**­­­­­­­Sorghum Epigenetics Project**

1. **Data Directory**

/sonas-hs/ware/hpc\_norepl/data/xwang/sorghum\_encode/chip\_fq/orig\_fq (will move offline)

1. **Experimental Design**

**Objective**: Identify epigenetic marks associated with root morphology/architecture and abiotic stress tolerance in sorghum.

**Design**: 2 genotypes (G1 and G3) x 2 treatments x 2 reps

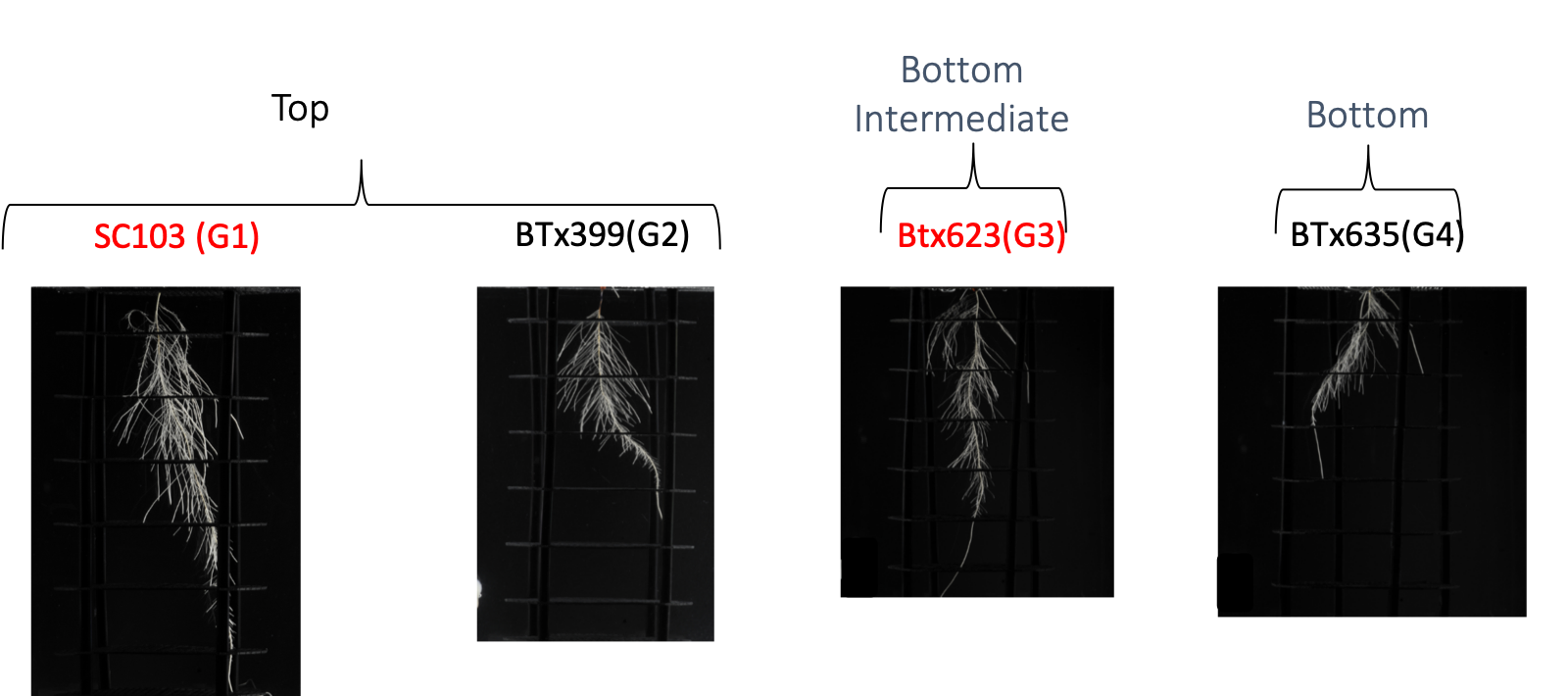
2 abiotic stresses: low and high phosphorus in hydroponics

3 Histone Marks: H3K4me3 (activation mark), H3K27me3 (repressive mark), and H3K27Ac



**Line selection**: key trait is root morphology and architecture but we also considered performance under stress: total root surface area.

**Root architecture of the sorghum lines**: to have root phenotype under phosphorus starving, given this picture showed here is for normal (high) phosphorus, based on the objective.

