

Robustness and evolvability of heterogeneous cell populations

Andrei Kucharavy^{a,b,c}, Boris Rubinstein^d, Jin Zhu^{a,b}, and Rong Li^{a,b,*}

^aCenter for Cell Dynamics, Department of Cell Biology, Johns Hopkins University School of Medicine, Baltimore, MD 21205; ^bDepartment of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, MD 21218; ^cUMR 7238 CNRS, Université Pierre et Marie Curie, Paris 75006, France; ^dStowers Institute for Medical Research, Kansas City, MO 64110

ABSTRACT Biological systems are endowed with two fundamental but seemingly contradictory properties: robustness, the ability to withstand environmental fluctuations and genetic variability; and evolvability, the ability to acquire selectable and heritable phenotypic changes. Cell populations with heterogeneous genetic makeup, such as those of infectious microbial organisms or cancer, rely on their inherent robustness to maintain viability and fitness, but when encountering environmental insults, such as drug treatment, these populations are also poised for rapid adaptation through evolutionary selection. In this study, we develop a general mathematical model that allows us to explain and quantify this fundamental relationship between robustness and evolvability of heterogeneous cell populations. Our model predicts that robustness is, in fact, essential for evolvability, especially for more adverse environments, a trend we observe in aneuploid budding yeast and breast cancer cells. Robustness also compensates for the negative impact of the systems' complexity on their evolvability. Our model also provides a mathematical means to estimate the number of independent processes underlying a system's performance and identify the most generally adapted subpopulation, which may resemble the multi-drug-resistant "persister" cells observed in cancer.

Monitoring Editor
Alex Mogilner
New York University

Received: Jan 31, 2018
Accepted: Mar 27, 2018

INTRODUCTION

Biological systems, even on the cellular level, are complex systems whose behaviors are emergent from the interaction of a large number of components (Kauffman, 1993; Chu, 2008; Mitchell, 2011). This complexity is thought to endow biological systems with both

stability in the face of perturbations (Carlson and Doyle, 2002) and the ability to adapt to the changing environment (de Visser *et al.*, 2003; Gerhart and Kirschner, 2007). Understanding the relationship between these seemingly conflicting properties can directly impact our ability to interpret the response of cellular systems to experimental manipulation or therapeutic intervention (Wagner, 2008; Draghi *et al.*, 2010). Adding to the difficulty is the fact that cellular systems are often heterogeneous on the population level due to stochasticity in the stoichiometry of many components (Eldar and Elowitz, 2010; Stewart-Ornstein *et al.*, 2012) as well as accumulated genetic variation (Lujan *et al.*, 2014), making the response of these systems to experimental or clinical manipulations difficult to predict (Altschuler and Wu, 2010). Cancer is a prime example of how a combination of heterogeneity, robustness, and evolvability makes a biological system particularly challenging to understand and intervene in. Cancer cell populations are notoriously heterogeneous (Alizadeh *et al.*, 2015) and carry a slew of genetic and copy number abnormalities affecting hundreds to thousands of genes (Giam and Rancati, 2015; Alvarez *et al.*, 2016). Unpredictable responses of tumors to targeted treatments and the frequent emergence of drug

This article was published online ahead of print in MBoC in Press (<http://www.molbiolcell.org/cgi/doi/10.1091/mbc.E18-01-0070>) on April 5, 2018.

The authors declare no competing financial interests.

Author contributions: A.K. and B.R. developed theoretical underpinning and performed the simulation; A.K. developed software and analyzed the primary data for *Saccharomyces cerevisiae* and breast cancer; J.Z. generated the primary data for *S. cerevisiae*; A.K. and R.L. wrote the manuscript; R.L. conceived and directed the project.

*Address correspondence to: Rong Li (rong@jhu.edu).

Abbreviations used: ESI, environment-specificity index; FGM, Fisher's geometrical model; IC50, inhibitory concentration 50%; OD, optical density.

© 2018 Kucharavy *et al.* This article is distributed by The American Society for Cell Biology under license from the author(s). Two months after publication it is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

"ASCB®," "The American Society for Cell Biology®," and "Molecular Biology of the Cell®" are registered trademarks of The American Society for Cell Biology.

resistance are major challenges in cancer treatment (Bozic *et al.*, 2012, 2013; Foo and Michor, 2014; McGranahan *et al.*, 2015; Ramirez *et al.*, 2016). While cancer is arguably the most notable example, similar challenges due to the interaction of heterogeneity, robustness, and evolvability arise in infectious diseases involving heterogeneous populations of microorganisms, from bacteria (Taubes, 2008; Deris *et al.*, 2013) to HIV (Pennings, 2012).

Rapidly evolving diseases could be better understood and treated if we have quantitative models more explicitly showing the interaction between heterogeneity, robustness, and evolvability. Whereas the tendency over the past decade has been to use network systems biology (Ideker *et al.*, 2001; Barabási and Oltvai, 2004) combined with graph theory to explore the interplay between these three factors (Draghi *et al.*, 2010; Raman and Wagner, 2011; Wagner, 2012), we endeavor to derive a model from more fundamental considerations. Cancer genomes carry not only massive gene mutation loads but also numerical and structural chromosome abnormalities (Gordon *et al.*, 2012; Weinstein *et al.*, 2013). These abnormalities can emerge rapidly (Ben-David *et al.*, 2017), and numerical abnormalities, referred to as aneuploidy, exert large effects on multiple cellular pathways in mitotically proliferating cell populations (Chen *et al.*, 2012; Gordon *et al.*, 2012; McGranahan *et al.*, 2012; Zhu *et al.*, 2012; Merino *et al.*, 2016). In a previous study of aneuploid yeast, we observed that the variance of fitness in a heterogeneous aneuploid cohort escalated with increasing stress magnitude regardless of stress type, leading to the emergence of adaptive variants even under the most toxic conditions (Chen *et al.*, 2015). Simulations based on simple assumptions linked to systems complexity and aneuploidy's broad effect on gene expression recapitulated this phenomenon; however, the model was not general enough to allow exploration of functional parameters that affect the adaptive potential of heterogeneous populations. Here we establish a general mathematical framework with analytical formulas and demonstrate its usefulness for investigating the fundamental principles governing the evolvability of complex heterogeneous systems.

RESULTS

A general model of adaptation of heterogeneous populations to stressful environments

Our modeling approach is based on the notion that the fitness of cells in each environment depends on their intermediate properties—intermediate traits representing molecular pathways that enable cell populations to maintain fitness in diverse environmental conditions. Biological systems are characterized by a high degree of modularity despite the fact that most pathways are connected (Hansen, 2003; Wagner *et al.*, 2007). This modularity allows biological systems to vary only in a small subset of traits at a time. Each subpopulation with an identical trait combination occupies a point in the multidimensional trait space. If the trait combination matches the optimal trait combination for a given environment, the subpopulation is at maximal fitness. As the mismatch between the trait combination and environmental optimum increases, the fitness decreases. We assume that all traits are continuous, as well as the performance determined by them. This assumption is justified by recent findings in genomewide association studies suggesting continuous variation of the traits by variants of many genes (Boyle *et al.*, 2017) and is consistent with the classical population genetics framework with the assumption of a single local optimum and highly pleiotropic mutations (Fisher, 1930; Orr, 2005a; Martin and Lenormand, 2006; Gros *et al.*, 2009).

