nature genetics

Genetic basis of individual differences in the response to small-molecule drugs in yeast

Ethan O Perlstein[1,2], Douglas M Ruderfer[3], David C Roberts[4], Stuart L Schreiber[1,2,5] & Leonid Kruglyak[3]

Individual response to small-molecule drugs is variable; a drug that provides a cure for some may confer no therapeutic benefit or trigger an adverse reaction in others. To begin to understand such differences systematically, we treated 104 genotyped segregants from a cross between two yeast strains with a collection of 100 diverse small molecules. We used linkage analysis to identify 124 distinct linkages between genetic markers and response to 83 compounds. The linked markers clustered at eight genomic locations, or quantitative-trait locus 'hotspots', that contain one or more polymorphisms that affect response to multiple small molecules. We also experimentally verified that a deficiency in leucine biosynthesis caused by a deletion of *LEU2* underlies sensitivity to niguldipine, which is structurally related to therapeutic calcium channel blockers, and that a natural coding-region polymorphism in the inorganic phosphate transporter *PHO84* underlies sensitivity to two polychlorinated phenols that uncouple oxidative phosphorylation. Our results provide a step toward a systematic understanding of small-molecule drug action in genetically distinct individuals.

Most therapeutic drugs are small molecules. The genetic basis of small-molecule activity in living cells can be complex and is often evolutionarily conserved across different taxa. In humans, many pharmacogenomic studies have been performed to assess the role of natural genetic variation in the cellular response to small-molecule drugs[1–8]. However, these studies are limited by small sample sizes and the inability to rapidly screen large numbers of drugs and phenotypes. *Saccharomyces cerevisiae* is a well studied unicellular yeast with extensive genetic and physiological homology to multicellular eukaryotes[9]. For this reason, budding yeast has been an attractive model for the study of many human diseases[10], including metabolic disorders[11] and neurodegeneration[12]. Yeast has also been effectively exploited in large-scale studies by consortia (for example, artificial gene deletion collections) to systematically assess individual gene function in response to small-molecule perturbation[13]. However, those studies have focused primarily on identifying the molecular targets of small molecules rather than on identifying the genes up- and downstream of these targets that modulate the physiological response (that is, resistance or sensitivity) to small molecules. A growing, complementary trend in yeast biology that is generalizable to other laboratory model organisms involves examining the effects of natural genetic variation, as opposed to engineered mutations, on the cellular phenotypes associated with complex traits such as high-temperature growth[14,15], sporulation[16,17] and genome-wide mRNA expression[18,19]. Expression traits have also been studied in mammalian cells[20,21]. Here, we greatly expand upon a previous study[22] of compound response traits in yeast by analyzing 104 genotyped segregants of *Saccharomyces cerevisiae* treated with 100 diverse compounds that we term small-molecule perturbagens (SMPs). Eighteen SMPs are Food and Drug Administration (FDA)-approved drugs that modulate evolutionarily conserved targets or cellular processes. The 104 segregants were derived from a cross between a laboratory strain (BY4716, hereafter 'BY') and a vineyard isolate (RM11-1a, hereafter 'RM') and were the focus of an extensive linkage study of steady-state mRNA expression traits[23]. The parent strains differ in DNA sequence at 0.6% of nucleotides[24].

RESULTS

Related compounds cluster based on cellular response

We first performed dose-response experiments on the two parent strains, BY and RM, in order to determine the half-maximal inhibitory concentration (IC_{50}) for each SMP (data not shown). We then measured in parallel the growth of all 104 segregants and both parent strains in the presence of one or more concentrations near the IC_{50} of each SMP at multiple time points post-inoculation (for complete list of SMPs, see **Supplementary Table 1** online). In this overlapping fashion, we better resolved the variance in segregant final yield. The median growth-inhibitory SMP concentration tested was 21.6 μM. We phenotyped seven SMPs—cetylpyridinium chloride (766 nM), cycloheximide (356 nM), clotrimazole (483 nM), flutrimazole (723 nM),

[1]Howard Hughes Medical Institute, Broad Institute of Harvard and MIT, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. [2]Department of Molecular and Cellular Biology, Harvard University, 7 Divinity Avenue, Cambridge, Massachusetts 02138, USA. [3]Lewis-Sigler Institute for Integrative Genomics and Department of Ecology and Evolutionary Biology, Princeton University, Princeton, New Jersey 08544, USA. [4]Theoretical Division and Center for Nonlinear Studies, Los Alamos National Laboratory, Los Alamos, New Mexico 87544, USA. [5]Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, Massachusetts 02138, USA. Correspondence should be addressed to S.L.S. (stuart_schreiber@harvard.edu) or L.K. (leonid@genomics.princeton.edu).

Received 6 November 2006; accepted 2 January 2007; published online 4 March 2007; doi:10.1038/ng1991

