Heritability and genetic basis of protein level variation in an outbred population

Leopold Parts,[1,4] Yi-Chun Liu,[1] Manu M. Tekkedil,[2] Lars M. Steinmetz,[2,3] Amy A. Caudy,[1,4] Andrew G. Fraser,[1,4] Charles Boone,[1,4] Brenda J. Andrews,[1,4] and Adam P. Rosebrock[1]

[1]Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, Toronto, M5S3E1, Canada; [2]European Molecular Biology Laboratory (EMBL), Genome Biology Unit, 69117 Heidelberg, Germany; [3]Department of Genetics, Stanford University School of Medicine, Stanford, California 94305, USA; [4]Department of Molecular Genetics, University of Toronto, Toronto, M5S3E1, Canada

The genetic basis of heritable traits has been studied for decades. Although recent mapping efforts have elucidated genetic determinants of transcript levels, mapping of protein abundance has lagged. Here, we analyze levels of 4084 GFP-tagged yeast proteins in the progeny of a cross between a laboratory and a wild strain using flow cytometry and high-content microscopy. The genotype of *trans* variants contributed little to protein level variation between individual cells but explained >50% of the variance in the population's average protein abundance for half of the GFP fusions tested. To map *trans*-acting factors responsible, we performed flow sorting and bulk segregant analysis of 25 proteins, finding a median of five protein quantitative trait loci (pQTLs) per GFP fusion. Further, we find that *cis*–acting variants predominate; the genotype of a gene and its surrounding region had a large effect on protein level six times more frequently than the rest of the genome combined. We present evidence for both shared and independent genetic control of transcript and protein abundance: More than half of the expression QTLs (eQTLs) contribute to changes in protein levels of regulated genes, but several pQTLs do not affect their cognate transcript levels. Allele replacements of genes known to underlie *trans* eQTL hotspots confirmed the correlation of effects on mRNA and protein levels. This study represents the first genome-scale measurement of genetic contribution to protein levels in single cells and populations, identifies more than a hundred *trans* pQTLs, and validates the propagation of effects associated with transcript variation to protein abundance.

[Supplemental material is available for this article.]

Efficient methods for genotyping large numbers of individuals have greatly improved our ability to map a heritable trait to the genome. While genome-wide association studies have identified more than 1200 genetic variants for nearly 170 complex traits in humans, these findings have not yet translated into substantially improved disease onset prediction or new molecular targets for prevention and cure (Visscher et al. 2012). Analyses of model organism crosses provide a tractable context in which to understand complex trait genetics. A particularly fruitful trait-mapping approach has been measurement of mRNA and protein expression levels as an intermediate for terminal phenotypes (Hubner et al. 2005; Emilsson et al. 2008; Chakravarti et al. 2013).

We and others have probed genetic regulation of RNA levels to considerable depth (Grundberg et al. 2012; Parts et al. 2012). In humans, the fraction of total mRNA level variation that can be attributed to additive effects of independent loci (narrow sense heritability) ranges between 0.15 and 0.35. Narrow sense heritability is even higher in yeast, prompting linkage and association studies to map the responsible variants (Gaffney 2013). Genotype around the promoter (*cis*) region influences transcript levels of many, if not most, genes in yeast and man (Brem et al. 2002; Stranger et al. 2012), and expression quantitative trait loci (eQTLs) affecting several genes at a distance (*trans* eQTLs) have been mapped in a variety of model organism crosses (Yvert et al. 2003; Mehrabian et al. 2005; Keurentjes et al. 2007). To date, eQTL mapping studies have made summary measurements from large populations of cells, generating phenotypes that represent a population average.

Tools for systematically measuring the proteome have lagged relative to methods for assaying the transcriptome. Efforts in yeast and other systems have begun to provide important insights about the relationship between mRNA and protein levels and their response to genetic and environmental perturbations. Mass spectrometry of protein fragments has evolved from quantification of a limited number of peptides (Foss et al. 2007; Lu et al. 2007; de Godoy et al. 2008; Ramakrishnan et al. 2009; Ghazalpour et al. 2011) to thousands of proteins (Marguerat et al. 2012; Picotti et al. 2013; Skelly et al. 2013; Wu et al. 2013). These advances have enabled mapping of alleles that contribute to the variation in protein abundance in yeast (Foss et al. 2007, 2011; Skelly et al. 2013), mouse (Ghazalpour et al. 2011; Holdt et al. 2013), and human (Johansson et al. 2013; Wu et al. 2013). Due to limited statistical power and inherently low-throughput measurement, mass spectrometry-based studies have mainly focused on the effects of *cis* variants and established that about half of *cis* alleles associated with a protein level also affect transcript abundance (Ghazalpour et al. 2011; Skelly et al. 2013; Wu et al. 2013). At present, mass spectrometric measurements are necessarily performed on populations of many cells.

Corresponding authors: leopold.parts@utoronto.ca, charlie.boone@utoronto.ca, brenda.andrews@utoronto.ca, adam.rosebrock@utoronto.ca

Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.170506.113.

Indirect measurement of fusion proteins provides an alternate method for protein level determination. Quantitative genome-scale experiments in yeast have been enabled by the availability of collections of strains comprised of individual open reading frames (ORFs) fused to green fluorescent protein (Huh et al. 2003) or a tandem affinity purification tag (Ghaemmaghami et al. 2003), *ORF-GFP* or *ORF-TAP*, respectively. Previous work on these collections has established the baseline level (Ghaemmaghami et al. 2003), localization (Huh et al. 2003), sources of individual variation (Newman et al. 2006), and response to perturbations (Tkach et al. 2012; Breker et al. 2013; Denervaud et al. 2013; Mazumder et al. 2013) for individual fusion proteins. Despite the availability of these resources, the effect of genotype on protein abundance has not been interrogated on a genome scale (Parts 2014).

Previous studies have demonstrated moderate correlation between protein and RNA levels, emphasizing the need to further explore the extent to which genetic mechanisms regulating protein and mRNA abundance overlap. Many *cis* signals regulate both mRNA and protein levels, but studies in outbred individuals have lacked power to detect associations with *trans* loci. To date, comprehensive dissection of protein level heritability and validation of *trans* eQTL signal propagation to protein abundance has been unexplored. Here we use a combination of flow cytometry and quantitative microscopy to establish the extent of genetic contribution of *trans* variants to protein levels in single cells and populations, map several responsible loci, and compare the genetic determinants of mRNA and protein level heritability.

