Using

to make Biomedical Research outcomes reproducible

Soroor Hediyeh-zadeh

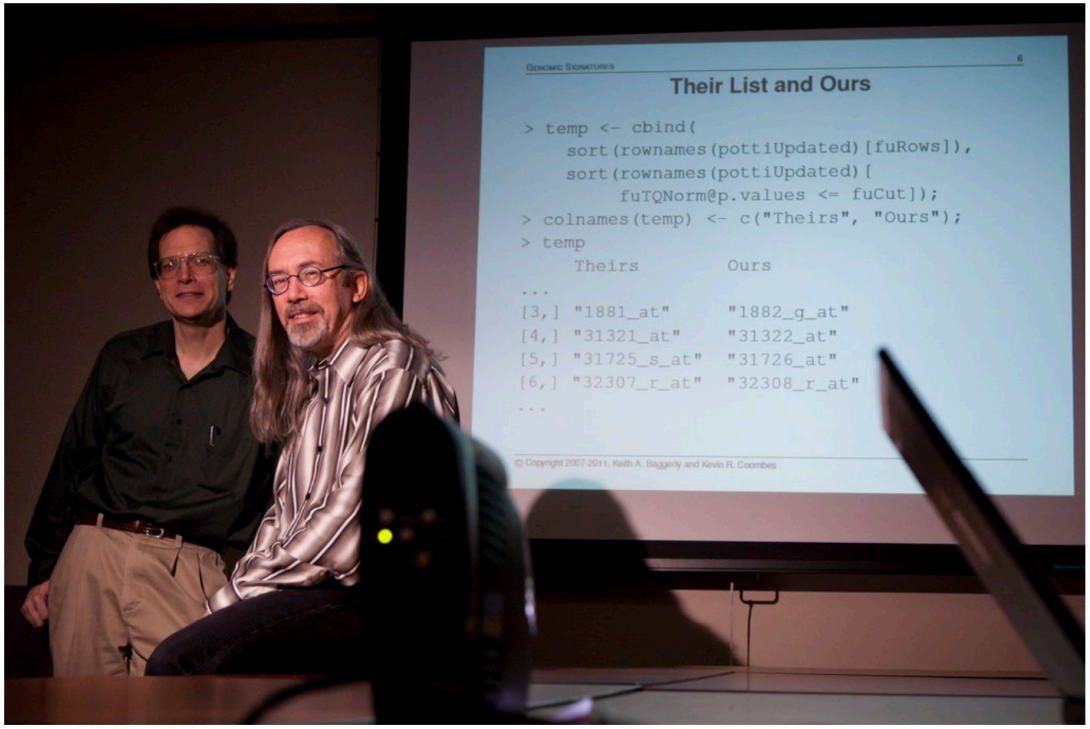
Research Assistant

Walter and Eliza Hall institute of Medical Research

The reproducibility crisis …

How Bright Promise in Cancer Testing Fell Apart

By GINA KOLATA JULY 7, 2011



Keith Baggerly, left, and Kevin Coombes, statisticians at M. D. Anderson Cancer Center, found flaws in research on tumors. Michael Stravato for The New York Times

RECENT COMMENTS

texas ta July 10, 2011
This story makes me so sad, since I work in this field as well. As a graduate student, I was witness to many flawed analyses with either...

tulipsinyard July 10, 2011
A mildly different perspective: after years teaching mathematics and statistics at universities in North America, I remain shocked by how...

Peter Melzer July 10, 2011
The widely-publicized promises of human genomics research for novel cures exert enormous pressure on the research community, paired with...

SEE ALL COMMENTS

According to recent editorials, the reproducibility crisis is still going on

nature

International weekly journal of science

Home | News & Comment | Research | Careers & Jobs | Current Issue | Archive | Audio & Video | For Au

Archive > Volume 533 > Issue 7604 > Editorial > Article

NATURE | EDITORIAL

Reality check on reproducibility

A survey of *Nature* readers revealed a high level of concern about the problem of irreproducible results. Researchers, funders and journals need to work together to make research more reliable.