Our model makes two additional assumptions of biological importance. First, we assume that the fitness decay is isotropic; that is,

all traits deviating from the optimal trait combination contribute to the fitness decay in the same way. Because the trait space is not directly observable, normalization ensures that this assumption is true. Our second assumption is that the population is distributed in the trait space according to an isotropic multidimensional Gaussian. This assumption corresponds to two possible biological realities. First, the genetic variation occurs from a founder subpopulation selected under an initial environment. If we assume that all traits are polygenic and are affected uniformly by the type of genetic variation, such as aneuploidy, whereby the expression of hundreds to thousands of genes is altered simultaneously, the assumption is satisfied on account of the central limit theorem (Feller, 1945). Second, the assumption can be satisfied as a result of the environment's selective pressure limiting the deviation from some reference phenotype, due to the maximum entropy nature of the Gaussian distribution for fixed mean and SD, making it the distribution we would expect to appear by default (Lisman and van Zuylen, 1972). The subpopulation located at the center of the Gaussian distribution, approximated by the arithmetic mean of the entire population in the trait space, would correspond to the founder genome or the trait combination selected by the initial environment. Given that all subpopulations others than this one correspond to adaptations to specific niches, to be consistent with traditional terminology in ecology and evolution, we will refer to this central population as “generalist.”

Figure 1A shows a representation of fitness and population distributions in the trait space, where fitness decreases as the Euclidean distance between the environmental optimum and the subpopulation's position in trait space (d) increases (Figure 1B). We use a general class of exponential functions to represent the fitness, $G_\gamma(d) = \exp\left(-\left(d/s\right)^{2\gamma}\right)$, where s is the characteristic fitness decay distance in the trait space and γ determines how sensitive the subpopulation's fitness is to small deviations from the environmental optimum. γ is, therefore, a measure of the cell system's inherent robustness—the higher it is, the smaller will be the drop in fitness before a critical value s is reached. This exponential class of functions is closely related to the Weibull distribution, used in the study of failures under mechanical loads. In that context, the parameter γ could be increased by introducing redundancy, such as by using a bundle cable rather than a bar of identical section (Weibull, 1939). s , on the other hand, is merely a scaling factor for d , the distance in the trait space, and hence can be set to any value we desired in the normalization step if related parameters are normalized as well (Martin and Lenormand, 2006).

This formalization separates the organism's robustness, controlled by the parameter γ , from the population heterogeneity, controlled by the parameter σ_A (Figure 1A). In the context of evolutionary selection, fitness relative to an ancestor or founder population is a more meaningful measure than absolute fitness. To represent this comparison, we introduced the comparative fitness $H = \ln(G_{\text{int}} / G_{\text{ref}})$, where G_{int} is the fitness of a subpopulation of interest and G_{ref} the fitness of the generalist subpopulation as defined above.

Robustness is an essential precondition for the emergence of adaptive subpopulations

On the basis of the formulation of the model, we derive a complete expression for the mean fitness (μ) and SD (σ) of relative fitness H of all subpopulations in the original heterogeneous population (Supplemental Material, Eqs. 3.3 and 3.6). While these expressions are analytical, they involve nontrivial functions (Supplemental

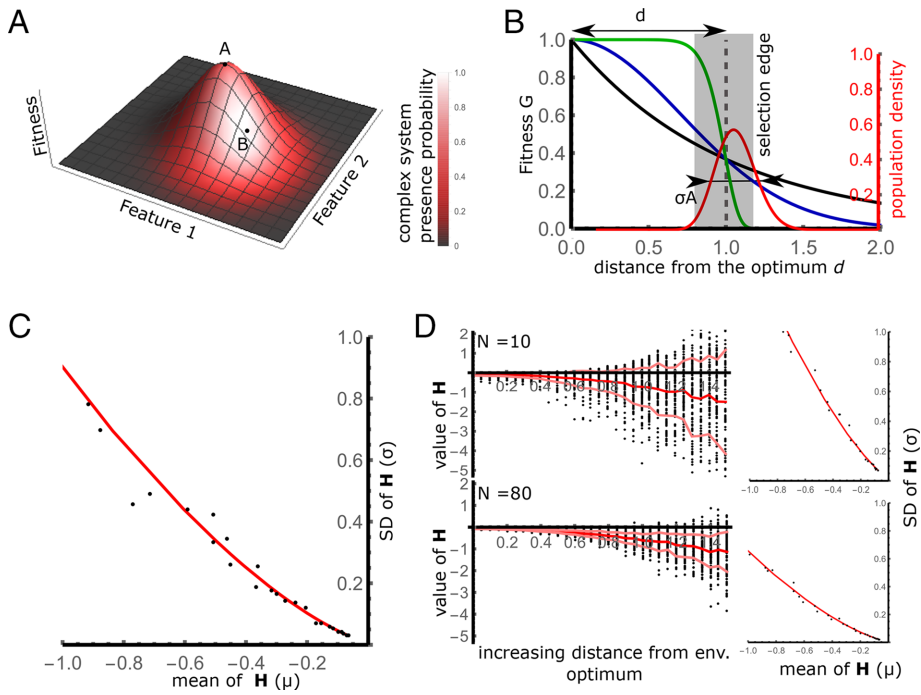


FIGURE 1: Model of adaptation of heterogeneous populations to stressful environments. (A) Schematic representation of an example trait space ($N = 2$). Point A represents the fitness optimum under a given environment. Color codes for the population density in the trait space. Point B is the position of the reference subpopulation. (B) Fitness (black, blue, green) and population distance to trait space optimum distribution (red). Gray denotes selection edge. $\gamma = 0.5$ (black), 1 (blue), 6 (green); $s = 1$. (C) Simulated (black) and theoretical (red) correlation between mean and SD of comparative fitness H computed with parameters $N = 40$, $\gamma = 2$. (D) Curse of complexity: simulations ($\gamma = 2$) showing that for larger N ($N = 80$) the number of selectable variants ($H > 0$) decreases. On the left, individual subpopulations (black), with mean (red) and SD (pink). On the right, H mean–SD correlation for the population (same colors as in C).

Material, Eqs. 3.3 and 3.6), and we therefore provide the following approximate expressions for μ and σ :

$$\mu = -\frac{\gamma(2\gamma + N - 2)}{N} \cdot \frac{(l/s)^2}{(l/\sigma_A)^2}$$

$$\sigma = \frac{2\gamma}{\sqrt{N}} \cdot \frac{(l/s)^{2\gamma}}{(l/\sigma_A)}$$

Here, σ_A , population heterogeneity, is the SD of the population distribution in the trait space (Figure 1A, white area radius). N is the dimensionality of the trait space, corresponding to the number of independent pathway groups allowing adaptation to a stress, and l is the distance between the environmental optimum and the generalist subpopulation in the trait space. The complete analytical expressions (see Supplemental Material, Eqs. 3.3 and 3.6) agree well with direct model simulations (Figure 1, C and D, and Supplemental Figures S1 and S2; see the Supplemental Material for more details), and approximate expressions, which allow much easier computation, are almost identical to the full expressions (Supplemental Figure S3) across a wide range of parameters.

The results show that the SD of the relative fitness (σ) increases monotonically as average relative fitness (μ) decreases for a larger departure from the environment optimum (Figure 1, C and D). In other words, under conditions close to the optimum, the population's fitness is represented by that of the generalist subpopulation

and exhibits minimal variation, but as the stress level increases (increasing distance from the optimum), the fitness of the population of systems scatters, so that subpopulations with higher fitness than the generalist emerge. This trend is consistent with the previous experimental observations (Chen *et al.*, 2015). We define evolvability as the presence of a subpopulation with a fitness significantly higher than that of the other subpopulations, allowing its genotype to spread through the entire population while under selection. This feature is better visible on the μ – σ parametric curve—the higher the curve slope, the larger SD compared with mean relative fitness and the higher is the chance of a niche subpopulation to exceed the generalist fitness. Analysis of the effect of various values of the robustness parameter γ on the μ – σ parametric curve revealed that evolvability exists if an organism is sufficiently robust ($\gamma > 1.5$) (Figure 2, A–C; Supplemental Figure S1). When $\gamma > 1.5$, cell fitness displays thresholdlike behavior in trait space, in which after an initial plateau the fitness falls sharply off at the selection edge ($s = l$; Figure 2A). Such systems are inherently robust because the fitness is stably maintained within a certain range of random perturbations (Ohta, 1992). Only then does the fitness difference become large enough across the population to overcome randomness in survival and reproduction, give rise to selection, and drive evolution (Figure 2, B and C; Supplemental Figure S1; Ohta, 2002). In contrast, when $\gamma < 1$, fitness falls

sharply off as soon as the condition deviates from the optimum (Figure 1A), and no adaptive subpopulation can be selected (Supplemental Figure S1).

