



Abstract

High throughput technology has been widely used by researchers to understand diseases at the molecular level. Database and servers for downloading and analyzing these publicly data is available as well. But there is still lacking tools for facilitating researchers to study the function of genes in pathways views by integrated public omics data.

Introduction

The broad applications of High throughput technology in biological and medical research yields a new perspective to understand diseases at the molecular level.

In the past decades, a large amount of patient-driven transcriptional omics data are accumulated in public databases such as TCGA and GEO. There are also many web servers for downloading and analyzing these publicly available data, such as cBioportal, Xena and GEPIA. However, these tools have not adequately addressed some particular needs from biomedical researchers to utilize public omics data in their research, these needs includes:

- 1. Patient-based evidence of molecular pathway alteration caused by a specific gene.
2. Discovery of potential upstreaming or downstream of a target gene.
3. Guidance on candidate cell line selection for gene function experiments.

Material and Methods

Genomics datasets used in this study are collected from TCGA, GEO and CCLE, data types including patients' transcriptional expression, mutation and clinical information. Literatures data were download from NCBI PubMed API. Django web application development framework is used following the ModelView-Controller (MVC) model. A new processes is implemented over Controller to handle GSEA analysis separately from the traditional MVC model, consisting of RabbitMQ and Celery.

Results

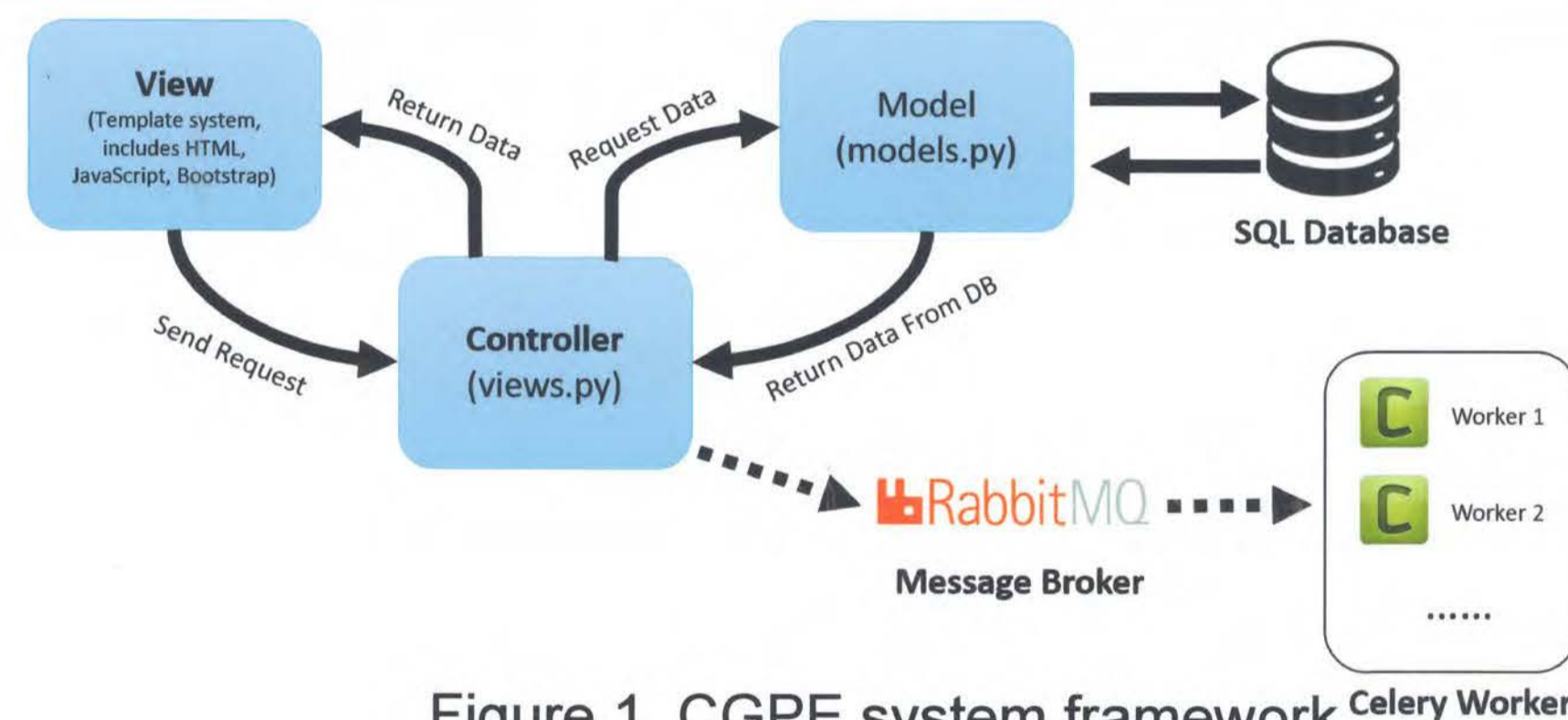


Figure 1. CGPE system framework.

