Unravelling variant-to-function relationships in sporulation efficiency through integrative analysis of SNP-specific gene expression networks and metabolic models

Srijith Sasikumar[1,2,3], Pavan Kumar S[2,3,4,5], Nirav Bhatt[2,3,4,5,6,7], Himanshu Sinha[1,2,3,6,*]

[1] Systems Genetics Lab, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[2] Centre for Integrative Biology and Systems Medicine (IBSE), Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[3] Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[4] BioSystems Engineering and Control (BiSECt) Lab, Department of Biotechnology, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[5] pCoE for Network Systems Learning, Control and Evolution, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[6] Wadhwani School of Data Science and Artificial Intelligence (WSAI), Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

[7] Department of Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, Tamil Nadu, India

*Corresponding author: sinha@iitm.ac.in

ABSTRACT

Elucidating genotype-phenotype or variant-to-function relationships remains a challenge in quantitative genetics. For quantitative traits, causal SNPs act either additively or epistatically, resulting in complex interactions that are difficult to dissect molecularly. Here, we developed gene co-expression networks and genome-scale metabolic models for all combinations of causal SNPs of yeast sporulation efficiency, a quantitative trait, to determine how genetic interactions drive phenotypic variation. Analysis of expression networks identified SNP-SNP interaction-dependent changes in the connectivity of crucial metabolic regulators. Using an efficient thresholding and model extraction algorithm, Localgini+iMAT, we integrated the gene expression data of all SNP combinations in the yeast genome-scale metabolic model, resulting in 16 SNP-specific metabolic models. Genome-scale differential flux analysis was used to identify differentially activated metabolic reactions across several metabolic pathways in each SNP-specific model. This analysis revealed causal reactions in six major metabolic pathways, explaining the differences in observed sporulation efficiency. Finally, the differential regulation of the pentose phosphate pathway in specific SNP combinations suggested autophagy as a pentose phosphate pathway-dependent compensatory mechanism for increasing sporulation efficiency. Our study highlights how genetic polymorphisms in transcription factors can interact to modulate metabolic pathways in a well-studied complex trait in yeast, providing insights into genome-wide association studies (GWAS) variants of metabolic traits.

AUTHOR SUMMARY

In this study with yeast, we did an integrative analysis using multiple approaches to understand how different genetic variations work together to affect a quantitative trait, sporulation efficiency. We used a network of co-expressed genes, or genes that are turned on and off together, to identify how different genetic variations interact. We found from many combinations of variants that specific combinations can make key genes in metabolic pathways more or less active. This differential metabolic pathway activity ultimately affects the overall efficiency of sporulation. To further understand how these genetic changes modulate metabolism, we integrated gene expression data into a genome-scale metabolic model of yeast to simulate the metabolic pathways in yeast. By comparing the metabolic pathways of different genetic combinations, we identified specific reactions that were causal for the observed changes in sporulation efficiency. Overall, this study provides a new framework to understand how gene variations affect our metabolism and could be particularly useful for identifying the genetic causes of metabolic diseases in humans.

KEYWORDS

single nucleotide polymorphisms, genome-scale metabolic models, yeast sporulation, gene expression networks, quantitative trait, metabolism, genetic interactions.

INTRODUCTION

Genome-wide association studies (GWAS) have significantly advanced our understanding of complex traits by identifying causal single nucleotide polymorphisms (SNPs) and their intricate interactions [1,2]. These studies reveal that multiple SNPs, each exerting a modest impact on a trait, contribute to the complexity of genetic effects through additive and epistatic interactions. Unravelling this complexity within the broader genetic and biological network poses a considerable challenge, necessitating the development of robust computational tools and strategies to explore SNP interactions [3–6].

It is crucial to examine intermediate phenotypes to address the challenge of understanding how SNPs and their interactions influence complex traits, particularly the variant-to-function (V2F) relationship [7]. Various techniques, such as expression (eQTLs), protein (pQTLs), and metabolite (mQTLs) quantitative trait loci, impact diverse phenotypes [8,9]. In addition, SNPs influence gene co-expression and interaction networks, particularly in complex metabolic traits where coding and non-coding region variants can alter enzymatic reactions and metabolic flux [10].

Genome-scale metabolic models (GSMMs) provide a systems biology framework to describe cellular metabolism [11]. While previous studies have explored the direct impact of SNPs on enzyme coding regions [12,13], understanding how SNPs in both coding and non-coding regions, especially in transcription factors, collectively reshape cellular metabolism remains a gap to be addressed. Integrating SNP-specific gene expression alterations into GSMMs becomes essential to decipher the effects of these variants and their interactions on metabolic flux variations. This study aims to fill this gap by leveraging GSMMs and integrating gene expression data of causal SNPs and their combinations.

Sporulation is a metabolically driven process where yeast cells, in the presence of a non-fermentable carbon source and the absence of nitrogen, undergo meiosis, generating spores. The entire course of sporulation is regulated by five major metabolic pathways - glutamate synthesis, TCA cycle, glyoxylate cycle, gluconeogenesis, and glycogenolysis [14–16]. Sporulation efficiency, the percentage of cells sporulating in a culture in yeast, is a quantitative trait with several SNPs identified as causal and is highly heritable across yeast strains [17,18]. Focusing on yeast sporulation as a model known for its metabolic regulation, we investigate how causal SNPs and their combinations modulate metabolic fluxes to influence sporulation efficiency. The findings reveal that SNP combinations uniquely drive divergent sets of differentially expressed genes and alter the connectivity of critical metabolic regulators. Context-specific GSMMs for each SNP and their combinations illustrate changes in flux distribution patterns and identify synergistic and antagonistic metabolic pathway interactions in the yeast sporulation process. Our results identify a role for autophagy in yeast sporulation, which is dependent on the pentose phosphate pathway, in an SNP-specific manner. Hence, integrating gene expression data into GSMMs is an effective technique for deciphering the intricate relationships between SNPs and cellular metabolism. Interpreting

these relationships can pave the way for a more nuanced understanding of therapeutic interventions.

RESULTS

We analysed gene expression data from an allele replacement panel of *Saccharomyces cerevisiae* [16] during sporulation to investigate how SNPs and their combinations contribute to phenotypic variation. The SNPs studied here included coding variants in sporulation genes *IME1*^{L325M} and *RSF1*^{D181G} and non-coding variants in *IME1nc*$^{A\text{-}548G}$ and *RME1*$^{indel\text{-}308A}$ that conferred high sporulation efficiency phenotype. First, we created SNP-specific differential gene expression networks highlighting sporulation, metabolic processes, and regulators. Second, we employed a GSM modelling approach to understand how these SNPs influence intracellular metabolic fluxes. These models identified causal metabolic reactions that could explain increased phenotype by specific SNPs and their combination, providing insight into how the SNPs interact to modulate a quantitative trait phenotype. The workflow, from the pre-processing of RNA-Seq data to the application of GSM modelling techniques, is described in Figure 1.

Transcriptomic analysis reveals SNP-specific differentially expressed genes

Based on the sporulation efficiency data of the 16 allele replacement strains containing causal oak alleles in the vineyard background strains [16], we classified IME1+IME1nc+RME1+RSF1, IME1+IME1nc+RME1, IME1+IME1nc+RSF1, IME1+RME1+RSF1, IME1nc+RME1+RSF1 allelic combination strains as 'high sporulating.' All two allele combination strains (IME1+IME1nc, IME1+RME1, IME1+RSF1, IME1nc+RME1, IME1nc+RSF1, RME1+RSF1) were 'medium sporulating strains', and all single SNP and wildtype were considered 'low sporulating strains.' Differential gene expression analysis, after adjusting for the false discovery rate (< 10% using the Benjamini-Hochberg method), identified a total of 380 significant differentially expressed genes (DEGs, upregulated and downregulated) across 15 SNP combinations (Figure 2A, Table S1). A comparison of the number of DEGs in each SNP combination showed that the strain with *IME1* coding and *IME1nc* combination had the highest number of DEGs. Interestingly, of all the SNP combinations, we also found a significant downregulation of several genes in the strain with the two non-coding alleles, *RME1* and *IME1nc*.

Since we were interested in determining the number of sporulation-related genes among DEGs between wildtype and allele replacement strains, we compared 380 DEGs with 362 genes known to influence sporulation efficiency phenotype [17]. Of the 362 sporulation-specific genes, we identified 48 sporulation-specific DEGs across all SNP combinations. Each SNP combination had a different set of 48 sporulation-specific DEGs, with the strain with all four alleles having the maximum number (Figure 2B, Figure S1). The presence of *RME1nc* resulted in the downregulation of *RME1* expression in all combinations except for RME1nc+IME1nc (Figure S1). This showed that the cis effect of the *RME1nc* allele downregulated its expression, enhancing sporulation efficiency in oak strains. Upregulation

of the *IME1*, whenever the *RSF1* allele was in combination with either *IME1* or *IME1nc* (Figure S1), might be due to a trans-effect of the *RSF1* variation on the expression of *RIM4*, a major regulator of early sporulation genes and regulates *IME1* expression.

Interestingly, no common genes were found to be upregulated across all five high sporulating strains (RME1+RSF1+IME1+IME1nc, RME1+IME1+IME1nc, IME1+IME1nc, RSF1+IME1+RME1+RSF1, RME1+RSF1+IME1nc), even though several genes were uniquely activated in each SNP combination strain (Figure 2C). However, *IME2*, a meiosis initiator gene, was activated and uniquely upregulated when all four SNP were combined. We identified five upregulated genes, *HSP26*, *IME1*, *REC8*, *RIM4*, and *YIR016W*, common across the subset of strains with *IME1* and *RSF1* variants in combination (Figure 2D). In particular, *REC8*, a gene that mediates sister chromatid cohesion, homologous recombination, and chromosome synapsis [18], was uniquely upregulated only in this subset of strains (Figure 2E).

Since yeast sporulation is a metabolically driven process, we considered all metabolic genes in the Yeast8 metabolic model and investigated if they were differentially expressed across different SNP combinations. Of the 1150 metabolic genes, 164 were differentially expressed across all SNP combinations (Figure S2). The highest number of differentially expressed metabolic genes were observed in IME1+IME1nc (Figure 2B). Interestingly, we also identified that *SRT1*, a gene involved in dolichol biosynthesis and whose null mutant showed reduced sporulation efficiency, was differentially expressed whenever the *RSF1* allele was in combination with *IME1* and *IME1nc*, i.e., in RSF1+IME1+IME1nc and RSF1+IME1+IME1nc+RME1 (Figure S2).

Analysis of the upregulated DEGs revealed that the meiosis nuclear division category was significantly enriched in the strains with the following SNP combinations: RME1+RSF1+IME1+IME1nc, RME1+RSF1+IME1, and RSF1+IME1nc+IME1 indicating that these strains have already been undergoing meiosis during the second hour of sporulation (Figure S3). Further, we observed that two non-coding SNPs (*RME1* and *IME1nc*), when combined, could activate genes enriched for cellular respiration, which included *TAZ1* (mitochondrial ATP synthesis coupled electron transport), *TOK1* (stabilisation of membrane potential) and *COX9* (oxidative phosphorylation, Figure S3, Table S1). These findings suggested that SNP interactions could affect the timing and progression of meiosis during sporulation.

