Local Adaptive Mapping of Karyotype Fitness Landscapes

Richard J Beck[1] and Noemi Andor[1]

[1]H. Lee Moffitt Cancer Center and Research Institute, Integrated Mathematical Oncology, Tampa, Florida

Abstract

Aneuploidy occurs in most solid tumors and has the potential to dramatically modify cellular phenotype and fitness. Despite the importance of aneuploidy in tumor evolution, quantitative understanding of the evolutionary landscape of aneuploidy is lacking. To address this question, we developed a method to infer the fitness landscape of either arm-level or whole-chromosome level karyotypes. Our method takes longitudinal single cell sequencing data from an evolving cell population as input, then estimates the fitness of thousands of karyotypes located near to the input data in karyotype space. The predictive ability of the method was validated using artificial data generated from an agent based model, as well as data from a selection of in vitro and in vivo passaged cell lines. We applied our pipeline to an in vitro dataset of serially passaged cells and - based on topological analysis of the fitness landscape around diploid and tetraploid karyotypes - found support for the hypothesis that whole genome doubling benefits tumour cells by conferring robustness against missegregations.

Introduction

Losses and gains of entire chromosomes or large sections thereof - termed aneuploidy - are a defining feature of solid tumors [1, 2]. Since an estimated 90% of solid tumours are aneuploid and because aneuploidy simultaneously alters the copy number of many genes, aneuploidy affects more of the cancer genome than any other genetic alteration [3]. DNA copy number is highly correlated to RNA expression and protein production [4, 5], providing a mechanism for aneuploidy to alter cellular phenotype. Therefore aneuploidy is thought to provide a substrate for tumour evolution [1, 6, 7]. This notion is supported by the observation that chromosomes rich in oncogenes are frequently amplified whilst those rich in tumour suppressor genes are often deleted [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]. Thus aneuploidy patterns are context dependent [3]. Environmental context plays a role in sculpting karyotype, 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 present a procedure to directly estimate fitness landscapes from single-cell copy number data. We first develop an agent based model (ABM) of karyotype evolution on artificially generated fitness landscapes. We use output data from the ABM to demonstrate the predictive ability of our inference methodology. We go on to validate the predictive power of our method on previously published data from several P53 deficient cell lines which exhibited substantial subclonal evolution across multiple passages in vivo or in vitro [21]. Finally, we compared the topology of diploid and tetraploid fitness landscapes, finding support for the hypothesis that whole genome doubling benefits tumour cells by conferring robustness against missegregations.

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 N_{it} with rows i corresponding to unique karotypes and columns t corresponding to the longitudinal sampling time points. The entries of N are the number of individual cells representing each karyotype that were detected at a particular sample time point. It will be useful to also define here U_{it} as the frequency of karyotype i among the population sample at time t, and n_{it} and u_{it} as model predictions for N_{it} and U_{it}, respectively.

Frequent karyotypes

"Frequent karyotypes" are operationally defined throughout the manuscript as those observed in at least five cells across all sampled time points. When longitudinal growth rate measurements ($f_{pop,t}$) are available, the fitness of the ith clone is estimated based on its observed frequency (U_{it}) as follows. First, a trial value of f_i is supplied. Together with $f_{pop,t}$ and $U_{i,t}$ this can be used to calculate the mean fitness of all other clones $f_{j\neq i,t}$ at each measurement time point, by solving:

$$f_{pop,t} = f_i U_{i,t} + f_{j \neq i,t}(1 - U_{i,t}) \tag{1}$$

.

To model the change of frequency of the ith clone between successive measurement time points, we take $f_{j \neq i}(t_n < t < t_{n+1})$ between measurement points as the average of the two adjacent measurements. f_i and $f_{j \neq i}(t)$ together with the initial ratio of clone i to all other clones (n_0) defines an exponential growth model:

$$n_i = n_0 exp(f_i t) \tag{2}$$

$$n_{i \neq j} = exp(f_{i \neq j}(t) t) \tag{3}$$

.

