Supplementary Discussion

Supplementary Discussion 1 – Genes with many eQTL

Some genes were affected by a large number of eQTL. The five genes with the maximum identified number of 21 eQTL included two pairs of genes that are physically located right next to each other, respectively. In each pair, the upstream gene (*NIT1* and *AQY2*, respectively) is annotated as a protein coding gene, while the downstream genes (YIL165C and YLL053C, respectively) is annotated as a "putative" protein. In several yeast strains other than the reference strain (a version of which is the BY strain we use here), these gene pairs each form a single open reading frame (ORF). In BY, this ORF is interrupted by a premature stop codon, resulting in truncated ORF annotations. The fact that BY tolerates the presence of a premature stop variant in these genes suggests that these genes are under low evolutionary constraint, which may help explain why these genes are also tolerant of regulatory influences from a large number of eQTL.

Supplementary Discussion 2 – Comparison of heritability to various gene features

We asked whether the heritability for each gene was correlated with various gene characteristics. Heritability was positively correlated with expression level (Supplementary Figure 2; Supplementary Table 1). This result may in part be caused by higher power in more highly expressed genes. Therefore, we controlled for expression level in a multivariate analysis that tested the correlation of various gene features with heritability while controlling for all other features, including gene expression level.

Genes with higher heritability were less likely to be essential, and had fewer protein-protein interaction and synthetic genetic interaction partners (Supplementary Table 1). This suggests that highly connected "hub" genes with many interaction partners may be less tolerant of regulatory variation. The amino acid sequence of genes with higher heritability evolved more slowly, perhaps reflecting higher sequence conservation.

Genes with higher heritability were enriched (T-test of heritability of genes in GO group vs. genes not in GO group as implemented in topGO[1]) for biological processes involved in mitochondrial function and cellular respiration (e.g. GO:0055114 "oxidation-reduction process": $p < 1e\text{-}30$; GO:0045333 "cellular respiration": $p = 5e\text{-}8$). In yeast, respiration is an optional and highly regulated means of energy production, and it is conceivable that the respiratory machinery provides a disproportionally large target for regulatory genetic variation.

Genes with lower heritability were enriched for processes involved in general biogenesis (e.g. GO:0042254 "ribosome biogenesis" $p < 1e\text{-}30$, GO:0010467 "gene expression" $p = 2e\text{-}25$, and GO:0000278 "mitotic cell cycle" $p < 1e\text{-}30$), suggesting that these cellular processes may be less tolerant of regulatory variation.

The 28 genes with the highest heritability of at least 0.9 were strongly enriched for genes involved in yeast mating. For example, they included five out of seven genes annotated as "regulation of mating-type specific transcription, DNA-templated" (GO:0007532, $p = 4e\text{-}11$). The yeast mating type pathway involves the **a** and alpha mating types that are determined by alternative alleles at the mating type locus. Each mating type expresses a set of highly abundant genes that are almost completely shut off in the other mating type. Our mapping population includes both mating types at equal frequency. Genes involved in mating are thus expected to fall into two genetically

determined groups in which their expression is either very high or nearly absent, resulting in high heritability across the segregant population.

Supplementary Discussion 3 – Relationship of eQTL number and heritability

The number of eQTL that influenced a given gene was correlated with heritability ($r = 0.56$, $p < 2.2e-16$, Supplementary Figure 2B), as expected if each additional eQTL adds to the genetic variance. However, among genes with the highest heritability (423 genes with $h^2 \geq 0.6$), heritability was negatively correlated with the number of eQTL ($r = -0.47$, $p < 2.2e-16$; Supplementary Figure 2B). For these genes, the strongest eQTL accounts for a progressively larger fraction of heritability ($r = 0.56$, $p < 2.2e-16$, Supplementary Figure 2C & D), while for genes with lower heritability this relationship was slightly negative ($r = -0.04$, $p = 0.001$). Thus, while the lower heritability typical of most genes tended to arise from multiple eQTL that each had small to intermediate effect, high heritability tended to arise from single, strong eQTL. The single eQTL that by themselves generated heritability of ≥ 0.9 were all local eQTL for genes in regions of the genome with high, structurally complex variation (e.g. the *ENA* locus[2] or subtelomeres[3]), as well as the gene *HO*, which is deleted in our RM but not in the BY strain.