Results

We quantified the extent to which protein levels differ between a near-clonal population of inbred and a genetically diverse population of outbred progeny in a model yeast cross for each of the 4084 *ORF-GFP* fusion strains in the yeast GFP collection (Huh et al. 2003). We used the synthetic genetic array (SGA) method (Tong and Boone 2006) to mate each member of the *ORF-GFP* collection to both an isogenic lab strain ("BY") and a divergent wild isolate, RM11 ("RM") (genotypes in Supplemental Table 1). We produced pools of ~120,000 haploid BY × BY and RM × BY meiotic progeny for each GFP-tagged ORF (Fig. 1A; Supplemental Methods). We marked BY and RM parents and haploid progeny with mCherry^{wt} and mCherry*, respectively, to permit unambiguous separation of parental genotype by red fluorescence intensity (Fig. 1C; Methods). We mixed the corresponding inbred (BY × BY) and outbred (RM × BY) segregant populations and grew them in the same liquid culture to control for environmental effects. At mid-log phase, we measured the fluorescence intensities of 20,000–50,000 individuals using high-throughput flow cytometry and imaged 200–500 cells from the same culture using high-content microscopy (Methods; Fig. 1B–D; Supplemental Fig. 1; Supplemental Table 2). We classified individual cells into inbred BY × BY and outbred RM × BY populations based on their mCherry level and quantified GFP fluorescence. We sequenced total RNA from inbred and outbred pools and found moderate correlation of mRNA and protein levels, consistent with previous work (Pearson's r = 0.56) (Methods; Supplemental Table 2; Supplemental Fig. 4; Marguerat et al. 2012).

Extent of *trans* variant influence on protein abundance

The *ORF-GFP* locus is fixed to the BY background in our crossing and selection design, thus all the differences between the inbred and outbred GFP levels are expected to be due to *trans* effects. We find that *trans* variants have a large effect on a subset of protein abundances in the BY × RM cross: 6% of fusion proteins (197/3251) differ by 20% or more between inbred and outbred progeny (Fig. 2A,B). Examples include GFP fusions to Hxt2 (hexose transporter) (Fig. 1B–D), Pdr5 (pleiotropic ABC transporter), and Mdh1 (malate dehydrogenase) (Fig. 2A). We find functional enrichment among proteins whose abundance varies in the cross, consistent with functions of known segregating regulators (Mense and Zhang

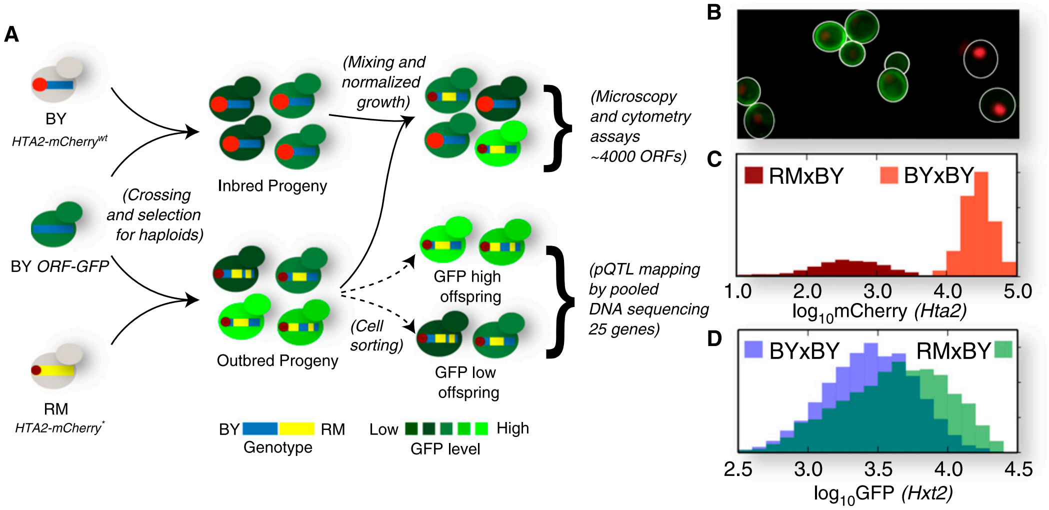

Figure 1. Overview of the experiment. (*A*) Experimental design. The GFP collection strains are crossed to near-isogenic (BY) and genetically diverged (RM) parental strains, the progeny measured using two assays, and sorted for pQTL mapping. (*B*) Section of an example microscopy image. Hxt2-GFP signal is shown in green and nuclear Hta2-mCherry signal in red; white outlines designate the cell boundaries. (*C*) Classifying cells into BY × BY inbred progeny and RM × BY segregants. The mCherry fluorescence (*x*-axis, $\log_{10}$ scale) of the RM parent (mCherry*, dark red) is on average ~100-fold lower than that of the BY parent (mCherry wild type, bright red), and is used to classify cells into RM × BY or BY × BY cross progeny. (*D*) Example cytometry readout. Hxt2-GFP level (*x*-axis, $\log_{10}$ scale) is measured for tens of thousands of individual cells from the inbred BY × BY population (blue) and segregating RM × BY population (green), and used to calculate the summary statistics of the Hxt2 protein level.

