Title:

Chromatin profiling of the repetitive and non-repetitive genome of the human fungal pathogen Candida albicans

Authors:

Robert Jordan Price[1], Esther Weindling[2], Judith Berman[2] and Alessia Buscaino[1*]

Running Title: *Candida albicans* chromatin profiling

[1] University of Kent, School of Biosciences, Kent Fungal Group, Canterbury Kent, CT2 7NJ. UK

[2] Department of Microbiology and Biotechnology, George S. Wise Faculty of Life Sciences, Tel Aviv University, Ramat Aviv, 69978, Israel

* To whom correspondence should be addressed. Tel: (+44) (0)1227824854; Email: A.Buscaino@kent.ac.uk

ABSTRACT

Background

Eukaryotic genomes are packaged into chromatin structures with pivotal roles in regulating all DNA-associated processes. Post-translational modifications of histone proteins modulate chromatin structure leading to rapid, reversible regulation of gene expression and genome stability which are key steps in environmental adaptation. *Candida albicans* is the leading fungal pathogen in humans, and can rapidly adapt and thrive in diverse host niches. The contribution of chromatin to *C. albicans* biology is largely unexplored.

Results

Here, we harnessed genome-wide sequencing approaches to generate the first comprehensive chromatin profiling of histone modifications (H3K4me^3, H3K9Ac, H4K16Ac and γ-H2A) across the *C. albicans* genome and relate it to gene expression. We demonstrate that gene-rich non-repetitive regions are packaged in canonical euchromatin associated with histone modifications that mirror their transcriptional activity. In contrast, repetitive regions are assembled into distinct chromatin states: subtelomeric regions and the rDNA locus are assembled into canonical heterochromatin, while Major Repeat Sequences and transposons are packaged in chromatin bearing features of euchromatin and heterochromatin. Genome-wide mapping of γH2A, a marker of genome instability, allowed the identification of potential recombination-prone genomic sites. Finally, we present the first quantitative chromatin profiling in *C. albicans* to delineate the role of the chromatin modifiers Sir2 and Set1 in controlling chromatin structure and gene expression.

Conclusions

This study presents the first genome-wide chromatin profiling of histone modifications associated with the *C. albicans* genome. These epigenomic maps provide an invaluable resource to understand the contribution of chromatin to *C. albicans* biology.

KEYWORDS: *Candida albicans*, Chromatin, Histone modifications, Human fungal pathogen, Heterochromatin, Euchromatin, Epigenetics, Genome Instability, Sir2, Set1

BACKGROUND

Packaging of genomes into chromatin is the key determinant of nuclear organization [1,2]. The basic unit of chromatin is the nucleosome, consisting of a histone octamer of two molecules each of histone H2A, H2B, H3, and H4, around which 147 bp of DNA are wrapped in almost two turns [3]. Histone proteins are subjected to a wide variety of post-translational modifications, known as histone marks, that decorate distinct chromatin regions [3]. Modification of chromatin structure controls a plethora of nuclear processes including gene expression, DNA repair and DNA replication [3,4]. Consequently, genome-wide maps of histone modifications have been instrumental in identifying functionally different regions of eukaryotic genomes [5,6]. Gene-rich, non-repetitive DNA is associated with active histone marks, forming euchromatin, a chromatin state permissive to transcription and recombination [7]. At euchromatic regions, promoters of active genes are enriched in histone H3 trimethylated on lysine 4 (H3K4me^3) and acetylated on lysine 9 (H3K9Ac), while gene bodies are enriched in a different set of histone modifications, such as acetylation of lysine 16 on histone H4 (H4K16Ac) [8–10]. In contrast, genomic regions enriched in repetitive DNA and low in gene density are assembled into

heterochromatin [7].These repetitive sequences (including tandem repeats, transposable elements and gene families) are a threat to genome stability. At repetitive elements, heterochromatin assembly promotes genome stability by repressing deleterious recombination events [7,11,12]. Heterochromatin is devoid of active histone marks (i.e. H3K4me^3, H3K9Ac and H4K16Ac) and is enriched in repressive histone marks such as methylation of lysine 9 on histone H3 (H3K9me) and methylation of lysine 27 on histone H3 (H3K27me) [1].

While euchromatin structure is largely conserved across organisms, histone marks associated with heterochromatic regions vary between organisms. For example, in the model system *Saccharomyces cerevisiae*, heterochromatin is devoid of H3K9me and H3K27me marks but nucleosomes are hypomethylated on H3K4 and hypoacetylated on H3K9 and H4K16 [1,13]. Phosphorylation of serine 129 on histone H2A (known as γ-H2A) is enriched at heterochromatin regions in *S. cerevisiae*, *Schizosaccharomyces pombe* and *Neurospora* crassa, independently of the cell cycle stage [14–17]. Given that γ-H2A is a hallmark of DNA double strand breaks, these findings suggest that heterochromatic regions are flagged for DNA damage. In contrast, in human cells, phosphorylation of H2AX, a modification functionally analogous to γ-H2A, does not decorate heterochromatic regions [18,19].

Chromatin modifications also play major roles in controlling genome stability by dictating pathways of DNA repair. Indeed, choices of DNA repair pathways (i.e. Non Homologous End Joining or Homologous Recombination) depend on the chromatin state of the genomic region undergoing repair and extensive chromatin changes, including γ-H2A, are linked to repair of DNA breaks [20]. Consequently, in unchallenged cells, γ-H2A mapping is used to identify unstable genomic regions (named γ-sites) that are prone to intrinsic DNA damage and recombination [14].

Chromatin modifications are reversible and specific histone modifiers maintain or erase the histone modification state associated with different chromatin regions. Among these, histone acetyltransferases (HATs) and histone deacetylases (HDACs) respectively maintain and erase histone acetylation, while histone methyltransferases (HMTs) and demethylases (HDMs) are responsible for the methylation state of histones [2,3]. Chromatin regulation rapidly and reversibly alters gene expression and genome stability and can, therefore, have a major impact on environmental adaptation of microbial organisms that need to rapidly adapt to sudden environmental changes [21,22].

One such organism is the human fungal pathogen *Candida albicans*. *C. albicans* is a commensal organism that colonises the mouth, the skin, and the uro-intestinal and reproductive tracts of most individuals without causing any harm. However, *C. albicans* is also the most common causative agent of invasive fungal infections and systemic infections are associated with high mortality rates (up to 50%) [23]. *C. albicans* is such a successful pathogen because it rapidly adapts and thrives in diverse host niches. The ability to switch among multiple specialised cell types, as well as its remarkable genome plasticity, is at the basis of *C. albicans* adaptation [24].

C. albicans is a diploid organism with a genome organised in 2 x 8 chromosomes containing 6408 protein-coding genes, in addition to a large number of non-coding RNAs [25–28]. The genome of *C. albicans* contains several classes of repetitive elements: telomeres/subtelomeres, the rDNA locus, Major Repeat Sequences (MRS) and transposable elements [29]. Telomeres are composed of tandemly repeating 23 bp units, while subtelomeres are enriched in long terminal repeats (LTR), retrotransposons and gene families [29,30].

The rDNA locus consists of a tandem array of a ~12 kb unit repeated 50 to 200 times. Each unit contains the two highly conserved 35 S and 5 S rRNA genes that are separated by two Non-Transcribed Spacer regions (NTS1 and NTS2), whose sequences are not conserved with other eukaryotes [29,31].

MRS loci are long tracts (10–100 kb) of nested DNA repeats found on 7 of the 8 *C. albicans* chromosomes [29,32]. These repetitive domains, found in *C. albicans* and in the closely related species *C. dubliensis* and *C. tropicalis*, are formed by large tandem arrays of 2.1 kb RPS unit flanked by non-repetitive HOK and RBP-2 elements. Each RBP-2 element contains a protein-coding gene, *FGR6,* important for morphological switches [32,33].

Several classes of retrotransposons are present in the *C. albicans* genome including 16 classes of LTR retro-transposons (Tca1-16) and Zorro non-LTR retrotransposons that are present in 5-10 copies per cell, dispersed along the chromosomes. Among those, Tca2, Tca4, Tca5, Zorro-2 and Zorro-3 are capable of transposition [34–36].

The *C. albicans* genome is remarkably plastic, and natural isolates exhibit a broad spectrum of genomic variations including Loss of Heterozygosity (LOH) events, chromosome rearrangements and aneuploidy [37]. Evolution experiments and analyses of clinical isolates have demonstrated that repetitive elements are hypermutable sites of the *C. albicans* genome and are prone to high rates of recombination [37,38].

Several studies have demonstrated that regulation of chromatin structure plays critical roles in regulating *C. albicans* gene expression and genome instability [39–42]. However, comprehensive profiling of histone modifications across the whole *C. albicans* genome is still lacking. Generation of these epigenomic maps will be

essential to truly understand the impact of chromatin regulation to *C. albicans* adaptation and development of virulence traits.

In this study, we used chromatin immunoprecipitation with massively parallel sequencing (ChIP-seq) technology to establish the first comprehensive genome-wide map of *C. albicans* histone modifications (H3K4me^3, H3K9Ac, H4K16Ac and γH2A), marking euchromatic, heterochromatic regions and potential recombination-prone unstable sites. Genome-wide mapping of RNA Polymerase II (RNAPII) and transcriptome expression profiling allowed us to unveil the link between histone modification states and transcriptional activity. We demonstrate that specific chromatin states are associated with the repetitive and non-repetitive *C. albicans* genome. While gene-rich regions are associated with active chromatin marks mirroring their transcriptional state, different types of repetitive elements are assembled into distinct chromatin types. Finally, we present the first *C. albicans* quantitative ChIP-seq (q-ChIP-seq) methodology that has permitted us to elucidate the roles of the HDAC Sir2 and the HMT Set1 in shaping the chromatin state of *C. albicans* genome and regulating gene expression.

RESULTS

Genome-wide histone modification profiling in C. albicans

The *C. albicans* genome contains two homologous pairs of divergently transcribed histone H2A and H2B genes, and histone H3 and H4 genes in addition to a single histone H3 gene (Fig S1 A). Sequence alignment demonstrated that the frequently modified amino acid residues H3K4, H3K9, H4K16 and H2AS129 are conserved in *C. albicans* (Fig S1 B).

