FILE S1

Sensitivity to evolutionary rate:

To compare evolutionary outcomes under positive selection with pleiotropy to those under stabilizing selection with genetic drift, we compared misregulation of the conserved trait in the pleiotropy model (Figure 1A) and the two-domain model (Figure 1B). We expect hybrid misregulation under pleiotropic constraint to be become increasingly likely under conditions that limit the availability of timely favorable mutations at the *cis*-regulatory locus of the directionally selected trait. Mutations at the TF site, while less beneficial because of their fitness costs at the stabilized trait, could still be better than none at all. Faster change in P_{opt} and decreased population size can impose these conditions.

Methods: We performed simulations at population sizes of 25, 50, 100, 200 and 400, crossed with evolutionary rates ranging from 1/2000 to 1/40000 generations. For this we used alleles of length n = 24 bits and G-P map 'd' of Figure 2. The 24-bit model decreases mutation effect size and allows greater opportunities for compensatory evolution. We varied the evolutionary rate, ΔP_{opt}, by varying the duration of the simulation as P_{opt} of the directionally selected trait changed gradually from 0 to 1.

Results: We found net F2 hybrid misregulation of both the directionally selected and stabilized traits, with an interaction between evolutionary rate and population size that differed between the models. In the directionally selected trait (Figure S1A), misregulation was highest in the two-domain model and constrained in the pleiotropic model. Decreasing population size and increasing evolutionary rate interacted to weaken the pleiotropic constraint. Population size and evolutionary rate had no effect in the two-domain model, where only linkage could contribute to the genetic correlation. These results are for the 24-bit model, and the 12-bit model (not shown) yielded the same pattern with slightly lower HI values overall.

For the trait under stabilizing selection (Figure S1B), F2 hybrid misregulation only appeared in the pleiotropic model, unless population sizes were very small and the evolutionary rate was very slow at the directionally selected trait. At a population size N = 25 and an evolutionary rate of ΔP_{opt} = 1/40000 generations, net F2 misregulation of the conserved trait in the pleiotropic and two-domain models was similar. In the pleiotropic model, high misregulation occurs because the pleiotropic TF locus provides some of the response to selection on the directionally selected trait, selecting in turn for compensatory mutants at the *cis* locus of the conserved trait. For the two-domain model, the misregulation pattern is consistent with compensatory evolution where the *cis*-regulatory locus of the conserved trait evolves essentially independently of the directionally selected locus, with perhaps a small contribution of linkage between domains at the TF site. In other words, the two-domain model behaved as if it were a simple two-locus model under stabilizing selection (TULCHINSKY *et al.* 2014). In contrast, for the pleiotropic model we found the same interaction between population size and evolutionary rate as in Figure S1A.

In some cases, the pleiotropic model showed more net F2 hybrid misregulation at high population size than at low population size (Figure S1B). Total F2 hybrid misregulation in these cases was nevertheless high, due to a large contribution from misregulation in reconstituted parental genotypes. At low population size, genetic load accumulates and can produce substantial misregulation in the parent populations, especially when directional selection to the final optimal phenotype is more rapid (ΔP_{opt} > 1/4000 generations).

Effects of fitness function and mutation effect size:

Binding regions vary in length among types of TFs and their promoters, and we suspected this would affect evolutionary rates and pleiotropic constraints. Under a given set of bioenergetic parameter combinations, mutations with smaller effect sizes at the pleiotropic (TF) locus will also have smaller effects on both downstream phenotypes and their marginal fitnesses. That may increase the chances that substitutions will occur at that locus, and in turn, lessen the pleiotropic constraint to allow more hybrid incompatibility to evolve. Here we test that scenario.

Motif length n is a good proxy for mutation effect size. To illustrate this, consider for reference the 12-bit G-P map or fitness landscape 'd' of Figure 2, which is based on bioenergetic parameter values of N_{TF} = 100, E_{diff} = -1 and ΔG_1 = -0.6125. A single mismatch of this map yields a phenotype of P = 0.978. By manipulating equation 2, we find that a 24-bit map with n = 2 mismatches has the same shape at N_{TF} = 100, E_{diff} = -1 provided the free energy of association is $\Delta G_1/2$.

Methods: The bioenergetic parameters determine differences in expression levels between genotypes, thus the shapes of the G-P map and fitness landscape (text Figure 2). These outcomes should also depend on environmental parameters determining marginal fitness. Higher values of σ_s^2 in the fitness function of a trait correspond to a flatter marginal fitness landscape around its optimum and therefore weaker selection on a given deviation from the optimum. Conversely, higher values of $1/\sigma_s^2$ correspond to a steeper fitness function and stronger selection. We present our results in units of steepness ($1/\sigma_s^2$) rather than flatness (σ_s^2) because we find it more intuitive. To study the effect of the fitness parameters directly, we varied σ_s^2 between 1/8 and 8 times the value used in the body of the paper, σ_s^2 = $2.5 * 10^{-3}$ at each trait.

