

## Emerging Topics in Cancer Evolution

Mohammed El-Kebir

*Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL 61801*  
*E-mail: melkebir@illinois.edu*

Quaid Morris

*Computational and Systems Biology, Memorial Sloan Kettering Cancer Center, New York, NY*  
*10065*

*E-mail: MorrisQ@mskcc.org*

Layla Oesper

*Department of Computer Science, Carleton College, Northfield, MN 55057*  
*E-mail: loesper@carleton.edu*

S. Cenk Sahinalp

*Cancer Data Science Laboratory, National Cancer Institute, NIH, Bethesda, MD 20892*  
*E-mail: cenk.sahinalp@nih.gov*

Cancer results from an evolutionary process that yields a heterogeneous tumor with distinct subpopulations and varying sets of somatic mutations. This perspective discusses computational methods to infer models of evolutionary processes in cancer that aim to improve our understanding of tumorigenesis and ultimately enhance current clinical practice.

*Keywords:* Intra-tumor heterogeneity; Phylogeny; Evolution; Precision medicine.

### 1. Introduction

Cancer results from an evolutionary process where somatic mutations accumulate in distinct populations of cells.<sup>1</sup> This theory has been corroborated by high-throughput sequencing studies of tumors in the last decade,<sup>2</sup> demonstrating that a tumor is not homogeneous but rather composed of *clones* with varying sets of somatic mutations. This phenomenon of *intra-tumor heterogeneity* is a major cause of relapse and resistance to treatment.<sup>3</sup> Metastasis, i.e. the migration of tumor cells to anatomical locations distinct from the primary tumor, is the primary cause of death in cancer.<sup>4</sup> Thus, the life history of a tumor is the end product of an evolutionary process characterized by cell division, cell mutation and cell migration. The emerging field of *cancer phylogenetics* views cancer through the lens of evolution and employs phylogenetic techniques to reconstruct, analyze and compare life histories of tumors.<sup>5</sup> This perspective will discuss recent advances in (i) computational methods for reconstructing cancer phylogenies from sequencing data, (ii) the identification of common evolutionary patterns and trajec-

ries in cancers, (iii) inference and deconvolution of mutational signatures of cancers, and (iv) adaptive therapies for treating cancer based on evolutionary models.

## 2. Reconstruction of cancer phylogenies from sequencing data

Not only are tumors distinct due to separate evolutionary processes, but each tumor contains multiple *subclones* that each share a common ancestor and set of somatic mutations. *Somatic mutations* are genetic and epigenetic alterations that occur during an individual's lifetime. The former range from *single-nucleotide variants* (SNVs), which alter a single nucleotide, to *copy-number aberrations* (CNAs), which duplicate or delete large genomic regions, as well as other *structural variants* (SVs), which include CNAs in addition to inversion and transposition/translocation of genomic segments. To study cancer evolution, researchers rely on computational methods that infer phylogenies from sequencing data, underpinned by an *evolutionary model* that constrains the types of allowed changes. Initially, *bulk DNA sequencing* was used, yielding short DNA sequences, or *reads*, from millions of cells with diverse genomes. Such data require deconvolution methods,<sup>6–10</sup> which yield many plausible phylogenies.<sup>11</sup> On the other hand, *single-cell DNA sequencing* (scDNA-seq) yields reads from individual cells, and generally does not require deconvolution, but has elevated error rates, requiring specialized methods to simultaneously infer trees and correct errors.<sup>12–15</sup> Recent methods have been proposed to infer tumor phylogenies from bulk and scDNA-seq data of the same tumor.<sup>16,17</sup> Importantly, current models for cancer evolution and corresponding phylogeny inference methods focus either only on SNVs<sup>6–10,12–15</sup> or on CNAs,<sup>18–20</sup> but not yet both in full generality. We anticipate that future research will focus on evolutionary models and computational methods that capture the interplay and evolutionary history of genomic SNVs, CNAs, SVs, and epigenomic alterations. Moreover, we expect that techniques that have been developed for tumor phylogeny inference will be applicable to lineage tracing.<sup>21</sup>

