Mathematical modeling of neuroblastoma associates evolutionary patterns with outcomes

Giulio Caravagna

A new study deciphers the origin and evolution of childhood neuroblastoma using genome sequencing data, mathematical models and statistical inference, showing how neuroblastoma evolution is an accurate predictor of outcome.

In 1976, Peter Nowell[1] proposed the clonal evolution model as a basic framework for understanding how cancers evolve over time. Körber et al.[2] now use this model to shed light on the initiation and evolution of neuroblastoma, a rare childhood cancer that develops from immature nerve cells called neuroblasts[3]. These tumors present mostly in children under the age of 5, accounting for about 7% of all childhood malignancies. Patients with high-risk neuroblastoma have survival rates of about 50%, and early diagnosis and accompanying prompt treatment may assist in improving their outcomes.

According to the clonal evolution model, tumors start from a single cell and acquire somatic mutations; clonal expansions are triggered by driver mutations that increase fitness – that is, the ability to produce surviving offspring in the current microenvironment[1]. Therefore, if we track mutations as in species evolution, we can understand the underlying evolutionary process[4]. Körber et al.[2] did this by using deeply sequenced bulk whole-genome data from two neuroblastoma cohorts (n = 100 patients and n = 86 patients) covering all clinical stages of the disease. The main challenge in this type of analysis is that the whole evolutionary process cannot be directly measured; instead some random sample of its outputs is measured.

Neuroblastomas sequenced in this work harbored point mutations and abundant copy-number alterations (CNAs) – gains and losses of genetic material that alter gene expression and contribute to genetic instability[5]. The authors found CNAs to be ubiquitous across the neuroblastoma cells (i.e., clonal), indicative of early acquisition of these changes[4]. They interpreted the mutational profiles of each neuroblastoma using a clever combination of mathematical models and statistical inference. First, they determined the temporal acquisition of segmental copy-number gains in chromosome regions 1q and 17q and of whole-chromosome amplifications of chromosomes 2, 7 and 17. To do this, they use neutral mutations that accumulate linearly with time[6] to build a molecular clock and identify two evolutionary groups of neuroblastomas: one with gains temporally close to the most recent cell from which all the sequenced neuroblasts descend (most recent common ancestor (MRCA)), and another in which a premalignant earlier common ancestor (ECA) acquired most gains before progressing to the malignant MRCA.

In the latter group, gains of regions 9 and 20q were timed after the ECA, indicating further evolution from the premalignant to the malignant state. Körber et al.[2] compared molecular clocks among primary and relapse samples to describe prototypical models with early and late MRCAs. This implies that distinct disease trajectories exist for the growth of resected neuroblastoma. Indeed, 95% of early-MRCA neuroblastomas were explained by aneuploidy, with little contribution from point mutations, whereas 55% of the late-MRCA tumors showed prolonged genetic evolution due to multiple point mutations and other somatic changes. An interesting observation came from MRCAs of metastases resected after initial diagnosis, which were timed to originate around when the tumor started to grow from its MRCA, resembling multifocal tumors such as glioblastoma[7].

Evolutionary trajectories can predict disease outcome[8], and the authors used a discovery cohort to show that MRCA timing translates into an accurate predictor of both event-free survival and overall survival, improving over clinical variables and neuroblastoma-specific

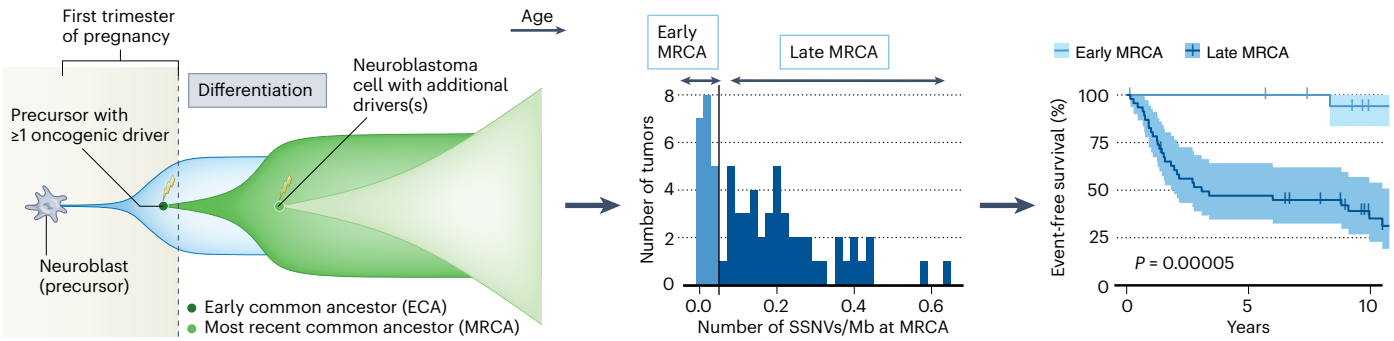

Fig. 1 | Clonal evolution model for neuroblastoma development and evolution. Using sequencing data, Körber et al.[2] classified neuroblastomas on the basis of their evolutionary features, showing that duration of evolution is an accurate predictor of outcome. SSNVs/Mb, somatic single nucleotide variants per megabase.

features such as telomere maintenance and RAS/p53-pathway mutations. The role of aneuploidy emerged clearly, with broader instability in late-MRCA tumors, which were also associated with worst prognosis, suggestive of a link between the duration of evolution and outcome. In this group, a clear role was also found for the acquisition of telomere maintenance mechanisms[9], even if the timing with which *TERT* mutations were acquired remained undetermined.