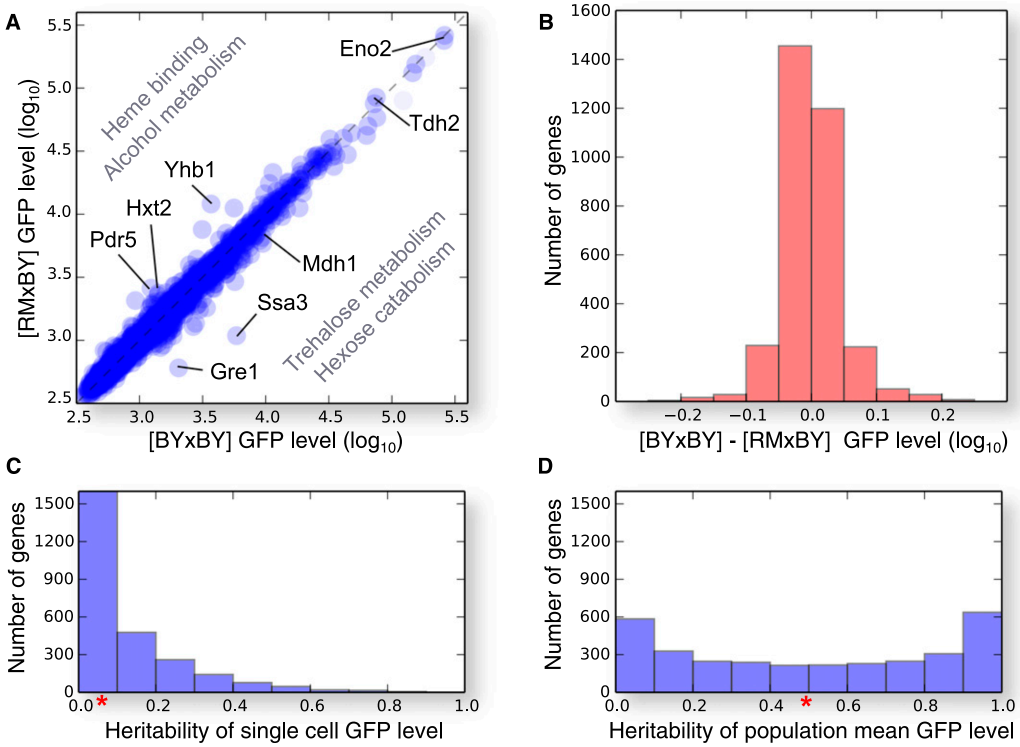

Figure 2. Genetic contribution to protein levels via *trans* variants. (*A*) Protein levels are similar in the inbred BY × BY and segregating RM × BY populations. Average protein level for the inbred BY × BY individuals (*x*-axis, $\log_{10}$ scale) and BY × RM segregants (*y*-axis, $\log_{10}$ scale) for 3173 genes (blue dots). Individual examples of genes with a large difference between inbred and outbred progeny, and genes discussed in the text are highlighted. (*B*) The distribution of differences in GFP level between inbred and outbred populations is tightly centered on zero. The effect of 0.1 on the $\log_{10}$ scale corresponds approximately to a difference of 25%. (*C*) Single-cell protein level heritabilities are low. Distribution of broad sense heritability (*x*-axis), calculated as the fraction of total variation explained by genotype in the pooled population of inbred BY × BY and outbred RM × BY segregants, assuming 50% contribution from both. Median heritability across genes is marked with a red asterisk. (*D*) Population average protein level heritabilities are high. Distribution of broad sense heritability (*x*-axis), calculated as the average squared difference of mean protein level between inbred BY × BY individuals and RM × BY segregants and the total variation in average protein level across replicates and populations. Median heritability across genes is marked with a red asterisk.

2006; Smith and Kruglyak 2008), enrichments of transcript level changes, and the different evolutionary backgrounds of the parental strains (Supplemental Table 3; Supplemental Results; Supplemental Methods; Mortimer et al. 1994; Warringer et al. 2011).

Genotype accounts for at least 50% of the variation in population mean protein abundance for 49% of the traits we measured (Fig. 2D; Supplemental Table 4; Methods). The magnitude of cell-to-cell variation in the measured populations was, however, much larger than a typical genetic signal (Fig. 1C,D; Supplemental Fig. 3). Thus, the distribution of the fraction of GFP level variation between single cells that could be attributed to genotype was skewed toward zero (median $H^2 = 0.03$) (Fig. 2C; Methods). We further quantified the maternal contribution to single-cell protein level variation by microscopy and found that mother and daughter cells have more similar protein levels compared to random pairs for 90% of the measured genes (Supplemental Fig. 5; Supplemental Table 2; Methods). This indicates that the large observed variation in protein abundance between individual cells is not completely stochastic in nature.

As we selected the *ORF-GFP* allele from the BY parent in our experiments, virtually no RM *cis* variants are present in outbred progeny. To explore the role of *cis* effects, we chose 52 ORFs based on our measurements of protein level heritability and number of previously described eQTLs (Smith and Kruglyak 2008) and constructed *ORF-GFP* fusions in the RM background. We performed a reciprocal experiment in which the RM *ORF-GFP* allele, and thus RM *cis* sequence, was selected (Methods) and found 26 genes (50%) for which the *cis* sequence made at least a 20% difference in protein abundance (Supplemental Table 5). This represents a significant ($p < 10^{-12}$, Fisher's exact test) eightfold increase over the 6% expected if *cis* and *trans* sequence had the same independent distribution of effect sizes. We further find that *cis* effects are greater in magnitude than *trans* effects for 42% (22 of 52) of the genes. From our data, we conclude that association studies which focus on the coding and *cis*-regulatory regions (due to statistical power considerations) are likely capturing a large fraction of genetic effects on protein expression, in concordance with previous reports (Keurentjes et al. 2007; Ghazalpour et al. 2011).

Mapping protein level QTLs

Although *cis* sequences strongly regulate protein abundance for many genes, our genome-wide analysis demonstrated the presence of *trans* effects. To map the responsible loci, we performed bulk segregant analysis (Ehrenreich et al. 2010; Parts et al. 2011; Cubillos et al. 2013) for genetic variation in protein levels (Methods). We used flow sorting to isolate 5% fractions of the brightest (GFP high) and dimmest (GFP low) cells of the outbred BY × RM progeny for crosses of 25 *ORF-GFP* fusions. Population average GFP levels of sorted fractions were stable for at least 10 generations (Supplemental Fig. 6). We compared parental contributions by measuring variant allele frequencies between the sorted populations by microarray genotyping and Illumina sequencing. In total, we found 156 genomic regions in which the RM allele frequency was at least 20% different between the GFP low and high populations and defined these as protein level quantitative trait loci (pQTLs; false discovery rate = 22%) (Supplemental Table 7; Supplemental Fig. 7; Methods). For a subset of traits, we identified nonsynonymous variants within the mapped region (Fig. 3B) and identified the causal gene by allele replacement and reciprocal hemizygosity test (Fig. 3C; Supplemental Results). The number of pQTLs per *ORF-GFP* varied from zero to 22, with a median of five. While technical factors contribute to the detection power in an individual cross, the number of loci we have mapped indicates that regulation of protein levels is not genetically simple.

Many of our pQTLs mapped to previously identified eQTLs (Fig. 3A; Smith and Kruglyak 2008). Eleven *trans* eQTL hotspot regions segregated in our experiment and regulated at least 30 transcripts; eight of them also had at least four pQTLs at a relaxed cutoff (1.0 expected) (Methods). The three eQTL-rich regions that did not harbor multiple pQTLs were located on chromosomes IX and XVI, and a chromosome region XV that includes the *IRA2* gene, discussed below.

