Vol 464|15 April 2010|doi:10.1038/nature08923

Dissection of genetically complex traits with extremely large pools of yeast segregants

Ian M. Ehrenreich[1,2,3], Noorossadat Torabi[1,4], Yue Jia[1,3], Jonathan Kent[1], Stephen Martis[1], Joshua A. Shapiro[1,2,3], David Gresham[1]†, Amy A. Caudy[1] & Leonid Kruglyak[1,2,3]

Most heritable traits, including many human diseases[1], are caused by multiple loci. Studies in both humans and model organisms, such as yeast, have failed to detect a large fraction of the loci that underlie such complex traits[2,3]. A lack of statistical power to identify multiple loci with small effects is undoubtedly one of the primary reasons for this problem. We have developed a method in yeast that allows the use of much larger sample sizes than previously possible and hence permits the detection of multiple loci with small effects. The method involves generating very large numbers of progeny from a cross between two *Saccharomyces cerevisiae* strains and then phenotyping and genotyping pools of these offspring. We applied the method to 17 chemical resistance traits and mitochondrial function, and identified loci for each of these phenotypes. We show that the level of genetic complexity underlying these quantitative traits is highly variable, with some traits influenced by one major locus and others by at least 20 loci. Our results provide an empirical demonstration of the genetic complexity of a number of traits and show that it is possible to identify many of the underlying factors using straightforward techniques. Our method should have broad applications in yeast and can be extended to other organisms.

Genome-wide association studies (GWAS) have recently detected many trait loci in humans[4]. Despite the large number of loci that have been identified by GWAS, case studies, such as human height[5], have shown that we remain unable to explain the genetic basis of complex traits in our population[2]. Controlled crosses in model organisms can shed light on this problem by elucidating basic principles that govern the genetic basis of trait variation. However, akin to the problem in humans, mapping studies in model organisms typically detect only a fraction of the loci underlying heritable traits, implying that they lack statistical power[3].

Very large mapping populations are needed to dissect comprehensively the genetic basis of highly complex traits. In many cases, genotyping and phenotyping on a sufficient scale will not be feasible without the use of methods that examine pools of individuals. One such method, bulk segregant analysis (BSA), was first proposed nearly twenty years ago as an expeditious approach for mapping quantitative trait loci (QTLs)[6], and its modern implementations are commonly used to map major effect QTLs and Mendelian loci[7–11]. However, BSA has yet to be effectively used to dissect a highly complex trait, even though simulations indicate that it should be capable of detecting numerous small-effect loci with high resolution when $>10^5$ cross progeny are used (Supplementary Figs 1 and 2). We have developed a powerful extension of BSA that can be used to map complex traits in yeast comprehensively. Extreme QTL mapping (X-QTL) has three key steps. The first is the generation of segregating populations of very large size. The second is selection-based phenotyping of these populations to recover large numbers of progeny with extreme trait values. This can be accomplished, for example, by selection for drug resistance or by cell sorting. The final step is quantitative measurement of pooled allele frequencies across the genome, by either microarray-based genotyping or massively parallel sequencing.

To generate the pools of segregants that form the starting point for X-QTL, we implemented the Synthetic Genetic Array (SGA) marker scheme[12,13], which enables the recovery of MAT**a** haploids from a cross of appropriately marked parental strains (Fig. 1a, b). We used the *Saccharomyces cerevisiae* strains BY4716 (hereafter referred to as BY), a laboratory strain, and RM11-1a (hereafter referred to as RM), a wine strain, as the progenitors of the pools. We crossed these strains to form a diploid, sporulated the diploid, and selected for $\sim10^7$ unique BY×RM MAT**a** haploid segregants. We designed an allele-specific genotyping microarray with isothermal probes[14] that assays ~18,000 single nucleotide polymorphisms (SNPs) between BY and RM. We tested the array by hybridizing the haploid and diploid progenitor strains, as well as multiple MAT**a** pools, and found that we could discriminate the parental strains and reproducibly identify deviations in allele frequencies associated with the SGA markers and other loci in the segregating pools (Fig. 1c–e). Comparable results were obtained by sequencing pools to ~180× coverage with the Illumina Genome Analyzer (Fig. 1e).

We first used X-QTL to map the genetic basis of sensitivity to 4-nitroquinoline (4-NQO), a DNA damaging agent. We previously showed that sensitivity to 4-NQO is a complex trait in the BY×RM cross[15]. BY×RM segregants show varying degrees of sensitivity, and the parental strains are both intermediate relative to their progeny, suggesting contributions of multiple alleles from each parent. Conventional QTL mapping with 123 genotyped segregants detected a single significant locus on chromosome 12, and subsequent experiments identified an amino acid substitution in the DNA repair gene *RAD5* as the underlying causative polymorphism. A backcrossing strategy identified a smaller contributing effect of a polymorphism in the gene *MKT1*. The BY allele of *RAD5* and the RM allele of *MKT1* conferred 4-NQO resistance, but these loci did not fully explain the observed 4-NQO responses of the segregants, implying that additional loci must exist.

