

DockerBIO user manual

Web application for efficient use of bioinformatics tools

Docker Pull Command

Docker name : netbuyer/dockerbio

Download : docker pull netbuyer/dockerbio

June 15, 2018

Version 0.1

Introducing projects

Bioinformatics tools are being developed and tested on various operating systems, with various program languages. Biologists and bioinformaticians want to build pipelines using a variety of tools, but it is not easy to compile, install and use them. DockerBIO is designed for biologists to easily use various bioinformatics tools from DockerHub, and easily make pipelines using these tools.

By default, DockerBIO has most of essential tools for NGS analysis such as Whole Genome analysis pipeline, RNA-Seq analysis pipeline, FastQC(quality check tool for fastq files), SnpShift(vcf annotation tool), Samtools, BWA(alignment tools), GATK(vcf calling tool), etc. However, users can register and use any tools and data on DockerBIO. For more information, please refer to Kwon et al., “DockerBIO: Web application for efficient use of bioinformatics Docker images”, under review in CMPB, 2018.

DockerBIO is available on Linux(both Centos and Ubuntu) and Mac OS. Windows are not fully supported due to hyper-v issues.

DockerBIO is a web program based on java spring created with two spring boots modules.

- SpringbootAsyn cJob - Asynchronous Module for Docker Implementation Monitoring
- SpringbootFileupload - Module to upload user files

Part1 – For Easy Installation

1. Install docker at your PC, workstation or server (<https://www.docker.com/>)
2. Enter following commands:

```
$ docker pull netbuyer/dockerbio
$ mkdir /data/dockerbio
```

3. (optional) You can download and unzip files containing the Reference DB / Snp DB files from <https://goo.gl/iivMWf>
 - 1) dbAll.tar.gz : a compressed version of all versions of dbSNP
 - 2) refDbAll.tar.gz : a reference databases containing hg19 and hg38
 - 3) dockerbio_extra.tar.gz : a compressed version of both Reference hg38 and dbSNP150(hg38 ver)

You can download one or more of these files.
Downloaded files should be at /data/dockerbio

4. Supposing you choose dockerbio_extra.tar.gz, you can unzip it through following command:

```
$ tar xvfz dockerbio_extra.tar.gz
```

The folder must be created in the form shown below.

```
/data/dockerbio/
.....upload/ (upload folder)
.....db/ (Snp db folder)
.....refDb/ (reference db folder)
.....sample/ (sample folder)
.....user/ (user file folder)
.....temp/ (temporary upload folder)
```

5. To run dockerBIO, enter following command:

```
$docker run -dit -h dockerbio --name dockerbio --privileged -v /data/dockerbio:/data/dockerbio -v /var/run/docker.sock:/var/run/docker.sock -p 8080:8080 -p 8090:8090 -p 8092:8092 -e HOSTIP=http://Service_Host_IP netbuyer/dockerbio bash
```

Service_Host_IP should be replaced with your (public) ip (example : 192.168.24.24).

```
$ docker exec dockerbio /opt/restart.sh
```

6. In the browser address bar, type `http://Service_Host_IP:8080`
7. Disk capacity requirements vary depending on how you use them.
 - Using dockers at the same time increases the capacity of temporary folders in the system.
 - The location of Docker's temporary folder is `/var/lib/docker`.
 - If the analysis file has a large capacity or a large number, resize the temporary upload folder
 - The location of the temporary upload folder is `/data/docker/dockerbio/temp`.

Part2 - For Easy Use

At the first time, you can log in as ID: `user@user.com` and Password: `User1234`. After first login, changing password is strongly recommended. You can make new account on the system.

After you successfully login into your DockerBIO page, you will be able to see "Register Docker" at which you can register new Docker and "Run Docker" menu at which you can execute the registered Docker from "Register Docker".

◆ Run Docker

First, you can drag and drop files you want to analyze into the UPLOAD USER FILE menu, and then check the files in the "user file Select" menu. Next, you need to execute the desired program. You can run the dockers registered in "Job Select" menu. For whole genome analysis, the user can select `_WGS` (netbuyer / wgs) in Job Select menu and select need files from user file Select menu and click the Run Docker menu.

Most options can be changed as needed, but we recommend to use the default options tested in "Register Docker" and change the version of REF DB or dbSNP as needed. If you need to modify it, you can create a new option in the Register Docker menu. We will show examples of the most commonly used tools, BWA, FastQC, SnpSift, WGS pipeline, and WTS pipeline.