The previous theoretical analysis suggested that robustness contributes to evolvability by allowing the accumulation of heterogeneity (Wagner, 2008). Our model is able to recapitulate that observation by adjusting the population heterogeneity parameter σ_A (Supplemental Figure S2, A–C), but also predicts that the effect of robustness on evolvability might be independent of population heterogeneity (σ_A) or population size (Supplemental Figure S2). Our model also suggests a solution to the well-known “cost of complexity” paradox in population genetics (Orr, 2000). This paradox suggests that an increase in complexity, represented in our model as the number of dimensions of the trait space (N), results in a smaller chance of a random step in the traits space leading to adaptation (Figures 1D and 2, E and F). Until now it was thought that biological systems’ modularity allows only a few dimensions to be affected by mutations, leading to a restriction on the number of trait space dimensions along which a mutation could move an organism (Martin and Lenormand, 2006; Wagner *et al.*, 2008). However, if the gain in complexity coincides with increased robustness, the higher γ can positively influence the emergence of selectable variance even with high trait-space complexity (Figure 2G). Our model also predicts that for a sufficiently robust organism ($\gamma > 1.5$), the slope of the σ – μ curve strongly depends on N , while a change in γ leads

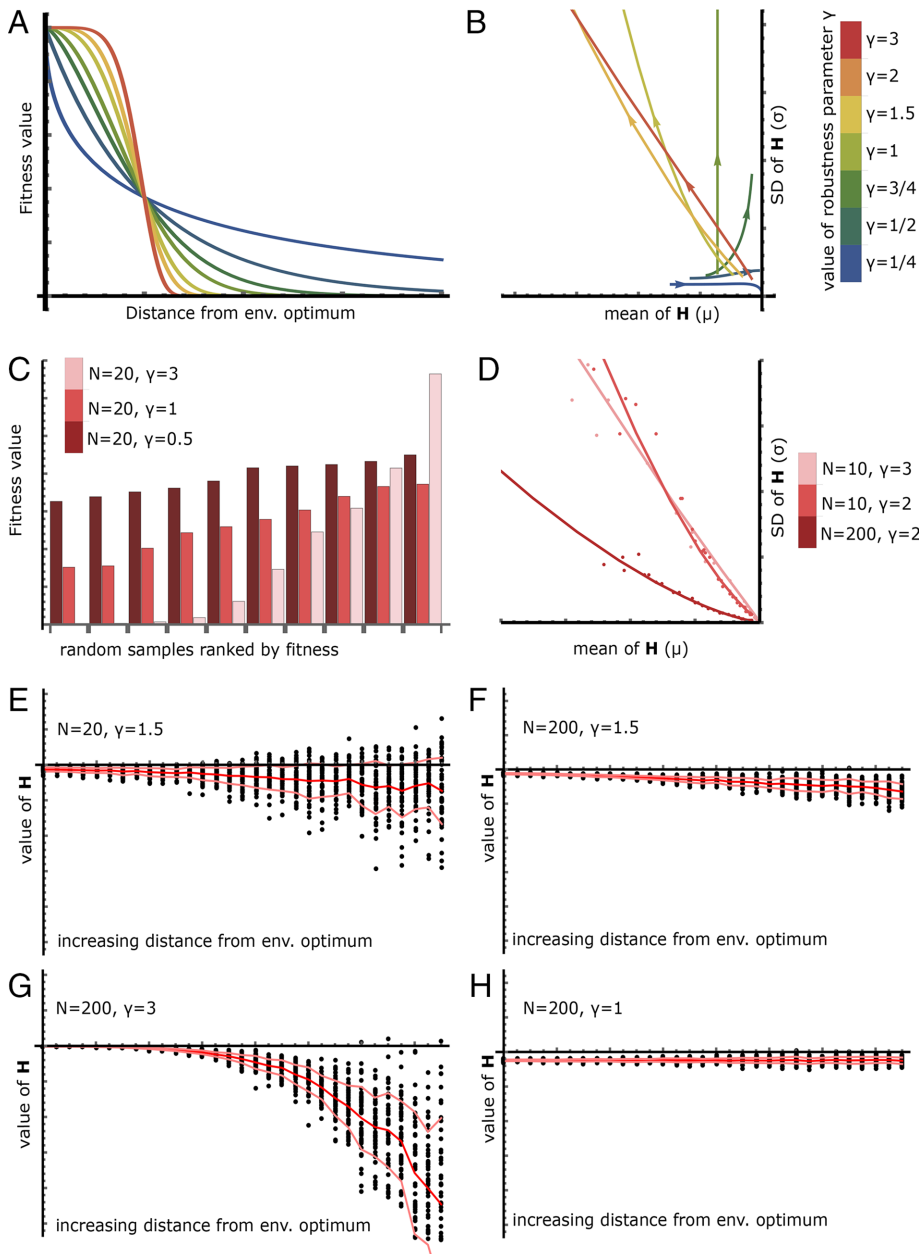


FIGURE 2: Robustness is an essential for the emergence of adaptive subpopulations. (A) Shape of the fitness function G for different values of the robustness parameter γ ; $s = 0.5$. (B) Different modes of the mean–SD relationship of the comparative fitness function H with varying γ ; $N = 40$. Color keys for E and F are shown on the right. (C) A selectable fitness difference occurs only for a high value of the robustness parameter γ ; $\gamma > 1.5$. Bars are raw fitness values for 10 randomly sampled subpopulations at the selection edge ($d = s$). (D) Slope of $\sigma(\mu)$ depends on N , while the curvature depends on γ . Simulated (points) and theoretical curve (continuous). (E) At low trait space complexity, intermediate robustness generates a sufficient difference in fitness in a new environment to lead to selection. (F) At high trait space complexity, intermediate robustness is unable to generate a sufficient difference in fitness in a new environment to lead to selection. (G) At high trait space complexity, high robustness is able to generate a large difference in fitness in a new environment to lead to selection of the fittest members of the population. (H) At high trait space complexity, low robustness generates an even lower difference in fitness between population members than intermediate robustness.

only to a curvature change (Figure 2D). This means that both the organism robustness γ and the complexity N can be estimated from the experimentally observed values of H , which to our knowledge is the first experimentally feasible method for performing such measurement (Orr, 2002).

A method for identifying the generalist subpopulation from experimental data

To make the model useful for the analysis of experimental data, there needs to be a way to accurately identify the generalist subpopulation. Because the trait space is not experimentally accessible, this method needs to rely on the distribution of fitness across subpopulations under different conditions. The central positioning within the heterogeneous population dictates that the core property of the generalist population is its relatively stable fitness in varied environments: regardless of the direction of shift of the environmental optimum, the generalist subpopulation always remains among the fittest ones, whereas the peripheral subpopulations would have much more contrasting fitness between different environments (Figure 3, A and B). There is, therefore, an inherent architecture of general adaptive potential within the heterogeneous population. To quantify how much of a generalist a subpopulation is, we developed the environment-specificity index (ESI), whose mathematical formulation is closely related to the Gini index (Gini, 1921), a highly robust measure of inequality and preference for a specific condition (Hurley and Rickard, 2009). For a subpopulation p with fitness $G_{p,i}$ in the environment i and a set m of total environments,

$$ESI(p) = \frac{\sum_{i \in [1,m]} \sum_{j \in [1,m]} |G_{p,i} - G_{p,j}|}{2n \sum_{i \in [1,m]} G_{p,i}}$$

Subpopulations whose fitness in a single environment or a few environments is significantly higher than in most others would exhibit high fitness inequality across the environments and hence high ESI, whereas those with a more even fitness distribution across environments would have low ESI (Figure 3C). Our model predicts that ESI is correlated to average fitness: the generalist subpopulation has the lowest ESI and the highest average fitness across various stressful environments, while subpopulations further away from it have higher ESI and lower average fitness (Figure 3D). These noncentral subpopulations fare well only in a small set of environments, and poorly in almost all others, behaving similarly to specialist species in an ecosystem.

