Galaxy Community Conference 2012

Integration of S-MART, a toolbox to aid RNA-seq data analysis in Galaxy

URGI INRA Versailles
yufei.luo@versailles.inra.fr

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT

INRA

High-throughput sequencing technologies

- Millions of reads
- Low cost
- cDNA samples (RNA-seq) :
  - transcriptome of eukaryotic genomes
  - gene expression measurement
  - unbiased and comprehensive manner for analyzing transcriptome
- Mapping high-throughput sequencing tools
- Mapped reads analysis tools?

URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART[1]

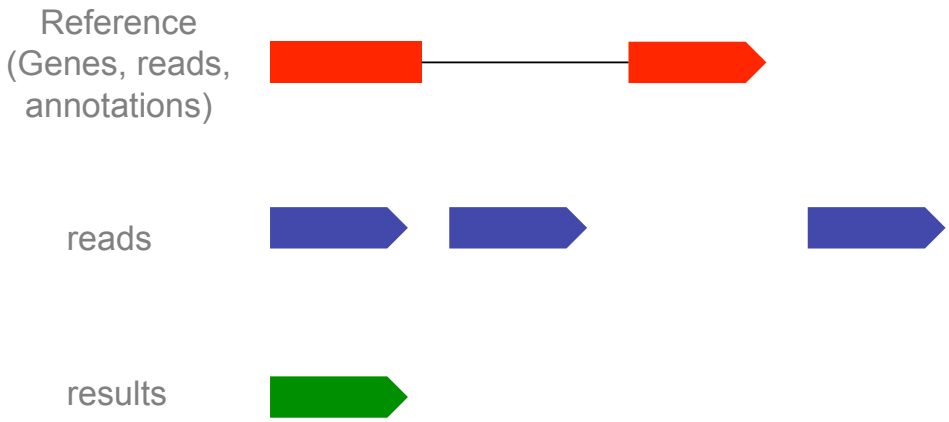
--analysis of mapped RNA-seq and CHIP-seq

- Conversion : Gff*, csv, sam, fastq, fasta, …
- WIG : exploit Wig information
- Merge
- Comparison
- Selection : Exon, Intron, Flanking, …
- Modification : genomic coordinates, sequence, adaptor
- Visualization

- [1] : *Zytnicki M, Quesneville H (2011) S-MART, A Software Toolbox to Aid RNA-seq Data Analaysis. PLoS ONE 6 (10):e25988.doi:10.1371/journal.pone.0025988*

S-MART
--Tools for RNA-seq

CompareOverlapping :

