Neuron
NeuroResource

CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells

Joyce van de Leemput,[1,4] Nathan C. Boles,[1,4] Thomas R. Kiehl,[1,2] Barbara Corneo,[1] Patty Lederman,[1] Vilas Menon,[3] Changkyu Lee,[3] Refugio A. Martinez,[3] Boaz P. Levi,[3] Carol L. Thompson,[3] Shuyuan Yao,[3] Ajamete Kaykas,[3] Sally Temple,[1,5,*] and Christopher A. Fasano[1,5,*]
[1]Neural Stem Cell Institute, Rensselaer, NY 12144, USA
[2]Department of Computer Science, College of Computing and Information, State University of New York, Albany, NY 12144, USA
[3]Allen Institute for Brain Science, Seattle, WA 98103, USA
[4]Co-first Authors
[5]Co-senior Authors
*Correspondence: sallytemple@neuralsci.org (S.T.), chrisfasano@neuralsci.org (C.A.F.)
http://dx.doi.org/10.1016/j.neuron.2014.05.013

SUMMARY

Many neurological and psychiatric disorders affect the cerebral cortex, and a clearer understanding of the molecular processes underlying human corticogenesis will provide greater insight into such pathologies. To date, knowledge of gene expression changes accompanying corticogenesis is largely based on murine data. Here we present a searchable, comprehensive, temporal gene expression data set encompassing cerebral cortical development from human embryonic stem cells (hESCs). Using a modified differentiation protocol that yields neurons suggestive of prefrontal cortex, we identified sets of genes and long noncoding RNAs that significantly change during corticogenesis and those enriched for disease-associations. Numerous alternatively spliced genes with varying temporal patterns of expression are revealed, including *TGIF1*, involved in holoprosencephaly, and *MARK1*, involved in autism. We have created a database (http://cortecon.neuralsci.org/) that provides online, query-based access to changes in RNA expression and alternatively spliced transcripts during human cortical development.

INTRODUCTION

The cerebral cortex is responsible for processing sensory input, encoding memories, coordinating motor movements, thought, and planning complex behaviors. Accomplishing these sophisticated tasks involves billions of intricately connected neurons in highly ordered layers and columns (Jones and Rakic, 2010). Cortical development involves a precisely orchestrated program of progenitor cell expansion, neuron differentiation, neuronal migration, and circuit formation to create its highly organized cytoarchitecture.

Knowledge of global gene expression patterns during corticogenesis has been largely based on data obtained in mouse systems (Belgard et al., 2011; Dougherty and Geschwind, 2005). Recent studies using RNA sequencing (RNA-seq) have uncovered novel coding, noncoding, and alternatively spliced transcripts expressed specifically in each cortical layer and at different developmental stages, providing a more comprehensive view of murine cortical development (Belgard et al., 2011; Hubbard et al., 2013). There are notable differences between mouse and human cortices, the most obvious being a 1,000-fold difference in size, and humans have a significantly larger and more complex cortex. This elaboration is reflected in the elongated gestation: cortical neurogenesis is approximately 70 days in humans (Caviness et al., 1995) but only 7–8 days in mice (Takahashi et al., 1996). Differences between mouse and human cortical development are also evident at the molecular level, with distinctly different gene expression profiles in the cortical germinal zones of mouse and human (Fietz et al., 2012); 30% of the layer-specific cortical markers are differentially expressed between mouse and human (Zeng et al., 2012).

Most human cerebral cortex transcriptome studies have been carried out on postmortem tissue using microarray techniques (Hawrylycz et al., 2012; Kang et al., 2011). Given the differences in mRNA expression, splicing, and editing observed throughout cerebral cortical development (Dillman et al., 2013) and the limitations of microarray technology, deep sequencing technology is needed to fully characterize the transcriptomic changes during corticogenesis. Differentiation of human embryonic stem cells (hESCs) into cortical progeny has been demonstrated to be a viable model of human cortical development (Gaspard et al., 2008; Johnson et al., 2007; Kim et al., 2011; Shi et al., 2012). In this study, we used an adapted protocol to differentiate hESCs into cortical neurons over 77 days; performed RNA-seq during the process to generate a comprehensive transcriptome database encompassing human in vitro corticogenesis; and conducted analyses for KEGG pathways, gene ontology (GO) categories, disease associations, and alternative splicing. We provide the database and analytical results as a resource released in raw data format as well as a

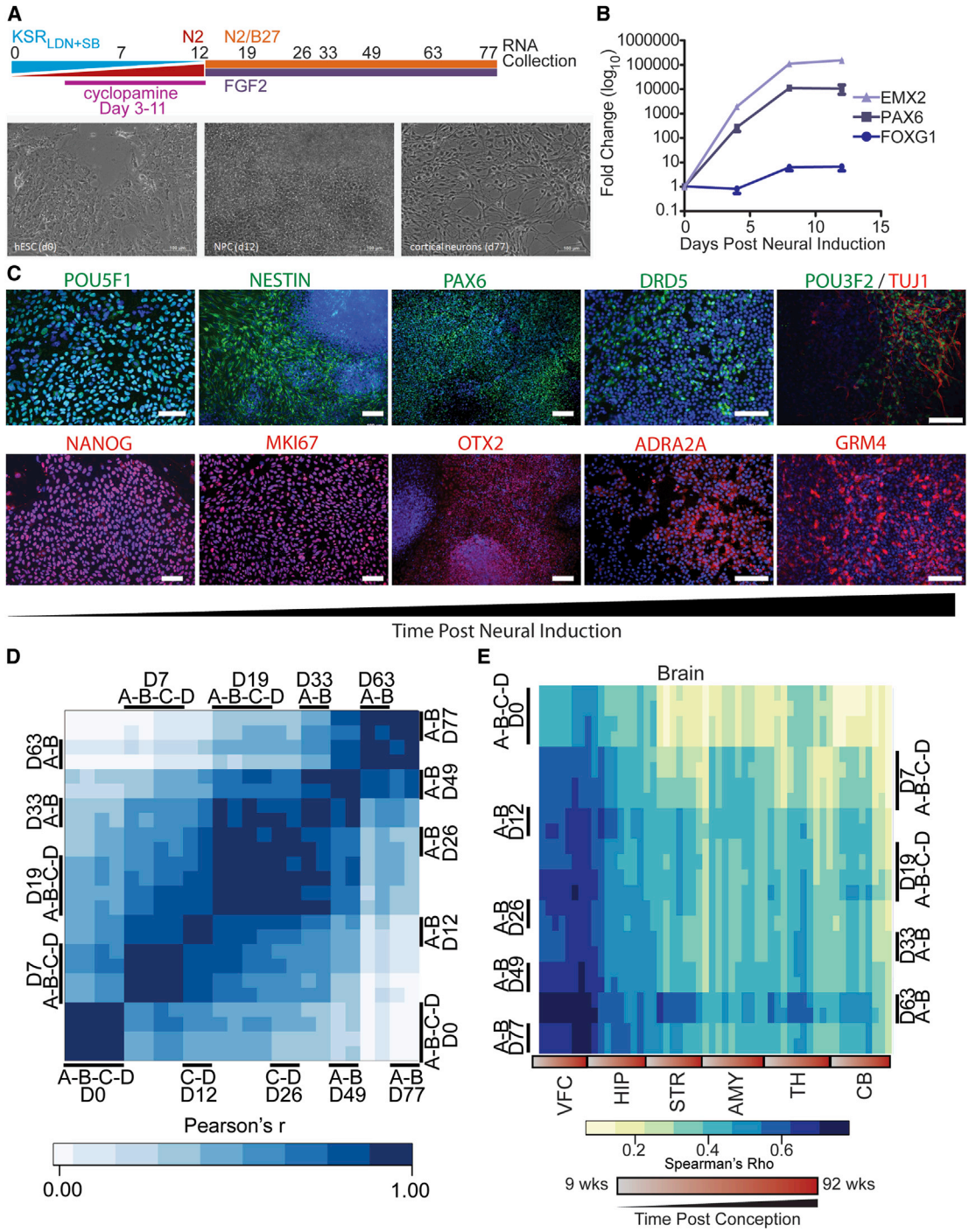


Figure 1. Establishing a Workflow for In Vitro Corticogenesis

(A) Schematic of cortical differentiation protocol, based on dual SMAD-inhibition (LDN = LDN193189; SB = SB431542; KSR = "Knockout Serum Replacement media"; N2 = DMEMF12 with N2 supplement; N2/N27 is DMEMF12 supplemented with N2 and B27) with each media represented by a different color. The collection times used for RNA-seq profiling are printed above the timeline. Shown below the schematic are phase images of cultures at days 0, 12, and 77 of the protocol.

(B) Real-time qPCR to verify differentiation of hESC to anterior neural progenitors. Prosencephalic marker, *PAX6*; Telencephalic marker, *FOXG1*; dorsal telencephalon marker *EMX2*; ventral telencephalic markers, *NKX2-1*, *DLX1*. No signal was observed for either of the posterior markers. All error bars are SD.

(C) Immunocytochemistry of known neural differentiation and cortical layer markers to establish validity of the protocol. Pluripotency markers: NANOG, POU5F1 (OCT4; day 0). Neural induction markers: NES (day 7), MKI67 (day 0), OTX2 (day 12), and PAX6 (day 12); ADRA2A (L6; day 19), DRD5 (L5; day 19), GRM4 (L4; day 35), and POU3F2 (L2–L4; previously BRN2; day 35).

(legend continued on next page)

web-based, searchable data set—http://cortecon.neuralsci.org/—with interactive access to the computational analyses.

RESULTS

Simulating Human Cortical Development In Vitro

We adapted previously published differentiation protocols (Chambers et al., 2009; Gaspard et al., 2008) to simulate cortical development in 2D adherent cultures. hESCs were grown and neural induction was initiated using the dual SMAD-inhibition protocol (Chambers et al., 2009). To ensure a dorsal telencephalic fate, the sonic hedgehog antagonist cyclopamine was added on day 3 of neural induction (Gaspard et al., 2008). Throughout the entire subsequent period of cortical differentiation, cultures were maintained in N2B27 medium (Gaspard et al., 2008) supplemented with FGF-2 (Figure 1A). To verify that this protocol was in good agreement with previously published results (Shi et al., 2012), we performed quantitative real-time PCR analysis during the first 12 days of the differentiation, focusing on several known markers of neural specification (Figure 1B). This revealed a steady increase in the prosencephalic marker *PAX6* (Zhang et al., 2010) and increased expression of the telencephalic marker *FOXG1*, indicating acquisition of a telencephalic fate (Hébert and Fishell, 2008). *EMX2* (Cecchi, 2002), a dorsal telencephalon marker, steadily increased over this period, indicating cerebral cortical fate acquisition (Figure 1B). Notably, no expression of the ventral telencephalic markers *NKX2-1* and *DLX1* (Hébert and Fishell, 2008) was detected.

The cell phenotypes produced in the cultures were identified using immunocytochemistry. Expression of the pluripotency markers NANOG and POU5F1 (OCT4) was high in the undifferentiated hESCs, and both proteins rapidly decreased after initiation of neural differentiation as expected (data not shown). Upon neural differentiation, the neural progenitor cell marker Nestin (NES) was induced, concomitant with high levels of the proliferative marker MKI67. Strong expression of PAX6 and OTX2 confirmed that a dorsal, cortical specification had occurred (Figure 1C).

Neurons destined for different cortical layers are formed in a set sequence. To assess whether the appropriate cortical cell layers were forming during the differentiation protocol, we carried out immunocytochemistry for markers selected from a recent study identifying several adult human layer-specific cortical markers by in situ hybridization in postmortem human brain (Zeng et al., 2012). Expression of adrenoreceptor alpha 2A (ADRA2A; layer VI) and dopamine receptor D5 (DRD5; layer V) confirmed deep layer formation, and expression of glutamate receptor, metabotropic 4 (GRM4; layer IV) and POU class 3 homeobox 2 (POU3F2; layer IV-II) indicated the presence of mid and upper layer cells. Together, these data show the 77 day protocol recapitulates human cortical development in vitro, encompassing all stages from pluripotency to upper layer neuron generation.

RNA-Seq Measures of In Vitro Human Cortical Development

We used RNA-seq to determine gene expression changes occurring during the span of corticogenesis. Analysis was carried out using R (Team, 2012) and packages available through Bioconductor (Gentleman et al., 2004). Data were derived from samples generated in two separate cortical differentiation experiments: a full differentiation (day 0, 7, 19, 33, 49, 63, and 77; two samples designated A and B) and a shorter differentiation (day 0, 7, 12, 19, and 26; two samples designated C and D). To verify that these two differentiations produced similar transcriptional profiles, we determined the Pearson product-moment correlation coefficient (Pearson's r) for all combinations of biological and experimental replicates (Figure 1D; Figure S1A available online) that showed a strong correlation between expression profiles generated on the same day of differentiation regardless of experimental origin, indicating strong reproducibility.

The cortical RNA-seq datasets were compared to data in the BrainSpan Atlas of the Developing Human Brain (Miller et al., 2014) (http://brainspan.org/), which contains transcriptome profiles generated by RNaseq from macrodissected and laser-microdissected human fetal brain regions from multiple time points during development (9, 12, 13, 17, 21, 24, 26, 56, and 92 postconception weeks). This analysis demonstrated the genes expressed by in vitro generated cortical cells correlated best with human fetal cerebral cortex transcriptomic profiles as compared to other brain regions (Figure 1E). Notably, the analysis indicates the cells generated by this protocol most closely resemble medial prefrontal cortex (Figure S1B) while showing the least similarity to primary somatosensory and auditory regions.

Global Temporal Transcriptome Analysis of In Vitro Corticogenesis

EdgeR and DESeq2 (Anders et al., 2013) were used to gain insight into the dynamic changes in the transcriptome during the cortical differentiation period. A total of 14,065 RNAs exhibited significant changes during the time course (adjusted p value <0.05 and a 2-fold change), common to both EdgeR and DESeq2, out of 31,041 expressed RNAs. A total of 1,729 of these were long intergenic, noncoding RNAs (lincRNAs) (Figure 2A; Table S1). To better understand the gene expression changes occurring, we utilized the GOseq package (Young et al., 2010) to determined enriched GO categories. CateGOrizer (Zhi-Liang et al., 2008), which parses GO analysis into a more user-friendly format, was used to count the number of enriched GO categories falling under a hand-annotated list of the developmental categories (Table S2). As expected, the major group of

(D) Pearson correlations of Cortecon replicates.

(E) Comparative analysis of RNA-seq data with BrainSpan Atlas of the Developing Human Brain of transcriptome profiles from macrodissected and laser-microdissected human fetal brain regions. Human sample ages range from 9 postconceptual weeks to 1 year old. Shown are the Spearman correlations with gross brain regions. Due to lack of data pertaining to the whole cortex in Brainspan, VFC was used to represent cortex in the brain-wide comparison because it closely correlated with other cortical regions. Ventrolateral Prefrontal Cortex (VFC); hippocampus (HIP); striatum (STR); amygdaloid complex (AMY); thalamus (TH); cerebellum (CB). See also Figures S1 and S2.

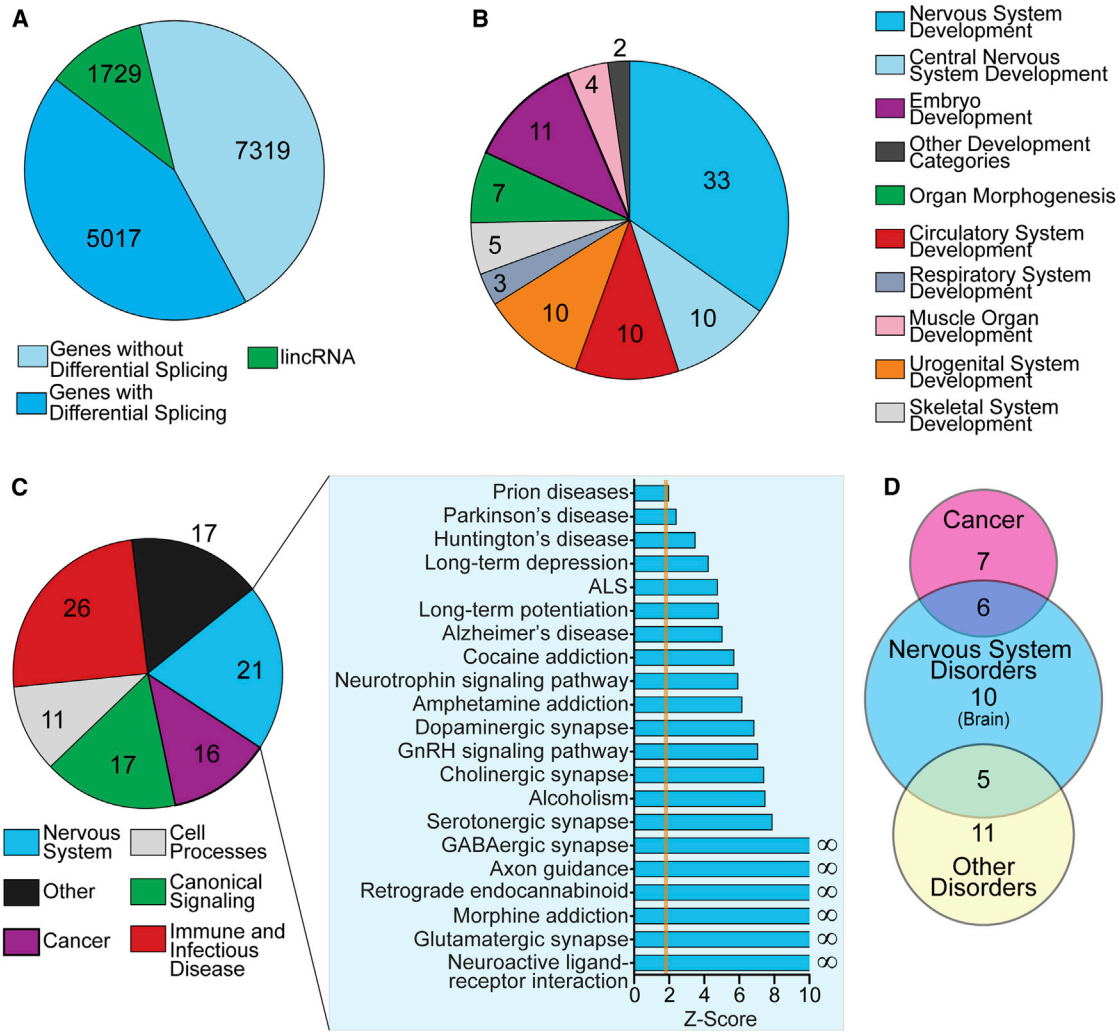


Figure 2. Global View of Temporal Expression Profiles during In Vitro Corticogenesis

RNA-seq was used to determine the transcriptome at frequent intervals during in vitro corticogenesis.

(A) Distribution of significantly changing transcripts.

(B) Significantly enriched developmentally related GO categories aggregated into more comprehensive terms.

(C) Significantly enriched KEGG pathways in the global gene set and their distribution among six categories. Nervous-system-related pathways are highlighted and displayed with calculated z scores.

(D) Significantly enriched diseases categorized as being a "nervous system disorders," brain-related disease, a type of "cancer," "other disorders," or a combination of these categories. See also Tables S1, S2, S3, and S4.

GO categories enriched in this global data set aligned with the "Nervous System Development" category, providing confidence in the differentiation and analysis, and smaller groups aligned with categories describing the development of a variety of other anatomical systems (Figure 2B; Table S2).

Key signaling pathways in the global RNA-seq data were elucidated using the Signaling Pathway Impact Analysis (SPIA) package (Tarca et al., 2009). SPIA utilizes the curated KEGG pathway repository to help determine the relative representation of specific pathways in expression data. Significantly enriched KEGG signaling pathways included several main categories: "Cell Processes," "Cancer," "Canonical Signaling," "Nervous System," and "Immune and Infectious Disease" (Figure 2C; Table S3). Many pathways implicated in normal neuronal development were highlighted by SPIA. For example, genes falling under the neuroactive-ligand receptor interaction-signaling pathway register as enriched. The glutamatergic synapse-signaling pathway is also enriched, which is encouraging, as the majority of endogenous cortical neurons are glutamatergic. In addition, several other synaptic pathways are enriched, such as the GABAergic, serotonergic, and dopaminergic pathways, signifying that a multitude of genes required for neuronal function increase expression over the time course. For example, neurotrophic factors (e.g., BDNF, Neurotrophin 3) play key roles in neuronal maturation, synaptogenesis, and cortical development (Zweifel et al., 2005) and show expected dynamic changes throughout

the time course (Figure S2). Genes contributing to synaptic function, such as SLC17A6 (Kaneko and Fujiyama, 2002), increase during the later stages as expected (Figure S2).

Genes showing differential expression during in vitro corticogenesis were analyzed using the DISEASES database (http://diseases.jensenlab.org/), which ranks disease-gene associations mined from the literature using Disease Ontology terms. Using GOseq along with the DISEASES database, we observed that several different categories of diseases were enriched in the global data set, the majority being related to "Cancer," "Nervous System Disorders," or a combination thereof (Figure 2D; Table S4).

Classification of Genes into Different Cortical Developmental Stages

We used fuzzy c-means clustering (Kumar and E Futschik, 2007) to group genes with significant changes in temporal expression into clusters based on similarity in their temporal expression patterns (Figure S3). This analysis generated 64 clusters ranging in size from 65 to 854 genes with an average of 220 genes per cluster (Figure S3).

The clusters were then classified by developmental stage, using five different categories: "Pluripotency" (PP), "Neural Differentiation" (ND), "Cortical Specification" (CS), "Deep Layer neuron generation" (DL), and "Upper Layer neuron generation" (UL). To establish these categories, we examined genes known to be associated with each stage and used their expression profiles to help establish milestones along the differentiation time course. Pluripotency genes such as *POU5F1* (*OCT4*), *NANOG*, *NODAL*, and *TDGF1* were found to be expressed highest at Day 0 and to decrease dramatically thereafter, as expected, thus setting clusters with highest expression at Day 0 as the PP stage. *PAX6*, a known early indicator of neural commitment, begins to be expressed at Day 7 and *Sox1*, another early neural marker, peaks at Day 7; hence, we set the start of the ND stage at Day 7. *EMX2*, a master regulator of cortical development (Cecchi, 2002), peaks at Day 12, setting the start of the CS stage at day 12. For the DL stage, we examined TBR1 expression, a layer 6 marker (Hevner et al., 2001), and found it upregulated at day 26. Additionally, a marker of layer 5 cortical projection neurons, BCL11B (CTIP2), rises in expression at day 33. Thus, we defined the start of DL neuron generation at day 26. For the UL stage, we looked at the recently identified human upper layer markers *CACNA1E*, *PRSS12*, and *CARTPT* (Zeng et al., 2012) and found their expression rises at day 63 and continues throughout, thus defining the start of the UL stage as Day 63.

Once we established the developmental stages, we grouped all the clusters with positive slopes of a certain degree (including peaks) during the ND thru UL stages. In the case of the PP stage, we grouped the clusters showing a negative slope after day 0. This was done using a heuristic algorithm based on the changing slopes of gene expression, rationalizing that positive slopes indicate a critical role at that point in the process. It is possible that genes could play critical roles at more than one period of corticogenesis; for example, some genes are expressed in both precursor cells and postmitotic neurons, so clusters exhibiting multiple peaks were classified into multiple stages. To take into account multiple stage membership, we generated a weighted average of clusters for each stage by first calculating the number of stages each cluster belongs to. We then gave less weight to those clusters belonging to more than one stage and calculated a representative expression profile for each developmental stage (Figure 3A).

We next examined the other genes and clusters associated with each developmental stage. The PP stage (containing 24 clusters) was exemplified by cluster 31 (Figure S3) and included not only the pluripotency genes already mentioned but also genes known to be vital to the differentiation of other lineages, such as SOX17. Clusters present in the ND stage include cluster 38 (Figure S3), which has the *LHX2* a gene implicated in neural differentiation (Hou et al., 2013), and cluster 13 (Figure S3), which contains *NPTX1*, a gene that plays a role in neural induction (Boles et al., 2014). In addition to the known CS gene *EMX2*, clusters with peaks at this stage (e.g., cluster 56; Figure S3) include *MDGA1* and *FGF8* (cluster 54; Figure S3), already implicated in cortical specification (Takeuchi et al., 2007; Toyoda et al., 2010).

Global Analysis of Each Stage of Corticogenesis

After classifying genes into a developmental stage, we functionally categorized them via a GO analysis using the GOseq package (Figure 3B; Table S6). We identified enriched GO categories for each stage and then used cateGOrizer to count the number of GO categories. As expected, we found "Nervous System Development" categories to be a major component of each stage, starting with ND (Figure 3B).

Using the DISEASES database (Frankild and Jensen, 2013), we determined disease affiliations that are overrepresented in the genes associated with each stage (Figure 3C; Table S7). For display purposes, we grouped diseases into four categories: "Eye Diseases," "Nervous System Diseases," "Cancers," and "Other Diseases." As pluripotent stem cells are inherently tumorigenic due to their propensity for teratoma formation, it was not surprising that the PP stage is enriched for several cancer types. Interestingly, both the ND and DL stages showed negligible disease enrichment. The majority of disease terms enriched in the CS stage are related to the nervous system and include the major category "Neurodegenerative Disease." The UL stage shows enrichment for the most disease terms, and genes associated with diseases such as schizophrenia and autism are enriched in this stage. Moreover, three of the four cancer terms associated with the UL stage are cancers of the nervous system. Finally, in examining the "Other Diseases" category, three of the 15 diseases have connections to cortical and nervous system defects (Bardet-Biedl syndrome, sudden infant death syndrome, and lipoidosis).

Changes in signaling pathways for each stage were revealed using the SPIA package. The SPIA method performs two measurements, the first utilizes expression data to determine if the components of pathways, as curated in the KEGG pathway database, are significantly overrepresented. The second measurement determines how likely it is that the current condition impacts the functioning of each pathway (activating or rendering it inactive). Pathways were categorized into six groups: "Nervous System," "Cell Processes," "Canonical Signaling," "Immune and Infectious Disease," "Cancer," and "Other"

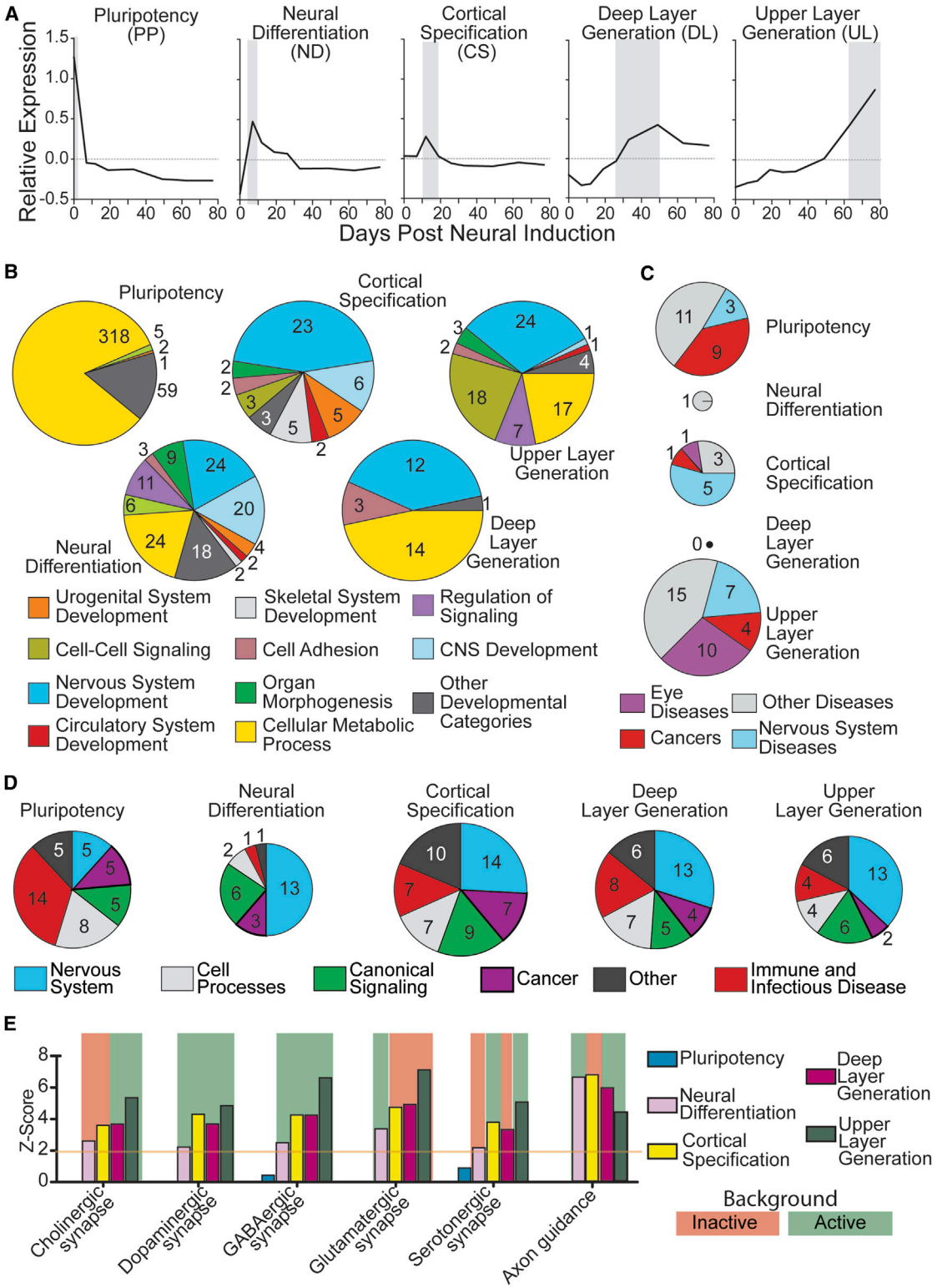


Figure 3. Developmental Stage Analysis of Temporal Expression Profiles

(A) Individual clusters were assigned to cortical developmental stages, classified as "Pluripotency" (PP), "Neural Differentiation" (ND), "Cortical Specification" (CS), "Deep Layer neuron generation" (DL), and "Upper Layer neuron generation." Plots for each stage are shown, with the time period of a stage illustrated by a gray band.

(legend continued on next page)

(Figure 3D; Table S8). As expected, the greatest number of pathways enriched in the ND through UL stages pertained to the "Nervous System" group, such as axon guidance and different neurotransmitter pathways (Figure 3E). Together, these data provide functional insight into genes expressed at the different stages of cortical development.

Functional Classification of Gene Clusters

We next examined each individual cluster (Figure S3) using GO, KEGG pathway, and DISEASES enrichment analysis, as described above. Clusters 18 (166 genes) and 34 (143 genes) showed the greatest number of enriched developmental categories, and both of these clusters were associated with the ND and CS stages (Figure 4A; Table S9). In cluster 18 nearly half (13/28) of the associated GO categories deal with nervous system development, and cluster 34 had most categories dealing with nervous system development (10/13), with over half of those more specifically involved in central nervous system development (6/10). Cluster 18 and 34 are similar in having an early peak that drops to a sustained level of expression; however, cluster 18 expression rises at a faster rate than cluster 34 at the early peak. Intriguingly, the composition of these clusters is quite different: cluster 18, which includes 166 genes (Figure S3), contains 42 (25%) transcription factors, whereas cluster 34, which includes 143 genes, contains 50 (35%) membrane proteins and 22 (15%) secreted molecules, illustrating two very different roles for these clusters in the process of corticogenesis. Another interesting cluster in this analysis is cluster 56 (152 genes), which is associated with the ND, CS, and UL stages. It has seven enriched categories, and all are involved in nervous system development. It includes 26 lincRNAs, 18 genes with transcription regulator activity, 21 secreted proteins, 39 transmembrane proteins, and assorted other genes. Of the 18 transcription regulatory factors, only four have previously been associated with cortical development, and it will be worthwhile investigating whether the remaining 14 genes are novel regulators of human corticogenesis.

SPIA pathway analysis highlighted a different set of clusters, with 32 clusters showing enrichment in at least one pathway (Figure 4B; Table S10). Cluster 47 (280 genes), which is associated with the PP and DL stages, was enriched for the greatest number of pathways, with two related to the nervous system, two related to cancer, and six related to the immune system. This cluster shows high expression in the PP stage and a shallow peak during DL formation, after which expression peters out, suggesting reduced involvement in mature cortical function. Genes associated with Huntington's disease and Alzheimer's disease pathways were also enriched in this cluster, and interestingly, these pathways are rated as inactive based on the specific genes present in cluster.

Cluster 58 (243 genes) and cluster 39 (370 genes) are specifically associated with the UL stage. For cluster 58, four of the six enriched pathways are concerned with the nervous system. Cluster 58 includes a significant number, 77 (31% of the cluster), of lincRNAs, the second highest lincRNA complement of any cluster. This cluster also includes 33 (14%) secreted factors and 60 (25%) transmembrane proteins. Interestingly, the other cluster specific to the UL stage, cluster 39 (370 genes), also has most of its enriched pathways associated with the nervous system (2 out of 3). Furthermore, the composition of the two clusters is similar, although cluster 39 contains fewer (46, 12%) lincRNAs. The combination of GO and KEGG analyses presented here provides functional insight into coordinated gene sets at different stages of cortical development and highlights several that would be worthwhile exploring further.

We next investigated the clusters for associations with disease (Figure 4C; Table S11) as described above. Only 11 clusters showed enrichment for genes involved in disease, of which two clusters are notable due to the number of associated disease terms. Cluster 31 (854 genes) is specifically associated with the PP stage and shows an enrichment for 14 disease terms, eight of which are cancer related. Many genes driving pluripotency also have roles in carcinogenesis (Lee et al., 2013), so this finding is consistent with pluripotent cell biology. In contrast, when looking at cluster 39, which in the KEGG pathway analysis is principally involved in nervous system pathways, only seven disease terms were associated with the nervous system, while 21 of 43 enriched terms were associated with cancer. Cluster 39 also showed enrichments for genes associated with developmental disorders that affect cortical function. For example, there is an enrichment in autism-associated genes, including *WNT2*, *GABRG1*, and *C4B*, and in schizophrenia-associated genes, including *HTR2A*, *TPH1*, and *ZNF804A* (Frankild and Jensen, 2013), indicating a potential role for other genes in this cluster in those diseases.

Alternatively Spliced Transcripts during Human Corticogenesis

Alternative splicing is an essential step in regulating the function of the majority of protein-coding genes, and studies have implicated abnormal splicing in several medical disorders (Padgett, 2012). Using the DEXseq package (Anders et al., 2012), we identified genes with evidence of alternative splicing in the Cortecon data set and then looked for disease associations (Table S12).

As shown earlier, 5,017 genes had differential splicing during in vitro corticogenesis (Figure 2A), with an average of 2.2 exons per gene changing expression during the differentiation time course. We used this subset of spliced genes, associated with stage and cluster, to look for enrichment of disease terms as we did for the total gene sets described above. Interestingly, in the PP and CS stages, the spliced gene subsets compared to the total stage gene sets were enriched in disease terms related to cancer (Figure 5A; Table S13). In concordance, cluster 31 showed an increase in the number of enriched disease terms

(B) GO enrichment for each of the cortical developmental stages.
(C) Enrichment for disease terms for each of the cortical developmental stages.
(D) KEGG pathway enrichment for each of the cortical developmental stages.
(E) Selected nervous-system-related KEGG pathways highlighted and displayed with calculated z scores. Backgrounds of each graph in (E) and (F) indicate if a pathway is active or inhibited by genes in each stage. See also Tables S6, S7, and S8.

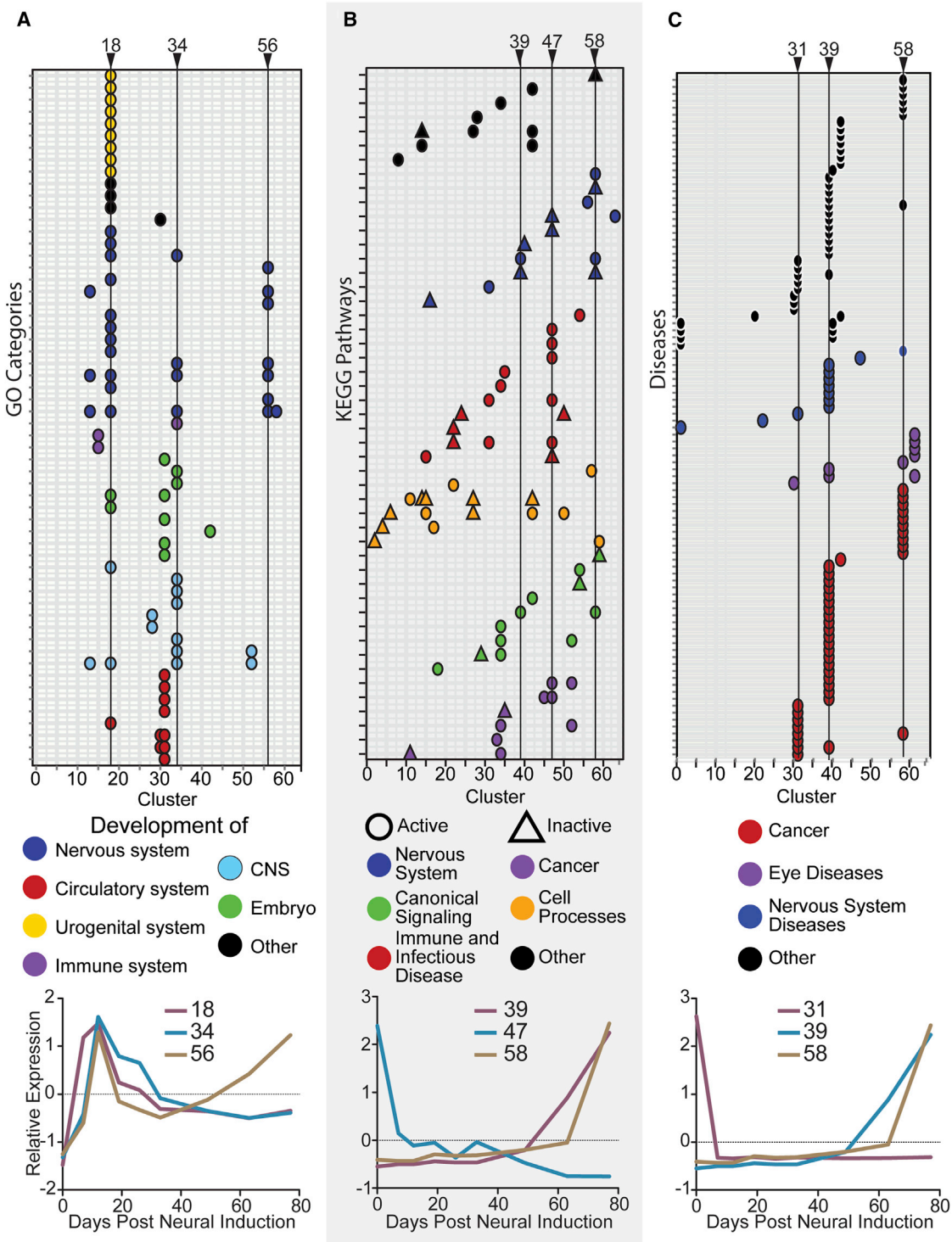


Figure 4. Individual Cluster-Based Analysis of Temporal Expression Profiles

(A) GO enrichment for each of the temporal expression profile-derived clusters. Black arrowheads with numbers indicate specific clusters.

(B) KEGG pathway enrichment for each of the temporal expression profile-derived clusters. Black arrowheads with numbers indicate specific clusters.

(C) Enrichment for diseases for each of the temporal expression profile-derived clusters. Black arrowheads leading into black lines with numbers indicate specific clusters. Expression profiles of clusters highlighted in each analysis is shown below each legend. See also Tables S9, S10, and S11 and Figure S3.

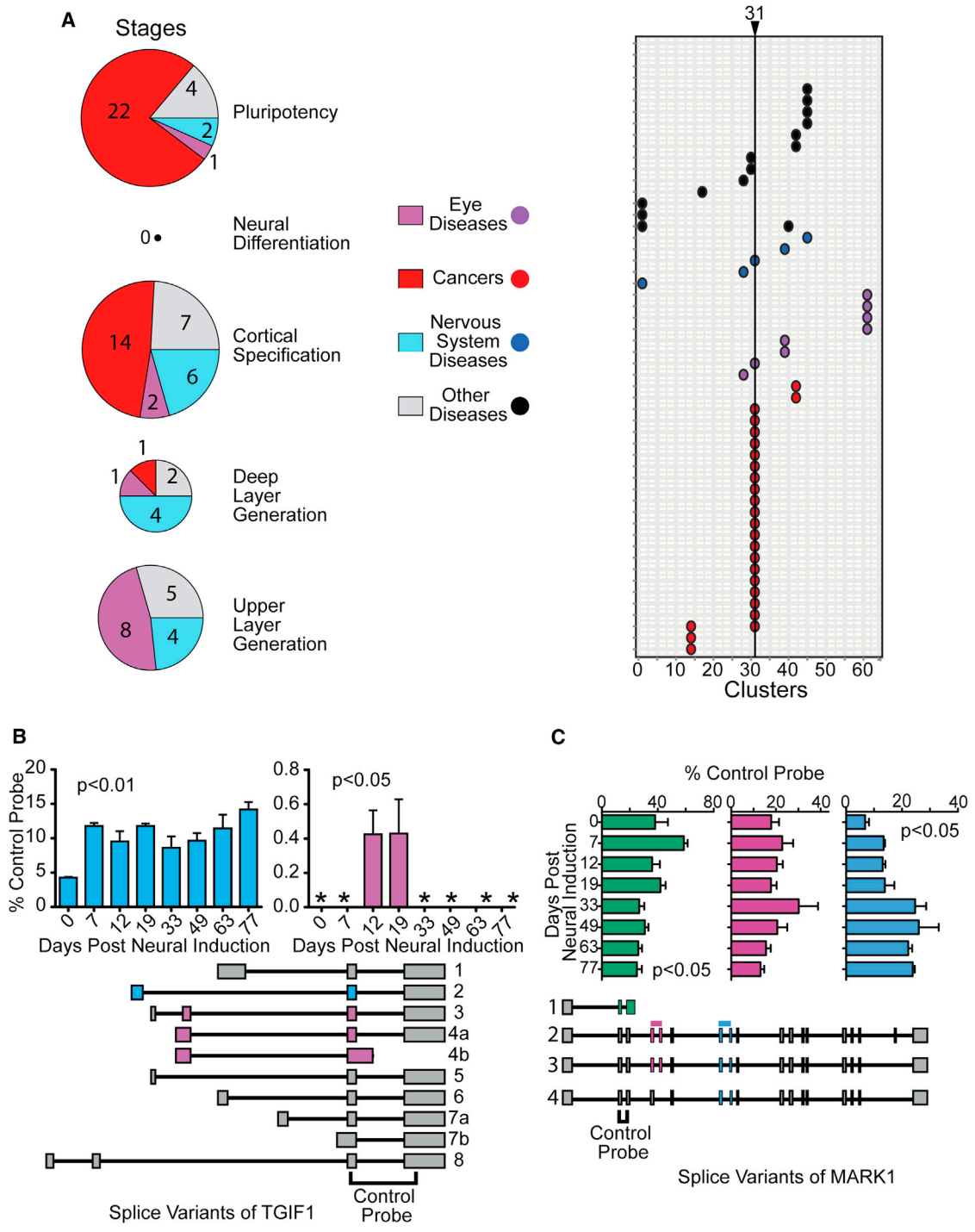


Figure 5. Splice Form Analysis Reveals Roles in Neural Diseases and Cancer

Identification of genes in the data set with significant temporal changes in alternative splice forms.

(A) Enrichment for diseases for genes with alternative splice forms for both cortical developmental stages (left panel) and temporal expression profile-derived clusters (right panel).

(B) Verification of alternative splice forms of *TGIF1* by qRT-PCR.

(C) Verification of alternative splice forms of *MARK1* by qRT-PCR. See also Tables S12, S13, and S14.

All error bars are SD.