The resulting predicted number of cells of the ith karyotype is then renormalized:

$$u_i = \frac{n_i}{\sum_j n_j}, \tag{4}$$

such that u_{it} represents the predicted frequency of the ith karyotype at each of the t sampled timepoints, given that the fitness was f_i. Assuming that the sampling process from the cell population is binomially distributed, the log likelihood of drawing N_{it} cells with the ith karyotype from the population is given by:

$$LL_i = \sum_t B(N_{it}, u_{it}). \tag{5}$$

Values of f_i and N_0 are optimised using standard gradient descent methods to maximise the likelihood according to E.q. 5.

The second inference method which does not rely on longitudinal growth rate measurements uses:

$$\vec{n_i}(t) = \vec{n_0} exp(\vec{f_i} t) \tag{6}$$

$$\vec{u_i}(t) = \vec{n_i}(t) / \sum \vec{n_i}(t). \tag{7}$$

The likelihood of each $\vec{u_i}(t)$ is again maximised according to e.q. 5. Note that a substantial difference between methods is the ability to optimise the parameters separately for each karyotype when accompanying growth rate data is available. All parameters must be optimised simultaneously when growth rate data is unavailable.

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. First, the fitness gradient between any frequent clones that are Von Neumann neighbours is calculated to form a prior distribution V (taken to be normal). Then for each of the i karyotypes whose fitness was already estimated in the prior step, the fitness of its j neighbours is estimated as follows:

1) Any of the frequent clones that are Von Neumann neighbours of the jth karyotype are treated as possible "sources" that could missegregate to generate the jth karyotype. Let these neighbours be indexed by k.

2) The per-division probability that each of the k neighbours mis-segregates to the jth karyotype is then determined. Since j and k are Von Neumann neighbours only one chromosome can differ between them. Let the copy number of this chromosome be α, and define $\tau = \alpha_k - \alpha_j$. The per-division probability can then be calculated using [23, 18]:

$$P(\tau|\alpha_k) = \sum_{odd/even:z_k=|\tau|}^{\alpha_k} \binom{\alpha_k}{z_k} \beta^{z_k}(1-\beta)^{\alpha_k - z_k} 0.5^{z_k} \binom{z_k}{\frac{z_k-\tau}{2}}, \tag{8}$$

where β is the mis-segregation rate per chromosome copy.

3) The expected number of cells with karyotype j can then be calculated based on the influx from its k Von Neumann neighbours and a supplied fitness value for the jth karyotype, i.e. f_j.

$$\frac{dn_j}{dt} = f_j n_j + \sum_k P(\tau|\alpha_k) f_k n_k \tag{9}$$

4) After solving e.q. 9 for n_j and converting the cell number into frequency u_j, the likelihood is calculated in a manner similar to the frequent clones but also incorporating the prior distribution:

$$LL_j = \sum_t B(N_{jt}, u_{jt}) + \sum_k V(f_j - f_k). \tag{10}$$

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, otherwise all arguments to the Krig function were set as default.

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 $\vec{f_a}$ 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 $\vec{f_b}$. Finally the R^2 value between $\vec{f_a}$ and $\vec{f_b}$ 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 value given by E.q. 8. 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 that maps each point in an n dimensional space to a fitness value. GRFs are defined here by a set of vectors $\vec{v}$. Each vector $\vec{v_i}$ acts as a point source for a wave of wavelength λ, the interference pattern of which creates the GRF. Thus for a karyotype x in the n dimensional space, fitness is evaluated according to:

$$\sum_i sin(\frac{||\vec{v} - \vec{x}||}{\lambda}). \tag{11}$$

Fitness landscape metrics

Ruggedness

Ruggedness has been defined in topographic analysis as a measure of local spatial variability in landscape elevation [24]. Extending that concept to the karyotype fitness landscape, we define a local roughness metric R_i for each karyotype as the mean difference between the fitness of a karyotype f_i and that of its Von Neumann neighbours f_j:

$$R_i = \frac{1}{N}\sum_j^N |f_i - f_j| \tag{12}$$

Local Moran's i

Local Moran's i is a local indicator of spatial association [25] related to spatial autocorrelation of values. It is defined:

$$I_i = N\frac{f_i - \bar{f}}{\sum_i^N (f_i - \bar{f})^2}\sum_j^N w_{ij}(f_j - \bar{f}), \tag{13}$$

where w_ij is a weight matrix encoding relationships between karyotypes on the fitness landscape. Here we use:

$$w_{ij} = \begin{cases} 0 & \text{if } i == j \\ \frac{1}{d_{ij}}, & \text{otherwise} \end{cases} \tag{14}$$

where d_{ij} is the euclidean distance between karyotypes indexed by i and j.

