Somatic Variant Richness in the Mitochondrial Genome

Ethan Ashby

9/4/2020

Introduction

Mitochondria have long thought to be implicated in cancer growth and progression, due to altered energy metabolism in carcinogenic tumors (Brandon, Baldi, and Wallace (2006); Zong, Rabinowitz, and White (2016); Hanahan and Weinberg (2011)). Additionally, mitochondria are involved in many cellular processes, ranging from biosynthesis to apoptosis, that are intrinsically linked to carcinogenesis. While the circular genome of the mitochondria is small (16.6 kb), it encodes several 13 proteins that form the cellular respiration machinery of the organelle. The mitochondrial genome also encodes several noncoding elements, predominantly transfer RNAs.

High depth, multi-sample molecular characterizations of the mitochondrial genome have illustrated a number of parallels between mtDNA and nDNA (Yuan et al. (2020)). Mutational hotspots were observed in the regulatory D loop and the *ND4* gene. *ND5* was most frequently mutated, *ND4* was most frequently mutated in prostate and lung cancers, and *COX1* was most frequently mutated in breast, cervical, and bladder cancers. Yuan et al. identified that truncating mutations were enriched in kidney, colorectal, and thyroid cancers, while copy number variants showed remarkable heterogeneity between cancers. In summary, like the nuclear genome, the mitochondrial genome contains significant mutational heterogeneity in cancer.

In the nuclear genome, there is remarkable heterogeneity in rare variants between genes and across cancer types. Greater than 90% of mutations in The Cancer Genome Atlas (TCGA) are singletons, meaning they appear once across the TCGA sequenced tumors. Previous research used a non-parametric Bayesian methods developed in the fields of ecology and computational linguistics to mine clinically relevant information from rare somatic mutations in major cancer genes (Chakraborty et al. (2019)). This study showed that probabilties of encountering previously unseen variants varied dramatically between nuclear genes and primary sites. This project represented an advancement in utility of the preponderance of previouslly neglected, rare mutational information for important clinical tasks like primary site classification.

There has been limited research on the frequencies of rare variation within the mitochondrial genome across several tissue types. Yuan et al. created The Cancer Mitochondria Atlas (TCMA), an open-access web portal to support research and education surrounding mitochondrial cancer genetics. The portal conatins their somatic variant, copy number, and nuclear transfer data available. This research applied the bayesian unseen variant probability estimation method to features in the mitochondrial genome to investigate whether rare mitochondrial variant patterns are tissue specific.

Methods

I applied the Smoothed Good-Turing frequency estimation functions sourced from the `varprobs` package written by Chakraborty et al.

