DVX - Differential Variation and eXpression analysis

This file was created on January 03, 2018 at 22:14:55.

Note: the report was generated automatically by DVX. Text in blue font was added manually to the report, and some plots were resized.

ExpressionSet metadata

name	N/A
lab	N/A
contact	N/A
title	16p11.2 copy number variation models: various brain regions
url	N/A
pubMedIds	21969575
samples	37
features	35556
abstract	Analysis of brain regions of C57BL/6N:129Sv animals harboring a df/+ deletion or dp/+ duplication in the chromosomal region corresponding to 16p11.2 in humans. Recurrent copy number variations (CNVs) of human 16p11.2 are associated with a variety of developmental/neurocognitive syndromes.

Dataset Summary - Gene Expression Statistics

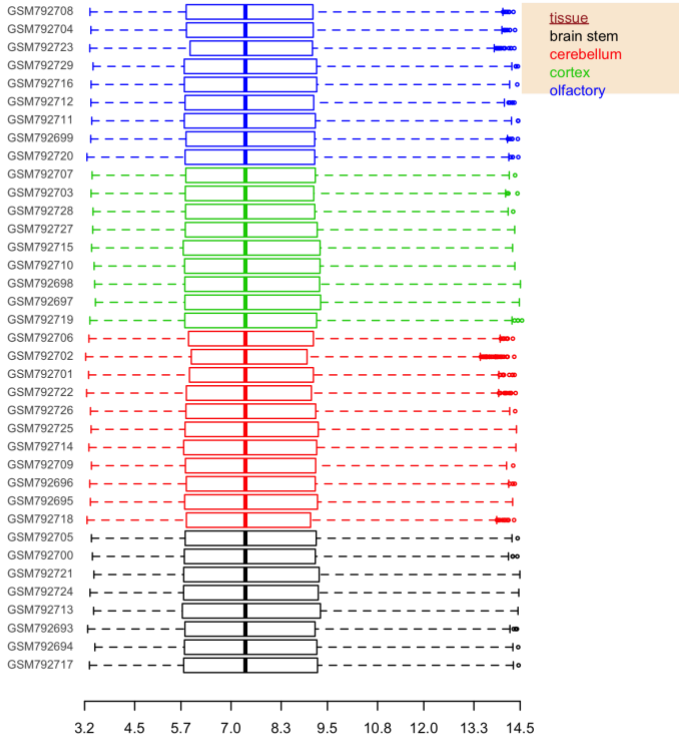
min	Q1	Median	Q3	max	Mean	Missing
3.23	5.82	7.37	9.20	14.55	7.57	0

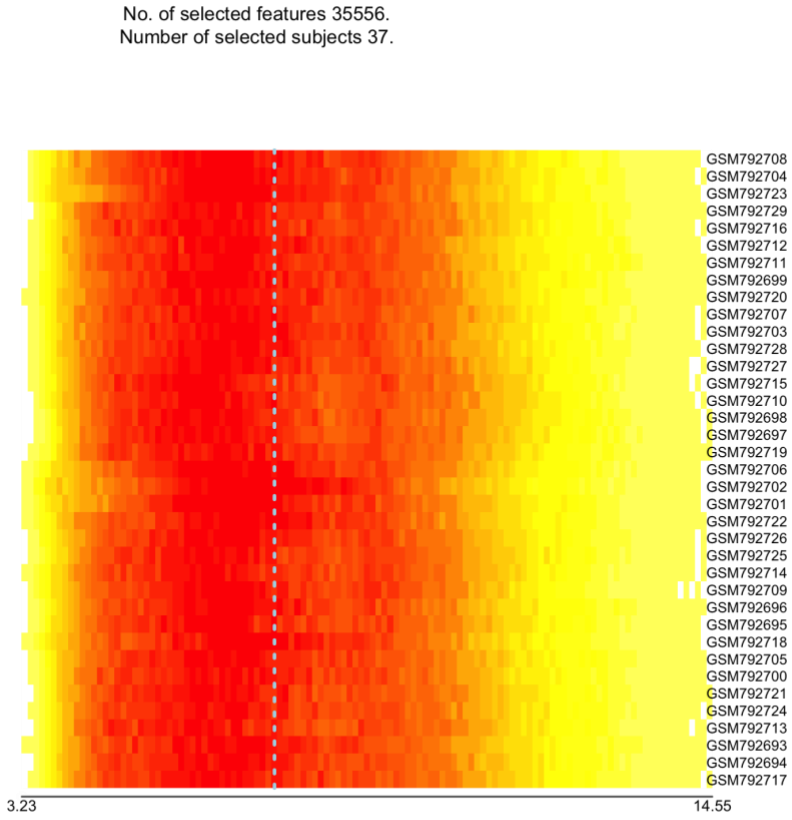
Phenotype data - Factors

Variable	Content (counts)
Tissue	brain stem (8), cerebellum (11), cortex (9), olfactory (9)
genotype/variation	+/+ wild type - 2 copies (15), df/+ deletion - 1 copy (10), dp/+ duplication - 3 copies (12)
Individual	101 (6), 102 (5), 121 (4), 62 (5), 63 (4), 88 (4), 89 (3), 90 (6)

