Pacific Symposium on Biocomputing 25:219-225(2020)

MUTATIONAL SIGNATURES: ETIOLOGY, PROPERTIES, AND ROLE IN CANCER

Mark D.M. Leiserson[1,*]
Department of Computer Science, University of Maryland
8125 Paint Branch Dr, College Park, MD 20740 USA
Email: mdml@cs.umd.edu
** Corresponding author*

Teresa M. Przytycka[1]
National Center of Biotechnology Information, National Library of Medicine, NIH
Bethesda, MD 20894 USA
Email: przytyck@ncbi.nlm.nih.gov

Roded Sharan[1]
Blavatnik School of Computer Science, Tel Aviv University
Tel Aviv, 69978 Israel
Email: roded@tau.ac.il

1. Introduction

Mutations are the ultimate source of genetic diversity within and across species, and the main drivers of evolution and disease. Mutations occur due to a combination of errors in DNA replication or damage that is unrepaired. Understanding the rate and mechanisms by which they occur is of great importance for studies of medical genetics and evolutionary biology. Recent advances in sequencing have now made it possible to generate large datasets to study mutational processes across individuals and diseases.

A chief application of mutation studies is in deciphering somatically acquired changes in the DNA of cancer cells. In addition to the mutations that confer a growth advantage, the so-called cancer drivers, cancer genomes accumulate a large number of somatic mutations resulting from normal DNA damage and repair processes as well as mutations triggered by carcinogenic exposures or cancer related aberrations of DNA maintenance machinery. These mutagenic processes often

[1] All authors contributed equally

produce characteristic mutational patterns called *Mutational Signatures*. However, cancer mutational signatures can be the consequence of interacting factors, such as carcinogenic exposures and potential deficiencies of the DNA repair mechanism. Further, mutational process activity changes over time, leading to unique patterns of mutations. While cancer biologists have known for years that mutational processes such as UV radiation produce characteristic mutations (e.g. [1]), the large-scale mining of cancer genomes for signatures of mutational processes has only begun recently.

The key drivers of the study of mutational signatures in cancer have been access to data and computational approaches. In the past decade, consortia such as The Cancer Genome Atlas and the International Cancer Genome Consortium have produced datasets of millions of somatic mutations from more than 35 cancer types. For the first time, these datasets enabled researchers to search for patterns of somatic mutations across thousands of tumors.

Seminal studies of mutational signatures focused on identifying the signatures and characterizing their etiology. Nik-Zainal et al. [2] and Alexandrov et al. [3] were the first to model a tumor's mutations as a mixture of hidden mutational signatures. Alexandrov et al. [4] applied such a model to a pan-cancer dataset, creating the first survey of mutational signatures in cancer. Recently, members of the Pan Cancer Analysis of Whole Genomes (PCAWG) Network have taken another large step forward in surveying mutational signatures combining cancer exome sequencing data with the whole genome sequencing data from the International Cancer Genome Consortium [5].

More recently, mutational signatures research has expanded to include applications towards personalized therapy and methods for understanding tumorigenesis. Researchers have found that mutational signatures may serve as prognostic [6] or predictive biomarkers [7]. In the latter case, the potential power of modeling mutational signatures comes from being able to identify patients with DNA damage/repair deficiencies, even in the absence of known driver mutations. Researchers have also begun to develop methods that use insights from mutational signatures to identify driver mutations [8-9] -- e.g. by modeling the “mutability” of different genes based on the signatures active in a given cancer type -- and better model cancer evolution [10-11] -- e.g. by characterizing the differences in signature activity among subclones.

2. New trends in mutational signatures research

This session highlights a keynote talk given by Dr. Ludmil Alexandrov, one of the leaders in this new burgeoning area of research. Dr. Alexandrov is a lead author on many of the seminal papers in the field, including the first pan-cancer survey of mutational signatures [4]. Dr. Alexandrov has continued to lead studies characterizing the mutational signatures of specific mutagens (e.g. tobacco smoke [12]) and one of the first studies exploring the use of mutational signatures for personalized therapy [13]. The original research papers accepted to the session span a broad array of active mutational signatures research topics.