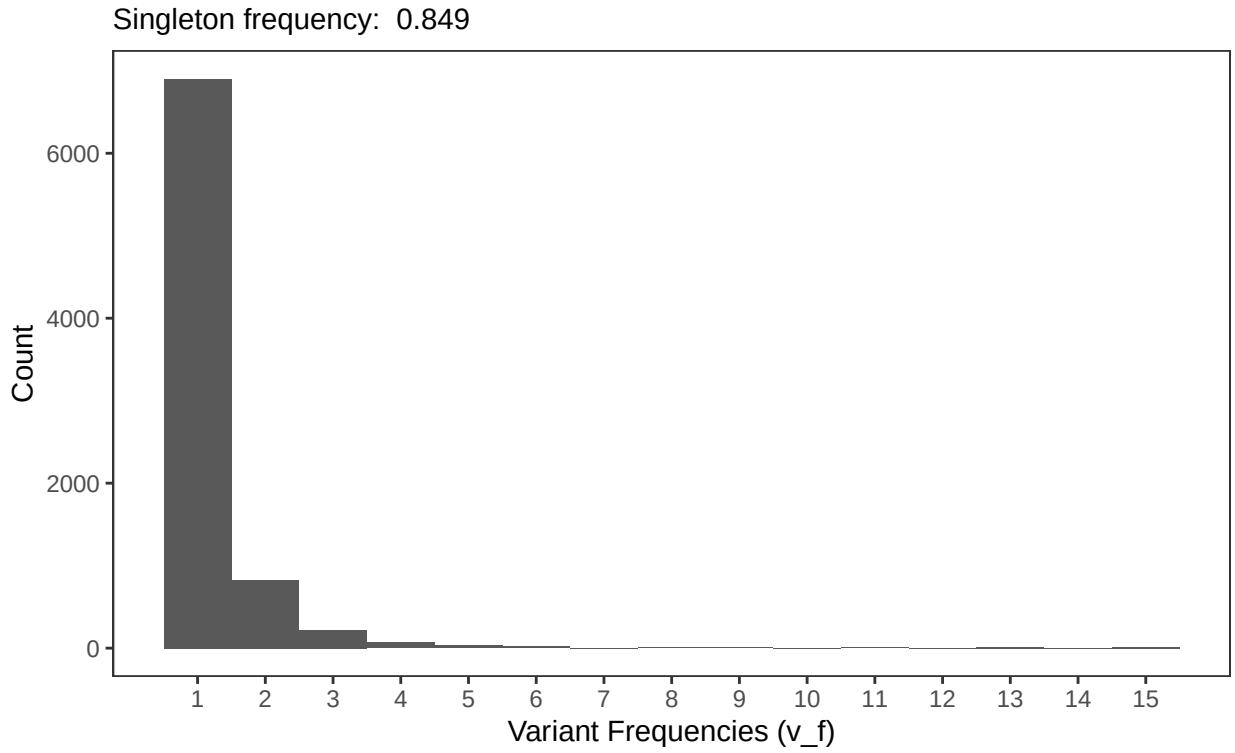

Figure 1: Mitochondrial single nucleotide variants are dominated by singleton variants

Results

Singleton Variants are Abundant in the Mitochondrial Genome

In *Figure 1*, a plot of the frequencies of frequencies indicates that the vast majority of variants (0.849 of total SNV) are singletons, meaning that they are observed only once in the 2173 tumor samples. This proportion is quite consistent with the proportion of singleton variants observed in the nuclear genome, as in The Cancer Genome Atlas, >90% of somatic variants were singletons. This finding reinforces the notion that somatic mutations in mitochondria are predominantly rare, and finding ways to extract clinical information from these rare variants is a worthy task.

Good-Turing Probabilities Show Gene-Specific and Tissue-Specific Patterns in Unseen Variant Probabilities

These singleton variant frequencies were broken down by genomic feature and tissue type, and used to calculate the Smoothed Good-Turing Probability, corresponding to encountering a previously uncatalogued (or new) variant in each gene and tissue type. Bubbleplots of the unseen variant probabilities (*Figure 2*) show variability in these probabilities, suggestive that they may contain gene and tissue-specific signals.

Conclusions

References

Brandon, M, P Baldi, and D C Wallace. 2006. "Mitochondrial mutations in cancer." *Oncogene* 25: 4647–62. https://doi.org/10.1038/sj.onc.1209607.