To map the genetic basis of sensitivity to 4-NQO using X-QTL, we first plated segregating pools across a range of drug doses to find a highly selective 4-NQO concentration. We then conducted 4-NQO selections at this concentration, while in parallel growing control populations on rich medium without the drug. 4-NQO-resistant and control pools were harvested, and the extracted DNA was hybridized to genotyping microarrays. To identify loci that confer resistance to 4-NQO, we scanned the genome for locations at which allele frequencies in selected pools were significantly different from the

[1]Lewis-Sigler Institute for Integrative Genomics, [2]Department of Ecology and Evolutionary Biology, [3]Howard Hughes Medical Institute, [4]Department of Molecular Biology, Princeton University, Princeton, New Jersey 08540, USA. †Present address: Center for Genomics and Systems Biology, New York University, New York, New York 10003, USA.

©2010 Macmillan Publishers Limited. All rights reserved

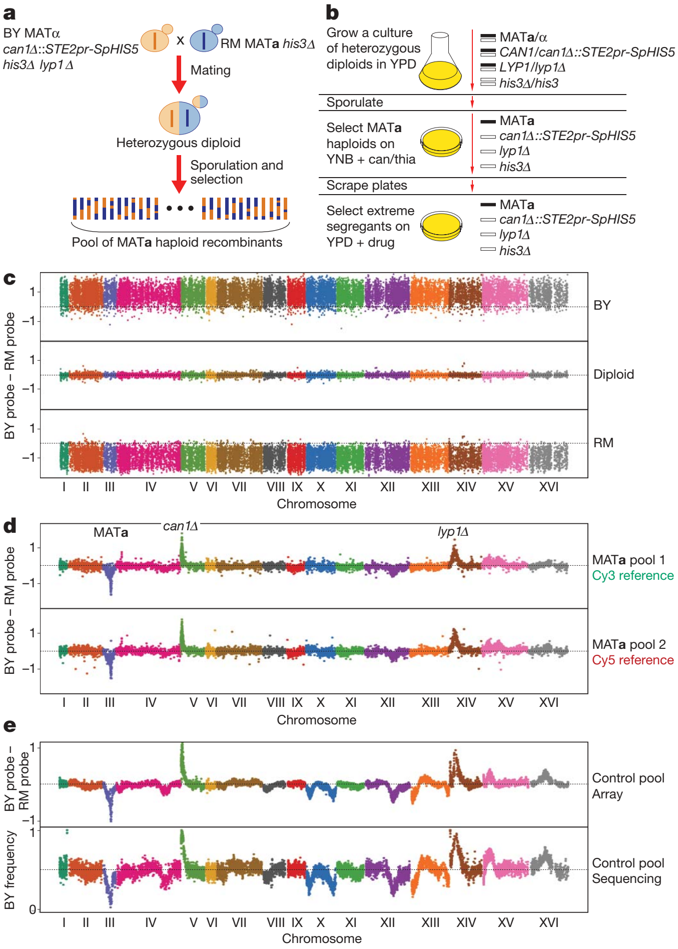

Figure 1 | X-QTL design and quantitative allele frequency measurement in DNA pools. a, **b**, The crossing design used for X-QTL is shown in **a**, whereas the selection scheme used to generate segregating pools is shown in **b**. can/thia, canavanine/thialysine. **c**–**e**, Genotyping of parental strains (**c**), two segregating pools (**d**) and an unselected control pool grown on rich medium (**e**) is shown. Dotted lines at zero indicate no difference between the $\log_{10}$ ratios of the BY and RM allele-specific probes. Enrichment of the BY allele is indicated by deviations above 0 and enrichment of the RM allele is indicated by deviations below 0. For the segregating pools, both the control loci involved in MAT**a** selection and the dye used for reference labelling are denoted. In **d**, we use a dye-swap experiment to show that the dye used for labelling does not cause any bias in allele frequency measurement. Panels **d** and **e** differ in that **d** shows a MAT**a** pool before plating on rich medium and **e** shows a MAT**a** pool after 2 days of growth on rich medium. In **e**, the same pool was hybridized to the genotyping microarray and was sequenced to ~180× coverage with the Illumina Genome Analyzer. The results in **c** and **d** are plots of raw data with no sites removed, whereas in **e** raw data was plotted with sites more than 1.5 standard deviations away from the local average of the 10 nearest data points removed for clarity.