To explore the chromatin signature of *C. albicans* repetitive and non-repetitive regions, we globally mapped the genomic locations of H3K4me^3, H3K9Ac, H4K16Ac

and γH2A by performing Chromatin ImmunoPrecipitation followed by high-throughput sequencing (ChIP-seq). Since nucleosomes are not equally distributed across genomes, we accounted for nucleosome occupancy by performing genome-wide profiling of unmodified histone H3 and histone H4. Finally, to correlate specific histone modification profiling with transcriptional activity, we mapped RNA polymerase II (RNAPII) occupancy genome-wide. In parallel, we performed transcriptome analysis by strand-specific RNA sequencing (RNA-seq) to profile gene expression levels.

For all samples, ChIP-seq was performed from *C. albicans* wild-type (WT) cells grown in standard laboratory growth conditions (YPAD 30 °C) using antibodies specific for modified or unmodified histones. Input (I) and Immunoprecipitated samples (Ip) were sequenced using the Illumina HiSeq2000 platform (single-end 50 bp reads; average coverage: 28x; Table S2, Dataset S1) and aligned to a custom haploid version of Assembly 22 of the *C. albicans* genome [25]. Unmodified histone H3 occupancy showed a strong positive correlation with histone H4 occupancy (Pearson correlation coefficient r = 0.97), with the exception of centromeric regions where the histone H3 variant Cse4^{CENP-A} replaces histone H3 (Fig 1A, 1C and S2). Furthermore, RNAPII occupancy showed a positive correlation with gene expression levels (Pearson correlation coefficient r = 0.72) (Fig 1B).

H3K4me^3, H3K9Ac and H4K16Ac mark C. albicans active genes

To delineate the chromatin signature of protein-coding *C. albicans* genes, enrichment profiles for each histone modification were compared to histone H4. Differential enrichment testing using DESeq2 allowed the identification of regions with statistically significant enrichment or depletion for particular histone marks compared to histone H4. We annotated these loci by proximity to annotated protein-

coding genes and non-coding RNAs [25–27]. For RNAPII, aligned reads from ChIP (IP) samples were normalised to aligned reads from the matching input (I) sample. Metagene analyses demonstrate that, as expected, RNAPII is enriched across all gene bodies while unmodified histone H3 is not significantly enriched or depleted relative to unmodified histone H4. In contrast, H3K4me^3 and H3K9Ac are more prominent at the transcriptional start site (TSS) and 5' regions of genes, and H4K16Ac is enriched at gene bodies. (Fig 2A).

To further explore the relationship between chromatin modifications and gene transcriptional states, we grouped all genes into four sets based on expression level (no expression, low expression, medium expression and high expression) as revealed by RNA-seq analysis (Fig S3). Enrichment profile plots of the levels of histone modifications for each of these gene sets demonstrated that H3K4me^3, H3K9Ac and H4K16Ac levels are very low at genes with low transcription rates. Levels of all modifications increase with increased gene expression reaching a maximum at highly transcribed genes (Fig 2B).

Therefore, in *C. albicans*, H3K4me^3, H3K9Ac and H4K16Ac correlate with gene transcription; H3K4me^3 and H3K9Ac are more enriched at the 5' of a gene and H4K16Ac at the gene bodies.

γ-H2A is enriched at convergent genes and in proximity of DNA replication origins

Having established that different regions of the *C. albicans* genome are marked by different chromatin modifications depending on their transcriptional state (Fig 2), we sought to systematically map the genome-wide profile of γH2A (γ-sites) in cycling undamaged cells, as this is a useful method to identify recombination-prone unstable sites [14]. Genome-wide ChIP-seq of γH2A identified 168 γ-sites where γH2A is

enriched compared to histone H4 (Dataset S1). *C. albicans* γ–sites are different from the γH2A -domain caused by irrecoverable DSBs as γ-sites have generally a single peak of enrichment and they are shorter (average length 850 bp) than the 50-kb length of the that γH2A domain surrounding DSBs [44].

Analysis of γ-sites indicates that they are present at three classes of genomic loci: *(i)* longer genes that are often convergent, *(ii)* origins of replication and *(iii)* subtelomeric regions (discussed below) (Fig 3A and 3B).

At convergent genes, γH2A enrichment is detected at both the gene bodies and intergenic regions, and no correlation was detected between gene expression levels and γH2A occupancy. Although we did not observe any correlation between γ-sites and histone H3 occupancy (Pearson Correlation r = 0.062), γ-sites are more likely to mark genomic regions that are acetylated on H4K16 and H3K9 (Pearson correlation r = 0.461 and 0.276, respectively). We also detect a weak negative correlation between γH2A occupancy and H3K4me^3 (Fig 3C).

The chromatin state of the C. albicans repetitive genome

Having determined the chromatin marks associated with *C. albicans* coding genes, we analysed the chromatin state of the *C. albicans* repetitive genome focusing on the major classes of DNA repeats: subtelomeric regions, the rDNA locus, MRS repeats and transposable elements (LTR and non-LTR retrotransposons). Sequence analysis of these elements can be problematic because of incomplete sequencing and their repetitive nature [25,29]. To estimate the chromatin modification state of these loci, we adopted a method previously applied to *S. cerevisiae* repeats and assumed that each repeat contributes equally to read-depth [45]. Consequently, reads that could not be uniquely mapped to one location were randomly assigned to copies of that repeat.

To investigate the chromatin state associated with the 16 subtelomeric regions in *C. albicans*, we analysed the ChIP-seq datasets in the 20-kb terminal regions of each chromosome arm. At these locations, occupancy of unmodified histone H3 was similar to histone H4 occupancy (Fig 4A, S4 and S5). In contrast, we detected large domains of chromatin that are hypomethylated on H3K4 and hypoacetylated on H3K9 and H4K16 (Fig 4A, S4 and S5). However, the H3K4 methylation and H3K9/H4K16 acetylation state of subtelomeres is not uniform as patches of high H3K4me^3, H3K9Ac and H4K16Ac are detected within each subtelomere (Fig 4A, S4 and S5). We detected statistically significant γH2A enrichment at 13/16 subtelomeres (Fig 4A, S4 and S5). We suspect that absence of γ-sites at ChrRR, Chr1R and Chr7L subtelomeric regions is due to incomplete genome assembly [25,29]. Subtelomeric γH2A enrichment is not uniform but present at distinct peaks within each subtelomere, which largely associate with hypoacetylated and hypomethylated chromatin (Fig 4A, S4 and S5).

Analysis of chromatin modifications associated with the rDNA locus demonstrate that the NTS1 and NTS2 regions are assembled into a chromatin structure resembling heterochromatin where nucleosomes are hypomethylated on H3K4 and hypoacetylated on H3K9 and H4K16 (Fig 4B). These findings are consistent with our published results demonstrating that these regions are assembled into transcriptionally silent heterochromatin [46]. Intriguingly, we detected two γ-sites at convergently transcribed genes surrounding the rDNA locus (Fig 4B).

This analysis also reveals that MRS repeats and retrotransposons (LTR and non-LTR) are associated with chromatin that is largely hypomethylated on H3K4 (Fig 4C, 4D). In contrast, H3K9Ac and H4K16Ac are similar to histone H4 levels (Fig 4C, 4D).

We did not detect any statistically significant enrichment of γ-H2A at either MRSs or retrotransposons.

Therefore, different *C. albicans* repetitive elements are associated with distinct chromatin states. Repetitive regions are more likely to be hypomethylated on H3K4, but are neither hypoacetylated on H3K9 and H4K16 nor enriched for γ-H2A.

The HDAC Sir2 governs the hypoacetylated state associated with C. albicans rDNA locus and subtelomeric regions

We have previously shown that, in *C. albicans*, the histone deacetylase Sir2 maintains the low level of H3K9Ac associated with the NTS regions of the rDNA locus [46]. To assess the role of the HDAC Sir2 in maintaining acetylation levels across the *C. albicans* genome, we performed H3K9Ac and H4K16Ac ChIP-seq analyses in WT and *sir2Δ/Δ* strains.

Traditional ChIP-seq are not inherently quantitative as it allows comparison of protein occupancies at different positions within a genome but it does not allow direct comparisons between samples derived from different strains [47–49]. To overcome this issue, we adapted *C. albicans* to a quantitative ChIP-seq (q-ChIP-seq) methodology [47–49]. To this end, WT and *sir2Δ/Δ* were spiked-in, at the time of fixation, with a single calibration sample from *S. cerevisiae* (Fig 5A). *S. cerevisiae* genome is a desirable exogenous reference for *C. albicans* cells because its genome is well studied and has a high-quality sequence assembly [50]. Moreover, reads originating from *C. albicans* or *S. cerevisiae* can be easily separated at the analysis level and our experiments reveal less than 2% of the total number of reads cannot be uniquely mapped (Table S2). Finally, histone proteins are well conserved between *C. albicans* and *S. cerevisiae* (Fig S1) and therefore the same histone

antibody is likely to immunoprecipitate *C. albicans* and *S. cerevisiae* chromatin with the same efficiency.

The q-ChIP-seq analyses identify only two regions of the *C. albicans* genome with increased H3K9Ac and H4K16Ac levels: subtelomeric regions and the NTS region of the rDNA locus (Fig 5B, 5C, S6, S7 and Dataset S1). Deletion of *SIR2* does not lead to increased histone acetylation levels at euchromatic regions or at other repetitive elements such as MRS and retrotransposons (Dataset S1). In agreement with these findings, the majority (83%) of gene expression changes observed in *sir2Δ/Δ* cells occur at the rDNA locus and subtelomeric regions (Fig 5B, 5C, Dataset S1 and [46])

We conclude that the *C. albicans* HDAC Sir2 acts exclusively at two genomic regions: the rDNA locus and subtelomeric regions. Our findings are consistent with the hypothesis that Sir2-mediated histone deacetylation represses gene expression at these locations.

