

CMSC499A: Mutation Visualization

Mark Keller *

Spring 2018

Abstract

Identification and analysis of patterns in data can be difficult without visualization tools. Datasets of somatic mutations in cancer are no exception. Recent whole-genome sequencing projects such as PanCancer Analysis of Whole Genomes (PCAWG) have produced large amounts of data to be explored. Of great importance is the classification of different mutational processes and the mutational signatures¹ they leave behind. As new mutational signatures continue to be discovered, observation of the levels of signature activity, called signature exposure, helps show how the underlying mutational processes differ across cancer types, time, and environmental variables.

1 Introduction

Slight modifications to data sources used in past studies of mutations and mutation signatures have the potential to solidify or weaken conclusions made based on limited data. For example, one might be interested in reproducing a past study across additional cancer types or in the context of additional variables, such as smoking status or age. Web-based interactive visualizations allow for this type of comparison to be done across data sets and features in a way that is accessible and fast. This semester, I have focused on building a visualization tool to do just this, using whole-genome somatic mutation datasets and mutation signature definitions from various sources.

The resulting tool, X (<http://link.to/it>), is a web-based mutation data browser that enables exploration of mutation signatures, cancer types, mutation types, and kataegis events. X allows users to choose sequencing project data sets by cancer type, as well as mutation signature combinations (of which can be selected based on cancer-type-specific presets) before plotting this data. Using the data-driven documents JavaScript library (D3.js)², plots can be customized down to the pixel. Showing donor clinical variables, such as smoking and alcohol usage, and their relationship to mutation signatures, requires this fine control that D3 provides. In addition, D3 contains APIs for easy implementation of custom interactive features, such as highlighting based on mouse position and zooming based on mouse dragging (“brushing”).

2 Case Study

This tool brings the power to visually analyze relationships between signature exposures and clinical variables. Interactive sorting and dynamic calculations of signature exposures allow trends

*mentored by Professor Max Leiserson

to be observed quickly. Connections generally assumed to exist can be verified and examined, such as specific signatures being associated with tobacco usage. Local datasets can be visualized, allowing functionality to extend beyond the public datasets we have processed.

3 Methods

References

- [1] Ludmil B Alexandrov, Serena Nik-Zainal, David C Wedge, Samuel AJR Aparicio, Sam Behjati, Andrew V Biankin, Graham R Bignell, Niccolo Bolli, Ake Borg, Anne-Lise Børresen-Dale, et al. Signatures of mutational processes in human cancer. *Nature*, 500(7463):415, 2013.
- [2] Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer. D³ data-driven documents. *IEEE transactions on visualization and computer graphics*, 17(12):2301–2309, 2011.