UCSCXenaShiny: an R package for exploring and analyzing UCSC Xena public datasets in web browser

Shixiang Wang

wangshx@shanghaitech.edu.cn

*

Yi Xiong

bob@example.com

Kai Gu

cat@example.com

Longfei Zhao

derek@example.com

Yin Li

cat@example.com

Fei Zhao

derek@example.com

Xuejun Li

derek@example.com

Xue-Song Liu

liuxs@shanghaitech.edu.cn

†

2020-07-05

Abstract

Motivation: UCSC Xena platform provides huge amounts of processed cancer omics data from big public projects like TCGA or individual reserach groups for enabling unprecedented research opportunities. In 2019, we developed UCSCXenaTools, an R package for retrieval of UCSC Xena data. However, an easier dataset exploration and analysis tool is still lack, especially for researchers without programming experience.

Results: We develop UCSCXenaShiny, an R Shiny package to quickly explore, download all datasets from UCSC Xena data hubs. In addiction, a module based analysis framework is constrcuted to analyze data.

Availability: `https://github.com/openbiox/UCSCXenaShiny`.

Contact: `wangshx@shanghaitech.edu.cn` or `liuxs@shanghaitech.edu.cn`.

Keywords: UCSC Xena; cancer genomics; TCGA

1   Introduction

Over the past decade, programs including TCGA, ICGC, PCAWG, GTEx, CCLE and etc. have generated large amounts of molecular data characterizing the landscape of more than ten thousands of tumors from genomic, epigenomic and proteomic aspects. The data have been preprocessed and stored at data hubs of UCSC Xena platform along with many public cancer datasets from individual research groups, providing unprecedented opportunities for eithor simple or systematic exploration of cancer behaviors and mechanisms at multiple molecular layers in individual caner type or across cancer types.

In 2019, we developed UCSCXenaTools, an open-source R package for retrieving and assembling public UCSC Xena data. UCSCXenaTools was developed to communicate with UCSC Xena data hubs for downloading datasets or dataset subsets, querying metadata of data hub, cohort or dataset. Despite UCSC Xena platform itself allows users to explore and analyze data, it is hard for researchers to quickly explore all available datasets, locate what they need in their research and download useful datasets. Besides, the analysis features provided by UCSC Xena platform mainly focus on individual cohort data, thus lack of full-feature functionality.

To this end, we develop an open-source R Shiny package UCSCXenaShiny for cancer community to allow researchers to explore and analyze datasets from UCSC Xena data hubs in web browser. In addiction, an extensible module based analysis framework is constrcuted to analyze data. Currently, several modules providing single gene expression analysis and visualization are implemented.

---

*Corresponding author; Email: wangshx@shanghaitech.edu.cn

†Corresponding author; Email: liuxs@shanghaitech.edu.cn

Here are two sample references: Feynman and Vernon Jr. [1963; Dirac, 1953]. Bibliography will appear at the end of the document.

2   Materials and methods

2.1   Dataset exploration

2.2   Single gene expression analysis

3   Results

	Data hub	Cohorts	Datasets	URL
1	tcgaHub	38	715	https://tcga.xenahubs.net
2	gdcHub	41	528	https://gdc.xenahubs.net
3	publicHub	36	109	https://ucscpublic.xenahubs.net
4	pcawgHub	2	53	https://pcawg.xenahubs.net
5	toilHub	5	50	https://toil.xenahubs.net
6	singlecellHub	18	54	https://singlecellnew.xenahubs.net
7	icgcHub	3	23	https://icgc.xenahubs.net
8	pancanAtlasHub	1	22	https://pancanatlas.xenahubs.net
9	treehouseHub	10	26	https://xena.treehouse.gi.ucsc.edu
10	atacseqHub	2	9	https://atacseq.xenahubs.net
11	kidsfirstHub	3	50	https://kidsfirst.xenahubs.net

Table 1: This is the table caption

Table 1 summarise the cohort and dataset number available at different UCSC Xena data hubs. TCGA project is the major dataset controbutor. There are total 1639 datasets available in UCSC Xena data hubs.

You can reference this figure as follows: Fig. 1.

