bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Fundamental principles governing sporulation efficiency: A network theory approach

Camellia Sarkar[1], **Saumya Gupta**[2,†], **Himanshu Sinha**[2,3,4,*], **and Sarika Jalan**[1,5,*]

[1]Centre for Biosciences and Biomedical Engineering, Indian Institute of Technology Indore, Khandwa Road, Simrol, Indore 453552, India
[2]Department of Biological Sciences, Tata Institute of Fundamental Research, Homi Bhabha Road, Colaba, Mumbai 400005, India
[3]Department of Biotechnology, Indian Institute of Technology Madras, Chennai 600036, India
[4]Initiative for Biological Systems Engineering, Indian Institute of Technology Madras, Chennai 600036, India
[5]Complex Systems Lab, Discipline of Physics, Indian Institute of Technology Indore, Khandwa Road, Simrol, Indore 453552, India
[*]Corresponding authors: Himanshu Sinha (sinha@iitm.ac.in); Sarika Jalan (sarikajalan9@gmail.com)
[†]Current address: Department of Cancer Biology, Dana-Farber Cancer Institute, Department of Genetics, Harvard Medical School, Boston, MA 02215, USA

ABSTRACT

Using network theory on an integrated time-resolved genome-wide gene expression data, we investigated intricate dynamic regulatory relationships of transcription factors and target genes to unravel signatures that contribute to extreme phenotypic differences in yeast, *Saccharomyces cerevisiae*. We performed comparative analysis of gene expression profiles of two yeast strains SK1 and S288c, which lie at extreme ends of sporulation efficiency. The results based on various structural attributes of the networks, such as clustering coefficient, degree-degree correlations and betweenness centrality suggested that a delay in crosstalk between functional modules can be construed as one of the prime reasons behind low sporulation efficiency of S288c strain. A more hierarchical structure in late phase of sporulation in S288c seemed to be an outcome of a delayed response, resulting in initiation of modularity, which is a feature of early sporulation phase. Further, weak ties analysis revealed meiosis-associated genes for the high sporulating SK1 strain, while for the low sporulating S288c strain it revealed mitotic genes. This was a further indication of delay in regulatory activities essential to initiate sporulation in S288c strain. Our results demonstrate the potential of this framework in identifying candidate nodes contributing to phenotypic diversity in natural populations.

Introduction

Understanding the molecular role of genetic variation is the current frontier in modern genetics. Even in well-studied model organism budding yeast (*S. cerevisiae*), we do not understand the mechanistic contribution of genetic variants in generating phenotypic variation in population.[1] The knowledge that transcript abundance is under genetic control[2] has paved way for multiple studies to investigate how genetic variation is mechanistically associated with gene expression changes that underlie physiological differences.[3,4] Yeast strains isolated from diverse ecological niches represent a useful resource to address how transcript abundance impacts phenotypic consequences.[5] The high genetic divergence among these yeast strains correlates well with high phenotypic variance observed when grown in multiple environments.[6] Underlying these growth differences, several molecular pathways have been identified by studying transcript abundance.[7] Recently, we used gene expression variation to study the effect of a genetic variant in the form of a single nucleotide polymorphism to elucidate its role in yeast sporulation efficiency variation.[8] While such studies have been useful, to gain even better insights into the molecular basis of complex traits, interdisciplinary approaches are required that can ascertain how genetic variants interact amongst themselves and with the environment to bring about phenotypic diversity in natural populations.[9]

Various methods have been proposed to perform comparative gene expression analysis such as clustering methods,[10] bootstrapping clustering,[11] four-stage Bayesian model,[12] Gaussian mixture models with a modified Cholesky-decomposed covariance structure,[13] etc. However, all these gene-centric methods tend to overlook local patterns where these genes are similar based on only a subset (subspace) of attributes, e.g. expression values. This led to implementation of pattern similarity based bi-clustering approaches to gene expression data that could find bi-clusters among co-regulated genes under different subset of experimental conditions.[14] However, the next step in interpreting gene expression profiles is to go beyond the gene-centric techniques by employing more global approaches to get a better understanding of how gene expression profiles are

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

specifically related to the regulatory circuitry of the genome.[15] Network theory provides an efficient framework for capturing structural properties and dynamical behavior of a range of systems spanning from society[16] to biology.[17,18] Furthermore, the architecture of networks has provided fundamental understanding to randomness and complexity in various biological, social and technological networks.[16,19–21] In this paper, we focus on understanding the behavior of complex systems in terms of the structural and functional relationships between the molecular entities[19,22,23] through various structural measures of the network, viz. degree-degree correlations,[24] hierarchy[25] and weak ties analysis.[26] The basic structural properties of networks are dependent on how the networks evolve, the inherent interdependencies of the nodes as well on the architectural constraints. While on one hand, these network measures help in identifying important nodes of the network, on the other hand, these measurements also enable realization of the impact of interactions on the behavior of the underlying system. Hence, studying network parameters have expanded our understanding of biological processes, for instance by identifying important genes for diseases,[17,18] elucidating the mechanism behind human diseases by analyzing relationships between disease phenotypes and genes,[27] and deciphering the common genetic origin of multiple diseases.[28]

In the current work, we used the network theory approach to investigate how transcriptional regulatory networks differ between two genetic backgrounds leading to extreme phenotypes in terms of their sporulation efficiency in the same environment. Sporulation in yeast, a developmental process initiated under extreme nutrient starvation, involves meiotic cell division followed by spore formation.[29,30] Several genome-wide transcriptome analyses have been performed to elucidate the cascade of transcriptional regulation during sporulation.[31,32] This has led to identification of critical regulatory nodes during sporulation that are responsible for cells transitioning between different stages in meiosis, viz. *IME1*, initiator of meiosis and *NDT80*, regulator of meiotic divisions. In yeast, most information about sporulation has been obtained by studying the SK1 strain, since it has high sporulation efficiency of 90% within 48h.[33] The standard laboratory strain S288c, genetically divergent from SK1, is generally not studied for sporulation as it has low sporulation efficiency (5-15% in 48h[33]). Both SK1 and another closely related strain, W303, have approximately 60% of genes showing correlated expression patterns during sporulation, with majority of these genes associated with gametogenesis.[32] Moreover, genetic studies have screened for sporulation defects in the deletion collection library constructed in S288c, and there is good agreement for the sporulation genes identified in this collection and SK1.[29,34–36] However, gene deletions in a single strain fail to reflect the extent of phenotypic variation observed in nature, which mostly arises from polymorphisms and gene duplications rather than gene loss.[37] Multiple linkage mapping studies have been performed between S288c and SK1 strains,[38,39] and also in natural isolates of yeast such as oak and wine strains.[40–42] They have collectively identified eleven genetic contributors of the sporulation efficiency variation, including known sporulation genes - *FKH2*, *IME1*, *PMS1*, *RAS2*, *RIM15*, *RIM101*, *RME1*, *RSF1* and *SWS2*, and a couple of non-sporulation genes, *MKT1* and *TAO3*. Interestingly the sporulation genes with causal variants are all known to affect the initial regulatory events during sporulation.[42] In this study, we propose to develop a framework using network analysis that can be used to understand the effect of genetic background differences on the underlying molecular network of individuals showing phenotypic variation. We use time-resolved transcriptomics data and integrate it with the known physical gene interaction network of yeast to create a dynamic yeast network at multiple time points. Using network parameters, we identify the molecular nodes that get highly perturbed during the sporulation process in two yeast strains showing extreme sporulation efficiencies.

Results and Discussion

We constructed a dynamic transcriptional regulatory network of SK1 strain during sporulation (see Methods) and noted the three phases of sporulation (Fig. 1a) by comparing the appearance of crucial meiosis regulators in the network (Fig. 1b) with their expression profiles described previously.[29,32] These three phases in SK1 have been named as early, middle and late phases of sporulation, respectively.[32] *NDT80* gets activated in the beginning of the middle phase of sporulation, around 2-3h in sporulation medium.[43] Concordantly, we observed increased expression of *NDT80* appearing in the time span from T_3 to T_{12} (from 3h till 12h in sporulation) in the dynamic sporulation network of SK1 (Supplementary Fig. S1). Interactions of *NDT80* increased from T_3, reaching a maximum in T_7 and then decreased as time progressed in sporulation (Supplementary Fig. S1). Based on the appearance of *NDT80* in the dynamic sporulation network of SK1, we classified T_1-T_3 (1-3h in sporulation) as the early phase, T_4-T_6 (4-6h in sporulation) as the middle sporulation phase and T_7 onwards (7h onwards in sporulation medium) as the late sporulation phase. The regulators of *NDT80* constituted the high degree nodes in SK1 such as *MSN4*, stress responsive transcriptional activator; *AFT1*, regulator of iron homeostasis; *FHL1*, regulator of ribosomal protein transcription (Supplementary Fig. S1 and Table 1).

