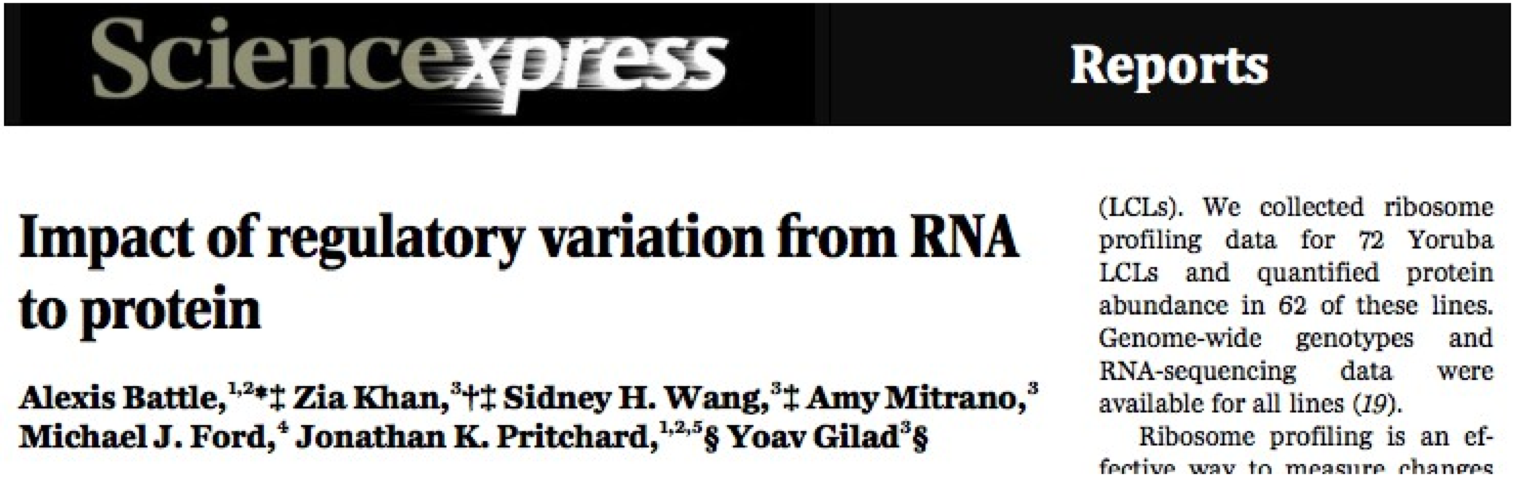


Sciencexpress

Reports

Impact of regulatory variation from RNA to protein

Alexis Battle,[1,2]*‡ Zia Khan,[3]†‡ Sidney H. Wang,[3]‡ Amy Mitrano,[3] Michael J. Ford,[4] Jonathan K. Pritchard,[1,2,5]§ Yoav Gilad[3]§

(LCLs). We collected ribosome profiling data for 72 Yoruba LCLs and quantified protein abundance in 62 of these lines. Genome-wide genotypes and RNA-sequencing data were available for all lines (*19*).

Ribosome profiling is an effective way to measure changes

“The majority of RNA expression differences between individuals have no connection to the abundance of a corresponding protein, report scientists. The results point to a yet-unidentified gene regulatory mechanism.”

ScienceDaily Dec 18, 2014

“The majority of RNA expression differences between individuals have no connection to the abundance of a corresponding protein, report scientists. The results point to a yet-unidentified gene regulatory mechanism.”

“The team confirmed variation in QTLs led to different levels of mRNA for a large number of genes. Yet, only one-third showed an accompanying change in protein levels. The majority of QTLs that affected mRNA levels did not have any effect on the amount of corresponding protein.”

ScienceDaily Dec 18, 2014

“The majority of RNA expression differences between individuals have no connection to the abundance of a corresponding protein, report scientists. The results point to a yet-unidentified gene regulatory mechanism.”

"The chief assumption for studies of RNA differences is that they ultimately reflect differences in an end product, which is protein," said senior study author Yoav Gilad, PhD, professor of human genetics at the University of Chicago. "But it turns out in most cases this may not be true."

“The team confirmed variation in QTLs led to different levels of mRNA for a large number of genes. Yet, only one-third showed an accompanying change in protein levels. The majority of QTLs that affected mRNA levels did not have any effect on the amount of corresponding protein.”

