



## Somatic mutation in noncoding regions: The sound of silence

Hua Tan

School of Biomedical Informatics, The University of Texas Health Science Center at Houston, Houston, TX 77030, USA



### ARTICLE INFO

#### Article History:

Received 6 October 2020  
Accepted 6 October 2020  
Available online xxx

Cancer develops primarily because of the sequential somatic alterations accumulating in the genome [1]. The identification of cancer drivers has traditionally focused on the protein-coding regions, which comprise less than 2% of the human genome, leaving the overwhelming noncoding regions largely unexplored. This is due to the currently high costs of whole-genome compared to exome-wide sequencing and poor understanding of the implications of noncoding mutations. While noncoding regions do not encode protein themselves, the noncoding elements can regulate the transcription and translation of protein-coding genes via various mechanisms. In particular, the cis-regulatory elements (e.g., promoters, enhancers, silencers and insulators) regulate the transcription of their target genes by controlling transcription regulatory factors binding, while noncoding RNAs (e.g., microRNAs (miRNAs) and long noncoding RNAs (lncRNAs)) further calibrate gene expression in a post-transcriptional manner [2].

Recently, there is an increasing consensus about the importance of noncoding mutations in human cancer, and whole-genome analysis has revealed significant driver mutations in noncoding regulatory elements including promoters, enhancers, and 5'- and 3'-untranslated regions (5'- and 3'-UTRs) [3]. In this article of *EBioMedicine*, Urbanek-Trzeciak et al. [4] systematically characterised the somatic mutations in miRNA genes (defined as sequences coding for the most crucial part of miRNA precursors) based on whole-exome sequencing data of 33 cancer types collected in The Cancer Genome Atlas (TCGA). They further demonstrated the biological and clinical significance of those overmutated miRNA genes via functional analysis, and discussed the ramifications of these miRNA mutations in oncogenesis. While further investigation (both *in silico* and in wet-lab, and favorably on larger scale of cancer genomics data) is warranted for elucidating the consequences and mechanisms of action of the miRNA mutations in cancer etiology/pathology, this study represents a good starting point for potential follow-up explorations.

The biological consequence of a somatic mutation in a miRNA gene is multifaceted, depending on the specific mutant site. Intuitively, a miRNA mutation might affect its binding sites/affinity on

target 3'-UTRs if mutant in seed region, or influence its own biogenesis if mutant in DROSHA/DICER1 cleavage sites [5]. Considering their functional importance, these mentioned regions are expected to be positively selected in oncogenic process. Surprisingly, this commented work reveals that mutations occur in all regions of the miRNA gene with little localization preference. This seemingly counterintuitive phenomenon might reflect a significant combinatorial mutational pattern called mutual exclusivity, in the sense that aberrant events tend to avoid co-occurrence if any individual alone can exert the same or similar function [6]. Since a mutation on any site of the miRNA gene can disrupt the normal production/function of its mature miRNA, no site needs to be preferably selected for mutation, leading to the observation of roughly even distribution of mutations across the whole miRNA gene. Similarly, since members from the same miRNA family have same seed sequence and similar physiological functions, combinatorial mutational profiles of a miRNA family may confound the association between the mutations on each single miRNA and relevant molecular/phenotypic traits. These hypotheses deserve a systematic test to determine the pattern and mechanism of the mutational processes in miRNA genes.

Accordingly, exploration of the combinatorial mutational pattern between a miRNA gene and its target genes may also bring new insights. The disruption of a specific miRNA-gene regulation pair can result from a mutation either on the miRNA or on the gene (most often, on the 3'-UTR), or mutations on both. In this sense, the coordinated mutational processes on both the miRNA genes and their target genes can be leveraged to investigate some dysregulated miRNA-gene interactions in cancer, such as the recently recognised miRNA-mediated gene activation [7, 8]. On the other hand, multiple mutations on the same miRNA gene can be combined together (collapsed into a single variable) to gain statistical power in the miRNA-gene and miRNA-phenotype association analysis. In this sense, rare but mutually exclusive variants may work complementarily to determine the phenotypical outcomes.