NDT80 is the prime initiator of sporulation in SK1, whereas, in S288c, the low sporulating strain, most of the cells do not enter meiosis at all and remain arrested in the stationary phase (G_1/G_0 phase).[8] This difference in the number of cells entering meiosis ultimately leads to a difference in sporulation efficiency, for SK1 it being greater than 90% in 48h and for S288c remaining very low at 10%. Interestingly, this low efficacy of S288c does not increase even when incubated in sporulation medium for one week[38] and therefore, it is unlikely to show distinct phases as observed in SK1. Thus, in order to determine

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the molecular differences that results in these two genetically divergent strains show sporulation efficiency variation, we chose SK1 sporulation phases as the basis for comparing their dynamic temporal profiles. Furthermore, the density of expression profiling time points varied between the two strains; linear time series with 1h gap in SK1 and log time series with denser time points early in sporulation for S288c. Thus, in order to compare across these two temporal time profiles, for S288c, T_1-T_5 (30m to 2h30m in sporulation medium), T_6-T_7 (3h50m to 5h40m) and T_8 (8h30m) time points were considered as corresponding time points of the early, middle and late phases of SK1, respectively. This comparison would allow us to determine differences in expression profiles of genes that get misregulated in S288c. We began with constructing the networks and comparing the general network properties of the two strains during sporulation (Fig. 2, Supplementary Tables S1, S2). The nodes of the networks are the differentially expressed genes, which were identified at each time point by setting the threshold value on log_2 fold differences as 1.0. Hence, genes that were considered overexpressed or repressed showed at least a 2-fold difference with respect to at the first time point (i.e. t_0 = 0h).[8]

The gene regulatory networks investigated here showed a heterogeneous degree distribution with a few of the nodes dominating the entire network, as observed in most real world networks.[22] SK1 exhibited a wider range of network sizes across different sporulation time points compared to S288c (Tables 1 and 2). The larger and denser networks were indicative of more extensive regulatory changes in the SK1 compared to the S288c. In order to follow these changes, we investigated the early, middle and late phases of sporulation in SK1 independently and compared them to the corresponding time points in S288c.[30] We found that there was a drastic increase in the number of genes having significantly high or low expression values in the consecutive time points at the onset of sporulation in both the strains (Tables 1 and 2), which could be due to cells transitioning from mitotic growth to initiate meiosis. This extensive reprogramming of gene expression early in sporulation as the cells prepare to enter meiotic cell division,[44] was revealed as an abrupt increase in the involvement of genes with sporulation progression in the early phase in both the strains. However, as the sporulation progressed, in the later phases of sporulation, the rate of change of network size reduced. Despite changes in the early sporulation phase in both the strains, the ratio of number of differentially expressed transcription factors (N_{TF}) and target genes remained almost constant across all the time points (Tables 1 and 2). The proportion of regulatory genes remaining constant throughout the sporulation indicated that it might be an intrinsic property of the sporulation process.

A change in the number of connections modulates the intrinsic properties of a network.[22] We investigated the impact of this change for both the strains during various sporulation phases. Similar to the network size, the number of connections (N_c) increased drastically in the early time points of sporulation in both the strains. However, this rate of increase in the number of connections was much higher in the case of SK1 as compared to S288c. For instance, in the earlier phases during sporulation, S288c had a two-fold increase in the number of connections, whereas, SK1 exhibited a four-fold increase (Tables 1 and 2). Since all interactions for both the strains are taken from the same repository base network, a change in the number of connections will only be possible if old nodes (genes) disappear and or new nodes arise in the networks. A higher rate of increase in the number of connections in SK1 as compared to the rate of increase in their size could be, therefore, attributed to the appearance of more number of high degree nodes in the second time point (Table 1). The nodes having high degree refers to genes that regulate a large number of genes. It is possible that there might also be a few feeble interactions of these highly interacting genes with other genes that are not significant. These highly interacting genes or nodes are known to be important in various cellular processes.[17] In the middle phase of sporulation, associated with processes involved in meiotic divisions,[31] the number of connections did not show considerable change for both the strains since we find that more than 75% of the genes remain same across the different time points in this phase in each strain. However, in the late sporulation phase, there was a change in the number of connections in S288c while for SK1, this number remained almost constant compared to the middle phase. From middle to late phase, a fall in the number of connections in S288c was observed. Incidentally, this decrease in the number of connections could be due to the disappearance of the high degree node *BAS1*, a Myb-related transcription factor involved in amino acid metabolism and meiosis.[45] Interestingly, *BAS1* contributed to approximately 50% of the connections in the early phase of S288c (Table 2) even though it is not one of the known regulators of sporulation,[30] and its disappearance in the middle phase was reflected in the number of connections. Furthermore, surprisingly, this gene was involved in the regulatory processes only in the early phase of sporulation and disappeared during the middle phase in both the strains. On one hand, this indicated the specific significance of this gene intrinsic to the early phase of sporulation; on the other hand it reflected the drastic changes in the regulatory activities from the early to the middle phase. Furthermore, in the late sporulation phase of S288c strain, the number of connections almost doubled and two known stress-responsive regulators, namely *MSN4* and *HSF1*[46] with a large number of edges appeared in this phase. However, in SK1, *MSN4* consistently appeared as one of the high degree nodes in both the early and middle phases implicating that it might be one the crucial signatures of high sporulation efficiency. It could be concluded that its absence in the early and late phases of S288c sporulation could be a reason of the cells poor ability to sporulate, however, it would be difficult to speculate if it is a cause or a consequence. Appearance of *MSN4* later than early sporulation phase (with respect to SK1) in S288c might be an indication of the delayed sporulation and indicating its important role in the regulation of sporulation.[47] The differences in the number of connections between the

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

strains in the three phases of sporulation, further motivated us to compare their general principles of regulatory interactions during sporulation.

So far, we focused only on the number of genes and the interactions in the networks. To understand how the interacting patterns impacted the overall structure of the underlying networks, we investigated the degree-degree mixing of the connected nodes across the three phases of sporulation in the two strains. Disassortativity is a parameter that measures the correlation in the degrees of the nodes in a network and provides understanding of the dislikelihood in connectivity of the underlying systems.[48] In gene regulatory networks, highly connected nodes avoid linking directly to each other and instead connect to proteins with only a few interactions, thus exhibiting disassortative topology.[49] This behavior of the nodes leads to a reduction in crosstalk between different functional modules and increase in the robustness of the networks by localizing the effects of deleterious perturbations.[50] The Pearson (degree-degree) correlation coefficient (r) was calculated for the networks at all time points in each of the strains (see Methods). As expected for gene regulatory networks, sporulation networks in both SK1 and S288c exhibited disassortativity at all time points (Fig. 2). A high value of this property was observed in both the strains during the early phase of sporulation, suggesting that the strains were more resilient to perturbations while carrying out early sporulation transcriptional events.[50] After the early phase, in SK1, disassortativity values reached a steady state at middle sporulation phase, while those of S288c still fluctuated (Fig. 2). Taken together, these observations implied that the necessary crosstalk between functional modules occurred early and then stabilized in SK1, while they were still going on or were random and unstable in the middle and late phases of S288c.

