The core meiotic transcriptome in budding yeasts

Michael Primig[1,4]*, Roy M. Williams[1,2]*, Elizabeth A. Winzeler[2], Gela G. Tevzadze[1], Andrew R. Conway[3], Seung Y. Hwang[2], Ronald W. Davis[2] & Rochelle Easton Esposito[1]

*These authors contributed equally to this work.

We used high-density oligonucleotide microarrays to analyse the genomes and meiotic expression patterns of two yeast strains, SK1 and W303, that display distinct kinetics and efficiencies of sporulation. Hybridization of genomic DNA to arrays revealed numerous gene deletions and polymorphisms in both backgrounds. The expression analysis yielded approximately 1,600 meiotically regulated genes in each strain, with a core set of approximately 60% displaying similar patterns in both strains. Most of these (95%) are *MAT*a/*MAT*α-dependent and are not similarly expressed in near-isogenic meiosis-deficient controls. The transcript profiles correlate with the distribution of defined meiotic promoter elements and with the time of known gene function.

Introduction

Meiosis is a key feature of eukaryotic sexual reproduction. Diploid parental cells replicate their DNA once and then undergo two rounds of chromosome segregation to produce haploid daughter cells. Gametogenesis in budding yeast has been a useful model for uncovering the genetic architecture of the meiotic process[1]. The major aspects of chromosome behaviour during meiosis are conserved from yeast to mammals and often use homologous gene functions[2–5]. Initiation and completion of meiosis in yeast depends on heterozygosity at the mating-type locus and deprivation of nitrogen and fermentable carbon sources. Classical approaches have thus far identified approximately 200 genes required for meiotic development. Transcriptional analyses of approximately 40 of these have defined several expression groups: very early, early, middle, mid-late and late, with complex induction and re-repression profiles[1,6]. Three major transcription factors (Ume6p/Ime1p (refs 7–11), Abf1p (refs 12,13) and Ndt80p (refs 14,15)) are known to activate the meiotic program. mRNA stability also varies during development, with early meiotic transcripts being highly unstable relative to those expressed later[16]. Several hundred additional induced genes (many of which contain promoter elements responsive to the factors noted above[17]) revealed by recent PCR microarray expression analysis generally support these prior group designations.

We present here a comparative genomic and meiotic transcriptome analysis of two genetically distinct yeast strains often used in meiotic studies (along with non-sporulating starvation controls) to determine the complete profile of regulated transcripts related to gametogenesis. One strain, W303 (closely related to the prototypical S288C whose genome was sequenced), is widely used in cell-cycle and other studies of vegetatively growing cells. The other, SK1 (genetically distant from W303), yields poor spore viability, often reflective of genome rearrangements in outcrosses to most other laboratory strains. Although both form asci with relatively high efficiency, SK1 sporulates faster, more synchronously and to higher levels than W303, and hence allows greater resolution of the timing of specific events in meiosis.

Our studies define a common set of meiotically regulated genes in both strains, identify genes displaying strain-specific meiotic expression patterns (which may account in part for their different sporulation properties) and distinguish genes involved in the response to nutrient deprivation from those specifically required for sporulation. The set of meiotically regulated genes described here contains most previously identified sporulation genes[1] and many hundreds of additional loci not found in an independent whole-genome expression study[17]. These data should facilitate the identification of all essential meiosis-specific components of this critical developmental pathway. Web sites in the United States (http://re-esposito.bsd.uchicago.edu and http://171.65.26.27:592/meiosis_2/) and Europe (http://www.igh.cnrs.fr/equip/primig-meiosis) allow downloading of raw data and access to a searchable, cross-referenced database.

Results

Genomic variation between SK1 and W303

An overview of the genetic differences between SK1 and W303 was determined by hybridizing genomic DNA to oligonucleotide microarrays made on the basis of the DNA sequence of S288C. The presence of a sequence is reflected by an increase in hybridization to an oligonucleotide that is a perfect match (PM), relative to a control mismatch (MM) sequence, in the array. For gene deletions, the PM signal is similar to the MM signal for most probes representing a gene. Single-nucleotide polymorphisms (SNPs) can be detected by decreased hybridization efficiency to specific PM sequences, depending on where the SNP is in the probe and how well it is accommodated in the duplex[18]. By these criteria, SK1 and W303 are

[1]*The University of Chicago, Department of Molecular Genetics and Cell Biology, Chicago, Illinois, USA.* [2]*Stanford University Medical School, Department of Biochemistry, Stanford, California, USA.* [3]*Silicon Genetics, San Carlos, California, USA.* [4]*Present address: Institut de Génétique Humaine, Montpellier, France. Correspondence should be addressed to R.E.E. (e-mail: re-esposito@uchicago.edu).*

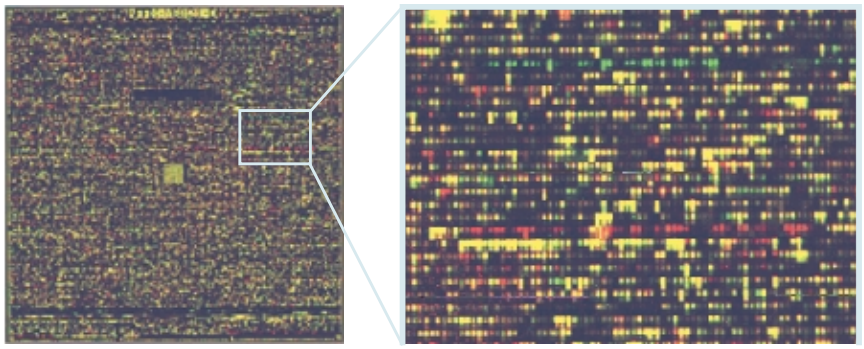

Fig. 1 Genomic variation in SK1 and W303. Genomic DNA prepared from SK1 and W303 was labelled with the fluorophor phycoerythrin and hybridized to high-density oligonucleotide arrays (Ye6100). False-colour images of the fluorescence signal patterns (SK1 in red and W303 in green) were superimposed to visualize genetic differences. A region of the array containing a signal pattern indicative of gene deletions is enlarged and genes deleted in SK1 and W303 indicated. Eight genes deleted from SK1 and meiotically induced in W303 are highlighted in green; one gene deleted from W303 and induced in SK1 is marked in red. Genes deleted in both strains are shown in italics.

SK1**a**/α Δs

YAL065C
YAR047C
YAR050W/FLO1
YAR069C
YAR070C
YAR073W
YBR012W
YCL021W
YDR036C
YDR038C/ENA5
YDR312W/SSF2
YDR544C
YER188W
YGL051W
YGL052W
YGL053W
YGR289C/AGT1
YGR290W
YHR043C/DOG2
YHR211W/FLO5
YIR042C
YIR043C
YJL113W
YJL114W
YLR155C/ASP3-1
YLR156W
YLR158C/ASP3-3
YLR159W
YLR160C/ASP3-4
YLR161W
YLR162W
YNL336W
YOL160W
YOL162W
YOL163W
YOL164W
YPR197C
YPR199C/ARR1
YPR200C

W303**a**/α Δs

YAR073W
YBR012W
YBR048W/RPS18B
YBR084C-a/RPL19B
YDR544C
YMR326C
YNR070W
YNR072W/HXT17

approximately 89% and approximately 93% similar to S288C, respectively (assuming a 100% hybridization efficiency for S288C), with respect to positively hybridizing oligonucleotides in the array (Fig. 1). Relative to S288C, we identified 39 gene deletions and 2025 polymorphisms in SK1, and 8 gene deletions and 318 polymorphisms in W303. The data show that, compared with S288C, SK1 has more sequence variation than W303. The relatively large amount of genetic variation (polymorphisms, deletions) detected in this analysis is a minimum estimate based on genomic DNA hybridization to oligonucleotides representing the ORFs and ignores potential differences in intergenic regions.

