Unraveling variant-to-function relationships by integrating 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], Himanshu Sinha[1,2,3,*]

[1] Systems Genetics Lab, Department of Biotechnology, Bhupat and Jyoti Mehta School of Biosciences, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[2] Centre for Integrative Biology and Systems Medicine (IBSE), Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[3] Robert Bosch Centre for Data Science and Artificial Intelligence (RBCDSAI), Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[4] BioSystems Engineering and Control (BiSECt) Lab, Department of Biotechnology, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[5] pCoE for Network Systems Learning, Control and Evolution, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

[6] School of Data Science and Artificial Intelligence, Indian Institute of Technology Madras, Chennai, Tamil Nadu 600036, India

*Corresponding author: sinha@iitm.ac.in

ABSTRACT

Elucidating genotype-phenotype or variant-to-function relationships remains a challenge in 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. These regulators revealed how some SNP combinations added to higher phenotypic value. Using IMAT and Localgini thresholding algorithm, 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 multiple metabolic pathways in each SNP-specific model. This analysis revealed causal reactions in six major metabolic pathways 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 discover causal metabolic pathways or reactions regulated by combinations of SNPs, providing insights into Genome-wide association studies (GWAS) variants of metabolic diseases.

AUTHOR SUMMARY

Understanding how genetic changes influence traits that vary in degree, such as human height, can be difficult to identify. In this study with yeast, we have developed an approach to understand how different genetic variations work together to affect a specific trait: sporulation efficiency. We used a network of genes that are co-expressed, or turned on and off together, to identify how different genetic variations interact with one another. We found that certain combinations of variations can make key genes in metabolic pathways more or less active, which can ultimately affect the overall efficiency of sporulation. To further understand how these genetic changes affect metabolism, we used a genome-scale model to simulate the metabolic pathways in yeast. By comparing the metabolic pathways of different genetic combinations, we identified specific reactions that were responsible for the observed changes in sporulation efficiency. Overall, this study provides a new way to understand how variations in our genes can affect our metabolism and could be particularly useful for identifying the genetic causes of metabolic diseases in humans.

KEYWORDS

Genome-scale metabolic models, Yeast sporulation, gene expression networks, quantitative trait, metabolism, genetic interactions, Single nucleotide polymorphisms.

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 quantitative trait loci (e-QTLs), protein expression loci (p-QTLs), and metabolite loci (m-QTLs), 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 collectively reshape cellular metabolism in both coding and non-coding regions, especially in transcription factors, remains a gap to be addressed. Integrating SNP-specific gene expression alterations into GSMMs becomes essential to precisely decipher the effects of variants and their interactions on metabolic flux variations.

80 This study aims to fill this gap by leveraging GSMMs and integrating gene expression data of
81 causal SNPs and their combinations. Focusing on yeast sporulation as a model known for its
82 metabolic nature, the analysis investigates how SNPs and their combinations modulate
83 metabolic fluxes to influence sporulation efficiency [14–16]. The findings reveal that SNP
84 combinations uniquely drive divergent sets of differentially expressed genes and alter the
85 connectivity of critical metabolic regulators. Context-specific GSMMs for each SNP and their
86 combinations illustrate changes in flux distribution patterns and identify synergistic and
87 antagonistic metabolic pathway interactions in the yeast sporulation process. Additionally, a
88 new role for autophagy in yeast sporulation, dependent on the pentose phosphate pathway,
89 is predicted in an SNP-specific manner. Hence, integrating gene expression data into GSMMs
90 is an effective technique for deciphering the intricate relationships between SNPs and cellular
91 metabolism, paving the way for a more nuanced approach to therapeutic intervention.

92 **RESULTS**

93 To investigate how SNPs and their combinations contribute to phenotypic variation, we
94 conducted a comprehensive analysis using gene expression data from an allele replacement
95 panel of *Saccharomyces cerevisiae* [16] during sporulation. The SNPs studied here included
96 coding variants in sporulation genes *IME1*^{L325M}, *RSF1*^{D181G}, and non-coding variants in
97 *IME1nc*$^{A\text{-}548G}$ and *RME1*$^{indel\text{-}308A}$) that conferred high sporulation efficiency phenotype. Here,
98 we present the results of our study, where we followed a multi-step approach. First, we
99 created SNP-specific differential gene expression networks highlighting sporulation,
100 metabolic processes, and regulators. Secondly, we employed a genome-scale metabolic
101 modeling approach to understand how these SNPs influence intracellular metabolic fluxes.
102 These analyses identified causal metabolic reactions that could explain increased phenotype

103 by specific SNPs and their combination, providing insight into how the SNPs interact to
104 modulate a quantitative trait phenotype. The entire workflow, from the pre-processing of
105 RNA-Seq data to the application of genome-scale metabolic modeling techniques, is
106 described in Figure 1.

107 **Transcriptomic analysis reveals SNP-specific differentially expressed genes**

108 Based on the sporulation efficiency data of the 16 allele replacement strains containing causal
109 oak alleles in the vineyard background strains [16], we classified *IME1_IME1nc_RME1_RSF1*,
110 *IME1_IME1nc_RME1*, *IME1_IME1nc_RSF1*, *IME1_RME1_RSF1*, *IME1nc_RME1_RSF1* allelic
111 combination strains as 'high sporulating.' All two allele combination strains (*IME1_IME1nc,*
112 *IME1_RME1, IME1_RSF1, IME1nc_RME1, IME1nc_RSF1, RME1_RSF1*) were 'medium
113 sporulating strains', and all single SNP and wildtype were considered 'low sporulating strains.'
114 Differential gene expression analysis, after adjusting for the false discovery rate (< 10% using
115 the Benjamini-Hochberg method), identified a total of 380 significant differentially expressed
116 genes (DEGs, upregulated and downregulated) across 15 SNP combinations (Figure 2A, Table
117 S1). A comparison of the number of DEGs in each SNP combination showed that the strain
118 with *IME1* coding and *IME1nc* combination had the highest number of DEGs. Interestingly, of
119 all the SNP combinations, we also found a significant downregulation of several genes in the
120 strain with the two non-coding alleles - *RME1* and *IME1nc*.

121 Since we were interested in determining the number of sporulation-related genes among
122 DEGs between wildtype and allele replacement strains, we compared DEGs with 362 genes
123 known to influence sporulation efficiency phenotype as described in the literature [17]. Of
124 the 362 genes, we identified 48 variable DEGs across all SNP combinations. Each SNP
125 combination had a different set of sporulation-specific DEGs, with the strain with all four

126 alleles having the maximum number (Figure 2B, Figure S1). The presence of *RME1nc* resulted
127 in the downregulation of *RME1* expression in all combinations except for *RME1nc_IME1nc*
128 (Figure S1). This showed that the cis effect of the *RME1nc* allele downregulated its expression,
129 enhancing sporulation efficiency in oak strains. Upregulation of the *IME1,* whenever the *RSF1*
130 allele was in combination with either *IME1* or *IME1nc* (Figure S1), might be due to a trans-
131 effect of the *RSF1* variation on the expression of *RIM4*, a major regulator of early sporulation
132 genes and regulates *IME1* expression.

133 Interestingly, no common genes were found to be upregulated across all five strains
134 (*IME1_IME1nc_RME1_RSF1*, *IME1_IME1nc_RME1*, *IME1_IME1nc_RSF1*, *IME1_RME1_RSF1*,
135 *IME1nc_RME1_RSF1*), even though several genes were uniquely activated in each SNP
136 combination strain (Figure 2C). However, *IME2*, a meiosis initiator gene, was activated and
137 uniquely upregulated when all four SNP were combined. We identified five upregulated
138 genes, *HSP26*, *IME1*, *REC8*, *RIM4*, and *YIR016W*, common across the subset of strains with
139 *IME1* and *RSF1* variants in combination (Figure 2D). In particular, *REC8*, a gene that mediates
140 sister chromatid cohesion, homologous recombination, and chromosome synapsis [18], was
141 uniquely upregulated only in this subset of strains, while *RIM4* was upregulated in an *RSF1*-
142 dependent manner (Figure 2E).

143 Since yeast sporulation is a metabolically driven process, we considered all metabolic genes
144 in the Yeast8 metabolic model and investigated if they were differentially expressed across
145 different SNP combinations. Of the 1150 metabolic genes, 164 were differentially expressed
146 across all SNP combinations (Figure S2). The highest number of differentially expressed
147 metabolic genes were observed in *IME1_IME1nc* (Figure 2B). Interestingly, we also identified
148 that SRT1, a gene involved in dolichol biosynthesis and whose null mutant showed reduced

149 sporulation efficiency, was differentially expressed whenever the RSF1 allele was in
150 combination with *IME1* and *IME1nc*, i.e., in *RSF1_IME1_IME1nc* and
151 *RSF1_IME1_IME1nc_RME1* (Figure S2).

152 GO analysis, using ClueGO, of the upregulated DEGs revealed that meiosis nuclear division
153 was significantly enriched in the strains with SNP combinations of *RME1_RSF1_IME1_IME1nc*,
154 *RME1_RSF1_IME1*, and *RSF1_IME1nc_IME1* indicating that these strains have already been
155 undergoing meiosis during the second hour of sporulation (Figure S3). Further, we observed
156 that two non-coding SNPs (*RME1* and *IME1nc*), when combined, could activate genes
157 enriched for cellular respiration (Figure S3). These findings suggested that SNP interactions
158 could affect the timing and progression of meiosis during sporulation.

