

WWIEM Bioinformatics Service

97 Lisburn Road|
Belfast, UK| BT9 7BL
+44*****

Report for RNA-seq Service

Customer Name,
Customer Address
City, Zip
UK



QUEEN'S
UNIVERSITY
BELFAST

WELLCOME-WOLFSON
INSTITUTE FOR
EXPERIMENTAL MEDICINE

Comprehensive Data Analysis Report

This report contains the results of a completed RNAseq experiment. RNA was isolated from 4 mouse samples and sequenced on a Sanger Illumina 1.9. An RNA-seq poly(A) library was generated according to Illumina's instructions using oligo-dT primers. Datasets was derived from ES cells after NeuroD1 induction for 48h (+Dox) or under non-induced condition (-Dox) in two biological replicates each.

Comparisons were completed as requested.

Quality check was performed using FASTQC v0.11.8, the output was aligned to the mouse genome (GRCm39). Salmon v.1.0.0 was used for mapping and calculating read counts per gene. The DESeq R package was used for normalized read counts generation and for differential gene expression analysis. The package called differentially expressed genes with FDR cutoff of 0.05 and FC ≥ 1 were considered as significant differentially expressed genes. GO terms and pathways analysis were performed using KEGG.

*Link to folder containing all data:
Sample_Link_To_Data_For_RNASeq_GEA*

Table of Contents

<i>Table of Contents</i>	<i>Page</i>
<i>Sample Key</i>	<i>4</i>
<i>QC Metrics</i>	<i>5</i>
<i>Summary of Mapping & Quantification</i>	<i>22</i>
<i>Differential Data Analysis</i>	<i>24</i>
<i>Pathway Data Analysis</i>	<i>30</i>

Samples & Comparison Information

This table provides a key to the sequenced samples and the comparison groups to which they belong. It also lists the comparisons that were made for this report.

Group Information	Samples	Group Comparison
A	<i>VT_380-CGATGT_L001_R1</i>	<i>Group B vs A</i>
	<i>VT_380-CGATGT_L002_R1</i>	
	<i>VT_380-CGATGT_L003_R1</i>	
	<i>VT_380-CGATGT_L004_R1</i>	
B	<i>VT_381-TGACCA_L001_R1</i>	
	<i>VT_381-TGACCA_L002_R1</i>	
	<i>VT_381-TGACCA_L003_R1</i>	
	<i>VT_381-TGACCA_L004_R1</i>	
A	<i>VT_382-ACAGTG_L001_R1</i>	
	<i>VT_382-ACAGTG_L002_R1</i>	
	<i>VT_382-ACAGTG_L003_R1</i>	
	<i>VT_382-ACAGTG_L004_R1</i>	
B	<i>VT_383-GCCAAT_L001_R1</i>	
	<i>VT_383-GCCAAT_L002_R1</i>	
	<i>VT_383-GCCAAT_L003_R1</i>	
	<i>VT_383-GCCAAT_L004_R1</i>	

QC Metrics

Detailed metrics can be found in the following link:
XXXXXXXXXXXXXXXXXXXX

Sample VT_380-CGATGT_L001_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_380_CGATGT_L001_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3076908
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_380-CGATGT_L002_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_380_CGATGT_L002_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3161892
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_380-CGATGT_L003_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_380_CGATGT_L003_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3073324
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_380-CGATGT_L004_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_380_CGATGT_L004_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	3362168
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_381_TGACCA_L001_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_381_TGACCA_L001_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2420015
Sequences flagged as poor quality	0
Sequence length	51
%GC	48

Sample VT_381_TGACCA_L002_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_381_TGACCA_L002_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2491972
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_381_TGACCA_L003_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_381_TGACCA_L004_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2414188
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_381_TGACCA_L004_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_381_TGACCA_L004_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2669241
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_382_ACAGTG_L001_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_382_ACAGTG_L001_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2189850
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_382_ACAGTG_L002_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_382_ACAGTG_L002_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2245727
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_382_ACAGTG_L003_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_382_ACAGTG_L003_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2178366
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_382_ACAGTG_L004_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_382_ACAGTG_L004_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2412516
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_383_GCCAAT_L001_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_383_GCCAAT_L001_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1995840
Sequences flagged as poor quality	0
Sequence length	51
%GC	48

Sample VT_383_GCCAAT_L002_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_383_GCCAAT_L002_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2057756
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_383_GCCAAT_L003_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_383_GCCAAT_L003_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	1986106
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Sample VT_383_GCCAAT_L004_R1