Application of the model to aneuploid budding yeast and breast cancer cell lines

We tested our model on two types of cells—a cohort of 38 different aneuploid *Saccharomyces cerevisiae* yeast strains submitted to a panel of stressful environments (Pavelka *et al.*, 2010; Supplemental

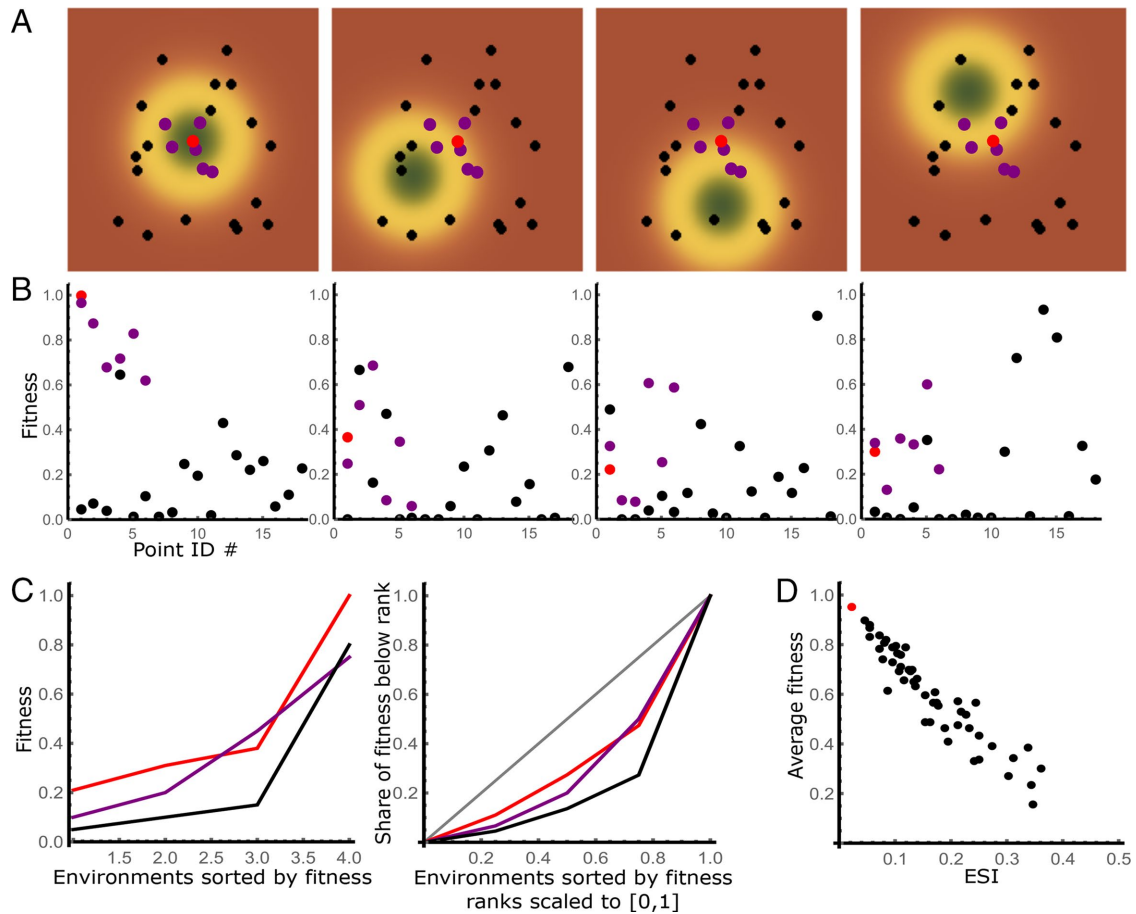


FIGURE 3: ESI and average fitness are correlated and allow identification of the reference subpopulation. (A) Variation in optimal fitness zone (green) in trait space around initial value leaves the reference subpopulation (red dot) in the acceptable zone (yellow) as other subpopulations (black and purple) switch between optimal (green) and unacceptable (orange) zones. (B) Fitness for the population above. The X axis has the subpopulation identification number for each dot. Fitness is somewhat stable for the reference subpopulation, whereas it varies more widely for other subpopulations. Subpopulations are more likely to reach deadly fitness (–0) if they are further from the reference (compare black dots with purple dots). (C) Sorted fitness values (left) and fitness Lorenz curve (right) for the reference subpopulation (red), subpopulations close to it in the trait space (purple), and a subpopulation at a significant distance from it (black). (D) The simulation shows a correlation between ESI and average fitness across environments for a population of 40 subpopulations in 20 environments.

Figure S4A) and a panel of 70 different breast cancer cell lines treated with a large number of drugs with diverse modes of action at their effective concentrations (Daemen *et al.*, 2013; Supplemental Figure S4B; see *Materials and Methods*). Aneuploid yeast cells have different chromosome copy number combinations, which result in large-scale perturbations to gene dosage stoichiometry, leading to distributed and highly pleiotropic phenotypic effects. Cancer cell lines, in addition to widespread aneuploidy, also possess numerous mutations.

For the heterogeneous cohort of aneuploid yeast strains exposed to stressful environments (Pavelka *et al.*, 2010; Chen *et al.*, 2015; Supplemental Table S1), ESI analysis based on the experimental data shows that the haploid population represents the generalist subpopulation (Figure 4A). This was expected, as the euploid strain was the ancestor population that had adapted well to the rich growth media and from which all aneuploid strains were derived and hence in this context can be considered as a founder. Using the haploid as the generalist subpopulation, the regression of our model fits well ($r^2 = 0.756$, $p < 5 \times 10^{-10}$) with the analytical prediction of the model with a complexity equivalent to the trait space

dimensionality (N) of 55 and inherent robustness equivalent to $\gamma = 2.2$ (Figure 4B).

We next applied the ESI analysis to the published data on 70 breast cancer cell lines treated with 90 drugs representing a variety of adverse environments (Heiser and Sadanandam, 2012; Daemen *et al.*, 2013) (Supplemental Table S2). Interestingly, the BT-483 cancer cell line, but not nontransformed human breast epithelial cells, was closest to being the generalist, with the lowest ESI and highest mean fitness across drug treatments (Figure 4C). This may be explained by the fact that nontransformed cells require specific growth signals in their environment to survive and proliferate, due to their role in a multicellular organism, whereas cancer cells were selected due to their clonal single-state fitness under suboptimal conditions and are often able to divide independent of growth-promoting factors. Using BT-483 as the generalist subpopulation, the data fit well ($r^2 = 0.61$, $p < 1 \times 10^{-10}$), with the model yielding a trait space dimensionality (N) of 90 and a robustness parameter $\gamma = 3$ (Figure 4D). Interestingly, even the looser fit for low μ and σ can be explained by our model—this pattern is expected when the population distribution in the trait space is wider among some traits than others

