REPORT

Epigenetic epistatic interactions constrain the evolution of gene expression

Solip Park1 and Ben Lehner1,2,*

1 EMBL-CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and University Pompeu Fabra (UPF), Barcelona, Spain and 2 Institució Catalana de Recerca i Estudis Avançats (ICREA), Barcelona, Spain

* Corresponding author. EMBL-CRG Systems Biology Research Unit, Centre for Genomic Regulation (CRG) and University Pompeu Fabra (UPF), Dr Aiguader 88, Barcelona 08003, Spain. Tel.: +34 933 160 194; Fax: +34 933 316 0199; E-mail: ben.lehner@crg.eu

Received 26.9.12; accepted 7.1.13

Reduced activity of two genes in combination often has a more detrimental effect than expected. Such epistatic interactions not only occur when genes are mutated but also due to variation in gene expression, including among isogenic individuals in a controlled environment. We hypothesized that these ‘epigenetic’ epistatic interactions could place important constraints on the evolution of gene expression. Consistent with this, we show here that yeast genes with many epistatic interaction partners typically show low expression variation among isogenic individuals and low variation across different conditions. In addition, their expression tends to remain stable in response to the accumulation of mutations and only diverges slowly between strains and species. Yeast promoter architectures, the retention of gene duplicates, and the divergence of expression between humans and chimps are also consistent with selective pressure to reduce the likelihood of harmful epigenetic epistatic interactions. Based on these and previous analyses, we propose that the tight regulation of epistatic interaction network hubs makes an important contribution to the maintenance of a robust, ‘canalized’ phenotype. Moreover, that epigenetic epistatic interactions may contribute substantially to fitness defects when single genes are deleted.

Molecular Systems Biology 9: 645; published online 19 February 2013; doi:10.1038/msb.2013.2

Subject Categories: functional genomics; chromatin & transcription

Keywords: epigenetics; epistasis; evolution; gene expression; genetic interaction

Introduction

Mutations do not have the same effects in all individuals. One reason for this is that the outcome of a mutation can depend on other genetic variation in a genome. Such genetic—or epistatic—interactions between mutations have been extensively mapped in model organisms, and contribute to disease phenotypes in humans (Lehner, 2007, 2011; Costanzo et al., 2011). Indeed in model organisms the effects of loss-of-function mutations in most genes can be enhanced by deletions in other loci (Costanzo et al., 2010). However, systematic genetic interaction mapping projects have also revealed that genes differ widely in the number of genetic interaction partners that they have (Pan et al., 2004; Tong et al., 2004; Schuldiner et al., 2005; Lehner et al., 2006; Costanzo et al., 2010). Although the inactivation of a typical gene only enhances the effects of mutations in a relatively small number of other loci, reduced activity of genes referred to as genetic interaction hubs (or buffers or capacitors) can enhance the phenotypic consequences of mutations in many different loci encoding genes with diverse functions (Costanzo et al., 2010, 2011).

Mutations are, however, just one cause of variation in gene activity. Two individuals may also differ in the activity of a particular gene simply because of differences in gene expression. For example, genes change in expression in response to the environment and during development. Moreover, gene expression also varies—often substantially—among isogenic individuals or cells in a common environment (Newman et al., 2006; Lopez-Maury et al., 2008). Both these environmentally triggered (Casanueva et al., 2012) and stochastic (Burga et al., 2011) changes in gene expression can also cause epistatic interactions with mutations: one protein’s activity is reduced because of mutation and the other because of insufficient gene expression, so triggering an epistatic interaction. We refer to these interactions as ‘epigenetic interactions’ or ‘epigenetic epistasis’ (Figure 1A).

