Functional networks of co-expressed genes to explore iron homeostasis processes in the pathogenic yeast Candida glabrata

Thomas Denecker, Youfang Zhou Li, Cécile Fairhead, Karine Budin, Jean-Michel Camadro, Monique Bolotin-Fukuhara, Adela Angoulvant, Gaëlle Lelandais, Youfang Zhou li

Published online 20 April 2020

NAR Genomics and Bioinformatics, 2020, Vol. 2, No. 2 **1**
doi: 10.1093/nargab/lqaa027

Functional networks of co-expressed genes to explore iron homeostasis processes in the pathogenic yeast *Candida glabrata*

Thomas Denecker[1,†], Youfang Zhou Li[2,†], Cécile Fairhead[2], Karine Budin[1], Jean-Michel Camadro[3], Monique Bolotin-Fukuhara[2], Adela Angoulvant[2,4,†] and Gaëlle Lelandais[1,*,†]

[1]Université Paris-Saclay, CEA, CNRS, Institut de Biologie Intégrative de la Cellule (I2BC), 91198, Gif-sur-Yvette, France, [2]Université Paris-Saclay, INRAE, CNRS, Génétique Quantitative et Évolution Le Moulon, 91400, Orsay, France, [3]Université de Paris, CNRS, Institut Jacques Monod (IJM), 75013, Paris, France and [4]Parasitology and Mycology Department, Bicêtre University Hospital, Univ. Paris-Sud/Univ. Paris Saclay, Le Kremlin-Bicêtre, France

Received November 07, 2019; Revised February 27, 2020; Editorial Decision March 30, 2020; Accepted April 06, 2020

ABSTRACT

Candida glabrata **is a cause of life-threatening invasive infections especially in elderly and immunocompromised patients. Part of human digestive and urogenital microbiota, *C. glabrata* faces varying iron availability, low during infection or high in digestive and urogenital tracts. To maintain its homeostasis, *C. glabrata* must get enough iron for essential cellular processes and resist toxic iron excess. The response of this pathogen to both depletion and lethal excess of iron at 30°C have been described in the literature using different strains and iron sources. However, adaptation to iron variations at 37°C, the human body temperature and to gentle overload, is poorly known. In this study, we performed transcriptomic experiments at 30°C and 37°C with low and high but sub-lethal ferrous concentrations. We identified iron responsive genes and clarified the potential effect of temperature on iron homeostasis. Our exploration of the datasets was facilitated by the inference of functional networks of co-expressed genes, which can be accessed through a web interface. Relying on stringent selection and independently of existing knowledge, we characterized a list of 214 genes as key elements of *C. glabrata* iron homeostasis and interesting candidates for medical applications.**

INTRODUCTION

Infections due to *Candida* yeast species cause serious problems in aging populations and patients with compromised immunity, e.g. as a result of cancer treatment, organ transplantation or HIV infection (1, 2). A major cause of morbidity and mortality in healthcare structures (3), the frequency of candidemia and invasive candidiasis is increasing worldwide. While *Candida albicans* is known as the most common cause, the epidemiology varies according to geographical region, the populations involved and the survey period (1,2). In the United States and Northern Europe, *Candida glabrata* has been reported as the second cause of candidiasis (1). *Candida glabrata* is a pathogenic yeast species whose haploid genome was described in 2004 (4). It is composed of 13 chromosomes with ∼5200 genes. Despite its name, *C. glabrata* is phylogenetically more closely related to the model yeast *Saccharomyces cerevisiae* than to the pathogenic yeast *C. albicans* (4). Notably, *C. glabrata* infections remain challenging to treat owing to delayed diagnosis, natural low susceptibility to azole antifungals and acquired resistance to echinocandins (5–7).

During host infection, pathogens face abrupt physiological changes in their immediate environment (8). In this context, promising therapies are expected to emerge from a better understanding of the homeostatic processes used by pathogens to protect (or restore) an internal stability in their cellular functions. A major player is iron homeostasis, as iron bioavailability is a key factor involved in the 'nutritional immunity' host-defense mechanism (9). Remarkably, iron is a two-faced oligo-element for living organisms. On the one hand, iron is essential, as part of heme- and iron-sulfur cluster (ISC)-containing proteins involved in a variety of vital functions including oxygen transport, DNA synthesis, metabolic energy or cellular respiration (see (10) for review). On the other hand, iron is toxic. Its excess triggers oxidative stress, lipid peroxidation and DNA damage that ultimately compromise cell viability and can promote

*To whom correspondence should be addressed. Tel: +33 1 69 82 46 80; Email: gaelle.lelandais@universite-paris-saclay.fr
†Contributed equally.

programmed cell death (11). Iron homeostasis is therefore essential to allow pathogens to maintain a balance between iron utilization, storage, transport and uptake in the host environment.

Molecular mechanisms of iron acquisition and consumption in fungi are well described in the literature (see (12) or (13) for reviews). The yeast species *S. cerevisiae* and *C. albicans* were long considered paradigms for non-pathogenic and pathogenic species, respectively, but the situation is changing. Several articles describe the regulatory mechanisms involved in *C. glabrata* iron homeostasis ((14–21) and see (22) for a comprehensive review). Although *C. glabrata* has conserved the classical fungal iron regulon, it has also remodeled its own functional networks to maintain iron homeostasis. So far, experimental studies have essentially described the impact of iron deficiency on gene expression in *C. glabrata*. Iron deficiency is indeed a relevant system for mimicking infections of the human host, during which access to iron for the pathogen is severely limited by the host defense mechanisms (23,24). Nonetheless, iron deficiency is not representative of all situations to which *C. glabrata* is exposed during its life cycle in the human body. *Candida glabrata* is either a commensal or a colonizer, at least transiently, of the digestive and urogenital tracts. When digestive or urinary epithelium are damaged due to hypoperfusion, medications or invasive procedures, *C. glabrata* can translocate to blood and then disseminate. *Candida glabrata* in the digestive tract faces either high or low iron concentrations depending on dietary sources, gut motility and even microbiota (25). In the urinary tract, *C. glabrata* faces iron concentrations which can be high and increase with the host's age (26,27). In the blood phagocytes, iron concentrations are low (28). The notion of 'homeostasis' is thus particularly important for *C. glabrata* as colonizer since the pathogen must maintain an internal iron balance despite external fluctuations in iron bioavailability in the immediate environment.

The aim of the present work is to specifically study iron homeostasis. In particular, we want to highlight key genes, which are systematically deregulated when *C. glabrata* faces decreased or increased bioavailability of iron. We performed transcriptomic experiments (microarray technology) to monitor gene expression changes of *C. glabrata* to ferrous iron (Fe^{2+}) deficient and overload conditions at 30°C and 37°C. These temperatures are respectively the usual temperature at which *C. glabrata* is cultivated in laboratories and the temperature at which *C. glabrata* develops in the human body. The resulting dataset was analyzed to (i) clarify the potential effect of temperature on iron homeostasis, (ii) identify iron responsive genes, i.e. genes significantly up- or downregulated in at least one iron imbalanced situation and (iii) define a new set of genes, referred to hereafter as 'iron homeostasis key genes' (iHKG, Figure 1). These genes are good candidates to be chief components of iron homeostasis. Our exploration of the datasets was facilitated by the inference of functional networks of co-expressed genes, which can be accessed through a web interface (https://thomasdenecker.github.io/iHKG/). The philosophy of this work is to empower experimental researchers by providing access to all transcriptomics data and by generating easily interpretable graphical outputs. This should facilitate deep exploration of genome-wide functional data in the pathogenic yeast *C. glabrata* to advance our global understanding of iron homeostasis.

MATERIALS AND METHODS

Transcriptome analyses

Yeast strains and growth conditions. The wild-type *C. glabrata* strain used for gene expression analysis under iron deficiency and overload conditions (see below) is ATCC 2001 (CBS 138), as described in (29). Cells were frozen in 40% glycerol at −80°C and used thereafter. They were first cultured on yeast extract-peptone-glucose (YPD) agar plates and then sub-cultured in YPD liquid medium at 30°C, on a rotating shaker (150 rpm) for 24 h, by inoculating 10^7 yeast cells in 10 ml of medium.

Conditions of cell culture for RNA isolation. All culture conditions were performed in YPD, starting from a sample of 2×10^6 cells, inoculated in 10 ml of YPD medium, then cultured under gentle shaking (150 rpm) at 30°C or 37°C for 4 h (log phase, determined from growth curves). Initial cultures were performed in standard conditions (referred to as 'Control', Figure 2 and Supplementary Data S1). Addition of 100 μM bathophenanthrolinedisulfonic acid (4,7-diphenyl-1,10-phenanthrolinedisulfonic acid disodium salt hydrate or BPS, SIGMA-ALDRICH® France) was used as iron deficiency condition (referred to as 'BPS', Figure 2 and Supplementary Data S1). The dose was chosen based on preliminary experiments showing that it affects but does not stop growth of yeasts. Addition of 500 μM of $FeSO_4$ heptahydrate (SIGMA-ALDRICH®, France) was used as iron excess condition (referred to as 'FeSO4', Figure 2 and Supplementary Data S1). Again, this was determined in preliminary experiments as the highest concentration that affects but does not stop yeast growth. Note that the concentration of 500 μM of $FeSO_4$ heptahydrate is a modest excess, close to iron concentrations in the human gut and urine. All conditions of cell culture (iron deficiency or excess, at 30°C or 37°C) were performed three times, starting from independent pre-cultures, to cover biological variations. After cell cultures in standard, iron deficiency or excess conditions (each at 30°C or 37°C), an aliquot of 1 ml with 2×10^7 cells was frozen at −80°C for RNA purification.

RNA preparation, labeling and microarray hybridization. RNAs were isolated using RNeasy Mini Kit for purification of total RNA (QUIAGEN, GmbH, Germany) according to the manufacturer's instructions. Concentration was determined using a Nanodrop 2000 instrument. RNA quality was checked using the Agilent 2 bioanalyzer Nanochip system according to the manufacturer's instructions. Microarrays were manufactured with eArray (https://earray.chem.agilent.com/earray/) from Agilent Technologies. There are described in the Gene Expression Omnibus (GEO) database (30) under 'Platform GPL27653'. Technical information is provided in Supplementary Note S8. Probe preparation, labeling and hybridization were performed with Agilent Technologies according to the manufacturer's instructions. The reference sample is a mixture of RNA extracted from all different growth conditions.