## 3. Identification of common evolutionary patterns and trajectories in cancers

While each cancer results from a different instantiation of an evolutionary process, the complexity of all cancers can be reduced to a small number of principles, so called *hallmarks of cancer*,<sup>22</sup> corresponding to different biological pathways. Nevertheless, there is an exponential number of combinations of mutations in which these traits can be acquired. Cancer subtypes divide primary cancers into smaller groups based on morphological and/or molecular features. There is increasing evidence that more granular cancer subtypes based on common evolutionary patterns better stratify patients in terms of survival and therapy response.<sup>23</sup> Recently, several methods have been introduced to infer evolutionary trajectories from cancer phylogeny cohorts.<sup>23–25</sup> Briefly, REVOLVER<sup>23</sup> is a machine-learning method that uses transfer learning to identify hidden evolutionary patterns in cancer cohorts. CONETT<sup>26</sup> is a combinatorial optimization method to detect recurrent tumor evolution trajectories using a consensus tree approach. Similarly, RECAP<sup>25</sup> is a computational method that resolves ambiguities in cancer sequencing data and detects subtypes of evolutionary trajectories by simultaneously (i) identifying a single tree among the solution space of trees for each patient, (ii) assigning patients to clusters and (iii) inferring a consensus tree for each cluster of patients. While current

approaches use mutations in matching genes across patients as a key signal, we anticipate further developments that will enable the inference of evolutionary trajectories composed of pathways or sets of mutated genes that are mutually interchangeable.

#### 4. Inference and deconvolution of mutational signatures of cancers

Exposures to endogenous (e.g. DNA mismatch repair deficiency) and exogenous (e.g. tobacco smoke or UV light) factors lead to characteristic patterns of SNVs, or *mutational signatures*.<sup>27</sup> That is, the set of somatic mutations in a tumor is the result of varying exposures to distinct mutational processes that can be represented by “signatures” that delineate the types of mutations associated with a given process. While initial work has focused on identifying the additive effects of mutational signatures using non-negative matrix factorization,<sup>27</sup> later work has focused on identifying non-additive secondary effects,<sup>28</sup> including interactions between DNA damage events and deficiencies in repair mechanisms. Recognizing that exposures to mutational signatures may vary during the evolution of a tumor, one line of research has focused on identifying the interplay between mutational processes and cancer evolution. TrackSig<sup>29</sup> and TrackSigFreq<sup>30</sup> aim to construct evolutionary trajectories of signature exposures, defined by changepoints in exposure ordered by pseudotime. In a similar vein, PhySigs<sup>31</sup> seeks to infer the clonal dynamics of mutational signatures in a tumor by identifying exposure shifts along the edges of a given cancer phylogeny. There are important therapeutic opportunities associated with studying mutational signatures in an evolutionary context. For example, mutational signatures can be used as a biomarker for perturbed DNA damage repair (DDR),<sup>32,33</sup> for which promising cancer therapies such as PARP inhibitors exist. By taking the clonality of DDR mutation signature exposures into account, one can identify cases where only a fraction of cells have a DDR deficiency and therefore the patient is unlikely to have a complete response to a therapy targeting that deficiency.

#### 5. Adaptive therapies for treating cancer based on evolutionary models

While many targeted cancer therapies have been introduced in the past two decades,<sup>34</sup> targeted therapies often have a response rate of less than 50% in solid tumors.<sup>35</sup> Resistance to therapy can be classified as either primary, where patients show no response to treatment, or secondary, where patients initially respond but later develop resistance. A prominent example of the latter is the treatment of melanoma with vemurafenib that targets the *BRAF* V600E mutation: despite an initial dramatic response, most patients eventually relapse with drug-resistant, fatal disease.<sup>36</sup> There is growing evidence that intra-tumor heterogeneity is a major driver of resistance to therapy.<sup>37</sup> To understand why, it is important to distinguish *clonal* somatic mutations, which are present in every tumor cell, from *subclonal* somatic mutations, which are present in only a subset of tumor cells. In chronic myeloid leukemia,<sup>38</sup> treatment that targets subclonal driver mutations has been associated to resistance and recurrence. Moreover, therapy itself is an evolutionary bottleneck and may lead to previously low-abundance, resistant clones becoming dominant. Therefore, it is important to take the clonal architecture of a tumor into account when designing treatment plans. Future research directions in this area include identifying effective combination therapies, where multiple targeted drugs are

used simultaneously, as well as adaptive containment strategies where targeted drugs are used alternately in order to constrain evolution in tumors with multiple clones.<sup>34</sup> Both directions will depend on accurate computational methods for tumor phylogeny inference, as well as high-throughput sequencing and monitoring strategies such as liquid biopsies.