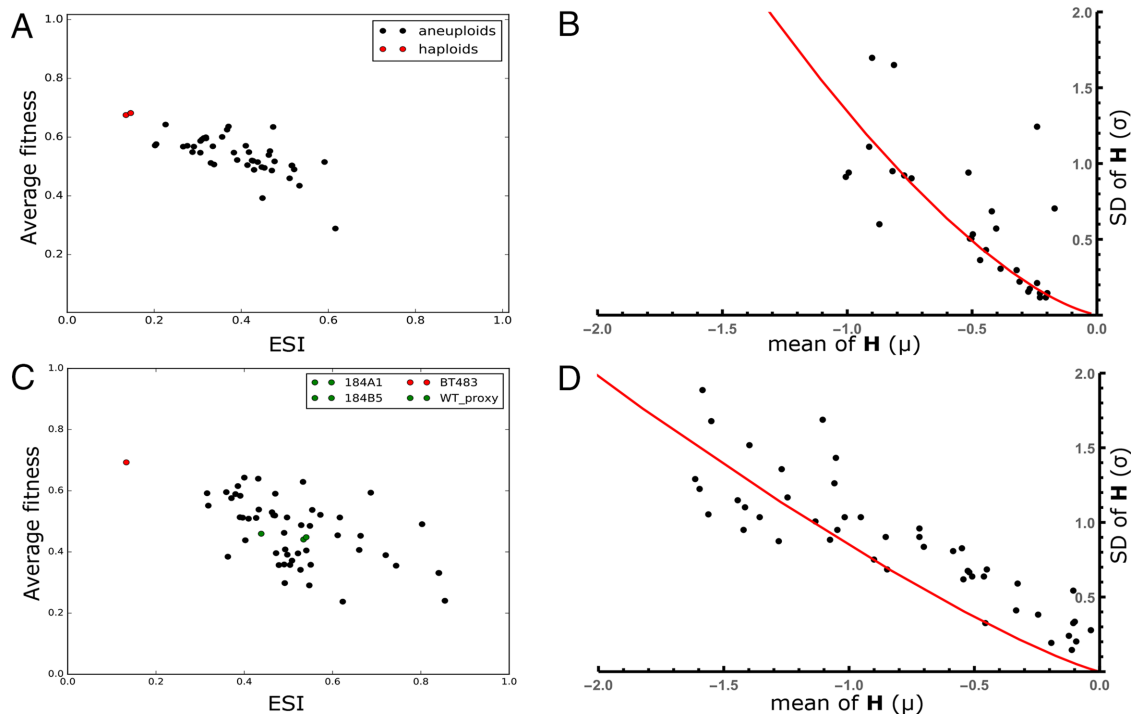


FIGURE 4: Application of the model to *S. cerevisiae* aneuploids and breast cancer cell lines. (A) The average fitness and ESI analysis of the data for aneuploid yeast responding to diverse stress conditions identify the haploid yeast strain (red dots, biological replicates) as the reference subpopulation. (B) Regressions of the growth data of aneuploid *S. cerevisiae* under diverse growth conditions suggest that for the trait space representing yeast fitness, $N \sim 55$, $\gamma \sim 2.2$ ($r^2 = 0.756$, 30 samples, $p < 10^{-10}$). (C) The average fitness and ESI analysis of the data of breast cancer cell lines responding to diverse drug treatments identified BT483 (red), rather than unaltered mammary duct epithelial cell lines (green), as the reference subpopulation. (D) Regressions of the data of breast cancer cell lines in response to diverse drugs predict $N \sim 90$, $\gamma \sim 3$ ($r^2 = 0.61$, 48 samples, $p = 5.81 \times 10^{-11}$), suggesting that cancer cell lines are more complex and more robust than yeast cells.

(Supplemental Figure S5; see *Materials and Methods*). This result shows that our model applies well to heterogeneous cell populations of different origins and that the trait space of human breast cancer cells is significantly more complex than that of yeast cells.

DISCUSSION

In this work, we introduced and formalized a mathematical model describing the adaptability of heterogeneous cell populations. The model was built to recapitulate several basic attributes of cellular systems—complexity, modularity, and robustness—without relying on details about specific molecular pathways. This model helps achieve a threefold conclusion. First, it predicts an essential role for robustness in the evolvability of cellular systems. When robustness increases with the complexity of organisms, evolvability on the cellular level increases accordingly, explaining why complex systems can be both robust and evolvable. Second, it implies the presence of an adaptive structure in a heterogeneous population and provides a means of finding a “generalist” subpopulation with the largest capacity to survive diverse stress conditions. Finally, by applying it to the drug response data for breast cancer cell lines, not only have we validated its prediction about the correlation between growth repression and phenotypic variation, but also we have identified a cell line that may be used as an experimental model for multidrug resistance in cancer.

Our theoretical approach shares a formalization close to Fisher’s geometrical model (FGM) in population genetics (Fisher, 1930; Orr, 2005a). FGM also links biological systems’ fitness to their position in

the phenotypic space relative to the optimum. After accounting for genetic drift (Kimura, 1989; Ohta, 1992), FGM allows analysis of the accumulation of both beneficial genome alterations (Gillespie, 1983, 1984) and deleterious ones (Charlesworth *et al.*, 1993). Historically, FGM was built to describe the accumulation of small-effect mutations from ancestor to descendants (Tenaillon, 2014) and thus reconcile Mendelian and biometric genetics, but more recently it has grown beyond that scope (Matuszewski *et al.*, 2014). In its classical formulation, it conflates the complexity of total phenotypic space and the average pleiotropy of mutations (Peck *et al.*, 1997; Martin and Lenormand, 2006) and focuses on a single environment (Orr, 2005b). The robustness degree γ within our model is closely related to the fitness norm of reaction shape (Simms, 2000) and is classically interpreted in FGM as factor characterizing the degree of interloci epistatic interaction (Martin and Lenormand, 2006; Gros *et al.*, 2009; Wagner and Zhang, 2011; Fraisse *et al.*, 2016). This interpretation is consistent with our view—if fitness decay due to a trait deviation from the optimum requires the simultaneous failure of several genes, the robustness would be higher and the falloff more abrupt (Weibull, 1939). By applying the values of parameters derived by our model for yeast into a previously derived formula for epistasis (Gros *et al.*, 2009), our model predicts an average epistasis of $\sim 2\%$ in yeast (see *Materials and Methods*), consistent with the observed 1.6% prevalence of synthetic-lethal interactions (Costanzo *et al.*, 2016).

Unlike the classical FGM models, our approach provides a direct solution to the “cost of complexity” paradox (Orr, 2002). An

extension of FGM suggested that this needs to be resolved by modularization of organism's traits (Martin and Lenormand, 2006; Wagner *et al.*, 2008). Our model suggests that the increased robustness of the organism could help restore the evolvability of systems affected by a high degree of complexity (large number of traits), consistent with the previous suggestion that selection may overcome the random drift in populations that are sufficiently large (Ohta, 1992) in a range of trait complexity (Gros *et al.*, 2009). Within the family of the FGMs, our model's ability to detect the generalist and to perform regression of the trait space complexity based on relatively small datasets is unique and could open up an avenue for the validation of FGM predictions, such as the link between robustness and trait space complexity in different organisms (Gros *et al.*, 2009).

In a more general context, our results on the role of robustness in evolvability are unique. Previous approaches that studied the role of robustness in evolvability approached it from the point of view of the diversity in the "state graph," where different graph nodes represented different phenotypes and were connected through mutation edges (Draghi *et al.*, 2010; Raman and Wagner, 2011; Wagner, 2012). In those models, robustness reduced the impact of phenotype on the fitness in a given environment, allowing a larger variety of nodes to be viable in several environments at the same time. In other words, the graph-based model suggested that robustness improves evolvability through increasing population diversity. Our model treats robustness and population diversity as features controlled by different, unrelated parameters and shows that robustness can enhance evolvability independent of population diversity.