We hypothesized that epigenetic epistasis could place an important constraint on gene expression and evolution. In particular, we hypothesized that the more genetic interaction partners a gene has, the stronger the selective constraint will be to maintain stable expression of that gene (Figure 1B).
is because, as a first approximation, the more genetic interaction partners a gene has, the higher the likelihood that at least one of them carries a detrimental mutation. Moreover, with more interaction partners there is also an increased likelihood that at least one interaction partner will have low expression in a particular individual. In short, the more genetic interaction partners a gene has, the greater the chance of an epigenetic interaction when the expression of that gene varies (Figure 1B).

In budding yeast, genome-wide measurements of gene expression can be used to quantify expression variation at many different scales, ranging from variation in expression levels among isogenic individuals in a common environment (‘noise’), through the responsiveness of expression to changes in conditions (‘plasticity’), to the divergence in expression between species (‘evolution’). This large resource of expression measurements allows us to test whether genes with more genetic interaction partners do indeed have less variable gene expression. We show here that at each of these scales this is the case and that genetic interaction network hubs have particularly tightly controlled and slowly evolving expression. We discuss the implications of this for the evolution of robust (or ‘canalized’ (Waddington, 1942)) phenotypes, and how the tight regulation of genetic network interaction hubs may both constrain short-term phenotypic evolution and facilitate it over a longer timescale.

**Results**

**Epistatic interaction network hubs have low expression noise and stable expression across environmental conditions**

We hypothesized that, in order to maintain phenotypic stability, the expression of genes with many potential epistatic interaction—hub genes in genetic interaction networks—should be tightly regulated. If a gene has many genetic interaction partners, then variation in the expression of that gene is more likely to cause an ‘epigenetic’ epistatic interaction (Figure 1). During evolution, therefore, the selective pressure to reduce gene expression variation may be stronger for genes with more genetic interaction partners.

Genetic interactions have been systematically mapped in budding yeast, identifying over 100,000 high-confidence negative genetic interactions between more than 4000 genes (Costanzo et al., 2010). In this data set, the number of genetic interactions for each gene ranges between 1 and 511, with a median of 22 interactions per gene (Costanzo et al., 2010). To relate the number of genetic interaction partners per gene to the variability of gene expression, we compiled data from multiple studies that analyzed genome-wide variation in gene expression at different scales (Figure 1B and Supplementary File 1).

We first considered how genes with different number of genetic interactions vary in expression among isogenic cells in a common environment—that is, their expression ‘noise’, as quantified using a library of GFP-tagged fusion proteins (Newman et al., 2006). As shown in Figure 2A, considering measurements across more than 1300 genes, expression noise is anticorrelated with genetic interaction degree ($r = -0.18$, $P = 4.54E-11$). Thus, genes with many potential genetic interaction partners have expression that is generally less variable from cell to cell in a single environment.

Next, we considered how genes vary in expression in response to changes in conditions. Ihmels et al (2002)
quantified how much each gene’s expression changes in a curated data set consisting over 1500 expression profiles across different experimental conditions, that is, the expression ‘responsiveness’ or ‘plasticity’ following external perturbations (Tirosh et al., 2006). As for expression noise, when considering more than 3600 yeast genes, the responsiveness of a gene’s expression is inversely related to the number of genetic interaction partners \( (\rho = -0.17, \ P = 4.67E^{-26}; \text{Figure 2B}) \). That is, genes with many genetic interaction partners typically have expression that changes very little across different environmental conditions. This result is also upheld when using alternative measures of expression plasticity (Supplementary Figure 1). Thus, the more genetic interactions a gene has, the more tightly controlled its expression is both within a condition and across different conditions. As for all of the other expression properties considered below, these relationships remain significant when controlling for the individual fitness effects of each gene deletion, gene expression levels, phenotypic capacity or multifunctionality (Supplementary Figures 3 and 6).

