Experimental annotation of the human pathogen Candida albicans coding and noncoding transcribed regions using high-resolution tiling arrays.
Sellam, Adnane; Hogues, Hervé; Askew, Christopher; Tebbji, Faiza; van het Hoog, Marco; Lavoie, Hugo; Kumamoto, Carol A.; Whiteway, Malcolm; Nantel, André

Publisher's version / Version de l'éditeur:
https://doi.org/10.1186/gb-2010-11-7-r71
Genome biology, 11, R71, pp. 1-14, 2010-01-01
Experimental annotation of the human pathogen *Candida albicans* coding and noncoding transcribed regions using high-resolution tiling arrays

Adnane Sellam*1,2, Hervé Hogues1, Christopher Askew1,3, Faiza Tebbji1,3, Marco van het Hoog1, Hugo Lavoie4, Carol A Kumamoto5, Malcolm Whiteway1,3 and André Nantel*1,2

**Abstract**

**Background:** Compared to other model organisms and despite the clinical relevance of the pathogenic yeast *Candida albicans*, no comprehensive analysis has been done to provide experimental support of its *in silico*-based genome annotation.

**Results:** We have undertaken a genome-wide experimental annotation to accurately uncover the transcriptional landscape of the pathogenic yeast *C. albicans* using strand-specific high-density tiling arrays. RNAs were purified from cells growing under conditions relevant to *C. albicans* pathogenicity, including biofilm, lab-grown yeast and serum-induced hyphae, as well as cells isolated from the mouse caecum. This work provides a genome-wide experimental validation for a large number of predicted ORFs for which transcription had not been detected by other approaches. Additionally, we identified more than 2,000 novel transcriptional segments, including new ORFs and exons, non-coding RNAs (ncRNAs) as well as convincing cases of antisense gene transcription. We also characterized the 5' and 3' UTRs of expressed ORFs, and established that genes with long 5' UTRs are significantly enriched in regulatory functions controlling filamentous growth. Furthermore, we found that genomic regions adjacent to telomeres harbor a cluster of expressed ncRNAs. To validate and confirm new ncRNA candidates, we adapted an iterative strategy combining both genome-wide occupancy of the different subunits of RNA polymerases I, II and III and expression data. This comprehensive approach allowed the identification of different families of ncRNAs.

**Conclusions:** In summary, we provide a comprehensive expression atlas that covers relevant *C. albicans* pathogenic developmental stages in addition to the discovery of new ORF and non-coding genetic elements.

**Background**

*Candida albicans* is an opportunistic pathogen responsible for various non-life-threatening infections, such as oral thrush and vaginitis, and accounts for more than half of all *Candida* infections [1,2]. This pathogen is also a major cause of morbidity and mortality in bloodstream infections, especially in immunosuppressed individuals. *C. albicans* can also colonize various biomaterials, such as urinary and vascular catheters, and ventricular assist devices, and readily forms dense biofilms that are resistant to most antifungal drugs [3]. The ability of this fungus to switch from yeast to filamentous forms (true hyphae or pseudohyphae) is also a crucial determinant for host invasion and thus virulence [4]. Because of the challenges of drug resistance [5-7] and the eukaryotic nature of *C. albicans*, which makes it similar to its human host, extensive efforts are being made to identify specific new drug targets for therapeutic intervention.

The *C. albicans* genome has been the subject of many curated annotations that have resulted in the current comprehensive physical genomic map [8-11]. Recently, the genome sequences of six further species from the
Candida clade have been released. Comparative analysis of these genomes revealed a significant expansion of gene families associated with virulence compared to non-pathogenic yeasts [12]. In addition, this work uncovered an unexpected divergence in the mechanisms controlling mating and meiosis in this clade. Given the high conservation of protein-coding sequence within the six Candida species, Butler et al. [12] undertook a comparative annotation to revise the genome sequence of C. albicans and identified 91 new or updated ORFs.

Genome sequencing followed by in silico-based annotation is the critical first step required to gain a comprehensive insight into the genetic features underlying different aspects of an organism's biology. To establish a more comprehensive and accurate layout of these features, in silico methods must be complemented by transcriptome or proteome investigations. Recent advances taking advantage of the high-throughput potential of whole-genome tiling microarrays or cDNA sequencing contributed significantly to the discovery of novel sites of active transcription missed by computational gene prediction (reviewed in [13-15]). Tiling array technology has revealed several unexpected hidden features of the eukaryotic transcriptome, including antisense (AS) transcription, non-coding RNAs (ncRNAs) as well as complex transcriptional architectures such as nested genes [16-22]. The use of tiling arrays has also been useful for mapping a variety of epigenetic marks in eukaryotes and uncovering the complex network of mechanisms involved in transcriptional regulation associated with chromatin dynamics [23-25]. Here we have undertaken a genome-wide experimental annotation using a strand-specific high-density tiling array that allows us to accurately uncover the transcriptional landscape of C. albicans. The main purposes of this work were: the experimental validation of computational-based genome annotations in C. albicans; the discovery of new coding and non-coding genetic elements for future studies; the identification of new functional features associated with the transcriptome organization; and the annotation of class I, II and III genes using an unbiased methodology that combines data from the genome-wide occupancy of different subunits of RNA polymerases (RNAPs) I, II and III with data from transcriptome studies.

Results and discussion
To illuminate the transcriptional landscape of the pathogenic fungus C. albicans, we tiled both Watson and Crick strands of the whole genome with 240,798 60-mer probes each overlapping by 1 bp. Total RNA was purified from cells growing under various conditions relevant to C. albicans pathogenicity; specifically growing as a biofilm, as hyphae and as a commensal within the mouse caecum. RNA from cells growing as yeast in YPD at 30°C were used as a reference for each condition.

Transcript mapping reveals extensive transcription in C. albicans
For each condition, thresholds were determined empirically based on the 95th percentile of signal intensities of non-conserved intergenic regions as described in the Materials and methods section. After combining expression data for all the tested conditions, transcription activity was detected for 72% of the 6,193 nuclear genes, including 4,402 ORFs, 4 pseudogenes, 67 tRNAs, 108 retrotransposons and 7 ncRNAs (5 small nuclear RNAs (snRNAs), 1 small nucleolar RNA (snoRNA) and the rRNA) (Table 1). The remaining 28% of the genomic features not detected in this study could be due to the fact that they are not used in our conditions, and an analysis of Gene Ontology (GO) functional categories of these unexpressed genes revealed a significant enrichment in functions related to the accomplishment of the parasexual cycle in C. albicans, including ascospore wall assembly ($P = 1.74e-05$), meiosis ($P = 1.33e-02$) and synapsis ($P = 8.64e-04$) (Additional file 1).