Distinct sporulation kinetics and efficiencies of SK1 and W303

Progression through the meiotic landmarks was monitored and shown to be consistent with prior studies[10,19] (Fig. 2). The data highlight the very rapid and efficient sporulation properties of SK1, and demonstrate that in W303 even the earliest of known landmarks (the initiation of DNA replication) is delayed, with completion of other events occurring progressively later. Sporulation-deficient SK1 *MAT*α/*MAT*α and W303 *MAT*a/*MAT*a control strains, which are near isogenic to their *MAT*a/*MAT*α counterparts, fail to enter meiosis (data not shown).

Identification of transcripts regulated during meiosis in SK1 and W303

We sampled sporulating and non-sporulating cultures of both strains throughout meiosis and spore development. Probes were prepared and hybridized to oligonucleotide microarrays[20,21]. We searched the resulting data sets for developmentally regulated genes on the basis of minimum and maximum fluorescence intensity signals measured for each gene over the entire time course. Criteria for inclusion were determined using known meiotic genes and set at a fourfold difference for the more synchronous SK1 strain, and at a 2.5-fold threshold for W303 to compensate for its lower degree of synchrony.

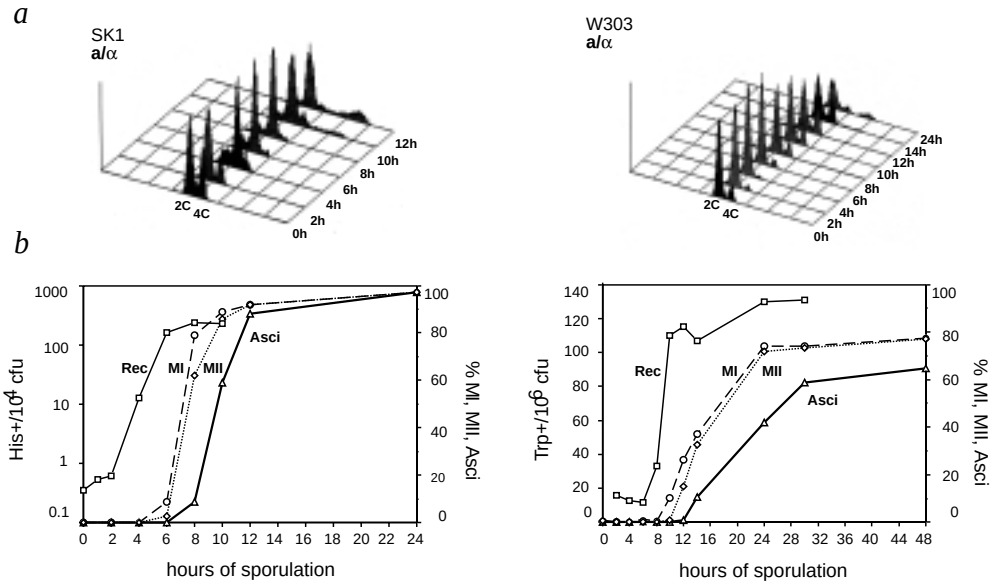

Fig. 2 Meiotic landmarks in SK1 and W303. ***a***, The kinetics of DNA replication monitored by FACS analysis. The DNA content of G1 and G2 cells is indicated as 2C and 4C, respectively. ***b***, Recombination frequencies (squares), and percentages of cells completing meiosis I (binucleates + tri/tetranucleates; circles), meiosis II (tri/tetranucleates; diamonds) and mature ascus formation (triangles) during sporulation. The induction of meiotic recombination in SK1 and W303 is shown on log and linear scales, respectively, reflecting the different frequencies of exchange of the marker systems used in each strain[10,19].

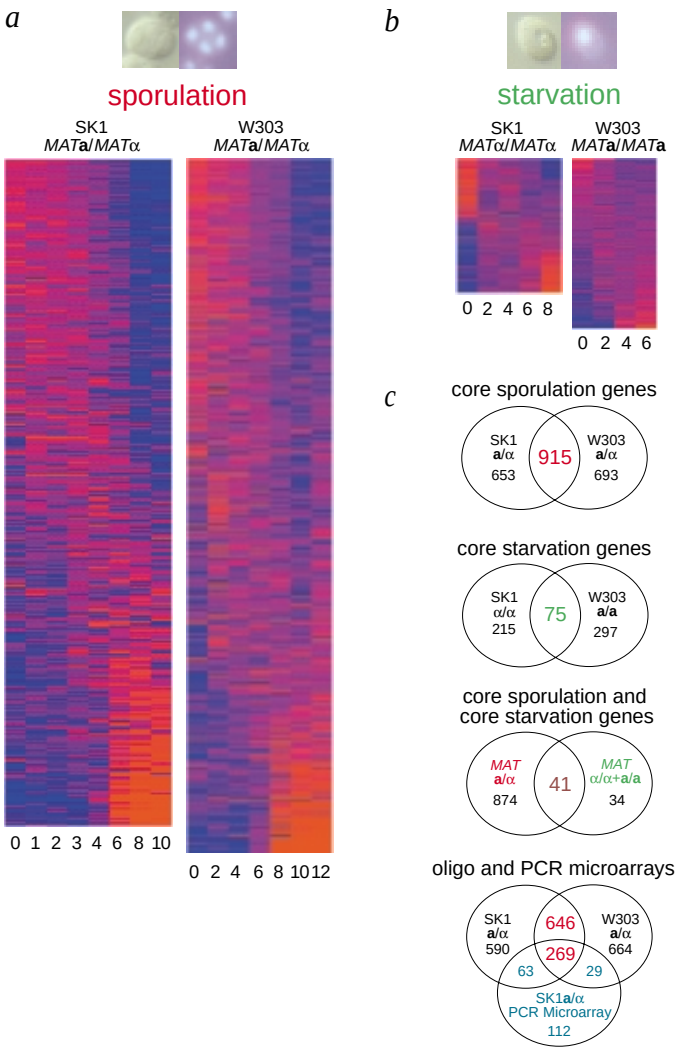

Fig. 3 Regulated transcription in SK1 and W303 during sporulation and starvation. GeneSpring 2.3 was used to visualize the pattern of expression of meiotically regulated genes in SK1 (1568) and W303 (1608), ordered by time of initial induction. Columns represent time during sporulation, shown in hours at the bottom. Horizontal lines represent expression of individual genes (not aligned in the two strains), with red and blue indicating highest and lowest levels of induction, respectively. ***a***, SK1 and W303 *MAT***a**/*MAT*α. ***b***, Starvation controls, SK1 *MAT*α/*MAT*α and W303 *MAT***a**/*MAT***a**. ***c***, Venn diagrams summarizing the numbers of genes regulated in common in each strain during sporulation and starvation, and comparison with previous data[17].

The expression patterns of approximately 1,600 genes varied over the course of the experiment in both strains (Fig. 3). Among the genes detected as transcriptionally regulated during sporulation of SK1 (1568) and W303 (1608), a core set of 915 display differential expression in both strains (Fig. 3*a*,*c*). This group includes 94 genes whose transcripts are detected in only sporulating cells (data not shown). The same analysis for starving cells yields 290 genes in SK1 *MAT*α*MAT*α and 372 genes in W303 *MAT*a/*MAT*a as differentially regulated (Fig. 3*b*). Among these, 75 are expressed similarly in both control strains, with 41 being also upregulated in sporulating *MAT*a/*MAT*α cells (Fig. 3*b*,*c*). Although up to 25% of all yeast genes appear to be meiotically regulated (depending on the genetic background), a smaller subset of approximately 900 core genes displays a strain-independent pattern of meiotic transcriptional regulation. Only a small percentage (5%) of the core meiotic genes (41/915) are induced in both starved, non-sporulating control strains, demonstrating the critical importance of heterozygosity at the *MAT* locus for meiotic gene expression.