Experimental data

The experimental data used here is prior published data derived from serially passaged cell lines [21]. Briefly, cell lines were serially passaged in vitro (hTERT) or in PDX models (SA535,SA1035,SA609) over a period of several months. Cell populations were each subject to single cell whole genome sequencing several times throughout the duration of the experiments. The data we received were single copy number profiles following quality control, according to the methodology outlined in the previously published article [21]. These copy number profiles were for 4375 bins across the genome of each cell: we further processed these data to obtain copy number profiles at either a per-chromosome or per-chromosome arm resolution. To do this we assigned the copy number per arm/chromosome as the modal copy number of all the bins present on that arm/chromosome.

Results

Clonal evolution on random fitness landscapes

To test our ability 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. Our GRF were generated via summation of multiple circular waves (Fig. 1A). Varying the wavelength parameter allows control of the complexity of the resulting field (Fig. 1B). ABM simulations of cell populations evolving on the GRF landscapes are characterized by expansion and contraction of karyotype-defined subclones (Fig. 1C) as fitter clones are generated and the fitness of the population increases (Fig. 1D).

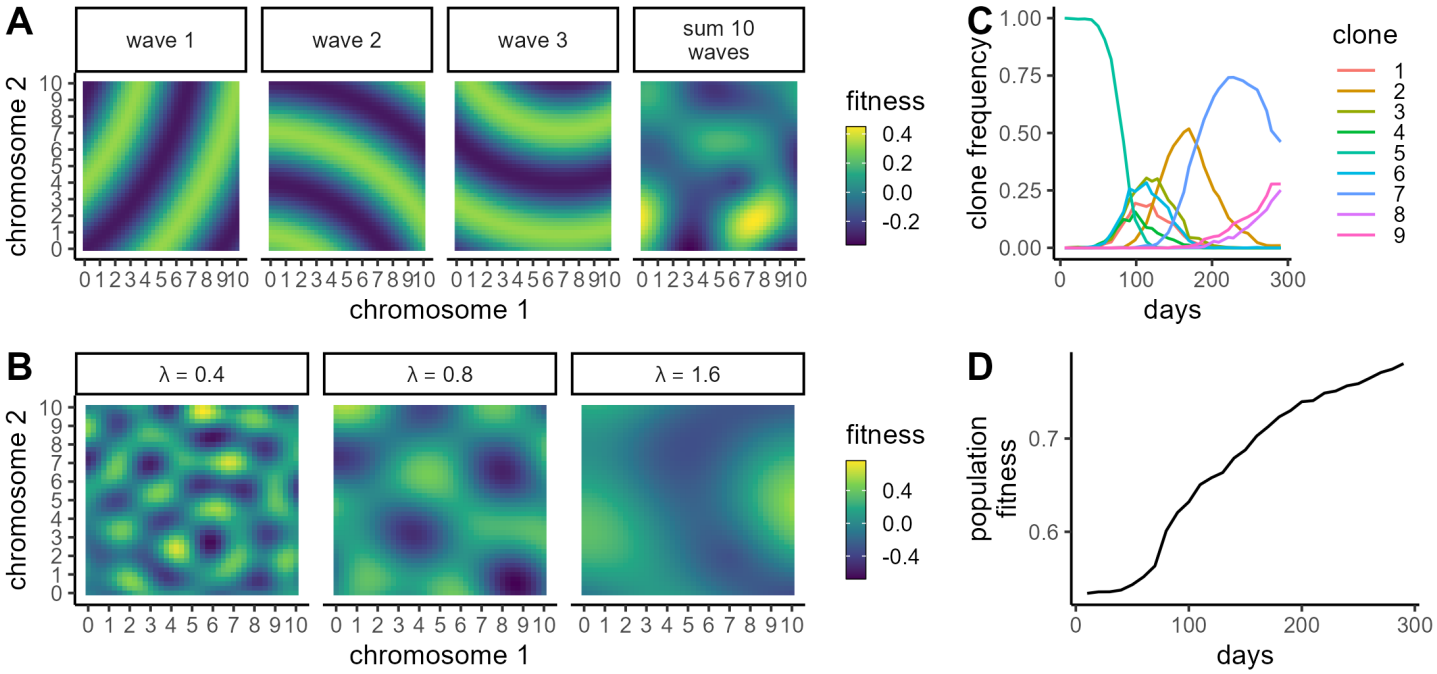

Figure 1: Clonal evolution on random fitness landscapes A) GRF are generated by summing multiple circular waves. The interference patterns generated by the waves result in complex unpredictable landscapes **B**) Increasing the wavelength results in GRF with decreasing complexity. **C**) Example simulation output for ABM model of cell population evolving on a GRF fitness landscape. Each coloured line represents the longitudinal frequency of a different karyotype. **D**) Mean population fitness (i.e. growth rate) of the evolving population shown in C.