The connectivity of critical metabolic regulators changes with SNP interactions

Given that the strain with all SNP combinations had the highest sporulation efficiency, we examined how the connectivity of regulators that control the expression of DEGs in this strain varied across other combinations of SNPs. Regulatory impact factor (RIF1 and RIF2) analyses helped determine the TFs that control the expression of DEGs in the strain with all 4 SNP combinations and also assigned a score to these TFs [21]. In brief, the RIF1 metric classified TFs that were differentially co-expressed with highly abundant DEGs, and RIF2

was used to identify the TFs with the altered ability to predict the abundance of DEGs. Using the RIF1 and RIF2 *z*-score cut-offs, we obtained 15 TFs as regulators of DEG expression (Table S2). The RIF1 analysis identified *GCR1* (*z*-score = 2.40), the highest-value transcriptional activator of the glycolysis [19] gene. The top two candidate transcription factors from RIF2 analysis were *ROF1* (z-score = 2.22), a putative TF, and *BAS1* (*z*-score = 2.18), a Myb-related TF involved in regulating basal and induced expression of genes of the purine and histidine biosynthesis pathways [20]. Of these 15 regulators, we were able to identify seven metabolic regulators: *GCR1, GCR2* in glycolysis, *ARG80, LYS14*, and *MET32* in amino acid metabolism, *UPC2* involved in sterol biosynthesis, and *BAS1* in histidine biosynthesis. We then constructed 15 networks to identify differences in gene co-expression profiles among the wild type and 14 allelic combinations using the partial correlation and information theory (PCIT) algorithm. The correlation of 6,307 genes with a cut-off greater than |0.95| retrieved varying numbers of correlated pairs across wildtype and allele-specific networks.

Filtering for co-expressed gene pairs that were DEGs between wildtype and all oak allele combination groups (143 genes) along with the 15 TFs from the RIF analyses ($R^2 > |0.95|$) allowed the construction of 15 allele-specific sub-networks with varying numbers of nodes and edges (Table S3). These allele-specific sub-networks permitted the determination of how these TFs regulated gene expression by varying their connectivity across the SNPs and their combinations (Figure 3A). We found that each metabolic regulator has different connectivity across SNP combinations. In particular, we could only identify that the connectivity of the *UPC2* gene was higher for high and medium sporulating strains than for low sporulating strains (Figure 3B). This finding underscores the notion that a regulator can exert distinct influences on different SNP combinations, even if they result in the same phenotypic outcome. This necessitates the need to understand the critical metabolic alterations that could be responsible for sporulation efficiency variation between allele replacement strains.

Generation of context-specific metabolic models for each SNP and their combinations

As previously established, several genes and pathways undergo deregulation during yeast sporulation [21,22]. Furthermore, our analysis showed that the SNPs studied here could activate different genes (Figure 2A). Hence, extracting only the active genes and pathways from the generalised yeast metabolic model was crucial to reconstructing SNP-specific metabolic models. We generated context-specific metabolic models by integrating the gene expression data of SNPs and their combination using three model extraction methods: iMAT [26], INIT [27] and FASTCORE [28], which do not require a specified objective function. Additionally, we applied two thresholding methods, Standep [29] and LocalGini [30], which could capture more housekeeping reactions than other thresholding methods. This resulted in six combinations (LocalGini+iMAT, LocalGini+INIT, LocalGini+FASTCORE, Standep+iMAT, Standep+INIT and Standep+FASTCORE) to extract SNP-specific models (Figure 4A). A comparison of the number of reactions extracted by each specific combination of model extraction and thresholding method showed that three combinations, namely Local-Gini+IMAT, Local-Gini+FASTCORE, and Standep+FASTCORE effectively captured most

of the meiosis-specific reactions [14] (Figure 4B). Further, through a comparative enrichment analysis using a hypergeometric test, we found that models derived with the LocalGini+iMAT combination exhibited higher enrichment for meiosis-specific reactions (Figure 4C, D). Therefore, we proceeded with the SNP-specific models extracted using the LocalGini+iMAT combination to gain further biological insights.

Each resulting SNP-specific model (M1 to M16) corresponded to its genotype and varied in the number of genes, reactions, and metabolites (Table 1). The metabolic heterogeneity among the SNP-specific models was captured by calculating the Jaccard distance between each model (Figure S4).

Table 1: Number of reactions, metabolites, and genes in each SNP-specific model generated using LocalGini+iMAT thresholding. 'O' denotes the oak allele, and 'W' indicates the vineyard allele.

Models	**Genes**	**Reactions**	**Metabolites**	**Genotypes**
M1	603	1726	1227	$RME1^{O}RSF1^{O}IME1^{O}IME1nc^{O}$
M2	595	1745	1246	$RME1^{W}RSF1^{O}IME1^{O}IME1nc^{O}$
M3	598	1737	1240	$RME1^{O}RSF1^{W}IME1^{O}IME1nc^{O}$
M4	557	1737	1242	$RME1^{O}RSF1^{O}IME1^{W}IME1nc^{O}$
M5	569	1696	1209	$RME1^{O}RSF1^{O}IME1^{O}IME1nc^{W}$
M6	579	1717	1222	$RME1^{W}RSF1^{W}IME1^{O}IME1nc^{O}$
M7	595	1742	1238	$RME1^{W}RSF1^{O}IME1^{W}IME1nc^{O}$
M8	570	1604	1139	$RME1^{W}RSF1^{O}IME1^{O}IME1nc^{W}$
M9	602	1750	1247	$RME1^{O}RSF1^{W}IME1^{W}IME1nc^{O}$
M10	515	1560	1088	$RME1^{O}RSF1^{W}IME1^{O}IME1nc^{W}$
M11	594	1733	1233	$RME1^{O}RSF1^{O}IME1^{W}IME1nc^{W}$
M12	565	1644	1155	$RME1^{O}RSF1^{W}IME1^{W}IME1nc^{W}$
M13	591	1731	1237	$RME1^{W}RSF1^{O}IME1^{W}IME1nc^{W}$
M14	571	1720	1219	$RME1^{W}RSF1^{W}IME1^{O}IME1nc^{W}$
M15	504	1247	1012	$RME1^{W}RSF1^{W}IME1^{W}IME1nc^{O}$
M16	596	1728	1231	$RME1^{W}RSF1^{W}IME1^{W}IME1nc^{W}$

We identified 401 reactions common across all the extracted models. We hypothesised that these reactions form the core meiosis network, which was crucial for the survival and sporulation of yeast strains in an acetate medium. As expected, these reactions were significantly enriched with pathways that play essential roles in sporulation, which include carbon, glyoxylate and dicarboxylate, glycerolipid and glycerophospholipid metabolism, pentose phosphate pathway, and amino acids biosynthesis (Figure S5).

SNPs activate unique subsystems to modulate sporulation efficiency variation

Since each SNP-specific model had different metabolic reactions, we hypothesised that the observed sporulation efficiency variation could be either due to the activation of new reactions in each model or variations in intracellular flux within the meiosis core network.

Using the wildtype (M16) model as a baseline, a subnetwork topology analysis of each SNP model helped investigate the contribution of each SNP combination to the sporulation efficiency variation. This analysis revealed that each SNP model had uniquely enriched subsystems, a set of reactions involved sharing a common function obtained using the flux enrichment analysis function *(FEA.m)* provided in the COBRA toolbox (Table S4). For example, the model with *RSF1* SNP (M13) showed enrichment of valine, leucine, and isoleucine degradation pathways, glycolysis, gluconeogenesis, pyruvate metabolism pathways, and the TCA cycle (Figure 4A). Using the subnetwork topology analysis, we found metabolic pathways uniquely enriched for respiration in the M13 model (Figure 4A), which was consistent with the previous reports highlighting the role of the *RSF1* gene in respiration during sporulation. Furthermore, purine metabolism is uniquely enriched only in the M12 model (RME1nc, Figure 5A).

The *IME1* coding variant, which affects the interactions between Rim11 and Ume6, is crucial for initiating the sporulation [25,26]. The study by Sudarsanam and Cohen [16] revealed that *IME1* coding and non-coding variants contribute to sporulation efficiency variation, with the coding variant explaining a more significant proportion. A comparison of SNP-specific metabolic models, IME1 (M14) and IME1nc (M15), revealed significant differences in the metabolic activity of these two variants. Enrichment analysis of active reactions indicated that the *IME1* coding variant activated autophagy, one carbon pool by folate, phagosome, glycerolipid metabolism, and propionate metabolism pathways (Figure S6A). In contrast, the *IME1nc* did not trigger any of these metabolic pathways. This highlighted the metabolic plasticity of the functional and non-functional forms of the Ime1 protein.

Next, we investigated whether, in comparison to their individual effects, interactions between SNPs had the potential to activate or deactivate particular cellular pathways selectively. In other words, we investigated whether these interactions could result in synergistic or antagonistic effects on specific metabolic reactions. To test this idea, we examined the combination of two coding SNPs, RSF1+IME1 (M8) and two non-coding SNPs, RME1+IME1nc (M9), as well as their single SNP models, namely RSF1 (M13), IME1 (M14), RME1 (M12), and IME1nc (M15). By comparing reactions that were activated exclusively in the RSF1+IME1 (M8) model but not in the RSF1 (M13), IME1 (M14), and wildtype (M16) models, we identified 113 synergistic reactions. These reactions were notably enriched in glycerophospholipid, glycerolipid, and pyrimidine metabolism pathways (Figure 5B, Table S5). Conversely, when identifying reactions that were inactive in the M8 model but active in the M13, M14, and M16 models, we determined 271 antagonistic reactions. These antagonistic reactions were enriched in various subsystems, including fatty acid biosynthesis, lipid metabolism, the phosphatidyl inositol signalling pathway, steroid biosynthesis, etc.

(Table S5, Figure S7). Similarly, when comparing the RME1 (M12), IME1nc (M15), and RME1+IME1nc (M9) models, we uncovered the combined role of *RME1* and *IME1nc* in synergistically modulating the one carbon pool by folate subsystems, encompassing the glycine cleavage complex and tetrahydrofolate aminomethyl transferase (Figure S8, Table S6).

Genome-scale differential flux analysis reveals the impact of SNP-SNP interactions on intracellular flux variation

The analysis of the unique sub-network topology showed that combinations of SNPs could activate shared metabolic pathways. However, whether SNPs could modify the flux distribution patterns of these pathways to modulate phenotypic variation was unclear. Therefore, flux variations in common reactions among SNP and wildtype models were analysed to investigate flux distribution patterns. While the flux balance analysis (FBA) [27] for a given constraint can find multiple potential flux values for reactions in a metabolic model, it cannot ensure the uniqueness of a solution. To overcome this limitation in FBA, we used flux sampling to assess variations in metabolic flux distribution using the genome-scale differential flux analysis (GS-DFA). We identified upregulated (Table S7) and downregulated (Table S8) reactions in each SNP model compared to the wildtype model. GS-DFA analysis showed that the upregulated reactions were primarily enriched in the six subsystems involved in the biosynthesis of amino acids, antibiotics, secondary metabolites, glycerophospholipid metabolism, glycerolipid metabolism, and pentose phosphate pathway in a strain with all four SNP combinations (M1, Figure 6B, Table S9).

As M1 had the highest sporulation efficiency, we checked if these six subsystems were present in the other high sporulation efficiency strains (M2-M5). We found that in M2-M5, the six subsystems were varying subsets of the M1 model, highlighting intracellular flux variation among these pathways as significant contributors to sporulation efficiency variation among these strains (Figure 6B). The two subsystems, glycerophospholipid and glycerolipid metabolism, present in all models suggested that they were essential for the overall fitness of the yeast strains during sporulation.

Further, we explored how the functional and non-functional forms of Ime1 protein could change the flux distribution patterns to modulate sporulation efficiency variation. From the GS-DFA analysis, we found that reactions involved in glycerolipid, glycerophospholipid metabolism, and biosynthesis of secondary metabolites were upregulated only in IME1+IME1nc (M6) and IME1 (M14) models but not in IME1nc (M15) model (Figure S6B). Interestingly, we could also predict the metabolic consequence of the non-functional form of *IME1* (M15), which was to upregulate the reactions involved in the peroxisome, fatty acid metabolism, fatty acid degradation, and fatty acid biosynthesis (Figure S7B).

