

MegaBOLT 生信分析加速器

高级用户手册

地 址： 中国深圳市盐田区北山工业区综合楼及11栋2楼
电 话： 4000-966-988
邮 箱： MGI-service@mgi-tech.com
网 址： www.mgi-tech.com

仅供科研使用

深圳华大智造科技股份有限公司

版本
4.0

关于本手册

本手册适用于 MegaBOLT 生信分析加速器（MegaBOLT），手册版本 4.0，软件版本 V2.3.0.0。

本手册及其包含的信息为深圳华大智造科技股份有限公司（以下简称深圳华大智造）的专有保密信息，未经深圳华大智造的书面许可，任何个人或组织不得全部或部分地对本手册进行重印、复制、修改、传播或公布给他人。本手册的读者为终端用户，其作为设备的一部分，由深圳华大智造授权终端用户予以使用。严禁未授权的个人使用本手册。

深圳华大智造对本手册不做任何种类的保证，包括（但不限于）用于特定目的的商业性和合理性的隐含保证。深圳华大智造已经采取措施，确保本手册的准确性。但是，深圳华大智造对错误或遗漏不承担责任，并保留任何对本手册和设备进行改进以提高其可靠性、功能或设计的权利。

本手册中的所有图片均为示意图，图片内容可能与实物有细微差异，请以购买的设备为准。

文中涉及的名称及商标属于各自所有者资产。

©2019-2020 深圳华大智造科技股份有限公司 版权所有。

发布日期：2021 年 7 月 1 日

制造商信息

生产厂家	深圳华大智造科技股份有限公司
生产地址	中国深圳市盐田区北山工业区综合楼及 11 栋 2 楼
技术支持厂家	深圳华大智造科技股份有限公司
技术支持联系电话	4000-966-988
技术支持联系邮箱	MGI-service@mgi-tech.com

版本记录

	修订日期	版本
修订	2021 年 7 月 1 日	4.0
	2020 年 12 月 30 日	3.0
	2020 年 7 月 17 日	2.0
编制	2019 年 11 月 30 日	1.0

目录

概述

1	概述.....	1
---	---------	---

深入了解 MegaBOLT

2	MegaBOLT 存储结构与目录结构.....	3
	系统存储结构	3
	MegaBOLT 软件目录结构	3
	Reference 目录结构和说明	4
	Bed 文件.....	4
3	MegaBOLT 子模块	5
	程序与工具	5
	开源软件源代码	6
4	MegaBOLT Server 与 Client.....	7
	Server 与 Client 简介	7
	MegaBOLT 系统服务	7
5	MegaBOLT 流量.....	9
	查询 MegaBOLT 流量.....	9
6	MegaBOLT 日志.....	10
	日志结构.....	10
	日志信息等级说明	11

常见 Linux 操作指引

7	移动存储设备挂载	13
	图形化界面挂载存储.....	13
	命令行方式挂载存储.....	13
8	网络配置	16
	打开网络配置选项	16
	配置网络连接	17
	用户验证.....	18
	网络自动连接	19
9	RAID5	20
	RAID5 阵列介绍	20
	RAID5 Resync Checking	20
	数据安全与冗余备份.....	21
10	存储设备加密	22
	数据盘加密与解密	22
	密码.....	22
	免责声明.....	24
附录		
	FAQ	26

概述

1

概述

本文档面向已熟悉 MegaBOLT 基本使用方法，需要进一步获取 MegaBOLT 详细信息和特性的用户。

首先，本文档介绍 MegaBOLT 存储结构、软件目录结构、ZLIMS 分析结果目录结构，MegaBOLT 的子模块构成，默认 Reference 与分析 WES 数据时使用的 bed 区间文件的相关信息，MegaBOLT 日志系统，以使用户深入了解 MegaBOLT，指导用户更方便、更灵活地使用 MegaBOLT 进行多样化数据分析需求与下游分析。同时，本文档介绍了 MegaBOLT 的 Server-Client 结构与 MegaBOLT 系统服务，指导用户通过系统服务查询与控制 MegaBOLT 的运行状态。

其次，文档还向用户介绍 linux 系统的常见操作指引，包括移动存储设备的挂载与网络设置。

最后，针对用户使用 MegaBOLT 产生的常见问题与错误，给出快速的解决方法。

深入了解 MegaBOLT

2 MegaBOLT 存储结构与目录结构

系统存储结构

MegaBOLT 配有两块 SSD 硬盘，分别用于存储操作系统和 MegaBOLT 软件。除此之外，MegaBOLT 工作站还配有一个大容量数据存储（Sata HDD），硬盘挂载目录以及功能如下表所示：

硬盘	挂载目录	功能
SSD	/	操作系统
SSD	/mnt/ssd	MegaBOLT 软件
Sata HDD	/data	用户数据存储

MegaBOLT 软件目录结构

MegaBOLT 软件根目录为 /mnt/ssd/MegaBOLT/，目录结构如下表所示：

目录	存放内容
bin	可执行文件
config	配置文件
docs	用户文档
env	运行环境依赖文件
lib	库文件
lic	Licence 文件
log	日志文件
test	测试案例
tmpDir	临时文件与中间文件
tools	第三方软件工具

Reference 目录结构和说明

MegaBOLT 提供三组 reference，分别是 hg19（默认）、hg38 以及 hs37d5。存储路径为 `/mnt/ssd/MegaBOLT/reference/`。其中主要文件及其说明如下表所示：

文件	说明
hg19.*	hg19 的 reference 文件及其索引文件
dbsnp_151.vcf.*	hg19 的配套 dbsnp 文件
1000G_phase1.*	knownSites 文件，可根据文件名中的 hg19、hg38 标签判断其配套 reference
Mills_and_1000G_gold_standard.indels.*	knownSites 文件，可根据文件名中的 hg19、hg38 标签判断其配套 reference
hg38.*	hg38 的 reference 文件及其索引文件
hs37d5/	hs37d5 的 reference 文件及其索引文件，配套 dbsnp 和 knownSites 文件
ReadMe	介绍 reference 及其配套文件的下载链接
build.sh	自动构建所有 reference 的索引文件的脚本程序
db	用于分析 WES 数据的 bed 文件