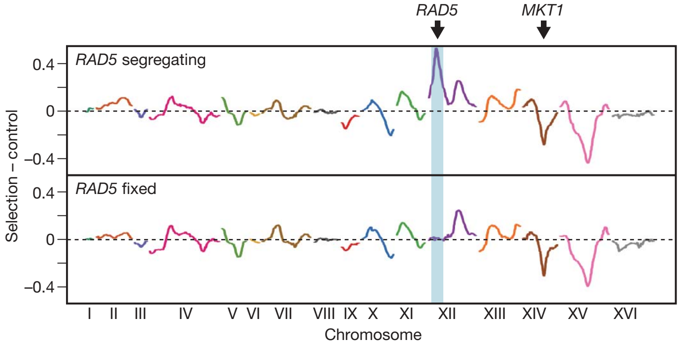

Figure 2 | X-QTL detection of loci for 4-NQO resistance. Results for 4-NQO resistance with *RAD5* segregating (top) and fixed (bottom) are shown. The difference between the average of the selections and the average of the controls generated on the same day is plotted, with enrichment of the BY allele indicated by deviations above 0 and enrichment of the RM allele indicated by deviations below 0. Sliding window averages (50 kb) are plotted. Arrows point to *MKT1* and *RAD5*. The *RAD5* fixed population was generated by using a RM parent strain in which the $RAD5^{RM}$ allele was replaced with a $RAD5^{BY}$ allele. This strain was constructed by crossing strain EAY1467 (ref. 15) to the RM parent strain used for X-QTL.

control pools (Supplementary Methods). Using this approach, we identified 14 loci in the 4-NQO selection at a false discovery rate (FDR) of 0.05. Similar deviations in allele frequency in the selected pools were observed when the genotyping step was carried out by either arrays or short-read sequencing (X-QTL-seq; Supplementary Fig. 3).

We examined whether the loci identified by X-QTL for 4-NQO resistance correspond to real biological effects. Using X-QTL, we observed peaks at *RAD5* and *MKT1*, with both loci selected in the expected direction (Fig. 2). We confirmed that the peak overlapping *RAD5* was actually due to this gene by repeating the BY×RM cross with an RM parent strain that had the BY version of *RAD5*. When 4-NQO resistance was mapped in the selected pool with *RAD5* fixed, the resulting segregating pool showed increased resistance to 4-NQO, and no *RAD5* peak was observed by X-QTL (Fig. 2). Next, we isolated 96 individual progeny from the same cross used to generate the segregating pools, phenotyped them for 4-NQO sensitivity, and genotyped them at the 14 loci identified by X-QTL. Nine of the loci showed significant effects in this independent data set ($P < 0.05$), five of which were highly significant ($P < 0.001$). The loci jointly explained 59% of the phenotypic variance in 4-NQO sensitivity in an additive model (Supplementary Fig. 4). Because we measured the heritability of this trait to be 0.84, the loci explained 70% of the genetic variance, indicating that we have explained most of the genetic basis of this trait with the loci detected by X-QTL.

We next applied X-QTL to resistance to 16 diverse chemical agents (Supplementary Table 1), including a detergent and a number of antifungal compounds, using the same methodology that was used for 4-NQO. At a global FDR of 0.05, we mapped 177 total loci for these 16 traits. We detected between 1 and 24 peaks in pools selected on these agents. Including 4-NQO, we detect an average of 11 peaks per trait, suggesting high genetic complexity for many traits. The 17 traits show marked differences in their genetic architectures (Fig. 3 and Supplementary Fig. 5). At the simpler end of the range, resistance to cadmium chloride, copper sulphate and ethanol is controlled by one major locus for each trait (Fig. 3a). At the other extreme, we identified more than 20 loci in the diamide, hydrogen peroxide and sodium dodecyl sulphate selections (Fig. 3b). Other traits show intermediate levels of complexity (Fig. 3c, d).

We compared the 191 peaks detected across the 17 traits. The genome was divided into 20-kilobase (kb) bins, and all loci within a bin were grouped together. Using this procedure, we found 123 distinct loci (Fig. 3e). Of these, 82 loci (~67%) were trait-specific. For instance, a peak was detected at *RAD5* on chromosome XII only in our analysis of resistance to 4-NQO. Similarly, the major locus for copper sulphate resistance, which was previously mapped in a screen for QTLs involved in resistance to small molecules in the BY×RM cross[16], coincides with the location of the *CUP1* genes on chromosome VIII and was detected only in the copper sulphate selection. Of the 41 remaining loci, 40 were detected for 2 to 5 traits, and 1 locus, which overlaps *MKT1*, was detected for 8 different compounds. An amino acid polymorphism in *MKT1* is known to be involved in a large number of trait differences between BY and other strains, including 4-NQO resistance[15], sensitivity to dipropyldopamine and phenylephrine[17], high temperature growth[18], sporulation efficiency[19], gene expression[20], and growth of petite colonies[21]. Our results suggest that in addition to these previously studied phenotypes, *MKT1* also has a broadly pleiotropic effect on drug resistance under the conditions of