Set1-dependent methylation of H3K4 impacts gene expression differentially at different repeats

Our data demonstrates that *C. albicans* repetitive elements are associated with chromatin that is hypomethylated on H3K4. However, at these regions, H3K4 methylation is not completely ablated as pockets of H3K4me^3 are detected (Fig 4). In *S. cerevisiae* and *S. pombe*, the H3K4 methyltransferase Set1 has been implicated in both gene repression and activation [51–56]. *S. cerevisiae* Set1 also maintains the transcriptional silencing associated with heterochromatic regions such as the telomeres and the rDNA locus [51–56]. *C. albicans* Set1 is important for efficient yeast-to-hyphae switching but its function in regulating chromatin structure and gene expression is unknown [57].

To gain insights into the role of *C. albicans* Set1, we performed H3K4me^3 q-ChIP-seq and RNA-seq analyses of WT and *set1Δ/Δ* strains. *C. albicans* Set1 clearly plays a major role in maintaining chromatin structure as 6846 loci, scattered throughout the genome, have a statistically significant reduction of H3K4me^3 in *set1Δ/Δ* compared to WT strain (Fig 6A and Dataset S1). RNA-seq analysis reveals that Set1 regulates gene expression both positively and negatively, as genes with a reduced H3K4me^3 pattern can be either upregulated (2320 genes/ ncRNAs) or downregulated (3184 genes/ ncRNAs) in *set1Δ/Δ* compared to WT (Fig 6A and Dataset S1). Analyses of the H3K4me^3 pattern and gene expression levels associated with repetitive elements demonstrates that Set1 has distinct roles at different repeats. At subtelomeric regions and the rDNA locus, deletion of the *SET1* gene leads to the reduced H3K4me^3 levels and is accompanied by the down-regulation of associated genes (Fig 6B, 6C and S8). In contrast, the reduced H3K4me^3 pattern associated with MRS repeats in the *set1Δ/Δ* strain leads to increased expression of coding and non-coding RNAs originating from MRS repeats (Fig 6D). Finally, deletion of *SET1* leads to decreased H3K4me^3 at retrotransposons without a significant impact on expression of retrotransposon-associated coding and non-coding RNAs.

We conclude that Set1 is the major H3K4-methyltransferase *in C. albicans* playing a key role in controlling chromatin structure and gene expression. Importantly, our analysis reveals that although deletion of *SET1* results in decreased H3K4 methylation across all repetitive elements, Set1 influences gene expression differentially at each repetitive element.

DISCUSSION

Here, we present the first comprehensive chromatin profiling of histone modifications associated with the *C. albicans* genome. Furthermore, we present the first *C.*

337 *albicans* quantitative ChIP-seq to delineate the role of the chromatin modifiers Sir2
338 and Set1 in *C. albicans*.

339 ***The chromatin state of the C. albicans repetitive and non-repetitive genome***

340 In all organisms, gene-rich genomic regions are associated with a histone
341 modification pattern mirroring their transcriptional state where H3K4me^3 and H3K9Ac
342 are enriched at active gene promoters and H4K16Ac is localised at gene bodies of
343 expressed genes [9,10]. Our first objective was to obtain "proof of concept"
344 epigenomic maps of chromatin modifications associated with gene rich regions of the
345 *C. albicans* genome. A robust histone modification profiling relies on *(i)* the use of
346 antibodies that recognise modified histones with high specificity and *(ii)* the use of
347 appropriate biological controls. Specificity of antibodies used in this study has been
348 tested in *S. pombe* or *S. cerevisiae* histone mutants lacking the modifiable amino
349 acid (H3K9, H4K16, H2AS129) [14,58]. To distinguish between nucleosome
350 occupancy and depletion/enrichment of specific histone modifications ChIP seq
351 analyses was also performed using antibodies recognising unmodified histone H3
352 and H4. This is an important control that should be included in all studies aimed to
353 analyse chromatin modification genome wide.
354 Our results confirm the validity of our experimental approach and conform to the
355 chromatin pattern reported in other organisms by showing that active genes,
356 associated with high levels of RNA Pol II, are assembled into canonical euchromatin
357 where H3K4me^3 and H3K9Ac are associated with promoters and H4K16Ac is
358 enriched at gene bodies. We conclude that in *C. albicans*, as in other organisms, a
359 specific histone modification pattern is predictive of active transcription.
360 Analyses of the chromatin profiling of repetitive elements demonstrate that repetitive
361 regions associated with the NTS regions of the rDNA locus and subtelomeric regions

are packaged into chromatin resembling the heterochromatic structure of other organisms, like the budding yeast *S. cerevisiae*, lacking H3K9me/H3K27me systems [45]. These findings are in agreement with our previous study demonstrating that the rDNA locus and subtelomeric regions are able to silence embedded marker genes, which is a hallmark of heterochromatic regions [46]. In contrast, we find that *C. albicans* retrotransposons and MRS repeats are assembled into a distinct chromatin state where nucleosomes are hypomethylated on H3K4me^3, but also acetylated on H3K9 and H4K16. Analyses of clinical isolates and *in vivo* evolution experiments have demonstrated that, in the host, MRSs and transposons are recombination hotspots as they are known sites of translocations [32,38,59]. Given the key roles of chromatin in regulating genome stability, it will be important to investigate whether the chromatin packaging of MRS and transposons regulates genome stability.

γ-H2A decorates C. albicans heterochromatic regions and potential recombination-prone unstable sites

The genome-wide γ-H2A profiling performed in this study reveals that this histone modification, a hallmark of DNA damage, is enriched at heterochromatic regions assembled into hypoacetylated chromatin that is also hypomethylated on H3K4. This is similar to observations in other fungal organisms where γ-H2A decorates heterochromatic regions [14–17]. In contrast, we did not detect any significant enrichment of γ-H2A at other repetitive elements such as MRS repeats and transposable elements. This is surprising because, in the host, MRS repeats are recombination hot-spots [32,38] and therefore a place where DSBs might be expected to accumulate. *C. albicans* genome instability is increased under host-relevant stresses [60,61] and therefore we propose that the recombination potential of MRSs is unlocked following exposure the specific host niche stresses.

Finally, we detected 168 additional γ-sites located in proximity of origins of replication or convergent genes that are often long. DNA replication origins are known replication fork barriers in many organisms and read-through transcription of convergent genes can also cause genome instability by, for example, R-loop formation [62]. Therefore, we propose that the γ-sites identified in this study represent novel recombination-prone unstable sites of the *C. albicans* genome.

The role of the histone deacetylase Sir2 and the histone methyltransferase Set1 in controlling the C. albicans epigenome

We present the first quantitative ChIP-seq in *C. albicans* that has allowed us to delineate the roles of the histone modifying enzymes Sir2 and Set1. We demonstrate that Sir2 maintains the hypoacetylated state of heterochromatic regions associated with the rDNA locus and subtelomeric regions. Sir2 deacetylation at these loci is linked to gene repression as shown by RNA-seq analysis. In contrast, we find that deletion of Sir2 does not lead to increased histone acetylation and gene expression at other genomic regions. Two possible scenarios could explain these findings: *(i*) Sir2 is specifically targeted to subtelomeres and the rDNA locus or *(ii)* other histone deacetylases act redundantly to Sir2 regulating hypoacetylation and gene expression at other genomic locations.

We present evidence demonstrating that the HMT Set1 is a major contributor to chromatin structure in *C. albicans*. Indeed, deletion of *SET1* leads to an almost complete ablation of H3K4 methylation and is linked to extensive gene expression changes. This demonstrates that *C. albicans* Set1 is the major H3K4 methyltransferase. It is particularly intriguing that deletion of *SET1* leads to decreased H3K4me^3 at all known repeats, yet its effect on gene expression can be

411 the opposite. Indeed, we demonstrate that at the rDNA locus and subtelomeric
412 regions Set1 represses gene expression while it activates gene expression at MRS
413 repeats. Further studies will untangle the role of Set1 at different genomic regions.

414 **CONCLUSIONS**

415 In this study we present the first epigenomic map of histone modifications associated
416 with the *C. albicans* genome. Given the key role of chromatin in regulating *C.*
417 *albicans* biology, the data generated in this study provide an invaluable resource to a
418 better understanding of this important human fungal pathogen.

419 **METHODS**

420 ***Yeast growth and manipulation***

421 Strains used in this study are listed in the Table S1. Yeast cells were cultured in
422 YPAD broth containing 1% yeast extract, 2% peptone, 2% dextrose, 0.1 mg/ml
423 adenine and 0.08 mg/ml uridine at 30^{o}C.

424 ***Antibody Information***

425 The following antibodies were used in this study: anti-H2AS129p (Millipore; Cat No:
426 07-745-I), anti-H3 (Abcam; Cat No: ab1791), anti-H4 (Millipore; Cat No: 05-858),
427 anti-H3K4me3 (Active Motif; Cat No: 39159), anti-H3K9ac (Active Motif; Cat No:
428 39137), anti-H4K16ac (Active Motif; Cat No: 39167), and anti-RNA Polymerase II
429 (BioLegend; Cat No: 664903).

430 ***ChIP-seq***

431 Chromatin immunoprecipitation with deep-sequencing (ChIP-seq) was performed as
432 follows: 5 ml of an overnight culture grown in YPAD was diluted into fresh YPAD and
433 grown until the exponential phase (OD_{600} = 0.6-0.8). 20 OD_{600} units of cells were
434 fixed with 1% formaldehyde (Sigma) for 15 minutes at room temperature. Reactions

were quenched by the addition of glycine to a final concentration of 125 mM. Cells were lysed using acid-washed glass beads (Sigma) and a DisruptorGenie (Scientific Industries) for four cycles of 30 minutes at 4°C with 5 minutes on ice between cycles. Chromatin was sheared to 200–500 bp using a BioRuptor sonicator (Diagenode) for a total of 20 minutes (30 seconds on, 30 seconds off cycle) at 4°C. Immunoprecipitation was performed overnight at 4°C using 2 µl of the appropriate antibody and 25 µl of protein G magnetic Dynabeads (Invitrogen). ChIP DNA was eluted, and cross-links reversed at 65°C in the presence of 1% SDS. All samples were then treated with RNaseA and proteinase K before being purified by phenol:chloroform extraction and ethanol precipitation. Libraries were prepared and sequenced as 50bp single end reads on an Illumina Hi seq2000 platform by the Genomics Core Facility at EMBL (Heidelberg, Germany). All ChIP-seq experiments were carried out in biological duplicates.

q-ChIP-seq

Quantitative chromatin immunoprecipitation with deep-sequencing (q-ChIP-seq) was performed similarly to the ChIP-seq method, except: 5 ml of an overnight culture of the *S. cerevisiae* reference strain BY4741 was grown alongside *C. albicans* in YPAD. These cultures were then diluted into fresh YPAD and grown until the exponential phase (OD_{600} = 0.6-0.8). 20 OD_{600} units of *C. albicans* cells were combined with 10 OD_{600} units of *S. cerevisiae* cells, and then fixed with 1% formaldehyde (Sigma) for 15 minutes at room temperature. After the cells had been fixed, the q-ChIP-seq sample was processed as a single ChIP-seq sample throughout the experiment until completion of DNA sequencing. All q-ChIP-seq experiments were carried out in biological duplicates.

