Genome-wide study of KNOX Regulatory Network in Shoot Meristem Function in Rice

Xinzhou Liu

**Introduction**

In contrast to animals that have completed their organ formation before birth, the plants continue to produce organs throughout their life cycle. The continuous formation of organs is very important to the plant development and growth; therefore understanding such feature can potentially improve the yield in many crop plants and benefit agriculture and food industry.

Shoot apical meristem (SAM) is responsible for the continous growth of plant organs. It

is a region in a growing shoot consisting of self-renewing stem cells that can be transitioned into different specialized cell types such as leaf and flower cells, therefore playing an important part in the formation of aboveground organs/tissues [1].

The undifferentiated status of SAM is found to be maintained by Class I knotted1-like homeobox (KNOX) transcription factors. Studies have show that KNOX genes are specifically expressed in SAM, and the overexpression of KNOX in leaves would affect the leaf growth pattern while loss of KNOX genes would result in failure of SAM formation and maintenance[2]. KNOW were also found to be involved in regulating growth-promoting hormone cytokinin biosynthesis, as well as in regulating biosynthesis of cell wall components such as lignin and gibberellin [2]. Yet, we have not completely understood how the KNOX genes contribute to the maintenance of indeterminacy of SAM.

OSH1 is one of the KNOX transcription factors (TFs) in rice (*Oryza sativa*). OSH1 was found to regulate the signaling pathway of brassinosteroids, a group of hormones promoting differentiation in meristem cells **[**3]. To gather a better understanding of this KNOX gene function, I planed to run a genome-wide analysis of the OSH1 downstream pathway hoping to find a comprehensive set of target genes of OSH1 and build its gene regulatory network.

**Workflow**

**1. Basic workflow**

To implement the genome-wide analysis, data from ChIP-Seq and RNA-Seq were used to build the gene network (Fig.1). The ChIP-Seq experiment consisted of two samples, one from chromatin immunoprecipitation of Japonica rice Nipponbare panicle sample using OSH1 antibody, and the other from the same using rabbit IgG as negative control. Each sample had two biological replicates. As for the RNA-Seq experiment, DNA were extracted from the wild-type leaf tissues and from the OSH1-overexpressing leaf tissues of the same rice line. Each treatment consisted of two biological replicates. [3]

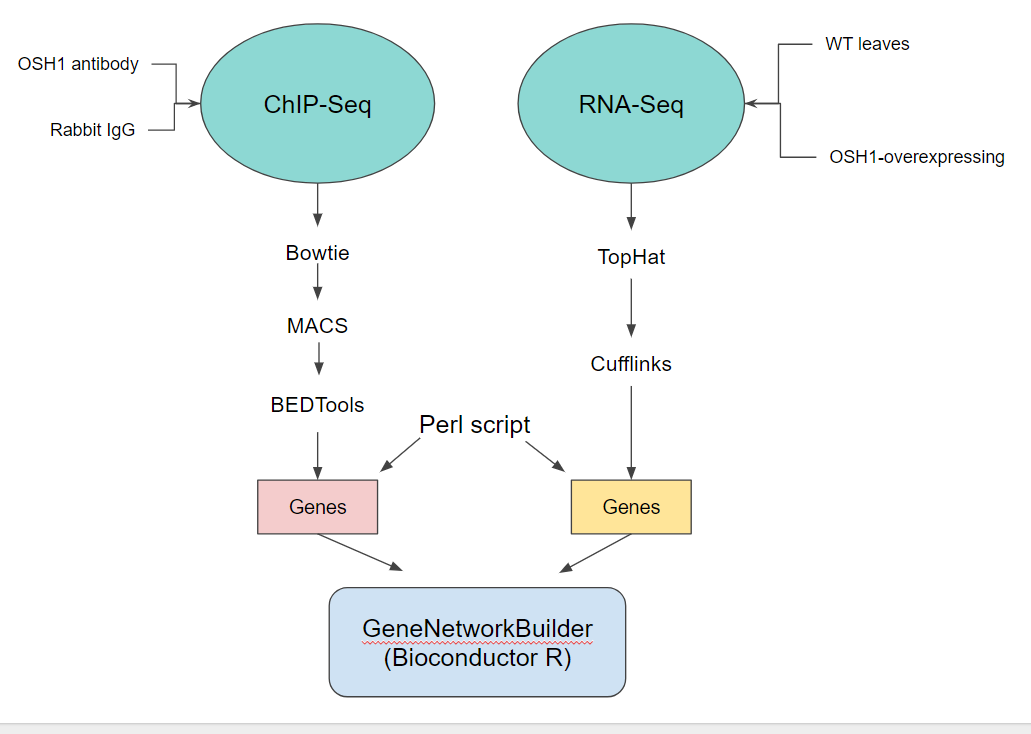


Figure 1. The workflow of building the gene regulatory network of OSH1.