After analyzing the global properties of the sporulation networks, the local properties of the networks, which were expected to reveal the impact of local architecture on the phenotypic profiles of the two strains, were investigated. Clustering coefficient is one such local property that measures the local cohesiveness between the nodes.[51] A high value of clustering coefficient of a node depicts high connectivity among the neighbors of that node. For the two strains, we evaluated the average value of clustering coefficient ($\langle C \rangle$) for each time point (see Methods). As expected for various biological networks,[22] a high value of $\langle C \rangle$ was observed for the networks at all time points in both the strains as compared to their corresponding random networks (Fig. 2, Supplementary Tables S1, S2) as expected.[51] Furthermore, keeping in view the manner in which we constructed the sporulation networks, a high $\langle C \rangle$ meant that many of the neighbor target genes of a transcription factor also acted as transcription factors for the other neighbor target genes of that same transcription factor. On comparing the average value of clustering coefficient between the strains, a sharp increase in $\langle C \rangle$ was observed three times for SK1 coinciding with the early, middle and late phases of sporulation, while for S288c only two such transitions were observed for this property (Fig. 2). Moreover, while the transitions between the three peaks were rapid in the SK1, a slower transition between the first and second peak was observed for S288c. High clustering in cellular networks is known to be associated with the emergence of isolated functional modules.[23] Our results of average clustering coefficient suggested that the increased time taken by S288c to form functional modules could be due to a delay in relaying signaling information from early to middle phase of its sporulation.

In order to further unravel the differences of the sporulation process between the two strains, we investigated how number of neighbors of nodes denoted by node degree was associated with their neighbor connectivities (interactions between the neighbors of the node of interest) evaluated in terms of clustering coefficient (see Methods). All the networks in SK1 and S288c exhibited negative degree-clustering coefficient correlation (Supplementary Figs. S2, S3) as observed in various other real world networks, indicating the existence of hierarchy in these underlying networks.[23] A hierarchical architecture implies that sparsely connected nodes are part of highly clustered areas, with communication between the different highly clustered neighborhoods being maintained by a few hubs. We quantified this hierarchy (h), also termed as global reaching centrality in the networks[25] (see Methods) and found that in both the strains, the networks were more hierarchical at the beginning of sporulation (Fig. 2). A high value of hierarchy has been associated with modularity in the network. For instance, in case of metabolic networks, hierarchical structure indicates that the sets of genes sharing common neighbor are likely to belong to the same functional class.[52] A low value of h indicates more random interactions in the underlying networks. A decrease in hierarchy was observed until the middle phase of sporulation in both the strains. While SK1 continued to exhibit diminishing hierarchy in the late phase, in S288c there was an increase in the hierarchy at the last time point, again suggesting an increase in modularity in later phase of sporulation in S288c. These results implied that since both the strains showed high values of disassortativity, average clustering coefficient $\langle C \rangle$ and h values early in sporulation, the nature of genes involved in transferring information from the early to middle and late phases of sporulation would be important for us to understand the phenotypic difference between them. Previous sporulation studies have shown that many causative sporulation-associated genetic variants are present in genes regulating early sporulation processes.[8,42] Therefore, next we identified the genes that would directly or indirectly be involved in bringing about the phenotypic differences in both the strains as sporulation progresses.

For a network, betweenness centrality (see Methods) is a measure of network resilience[53] and it estimates the number of shortest paths (the minimum number of edges traversed between each of the pairs of nodes) that will increase if a node is removed from the network.[54] Usually nodes with high degree have high betweenness centrality[16] and are known to bridge different communities in the network. However, in a network, there exist some nodes, which despite having low degree have

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

relatively high betweenness centrality.[16] In the case of gene regulatory networks, such nodes (genes or transcription factors), while are involved in less number of regulatory interactions but these interactions are with different signaling pathways. Thus, these nodes are expected to have special significance in the underlying networks as their removal can result in a major breakdown in the pathways controlling the sporulation process. Furthermore, in very few cases, a target gene, known to have low degree may also have relatively higher betweenness centrality than the other target genes if it is simultaneously being regulated by several transcription factors. We identified a few important sporulation genes showing this property of low degree and high betweenness centrality in both SK1 and S288c (Fig. 3). In the SK1 networks, these genes were known regulators of respiratory stress and starvation, namely *STP2*,[55] *PMA1*[56] and *RPL2B*,[57] while in S288c these were *IME1*[8] and *TOS4*,[58] genes involved in initiation of meiosis and DNA replication checkpoint response, respectively. Generally, sporulation genes appeared to show this property in the early phase of SK1 but during the middle to late phase in S288c (Supplementary Tables S3, S4). These results suggested that this late appearance of important early sporulation genes as bridges that could transfer information between regulatory modules during early sporulation might be the cause for sporulation not proceeding in S288c. Thus, above analyses helped us to identify influential genes underlying the differential sporulation process. We next identified a few interactions that might be instrumental in regulating the sporulation process by considering an important proposition from sociology, Granovetter's Weak ties hypothesis.[59] This hypothesis states that the degree of overlap of two individuals' friendship networks varies directly with the strength of their tie to one another. In the networks, the ties having low overlap in their neighborhoods (i.e. less number of common neighbors) are termed as the weak ties.[26] The weak ties that have high link betweenness centrality (see Methods) are the ones known to bridge different communities.[60] Such weak ties revealed through our analysis of different sporulation networks are listed in Tables 3 and 4. Interestingly, we found repetitive occurrence of the same weak ties in consecutive time points for both the strains indicating their phase-specific importance in yeast sporulation. For instance, *BAS1-RTT107*, *BAS1-TYE7*, *YAP6-BAS1* and *ASK10-HMO1* were repetitive weak ties with high link betweenness centrality in consecutive time points of S288c networks while in SK1 networks *DAL81-ACE2* and *CDC14-ACE2* were such ties. In order to assess the functional importance of these weak ties, we investigated the characteristic properties of the end nodes of these weak ties. Unlike social networks where the end nodes of weak ties are low degree nodes,[61] in the sporulation networks of both the strains, the nodes forming weak ties were high degree nodes. An example of this was again *BAS1*, which as discussed above, is a Myb-related transcription factor involved in amino acid metabolism and meiosis.[45] In addition to *BAS1*, other important sporulation regulatory genes were identified in SK1, such as *RIM101*, a pH-responsive regulator of an initiator of meiosis;[62] *IME2*, a serine-threonine kinase activator of *NDT80* and meiosis;[63] *CDC14*, a protein phosphatase required for meiotic progression;[64] *HCM1*, an activator of genes involved in respiration.[65] Whereas in S288c, apart from *BAS1*, genes associated with mitotic functions such as *TYE7* for glycolytic gene expression,[66] *YAP6* for carbohydrate metabolism,[67] *RTT107* for DNA repair,[68] *ASK10* for glycerol transport[69] and *HMO1* for DNA structure modification[70] were identified. These results showed that while in SK1 meiosis-associated genes formed important bridges, in S288c these bridges were formed by genes involved in mitotic functions. This implied how differences in weak ties in regulatory networks can help us understand the dramatic differences observed in phenotypes. Moreover, *DAL81*, a nitrogen starvation regulator[71] and *ACE2*, a regulator of G_1/S transition in the mitotic cell cycle[72] were identified as end nodes of repetitive weak ties in SK1, suggesting their probable regulatory role in the sporulation process that requires further investigation.

Conclusion

This study presents a novel framework for assessing the molecular underpinnings of the phenotypic variation across strains due to the genetic differences between them. We studied the combined effect of genetic variants on the dynamic yeast sporulation network and used comparative analysis of various network parameters between two yeast strains showing extreme phenotypic differences. This framework helped reveal the characteristic signatures of the phenotype of interest and identified candidate genes contributing to phenotypic variation. Using this framework, we showed that the comparative analysis of parameters measuring the network connectivity and degree-degree mixing were the best in identifying differences between two yeast strains showing diverse sporulation efficiency. Comparing the basic structural attributes of the dynamic sporulation networks of the two strains revealed that a delayed crosstalk between functional modules of the low sporulating S288c might be the plausible reason behind its low sporulation efficiency. The end nodes of the repetitive weak ties, which are instrumental in bridging communities, were meiosis-associated genes for SK1 while these nodes in S288c were involved in mitotic functions, thus outlining the importance of this parameter in unraveling the molecular differences between the two strains.