159 **Connectivity of critical metabolic regulators changes as SNP interact**

160 Given that the strain with all SNP combinations has the highest sporulation efficiency, we
161 examined how the connectivity of regulators that control the expression of differentially
162 expressed genes in this strain varies across other combinations of SNPs. Regulatory impact
163 factor (RIF1 and RIF2) analyses helped determine the TFs that control the expression of DEGs
164 in the strain with all 4 SNP combinations and also assigned a score to these TFs. In brief, the
165 RIF1 metric classified TFs that were differentially co-expressed with highly abundant DEGs,
166 and RIF2 was used to identify the TFs with the altered ability to predict the abundance of
167 DEGs. Thus, the RIF1 metric identified TFs most differentially co-expressed with highly
168 abundant DE genes. On the other hand, the RIF2 measure was used to determine the TFs with
169 the most significant altered capacity to predict the abundance of DEGs. Using the RIF1 and
170 RIF2 *z*-score cut-offs, we obtained 15 TFs as regulators of DEG expression (Table S2). The RIF1
171 analysis identified *GCR1* (*z*-score = 2.40), a transcriptional activator of glycolysis [19] gene

172 with the highest value. The top two candidate transcription factors from RIF2 analysis were
173 *ROF1* (z-score = 2.22), a putative TF, and *BAS1* (*z*-score = 2.18), a Myb-related TF involved in
174 regulating basal and induced expression of genes of the purine and histidine biosynthesis
175 pathways [20]. Of these 15 regulators, we were able to identify seven metabolic regulators:
176 *UPC2* involved in sterol biosynthesis, *GCR1, GCR2* in glycolysis, *BAS1* in histidine biosynthesis,
177 and *ARG80, LYS14,* and *MET32* in amino acid metabolism. We then constructed fifteen
178 networks to identify differences in gene co-expression profiles among the wild type and 14
179 allelic combinations using the partial correlation and information theory (PCIT) algorithm. The
180 correlation of 6,307 genes with a cut-off greater than |0.95| retrieved varying numbers of
181 correlated pairs across wildtype and allele-specific networks (Table 1).

182 Filtering for co-expressed gene pairs that were DEGs between wildtype (null model) and all
183 oak allele combination groups (143 genes) along with the 15 TFs from the RIF analyses (R^2 >
184 |0.95|) allowed the construction of 15 allele-specific sub-networks with varying numbers of
185 nodes and edges (Table S3). These allele-specific sub-networks permitted the determination
186 of how these TFs regulated gene expression by varying their connectivity across the SNPs and
187 their combinations (Figure 3A). Further, we found that the connectivity of *UPC2* differed
188 between high, moderate, and low sporulating strains (Figure 3B). These findings confirm that
189 SNP combinations significantly impact overall transcriptional activity and highlight the
190 possibility that critical metabolic alterations could be responsible for sporulation efficiency
191 variation between allele replacement strains.

192 **SNP-specific metabolic models show metabolic heterogeneity**

193 As previously established, numerous genes and pathways undergo deregulation during yeast
194 sporulation [21,22]. Additionally, our analysis showed that the SNPs studied here could

activate different genes (Figure 2A). Hence, extracting only the active genes and pathways from the generalized yeast metabolic model is crucial to reconstructing SNP-specific metabolic models. We utilized 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 1). 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.

Table 1: 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	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}

208

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

217 **SNPs activate unique subsystems to modulate sporulation efficiency variation**

218 Using the wildtype (M16) model as a baseline, a subnetwork topology analysis on each SNP
219 model helped investigate the contribution of each SNP combination to the variation in
220 sporulation efficiency. This analysis revealed that each SNP model had uniquely enriched
221 subsystems (a set of reactions involved sharing a common function, Table S4). For example,
222 the M13 model with *RSF1* SNP showed enrichment of valine, leucine, and isoleucine
223 degradation pathways, glycolysis, gluconeogenesis, pyruvate metabolism pathways, and the
224 TCA cycle (Figure 4A). Identifying metabolic pathways required during respiration in the M13
225 model (*RFS1*, Figure 4A) using the subnetwork topology analysis was consistent with the
226 function of the *RSF1* gene as a gene responsible for respiration, particularly during the early
227 sporulation [23,24]. Furthermore, purine metabolism is uniquely enriched only in the M12
228 model (*RME1nc*, Figure 4A).

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

239 Next, we investigated whether, in comparison to their individual effects, interactions
240 between SNPs had the potential to activate or deactivate particular cellular pathways
241 selectively. In other words, we investigated whether these interactions could result in
242 synergistic or antagonistic effects on specific metabolic reactions. To demonstrate this idea,
243 we examined the combination of two coding SNPs (*RSF1_IME1* (M8)) and two non-coding
244 SNPs *RME1_IME1nc* (M9) as well as their single SNP models – *RSF1* (M13), *IME1* (M14),
245 *RME1* (M12), and *IME1nc* (M15). By comparing reactions that were activated exclusively in
246 the M8 (*RSF1_IME1*) model but not in the M13 (*RSF1*), M14 (*IME1*), and M16 (wildtype)
247 models, we identified 113 synergistic reactions. These reactions were notably enriched in
248 glycerophospholipid, glycerolipid, and pyrimidine metabolism pathways (Figure 4B, Table S5).
249 Conversely, when identifying reactions that were inactive in the M8 model but active in the
250 M13, M14, and M16 models, we determined 271 antagonistic reactions. These antagonistic
251 reactions were enriched in various subsystems, including fatty acid biosynthesis, lipid

252 metabolism, the phosphatidyl inositol signalling pathway, steroid biosynthesis, etc. (Table S5,
253 Figure S7). Similarly, when comparing the *RME1* (M12), *IME1nc* (M15), and *RME1_IME1nc*
254 (M9) models, we uncovered the role of *RME1_IME1nc* in synergistically modulating the "one
255 carbon pool by folate" subsystems, encompassing the "glycine cleavage complex" and
256 "tetrahydrofolate aminomethyl transferase" (Figure S8, Table S6).

257 **GS-DFA analysis reveals the impact of SNP-SNP interactions on intracellular flux variation**

258 The analysis of the unique sub-network topology showed that combinations of SNPs could
259 activate shared metabolic pathways. However, whether SNPs could modify the flux
260 distribution patterns of these pathways to modulate phenotypic variation was unclear.
261 Therefore, flux variations in common reactions among SNP and wildtype models were
262 analyzed to investigate flux distribution patterns. While the flux balance analysis (FBA, [27]),
263 for a given constraint, can find multiple potential flux values for reactions in a metabolic
264 model, it cannot ensure the uniqueness of a solution. To overcome this limitation in FBA, we
265 used flux sampling to assess variations in metabolic flux distribution using the genome-scale
266 differential flux analysis (GS-DFA) described in the Methods section. Using GS-DFA analysis,
267 we identified upregulated and downregulated reactions in each SNP model compared to the
268 wildtype model. The complete list of upregulated and downregulated reactions is provided in
269 Table S7 and Table S8, respectively.

270 GS-DFA analysis showed that the upregulated reactions were primarily enriched in the
271 following six subsystems involved in the biosynthesis of amino acids, antibiotics, secondary
272 metabolites, glycerophospholipid metabolism, glycerolipid metabolism, and pentose
273 phosphate pathway in a strain with all four SNP combinations (M1, Figure 5B, Table S9).

274 As M1 had the highest sporulation efficiency, we asked if these six subsystems were present
275 in the other high sporulation efficiency strains (M2-M5). We found that in M2-M5, the six
276 subsystems were variable subsets of the M1 model, highlighting intracellular flux variation
277 among these pathways as significant contributors to sporulation efficiency variation among
278 these strains (Figure 5B). The two subsystems, glycerophospholipid and glycerolipid
279 metabolism, present in all models suggested that they were essential for the overall fitness
280 of the yeast strains during sporulation.

281 Further, we explored how the functional and non-functional forms of Ime1 protein could
282 change the flux distribution patterns to modulate sporulation efficiency variation. Using the
283 GS-DFA analysis, we found that reactions involved in glycerolipid metabolism,
284 glycerophospholipid metabolism, and biosynthesis of secondary metabolism were
285 upregulated in *IME1_IME1nc* (M6) and *IME1* (M14) models but not in *IME1nc* (M15) model
286 (Figure S6B). Interestingly, we could also predict the metabolic consequence of the non-
287 functional form of *IME1* (M15), which was to upregulate the reactions involved in the
288 peroxisome, fatty acid metabolism, fatty acid degradation, and fatty acid biosynthesis (Figure
289 S6B).

290 We found that models were clustered into two groups upon hierarchically clustering the Flux
291 Change (FC) values of upregulated reactions in the M1 model across other models (Figure S9).
292 Hence, we were interested in exploring differentially upregulated subsystems in each SNP-
293 specific model. We found that SNP combinations without the *IME1nc* variant, i.e., M5, M8,
294 M10, M11, M12, M14, and the M1 models (which contains all four SNPs), led to an increased
295 flux through reactions involved in the pentose phosphate pathway (Table 2, Table S9). This
296 result was consistent with our flux variability analysis (see Supplementary Methods section,

297 [28]), where we found that the PPP reactions were upregulated in these models (Figure S10).
298 The increased flux through PPP reactions was inversely coupled with the glucose-6-phosphate
299 isomerase reaction. This indicated that the flux was redirected through the pentose
300 phosphate pathway for biosynthesis rather than glycolysis for energy generation (Figure 5C).
301 In highly sporulating SNP-specific models M2, M3, and M4, which had *IME1nc* variant and
302 where the pentose phosphate pathway was not upregulated (Figure 5C), showed enrichment
303 of reactions involved in autophagy and phagosomes (Figure 5A, 5B). These findings were
304 interesting since previous studies have established a correlation between autophagy and
305 yeast sporulation, and the deletion of several autophagy-related genes has shown reduced
306 sporulation [29,30]. Moreover, we found that the upregulation of reactions involved in amino
307 acid biosynthesis was observed only in the M1, M4, and M11 models (Table S5), highlighting
308 the effect of SNP combinations *RME1_RSF1* and their interaction with *IME1nc* variants on flux
309 distribution patterns. These findings demonstrate the significant impact of SNPs and their
310 interactions on the metabolic flux distribution, explaining how these interactions affected
311 sporulation efficiency.