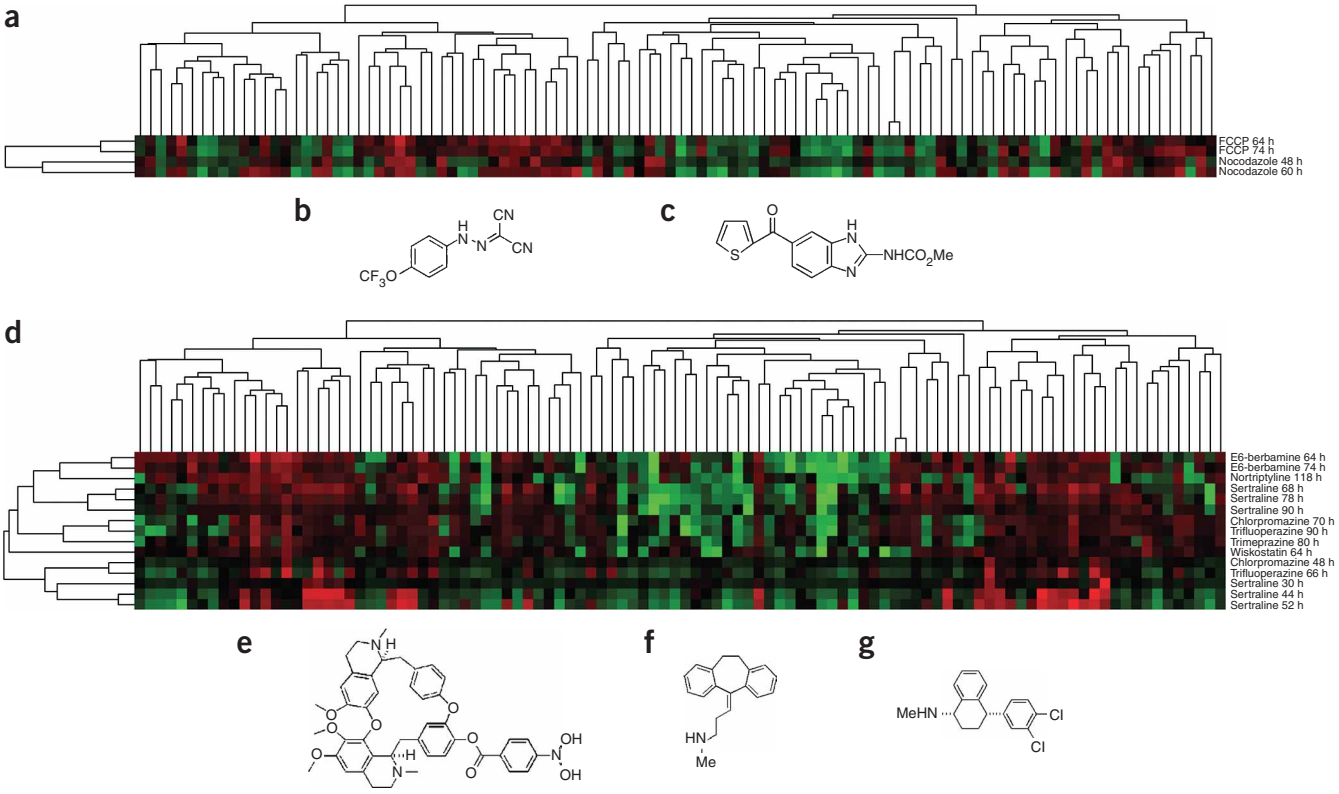

Figure 1 Hierarchical clustering of final yield measurements of 104 BY and RM segregants treated with selected SMPs shows clustering of functional analogs, including a cluster of psychiatric disease drugs. Red indicates resistance; green indicates sensitivity. Columns represent segregants; rows represent responses to SMPs. SMP names and time points are listed on the right-hand side of the clustergram. (**a**) Clustergram containing nocodazole and FCCP. (**b**) Structure of nocodazole. (**c**) Structure of FCCP. (**d**) Clustergram containing psychoactive therapeutic drugs. (**e**) Structure of the natural product E6-berbamine. (**f**) Structure of the tricyclic antidepressant nortriptyline. (**g**) Structure of the selective serotonin reuptake inhibitor sertraline.

ketoconazole (940 nM), rapamycin (50 nM) and tunicamycin (592 nM)—at submicromolar concentrations. In order to investigate the underlying genetic basis of compound response in the 104 segregants, we first performed two-dimensional hierarchical clustering on all 324 phenotypes (100 compounds at multiple time points and concentrations) after regressing out the effect of growth in SMP-free medium (for complete clustergram and raw data, see **Supplementary Fig. 1** and **Supplementary Table 2** online, respectively).

As expected, SMPs having common physiological effects on cells yet lacking structural similarity (so-called functional analogs) clustered together. For example, one cluster contained cycloheximide, anisomycin and rapamycin. These three SMPs are structurally unrelated, and they target different proteins, but their common physiological effect on cells is inhibition of protein translation. Another cluster contained FCCP, a proton ionophore, and nocodazole, a microtubule depolymerizer (**Fig. 1a–c**). A previously published report states that chronic exposure of mammalian cells to FCCP depolymerizes microtubules in a mitochondrial-dependent fashion through its destabilizing effects on the mitochondrial proton gradient[25]. Notably, another cluster contained FDA-approved therapeutic drugs of three distinct pharmacological classes, as well as the structurally unrelated natural product E6 berbamine, which is found in Chinese herbal remedies (**Fig. 1d–g**). The three pharmacological classes are the phenothiazines (for example, chlorpromazine), which are clinical antipsychotics, the selective serotonin reuptake inhibitors (for example, sertraline) and the tricyclics (for example, nortriptyline); the latter two classes are clinical antidepressants. Although the precise mechanism of action of antidepressant cytotoxicity in yeast is unknown, the clustering results suggest that there are evolutionarily conserved pathways associated with single-cell physiology that are relevant to multicellular human psychiatric disease.

Linkage analysis identifies loci that affect response to SMPs

We carried out linkage analysis between segregant final yield at each SMP concentration and time point (324 total phenotypes) and 2,956 genetic markers previously genotyped in the segregants[23]. We identified 219 QTLs with a logarithm of the odds (lod) score ≥ 4 (**Supplementary Table 3** online). Only 3.8 loci are expected by chance at this significance threshold based on empirical permutation tests; thus, the detected loci had a false discovery rate of $<2\%$. Because all time points and concentrations of each compound were used as independent phenotypes, some of the 219 QTLs were duplicate linkages of the same SMP to the same locus. We collapsed all such duplicate loci for a given SMP with overlapping confidence intervals (1 lod drop) to obtain a set of 124 unique SMP-locus pairs (**Supplementary Tables 4** and **5** online). Eighty-three of the tested SMPs linked to at least one locus, 25 SMPs linked to two loci each and eight SMPs linked to three loci each. At lower lod score thresholds, we observed many more additional loci at rates much higher than chance, providing further evidence that response to most SMPs is a genetically complex trait. For 17 of the 33 compounds with multiple detected QTLs, we observed both BY- and RM-derived genotypes that

increased resistance. It should be noted that differential parental sensitivity to a SMP is not a prerequisite for linkage of response to that SMP in the segregants; this indicates that response to compounds often demonstrates transgressive segregation, a phenomenon previously observed in the case of transcript abundance in this cross[23].

