Extreme robustness requires extreme phenotypes

Felipe Bastos Rocha[1*] and Louis Bernard Klaczko[2]

1 Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo – USP, SP, Brazil.

2 Departamento de Genética, Evolução e Bioagentes, Universidade Estadual de Campinas – Unicamp, Campinas, Brazil.

* Corresponding author:

e-mail: rochaf@ib.usp.br

An outstanding feature of living organisms is their ability to develop in shifting and often perturbing conditions and still yield a defined set of traits with stable phenotypes. Here we show that this ability – phenotypic robustness – may be achieved via a possibly universal mechanism that emerges from basic rules relating gene activity and development. First, we demonstrate that phenotypic stability is strongly associated with extreme phenotypes using a comprehensive dataset of reaction norms of plants, animals, and bacteria from the literature[1]. This association suggests that genetic variation affects robustness mainly through saturation or depletion – and not stabilization – of developmental systems. Second, we build a simple developmental model to assess how genetic variation for gene activity may affect the pattern of phenotypic response of a given trait to perturbations of transcription. The results show that organisms may achieve phenotypic robustness in the absence transcriptional stability by yielding extreme phenotypes. Third, we show that this model may offer a simple and coherent explanation for experimental results (from Crocker et al.[2]) associating *cis*-regulatory sequence variation to phenotypic robustness. These results suggest that, even though specific systems of developmental buffering may play a role in phenotypic robustness, natural selection often resorts to the genetic variation controlling the level of trait induction to yield phenotypic stability.

Gene activity is the primary trigger of the developmental processes that lead to the formation of a trait. Such processes are influenced by the internal cellular environment, the individual's genetic makeup, and the surrounding environment. Hence, the final phenotype of an individual depends on the resultant of these factors – hereafter the *level of trait induction*. In contrast to laboratory mutants, wild-type organisms can typically fully develop and yield fairly constant phenotypes even when submitted to variable genetic backgrounds and environmental conditions. At least since Waddington[3], this observation is seen as evidence that wild-type organisms can compensate or avoid changes in the level of trait induction when submitted to perturbations (Fig. 1a, b). Typically, this is assumed to be due to specific systems that evolved by natural selection to perform such buffering[4–6]. However, with few exceptions (e.g., the chaperone Hsp90[7,8]), genes whose function is to confer phenotypic stability against perturbations are seldom found.

Is it possible to achieve phenotypic robustness in the absence of buffering genes? A possible answer lies in the effect that two developmental properties have on the relationship between trait induction and phenotype formation (see Fig. 1c – analogous rationales are found in Refs[9,10] and date back to Wright[11] and Plunkett[12]). The first is the fact that no trait can increase indefinitely. The second is the fact that most traits require the level of induction to reach a non-zero value to yield phenotypic increments from the minimum possible phenotype. These properties create two regions in the induction-phenotype curve where changes in the level of trait induction, regardless of their cause, will not lead to phenotypic changes because the developmental system is either depleted or saturated (see Fig. 1c).

What is the relevance of this dynamics of saturation/depletion in the observed variation of robustness? To address this question, one could assess how the amount of response to a given perturbation varies across the phenotypic space: while saturation/depletion will necessarily yield phenotypic robustness only near extreme phenotypes, buffering may potentially confer robustness at any phenotypic value within the limits of the phenotypic space.

Following this rationale, we analyzed phenotypic reaction norms – the arrays of phenotypes produced by different genotypes in response to a given array of environmental conditions – from a dataset compiled by Murren et al.[1]. After filtering (see Methods) we analyzed 856 reaction norms from 29 plants, 43 animal species and three bacteria. To make these diverse responses comparable, we first normalized all phenotypic values by species and trait and all environmental values by experiment, so that all reaction norms fitted a 1x1 plot. Then, to circumvent the complexity of forms that reaction norms may assume (see [13,14]), we used Global Plasticity (GP), a shape-independent parameter[15] to quantify the total amount of phenotypic response in each reaction norm. Thus, we characterized phenotypic robustness, regardless of trait, species, and experiment, as near-zero GP values, and high phenotypic plasticity as near-one GP values. For each trait in each species, we estimated the phenotypic space by the interval of values between the maximum and minimum observed phenotypes. Then, to characterize the position of each reaction norm in the phenotypic space, we took the distance between the across-environment mean phenotype and the closest extreme phenotype (see Methods for details). Thus, the relevance of saturation/depletion could be assessed by whether GP varied independently from the distance from the limits of the phenotypic space.