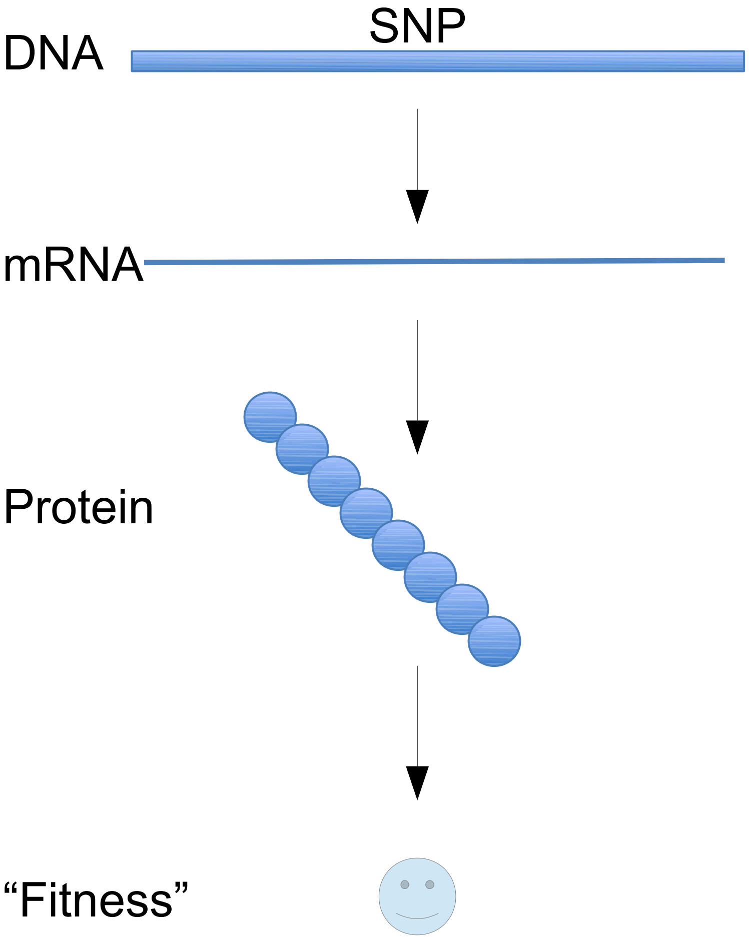
ScienceDaily Dec 18, 2014

QTL- Quantitative Trait Loci

Quantitative trait loci (QTLs) are stretches of DNA containing or linked to the genes that underlie a quantitative trait. Mapping regions of the genome that contain genes involved in specifying a quantitative trait is done using e.g. SNPs.

Quantitative traits refer to phenotypes (characteristics) that vary in degree and can be attributed to polygenic effects, i.e., product of two or more genes, and their environment.

Modified from Wikipedia



DNA

SNP

mRNA

Protein

“Fitness”

Genotype data (SNP)

“Transcript expression” phenotype

“Ribosomal profiling” phenotype

“Steady state protein levels” phenotype

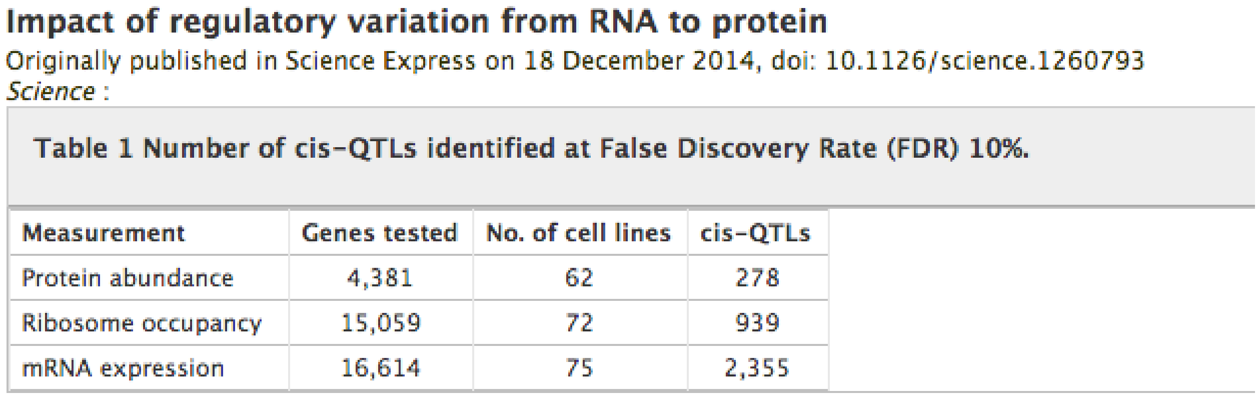
3 data sets

- Lymphoblastoid cell lines (LCLs)(from Nigerian individuals, HapMap project)
  - Transcript expression
  - Ribosomal profiling
  - Steady state protein levels

N=75-62

QTL analysis 1

- For all loci with minor allele frequency > 10%
- 20 kb window around corresponding gene
- For each phenotype separately