The three sharp transitions in the average clustering coefficient in the SK1 indicating formation of functional modules correlate very well with the known early, middle and late phases of sporulation.[15] This three-tiered modularity was not observed in the S288c with a delayed appearance of the second peak of average clustering coefficient. These observations in S288c imply that a probable delay in cross talk between the early phase genes results in delayed formation of a functional module in later phases. This speculation is especially interesting since most causative genetic variants known to contribute to

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

sporulation efficiency variation have been observed in genes either showing early role in sporulation[73] or affecting genes with early regulatory role in sporulation.[8,44]

Application of genome-wide strategies to elucidate the molecular networks in multiple genetic backgrounds provides us with the opportunity to understand the impact of natural variation. Studying these network properties for variation in causal genes would further help in understanding specific molecular effects in the different temporal phases of the phenotype. The strategies adopted in this work can be extended to assess the impact of molecular perturbations in the already known core interaction network of an organism.[1,74] Moreover, application of such a network analysis on gene expression datasets for disease progression in complex diseases such as cancer and metabolic diseases can help identify specific nodes perturbing the underlying molecular pathways that can be focus of personalized medicine and drug target discovery.

Methods

Network construction

For constructing the transcriptional regulatory sporulation network, the known static regulatory interactions were overlaid on the time-resolved transcriptomics data of the two strains. This created the dynamic integrated sporulation network. The static network known for yeast contains all the known regulatory interactions between all the yeast transcription factors (TF) and their target genes (TG). These interactions were obtained from YEASTRACT database,[75] a curated repository of regulatory associations in *S. cerevisiae*, based on more than 1,200 literature references.

Gene expression data for yeast strains SK1[76] and S288c[8] was obtained from previously published studies. These datasets contained gene expression of 6,926 genes across 13 different time points in linear scale (0h to 12h with 1h intervals termed as T_0 to T_{12}, respectively) in SK1 and 9 different time points in logarithmic scale (0h, 30m, 45m, 1h10m, 1h40m, 2h30m, 3h50m, 5h40m, 8h30m termed as T_0 to T_8, respectively) in S288c. Gene expression analysis was performed as described previously.[8] In brief, all time points were normalized together using *vsn*[77] and the log_2 transformed expression values obtained after normalization were smoothed using $locfit$.[78] Fold differences in expression values were calculated for all the time-points relative to t = 0h (t_0), as follows:

$$Y'_{SK1(t_n)} = Y_{SK1(t_n)} - Y_{SK1(t_0)} \tag{1}$$

$$Y'_{S288c(t_n)} = Y_{S288c(t_n)} - Y_{S288c(t_0)} \tag{2}$$

such that Y is the expression value of a transcript for a strain (SK1 or S288c) at a specific time point n and Y' is the transformed expression value.

Differentially expressed genes were identified at each time point by setting the threshold value on log_2 fold differences as 1.0. Hence, genes that were considered overexpressed or repressed showed at least a 2-fold difference with respect to the first time point t_0 (i.e. t = 0h).[8]

The dynamic sporulation network was constructed by overlaying the experimentally determined yeast sporulation-specific gene expression values on the yeast static network. For each time point of each strain, only those TF-TG pairs were considered that both showed either overexpression or repression. These pairs were included in the subnetwork for that specific time point and thus, subnetworks for each time point were constructed for each strain. For comparison of the gene names obtained from YEASTRACT and the sporulation gene expression data, aliases were obtained from Saccharomyces Genome Database.[79]

Data availability

The adjacency matrices of the networks constructed using time-resolved sporulation data drawn from SK1 and S288c strains, the corresponding gene indices and transcription factors are freely available online at Figshare.[80]

Structural parameters

Several statistical measures are proposed to understand specific features of the network.[19,22] The number of connections possessed by a node is termed as its degree. The spread in the degrees is characterized by a distribution function $P(k)$, which gives the probability that a randomly selected node has exactly k edges. The degree distribution of a random graph is a Poisson distribution with a peak at $P(\langle k \rangle)$. However, in most large networks such as the World Wide Web, the Internet or the metabolic networks, the degree distribution significantly deviates from a Poisson distribution but has a power-law tail $P(k) \sim k^{-\gamma}$. The inherent tendency of social networks to form clusters representing circles of friends or acquaintances in which every member knows every other member, is quantified by the clustering coefficient.[51] We categorize the nodes as high and low degree nodes by arranging all the nodes in a network in descending order of degrees and keep assigning the nodes as high degree nodes until the next lower degree node differs by nearly 1.5-fold from the former in terms of the degree. Clustering coefficient of a

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

node i denoted as C_i, is defined as the ratio of the number of links existing between the neighbors of the node to the possible number of links that could exist between the neighbors of that node[81] and is given by

$$C_i = \frac{2\sum_{j_2=1}^{k_i}\sum_{j_1=1}^{k_i}(A_{ij_1}A_{j_1j_2}A_{j_2i})}{k_i(k_i-1)} \tag{3}$$

where i is the node of interest and j_1 and j_2 are any two neighbors of the node i and k_i is the degree of the node i. The average clustering coefficient of a network corresponding to a particular condition ($\langle C\rangle$) can be written as

$$\langle C\rangle = \frac{1}{N}\sum_{i=1}^{N} C_i \tag{4}$$

We define the betweenness centrality of a node i, as the fraction of shortest paths between node pairs that pass through the said node of interest.[54]

$$x_i = \sum_{st}\frac{n_{st}^i}{g_{st}} \tag{5}$$

where n_{st}^i is the number of geodesic paths from s to t that passes through i and g_{st} is the total number of geodesic paths from s to t. All the nodes were plotted and the top 5% of the nodes (genes) with high betweenness centrality but low degree were identified.

We quantify the degree-degree correlations of a network by considering the Pearson (degree-degree) correlation coefficient, given as[48]

$$r = \frac{[M^{-1}\sum_{i=1}^{M} j_ik_i] - [M^{-1}\sum_{i=1}^{M}\frac{1}{2}(j_i+k_i)^2]}{[M^{-1}\sum_{i=1}^{M}\frac{1}{2}(j_i^2+k_i^2)] - [M^{-1}\sum_{i=1}^{M}\frac{1}{2}(j_i+k_i)^2]}, \tag{6}$$

where j_i, k_i are the degrees of nodes at both the ends of the ith connection and M represents the total connections in the network.

Link betweenness centrality is defined for an undirected link as

$$\beta_L = \sum_{v\in V_s}\sum_{w\in V/v}\sigma_{vw}(e)/\sigma_{vw} \tag{7}$$

where $\sigma_{vw}(e)$ is the number of shortest paths between v and w that contain e, and σ_{vw} is the total number of shortest paths between v and w.[26]

The overlap of the neighborhood of two connected nodes i and j is defined as[26]

$$O_{ij} = \frac{n_{ij}}{(k_i-1)+(k_j-1)-n_{ij}} \tag{8}$$

where n_{ij} is the number of neighbors common to both nodes i and j. Here k_i and k_j represent the degree of the ith and j^{th} nodes.

Hierarchy can be defined as the heterogeneous distribution of local reaching centrality of nodes in the network. The local reaching centrality, (C_R), of a node i is defined as[25]

$$C_R(i) = \frac{1}{N-1}\sum_{j:0<d(i,j)<\infty}\frac{1}{d(i,j)}, \tag{9}$$

where $d(i,j)$ is the length of the shortest path between any pair of nodes i and j. The measure of hierarchy (h), termed as global reaching centrality is given by