Upon hierarchically clustering the flux change (FC) values of upregulated reactions in the M1 model across other models, we found that models were clustered into two groups (Figure S9). The presence or absence of the *IME1nc* variant primarily determined this clustering. Further,

We found that SNP combinations without the *IME1nc* variant, i.e., M5, M8, M10, M11, M12, M14, and the M1 models (which contains all four SNPs), led to an increased flux through reactions involved in the pentose phosphate pathway (PPP, Table 2, Table S9). This result was consistent with our flux variability analysis (Supplementary Methods section) [28], where we found that the PPP reactions were upregulated in these models (Figure S10). The increased flux through PPP reactions was inversely coupled with the glucose-6-phosphate isomerase reaction. This indicated that the flux was redirected through the PPP for biosynthesis rather than glycolysis for energy generation (Figure 6C). In highly sporulating SNP-specific models M2, M3, and M4, which had *IME1nc* variant and where the PPP was not upregulated (Figure 6C), showed enrichment of reactions involved in autophagy and phagosomes (Figure 5B, 6A). These findings were interesting since previous studies have established a correlation between autophagy and yeast sporulation, and the deletion of several autophagy-related genes has shown reduced sporulation [29,30]. Moreover, we found that the upregulation of reactions involved in amino acid biosynthesis was observed only in the M1, M4, and M11 models (Table S5), highlighting the effect of SNP combination *RME1* and *RSF1* and their interaction with *IME1nc* variants on flux distribution patterns. These findings demonstrated the significant impact of SNPs and their interactions on the metabolic flux distribution, explaining how these interactions affected sporulation efficiency.

Table 2: Differential regulation of pentose phosphate pathway subsystem across SNP models in an *IME1nc*-dependent manner (highlighted in bold).

PPP status	**Models**
PPP normal	RME1+**IME1nc** IME1+**IME1nc** RME1+IME1+**IME1nc** RME1+RSF1+**IME1nc** IME1+RSF1+**IME1nc** RSF1 **IME1nc** RSF1+**IME1nc**
PPP upregulation	RME1 IME1 RME1+IME1 RME1+RSF1 IME1+RSF1 RME1+RSF1+IME1 RME1+RSF1+IME1+IME1nc

DISCUSSION

Here, we attempted to bridge the V2F challenge for a well-studied complex trait, sporulation efficiency in yeast. By analysing the gene expression data from an allele replacement panel of four SNPs, we showed how SNPs interacted and modulated the metabolic pathways to drive

sporulation efficiency variation. We showed that certain SNP combinations could activate or deactivate the flux through a particular reaction. We identified that nucleotide metabolism (purine and pyrimidine metabolism) was synergistically modulated, while steroid biosynthesis was antagonistically modulated when coding variants of *RSF1* and *IME1* were combined. These insights gave us a deeper understanding of how SNP-SNP interactions could redirect metabolic flux, prioritising nucleotide synthesis over steroid synthesis in the early stages of meiotic progression during sporulation. Previous studies have underscored the pivotal role of pyrimidine metabolism in facilitating efficient meiotic progression [15]. Further, the SNP-specific modulation of the glycine cleavage system was established in this study, where the SNPs of *RME1* and *IME1nc* were synergistic.

We characterised the intracellular metabolic flux states of each SNP-specific model without requiring an objective function as previously described [31,32]. Using GS-DFA analysis, we identified flux variation in six major pathways as significant contributors to sporulation efficiency variation. Interestingly, our GS-DFA analysis did not identify the fundamental reactions that were known to constitute the meiosis-specific metabolic model [14]. These included the TCA cycle, glyoxylate cycle, and acetate uptake, which have been shown to influence sporulation at an early stage through gene-knockout approaches. However, we were able to capture these pathways in our core meiosis-specific network by comparing the common reactions shared among SNP-specific models. These findings suggested that SNPs and their interactions modulate sporulation efficiency variation by influencing metabolic reactions beyond the essential ones required for sporulation. This highlights the specific effects of SNP levels, which are usually masked in gene deletion studies. Our analysis also predicted the differential regulation of the pentose phosphate pathway in an *IME1nc*-specific manner. The pentose phosphate pathway is critical in supplying the cell with the necessary building blocks and reducing equivalents needed for the biosynthesis of nucleotides and amino acids [33]. Our hypothesis suggested that autophagy could produce the essential precursors for nucleotide and amino acid biosynthesis by degrading and recycling cellular components compensating for PPP, particularly in models where PPP upregulation was absent (Figure 6D). Previous studies have shown autophagy as a compensatory mechanism for the PPP in tumour cells [34]. It would be interesting to experimentally validate whether autophagosomes are induced when the PPP reaction is downregulated. The reactions identified as upregulated or downregulated based on flux change do not necessarily correspond to differential gene expression. These changes can also be attributed to the post-transcriptional regulation of genes. Thus, integrating SNP and their combination-specific gene expression data in GSMM provided several novel mechanisms by which phenotypic variation in a well-studied trait was regulated at the molecular level.

Several studies have reported complex genetic interactions between identified SNPs in metabolic diseases [35,36]. A recent study has used GWAS variant data to develop personalised organ-specific metabolic models for 524,615 individuals of the INTERVAL and UK Biobank cohorts to clarify the impact of genetic variations on the metabolic processes implicated in coronary artery disease [37]. Studying the population-level data by clustering based on variant combinations rather than phenotype along with GS-DFA analysis can

provide additional novel targets for therapeutic interventions in complex human metabolic diseases. Applying SNP-specific genome-scale metabolic models that integrate transcriptomic and metabolomic data will help identify known and novel pathways perturbed by SNPs and their combinations. With scRNASeq becoming common and generated for diseases such as cancers, single-cell GSMM [38,39] can open a new level of metabolic heterogeneity corresponding to cellular heterogeneity. Combining this with the power of scRNASeq with context-specific GSMM can provide novel insights into cellular heterogeneity and uncover causes of disease, drug response, drug resistance, etc. Therefore, our approach to studying SNP-specific effects using GSMM can provide candidate genes for further validation in the molecular processes and phenotype-genotype relationship, leading to clearer variant-to-function pathways.

METHODS

RNA-Seq data, pre-processing, and differential expression analysis

Raw RNA sequence read data for single SNP and their combinations was downloaded from NCBI GEO (accession number GSE55409) [16] for 16 isogenic allele replacement strains of the oak alleles of four SNPs *IME1*, *RME1*, *RSF1*, and *IME1nc*. The overview of the dataset is discussed in Supplementary Methods.

The reads were converted to Fastq format using the command 'Fastqdump' in the SRA toolkit. The FastQC and MutiQC were performed for quality control [40]. The reads were mapped to reference *Saccharomyces cerevisiae* R64-1-1.105 downloaded from the Ensembl database (https://ftp.ensembl.org/pub/release110/fasta/saccharomyces_cerevisiae/cdna/Saccharomyces_cerevisiae.R64-1-1.cdna.all.fa.gz) using Salmon (an alignment-free tool), accounting for GC and sequence bias [41]. The transcript TPM values obtained from Salmon were converted to gene-level counts using Tximport v1.4.0 [50]. The effect of day on the gene expression values was removed using ComBatseq [51].

To identify the genes that are differentially expressed between the wildtype and each allele replacement strain, we performed differential gene expression analysis using DESeq2 [52].In brief, the batch-corrected untransformed raw counts were given as input to the DESeq2. After filtering for low counts with less than 100 counts in all 63 samples, 6305 genes out of 6571 were selected for further processing. The effect of the day was included in the design model of DESeq2 to account for the contribution of covariates to the total gene expression. log2 fold changes (LFC) shrinkage using the alpegm method [53], with a false discovery rate of 10% (p-value adjusted using the Benjamini-Hochberg method), was used to identify differentially expressed genes (DEGs). Genes with LFC values greater than 0.5 and less than -0.5 are considered upregulated and down-regulated, respectively. The upregulated DEGs obtained for each allele replacement strain are given as input to ClueGO [54] from Cytoscape [47] for gene ontology and functional annotations analysis.

Regulatory impact factor (RIF) analysis

The CeTF package in R was utilised to perform the Regulatory Impact Factors analysis [48,49]. The Variance stabilising transformation (VST) was performed on the gene expression data of wildtype and strain with four SNPs. It was given as input for the RIF analysis as it required two contrasting conditions. The transcription factors (TFs), downloaded from the Yeastract database (http://www.yeastract.com), [50] were contrasted to the list of DEGs identified by comparing the wildtype strain (i.e., a strain with no SNPs) to the strain with all SNP combinations. RIF analysis classifies transcription factors (TFs) into two categories: RIF1 and RIF2. RIF1 focuses on TFs consistently exhibiting the strongest differential co-expression with highly abundant and differentially expressed genes. In contrast, RIF2 identifies TFs that provide predictive evidence for changes in the abundance of DE genes. The TFs identified using the RIF analysis with scores deviating ± 1.96 SD from the mean were considered significant and termed 'critical regulators.'

PCIT and connectivity analysis

The Partial Correlation and Information Theory (PCIT) approach [51] was used to arrive at co-expression networks for wildtype and 14 SNP combinations except for strain with *IME1*, which had only three replicates. PCIT evaluated all possible triplets of genes, identifying the strongly associated pair after screening all the network's genes. The correlation of |0.95| and based on the p-value < 0.05 was used as a threshold for arriving at a significant association (this threshold was chosen to arrive at a maximum number of significant edges) [52]. Further, the DEGs (specific for wild type vs. four SNP combination strains) and critical regulators identified using the RIF analysis are filtered to arrive at sub-networks for each SNP-specific network. The individual co-expression sub-networks for wildtype and 14 allele replacement strains were visualised using Cytoscape [47] version (3.9.1). Each subnetwork was analysed for network properties using the Cytoscape plugin NetworkAnalyser 4.4.8. The connectivity for each critical regulator was identified by dividing the degree of that regulator in any specific network with the highest degree in that network [59,60]. The connectivity of each critical regulator was then compared across each SNP combination and was visualised using the Pheatmap package in R (v1.0.12, https://CRAN.R-project.org/package=pheatmap).

An evaluation of model extraction and thresholding methods for reconstructing SNP-specific metabolic models

Yeast 8.4.2 [61], a commonly used yeast GSM model, was used in this study (https://github.com/sysbiochalmers/yeast-gem). This GSMM had 1,150 genes, 4,058 reactions, and 2,742 metabolites. The bounds of exchange reactions of the GSMM were changed to appropriately reflect the early stages of the sporulation state. For this, we restricted the intake of glucose and nitrogen (Lower bound = Upper bound = 0 [mmol/(g DW h)]) while allowing an unrestricted supply of oxygen and acetate (Lower bound = -1000 [mmol/(g DW h)], 14).

The gene expression data (TPM values) averaged across replicates of wild type and 15 allele replacement strains were pre-processed using Localgini thresholding [30] and Standep [29] and integrated into Yeast 8.4.2 using the model extraction methods (MEMs), iMAT [26], INIT [27] and FASTCORE [28]. The detailed methodology for reconstructing an SNP-specific model using six combinations of thresholding methods and MEMs is described in Supplementary Methods. This resulted in the reconstruction of 15 SNP-specific (M1 to M15) and wildtype (M16) models for each thresholding and MEM combination. Further, we evaluated the performance of each method for its ability to capture most of the meiosis-specific reactions [14] in each SNP-specific model. Then, we employed a hypergeometric test to compare the enrichment of the meiosis-specific reactions in each method. The models generated from the optimal thresholding and MEM identified in our evaluation were used for further analysis.