 **注意** hg19,hg38 及其配套文件均从 NCBI 下载。

Bed 文件

用于分析 WES 数据的 bed 文件路径为：

`/mnt/ssd/MegaBOLT/reference/db/db_BED/`

MegaBOLT 的 `--bed` 参数支持的 bed 文件简称与 bed 文件对应关系列表，可通过如下文件查看：

`/mnt/ssd/MegaBOLT/reference/db/db_BED/db.list`

3

MegaBOLT 子模块

程序与工具

MegaBOLT 的可执行文件存放路径为: `/mnt/ssd/MegaBOLT/bin/`

该目录存放了可供 MegaBOLT 调用的子程序, 用户也可以单独调用这些子程序以进行定制化分析流程开发, 其中主要的程序如下表所示:

程序	说明
MegaBOLT	MegaBOLT client 程序, 用于投递任务
MegaBOLT_server	MegaBOLT server 程序, 用于运行由 client 投递的任务
hc4HaplotypeCaller.jar	Germline 变异检测程序
BQSR.jar	碱基质量值重校正程序
gatk4_mutect2_acc.jar	Somatic 变异检测程序
bwa	BWA 比对程序
mm2Alignment	Minimap2 比对程序
picard.jar	Picard 程序
samtools	Samtools 程序
Sort	排序程序
SortMarkDup	排序标重程序

除了上述 MegaBOLT 子程序之外, MegaBOLT 中还包含了部分第三方应用程序, 为了统一管理这些程序, 将第三程序安装到如下路径:

`/mnt/ssd/MegaBOLT/tools/`

该目录下的所有软件工具如下表所示:

目录	说明
GATK	GATK 软件, 包含 GATK3.8 与 GATK4.0
biobambam2	Biobambam2 工具集
DeepVariant	基于深度学习的变异检测程序
DNAPipe	生信分析相关工具集

开源软件源代码

MegaBOLT 的部分子程序是在开源软件的基础上，由 MegaBOLT 开发团队进行二次开发、性能优化以及准确性优化而得，这部分子程序的原始开源代码存放于如下路径：

`/mnt/ssd/MegaBOLT/tools/src/`

4

MegaBOLT Server 与 Client

Server 与 Client 简介

MegaBOLT 采用 Server-Client 结构设计，server 程序在后台常驻，当 client 投递任务，server 将任务解析后统一执行。

MegaBOLT server 由 MegaBOLT 系统服务监控，该服务随操作系统启动，并自动启动 MegaBOLT server，用户投递任务时无须关注 server 程序的运行状态。

MegaBOLT 系统服务

MegaBOLT 系统服务用于监控 server 程序，以保证 server 程序正常、稳健地运作。用户可以通过 MegaBOLT 系统服务查询 server 运行状态，启动、关闭和重启 server 程序。



注意

MegaBOLT server 是 MegaBOLT 系统正常运行的必要条件。除非您明确知晓如下操作的后果，否则请在指导下进行操作。

1. 查询 server 运行状态

执行如下命令，查询 server 运行状态：

```
service megabolt status
```

输出结果示例：

```
● megabolt.service - SYSV: start/stop/restart/reload
  daemon of MegaBOLT server
   Loaded: loaded (/etc/rc.d/init.d/megabolt; bad;
  vendor preset: disabled)
   Active: active (running) since Mon 2019-11-04
  09:58:40 CST; 7h ago
     Docs: man:systemd-sysv-generator(8)
  Memory: 63.9G
    CGroup: /system.slice/megabolt.service
            └─ 1972 /usr/bin/megabolt
               └─ 2024 /mnt/ssd/MegaBOLT_scheduler/bin/
MegaBOLT_server -restart
```

从输出结果中的“active (running)”可以看出，当前 server 正在运行中。

2. 停止 server

执行如下命令，关闭 server:

```
sudo service megabolt stop
```

输入 SU 密码之后，输出如下:

```
Stopping megabolt (via systemctl): [ OK ]
```

MegaBOLT 系统服务和 server 都已经停止，此时投递 MegaBOLT 任务将无法执行。

3. 启动 server

执行如下命令，启动 MegaBOLT 系统服务和 server:

```
sudo service megabolt start
```

输入 SU 密码之后，输出如下:

```
Starting megabolt (via systemctl): [ OK ]
```

MegaBOLT 系统服务和 server 都已经启动。

4. 重启 server

当修改了 MegaBOLT 配置文件时，需要重启 server 以生效更改，执行如下命令，重启 MegaBOLT 系统服务和 server:

```
sudo service megabolt restart
```

输入 SU 密码之后，输出如下:

```
Restarting megabolt (via systemctl): [ OK ]
```

MegaBOLT 系统服务和 server 都已经重启。

5

MegaBOLT 流量

查询 MegaBOLT 流量

MegaBOLT 流量可以通过终端命令查询，打开终端后，执行如下命令：

```
MegaBOLT --traffic
```

输出结果样例如下：

```
TrafficStatistics:
```

```
Total Traffic : 100500000000000 bp
```

```
Used Traffic : 4998200 bp
```

```
Left Traffic : 100499995001800 bp.
```

字段释义：

Total Traffic: 总流量，即出厂时预装的流量

Used Traffic: 已使用的流量

Left Traffic: 当前剩余可用流量

6

MegaBOLT 日志

日志结构

MegaBOLT 运行时产生三种日志，分别是 **server** 日志、**client** 日志和任务日志。三种日志分别记录三种不同的运行信息，分等级进行记录，帮助用户快速查看任务信息或定位错误。

Server 日志

Server 日志分为四个，所在目录为 `/mnt/ssd/MegaBOLT_scheduler/log`。这四个日志的作用如下表所示：

日志文件	说明
<code>megabolt_server.err</code>	server 程序标准输出流 / 标准错误流信息
<code>megabolt_server.log</code>	server 程序运行时日志
<code>megabolt_rocket.log</code>	加密锁流量日志
<code>megabolt_task.log</code>	server 任务日志

Client 日志

Client 日志有两个，分别是 `megabolt.log` 和 `megabolt.out`，存放于 MegaBOLT 分析任务的输出目录下，这两个日志的作用如下表所示：