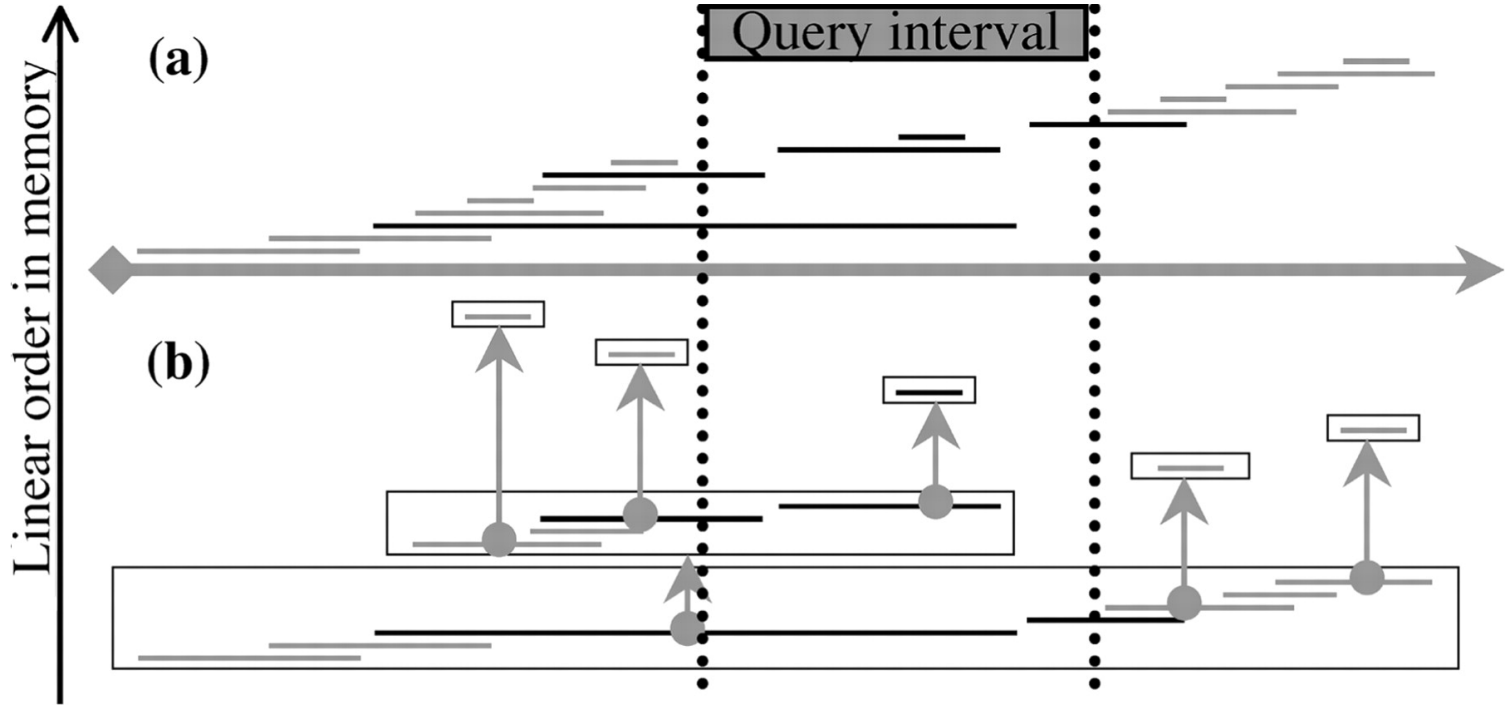


Faster, Lighter Algorithm

URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART
--Tools for RNA-seq

Nested Containment List (NCList)[2]



Storage and querying for interval overlap.

Alekseyenko A V , Lee C J Bioinformatics 2007;23:1386-1393

[2] A.Alekseyenko & J.Lee Nested; Containment List (NCList) : a new algorithm for acceleration interval query of genome alignment and interval databases; Jan 18, 2007; Bioinformatics

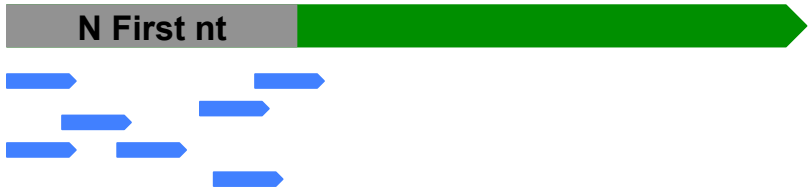
S-MART

--Tools for RNA-seq

CompareOverlapping : O(N), where N is the number of short reads. 1h30->20 million reads VS. 250 thousand ref transcripts.

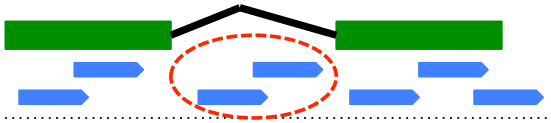
Options :

Restrict to N first nucleotides :

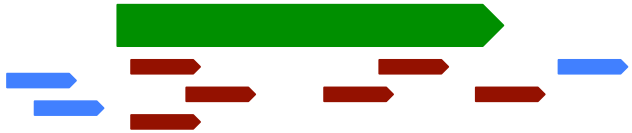


Extension on 5' or 3' direction :