Several QTLs contain candidate genes known to affect the linked SMP response. For example, response to the mating pheromone alpha factor (αF) links with a lod score of 53 to the *MATALPHA* locus, with the BY genotype conferring resistance. This is expected, given that BY is of the MATα mating type, which is completely insensitive to the growth-arresting effects of αF. Response to copper sulfate ($CuSO_4$) links, with a lod score of 16, to a QTL that contains the copper-binding metallothionein *CUP1*, whose copy number is known to vary between yeast strains[26]. Notably, the BY parent strain, which carries the resistance genotype, possesses two copies of *CUP1* (*YHR053C* and *YHR055C*), whereas the RM parent strain, which is sensitive to $CuSO_4$, possesses four copies, and *CUP1* expression levels are higher in RM, suggesting a more complex mechanism of resistance than stoichiometric quenching of excess copper ions by *CUP1*. The sequences of different copies of *CUP1* in BY and RM are similar but not identical, and further experiments are needed to clarify the role of *CUP1* copy number and sequence polymorphism in response to $CuSO_4$.

The molecular targets of several SMPs in the data set are known: rapamycin targets the TOR proteins (Tor1p and Tor2p); latrunculin B targets the cytoskeletal protein actin (Act1p); tunicamycin targets the glycosylation protein Alg7p and ketoconazole targets the ergosterol-biosynthesis protein Erg11p. However, the amino acid coding sequences of the genes encoding these five proteins are identical in the two parent strains, and none of these five genes is located near QTLs linking to response to the SMPs that target their respective protein products. Therefore, response to small molecules is frequently affected by polymorphisms in genes other than the molecular targets themselves.

Linkage hotspots influence response to multiple SMPs

To detect loci that affect response to multiple compounds, we divided the genome into 611 bins of 20 kb each and counted the number of compounds with linkage to markers in each bin. The average number of linkages in a bin was 0.2, and none of the bins was expected to contain linkages to more than three compounds by chance (Poisson $P < 0.05$ after Bonferroni correction for 611 bins). In fact, we observed eight bins with linkages to between 4 and 25 compounds (**Fig. 2**). When we performed the analysis with equal bins on the genetic map, none of the hotspots changed; the hotspots were also not higher than average in gene density (see **Methods**). We designated these bins as linkage or QTL hotspots (for a complete list of linking SMPs, see **Supplementary Table 6** online). Each hotspot must contain either a single pleiotropic polymorphism or several closely linked polymorphisms affecting the response to multiple compounds. At some hotspot loci, the genotype of the interval from one parent consistently conferred higher sensitivity to all linking compounds, whereas at others, a given genotype conferred higher sensitivity to some linking compounds and higher resistance to others (**Fig. 2**).

Seven of the eight QTL hotspots corresponded to loci that were previously shown to affect the abundance of multiple transcripts in this cross[23], with the exception of the hotspot on chromosome I. This overlap is highly unlikely by chance (binomial $P < 10^{-10}$) and suggests (but does not prove) that the same polymorphisms affect both expression and compound response. Hotspots on chromosomes III and XIII are discussed in greater detail below. Hotspots on

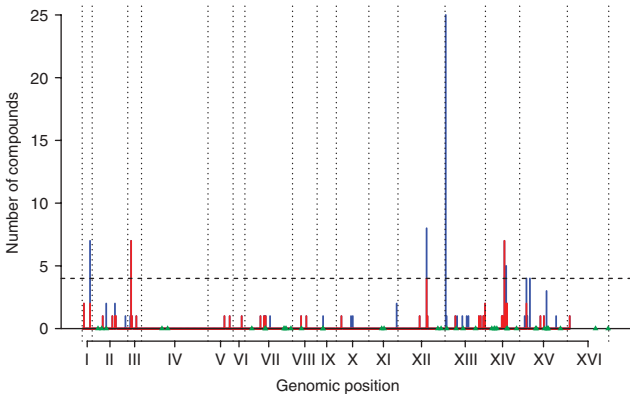

Figure 2 QTL hotspots. The number of linking compounds is plotted against genome location. Chromosomes (1–16) are arrayed end to end along the x-axis in ascending order from left to right. Tick marks denote midpoints (centromeres); dotted lines denote telomeres. Bar height indicates the number of compounds linking to the corresponding 20-kb bin; red segment of the bar indicates the number of compounds for which the BY genotype increases resistance; blue segment of the bar indicates the number of compounds for which the RM genotype increases resistance. Green triangles mark the locations of genes in the GO category 'response to drugs.' Horizontal dashed line indicates the minimum number of compounds (four) not expected to link by chance to any one bin.

chromosomes XIV and XV (two closely linked hotspots in each case) have been linked previously to expression levels of groups of genes with functions in the mitochondria and the stress response, respectively. The hotspot on chromosome XII has been shown previously to link to transcript abundance of a number of genes involved in sterol metabolism[18]. It links to response to eight compounds, including two structurally similar SMPs, cetylpyridinium chloride and benzalkonium chloride, which both contain charged heads and aliphatic tails and may affect membrane homeostasis. This hotspot contains the candidate gene *HAP1* (YLR256W), which encodes a zinc-finger transcription factor whose expression varies according to oxygen levels in the cell. The BY allele of *HAP1* contains a Ty1-insertion element at its 3′ end that lengthens the translated protein product by 13 amino acids, whereas the RM allele is wild-type[27]. Comparative transcriptional profiling of a strain with a wild-type *HAP1* allele and a strain with a *HAP1* allele with a Ty1 insertion has shown that sterol biosynthesis genes are downregulated in the strain with the Ty1 insertion allele[27]. The 20 genes most correlated with response to cetylpyridinium chloride and benzalkonium chloride are both highly enriched for the Gene Ontology (GO) cellular process 'sterol metabolism' ($P < 10^{-11}$ and $P < 10^{-7}$, respectively). Notably, the BY genotype confers greater resistance to four of the eight linking compounds but greater sensitivity to the other four; further experiments are needed to determine whether these effects are due to the *HAP1* polymorphism.