$$h = \frac{\sum_{i\in V}[C_R^{max} - C_R(i)]}{N-1} \tag{10}$$

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. Gasch, A. P., Bret, A. P. & John, E. P. The Power of Natural Variation for Model Organism Biology. *Trends Genet.* **32** (3), 147-154 (2016).
2. Brem, R. B., Yvert, G., Clinton, R., & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296** (5568), 752-755 (2002).
3. Cubillos, F. A. *et al*. Assessing the complex architecture of polygenic traits in diverged yeast populations. *Mol. Ecol.* **20** (7), 1401-1413 (2011).
4. Ehrenreich, I. M. *et al*. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464**(7291), 1039-1042 (2010).
5. Thompson, D. A. & Francisco A. C. Natural gene expression variation studies in yeast. *Yeast* (2016).
6. Liti, G. *et al*. Population genomics of domestic and wild yeasts. *Nature* **458** (7236), 337-341 (2009).
7. Gagneur, J. *et al*. Genotype-environment interactions reveal causal pathways that mediate genetic effects on phenotype. *PLoS Genet.* **9** (9), e1003803 (2013).
8. Gupta S. *et al*. Temporal Expression Profiling Identifies Pathways Mediating Effect of Causal Variant on Phenotype. *PLoS Genet.* **11**, 1-23 (2015).
9. Skelly, D. A. *et al*. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res.* **23** (9), 1496-1504 (2013).
10. 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).
11. Kerr, M. K. & Churchill, G. A. Bootstrapping cluster analysis: assessing the reliability of conclusions from microarray experiments. *Proc. Natl. Acad. Sci. USA* **98**, 8961-8965 (2001).
12. Wakefield, J. C., Zhou, C. & Self, S. G. Modelling Gene Expression Data over Time: Curve Clustering with Informative Prior Distributions in *Bayesian Statistics*, Vol. 7 (eds Bernardo J. M. *et al*.) 721-732 (Oxford University Press, 2003).
13. McNicholas, P. D. & Murphy, T. B. Model-based clustering of longitudinal data. *Canad. J. Statist.* **38**, 153-168 (2010).
14. Roy, S., Bhattacharyya, D. K. & Kalita, J. K. Cobi: pattern based co-regulated biclustering of gene expression data. *Pattern Recogn. Lett.* **34**, 1669-1678 (2013).
15. Huang, S. Gene expression profiling, genetic networks, and cellular states: an integrating concept for tumorigenesis and drug discovery. *J. Mol. Med.* **77**, 469-480 (1999).
16. Jalan, S., Sarkar, C., Madhusudanan, A. & Dwivedi, S. K. Uncovering randomness and success in Society. *PLoS ONE* **9**, e88249 (2014).
17. Rai, A., Menon, A. V. & Jalan, S. Randomness and preserved patterns in cancer network. *Sci. Rep.* **4** (2014).
18. Shinde, P. Yadav, A., Rai, A. & Jalan, S. Dissortativity and duplications in oral cancer. *EPJ B* **88**, 1-7 (2015).
19. Boccaletti, S., Latora, V., Moreno, Y., Chavez, M. & Hwang, D. U. Complex networks: Structure and dynamics. *Phys. Rep.* **424**, 175-308 (2006).
20. Goltsev, A. V., Dorogovtsev, S. N., Oliveira, J. G., & Mendes, J. F.. Localization and spreading of diseases in complex networks. *Phys. Rev. Lett.* **109**(12), 128702 (2012).
21. Yadav, A. & Jalan, S. Origin and implications of zero degeneracy in networks spectra. *Chaos* **25**(4), 043110 (2015).
22. Albert, R. & Barabási, A. L. Statistical mechanics of complex networks. *Rev. Mod. Phys.* **74**, 47 (2002).
23. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* **5**, 101-113 (2004).
24. Jalan, S. & Yadav, A. Assortative and disassortative mixing investigated using the spectra of graphs. *Phys. Rev. E* **91** (1), 012813 (2015).
25. Mones, E., Vicsek, L. & Vicsek, T. Hierarchy measure for complex networks. *PLoS ONE* **7**, e33799 (2012).
26. Onnela, J.P. *et al*. Analysis of a large-scale weighted network of one-to-one human communication. *New J. Phys* **9**, 179 (2007).
27. Alon, U. Biological networks: the tinkerer as an engineer. *Science* **301**, 1866-1867 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

28. Goh, K.I., Cusick, M.E., Valle, D., Childs, B., Vidal, M. & Barabási, A.L. The human disease network. *Proc. Natl. Acad. Sci. USA* **104**, 8685-8690 (2007).

29. Neiman, A.M. Ascospore formation in the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **69**, 565-584 (2005).

30. Neiman A.M., Sporulation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **189**, 737-765 (2011).

31. Chu, S. *et al*. The Transcriptional Program of Sporulation in Budding Yeast. *Science* **282**, 699-705 (1998).

32. Primig, M. *et al*. The core meiotic transcriptome in budding yeasts. *Nat. Genet.* **26**, 415-423 (2000).

33. Keeney, S. Meiosis. Volume 1, molecular and genetic methods. Preface. *Methods in molecular biology* (Clifton, NJ) **557**, v-vi (2009).

34. Rabitsch, K. P. *et al*. A screen for genes required for meiosis and spore formation based on whole-genome expression. *Curr. Biol.* **11** (13), 1001-1009 (2001).

35. Deutschbauer, A. M., Williams, R. M., Chu, A. M., & Davis, R. W. Parallel phenotypic analysis of sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* **99** (24), 15530-15535 (2002).

36. Enyenihi, A. H. & Saunders, W. S. Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*. *Genetics* **163** (1), 47-54 (2003).

37. Nieduszynski, C. A. & Liti, G. From sequence to function: insights from natural variation in budding yeasts. *Biochim. Biophys. Acta (BBA)-General Subjects* **1810** (10), 959-966 (2011).

38. Deutschbauer, A. M. & Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* **37** (12), 1333-1340 (2005).

39. Ben-Ari, G. *et al*. Four linked genes participate in controlling sporulation efficiency in budding yeast. *PLoS Genet.* **2** (11), e195 (2006).

40. Gerke, J. P., Chen, C. T. & Cohen, B. A. Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* **174** (2), 985-997 (2006).

41. Gerke, J., Lorenz, K., & Cohen, B. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323** (5913), 498-501 (2009).

42. Lorenz, K. & Cohen, B. A. Causal variation in yeast sporulation tends to reside in a pathway bottleneck. *PLoS Genet.* **10** (9), e1004634 (2014).

43. Tsuchiya, D., Yang, Y. & Lacefield, S. Positive Feedback of NDT80 Expression Ensures Irreversible Meiotic Commitment in Budding Yeast. *PLoS Genet.* **10**, 1-15 (2014).

44. Gupta, S. *et al*. Meiotic Interactors of a Mitotic Gene TAO3 Revealed by Functional Analysis of its Rare Variant. *G3 Genes|Genomes|Genetics* doi:10.1534/g3.116.029900 (2016).

45. Mieczkowski, P. A. *et al*. Global analysis of the relationship between the binding of the Bas1p transcription factor and meiosis-specific double-strand DNA breaks in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **26**, 1014-1027 (2006).

46. Görner, W. *et al*. Nuclear localization of the C2H2 zinc finger protein Msn2p is regulated by stress and protein kinase A activity. *Gene. Dev.* **12**, 586-597 (1998).

47. Granek, J. A., Kayıkçı, Ö. & Magwene, P. M. Pleiotropic signaling pathways orchestrate yeast development. *Curr. Opin. Microbiol.* **14** (6), 676-681 (2011).

48. Newman, M. E. J. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).

49. Yook, S. H., Radicchi, F. & Meyer-Ortmanns, H. Self-similar scale-free networks and disassortativity. *Phys. Rev. E* **72**, 045105 (2005).

50. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910-913 (2002).

51. Watts, D. J. & Strogatz, S. H. Collective dynamics of 'small-world' networks. *Nature* **393**, 440-442 (1998).

52. Ravasz, E., Somera, A. L., Mongru, D. A., Oltvai, Z. N. & Barabási, A-L. Hierarchical organization of modularity in metabolic networks. *Science* **297**, 1551-1555 (2002).

53. Newman, M. E. J. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 016132 (2001).

54. Newman, M. E. J. The Structure and Function of Complex Networks. *SIAM Rev.* **45**, 167-256 (2003).

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

55. Merz, S. & Westermann, B. Genome-wide deletion mutant analysis reveals genes required for respiratory growth, mitochondrial genome maintenance and mitochondrial protein synthesis in *Saccharomyces cerevisiae*. *Genome Biol.* **10**, R95 (2009).

56. Ding, J. *et al*. Tolerance and stress response to ethanol in the yeast *Saccharomyces cerevisiae*. *Appl. Microbiol. Biotechnol.* **85**, 253-263 (2009).

57. Davey, H. M. *et al*. Genome-wide analysis of longevity in nutrient-deprived *Saccharomyces cerevisiae* reveals importance of recycling in maintaining cell viability. *Environ. Microbiol.* **14**, 1249-1260 (2012).