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

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*

314

315 **DISCUSSION**

316 Here, we attempted to bridge the V2F challenge for a well-studied complex trait in yeast –
317 sporulation efficiency. By harnessing the gene expression data from an allele replacement
318 panel of four SNPs, we showed how SNPs interacted and modulated the metabolic pathways
319 to drive sporulation efficiency variation. We showed that certain SNP combinations could
320 activate or deactivate the flux through a particular reaction. We identified nucleotide
321 metabolism (purine and pyrimidine metabolism) was synergistically modulated while 'steroid
322 biosynthesis' was antagonistically modulated when coding variants of *RSF1* and *IME1* were in
323 combination. This insight gave us a deeper understanding of how SNP-SNP interactions could
324 redirect metabolic flux, prioritizing nucleotide synthesis over steroid synthesis in the early
325 stages of meiotic progression during sporulation. Previous studies have underscored the
326 pivotal role of pyrimidine metabolism in facilitating efficient meiotic progression [15].
327 Further, the SNP-specific modulation of the glycine cleavage system was established in this
328 study, where the SNPs of *RME1* and *IME1nc* were synergistic.

329 We characterised the intracellular metabolic flux states of each SNP-specific model without
330 requiring an objective function as previously described [31,32]. Using GS-DFA analysis, we
331 identified flux variation in six major pathways as significant contributors to sporulation
332 efficiency variation. Interestingly, our GS-DFA analysis did not identify the fundamental
333 reactions that were known to constitute the meiosis-specific metabolic model [14]. These

334 included the TCA cycle, glyoxylate cycle, and acetate uptake, which have been shown to
335 influence sporulation at an early stage through gene-knockout approaches. However, we
336 were able to capture these pathways in our core meiosis-specific network by comparing the
337 common reactions shared among SNP-specific models. These findings suggested that SNPs
338 and their interactions might modulate sporulation efficiency variation by influencing
339 metabolic reactions beyond the essential ones required for sporulation, highlighting the
340 specific effects of genetic variations at the SNP level instead of the broader gene deletion
341 effects. Our analysis also predicted the differential regulation of the pentose phosphate
342 pathway in an *IME1nc*-specific manner. The pentose phosphate pathway is critical in
343 supplying the cell with the necessary building blocks and reducing equivalents needed for the
344 biosynthesis of nucleotides and amino acids [33]. Our hypothesis suggested that autophagy
345 could produce the essential precursors for nucleotide and amino acid biosynthesis by
346 degrading and recycling cellular components compensating for PPP, particularly in models
347 where PPP upregulation was absent (Figure S11). Previous studies have shown autophagy as
348 a compensatory mechanism for the PPP in tumour cells [34]. It would be interesting to
349 experimentally validate whether autophagosomes are induced when the PPP reaction is
350 downregulated. Thus, integrating SNP and their combination-specific gene expression data in
351 GSMM has provided several novel mechanisms by which phenotypic variation in a well-
352 studied trait is regulated at the molecular level.

353 Several studies have reported complex genetic interactions between identified SNPs in
354 metabolic diseases [35,36]. A recent study has used GWAS variant data to develop
355 personalized organ-specific metabolic models for 524,615 individuals of the INTERVAL and UK
356 Biobank cohorts to clarify the impact of genetic variations on the metabolic processes

357 implicated in coronary artery disease [37]. Studying the population-level data by clustering
358 based on variant combinations rather than phenotype along with GS-DFA analysis can provide
359 additional novel targets for therapeutic interventions in complex human metabolic diseases.
360 Applying SNP-specific genome-scale metabolic models that integrate transcriptomic and
361 metabolomic data will help identify known and novel pathways perturbed by SNPs and their
362 combinations. With scRNASeq becoming common and generated for diseases such as
363 cancers, single-cell GSMM [38,39] can open a new level of metabolic heterogeneity
364 corresponding to cellular heterogeneity. Combining this with the power of scRNASeq with
365 context-specific GSMM can provide novel insights into cellular heterogeneity and uncover
366 causes of disease, drug response, drug resistance, etc. Therefore, this can be a robust
367 framework to provide candidate genes for further validation in the molecular processes and
368 phenotype-genotype relationship, leading to clearer variant-to-function pathways.

369 **METHODS**

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

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

375 The reads were converted to Fastq format using the command 'Fastqdump' in the SRA toolkit.
376 The FastQC and MutiQC were performed for quality control [40]. The reads were mapped to
377 reference *Saccharomyces cerevisiae* R64-1-1.105 downloaded from the Ensembl database
378 (https://ftp.ensembl.org/pub/release110/fasta/saccharomyces_cerevisiae/cdna/Saccharom

379 yces_cerevisiae.R64-1-1.cdna.all.fa.gz) using Salmon (an alignment-free tool), accounting for
380 GC and sequence bias [41]. The transcript TPM values obtained from Salmon were converted
381 to gene-level counts using Tximport v1.4.0 [42]. The effect of day on the gene expression
382 values was removed using ComBatseq [43].

383 Differential gene expression analysis compared each allele replacement strain with the
384 wildtype strain using DESeq2 [44]. The batch-corrected untransformed raw counts were given
385 as input to the DESeq2. After filtering for low counts with less than 100 counts in all 63
386 samples, 6305 genes out of 6571 were selected for further processing. The effect of the day
387 was included in the design model of DESeq2 to account for the contribution of covariates to
388 the total gene expression. log2 fold changes (LFC) shrinkage using the alpegm method [45],
389 with a false discovery rate of 10% (p-value adjusted using the Benjamini-Hochberg method),
390 was used to identify differentially expressed genes (DEGs). Genes with LFC values greater than
391 0.5 and less than -0.5 are considered upregulated and down-regulated, respectively. The
392 upregulated DEGs obtained for each allele replacement strain are given as input to ClueGO
393 [46] from Cytoscape [47] for gene ontology and functional annotations analysis.

394 **Regulatory impact factor (RIF) analysis**

395 CeTF package in R was utilized to perform the Regulatory Impact Factors analysis [48,49]. The
396 Variance stabilizing transformation (VST) was performed on the gene expression data of
397 wildtype and strain with four SNPs. It was given as input for the RIF analysis as it required two
398 contrasting conditions. The Transcription factors (TFs), downloaded from the Yeastract
399 database (http://www.yeastract.com/, [50]), were contrasted to the list of DEGs identified by
400 comparing the wildtype strain (i.e., a strain with no SNPs) to the strain with all SNP

401 combinations. The TFs identified using the RIF analysis with scores deviating ± 1.96 SD from
402 the mean were considered significant and termed 'critical regulators.'

403 **PCIT and connectivity analysis**

404 The partial correlation and information theory (PCIT) approach [51] was used to arrive at co-
405 expression networks for wildtype and 14 SNP combinations except for strain with *IME1*, which
406 had only three replicates. PCIT evaluated all possible triplets of genes, identifying the strongly
407 associated pair after screening all the network's genes. The correlation of |0.95| and based
408 on the p-value < 0.05 was used as a threshold for arriving at a significant association (this
409 threshold was chosen to arrive at a maximum number of significant edges, [52]). Further, the
410 DEGs (specific for wild type vs. four SNP combination strains) and critical regulators identified
411 using the RIF analysis are filtered to arrive at sub-networks for each SNP-specific network. The
412 individual co-expression sub-networks for wild type and 14 allele replacement strains were
413 visualized using Cytoscape [47] version (3.9.1). Each subnetwork was analyzed for network
414 properties using the Cytoscape plugin NetworkAnalyser 4.4.8. The connectivity for each
415 critical regulator was identified by dividing the degree of that regulator in any specific
416 network with the highest degree in that network [52,53]. The connectivity of each critical
417 regulator was then compared across each SNP combination and was visualized using the
418 Pheatmap package in R (v1.0.12, https://CRAN.R-project.org/package=pheatmap).

419 **Reconstruction of SNP-specific metabolic models**

420 Yeast 8.4.2 [54], a commonly used yeast Genome-scale metabolic model, was used in this
421 study. It is accessible via GitHub (https://github.com/sysbiochalmers/yeast-gem). This GSMM
422 had 1,150 genes, 4,058 reactions, and 2,742 metabolites. The bounds of exchange reactions
423 of the GSMM were changed to reflect the early stages of the sporulation state appropriately.

424 For this, we restricted the intake of glucose and nitrogen (Lower bound=upper bound=0
425 [mmol/(g DW h)]) while allowing an unrestricted supply of oxygen and acetate (lower bound=
426 -1000 [mmol/(g DW h)], [14]).

427 The gene expression data (TPM values) averaged across replicates of wild type and 15 allele
428 replacement strains were pre-processed using Localgini thresholding [55] and integrated into
429 Yeast 8.4.2 using the model extraction method, iMAT [56]. The detailed methodology for
430 reconstructing an SNP-specific model using Localgini thresholding and iMAT is discussed in
431 Supplementary Methods. This led us to reconstruct 15 SNP-specific models (M1 to M15) and
432 a wildtype model (M16).