We compared these data with those reported for induced genes from PCR microarray analysis[17] (Fig. 3*c*, Venn diagram). Nearly all of the genes detected in the PCR microarray study were upregulated in our analyses, but some at levels below the minimum standard set for inclusion for either SK1 or W303. Approximately 76% (361/473) met the criteria for a meiotically regulated gene, with 57% (269/473) being detected in the core in both strains. In addition, this study identifies approximately 650 previously unreported meiotically induced genes (gene lists from these comparisons are available at our web sites). It should be noted that *IME1*, a key positive regulator of meiosis (upregulated in early stationary phase and meiosis), did not meet the criteria for inclusion in either this study or the PCR microarray analysis[17] because its transcript is already present at high levels at the time of transfer of cells to sporulation medium.

Expression classes during sporulation and starvation

The expression data were next analysed by using pair-wise correlation statistics[22] (clustering) followed by gene-by-gene comparison. First, the 1,568 meiotic expression profiles included in the SK1 data set were correlated and clusters delineated based on the positions of the major branch points in the dendrogram. We then ordered the resulting clusters with respect to initial time of induction. Genes from sporulating W303 and starving controls, corresponding to those present in the SK1 clustergram (which includes the 915 core meiotic genes), were subsequently aligned (Fig. 4, horizontal bars). This comparison reveals a very similar global transcriptional response in both sporulating strains.

We identified seven broad clusters. As expected, many metabolic and stress response genes are induced first (clusters 1 and 2), followed by genes required for DNA replication, recombination and synaptonemal complex formation (3 and 4), for the meiotic divisions (5 and 6) and for ascus formation/germination (7). Within clusters 5, 6 and 7 at least two subclasses (a and b) are evident based on peak levels of expression and other pattern features such as re-repression, which may reflect different promoter strength and/or structure. The timing of transcription and time of function correlate well for meiotic genes that have important roles in different landmark events. Finally, nearly all of the early, middle and mid-late meiotic genes either fail to be induced or display altered expression in starving controls, revealing a global pattern of strain-independent transcriptional upregulation that is largely absent from non-sporulating cells.

Correlation of expression in SK1 and W303

The transcription profiles obtained from SK1 and W303 were compared by calculating the Pearson correlation coefficient for each pair of expression patterns for the approximately 900 core genes (independent of induction time, which is often delayed in W303) and sorted into groups (Fig. 5). Approximately 60% of the genes showed well-correlated expression patterns (displaying a Pearson correlation coefficient of >0.5) in both backgrounds. As this group contains a majority (60%) of genes previously shown to be important in gametogenesis[1], many of the highly correlated uncharacterized genes may have functional significance as well.

We also determined the mean expression patterns (by cluster) for core meiotic genes whose patterns are particularly well correlated (coefficients equal to or higher than 0.8; Fig. 6). Clusters 1, 2

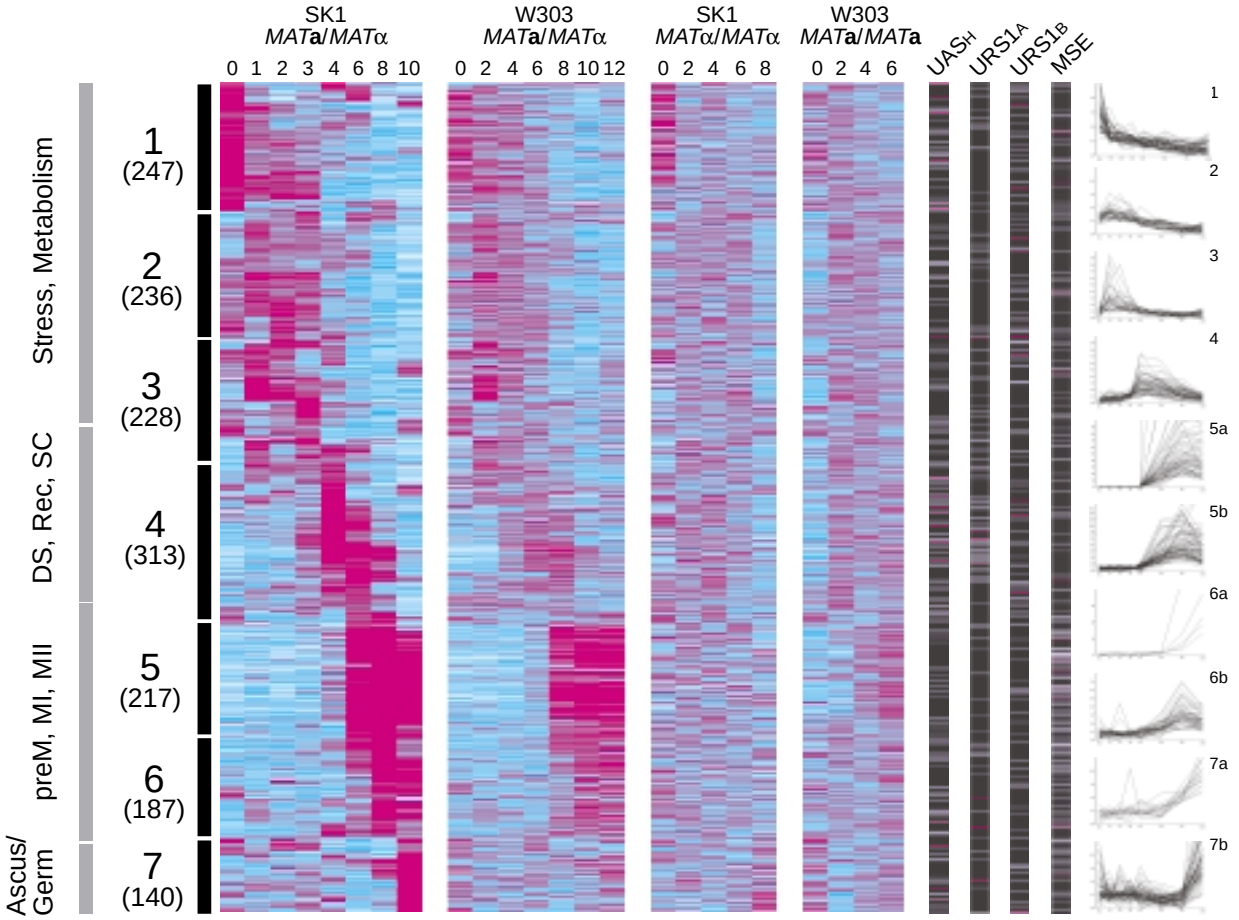

Fig. 4 Clustered patterns of gene expression during sporulation. The 1,568 genes differentially regulated in sporulating SK1 were clustered after pair-wise correlation analysis, and then aligned with the expression data from the same genes in sporulating W303 and starved SK1 *MAT*α/*MAT*α and W303 *MAT***a**/*MAT***a** cells using programs written in Matlab 5.0. The data is organized into vertical clustergrams for each strain. Columns reflect hours of sporulation and horizontal rows, the expression of a particular gene. Each time point is coloured following a scale of signal strength from low (blue) to high (pink). The clustergram is divided into seven clusters based on groups of genes with similar expression patterns (clusters 1–7, as indicated). The functions attributed to the clusters, that is, DNA synthesis (DS), recombination (Rec), synaptonemal complex (SC) and ascus formation/germination (Ascus/Germ), are given as a general guide, and are not meant to represent a set of genes exclusively involved in a response. Horizontal bars (right) represent core genes containing the promoter elements UAS_H (green), $URS1_A$ (red), $URS1_B$ (blue) and MSE (yellow). A graphical view of expression is shown at the far right; for clusters 5–7 the subgroups (a and b) are shown, which display similar patterns, but distinct expression profiles.

and 3 contain numerous stress-response and metabolic genes required for nitrogen uptake, amino acid biosynthesis and mitochondrial function. Cluster 1 genes are also often upregulated in starving cells, whereas clusters 2 and 3, which are induced later, both require *MAT*a/*MAT*α. Cluster 3 is distinguished by somewhat higher and significantly different expression patterns. Cluster 4 contains key known early genes required for pre-meiotic DNA replication (*RFA1*, *RFA2*), recombination (*SPO11*, *REC102*, *REC104*, *REC114*, *RAD52*, *MSH5*, *MEI5*), SC formation (*HOP1*, *HOP2*, *RED1*), sister chromatid and centromere cohesion (*HHT1*, *SCC2*, *SPO13*), and pachytene checkpoint functions (*PCH2*). They have similar induction and re-repression patterns in both sporulating strains, with signals consistently weaker and slightly delayed in W303. Although low-level induction of some of these genes also occurs in non-sporulating control strains, most upregulation is *MAT*a/*MAT*α dependent, consistent with prior reports that some early meiotic genes display low transient transcriptional induction during starvation[23,24].