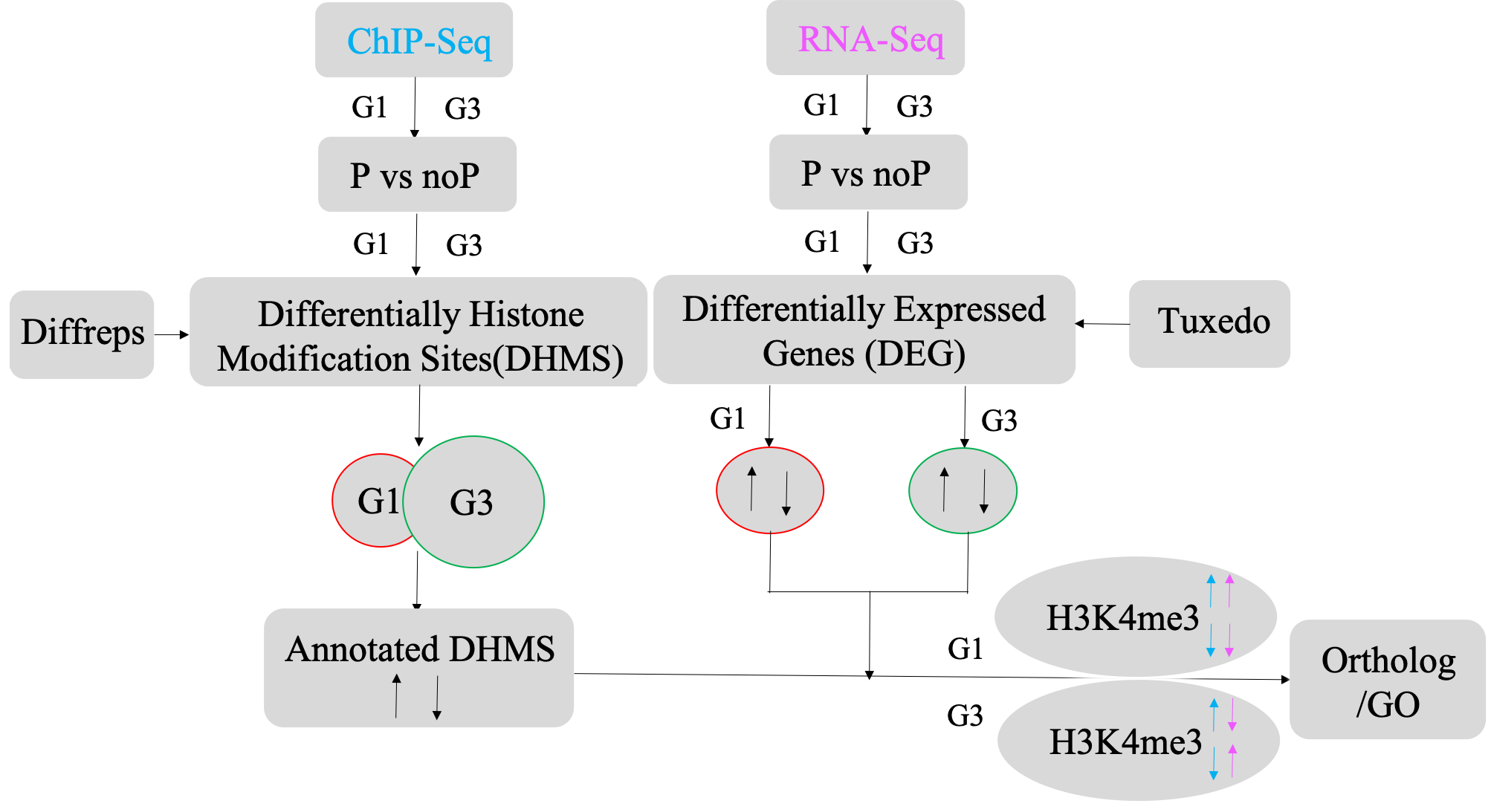


1. **Workflows**

**Workflow for ChIP-Seq**



**Workflow for integrating ChIP-Seq and RNA-Seq** (DEG results are from Kapeel)

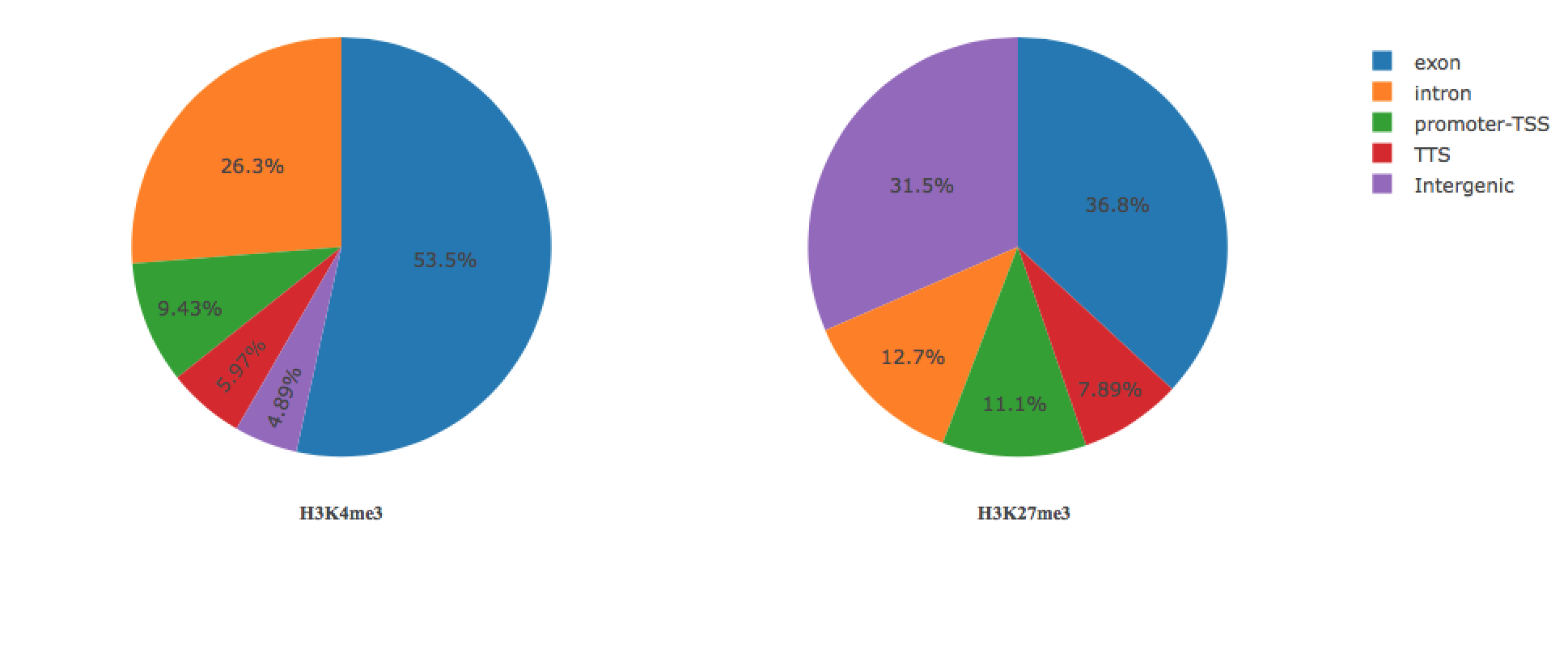
****

1. **Results**

**ChIP-Seq Library Sequencing Alignments and Marks Statistics (V3)**

****

**Distribution of Histone