Multi-SNP combination genome-scale metabolic models identify molecular mechanisms determining genetic interactions in a quantitative trait

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

[1]Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, India

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

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

*Corresponding authors: niravbhatt@iitm.ac.in, sinha@iitm.ac.in

ABSTRACT

Elucidating genotype-phenotype or variant-to-function relationships remains a pivotal challenge in genetics. For quantitative traits, causal SNPs act either additively or epistatically, resulting in complex interactions which 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. These regulators revealed how some SNP combinations added to higher phenotypic value. Genome-scale differential flux analysis was used to identify differentially activated metabolic reactions across multiple metabolic pathways in each SNP-specific model. This analysis identified causal reactions explaining the observed sporulation efficiency differences. 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 presents a modelling framework that has the potential to unravel specific causal metabolic pathways and reactions regulated by SNPs and their combinations, thus providing insights into GWAS variants of metabolic diseases.

INTRODUCTION

Genome-wide association studies have successfully identified the causal single nucleotide polymorphisms (SNPs) responsible for complex traits. However, understanding the underlying mechanism by which GWAS variants influence complex traits (i.e., variant to function (V2F)) remains challenging. Furthermore, GWAS identifies multiple SNPs, each having a small individual impact on a complex trait, with these effects being influenced by the additive and epistatic interactions amongst these variants. This intricate interplay makes studying how these variants interact within a more extensive genetic and biological network challenging. Numerous studies have highlighted genetic interactions among variants in diverse disease contexts[3,4] and the development of robust computational tools and strategies intended to examine interactions among SNPs[5,6]. Hence, understanding the causal

mechanism of SNPs and their combined effects that influence the complex traits can give insights for developing more appropriate drug targets for therapeutic interventions.

Perturbations in gene expression can directly or indirectly impact cellular metabolism, leading to fluctuations in intracellular fluxes and consequent alterations in metabolic phenotypes. Genome-scale metabolic models (GSMMs), which connect metabolic genes, proteins, and metabolites based on gene-protein-reaction association and stoichiometric matrix, are practical systems biology tools for describing cellular metabolism[7]. Several studies have integrated the gene expression data into GSMMs to get valuable insights into understanding human diseases and infections[8–13] and the knockout effects of genes[14]. Recent studies have utilised GSMM to reconstruct personalised organ-specific metabolic models to address the V2F challenge[15,16].

A few studies have also explored the SNP-specific effects on the metabolism using GSMM[17,18]. However, these studies captured only the *cis* effect of enzyme coding regions, and it remains to be explored how SNPs in non-coding and coding regions of transcription factors individually and in combination can have a *trans* effect in reshaping the metabolic landscape of the cell. This necessitates the integration of SNP-specific gene expression alterations in the GSMM to decipher the *trans* effects of these variants and their interactions in metabolic flux variation.

Sporulation is a metabolically driven process where yeast cells undergo meiosis, generating spores[19]. The entire course of sporulation is orchestrated by five major metabolic pathways - glutamate synthesis, TCA cycle, glyoxylate cycle, gluconeogenesis, and glycogenolysis. Sporulation efficiency, the percentage of cells sporulating in a culture in yeast, is a quantitative trait[20–26] and is highly heritable across yeast strains. Studies have also highlighted metabolic mediators associated with a single causal SNP in driving sporulation efficiency[27,28]. A previous study[22] has identified four causal SNPs that confer high sporulation efficiency phenotype, namely, coding variants in sporulation genes *IME1*^{L325M}, *RSF1*^{D181G}, and non-coding variants in *IME1nc*$^{A\text{-}548G}$ and *RME1*$^{\text{indel-308A}}$. When incorporated singly or in all possible combinations in a low sporulating parental strain, the allele replacement panel[24,29,30] of these strains showed additive phenotype. Gene expression analysis of these allele replacement panels at an early time point revealed large-scale differential gene expression patterns[26]. We leveraged this gene expression data of the allele replacement panel[26] to develop context-

specific GSMM models to address the impact of these SNPs and their combinations on the intracellular metabolic fluxes to drive sporulation efficiency variation. We show SNPs in combinations by uniquely driving divergent sets of differentially expressed genes and changing the connectivity of critical metabolic regulators. Furthermore, with context-specific GSMM for each SNP and their combinations, we show that SNP-SNP interactions could change flux distribution patterns of reactions in six major subsystems. These context-specific GSMM models allowed us to determine synergistic and antagonistic metabolic pathway interactions among various allelic combinations. Further, this led us to predict a crucial role for autophagy in yeast sporulation as a pentose phosphate pathway-dependent compensatory mechanism in a SNP-specific manner. Thus, using context-specific GSMM, we demonstrated the potential of genome-scale metabolic modelling in elucidating the phenotypic effects of variant-variant interactions that help in a more comprehensive understanding of V2F maps.

RESULTS

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[26], 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 DEGs with 362 genes

involved in sporulation efficiency[20]. Of the 362 genes, we identified 48 DEGs that were variable across all SNP combinations. Each SNP combination had a different set of 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 strains (*IME1_IME1nc_RME1_RSF1*, *IME1_IME1nc_RME1*, *IME1_IME1nc_RSF1*, *IME1_RME1_RSF1*, *IME1nc_RME1_RSF1*), 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[31] was uniquely upregulated only in this subset of strains, while *RIM4* was upregulated in an *RSF1*-dependent manner (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).

GO analysis, using ClueGO, of the upregulated DEGs revealed that meiosis nuclear division was significantly enriched in the strains with SNP combinations of *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 (Figure S3). These findings suggested that SNP interactions could affect the timing and progression of meiosis during sporulation.

Connectivity of critical metabolic regulators changes as SNP interact

Given that the strain with all SNP combinations has the highest sporulation efficiency, we were interested in asking how the connectivity of regulators that control the expression of differentially expressed genes in this strain varies 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. 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. Thus, the RIF1 metric identified TFs most differentially co-expressed with highly abundant DE genes. On the other hand, the RIF2 measure was used to determine the TFs with the most significant altered capacity 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), a transcriptional activator of glycolysis[32], as the gene with the highest value. The top two candidates TFs 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[33]. Of these 15 regulators, we were able to identify seven metabolic regulators: *UPC2* involved in sterol biosynthesis, *GCR1, GCR2* in glycolysis, *BAS1* in histidine biosynthesis, and *ARG80, LYS14,* and *MET32* in amino acid metabolism. We then constructed fifteen 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 (Table 1).

Filtering for co-expressed gene pairs that were DEGs between wildtype (null model) 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 1). 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). Further, we found that the connectivity of *UPC2* differed between high, moderate, and low sporulating strains (Figure 3B). These findings confirm that SNP combinations significantly impact overall transcriptional activity and highlight the possibility that critical metabolic alterations could be responsible for sporulation efficiency variation between allele replacement strains.

SNP-specific metabolic models show metabolic heterogeneity

As previously established, numerous genes and pathways undergo deregulation during yeast sporulation[34,35]. Additionally, our analysis showed that the SNPs studied here could activate different genes (Figure 2A). Hence, it is crucial to extract only the active genes and pathways from the generalised yeast metabolic model to reconstruct SNP-specific metabolic models. We utilised Localgini thresholding and iMAT to integrate gene expression data from each allele replacement strain with the Yeast8 genome-scale metabolic model following the procedure outlined in Figure 1. Each resulting SNP-specific model (M1 to M16) corresponded to its genotype and varied in the number of genes, reactions, and metabolites (Table 2). The metabolic heterogeneity among the SNP-specific models was captured by calculating the Hamming distance between each model (Figure S4). For example, it was clear from this analysis that M15 with *IME1nc* had a unique set of metabolic reactions compared to the rest of the models.