To understand the metabolic heterogeneity of each extracted model, we identified the Jaccard distance between each SNP-specific model. Further, we identified reactions in each SNP-specific model that were not in the wildtype model. These reactions, active in SNP models while not in wildtype, were searched for significantly enriched subsystems using the flux enrichment analysis function $(FEA.m)$ provided in the COBRA toolbox [62]. In brief, $FEA.m$ used a hypergeometric 1-sided test (α = 0.05) and FDR correction (adjusted p-value < 0.05) to obtain significantly enriched subsystems.

Genome-scale differential flux analysis (GS-DFA) using a flux sampling approach

To discern how each SNP and their combinations can modulate the intracellular fluxes to show sporulation efficiency variation, we leveraged the random flux sampling approach, which acquires a significant number of evenly distributed solutions over the whole solution space and does not require an objective function. We adopted a previously developed protocol [32] to identify the most significantly dysregulated reactions in the presence of a particular SNP and their combination by comparing the wildtype (M16) and each SNP model (M1 to M15). However, we modified the pipeline to use an optGpSampler [58] instead of an ACHR sampler. This modification was based on previous studies that demonstrated the optGpSampler's advantage in faster convergence and reduced run time compared to the ACHR sampler [64]. The optGpSampler in COBRApy version 0.10.1[65] was used in this study to sample 10000 flux solutions, and the thinning factor was set to 100 for all models. This was to ensure the spanning of the whole flux space and to obtain sparse and uncorrelated fluxes [32]. Let X_{SNP} and $X_{Wildtype}$ be the flux distributions of a particular reaction in SNP-specific and wildtype models, respectively. We identified the significantly deregulated reactions in each SNP model compared to the wild type using a two-sided Kolmogorov-Smirnov test with a significance level of 0.05 (adjusted p-value using the Benjamini Hochberg method). Upregulated and downregulated reactions were identified from these dysregulated reactions using the flux change value obtained. The reactions with flux change greater than 0.82 were considered upregulated, and flux change less than -0.82 were considered downregulated, indicating ten times higher or lower fold change in flux in the SNP model compared to the wildtype model.

The flux change was calculated for each reaction as follows:

$$Flux\ change = \frac{\bar{X}_{SNP} - \bar{X}_{Wildtype}}{\left|\bar{X}_{SNP} + \bar{X}_{Wildtype}\right|}$$

where, $\bar{X}_{SNP}$ and $\bar{X}_{Wildtype}$, the arithmetic mean of the flux distributions for a given reaction in SNP and wildtype conditions, respectively.

COBRA toolbox and COBRApy setup

COBRA Toolbox version 3.0.0 and COBRApy version 0.10.1 on MATLAB R2021b and Python version 3.5.2, respectively. The optimiser used for linear programming and mixed-integer linear programming optimisation was Gurobi, version 8.0.0. optGpSampler was run using parallel processes. We used customised Python scripts for statistical analysis and visualisation of flux sampling results.

CODE AVAILABILITY

The codes used in this paper are available in the following GitHub repository (https://github.com/HimanshuLab/Modelling-and-transcriptomics-analysis-of-QTNs), along with all generated models.

ACKNOWLEDGEMENTS

S.S. and P.K.S. are supported by a fellowship from the Indian Institute of Technology Madras; P.K.S. was also supported by a grant from the Centre for Integrative Biology and Systems Medicine (IBSE), IIT Madras (BIO/18-19/304/ALUM/KARH) to H.S. We thank Dr Karthik Raman for critically reading our manuscript.

AUTHOR CONTRIBUTIONS

Conceptualisation: S.S., and H.S.; Formal Analysis: S.S., P.K.S., and N.B.; Investigation: S.S., P.K.S., N.B., and H.S.; Writing original draft: S.S. and H.S.; Supervision: H.S.; All authors reviewed and approved the final paper.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Visscher PM, Wray NR, Zhang Q, Sklar P, McCarthy MI, Brown MA, et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. American Journal of Human Genetics. 2017;101: 5–22. doi:10.1016/j.ajhg.2017.06.005
2. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nature Reviews Genetics. 2019;20: 467–484. doi:10.1038/s41576-019-0127-1
3. Cordell HJ. Detecting gene-gene interactions that underlie human diseases. 2009;10: 392–404. doi:10.1038/nrg2579
4. Martin MP, Gao X, Lee J-H, Nelson GW, Detels R, Goedert JJ, et al. Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. 2002;31: 429–434. doi:10.1038/ng934
5. Wan X, Yang C, Yang Q, Xue H, Fan X, Tang NLS, et al. BOOST: A Fast Approach to Detecting Gene-Gene Interactions in Genome-wide Case-Control Studies. 2010;87: 325–340. doi:10.1016/j.ajhg.2010.07.021
6. Fang G, Wang W, Paunic V, Heydari H, Costanzo M, Liu X, et al. Discovering genetic interactions bridging pathways in genome-wide association studies. Nature Communications. 2019;10. doi:10.1038/s41467-019-12131-7
7. Cano-gamez E, Trynka G. From GWAS to Function : Using Functional Genomics to Identify the Mechanisms Underlying Complex Diseases. 2020;11: 1–21. doi:10.3389/fgene.2020.00424
8. Battle A, Khan Z, Wang SH, Mitrano A, Ford MJ, Pritchard JK, et al. Genomic variation. Impact of regulatory variation from RNA to protein. Science. 2015;347: 664–667. doi:10.1126/science.1260793
9. Olayinka OA, O'Neill NK, Farrer LA, Wang G, Zhang X. Molecular Quantitative Trait Locus Mapping in Human Complex Diseases. Curr Protoc. 2022;2: e426. doi:10.1002/cpz1.426
10. Tong H, Küken A, Razaghi-Moghadam Z, Nikoloski Z. Characterization of effects of genetic variants via genome-scale metabolic modelling. Cell Mol Life Sci. 2021;78: 5123–5138. doi:10.1007/s00018-021-03844-4
11. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale metabolic models. Genome Biol. 2019;20: 121. doi:10.1186/s13059-019-1730-3
12. Øyås O, Borrell S, Trauner A, Zimmermann M, Feldmann J, Liphardt T, et al. Model-based integration of genomics and metabolomics reveals SNP functionality in Mycobacterium tuberculosis. Proceedings of the National Academy of Sciences. 2020;117: 8494–8502. doi:10.1073/pnas.1915551117
13. Sarkar D, Maranas CD. SNPeffect: identifying functional roles of SNPs using metabolic networks. The Plant Journal. 2020;103: 512–531. doi:10.1111/tpj.14746
14. Ray D, Ye P. Characterization of the Metabolic Requirements in Yeast Meiosis Plos One. 2013;8. doi:10.1371/journal.pone.0063707
15. Walther T, Létisse F, Peyriga L, Alkim C, Liu Y, Lardenois A, et al. Developmental stage-dependent metabolic regulation during meiotic differentiation in budding yeast. BMC Biol. 2014;12: 1–22. doi:10.1186/s12915-014-0060-x

16. Sudarsanam P, Cohen BA. Single Nucleotide Variants in Transcription Factors Associate More Tightly with Phenotype than with Gene Expression. PLoS Genetics. 2014;10. doi:10.1371/journal.pgen.1004325
17. Deutschbauer AM, Davis RW. Quantitative trait loci mapped to single-nucleotide resolution in yeast. Nature Genetics. 2005;37: 1333–1340. doi:10.1038/ng1674
18. Gerke JP, Chen CTL, Cohen BA. Natural isolates of Saccharomyces cerevisiae display complex genetic variation in sporulation efficiency. Genetics. 2006;174: 985–997. doi:10.1534/genetics.106.058453
19. Deutschbauer AM, Williams RM, Chu AM, Davis RW. Parallel phenotypic analysis of sporulation and postgermination growth in Saccharomyces cerevisiae. Proceedings of the National Academy of Sciences of the United States of America. 2002;99: 15530–15535. doi:10.1073/pnas.202604399
20. Yoon S-W, Lee M-S, Xaver M, Zhang L, Hong S-G, Kong Y-J, et al. Meiotic prophase roles of Rec8 in crossover recombination and chromosome structure. Nucleic Acids Research . 2016;44: 9296–9314. doi:10.1093/nar/gkw682
21. Reverter A, Hudson NJ, Nagaraj SH, Pérez-Enciso M, Dalrymple BP. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. Bioinformatics. 2010;26: 896–904. doi:10.1093/bioinformatics/btq051
22. Holland MJ, Yokoi T, Holland JP, Myambo K, Innis MA. The GCR1 Gene Encodes a Positive Transcriptional Regulator of the Enolase and Glyceraldehyde-3-Phosphate Dehydrogenase Gene Families in Saccharomyces cerevisiae. Molecular and Cellular Biology. 1987;7: 813–820. doi:10.1128/mcb.7.2.813-820.1987
23. Daignan-Fornier B, Fink GR. Coregulation of purine and histidine biosynthesis by the transcriptional activators BAS1 and BAS2. Proceedings of the National Academy of Sciences of the United States of America. 1992;89: 6746–6750. doi:10.1073/pnas.89.15.6746
24. Primig M, Williams RM, Winzeler EA, Tevzadze GG, Conway AR, Hwang SY, et al. The core meiotic transcriptome in budding yeasts. Nature Genetics. 2000;26: 415–423. doi:10.1038/82539
25. Enyenihi AH, Saunders WS. Large-scale functional genomic analysis of sporulation and meiosis in Saccharomyces cerevisiae. Genetics. 2003;163: 47–54.
26. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. Bioinformatics. 2010;26: 3140–3142. doi:10.1093/bioinformatics/btq602
27. Agren R, Bordel S, Mardinoglu A, Pornputtapong N, Nookaew I, Nielsen J. Reconstruction of Genome-Scale Active Metabolic Networks for 69 Human Cell Types and 16 Cancer Types Using INIT. PLOS Computational Biology. 2012;8: e1002518. doi:10.1371/journal.pcbi.1002518
28. Vlassis N, Pacheco MP, Sauter T. Fast Reconstruction of Compact Context-Specific Metabolic Network Models. PLOS Computational Biology. 2014;10: e1003424. doi:10.1371/journal.pcbi.1003424
29. Joshi CJ, Schinn S-M, Richelle A, Shamie I, O'Rourke EJ, Lewis NE. StanDep: Capturing transcriptomic variability improves context-specific metabolic models. PLOS Computational Biology. 2020;16: e1007764. doi:10.1371/journal.pcbi.1007764

