Adaptive Fitness Landscapes For Aneuploid Karyotypes

Abstract

Aneuploidy is prevalent in most solid tumors and can significantly alter cellular phenotype and fitness. Despite its critical role in tumor evolution, a detailed quantitative understanding of the evolutionary dynamics of aneuploidy remains elusive. To bridge this gap, we developed a novel method to infer the fitness landscape of chromosome-level karyotypes. This method utilizes longitudinal single-cell karyotype data from evolving cell populations to estimate the fitness of thousands of karyotypes proximal to the observed data in karyotype space. We validated the predictive capability of our approach using both synthetic data from an agent-based model and empirical data from in vitro and in vivo passaged cell lines. Our analysis of the fitted landscapes suggests several key insights: (1) Whole genome doubling promotes aneuploidy by increasing the spectrum of potentially beneficial copy number alterations; (2) Cisplatin treatment alters the fitness impact of these alterations; (3) The fitness consequences of a copy number alteration (CNA) vary depending on the parental karyotype; (4) The rate of chromosome missegregation can dictate the predominant karyotypes in an evolving population.

Introduction

Losses and gains of entire chromosomes or large sections thereof, known as aneuploidy, are a defining feature of solid tumors[1, 2]. Affecting more of the cancer genome than any other genetic alteration [3], aneuploidy is present in an estimated 90% of solid tumors and modifies the copy number of many genes, thereby altering cellular phenotype through correlated changes in RNA expression and protein production[4, 5]. Aneuploidy provides a substrate for tumor evolution[1, 6, 7], often enriching chromosomes with oncogenes while deleting those with tumor suppressor genes[8].

The factors which explain aneuploidy patterns in cancers are not limited to the density of driver or suppressor genes on a particular chromosome. Aneuploidy is usually detrimental to cell fitness [9], in the first instance due to proteins such as P53 which cause apoptosis or cell cycle arrest in response to chromosome mis-segregations [10]. According to the *gene doseage* hypothesis, aneuploidy can also reduce fitness by upsetting the balance of protein levels within cells: leading to negative effects such as impaired formation of stoichiometry dependent protein complexes, or protein aggregates that overwhelm protein quality-control mechanisms [9, 10]. Environmental context plays a role in sculpting karyotype [3], since the specific pressures of an environment will determine whether the fitness advantages of a particular CNA outweigh the costs. Evidence for the role of environment in determining karyotype includes the selective advantage of particular karyotypes under stressful conditions in yeast [4] and distinct patterns of aneuploidies between cancer types [11, 12, 13]. Genomic context also plays a role in sculpting karyotype because a given CNA may only be favorable if other mutations or CNAs are already present within the cell. Evidence for the importance of genomic context includes observations that CNAs which are not independently significant predict survival when co-occurring [14], and defined temporal ordering of CNAs observed in a patient derived xenograft model [15, 16].

Aneuploidy remains difficult to study, for reasons which include the difficulty of experimentally inducing aneuploidy and the difficulty of distinguishing the effects of aneuploidy from those of chromosomal instability, the process which causes aneuploidy [3]. *In silico* models will be an important tool to further our knowledge of aneuploidy. Gusev and colleagues developed the first model describing whole-chromosome mis-segregations [17, 18]. This model laid the mathematical foundations for describing segregation errors and explained patterns of aneuploidy in experimental data as a consequence of variable chromosome mis-segregation rate. A limitation of this model was that a fitness landscape defining the effect of aneuploidy on cell fitness was not considered, beyond a constraint that cells losing all copies of any chromosome were not considered viable. This limitation was later addressed by others who assumed that the fitness effect of changing the copy number of a particular chromosome was dependent on the number of oncogenes or tumor suppressor genes expressed on that chromosome [19, 20]. In these models cell fitness could be increased by gaining additional copies of chromosomes with many oncogenes, or losing copies of chromosomes with many tumor suppressor genes. These models predicted an optimal mis-segregation rate (in the sense of minimising total cell death) that matched experimental observations, and resulted in a near-triploid karyotype that is frequently observed in tumour cells. All these previous models of aneuploidy share the limitation that they ignore the context dependency of the mapping between karyotype and fitness. This is perhaps unsurprising, since the vast number of possible karyotypes is challenging enough to map even without the additional variability introduced by context. However, the burgeoning quantity of single cell copy number data [21, 22] now permits tracking of subclonal evolution at unprecedented resolution, giving the potential to refine our understanding of the relationships between karyotype and cellular fitness.

Here, we introduce ALFA-K, a method to directly estimate fitness landscapes from single-cell copy number data. After training and validating ALFA-K with synthetic data from an agent-based model, we further confirmed its efficacy with empirical data from P53-deficient cell lines showing extensive subclonal evolution. Our analysis suggests several key insights, including the potential roles of

whole genome doubling, cisplatin treatment, and the rate of chromosome missegregation in shaping the evolutionary trajectory of tumor cells.

Methods

Fitness landscape inference

Our landscape inference method consists of three steps, followed by a final validation procedure as outlined below. The input for the inference method is a matrix Y, with elements y_{it} which represent the number of times the ith karyotype appears in the longitudinal sample at time t. It will also be useful to also define here the matrix U as the normalized version of Y where each column is divided by its sum.

Frequent karyotypes

"Frequent karyotypes" are operationally defined throughout the manuscript as those observed in at least N cells across all sampled time points, where N is a user supplied hyperparameter. We consider that Y is generated via sampling from a latent population growing exponentially such that:

$$v_{it} = v_0 e^{f_i t}, \tag{1}$$

where v_{it} is the number of cells in the latent population with karyotype i at time t, and f_i is their fitness. These are normalized to form our prediction matrix $\hat{U}$, with elements:

$$\hat{u}_{it} = v_{it} / \sum_i v_i. \tag{2}$$

If $\hat{U}$ is sampled according to a binomial process to generate Y, then the probability of generating each y_{it} is:

$$l_{it} = P(y_{it}; \hat{u}_{it}, \sum_i y_{it}). \tag{3}$$

Defining S the set of indices belonging to frequent karyotypes, we find values of $f_{i \in S}$ and v_0 which maximise $\sum_{i \in S} \sum_t \log l_{it}$, i.e. the log likelihood, using a multi-start procedure and standard gradient descent methods.

Nearest neighbours

This step of the inference pipeline estimates the fitness of all the Von Neumann neighbours of the frequent clones estimated in the prior step. The growth of these neigbhbour karyotypes is governed by:

$$\frac{dv_i}{dt} = f_i v_i + \sum_j P(\alpha_i | \alpha_j) f_j v_j, \tag{4}$$

where the second term represents influx due to missegregations from the frequent karyotypes. $P(\alpha_i | \alpha_j)$ is the probability that the karyotype α_j missegregates to generate a cell with karyotype α_i (see [23, 18] for details). 4 is solved numerically for the sampled timepoints, then normalized to find $\hat{u}_{it}$. The likelihood is computed similarly to the frequent clones, using:

$$l_{it} = P(y_{it}; \hat{u}_{it}, \sum_i y_{it}) + \sum_j Q(f_i - f_j), \tag{5}$$

where $Q(f_i - f_j)$ is a prior distribution for the fitness gradient between karyotypes (taken to be normal). The parameters of the prior are estimated from the previously estimated frequent karyotyopes. Thus we find values of $f_{i \notin S}$ and v_0 which maximise $\sum_{i \notin S} \sum_t \log l_{it}$.

All other karyotypes

Kriging (gaussian process regression) was used to infer fitness of all other karyotypes. Kriging was implemented using the Krig function in the R package fields. Karyotypes of all frequent clones and their Von Neumann neighbours were used as the input matrix of independent variables and their accompanying fitness estimates were used as the vector of dependent variables. The model was fit with no drift component (function parameter m=1), otherwise all arguments to the Krig function were set to their defaults.

Cross validation procedure

.

We developed a leave-one-out cross validation procedure as a metric to judge the performance of our inference pipeline. For this procedure the fitness of the frequent karyotypes is estimated as normal. Since these fitness estimates $f_{i \in S}$ are relatively robust we treat them as a ground truth for the cross validation procedure. Then, steps 2 and 3 of the pipeline are repeatedly applied while leaving out each of the frequent karyotypes in turn. Each pipeline trained on the reduced data set is then used to predict the fitness of the omitted clone, resulting in a second set of fitness estimates for the frequent karyotypes $g_{i \in S}$. Finally the R^2 value between $f_{i \in S}$ and $g_{i \in S}$ is calculated and used as the metric to evaluate the inference pipeline performance.

Agent Based Model

We developed a stochastic, agent-based model of karyotype evolution. The model treats cells as individual agents with their own karyotype and fitness value. The model advances in fixed time steps Δt. At each timestep, cells may divide at a rate determined by their karyotype specific fitness value (which is treated directly as division rate). Dividing cells may mis-segregate, with the probability of each chromosome of a dividing cell mis-segregating to a specified as specified previously [23, 18]. The ABM features exponential growth and no cell death. To maintain a finite size population, we implemented a 'bottleneck' procedure. We measured population size at each timestep and if the population size exceeded a fixed threshold, we would select a fixed fraction of cells at random (typically 90%) and immediately remove them from the simulation. This mimics cell growth in culture conditions, where cells grow exponentially in ideal conditions then are passaged to remove a majority of cells before confluence is reached. The model was developed in C++ using standard library headers.