Concordance of genotype effects on mRNA and protein level

The genetic determinants of mRNA levels likely also affect corresponding protein abundance. To test whether this is the case for our mapped varying traits, we compared the previously identified eQTLs (Smith and Kruglyak 2008) to our pQTLs. We calculated the eQTL effect size from a multivariate regression model that includes all significant eQTLs (Methods; Smith and Kruglyak 2008) and used the difference in RM and BY allele frequency at the pQTL locus as the pQTL effect size. There were 19 significant eQTLs (log-odds score > 5) for our 25 interrogated *ORF-GFP* genes, 12 of which had a concordant pQTL of effect at least 0.1 (Fig. 4A; Methods). We also found several strong pQTLs that did not show an eQTL signal (Supplemental Fig. 8). For example, of the 10 strongest pQTLs, three mapped to a 280-kb locus on chromosome XI, with the RM allele linked to high levels of Tdh2-GFP, Eno2-GFP, and Sed1-GFP proteins. Post-transcriptional regulation of glycolytic gene products, including Tdh2 and Eno2, has been reported previously (Bruckmann et al. 2007), suggesting a mechanism regulating central carbon metabolic protein levels independent of transcript levels. It is possible that the same process underlies some of the other large heritable changes we found for other proteins involved in glycolysis (Supplemental Table 2). Our genome-wide analysis and subsequent pQTL mapping suggest functionally coherent regulation of protein abundance above regulation of mRNA levels.

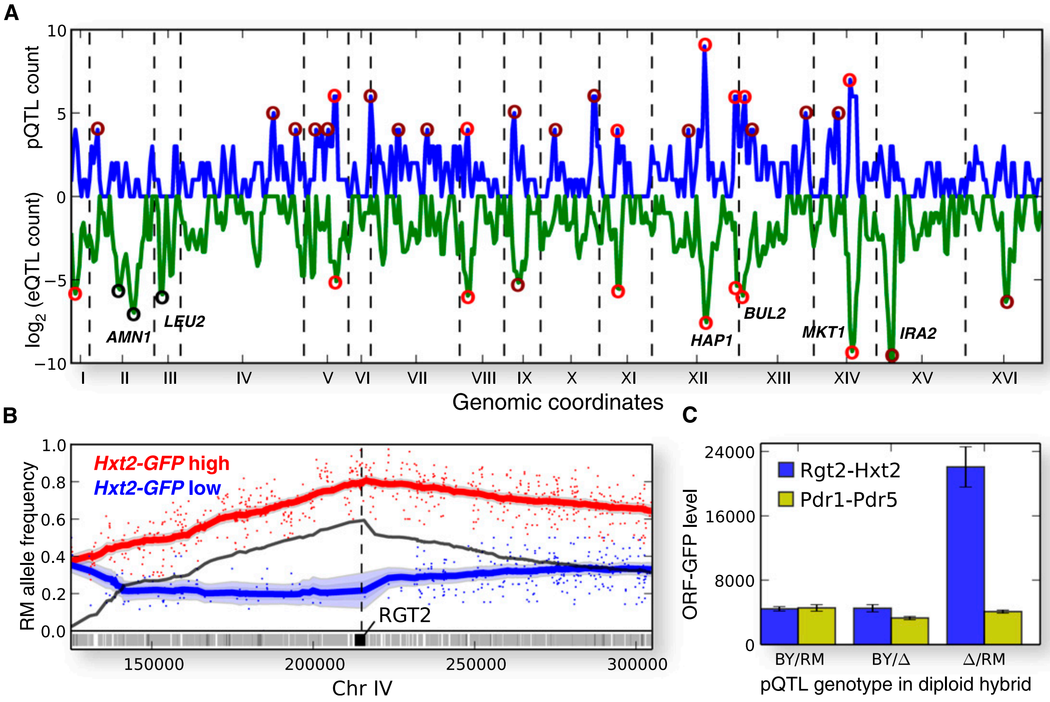

Figure 3. Protein level QTLs. (*A*) Genome-wide density of pQTLs and eQTLs. Number of pQTLs (*y*-axis, linear scale, positive), and eQTLs (*y*-axis, log$_2$ scale, negative) in 50% overlapping 25-kb bins along the genome (*x*-axis). eQTL peaks with at least 30 linkages, and pQTL peaks with at least four pQTLs are highlighted with either bright red markers (peak present for both pQTLs and eQTLs), dark red markers (peak present for pQTLs or eQTLs, but not both), or black markers (eQTL peaks due to alleles not segregating in our cross). Known QTL hotspots are marked with the causal gene. (*B*) Chromosome IV QTL for Hxt2-GFP abundance overlapping the *RGT2* locus. The allele frequencies (*y*-axis) in the GFP high pool (red) and GFP low pool (blue) along a 200-kb region of chromosome IV (*x*-axis). Each dot corresponds to one segregating site between the BY and RM parents, with its *y*-coordinate equal to the fraction of sequencing reads with the RM allele mapped to the site. Solid lines correspond to window-averaging of the dots and approximate the underlying population allele frequency, with the 5% confidence intervals given by shaded regions. The estimated difference between the RM allele frequencies in the GFP high and low pools is overlaid with a solid black line, and the local allele frequency difference maximum with a dashed black vertical line. Yeast genes in the region are displayed *below* the allele frequencies as rectangles, with the *RGT2* gene highlighted. (*C*) $RGT2^{RM}$ (blue) and $PDR1^{RM}$ (yellow) are causal alleles for Hxt2 and Pdr5 protein level differences, respectively, by reciprocal hemizygosity. Two independent clones were measured in triplicate for the BY × RM hybrid (*left*), the same hybrid with the BY allele deleted (*middle*), and the RM allele deleted (*right*).