2.1. *Mutational signatures and cancer evolution*

Two of the accepted papers examine the interplay between mutational processes and cancer evolution. Knowing *when* mutational processes were active is crucial context for fully understanding a tumor's evolutionary history. An understanding of cancer evolution can also improve estimates of signature exposures, as cells in a tumor do not share all mutations. Harrigan et al. introduce TrackSigFreq for constructing evolutionary "trajectories" of signature exposures, which consist of changepoints in exposure ordered by pseudotime. TrackSigFreq improves upon the authors' earlier method, TrackSig [10], by using variant allele frequencies to distinguish between subclones with the same signature exposure. Harrigan et al. evaluate TrackSigFreq on simulated data, and even with the advantage of distinguishing between subclones with the same exposures, they find TrackSigFreq provides at least as accurate trajectory reconstructions compared to TrackSig, depending on the simulated scenario. Christensen et al. introduce PhySigs, an algorithm for inferring exposures for subtrees for cancer phylogenies given as input. In multi-sample sequencing data from lung cancer patients, Christensen et al. find that 22% (20/91) of the cancer phylogenies have evidence of exposure "shifts", and show how understanding exposure shifts can improve understanding of driver mutations and be used to distinguish among alternative proposed phylogenies for a tumor.

2.2. *Distinguishing signatures of DNA damage and repair*

Cancer mutational signatures can be seen as the end-effect of several interplaying factors: the nature of DNA damage including specific properties of the lesion (e.g. DNA break, covalent modification, bulky adducts), the properties and distribution of sites that are vulnerable to the damage and the properties of repair mechanism responsible for repairing the primary damage including potential cancer-related deficiencies of this mechanism. In addition, DNA stress, double strand breaks (DSB) or other cancer related changes create mutation opportunities that could lead

to unique patterns of DNA primary damage. Thus, there are many factors that jointly contribute to the final catalog of patient's mutation and their contribution is not always additive [14]. However, the leading methods for signature discovery, such as NMF, cannot model non-additive effects. In order to disambiguate the atomic components that contribute to the final signature, new methods are required. As a step in this direction, the paper by Wojtowicz et al. presented in this session introduces a simple yet powerful descriptor of mutational signatures, called DNA Repair Footprint (RePrint). They show that it can capture common properties of repair mechanisms contributing to diverse signatures. These results demonstrate that large mutational datasets contain information that allows us to go beyond uncovering mutational signatures and begin the discovery of elementary components of the processes that generated them.

2.3. *The distribution of somatic mutations across genome*

Mutational processes do not act uniformly on the genome, but depend on many factors such as the genome's chromatin structure, epigenetics, and genomics region (e.g. genic versus intergenic, promoter, replication time, etc). These biases are typically specific to a mutagenic process [15]. Thus, similarly as in the case of mutational signatures, the distribution of mutations is a result of the interplay between DNA damage and repair, but also the genomic context. For example APOBEC enzyme introduces mutations on single-stranded DNA (ssDNA). During replication, the lagging strand is particularly susceptible to APOBEC mutations [16]. Similarly, non-canonical DNA structures, like cruciforms expose ss-DNA making corresponding regions prone to APOBEC mutagenicity [17]. As another example, mismatch repair acts more efficiently in early-replicating areas while the base excision repair pathway has lower efficiency at nucleosomes than linkers. This heterogeneity raises the question of whether the distribution of mutations alone is predictive of cancer type. Over the past several years, multiple studies have shown that it is [18-19]. The study by Jiao et al. [19] shows that it is possible to use machine learning to classify primary and metastatic cancers based on patterns of passenger mutations. In addition, the non-uniform and signature dependent distribution of mutations suggest that the different genomic locations can provide different information about the relation of mutations in this region and tumor type. In the paper by Young at al. presented in this session, the authors develop an information theory driven, dynamic programming algorithm for associating regional mutation density with cancer type. They show that their algorithm provides an efficient method for finding a partition of the genome into regions with mutation density strongly associated with a phenotype, and can thus be used to predict cancer subtypes.

