



## OPEN

## SUBJECT AREAS:

INDUSTRIAL  
MICROBIOLOGY

COMPARATIVE GENOMICS

Received  
9 July 2014Accepted  
10 September 2014Published  
9 October 2014

Correspondence and requests for materials should be addressed to Z.L.L. (zlewis.liu@ars.usda.gov) or J.X. (xujian@qibebt.ac.cn)

\* These authors contributed equally to this work.

# Genomic and transcriptome analyses reveal that MAPK- and phosphatidylinositol-signaling pathways mediate tolerance to 5-hydroxymethyl-2-furaldehyde for industrial yeast *Saccharomyces cerevisiae*

Qian Zhou<sup>1\*</sup>, Z. Lewis Liu<sup>2\*</sup>, Kang Ning<sup>1\*</sup>, Anhui Wang<sup>3</sup>, Xiaowei Zeng<sup>1</sup> & Jian Xu<sup>1</sup>

<sup>1</sup>Single-Cell Center, CAS Key Laboratory of Biofuels and Shandong Key Laboratory of Energy Genetics, Qingdao Institute of BioEnergy and Bioprocess Technology, Chinese Academy of Sciences, Qingdao, Shandong, China 266101, <sup>2</sup>Bioenergy Research Unit, National Center for Agricultural Utilization Research, USDA-ARS, Peoria, IL, USA 61604-3902, <sup>3</sup>College of Computer and Information Technology, China Three Gorges University, Yichang, Hubei, China.

The industrial yeast *Saccharomyces cerevisiae* is a traditional ethanologenic agent and a promising biocatalyst for advanced biofuels production using lignocellulose materials. Here we present the genomic background of type strain NRRL Y-12632 and its transcriptomic response to 5-hydroxymethyl-2-furaldehyde (HMF), a commonly encountered toxic compound liberated from lignocellulosic-biomass pretreatment, in dissecting the genomic mechanisms of yeast tolerance. Compared with the genome of laboratory model strain S288C, we identified more than 32,000 SNPs in Y-12632 with 23,000 missense and nonsense SNPs. Enriched sequence mutations occurred for genes involved in MAPK- and phosphatidylinositol (PI)- signaling pathways in strain Y-12632, with 41 and 13 genes containing non-synonymous SNPs, respectively. Many of these mutated genes displayed consistent up-regulated signature expressions in response to challenges of 30 mM HMF. Analogous single-gene deletion mutations of these genes showed significantly sensitive growth response on a synthetic medium containing 20 mM HMF. Our results suggest at least three MAPK-signaling pathways, especially for the cell-wall integrity pathway, and PI-signaling pathways to be involved in mediation of yeast tolerance against HMF in industrial yeast *Saccharomyces cerevisiae*. Higher levels of sequence variations were also observed for genes involved in purine and pyrimidine metabolism pathways.

Production of advanced biofuels including cellulosic ethanol has attracted growing interest with increased concerns over a sustainable energy supply and a cleaner environment<sup>1,2</sup>. Development of the next-generation biocatalyst with robust and tolerance characteristics is a necessity and among the significant challenges for cost-efficient bio-based economy. The traditional industrial ethanologenic yeast *Saccharomyces cerevisiae* is a candidate for the next-generation biocatalyst development due to its high ethanol titer and robust performance in ethanol conversion. Numerous efforts have been made to improve *S. cerevisiae* for advanced biofuels applications, however, results obtained from laboratory strains and industrial yeast strains are not always consistent. For example, lab model strains of *S. cerevisiae* typically show transient expression responses to various environmental stimuli<sup>3,4</sup> while industrial yeast strains, on the other hand, displayed more persistent expressions against biomass inhibitory compounds<sup>5,6</sup>. Using a lab strain as a host, poor or negative function of heterologous xylose transporter genes was observed<sup>7,8</sup>. In contrast, when an industrial strain was used as a host, several heterologous xylose transporters were found to exhibit strong expressions and significant functional improvement of xylose uptake



and utilization<sup>9</sup>. Speculations behind these differences are many and most of these remain unresolved. The lab strains equipped with many genetic markers for phenotype selections are convenient, however, such strains often have mutated gene copies and allelic variations (<http://www.yeastgenome.org/>) that may not necessarily reflect response or performance of an industrial yeast<sup>10</sup>. For ethanogenic yeast, many genes and metabolic pathways were found to be responsible for varied stress response<sup>11,12</sup>. Sequence rearrangements and copy number changes of adaptive genes are likely related to unique metabolic pathways of industrial yeasts<sup>13,14</sup>. However, knowledge about genome background of industrial yeast especially the genome structure in relationship to stress tolerance, including inhibitor tolerance, remains largely unknown. Using known mechanisms of well-characterized model yeast to interpret industrial yeast performance often is false or misleading. The lack of clearly defined knowledge on genome origin of different yeast strains hinders development of the next-generation biocatalysts for advanced biofuels production.

In this study, using an industrial yeast strain NRRL Y-12632 as an example, we investigated its genome sequence variations and transcriptome response in comparison with a model strain S288C against 5-hydroxymethyl-2-furaldehyde (HMF), a representative toxic compound derived from lignocellulose biomass pretreatment. Strain Y-12632 was selected based on its more tolerant phenotypes and response to inhibitory compounds from 16 industrial yeast strains evaluated<sup>15</sup>. It is a commonly recognized industrial type strain also known as CBS1171 and was originally isolated from the brewer's top yeast in Netherlands in 1925 and deposited at the Centraalbureau voor Schimmelcultures (CBS), Utrecht, The Netherlands<sup>16</sup>. As a type strain, it has been collected worldwide and is also known as ATCC 18824, AWRI74, CCRC 21447, DBVPG 6173, DSM 70449, IFO 10217, IGC 4455, JCM 7255 and NCYC 505<sup>17</sup>. Novel aldehyde reductase genes and candidate tolerant genes to biomass fermentation inhibitors were identified from strain NRRL Y-12632<sup>18,19</sup>. Key regulatory elements in genomic adaptation of strain Y-12632 to HMF were reported<sup>5</sup>. Using an integrated approach of genome and transcriptome analyses, we identified extensive genome-wide sequence variations of the industrial yeast Y-12632 when compared with the model strain S288C. Our transcriptome analysis of Y-12632 revealed the first insight into significant roles of tolerant signaling pathways in yeast against biomass pretreatment inhibitory compounds. The unique genomic background of strain Y-12632, the high levels of its genome sequence variations and the signature expressions uncovered by this study provide new knowledge to understand tolerance of industrial yeast. Results of this study aid development of the next-generation biocatalysts for advanced biofuels production.

