

GeneLab Analysis Working Group Pipelines

National Aeronautics and
Space Administration



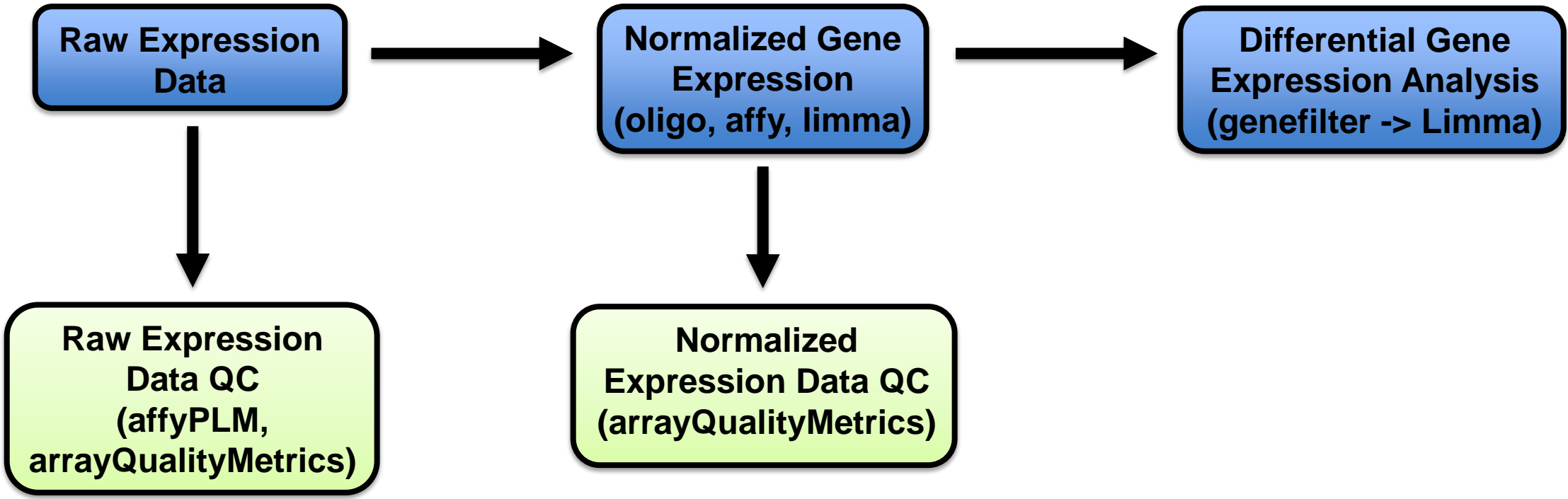
Recap of pipelines that GeneLab currently used for data processing (microarray/RNAseq)

Status update on ongoing pipeline development (metagenome profiling)

Proposed cross-species analysis pipeline

Open discussion on current and future pipelines:

- What is the value and utility/constraints of consensus pipelines to AWG work?
- Are the AWG bound by pipelines?
- Justifying exceptions to standard pipeline use
- Discussion about Reference Genome use for different pipelines
- Are there cases where we need to develop species-specific pipelines?
- Pipeline priorities (extensions of RNA-seq vs. proteomics vs. epigenomics vs. spatial vs. single cell)

