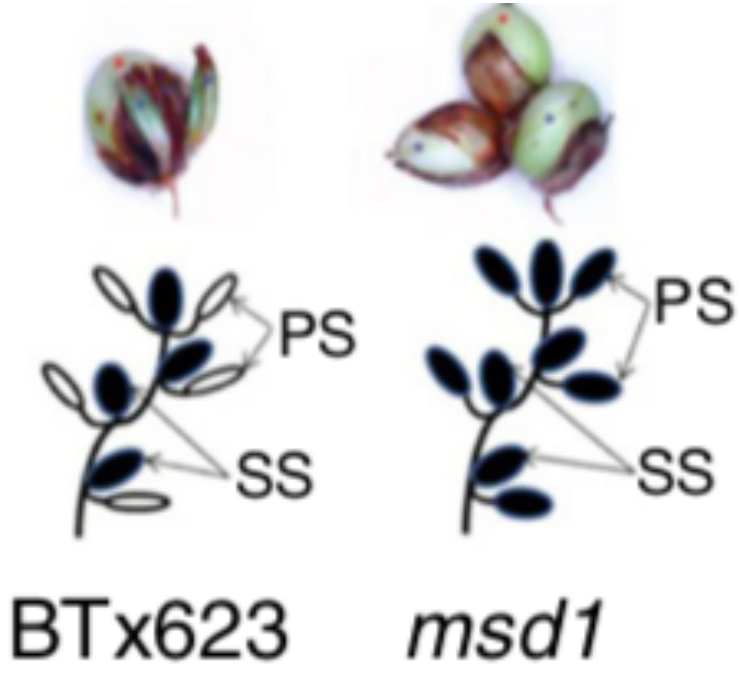
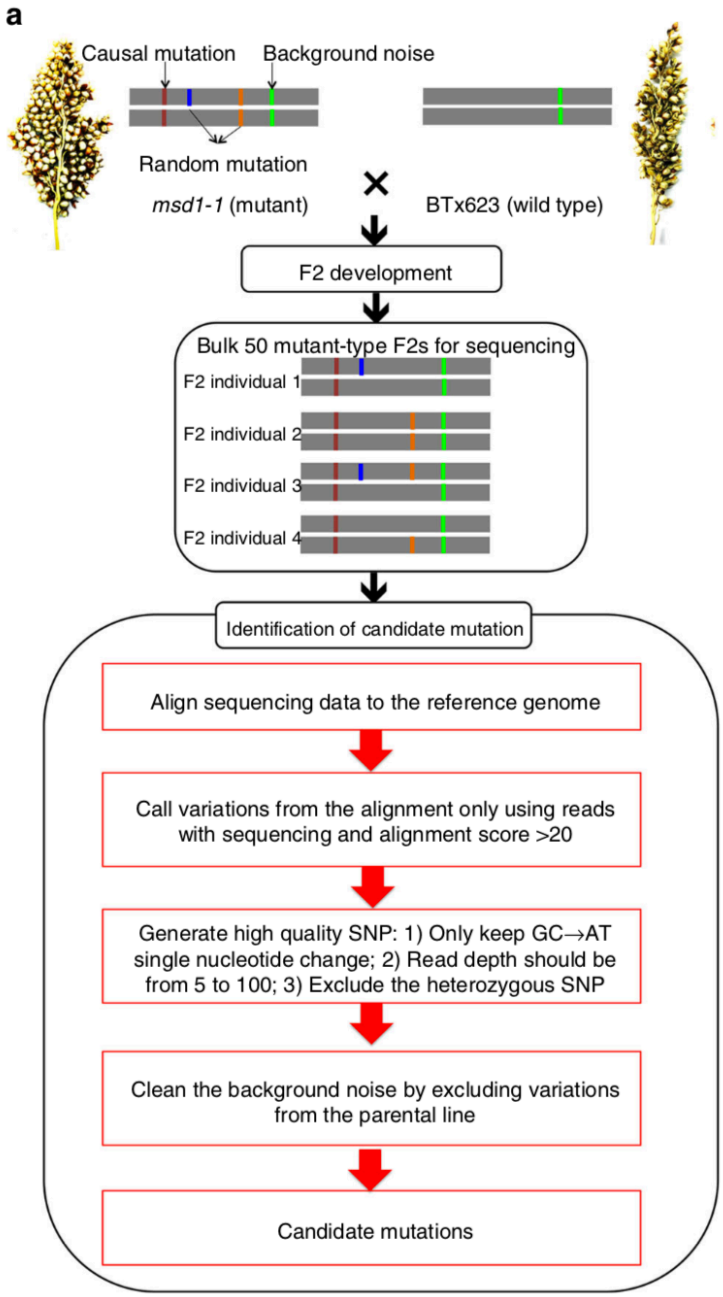
MSD1 regulates pedicellate spikelet fertility in sorghum through the jasmonic acid pathway

