Title
Temporal Regulation of Gene Expression Profiles in Rat Brains Following Ischemic Stroke

Author
Omotayo, Hakeem

Publication Date
2019

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
RIVERSIDE

Temporal Regulation of Gene Expression Profiles in Rat Brains Following Ischemic Stroke

A Thesis submitted in partial satisfaction
of the requirements for the degree of

Master of Science

in

Bioengineering

by

Hakeem Omotayo

December 2019

Thesis Committee:
Dr. Byron D. Ford, Chairperson
Dr. Victor G. J. Rodgers
Dr. Boris-Hyle Park

The Thesis of Hakeem Omotayo is approved:

Committee Chairperson

University of California, Riverside

ABSTRACT OF THE THESIS

Temporal Regulation of Gene Expression Profiles in Rat Brains Following Ischemic Stroke

by

Hakeem Omotayo

Master of Science, Graduate Program in Bioengineering
University of California, Riverside, December 2019
Dr. Byron D. Ford, Chairperson

To improve understanding of the molecular mechanisms that underlie ischemic stroke, we analyzed early gene expression profiles in permanent middle cerebral artery occlusion (pMCAO) stroke model. Rats were allocated into 4 groups: control, 3h MCAO, 6h MCAO, and 12h MCAO. Cortical brain tissue was collected after stroke and subjected to microarray analysis. This information was then analyzed using a series of bioinformatic tools including Transcriptome Analysis Console (TAC), Genesis, Short time-series Expression Miner (STEM), and STRING. A number of differentially expressed genes were found to be either upregulated or downregulated in a temporal manner. At 3 hours post-stroke, 230 genes were upregulated and 21 downregulated, 635 were upregulated and 238 downregulated at 6 hours, 1033 were upregulated and 741 downregulated at 12 hours following MCAO. STEM analysis highlighted several distinct patterns of gene activity. Genes that were temporally upregulated were highly associated with inflammatory response and apoptosis, while genes that were temporally downregulated

after stroke were strongly associated with cell membrane components and neurons. Genes that decreased at 3 hours and increased afterwards were also associated with inflammatory response. Genes that increased the 1st 6 hours and decreased slightly at 12 hours were associated with RNA binding with indication that a key regulator THOC1 may also be involved in apoptosis. Lastly genes that decreased at 6 hours but remained flat before and after were associated with hydroxy compound transport. The temporally upregulated genes separated into two distinct hubs: the 1st was mostly associated with cytokines and chemokines, and the 2nd was mostly made up of RGD, Rps, Eif, and ESNROG proteins which are mostly ribosomal or associated with translation. The main regulator in the 2nd hub Rsl1d1 is associated with apoptosis. The main regulators of inflammatory response appeared to be Il6, Il1b, and Ccl2. STRING analysis revealed no distinct pattern for the temporally downregulated genes. These results provide evidence that in addition to the neuronal death and inflammatory responses following stroke, regulation of translation may play an important role. Understanding the transcriptional mechanisms following ischemia may provide therapeutic targets for treating stroke.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Introduction

Stroke

Stroke is the sudden interruption in the blood supply to the brain(1). Stroke is the leading cause of adult disability and the 5th leading cause of death in the United States(2). Roughly 795,000 people in the United States have as stroke each year(3). Ischemic stroke occurs when an artery supplying blood to the brain becomes blocked. Roughly 87% of all strokes are ischemic (3, 4). The remaining 13% are hemorrhagic, caused by rupture of a blood vessel(4).

Ischemic Stroke

Due to their higher likelihood, ischemic stroke is the more commonly studied and therefore the type we modeled for this study. Ischemic stroke causes an area of initial damage due to oxygen and nutrient deprivation which occurs within minutes to hours(5) called the ischemic core. The core is characterized by an early onset neuronal death that begins within minutes following stroke. This area of immediate brain injury is characterized by low cerebral blood flow, energy failure and excitotoxicity(6, 7). This is followed by a much larger area of inflammatory secondary damage which occurs after hours to days called the penumbra. The resulting ischemic brain injury in the penumbra is accompanied by increased synthesis of inflammatory molecules and cytokines in neurons, glia and in the cerebral vasculature(6, 8-11). This inflammatory response endangers brain cells in the surrounding penumbra where blood supply is compromised but not completely interrupted (11-13).

Stroke and inflammatory response

In the infarct core, brain ischemia causes early failure of ion pumps, loss of membrane integrity and necrotic cell death(14). Additionally, arterial occlusion leads to intravascular hypoxia, changes in shear stress and production of reactive oxygen species which activate the coagulation cascade causing further occlusion(14).

Early neuronal death leads to the release of danger associated molecular patterns (DAMPs) such as extracellular ATP, heat shock proteins , HMGB1 and others which lead to activation of immune response(15).

Neuronal injury and death also lead to excess glutamate release and excitotoxicity in regions of low blood flow. Astrocytes take in excess glutamate and other ions, to try to restore balance, but quickly become overwhelmed. These astrocytes may enter a proinflammatory state releasing inflammatory cytokines and exacerbating blood brain barrier disruption.

Regardless of the cell type, a variety of proinflammatory factors are released following stroke which may increase the area of damage.

Stroke Microarray Studies

Early stroke studies only looked at mRNA expression for small number of genes of interest individually. However, this single gene approach does not capture the complex interactions or redundancy within the genome. Instead, a transcriptome analysis approach of looking at a large number of gene changes simultaneously offers the opportunity to home in on possible regulating gene groups. Microarray technology allows for the rapid analysis of a large amount of gene data(16). In a microarray, mRNA from a sample is hybridized in parallel to a large number of DNA sequences and immobilized on a solid surface in an ordered array(16). Next mRNA from the sample is extracted, converted to DNA and labelled,

hybridized to the DNA elements of the array, and detected by fluorescence scanning or phosphor-imaging(16).

Several prior microarray studies have been performed to understand gene expression following ischemic stroke. The first microarray ischemia study looked at changes in the striatum and cortex 3h after ischemia and found that of the 24 genes differentially regulated 2 fold or more, most were immediate early genes such as c-fos, NGFI-A, NGFI-C, Krox-20, and Arc(17). Another study looked at DNA regulation in cortical slices 6h after MCAO and identified several genes such as IFN, NDGAP-1, and NPR which had not previously been shown to be modulated following focal ischemia(18). Yet another study looked at RNA samples from the periinfarct cortex of rats 24h after permanent middle cerebral artery occlusion (pMCAO) and found that of 328 differentially expressed genes in ischemia compared to SHAM 163 of these had not previously been reported in stroke(19).The study also linked the genes to 14 different functional categories including metabolism related genes, stress response proteins, neurotransmitter/hormone related genes, and cytokine/chemokine related genes(19). Some of our previous studies have even shown differences in gene expression between the transient middle cerebral artery occlusion (tMCAO) and pMCAO models 24h after stroke(20, 21) with pMCAO more associated with neurotransmitter receptors, ion channels and growth factors and tMCAO more associated with inflammation and apoptosis(20). The pMCAO model may be more similar to an untreated stroke, as it may take days for reperfusion to occur by the body's endogenous processes(20).

Each of these studies was able to give a single snapshot, but many of these genes have dynamic expression patterns. Understanding these may help us better understand the mechanisms behind gene expression in stroke.

Clustering Techniques

General methods such as k-means clustering(22), hierarchical clustering(23), and self-organizing maps(24) are often used to analyze gene expression data(25). However, these methods do not account for temporal dependency among successive time points(25). It is also important to note that due to the large data size, many genes may have the same expression pattern due to random chance(25, 26). Short Time-series Expression Miner (STEM) is a program that clusters, compares and visualizes microarray data for a small series of 3-8 timepoints (25). STEM takes advantage of the large number of genes and small number of timepoints to identify statistically significant temporal expression profiles and genes associated with these profiles(26). Additionally, STEM can link Gene Ontologies to sets of genes that have the same temporal expression patterns(25).

STEM Clustering in Other Types of Studies

STEM has already been used in a time-series spinal cord injury model paper, which looked at gene fold-change values at different time points after injury(27). In that study STEM was used to associate deregulated transcripts into eight groups and associate Gene Ontologies with each of them(27). In a study on inflammatory responses in amnion cells after cytokine stimulation, STEM was used because hierarchical clustering was not sufficient to group the large variation in co-expressed genes based on relative levels of induction(28). STEM was also used to associate similar groups to Gene Ontologies in that study(28). In a study of gene expression responses in the brain and heart to Ephedra herba, STEM was used to select temporally expressed genes(29).

Scope of this Study

Though several studies have used microarray analysis to look at gene changes at a single timepoint. This is the 1st stroke study using STEM to evaluate temporal expression of genes. Because microarrays encompass a large number of genes, this method allows for clustering of genes based on temporal expression, which may give insight to possible mechanisms of gene regulation following stroke. This information may allow us to look for potential therapeutic targets for stroke treatment.

Methods

Animals and Ischemia Induction

All animals were treated humanely and with regard for alleviation of suffering and pain and all surgical protocols involving animals were performed by sterile/aseptic techniques and were approved by the Institutional Animal Care and Use Committee at Morehouse School of Medicine prior to the initiation of experimentation. Male adult Sprague - Dawley rats (250-300g; Charles River Laboratory International, Inc., USA) were housed in standard cages in a temperature-controlled room (22 ± 2°C) on a 12 h reverse light-dark cycle. Food and water were provided *ad libitum*.

Animals were randomly allocated into 4 groups: SHAM(control), 3, 6 or 12 hour MCAO. All rats except the SHAM group were subjected to a left permanent middle cerebral artery occlusion (pMCAO). After anesthesia administration, a rectal probe monitored the core body temperature and a Homoeothermic Blanket Control Unit (Harvard Apparatus, Hollister, MA) was used to ensure the body temperature maintained at 37 degrees Celsius. Cerebral blood flow was monitored throughout the length of the

surgery by a continuous laser Doppler flowmeter (Perimed, Ardmore, PA), with a laser Doppler probe placed 7mm lateral and 2mm posterior to bregma in a thinned cranial skull window. MCAO was induced by the intraluminal suture method as we previously described (30). Briefly, a 4 cm length 4-0 surgical monofilament nylon suture coated with silicon (Doccol Corp., Sharon, MA, diameter 0.37mm, length 2.3-2.5 mm) was inserted from the external carotid artery (ECA) into the internal carotid artery (ICA) and then into the Circle of Willis, to occlude the origin of the left middle cerebral artery (MCA). Rats in the sham control group underwent the same procedure as those in the injury group, but a filament was not inserted into the ICA. Animals were sacrificed three hours, six hours, or twelve hours after MCAO.

Microarray Analysis

Animals in the control group (SHAM) were sacrificed 3 hours following surgery. While the 3 hour, 6 hour, and 12 hour groups were sacrificed 3 hours, 6 hours, and 12 hours after MCAO respectively. Brains were extracted and sectioned into 2 mm coronal sections (approximately +3.0 to −5.0 from bregma) using a brain matrix. The brains were separated at midline and the injured (left) cortical tissue was isolated. A left hemi-cortical tissue from the sham was used as the control. Total RNA was extracted with TRIzol Reagent (Life Technologies Corp, Carlsbad, CA), quality controlled and quantified by Agilent 2100 Bioanalyzer (Agilent Technology, Santa, Clara, CA). Microarrays were completed according to manufacturing guidelines (Affymetrix Inc., Santa, Clara, CA), with cRNA hybridized to an Affymetrix Rat Genome 2.0st Gene Chip (Affymetrix Inc.).

Transcriptome Analysis Console (TAC)

CEL files from the Affymetrix Rat Genome 2.0 Gene Chip were imported to TAC 4.0 for quality control and to normalize the data through multi array averages and identify differentially expressed genes. For differential expression analyses a cutoff of 2-fold change and p-value of < 0.05 were used. TAC also provided principle component analysis information for the experimental samples. This information was exported as a .txt file. Since there were a number of transcript IDs that did not have corresponding genes, a homemade MATLAB script was used to sort out these transcripts that did not correspond to genes. This list was then saved as an Excel file.

Genesis

Genesis was used to cluster genes that showed similar expression patterns among the treatment groups and more easily visualize them. In order to analyze the chart information in Genesis it had to be converted to "Stanford" format, which meant normalizing the Log base 2 expression data, by subtracting the SHAM baseline from each group. The original file, the Stanford file, and the file containing both were all saved in the same EXCEL workbook under separate sheets. Genesis provided hierarchical clustering and K-means clustering information as well as heatmaps of the gene expression. Each individual cluster was saved as a ".txt" file containing the genes in each cluster.

Enrichr

Enrichr is a free online resource where one can input a list of genes to be associated with Ontologies that were highly enriched among them

http://amp.pharm.mssm.edu/Enrichr/. Ontologies were show on the organismal, cellular, and molecular levels. Enrichr pulls its ontologies from 8 different databases: GO Biological Process, GO Molecular Function, GO Cellular Component, MGI Mammalian Phenotype, Human Phenotype Ontology, Jensen TISSUES, Jensen COMPARTMENTS, and Jensen DISEASES. This program was mainly used to analyze the Genesis dataset, as several other programs had a gene ontology function already built in. Each cluster from Genesis was analyzed in Enrichr as a .txt file. Enrichr provided information on Gene Ontologies linked to each cluster.

Enrichr displayed the top ontologies associated with each cluster as well as four scores: p-value, q-value, z-score, and combined score. The p-value is computed using Fisher's exact test assuming a binomial distribution and independence for probability of any gene belonging to any set. The q-value is an adjusted p-value using the Bejamini-Hocherg method for correction for multiple hypotheses testing. The z-score is computed using a modification to the Fischer's exact test, since Fischer's exact test would produce lower p-values for longer lists, Enrichr precomputes a background expected rank for each term in each gene set library. The z-score is the standard deviation from this expected rank. And the combined score is the z-score multiplied by the natural log of the p-value.

STEM

Stem was used to cluster differentially expressed genes based on their temporal profiles.

The text file from TAC (with blanks removed using MATLAB script), was analyzed by STEM. The data was normalized. Spot IDs were included in the data file, so the box

remains checked. And the bullet next to "normalize data" is filled. Under "Gene info" the Gene Annotation Source and Cross Reference Source should both be "user provided". Using rgd.gaf as the gene annotation file and goa_rat_rna.gpi.gz as the cross reference file. These files can be found http://geneontology.org/page/download-annotations and ftp://ftp.ebi.ac.uk/pub/databases/GO/goa/ respectively. Under options, the default controls are left (STEM Clustering method, 50 model profiles, 2 maximum unit change in model profiles between time points).

Alternatively, the data file can be formatted without Spot IDs. In which case, bubble in "Normalize data", and leave a blank next to "Spot IDs included in the data file".

STRING

String is a free online database of known and predicted protein-protein interactions https://string-db.org/. String was used to identify potential "Hub" genes that may regulate pathways associated with a gene set. String was used to analyze the data sets from the STEM analysis and was used on individual profiles as well as superclusters.

Results

Transcriptome Analysis Console Analysis

To determine gene expression following treatment, microarray analysis was performed on cortical tissues from each experimental group using the Transcriptome Analysis Console (TAC). As a quality control step, principal component analysis (PCA) was performed. As a qualitative assessment, PCA showed similar grouping between

samples from the same experimental condition, as well as a temporal progression away from SHAM for the stroke groups *Figure 1*. Of the 36,685 gene probes on the Affymetrix rat 2.0 chip, 10,464 were annotated for the rat genome. Of those 251 were differentially regulated between 3h MCAO and SHAM with 230 upregulated and 21 downregulated, 873 were differentially regulated between 6h MCAO and SHAM with 635 upregulated and 238 downregulated, finally 1774 were differentially regulated between 12h MCAO and SHAM with 1033 upregulated and 741 downregulated *Figures 2-5*. Thus, the number of upregulated genes and the number of downregulated genes both increase over time. The Venn diagram shows how many differentially expressed genes overlapped between each set of comparisons as well as how many were unique to each condition *Figure 6*. Of the 251 genes differentially expressed between 3h MCAO and SHAM, only 51 were unique to 3h, 302 overlapped with just the 6h comparison, 35 overlapped with just the 12h comparison and 130 genes overlapped across all 3 timepoints. Of the 873 genes differentially expressed between 6h MCAO and SHAM, 302 were unique to 6h, while 406 overlapped with just the 12h condition. Finally of the 1774 genes that were differentially expressed between 12h MCAO and SHAM, 1203 genes were unique to 12h. The gene information from TAC was then analyzed with a variety of different tools to further gain insight about the cellular and molecular nature of the expression patterns.