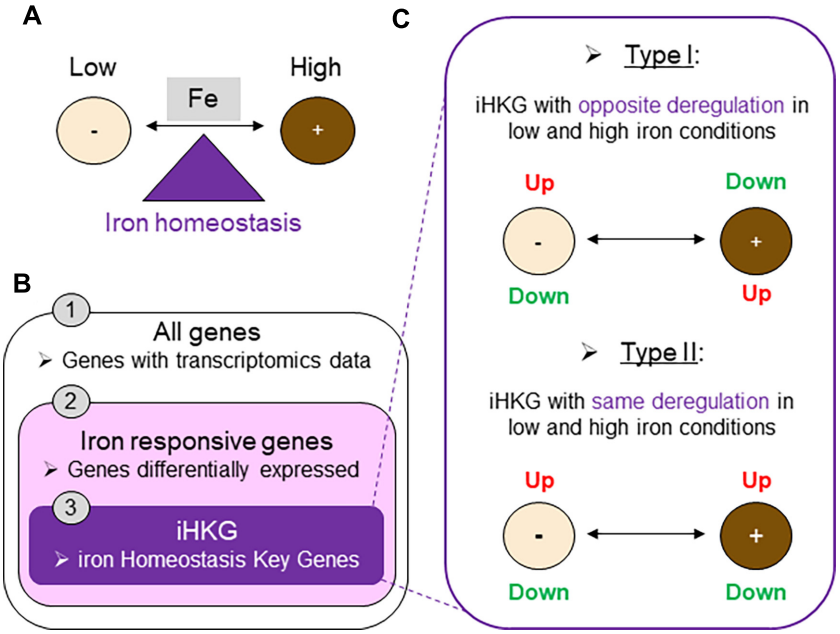

Figure 1. Definitions and working hypothesis. (**A**) Schematic representation of extracellular changes in iron availability, faced by the pathogenic yeast *C. glabrata*. 'Low' means that the iron availability is low, the cell must adapt to iron deficiency. 'High' means that the iron availability is high, the cell must adapt to iron overload. Iron homeostasis is represented here as the central physiological process to maintain an internal cellular environment in a constant state balance, despite the external changes. (**B**) Three classes of genes studied in this article. The first class 'All genes' refers to all the genes for which transcriptomics data is obtained, the second class 'Iron responsive genes' refers to the genes for which expression changes are observed in at least one transcriptomic experiment. At last, the third class 'iHKG' refers to a new set of genes with particular functions important for the cell to counterbalance external fluctuations in iron availability, in any direction (low or high). (**C**) Schematic representation of the two types of iHKG based on the de-regulations observed in our dataset, respectively in 'low' (−) and 'high' (+) iron conditions. 'Type I' are iHKG with opposite deregulations in low and high iron conditions, whereas 'Type II' are iHKG with constant (or parallel) deregulation in low and high iron conditions.

Microarray data acquisition, inter-channel normalization and public access. Microarrays were scanned using the Agilent Array Scanner according to manufacturer's instructions. Pre-processing and inter-channel (Cy5 and Cy3) fluorochrome normalization were performed with MAnGO software (31). Artefactual spots were eliminated from the analysis, duplicated spots were averaged. The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (30) and are accessible through GEO Series accession number GSE139363 (https://www.ncbi.nlm.nih.gov/geo/).

Bioinformatics analyses

Source code availability, access to intermediate result files and reproduction of all analyses. All the source code written for this project is available in the Github repository https://github.com/thomasdenecker/iHKG/. The analysis starts with reading the MAnGO data file deposited in GEO (see previous section). This ensures continuity between the work on raw data (standard results of microarrays processed by a technical platform) and the work of data mining and analyses performed for this article (our contribution). To reproduce our data output tables, our selections of genes and our output graphs, a docker image is provided, with RStudio software (https://rstudio.com/) and all the necessary R packages installed: https://cloud.docker.com/repository/docker/tdenecker/ihkg/general. The website to explore the functional networks of co-expressed genes is available at https://thomasdenecker.github.io/iHKG/.

Statistics for differential expression. Our transcriptomic results are from two-color microarray experiments with a common reference, *i.e.* a mixture of RNA extracted from all the other growth conditions (see previous section). To compare gene expression levels between 'BPS' and 'Control' on the one hand, and between 'FeSO4' and 'Control' on the other hand, a design matrix was defined as explained in the LIMMA user guide available online https://bioconductor.org/packages/release/bioc/vignettes/limma/inst/doc/usersguide.pdf. This matrix allows identification in the entire dataset of the different conditions to be compared and the associated biological replicates. Statistics for differential expression can thus be calculated, applying the LIMMA linear model (32). Results include log2 fold changes (logFC), *t*-statistics and adjusted *P*-values (Benjamini–Hochberg correction). Note that statistics for differential expression were calculated independently using results of microarrays obtained either at 30°C or 37°C. This allowed us to obtain statistics for all the genes in four different conditions referred to as C1 (low Fe–30°C), C2 (low Fe–37°C), C3 (high Fe–30°C) and C4 (high Fe–37°C). Gene expression levels were also compared between 30°C and 37°C using 'Control' samples obtained at each temperature (referred to as C30–37, Figure 2). Variability observed in logFC values for C1, C2, C3, C4 and C30–37 conditions was standardized, calculating *Z*-Score

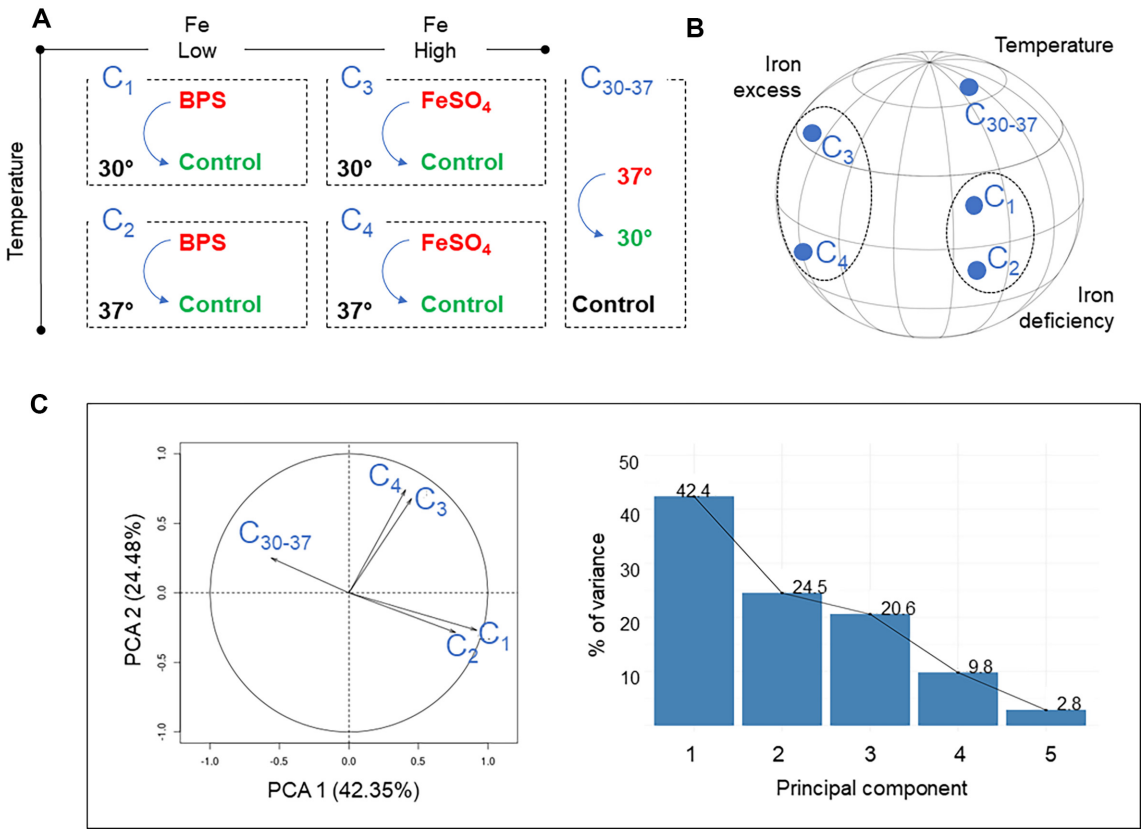

Figure 2. Global analysis of transcriptome changes in iron deficiency or iron overload conditions at 30°C and 37°C. (**A**) Schematic representation of the experimental design used for comparing mRNA abundance with microarray technology. Five conditions named C_1, C_2, C_3, C_4 and $C_{30–37}$ were defined. Z-Score values were derived for each condition, comparing samples written in red (respectively, 'BPS', 'FeSO4' and '37°C') to samples written in green (respectively, 'Control' and '30°C'). (**B**) Spherical representation of the correlation matrix between Z-Score values for the five conditions shown in (A). The smaller the distance between the points, the greater the correlation values. (**C**) Biplot representing the five conditions using the coordinate system defined by principal components 1 and 2 (left), and a histogram showing the percentage of variance represented by each principal component. Together, the principal components 1 and 2 account for more than 66%.

values, i.e. $Z-\text{Score} = \frac{\text{logFC} - \text{mean}}{\text{stand. dev.}}$ where 'mean' is the average value for logFC of all genes in a particular condition (this value is around 0 due to microarray raw data normalization and hence no significant difference is observed between conditions) and 'stand. dev.' is the standard deviation of logFC of all genes. This is the parameter that must be normalized (standard deviation equal 1 in *Z*-Score distribution) to ensure a constant stringency of gene selection in all conditions (see below).

Principal component analysis (PCA). Values of *Z*-Scores for all genes in all conditions were analyzed by principal component analysis (PCA) using the librairies FactoMineR (https://cran.r-project.org/package=FactoMineR) and psy (https://cran.r-project.org/package=psy) with default parameters. Detailed interpretations of the 3D sphere and the biplots can be found in (33,34).