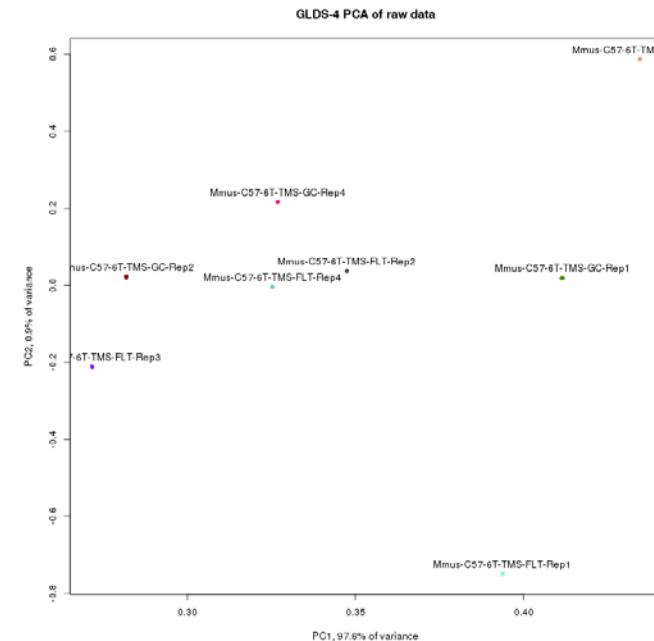
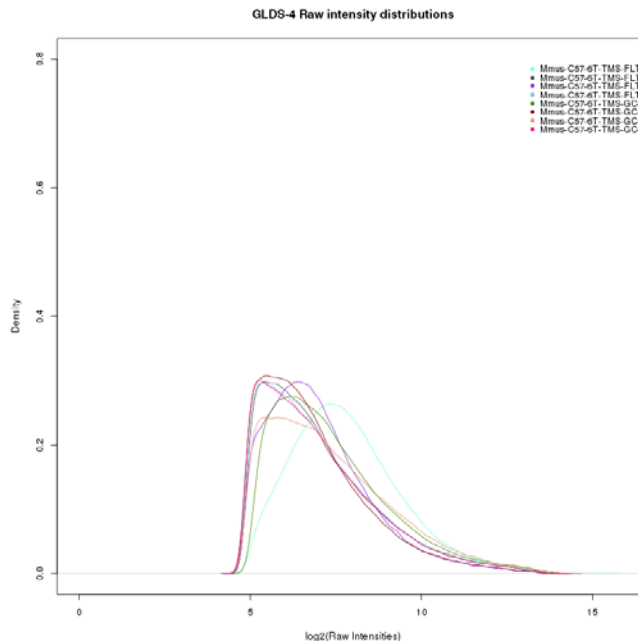
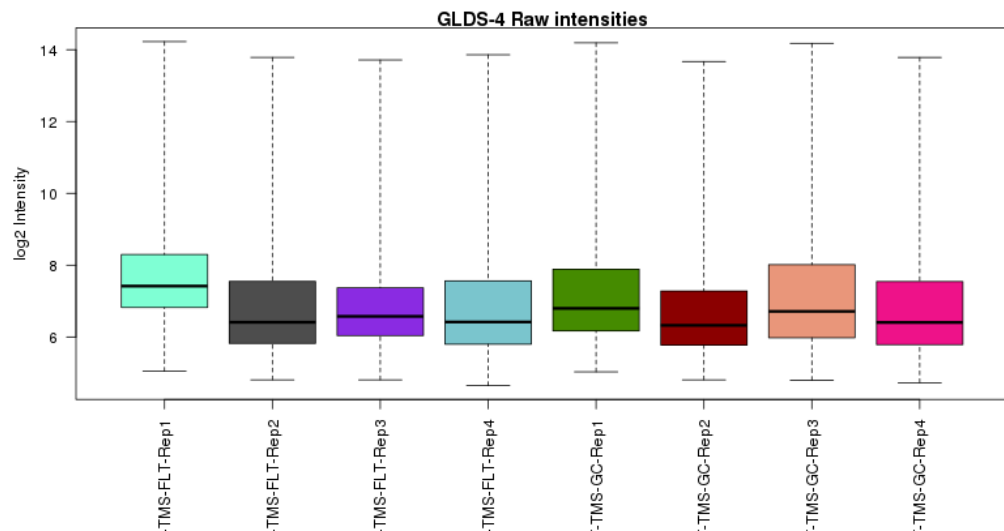