Table 1. Selected Genes from Each Stage

Entrez Gene ID	Gene Symbols	Stage	Stage Eigenvector	Stage Betweenness	Associated Disease	Max/Min	Max/Threshold
183	AGT	CS	0.8515	0.0162	Diabetes mellitus	626.88	227.72
		UL	0.8571	0.0400	Brain disease		
185	AGTR1	CS	0.7993	0.0190	Diabetic retinopathy	Inf	7.85
		UL	0.7929	0.0067	Brain disease		
1565	CYP2D6	ND	0.5859	0.4384	Alzheimer's disease	6.79	9.46
1571	CYP2E1	ND	0.9465	0.1982	Parkinson's disease	24.94	2.66
1636	ACE	CS	1.0000	0.1935	Central nervous system disease	102.35	62.99
		DL	0.1008	0.1516	Migraine		
		UL	1.0000	0.2366	Alzheimer's disease		
1813	DRD2	ND	0.5572	0.8659	Schizophrenia	22.87	56.87
1815	DRD4	ND	0.4424	0.0204	Autism	68.99	1605.65
2153	F5	UL	0.5776	0.0877	Migraine	274.48	66.40
2944	GSTM1	ND	1.0000	0.9109	Carcinoma	2.50	572.52
2947	GSTM3	ND	0.5304	0.0301	Childhood brain tumor	26.81	537.75
3106	HLA-B	DL	0.6227	0.0385	Neurodegenerative disease	4.84	143.83
3113	HLA-DPA1	DL	0.0840	0.0016	Amyloidosis	4.28	10.11
3119	HLA-DQB1	DL	0.9222	0.0000	Pilocytic astrocytoma	7.30	157.24
3123	HLA-DRB1	DL	1.0000	0.0590	NA	7.24	45.24
3383	ICAM1	DS	0.3573	0.1023	Blindness	27.60	1451.80
3552	IL1A	DL	0.1026	0.0358	Alzheimer's disease	15.28	98.80
3553	IL1B	CS	0.6087	0.5317	Alzheimer's disease	Inf	8.02
		DL	0.1223	0.4333	Schizophrenia		
3557	IL1RN	UL	0.3693	0.0141	Brain disease	Inf	1.14
3569	IL6	UL	0.6628	1.0000	Alzheimer's disease	Inf	34.19
3576	IL8	CS	0.4216	0.1884	Brain neoplasm	Inf	71.94
3990	LIPC	ND	0.4340	0.0592	Age-related macular degeneration	Inf	4.47
4023	LPL	ND	0.4934	0.8059	Cancer	44.81	1630.98
4524	MTHFR	CS	0.6716	1.0000	Down syndrome	6.06	226.72
		DL	0.0938	0.8549	Alzheimer's disease		
		UL	0.7463	0.8106	Schizophrenia		
4846	NOS3	CS	0.6416	0.2358	Alzheimer's disease	4.95	62.27
4968	OGG1	ND	0.6267	0.0546	Neurodegenerative disease	3.52	291.74
5468	PPARG	UL	0.3666	0.3778	Alzheimer's disease	8.73	9.48
5683	PSMA2	PP	0.9990	0.0016	Alveolar echinococcosis	3.90	727.40
5684	PSMA3	PP	0.9995	0.0104	Chronic fatigue syndrome	5.05	5225.13
5685	PSMA4	PP	0.9904	0.0090	Keratoconus	3.87	2397.07
5687	PSMA6	PP	1.0000	0.0098	Heart disease	2.93	1216.66
5688	PSMA7	PP	0.9977	0.0067	Cancer	3.67	12907.39
5689	PSMB1	PP	0.9819	0.0014	Visceral leishmaniasis	2.67	2730.82
5690	PSMB2	PP	0.9845	0.0008	NA	2.95	693.59
5692	PSMB4	PP	0.9819	0.0026	Neurotic depression	2.19	7890.62
5693	PSMB5	PP	0.9857	0.0002	Multiple myeloma	2.51	6924.69
5701	PSMC2	PP	0.9897	0.0193	NA	2.35	1190.04
7040	TGFB1	CS	0.5385	0.4873	Holoprosencephaly	12.40	983.38
		DL	0.0780	0.2332	Alzheimer's disease		
		UL	0.5636	0.4556	Glioma		

(*Continued on next page*)

Table 1. *Continued*

Entrez Gene ID	Gene Symbols	Stage	Stage Eigenvector	Stage Betweenness	Associated Disease	Max/Min	Max/Threshold
7422	VEGFA	CS	0.4208	0.5852	Astrocytoma	5.82	401.39
		UL	0.3529	0.3714	Neuroblastoma		
7515	XRCC1	ND	0.8439	0.9213	Carcinoma	2.39	1766.83
100507436	MICA	DL	0.0987	0.0137	Carcinoma	6.47	121.50

Shown are the top ten genes from each stage according to eigenvector centrality scores as derived from the network analysis. The fold change for each gene from its peak to its trough in the time course is shown in the Max/Min column. The Max/Threshold column shows change between the calculated threshold for each gene and the max value. The remaining columns show the eigenvector centrality scores (EV) of the gene and betweenness centrality score for its stage. Betweenness scores were normalized against the top betweenness score in its stage so that a score of 1 represents the highest betweenness score.

when comparing all genes in the cluster to those genes that are spliced, and a higher percentage of the terms are cancer related in the spliced genes (91% versus 57%) (Figure 5A; Table S14). Of the genes spliced differently over time, a number were associated with cortical-associated diseases.

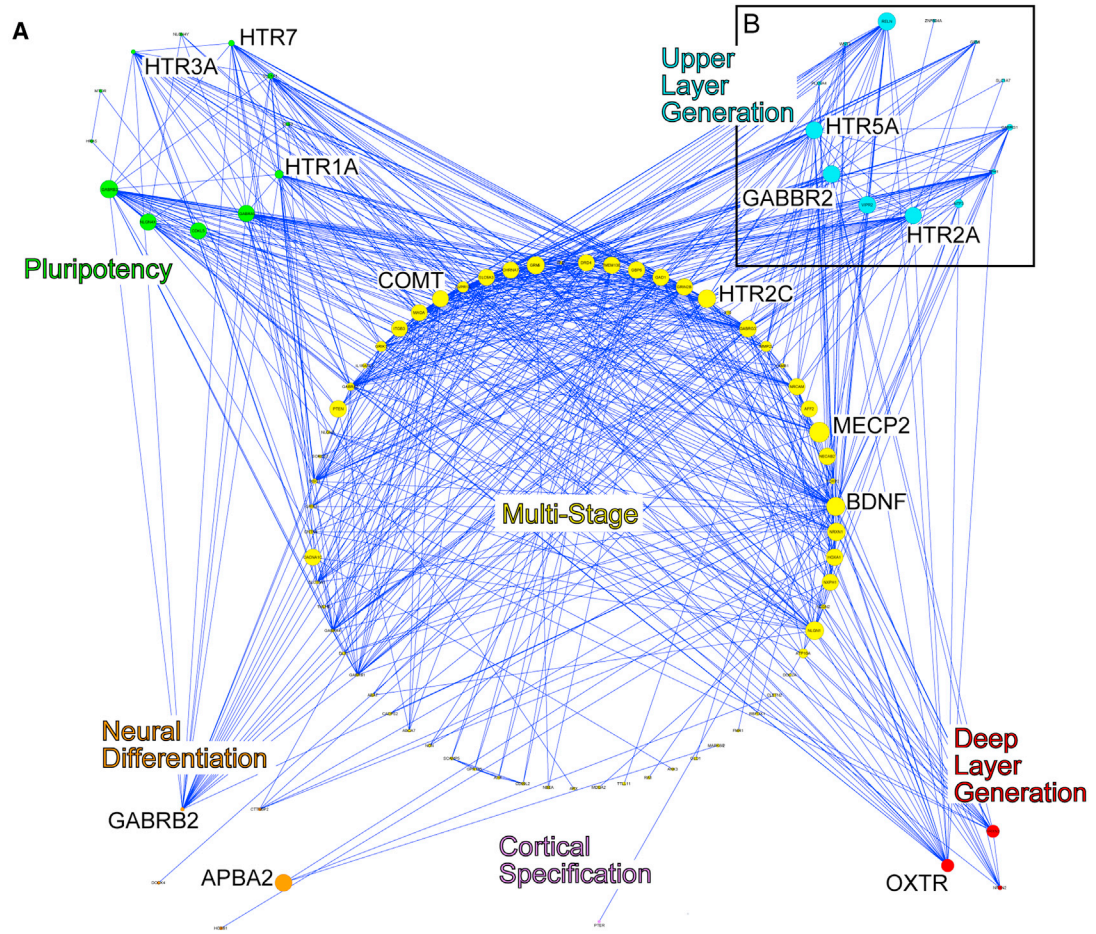
To confirm alternative splicing at different stages, we carried out qRT-PCR targeting specific splice forms for select examples. We chose to examine *TGFB-induced factor homeobox 1* (*TGIF1*; cluster 14) due to its association with holoprosencephaly (El-Jaick et al., 2007). We designed probes spanning exon junctions with one probe belonging to nearly all splice forms of *TGIF1* to act as a control and two spanning exon junctions of selected *TGIF1* isoforms and performed qRT-PCR. The expression values were plotted as a percent expression of the control probe. One exon junction demonstrated expression specifically at days 12 and 19 of differentiation (CS stage), while no expression was detected at any of the other time points (Figure 5B, "purple"). This junction is associated with three splice forms of *TGIF1*, one of which (splice form 4a) is the predominant *TGIF1* splice form expressed in a range of human tissues, including brain (Hamid et al., 2008). Thus, the expression of this exon junction specifically during cortical specification, with low expression at later times, may indicate a need for one of these three splice forms during cortical patterning. One might speculate that as holoprosencephly occurs early in nervous system development, these splice forms may be specifically involved in that pathology. The other exon junction tested was unique to a specific *TGIF1* splice forms (Figure 5B, "blue"), and it demonstrated changing expression throughout the time course, with expression being lowest during PP but increasing throughout the remainder of corticogenesis (Figure 5B), suggesting a wider role during the process. Together, these data demonstrate specific isoforms of TGIF1 are expressed at specific times during cortical development, so that defects in the splicing events at these times could contribute to developmental disorders.

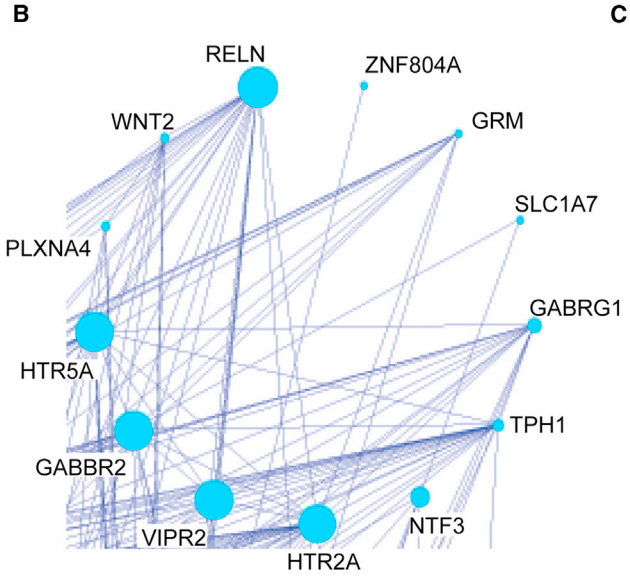
We next examined *MAP/microtubule affinity-regulating kinase 1* (*MARK1*) due to its implication in autism spectrum disorder (ASD) (Maussion et al., 2008). We used four probes spanning exon junctions: one spanning all to serve as a control, a second to test a short *MARK1* splice forms, and a third and fourth probe that split the remaining variants into two groups (Figure 5C). At the start of the differentiation process, variant 1, the short splice forms, makes up approximately 35% of the total MARK1 message, increases to approximately 60% on day 7, reduces to 35% at day 12–19 (Figure 5C), and thereafter steadily decreases to 25% by day 77. Other probes indicate an increase in the proportion of variant 4 splice forms as corticogenesis progresses, while splice forms 2 and 3 maintain a similar proportion throughout. This suggests the short *MARK1* splice forms (variant 1) might play a critical role during early neural development and cortical specification, while the *MARK1* variant 4 splice forms might play a greater role at later stages of cortical development. It will be interesting to explore these different splice forms to determine if one in particular is critical for the progression of autism-related pathologies.

Network Analysis Shows Autism-Related Genes in the Context of Cortical Developmental

As development proceeds, waves of gene expression must work coordinately to establish a fluid, systematic, neural progenitor expansion and differentiation process. To further demonstrate the utility of this data set and to obtain a greater understanding of how genes in each cluster and stage relate to each other throughout cortical development, we created a network based on cocitations (see Supplemental Experimental Procedures, available for download from http://cortecon.neuralsci.org/). The resulting network was then filtered against lists of genes belonging to a particular cluster or stage to obtain subgraphs. Graph theoretic measures of "centrality" were then utilized to find important nodes (genes) based on the relationships defined within the structure of these subgraphs.

Using the igraph package (Csardi and Nepusz, 2006), we utilized two methods for computing centrality, one more quantitative (eigenvector centrality) and one more heuristic (betweenness centrality) to determine genes pivotal to each cluster and stage (Tables 1, S15 and S16). Utilizing these two distinct measures of centrality, we provide insight into the key players for each cluster and developmental stage. To illustrate this method of examining the data, we chose to focus on genes associated with ASD from the DISEASES database (Frankild and Jensen, 2013) that change significantly in our data set. ASD encompasses a range of neurodevelopmental disorders, characterized by social and communication deficits and repetitive behaviors (Halfon and Kuo, 2013). We created a subgraph comprising only those 171 genes related to ASD that change significantly over the time course of in vitro corticogenesis (Table S17). After constructing this subgraph, the genes in the graph were then grouped by stage to visualize a potential ordering of gene





C

Gene	Stage	Stage Eigen	Stage Between	Autism Eigen	Autism Between
BDNF	CS	0.124	0.130	0.548	0.522
	DL	0.116	0.100		
	UL	0.012	0.088		
COMT	CS	0.300	0.459	1.000	0.010
	DL	0.030	0.129		
	UL	0.216	0.395		
MECP2	ND	0.005	0.024	0.016	0.994
	CS	0.003	0.014		
	DL	0.002	0.005		
	UL	1.6e-4	0.004		
GABRB2	ND	0.050	0.010	0.137	0.009
APBA2	ND	1.5e-4	0.009	1.4e-4	0.298
OXTR	DL	0.004	4.6e-7	0.050	0.127
GABBR2	UL	0.026	0.012	0.049	0.965
HTR2A	UL	0.156	0.109	0.865	0.050
HTR2C	CS	0.066	0.005	0.544	1.000
	UL	0.070	0.004		

Figure 6. Literature-Based Network Analysis for ASD-Associated Genes in Data set

(A) Cocitation network for ASD-associated genes (DISEASES database) with significant temporal expression in RNA-seq data set. Larger nodes representing genes of more importance to the network by betweeness centrality. Network is set up according to the cortical developmental stages: "Pluripotency" (PP; green), "Neural Differentiation" (ND; orange), "Cortical Specification" (CS; purple), "Deep Layer neuron generation" (DL; red), "Upper Layer neuron generation" (UL; blue), and "multi-stage" ("linked": yellow).

(legend continued on next page)

interactions during the developmental process. Genes unique to a single stage (PP, ND, CS, DL, and UL) are each arranged in five groups with a sixth group comprised of genes associated with multiple stages (Figures 6A and 6B). Distributed around the central group are groups of genes unique to a single stage, arranged in a counter-clockwise fashion with the "Pluripotency stage" at 10 o'clock and each following stage thereafter. Genes that have high centrality are highlighted in the network (Figure 6C). This schematic (Figures 6A and 6B) gives an overall impression of (a) the activity of genes over time (according to the stages), (b) the importance of genes according to their measure of centrality, and (c) the relationships/connections between genes associated with autism.

A general impression of the developmental autism network is that a myriad of receptors and ligands are at the center of each stage (Figure 6). The gene with the highest eigenvector score is *catechol-O-methyltransferase* (*COMT*), one of several enzymes that degrade catecholamines such as dopamine, epinephrine, and norepinephrine. Dopamine degradation by COMT is of particular importance to the functions of the prefrontal cortex, a region that our differentiation protocol enriches for (Figure 1E), and polymorphisms in COMT affect working memory (Matsumoto et al., 2003). A serotonin receptor, *HTR2C*, has the highest betweenness score in the network. Dysfunction in serotonin signaling has been linked to autism (Scott and Deneris, 2005), and several members of the serotonin receptor-family in addition to *HTR2C* are also prominent at different cortical developmental stages in the network: "PP" (green), *HTR3A*, *HTR1A*, and *HTR7* and "UL" (blue), *HTR5A*, and *HTR2A*. *HTR2A*, in addition to having the fourth highest eigenvector score in this network, is also in the top 1% of eigenvector scores in the UL stage and has the highest eigenvector score for cluster 39. Previous studies have shown serotonin plays a role in cortical development, and loss of serotonin results in abnormal development of neurons in cortical layers 3 and 5 in rats (Vitalis et al., 2007). Our data supports these studies and lends further credence to the importance of serotonergic signaling during cortical development and in ASD. Thus, the use of these networks can highlight the interconnectivity of genes changing during the different stages of cortical development and provide insight into which stages might be most important for a specific disease, such as ASD. By using the data provided in this resource, similar networks can be created for other diseases to highlight disease associated processes.

Resource Utilization

To enhance the utility of this resource, we present the data and analytical results in a searchable, web-based resource (http://cortecon.neuralsci.org). Users can search for and navigate genes, clusters, diseases, pathways, and ontology categories of interest. The Cortecon resource provides a valuable standard for comparison of other data sets. While the protocol described here can robustly generate human cortical cells, culture variations exist from lab-to-lab, with two common variables being (1) the variability of pluripotent stem cell (iPSC and hESC) lines and (2) the starting cell density of the pluripotent stem cells. As an example, when cortical differentiation was performed using the same hESC line but with a lower cell density (approximately 60% compared to 85%) at the initiation of neural differentiation, changes were observed in the resultant time course of corticogenesis, based upon the expression profile (Figure 7). The cells were cultured for the full span of 77 days, with RNA sampled at the same time points described in Figure 1 for RNA-seq.

Using Spearman rank correlations (Figures 7A and 7B), we compared the original Cortecon data set to the lower density test run data set and saw strong correlations at day 0 and day 7 between the two runs; however, correlations decreased at later time points (Figure 7A). In order to assess whether the rate of differentiation was different, we used the top 1,000 changing genes in the Cortecon data as landmarks for the temporal progression of corticogenesis (Figure 7B). Based upon these landmarks, the test run differentiation was accelerated throughout the early time course compared to the Cortecon data set (i.e., day 33 of the test run showed strong correlations with day 33–63 of the Cortecon data). We next examined the overlap of significantly changing genes (Figure 7C) and found that approximately 70% of the Cortecon genes were changed in the test case. After examining the enriched GO Biological Processes categories from both data sets, we found a high overlap between each differentiation (Figure 7D). Finally, we compared the percentage of deep and upper layer markers present in each data set and noticed that while decreased compared to the Cortecon data set, the test protocol did robustly generate cortical neurons (Figure 7E). Importantly, we were able to demonstrate that by using simple correlation analyses, researchers can perform a temporal alignment of their data to this reference set for corticogenesis, regardless of cell density.

We also compared the Cortecon data set to a similar data set generated from mouse embryonic stem cells (Hubbard et al., 2013) (Figure S4). Using the human/mouse homologs, the Pearson product-moment correlation coefficient (Pearson's r) was calculated between the time points of each data set. Within each data set, adjacent time points were highly correlated, while overall there was low correlation between the mouse and human data, although the earlier time points were more correlated (Figure S4A). We next examined the expression of select genes involved in neural differentiation and corticogenesis. Normalized counts for each gene were scaled to the average expression of the gene over the time course for each data set to facilitate the comparison of expression profiles (Figure S4B). *FOXG1* nearly has identical profiles in mice and human corticogenesis, and *PAX6* and *EMX2* have very similar profiles. Finally, GAD1 and GRIK4 trend the same in both mouse and human, but their profiles are different (Figure S4B). Thus, "Cortecon" is a valuable tool for researchers to identify similarities or differences even across species over the differentiation time course.

(B) ASD-associated gene network for "Upper Layers" (UL) stage highlighted.

(C) Table with genes that display highest Eigenvector centralities and Betweenness centralities within specific developmental stages of the ASD-gene network. Eigenvector and Betweenness values for the genes within overall cortical developmental stages are given as well. See also Tables S15, S16, and S17 and Figure S5.

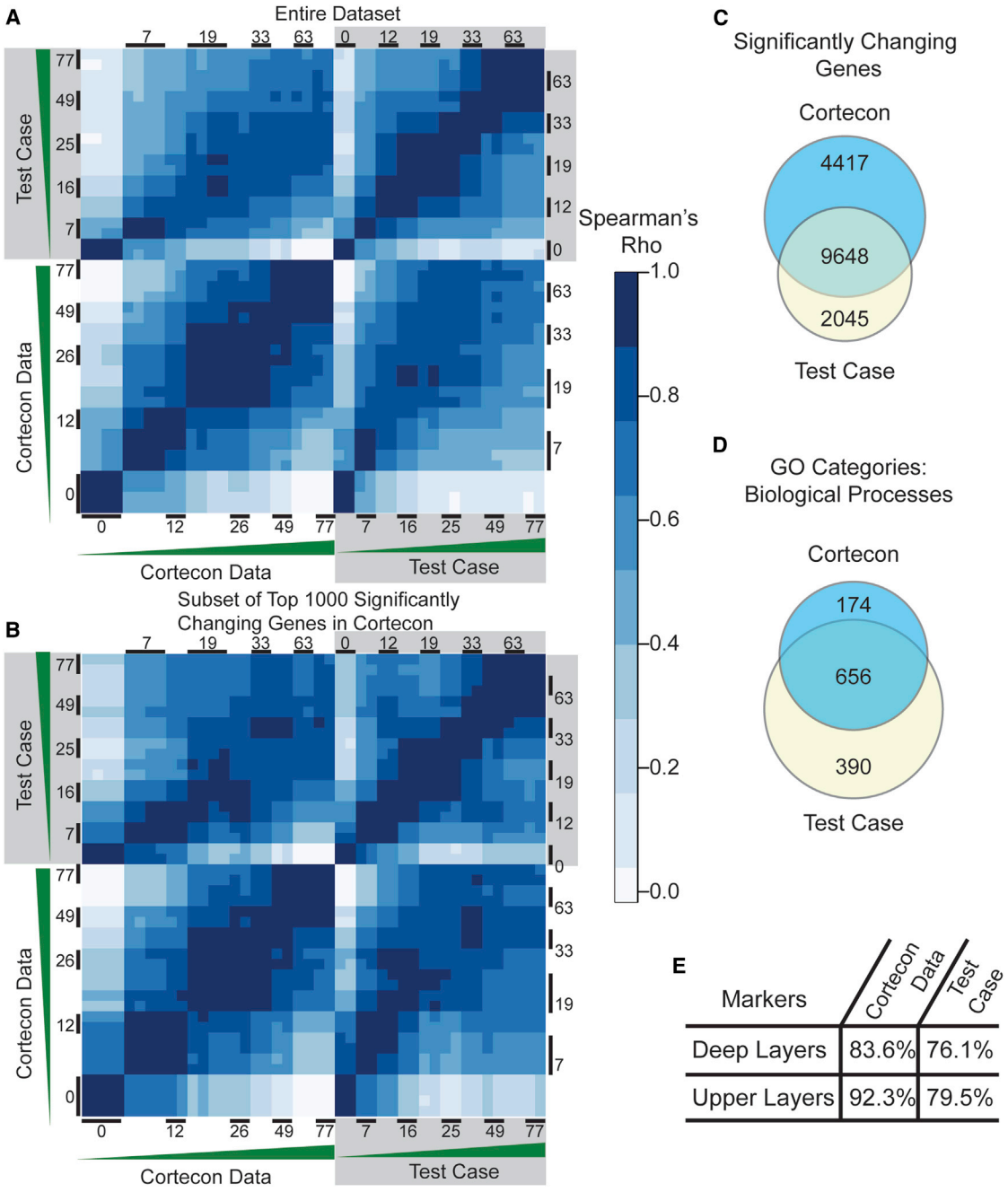


Figure 7. Comparison of Cortecon Data Set to a Modified Cortical Protocol

A separate cortical differentiation was carried out with the starting cells at a reduced confluency of 60%–80% (test case).

(A and B) Spearman's correlation was used to compare the Cortecon data and the test case. Heatmaps of the Spearman's rho values describing the correlation between every sample in the data set (A) using all genes and (B) using the top 1,000 significantly changing genes (the test case sample labels are indicated by a gray background).

(C) A Venn diagram showing the overlap of significantly changing genes between the test case and the Cortecon data.

(D) A Venn diagram showing the overlap of GO categories between the test case and the Cortecon data.

(E) Percentage of layer markers (Zeng et al., 2012) expressed during the time course in the Cortecon data versus the test case. See also Figure S4.

DISCUSSION

Until recently, insights into the molecular determinants of human corticogenesis have primarily been inferred from non-human model systems. Since the first studies describing the differentiation of hESCs into characteristic cortical neurons (Johnson et al., 2007; Kim et al., 2011; Shi et al., 2012), this system is gaining recognition as a valuable "development-in-a-dish" model of

human corticogenesis. Here, we describe a resource "Cortecon" that we have created to help researchers exploring this topic by cataloguing how the transcriptome changes during human in vitro corticogenesis using RNA-seq technology, demonstrating a variety of downstream analyses to show its utility and value, and creating an online, readily searchable website to access the resource.

Through a comparison to BrainSpan RNA-seq data sets, the cortical neurons generated using this protocol were found to be most closely associated with prefrontal cortical regions. This is an important area to model in vitro, given its association with a number of significant diseases and disorders, including schizophrenia (Pomarol-Clotet et al., 2010), autism (Courchesne et al., 2011), and age-related impairment in memory consolidation and retrieval (Mander et al., 2013). While strongly suggestive of a prefrontal cortical identity, this area takes longer to mature than others such as somatosensory or visual cortical areas (Gogtay et al., 2004). Thus, the high correlation to MFC could also be attributed to the late persistence of genes associated with cortical development, which will be interesting to explore further.

Using the Cortecon data set, we can assess mechanisms of cortical development established in other model systems for their role in human. The radial glial marker NES was expressed throughout the period of progenitor cell generation, as expected (Figure S2). In general, neurons are born in order (Shen et al., 2006), as they are in other species, confirming results in humans. For example, Reelin is a marker of early born Cajal-Retsius cells and is an important secreted molecule that guides organization of the mouse cortex (Rice and Curran, 2001). In the Cortecon data set, its expression increases early in the time course, consistent with early-born Cajal-Retsius cell production. Interestingly, this is followed by a much larger increase during the UL stage, consistent with reports in the mouse of a later period of generation of Reelin+ cortical cells (Förster et al., 2010). After neurogenesis, gliogenesis occurs, and consistent with this, the astrocyte marker GFAP greatly increases at the last time point sampled (Figure S2).

FGF and WNT signaling play key roles in mouse corticogenesis, and several of the related genes show similar trends in expression in the Cortecon database to those reported in mouse. For example, FGF8 plays a crucial role in mouse cortical areal patterning (Toyoda et al., 2010) and in generating deep layer neurons (Chalmers et al., 2002). In the Cortecon data set, FG8 expression increases throughout the CS and DL stages supporting a possible involvement during these early stages of human corticogenesis. The FGF receptors (FGFR1, FGFR2, and FGFR3) have been demonstrated to be essential for corticogenesis (Rash et al., 2011), and their expression peaks during the early stages of cortical development (Figure S2). Finally, FGF10, which in mouse controls the differentiation of neuroepithelial cells into radial glia, has very low expression until upper layer neurons are being generated (Sahara and O'Leary, 2009), indicating a different role for FGF10 in human cortical development.

To assess how the WNT pathway components changed, we looked for WNT regulators in the Cortecon data set. WNT7B and WNT8B play key roles in cortex formation in mouse, (Garda et al., 2002), and in the Cortecon data set, their expression increased during cortical specification (Figure S2). In addition to pathway stimulation, WNT antagonism plays a critical role in cortical development; for example, the WNT antagonist DKK1 impacts the early decision for neurectoderm to adopt a telencephalic fate (del Barco Barrantes et al., 2003). Using the Cortecon data set, we assessed the expression profile of two WNT antagonists, DKK1 and FRZB. DKK1 sharply increased during the ND stage, sharply decreased during the CS stage, and peaked again at the UL stage. FRZB showed a similar pattern; however, it had a much greater peak during the UL stage and a much smaller peak during the ND stage (Figure S2).

Exploration of the Cortecon data set from different perspectives will allow researchers to examine hypotheses centered on human cortical development. One of the most interesting discoveries made during this analysis was the catalog of alternatively spliced transcripts present at different stages of cortical development. Interestingly, the splice forms gene sets can consolidate the disease enrichment of a gene cluster into more narrow categories. For example, cluster 31, which fits the pluripotency PP expression profile, had roughly 60% of enriched diseases associated with cancer when looking at all significantly changing genes. However, when focusing on genes demonstrating alternative splicing, the disease enrichment was now almost exclusively cancer related (Figures 3, 4, and 5). These findings are in line with recent evidence showing overlapping mechanisms underlying stem cell reprogramming and tumorigenesis (Bernhardt et al., 2012), indicating that the differentially expressed splice forms in cluster 31 not currently associated with cancer are potential oncogenic candidates for further study. In relation to cortical development, we identified different isoforms expressed at different stages of in vitro corticogenesis for *TGIF1*, which is associated with holoprosencephaly (El-Jaick et al., 2007), and *MARK1*, which has been implicated in ASD (Maussion et al., 2008) (Figure 5). These data suggest that specific splice forms might be more relevant to certain diseases based on the time they are expressed in the developmental program. For example, three splice forms of TGIF1 were expressed specifically at the early stages of cortical development as the cortical field was being specified. It could be that these transcripts are the ones critical for cortical development, and subsequently the progression of holoprosencephaly. Hence, the Cortecon resource will be valuable for those interested in splicing mechanisms and how they contribute to development, to changes in specific gene expressions, and to a variety of neurological diseases.

Researchers can also use this resource to gain insight into developmental pathways by applying network analysis strategies. For example, the developmental autism network created revealed that receptors and ligands associated with ASD are key players at all stages of cortical development. Most notable are genes related to serotonin signaling and GABAergic signaling. Serotonin has been previously shown to be critical in cortical development (Janusonis et al., 2004) and in autism (Anderson et al., 2009; Janusonis et al., 2004). The ASD network constructed using these data shows various serotonin receptors as well as a serotonin-pathway-related enzymes to be important at the PP and UL stages (Figure 6). Interestingly, one hypothesis

for ASD is based on a vital role for serotonin in the development of upper cortical layers 2–4, in which perturbed serotonin signaling results in altered layer formation, possibly contributing to ASD pathology (Altamura et al., 2007). In addition, the autism network indicates several GABA-receptors are involved in different stages of in vitro corticogenesis, with individual subtypes showing distinct temporal expression patterns (Figure 6). Many genes encoding presynaptic and postsynaptic proteins located at GABAergic synapses have been implicated in autism pathology (Delorme et al., 2013), and as indicated by our data set, many of these genes also appear to play a role in corticogenesis.

Application of this network approach enables different clusters to be connected to each other, allowing interrogation of interactions between clusters. For example, genes in cluster 18, which shows an early peak and is largely composed of transcriptional regulators, may regulate the expression of genes in cluster 34, which also shows an early peak but is principally composed of signaling molecules. To test this hypothesis, we examined a network composed of both clusters and observed the transcription factors in cluster 18 were highly connected with the signaling molecules in cluster 34 (Figure S5). In fact, in a network constructed from the genes in cluster 34 and the transcription factors from cluster 18, we found that a cluster 18 transcription factor (*SOX5*) had the highest betweenness and closeness centrality scores in the network, and this corroborates the critical role of Sox5 in mouse cortical development (Lai et al., 2008). Moreover, *SOX5* is directly linked to Myocyte Enhancer Factor 2C (*MEF2C*), which is a transcription factor belonging to cluster 34 and has the second highest betweenness and closeness centrality scores in this network. Other cluster 18 transcription factors are similarly linked to cluster 34 genes. This underscores the potential importance of cluster 18 in the correct expression of genes in cluster 34. Using this network approach to search for temporally relevant regulators of each cluster can provide insight into how each cluster is connected and regulated and focus attention on particular pathways to uncover mechanisms underlying human cortical development in health and disease.

EXPERIMENTAL PROCEDURES

Human ESC Differentiation

Human ESCs (WA-09, WiCell) were maintained on irradiated mouse embryonic fibroblasts (GobalStem) according to standard culture procedures. In some cases, hESCS were maintained using StemBeads FGF2 (StemCulture). Induction of neural progenitors was initiated using an adaptation of the dual SMAD-inhibition protocol (Chambers et al., 2009). Cyclopamine was added from day 3 of neural induction to ensure an anterior cortical fate (Gaspard et al., 2008). Throughout the entire subsequent cortical differentiation, cultures were maintained in N2B27-medium (Gaspard et al., 2008) supplemented with 10 ng/ml FGF-2. See Supplemental Experimental Procedures.

Quantitative RT-PCR

Quantitative RT-PCR was performed using TaqMan gene expression assays (Life Technologies); see Supplemental Experimental Procedures.

Immunocytochemistry

Immunocytochemistry was carried out using standard procedures; see Supplemental Experimental Procedures.

RNA-Seq Application

Samples were collected using RNAprotect cell reagent (QIAGEN), and total RNA was extracted using RNeasy plus mini kit (QIAGEN), according to the manufacturer's protocol. RNA samples were subject to RNA-seq by Expression Analysis (Durham) or Covance (Seattle): TruSeq RNA sample prep kit (v2), HiSeq2000 platform (Illumina), paired-end, 50 cycles, multiplex (EA 4-plex; Covance 5-plex).

RNA-Seq Data Analysis

FASTQ files were mapped to the hg19 assembly via BFAST (Homer et al., 2009) and analyzed using R (Team, 2012), an open source programming environment for statistical computing and graphics, and a variety of packages available through Bioconductor (Gentleman et al., 2004), an open development project. FastQC was used to assess the quality of the sequencing (Table S18). All code and R sessions are provided for download from http://cortecon.neuralsci.org/. Files can also be downloaded from GEO: GSE56796. See Supplemental Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five figures, eighteen tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2014.05.013.

AUTHOR CONTRIBUTIONS

J.v.d.L. performed cell culture assays, prepared RNA for sample processing, performed immunocytochemistry and qRT-PCR, assisted with data interpretations, and helped write the manuscript. N.C.B. performed all computational analyses, assisted with data interpretations, and helped write the manuscript. T.R.K. assisted with computational analyses, designed the online search tool, and helped with data interpretation and manuscript preparation. B.C. performed cell culture assays. P.L. assisted with cell culture assays and immunocytochemistry. R.M. and C.L. assisted with sample preparation and analysis, respectively. V.M., B.P.L., C.T., and S.Y. assisted with overseeing experiments and data interpretations and helped write the manuscript. S.T. conceived the project; helped establish the collaboration; and also helped oversee the experiments, data interpretation, and manuscript writing. C.A.F. conceived differentiation protocols and helped oversee the experiments, data interpretations, and writing of the manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from NINDS (NS072434-01A1 to C.A.F.) and the Regenerative Research Foundation. This collaborative work was supported in part by the Allen Institute for Brain Science. Its authors wish to thank the Allen Institute founders, Paul G. Allen and Jody Allen, for their vision, encouragement, and support.

Accepted: May 6, 2014
Published: July 2, 2014

REFERENCES

Altamura, C., Dell'Acqua, M.L., Moessner, R., Murphy, D.L., Lesch, K.P., and Persico, A.M. (2007). Altered neocortical cell density and layer thickness in serotonin transporter knockout mice: a quantitation study. Cereb. Cortex *17*, 1394–1401.

Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome Res. *22*, 2008–2017.

Anders, S., McCarthy, D.J., Chen, Y., Okoniewski, M., Smyth, G.K., Huber, W., and Robinson, M.D. (2013). Count-based differential expression analysis of RNA sequencing data using R and Bioconductor. Nat. Protoc. *8*, 1765–1786.

Anderson, B.M., Schnetz-Boutaud, N.C., Bartlett, J., Wotawa, A.M., Wright, H.H., Abramson, R.K., Cuccaro, M.L., Gilbert, J.R., Pericak-Vance, M.A.,

and Haines, J.L. (2009). Examination of association of genes in the serotonin system to autism. Neurogenetics *10*, 209–216.

Belgard, T.G., Marques, A.C., Oliver, P.L., Abaan, H.O., Sirey, T.M., Hoerder-Suabedissen, A., García-Moreno, F., Molnár, Z., Margulies, E.H., and Ponting, C.P. (2011). A transcriptomic atlas of mouse neocortical layers. Neuron *71*, 605–616.

Bernhardt, M., Galach, M., Novak, D., and Utikal, J. (2012). Mediators of induced pluripotency and their role in cancer cells - current scientific knowledge and future perspectives. Biotechnol. J. *7*, 810–821.

Boles, N.C., Hirsch, S.E., Le, S., Corneo, B., Najm, F., Minotti, A.P., Wang, Q., Lotz, S., Tesar, P.J., and Fasano, C.A. (2014). NPTX1 Regulates Neural Lineage Specification from Human Pluripotent Stem Cells. Cell Rep. *6*, 724–736.

Caviness, V.S., Jr., Takahashi, T., and Nowakowski, R.S. (1995). Numbers, time and neocortical neuronogenesis: a general developmental and evolutionary model. Trends Neurosci. *18*, 379–383.

Cecchi, C. (2002). Emx2: a gene responsible for cortical development, regionalization and area specification. Gene *291*, 1–9.

Chalmers, A.D., Welchman, D., and Papalopulu, N. (2002). Intrinsic differences between the superficial and deep layers of the Xenopus ectoderm control primary neuronal differentiation. Dev. Cell *2*, 171–182.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Courchesne, E., Mouton, P.R., Calhoun, M.E., Semendeferi, K., Ahrens-Barbeau, C., Hallet, M.J., Barnes, C.C., and Pierce, K. (2011). Neuron number and size in prefrontal cortex of children with autism. JAMA *306*, 2001–2010.

Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems *1695*.

del Barco Barrantes, I., Davidson, G., Gröne, H.J., Westphal, H., and Niehrs, C. (2003). Dkk1 and noggin cooperate in mammalian head induction. Genes Dev. *17*, 2239–2244.

Delorme, R., Ey, E., Toro, R., Leboyer, M., Gillberg, C., and Bourgeron, T. (2013). Progress toward treatments for synaptic defects in autism. Nat. Med. *19*, 685–694.

Dillman, A.A., Hauser, D.N., Gibbs, J.R., Nalls, M.A., McCoy, M.K., Rudenko, I.N., Galter, D., and Cookson, M.R. (2013). mRNA expression, splicing and editing in the embryonic and adult mouse cerebral cortex. Nat. Neurosci. *16*, 499–506.

Dougherty, J.D., and Geschwind, D.H. (2005). Progress in realizing the promise of microarrays in systems neurobiology. Neuron *45*, 183–185.

El-Jaick, K.B., Powers, S.E., Bartholin, L., Myers, K.R., Hahn, J., Orioli, I.M., Ouspenskaia, M., Lacbawan, F., Roessler, E., Wotton, D., and Muenke, M. (2007). Functional analysis of mutations in TGIF associated with holoprosencephaly. Mol. Genet. Metab. *90*, 97–111.

Fietz, S.A., Lachmann, R., Brandl, H., Kircher, M., Samusik, N., Schröder, R., Lakshmanaperumal, N., Henry, I., Vogt, J., Riehn, A., et al. (2012). Transcriptomes of germinal zones of human and mouse fetal neocortex suggest a role of extracellular matrix in progenitor self-renewal. Proc. Natl. Acad. Sci. USA *109*, 11836–11841.

Förster, E., Bock, H.H., Herz, J., Chai, X., Frotscher, M., and Zhao, S. (2010). Emerging topics in Reelin function. Eur. J. Neurosci. *31*, 1511–1518.

Frankild, S., and Jensen, L.J. (2013). DISEASES: Disease-gene associations mined from literature. http://diseases.jensenlab.org/Search.

Garda, A.L., Puelles, L., Rubenstein, J.L., and Medina, L. (2002). Expression patterns of Wnt8b and Wnt7b in the chicken embryonic brain suggest a correlation with forebrain patterning centers and morphogenesis. Neuroscience *113*, 689–698.

Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N., et al. (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature *455*, 351–357.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., et al. (2004). Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. *5*, R80.

Gogtay, N., Giedd, J.N., Lusk, L., Hayashi, K.M., Greenstein, D., Vaituzis, A.C., Nugent, T.F., 3rd, Herman, D.H., Clasen, L.S., Toga, A.W., et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci. USA *101*, 8174–8179.

Halfon, N., and Kuo, A.A. (2013). What DSM-5 could mean to children with autism and their families. JAMA Pediatr. *167*, 608–613.

Hamid, R., Patterson, J., and Brandt, S.J. (2008). Genomic structure, alternative splicing and expression of TG-interacting factor, in human myeloid leukemia blasts and cell lines. Biochim. Biophys. Acta *1779*, 347–355.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., et al. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. Nature *489*, 391–399.

Hébert, J.M., and Fishell, G. (2008). The genetics of early telencephalon patterning: some assembly required. Nat. Rev. Neurosci. *9*, 678–685.

Hevner, R.F., Shi, L., Justice, N., Hsueh, Y., Sheng, M., Smiga, S., Bulfone, A., Goffinet, A.M., Campagnoni, A.T., and Rubenstein, J.L. (2001). Tbr1 regulates differentiation of the preplate and layer 6. Neuron *29*, 353–366.

Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. PLoS ONE *4*, e7767.

Hou, P.S., Chuang, C.Y., Kao, C.F., Chou, S.J., Stone, L., Ho, H.N., Chien, C.L., and Kuo, H.C. (2013). LHX2 regulates the neural differentiation of human embryonic stem cells via transcriptional modulation of PAX6 and CER1. Nucleic Acids Res. *41*, 7753–7770.

Hubbard, K.S., Gut, I.M., Lyman, M.E., and McNutt, P.M. (2013). Longitudinal RNA sequencing of the deep transcriptome during neurogenesis of cortical glutamatergic neurons from murine ESCs. F1000Res. *2*, 35.

Janusonis, S., Gluncic, V., and Rakic, P. (2004). Early serotonergic projections to Cajal-Retzius cells: relevance for cortical development. J. Neurosci. *24*, 1652–1659.

Johnson, M.A., Weick, J.P., Pearce, R.A., and Zhang, S.C. (2007). Functional neural development from human embryonic stem cells: accelerated synaptic activity via astrocyte coculture. J. Neurosci. *27*, 3069–3077.

Jones, E.G., and Rakic, P. (2010). Radial columns in cortical architecture: it is the composition that counts. Cereb. Cortex *20*, 2261–2264.

Kaneko, T., and Fujiyama, F. (2002). Complementary distribution of vesicular glutamate transporters in the central nervous system. Neurosci. Res. *42*, 243–250.

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M., Pletikos, M., Meyer, K.A., Sedmak, G., et al. (2011). Spatio-temporal transcriptome of the human brain. Nature *478*, 483–489.

Kim, J.E., O'Sullivan, M.L., Sanchez, C.A., Hwang, M., Israel, M.A., Brennand, K., Deerinck, T.J., Goldstein, L.S., Gage, F.H., Ellisman, M.H., and Ghosh, A. (2011). Investigating synapse formation and function using human pluripotent stem cell-derived neurons. Proc. Natl. Acad. Sci. USA *108*, 3005–3010.

Kumar, L., and E Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5–7.

Lai, T., Jabaudon, D., Molyneaux, B.J., Azim, E., Arlotta, P., Menezes, J.R., and Macklis, J.D. (2008). SOX5 controls the sequential generation of distinct corticofugal neuron subtypes. Neuron *57*, 232–247.

Lee, A.S., Tang, C., Rao, M.S., Weissman, I.L., and Wu, J.C. (2013). Tumorigenicity as a clinical hurdle for pluripotent stem cell therapies. Nat. Med. *19*, 998–1004.

Mander, B.A., Rao, V., Lu, B., Saletin, J.M., Lindquist, J.R., Ancoli-Israel, S., Jagust, W., and Walker, M.P. (2013). Prefrontal atrophy, disrupted NREM slow waves and impaired hippocampal-dependent memory in aging. Nat. Neurosci. *16*, 357–364.

Matsumoto, M., Weickert, C.S., Akil, M., Lipska, B.K., Hyde, T.M., Herman, M.M., Kleinman, J.E., and Weinberger, D.R. (2003). Catechol O-methyltransferase mRNA expression in human and rat brain: evidence for a role in cortical neuronal function. Neuroscience *116*, 127–137.

Maussion, G., Carayol, J., Lepagnol-Bestel, A.M., Tores, F., Loe-Mie, Y., Milbreta, U., Rousseau, F., Fontaine, K., Renaud, J., Moalic, J.M., et al. (2008). Convergent evidence identifying MAP/microtubule affinity-regulating kinase 1 (MARK1) as a susceptibility gene for autism. Hum. Mol. Genet. *17*, 2541–2551.

Miller, J.A., Ding, S.L., Sunkin, S.M., Smith, K.A., Ng, L., Szafer, A., Ebbert, A., Riley, Z.L., Royall, J.J., Aiona, K., et al. (2014). Transcriptional landscape of the prenatal human brain. Nature *508*, 199–206.

Padgett, R.A. (2012). New connections between splicing and human disease. Trends Genet. *28*, 147–154.

Pomarol-Clotet, E., Canales-Rodríguez, E.J., Salvador, R., Sarró, S., Gomar, J.J., Vila, F., Ortiz-Gil, J., Iturria-Medina, Y., Capdevila, A., and McKenna, P.J. (2010). Medial prefrontal cortex pathology in schizophrenia as revealed by convergent findings from multimodal imaging. Mol. Psychiatry *15*, 823–830.

Rash, B.G., Lim, H.D., Breunig, J.J., and Vaccarino, F.M. (2011). FGF signaling expands embryonic cortical surface area by regulating Notch-dependent neurogenesis. J. Neurosci. *31*, 15604–15617.

Rice, D.S., and Curran, T. (2001). Role of the reelin signaling pathway in central nervous system development. Annu. Rev. Neurosci. *24*, 1005–1039.

Sahara, S., and O'Leary, D.D. (2009). Fgf10 regulates transition period of cortical stem cell differentiation to radial glia controlling generation of neurons and basal progenitors. Neuron *63*, 48–62.

Scott, M.M., and Deneris, E.S. (2005). Making and breaking serotonin neurons and autism. Int. J. Dev. Neurosci. *23*, 277–285.

Shen, Q., Wang, Y., Dimos, J.T., Fasano, C.A., Phoenix, T.N., Lemischka, I.R., Ivanova, N.B., Stifani, S., Morrisey, E.E., and Temple, S. (2006). The timing of cortical neurogenesis is encoded within lineages of individual progenitor cells. Nat. Neurosci. *9*, 743–751.

Shi, Y., Kirwan, P., Smith, J., Robinson, H.P., and Livesey, F.J. (2012). Human cerebral cortex development from pluripotent stem cells to functional excitatory synapses. Nat. Neurosci. *15*, 477–486.

Takahashi, T., Nowakowski, R.S., and Caviness, V.S., Jr. (1996). The leaving or Q fraction of the murine cerebral proliferative epithelium: a general model of neocortical neuronogenesis. J. Neurosci. *16*, 6183–6196.

Takeuchi, A., Hamasaki, T., Litwack, E.D., and O'Leary, D.D. (2007). Novel IgCAM, MDGA1, expressed in unique cortical area- and layer-specific patterns and transiently by distinct forebrain populations of Cajal-Retzius neurons. Cereb. Cortex *17*, 1531–1541.

Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., and Romero, R. (2009). A novel signaling pathway impact analysis. Bioinformatics *25*, 75–82.

Team, R.C. (2012). R: A Language and Environment for Statistical Computing. (Vienna: R Foundation for Statistical Computing).

Toyoda, R., Assimacopoulos, S., Wilcoxon, J., Taylor, A., Feldman, P., Suzuki-Hirano, A., Shimogori, T., and Grove, E.A. (2010). FGF8 acts as a classic diffusible morphogen to pattern the neocortex. Development *137*, 3439–3448.

Vitalis, T., Cases, O., Passemard, S., Callebert, J., and Parnavelas, J.G. (2007). Embryonic depletion of serotonin affects cortical development. Eur. J. Neurosci. *26*, 331–344.

Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome Biol. *11*, R14.

