DockerBIO user manual

Web application for efficient use of bioinformatics tools

Docker Pull Command

Docker name   : netbuyer/dockerbio
Download1   : docker pull netbuyer/dockerbio

Download2   : docker pull netbuyer/dockerbio:0.3

* Download1 includes only DockerBIO

* Download2 includes a reference(hg38), dbSNP and DockerBIO.

Oct 10, 2018

Version 0.3

Introducing projects

Bioinformatics tools are being developed with various program languages and tested on various operating systems.. Biologists and bioinfomaticians want to build pipelines using a variety of tools, but it is not easy to compile, install and incorporate them into pipelines. 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 the 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".

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

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

- SpringbootAsyn cJob - Asynchronous Module for Docker Implementation Monitoring

- SpringbootFileupload - Module to upload user files

Part 1 – For Easy Installation

1. Install Docker at your PC, workstation, or server (https://www.docker.com/)

2. Enter the 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 following repositories, and place them at /data/dockerbio.

Google drive: https://drive.google.com/drive/folders/16jkIWSJbFYmrZ0GjJrMDq7trW714JYvI
Sharefile: https://mygenomeboxmygenomebox.sharefile.com/share/view/sdcce93756ad45978

   In above repositories, you can download one or more of following files:

   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)

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

       $ tar xvzf dockerbio_extra.tar.gz

   After unzipping, the folders will be created in the form shown below.

       /data/dockerbio/
       ....upload/ (upload folder)
       ....db/ (Snp db folder)
       ....refDb/ (refrence db folder)

...........................sample/ (sample folder)
...........................user/ (user file folder)
...........................temp/ (temporary upload folder)

5. To run dockerBIO, enter the 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:0.1 bash

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

$ docker exec dockerbio /opt/restart.sh

7. In the browser address bar, type http://Service_Host_IP:8080

Part 2 - For Easy Use

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

After you successfully login to your DockerBIO page, you will be able to see "RegisterDocker" 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, 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.

Second, execute the desired program. You can run the dockers registered in the "Job Select" menu. For whole genome analysis, select _WGS (netbuyer/wgs) in the Job Select menu and select needed files from the user file Select menu, and then click the Run Docker menu.

Most options can be changed as needed, but we recommend using the default options tested in "Register Docker" and changing 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 the FastQC tool

1) Job Select ➔ FASTQC(conradstoerker/fastqc)

2) user file Select ➔ chr1_1.1000.fq (**Please be sure to select only one file;** if you need two files, click the +(plus button).)

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 *_fastqc.zip and *_fastqc.html files. When you create a file in Register Docker, you must select DIR if multiple files are created.)

If you are an expert in fastqc, you can change the Options freely.

If you want to use option -k (kmers) or -t (threads), type the desired option between fastqc and [USER_OPT]. If you click the "Check Docker Command" button, you can see your full command set.

Example 1: If you use option -k 2 : fastqc **-k 2** [USER_OPT] -o [RESULT_OPT].

Example 2: To use option -k 2 and -t 4 at the same time : fastqc **-k 2 -t 4** [USER_OPT] -o [RESULT_OPT].

- How to run the 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 with the bwa tool, you can change the Options freely.

If you want to use option -B (penalty for a mismatch) or -t (threads), type the desired option between "bwa mem" and "-R" command. If you click the "Check Docker Command" button, you can see your full command set. If you have a problem, familiarize yourself with the original command and follow it.

Example 1: if you use option -B 5 : bwa mem **-B 5** -R "@RG\\tID:Id1\\tSM:Sample1\\tPL:ILLUMINA" ~.

Example 2: Using option -t 4 : bwa mem **-t 4** -R "@RG\\tID:Id1\\tSM:Sample1\\tPL:ILLUMINA" ~.

- How to run the SnpSift tool

1) Job Select   ➔ SnpSift annotate(alexcoppe/snpsift)

2) dbSNP Select   ➔ dbsnp.150.hg38.vcf.gz (Please select the desired dbSNP version.)

3) user file Select   ➔ test.1KG.vcf (You must select a vcf file for annotation.)

4) Run Docker   ➔ Click the "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 the SnpSift annotate tool)

If you are an expert with the SnpSift tool, you can change the Options freely.

- How to run the 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 the 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 the "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 and filtered_snps.vcf files. When you create a file in Register Docker, you must select DIR if multiple files are created.)

