**Pan-Cancer Analysis Interface with TCGAnalyzeR**

metin, noel ağacı, noel içeren bir resim

Açıklama otomatik olarak oluşturuldu

**Computer Engineering**

**Senior Project Report**

**Advisor:** Tuğba Önal SÜZEK, PhD

**Ayşe Nur Özer**

**Date:** 30/01/2024

**Pan-Cancer Analysis Interface with TCGAnalyzeR**

Ayşe Nur Özer

Computer Engineering

**Summary**

Cancer is a disease resulting from the uncontrolled growth and multiplication of normal cells in the body, often occurring when the normal life cycle of cells is disrupted. Despite ongoing research utilizing new technologies, our understanding of the mechanisms and distribution of cancer remains insufficient.

The pan-cancer approach focuses on understanding common features among different types of cancer to develop general cancer treatments. Pan-cancer analysis examines genomic and cellular changes across various cancer types. It aims to identify genes causing oncogenesis, detect different genomic anomalies, and analyze mutations irrespective of tumor origins.

The Cancer Genome Atlas (TCGA), a groundbreaking cancer genomics program covering 33 cancer types in over 11,000 patients, plays a pivotal role in pan-cancer research.

Muğla Sıtkı Koçman University has developed the TCGAnalyzeR[1] web interface, utilizing the Shiny R package, allowing clinical oncologists and cancer researchers to filter and visualize analyses based on subgroups and/or gene sets. Our Project aims to add a Pan-Cancer interface to TCGAnalyeR.

In this project, a functional pan-cancer analysis module using the open-source R package Shiny is developed and integrated into the TCGAnalyzeR interface. This integration facilitates efficient analysis of individual gene statistics across all 33 cancers. The enhanced sensitivity in data analysis aims to provide a deeper understanding of genetic and cellular differences in various tumor types.

Our project stands out by creating a web tool for worldwide pan-cancer analysis with TCGA data. It helps researchers understand genetic changes in various cancers, identifying common factors for developing general cancer treatments. The user-friendly web interface allows interactive visualization of processed data, making analysis more accessible.

The development of this pan-cancer interface module will enable clinicians to conduct more effective individual gene and cohort-based statistical analyses, providing researchers with a fresh perspective on understanding genetic changes in different cancer types.

By integrating this interface with TCGAnalyzeR, it will offer more comprehensive access to genetic data obtained from large sources like The Cancer Genome Atlas (TCGA)[2]. This integration not only provides researchers with the powerful analysis tools of TCGAnalyzeR but also offers a customized solution for pan-cancer analysis. It enhances cancer research by providing a more inclusive perspective based on a broad dataset.

**Table of Contents**

**1. Introduction4**

**2. Methods4**

2.1 Data Preprocessing4

2.2 Creating User Interface5

2.3 User Interface Integration5

2.4 Comparison of Data with TCGAnalyzeR5

**3. Results6**

**4. Conclusion8**

**5. References9**

1. **Introduction**

In cancer research, clinicians and molecular oncologists use software tools to understand a patient's risk and plan treatments. Making these tools user-friendly is crucial as science advances and more data becomes available. Our project focuses on creating an easy-to-use interface that combines data from TCGAnalyzeR, allowing clinicians and molecular oncologists to analyze genetic changes in 33 different cancers more effectively.

By connecting our interface with TCGAnalyzeR, we give researchers access to a wide range of genetic data. This not only includes TCGAnalyzeR's powerful tools but also a solution for studying genetic changes in various cancers.

Our goal is to help researchers quickly and efficiently analyze genetic differences in different cancer types. This will enable them to make better decisions and optimize treatment strategies. In the bigger picture, our project aims to contribute to improved cancer understanding, more effective treatments, and ultimately, disease prevention.

1. **Methods**

**2.1 Data Preprocessing**

We will perform data preprocessing utilizing the dplyr[3] library in R and the pandas[4] library in Python.