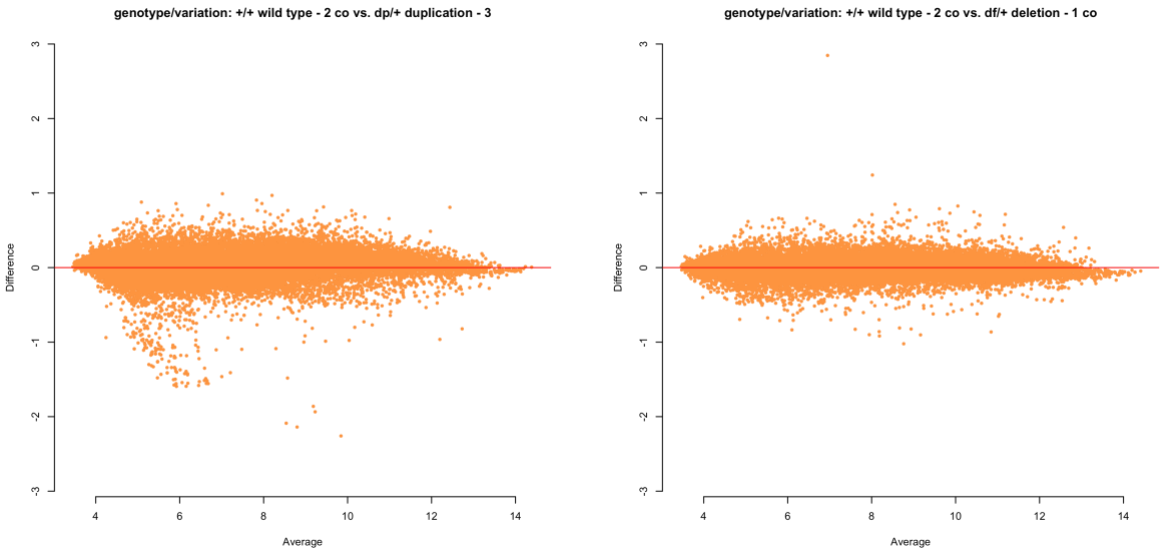
Plots

The distribution of the 35,556 expression values for each of the 37 samples is presented as boxplots, and as a "flat histogram plot" (next page)

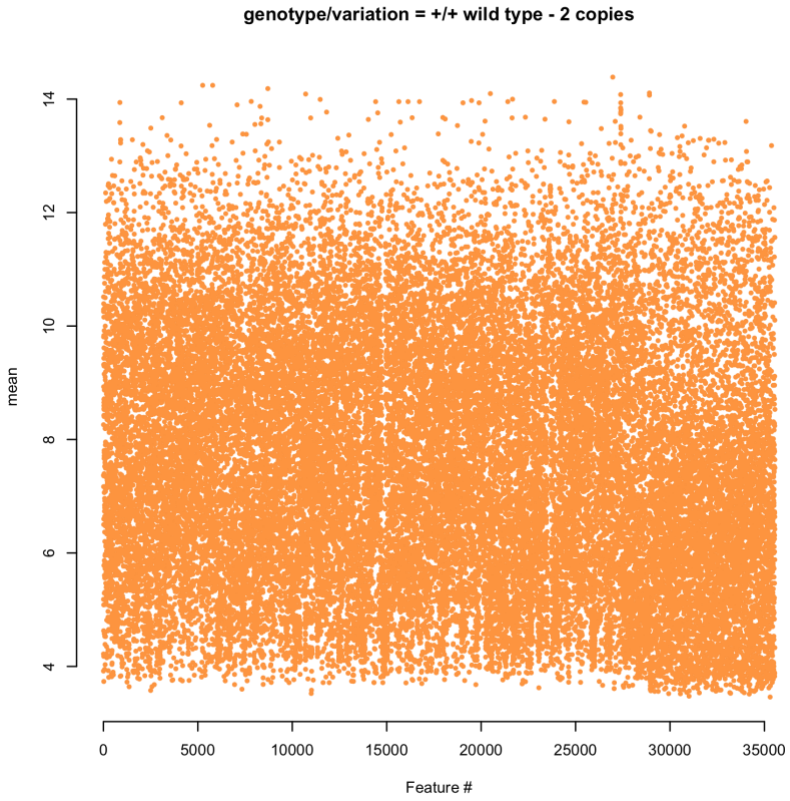




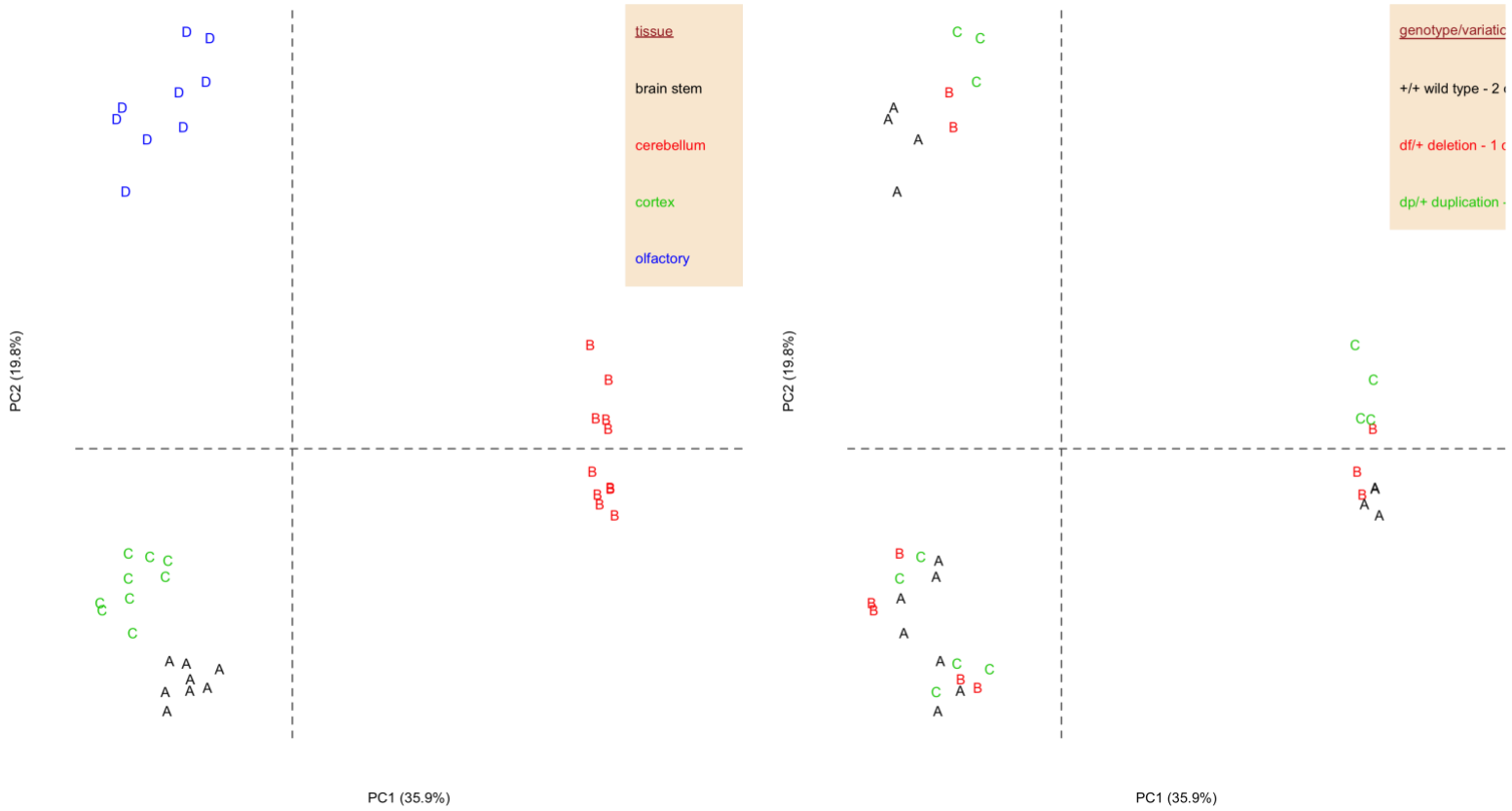
The vertical dashed line represents the overall median. The data was centered so that all samples have the same median.
The distributions appear slightly skewed. The data was already normalized by the submitters, but we may wish to consider further transformation or filtering.