©2010 Macmillan Publishers Limited. All rights reserved

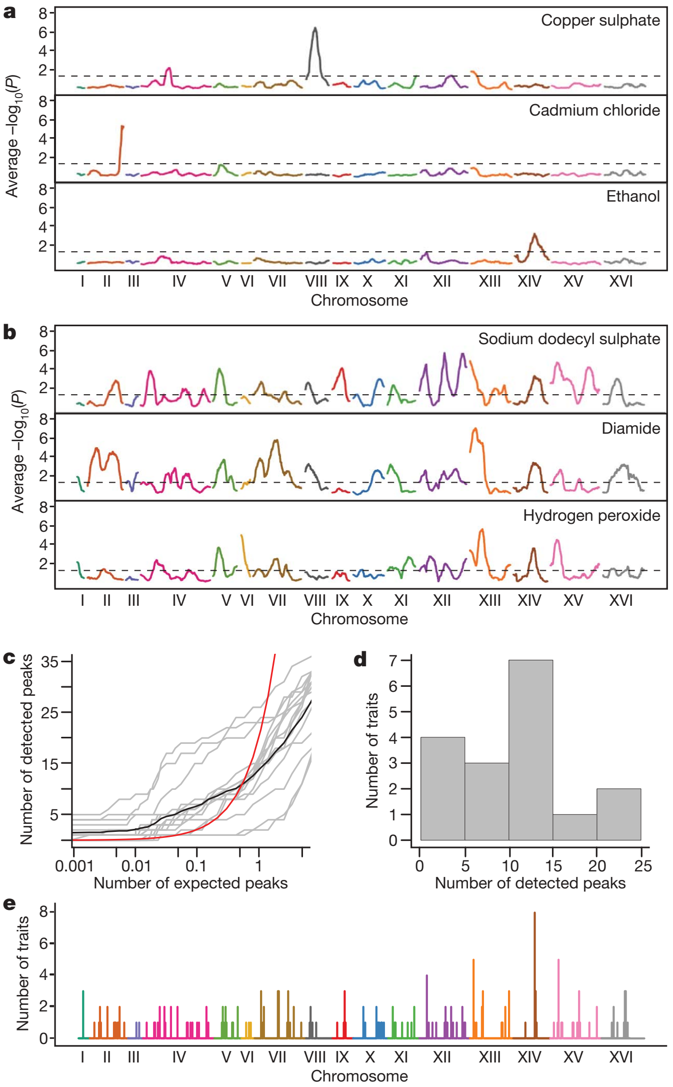

Figure 3 | Genetic architecture of chemical resistance traits. **a**–**e**, Examples of genetically simple traits (**a**), examples of genetically complex traits (**b**), relationship between the number of expected and detected peaks (**c**), the number of loci detected per trait (**d**), and a map of compound-specific and pleiotropic loci across the genome (**e**). In **a** and **b** the $-\log_{10}(P)$ values are shown for *t*-tests comparing selected samples to control samples. The sliding averages within 50-kb windows for these tests are plotted. Peaks above the dashed lines in **a** and **b** are significant at an FDR of 0.05. In **c**, the average relationship between expected and detected peaks is plotted as a black line and the trait-specific relationships are plotted as grey lines. The red line plots the theoretical relationship between expected and detected peaks at an FDR of 0.05. The expected counts were generated from permutations of the chemical resistance data set. The histogram in **d** was made using loci significant at a global FDR of 0.05. In **e**, detected loci were grouped within 20-kb windows across the genome.

our study. Furthermore, our results suggest that X-QTL detects loci at a fine resolution, as the locations of the peaks corresponding to *MKT1* and *RAD5* were estimated to be within 2 kb of these genes themselves (Supplementary Table 2). The loci we have detected across 17 compounds thus provide a foundation for comprehensively studying the molecular mechanisms that shape phenotypic variation in response to chemical agents among yeast strains.

Selections for resistance to chemical agents permit only one extreme tail of the phenotype distribution to be sampled. Additional insights can be gained from selections where both high and low extreme segregants can be recovered. Fluorescence-activated cell sorting (FACS) provides a straightforward approach to such two-tailed selections, as large numbers of individuals exhibiting high and low values for a stain or reporter can easily be recovered. To pilot this approach, we used the dye Mitotracker red, which stains cells depending on the mitochondrial proton gradient and mitochondrial volume. We harvested a MAT**a** pool, stained it with Mitotracker red, and then sorted out extreme cells by FACS. We sorted a population of $\sim 5 \times 10^6$ cells and selected 3×10^4 cells from each tail. These selected cells were then grown up on agar plates with rich medium to generate enough cells from which to extract DNA. DNA pools from both tails, as well as from a subsample of the whole population, were hybridized to the genotyping microarray.

Comparison of the high and low extremes found multiple major peaks at an FDR of 0.05 (Fig. 4). These peaks showed similar heights but opposite directions in the two tails. The location of one of the peaks provided a strong candidate for the causal gene. The peak on chromosome XII spans *HAP1*, a zinc finger transcription factor involved in response to oxygen. *HAP1* was previously shown to be a hotspot for *trans* regulation of gene expression differences in the BY×RM cross[22,23]. BY has a partially functional allele of *HAP1* due to a Ty transposon insertion in the *HAP1* coding region[24], whereas RM has a fully functional *HAP1* allele. Consistent with *HAP1*'s function, segregants carrying the RM allele of *HAP1* show increased oxidative capacity based on X-QTL mapping. Comparison of BY with a partially functional *HAP1* to BY with a fully functional *HAP1* shows that *HAP1* has a causal role in variation in Mitotracker red staining (Supplementary Fig. 6).