## Results and Discussion

### Genome sequence analysis of industrial type strain NRRL Y-12632.

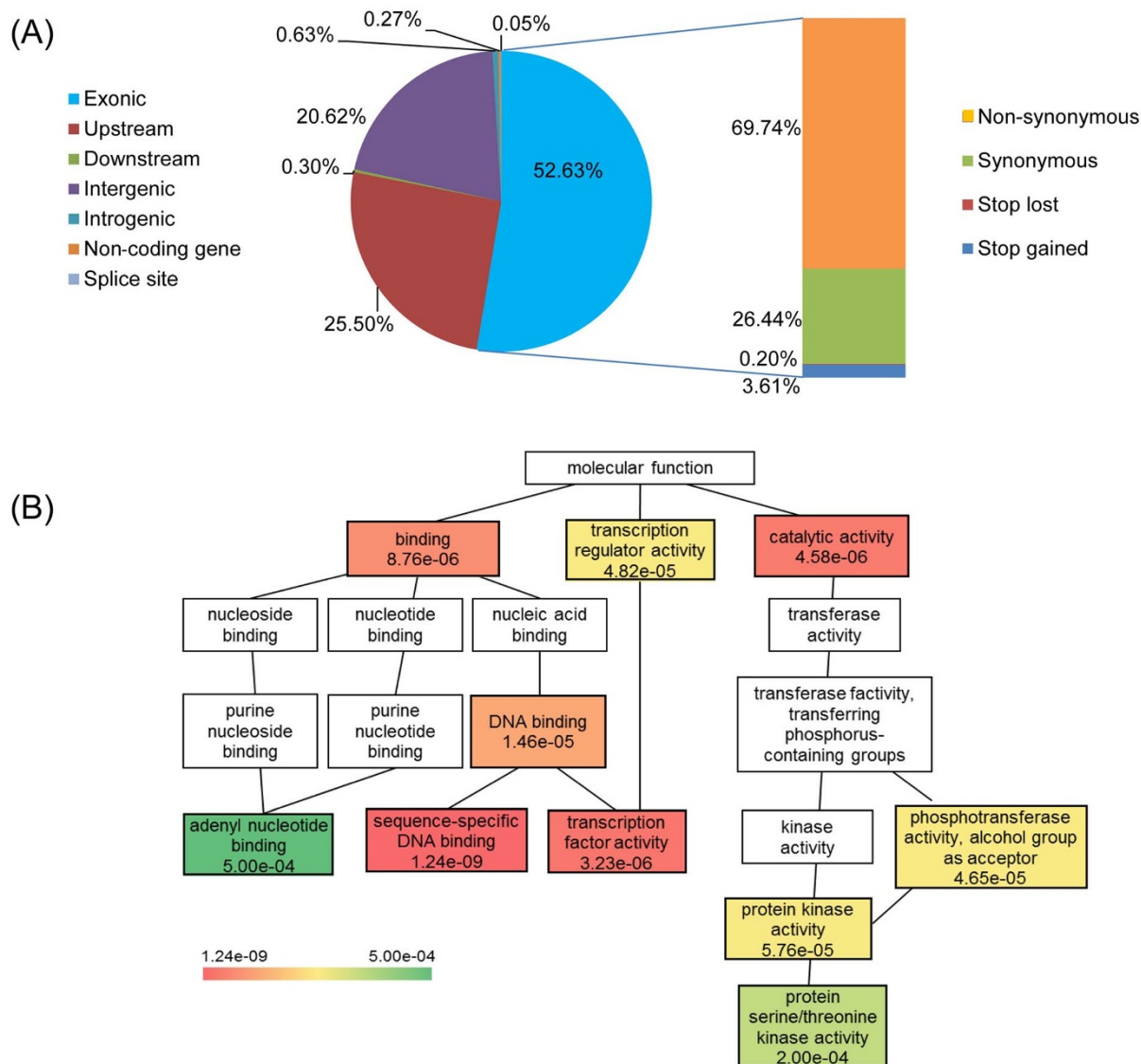
The Y-12632 genome was sequenced using Illumina technology GA-IIx. High quality sequencing data of 23 million reads were obtained by stringent trimming with 160-fold coverage of the genome. The *de novo* assembly was performed by Velvet 1.0.14<sup>20</sup>, producing 238 contigs, with N50 of 94.2 kbp and a total length of 11.6 Mbp (Table S1). Based on the assembly, 5,380 polypeptide-encoding Open Reading Frames (ORFs) were predicted with gene model of S288C as a reference. Through protein sequence alignment analysis, we found 98.6% and 98.3% of these ORFs showing high degrees of homology to entries in non-redundant protein (NR, <http://www.ncbi.nlm.nih.gov/>) database and Kyoto Encyclopedia of Genes and Genomes (KEGG, <http://www.kegg.jp/>) database, respectively. A total of 3,605 protein encoding genes were annotated by InterProScan<sup>21</sup>. We observed an ORF for gene *FPG1* which is not present in the S288C genome. *FPG1* was recently reported to encode a mannoprotein precursor promoting foam in wine strains<sup>22</sup>. A few other ORFs not present in S288C

were found encoding hypothetical proteins or unknown functions. These sequences were also commonly observed in other wine or beer yeast strains such as EC1118<sup>23</sup>. Based on the genome sequences (SGD, <http://www.yeastgenome.org/>) of eight yeast strains from a variety of sources and application scopes, including laboratory, bakery, wine, industrial bioethanol, sake and natural isolates, as well as that of Y-12632, 942 single-copy orthologous gene-sets were identified and a phylogenetic tree was generated accordingly (Fig. S1). It is known that significant genetic variability exists among a large number of *S. cerevisiae* strains from wild type, laboratory, and industrial sources. In general, cluster of yeast strains based on sequence variations was more closely related to the area of technological applications rather than geographical origin<sup>24</sup>. Consistent with this observation, our phylogenetic analysis showed that Y-12632 was not closely related to the laboratory model, wine and other types of strains (Fig. S1).

Using the S288C genome as a reference, we identified a total of 32,811 single nucleotide polymorphisms (SNPs) in Y-12632 with stringent filtration of high sequencing depth (>70) and quality value (>2100) (Supplementary Dataset 1; Fig. S2). PCR sequencing verifications of 69 selective SNPs (from 79 loci) confirmed 87% accuracy of the global map of SNPs (Methods). The SNPs in Y-12632 were observed to be distributed widely across 16 chromosomes and the mitochondrion genome of the yeast (Supplementary Dataset 1). Approximate 16,000 detected SNPs in Y-12632 were found in exonic regions and 6,691 SNPs located at intergenic regions (Fig. 1A). More than 8,000 mutations were detected to reside within 5-kb upstream of the coding regions, which potentially affect protein binding motifs that interfere with gene transcription regulatory functions. Approximately 70% of exonic SNPs were non-synonymous mutations involved in 3,740 genes. Genes harboring these variations were found significantly enriched under molecular functions and categories of sequencing-specific DNA binding ( $P = 1.24e-09$ ), transcription regulator activity ( $P = 3.23e-06$ ), adenylyl nucleotide binding ( $P = 5.00e-04$ ), and protein serine/threonine kinase activity ( $P = 2.00e-04$ ) (Fig. 1B). A summary of all SNPs identified in Y-12632 with overlapping protein domain and GO terms are presented (Supplementary Dataset 1, Table S2). More than 700 SNPs showed variations involved in stop codons, which accounted for approximately 4% of the total SNPs (Fig. 1A).

**Sequence variation impacts global metabolism pathways.** At the genomic level, over 23,000 missense and nonsense SNPs were associated with protein-coding genes from 94 KEGG pathways (Fig. 2). To probe the functional implication of these SNPs, we analyzed a set of time-course data on global gene expression in response to HMF for Y-12632. Differential gene expression during the lag phase is considered as an adaptive response of the yeast to the toxic compounds liberated by the biomass pretreatment<sup>5,6</sup>. Using Significance Analysis of Microarrays (SAM)<sup>25</sup> with a time-course mode, we identified a total of 2,736 genes that were up-regulated upon HMF challenges (Table S3). These genes were involved in 90 KEGG pathways, among which, 60 metabolic pathways were found containing both enhanced expressed genes and genes with non-synonymous SNPs involving more than 1,700 genes (Fig. 2; Table S4).