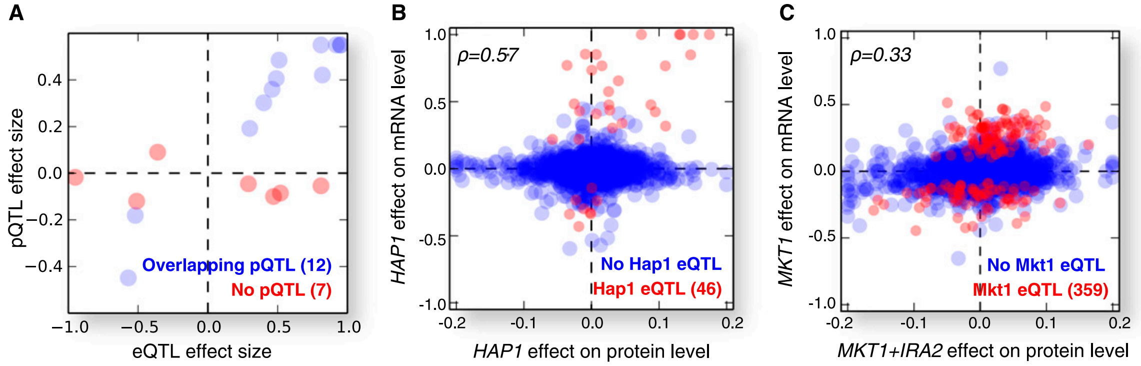

Figure 4. Concordance of genotype effects on mRNA and protein level. (*A*) eQTL and pQTL effect directions and magnitudes are concordant. Difference between mRNA level in the segregants with the BY and RM allele at the eQTL locus (*x*-axis) is concordant with the difference between the RM allele frequencies between the corresponding ORF-GFP high and low pools at the eQTL locus (*y*-axis) for 12 eQTLs (blue dots), and not detectable for the remaining seven (red dots). (*B*) Comparison of allele effect sizes for *HAP1* locus. Change in GFP level in response to substituting the *HAP1*^{RM} allele into the BY background (*x*-axis) is correlated with the difference between average mRNA levels of segregants with the *HAP1*^{RM} and *HAP1*^{BY} alleles (*y*-axis) (Smith and Kruglyak 2008) for genes with an mRNA level linkage to the *HAP1* locus (red markers). Changes in all other genes (blue markers) are shown as a reference. (*C*) As in *B*, but for both *MKT1*^{RM} and *IRA2*^{RM} alleles introduced to the BY background and genes with *MTK1* mRNA linkages highlighted in red. See also Supplemental Figure 9.

Multiple pleiotropic regulators of mRNA levels have been mapped in the BY × RM cross (Brem and Kruglyak 2005; Smith and Kruglyak 2008), with *HAP1*, *MKT1*, and *IRA2* loci showing the most profound effects (Fig. 3A). We examined the contribution of known regulators of mRNA abundance on protein level by building RM alleles of these three genes into the BY parent background in series and repeated the cross to generate BY × BY$^{\mathrm{RMregulator}}$ progeny in which these regulators were the only segregating loci. We then compared the effect of the allele swaps on protein levels to the effect of corresponding locus genotypes on mRNA abundance. We found high concordance between effects of the *HAP1* locus genotype on GFP and mRNA levels (Spearman rho = 0.57) and moderate concordance between the *MKT1* locus effects (Spearman rho = 0.33) (Fig. 4B,C; Supplemental Table 8; Methods). The *IRA2* locus also affected both mRNA and protein levels, but its effects on the protein levels in our study were anticorrelated with those from the previous mRNA results (Spearman rho = –0.49) (Supplemental Fig. 9). Overall, these three known major effect loci partially explain nearly half the genes with >20% GFP level change (84/197, 43%), with the remaining 113 genes not strongly enriched in specific GO categories ($p > 10^{-3}$) (Supplemental Table 3). Our results confirm indirect control of protein abundance via mRNA abundance, and further identify nearly 120 heritable protein traits that cannot be accounted for by these major effect eQTLs.

Discussion

The ability to quantify how genetic signals are propagated from genome to phenotype has been limited by the lack of sensitive high-throughput assays of intermediate phenotypes. We have overcome this limitation for the examination of protein levels by use of large-scale crosses and high-throughput phenotyping of single cells. We have shown that genotype-dependent changes in mRNA levels, the immediate product of DNA, are to a large extent carried through to respective protein abundances. As anticipated but not previously demonstrated (Foss et al. 2007, 2011; Wu et al. 2013), the effects of the loci that contribute to the differences are shared across both expression and protein level phenotypes. A recent pQTL mapping study of the BY × RM cross found that the majority of eQTLs have a corresponding pQTL, and that *cis* pQTLs exist for about half of the tested genes (Albert et al. 2014). This result is complementary to the present work, as only three protein traits overlapped between studies. Given the broad similarities in *cis* genetic control of mRNA and protein levels (Skelly et al. 2013; Wu et al. 2013; Albert et al. 2014), it is reasonable to assume that a large fraction of described mRNA level regulators in humans, including variants associated with common diseases, affect the levels of corresponding proteins (Nica et al. 2010).

While we demonstrated that *cis* signals on mRNA level alone are not sufficient to explain the heritable variation in genetically complex protein levels, both our data and previous studies (Keurentjes et al. 2007; Ghazalpour et al. 2011; Albert et al. 2014) suggest that the effects of *trans* variants on the proteome are generally smaller, perhaps due to purifying selection acting on alleles that have a large effect on multiple transcripts or proteins (Battle et al. 2014). The extent to which common genetic variation in *cis* and *trans* contributes to translation, protein degradation, mRNA turnover, and other important regulatory layers remains to be precisely quantified, but we have shown that several pleiotropic regulators of mRNA abundance influence levels of the corresponding proteins through one of these mechanisms.

Gene expression heterogeneity within a population can be due to the influence of cell cycle position, stochastic expression, epigenetic effects, or other factors. Such variation, sometimes termed noise, has been demonstrated to have a complex genetic basis for some traits (Ansel et al. 2008), and postulated to underlie bet-hedging strategies (Levy et al. 2012). Genetic influences of *trans* variants that we have focused on could be reflected in increasing or suppressing variation in the inbred or segregating population without affecting the mean expression. However, as the inbred and segregant population GFP level variances do not show large differences (Supplemental Fig. 3C) and the heritability calculations take these variances into account, our data do not provide strong evidence toward genetic control of gene expression variance independently of the average abundance. Substantial influence of genetic interactions has been found for mRNAs (Brem et al. 2005; Smith and Kruglyak 2008) and other traits (Gerke et al. 2009; Costanzo et al. 2010), and we expect to uncover interactions between loci in *trans* as well as between the genotype of the promoter and distal regulators in future experiments.