A large number of transcribed segments, or transfrags [26], were detected in intergenic regions devoid of existing annotation. Transfrags were identified on the basis of two or more consecutive probes exhibiting intensities above the threshold, together with separation by at least 120 bp from any currently annotated elements. Using

| Features                  | CGD  | This study | Coverage |
|---------------------------|------|------------|----------|
| ORF                       | 6,197| 4,402      | 71%      |
| Retrotransposon and LTR   | 129  | 108        | 83%      |
| Pseudogene                | 8    | 5          | 62.5%    |
| tRNA                      | 156  | 67         | 43%      |
| snRNA                     | 5    | 5          | 100%     |
| snoRNA                    | 1    | 1          | 100%     |

*CGD, Candida Genome Database; LTR, long terminal repeat.*
these criteria, a total of 2,172 transfrags were detected and mapped (Additional file 2). Interestingly, 31% of the intergenic transcribed units (680 transfrags) display significant sequence conservation (e-value < 10^{-10}) with Candida dubliniensis, suggesting the existence of functional genetic elements.

**Features of transcribed regions in the C. albicans genome**

As shown in Figure 1, a clear correlation can be seen between the annotated ORFs and the signal intensities of probes. In general, the obtained data are in agreement with the current *Candida* Genome Database (CGD) annotation [27]. At the gene level, our data allowed us to confirm the presence of introns in a number of ORFs, as shown for INO4 (ORF19.837.1) and EFB1 (ORF19.3838) (Figure 2b, f). Although the resolution of our tiling array shown for INO4 was not high enough to delimit precisely intron boundaries, we were able to confirm the introns previously annotated in the *C. albicans* genome [28]. Moreover, extensions of transcripts corresponding to potential upstream ORFs (for example, CLN3; Figure 2g) or 5' and 3' UTRs (for example, ZCF37; Figure 2h) were identified in several locations. Genetic elements displaying complex transcriptional architectures, such as nested genes (TLO34 and ORF9.2662; Figure 2a; Additional file 3) or intronic nested genes (snR18 hosted by the EFB1 intron; Figure 2f), were identified. Additionally, a large number of sense-AS transcript pairs have been detected (PFK1 and EFB1; Figure 2d, f). Intriguingly, in some cases, AS transcription was found on the opposite strand rather than the annotated strands (CHR12 and CCW14; Figure 2d, e). Previously unannotated ORFs and ncRNAs were also uncovered (ORF19.6853.1 and snR18; Figure 2c, f). To illustrate the annotation concept, some of the most relevant *C. albicans* genome features will be highlighted throughout the manuscript.

**Revisiting the C. albicans ORFeome**

Based on the last CGD update (24 December 2009), the existing ORF catalogue of *C. albicans* consists of a total of 6,197 ORFs, of which 1,084 were experimentally verified, 4,933 functionally uncharacterized and 180 considered as dubious. In our current analysis, we have been able to detect the expression of 4,588 ORFs. Compared to other model organisms and despite the clinical relevance of the pathogenic yeast *C. albicans*, no comprehensive analysis has been done to provide experimental support to the *in silico*-based annotation. Our study thus provides such a genome-wide experimental validation for a large number of predicted ORFs for which transcription had not yet been confirmed by other approaches. Recently, using a comparative annotation approach, Butler et al. [12] identified 91 new ORFs, of which 80% are specific to the *Candida* clade. In the present study, 52% of those new ORFs (48 ORFs) were expressed above the background in our conditions, thus validating their functionality (Additional file 4). Furthermore, our data raised questions about 34 ORFs previously annotated as spurious or dubious [8] (Additional file 4). We also annotated 11 ORFs when screening the 2,172 expressed intergenic segments for their protein-coding potential (Additional file 4).

**Characterization of UTR regions**

UTRs are known to play key roles in the post-transcriptional regulation of gene expression, influencing mRNA transport, mRNA subcellular localization, and RNA turnover [29]. Therefore, annotation of *C. albicans* UTRs has the potential to provide important insights into gene regulatory mechanisms underlying the biology and the pathogenicity of this fungus. To define *C. albicans* UTRs, we scanned the expression maps under different conditions, identified unannotated segments exhibiting an unbroken signal intensity connected to nuclear-encoded genes. A total of 481 5' UTRs and 846 3' UTRs longer than 240 bp were identified (Additional file 5). Compared to *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* [16,18,30], where the 3' UTRs are longer than 5' UTRs, the median length of both 5' and 3' UTRs was almost the same (the mean length of 5' and 3' UTRs was
88 bp and 84 bp, respectively, with a range of 0 to 3 kb for both 5' and 3' UTRs.

Genes with long 5' UTRs (>330 bp) were significantly enriched in regulatory functions, including transcription and signal transduction (Table 2; Additional file 6). A similar result was observed in S. pombe for both functions [31], and in S. cerevisiae for signal transduction [16]. In many eukaryotes, including the fission yeast S. pombe, it is well known that the most stable transcripts have short 5' UTRs, while the least stable transcripts have both long 5' and 3' UTRs [32,33].

Intriguingly, a large number of transcripts with long 5' UTRs are key regulators of filamentous growth in C. albicans, including the transcription factors EFG1, RFG1, CPH1, CPH2, CZF1, CRZ1, CRZ2, SSN6, NRG1 and FCR1, and the phosphatases YVH1, PTC8 and CPP1 (Additional file 6). The regulation of RNA stability is a critical issue in modulating gene expression, in particular for transiently expressed regulatory genes such as those encoding transcription factors and phosphatases. Therefore, fine-tuning RNA turnover rates for those transcripts is potentially a key regulatory process involved in control

**Table 2: Gene Ontology analysis of genes with long 5' UTR regions (>330 bp)**

| GO terms                      | P-value       | Median UTR length (bp) |
|-------------------------------|---------------|------------------------|
| DNA binding                   | 2.36e-05      | 540                    |
| Transcription factor activity | 3.68e-05      | 540                    |
| Phosphoprotein phosphatase activity | 1.7e-04    | 420                    |
| Hyphal growth                 | 2.60e-05      | 450                    |
| Filamentous growth            | 1.30e-10      | 480                    |
| Growth                        | 1.54e-10      | 480                    |
| Regulation of biological process | 4.70e-11    | 480                    |
| Cellular bud neck            | 3.6e-04       | 390                    |
of the yeast-to-hyphae transition in C. albicans. A high rate of RNA decay of transcripts involved in regulatory systems has been reported in S. cerevisiae as well [34]. Intriguingly, of the 38 RNAs identified recently as She3-transported in C. albicans during hyphal growth [35], 9 were found to exhibit long 5’ UTRs (P = 4.3e-04). This leads us to speculate that long 5’ UTRs are probably required for RNA transport to cellular locations where hyphal buds are produced.

Widespread occurrence of antisense transcription in C. albicans

Large-scale transcript mapping studies revealed the common occurrence of overlapping cis-natural AS transcripts in different model organisms [16-19,36]. In a recent study, Perocchi et al. [37] have shown that about half of all annotated antisense (AS) transcripts detected by tiling arrays in S. cerevisiae were experimental artifacts related to spurious synthesis of second-strand cDNAs that occurred during reverse transcription (RT) [37,38]. These authors showed that these RT artifacts were efficiently resolved by using the transcription inhibitor actinomycin D. In light of their finding, we have used actinomycin D to prevent the appearance of these artifacts. Indeed, as shown in Figure 3a, b, the use of actinomycin D reduced, in part, the dependence of AS signal intensity on the sense expression level.

