



Published in final edited form as:

J Med Artif Intell. 2020 March ; 3(7): . doi:10.21037/jmai.2020.01.01.

Machine learning can accelerate discovery and application of cyber-molecular cancer diagnostics

David S. Campo, Yury Khudyakov

Division of Viral Hepatitis, Centers for Disease Control and Prevention, Atlanta, GA, USA

Accurate and early cancer diagnosis is fundamental for clinical management and public health. Unfortunately, the biological complexity of cancer confounds the development of effective diagnostic approaches to its detection. Histological examination of tissue samples obtained by biopsy directly from solid tumors and imaging technologies remain as the mainstays of cancer diagnostics. The liquid biopsy concept aims to overcome the shortcomings of these onco-diagnostics by detecting tumor-derived biomarkers such as circulating tumor cells, extracellular vesicles, nucleosomes, proteins, antigens, and extracellular nucleic acids in blood (1).

Among many, mitochondrial DNA (mtDNA) is one of the most promising biomarkers of liquid biopsy. Mitochondria are highly abundant in human body, exceeding the number of human cells by 100–10,000 times. They play an essential role in the whole-body physiology, being involved in bioenergetics, apoptosis, innate immunity, networks of communication with different cell types and metabolic coordination. Owing to such fundamental involvement of mitochondria in human physiology, mtDNA mutations in general have a highly detrimental effect on cell viability. Nevertheless, the astronomical mitochondrial population size, lack of genetic mechanisms for effective control of mutations, genetic complementation and vegetative segregation of mtDNA establish an environment that supports a significant intra-host mitochondrial genetic heterogeneity, known as heteroplasmy (2).

Some health conditions, such as cancer are potentially conducive to maintaining heteroplasmy. The intra-host mitochondrial genetic diversity detected in blood is very dynamic and may change at the rate usually observed in intra-host viral populations, rapidly responding to, for example, progression of cancer or hepatitis C virus infection (3–5). The dynamic nature of mitochondrial genetic heterogeneity in blood offers potential diagnostic opportunities for the detection of cancer and other health conditions (3,6). The fluid biopsy concept takes advantage of such opportunities and provides guiding principles for diagnosing and managing cancer using blood rather than solid tumor tissue, with several

Correspondence to: David S. Campo. 1600 Clifton Rd., Atlanta, GA 30333, MS H18-4, USA. fyv6@cdc.gov.

Conflicts of Interest: The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

Publisher's Disclaimer: *Disclaimer:* The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

molecular approaches being developed for the direct detection of circulating mtDNA variants associated with cancer (7,8).

We recently showed that heterogeneity at specially selected polymorphic mtDNA sites can be efficiently associated with liver cancer by means of machine learning, suggesting a different research direction towards development of novel cyber-molecular diagnostics (6). Such assays are basically complex computational models capable of extracting diagnostically relevant information from molecular data obtained using Ultra-Deep Sequencing (UDS) technologies. Molecular wet-laboratory assays generate diagnostic information directly from blood by detection of circulating specific sequence variants. Performance of wet-laboratory assays can be greatly afflicted by a limited presentation of tumor-derived genetic markers in blood (9), owing either to a low level of specific variants or to abundance of different mutations associated with cancer, reflecting the complex biological nature of this disease. Assays based on the identification of patterns in molecular data are potentially less sensitive to dilution of tumor-derived markers in blood and make the most of this biological complexity rather than treat it as impediment for the detection.

The major difference between molecular and cyber-molecular assays is in the source of sensitivity and specificity of detection. While conventional liquid biopsies target individual variants that have been found in tumors, the detection targets of cyber-molecular assays are patterns of mutations derived from molecular data. Such patterns result from coordinated evolution driven by epistatic interactions among polymorphic sites (10). Patterns of mutations encoding epistatic interactions among selected sites represent a new source of diagnostically relevant information and can be obtained using UDS of entire genomes or small genomic regions. Here, we explored patterns of site entropy across mtDNA with the rationale being that entropy is highly generalizable because it is not reliant on specific sequence variants or mitochondrial haplogroups. This is but one of a few possible approaches for assessing epistatic interactions from short UDS reads. Increased length of uninterrupted sequences of the entire mtDNA that can be obtained using other technologies will enable a more efficient detection of epistatic associations with cancer using variety of machine learning techniques.

Extension of the fluid biopsy concept from the direct detection of tumor-derived molecules in blood to the detection of patterns of epistatic interactions among mtDNA sites and their association with cancer and other health conditions is a promising venue of research towards efficient onco-diagnostics and cyber-molecular diagnostics of diseases. Shift from straight identification of disease markers using laboratory techniques to extraction of diagnostically relevant information from complex genetic associations in molecular data obtained from biological samples is a hallmark of the disease cyber-diagnostics, with machine learning being the engine of the concept. Application and development of machine learning techniques is central to discovery of new disease markers and advancement of diagnostics.

References

1. Yong E Cancer biomarkers: Written in blood. *Nature* 2014;511:524–6. [PubMed: 25079538]
2. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* 2015;16:530–42. [PubMed: 26281784]

3. Campo DS, Roh HJ, Pearlman BL, et al. Increased Mitochondrial Genetic Diversity in Persons Infected With Hepatitis C Virus. *Cell Mol Gastroenterol Hepatol* 2016;2:676–84. [PubMed: 28174739]
4. Kustanovich A, Schwartz R, Peretz T, et al. Life and death of circulating cell-free DNA. *Cancer Biol Ther* 2019;20:1057–67. [PubMed: 30990132]
5. Wallace DC, Chalkia D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Biol* 2013;5:a021220. [PubMed: 24186072]
6. Campo DS, Nayak V, Srinivasamoorthy G, et al. Entropy of mitochondrial DNA circulating in blood is associated with hepatocellular carcinoma. *BMC Med Genomics* 2019;12:74. [PubMed: 31167647]
7. He Y, Wu J, Dressman DC, et al. Heteroplasmic mitochondrial DNA mutations in normal and tumour cells. *Nature* 2010;464:610–4. [PubMed: 20200521]
8. Uzawa K, Baba T, Uchida F, et al. Circulating tumor-derived mutant mitochondrial DNA: a predictive biomarker of clinical prognosis in human squamous cell carcinoma. *Oncotarget* 2012;3:670–7. [PubMed: 22832574]
9. Weerts MJA, Timmermans EC, van de Stolpe A, et al. Tumor-Specific Mitochondrial DNA Variants Are Rarely Detected in Cell-Free DNA. *Neoplasia* 2018;20:687–96. [PubMed: 29842994]
10. Campo D, Dimitrova Z, Mitchell R, et al. Coordinated evolution of the hepatitis C virus. *PNAS* 2008;105:9685–90. [PubMed: 18621679]