Despite the extensive heterogeneity of organisms, traits, and perturbations included in this analysis, more than 50% of GP variation could be explained by a linear function that increased from near-zero (high robustness) to near-one (high plasticity) as the distance from the trait limits varied from minimal to maximal ($R^2 = 0.520$; $F = 930.122$; $p < 0.00001$ - Fig. 2a). Likewise, extreme robustness, as characterized by reaction norms with GP below 0.2 (i.e., whose response was lower than 1/5 of the phenotypic space), were strongly concentrated towards the phenotypic extremes (Fig. 2b – $D = 0.309$, $p < 0.001$). These findings are substantial evidence that the saturation/depletion dynamics is the leading cause of phenotypic robustness in this dataset. This scenario is somewhat conflicting with the paradigmatic view of organisms as masters of their

development. Instead, it might be the case, as Smith-Gill[16] once stated, that organisms may often be doing the best they can to minimize environmentally induced changes, and failure to do so could often be what we call phenotypic plasticity.

How can such a clear pattern emerge from such a diverse dataset? Recent studies have repeatedly found that *cis*-regulatory sequences affect phenotypic robustness following a consistent pattern in various systems. Genotypes enriched for *cis*-regulatory sequences that drive similar patterns of expression (shadow enhancers and homotypic clusters of transcription factor binding sites) yield stability of high phenotypic values[2,17–23]. Conversely, stability of low phenotypic values seems to be achieved by the loss of such sequences and gain of sequences with negative effect on gene activity[24]. These results are remarkably consistent with the pattern of reaction norm variation we found and imply a fundamental biological principle – the direct control of trait induction by gene activity – in the control of phenotypic robustness by saturation/depletion.

Oddly, however, the relationship between the multiplicity of functionally similar *cis*-regulatory sequences and robustness may work in the opposite direction for transcription at the whole-genome scale[25]. To assess how genetic variation for gene activity may shape the patterns of robustness variation at the transcriptional and phenotypic levels, we built a simple model of trait development that generates phenotypic reaction norms from the effect of a perturbation that directly affects transcription. Briefly, our model describes a trait formed by a limited group of cells that differentiate when an inducer gene is expressed at or above a threshold. Perturbation was introduced by linearly reducing the probability of transcription initiation at a constant rate for all genotypes. Given the stochastic nature of gene activity[26,27] and evidence that multiple similar *cis*-regulatory sequences cumulatively increase their target gene activity[28–30], genetic variation was introduced by varying the probability of transcription initiation at optimal conditions. We simulated both transcriptional and phenotypic reaction norms (Fig. 3a and b) and then analyzed

their patterns of GP variation as a function of the level of trait induction, as given by mean probability of gene activation.

Both transcriptional and phenotypic GP values increased towards genotypes with intermediary levels of trait induction and decreased towards extreme levels of trait induction (Fig. 3c). This feature was independent from the developmental setup and the intensity of perturbation (Extended Data Fig. 1). Nonetheless, despite the similarity of forms, the curves of GP variation for transcription and phenotype were not related one-to-one, as many transcriptionally plastic genotypes were phenotypically robust (see Fig. 3c). This pattern arises because phenotypic stability was not the result of stability of transcription, but rather the outcome of the saturation or depletion of the developmental system leading to the trait formation (Fig. 3d). Thus, a genotype with stable phenotype is one whose transcription, even if perturbed, is either below the depletion threshold or above the saturation threshold, not necessarily one that is transcriptionally robust to perturbations (see Figs. 1b and 3d). Of course, the same principle pertains to any kind perturbation that might affect a gene's activity, and a genotype that is phenotypically robust due to saturation or depletion will be robust to multiple forms of perturbations: within-locus genetic perturbation (where robustness equals dominance), between-loci genetic perturbation, and multi-variate environmental perturbation.

We then set out to evaluate whether this model could fit to the pattern of robustness variation that is caused by multiplicity of similar *cis*-regulatory sequences. We used RNs described by Crocker et al.[2], which portray the response of the number of larval trichomes of eight genotypes of *Drosophila melanogaster* to temperature. Each reaction norm corresponds to a genotype with a specific arrangement of homotypic binding sites for the Ubx-Exd transcription factors in an enhancer driving expression of the gene *shavenbaby* (*svb*). To simulate the GP values using our model, we assumed that all genotypes suffered the same impact of temperature on *svb* activity and the only cause of response variation was saturation/depletion. The model offered an excellent

prediction of the observed GP values (R^2 = 0.829 – Fig. 4a), suggesting that the observed amount of response of the number of larval trichomes varied mainly due to saturation or depletion. This result suggests that the number of Ubx binding sites in the enhancer had a cumulative (rather than redundant) effect on the level of induction of larval trichomes. Accordingly, the across-temperature mean phenotype was significantly correlated with the number of unmutated Ubx-Exd binding sites (Fig. 4b). These findings seem to diverge from the authors' interpretation that the observed variation in phenotypic robustness was due to a positive association between redundancy (multiplicity) of Ubx binding sites and robustness of gene activity.