Basic Statistics

This table shows some simple composition statistics for the file analysed

Measure	Value
File Name	VT_383_GCCAAT_L004_R1_002.fastq.gz
File Type	Conventional base calls
Encoding	Sanger / Illumina 1.9
Total Sequences	2226561
Sequences flagged as poor quality	0
Sequence length	51
%GC	47

Summary of Mapping

Sample	VT_380_L001	VT_380_L002	VT_380_L003	VT_380_L004
Group	A	A	A	A
Index	CGATGT	CGATGT	CGATGT	CGATGT
Reads	3076908	3161892	3073324	3362168
Total Mapped Reads	2918824	3003582	2775231	3184036
Mapping Rate	94.8622 %	94.9932 %	90.3006 %	94.7019 %

Sample	VT_381_L001	VT_381_L002	VT_381_L003	VT_381_L004
Group	B	B	B	B
Index	TGACCA	TGACCA	TGACCA	TGACCA
Reads	2420015	2491972	2414188	2669241
Total Mapped Reads	2346518	2415603	2232398	2580121
Mapping Rate	96.963 %	96.9354 %	92.4699 %	96.6612 %

Sample	VT_382_L001	VT_382_L002	VT_382_L003	VT_382_L004
Group	A	A	A	A
Index	ACAGTG	ACAGTG	ACAGTG	ACAGTG
Reads	2189850	2245727	2178366	2412516
Total Mapped Reads	2089261	2142160	1957814	2294061
Mapping Rate	95.4066 %	95.3883 %	89.8753 %	95.09 %

Sample	VT_383_L001	VT_383_L002	VT_383_L003	VT_383_L004
Group	B	B	B	B
Index	GCCAAT	GCCAAT	GCCAAT	GCCAAT
Reads	1995840	2057756	1986106	2226561
Total Mapped Reads	1915307	1974462	1800009	2129790
Mapping Rate	95.965 %	95.9522 %	90.6301 %	95.6538 %

Differential Data Analysis

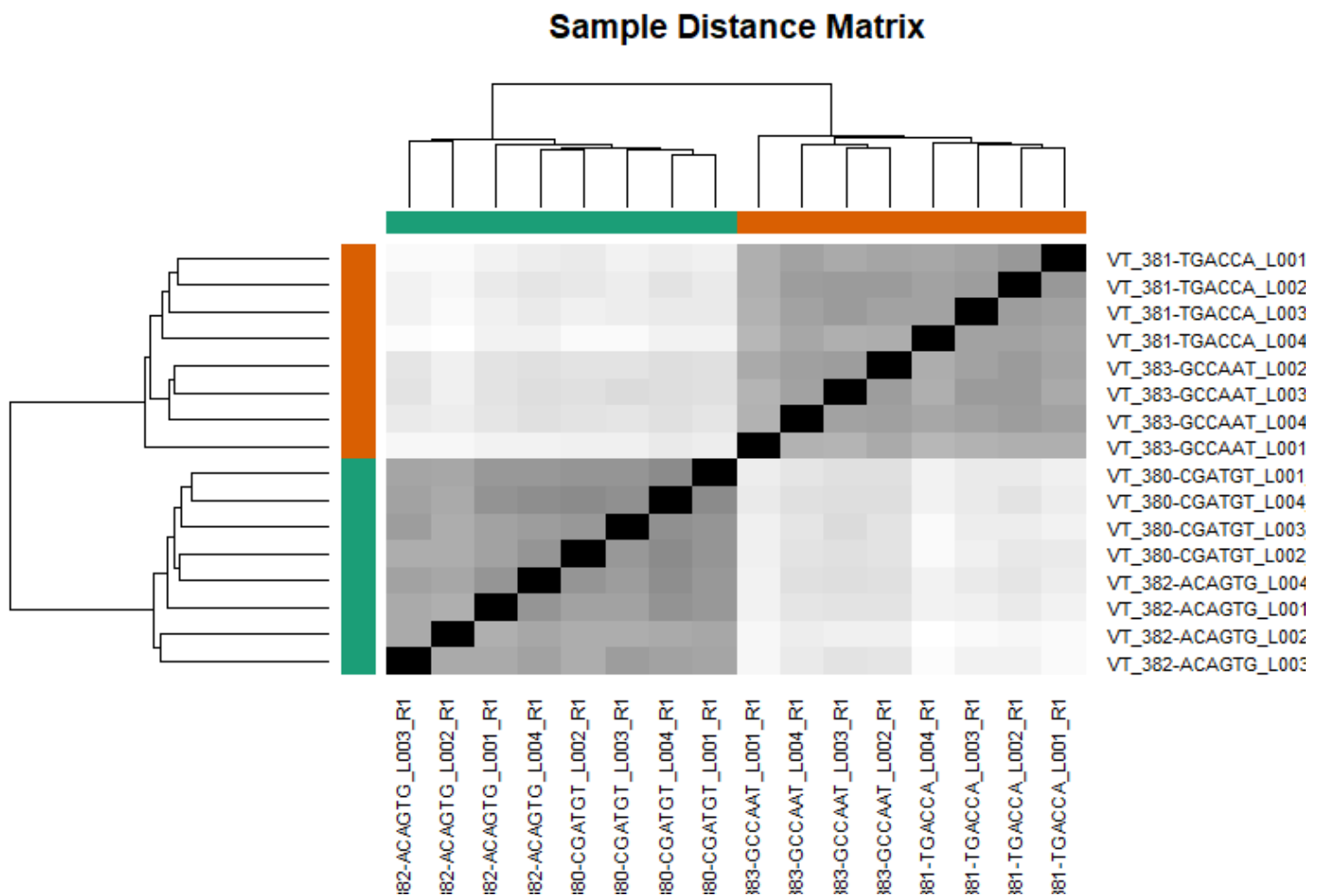
Supplementary result files can be found following this link:

XXXXXXXXXXXXXXXXXX

Exploring results

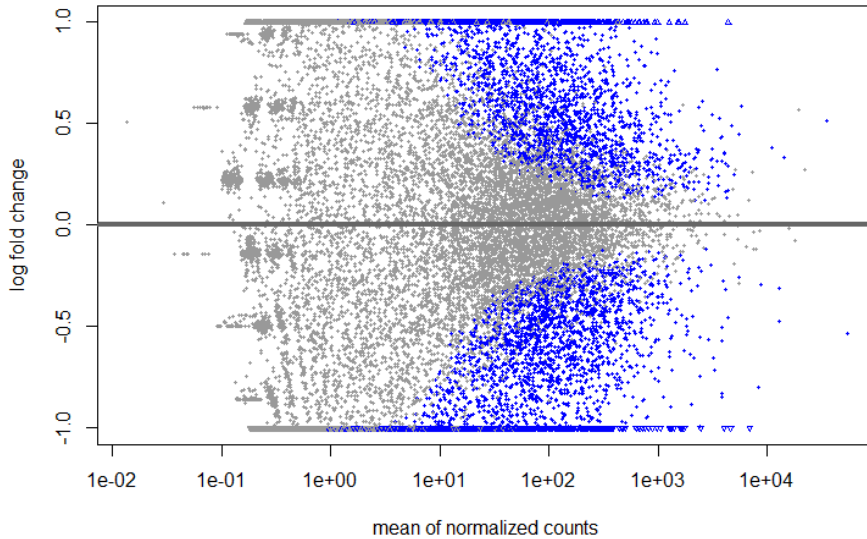
Heatmap of the sample to sample distances

The heatmap below give us an overview over similarities and dissimilarities between samples.



MA-plot

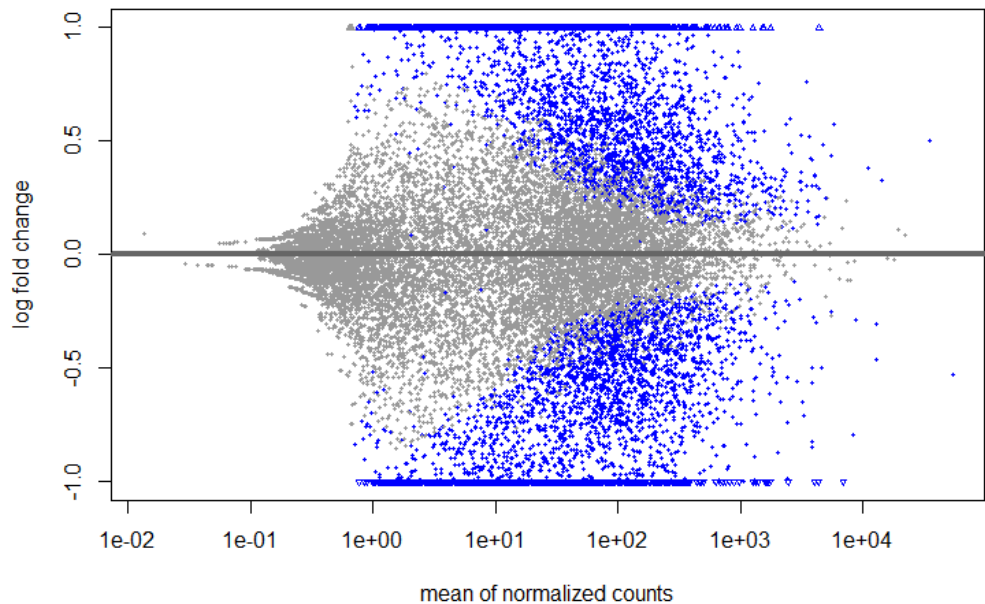
The plot below shows the log₂ fold change over the mean of normalized counts for all the samples. Points plotted as open triangles either up or down are points that fall out of the window.