日志文件	说明
<code>megabolt.out</code>	client 程序标准输出流 / 标准错误流信息
<code>megabolt.log</code>	client 程序运行时日志

任务日志

MegaBOLT 以一个样本作为一个分析任务，对应样本 list 文件中的一行，每一个样本的分析结果和任务日志都会输出到以样本名命名的目录下。任务日志有两个，均以样本名命名，当样本名没有被指定并且 `--outputprefix` 参数未被指定时，将以“output”命名，这两个日志分别是 `samplename.log` 和 `samplename.out`，它们的作用如下表所示：

日志文件	说明
<code>samplename.log</code>	当前样本分析任务日志
<code>samplename.out</code>	处理当前样本分析任务的子程序运行时日志

日志信息等级说明

MegaBOLT 日志信息根据严重程度划分不同的等级，越严重的问题产生的日志记录等级越高，下面按严重等级从低到高介绍 MegaBOLT 日志信息等级及含义：

日志等级	说明
debug	调试信息，默认不开启调试模式，不输出
info	普通日志信息
warning	警告信息
error	错误信息，当前任务失败时产生，但不影响其它任务的执行
critical	致命错误，MegaBOLT server 或 client 程序异常退出

常见 Linux 操作指引

7

移动存储设备挂载

MegaBOLT 服务器采用 CentOS7 操作系统，硬盘的挂载方式与 Windows 操作系统不同，下面按照图形化界面和命令行两种方式分别介绍，CentOS 移动存储设备挂载。

图形化界面挂载存储

CentOS 原生支持 VFAT、FAT32、NTFS 格式的移动存储设备，只需将设备插入任意 USB 接口，CentOS 将自动识别和自动挂载存储到操作系统中，并在桌面生成插入设备的快捷访问方式，如下图所示：

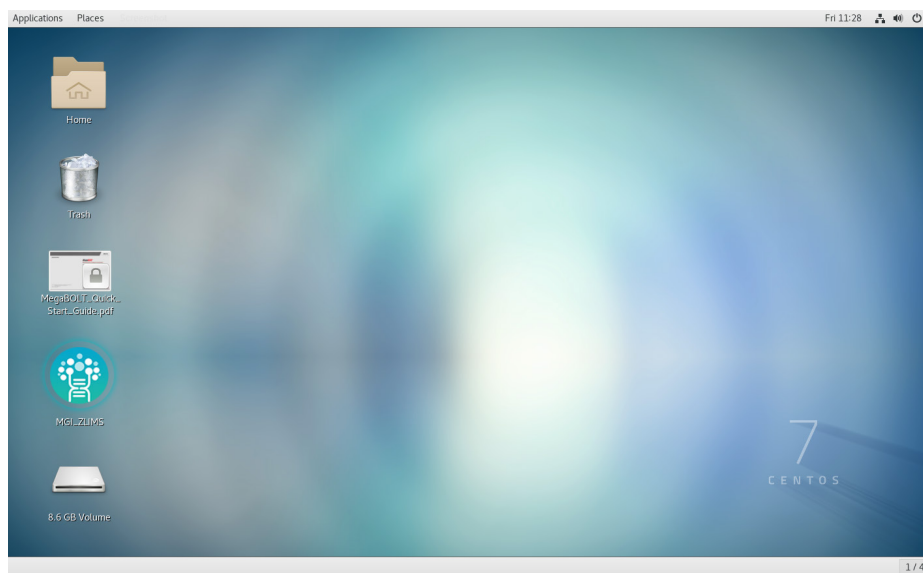


图 1 CentOS7磁盘自动挂载

而 exFAT、ext4 格式的存储设备无法被 CentOS 直接识别，需要使用命令行方式挂载。

命令行方式挂载存储

在非图形化界面以及需要挂载 exFAT，ext4，格式移动存储设备时，则需要使用命令行方式挂载存储。下面以挂载一个容量约为 16GB 的 U 盘为例，分步骤介绍命令行挂载存储设备。

1. 查看磁盘列表

执行如下命令，查看当前服务器的存储设备列表：

```
sudo fdisk -l
```

命令输入结果示例：

```
.....
Disk /dev/nvme1n1: 1024.2 GB, 1024209543168 bytes,
2000409264 sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x659fab7d

Device Boot      Start          End      Blocks   Id
System
/dev/nvme1n1p1    2048    2000409263   1000203608
83  Linux

Disk /dev/sda: 15.1 GB, 12501054772 bytes, 24025497
sectors
Units = sectors of 1 * 512 = 512 bytes
Sector size (logical/physical): 512 bytes / 512 bytes
I/O size (minimum/optimal): 512 bytes / 512 bytes
Disk label type: dos
Disk identifier: 0x4f8af734

Device Boot      Start          End      Blocks   Id
System
/dev/sda1        16779264    240254975   111737856    7
HPFS/NTFS/exFAT
.....
```

从输出结果中获取到 U 盘的设备名为 “/dev/sda”，文件系统格式为 exFAT，并且只有一个硬盘分区 “/dev/sda1”。

2. 挂载磁盘到操作系统

如果磁盘采用 exFAT 文件系统，使用如下命令挂载：

```
mkdir /mnt/usb
mount.exfat /dev/sda1 /mnt/usb
```

如果磁盘采用 NTFS 文件系统，使用如下命令挂载：

```
mkdir /mnt/usb
mount.ntfs /dev/sda1 /mnt/usb
```

如果磁盘采用 FAT32、ext4 以及 VFAT 文件系统，使用如下命令挂载：

```
mkdir /mnt/usb  
mount /dev/sda1 /mnt/usb
```