We may thus outline a coherent scenario of how natural selection may shape regulatory variation when robustness is favored. Frequently (and as observed), selection for phenotypic robustness may lead to changes in the level of trait induction via modulation of gene activity. *Cis*-regulatory sequences of genes promoting traits will be enriched for shadow enhancers and homotypic binding sites when high phenotypic values are favored. This pattern will be inverted when low phenotypes are selected. In some (perhaps rare) cases, however, two different outcomes of selection for robustness may be observed. First, genes whose activity inhibits the trait formation will display opposite *cis*-regulatory patterns: enrichment of enhancers and TF binding sites when the favored phenotype is low and reduction when the favored phenotype is high. Second, if selection favors robustness of intermediary phenotypes, alleles creating narrower phenotypic limits will be favored, which could be achieved via trans-regulatory changes. Lastly, if selection for phenotypic constancy is ubiquitous and there is, as it seems, a universal mechanism to achieve this property, selection for robustness may be a major force shaping non-coding regulatory DNA.

Methods

Analysis of reaction norms extracted from the dataset compiled by Murren et al. (2014a):

We retrieved the data compiled by Murren et al.[1] from Dryad[31] and filtered for reaction norms that had continuous (non-qualitative) environmental variables and more than one reaction norm by trait in each species. This preliminary assessment resulted in a dataset of 858 reaction norms. We then normalized all environmental values by experiment and all phenotypic values by trait and species so that all reaction norms were fitted within a 1x1 plot. Global Plasticity values were estimated by the reaction norm range (the difference between the maximum and minimum phenotypes observed in each reaction norm). For each species-specific trait, the upper and lower extreme values were estimated by the maximum and minimum observed phenotypes regardless of genotype or environment. To characterize how distant each genotype was from either limit of the phenotypic space, we calculated the difference between the across-environment mean phenotype of each reaction norm and the closest extreme value as $(0.5 - |\bar{P} - 0.5|)$, where $\bar{P}$ is the across-environment mean phenotype of each reaction norm. We tested the dependency of Global Plasticity on the distance from the closest extreme phenotype by performing a least-squares linear regression and the distribution of highly robust genotypes (i.e., the reaction norms with GP lower or equal to 0.2) against a uniform distribution using a Kolmogorov-Smirnov test.

Model

The model describes a trait that is formed by a set of cells that may or may not follow a differentiation path; the individual phenotype is given by the number of cells that differentiate. Cell differentiation is triggered by the expression of an inducer gene (IG) at or above a threshold value (t). Each individual has a constant number of 100 cells that may or not express the IG, representing a constant trans-regulatory landscape of cells that constitutively express the transcription factor (TF) necessary for the IG activity. There is a developmental time-window in

which the IG expression may trigger differentiation. The maximum number of times each cell may or may not activate the transcription of the IG (rounds of activation) during this developmental time window is given by n. Whether transcription is activated at each round of activation depends on the probability of activation (p). This probability is given by a negative linear function of an environmental variable (E) that varies from 0 to 1 with eleven values used to estimate each genotype's response. The equation of environmental perturbation is $p = A + s\,E$, where A is the probability of transcription activation at $E = 0$ and s is the slope that determines how the environmental variable affects the A; p was limited to vary between zero and one. The realized transcription (RT) in each cell is given by the product of the number of times that IG is activated (k) within the developmental time-window, i.e., we set the rate of transcription to be a constant. Therefore, the phenotypic value of an individual is given by the number of cells activating the IG at or above the minimum number of activations necessary to reach t. The probability of each cell to activate IG k times was given by the binomial distribution: $Pr(k; n, p) = \binom{n}{k} p^k (1-p)^{n-k}$. Since p is constant for each genotype in each environmental value, we may assume that the proportion of cells activating IG k times in an infinite number of individuals raised in each environmental value is given by the binomial probability $Pr(k;n,p)$.

Therefore, transcriptional reaction norms were simulated by calculating the mean Realized Transcription ($\overline{RT}$) as the mean number of activations ($p \cdot n$). The mean phenotype in a given environment may be calculated from the proportion of cells activating IG at or above $k_{\min}$. We calculated this from the cumulative probability function of the binomial equation:

$$\bar{P} = Pr(X \geq t) = 1 - \left(\sum_{i=0}^{\lfloor k_{min} \rfloor} \binom{n}{i} p^i (1-p)^{n-i}\right).$$

We explored different developmental setups (given by t and n) using, in each setup, a variation of A that was sufficient to yield both phenotype saturation and depletion for s values ranging from 0 to -1 (Extended Data Fig. 1a gives an example for $t = 50$ and $n = 101$).

Adjustment to the dataset extracted from Crocker et al.[2]'s reaction norms:

In order to fit our model to the set of reaction norms from Crocker et al.[2], we captured each reaction norm point available in Figure 6 from their study using tps.dig[33] and converted the values to normalized phenotypic and environmental values. As the reaction norms in their data are non-monotonic and our model does not consider changes in the reaction norm direction, we analyzed only the lower segment (17-25° C). We then used the localized slope of this segment of each reaction norm (Local Plasticity – LP, from Ref[32]) to estimate GP.