Selection of iron responsive genes and iron homeostasis key genes (iHKG). Selection of iron responsive genes, as defined in Figure 1, was based on statistics for differential expression, i.e. *Z*-Score values and adjusted *P*-values. Upregulated (or induced) genes are those with a *Z*-Score value greater than two and an adjusted *P*-value lower than 5%, whereas downregulated (or repressed) genes are those with a *Z*-Score value lower than −2 and an adjusted *P*-value lower than 5%. Each gene that was observed as up- or downregulated in at least one of the conditions C1, C2, C3 or C4 was included in the final list of 'iron responsive genes'. From all the iron responsive genes, iHKG genes, as defined in Figure 1, were selected by applying the following successive rules: (i) adjusted *P*-value < 0.01, (ii) *Z*-Score value higher than 1.5 or lower than −1.5, (iii) validation of criteria (i) and (ii) in conditions 'low Fe' (C1 or C2) and in conditions 'high Fe' (C3 or C4). In iHKG, Type I genes are those which were found upregulated (respectively, downregulated) in 'low Fe' conditions (C1 or C2) and downregulated (respectively, upregulated) in 'high Fe' conditions (C3 or C4). Type II genes are those which were found upregulated (respectively, downregulated) in 'low Fe' conditions (C1 or C2) and also upregulated (respectively, downregulated) in 'high Fe' conditions (C3 or C4).

Data mining of GO terms, definition of Meta-GO and allocation of iron responsive genes to general functions. Gene Ontology (GO) terms were used to define the Meta-GO (Figure 3). They were extracted from the 'Biological Process' section of the GO database (35). The li-

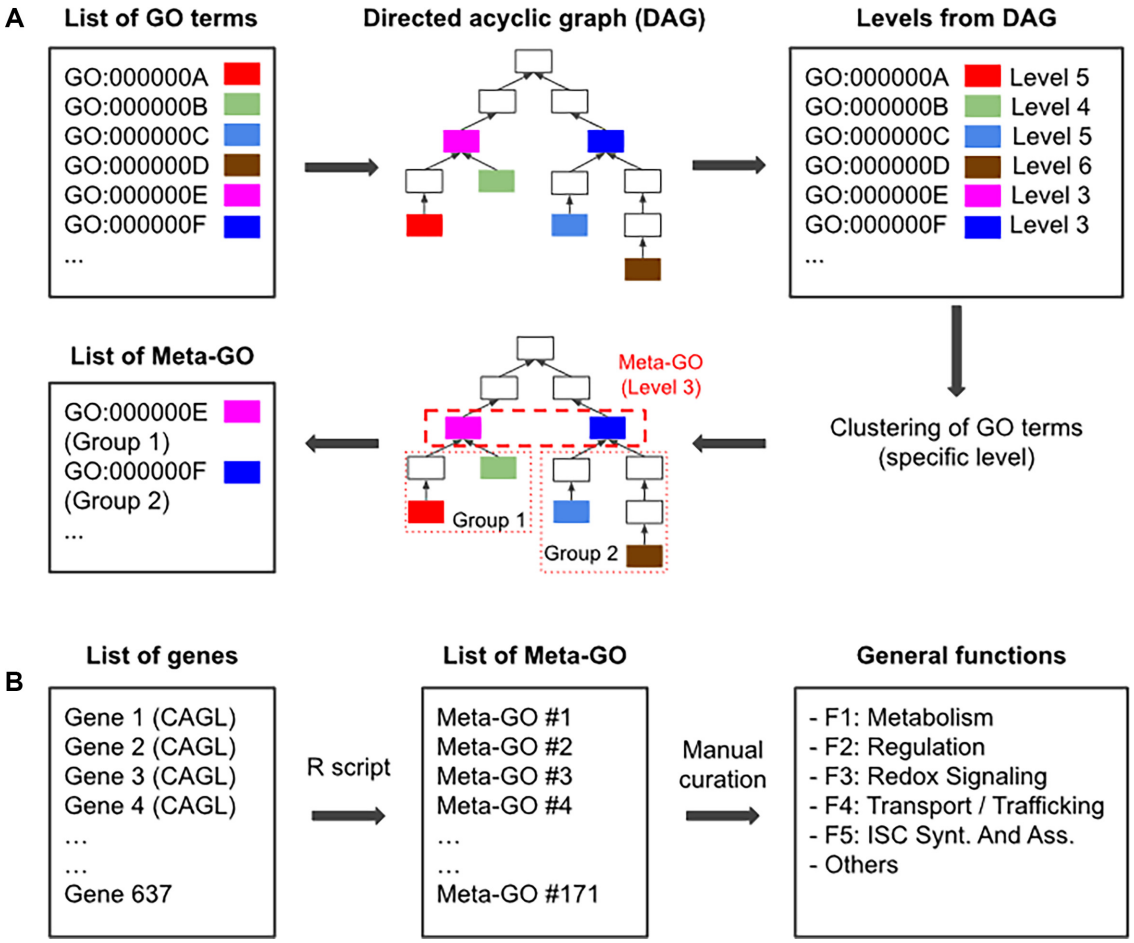

Figure 3. Data mining of GO terms to define a limited number of general functions to which all iron responsive genes can be assigned. (**A**) Schematic representation of the process applied to manage GO terms in this study. Starting from a list of GO terms (all terms in 'Biological Process' here), they are located in the DAG provided by the GO database. Levels are associated to each GO term according to their position in the DAG (see 'Materials and Methods' section and Supplementary Figures S10). This is illustrated here with colored boxes. GO terms are next clustered according to a specific level (Level 3 in this example) in the GO hierarchy. This produces 'Meta-GO', i.e. groups of GO terms which share a common ancestor in the DAG. Note that in the analysis of iron responsive genes presented in this article, Level 4 was used to create Meta-GO (see 'Materials and Methods' section and Supplementary Data S3). (**B**) General functions defined in this study to highlight physiological roles of iron responsive genes. They are named 'Metabolism' (F1), 'Regulation' (F2), 'Redox signaling' (F3), 'Transport/trafficking' (F4), 'Iron sulfur cluster synthesis and assembly' (F5) and 'Others'.

brary OntologyIndex (https://cran.r-project.org/package=ontologyIndex) was used to associate each term with a 'Level'. By definition, a level is the length of the longest path, which exist between a GO term (e.g. 'iron ion transmembrane transport'—GO:0034755) and the term 'biological process' (GO:0008150) that is located at the root of the GO hierarchy (see Supplementary Figures S10 for an illustration). Once the levels were assigned, only the terms related to at least one of the iron responsive genes (see previous section) were retained and used for Meta-GO assignments. Because we observed that GO terms in *C. glabrata* were not accurate enough, the GO terms associated with orthologous genes in *S. cerevisiae* were preferred at this step. Meta-GO assignment consisted to group the GO terms at a specific level in the GO hierarchy and assign all of them to the associated common ancestor in the DAG (see Figure 3A for an illustration). The Meta-GO defined in this study were obtained at Level 4. They were allocated (manually) to general functions named 'Metabolism' (F1), 'Regulation' (F2), 'Redox signaling' (F3), 'Transport/trafficking' (F4), 'Iron-sulfur cluster synthesis and assembly' (F5) and 'Others' (Figure 3B). This procedure allowed to further classify (through the Meta-GO) the iron responsive genes into the general functions. Note that in case of multiple functions for a gene (a gene can be associated with several GO terms, in several Meta-GO, in different general functions) the general function in which the gene had the highest number of associated GO terms was chosen.

Networks of co-expressed genes. Gene networks were inferred using a simple approach, which is based on distance calculations between gene expression measurements in conditions C1, C2, C3, C4. A schematic representation of the method can be found in Figure 4. In this work, Euclidean distances were calculated between all pairwise gene expression profiles (R function 'dist') and the threshold to connect nodes on the network was fixed such that only the 5% of pairs of genes with the smallest distances between

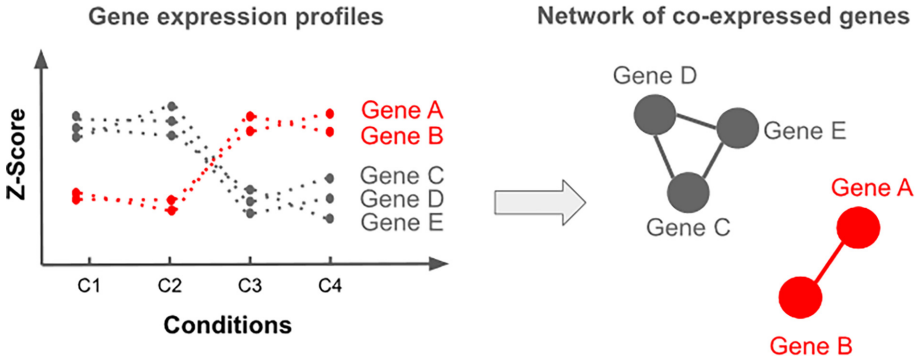

Figure 4. Schematic representation of a network of co-expressed genes. In such a network, genes are represented by nodes and similarities between gene expression profiles of two genes are represented by edges. To infer a network it is necessary to (i) calculate a measure of distance between gene expression profiles (Euclidean distance for instance), (ii) define a threshold value such that if the distance between two gene expression profiles is below the threshold, the corresponding nodes on the network are connected with an edge and (iii) apply dedicated algorithms to calculate node positions in a two dimensional space, such that co-expressed genes are represented by nodes positioned side by side as illustrated here with genes A and B (on the right) and genes C, D and E (on the left).

expression profiles would be connected. The network of co-expressed genes were represented using a combination of the library igraph (https://cran.r-project.org/package=igraph) for the calculation of the positions of nodes (R function 'layout_nicely') and the software Cytoscape (36) for dynamic network exploration in https://thomasdenecker.github.io/iHKG/.

Source of gene descriptions and annotations, sequence data and orthologous relationships between genes. Gene description and annotations were retrieved from the Candida Genome Database (CGD) (37) and the GRYC database (http://gryc.inra.fr/). Functional enrichment of GO terms was performed with the version of GOterm Finder (38) available in CGD (http://www.candidagenome.org/cgi-bin/GO/goTermFinder) and GO SLIM Mapper (http://www.candidagenome.org/cgi-bin/GO/goTermMapper). Orthologous relationships between *C. glabrata* and *S. cerevisiae* genes were downloaded from CGD (http://www.candidagenome.org/download/homology/orthologs/C_glabrata_CBS138_S_cerevisiae_by_CGOB/), as well as orthologous relationships between *C. glabrata* and *C. albicans* (http://www.candidagenome.org/download/homology/orthologs/C_glabrata_CBS138_C_albicans_SC5314_by_CGOB/).