A set of fundamental reactions consists of the core meiosis-specific network shared by all the extracted models and is crucial for the survival and sporulation of yeast strains in an acetate medium. We identified 401 reactions common in all extracted models and hypothesised that these reactions form the core meiosis network. As expected, these reactions are significantly enriched with pathways that play essential roles in sporulation (Figure S5). Since each SNP-specific model had different sets of metabolic reactions, we hypothesised that this could be either due to the activation of novel reactions in each model or variations in intracellular flux within the meiosis core network.

SNPs activate unique subsystems to modulate sporulation efficiency variation

Using the wildtype (M16) model as a baseline, a subnetwork topology analysis on each SNP model helped investigate the contribution of each SNP combination to the variation in sporulation efficiency. This analysis revealed that each SNP model had uniquely enriched subsystems (a set of reactions involved sharing a common function, Table S3). For example, the M13 model with *RSF1* SNP showed enrichment of valine, leucine, and isoleucine degradation pathways, glycolysis, gluconeogenesis, pyruvate metabolism pathways, and the TCA cycle (Figure 4A). Identifying metabolic pathways required during respiration in the M13 model (*RFS1*, Figure 4A) using the subnetwork topology analysis was consistent with the function of the *RSF1* gene as a gene responsible for respiration, particularly during the early sporulation[36,37] Furthermore, purine metabolism is uniquely enriched only in the M12 model (*RME1nc*, Figure 4A).

The *IME1* coding variant, which affects the interactions between Rim11 and Ume6, is crucial for initiating the sporulation[23,38]. The study by Sudarsanam and Cohen[26] 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 the 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 demonstrate 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 – *RSF1* (M13), *IME1* (M14), *RME1* (M12), and *IME1nc* (M15). By comparing reactions that were activated exclusively in the M8 (*RSF1_IME1*) model but not in the M13 (*RSF1*), M14 (*IME1*), and M16 (wildtype)

models, we identified 113 synergistic reactions. These reactions were notably enriched in glycerophospholipid, glycerolipid, and pyrimidine metabolism pathways (Figure 4B, Table S4). Conversely, when identifying reactions 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 S4, Figure S7). Similarly, when comparing the *RME1* (M12), *IME1nc* (M15), and *RME1_IME1nc* (M9) models, we uncovered the role of *RME1_IME1nc* in synergistically modulating the "one carbon pool by folate" subsystems, encompassing the "glycine cleavage complex" and "tetrahydrofolate aminomethyl transferase" (Figure S8, Table S5).

GS-DFA 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)[39], 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) described in the Methods section. Using GS-DFA analysis, we identified upregulated and downregulated reactions in each SNP model compared to the wildtype model. The complete list of upregulated and downregulated reactions is provided in Table S6 and Table S7, respectively.

GS-DFA analysis showed that the upregulated reactions were primarily enriched in the following 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 5B, Table S8).

As M1 had the highest sporulation efficiency, we asked 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 variable subsets of the M1 model, highlighting intracellular flux variation among these pathways as significant contributors to sporulation efficiency variation among these strains (Figure 5B). 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. Using the GS-DFA analysis, we found that reactions involved in glycerolipid metabolism, glycerophospholipid metabolism, and biosynthesis of secondary metabolism were upregulated 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 S6B).

We found that models were clustered into two groups on clustering the Flux change (FC) values of upregulated reactions in the M1 model across other models (Figure S9). Hence, we were interested in exploring differentially upregulated subsystems in each SNP-specific model. 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 (Table 3, Table S8). This result was consistent with our flux variability analysis (Supplementary Methods)[40], 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 pentose phosphate pathway for biosynthesis rather than glycolysis for energy generation (Figure 5C). In highly sporulating SNP-specific models M2, M3, and M4, which had *IME1nc* variant and where the pentose phosphate pathway was not upregulated (Figure 5C), showed enrichment of reactions involved in autophagy and phagosomes (Figure 5A, 5B). 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[41,42].

Moreover, we found that the upregulation of reactions involved in amino acid biosynthesis was observed only in the M1, M4, and M11 models (Table S4), highlighting the effect of SNP combinations *RME1_RSF1* and their interaction with *IME1nc* variants on flux distribution patterns. These findings demonstrate the significant impact of SNPs and their interactions on the metabolic flux distribution, explaining how these interactions affected sporulation efficiency.

DISCUSSION

Here, we attempted to bridge the V2F challenge by using a well-studied complex trait in yeast – sporulation efficiency. By harnessing 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 nucleotide metabolism (purine and pyrimidine metabolism) was synergistically modulated while 'steroid biosynthesis' was antagonistically modulated when coding variants of *RSF1* and *IME1* were in combination. This insight 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[30]. 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[13,43]. 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[29]. 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 might modulate sporulation efficiency variation by influencing

metabolic reactions beyond the essential ones required for sporulation, highlighting the specific effects of genetic variations at the SNP level instead of the broader gene deletion effects. 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[44]. 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 S11). Previous studies have shown autophagy as a compensatory mechanism for the PPP in tumour cells[45]. It would be interesting to experimentally validate whether autophagosomes are induced when the PPP reaction is downregulated. Thus, integrating SNP and their combination-specific gene expression data in GSMM has provided several novel mechanisms by which phenotypic variation in a well-studied trait is regulated at the molecular level.

Several studies have reported complex genetic interactions between identified SNPs in metabolic diseases[46,47]. 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[15]. 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[48,49] 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, this can be a robust framework to provide candidate genes for further validation in the molecular processes and phenotype-genotype relationship, leading to clearer variant-to-function pathways.

METHODS

To investigate how single nucleotide polymorphisms (SNPs) and their combinations contribute to phenotypic variation, we conducted a comprehensive analysis using gene expression data from an allele replacement panel of Saccharomyces cerevisiae[26] during sporulation. Our study followed a multi-step approach. Firstly, we identified the SNP-specific differentially expressed genes (DEGs) associated with sporulation and metabolic processes. Secondly, to identify the regulators that control DEGs, we determined their connectivity changes as SNP interacts. Lastly, we employed a genome-scale metabolic modelling approach to understand how these SNPs influence the intracellular metabolic fluxes. The entire workflow, from the pre-processing of RNA-Seq data to the application of genome-scale metabolic modelling techniques, is described in Figure 1.

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)[26] 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[50]. 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[51]. The transcript TPM values obtained from Salmon were converted to gene-level counts using Tximport v1.4. 0[52]. The effect of day on the gene expression values is removed using ComBatseq[53].

Differential gene expression analysis compared each allele replacement strain with the wildtype strain using DESeq2[54]. 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 foldchanges (LFC) shrinkage using the alpegm method[55], 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[56] from Cytoscape[57] for gene ontology and functional annotations analysis.

Regulatory impact factor (RIF) analysis

CeTF package in R was utilised to perform the Regulatory Impact Factors analysis[58,59]. 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/)[60], 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. 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[61] 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)[62]. 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 wild type and 14 allele replacement strains were visualised using Cytoscape[57] 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[62,63]. 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).

Reconstruction of SNP-specific metabolic models

Yeast 8.4.2[64], a commonly used yeast Genome-scale metabolic model, was used in this study. It is accessible via GitHub (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 (LB=UB=0 [mmol/(g DW h)]) while allowing an unrestricted supply of oxygen and acetate (LB= -1000 [mmol/(g DW h)])[29].

The gene expression data (TPM values) averaged across replicates of wild type, and 15 allele replacement strains were pre-processed using Localgini thresholding[65] and integrated into Yeast 8.4.2 using the model extraction method, iMAT[66]. The detailed methodology for reconstructing a SNP-specific model using Localgini thresholding and iMAT is discussed in Supplementary Methods. This led us to reconstruct 15 SNP-specific models (M1 to M15) and a wildtype model (M16).