2.4. *Mutational signatures can reveal sources of cancer drivers*

Mutational signatures help to identify mutational processes acting on the genome. As the research outlined above indicates, some regions are more vulnerable to disruption by some mutagenic precesses than others. Recent research demonstrates clear links between mutational processes and driver mutations [9, 20]. For example, a famous cancer driver, the PIK3C gene, has a mutational hotspot consistent with APOBEC activity [20-21], and Fredriksson et al. [22] found that many recurrent promoter mutations in melanoma likely occur due to a particular vulnerability to a UV radiation mutational process [22]. Given the genome-wide activity of mutational signatures, it is also expected that mutagenic processes can also have a signature-specific effect on binding sites and microRNA activity. In a paper by Stamoulakatou et al. presented in this session, the authors investigate how mutational signatures may disrupt microRNA activity. Mutational processes have the potential to alter either microRNA seed regions or their targets, in turn leading to gene dysregulation. The authors introduce a probabilistic framework to evaluate the "impact" of a given signature on each microRNA and its target sequences, both in general and in a particular tumor. To validate their method, the authors ranked microRNAs for different cancers based on their predicted impact, showing that top ranking microRNAs tend to be associated with the cancer type for which they were predicted. The authors further show a positive correlation between the predicted impact on a microRNA and the number of somatic mutations in its response elements.

Acknowledgments

We thank all the researchers who submitted their work and the reviewers who provided insightful reviews of submitted papers. We thank the organizers of PSB 2020 and Tiffany Murray. We also thank the PSB chairs for handling the review process for papers co-authored by session chairs. TMP is supported by Intramural Research Program of the National Library of Medicine, NIH.

References

1. Miller, J. (1985). Mutagenic specificity of ultraviolet light. *Journal of Molecular Biology* **182**(1), 45-65. https://dx.doi.org/10.1016/0022-2836(85)90026-9
2. Nik-Zainal, S., Alexandrov, L., Wedge, D., Van Loo, P., Greenman, C., et al. (2012). Mutational Processes Molding the Genomes of 21 Breast Cancers. *Cell* **149**(5), 979-993. https://dx.doi.org/10.1016/j.cell.2012.04.024
3. Alexandrov, L., Nik-Zainal, S., Wedge, D., Campbell, P., Stratton, M. (2013). Deciphering Signatures of Mutational Processes Operative in Human Cancer. *Cell Reports* **3**(1), 246-259. https://dx.doi.org/10.1016/j.celrep.2012.12.008