433 To understand the metabolic heterogeneity of each extracted model, we created a Boolean
434 reaction vector for each SNP-specific model. In brief, for each model, we made a Boolean
435 vector of length 4058 (corresponding to all the reactions in the Yeast8 metabolic model) with
436 1s and 0s, indicating whether a reaction was included in the SNP-specific model. Further, we
437 identified reactions in each SNP-specific model that were not in the wildtype model. These
438 reactions, active in SNP models while not in wildtype, were searched for significantly enriched
439 subsystems using the flux enrichment analysis function $(FEA.m)$ provided in the COBRA
440 toolbox [57]. In brief, $FEA.m$ used a hypergeometric 1-sided test (α = 0.05) and FDR
441 correction (adjusted p-value < 0.05) to obtain significantly enriched subsystems.

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

443 To discern how each SNP and their combinations can modulate the intracellular fluxes to
444 show sporulation efficiency variation, we leveraged the random flux sampling approach,
445 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 [59]. The optGpSampler in COBRApy version 0.10.1 [60] 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 [32]. 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 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.

468 **COBRA toolbox and COBRApy setup**

469 COBRA Toolbox version 3.0.0 and COBRApy version 0.10.1 on MATLAB R2021b and Python
470 version 3.5.2, respectively. The optimizer used for linear programming and mixed-integer
471 linear programming optimization was Gurobi, version 8.0.0. optGpSampler was run using
472 parallel processes. We used customized Python scripts for statistical analysis and visualization
473 of flux sampling results.

474 **CODE AVAILABILITY**

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

478 **ACKNOWLEDGEMENTS**

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

483 **AUTHOR CONTRIBUTIONS**

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

487 **DECLARATION OF INTERESTS**

488 The authors declare no competing interests.

489 **REFERENCES**

490 1. Visscher PM, Wray NR, Zhang Q, et al. 10 years of GWAS discovery: Biology, function, and
491 translation. *Am J Hum Genet*. 2017;101(1):5-22. doi:10.1016/j.ajhg.2017.06.005

492 2. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-
493 wide association studies. *Nat Rev Genet*. 2019;20(8):467-484. doi:10.1038/s41576-019-
494 0127-1

495 3. Cordell HJ. Detecting gene–gene interactions that underlie human diseases. *Nat Rev*
496 *Genet*. 2009;10(6):392-404. doi:10.1038/nrg2579

497 4. Martin MP, Gao X, Lee JH, et al. Epistatic interaction between KIR3DS1 and HLA-B delays
498 the progression to AIDS. *Nat Genet*. 2002;31(4):429-434. doi:10.1038/ng934

499 5. Wan X, Yang C, Yang Q, et al. BOOST: A fast approach to detecting gene-gene interactions
500 in genome-wide case-control studies. *Am J Hum Genet*. 2010;87(3):325-340.
501 doi:10.1016/j.ajhg.2010.07.021

502 6. Fang G, Wang W, Paunic V, et al. Discovering genetic interactions bridging pathways in
503 genome-wide association studies. *Nat Commun*. 2019;10(1):4274. doi:10.1038/s41467-
504 019-12131-7

505 7. Cano-Gamez E, Trynka G. From GWAS to function: Using functional genomics to identify
506 the mechanisms underlying complex diseases. *Front Genet*. 2020;11:424.
507 doi:10.3389/fgene.2020.00424

508 8. Battle A, Khan Z, Wang SH, et al. Genomic variation. Impact of regulatory variation from
509 RNA to protein. *Science*. 2015;347(6222):664-667. doi:10.1126/science.1260793

510 9. Olayinka OA, O'Neill NK, Farrer LA, Wang G, Zhang X. Molecular quantitative trait locus
511 mapping in human complex diseases. *Curr Protoc*. 2022;2(5):e426. doi:10.1002/cpz1.426

512 10. Tong H, Küken A, Razaghi-Moghadam Z, Nikoloski Z. Characterization of effects of genetic
513 variants via genome-scale metabolic modelling. *Cell Mol Life Sci*. 2021;78(12):5123-5138.
514 doi:10.1007/s00018-021-03844-4

515 11. Gu C, Kim GB, Kim WJ, Kim HU, Lee SY. Current status and applications of genome-scale
516 metabolic models. *Genome Biol*. 2019;20(1):121. doi:10.1186/s13059-019-1730-3

517 12. Øyås O, Borrell S, Trauner A, et al. Model-based integration of genomics and
518 metabolomics reveals SNP functionality in *Mycobacterium tuberculosis*. *Proc Natl Acad*
519 *Sci U S A*. 2020;117(15):8494-8502. doi:10.1073/pnas.1915551117

520 13. Sarkar D, Maranas CD. SNPeffect: identifying functional roles of SNPs using metabolic
521 networks. *Plant J*. 2020;103(2):512-531. doi:10.1111/tpj.14746

522 14. Ray D, Ye P. Characterization of the metabolic requirements in yeast meiosis. *PLoS One*.
523 2013;8(5):e63707. doi:10.1371/journal.pone.0063707

524 15. Walther T, Létisse F, Peyriga L, et al. Developmental stage dependent metabolic regulation
525 during meiotic differentiation in budding yeast. *BMC Biol*. 2014;12(1):60.
526 doi:10.1186/s12915-014-0060-x

527 16. Sudarsanam P, Cohen BA. Single nucleotide variants in transcription factors associate
528 more tightly with phenotype than with gene expression. *PLoS Genet*.
529 2014;10(5):e1004325. doi:10.1371/journal.pgen.1004325

530 17. Deutschbauer AM, Williams RM, Chu AM, Davis RW. Parallel phenotypic analysis of
531 sporulation and postgermination growth in *Saccharomyces cerevisiae*. *Proc Natl Acad Sci*
532 *U S A*. 2002;99(24):15530-15535. doi:10.1073/pnas.202604399

533 18. Yoon SW, Lee MS, Xaver M, et al. Meiotic prophase roles of Rec8 in crossover
534 recombination and chromosome structure. *Nucleic Acids Res*. 2016;44(19):9296-9314.
535 doi:10.1093/nar/gkw682

536 19. Holland MJ, Yokoi T, Holland JP, Myambo K, Innis MA. The *GCR1* gene encodes a positive
537 transcriptional regulator of the enolase and glyceraldehyde-3-phosphate dehydrogenase
538 gene families in *Saccharomyces cerevisiae*. *Mol Cell Biol*. 1987;7(2):813-820.
539 doi:10.1128/mcb.7.2.813-820.1987

540 20. Daignan-Fornier B, Fink GR. Coregulation of purine and histidine biosynthesis by the
541 transcriptional activators *BAS1* and *BAS2*. *Proc Natl Acad Sci U S A*. 1992;89(15):6746-
542 6750. doi:10.1073/pnas.89.15.6746

543 21. Primig M, Williams RM, Winzeler EA, et al. The core meiotic transcriptome in budding
544 yeasts. *Nat Genet*. 2000;26(4):415-423. doi:10.1038/82539

545 22. Enyenihi AH, Saunders WS. Large-scale functional genomic analysis of sporulation and
546 meiosis in *Saccharomyces cerevisiae*. *Genetics*. 2003;163(1):47-54.
547 doi:10.1093/genetics/163.1.47

548 23. Lu L, Roberts G, Simon K, Yu J, Hudson AP. Rsf1p, a protein required for respiratory growth
549 of *Saccharomyces cerevisiae*. *Curr Genet*. 2003;43(4):263-272. doi:10.1007/s00294-003-
550 0398-z

551 24. Roberts GG 3rd, Hudson AP. Rsf1p is required for an efficient metabolic shift from
552 fermentative to glycerol-based respiratory growth in *S. cerevisiae*. *Yeast*. 2009;26(2):95-
553 110. doi:10.1002/yea.1655

554 25. Rubin-Bejerano I, Mandel S, Robzyk K, Kassir Y. Induction of meiosis in *Saccharomyces*
555 *cerevisiae* depends on conversion of the transcriptional repressor Ume6 to a positive
556 regulator by its regulated association with the transcriptional activator Ime1. *Mol Cell Biol*.
557 1996;16(5):2518-2526. doi:10.1128/MCB.16.5.2518

26. Gertz J, Gerke JP, Cohen BA. Epistasis in a quantitative trait captured by a molecular model of transcription factor interactions. *Theor Popul Biol*. 2010;77(1):1-5. doi:10.1016/j.tpb.2009.10.002

27. Orth JD, Thiele I, Palsson BØ. What is flux balance analysis? *Nat Biotechnol*. 2010;28(3):245-248. doi:10.1038/nbt.1614

28. Gudmundsson S, Thiele I. Computationally efficient flux variability analysis. *BMC Bioinformatics*. 2010;11(1):489. doi:10.1186/1471-2105-11-489

29. Barve G, Manjithaya R. Cross-talk between autophagy and sporulation in *Saccharomyces cerevisiae*. *Yeast*. 2021;38(7):401-413. doi:10.1002/yea.3556

30. Wen FP, Guo YS, Hu Y, et al. Distinct temporal requirements for autophagy and the proteasome in yeast meiosis. *Autophagy*. 2016;12(4):671-688. doi:10.1080/15548627.2016.1149659

31. Scott WT Jr, Smid EJ, Block DE, Notebaart RA. Metabolic flux sampling predicts strain-dependent differences related to aroma production among commercial wine yeasts. *Microb Cell Fact*. 2021;20(1):204. doi:10.1186/s12934-021-01694-0

32. Nanda P, Ghosh A. Genome Scale-Differential Flux Analysis reveals deregulation of lung cell metabolism on SARS-CoV-2 infection. *PLoS Comput Biol*. 2021;17(4):e1008860. doi:10.1371/journal.pcbi.1008860

33. Bertels LK, Fernández Murillo L, Heinisch JJ. The pentose phosphate pathway in yeasts-more than a poor cousin of glycolysis. *Biomolecules*. 2021;11(5):725. doi:10.3390/biom11050725