AS transcription was observed for 724 genes, of which 623 are ORFs, 16 ncRNAs and 85 retrotransposons (Table S5 in Additional file 7). With few exceptions, all C. albicans AS transcripts belong to the completely overlapping natural AS transcript category. Based on sense/AS signal intensity ratio, AS transcripts were separated into two classes as was described for S. cerevisiae [37]. In the first class of AS transcripts, the hybridization signal intensity of the annotated features is higher and proportional to its AS counterpart (Figure 3a, b). This class contains the majority (79%) of the detected AS transcripts. Genes with this pattern are highly expressed in all conditions and GO analysis showed a preferential enrichment in housekeeping functions, including translation (P = 1.11e-38), cell surface proteins (P = 1.63e-13), glycolysis (P = 1.18e-12) and nucleosomes (P = 5.27e-08) (Figure 3c). Similar findings have been reported by experimental-based annotation of AS transcripts in wheat [39], rice [40] and S. cerevisiae [16], as well as by in silico approaches in other model organisms [41].

The second class of AS transcripts, where the average activity for the AS strand was much higher than the sense strand, contains only 37 genes (Figure 2d, e; Table S5 in Additional file 7). Strand-specific RT-PCR validated the expression of eight of these genes at the AS strand (Figure 3d). No functional enrichment was obtained for those transcripts. However, this AS category includes the transcription factor gene encoding the ortholog of S. cerevisiae Kar4p that plays a critical role in karyogamy during the mating process [42,43]. Overexpression of KAR4 in S. cerevisiae during vegetative growth causes a severe growth defect as a consequence of accumulation of cells arrested at G1 and G2/M stages [44]. Thus, if Kar4p plays a similar role in C. albicans, the AS transcription at this locus might be required for repression of the sense transcript during vegetative growth. A similar scenario was
reported in *S. cerevisiae* where AS transcription opposite to *IME4* has been shown to play a critical role in controlling entry into meiosis [45].

**RNAP-guided annotation of new C. albicans ncRNAs**

Ongoing investigations on the function of ncRNAs established their specific roles in processes that require highly specific nucleic acid recognition without complex catalysis, such as guiding rRNA or tRNA covalent modifications [46,47] or guiding chromatin-modifying complexes to specific locations within the nucleus [48]. Given the central role of ncRNAs in such crucial biological processes, their genomic annotation is of great importance. However, annotating ncRNAs is a non-trivial task since their primary sequences are poorly conserved even between evolutionarily similar organisms. Here we adapted a strategy in which genome-wide occupancy of different subunits of RNAPs I, II and III is combined with expression data to annotate ncRNAs resulting from real transcriptional events. For this purpose we have performed chromatin immunoprecipitation on chip (ChIP-chip) of subunits that represent the three RNAP machines in *C. albicans* cells growing in rich media (YPD) at 30°C.

**RNAP I-associated ncRNAs**

RNAP I targets were determined by mapping the genomic location of the largest RNAP I subunit, Rpa190p (ORF19.1839). The results obtained show that Rpa190p occupancy was restricted to the rDNA locus where it binds the 18 S, the 5.8 S and the 28 S precursor gene promoters as well as internal transcribed regions (Additional file 8).

**RNAP II-associated ncRNA**

In *S. cerevisiae* RNAP II occupancy was evaluated by performing ChIP-chip of the two subunits Rpo21p (ORF19.7655) and Rpb3p (ORF19.1248). Among the CGD-annotated ncRNAs, the snRNAs U1, U2, U4 and U5, associated with the spliceosomal machinery, were found to fit the established criteria. When Rpo21p and Rpb3 binding sites were matched to the 2,161 non-coding intergenic transfrags, 425 actively transcribed putative ncRNAs were found. A search of these 425 transfrags using the *S. cerevisiae* ncRNA database returned only four matches that corresponded to snoRNAs. To generate an exhaustive list of *C. albicans* snoRNAs among the 2,161 ncRNA candidates, Snoscan [49] and snoGPS [50] servers were used to detect both C/D and H/ACA box snoRNA families, respectively. A total of 27 C/D box and 35 H/ACA box snoRNA candidates were identified. Most of the detected snoRNAs possess a canonical secondary structure and conserved C, D, A and ACA consensus motifs (Table S6 in Additional file 7). A comparison of these snoRNAs with entries in the Rfam database [51] returned 18 hits (4 H/ACA box and 14 C/D box) that match significantly to *S. cerevisiae* characterized snoRNAs. Orthologs of *S. cerevisiae* essential snoRNAs required for the cleavage of rRNA transcripts, namely U3a (snR17a), U3b (snR17b), U14 (snR128) and the snoRNA MRP NME1, were also detected and annotated in this study (Table S6 in Additional file 7). Interestingly, our results show that the U5 spliceosomal RNA (SNRNAU5) exhibits an extended transcriptional activity beyond its 3’ terminal end, suggesting that *C. albicans*, like *S. cerevisiae*, possesses a long form of SNRNAU5 (U5L). Using 3’ rapid amplification of cDNA ends (RACE), Mitrovich and Guthrie [52] have shown that, in addition to the vast majority of products that correspond to the short form of SNRNAU5 (U5S), a small amount of the long form was detected. In accordance with this, we found that the U5L transfrag was weakly transcribed compared to the U5 S. We also detected the previously characterized but unmapped *C. albicans* telomerase ncRNA TER1 [53] (Table S6 in Additional file 7). A total of 35 putative non-coding transfrags were randomly selected and their expression was confirmed using quantitative PCR (qPCR; Table S7 in Additional file 7). No obvious functions were attributed to the remaining 361 putative ncRNAs. Many large-scale gene expression mapping studies in mammals have suggested widespread transcription in intergenic regions that represent 47% to 80% of the transcribed features [54]. This ‘dark matter’ transcription has been accredited to previously undetected non-coding genes, ‘junk’ transcription, or experimental artifacts (reviewed in [15,55]). A recent report has demonstrated that the number and abundance of intergenic transcribed fragments from a large variety of different human and mouse tissue types were lower than shown earlier [54]. Using RNA-seq, van Bakel *et al.* [54] showed clearly that a significant number of these transcripts are associated with known genes and include many previously unidentified exons and alternative promoters. Though the majority of the ‘dark matter’ transcription seems to be artifactual, many conserved and presumably functional intergenic transcribed fragments remain to be characterized. In our work, many transfrags are conserved and expressed reproducibly in different conditions, suggesting a potential for a function and making them priority candidates for genetic perturbation and phenotypic characterization.

Additionally, to gain an insight into the function of these ncRNAs and their transcriptional regulation, we mapped the location of different transcription factors described in the literature for which genomic occupancies were determined using ChIP-chip. With the exception of Tfb1p, a master regulator of ribosomal protein expression in *C. albicans* [56,57], no transcription factors have been found associated with the promoter sequences of putative ncRNAs. Remarkably, in addition the occupancy of ribosomal protein genes and rRNA *cis-regula-
tory regions, Tbf1p was found to be associated with the promoter of six snoRNAs annotated in this work. This finding implies that Tbf1p coordinates transcriptional activation of both structural components of the ribosome (rRNA and ribosomal protein genes) [56] in addition to the snoRNAs that guide methylation and pseudouridylylation modifications required for ribosome maturation and functionality. Recently, Preti et al. [58] showed that Tbf1p in S. cerevisiae is required for the activation of snoRNA, implying a similar role in C. albicans. Similar findings were also obtained in the plant model Arabidopsis thaliana where the Tbf1p motif (ACCCCTA) was significantly enriched in upstream snoRNAs (P = 4.64e-20), suggesting a highly conserved role for this factor.