Finally, the authors gathered all of their information and developed a birth–death process model for the dynamics of neuronal precursors, ECA and MRCA. This population genetics model was parameterized from high-risk late-MRCA neuroblastomas, and intelligently linked to reproduce the incidence of the disease in the human population (~1 in 105 children). Using approximate Bayesian computation[10], the authors inferred that these neuroblastomas acquire ~3.2 mutations per day, with oncogenic events triggering ECA–MRCA transitions once per million cell divisions. These estimates were calibrated against real time to position ECA origin within the first trimester of pregnancy, when the adrenal medulla forms from sympathetic neuroblasts, also suggesting that about 1 in 10 cell divisions of ECAs results in tumor growth of neuroblastomas with acquired telomere maintenance mechanisms, consistent with the extensive cell death observed histopathologically.

Building on Nowell's clonal evolution model[1], the field of tumor evolution has recently emerged as a consolidated approach to understanding, from sequencing data, how cancer cells change over time and evolve in response to various selective pressures, such as genetic mutations, environmental factors and treatment.

Mathematical models based on differential equations and stochastic processes have a longstanding history of successful application in biology, especially in the fields of ecology and epidemiology[11]. Their direct application to cancer evolution is more laborious, however, because high-dimensional, noisy sequencing data are difficult to model, and the underlying evolutionary process is only partially observable and subject to various sources of sampling bias, in time and space. However, identifying statistical patterns from measurements with many dimensions, while resisting the effects of noise and other confounders, is a primary aim of real-world machine learning[12]. For this reason, cancer genomics needs machine learning, and algorithm-based signal deconvolution, feature selection and dimensionality reduction have become common tools of the field[13]. In this framework, combining mathematical encoding of the clonal evolution principles and advanced statistical models could prove invaluable[14]. The new work[2] is an example of this, as it interprets the statistical signal of mutations in light of tumor evolutionary principles, finally using population genetics to quantify clonal expansion patterns in neuroblastomas. In my opinion, the best results of both mathematical modeling and machine learning are reached when the two approaches are intermixed. One possibility, for example, is to derive statistical distributions for quantities of interest using a mathematical model, and then utilize them to instantiate a machine learning algorithm[15]. In general, by linking the two approaches, one can attempt to extract and interpret statistical signals directly within a model that is interpretable, refutable and falsifiable. One of the most rewarding consequences of these model-based paradigms could be the translation of tumor evolution statistics into clinical predictors. By comparing tumors in regard to how they evolve, and not just the mutations they have[8], we might lay the ground for 'predicting' tumor evolutionary dynamics, and stay one step ahead of cancer Fig. 1.

Giulio Caravagna
Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy.
e-mail: gcaravagna@units.it

References

1. Nowell, P. C. *Science* **194**, 23–28 (1976).
2. Körber, V. et al. *Nat. Genet*. https://doi.org/10.1038/s41588-023-01332-y (2023).
3. Brodeur, G. M. *Nat. Rev. Cancer* **3**, 203–216 (2003).
4. Turajlic, S., Sottoriva, A., Graham, T. & Swanton, C. *Nat. Rev. Genet.* **20**, 404–416 (2019).
5. Sansregret, L. & Swanton, C. *Cold Spring Harb. Perspect. Med.* **7**, a028373 (2017).
6. Alexandrov, L. B. et al. *Nature* **500**, 415–421 (2013).
7. Spiteri, I. et al. *Ann. Oncol.* **30**, 456–463 (2019).
8. Caravagna, G. et al. *Nat. Methods* **15**, 707–714 (2018).
9. Gao, J. & Pickett, H. A. *Nat. Rev. Cancer* **22**, 515–532 (2022).
10. Beaumont, M. A., Zhang, W. & Balding, D. J. *Genetics* **162**, 2025–2035 (2002).
11. Otto, S. P. & Day, T. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution* (Princeton Univ. Press, 2011).
12. Bishop, C. M. *Pattern Recognition and Machine Learning* (Springer Nature, 2006).
13. Topol, E. J. *Nat. Med.* **25**, 44–56 (2019).
14. Beerenwinkel, N., Schwarz, R. F., Gerstung, M. & Markowetz, F. *Syst. Biol.* **64**, e1–e25 (2015).
15. Caravagna, G. et al. *Nat. Genet.* **52**, 898–907 (2020).

Acknowledgements

The author acknowledges funding from the Italian Foundation for Cancer Research (AIRC) under MFAG 2020, ID 24913 project, PI Giulio Caravagna.

Competing interests

The author declares no competing interests.