Genes with many genetic interactions have expression that is less responsive to genetic change

The above results imply that the expression of genetic interaction hubs responds less to stochastic and environmental stimuli. We next asked whether they also respond less to genetic perturbations, that is, whether their expression is less sensitive to mutation. We used two data sets: first, the variance of expression across yeast cell lines that had been allowed to accumulate random mutations during 4000 generations of growth under conditions of minimal selection due to population bottlenecks (mutational variance, \( V_m \); Landry et al., 2007), and second, a measure of the variance of each gene’s expression caused by different trans-acting genetic loci in recombinant inbred lines generated in a cross between two strains (trans variability; Brem et al., 2002; Choi and Kim, 2008). In both cases we observe that genes with more genetic interactions have expression that varies less across genotypes \( (\rho = -0.15, \ P = 4.40E^{-19} \text{ for mutation variation}, \rho = -0.15, \ P = 4.96E^{-16} \text{ for trans variability; Figure 3A and B}) \). Thus, genetic interaction hubs also have expression that is intrinsically less ‘evolvable’ when either random mutations accumulate in a genome or in response to combinations of natural variants.

The expression of genetic interaction hubs diverges slowly within and between species

The reduced evolvability of the expression of genetic interaction hubs predicts that their expression might also diverge more slowly during evolution. We tested this in two ways—by considering expression divergence (ED) among strains of one species, and by considering ED among different yeast species. In both cases, the levels of divergence inversely correlate with the number of genetic interactions \( (\rho = -0.15, \ P = 1.69E^{-20} \text{ for inter-strain variation}, \rho = -0.12, \ P = 1.00E^{-09} \text{ for ED between species; Figure 4A and B}) \). Genes with many genetic interaction partners, therefore, tend to have expression that evolves slower than other genes.

Genetic interaction hubs have biased promoter architectures

Taken together, these analyses show that genetic hubs have more stable expression levels at many different scales: among individuals, in response to environmental change, in response to mutations and during evolution. These different timescales of expression variability are correlated across genes (Supplementary Figure 9; Tirosh et al., 2006; Landry et al., 2007; Choi and Kim, 2009; Lehner, 2010a). What are the molecular mechanisms that underlie this constrained gene expression? Previous studies have linked the presence of a TATA box and high promoter nucleosome occupancy to increased variability of expression (Tirosh et al., 2006; Landry et al., 2007; Choi and Kim, 2009; Lehner, 2010a). Consistent with this, we find that the more genetic interaction partners a gene has, the less likely it is to have a TATA box element in its promoter (Kolomogorov–Smirnov (KS) statistic,
et al. 2012; Koch et al., 2012; Ryan et al., 2012). Therefore, we assigned the interaction degree of more than 2500 yeast genes to their human orthologs (Figure 7A). This revealed that, just as in yeast, these predicted genetic hubs in human also have expression that is constrained during evolution, showing reduced ED between human and chimpanzee ($\rho = -0.10$, $P = 1.57E-06$; Figure 7B). This relationship remains significant when controlling for the individual fitness effects of each gene deletion ($\rho = -0.05$, $P = 1.47E-02$) or gene expression levels ($\rho = -0.09$, $P = 2.17E-05$) (Supplementary Figure 5). We concluded that the principle that epistatic interactions constrain expression variation is likely to be conserved across species.

### Discussion

**Evidence that epigenetic epistatic interactions constrain the evolution of gene expression**

We have shown here that genes with many genetic interaction partners have less variable gene expression across multiple species. However, we have also observed that highly connected genes with high expression variation are more likely to have gene duplicates than highly connected genes with low expression variation. This suggests that the presence of duplicates may be a mechanism to escape the detrimental effects of high expression variation. Indeed, we find that highly connected genes with variable expression are likely to have more gene duplicates than highly connected genes with less variable expression (Figure 6). This suggests that the highly variable expression of a subset of genetic hubs may be tolerated because of the presence of duplicates.

