

Catch-up

- [Environment setup](#)
 - [Directories and symlinks](#)
 - [.bashrc setup](#)
 - [Environment variables](#)
- [TACC intro](#)
 - [Commands files](#)
 - [Start an idev session](#)
- [Working with FASTQ](#)
 - [Yeast data](#)
 - [ATACseq data for MultiQC](#)
 - [FASTQ files for cutadapt](#)
- [Alignment workflow](#)
 - [Alignment workflow setup](#)
 - [References](#)
 - [BWA PE alignment of yeast data](#)
 - [samtools manipulation of aligned yeast data](#)
- [SAMTools and BEDTools](#)
 - [Setup for samtools](#)

Environment setup

Directories and symlinks

Directories and links needed in your **home** directory.

```
cd
ln -s -f $SCRATCH scratch
ln -s -f $WORK2 work2
ln -s -f /work2/projects/BioITeam
ln -s -f /work2/projects/BioITeam/projects/courses/Core_NGS_Tools CoreNGS

mkdir -p ~/local/bin
cd ~/local/bin
ln -s -f /work2/projects/BioITeam/common/bin/launcher_creator.py
ln -s -f /work2/projects/BioITeam/common/bin/launcher_maker.py
```

.bashrc setup

If you already have a [.bashrc](#) set up, make a backup copy first. You can restore your original login script after class is over.

```
cd
cp .bashrc .bashrc.beforeNGSTools
```

Copy and configure the login profile for this class

```
cd
cp /work2/projects/BioITeam/projects/courses/Core_NGS_Tools/tacc/bashrc.corengs.stamped2 .bashrc
chmod 600 .bashrc

# or, using your symlink
cd
cp ~/CoreNGS/tacc/bashrc.corengs.stamped2 .bashrc
chmod 600 .bashrc
```

Source it to make it active (if this doesn't work, log off then log back in):

Copy a pre-configured login script

```
source ~/.bashrc
```

Environment variables

General

```
export ALLOCATION=UT-2015-05-18
export BIWORK=/work2/projects/BioITeam
export CORENGS=$BIWORK/projects/courses/Core_NGS_Tools

export PATH=.:$HOME/local/bin:$PATH
```

Turn on coloring by file type in the shell:

```
# For better colors using a dark background terminal, un-comment this line:
export LS_COLORS=$LS_COLORS:'di=1;33:fi=01:ln=01;36:'

# For better colors using a white background terminal:
export LS_COLORS=$LS_COLORS:'di=1;34:fi=01:ln=01;36:'

# May or may not be needed
export LS_OPTIONS='-N --color=auto -T 0'
```

TACC intro

Commands files

Simple commands

```
mkdir -p $SCRATCH/core_ngs/slurm/simple
cd $SCRATCH/core_ngs/slurm/simple
cp $CORENGS/tacc/simple.cmds .
```

Wayness commands

```
mkdir -p $SCRATCH/core_ngs/slurm/wayness
cd $SCRATCH/core_ngs/slurm/wayness
cp $CORENGS/tacc/wayness.cmds .
```

Start an idev session

To start a 3-hour **idev** (interactive **development**) session:

Start an idev session

```
idev -p normal -m 120 -N 1 -n 68 -A UT-2015-05-18 --reservation=BI0_DATA_week_1
```



You can tell you're in a **idev** session because the **hostname** command will return a compute node name (e.g. **c401-041.stamped2.tacc.utexas.edu**) instead of a login node name (e.g. **login2.stamped2.tacc.utexas.edu**).

The **n idev** session will terminate when the requested time has expired, or you use the **exit** command.