Modifications within Different Regions of the Sorghum Genome**

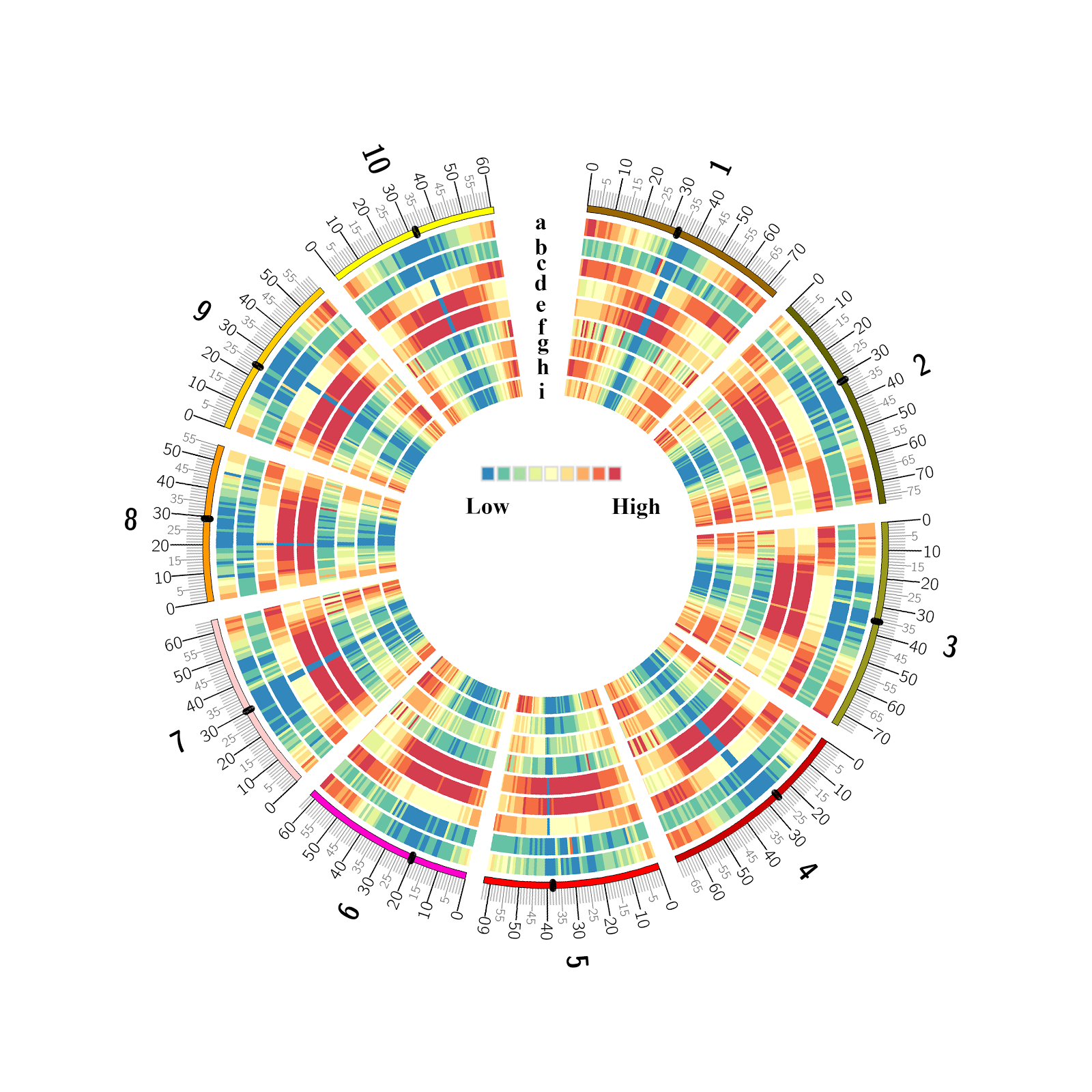
****

**Circle diagram of all 10 chromosomes**

The tracks represent sequence scaffolds with centromeric regions labeled in black.

1. gene density;
2. gene express level;
3. CHH;
4. CHG;
5. CG;
6. srna21;
7. srna24;
8. H3K4me3;
9. H3K27me3