Clusters 5a and 5b contain two groups induced at the same time (in late pachytene during commitment to meiosis), but display distinct expression levels. The induction of nearly all genes is dependent on *MAT*a/*MAT*α, but their expression is often not limited to sporulation, reflecting the participation of mitotic functions like B-type cyclins (*CLB3*, *CLB4*), anaphase-promoting complex (*APC2*, *APC4*, *APC5*, *CDC26*, *CDC27*) and other factors involved in cell division (*PDS1*, *CDC5*, *CDC14*, *KAR1*, *SPC42*, *SPO12*) during meiotic M-phase. Many known meiosis-specific genes (*NDT80*, *SPR1*, *SPR3*, *SPR6*, *SPR28*, *SSP2*) and factors ultimately required for proper spore maturation (*SPS1*, *SMK1*, *SPO19*, *SPO20*) are also in this group. A pronounced delay of transcriptional onset in W303 was reiterated in subsequent expression clusters, indicating that the slow meiotic progression of W303 may result from delayed induction of genes during commitment to sporulation.

Cluster 6a shows massive induction and remains highly expressed. *DIT1*, needed for spore-wall maturation[25], occurs in this group. Cluster 6b peaks at lower levels and shows re-repression during exit from M-phase. It contains genes acting in re-repression of early and middle loci (*IME2*), transition through meiosis II (*SPO14*, *CLB5*), cell structure (*CWP1*, *ECM37*) and various metabolic activities (*PHO4*, *ALG5*, *EHD2*, *ERO1*). It should be noted that *IME2*, previously classified as an early gene, displays biphasic expression in this analysis; it is first induced in SK1 at approximately three hours and reaches maximal levels at eight hours, causing the

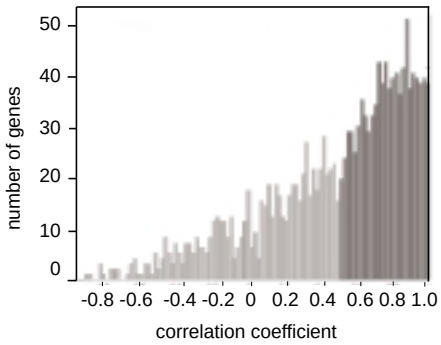

Fig. 5 Correlation of gene expression in SK1 and W303. Correlation coefficients were calculated and plotted for the 915 genes regulated during meiosis in both SK1 and W303. Genes showing a correlation coefficient indicating highly similar expression patterns (>0.5) are shown in dark grey; those with low coefficients are in light grey.

clustering algorithm to place it in this group. Finally, clusters 7a and 7b are upregulated at low levels late in meiosis, persisting through spore development. They contain known genes expressed in late log phase that may act in stress response as well as ascus maturation and germination, required for maintenance of cell structure (*CWP2*, *ECM1*), transcription (*RBP10*), the spindle checkpoint (*MAD3*) and translation (*RPS2*, *RPS8A*, *RPS12*, *RPS26b*, *RPS27b*).

Low-abundance early mRNAs are more efficiently detected in SK1 than in W303

The comparison of meiotically induced genes in SK1 and W303 yielded a large number of strain-specific developmentally regulated loci. We generated a cluster distribution (Fig. 7*a*) of the total number of meiotically regulated genes in SK1 (1568), including core genes regulated similarly in both strains (915), and those which appear to be regulated only in SK1 (653; data for genes expressed only in W303 not shown). Cluster 4, corresponding to the classical early meiotic expression class known to encode highly unstable and relatively lower-abundance mRNAs (ref. 16), shows the biggest strain difference (168 of 313 genes are detected only in SK1). Cluster 5, containing highly induced middle meiotic genes, display the smallest strain difference (only 44 of 217 genes have SK1-specific induction). The distributions of core- and SK1-specific genes in clusters 1, 2, 3, 6 and 7 are similar (see web sites for details). We also compared these data with the previous microarray analysis of SK1 (ref. 17), in which the more highly induced