Impact of regulatory variation from RNA to protein

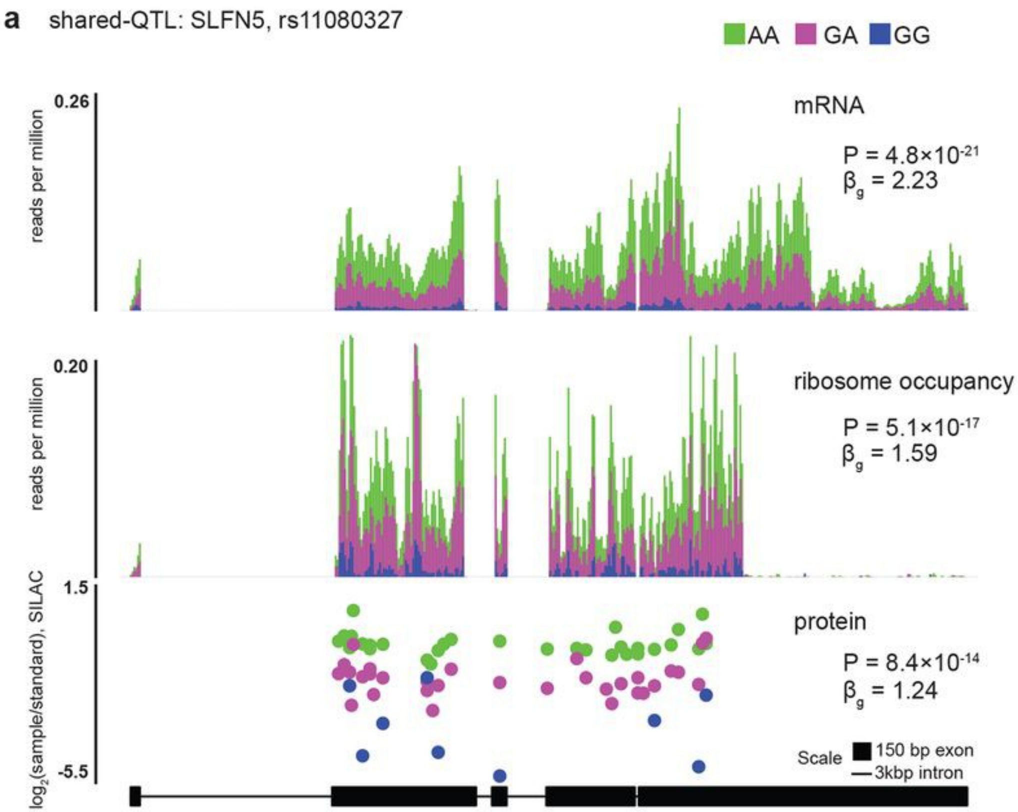
Originally published in Science Express on 18 December 2014, doi: 10.1126/science.1260793

Science :

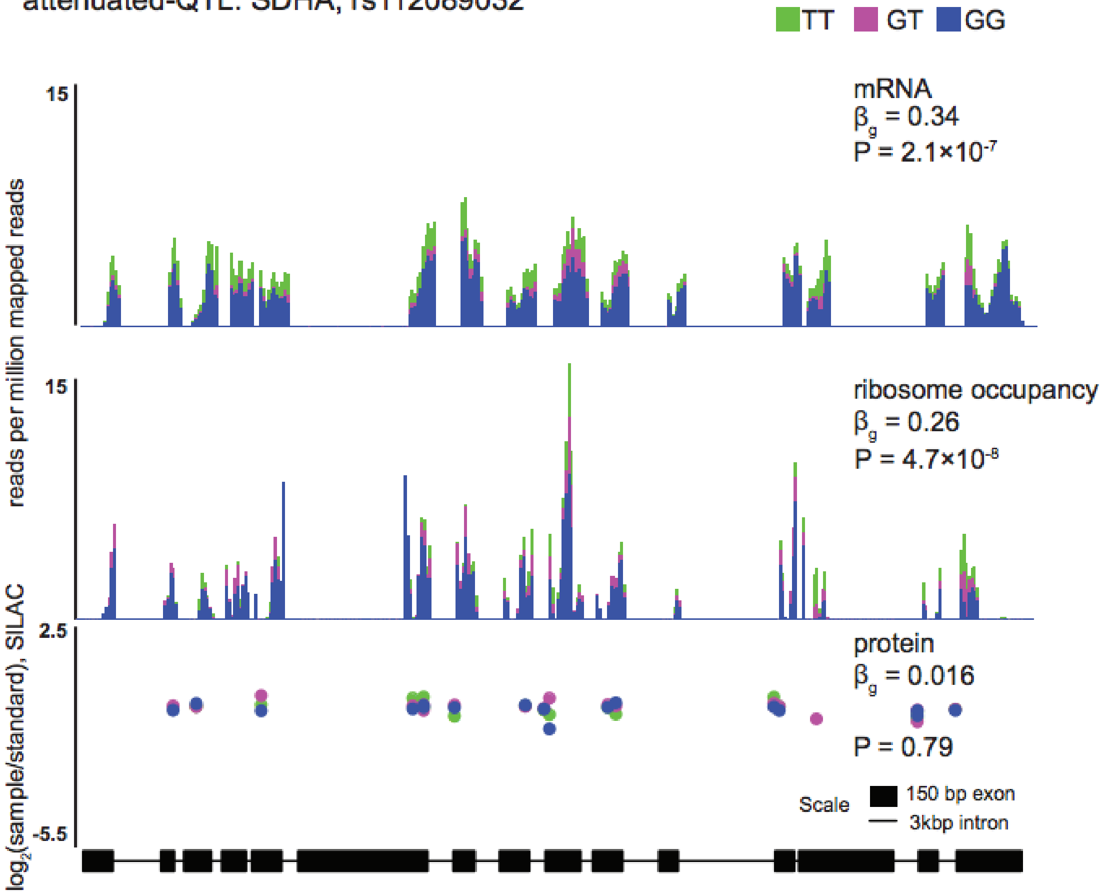
Table 1 Number of cis-QTLs identified at False Discovery Rate (FDR) 10%.