34. Parkhitko AA, Priolo C, Coloff JL, et al. Autophagy-dependent metabolic reprogramming sensitizes TSC2-deficient cells to the antimetabolite 6-aminonicotinamide. *Mol Cancer Res*. 2014;12(1):48-57. doi:10.1158/1541-7786.MCR-13-0258-T

35. Lin E, Kuo PH, Liu YL, Yang AC, Kao CF, Tsai SJ. Association and interaction of APOA5, BUD13, CETP, LIPA and health-related behavior with metabolic syndrome in a Taiwanese population. *Sci Rep*. 2016;6(1). doi:10.1038/srep36830

36. Cadena-López RO, Hernández-Rodríguez LV, Aguilar-Galarza A, et al. Association between SNPs in Leptin pathway genes and anthropometric, biochemical, and dietary markers related to obesity. *Genes (Basel)*. 2022;13(6):945. doi:10.3390/genes13060945

37. Foguet C, Xu Y, Ritchie SC, et al. Genetically personalised organ-specific metabolic models in health and disease. *Nat Commun*. 2022;13(1):7356. doi:10.1038/s41467-022-35017-7

38. Gustafsson J, Anton M, Roshanzamir F, 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*. 2023;120(6):e2217868120. doi:10.1073/pnas.2217868120

39. Huang Y, Mohanty V, Dede M, et al. Characterizing cancer metabolism from bulk and single-cell RNA-seq data using METAFlux. *Nat Commun*. 2023;14(1):4883. doi:10.1038/s41467-023-40457-w

40. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047-3048. doi:10.1093/bioinformatics/btw354

41. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. *Nat Methods*. 2017;14(4):417-419. doi:10.1038/nmeth.4197

42. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Res*. 2015;4:1521. doi:10.12688/f1000research.7563.1

43. Zhang Y, Parmigiani G, Johnson WE. ComBat-seq: batch effect adjustment for RNA-seq count data. *NAR Genom Bioinform*. 2020;2(3):lqaa078. doi:10.1093/nargab/lqaa078

44. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol*. 2014;15(12):550. doi:10.1186/s13059-014-0550-8

45. Zhu A, Ibrahim JG, Love MI. Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. *Bioinformatics*. 2019;35(12):2084-2092. doi:10.1093/bioinformatics/bty895

46. Bindea G, Mlecnik B, Hackl H, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. *Bioinformatics*. 2009;25(8):1091-1093. doi:10.1093/bioinformatics/btp101

47. Shannon P, Markiel A, Ozier O, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*. 2003;13(11):2498-2504. doi:10.1101/gr.1239303

48. Oliveira de Biagi CA Jr, Nociti RP, Brotto DB, 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):624. doi:10.1186/s12864-021-07918-2

49. 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(7):896-904. doi:10.1093/bioinformatics/btq051

50. Teixeira MC, Viana R, Palma M, et al. YEASTRACT+: a portal for the exploitation of global transcription regulation and metabolic model data in yeast biotechnology and pathogenesis. *Nucleic Acids Res*. 2023;51(D1):D785-D791. doi:10.1093/nar/gkac1041

51. 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*. 2010;26(3):411-413. doi:10.1093/bioinformatics/btp674

52. Banerjee P, Balraj P, Ambhore NS, et al. Network and co-expression analysis of airway smooth muscle cell transcriptome delineates potential gene signatures in asthma. *Sci Rep*. 2021;11(1):14386. doi:10.1038/s41598-021-93845-x

53. 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. *Front Genet*. 2020;11:517. doi:10.3389/fgene.2020.00517

54. Lu H, Li F, Sánchez BJ, et al. A consensus S. cerevisiae metabolic model Yeast8 and its ecosystem for comprehensively probing cellular metabolism. *Nat Commun*. 2019;10(1):3586. doi:10.1038/s41467-019-11581-3

55. Pavan Kumar S, Bhatt NP. Localgini: A method for harnessing inequality in gene expression to improve the quality of context-specific models. *bioRxiv*. Published online 17 September 2023. doi:10.1101/2023.09.14.557840

56. Zur H, Ruppin E, Shlomi T. iMAT: an integrative metabolic analysis tool. *Bioinformatics*. 2010;26(24):3140-3142. doi:10.1093/bioinformatics/btq602

57. Heirendt L, Arreckx S, Pfau T, et al. Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v.3.0. *Nat Protoc*. 2019;14(3):639-702. doi:10.1038/s41596-018-0098-2

58. 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(2):e86587. doi:10.1371/journal.pone.0086587

59. Herrmann HA, Dyson BC, Vass L, Johnson GN, Schwartz JM. Flux sampling is a powerful tool to study metabolism under changing environmental conditions. *NPJ Syst Biol Appl*. 2019;5(1):32. doi:10.1038/s41540-019-0109-0

60. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR. COBRApy: COnstraints-Based Reconstruction and analysis for Python. *BMC Syst Biol*. 2013;7(1):74. doi:10.1186/1752-0509-7-

659 **FIGURE LEGENDS**

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

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

669 **Figure 2: Transcriptional differences between the allele replacement strains**

670 (A) Number of upregulated and downregulated genes (LFC>0.5 and LFC<-0.5) identified using
671 DeSeq2 between wildtype vs. each SNP combination. (B) The number of differentially
672 expressed ((LFC > 0.5 and LFC < 0.5) sporulation-specific and metabolic genes identified using
673 DeSeq2 between wildtype vs. each SNP. (C) Venn diagram plot for upregulated genes
674 identified using DeSEq2 with LFC > 0.5 and p-adj > 0.1 for three SNP combination strains
675 (*RME1_RSF1_IME1*, *RME1_IME1_IME1nc*, *RSF1_IME1_IME1nc* and *RME1_RSF1_IME1nc*) and
676 four SNP combination strain (*RME1_RSF1_IME1_IME1nc*) strains in comparison with the
677 wildtype. (D) Venn diagram analysis for *RSF1* and *IME1* SNP combinations. (E) Heatmap
678 representing the LFC values of *IME1*, *RIM4*, and *REC8* genes across all SNP combination strains
679 compared to the wildtype strain

680 **Figure 3: Analysis of connectivity of critical metabolic regulators**

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

685 **Figure 4: SNPs and their interaction can activate unique reactions and subsystems**

686 (A) Metabolic subsystems enrichment analysis for reactions that are over-represented in each
687 comparison that are present exclusively in single SNP models compared with the wildtype
688 model. (B) Synergistic effect of *RSF1* and *IME1* in modulating nucleotide metabolism. Here,
689 the violin plots represent the flux distribution obtained by optGpSampler for key reactions
690 involved in purine and pyrimidine metabolism compared across wildtype (M16), *RSF1+IME1*
691 (M8), *RSF1* (M13) and *IME1* (M14) models.

692 **Figure 5: Pentose phosphate pathway is differentially regulated in SNP models**

693 (A) Flux enrichment plots for unique reactions in highly sporulating strains compared to the
694 wildtype strain. (B) Flux enrichment plots for upregulated reactions in highly sporulating
695 strains compared to the wildtype strain obtained after GF-DFA analysis. (C) Box plots
696 representing the flux distribution of key reactions involved in the pentose phosphate pathway
697 obtained from optGpSampler for three and four SNP models (M1 to M5) compared to the
698 wildtype (M16) model.

699

700 **SUPPLEMENTARY METHODS**

701 Supplementary methods accompany this manuscript.

702 **SUPPLEMENTARY TABLE LEGENDS**

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

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

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

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

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

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

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

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

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

721 **SUPPLEMENTARY FIGURE LEGENDS**

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

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

726 **Figure S3:** Over-represented pathways of upregulated differentially expressed genes between
727 wildtype and (A) *RME1_RSF1_IME1_IME1nc*, (B) *RME1_RSF1_IME1*, (C) *RSF1_IME1_IME1nc*,
728 (D) *RME1_IME1nc* and (E) *IME1_IME1nc* strains. The circle size for a specific GO Term
729 (pathway) varies with the order of significance (p-value) level. The most prominent node
730 represents the most significant pathway. The pathways were constructed using Cytoscape
731 v3.8.0 (https://cytoscape.org/) using the plugin ClueGO v2.5.41.

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

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

736 **Figure S6: Metabolic subsystems enrichment analysis reveals the metabolic role of *IME1***
737 **functional and non-functional forms.** (A) Flux enrichment analysis for reactions that are
738 exclusively in each model in comparison with the wildtype (B) Flux enrichment analysis of
739 upregulated reactions obtained using GS-DFA analysis between wildtype and *IME1*, *IME1nc*

740 and *IME1_IME1nc* strains. The presented reactions have an FC value (Flux Change) greater
741 than 0.8 (as described in materials and methods) and have an adjusted p-value less than 0.05.

742 **Figure S7: Antagonistic effect of *RSF1* and *IME1* in modulating 'Steroid biosynthesis'.** The
743 violin plots represent the flux distribution obtained by optGpSampler for key reactions
744 involved in **'Steroid biosynthesis'** compared across wildtype (M16), *RSF1_IME1* (M8), *RSF1*
745 (M13) and *IME1* (M14) models.

746 **Figure S8: Synergistic effect of *RME1* and *IME1nc* in modulating 'one carbon pool by folate'**
747 **metabolism.** The violin plots represent the flux distribution obtained by optGpSampler for
748 key reactions involved in **'one carbon pool by folate' metabolism** compared across wildtype
749 (M16), *RME1_IME1nc* (M9), *RME1* (M12) and *IME1c* (M15) models.