Many of these genes were found to be associated with multiple pathways such as carbohydrate, amino acid and nucleotide metabolism (Fig. 2; Table S4). Tolerant responses of industrial yeast to pretreatment inhibitors have been frequently reported in association with pathways of glycolysis, butanate metabolism, alanine, aspartate and glutamate metabolism, valine, leucine and isoleucine biosynthesis, and the metabolisms of arginine, proline, histidine, tyrosine, tryptophan and glutathione<sup>12,26</sup>. Findings in this study were consistent with those previously reported. In addition, we found a large amount of single nucleotide variations to be associated with nucle-



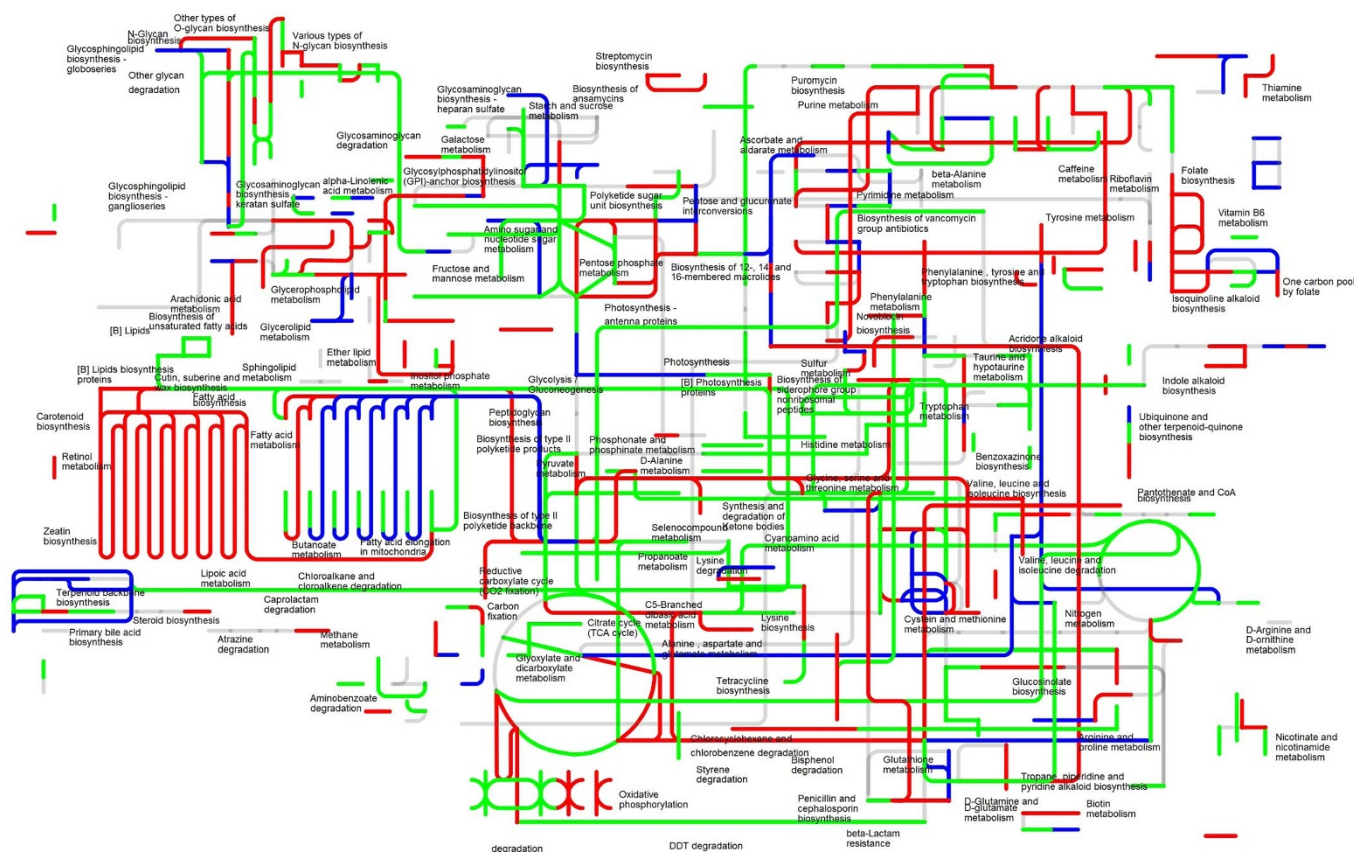
**Figure 1** | Genome sequence analysis of industrial yeast *Saccharomyces cerevisiae* type strain NRRL Y-12632. (A) Distribution of single nucleotide variations in the Y-12632 genome. (B) Functional categories of genes containing SNPs in Y-12632 genome with significant higher levels of variations marked against reference genome S288C.

tide metabolism, particularly in purine metabolism and pyrimidine metabolism pathways, each had 35 and 23 SNPs, respectively. Purines and pyrimidines are two of the building blocks of nucleic acids. The synthesis pathways of these nucleotides involve generation of many kinds of energy molecules such as commonly used ATP in most reactions, GTP in protein synthesis, UTP for activating glucose and galactose, CTP in lipid metabolism and AMP as a structural component of NAD and coenzyme A. These pathways are also closely related to pentose phosphate pathways and many amino acid metabolism pathways and worth further investigation for yeast tolerance studies. Furthermore, our study revealed two tolerant signal transduction pathways, the mitogen-activated protein kinases (MAPK) signaling pathway and the phosphatidylinositol signaling system, that have a great impact on yeast performance against HMF and possible other biomass pretreatment inhibitors.

**Tolerant MAPK signaling pathways.** Numerous metabolic pathways have been reported affecting yeast tolerance by high throughput assays including transcriptome and proteome analysis<sup>5,27,28</sup>. However, roles of MAPK pathway, especially in tolerance to toxic compounds liberated from biomass pretreatment, remain largely

unknown. MAPK signaling pathway regulates a variety of cellular activities including cell proliferation, differentiation, survival and death. It contains sensor receptors, transduction proteins and three protein kinases acting in a series of a MAP kinase kinase kinase (MAPKKK), a MAP kinase kinase (MAPKK), followed by a MAP kinase (MAPK)<sup>29</sup>. In *S. cerevisiae*, five functionally distinct MAPK cascades were characterized and each MAPK cascade performs important functions in the regulation of diverse gene expression by phosphorylation on transcriptional factors and other target genes<sup>30</sup>. In Y-12632, non-synonymous sequence variation and differential expression of genes were found to be involved in at least three MAPK pathways: filamentous growth pathway (FG), high osmolarity glycerol pathway (HOG) and cell wall integrity pathway (CWI).

In the FG, HOG and CWI pathways, we found more than 160 non-synonymous SNPs contained in at least 41 genes, with up to 19 SNPs within a gene. Our quantitative gene expression analysis singled out 25 genes in these three MAPK pathways that displayed consistently higher levels of signature expressions against the HMF challenge compared with the untreated control (Fig. 3B). Among which, 21 genes were overlapping with those non-synonymous SNP containing genes (Fig. 3A). For example, transmembrane sensor genes *SLN1*,