Measurement	Genes tested	No. of cell lines	cis-QTLs
Protein abundance	4,381	62	278
Ribosome occupancy	15,059	72	939
mRNA expression	16,614	75	2,355

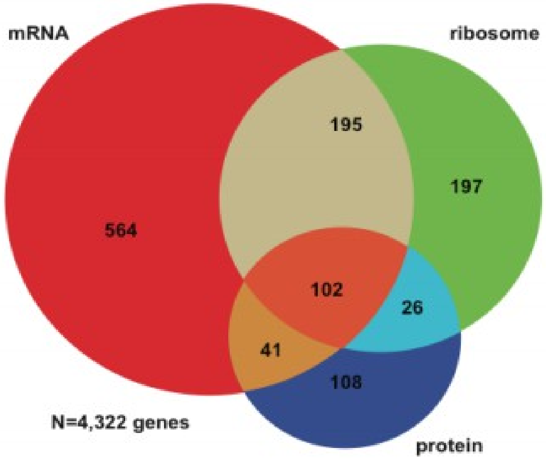
Many QTLs exhibit shared effects across mRNA, ribosome occupancy and protein.



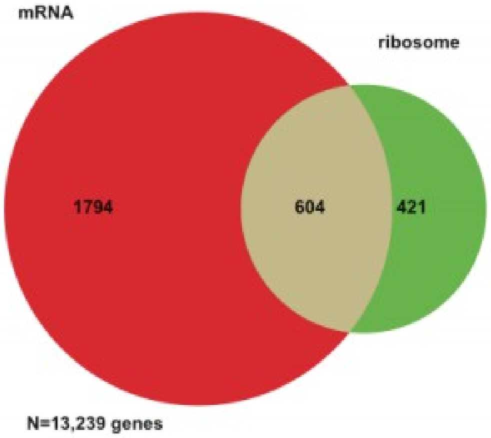
attenuated-QTL: SDHA, rs112089032



a strict overlap between eQTLs, rQTLs, and pQTLs found in genes measured across all 3 phenotypes



b strict overlap between eQTLs and rQTLs across genes measured in RNA and ribosome occupancy

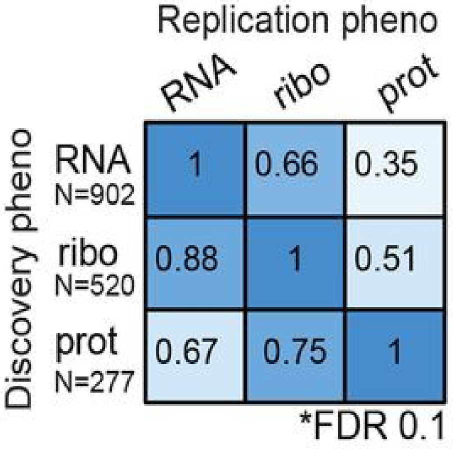


Comparisons of QTLs at three levels of gene regulation.

Analysis 2: (to minimize affect of different statistical power with different data sets)

For each discovered SNP-gene pair in each discovery phenotype
– look at same specific pair i the other (replication) phenotypes

b replication rates of cis-QTLs across phenotypes



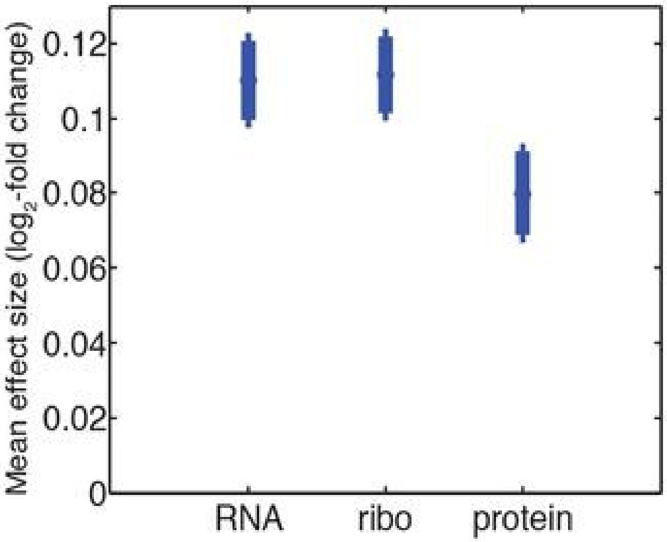
Discovery pheno	Replication pheno: RNA	ribo	prot
RNA N=902	1	0.66	0.35
ribo N=520	0.88	1	0.51
prot N=277	0.67	0.75	1

*FDR 0.1

Highest rep.rate: QTL from Ribosomal profiling in RNA
Lowest: QTL from RNA in protein

Analysis 3: Using an independent set of SNP-gene pairs (from GEUVADIS study)
The same analyis as in 2, but only including SNP which were present in current dataset and where data was complete.
The independent dataset was used to avoid biasing the results (Winner's Curse).

C effect size of eQTLs ascertained from the GEUVADIS study



Using independent QTL they conclude that effect sizes are smaller in protein than in RNA and protein.