These are two Bland-Altman plots, showing the Difference between expression levels between two groups, versus the average expression levels of the two groups. On the left is the plot for the comparison between the wild type and the "dp/+" group (duplication, with 3 copies). On the right is the plot for wild type vs. the "df/+" group (deletion, with one copy).
The comparison between wild type and the three copies group shows that the largest difference (in absolute values) are found, mostly, for genes which have low abundance, on average. We may consider filtering low-abundance genes from the analysis.



The plot of the mean expression levels of all subjects shows that many of the ~7,000 genes on the right may have lower values than the rest of the array. We may consider using a subset of genes with minimum mean expression level.



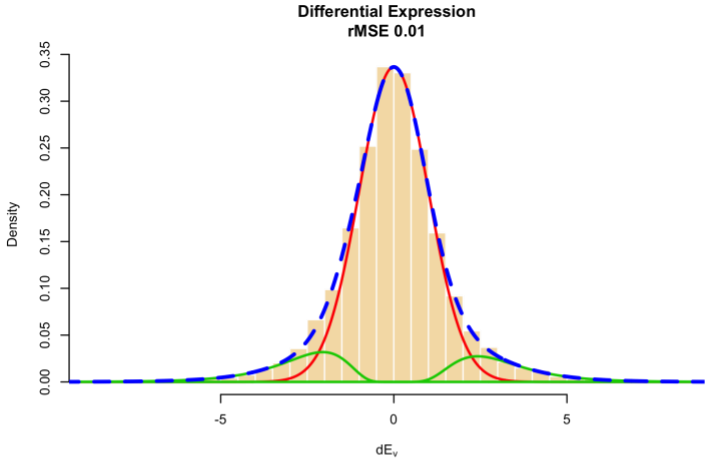
The principal component plots show PC1 vs. PC2, and the colors and labels correspond to different levels of the selected factor. It is clear that the expression levels within each brain tissue are correlated, as evident by the clusters in the plot on the left. However, within each cluster, the genotype/variation levels are balanced. This suggests that the overall effect of the brain tissue factor will cancel out when we perform the differential analysis.

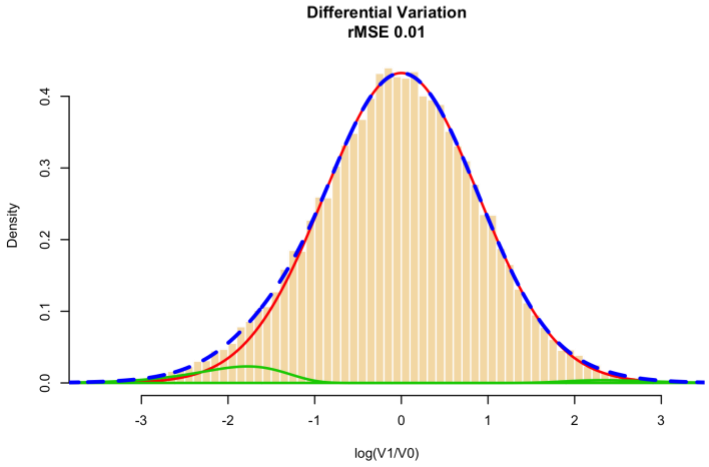
The following pages contain the statistical analyses for two comparisons, based on the Copy Number Variation factor: CNV=2 vs. CNV=1 (deletion), and CNV=2 vs. CNV=3 (duplication). For each comparison, we apply all three models, namely L2N, N3, and limma.
The complete lists of genes with the corresponding statistics and p-values from each analysis are provided in separate .csv files.