RESULTS

Transcriptomic experiments in two unbalanced iron conditions (low and high) to reveal the 'key genes' involved in iron homeostasis

Definitions, working hypothesis and experimental design. In this study, our aim was to identify the 'key genes' required for iron homeostasis in the pathogenic yeast *C. glabrata*. Referred to as iHKG, these genes are expected to have an important activity for iron homeostasis that we define as the property of the cells to counterbalance intracellular physiological consequences of external iron changes (Figure 1A). In particular, we wanted to discriminate iHKG from the set of iron-responsive genes, which are all the genes exhibiting up- or downregulation in response to iron fluctuations. In this context, our working hypothesis was the following. Because iHKG are involved in iron homeostasis specifically, and that homeostasis is associated with a natural resistance to bidirectional changes (low or high iron availability), iHKG should exhibit de-regulation in both low and high iron source conditions. Other iron responsive genes should exhibit de-regulation in only one of the conditions, either low or high. The iHKG thus represent a subclass of the iron responsive genes (Figure 1B). Based on this, we defined two types of iHKG, 'Type I' and 'Type II' (Figure 1C). Type I genes are those for which the de-regulation is opposite in low and high iron conditions (respectively up- and downregulated), whereas Type II genes are those for which the de-regulation is identical in low and high iron conditions (constantly up- or downregulated). Note that we could anticipate, based on these definitions, that Type II genes would also include genes involved in the general stress response of cells (ESR genes for instance (39)) and hence would be less specific to iron physiological processes.

To reveal iHKG, we designed an experimental plan (Figure 2A) in which cells were cultured under two iron conditions: iron-deficiency (Fe 'low') and iron-overload (Fe 'high'), at two different temperatures: 30°C and 37°C. Thirty degrees is the usual temperature at which *C. glabrata* is cultivated in laboratories, while 37° is the temperature at which *C. glabrata* develops in the human body. Since iron homeostasis is an important physiological process involved in host-pathogen interactions, it was important to assess the potential effect of the temperature. Transcriptomics experiments were performed with microarray technology, to quantify gene expression changes with respect to appropriate controls (see 'Materials and Methods' section). As a result, experimental data were obtained in the five conditions referred to hereafter as C1, C2, C3, C4 and C30–37, for all *C. glabrata* genes that could be monitored with microarrays.

Temperature shift has a limited impact on gene expression changes associated with cell responses to unbalanced iron conditions. In the first step of the analysis, our aim was to evaluate the potential effect of temperature shift (30°C–37°C) on gene expression changes monitored in 'low' and

'high' iron conditions. For that, we applied PCA to study the correlation matrix between normalized expression measurements obtained for all genes in conditions C1, C2, C3, C4 and C30–37 (see 'Materials and Methods' section). Results are presented in Figure 2B and C. The five experimental conditions are represented by individual variables and they are on a unit hypersphere on which the distances between them are directly proportional to the initial correlation values stored in the matrix (34). Notably, we observed (i) that gene expression measurements are highly correlated between conditions C1 and C2 in the one hand, and conditions C3 and C4 on the other hand, and (ii) that the condition C30–37 is clearly separated from the others (Figure 2B and C). This observation indicates that if the temperature shift has an effect on the expression of *C. glabrata* genes, this effect is unrelated to any effects caused by extracellular iron imbalance. In other words, most of the 'iron responsive genes', which are genes de-regulated in response to iron deficiency or iron overload at 30°C, are also de-regulated at 37°C. This is an interesting result, which allows us to consider the generalization of experimental results obtained in a laboratory at 30°C to a more realistic living situation for *C. glabrata*, as a human pathogen. In the rest of this work, the transcriptomics data we obtained at 30°C and 37°C were combined in order to simplify iHKG identification. Note that another interesting property could be drawn from this PCA. We observed that projections of the initial condition vectors (C1-C2 and C3-C4) were orthogonal (90° angles, Figure 2C). This means that gene expression changes in response to 'low iron' conditions are essentially unrelated to gene expression changes in response to 'high iron'. This observation is important with regard to the respective roles of Type I and II genes (see next sections).

Type I and type II iHKG represent small sub-classes of the set of all iron responsive genes. Among the genes for which expression data were available, we identified those displaying significant changes in mRNA levels, respectively at 'low' or 'high' levels of extracellular iron availability. We used the LIMMA statistical procedure to identify differentially expressed genes based on replicated experiments in C1, C2, C3 and C4 (see 'Materials and Methods' section). As a result, we found that 637 genes were significantly up- or downregulated in at least one condition. Together, these genes represent our complete set of 'iron responsive genes' (Figure 1B). Among these genes, we found 214 iHKG of which 73 were Type I and 141 were of Type II (see 'Materials and Methods' section). Together iHKG represent 33% of iron responsive genes. They constitute, as anticipated from PCA analysis, a fairly small sub-class of genes, especially with regard to type I genes (11%). A detailed list of iron responsive genes and iHKG can be found in Supplementary Data S2.

Representation of iron responsive genes in functional networks of co-expressed genes

Definition and objectives. Starting from the list of 637 genes up- or downregulated in at least one condition (the iron responsive genes, see previous sections), we can infer functional networks of co-expressed genes. By 'functional networks', we mean graphs in which genes (represented as nodes) are (i) involved in a common cellular function and (ii) are connected by edges if they react similarly during iron homeostasis. On these graphs, we expected the iHKG to be easily highlighted and thus integrated into a more comprehensive functional context. Our computational strategy to infer and visualize the functional networks was divided into three steps: (i) placing all the iron responsive genes in a cellular function to which they contribute, (ii) representing them, in each function, by applying methods for gene co-expression network inference, (iii) developing an interactive web viewer for network exploration, rapid location of any gene and retrieval of its biological context. These three steps are detailed below.

Step 1: Computational strategy to place all the iron responsive genes in a small number of functional categories. Unlike the classical approach which consists in identifying functional categories (generally GO terms) significantly enriched in our list of genes (38), we wanted here to define a small number of cellular functions (fewer than 10) and assign all of the 637 iron responsive genes to only one of the pre-defined functions. This limited number of functions was required to allow subsequent gene network inference and visualization. To define our functions, we first classified all GO terms (more than 40 000) available in the 'Biological Process' section of the GO database into Meta-GO groups (see 'Materials and Methods' section). As illustrated Figure 3A, the 'Meta-GO' are GO terms, which share a specific level in the GO directed acyclic graph (DAG), the hierarchical organization of terms defined by the Gene Ontology Consortium (35). They serve as representatives for all other related terms in the DAG. Next, we selected all Meta-GO (level 4) in which at least one GO term was assigned to one iron responsive gene (one of the 637 genes). We found at this step 171 Meta-GO. They were inspected and manually classified into the general functions named 'Metabolism' (F1), 'Regulation' (F2), 'Redox Signaling' (F3), 'Transport/trafficking' (F4), 'Iron sulfur cluster synthesis and assembly' (F5) or 'Others' (Figure 3B). These general functions are emblematic of the key cellular processes by which yeast cells adapt their functioning to iron deficiency or iron overload. They are detailed in Table 1. All information regarding data mining of GO terms, Meta-GO assignments and clustering into general functions is available in Supplementary Data S3.

Step 2: Computational strategy to derive functional networks from gene expression measurements. We calculated the allocation of the 637 iron responsive genes into the functional categories described in the previous section. We observed 28% of genes in F1 (Metabolism), 17% in F2 (Regulation), 18% in F3 (Redox signaling), 13% in F4 (Transport/trafficking) and 3% in F5 (ISC synthesis and assembly). Altogether almost 80% of the iron responsive genes were classified in one of the general functions based on Meta-GO exploration. We also found that iHKG are well-represented in each function: 34% in Metabolism (F1), 29% in Regulation (F2), 42% in Redox Signaling (F3), 25% in Transport/trafficking (F4) and 36% in ISC synthesis and assembly (F5). To derive gene networks within each function, we calculated co-expression graphs (see 'Materials and

Table 1. Functional categories defined in this work to represent iron responsive genes using functional networks of co-expressed genes

Label	Name	Includes Meta-Go related to:	# of Meta-GO
F1	Metabolism	Nucleic acid, amino acid, fatty acid and lipid metabolism ; Carbon metabolism/energy production from respiratory and non respiratory origin ; Mitochondria functions (including biogenesis, mitophagy and functions other than redox signaling that are included in function F3, see below) ; Membrane and cell wall biogenesis ; Ribosomal biogenesis (biosynthesis, RNA processing and translation).	41
F2	Regulation	Transcriptional regulation (including general transcription) ; Post-translational modifications (including protein phosphorylation, glycosylation, structural modification and degradation) ; Ribosome activity (including translation) ; Signal transduction ; RNA and protein fate.	8
F3	Redox signaling	Proteins with functions directly or indirectly linked with redox mechanisms, i.e. oxygen dependant enzymes, flavo-hemoproteins including Cytochrome P dependent, membrane iron reductase, NADP/NADPH-dependent enzymes, metalloenzymes. These proteins are thus involved in thiol and redox signaling pathways, respiratory chain components, peroxisome activity, ROS detoxification, carrier proteins, folate biosynthesis, heme biosynthesis, etc.	7
F4	Transport/trafficking	Cell exchanges between intra and extracellular compartments (directed movement of substances such as macromolecules, small molecules, ions, etc.) ; Processes for internal cell trafficking.	3
F5	Iron sulfur cluster synthesis and assembly	Chemical reactions and pathways involving sulfur or compounds that contain sulfur.	1
OTHERS	Membrane/cell wall| Pathogenesis| Stress response| Unclassified	Remaining Meta-GO for which no classification was clear or not associated specifically to iron homeostasis: • 'Unclassified' (73 Meta-GO) • 'Stress Response' (26 Meta-GO) • 'Membrane/Cell Wall' (8 Meta-GO) • 'Pathogenesis' (5 Meta-GO).	112

These functions were meant to be representative of the cellular processes by which yeast cells adapt to iron deficiency or iron overload. *Candida glabrata* indeed tunes its response by adapting the biological systems devoted either to iron acquisition or to the mobilization of cellular iron storage (represented with the function 'Transport/trafficking'). Cellular processes that require iron for function are also modified (functions 'Metabolism' and 'Redox signaling'). The underlying regulatory pathways are well described, including transcriptional and post-transcriptional mechanisms (function 'Regulation'). At last, critical roles played by iron-sulfur clusters are known (function 'Iron sulfur cluster synthesis and assembly') and complex relationships between iron homeostasis and oxidative stress response is often emphasized (function 'Stress response' in 'Others' category). Methodology to explain the Meta-GO is presented Figure 3 and in 'Materials and Methods' section. Note that the category 'Others' comprise all remaining Meta-GO for which no classification was clear or not associated specifically to iron homeostasis.

