

RNA-Seq Workshop

Using Tuxedo and Trinity for Transcript Reconstruction and Analysis

Software and Data Installation Using
VirtualBox

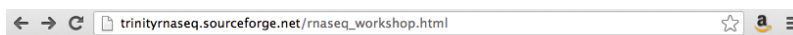
Download and Install VirtualBox

<http://virtualbox.org>



Download the RNA-Seq Workshop Virtual Machine (VM) Image and Related Materials

http://trinityrnaseq.sourceforge.net/rnaseq_workshop.html



RNA-Seq Workshop

Do you want to learn the basics of how to analyze RNA-Seq data, including genome-based and genome-free transcript reconstruction and differential expression analysis? Below, we provide materials to help get you started. A VirtualBox virtual machine image is provided that includes several popular tools for analyzing RNA-Seq, including Tuxedo for genome-based and Trinity for genome-free RNA-Seq studies, and small sample (simulated) RNA-Seq data sets. Hands-on tutorials are provided to guide users in a step-by-step analysis of these data. Note, the only purpose of these simulated data is to give the user hands-on experience in using the various analysis tools and be sufficiently small in size such that the data can be processed on an average desktop computer in a small amount of time (minutes rather than hours). Future work may have an increased focus on learning interesting biology from real data sets, in addition to familiarizing users with use of these applications.

To get started, you must first install [VirtualBox](#) and import the [Ubuntu Linux image provided](#) that's preinstalled with the required data and software. A [basic installation guide](#) with screenshots is provided. VirtualBox should allow a user to experiment with this Ubuntu Linux from within their own operating system, including from Mac or Windows operating systems.

The accompanying Trinity and Tuxedo walk-throughs are available from the Trinity ftp site at: http://sourceforge.net/projects/trinityrnaseq/files/misc/RNASEQ_WORKSHOP/.

In summary, these are the requirements:

- Install Virtualbox from <http://virtualbox.org>
- Download the [Ubuntu Linux image with pre-installed software and data](#). (Ubuntu_RNASeqWorkshop.ova)
- Install the Ubuntu image following [these instructions](#).
- Run through the Trinity and/or the Tuxedo [tutorials](#).

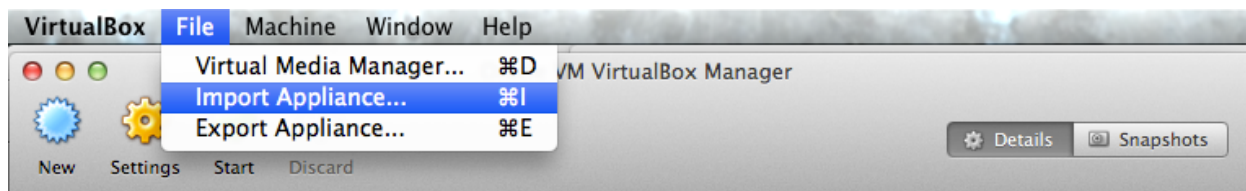
Last updated 2013-06-17 14:03:40 EDT

Note: This is the installation guide mentioned above.