58. Horak, C. E. *et al*. Complex transcriptional circuitry at the G1/S transition in *Saccharomyces cerevisiae*. *Genes Dev.* **16** (23), 3017-3033 (2002).

59. Granovetter, M. S. The strength of weak ties. *Am. J. Sociol.* 1360-1380 (1973).

60. Szell, M. & Stefan, T. Measuring social dynamics in a massive multiplayer online game. *Soc. Net.* **32**, 313-329 (2010).

61. Sarkar, C., Yadav, A. & Jalan, S. Multilayer network decoding versatility and trust. *EPL* **113**, 18007 (2016).

62. Su, S. S. & Mitchell, A. P. Identification of functionally related genes that stimulate early meiotic gene expression in yeast. *Genetics* **133**, 67-77 (1993).

63. Honigberg, S. M. & Purnapatre, K. Signal pathway integration in the switch from the mitotic cell cycle to meiosis in yeast. *J. Cell Sci.* **116**, 2137-2147 (2003).

64. McDonald, C. M., Cooper, K. F. & Winter, E. The Ama1-directed anaphase-promoting complex regulates the Smk1 mitogen-activated protein kinase during meiosis in yeast. *Genetics* **171**, 901-911 (2005).

65. Rodriguez-Colman M. J. *et al*. The forkhead transcription factor Hcm1 promotes mitochondrial biogenesis and stress resistance in yeast. *J. Biol. Chem.* **285**, 37092-37101 (2010).

66. Sato, T. *et al*. The E-box DNA binding protein Sgc1p suppresses the gcr2 mutation, which is involved in transcriptional activation of glycolytic genes in *Saccharomyces cerevisiae*. *FEBS Lett.* **463**, 307-311 (1999).

67. Hanlon, S. E., Rizzo, J. M., Tatomer, D. C., Lieb, J. D. & Buck, M. J. The stress response factors Yap6, Cin5, Phd1, and Skn7 direct targeting of the conserved co-repressor Tup1-Ssn6 in *S. cerevisiae*. *PLoS ONE* **6**, e19060 (2011).

68. Leung, G. P., Lee, L., Schmidt, T. I., Shirahige, K. & Kobor, M. S. Rtt107 is required for recruitment of the SMC5/6 complex to DNA double strand breaks. *J. Biol. Chem.* **286**, 26250-26257 (2011)..

69. Beese, S. E., Negishi, T. & Levin, D. E. Identification of positive regulators of the yeast fps1 glycerol channel. *PLoS Genet.* **5**, e1000738 (2009).

70. Murugesapillai, D. *et al*. DNA bridging and looping by *HMO1* provides a mechanism for stabilizing nucleosome-free chromatin. *Nucleic Acids Res.* **42**, 8996-9004 (2014).

71. Marzluf, G. A. Genetic regulation of nitrogen metabolism in the fungi. *Microbiol. Mol. Biol. Rev.* **61**, 17-32 (1997).

72. Spellman, P. T. *et al*. Comprehensive identification of cell cycle–regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol. Biol. Cell* **9**, 3273-3297 (1998).

73. Lander, E. S. *et al*. Initial sequencing and analysis of the human genome. *Nature* **409**(6822), 860-921 (2001); Bassett, D. E., Boguski, M. S. & Hieter, P. Yeast genes and human disease. *Nature* **379**, 589-590 (1996).

74. Carter, H., Hofree, M., & Ideker, T. Genotype to phenotype via network analysis. *Curr. Opin. Genet. Dev.* **23**(6), 611-621 (2013).

75. Teixeira M. C. *et al.*. The YEASTRACT database: an upgraded information system for the analysis of gene and genomic transcription regulation in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* gkt1015, 1-6 (2013).

76. Lardenois A. *et al*. Execution of the meiotic noncoding RNA expression program and the onset of gametogenesis in yeast require the conserved exosome subunit Rrp6. *Proc. Natl. Acad. Sci. USA* **108**, 1058-1063 (2011).

77. Huber W., Heydebreck von A., Sültmann H., Poustka A. & Vingron M. Variance stabilization applied to microarray data calibration and to the quantification of differential expression. *Bioinformatics* **18** Suppl 1, S96-104 (2002).

78. Loader C. Locfit: Local regression, likelihood and density estimation. R package version 1.5-9.1 Merck, Kenilworth, N. J. (2013).

79. Cherry, J. M. *et al*. Saccharomyces Genome Database: the genomics resource of budding yeast. *Nucleic Acids Res.* gkr1029, 1-6 (2011).

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

80. Sarkar, C., Gupta, S., Sinha, H. & Jalan, S. Sporulation data: Fundamental principles governing sporulation efficiency: A network theory approach. *figshare* https://dx.doi.org/10.6084/m9.figshare.3457508.v1 (2016).

81. Newman, M. E. J., Strogatz, S. H. & Watts, D. J. Random graphs with arbitrary degree distributions and their applications. *Phys. Rev. E* **64**, 026118 (2001).

Acknowledgements

SJ acknowledges Department of Science and Technology (DST), Govt. of India grant EMR/2014/000368 and Council of Scientific and Industrial Research (CSIR), Govt. of India grant 25(0205)/12/EMR-II. HS acknowledges Tata Institute of Fundamental Research 12P-0120 intra-mural grant.

Author contributions statement

SJ conceived the idea. SJ and HS designed and supervised the project. CS constructed the networks and analyzed the structural properties. SG and CS analyzed the functional properties. All the authors wrote and approved the manuscript.

Additional Information

Supplementary Information

accompanies this paper at http://www.nature.com/naturescientificreports

Competing financial interests statement

The authors declare no competing financial interests.

bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure 1. (a) Schematic diagram for the sporulation process. Upon nutrient starvation, the yeast cell exits mitosis (cell division) and initiates meiosis and sporulation. This developmental process of meiosis and sporulation is divided in three phases: early, middle and late, with each stage having distinct functions and crucial genes being activated. (Adapted from Chu *et al.*[31]) (b) Heatmap of gene expression profiles of crucial meiosis and sporulation regulators in early, middle and late phases of sporulation in SK1 and S288c. Early, middle and late phases with their corresponding time points, demarcated by dashed lines, are defined based on SK1 profile. For SK1, early, middle and late regulators are indicated by orange, green and gray bars, respectively.

Figure 2. Comparison of temporal profiles structural properties of the sporulation networks of SK1 (red circles) and S288c (blue squares) across early, middle and late phases of sporulation. Early, middle and late phases, demarcated by dashed lines, are defined based on SK1 profile. (a) Average degree ($\langle k \rangle$), (b) average clustering coefficient ($\langle C \rangle$), (c) Pearson degree-degree correlation coefficient (r), (d) global reaching centrality (h).

Figure 3. Plots of degree (k) as a function of betweenness centrality (β_C) in (a) SK1 (red) and (b) S288c (blue) networks. Each dot represents a gene and genes with low degree but higher betweenness centrality in their respective time points are marked as black and named.

Table 1. Structural attributes of SK1 sporulation networks. N, N_c and N_{TF} respectively denote the network size, number of connections and number of transcription factors in a particular time point. Catalog of high degree nodes (degree mentioned in brackets) and degree of *NDT80* gene are provided for each time point of SK1 sporulation networks.