To understand the metabolic heterogeneity of each extracted model, we created a Boolean reaction vector for each SNP-specific model. In brief, for each model, we made a Boolean vector of length 4058 (corresponding to all the reactions in the Yeast8 metabolic model) with 1s and 0s, indicating whether a reaction was included in the 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[67]. In brief, $FEA.m$ used a hypergeometric 1-sided test (α = 0.05) and FDR correction (p-adj < 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[13], 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 using an optGpSampler[68] 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[69]. The optGpSampler in COBRApy version 0.10.1[63] was used in this study to sample 10,000 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 un-correlated fluxes[13]. Let X_{SNP} and $X_{Wildtype}$ be the flux distributions of a particular reaction in SNP 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 10 times higher or lower 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

Complete analysis associated with this study is available in a Jupyter Notebook hosted in a GitHub repository <https://github.com/bt19d403/Modelling-and-transcriptomics-analysis-of-QTNs>, along with all generated models and code related to this study.

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).

AUTHOR CONTRIBUTIONS

Conceptualisation: S.S., and H.S.; Formal Analysis: S.S., and P.K.S.; 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.

COMPETING INTERESTS

The authors declare no competing interests.

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 and 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) 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: 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 5: Pentose phosphate pathway is differentially regulated in SNP models

(A) Flux enrichment plots for unique reactions in highly sporulating strains compared to the wildtype strain (B) Flux enrichment plots for upregulated reactions in highly sporulating strains compared to the wildtype strain obtained after GF-DFA analysis. (B) 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.

TABLES

Table 1: 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.

Before Filtering for DEGs and RIFs			**After Filtering for DEGs and RIFs**		
Networks	**Nodes**	**Edges**	**Networks**	**Nodes**	**Edges**
N1	6305	117647	N1	3478	5540
N2	6305	114007	N2	3808	5333
N3	6305	128798	N3	3908	6039
N4	6305	113923	N4	3685	5486
N5	6305	119963	N5	3660	5697
N6	6305	122118	N6	4094	5940
N7	6305	114801	N7	3685	5590
N8	6305	124040	N8	3717	5540
N9	6305	115132	N9	3635	5521
N10	6305	115314	N10	3770	5617
N11	6305	120083	N11	3902	5628
N12	6305	123357	N12	3918	5725
N13	6305	119868	N13	3871	5833
N15	6305	117678	N15	3859	5707
N16	6305	121519	N16	3892	6189

Table 2: Number of reactions, metabolites, and genes in each SNP-specific model generated using iMAT+Localgini thresholding. 'O' denotes the oak allele, and 'W' indicates the vineyard allele.

Models	Genes	Reactions	Metabolites	Genotype
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	$RME1^{W}RSF1^{W}IME1^{W}IME1nc^{W}$

Table 3: Differential regulation of Pentose phosphate pathway subsystem across SNP models in an *IME1nc*-dependent manner (highlighted in bold).

PPP normal	**PPP upregulation**
*RME1_**IME1nc***	*RME1*
*IME1_**IME1nc***	*IME1*
*RME1_IME1_**IME1nc***	*RME1_IME1*
*RME1_RSF1_**IME1nc***	*RME1_RSF1*
*IME1_RSF1_**IME1nc***	*IME1_RSF1*
RSF1	*RME1_RSF1_IME1*
IME1nc	*RME1_RSF1_IME1_IME1nc*
*RSF1_**IME1nc***	

SUPPLEMENTARY INFORMATION

Supplementary Methods

Supplementary Tables

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

Table S4: Synergistic and antagonistic reactions and their subsystem enrichment for *RSF1+IME1* (M8) SNP combination.

Table S5: Synergistic and antagonistic reactions and their subsystem enrichment for *RME1+IME1nc* (M9) SNP combination.

Table S6: List of upregulated reactions identified by GS-DFA analysis for all SNP models compared with wildtype model

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

Table S8: Subsystems enriched for upregulated reactions in each SNP model compared to wildtype identified using GS-DFA analysis.

Supplementary Figures

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 Hamming 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: 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 S11: Proposed mechanism of autophagy as a pentose-phosphate pathway-dependent compensatory mechanism during the early phase of yeast sporulation.

REFERENCES

1. Visscher, P. M. *et al.* 10 years of GWAS discovery: biology, function, and translation. *Am J Hum Genet* **101**, 5–22 (2017).
2. Tam, V. *et al.* Benefits and limitations of genome-wide association studies. *Nat Rev Genet* **20**, 467–484 (2019).
3. Cordell, H. J. Detecting gene-gene interactions that underlie human diseases. *Nat Rev Genet* **10**, 392–404 (2009).
4. Martin, M. P. *et al.* Epistatic interaction between KIR3DS1 and HLA-B delays the progression to AIDS. *Nat Genet* **31**, 429–434 (2002).
5. Fang, G. *et al.* Discovering genetic interactions bridging pathways in genome-wide association studies. *Nat Commun* **10**, 4274 (2019).
6. Wan, X. *et al.* BOOST: A Fast approach to detecting gene-gene interactions in genome-wide case-control studies. *Am J Hum Genet* **87**, 325–340 (2010).
7. Gu, C., Kim, G. B., Kim, W. J., Kim, H. U. & Lee, S. Y. Current status and applications of genome-scale metabolic models. *Genome Biol* **20**, 121 (2019).
8. Agren, R. *et al.* Reconstruction of genome-scale active metabolic networks for 69 human cell types and 16 cancer types using INIT. *PLoS Comput Biol* **8**, e1002518 (2012).
9. Bhalla, P., Rengaswamy, R., Karunagaran, D., Suraishkumar, G. K. & Sahoo, S. Metabolic modeling of host–microbe interactions for therapeutics in colorectal cancer. *NPJ Syst Biol Appl* **8**, 1–11 (2022).
10. Islam, M. M., Goertzen, A., Singh, P. K. & Saha, R. Exploring the metabolic landscape of pancreatic ductal adenocarcinoma cells using genome-scale metabolic modeling. *iScience* **25**, 104483 (2022).

11. Jenior, M. L., Dickenson, M. E. & Papin, J. A. Genome-scale metabolic modeling reveals increased reliance on valine catabolism in clinical isolates of *Klebsiella pneumoniae*. *NPJ Syst Biol Appl* **8**, 1–8 (2022).
12. Larsson, I., Uhlén, M., Zhang, C. & Mardinoglu, A. Genome-scale metabolic modeling of glioblastoma reveals promising targets for drug development. *Front Genet* **11**, 381 (2020).
13. Nanda, P. & Ghosh, A. Genome scale-differential flux analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection. *PLoS Comput Biol* **17**, e1008860 (2021).
14. Granados, J. C. *et al.* AHR is a master regulator of diverse pathways in endogenous metabolism. *Sci Rep* **12**, 16625 (2022).
15. Foguet, C. *et al.* Genetically personalised organ-specific metabolic models in health and disease. *Nat Commun* **13**, 7356 (2022).
16. Lewis, J. E., Forshaw, T. E., Boothman, D. A., Furdui, C. M. & Kemp, M. L. Personalized genome-scale metabolic models identify targets of redox metabolism in radiation-resistant tumors. *Cell Syst* **12**, 68-81.e11 (2021).
17. Øyås, O. *et al.* Model-based integration of genomics and metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. *Proc Natl Acad Sci U S A* **117**, 8494–8502 (2020).
18. Sarkar, D. & Maranas, C. D. SNPeffect: identifying functional roles of SNPs using metabolic networks. *Plant J* **103**, 512–531 (2020).
19. Neiman, A. M. Sporulation in the budding yeast *Saccharomyces cerevisiae*. *Genetics* **189**, 737–765 (2011).
20. 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 U S A* **99**, 15530–15535 (2002).
21. Deutschbauer, A. M. & Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat Genet* **37**, 1333–1340 (2005).
22. Gerke, J. P., Chen, C. T. L. & Cohen, B. A. Natural isolates of *Saccharomyces cerevisiae* display complex genetic variation in sporulation efficiency. *Genetics* **174**, 985–997 (2006).
23. Gertz, J., Gerke, J. P. & Cohen, B. A. Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. *Theor Popul Biol* **77**, 1–5 (2010).
24. J, G., K, L. & B, C. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323**, 498–501 (2009).
25. Lorenz, K. & Cohen, B. A. Causal variation in yeast sporulation tends to reside in a pathway bottleneck. *PLoS Genet* **10**, e1004634 (2014).
26. Sudarsanam, P. & Cohen, B. A. Single nucleotide variants in transcription factors associate more tightly with phenotype than with gene expression. *PLoS Genet* **10**, e1004325 (2014).
27. Gupta, S. *et al.* Temporal expression profiling identifies pathways mediating effect of causal variant on phenotype. *PLoS Genet* **11**, e1005195 (2015).
28. Gupta, S. *et al.* Meiotic interactors of a mitotic gene Tao3 revealed by functional analysis of its rare variant. *G3 (Bethesda)* **6**, 2255–2263 (2016).