MA-plot for shrunken log₂ fold changes

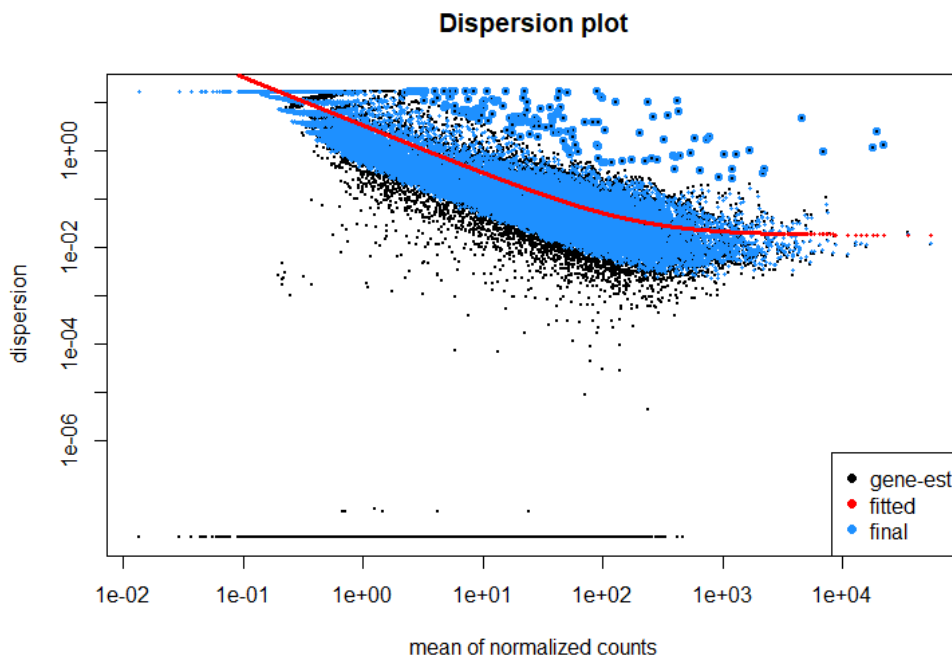
The graph below visualizes the MA-plot for the shrunken log₂ fold changes which remove the noise associated with log₂ fold changes from low count genes.

MA-Plot for shrunken log₂ fold changes: Condition B vs A



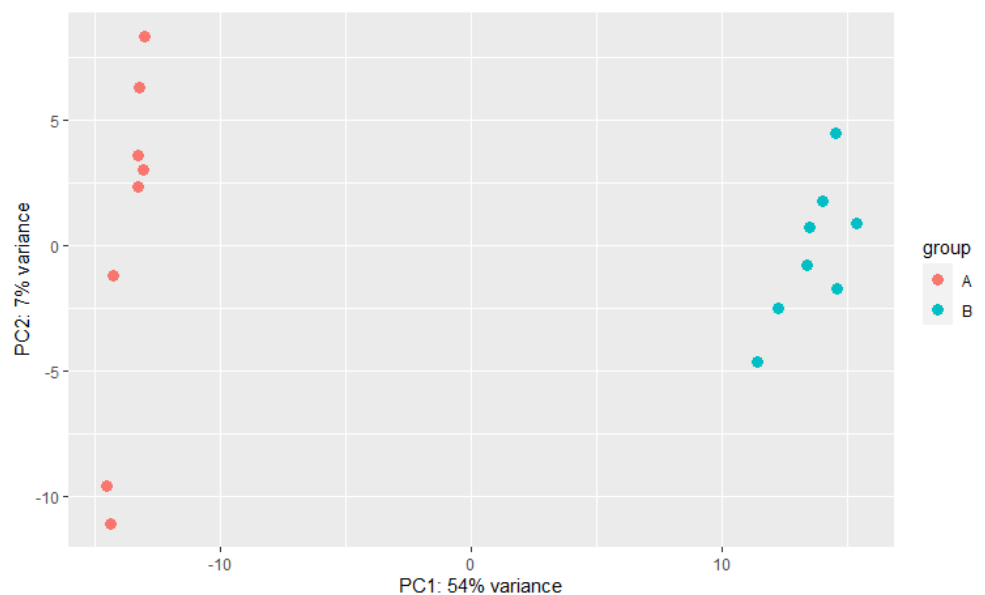
Dispersion plot

The dispersion plot below shows the final estimates shrunk from gene-wise estimates towards fitted estimates. Some gene-wise estimates are flagged as outliers and not shrunk towards fitted value.