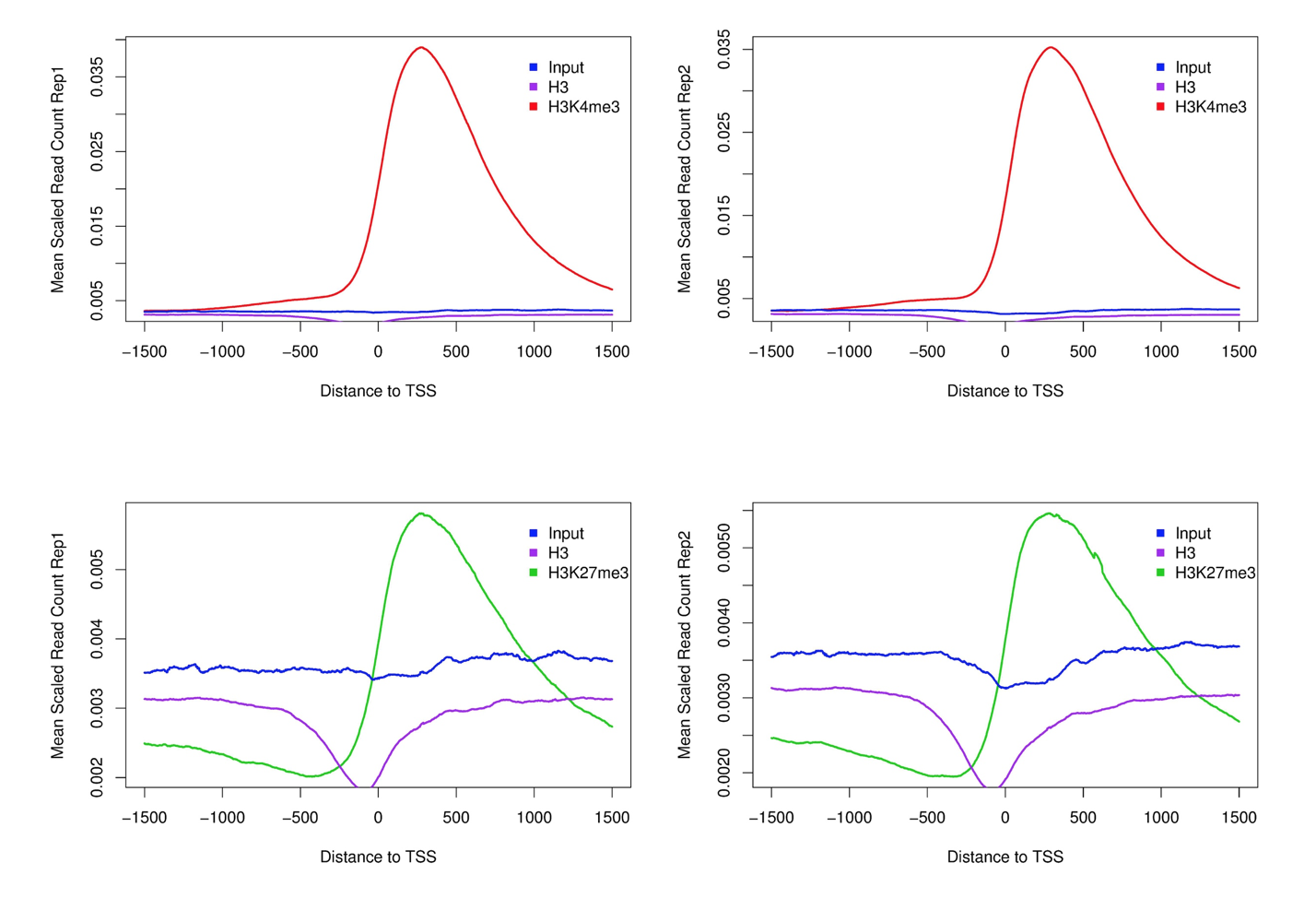
Centromeres Position: Cen38, sorghum-specific centromeric repeat

****

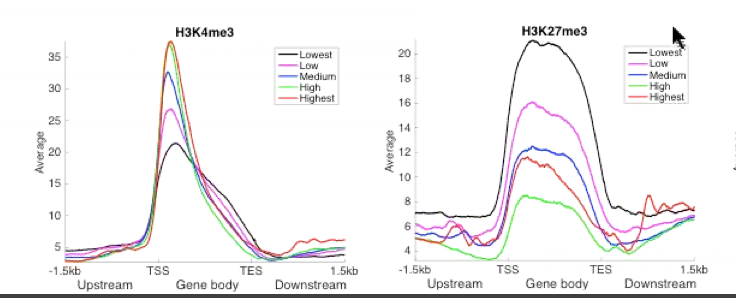
****

**Histone modification profiles across genes (TSS, Gene body, and TES) (G3\_P)**

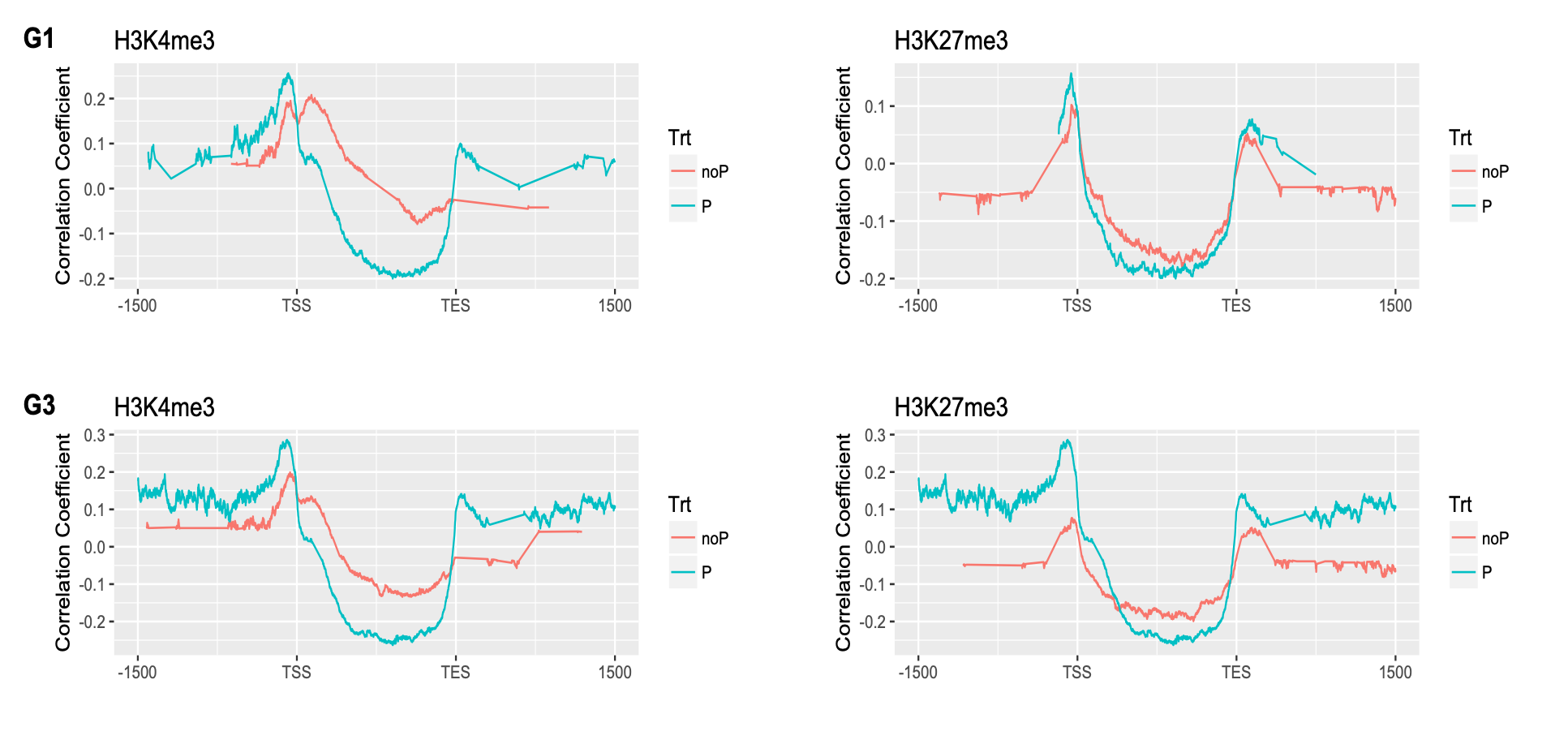
Histone modification profiles for each replicates and histone marks:



Spearman coefficients for correlation of gene expression data (RNA-Seq) and the presence of a modification (Liya):

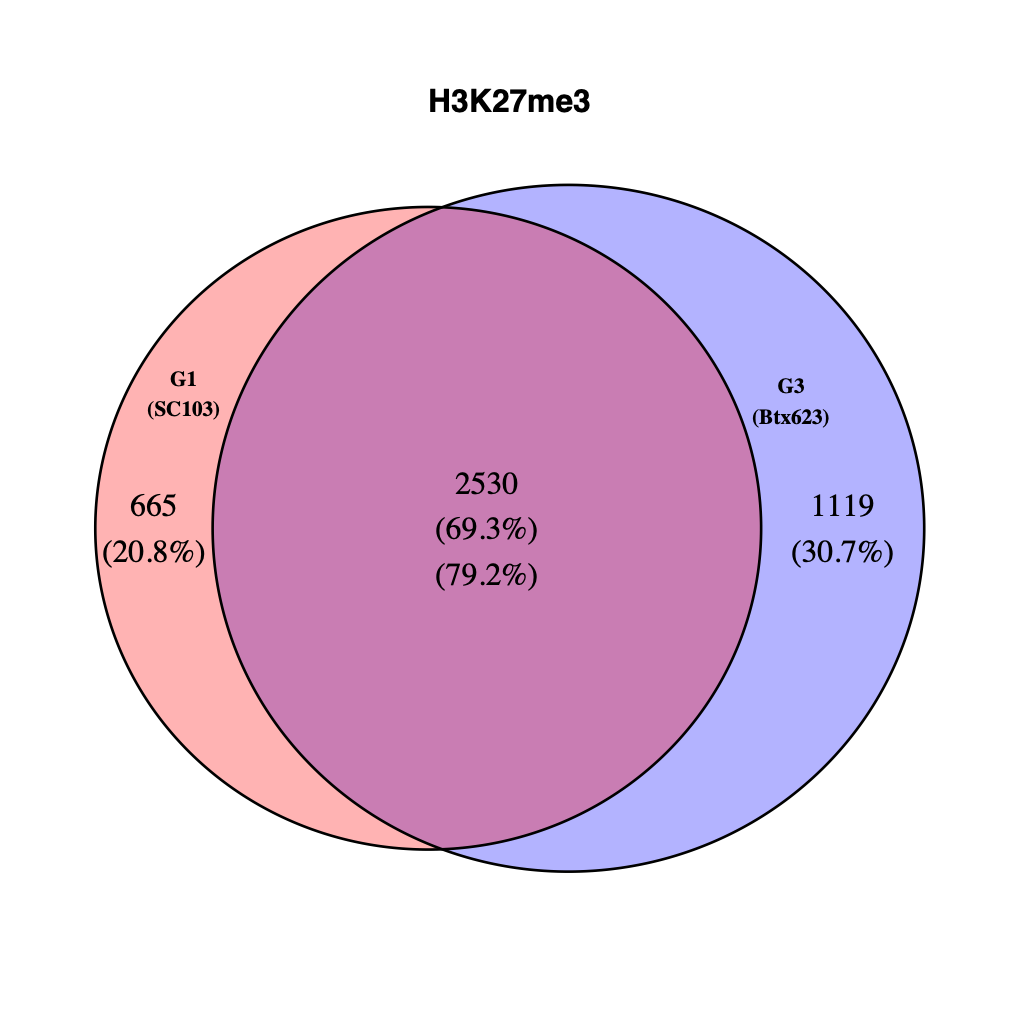


**Correlation of gene expression and the presence of modification**



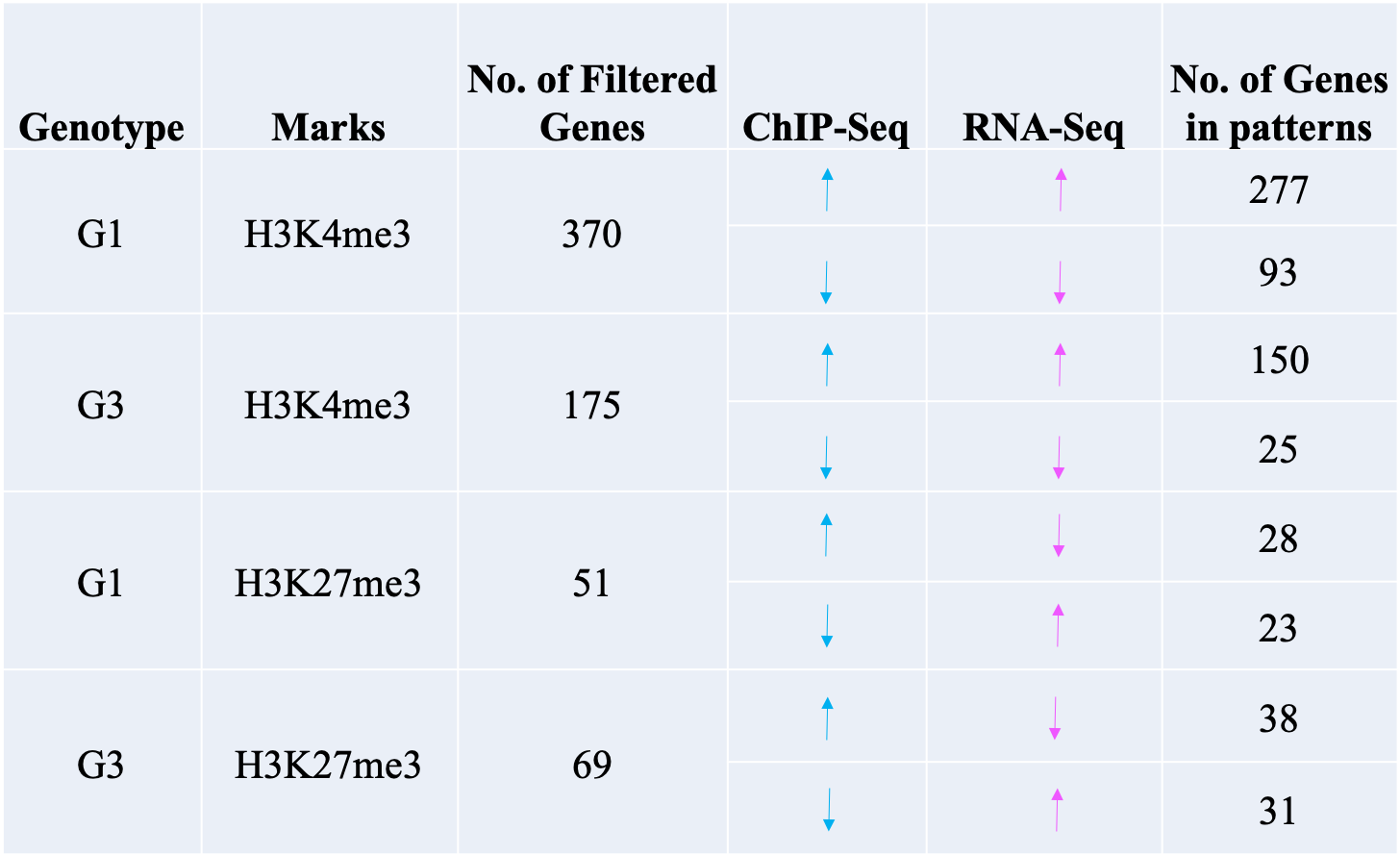
**Epigenetic study related to root system architecture in Sorghum bicolor**

Specified DHMS for each Genotype: The venn diagrams here are showing the specific DHMS for each of genotype eliminating the overlapping peaks between them. Percentages are given of the relative number of DHMS for each part compared to the total number of DHMS for each of genotype.



**Number of filtered genes based on the pipeline strategy**

Table indicating the number of filtered genes depending on the pipeline strategy. Last column is the number of genes in specific pattern, based on the fact that H3K4me3 marks actively transcribed genes and H3K27me3 repressively transcribed genes.



**Ortholog and GO terms (the functional domain included)**

G1\_H3K4me3\_diffPeak\_diffGene\_intersect\_ortholog.xlsx

G1\_H3K27me3\_diffPeak\_diffGene\_intersect\_ortholog.xlsx

G3\_H3K4me3\_diffPeak\_diffGene\_intersect\_ortholog.xlsx

G3\_H3K27me3\_diffPeak\_diffGene\_intersect\_ortholog.xlsx

**Appendix results**

Number of Histone Mark peaks (V1) for several runs to find out why the peaks number for H3K27Ac is pretty low.

Note: For controls they use both input (genomic DNA) and H3. But there is no significant difference in results between them, so we move forward calling peaks against unmodified H3 as background (double check). The above table (using V1 reference) is the peaks number results based on both H3 and input as control.

There are several runs to find out why the peaks number for H3K27Ac is pretty low. No matter which peak caller is used, the number is low for H3K27Ac, especially for G3\_noP\_K27Ac\_rep1 and G3\_P\_K27Ac\_rep2.