Raw Data Files

- raw.CEL
- raw.txt
- raw.CHP
- annotation.adf.txt
- image.png

Raw Data QC Reports

- rawBoxplot.png
- rawDensityDistributions.png
- rawPCA.png



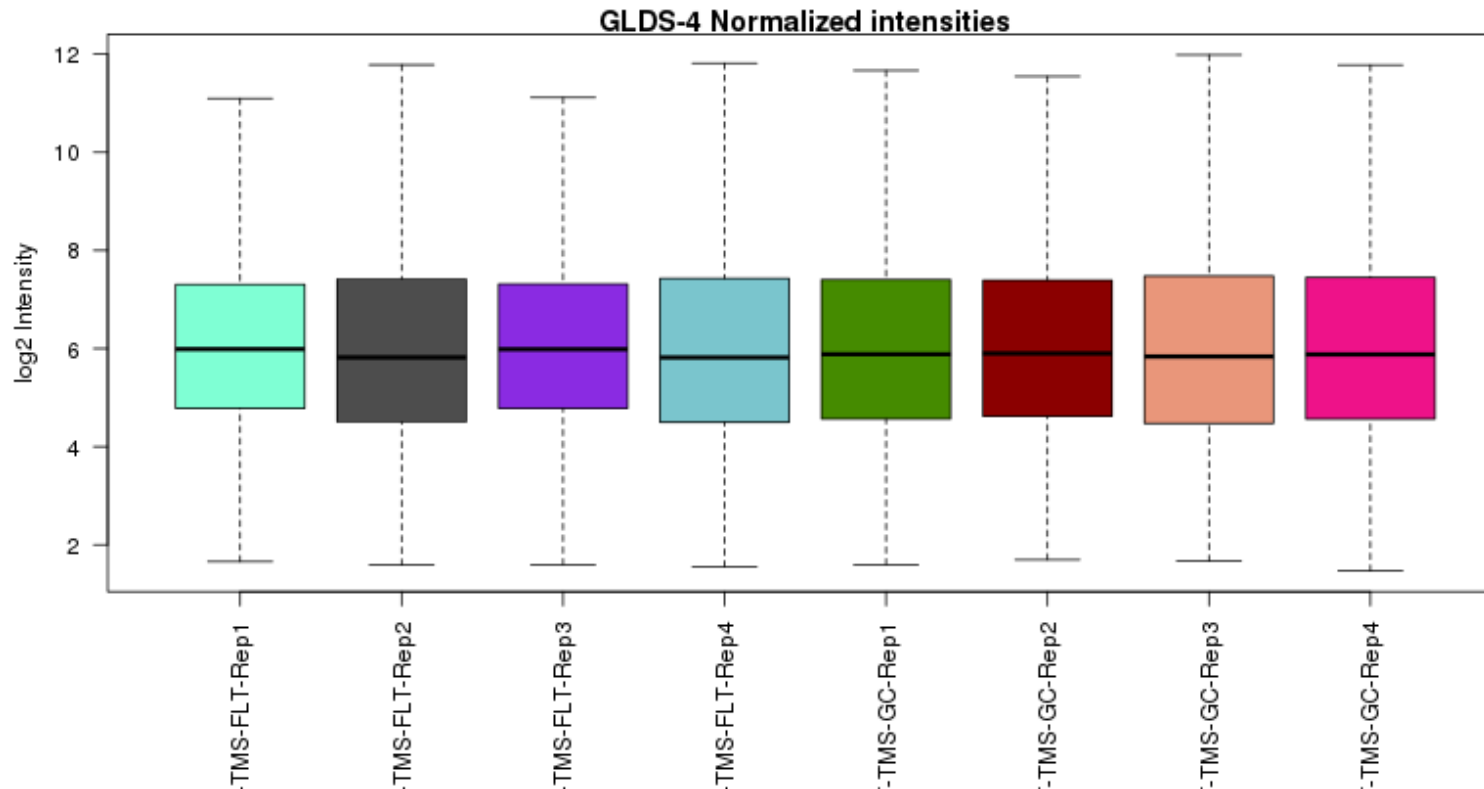
- md5sum.txt
- arrayInfo.txt

Normalized Data Files

- normalized-annotation.rda
- normalized-annotation.txt
- normalized.txt

Normalized Data QC Reports

- normBoxplot.png



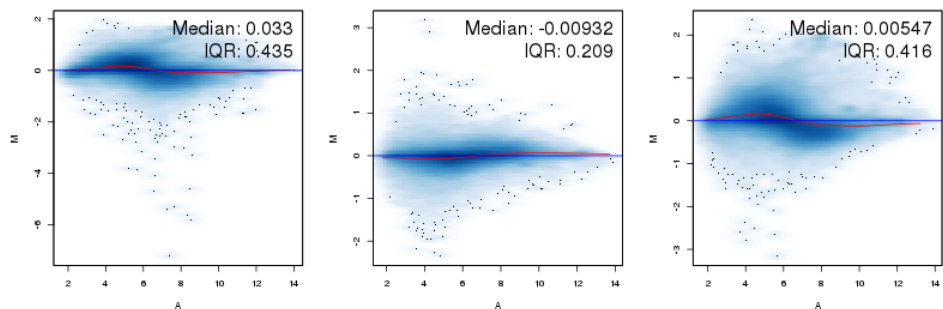