In our model, each specific slope of perturbation (s) generated a specific curve of GP variation as a function of the genotype's level of trait induction, which may be characterized by the mean phenotype across different conditions. These curves converge at extreme phenotypes (due to saturation and depletion) and diverge at intermediary mean phenotypes, where the peak of GP values occur (Extended Data Fig.1b). Therefore, we took the reaction norm with most intermediary mean phenotype and determined which value of s was necessary to yield a curve of equivalent GP, following a linear variation of GP by s when the mean phenotype is 0.5 in the model ($s = 0.0015216 - 0.134379\ GP$). We then used this s (-0.062813) to calibrate our model (keeping t at 50 and n at 101) and estimate GP for the remaining genotypes (i.e., those not involved in the model calibration) based on their mean phenotype only. We calculated the R^2 between observed and predicted values to evaluate the goodness of fit of our model to their experimental results.

Data availability

The set of data extracted from the reaction norms compiled by Murren et al. 2014 used in the analysis is available in Supplementary Table 1. Additional data supporting the findings of this study are available from the corresponding author upon reasonable request.

References

1. Murren, C. J. *et al.* Evolutionary Change in Continuous Reaction Norms. *Am. Nat.* **183,** 453–467 (2014).

2. Crocker, J. *et al.* Low affinity binding site clusters confer HOX specificity and regulatory robustness. *Cell* **160,** 191–203 (2015).

3. Waddington, C. H. Canalization of development and the inheritance of acquired characters. *Nature* **150,** 563–565 (1942).

4. Debat, V. & David, P. Mapping phenotype: canalization, plasticity and developmental stability. *Trends Ecol. Evol.* **16,** 555–561 (2001).

5. Flatt, T. The Evolutionary Genetics of Canalization. *Q. Rev. Biol.* **80,** 287–316 (2005).

6. Queitsch, C., Carlson, K. D. & Girirajan, S. Lessons from Model Organisms: Phenotypic Robustness and Missing Heritability in Complex Disease. *PLoS Genet.* **8,** e1003041 (2012).

7. Queitsch, C., Sangstert, T. A. & Lindquist, S. Hsp90 as a capacitor of phenotypic variation. *Nature* **417,** 618–624 (2002).

8. Rutherford, S. L. & Lindquist, S. Hsp90 as a capacitor for morphologival evolution. *Nature* **396,** 336–342 (1998).

9. Félix, M. A. & Barkoulas, M. Pervasive robustness in biological systems. *Nat. Rev. Genet.* **16,** 483–496 (2015).

10. Meiklejohn, C. D. & Hartl, D. L. A single mode of canalization. *Trends Ecol. Evol.* **17,** 468–473 (2002).

11. Wright, S. Physiological and Evolutionary Theories of Dominance. *Am. Nat.* **68,** 24–53 (1934).

12. Plunkett, C. R. A contribution to the theory of dominance. in *Am Nat* 84–85 (1933).

13. Rocha, F., Medeiros, H. F. & Klaczko, L. B. The reaction norm for abdominal pigmentation and its curve in Drosophila mediopunctata depend on the mean phenotypic value. *Evolution (N. Y).* **63,** 280–287 (2009).

14. Rocha, F. B. & Klaczko, L. B. Connecting The Dots Of Nonlinear Reaction Norms Unravels The Threads Of Genotype-Environment Interaction In Drosophila. *Evolution (N. Y).* **66,** 3404–3416 (2012).

15. Rocha, F. B. & Klaczko, L. B. Untangling the Components of Phenotypic Plasticity in Nonlinear Reaction Norms of Drosophila mediopunctata Pigmentation. *bioRxiv* (2016). doi:10.1101/070599

16. Smith-Gill, S. J. Developmental Plasticity: DevelopmentalConversion versus Phenotypic Modulation. *Amer Zool* **23,** 47–55 (1983).

17. Perry, M. W., Boettiger, A. N., Bothma, J. P. & Levine, M. Shadow enhancers foster robustness of Drosophila gastrulation. *Curr. Biol.* **20,** 1562–1567 (2010).

18. Frankel, N. *et al.* Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature* **466,** 490–493 (2010).

19. Ludwig, M. Z., Manu, Kittler, R., White, K. P. & Kreitman, M. Consequences of eukaryotic

enhancer architecture for gene expression dynamics, development, and fitness. *PLoS Genet.* **7,** e1002364 (2011).

20. Cannavò, E. *et al.* Shadow Enhancers Are Pervasive Features of Developmental Regulatory Networks. *Curr. Biol.* **26,** 38–51 (2016).