**RNAP III-associated ncRNAs**

In eukaryotic cells, RNAP III transcribes genes encoding tRNAs, 5 S rRNA and other ncRNAs, such as the RNA component of RNase P (RPR1) and the U6 snRNA (SNR6) [59-61]. To investigate the targets of the RNAP III machinery in C. albicans, we performed ChIP-chip with the subunit Rcp82p (ORF19.2847). Based simply on signal intensities of the ChIP-chip, Rcp82p targets can be divided in two categories. The first category includes loci with a high level of occupancy (between 6- and 45-fold enrichment): this category contains 120 tRNAs and the 5 S rRNA (Table S8 in Additional file 9) alongside the well-known non-tRNA genes transcribed by RNAP III (RPR1, SNR6, snR52, SCR1), which were characterized [62,63] but not mapped (Additional file 10). For all these binding events significant transcriptional hybridization signals were detected at least in two different conditions for 67 tRNAs, RPR1, SNR6, snR52, SCR1 and the 5 S rRNA. The second category includes loci with a low level of occupancy (between 2- and 4.5-fold enrichment): with a few exceptions, all these loci were expressed and correspond to repetitive DNA elements associated with retrotransposons. Since long terminal repeat (LTR) retrotransposons are present in the C. albicans genome in multiple copies (Figure 5), it has been demonstrated that yeasts S. cerevisiae LTR retrotransposons Ty1 and Ty3 strictly target regions in the vicinity of tRNAs [64,65]. This conserved strategy is most likely adopted to avoid deleterious integrations into coding sequences. In the social amoeba Dictyostelium discoideum, Siol et al. [66] have demonstrated that the general transcription factor TFIIIC of the RNAP III machinery is actively required for targeted integration of the retrotransposon TRE5-A [66]. This finding supports that, in our study, some Rcp82p-retrotransposon-occupied loci might be real binding events. Indeed, based on binding intensity, it is probably the case for two loci where Rcp82p was found to bind the repetitive DNA elements beta-1a and beta-1c of the retrotransposon Tca8 with an occupancy level similar to that seen for tRNAs (Table S8 in Additional file 9).

**Subtelomeric regions are transcriptionally active and express a cluster of ncRNAs**

We found that clustered transcribed segments (52 transfrags) with no protein-coding potential were located at the subtelomeric regions of all chromosomes (Figure 4a). This finding is in accordance with early work in mammals that established that telomeres, originally thought to be transcriptionally silent, bore actively transcribed ncRNAs [67,68]. Based on sequence similarity, these telomere-associated ncRNAs (TelRs) can be divided into eight classes (TelR A to H; Figure 4; Table S9 in Additional file 9). With no exception, all TelRs from class A are AS of TLO genes, overlapping with their 5' ends. The class B TelRs correspond to the telomeric element CARE-2 [69], which is composed, in part, of the LTR retrotransposon. TelRs are specific to C. albicans and their sequences are not conserved throughout the clades represented in the CTG. Furthermore, when TelR sequences of the SC5314 strain were compared to their counterparts in the WO1 strain, we noticed a significant degree of polymorphism. Subtelomeric regions are suggested to be potential locations of gene amplification since one telomere might be functionally exchanged with another [70]. Thus, in addition to TLO genes, TelRNAs seem to be members of a new family of multi-copy subtelomeric ncRNAs.

**Differentially regulated transfrags during pathogenic-related growth**

As an opportunistic fungus, C. albicans must activate numerous transcriptional outputs to promote host colonization or virulence [71]. To elucidate the transcriptional patterns of annotated features in the different tested conditions, signal intensities of transfrags detected in cells growing as hyphae, biofilms and in the mouse caecum were compared to their counterparts in yeast cells (the control condition). GO analysis was used to assess the average expression levels of genes encoding specific classes of proteins in the three tested conditions (Figure 5; Additional files 11 and 12). In general, our results demonstrated a large overlap in transcripts present in hyphae or biofilms that were found in other studies. For instance, many differentially expressed genes in the three tested conditions encode adhesins and fungal cell wall proteins, consistent with their described roles during the interaction with the host and biofilm formation [71-73]. Unexpectedly, classes of genes involved in ncRNA metabolic processes, such as small nucleolar ribonucleoprotein (snoRNP) assembly complexes, were found differentially expressed in hyphae and in cells recovered from the caecum (Figure 5). Similarly, several genes that had never
been detected before in *C. albicans* biofilms, including genes encoding tRNAs (GO term 'translation elongation'; \( P = 1.57 \times 10^{-59} \)), were found to be significantly consistently repressed with the repression of ribosomal genes, as reported in other biofilm models [74,75].

Interestingly, we found that genes encoding proteins involved in heme binding were actively transcribed in *C. albicans* cells recovered from the caecum (Figure 5a), suggesting that the caecum is an iron-poor niche. These genes include hemoglobin-receptors *RBT5, PGA10*,...
CSAI, and DAPI, as well as the heme-degradation oxygenase HMX1. During this commensal growth, C. albicans also activates genes related to carbohydrate catabolism, as was reported in other in vivo infection models [71]. qPCR confirmed the activation of selected genes representing carbohydrate catabolism and heme binding functions in two independent biological replicates (Additional file 13).

To discover candidate ncRNAs potentially associated with host-dependant growth, we defined differentially expressed intergenic transfrags in C. albicans cells growing in the caecum as well as in cells undergoing hyphal and biofilm growth. Using a stringent cutoff (see Materials and methods), 264, 47, and 64 transfrags were found differentially regulated in caecum-grown cells, hyphae and biofilm cells, respectively (Additional file 14). Many of them are bound by the RNAP II or are conserved with other species from the Candida clade (Additional file 14), suggesting a significant potential for function.

Conclusions
We provide a comprehensive expression map that covers a set of conditions relevant to C. albicans pathogenic developmental stages. The identification of unannotated transcribed regions was the main motivation of this study. Using multiple genome-scale measurements (expression profiling and RNAP occupancy), we have characterized and annotated a number of ncRNAs hidden in the ‘dark matter’ of the C. albicans genome. These ncRNAs candidates constitute an interesting framework for future functional studies and will contribute to our understanding of the role of the C. albicans non-coding genome. Furthermore, our work has uncovered different genetic features, including extensive AS transcription, 5’ and 3’ UTRs and expression at subtelomeric regions. One particular feature was the enrichment of genes with long 5’ UTRs in regulatory function associated with hyphal development. This feature might imply noteworthy regulation at the post-transcriptional level of the C. albicans yeast-to-hyphae switch and should be clarified in the near future. Transcript mapping data and RNAP occupancies will be available at the CGD database [76] displayed via a genome browser interface (Gbrowse), enabling the inspection of any locus of interest.