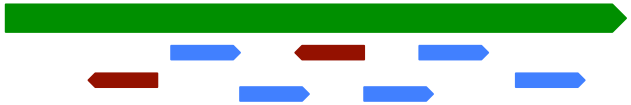
Report introns :



Invert selection :



Colinear/Anti-sense :



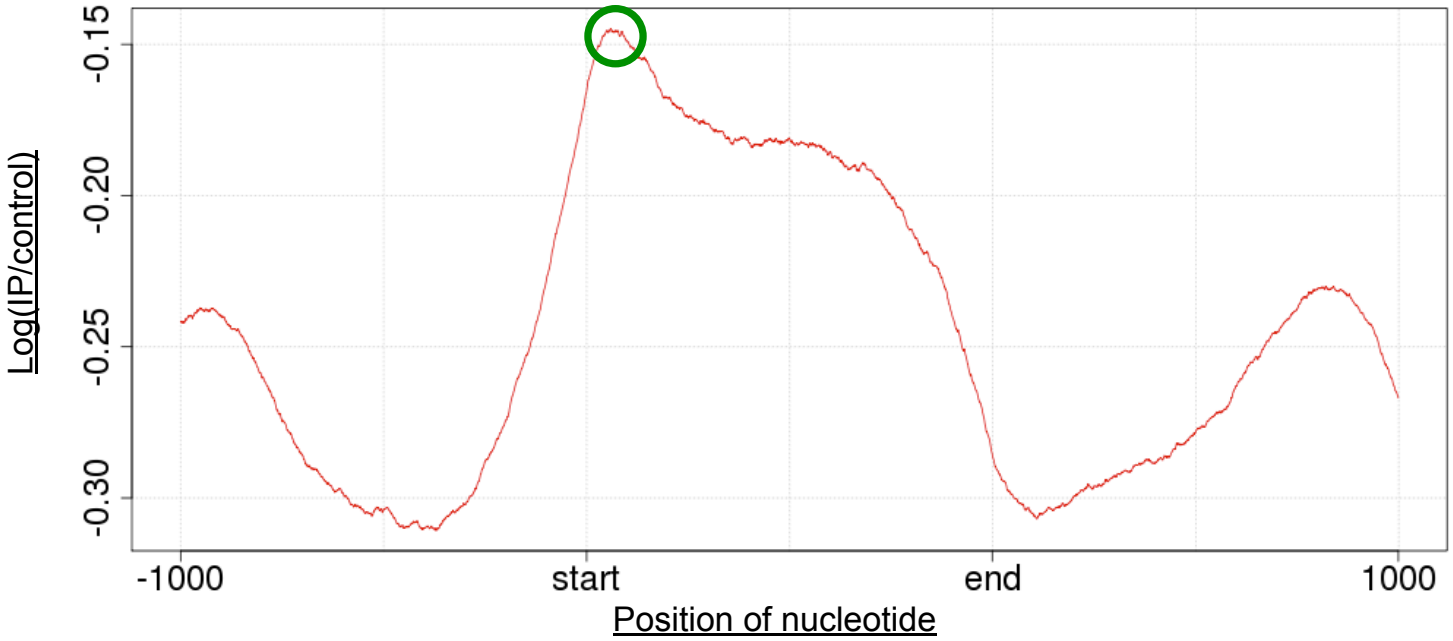
Included :

INRA URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART
--Tools for ChIP-seq