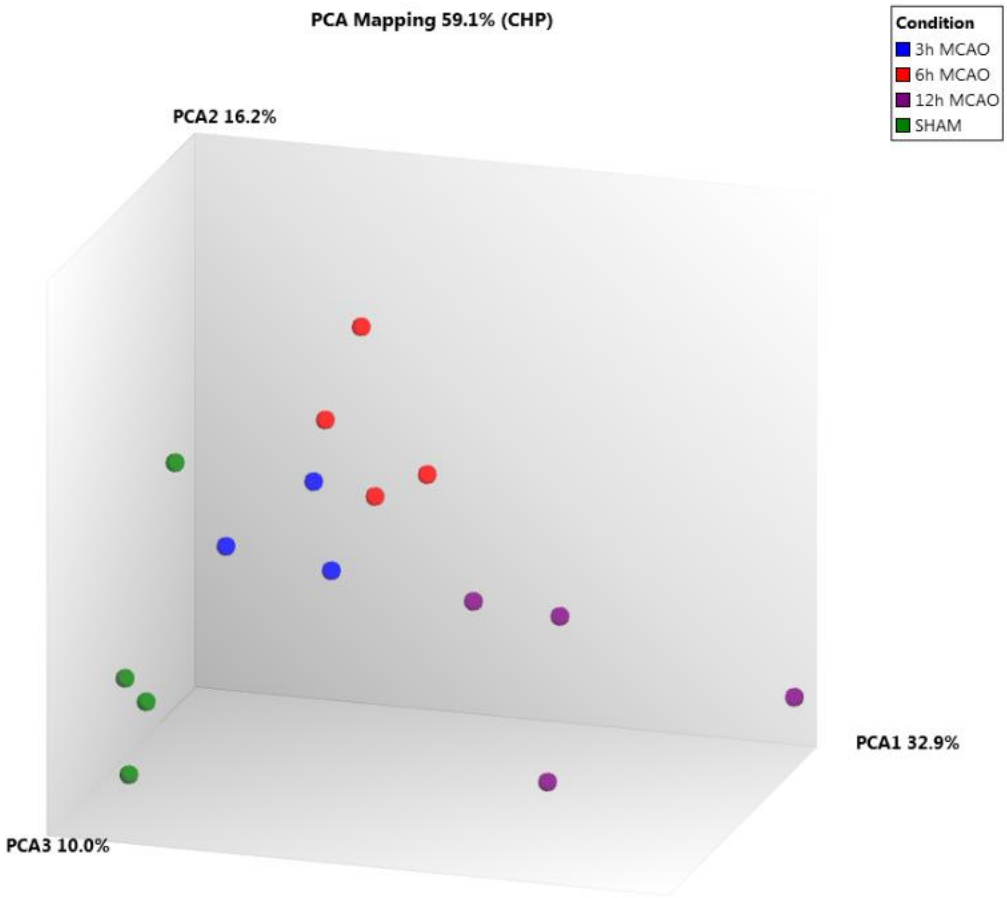


Figure 1. 3D principle component analysis (PCA) graph. SHAM (n=4) animals are represented in green. 3h MCAO (n=3) are represented in blue. 6h MCAO (n=4) are represented in red. And 12h MCAO (n=4) are represented in purple.

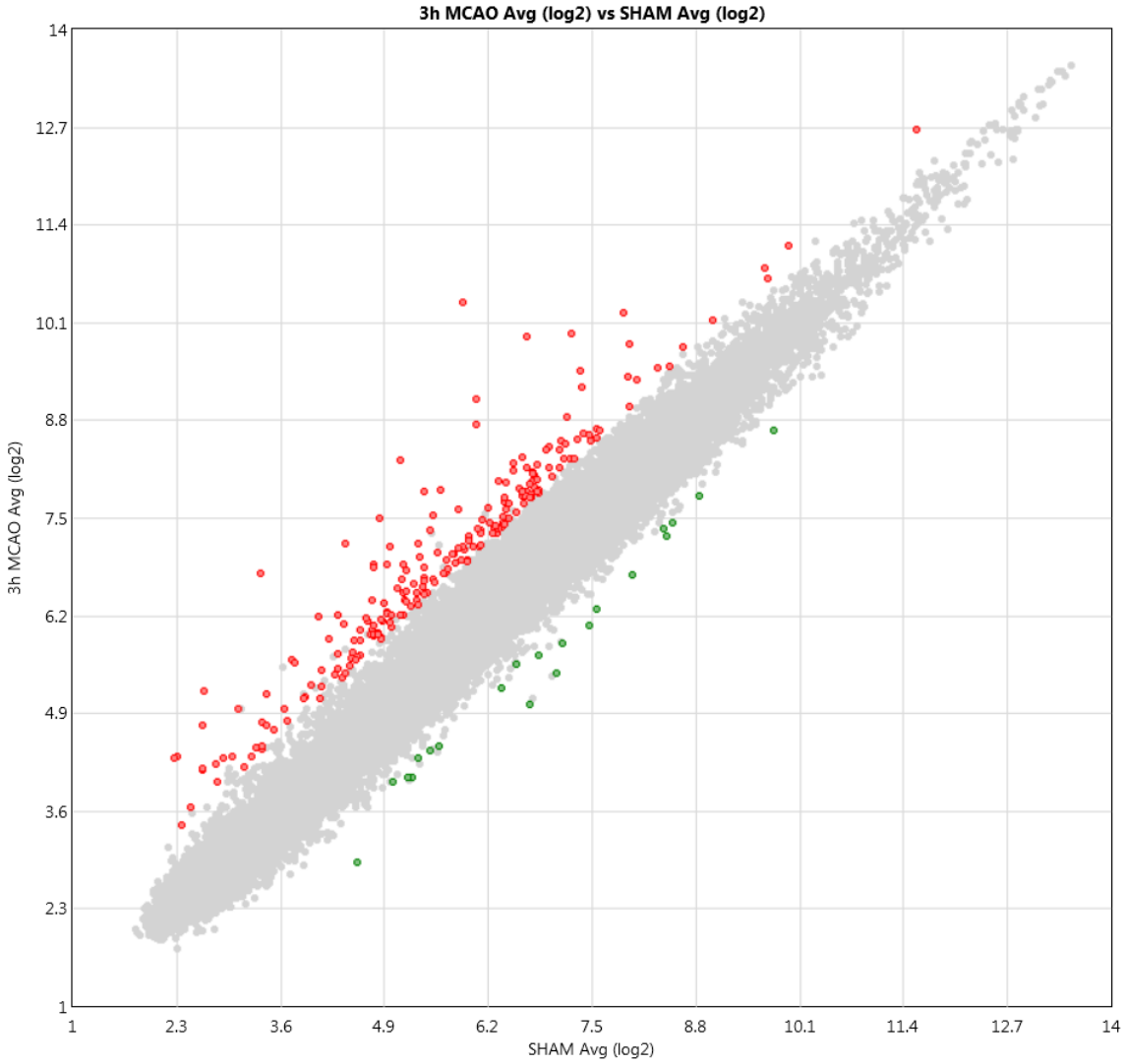


Figure 2. Scatter plot showing the average hybridization signal intensity of the genes in the 3h MCAO (n=3), Compared to the SHAM (n=4) group. The red dots indicate genes upregulated 2-fold or more. The green dots indicate genes downregulated 2-fold or more. And the grey dots indicate genes that were not differentially regulated with at least 2-fold.

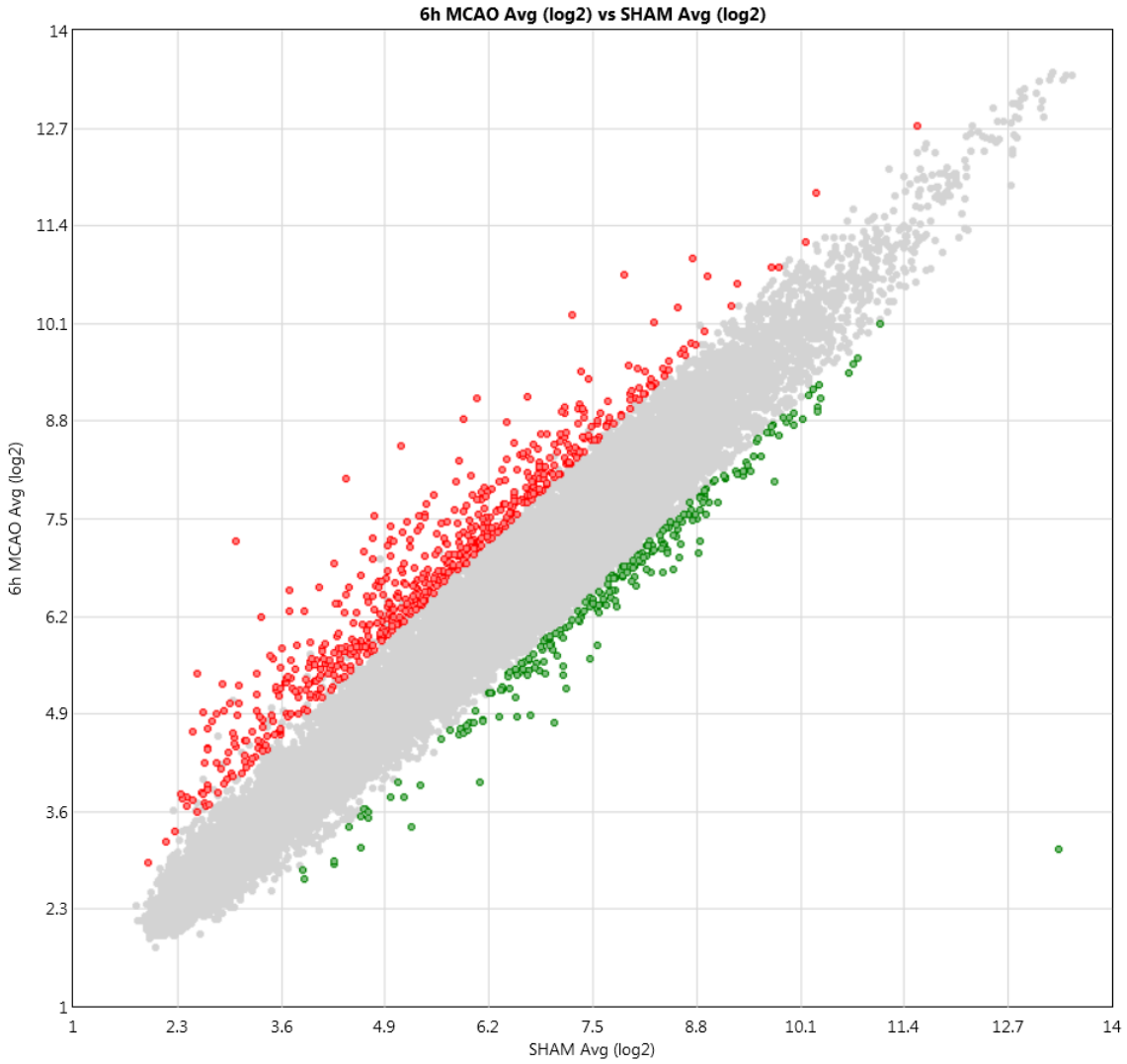


Figure 3. Scatter plot showing the average hybridization signal intensity of the genes in the 6h MCAO (n=4), Compared to the SHAM (n=4) group. The red dots indicate genes upregulated 2-fold or more. The green dots indicate genes downregulated 2-fold or more. And the grey dots indicate genes that were not differentially regulated with at least 2-fold.

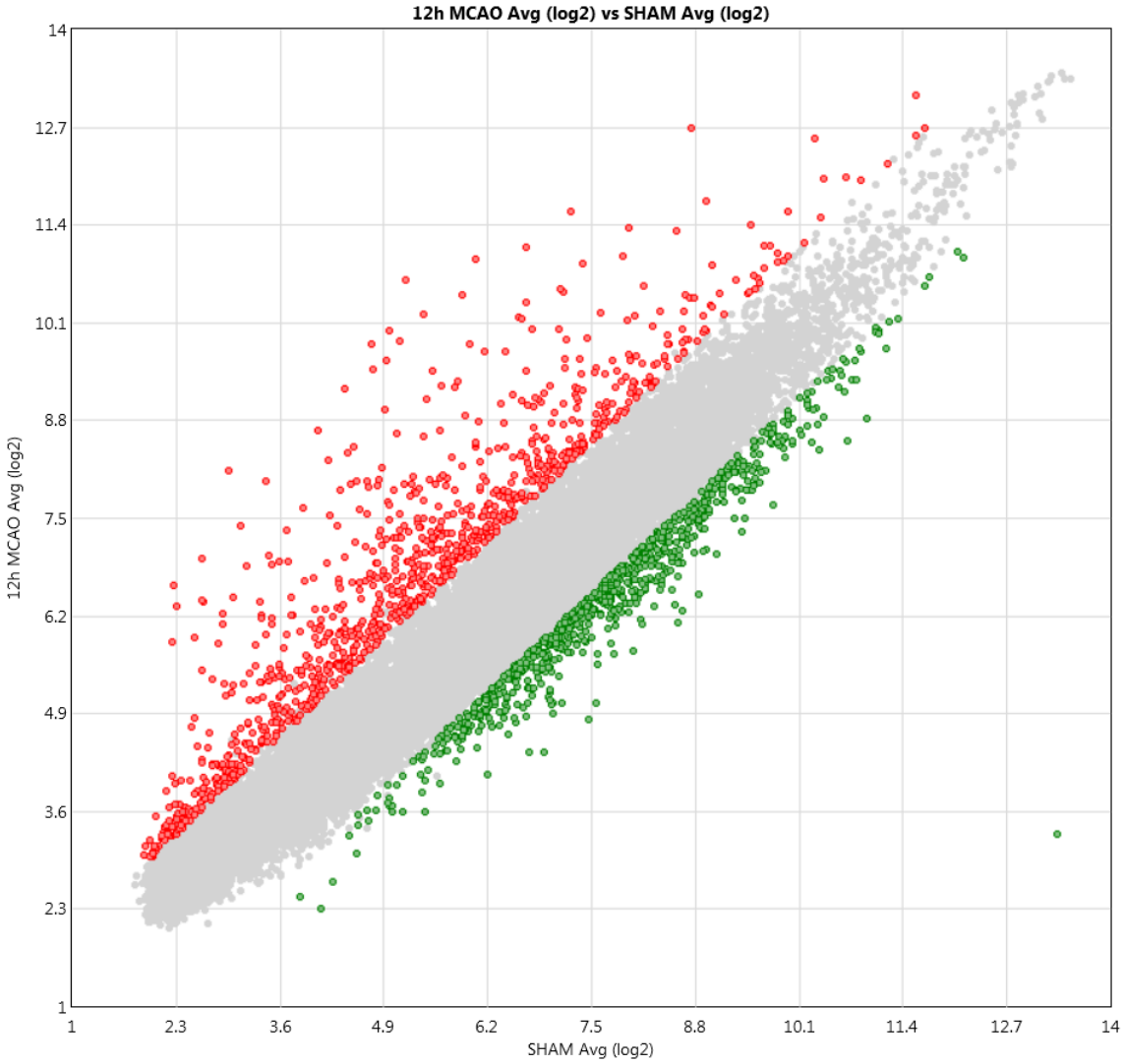
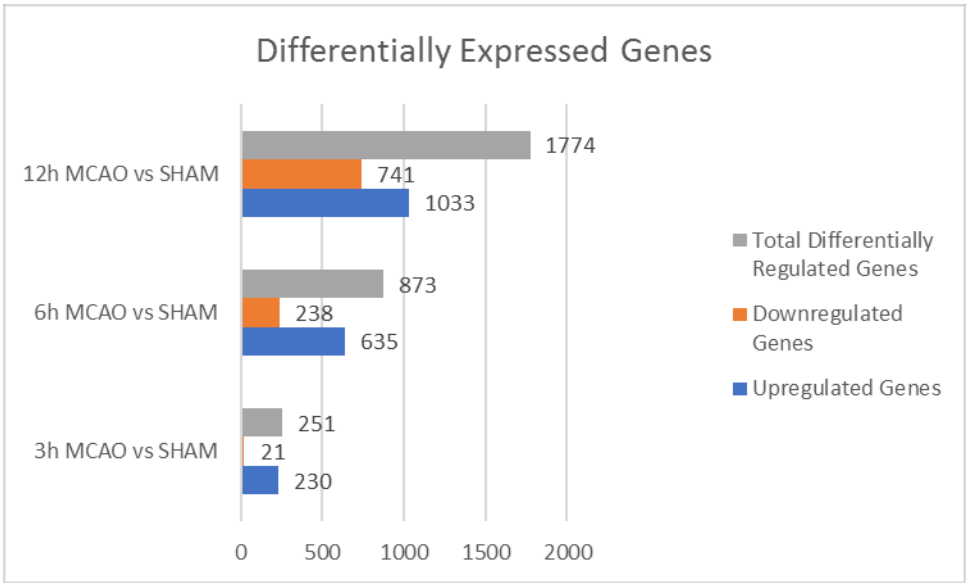


Figure 4. Scatter plot showing the average hybridization signal intensity of the genes in the 12h MCAO (n=4), Compared to the SHAM (n=4) group. The red dots indicate genes upregulated 2-fold or more. The green dots indicate genes downregulated 2-fold or more. And the grey dots indicate genes that were not differentially regulated with at least 2-fold.