Materials and methods
Growth media and conditions
Strains used in this study are listed in Additional file 15. For general propagation and maintenance conditions, the strains were cultured at 30°C in yeast-peptone-dextrose (YPD) medium supplemented with uridine (2% Bacto peptone, 1% yeast extract, 2% dextrose, and 50 μg/ml uridine, with the addition of 2% agar for solid medium). Cell growth, transformation and DNA preparation were carried out using standard yeast procedures.

For gene expression profiling of yeast-form cells, saturated overnight cultures of the SC5314 strain were diluted to a starting OD₆₀₀ of 0.1 in 50 ml fresh YPD and grown at 30°C to an OD₆₀₀ of 0.8. Hyphae were induced by growing Candida cells in YPD plus 10% fetal bovine serum at 37°C to an OD₆₀₀ of 0.8. Cultures were harvested by centrifugation at 3,000 x g for 5 minutes, and the pellet rapidly frozen in liquid nitrogen. Biofilms were grown in
RPMI medium at 37°C as described [77]. For RNA extracted from caecum-grown cells, female C57BL/6 mice (5 to 7 weeks old) were treated with tetracycline (1 mg/ml), streptomycin (2 mg/ml) and gentamicin (0.1 mg/ml) added to their drinking water for the duration of the experiment, beginning 4 days prior to inoculation. C. albicans cells (5 × 10⁷ cells) were orally inoculated into the mice by gavage. Three days post-inoculation, the mice were sacrificed and the contents of the caecum were recovered and frozen in RNALater (Ambion, Austin, TX, USA) at -80°C. Caecum contents were filtered through 500 μm polypropylene mesh (Small Parts, Inc., Miramar, FL, USA) to remove large particles and RNA was extracted by bead beating with 0.5 mm zirconia/silica beads in TRIzol (Invitrogen, Carlsbad, CA, USA). After the TRIzol RNA purification procedure described by the manufacturer, RNA was further purified on Qiagen RNeasy protocol was followed as recommended.

To extract RNA from cells, samples stored at -80°C were thawed in the buffer with vortexing briefly at high speed. The resuspended pellet was placed back on ice and divided into 1 ml aliquots in 2 ml screw cap microcentrifuge tubes containing 0.6 ml of 3 mm diameter acid-washed glass beads. Samples were homogenized 5 times, 1 minute each, at 4,200 RPM using Beadbeater. Samples were placed on ice for 1 minute after each homogenization step. After the homogenization the Qia-gen RNeasy protocol was followed as recommended. Total RNA samples were eluted in RNAse free H₂O. RNA quality and integrity were assessed using an Agilent Technologies eArray service.

Microarray experiments
To extract RNA from cells, samples stored at -80°C were placed on ice and RNAeasy buffer RLT was added to pellets at a ratio of 10:1 (vol/vol) buffer/pellet. The pellet was allowed to thaw in the buffer with vortexing briefly at high speed. The resuspended pellet was placed back on ice and divided into 1 ml aliquots in 2 ml screw cap microcentrifuge tubes containing 0.6 ml of 3 mm diameter acid-washed glass beads. Samples were homogenized 5 times, 1 minute each, at 4,200 RPM using Beadbeater. Samples were placed on ice for 1 minute after each homogenization step. After the homogenization the Qia-gen RNeasy protocol was followed as recommended. Total RNA samples were eluted in RNAse free H₂O. RNA quality and integrity were assessed using an Agilent 2100 bioanalyzer.

cDNA labeling and microarray production were performed as described [79]. Briefly, 20 μg of total RNA was reverse transcribed using 9 ng of oligo(dT)₂₁ and 15 ng of random octamers (Invitrogen) in the presence of Cy3 or Cy5-dCTP (Invitrogen) and 400 U of Superscript III reverse transcriptase (Invitrogen). Actinomycin D was used to inhibit synthesis of the second cDNA strand to a final concentration of 6 μg/ml. To assess actinomycin D efficiency in resolving spurious AS transcripts, signal intensities of annotated feature (from yeast and hyphae experiments) probes exhibiting an AS transcript expressed above the background were considered. The signals of every probe representing either sense or AS transcripts for each hybridization, performed with or without actinomycin D, were plotted (Figure 3a, b).

After cDNA synthesis, template RNA was degraded by adding 2.5 units RNase H (Promega, Madison, WI, USA) and 1 μg RNase A (Pharmacia, Uppsala, Sweden) followed by incubation for 15 minutes at 37°C. The labeled cDNAs were purified with a QIAquick PCR Purification Kit (Qiagen). Prior to hybridization, Cy3/Cy5-labeled cDNA was quantified using a ND-1000 UV-VIS spectrophotometer (NanoDrop, Wilmington, DE, USA) to confirm dye incorporation. DNA microarrays were processed and analyzed as previously described [80].

Whole-genome location profiling by ChIP-chip and data analysis
RPA190 (ORF19.1839), RPC82 (ORF9.2847), RPB3 (ORF19.1248) and RPO21 (ORF19.7655) were TAP-tagged in vivo with a TAP-URA3 PCR product as described [81]. Transformants were selected on YPD-ura plates and correct integration of the TAP-tag was checked by PCR and sequencing. Cells were grown to an OD₆₀₀ of 2 in 40 ml of YPD. The subsequent steps of DNA cross-linking, DNA shearing, chromatin immunoprecipitation and DNA labeling with Cy dyes were conducted exactly as described by Lavoie et al. [81]. Tiling arrays were co-hybridized with tagged immunoprecipitated (Cy5-labeled) and mock immunoprecipitated (untagged BWP17 strain; Cy3-labeled) DNA samples. Microarray hybridization, washing and scanning were performed as described above. The significance cut-off was determined using the distribution of log-ratios for each factor. It was set at 2 standard deviations from the mean of log-transformed fold enrichments. Values shown are an average of two biological replicates derived from independently isolated transformants of tagged and mock constructs. Peak detection was performed using Gaussian edge detection applied to the smoothed probe signal curve as described [82].

Expression analysis by real-time quantitative PCR
For qPCR, cDNA was synthesized from 5 μg of total RNA using the RT system (50 mM Tris-HCl, 75 mM KCl, 5 mM dithiothreitol, 3 mM MgCl₂, 400 nM oligo(dT)₃₅, 20 ng random octamers, 0.5 mM dNTPs, 200 units Super-script III reverse transcriptase; Invitrogen). The mixture was incubated for 60 minutes at 50°C. cDNAs were then treated with 2 U of RNase H (Promega) for 20 minutes at 37°C followed by heat inactivation of the enzyme at 80°C.
for 10 minutes. Aliquots were used for qPCR, which was performed using the MX3000P QPCR System (Agilent, Santa Clara, CA, USA) with the QuantiTect SYBR Green PCR master mix (Qiagen). Cycling was 10 minutes at 95°C followed by 40 cycles (95°C, 10 s; 58°C, 15 s; 72°C, 15 s). Samples were done in triplicate and means were used for calculations. Fold changes were estimated using the coding sequence of the C. albicans ACT1 ORF as a reference. Fold enrichments of the tested coding sequences were estimated using the comparative ΔΔCt method as described [83]. Primers used for qPCR are summarized in Additional file 16.