Methods' section). The main idea is to measure the similarity between gene expression measurements of all genes in conditions C1–C4. If the similarity is high enough between two genes, these two genes are represented by connected nodes on the graph (see Figure 4). Graphs of co-expressed genes obtained for the functional categories F1–F5 are presented in Figure 5. Note that if a unique graph was obtained for a function then two complementary representations were derived by coloring the nodes according to gene expression measurements obtained in 'low' or 'high' iron conditions. This allowed us to observe the symmetrical de-regulations that characterize Type I and Type II iHKG (Figure 1).

Step 3: Computational strategy to interactively explore the functional networks of co-expressed genes. The main interest of Figure 5 is to give a global overview of the gene expression processes used by the pathogenic yeast *C. glabrata* to face low and high iron conditions. We integrated here more than 100 000 qualitative and quantitative heterogeneous kinds of information (GO terms, microarray data, gene annotations in public databases, etc.) to obtain a unified picture of multiple cellular processes. We added to Figure 5 the names of around fifty genes, which we found to agree with current knowledge of iron homeostasis in *C. glabrata* (see the next section). It is clear, however, that further exploration of these networks could be considered, examining for instance the disconnected (isolated) genes, or searching for the location of genes with new putative functions related to iron homeostasis. To ensure the diffusion of our data, as well as their subsequent exploitation by ourselves and by others, we developed an interactive web viewer (called 'iHKG viewer'). This website is publicly accessible (see 'Materials and Methods' section). Several features were implemented to (i) enable dynamic exploration of each functional network of co-expressed genes, (ii) enable rapid location of any genes in the network and (iii) obtain for any gene, all the available annotations available in the public databases GRYC and CGD (Figure 6).

Biological relevance of iron homeostasis key genes

Accordance with knowledge of iron homeostasis in Candida glabrata. The biological relevance of our results was manually verified based on a recent review of the regulation of iron homeostasis in *C. glabrata* (22). The genes cited in this article for their role in iron homeostasis were searched in our functional networks of co-expressed genes. As expected, we found orthologs of genes described in *S. cerevisiae* as involved in (i) iron-dependant cellular functions

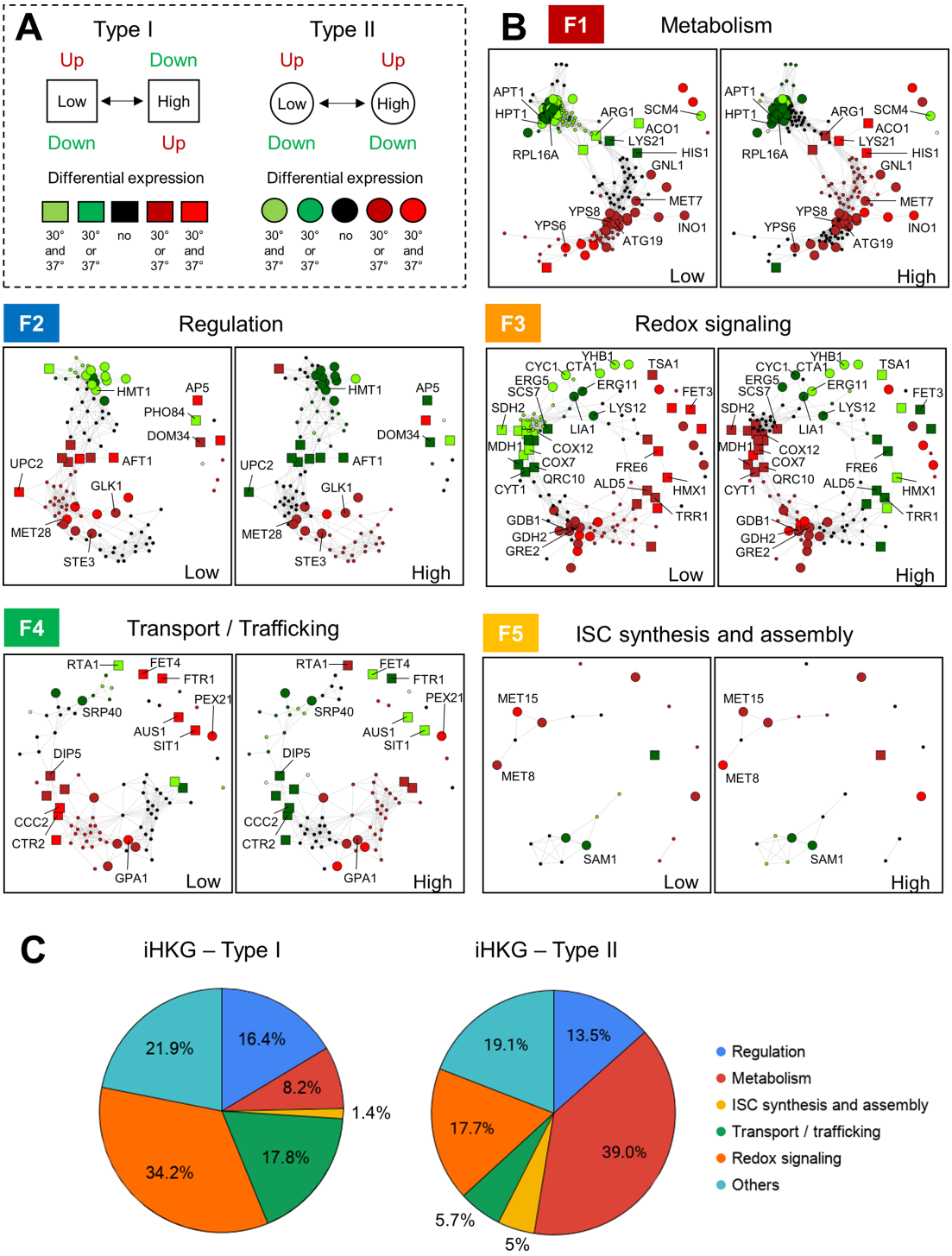

Figure 5. Functional networks of co-expressed genes derived from iron responsive genes. (**A**) Color coding and graphical style used to represent the vertices in the co-expression graphs shown in (**B**). Type I genes are represented by squares, while type II genes are represented by circles. Color is based on the deregulation status of the genes: upregulation in red and downregulation in green. (**B**) Co-expression graphs obtained based on distance calculations between gene expression profiles (C_1–C_4, Figure 4 and 'Materials and Methods' section). Genes were separated here according to their assignment into the general function 'Metabolism' (F1, 176 genes), 'Regulation' (F2, 106 genes), 'Redox signaling' (F3, 118 genes), 'Transport trafficking' (F4, 84 genes) and 'ISC synthesis and assembly' (F5, 22 genes). These two-faced functional networks highlight iHKG in the pathogenic yeast *C. glabrata.* An emblematic illustration of the typical homeostatic role played by iHKG is the co-expression graphs obtained for the function 'Transport/trafficking', in which most of the genes match the definition of iHKG of Type I. Gene names were placed on the graphs, when they were available, according to the CGD database. (**C**) Allocation of Type I and Type II genes in general functions defined in this work, with the percentage of genes assigned to each function, considering iHKG of Type I (left) or type II (right).