Comparison	Upregulated	Downregulated
3h MCAO vs SHAM	230	21
6h MCAO vs SHAM	635	238
12h MCAO vs SHAM	1033	741

Figure 5. TAC expression data, based on 2-fold change cutoff. Each group compared to SHAM (animals given surgery but no stroke). A p-value cutoff of 0.05 was used. The groups are as follows: SHAM (n=4), 3h MCAO (n=3), 6h MCAO (n=4), 12h MCAO (n=4). Ebayes Anova Method was used.

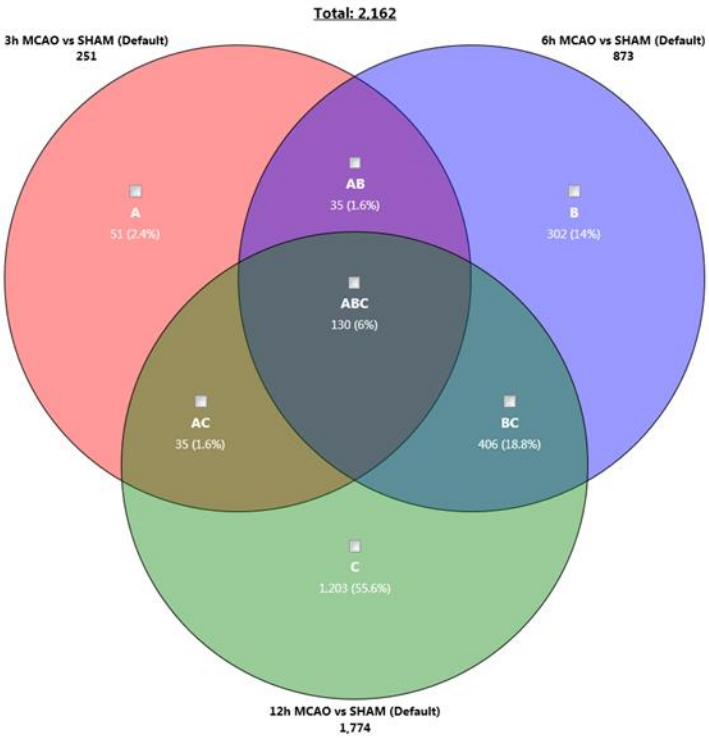


Figure 6. The number of differentially expressed genes in each category compared to SHAM. With the overlapping areas being genes that remain differentially expressed compared to SHAM between different experimental conditions. 2 fold cutoff with $p \leq 0.05$ was used. For example, the pink circle 251 genes that are differentially expressed between 3h MCAO and SHAM. The 51 genes in **A** are only differentially expressed at 3h MCAO, while the 35 genes in **AB** are differentially expressed compared to SHAM at both 3h and 6h. And the genes in **ABC** are differentially expressed compared to SHAM at 3h, 6h, and 12h.

Genesis Analysis

As an early sorting mechanism, k-means clustering with Genesis was used to separate differentially expressed genes into 10 different clustering patterns *Figure 7*. Clusters 1 and 5 contained genes that were continuously sharply upregulated over time and contained 327 and 50 genes respectively. Cluster 8 contained 856 genes that were continuously upregulated but relatively lower. Clusters 6 and 10 contained genes that were continuously downregulated over time and contained 343 and 1131 genes respectively. Cluster 2 contains 1357 genes that decrease at 3h and 6h and then remain relatively flat. Cluster 3 contains 1080 genes that remain close to baseline for 3h and 6h

and then increase at 12h. Cluster 4 contains 973 genes that remain close to baseline for 3h and 6h and then decrease at 12h. Cluster 7 contains 888 genes that increase for 3 and 6h then decrease slightly at 12h. Cluster 9 contains 598 genes that increase for 3h and 6h then return to baseline at 12h.

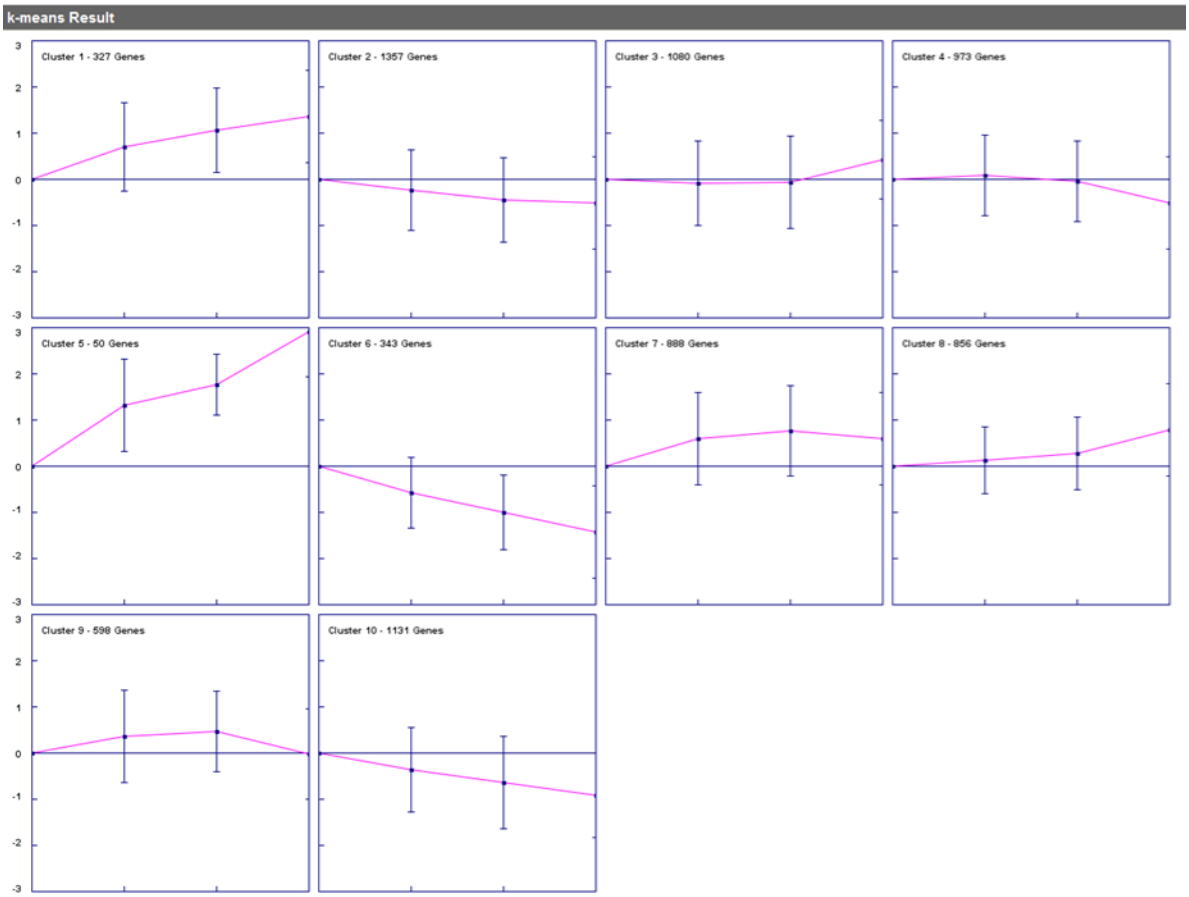


Figure 7. Genesis analysis results. The graphs shown are k-means clustering Stanford plots of the average normalized Log base 2 expression for each cluster. The error bars are measuring how well genes within that cluster fit with the average shape of that cluster. The top left of each box has the cluster name and number of genes in the cluster.

Enrichr Analysis of Gene Ontologies from Genesis Clusters

Enrichr was then used to determine Gene Ontologies associated with each of the 10 Genesis clusters *Figures 8-17*. Enrichr displayed the top ontologies associated with each cluster as well as four scores: p-value, q-value, z-score, and combined score. The continuously sharply upregulated Genesis clusters 1 and 5 both have molecular functions associated with cytokine/chemokine activity and immune function. Cluster 1 is also associated with RNA binding. The continuously slightly upregulated cluster 8 was associated with molecular functions related to with RNA binding and protein synthesis. On the other hand, the continuously downregulated clusters 6 and 10 were both associated with cellular ontologies dealing with neuronal cell components and membrane transport.

				Cluster 1 ontologies	
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	regulation of phosphorylation of RNA polymerase II C-terminal domain	0.00002234	0.00855	-2.89	30.97
GO Molecular Function 2018	C-C chemokine binding	0.00753	0.134	-3.33	16.29
	RNA binding	0.00001086	0.00184	-1.37	15.63
	cytokine activity	0.000004282	0.00145	-1.23	15.23
GO Cellular Component 2018	extrinsic component of endosome membrane	0.005915	0.1764	-2.75	14.13
	N-terminal protein acetyltransferase complex	0.01128	0.1764	-2.85	12.76
MGI Mammalian Phenotype 2017	abnormal tumor necrosis factor secretion	0.00006463	0.02697	-2.65	25.55
	increased fibrosarcoma incidence	0.0005043	0.06427	-3.08	23.4
	impaired neutrophil recruitment	0.00002708	0.02697	-3.12	21.3
Human Phenotype Ontology	Cerebral hemmorhage	0.01568	0.6716	-2.12	8.81
	Inappropriate behavior	0.02629	0.6716	-2.27	8.27
Jensen Tissues	Immune system	4.014 e-12	2.352 e-9	-6.92	181.53
	Parenchyma	1.269 e-10	2.478 e-8	-7.66	174.54
	Vascular system	1.361 e-11	3.988 e-9	-6.28	157.02
Jensen Compartments	NF kappa B complex	4.492 e-12	5.199 e-9	-6.23	162.74
	Bcl-2 family protein complex	9.745 e-9	5.652 e-6	-8.61	158.8
Jensen Diseases	Retinal ischemia	0.00003268	0.00447	-3.84	39.65

Figure 8. Enrichr Gene Ontology results for Genesis Cluster 1. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 2 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	regulation of cardiac muscle cell membrane	0.001164	0.2595	-3.09	20.91
	vesicle mediated transport	4.379 e-6	0.01562	-1.67	20.59
	neurotransmitter receptor transport	0.0004571	0.1875	-2.6	20.01
GO Molecular Function 2018	actin binding	0.00004944	0.0381	-2.44	24.15
	calcium transporting ATPase activity	0.0002765	0.05275	-2.65	21.69
	spectrin binding	0.000224	0.05275	-2.4	20.17
GO Cellular Component 2018	lysosome	2.037 e-7	6.37E-05	-1.17	18.05
	axon initial segment	0.003129	0.1818	-2.98	17.16
	lysosome membrane	6.733 e-6	0.0007	-1.37	16.29
	lytic vacuole membrane	3.007 e-6	0.000469	-1.22	15.51
MGI Mammalian Phenotype 2017	abmornal foramen magnum morphology	0.0006148	0.07863	-3.02	22.34
Human Phenotype Ontology	malar flattening	2.988 e-6	0.002062	-2.41	30.6
	single transverse crease	1.028 e-6	0.001419	-2.16	29.72
Jensen Tissues	corpus callosum	7.025 e-46	7.601 e-43	-9.55	992.39
	temporal lobe	2.968 e-45	1.606 e-42	-8.29	849.61
Jensen Compartments	cellular component	6.377 e-8	1.86E-05	-11.38	188.5
	neuron part	1.794 e-11	3.137 e-8	-7.14	176.69

Figure 9. Enrichr Gene Ontology results for Genesis Cluster 2. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 3 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	ribososmall subunit assembly	0.003308	1	-2.3	13.13
	basement membrane organization	0.009827	1	-2.84	13.12
GO Molecular Function 2018					
GO Cellular Component 2018					
MGI Mammalian Phenotype 2017	dilated esophagus	0.008138	1	-4.86	23.39
	absent skeletal muscle	0.008138	1	-4.58	22.04
Human Phenotype Ontology	erythroderma	0.01076	1	-1.92	8.69
	aplasia/hypoplasia of the middle phalanges of the hand	0.02241	1	-2.24	8.5
Jensen Tissues	synovial tissue	0.004743	1	-3.85	20.62
Jensen Compartments	keratohyalin granule	0.01109	1	-2.45	11.03
	cytosolic small ribosomal subunit	0.01599	1	-2.35	9.72
Jensen Diseases	conjuctivochalasis	0.02526	0.9947	-4.05	14.9

Figure 10. Enrichr Gene Ontology results for Genesis Cluster 3. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 4 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	axonogenesis	3.297 e-6	0.003605	-1.71	21.64
GO Molecular Function 2018	DNA insertion or deletion binding	0.003297	0.2253	-3.71	21.22
GO Cellular Component 2018	axolemma	0.0001617	0.04414	-2.92	25.46
MGI Mammalian Phenotype 2017	abnormal GABA-mediated receptor currents	1.508 e-5	0.01219	-2.82	31.32
Human Phenotype Ontology	Dialeptic seizures	3.594 e-5	0.03912	-2.1	21.53
	Focal seizures	6.967 e-5	0.03912	-2.15	20.54
Jensen Tissues	Hypothalamus	1.833 e-26	1.575 e-23	-7.41	439.06
Jensen Compartments	Neuron part	9.126 e-12	3.643 e-9	-6.98	177.3
	Synapse	1.591 e-15	2.541 e-12	-5.19	176.68
Jensen Diseases	kidney cancer	1.510 e-15	1.208 e-12	-0.71	24.23

Figure 11. Enrichr Gene Ontology results for Genesis Cluster 4. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 5 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	Positive regulation of calcium ion import	2.112 e-6	0.000267	-2.84	37.09
	regulation of natural killer cell chemotaxis	7.204 e-7	0.000213	-2.53	35.81
GO Molecular Function 2018	CCR chemokine receptor binding	0.0001116	0.005077	-1.91	17.39
	cytokine activity	1.195 e-6	0.000174	-1.23	16.22
GO Cellular Component 2018	ISWI-type complex	0.02843	0.13	-2.7	9.6
	specific granule lumen	0.01004	0.07435	-2.02	9.28
MGI Mammalian Phenotype 2017	abnormal wound healing	4.636 e-6	0.002469	-2.36	29.03
	ectopic cartilage	0.007183	0.05268	-5.61	27.68
Human Phenotype Ontology	growth hormone excess	0.03076	0.1653	-2.09	7.27
Jensen Tissues	Immune system	1.220 e-11	1.967 e-9	-6.92	173.84
Jensen Compartments	NF-kappaB complex	3.448 e-14	1.113 e-11	-6.23	193.05
	Bcl-2 family protein complex	1.646 e-9	2.371 e-7	-8.56	173.06
	VEGF-A complex	5.154 e-14	1.113 e-11	-5.61	171.69
Jensen Diseases	Uteral disease	1.641 e-9	9.436 e-8	-2.65	53.55