We examined the relationship between QTL hotspots and the genes that potentiate the pleiotropic drug response (PDR), the mammalian cognate of which, multidrug resistance (MDR), is the most commonly described determinant of response to drugs[28]. Only two hotspots mapped close to PDR genes. The gene *ATR1*, which encodes a multidrug efflux pump, is located near the major hotspot on chromosome XIII, but it contains no coding polymorphisms between the parent strains and falls outside the 95% linkage confidence region for the most strongly linked SMPs; we show below that a polymorphism in another gene in the region is responsible for differences in response to these SMPs. A candidate gene located near the second of a

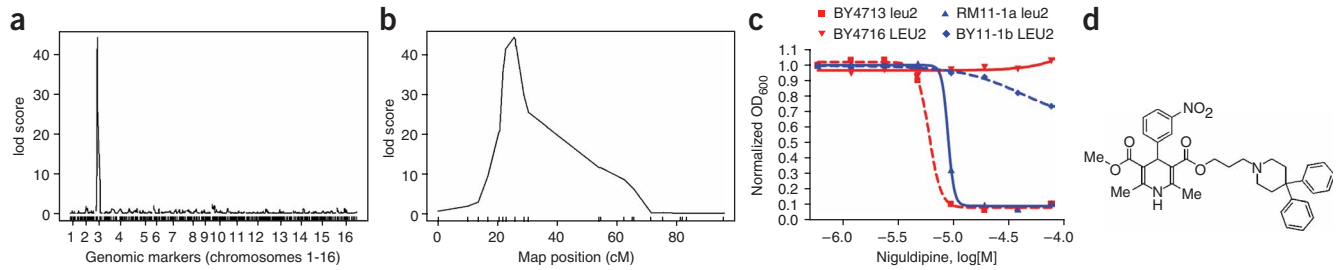

Figure 3 Functional verification that the *LEU2* locus determines response to the dihydropyridine niguldipine at chromosome III hotspot. (**a**) Genome-wide linkage plot showing a peak marker with lod score 44 on chromosome III at base pair 82466 that is linked to niguldipine. Chromosomes are arrayed end to end from left to right, and tick marks designate chromosomal centromeres. (**b**) Linkage plot of chromosome III interval. Tick marks denote individual markers. (**c**) Normalized, averaged OD_{600} absorbance measurements plotted against twofold dilution series of niguldipine. Inverted red triangles represent BY4716 (LEU2 *lys2* URA3), blue diamonds represent RM11-1b (LEU2 *lys2 ura3*), blue triangles represent RM11-1a (*leu2* LYS2 *ura3*) and red squares represent BY4713 (*leu2* LYS2 URA3). (**d**) Chemical structure of the dihydropyridine niguldipine.

pair of hotspots on chromosome XIV is *AQR1*, which encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine[29] and differs between the parent strains by an addition of 25 amino acids at the C-terminal end in BY. The BY genotype confers greater sensitivity to two SMPs and greater resistance to three SMPs. However, the SMPs that link there do not share obvious structural motifs with the compounds known to be effluxed by *AQR1*, suggesting that another gene in the interval encodes the resistance-conferring polymorphism. There are 39 genes in the yeast genome that belong to the GO class 'response to drugs'. Thirteen of these fall within 10 kb of the linkage peaks for compound response, an enrichment of more than threefold that was not observed in 10,000 random gene sets. We observed an even more marked enrichment for the GO class 'drug transport', with 8 of 13 genes located near linkage peaks. However, only the two genes described above fell near QTL hotspots (**Fig. 2**). We conclude that polymorphisms in genes known to affect drug response have some role in the differences in sensitivity segregating in this cross but that they are not responsible for most of the QTL hotspots.

A defect in leucine biosynthesis influences response to SMPs

A QTL hotspot on chromosome III affects response to seven compounds, including the dihydropyridine niguldipine, a reported L-type calcium channel antagonist. Response to niguldipine links to a marker adjacent to the gene *LEU2* (YCL018W) with a lod score of 44, the second highest observed in the entire linkage analysis (**Fig. 3**). The most significantly enriched GO process for the 20 genes whose expression is most highly correlated with response to niguldipine is 'branched chain family amino acid biosynthesis' ($P < 10^{-11}$). The *LEU2* gene is intact in the BY strain but was artificially deleted in the RM strain in order to enable crossing in the laboratory. RM is sensitive to niguldipine, whereas BY is completely resistant up to 100 μM (**Fig. 3c**). We hypothesized that niguldipine inhibits a leucine permease, as its antiproliferative effects were exacerbated in synthetic medium as compared with rich medium. We experimentally confirmed that deletion of *LEU2* confers sensitivity to niguldipine in the BY background and that the presence of *LEU2* restores resistance in the RM background (**Fig. 3c**). Moreover, a chemical complementation experiment showed that increasing the concentration of leucine, but not lysine or calcium, suppressed the antiproliferative effects of niguldipine in both BY *leu2Δ* and RM *leu2Δ* strains (data not shown). The QTL hotspot on chromosome III also linked to response to the translation-regulation inhibitor rapamycin, suggesting the existence of feedback mechanisms tying leucine availability and translational throughput. This hypothesis is supported by studies that show that exogenous leucine stimulates translation[30].

A missense polymorphism in *PHO84* underlies QTL hotspot

The major QTL hotspot on chromosome XIII affects response to 25 structurally and mechanistically diverse SMPs. This represents 25% of all tested SMPs and 30% of those for which we found significant linkage. The RM parent genotype conferred resistance to all these SMPs. We observed the strongest linkage for response to two polychlorinated aromatics, pentachlorophenol[31], an uncoupler of oxidative phosphorylation, and tetrachloroisophthalonitrile[32], an oxidative stressor, which have lod scores of 17 and 26, respectively (**Fig. 4**). Response to the proton ionophore FCCP, another uncoupler of oxidative phosphorylation, also linked to this hotspot. The RM parental strain was completely resistant to tetrachloroisophthalonitrile up to 10 μM and was approximately three times as resistant to pentachlorophenol than the BY parental strain (**Fig. 4**). The gene whose expression was most highly correlated with resistance to tetrachloroisophthalonitrile at 60 h and 84 h post-inoculation was *PHO84*, which is located in this QTL hotspot. The expression level of *PHO84* was also correlated to resistance to pentachlorophenol. Although we used this expression pattern to select *PHO84* as a candidate for further testing, it should be noted that, in general, searching for differential expression of genes in the linked regions will identify only a subset of causative genes. *PHO84* encodes a high-affinity inorganic phosphate transporter in the plasma membrane whose expression is induced by phosphate starvation[33,34].