The ChIP-seq raw data ( DRA000206 and DRA000313) were quality-controlled by FastQC, and aligned to the rice genome(MSU7.0) using Bowtie. MACS software was used for peak calling. The list of the nearest genes to the significant peaks were extracted using BEDTools. The list of genes, which were the potential targets of OSH1, was used for building the gene network.

The RNA-Seq reads (DRA002287) were quality-controlled by FastQC and aligned to the rice genome (MSU7.0) using TopHat. Cufflinks was used to identify differentially expressed genes provided with a GFF3 annotation file. The output file of Cufflinks was used to build the gene network.

The gene regulatory network of OSH1 was generated using GeneNetworkBuilder (Bioconductor R),inputting the bound list identified by ChIP-Seq and the gene expression result.

Additional scripts were written to format data in transition from one software to another,

**2. ChIP-Seq analysis steps**

- Set working directory to: /courses/INFOINST8150\_S2018/xl6v8/Data\_NEW/final\_project

- Indexing the genome:

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/bowtie2-2.2.4/bowtie2-build genome/rice\_genome.fa genome/rice\_genome

- Aligning the reads and output as .sam:

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/bowtie2-2.2.4/bowtie2 -x rice\_genome -1 osh1\_1.fastq.bz2 -2 osh1\_2.fastq.bz2 -S chipseq/osh1.sam

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/bowtie2-2.2.4/bowtie2 -x rice\_genome -1 ctrl\_1.fastq.bz2 -2 ctrl\_2.fastq.bz2 -S chipseq/ctrl.sam

(No replicate, otherwise too much time consumption)

- Run MACS analysis using .sam as input files.

$ /courses/INFOINST8150\_S2018/Software/MACS-1.4.2/bin/macs14 -t chipseq/osh1.sam -c chipseq/ctrl.sam -f SAM -g 3.73e8 -n chipseq/chipseq

- Select peaks with FDR < 1% and fold-enrichment > 120

Use chipseq/filter.pl

- BEDTool select the closest genes to the peaks:

First, extract just gene annotations from all.gff3 file

$ awk '{if($3 ~ /gene/) print $0}' genome/all.gff3 > genome/genes.gff3

Then sort the bed files because “bedtools closest” requires that all input files are presorted data by chromosome and then by start position (e.g., sort -k1,1 -k2,2n in.bed > in.sorted.bed for BED files).

$ sort -k 1,1 -k 2,2n chipseq/chipseq\_peaks\_filtered.bed > chipseq/sorted\_peaks.bed

$ sort -k 1,1 -k 4,4n genome/genes.gff3 > genome/sorted\_genes.gff3

Then use “bedtools closest”

$ ~/bedtools2/bin/bedtools closest -io -a chipseq/sorted\_peaks.bed -b genome/sorted\_genes.gff3 > chipseq/nearest.out

**3. RNA-Seq analysis steps**

- Set working directory to: /courses/INFOINST8150\_S2018/xl6v8/Data\_NEW/final\_project

- Index transcriptome:

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/tophat/tophat2 -G genome/all.gff3 --transcriptome-index=transcriptome/all\_transcripts genome/rice\_genome

- Tophat alignment:

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/tophat/tophat2 --transcriptome-index=transcriptome/all\_transcripts --library-type fr-secondstrand genome/rice\_genome wt\_rep1.fastq

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/tophat/tophat2 --transcriptome-index=transcriptome/all\_transcripts -o wt\_rep2 -p6 --library-type fr-secondstrand genome/rice\_genome wt\_rep2.fastq

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/tophat/tophat2 --transcriptome-index=transcriptome/all\_transcripts -o osh1\_rep1 -p6 --library-type fr-secondstrand genome/rice\_genome osh1\_rep1.fastq