Strand-specific RT-PCR
Strand-specific RT was performed as for the qPCR experiment. The RT reaction used 2 pmol of gene-specific primers (Additional file 16) designed to anneal to the AS transcript. Strand-specific RT-PCR was performed using 1 μl of the RT reaction. Cycling was 10 minutes at 95°C followed by 30 cycles (95°C, 10 s; 60°C, 55 s; 72°C, 30 s). As a negative control, RT-PCR was performed using RT and detection were performed as exactly described by Lavoie et al. [57].

GO annotation was performed using the GO Term Finder at the CGD website [27]. The P-value was calculated using hypergeometric distribution, as described on the GO Term Finder website. Motif detection of A. thaliana snoRNA promoters was performed using the TAIR Motif Finder tool [86].

Accession codes
Microarray data have been submitted to the NCBI Gene Expression Omnibus (GEO) under accession number [GEO:GSE22625].

Additional material

Additional file 1 Figure S1. GO analysis of the 28% of nuclear genes not expressed in this study.
Additional file 2 Table S1. Genome-scale detection of unannotated transcribed segments in C. albicans growing in different conditions.
Additional file 3 Table S15. List of nested or overlapping genes validated in this work.
Additional file 4 Table S2. List of detected ORFs and pseudogenes.
Additional file 5 Table S3. List of ORFs exhibiting long 5' and 3' UTRs (>240 bp).
Additional file 6 Table S4. Gene Ontology analysis of ORFs with long 5' and 3' UTR regions (>330 bp).
Additional file 7 Tables S5, S6, and S7. Genome-wide detection of ncRNAs. Table S5, AS transcripts; Table S6, housekeeping ncRNAs; and Table S7, RT-qPCR validation of randomly selected ncRNAs.
Additional file 8 Figure S2. Transcription and RNAP I and III occupancies within the rDNA locus.
Additional file 9 Tables S8 and S9. Detection of RNAP III binding peaks (Table S8) and genomic organization and coordinates of telomeric ncRNA (TelRs; Table S9).
Additional file 10 Figure S3. Transcription and RNAP III occupancy of ncRNAs. tRNAs (a, b), RPR1 (b) and an unknown ncRNA (c) are represented.
Additional file 11 Table S10. List of differentially expressed ORFs in hyphae, biofilm and caecum-grown cells.
Additional file 12 Table S11. List of differentially expressed ORFs in Candida Genome Database (CGD) GO Term Finder [27].
Additional file 13 Figure S4. Real-time quantitative PCR validation of candidate genes differentially expressed in caecum-grown Candida cells. Both, heme-binding (a) and carbohydrate catabolism genes (b) were considered.
Additional file 14 Table S12. Genome-scale detection of differentially expressed unannotated transfrags in C. albicans.
Additional file 15 Table S13. C. albicans strains used in the study [87,88].
Additional file 16 Table S14. Primers used in this study.

Abbreviations
AS: antisense; bp: base pair; CGD: Candida Genome Database; ChIP-chip: chromatin immunoprecipitation on chip; GO: Gene Ontology; LTR: long terminal repeat; ncRNA: non-coding RNA; ORF: open reading frame; qPCR: quantitative PCR; RNAP: RNA polymerase; RT: reverse transcription; snRNA: small nuclear RNA; snoRNA: small nucleolar RNA; TelRs: telomere-associated ncRNA; UTR: untranslated region.
Authors' contributions
AS and AN conceived and designed the experiments. AS performed the experiments with the help of CA and FT. AS and HH analyzed the data. CK, MuHlH and HL contributed reagents, materials and analysis tools. AS wrote the paper. AN and MW reviewed and edited the paper. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by a team grant from the Canadian Institutes of Health Research (CIHR) to AN, MW and others (CTP 79843). CA was supported by a Natural Sciences and Engineering Research Council of Canada and MvhH for the preparation of RNA from mouse caecum. This is NRC publication number 50694.

Author Details
1Biotecnology Research Institute, National Research Council of Canada, 6100 Royalmount, Montréal, Québec, H4P 2R2, Canada, 2Department of Anatomy and Cell Biology, McGill University, 3640 University Street, Montréal, Québec, H3A 1B1, Canada, 3Department of Biology, McGill University, 1205 Docteur Penfield, Montréal, Québec, H3A 1B1, Canada, 4Intracellular Signaling Laboratory, Institute of Research in Immunology and Cancer (IRIC), University of Montreal, 2900 boulevard Edouard-Montpetit, Montréal, Québec, H3C 3J7, Canada and 5Department of Molecular Biology and Microbiology, Tufts University, 136 Harrison Avenue, Boston, MA 02111, USA

Received: 22 April 2010 Revised: 7 June 2010 Accepted: 9 July 2010 Published: 9 July 2010