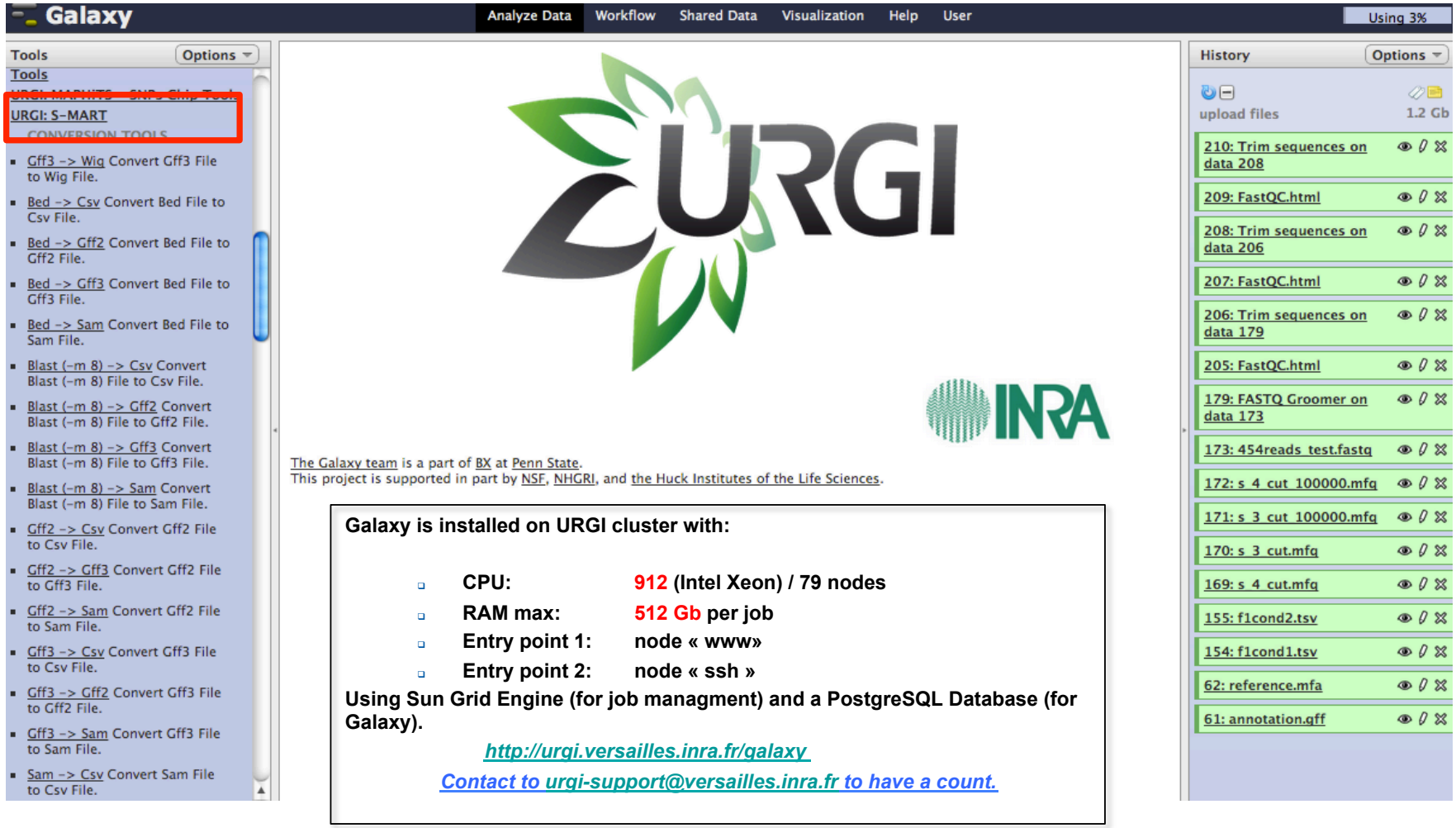
- **getWigProfile**

  Wiggle (WIG): display of dense, continuous data (GC percent, probability scores, transcriptome data), the value of each genome nucleotide



S-MART in Galaxy

--pipelines



Galaxy is installed on URGI cluster with:

- **CPU:** **912 (Intel Xeon) / 79 nodes**
- **RAM max:** **512 Gb per job**
- **Entry point 1:** **node « www»**
- **Entry point 2:** **node « ssh »**

Using Sun Grid Engine (for job managment) and a PostgreSQL Database (for Galaxy).

http://urgi.versailles.inra.fr/galaxy

Contact to urgi-support@versailles.inra.fr to have a count.