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/tophat/tophat2 --transcriptome-index=transcriptome/all\_transcripts -o osh1\_rep2 -p6 --library-type fr-secondstrand genome/rice\_genome osh1\_rep2.fastq

- Cuffdiff to calculate differential gene expression:

$ /courses/INFOINST8150\_S2018/Software/F2015\_Software/cufflinks/cuffdiff -o cuffdiff\_out transcriptome/all\_transcripts.gff wt\_rep1/accepted\_hits.bam,wt\_rep2/accepted\_hits.bam osh1\_rep1/accepted\_hits.bam,osh1\_rep2/accepted\_hits.bam

- Keep only the significantly differential expressions:

$ grep -i yes cuffdiff\_out/gene\_exp.diff > gene\_exp\_significant.diff

**4. Build gene network**

- Set working directory to /courses/INFOINST8150\_S2018/xl6v8/Data\_NEW/final\_project/geneNetwork

- Copy results of ChIP-Seq and RNA-Seq analyses to the current directory:

$ cp ../chipseq/nearest.out .

$ cp ../cuffdiff\_out/gene\_exp\_significant.diff

- Create the input of OSH1-binding list, “bind\_list.csv”, using the list of nearest genes to the peaks:

$ create\_bind\_list.pl nearest.out

- Create the input of the differential expression data, “exprdata.csv”, and remove infinite number:

$ create\_expr\_input.pl gene\_exp\_significant.diff

$ grep -v -i inf exprdata.csv > exprdata\_temp.csv

$ mv exprdata\_temp.csv exprdata.csv

- Create dummy miRNA.map (not used, just beacuse it is GeneNetworkBuilder requirement)

- Use experimentally verified protein-protein interactions from

http://bis.zju.edu.cn/prin/download.do

- Run the osh1\_network.R (See “Results” section for the plots)

- Extract function annotations of the downstream genes in the network:

$ perl extract\_targets.pl cifNetwork.csv ../genome/genes.gff3

**Results**

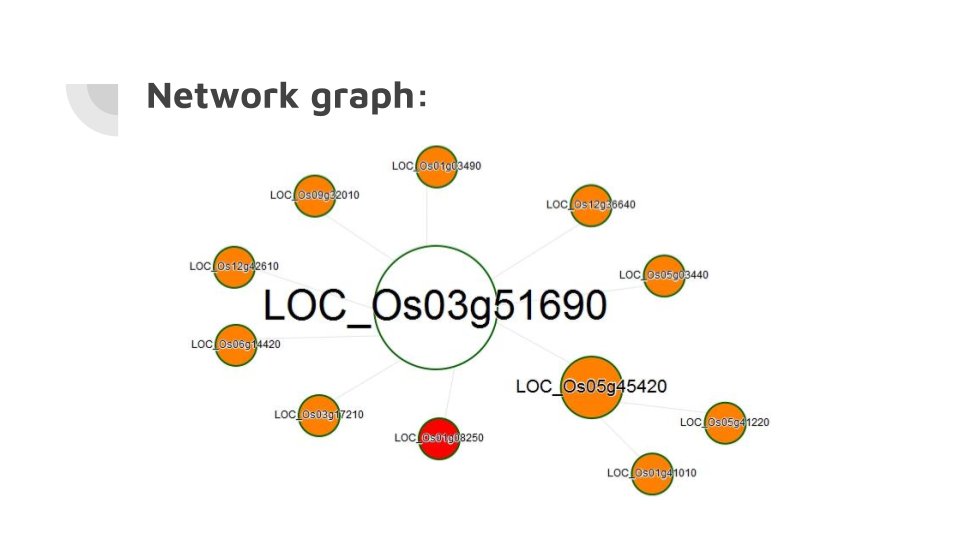
****

Figure 1. The OSH1-related genes interaction network. LOC\_Os0351690 (center) is the ID of OSH1.