- How to run FastQC tool

- 1) Job Select → FASTQC(conradstoerker/fastqc)
- 2) user file Select → chr1_1.1000.fq (**Please be sure to select only one file**)
- 3) Run Docker → Click "Run Docker" button

other option : Please use the default option.

Tags → latest (The latest version of docker),

Output Type → DIR (Select **DIR**ectory because there are several files such as the `*_fastqc.zip`, `*_fastqc.html` file. When you create it in the Register Docker, you must select DIR if multiple files are created.)

If you are an expert in fastqc, you can change the Options freely, but I do not recommend it.

- How to run BWA tool

- 1) Job Select → bwa mem(comics/bwa)
- 2) REF DB Select → hg38.fasta (Please select a desired Reference Database)
- 3) user file Select → First, select chr1_1.1000.fq, push down +(Plus button) and then chr1_2.1000.fq (**Please be sure to select two files**)
- 4) Run Docker → Click “Run Docker” button

other option : Please use the default option.

Tags → latest (The latest version of docker),

Output Type → LOG (Select **LOG** to view and download the resulting file with the log files..)

Output Ext → sam (Extension of result file generated after running bwa tool)

If you are an expert in bwa tool, you can change the Options freely, but I do not recommend it.

- How to run SnpSift tool

- 1) Job Select → SnpSift annotate(alexcope/snpsift)
- 2) dbSNP Select → dbsnp.150.hg38.vcf.gz (Please select a desired dbSNP version)
- 3) user file Select → test.1KG.vcf (You must select a vcf file for annotation)
- 4) Run Docker → Click “Run Docker” button

other option : Please use the default option.

Tags → latest (The latest version of docker),

Output Type → LOG (Select **LOG** to view and download the resulting file with the log files..)

Output Ext → annotate.vcf (Result file generated after running SnpSift annotate tool)

If you are an expert in SnpSift tool, you can change the Options freely, but I do not recommend it.

- How to run WGS pipeline tool

- 1) Job Select → `_WGS(netbuyer/wgs)`
- 2) REF DB Select → `hg38.fasta`
- 3) dbSNP Select → `dbsnp.150.hg38.vcf.gz` (Please select a desired dbSNP version)
- 4) user file Select → First, select `chr1_1.1000.fq`, push down +(Plus button) and then `chr1_2.1000.fq` (**Please be sure to select two files**)
- 5) Run Docker → Click “Run Docker” button

other option : Please use the default option.

Tags → 0.4 (The latest version of docker),

Output Type → DIR (Select **DIR**ectory because there are several files such as the raw_snps.vcf, filtered_snps.vcf file. When you create it in the Register Docker, you must select DIR if multiple files are created.)

If you are an expert in WGS pipeline, you can change the Options freely, but I do not recommend it.

- How to run RNA-Seq(WTS) pipeline tool

- 1) Job Select → `_RNA_SEQ(netbuyer/rna_seq)`
- 2) REF DB Select → `RNA_hg38/genome`
- 3) dbSNP Select → `RNA_hg38.annotated.gtf` (Please select a desired annotation file for RNA_Seq analysis)
- 4) user file Select → First, select `chr1_1.1000.fq`, push down +(Plus button) and then `chr1_2.1000.fq` (**Please be sure to select two files**)
- 5) Run Docker → Click “Run Docker” button

other option : Please use the default option.

Tags → latest (The latest version of docker),

Output Type → DIR (Select **DIR**ectory because there are several files such as the merged.annotated.gtf, merged.stats and merged.tracking files. When you create it in the Register Docker, you must select DIR if multiple files are created.)

If you are an expert in WTS pipeline, you can change the Options freely, but I do not recommend it.

- How to run other tools

Please click the “Run Docker” by selecting the input file (user file Select) referring to the above tools.

◆ Register Docker

This page functions to download a number of bioinformatics tools registered in dockerhub and register with the DockerBIO. This page consists of two menus: Docker LIST and Docker Info Register. Docker LIST is a menu that tests whether the registered option is edited and works. The Docker Info Register searches the docker in dockerhub and downloads it to the current system. It then functions to set various options.