25 May 2016

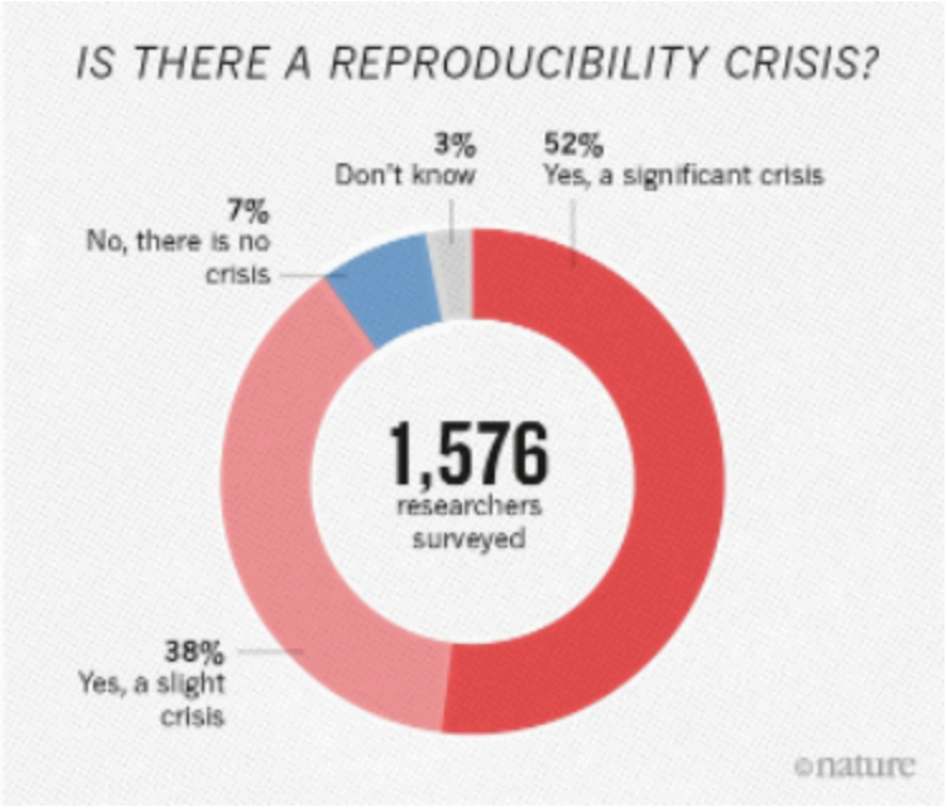
PDF | Rights & Permissions

Is there a reproducibility crisis in science? Yes, according to the readers of *Nature*. Two-thirds of researchers who responded to a survey by this journal said that current levels of reproducibility are a major problem.

The ability to reproduce experiments is at the heart of science, yet failure to do so is a routine part of research. Some amount of irreproducibility is inevitable: profound insights can start as fragile signals, and sources of variability are infinite. But, the survey