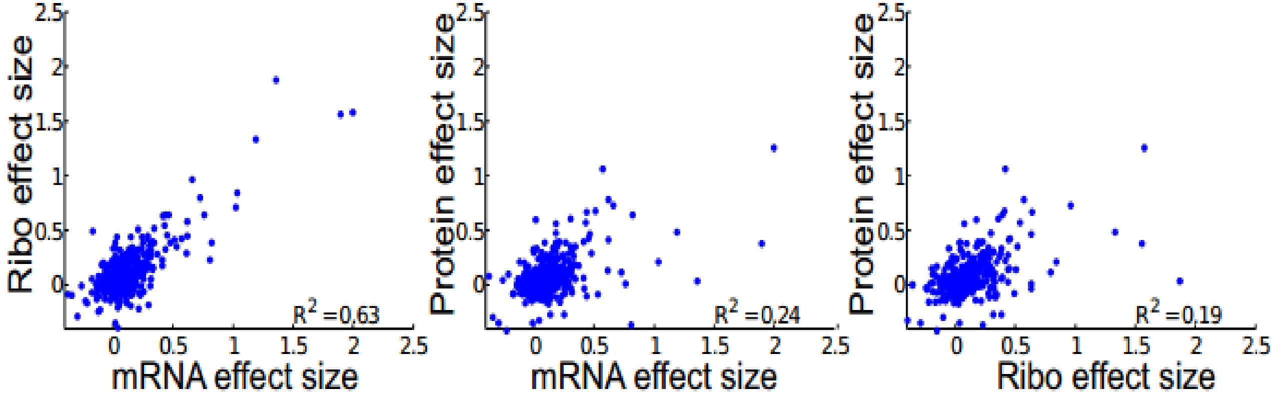
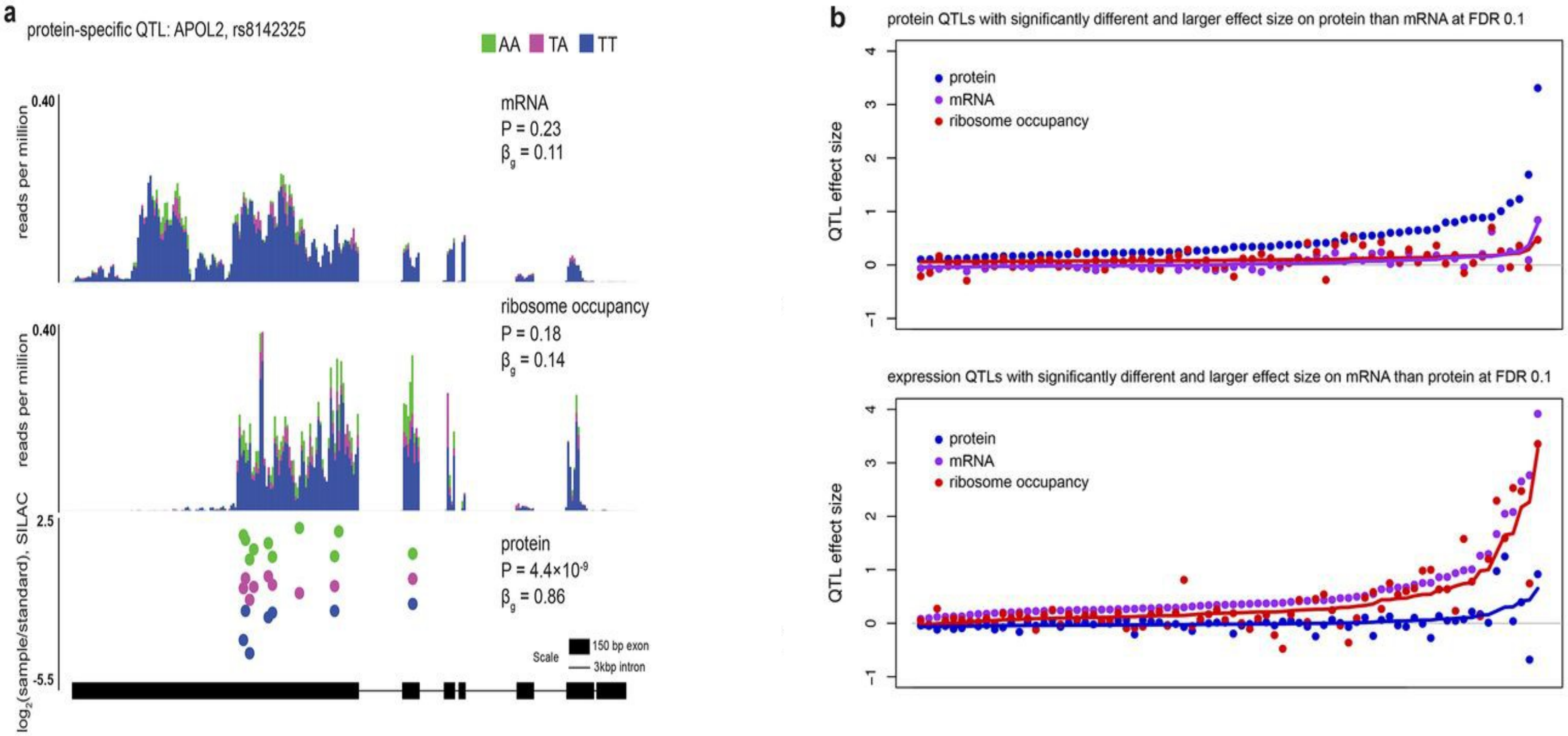