Highly connected genes with variable expression are enriched for gene duplicates

Despite the overall anticorrelation between genetic interaction degree and variation in gene expression, some highly connected genes still show substantial variation in gene expression. We were interested in whether these genes have an alternative mechanism to reduce the detrimental effects of variation in expression. One general mechanism to escape from adaptive conflict is gene duplication (Hughes, 1994; Des Marais and Rausher, 2008). Based on previous work (Lehner, 2010a), we hypothesized that the variable gene expression of some highly connected genes might be compensated for by the expression of partially redundant gene duplicates. Indeed we find that highly connected genes with variable expression are likely to have more gene duplicates than highly connected genes with less variable expression (Figure 6). This suggests that the highly variable expression of a subset of genetic hubs may be tolerated because of the presence of duplicates.

Epistatic interactions may constrain gene expression in other species

In yeast we observed that genes with more genetic interaction partners have more constrained expression than other genes. We were interested in whether this result also applies to other species. Genetic interaction hubs appear to be well conserved across species, as does the genetic interaction degree of individual genes (Roguev et al., 2007; Dixon et al., 2008; Frost et al., 2012; Koch et al., 2012; Ryan et al., 2012). Therefore, we...
different timescales. Genetic interaction hubs have expression that is less noisy, less responsive to changes in conditions, less evolvable, and diverging more slowly during evolution. This reduced expression variation is linked to differences in promoter architecture, with interaction hubs having transcription that is most likely intrinsically less ‘bursty’ (Hornung et al., 2010) and so less sensitive to perturbations.

The tight regulation of genetic hubs may contribute to phenotypic robustness

An important concept in biology is phenotypic robustness or canalization (Waddington, 1942; Wagner, 2000; Stelling et al., 2004). Based on our results, we propose that the stable expression of genetic interaction hubs makes an important contribution to this robustness. Of course, the relationship between genetic interaction degree and expression variation may also be indirect. However, other observations are also consistent with our hypothesis, such as the tendency for two genetic partners not to simultaneously have low expression (Supplementary Figure 8). In addition, yeast strains in which genetic interaction hubs have been deleted have highly variable morphological phenotypes (Ohya et al., 2005; Levy and Siegal, 2008; Lehner, 2010b), and strains in which genetic hubs have been deleted are also more sensitive to environmental perturbations (Lehner, 2010b). Both results are consistent with a loss of robustness or a ‘de-canalization’ — in the absence of a hub, otherwise harmless fluctuations in gene expression start to have phenotypic effects (Burga et al., 2011; Casanueva et al., 2012). Thus, the stable expression of genetic interaction hubs may serve to buffer genetic, environmental, and stochastic perturbations.

Epigenetic epistasis and evolvability

In summary, the analyses presented here support a model in which there has been a selective pressure to reduce the variability of expression of genes with many genetic interaction partners. Together with previous work showing that expression variation can cause harmful epigenetic epistatic interactions (Burga et al., 2011; Casanueva et al., 2012), and that genetic hubs promote phenotypic stability (Levy and Siegal, 2008; Lehner, 2010b), this suggests that the tight regulation of genetic interaction hubs is an important aspect of phenotypic robustness or canalization. Thus, in the short term the tight regulation of genetic hubs may therefore constrain evolution by reducing phenotypic variation. However, it has also been proposed that over long evolutionary periods increased phenotypic robustness promotes evolution by facilitating a wider exploration of viable genotypes (Wagner, 2005), and so the tight regulation of epistatic interaction hub may facilitate the evolution of complexity. Using laboratory evolution experiments it might be possible to directly test some of these ideas.

Materials and methods

Genetic interactions

Genetic interactions were systematically identified by (Costanzo et al., 2010) for 3458 array genes. Only high-confidence negative interactions with $|e|<0.08$ and $P$-value $<0.05$ were considered, as suggested by the authors (Costanzo et al., 2010). The fitness defects of the same strains were determined by the same authors (Baryshnikova et al., 2010; Costanzo et al., 2010). For an independent test (Supplementary Figure 10), all genetic interaction data sets, except Costanzo et al. (2010) and Tong et al. (2004), were downloaded from the BioGRID (www.thebiogrid.org, version 3.1) (Stark et al., 2011).