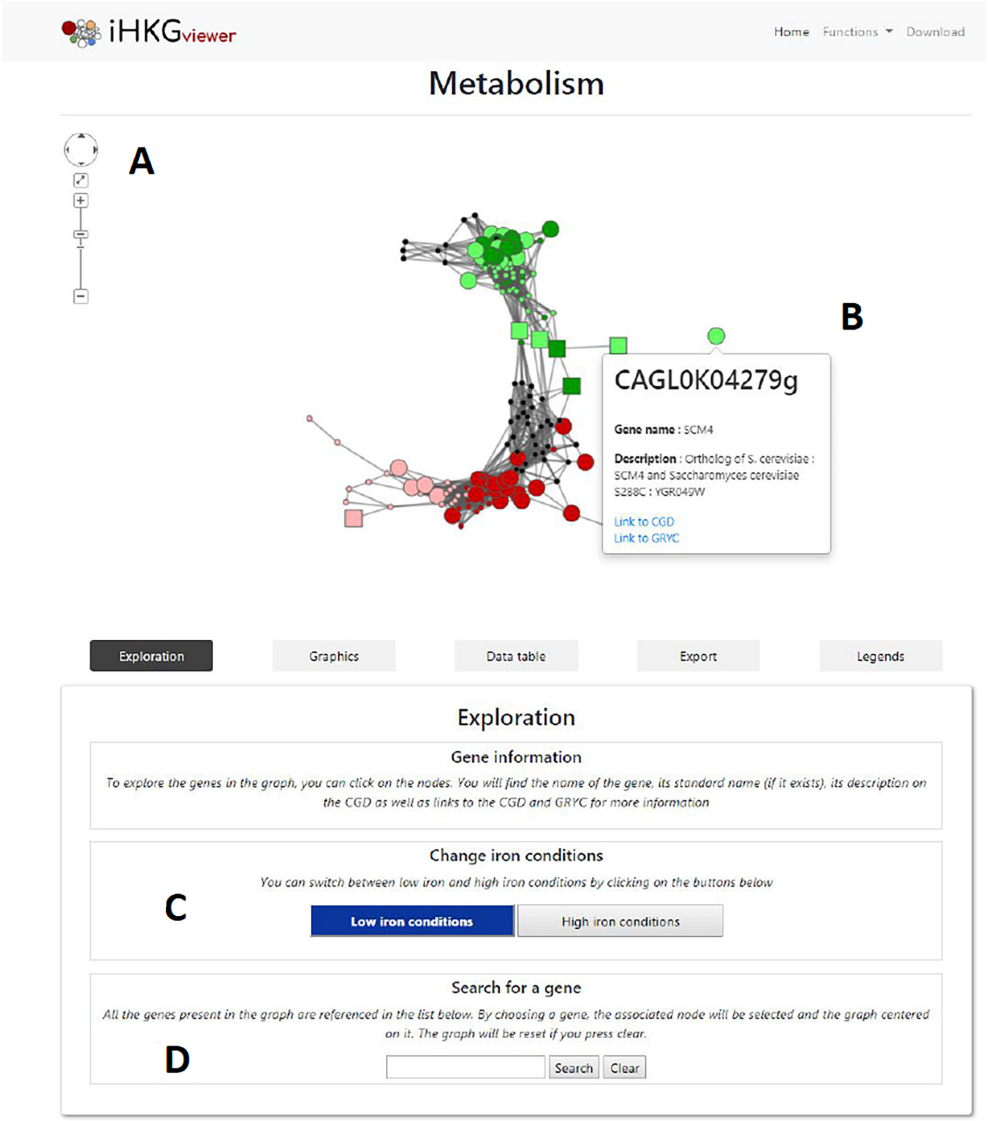

Figure 6. Screenshot of the iHKG viewer. Several functionalities were implemented to easily explore the functional networks of co-expressed genes described in this article: (**A**) Possibility to zoom in or out the graph, (**B**) Possibility to click a node with the mouse and obtain gene name, description and direct web links to CGD and GRYC databases. (**C**) Possibility to switch between low iron condition and high iron condition and (**D**) Possibility to search for a particular gene in the network. The functional network 'Metabolism' is shown here as an illustration.

such as respiration, e.g. *QCR2, QCR6, QCR7, QCR10, COX4, COX5B, COX6, COX7, COX9, COX12, COX15* (all in the functional network 'Redox Signaling'), *ACO1* (functional network 'Metabolism') or COX23 (functional network 'Unclassified'), (ii) genes that encode metalloproteins, e.g. *SDH2, CCP1, RIP1, CYT1*, *LIA1, CYC1, GLT1*, *YHB1* (all in the functional network 'Redox Signaling'), *RLI1* (functional network 'Regulation'), *ILV3* (functional network 'Metabolism') or (iii) genes involved in autophagy (general or mitochondria), e.g. *ATG19*, *ATG32*, *ATG41* (functional network 'Metabolism'). Genes involved in Fe-S clusters or heme biosynthesis and assembly were spread into the functional networks 'ISC synthesis and assembly' (e.g. *ISA1, CFD1*), 'Regulation' (e.g. *GRX4*), 'Metabolism' (e.g. *HEM4*, *HEM15*), 'Redox Signaling' (e.g. *HEM13*, *COX15*). This splitting between different networks arises from our decision to assign genes to only one functional category. Genes encoding proteins with functions directly or indirectly related to redox mechanisms were thus separated from the metabolic pathway in which they participate (as illustrated here with the heme biosynthetic pathway). The reader who is disturbed by this separation of functions can refer to the global network accessible here https://thomasdenecker.github.io/iHKG/all.html. As expected also, the genes *AFT1* (CAGL0H03487g) and *YAP5* (CAGL0K08756g), which encode transcriptional factors, were found in the functional network ('Regulation') along with CAGL0E01243g (the ortholog of *S. cerevisiae* genes *CTH1*/*CTH2*, which encode RNA binding proteins that promote degradation of mRNA of iron-associated genes). Also, we noticed (i) in the functional network 'Redox Signaling' the gene CAGL0A03905g (the ortholog of

S. cerevisiae gene *HMX1*), which encodes a heme oxygenase and thus allows iron recycling from heme and the gene *ERG11* (CAGL0E04334g), which encodes a heme-binding protein required for sterol biosynthesis, (ii) in the functional network 'Metabolism' the gene *DOM34* (CAGL0B04675g), which encodes a ribosome recycling factor, and (iii) in the functional network 'Others' the gene CAGL0G06798, the ortholog of *S. cerevisiae* gene *LSO1* for which the function in still undefined in *S. cerevisiae*. At last, we found genes involved in the three main routes for iron uptake in yeasts, i.e. the high-affinity system, e.g. *FTR1* (CAGL0I06743g), *FET3* (CAGL0F06413g), *CCC2* (CAGL0M08602g), the low-affinity system, e.g. *FET4* (CAGL0F00187g) and the uptake mediated by the capture of xeno-siderophores, e.g. CAGL0E04092g (the ortholog of the *S. cerevisiae* gene *ARN1*). These genes were all in the functional network 'Transport/trafficking'. The gene CAGL0C03333g (the ortholog of *S. cerevisiae* genes *FRE6*/*FRE4*), which is required for the appropriate functioning of the high-affinity uptake system, was found in the functional network 'Redox signaling'. It encodes a protein with a ferric reductase activity that allows an extracellular reduction of ferric iron (Fe3+) to ferrous iron (Fe2+). Genes involved in the copper transfer toward the protein Fet, e.g. the gene CAGL0J07980g (ortholog of *S. cerevisiae* gene *ATX1*), CAGL0M08602g (ortholog of the *S. cerevisiae* gene *CCC2*), CAGL0D04708g (ortholog of the *S. cerevisiae* gene *CTR1*) and CAGL0I02508g (ortholog of the *S. cerevisiae* gene *CTR2*) were also identified in the functional network 'Transport/trafficking', as well as genes coding for actors in the vacuolar or mitochondrial iron export system, e.g. CAGL0H08822g (ortholog of the *S. cerevisiae* gene *MMT1*), *FTH1* (CAGL0M05511g), *SMF3* (CAGL0A03476g).

Gene annotations in public databases. To further explore our list of iron responsive genes, independently of the literature, we used gene annotations available in public databases. We could thus identify a set of genes which (i) were classified as iHKG and (ii) were given a 'Standard Name' (for example, *TRR1* or *TSA1*). This means that the functional information available in CGD or GRYC for these genes was assigned on the basis of genetic, biochemical, or molecular characterization (standard gene names are optional, and genes that are completely uncharacterized generally only have systematic names, for example CAGL0H09592g or CAGL0I00286g). We thus found 11 genes with oxidoreductase activity (i.e. *ALD5*, *CTA1*, *ERG5*, *GDH2*, *GRE2*, *LYS12*, *MDH1*, *MET8*, *SCS7*, *TRR1* and *TSA1*), nine genes with transferase activity (i.e. *APT1*, *GDB1*, *GLK1*, *HIS1*, *HMT1*, *HPT1*, *LYS21*, *MET15* and *SAM1*) and four genes with transporter activity (i.e. *AUS1*, *DIP5*, *GAP1* and *PHO84*). Several genes are also involved in response to stress (i.e. *CTA1*, *ERG5*, *HSP12*, *PHO84*, *RSB1*, *RTA1*, *TRR1*, *TSA1* and *UPC2*), filamentous growth (i.e. *ERG5* and *PHO84*), cellular respiration (i.e. *MDH1*) or cellular homeostasis (i.e. *GLK1*, *TSA1*). They were all noted on the functional networks presented in Figure 5.

Potential specificities of Type I and Type II gene functions. To conclude our description of iHKG and associated functional networks, we examined the repartition of iHKG of Type I and Type II in each functional network. Results are presented in Figure 5C. We observed that the allocation of iHKG greatly differ between Type I and Type II. Differences were particularly important for networks 'Metabolism' (8% in Type I compared to 39% in Type II), 'Transport/trafficking' (18% in Type I compared to 6% in Type II) and 'ISC synthesis and assembly' (1.4% in Type I compared to 5% in Type II). To better explain this observation, we searched for significantly enriched GO terms in lists of Type I and Type II genes, respectively, using the GO database sections 'Cellular component' and 'Molecular function' ('Biological Process' was already used to define Meta-GO, Figure 3). Detailed results are available in Supplementary Data S4. In Type I genes, we found significantly enriched the terms 'membrane part' (33% of genes in the list, P-value = 0.0033), 'cell periphery' (30%, P-value = 0.00156) and 'transporter activity' (22%, P-value = 0.00011). This underlines the particular role of Type I genes in controlling the cell's interactions with its environment, in order to adjust the activity of its transporter proteins. This is relevant to the observation that Type I genes are more represented in the network 'Transport/trafficking' than Type II genes. For Type II genes, we rather found significantly enriched the terms 'ribonucleoprotein complex' (21%, P-value = 0.00277), 'cytosol' (17%, P-value = 0.01141), 'ribosome' (13%, P-value = 0.00016) or 'heme binding' (4%, P-value = 0.00076). It is the particular role of Type II genes to protect internal cellular functions which is highlighted, through the use of traditional mechanisms of stress response (ribosomal proteins) and stabilization of key processes that critically depend on iron (heme utilization or the pathway for iron-sulfur cluster synthesis and assembly). Again, this is relevant to the observation that Type II genes are more represented in the networks 'Metabolism' and 'ISC synthesis and assembly' than Type I genes.

To summarize this section, we observed in our functional networks of co-expressed genes, many genes already known for their role in iron homeostasis in *C. glabrata*. Such an observation gives credence to the biological interest of the transcriptomics dataset we produced in this study. We also described genes for which annotations were available in public databases and whose functions remain coherent with iron homeostasis processes in *C. glabrata*. At last, we observed interesting specificities for the functions supported by iHKG respectively of Type I and Type II. Of course, these observations merit further experiments. But the important point is that we provide with this work an original source of transcriptomics data, which can be interactively and comprehensively explored at any time by anyone, in any context related to iron homeostasis studies.