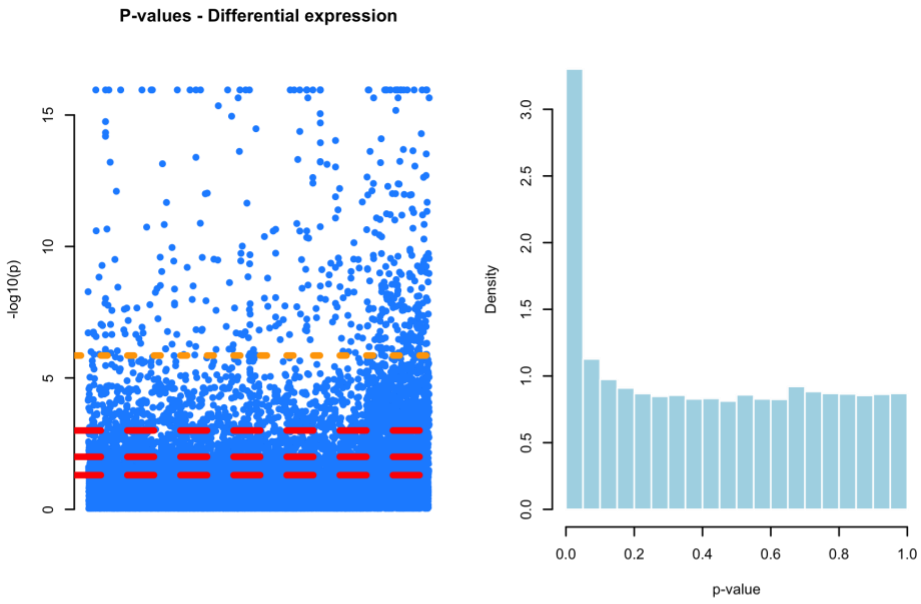
WT vs. deletion - L2N

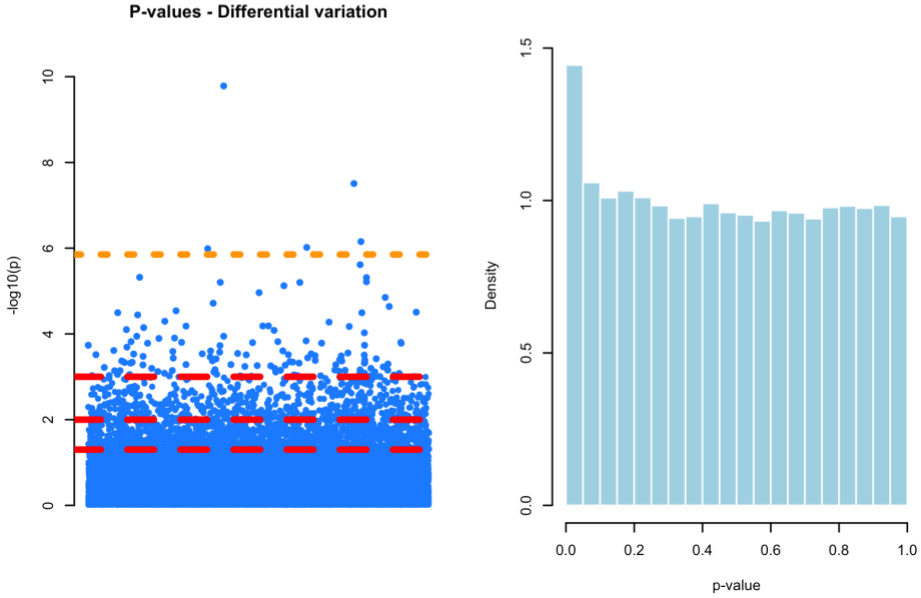
Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: df/+ deletion - 1 copy
Model: L2N
Predictors: individual





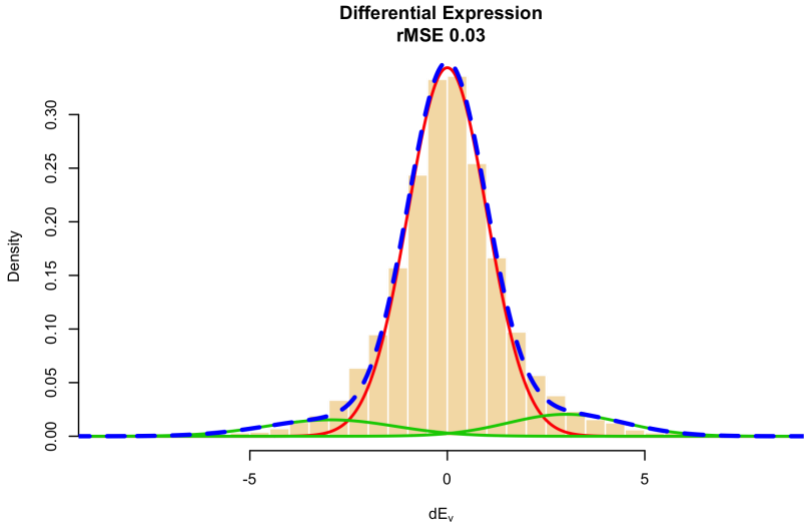

P-values - Differential expression
-log10(p)
0
5
10
15
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value

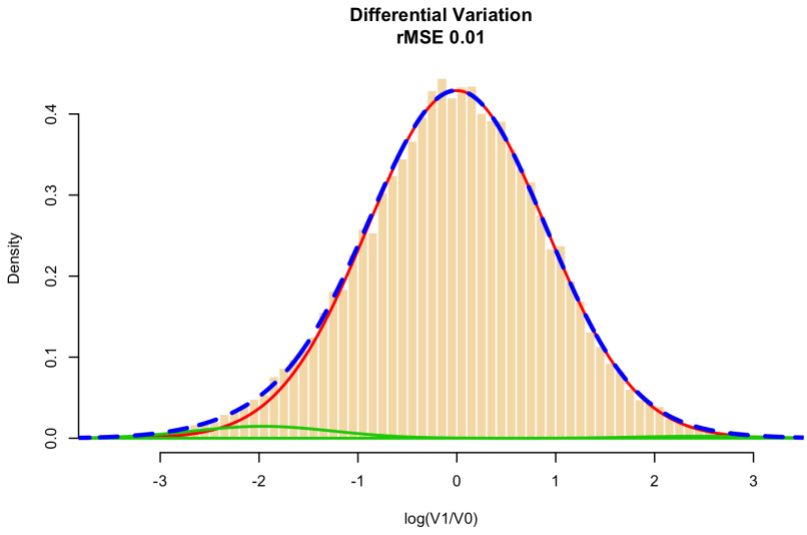

P-values - Differential variation
-log10(p)
0
2
4
6
8
10
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
p-value