Related stories

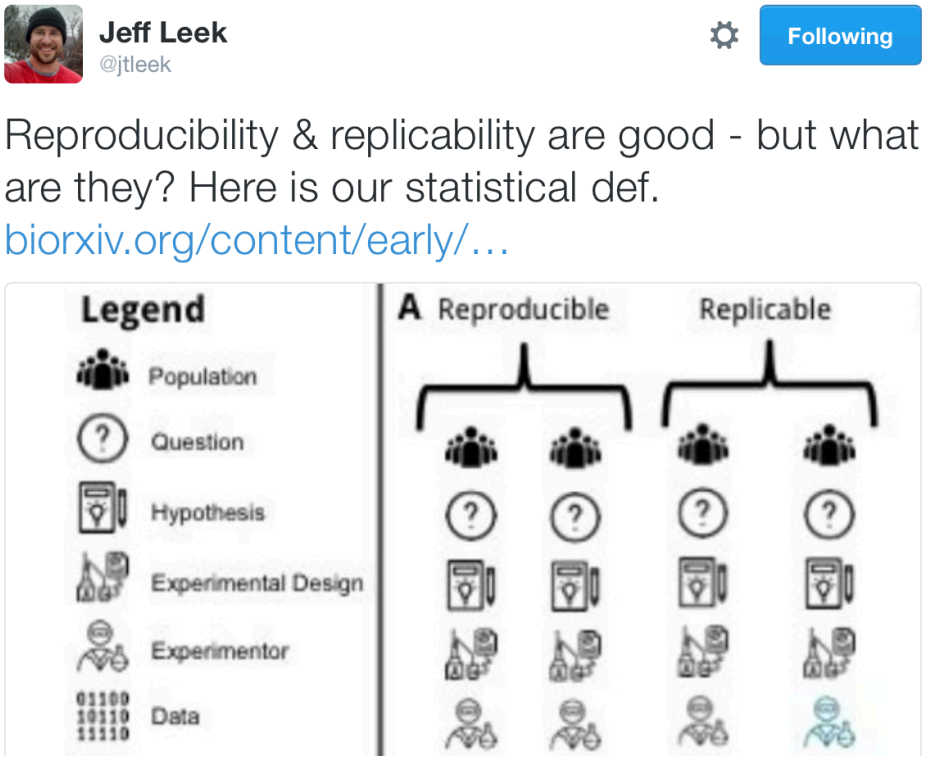
- The pressure to publish pushes down quality
- Research data: Share



Nature, May 2016

Reproducibility versus Replicability

- Conflicting/inconsistent definition for reproducibility and replicability *
- **Reproducibility** is the ability to reproduce the same results as the author(s), when an individual runs the same codes and data
- **Replicability** is the ability to get consistent results, when a different set of data is used



*http://biorxiv.org/content/early/2016/07/29/066803

Reproducible Research

- Challenges
  - Different Operating Systems (OS)
  - Different software versions/system requirements