执行上述命令后，通过 `/mnt/usb` 目录就可以使用已挂载磁盘的内容了。

当前磁盘挂载方式作出的修改是临时的，操作系统重新启动之后，挂载将失效。

3. 卸载磁盘

当磁盘使用完毕，应先卸载磁盘，再移除移动存储设备，以避免造成存储设备数据损坏。执行如下命令，卸载磁盘：

```
umount /mnt/usb
```

8

网络配置

MegaBOLT 服务器采用 CentOS7 操作系统，可按照本引导完成网络设置。

打开网络配置选项

1. 点击桌面右上角的喇叭图标，从下拉选项中选择【Ethernet】，再选择【Wired Settings】。

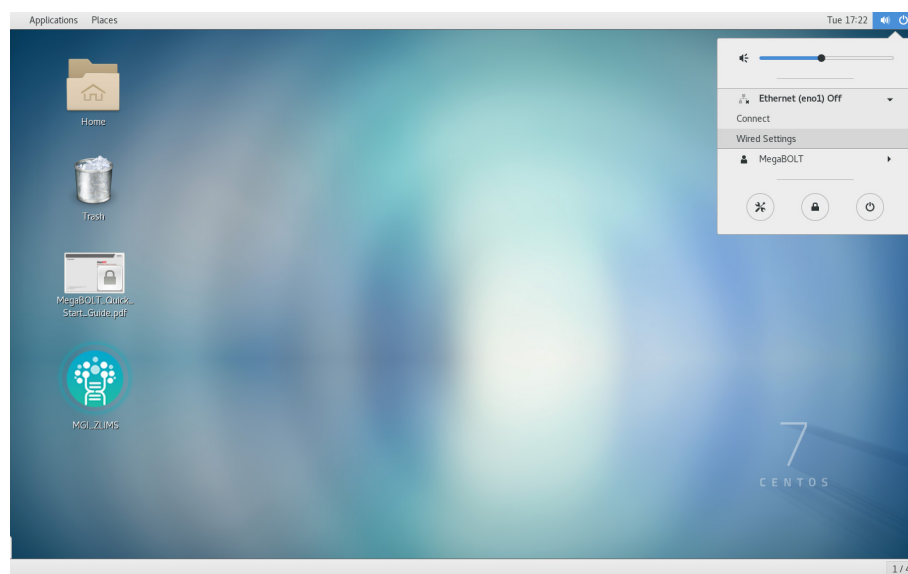


图 2 CentOS7快捷方式

2. 在 Wired Settings 窗口中，窗口左侧展示了当前设备可用的以太网端口，接通网线的端口将在右侧显示网络带宽信息，并且开关处于可操作状态。

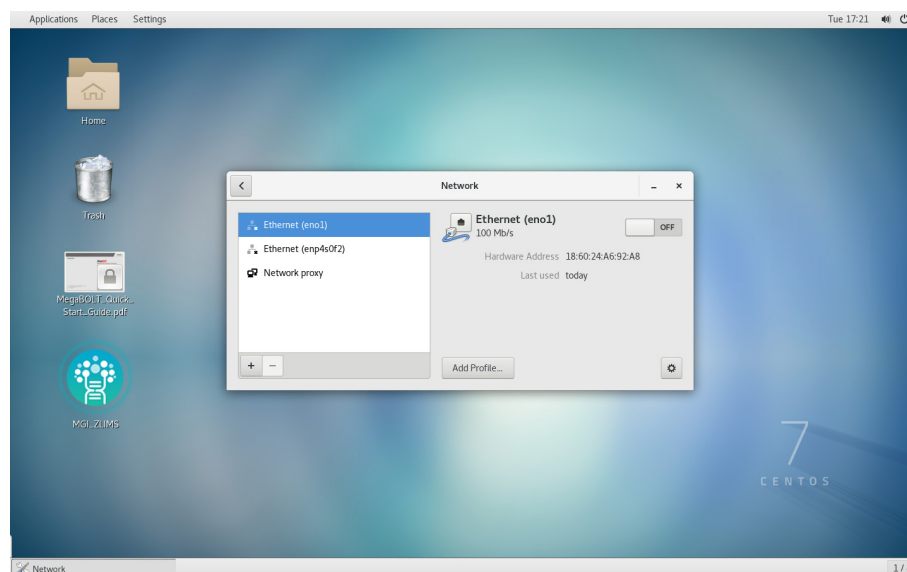


图 3 Network设置

配置网络连接

1. 动态网络连接

如果目标网络支持 DHCP 服务，打开点击网络连接开关，稍等片刻即可自动连接成功。

2. 静态网络连接

如果用户需要配置静态网络，可先通过网络设置中的 IPv4 进行设置，将【Addresses】选项设置为【Manual】，然后将目标网络静态 IP 填入【Address】，子网掩码填入【Netmask】，网关填入【Gateway】，应用修改后，在图 4 所示的界面中，再打开网络连接开关即可生效。