Principal Component plot

The PCA plot below shows the samples in 2D plane spanned by their first two principal components. The samples scored a 54% variance in the first component, while 7% variance was noticed for the second component.



Top Differentially Expressed Genes

The table shows the top 25 genes with the greatest differential expression and an adjusted p-value ≤ 0.05 are shown below. The adjusted p-values were calculated using the Benjamini-Hochberg false discovery rate adjustment.

Gene	Linear Fold Change	P-value	Adjusted p-value
ENSMUSG00000002265	2.794368	0	0
ENSMUSG000000068735	3.435468	0	0
ENSMUSG000000025867	5.227207	4.62E-248	2.45E-244
ENSMUSG000000020844	2.587876	3.49E-226	1.11E-222
ENSMUSG000000034701	4.030701	1.17E-223	3.10E-220
ENSMUSG000000045991	5.826352	6.24E-207	1.42E-203
ENSMUSG000000052727	2.488145	2.90E-200	5.77E-197
ENSMUSG000000062380	4.159948	4.91E-196	8.66E-193
ENSMUSG000000069662	2.041123	1.84E-195	2.93E-192
ENSMUSG000000026728	2.137485	5.16E-186	6.83E-183
ENSMUSG000000032394	2.990115	1.05E-172	1.28E-169
ENSMUSG000000063632	2.175593	1.07E-165	1.21E-162
ENSMUSG000000067071	2.816204	2.37E-144	2.09E-141
ENSMUSG000000028654	2.897297	4.64E-129	3.68E-126
ENSMUSG000000051747	3.031766	5.82E-129	4.40E-126
ENSMUSG000000024304	2.07102	1.21E-126	8.72E-124
ENSMUSG000000032291	4.354454	3.01E-126	2.08E-123
ENSMUSG000000029838	3.424072	2.36E-124	1.44E-121
ENSMUSG000000034486	3.285496	2.30E-118	1.18E-115
ENSMUSG000000049532	2.200979	5.00E-118	2.48E-115
ENSMUSG000000042436	5.95148	5.62E-108	2.41E-105
ENSMUSG000000024548	2.433491	1.48E-106	6.19E-104
ENSMUSG000000047261	3.271865	7.19E-106	2.93E-103
ENSMUSG000000020644	2.653197	1.49E-104	5.93E-102
ENSMUSG000000026185	4.034456	1.68E-104	6.51E-102