Zeng, H., Shen, E.H., Hohmann, J.G., Oh, S.W., Bernard, A., Royall, J.J., Glattfelder, K.J., Sunkin, S.M., Morris, J.A., Guillozet-Bongaarts, A.L., et al. (2012). Large-scale cellular-resolution gene profiling in human neocortex reveals species-specific molecular signatures. Cell *149*, 483–496.

Zhang, X., Huang, C.T., Chen, J., Pankratz, M.T., Xi, J., Li, J., Yang, Y., Lavaute, T.M., Li, X.J., Ayala, M., et al. (2010). Pax6 is a human neuroectoderm cell fate determinant. Cell Stem Cell *7*, 90–100.

Zhi-Liang, H., B, J., and R, J.M. (2008). CateGOrizer: A Web-Based Program to Batch Analyze Gene Ontology Classification Categories. Online J. Bioinformatics *9*, 108–112.

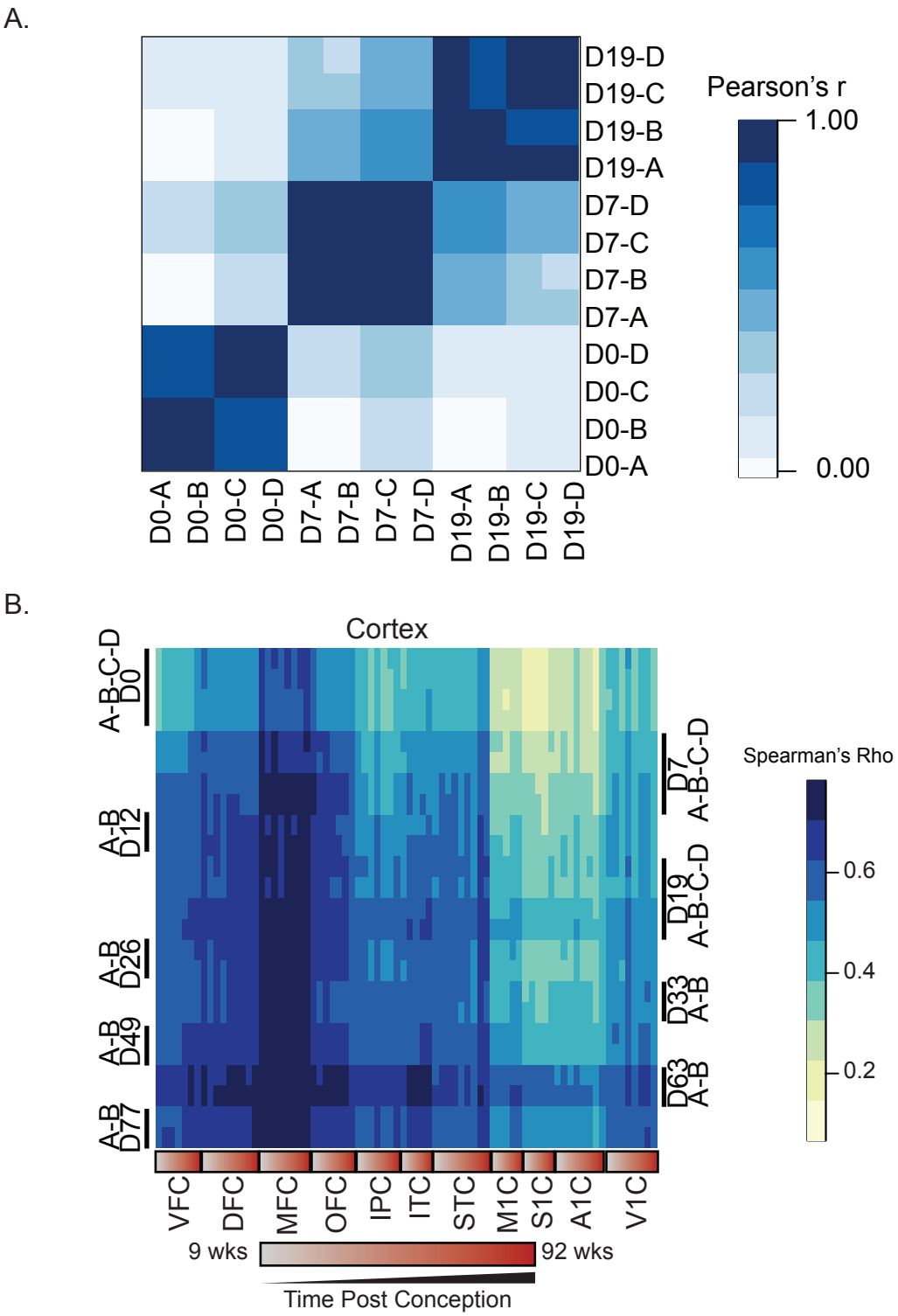
Zweifel, L.S., Kuruvilla, R., and Ginty, D.D. (2005). Functions and mechanisms of retrograde neurotrophin signalling. Nat. Rev. Neurosci. *6*, 615–625.

Neuron, Volume *83*

Supplemental Information

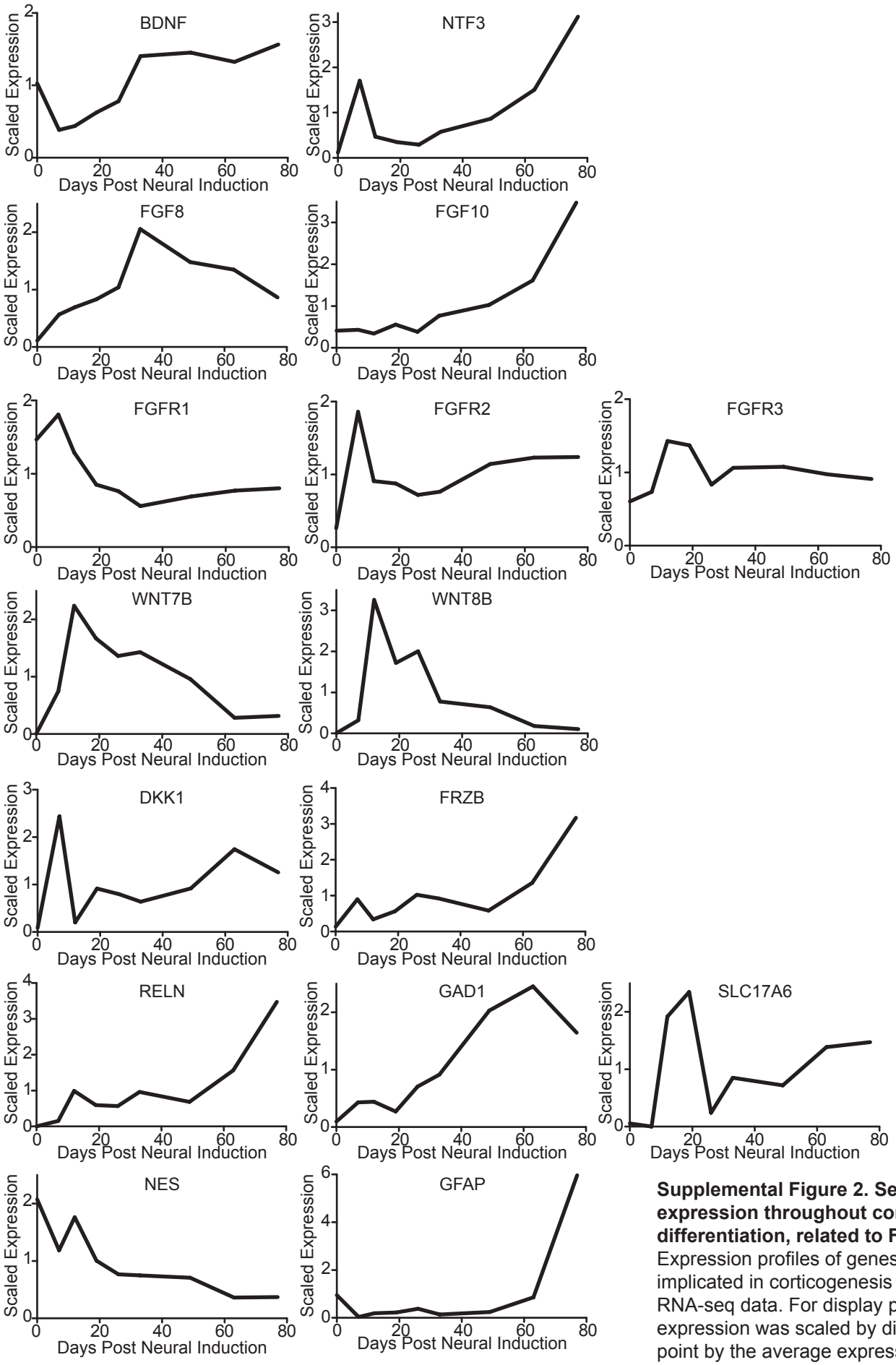
CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells

Joyce van de Leemput, Nathan C. Boles, Thomas R. Kiehl, Barbara Corneo, Patty Lederman, Vilas Menon, Changkyu Lee, Refugio A. Martinez, Boaz P. Levi, Carol L. Thompson, Shuyuan Yao, Ajamete Kaykas, Sally Temple, and Christopher A. Fasano



Supplementary Figure 1. Correlations of Cortecon Replicates and Cortecon vs. Brainspan cortical regions, related to Figure 1.

A.) Pearson's r value between expression profiles from samples (day 0, 7, 19) generated in the two separate cortical differentiation experiments: shorter, 26-days (C, D) and longer, 77-days (A, B). (B.) Comparative analysis of RNA-Seq data with BrainSpan Atlas of the Developing Human Brain of transcriptome profiles from macrodissected and laser-microdissected human fetal brain regions. Human sample ages range from as early as 9 postconceptual weeks to 1 year old. Correlation with specific cortical regions. Ventrolateral Prefrontal Cortex (VFC); Dorsolateral Prefrontal Cortex (DFC); Medial Prefrontal Cortex (MFC); Orbital Frontal Cortex (OFC); Posteroventral (inferior) parietal cortex (IPC); Posterior (caudal) Superior Temporal Cortex (STC); Inferolateral Temporal Cortex (ITC); Primary Motor Cortex (M1C); Primary Somatosensory Cortex (S1C); Primary Auditory Cortex (A1C); Primary Visual Cortex (V1C).

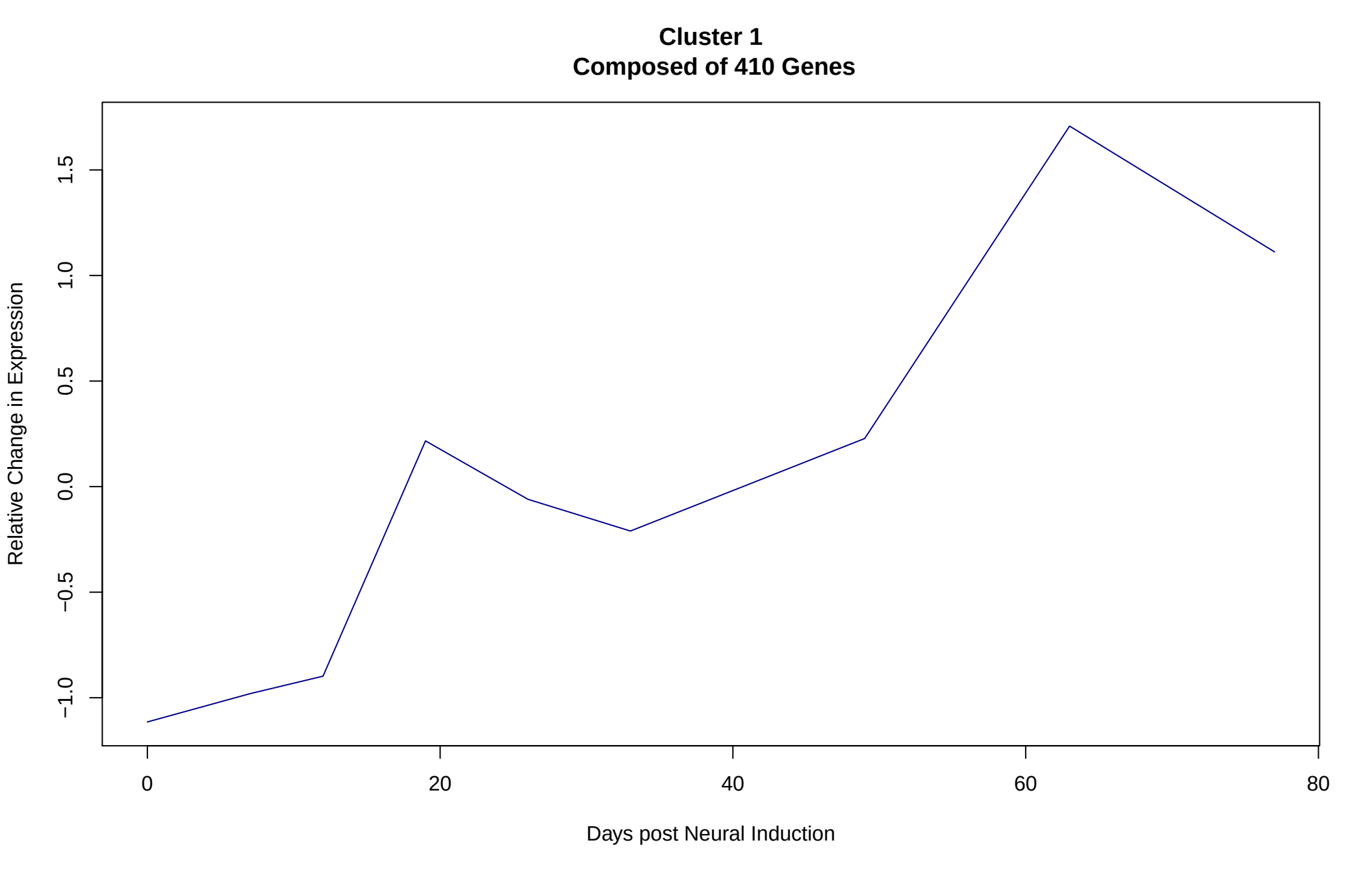


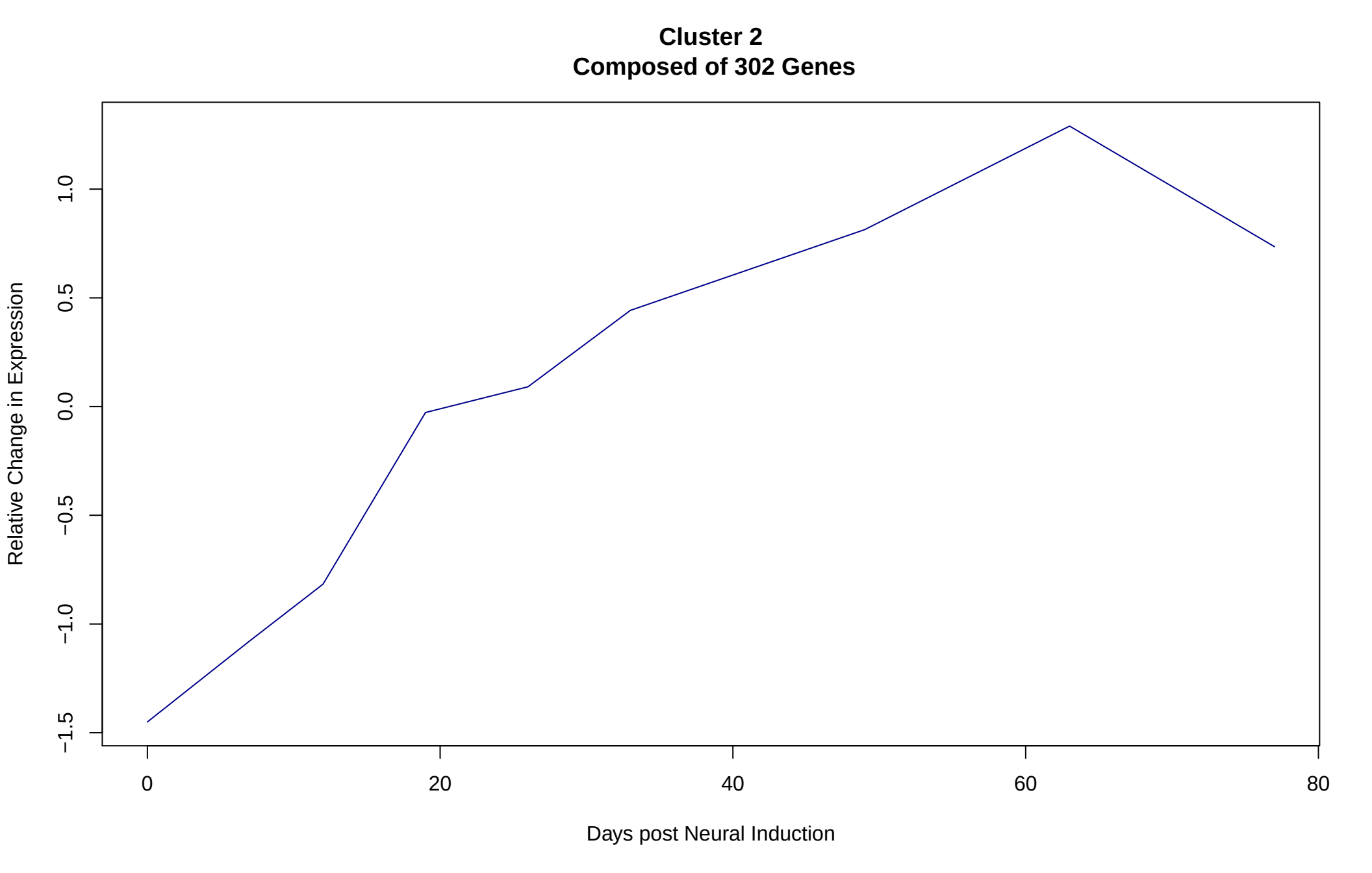
Supplemental Figure 2. Select gene expression throughout cortical differentiation, related to Figure 2. Expression profiles of genes previously implicated in corticogenesis from the RNA-seq data. For display purposes gene expression was scaled by dividing each time point by the average expression over the entire time course.

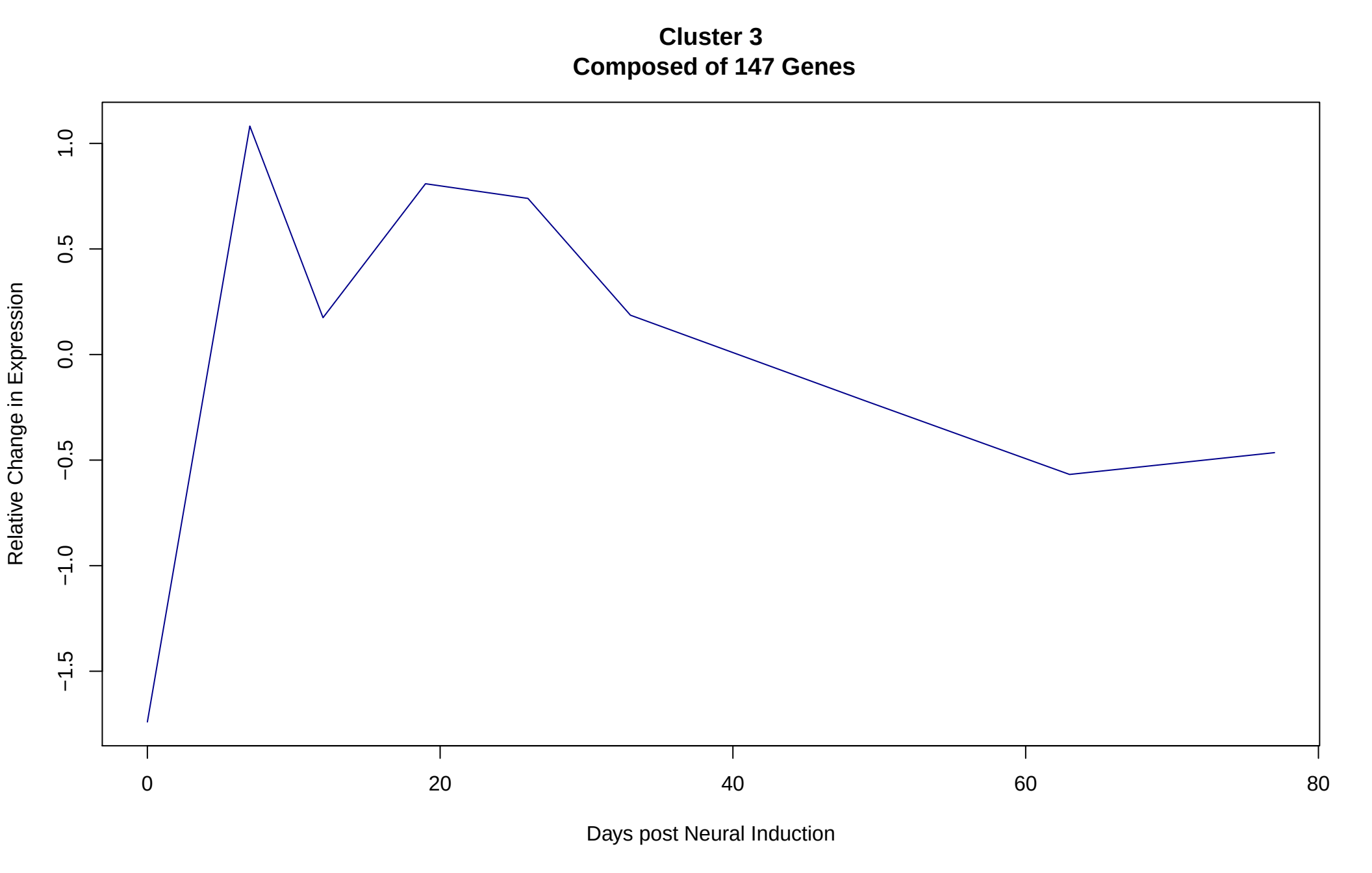
Stage Membership of each Cluster												
Cluster	PP	ND	CS	DL	UL		Cluster	PP	ND	CS	DL	UL
1	Absent	Absent	Present	Absent	Present		33	Absent	Present	Present	Present	Absent
2	Absent	Present	Present	Present	Present		34	Absent	Present	Present	Absent	Absent
3	Absent	Present	Present	Present	Absent		35	Present	Absent	Absent	Absent	Present
4	Absent	Present	Present	Present	Present		36	Absent	Present	Present	Present	Absent
5	Present	Absent	Absent	Absent	Absent		37	Absent	Present	Absent	Absent	Absent
6	Present	Absent	Absent	Present	Absent		38	Absent	Present	Present	Present	Present
7	Present	Absent	Absent	Absent	Present		39	Absent	Absent	Absent	Absent	Present
8	Absent	Absent	Present	Absent	Present		40	Absent	Present	Present	Absent	Present
9	Present	Absent	Present	Absent	Absent		41	Absent	Present	Absent	Absent	Absent
10	Present	Absent	Present	Absent	Absent		42	Present	Absent	Present	Absent	Absent
11	Absent	Absent	Absent	Present	Absent		43	Absent	Present	Present	Present	Present
12	Absent	Present	Present	Absent	Absent		44	Absent	Present	Present	Present	Absent
13	Absent	Present	Absent	Absent	Present		45	Absent	Absent	Present	Present	Present
14	Absent	Absent	Present	Absent	Absent		46	Absent	Absent	Present	Present	Absent
15	Present	Absent	Absent	Absent	Absent		47	Present	Absent	Absent	Present	Absent
16	Absent	Present	Absent	Present	Present		48	Absent	Present	Absent	Absent	Absent
17	Absent	Absent	Present	Present	Present		49	Absent	Present	Present	Present	Absent
18	Absent	Present	Present	Absent	Absent		50	Present	Absent	Absent	Absent	Absent
19	Absent	Present	Absent	Present	Present		51	Present	Absent	Present	Present	Absent
20	Present	Absent	Absent	Present	Absent		52	Absent	Present	Absent	Absent	Absent
21	Absent	Present	Present	Present	Absent		53	Present	Absent	Present	Present	Present
22	Present	Absent	Absent	Absent	Absent		54	Absent	Absent	Present	Present	Absent
23	Absent	Present	Absent	Present	Present		55	Absent	Present	Present	Present	Absent
24	Absent	Absent	Present	Present	Present		56	Absent	Present	Present	Absent	Present
25	Absent	Present	Absent	Present	Absent		57	Present	Absent	Present	Present	Present
26	Present	Absent	Present	Present	Absent		58	Absent	Absent	Absent	Absent	Present
27	Present	Absent	Absent	Absent	Absent		59	Present	Absent	Absent	Present	Absent
28	Absent	Absent	Present	Absent	Present		60	Present	Absent	Absent	Present	Absent
29	Absent	Present	Absent	Absent	Present		61	Absent	Absent	Present	Absent	Present
30	Absent	Absent	Present	Present	Present		62	Absent	Absent	Present	Absent	Present
31	Present	Absent	Absent	Absent	Absent		63	Present	Absent	Present	Absent	Absent
32	Absent	Present	Present	Present	Absent		64	Absent	Present	Absent	Absent	Absent

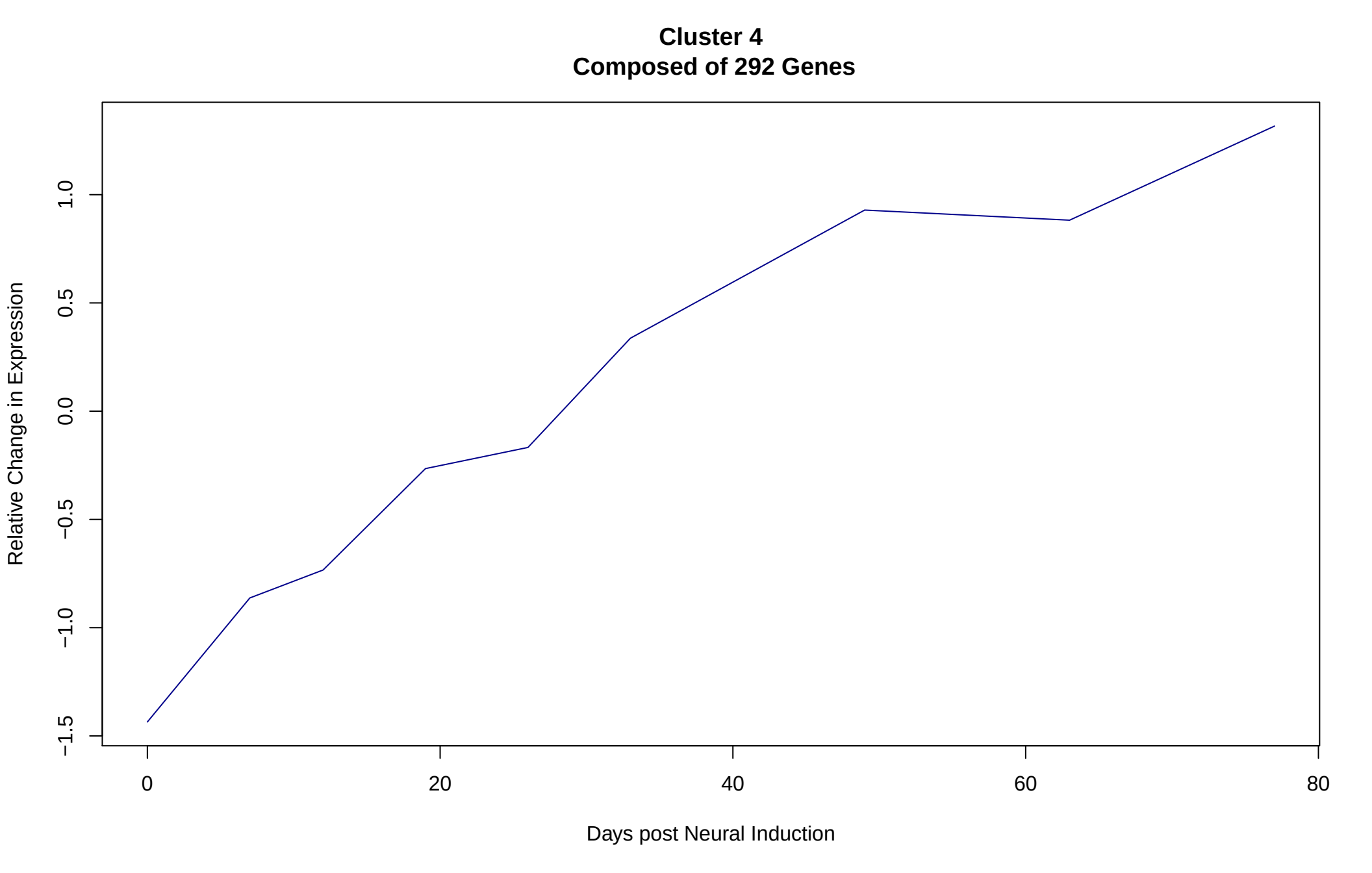
Figure S3. Temporal expression profile-derived clusters.

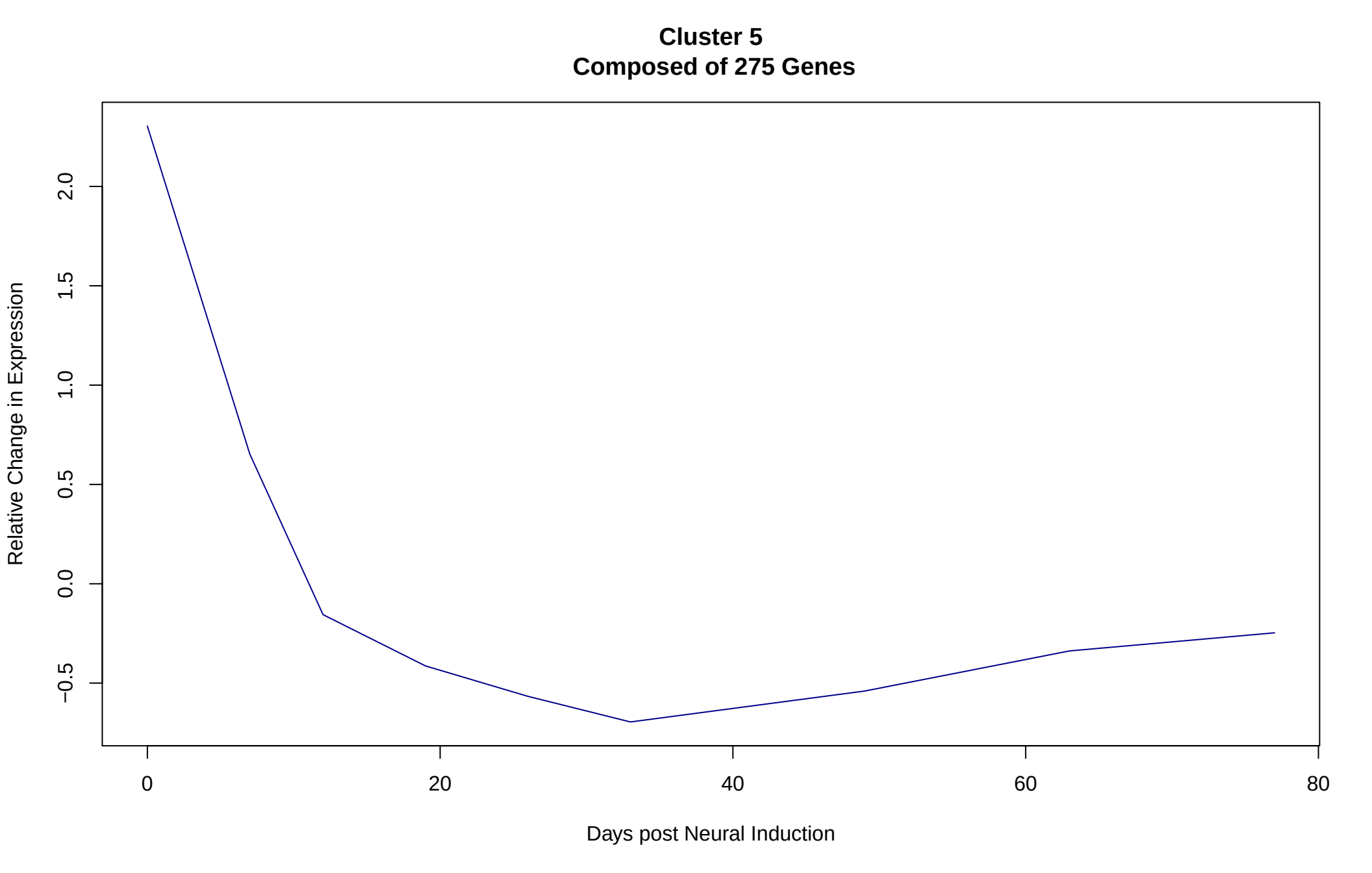
Significantly changing genes were clustered by c-means clustering based on their expression across the time course. Shown above are the stage memberships of each cluster, and following are graphs of each cluster's shape.