Time points	N	N_c	N_{TF}	High degree nodes	Degree of *NDT80*
T_1	360	529	13	*BAS1* (283), *RIM101* (39)	-
T_2	791	2107	34	*BAS1* (480), *MSN4* (300), *KAR4* (132), *FKH1* (109), *AFT2* (75)	-
T_3	1096	4033	46	*ASH1* (612), *BAS1* (560), *MSN4* (363), *HMS1* (168)	47
T_4	1495	6540	63	*ACE2* (1250), *ASH1* (770), *MSN4* (428), AFT_1 (369), *HMS1* (209)	69
T_5	1640	7720	73	*ASH1* (926), *MSN4* (499), AFT_1 (437), *GCR1* (399), *HMO1* (248)	87
T_6	1900	8978	77	*ASH1* (1074), *MSN4* (580), AFT_1 (486), *ISW2* (292)	101
T_7	1928	8012	75	*ASH1* (1118), AFT_1 (504), *ISW2* (309)	102
T_8	1886	8281	76	*ASH1* (1094), AFT_1 (484), *STE12* (432), *FHL1* (392), *ISW2* (299), *HMO1* (289)	95
T_9	2024	9213	74	*ACE2* (1700), *ASH1* (1039), AFT_1 (461), *STE12* (421), *FHL1* (362), *ISW2* (266)	87
T_{10}	1880	8410	69	*ACE2* (1579), *ASH1* (955), AFT_1 (435), *STE12* (402), *FHL1* (322), *ISW2* (243)	79
T_{11}	1711	7188	66	*ACE2* (1429), *ASH1* (870), AFT_1 (428), *STE12* (386), *FHL1* (291), *HMS1* (222)	69
T_{12}	1745	7206	61	*ACE2* (1472), *ASH1* (886), AFT_1 (481), *GAL4* (299), *HMS1* (246), *SIN4* (216), *INO4* (177)	77

Table 2. Structural attributes of S288c sporulation networks. N, N_c and N_{TF} respectively denote the network size, number of connections and number of transcription factors in a particular time point. Catalog of high degree nodes (degree mentioned in brackets) and degree of *IME1* gene are provided for each time point of S288c sporulation networks.

Time points	N	N_c	N_{TF}	High degree nodes	Degree of *IME1*
T_1	434	605	15	*BAS1* (356), *HAP4* (92), *TUP1* (48)	10
T_2	706	1327	24	*BAS1* (523), *STE12* (158), *HAP4* (134), *CUP2* (117), *TUP1* (72)	14
T_3	691	1280	26	*BAS1* (524), *HAP4* (136), *CUP2* (114), *TUP1* (76), PUT_3 (70)	13
T_4	569	1025	25	*BAS1* (442), *HAP4* (108), *RLM1* (93)	13
T_5	582	1049	24	*BAS1* (455), *HAP4* (112), *SWI5* (92)	15
T_6	415	733	23	*HMO1* (138), *HAP4* (125), *SWI5* (98)	16
T_7	409	686	18	*HMO1* (131), *HAP4* (124), *SWI5* (97)	19
T_8	513	1195	23	*MSN4* (261), *HSF1* (246), *HMO1* (129), *SWI5* (95), *INO4* (75)	21

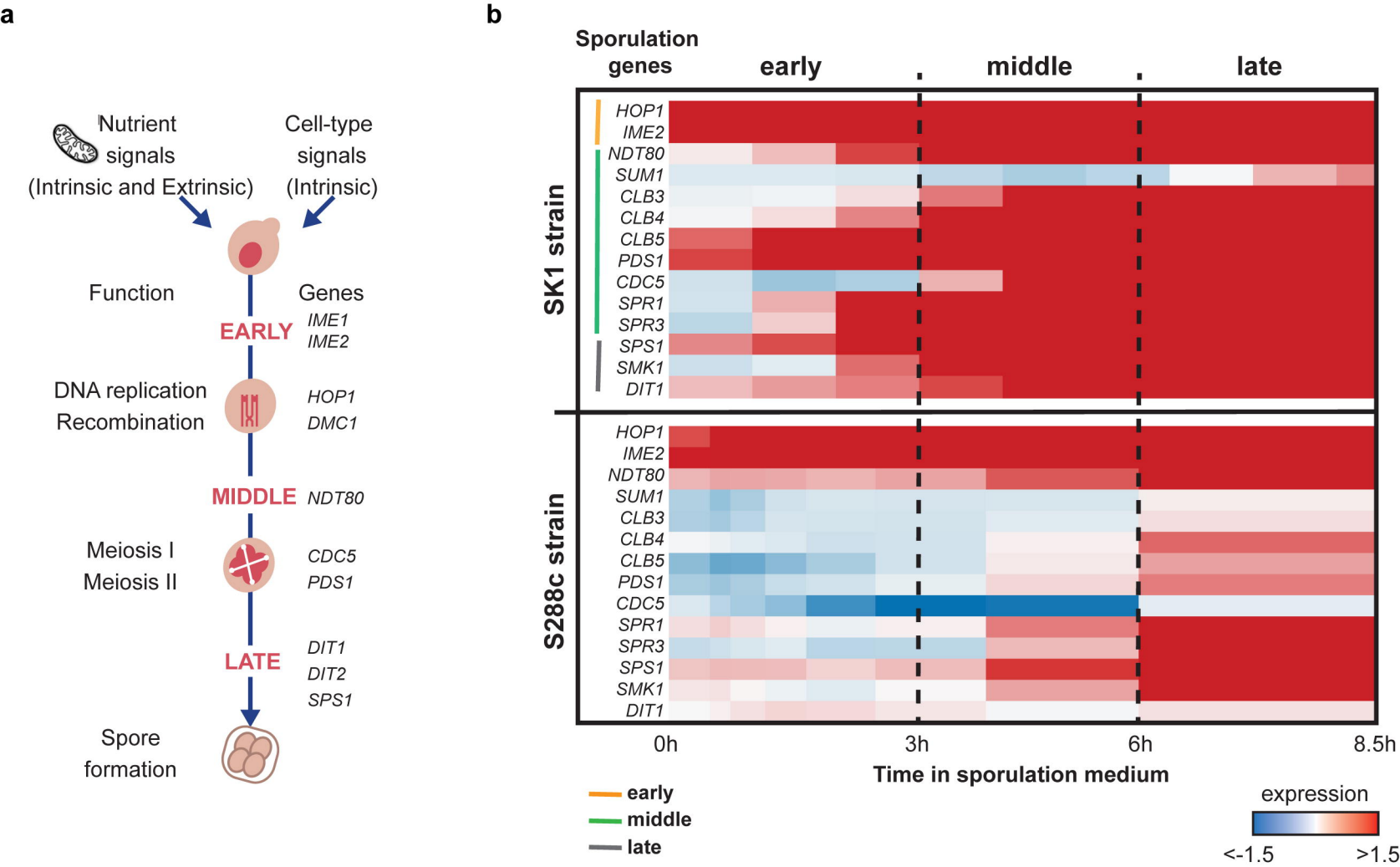
bioRxiv preprint doi: https://doi.org/10.1101/068270; this version posted November 30, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

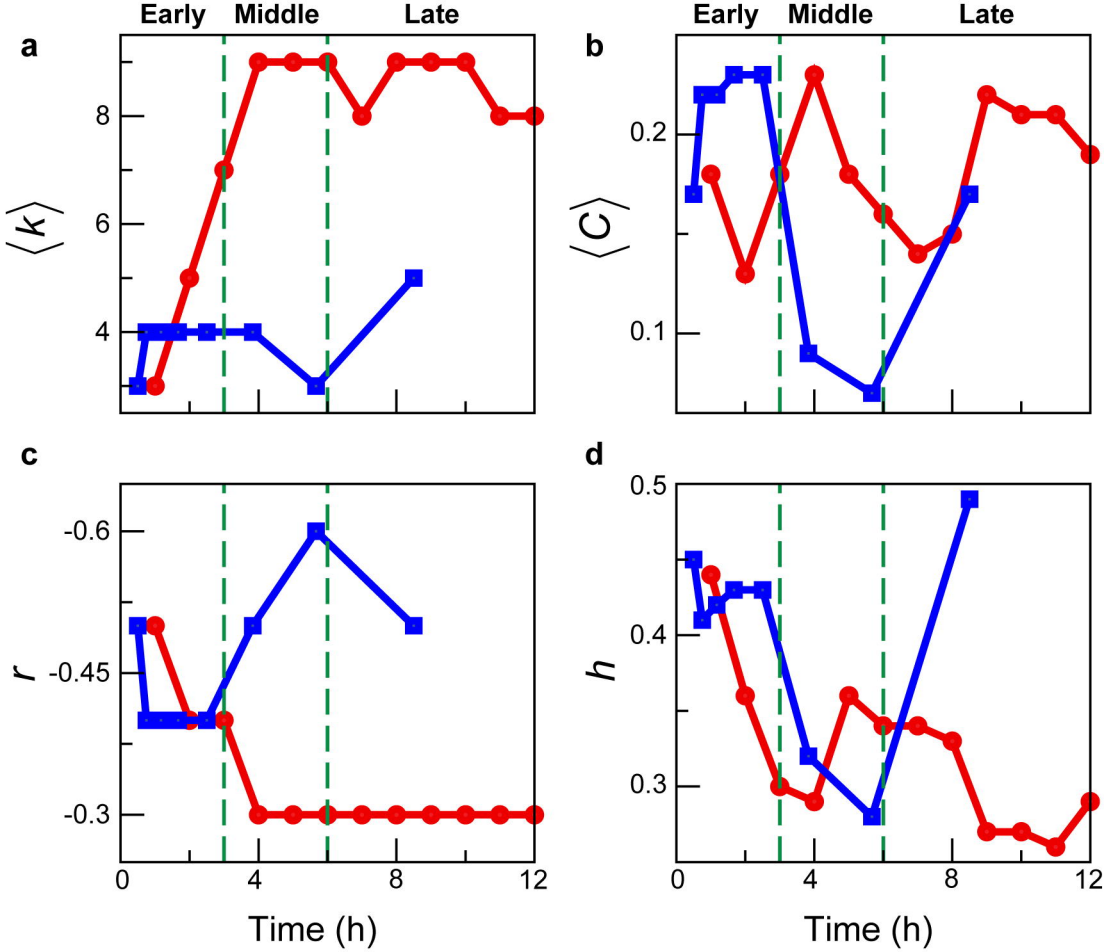
Table 3. Pairs of interacting genes which have low overlap (O) and high link betweenness centrality (β_L) in SK1 networks. Their corresponding indices as given in Supplementary Fig. S4.