If you are an expert with WGS pipeline, you can change the Options freely.

If you want to use option -t (Number of threads) or -k (Minimum seed length), type the desired option between "wgsPipeline.sh" and "hg38.fasta" command. If you click the "Check Docker Command" button, you can see your full command set. If you have a problem, familiarize yourself with the original command and follow it.

Example 1: If you use addition option -t 6 : wgsPipeline.sh **–t 6** hg38.fasta   dbsnp.150.hg38.vcf.gz s2_r1.fastq   /resultFolder.

- How to run the 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 the 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 a file in Register Docker, you must select DIR if multiple files are created.)

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

If you want to use option -5 (Trim $<int>$ bases from 5" (left) end of each read before alignment) and -3 (Trim $<int>$ bases from 3" (right) end of each read before alignment), type the desired option between the "wtsPipeline.sh" and "RNA_hg38/genome" command. If you click the "Check Docker Command" button, you can see your full command set. If you have a problem, familiarize yourself with the original command and follow it.

Example 1: If you use addition option -5 4 : wtsPipeline.sh **-5 4** RNA_hg38/genome RNA_hg38.annotated.gtf s2_r1.fastq s2_r1.fastq　/resultFolder.

- How to run the samtools tool

1) Job Select　　　　　➔ samtools view (biocontainers/samtools)

2) REF DB Select　　　➔ hg38.fasta

3) user file Select ➔ chr1_1.1000.sam (You must select a sam file)

4) Run Docker ➔ Click the "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 ➔ chr1_1.1000.bam (Result file generated after running samtools view tool)

If you are an expert with the samtools tool, you can change the Options freely and you can type the desired option between "samtools view" and "-Sb" command.

Example : If you use the addition option –**C**(CRAM format) : samtools view **-C** -Sb hg38.fasta chr1_1.1000.sam > /resultFolder/1.1000.bam

◆ **Register Docker**

This page provides functions for downloading a number of bioinformatics tools registered in DockerHub and registering them with DockerBIO. This page consists of two menus: Docker LIST and Docker Info Register. Docker LIST is a menu to test whether the registered options   works. The Docker Info Register can be used to search for Docker images in DockerHub and download them to the current system. It also provides functions for setting various options.
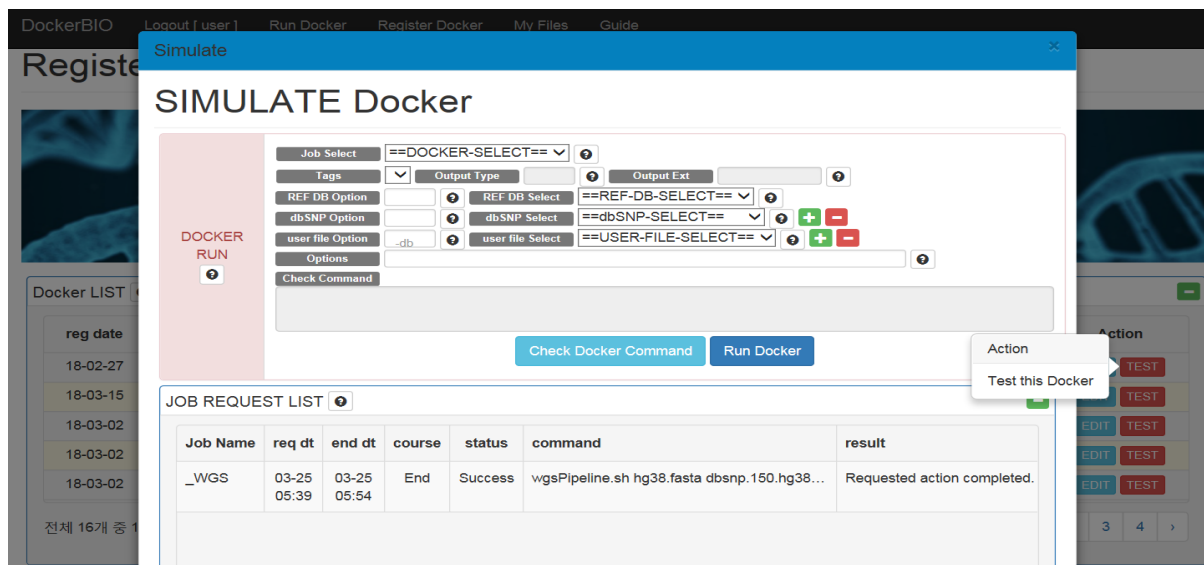
● Docker LIST

If you click the EDIT button in the action menu, you can see the menus in the Docker Info Register and modify options 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 Run Docker. If you want to change the current Run Docker options, you can modify them here.