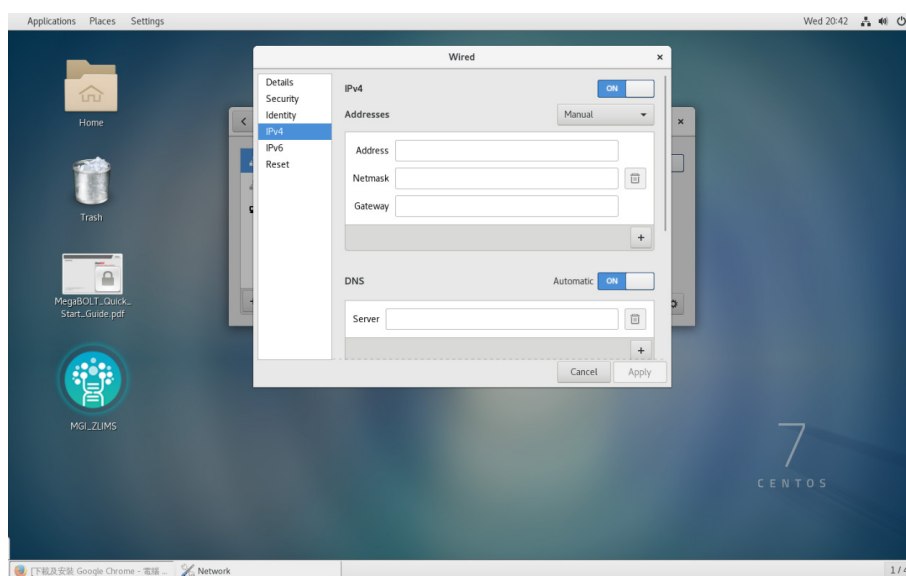


图 4 静态网络设置

用户验证

如果目标网络需要用户验证，在图 4 所示【Network 设置】界面中，选择右下方的齿轮图标，进入网络安全设置。

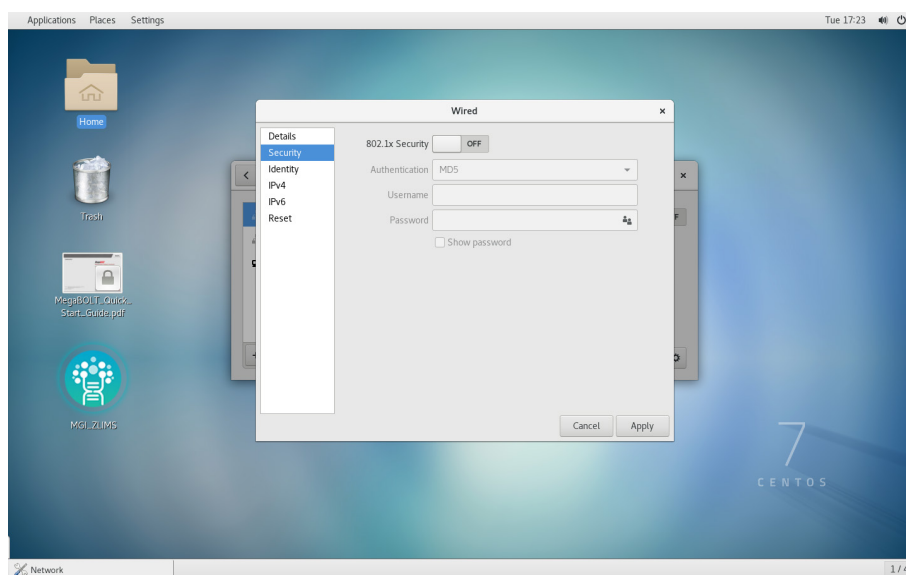


图 5 网络用户验证

继续选择【Security】选项，在窗口右侧选择打开 Security 开关，并根据目标网络用户验证方式，选择在【Authentication】中对应的用户验证方式，输入用户名与密码后，点击右下方的【Apply】应用修改，最后在 Wired Settings 窗口右侧，打开网络连接开关，稍等片刻即可连接成功。

网络自动连接

CentOS 7 操作系统支持开机自动连接网络，完成网络用户验证之后，再次进入进入网络安全设置，点击窗口左侧的【Identity】，勾选窗口下方的【Connect Automatically】选项，应用修改后，再次开机时即可自动连接网络。

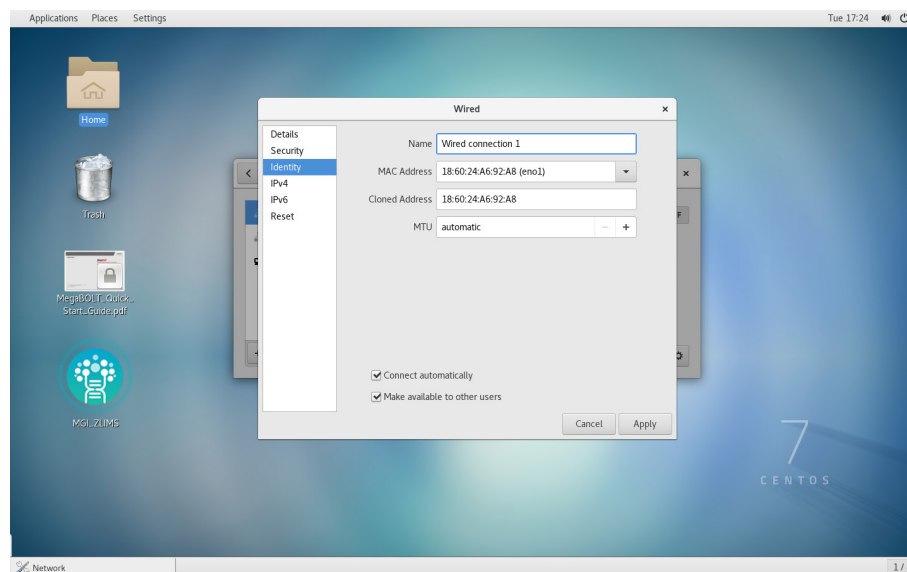


图 6 自动连接选项

9

RAID5

RAID5 阵列介绍

为兼顾存储性能、数据安全和存储成本，MegaBOLT 工作站采用 RAID5 对 5 块容量为 6TB 的散列机械硬盘进行 RAID5 阵列构建，阵列构建后，可用容量约为 21TB。RAID5 阵列挂载点目录为 /data。

RAID5 Resync Checking

RAID5 阵列在正常工作时，会定时执行同步检查，时间为每年的 1、4、7、10 四个月的 1 号凌晨 1 点执行。同时，如果发生了意外断电情况，例如强制关机，意外断电，也会执行同步检查。每次同步检查将持续 12 小时，在同步检查过程中，RAID5 阵列的读写操作会出现高延迟、低 IOPS 的现象，这是正常的，等检查结束后即可恢复正常速度。

当 RAID5 阵列在进行同步检查时，恰好需要使用本机进行紧急的分析任务，如果需要避免高延迟、低 IOPS 对分析时间造成影响，则可以使用如下终端命令暂停同步检查：

```
sudo echo frozen > /sys/block/md126/md/sync_action  
sudo echo idle > /sys/block/md126/md/sync_action
```

✦ **注意** 当出现如下任意一种情况时，禁止手动停止同步检查。

- (1) 同步检查是由意外断电引起的，如强制重启、停电
- (2) 同步检查是由更换硬盘，修复掉盘引起的
- (3) 新建阵列时
- (4) 当使用 mdadm -D 命令查询阵列，发现 Resync Status 为正在进行中
- (5) 当使用 mdadm -D 命令查询阵列，发现阵列 State 为 Resyncing

当紧急分析任务完毕，服务器空闲时，必须执行如下命令，继续进行同步检查以完成数据完整性验证，保证数据安全：

```
sudo /usr/sbin/raid-check &
```