459 ***RNA-seq***

460 RNA was extracted from exponential cultures (OD_{600} = 0.6-0.8) using a yeast RNA
461 extraction kit (E.Z.N.A. Isolation Kit RNA Yeast; Omega Bio-Tek) following the
462 manufacturer's instructions. RNA quality was checked by electrophoresis under
463 denaturing conditions in 1% agarose, 1x HEPES, 6% formaldehyde (Sigma). RNA
464 concentration was measured using a NanoDrop ND-1000 spectrophotometer.
465 Strand-specific cDNA Illumina barcoded libraries were generated from 1 µg of total
466 RNA and sequenced as 50bp single end reads using an Illumina Hi seq2000
467 sequencer by the Genomics Core Facility at EMBL (Heidelberg, Germany). All RNA-
468 seq experiments were carried out in biological duplicates.

469 ***Analysis of high-throughput sequencing***

470 All datasets generated and analysed during the current study are available in the
471 BioProject NCBI repository (https://www.ncbi.nlm.nih.gov/bioproject) under the
472 BioProject ID (PRJNA503946).

473 *ChIP-seq Analysis:*

474 Illumina reads were mapped using Bowtie2 [63] to a custom haploid version of
475 assembly 22 of the *C. albicans* genome (Table S2). Reads that mapped to repeated
476 sequences were randomly assigned to copies of that repeat, allowing for an
477 estimation of enrichment at the repetitive elements of the genome. Peak calling was
478 performed using MACS2 [64] on the default settings, except that no model was used
479 with all reads extended to 250 bp. MACS2 was run separately on both biological
480 replicates for each ChIP-seq sample. For each sample analysed with MACS2, the IP
481 sample was the "treatment," and the input sample was the "control." We defined
482 peaks as reproducible if they were called in both data sets. Read counts within peak
483 intervals were generated using featureCounts [65]. For each interval, biological

duplicate counts were compared between each histone modification and unmodified histone H4 samples using DESeq2, with an adjusted p-value threshold of <0.05 being used to identify significant differences. Replicates were compared by generating a raw alignment coverage track and performing a Pearson correlation between them using the multiBamSummary and plotCorrelation tools as part of the deepTools2 package (Fig S9) [66]. Genome coverage tracks were made using the pileup function of MACS2 [64] and tracks from biological replicates were averaged after the replicates were deemed to be sufficiently correlative ($r > 0.9$). For each coverage track, reads per million (RPM) were calculated. The histone modification coverage tracks were normalised to unmodified histone H4, and the RNAPII track was normalised to the respective input sample. All coverage tracks were visualised using IGV [67]. Metaplots and heatmaps were made using computeMatrix, plotProfile and plotHeatmap tools as part of the deepTools2 package [66].

q-ChIP-seq Analysis:

To isolate the reads that uniquely aligned to the *C. albicans* genome, the full datasets were first aligned to the *S. cerevisiae* genome (sacCer3). The unaligned reads were output as separate fastq files, and then these reads were aligned to a custom haploid version of assembly 22 of the *C. albicans* genome (Table S2). The same strategy was used to isolate reads that uniquely aligned to *S. cerevisiae* (Table S2). All alignments were performed using Bowtie2 [63]. The unique *S. cerevisiae* reads were then used to calculate the normalisation factor (normalisation factor = 1 / [unique reference reads/1,000,000]), according to Orlando et al. [48]. Reads that mapped to repeated sequences in the *C. albicans* genome were randomly assigned to copies of that repeat. Peak calling was performed using MACS2 [64] on the default settings, except that no model was used with all reads extended to 250bp.

MACS2 was run separately on both biological replicates for each ChIP-seq sample. For each sample analysed with MACS2, the IP sample was the "treatment," and the input sample was the "control." Peaks called in both replicate datasets for mutant and WT samples were combined into one peak set for each histone modification. Read counts within these peak intervals were generated using featureCounts (Liao et al. 2014), which were then scaled by the normalisation factor to obtain the reference reads per million (RRPM). For each interval, RRPM values were compared between the mutant and WT samples using a two-sample t-test, with a p-value threshold of <0.05 being used to identify significant differences. Replicates were compared by generating a raw alignment coverage track and performing a Pearson correlation between them using the multiBamSummary and plotCorrelation tools as part of the deepTools2 package (Fig S9) [66]. Genome coverage tracks were made using the pileup function of MACS2 [64] and for each track, RRPM values were calculated using the normalisation factor. Coverage tracks from biological replicates were averaged after the replicates were deemed to be sufficiently correlative ($r > 0.9$), and the mutant strain coverage tracks were normalised to the WT coverage. All tracks were visualised using IGV [67]. Metaplots and heatmaps were made using computeMatrix, plotProfile and plotHeatmap tools as part of the deepTools2 package [66].

RNA-seq Analysis:

Reads were aligned to a custom haploid version of assembly 22 of the *C. albicans* genome using HISAT2 (Table SX) [68], and per-gene transcript quantification was performed using featureCounts, which discards multi-mapped read fragments; therefore, only uniquely mapped reads were included for the expression analysis [65]. Differential expression testing was performed using DESeq2, with an adjusted

534 p-value threshold of <0.05 being used to determine statistical significance.
535 Replicates were compared by generating a raw alignment coverage track and
536 performing a Pearson correlation between them using the multiBamSummary and
537 plotCorrelation tools as part of the deepTools2 package (Fig S10) [66]. Scatterplots
538 and correlation analyses were performed in R using Pearson correlation.
539

540 **DECLARATION**

541 *Ethics approval and consent to p*articipate: 'Not applicable'; *Consent for publication*:
542 'Not applicable'; *Availability of data and material:* The datasets generated and
543 analysed during the current study are available in the BioProject NCBI repository
544 (https://www.ncbi.nlm.nih.gov/bioproject) under the BioProject ID (PRJNA503946).;
545 *Competing interests*: 'Not applicable'**;** *Funding*: This work was supported by MRC
546 (MR/M019713/1 to A.B., R.J.P.) and an ERC Grant (340087, RAPLODAPT to J.B.)
547 *Authors' contributions* R.J.P conducted the ChIP seq , the RNA-seq and the
548 bioinformatics analyses. E.W performed RNAseq experiments of *sir2* Δ/Δ strain. A.B
549 and J.B conceived the project, designed the experiments and wrote the manuscript.
550 *Acknowledgements*: We thank members of the Kent Fungal Group, Jan Soetaert and
551 Alison Pidoux for discussion and critical reading of the manuscript. We thank the
552 Gene Core Facility at EMBL (Heidelberg-Germany) for Illumina Sequencing.

553 **REFERENCES**

554 1. Allshire RC, Madhani HD. Ten principles of heterochromatin formation and
555 function. Nat. Rev.;19:229–44.
556 2. Strahl BD, Allis CD. The language of covalent histone modifications. Nature.
557 2000;403:41–5.
558 3. Allis CD, Jenuwein T. The molecular hallmarks of epigenetic control. Nat. Rev.

559 Genet. 2016;17:487–500.

560 4. Van HT, Santos MA. Histone Modifications and the DNA Double-Strand Break

561 Response. Cell Cycle Taylor & Francis; 2018;0:15384101.2018.1542899.

562 5. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al.

563 Mapping and analysis of chromatin state dynamics in nine human cell types. Nature.

564 2011;473:43–9.

565 6. Guttman M, Amit I, Garber M, French C, Lin MF, Feldser D, et al. Chromatin

566 signature reveals over a thousand highly conserved large non-coding RNAs in

567 mammals. Nature 2009;458:223–7.

568 7. Janssen A, Colmenares SU, Karpen GH. Heterochromatin: Guardian of the

569 Genome. Annu. Rev. Cell Dev. Biol. 2018;34:annurev-cellbio-100617-062653.

570 8. Howe FS, Fischl H, Murray SC, Mellor J. Is H3K4me3 instructive for transcription

571 activation? BioEssays. 2017;39:1–12.

572 9. Wang Z, Zang C, Rosenfeld JA, Schones DE, Barski A, Cuddapah S, et al.

573 Combinatorial patterns of histone acetylations and methylations in the human

574 genome. Nat. Genet. 2008;40:897–903.

575 10. Heintzman ND, Stuart RK, Hon G, Fu Y, Ching CW, Hawkins RD, et al. Distinct

576 and predictive chromatin signatures of transcriptional promoters and enhancers in

577 the human genome. Nat. Genet. 2007;39:311–8.

578 11. Padeken J, Zeller P, Gasser SM. Repeat DNA in genome organization and

579 stability. Curr. Opin. Genet. Dev.; 2015;31:12–9.

580 12. Nair N, Shoaib M, Sørensen CS. Chromatin dynamics in genome stability: Roles

581 in suppressing endogenous DNA damage and facilitating DNA repair. Int. J. Mol. Sci.

582 2017;18:1–21.