- 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 keywords, more than 300 bwa entries registered in DockerHub are searched. After selecting a docker (if you are not sure which one to choose, it will be good to choose one with a large number of downloads or a large number of stars), click the select button to register it in Docker ID.

Next, you can set Docker Name, and output TYPE. There are three types: LOG, FILE, and DIR. LOG selects the log file of the result file if desired, FILE is used when the result should be used to create a file, and DIR selects "log file Select" to save the resulting file.

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

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 in 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; however, you will need to learn how to test basic tools yourself.

- How to add (register) the bwa tool

1) Search-Docker-ID ➜ Search for the tool you want to add and select the check button for the desired item.

2) Docker Name ➜ It generally has the same name as the Docker ID, but can be changed.

3) Publish ➜ The tool you want to run in the Run Docker must be changed to Y (yes). We recommend changing to the Y option only after running many tests with SIMULATE.

4) Reference DB ➜ After selecting 1, select hg38.fasta in RefDB.

5) dbSNP ➜ N (Don"t select)

6) Options   ➔ Change the "command" keyword to "bwa" or "bwa mem" keyword.

7) Run Docker   ➔ After checking your command with the Check Docker Command, press the Create button to register it in Run Docker. It is recommended to distribute it as Run Docker only after performing many simulations using the TEST function, rather than sending it to Run Docker at once.

* other options : Please use the default option.

Tags   ➔ latest (The latest version of docker).

output TYPE   ➔ FILE (Select **FILE** to view and download the resulting file with the bam files).

output Ext   ➔ bam (Convert sam to bam using the samtools view tool.)

success condition   ➔ In the success condition, select "file size" if you want the file to be created above a certain size, or select "log file" if you want to keep the log.

If you are an expert with this tool, you can freely change the Options on the Run Docker page.

- How to add (register) the fastqc tool

1) Search-Docker-ID ➔ Search for the "fastqc" keyword and select the check button for the desired item.

2) Docker Name   ➔ It generally has the same name as the Docker ID, but can be changed.

3) Publish   ➔ The tool you want to run in Run Docker must be changed to Y (yes). We recommend changing to the Y option only after running many tests with SIMULATE.

4) Reference DB   ➔ N (Don"t select).

5) dbSNP   ➔ N (Don"t select)

6) Options   ➔ Change the "command" keyword to "tophat2" keyword.

7) Run Docker   ➔ After checking your command with the Check Docker Command, press the Create button to register it in Run Docker. It is recommended to distribute it as Run Docker only after performing many simulations using the TEST function (SIMULATE) rather than sending it to Run Docker at once.

* other options : 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 *_fastqc.zip and *_fastqc.html files. You must select DIR if multiple files are created).

success condition ➔ In the success condition, select "file size" if you want the file to be created above a certain size, or select "log file" if you want to keep the log.

If you click the "Check Docker Command" button, you can see your full command set. If you are an expert with this tool, you can freely change the Options on Run Docker page.

If you want to use option -k (kmers) or -t (threads), type the desired option between fastqc and userFile in Options command.

Example 1: If you use option -k 2 : fastqc **-k 2** userFile /resultFolder.

Example 2: To use option -k 2 and -t 4 at the same time : fastqc **-k 2 -t 4** userFile /resultFolder.

Docker Info Register ❓

Docker ID	maxulysse/fastqc 🔍 Search-Docker-ID 🔗 Docker Link ❓
Docker Name	maxulysse/fastqc ❓ Tags latest ▾ Publish Y ▾ ❓
output TYPE	DIR ▾ ❓ success condition file size ▾ ❓
Reference DB	N ▾ ❓ ➕
dbSNP	N ▾ ❓ ➕
Options	fastqc [USER_OPT] [RESULT_OPT] ❓
Check Command	

fastqc userFile /resultFolder

Check Docker Command   Create   Update   Delete