The amino acid sequences of Pho84p in BY and RM differ by two substitutions, only one of which is nonconservative: the RM allele of *PHO84* encodes proline at amino acid position 259, whereas the BY allele encodes leucine. Alignment of the *PHO84* amino acid sequence with its orthologs in other fungal species showed uniform conservation of Pro259. In fact, the region containing residue 259 of Pho84p is orthologous to human xenobiotic transporters belonging to the solute carrier transporter (SLC) superfamily (for example, SLC22A2, which also encodes proline at this residue; data not shown). Moreover, the proline-to-leucine substitution at position 259 (L259P) occurs in the middle of one of nine predicted alpha-helical transmembrane domains, and proline-induced kinks in transmembrane spans have been shown to be essential to protein function[35]. We engineered a BY strain carrying Pro259 and observed that this single substitution restored complete resistance to tetrachloroisophthalonitrile and partial

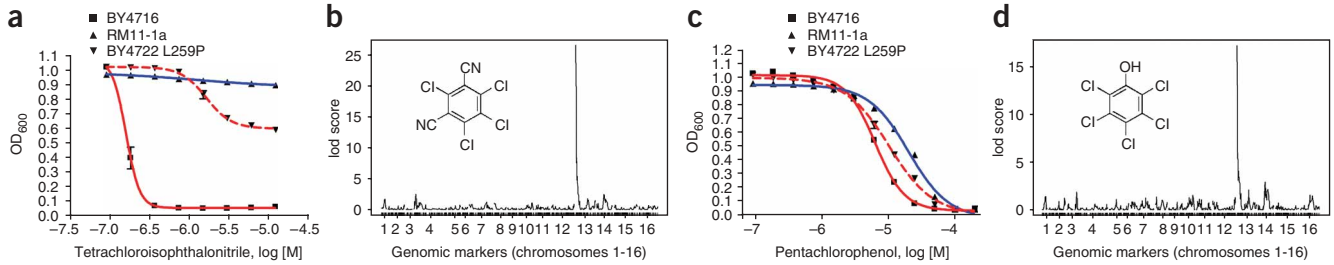

Figure 4 A missense mutation (leading to L259P) in *PHO84* confers resistance to two polychlorinated phenols. (**a**) Averaged OD_{600} absorbance growth measurements of BY4716, BY4722 L259P and RM11-1a plotted against twofold dilution series of tetrachloroisophthalonitrile. (**b**) Genome-wide linkage plot uncovers a peak marker with lod score of 26 on chromosome XIII at base pair 27644 that is linked to tetrachloroisophthalonitrile (the chemical structure in the inset). (**c**) Averaged OD_{600} absorbance growth measurements of BY4716, BY4722 L259P and RM11-1a plotted against twofold dilution series of pentachlorophenol. (**d**) Genome-wide linkage plot uncovers a peak marker with lod score of 17 on chromosome XIII at base pair 27644 that is linked to pentachlorophenol (the chemical structure in the inset).

resistance to pentachlorophenol (**Fig. 4a–d**). We also showed that the second missense mutation in *PHO84*, leading to an aspartate-to-asparagine substitution at position 55 (D55N), does not, either on its own or in combination with L259P, confer resistance to the two polychlorinated phenols (data not shown).

DISCUSSION

By measuring the growth of 104 genotyped yeast segregants in response to treatment with 100 diverse small molecules, we identified 124 distinct linkages between genetic markers and response to 83 compounds. Clustering segregants according to their growth in the presence of each SMP demonstrated that natural genetic variation in yeast consistently groups functional analogs. Our approach is complementary to previously published compound classification strategies[36,37]. In general, the continued annotation of biologically active small molecules with our collection of genotyped segregants will create an invaluable resource for compound classification, in particular as an easy-to-use, first-pass tool for synthetic chemists synthesizing libraries of novel small molecules of unknown function or for pharmacologists wishing to gain greater insight into the primary and off-target effects of existing therapeutic drugs.

Another advantage to using a collection of richly annotated genotyped segregants—which is not explored in this study—is the ability to compare disparate phenotypes. For example, the segregants described here were the focus of an extensive study of steady-state mRNA expression traits[23]. Expression levels have been used[38] to predict compound response. We have preliminary evidence that expression- and genotype-based prediction of the response to some small-molecule drugs is possible. For example, sensitivity or resistance to rapamycin, which inhibits the TOR proteins, can be predicted with 96% accuracy both from the expression level of the gene *GCN20* (YFR009W) and from the genotype of the single most strongly linked marker (see Methods). Two other genes whose expression level strongly correlates with response to rapamycin are *GCN1* (YGL195W) and *DHH1* (YDL160C). The protein encoded by *GCN1* forms a complex with the protein encoded by *GCN20*, and this Gcn20p-Gcn1p complex positively regulates the protein encoded by *GCN2* (YDR283C), which in turn negatively regulates translation elongation at ribosomes in response to amino acid starvation, a physiological state approximated by treatment of cells with rapamycin[39]. As expected, segregants with low steady-state expression of *GCN20* and *GCN1* are resistant to the growth-arresting effects of rapamycin. Dhh1p is also known to repress translation through its regulation of mRNA decapping[40]. Rapamycin represents a very favorable case for prediction—one in which genotype and expression are equally good predictors. A more general evaluation of predictive power and of the relationship between expression- and genotype-based prediction, including the use of multiple transcripts and markers, will be reported elsewhere.

We have identified eight QTL hotspots linked to response to multiple small molecules, seven of which have been shown previously to affect the abundance of multiple transcripts in this cross. The only exception is the hotspot on chromosome I, which contains no obvious polymorphic candidates (**Supplementary Note** online). Thus, potentially pleiotropic regulators, rather than 'drug response' genes, are responsible for differential sensitivities to multiple compounds. However, 47 loci outside the hotspots each influence response to one to three compounds, and polymorphisms in 'drug response' genes probably account for some of these. We may have failed to observe linkage hotspots at so-called 'drug response' genes because fewer of these genes contain functional polymorphisms between the parent strains and because this GO category may be incomplete. However, we have provided direct evidence (*LEU2*) and indirect evidence (*HAP1*) that known underlying regulators of expression levels also affect response to small molecules. Often, but not always, the BY laboratory strain carries the derived allele (not shared by other *S. cerevisiae* strains) of the responsible polymorphisms (for example, transposon insertion in *HAP1*), an observation that has been previously noted[19]. In the future, it will be worthwhile to cross non-laboratory strains that have not experienced population bottlenecks or *in vitro* selection associated with domestication.