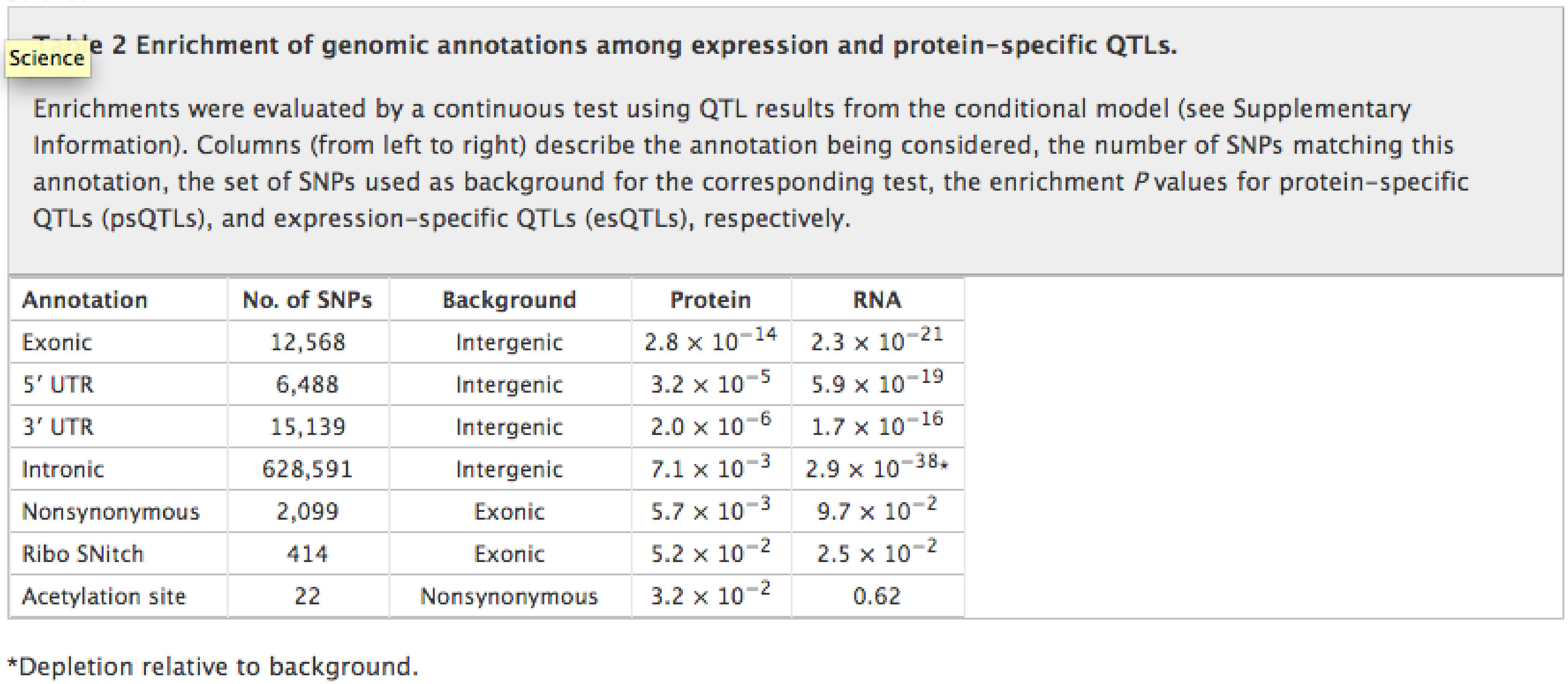