X-QTL represents a powerful method for rapidly and cost-effectively mapping the multiple QTLs underlying a trait difference between two yeast strains. We have used X-QTL to demonstrate empirically that many traits have a highly complex genetic basis. These results are consistent with previous studies in yeast, such as those focused on transcript levels[3], protein abundance[25] and sensitivity to chemical agents[16], in which genetic complexity was inferred from trait distributions and a lack of mapped loci, rather than from direct detection of multiple loci as we have accomplished here. Our results agree with those from the comprehensive genetic dissection of a small number of traits in other model organisms, such as bristle number in *Drosophila*[26] and flowering time in maize[27], which have shown that dozens of loci can underlie a difference between two individuals. Notably, whereas these studies required substantial labour, time and resources, X-QTL is a quick and easy approach to achieve a comparable level of genetic dissection. The levels of complexity observed here (for example, 14 loci explaining 70% of the genetic variance for 4-NQO resistance) are still markedly lower than those seen for some human traits in GWAS (for example, 40 loci explaining 5% of the variance for height[2,5]). One obvious explanation is the difference in experimental designs (line crosses versus population association studies), but differences in genetic architectures among species and traits may also contribute. The comprehensive genetic dissection of complex traits by X-QTL makes it possible to answer empirically many of the basic questions about the genetic architecture of complex traits, including

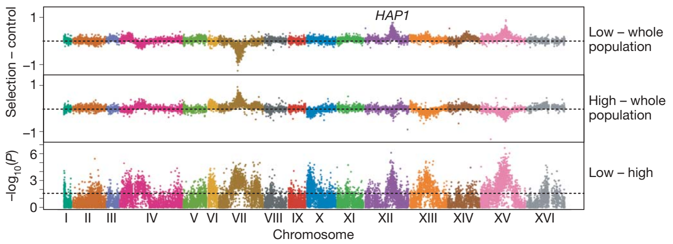

Figure 4 | X-QTL mapping of mitochondrial activity by cell sorting. Segregants were stained with the dye Mitotracker red. The comparisons of high and low pools to the entire population are shown, in addition to $-\log_{10}(P)$ values for the difference between these groups. The dashed lines in the high or low minus control plots indicate zero difference in a comparison, whereas the dashed line in the low minus high plot indicates the probe-level threshold for an FDR of 0.05.

©2010 Macmillan Publishers Limited. All rights reserved

the number of loci underlying a trait and the distribution of their allele frequencies in a population. High-resolution mapping of these loci also enables identification of the underlying genes and sequence variants, as well as investigation of allelic effect sizes and genetic interactions. We anticipate that general insights from such studies will be applicable to understanding the genetics of complex traits in other organisms, including humans, and that variants of X-QTL can be developed for other species.

METHODS SUMMARY

Microarray hybridizations. DNA was extracted from segregating pools using Qiagen Genomic-tip 100/G columns. DNA was labelled using array comparative genomic hybridization reagents from Invitrogen and Cy3- or Cy5-labelled dUTP from Enzo. Hybridization, scanning and feature extraction were done using Agilent equipment and software. Normalization of arrays was done using the rank invariant method within the Agilent software.

Statistical analysis. For a given SNP, the difference in $\log_{10}$ ratios of the intensities of the BY and RM allele-specific probes on a single array was computed, and this metric was used in downstream analyses. In cases where a SNP was represented by two probe sets, the probe sets were used as separate data points. For the drug selections, selection and control experiments were compared using *t*-tests with equal variances. A regression-based peak-finding approach was then used, which scans the genome for locations where the slope in $-\log_{10}(P)$ values changes signs. Significance levels were determined by permutation (Fig. 3c). For the Mitotracker red study, the high- and low-staining pools were compared using *t*-tests with equal variances. QVALUE[28] was then used to determine an FDR based on the observed *P* values.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 November 2009; accepted 16 February 2010.