29. Ray, D. & Ye, P. Characterization of the metabolic requirements in yeast meiosis. *PLoS One* **8**, e63707 (2013).
30. Walther, T. *et al.* Developmental stage-dependent metabolic regulation during meiotic differentiation in budding yeast. *BMC Biol* **12**, 1–22 (2014).
31. Yoon, S.-W. *et al.* Meiotic prophase roles of Rec8 in crossover recombination and chromosome structure. *Nucleic Acids Res* **44**, 9296–9314 (2016).
32. Holland, M. J., Yokoi, T., Holland, J. P., Myambo, K. & Innis, M. A. The *GCR1* gene encodes a positive transcriptional regulator of the enolase and glyceraldehyde-3-phosphate dehydrogenase gene families in *Saccharomyces cerevisiae*. *Mol Cell Biol* **7**, 813–820 (1987).
33. Daignan-Fornier, B. & Fink, G. R. Coregulation of purine and histidine biosynthesis by the transcriptional activators *BAS1* and *BAS2*. *Proc Natl Acad Sci U S A* **89**, 6746–6750 (1992).
34. Primig, M. *et al.* The core meiotic transcriptome in budding yeasts. *Nat Genet* **26**, 415–423 (2000).
35. Enyenihi, A. H. & Saunders, W. S. Large-scale functional genomic analysis of sporulation and meiosis in *Saccharomyces cerevisiae*. *Genetics* **163**, 47–54 (2003).
36. Lu, L., Roberts, G., Simon, K., Yu, J. & Hudson, A. P. Rsf1p, a protein required for respiratory growth of *Saccharomyces cerevisiae*. *Curr Genet* **43**, 263–272 (2003).
37. Roberts III, G. G. & Hudson, A. P. Rsf1p is required for an efficient metabolic shift from fermentative to glycerol-based respiratory growth in *S. cerevisiae*. *Yeast* **26**, 95–110 (2009).
38. Rubin-Bejerano, I., Mandel, S., Robzyk, K. & Kassir, Y. Induction of meiosis in *Saccharomyces cerevisiae* depends on conversion of the transcriptional represssor Ume6 to a positive regulator by its regulated association with the transcriptional activator Ime1. *Mol Cell Biol* **16**, 2518–2526 (1996).
39. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis? *Nat Biotechnol* **28**, 245–248 (2010).
40. Gudmundsson, S. & Thiele, I. Computationally efficient flux variability analysis. *BMC Bioinformatics* **11**, 489 (2010).
41. Barve, G. & Manjithaya, R. Cross-talk between autophagy and sporulation in *Saccharomyces cerevisiae*. *Yeast* **38**, 401–413 (2021).
42. Wen, F.-P. *et al.* Distinct temporal requirements for autophagy and the proteasome in yeast meiosis. *Autophagy* **12**, 671–688 (2016).
43. Scott, W. T., Smid, E. J., Block, D. E. & Notebaart, R. A. Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts. *Microb Cell Fact* **20**, 204 (2021).
44. Bertels, L.-K., Fernández Murillo, L. & Heinisch, J. J. The pentose phosphate pathway in yeasts–more than a poor cousin of glycolysis. *Biomolecules* **11**, 725 (2021).
45. Parkhitko, A. A. *et al.* Autophagy-dependent metabolic reprogramming sensitizes *TSC2*-deficient cells to the antimetabolite 6- *Mol Cancer Res* **12**, 48–57 (2014).

46. Lin, E. *et al.* Association and interaction of *APOA5*, *BUD13*, *CETP*, *LIPA* and health-related behavior with metabolic syndrome in a Taiwanese population. *Sci Rep* **6**, 36830 (2016).
47. Cadena-López, R. O. *et al.* Association between SNPs in leptin pathway genes and anthropometric, biochemical, and dietary markers related to obesity. *Genes* **13**, 945 (2022).
48. Gustafsson, J. *et al.* Generation and analysis of context-specific genome-scale metabolic models derived from single-cell RNA-Seq data. *Proc Natl Acad Sci U S A* **120**, e2217868120 (2023).
49. Huang, Y. *et al.* Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux. *Nat Commun* **14**, 4883 (2023).
50. Ewels, P., Magnusson, M., Lundin, S. & Käller, M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics* **32**, 3047–3048 (2016).
51. Patro, R., Duggal, G., Love, M. I., Irizarry, R. A. & Kingsford, C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods* **14**, 417–419 (2017).
52. Soneson, C., Love, M. I. & Robinson, M. D. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F100Res* **4**, 1521 (2015).
53. Zhang, Y., Parmigiani, G. & Johnson, W. E. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform* **2**, lqaa078 (2020).
54. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**, 550 (2014).
55. Zhu, A., Ibrahim, J. G. & Love, M. I. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics* **35**, 2084–2092 (2019).
56. Bindea, G. *et al.* ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics* **25**, 1091–1093 (2009).
57. Shannon, P. *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504 (2003).
58. Oliveira de Biagi, C. A. *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* **22**, 1–8 (2021).
59. Reverter, A., Hudson, N. J., Nagaraj, S. H., Pérez-Enciso, M. & Dalrymple, B. P. Regulatory impact factors: unraveling the transcriptional regulation of complex traits from expression data. *Bioinformatics* **26**, 896–904 (2010).
60. Teixeira, M. C. *et al.* YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. *Nucleic Acids Res* **51**, D785–D791 (2023).
61. Watson-Haigh, N. S., Kadarmideen, H. N. & Reverter, A. PCIT: An R package for weighted gene co-expression networks based on partial correlation and information theory approaches. *Bioinformatics* **26**, 411–413 (2009).
62. Banerjee, P. *et al.* Network and co-expression analysis of airway smooth muscle cell transcriptome delineates potential gene signatures in asthma. *Sci Rep* **11**, 1–16 (2021).

63. Lau, L. Y. *et al.* Dynamics of gene co-expression networks in time-series data : A case study in *Drosophila melanogaster* embryogenesis. *Front Genet* **11**, 517 (2020).
64. Lu, H. *et al.* A consensus *S. cerevisiae* metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun* **10**, 3586 (2019).
65. Kumar, S.P., & Bhatt,N. Localgini: A method for harnessing inequality in gene expression to improve the quality of context-specific models. *bioRxiv* 2023.09.14.557840 doi:10.1101/2023.09.14.557840 (2023).
66. Zur, H., Ruppin, E. & Shlomi, T. iMAT: an integrative metabolic analysis tool. *Bioinformatics* **26**, 3140–3142 (2010).
67. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc* **14**, 639–702 (2019).
68. Megchelenbrink, W., Huynen, M. & Marchiori, E. optGpSampler: An improved tool for uniformly sampling the solution-space of genome-scale metabolic networks. *PLoS One* **9**, e86587 (2014).
69. Herrmann, H. A., Dyson, B. C., Vass, L., Johnson, G. N. & Schwartz, J.-M. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ Syst Biol Appl* **5**, 1–8 (2019).