We tested the role of mutation effect size by using a binding motif of length n = 12, which has a larger mutation effect size than a motif of n = 24 bits. We used the genotype-phenotype (G-P) maps and fitness landscapes labeled 'd' Figure 2, scaled to the 24-bit case by dividing ΔG_1 by 2. The resulting G-P map shapes are identical, except that the horizontal axis extends to 24 rather than 12 mismatches. To study the interaction of mutation rate and mutation effect size, we repeated the analyses of text Figure 4, which is based on the 12-bit motif, using the 24-bit motif. We likewise examined the interaction between selection strength and mutation effect size by comparing results from the 12-bit and 24-bit motifs.

Results: We found that the extent that hybrid misregulation evolved under a given fitness function depended on relative strengths of selection acting through the phenotypic and fitness effects of these parameters.

The effect of the steepness of the fitness function on net F2 misregulation is shown in Figure S2. The strength of selection had no effect on misregulation in the two-domain model (gray bars), but misregulation was constrained to lower levels in the pleiotropic model (white bars). For the directionally selected trait (Figure S2, A and C), the steepness of the fitness function had only a minor effect. However, increasing selection on the stabilized trait increased the pleiotropic constraint, such that misregulation decreased and ultimately disappeared (Figure S2, B and D). Reducing the mutation effect size by adopting a 24-bit motif had no effect in the two-domain model. In the pleiotropic model, it only slightly reduced the constraint on misregulation imposed by the 12-bit motif.

A similarly small influence of mutation effect size appeared in relation to variation in mutation rate. In the pleiotropic model, decreasing the mutation effect size by setting n = 24 lessened the constraint on the evolution of misregulation relative to the 12-bit case (Figure S3 A and B, vs. Figure 4 A and B). Mutation effect size and mutation rate interacted, such that median net misregulation in the 24-bit case was 1.2x higher at the lowest mutation rate and 5x higher at the highest mutation rates. The interaction effect on misregulation of the conserved trait was similar but not as pronounced (Figure S3 C vs. Figure 4C). In the two-domain model, mutation effect size again played no role in the extent of misregulation.

These results confirm our prediction that lowing mutation effect sizes will reduce the pleiotropic constraint on hybrid incompatibility, permitting more to evolve.

Earlier work by Johnson and Porter (2007) examining the evolutionary dynamics of a pleiotropic locus that simultaneously regulates positively selected and conserved traits also found that increasing the strength of stabilizing selection decreased divergence at the pleiotropic locus. In that model, which did not include bioenergetic parameters, the pleiotropic locus experienced divergence sufficient to produce hybrid incompatibility even under strong stabilizing selection. The following differences in model assumptions likely explain this result. First, the mutation effect size was effectively smaller in the Johnson and Porter (2007) model, in that mutation effects followed a Gaussian distribution with no minimum effect size. This allowed compensatory evolution to proceed in smaller steps. Second, because of a detail in how allelic values were represented, a large fraction of all potentially compensatory mutations successfully compensated for a given pleiotropic mutation. Though compensation is more difficult in the bioenergetic model, our results in general agree with Johnson and Porter (2007) that selection due to pleiotropy drives the evolution of HI in a conserved regulatory interaction under a broad range of parameter values (Figures 3 and S1).

Effects of other bioenergetic parameters:

Motif length is a good proxy for mutation effect size as we described in the section above, provided we substitute $\Delta G_1/2$ for ΔG_1 of the 12-bit G-P map. The same logic applies if we substitute for other bioenergetic parameters, and the same outcomes derive from those effects. The same phenotype of P = 0.978 for G-P map 'd' in Figure 2 can be found for a 24-bit motif by changing any of the bioenergetic parameters. For example, a 24-bit G-P map with this phenotype at $\Delta G_1/2$, but having only n = 1 mismatch, exists when E_{diff} = -1.896. We also get the same phenotype using E_{diff} = -1 and ΔG_1 = -0.6125 on the 24-bit map, provided that N_{TF} = 41.

The result is that the same types of mutation-effect constraints apply, but scaled differently. For example, the effect of a single substitution in a 12-bit motif with N_{TF} = 100 can be achieved with a single substitution in a 24-bit motif and N_{TF} = 41. Roughly, doubling the genome size will produce the same E_{diff} for a 24-bit motif as you would find in a 12-bit motif. Thus, re-regulating to get the same evolutionary effect of a substitution at a TF or *cis*-regulatory site can be achieved by down-regulating N_{TF} through upstream evolutionary change.