We implemented a working version of the proposed CCPE with the following major functions:

- 1. Gene Hot Index: summarized visualized publication statistics for each gene.
2. onlineGSEA: perform GSEA analysis on integrated public and in-house data.
3. CellLine Selector: help with picking the most appropriate cell lines given an interested gene name and cancer type

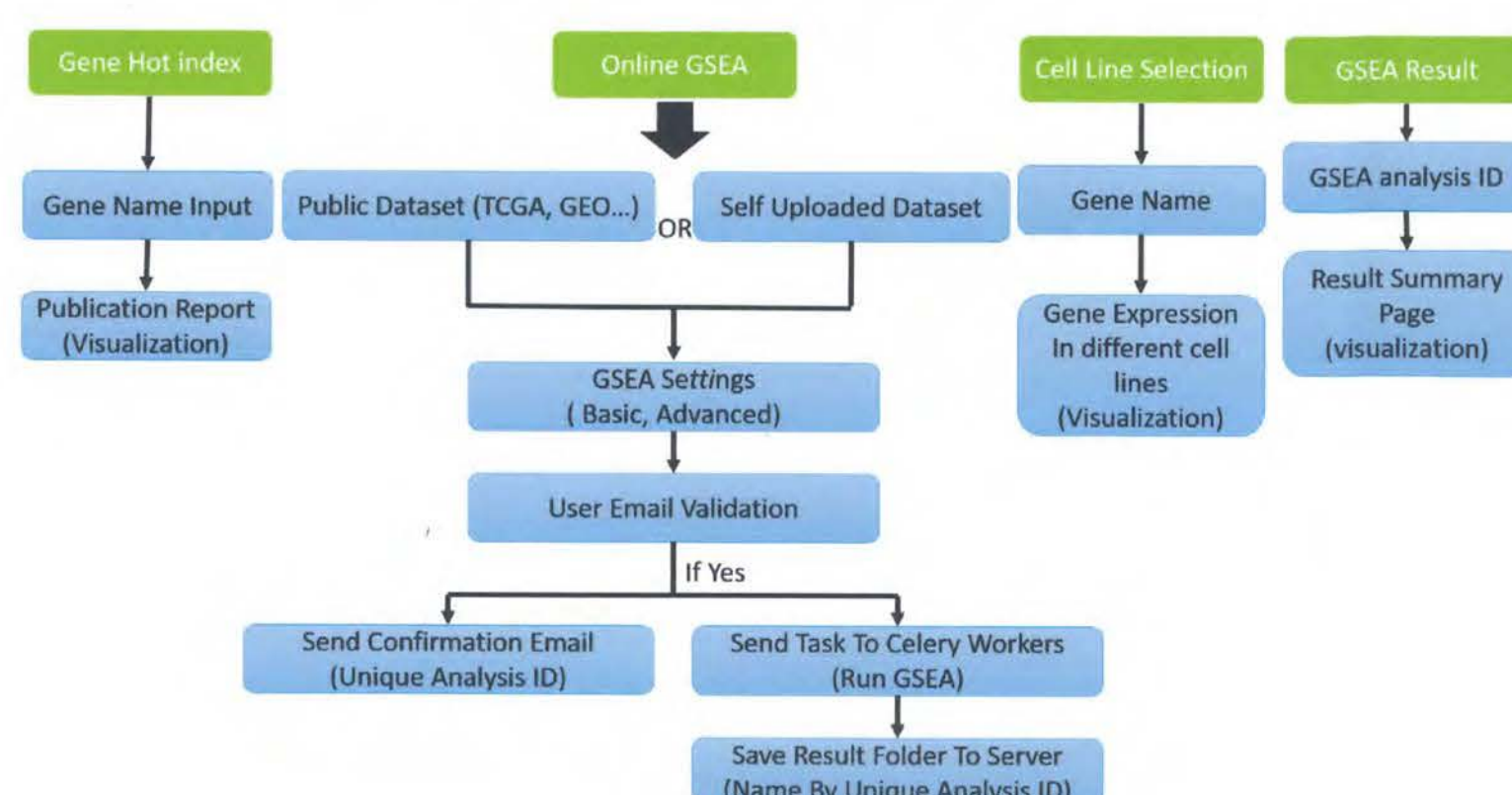


Figure 2. CGPE website structure.