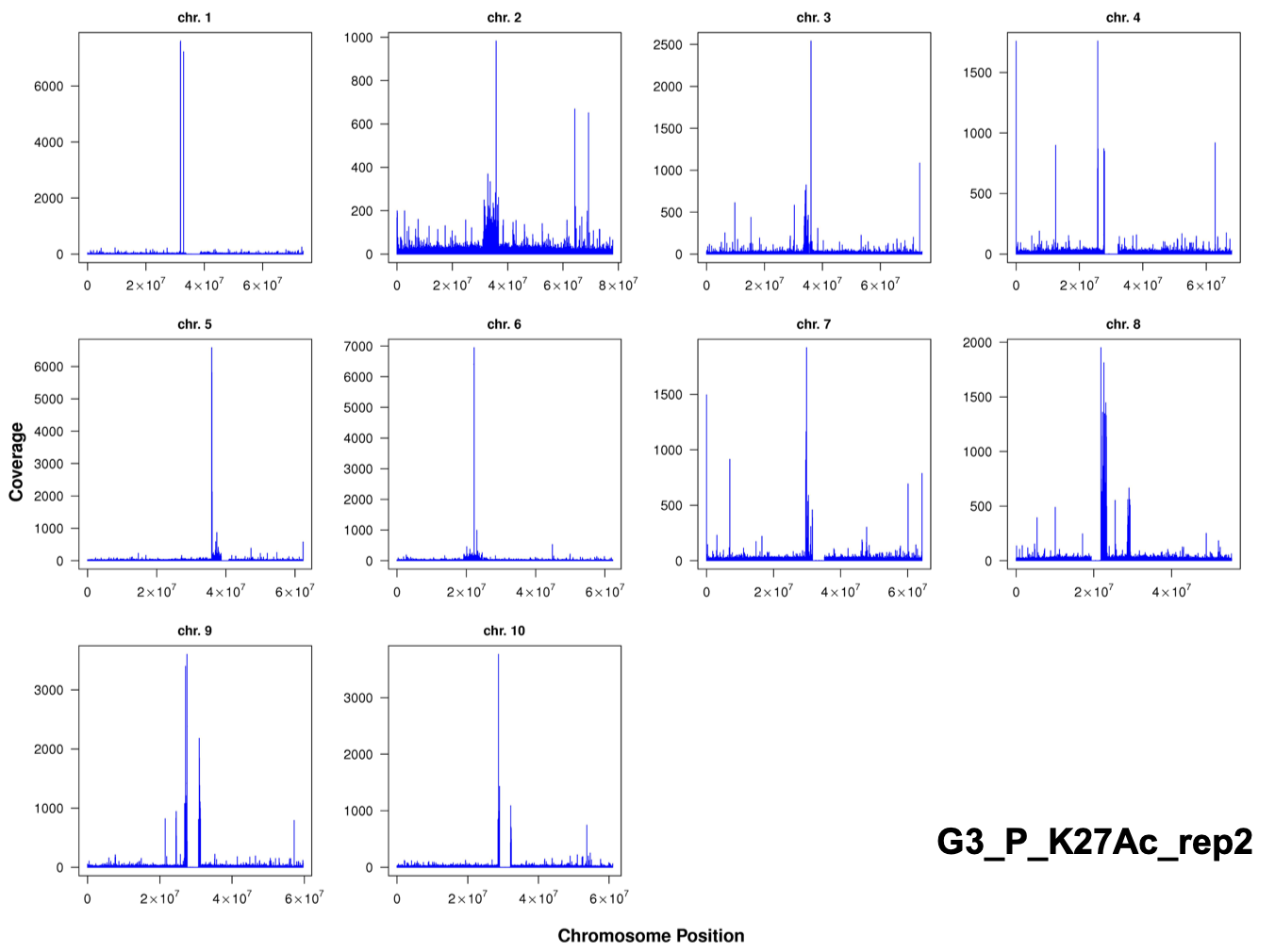
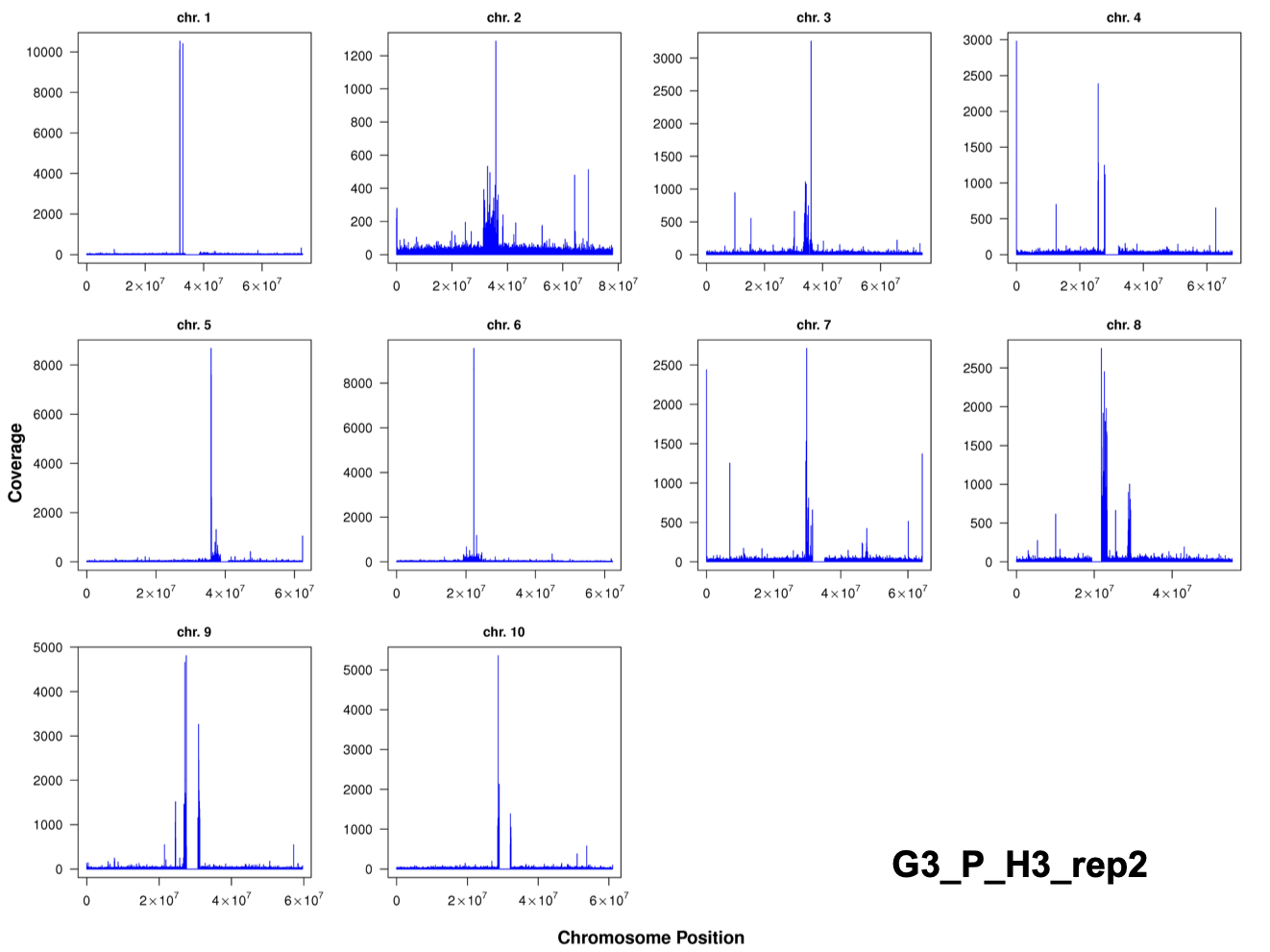