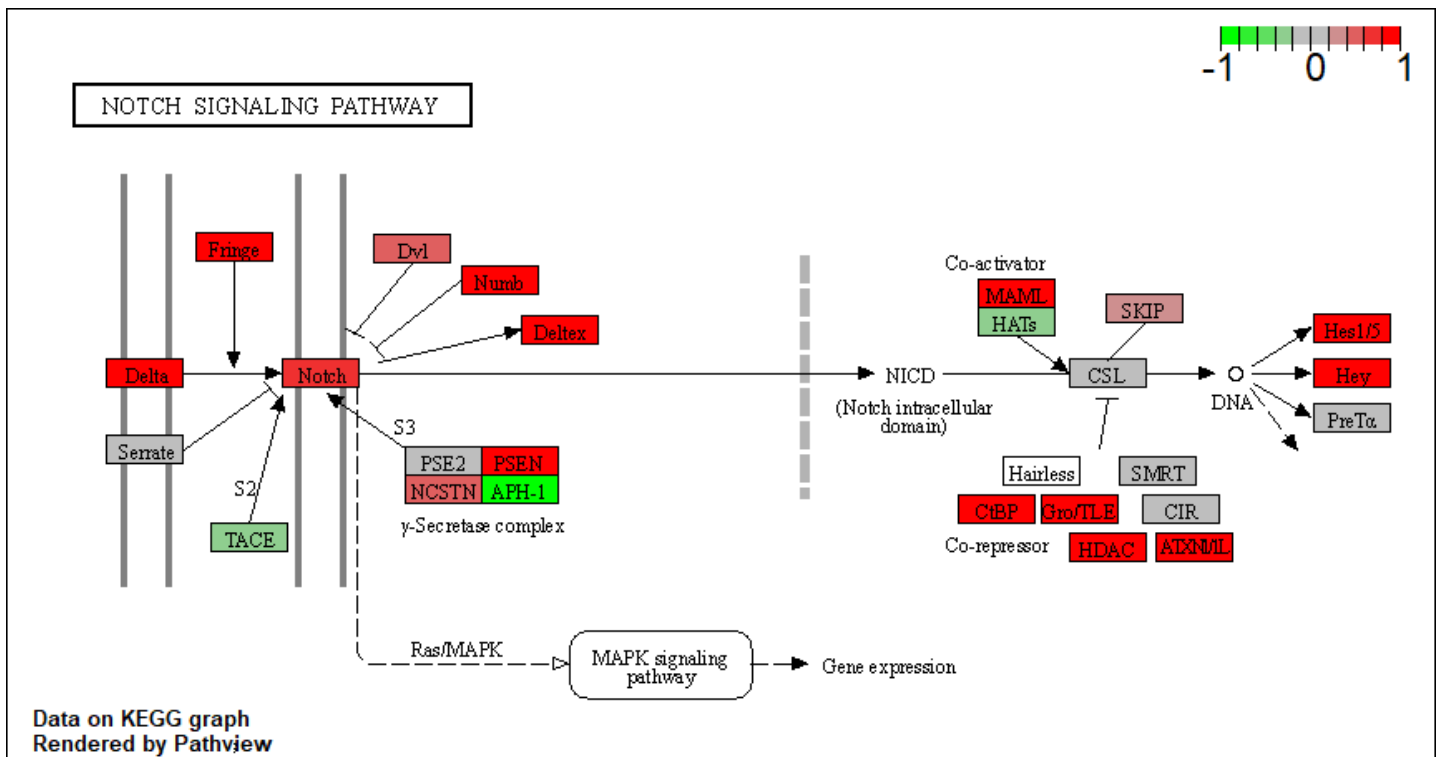
Pathway Data Analysis

KEGG Pathway Analyses

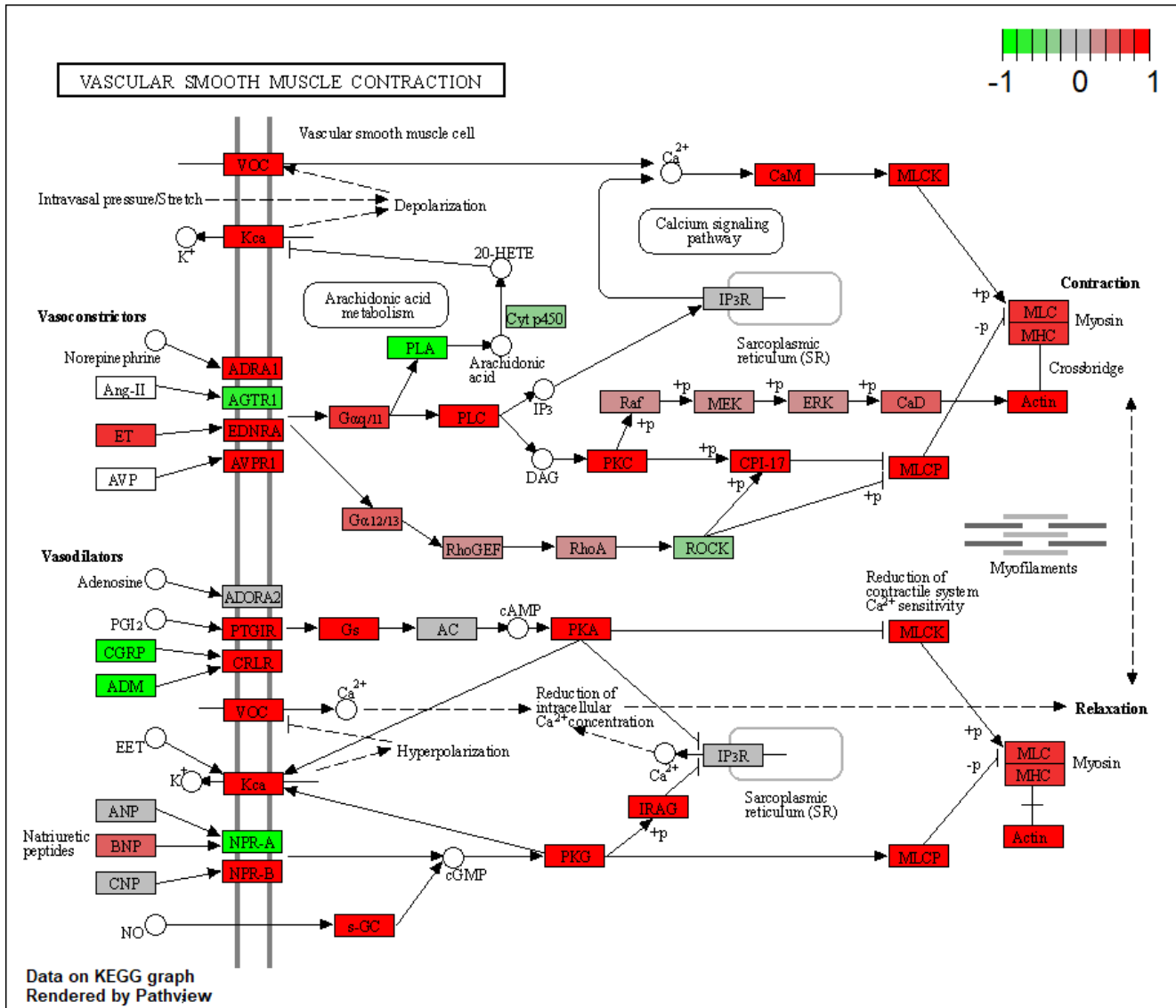
The log2 fold changes for all genes were used to analyze Kegg pathways. Pathways were first filtered by p-value ($p \leq 0.05$) and then by mean of logFC. The top 5 upregulated pathways were pulled out.