Figure 12. Enrichr Gene Ontology results for Genesis Cluster 5. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 6 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	sodium-independent organic anion transport	3.095 e-6	0.005066	-1.52	19.27
GO Molecular Function 2018	secondary active transmembrane transport	0.0001399	0.04528	-2.79	24.79
GO Cellular Component 2018	dendrite	0.001225	0.2119	-2.16	14.49
MGI Mammalian Phenotype 2017	abnormal compact bone lamella structure	0.0008548	0.1733	-4.12	29.11
	enhanced NMDA mediated synaptic currents	0.0008548	0.1733	-4.03	28.5
Human Phenotype Ontology	delayed gross motor development	0.0003907	0.2203	-2.08	16.32
Jensen Tissues	temporal lobe	1.515 e-16	7.455 e-14	-8.32	302.96
	corpus callosum	1.698 e-13	2.784 e-11	-9.48	278.82
Jensen Compartments	Intrinsic component of membrane	1.885 e-8	1.033 e-5	-11.38	202.39
	membrane part	4.144 e-10	4.542 e-7	-9.34	201.74
	integral component of membrane	9.524 e-8	2.610 e-5	-10.84	175.3
Jensen Diseases	Laurence-Moon syndrome	0.002785	0.2128	-3.74	21.98

Figure 13. Enrichr Gene Ontology results for Genesis Cluster 6. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 7 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	mRNA processing	7.388 e-11	2.144 e-7	-1.48	34.5
GO Molecular Function 2018	RNA binding	2.652 e-22	1.618 e-19	-1.37	68.19
GO Cellular Component 2018	nucleolus	1.619 e-8	4.938 e-6	-1.57	28.2
MGI Mammalian Phenotype 2017	increased thyroid tumor incidence	0.005619	0.6802	-4.88	25.29
Human Phenotype Ontology	Neurodegeneration	0.004415	0.8415	-2.2	11.93
	Intellectual disability mild	0.002294	0.8415	-1.9	11.57
Jensen Tissues	Parietal lobe	2.458 e-29	6.514 e-27	-8.62	568.01
	B-lymphocyte	1.308 e-28	2.080 e-26	-8.75	562.04
Jensen Compartments	Macromolecular complex	4.055 e-18	6.804 e-15	-9.61	385.01
Jensen Diseases	relapsing-remitting multiple sclerosis	5.777 e-5	0.04171	-2.53	24.66

Figure 14. Enrichr Gene Ontology results for Genesis Cluster 7. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 8 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	protein targeting to ER	1.353 e-12	1.081 e-9	-2.09	57.03
	SRP-dependent cotranslational protein targeting	1.779 e-13	4.265 e-10	-1.84	54.06
GO Molecular Function 2018	small ribosomal subunit rRNA binding	0.003457	0.5381	-4.3	24.34
GO Cellular Component 2018	cytosolic part	2.075 e-11	4.711 e-9	-1.64	40.24
	ribosome	9.125 e-8	6.905 e-6	-2.21	35.86
MGI Mammalian Phenotype 2017	cecum inflammation	2.440 e-6	0.004709	-3.35	43.35
Human Phenotype Ontology	coarctation of aorta	0.002733	0.9731	-2.2	12.96
Jensen Tissues	vascular system	2.715 e-9	1.996 e-6	-6.31	124.38
Jensen Compartments	cytosolic ribosome	4.500 e-10	6.386 e-7	-2.48	53.43
	cytosolic part	8.583 e-9	6.090 e-6	-2.69	49.92
Jensen Diseases	Diamond-Blackfan anemia	0.0002638	0.07648	-2.82	23.22
	epidymo-orchitis	0.001315	0.2261	-3.08	20.46

Figure 15. Enrichr Gene Ontology results for Genesis Cluster 8. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 9 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	positive regulation of mitochondrial membrane permiablity involved in apoptotic proces	0.001842	0.4285	-3.19	20.09
GO Molecular Function 2018	neurotransmitter receptor activity involved	0.0003695	0.05018	-2.05	16.18
GO Cellular Component 2018	pericentric heterochromatin	0.005747	0.235	-2.73	14.1
	nuclear heterochromatin	0.002935	0.235	-2.28	13.29
MGI Mammalian Phenotype 2017	visceral vascular congestion	0.0002395	0.1859	-3.37	28.06
Human Phenotype Ontology	Aplasia/Hypoplasia of the ulna	0.001193	0.4117	-2.3	15.46
Jensen Tissues	Hypothalamus	2.351 e-13	1.352 e-10	-7.41	215.44
Jensen Compartments	Bounding membrane of organelle	0.005942	0.3248	-6.51	33.38
Jensen Diseases	alpha thalassemia	0.000802	0.1961	-2.83	20.15

Figure 16. Enrichr Gene Ontology results for Genesis Cluster 9. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

	Cluster 10 ontologies				
Process	**Top function(s)**	**p-value**	**q-value**	**z-score**	**combined score**
GO Biological Process 2018	neuron projection morphogenesis	5.564 e-10	1.743 e-6	-1.59	33.8
GO Molecular Function 2018	exo-alpha-sialidase activity	0.0003039	0.0151	-4.93	39.92
	alpha-sialidase activity	0.0003039	0.0151	-4.88	39.51
GO Cellular Component 2018	dendrite	1.617 e-7	2.296 e-5	-2.14	33.48
MGI Mammalian Phenotype 2017	abnormal NMDA-mediated synaptic currents	2.710 e-5	0.007641	-2.69	28.29
Human Phenotype Ontology	Dementia	0.0003358	0.2452	-1.98	15.86
	Wide nasal bridge	0.001222	0.4273	-2.35	15.76
Jensen Tissues	Temporal lobe	1.300 e-50	1.168 e-47	-8.32	955.37
Jensen Compartments	Neuron part	2.712 e-24	2.245 e-21	-7.09	384.49
Jensen Diseases	Hemiplegia	0.004728	0.6777	-2.9	15.51

Figure 17. Enrichr Gene Ontology results for Genesis Cluster 10. Listing the GO database, the top ontologies in the cluster associated with each. And the statistical confidence of each ontology. The z-score deviation from expected rank and is computed using a modified Fischer's exact test. The combined score is the z-score multiplied by the natural log of the p-value.

Short Time-Series Expression Miner (STEM) Analysis

Short Time-Series Expression Miner (STEM) was selected to reduce possible statistical noise from the large number of genes for our short time study. If the same Genesis analysis was run multiple times, there would be minor fluctuation in the size of the profiles. However, Genesis does not account for random error that occurs by having a

large set of genes. In order to account for these discrepancies, STEM was used to analyze the TAC dataset. STEM was used to cluster the genes gathered from the probe set from TAC. Out of a possible 50 model profiles, it divided the differentially expressed genes into 9 significant clusters based on their expression over time *Figure 18*. Colored profiles were statistically significant. Profile 9 contained 419 genes that had a trend of continuous downregulation. Profile 11 had 104 genes with a trend of continuous downregulation. Profile 1 had 67 genes and a similar trend. Profile 40 contained 117 genes with a trend of continuous upregulation. Profile 42 contained 93 genes that were continuously upregulated. Profile 48 contained 103 genes that were continuously upregulated. Profile 49 contained 117 genes that were sharply upregulated up to 6h then slightly less upregulated at 12h. Profile 18 contained 65 genes that were downregulated at 3h and upregulated afterwards. Lastly profile 23 contained 28 genes that were downregulated at 6h but stayed relatively flat before and after. Of these, some clusters were grouped together into larger superclusters that behaved similarly and were displayed in the same color. The red supercluster (profiles 1, 9, and 11) contained genes that were continuously downregulated. Whereas the green supercluster (profiles 40, 42, 48) contained genes that were continuously upregulated.

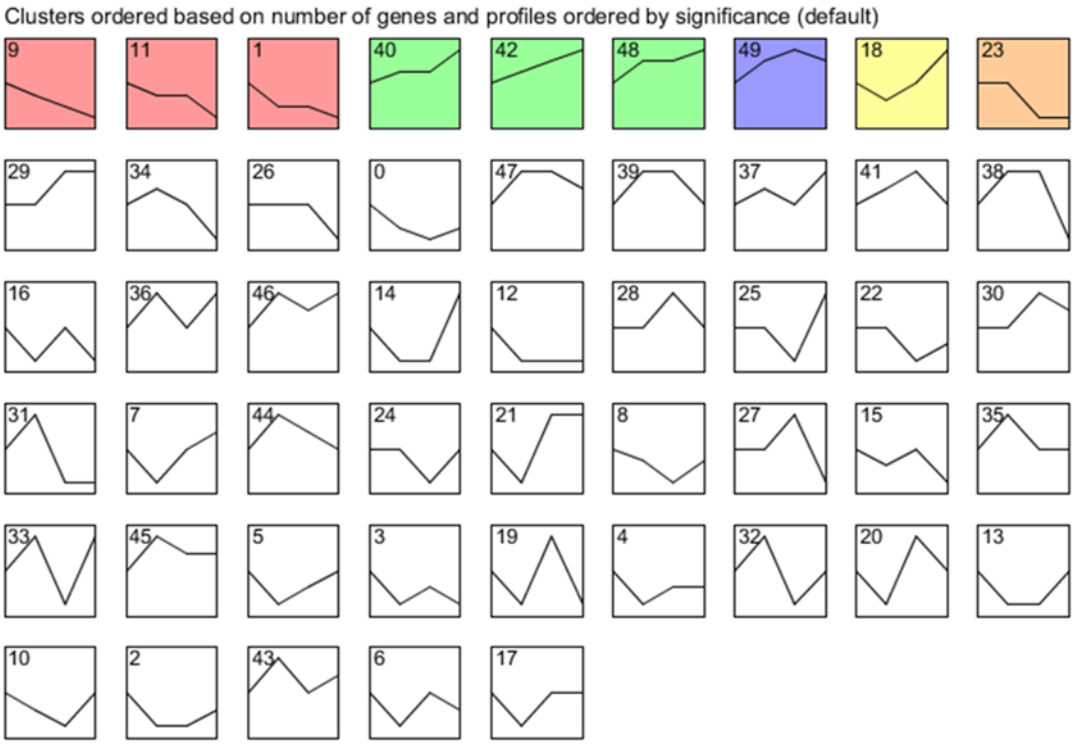


Figure 18. STEM profile summary. Each gene is assigned a profile. The colored profiles are statistically significant. And profiles of the same color belong to the same larger supercluster. The number on the top left of the box is the profile number. Within each box is a graph of normalized Log base 2 gene expression that goes from left to right SHAM, 3h, 6h, 12h.

Additionally, each cluster or supercluster was then associated with enriched Gene Ontologies *Figures 19-23*. The corrected p-value uses a randomization test where samples of the same size of the set being analyzed is drawn with a Bonferroni correction being used when the p-value enrichment is based on the expected size of the set of genes. 500 samples are used for randomized multiple hypothesis corrected enrichment p-values. Significant ontologies for STEM were considered to be ones with $p<0.05$ for the corrected p-value. The expected number of genes assigned is the average number of genes assigned over all permutations.

Among the 73 significant ontologies in the red supercluster which contained continuously downregulated genes: plasma membrane parts, ion transport, and ontologies related to neurons were all highly enriched *Figure 19*.

Red Supercluster Gene Ontologies			
Category Name	#Genes Assigned	#Genes Expected	Corrected p-value
plasma membrane part	144	83.9	<0.001
plasma membrane region	90	45.9	<0.001
cell-cell signaling	103	56.4	<0.001
ion transport	99	53.5	<0.001
synapse part	91	47.6	<0.001
anterograde trans-synaptic signaling	70	32.9	<0.001
chemical synaptic transmission	70	32.9	<0.001
dendrite	67	30.9	<0.001
dendritic tree	67	31	<0.001
trans-synaptic signaling	70	33.4	<0.001

Figure 19. Top 10 out of 73 significant Gene Ontologies associated with the red supercluster. Showing the number of genes in the profile or supercluster assigned to each ontology, the expected number of genes if based on random chance, and the corrected p-value for each assigned ontology.

Among the 262 ontologies in the green supercluster which contained continuously upregulated genes: inflammatory and apoptotic processes were highly enriched *Figure 20*.

Green Supercluster Gene Ontologies			
Category Name	#Genes Assigned	#Genes Expected	Corrected p-value
inflammatory response	42	10.8	<0.001
cytokine activity	20	2.3	<0.001
cell death	82	34.4	<0.001
response to biotic stimulus	60	20.7	<0.001
response to cytokine	53	16.9	<0.001
response to other organism	59	20.4	<0.001
response to external biotic stimulus	59	20.4	<0.001
defense response	59	20.6	<0.001
regulation of programmed cell death	68	26.1	<0.001
regulation of cell death	72	28.9	<0.001

Figure 20. Top 10 out of 262 significant Gene Ontologies associated with the green supercluster. Showing the number of genes in the profile or supercluster assigned to each ontology, the expected number of genes if based on random chance, and the corrected p-value for each assigned ontology.

Profile 18, which was downregulated at 3h and upregulated afterwards, contained genes associated with defense response *Figure 21*. This ontology overlapped with the ontologies of the green supercluster.

Profile 18 Gene Ontologies			
Category Name	#Genes Assigned	#Genes Expected	Corrected p-value
defense response	14	4.3	0.042

Figure 21. Gene Ontologies associated with the profile 18. Showing the number of genes in the profile or supercluster assigned to each ontology, the expected number of genes if based on random chance, and the corrected p-value for each assigned ontology.

Profile 49 was always upregulated but was less at 12h than 6h *Figure 22*. In this profile, RNA binding and nucleic acid binding were highly enriched.

Profile 49 Gene Ontologies			
Category Name	#Genes Assigned	#Genes Expected	Corrected p-value
RNA binding	21	6.5	<0.001
nucleic acid binding	36	17.4	0.014

Figure 22. Gene Ontologies associated with the profile 49. Showing the number of genes in the profile or supercluster assigned to each ontology, the expected number of genes if based on random chance, and the corrected p-value for each assigned ontology.

Lastly, profile 23 contained genes that remained at baseline, dropped at 6h, and then maintained a new lower baseline. Among these genes transport was enriched, especially organic hydroxy compound transport *Figure 23*.

Profile 23 Gene Ontologies			
Category Name	#Genes Assigned	#Genes Expected	Corrected p-value
organic hydroxy compound transport	5	0.4	0.012

Figure 23. Gene Ontologies associated with the profile 23. Showing the number of genes in the profile or supercluster assigned to each ontology, the expected number of genes if based on random chance, and the corrected p-value for each assigned ontology.

Inflammatory Response Ontologies Expression

Since profile 18 (that decreases at 3h and increases afterwards) and the continuously upregulated green supercluster both shared similar ontologies, we decided to look into the inflammatory response genes in each of those profiles *Figures 24-27*. The 14

inflammatory response genes associated with profile 18 included ITGAX, LSP1, SELE, PTPRC, TNIP2, CXCL6, LGALS3, PRKCD, S1PR3, CDH5, SLPI, SERPINA3N, SBNO2, and CCL20. CCL20 and SERPINA3N were the most highly expressed in this set *Figure 24*. The 24 inflammatory response genes associated with profile 40 include ENPP3, ITGAM, ANXA1, CCL2, CCL7, KDM6B, STAT3, SOCS3, SERPINE1, TLR1, CXCL2, CD14, ITGB2, IL1RN, LBP, CD44, IL1A, IL1B, A2M, IL6, PDPN, ICAM1, IL1R2, TIMP1. CCL2 and TIMP1 were the most highly expressed in this set *Figure 25*. The 10 inflammatory response genes associated with profile 42 include ZFP36, CCL3, FCGR2B, SPP1, EDNRB, PTPN2, HMOX1, S100A8, S100A9 and CASP4. CCL3 and SPP1 were the most highly expressed in this set *Figure 26*. The 8 inflammatory response genes associated with profile 48 include JAK2, CCL4, PTGS2, IER3, PTGES, JUN, and NAMPT. PTGS2 and CCL4 are the most highly expressed in this set *Figure 27*. Each of these profiles had a CCL among its top 2 highly expressed genes.