Normalized Data QC Reports

- normPlotMA.png

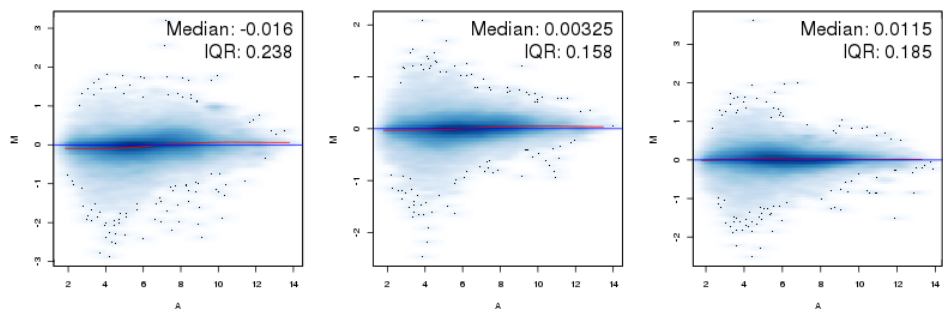
- normDensityDistributions.png

- normPCA.png

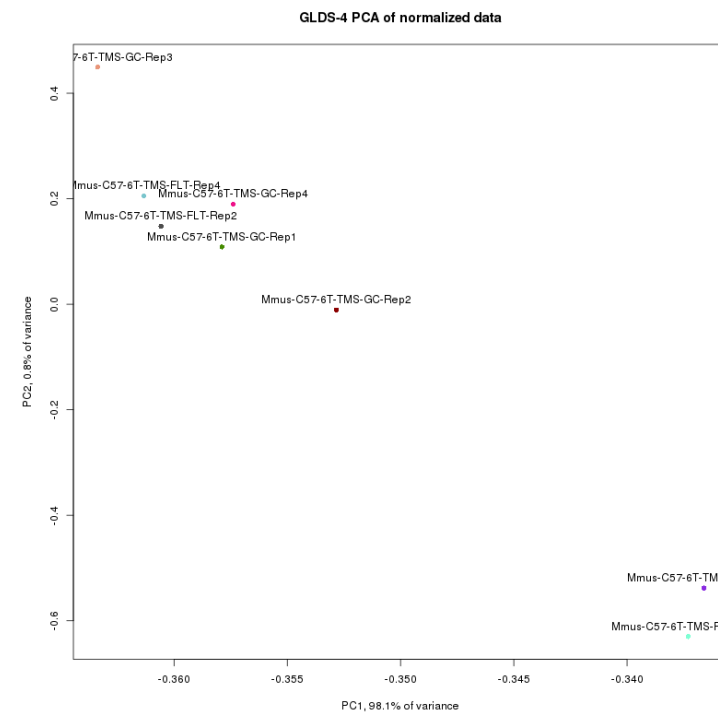
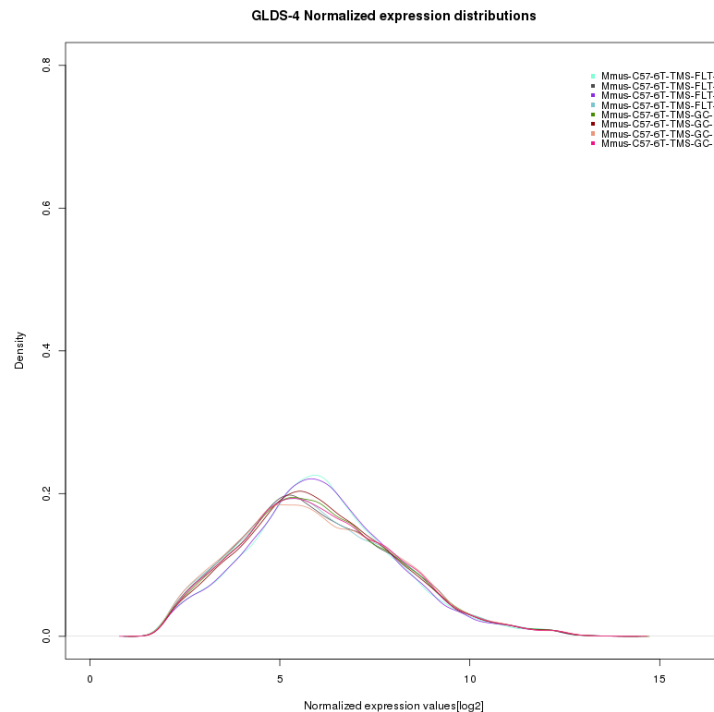
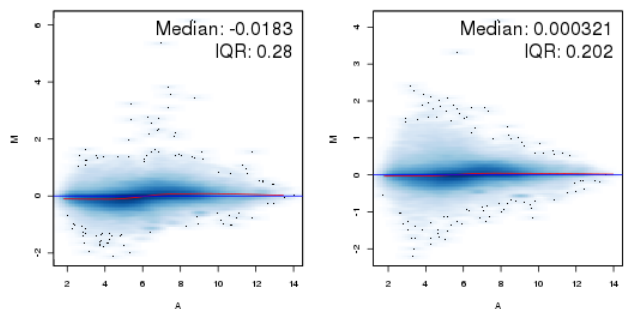
Mmus-C57-6T-TMS-FLT-Rep1 vs pseudo-median reference dMmus-C57-6T-TMS-FLT-Rep2 vs pseudo-median reference eMmus-C57-6T-TMS-FLT-Rep3 vs pseudo-median reference c