The generalist subpopulation occupies a central role in our model. Whereas we expect that, in the case of a heterogeneous population derived by random variation from a single founding subpopulation, the generalist subpopulation will coincide with the founding subpopulation, this may not be the case in general. More generally, the generalist subpopulation is defined by its central position in the trait space. As the arithmetic mean of the population, it is maximally representative of the whole population, and unlike other subpopulations, it is not specialized toward a specific environment. From the evolutionary point of view, this distinction between the generalist subpopulation that is slowly evolving at long timescales and niche subpopulations that can rapidly take over the population in response to a particular stress could potentially explain the time-scale-dependent evolution speed paradox (Cavunis *et al.*, 2015). Accordingly, genomes of organisms that are observed to adapt on short timescales, such as during experimental evolution, are changing much faster than genomes of organisms on historical timescales, such as changes occurring during speciation. With respect to the breast cancer cell lines, the multidrug resistance of BT483 cells did not seem to be dependent on drug efflux pumps (Christgen *et al.*, 2007), which may be consistent with our model, which explains general resistance to stress (drugs) due to a generalist-like position in the trait space and not due to a specific feature. BT-483 may therefore be a useful model for understanding the origin and mechanism of multi-drug-resistant "cancer initiating cells" (Sharma *et al.*, 2010; Greaves, 2015; Laughney *et al.*, 2015; Ramirez *et al.*, 2016).

MATERIALS AND METHODS

Isotropy of trait space

A potential limitation of our model stems from the assumption that the distribution of population in trait space remains isotropic Gaussian following the normalization of the trait space to have equal effects on performance. Although multidimensional Gaussianity of

the distribution is reasonable, variances along all dimensions may not be equal. Analytical consideration of this problem is closely related to Wishart matrix trace calculation (Graczyk *et al.*, 2003) and to our knowledge is intractable, although tail behavior analysis (Jaschke *et al.*, 2004; Martin and Lenormand, 2008) suggests that the fitness distribution would be similar to the distribution described here, except with a lower apparent complexity of the trait space. This is consistent with our simulations (see Supplemental Figure S5). The excellent fitting between model predictions and regression of biological data suggests that our assumption above is unlikely to be far off. It is interesting to note that for anisotropic distribution of the population in trait space, we would expect the trait-space dimension to be underestimated, with dimensions with smaller variation being hidden by ones with larger. Also, due to a larger number of dimensions available for exploration at a small distance away from the optimum, the parametric μ - σ plot fit, such as those in Supplemental Figure S5, A and C, has σ higher than predicted by our model for low μ . This was indeed observed (Figure 2K).

ESI index

The ESI index is closely related to the Gini index, which was introduced in economics as a measure of income inequality in populations (Gini, 1921). More recently, its usage has been expanded to account for the degree of heterogeneity in a population and to measure how specific a variable is for a given environment (Hurley and Rickard, 2009). The Gini index is defined as the difference between the fraction of total wealth owned by every individual and the fraction he or she would own if wealth were distributed equally among all. Formalizing this, in a population of size n , where w_i is wealth owned by the individual i , the Gini index is expressed as follows:

$$\text{Gini} = \frac{\sum_{i=1}^n \sum_{j=1}^n |w_i - w_j|}{2n \sum_{i=1}^n w_i}$$

Because we are interested in the specificity of the fitness to a particular environment for a given system, we need to use the fraction of total fitness in a given environment as opposed to the case where the fitness is distributed equally across all environments. Formalizing this, if we take a member of a population p , with fitness in an environment i denoted as $G_{p,i}$,

$$\text{ESI}(p) = \frac{\sum_{i \in \text{environments}} \sum_{j \in \text{environments}} |G_{p,i} - G_{p,j}|}{2n \sum_{i \in \text{environments}} G_{p,i}}$$

Phenotypic profiling of aneuploid yeast strains

Phenotypic profiling was performed as described in (Pavelka *et al.*, 2010). Briefly, 4 μ l of the normalized ($\text{OD}_{600} = 0.1$) aneuploid and euploid cultures was spotted onto various Nunc OmniTray agar plates with a Biomek FX liquid handler. A list of chemicals and conditions used can be found in Supplemental Table S1. After 3–10 d of growth, the plates were scanned and the images were processed with a custom R script (Shah *et al.*, 2007) to obtain the mean spot intensity.

Data processing of aneuploid *Saccharomyces cerevisiae* response to stress conditions

The initial raw data for *S. cerevisiae* aneuploid growth was obtained from in-lab archives resulting from experiments performed for

Pavelka et al. (2010). The image analysis and the assembly pipeline were kept intact. Consistent with Pavelka et al. (2010), the density of the colony on the plate following 7 d of growth was used as a proxy for the fitness. The raw data are available in Supplemental Table S2 (*final_growth_values.csv*), with strain names and stress names consistent with the original publication. This data extraction was performed using custom Python 2.7 code located at https://github.com/chiffa/Screening_analysis, commit *ebe0eb8*. First the spot intensity values were retrieved by running *growth_assay_fitting.py*.

The resulting data, summarized in *lastvalues.csv*, were analyzed with custom Python 2.7 and Wolfram Mathematica code, located at https://github.com/chiffa/General_Adaptability_Model, commit *624f1f7*. Figure 2E and Supplemental Figure 4A were generated directly by running *Jin_data_re_analysis.py*, whereas Figure 2F was generated directly by running *Jin_Data_Regression.nb* on *Jin_out_haploid-gen.csv* data output by *Jin_data_re_analysis.py*.

p values for Figure 2F were calculated from the r^2 correlation coefficient and sample number. The formula $t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$ was used to find the t -statistic value and the p value corresponding to a two-tailed Student's test was used to determine the p value from the t value (Miles and Banyard, 2007). The same procedure was applied to retrieve the p value from the r^2 correlation coefficient in the breast cancer cell lines analysis.

Breast cancer cell lines collection response to the drug treatments

The initial data used for breast cancer cell lines analysis were obtained from Daemen et al. (2013). The additional file 9 – “*gb-2013-14-10-r110-s9.tsv*” was downloaded, renamed “*gb-breast_cancer.tsv*,” and analyzed with a custom Python 2.7 pipeline located at https://github.com/chiffa/Pharmacosensitivity_growth_assays, commit *b31eaa1*. The first run with the “*readers.py*” script extracted the dilution series for every drug–cell line combination pair. The second run with “*post_processing.py*” used the data structures created by the first run to generate Figure 2G and Supplemental Figure 4B. Once BT483 was identified as a generalist, normalization to it was performed and the mean and SD of normalized log-fitness were computed and stored in “*BC_analysis.csv*” file (Supplemental Table S3). This file was fed into the same *Jin_Data_Regression.nb* script as in the case of *Saccharomyces cerevisiae*, which generated Figure 2H. p values were generated in the same way as for *aneuploid S. cerevisiae* data.

Conceptually, the “*readers.py*” script performed the following steps:

Read the OD readings for blank wells on each plate, use it to calculate the instrument noise for a given plate, and discard plates that had too high an instrument noise. Pooled noise from the plates was used to estimate the instrument-based error margin of the measurements.

Read the ODs for wells without drugs. Based on them, the normalized maximum is calculated. To determine the fitness, we used the total decrease in OD compared with the no-drug conditions.

Due to the large size of the confidence interval for the cell line fitness in response to a drug at a given concentration, as well as a strong batch effect, ODs from wells that contained the same cell lines submitted to the same drug at the same concentration were pooled together. Plates where the difference on a cell line

between no drug and maximum drug concentration had a dynamic range of less than 10 were omitted from the pooling.

To be able to compare cell line response across several environments, all cell lines that were tested on fewer than 10 drugs were dropped.

To preserve the additional information beyond the IC50 (Fallahi-Sichani et al., 2013), we calculated the cell line–drug effectiveness score based on the cell line fitness at the discriminatory drug concentration. A discriminatory drug concentration was defined as a concentration of the drug for which at least 75% of cell lines tested were at relative fitness below 0.9 and at most 25% were at relative fitness below 0.1.