We asked whether longitudinal measurements of karyotype frequency (e.g. Fig. 1C) could be used to infer the underlying fitness landscape. To this end we developed an inference pipeline to estimate the topology of the local fitness landscape around the observed karyotypes. The first step of our method is fitness inference for those clones present in sufficient frequency that direct estimation from the experimental data is feasible. We tested two methods, the first requiring longitudinal population growth rate measurements accompanying the longitudinal clone frequency data, the second not requiring these. Both methods were able to estimate the relative fitness of the most common clones from our ABM dataset with a high degree of accuracy (Fig. 2A). The second step involves making an estimate of the fitness of the nearest neighbours of karyotypes estimated in step 1. Here the predictive power depends on the roughness of the fitness landscape (Fig. 2B). Finally we used Kriging interpolation to extrapolate fitness estimates for the entire landscape. Testing the Kriging model on karyotypes at a Manhattan distance of 2 away from any of the frequent clones observed in step 1 yielded increasingly good results for simpler fitness landscapes (Fig. 2C). Note that for 22 chromosomes as tested here, each karyotype has 44 distance 1 neighbours,

and almost 2000 distance 2 neighbours. Thus based on the dynamics of only a few clones, we are able to extrapolate estimates of fitness of of thousands of karyotypes with a high degree of accuracy, provided the underlying fitness landscape is sufficiently smooth. Finally, because in some cases our method did not produce good estimates we asked how one could verify the accuracy of the inferred landscapes. One useful metric results from a leave-one-out cross validation procedure, whereby each clone estimated in step one is omitted from the Kriging training data, then it's fitness as predicted from the trained Kriging model is compared with the fitness predicted in step 1. The R^2 resulting from this validation procedure was strongly correlated with the accuracy of the fitted fitness landscape overall (Fig. 2C).