Gaussian random fields

Gaussian random fields (GRF) were used to generate artificial fitness landscapes of varying complexity. Each GRF is a continuous field resulting from mapping each point in K dimensional karyotype space to a fitness value. GRF are defined here by a matrix R with rows r_i corresponding to points in the K dimensional space. Each r_i acts as a point source for a wave of wavelength λ, the interference pattern of which creates the GRF. Thus for a karyotype α in the K dimensional space, fitness is evaluated according to:

$$\sum_i sin(\frac{||r_i - \alpha||}{\lambda}).$$ (6)

Results

Clonal evolution on random fitness landscapes

The ALFA-K methodology predicts single-cell karyotype evolution using longitudinal data as input. Sequencing data from individual cells is used to identify temporal changes in karyotype frequency. The method estimates the fitness of common karyotypes from changes in their observed frequencies and extends these estimates to rarer ones, constructing a local fitness landscape via Gaussian process regression (Fig. 1A). To test the ability of ALFA-K to infer fitness landscapes, we developed ABM simulations of karyotype diversification and selection on randomly generated fitness landscapes. We used Gaussian random fields (GRF) as fitness landscapes, which were generated via summation of multiple spherical waves (Fig. 1B). Varying the wavelength parameter (λ) allows control of the complexity of the resulting landscape (Fig. 1C). ABM simulations of cell populations evolving on the GRF landscapes are characterized by expansion and contraction of karyotype-defined subclones (Fig. 1D,E) as fitter clones are generated and the fitness of the population increases (Fig. 2A). In these test simulations, populations on landscapes with low λ (Fig.1D) tended to experience punctuated evolution, whereas populations evolving on landscapes with high λ exhibited gradual increases in fitness over time (Fig.1E).

Validating ALFA-K against output of ABM simulations

Data from the first 200 days of each simulated population was used to train ALFA-K. We trained ALFA-K varying both the number of longitudinal samples and the hyperparameter N, then evaluated the results using a cross validation procedure (Fig.2B) which tests the ability of ALFA-K to infer the fitness of karyotypes withheld from the input data (see Methods and Supplemental Text). ALFA-K performance was not sensitive to the value of the hyperparameter N within the tested range. It was however sensitive to landscape complexity (λ) and the number of longitudinal samples in the input, with at least 4 samples needed to obtain satisfactory results. We next tested the ability of ALFA-K to predict population evolution for the time period from 200-300 days that was withheld from the training data. We used the angle metric (SX Fig.) to evaluate predictive performance, in which values below 90 degrees are taken as good predictions. Landscapes with positive cross-validation scores ($R^2 > 0$) predicted future population evolution much better than those with negative scores (Fig. 2C). The results from the forward prediction tests agreed with the cross validation test in terms of sensitivity to landscape complexity, number of sampled timepoints, and the value of the hyperparameter N (Fig. 2D). Finally,

Figure 1: Clonal evolution on random fitness landscapes. A) Schematic flowchart of steps in ALFA-K pipeline. B) 2D example of GRF generation via summation of multiple spherical waves. The interference patterns generated by the waves result in complex unpredictable landscapes. C) Increasing the wavelength (λ) results in GRF with decreasing complexity. D-E) Example simulation output for ABM cell populations evolving on GRF fitness landscape with D) $\lambda = 0.2$ or E) $\lambda = 1.6$. Each coloured line represents the longitudinal frequency of a different karyotype.

we evaluated the robustness of ALFA-K to different values of missegregation rate (Fig. 2E). The procedure was robust to a wide range of missegregation rates up to a threshold value, which occurred when karyotype became too unstable to estimate the fitness of subpopulations across multiple longitudinal timepoints. Up to this threshold however, increasing missegregation rate benefits ALFA-K by allowing a larger region of the fitness landscape to be charted.
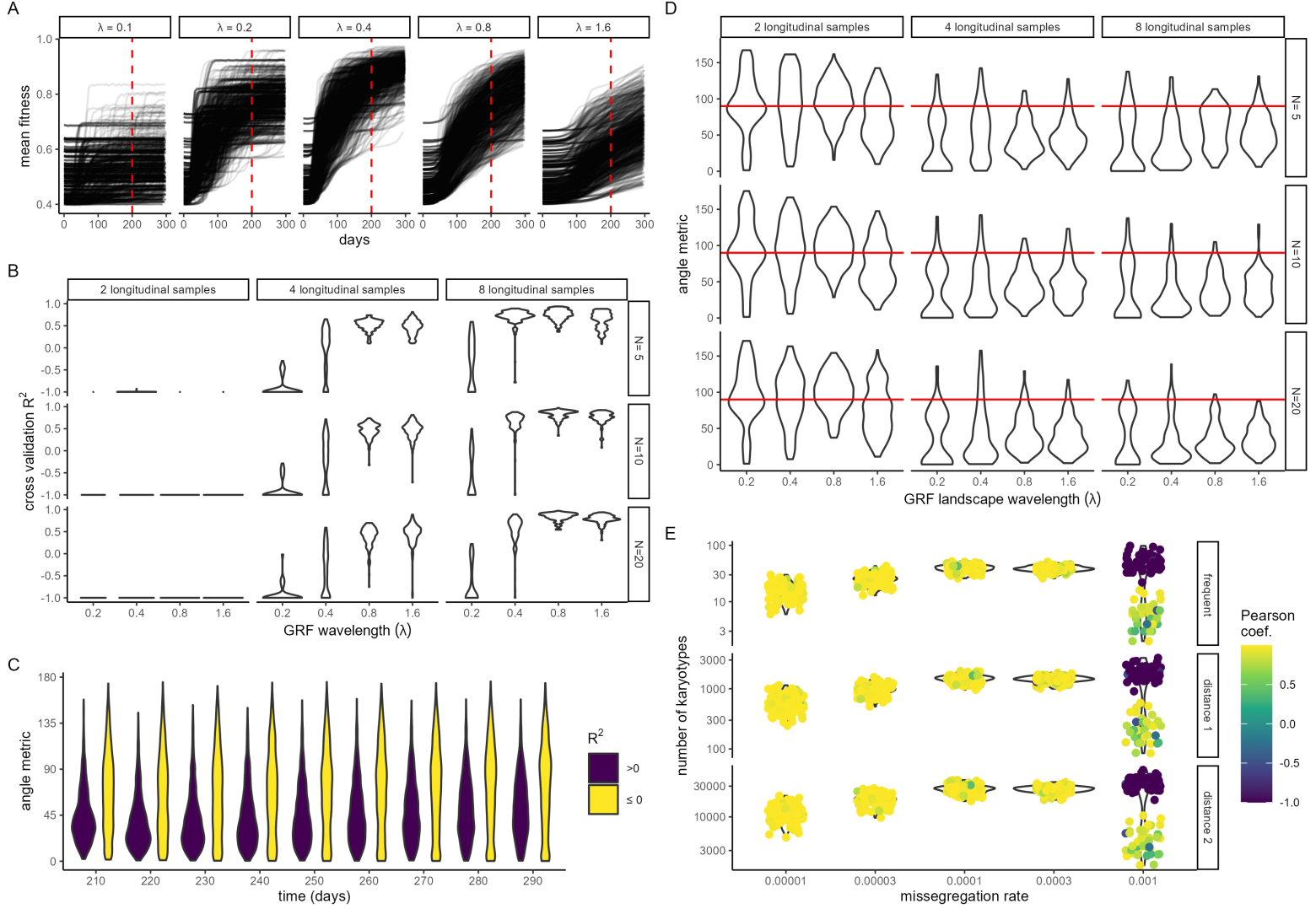

Figure 2: Validation of ALFA-K against synthetic data. A) Mean fitness of ABM cell populations evolving on artificial fitness landscapes of varying complexity (as determined by λ). Data from the first 200 days of each simulation (as demarcated by the dashed vertical lines) was used to train ALFA-K. B) Cross validation results of ALFA-K for varying numbers of input longitudinal samples and values of the hyperparameter N. c) Evolutionary prediction results of ALFA-K aggregated across all landscapes at various times in the validation period. Results are grouped by performance on the cross validation test and summarized by the angle metric. D) Evolutionary prediction results of ALFA-K for varying numbers of input longitudinal samples and values of the hyperparameter N. Prediction results are summarized by the angle metric. E) ALFA-K was trained on output from simulations with various missegregation rates. Performance was judged based on the Pearson correlation between ground truth and ALFA-K estimated fitness. Each point represents a single fitted landscape.

Fitting ALFA-K model to passaging data

ALFA-K was fit to all lineages of data from passaging experiments by Salehi et al. A lineage was defined as two or more successive samples from the same cell line. Thus e.g. for SA532 there are 7 2-sample lineages, 6 3-sample lineages etc. Fig. 3A shows the relationship between samples for all 5 cell lines. Points are colored according to the R^2 metric score obtained after fitting ALFA-K to the longest lineage terminating at that point. We were not able to obtain good fits to SA532 or SA906 lineage B. Examining R^2 metric scores for all lineages obtained from the data, we found in agreement with our ABM results that longer lineages result in better fits (Fig. 3B). We next applied ALFA-K to lineages with descendants, simulated population evolution, then compared the simulated populations to the withheld samples from subsequent passages using the angle metric(Fig. 3C). We were able to obtain good predictions for the next passage, although predictions were worse for lineages that changed cisplatin treatment status during the interval across which predictions were made. We also examined angle metric scores for sample population pairs within the experimental data, at one passage following lineage splitting (Fig. 3D). Population changes were more similar when both lineages shared the same cisplatin treatment status, in agreement with the ABM/ALFA-K predictions. The distribution of angle metrics were similar to the