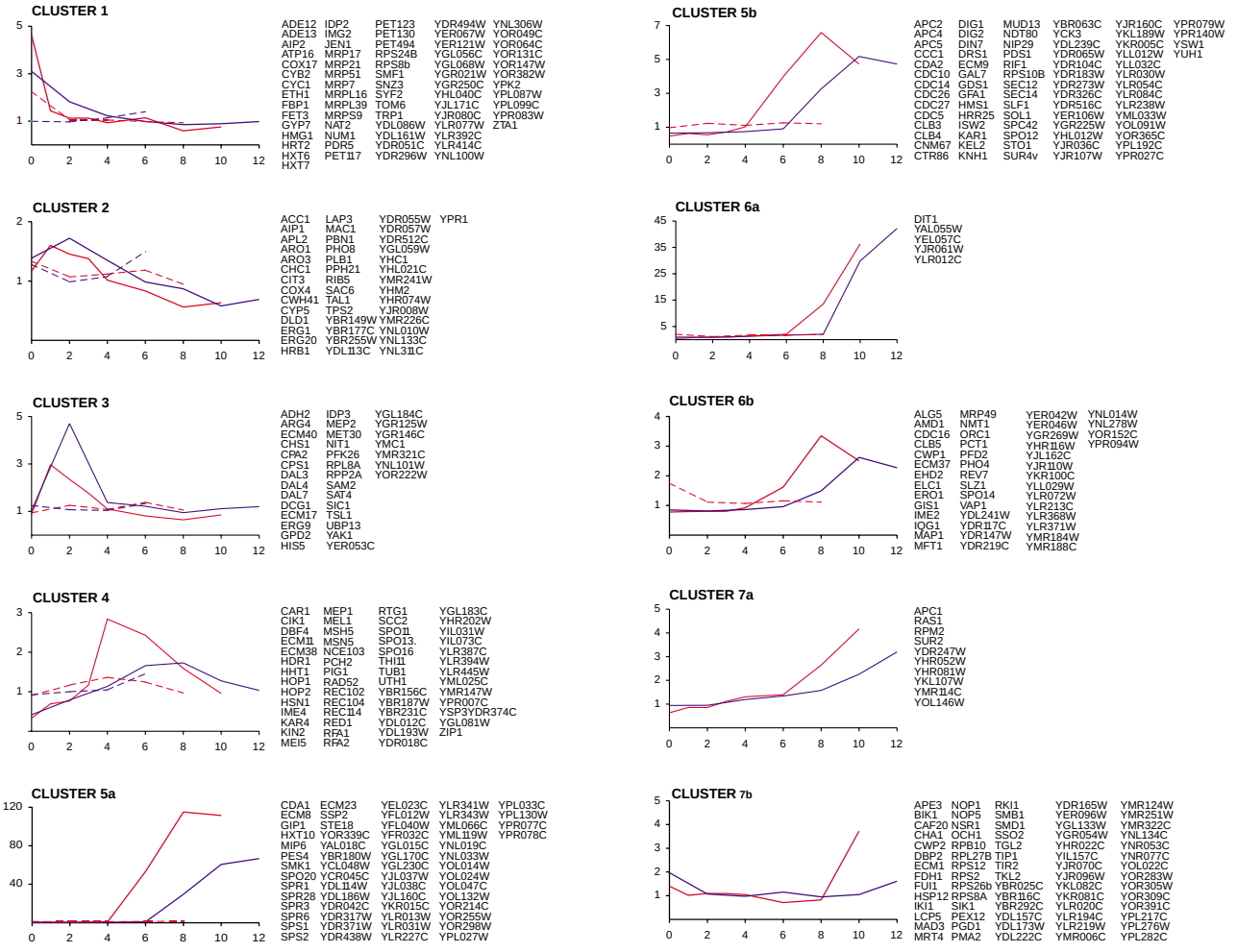

Fig. 6 Average expression profiles of well-correlated genes in SK1 and W303. Ten distinct meiotic expression patterns, plotted by hybridization fluorescence intensity (*y* axis) during sporulation (*x* axis), are shown for 477 selected genes with a correlation coefficient ≥0.8; *MAT***a**/*MAT*α SK1 (red) and W303 (blue), starvation controls *MAT*α/*MAT*α SK1 (dashed red line) and *MAT***a**/*MAT***a** W303 (dashed blue line). The graphs representing various clusters or sub-clusters are given on different scales to emphasize distinct levels of expression.

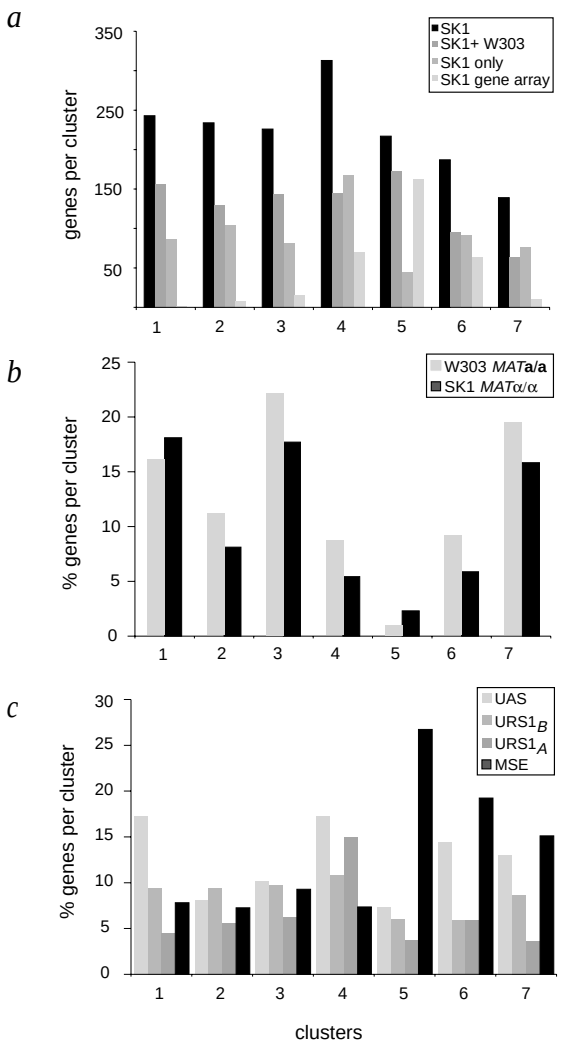

Fig. 7 Cluster comparisons. ***a***, Distribution by cluster of meiotically induced loci detected in SK1 (total), in both SK1 and W303 (core), in only SK1 (strain-specific) and in an independent gene array analysis[17]. ***b***, Distribution by cluster of starvation-induced genes. ***c***, Distribution by cluster of known regulatory elements as indicated.

middle and mid-late genes were detected most efficiently (Fig. 7*a*). Stress-response and starvation genes (clusters 1, 2 and 3) were virtually absent, and early and late genes (clusters 4 and 7, respectively) were under-represented, presumably because lower synchrony and sporulation efficiency prevented their detection. These comparisons suggest that the major differences in mRNA detection may be related to the level of population synchrony during meiosis, particularly in revealing lower abundance, or transient, species.

Genes transcribed during starvation are predominantly in stress-response expression clusters

The distribution of starvation genes expressed in SK1 (*MAT*α/*MAT*α) and W303 (*MAT*a/*MAT*a) is shown in Fig. 7*b*. Most of them are in clusters containing known genes involved in starvation and stress responses, such as those required for nitrogen metabolism and methionine biosynthesis (cluster 3). Early and middle classes expressed in sporulating cells are under-represented (clusters 4 and 5). Several genes involved in amino-acid and RNA metabolism are induced late in both sporulating and starving cells (cluster 7). This indicates the presence of a regulatory system governing the expression of genes required for various metabolic functions during starvation, independent of mating type.

Transcription factor binding sites upstream of developmentally regulated core genes

The distributions of known DNA promoter motifs was determined (within 600 bp upstream of the ATG) among the 1,568 meiotically regulated genes in SK1. Those for proteins implicated in meiotic progression, Abf1p, Ume6p and Ndt80p, are shown for the 915 core genes (Figs 4 and 7*a*). The Abf1p binding site (UAS_H, 5′–RTCRYY*N*NACG–3′; refs 12,26,27) occurs at slightly higher frequency among stress response, early meiotic, and certain middle and late meiotic genes (20% versus 13% in the genome), consistent with the known role of Abf1p in the transcriptional regulation of numerous metabolic and meiotic genes[12,28–30]. The Ume6p protein recognizes two related target sites, one of which (designated $URS1_A$, 5′–DSGGCGGCND–3′; ref. 31), shown to be critical in regulating *SPO13* (ref. 24) and other early meiotic genes[1], is highly over-represented in the upstream regions of genes in cluster 4 (16% versus 4.5% in the genome). The other element (termed URS_B, 5′–SGWGGMRRNANW–3′), identified in the promoter of a DNA-damage-response gene[32], is only moderately over-represented in cluster 4 (12% versus 8% in the genome). Some $URS1_A$ and URS_B-containing genes also display late expression patterns, suggesting that Ume6p functions more broadly during sporulation than previously thought[7,10]. Lastly, the Ndt80p target sequence (MSE, 5′–VKNCRCAAAWD–3′; refs 13,33) is particularly prevalent in the upstream regions of genes in clusters 4–7 (~30% versus 10.5% in the genome). Their expression patterns support the idea that Ndt80p regulates transcription of many genes involved in the meiotic divisions and later steps of spore formation[14]. Apart from these well-characterized regulatory elements, the sequences 5′–TATATA–3′ and poly$(A)_8$ are also moderately over-represented in the middle expression classes. The above binding-site survey correlates the presence of known meiotic regulatory motifs to specific meiotic expression patterns, but numerous meiotic genes do not contain matches to any of these sites, and many genes with these sites are not induced during sporulation. The promoter sequences and transcription factors governing these genes remain to be identified.

Discussion

Getting to the core of the meiotic question. We have defined the meiotic transcriptome of budding yeast by high-density oligonucleotide microarray analysis. Using the described criteria, we found approximately 1,600 genes to be meiotically regulated in sporulating SK1 or W303. This number is compatible with an independent estimate based on genomic analysis of *lac*Z translational fusions during sporulation[34]. Gene-by-gene comparison in SK1 and W303 identified approximately 900 "core" loci meiotically regulated in both strains (including ~650 meiotically expressed genes not previously reported[17]) that comprise most known sporulation genes and other ORFs essential for the process (European Functional Analysis Project[35]; A. Nicolas, pers. comm., and R.M.W. and R.W.D., unpublished data). Besides the core loci, several hundred SK1- or W303-specific regulated genes were uncovered. The presence of SK1-specific transcripts may be explained, in part, by the strain's highly synchronous sporulation properties, which yield more abundant populations of induced messages. The detection of approximately 700 genes regulated exclusively in W303, in which meiosis proceeds more slowly and with less synchrony, indicates that promoter sequence polymorphisms may also contribute to

strain-specific transcription (in both W303 and SK1). We found 8 of 39 genes deleted in SK1 to be transcriptionally induced in sporulating W303 cells, and 1 of 8 genes deleted in W303 to be upregulated in sporulating SK1 cells, indicating that these loci may also be important for controlling the different sporulation properties of the strains. Finally, *RME1*, a negative regulator of meiosis[36] expressed more than 10–20-fold higher in W303 than in SK1, may also participate in delaying the onset and progression of meiosis in this strain.