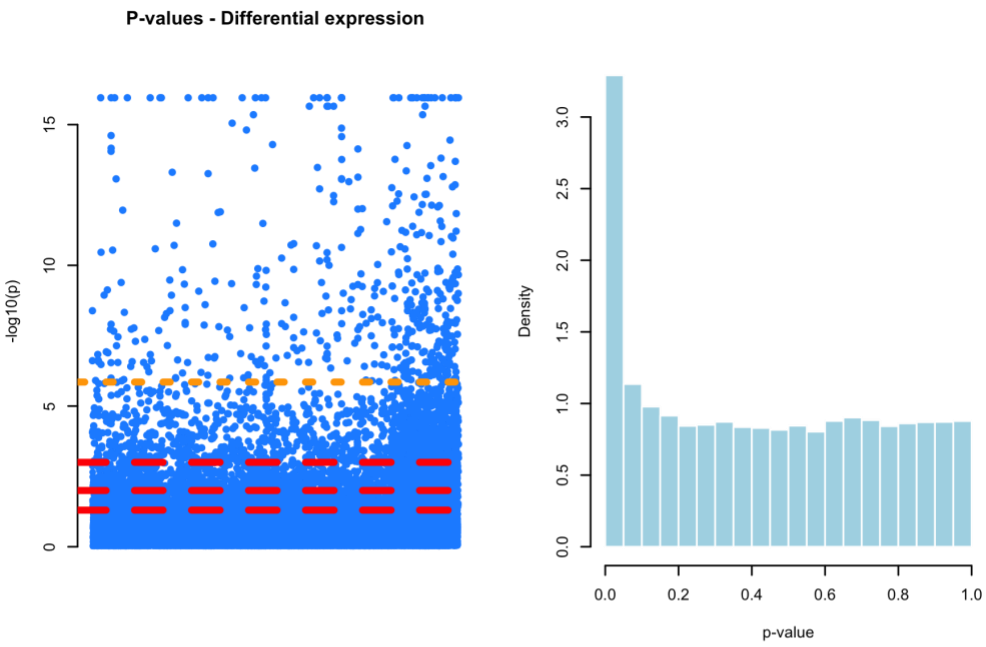
WT vs. deletion - N3

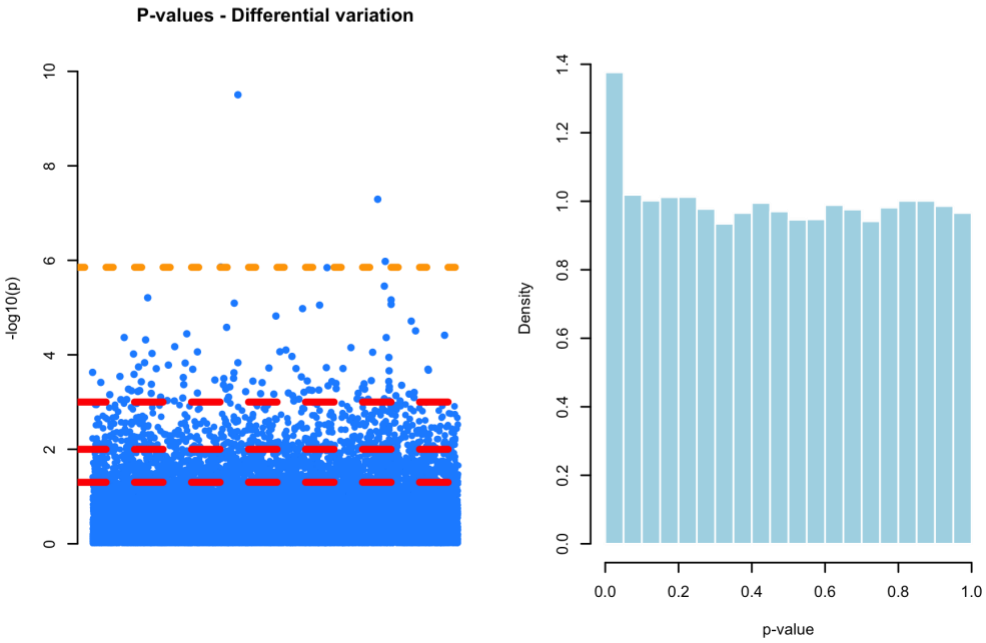
Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: df/+ deletion - 1 copy
Model: N3
Predictors: individual





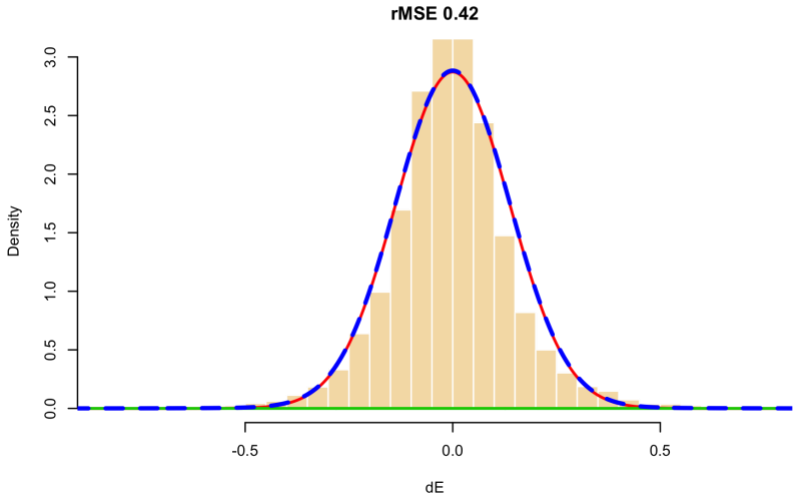

P-values - Differential expression
-log10(p)
0
5
10
15
Density
0.0
0.5
1.0
1.5
2.0
2.5
3.0
0.0
0.2
0.4
0.6
0.8
1.0
p-value