We experimentally verified that a deficiency in leucine biosynthesis underlies sensitivity to a dihydropyridine, niguldipine, that is structurally related to therapeutic calcium-channel blockers. Recently, niguldipine has been showed to inhibit human P-glycoprotein transporters, which are involved in multidrug efflux[41]. Given the polyspecificity of niguldipine for membrane transporters, it is conceivable that it also inhibits a leucine permease. We also experimentally verified that a natural polymorphism in the inorganic-phosphate transporter *PHO84* underlies sensitivity to two polychlorinated phenols that uncouple oxidative phosphorylation. Linkage of many other compounds to this locus with the same direction of allelic effect strongly suggests that this polymorphism also affects response to many other small molecules. One explanation for this result derives from the fact that *PHO84* is orthologous to well-known human xenobiotic transporters belonging to the solute carrier transporter (SLC) superfamily. Polymorphisms in

those genes have been shown to alter drug sensitivity in humans[42]. The proline (RM) allele of *PHO84* that confers drug resistance in yeast is conserved in the orthologous human transporters, suggesting that the leucine (BY) allele is a loss-of-function mutation. In addition to its role as an inorganic phosphate transporter, *PHO84* has also been shown to transport the metal ion manganese[43] and may possess additional, unappreciated selectivity for structurally unrelated xenobiotic substrates. Therefore, the protein encoded by the BY allele of *PHO84* may either be incapable of effluxing xenobiotic substrates itself or may have compromised inorganic phosphate or manganese transport, which indirectly affects membrane permeability of cells to xenobiotics.

Both experimentally verified polymorphisms affecting response to multiple compounds involve transport of nutrients and/or xenobiotic substrates. It will be interesting to evaluate the role of transporters in cases where single SMPs map to one or more unique QTLs. Identifying additional functional polymorphisms that underlie response to small molecules of general therapeutic interest (for example, rapamycin and sertraline) is a prime direction for future study. The identification of loci or polymorphisms that are pharmacodeterminants in our model system of a segregating population of genetically diverse yeast strains (for example, *PHO84*) may serve as a springboard from which to perform comparable studies in humans. Also, the identification of polymorphisms that affect the cellular response of budding yeast to therapeutic antifungal drugs may increase understanding of drug resistance mechanisms in fungal pathogens. In sum, our results provide an initial step toward predictive tailoring of drug treatments to individuals to maximize therapeutic benefit and minimize adverse side effects.

METHODS

Yeast strains and media. BY4713 (MATα *leu2Δ0*) and BY4741 (MAT**a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0*) were obtained from ATCC. RM11-1a (MAT**a** *leu2Δ ura3Δ*), RM11-1b (MATα *lys2Δ ura3Δ*) and BY4716 (MATα *lys2Δ*) were gifts of B. Garvik (Fred Hutchison Cancer Research Center). RM11-1a is one of the two parent strains and is referred to throughout as 'RM'; BY4716 is the other parent strain and is referred to throughout as 'BY'. Rich medium (YPD) is 2% yeast extract, 2% peptone and 2% glucose. Complete synthetic medium (CSM) is 6.7 g/l yeast nitrogen base (YNB), 0.05% ammonium sulfate (AS) and 2% glucose.

Small-molecule perturbagens. Powder stocks were purchased from either Biomol, Sigma-Aldrich or Microsource Discovery and were then resuspended in DMSO and stored at –20 °C. A complete list of SMPs used in this study has been included in **Supplementary Table 1**.

Phenotyping and linkage analysis. We manually arrayed 104 BY and RM segregants into a stock plate. We included inocula of three separate colonies of each BY and RM segregant as biological replicates, and each experiment included an experimental replicate. We then used the stock plates and sterile polypropylene 384-pin replicators (Genetix, #X5050) to inoculate NUNC 384-well clear-bottom untreated sterile plates (VWR, #62409-604) containing rich medium and either vehicle solvent (DMSO, as a control for intrinsic growth differences among the segregants) or an SMP. DMSO-treated plates were analyzed side by side with SMP-treated plates. Inoculated plates were grown without agitation on the benchtop at ambient temperature (22–25 °C) for up to 120 h, during which time multiple OD_{600} measurements were made. At each of these time points, plates were agitated on a standard tabletop Vortex (VWR) for 10–30 s before measurement in a SpectraMax plate reader (Molecular Devices) set to 600 nm emission. All linkage analysis was performed using the R/Qtl software package[44]. To test significance, segregant final yield in each compound treatment was randomly permuted among the segregants, and genome-wide linkage analysis was performed. The entire process was repeated 20 times, and false discovery rates were calculated based on the average number of loci exceeding a given lod score in the permuted data sets. The analysis was done using a set of 104 segregants and 2,956 markers. Phenotypes and genotypes were arranged into CSV file format for direct input into R/Qtl. Phenotypes (compound responses) were scanned under the default (normal) single-QTL model using the standard EM algorithm. All OD_{600} measurements under control conditions (DMSO-only) were used as additive covariates in order to eliminate their effect on linkage. Linkage 'hotspot' analysis was performed on the physical map for simplicity and because the physical locations of the markers are known exactly, whereas the genetic map locations are only estimates. The number of compounds expected to map to each bin by chance is 0.2, as determined by dividing the total number of linkages (124) by the total number of bins (611). The probability of observing four or more linkages to one bin was computed to be 5.8×10^{-5} based on the Poisson distribution. The probability that any of the 611 bins would contain four or more linkages was computed by applying a Bonferroni correction for 611 tests to this value to obtain $P < 0.05$.