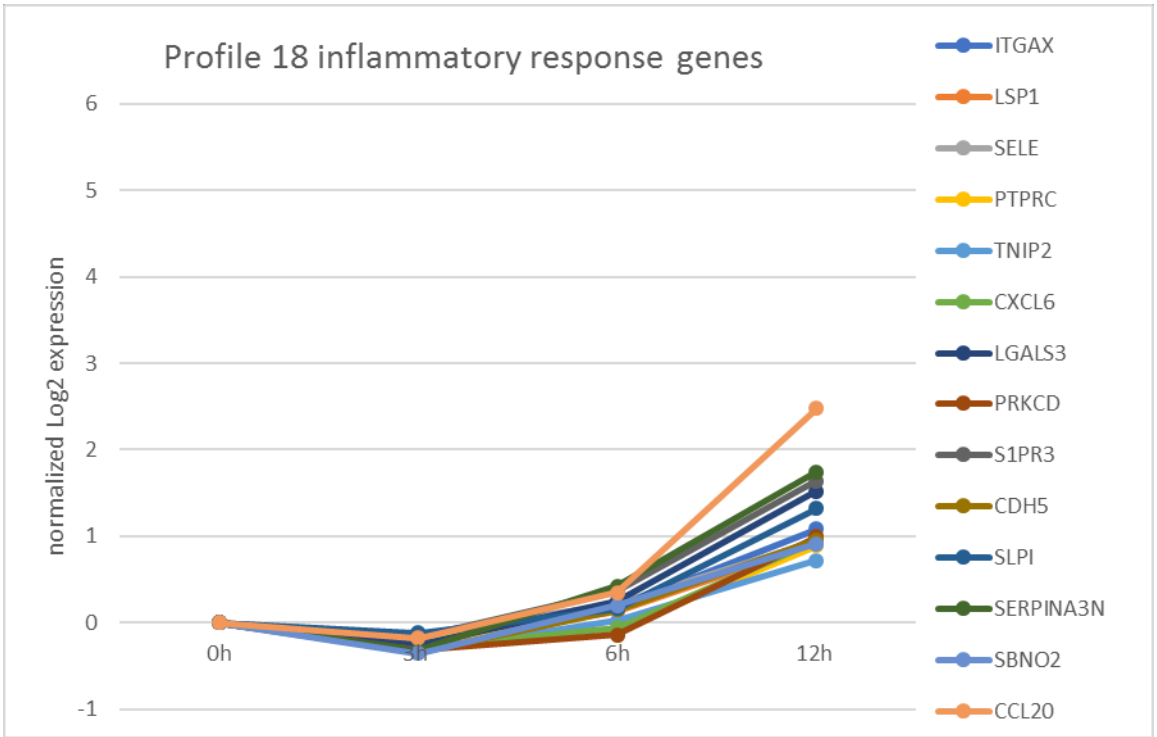


Figure 24. Normalized Log2 expression of genes in inflammatory response ontology for profile 18.

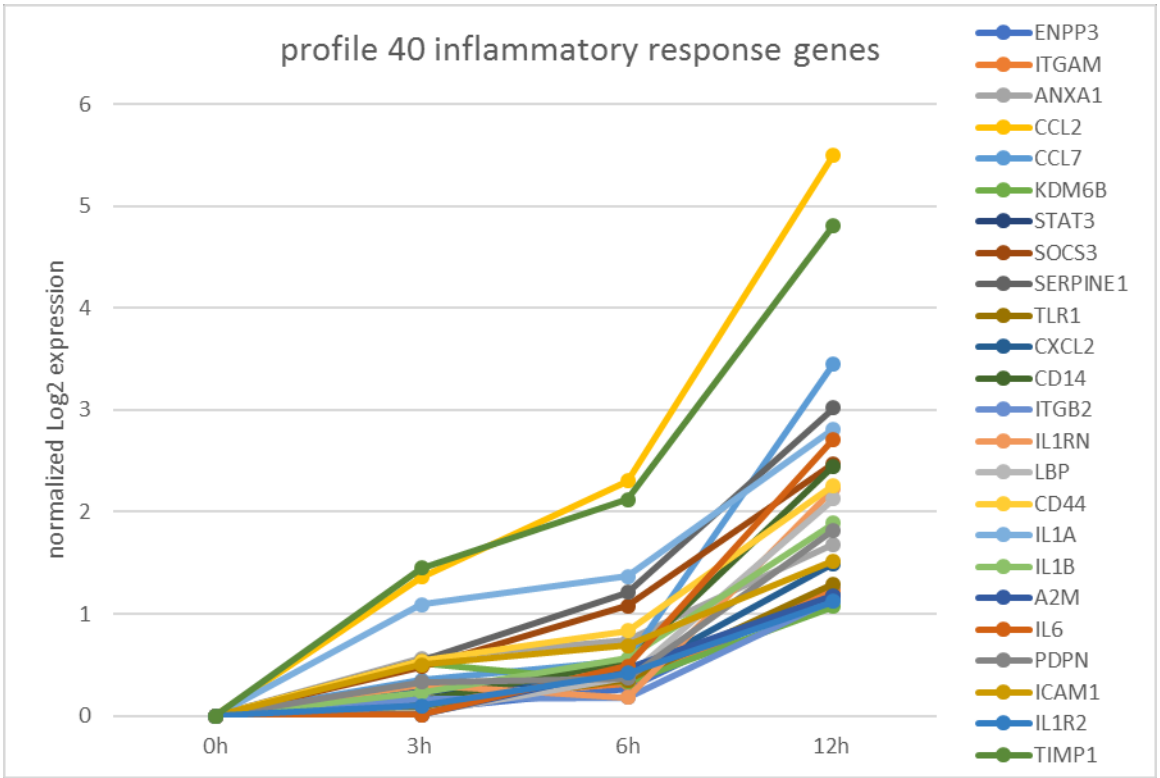


Figure 25. Normalized Log2 expression of genes in inflammatory response ontology for profile 40.

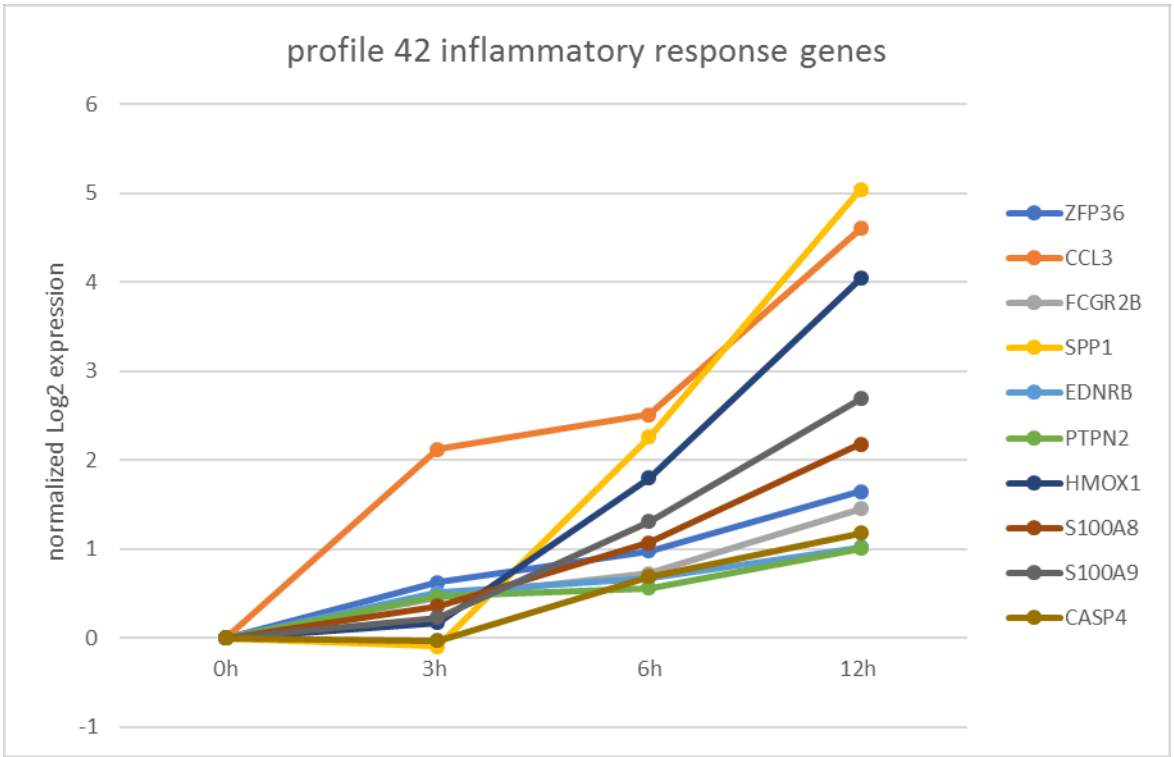


Figure 26. Normalized Log2 expression of genes in inflammatory response ontology for profile 42.

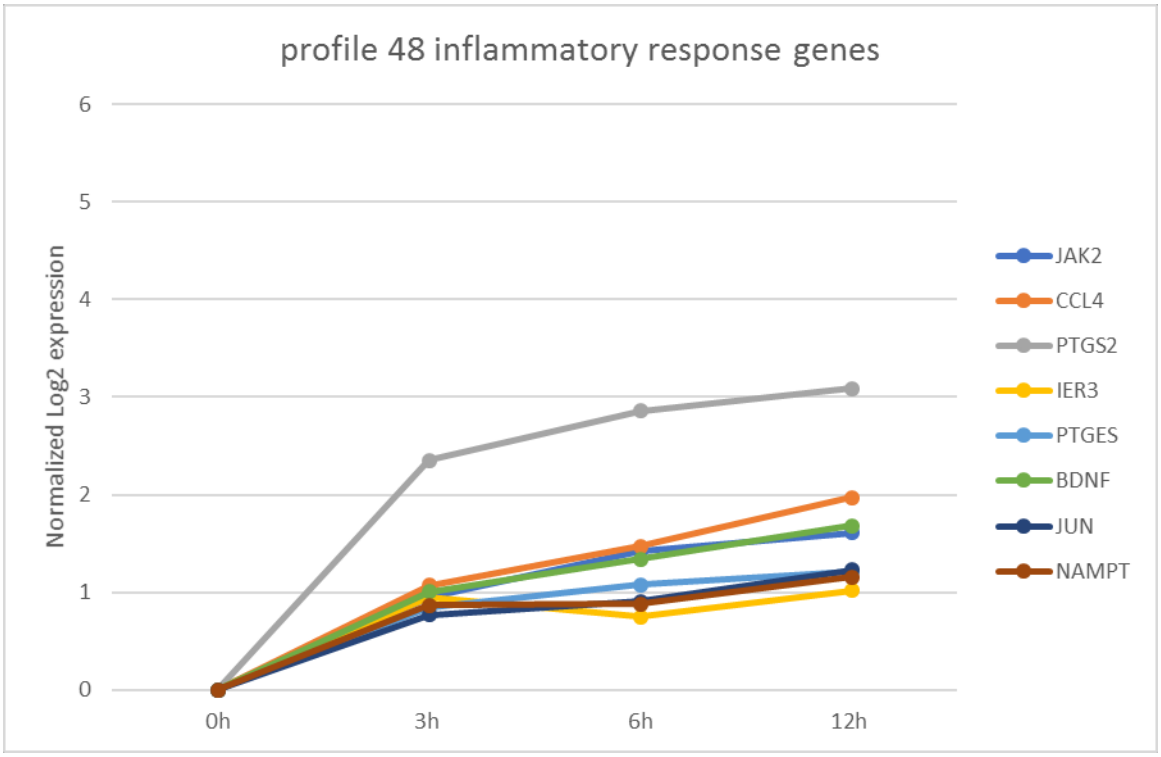


Figure 27. Normalized Log2 expression of genes in inflammatory response ontology for profile 48.

STRING Analysis

Following STEM analysis, the different STEM clusters were placed in String for further analysis. String was used to create protein networks associated with the given genes of interest. String was used to analyze each of the significant profiles and clusters *Figures 28-33*.

Due to their similar Ontologies, the combination of profile 18 and the green supercluster was the 1st to be examined with String *Figure 28a*. This group created a network of 350 genes with an average of 8.82 connections. This network contained two distinct hubs where the genes contained each had a significantly large number of connections compared to the genes outside these hubs. The larger hub mostly contained

cytokines and chemokines *Figure 28b*. Within the larger hub the most highly connected gene by far was IL6 with 79 connections *Table 1*. While the smaller hub contained proteins related to RNA and protein synthesis, including many ribosomal proteins and translation initiation factors *Figure 28c*. The most highly connected genes in the small hub were Rsl1d1 with 32 connections and Etf1 with 31 connections *Table 2*. Rsl1d1 is a ribosomal domain coding gene that acts as pro-apoptotic regulator in response to DNA damage. Etf1 is part of a complex that promotes nonsense mediated mRNA decay. The small hub also contained several Rps, ENSRNOG and Rpl genes which are ribosomal proteins and Eif genes which are translation initiation factors *Table 2*.