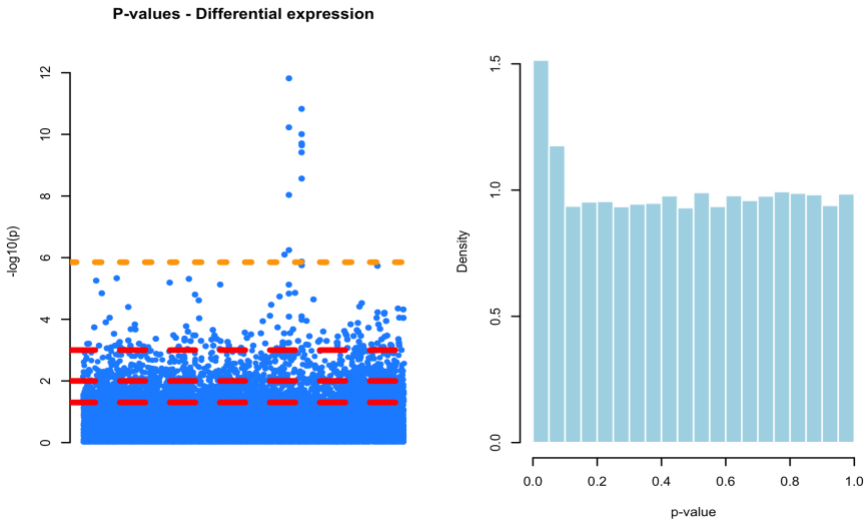

P-values - Differential variation
-log10(p)
0
2
4
6
8
10
Density
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
0.0
0.2
0.4
0.6
0.8
1.0
p-value

WT vs. deletion - limma

Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: df/+ deletion - 1 copy
Model: limma
Predictors: individual





Summary of the WT vs. CNV=1 analysis, using q-value threshold of 0.05 to determine differential genes:

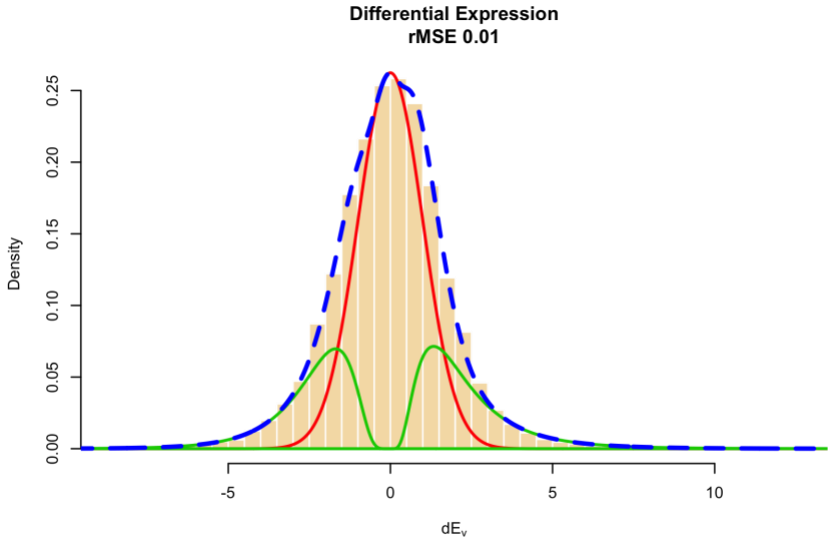
	Differential variation		Differential expression	
	rMSE	Genes with q≤ 0.05	rMSE	Genes with q≤ 0.05
L2N	0.01	16	0.01	2,479
N3	0.01	14	0.03	2,456
limma	NA	NA	0.42	32

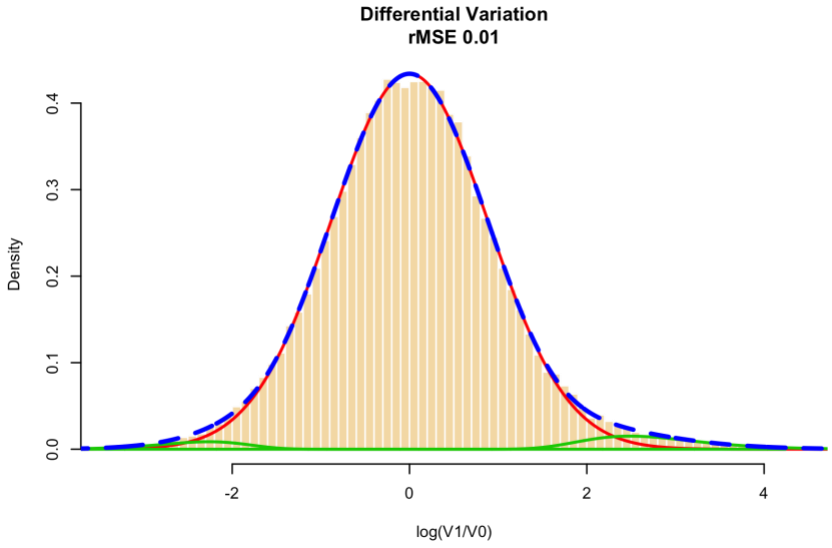
L2N has the smallest rMSE, although very similar to N3.
Note that limma does not provide differential variation analysis.