30. Pavan Kumar S, Nirav Bhatt. Localgini: A method for harnessing inequality in gene expression to improve the quality of context-specific models. bioRxiv. 2023; 2023.09.14.557840. doi:10.1101/2023.09.14.557840
31. Lu L, Roberts G, Simon K, Yu J, Hudson AP. Rsf1p, a protein required for respiratory growth of Saccharomyces cerevisiae. Curr Genet. 2003;43: 263–272. doi:10.1007/s00294-003-0398-z
32. Roberts III GG, Hudson AP. Rsf1p is required for an efficient metabolic shift from fermentative to glycerol-based respiratory growth in S. cerevisiae. Yeast. 2009;26: 95–110. doi:10.1002/yea.1655
33. 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. Molecular Cell Biology. 1996;16: 2518–2526. doi:10.1128/MCB.16.5.2518
34. Gertz J, Gerke JP, Cohen BA. Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. Theoretical Population Biology. 2010;77: 1–5. doi:10.1016/j.tpb.2009.10.002
35. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? Nature Biotechnology. 2010;28: 245–248. doi:10.1038/nbt.1614
36. Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. BMC Bioinformatics. 2010;11: 489. doi:10.1186/1471-2105-11-489
37. Barve G, Manjithaya R. Cross-talk between autophagy and sporulation in Saccharomyces cerevisiae. Yeast. 2021;38: 401–413. doi:10.1002/yea.3556
38. Wen F-P, Guo Y-S, Hu Y, Liu W-X, Wang Q, Wang Y-T, et al. Distinct temporal requirements for autophagy and the proteasome in yeast meiosis. Autophagy. 2016;12: 671–688. doi:10.1080/15548627.2016.1149659
39. Scott WT, Smid EJ, Block DE, Notebaart RA. Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts. Microbial Cell Factories. 2021;20: 204. doi:10.1186/s12934-021-01694-0
40. Nanda P, Ghosh A. Genome Scale-Differential Flux Analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection. PLoS Computational Biology. 2021;17: e1008860. doi:10.1371/journal.pcbi.1008860
41. Bertels L-K, Fernández Murillo L, Heinisch JJ. The Pentose Phosphate Pathway in Yeasts–More Than a Poor Cousin of Glycolysis. Biomolecules. 2021;11: 725. doi:10.3390/biom11050725
42. Parkhitko AA, Priolo C, Coloff JL, Yun J, Wu JJ, Mizumura K, et al. Autophagy-Dependent Metabolic Reprogramming Sensitizes TSC2-Deficient Cells to the Antimetabolite 6-Aminonicotinamide. Molecular Cancer Research. 2014;12: 48–57. doi:10.1158/1541-7786.MCR-13-0258-T
43. Lin E, Kuo P-H, Liu Y-L, Yang AC, Kao C-F, Tsai S-J. Association and interaction of APOA5, BUD13, CETP, LIPA and health-related behavior with metabolic syndrome in a Taiwanese population. Scientific Reports. 2016;6: 36830. doi:10.1038/srep36830
44. Cadena-López RO, Hernández-Rodríguez LV, Aguilar-Galarza A, García-Muñoz W, Haddad-Talancón L, Anzures-Cortes M de L, et al. Association between SNPs in Leptin

Pathway Genes and Anthropometric, Biochemical, and Dietary Markers Related to Obesity. Genes. 2022;13: 945. doi:10.3390/genes13060945
45. Foguet C, Xu Y, Ritchie SC, Lambert SA, Persyn E, Nath AP, et al. Genetically personalised organ-specific metabolic models in health and disease. Nature Communications. 2022;13: 7356. doi:10.1038/s41467-022-35017-7
46. Gustafsson J, Anton M, Roshanzamir F, Jörnsten R, Kerkhoven EJ, Robinson JL, et al. Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. Proceedings of the National Academy of Sciences. 2023;120: e2217868120. doi:10.1073/pnas.2217868120
47. Huang Y, Mohanty V, Dede M, Tsai K, Daher M, Li L, et al. Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux. Nature Communications. 2023;14: 4883. doi:10.1038/s41467-023-40457-w
48. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016;32: 3047–3048. doi:10.1093/bioinformatics/btw354
49. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nature Methods. 2017;14: 417–419. doi:10.1038/nmeth.4197
50. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Research; 2016. doi:10.12688/f1000research.7563.1
51. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. NAR Genomics and Bioinformatics. 2020;2: lqaa078. doi:10.1093/nargab/lqaa078
52. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology. 2014;15: 550. doi:10.1186/s13059-014-0550-8
53. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics. 2019;35: 2084–2092. doi:10.1093/bioinformatics/bty895
54. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25: 1091–1093. doi:10.1093/bioinformatics/btp101
55. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. Genome Research. 2003;13: 2498–2504. doi:10.1101/gr.1239303
56. Oliveira de Biagi CA, Nociti RP, Brotto DB, Funicheli BO, Cássia Ruy P de, Bianchi Ximenez JP, et al. CeTF: an R/Bioconductor package for transcription factor co-expression networks using regulatory impact factors (RIF) and partial correlation and information (PCIT) analysis. BMC Genomics. 2021;22: 1–8. doi:10.1186/s12864-021-07918-2
57. Teixeira MC, Viana R, Palma M, Oliveira J, Galocha M, Mota MN, et al. YEASTRACT+: a portal for the exploitation of global transcription regulation and

metabolic model data in yeast biotechnology and pathogenesis. Nucleic Acids Research. 2023;51: D785–D791. doi:10.1093/nar/gkac1041

58. Watson-Haigh NS, Kadarmideen HN, Reverter A. PCIT: An R package for weighted gene co-expression networks based on partial correlation and information theory approaches. Bioinformatics. 2009;26: 411–413. doi:10.1093/bioinformatics/btp674
59. Banerjee P, Balraj P, Ambhore NS, Wicher SA, Britt RD, Pabelick CM, et al. Network and co-expression analysis of airway smooth muscle cell transcriptome delineates potential gene signatures in asthma. Scientific Reports. 2021;11: 1–16. doi:10.1038/s41598-021-93845-x
60. Lau LY, Reverter A, Hudson NJ, Naval-sanchez M, Fortes MRS, Alexandre PA. Dynamics of Gene Co-expression Networks in Time-Series Data : A Case Study in Drosophila melanogaster Embryogenesis. 2020;11. doi:10.3389/fgene.2020.00517
61. Lu H, Li F, Sánchez BJ, Zhu Z, Li G, Domenzain I, et al. A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. 2019;10: 3586. doi:10.1038/s41467-019-11581-3
62. Heirendt L, Arreckx S, Pfau T, Mendoza SN, Richelle A, Heinken A, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. 2019;14: 639–702. doi:10.1038/s41596-018-0098-2
63. Megchelenbrink W, Huynen M, Marchiori E. optGpSampler: An Improved Tool for Uniformly Sampling the Solution-Space of Genome-Scale Metabolic Networks. PLoS One. 2014;9: e86587. doi:10.1371/journal.pone.0086587
64. Herrmann HA, Dyson BC, Vass L, Johnson GN, Schwartz J-M. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. 2019;5: 1–8. doi:10.1038/s41540-019-0109-0
65. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and Analysis for Python. BMC Systems Biology. 2013;7: 74. doi:10.1186/1752-0509-7-74

FIGURE LEGENDS

Figure 1: Schematic representation outlines the key steps and methodologies employed in our study

The raw transcriptomic data of 16 allele replacement panels were pre-processed and corrected for batch effects using CombatSeq. Differentially expressed genes were obtained using DeSeq2, and the critical regulators were obtained from RIF analysis and further studied for connectivity using PCIT analysis. Integration of transcriptomic data into the genome-scale metabolic model of yeast was performed using Localgini+iMAT to reconstruct SNP-specific metabolic models. The change in intracellular fluxes between the wildtype and the SNP models was studied using GS-DFA analysis. Figure 1 was created with BioRender.com.

Figure 2: Transcriptional differences between the allele replacement strains

(A) Number of upregulated and downregulated genes (LFC > 0.5 and LFC < -0.5) identified using DeSeq2 between wildtype vs. each SNP combination. (B) The number of differentially expressed ((LFC > 0.5 and LFC < 0.5) sporulation-specific and metabolic genes identified using DeSeq2 between wildtype vs. each SNP. (C) Venn diagram plot for upregulated genes identified using DeSEq2 with LFC > 0.5 and p-adj > 0.1 for three SNP combination strains (RME1+RSF1+IME1, RME1+IME1+IME1nc, RSF1+IME1+IME1nc and RME1+RSF1+IME1nc) and four SNP combination strain (RME1+RSF1+IME1+IME1nc) strains in comparison with the wildtype. (D) Venn diagram analysis for *RSF1* and *IME1* SNP combinations. (E) Heatmap representing the LFC values of *IME1*, *RIM4*, and *REC8* genes across all SNP combination strains compared to the wildtype strain

Figure 3: Analysis of connectivity of critical metabolic regulators

(A) Connectivity of critical metabolic regulators identified using RIF analysis across combinations of SNPs. (B) Comparing the connectivity of critical metabolic regulators based on the sporulation efficiency of the strains. Significance was calculated using a t-test (ns represents non-significant).

Figure 4: Localgini+iMAT shows higher enrichment for meiosis-specific reactions

(A) Number of reactions extracted from each MEM and thresholding method; (B) Number of meiosis-specific reactions extracted from each MEM and thresholding method; (C) Heatmap representing the p-values obtained using a hypergeometric test for models enriched for meiosis-specific reactions; (D) Distribution of p-values obtained using a hypergeometric test for models enriched for meiosis-specific reactions

Figure 5: SNPs and their interaction can activate unique reactions and subsystems

(A) Metabolic subsystems enrichment analysis for reactions that are over-represented in each comparison that are present exclusively in single SNP models compared with the wildtype

model. (B) Synergistic effect of *RSF1* and *IME1* in modulating nucleotide metabolism. Here, the violin plots represent the flux distribution obtained by optGpSampler for key reactions involved in purine and pyrimidine metabolism compared across wildtype (M16), RSF1+IME1 (M8), RSF1 (M13) and IME1 (M14) models.

Figure 6: Pentose phosphate pathway is differentially regulated in SNP models

(A) Flux enrichment plots for unique reactions in highly sporulating strains compared to the wildtype. (B) Flux enrichment plots for upregulated reactions in highly sporulating strains compared to the wildtype strain obtained after GF-DFA analysis. (C) Box plots representing the flux distribution of key reactions involved in the pentose phosphate pathway obtained from optGpSampler for three and four SNP models (M1 to M5) compared to the wildtype (M16) model (D) Proposed mechanism of autophagy as a pentose-phosphate pathway-dependent compensatory mechanism during the early phase of yeast sporulation (created with Biorender.com).

SUPPLEMENTARY METHODS

Supplementary methods accompany this manuscript.

SUPPLEMENTARY TABLE LEGENDS

Table S1: Differentially expressed genes identified using DESeq2 between wildtype and 15 allele replacement strains.

Table S2: The critical regulators identified by regulatory impact factor analysis. The metabolic regulators are shown in bold.

Table S3: Number of nodes and edges in wildtype and SNP-specific networks before and after filtering DEGs and RIFs. We could not construct networks for the strain with the IME1 variant due to the limited number of replicates to perform PCIT analysis.

Table S4: Subsystems enriched for unique reactions in each SNP model compared to wildtype.

Table S5: Synergistic and antagonistic reactions and their subsystem enrichment for RSF1+IME1 (M8) SNP combination.

Table S6: Synergistic and antagonistic reactions and their subsystem enrichment for RME1+IME1nc (M9) SNP combination.

Table S7: List of upregulated reactions identified by GS-DFA analysis for all SNP models compared with wildtype model.

Table S8: List of downregulated reactions identified by GS-DFA analysis for all SNP models compared with wildtype model.

Table S9: Subsystems enriched for upregulated reactions in each SNP model compared to wildtype identified using GS-DFA analysis.

SUPPLEMENTARY FIGURE LEGENDS

Figure S1: Clustering of differentially expressed sporulation-related genes based on LFC values for each allele replacement strain comparison with the wildtype strain.

Figure S2: Clustering of differentially expressed metabolic genes based on the LFC values for each allele replacement strain comparison with the wildtype strain.

Figure S3: Over-represented pathways of upregulated differentially expressed genes between wildtype and (A) RME1+RSF1+IME1+IME1nc, (B) RME1+RSF1+IME1, (C) RSF1+IME1+IME1nc, (D) RME1+IME1nc and (E) IME1+IME1nc strains. The circle size

for a specific GO Term (pathway) varies with the order of significance (p-value) level. The most prominent node represents the most significant pathway. The pathways were constructed using Cytoscape v3.8.0 (https://cytoscape.org/) using the plugin ClueGO v2.5.41.

Figure S4: The Jaccard distance between all the generated context-specific models based on the differences in reactions in the model.

Figure S5: Enriched subsystems (pathways) for the core reactions in all extracted SNP-specific models.

