Extended Experimental Procedure: Model

Initial assumptions

The model is based on a set of simple assumptions and their immediate logical derivations are as follows (all the elements in the model are shown in a schematic manner in Figure S1):

1) Fitness, as measured by cell growth in defined time in the experimental study, is governed by N ($N > 1$) independent pathways (e.g. those governing different processes such as energy production, DNA replication, mitosis, etc.).

2) Given the fact that aneuploidy affects the expression of hundreds of genes across the genome thus likely alters many components of each individual pathway, the activity of each pathway is affected by aneuploidy, encompassing diverse karyotypes, as a random process.

3) Because any karyotype change could encompass gene expression changes that either increase or decrease a pathway activity, the ensemble effect of aneuploidy on a pathway is a random variable. Based on central limit theorem applied to these random variables, the distribution of activity of pathway i across karyotypes assumes a 1D normal distribution having mean m_i and width (standard deviation) σ_i.

4) Assuming that i-th pathway activity of the euploid genome has been selected for optimal fitness under the present stress-free condition, the mean of the 1D normal distribution (m_i) corresponds to the mean activity of the euploid distribution.

5) The total pathway activity that defines growth fitness is a product of the activity distributions of all N pathways, represented by the multivariate normal distribution with the maximum at the point $\mathbf{m} = \{m_1, m_2, \ldots, m_N\}$ in the N-dimensional space, corresponding to the maximal fitness of the euploid in the absence of stress (Figure S1A). Fitness of aneuploid genomes, by contrast, due to sub-optimal activity of the affected pathways, corresponds to points away from $\mathbf{m}$, in line with observations that under optimal growth conditions, aneuploids are in general less fit than euploid (22,33,44) (Figure S1A)

6) The effect of a random stress condition, defined by type k and magnitude l, is represented by a random shift of maximum from point $\mathbf{m}$ to $\mathbf{m}_{k,l}$, the distance between which reflects the modulation necessary in the affected pathway activities in order to regain maximal fitness (Figure S1B).

7) Normalized cell growth, our measure of fitness, is defined as $G = 1 - (\Delta d)^2$, where Δd denotes the Euclidean distance in the N-dimensional space from the activity position of a given

aneuploidy karyotype to point $\mathbf{m}$ (without stress) or $\mathbf{m}_{k,l}$ (under stress). Thus, for euploid without stress, $\Delta d = 0$ and $G_{max} = 1$. Aneuploidy and stress independently impair fitness as $\Delta d > 0$, and thus $G < 1$.

Mathematical Background

The critical assumption of our approach is to consider stress and aneuploidy as random statistical processes forcing the cell population to drift from their optimal fitness positions ($\mathbf{m}$ under the stress-free condition; $\mathbf{m}_{k,l}$ under stress conditions), either by stress or by aneuploidy or both.

To model the effect of random stress, values $I_{i,S}$ (positive or negative) which represents the activitiy of the i-th pathway needs to be modulated to reach the optimal fitness are introduced. The vector $\mathbf{I}_S = \{I_{1,S}, I_{2,S}, \ldots, I_{N,S}\}$ with length l and direction k determines the shift of the optimal fitness position from $\mathbf{m}$ to $\mathbf{m}_{k,l}$, so that $\mathbf{I}_S = \mathbf{m}_{k,l} - \mathbf{m}$. We define stress adaptation margin as a norm of this vector, $l = \Delta d_S = \|\mathbf{I}_S\|$, so that

$$(\Delta d_S)^2 = \sum_{i=1}^{N} I_{i,S}^2. \tag{1}$$

Reuse the above reasoning for aneuploidy effect, and replace the shift required for adaptation to stress by the shift caused by gene dosage alterations, we can similarly define the fitness deficit caused by aneuploidy. Then the joint effect of stress and aneuploidy can be expressed as:

$$(\Delta d_{S,A})^2 = \sum_{i} I_{i,S,A}^2, \quad I_{i,S,A} = I_{i,A} + I_{i,S}. \tag{2}$$

where $I_{i,A}$ are distributed normally across the pathways/dimensions.

Now we can introduce a relation between the growth fitness G and the joint effect of stress and aneuploidy $\Delta d_{S,A}$. We expect and verify that many functions $G(\Delta d)$ with an appropriate behavior (a smooth and monotonically decreasing function with $G = 1$ at $\Delta d = 0$) will yield the observed trend in our experimental data. Here, $G = 1 - (\Delta d)^2$ is used, with the assumption that for $\Delta d > 1$ the cells do not survive and are removed from the calculation. Thus, for aneuploids under the stress we find the growth fitness

$$G_{S,A} = 1 - (\Delta d_{S,A})^2, \tag{3}$$

and calculate r as log_2 of the ratio of growth fitness between an aneuploid and the euploid (both under the same stress):

$$r = \log_2 \frac{1 - (\Delta d_{S,A})^2}{1 - (\Delta d_S)^2}, \tag{4}$$

where $\Delta d_{S,A}, \Delta d_S < 1$.

Simulations

To verify the predictions of the model and compare them to the experimental observations we performed a set of simulations.

The aneuploids were generated as points in the N-dimensional space by pulling a random vector from multivariate normal distribution with zero mean and the covariance matrix equal to the $N \times N$ unit matrix I_N times standard deviation a and divided by the dimension number $(a/N)I_N$. The division by N is required to compensate for the fact that the mean of norm of random vectors pulled from the abovementioned distribution is proportional to N. The average size of these random vectors is set to the value a corresponding to the mean aneuploidy effect under the stress-free condition. The stresses were generated as vectors in the N-dimensional space with the magnitude l within the range $\{0, l_{max}\}$ and randomly distributed direction (type) k.

The simulated data sets for each value of l were used to compute the absolute value of mean $|\mu|$ and the standard deviation σ of r, and then the linear regression was performed on these values. The results shown in Figure 1G were generated for the parameters $l_{max} = 0.9$, $a = 0.3$, and $N = 8, 16, 24, 48, 96$, and compared to the experimental data on the yeast strains.

In the analysis of the cancer cell lines data we could not use the normalization procedure as the data sets do not contain the normal (euploid) cell lines. In this case instead of $\log_2$ ratio r we used the logarithm of the absolute growth rate $G_{S,A}$ defined in (3) and computed $|\mu|$ and σ. The corresponding linear regression results are shown in Figure 2D for the same parameter values as in Figure 1G.

Additional simulations were performed to demonstrate the evolution trap method validity. To this end we selected $N = 24$, and generated 1000 aneuploids with $a = 0.3$. For a single randomly directed stress $\mathbf{X}$ with the magnitude $|\mathbf{X}| = 0.9$, we selected top 5% of the survived anuploids with highest relative growth rates (Figure 3A,B). Then we put these selected strains under no stress conditions and found that all of them predictably grew worse than the euploid (Figure 3B). It appeared that under action of the opposite stress $\mathbf{Y} = -\mathbf{X}$, all these strains died, while for a stress of the same magnitude but with a random direction we obtained the whole spectrum of the growth rates compared to the euploid strain (Figure 3B).