Top 5 Significant Pathways

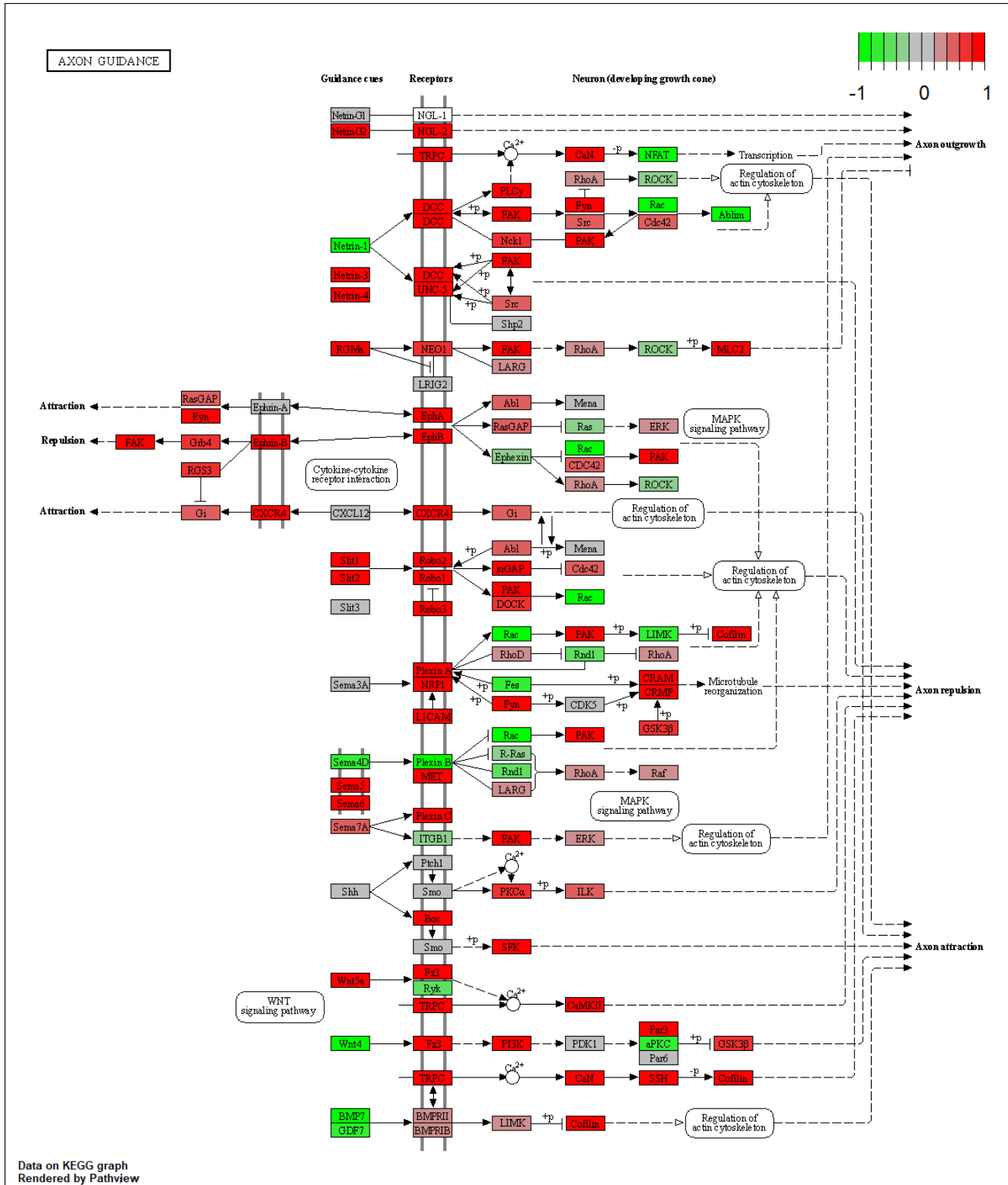
Notch Signaling pathway



Vascular Smooth Muscle Contraction