S-MART in Galaxy

--pipelines

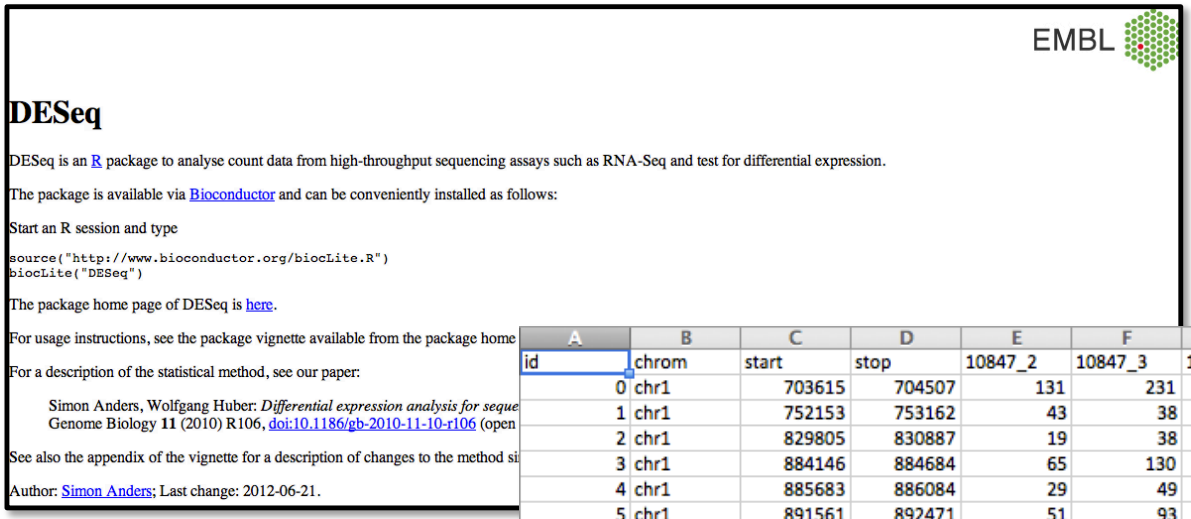
URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART in Galaxy
--pipelines

Pipeline of Differential expression analysis using DESeq

MATLAB

CompareOverlapping (S-MART)