Mmus-C57-6T-TMS-FLT-Rep4 vs pseudo-median reference cMmus-C57-6T-TMS-GC-Rep1 vs pseudo-median reference cMmus-C57-6T-TMS-GC-Rep2 vs pseudo-median reference c



Mmus-C57-6T-TMS-GC-Rep3 vs pseudo-median reference cMmus-C57-6T-TMS-GC-Rep4 vs pseudo-median reference c



- annotReport.txt

Viewable DGE Files

- contrasts.csv

- differential_expression.csv
(reviewed in next slides)

	A	B	C
1		(Ground Control)v(Space Flight)	(Space Flight)v(Ground Control)
2	Ground.Control	1	-1
3	Space.Flight	-1	1

Hidden DGE Files

- visualization_output_table.csv
- visualization_PCA_table.csv

	A	B	C	D	E	F	G
1	REFSEQ	SYMBOL	GENENAME	ENSEMBL	ENTREZID	STRING_id	GOSLIM_IDS
2	NM_001355712	Lypla1	lysophospholipase 1	ENSMUSG00000025903	18777	NA	GO:0003824, GO:
3	NM_001310442	Atp6v1h	ATPase, H+ transport	ENSMUSG00000033793	108664	NA	GO:0003824, GO:
4	NM_001204371	Oprk1	opioid receptor, kappa	ENSMUSG00000025905	18387	NA	GO:0003008, GO:
5	NM_009826	Rb1cc1	RB1-inducible coiled-	ENSMUSG00000025907	12421	NA	GO:0005488, GO:
6	NM_001195732	Alka1	ALK and LTK ligand 1	ENSMUSG00000087247	620393	NA	NA
7	NM_001244692	St18	suppression of tumo	ENSMUSG00000033740	240690	NA	GO:0003676, GO:
8	NM_183028	Pcmt1	protein-L-isoaspartat	ENSMUSG00000051285	319263	NA	GO:0003824, GO:
9	NM_021511	Rrs1	ribosome biogenesis	ENSMUSG00000061024	59014	NA	GO:0005622, GO:
10	NM_001357376	Adhfe1	alcohol dehydrogena	ENSMUSG00000025911	76187	NA	GO:0003824, GO:
11	NM_178399	Vxn	vexin	ENSMUSG00000067879	76982	NA	NA
12	NM_001037759	Sgk3	serum/glucocorticoid	ENSMUSG00000025915	170755	NA	GO:0003824, GO:
13	NM_177722	Mcmdc2	minichromosome ma	ENSMUSG00000046101	240697	NA	GO:0003676, GO:



Microarray DGE Output Table (GLDS-4)