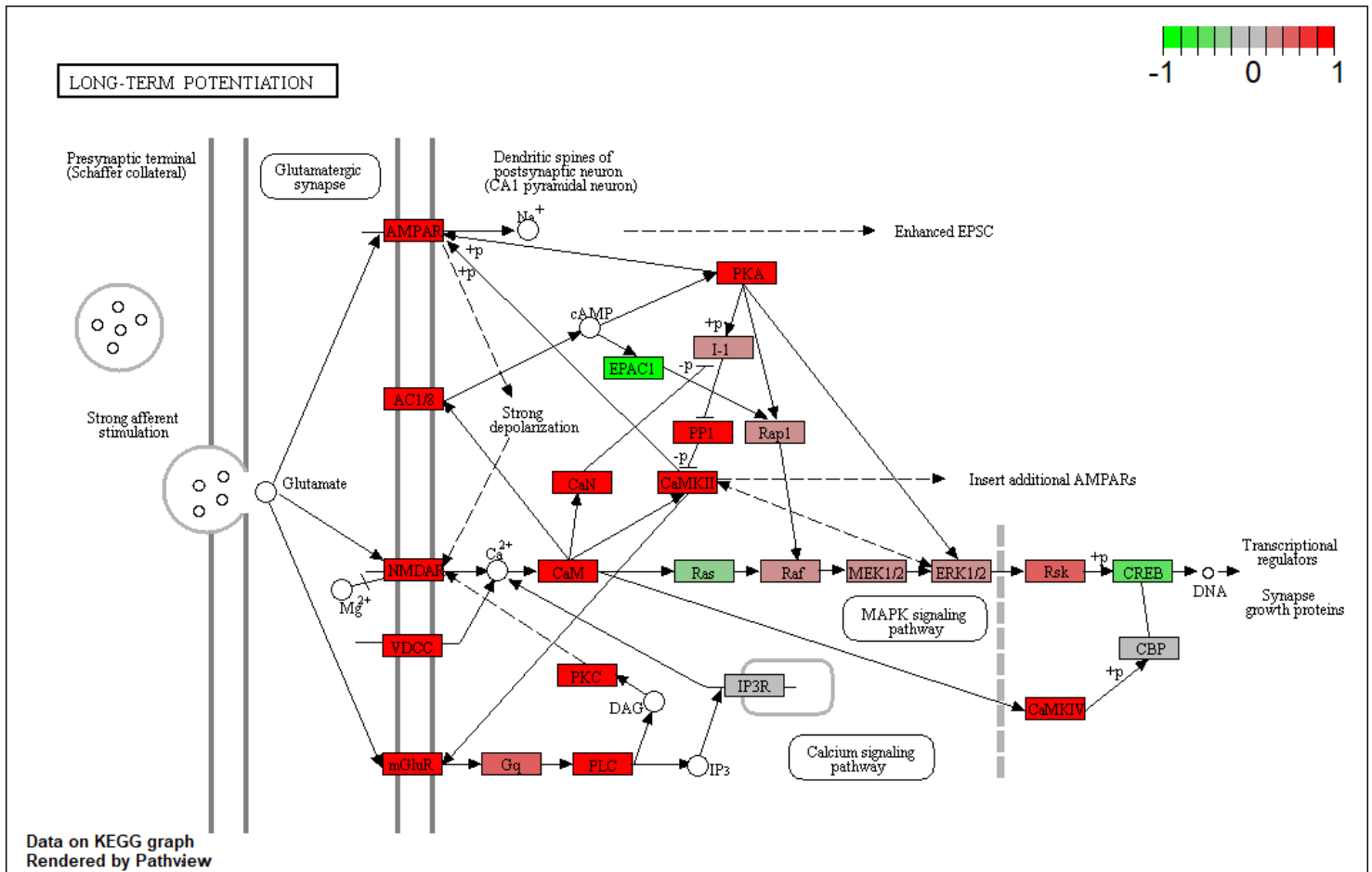


Axon Guidance



Data on KEGG graph
Rendered by Pathview

Long-term Potentiation



Neurotrophin Signaling Pathway