Clustering and fungal sequence alignments. Hierarchical clustering was performed with Cluster 3.0 (M. Eisen, Stanford University) using the average linkage clustering method. Clustergrams were visualized with Java TreeView v. 1.0.12 (Alok) with default settings. Sequence alignments were performed using either BLAST or the Saccharomyces Genome Database (SGD). Gene Ontology (GO) enrichment analysis was performed on the SGD website using a hypergeometric *P* value.

Prediction of response to rapamycin. Segregants with residual final yields (after removing the growth effects of the nine control (DMSO) replicates by using them as covariates in a linear regression) over 0.2 were classified as resistant; those with residual final yields less than –0.2 were classified as sensitive and those falling between –0.2 and 0.2 were left unclassified. For the selected time point and concentration (115 h and 25 nM), 29 segregants were classified as sensitive or resistant. We performed leave-one-out cross-validation over all sensitive and resistant segregants to build a predictor using data for all but one segregant (the training set) and then tested this predictor on the one excluded segregant (the test set). This was repeated with each segregant playing the role of the test set. For each training set, we computed the correlation to yield of genotype at each marker and of expression level of each transcript over all segregants in that training set and all unclassified segregants (excluding those segregants with missing data for a given marker or transcript) and then selected the most correlated marker or transcript to be the predictor for genotype or expression, respectively. The test segregant was always excluded when computing the correlations. We used the genotypes of the single most correlated marker at each segregant to determine which allele more often conferred resistance or sensitivity, and we predicted the phenotype of the test segregant based on its genotype at this marker as the more frequent phenotype for the observed allele. For prediction based on expression, we sorted the segregants by expression and took the average expression value between the two segregants that best separated resistant segregants from sensitive segregants as our threshold. We then used the same prediction strategy as with genotype, with expression above or below the threshold taking the role of the marker alleles. The most correlated marker based on genotype in every cross-validation replicate was NNL035W at chromosome XIV position 449639, which was linked to segregant yield in rapamycin with a lod score of 12 and which had a typical correlation coefficient of ~0.65. This marker correctly predicted 27/28 (96.4%) classified segregants (one segregant lacked genotype information at this marker). The most correlated gene in every replicate was *GCN20* (YFR009W), with a typical correlation coefficient of ~0.64. High or low expression of this gene correctly predicts response in 28/29 (96.6%) classified segregants.

Allele swapping and dose response experiments. We used the previously described modified 'delitto perfetto' method to create the L259P and D55N *PHO84* allele swap strains[45]. See **Supplementary Table 7** online for a complete list of primers. Dose response experiments were performed in multi-well (96- or 384-well) plates as described above. Serial dilutions were carried out in replicates, and the resulting growth curves and accompanying IC_{50} values were generated with GraphPad Prism (v. 4.01).

URLs. BLAST: http://www.ncbi.nlm.nih.gov/BLAST/; SGD: http://www.yeastgenome.org.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

E.O.P. acknowledges F. Storici for technical advice and D. Altshuler for useful discussions. D.C.R acknowledges discussions with R. Schapire, S. Kulkarni and W. Schoendorf. RM11-1a (MAT**a** *leu2Δ ura3Δ*), RM11-1b (MATα *lys2Δ ura3Δ*) and BY4716 (MATα *lys2Δ*) were gifts of B. Garvik (Fred Hutchison Cancer Research Center). This work was supported by the US National Institute of General Medicine Sciences (S.L.S.) and the US National Institute of Mental Health (L.K.). Work at Princeton was supported in part by a Center grant P50GM071508 from the US National Institute of General Medical Science/US National Institutes of Health. L.K. is a James S. McDonnell Centennial Fellow. S.L.S. is an Investigator at the Howard Hughes Medical Institute.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions

1. Weinstein, J.N. *et al.* An information-intensive approach to the molecular pharmacology of cancer. *Science* **275**, 343–349 (1997).
2. Dolan, M.E. *et al.* Heritability and linkage analysis of sensitivity to cisplatin-induced cytotoxicity. *Cancer Res.* **64**, 4353–4356 (2004).
3. Watters, J.W., Kraja, A., Meucci, M.A., Province, M.A. & McLeod, H.L. Genome-wide discovery of loci influencing chemotherapy cytotoxicity. *Proc. Natl. Acad. Sci. USA* **101**, 11809–11814 (2004).
4. Shukla, S.J. & Dolan, M.E. Use of CEPH and non-CEPH lymphoblast cell lines in pharmacogenetic studies. *Pharmacogenomics* **6**, 303–310 (2005).
5. Le Morvan, V. *et al.* Relationships between genetic polymorphisms and anticancer drug cytotoxicity vis-à-vis the NCI-60 panel. *Pharmacogenomics* **7**, 843–852 (2006).
6. Moisan, F., Longy, M., Robert, J. & Le Morvan, V. Identification of gene polymorphisms of human DNA topoisomerase I in the National Cancer Institute panel of human tumour cell lines. *Br. J. Cancer* **95**, 906–913 (2006).
7. Yarosh, D.B., Pena, A. & Brown, D.A. DNA repair gene polymorphisms affect cytotoxicity in the National Cancer Institute Human Tumour Cell Line Screening Panel. *Biomarkers* **10**, 188–202 (2005).
8. Stoehlmacher, J. *et al.* A multivariate analysis of genomic polymorphisms: prediction of clinical outcome to 5-FU/oxaliplatin combination chemotherapy in refractory colorectal cancer. *Br. J. Cancer* **91**, 344–354 (2004).
9. Hohmann, S. The Yeast Systems Biology Network: mating communities. *Curr. Opin. Biotechnol.* **16**, 356–360 (2005).
10. Foury, F. Human genetic disease: a cross-talk between man and yeast. *Gene* **195**, 1–10 (1997).
11. Steinmetz, L.M. *et al.* Systematic screen for human disease genes in yeast. *Nat. Genet.* **31**, 400–404 (2002).
12. Outeiro, T.F. & Giorgini, F. Yeast as a drug discovery platform in Huntington's and Parkinson's disease. *Biotechnol. J.* **1**, 258–269 (2006).
13. Ooi, S.L. *et al.* Global synthetic-lethality analysis and yeast functional profiling. *Trends Genet.* **22**, 56–63 (2006).
14. Sinha, H., Nicholson, B.P., Steinmetz, L.M. & McCusker, J.H. Complex genetic interactions in a quantitative trait locus. *PloS Genet.* **2**, e13 (2006).
15. Steinmetz, L.M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416**, 326–330 (2002).
16. Deutschbauer, A.M. & Davis, R.W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nat. Genet.* **37**, 1333–1340 (2005).
17. Ben-Ari, G., Zenvirth, D., Sherman, A. & Klutstein, D.L. Four linked genes participate in controlling sporulation efficiency in budding yeast. *PLoS Genet.* (in the press) (2006).
18. Brem, R.B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296**, 752–755 (2002).
19. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat. Genet.* **35**, 57–64 (2003).
20. Morley, M. *et al.* Genetic analysis of genome-wide variation in human gene expression. *Nature* **430**, 743–747 (2004).
21. Monks, S.A. *et al.* Genetic inheritance of gene expression in human cell lines. *Am. J. Hum. Genet.* **75**, 1094–1105 (2004).
22. Perlstein, E.O. *et al.* Revealing complex traits with small molecules and naturally recombinant yeast strains. *Chem. Biol.* **13**, 319–327 (2006).
23. Brem, R.B. & Kruglyak, L. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc. Natl. Acad. Sci. USA* **102**, 1572–1577 (2005).
24. Ruderfer, D.M., Pratt, S.C., Seidel, H.S. & Kruglyak, L. Population genomc analysis of outcrossing and recombination in yeast. *Nat. Genet.* **38**, 1077–1081 (2006).
25. Maro, B., Marty, M.C. & Bornens, M. *In vivo* and *in vitro* effects of the mitochondrial uncoupler FCCP on microtubules. *EMBO J.* **1**, 1347–1352 (1982).
26. Fogel, S., Welch, J.W. & Maloney, D.H. The molecular genetics of copper resistance in *Saccharomyces cerevisiae* – a paradigm for non-conventional yeasts. *J. Basic Microbiol.* **28**, 147–160 (1988).
27. Tamura, K. *et al.* A hap1 mutation in a laboratory strain of *Saccharomyces cerevisiae* results in decreased expression of ergosterol-related genes and cellular ergosterol content compared to sake yeast. *J. Biosci. Bioeng.* **98**, 159–166 (2004).
28. Rogers, B. *et al.* The pleiotropic drug ABC transporters from *Saccharomyces cerevisiae*. *J. Mol. Microbiol. Biotechnol.* **3**, 207–214 (2001).
29. Tenreiro, S. *et al.* AQR1 gene (ORF YNL065w) encodes a plasma membrane transporter of the major facilitator superfamily that confers resistance to short-chain monocarboxylic acids and quinidine in *Saccharomyces cerevisiae*. *Biochem. Biophys. Res. Commun.* **292**, 741–748 (2002).
30. Bolster, D.R., Vary, T.C., Kimball, S.R. & Jefferson, L.S. Leucine regulates translation initiation in rat skeletal muscle via enhanced eIF4G phosphorylation. *J. Nutr.* **134**, 1704–1710 (2004).
31. Weinbach, E.C. Biochemical basis for the toxicity of pentachlorophenol. *Proc. Natl. Acad. Sci. USA* **43**, 393–397 (1957).
32. Somerville, L. The metabolism of fungicides. *Xenobiotica* **16**, 1017–1030 (1986).
33. Bun-Ya, M., Nishimura, M., Harashima, S. & Oshima, Y. The PHO84 gene of *Saccharomyces cerevisiae* encodes an inorganic phosphate transporter. *Mol. Cell. Biol.* **11**, 3229–3238 (1991).
34. Thomas, M.R. & O'Shea, E.K. An intracellular phosphate buffer filters transient fluctuations in extracellular phosphate levels. *Proc. Natl. Acad. Sci. USA* **102**, 9565–9570 (2005).
35. Cordes, F.S., Bright, J.N. & Sansom, M.S. Proline-induced distortions of transmembrane helices. *J. Mol. Biol.* **323**, 951–960 (2002).
36. Yeh, P., Tschumi, A.I. & Kishony, R. Functional classification of drugs by properties of their pairwise interactions. *Nat. Genet.* **38**, 489–494 (2006).
37. Borissy, A.A. *et al.* Systematic discovery of multicomponent therapeutics. *Proc. Natl. Acad. Sci. USA* **100**, 7977–7982 (2003).
38. Staunton, J.E. *et al.* Chemosensitivity prediction by transcriptional profiling. *Proc. Natl. Acad. Sci. USA* **98**, 10787–10792 (2001).
39. Marton, M.J., Vazquez de Aldana, C.R., Qui, H., Chakraburtty, K. & Hinnebusch, A.G. Evidence that GCN1 and GCN20, translational regulators of GCN4, function on elongating ribosomes in activation of eIFalpha kinase GCN2. *Mol. Cell. Biol.* **17**, 4474–4489 (1997).
40. Coller, J. & Parker, R. General translational repression by activators of mRNA decapping. *Cell* **122**, 875–886 (2005).
41. Zhou, X.F., Yang, X., Wang, Q., Coburn, R.A. & Morris, M.E. Effects of dihydropyridines and pyridines on multidrug resistance mediated by breast cancer resistance protein: *in vitro* and *in vivo* studies. *Drug Metab. Dispos.* **33**, 1220–1228 (2005).
42. Urban, T.J. *et al.* Functional genomics of membrane transporters in human populations. *Genome Res.* **16**, 223–230 (2006).
43. Jensen, L.T., Ajua-Alemanji, M. & Culotta, V.C. The *Saccharomyces cerevisiae* high affinity phosphate transporter encoded by PHO84 also functions in manganese homeostasis. *J. Biol. Chem.* **278**, 42036–42040 (2003).
44. Broman, K.W., Wu, H., Sen, S. & Churchill, G.A. R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890 (2003).
45. Storici, F., Durham, C.L., Gordenin, D.A. & Resnick, M.A. Chromosomal site-specific double-strand breaks are efficiently targeted for repair by oligonucleotides in yeast. *Proc. Natl. Acad. Sci. USA* **100**, 14994–14999 (2003).