ABM/ALFA-K (Comparing between Figs 3C-D), indicating that our predictions were accurate within the level variability observed in the data.
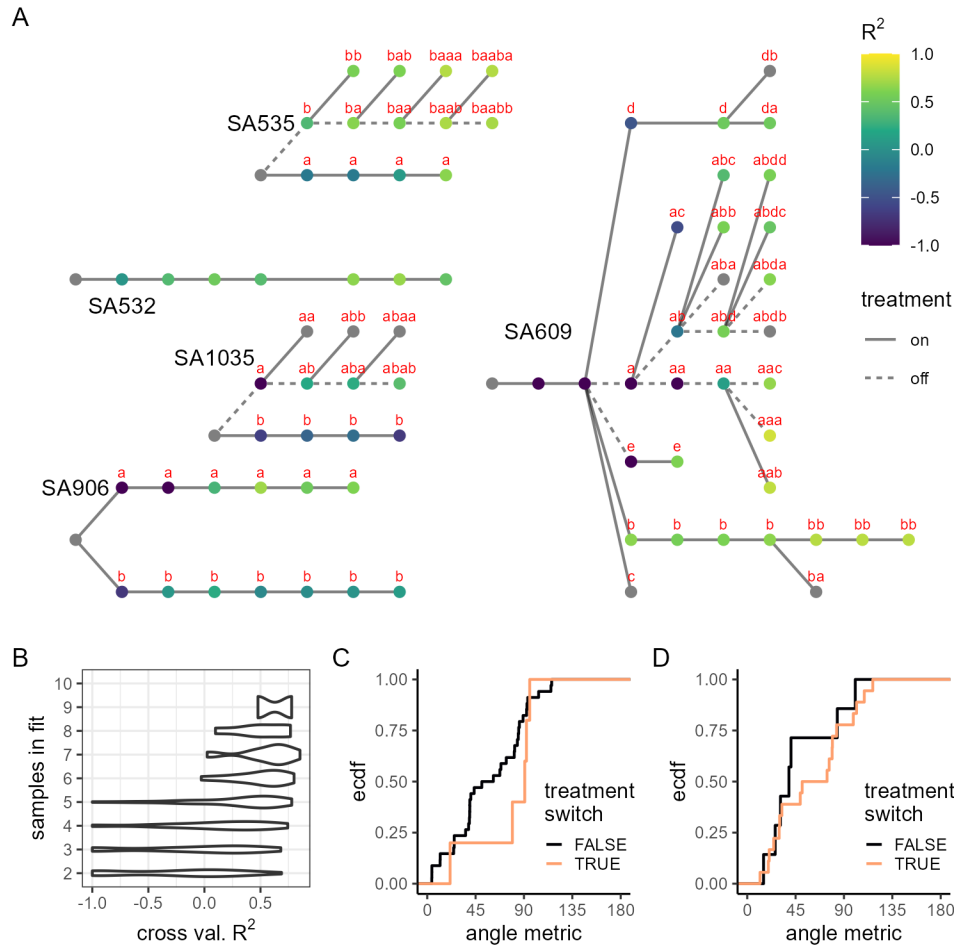

Figure 3. Fits to cell passaging data A) Maps showing the relationship cell passages that were sequenced and used as inputs to ALFA-K. Each filled circle represents a sequenced sample. Lines connect parental passages to children, with line type indicating the presence or absence of cisplatin treatment in the intervening period. Circles are colored according to the R^2 metric value for fitted landscapes including that sample and all its ancestors. B) R^2 metric scores for all sublineages in the dataset, grouped according to the number of samples used in the fitting. C) Cumulative distribution of angle metric scores between ALFA-K/ABM predicted populations and cultured cell populations. Distributions are grouped according to whether the cisplatin treatment status of the cultured population changed during the prediction interval. ALFA-K landscapes with negative R^2 metric score were excluded. D) Cumulative distribution of angle metric scores between cultured populations after lineage branching points. Distributions are grouped according to whether the cisplatin treatment status of lineage pairs differed.

Predicting novel karyotypes

It was previously shown **??** using the same dataset that the evolution of karyotype-defined subpopulations could be predicted. The previous prediction extrapolated the dynamics of previously observed karyotypes, but was not designed to predict the fitness or emergence of novel karyotypes not yet observed in the data. We asked whether ALFA-K can help predict the emergence of such novel karyotypes. Denote Θ the set of karyotypes with fitness estimates from ALFA-K, ζ the subset of Θ observed in a given longitudinal sample. Then we would like to predict Ψ (the subset of Θ that will be present in a future sample) (Fig. 4A). In particular we wish to predict which new karyotypes will emerge in the next sample, $'\zeta \cap \Psi$ (Fig. 4B). The probability of any novel karyotype actually emerging presumably depends both on its fitness and its number of neighbours in the preceding generation. Therefore for each member of $'\zeta$ we computed the fraction of karyotypes in ζ that were between 1-5 missegregations distant (Fig. 4C). These were used as variables which, together with the fitness estimate, were used to preduct whether the karyotype would emerge. For prediction we used binomial logistic regression, then assessed whether each predictor variable was significantly correlated with the response variable. As expected, the fraction of ζ that were distance-1 neighbours (d_1) was a very signficant predictor of novel karyotype emergence (Figs. 4D-E). The fitness estimates from ALFA-K also contributed significantly to the prediction, being significant in 30/45 tests (Fig. 4D) and th most significant predictor in 12/45 tests (Fig. 4E). These results indicate that ALFA-K fitness estimates can help predict emergence of new karyotypes.
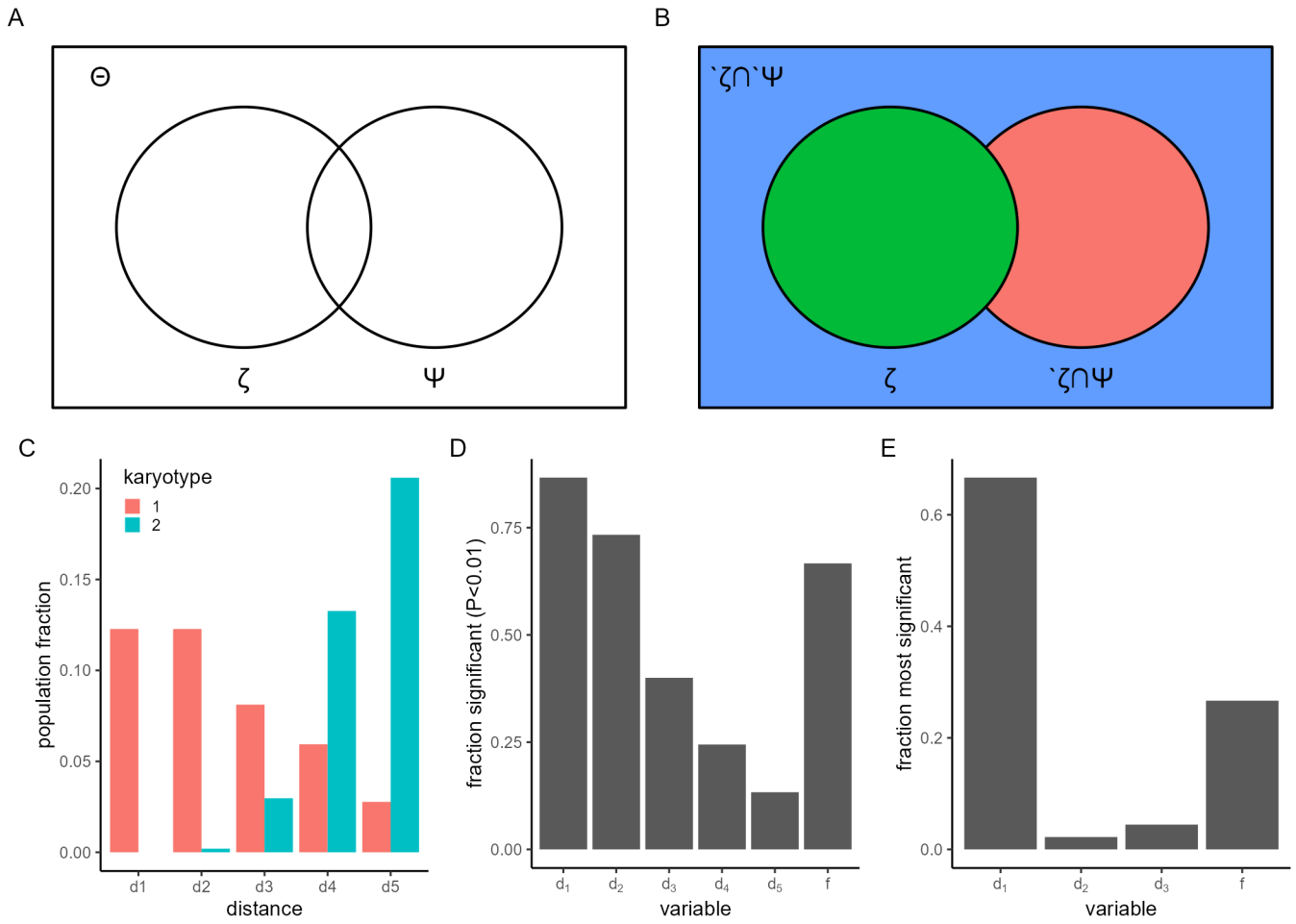
Figure 4 Predicting novel karyotypes. A) Venn diagram representing all karyotypes in the fitted landscape (Θ), the subset observed in the latest sample (ζ), and the subset that will be present in a future sample (Ψ). B) Θ is separated into 3 disjoint subsets. The aim is to predict $'\zeta \cap \Psi$. C) Distance feature vector assigned for two example karyotypes. d_i is the fraction of karyotypes in ζ that are i missegregations away from the karyotype. D) Fraction of fitted lineages in which each variable was significant (P>0.01) E) Fraction of fitted lineages in which each variable was most significant. Significance (D-E) was determined by the P-value of the test statistic for each fitted parameter.