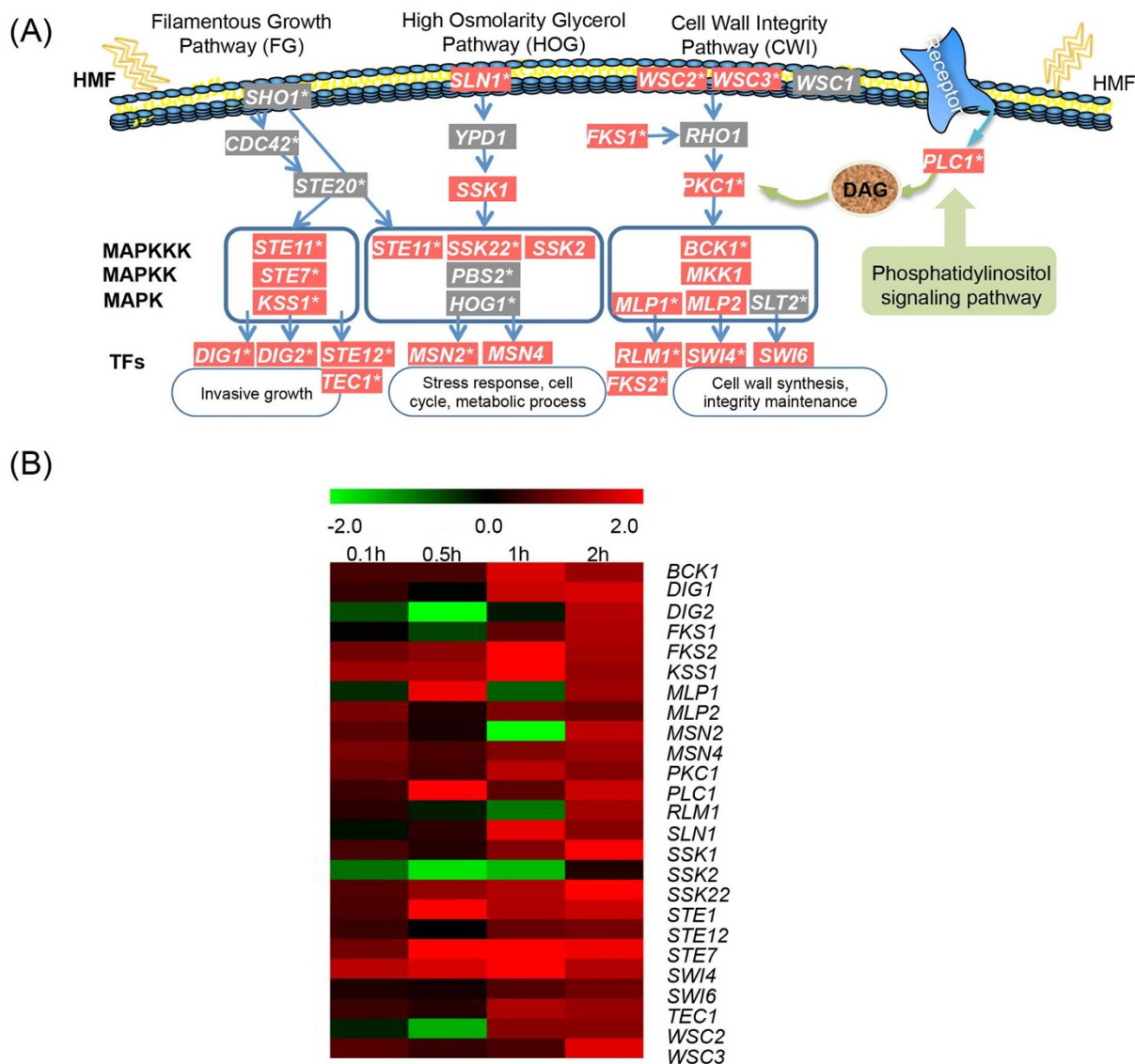


**Figure 2** | The global metabolic pathways of industrial yeast *Saccharomyces cerevisiae* type strain NRRL Y-12632. Pathways with genes showing consistent up-regulated transcription levels in response to HMF challenge are highlighted in blue, those with genes harboring non-synonymous SNPs are highlighted in green, while those genes that show consistent up-regulated transcription levels in response to HMF challenge and also harbor non-synonymous SNPs are highlighted in red.

WSC2 and WSC3 each had nine, five and three non-synonymous mutations, respectively (Table S5), which were all confirmed using independent PCR sequencing verifications. These up-regulated sensor genes in HOG and CWI pathways appeared to be interacted effectively with downstream genes such as *SSK1*, *FKS1* and *PKC1* to activate their respective MAPK cascades as direct signal transduction channels in response to HMF challenge. This includes the enhanced signature expressions of *BCK1*, *MKK1*, *MLP1*, *MLP2* and *PLC1* specifically in the CWI pathway and *SSK2*, *SSK22* and *STE11* in the HOG pathway, respectively (Fig. 3A, B). Sequence variations have been observed to affect gene performance, for example, non-synonymous SNPs of *MSN4* altered gene expression<sup>31,32</sup>. Mutation of amino acids in the extracellular domain of *SLN1* was reported to significantly affect downstream proteins such as Hog1p, causing protein dephosphorylation<sup>33</sup>. Since all of the nine non-synonymous SNPs for *SLN1* observed in this study were located in the extracellular domain, the altered amino acids are expected to impact the downstream interplay in the HOG pathway. Moreover, *SLN1* exhibited a steadily enhanced signature expression in response to HMF. The large amount of sequence variations as measured by SNPs for those up-regulated genes was significant statistically. In the FG pathway, a normally expressed membrane sensor path activated MARK cascades showing up-regulated signature expressions of *STE11*, *STE7* and *KSS1*. All of these genes involved in this pathway showed sequence mutations. It is important to point out that normally expressed genes under the inhibitory challenge conditions, such as *SHO1*, *CDC42*, and *STE20*, are important and necessary for functional cells to keep the interaction flow globally. We observed that this kind of genes in a non-tolerant wild-type strain lost functions 48 h after inhibitor challenge<sup>11</sup>. Thus, persistent expression under

the stress should be considered as a tolerant characteristic of the gene. Most of these normally expressed genes were also found to contain non-synonymous SNPs (Fig. 3A). Whether these sequence variations are involved in the fine-tuned expression response is currently unknown.

We found that seven SNPs of *MLP2* in the CWI pathway MAPK cassette for Y-12632 were particular interesting. These mutations were located at the upstream regions at -5, -46, -119, -317, -394, -1557, and -1722 positions, respectively. UTR sequences, especially 5' UTRs, play a concerted role in regulation of gene expression under stress conditions<sup>34</sup>. Among the 5'UTR SNPs observed, the mutation at the -5 locus resulted in a transcription factor binding motif of CGGNS for Stb5p. Transcription factor Stb5p is an important regulator mediating multidrug resistance and oxidative stress response, which is a key element of yeast tolerance to biomass pretreatment inhibitors. In addition, SNPs at -119 and -317 loci resulted in two duplicated *Azf1p* binding motifs of AAAAGAAA. Thus Y-12632 has a total of four repetitive binding motifs of *Azf1p* in the upstream of the *MLP2* coding region, which is twice the dose of S288C. *Azf1p* is a zinc-finger transcription factor, which activates genes involved in maintenance of cell wall integrity<sup>35,36</sup>. The doubled amount of motifs is expected to enhance protein binding activity and enable stronger expressions of the tolerance gene for a tolerant response against the toxic chemicals. We also observed that *PKC1* and *PLC1*, key-connecting nodes of MAPK-CWI pathway, harbored multiple SNPs in their pre-coding regions. However, these mutations did not generate any transcription factor binding motifs and their possible functional impact is not clear. Nonetheless, the significantly activated CWI pathway responses observed in this study supports CWI as an important component underpinning Y-12632 performance in res-



