

SUPPLEMENTARY MATERIAL FOR THE PLANT JOURNAL PUBLICATION

Liu, L., Ramsay, T., Zinkgraf, M., Street, N.R., Filkov, V., Groover, A. (2015) A resource for characterizing genome-wide binding and target genes of transcription factors expressed during secondary growth in *Populus*. The Plant Journal

PURPOSE

This readme provides the tools and data used for the facts and figures in the paper. The accompanying bash and R scripts are easily modified to work with any ChIP-seq data.

VIRTUAL MACHINE REQUIREMENTS

Standalone Virtual Machine

If you are building the Virtual Machine on a standalone system you can obtain the Virtualbox Image at the following url: <http://www.cs.ucdavis.edu/~filkov/software/iPlant/>

IMPORTANT: This image does not have a boot loader and you'll need to add one to get it working within virtual box. Once configured the root password has been set to 'atmosphere'. Otherwise, you can mount this disk onto an existing virtual box image.

Virtual Machine on Atmosphere

You will only need to get an account and get access to Atmosphere. The image has the name: Populus_ChIPseq_VM_1.0. If being used with the Populus_ChIPseq data, please create an instance with at least 160 GB of hard drive space.

DATA RESOURCES

Sequence Data

If reproducing the results in the paper, download the data from NCBI SRA:

<http://www.ncbi.nlm.nih.gov/sra>

If you wish to individually view the files on NCBI, use the following accession numbers: ARK1 ChIP-seq can be found by searching for SRX561124. inside are the following test runs:

SRR1315545, SRR1315546, SRR1315547, SRR1315548, SRR1315549, SRR1315550

ARK2, BLR, PCN, and PRE ChIP-seq can be found by searching for SRP053368. inside are the following test runs:

SRR1575896, SRR1575630, SRR1575634, SRR1575633

But, the image also includes all that is required to download the files to the image by following the HOWTO.

HOWTO: Download the files from the NCBI using the VM image:

1. Execute the following command to provide access to the SRA toolkit for downloading:
`export PATH=/opt/sratoolkit.2.4.4-ubuntu64:$PATH`
2. Copy the contents of /Repository to your home directory
3. Execute the following command to download the files from NCBI, this download will take a very long time:

prefetch --option-file SRA_list.txt

4. An NCBI folder will be created containing all of the .sra files. Execute the following command for each of the downloaded files to have the fastq format files needed for the IDR_analysis.sh script.

fastq-dump <accession number>

Reference

Use the following link to access the manual for the SRA toolkit functions used in the HOWTO:

http://www.ncbi.nlm.nih.gov/Traces/sra/?view=toolkit_doc

Virtual Machine and Scripts

The scripts used for processing the ChIP-seq data are provided at the following url:

<http://www.cs.ucdavis.edu/~filkov/software/iPlant/>

IDR Analysis of ARK1 ChIP-seq Data

For the source code used for IDR data processing please reference the supplementary notes from the following publication:

Liu, L., Zinkgraf, M., Petzold, H. E., Beers, E. P., Filkov, V. and Groover, A. (2015), The Populus ARBORKNOX1 homeodomain transcription factor regulates woody growth through binding to evolutionarily conserved target genes of diverse function. New Phytologist, 205: 682–694. doi: 10.1111/nph.13151

<http://onlinelibrary.wiley.com/store/10.1111/nph.13151/asset/supinfo/nph13151-sup-0003-NotesS1-S2.pdf?v=1&s=161d5948c240cd3d3acf9ff20ac2649d444b2e8c>

Supporting Information Notes S1 Commands used for peak calling and IDR analysis.

The IDR source code has also been included in the folder /Repository/idrCode. Reference

<https://sites.google.com/site/anshulkundaie/projects/idr> for additional information.

ARK1_IDR_analysis.sh

This file contains the steps necessary to complete processing from raw ChIP-seq data to IDR peak data. The file is a copy from the supporting information script provided with the publication, and it provides analysis for a 3 replicate condition. Reference

<https://sites.google.com/site/anshulkundaie/projects/idr> for additional information.

Additionally, the IDR source code has also been included in the folder idrCode for use in the analysis.

ARK2 ChIP-seq_CHIPpeakAnno and GStat.txt

This file contains the R script necessary to complete CHIPpeakAnno and GStats analysis with completed ChIP-seq peak data on the ARK2 ChIP-seq data used in the study.

GO_Ontology_BP_Heatmaps.R

This file contains the R script necessary to create the heatmaps presented in the paper from GO Ontology data collected from the GStats analysis.

SOFTWARE AND R PACKAGES

Software

R 3.1.2, scythe 0.950, sickle 1.33, macs2 2.1.0.2040616, htseq 0.6.0, samtools 0.1.19, bowtie2 2.2.4, tophat 2.0.13, Python 2.7, gzip 1.4

R Packages

idr 1.2, gplots 2.15.0, GOstats 2.32.0, GSEABase 1.28.0, biomaRt 2.22.0, ChIPpeakAnno 2.16.4, edgeR 3.8.5, Biostrings 2.34.1