Methods

Strain construction and crossing

Starting strains were generated from the BY4742 (BY) or LK1552 (RM11-A) strain (Supplemental Table 1). Query *MAT*α strains carrying the *HTA2-mCherry:URA3MX* allele were crossed to the GFP array *MAT*a strains using a modified synthetic genetic array procedure (Tong and Boone 2006) to create large pools of inbred (BY query × BY GFP array) and segregating outbred (RM query × BY GFP array) haploid individuals. We developed and used a low-fluorescent mCherry* allele (Y198C) in the RM parent to distinguish progeny of inbred BY × BY and segregating RM × BY crosses based on red fluorescence signal (Fig. 1C). To quantify *cis* effects, we generated 52 ORFs with GFP fusion strains in the RM background (RM^G) parallel to the BY collection (BY^G). We crossed the BY and RM query strains to the GFP fusion strains in both backgrounds, yielding BY × BY^G, BY × RM^G, RM × BY^G, and RM × RM^G crosses. Additional details on strain construction and production of diverse haploid populations are available in Supplemental Methods.

Cytometry

Inbred BY × BY and segregating RM × BY populations were pooled in the liquid culture; *cis* effect screens and reciprocal hemizygosity confirmation were grown separately. For each culture, measurements from up to 50,000 cells were collected using an LSR II flow cytometer (BD Biosciences) following sonication. Scatter and mCherry measurements were used to filter the data and cluster the cells into budded vs. unbudded and mCherry high (mCherry wild type, BY × BY cross) vs. mCherry low (mCherry*, RM × BY cross) classes (schematized in Supplemental Fig. 1). Unbudded (side-scatter low) populations were used to calculate GFP abundance summary statistics (mean, median, variance, count) both on raw data, and $\log_{10}$-transformed data on all the cells, as well as the middle 10% of the cells centered on the median cell size (FSC-A), followed by normalization across replicates (Supplemental Tables 2, 5). To quantify the *cis* effect of the RM allele, we calculated the difference between the BY × RM^G and RM × BY^G normalized GFP level. Methodological details, including relevant instrument configurations and data preprocessing, are in Supplemental Methods.

Microscopy

We imaged the cell cultures on the Opera High Content Screening Platform (PerkinElmer) using Dextran Alexa647 fluor (Molecular Probes) in the media to distinguish cells from the background. The images were processed with CellProfiler (Stoter et al. 2013) (pipeline in Supplemental Data set 3). Segmented cells were classified into inbred BY × BY and outbred RM × BY populations according to the mCherry signal, and GFP abundance statistics were calculated from the populations using average cellular pixel GFP intensity as a readout. Details on image acquisition and processing are available in Supplemental Methods.

Heritability estimation

For single-cell levels, we assumed the entire considered population to consist of 50% RM × BY segregants and 50% of inbred BY × BY individuals, which produces an estimate for the total variance in the population σ^2 as $0.5(\sigma^2_{\text{inbred}} + \sigma^2_{\text{segregants}} + 0.5\Delta^2)$, where the σ^2_{inbred} and $\sigma^2_{\text{segregants}}$ are calculated directly from cytometry data, and Δ is the difference between the corresponding means. We then computed the heritability as $\max(0, 1 - \sigma^2_{\text{inbred}}/\sigma^2)$. The assumptions of the population composition and log-transformation of the data affect the distribution of heritabilities but not the qualitative conclusions drawn (Supplemental Fig. 3D,F). For population average level heritability, we estimated the variances as average squared errors of the population average measurements across replicate experiments (Supplemental Fig. 3E,G). Additional details on heritability calculations are in Supplemental Methods and Supplemental Figure 3H–K.

Cell sorting

We chose 25 ORFs with large GFP level differences between inbred BY × BY and RM × BY segregants in our whole-genome screen and used flow sorting (FACSAria, BD Biosciences) to isolate 20,000 cells in the top 5% ("GFP high") and bottom 5%–10% of the GFP levels ("GFP low") for two replicates of each of the corresponding ORF^{BY}*-GFP* segregant pools. Sorted cells were grown as single colonies, pooled, and re-measured by flow cytometry to confirm that the selected difference in average GFP level was inherited (Supplemental Fig. 6). Culture pregrowth and sorting setup is described in Supplemental Methods.

Sequencing and microarrays

DNA was extracted from the sorted pools, following Nextera XT library preparation, and sequencing was done using 100-bp paired-end reads on the Illumina platform. RNA was extracted from two replicates each of a pool of 96 BY × BY and RM × BY crosses, followed by Illumina sequencing. For each sequenced DNA sample from a sorted pool, the posterior distribution of the RM allele frequency at each locus was calculated as described (Parts et al. 2011). ORF mRNA abundance was calculated as the average read coverage of the ORF. Microarray measurements were performed with Agilent SurePrint oligonucleotide microarrays with 60-nt probes (Agilent Technologies), and the difference between the BY and RM probe intensities for each sample at each locus was used as the genotype signal. Details of sequencing library preparation, array design and hybridization, and data processing are provided in Supplemental Methods.

pQTL calling

pQTLs were called as regions for which the posterior allele frequency difference between the GFP very high and low populations was at least 20%, and at least five standard deviations, and the pQTL peak center was picked as the site with the highest combined allele frequency difference. The frequency of expected false positives was calculated as the fraction of all sites in the other replicate that exhibited allele frequency change of at least 10% in the same direction within 500 bp of a randomly picked QTL peak center. Comparison of pQTLs and eQTLs is described in Supplemental Methods.

Analyzing effects of alternate alleles of transcriptional regulators

Allele replacements of $HAP1^{RM}$, $MKT1^{RM}$, and $IRA2^{RM}$ were performed in the BY background. The population average GFP level was measured using a Tecan Infinite M1000 fluorometer (Supplemental Table 8). Full details of strain construction and analysis are in Supplemental Methods.

Data access

The sequencing data from this study have been submitted to the European Nucleotide Archive (ENA; http://www.ebi.ac.uk/ena/)

under accession number PRJEB5268. The microarray data are available in the ArrayExpress database (www.ebi.ac.uk/arrayexpress) under accession number E-MTAB-2269. The data and source code used for analysis are available in the Supplemental Material.

