ALFA-K results

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.

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 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 [24, 25], 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] 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.

[25] 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.