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

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

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

Figure 1

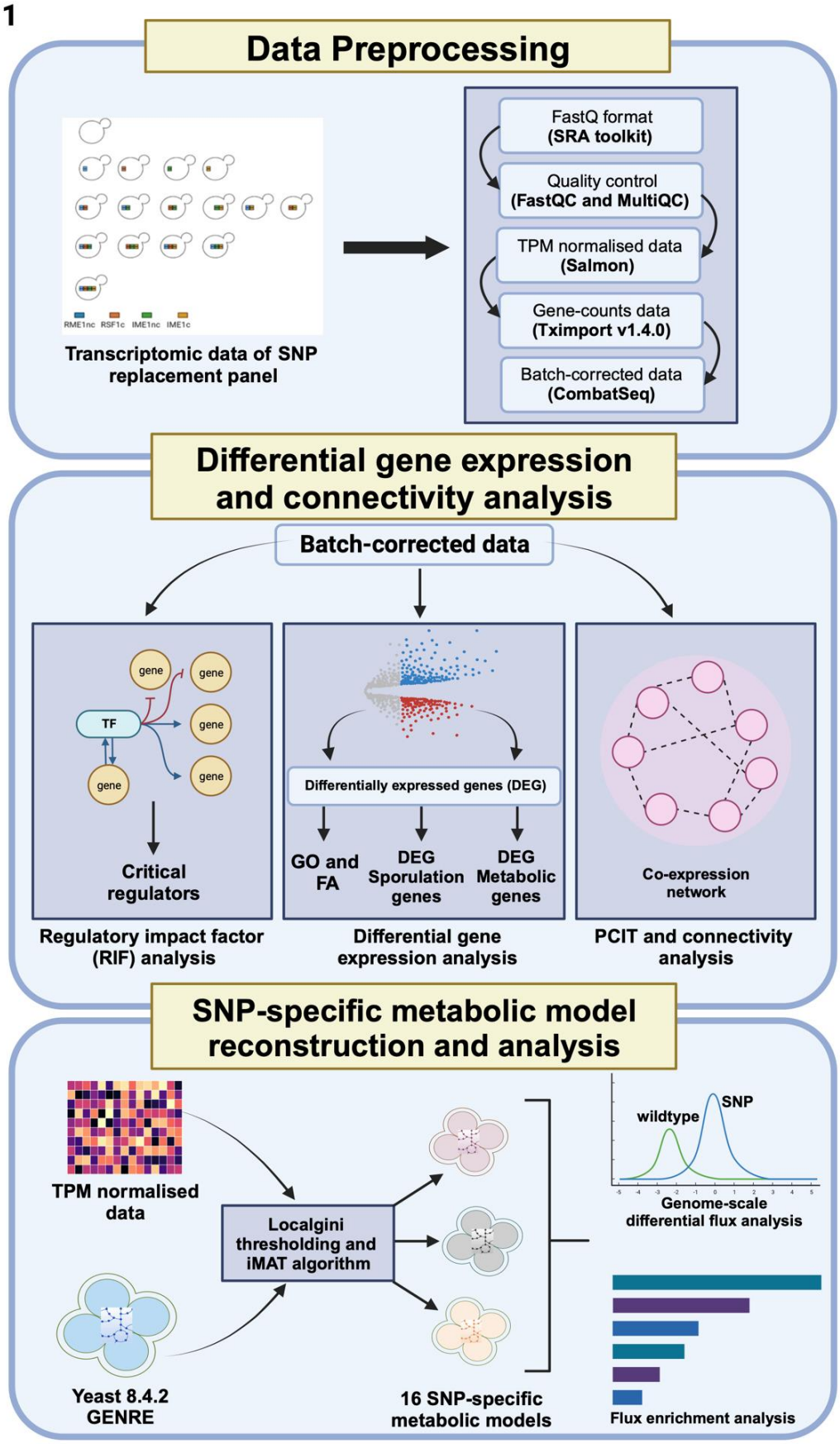


Figure 1: Schematic representation outlines the key steps and methodologies employed in our study

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. (created with BioRender.com)

Figure 2

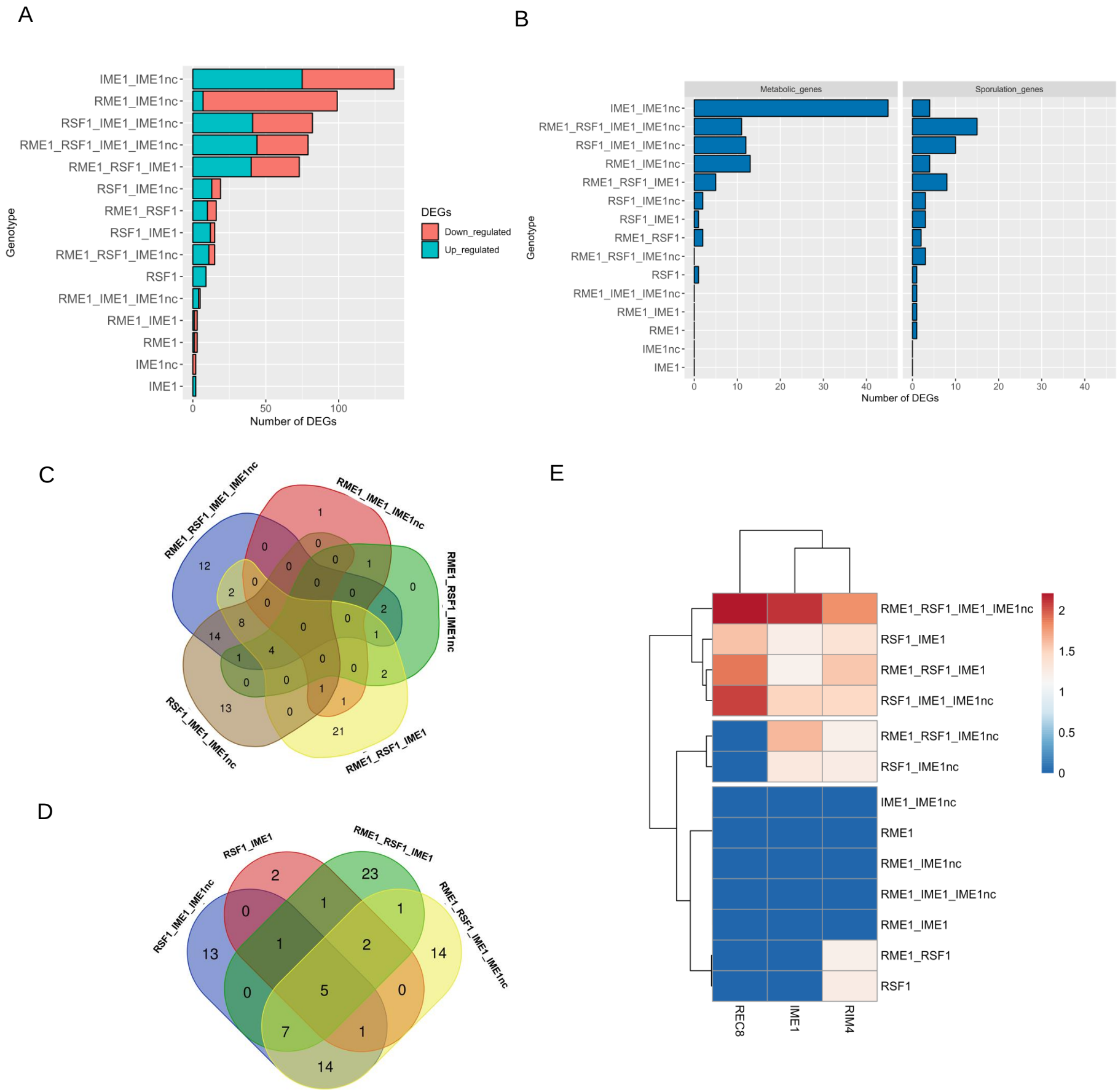
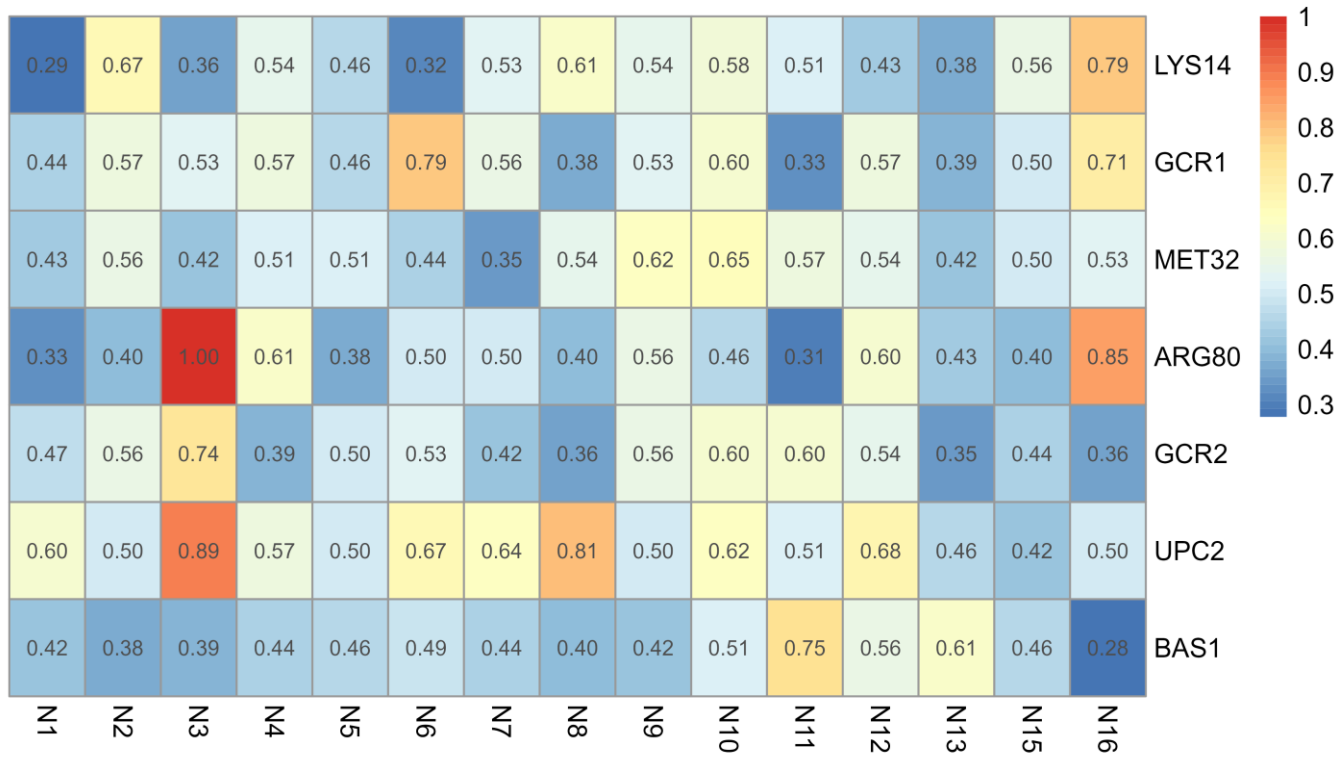