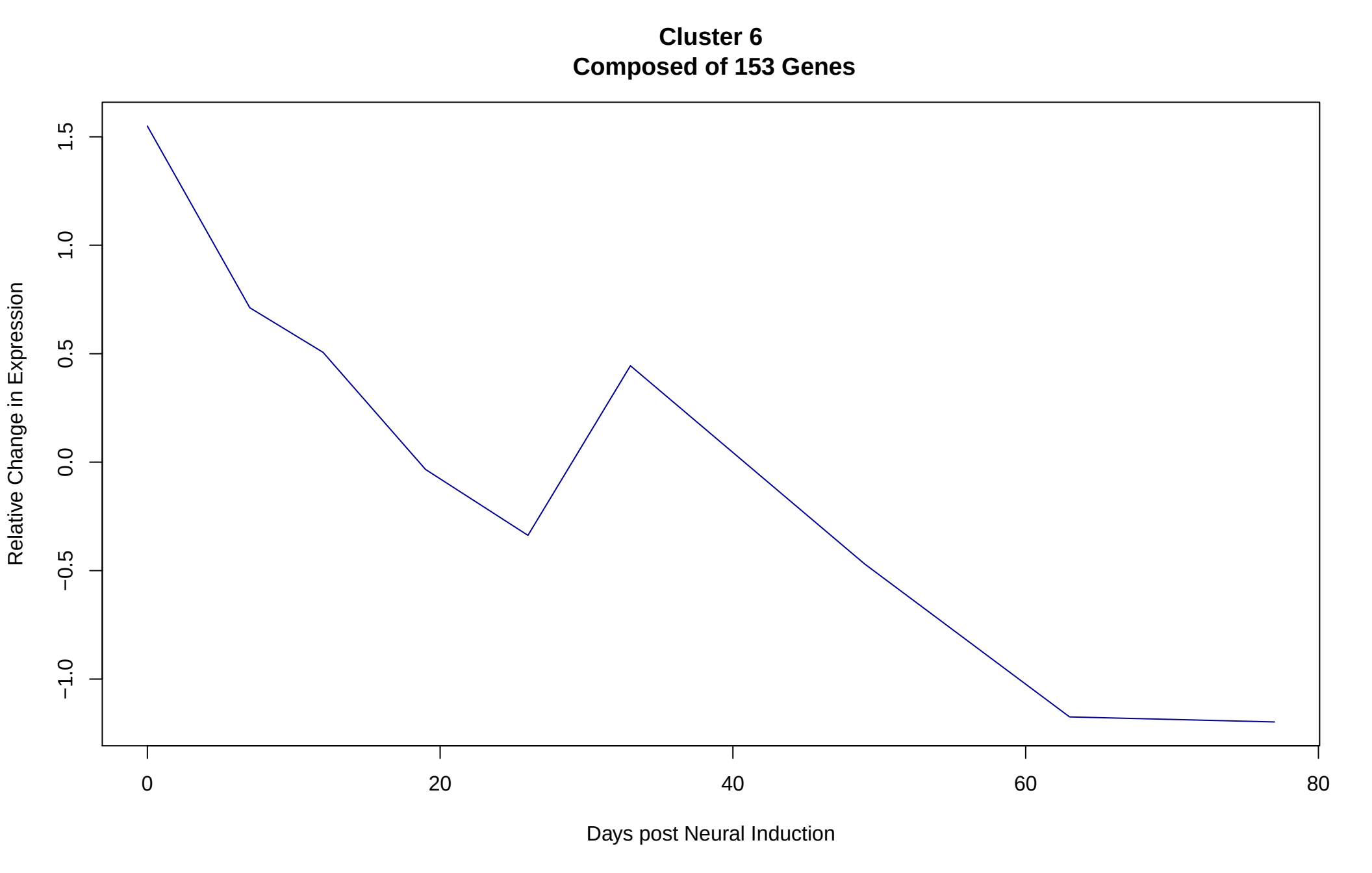

Cluster 1
Composed of 410 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

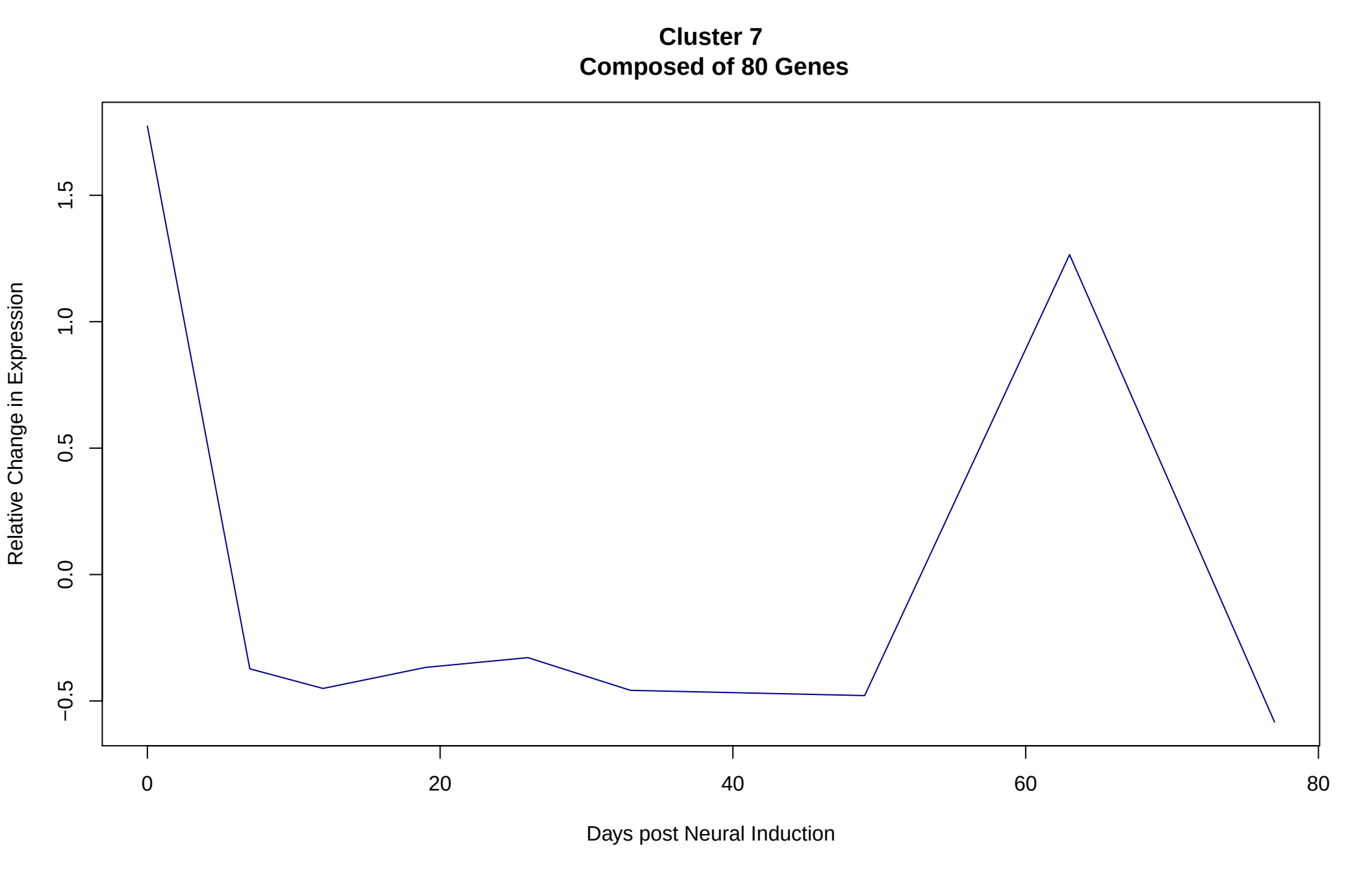

Cluster 2
Composed of 302 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

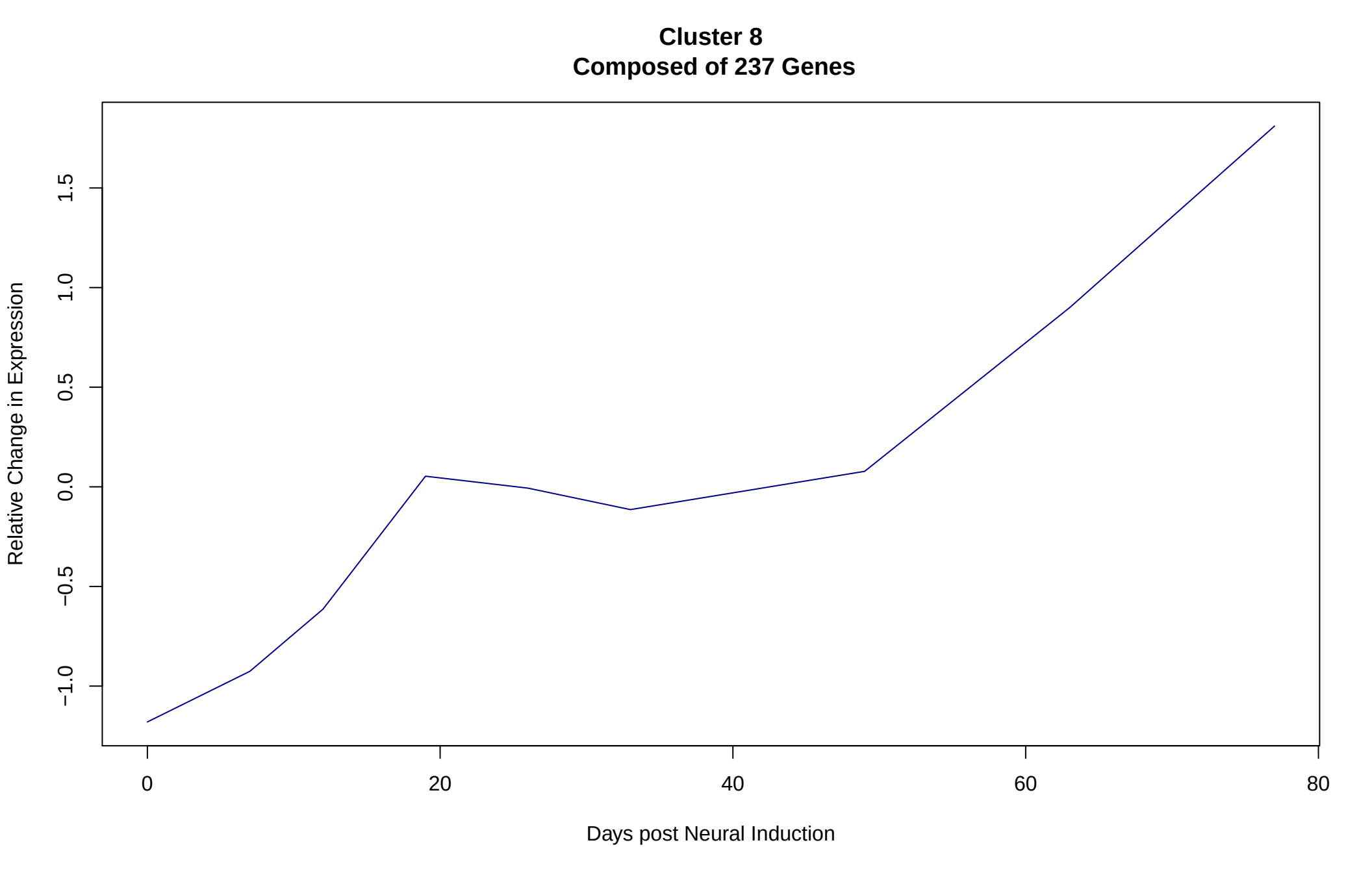

Cluster 3
Composed of 147 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

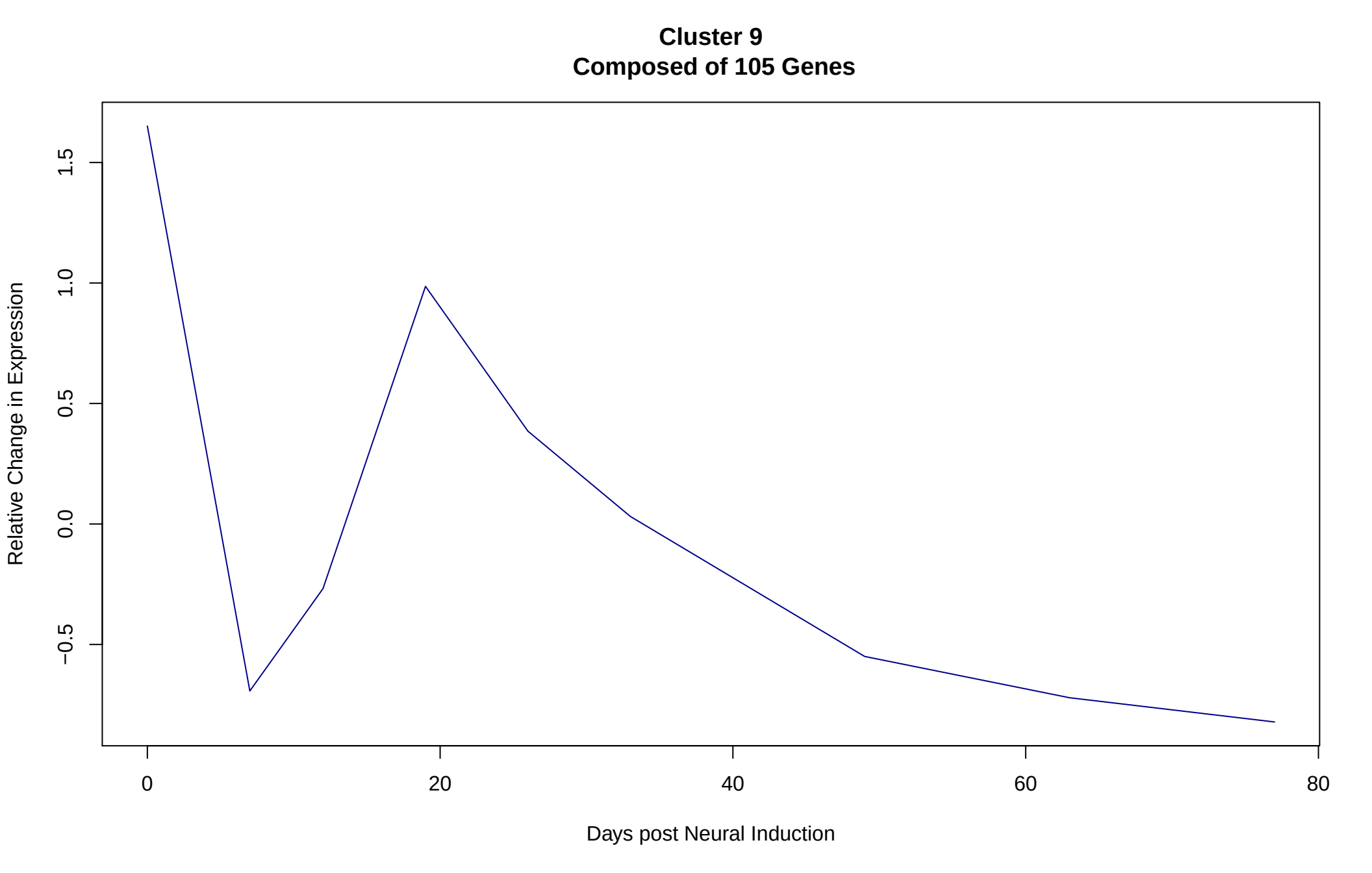

Cluster 4
Composed of 292 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

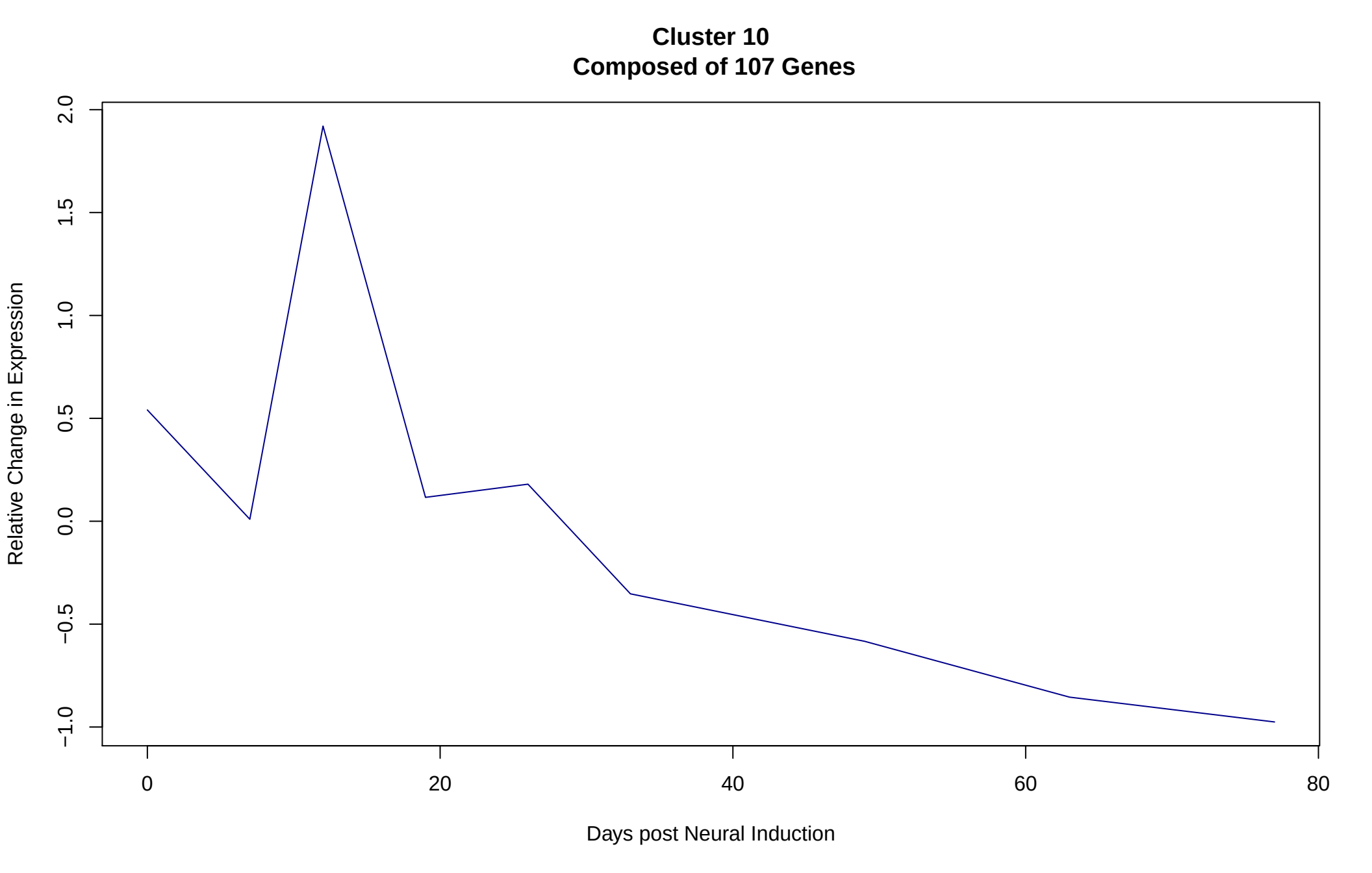

Cluster 5
Composed of 275 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

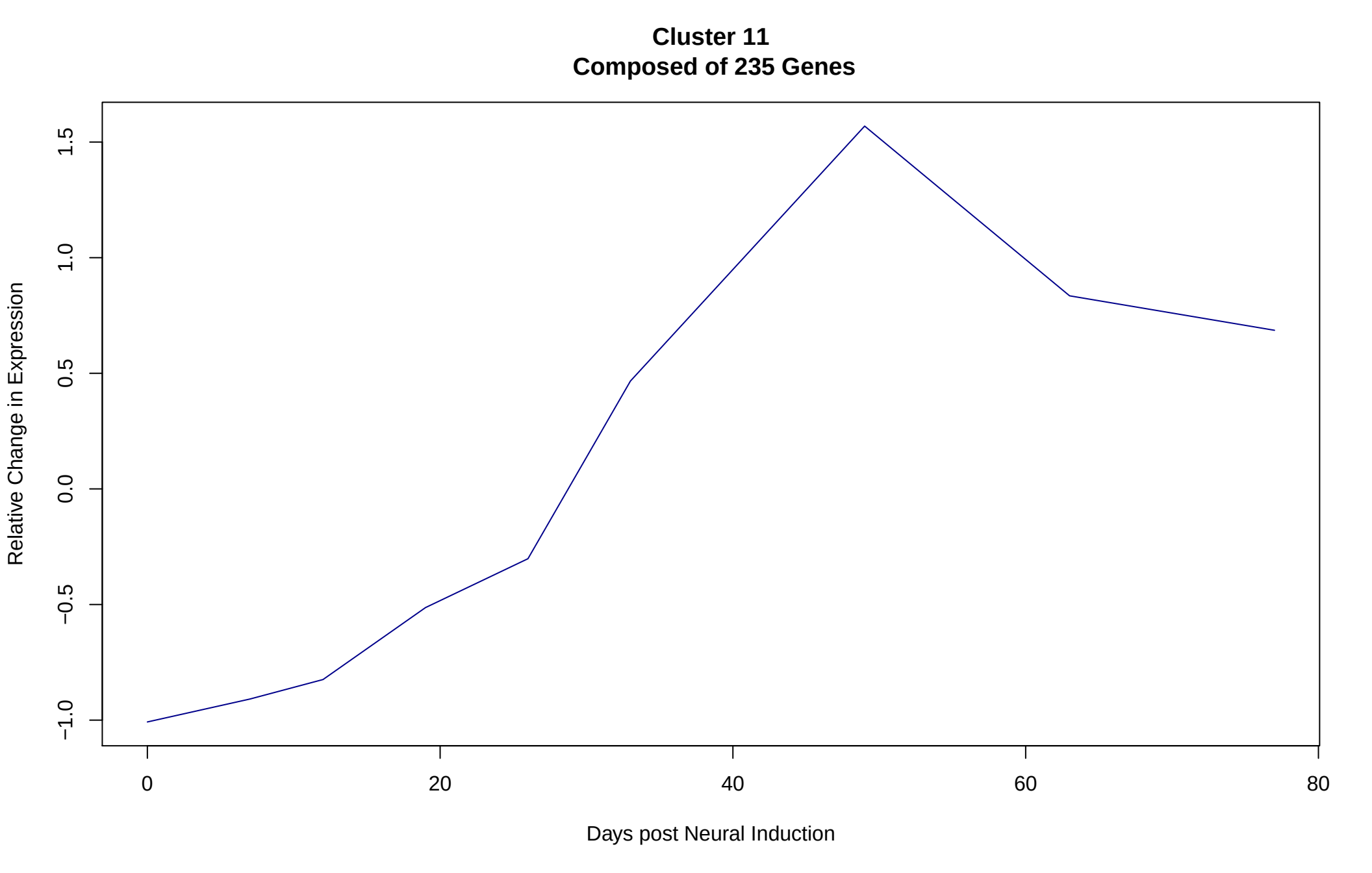

Cluster 6
Composed of 153 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

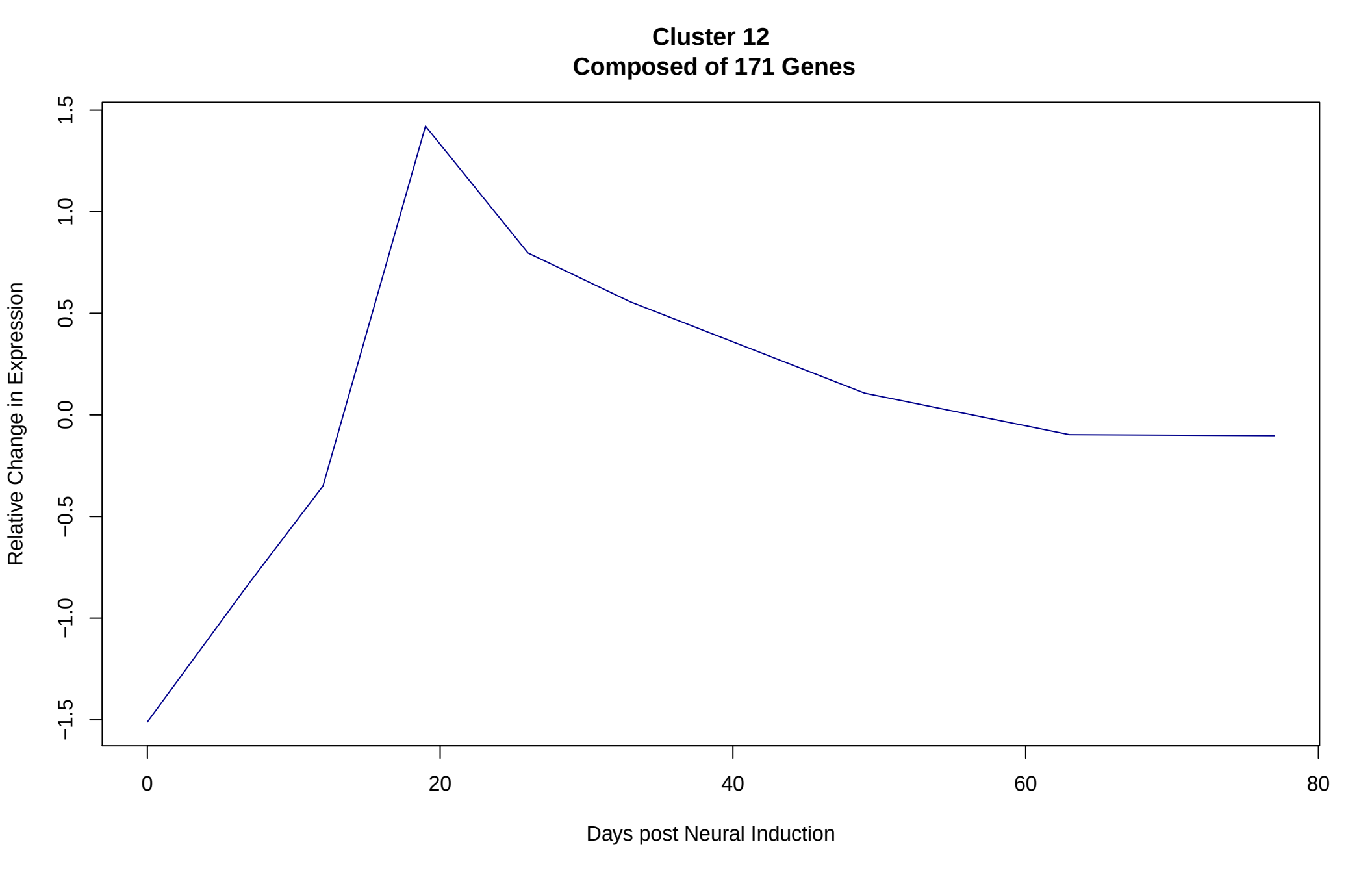

Cluster 7
Composed of 80 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

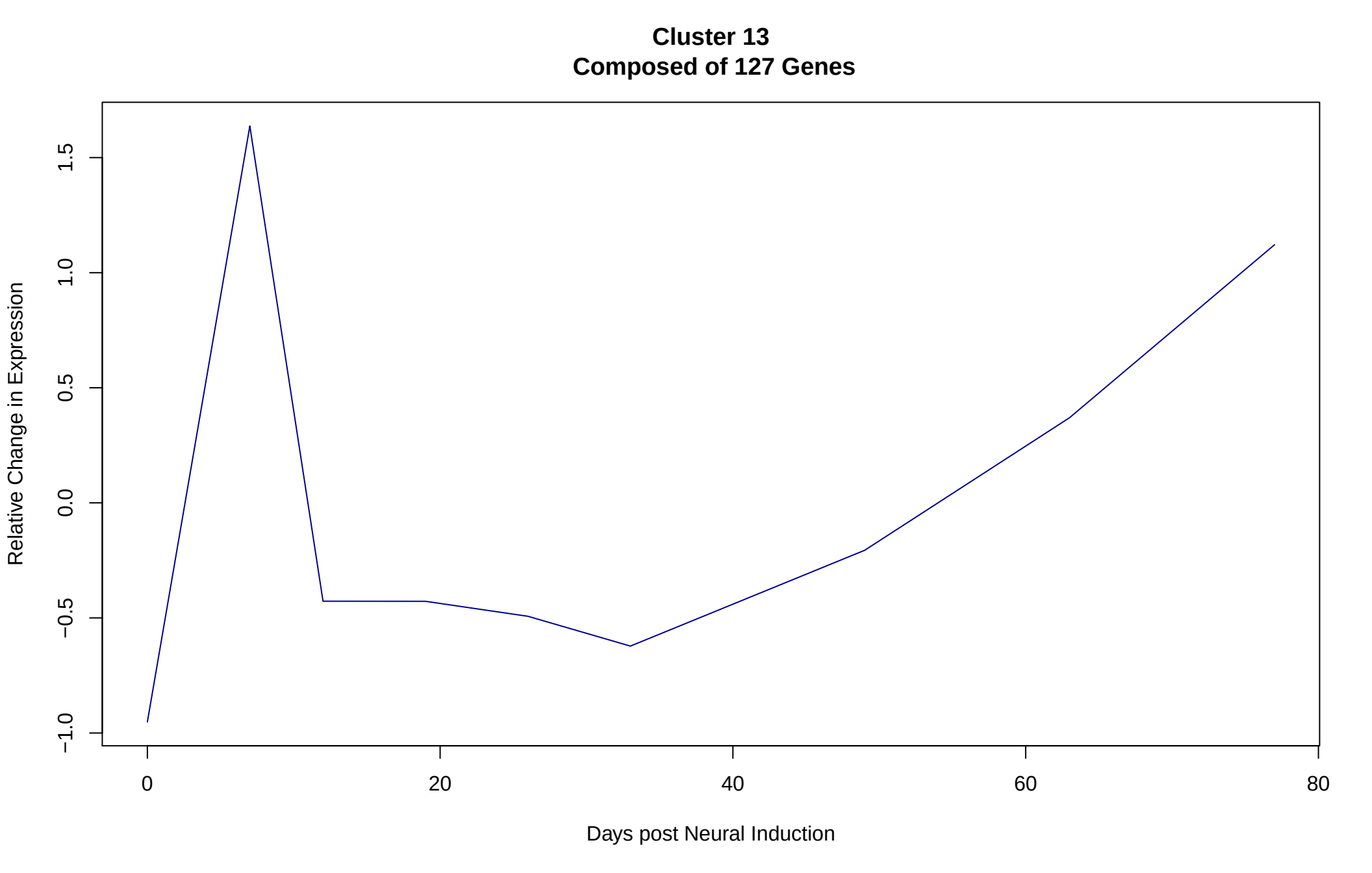

Cluster 8
Composed of 237 Genes
Relative Change in Expression
−1.0
−0.5
0.0
0.5
1.0
1.5
0
20
40
60
80
Days post Neural Induction