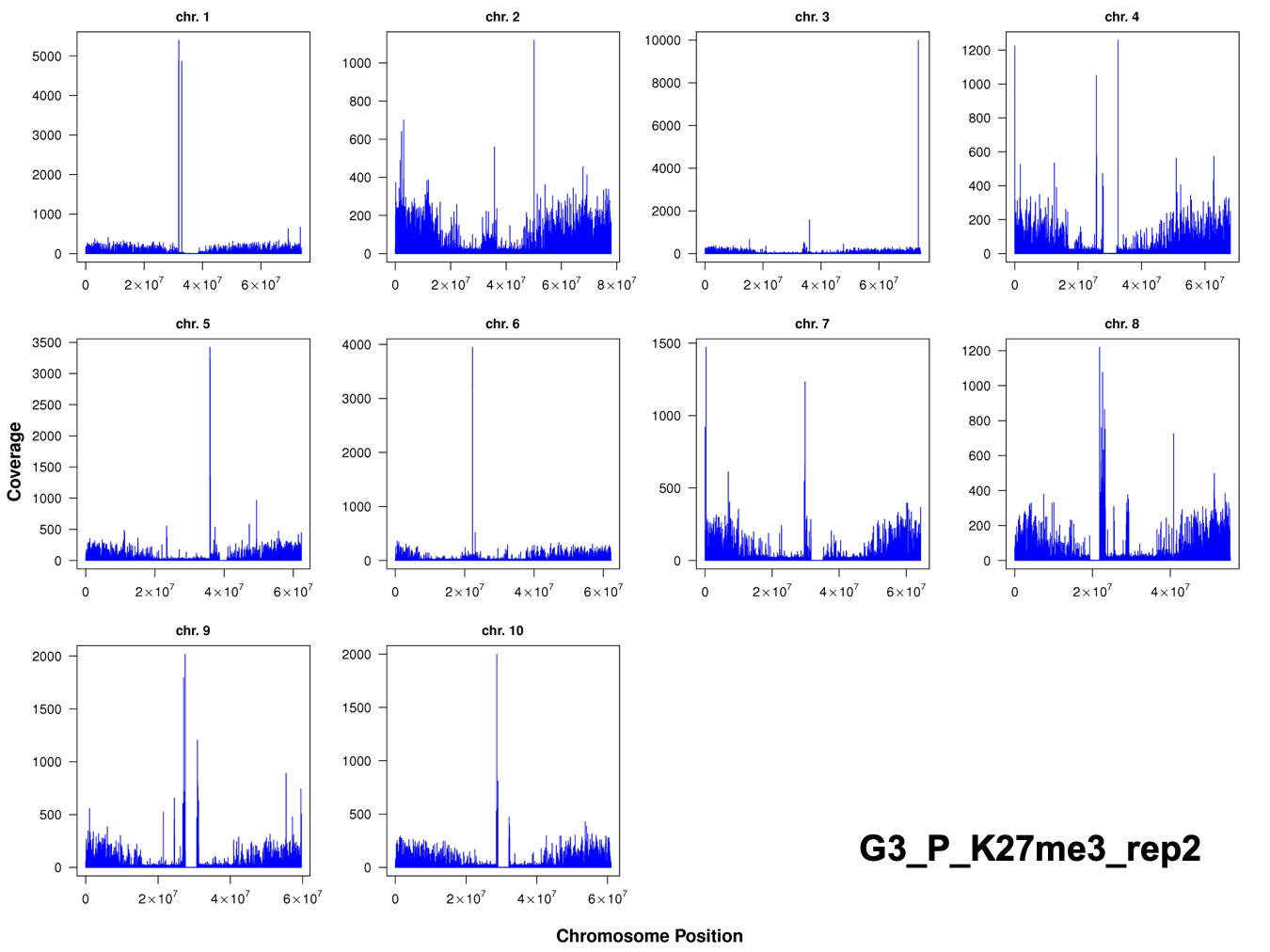