Whole genome doubling is associated with faster evolution

It has been hypothesized that WGD facilitates aneuploidy tolerance [?]. We explored this using landscapes fit to the SA906 lineages, which were unique among the cell lines in our dataset in that a WGD+ subpopulation emerged during the experiment. In both lineages, fitness increased with distance from a euploid state (Fig. XXA)(note that ALFA-K does not chart fitness landscape regions that are not explored by the cells, so this result does not preclude the existence of low fitness karyotypes). Interestingly, at around 5 missegregations distant from the diploid state aneuploidy associated fitness increases for the WGD- subpopulations appeared to plateau, indicating that no easily accessible karyotype states would yield a fitness benefit. There was no evidence of such a plateau for WGD+ cells. This result was supported by a comparison of the estimated fitness effects of karyotype changes for the frequently observed subclones in the experimental data (Fig. XXb). These effects different significantly for WGD+ cells (KS test: $P < 10^{-8}$), with more beneficial karyotype changes available than their WGD- counterparts. Overall these results support the notion that WGD faciliates aneuploidy by increasing the number of advantageous karyotype states.
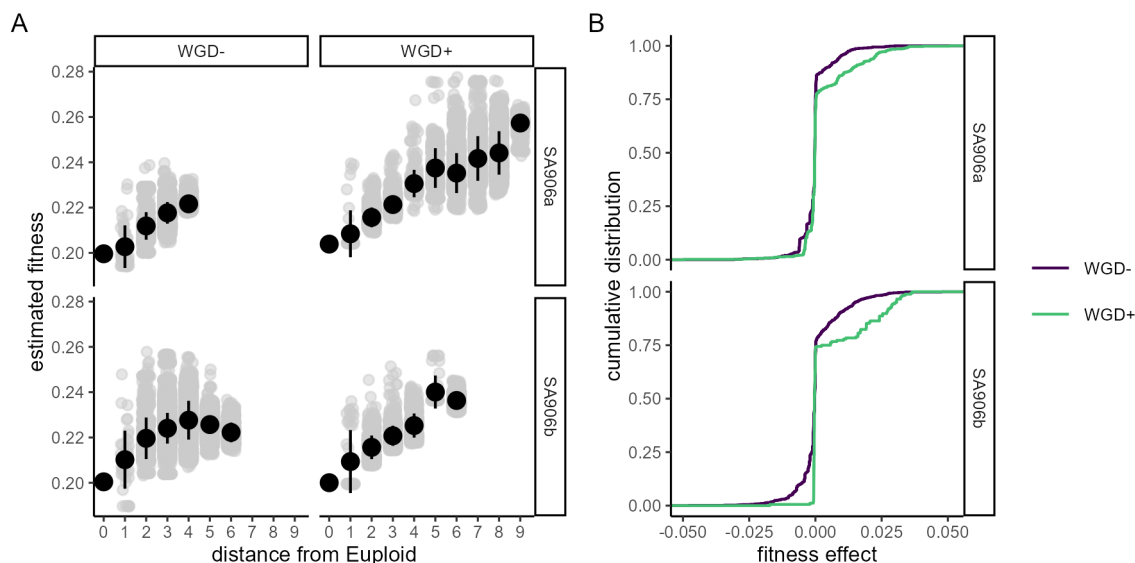

Figure XX Effects of WGD on fitness landscapes. A) Fitness of all karyotypes estimated with SA906 lineages a and b. Bars indicate mean +/- s.d. B) Empirical cumulative distribution function of fitness changes associated with all single-chromosome mis-segregations evaluated for frequent karyotypes in SA906a and SA906b.

Karyotypic background determines fitness effects of CNAs

We sought to compare fitness landscapes derived from the various in vitro and PDX datasets. For individual karyotypes within these landscapes, we computed copy number alteration profiles (CNAPs) i.e., the fitness deltas (Δf) resulting from each possible CNA (Fig.5A). To assess whether there were any broad changes in CNAPs across experimental conditions, we aggregated the Δf for karyotypes per condition. We observed larger Δf in PDX vs. in vitro landscapes (Fig.5B). In PDX landscapes, cisplatin-treated lineages showed greater fitness deltas (Δf) than those untreated (Fig.5B). This indicates the strongest selection pressure is in cisplatin-treated PDX landscapes, followed by untreated PDX, and is least within the in vitro landscapes. Overall, the results indicate that ALFA-K fitted landscapes could be used to compare the stringency of selection various contexts.

We then explored how a CNA's impact on fitness varies with the karyotypic context using our model landscapes. We compared pairs of karyotypes within the same landscape and analyzed the correlation between their CNAPs. The correlation strength was inversely proportional to the manhattan distance between the copy number profiles of the karyotype pair (Fig.5C). These results indicate that within a fixed biological context, similar karyotypes will experience similar fitness effects from a given CNA. However, if the karyotypes are not similar the fitness effects of a particular CNA are unlikely to correlate. We asked whether CNAPs from similar karyotypes derived from different fitness landscapes would correlate. We first used landscapes fitted to data from different PDX cell lines (Fig.5D). We saw a small but significant positive correlation, of comparable magnitude to karyotype pairs taken from identical landscapes. Unfortunately however, there were no karyotypes separated by a manhattan distance less than 5, so we could not assess whether the higher correlations observed for very similar karyotypes were also present across cell lines. We also checked whether CNAPs from landscapes fit to independent lineages within a single cell line would correlate. For the SA609 cell line we had three independent lineages which passed quality control, two untreated and one treated (Fig. 5E). CNAPs across the two untreated lineages exhibited much more evidence of correlation than CNAPS compared between either untreated lineage and the cisplatin treated lineage (Fig.5F). For SA535 we had available two lineages, one untreated and the other cisplatin treated(Fig. 5E). CNAPs exhibited significant correlation across these two lineages (Fig.5G). Finally we used the angle metric to compare the similarity of the evolutionary paths taken by karyotype populations from the previously analysed lineages (i.e. Figs 5F-G). For SA609, low angle metrics between the untreated lineages indicate similar evolutionary paths whilst high angle metrics around 90 degrees indicated a different evolutionary path for the untreated lineage (Fig.5H). For SA535, the low angle metric values indicate similar evolutionary paths between untreated and cisplatin treated lineages

(Fig. 5I). Overall, our findings strongly suggest that a CNA's fitness effects rely on the parent cell's karyotype. Our results also give some indication that karyotypic fitness is dependent on cellular context. However, more data from populations with similar karyotypes in varying biological contexts would be needed to fully address this question.

Missegregation rate influences karyotype dominance

In most circumstances the fittest karyotype is expected to eventually become dominant within a population. We asked whether this status quo could be altered simply by changing the missegregation rate. We envisage two ways this could occur (Fig.6A): 1) karyotypes with more high-fitness neighbors should have an advantage when missegregation rate is high, since their numbers are bolstered by a continuous (albeit small) influx from their neighbors. 2) Low ploidy cells should have an advantange when missegregation rate is high, since their reduced total number of chromosomes mean a reduced per-cell risk of missegregation, so their populations should be more stable. Overall if the fittest karyotype has few neighbors or has high ploidy, it may not be dominant when the missegregation rate is high.

To identify possible missegregation dependendent switches in karyotype distribution, we generated approximate transisition matrices connecting all karyotypes in the charted region of each ALFA-K fitted landscape. We then determined the dominant eigenvectors of these transition matrices at varying missegregation rates to find the steady state karyotype distributions. For a subset of fitted landscapes we indeed found a missegregation rate dependentent switch in karyotype distribution (Fig.6 B), where karyotypes that were fitter (Fig. 6 C,D) were dominant at lower missegregation rates but were displaced by less fit karyotypes at increasing missegregation rates. For SA906 the fitter group had significantly higher ploidy (Fig. 6E) but similar sized fitness peak to the less fit group (Fig. 6F), indicating that this was an example of a ploidy dependent frequency shift. For SA535 both groups had similar ploidy (Fig. 6E) but the fitter group had less members (Fig. 6F), indicating that this was an example of peak size dependent freuqency shift. Finally, we simulated population evolution in our ABM as a validation of our transition matrix based approximation (Fig. 6G). For SA535 we saw a shift in clonal dominance as predicted. For SA906, the karyotype population was unable to remain within the charted region of thie fitness landscape (Fig. 6H), although the karyotypes predicted to dominate at higher missegregation rates disappeared more slowly than the high fitness karyotypes which dominated when the missegregation rate was low. Overall, these results identify two ways which missegregation rate can interact with fitness to determine clonal dominance and found evidence of their occurence in cultured cell lines.

Discussion

We and others have previously modeled karyotypic evolution through the process of missegregation, employing various assumptions about the fitness associated with specific karyotypes. These assumptions range from considering all karyotypes equally fit[18], associating fitness with the density of driver or suppressor genes on each chromosome[19, 20], to correlating fitness negatively with deviations from a euploid state[23]. These models, however, have typically neglected the context in which karyotypic fitness is embedded—factors such as genetic background, tumor microenvironment, and immune interactions profoundly shape the fitness landscape. The vast number of possible karyotypes further complicates the reconstruction of fitness landscapes. Our mathematical model introduces the flexibility needed to begin reconstructing adaptive fitness landscapes, allowing us to extrapolate the fitness of thousands of karyotypes based on the dynamics of just a few subclones.

Whilst the fitness landscapes reconstructed by ALFA-K are not mechanistic, they can be leveraged to provide mechanistic insights into the evolutionary dynamics of the underlying cell populations. It has been observed that cells undergoing whole genome doubling (WGD) exhibit a higher rate of chromosomal alterations than non-WGD cells[24]. The reasons behind this accelerated evolution remain unclear: it could be due to either an increased rate of CNA generation or a heightened tolerance to the deleterious effects of CNAs in WGD cells. We explored this question using an empirical dataset from a cell line that underwent WGD during passaging[21]. Our analysis of the fitness landscape in this line suggests that WGD karyotypes display a broader spectrum of CNAs with positive fitness effects and fewer deleterious CNAs. This supports the notion that WGD may facilitate chromosomal evolution by enhancing cellular robustness to CNAs.

We also aimed to assess the extent to which CNAs would yield similar fitness effects in different karyotypes, under different treatment conditions, and across different cell lines. We found that CNA fitness effects were strongly contingent on parental karyotype, such that changing the parental karyotype by only a few missegregations was enough to completely change the set of favourable CNAs. This hampered our ability to test whether CNAs would yield similar karyotypes across cell lines, which were too far apart in karyotype space to produce a strong result. We found CNAs had similar fitness effects in parallel replicates from the same cell line. In agreement previous analysis [21] we found that parallel replicates sharing cisplatin treatment status were more similar than those that did not.