## 6. Discussion

This perspective discussed emerging topics in cancer evolution, focusing on recent advances in tumor phylogeny inference, identification of evolutionary trajectories, mutational signatures as well as cancer therapy in light of evolution. While the field has mainly focused on reconstructing a tumor's evolutionary history in terms of genetic alterations, it is important to also take the evolution of epigenetic mutations into account. Continued interdisciplinary collaboration between cancer biologists, computational method developers and clinicians on these topics will be essential towards gaining a more thorough understanding of tumor emergence, proliferation and metastatic expansion and response to therapy. Ultimately this will be an essential step towards achieving the goal of improved clinical treatments.

**Acknowledgments.** M.E-K. was supported by the National Science Foundation (grants: CCF-1850502 and CCF-2046488) as well as funding from the Cancer Center at Illinois. Q.M. was supported by the National Institutes of Health (grants: P30 CA 008748). L.O. was supported by the National Science Foundation (grants: IIS-1657380 and IIS-2046011). S.C.S. was supported by the intramural research program of the National Cancer Institute, National Institutes of Health.

## References

1. P. C. Nowell, The clonal evolution of tumor cell populations, *Science* **194**, 23 (1976).
2. P. J. Campbell *et al.*, Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing, *Proceedings of the National Academy of Sciences* **105**, 13081 (2008).
3. R. Fisher, L. Pusztai and C. Swanton, Cancer heterogeneity: implications for targeted therapeutics, *British journal of cancer* **108**, 479 (2013).
4. G. P. Gupta and J. Massagué, Cancer metastasis: building a framework, *Cell* **127**, 679 (2006).
5. R. Schwartz and A. A. Schäffer, The evolution of tumour phylogenetics: principles and practice, *Nature Reviews Genetics* **18**, p. 213 (2017).
6. F. Strino, F. Parisi, M. Micsinai and Y. Kluger, Trap: a tree approach for fingerprinting subclonal tumor composition, *Nucleic Acids Res* **41**, p. e165 (2013).
7. S. Malikic, A. W. McPherson, N. Donmez and C. S. Sahinalp, Clonality inference in multiple tumor samples using phylogeny., *Bioinformatics* **31**, 1349 (2015).
8. V. Popic, R. Salari, I. Hajirasouliha, D. Kashef-Haghighi, R. B. West and S. Batzoglou, Fast and scalable inference of multi-sample cancer lineages, *Genome biology* **16**, p. 91 (2015).
9. A. G. Deshwar *et al.*, PhyloWGS: Reconstructing subclonal composition and evolution from whole-genome sequencing of tumors, *Genome Biology* **16**, p. 35 (2015).
10. M. El-Kebir, L. Oesper, H. Acheson-Field and B. J. Raphael, Reconstruction of clonal trees and tumor composition from multi-sample sequencing data., *Bioinformatics* **31**, i62 (2015).
11. Y. Qi, D. Pradhan and M. El-Kebir, Implications of non-uniqueness in phylogenetic deconvolution of bulk DNA samples of tumors, *Algorithms for Molecular Biology* **14**, p. 19 (2019).
12. M. El-Kebir, SPhyR: tumor phylogeny estimation from single-cell sequencing data under loss and error, *Bioinformatics* **34**, i671 (2018).