FLT

GC

H	I	J	K	L	M	N	O
Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T	Mmus-C57-6T-T
8.94207578	9.46813253	9.212214603	9.41423311	9.50678711	9.364989464	9.379174386	9.451228494
7.993203537	8.430513286	8.214404301	8.559117139	8.521083776	8.46005033	8.294820919	8.329526007
3.892463048	4.113050799	4.236066374	3.673600579	4.171318208	4.231675859	3.689274732	3.571903696
8.329472609	8.507682803	8.327480432	8.451815852	8.30189246	8.345079482	8.109227417	8.109114561
3.533509315	3.292035686	3.958079091	3.566297334	3.561110689	3.504118022	3.541423336	3.492351412
5.527285822	5.316749183	4.917606596	5.224871708	5.120084602	4.969627168	4.846140097	5.014355134
8.570359137	8.979656841	8.672800351	8.856555151	8.72261036	8.92404142	8.906917008	8.682484484
7.331069045	7.308107749	7.128373591	7.339554764	7.517422532	7.379427448	7.13914564	7.185443669
7.073673101	8.10845166	8.068972476	7.786711201	8.042140934	8.184687239	8.642181967	8.040036423
4.540142268	4.485612161	4.455309728	4.139137682	4.113947171	4.657196649	3.995546005	4.232743672
7.80680137	8.166232735	7.907231912	8.149754759	8.314871956	8.1636341	8.19672727	8.20839631
4.025952867	3.8535043	3.995937683	3.662130869	3.926143246	4.237139359	3.883649086	3.991220651

P	Q	R	S	T	U	V
All.mean	All.stdev	F.p.value	Group.Mean_(Space	Group.Mean_(Ground	Group.Stdev_(Space	Group.Stdev_(Ground
9.34235443	0.03285599	0.2189058	9.425544863	9.259164006	4.712772432	4.629582003
8.35033991	0.03592015	0.46118532	8.401370258	8.299309566	4.200685129	4.149654783
3.94741916	0.06115463	0.72599956	3.916043124	3.9787952	1.958021562	1.9893976
8.3102207	0.02340124	0.10841635	8.21632848	8.404112924	4.10816424	4.202056462
3.55611561	0.03685987	0.65238368	3.524750865	3.587480357	1.762375432	1.793740178
5.11709004	0.0368555	0.08079719	4.98755175	5.246628327	2.493775875	2.623314164
8.78942809	0.02978135	0.75379349	8.809013318	8.76984287	4.404506659	4.384921435
7.29106805	0.02770171	0.81225039	7.305359822	7.276776287	3.652679911	3.638388144
7.99335687	0.09590328	0.05418532	8.227261641	7.759452109	4.11363082	3.879726055
4.32745442	0.046655	0.32984653	4.249858374	4.40505046	2.124929187	2.20252523
8.1142063	0.02654415	0.08898839	8.220907409	8.007505194	4.110453704	4.003752597
3.94695976	0.03068554	0.33244738	4.009538085	3.88438143	2.004769043	1.942190715

W	X	Y	Z	AA	AB
Log2fc_(Ground Con	Log2fc_(Space Fli	P.value_(Ground Con	P.value_(Space Fli	Adj.p.value_(Ground Con	Adj.p.value_(Space Fli
0.166380858	-0.166380858	0.2189058	0.2189058	0.599192969	0.599192969
0.102060692	-0.102060692	0.461185324	0.461185324	0.761770048	0.761770048
-0.062752077	0.062752077	0.725999561	0.725999561	0.894928857	0.894928857
-0.187784444	0.187784444	0.108416348	0.108416348	0.511285762	0.511285762
-0.062729492	0.062729492	0.652383677	0.652383677	0.860339486	0.860339486
-0.259076577	0.259076577	0.080797195	0.080797195	0.488401839	0.488401839
0.039170448	-0.039170448	0.753793487	0.753793487	0.904897061	0.904897061
0.028583535	-0.028583535	0.812250395	0.812250395	0.930106109	0.930106109
0.467809531	-0.467809531	0.054185323	0.054185323	0.463152639	0.463152639
-0.155192086	0.155192086	0.329846526	0.329846526	0.678367538	0.678367538
0.213402215	-0.213402215	0.088988388	0.088988388	0.496332813	0.496332813
0.125156656	-0.125156656	0.332447378	0.332447378	0.679510477	0.679510477