Figure S14. Effect sizes of GEUVADIS eQTLs compared in mRNA, ribosome occupancy, and protein. Here, we compare effect sizes for each individual eQTL previously ascertained in the GEUVADIS study, with comparisons between each pair of expression phenotypes (plots show fold change on $\log_2$ scale). We measured effect size as the regression coefficient obtained from linear regression using raw data for each of our three phenotypes (no quantile normalization or PC-correction, but mRNA and ribosome occupancy values were log-transformed). One outlier data point is not shown on the protein scatter plots, with an effect size of 3.5 in protein, and 1.0 and 0.7 in mRNA and ribo respectively. Strong correlation and no effect size compression is observed between mRNA and ribosome occupancy, whereas effect sizes appear compressed in protein data.

They also found examples where QTL where only found in protein



Enrichment analysis



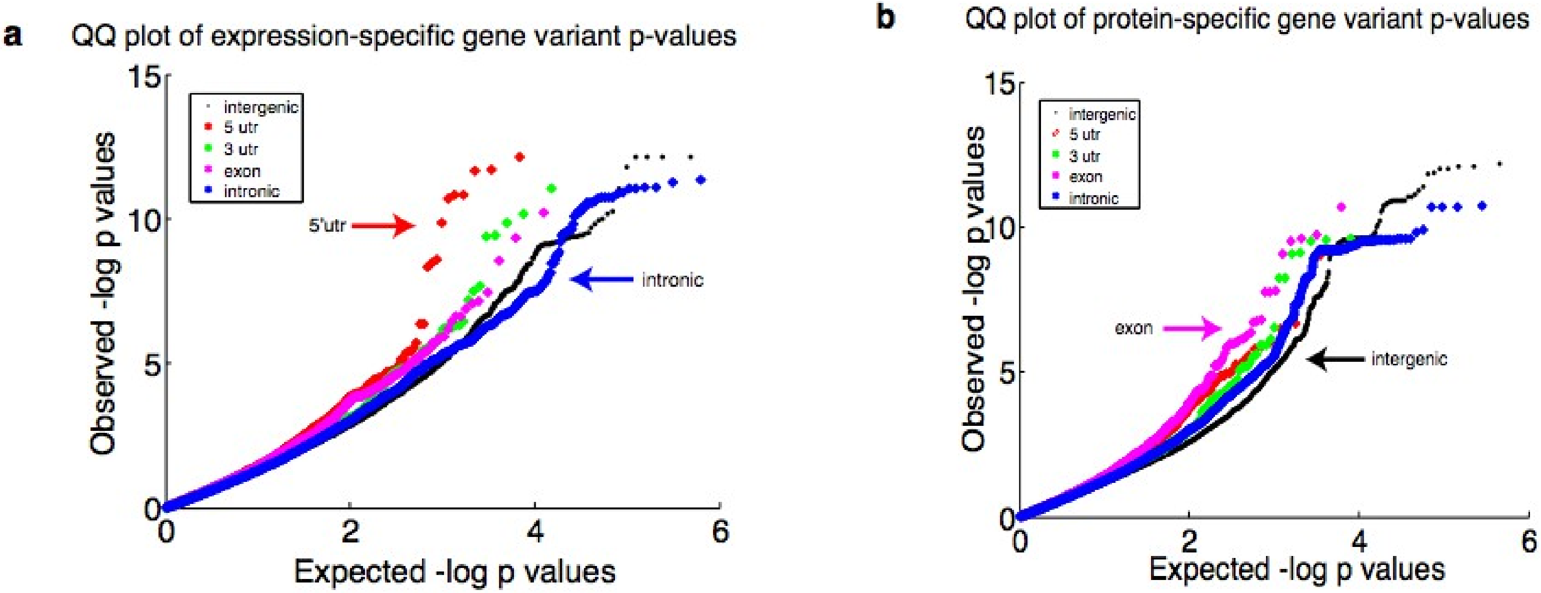
Science

Table 2 Enrichment of genomic annotations among expression and protein-specific QTLs.

Enrichments were evaluated by a continuous test using QTL results from the conditional model (see Supplementary Information). Columns (from left to right) describe the annotation being considered, the number of SNPs matching this annotation, the set of SNPs used as background for the corresponding test, the enrichment *P* values for protein-specific QTLs (psQTLs), and expression-specific QTLs (esQTLs), respectively.

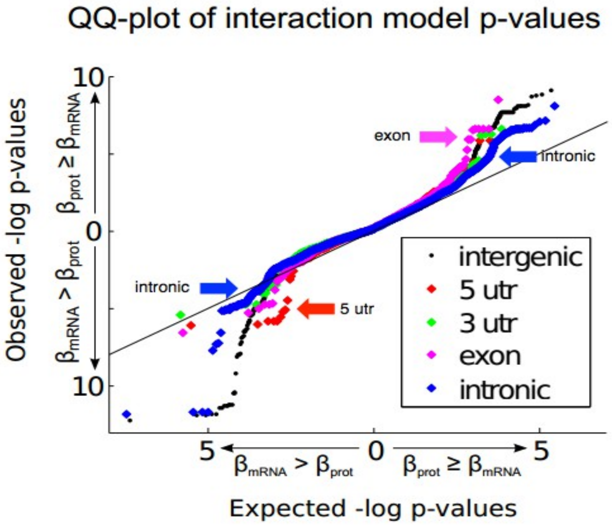
Annotation	No. of SNPs	Background	Protein	RNA
Exonic	12,568	Intergenic	2.8×10^{-14}	2.3×10^{-21}
5′ UTR	6,488	Intergenic	3.2×10^{-5}	5.9×10^{-19}
3′ UTR	15,139	Intergenic	2.0×10^{-6}	1.7×10^{-16}
Intronic	628,591	Intergenic	7.1×10^{-3}	2.9×10^{-38}*
Nonsynonymous	2,099	Exonic	5.7×10^{-3}	9.7×10^{-2}
Ribo SNitch	414	Exonic	5.2×10^{-2}	2.5×10^{-2}
Acetylation site	22	Nonsynonymous	3.2×10^{-2}	0.62

*Depletion relative to background.