References
1. Leroy O, Gangneux JP, Montravers P, Mira JP, Cousin F, Sollet JP, Carlet J, Reynolds J, Rosenheim M, Regnier B, Lortholly O: Epidemiology, management, and risk factors for death of invasive Candida infections in critical care: a multicenter, prospective, observational study in France (2005-2006). Curr Care Med 2009, 37:612-1618.
2. Wisplinghoff H, Bischoff T, Silventi SM, Hinz PW, Wenzel RP, Edmond MB: Nosocomial bloodstream infections in US hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. Clin Infect Dis 2004, 39:309-317.
3. Kojic EM, Darouiche RO: Candida infections of medical devices. Clin Microbiol Rev 2004, 17:255-267.
4. Biswas S, Van Dijk P, Datta A: Environmental sensing and signal transduction pathways regulating morphophenotypic determinants of Candida albicans. Microbiol Mol Biol Rev 2007, 71:348-376.
5. Kontoyiannis DP, Lewis RE: Antifungal drug resistance of pathogenic fungi. Lancet 2002, 359:113-114.
6. Sanges D, Coste A, Ferrari S: Antifungal drug resistance mechanisms in fungal pathogens from the perspective of transcriptional gene regulation. EMBO Rep 2009, 10:1029-1036.
7. Morschhauser J: Regulation of multidrug resistance in pathogenic fungi. Fungal Genet Biol 2009, 46:97-104.
8. Braun BR, van Het Hoog M, d’Enfert C, Martchenko M, Dungan J, Koo A, Inglis DO, Uhl MA, Hougou B, Herrman B, Lorenz M, Leven M, Auerhoml Z, Bachewich C, Harcus D, Marci A, Dgnard D, Iouk T, Zizo R, Frangeul L, Tekkaia F, Rutherford K, Wang E, Munro CA, Bates S, Gove NA, Hoeyer LL, Kohler G, Morschhauser J, Newport G, et al.: A human-curated annotation of the Candida albicans genome. PLoS Genet 2005, 1:36-57.
9. van het Hoog M, Rast T, Martchenko M, Grindele S, Dgnard D, Hougou B, Cuomo C, Herrman M, Scherer S, Magee BB, Whiteway M, Chibana H, Nantel A, Magee PT: Assembly of the Candida albicans genome into sixteen supercontigs aligned on the eight chromosomes. Genome Biol 2007, 8:R52.
10. Jones T, Feddersen NA, Chibana H, Dungan J, Kalmann S, Magee BB, Newport G, Thorstenson YR, Agabian N, Magee PT, Davis RW, Scherer S: The diploid genome sequence of Candida albicans. Fungal Genet Biol 2006, 43:311-315.
11. Nantel A: The long hard road to a completed Candida albicans genome. Fungal Genet Biol 2007, 44:4453-4458.
12. Butler G, Rasmussen MD, Lin MF, Santos MA, Sakthikumar S, Munro CA, Rhenbey E, Grabber M, Forche A, Reedy JL, Agrafioti I, Arnaud MB, Bates S, Brown AJ, Brune S, Costanzo MC, Fitzpatrick DA, de Groot PW, Harris D, Hoeyer LL, Hube B, Klis FM, Kodira C, Lennard N, Logue ME, Martin R, Neiman AM, Nikolau E, Quail MA, Quinn J, et al.: Evolution of pathogenicity and sexual reproduction in eight Candida genomes. Nature 2009, 458:667-672.
13. Bertone P, Gerstein M, Snyder M: Applications of DNA tiling arrays to experimental genome annotation and regulatory pathway discovery. Chromosome Res 2005, 13:259-274.
14. Yazaki J, Gregory BD, Ecker JR: Mapping the genome landscape using tiling array technology.Curr Opin Plant Biol 2007, 10:534-542.
15. Johnson JM, Edwards S, Shoemaker D, Schadt EE: Mapping the conserved secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat Biotechnol 2005, 23:1383-1390.
16. Steigle S, Huber W, Storcus C, Studler PF, Nieselt K: Comparative analysis of structured RNAs in S. cerevisiae indicates a multitude of different functions. BMC Biol 2007, 5:23.
17. Bertone P, Stolc V, Royce TE, Rozowsky JS, Urban AE, Zhubu Q, Whorl RB, Lowsan JL, Sussman MR: Identification of transcribed sequences in Arabidopsis thaliana using high-resolution genome tiling arrays. Proc Natl Acad Sci USA 2005, 102:4458-4459.
18. Dutrow N, Nix DA, Hoo D, Milash B, Dalley B, Westbroek E, Parnell TJ, Cairns BR: Dynamic transcriptome of Schizosaccharomyces pombe shown by RNA-DNA hybrid mapping. Nat Genet 2008, 40:977-986.
19. Li L, Wang X, Stolc V, Li X, Zhang D, Su N, Tongprasit W, Li S, Cheng Z, Wang J, Deng XW: Genome-wide transcript analyses in rice using tiling microarrays. Nat Genet 2006, 38:124-129.
20. Washietl S, Hofacker IL, Lukasser M, Hutterhofer A, Studler PF, Mapping the conserved secondary structures predicts thousands of functional noncoding RNAs in the human genome. Nat Biotechnol 2005, 23:1383-1390.
21. Shinn P, Pellegrini M, Jacobsen SE, Ecker JR: Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell 2006, 126:1189-1201.
22. Shinn P, Arnaud MB, Costanzo MC, Inglis DO, Shah P, Binkley G, Miyazato SR, Sherlock G: New tools at the Candida Genome Database: biochemical pathways and full-text literature search. Nucleic Acids Res 2010, 38:D428-432.
23. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, Henderson IR, Shirin P, Pellegrini M, Jacobsen SE, Ecker JR: Genome-wide high-resolution mapping and functional analysis of DNA methylation in Arabidopsis. Cell 2006, 126:1189-1201.
24. Pokholok DK, Harbison CT, Levine S, Cole M, Hannon NM, Lee T, Bell GW, Walker K, Rolle PA, Herbolsheimer E, Zeitlinger J, Lewitter F, Gifford DK, Young RA: Genome-wide map of nucleosome acetylation and methylation in yeast. Cell 2005, 122:517-527.
25. Bernstein BE, Miserres A, Lander ES: The mammalian epigenome. Cell 2007, 128:669-681.
26. He H, Wang J, Liu T, Liu XS, Li T, Wang Y, Qian Z, Zhou H, Wang X, Shi B, Deng W, Zhou W, Skogerbo G, Chen R: Mapping the C. elegans noncoding transcriptome with a whole-genome tiling microarray. Genome Res 2007, 17:1471-1477.
27. Skrypnek MS, Arnaud MB, Costanzo MC, Inglis DO, Shah P, Binkley G, Miyazato SR, Sherlock G: New tools at the Candida Genome Database: biochemical pathways and full-text literature search. Nucleic Acids Res 2010, 38:D428-432.
28. Mitrovich QM, Tuch BB, Guthrie C, Johnson AD: Computational and experimental approaches double the number of known introns in the pathogenic yeast Candida albicans. Genome Res 2007, 17:492-502.
29. Mignone F, Giusi C, Lian M, Pesole G: Untranslated regions of mRNAs. Genome Biol 2002, 3(REVIEW)0004.
30. Nagalakshmi U, Wang Z, Waen K, Shou C, Raha D, Gerstein M, Snyder M: The transcriptional landscape of the yeast genome defined by RNA sequencing. Science 2008, 320:1344-1349.
31. Wilhelm BT, Marguerat S, Watt S, Gros P, Goodiss I, Tenkaitt CJ, Roberts I, Bahler J: Dynamic repertoire of a eukaryotic transcriptome surveyed at single-nucleotide resolution. Nature 2008, 453:1239-1243.
32. Lackner DH, Beilharz TH, Marguerat S, Mata J, Watt S, Schubert F, Preis T, Bahler J: A network of multiple regulatory layers shapes gene expression in fission yeast. Mol Cell 2007, 26:145-155.
33. Davuluri RV, Suzuki Y, Sugano S, Zhang MQ: CART classification of human 5’ UTR sequences. Genome Res 2008, 18:1807-1816.
34. Wang Y, Liu CL, Storey JD, Tishbirani RJ, Herschlag D, Brown PO. Precision and functional specificity in mRNA decay. *Proc Natl Acad Sci USA* 2002, 99:5860-5865.

35. Elson SL, Noble SM, Solis NV, Filler SG, Johnson AD. An RNA transport system in *Candida albicans* regulates hyphal morphology and invasive growth. *PLoS Genet* 2009, 5:e1000664.

36. Lehner B, Williams G, Campbell RD, Sanderson CM. Antisense transcripts in the human genome. *Trends Genet* 2002, 18:65-69.

37. Porecchi T, Xu Z, Claudio-Munster S, Steinmetz LM. Antisense artifacts in transcriptome microarray experiments are resolved by actinomycin D. *Nucleic Acids Res* 2007, 35:128.

38. Beiter T, Reich E, Williams RW, Simon P. Antisense transcription: a critical look in both directions. *Cell Mol Life Sci* 2009, 66:94-112.