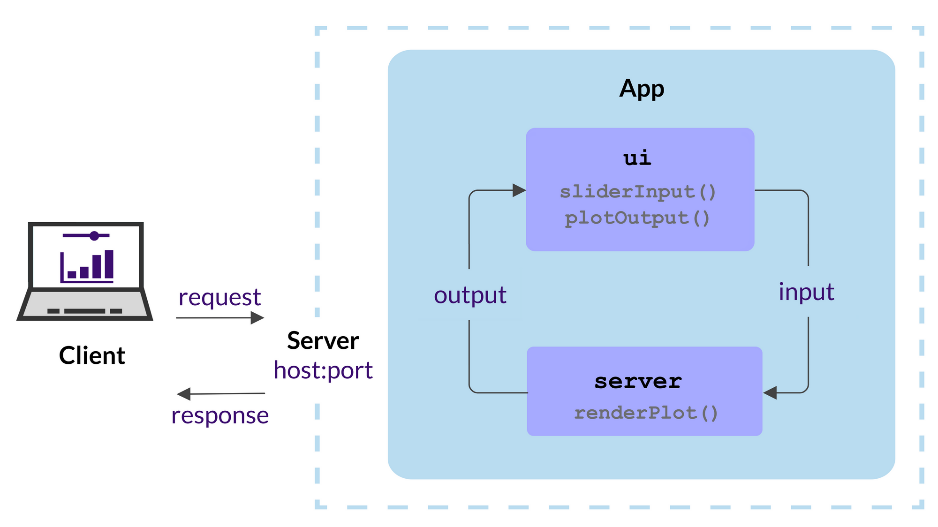
metin, daire içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Data that can be found in TCGA[5].*

**2.2 Creating User Interface**

We will develop a user-friendly interface using the Shiny[5] library in R.



*Workflow chart of Shiny web application[6].*

**2.3 User Interface Integration**

The developed user interface will be integrated into TCGAnalyzeR.

Since the developed interactive pan-cancer interface will be installed on an independent and different server, it will be seamlessly integrated into the TCGAnalyzeR web application.

**2.4** **Comparison of Data with TCGAnalyzeR**

The developed interface will compare genomic data with the TCGAnalyzeR interface, aiming to detect errors and understand pan-cancer characteristics. It will use statistical methods to identify significant genetic variations, especially between different cancer cells.

metin, diyagram, ekran görüntüsü, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Example classification table generated using TCGA data [7].*

1. **Results**

metin, ekran görüntüsü, yazı tipi, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, diyagram içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, çizgi, yazı tipi içeren bir resim

Açıklama otomatik olarak oluşturuldu

metin, ekran görüntüsü, yazı tipi, öykü gelişim çizgisi; kumpas; grafiğini çıkarma içeren bir resim

Açıklama otomatik olarak oluşturuldu

This interface will be integrated with the TCGAnalyzeR interface, providing more comprehensive access to genetic data obtained from big data sources such as The Cancer Genome Atlas (TCGA).

The integration will provide researchers with a customized solution for pan-cancer analysis in addition to the powerful analysis tools offered by TCGAnalyzeR, providing a more holistic perspective on cancer research based on a broad dataset.

**4****. Conclusion**

After preprocessing our TCGA data using multiple methods, we will create a web application interface using Shiny. This interface will help clinicians and molecular oncologists in treatment planning by identifying common features of different cancer types on a single gene-based or cohort-based basis and planning common targeted treatments.

Since TCGA data is constantly updated, data integration will be carried out regularly to prevent inaccuracies in the data.

Output of the interface will create using plotly and complexheatmap libraries from R.

metin, diyagram, ekran görüntüsü, çizgi içeren bir resim

Açıklama otomatik olarak oluşturuldu

*Representation of the Project*

**5. References**

*[1]* [*http://tcganalyzer.mu.edu.tr*](http://tcganalyzer.mu.edu.tr)

*[2]* ["The Cancer Genome Atlas Pan-Cancer analysis project."](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3919969) *Nature Genetics.*

*[3]* [*https://dplyr.tidyverse.org*](https://dplyr.tidyverse.org)

*[4]* [*https://pandas.pydata.org*](https://pandas.pydata.org)

*[5]* [*https://shiny.posit.co*](https://shiny.posit.co)

*[6]* [*https://www.cusabio.com*](https://www.cusabio.com) *TCGA Database- A Good Partner of Researchers*

*[7]* [*https://hosting.analythium.io/the-anatomy-of-a-shiny-application/*](https://hosting.analythium.io/the-anatomy-of-a-shiny-application/)