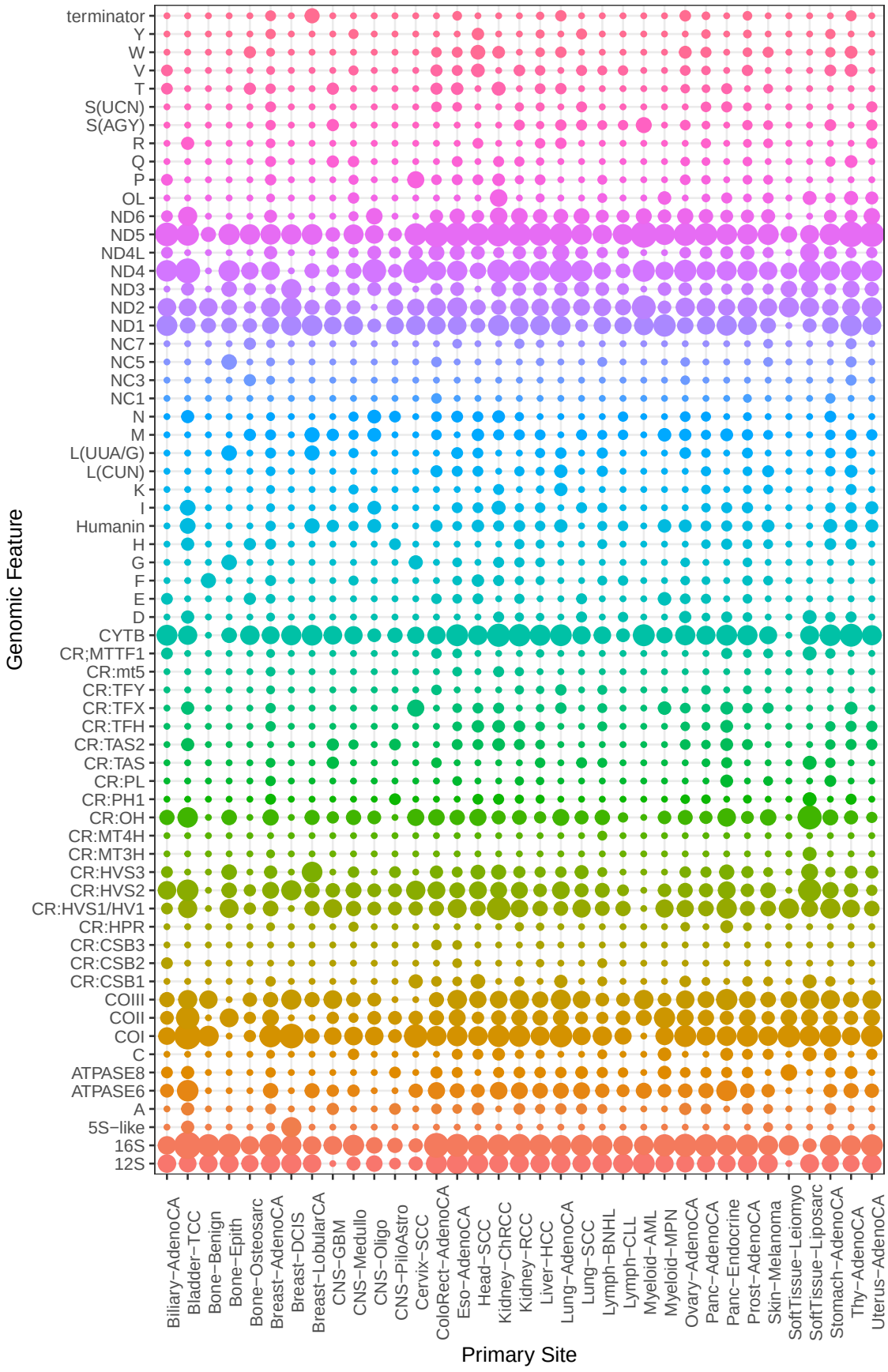

Figure 2: Unseen variant probabilities show tissue specific patterns in several mitochondrial genomic features.

Chakraborty, Saptarshi, Arshi Arora, Colin B. Begg, and Ronglai Shen. 2019. "Using somatic variant richness to mine signals from rare variants in the cancer genome." *Nature Communications* 10: 1–9. https://doi.org/10.1038/s41467-019-13402-z.

Hanahan, Douglas, and Robert A Weinberg. 2011. "Review Hallmarks of Cancer : The Next Generation." *Cell* 144 (5): 646–74. https://doi.org/10.1016/j.cell.2011.02.013.

Yuan, Yuan, Young Seok Ju, Youngwook Kim, Jun Li, Yumeng Wang, Christopher J Yoon, Yang Yang, et al. 2020. "Complete molecular characterization of mitochondrial genomes in human cancers." *Nature Genetics* 52 (March): 342–52. https://doi.org/10.1038/s41588-019-0557-x.

Zong, Wei-xing, Joshua D Rabinowitz, and Eileen White. 2016. "Mitochondria and Cancer." *Molecular Cell* 61 (5): 667–76. https://doi.org/10.1016/j.molcel.2016.02.011.