ABE516 Project 2
Chen Chen and Ken Youens-Clark

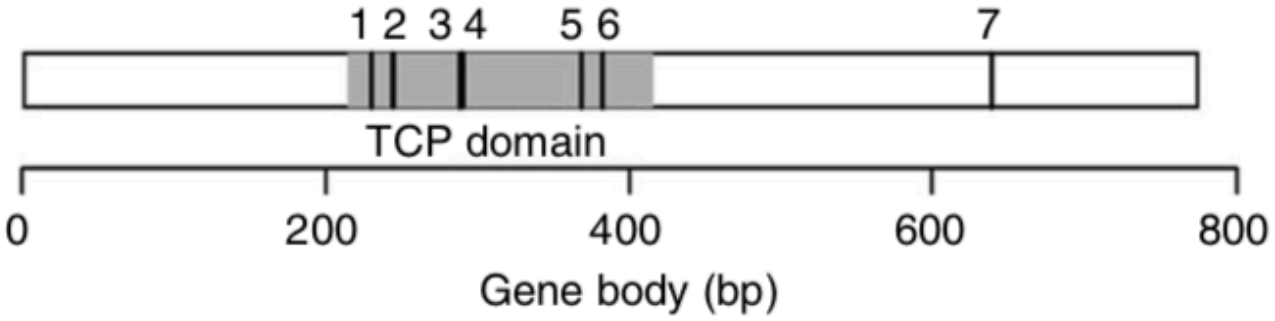
Sorghum

BTx623 v *msd1*



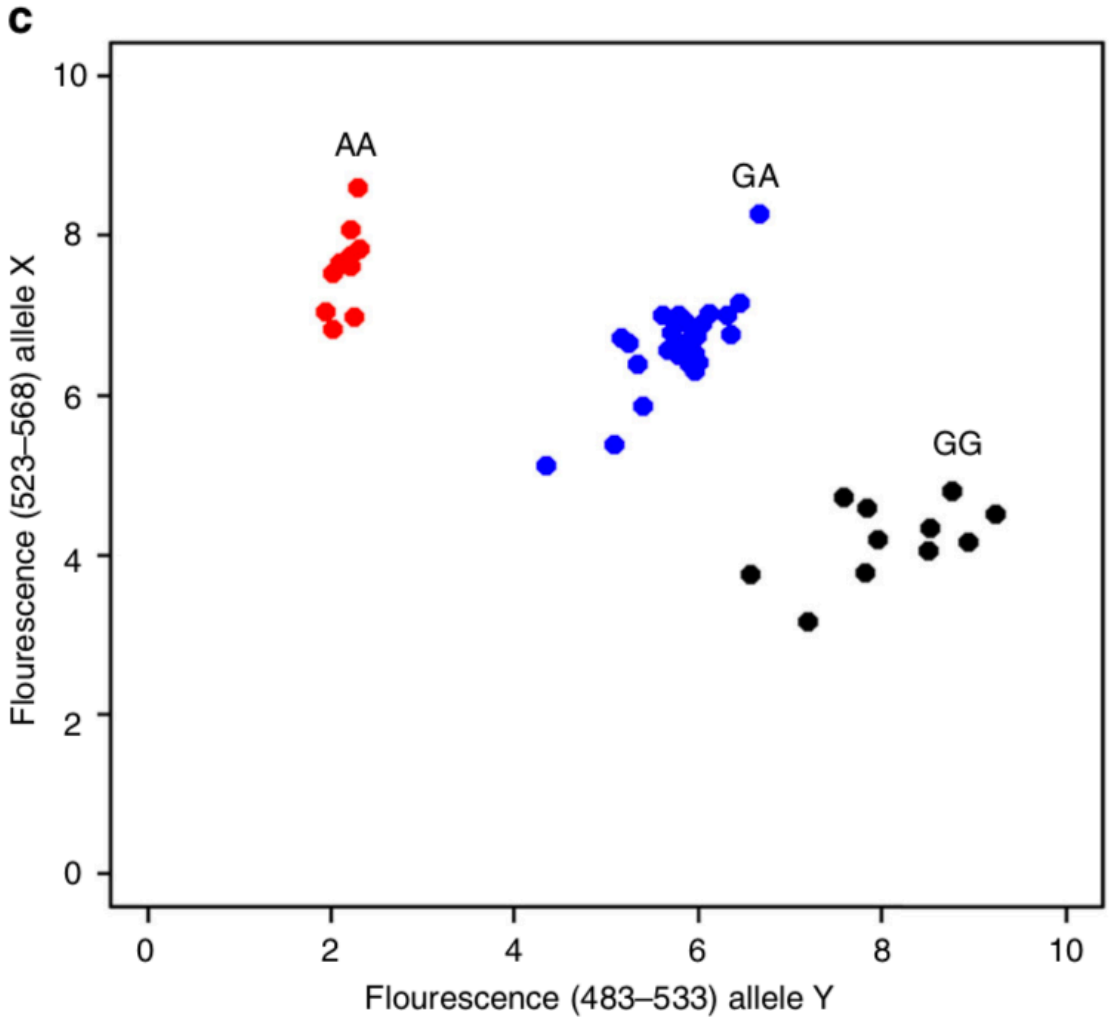

a
Causal mutation
Background noise
Random mutation
msd1-1 (mutant)
BTx623 (wild type)
F2 development
Bulk 50 mutant-type F2s for sequencing
F2 individual 1
F2 individual 2
F2 individual 3
F2 individual 4
Identification of candidate mutation
Align sequencing data to the reference genome
Call variations from the alignment only using reads with sequencing and alignment score >20
Generate high quality SNP: 1) Only keep GC→AT single nucleotide change; 2) Read depth should be from 5 to 100; 3) Exclude the heterozygous SNP
Clean the background noise by excluding variations from the parental line
Candidate mutations