Meiosis-specific genes. The starvation control experiments revealed that most meiotically induced genes are dependent on the *MAT*a/*MAT*α locus (>95%). Among the approximately 300 genes transcriptionally induced in starving *MAT*α/*MAT*α or *MAT*a/*MAT*a cells, only a small fraction expressed in both controls (45) are found among the approximately 900 'core' meiotic genes described above (particularly genes required for methionine biosynthesis). The presence of a large number of starvation-regulated genes that vary between the two controls (~200) may arise from mating-type–specific starvation responses and/or other variation between strains. It should be noted that SK1 and W303 have numerous genetic differences that may contribute to strain-specific expression profiles (Fig. 1). Similarly, the sporulating and non-sporulating counterparts for each background, although highly related, are not completely isogenic. Although appropriate nutrients are added to sporulation media to minimize effects of different auxotrophic markers, these and possibly other undetected point mutations may also influence expression profiles. Given these considerations, the data thus provide a minimum estimate of the overlap or core genes in each case.

Among the genes identified, those that are meiosis-specific (transcribed exclusively in meiotic cells) are of particular interest because they may encode critical determinants that reprogram mitotic cells to undergo meiotic development. Our analysis revealed that approximately 250 of the approximately 1,600 meiotically regulated genes in SK1 are expressed in a meiosis-specific manner (after excluding genes expressed in starving and mitotically growing cells). Among these, approximately 70 appear in the core meiotic group. They include many known essential sporulation genes (for example, *SPO13*, *HOP1*, *DMC1*), underscoring the biological importance of this class of genes. As many of the non-core meiosis-specific genes in SK1 may also be relevant, but of too low abundance to be detected in W303, we estimate that as many as 150 genes in the yeast genome may be meiosis-specific. This estimate is close to that made based on recovery of sporulation-specific mutants (~200 genes) more than 20 years ago[37,38].

Algorithms and meiosis. Several lines of evidence support the proposed categorization into seven broad expression clusters and their functional designations (Figs 4 and 7). First, expression clusters are reproducibly detectable in sporulating, but not starving, control SK1 and W303 strains. Second, the well-defined meiotic clusters 4 and 5 contain many known sporulation genes that fail to be induced in sporulation-deficient control cells, whereas the starvation and largely metabolic clusters 1, 2 and 3 contain a substantial number of genes upregulated in starving cells. Third, there is a good correlation between the presence of specific motifs in the promoters of genes that are co-regulated in specific clusters. These and other studies demonstrate that genomic expression can reproducibly and reliably provide a broad overview of mRNA abundance associated with a biological process. It is also important to note potential limitations of such analyses. mRNA measurements based on average transcript levels in a population do not necessarily reflect what happens in a single cell. Moreover, changes in a population's pattern of mRNAs may result from other factors besides transcriptional regulation. For example, the 50–70% decrease in RNA recovery typically seen during sporulation (due to changes in RNA extractability and/or intracellular degradation), may itself alter the number of RNA species detected, even though the same total amount of labelled RNA produced the same sum of hybridization intensities.

From the message to the function. The general correlation between specific clusters and meiotic function suggests that cluster designations of unknown genes may be useful in providing clues about their function. Indeed, we found that two previously uncharacterized genes (*YILO73C* and *YPL130W*, now designated *SPO22* and *SPO19*, respectively), expressed as part of the early and middle groups (clusters 4 and 5), are essential for proper timing of MI and for completion of the nuclear divisions, respectively, similar to several other genes in these clusters (G.G.T. and R.E.E., unpublished data).

Earlier studies[39] and whole-genome analysis of deletions have demonstrated that the presence of transcriptional induction during meiosis does not necessarily imply an essential role for a gene in the meiotic process. Why do genes that appear to have no essential role in sporulation show meiotically regulated transcription? One explanation is genetic redundancy and/or the ability of some gene functions to substitute for others (for example, cyclins). Another is that promoters may become activated cotemporally with meiotically important genes because their regulatory motifs have evolved a 'meiotic' role during the course of evolution. Given the similarity of meiotic chromosome behaviour and homology of genes in different organisms, the genome-wide identification of meiotically regulated genes in budding yeast, together with deletion analysis to evaluate their relevance to gametogenesis, should ultimately lead to a more comprehensive view of how this process occurs in all eukaryotes.

Methods

Yeast strains. The sporulation-proficient W303 strain (REE3201/REE3182: *MATa/MATα ade2/ADE2, can1-100/CAN1, CYH2/cyh2, his3-11,15/his3-11,15, LEU1/leu1-c, LEU2/leu2-3,112, trp1-1:URA3:trp1-3′Δ/trp1-1, ura3-1/ura3-1*) is near isogenic to the non-sporulating W303 control strain (REE3063: *MAT*a/*MAT*a, *ade2/ade2, can1-100/can1-100, his3-11,15/his3-11,15, leu2-3,112/leu2-3,112, trp1-1/trp1-1, ura3-1/ura3-1*). REE3201 was derived from W303-1A by integration of pRS19 (ref. 40) at *trp1-1* to create a *trp1* duplication for monitoring recombination. REE 3182 was derived by multiple backcrosses between W303-1B and K264-10D (a strain highly related to W303 from our laboratory). W303-1A, W303-1B and W303 *MAT*a/*MAT*a were provided by R. Rothstein. The sporulation-proficient SK1 strain (NKY1551: *MAT*a/*MAT*α *ho:LYS2/ho:LYS2, ura3/ura3, lys2/lys2, leu2:hisG/leu2:hisG, arg4-Nsp/arg4-Bgl, his4x:LEU2-URA3 /his4B:LEU2*) is near isogenic to the non-sporulating SK1 control strain (NKY471: *MATα/MATα ho:LYS2/ho:LYS2 ura3/ura3 lys2/lys2*, differing by the presence of integrated markers for monitoring recombination (provided by N. Kleckner).

Media and culture conditions. We used rich medium (YPD), pre-sporulation medium (YPA) and sporulation medium (SPII and SPIII), supplemented with specific amino acids and nitrogen bases (each at 50 μg/ml) to compensate for auxotrophic markers in the strains, following standard protocols[41]. Induction of sporulation was carried out as described[41,42] with modifications (see our web sites). We took aliquots of sporulating cells from the W303 culture at 0, 2, 4, 6, 8, 10 and 12 h and from the SK1 culture at 0, 1, 2, 3, 4, 6, 8 and 10 h. These time points were chosen to reflect the progression of meiotic landmarks in each strain. Non-sporulating control samples were collected from SK1 *MAT*α/*MAT*α at 0, 2, 4, 6 and 8 h, and from W303 *MAT*a/*MAT*a at 0, 2, 4 and 6 h, respectively.

FACS analysis. We fixed 1×10^7 cells with 70% ethanol, resuspended in sodium citrate (50 mM) and sonicated for 10 s at 20% power with a Branson Sonifier 250. Samples were digested with RNase H (250 μg; Boehringer) for 1 h at 50 °C, stained with propidium iodide (20 μg/ml), diluted 1:10 in 1×PBS and analysed by FACS (Becton-Dickinson).