13. K. Jahn *et al.*, Tree inference for single-cell data, *Genome biology* **17**, p. 86 (2016).
14. E. M. Ross and F. Markowetz, OncoNEM: inferring tumor evolution from single-cell sequencing data., *Genome biology* **17**, p. 69 (2016).
15. H. Zafar *et al.*, SiCloneFit: Bayesian inference of population structure, genotype, and phylogeny of tumor clones from single-cell genome sequencing data, *Genome Research* **29**, 1847 (2019).
16. S. Malikic *et al.*, Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data, *Nature communications* **10**, 1 (2019).
17. S. Malikic *et al.*, PhISCS: a combinatorial approach for subperfect tumor phylogeny reconstruction via integrative use of single-cell and bulk sequencing data, *Genome Research* **29** (2019).
18. M. El-Kebir *et al.*, Complexity and algorithms for copy-number evolution problems, *Algorithms for Molecular Biology* **12**, p. 13 (2017).
19. S. Zaccaria, M. El-Kebir, G. W. Klau and B. J. Raphael, Phylogenetic Copy-Number Factorization of Multiple Tumor Samples., *Journal of Computational Biology* **25**, 689 (2018).
20. S. Zaccaria and B. J. Raphael, Characterizing allele- and haplotype-specific copy numbers in single cells with CHISEL, *Nature Biotechnology* (2020).
21. W. Gong *et al.*, Benchmarked approaches for reconstruction of in vitro cell lineages and in silico models of *C. elegans* and *M. musculus* developmental trees, *Cell Systems* **12**, 810 (2021).
22. D. Hanahan and R. A. Weinberg, The Hallmarks of Cancer, *Cell* **100**, 57 (2000).
23. G. Caravagna *et al.*, Detecting repeated cancer evolution from multiregion tumor sequencing data, *Nature methods* **15**, 707 (2018).
24. S. Khakabimamaghani, S. Malikic, J. Tang, D. Ding, R. Morin, L. Chindelevitch and M. Ester, Collaborative intra-tumor heterogeneity detection, *Bioinformatics* **35**, i379 (2019).
25. S. Christensen, J. Kim, N. Chia, O. Koyejo and M. El-Kebir, Detecting evolutionary patterns of cancers using consensus trees, *Bioinformatics* **36**, i684 (2020).
26. E. Hodzic, R. Shrestha, S. Malikic, C. C. Collins, K. Litchfield, S. Turajlic and S. C. Sahinalp, Identification of conserved evolutionary trajectories in tumors, *Bioinformatics* **36**, i427 (2020).
27. L. B. Alexandrov *et al.*, Signatures of mutational processes in human cancer, *Nature* **500** (2013).
28. D. Wojtowicz *et al.*, Repairsig: Deconvolution of dna damage and repair contributions to the mutational landscape of cancer, *Cell Systems* (2021).
29. PCAWG Evolution and Heterogeneity *et al.*, Reconstructing evolutionary trajectories of mutation signature activities in cancer using TrackSig, *Nature Communications* **11**, p. 731 (2020).
30. C. F. Harrigan *et al.*, TrackSigFreq: subclonal reconstructions based on mutation signatures and allele frequencies, in *Biocomputing 2020*, (Kohala Coast, Hawaii, USA, 2019).
31. S. Christensen *et al.*, PhySigs: Phylogenetic Inference of Mutational Signature Dynamics, in *Pacific Symposium on Biocomputing 2020*, (Kohala Coast, Hawaii, USA, 2019).
32. H. Davies, D. Glodzik, S. Morganella, L. R. Yates *et al.*, HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures, *Nature Medicine* **23**, 517 (2017).
33. A. V. Hoeck, N. H. Tjoonk, R. v. Bostel and E. Cuppen, Portrait of a cancer: mutational signature analyses for cancer diagnostics, *BMC Cancer* **19**, p. 457 (2019).
34. K. L. Pogrebniak and C. Curtis, Harnessing Tumor Evolution to Circumvent Resistance, *Trends in Genetics* **34**, 639 (2018).
35. C. Holohan, S. Van Schaeybroeck, D. B. Longley and P. G. Johnston, Cancer drug resistance: an evolving paradigm, *Nature Reviews Cancer* **13**, 714 (2013).
36. M. Das Thakur *et al.*, Modelling vemurafenib resistance in melanoma reveals a strategy to forestall drug resistance, *Nature* **494**, 251 (2013).
37. A. Marusyk, V. Almendro and K. Polyak, Intra-tumour heterogeneity: a looking glass for cancer?, *Nature Reviews Cancer* **12**, 323 (2012).
38. M. Gerlinger and C. Swanton, How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine, *British Journal of Cancer* **103**, 1139 (2010).