In a novel application of ALFA-K, we investigated the impact of CIN on chromosomal evolution by increasing CIN rates until populations were no longer sustainable on the highest fitness peaks of the landscape. This approach highlighted selection pressures against karyotype populations residing on narrow fitness peaks or in high ploidy regions of the landscape. These findings emerged at missegregation rates of 5-20% per cell division, a range considered plausible for CIN cells[25]. Given that CIN is therapeutically modifiable, our results underscore the potential of using evolutionary principles to steer tumor evolution for long-term control[26]. The ploidy-dependent mechanism builds on prior models, which generally assume equal missegregation rates per chromosome, causing high ploidy karyotypes to missegregate more frequently overall. While this assumption aligns with observations that high ploidy cells missegregate more frequently[24, 27], it warrants further scrutiny. Future models could benefit from integrating metrics such as interferon gamma as a
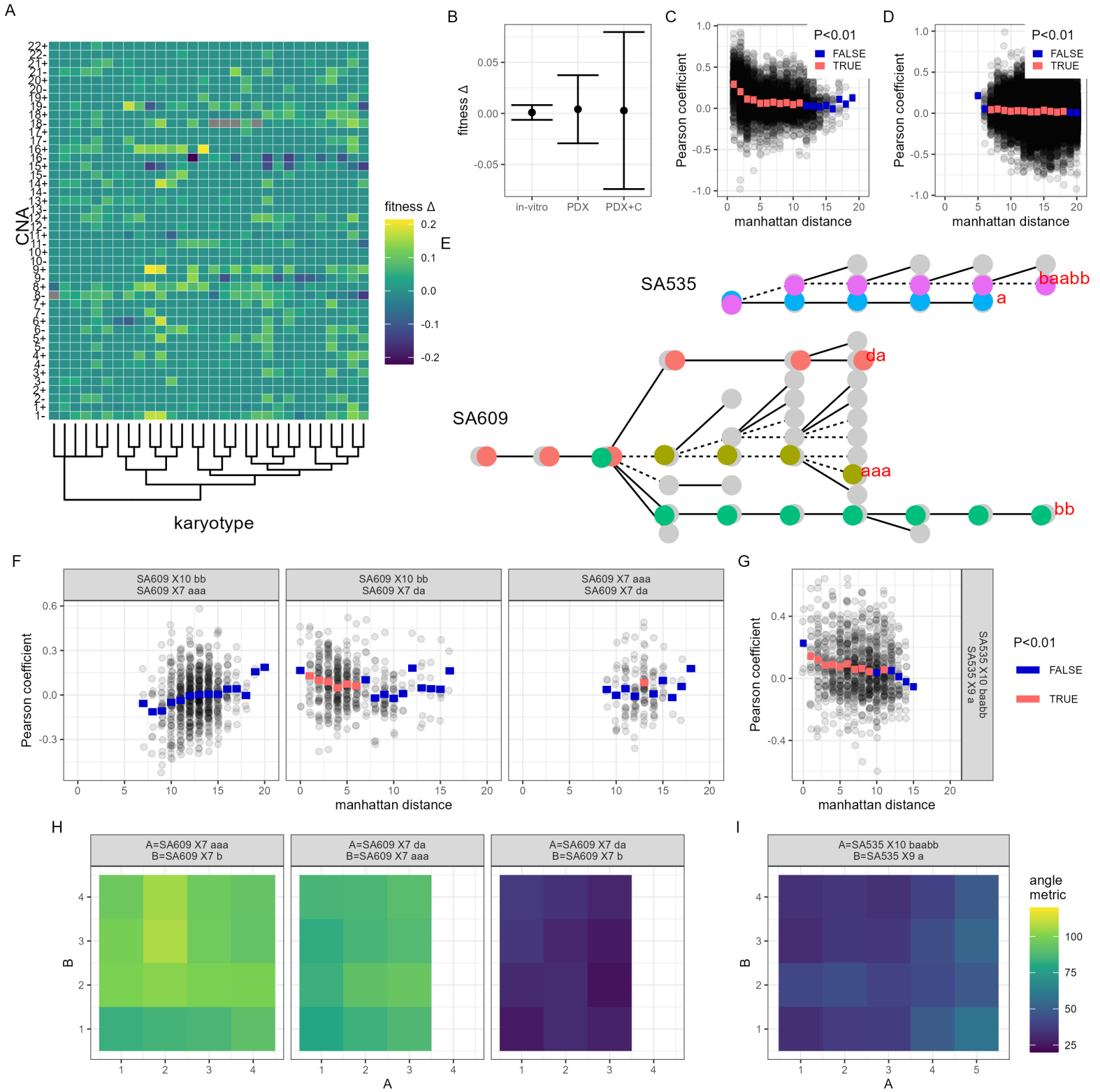
Figure 5. Comparison of fitted landscapes. A) CNAPs for karyotypes taken from a single fitness landscape. Each column in the image represents a CNAP. Karyotypes are clustered based on the manhattan distance between their copy number profiles. B) (Δf) comparison between in vitro (cultured) untreated PDX karyotypes, and PDX with cisplatin (PDX+C). C) Correlation between CNAPs within the same landscape. Each point represents the pearson correlation coefficient between two CNAPs, presented according to the manhattan distance between the two karyotypes in karyotype space. D) Correlation between CNAPs across cell lines. E) Independent, non-overlapping lineages analyzed in F-I are highlighted. F) correlation of CNAPS across independent lineages of SA609. Untreated lineages are da and bb, cisplatin treated lineage is aaa. G) correlation of CNAPS across independent lineages of SA535. P-value (F-G) from T test with null hypothesis mean=0. Only fitted landscapes with metric values $R^2 > 0.3$ were included in the analysis (C-D,F-G). H-I) Angle metrics indicate similarity of evolutionary paths in the empirical data for SA609 lineages (H) and SA535(I).
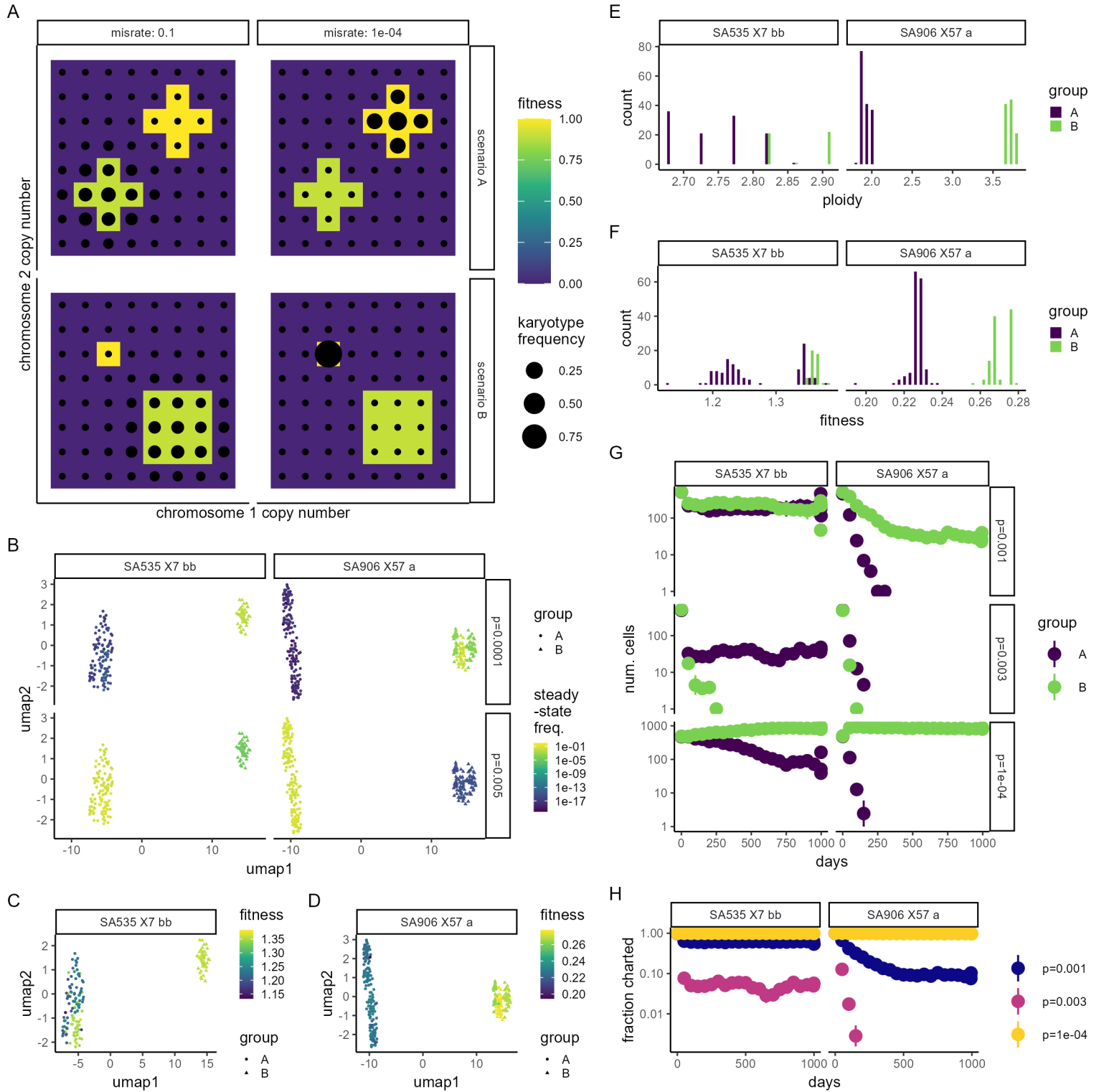
Figure 6 Influence of missegregation rate on karyotype selection. A) Simple hyothetical example fitness landscapes where karyotype frequency dependends on missegregation rate. Karyotype frequency (point size) shifts from high to low ploidy (top row) or broader to narrower peaks (bottom row) as missegregation rate changes. (B) UMAPs showing the relative frequency of karyotypes representing at least 2% of the predicted steady-state population for SA906_x57_a and SA535_X7_bb at different missegregation rates. (C,D) ALFA-K estimated fitness of karyotypes shown in B-C for SA906_x57_a (C) and SA535_X7_bb (D). E) ploidy and F) fitness distributions for the top 2% most frequent clones. G) ABM simulation of karyotype evolution on fitted landscapes for SA906_x57_a and SA535_X7_b. H) Fraction of the population remaining on charted regions of the fitness landscape during ABM simulations.

measure of missegregation rate[23], helping to discern whether a permissive fitness landscape or a propensity to missegregate underlies the relationship between WGD and aneuploidy.