Recombination assays. We determined recombination frequencies in W303 and SK1 using auxotrophic tryptophan and histidine gene duplication markers (*trp1-1:URA3:trp1-3′Δ* and *his4x:LEU2-URA3/his4B:LEU2*, respectively) as described[10,42]. At the intervals indicated, ~5×10^7 cells were removed, sonicated briefly with a Branson Sonifier 250 at 20% for 5 s, washed in sterile water, and plated on media lacking tryptophane or histidine to select for prototrophic recombinants, and on synthetic complete media to determine viability.

Light and fluorescence microscopy. Cells were fixed with formaldehyde, and ascus formation monitored by Nomarski light microscopy. Samples were also fixed in ethanol and mounted in Moviol 4.88 containing Hoechst 33258 (1 μg/ml; CalBiochem) to detect MI and MII chromatin segregation, using a Zeiss Axiophot microscope equipped with a Xenon ultraviolet lamp. We counted ~300 cells per sample in each case.

RNA preparation and DNA probe synthesis. We collected aliquots (400 ml) and stored them as pellets at –80 °C. Each pellet was thawed rapidly and resuspended in phenol (20 ml; Amersham), 20 ml breakage buffer (75 mM ammonium acetate, 10 mM EDTA), 10% sodium dodecyl sulphate (2 ml; SDS) and 5 g glass beads (0.3–0.5 μm; Thomas Scientific). The cells were vortexed vigorously for 2 min, incubated in a 65 °C water bath for 15 min, re-vortexed for 1 min and spun at 10,000 r.p.m. for 10 min. After re-extraction with phenol/chloroform (20 ml; 1:1), total RNA was precipitated (750 mM ammonium acetate, 2 volumes of ethanol), resuspended in Tris (10 mM, pH 7.5) and stored at –80 °C. Total RNA (~2 mg per sample) was applied to a Qiagen Oligotex maxi column for poly(A)$^+$ RNA isolation, following the manufacturer's recommended protocol. Poly(A)$^+$ RNA (~20 μg) was reverse transcribed using SuperScript II (GIBCO); the cDNA was digested with DNase I (Gibco) and biotinylated with a ddATP analogue NEL-508 (NEN) using Terminal Transferase (Boehringer). The RNA yield from these samples indicated a 50–70% drop in intracellular total RNA (including tRNA and rRNA) during the course of sporulation. When quantified on an array, however, equivalent amounts of labelled RNA (1% of total RNA) produced the same sum of hybridization intensities at the start and completion of sporulation.

Genomic DNA preparation. We grew W303 and SK1 *MAT*a/*MAT*α strains to 5×10^7 cells/ml in 100-ml YPD cultures. The cells were collected and washed once with sterile water. We isolated genomic DNA using Genomic-tip 500/G columns (Qiagen) according to the manufacturer's instructions.

GeneChip probe array hybridizations. The labelled cDNA probes were hybridized to antisense Ye6100 GeneChip probe arrays (Affymetrix) as described[20]. Genomic DNA was digested, labelled as described above and hybridized to sense Y6301 GeneChips[18].

Expression data analysis. We analysed GeneChips in a Hewlett-Packard Gene Array Scanner and collected primary data using GeneChip 3.0 software. Data analysis was done in three stages. First, the average difference was calculated for each gene probed on the array for all hybridizations (26 time points×4 arrays) using GeneChip 3.0. The average difference has been shown to quantitatively reflect the abundance of a particular mRNA molecule in a population, and is based on the mean of the differences between the hybridization intensity for the perfect match (PM) features and the mismatch (MM) features for a particular gene[43]. Second, the average difference values were normalized: for each array, the average difference of every gene was divided by the mean of all average difference values probed by that array. This procedure accounted for any variation in hybridization intensity between the individual arrays. Finally, these normalized data were analysed using a computer script to identify a set of ~1,600 genes comprising the best candidates for meiotically regulated genes. Criteria for inclusion were based on the ratio of the normalized average difference for each transcript to the mean average difference in the array (where >2 indicates an abundant transcript, and <2 indicates a low-abundance transcript). We selected genes expressing abundant transcripts that showed at least a four-fold change (for SK1) or a 2.5-fold change (for W303) in normalized average difference between the minimum and maximum normalized values during the time course. These criteria were chosen based on the expression data of a set of well-characterized loci regulated during meiosis[1] (see our web sites). The less-stringent criterion of 2.5-fold for W303 compensates for its lower sporulation synchrony, which decreases signal intensity. We used a more rigorous criterion for selection of low-abundance transcripts to recover genes with higher expression over basal levels. In this case, to be considered significant, the ratio between minimum and maximum expression levels had to be at least fourfold greater than the maximum for SK1 and 2.5-fold greater for W303. Genes whose maximum values during sporulation were less than the mean value for the array (ratio >1) were excluded from further analysis (although some meiotically regulated loci may be omitted by this procedure, these data were considered less reliable). The expression patterns of meiotically induced genes were statistically analysed (clustered) and processed for presentation using GeneSpring 2.3.4 and software written in Matlab 5.0 (The MathWorks). Clustering was carried out using a standard correlation coefficient of 0.95 as the measure for significant statistical similarity. The branching behaviour of the tree was controlled by a minimum distance setting of 0.001 and a separation ratio between clusters of 0.5. The minimum distance dictates how far down the tree discreet branches are depicted and reflects the fact that the expression patterns for most genes are not more closely correlated than 0.001. The separation ratio determines how large the correlation differences between groups of clustered genes has to be for them to be considered discrete groups and to be represented separately (set between 0 and 1). Expression clustergrams were visually examined and manually ordered over initial time of induction. Processed expression data of ~1,600 yeast genes obtained using sporulating and starving cells were organized in a relational database (see web sites). Meiosis-specific genes (not expressed during vegetative growth) were detected using the fluorescence signal values at t=0 of known sporulation genes as selection criteria (*SPO13*, *REC102*, *MEI4*, *SMK4*, *REC104*, *MEK1*, *SPS1*, *DMC1*, *HOP1*, *HOP2*, *NDJ1*, *IME2*, *SPS4*, *SPS18*, *SPO11*, *ZIP1*, *RED1*, *REC114*).

Search for regulatory DNA motifs. Regulatory motifs were detected using the 'Find-Regulatory-Sequences' subroutine in GeneSpring 2.3.4, which iteratively scans upstream sequences from four nucleotides to a specified maximum length. It reports the position of each sequence (with respect to the ATG start codon), expected probability of occurrence, observed frequency, and *P* value for those with a significantly small *P* value (default=0.05).

Acknowledgements

We thank L.H. Rutkowski and B.K. Washburn for critical reading of the manuscript; S. Fuller for on-line support of GeneSpring; J. Clark, C. Sarrauste de Menthière and B. Masdoua for help in designing the web sites and constructing the database; and J. Demaille, A. Fernandez and N. Lamb at the IGH for hosting M.P. during the final stages of the data analysis. This work was supported by NIH grants 1R01GM29182 (to R.E.E.) and 1RO1HG01633 (to R.W.D.). M.P. was supported by a Max Kade postdoctoral fellowship and in part by grant RG0533 (to A.F. and N.L.).

Received 9 March; accepted 20 October 2000.