Supplementary Discussion 4 – Allele-specific expression analyses

To estimate the fraction of local eQTL that act in *cis* vs. in *trans*, we compared the local eQTL to RNA-Seq data from a diploid BY/RM hybrid. In the hybrid, *trans*-acting genetic variation influences both alleles at a gene similarly. By contrast, *cis* effects can be detected as an allelic imbalance in expression (also called "allele specific expression", ASE) between the BY and the RM allele. We analyzed two independent BY/RM hybrid datasets and quantified ASE for 3,340 genes that have at least one variant in their coding sequence. Of these, 1,974 (59%) had a genome-wide significant local eQTL.

Genes with significant ASE (Bonferroni-corrected $p < 0.05$) had more local eQTL than expected by chance (at least one significant ASE dataset: 451 / 598 genes, $p < 2.2e-16$, odds ratio (OR) = 2.5; both ASE datasets significant: 100 / 121; $p = 3e-8$, OR = 3.4). Most genes with ASE but without a local eQTL had ASE of very small magnitude. For these genes, the local eQTL that would be expected to correspond to the ASE may have been missed in spite of high power in our current dataset. The remaining genes with ASE but without a local eQTL were located very close to strong eQTL that may have been misclassified as distant (Supplementary Table 2 & Supplementary Figure 3A & B). Overall, significant ASE typically resulted in a detectable local eQTL.

We next examined what fraction of the local eQTL arises in *cis* from ASE vs. from local *trans* acting variation. A simulation analysis of the diploid hybrid data showed that even if there is perfect agreement between ASE and local eQTL, the statistical power to detect ASE of the magnitude typical for most local eQTL is limited (Methods, Supplementary Figure 4). Indeed, 77% (1,523 / 1,974) of the local eQTL did not have significant ASE. Those local eQTL that had significant ASE had larger effects than those that did not (Supplementary Figure 3C). Of the 35 local eQTL for which power to detect ASE was at least 80%, 30 had ASE in at least one ASE dataset (26 in both datasets; Supplementary Figure 3D). Thus, most cases of missing ASE were

probably due to low statistical power, rather than caused by local *trans* effects. We present genes with the strongest discrepancies between ASE and local eQTL in Supplementary Tables 2 – 4.

Regulatory sequence variants in upstream regulatory regions such as the promoter are expected to act in *cis*, and should result in both ASE and local eQTL. Genes vary in the number of variants in their upstream regions. While the number of upstream variants was only weakly correlated with ASE (Spearman's rho = 0.04, p = 0.01), there was a stronger correlation with the fold changes of local eQTL (rho = 0.23, p < 2.2e-16). This analysis is imperfect because not every upstream variant has effects on expression, and because multiple variants with opposite effects could cancel each other's effect on expression. Therefore, *cis* effects are not expected to be a simple function of upstream variant number. Nevertheless, better correlation of variant number with local eQTL than with ASE is consistent with *cis* regulatory effects that were better detected in the current well-powered eQTL data than in the available ASE data.

The absolute values of the fold changes agreed remarkably well between ASE and local eQTL. The standardized major axis (SMA) slope for all local eQTL effects compared to ASE was 0.94 ($r^2 = 0.45$, p < 2.2e-16). Thus, the effects of *cis*-acting variation on allelic expression were typically carried forward to local eQTL of nearly the same magnitude. The SMA slope was just less than one (confidence interval 0.91 – 0.97), which may indicate a small tendency for local *trans*-acting variation to buffer some *cis*-acting variants[4].

Supplementary Discussion 5 – Comparison of eQTL and protein QTL (pQTL)

We focused our comparison of mRNA vs protein variation on distant rather than local eQTL and pQTL, because of the larger number of distant compared to local QTL. Local QTL yielded similar results. The main text presents a comparison focused on the strongest eQTL and pQTL. The results for all genome-wide significant eQTL and pQTL are presented here. The 154 genes that were present in both our current dataset and our earlier X-pQTL data[5] had 1,059 distant eQTL and 1,024 distant pQTL. Of these eQTL, 30% (321) overlapped a pQTL, while of the pQTL, 31% (314) overlapped an eQTL. The number of overlapping QTL differed between the two comparisons because QTL in one dataset can overlap two neighboring QTL in the other due to wide confidence intervals of weak QTL. Of the overlapping QTL, 77% (254 / 331) had the same direction of effect (Supplementary Figure 8).