1. Plomin, R., Haworth, C. M. A. & Davis, O. S. P. Common disorders are quantitative traits. *Nature Rev. Genet.* **10,** 872–878 (2009).
2. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461,** 747–753 (2009).
3. 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).
4. Hindorff, L. A. *et al.* Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA* **106,** 9362–9367 (2009).
5. Visscher, P. M. Sizing up human height variation. *Nature Genet.* **40,** 489–490 (2008).
6. Michelmore, R. W., Paran, I. & Kesseli, R. V. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. *Proc. Natl Acad. Sci. USA* **88,** 9828–9832 (1991).
7. Wolyn, D. J. *et al.* Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana*. *Genetics* **167,** 907–917 (2004).
8. Brauer, M. J., Christianson, C. M., Pai, D. A. & Dunham, M. J. Mapping novel traits by array-assisted bulk segregant analysis in *Saccharomyces cerevisiae. Genetics* **173,** 1813–1816 (2006).
9. Segrè, A. V., Murray, A. W. & Leu, J. Y. High-resolution mutation mapping reveals parallel experimental evolution in yeast. *PLoS Biol.* **4,** e256 (2006).
10. Lai, C. Q. *et al.* Speed-mapping quantitative trait loci using microarrays. *Nature Methods* **4,** 839–841 (2007).
11. Schneeberger, K. *et al.* SHOREmap: simultaneous mapping and mutation identification by deep sequencing. *Nature Methods* **6,** 550–551 (2009).
12. Tong, A. H. *et al.* Systematic genetic analysis with ordered arrays of yeast deletion mutants. *Science* **294,** 2364–2368 (2001).
13. Tong, A. H. & Boone, C. High-throughput strain construction and systematic synthetic lethal screening in *Saccharomyces cerevisiae*. In *Yeast Gene Analysis* 2nd edn *Methods in Microbiology* Vol. 36, 369–707 (Elsevier, 2007).
14. Gresham, D. *et al.* Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. *Proc. Natl Acad. Sci. USA* **107,** 1482–1487 (2010).
15. Demogines, A., Smith, E., Kruglyak, L. & Alani, E. Identification and dissection of a complex DNA repair sensitivity phenotype in Baker's yeast. *PLoS Genet.* **4,** e1000123 (2008).
16. Perlstein, E. O., Ruderfer, D. M., Roberts, D. C., Schreiber, S. L. & Kruglyak, L. Genetic basis of individual differences in the response to small-molecule drugs in yeast. *Nature Genet.* **39,** 496–502 (2007).
17. Kim, H. S. & Fay, J. C. A combined cross analysis reveals genes with drug-specific and background-dependent effects on drug sensitivity in *Saccharomyces cerevisiae*. *Genetics* **183,** 1141–1151 (2009).
18. Steinmetz, L. M. *et al.* Dissecting the architecture of a quantitative trait locus in yeast. *Nature* **416,** 326–330 (2002).
19. Deutschbauer, A. M. & Davis, R. W. Quantitative trait loci mapped to single-nucleotide resolution in yeast. *Nature Genet.* **37,** 1333–1340 (2005).
20. Smith, E. N. & Kruglyak, L. Gene-environment interaction in yeast gene expression. *PLoS Biol.* **6,** e83 (2008).
21. Dimitrov, L. N., Brem, R. B., Kruglyak, L. & Gottschling, D. E. Polymorphisms in multiple genes contribute to the spontaneous mitochondrial genome instability of *Saccharomyces cerevisiae* S288C strains. *Genetics* **183,** 365–383 (2009).
22. Brem, R. B., Yvert, G., Clinton, R. & Kruglyak, L. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296,** 752–755 (2002).
23. Yvert, G. *et al.* Trans-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nature Genet.* **35,** 57–64 (2003).
24. Gaisne, M., Becam, A. M., Verdiere, J. & Herbert, C. J. A 'natural' mutation in *Saccharomyces cerevisiae* strains derived from S288c affects the complex regulatory gene *HAP1* (*CYP1*). *Curr. Genet.* **36,** 195–200 (1999).
25. Foss, E. J. *et al.* Genetic basis of proteome variation in yeast. *Nature Genet.* **39,** 1369–1375 (2007).
26. Mackay, T. F. & Lyman, R. F. *Drosophila* bristles and the nature of quantitative genetic variation. *Phil. Trans. R. Soc. Lond. B* **360,** 1513–1527 (2005).
27. Buckler, E. S. *et al.* The genetic architecture of maize flowering time. *Science* **325,** 714–718 (2009).
28. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100,** 9440–9445 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors are grateful to members of the Kruglyak laboratory for comments on this manuscript; D. Botstein, J. Gerke, G. Lang and E. Perlstein for input regarding experiments; J. Bloom, L. Parsons, D. Sangurdekar and J. Storey for advice regarding analyses; and E. Alani for sharing the EAY1467 strain. This work was supported by NIH grant R37 MH59520, a James S. McDonnell Centennial Fellowship, and the Howard Hughes Medical Institute (L.K.), a NIH postdoctoral fellowship F32 HG51762 (I.M.E.), and NIH grant P50 GM071508 to the Center for Quantitative Biology at the Lewis-Sigler Institute of Princeton University.

Author Contributions Experiments were designed by I.M.E., A.A.C. and L.K. Strains were constructed by I.M.E., Y.J., S.M. and A.A.C. Microarrays were designed by I.M.E. and D.G. Experiments were performed by I.M.E., N.T., Y.J., J.K., S.M. and A.A.C. Simulation scripts were written by I.M.E. and J.A.S. Analyses were conducted by I.M.E. The manuscript was written by I.M.E. and L.K., and incorporates comments by all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to L.K. (leonid@genomics.princeton.edu).