1. Kupiec, M., Byers, B., Esposito, R.E. & Mitchell, A.P. Meiosis and sporulation. in *Saccharomyces cerevisiae* 889–1036 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1997).
2. Freire, R. *et al.* Human and mouse homologs of *Schizosaccharomyces pombe rad1*(+) and *Saccharomyces cerevisiae RAD17:* linkage to checkpoint control and mammalian meiosis. *Genes Dev.* **12**, 2560–2573 (1998).
3. McKim, K.S. & Hayashi-Hagihara, A. mei-W68 in *Drosophila melanogaster* encodes a Spo11 homolog: evidence that the mechanism for initiating meiotic recombination is conserved. *Genes Dev.* **12**, 2932–2942 (1998).
4. Edelmann, W. *et al.* Mammalian MutS homologue 5 is required for chromosome pairing in meiosis. *Nature Genet.* **21**, 123–127 (1999).
5. Zalevsky, J., MacQueen, A., Duffy, J., Kemphues, K. & Villeneuve, A. Crossing over during *Caenorhabditis elegans* meiosis requires a conserved MutS-based pathway that is partially dispensable in budding yeast. *Genetics* **153**, 1271–1283 (1999).
6. Mitchell, A.P. Control of meiotic gene expression in *Saccharomyces cerevisiae*. *Microbiol. Rev.* **58**, 56–70 (1994).
7. Strich, R. *et al. UME6* is a key regulator of nitrogen repression and meiotic development. *Genes Dev.* **8**, 796–810 (1994).
8. Anderson, S.F., Steber, C.M., Esposito, R.E. & Coleman, J.E. *UME6*, a negative regulator of meiosis in *Saccharomyces cerevisiae*, contains a C-terminal Zn2Cys6 binuclear cluster that binds the URS1 DNA sequence in a zinc-dependent manner. *Protein Sci.* **4**, 1832–1843 (1995).
9. Bowdish, K.S., Yuan, H.E. & Mitchell, A.P. Positive control of yeast meiotic genes by the negative regulator *UME6*. *Mol. Cell. Biol.* **15**, 2955–2961 (1995).
10. Steber, C.M. & Esposito, R.E. *UME6* is a central component of a developmental regulatory switch controlling meiosis-specific gene expression. *Proc. Natl Acad. Sci. USA* **92**, 12490–12494 (1995).
11. Rubin-Bejerano, I., Mandel, S., Robzyk, K. & Kassir, Y. Induction of meiosis in Saccharomyces cerevisiae depends on conversion of the transcriptional represssor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol. Cell. Biol.* **16**, 2518–2526 (1996).
12. Gailus-Durner, V., Xie, J., Chintamaneni, C. & Vershon, A.K. Participation of the yeast activator Abf1 in meiosis-specific expression of the *HOP1* gene. *Mol. Cell. Biol.* **16**, 2777–2786 (1996).
13. Ozsarac, N., Straffon, M.J., Dalton, H.E. & Dawes, I.W. Regulation of gene expression during meiosis in *Saccharomyces cerevisiae*: *SPR3* is controlled by both ABFI and a new sporulation control element. *Mol. Cell. Biol.* **17**, 1152–1159 (1997).
14. Chu, S. & Herskowitz, I. Gametogenesis in yeast is regulated by a transcriptional cascade dependent on Ndt80. *Mol. Cell* **1**, 685–696 (1998).
15. Hepworth, S.R., Friesen, H. & Segall, J. *NDT80* and the meiotic recombination checkpoint regulate expression of middle sporulation-specific genes in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 5750–5761 (1998).
16. Surosky, R.T. & Esposito, R.E. Early meiotic transcripts are highly unstable in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **12**, 3948–3958 (1992).
17. Chu, S. *et al.* The transcriptional program of sporulation in budding yeast. *Science* **282**, 699–705 (1998).
18. Winzeler, E.A. *et al.* Direct allelic variation scanning of the yeast genome. *Science* **281**, 1194–1197 (1998).
19. Padmore, R., Cao, L. & Kleckner, N. Temporal comparison of recombination and synaptonemal complex formation during meiosis in *S. cerevisiae*. *Cell* **66**, 1239–1256 (1991).
20. Cho, R.J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
21. Wodicka, L., Dong, H., Mittmann, M., Ho, M.H. & Lockhart, D.J. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nature Biotechnol.* **15**, 1359–1367 (1997).
22. Eisen, M.B., Spellman, P.T., Brown, P.O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).
23. Atcheson, C.L., DiDomenico, B., Frackman, S., Esposito, R.E. & Elder, R.T. Isolation, DNA sequence, and regulation of a meiosis-specific eukaryotic recombination gene. *Proc. Natl Acad. Sci. USA* **84**, 8035–8039 (1987).
24. Buckingham, L.E. *et al.* Nucleotide sequence and promoter analysis of *SPO13*, a meiosis-specific gene of Saccharomyces cerevisiae. *Proc. Natl Acad. Sci. USA* **87**, 9406–9410 (1990).
25. Briza, P., Breitenbach, M., Ellinger, A. & Segall, J. Isolation of two developmentally regulated genes involved in spore wall maturation in *Saccharomyces cerevisiae*. *Genes Dev.* **4**, 1775–1789 (1990).
26. Dorsman, J.C. *et al.* An ARS/silencer binding factor also activates two ribosomal protein genes in yeast. *Nucleic Acid Res.* **17**, 4917–4923 (1989).
27. Della Seta, F., Treich, I., Buhler, J.M. & Sentenac, A. ABF1 binding sites in yeast RNA polymerase genes. *J. Biol. Chem.* **265**, 15168–15175 (1990).
28. Planta, R.J., Goncalves, P.M. & Mager, W.H. Global regulators of ribosome biosynthesis in yeast. *Biochem. Cell Biol.* **73**, 825–834 (1995).
29. Rolfes, R.J., Zhang, F. & Hinnebusch, A.G. The transcriptional activators BAS1, BAS2, and ABF1 bind positive regulatory sites as the critical elements for adenine regulation of *ADE5,7*. *J. Biol. Chem.* **272**, 13343–13354 (1997).
30. Pierce, M. *et al.* Transcriptional regulation of the *SMK1* mitogen-activated protein kinase gene during meiotic development in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **18**, 5970–5980 (1998).
31. Sumrada, R.A. & Cooper, T.G. Ubiquitous upstream repression sequences control activation of the inducible arginase gene in yeast. *Proc. Natl Acad. Sci. USA* **84**, 3997–4001 (1987).
32. Sweet, D.H., Jang, Y.K. & Sancar, G.B. Role of *UME6* in transcriptional regulation of a DNA repair gene in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **17**, 6223–6235 (1997).
33. Hepworth, S.R., Ebisuzaki, L.K. & Segall, J. A 15-base-pair element activates the *SPS4* gene midway through sporulation in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **15**, 3934–3944 (1995).
34. Ross-Macdonald, P. *et al.* Large-scale analysis of the genome by transposon tagging and gene disruption. *Nature* **402**, 413–418 (1999).
35. Dujon, B. European Functional Analysis Network (EUROFAN) and the functional analysis of the *Saccharomyces cerevisiae* genome. *Electrophoresis* **19**, 617–624 (1998).
36. Covitz, P.A. & Mitchell, A.P. Repression by the yeast meiotic inhibitor *RME1*. *Genes Dev.* **7**, 1598–1608 (1993).
37. Esposito, M.S. & Esposito, R.E. The genetic control of sporulation in *Saccharomyces*. I. The isolation of temperature-sensitive sporulation-deficient mutants. *Genetics* **61**, 79–89 (1969).
38. Esposito, R.E. & Klapholz, S. Meiosis and ascospore development. in *The Molecular Biology of the Yeast Saccharomyces: Life Cycle and Inheritance* (eds Stratern, J.N., Jones, E.W. & Broach, J.R.) 211–287 (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 1981).
39. Kaback, D.B. & Feldberg, L.R. *Saccharomyces cerevisiae* exhibits a sporulation-specific temporal pattern of transcript accumulation. *Mol. Cell. Biol.* **5**, 751–761 (1985).
40. Strich, R., Woontner, M. & Scott, J. Mutations in ARS1 increase the rate of simple loss of plasmids in *Saccharomyces cerevisiae*. *Yeast* **2**, 169–178 (1986).
41. Klapholz, S., Waddell, C.S. & Esposito, R.E. The role of the *SPO11* gene in meiotic recombination in yeast. *Genetics* **110**, 187–216 (1985).
42. Bishop, D.K., Park, D., Xu, L. & Kleckner, N. DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell* **69**, 439–456 (1992).
43. Lockhart, D.J. *et al.* Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nature Biotechnol.* **14**, 1675–1680 (1996).