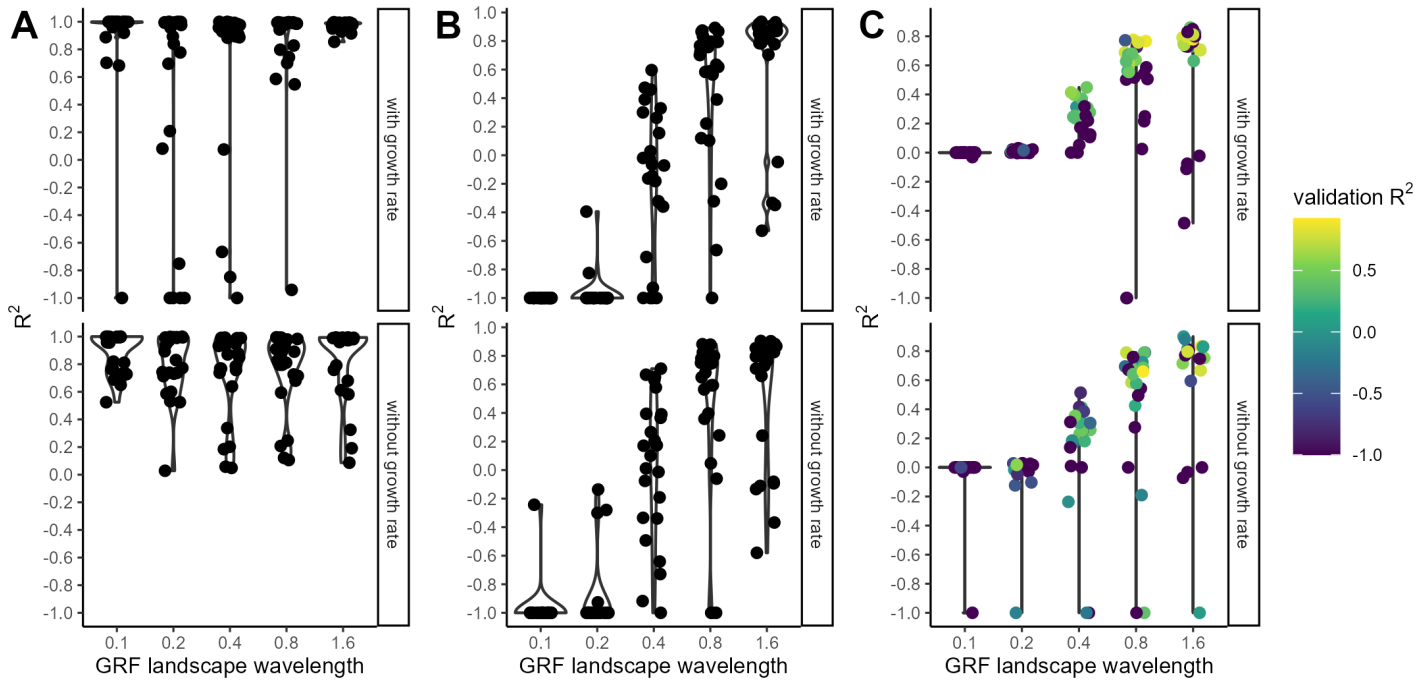

Figure 2: Inference validation on artificial data (**A**) R^2 values for inferred fitness of frequent karyotypes (defined as karyotypes observed at least 5 times across all measured timepoints) for various GRF wavelengths, with or without population growth rate measurements (columns) (**B**) R^2 values for inferred fitness of karyotypes in the Von Neumann neighbourhood of frequent karyotypes, for various GRF wavelengths, with or without population growth rate measurements (columns). (**C**) R^2 values for inferred fitness of karyotypes at a Manhattan distance of two from frequent karyotypes, for various GRF wavelengths, with or without population growth rate measurements (columns). Colour indicates R^2 score on the cross-validation procedure.

karyotype fitness landscapes of breast cancer cell lines

We applied the pipeline to longitudinal data derived from serial passaging experiments [21]. Cross validation results indicated the pipeline performed well on 4/5 datasets (Fig. 3A). To investigate why the cross validation result was so poor on the SA1035 cell line, we visualised the fit to the frequent clone data. In contrast to SA535 which had a high cross validation score and fit the frequent clone data well (Fig. 3B), SA1035 fit the frequent clone data badly (Fig. 3C). This indicates a problem either with the assumptions of the model or the handling of the experimental data. To address this, we repeated the inference procedure and cross validation using arm-level copy numbers instead of chromosome-level. This resulted in a much better fit to the frequent karyotypes (Fig. 3D) and an improved cross validation score of 0.44. Taken together, these results demonstrate the ability of our inference method to predict relative fitness of of karyotypes across several breast cancer cell lines.