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | peak size | minDist | -region | -style histone |
| 2nd run | 1000 | 2500 | Yes | No |
| 3rd run | Variable | Variable | Yes | No |
| 4th run | 500 | 1000 | Yes | Yes |
| 5th run | 1000 | 2500 | Yes | Yes |

**Trouble shooting K27Ac**

I summarized the statistics for mapped reads (alignments), but did not find there is abnormal things. But, the total tags in peaks are very little, refer to the peaks stats table above, suggesting that there are no enrichment on expected peaks, especially for G3\_noP\_K27Ac\_rep1 and G3\_P\_K27Ac\_rep2.

Then, I plot genome coverage for 3 samples, one is with positive signals (G3\_P\_K27me3\_rep2), one is with negative signals (G3\_P\_K27Ac\_rep2), and one is the H3 control (G3\_P\_H3\_rep2). The plot for histone marks are expected to have spikes across the genome like G3\_P\_K27me3\_rep2 as below. But, it turns out the plot of G3\_P\_K27Ac\_rep2 is in very similar pattern to G3\_P\_H3\_rep2 (control), suggesting that it is highly possible that there is no enrichment for G3\_P\_K27Ac\_rep2.



Build genome browser tracks (To bigwig)

Bigwig files:

/sonas-hs/ware/hpc\_norepl/data/xwang/sorghum\_encode/macs2\_peaks/K27Ac\_bigwig

Refer to the URL:

<https://genome.ucsc.edu/goldenpath/help/twoBit.html>

<https://github.com/taoliu/MACS/wiki/Build-Signal-Track#Fix_the_bedGraph_and_convert_them_to_bigWig_files>