One notable limitation of our study is the scarcity of longitudinal data available for training and validating ALFA-K. To mitigate this, we relied heavily on synthetic data and developed metrics to assess the accuracy of the inferred landscapes. However, the absence of independent biological replicates restricted our ability to fully evaluate the quality of our model predictions, as the inherent variability in the system was not well characterized. Future studies with identical biological replicates are essential to further validate ALFA-K[22]. Additionally, data from evolving cell populations inherently focusses on high-fitness regions, making estimation of the fitness impact of deleterious CNAs challenging. ALFA-K attempts to estimate fitness for unobserved karyotypes, using their absence to infer upper fitness limits. However, our estimates for the fitness impacts of deleterious CNAs are likely biased upwards and should be interpreted with caution. Unbiased screens that can accurately estimate the fitness impact of deleterious CNAs would significantly enhance our understanding[?]. Moving forward, incorporating fitness estimates from individual samples[28, 29], in addition to or instead of longitudinal data, could greatly expand the applicability of ALFA-K.

Future applications of this model will include more detailed studies of the fitness costs and benefits of high ploidy. In addition to expanding our characterisation of the relation between WGD and aneuploidy tolerance, our model will also help quantify the energetic requirements of high ploidy cells. Ultimately our model will be a powerful tool for studying karyotype evolution, revealing how selection acts upon coexisting karyotypes in various environments.

References

[1] Samuel F Bakhoum and Lewis C Cantley. The multifaceted role of chromosomal instability in cancer and its microenvironment. *Cell*, 174(6):1347–1360, September 2018.

[2] Stefano Santaguida and Angelika Amon. Short- and long-term effects of chromosome mis-segregation and aneuploidy. *Nat. Rev. Mol. Cell Biol.*, 16(8):473–485, August 2015.

[3] Uri Ben-David and Angelika Amon. Context is everything: aneuploidy in cancer. *Nat. Rev. Genet.*, 21(1):44–62, January 2020.

[4] Norman Pavelka, Giulia Rancati, Jin Zhu, William D Bradford, Anita Saraf, Laurence Florens, Brian W Sanderson, Gaye L Hattem, and Rong Li. Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature*, 468(7321):321–325, November 2010.

[5] Eduardo M Torres, Tanya Sokolsky, Cheryl M Tucker, Leon Y Chan, Monica Boselli, Maitreya J Dunham, and Angelika Amon. Effects of aneuploidy on cellular physiology and cell division in haploid yeast. *Science*, 317(5840):916–924, August 2007.

[6] Guangbo Chen, Boris Rubinstein, and Rong Li. Whole chromosome aneuploidy: big mutations drive adaptation by phenotypic leap. *Bioessays*, 34(10):893–900, October 2012.

[7] Juliann Shih, Shahab Sarmashghi, Nadja Zhakula-Kostadinova, Shu Zhang, Yohanna Georgis, Stephanie H Hoyt, Michael S Cuoco, Galen F Gao, Liam F Spurr, Ashton C Berger, Gavin Ha, Veronica Rendo, Hui Shen, Matthew Meyerson, Andrew D Cherniack, Alison M Taylor, and Rameen Beroukhim. Cancer aneuploidies are shaped primarily by effects on tumour fitness. *Nature*, June 2023.

[8] Teresa Davoli, Andrew Wei Xu, Kristen E Mengwasser, Laura M Sack, John C Yoon, Peter J Park, and Stephen J Elledge. Cumulative haploinsufficiency and triplosensitivity drive aneuploidy patterns and shape the cancer genome. *Cell*, 155(4):948–962, November 2013.

[9] Eduardo M Torres, Bret R Williams, and Angelika Amon. Aneuploidy: cells losing their balance. *Genetics*, 179(2):737–746, June 2008.

[10] Yun-Chi Tang and Angelika Amon. Gene copy-number alterations: a cost-benefit analysis. *Cell*, 152(3):394–405, January 2013.

[11] ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. Pan-cancer analysis of whole genomes. *Nature*, 578(7793):82–93, February 2020.

[12] Bastien Nguyen, Christopher Fong, Anisha Luthra, Shaleigh A Smith, Renzo G DiNatale, Subhiksha Nandakumar, Henry Walch, Walid K Chatila, Ramyasree Madupuri, Ritika Kundra, Craig M Bielski, Brooke Mastrogiacomo, Mark T A Donoghue, Adrienne Boire, Sarat Chandarlapaty, Karuna Ganesh, James J Harding, Christine A Iacobuzio-Donahue, Pedram Razavi, Ed Reznik, Charles M Rudin, Dmitriy Zamarin, Wassim Abida, Ghassan K Abou-Alfa, Carol Aghajanian, Andrea Cercek, Ping Chi, Darren Feldman, Alan L Ho, Gopakumar Iyer, Yelena Y Janjigian, Michael Morris, Robert J Motzer, Eileen M O'Reilly, Michael A Postow, Nitya P Raj, Gregory J Riely, Mark E Robson, Jonathan E Rosenberg, Anton Safonov, Alexander N Shoushtari, William Tap, Min Yuen Teo, Anna M Varghese, Martin Voss, Rona Yaeger, Marjorie G Zauderer, Nadeem Abu-Rustum, Julio Garcia-Aguilar, Bernard Bochner, Abraham Hakimi, William R Jarnagin, David R Jones, Daniela Molena, Luc Morris, Eric Rios-Doria, Paul Russo, Samuel Singer, Vivian E Strong, Debyani Chakravarty, Lora H Ellenson, Anuradha Gopalan, Jorge S Reis-Filho, Britta Weigelt, Marc Ladanyi, Mithat Gonen, Sohrab P Shah, Joan Massague, Jianjiong Gao, Ahmet Zehir, Michael F Berger, David B Solit, Samuel F Bakhoum, Francisco Sanchez-Vega, and Nikolaus Schultz. Genomic characterization of metastatic patterns from prospective clinical sequencing of 25,000 patients. *Cell*, 185(3):563–575.e11, February 2022.

[13] Ruben M Drews, Barbara Hernando, Maxime Tarabichi, Kerstin Haase, Tom Lesluyes, Philip S Smith, Lena Morrill Gavarró, Dominique-Laurent Couturier, Lydia Liu, Michael Schneider, James D Brenton, Peter Van Loo, Geoff Macintyre, and Florian Markowetz. A pan-cancer compendium of chromosomal instability. *Nature*, 606(7916):976–983, June 2022.

[14] Ankit Shukla, Thu H M Nguyen, Sarat B Moka, Jonathan J Ellis, John P Grady, Harald Oey, Alexandre S Cristino, Kum Kum Khanna, Dirk P Kroese, Lutz Krause, Eloise Dray, J Lynn Fink, and Pascal H G Duijf. Chromosome arm aneuploidies shape tumour evolution and drug response. *Nat. Commun.*, 11(1):449, January 2020.

[15] Kasper Karlsson, Moritz Przybilla, Hang Xu, Eran Kotler, Kremena Karagyozova, Alexandra Sockell, Katherine Liu, Amanda Mah, Yuan-Hung Lo, Bingxin Lu, Kathleen E Houlahan, Aziz Khan, Zhicheng Ma, Carlos J Suarez, Chris P Barnes, Calvin J Kuo, and Christina Curtis. Experimental evolution in TP53 deficient human gastric organoids recapitulates tumorigenesis. April 2022.

[16] Timour Baslan, John P Morris, 4th, Zhen Zhao, Jose Reyes, Yu-Jui Ho, Kaloyan M Tsanov, Jonathan Bermeo, Sha Tian, Sean Zhang, Gokce Askan, Aslihan Yavas, Nicolas Lecomte, Amanda Erakky, Anna M Varghese, Amy Zhang, Jude Kendall, Elena Ghiban, Lubomir Chorbadjiev, Jie Wu, Nevenka Dimitrova, Kalyani Chadalavada, Gouri J Nanjangud, Chaitanya Bandlamudi, Yixiao Gong, Mark T A Donoghue, Nicholas D Socci, Alex Krasnitz, Faiyaz Notta, Steve D Leach, Christine A Iacobuzio-Donahue, and Scott W Lowe. Ordered and deterministic cancer genome evolution after p53 loss. *Nature*, August 2022.

[17] Y Gusev, V Kagansky, and W C Dooley. A stochastic model of chromosome segregation errors with reference to cancer cells. *Math. Comput. Model.*, 32(1):97–111, July 2000.

[18] Y Gusev, V Kagansky, and W C Dooley. Long-term dynamics of chromosomal instability in cancer: A transition probability model. *Math. Comput. Model.*, 33(12):1253–1273, June 2001.

[19] Sergi Elizalde, Ashley M Laughney, and Samuel F Bakhoum. A markov chain for numerical chromosomal instability in clonally expanding populations. *PLoS Comput. Biol.*, 14(9):e1006447, September 2018.

[20] Ashley M Laughney, Sergi Elizalde, Giulio Genovese, and Samuel F Bakhoum. Dynamics of tumor heterogeneity derived from clonal karyotypic evolution. *Cell Rep.*, 12(5):809–820, August 2015.