Identification of SNPs in MSD1 gene

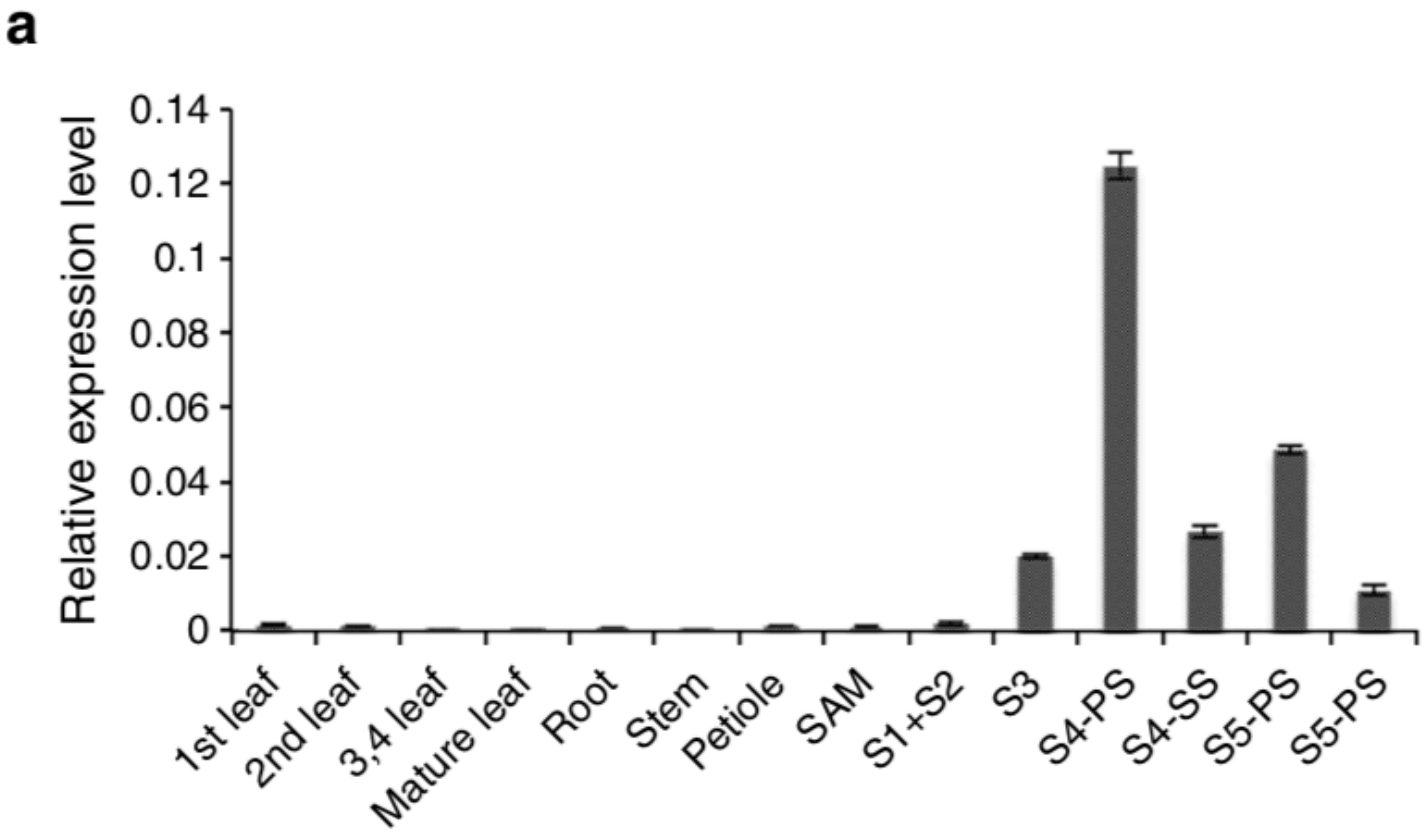


SNP ID	Mutant	Chromosome	Location	Variation	Mutation effect	Amino acid change
1	*msd1-7*	7	54912926	G/A	Missense_variant	R/Q
2	*msd1-5*	7	54912940	G/A	Missense_variant	D/N
3	*msd1-6*	7	54912985	C/T	Missense_variant	R/W
4	*msd1-1*	7	54912986	G/A	Missense_variant	R/Q
5	*msd1-2*	7	54913064	C/T	Missense_variant	T/M
6	*msd1-3*	7	54913078	C/T	Missense_variant	L/F
7	*msd1-4*	7	54913335	G/A	Stop gained	W/*

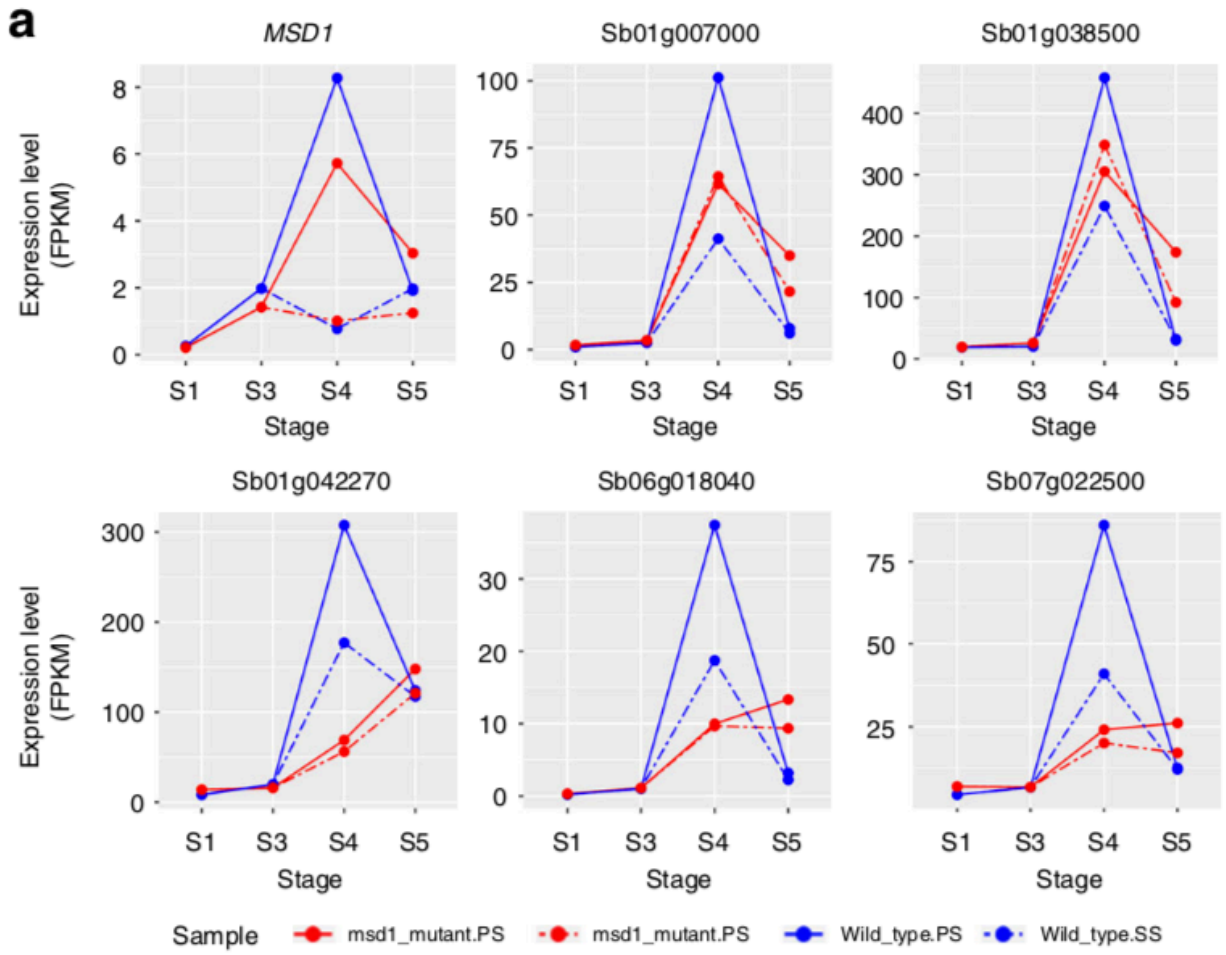
Homozygous mutation (AA) mutation leads to *msd1* panicles