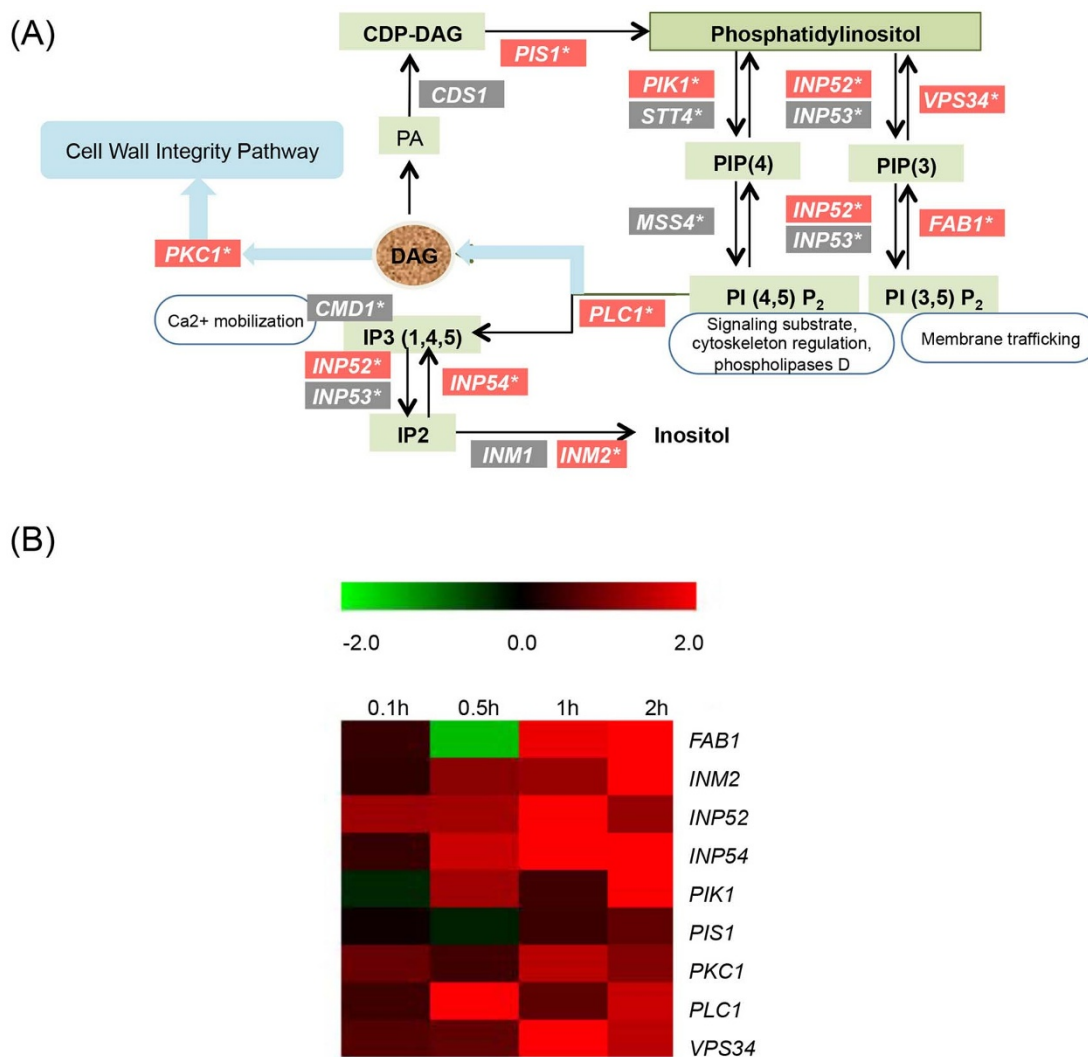
**Figure 3 | Genomic and phenotype analysis of selective MAPK-signaling pathways against HMF in *S. cerevisiae* Y-12632.** (A) Illustrative networks showing normally expressed (in gray) and enhanced expressed genes (in red). Genes with functional SNPs are indicated with a star (\*). (B) Gene expression over time during the lag phase for selective genes in response to HMF.

ponse to HMF challenge. The enriched sequence variation for many genes in this pathway and their enhanced signature expressions under HMF exposure suggest a potential role of this pathway in yeast adaptation to HMF stress.

To confirm functions involved in HMF tolerance of these candidate genes, we examined growth response of analogous single gene deletion mutations for the following selective genes: *SHO1*, *STE7*, *KSS1*, *DIG1*, *DIG2*, *TEC1* from FGP pathway, *SSK1*, *STE11*, *SSK22*, *SSK2*, *MSN2* and *MSN4* from HOG pathway, and *WSC2*, *FKS1*, *BCK1*, *MLP1*, *MLP2*, *PLC1*, *SWI4*, *SWI6*, *RLM1* and *GSC2/FKS2* from CWI pathway. On a medium without HMF, a wild type strain BY4742 and all single gene deletion mutations showed normal growth and reached a station phase approximately at 24 h after incubation (Fig. S3). When 20 mM HMF was added in the medium, the control strain was still able to grow but at a slower rate, reaching a station phase at 32 h. In contrast, all mutations were severely repressed on the HMF containing medium and no growth response was observed until 48 h, except for  $\Delta$ *WSC3* showing a recovery of growth after a 24 h lag phase. Most single gene deletion mutations showed no growth response on HMF-containing medium even until 72 h after incubation. These results indicated that each of these tested

candidate gene was essential for the yeast survival and growth in response to HMF challenge.

For many genes involved in varied downstream interactions to MAPK pathways, we also observed high levels of sequence variations accompanied with greater gene expressions in response to HMF challenge (Fig. 3A). These genes are transcription factor genes, including *RLM1*, *SWI6*, *SWI4*, *FKS2*, *MSN2*, *MSN4*, *DIG1*, *DIG2*, *STE12* and *TEC1*, that mediate a wide range of biological functional categories relevant to yeast tolerance such as stress response, cell cycle and cell wall integrity. In general, expression levels of these genes increased over time with exposure of HMF (Fig. 3B). The HOG and CWI pathways have been reported to play important roles in adapting to hyperosmotic stress and maintaining cell wall integrity under stressful conditions<sup>37</sup>. High levels of protein expression in HOG pathway were reported in response to pretreatment inhibitors<sup>38</sup>. HOG1 pathway was also reported to be involved in tolerance of furfural and acetic acid<sup>39,40</sup>. Our observations in this study concur with those previously reported. As mentioned above, the CWI pathway was especially noteworthy in response to HMF challenge and can be a key component for the yeast tolerance to HMF and other pretreatment inhibitors. Its important role for tolerance will be fur-



**Figure 4 | Genomic and phenotype analysis of phosphatidylinositol signaling system against HMF in *S. cerevisiae* Y-12632.** (A) Illustrative networks showing normally expressed (in gray) and enhanced expressed genes (in red). Genes with functional SNPs are indicated with a star (\*). (B) Gene expression over time during the lag phase for selective genes in response to HMF.

ther discussed in the following section in close relationship with phosphatidylinositol signaling pathway.

**The unique phosphatidylinositol signaling pathways.** Potential functional alterations derived from DNA sequence variations also appeared to exist in the phosphatidylinositol signaling pathways. Phosphoinositides (PIs), derived from phosphatidylinositol by phosphorylation, are regulatory lipids that function in signal transduction and mediate numerous physiological processes in eukaryotic organisms, such as growth, cytoskeletal rearrangement and membrane trafficking<sup>41</sup>. Thirteen genes in the PI signaling system of Y-12632 were found to harbor numerous non-synonymous SNPs (Fig. 4A). Among which, nine genes, including *VPS34*, *FAB1*, *INP52*, *INP54*, *PIK1*, *PLC1*, *PIS1*, *INM2*, and *PKC1*, showed consistently enhanced signature expressions in response to HMF challenge (Fig. 4B). For example, *VPS34* (encoding PI(3) kinase which is crucial for phosphatidylinositol-3-phosphate [PIP(3)] synthesis) and its downstream membrane kinase *FAB1* each displayed ten and nine non-synonymous mutations, respectively (Table S5). Both genes showed enhanced signature expression facilitating pathways of PI(3,5)P<sub>2</sub> synthesis. Accumulation of PI(3,5)P<sub>2</sub> has been observed in association with osmotic stress in yeast<sup>42</sup>. The active expression for PI(3,5)P<sub>2</sub> biosynthesis indicates a

tolerant response to the imposed inhibitor stress. In a parallel path of PIP(4) synthesis, a kinase encoded by *PIK1* catalyzed reaction was also highly activated. *PIK1* has been recognized as an essential gene for synthesis of PIP(4) used in the secretory machinery of yeast cell<sup>43</sup>. All these three genes in Y-12632 can be distinguished from S288C in sequence variations with non-synonymous mutations in their exonic regions. *INP52* and *INP54* were also observed to be up-regulated in favor of biosynthesis of IP1/IP3. These genes encode inositol polyphosphate 5-phosphatase, which pose dual functions of 5-phosphatase and polyphosphoinositide phosphatase activity that are essential in mediating metabolism of PIs. Similarly with confirmations of gene functions for MARK pathways, we evaluated selective genes for their growth response to HMF using corresponding single gene deletion mutations. The highly sensitive growth response of the mutations against HMF confirmed the tolerant functions of these candidate genes (Fig. S4).

Using independent sequencing analysis by PCR amplification, we confirmed six non-synonymous SNPs of *INP54* and detected five SNPs were located in its polyphosphate 5-phosphatase domain. One mutation resulted in an amino-acid alteration from Lys264 to Met264, which was predicted as a non-neutral substitution by SNAP<sup>44</sup>. Recently, we demonstrated that amino acid substitution from Val285 to Asp285 of Gre2p significantly increased aldehyde



reduction activity using additional cofactor NADPH in yeast<sup>45</sup>. Thus, DNA sequence variation in industrial yeast can lead to amino-acid substitutions that potentially affect gene functions and gene interactions. Interestingly, these mutated genes in Y-12632 exhibited highly activated signature expressions under the HMF stress.

