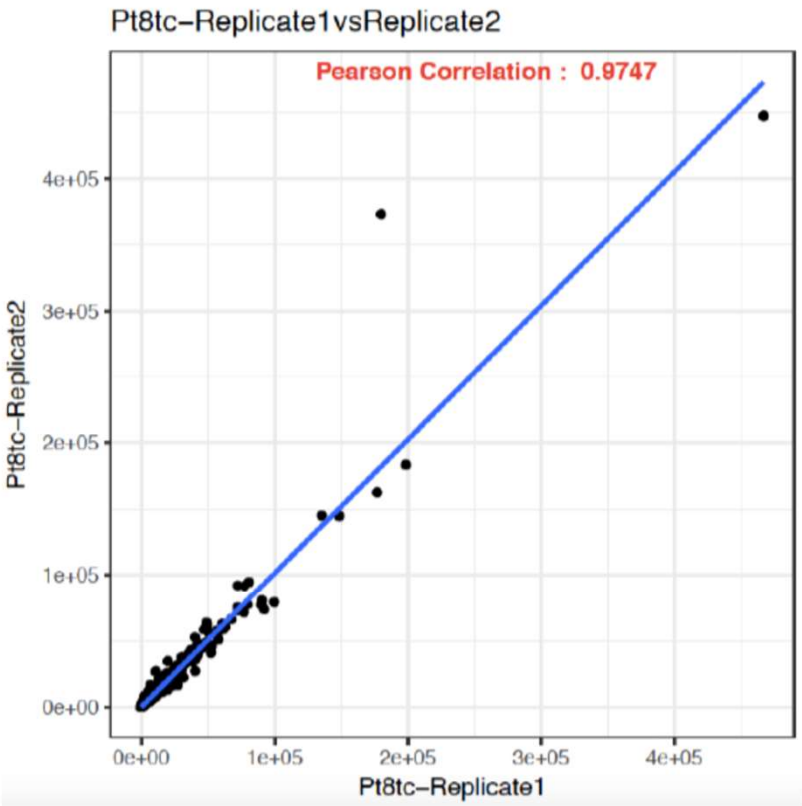

a
H3K27me3 coverage (%)
15
14.5
14
13.5
13
12.5
FM
TM
b
Normalized read counts
120
100
80
60
40
20
0
FM
TM
-500 bps
Gene Body
+500 bps
ATG
*
c
H3K27me3 relative enrichment
200
150
100
50
0
TM
FM
TM mock
TM mock
J01910
J31617
J34600
J44413
EG00164
J49062
d
GENES Differential Expression (Log2FC)
10
0
−10
FM [Ez(KO) vs WT]

Figure S3


Pt8tc−Replicate1vsReplicate2
Pearson Correlation : 0.9747
Pt8tc−Replicate2
Pt8tc−Replicate1
0e+00
1e+05
2e+05
3e+05
4e+05

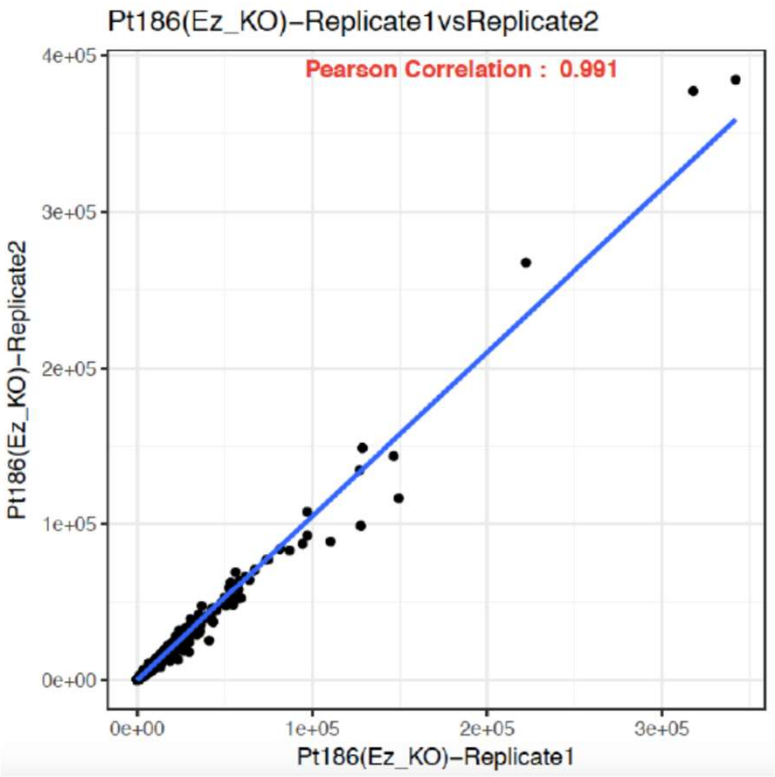

Pt186(Ez_KO)−Replicate1vsReplicate2
Pearson Correlation : 0.991
Pt186(Ez_KO)−Replicate2
Pt186(Ez_KO)−Replicate1
0e+00
1e+05
2e+05
3e+05
4e+05