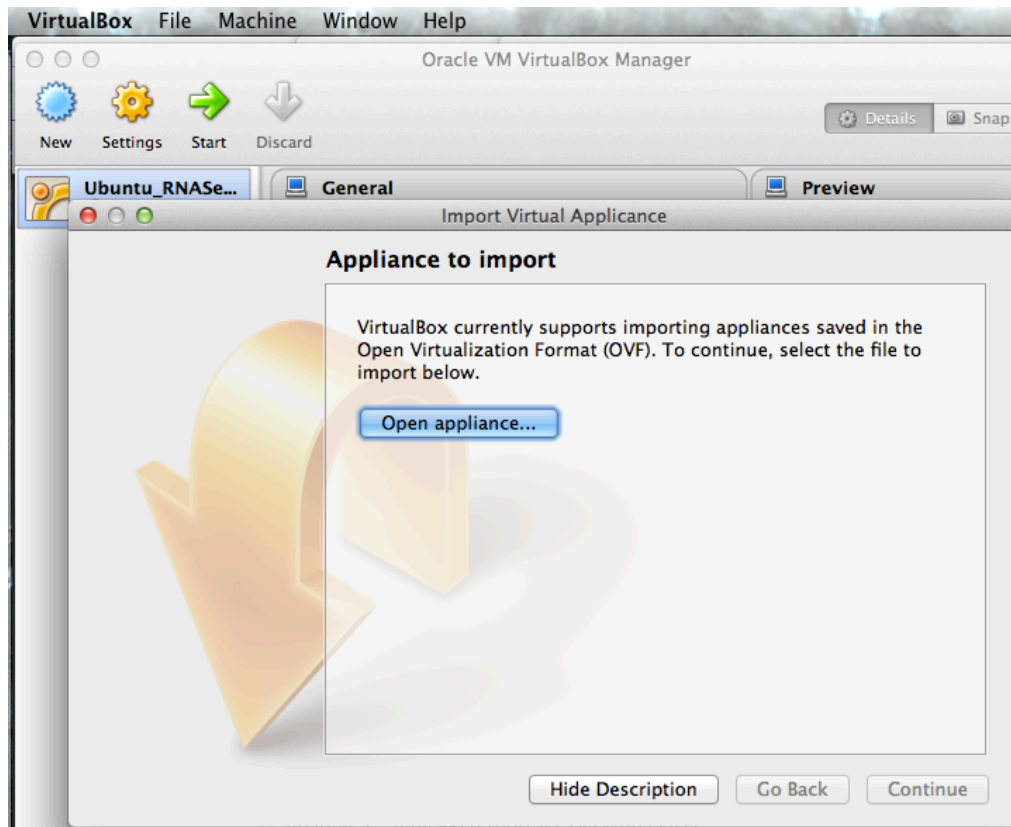
Download
Materials



Run VirtualBox and Import Appliance

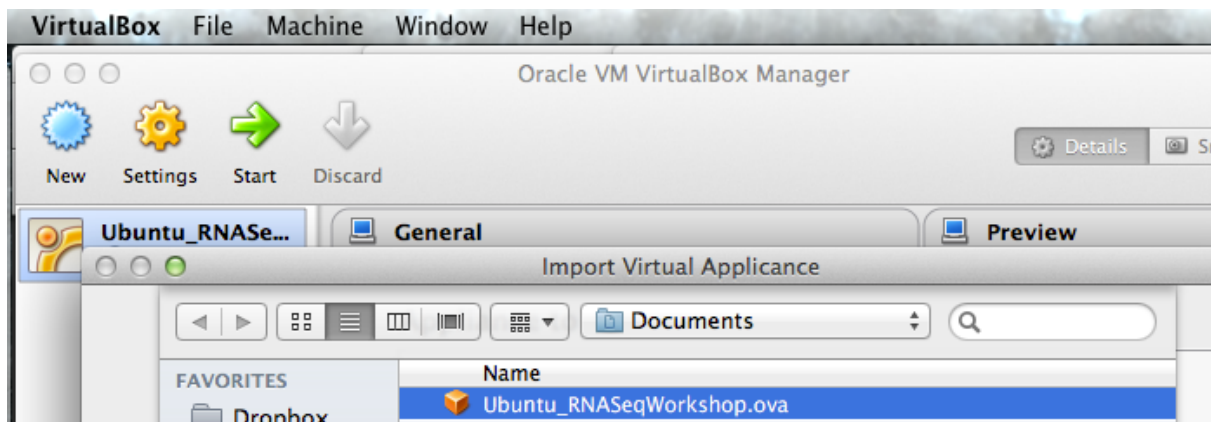


Next, click the 'Open appliance...' button

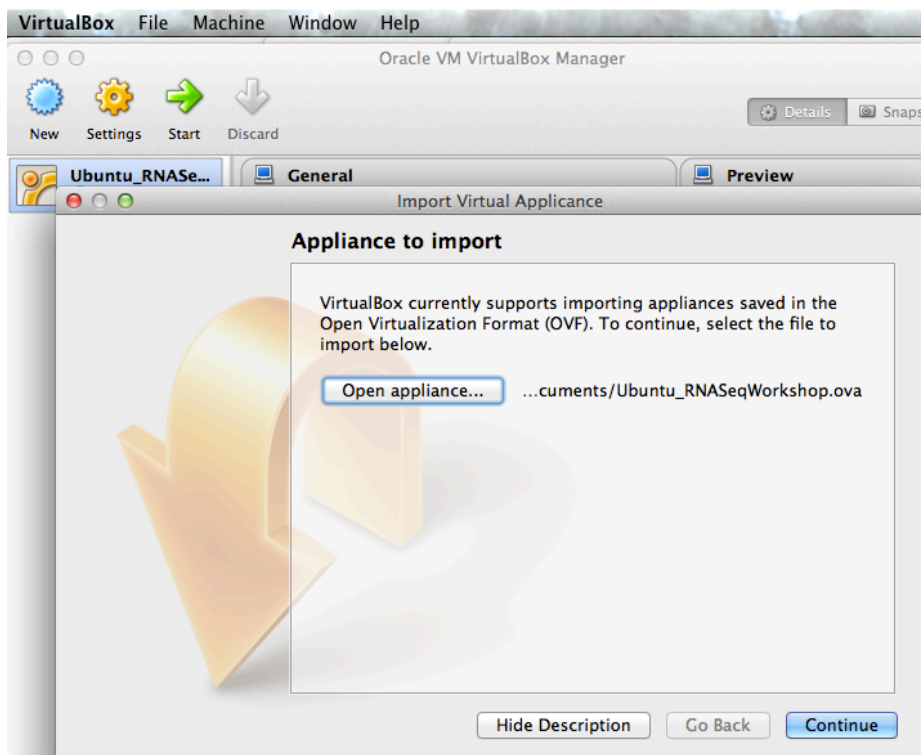