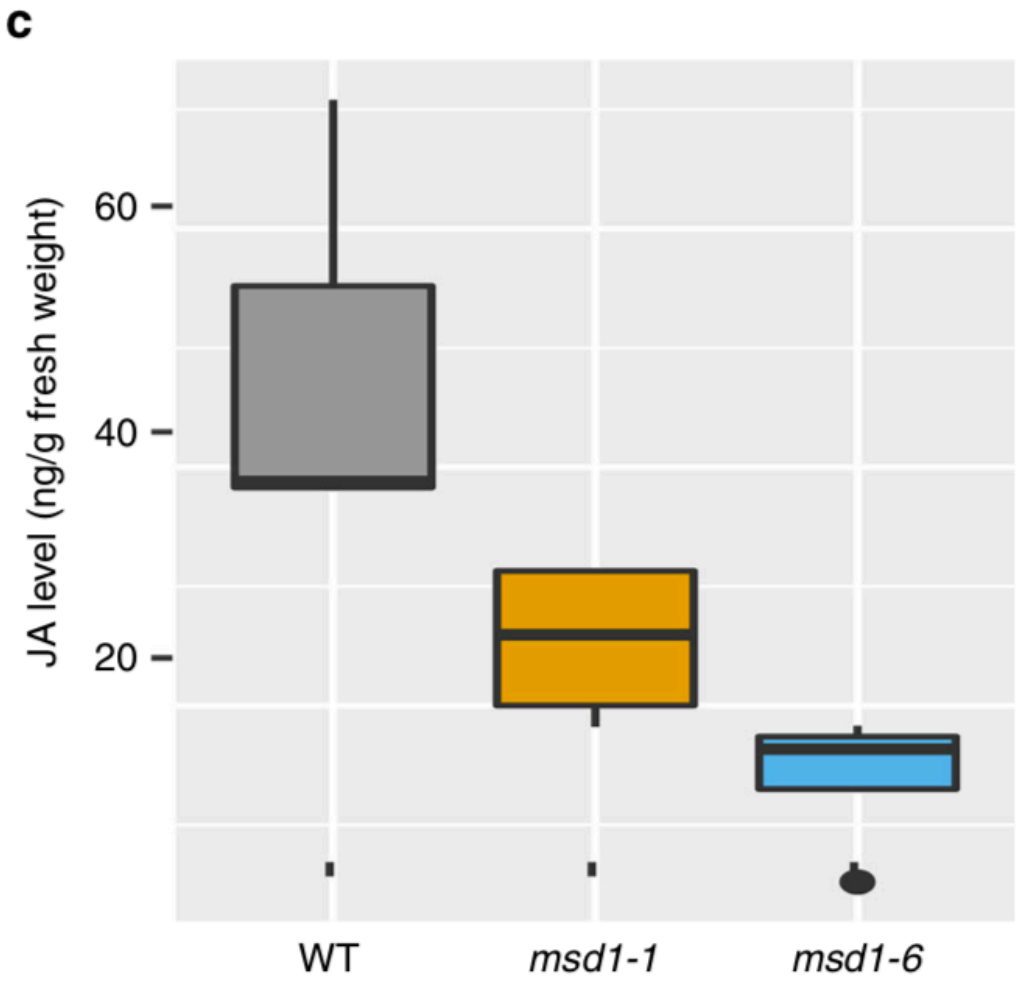
Relative expression of MSD1 from qRT-PCR in various tissue types