DISCUSSION

In this study, our goal was to investigate the mechanisms underlying iron homeostasis in the pathogenic yeast *C. glabrata*. In particular, we wanted to identify the 'key genes' whose role is to counterbalance the consequences for the cell not only of iron deficiency, but also of iron overload. In the literature the word 'homeostasis' is found in association with multiple cellular processes, such as 'energy homeosta-

sis' (40), 'pH homeostasis' (41), 'cellular redox homeostasis' (42) or 'cell size homeostasis' (43). In each case, homeostasis refers to mechanisms of return to equilibrium, necessary for the cell to maintain its biological functions in a state compatible with its survival. By definition, homeostasis is thus a dynamic process, which is reversible and requires multiple degrees of freedom to be effective in a panel of different situations. In this context, we believe that the iHKG identified in this study are of particular interest as fundamental actors of iron homeostasis *per se*.

In order to capture iHKG, we produced an original set of microarray experiments in which 'low' and 'high' iron conditions were faced, at two different temperatures (Figure 2). The microarrays we used allowed us to monitor more than 98% of 'verified' and 'uncharacterized' open reading frames in *C. glabrata* genome (see Supplementary Note S8). It is a high percentage that guarantees realistic snapshots of the transcriptome changes. In our experiments, we wanted to mimic the host environment of *C. glabrata* regarding iron sources. We thus used ferrous iron ($FeSO_4$) instead of ferric iron ($FeCl_3$) since in the human body, Fe^{3+} is very scarce or not available because it is rapidly reduced to Fe^{2+} or bound to proteins like transferrin, or ferritin. Fe^{2+} is therefore the major available form of iron for *C. glabrata* cells, both in extra (digestive or urogenital epithelia) and intracellular (macrophage) compartments. We paid special attention to subject the cells to limited environmental changes, in order not to induce cell death or irreversible changes in cell physiology. The concentrations of BPS and $FeSO_4$ used to generated respectively 'low' and 'high' iron conditions remain are equivalent or slighter than those applied in previous studies (see Supplementary Data S5). Also, the same yeast strain was systematically used, as well as the same protocols for RNA preparation, slide hybridization, raw data normalization, etc. This is an important added value for subsequent data analyses, greatly limiting potential sources of experimental noise. At last, we performed the experiments at two different temperatures, i.e. 30°C and 37°C. It was important to verify that at 37° (the human body temperature), the transcriptomic responses remained consistent with those described at 30°, which is the classical temperature used for *C. glabrata* cell cultures in laboratories. To our knowledge, this represents the first transcriptomic dataset published for *C. glabrata* that allows gene expression comparison between 30° and 37°. Even if we observed that the temperature did not have a critical effect on the transcriptomic response which arises either in low or in high iron conditions, we still observed several hundred genes with differential expression between the two conditions (Supplementary Data S1). These results will be a valuable resource to control in any transcriptomics project that genes identified based on laboratory experiments performed at 30°C, are not completely de-regulated at 37°C.

In this work, our mining of the data relies greatly on the representation of the *C. glabrata* genes in functional networks of co-expressed genes (Figures 4 and 5). To infer these networks, we combined a large amount of information, both numerical (measurements of gene expression) and descriptive (GO terms, Types I or II classifications, gene names, etc.). We divided the global network of co-expressed genes (which initially comprised the 636 iron responsive genes) into six functional sub-networks, referred to as 'Metabolism', 'Regulation', 'Redox Signaling', 'Transport trafficking', 'ISC synthesis and assembly' and 'Others' (https://thomasdenecker.github.io/iHKG/). Our aims were (i) to limit the number of genes in each network, thus improving the visualization and exploration of graphs and (ii) to provide a global picture of the cellular functions in which substantial gene expression rewiring was observed. Of course, other choices for gene assignment in subnetworks could have been made (see Supplementary Note S9), but they are, in any case, totally disconnected from the definition and the identification of the iHKG.

We identified with this work a list of 214 genes as good candidates for being key elements of iron homeostasis (iHKG of Type I and II, Supplementary Data S2). Because they rely on stringent selection, iHKG can be considered trustworthy information, which is, notably, independent of existing knowledge obtained in the model species *S. cerevisiae* and *C. albicans*. We present in Supplementary Note S7 examples of interesting biological insights that arise from our observations: (i) We propose a shortlist of 27 genes that did not appear in previous emblematic publications in the field (15,22). For 10 of them, we observed interesting variations in expression levels reported in recent multi-omics datasets (RNAseq and mass spectrometry technologies). (ii) We detected differentiated regulatory mechanisms underlying transcription of the iHKG of Type I and Type II. In Type II genes, we found as enriched DNA sequence in promoters, the motif AGGG. It is the stress response element (STRE), i.e. the DNA binding site of transcription factors Msn2 and Msn4 (44). Such observation is relevant with the manner Type II genes were defined in this study (Figure 1) and our anticipation they could include genes involved in the general stress response of cells (ESR genes for instance (39)). In Type I genes, we found as enriched motif, the sequence TGCACCC. It corresponds to the DNA binding site of the transcription factor Aft1 (45), one of the main regulators of iron homeostasis (15). Type I genes are very strongly linked to the Aft1 regulon in *C. glabrata* and certainly includes new Aft1 targets so far not described in the literature (e.g. CAGL0A01199g or CAGL0K06259g, Supplementary Note S7). (*iii*) We noticed in our network of co-expressed genes (function 'Regulation') that the gene coding the transcription factor Hap1 (CAGL0B03421g) was in the immediate neighbourhood of the transcription factor Aft1 (Supplementary Figures S10). Hap1 is the heme activated transcription factor. It is thus not surprising to find the gene *HAP1* in our list of iHKG, because heme biosynthesis requires iron. Hap1 is thus expected to be an indirect iron sensor and its activity is known to be affected by iron deficiency (46). Although the main role of Hap1 is globally conserved in yeast species, a dual role may exist as described for *Kluyveromyces lactis* (47). Our results show for the first time that, in *C. glabrata*, *HAP1* is also deregulated in excess iron conditions. At last, even if we observed the global responses at 30°C and 37°C are similar, we could see several iKHG that were more de-regulated at 37°C, which is the human host temperature. Among them, we found *AWP2* (CAGL0K00110g) and *ERG5* (CAGL0M07656g). We provide here interesting transcriptomic data showing that they may be essential for the yeast homeostasis in iron varying

conditions, and hence for the *C. glabrata* adaptation in the human body compartments.

To conclude, the data we provide in this work is full of new research perspectives. It can help to improve the functional annotation of *C. glabrata* genes (only 5% of genes are 'verified' in the CGD database) and increase our understanding of the specificities of this human pathogen. In that respect, we found in our data several genes for which no clear orthologous genes could be identified in *S. cerevisiae* and *C. albicans* (Supplementary Data S6). A natural further direction would be to search for genomic mutations in these genes, using *C. glabrata* strains isolated from patients (48). The rapid increase in genomic sequence availability together with the decreasing cost for deep sequencing represent unprecedented opportunities for population genomics studies. We finally hope our web server will be valuable resources for such exciting analyses, particularly in perspective of medical applications.

DATA AVAILABILITY

Microarray raw data files are accessible through GEO Series accession number GSE139363 (https://www.ncbi.nlm.nih.gov/geo/). Source code and result data files can be found here: https://github.com/thomasdenecker/iHKG/. The website to explore the functional networks of co-expressed genes is available here: https://thomasdenecker.github.io/iHKG/.

GEO Series accession number is GSE139363.

SUPPLEMENTARY DATA

Supplementary Data are available at NARGAB Online.

ACKNOWLEDGEMENTS

G.L. would like to thank Linda Sperling for careful reading of the manuscript and comments, the research team 'Mitochondria, Metals and Oxidative Stress' (Jacques Monod Institute, Paris), the research team 'Fungal Epigenomics and Development' (I2BC, Orsay) and the partners from ANR CANDIHUB for helpful discussions.

FUNDING

Ecole doctorale de l'Université Paris-Saclay (SDSV http://www.ed-sdsv.u-psud.fr/); Agence Nationale pour la Recherche (CANDIHUB project, Grant Number ANR–14–CE14–0018–02).

Conflict of interest statement. None declared.