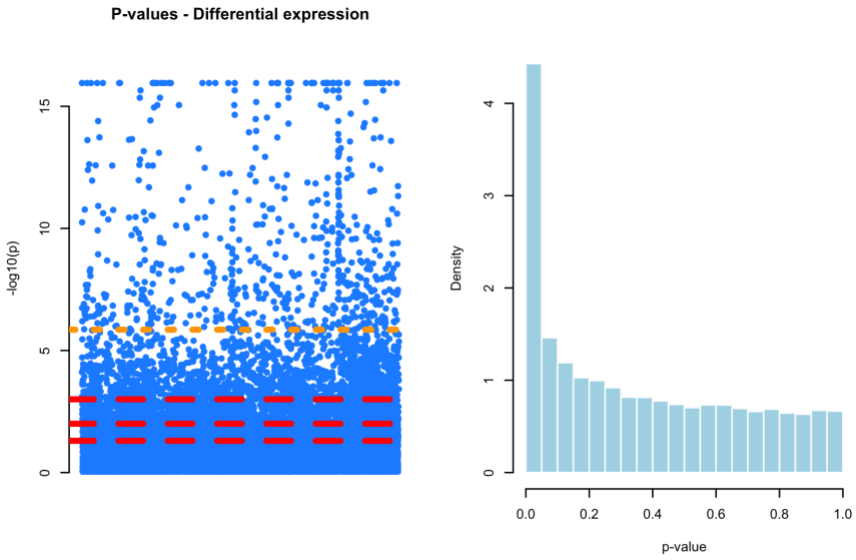
WT vs. duplication - L2N

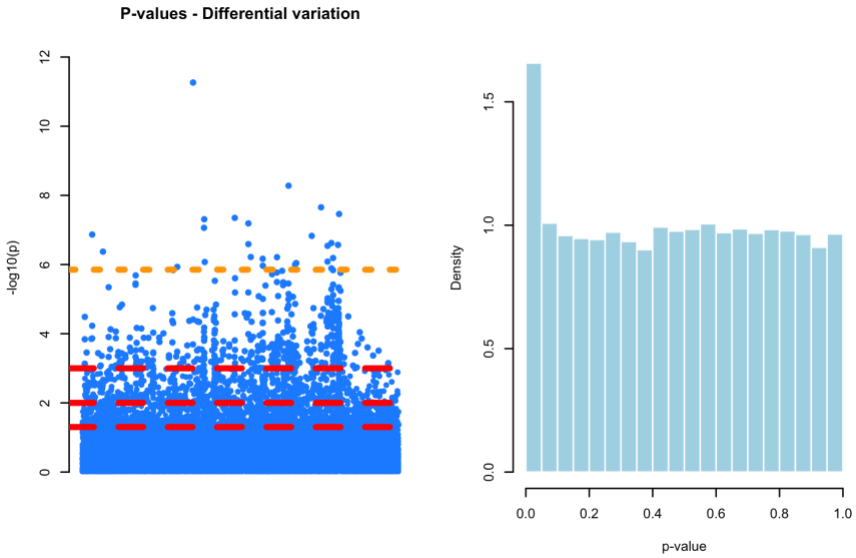
Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: dp/+ duplication - 3 copies
Model: L2N
Predictors: individual





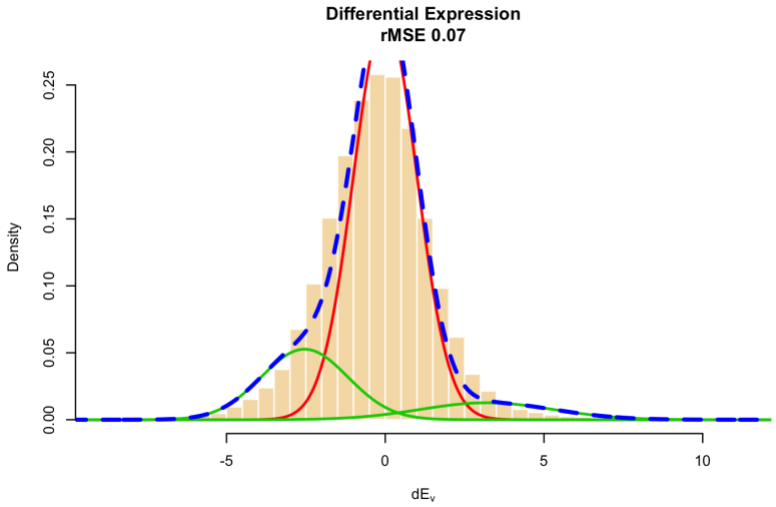

P-values - Differential expression
-log10(p)
0
5
10
15
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
p-value

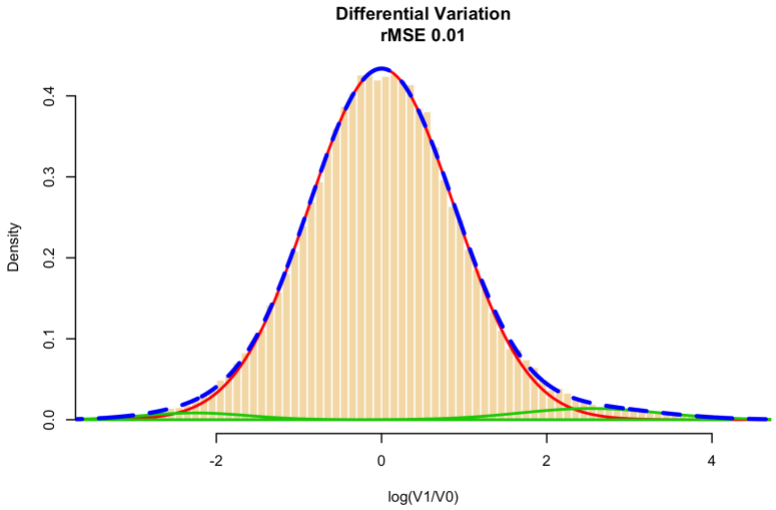

P-values - Differential variation
-log10(p)
0
2
4
6
8
10
12
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
p-value