Working with FASTQ

Yeast data

Working with some yeast ChIP-seq FASTQ data:

```
# Create a $SCRATCH area to work on data for this course,
# with a sub-directory for pre-processing raw fastq files
mkdir -p $SCRATCH/core_ngs/fastq_prep

# Make symbolic links to the original yeast data:
cd $SCRATCH/core_ngs/fastq_prep
ln -s -f $CORENGS/yeast_stuff/Sample_Yeast_L005_R1.cat.fastq.gz
ln -s -f $CORENGS/yeast_stuff/Sample_Yeast_L005_R2.cat.fastq.gz

# Copy over a small FASTQ file
cd $SCRATCH/core_ngs/fastq_prep
cp $CORENGS/misc/small.fq .
```

ATACseq data for MultiQC

Get some FastQC reports for [MultiQC](#):

```
mkdir -p $SCRATCH/core_ngs/multiqc/fqc.atacseq
cd $SCRATCH/core_ngs/multiqc/fqc.atacseq
cp $CORENGS/multiqc/fqc.atacseq/*.html .
```

FASTQ files for cutadapt

For command-line [cutadapt](#) exploration:

```
cd $SCRATCH/core_ngs/fastq_prep
cp $CORENGS/human_stuff/Sample_H54_miRNA_L004_R1.cat.fastq.gz .
cp $CORENGS/human_stuff/Sample_H54_miRNA_L005_R1.cat.fastq.gz .
zcat Sample_H54_miRNA_L004_R1.cat.fastq.gz | head -2000 > miRNA_test.fq
```

For batch [cutadapt](#) processing:

```
mkdir -p $SCRATCH/core_ngs/cutadapt
cd $SCRATCH/core_ngs/cutadapt
cp $CORENGS/human_stuff/Sample_H54_miRNA_L004_R1.cat.fastq.gz .
cp $CORENGS/human_stuff/Sample_H54_miRNA_L005_R1.cat.fastq.gz .
cp $CORENGS/yeast_stuff/Yeast_RNAseq_L002_R1.fastq.gz .
cp $CORENGS/yeast_stuff/Yeast_RNAseq_L002_R2.fastq.gz .
cp $CORENGS/tacc/cuta.cmds .
```

Alignment workflow

Alignment workflow setup

Starting files:

```
# FASTA (for building references)
mkdir -p $SCRATCH/core_ngs/references/fastq
cp $CORENGS/references/*.fa $SCRATCH/core_ngs/references/fastq/

# FASTQ (to align)
mkdir -p $SCRATCH/core_ngs/alignment/fastq
cp $CORENGS/alignment/*fastq.gz $SCRATCH/core_ngs/alignment/fastq/
```

References

Get a copy of all references we build in the exercises (including FASTA):

```
mkdir -p $SCRATCH/core_ngs/references
rsync -ptlvrP $CORENGS/references/ $SCRATCH/core_ngs/references/
```

BWA PE alignment of yeast data

To jump into aligning PE yeast data with BWA

```
# Pre-built references
mkdir -p $SCRATCH/core_ngs/references
rsync -avrP $CORENGS/references/ $SCRATCH/core_ngs/references/

# FASTQ (to align)
mkdir -p $SCRATCH/core_ngs/alignment/fastq
cp $CORENGS/alignment/*fastq.gz $SCRATCH/core_ngs/alignment/fastq/

# Alignment directory
mkdir -p $SCRATCH/core_ngs/alignment/yeast_bwa
cd $SCRATCH/core_ngs/alignment/yeast_bwa
ln -s -f ../fastq
ln -s -f ../../references/bwa/sacCer3

module load biocontainers # takes a while
module load bwa
module load samtools
```

samtools manipulation of aligned yeast data

To jump into post-alignment manipulation of the [yeast_pairedend.bam](#) with [samtools](#):

```
mkdir -p $SCRATCH/core_ngs/alignment/yeast_bwa
cd $SCRATCH/core_ngs/alignment/yeast_bwa
cp $CORENGS/catchup/yeast_bwa/yeast_pairedend.bam .

module load biocontainers # takes a while
module load samtools

# If the sorted, indexed BAM is needed:
cp $CORENGS/catchup/yeast_bwa/yeast_pairedend.sort* .
```

SAMTools and BEDTools

Setup for samtools

Setup for samtools exercises

```
mkdir -p $SCRATCH/core_ngs/samtools
cd $SCRATCH/core_ngs/samtools
cp $CORENGS/catchup/for_samtools/* .

module load biocontainers # takes a while
module load samtools
```