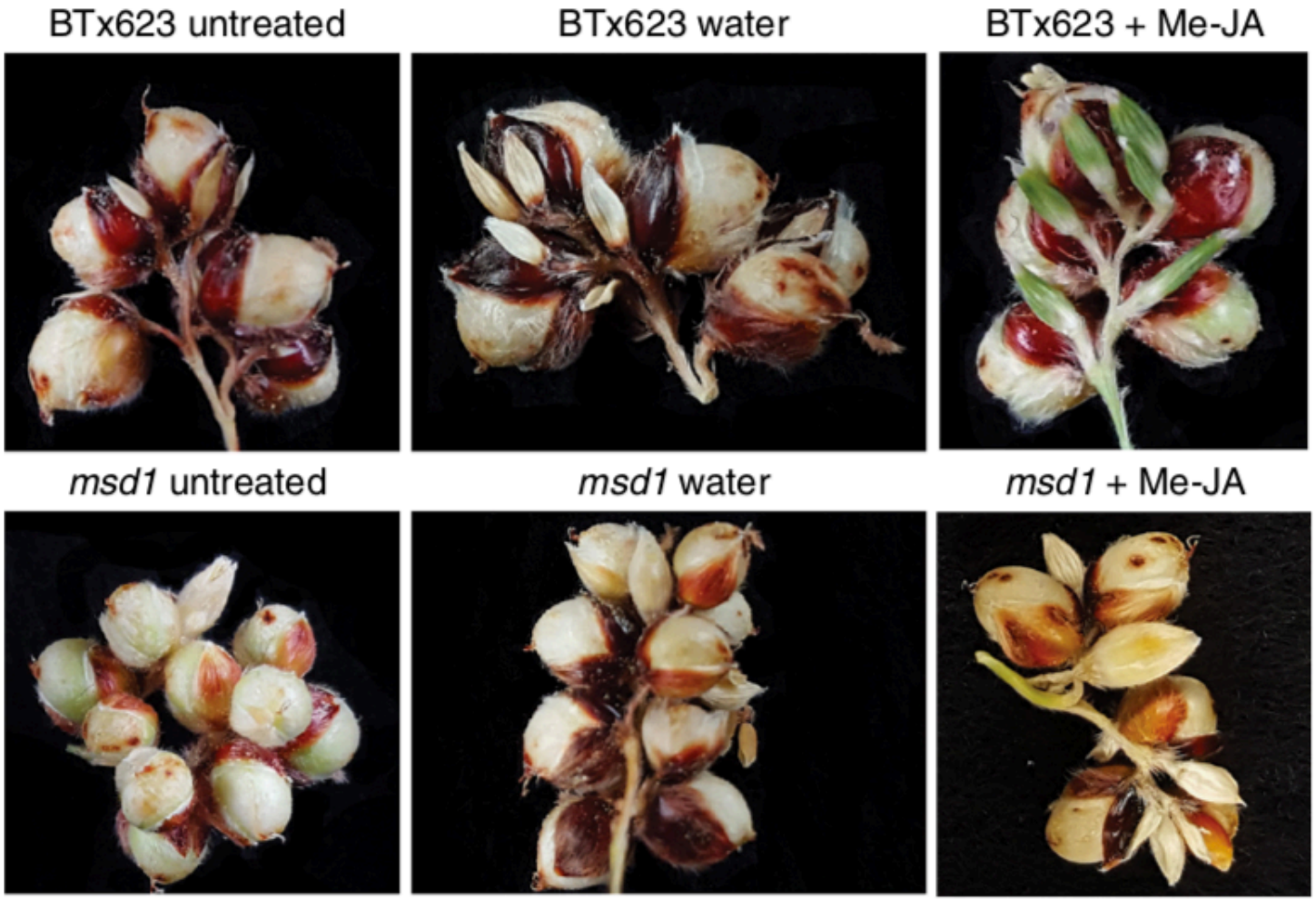
Expression pattern of 5 target genes involved in JA synthesis

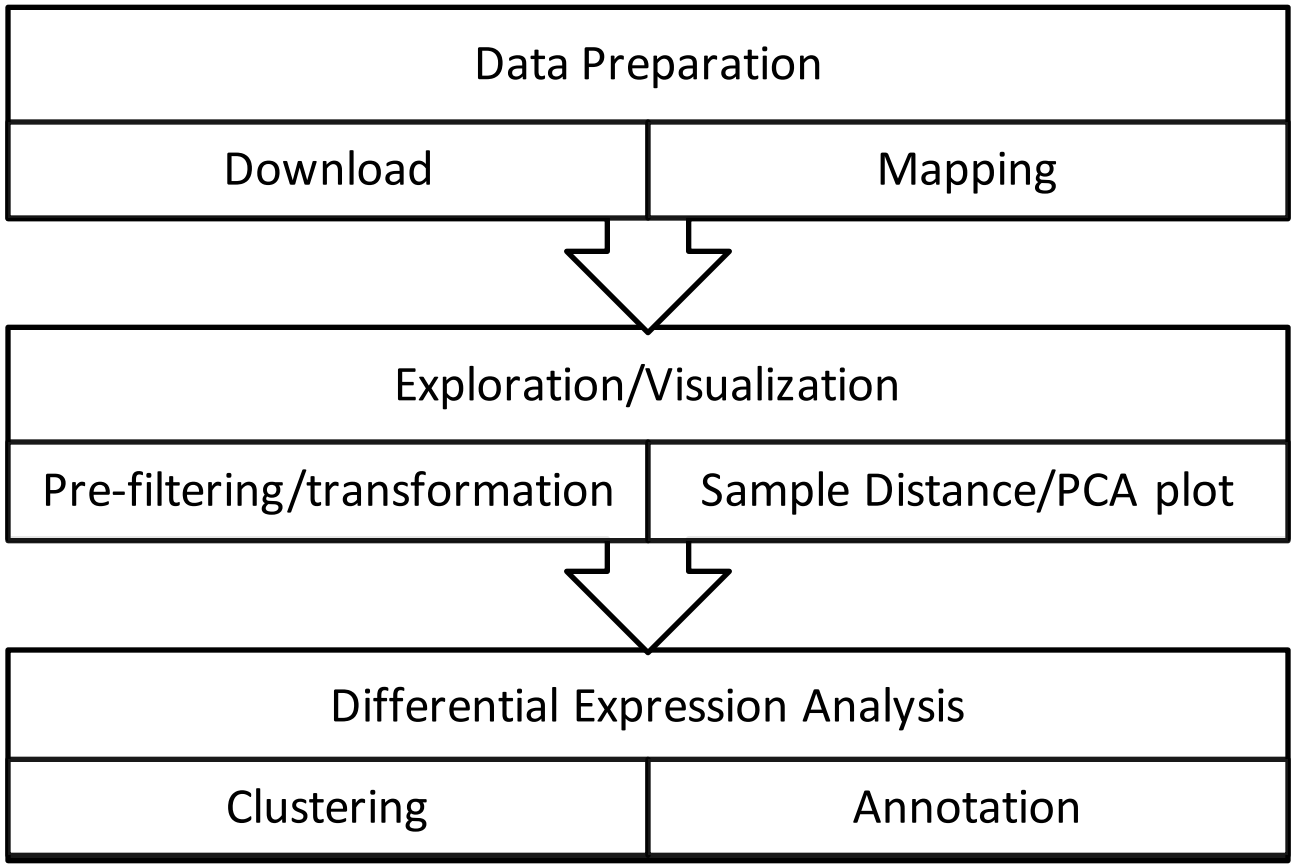


JA levels are significantly different in WT v *msd1*



msd1 can be "rescued" with application of JA




Data Preparation
Download
Mapping
Exploration/Visualization
Pre-filtering/transformation
Sample Distance/PCA plot
Differential Expression Analysis
Clustering
Annotation

Download SRA data from NCBI website and use SRAToolkit to convert them to fastq format.