Figure 2: Transcriptional differences between the allele replacement strains
(A) Number of upregulated and downregulated genes (LFC>0.5 and LFC<-0.5) identified using DeSeq2 between wildtype vs. each SNP combination. (B) Number of differentially expressed ((LFC > 0.5 and LFC < 0.5) sporulation-specific 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

A



B

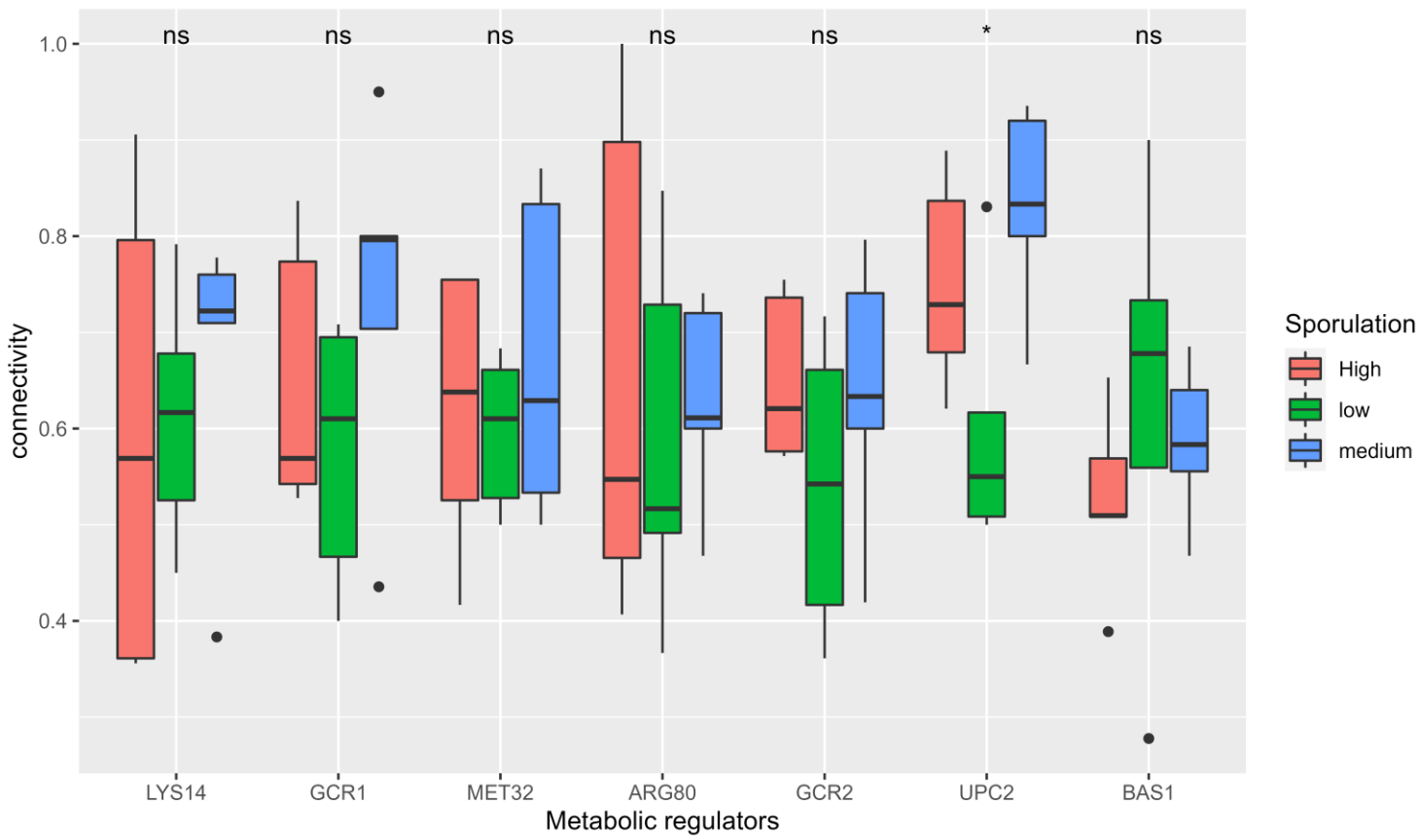


Figure 3: Analysis of connectivity of critical metabolic regulators
(A) Connectivity of critical metabolic regulators identified using RIF analysis across combinations of SNPs. (B) Comparing the connectivity of critical metabolic regulators based on the sporulation efficiency of the strains. Significance was calculated using a t-test (ns represents non-significant).

Figure 4

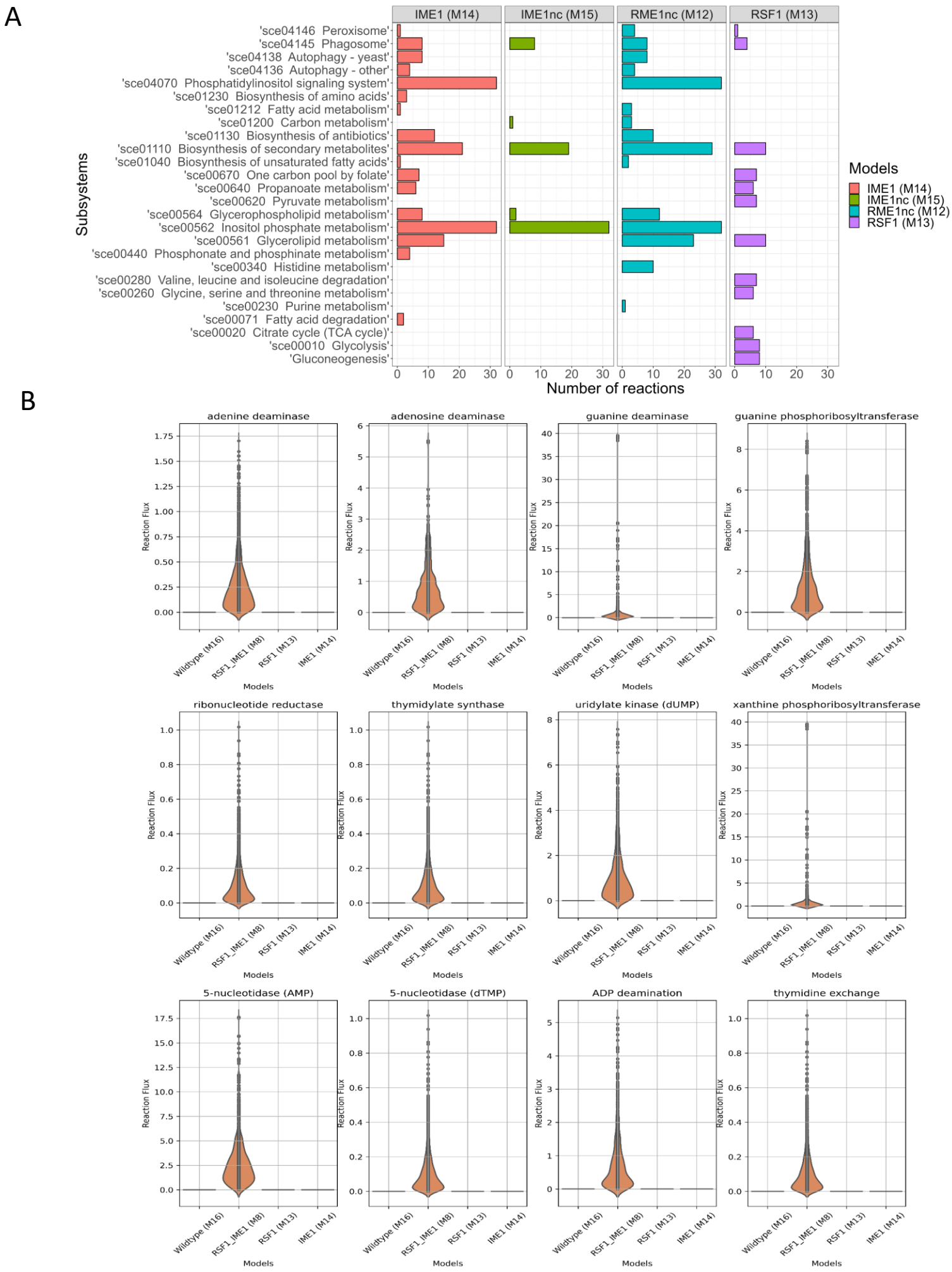
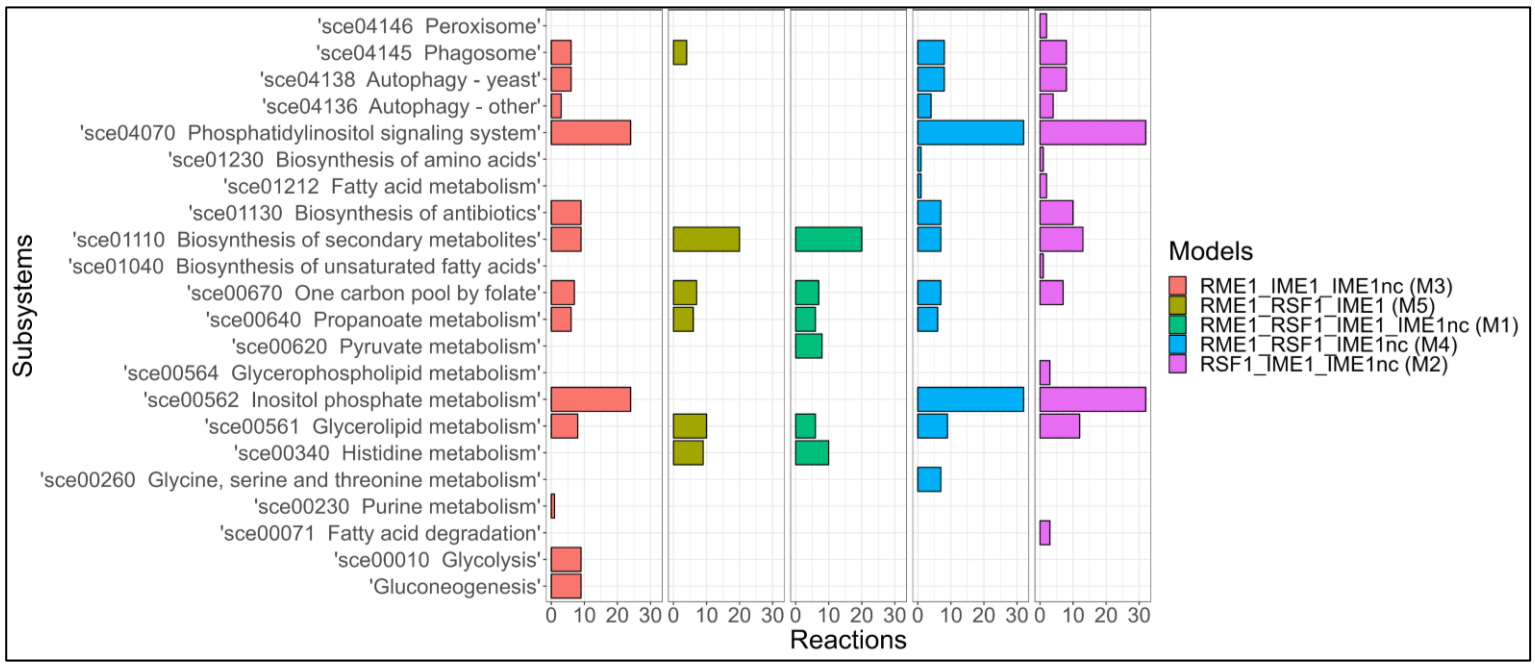