21. Kitano, H. Biological robustness. *Nat. Rev. Genet.* **5,** 826–837 (2004).

22. Payne, J. L. & Wagner, A. Mechanisms of mutational robustness in transcriptional regulation. *Front. Genet.* **6,** 1–10 (2015).

23. Osterwalder, M. *et al.* Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554,** 239–243 (2018).

24. Preger-Ben Noon, E., Davis, F. P. & Stern, D. L. Evolved Repression Overcomes Enhancer Robustness. *Dev. Cell* **39,** 572–584 (2016).

25. Paixão, T. & Azevedo, R. B. R. Redundancy and the Evolution of *Cis* -Regulatory Element Multiplicity. *PLoS Comput. Biol.* **6,** e1000848 (2010).

26. Walters, M. C. *et al.* Enhancers increase the probability but not the level of genel of gene expression. *Proc. Natl. Acad. Sci.* **92,** 7125–7129 (1995).

27. Fukaya, T., Lim, B. & Levine, M. Enhancer Control of Transcriptional Bursting. *Cell* **166,** 358–368 (2016).

28. Sharon, E. *et al.* Inferring gene regulatory logic from high-throughput measurements of thousands of systematically designed promoters. *Nat. Biotechnol.* **30,** 521–530 (2012).

29. Smith, R. P. *et al.* Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat. Genet.* **45,** 1021–1028 (2013).

30. Lam, D. D. *et al.* Partially Redundant Enhancers Cooperatively Maintain Mammalian Pomc Expression Above a Critical Functional Threshold. *PLoS Genet.* **11,** e1004935 (2015).

31. Murren, C. J. *et al.* Evolutionary change in continuous reaction norms. Dryad Digital Repository.

32. Rocha, F. B. & Klaczko, L. B. Undesirable consequences of neglecting nonlinearity: Response to comments by Liefting et al. (2013) on rocha & klaczko (2012). *Evolution (N. Y).* **68,** 1548–1551 (2014).

33. Rohlf, F. J. tpsDig, digitize landmarks and outlines. (2017).

Acknowledgements

We thank the financial support of the following: Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Fundação de Amparo ao Ensino, à Pesquisa e à Extensão (FAEPEX), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP). FBR was supported by a postdoctoral fellowship from FAPESP (processo nº 2013/04980-0) and a fellowship from the Programa Nacional de Pós-Doutorado from CAPES. FBR thanks F. Uno, I.M. Ventura and Prof. M. D. Vibranovski for comments on the manuscript.

Author Contributions

FBR conceived and carried out the analyses of phenotypic reaction norms, designed the model, performed the numerical simulations and adjustments and wrote the manuscript.

LBK supervised the analysis of reaction norms and contributed to the final version of the manuscript.

Author Information

Reprints and permissions information is available at www.nature.com/reprints
The authors declare no conflict of interest.
Correspondence and requests for materials should be addressed to FBR (rochaf@ib.usp.br).