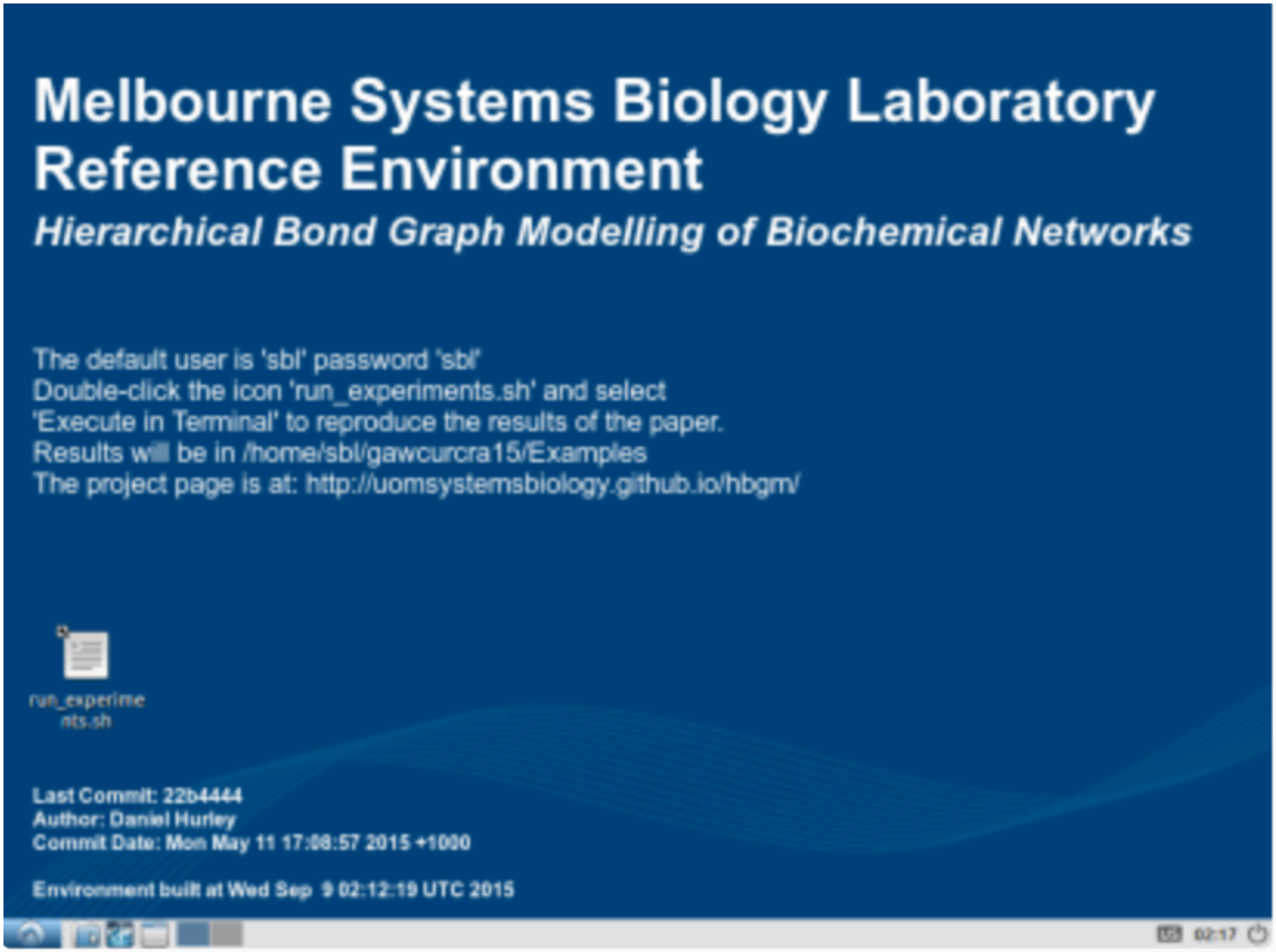
- Tools

  There exists many. Colleagues at the Systems Biology lab (University of Melbourne) introduced ***Reference Environments*,***

  a virtual machine that re-generates all study results.

Hurley, D. G., Budden, D. M., & Crampin, E. J. (2014).
*Virtual Reference Environments: a simple way to make research reproducible. *Briefings in Bioinformatics*, bbu043. doi:10.1093/bib/bbu043*

A screen capture of a Reference Environment



Docker

- Docker is very similar to a virtual machine. They both provide an environment to run a program, without necessarily having the system requirements of that program actually installed on your machine.

- An **advantage** of Docker **over Reference Environment (RE**) is that the build/run process has less software dependencies.

  - Packer and Vagrant are required to develop an RE.

    But Docker doesn't depend on any software

Docker and RStudio

- We can program a Docker image to run a (password - protected) instance of Rstudio server
- We can make the data and scripts available through this Rstudio-server instance, so that the user can interact with the code and data
- The programming of a Docker image is done through writing a `Dockerfile`

The `Dockerfile`

- To get Rstudio server running on Docker, we first need an OS system, such as **Ubuntu**.
- The required software (e.g. R, Git, wget, RStudio) are then installed on Ubuntu. This is our **'base'** docker **image**.
- We then **customise the Rstudio-server instance** on this base image.

The base **dockerfile**

```
1  FROM ubuntu:14.04
2  MAINTAINER Soroor Hediyeh Zadeh <hediyehzadeh.s@wehi.edu.au>
3  RUN echo "deb http://cran.ms.unimelb.edu.au/bin/linux/ubuntu trusty/" >> /etc/apt/sources.list
4  RUN apt-key adv --keyserver keyserver.ubuntu.com --recv-keys E084DAB9
5  RUN apt-get update
6  RUN apt-get install -y software-properties-common libxml2-dev
7  RUN add-apt-repository ppa:marutter/rdev
8  RUN apt-get update
9  RUN apt-get upgrade -y
10 RUN apt-get install -y -q r-base r-base-dev gdebi-core libapparmor1 supervisor wget git
11 RUN (wget https://download2.rstudio.org/rstudio-server-0.99.902-amd64.deb && gdebi -n rstudio-server-0.99.902-amd64.deb)
12 RUN rm /rstudio-server-0.99.902-amd64.deb
13 RUN (adduser --disabled-password --gecos "" davislab && echo "davislab:davislab"|chpasswd)
14 RUN mkdir -p /var/log/supervisor
15 ADD supervisord.conf /etc/supervisor/conf.d/supervisord.conf
16 RUN chown -R davislab:davislab /home/davislab
17 RUN chmod 700 /home/davislab
18 EXPOSE 8787
19 CMD ["/usr/bin/supervisord"]
```