Quality Check

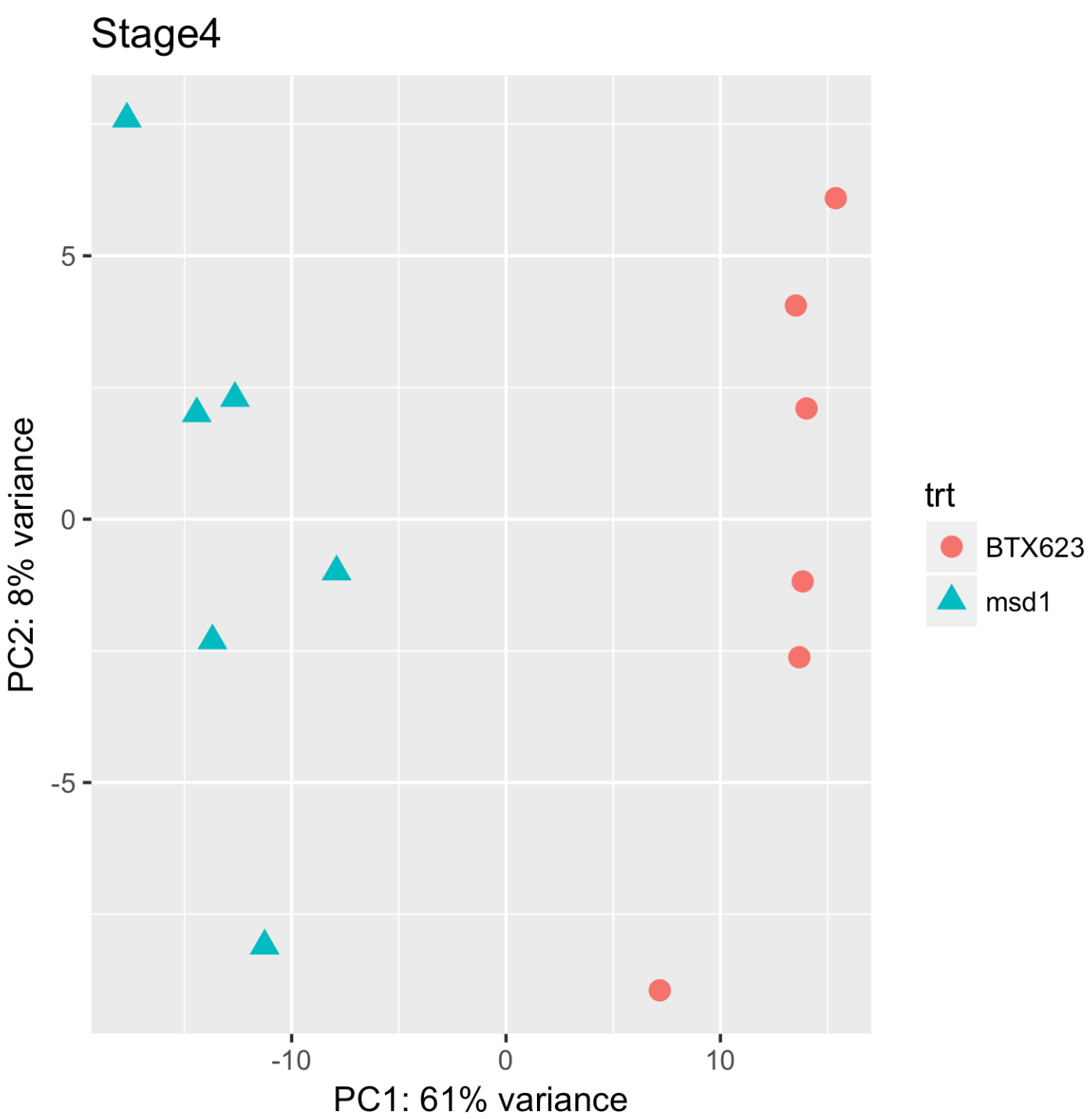
Basic Statistics

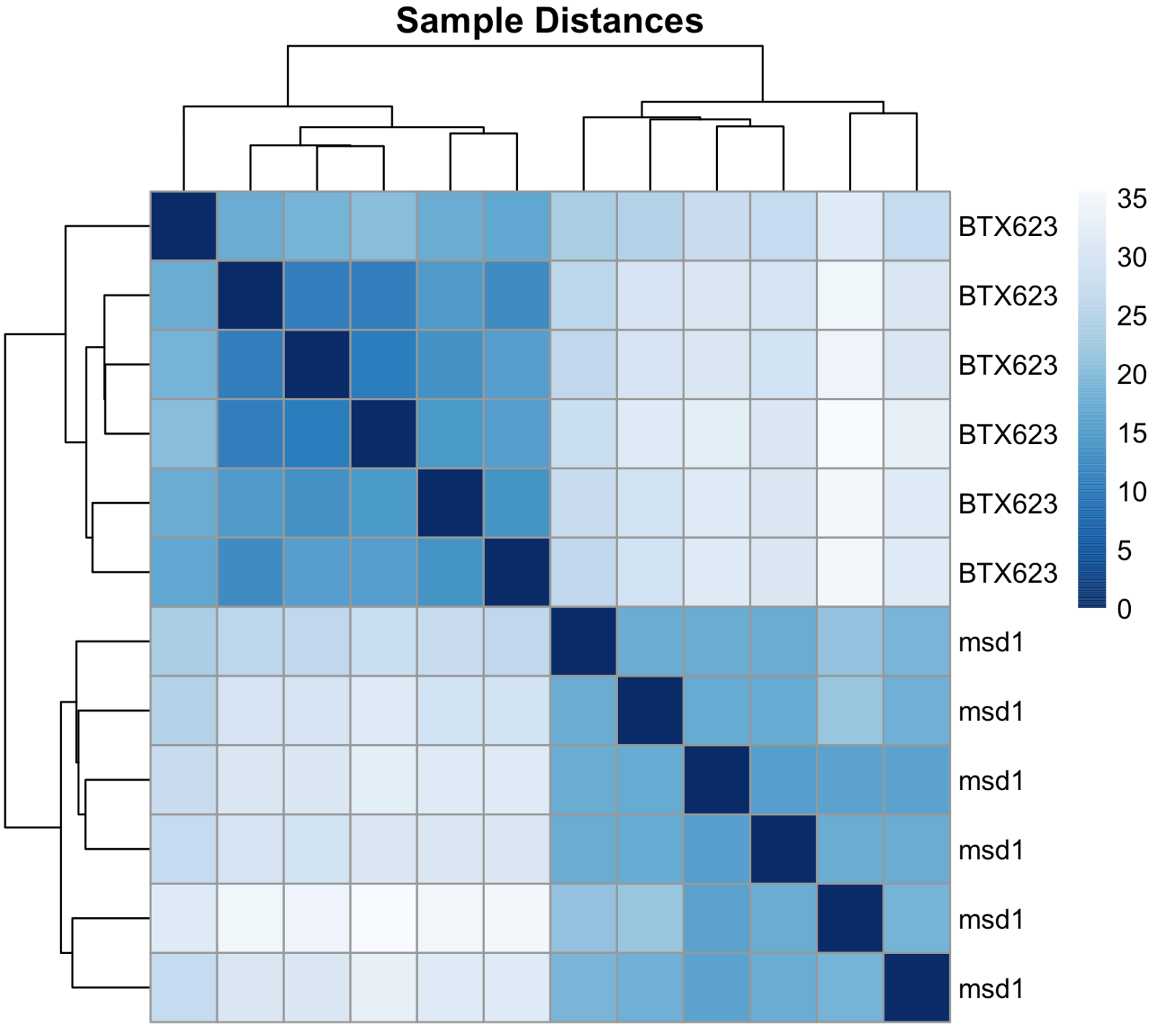
Measure	Value
Filename	SRR6431607.fastq
File type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	35654668
Sequences flagged as poor quality	0
Sequence length	250
%GC	44