Select the Ubuntu_RNASeqWorkshop.ova file you downloaded earlier.

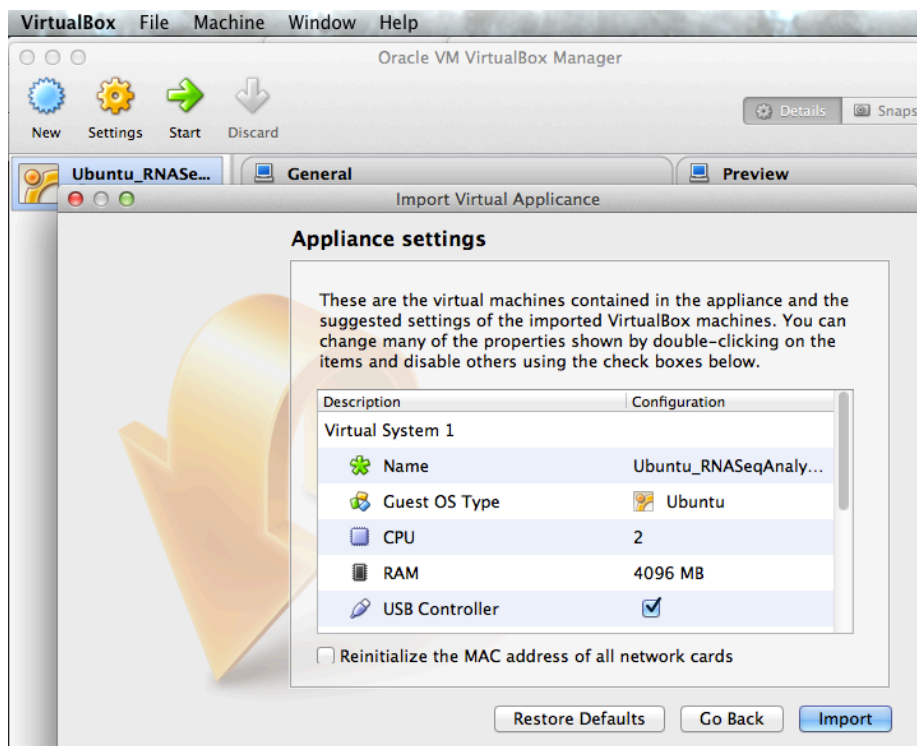
(note, the .ova file may be named differently depending on workshop VM release)