You can reference this table as follows:

You can cross-reference sections and subsections as follows: Section 2 and Section ??.
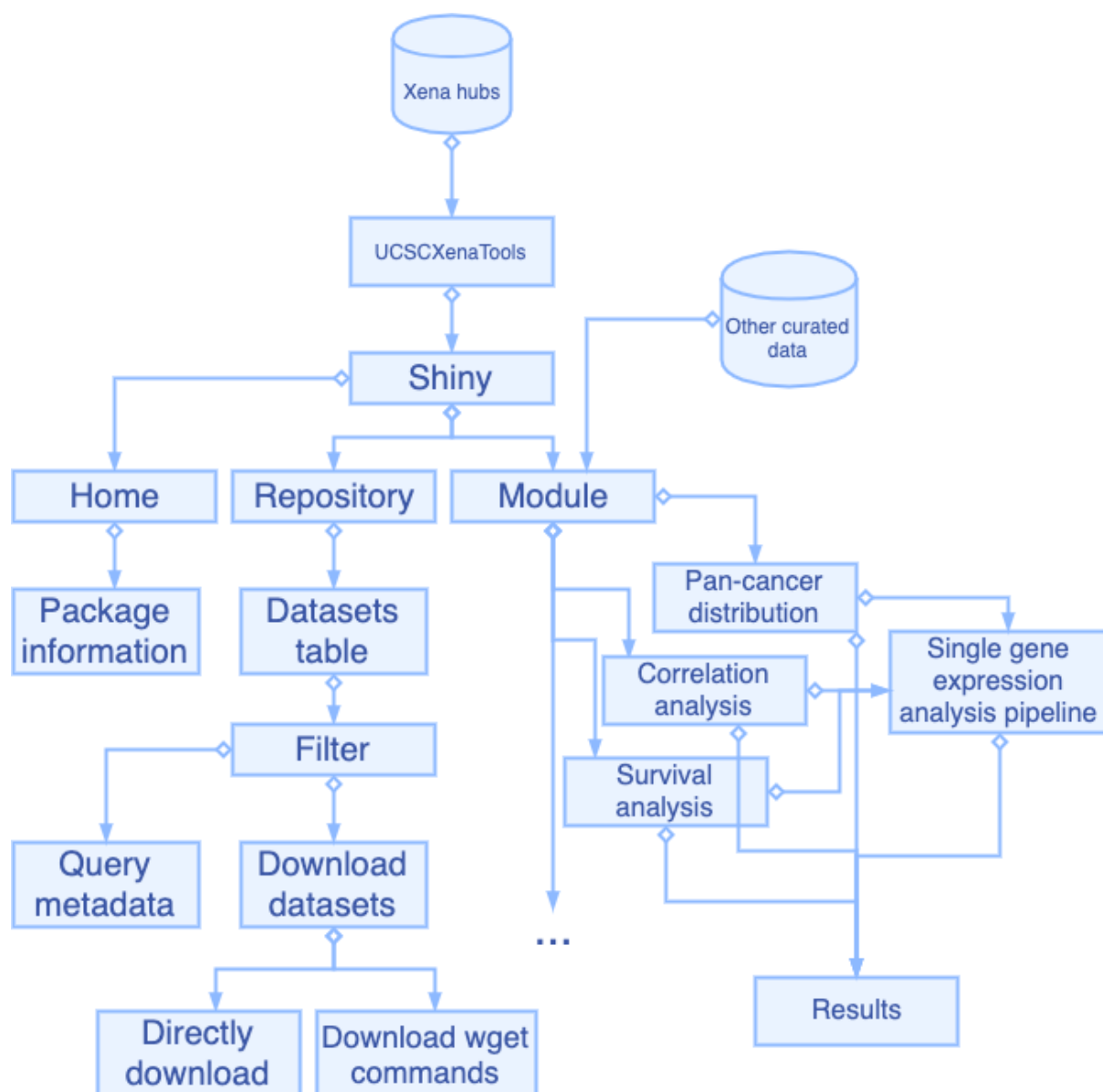
Acknowledgements

Figure 1: Software architecture and functional flowchart of UCSCXena Shiny

References

P.A.M. Dirac. The lorentz transformation and absolute time. *Physica*, 19(1-–12):888–896, 1953. doi: 10.1016/S0031-8914(53)80099-6.

R.P Feynman and F.L Vernon Jr. The theory of a general quantum system interacting with a linear dissipative system. *Annals of Physics*, 24:118–173, 1963. doi: 10.1016/0003-4916(63)90068-X.