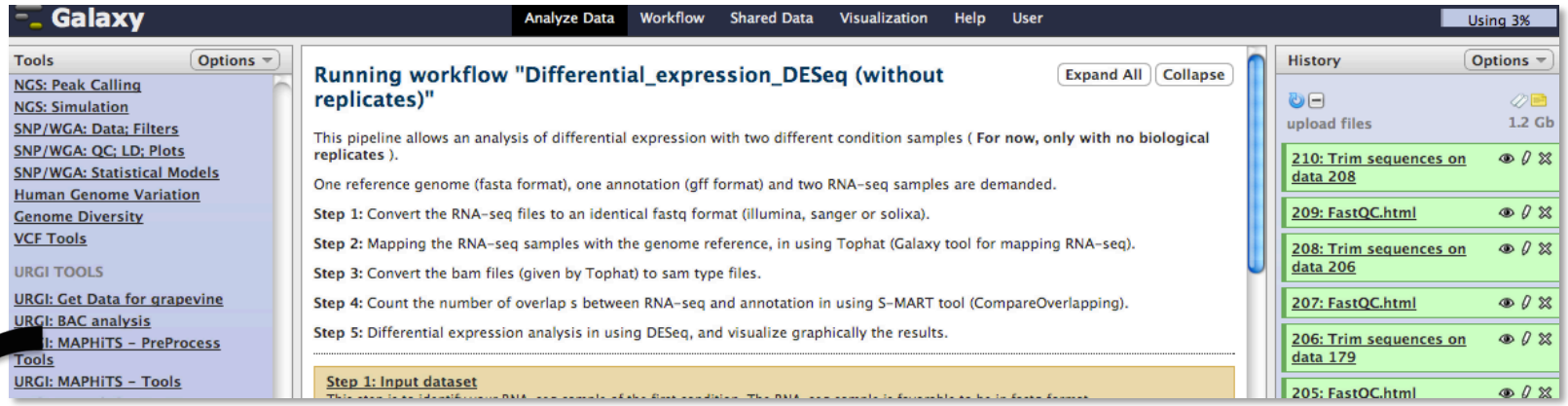
DESeq input file

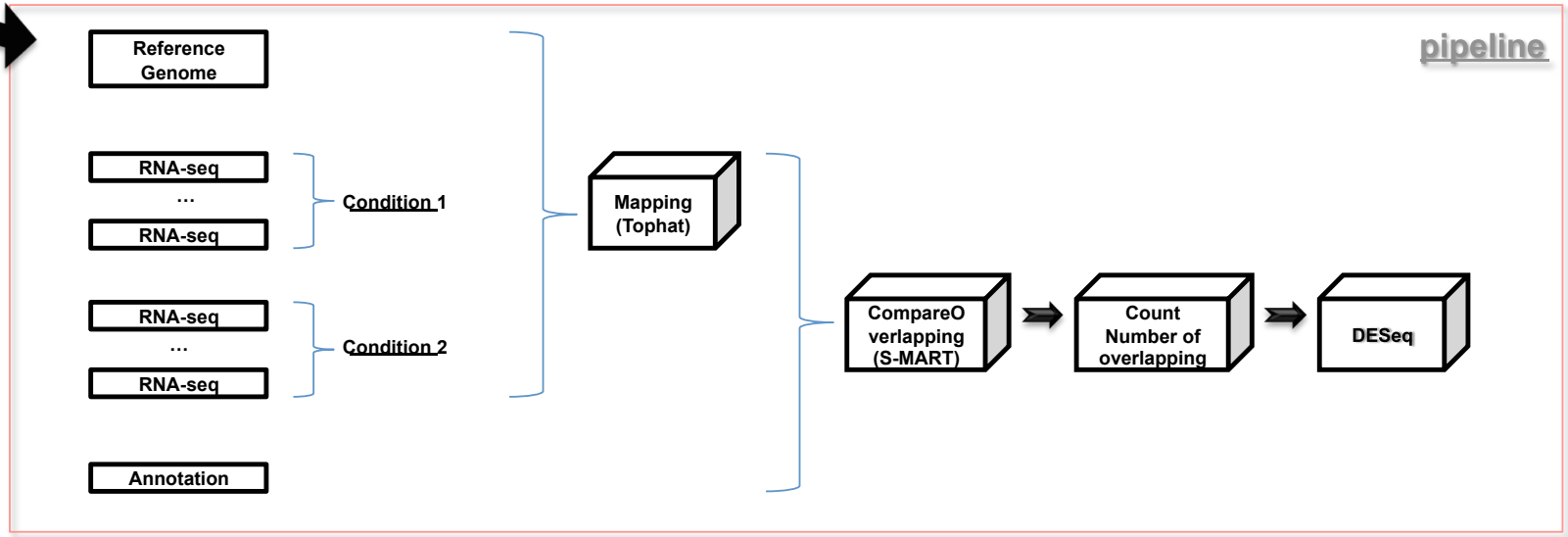
	A	B	C	D	E	F	G	H	I	J
	id	chrom	start	stop	10847_2	10847_3	10847_4	12878_1	12878_2	12878_3
0		chr1	703615	704507	131	231	158	108	71	74
1		chr1	752153	753162	43	38	67	32	31	25
2		chr1	829805	830887	19	38	25	46	41	26
3		chr1	884146	884684	65	130	82	44	41	27
4		chr1	885683	886084	29	49	16	17	9	7
5		chr1	891561	892471	51	93	30	32	16	13
6		chr1	925116	926314	111	151	52	28	31	28
7		chr1	927015	927339	33	31	14	9	8	9
8		chr1	938130	941126	478	835	537	161	121	169
9		chr1	944510	945795	67	147	71	31	18	16
10		chr1	958289	958813	45	86	22	17	13	14
11		chr1	965861	966209	26	45	12	7	10	0
12		chr1	984296	985192	56	123	36	28	22	7
13		chr1	988849	989944	92	95	60	59	40	45
14		chr1	994173	994726	38	66	15	21	11	19

URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART in Galaxy

--pipelines

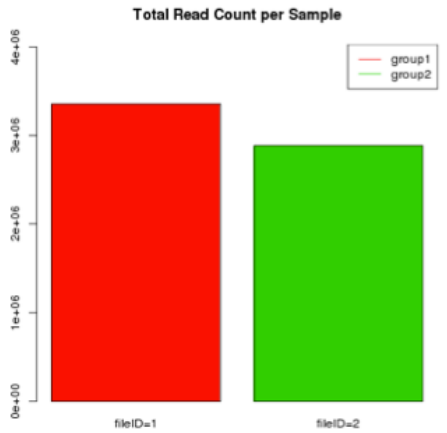


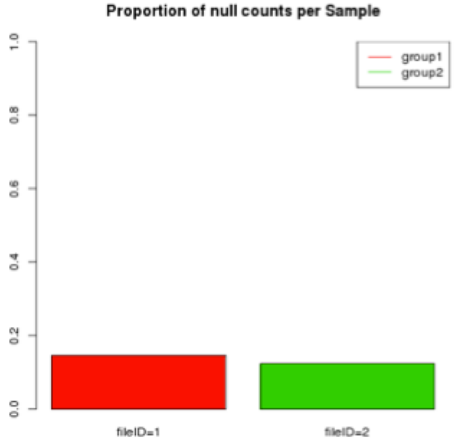


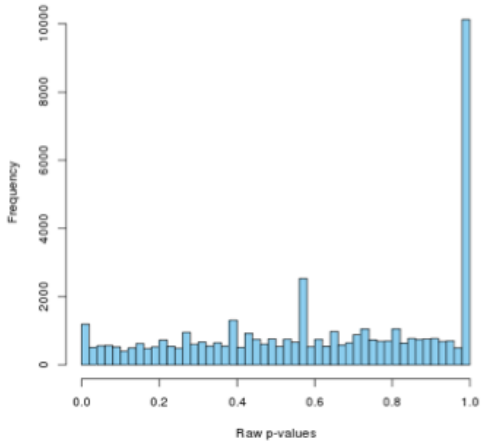
URGI INRA Versailles yufei.luo@versailles.inra.fr

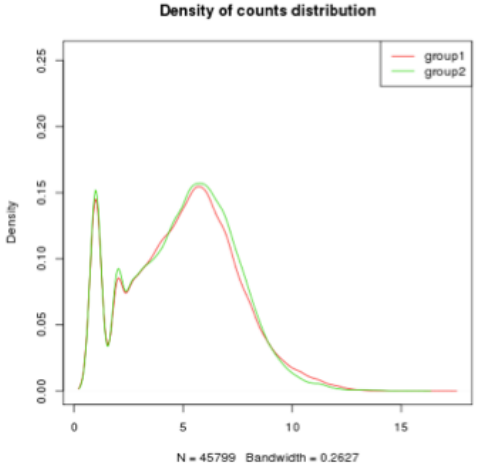
S-MART in Galaxy

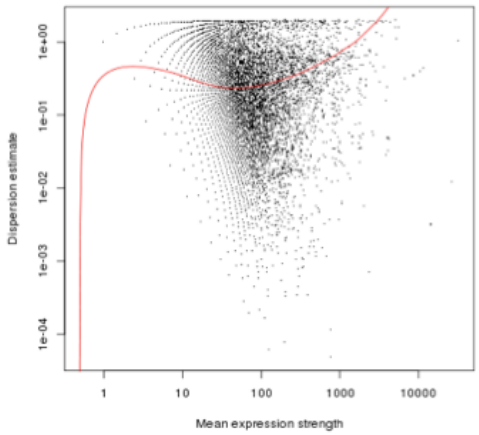
--pipelines (Maize sample, 2 different DAP conditions, neither biological nor technical replicates)