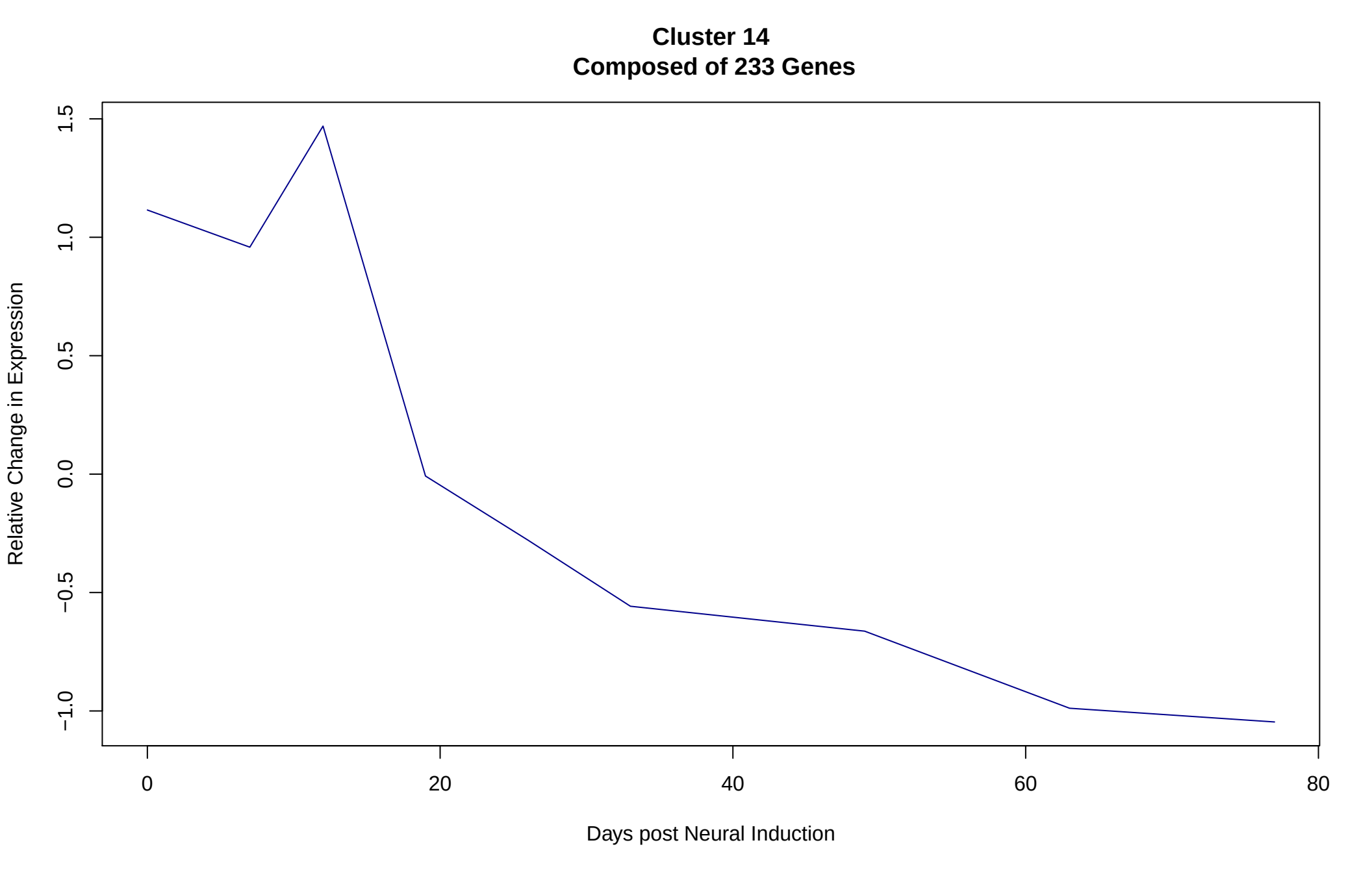

Cluster 9
Composed of 105 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

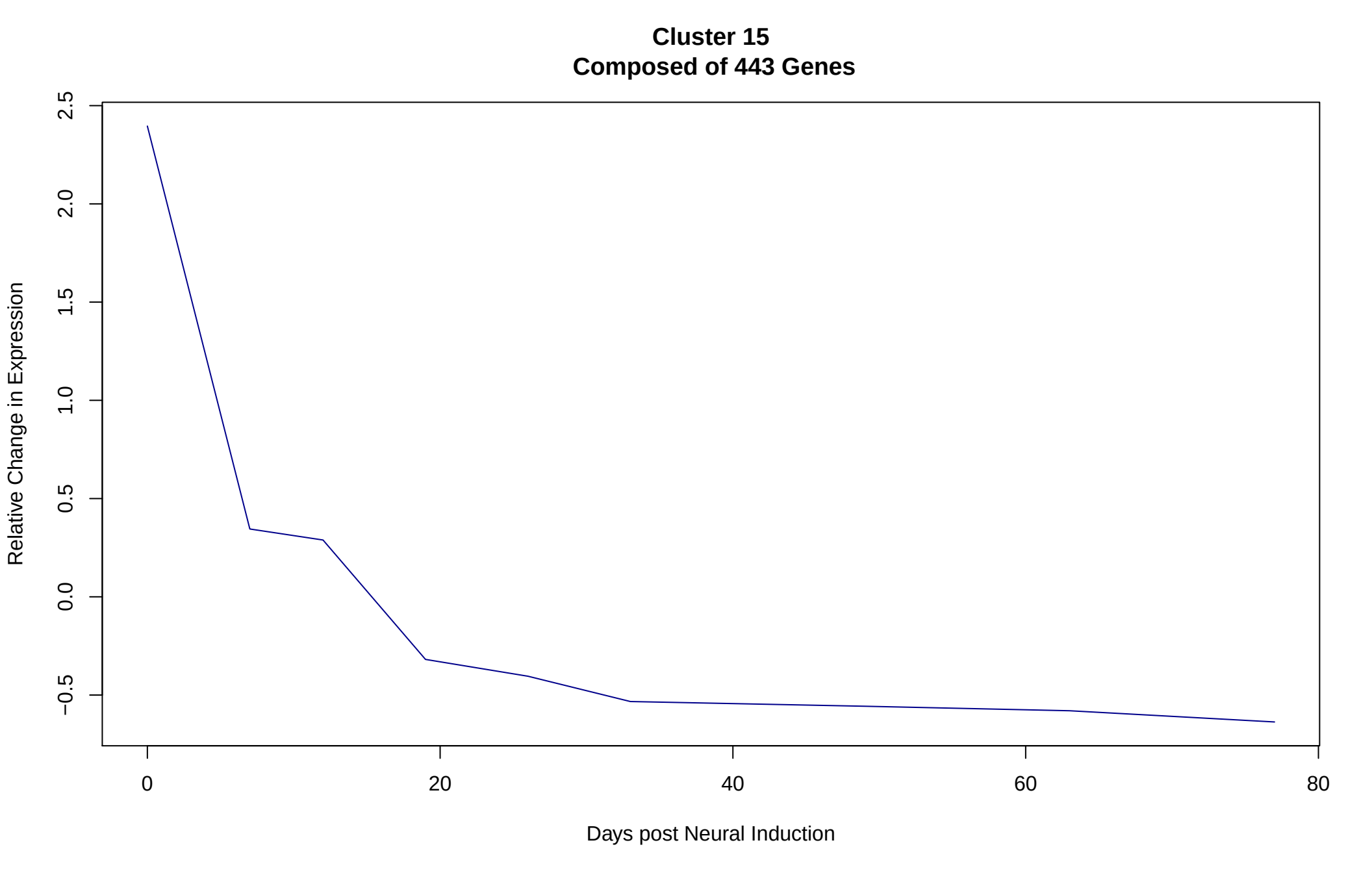

Cluster 10
Composed of 107 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

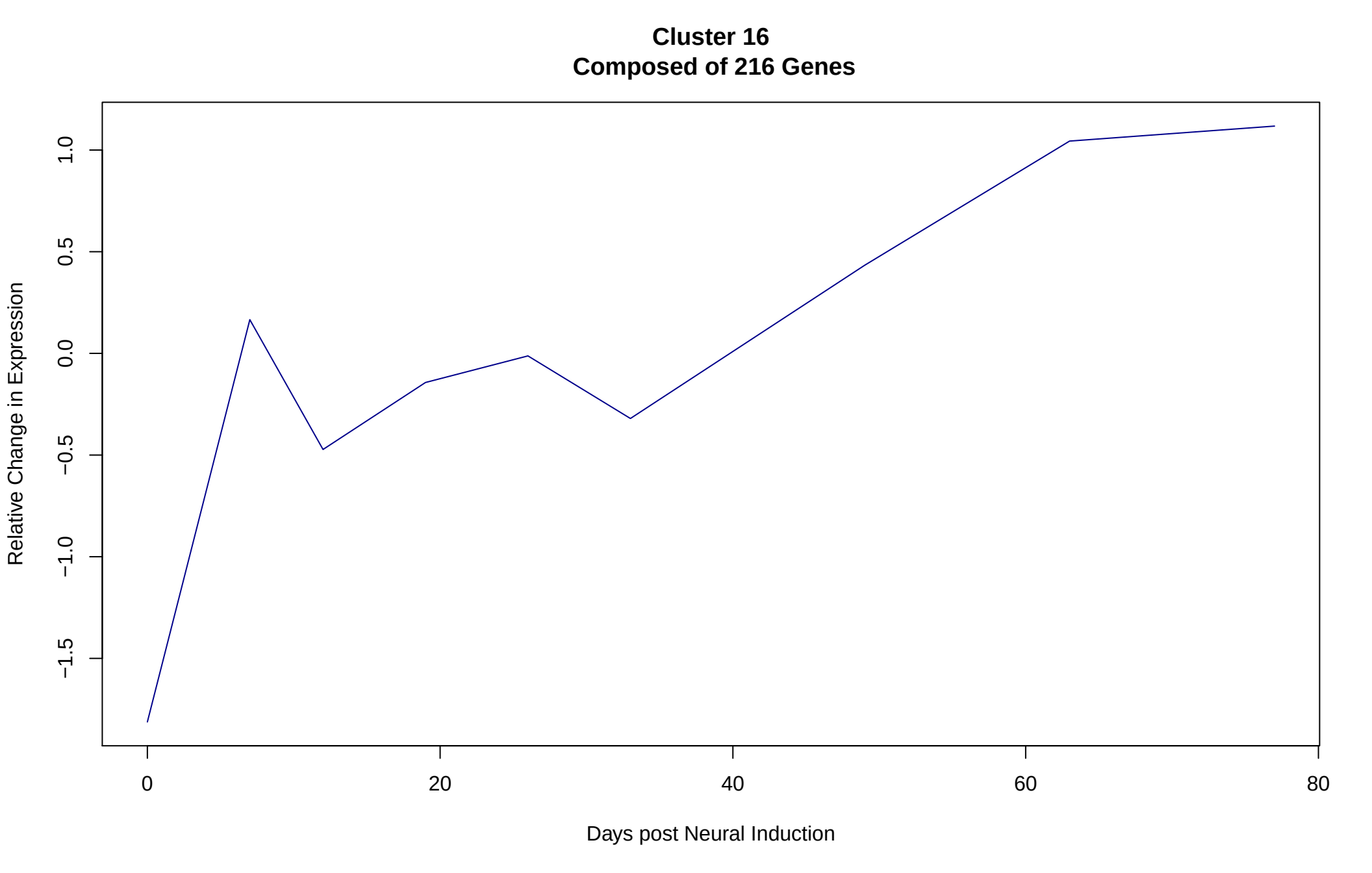

Cluster 11
Composed of 235 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

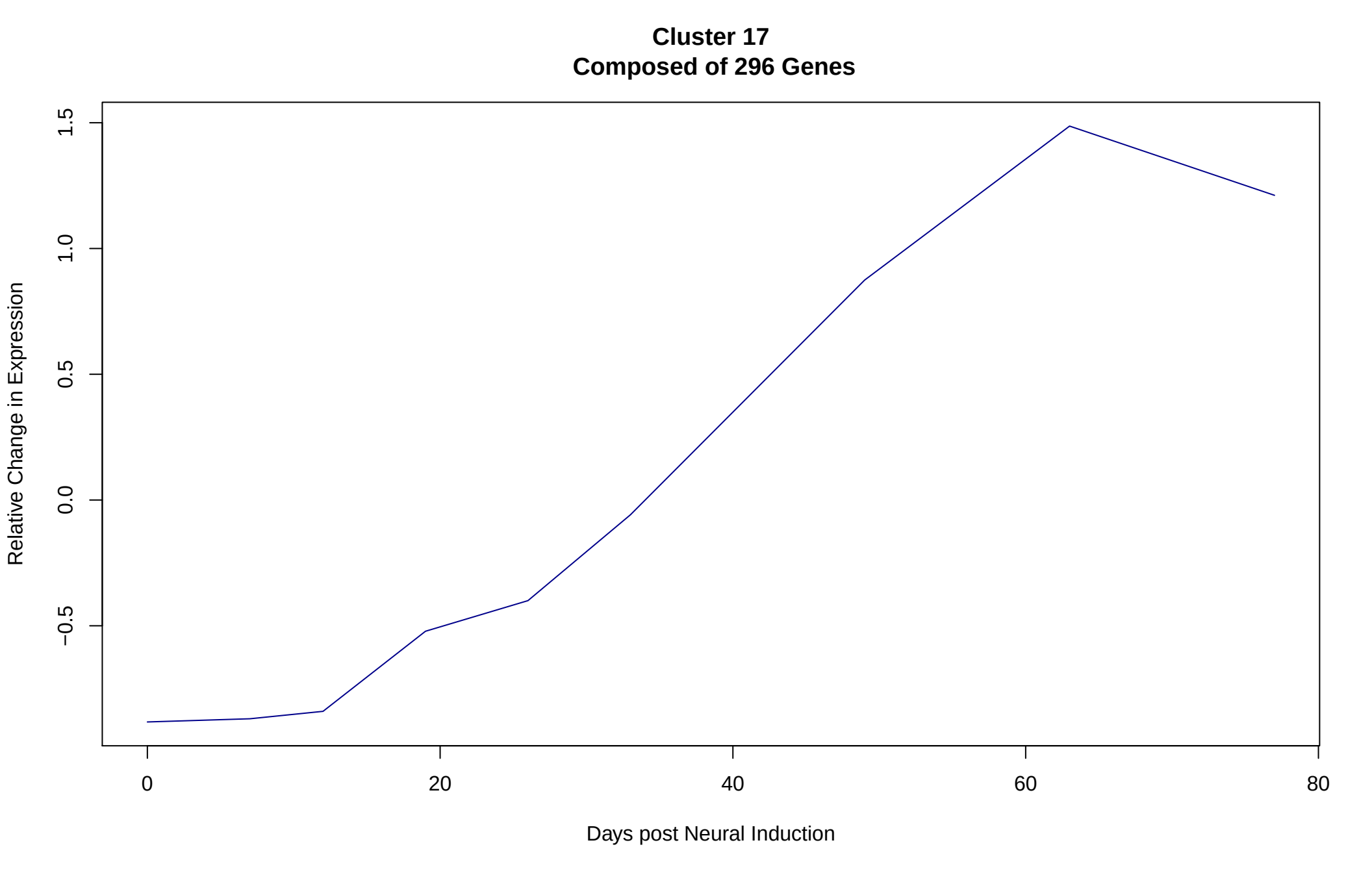

Cluster 12
Composed of 171 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

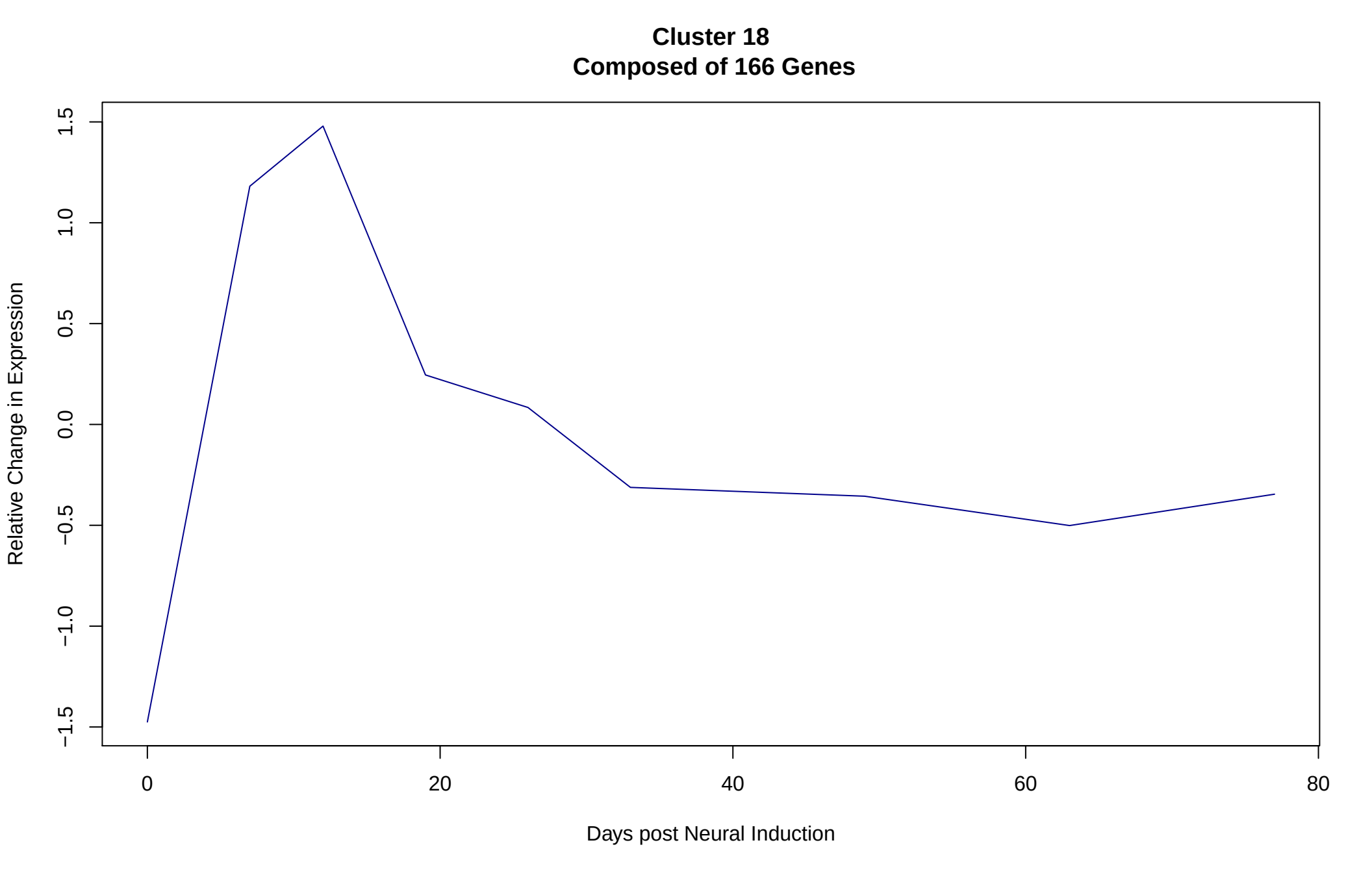

Cluster 13
Composed of 127 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

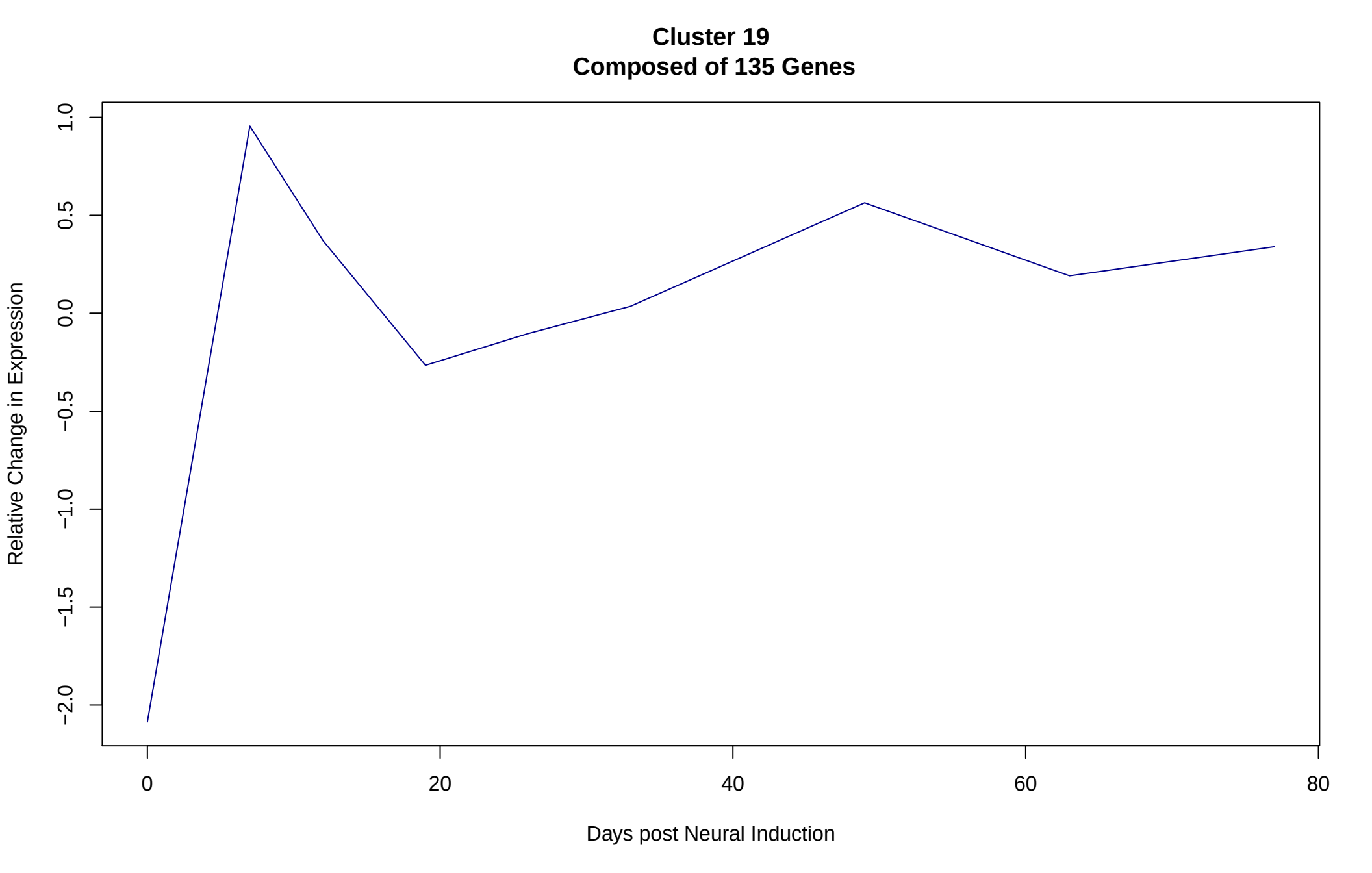

Cluster 14
Composed of 233 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

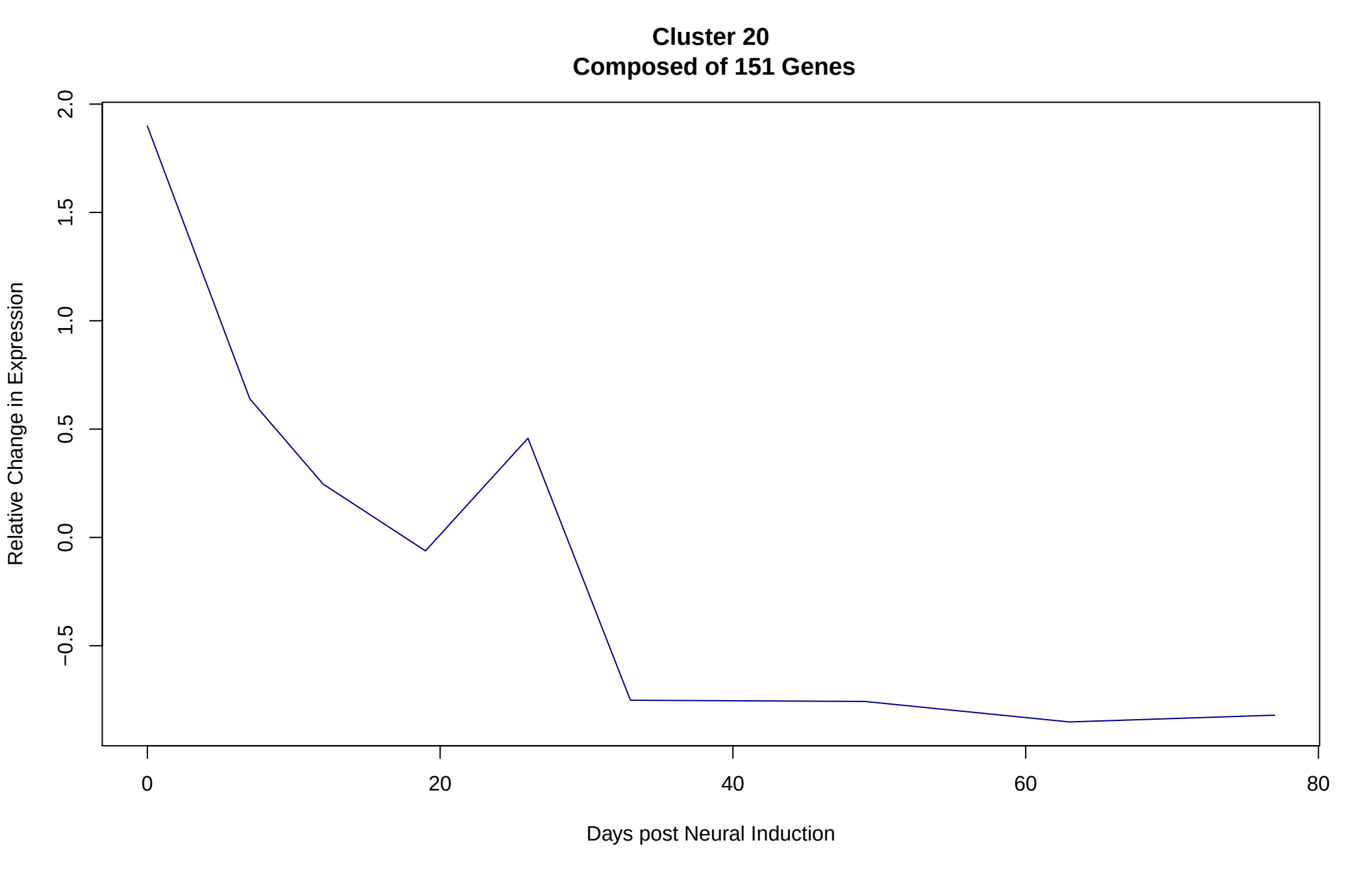

Cluster 15
Composed of 443 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

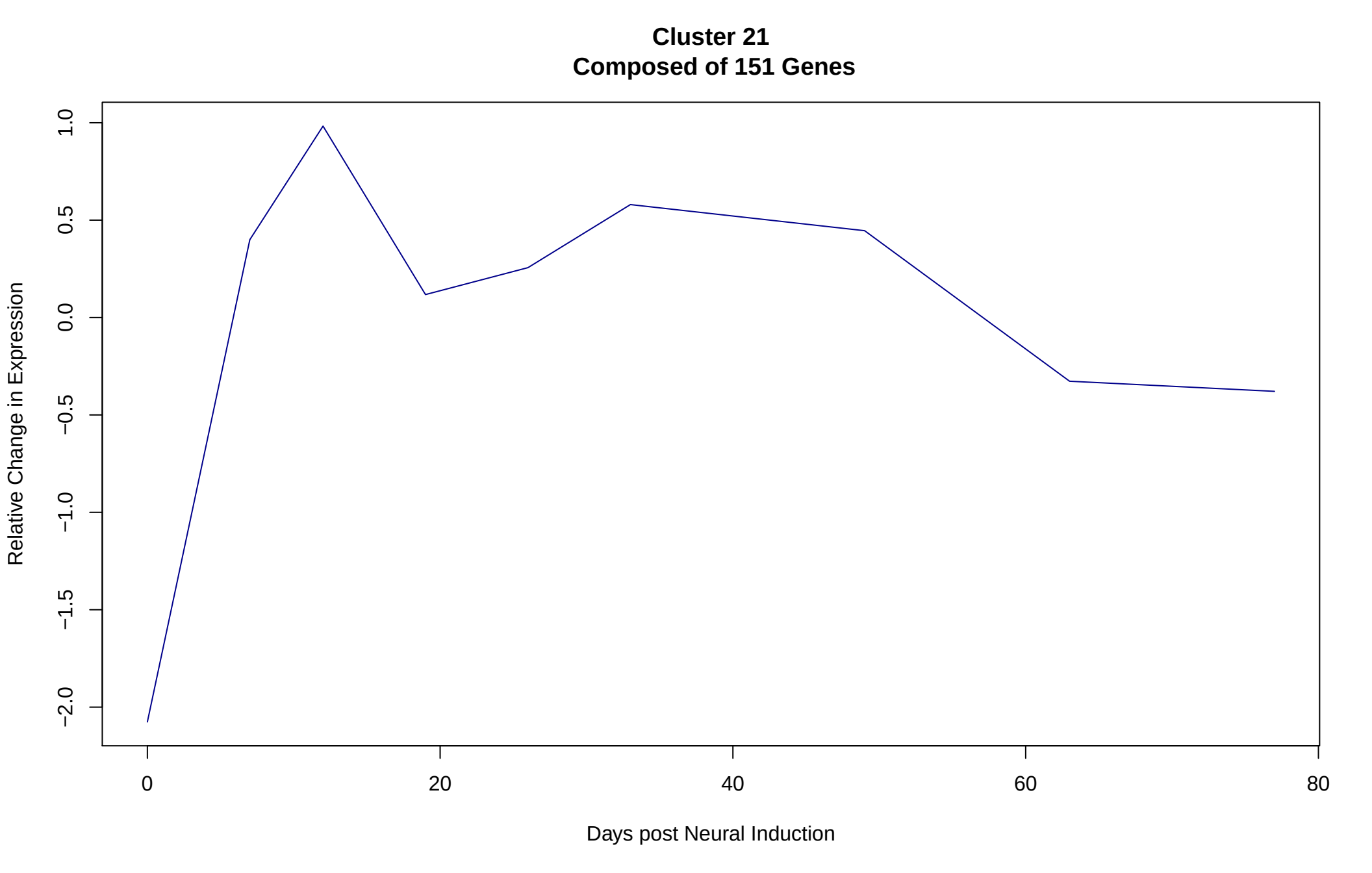

Cluster 16
Composed of 216 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

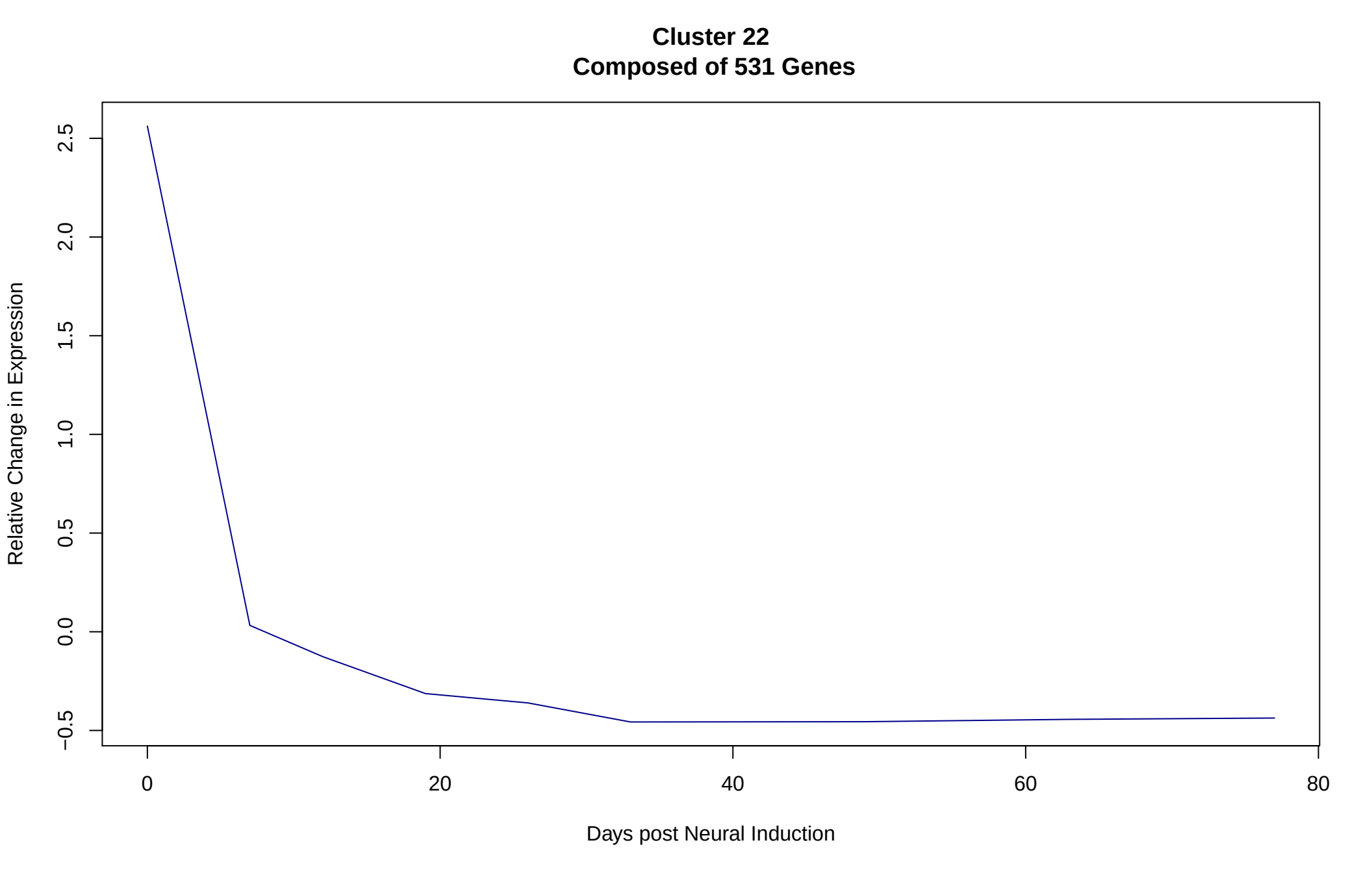

Cluster 17
Composed of 296 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

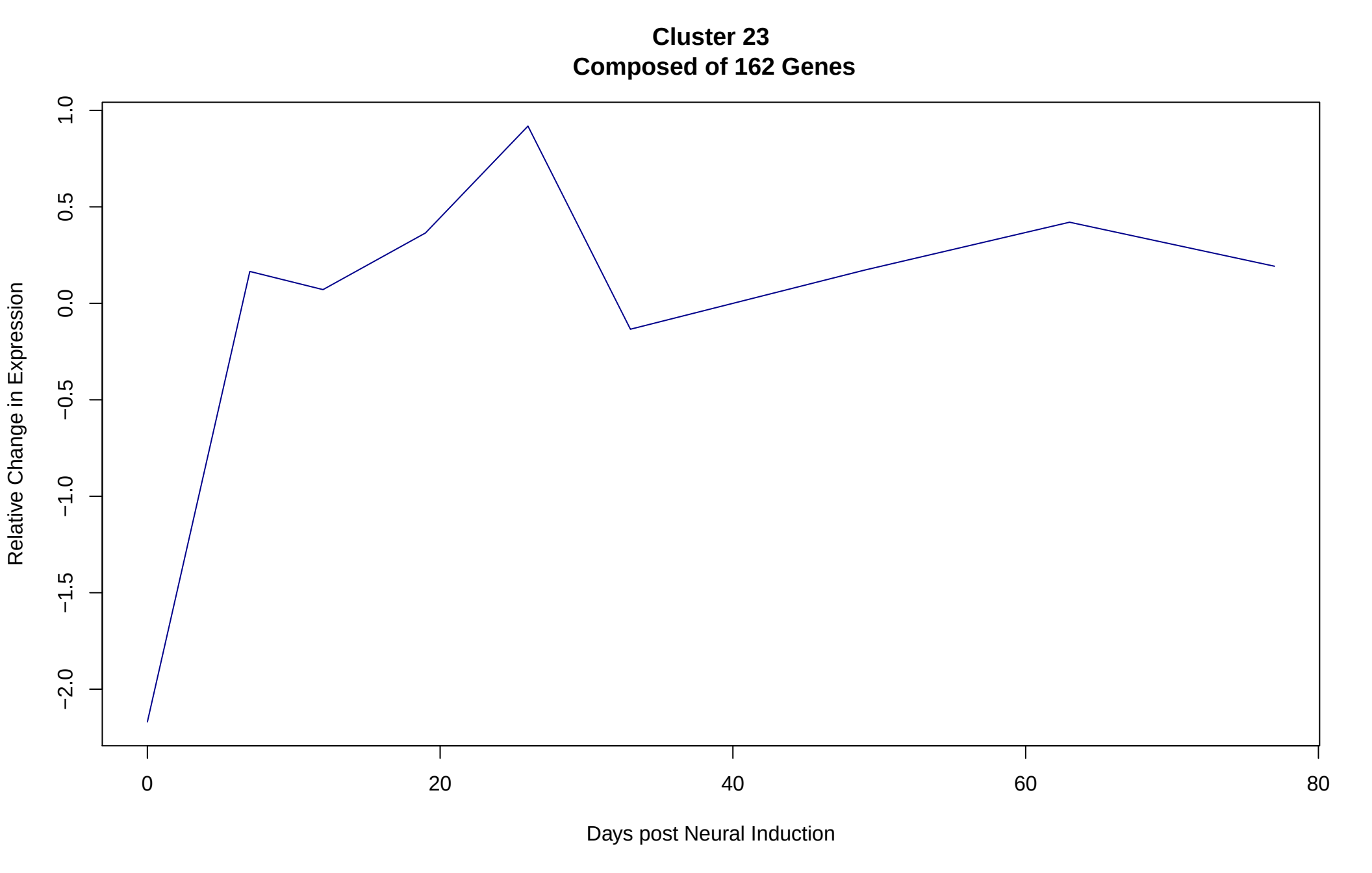

Cluster 18
Composed of 166 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

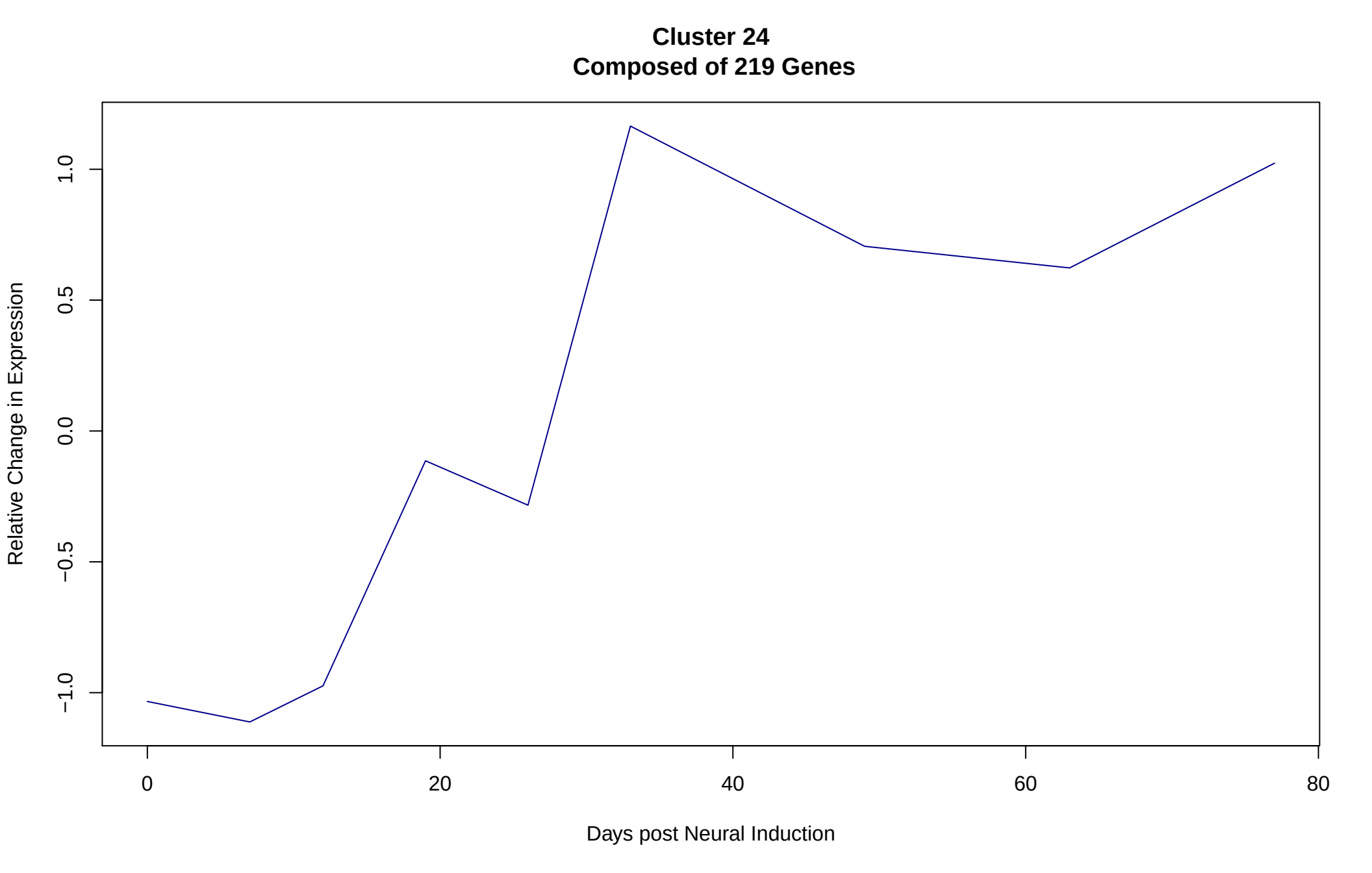

Cluster 19
Composed of 135 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

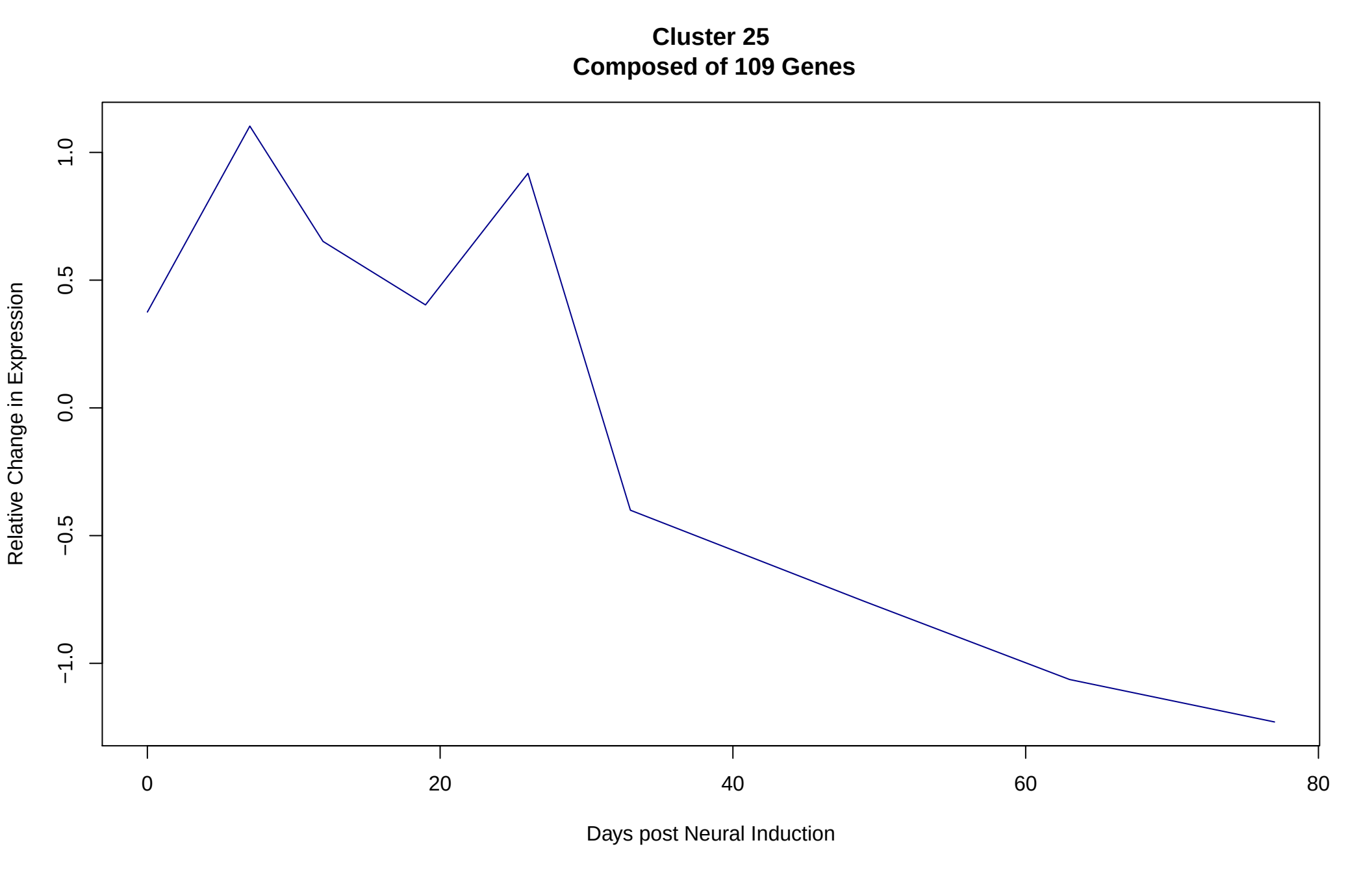

Cluster 20
Composed of 151 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

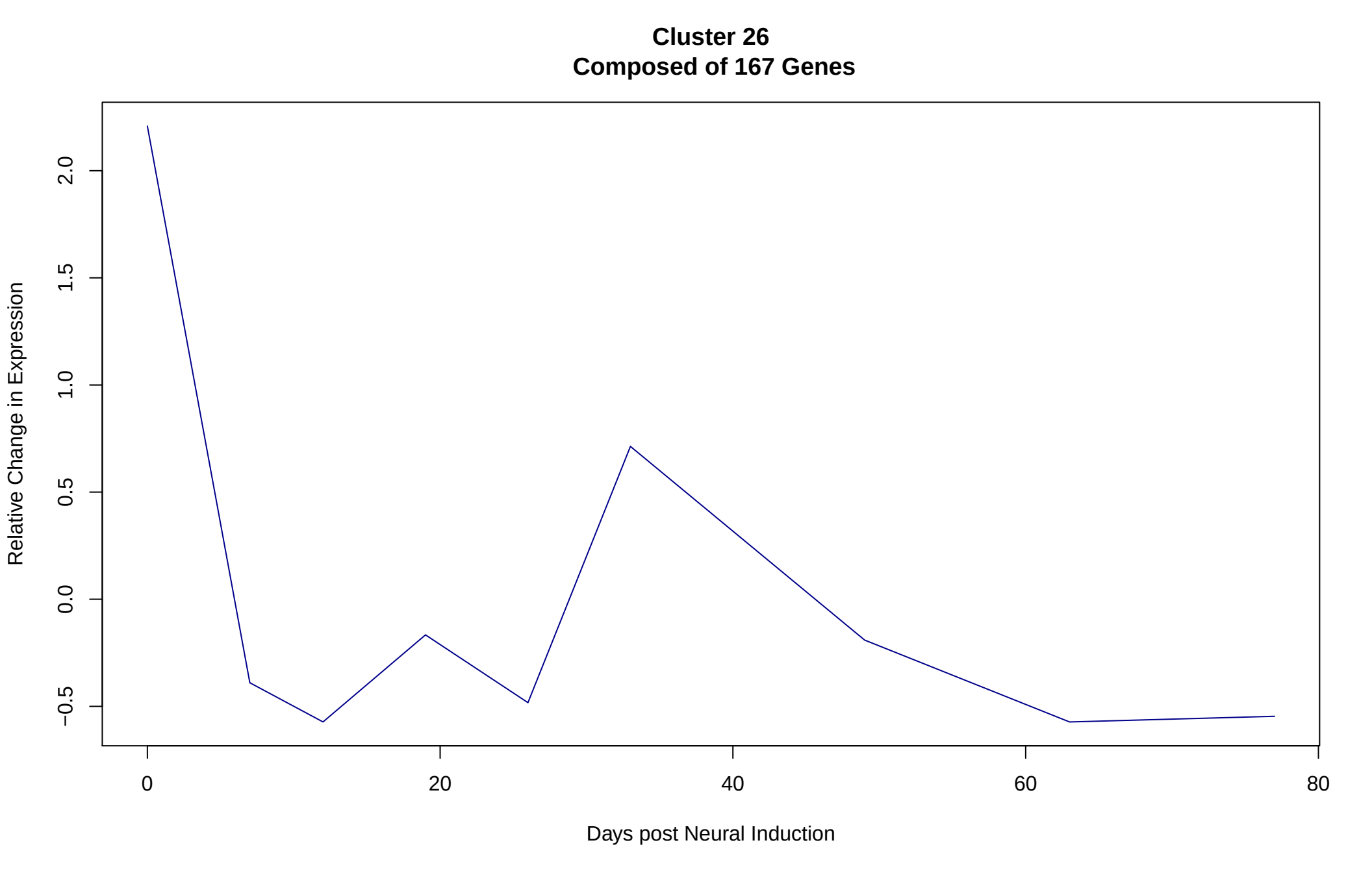

Cluster 21
Composed of 151 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

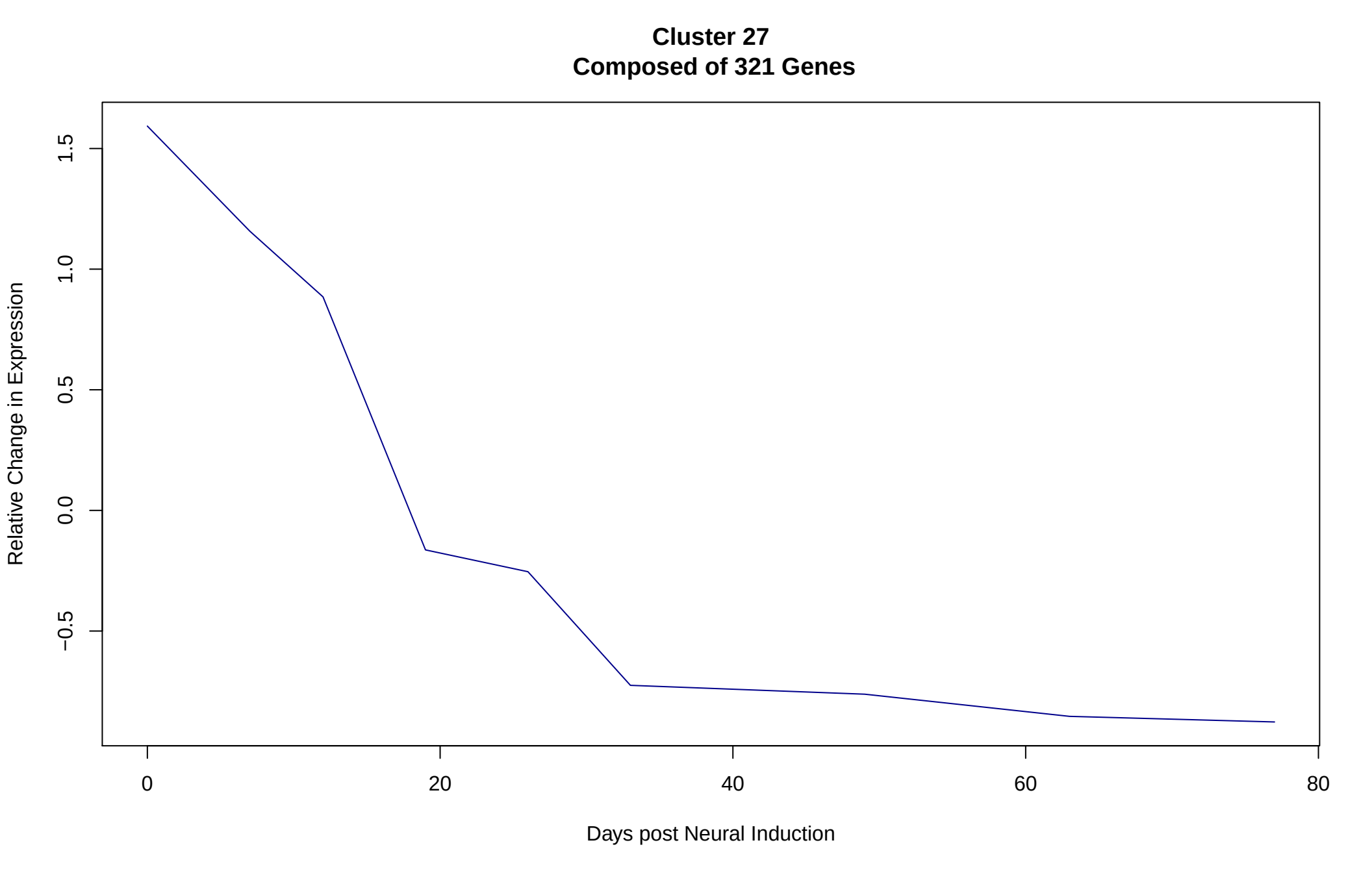

Cluster 22
Composed of 531 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

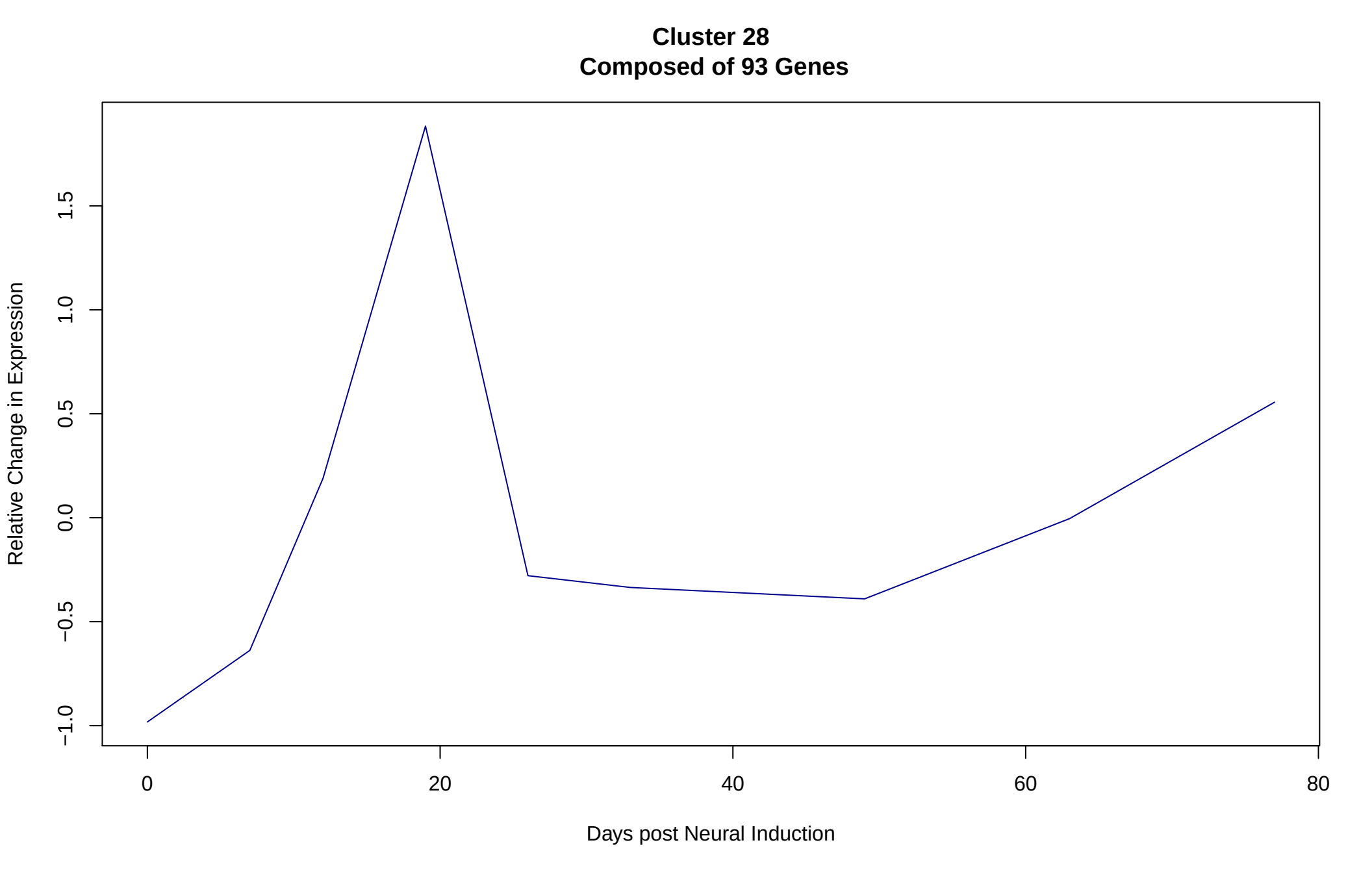

Cluster 23
Composed of 162 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

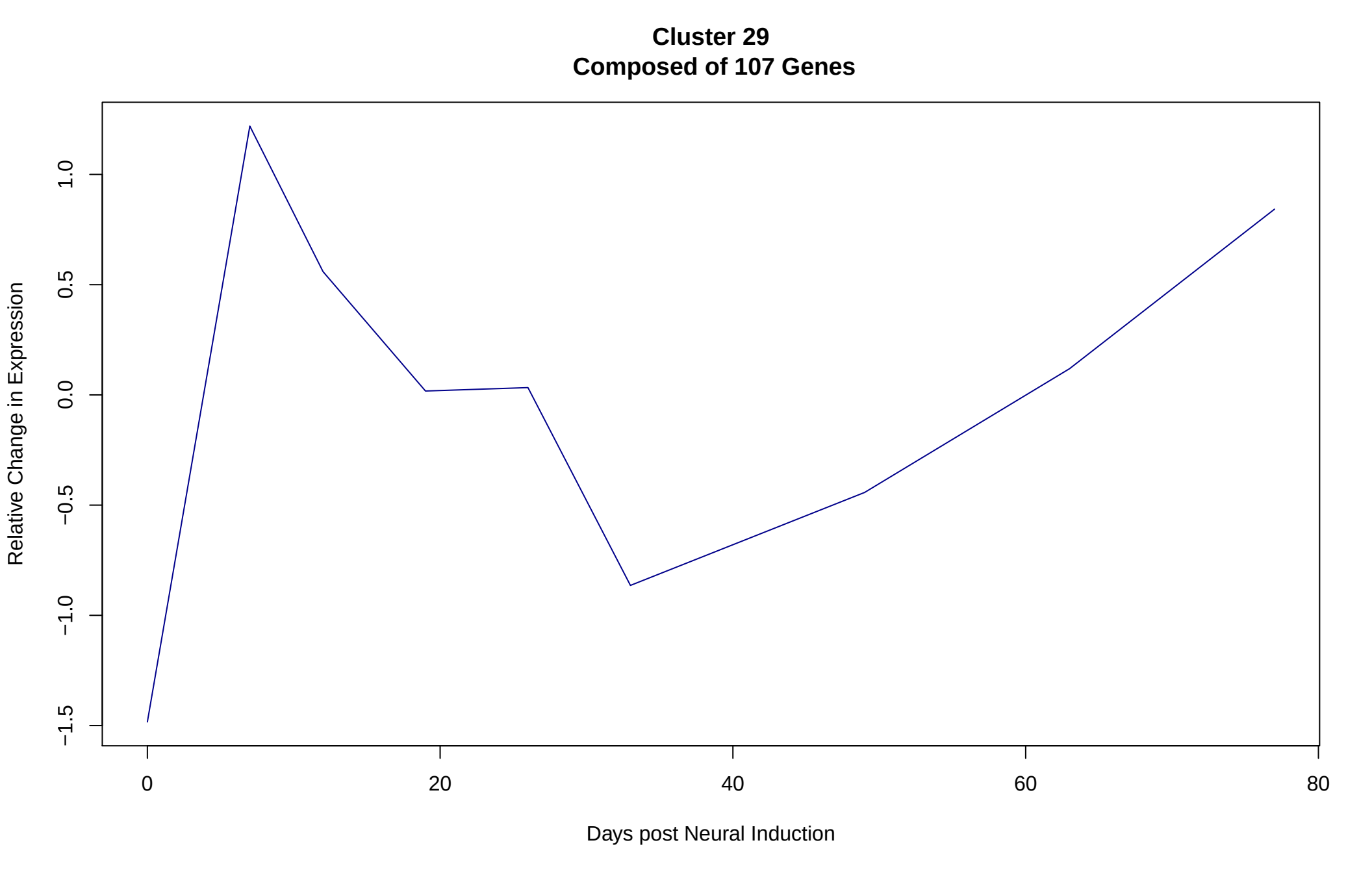

Cluster 24
Composed of 219 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

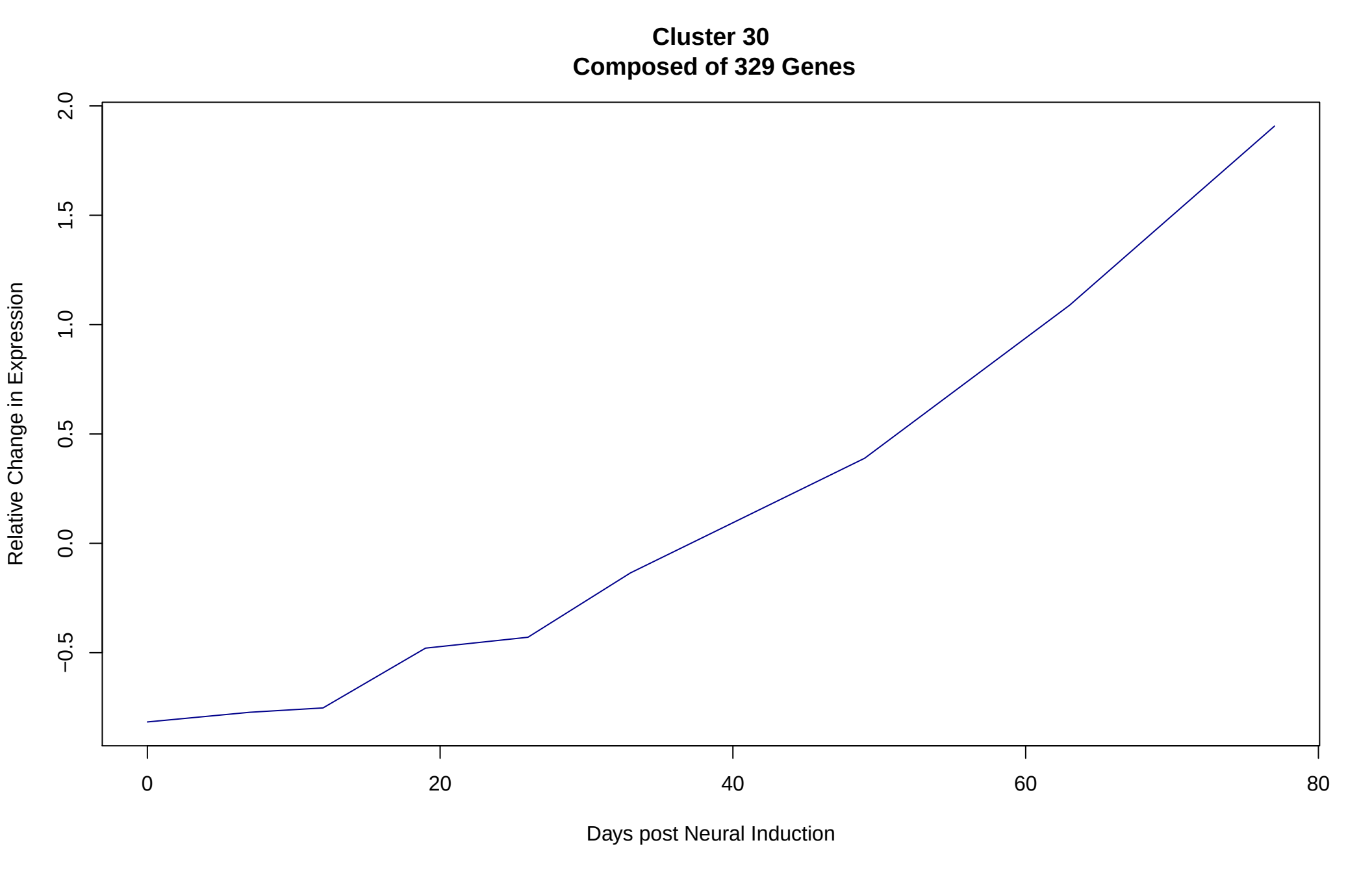

Cluster 25
Composed of 109 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

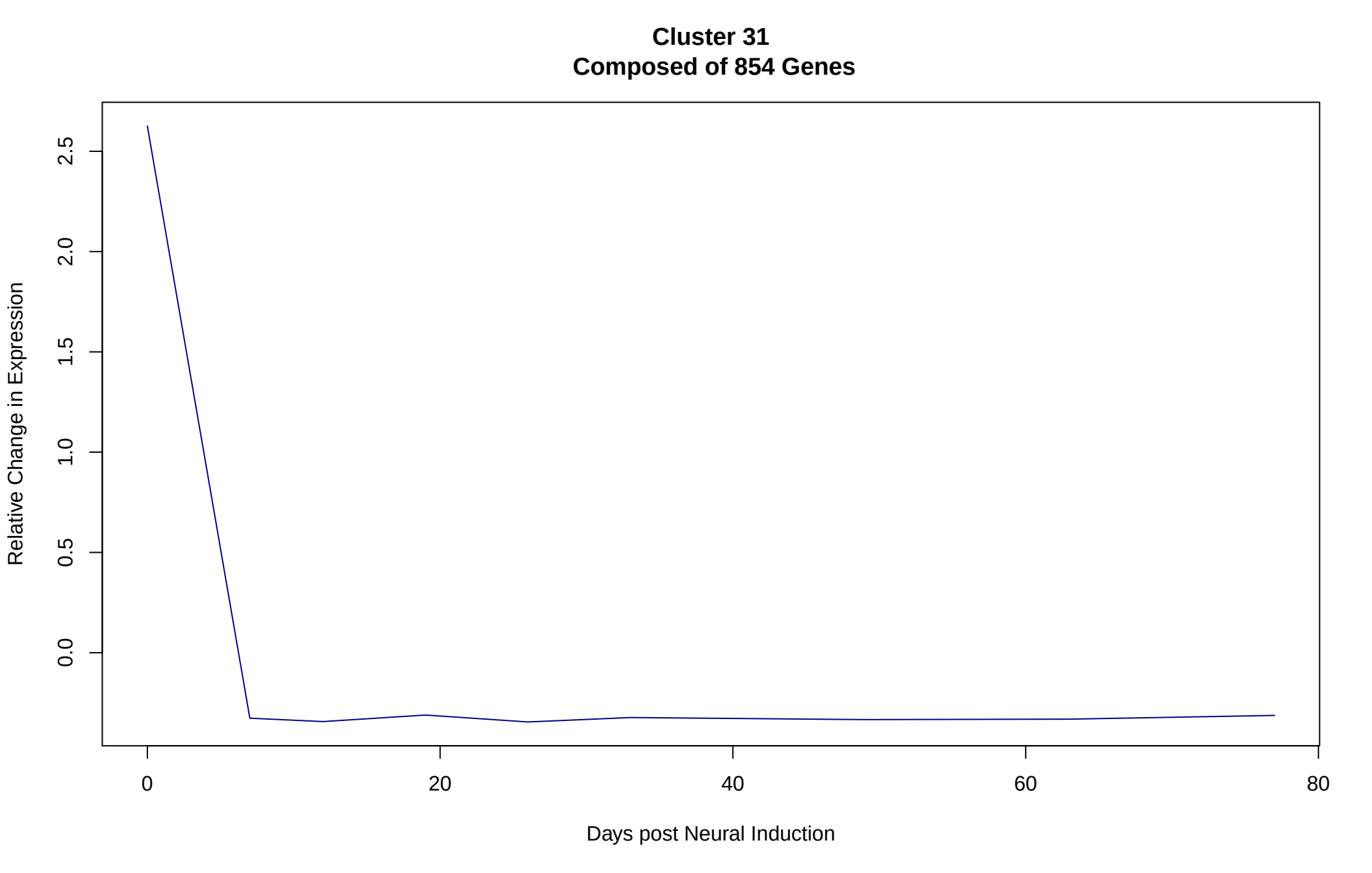

Cluster 26
Composed of 167 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

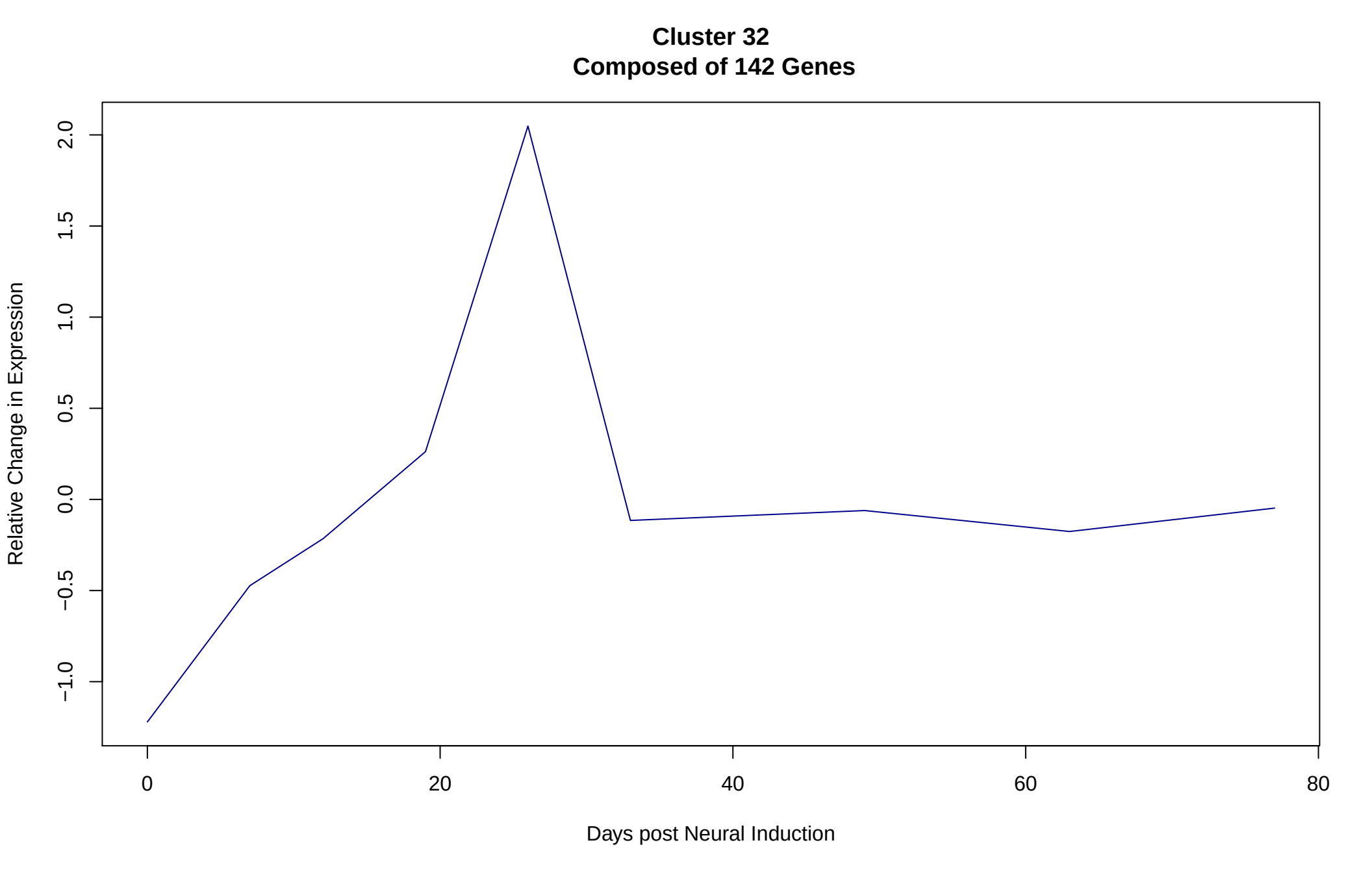

Cluster 27
Composed of 321 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

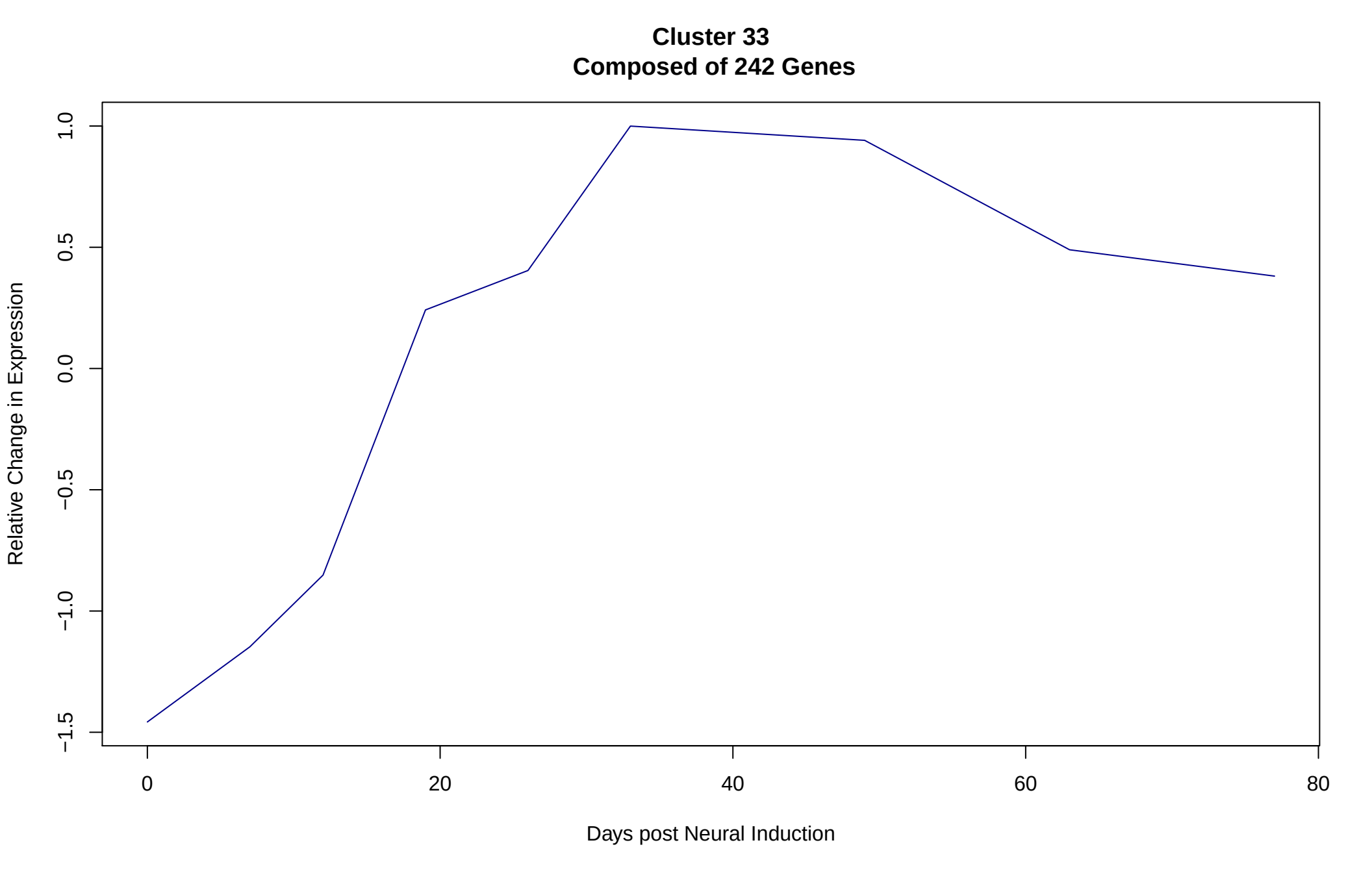

Cluster 28
Composed of 93 Genes
Relative Change in Expression
−1.0
−0.5
0.0
0.5
1.0
1.5
0
20
40
60
80
Days post Neural Induction