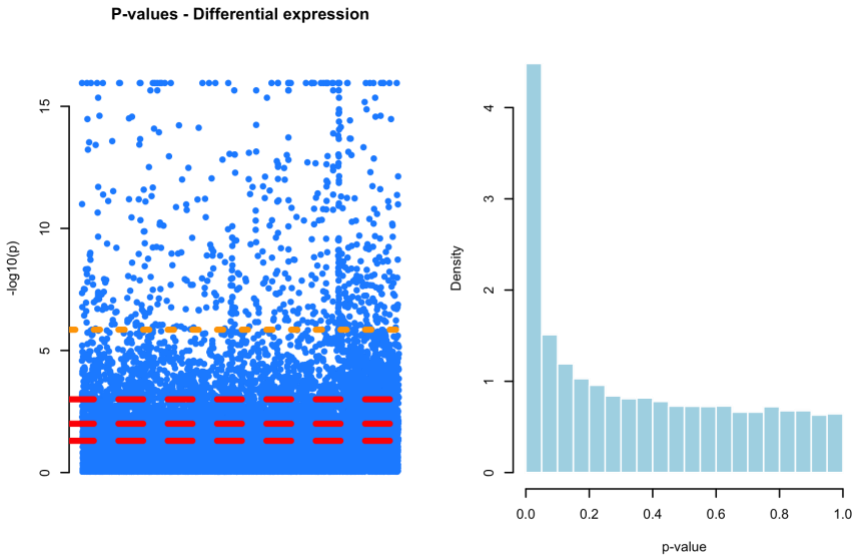
WT vs duplication - N3

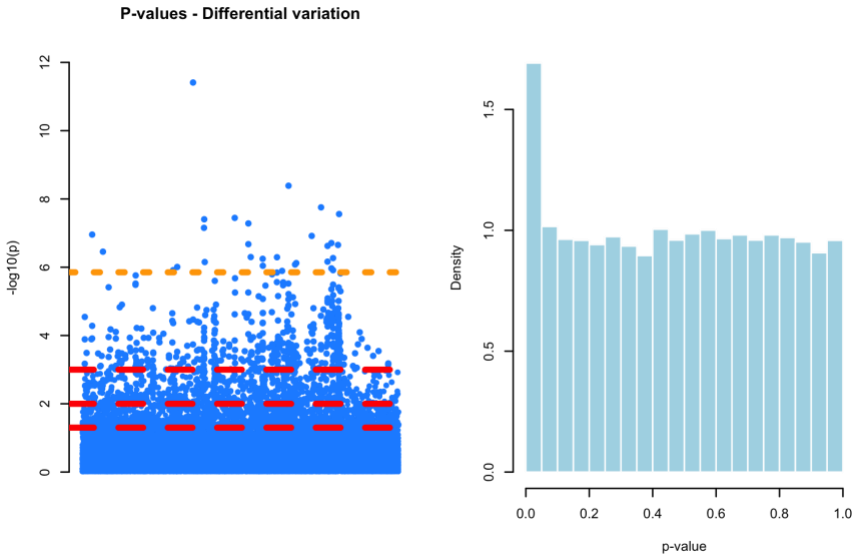
Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: dp/+ duplication - 3 copies
Model: N3
Predictors: individual





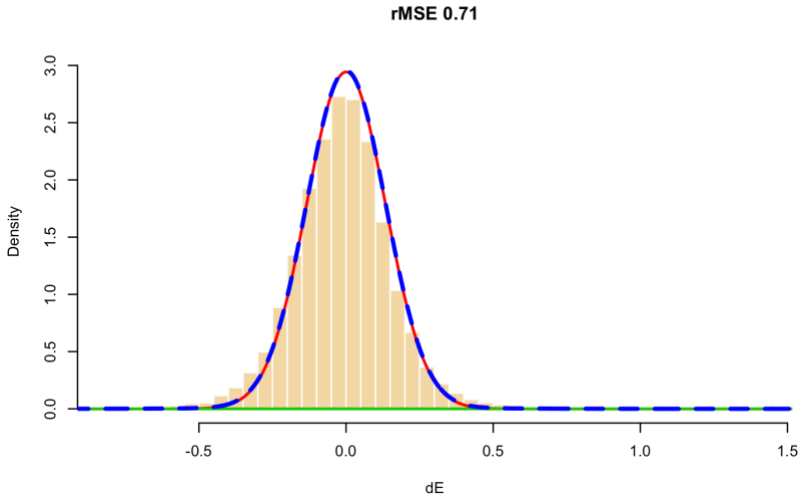

P-values - Differential expression
-log10(p)
0
5
10
15
Density
0
1
2
3
4
0.0
0.2
0.4
0.6
0.8
1.0
p-value

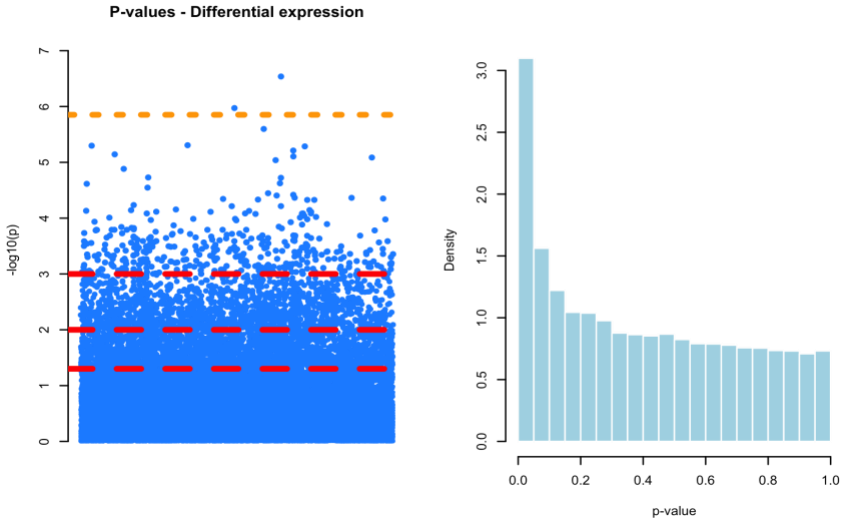

P-values - Differential variation
-log10(p)
0
2
4
6
8
10
12
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
p-value

WT vs. duplication - limma

Model Specification

Dataset: eSetGDS4430.RData
Differential factor: genotype/variation
Baseline level: +/+ wild type - 2 copies
Treatment level: dp/+ duplication - 3 copies
Model: limma
Predictors: individual





Summary of the WT vs. CNV=3 analysis, using q-value threshold of 0.05 to determine differential genes:

	Differential variation		Differential expression	
	rMSE	Genes with q≤ 0.05	rMSE	Genes with q≤ 0.05
L2N	0.01	330	0.01	4,163
N3	0.01	349	0.07	4,346
limma	NA	NA	0.71	44

L2N has the smallest rMSE. Both L2N and N3 model suggest that duplication of the allele is associated with significant differential variation.