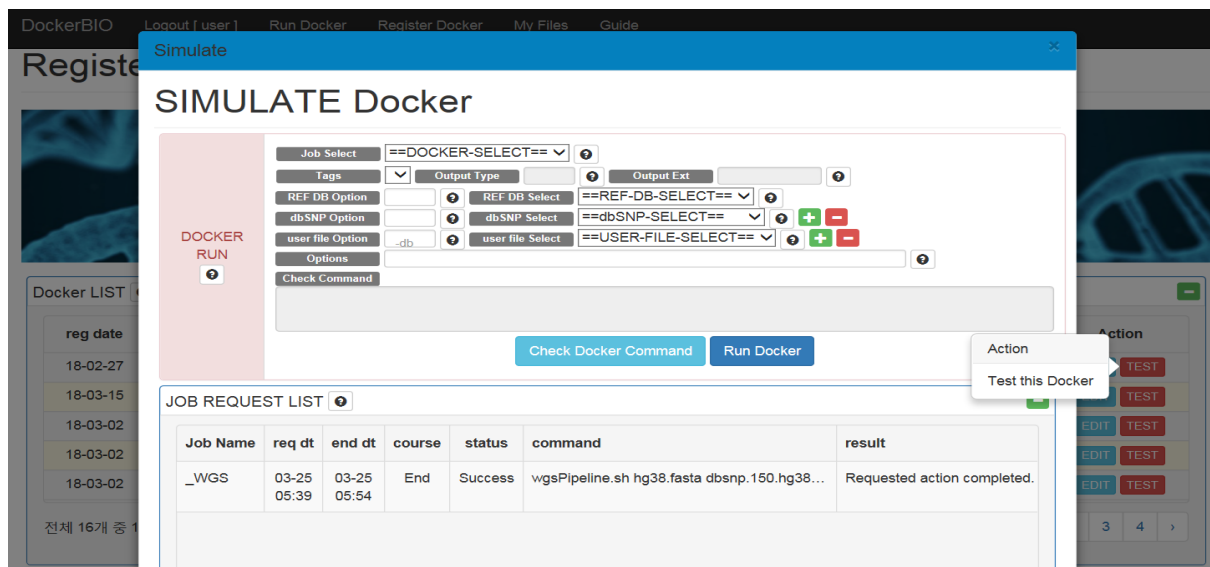
- Docker LIST

If you click the EDIT button in the action menu, you can see the menu set in the Docker Info Register and modify it if necessary. After setting the options, you need to click the TEST button to set the best option in the SIMULATE Docker window. You can test whether the docker is working using your options and see it in the Run Docker. If you want to change the current Run Docker option, you can modify it here and use the Update option.

reg date	Job Name	Docker ID	whole param	ref	db	Action
18-02-27	_WGS	netbuyer/wgs	wgsPipeline.sh [REFDB_OPT] [DB_OPT] [U...	1	1	EDIT TEST
18-03-15	_RNA_SEQ	netbuyer/rna_seq	wtsPipeline.sh [REFDB_OPT] [DB_OPT] [U...	1	1	EDIT TEST
18-03-02	GATK HaplotypeCaller	netbuyer/wgs	java -jar /tools/GenomeAnalysisTK.jar -T Ha...	1	N	EDIT TEST
18-03-02	GATK PrintReads	netbuyer/wgs	java -jar /tools/GenomeAnalysisTK.jar -T Pri...	1	N	EDIT TEST
18-03-02	GATK BaseRecalibrator post recal	netbuyer/wgs	java -jar /tools/GenomeAnalysisTK.jar -T Ba...	1	1	EDIT TEST

전체 16개 중 1~5번째 데이터 출력, 페이지 당 5 개 데이터 출력

1 2 3 4



● Docker Info Register

Please click Search-Docker-ID to open the docker search window and search for the desired search term. If you search for bwa keyword, more than 300 bwa registered in dockerhub are searched. After selecting a docker with a large number of downloads or a large number of stars, click the select button to register it in the Docker ID.

Type the Docker Name, and set the output TYPE. There are three types of LOG, FILE, and DIR. LOG selects the log file of the result file if desired, FILE selects when the result is created as a file, and DIR selects the log file Select to save the resulting file.

The success condition allows you to choose between file size and log file. If it fails, select log file if there is a fail in the log file, and select file size if the file size becomes smaller. Most tools are available with log and file size.

Reference DB is a menu for selecting Reference, and now only one DB can be selected.

dbSNP is a menu to select the SNP database, and you can specify several.

The Options menu can be set by typing additional options when you need to type them.

When all options are set, you can check the command with Check Docker Command.

If you click the Create button, a new job will be registered in the Docker List and you can run the test.

If you follow the whole procedure through the manual, you will find a lot of tools registered in dockerhub. You will be able to use it easily after registering and testing without difficulty of installation. But you need to learn to test basic tools yourself.