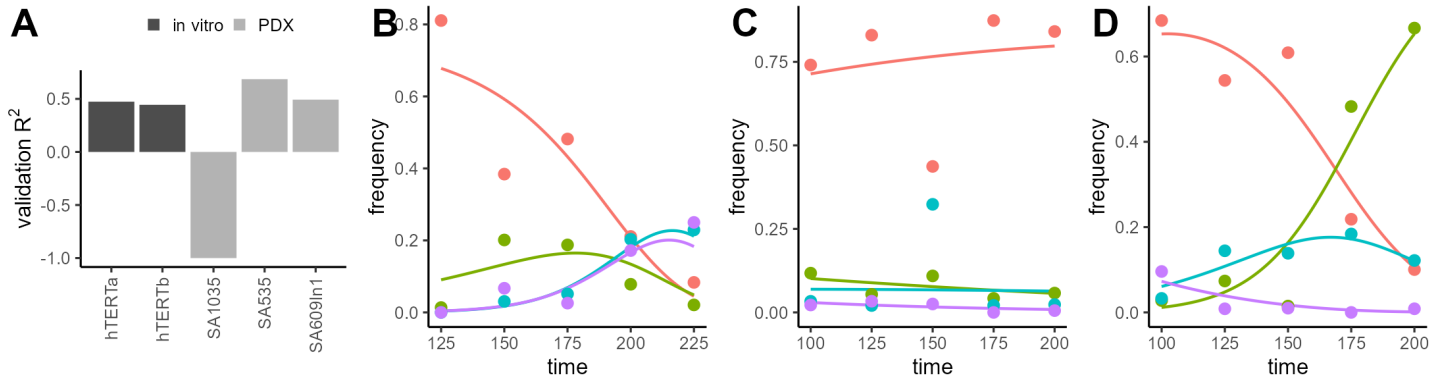

Figure 3: Inference validation with breast cancer cell lines (**A**) Cross validation scores of various cell lines. (**B-D**) show ODE model fits (lines) to frequency of frequent karyotypes (points). Different karyotypes are represented by colour. **B** Fit to SA535 at chromosome-level copy number resolution. **C** Fit to SA1035 at chromosome-level copy number resolution. **D** Fit to SA1035 at arm-level copy number resolution.

Whole genome doubling is a common occurrence among some tumour types, which is thought to benefit cells by conferring robustness to tolerate chromosomal instability. If so then these differences should be reflected in the topology of the fitness landscapes for 2N and 4N cells. We employed two metrics - local Moran's i (Fig. 4A), and "ruggedness" (Fig. 4B) - to identify differences in fitness landscapes inferred from our ABM simulations. Moran's i is a measure of local autocorrelation and as such takes higher values for simpler landscapes, whereas ruggedness measures the magnitude of fitness differences between immediately neighbouring karyotypes and takes higher values for more complex landscapes. Each metric requires specification of karyotypes to serve as focal points for metric evaluation. Frequently observed karyotypes are obvious choices for focal points since our fitness estimates are most robust here, with frequent karyotypes and their Von Neumann neighbours being the next most obvious choice since reduced accuracy of fitness estimates for the Von Neumann neighbours is traded off against the larger number of data points providing additional statistical power for indentifying differences between landscapes. For each choice of metric and set of focal points we asked how many frequent karyotypes we would need to observe before detecting a significant difference between fitness landscapes (Fig. 4 C,D). An advantage of using Moran's i with frequent clone focal points is that since frequent clone fitness was uniformly well estimated we could compare simulations from all five GRF wavelengths (Fig. 4 C). By contrast the ruggedness metric still requires knowledge of Von Neumann neighbours even when evaluating only frequent karyotypes (Fig. 4 D), thus we had to filter out estimates from simulations where the cross validation metric indicated that the latter stages of the pipeline had failed. Overall including Von Neumann neighbours reduced the number of observations expected for significance (Fig. 4 C,D), therefore we applied this analysis to the longitudinal data from the hTERT cell line, which had undergone whole genome doubling during the experiments and thus had both 2N and 4N karyotypes available to analyse. Moran's i indicated that the fitness landscape around 4N cells was significantly smoother than 2N cells (Fig. 4E; p-value $< 2.2e\text{-}16$ - Welch two sample t-test) in line with the hypothesis that WGD ameliorates fitness effects of mis-segregation, although no significant difference was detected based on the ruggedness metric (Fig. 4F; p-value < 0.95).