数据安全与冗余备份

MegaBOLT 工作站使用 Intel® VROC 技术构建 RAID5 阵列，该阵列的稳定性与主板寿命、硬盘寿命直接相关，两者任意一个损坏都有可能导致数据损坏或丢失，因此用户必须定期进行重要业务数据、数据库、文档文稿、应用数据的冗余备份，保证数据安全性，时间间隔可由用户业务自行决定，例如一个月，三个月等。

10 存储设备加密

数据盘加密与解密

MegaBOLT 工作站采用 CentOS 系统自带 dm-crypt 组件对数据盘进行加密，加密格式为 LUKS（Linux Unified Key Setup，Linux 统一密钥管理）。LUKS 作为 Linux 硬盘加密的标准被广泛的应用到各 Linux 发行版中，当加密设置完成后，如果一个非法用户持有该加密硬盘，那么他将无法读写这块硬盘，除非他知道对应的密钥。

密码

MegaBOLT 工作站加密的存储设备为大容量存储，用于用户数据临时存储，该设备采用 RAID5 对多块硬盘进行组织，详情请参阅 XXX 章节。加密设备的初始密码为：`megabolt`

为了保证用户数据安全，请在首次进入 MegaBOLT 操作系统后，修改默认密码。请执行如下步骤。

- 1) 按如下路径，打开 Linux Disks 工具
桌面 → Applications → Utilities → Disks
- 2) 选中目标加密硬盘，点击设置图标，并在弹出的选框中选择【Change Passphrase】

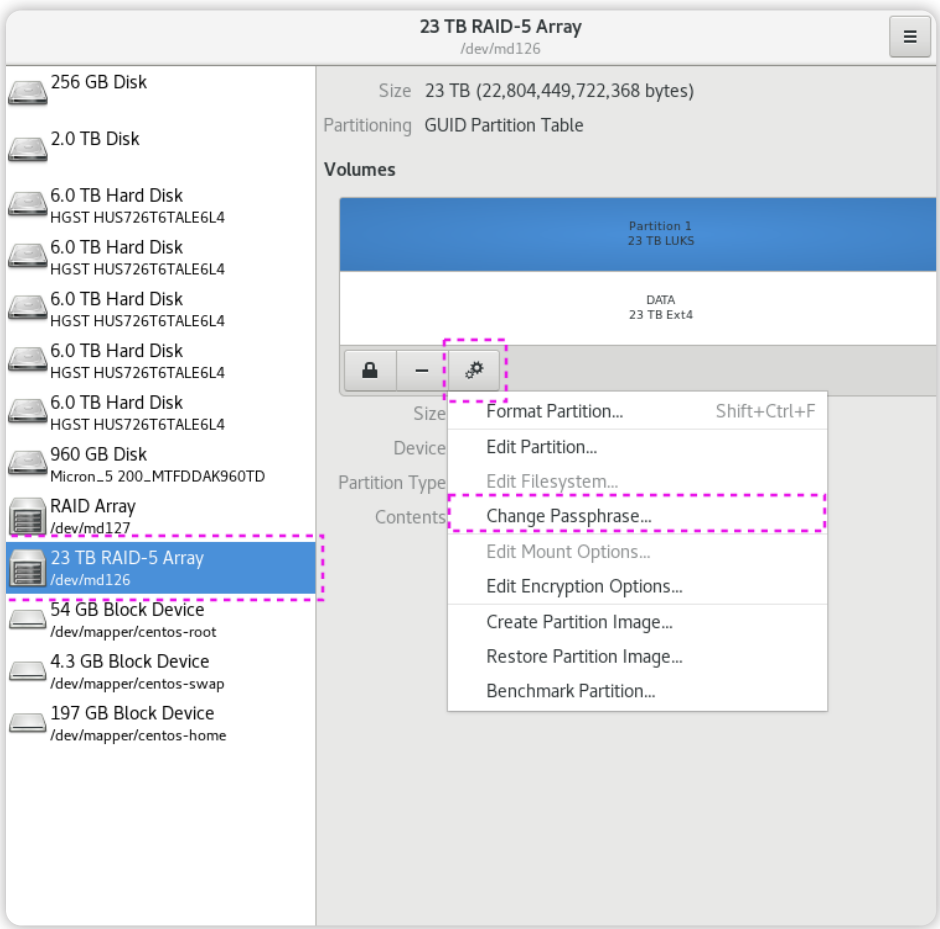
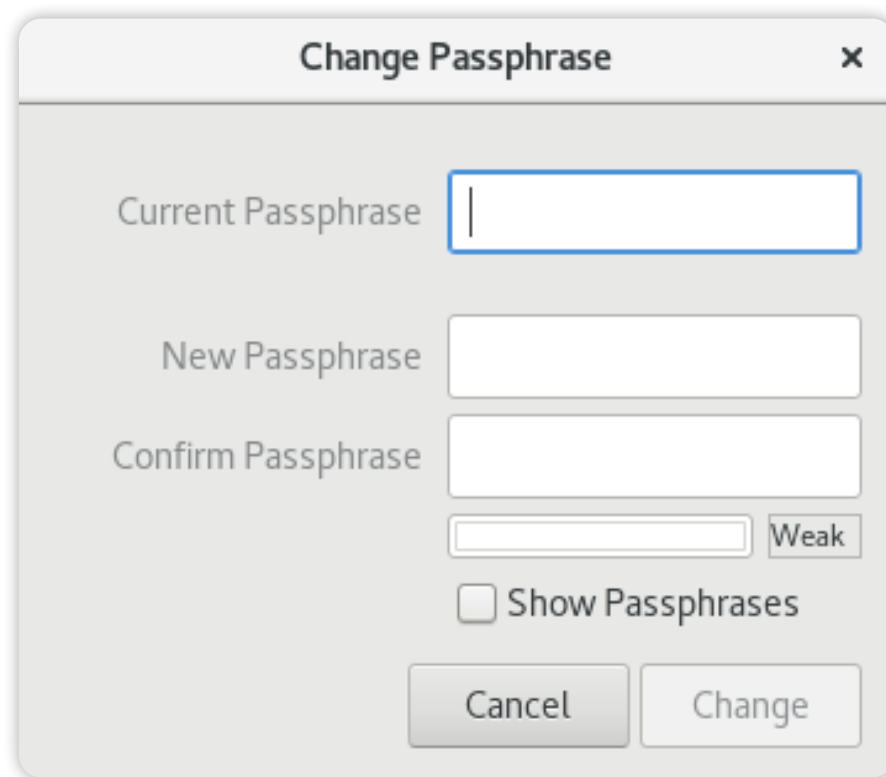