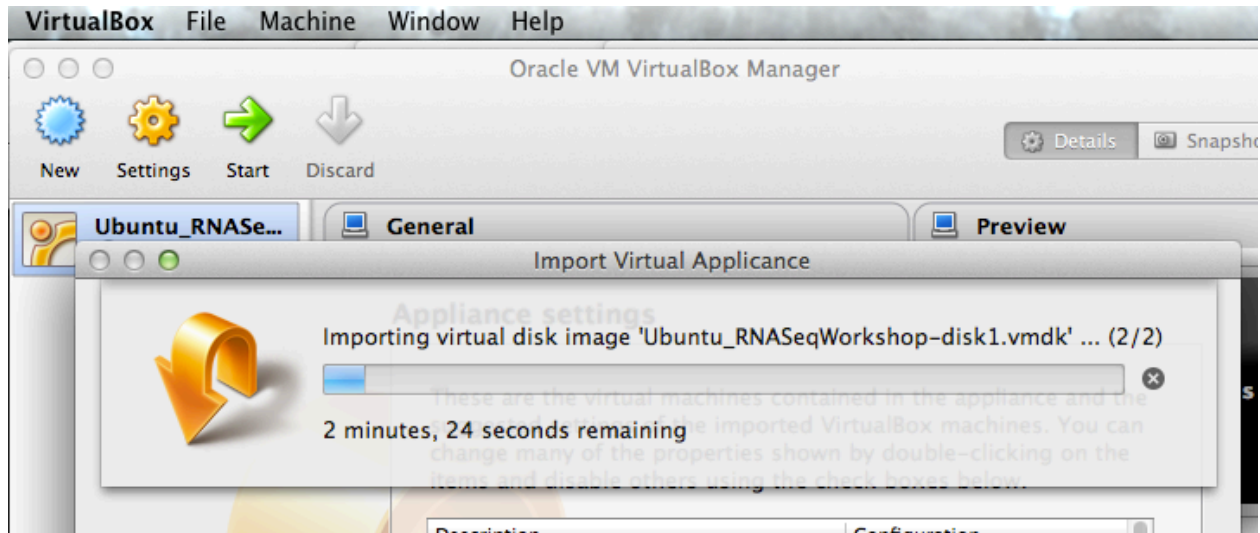
Click 'Continue'



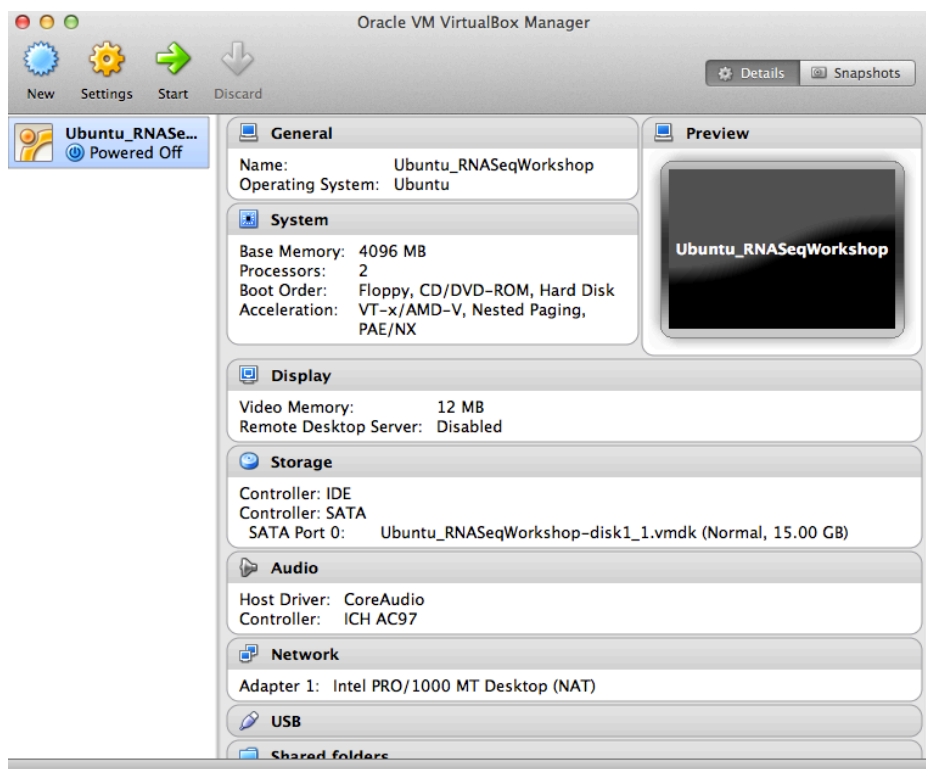
Click 'Import'



This import step can take a few minutes.