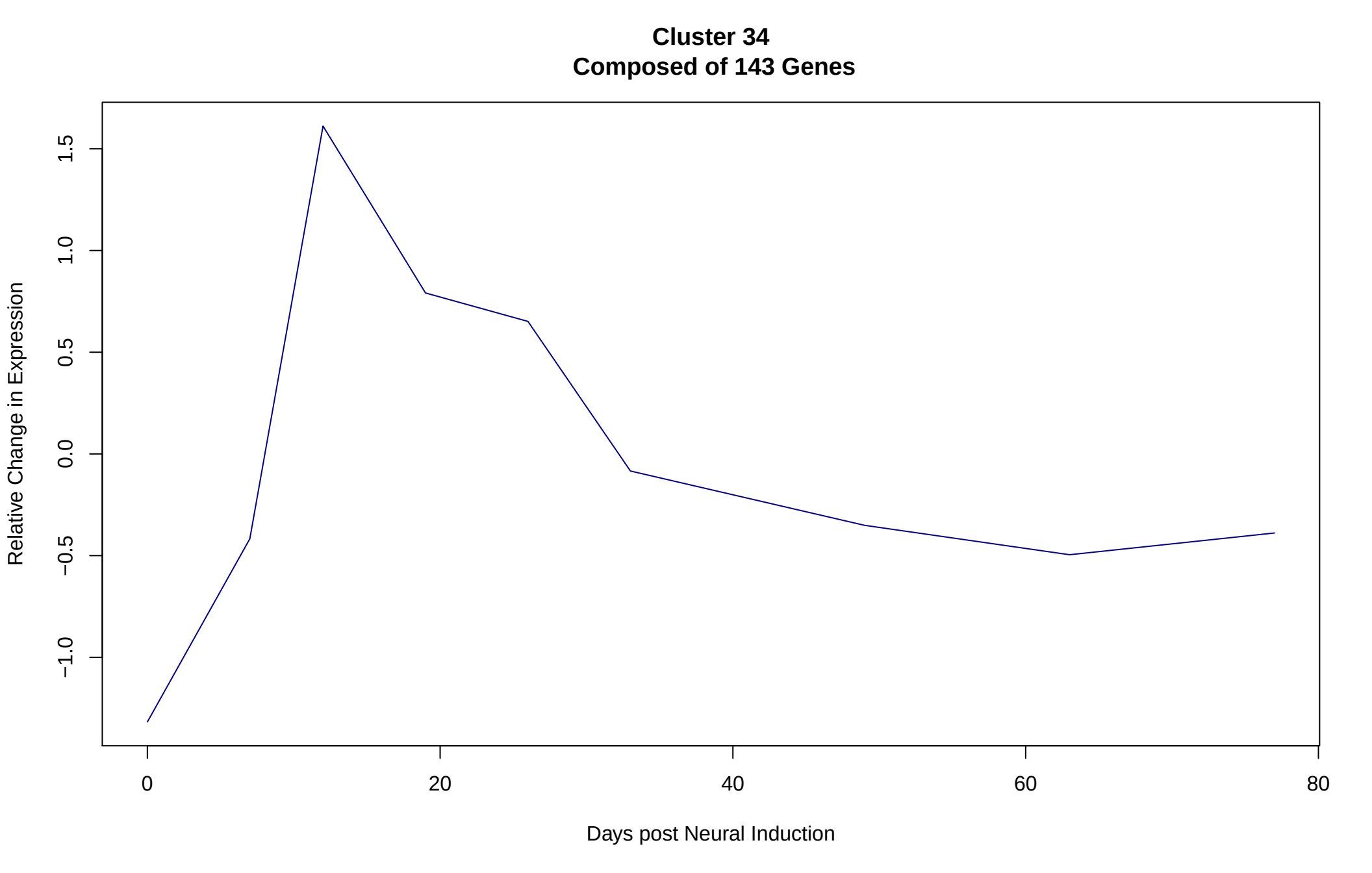

Cluster 29
Composed of 107 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

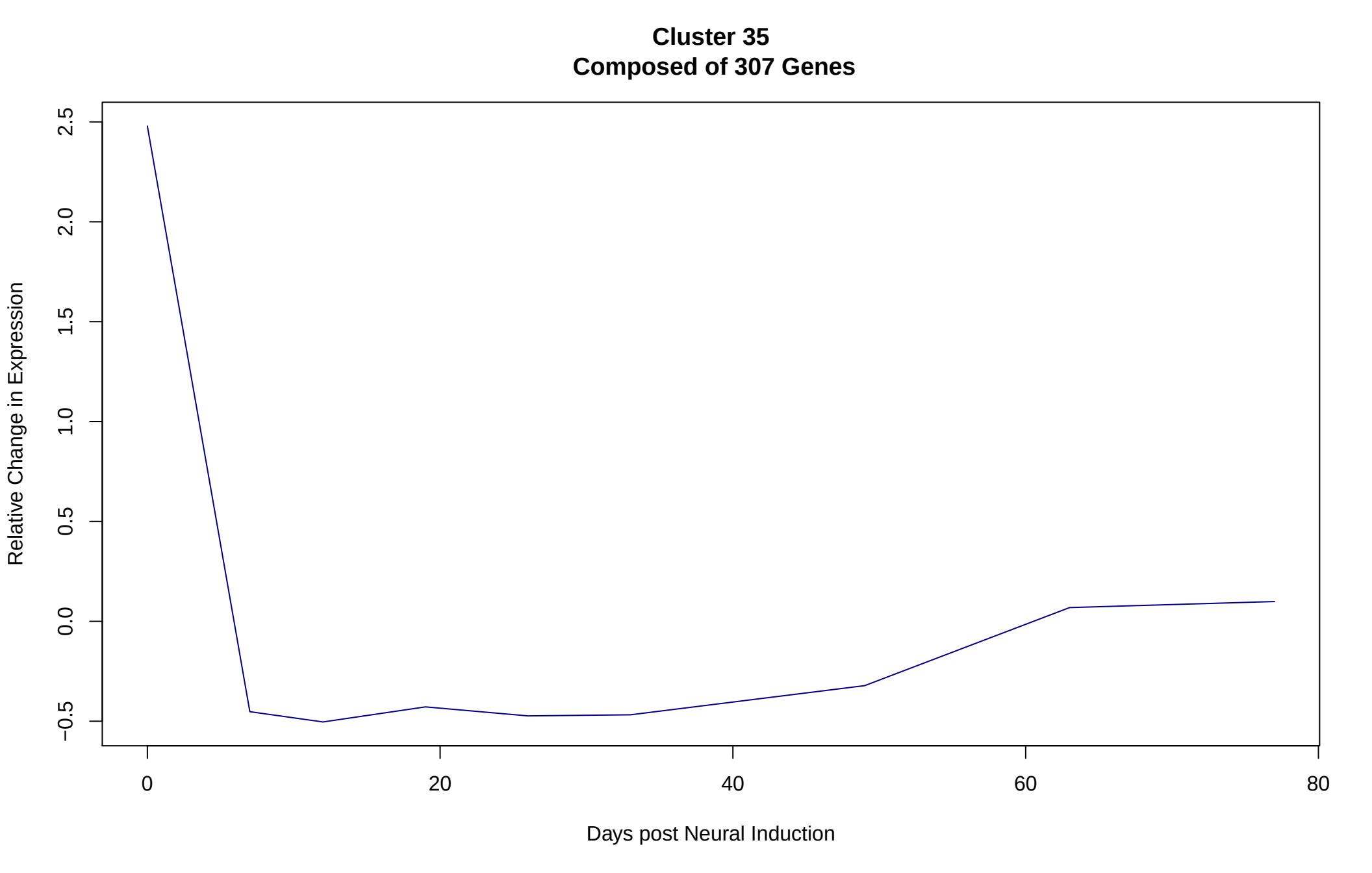

Cluster 30
Composed of 329 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

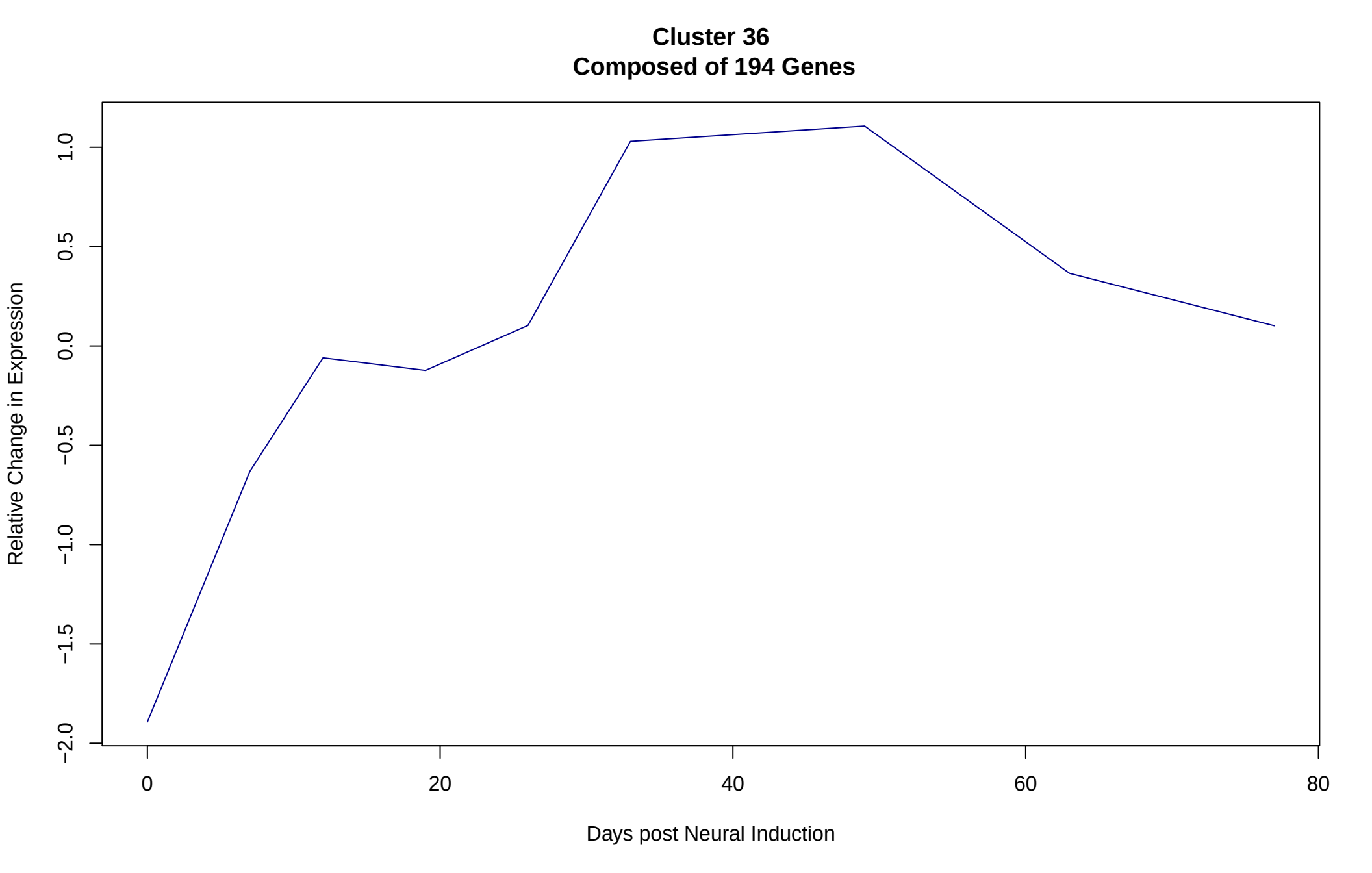

Cluster 31
Composed of 854 Genes
Relative Change in Expression
0.0
0.5
1.0
1.5
2.0
2.5
0
20
40
60
80
Days post Neural Induction

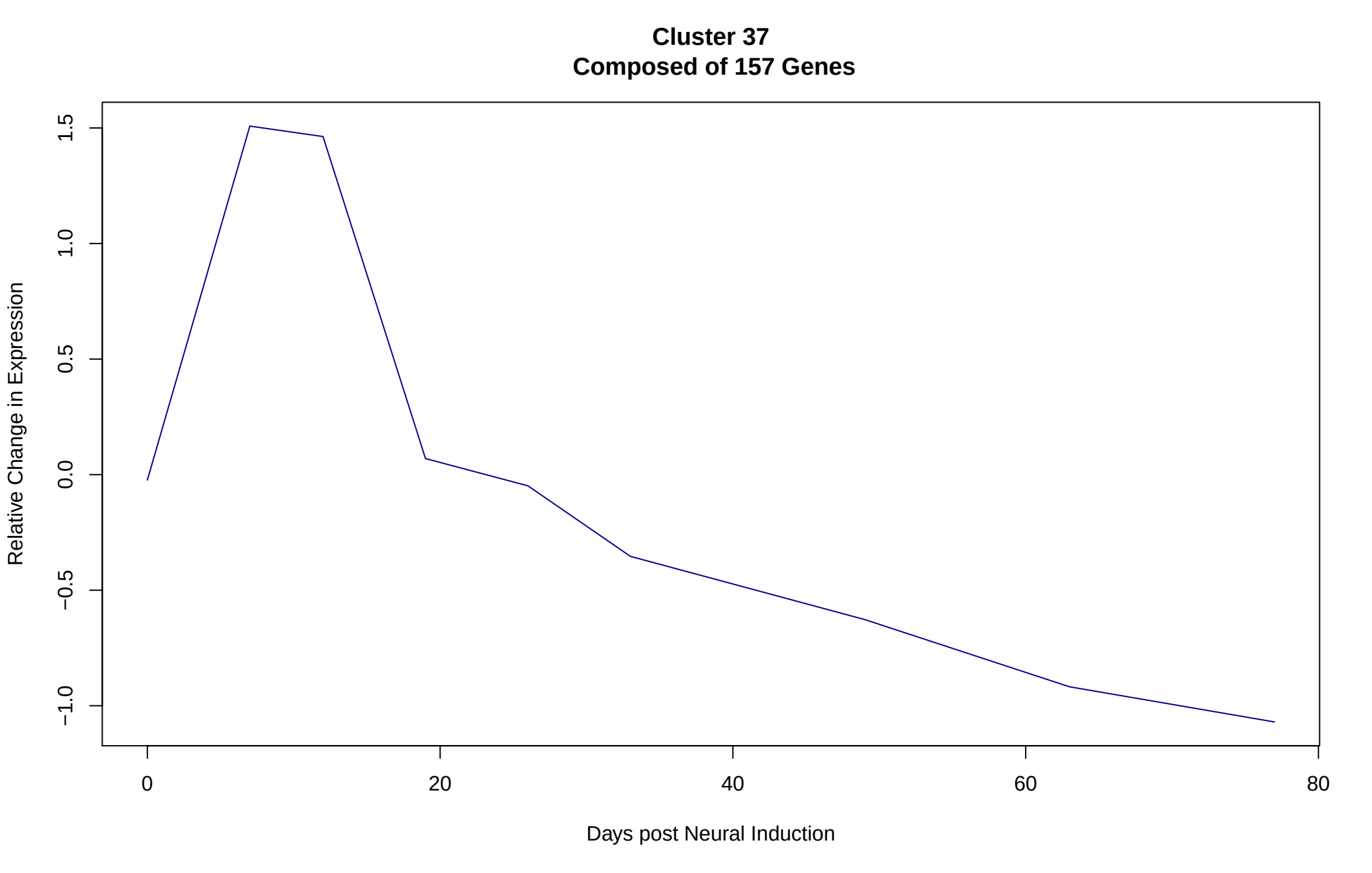

Cluster 32
Composed of 142 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

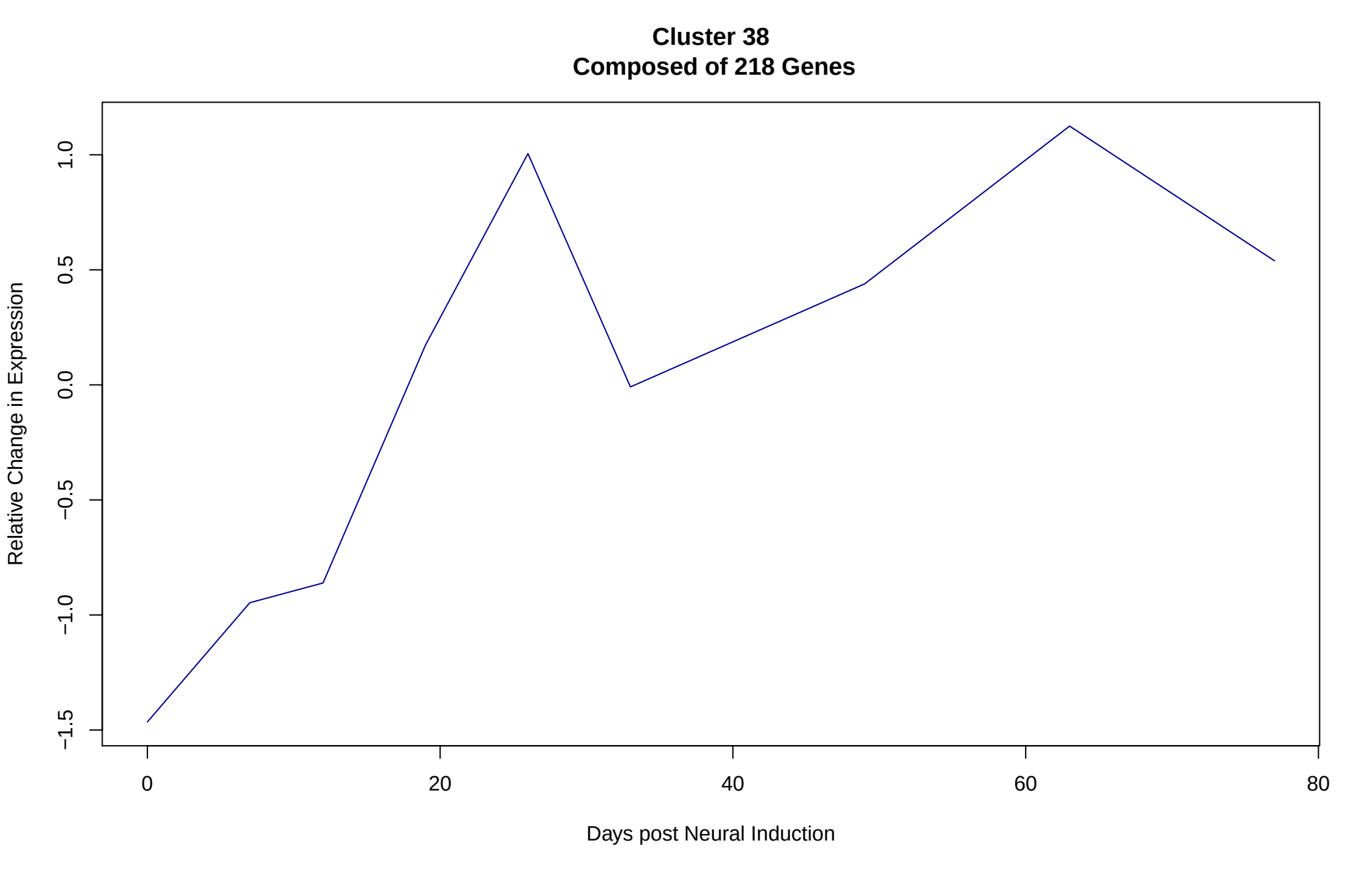

Cluster 33
Composed of 242 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

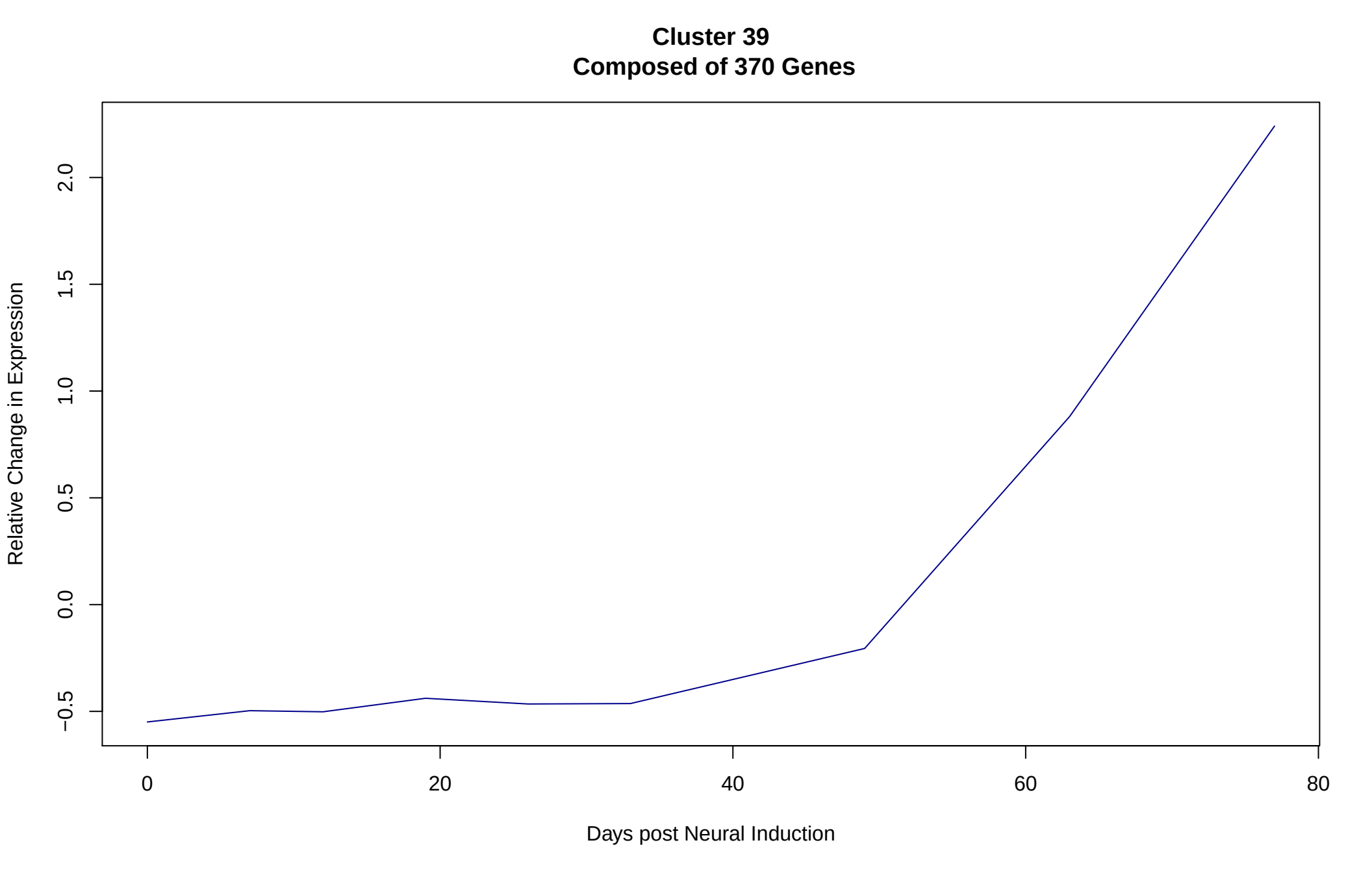

Cluster 34
Composed of 143 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

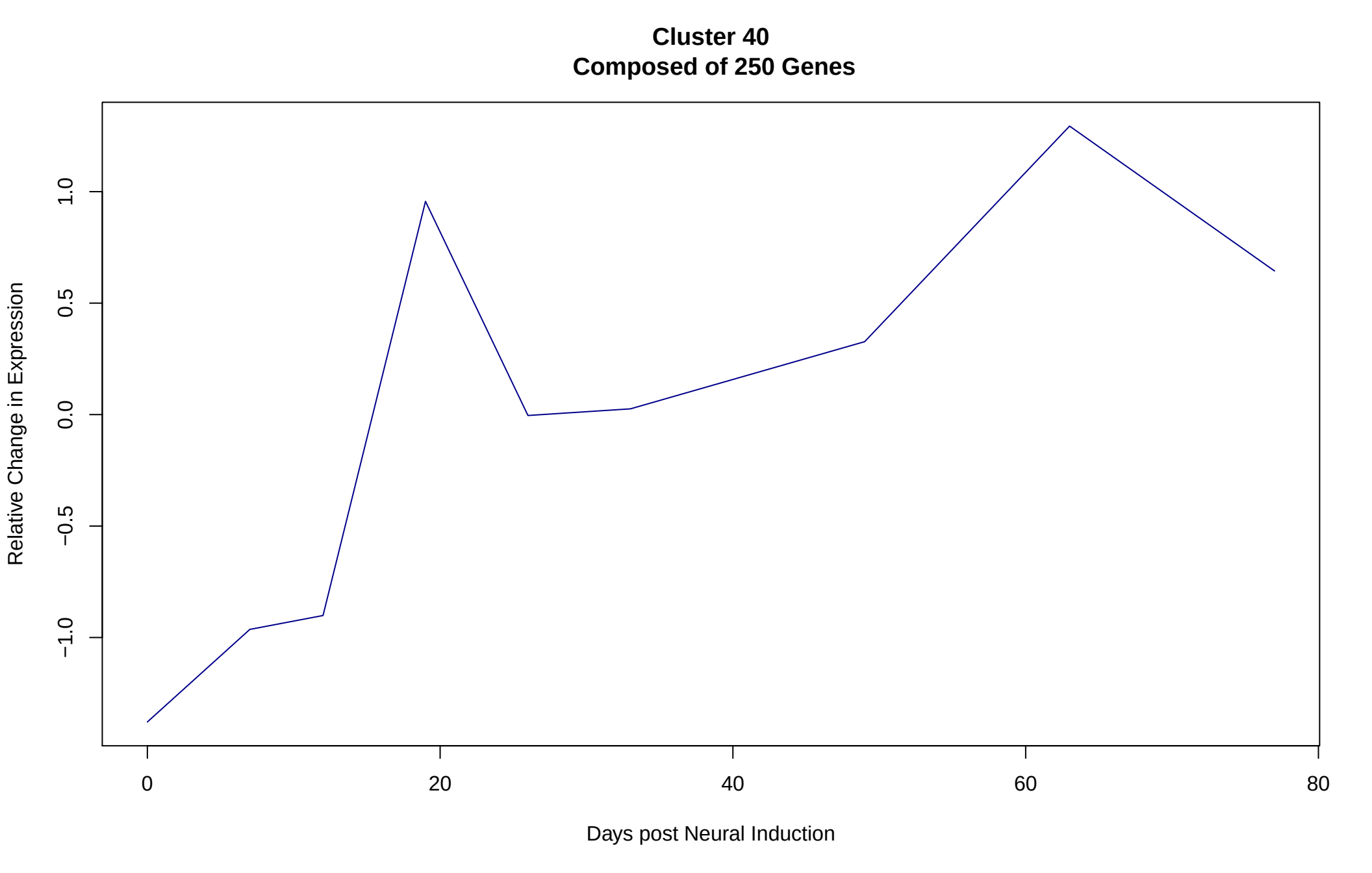

Cluster 35
Composed of 307 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

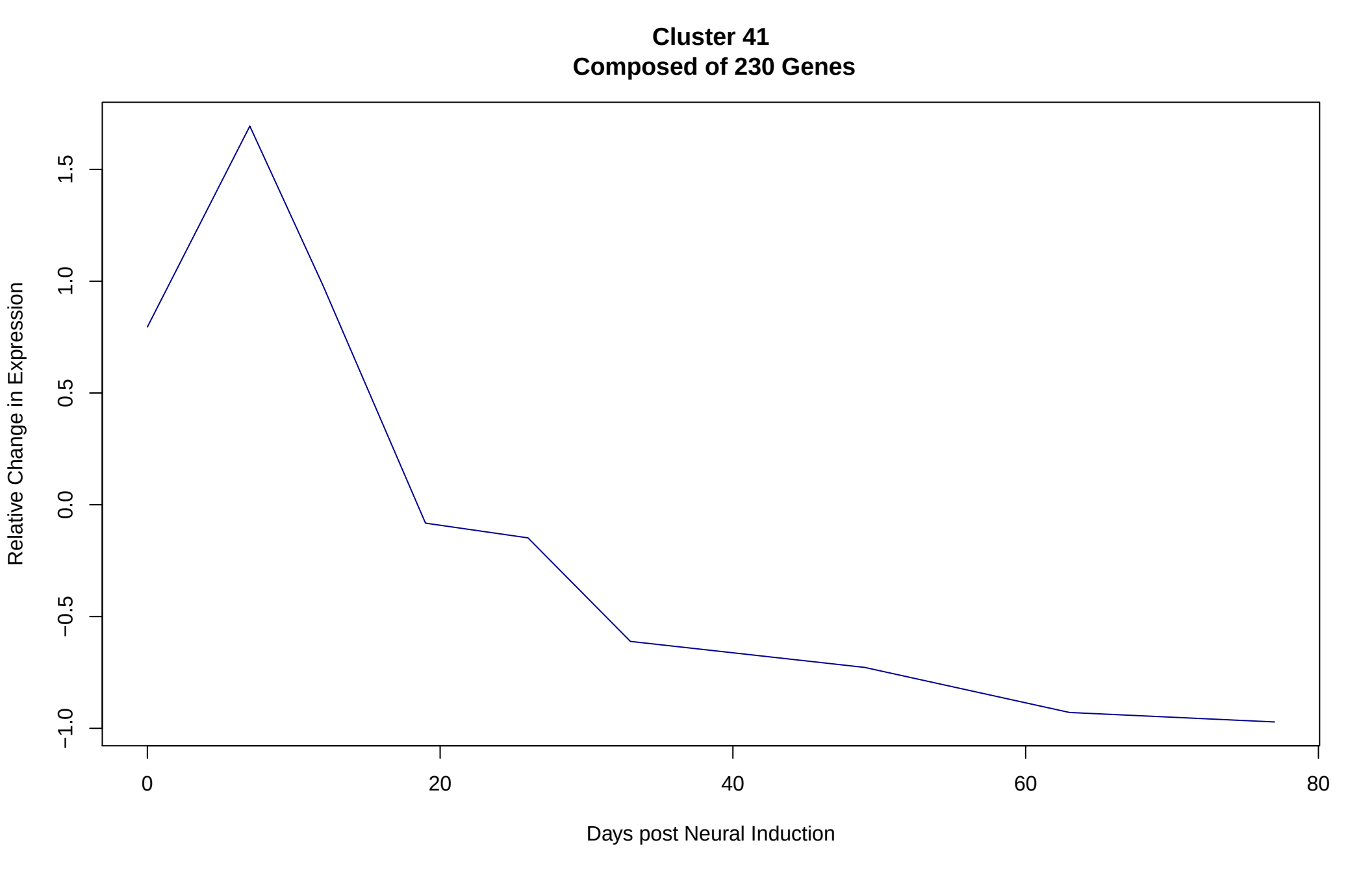

Cluster 36
Composed of 194 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

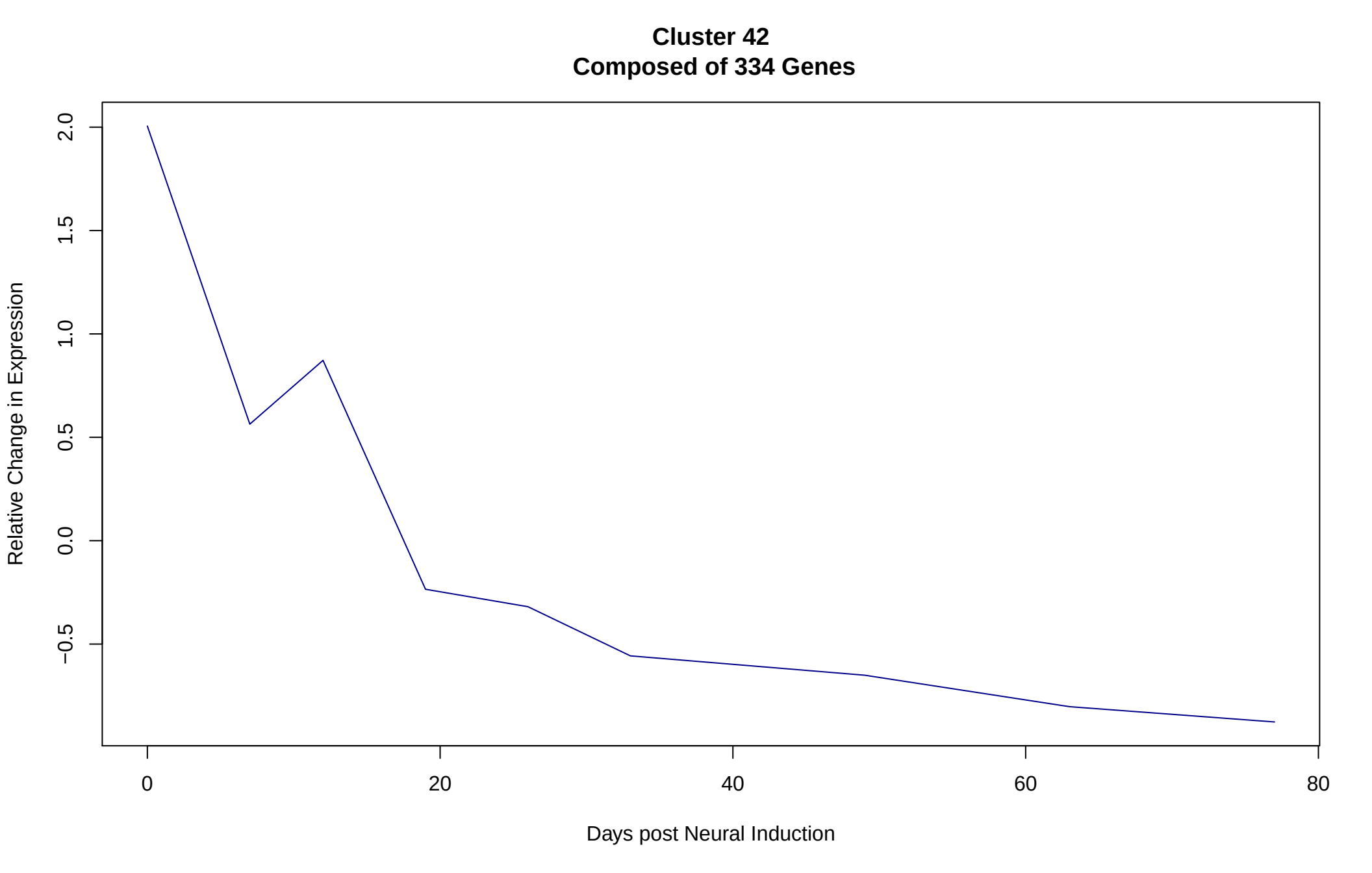

Cluster 37
Composed of 157 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

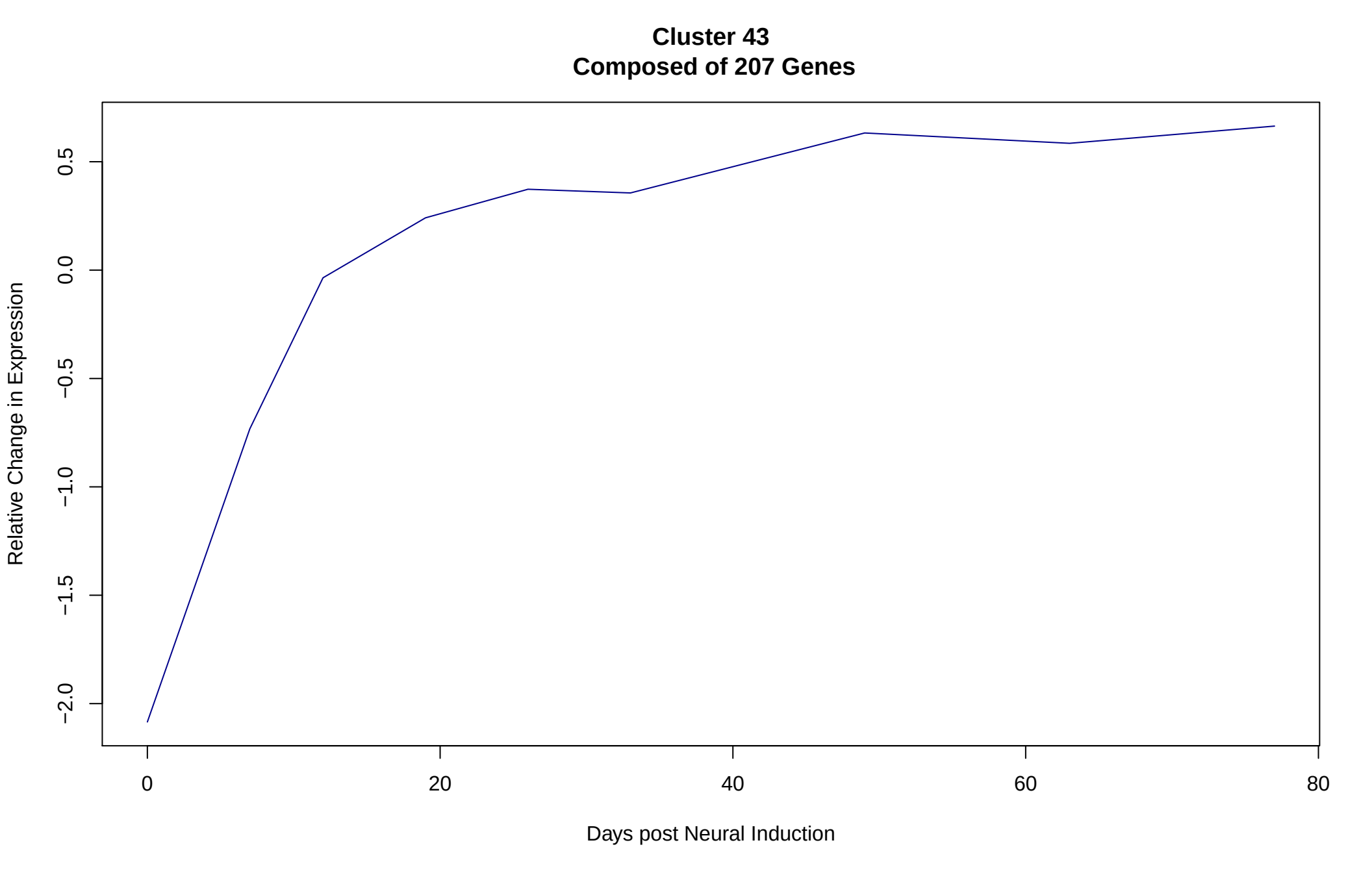

Cluster 38
Composed of 218 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

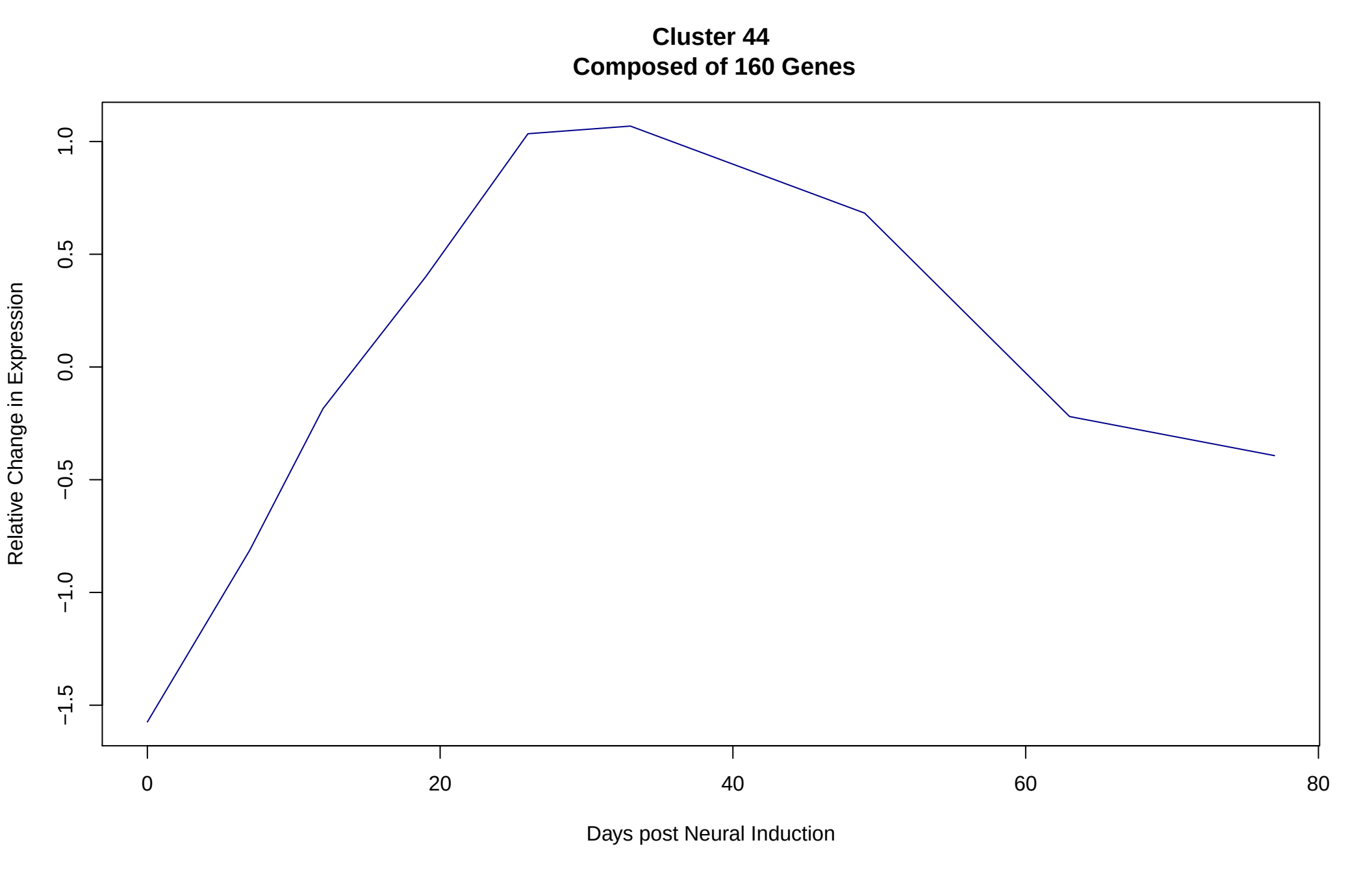

Cluster 39
Composed of 370 Genes
Relative Change in Expression
−0.5
0.0
0.5
1.0
1.5
2.0
0
20
40
60
80
Days post Neural Induction