Every drug at effective concentration was considered a separate environment, and the OD at that concentration after the growth period, relative to the no-drug condition, was considered as the fitness for a given cancer cell line in that environment.

Simulation

The simulations were performed with a custom Mathematica script, available at https://github.com/chiffa/General_Adaptability_Model, commit *d66c8e0*, *Supporting_Simulations.nb* file. The simulations consisted of a sampling of positions in the heterogeneous population in the N -dimensional trait space, based on a multinormal distribution. For the simulations of anisotropic space, the covariance matrix was constructed by sampling from a lognormal distribution with a mean of 0 and a dispersion of 1. After the positions were computed and saved, a set of random new optimum values were generated, with a fixed distance from the center of the population. The distance between the positions of subpopulations and optimums were computed and fed into the fitness function G . Direct output results were used in Figure 1G. For all the other functions, the transformation through the relative log fitness with the function H was performed, and then mean relative log fitness and SD were calculated and plotted.

Average epistasis prediction

Because our model bears a formal resemblance to the model developed by Gros et al. (2009), we are using the formula for epistasis provided in that paper: $e \approx (2 - 2^{s/2})^{\sigma s}$. In our model, we use the parameterization $\gamma = \sigma/2$, leading to the formula $e \approx (2 - 2^\gamma)^{\sigma \gamma}$. Assuming that σ , the average mutation effect, is of the order of magnitude of $\sigma_A/3$, consistent with the observation of an average aneuploidy deficit of ~0.75% in aneuploid yeast (Pavelka et al., 2010) and ~0.93% for nonessential genes deletion (Costanzo et al., 2016), for $\Upsilon = 2.2$ and $\sigma = 1/3$, the estimated e value is -0.01975 , corresponding to ~2% negative epistasis.

ACKNOWLEDGMENTS

We thank O. Tenaillon, S. Matuszewski, G. Fischer, J.E. Gern, R. Yu, D. Dahlen, A. Baryshnikova, and D. Georgess for insightful discussions and O. Tenaillon, H.-J. Tsai, S. Matuszewski, and G. Fischer for comments on the manuscript. This work was supported by Grant R35 GM118172 from the National Institutes of Health to R.L. Any custom software used in this publication and data sets used to obtain the results in this manuscript are listed in *Materials and Methods* and are in open access in the indicated GitHub repositories.

REFERENCES

Alizadeh AA, Aranda V, Bardelli A, Blanpain C, Bock C, Borowski C, Caldas C, Califano A, Doherty M, Elsner M, et al. (2015). Toward understanding and exploiting tumor heterogeneity. *Nat Med* 21, 846–853.

- Altschuler SJ, Wu LF (2010). Cellular heterogeneity: do differences make a difference? *Cell* 141, 559–563.
- Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Califano A (2016). Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat Genet* 48, 838–847.
- Barabási A-L, Oltvai ZN (2004). Network biology: understanding the cell's functional organization. *Nat Rev Genet* 5, 101–113.
- Ben-David U, Ha G, Tseng Y-Y, Greenwald NF, Oh C, Shih J, McFarland JM, Wong B, Boehm JS, Beroukhi R, et al. (2017). Patient-derived xenografts undergo murine-specific tumor evolution. *Nat Genet* 49, 1567–1575.
- Boyle EA, Li Yi, Pritchard JK (2017). An expanded view of complex traits: from polygenic to omnigenic. *Cell* 169, 1177–1186.
- Bozic I, Allen B, Nowak MA (2012). Dynamics of targeted cancer therapy. *Trends Mol Med* 18, 311–316.
- Bozic I, Reiter JG, Allen B, Antal T, Chatterjee K, Shah P, Moon YS, Yaqubie A, Kelly N, Le DT, et al. (2013). Evolutionary dynamics of cancer in response to targeted combination therapy. *Elife* 2, e00747.
- Carlson JM, Doyle J (2002). Complexity and robustness. *Proc Natl Acad Sci USA* 99, 2538–2545.
- Carvunis AR, Wang T, Skola D, Yu A, Chen J, Kreisberg JF, Ideker T (2015). Evidence for a common evolutionary rate in metazoan transcriptional networks. *Elife* 4, e11615.
- Charlesworth B, Morgan BT, Charlesworth D (1993). The effect of deleterious mutations on neutral molecular variation. *Genetics* 134, 1289–1303.
- Chen G, Mulla WA, Kucharavy A, Tsai H-J, Rubinstein B, Conkright J, McCroskey S, Bradford WD, Weems L, Haug JS, et al. (2015). Targeting the adaptability of heterogeneous aneuploids. *Cell* 160, 771–784.
- Chen G, Rubinstein B, Li R (2012). Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *BioEssays* 34, 893–900.
- Christgen M, Ballmaier M, Bruchhardt H, Wasielewski R, Kreipe H, Lehmann U (2007). Identification of a distinct side population of cancer cells in the Cal-51 human breast carcinoma cell line. *Mol Cell Biochem* 306, 201–212.
- Chu D (2008). Criteria for conceptual and operational notions of complexity. *Artif Life* 14, 313–323.
- Costanzo M, VanderSluis B, Koch EN, Baryshnikova A, Pons C, Tan G, Wang W, Usaj M, Hanchard J, Lee SD, et al. (2016). A global genetic interaction network maps a wiring diagram of cellular function. *Science* 353, aaf1420.
- Daemen A, Griffith OL, Heiser LM, Wang NJ, Enache OM, Sanborn Z, Pepin F, Durinck S, Korkola JE, Griffith M, et al. (2013). Modeling precision treatment of breast cancer. *Genome Biol* 14, R110.
- de Visser JAGM, Hermisson J, Wagner GP, Ance Meyers L, Bagheri-Chaichian H, Blanchard JL, Chao L, Cheverud JM, Elena SF, Fontana W, et al. (2003). Perspective: evolution and detection of genetic robustness. *Evolution* 57, 1959–1972.
- Deris JB, Kim M, Zhang Z, Okano H, Hermsen R, Groisman A, Hwa T (2013). The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science* 342, 1237435.
- Draghi JA, Parsons TL, Wagner GP, Plotkin JB (2010). Mutational robustness can facilitate adaptation. *Nature* 463, 353–355.
- Eldar A, Elowitz MB (2010). Functional roles for noise in genetic circuits. *Nature* 467, 167–173.
- Fallahi-Sichani M, Honarnejad S, Heiser LM, Gray JW, Sorger PK (2013). Metrics other than potency reveal systematic variation in responses to cancer drugs. *Nat Chem Biol* 9, 708–714.
- Feller W (1945). The fundamental limit theorems in probability. *Bull Am Math Soc* 51, 800–833.
- Fisher RA (1930). *The Genetical Theory of Natural Selection*, London: Oxford University Press.
- Foo J, Michor F (2014). Evolution of acquired resistance to anti-cancer therapy. *J Theor Biol* 355, 10–20.
- Fraisse C, Gunnarsson PA, Roze D, Bierre N, Welch JJ (2016). The genetics of speciation: insights from Fisher's geometric model. *Evolution* 70, 1450–1464.
- Gerhart J, Kirschner M (2007). The theory of facilitated variation. *Proc Natl Acad Sci USA* 104, 8582–8589.
- Giam M, Rancati G (2015). Aneuploidy and chromosomal instability in cancer: a jackpot to chaos. *Cell Div* 10, 3.
- Gillespie JH (1983). A simple stochastic substitution model. *Theor Popul Biol* 23, 202–215.
- Gillespie JH (1984). Molecular evolution over the mutational landscape. *Evolution* 38, 1116–1129.
- Gini C (1921). Measurement of inequality of incomes. *Econ J* 31, 124–126.
- Gordon DJ, Resio B, Pellman D (2012). Causes and consequences of aneuploidy in cancer. *Nat Rev Genet* 13, 189–203.
- Graczyk P, Letac G, Massam H (2003). The complex Wishart distribution and the symmetric group. *Ann Stat* 31, 287–309.
- Greaves M (2015). Evolutionary determinants of cancer. *Cancer Discov* 5, 806–820.
- Gros PA, Nagard HL, Tenaillon O (2009). The evolution of epistasis and its links with genetic robustness, complexity and drift in a phenotypic model of adaptation. *Genetics* 182, 277–293.
- Hansen TF (2003). Is modularity necessary for evolvability? Remarks on the relationship between pleiotropy and evolvability. *BioSystems* 69, 83–94.
- Heiser L, Sadanandam A (2012). Subtype and pathway specific responses to anticancer compounds in breast cancer. *Proc Natl Acad Sci USA* 109, 2724–2729.
- Hurler N, Rickard S (2009). Comparing measures of sparsity. *IEEE Trans Inf Theory* 55, 4723–4741.
- Ideker T, Galitski T, Hood L (2001). A new approach to decoding life: systems biology. *Annu Rev Genomics Hum* 2, 343–72.
- Jaschke S, Klüppelberg C, Lindner A (2004). Asymptotic behavior of tails and quantiles of quadratic forms of Gaussian vectors. *J Multivar Anal* 88, 252–273.
- Kauffman SA (1993). *The Origins of Order: Self-Organization and Selection in Evolution*, New York: Oxford University Press.
- Kimura M (1989). The neutral theory of molecular evolution and the world view of the neutralists. *Genome* 31, 24–31.
- Laughney AM, Elizalde S, Genovese G, Bakhoun SF (2015). Dynamics of tumor heterogeneity derived from clonal karyotypic evolution. *Cell Rep* 12, 809–820.
- Lisman JHC, van Zuylen MCA (1972). Note on the generation of most probable frequency distributions. *Stat Neerl* 26, 19–23.
- Lujan SA, Clausen AR, Clark AB, MacAlpine HK, MacAlpine DM, Malc EP, Mieczkowski PA, Burkholder AB, Fargo DC, Gordenin DA, et al. (2014). Heterogeneous polymerase fidelity and mismatch repair bias genome variation and composition. *Genome Res* 24, 1751–1764.
- Martin G, Lenormand T (2006). A general multivariate extension of Fisher's geometrical model and the distribution of mutation fitness effects across species. *Evolution* 60, 893–907.
- Martin G, Lenormand T (2008). The distribution of beneficial and fixed mutation fitness effects close to an optimum. *Genetics* 179, 907–916.
- Matuszewski S, Hermisson J, Kopp M (2014). Fisher's geometric model with a moving optimum. *Evolution* 68, 2571–2588.
- McGranahan N, Burrell RA, Endesfelder D, Novelli MR, Swanton C (2012). Cancer chromosomal instability: therapeutic and diagnostic challenges. *EMBO Rep* 13, 528–538.
- McGranahan N, Swanton C, Abkevich V, Timms KM, Hennessy BT, Potter J, Carey MS, Meyer LA, Smith-McCune K, Broadus R, et al. (2015). Biological and therapeutic impact of intratumor heterogeneity in cancer evolution. *Cancer Cell* 27, 15–26.
- Merino VF, Nguyen N, Jin K, Sadik H, Cho S, Korangath P, Han L, Foster YMN, Zhou XC, Zhang Z, et al. (2016). Combined treatment with epigenetic, differentiating, and chemotherapeutic agents cooperatively targets tumor-initiating cells in triple-negative breast cancer. *Cancer Res* 76, 2013–2024.
- Miles J, Banyard P (2007). *Understanding and Using Statistics in Psychology: A Practical Introduction*, London: Sage Publications.
- Mitchell M (2011). *Complexity, a Guided Tour*, New York: Oxford University Press.
- Ohta T (1992). The nearly neutral theory of molecular evolution. *Annu Rev Ecol Syst* 23, 263–286.
- Ohta T (2002). Near-neutrality in evolution of genes and gene regulation. *Proc Natl Acad Sci USA* 99, 16134–16137.
- Orr HA (2000). Adaptation and the cost of complexity. *Evolution* 54, 13–20.
- Orr HA (2002). The population genetics of adaptation: the adaptation of DNA sequences. *Evolution* 56, 1317–1330.
- Orr HA (2005a). The genetic theory of adaptation: a brief history. *Nat Rev Genet* 6, 119–127.
- Orr HA (2005b). Theories of adaptation: what they do and don't say. *Genetica* 123, 3–13.
- Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, Sanderson BW, Hattem GL, Li R (2010). Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* 468, 321–325.
- Peck JR, Barreau G, Heath SC (1997). Imperfect genes, Fisherian mutation and the evolution of sex. *Genetics* 145, 1171–1199.
- Pennings PS (2012). Standing genetic variation and the evolution of drug resistance in HIV. *PLoS Comput Biol* 8, e1002527.