Supplementary References

1. Alexa, A. & Rahnenführer, J. topGO: Enrichment analysis for Gene Ontology. (2016). doi:10.18129/B9.bioc.topGO
2. Treusch, S., Albert, F. W., Bloom, J. S., Kotenko, I. E. & Kruglyak, L. Genetic mapping of MAPK-mediated complex traits Across S. cerevisiae. *PLoS Genet* **11,** e1004913 (2015).
3. Bergstroem, A. *et al*. A High-Definition View of Functional Genetic Variation from Natural Yeast Genomes. *Mol. Biol. Evol.* **31,** 872–888 (2014).
4. Bader, D. M. *et al*. Negative feedback buffers effects of regulatory variants. *Mol Syst Biol* **11,** –785 (2015).
5. Albert, F. W., Treusch, S., Shockley, A. H., Bloom, J. S. & Kruglyak, L. Genetics of single-cell protein abundance variation in large yeast populations. *Nature* **506,** 494–497 (2014).

Supplementary Tables

Supplementary Table 1 – Multiple regression of heritability on various gene features.

Feature	Regression slope (change in heritability in units of percent per one unit of change in the given predictor)	Sums of squares	Degrees of freedom	F-value	p-value
Total expression variance	7.6	5.9	1	264	< 2.2e-16
Expression level[1]	2.1	2.5	1	111	< 2.2e-16
Essential (yes/no)	-8.2	2.2	1	98	< 2.2e-16
dNdS	-23	0.5	1	20	6e-6
Number of protein-protein interactions	-0.01	0.2	1	8	0.005
Number of genetic interactions	-0.01	0.3	1	15	1e-4
Transcription factor (yes/no)	0.4	0.001	1	0.06	0.8
Human homolog (yes/no)	-0.6	0.02	1	1.1	0.3
residuals	N/A	61.5	2739	N/A	N/A

Sums of squares, degrees of freedom and F-values were computed using Type II analysis of variance as implemented in the R car package.

[1]log2(TPM)

Supplementary Table 2 – Genes with strong (more than 2-fold) and significant ASE in both datasets but no local eQTL

Gene	Local eQTL LOD[2]	Local eQTL log2(fold change)	ASE p-value[1]	ASE log2(fold change)[3]
OPT2[4]	87	2.35	3e-13	2.8
CIN5	0.8	0.08	3e-14	-1.8
UIP3[4]	286	-1.65	1e-14	-1.4
YAR028W[4]	300	-2.52	1e-5	-1.3

[1] Shown is the less significant p-value from the two ASE datasets.

[2] The LOD score at the gene position itself irrespective of whether this eQTL is significant.

[3] Positive values indicate higher expression in RM compared to BY.

[4] These genes have strong eQTL close to the gene, but with a confidence interval that just excludes the gene. The may be influenced by *cis* acting local eQTL where the causal variant is located further away from the gene than captured by our definition of upstream regulatory regions as 1,000 base pairs upstream of the start codon.

Supplementary Table 3 – Genes with a local eQTL but no ASE in spite of ≥80% power to detect ASE

Gene	Local eQTL LOD	Local eQTL log2(fold change) [1]	ASE p-value[2,3]	ASE log2(fold change) [1]
TIF1	405	-0.87	0.003	-0.02
CBF1	349	-1.36	0.8	1e-5
VPS63	200	-4.4	0.09	-0.36
UBA1	128	-0.34	0.005	-0.01
TPO4	108	0.97	9e-5	0.08

[1] Positive values indicate higher expression in RM compared to BY.

[2] Shown is the more significant p-value from the two ASE datasets.

[3] The nominally significant p-values in this column do not pass Bonferroni cutoff for significance. Therefore, ASE at these genes was not identified as significant.

Supplementary Table 4 – Genes with a local eQTL and significant ASE, and discordant direction of effect

Gene	Local eQTL LOD	Local eQTL log2(fold change)[1]	ASE p-value[2]	ASE log2(fold change)[1,3]
TDH3	88	-0.33	4e-239	0.37
YTA12	5.1	-0.06	1e-6	0.25
DBP5	3.6	-0.03	3e-6	0.18

[1] Positive values indicate higher expression in RM compared to BY.

[2] Shown is the less significant p-value from the two ASE datasets.

[3] The table shows only genes where both ASE datasets agreed in the direction of effect. Shown is the average effect.

Supplementary Table 5 – Multiple logistic regression of genes located in hotspots on various gene features.