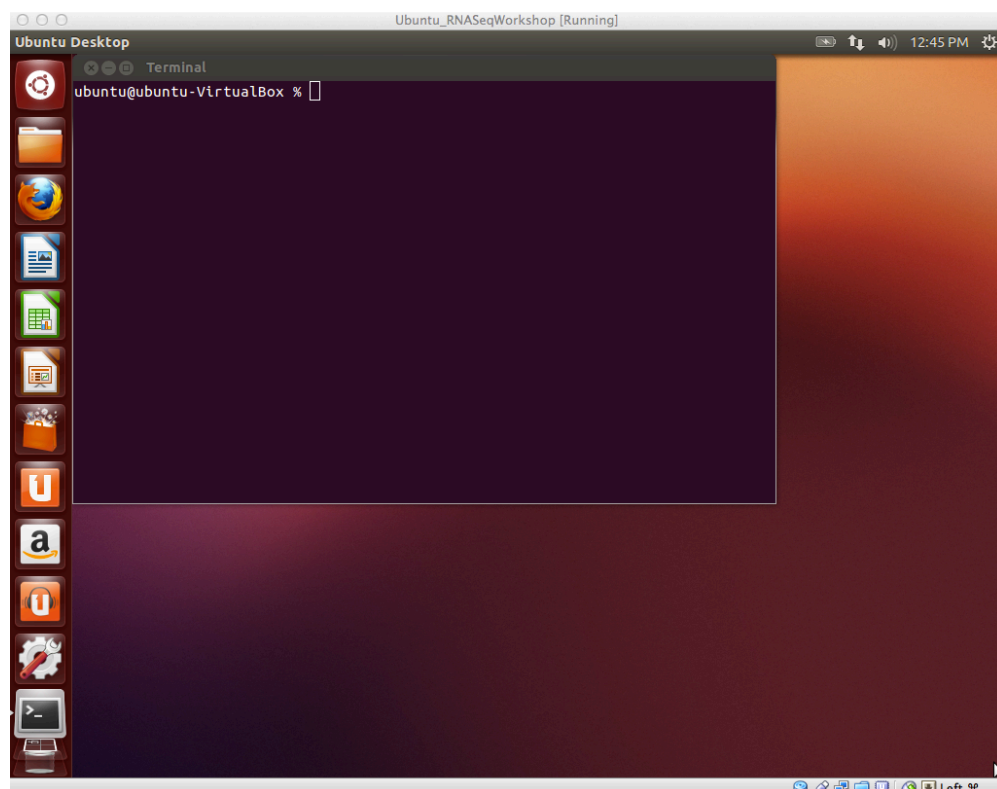
Select the Ubuntu_RNASeqWorkshop entry on the left, and click the Start button at top.



Ubuntu Linux should load. Click the 'terminal' icon on the left to open a terminal window.

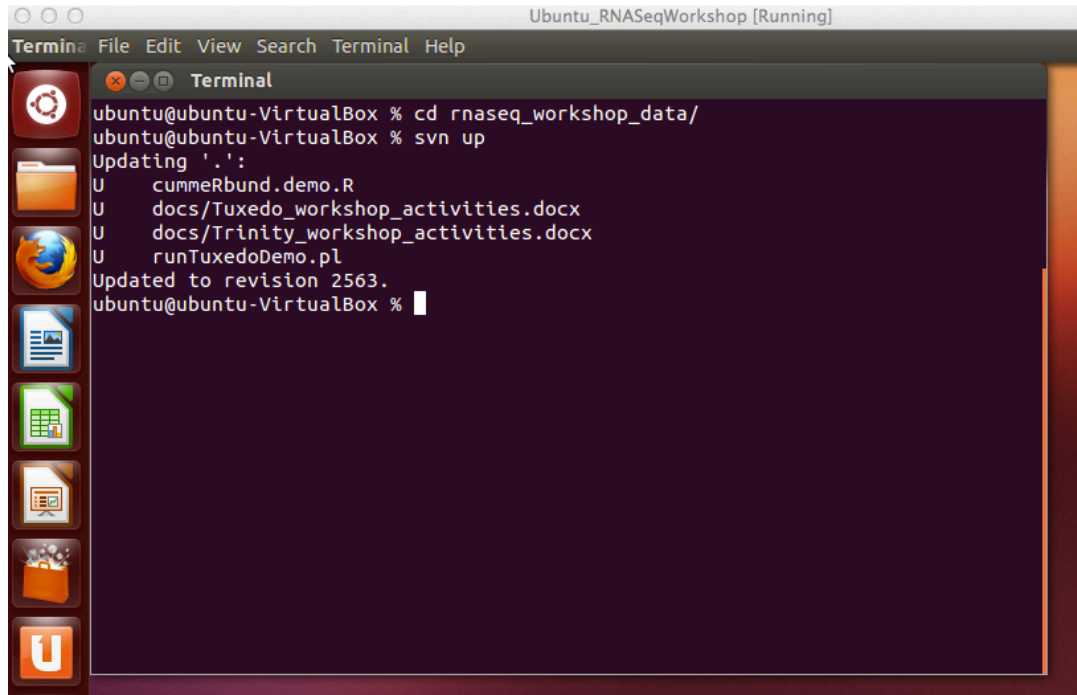


Terminal window should appear.