583 13. Shaver S, Casas-Mollano JA, Cerny RL, Cerutti H. Origin of the polycomb

584 repressive complex 2 and gene silencing by an E(z) homolog in the unicellular alga
585 Chlamydomonas. Epigenetics. 2010;5:301–12.
586 14. Szilard RK, Jacques P-E, Laramée L, Cheng B, Galicia S, Bataille AR, et al.
587 Systematic identification of fragile sites via genome-wide location analysis of
588 gamma-H2AX. Nat. Struct. Mol. Biol. 2010;17:299–305.
589 15. Kitada T, Schleker T, Sperling AS, Xie W, Gasser SM, Grunstein M. γH2A is a
590 component of yeast heterochromatin required for telomere elongation. Cell Cycle.
591 2011;10:293–300.
592 16. Rozenzhak S, Mejía-Ramírez E, Williams JS, Schaffer L, Hammond J a, Head
593 SR, et al. Rad3 decorates critical chromosomal domains with gammaH2A to protect
594 genome integrity during S-Phase in fission yeast. PLoS Genet. 2010 ;6:e1001032.
595 17. Sasaki T, Lynch KL, Mueller C V., Friedman S, Freitag M. et al Heterochromatin
596 controls γH2A localization in Neurospora crassa. Eukaryot. Cell. 2014;13:990–1000.
597 18. Meier A, Fiegler H, Muñoz P, Ellis P, Rigler D, Langford C, et al. Spreading of
598 mammalian DNA-damage response factors studied by ChIP-chip at damaged
599 telomeres. EMBO J.; 2007;26:2707–18.
600 19. Iacovoni JS, Caron P, Lassadi I, Nicolas E, Massip L, Trouche D, et al. High-
601 resolution profiling of gammaH2AX around DNA double strand breaks in the
602 mammalian genome. EMBO J.; 2010;29:1446–57.
603 20. Clouaire T, Rocher V, Lashgari A, Arnould C, Aguirrebengoa M, Biernacka A, et
604 al. Comprehensive Mapping of Histone Modifications at DNA Double-Strand Breaks
605 Deciphers Repair Pathway Chromatin Signatures. Mol. Cell . 2018;1–13.
606 21. Wang J, Jia ST, Jia S. New Insights into the Regulation of Heterochromatin.
607 Trends Genet. 2016;32:284–94.
608 22. De Las Peñas A, Juárez-Cepeda J, López-Fuentes E, Briones-Martín-del-Campo

609 M, Gutiérrez-Escobedo G, Castaño I. Local and regional chromatin silencing in
610 *Candida glabrata*□: consequences for adhesion and the response to stressFEMS
611 Yeast Res. 2015;15:fov056.
612 23. Brown GD, Denning DW, Gow NAR, Levitz SM, Netea MG, White TC. Hidden
613 Killers: Human Fungal Infections. Sci. Transl. Med. 2012;4:165rv13-165rv13.
614 24. Selmecki A, Forche A, Berman J. Genomic plasticity of the human fungal
615 pathogen Candida albicans. Eukaryot. Cell. 2010;9:991–1008.
616 25. Muzzey D, Schwartz K, Weissman JS, Sherlock G. Assembly of a phased diploid
617 Candida albicans genome facilitates allele-specific measurements and provides a
618 simple model for repeat and indel structure. Genome Biol. 2013;14:R97.
619 26. Bruno VM, Wang Z, Marjani SL, Euskirchen GM, Martin J, Sherlock G, et al.
620 Comprehensive annotation of the transcriptome of the human fungal pathogen
621 Candida albicans using RNA-seq. Genome Res. 2010;20:1451–8.
622 27. Sellam A, Hogues H, Askew C, Tebbji F, van Het Hoog M, Lavoie H, et al.
623 Experimental annotation of the human pathogen Candida albicans coding and
624 noncoding transcribed regions using high-resolution tiling arrays. Genome Biol.
625 2010;11:R71.
626 28. Tuch BB, Mitrovich QM, Homann OR, Hernday AD, Monighetti CK, De La Vega
627 FM, et al. The transcriptomes of two heritable cell types illuminate the circuit
628 governing their differentiation. Copenhaver GP, editor. PLoS Genet.
629 2010;6:e1001070.
630 29. van het Hoog M, Rast TJ, Martchenko M, Grindle S, Dignard D, Hogues H, et al.
631 Assembly of the Candida albicans genome into sixteen supercontigs aligned on the
632 eight chromosomes. Genome Biol. 2007;8:R52.
633 30. McEachern MJ, Hicks JB. Unusually large telomeric repeats in the yeast Candida

albicans. Mol. Cell. Biol. 1993;13:551–60. A

31. Jones T, Federspiel NA, Chibana H, Dungan J, Kalman S, Magee BB, et al. The diploid genome sequence of Candida albicans. Proc. Natl. Acad. Sci. U. S. A. 2004;101:7329–34.

32. Chibana H, Magee PT. The enigma of the major repeat sequence of Candida albicans. Future Microbiol. 2009;4:171–9.

33. Uhl MA, Biery M, Craig N, Johnson AD. Haploinsufficiency-based large-scale forward genetic analysis of filamentous growth in the diploid human fungal pathogen C.albicans. EMBO J. 2003;22:2668–78.

34. Goodwin TJD. Multiple LTR-Retrotransposon Families in the Asexual Yeast Candida albicans. Genome Res. 2000 ;10:174–91.

35. Zhang L, Yan L, Jiang J, Wang Y, Jiang Y, Yan T, et al. The structure and retrotransposition mechanism of LTR-retrotransposons in the asexual yeast Candida albicans. Virulence. 2014;5:1–10.

36. Goodwin TJD, Ormandy JE, Poulter RTM. L1-like non-LTR retrotransposons in the yeast Candida albicans. Curr. Genet. 2001;39:83–91.

37. Hirakawa MP, Martinez DA, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in Candida albicans. Genome Res. 2015;25:413–25.

38. Ene I V., Farrer RA, Hirakawa MP, Agwamba K, Cuomo CA, Bennett RJ. Global analysis of mutations driving microevolution of a heterozygous diploid fungal pathogen. Proc. Natl. Acad. Sci. 2018;115:201806002.

39. Freire-Benéitez V, Price RJ, Buscaino A. The Chromatin of Candida albicans Pericentromeres Bears Features of Both Euchromatin and Heterochromatin. Front. Microbiol. 2016;7:759.

40. Peterson M, Price RJ, Gourlay S, May A, Tullet J, Buscaino A. The fungal-specific Hda2 and Hda3 proteins regulate morphological switches in the human fungal pathogen Candida albicans. bioRxiv. Cold Spring Harbor Laboratory; 2018;340364.

41. Anderson MZ, Bennett RJ. Budding off: Bringing functional genomics to Candida albicans. Brief. Funct. Genomics. 2016;15:85–94.

42. Kim J, Lee J-E, Lee J-S. Histone deacetylase-mediated morphological transition in Candida albicans. J. Microbiol. 2015;53:805–11.

43. Thakur J, Sanyal K. Efficient neocentromere formation is suppressed by gene conversion to maintain centromere function. Genome Research. 2013;638–52.

44. Shroff R, Arbel-Eden A, Pilch D, Ira G, Bonner WM, Petrini JH, et al. Distribution and Dynamics of Chromatin Modification Induced by a Defined DNA Double-Strand Break. Curr. Biol; 2004;14:1703–11.

45. Ellahi A, Thurtle DM, Rine J. The Chromatin and Transcriptional Landscape of Native Saccharomyces cerevisiae Telomeres and Subtelomeric Domains. Genetics. 2015;200:505–21.

46. Freire-Benéitez V, Price RJ, Tarrant D, Berman J, Buscaino A. Candida albicans repetitive elements display epigenetic diversity and plasticity. Sci. Rep. 2016;6:22989.

47. Hu B, Petela N, Kurze A, Chan KL, Chapard C, Nasmyth K. Biological chromodynamics: A general method for measuring protein occupancy across the genome by calibrating ChIP-seq. Nucleic Acids Res. 2015;43.

48. Orlando DA, Chen MW, Brown VE, Solanki S, Choi YJ, Olson ER, et al. Quantitative ChIP-Seq normalization reveals global modulation of the epigenome. Cell Rep; 2014;9:1163–70.

684 49. Helman E, Lawrence MS, Stewart C, Sougnez C, Getz G, Meyerson M, et al.
685 Quantifying ChIP-seq data□: a spiking method providing an internal reference for A
686 cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification.
687 Genome Res. 2014;24:1157–68.
688 50. Schwartz K, Sherlock G. High-Throughput Yeast Strain Sequencing. Cold Spring
689 Harb. Protoc. 2016:pdb.top077651.
690 51. Krogan NJ, Dover J, Khorrami S, Greenblatt JF, Schneider J, Johnston M, et al.
691 COMPASS, a histone H3 (Lysine 4) methyltransferase required for telomeric
692 silencing of gene expression. J. Biol. Chem. 2002;277:10753–5.
693 52. Bryk M, Briggs SD, Strahl BD, Curcio MJ, Allis CD, Winston F. Evidence that
694 Set1, a factor required for methylation of histone H3, regulates rDNA silencing in S.
695 cerevisiae by a Sir2-independent mechanism. Curr. Biol.. 2002;12:165–70.
696 53. Margaritis T, Oreal V, Brabers N, Maestroni L, Vitaliano-Prunier A, Benschop JJ,
697 et al. Two Distinct Repressive Mechanisms for Histone 3 Lysine 4 Methylation
698 through Promoting 3′-End Antisense Transcription. PLoS Genet. 2012;8.
699 54. Ramakrishnan S, Pokhrel S, Palani S, Pflueger C, Parnell TJ, Cairns BR, et al.
700 Counteracting H3K4 methylation modulators Set1 and Jhd2 co-regulate chromatin
701 dynamics and gene transcription. Nat. Commun. 2016;7:11949.
702 55. Weiner A, Hsieh THS, Appleboim A, Chen H V., Rahat A, Amit I, et al. High-
703 resolution chromatin dynamics during a yeast stress response. Mol. Cell.
704 2015;58:371–86.
705 56. Lorenz DR, Meyer LF, Grady PJR, Meyer MM, Cam HP. Heterochromatin
706 assembly and transcriptome repression by Set1 in coordination with a class II
707 histone deacetylase. Elife. 2014;3:1–17.
708 57. Raman SB, Nguyen MH, Zhang Z, Cheng S, Jia HY, Weisner N, et al. Candida

albicans SET1 encodes a histone 3 lysine 4 methyltransferase that contributes to the pathogenesis of invasive candidiasis. Mol. Microbiol. 2006;60:697–709.

58. Mellone BG, Ball L, Suka N, Grunstein MR, Partridge JF, Allshire RC. Centromere Silencing and Function in Fission Yeast Is Governed by the Amino Terminus of Histone H3. Curr. Biol. Cell Press; 2003;13:1748–57.