[21] Sohrab Salehi, Farhia Kabeer, Nicholas Ceglia, Mirela Andronescu, Marc J Williams, Kieran R Campbell, Tehmina Masud, Beixi Wang, Justina Biele, Jazmine Brimhall, David Gee, Hakwoo Lee, Jerome Ting, Allen W Zhang, Hoa Tran, Ciara O'Flanagan, Fatemeh Dorri, Nicole Rusk, Teresa Ruiz de Algara, So Ra Lee, Brian Yu Chieh Cheng, Peter Eirew, Takako Kono, Jenifer Pham, Diljot Grewal, Daniel Lai, Richard Moore, Andrew J Mungall, Marco A Marra, IMAXT Consortium, Andrew McPherson, Alexandre Bouchard-Côté, Samuel Aparicio, and Sohrab P Shah. Clonal fitness inferred from time-series modelling of single-cell cancer genomes. *Nature*, 595(7868):585–590, July 2021.

[22] Kasper Karlsson, Moritz J Przybilla, Eran Kotler, Aziz Khan, Hang Xu, Kremena Karagyozova, Alexandra Sockell, Wing H Wong, Katherine Liu, Amanda Mah, Yuan-Hung Lo, Bingxin Lu, Kathleen E Houlahan, Zhicheng Ma, Carlos J Suarez, Chris P Barnes, Calvin J Kuo, and Christina Curtis. Deterministic evolution and stringent selection during preneoplasia. *Nature*, 618(7964):383–393, June 2023.

[23] Gregory J Kimmel, Richard J Beck, Xiaoqing Yu, Thomas Veith, Samuel Bakhoum, Philipp M Altrock, and Noemi Andor. Intra-tumor heterogeneity, turnover rate and karyotype space shape susceptibility to missegregation-induced extinction. *PLoS Comput. Biol.*, 19(1):e1010815, January 2023.

[24] Emma V Watson, Jake June-Koo Lee, Doga C Gulhan, Giorgio E M Melloni, Sergey V Venev, Rayna Y Magesh, Abdulrazak Frederick, Kunitoshi Chiba, Eric C Wooten, Kamila Naxerova, Job Dekker, Peter J Park, and Stephen J Elledge. Chromosome evolution screens recapitulate tissue-specific tumor aneuploidy patterns. *Nat. Genet.*, February 2024.

[25] Natalia Vargas-Rondón, Victoria E Villegas, and Milena Rondón-Lagos. The role of chromosomal instability in cancer and therapeutic responses. *Cancers*, 10(1), December 2017.

[26] Robert A Gatenby, Ariosto S Silva, Robert J Gillies, and B Roy Frieden. Adaptive therapy. *Cancer Res.*, 69(11):4894–4903, June 2009.

[27] Simon Gemble, René Wardenaar, Kristina Keuper, Nishit Srivastava, Maddalena Nano, Anne-Sophie Macé, Andréa E Tijhuis, Sara Vanessa Bernhard, Diana C J Spierings, Anthony Simon, Oumou Goundiam, Helfrid Hochegger, Matthieu Piel, Floris Foijer, Zuzana Storchová, and Renata Basto. Genetic instability from a single S phase after whole-genome duplication. *Nature*, 604(7904):146–151, April 2022.

[28] Noemi Andor, Billy T Lau, Claudia Catalanotti, Anuja Sathe, Matthew Kubit, Jiamin Chen, Cristina Blaj, Athena Cherry, Charles D Bangs, Susan M Grimes, Carlos J Suarez, and Hanlee P Ji. Joint single cell DNA-seq and RNA-seq of gastric cancer cell lines reveals rules of in vitro evolution. *NAR Genom Bioinform*, 2(2):lqaa016, June 2020.

[29] Olivia Lucas, Sophia Ward, Rija Zaidi, Abigail Bunkum, Alexander M Frankell, David A Moore, Mark S Hill, Wing Kin Liu, Daniele Marinelli, Emilia L Lim, Sonya Hessey, Cristina Naceur-Lombardelli, Andrew Rowan, Sukhveer Mann, Haoran Zhai, Michelle Dietzen, Boyue Ding, Gary Royle, Nicholas McGranahan, Mariam Jamal-Hanjani, Nnennaya Kanu, Charles Swanton, and Simone Zaccaria. Abstract PR010: Linking proliferation rate to the evolution of single-cell primary and metastatic tumor clones. *Cancer Res.*, 84(3_Supplement_2):PR010–PR010, February 2024.

Supplementary Text

We developed a number of metrics to evaluate the accuracy of fitness landscapes generated by ALFA-K. Metrics were developed using ALFA-K fits to data from simulations, so that ground truth fitnesses would be available from the GRF artificial landscapes. Across all fits we saw a Pearson correlation median 0.83 between predicted and true fitness. Accuracy was positively correlated with the number of sample timepoints (Fig. S1A) and with λ (Fig. S1B), was negatively correlated with distance from frequent karyotypes (Fig. S1C), but had little correlating with the frequency threshold - a hyperparameter used in fitting (Fig. S1D). ALFA-K did fail to produce accurate fits in some cases (Fig. S1E). Our main concern was to design metrics to identify such instances, for applications where no ground truth fitness estimates will be available.
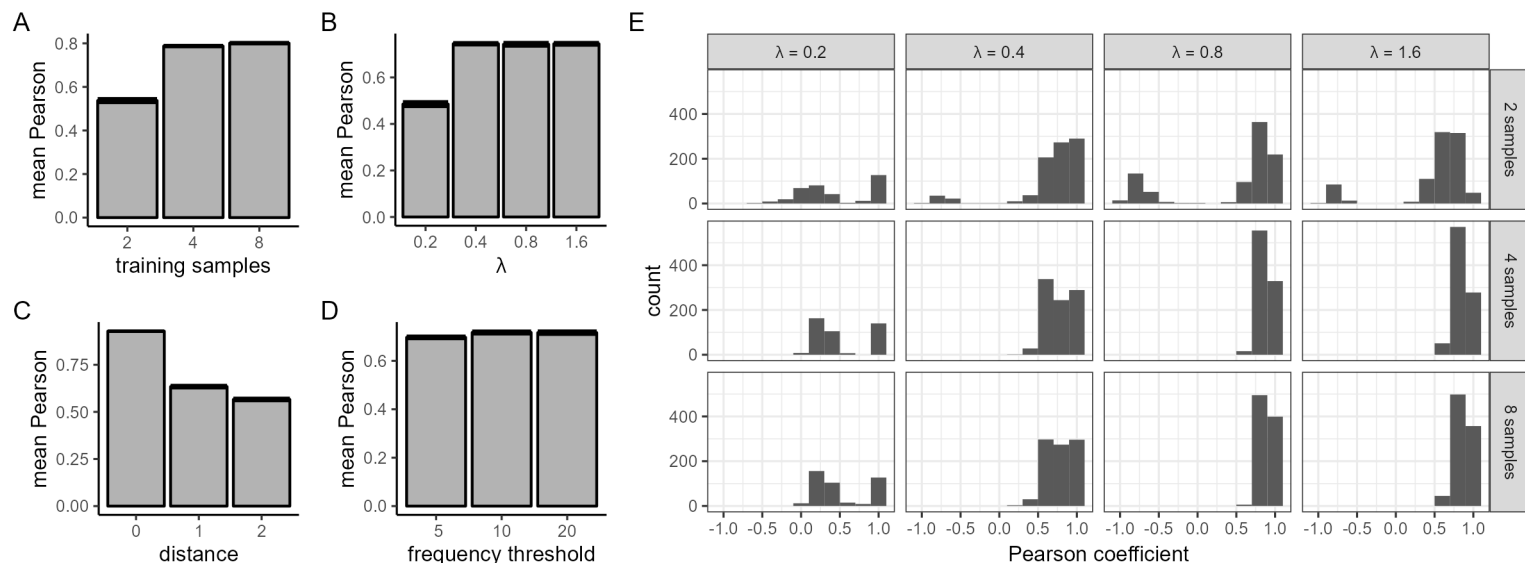

Figure S1. A-D) Pearson correlation coefficient of ALFA-K estimated fitness with ground truth fitness, grouped by A) Distance of karyotype from the nearest 'frequent karyotype' estimated in step 1 of the fit; B) Number of longitudinal samples used to train ALFA-K; C) frequency threshold for frequent karyotypes; D) GRF wavelength parameter λ. E) Pearson coefficients between actual fitness and ALFA-K estimated fitness of karyotypes.

We employed a leave-one-out cross validation procedure to test the ability of ALFA-K to estimate the fitness of karyotypes not present in the input data (Fig. S2). The fitness estimates from the first step of the ALFA-K pipeline (based on frequency changes of common karyotypes) are treated as "ground truth" and used as inputs for the cross validation test.
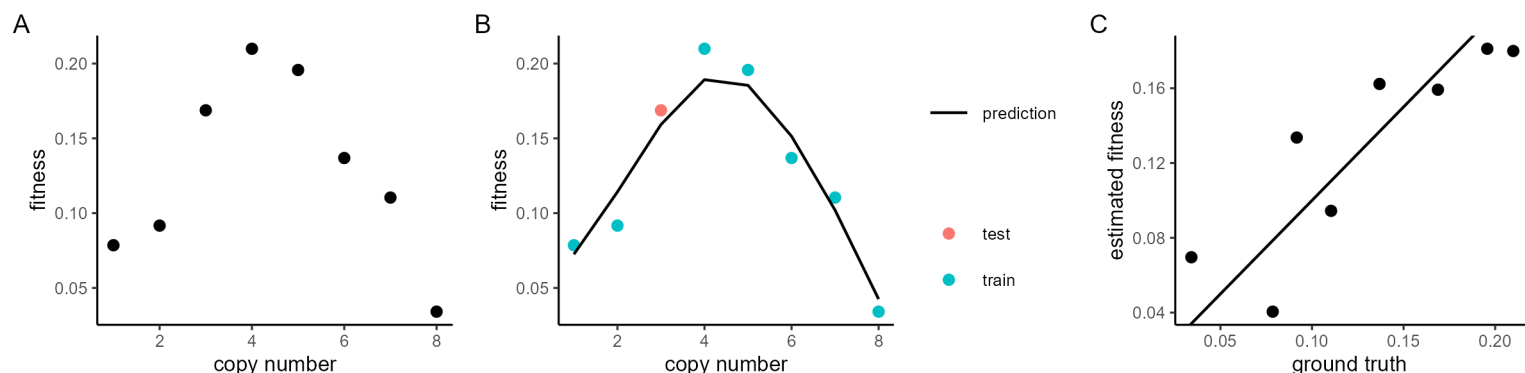

Figure S2 A) A simplified hypothetical example in which the fitness of a karyotype depends on the copy number of a single chromosome. B) In the cross validation procedure, each data-point is omitted from the training set in turn. The model is trained on the reduced dataset and then used to predict the fitness of the omitted clone. C) Cross validation predicted fitnesses are compared to the ground truth.