It is important to point out that the PI signaling system is closely related to MAPK pathways, especially to the CWI pathway, as bridged through interactions mediated by *PLC1* and *PKC1*. It is striking that both genes contained substantial amino acid substitutions. Amino-acid changes caused by SNP in *HOG1* and *PBS2* have been observed to cause abnormal cross-talk between MAPK-HOG pathways in activation of osmolarity pheromone response<sup>46</sup>. Highly activated expressions of *PLC1* and *PKC1*, key genes linking PI signaling system and CWI pathway, are expected to reflect the impact of sequence mutations and associated amino-acid substitutions.

## Conclusions

Research and development of yeast tolerance to pretreatment inhibitors is a rapid growing field for advanced biofuels applications. Many candidate genes and regulatory elements have been identified in yeast response to inhibitory compounds such as furfural and HMF<sup>5,18,19,47</sup>. However, little is known about the signaling transduction pathways regarding yeast tolerance to biomass pretreatment inhibitors. This study provides the first insight into the unique genomic background of an industrial yeast type strain regarding to chemical stress tolerance in comparison to a model reference genome. Two major technical challenges exist for developing next-generation biocatalyst for advanced biofuels production: improved stress tolerance including inhibitor tolerance, and efficient and balanced utilization of C-5 and C-6 biomass sugars using lignocellulosic hydrolytes. Yeast strain improvement has been widely studied using lab strain and industrial yeast strains. Inconsistent phenotypic results from model lab strain and industrial yeast strains exist and it poses challenges to effectively addressing these technical difficulties based on their different genomic background. Utilizing a model lab strain has a great advantage of convenient genetic tools. However, a careful interpretation of results derived from study of lab strains is needed since it may not necessarily reflect an industrial yeast response or performance. A recently reported significant improvement of heterologous xylose transporter function in an industrial strain of *S. cerevisiae* suggested industrial yeast as a more suitable host for xylose utilization improvement<sup>9</sup>. In this study, the findings of the tolerant MAPK and PI signaling pathways in Y-12632 against biomass pretreatment inhibitors provide a strong basis for interpretation and rationale for utilizing industrial yeast for tolerant strain development. Thus an increased use of industrial yeasts as both workhorse and research model cannot be overemphasized in the development of the next-generation biocatalysts for advanced biofuel applications.

## Methods

**Yeast strains and culture conditions.** *S. cerevisiae* strain NRRL Y-12632 (Agricultural Research Service Culture Collection, Peoria, IL, USA) was maintained and cultured on a synthetic complete medium. Nonessential haploid *S. cerevisiae* deletion mutations generated by the Saccharomyces Genome Deletion Project and the parental strain BY4742 (MAT $\alpha$  his3 $\Delta$ 1 leu2 $\Delta$ 0 lys2 $\Delta$ 0 ura3 $\Delta$ 0) were obtained from Open Biosystems (Huntsville, AL). BY4742 is a designed deletion strain directly derived from *Saccharomyces cerevisiae* S288C, which is widely selected as parent strains for the international systematic *Saccharomyces cerevisiae* gene disruption project<sup>48</sup>.

Freshly grown cells harvested at logarithmic growth phase were used as inoculate after 16 h incubation with agitation of 250 rpm at 30°C. Cells were incubated on SC medium using a flask fermentation system at 30°C as previously described<sup>15</sup>.

**Genome sequencing and analysis.** Genomic DNA of Y-12632 was isolated and sequenced by Solexa GA-IIx with a pair-ended strategy. Paired reads with an average length of 75 bp and pair distance of 300 bp were collected and assembled using Velvet (v1.1.07). Contigs shorter than 500 bp were filtered and removed from the final assembly. Gene structure was predicted using Augustus (v2.5.5) with the gene models of *S. cerevisiae* S288C as a reference. Orthology and sequence identity of each predicted protein was evaluated by searching against Saccharomyces Genome

Database (SGD; <http://www.yeastgenome.org/>), NCBI-NR (<http://www.ncbi.nlm.nih.gov/>) and KEGG database (<http://www.kegg.jp/>) using BlastP (E-value < 1e-5) and the best-hit was used for the final annotation. The functional domains and GO annotations were obtained by InterProScan.

To identify unique genes in Y-12632, the predicted ORFs in Y-12632 genome were searched against all proteins of S288C by BlastP using an E-value cutoff of 1e-5 and identity (length of alignment/length of query) cutoff of 80%. To avoid omitting alignments by gene prediction error, the unaligned proteins from Y-12632 were searched against genome of S288C by tBlastN with E-value cutoff at 1e-5 and identity cutoff of 80%. Proteins in Y-12632 that failed to align to S288C genome by either BlastP or tBlastN were considered unique genes in Y-12632.

**Phylogenetic analysis.** The orthologs among the selected yeast strains were identified by a Markov Clustering algorithm (OrthoMCL v.4)<sup>49</sup> with inflation index of 1.5. The protein-coding gene sets for each of the genomes were searched against all genes in the genomes (including their own genome) by blast with E-value cutoff value 1e-5. Then the orthologous groups were generated by MCL with inflation index of 1.5, in which each of the genes is an ortholog to all other members of the same group. For each of the orthologous gene-sets, the protein sequences encoded were aligned by MUSCLE (v3.7)<sup>50</sup>. The alignments were curated by GBlock (v0.91b)<sup>51</sup> to filter out poorly aligned positions. The curated alignments were then analyzed by PhyML (v3.0)<sup>52</sup> to generate the phylogenetic trees based on maximum likelihood. A consensus tree was then constructed based on all of the single copy orthologous gene-sets.

**SNP identification and analysis.** Sequence reads were mapped to the *S. cerevisiae* S288C genome using BWA<sup>53</sup>. The variants calling was performed using SAMtools v0.1.16<sup>54</sup> and GATK v1.0<sup>55</sup> with default parameters, respectively. Then a custom python script was developed to produce an intersection of the SNPs predicted by SAMtools and GATK. The results were then subject to additional filtering to remove SNPs with low quality value and low read coverage. Functional classification and annotation of the predicted SNPs were performed using NGS-SNP tool<sup>56</sup>. Databases Ensembl (release 63), EntrezGene, and Gene Ontology were used as reference for annotation. Several fields of other information, such as descriptions of the influenced transcripts and proteins were also provided where applicable. Potential effects of single amino acid substitutions on protein function were predicted by SNAP<sup>44</sup>.

To confirm Y-12632 mutant protein sequences, queries were aligned to structural templates in Protein Data Bank ([www.rcsb.org/pdb/](http://www.rcsb.org/pdb/)) using BlastP. Transcription factor binding sites in promoter sequences were identified by searching Yeasttract ([www.yeasttract.com](http://www.yeasttract.com)) and the Promoter Database of *Saccharomyces cerevisiae* (<http://rulai.cshl.edu/SCPD>).

**Functional enrichment analysis.** All genes with SNPs were classified by their annotations of GO terms. Functional enrichment analysis was performed using WebGestalt<sup>57</sup>. Briefly, we used WebGestalt to implement the hypergeometric test for enrichment of GO terms and metabolic pathways for all candidate genes. The entire yeast S288C genome was used as a reference list. A statistic cutoff *P*-value for significantly enriched terms was set up at *P* < 0.05. The *P* value was calculated using Fisher's exact test and then adjusted using the R function *p.adjust* by WebGestalt.

**Metabolic pathway analysis.** KEGG IDs associated with each functional SNP associated gene in *Saccharomyces cerevisiae* were obtained, when applicable, by searching the protein sequence against KEGG database with E-value cutoff at 1e-5. Best hits were recorded, which allowed the mapping of genes to KEGG pathways. The pathways related to the identified genes were then displayed using the iPath<sup>58</sup>.