Figure S6: Metabolic subsystems enrichment analysis reveals the metabolic role of *IME1* functional and non-functional forms. (A) Flux enrichment analysis for reactions that are exclusively in each model in comparison with the wildtype (B) Flux enrichment analysis of upregulated reactions obtained using GS-DFA analysis between wildtype and IME1, IME1nc and IME1+IME1nc strains. The presented reactions have an FC value (Flux Change) greater than 0.8 (as described in materials and methods) and have an adjusted p-value less than 0.05.

Figure S7: Antagonistic effect of *RSF1* and *IME1* in modulating 'Steroid biosynthesis'. The violin plots represent the flux distribution obtained by optGpSampler for key reactions involved in **Steroid biosynthesis** compared across wildtype (M16), RSF1+IME1 (M8), RSF1 (M13) and IME1 (M14) models.

Figure S8: Synergistic effect of *RME1* and *IME1nc* in modulating 'one carbon pool by folate' metabolism. The violin plots represent the flux distribution obtained by optGpSampler for key reactions involved in **one carbon pool by folate metabolism** compared across wildtype (M16), RME1+IME1nc (M9), RME1 (M12) and IME1c (M15) models.

Figure S9: Hierarchical clustering of FC values of upregulated reactions in the M1 model across other models shows two specific groups. The presented reactions have an FC value greater than 0.8 in M1 models compared to the wildtype model (M16). The FC values have been scaled for better representation.

Figure S10: (A) Enriched pathways (subsystems) in each SNP model were revealed by Flux Variability analysis. The presented reactions have a flux span ratio (as described in materials and methods) greater than two and have an adjusted p-value less than 0.05. (B) Clustering of FSR values of the upregulated reactions specific for the M1 model that are compared across all SNP models.

Figure 1

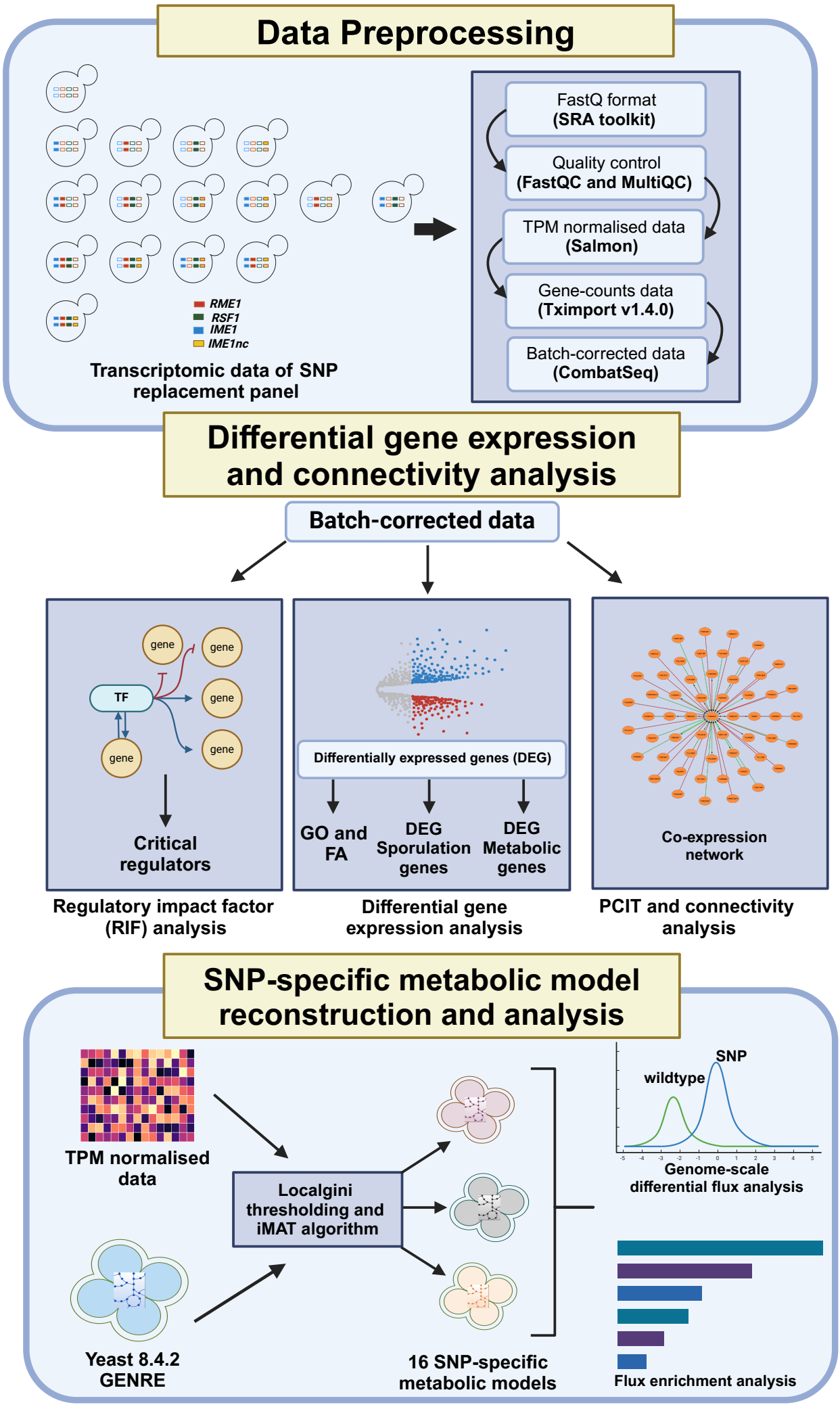

Data Preprocessing
RME1
RSF1
IME1
IME1nc
Transcriptomic data of SNP replacement panel
FastQ format (SRA toolkit)
Quality control (FastQC and MultiQC)
TPM normalised data (Salmon)
Gene-counts data (Tximport v1.4.0)
Batch-corrected data (CombatSeq)
Differential gene expression and connectivity analysis
Batch-corrected data
gene
TF
Critical regulators
Regulatory impact factor (RIF) analysis
Differentially expressed genes (DEG)
GO and FA
DEG Sporulation genes
DEG Metabolic genes
Differential gene expression analysis
Co-expression network
PCIT and connectivity analysis
SNP-specific metabolic model reconstruction and analysis
TPM normalised data
Localgini thresholding and iMAT algorithm
Yeast 8.4.2 GENRE
16 SNP-specific metabolic models
wildtype
SNP
Genome-scale differential flux analysis
Flux enrichment analysis

Figure 2

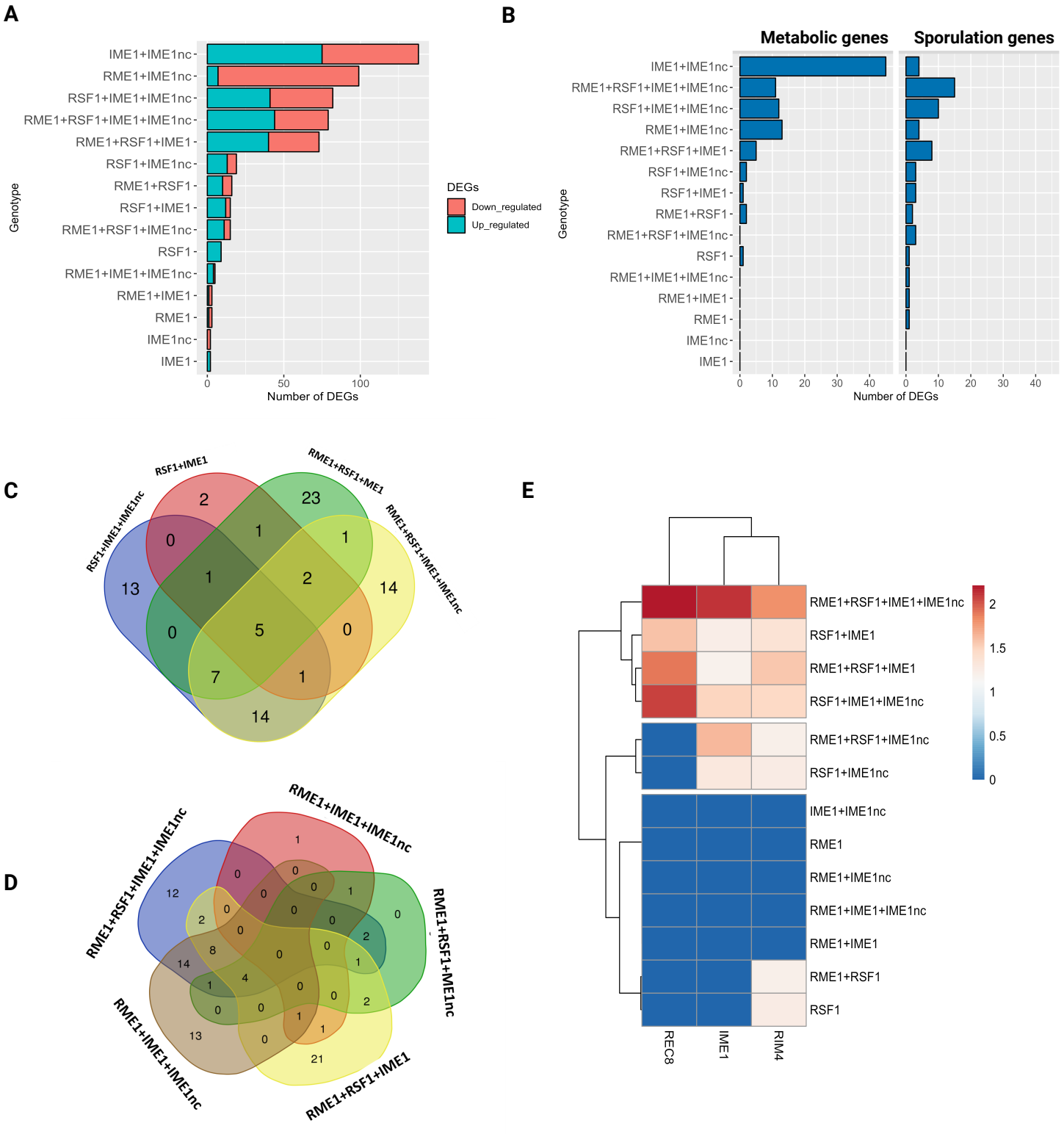

A
IME1+IME1nc
RME1+IME1nc
RSF1+IME1+IME1nc
RME1+RSF1+IME1+IME1nc
RME1+RSF1+IME1
RSF1+IME1nc
RME1+RSF1
RSF1+IME1
RME1+RSF1+IME1nc
RSF1
RME1+IME1+IME1nc
RME1+IME1
RME1
IME1nc
IME1
Genotype
0
50
100
Number of DEGs
DEGs
Down_regulated
Up_regulated
B
Metabolic genes
Sporulation genes
IME1+IME1nc
RME1+RSF1+IME1+IME1nc
RSF1+IME1+IME1nc
RME1+IME1nc
RME1+RSF1+IME1
RSF1+IME1nc
RSF1+IME1
RME1+RSF1
RME1+RSF1+IME1nc
RSF1
RME1+IME1+IME1nc
RME1+IME1
RME1
IME1nc
IME1
Genotype
0
10
20
30
40
Number of DEGs
C
RSF1+IME1
RME1+RSF1+ME1
RSF1+IME1+IME1nc
RME1+RSF1+IME1+IME1nc
2
23
0
1
1
13
1
2
14
0
5
0
7
1
14
D
RME1+IME1+IME1nc
RME1+RSF1+IME1+IME1nc
RME1+RSF1+ME1nc
RME1+IME1+IME1nc
RME1+RSF1+IME1
12
14
13
21
E
RME1+RSF1+IME1+IME1nc
RSF1+IME1
RME1+RSF1+IME1
RSF1+IME1+IME1nc
RME1+RSF1+IME1nc
RSF1+IME1nc
IME1+IME1nc
RME1
RME1+IME1nc
RME1+IME1+IME1nc
RME1+IME1
RME1+RSF1
RSF1
REC8
IME1
RIM4
2
1.5
1
0.5
0