Figure 1

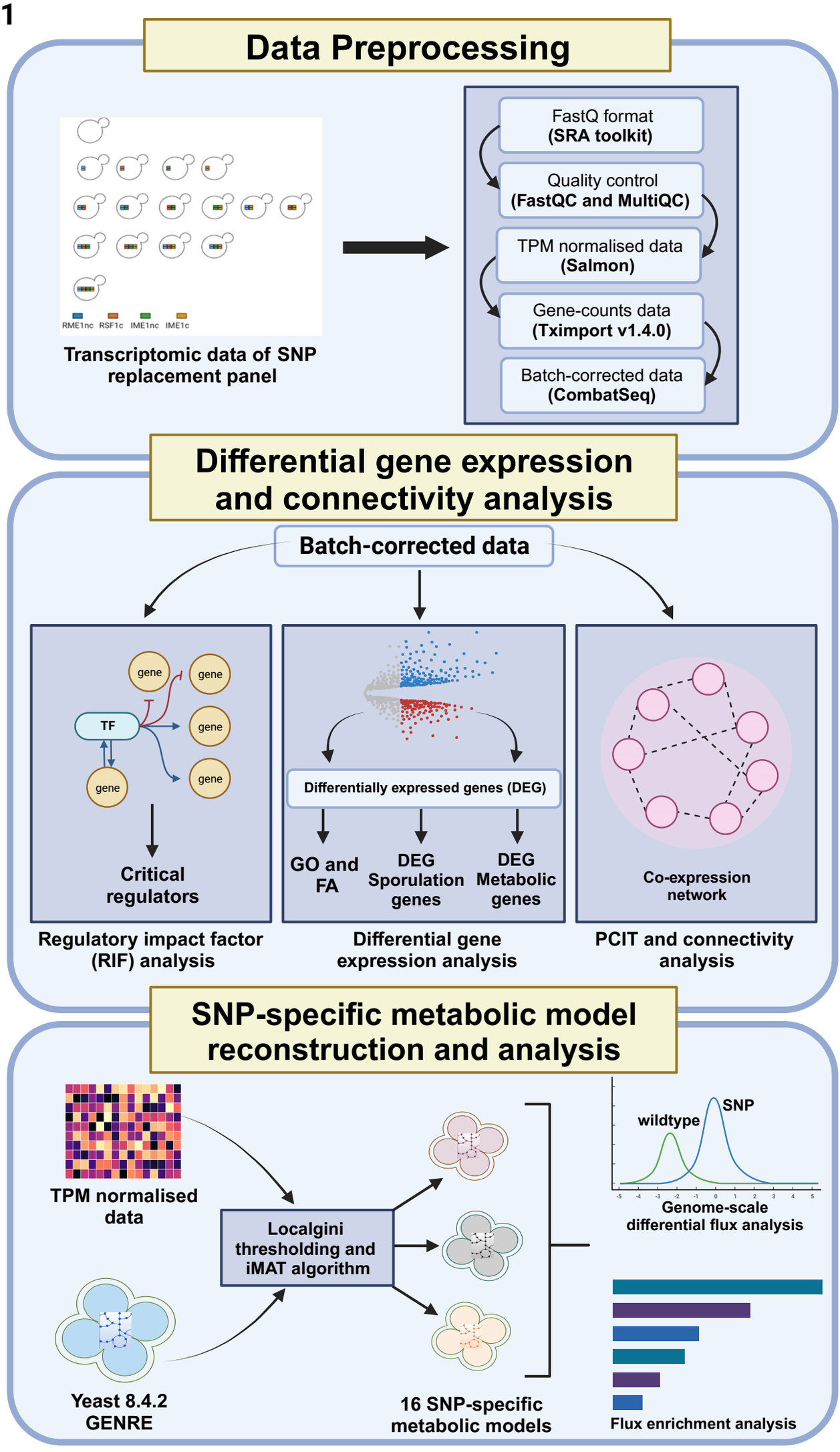
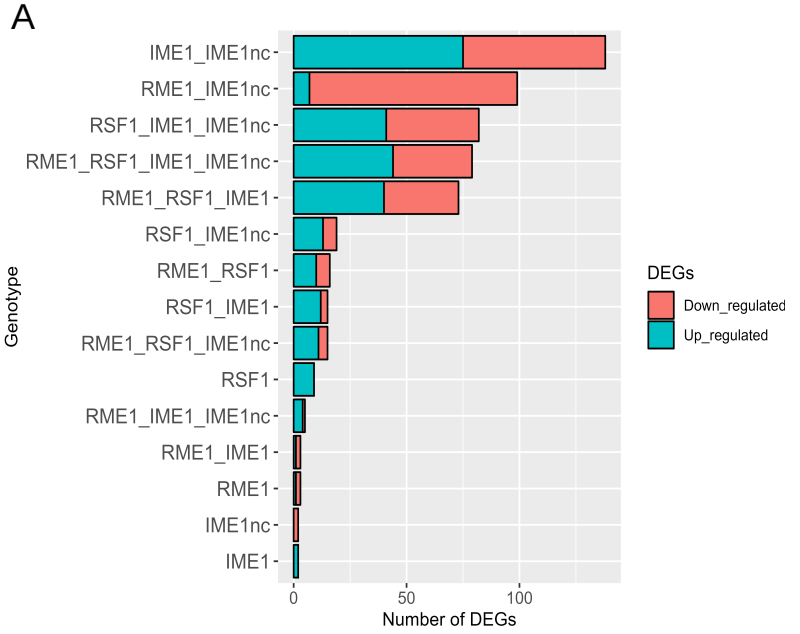
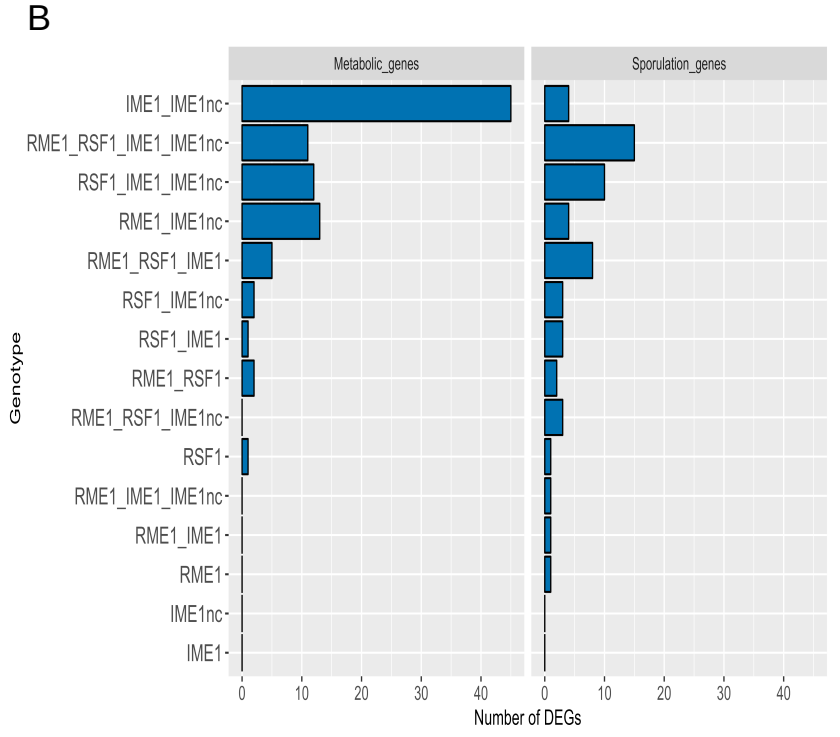
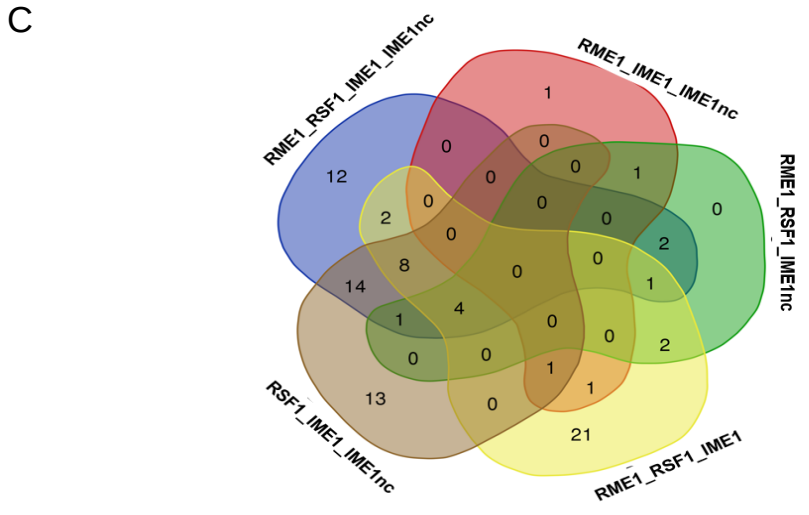
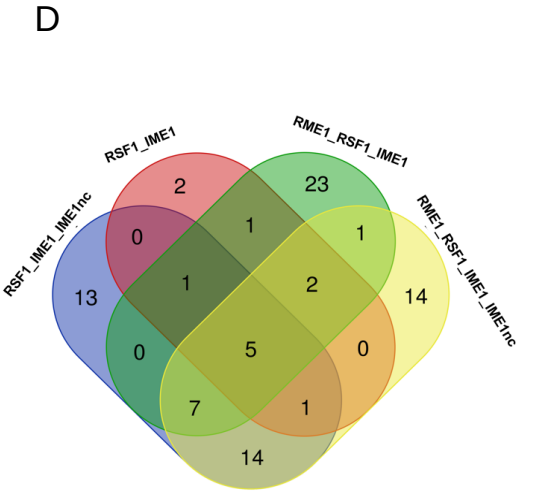