Figures

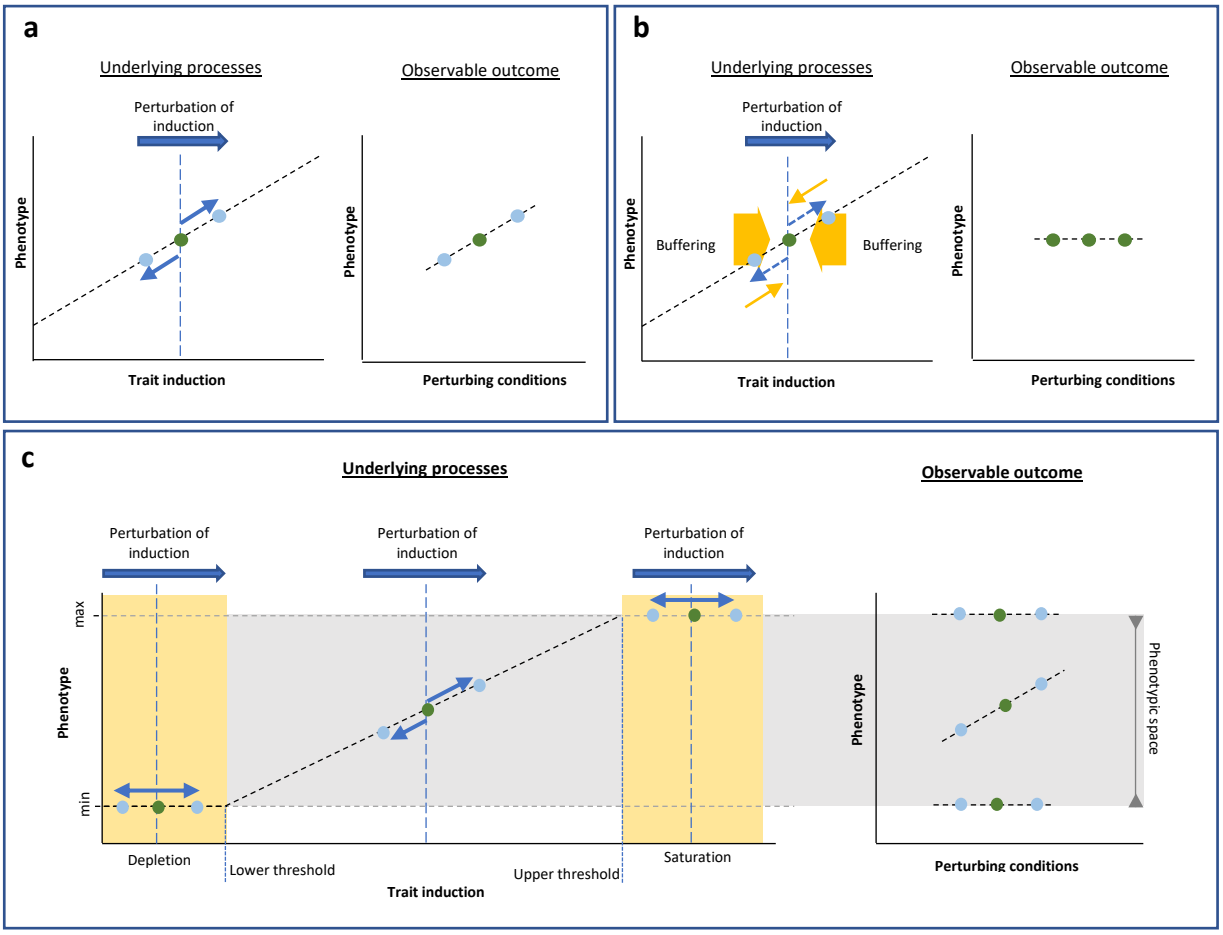

Figure 1. Phenotypic robustness may be the outcome of two contrasting underlying mechanisms. a, perturbations of the level of trait induction (left) due to alleles or environments that either inhibit or promote the trait's formation are prone to yield observable phenotypic responses (right). **b**, developmental buffering (left), which avoids (or cancels out) the changes in the level of trait induction potentially caused by such perturbations are often thought to underlie observable phenotypic robustness (right). **c**, when there is no such buffering system, basic developmental principles entail a nonlinear relationship between trait induction and the phenotype (left). This creates two regions of phenotypic robustness where even individuals submitted to perturbing conditions will be phenotypically stable. The same perturbation, when applied to genotypes conferring intermediary levels of trait induction will yield an unavoidable phenotypic response, thus yielding a pattern of association between the amount of phenotypic response and the position of a genotype in the phenotypic space (right).