Figure 3

A

	LYS4	GCR1	MET32	ARG80	GCR2	UPC2	BAS1
N1	0.36	0.54	0.53	0.41	0.58	0.73	0.51
N2	0.91	0.77	0.75	0.55	0.75	0.68	0.51
N3	0.36	0.53	0.42	1.00	0.74	0.89	0.39
N4	0.80	0.84	0.76	0.90	0.57	0.84	0.65
N5	0.57	0.57	0.64	0.47	0.62	0.62	0.57
N6	0.38	0.95	0.53	0.60	0.63	0.80	0.58
N7	0.76	0.80	0.50	0.72	0.60	0.92	0.64
N8	0.71	0.44	0.63	0.47	0.42	0.94	0.47
N9	0.72	0.70	0.83	0.74	0.74	0.67	0.56
N10	0.78	0.80	0.87	0.61	0.80	0.83	0.69
N11	0.62	0.40	0.68	0.37	0.72	0.62	0.90
N12	0.53	0.69	0.66	0.73	0.66	0.83	0.68
N13	0.45	0.47	0.50	0.52	0.42	0.55	0.73
N15	0.68	0.61	0.61	0.49	0.54	0.51	0.56
N16	0.79	0.71	0.53	0.85	0.36	0.50	0.28

Connectivity: 1, 0.9, 0.8, 0.7, 0.6, 0.5, 0.4, 0.3

Metabolic regulators: LYS4, GCR1, MET32, ARG80, GCR2, UPC2, BAS1

B

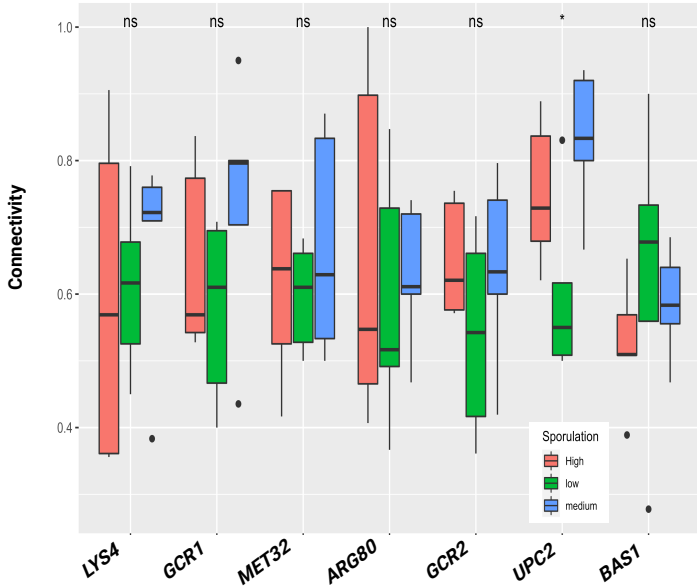


Figure 4

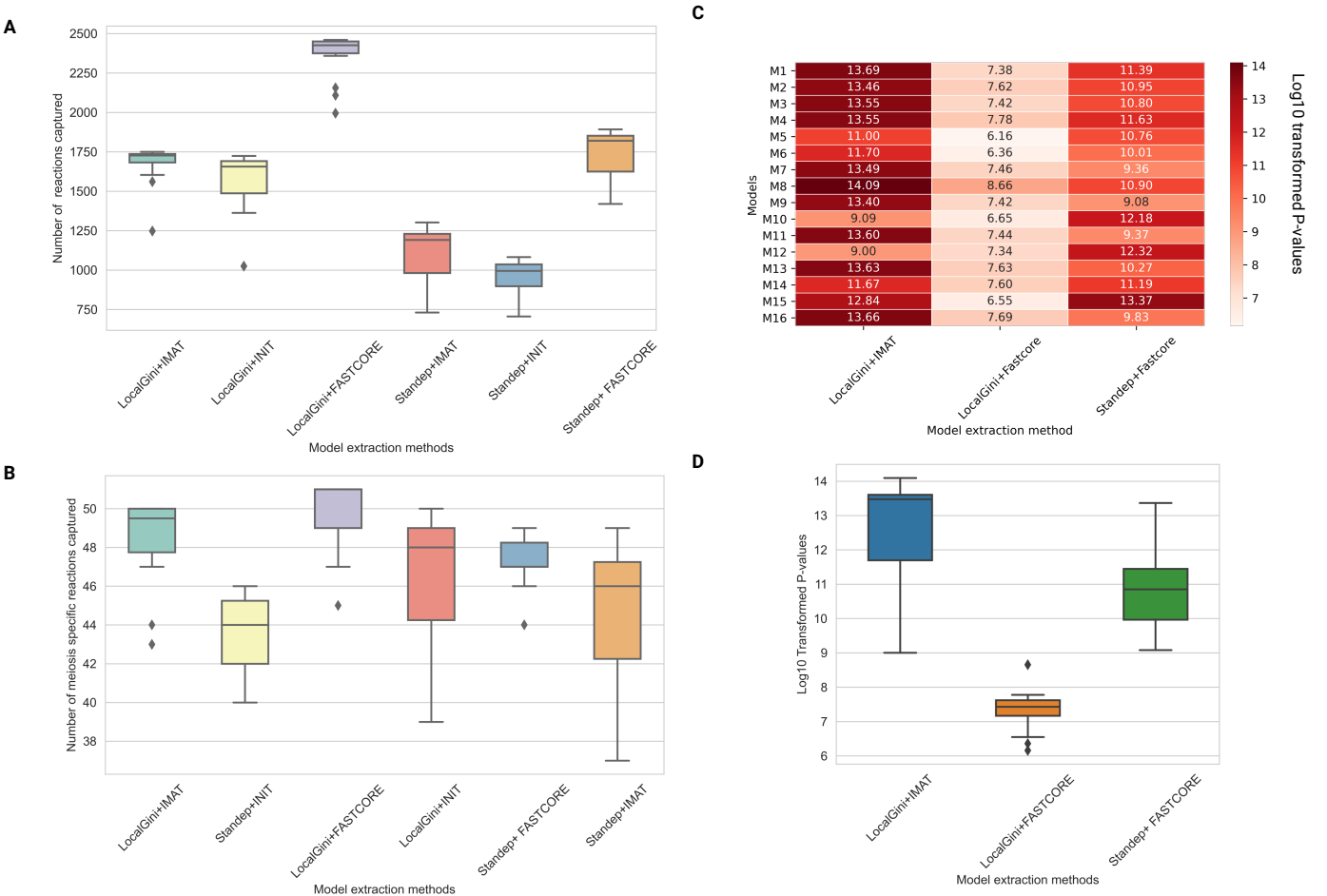
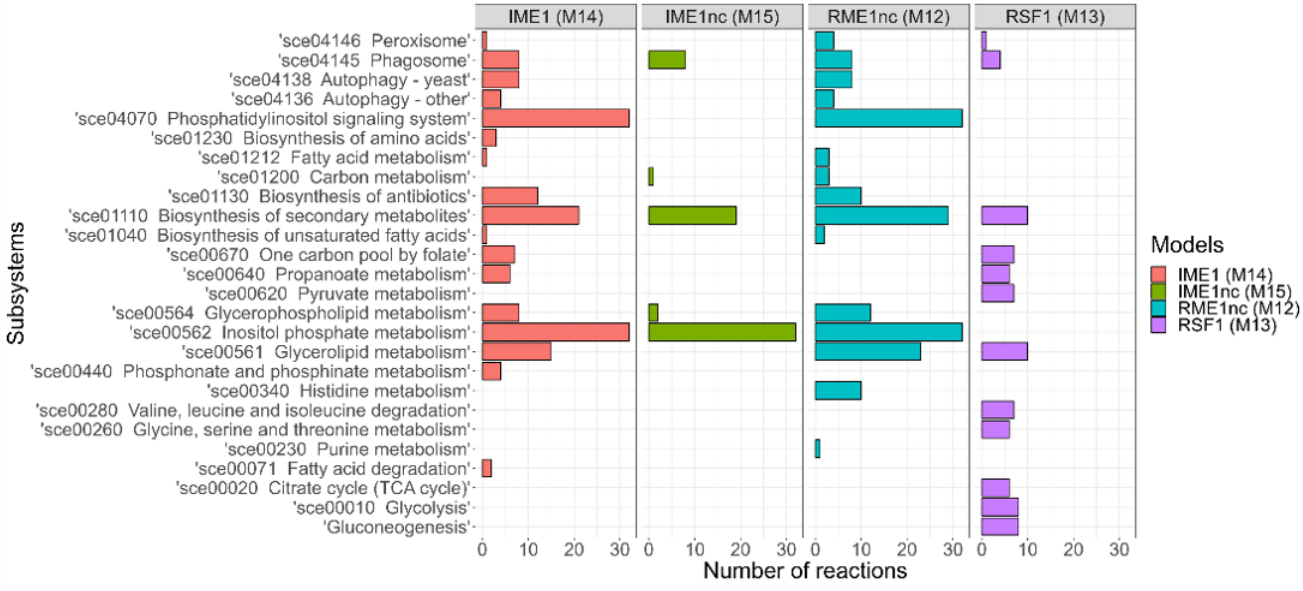

A
Number of reactions captured
2500
2250
2000
1750
1500
1250
1000
750
LocalGini+IMAT
LocalGini+INIT
LocalGini+FASTCORE
Standep+IMAT
Standep+INIT
Standep+ FASTCORE
Model extraction methods
B
Number of meiosis specific reactions captured
50
48
46
44
42
40
38
LocalGini+IMAT
Standep+INIT
LocalGini+FASTCORE
LocalGini+INIT
Standep+ FASTCORE
Standep+IMAT
Model extraction methods
C
Models
M1 13.69 7.38 11.39
M2 13.46 7.62 10.95
M3 13.55 7.42 10.80
M4 13.55 7.78 11.63
M5 11.00 6.16 10.76
M6 11.70 6.36 10.01
M7 13.49 7.46 9.36
M8 14.09 8.66 10.90
M9 13.40 7.42 9.08
M10 9.09 6.65 12.18
M11 13.60 7.44 9.37
M12 9.00 7.34 12.32
M13 13.63 7.63 10.27
M14 11.67 7.60 11.19
M15 12.84 6.55 13.37
M16 13.66 7.69 9.83
LocalGini+IMAT
LocalGini+Fastcore
Standep+Fastcore
Model extraction method
Log10 transformed P-values
14
13
12
11
10
9
8
7
D
Log10 Transformed P-values
14
13
12
11
10
9
8
7
6
LocalGini+IMAT
LocalGini+FASTCORE
Standep+ FASTCORE
Model extraction methods

Figure 5

A


IME1 (M14)
IME1nc (M15)
RME1nc (M12)
RSF1 (M13)
'sce04146 Peroxisome'
'sce04145 Phagosome'
'sce04138 Autophagy - yeast'
'sce04136 Autophagy - other'
'sce04070 Phosphatidylinositol signaling system'
'sce01230 Biosynthesis of amino acids'
'sce01212 Fatty acid metabolism'
'sce01200 Carbon metabolism'
'sce01130 Biosynthesis of antibiotics'
'sce01110 Biosynthesis of secondary metabolites'
'sce01040 Biosynthesis of unsaturated fatty acids'
'sce00670 One carbon pool by folate'
'sce00640 Propanoate metabolism'
'sce00620 Pyruvate metabolism'
'sce00564 Glycerophospholipid metabolism'
'sce00562 Inositol phosphate metabolism'
'sce00561 Glycerolipid metabolism'
'sce00440 Phosphonate and phosphinate metabolism'
'sce00340 Histidine metabolism'
'sce00280 Valine, leucine and isoleucine degradation'
'sce00260 Glycine, serine and threonine metabolism'
'sce00230 Purine metabolism'
'sce00071 Fatty acid degradation'
'sce00020 Citrate cycle (TCA cycle)'
'sce00010 Glycolysis'
'Gluconeogenesis'
Subsystems
0 10 20 30
Number of reactions
Models
IME1 (M14)
IME1nc (M15)
RME1nc (M12)
RSF1 (M13)