Data Preprocessing
RME1nc
RSF1c
IME1nc
IME1c
Transcriptomic data of SNP replacement panel
FastQ format (SRA toolkit)
Quality control (FastQC and MultiQC)
TPM normalised data (Salmon)
Gene-counts data (Tximport v1.4.0)
Batch-corrected data (CombatSeq)
Differential gene expression and connectivity analysis
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
RME1_RSF1_IME1_IME1nc
RME1_IME1_IME1nc
RME1_RSF1_IME1nc
RSF1_IME1_IME1nc
RME1_RSF1_IME1


D
RSF1_IME1
RME1_RSF1_IME1
RSF1_IME1_IME1nc
RME1_RSF1_IME1_IME1nc

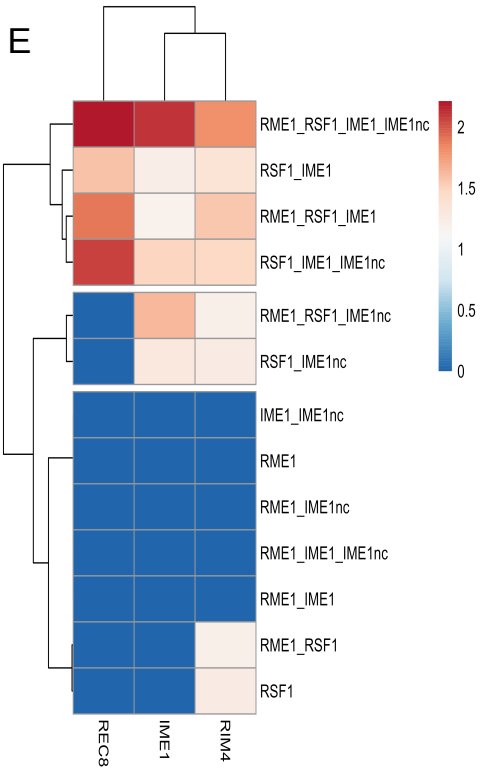

E
RME1_RSF1_IME1_IME1nc
RSF1_IME1
RME1_RSF1_IME1
RSF1_IME1_IME1nc
RME1_RSF1_IME1nc
RSF1_IME1nc
IME1_IME1nc
RME1
RME1_IME1nc
RME1_IME1_IME1nc
RME1_IME1
RME1_RSF1
RSF1
REC8
IME1
RIM4
2
1.5
1
0.5
0

Figure 3

A

	N1	N2	N3	N4	N5	N6	N7	N8	N9	N10	N11	N12	N13	N15	N16
LYS14	0.29	0.67	0.36	0.54	0.46	0.32	0.53	0.61	0.54	0.58	0.51	0.43	0.38	0.56	0.79
GCR1	0.44	0.57	0.53	0.57	0.46	0.79	0.56	0.38	0.53	0.60	0.33	0.57	0.39	0.50	0.71
MET32	0.43	0.56	0.42	0.51	0.51	0.44	0.35	0.54	0.62	0.65	0.57	0.54	0.42	0.50	0.53
ARG80	0.33	0.40	1.00	0.61	0.38	0.50	0.50	0.40	0.56	0.46	0.31	0.60	0.43	0.40	0.85
GCR2	0.47	0.56	0.74	0.39	0.50	0.53	0.42	0.36	0.56	0.60	0.60	0.54	0.35	0.44	0.36
UPC2	0.60	0.50	0.89	0.57	0.50	0.67	0.64	0.81	0.50	0.62	0.51	0.68	0.46	0.42	0.50
BAS1	0.42	0.38	0.39	0.44	0.46	0.49	0.44	0.40	0.42	0.51	0.75	0.56	0.61	0.46	0.28