We used the STAR read aligner (Dobin et al. 2013) to align the reads for our current experiment to the Sorghum Bicolor reference genome.

BTX623 (wild type)	Msd1 (mutant)
SRR6431605.bam	SRR6431615.bam
SRR6431606.bam	SRR6431616.bam
SRR6431607.bam	SRR6431621.bam
SRR6431608.bam	SRR6431622.bam
SRR6431609.bam	SRR6431627.bam
SRR6431610.bam	SRR6431628.bam

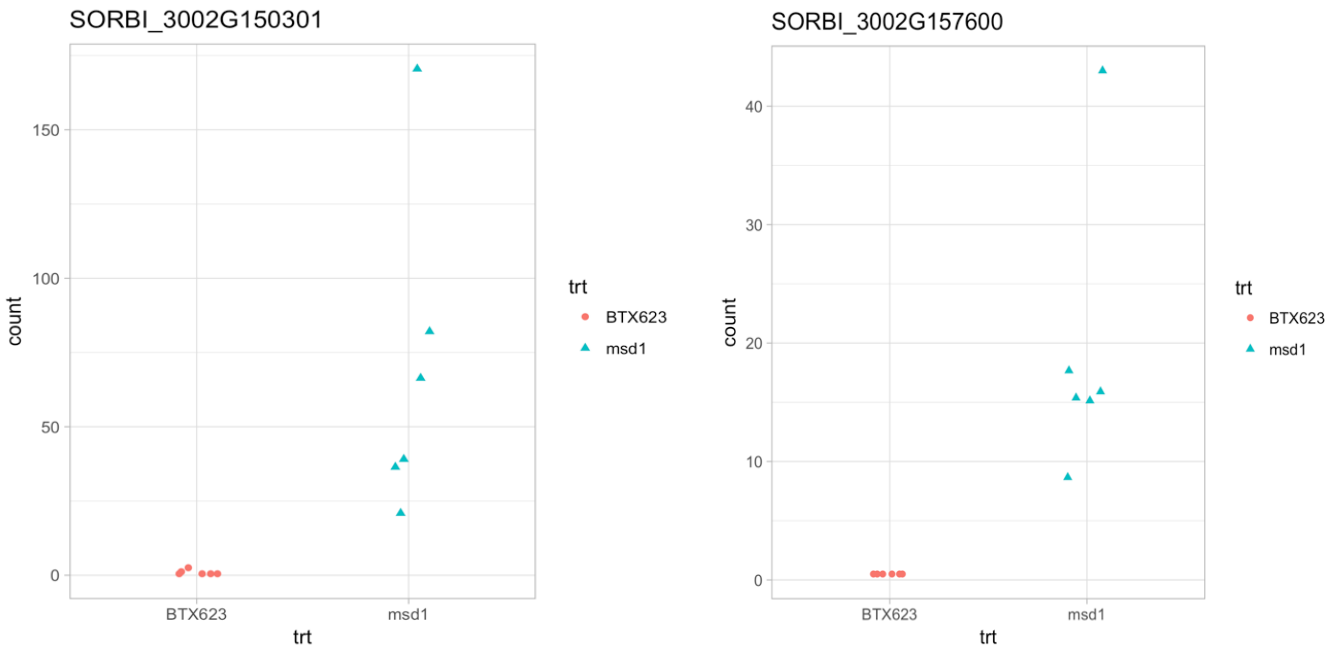
For this analysis, we used 2M sequences for each of the 12 samples.