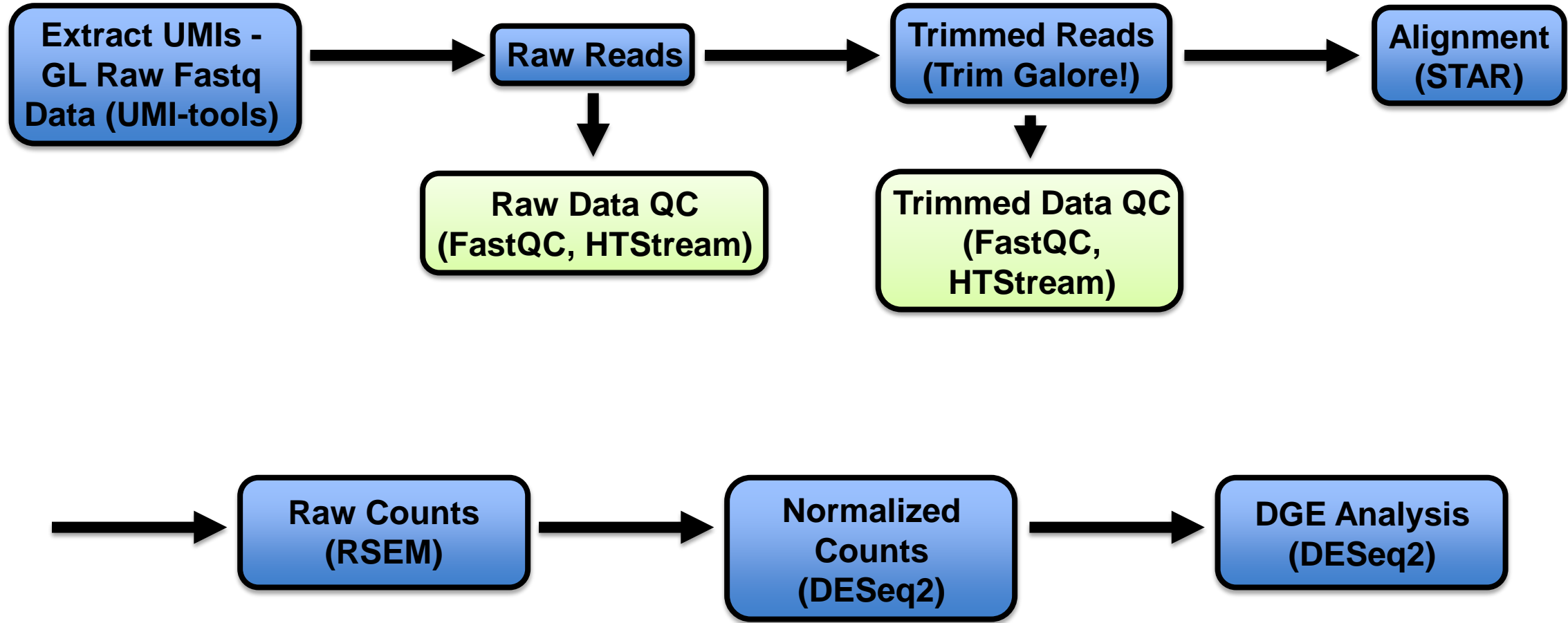
GLDS-4: Microarray Analysis of Space-flown Murine Thymus Tissue Version 1

Select a Version: ▼

Source Accession Number [GSE18388](#)
 Total Data Volume: 144.3 MB

Submitted Date: 02-Oct-2009
 Release Date: 10-Mar-2010

DESCRIPTION	PROTOCOLS	SAMPLES	ASSAYS/MEASUREMENTS	PUBLICATIONS	STUDY FILES
<p>To view files, click on the folder of interest.</p> <ul style="list-style-type: none"> ◀ All Files <ul style="list-style-type: none"> Study Metadata Files Microarray Data Files ◀ GeneLab Processed Microarray Data Files <ul style="list-style-type: none"> Raw Data Raw Data/Plots and Images Raw Data/Supplemental Materials Raw Data/QC Reports Normalized Data Normalized Data/Supplemental Materials Normalized Data/Plots and Images Differential Expression Analyses 					<p>0 files selected</p> <div style="background-color: #ccc; padding: 5px; text-align: center;">Download Selected Files</div>



Raw Reads:

*raw.fastq.gz

Raw Data QC:

*raw_fastqc.html

*raw_fastqc.zip

/raw_multiqc_report

*raw_multiqc_report.html

/raw_multiqc_data

*multiqc_data.json

*multiqc_fastqc.txt

*multiqc_general_stats.txt

*multiqc.log

*multiqc_sources.txt

Trimmed Reads:

*trimmed.fastq.gz

Trimmed Data QC:

/Trimming_Reports

*trimming_report.txt

/FastQC_Reports

*raw_fastqc.html

*raw_fastqc.zip

/raw_multiqc_report

*raw_multiqc_report.html

/raw_multiqc_data

*multiqc_data.json

*multiqc_fastqc.txt

*multiqc_general_stats.txt

*multiqc.log

*multiqc_sources.txt

Alignment:

- *Aligned.sortedByCoord.out.bam
- *Aligned.toTranscriptome.out.bam
- *SJ.out.tab

Raw Counts:

- *genes.results
- *isoforms.results
- *Unnormalized_Counts.csv

Normalized Counts:

- *Normalized_Counts.csv
- /ERCC-NormCounts**
- *ERCC-Normalized_Counts.csv

Alignment QC:

- *Log.final.out

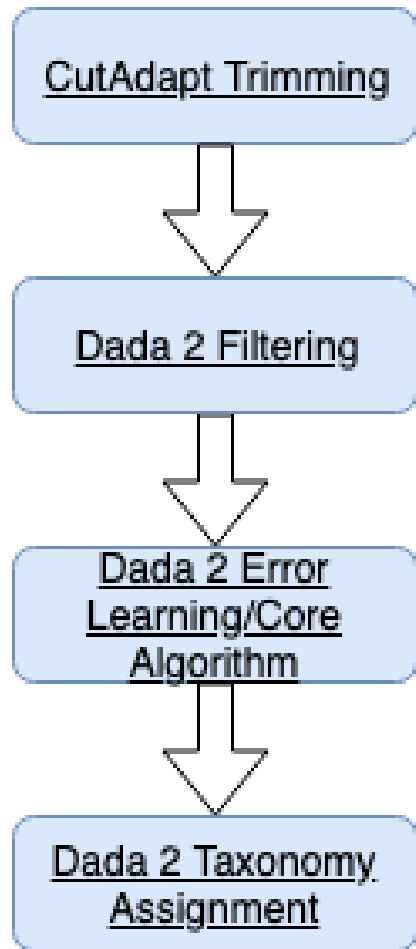
DGE Analysis:

- *contrasts.csv
- *Group1_v_Group2_DGE.csv
- *Group1_v_Group3_DGE.csv
- *Group2_v_Group3_DGE.csv
- *differential_expression.csv
- /ERCC-NormDGE**
- *Group1_v_Group2_DGE.csv
- *Group1_v_Group3_DGE.csv
- *Group2_v_Group3_DGE.csv
- *differential_expression.csv

(Hidden) Visualization Files:

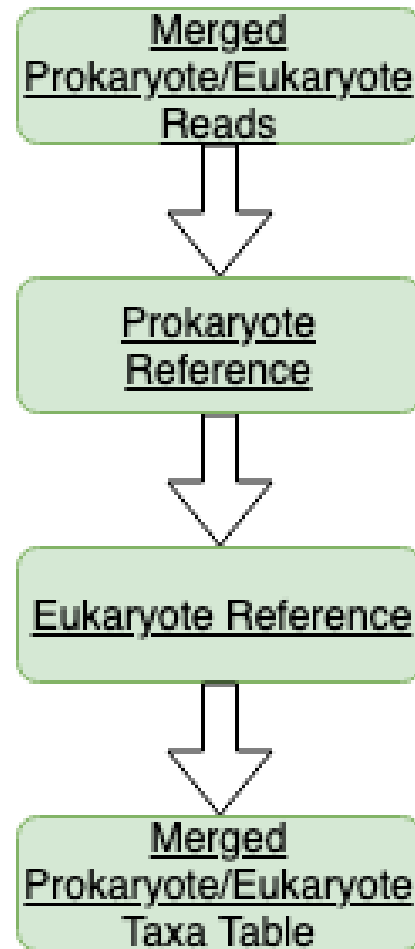
- *visualization_output_table.csv
- *visualization_PCA_table.csv
- /ERCC-NormDGE**
- *visualization_output_table.csv
- *visualization_PCA_table.csv

Current Pipeline (Under Review)



Options for Processing Eukaryotic vs. Prokaryotic Primers/Reads

Option 1



Option 2