Results continued

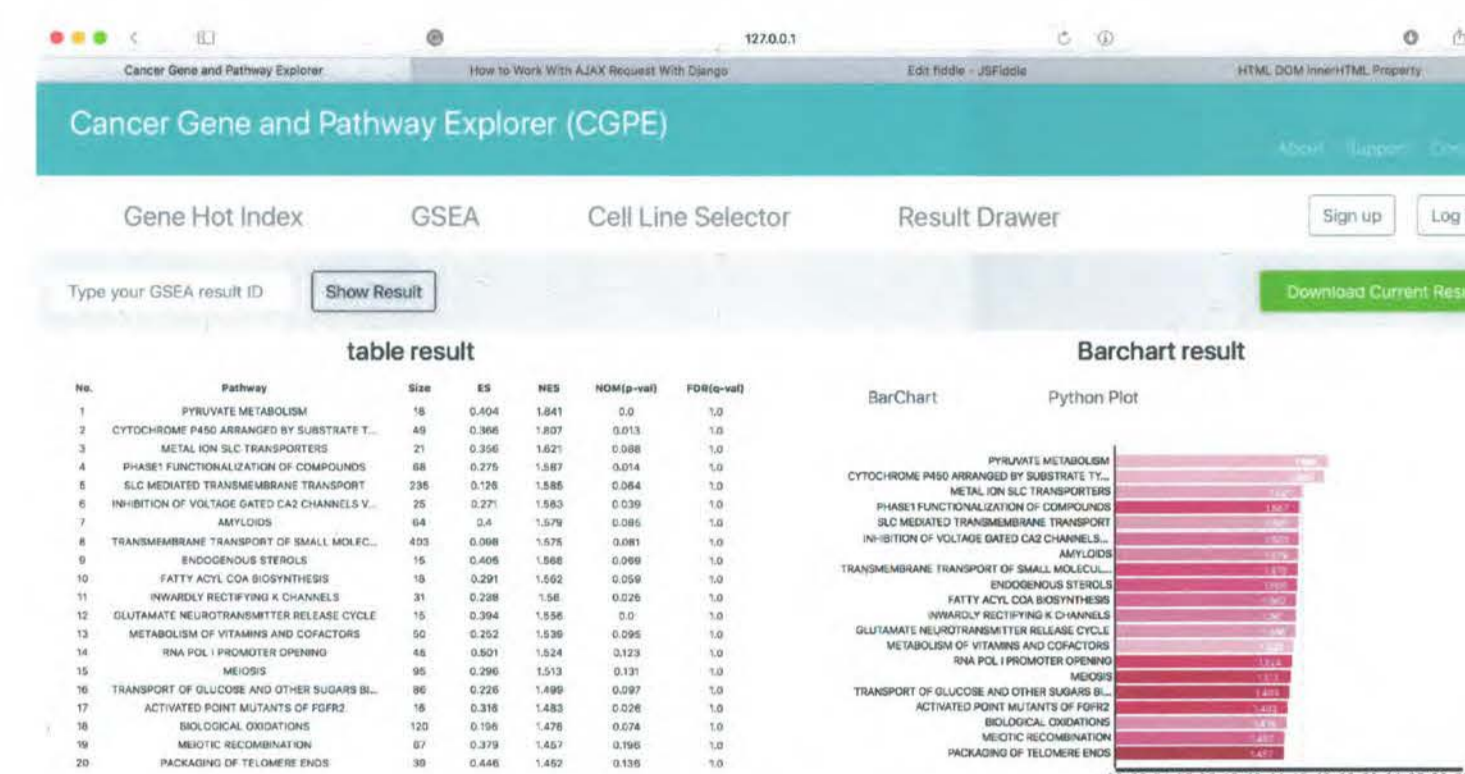


Figure 3. CGPE onlineGSEA result page.

All the analyze results are visualized with different charts including GSEA results.

Conclusions

The new CGPE provides:

- 1. convenient tool for biomedical researchers as a discovery tool to analyze the available public cancer-related data by performing state-of-art algorithms like GSEA.
2. guidance on exploring potentially related target genes for the next step of work.
3. Providing further evidence for guiding vitro experiment in cancer research.

The development of CGPE will further accelerate cancer biomedical research by providing insights from a genomic perspective.

References

List of references including: Tang, Zefang, et al. "GEPIA: a web server for cancer and normal gene expression profiling and interactive analyses." Nucleic acids research 45.W1 (2017): W98-W102.