©2010 Macmillan Publishers Limited. All rights reserved

METHODS

Construction of segregating pools. To construct segregating pools, we use the Synthetic Genetic Array (SGA) marker system[12,13]. In our cross, we use a BY parent that is MATα *can1Δ::STE2pr-SpHIS5 lyp1Δ his3Δ1* and an RM parent that is MAT**a** *AMN1^{BY} his3Δ0::NatMX ho::HphMX*. These strains were crossed and a diploid zygote was recovered.

To create the segregating pools, a single colony of the diploid progenitor was inoculated into 100 ml YPD and grown to stationary phase. The diploid culture was spun down and the supernatant was decanted. The diploid pellet was then resuspended in 200 ml Spo++ sporulation medium (http://www.genomics.princeton.edu/dunham/sporulationdissection.htm). The sporulation was kept at room temperature (~22 °C) with shaking and monitored for the fraction of diploids that had sporulated. Once more than 50% of the diploids had sporulated, the culture was deemed suitable for downstream use.

The next step in the generation of segregating pools was to select for MAT**a** haploids. Fifty millilitres of the sporulation were spun down and then the supernatant was decanted. The pellet was resuspended in 1 ml water. Three-hundred microlitres β-glucoronidase (Sigma; G7770) were added to the preparation and the mixture was incubated at 30 °C for 1 h. Approximately 50 μl of glass beads (Sigma; G8772) were then added and the sample was vortexed for 2 min. The sample was incubated for an additional hour at 30 °C, followed by a second round of vortexing for 2 min. Water was added to the sample so that the total volume was 20 ml. The spore preparation was spread onto YNB + canavanine/thialysine (Sigma; C9758 for canavanine (L-canavanine sulphate salt); A2636 for thialysine (*S*-(2-aminoethyl)- L-cysteine hydrochloride)), with 100 μl of sample going onto each plate. The plates were incubated at 30 °C. MAT**a** haploids typically grew up after ~2 days.

The final step in pool creation was to mix together MAT**a** segregants selected on different plates. Ten millilitres of water were poured onto a plate and a sterile spreader was used to remove the segregants from the plate. The cell mixtures from each plate were then pipetted off the plates into a separate container. The pool was spun down and the water decanted. For drug selections, the cell pellet was resuspended in 1.5 ml YPD per scraped plate. The segregant pool was incubated at 30 °C for 1 h. One-hundred microlitres of this segregant pool was then spread onto each selection or control plate. For sorting of Mitotracker red-stained cells, haploid segregants selected on YNB + canavanine/thialysine were scraped from plates and inoculated into YNB + canavanine liquid medium at a concentration of around $\sim 3 \times 10^6$ cells ml^{-1}. The cells were grown for approximately three generations to a density of $\sim 2 \times 10^7$ cells ml^{-1}.

Drug selections with segregating pools. X-QTL should be most powerful when selections are stringent, as this implies that one is enriching for segregants that are phenotypically extreme and are likely to possess multiple alleles that affect a trait in the same direction. For cell sorting, such selections are straightforward, as individual cells exhibiting a trait value within a specified range can be isolated. For chemical resistance mapping, achieving a stringent selection is slightly more challenging, as a whole population of cells is plated and one can only enrich for segregants with high trait values.

Drug selections with segregating pools require finding the optimal concentration to use for a particular compound before X-QTL mapping. To do this, we plate segregating pools across a range of concentrations. The concentration at which we start to resolve individual colonies on plates is the concentration that we use for X-QTL mapping. The fact that we observe $\sim 5 \times 10^2$ to $\sim 5 \times 10^3$ individual colonies when we plate more than 10^6 individuals implies that we are selecting far into the resistance tail of the phenotype distribution. Final concentrations used for the chemical selections are in Supplementary Table 1. After selection was completed, several replicate selection plates were scraped, pooled and frozen at −80 °C. Control experiments were also conducted by plating segregating pools on YPD without any drug added, and pools were collected and stored in the same manner as the selections.

We attempted to combine MAT**a** selections with our chemical resistance selections by incorporating a chemical of interest into our YNB + canavanaine/thialysine plates. We found that this approach worked far worse than separating the selection of MAT**a** haploids and the selection of resistant segregants into two steps.

Microarray description. We designed our array using 21,994 BY and RM allele-specific probe pairs. These pairs cover 17,566 SNPs that differentiate BY and RM, at an average spacing of one marker every ~700 bp. The BY-specific probes were designed as part of a separate study of optimal probe design parameters for DNA genotyping arrays and were chosen to minimize the variance in T_m values across probes[14]. For this study, we used the previously designed BY-specific probes and made an additional probe specific to the RM sequence. To maximize the sensitivity of our genotyping array, probes were chosen to have the interrogated SNP within the middle five bases of a given probe. Our custom two-colour microarray was manufactured by Agilent.