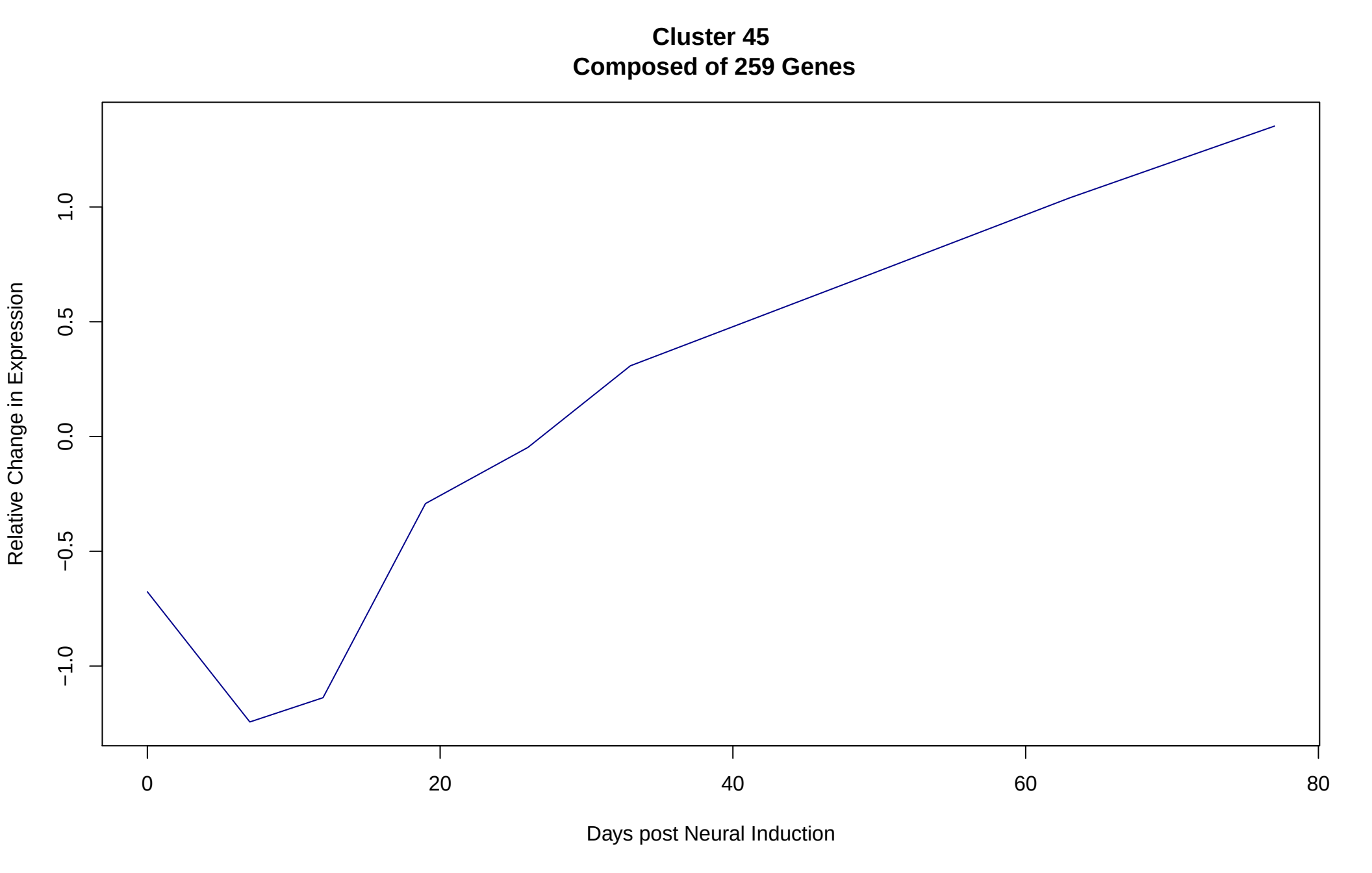

Cluster 40
Composed of 250 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

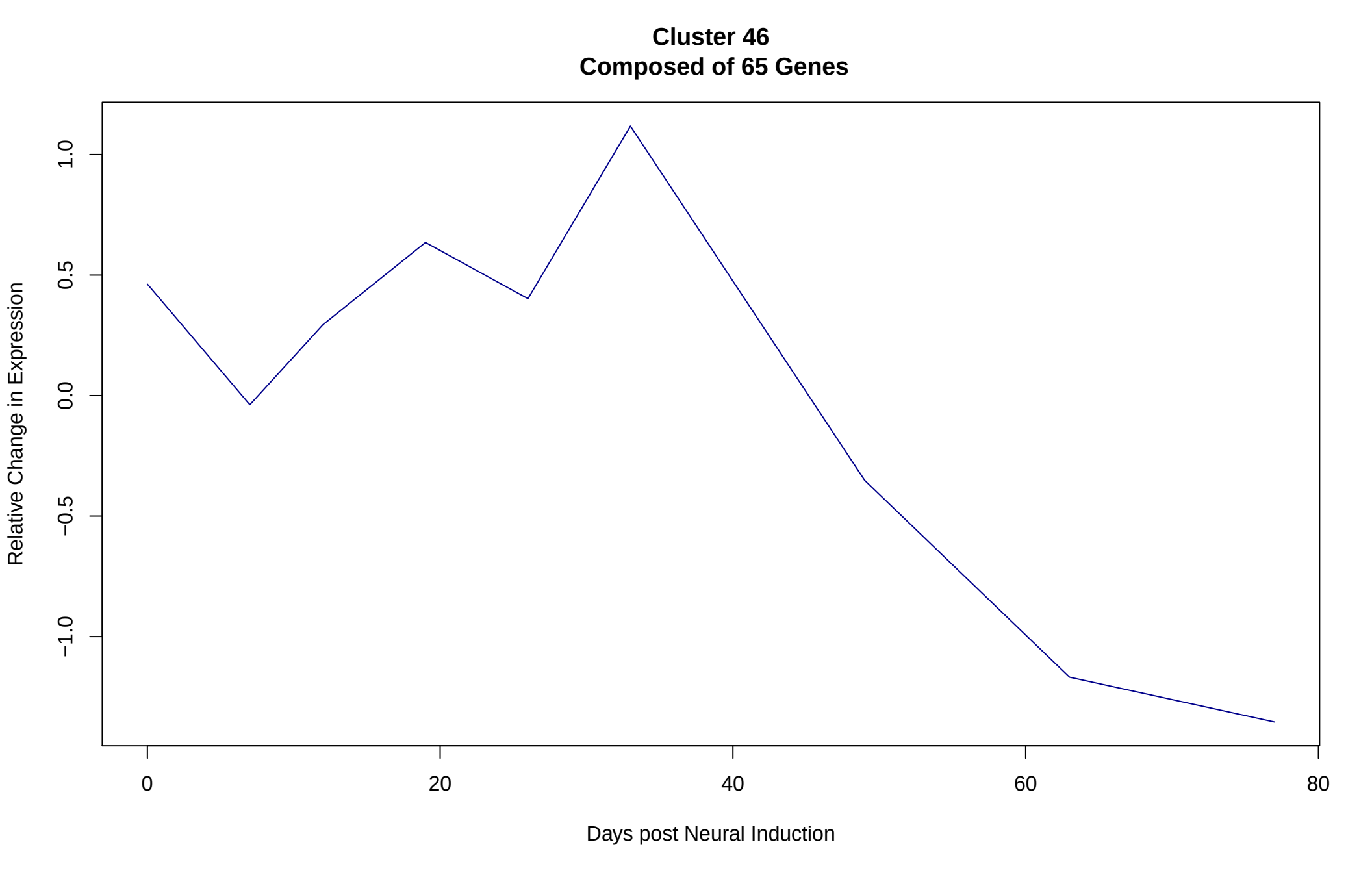

Cluster 41
Composed of 230 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

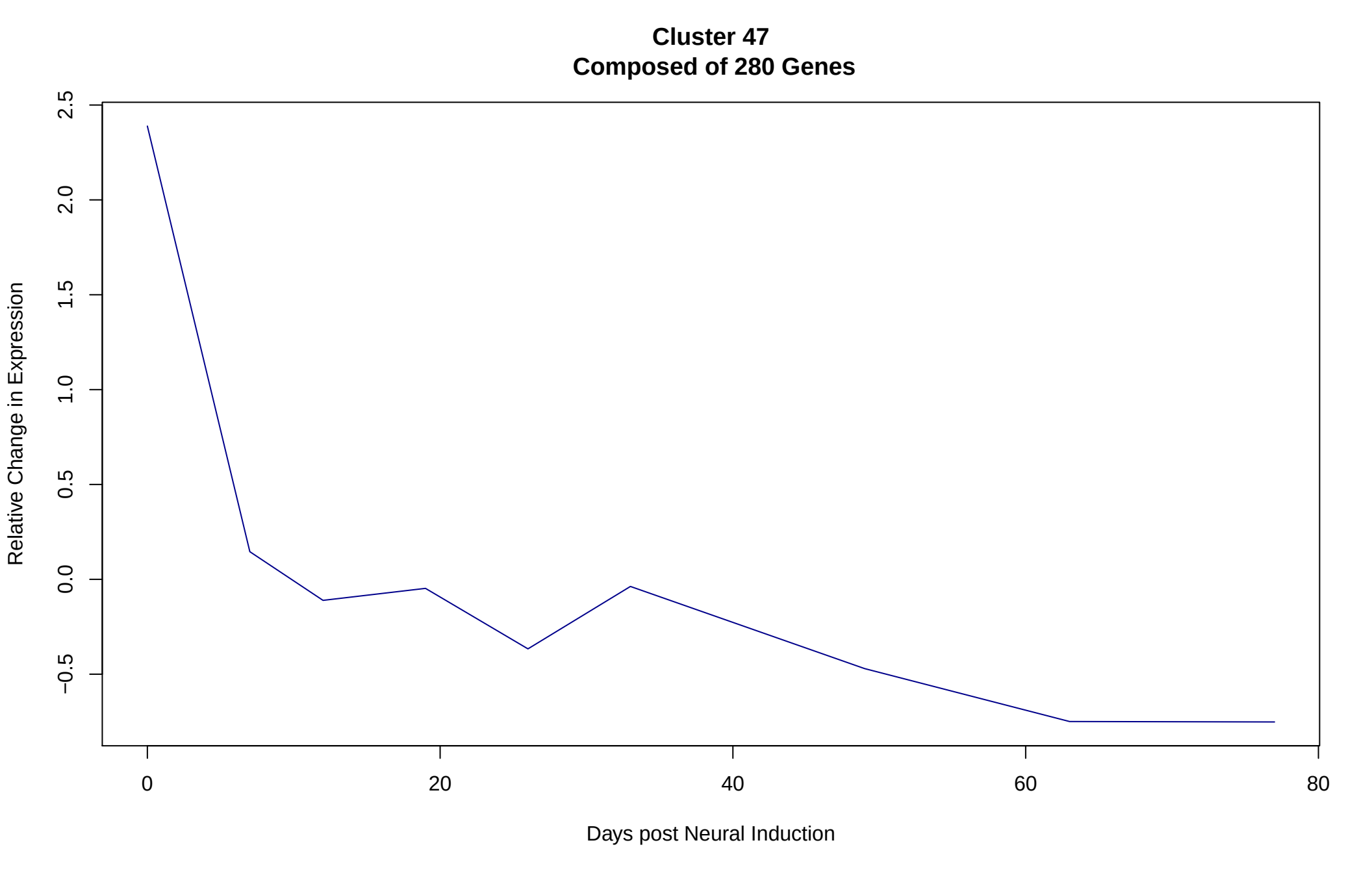

Cluster 42
Composed of 334 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

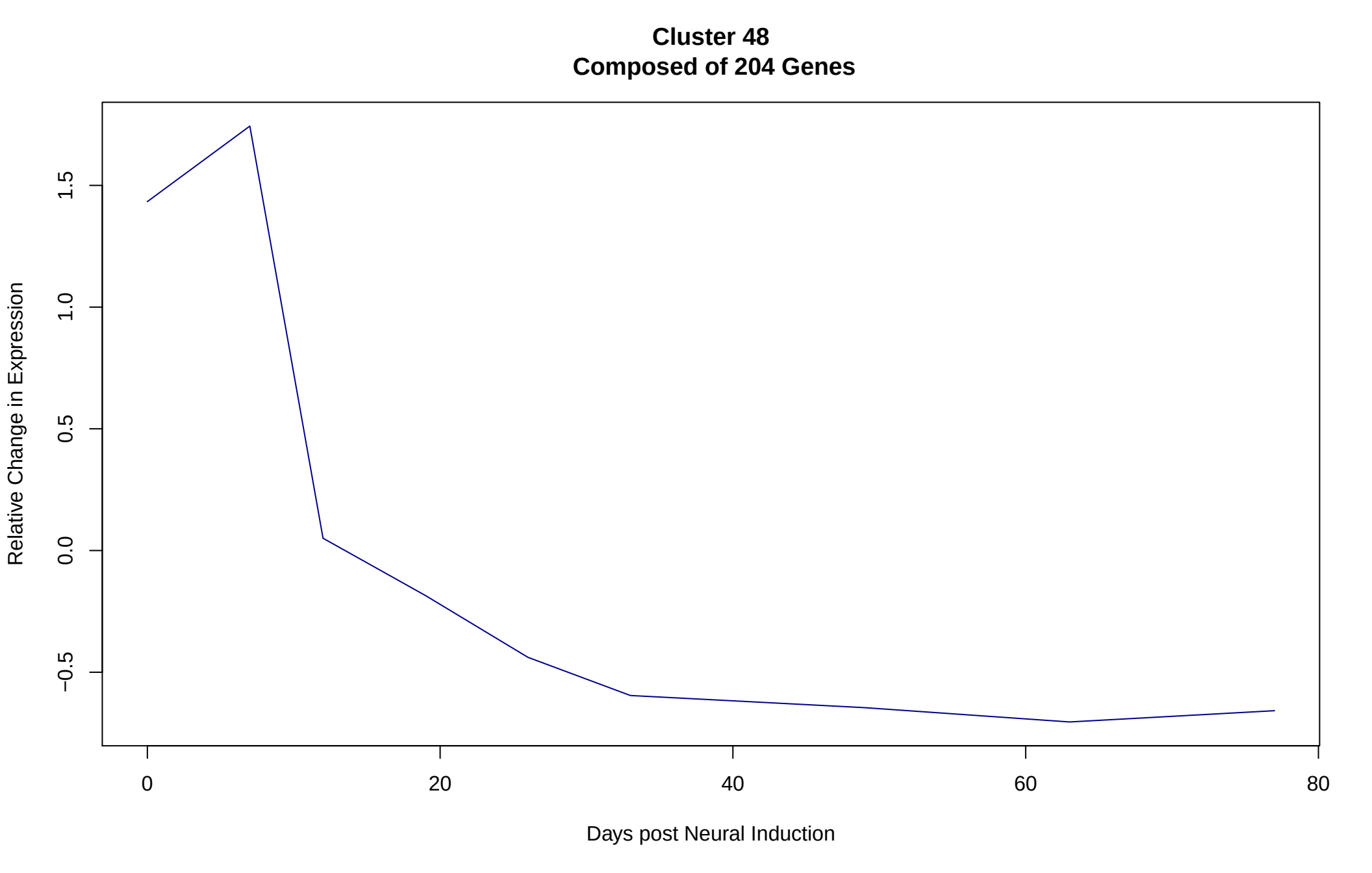

Cluster 43
Composed of 207 Genes
Relative Change in Expression
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

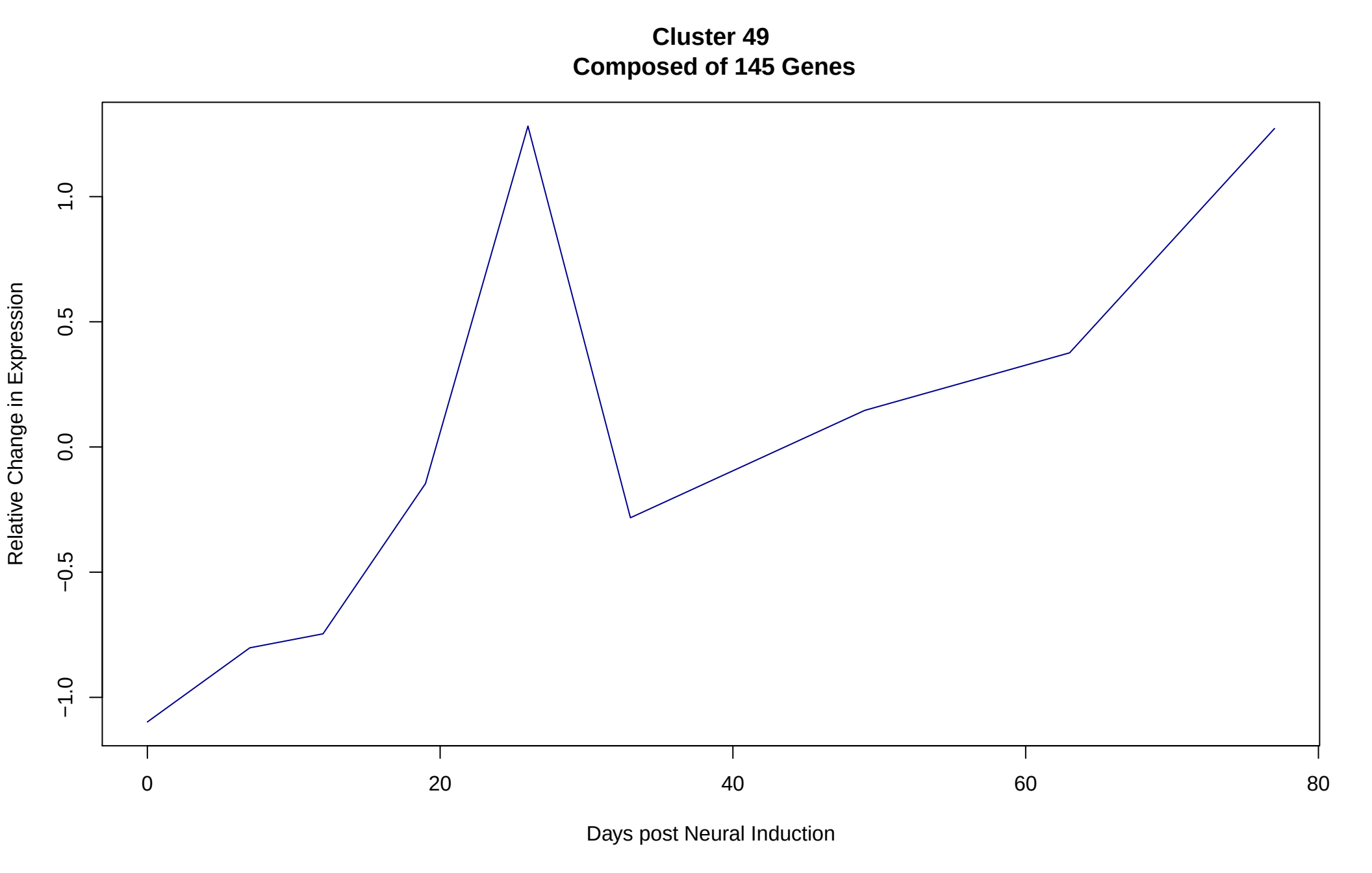

Cluster 44
Composed of 160 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
0
20
40
60
80
Days post Neural Induction