Analogous to the scenario in protein-coding regions, somatic mutations in noncoding regions such as the miRNA genes herein can also be either disease-causing drivers or just neutral passengers. And most likely, the passengers largely outnumber the drivers. Therefore, the major challenge in next step is to identify those noncoding drivers from the bulk of passengers in an accurate and robust way. The challenge might be even greater than that for coding mutations owing to multiple factors such as the substantial sequencing and mapping artefacts in current whole genome sequencing (WGS), the incomplete annotation of noncoding regions, as well as the inaccurate estimation

E-mail address: [Hua.Tan@uth.tmc.edu](mailto:Hua.Tan@uth.tmc.edu)

<https://doi.org/10.1016/j.ebiom.2020.103084>

2352-3964/© 2020 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

of background mutation rate that is highly heterogeneous across a genome. These noncoding sequence-related attributes will inevitably diminish the prediction power of the tools that heavily depend on mutation frequency. As the WGS data increases, the linkage between somatic noncoding mutations and cancer susceptibility can be probed by the genome-wide association studies (GWAS)-like strategy that has been widely employed for the inherited germline variants. Another promising solution lies on the machine learning based approaches that interrogate the local genomic compositions surrounding the mutation site by learning the inherent sequence features that dictate the mutation's functional role as a driver or passenger. Such kind of learning models can be either a traditional framework with manual feature engineering (see [9] and references therein) or a deep learning algorithm that extracts genomic features in a completely automatic manner (reviewed in [10]).

We are witnessing an era when the whole genome, not just the coding region, is seriously taken into account when studying the causative genetic and epigenetic aberrations of human complex disorders, especially cancer. With increasing focus and effort (sequencing technologies, bioinformatics analysis algorithms and pipelines) devoted to the noncoding DNA domain, we are beginning to hear the sound of silence. Probably, in the near future, no single piece of DNA will be deemed junk anymore. Instead, all base pairs of the whole genome orchestrate to play a symphony.

#### Declaration of interests

No conflict of interest to declare.

#### Contributors

Hua Tan conceived the concepts and wrote the manuscript.

#### Acknowledgments

The author apologizes to the many researchers whose work was not specifically referenced due to space limitations.

#### References

- [1] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 2011;144(5):646–74.
- [2] Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. *Nat Rev Genet* 2016;17(2):93–108.
- [3] Rheinbay E, Nielsen MM, Abascal F, Wala JA, Shapira O, Tiao G, et al. Analyses of non-coding somatic drivers in 2,658 cancer whole genomes. *Nature* 2020;578(7793):102–11.
- [4] Urbanek-Trzeciak MO, Galka-Marciniak P, Nawrocka PM, Kowal E, Szwec S, Giefing M, et al. Pan-Cancer analysis of somatic mutations in miRNA genes. *EBioMedicine* 2020. doi: 10.1016/j.ebiom.2020.103051.
- [5] Bartel DP. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 2004;116(2):281–97.
- [6] Tan H, Bao J, Zhou X. Genome-wide mutational spectra analysis reveals significant cancer-specific heterogeneity. *Sci Rep* 2015;5:12566.
- [7] Tan H, Kim P, Sun P, Zhou X. miRactDB characterizes miRNA-gene relation switch between normal and cancer tissues across pan-cancer. *Brief Bioinform*. 2020. doi: 10.1093/bib/bbaa089.
- [8] Tan H, Huang S, Zhang Z, Qian X, Sun P, Zhou X. Pan-cancer analysis on microRNA-associated gene activation. *EBioMedicine* 2019;43:82–97.
- [9] Tan H, Bao J, Zhou X. A novel missense-mutation-related feature extraction scheme for 'driver' mutation identification. *Bioinform* 2012;28(22):2948–55.
- [10] Eraslan G, Avsec Z, Gagneur J, Theis FJ. Deep learning: new computational modeling techniques for genomics. *Nat Rev Genet* 2019;20(7):389–403.