39. Coram TE, Settles ML, Chen X. Large-scale analysis of antisense transcription in wheat using the Affymetrix GeneChip Wheat Genome Array. *BMC Genomics* 2009, 10:253.

40. Zhou X, Sunkar R, Jin H, Zhu JK, Zhang W. Genome-wide identification and analysis of small RNAs originated from natural antisense transcripts in *Oryza sativa*. *Genome Res* 2009, 19:70-78.

41. Zhang Y, Liu XS, Liu QR, Wei L. Genome-wide in silico identification and analysis of cis natural antisense transcripts (cis-NATs) in ten species. *Nucleic Acids Res* 2006, 34:3465-3475.

42. Lockhart SR, Zhao R, Daniels KJ, Soll DR. Alpha-pheromone-induced "shmooing" and gene regulation require white-opaque switching during *Candida albicans* mating. *Eukaryot Cell* 2003, 2:847-855.

43. Kurthara LJ, Beh CT, Latheren M, Schekman R, Rose MD. Nuclear congression and membrane fusion: two distinct events in the yeast karyogamy pathway. *J Cell Biol* 1994, 126:911-923.

44. Gammie AE, Stewart BG, Scott CF, Rose MD. The two forms of karyogamy transcription factor Kar4p are regulated by differential initiation of transcription, translation, and protein turnover. *Mol Cell Biol* 1999, 19:817-825.

45. Hongay CF, Grisafi PL, Gaitioki T, Fink GR. Antisense transcription controls cell fate in *Saccharomyces cerevisiae*. *Cell* 2007, 127:735-745.

46. Eddy SR. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2001, 2:919-929.

47. Matera AG, Terns RM, Terns MP. Non-coding RNAs: lessons from the small nuclear and small nucleolar RNAs. *Nat Rev Mol Cell Biol* 2007, 8:209-220.

48. Scott MJ, Li F. How do ncRNAs guide chromatin-modifying complexes to specific locations within the nucleus?. *RNA Biol* 2008, 5:13-16.

49. Lowe TM, Eddy SR. A computational screen for methylation guide RNA and RNA surveillance factors at mammalian chromosome ends. *Science* 2007, 318:708-801.

50. Schoeftner S, Blasco MA. Developmentally regulated transcription of mammalian telomeres by DNA-dependent RNA polymerase II. *Nat Cell Biol* 2008, 10:228-236.

51. Thrash-Ringham C, Gorman JA. Identification, characterization and sequence of Candida albicans repetitive DNAs Rel-1 and Rel-2. *Curr Genet* 1993, 23:455-462.

52. Louis EJ. The chromosome ends of *Saccharomyces cerevisiae*. *Yeast* 1995, 11:1553-1573.

53. Kumamoto CA. Niche-specific gene expression during *C. albicans* infection. *Curr Opin Microbiol* 2008, 11:325-330.

54. Brown AJ, Oddis FC, Gow NA. Infection-related gene expression in *Candida albicans*. *Curr Opin Microbiol* 2007, 10:307-313.

55. ten Cate JM, Klis FM, Perea-Cenci T, Cremers, W, de Groot PW. Molecular and cellular mechanisms that lead to *Candida* biofilm formation. *J Dent Res* 2009, 88:B105-115.

56. Sellam A, Niemi M, Micinney K, Brumfield S, Nantel A, Suci PA. A Candida albicans early stage biofilm detachment event in rich medium. *MBio Microbial Biol* 2009, 925.

57. Nett JE, Lepak AJ, Marchillo K, Andes DR. Time course global gene expression analysis of an in vivo *Candida* biofilm. *J Infect Dis* 2009, 200:307-313.

58. *Candida* Genome Database. [http://www.Candidagenome.org/](http://www.Candidagenome.org/)

59. Azzalin CM, Reichenbach P, Khoriul L, Giulotto E, Lingner J. Telomerase core components protect mammalian telomeres by DNA-dependent RNA polymerase II. *Mol Cell Biol* 2007, 27:855-868.

60. Kachouri R, Stibrinskis V, Zhu Y, Ramos KS, Westhof E, Li Y. A surprisingly large RNA Pol II RNA in *Candida glabrata*. RNA 2005, 11:1064-1072.

61. Boeke JD, Devine SE. Yeast retrotransposons: finding a quiet neighborhood. *Cell* 1998, 93:1087-1089.

62. Sandmeyer S. Targeting transposition: at home in the genome. *Genome Res* 1998, 8:816-818.

63. Hsu M, McEachern MJ, Dandjinou AT, Tzfati Y, Orr E, Blackburn EH, Lue NF. The RNA polymerase III-dependent family of genes in *Saccharomyces cerevisiae*: comparative RNomics, decoding strategies, transcription and evolutionary implications. *Nucleic Acids Res* 2006, 34:1816-1835.

64. Marck C, Kachouri-Lafond R, Lafontaine, I, Westhof E, Li Y. A surprisingly large RNA Pol II RNA in *Candida glabrata*. RNA 2005, 11:1064-1072.

65. Roberts DN, Stewart AJ, Huff JT, Cairns BR. The RNA polymerase III transcriptome revealed by genome-wide localization and activity-occupancy relationships. *Proc Natl Acad Sci USA* 2003, 100:14695-14700.

66. Mohtarli Z, Struhl K. Genome-wide occupancy profile of the RNA polymerase III machinery in *Saccharomyces cerevisiae* reveals loci with incomplete transcription complexes. *Mol Cell Biol* 2004, 24:4118-4127.

67. Guarniera E, Naef F, Shore D, Dieci G. Evolutionary tinkering with conserved components of a transcriptional regulatory network. *Cell Mol Life Sci* 2009, 66:94-112.

68. Bagchi P, Bagchi D, Bagchi S, Mitra S. Time course global gene expression analysis of an in vivo *Candida* biofilm. *J Infect Dis* 2009, 200:307-313.

69. Nantel A: Identification, characterization and sequence of Candida albicans repetitive DNAs Rel-1 and Rel-2. *Curr Genet* 1993, 23:455-462.

70. Sellam A, Al-Niemi T, McInnerney K, Brumfield S, Nantel A, Suci PA. A Candida albicans early stage biofilm detachment event in rich medium. *MBio Microbial Biol* 2009, 925.

71. Nett JE, Lepak AJ, Marchillo K, Andes DR. Time course global gene expression analysis of an in vivo *Candida* biofilm. *J Infect Dis* 2009, 200:307-313.
84. Candida dubliniensis genome sequence [http://www.sanger.ac.uk/sequencing/Candida/dubliniensis/]
85. TAIR Motif Finder [http://www.arabidopsis.org/tools/bulk/motiffinder/index.jsp]
86. Blast [ftp://ftp.ncbi.nlm.nih.gov/blast/]
87. Wilson RB, Davis D, Mitchell AP: Rapid hypothesis testing with Candida albicans through gene disruption with short homology regions. J Bacteriol 1999, 181:1868-1874.
88. Gillum AM, Tsay EY, Kirsch DR: Isolation of the Candida albicans gene for orotidine-5'-phosphate decarboxylase by complementation of S. cerevisiae ura3 and E. coli pyrF mutations. Mol Gen Genet 1984, 198:179-182.