**Genome-wide gene expression profiling using microarray.** Genome microarray was fabricated with a version of 70-mer oligo set representing 6388 genes using OmniGrid 300 Gene Machine with embedded universal RNA reference and restricted quality control measurements<sup>58,59</sup>. Cultures of *S. cerevisiae* strain NRRL Y-12632 were treated with HMF at a final concentration of 30 mM 6 h after the inoculation. A time course study was carried on and samples taken at 0, 10, 30, 60, 120 min after the HMF treatment. Cultures grown under the same conditions without the HMF treatment served as a control. Two replicated experiments were carried out for each condition. Cells harvesting, RNA isolation, labeling and hybridization were carried out as previously described<sup>5,59,60</sup>. Microarray data were normalized using the quality control gene CAB and deposited in the Gene Expression Omnibus database under accession number GSE22939. Differentially expressed genes were identified using Significance Analysis of Microarrays (SAM)<sup>25</sup> with the time-course mode to identify genes with a consistent increase over time. Significant gene changes were arbitrarily selected at SAM (d)score values equal or greater than 0.23 with the lowest false discovery rate (FDR < 0.1). All identified regulated genes were mapped to the *S. cerevisiae* biological pathways in KEGG database and annotated with GO database.

**Growth response of single-gene-deletion mutants to HMF.** Twenty-eight selective single gene deletion mutations from Saccharomyces Genome Deletion Sets were used for examination of growth response to HMF. Since currently there is no deletion mutant library available for Y-12632 strain, gene knockout mutants of a parental wild type strain BY4742 were used as an analogue to verify the associated functions of candidate genes.

The selected genes are mainly involved in MAPK signaling pathway, phosphatidylinositol signaling system and immediate interplays, including available non-



essential genes and transcription factor genes *WSC2*, *INP54*, *MSN4*, *SWI6*, *MLP1*, *FKS1*, *DIG1*, *RLM1*, *INM2*, *WSC3*, *BCK1*, *STE7*, *SSK2*, *INP52*, *MLP2*, *KSS1*, *TEC1*, *FKS2*, *SSK22*, *FAB1*, *MSN2*, *DIG2*, *SHO1*, *SWI4*, *STE11*, *PLC1*, and *VPS34*. Each tested strain was grown on a 4 ml synthetic medium in a 15-ml tube at 30°C with agitation of 250 rpm [10]. The initial OD at 600 nm of the inoculated medium for each deletion strain culture was adjusted to the same level and inoculated onto the medium with or without a final HMF concentration of 20 mM. Cell growth was monitored by absorbance at OD<sub>600</sub> (n=3).