1 0.9 0.8 0.7 0.6 0.5 0.4 0.3

B

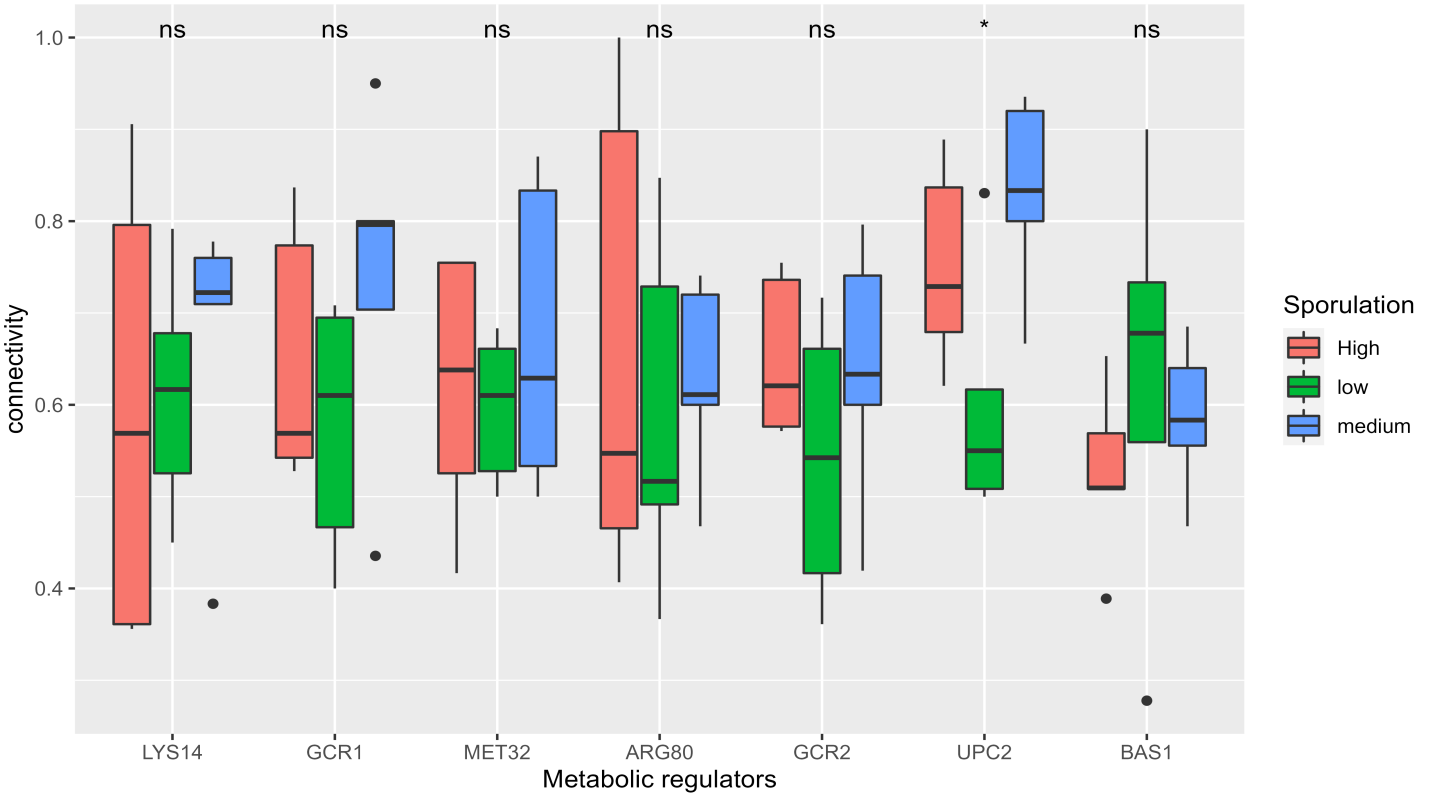


Figure 4

A

IME1 (M14)
IME1nc (M15)
RME1nc (M12)
RSF1 (M13)
Subsystems
'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'
0 10 20 30
Number of reactions
Models
IME1 (M14)
IME1nc (M15)
RME1nc (M12)
RSF1 (M13)

B

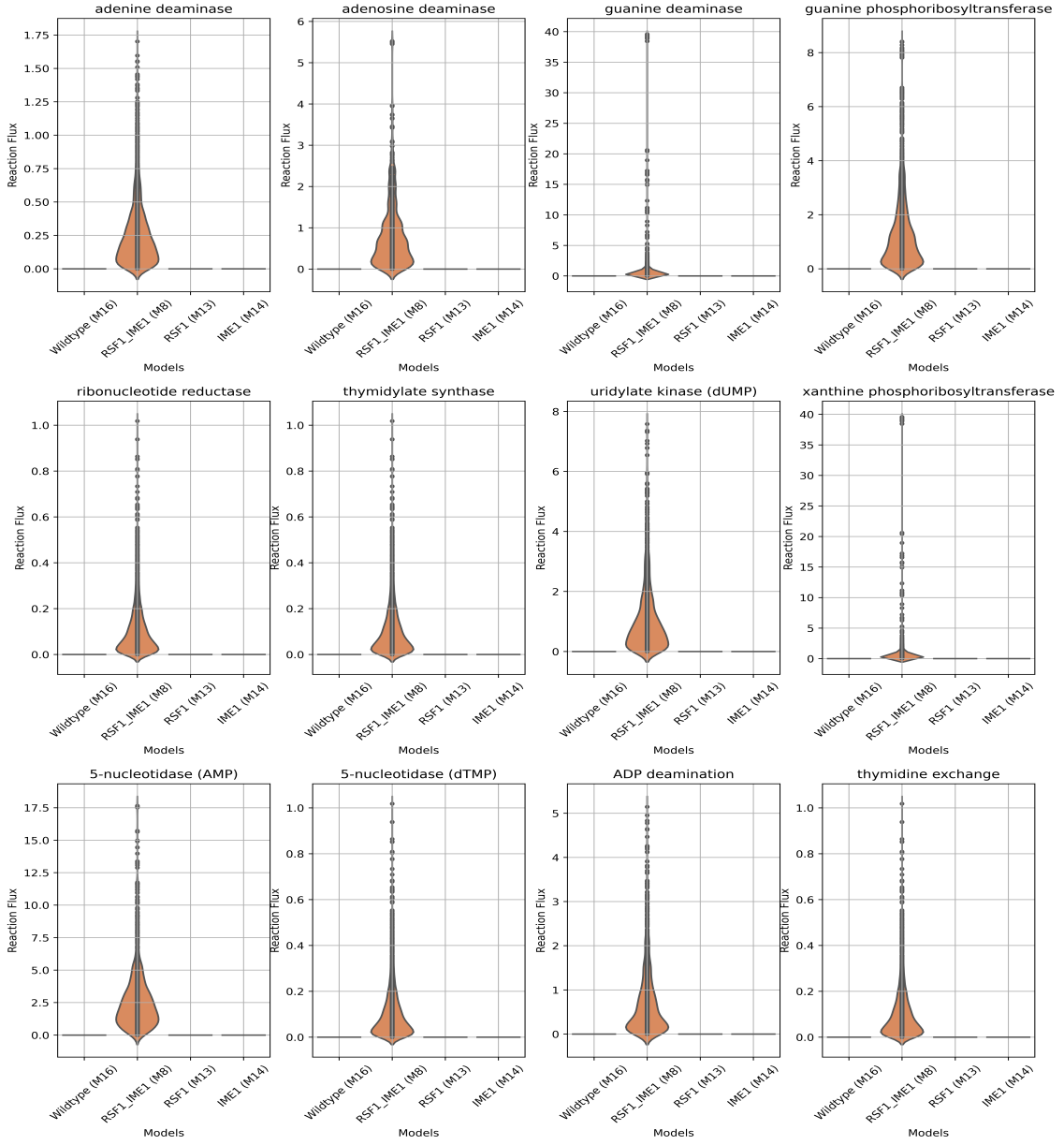
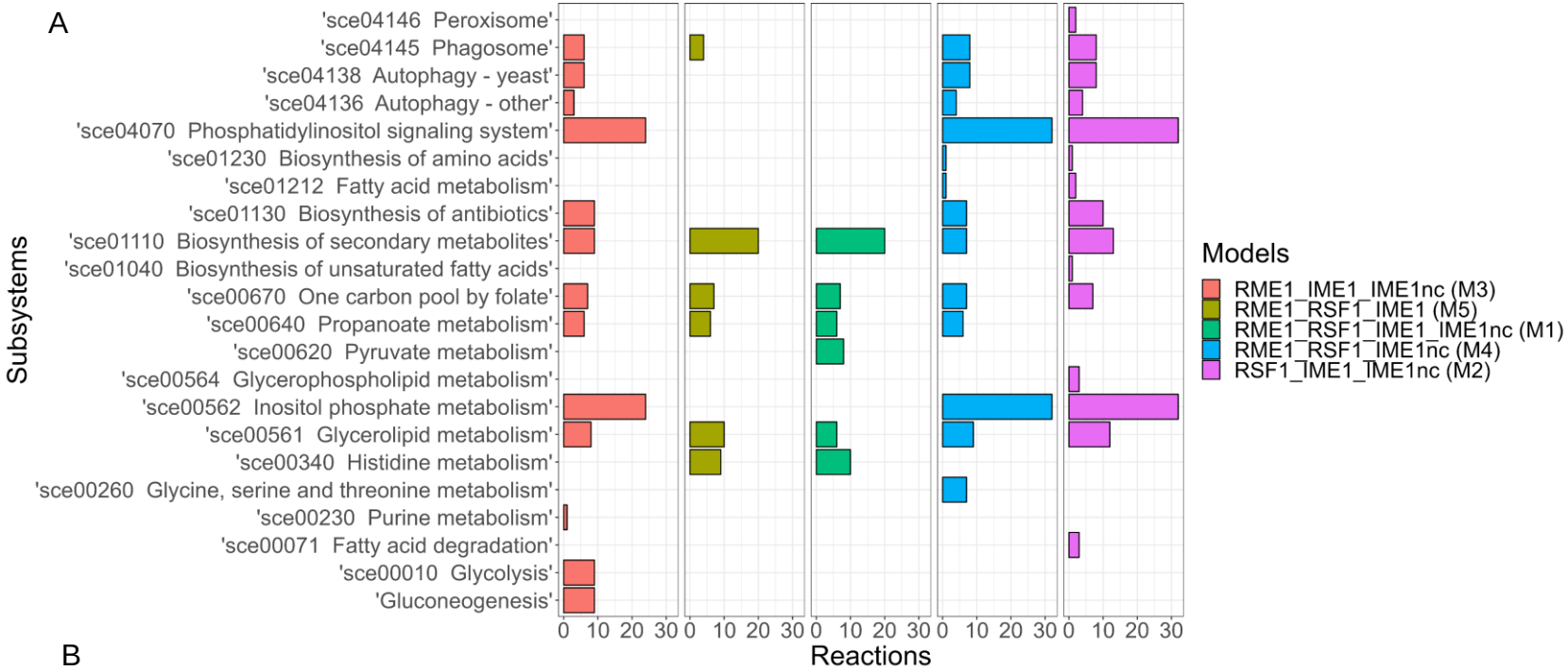
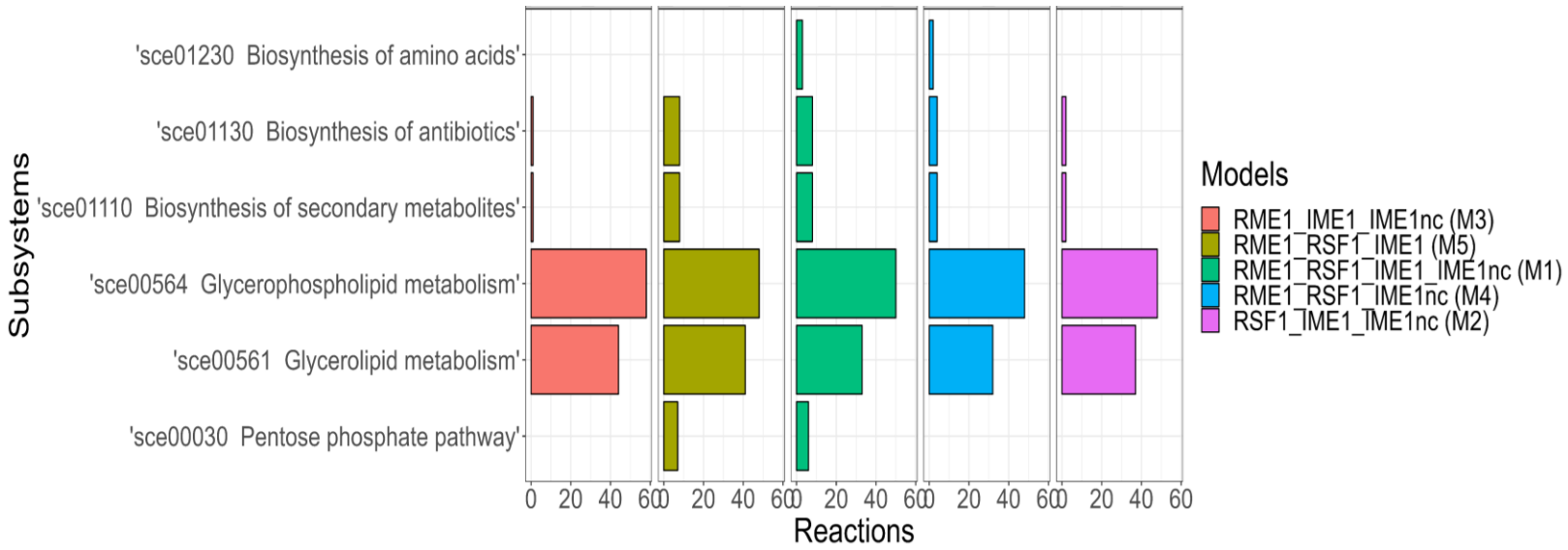