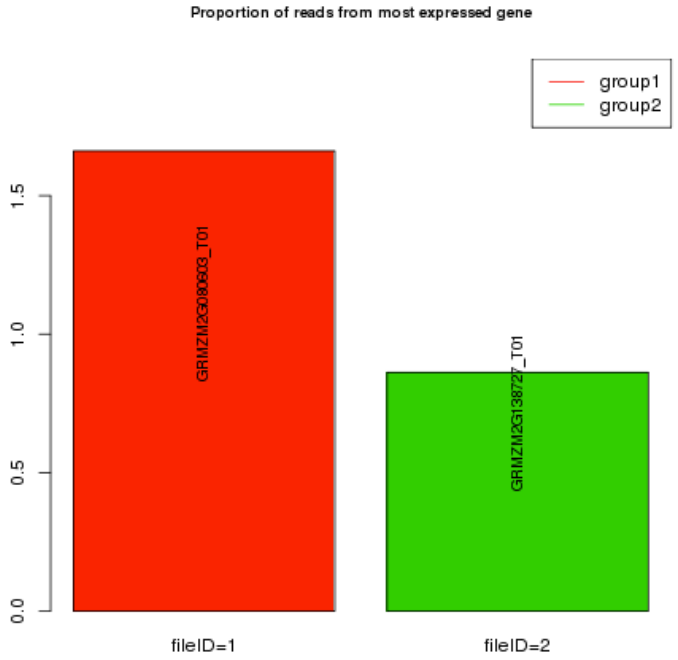


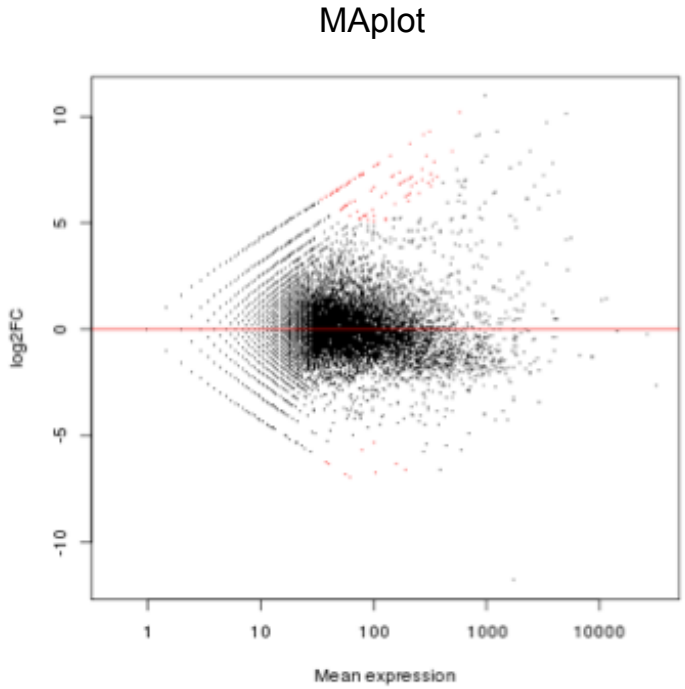


No replicates
Raw data
Results

URGI INRA Versailles yufei.luo@versailles.inra.fr

S-MART in Galaxy
--pipelines (Maize sample, 2 different DAP conditions)





Normalized data
Results

URGI INRA Versailles yufei.luo@versailles.inra.fr

Acknowledgment

URGI INRA Versailles yufei.luo@versailles.inra.fr

Thank you for your attention !!!

URGI INRA Versailles
yufei.luo@versailles.inra.fr

ALIMENTATION
AGRICULTURE
ENVIRONNEMENT