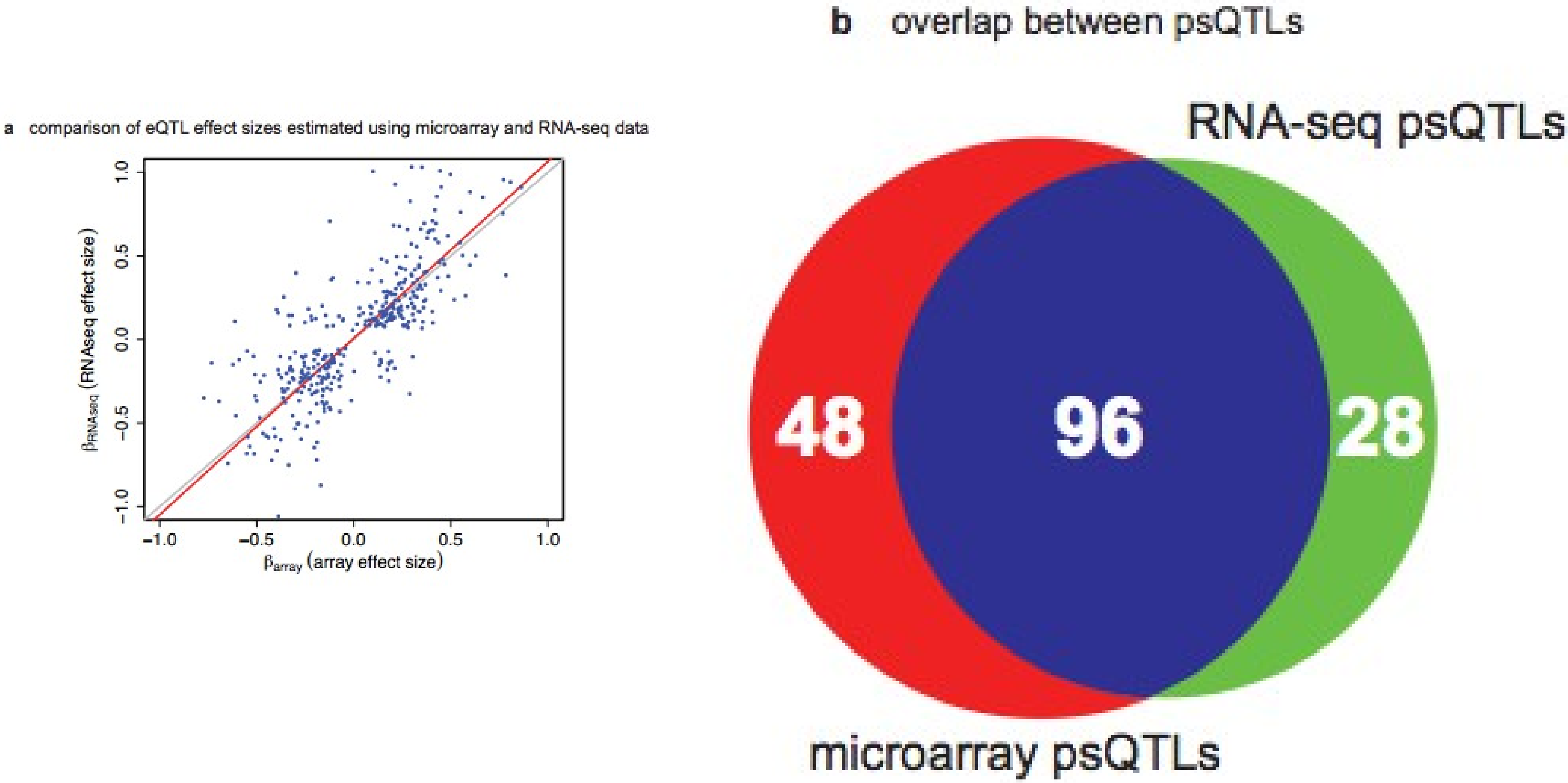


a QQ plot of expression-specific gene variant p-values

b QQ plot of protein-specific gene variant p-values


QQ-plot of interaction model p-values
Observed -log p-values
$\beta_{mRNA} > \beta_{prot}$
$\beta_{prot} \geq \beta_{mRNA}$
10
0
10
exon
intronic
intronic
5 utr
● intergenic
◆ 5 utr
◆ 3 utr
◆ exon
◆ intronic
5
0
5
$\beta_{mRNA} > \beta_{prot}$
$\beta_{prot} \geq \beta_{mRNA}$
Expected -log p-values

To check if findings are stable over different technologies, they compared the RNA-seq results with microarray results



Summary

Careful statistics/bioinformatics work has been done to minimize systematic biases etc.
Few things are not covered/explained. (scale of data,effect distribution, differences between technologies, model fit etc)
I also miss comments on protein degradation, ribosomal stalling etc. Isn't this known already?

The biological finding that there most exist some unknown mechanism is interesting, but only as starting point. This goes for both the cases where the protein is more stable and when it is less stable.

“Gilad and his colleagues hypothesize a yet-unidentified buffering mechanism prevents dramatic shifts in protein levels when mRNA levels and ribosome function are in flux.”

The study nevertheless highlights something very important. Regardless of the cause of the discrepancy one should be aware of it.

"This underscores the importance of being cautious," Gilad said. "We need to make sure that genes we think are important for disease or that give us insight into biology are linked to protein products and not just RNA."

We tested for phenotype-specific QTLs using a likelihood ratio test (LRT) to identify SNPs significantly associated with the phenotype of interest while including other phenotypes as covariates to account for effects fully mediated by another phenotype. Let P_{tij} represent the expression level of the phenotype t in individual i and gene j, g_{is} be the genotype for that individual at SNP s, and P_{uij} be the expression level in another phenotype u that we seek to control for. We compared two linear models

$$P_{tij} = \mu_{tj} + \beta_g g_{is} + \beta_p P_{\mathrm{u}ij} + \epsilon_{i,j}$$

$$P_{tij} = \mu_{tj} + \beta_p P_{\mathrm{u}ij} + \epsilon_{i,j}$$

using a likelihood ratio test to determine the contribution of SNP s not mediated by phenotype P_u. To identify