Import Ubuntu

Add the link to download R to the path

Install software required to install R on Ubuntu

Install R, Git, wget

Download Rstudio server

Set up a user account named davislab, set the password

Copy some scripts related to user account set up into the image

Allocate port 8787 to connect to Rstudio server

Execute the scripts added for user account setup

The template is available on GitHub

Main `Dockerfile`

```
1  FROM davislaboratory/docker-rstudio-server
2  MAINTAINER soroorh <hediyehzadeh.s@wehi.edu.au>
3  RUN git clone https://soroorh@bitbucket.org/soroorh/mforoutan_tgfb_paper_2016.git
4  RUN ln -s /mforoutan_tgfb_paper_2016/output /home/davislab/output
5  RUN ln -s /mforoutan_tgfb_paper_2016/scripts /home/davislab/scripts
6  RUN ln -s /mforoutan_tgfb_paper_2016/data /home/davislab/data
7  RUN chown -R davislab:davislab /home/davislab/output/figures
8  RUN chmod 700 /home/davislab/output/figures
9  RUN chown -R davislab:davislab /home/davislab/output/results
10 RUN chmod 700 /home/davislab/output/results
11 RUN chown -R davislab:davislab /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
12 RUN chmod 700 /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
13 RUN git clone https://soroorh@bitbucket.org/soroorh/mforoutan_paper_rdata.git
14 RUN mv /mforoutan_paper_rdata/randRanks_up.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
15 RUN mv /mforoutan_paper_rdata/CI99_up.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
16 RUN mv /mforoutan_paper_rdata/CI99_down.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
17 RUN mv /mforoutan_paper_rdata/down_permute_dens.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
18 RUN mv /mforoutan_paper_rdata/up_permute_dens.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
19 RUN git clone https://soroorh@bitbucket.org/soroorh/mforoutan_paper_rdata_2.git
20 RUN mv /mforoutan_paper_rdata_2/randRanks_down.RData /home/davislab/data/comparative_analysis/probe_gene_mapping/out_10data_check
21 RUN chown -R davislab:davislab /home/davislab/data/integrative_analysis/out_10data_check
22 RUN chmod 700 /home/davislab/data/integrative_analysis/out_10data_check
23 RUN (Rscript -e 'install.packages(c("dplyr","hexbin","colorRamps","survival","XML","ggplot2"), repos="http://cran.rstudio.com/")')
24 RUN (Rscript -e 'source("http://bioconductor.org/biocLite.R"); biocLite(c("limma","GSVA", "sva"))')
25 WORKDIR /mforoutan_tgfb_paper_2016
26 RUN mv generate_all_experiments.R /home/davislab
```