The cross validation metric performed extremely well at filtering out poor fits, with almost no fits having a positive R^2 but negatively correlating with the ground truth. However, the metric was perhaps overzealous in the sense that many simulations scored poorly despite having predicted fitnesses that correlated extremely strongly with the ground truth (it would be nice to explain why this happens, but I can't figure out why).

We also sought to identify metrics based on the combined ability of the ABM and ALFA-K to predict future karyotype population evolution. We developed and tested two metrics: the Wasserstein metric, and the angle metric. Suppose two initially indentical karyotype populations evolve independently across two fitness landscapes, which could be either identical, or different (as shown, Fig.

S3A-B). Two vectors are computed connecting the centroid of the initial population to the centroids of each of the evolved populations (Fig. S3C). The angle metric is the angle between these two vectors. Alternatively the Wasserstein distance between the two evolved populations can be computed (Fig. S3D, d_3). If the populations are evolving slowly relative to the timescale of measurement, the distance d_3 will automatically be low. Therefore we normalise by the mean distance each population has travelled from the founder, giving the Wasserstein metric $M_w = 2d_3(d1 + d2)^{-1}$.
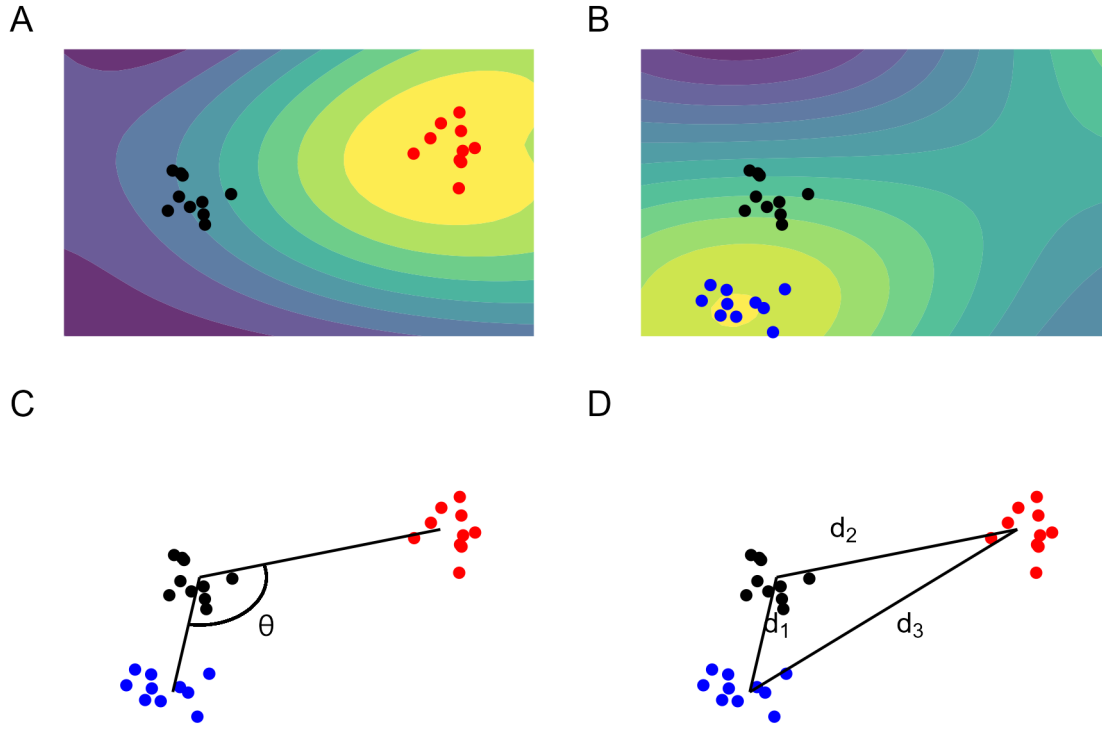

Figure S3. Metrics comparing similarity of population evolution. A-B) Examples of an initially identical founder population (black) evolving on two different landscapes. C) The vector angle between the centroids of the two evolved populations gives the angle metric. D) The Wasserstein metric is computed as the Wasserstein distance between the two evolved populations (d_3) normalised by the average distance the populations have travelled (d_1,d_2).

To aid intepretation of the Wasserstein and angle metrics, we generated pairs of vectors uniformly distributed on the surface of the 22-dimensional hypersphere and computed the values of the metrics for each vector pair (Fig. S4). The expected value of the angle metric was 90° while the Wasserstein metric was $\sqrt{2}$. Therefore one can say that recovering values less than these thresholds indicates better agreement than random chance.
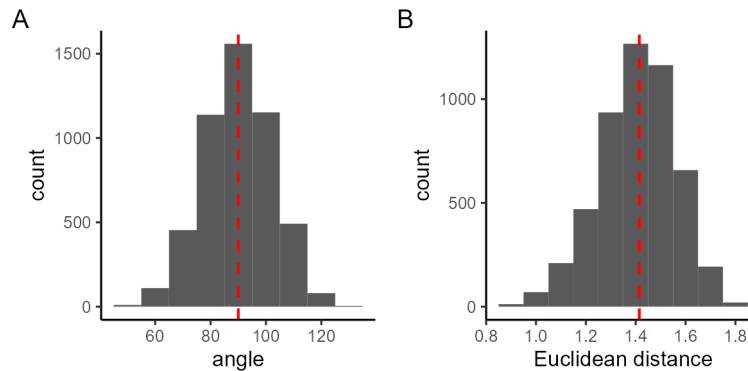

Figure S4. Metric expectation. Pairs of vectors were generated uniformly on the surface of a 22-dimensional hypersphere. The distribution of A) angles, and B) distances between each pair of vectors is shown.

We tested the ability of each metric to characterise whether an ALFA-K estimated landscape was accurate or not. We applied the R^2 metric to ALFA-K landscapes fit to data from our ABM test simulations. These fitted landscapes were also used as input to a second round of ABM simulations, to which we applied the angle and Wasserstein metrics. All three metrics were able to discriminate between landscapes that correlated strongly with the ground truth and those that did not.
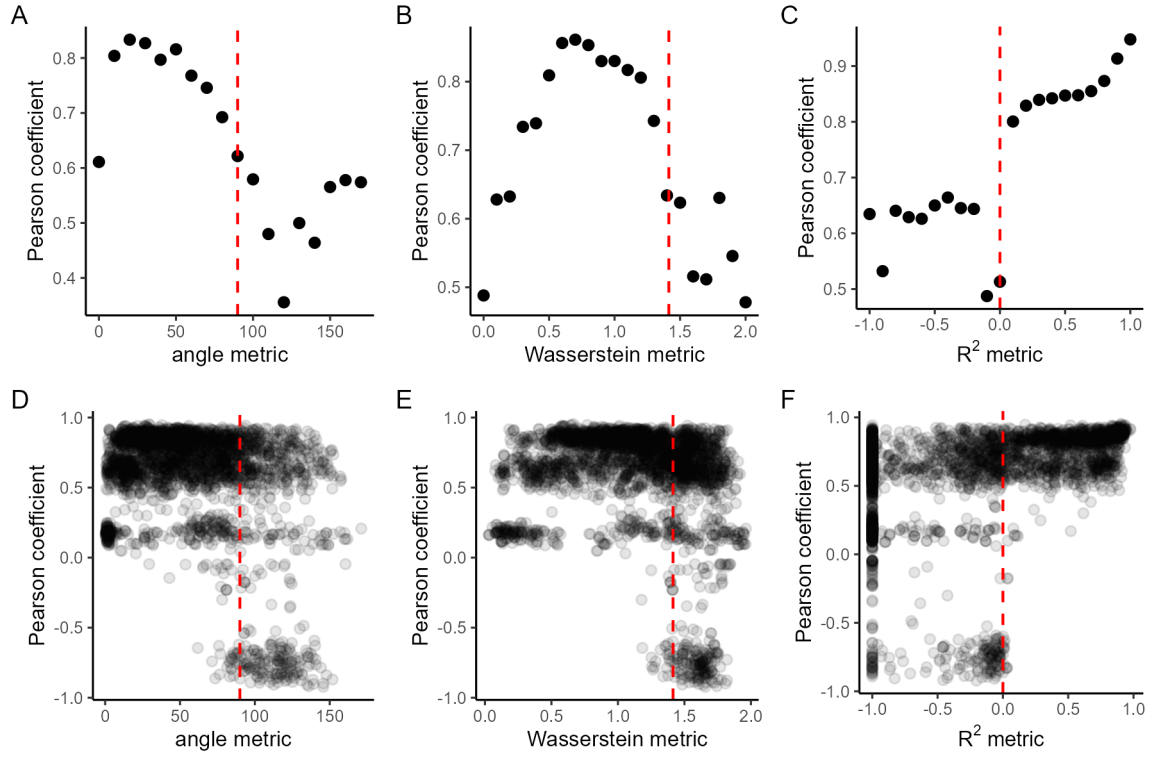
Figure S5. Evaluation of metrics. A-C) The mean value of Pearson correlation between ground truth and ALFA-K fitted landscape, for corresponding values of A) angle metric, B) Wasserstein metric, and C) R^2 metric. D-F) Comparison of Pearson correlation and D) angle metric, E) Wasserstein metric, and F) R^2 metric for all fitted landscapes.