REFERENCES

1. Pfaller,M.A. and Diekema,D.J. (2007) Epidemiology of invasive candidiasis: a persistent public health problem. *Clin. Microbiol. Rev.*, **20**, 133–163.
2. Goemaere,B., Lagrou,K., Spriet,I., Hendrickx,M. and Becker,P. (2018) Clonal spread of candida glabrata bloodstream isolates and fluconazole resistance affected by prolonged exposure: a 12-year single-center study in Belgium. *Antimicrob. Agents Chemother.*, **62**, e00591-18.
3. Epelbaum,O. and Chasan,R. (2017) Candidemia in the intensive care unit. *Clin. Chest Med.*, **38**, 493–509.
4. Dujon,B., Sherman,D., Fischer,G., Durrens,P., Casaregola,S., Lafontaine,I., Montigny,J. de, Marck,C., Neuvéglise,C., Talla,E. *et al.* (2004) Genome evolution in yeasts. *Nature*, **430**, 35–44.
5. Pfaller,M.A., Castanheira,M., Lockhart,S.R., Ahlquist,A.M., Messer,S.A. and Jones,R.N. (2012) Frequency of decreased susceptibility and resistance to echinocandins among fluconazole-resistant bloodstream isolates of candida glabrata. *J. Clin. Microbiol.*, **50**, 1199–1203.
6. Pham,C.D., Iqbal,N., Bolden,C.B., Kuykendall,R.J., Harrison,L.H., Farley,M.M., Schaffner,W., Beldavs,Z.G., Chiller,T.M., Park,B.J. *et al.* (2014) Role of FKS Mutations in Candida glabrata: MIC Values, Echinocandin Resistance, and Multidrug Resistance. *Antimicrob. Agents Chemother.*, **58**, 4690–4696.
7. Vallabhaneni,S., Cleveland,A.A., Farley,M.M., Harrison,L.H., Schaffner,W., Beldavs,Z.G., Derado,G., Pham,C.D., Lockhart,S.R. and Smith,R.M. (2015) Epidemiology and Risk Factors for Echinocandin Nonsusceptible Candida glabrata Bloodstream Infections: Data From a Large Multisite Population-Based Candidemia Surveillance Program, 2008–2014. *Open Forum Infect. Dis.*, **2**, ofv163.
8. Brunke,S. and Hube,B. (2013) Two unlike cousins: Candida albicans and C. glabrata infection strategies. *Cell. Microbiol.*, **15**, 701–708.
9. Sutak,R., Lesuisse,E., Tachezy,J. and Richardson,D.R. (2008) Crusade for iron: iron uptake in unicellular eukaryotes and its significance for virulence. *Trends Microbiol.*, **16**, 261–268.
10. Wang,J. and Pantopoulos,K. (2011) Regulation of cellular iron metabolism. *Biochem. J.*, **434**, 365–381.
11. Nakamura,T., Naguro,I. and Ichijo,H. (2019) Iron homeostasis and iron-regulated ROS in cell death, senescence and human diseases. *Biochim. Biophys. Acta Gen. Subj.*, **1863**, 1398–1409.
12. Bairwa,G., Jung,W.H. and Kronstad,J.W. (2017) Iron acquisition in fungal pathogens of humans. *Metallomics*, **9**, 215–227.
13. Gerwien,F., Skrahina,V., Kasper,L., Hube,B. and Brunke,S. (2018) Metals in fungal virulence. *FEMS Microbiol. Rev.*,**42**, doi:10.1093/femsre/fux050.
14. Thiébaut,A., Delaveau,T., Benchouaia,M., Boeri,J., Garcia,M., Lelandais,G. and Devaux,F. (2017) The CCAAT-binding complex controls respiratory gene expression and iron homeostasis in Candida Glabrata. *Sci. Rep.*, **7**, 1–10.
15. Gerwien,F., Safyan,A., Wisgott,S., Hille,F., Kaemmer,P., Linde,J., Brunke,S., Kasper,L. and Hube,B. (2016) A novel hybrid iron regulation network combines features from pathogenic and nonpathogenic yeasts. *Mbio*,**7**, e01782-16.
16. Gerwien,F., Safyan,A., Wisgott,S., Brunke,S., Kasper,L. and Hube,B. (2017) The fungal pathogen Candida glabrata does not depend on surface ferric reductases for iron acquisition. *Front. Microbiol.*, **8**, 1055.
17. Sharma,V., Purushotham,R. and Kaur,R. (2016) The Phosphoinositide 3-Kinase Regulates Retrograde Trafficking of the Iron Permease CgFtr1 and Iron Homeostasis in Candida glabrata. *J. Biol. Chem.*, **291**, 24715–24734.
18. Srivastava,V.K., Suneetha,K.J. and Kaur,R. (2015) The mitogen-activated protein kinase CgHog1 is required for iron homeostasis, adherence and virulence in Candida glabrata. *FEBS J.*, **282**, 2142–2166.
19. Srivastava,V.K., Suneetha,K.J. and Kaur,R. (2014) A systematic analysis reveals an essential role for high-affinity iron uptake system, haemolysin and CFEM domain-containing protein in iron homoeostasis and virulence in Candida glabrata. *Biochem. J.*, **463**, 103–114.
20. Seider,K., Gerwien,F., Kasper,L., Allert,S., Brunke,S., Jablonowski,N., Schwarzmüller,T., Barz,D., Rupp,S., Kuchler,K. *et al.* (2014) Immune evasion, stress resistance, and efficient nutrient acquisition are crucial for intracellular survival of Candida glabrata within macrophages. *Eukaryot. Cell*, **13**, 170–183.
21. Hosogaya,N., Miyazaki,T., Nagi,M., Tanabe,K., Minematsu,A., Nagayoshi,Y., Yamauchi,S., Nakamura,S., Imamura,Y., Izumikawa,K. *et al.* (2013) The heme-binding protein Dap1 links iron homeostasis to azole resistance via the P450 protein Erg11 in Candida glabrata. *FEMS Yeast Res.*, **13**, 411–421.
22. Devaux,F. and Thiébaut,A. (2019) The regulation of iron homeostasis in the fungal human pathogen Candida glabrata. *Microbiol. Read. Engl.*, **165**, 1041–1060.

23. Cassat,J.E. and Skaar,E.P. (2013) Iron in Infection and Immunity. *Cell Host Microbe*, **13**, 509–519.
24. Nairz,M., Schroll,A., Sonnweber,T. and Weiss,G. (2010) The struggle for iron—a metal at the host-pathogen interface: Iron at the host-pathogen interface. *Cell. Microbiol.*, **12**, 1691–1702.
25. Yilmaz,B. and Li,H. (2018) Gut microbiota and iron: the crucial actors in health and disease. *Pharmaceuticals*, **11**, E98.
26. Pfrimer,K., Micheletto,R.F., Marchini,J.S., Padovan,G.J., Moriguti,J.C. and Ferriolli,E. (2014) Impact of aging on urinary excretion of iron and zinc. *Nutr. Metab. Insights*, **7**, 47–50.
27. van Raaij,S.E.G., Srai,S.K.S., Swinkels,D.W. and van Swelm,R.P.L. (2019) Iron uptake by ZIP8 and ZIP14 in human proximal tubular epithelial cells. *Biometals*, **32**, 211–226.
28. Abreu,R., Quinn,F. and Giri,P.K. (2018) Role of the hepcidin-ferroportin axis in pathogen-mediated intracellular iron sequestration in human phagocytic cells. *Blood Adv.*, **2**, 1089–1100.
29. Gabaldón,T., Martin,T., Marcet-Houben,M., Durrens,P., Bolotin-Fukuhara,M., Lespinet,O., Arnaise,S., Boisnard,S., Aguileta,G., Atanasova,R. *et al.* (2013) Comparative genomics of emerging pathogens in the Candida glabrata clade. *BMC Genomics*, **14**, 623.
30. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res.*, **30**, 207–210.
31. Marisa,L., Ichanté,J.-L., Reymond,N., Aggerbeck,L., Delacroix,H. and Mucchielli-Giorgi,M.-H. (2007) MAnGO: an interactive R-based tool for two-colour microarray analysis. *Bioinform. Oxf. Engl.*, **23**, 2339–2341.
32. Ritchie,M.E., Phipson,B., Wu,D., Hu,Y., Law,C.W., Shi,W. and Smyth,G.K. (2015) limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.*, **43**, e47.
33. Ringnér,M. (2008) What is principal component analysis? *Nat. Biotechnol.*, **26**, 303–304.
34. Falissard,B. (1996) A spherical representation of a correlation matrix. *J. Classif.*, **13**, 267–280.
35. Ashburner,M., Ball,C.A., Blake,J.A., Botstein,D., Butler,H., Cherry,J.M., Davis,A.P., Dolinski,K., Dwight,S.S., Eppig,J.T. *et al.* (2000) Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat. Genet.*, **25**, 25–29.
36. Shannon,P., Markiel,A., Ozier,O., Baliga,N.S., Wang,J.T., Ramage,D., Amin,N., Schwikowski,B. and Ideker,T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.*, **13**, 2498–2504.
37. Skrzypek,M.S., Binkley,J., Binkley,G., Miyasato,S.R., Simison,M. and Sherlock,G. (2017) The Candida Genome Database (CGD): incorporation of Assembly 22, systematic identifiers and visualization of high throughput sequencing data. *Nucleic Acids Res.*, **45**, D592–D596.
38. Boyle,E.I., Weng,S., Gollub,J., Jin,H., Botstein,D., Cherry,J.M. and Sherlock,G. (2004) GO::TermFinder–open source software for accessing Gene Ontology information and finding significantly enriched Gene Ontology terms associated with a list of genes. *Bioinform. Oxf. Engl.*, **20**, 3710–3715.
39. Gasch,A.P., Spellman,P.T., Kao,C.M., Carmel-Harel,O., Eisen,M.B., Storz,G., Botstein,D. and Brown,P.O. (2000) Genomic expression programs in the response of yeast cells to environmental changes. *Mol. Biol. Cell*, **11**, 4241–4257.
40. Zhang,J., Vemuri,G. and Nielsen,J. (2010) Systems biology of energy homeostasis in yeast. *Curr. Opin. Microbiol.*, **13**, 382–388.
41. Eskes,E., Deprez,M.-A., Wilms,T. and Winderickx,J. (2018) pH homeostasis in yeast; the phosphate perspective. *Curr. Genet.*, **64**, 155–161.
42. Ayer,A., Gourlay,C.W. and Dawes,I.W. (2014) Cellular redox homeostasis, reactive oxygen species and replicative ageing in Saccharomyces cerevisiae. *FEMS Yeast Res.*, **14**, 60–72.
43. Millar,J.B.A. (2002) A genomic approach to studying cell-size homeostasis in yeast. *Genome Biol.*, **3**, REVIEWS1028.
44. Martínez-Pastor,MT., Marchler,G., Schüller,C., Marchler-Bauer,A., Ruis,H. and Estruch,F. (1996) The Saccharomyces cerevisiae zinc finger proteins Msn2p and Msn4p are required for transcriptional induction through the stress response element (STRE). *EMBO J.*, **15**, 2227–2235
45. Conde e Silva,N., Gonçalves,I.R., Lemaire,M., Lesuisse,E., Camadro,J.M. and Blaiseau,P.L. (2009) KlAft, the Kluyveromyces lactis ortholog of Aft1 and Aft2, mediates activation of iron-responsive transcription through the PuCACCC Aft-type sequence. *Genetics.*, **183**, 93–106.
46. Ihrig,J., Hausmann,A., Hain,A., Richter,N., Hamza,I., Lill,R. and Mühlenhoff,U. (2010) Iron regulation through the back door: iron-dependent metabolite levels contribute to transcriptional adaptation to iron deprivation in Saccharomyces cerevisiae. *Eukaryot. Cell*, **9**, 460–471.
47. Bao,W.G., Guiard,B., Fang,Z.A., Donnini,C., Gervais,M., Passos,F.M., Ferrero,I. and Fukuhara,H. (2008) Oxygen-dependent transcriptional regulator Hap1p limits glucose uptake by repressing the expression of the major glucose transporter gene RAG1 in Kluyveromyces lactis. *Eukaryot. Cell.*, **7**, 1895–1905.
48. Carreté,L., Ksiezopolska,E., Gómez-Molero,E., Angoulvant,A., Bader,O., Fairhead,C. and Gabaldón,T. (2019) Genome comparisons of Candida glabrata serial clinical isolates reveal patterns of genetic variation in infecting clonal populations. *Front. Microbiol.*, **10**, 112.