- Raman K, Wagner A (2011). Evolvability and robustness in a complex signaling circuit. *Mol Biosyst* 7, 1081–1092.
- Ramírez M, Rajaram S, Steininger RJ, Osipchuk D, Roth MA, Morinishi LS, Evans L, Ji W, Hsu C-H, Thurley K, et al. (2016). Diverse drug-resistance mechanisms can emerge from drug-tolerant cancer persister cells. *Nat Commun* 7, 10690.
- Shah NA, Laws RJ, Wardman B, Zhao LP, Hartman JL (2007). Accurate, precise modeling of cell proliferation kinetics from time-lapse imaging and automated image analysis of agar yeast culture arrays. *BMC Syst Biol* 1, 3.
- Sharma SV, Lee DY, Li B, Quinlan MP, Takahashi F, Maheswaran S, McDermott U, Azizian N, Zou L, Fischbach MA, et al. (2010). A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell* 141, 69–80.
- Simms EL (2000). Defining tolerance as a norm of reaction. *Evol Ecol* 14, 563–570.
- Stewart-Ornstein J, Weissman JS, El-Samad H (2012). Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol Cell* 45, 483–493.
- Taubes G (2008). The bacteria fight back. *Science* 321, 356–361.
- Tenaillon O (2014). The utility of Fisher's geometric model in evolutionary genetics. *Annu Rev Ecol Evol Syst* 45, 179–201.
- Wagner A (2008). Robustness and evolvability: a paradox resolved. *Proc R Soc B* 275, 91–100.
- Wagner A (2012). The role of robustness in phenotypic adaptation and innovation. *Proc Biol Sci* 279, 1249–1258.
- Wagner GP, Kenney-Hunt JP, Pavlicev M, Peck JR, Waxman D, Cheverud JM (2008). Pleiotropic scaling of gene effects and the “cost of complexity.” *Nature* 452, 470–472.
- Wagner GP, Pavlicev M, Cheverud JM (2007). The road to modularity. *Nat Rev Genet* 8, 921–931.
- Wagner GP, Zhang J (2011). The pleiotropic structure of the genotype—phenotype map: the evolvability of complex organisms. *Nat Rev Genet* 12, 204–213.
- Weibull W (1939). A statistical theory of the strength of materials. *Ingenjörsvetenskapsakademiens* 151, 1–45.
- Weinstein JN, Collisson EA, Mills GB, Shaw KRM, Ozenberger BA, Ellrott K, Shmulevich I, Sander C, Stuart JM (2013). The Cancer Genome Atlas pan-cancer analysis project. *Nat Genet* 45, 1113–1120.
- Zhu J, Pavelka N, Bradford WD, Rancati G, Li R (2012). Karyotypic determinants of chromosome instability in aneuploid budding yeast. *PLoS Genet* 8, e1002719.