SCHOOL OF INFORMATICS AND COMPUTING

INDIANA UNIVERSITY Department of BioHealth Informatics IUPUI CGPE: A user-friendly gene and pathway explore webserver for public cancer transcriptional data Jiannan Liu, Chuanpeng Dong, Yunlong Liu, Huanmei Wu Indiana University School of Informatics and Computing Indiana University-Purdue University Indianapolis



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 availableas well. But there is still lacking tools for facilitaing researchers to study the function of genes in pathways views by integrated public omics data.

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 continued

Passan Dana and Delivery Distance			And a strate of an end of a strategy of the				warman metani		and a more all with a more than the
Cancer Gene and ratinway Explorer			How to work with work wednest with baingo				EQLUDDe - Da	FIDDIE	HTML DOM MINEHTINL PREPARTY
Ca	ancer Gene and Pathy	vay	Explo	orer (CGPE)				
Gene Hot Index		GSEA			Cell Line Selector		Result Drawer		Sign up Log in
Туре	e your GSEA result ID Show Re	esuit							Download Current Result
	table result						Barchart result		
No.	Pathway	Size	ES	NES	NOM(p-val)	FOR(q-val)	BacChart	Buthon Blot	
1	PYRUVATE METABOLISM	18	0.404	1.841	0.0	1.0	Dal Ghan	Python Pior	
2	CYTOCHROME P450 ARRANGED BY SUBSTRATE T	49	0.366	1.807	0.013	1.0			
3	METAL ION SLC TRANSPORTERS	21	0.356	1.621	0.088	1.0		1	
4	PHASE1 FUNCTIONALIZATION OF COMPOUNDS	68	0.275	1.587	0.014	10	PPRUVATE METABOLISM CYTOCHROME P450 ARRANGED BY SUBSTRATE TY METALION SLC TRANSPORTERS PHASET FUNCTIONALIZATION OF COMPOUNDS SLC MEDIATED TRANSMAINARE TRANSPORT		
6	SLC MEDIATED TRANSMEMBRANE TRANSPORT	235	0.128	1,585	0.084	1.0			
6	INHIBITION OF VOLTAGE GATED CA2 CHANNELS V.	25	0.271	1.583	0.039	1.0			1967
7	AMYLOIDS	64	0,4	1.679	0.086	1.0			1.00
8	TRANSMEMBRANE TRANSPORT OF SMALL MOLEC	403	0.098	1.575	0.081	1.0	INHIBITION OF VOLTAGE DATED CA2 CHANNELS		1000
9	ENDOGENOUS STEROLS	15	0.405	1.568	0.069	1.0	TRANSME MIDDAME TRANSPORT OF CHAPT MODE FOR		101
10	FATTY ACYL COA BIOSYNTHESIS	18	0.291	1.562	0.059	7.0	ENDOGENOUS STERCLS		1417
11	INWARDLY RECTIFYING K CHANNELS	31	0.238	1.5E	0.026	1.0	FATTY ACYL COA BIOSYNTHESIS		
12	DEUTAMATE NEUROTRANSMITTER RELEASE CYCLE	15	0.394	1.556	0.0	3.0	INWARDLY RECTIFYING K CHANNELS		
13	METABOLISM OF VITAMINS AND COFACTORS	50	0.252	1.539	0.095	1.0	GLUTAMATE NEUROTRANSMITTER RELEASE CYCLE		- 1940
14	RNA POL I PROMOTER OPENING	46	0.501	1.524	0,123	1.0	METABOLISM OF VITAMINS AND COFACTORS		
15	MEIOSIS	95	0.296	1.513	0.131	1.0	ARVA PULL I PROMOTER O DESINIO MEDORIS TRANSPORT OF GLUCOSE AND OTHER SUGARS BL. ACTIVATED POINT MU L'ANTE OF FORIEZ ACTIVATED POINT MU L'ANTE OF FORIEZ BULDOREA: COMPATIONS		Contraction of the second s
16	TRANSPORT OF GLUCOSE AND OTHER SUGARS BI-	86	0.226	1.499	0.097	1.0			1400
17	ACTIVATED POINT MUTANTS OF FGFR2	18	0.316	1.483	0.026	1.0			Sales -
									and the second se

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 includs:





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.

- 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

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

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



 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

- 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.
- Gao, Jianjiong, et al. "Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal." *Sci. Signal.* 6.269 (2013): pl1-pl1.
- Subramanian, Aravind, et al. "Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles." *Proceedings of the National Academy of Sciences* 102.43 (2005): 15545-15550.
- Tomczak, Katarzyna, Patrycja Czerwińska, and Maciej Wiznerowicz. "The Cancer Genome Atlas (TCGA): an



Figure 2. CGPE website structure.