Figure S4

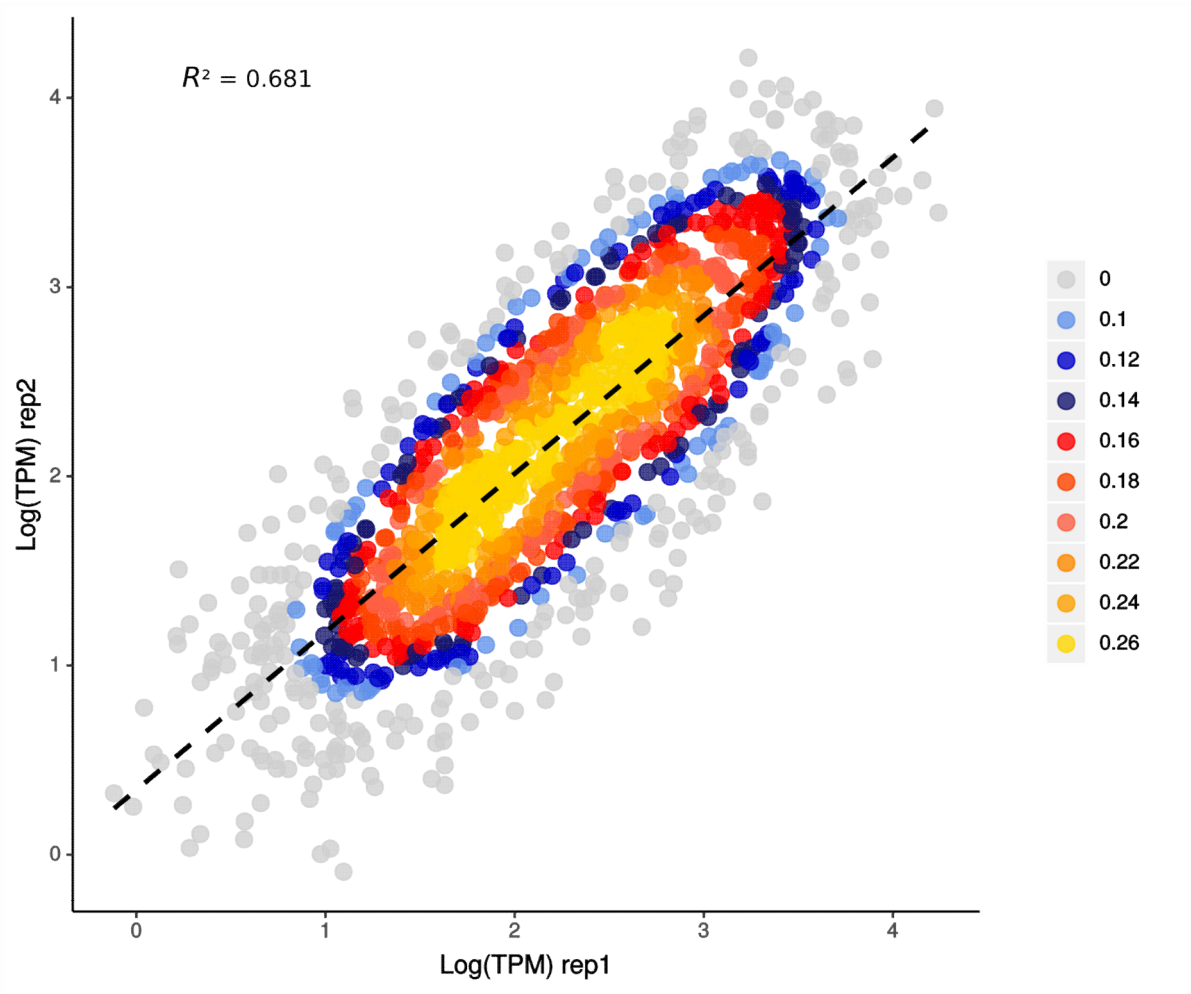

R^2 = 0.681
Log(TPM) rep2
Log(TPM) rep1
0
1
2
3
4
0
0.1
0.12
0.14
0.16
0.18
0.2
0.22
0.24
0.26

Figure S5

Table 1. Cell counts of different morphologies and cell size measurements in each wild type morphotype and knock out of *E(z)*

	FM	TM	CM	OM	Majority	Percentage	Size
FM	107	0	0	0	FM	100%	26.58μm
TM	0	100	0	0	TM	100%	L:13.59μm; W:11.48μm
CM	4	31	66	0	CM	67.34%	-
OM	0	0	0	100	OM	100%	-
TM-N	63	55	0	0	FM	53.38%	-
T1	0	219	0	0	TM	100%	-
DEL6	74	0	0	0	FM	100%	19.91μm
TS12	41	0	0	0	FM	100%	-
M2-10	65	0	0	0	FM	100%	-
M2-11	74	0	0	0	FM	100%	21.04μm
M20	32	0	0	0	FM	100%	23.59μm

Table 2. Mass spectrometry quantification of Di- and Tri-methylation of H3K27me3

peptide sequence	Protein Modification (without Nterm Met)	peptide used for normalization	log2(WT/Mutant)	CI 2.5%	CI 97.5%	adjusted p value
KSAPATGGVK	Dimethyl (K:27)	STDLLIR	1.52038639	1.09023167	1.95054112	2.10E-07
KSAPATGGVK	Trimethyl (K:27)	STDLLIR	3.5091078	2.7625983	4.25561731	7.82E-10