59. Hirakawa MP, Martinez D a, Sakthikumar S, Anderson MZ, Berlin A, Gujja S, et al. Genetic and phenotypic intra-species variation in Candida albicans. 2015;1–13.

60. Forche A, Abbey D, Pisithkul T, Weinzierl MA, Ringstrom T, Bruck D, et al. Stress alters rates and types of loss of heterozygosity in Candida albicans. MBio. 2011;2.

61. Freire-Benéitez V, Gourlay S, Berman J, Buscaino A. Sir2 regulates stability of repetitive domains differentially in the human fungal pathogen Candida albicans. Nucleic Acids Res. 2016;44.

62. Santos-Pereira JM, Aguilera A. R loops: new modulators of genome dynamics and function. Nat. Rev. Genet. 2015;16:583–97.

63. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat. Methods [Internet]. 2012;9:357–9.

64. Feng J, Liu T, Qin B, Zhang Y, Liu XS. Identifying ChIP-seq enrichment using MACS. Nat. Protoc. 2012;7:1728–40.

65. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics. 2014;30:923–30.

66. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, et al. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res. 2016;44:W160-5.

67. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV):

high-performance genomics data visualization and exploration. Brief. Bioinform. 2013;14:178–92.

68. Kim D, Langmead B, Salzberg SL. HISAT: a fast spliced aligner with low memory requirements. Nat. Methods. 2015;12:357–60.

ADDITIONAL FILES

Additional file 1- 10: Fig S1 to S10. pdf

Additional file 11: Table S1: Strains used in this study

Additional file 12: Table S2.xsl ; Sequencing and coverage information

Additional file 13: Dataset S1.xsl ; Datasets generated in this study

FIGURE LEGENDS

Figure 1 Histones and RNA Polymerase II occupancy **(A)** Correlation between H3 and H4 occupancy (log10 counts at 1 kb bins) across the *C. albicans* genome. **(B)** Correlation between RNAPII occupancy (log10 counts) and transcriptional levels (RNA-seq; log10 counts) at protein-coding genes. **(C)** Histone H3 is depleted at centromeric regions. Fold enrichment (log2) of histone H3 relative to unmodified H4 across CEN1 (Chr1) and CEN5 (Chr5) centromeric and pericentromeric regions in *C. albicans*. The Cse4^{CENP-A} enrichment profile [43] is shown as comparison. The blue bar indicates statistically significant depleted regions for histone H3.

Figure 2 *C. albicans* chromatin modifications mirror their transcriptional state **(A)** Chromatin signature of *C. albicans* genes (n=6408). Average profiles and heatmaps of histone modification signatures around the Transcriptional Start Sites (TSS) of genes. The relative fold enrichment (log2) for each histone modification normalised to unmodified histone H4, or aligned reads of immunoprecipitated (IP) sample normalised to aligned reads of Input sample (for RNAPII ChIP-seq) is displayed within a region spanning ±0.5 kb around TSS. The gradient blue-to-red colour

indicates high-to-low enrichment in the corresponding region. Min: - 1.5 log2. Max: + 1.5 log2. **(B)** Average profiles of histone modifications and RNAPII occupancy across gene sets of different expression level (no [n = 416], low [n = 1369], medium [n = 3570] and high [n = 983] expression). For each histone modification, the fold enrichment (log2) relative to unmodified H4 is shown. For RNAPII the IP/I enrichment (log2) is shown.

Figure 3 Identification of *C. albicans* γ-sites **(A)** Locations and frequencies of γ-sites throughout the *C. albicans* genome. **(B)** γ-sites map to longer genes. Histogram showing the gene lengths of γ-sites (red) compared to the genome average (grey) **(C)** Chromatin signature of γ-sites throughout the *C. albicans* genome. Average profiles and heatmaps of histone modification signatures at γ-sites. The relative fold enrichment (log2) for each histone modification normalised to unmodified histone H4 is displayed within a region spanning ±2 kb around the γH2A peak summits. The gradient blue-to-red colour indicates high-to-low enrichment in the corresponding region. Min: - 1.5 log2. Max: + 1.5 log2.

Figure 4 Chromatin signature of *C. albicans* repetitive elements **(A)** *Top*: Fold enrichment (log2) of H3K4me^3, H3K9Ac, H4K16Ac, γH2A and H3 relative to unmodified H4 across the 20 kb right terminal region of chromosome 6 (Chr6). *Middle*: Diagram of coding genes found at these regions, according to assembly 22. *Bottom*: Diagram depicting statistically significant enriched (red) or depleted (blue) domains for each histone modification. **(B)** *Top*: Fold enrichment (log2) of H3K4me3, H3K9Ac, H4K16Ac, γH2A and H3 relative to unmodified H4 at the rDNA locus and flanking regions (ChrR). *Middle*: Diagram of coding genes (white) and ncRNA (grey) found at this region, according to assembly 22. *Bottom*: Diagram depicting statistically significant enriched (red) or depleted (blue) domains for each histone

784 modification **(C)** Average profiles of histone modifications at MRS repeats, and
785 upstream and downstream sequences. The grey arrow indicates the location of the
786 FGR gene. For each histone modification, the fold enrichment (log2) relative to
787 unmodified H4 is shown. **(D)** *Left:* Diagrams of the structure of the *C. albicans* LTR
788 and non-LTR retrotransposons. *Right:* Chromatin signature of LTR and non-LTR
789 retrotransposons. Average profiles and heatmaps of histone modification signatures
790 across each sequence. The relative fold enrichment (log2) for each histone
791 modification normalised to unmodified histone H4, or aligned reads of
792 immunoprecipitated (IP) sample normalised to aligned reads of Input sample (for
793 RNAPII ChIP-seq) is displayed. The gradient blue-to-red colour indicates high-to-low
794 enrichment in the corresponding region. Min: - 1.5 log2. Max: + 1.5 log2.
795 **Figure 5** The HDAC Sir2 controls the chromatin state of subtelomeres and the rDNA
796 locus **(A)** Schematic of the quantitative ChIP-seq experimental and analytical
797 workflow. **(B)** *Top*: Fold enrichment (log2) of H3K9Ac and H4K16Ac relative to
798 unmodified H4 in WT cells, and relative to WT in *sir2Δ/Δ* cells, across the rDNA loci
799 of chromosome R (ChrR). *Middle*: Diagram of transcripts found at this region,
800 according to assembly 22. *Bottom*: Heatmap depicting changes in gene and ncRNA
801 expression across the rDNA region in *sir2Δ/Δ* cells relative to WT. The gradient
802 yellow-to-blue colour indicates high-to-low expression. Min: - 2 log2. Max: + 2 log2.
803 **(C)** *Left*: Fold enrichment (log2) of H3K9Ac and H4K16Ac relative to unmodified H4
804 in WT cells, and relative to WT in *sir2Δ/Δ* cells, across the 20 kb left and right
805 terminal regions of chromosome 3 (Chr3). Diagrams of coding genes (TLO: grey)
806 found at these regions, according to assembly 22, are below. *Right*: Heatmap
807 depicting changes in gene and ncRNA expression in *sir2Δ/Δ* cells relative to WT at

808 the 10 kb terminal regions of all *C. albicans* chromosomes. The gradient yellow-to-
809 blue colour indicates high-to-low expression. Min: - 2 log2. Max: + 2 log2.

810 **Figure 6** Chromatin and gene expression changes of set1Δ/Δ strain **(A)** Heatmap
811 depicting changes in expression of genes and ncRNA associated with statistically
812 significant changes in H3K4me^3 enrichment in *set1Δ/Δ* cells relative to WT cells. The
813 gradient yellow-to-blue colour indicates high-to-low enrichment/expression. Min: - 4
814 log2. Max: + 4 log2 **(B)** *Top*: Fold enrichment (log2) of H3K4me^3 relative to
815 unmodified H4 in WT cells, and relative to WT in *set1Δ/Δ* cells, across the 20 kb left
816 and right terminal regions of chromosome 3 (Chr3). Diagrams of coding genes (TLO:
817 grey) found at these regions, according to assembly 22, are below. *Bottom*: Heatmap
818 depicting changes in gene and ncRNA expression in *set1Δ/Δ* cells relative to WT at
819 the 10 kb terminal regions of all *C. albicans* chromosomes. The gradient yellow-to-
820 blue colour indicates high-to-low expression. Min: - 2 log2. Max: + 2 log2. **(C)** *Left*:
821 Fold enrichment (log2) of H3K4me^3 relative to unmodified H4 in WT cells, and
822 relative to WT in *set1Δ/Δ* cells, across the rDNA loci of chromosome R (ChrR).
823 Diagrams of coding genes and ncRNA (grey) found at this region, according to
824 assembly 22, are below. *Right*: Heatmap depicting changes in gene and ncRNA
825 expression across the rDNA region in *set1Δ/Δ* cells relative to WT. The gradient
826 yellow-to-blue colour indicates high-to-low expression. Min: - 2 log2. Max: + 2 log2.
827 **(D)** *Left*: Profiles of fold enrichment (log2) of H3K4me^3 relative to unmodified H4 in
828 WT cells, and relative to WT in *set1Δ/Δ* cells averaged across the MRS repeats, and
829 the upstream and downstream sequences. The grey arrow indicates the location of
830 the FGR gene. *Right*: Heatmap depicting changes in gene and ncRNA expression
831 across all of the MRS regions in *set1Δ/Δ* cells relative to WT. The gradient yellow-to-
832 blue colour indicates high-to-low expression. Min: - 2 log2. Max: + 2 log2.