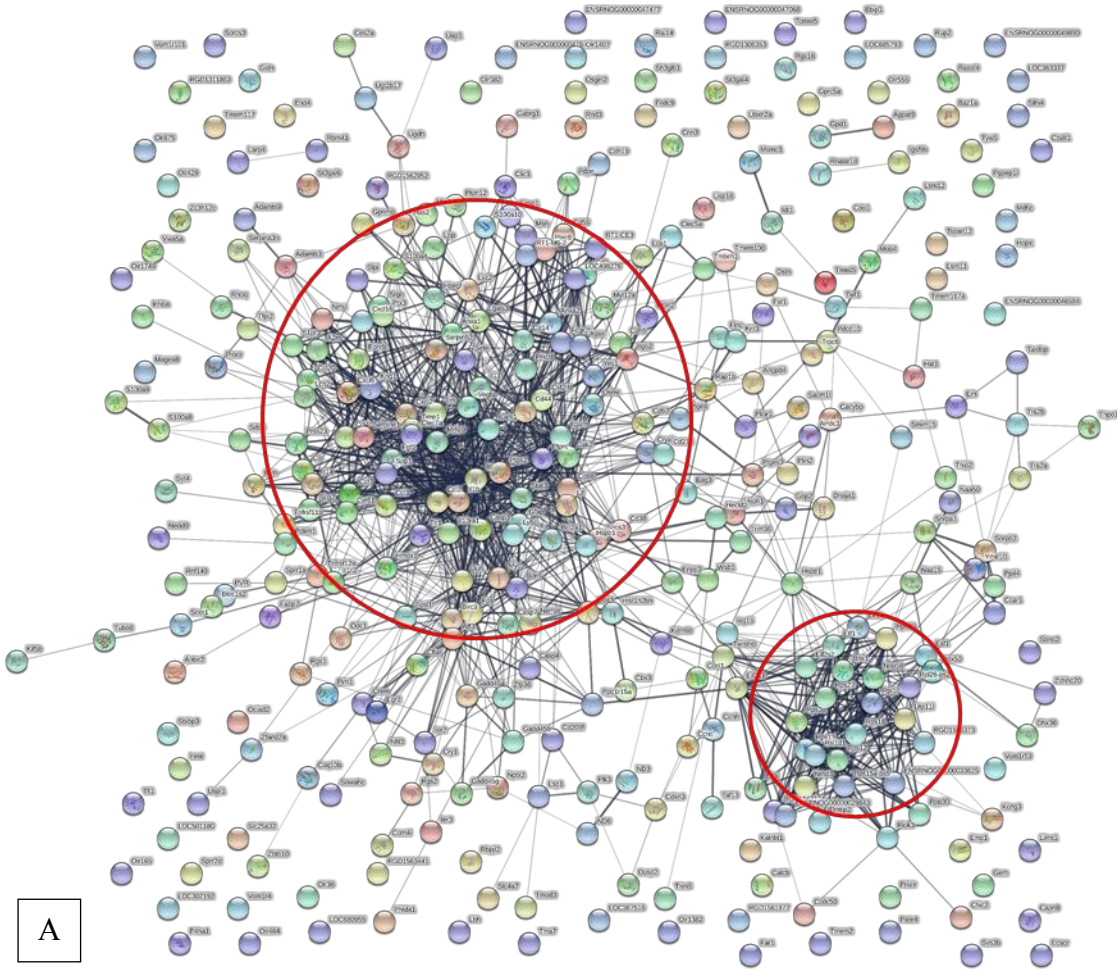
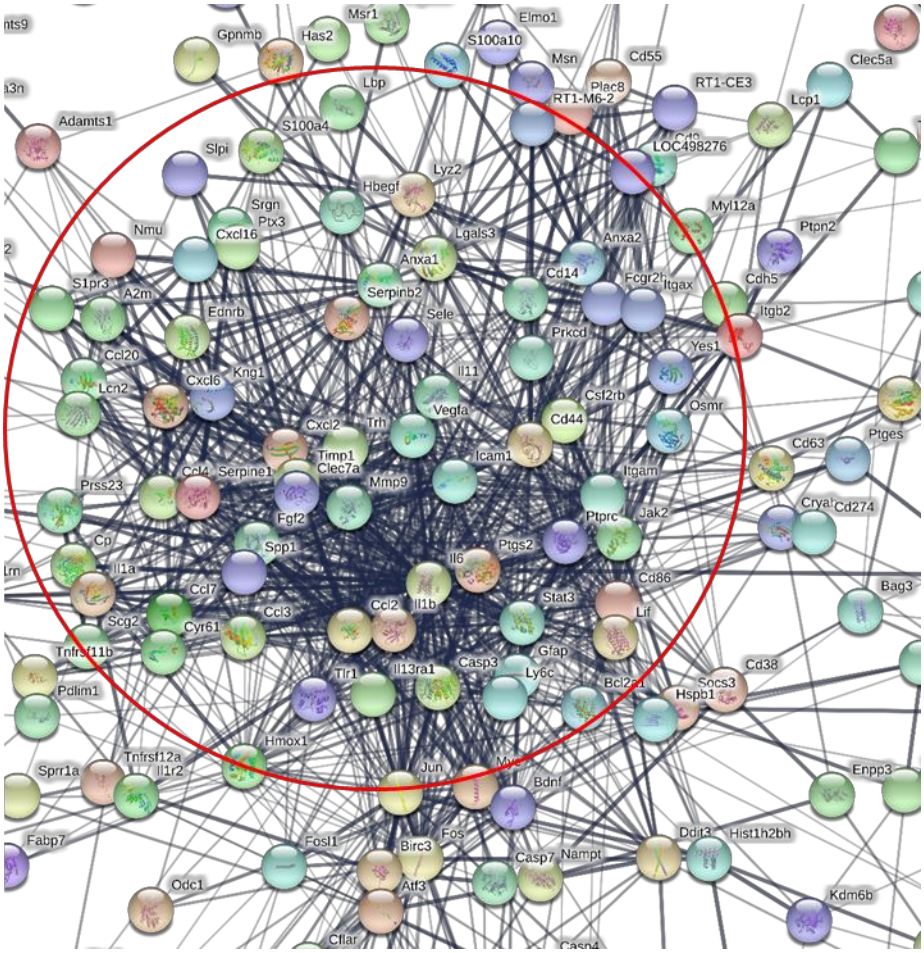
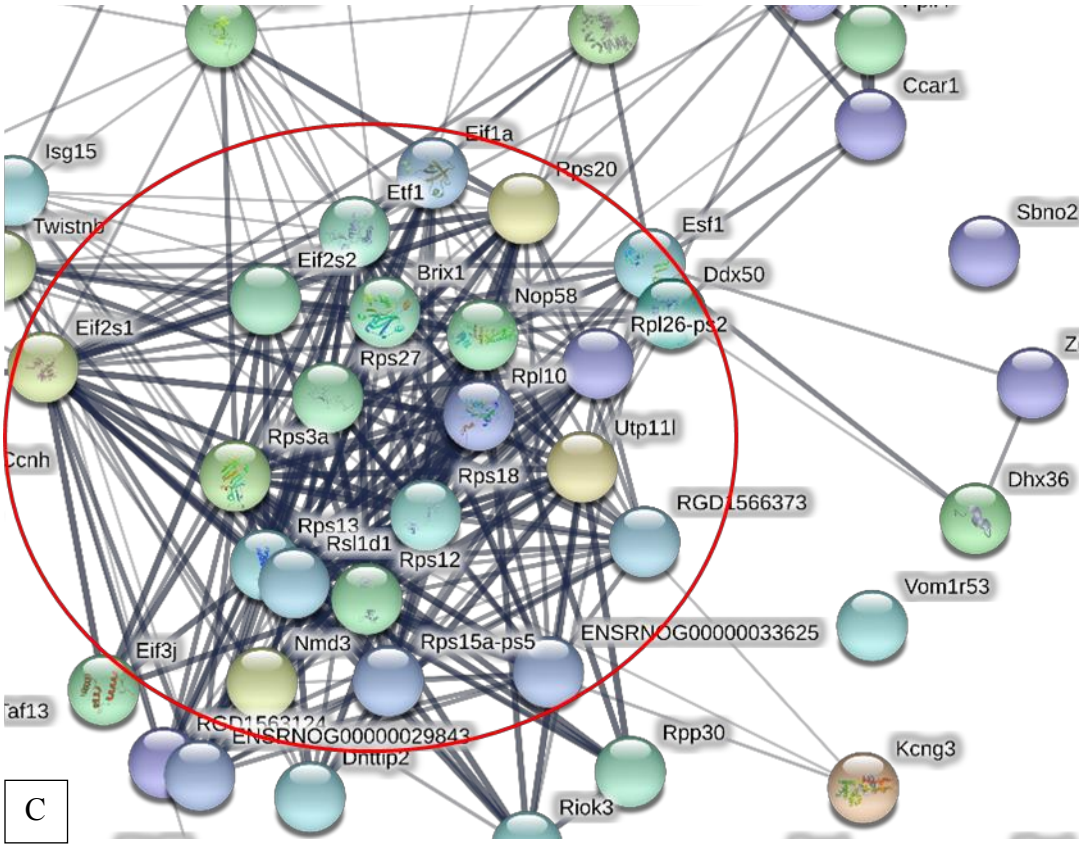


Figure 28. Figure 28a shows the gene network associated with the genes from the combined green supercluster and profile 18. The 2 red circles indicate hubs that contain a large number of highly connected genes. Figure 28b is a zoomed in version of the larger hub, while Figure 28c is a zoomed in version of the smaller hub.


Msr1
Elmo1
Gpnmb
Has2
S100a10
Msn
Cd55
Clec5a
Lbp
Plac8
RT1-CE3
RT1-M6-2
Lcp1
Adamts1
S100a4
LOC498276
Slpi
Lyz2
Hbegf
Srgn
Ptx3
Myl12a
Ptpn2
Nmu
Cxcl16
Lgals3
Anxa2
Anxa1
Cd14
Fcgr2b
Itgax
Cdh5
S1pr3
A2m
Serpinb2
Sele
Itgb2
Ednrb
Prkcd
Yes1
Ccl20
Il11
Kng1
Cxcl6
Lcn2
Csf2rb
Osmr
Vegfa
Cd44
Cxcl2
Trh
Ptges
Cd63
Timp1
Clec7a
Icam1
Ccl4
Serpine1
Itgam
Prss23
Mmp9
Cryab
Cd274
Fgf2
Ptprc
Jak2
Cp
Spp1
Ptgs2
Il6
Il1a
Cd86
Bag3
Ccl7
Ccl3
Ccl2
Il1b
Stat3
Lif
Scg2
Cyr61
Tnfrsf11b
Gfap
Casp3
Cd38
Tlr1
Il13ra1
Ly6c
Bcl2a1
Socs3
Pdlim1
Hspb1
Hmox1
Tnfrsf12a
Il1r2
Sprr1a
Jun
Bdnf
Enpp3
Fabp7
Fosl1
Fos
Birc3
Ddit3
Hist1h2bh
Casp7
Nampt
Odc1
Atf3
Kdm6b
Cflar
Casp4

B


Ccar1
Isg15
Eif1a
Rps20
Etf1
Twistnb
Esf1
Sbno2
Eif2s2
Brix1
Ddx50
Nop58
Eif2s1
Rpl26-ps2
Rps27
Rpl10
Rps3a
Utp11l
Ccnh
Rps18
Dhx36
RGD1566373
Rps13
Rsl1d1
Rps12
Vom1r53
Nmd3
Rps15a-ps5
ENSRNOG00000033625
Eif3j
Taf13
ENSRNOG00000029843
Rpp30
Kcng3
Dnttip2
Riok3
C

Gene Name	Node degree		Gene Name	Node degree		Gene Name	Node degree
Il6	79		Lgals3	28		LOC498276	15
Vegfa	53		Bdnf	28		Birc3	15
Mmp9	53		Ccl3	27		Cxcl16	14
Il1b	53		Jak2	26		Bcl2a1	14
Stat3	53		Itgax	25		Fosl1	14
Ccl2	52		Cd14	25		Gadd45a	14
Ptprc	51		Gfap	24		Ptx3	13
Cd44	50		Sele	23		Cdh5	13
Icam1	49		Anxa1	23		IL1rn	13
Itgam	47		Kng1	23		Cd63	13
Timp1	46		Socs3	23		A2m	12
Ptgs2	44		Lcn2	22		Ccl20	12
Fos	43		Ccl7	22		Trh	12
Jun	42		Cp	19		Cd9	11
Cxcl2	41		Il1a	19		Il1r2	11
Casp3	41		Anxa2	18		Cd38	11
Fgf2	40		Itgb2	18		S100a4	10
Cd86	37		Fcgr2b	18		Nmu	10
Serpine1	35		Il11	18		Lyz2	10
Myc	35		Hbegf	17		Serpinb2	10
Cxcl6	32		Lif	17		Scg2	10
Hmox1	31		Hspb1	17		Csf2rb	10
Atf3	31		Cyr61	16		Tnfrsf11b	10
Spp1	30		Ly6c	16		Cryab	9
Ccl4	29						

Table 1. Chart of number of connections for genes in network. Network is from STRING analysis of network made up of combination of profile 18 and green supercluster from the larger hub (shown in *Figure 28b*). The genes listed all had more connections than the average for the network (above 8.82 connections). Only the number of connections was counted and not the relative confidence of each connection.

Gene Name	Node Degree
Rsl1d1	32
Etf1	31
Rps3a	29
Eif2s1	27
Rps13	27
Nop58	25
Rps12	25
Rps18	24
Rps15a-ps5	23
Rpl26-ps2	22
Rps27	22
Eif1a	21
Rpl10	21
Rps20	20
ENSRNOG00000033625	19
Isg15	18
RGD1566373	18
RGD1563124	18
ENSRNOG00000029843	18
Brix1	17
Utp11l	17
Eif2s2	16
Esf1	14
Twistnb	14
Nmd3	14
Dnttip2	12
Riok3	12
Ddx50	10
Eif3j	9

Table 2. Chart of number of connections for genes in network. Network is from STRING analysis of network made up of combination of profile 18 and green supercluster from the smaller hub (shown in *Figure 28c*). The genes listed all had more connections than the average for the network (above 8.82 connections). Only the number of connections was counted and not the relative confidence of each connection.

Since the larger hub was mostly associated with inflammation ontologies, String was also used to look at the subset of genes in the combination of profile 18 and the green supercluster from the inflammatory response ontology *Figure 29*. This group had 56 genes with an average of 12.6 connections. Of these, 25 were highly connected with a node degree above the average for that network *Table 3*. Of the inflammatory response genes CCL2, IL1B, and IL6 all appear to be distinct hub genes with 33 or more connections each *Table 3*.

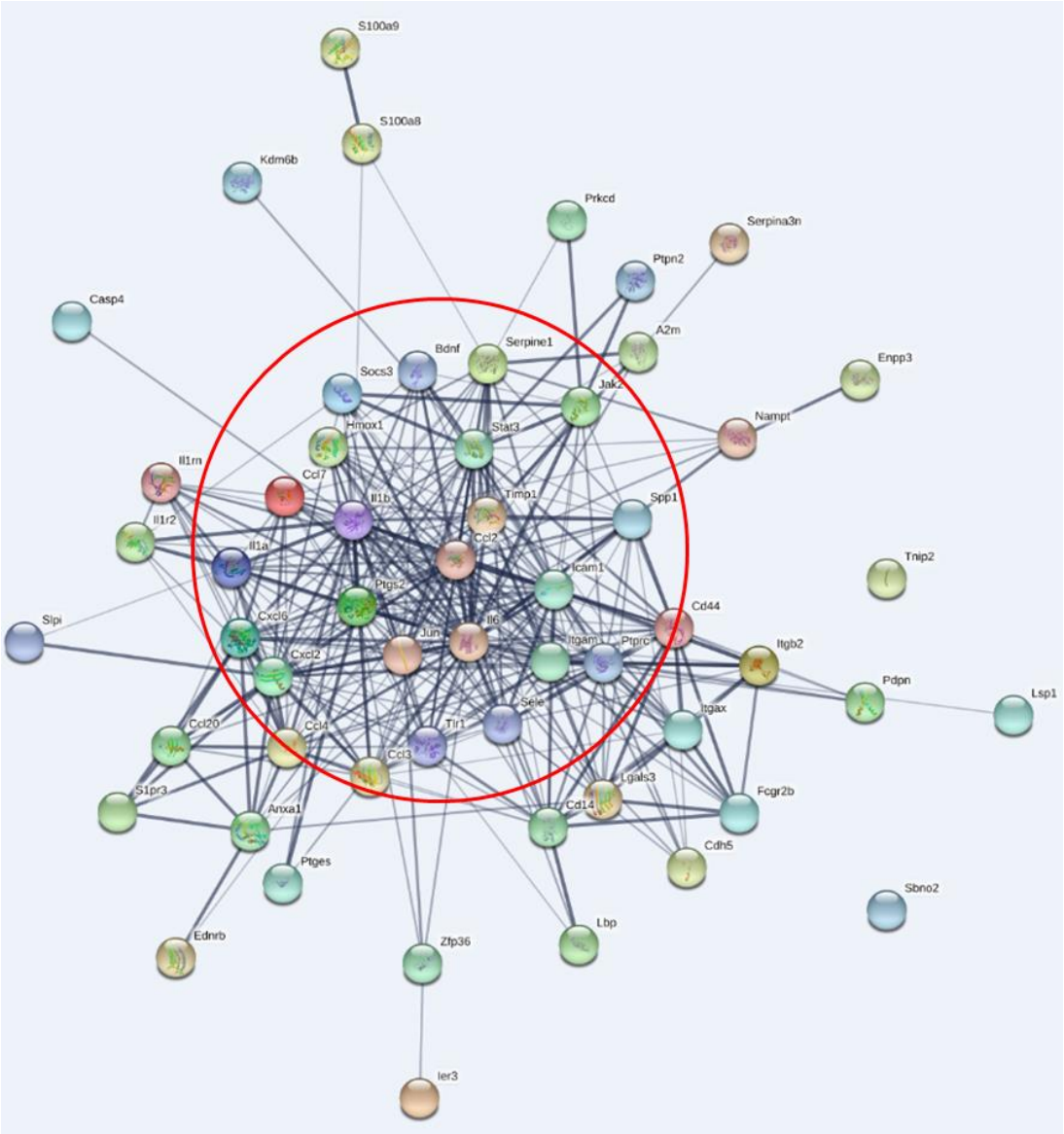


Figure 29. The gene network associated with the "Inflammatory response" ontology genes from the combined green supercluster and profile 18. The red circle indicates a hub of highly connected genes.