B

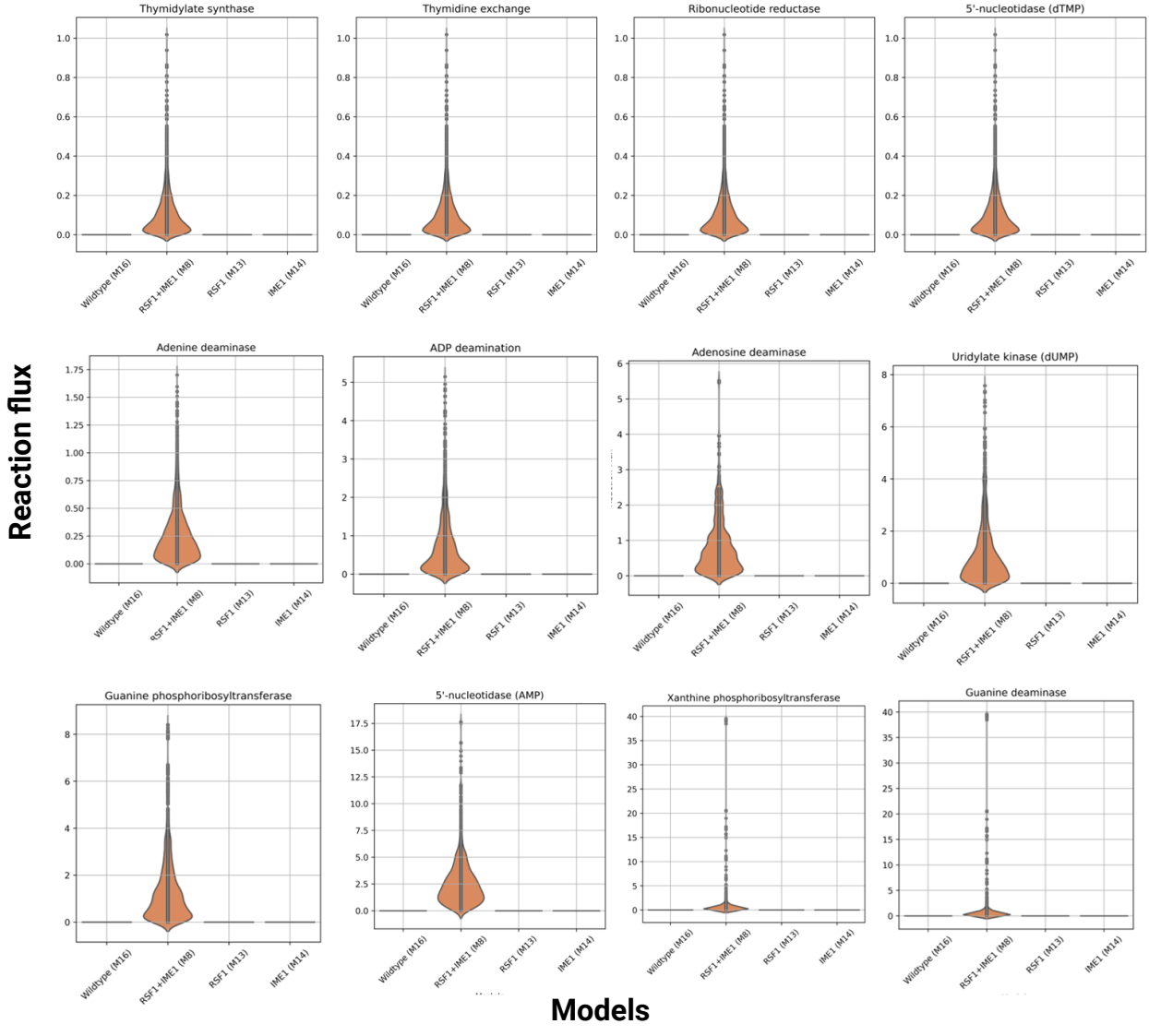
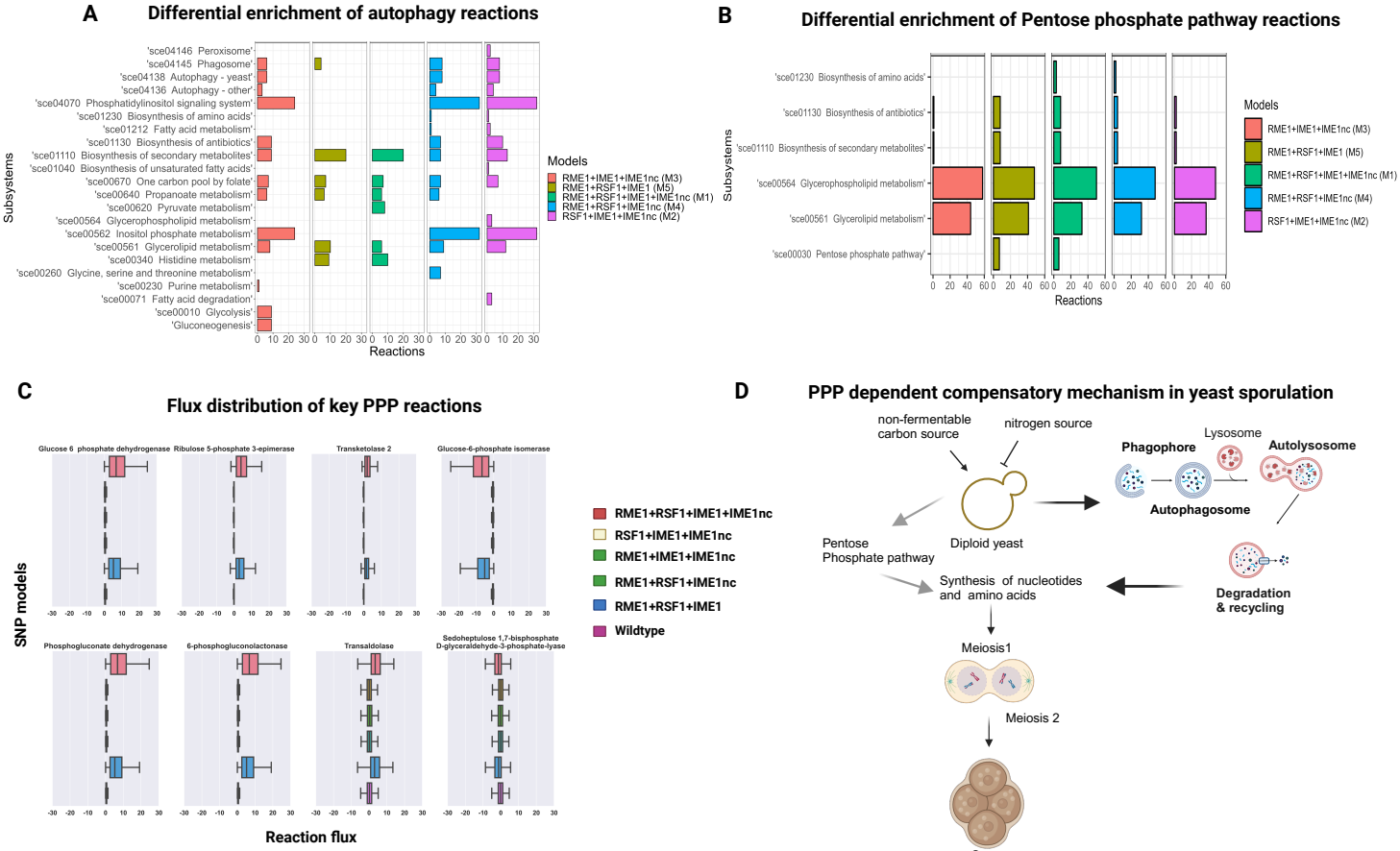

Thymidylate synthase
Thymidine exchange
Ribonucleotide reductase
5'-nucleotidase (dTMP)
Adenine deaminase
ADP deamination
Adenosine deaminase
Uridylate kinase (dUMP)
Guanine phosphoribosyltransferase
5'-nucleotidase (AMP)
Xanthine phosphoribosyltransferase
Guanine deaminase
Wildtype (M16)
RSF1+IME1 (M8)
RSF1 (M13)
IME1 (M14)
Reaction flux
Models

Figure 6


A
Differential enrichment of autophagy reactions
Subsystems
'sce04146 Peroxisome'
'sce04145 Phagosome'
'sce04138 Autophagy - yeast'
'sce04136 Autophagy - other'
'sce04070 Phosphatidylinositol signaling system'
'sce01230 Biosynthesis of amino acids'
'sce01212 Fatty acid metabolism'
'sce01130 Biosynthesis of antibiotics'
'sce01110 Biosynthesis of secondary metabolites'
'sce01040 Biosynthesis of unsaturated fatty acids'
'sce00670 One carbon pool by folate'
'sce00640 Propanoate metabolism'
'sce00620 Pyruvate metabolism'
'sce00564 Glycerophospholipid metabolism'
'sce00562 Inositol phosphate metabolism'
'sce00561 Glycerolipid metabolism'
'sce00340 Histidine metabolism'
'sce00260 Glycine, serine and threonine metabolism'
'sce00230 Purine metabolism'
'sce00071 Fatty acid degradation'
'sce00010 Glycolysis'
'Gluconeogenesis'
0 10 20 30
Reactions
Models
RME1+IME1+IME1nc (M3)
RME1+RSF1+IME1 (M5)
RME1+RSF1+IME1+IME1nc (M1)
RME1+RSF1+IME1nc (M4)
RSF1+IME1+IME1nc (M2)
B
Differential enrichment of Pentose phosphate pathway reactions
Subsystems
'sce01230 Biosynthesis of amino acids'
'sce01130 Biosynthesis of antibiotics'
'sce01110 Biosynthesis of secondary metabolites'
'sce00564 Glycerophospholipid metabolism'
'sce00561 Glycerolipid metabolism'
'sce00030 Pentose phosphate pathway'
0 20 40 60
Reactions
Models
RME1+IME1+IME1nc (M3)
RME1+RSF1+IME1 (M5)
RME1+RSF1+IME1+IME1nc (M1)
RME1+RSF1+IME1nc (M4)
RSF1+IME1+IME1nc (M2)
C
Flux distribution of key PPP reactions
Glucose 6 phosphate dehydrogenase
Ribulose 5-phosphate 3-epimerase
Transketolase 2
Glucose-6-phosphate isomerase
Phosphogluconate dehydrogenase
6-phosphogluconolactonase
Transaldolase
Sedoheptulose 1,7-bisphosphate D-glyceraldehyde-3-phosphate-lyase
SNP models
-30 -20 -10 0 10 20 30
Reaction flux
RME1+RSF1+IME1+IME1nc
RSF1+IME1+IME1nc
RME1+IME1+IME1nc
RME1+RSF1+IME1nc
RME1+RSF1+IME1
Wildtype
D
PPP dependent compensatory mechanism in yeast sporulation
non-fermentable carbon source
nitrogen source
Phagophore
Lysosome
Autolysosome
Autophagosome
Pentose Phosphate pathway
Diploid yeast
Synthesis of nucleotides and amino acids
Degradation & recycling
Meiosis1
Meiosis 2
Spore