Acknowledgments

We thank Jared Simpson, Gianni Liti, Jüri Reimand, Oliver Stegle, and Anton Khmelinskii for comments on the manuscript, Frederick P. Roth, Joseph Mellor, and Atina Cote for help with sequencing library preparation, Dax Torti and Danica Leung for DNA sequencing, Ji-Young Youn and EMBL GeneCore for RNA sequencing help, Dionne White for performing the cell sorting, Erin Smith for discussing the *IRA2* genotype, Helena Friesen for help with designing and carrying out experiments, and Michael J. Cox for critical feedback and help with design and carrying out of experiments. This work was funded by startup funds from the Faculty of Medicine (A.A.C., A.P.R.). The research leading to these results has received funding from the National Institutes of Health and the European Research Council under the European Union's Seventh Framework Programme (FP7/2007-2013/ERC Grant agreement no. AdG-294542 to L.M.S.). L.P. is a Canadian Institute for Advanced Research Global Scholar, and a Marie Curie International Outgoing Fellow. A.A.C. is supported by the Ontario Early Researcher Award, by the Canadian Foundation for Innovation and the Ontario Leader's Opportunity Fund, by the Canadian Institutes for Health Research, and by the Natural Sciences and Engineering Research Council of Canada. L.P., C.B., and B.J.A. were supported by Canadian Institutes of Health Research grants MOP-97939 and MOP-102629, and the Ontario Research Fund (GL2). A.P.R. is supported by a Connaught Innovator Award.

Author contributions: L.P. and A.G.F. conceived the project. L.P., B.J.A., C.B., and A.P.R. designed the experiments. L.P., Y.-C.L., M.M.T., A.A.C., and A.P.R. performed the experiments. L.M.S., A.A.C., C.B., B.J.A., and A.P.R. provided the reagents, materials, and analysis tools. L.P. analyzed the data. C.B., B.J.A., and A.P.R. supervised the project. L.P., C.B., B.J.A., and A.P.R. wrote the manuscript with input from all the authors.

References

Albert FW, Treusch S, Shockley AH, Bloom JS, Kruglyak L. 2014. Genetics of single-cell protein abundance variation in large yeast populations. *Nature* **506:** 494–497.

Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, Francois J, Yvert G. 2008. Cell-to-cell stochastic variation in gene expression is a complex genetic trait. *PLoS Genet* **4:** e1000049.

Battle A, Mostafavi S, Zhu X, Potash JB, Weissman MM, McCormick C, Haudenschild CD, Beckman KB, Shi J, Mei R, et al. 2014. Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922 individuals. *Genome Res* **24:** 14–24.

Breker M, Gymrek M, Schuldiner M. 2013. A novel single-cell screening platform reveals proteome plasticity during yeast stress responses. *J Cell Biol* **200:** 839–850.

Brem RB, Kruglyak L. 2005. The landscape of genetic complexity across 5,700 gene expression traits in yeast. *Proc Natl Acad Sci* **102:** 1572–1577.

Brem RB, Yvert G, Clinton R, Kruglyak L. 2002. Genetic dissection of transcriptional regulation in budding yeast. *Science* **296:** 752–755.

Brem RB, Storey JD, Whittle J, Kruglyak L. 2005. Genetic interactions between polymorphisms that affect gene expression in yeast. *Nature* **436:** 701–703.

Bruckmann A, Hensbergen PJ, Balog CI, Deelder AM, de Steensma HY, van Heusden GP. 2007. Post-transcriptional control of the *Saccharomyces cerevisiae* proteome by 14-3-3 proteins. *J Proteome Res* **6:** 1689–1699.

Chakravarti A, Clark AG, Mootha VK. 2013. Distilling pathophysiology from complex disease genetics. *Cell* **155:** 21–26.

Costanzo M, Baryshnikova A, Bellay J, Kim Y, Spear ED, Sevier CS, Ding H, Koh JL, Toufighi K, Mostafavi S, et al. 2010. The genetic landscape of a cell. *Science* **327:** 425–431.

Cubillos FA, Parts L, Salinas F, Bergstrom A, Scovacricchi E, Zia A, Illingworth CJ, Mustonen V, Ibstedt S, Warringer J, et al. 2013. High-resolution mapping of complex traits with a four-parent advanced intercross yeast population. *Genetics* **195:** 1141–1155.

de Godoy LM, Olsen JV, Cox J, Nielsen ML, Hubner NC, Frohlich F, Walther TC, Mann M. 2008. Comprehensive mass-spectrometry-based proteome quantification of haploid versus diploid yeast. *Nature* **455:** 1251–1254.

Denervaud N, Becker J, Delgado-Gonzalo R, Damay P, Rajkumar AS, Unser M, Shore D, Naef F, Maerkl SJ. 2013. A chemostat array enables the spatio-temporal analysis of the yeast proteome. *Proc Natl Acad Sci* **110:** 15842–15847.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. *Nature* **464:** 1039–1042.

Emilsson V, Thorleifsson G, Zhang B, Leonardson AS, Zink F, Zhu J, Carlson S, Helgason A, Walters GB, Gunnarsdottir S, et al. 2008. Genetics of gene expression and its effect on disease. *Nature* **452:** 423–428.

Foss EJ, Radulovic D, Shaffer SA, Ruderfer DM, Bedalov A, Goodlett DR, Kruglyak L. 2007. Genetic basis of proteome variation in yeast. *Nat Genet* **39:** 1369–1375.

Foss EJ, Radulovic D, Shaffer SA, Goodlett DR, Kruglyak L, Bedalov A. 2011. Genetic variation shapes protein networks mainly through non-transcriptional mechanisms. *PLoS Biol* **9:** e1001144.

Gaffney DJ. 2013. Global properties and functional complexity of human gene regulatory variation. *PLoS Genet* **9:** e1003501.

Gerke J, Lorenz K, Cohen B. 2009. Genetic interactions between transcription factors cause natural variation in yeast. *Science* **323:** 498–501.

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS. 2003. Global analysis of protein expression in yeast. *Nature* **425:** 737–741.

Ghazalpour A, Bennett B, Petyuk VA, Orozco L, Hagopian R, Mungrue IN, Farber CR, Sinsheimer J, Kang HM, Furlotte N, et al. 2011. Comparative analysis of proteome and transcriptome variation in mouse. *PLoS Genet* **7:** e1001393.

Grundberg E, Small KS, Hedman AK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al. 2012. Mapping *cis*- and *trans*-regulatory effects across multiple tissues in twins. *Nat Genet* **44:** 1084–1089.

Holdt LM, von Delft A, Nicolaou A, Baumann S, Kostrzewa M, Thiery J, Teupser D. 2013. Quantitative trait loci mapping of the mouse plasma proteome (pQTL). *Genetics* **193:** 601–608.