Gene Name	Node Degree
Il6	39
Il1b	34
Ccl2	33
Icam1	31
Ptprc	28
Ptgs2	27
Cxcl2	27
Stat3	26
Itgam	26
Timp1	23
Cxcl6	22
Tlr1	22
Cd44	21
Ccl4	21
Jun	20
Ccl3	19
Serpine1	17
Sele	17
Cd14	17
Hmox1	16
Ccl7	16
Spp1	15
Il1a	15
Jak2	14
Lgals3	13

Table 3. Chart of number of connections for genes in network. Network is from STRING analysis of inflammatory response gene ontology from combination of profile 18 and green supercluster (shown in *Figure 29*). The genes listed all had more connections than the average for the network (above 12.6 connections). Only the number of connections was counted and not the

The green supercluster was a network of 294 genes *Figure 30*. It showed a similar pattern to the network that combined it with profile 18. Similarly, to the combined network, it grouped into 2 distinct hubs, though this was slightly less pronounced. The

larger hub in the green supercluster network mostly containing cytokines and chemokines. While the smaller hub once again contained several Rps, ENSRNOG and Rpl genes which are ribosomal proteins and Eif genes which are translation initiation factors.

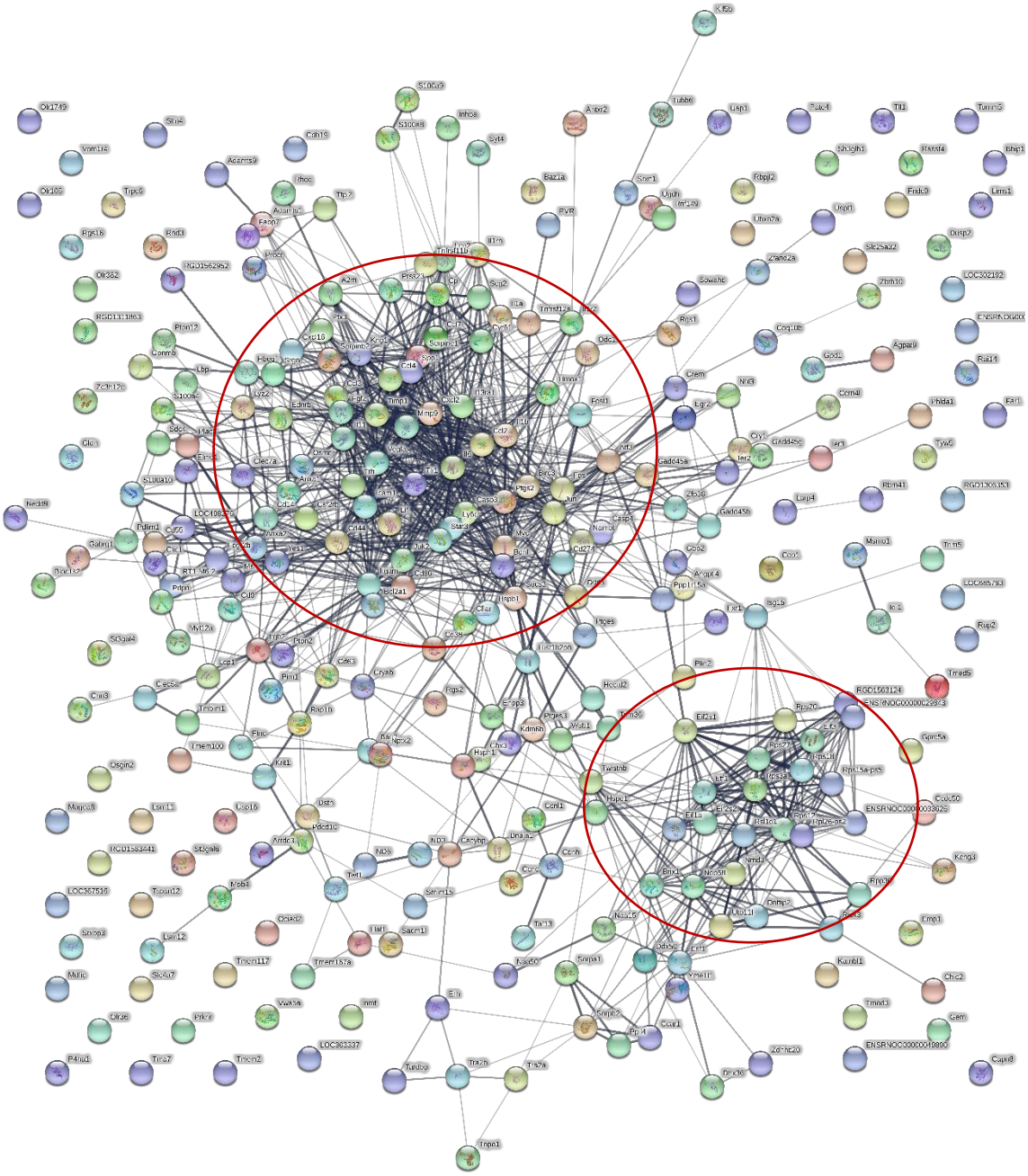


Figure 30. The gene network associated with the genes from the green supercluster. The 2 red circles indicate hubs that contain a large number of highly connected genes.

The red supercluster did not have much distinct grouping in its networks despite being the largest network of 579 genes *Figure 23*. This seems to indicate lack of a controlled connecting mechanism. Since various cellular processes are all generally downregulated and there appears to lack a controlling hub, this general decrease in expression may be due to general cell death.

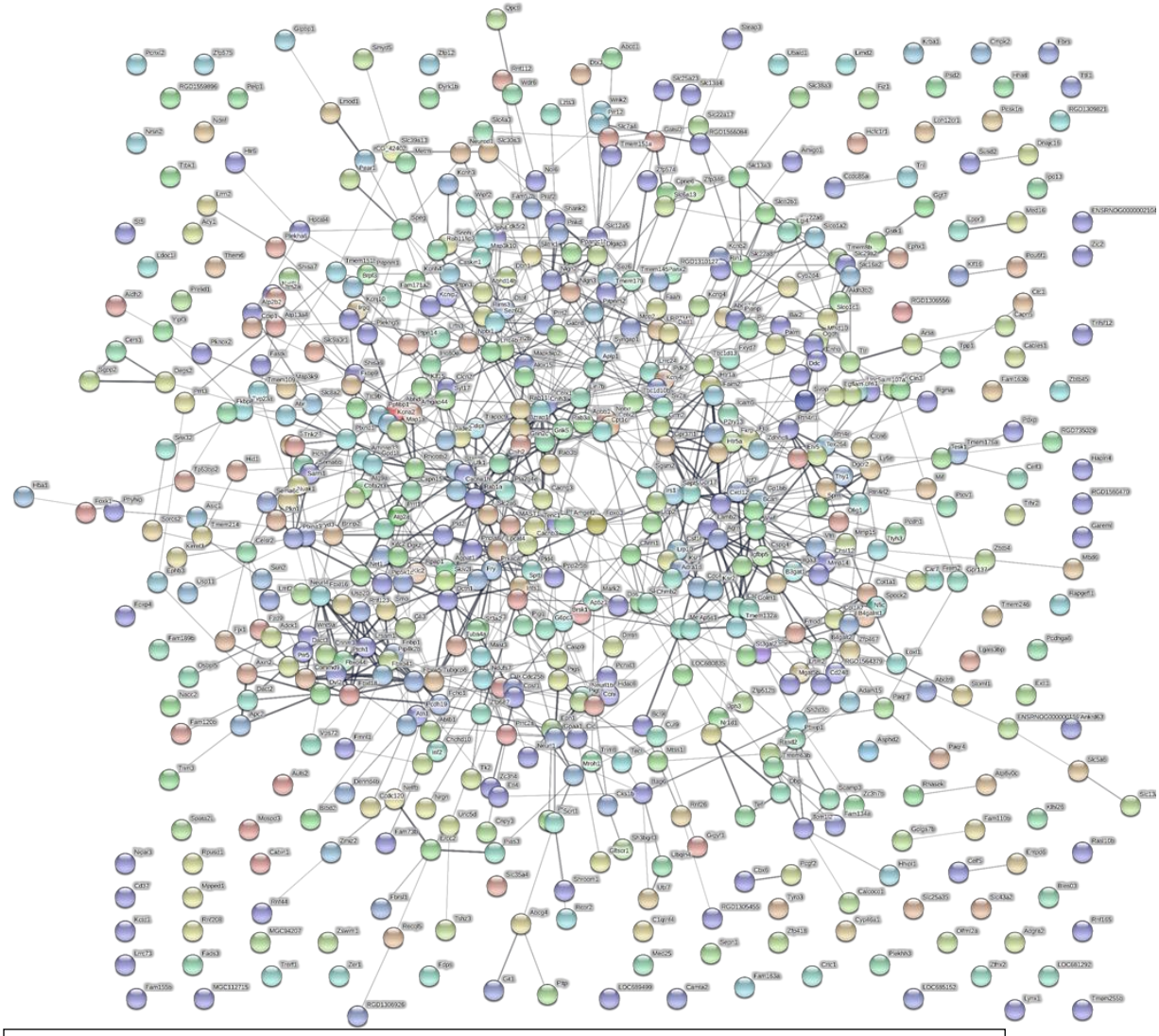


Figure 31. The gene network associated with the genes from the red supercluster. Despite the large number of genes that make up the supercluster, there does not appear to be a distinct hub.

The network associated with profile 49 was made up of 115 genes, roughly half of which were connected to 1 or more genes *Figure 32*. The proteins: Smek2, Thoc1, Fyttd1, Upf3b and Prfp39 had the most connections, each with 7 or more connections each *Table 4*.

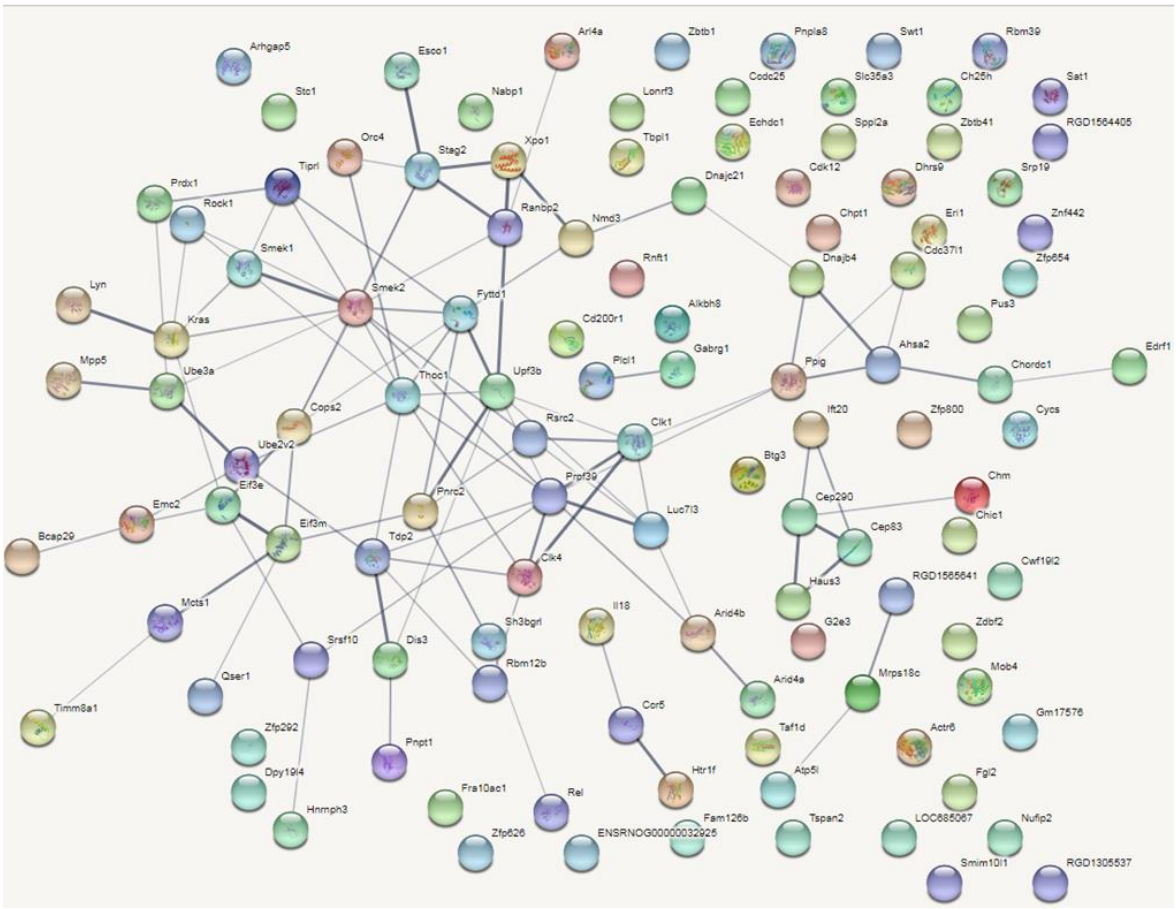


Figure 32. The gene network associated with the genes from profile 49. There appear to be a few distinct genes with significantly more connections than the rest (7 or more). These include Smek2, Thoc1, Fyttd1, Upf3b and Prpf39.

Gene Name	Node Degree
Smek2	12
Prpf39	10
Thoc1	9
Upf3b	8
Fyttd1	7

Table 4. Chart of number of connections for genes in network. Network is from STRING analysis profile 49 (shown in *Figure 31*). The genes listed all had more connections than the average for the network (above 1.79 connections). Only the number of connections was counted and not the relative confidence of each connection.

The network associated with profile 23 was a small one of roughly 25 genes *Figure 33*. Of these the acetylcholine receptor Chrm4 and the dopamine receptor Drd2 had the most connections (3 each).

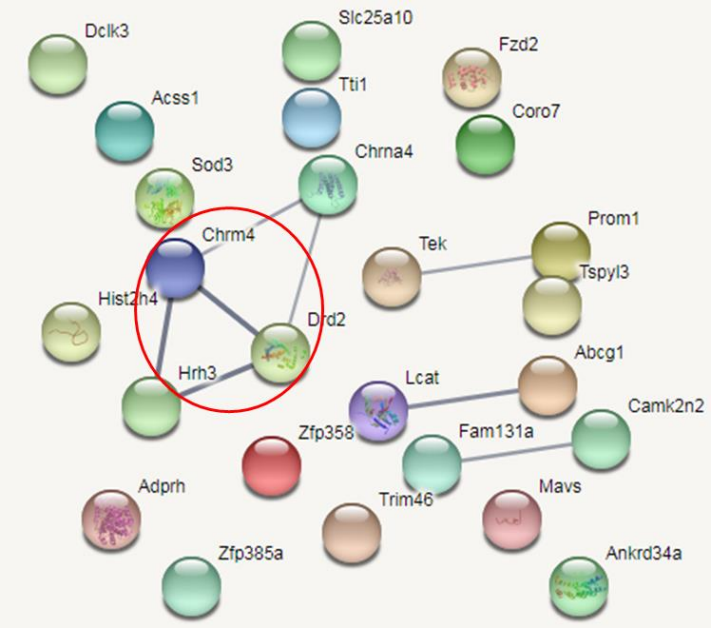


Figure 33. The gene network associated with the genes from profile 23. The red circle indicates the 2 genes (Chrm4 and Drd2) with the most connections in the network (3 each). These appear to be connected with 2 other genes (Hrh3 and Chrna4) that also seem to have more connections than average (2 each).

Discussion

From the TAC analysis the number of upregulated genes as well as the number of downregulated genes increase over time following stroke. Meaning that cell death alone does not account for all of the gene regulation following stroke, and that there must be numerous active regulatory responses as well.

Both Gene Ontology and String analysis of the green supercluster implicate inflammatory response and apoptosis as major pathways. This timescale fits well as what is currently understood about the immunology of stroke(6, 8-11) as much of the inflammation tends to increase during the 1st 6-12 hours following stroke. While at first glance, upregulation of both of these appears somewhat contrary to our previous EASE analysis paper which associated downregulation of both of these to pMCAO(20). In actuality, certain inflammatory factors, such as interleukin-1β have been shown to peak at 3–12 h, then decline over the next several hours and days (31, 32). Of the inflammatory response genes CCL2, IL1B, and IL6 all appear to be distinct hub genes with 33 or more connections each *Table 2*. IL1B was implicated previously in one of our tMCAO models(33). While IL6 was previously implicated in our pMCAO nonhuman primate model(34). Jun was previously implicated as well(19, 21).