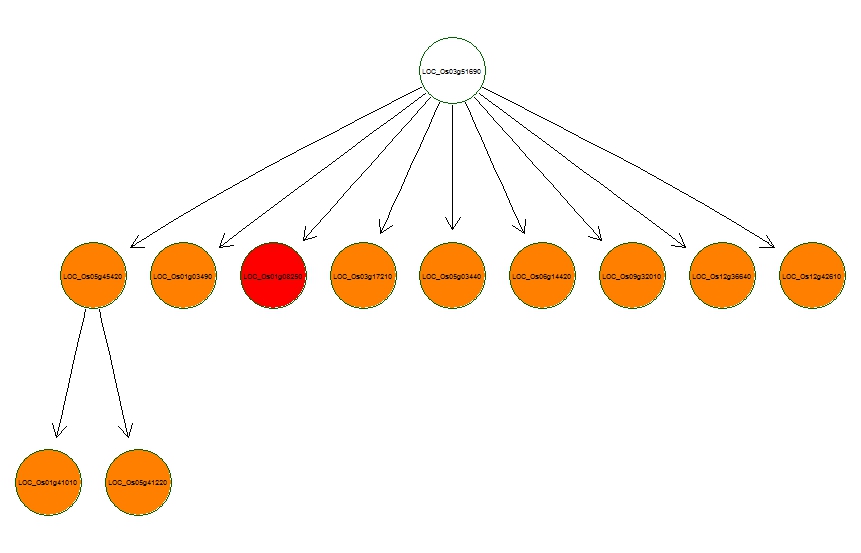
****

Figure 2. The network is plotted by interaction level (direct or indirect).

|  |  |
| --- | --- |
| Gene\_ID | Function\_Annotation |
| LOC\_Os01g03490 | Heavy-metal-associated domain-containing protein, putative, expressed |
| LOC\_Os01g08250 | Retrotransposon protein, putative, unclassified, expressed |
| LOC\_Os01g41010 | DUF581 domain containing protein, expressed |
| LOC\_Os03g17210 | Expressed protein |
| LOC\_Os05g03440 | Expressed protein |
| LOC\_Os05g41220 | SNF1-related protein kinase regulatory subunit beta-1, putative, expressed |
| LOC\_Os05g45420 | CAMK\_KIN1/SNF1/Nim1\_like\_AMPKh.3 - CAMK includes calcium/calmodulin depedent protein kinases, expressed |
| LOC\_Os06g14420 | Hydrolase NUDIX family, domain containing protein, expressed |
| LOC\_Os09g32010 | Ternary complex factor MIP1, putative , expressed |
| LOC\_Os12g36640 | Universal stress protein domain containing protein, putative, expressed |
| LOC\_Os12g42610 | YABBY domain containing protein, putative, expressed |

Table 1. The function annotations of the genes in the OSH1 network.

**Discussion**

From the gene network we obtained a list of genes potentially interacting with OSH1. The gene functions include metabolite transport (LOC\_Os01g03490), signaling transduction (LOC\_Os05g45420, LOC\_Os01g41010, LOC\_Os05g41220), and stress response (LOC\_Os12g36640). Since OSH1 is involved in the maintenance of undifferentiated status of SAM, these genes are likely involved in the relevant downstream pathways of OSH1.

The peak selection was stract, leading to the discovery of only about a dozen genes in the network, and all these genes were upregulated by OSH1. Less strict selection would discover more genes, among which some might be downregulated, equally interesting. In fact, lowering the selection threshold to (FDR < 5% and fold enrichment > 100) generated a network of 1000+ genes (data not shown). A further GO enrichment analysis would help categorize the gene groups. Moreover, examining the binding sites of OSH1 to these genes can reveal more information about their interactions.

In conclusion, this study built a gene regulatory network of OSH1 based the ChIP-Seq and transcriptomics data, and extracted functional annotations of the downstream genes from OSH1. Future analyses can lower the peak selection threshold to obtain a bigger network and categorize the gene functions.

**References**

1. Steeves, T.A., and Sussex, I.M. (1989). Patterns in Plant Development. (Cambridge, UK: Cambridge University Press)
2. Tsuda K, Ito Y, Sato Y, Kurata N. Positive Autoregulation of a KNOX Gene Is Essential for Shoot Apical Meristem Maintenance in Rice. The Plant Cell. 2011;23(12):4368-4381. doi:10.1105/tpc.111.090050.
3. Tsuda K, Kurata N, Ohyanagi H, Hake S. Genome-Wide Study of KNOXRegulatory Network Reveals Brassinosteroid Catabolic Genes Important for Shoot Meristem Function in Rice. The Plant Cell. 2014;26(9):3488-3500. doi:10.1105/tpc.114.129122.