DNA extraction, labelling and microarray hybridization. DNA was extracted from parental strains and segregating pools using Genomic-tip 100/G columns (Qiagen; 10243). DNA was labelled using the BioPrime Array CGH Genomic Labeling Module (Invitrogen; 18095-012) with the sample being labelled with Cy3 dUTP and the reference being labelled with Cy5 dUTP in most cases. We used a BY/RM diploid as the reference for all hybridizations. Hybridization intensities were extracted and normalized using the rank invariant method in the Agilent Feature Extraction software package.

Comparison of microarray data to sequencing data. DNA from the same control and 4-NQO-selected segregating pools was hybridized to the microarray and sequenced on the Illumina Genome Analyzer using 75-bp reads. Two biological replicate control and two biological replicate 4-NQO-selected pools were sequenced. Except for one of the replicate controls that was sequenced in a single lane, each sample was sequenced in two lanes. To analyse the Genome Analyser data, sequencing reads were mapped to the BY genome using ELAND and the Illumina EXPORT files were converted into SAM format using SAMTOOLS[29]. The PILEUP function in SAMTOOLS was used to reformat the sequence data. Sequence data at polymorphic sites included on the genotyping microarray were extracted from the PILEUP file and only these sites were analysed. The polymorphic sites were subjected to a quality filter, with only sites having a quality score of 10 or higher used. The coverage was ~60× per site in each lane. Supplementary Fig. 3 shows only one lane (~60×) of sequence data from a 4-NQO selection, four aggregated lanes of sequence data (~240×) from both 4-NQO selections, and a single microarray. Even at 60× sequencing coverage, peaks are discernible, although the variance in measured allele frequencies is high. 240× coverage provides results comparable to the genotyping microarray. Our results suggest that both X-QTL and X-QTL-seq are useful approaches to genetic mapping in pools of cross progeny.

Mapping results for drug traits. Before analysis, each array was subjected to a quality check that both allele-specific probes for a given probe set had successfully hybridized. Bad probe sets were excluded from downstream analyses. We conducted separate analyses for the drug selection and FACS-based selection experiments.

The difference in the $\log_{10}$ ratios of the intensities of the BY and RM allele-specific probes on a single array was computed for a given SNP, and this metric was used in downstream analyses. In cases where a SNP was represented by two probe sets, the probe sets were used as separate data points. For the drug selections, *t*-tests were conducted comparing results from two independent selection experiments to results from 13 independent control experiments. *t*-tests were conducted with the variances of the two groups set to be equal. The $-\log_{10}(P)$ values were then used for unsupervised peak calling. We found that an approach that scanned the genome for inflection points in the slope of the average $-\log_{10}(P)$ values worked best. By definition, a peak is a point at which the slope of the data changes sign. When scanning $-\log_{10}(P)$ values, which are always positive, a peak is represented by a positive to negative sign change.

To identify inflection points, we first smoothed the data by averaging the $-\log_{10}(P)$ values within 50-kb sliding windows. We then scanned the genome chromosome-by-chromosome by resistance trait using sliding window linear regression. We fit linear regressions over 100-kb sliding windows and used the slope of these regressions to estimate the locations of peaks. A special case was allowed at the ends of chromosomes in which peaks were recorded if the slope was negative at the top of the chromosome or positive at the bottom of the chromosome. The average $-\log_{10}(P)$ value at estimated peaks was recorded and used for thresholding. The same approach was used to analyse 1,000 permutations of the chemical resistance data set, in which two randomly chosen arrays ('selections') were compared to 13 randomly chosen arrays ('controls'). A requirement was set on the permuted data sets that the selection arrays never be biological replicates of the same real trait selection. Because of uncertainty about what constitutes a distinct peak under cases of close linkage, we set a requirement that two peaks could not occur within 200 kb of each other. Increasing or decreasing this proximity threshold results in a slightly different number of called peaks, but does not affect the general findings of the paper. Inflection points detected in the permutations were used to set an empirical FDR threshold of 0.05. We used a global FDR threshold, as opposed to a trait-level FDR, as most observed expected–observed peak relationships at the trait level were very close to the global relationship (Fig. 3d). Average $-\log_{10}(P)$ plots, as well as significant peaks, are provided for each trait in Supplementary Fig. 5a–q.

For the FACS experiment, three low, three high and three whole-population biological replicates were generated. Because of the small number of arrays in the experiment, permutations were unlikely to be useful for setting an empirical FDR threshold. Furthermore, because the data structure of the FACS experiment, which used two tails of the segregant distribution, was different from the drug selections, which used only one tail of the segregant distribution, we could not use the drug selections in permutations of the FACS data. For these reasons, we

©2010 Macmillan Publishers Limited. All rights reserved

used QVALUE[28], which estimates the FDR using the distribution of *P* values in an experiment, to determine probes that were significant at an FDR of 0.05. We show this threshold in Fig. 4.

All analyses were conducted in R.

29. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25,** 2078–2079 (2009).

©2010 **Macmillan Publishers Limited. All rights reserved**