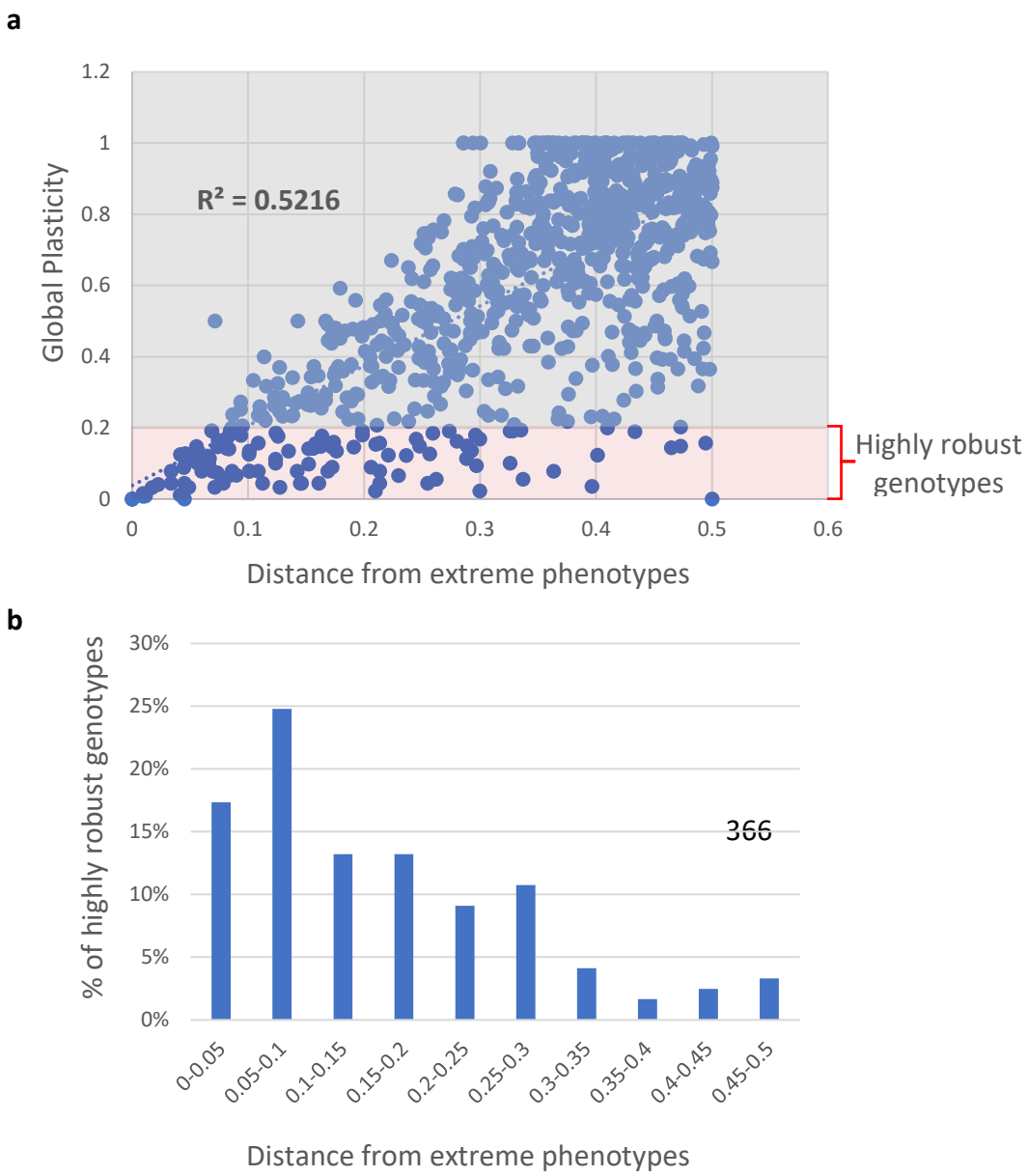

Figure 2. Phenotypic robustness variation in a large sample of 856 reaction norms of diverse animal, plant and bacteria species is mostly explained by the position of the genotype in the phenotypic space. a, observed Global Plascitiy increases with the distance between the genotype's mean phenotype and the trait limits in the dataset extracted from Murren et al. (2014). **b**, highly robust genotypes are significantly concentrated near the phenotypic extremes in this sample of reaction norms.