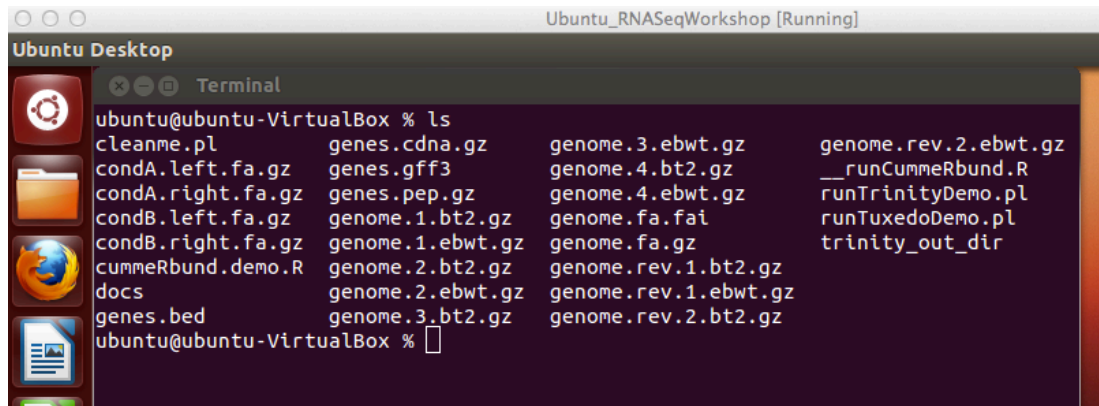
Enter the `rnaseq_workshop_data/` directory, and update the contents:

(note, may be `rnaseq_workshop_{year}/` in newer versions)



```
Ubuntu_RNASeqWorkshop [Running]
Terminal
ubuntu@ubuntu-VirtualBox % cd rnaseq_workshop_data/
ubuntu@ubuntu-VirtualBox % svn up
Updating '.':
U    cummeRbund.demo.R
U    docs/Tuxedo_workshop_activities.docx
U    docs/Trinity_workshop_activities.docx
U    runTuxedoDemo.pl
Updated to revision 2563.
ubuntu@ubuntu-VirtualBox %
```

Data and demo scripts provided:



```
Ubuntu_RNASeqWorkshop [Running]
Ubuntu Desktop
Terminal
ubuntu@ubuntu-VirtualBox % ls
cleanme.pl          genes.cdna.gz       genome.3.ebwt.gz    genome.rev.2.ebwt.gz
condA.left.fa.gz    genes.gff3          genome.4.bt2.gz     __runCummeRbund.R
condA.right.fa.gz   genes.pep.gz        genome.4.ebwt.gz    runTrinityDemo.pl
condB.left.fa.gz    genome.1.bt2.gz     genome.fa.fai       runTuxedoDemo.pl
condB.right.fa.gz   genome.1.ebwt.gz    genome.fa.gz        trinity_out_dir
cummeRbund.demo.R   genome.2.bt2.gz     genome.rev.1.bt2.gz
docs                genome.2.ebwt.gz    genome.rev.1.ebwt.gz
genes.bed           genome.3.bt2.gz     genome.rev.2.bt2.gz
ubuntu@ubuntu-VirtualBox %
```

Lots of files, but the most important data files used here are:

- Paired-end RNA-Seq reads for two conditions (A, B):
 - condA.left.fa and condA.right.fa
 - condB.left.fa and condB.right.fa
- Genome sequence:
 - genome.fa

Notes on the VM configuration

- All required software tools come pre-installed
 - Transcript reconstruction:
 - Trinity and Tuxedo (bowtie, tophat, cufflinks, cuffdiff, cummeRbund)
 - Visualization
 - IGV
 - Utilities leveraged during RNA-Seq processing and analysis:
 - samtools, GMAP, R/Bioconductor, etc.

Notes on the VM Environment

- All required software tools are available via the PATH setting. (so just type 'tophat' or 'cufflinks' at the command prompt).
- Additional environmental variables are set to facilitate access to Trinity and IGV
 - \$TRINITY_HOME = /path/to/Trinity/installation/directory
 - \$IGV = /path/to/IGV/installation/directory

These environmental variables are used throughout the hands-on tutorials.

Automating Tutorial Execution

*See documents for hands-on activities for Trinity and Tuxedo

Automating Trinity Activities:

```
% runTrinityDemo.pl
```

Automating Tuxedo Activities:

```
% runTuxedoDemo.pl
```