图 7 Change PWD 1

- 3) 输入自定义密码后，可点击【Show Passphrases】复选框检查密码，确认无误后点击右下角的【Change】图标完成修改。

A macOS-style dialog box titled "Change Passphrase" with a close button (X) in the top right corner. It contains three text input fields: "Current Passphrase", "New Passphrase", and "Confirm Passphrase". Below the "Confirm Passphrase" field is a strength indicator bar and a "Weak" label. At the bottom, there is a checkbox labeled "Show Passphrases" and two buttons: "Cancel" and "Change".

Change Passphrase

Current Passphrase

New Passphrase

Confirm Passphrase

Weak

☐ Show Passphrases

Cancel Change

图 8 Change PWD 2



警告

请妥善保管自定义密码，并在多个管理员之间共享，一旦密码丢失，则加密存储设备中的所有数据文件将无法恢复，如果需要使用硬盘，则需要全盘格式化后才能使用。

免责声明

存储设备的数据安全需要华大智造与客户共同维护。华大智造作为 MegaBOLT 产品的制造方，为产品本身的安全提供支持服务；客户作为使用 MegaBOLT 去处理、存储基因数据及对应的数据主体的其他个人数据的控制者或处理者，应考虑处理可能给数据主体带来的风险，并采取适当的组织措施和技术措施保护数据主体信息资产的安全。

附录

FAQ

机器相关

- 1 Q 机器上有大容量存储吗？

 A MegaBOLT 机器工作站版默认包含 30 TB 数据存储。
- 2 Q 临时计算存储大小能够分析多大数据量？

 A 按照“SSD 大小 /5”，来计算最大能高效分析的数据量。超过也能分析，只是速度变慢。如：2 TB SSD 可分析 400 GB 以内的数据。
- 3 Q 两个任务能同时运行么？

 A 可以，MegaBOLT 通过任务调度系统调度任务运行。
- 4 Q 可以在机器上运行其他程序吗？

 A 不建议，但可以在不影响 MegaBOLT 运行情况下增加其他应用。
- 5 Q 可以在运行 MegaBOLT 时同时运行其他程序吗？

 A 不建议，运行 MegaBOLT 时会占用大部分硬件资源，除非您充分悉知这么做带来的风险。同时运行其它程序可能会引起 MegaBOLT 分析任务失败，甚至服务器宕机。
- 6 Q 可以在 MegaBOLT 缓存硬盘下存储文件吗？

 A 不建议，运行 MegaBOLT 时会产生大量中间文件，需要有充足的缓存硬盘空间。
- 7 Q 可以提供 root 权限吗

 A MegaBOLT 用户含有管理员权限，使用命令时在命令前加上 `sudo`，输入 MegaBOLT 用户密码就能以管理员方式执行命令。

- 8 Q 可以升级操作系统吗?
- A 不建议。MegaBOLT 在当前操作系统版本获得最好的性能与稳定性，更新系统需要重新编译驱动，如需更新，请联系当地销售代表或制造商。

流程相关

- 9 Q 当前版本支持不等长 Read 么？支持 Reads 长度是多少？
- A 当前版本支持不等长 Reads，Reads 长度 25 bp~500 bp。
- 10 Q 当前版本支持 SE 数据么？
- A 当前版本支持 SE 数据 Reads。
- 11 Q 当前版本是否支持多个 Read 文件输入
- A 支持 list 文件 Reads 输入方法，详情查看 MegaBOLT 用户手册。
- 12 Q 如何生成 gVcf 结果
- A 设置参数 `--ERC GVCF`。
- 13 Q PCR-free 数据如何设置
- A HaplotypeCaller 3.8 和 4.0 可设置参数 `--pcr-indel-model NONE`。DeepVariant 可设置参数 `--WGS_mode PCR-free`。
- 14 Q 可以通过设置多个 `--vcf` 参数输入 vcf 文件（dbSNP）吗？
- A 目前不支持，如果用户输入多个 vcf 文件，则只会使用最后一个 vcf 文件。
- 15 Q HaplotypeCaller 能同时输入多个 bam 文件作为 input 么？
- A 当前版本不能输入多个 bam 文件。
- 16 Q 如何使用 DeepVariant 进行变异检测
- A 设置参数 `--deepvariant 1`。
- 17 Q 如何运行 WES 流程

A 设置参数 `--runtype WES`。

18 Q 如何分析 Single End 数据

A 设置参数 `--se 1`。

19 Q 当前版本 basic 流程是否有 bam 和 vcf 的统计信息

A MegaBOLT 当前版本 basic 步骤暂时没有 bam 和 vcf 的统计信息，如果需要统计信息，请运行 full 流程。

20 Q 如何快速运行其他物种的分析

A 使用 `--ref ref.fa` 与 `--vcf db.vcf` 参数及模式运行。

21 Q 单独运行子模块时都是输出 `output.*` 的结果，可以指定其他名字么？

A 可以设置 `--outputprefix` 进行输出名字的设置。

22 Q 输入的文件或者目录一定要使用绝对路径吗？

A 我们推荐用户使用绝对路径。

23 Q 流程支持 CombineGVCFs 和 GenotypeGVCFs 么？

A 流程暂时不支持 CombineGVCFs。但支持 GenotypeGVCFs，可用于 GVCF 转换成 VCF。

24 Q 支持如下几种分析吗？（CNV/SV/annotation/RNA/metagenomics/denovo genome assembly）

A 默认版本暂不支持。

25 Q MegaBOLT 一次最多可以投递几个分析任务？

A 当前版本不限制单次投递任务数，用户设置在 sample list 文件中的所有样本都会按顺序逐个处理。

错误相关

26 Q 如何排查失败的分析任务

- A
- a) 确认 megabolt 服务运行状态，确认 MegaBOLT_server 运行状态
 - b) 结合 client 日志（megabolt.log）和 task 日志（sample.log）中首先出现 warning/error/critical 的位置，综合判断是 client 端还是任务端触发异常处理；
 - c) 结合 megabolt.out、sample.out 查看具体报错信息