adenine deaminase
adenosine deaminase
guanine deaminase
guanine phosphoribosyltransferase
ribonucleotide reductase
thymidylate synthase
uridylate kinase (dUMP)
xanthine phosphoribosyltransferase
5-nucleotidase (AMP)
5-nucleotidase (dTMP)
ADP deamination
thymidine exchange
Reaction Flux
Wildtype (M16)
RSF1_IME1 (M8)
RSF1 (M13)
IME1 (M14)
Models

Figure 5

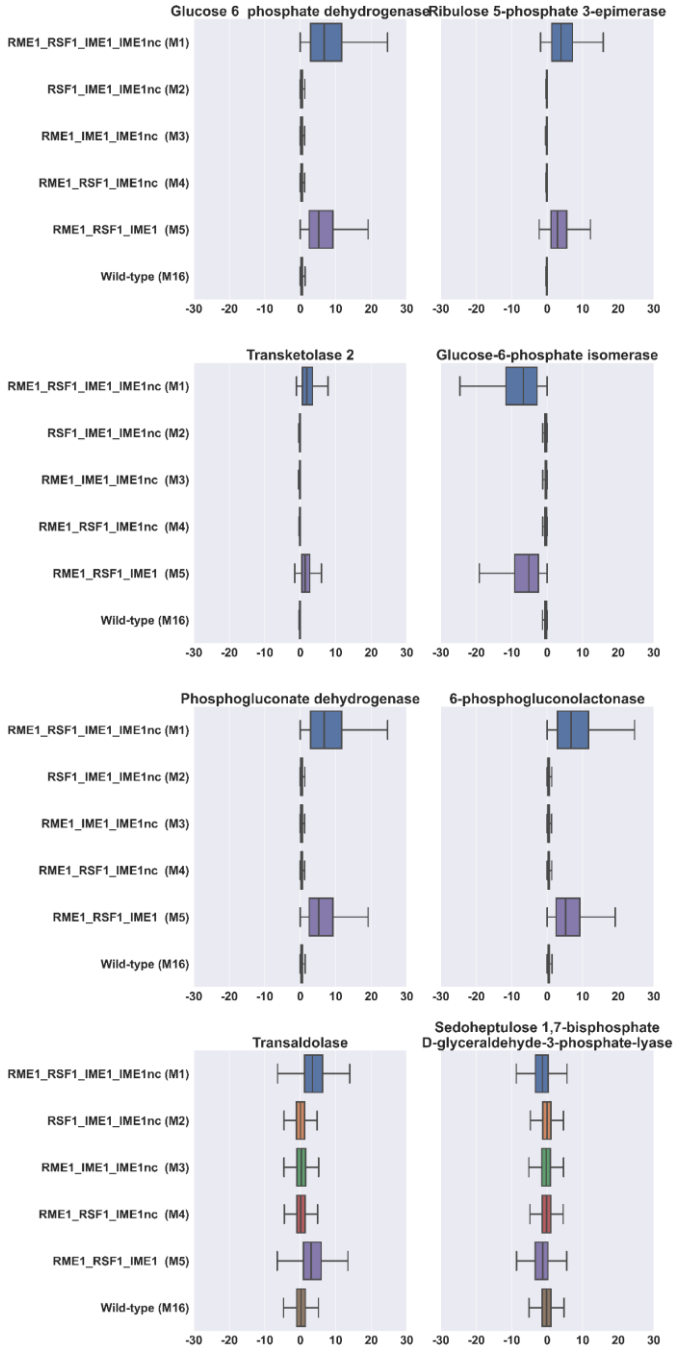
A


'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'
Subsystems
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


'sce01230 Biosynthesis of amino acids'
'sce01130 Biosynthesis of antibiotics'
'sce01110 Biosynthesis of secondary metabolites'
'sce00564 Glycerophospholipid metabolism'
'sce00561 Glycerolipid metabolism'
'sce00030 Pentose phosphate pathway'
Subsystems
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


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
RME1_RSF1_IME1_IME1nc (M1)
RSF1_IME1_IME1nc (M2)
RME1_IME1_IME1nc (M3)
RME1_RSF1_IME1nc (M4)
RME1_RSF1_IME1 (M5)
Wild-type (M16)
-30 -20 -10 0 10 20 30