Feature	Regression slope	Likelihood ratio	Degrees of freedom	p-value
Expression level[1]	0.15	0.9	1	0.3
Essential (yes/no)	-1.96	5.3	1	0.02
dNdS	-6.94	2.9	1	0.09
Number of protein-protein interactions	0.001	0.06	1	0.8
Number of genetic interactions	0.002	1.2	1	0.3
Transcription factor (yes/no)	2.73	21	1	4e-6
Human homolog (yes/no)	-1.30	6.1	1	0.01

Likelihood ratios, degrees of freedom and p-values were computed using Type II analysis of variance as implemented in the R car package.

[1]log2(TPM)

Supplementary Table 6 – Strong eQTL without pQTL

Gene	Chromosome	Position (bp)	eQTL LOD	eQTL effect[1]	pQTL effect[1]
MMT1	XII	657,792	33.81	-0.38	-0.10
GPP2	XII	659,350	33.24	0.37	0.06
ATP5	XII	660,371	36.82	0.39	0.08
CDC10	XII	661,927	31.37	0.36	0.02
ILV6	XIV	376,313	61.21	-0.49	0.19
UTP4	XIV	376,313	29.12	0.35	-0.06
ARO8	XIV	377,751	39.94	-0.41	0.04
TRP5	XIV	377,751	29.89	-0.36	-0.06
SOP4	XIV	393,050	32.33	-0.37	0.00
STT3	XIV	449,640	33.23	-0.37	0.06
CAM1	XIV	462,478	54.46	-0.47	-0.10
ADO1	XIV	466,588	31.96	-0.37	-0.06
ARO1	XIV	466,588	28.89	0.35	-0.04
COX17	XIV	466,588	41.59	0.42	0.00
SEY1	XIV	466,588	145.00	0.70	0.04
PAT1	XIV	467,028	90.02	0.58	0.00

[1] Positive values indicate higher expression in RM compared to BY.

Supplementary Table 7 – Strong mRNA and protein QTL with opposite effect

Gene	Chromosome	Position (bp)	eQTL LOD	pQTL LOD	eQTL effect[1]	pQTL effect[1]
HEM1	XII	657,022	25.33	13.97	-0.33	0.17
HXK2	XII	657,022	10.62	10.74	0.22	-0.16
BDH1	XII	657,792	11.28	13.17	0.22	-0.16
GPD1	XII	662,515	13.50	27.24	0.24	-0.25
RPS17A	XII	662,515	5.75	19.3	-0.16	0.22
MDH1	XIII	74,632	7.20	18.64	0.18	-0.18
CYC1	XIII	338,431	9.43	28.95	-0.20	0.27
NEW1	XIV	368,183	5.13	26.49	0.15	-0.24
ILV6	XIV	464,117	8.04	38.18	-0.19	0.31
OLE1	XIV	465,187	6.62	13.4	0.17	-0.15
BDH1	XIV	466,588	16.14	22.36	0.27	-0.21
RPL13B	XIV	466,588	17.14	29.06	-0.27	0.24
RPL19A	XIV	466,588	15.48	36.39	-0.26	0.28
RPS17A	XIV	466,588	11.10	18.55	-0.22	0.19
RPS25A	XIV	466,588	13.54	13.09	-0.24	0.19
SEC16	XIV	466,588	7.86	16.52	0.19	-0.19
SLA1	XIV	466,588	18.64	23.48	0.29	-0.21
GCV3	XIV	467,028	17.55	9.04	-0.28	0.16
CPA2	XIV	469,224	10.41	70.15	0.22	-0.50

[1] Positive values indicate higher expression in RM compared to BY.

Supplementary Table 8 – Strong pQTL without eQTL

Gene	Chromosome	Position (bp)	pQTL LOD	pQTL effect[1]	eQTL effect[1]
RPS17A	II	137,197	38.83	-0.30	-0.01
TIP1	V	378,304	44.28	0.30	0.02
GPD1	V	504,305	123.79	-0.54	-0.02
TPO1	VII	475,295	51.83	0.35	0.03
TIP1	XI	269,056	39.08	0.30	-0.07
SSA1	XI	270,756	95.23	0.50	-0.02
LEU1	XII	676,798	123.29	0.57	0.04
CAR2	XIII	112,600	46.66	-0.33	-0.02
GCN1	XIV	457,698	94.83	-0.48	0.03
ATP2	XIV	460,398	104.94	0.52	0.07
CDC60	XIV	464,098	45.72	-0.31	0.05
GLN1	XV	167,400	43.03	-0.31	-0.03

[1] Positive values indicate higher expression in RM compared to BY.