Fig 1

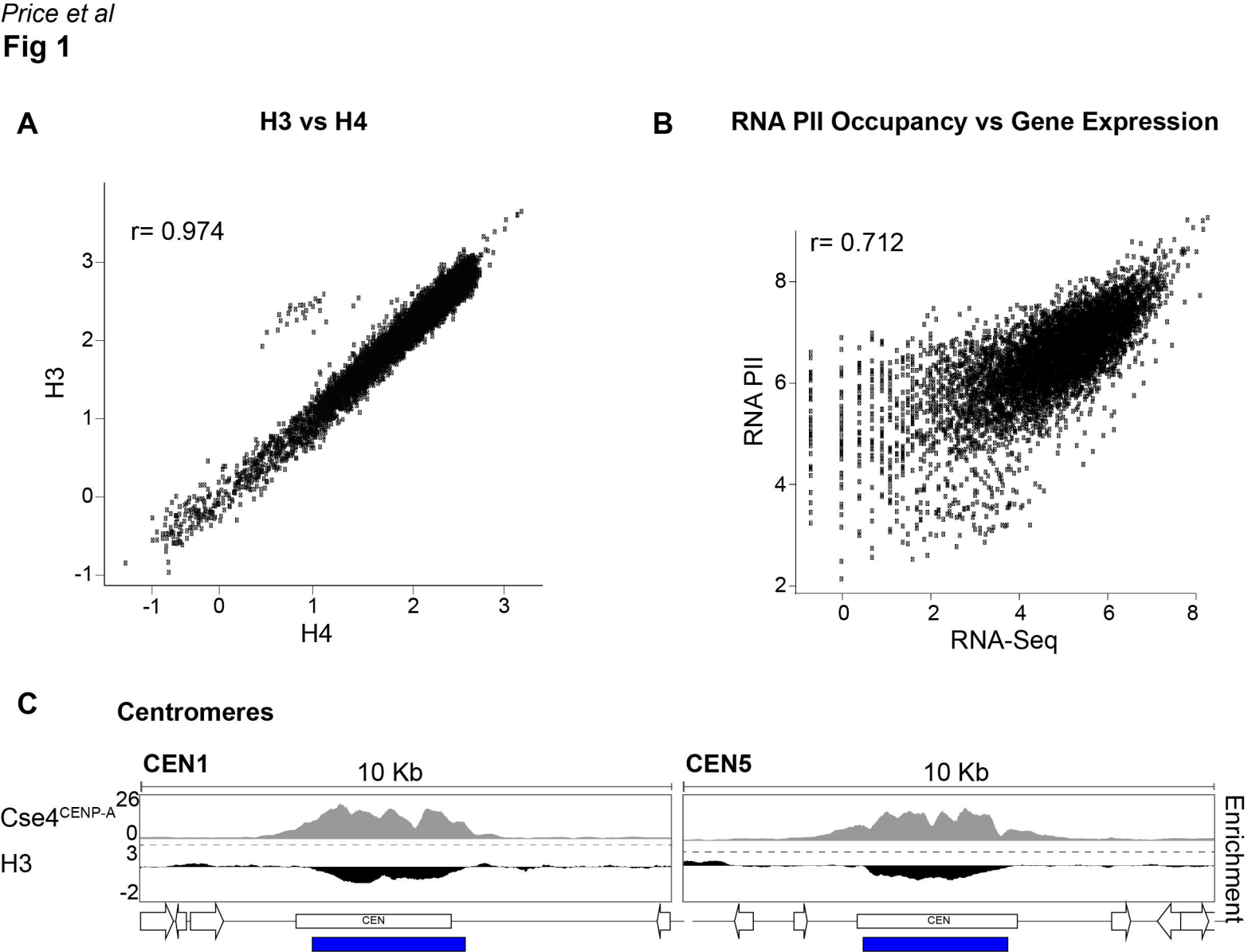

A
H3 vs H4
r= 0.974
H3
H4
B
RNA PII Occupancy vs Gene Expression
r= 0.712
RNA PII
RNA-Seq
C
Centromeres
CEN1
10 Kb
CEN5
10 Kb
Cse4CENP-A
H3
Enrichment
CEN
CEN

Fig 2

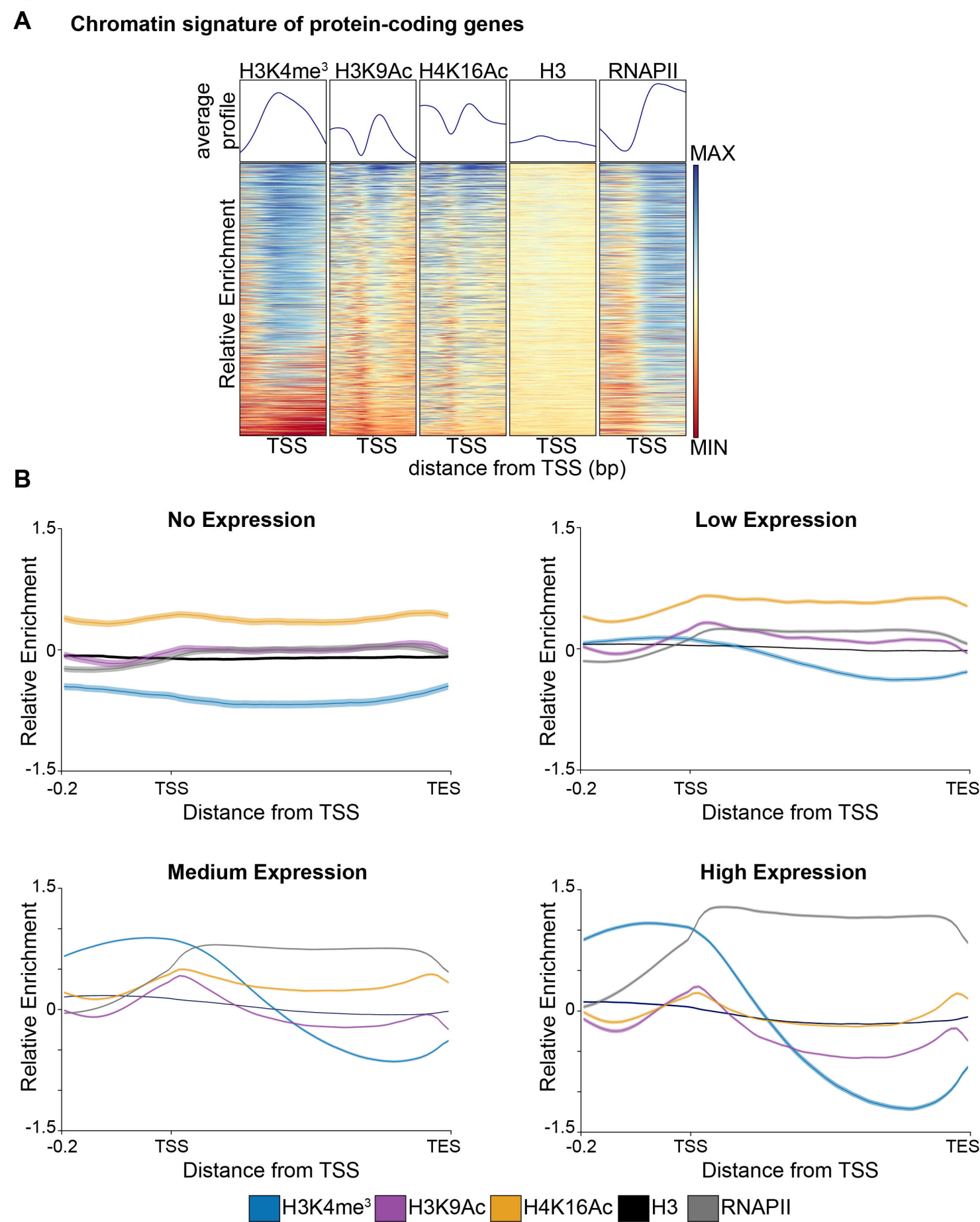
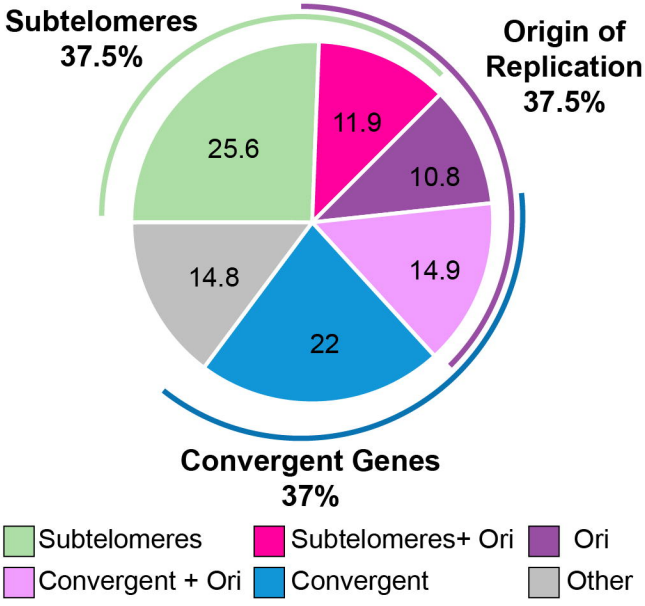

A
Chromatin signature of protein-coding genes
H3K4me3
H3K9Ac
H4K16Ac
H3
RNAPII
average profile
Relative Enrichment
MAX
MIN
TSS
TSS
TSS
TSS
TSS
distance from TSS (bp)
B
No Expression
Low Expression
Medium Expression
High Expression
Relative Enrichment
1.5
0
-1.5
-0.2
TSS
TES
Distance from TSS
H3K4me3
H3K9Ac
H4K16Ac
H3
RNAPII