- Vertès, A. Q. N., Yukawa, H. & Blaschek, H. *Biomass to Biofuels: Strategies for Global Industries*. (Wiley, 2010).
- Wall, J. D. H. C. & Demain, A. *Bioenergy*. (ASM Press, 2008).
- Gasch, A. P. & Werner-Washburne, M. The genomics of yeast responses to environmental stress and starvation. *Funct Integr Genomics* **2**, 181–192 (2002).
- Gasch, A. P. *et al.* Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**, 4241–4257 (2000).
- Ma, M. & Liu, Z. L. Comparative transcriptome profiling analyses during the lag phase uncover YAP1, PDR1, PDR3, RPN4, and HSF1 as key regulatory genes in genomic adaptation to the lignocellulose derived inhibitor HMF for *Saccharomyces cerevisiae*. *BMC Genomics* **11**, 660 (2010).
- Liu, Z. Genomic adaptation of ethanogenic yeast to biomass conversion inhibitors. *Appl Microbiol Biotechnol* **73**, 27–36 (2006).
- Runquist, D., Hahn-Hagerdal, B. & Radstrom, P. Comparison of heterologous xylose transporters in recombinant *Saccharomyces cerevisiae*. *Biotechnol Biofuels* **3**, 5 (2010).
- Young, E., Poucher, A., Comer, A., Bailey, A. & Alper, H. Functional Survey for Heterologous Sugar Transport Proteins, Using *Saccharomyces cerevisiae* as a Host. *Appl Environ Microbiol* **77**, 3311–3319 (2011).
- Moon, J., Lewis Liu, Z., Ma, M. & Slininger, P. J. New genotypes of industrial yeast *Saccharomyces cerevisiae* engineered with YXI and heterologous xylose transporters improve xylose utilization and ethanol production. *Biocatal Agric Biotechnol* **2**, 247–254 (2013).
- Lewis, J. A., Elkon, I. M., McGee, M. A., Higbee, A. J. & Gasch, A. P. Exploiting Natural Variation in *Saccharomyces cerevisiae* to Identify Genes for Increased Ethanol Resistance. *Genetics* **186**, 1197–1205 (2010).
- Liu, Z. L., Ma, M. & Song, M. Evolutionarily engineered ethanogenic yeast detoxifies lignocellulosic biomass conversion inhibitors by reprogrammed pathways. *Mol Genet Genomics* **282**, 233–244 (2009).
- Shima, J. & Takagi, H. Stress-tolerance of baker's-yeast (*Saccharomyces cerevisiae*) cells: stress-protective molecules and genes involved in stress tolerance. *Biotechnol Appl Biochem* **53**, 155–164 (2009).
- Argueso, J. L. *et al.* Genome structure of a *Saccharomyces cerevisiae* strain widely used in bioethanol production. *Genome Res* **19**, 2258–2270 (2009).
- Stambuk, B. U., Dunn, B., Alves Jr, S. L., Duval, E. H. & Sherlock, G. Industrial fuel ethanol yeasts contain adaptive copy number changes in genes involved in vitamin B1 and B6 biosynthesis. *Genome Res* **19**, 2271–2278 (2009).
- Liu, Z. L. *et al.* Adaptive response of yeasts to furfural and 5-hydroxymethylfurfural and new chemical evidence for HMF conversion to 2,5-bis-hydroxymethylfuran. *J Ind Microbiol Biotechnol* **31**, 345–352 (2004).
- Mannazzu, I., Clementi, F. & Ciani, M. Strategies and criteria for the isolation and selection of autochthonous starters. *Biodivers Biotechnol Wine Yeasts*, 19–35 (2002).
- Bradbury, J. E. *et al.* A homozygous diploid subset of commercial wine yeast strains. *Antonie van Leeuwenhoek* **89**, 27–37 (2006).
- Liu, Z. L. & Moon, J. A novel NADPH-dependent aldehyde reductase gene from *Saccharomyces cerevisiae* NRRL Y-12632 involved in the detoxification of aldehyde inhibitors derived from lignocellulosic biomass conversion. *Gene* **446**, 1–10 (2009).
- Sehnm, N. T. *et al.* 5-Hydroxymethylfurfural induces ADH7 and ARI1 expression in tolerant industrial *Saccharomyces cerevisiae* strain P6H9 during bioethanol production. *Bioresour Technol* **133**, 190–196 (2013).
- Zerbin, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829 (2008).
- Quevillon, E. *et al.* InterProScan: protein domains identifier. *Nucleic Acids Res* **33**, W116–W120 (2005).
- Blasco, L., Veiga-Crespo, P. & Villa, T. G. FPG1, a gene involved in foam formation in *Saccharomyces cerevisiae*. *Yeast* **28**, 437–451 (2011).
- Borneman, A. R. *et al.* Whole-genome comparison reveals novel genetic elements that characterize the genome of industrial strains of *Saccharomyces cerevisiae*. *PLoS Genet* **7**, e1001287 (2011).
- Schacherer, J., Shapiro, J. A., Ruderfer, D. M. & Kruglyak, L. Comprehensive polymorphism survey elucidates population structure of *Saccharomyces cerevisiae*. *Nature* **458**, 342–345 (2009).
- Tusher, V. G., Tibshirani, R. & Chu, G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci* **98**, 5116 (2001).
- Stanley, D., Chambers, P. J., Stanley, G. A., Borneman, A. & Fraser, S. Transcriptional changes associated with ethanol tolerance in *Saccharomyces cerevisiae*. *Appl Microbiol Biotechnol* **88**, 231–239 (2010).
- Lin, F. M., Qiao, B. & Yuan, Y. J. Comparative proteomic analysis of tolerance and adaptation of ethanogenic *Saccharomyces cerevisiae* to furfural, a lignocellulosic inhibitory compound. *Appl Environ Microbiol* **75**, 3765–3776 (2009).
- Liu, Z. L. in *Microbial Stress Tolerance for Biofuels* Vol. 22 *Microbiology Monographs* (ed Zonglin Lewis Liu) Ch. 12, 279–299 (Springer Berlin Heidelberg, 2012).
- Herskowitz, I. MAP kinase pathways in yeast: For mating and more. *Cell* **80**, 187–197 (1995).
- Chen, R. E. & Thorner, J. Function and regulation in MAPK signaling pathways: Lessons learned from the yeast *Saccharomyces cerevisiae*. *Biochim Biophys Acta-Mol Cell Res* **1773**, 1311–1340 (2007).
- Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* **35**, 57–64 (2003).
- Sadeh, A., Baran, D., Volokh, M. & Aharoni, A. Conserved motifs in the Msn2-activating domain are important for Msn2-mediated yeast stress response. *J Cell Sci* **125**, 3333–3342 (2012).
- Ostrander, D. B. & Gorman, J. A. The Extracellular Domain of the *Saccharomyces cerevisiae* Sln1p Membrane Osmolarity Sensor Is Necessary for Kinase Activity. *J Bacteriol* **181**, 2527–2534 (1999).
- Lawless, C. *et al.* Upstream sequence elements direct post-transcriptional regulation of gene expression under stress conditions in yeast. *BMC Genomics* **10**, 7 (2009).
- Slattery, M. G., Liko, D. & Heideman, W. The Function and Properties of the Azf1 Transcriptional Regulator Change with Growth Conditions in *Saccharomyces cerevisiae*. *Eukaryot Cell* **5**, 313–320 (2006).
- Mirzaei, H. *et al.* Systematic measurement of transcription factor-DNA interactions by targeted mass spectrometry identifies candidate gene regulatory proteins. *Proc Natl Acad Sci* **110**, 3645–3650 (2013).
- Posas, F., Takekawa, M. & Saito, H. Signal transduction by MAP kinase cascades in budding yeast. *Curr Opin Biotechnol* **1**, 175–182 (1998).
- Lin, F. M., Tan, Y. & Yuan, Y. J. Temporal quantitative proteomics of *Saccharomyces cerevisiae* in response to a nonlethal concentration of furfural. *Proteomics* **9**, 5471–5483 (2009).
- Mollapour, M. & Piper, P. Hog1p mitogen-activated protein kinase determines acetic acid resistance in *Saccharomyces cerevisiae*. *FEBS Yeast Res* **6**, 1274–1280 (2006).
- Park, J., Kim, J., Park, S. & Kim, D. Characterization of CpSte11, a MAPKKK gene of *Cryphonectria parasitica*, and initial evidence of its involvement in the pheromone response pathway. *Mol Plant Pathol* **13**, 240–250 (2012).
- Wera, S., Bergsma, J. C. T. & Thevelein, J. M. Phosphoinositides in yeast: genetically tractable signalling. *FEMS Yeast Res* **1**, 9–13 (2001).
- Dove, S. K. *et al.* Osmotic stress activates phosphatidylinositol-3,5-bisphosphate synthesis. *Nature* **390**, 187–192 (1997).
- Walch-Solimena, C. & Novick, P. The yeast phosphatidylinositol-4-OH kinase Pik1 regulates secretion at the Golgi. *Nat Cell Biol* **1**, 523–525 (1999).
- Johnson, A. D. *et al.* SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics* **24**, 2938–2939 (2008).
- Moon, J. & Liu, Z. L. Engineered NADH-dependent GRE2 from *Saccharomyces cerevisiae* by directed enzyme evolution enhances HMF reduction using additional cofactor NADPH. *Enzyme Microb Technol* **50**, 115–120 (2012).
- O'Rourke, S. M. & Herskowitz, I. The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in *Saccharomyces cerevisiae*. *Genes Dev* **12**, 2874 (1998).
- Bowman, M. J. *et al.* Stereochemistry of furfural reduction by a *Saccharomyces cerevisiae* aldehyde reductase that contributes to in situ furfural detoxification. *Appl Environ Microbiol* **76**, 4926–4932 (2010).
- Baker Brachmann, C. *et al.* Designer deletion strains derived from *Saccharomyces cerevisiae* S288C: A useful set of strains and plasmids for PCR-mediated gene disruption and other application. *Yeast* **14**, 115–132 (1998).
- Li, L., Stoeckert, C. J. & Roos, D. S. OrthoMCL: Identification of Ortholog Groups for Eukaryotic Genomes. *Genome Res* **13**, 2178–2189 (2003).
- Edgar, R. C. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* **32**, 1792–1797 (2004).
- Talavera, G. & Castresana, J. Improvement of Phylogenies after Removing Divergent and Ambiguously Aligned Blocks from Protein Sequence Alignments. *Syst Biol* **56**, 564–577 (2007).
- Guindon, S. *et al.* New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. *Syst Biol* **59**, 307–321 (2010).
- Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
- Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
- DePristo, M. A. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* **43**, 491–498 (2011).
- Grant, J. R., Arantes, A. S., Liao, X. & Stothard, P. In-depth annotation of SNPs arising from resequencing projects using NGS-SNP. *Bioinformatics* **27**, 2300–2301 (2011).
- Duncan, D., Prodduturi, N. & Zhang, B. WebGestalt2: an updated and expanded version of the Web-based Gene Set Analysis Toolkit. *BMC Bioinformatics* **11**, 1–1 (2010).
- Letunic, I., Yamada, T., Kanehisa, M. & Bork, P. iPath: interactive exploration of biochemical pathways and networks. *Trends Biochem Sci* **33**, 101–103 (2008).





59. Liu, Z. L. & Slininger, P. J. Universal external RNA controls for microbial gene expression analysis using microarray and qRT-PCR. *J Microbiol Methods* **68**, 486–496 (2007).
60. Hegde, P. *et al.* A concise guide to cDNA microarray analysis. *BioTechniques* **29**, 548–550, 552–544, 556 passim (2000).

## Acknowledgments

This work was supported by Young Investigator Program (31401076; 61103167; 6130316) from National Natural Science Foundation of China, National Basic Research Program from Ministry of Science and Technology of China (2011CB200902; 2012CB721101), International Research Collaboration Program (31010103907), International Innovation Partnership Program from Chinese Academy of Sciences and Ministry of Science and Technology's high-tech (863) grant (grant number 2012AA02A707 and 2014AA021502); and NIFA National Research Initiative Competitive Award project 2006-35504-17359 to ZLL. The Whole Genome Shotgun projects were deposited at NCBI: PRJNA201005 for NRRL Y-12632.

## Author contributions

J.X. and Z.L.L. conceived the study. Q.Z. and K.N. performed genomic and transcriptomic data analyses. Z.L.L. fabricated yeast DNA 70-mer oligo microarray, conducted microarray

experiments, performed deletion mutation assay, verified selective S.N.P. and analyzed data. A.W. wrote some scripts for data processing. X.Z. contributed to genome sequencing. Q.Z., Z.L.L. and J.X. wrote the paper. All authors read and approved the final manuscript.

## Additional information

**Supplementary information** accompanies this paper at <http://www.nature.com/scientificreports>

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Zhou, Q. *et al.* Genomic and transcriptome analyses reveal that MAPK- and phosphatidylinositol-signaling pathways mediate tolerance to 5-hydroxymethyl-2-furaldehyde for industrial yeast *Saccharomyces cerevisiae*. *Sci. Rep.* **4**, 6556; DOI:10.1038/srep06556 (2014).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder in order to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/4.0/>