Discussion

We and others have previously modelled karyotypic evolution via the process of missegregation [17, 18, 19, 20, 23]. Across these modelling efforts, various assumptions have been made regarding fitness associated with specific karyotypes: either that all karyotypes are equally fit [18], that fitness is associated with density of driver or suppressor genes on each chromosome [19, 20], or that fitness is negatively correlated

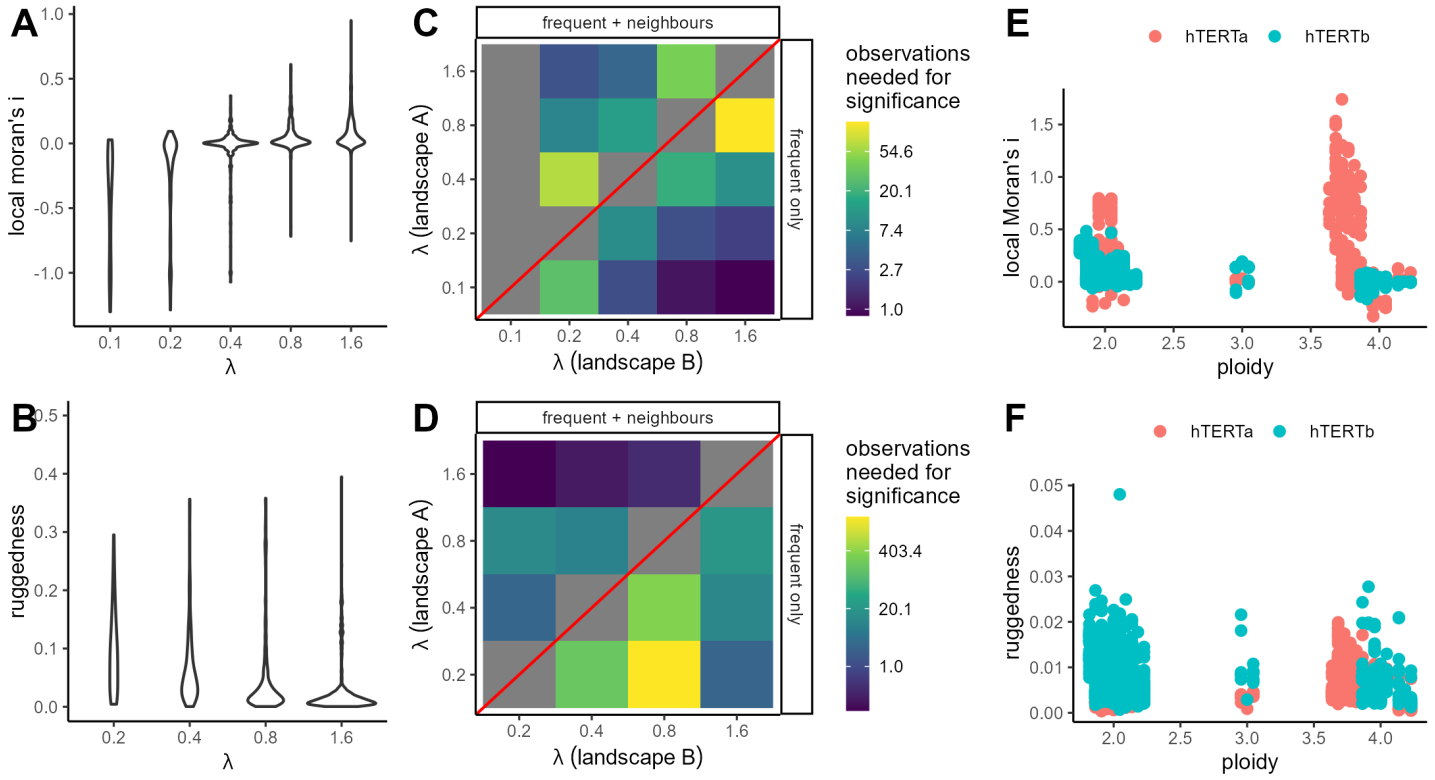

Figure 4: Characterising landscape topology (**A**) local Moran's i evaluated for frequent clones from ABM simulations (**B**) Ruggedness metric evaluated for frequent clones from ABM simulations (**C**) Expected number of frequent karyotypes to detect difference in fitness landscapes using Moran's i. (**D**) Expected number of frequent karyotypes to detect difference in fitness landscapes using ruggedness metric. Heatmap squares (C,D) above the line represent the expected number of frequent karyotypes required for significance if frequent karyotypes and Von Neumann neighbours are used as focal points, squares below the line represent the expected number if only frequent karyotypes are used. (**E**) Moran's i evaluated for frequent karyotypes and their Von Neumann neighbours from hTERT cell line. (**F**) Ruggedness evaluated for frequent karyotypes and their Von Neumann neighbours from hTERT cell line.