Gene expression data sets

Stochastic noise in protein expression was measured from single-cell profiling of fluorescently tagged proteins by Newman et al. (2006). To exclude the influence of protein abundance, the coefficient of variation was converted to the distance-to-median metric (DM) for each gene (Newman et al., 2006). Responsiveness of mRNA expression to environmental perturbations was derived from more than 1500 gene expression profiles (Ihmels et al., 2002), and calculated as the sum of squares of the log2-ratios over the conditions (Tirosh et al., 2006). A second measure of expression variation across environmental conditions was obtained from Edgar et al. (2002) and Koch et al. (2012). Each gene’s percentile of variation was calculated and then assigned to the average percentile across all microarray data sets. Responsiveness across stress conditions was calculated as the variance in gene expression (Casch et al., 2000; Choi and Kim, 2009). Trans viability was defined as the responsiveness to trans-acting genetic loci in recombinant inbred lines (Brem et al., 2002) and was obtained from Brem et al. (2002) and Choi and Kim (2008). Mutual variance ($V_m$) is a measure of expression variance among four mutation accumulation lines by Landry et al. (2007). Inter-strain variation was measured across four natural isolates of S. cerevisiae (Townsend et al., 2003). ED between closely related yeast species under controlled environmental conditions was taken from Tirosh et al. (2006). Gene expression levels of yeast genes were obtained from Holstege et al. (1998).

Promoter features

The presence and absence of TATA boxes was obtained from Basehoar et al. (2004), with TATA boxes identified in promoter regions by scanning each gene’s promoter region (–70 to –310) for the
commonly found site TATA(A/T)(A/T)(A/G). A total of 1090 genes initiating from TATA-box promoters and 4581 genes from non-TATA box promoters are considered. Nucleosome occupancy in promoter regions was taken from in vivo nucleosome occupancy data (Lee et al., 2007) in the 100 base pairs upstream of each gene, as defined in Tirosh and Barkai (2008). According to this criterion, a total of 1082 ‘occupied proximal nucleosome’ (OPN) promoters and 1940 ‘depleted proximal nucleosome’ (DPN) proximal promoters are analyzed.

**Yeast gene duplicates**

Gene duplicates were defined using the SYNERGY algorithm (Wapinski et al., 2007), which uses gene trees based on sequence similarity.

**Expression divergence of human genes**

Human orthologs of yeast genes were obtained from Ensembl Compara through BioMart (Haider et al., 2009; Vilella et al., 2009). ED of human genes was calculated as expression variation between human and chimpanzee across five tissues, as reported in Tirosh et al. (2006) using gene expression from Khaitovich et al. (2005).

**Multifunctionality**

Multifunctionality was defined as the number of ‘biological process’ terms of Gene Ontology (GO) as reported in Costanzo et al. (2010), which uses a multifunctionality index (Myers et al., 2006).

**Phenotypic capacitance**

The phenotypic capacitance was quantified morphological variability (Ohya et al., 2005) upon deletion of nonessential genes and was used directly from Levy and Siegal (2008).

**Haploinsufficient genes**

Overall, 184 haploinsufficient genes were identified by a heterozygous deletion screen (Deutschbauer et al., 2005).

**Supplementary information**

Supplementary information is available at the Molecular Systems Biology website (www.nature.com/msb).

**Acknowledgements**

We thank JS Yang, J Semple and A Burga for helpful discussions, Itay Tirosh for sharing ED data of human genes. This work was funded by the European Research Council (ERC), MINECO Plan Nacional grants BFU2008-03065 and BFU2011-26206, ERASysBio + ERANET project EU2009-04059 CRAPPLE, the European Molecular Biology Organization (EMBO) Young Investigator Program, EU Framework 7 project 277899 4DCellFate, and the EMBL/CRG Systems Biology Program.

**Conflict of interest**

The authors declare that they have no conflict of interest.

**References**