4. Alexandrov, L., Nik-Zainal, S., Wedge, D., Aparicio, S., Behjati, S., et al. (2013). Signatures of mutational processes in human cancer. *Nature* **500**(7463), 415-421. https://dx.doi.org/10.1038/nature12477
5. Alexandrov, L., Kim, J., Haradhvala, N., Huang, M., Ng, A., et al. (2018). The Repertoire of Mutational Signatures in Human Cancer. *bioRxiv*. https://dx.doi.org/10.1101/322859
6. Trucco, L., Mundra, P., Hogan, K., Garcia-Martinez, P., Viros, A., et al. (2018). Ultraviolet radiation–induced DNA damage is prognostic for outcome in melanoma. *Nature Medicine* **25**, 221–224. https://dx.doi.org/10.1038/s41591-018-0265-6
7. Davies, H., Glodzik, D., Morganella, S., Yates, L., Staaf, J., et al. (2017). HRDetect is a predictor of BRCA1 and BRCA2 deficiency based on mutational signatures. *Nature Medicine* **23**(4), 517-525. https://dx.doi.org/10.1038/nm.4292
8. Brown, A., Li, M., Goncearenco, A., Panchenko, A. (2019). Finding driver mutations in cancer: Elucidating the role of background mutational processes. *PLOS Computational Biology* **15**(4), e1006981. https://dx.doi.org/10.1371/journal.pcbi.1006981
9. Temko, D., Tomlinson, I., Severini, S., Schuster-Böckler, B., Graham, T. (2018). The effects of mutational processes and selection on driver mutations across cancer types. *Nature Communications* **9**(1), 1857. https://dx.doi.org/10.1038/s41467-018-04208-6
10. Rubanova, Y., Shi, R., Li, R., Wintersinger, J., Sahin, N., Deshwar, A., Morris, Q., PCAWG Evolution and Heterogeneity Working Group, PCAWG network. (2018). TrackSig: reconstructing evolutionary trajectories of mutations in cancer. *bioRxiv.* https://dx.doi.org/10.1101/260471
11. Dentro, S., Leshchiner, I., Haase, K., Tarabichi, M., Wintersinger, J., et al. (2018). Portraits of genetic intra-tumour heterogeneity and subclonal selection across cancer types. *bioRxiv.* https://dx.doi.org/10.1101/312041
12. Alexandrov, L., Ju, Y., Haase, K., Loo, P., Martincorena, I., et al. (2016). Mutational signatures associated with tobacco smoking in human cancer. *Science* **354**(6312), 618-622. https://dx.doi.org/10.1126/science.aag0299
13. Alexandrov, L., Nik-Zainal, S., Siu, H., Leung, S., Stratton, M. (2015). A mutational signature in gastric cancer suggests therapeutic strategies. *Nature Communications* **6**, 8683. https://dx.doi.org/10.1038/ncomms9683
14. Volkova, N., Meier, B., Huici, V., Bertolini, S., Rosado, S., et al. (2019). Mutational signatures are jointly shaped by DNA damage and repair. *bioRxiv*. https://dx.doi.org/10.1101/686295
15. Gonzalez-Perez, A., Sabarinathan, R., Lopez-Bigas, N. (2019). Local Determinants of the Mutational Landscape of the Human Genome. *Cell* **177**(1), 101-114. https://dx.doi.org/10.1016/j.cell.2019.02.051
16. Seplyarskiy, V., Soldatov, R., Popadin, K., Antonarakis, S., Bazykin, G., Nikolaev, S. (2016). APOBEC-induced mutations in human cancers are strongly enriched on the lagging DNA strand during replication. *Genome Research* **26**(2), 174-182. https://dx.doi.org/10.1101/gr.197046.115
17. Zou, X., Morganella, S., Glodzik, D., Davies, H., Li, Y., et al. (2017). Short inverted repeats contribute to localized mutability in human somatic cells. *Nucleic Acids Research* **45**(19), 11213–11221. https://dx.doi.org/10.1093/nar/gkx731

18. Polak, P., Karlić, R., Koren, A., Thurman, R., Sandstrom, R., et al. (2015). Cell-of-origin chromatin organization shapes the mutational landscape of cancer. *Nature* **518**(7539), 360-364. https://dx.doi.org/10.1038/nature14221
19. Jiao, W., Atwal, G., Polak, P., Karlic, R., Cuppen, E., et al. (2019). A deep learning system can accurately classify primary and metastatic cancers based on patterns of passenger mutations. *bioRxiv*. https://dx.doi.org/10.1101/214494
20. Poulos, R., Wong, Y., Ryan, R., Pang, H., Wong, J. (2018). Analysis of 7,815 cancer exomes reveals associations between mutational processes and somatic driver mutations. *PLOS Genetics* **14**(11), e1007779. https://dx.doi.org/10.1371/journal.pgen.1007779
21. Henderson, S., Chakravarthy, A., Su, X., Boshoff, C., Fenton, T. (2014). APOBEC-Mediated Cytosine Deamination Links PIK3CA Helical Domain Mutations to Human Papillomavirus-Driven Tumor Development. *Cell Reports* **7**(6), 1833-1841. https://dx.doi.org/10.1016/j.celrep.2014.05.012
22. Fredriksson, N., Elliott, K., Filges, S., Eynden, J., Ståhlberg, A., Larsson, E. (2017). Recurrent promoter mutations in melanoma are defined by an extended context-specific mutational signature. *PLOS Genetics* **13**(5), e1006773. https://dx.doi.org/10.1371/journal.pgen.1006773