with deviation from a euploid state [23]. These attempts to model karyotypic fitness landscapes all ignore context, i.e. that the fitness landscape is sculpted by multiple factors including genetic background, tumour microenvironment, immune interactions and more[3]. This context dependency along with the large number of possible karyotypes makes reconstruction of fitness landscapes a challenging task. The mathematical model presented here offers the flexibility necessary to begin reconstructing adaptive fitness landscapes. Based on the dynamics of just a few subclones, our method is able to extrapolate the fitness of thousands of karyotypes.

One application of our methodology is the study of fitness landscape topology. To that end we adapted metrics from geostatistics to study the fitness landscape of a P53 deficient 184-hTERT diploid breast epithelial cell line[21]. Over several months of passaging these cells developed substantial aneuploidy and underwent WGD, allowing us to compare the topology of the fitness landscape between the diploid and tetraploid states. It is thought that whole genome doubling promotes aneuploidy tolerance [26, 27], a hypothesis which if true predicts a smoother fitness landscape around tetraploid landscapes compared to diploid counterparts. Indeed our topological analysis revealed a significantly larger Moran's i in the tetraploid landscape, in line with the notion that WGD ameliorates the fitness impacts associated SCNAs.

One limitation of our study is the lack of experimental data upon which our model predictions can be validated. In principle our model should be able to predict karyotypic evolution among aneuploid cell populations, however in practise this was prohibited by the paucity of data available to us. Nevertheless

we do present a cross validation procedure that serves as a useful heuristic to evaluate the success of our method. Using this cross validation procedure we were able to demonstrate the ability of our method to predict the fitness of unobserved karyotypes across several cell lines with R^2 values in the range 0.4-0.6. Future work will incorporate data from studies with identical biological replicates [22], which will offer opportunities for improved validation of model predictions.

A second limitation of our model is the assumption that missegregation rate is homogeneous throughout a given fitness landscape. In particular, high ploidy cells may exhibit greater missegregation rates than low ploidy cells which could influence our conclusions surrounding the topology of the fitness landscapes near diploid and tetraploid karyotypes. Previously we have used interferon gamma as a measure of missegregation rate[23]. Incorporating such a metric in future work could help us quantify the extent to which a permissive fitness landscape, rather than propensity to missegregate, explains observed associations between WGD and aneuploidy.

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] John B Lindsay, Daniel R Newman, and Anthony Francioni. Scale-Optimized surface roughness for topographic analysis. Geosci. J., 9(7):322, July 2019.

[25] Luc Anselin. Local indicators of spatial association-LISA. Geogr. Anal., 27(2):93–115, September 2010.

[26] Anastasia Y Kuznetsova, Katarzyna Seget, Giuliana K Moeller, Mirjam S de Pagter, Jeroen A D M de Roos, Milena Dürrbaum, Christian Kuffer, Stefan Müller, Guido J R Zaman, Wigard P Kloosterman, and Zuzana Storchová. Chromosomal instability, tolerance of mitotic errors and multidrug resistance are promoted by tetraploidization in human cells. Cell Cycle, 14(17):2810–2820, 2015.

[27] Kavya Prasad, Mathew Bloomfield, Hagai Levi, Kristina Keuper, Sara V Bernhard, Nicolaas C Baudoin, Gil Leor, Yonatan Eliezer, Maybelline Giam, Cheng Kit Wong, Giulia Rancati, Zuzana Storchová, Daniela Cimini, and Uri Ben-David. Whole-Genome duplication shapes the aneuploidy landscape of human cancers. Cancer Res., 82(9):1736–1752, May 2022.