Stage4
5
0
-5
PC2: 8% variance
-10
0
10
PC1: 61% variance
trt
BTX623
msd1

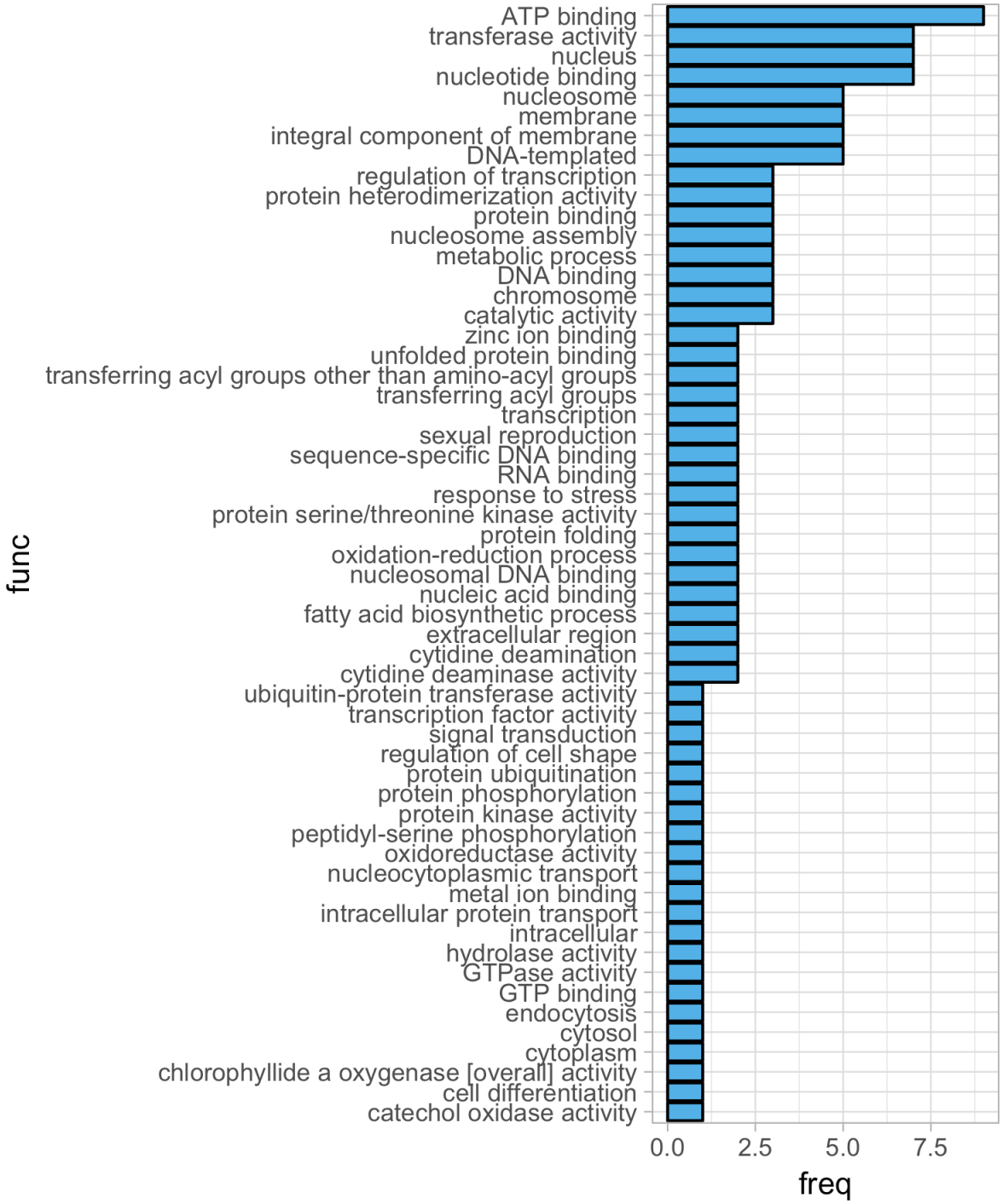

Sample Distances
BTX623
BTX623
BTX623
BTX623
BTX623
BTX623
msd1
msd1
msd1
msd1
msd1
msd1
35
30
25
20
15
10
5
0

We used DESeq2 to do differential expression analysis with FDR=0.05.

With 556 non-zero read counts	
LFC>0 (up-regulated)	73 (13%)
LFC<0 (down-regulated)	60 (11%)




Gene Clusters
SORBI_3010G209100
SORBI_3010G159900
SORBI_3001G301500
SORBI_3001G301600
SORBI_3005G125000
SORBI_3005G124200
SORBI_3003G178500
SORBI_3001G420100
SORBI_3001G418600
SORBI_3002G150301
SORBI_3010G144700
SORBI_3002G157600
SORBI_3009G096850
SORBI_3010G140500
SORBI_3001G371500
SORBI_3004G144500
SORBI_3004G113850
SORBI_3002G150100
SORBI_3007G098950
SORBI_3008G155000
SORBI_3001G261250
SORBI_3008G075350
ENSRNA049485572
SORBI_3010G047200
SORBI_3010G047100
4
2
0
-2
-4
BTX623
BTX623
BTX623
BTX623
BTX623
BTX623
msd1
msd1
msd1
msd1
msd1
msd1


ATP binding
transferase activity
nucleus
nucleotide binding
nucleosome
membrane
integral component of membrane
DNA-templated
regulation of transcription
protein heterodimerization activity
protein binding
nucleosome assembly
metabolic process
DNA binding
chromosome
catalytic activity
zinc ion binding
unfolded protein binding
transferring acyl groups other than amino-acyl groups
transferring acyl groups
transcription
sexual reproduction
sequence-specific DNA binding
RNA binding
response to stress
protein serine/threonine kinase activity
protein folding
oxidation-reduction process
nucleosomal DNA binding
nucleic acid binding
fatty acid biosynthetic process
extracellular region
cytidine deamination
cytidine deaminase activity
ubiquitin-protein transferase activity
transcription factor activity
signal transduction
regulation of cell shape
protein ubiquitination
protein phosphorylation
protein kinase activity
peptidyl-serine phosphorylation
oxidoreductase activity
nucleocytoplasmic transport
metal ion binding
intracellular protein transport
intracellular
hydrolase activity
GTPase activity
GTP binding
endocytosis
cytosol
cytoplasm
chlorophyllide a oxygenase [overall] activity
cell differentiation
catechol oxidase activity
func
0.0
2.5
5.0
7.5
freq

Conclusion:

Using the first 10 million sequences out of 36 million sequences, we detected 133 differentially expressed genes between stage-4 BTX623 and msd1 samples.

Among them, 73 genes are upregulated and 60 are downregulated.

With clustering method and gene ontology we found some interesting patterns.

Clustering of samples is very consistent with the background information.

Further analysis like pathway analysis and network analysis is needed.