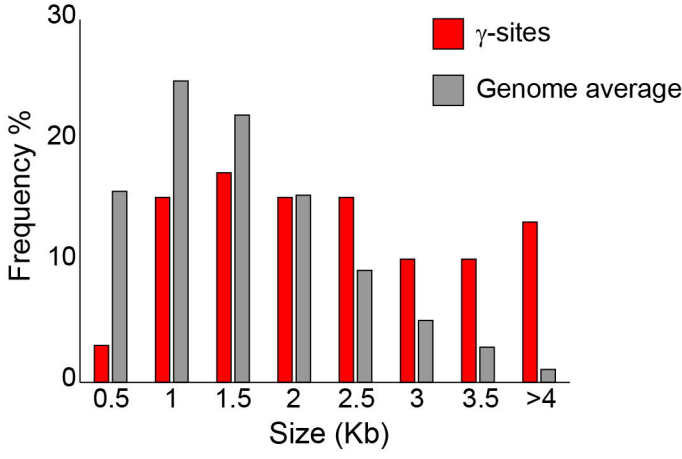
Fig 3

A

***C. albicans* γ-sites**



B



C

Chromatin signature of γ-sites

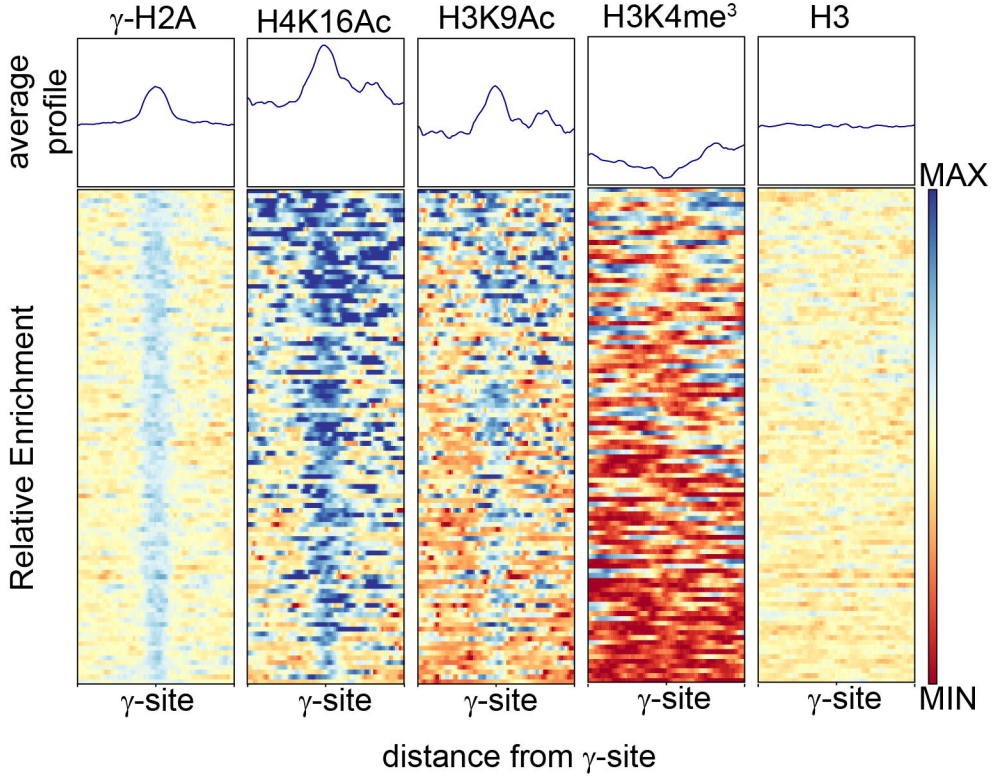
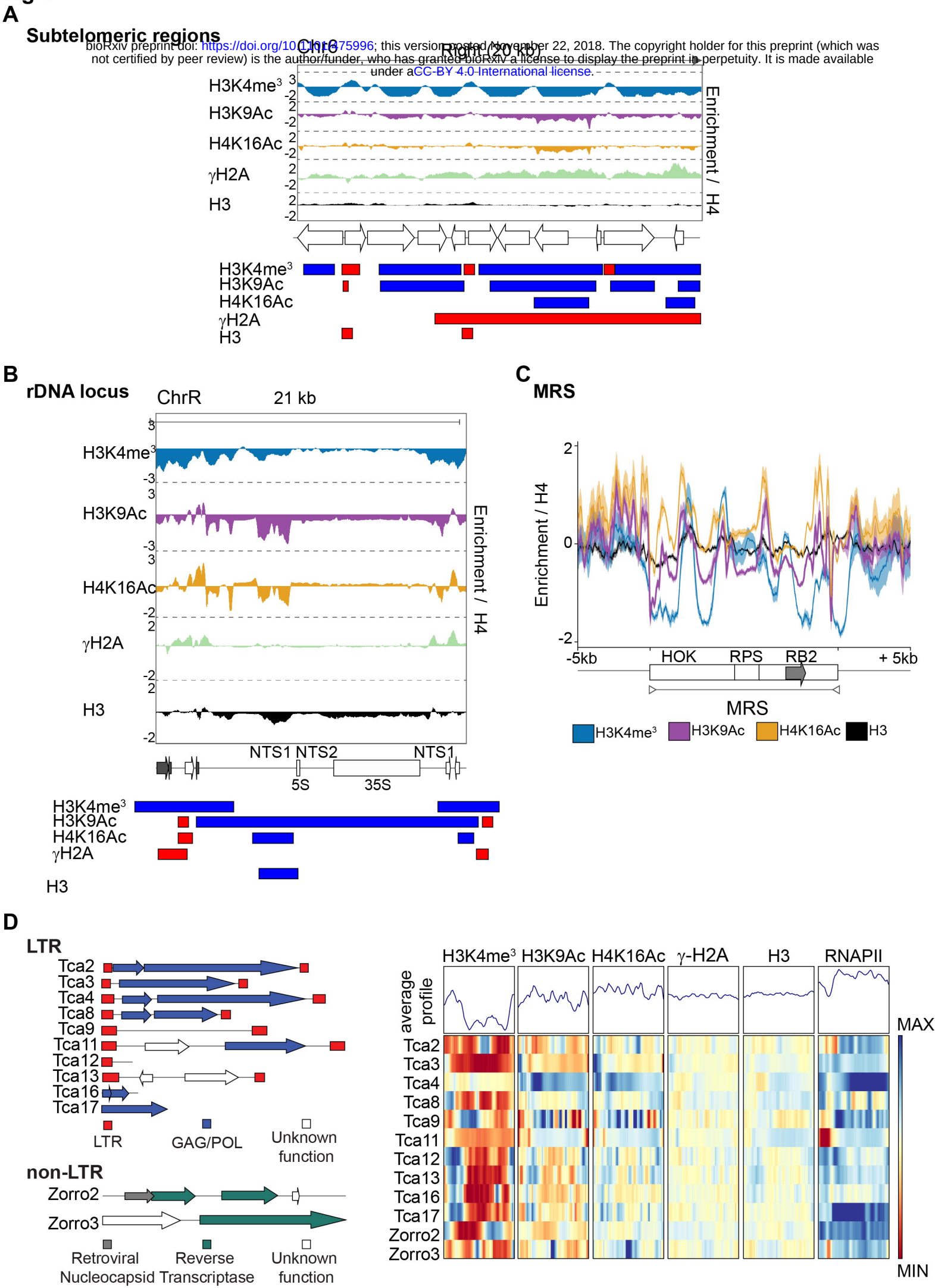


Fig 4



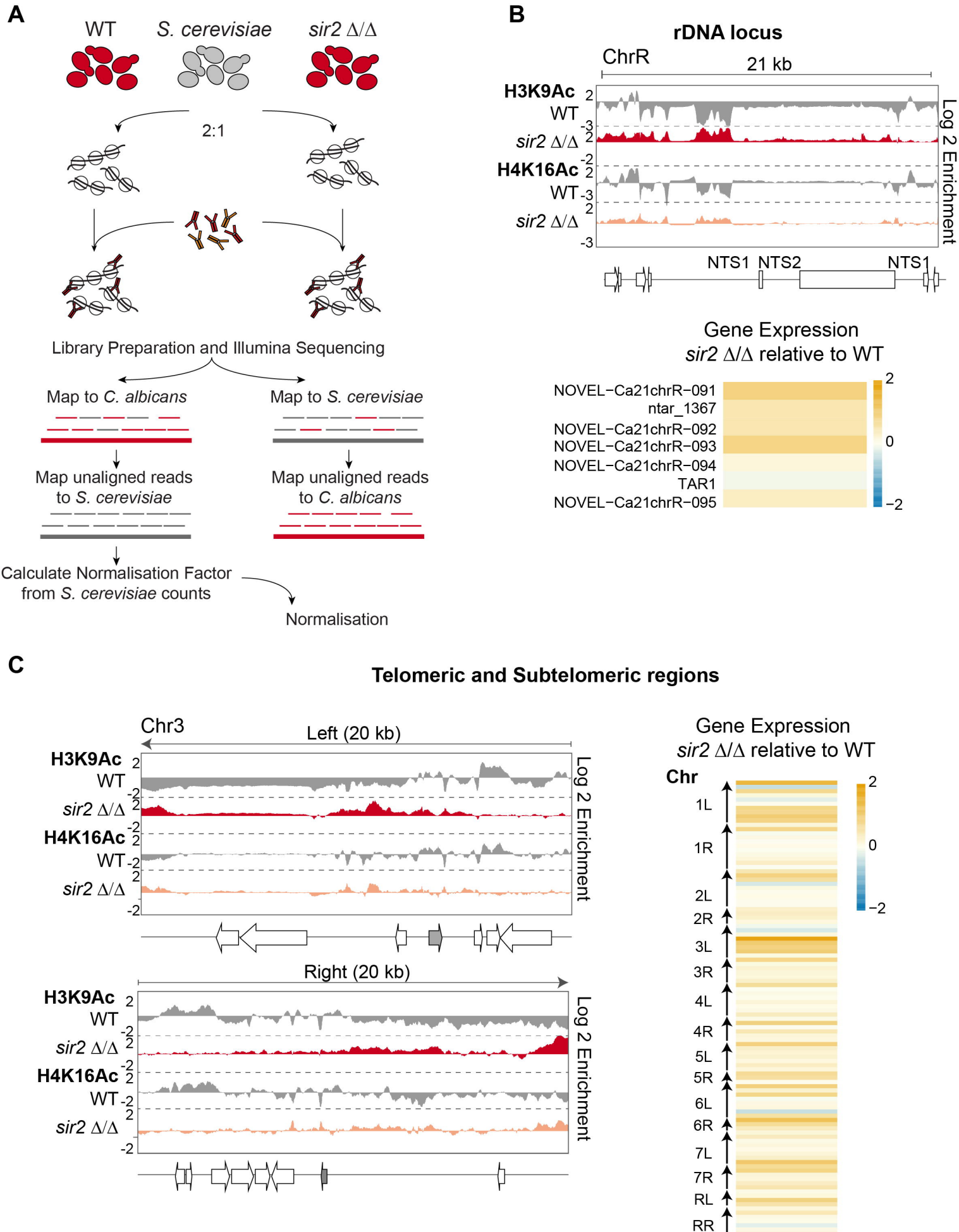


Fig 6