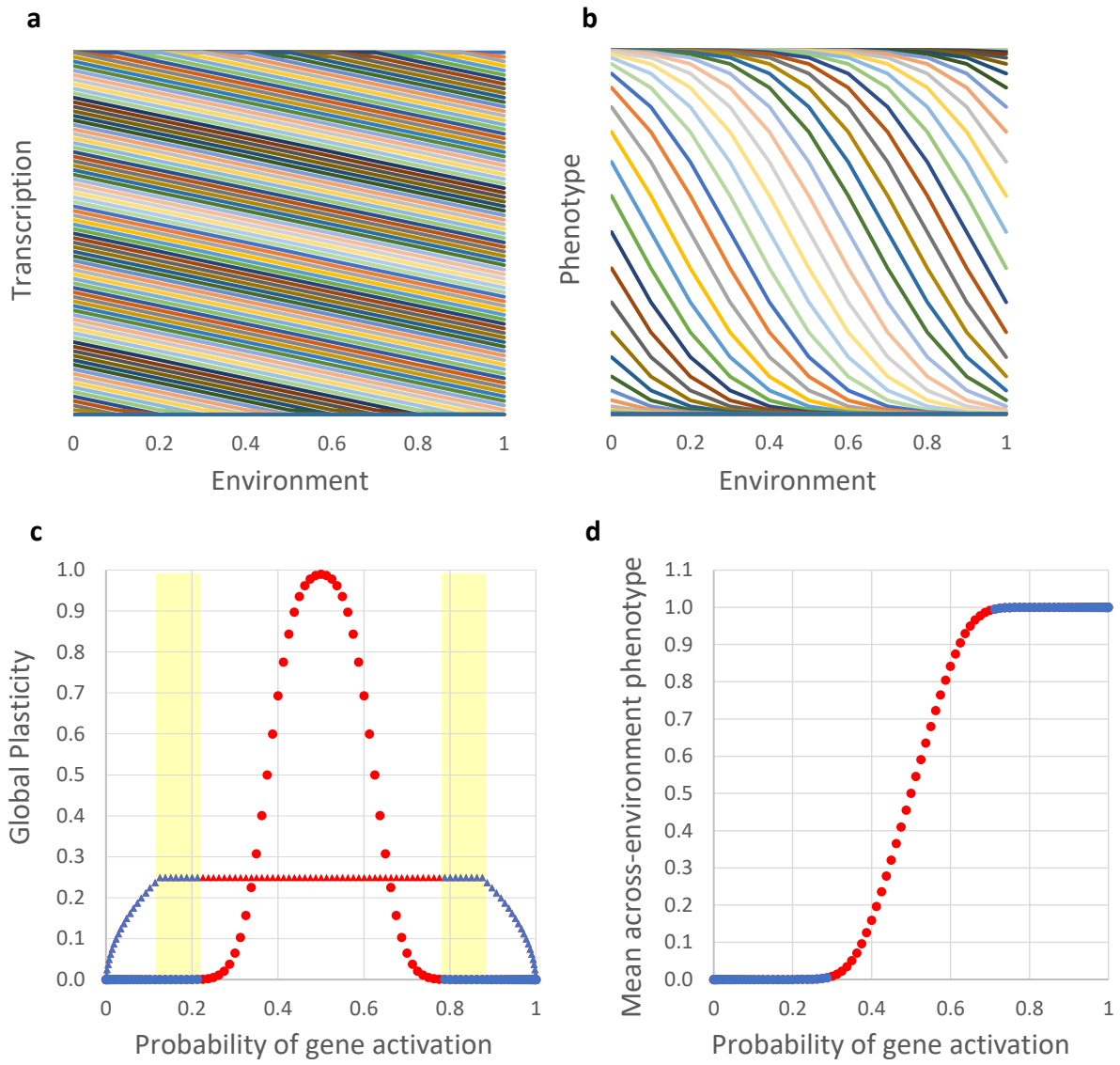

Figure 3. A simple model of trait development shows that phenotypic robustness may be achieved without transcriptional robustness. a, simulated transcriptional reaction norms. **b,** simulated phenotypic reaction norms. **c,** Global Plasticity variation for transcriptional (triangles) and phenotypic (circles) reaction norms for genotypes with varying probability of transcription initiation. Yellow shaded areas highlight "genotypes" with total phenotypic robustness (zero GP) despite having maximum transcriptional plasticity. **d,** mean phenotype as a function of probability of gene activation gives a depletion/saturation curve due to the minimal threshold of gene activity and the limitedness of the phenotype. **c** and **d**, red – "genotypes" with phenotypically responsive reaction norms; blue – "genotypes" with phenotypically robust reaction norms. Values shown correspond to t = 50, n = 101 and s = -0.25.

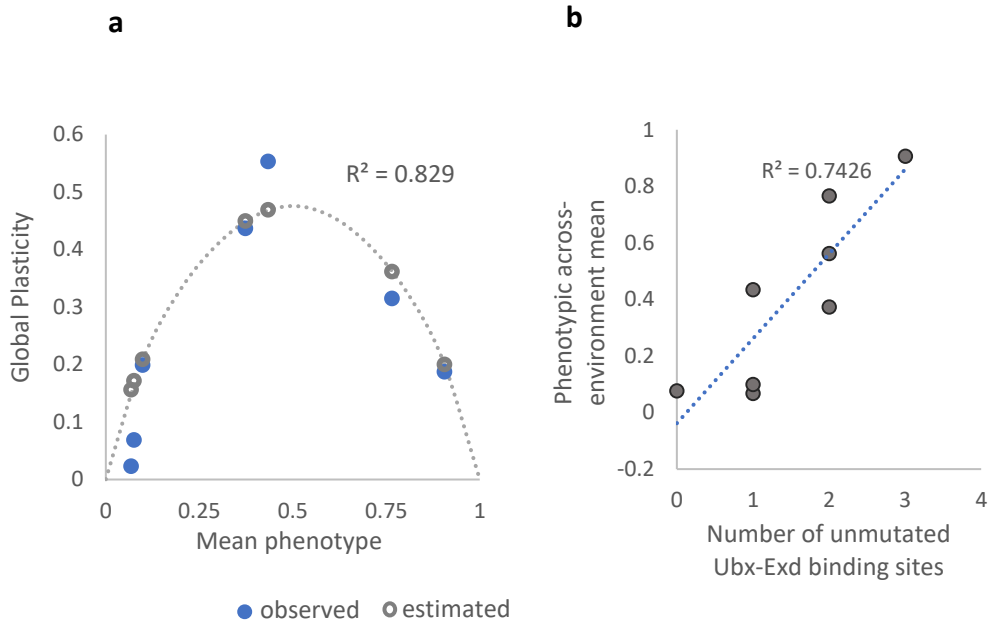

Figure 4. A model of robustness by saturation and depletion explains the pattern of association between *cis*-regulatory multiplicity and robustness described by Crocker et al. (2015). a, a model adjustment assuming that all genotypes suffer the same impact of temperature on transcription allows to describe 84% of Global Plasticity variation as a function of the across-environment mean phenotype. **b**, the across-environment mean phenotype is positively associated with the number of unmutated Ubx-Exd binding sites ($R^2 = 0.70$; $p = 0.005942$). [Data extracted from Crocker et al. (2015) Fig. 6.]