Import the customized Ubuntu- the base image

Download data from git repository

Make symbolic links (pointers) to /data, / scripts, /output folders

Set write permissions to davislab user account

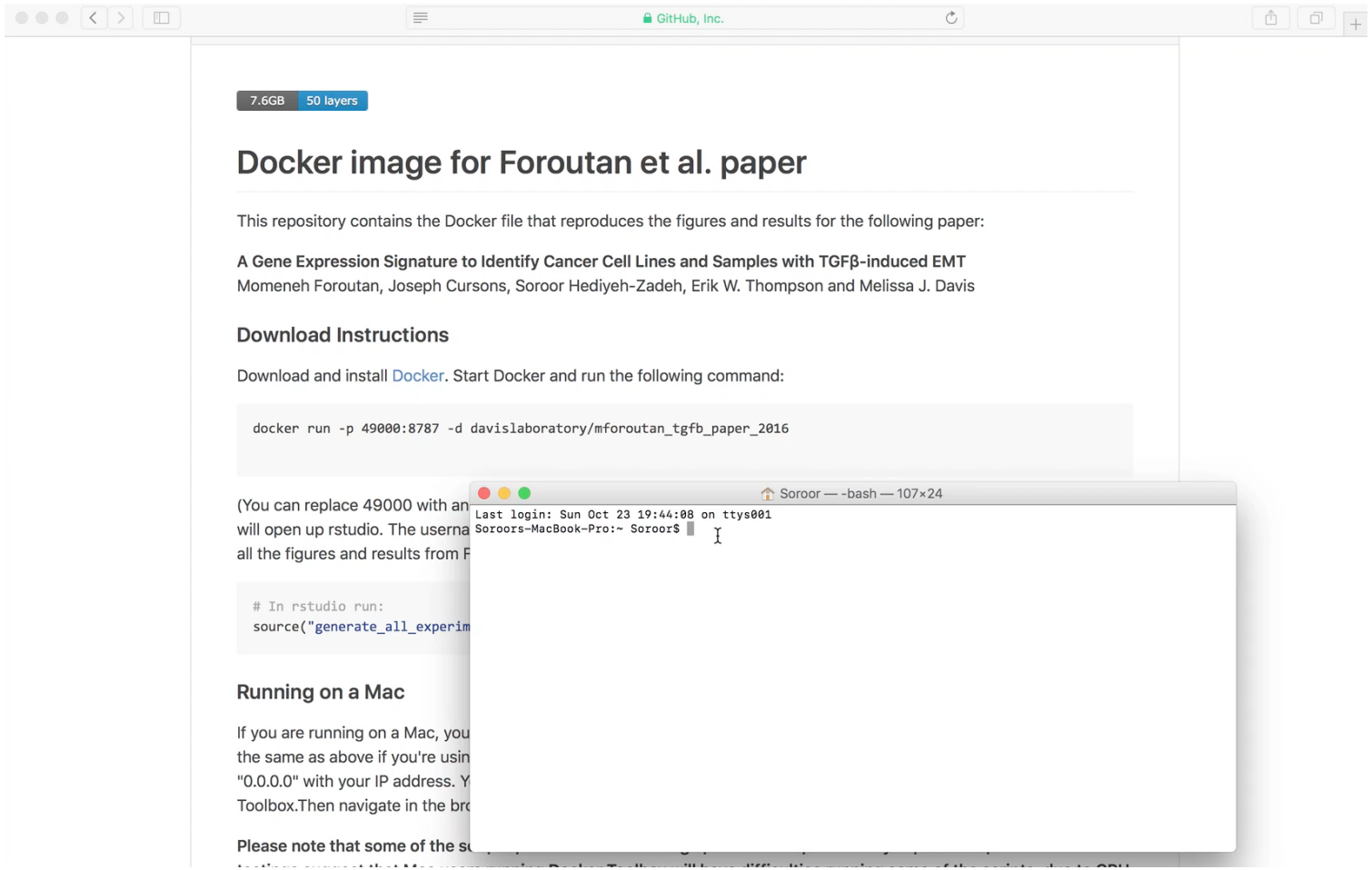
Add R packages that the analysis use

Download bioconductor packages

Change the working directory

Copy the main script to user ‘davislab’

Docker Demo


GitHub, Inc.
7.6GB
50 layers
Docker image for Foroutan et al. paper
This repository contains the Docker file that reproduces the figures and results for the following paper:
A Gene Expression Signature to Identify Cancer Cell Lines and Samples with TGFβ-induced EMT
Momeneh Foroutan, Joseph Cursons, Soroor Hediyeh-Zadeh, Erik W. Thompson and Melissa J. Davis
Download Instructions
Download and install Docker. Start Docker and run the following command:
docker run -p 49000:8787 -d davislaboratory/mforoutan_tgfb_paper_2016
(You can replace 49000 with an
will open up rstudio. The userna
all the figures and results from F
In rstudio run:
source("generate_all_experim
Running on a Mac
If you are running on a Mac, you
the same as above if you're usin
"0.0.0.0" with your IP address. Y
Toolbox.Then navigate in the br
Please note that some of the su
Soroor — -bash — 107×24
Last login: Sun Oct 23 19:44:08 on ttys001
Soroors-MacBook-Pro:~ Soroor$