27 Q 遇到如下加密锁异常

```
[Rockey] Error: enum Rockey ARM Failed! Return value: 0xF0000001
[Rockey] Error: cannot find valid rockey, 0x0003.
[Rockey] Close file.
[Rockey] Error: decode license failed.
[Rockey] Dongle ARM is already closed.
ERROR 09:10:49,717 RockeyValidateRunnable - Validating failed.
```

- A 请重新插拔加密锁以及重启机器，无法解决请联系当地销售代表或者供应商。

28 Q BQSR 报错 “Exception in thread “main” java.lang.
UnsupportedClassVersionError: bqsr/BaseQualityScoreRecalibration :
Unsupported major.minor version 52.0”

- A 用户在 .bashrc 中指定了 java 路径，且 java 版本过低，我们最低支持的 java 版本为 JDK1.8（请用户屏蔽 .bashrc 中的 java 路径，重新运行任务）。

29 Q MegaBOLT 运行至 HaplotypeCaller 出错，“Input files dbsnp and reference have incompatible contigs.” 如下：

```
##### ERROR #####
##### ERROR A USER ERROR has occurred (version 3.7-0-gcfed67):
##### ERROR
##### ERROR This means that one or more arguments or inputs in your command are incorrect.
##### ERROR The error message below tells you what is the problem.
##### ERROR
##### ERROR If the problem is an invalid argument, please check the online documentation guide
##### ERROR (or rerun your command with --help) to view allowable command-line arguments for this tool.
##### ERROR
##### ERROR Visit our website and forum for extensive documentation and answers to
##### ERROR commonly asked questions https://software.broadinstitute.org/gatk
##### ERROR
##### ERROR Please do NOT post this error to the GATK forum unless you have really tried to fix it yourself.
##### ERROR
##### ERROR MC53AGC: Input files dbsnp and reference have incompatible contigs. Please see https://software.broadinstitute.org
##### ERROR contig dbsnp is named chrM with length 16571
##### ERROR contig reference is named chrM with length 16569 and MD5 c68f52674c9fb33aef52dcf399755519.
##### ERROR dbsnp contigs = [chr1, chr2, chr3, chr4, chr5, chr6, chr7, chr8, chr9, chr10, chr11, chr12, chr13, chr14, chr15,
##### ERROR reference contigs = [chr1, chr2, chr3, chr4, chr5, chr6, chr7, chr8, chr9, chr10, chr11, chr12, chr13, chr14, ch
##### ERROR
##### ERROR 06:43:49,420 FunctionEdge - Error: 'java' '-Xmx4096m' '-XX:+UseParallelOldGC' '-XX:ParallelGCThreads=4' '-XX:GCTime
```

- A 使用的 reference 和 dbsnp 版本不一致，请确保输入相同版本的 reference 和 dbsnp。

- 30 Q 程序主脚本报错，但没有详细信息，如何查看详细错误信息。
- A 可查看输出目录下错误日志 *samplename.out*、*samplename.log* 以及 *megabolt.log* 进行查看详细信息。
- 31 Q 任务中止之后的任务投递不上去
- A 可能由于任务中止后，导致机器死机导致，可能需要重启节点。
- 32 Q 运行 MegaBOLT 任务出错，“*Socket exception - Client connect server failed, strerror: Connection refused*”
- A MegaBOLT server 程序未启动，请等待 1 分钟后重新运行分析。如果仍然出现此问题，请使用命令 `sudo service megabolt restart` 启动 MegaBOLT server 程序，然后再重新投递任务。
- 33 Q 运行 MegaBOLT full 任务出错，报错信息为 "ERROR: split failed, QC1_split.sh error", "pigz: abort: internal threads error"

```
SOAPnukePara reform: -n 0.1 -q 0.5 -l 12 -Q 2 -G 2 -M 2 -T 1
Read length: 100
Split line: 428948
Generate split scripts done.
ERROR: split failed, QC1_split.sh error
pigz: skipping: read1.fq.gz: corrupted -- incomplete deflate data
pigz: abort: internal threads error
pigz: skipping: read2.fq.gz: corrupted -- incomplete deflate data
pigz: abort: internal threads error
```

- A 原始 reads 不完整，请检查原始文件
- 34 Q 运行 MegaBOLT full 任务出错，报错信息为 "ERROR: Sample-name.bam_stats.sh error"

```
ERROR: test.bam_stats.sh error
[Thu Dec 08 15:25:54 CET 2020] picard.analysis.CollectInsertSizeMetrics HISTOGRAM_FILE=/mnt/ssd/MegaBOLT_scheduler/tmpDir/1/1/test/b
[Thu Dec 08 15:25:54 CET 2020] Executing as megabolt@cngb-fpga-f14-1.cngb.ac.hpc on Linux 3.10.0-693.el7.x86_64 amd64/ OpenJDK 64-bit
INFO 2020-12-08 15:25:54 Executor      Executing R script via command: Rscript /tmp/megabolt/executor/RT10554303757830553.R /m
INFO 2020-12-08 15:25:55 ProcessExecutor [1] "Not creating insert size PDF as there are duplicated header names: All Reads"
INFO 2020-12-08 15:25:55 ProcessExecutor [2] "Not creating insert size PDF as there are duplicated header names: All Reads"
[Thu Dec 08 15:25:55 CET 2020] picard.analysis.CollectInsertSizeMetrics done, elapsed time: 0.33 minutes,
```

- A 这个问题有两个原因导致，请分别检查：
- a) 多文库合并分析，不同 Insert size 数据合并分析导致
- b) List 错误，例如 read1 和 read2 都写成 read1，例如 "samplename read1.fq.gz read1.fq.gz"；不同 barcode 的数据写错在一起分析，例如 "samplename 33_1.fq.gz 35_2.fq.gz"