The String analysis of the red supercluster seems to indicate lack of a controlled connecting mechanism as various cellular processes are all generally downregulated but without any central gene hubs. This combined with the Gene Ontology results from STEM imply that these processes may be decreasing due to general cell death. And since

many of these processes are related to neurons it hints at neuronal death specifically. This corresponds with early neuronal death following stroke(20, 30, 35-40).

Overall the network associated with profile 49 was not very highly connected. String analysis of profile 49 linked it to RNA transport and the nucleolus which fits well with the Gene Ontologies of RNA binding and nucleic acid binding associated to it in STEM. The most connected proteins in the String analysis for this profile were Smek2, Thoc1, Fyttd1, Upf3b and Prfp39 with 7 or more connections each. These are mostly connected with mRNA splicing and export. This implies that these processes are highly upregulated for the 1st 6h following stroke but resources are diverted from these processes afterwards. Smek2 is a protein phosphatase, Prpf39 is a pre-mRNA splicing factor, Thoc1 is associated with mRNA transcription and transport but also participates in the apoptotic pathway, Upf3b decays nonsense mRNAs and Fyttd1 is required for mRNA transport from the nucleus to the cytoplasm. Thoc1 was in cluster 9 of the supporting information of our previous study, but was not identified as a gene of interest(41). Smek2, Fyttd1, Upf3b and Prfp39 do not appear to be previously identified in relation to stroke.

The String analysis of profile 23 indicates that this small group of genes is not a uniformly connected network. But the two genes with the most connections Chrm4 and Drd2 are an acetylcholine and dopamine receptor respectively.

Though the Genesis clusters do not directly match to the STEM profiles. Both had sets of continuously upregulated and continuously downregulated genes. The ontologies associated with the continuously sharply upregulated Genesis clusters 1 and 5 were

associated with molecular ontologies for cytokines and chemokines much like the STEM green supercluster of continuously upregulated genes. Whereas the continuously slightly upregulated Genesis cluster 8 was associated with the molecular ontology of RNA binding much like the mostly upregulated STEM profile 49. Conversely, the continuously sharply downregulated Genesis clusters 6 and 10 had cellular ontologies associated with dendrites, which complements the STEM red supercluster of continuously downregulated genes that had ontologies associated with neuronal processes.

String analysis of the combination of profile 18 and the green supercluster produced 2 distinct hubs that were roughly replicated in the string analysis of the green supercluster. The smaller of these 2 distinct hubs was associated with ribosomal and translation initiation processes. Many members of this group were proteins that made up the ribosome including Rsl1d1, members of the Rps, RGD and ENSRNOG families, as well as several translation initiation factors such as members of the Eif family. The most highly connected gene Rsl1d1 acts as a pro-apoptotic regulator.

Likewise, profile 49 was made up of genes associated with different aspects of protein regulation, with one of the more highly connected genes Thoc1 associated with the apoptotic pathway as well.

Taken together these may imply that this upregulation of genes associated with protein expression may be involved in apoptosis following stroke. Several prior studies have found associations between regulation of protein synthesis and apoptosis(42-46). And potential regulatory mechanisms have previously been suggested(42). Thus, it is likely that in addition to inflammatory response, regulation of apoptosis through control

of protein synthesis could be a major controlling factor over the early 12-hour window. This apoptosis is likely upregulated in response to energy stress following stroke(47).

Conclusion

The results of this study demonstrate indicate distinct temporal patterns of gene regulation following stroke. Genes continuously upregulated for 12 hours following stroke fall into two distinct patterns. The 1st contains genes that control inflammation and apoptosis where CCL2, IL1B, and IL6 all appear to be important regulatory factors. The 2nd is associated with RNA regulation with the top 2 genes being Rsl1d1 a ribosomal protein associated with apoptosis and Etf1 which promotes nonsense mediated mRNA decay along with mostly RGD, Rps, and ESNROG ribosomal proteins and Eif translation initiation factors which had not previously been associated with stroke. Meanwhile, there is another set of genes involved in protein regulation that increase for the 1st 6 hours and decrease slightly at 12 hours but remain above baseline the entire time. These include Smek2, Prpf39, Thoc1, Upf3b, and Fyttd1. Understanding these regulatory patterns may be useful in designing treatments for stroke.

References

1. Stroke NIoNDa. The Ischemic Penumbra: NIH Specialized Programs of Translational Research in Acute Stroke Network; 2017. Available from: http://www.strokecenter.org/professionals/brain-anatomy/cellular-injury-during-ischemia/the-ischemic-penumbra/.
2. Association AH. About Stroke: American Stroke Association; 2019. Available from: https://www.strokeassociation.org/STROKEORG/AboutStroke/About-Stroke_UCM_308529_SubHomePage.jsp.
3. Services USDoHaH. Stroke Information | cdc.gov2017.
4. Association NS. What is stroke? 2014 [updated 2014-07-16]. Available from: http://www.stroke.org/understand-stroke/what-stroke.
5. Sarabi AS. Gene expression patterns in mouse cortical penumbra after focal ischemic brain injury and reperfusion. Journal of Neuroscience Research. 2008;86:2912-24.
6. Dirnagl U, Constantino I, Moskowitz MA. Pathobiology of ischaemic stroke: an integrated view. Trends in Neuroscience. 1999;22(9):391-7. doi: 10.1016/S0166-2236(99)01401-0.
7. Barone FC, Feuerstein GZ. Inflammatory Mediators and Stroke: New Opportunities for Novel Therapeutics:. Journal of Cerebral Blood Flow and Metabolism. 1999;19(8):819-34. doi: 10.1097_00004647-199908000-00001.
8. Arvin B, Neville LF, Barone FC, Feuerstein GZ. The Role of Inflammation and Cytokines in Brain Injury. Neuroscience & Biobehavioral Reviews. 1996;20(3):445-52. doi: 10.1016/0149-7634(95)00026-7.
9. Zoppo Gd, Ginis I, Hallenbeck JM, Iadecola C, Wang X, Feuerstine GZ. Inflammation and Stroke: Putative Role for Cytokines, Adhesion Molecules and iNOS in Brain Response to Ischemia. Brain Pathology. 2000;10(1):95-112.
10. Iadecola C, Anrather J. The immunology of stroke: from mechanisms to translation. Nat Med. 2011;17(7):796-808. doi: 10.1038/nm.2399. PubMed PMID: 21738161; PMCID: 3137275.
11. Iadecola C, Alexander M. Cerebral Ischemia and inflammation. Current Opinion in Neurology. 2001;14(1):89-94.
12. Zoppo Gd, Ginis I, Hallenbeck JM. Inflammation and Stroke: Putative Role for Cytokines, Adhesion Molecules and iNOS in Brain Response to Ischemia. Brain Pathology. 2000;10(1):95-112.
13. Stoll G, Jander S, Schroeter M. Inflammation and glial responses in ischemic brain lesions. Progress in Neurobiology. 1998;56(2):149–71. doi: 10.1016/S0301-0082(98)00034-3.
14. Kamel H, Iadecola C. Brain-Immune Interactions and Ischemic Stroke: Clinical Implications. Archives of Neurology. 2012;69(5):576-81. doi: 10.1001/archneurol.2011.3590.
15. Famakin BM. The Immune Response to Acute Focal Cerebral Ischemia and Associated Post-stroke Immunodepression: A Focused Review. Aging Dis2014. p. 307-26.
16. Schulze A, Downward J. Navigating gene expression using microarrays — a technology review. Nature Cell Biology. 2011;3(8):E190-E5. doi: doi:10.1038/35087138.

17. Soriano MA, Tessier M, Certa U, Gill R. Parallel Gene Expression Monitoring Using Oligonucleotide Probe Arrays of Multiple Transcripts With an Animal Model of Focal Ischemia. Journal of Cerebral Blood Flow. 2000;20:1045-55.
18. Kim Y-D, Sohn NW, Kang C, Soh Y. DNA array reveals altered gene expression in response to focal cerebral ischemia. Brain Research Bulletin. 2002;58(5):491-8. doi: 10.1016/S0361-9230(02)00823-7.
19. Lu A, Tang Y, Ran R, Clark JF, Aronow BJ, Sharp FR. Genomics of the Periinfarction Cortex after Focal Cerebral Ischemia. Journal of Cerebral Blood Flow and Metabolism. 2003;23(7):786-810. doi: 10.1097_01.WCB.0000062340.80057.06.
20. Ford G. Expression Analysis Systematic Explorer (EASE) analysis reveals differential gene expression in permanent and transient focal stroke rat models2006;1071(1):226–36. doi: 10.1016/j.brainres.2005.11.090.
21. Pulliam JVK, Xu ZF, Ford GD, Liu CM, Li YG, Stovall KC, Cannon VS, Tewolde T, Moreno CS, Ford BD. Computational identification of conserved transcription factor binding sites upstream of genes induced in rat brain by transient focal ischemic stroke. Brain Research. 2013;1495:76-85. doi: 10.1016/j.brainres.2012.11.052. PubMed PMID: WOS:000314624500008.
22. Tavazoie S, Hughes JD, Campbell MJ, Cho RJ, Church GM. Systematic determination of genetic network architecture. Nature Genetics. 1999:281-5.
23. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. PNAS. 1998;95(25):14863-8.
24. Tamayo P, Slonim D, Mesirov J, Zhu Q, Kitareewan S, Dmitrovsky E, Lander ES, Golub TR. Interpreting patterns of gene expression with self organizing maps: Methods and applications to hematopoietic differentiation. PNAS. 1998;96:2907-12.
25. Ernst J, Bar-Joseph Z. STEM: a tool for the analysis of short time series gene expression data. BMC Bioinformatics. 2006;7(1):191. doi: 10.1186/1471-2105-7-191.
26. Ernst J, Nau GJ, Bar-Joseph Z. Clustering short time series gene expression data. Bioinformatics. 2005;21(suppl_1). doi: 10.1093/bioinformatics/bti1022.
27. Chamankhah M, Eftekharpour E, Karimi-Abdolrezaee S, Boutros PC, San-Marina S, Fehlings MG. Genome-wide gene expression profiling of stress response in a spinal cord clip compression injury model. BMC Genomics. 2013;14(1):583. doi: doi:10.1186/1471-2164-14-583.
28. Li R, IV WEA, Summerfield TL, Yu L, Gulati P, Zhang J. Inflammatory Gene Regulatory Networks in Amnion Cells Following Cytokine Stimulation: Translational Systems Approach to Modeling Human Parturition. Plos One. 2011;6(6):1-19. doi: 10.1371/journal.pone.0020560.
29. Kim B-Y, Cao LH, Kim JY. Common Responses in Gene Expression by Ephedra herba in Brain and Heart of Mouse. Phytotherapy Research. 2011;25(10):1440-6.
30. Li Y, Xu Z, Ford GD, Croslan DR, Cairobe T, Li Z, Ford BD. Neuroprotection by Neuregulin-1 in a Rat Model of Permanent Focal Cerebral Ischemia. Brain Res. 2007;1184:277-83. doi: 10.1016/j.brainres.2007.09.037. PubMed PMID: 17961519; PMCID: 2374743.
31. Buttini M, Suater A, Boddeke HWGM. Induction of interleukin-1β mRNA after focal cerebral ischaemia in the rat. Molecular Brain Research. 1994;23(1-2):126-34. doi: 10.1016/0169-328X(94)90218-6.
32. X. W, F.C. B, N.V. A, G.Z. F. Interleukin-1 receptor and receptor antagonist gene expression after focal stroke in rats. Europe PMC. 1997;28(1):155-62. doi: https://doi.org/10.1161/01.STR.28.1.155.

33. Simmons LJ, Surles-Zeigler MC, Li Y, Ford GD, Newman GD, Ford BD. Regulation of inflammatory responses by neuregulin-1 in brain ischemia and microglial cells in vitro involves the NF-kappa B pathway. J Neuroinflammation2016.
34. Rodriguez-Mercado R, Ford GD, Xu Z, Kraiselburd EN, Martinez MI, Eterović VA, Colon E, Rodriguez IV, Portilla P, Ferchmin PA, Gierbolini L, Rodriguez-Carrasquillo M, Powell MD, Pulliam JV, McCraw CO, Gates A, Ford BD. Acute Neuronal Injury and Blood Genomic Profiles in a Nonhuman Primate Model for Ischemic Stroke. Comp Med2012. p. 427-38.
35. Ramos-Cejudo J, Gutierrez-Fernandez M, Rodriguez-Frutos B, Alcaide ME, Sanchez-Cabo F, Dopazo A, Diez-Tejedor E. Spatial and Temporal Gene Expression Differences in Core and Periinfarct Areas in Experimental Stroke: A Microarray Analysis. Plos One. 2012;7(12):1-15. doi: 10.1371/journal.pone.0052121.
36. Tokita Y, Keino H, Matsui F, Aono S, Ishiguro H, Higashiyama S, Oohira A. Regulation of Neuregulin Expression in the Injured Rat Brain and Cultured Astrocytes. Neuroscience. 2001;21(4):1257-64. doi: 10.1523/JNEUROSCI.21-04-01257.2001.
37. Croslan DR, Schoell MC, Ford GD, Pulliam JV, Gates A, Clement CM, Harris AE, Ford BD. Neuroprotective effects of Neuregulin-1 on B35 Neuronal Cells following Ischemia. Brain Res. 2008;1210:39-47. doi: 10.1016/j.brainres.2008.02.059. PubMed PMID: 18410912; PMCID: 2442468.
38. Lu XCM, Williams AJ, Yao C, Berti R, Hartings JA. Microarray analysis of acute and delayed gene expression profile in rats after focal ischemic brain injury and reperfusion. Journal of Neuroscience Research. 2004;77(6):843-57.
39. Lo EH, Dalkara T, Moskowitz MA. Mechanisms, challenges and opportunities in stroke. Nature Reviews Neuroscience. 2003;4(5):399-414. doi: doi:10.1038/nrn1106.
40. Danton GH, Dietrich WD. Inflammatory Mechanisms after Ischemia and Stroke. Journal of Neuropathology & Experimental Neurology. 2003;62(2):127-36. doi: 10.1093/jnen/62.2.127.
41. Surles-Zeigler MC, Li Y, Distel TJ, Omotayo H, Ge S, Ford BD. Transcriptomic analysis of neuregulin-1 regulated genes following ischemic stroke by computational identification of promoter binding sites: A role for the ETS-1 transcription factor. PLoS One2018.
42. Clemens MJ, Bushell M, Jeffrey IW, Pain VM, Morley SJ. Translation initiation factor modifications and the regulation of protein synthesis in apoptotic cells. Cell Death & Differentiation. 2000;7(7):603-15. doi: doi:10.1038/sj.cdd.4400695.
43. Tatton WG, Ju WYL, Holland DP, Tai C, Kwan M. (-)-Deprenyl Reduces PC12 Cell Apoptosis by Inducing New Protein Synthesis. Journal of Neurochemistry. 1994;63(4):1572-5.
44. Martin SJ, Lennon SV, Bonham AM, Cotter TG. Induction of apoptosis (programmed cell death) in human leukemic HL-60 cells by inhibition of RNA or protein synthesis. The Journal of Immunology. 1990;145(6):1859-67.
45. Plunovsky VA, Wendt CH, Ingbar DH, Peterson MS, Bitterman PB. Induction of Endothelial Cell Apoptosis by TNFα: Modulation by Inhibitors of Protein Synthesis. Experimental Cell Research. 1994;214(2):584-94. doi: 10.1006/excr.1994.1296.
46. Wyllie AH, Morris RG, Smith AL, Dunlop D. Chromatin Cleavage in Apoptosis: Association with Condensed Chromatin Morphology and Dependence on Macromolecular Synthesis. Journal of Pathology. 1984;142:67-77.
47. Shaw RJ, Kosmatka M, Bardeesy N, Hurley RL, Witters LA, DePinho RA, Cantley LC. The tumor suppressor LKB1 kinase directly activates AMP-activated kinase and regulates apoptosis in response to energy stress. PNAS. 2004;101(10):3329-35. doi: 10.1073/pnas.0308061100.