Hubner N, Wallace CA, Zimdahl H, Petretto E, Schulz H, Maciver F, Mueller M, Hummel O, Monti J, Zidek V, et al. 2005. Integrated transcriptional profiling and linkage analysis for identification of genes underlying disease. *Nat Genet* **37:** 243–253.

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. *Nature* **425:** 686–691.

Johansson A, Enroth S, Palmblad M, Deelder AM, Bergquist J, Gyllensten U. 2013. Identification of genetic variants influencing the human plasma proteome. *Proc Natl Acad Sci* **110:** 4673–4678.

Keurentjes JJ, Fu J, Terpstra IR, Garcia JM, van den Ackerveken G, Snoek LB, Peeters AJ, Vreugdenhil D, Koornneef M, Jansen RC. 2007. Regulatory network construction in *Arabidopsis* by using genome-wide gene expression quantitative trait loci. *Proc Natl Acad Sci* **104:** 1708–1713.

Levy SF, Ziv N, Siegal ML. 2012. Bet hedging in yeast by heterogeneous, age-correlated expression of a stress protectant. *PLoS Biol* **10:** e1001325.

Lu P, Vogel C, Wang R, Yao X, Marcotte EM. 2007. Absolute protein expression profiling estimates the relative contributions of transcriptional and translational regulation. *Nat Biotechnol* **25:** 117–124.

Marguerat S, Schmidt A, Codlin S, Chen W, Aebersold R, Bahler J. 2012. Quantitative analysis of fission yeast transcriptomes and proteomes in proliferating and quiescent cells. *Cell* **151:** 671–683.

Mazumder A, Pesudo LQ, McRee S, Bathe M, Samson LD. 2013. Genome-wide single-cell-level screen for protein abundance and localization changes in response to DNA damage in *S. cerevisiae*. *Nucleic Acids Res* **41:** 9310–9324.

Mehrabian M, Allayee H, Stockton J, Lum PY, Drake TA, Castellani LW, Suh M, Armour C, Edwards S, Lamb J, et al. 2005. Integrating genotypic and expression data in a segregating mouse population to identify 5-lipoxygenase as a susceptibility gene for obesity and bone traits. *Nat Genet* **37:** 1224–1233.

Mense SM, Zhang L. 2006. Heme: a versatile signaling molecule controlling the activities of diverse regulators ranging from transcription factors to MAP kinases. *Cell Res* **16:** 681–692.

Mortimer RK, Romano P, Suzzi G, Polsinelli M. 1994. Genome renewal: a new phenomenon revealed from a genetic study of 43 strains of *Saccharomyces cerevisiae* derived from natural fermentation of grape musts. *Yeast* **10:** 1543–1552.

Newman JR, Ghaemmaghami S, Ihmels J, Breslow DK, Noble M, DeRisi JL, Weissman JS. 2006. Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* **441:** 840–846.

Nica AC, Montgomery SB, Dimas AS, Stranger BE, Beazley C, Barroso I, Dermitzakis ET. 2010. Candidate causal regulatory effects by integration of expression QTLs with complex trait genetic associations. *PLoS Genet* **6:** e1000895.

Parts L. 2014. Genome-wide mapping of cellular traits using yeast. *Yeast* doi: 10.1002/yea.3010.

Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail MA, et al. 2011. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res* **21:** 1131–1138.

Parts L, Hedman AK, Keildson S, Knights AJ, Abreu-Goodger C, van de Bunt M, Guerra-Assuncao JA, Bartonicek N, van Dongen S, Magi R, et al. 2012. Extent, causes, and consequences of small RNA expression variation in human adipose tissue. *PLoS Genet* **8:** e1002704.

Picotti P, Clement-Ziza M, Lam H, Campbell DS, Schmidt A, Deutsch EW, Rost H, Sun Z, Rinner O, Reiter L, et al. 2013. A complete mass-spectrometric map of the yeast proteome applied to quantitative trait analysis. *Nature* **494:** 266–270.

Ramakrishnan SR, Vogel C, Prince JT, Li Z, Penalva LO, Myers M, Marcotte EM, Miranker DP, Wang R. 2009. Integrating shotgun proteomics and mRNA expression data to improve protein identification. *Bioinformatics* **25:** 1397–1403.

Skelly DA, Merrihew GE, Riffle M, Connelly CF, Kerr EO, Johansson M, Jaschob D, Graczyk B, Shulman NJ, Wakefield J, et al. 2013. Integrative phenomics reveals insight into the structure of phenotypic diversity in budding yeast. *Genome Res* **23:** 1496–1504.

Smith EN, Kruglyak L. 2008. Gene-environment interaction in yeast gene expression. *PLoS Biol* **6:** e83.

Stoter M, Niederlein A, Barsacchi R, Meyenhofer F, Brandl H, Bickle M. 2013. CellProfiler and KNIME: open source tools for high content screening. *Methods Mol Biol* **986:** 105–122.

Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, et al. 2012. Patterns of *cis* regulatory variation in diverse human populations. *PLoS Genet* **8:** e1002639.

Tkach JM, Yimit A, Lee AY, Riffle M, Costanzo M, Jaschob D, Hendry JA, Ou J, Moffat J, Boone C, et al. 2012. Dissecting DNA damage response pathways by analysing protein localization and abundance changes during DNA replication stress. *Nat Cell Biol* **14:** 966–976.

Tong AH, Boone C. 2006. Synthetic genetic array analysis in *Saccharomyces cerevisiae*. *Methods Mol Biol* **313:** 171–192.

Visscher PM, Brown MA, McCarthy MI, Yang J. 2012. Five years of GWAS discovery. *Am J Hum Genet* **90:** 7–24.

Warringer J, Zorgo E, Cubillos FA, Zia A, Gjuvsland A, Simpson JT, Forsmark A, Durbin R, Omholt SW, Louis EJ, et al. 2011. Trait variation in yeast is defined by population history. *PLoS Genet* **7:** e1002111.

Wu L, Candille SI, Choi Y, Xie D, Jiang L, Li-Pook-Than J, Tang H, Snyder M. 2013. Variation and genetic control of protein abundance in humans. *Nature* **499:** 79–82.

Yvert G, Brem RB, Whittle J, Akey JM, Foss E, Smith EN, Mackelprang R, Kruglyak L. 2003. *Trans*-acting regulatory variation in *Saccharomyces cerevisiae* and the role of transcription factors. *Nat Genet* **35:** 57–64.

Received December 4, 2013; accepted in revised form May 6, 2014.