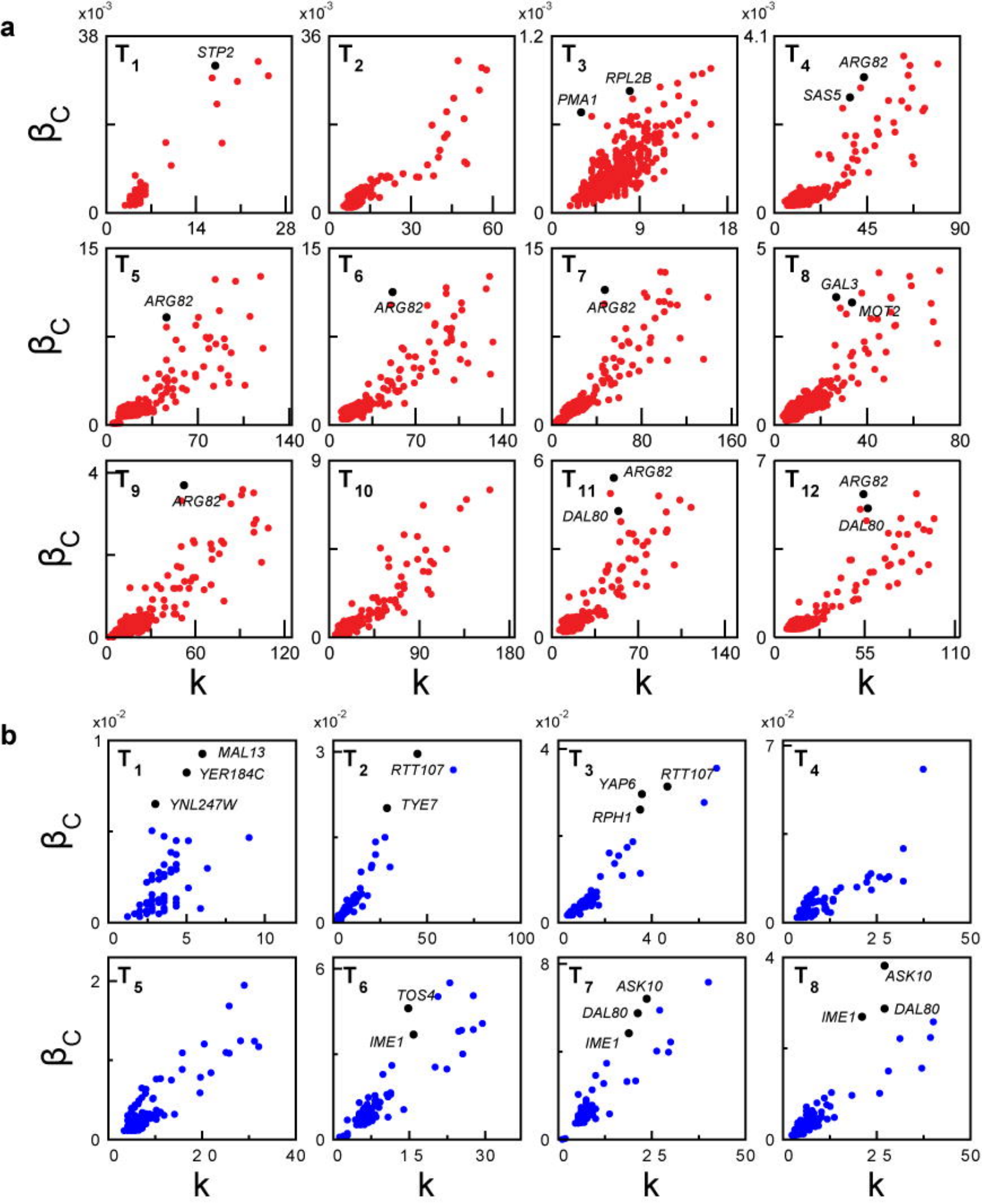
Time points	Gene pair
T_1	*STP2-BAS1*, *HCM1-BAS1*, *BAS1-RIM101*
T_9	*ACE2-MGA1*, *ACE2-SIR2*, *DAL81-ACE2*, *CDC14-ACE2*
T_{10}	*CDC14-ACE2*, *ACE2-ARG82*, *DAL81-ACE2*
T_{11}	*DAL81-ACE2*, *ACE2-MGA1*, *CDC14-ACE2*
T_{12}	*YPL056C-ARG82*, *DAL80-ACE2*, *RIF1-TPK1*

Table 4. Catalog of links having low overlap (O) yet relatively high link betweenness centrality (β_L) in each time point for S288c. Their corresponding indices as given in Supplementary Fig. S5.

Time points	Nodes with low O high β_L
T_1	*BAS1-RTT107*, *BAS1-TYE7*
T_2	*BAS1-RTT107*, *BAS1-TYE7*
T_3	*BAS1-RTT107*, *BAS1-RPH1*, *YAP6-BAS1*
T_4	*YAP6-BAS1*
T_6	*ASK10-HMO1*, *HMO1-XBP1*
T_7	*ASK10-HMO1*, *YAP6-HMO1*
T_8	*HSF1-IME1*, *ASK10-HMO1*


a
Nutrient signals (Intrinsic and Extrinsic)
Cell-type signals (Intrinsic)
Function
Genes
EARLY
IME1
IME2
DNA replication
Recombination
HOP1
DMC1
MIDDLE
NDT80
Meiosis I
Meiosis II
CDC5
PDS1
LATE
DIT1
DIT2
SPS1
Spore formation
b
Sporulation genes
early
middle
late
SK1 strain
S288c strain
HOP1
IME2
NDT80
SUM1
CLB3
CLB4
CLB5
PDS1
CDC5
SPR1
SPR3
SPS1
SMK1
DIT1
0h
3h
6h
8.5h
Time in sporulation medium
early
middle
late
expression
<-1.5
>1.5


Early
Middle
Late
a
b
c
d
⟨k⟩
⟨C⟩
r
h
8
6
4
0.2
0.1
-0.6
-0.45
-0.3
0.5
0.4
0.3
0
4
8
12
Time (h)
Time (h)


a
T1
STP2
T2
T3
RPL2B
PMA1
T4
ARG82
SAS5
T5
ARG82
T6
ARG82
T7
ARG82
T8
GAL3
MOT2
T9
ARG82
T10
T11
ARG82
DAL80
T12
ARG82
DAL80
βC
k
b
T1
MAL13
YER184C
YNL247W
T2
RTT107
TYE7
T3
YAP6
RTT107
RPH1
T4
T5
T6
TOS4
IME1
T7
ASK10
DAL80
IME1
T8
ASK10
IME1
DAL80
βC
k