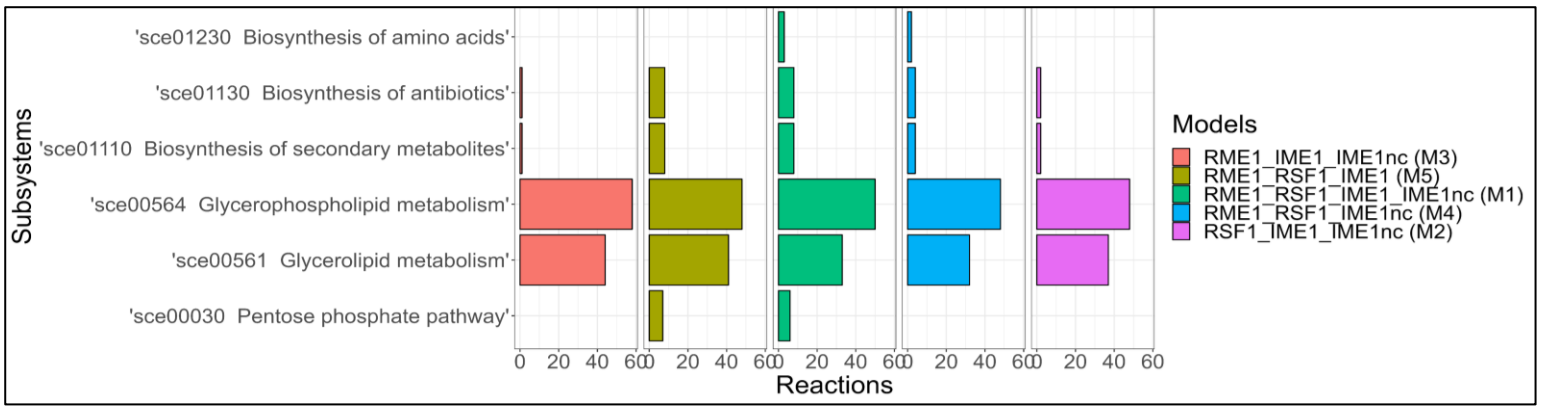
Figure 4: SNPs and their interaction can activate unique reactions and subsystems
(A) Metabolic subsystems enrichment analysis for reactions that are over-represented in each comparison that are present exclusively in single SNP models compared with the wildtype model. (B) Synergistic effect of *RSF1* 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

A



B



C

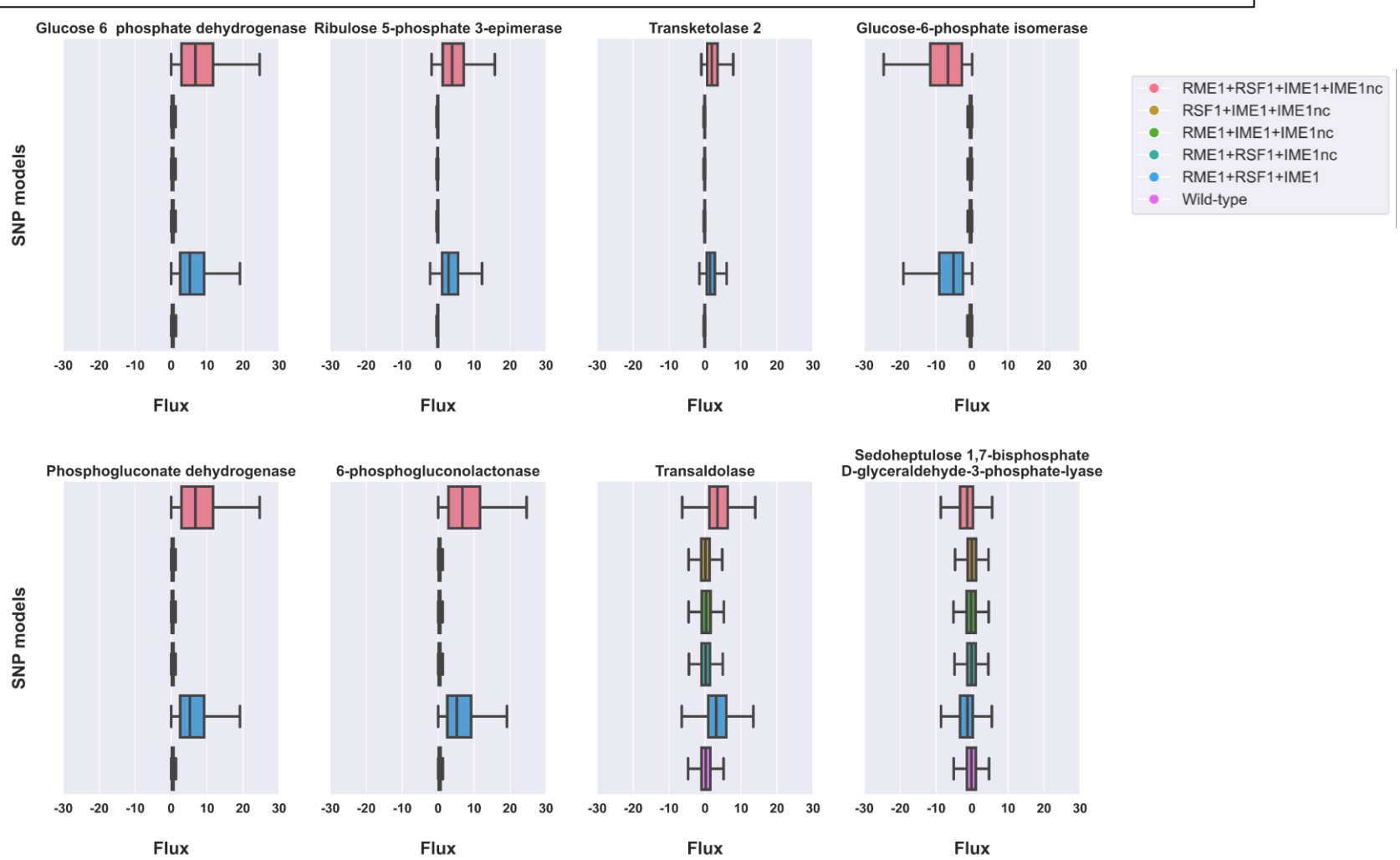


Figure 5: Pentose phosphate pathway is differentially regulated in SNP models (A) Flux enrichment plots for unique reactions in highly sporulating strains compared to the wildtype strain (B) Flux enrichment plots for upregulated reactions in highly sporulating strains compared to the wildtype strain obtained after GF-DFA analysis. (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.