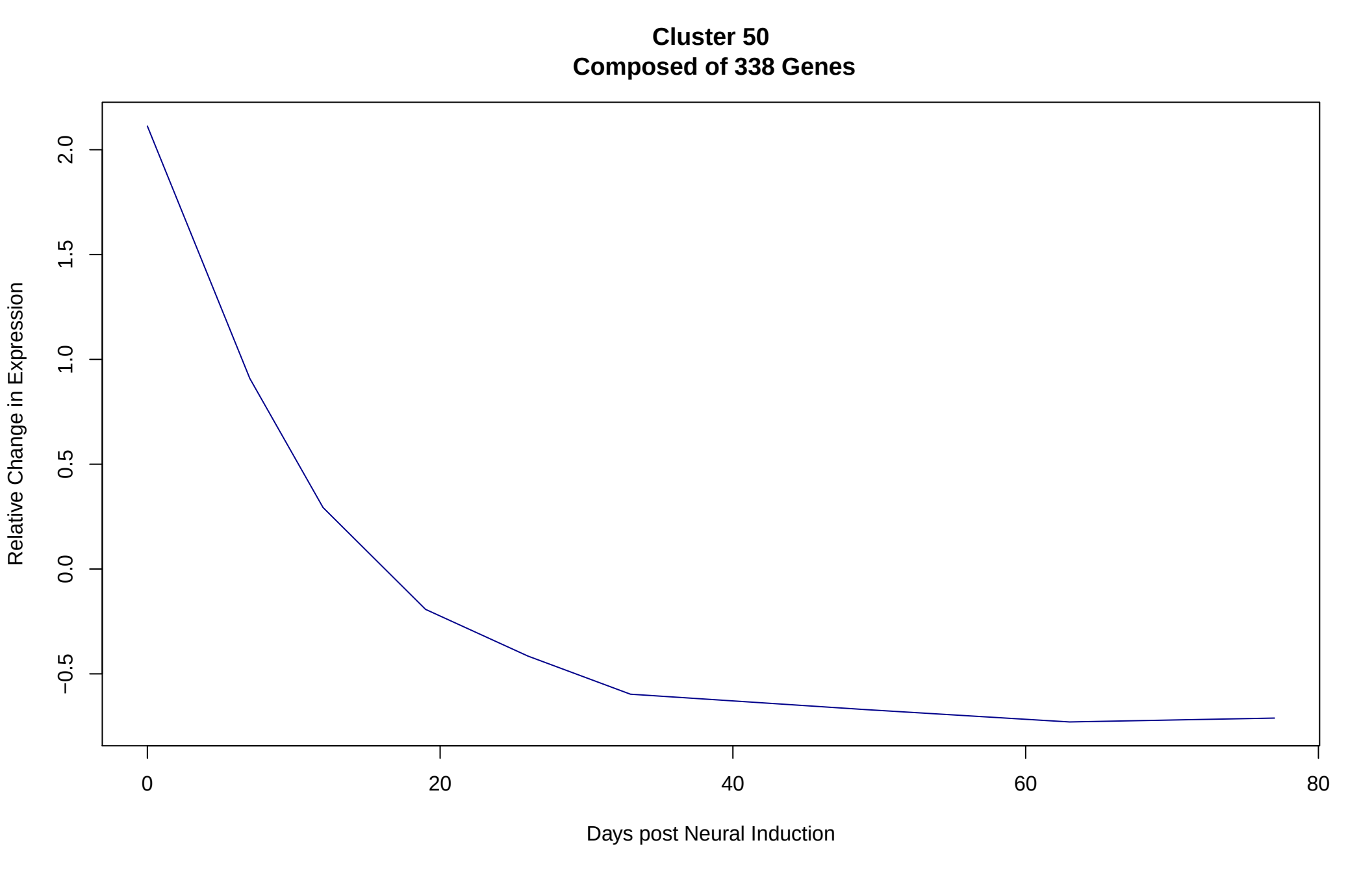

Cluster 45
Composed of 259 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

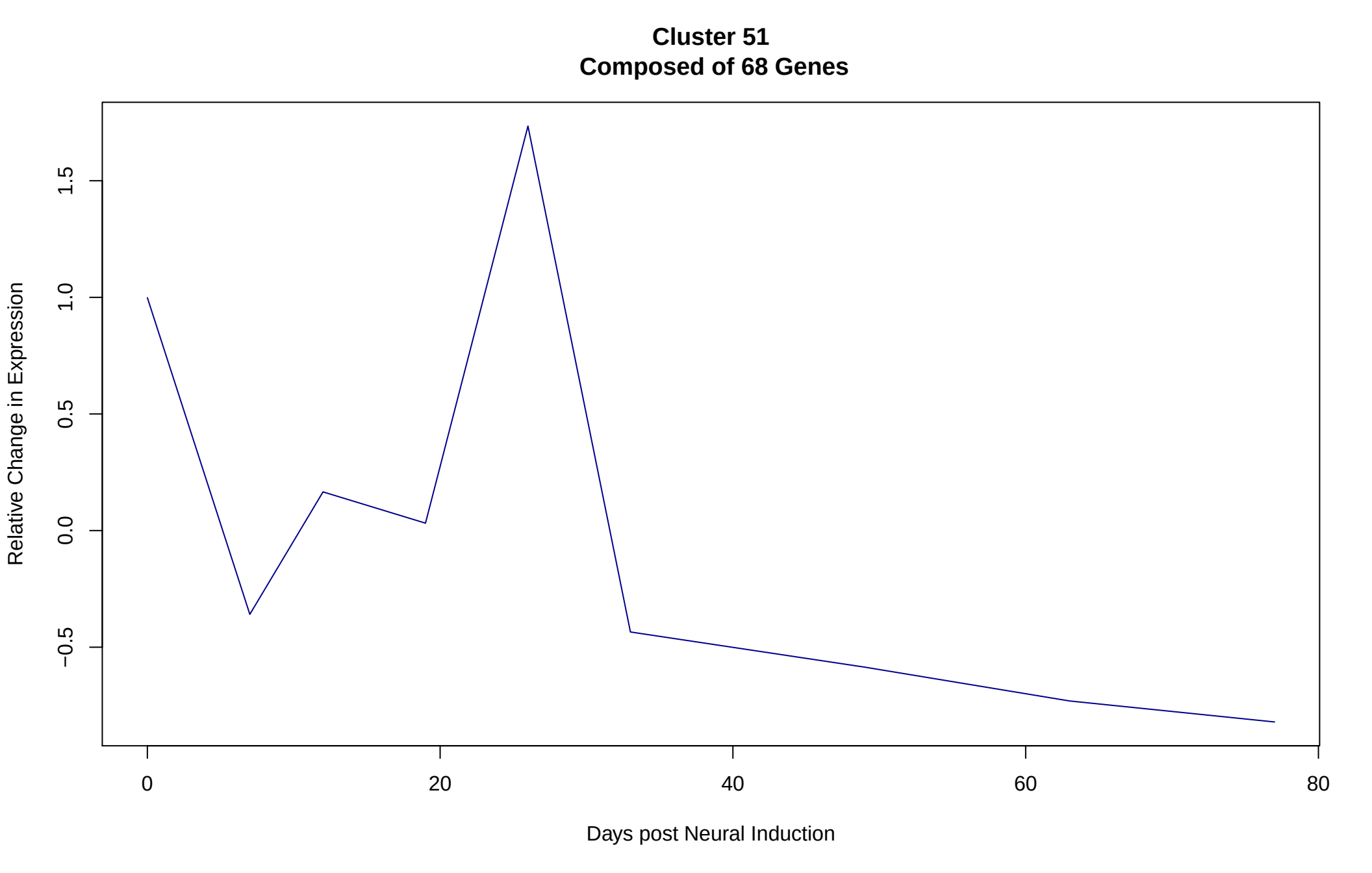

Cluster 46
Composed of 65 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

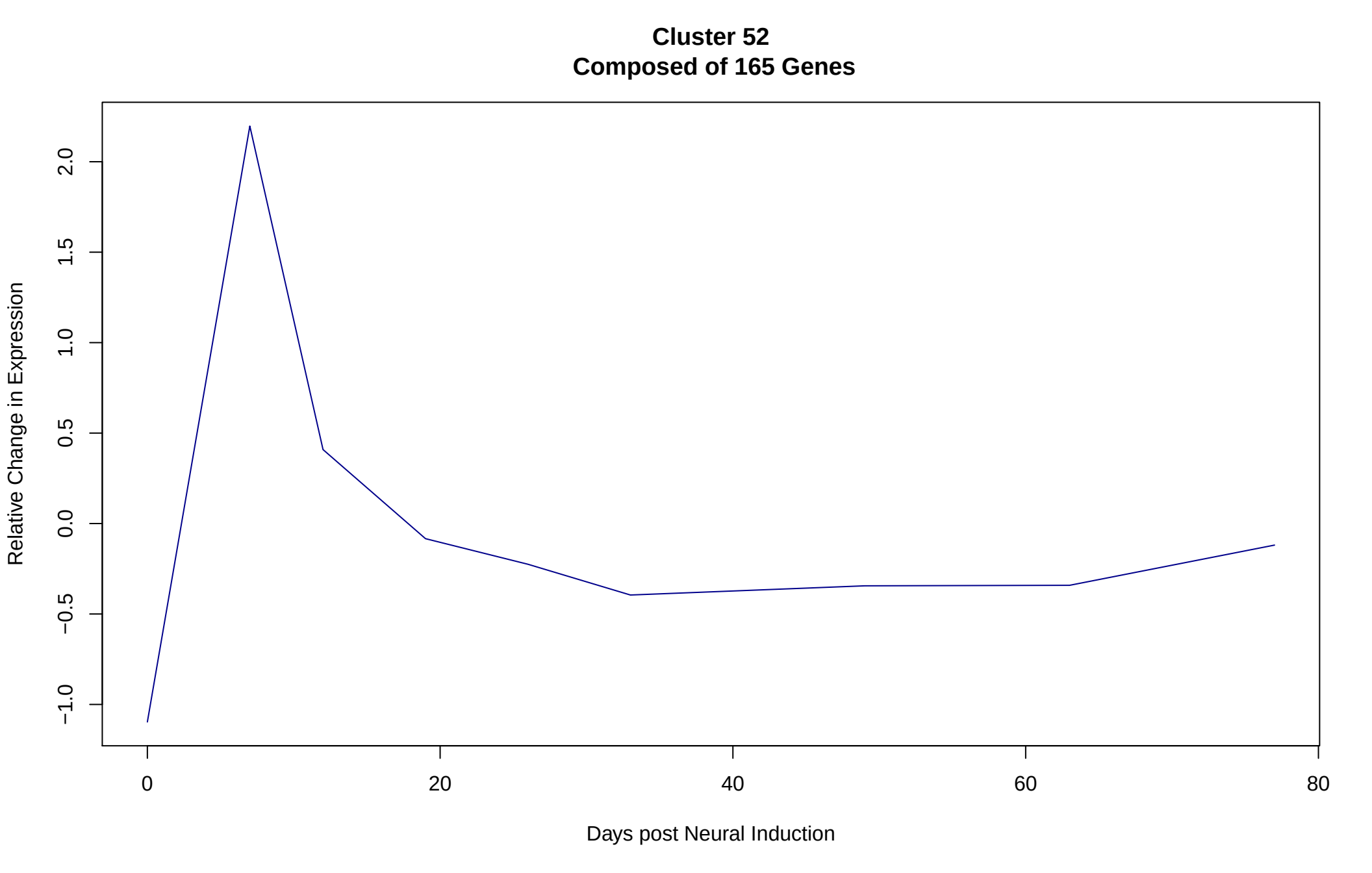

Cluster 47
Composed of 280 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

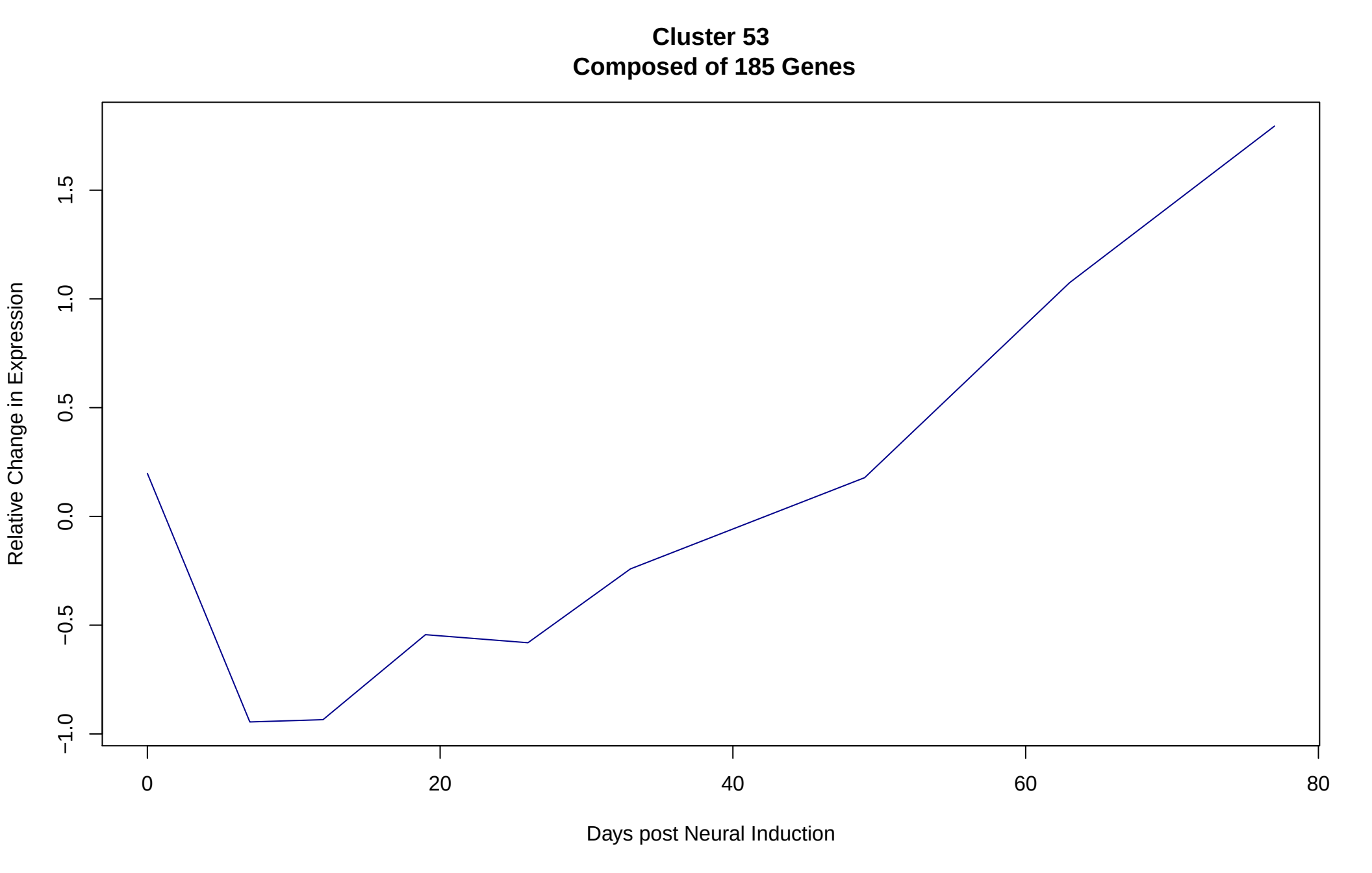

Cluster 48
Composed of 204 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

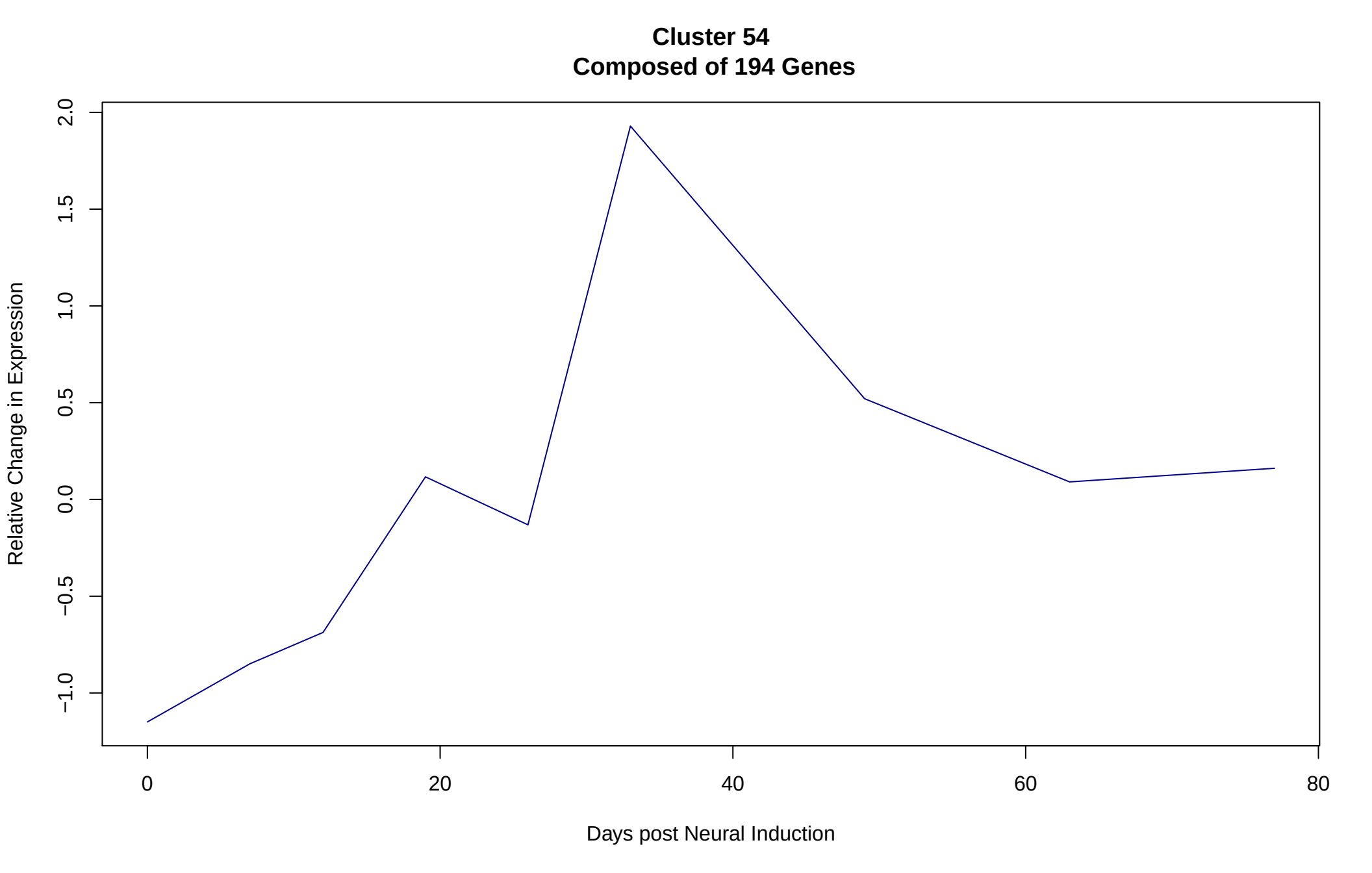

Cluster 49
Composed of 145 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

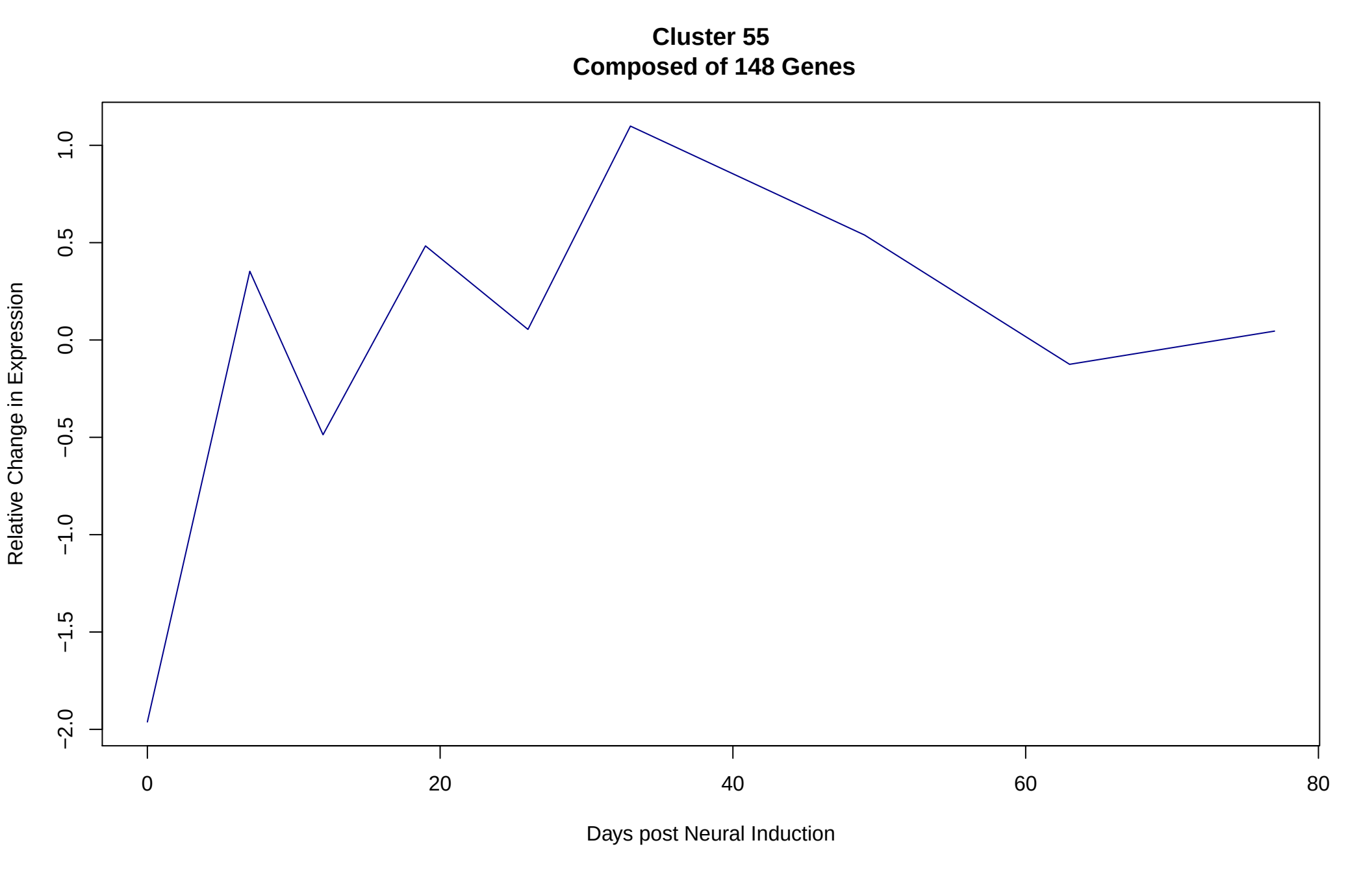

Cluster 50
Composed of 338 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

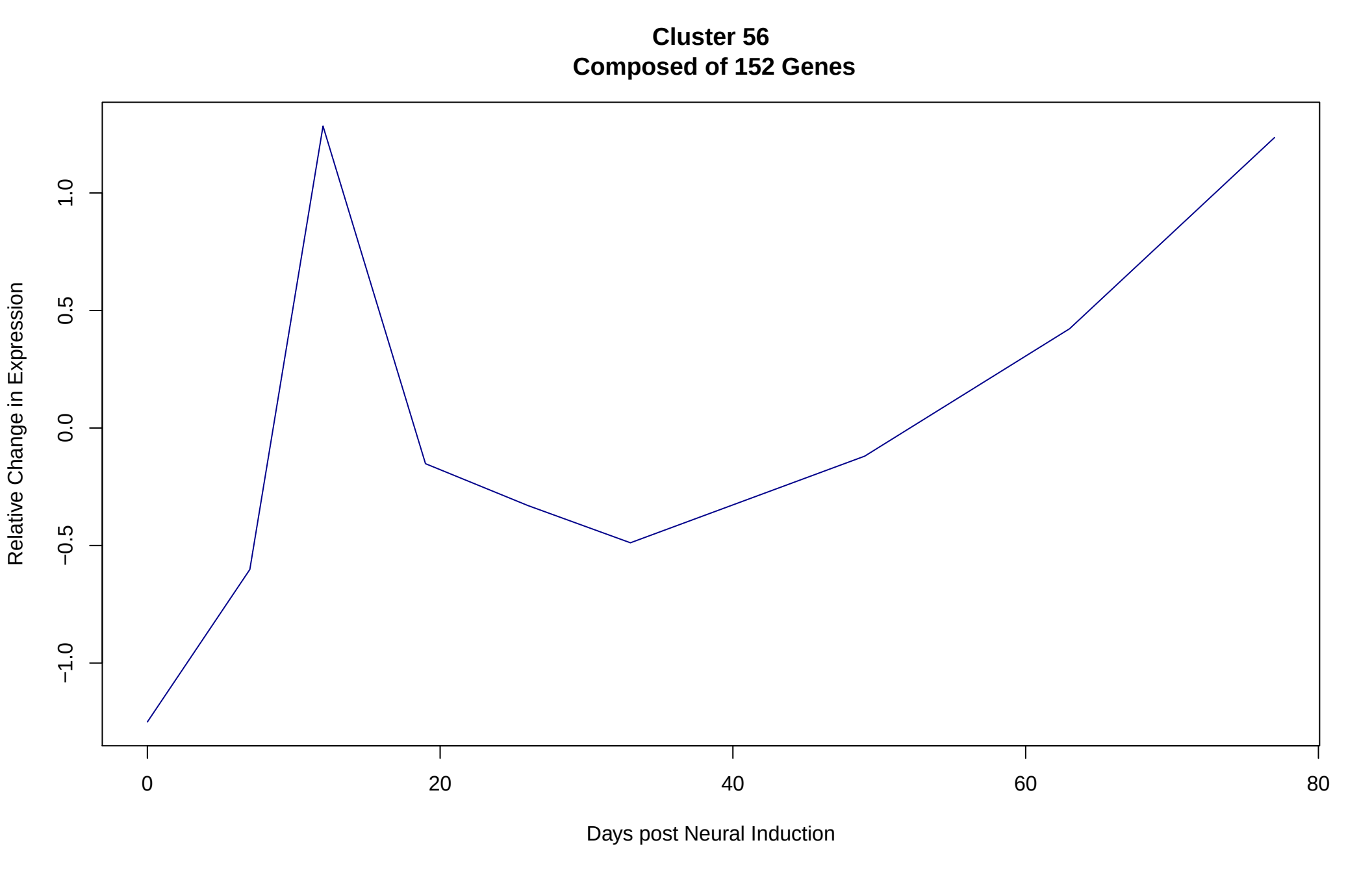

Cluster 51
Composed of 68 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

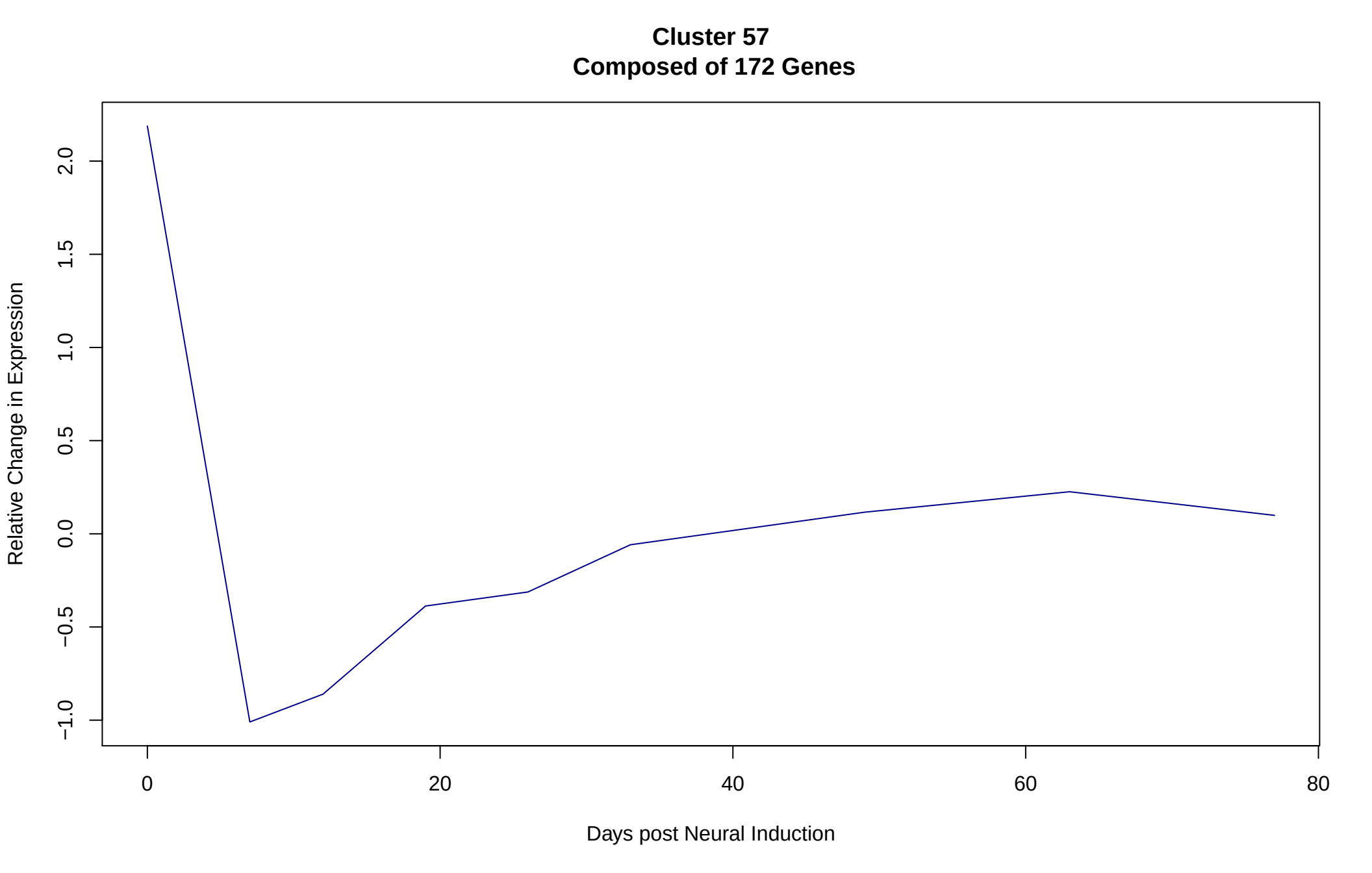

Cluster 52
Composed of 165 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

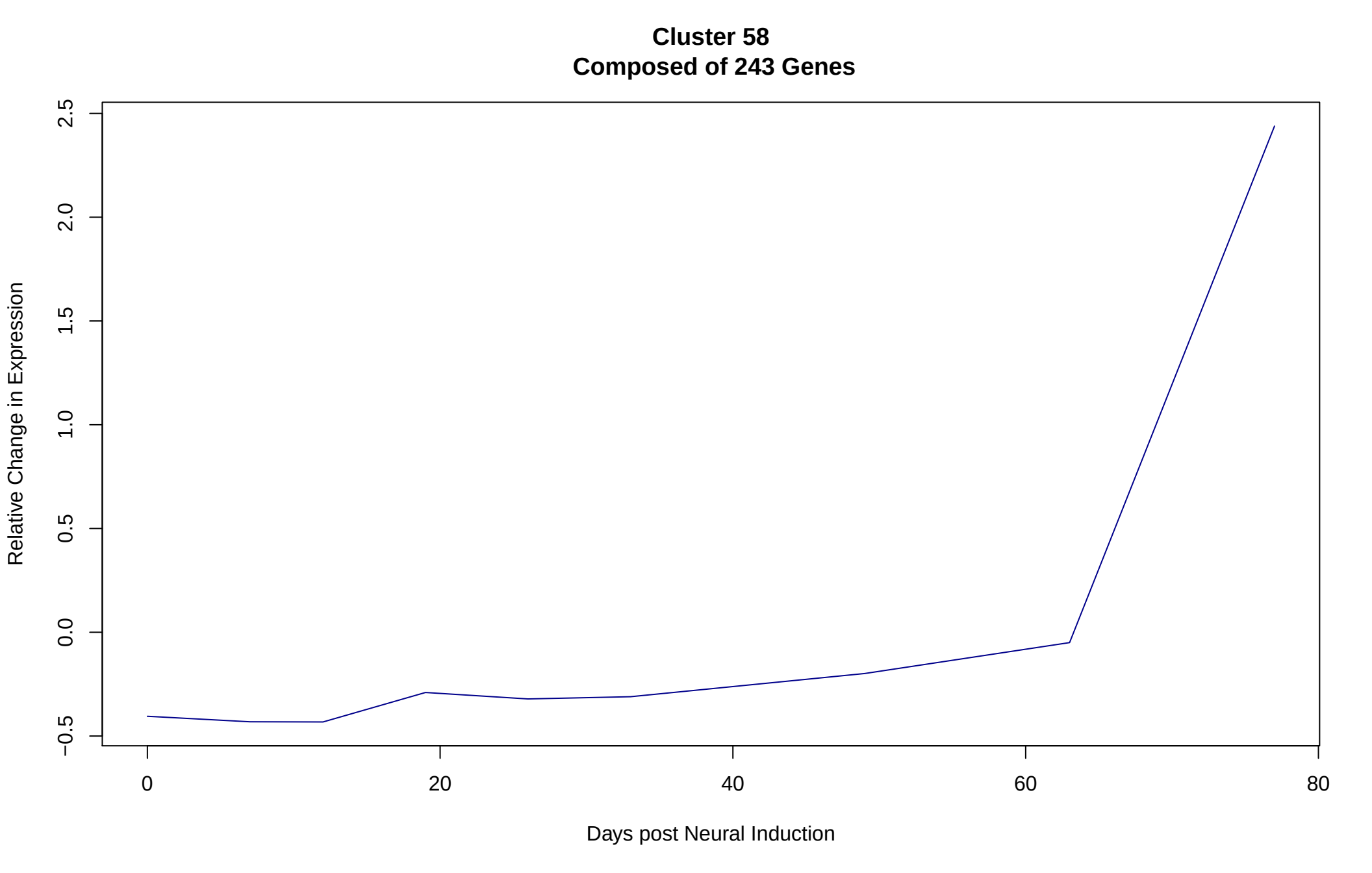

Cluster 53
Composed of 185 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

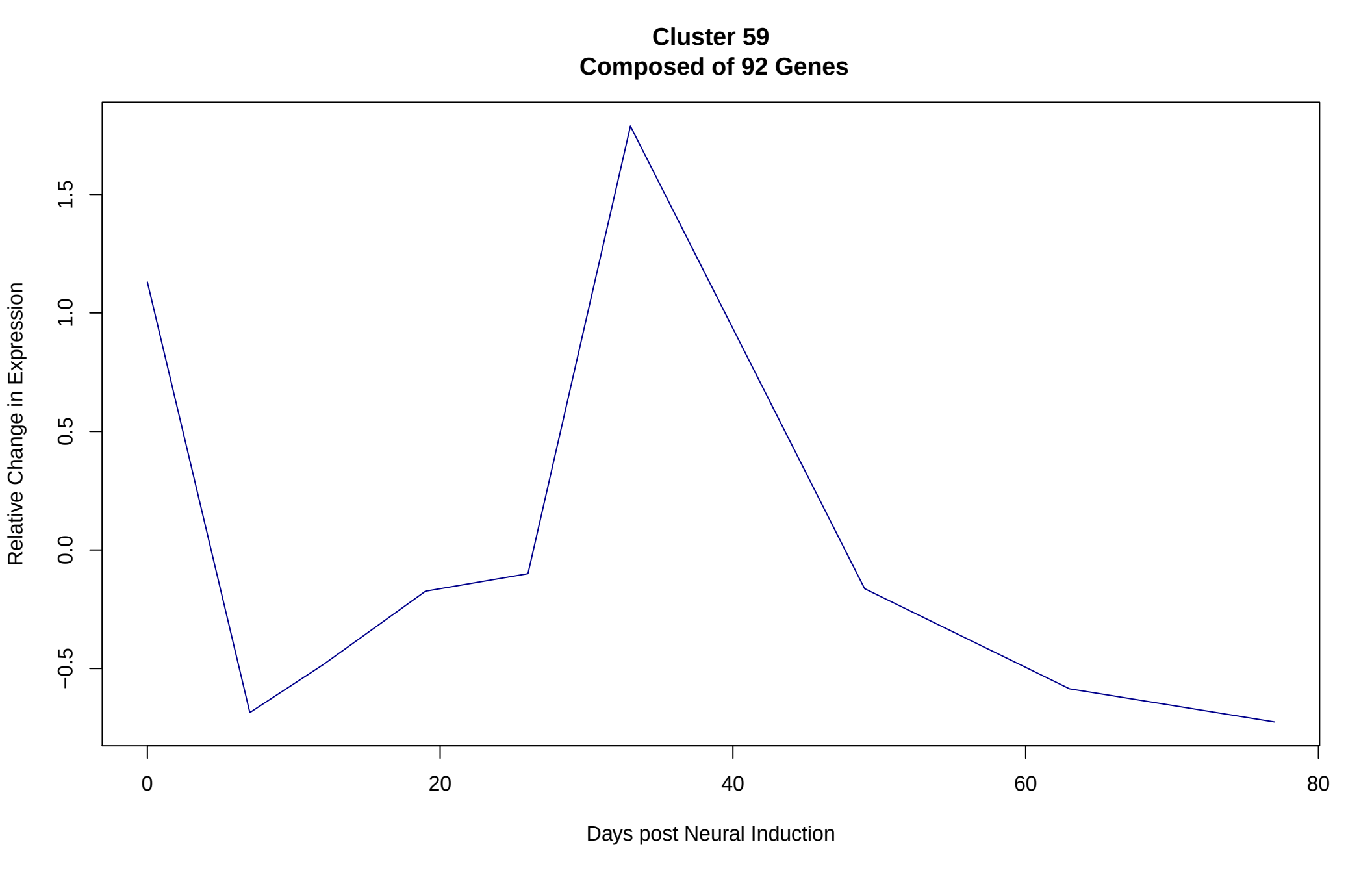

Cluster 54
Composed of 194 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

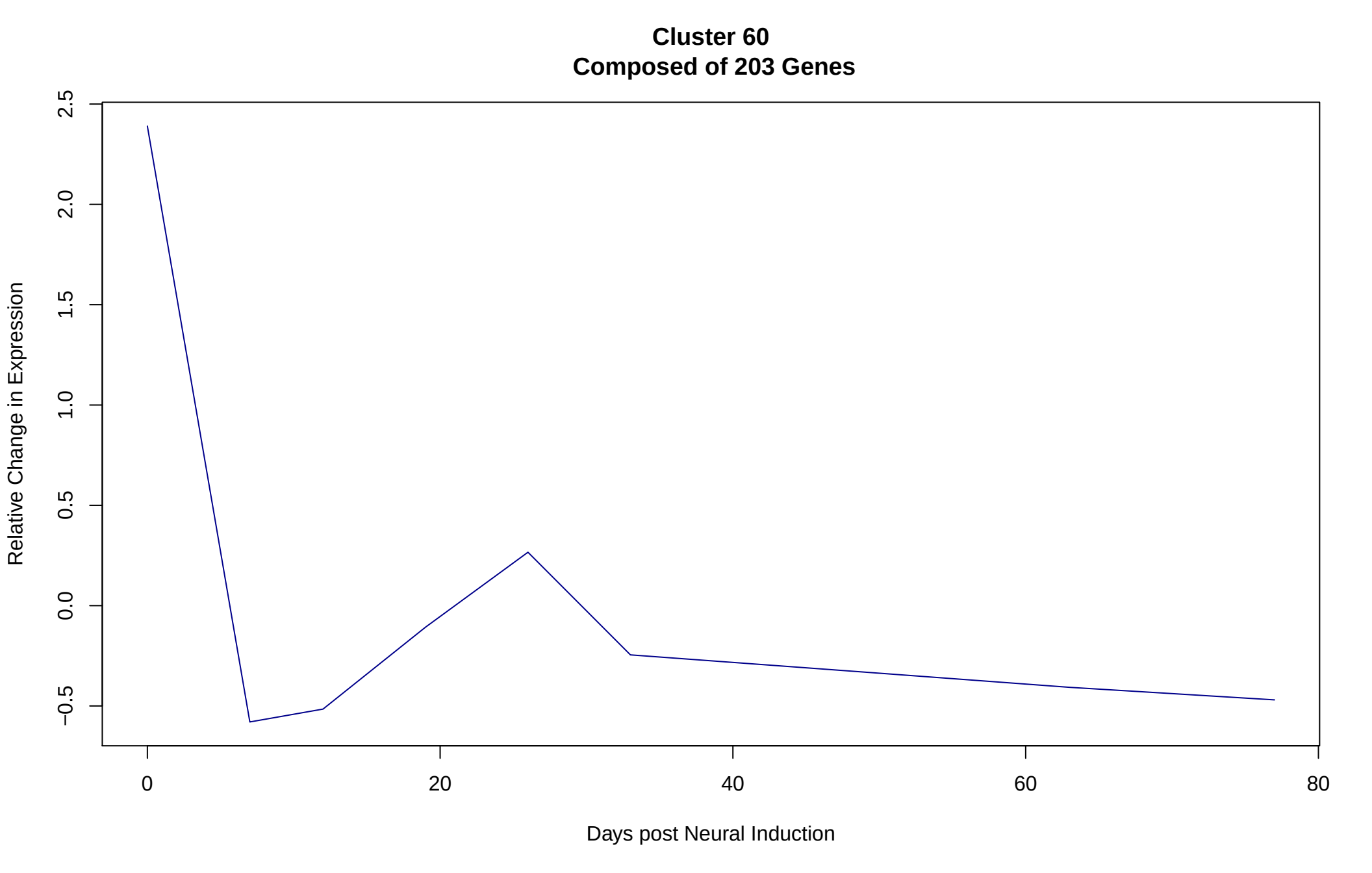

Cluster 55
Composed of 148 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
0
20
40
60
80
Days post Neural Induction

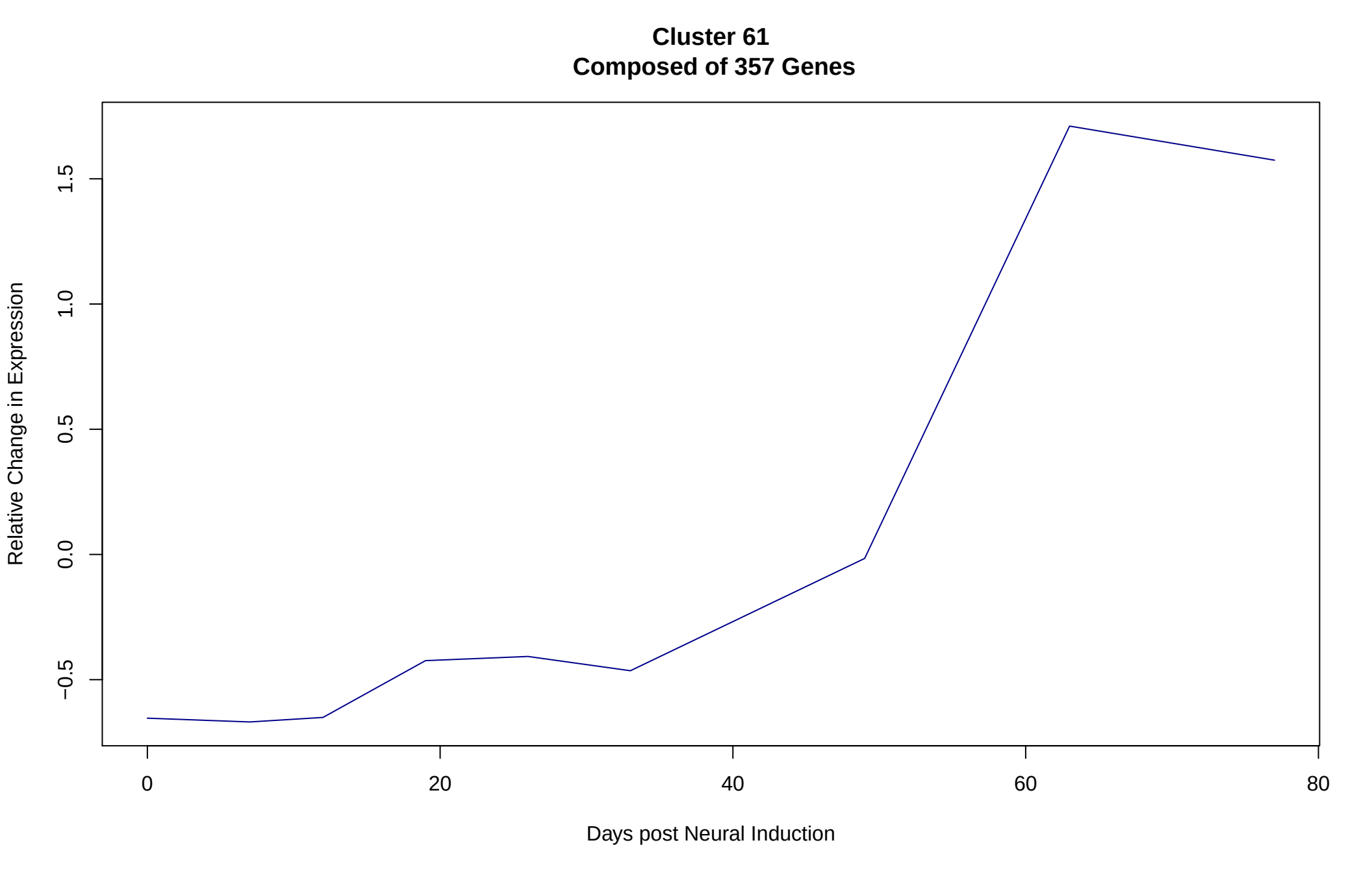

Cluster 56
Composed of 152 Genes
Relative Change in Expression
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

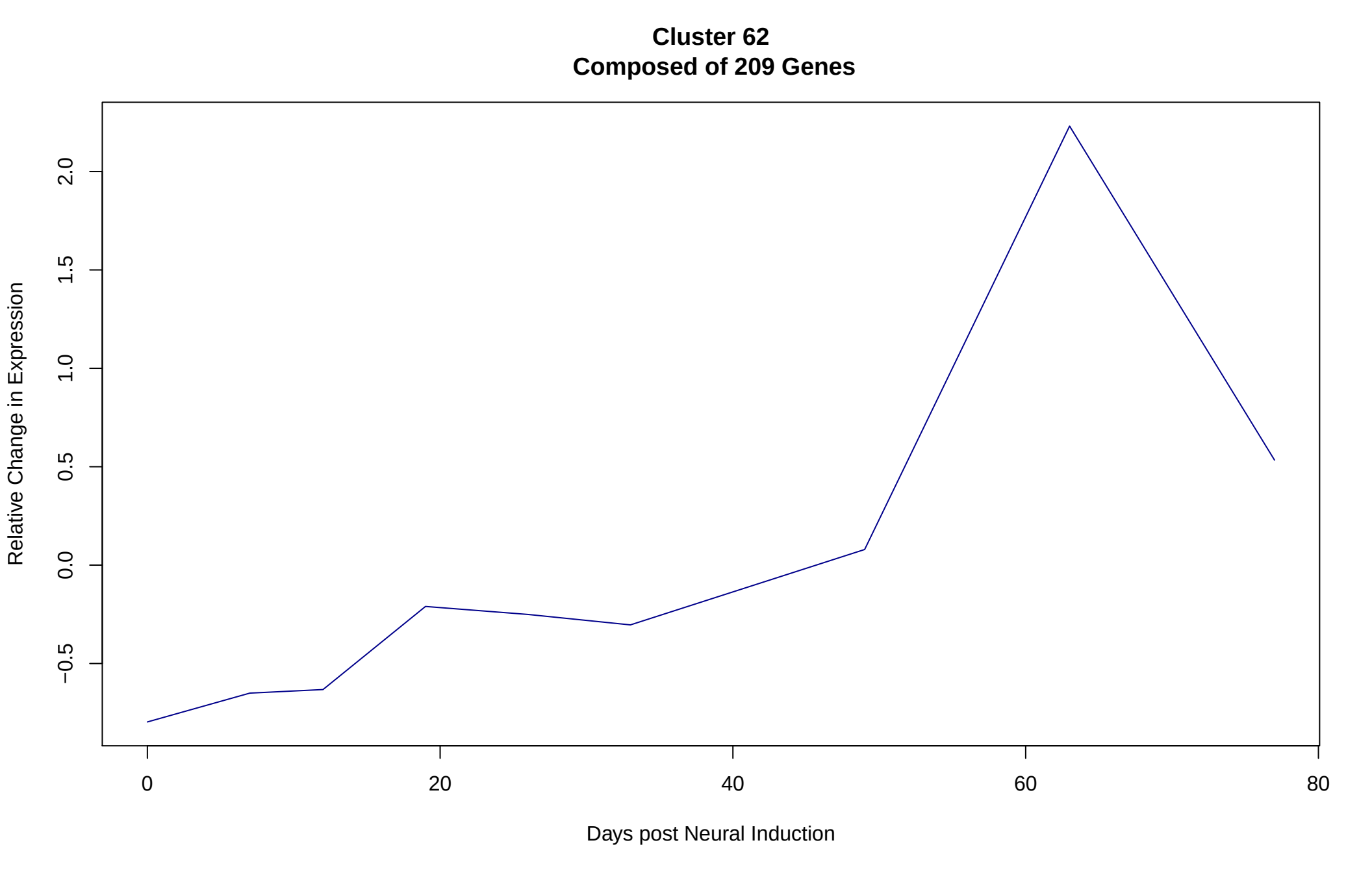

Cluster 57
Composed of 172 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
0
20
40
60
80
Days post Neural Induction

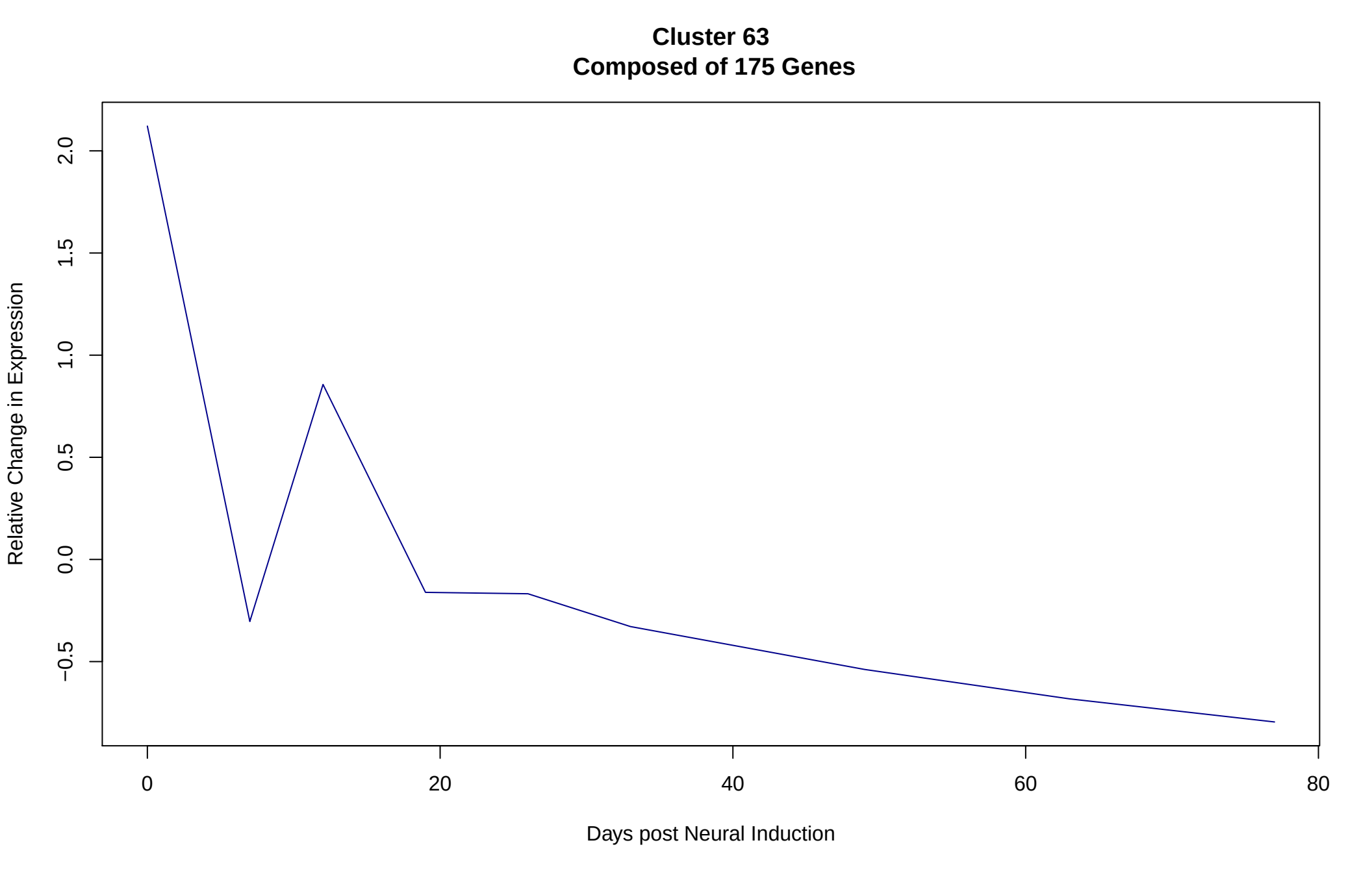

Cluster 58
Composed of 243 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

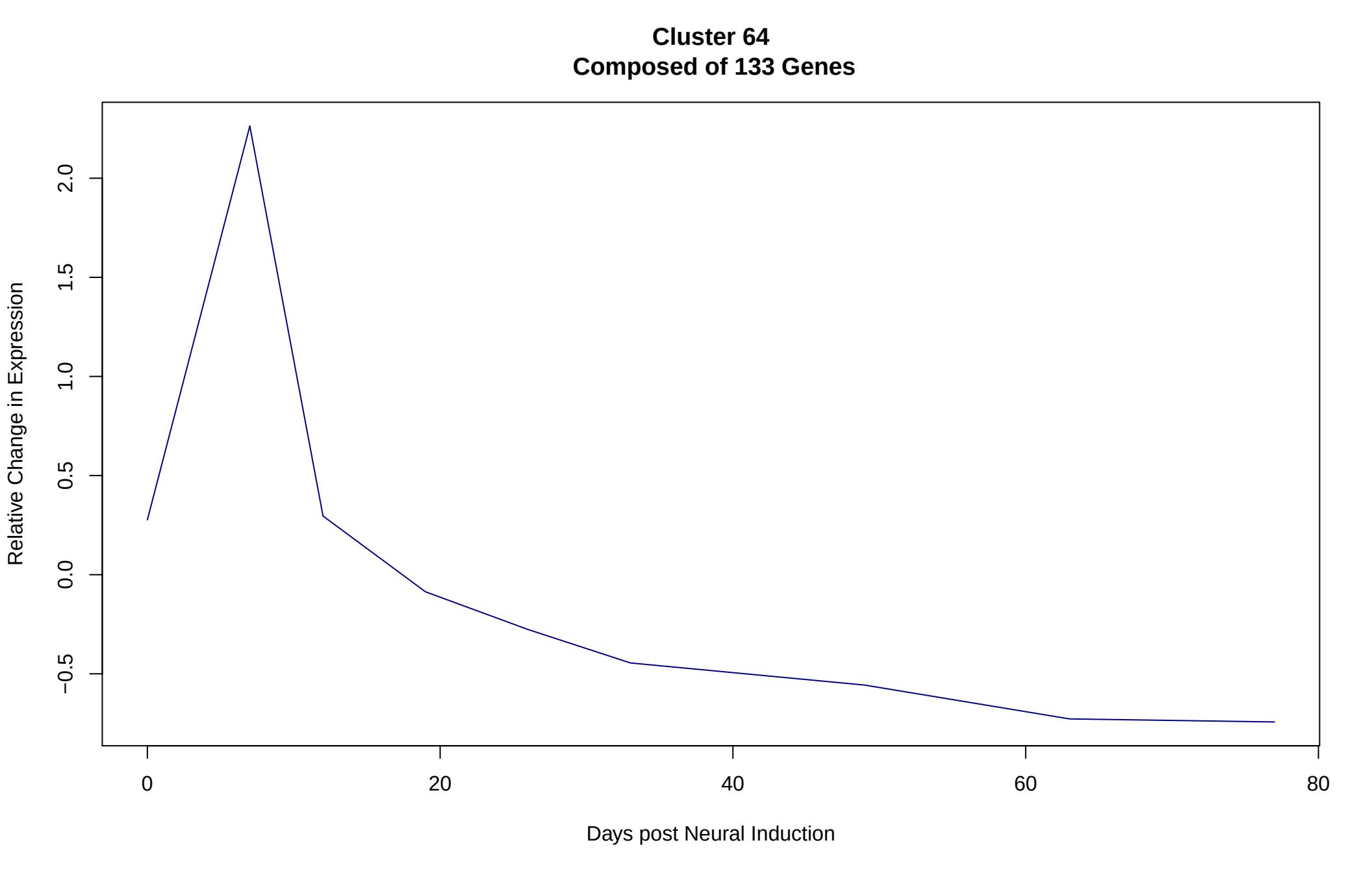

Cluster 59
Composed of 92 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

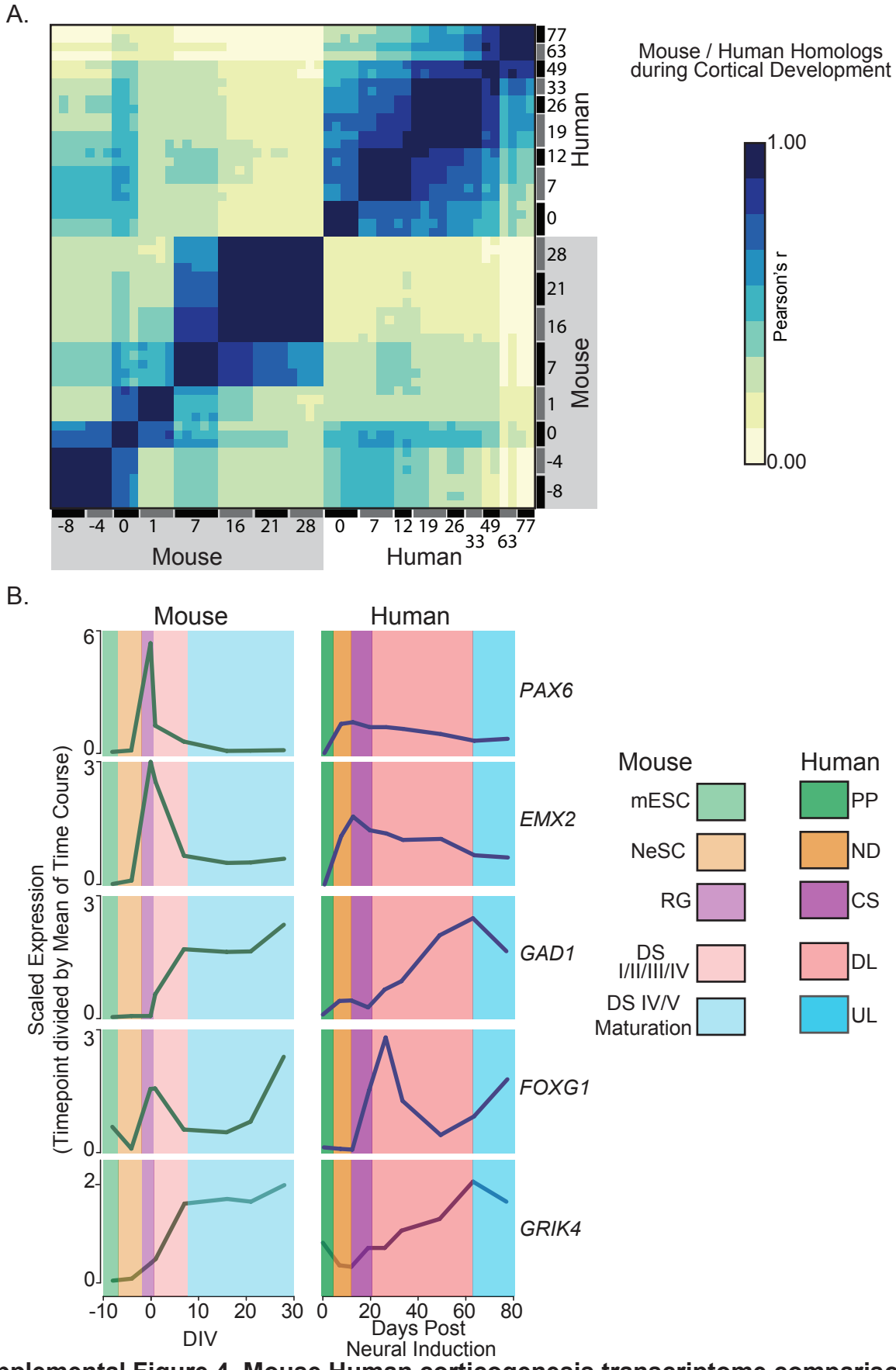

Cluster 60
Composed of 203 Genes
Relative Change in Expression
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction


Cluster 61
Composed of 357 Genes
Relative Change in Expression
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction


Cluster 62
Composed of 209 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction

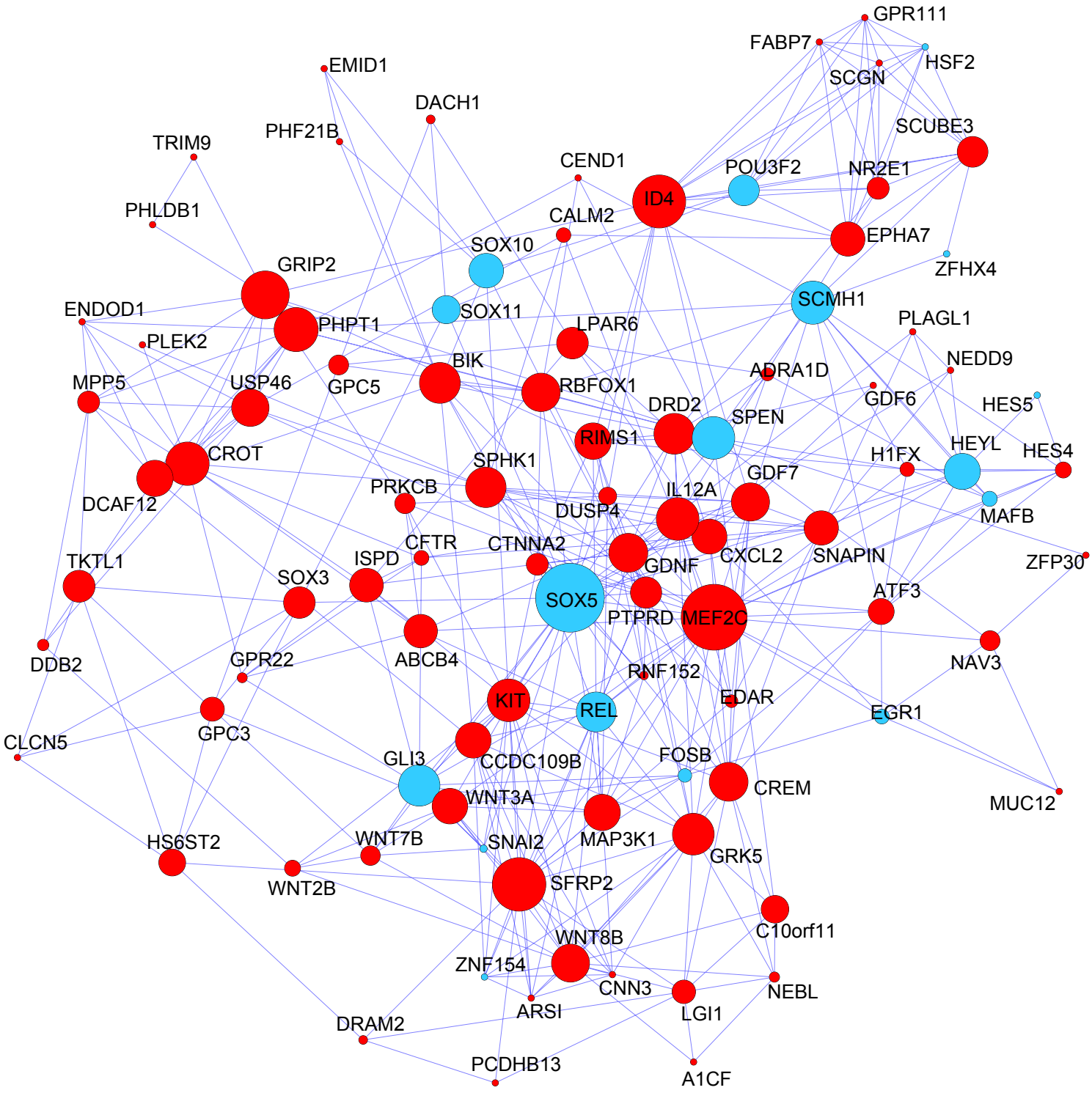

Cluster 63
Composed of 175 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction


Cluster 64
Composed of 133 Genes
Relative Change in Expression
2.0
1.5
1.0
0.5
0.0
−0.5
0
20
40
60
80
Days post Neural Induction



Supplemental Figure 4. Mouse Human corticogenesis transcriptome comparison, related to Figure 7.

Transcriptome data from a mouse in vitro corticogenesis differentiation (Hubbard 2013), was obtained and the human homologs were isolated. (A.) Heatmap of Pearson correlations between the human and mouse datasets containing only the mouse/human homologs.(B.) Comparison of select cortical genes from the mouse and human cortecon data. (mouse embryonic stem cells, **mESC**; neuroepithelial stem cells, **NeSC**; radial glia, **RG**; developmental stage (DS) I/II/III/IV neurons, **DS I/II/III/IV**; maturing DS IV/V neurons,**DS IV/V maturation**; Pluripotency, **PP**; Neural Differentiation, **ND**; Cortical Specification, **CS**; Deep Layer Generation, **DL**; Upper Layer Generation, **UL**).



Supplemental Figure 5. Cluster 18 transcription factor and cluster 34 network, related to Figure 6. A subgraph was created from transcription factors of cluster 18 (shown in blue) and cluster 34 (shown in red). SOX5 has the highest betweenness and closeness centrality scores of the subgraph indicating it as a key gene in this network. Size of node represents betweenness centrality score with larger nodes having higher scores.

Supplemental Experimental Procedures

Human Embryonic Stem Cell Culture

Human ES cells (H9, WiCell) were maintained on irradiated mouse embryonic fibroblasts (MEFs; GobalStem) according to standard culture procedures(Fasano et al., 2010). H9 cells were cultured in HES-medium (all reagents were obtained from Life Technologies): DMEM/F12 (1:1) with 20% (v/v) KSR, 1 mM L-glutamine, 100 µM non-essential amino acids, 55 µM 2-mercaptoethanol, and 10 ng ml^{-1} FGF-2.

Neural Induction Human ESCs

Human ES cells were dissociated to singe cells with dispase (StemCell), and isolated from MEFs by a 1hr, 37°C incubation on gelatin (StemCell)-coated plates in HES-medium. Non-adherent hESCs were washed and plated onto matrigel (BD biosciences)-coated 12-well or 24-well plates in HES-medium conditioned on MEFs, supplemented with ROCK inhibitor (Stemgent). When cultures were around 90% confluent, induction of neural progenitors was initiated using an adaptation of the 12 day protocol (Chambers et al., 2009). The culture medium was changed to KSR-medium (all reagents were obtained from Life Technologies unless indicated otherwise): knockout-DMEM with 15% (v/v) KSR, 2mM L-glutamine, 100 µM non-essential amino acids, 55 µM 2-mercaptoethanol, supplemented with 100 nM LDN193189 (Stemgent) and 10 µM SB431542 (Tocris). Medium was changed daily, on the third day 1µM cyclopamine (Stemgent) was added. Medium was changed every other day, as of day five increasing amounts of N2-medium (25%, 50%, 75%, 100%) was added, cyclopamine concentration was maintained at 1 µM. N2-medium (all reagents were obtained from Life Technologies unless indicated otherwise): DMEM/F12 (1:1) with N-2 supplement and 0.775 g glucose (Sigma).

Cortical Differentiation Human ESCs

On day 13 of initial neural induction, medium was changed to a culture medium supporting cortical differentiation, adapted from Gaspard et al. (Gaspard et al., 2008), and changed every other day. N2/B27-medium consists of equal parts N2 and B27. N2 (all reagents were obtained from Life Technologies): DMEM/F12 (1:1) with N-2 supplement, 2 mM L-glutamine, 0.5 mg ml^{-1} bovine albumin fraction V solution, 110 µM 2-mercaptoethanol, supplemented with 10 ng ml^{-1} FGF-2. B27 (all reagents were obtained from Life Technologies): neurobasal with B-27 supplement, 2 mM L-glutamine, supplemented with 10 ng/ml FGF-2 to promote cortical progenitor proliferation and neurogenesis (Qian et al., 1997)

Quantitative RT-PCR

Samples were collected using RNAprotect cell reagent (Qiagen) and total RNA extracted using RNeasy plus mini kit (Qiagen), according to the manufacturer's protocol. cDNA was generated using iScript reverse transcription supermix (Biorad), and quantitative RT-PCR performed using SsoFast probes supermix (Biorad) with TaqMan gene expression assays (Life Technologies) on a Mastercycler ep realplex^2 S (Eppendorf) system, all reactions were carried out according to the manufacturer's protocol. VIC-labeled HPRT was used as an endogenous reference gene in multiplex assays with FAM-labeled gene of interest: *DLX1* (Hs00698288_m1*), *EMX2* (Hs00244574_m1*), *FOXG1* (Hs01850784_s1), *NKX2-1* (Hs00163037_m1*), *OTX2* (Hs00222238_m1), *PAX6* (Hs01088112_m1).

Immunocytochemisty

Cultures were fixed in 4% (w/v) paraformaldehyde (Santa Cruz Biotechnology, Inc.) in phosphate-buffered saline (Fisher Bioreagents) for 10 min at room temperature, and washed thrice with phosphate-buffered saline. Fixed cells were incubated for 1 hr at room temperature while gently rocked in block solution: phosphate buffered saline with 3% (w/v) bovine serum albumin (Santa Cruz Biotechnology, Inc.), 0.25% (v/v) triton X-100 (Sigma), 10% (v/v) normal donkey serum (Jackson ImmunoResearch). Fixed cells were incubated overnight at +4°C while gently rocked with

primary antibody diluted in block solution. Next day the fixed cultures were rinsed thrice with phosphate-buffered saline (Fisher Bioreagents), and incubated for 1 hr at room temperature with secondary antibody diluted in block solution. Following, the cultures were rinsed five times with phosphate-buffered saline (Fisher Bioreagents), and counterstained with DAPI nucleic acid stain (Life Technologies). Immunocytochemistry cultures were imaged using a Zeiss AxioObserver.D1 inverted microscope. Primary antibodies used: ADRA2A (abcam, ab85570), DRD5 (abcam, 32620), GRM4 (abcam, ab53088), MKI67 (BD Pharmingen, 550609), NANOG (R&D systems, AF1997), NES (Millipore, AB5922), OTX2 (Millipore, AB9566), PAX6 (DSHB, PAX6), POU3F2 (previously BRN2) (Santa Cruz, sc-6029), POU5F1 (previously Oct4) (Santa Cruz, sc-9081), TUBB3 (Sigma-Aldrich, T8578). Secondary antibodies used were species-specific Alexa Fluor dyes (Jackson ImmunoResearch).

RNA-Seq Application

Samples were collected using RNAprotect cell reagent (Qiagen), homogenized with QIAshredder (Qiagen), and total RNA extracted using RNeasy plus mini kit (Qiagen), according to manufacturer's protocol. Sample quality control was carried out using a 2100 Bioanalyzer (Agilent), only samples with an RNA integrity number (RIN) above 0.9 were selected for RNA-Seq. Selected samples were send out for RNA-Seq application (including library preparation, bar-coding, etcetera) to Expression Analysis (Durham, NC, USA), or Covance (Seattle, WA, USA), samples day 63 only: TruSeq RNA sample prep kit (v2), HiSeq2000 platform (Illumina), paired-end, 50 cycles, multiplex (EA 4-plex; Covance 5-plex).

RNA-Seq Data Analysis

FASTQ files was mapped (80-85%) to human genome 19 using BFAST(Homer et al., 2009). Mapped files were analyzed utililzing R (Team, 2012) in conjunction with multiple packages: GenomicFeatures, rtracklayer, Rsamtools, GenomicRanges, edgeR, Mfuzz, DESeq, DESeq2, goseq, SPIA, ggplot2, igraph, RcolorBrewer, GO.db, org.Hs.eg.db, foreach, parallel, multicore,

doMC ((Anders and Huber, 2010; Anders et al., 2012; Csardi and Nepusz, 2006; Kumar and M, 2007; Lawrence et al., 2009; Lawrence et al., 2013; Robinson et al., 2010; Tarca et al., 2009; Wickham, 2009; Young et al., 2010). For display purposes, data was normalized for library size using the DEseq package. All R code and sessions are available for download from the Corecton website (http://cortecon.neuralsci.org/).

Network creation and analysis

We created a network of gene associations based on mappings between Entrez gene ids and Pubmed ids. These mappings are available in the org.Hs.eg.db R package (Carlson et al.). Using this package, a network was constructed by connecting genes (the nodes in the network) to one another if they were co-cited in published articles (edges in the network). Connections between genes were weighted depending on the number of co-citations to strengthen confidence in the associations. Constructing the network in this manner indicates plausible interactions between genes, with these interactions weighted by the cohort of shared literature (Available for download from http://cortecon.neuralsci.org). The resulting network was then filtered against lists of genes belonging to a particular cluster or stage. This filtering yielded subgraphs (smaller networks) containing only genes relevant to each cluster or stage. Graph theoretic measures of “centrality” were then utilized to find important nodes (genes) based on the relationships defined within the structure of these subgraphs. Centrality is one measure of how important a particular node is, indicating how likely it is that a particular node occurs along the most efficient path between two other nodes. On a map, we can compute the centrality of a location of interest based on how likely it is that the shortest route between any two other locations passes through it. In our case, centrality measures the likelihood that a gene in the network is found in a pathway of interactions connecting two other genes. Cytoscape was used for network visualization (Shannon et al., 2003).

Comparison of in vitro data to in vivo Brainspan developmental data

RNAseq data obtained at each time point from the differentiation protocol was compared to the Allen Institute's Brainspan database, a publicly available RNAseq data set from the developing human brain, to identify in vivo brain regions with the highest molecular similarity to the in vitro samples. Two comparisons were performed: a brain-wide comparison and a cortex-specific comparison. For the brain-wide comparison, seven regions were selected from Brainspan in vivo data (Ganglionic eminences, Ventrolateral prefrontal cortex, Hippocampus, Striatum, Amygdala, Thalamus, and Cerebellum), and for the cortex-specific comparison, eleven cortical areas (Ventrolateral prefrontal, Dorsolateral prefrontal, Medial prefrontal, Orbital frontal, Inferior parietal, Inferolateral temporal, Superior temporal, Primary motor, Primary somatosensory, Primary auditory, and Primary visual) were chosen. For most of these regions, we selected data from multiple ages of human brain development, as is available in the Brainspan database.

In order to find a unique molecular signature for each region, the top 300 genes were selected whose expression was enriched in that region as compared to all other regions regardless of time (all brain regions for the brainwide comparison, all cortical regions for the cortex-only comparison; thus the gene sets used to define a signature for the VFC region vary depending on whether it is compared to brain-wide regions or to cortical regions). Genes were ranked based on enrichment p-values, with the restriction that all genes selected had a corrected p-value<0.01. The Spearman correlation was calculated between the expression levels of these genes in each of the *in vitro* samples and in the relevant *in vivo* region, and plotted these values to find the highest correlations between *in vivo* brain regions and *in vitro* time points. It is important to note that the gene sets used to calculate the correlation are different for each brain region, although there is some overlap. As different genes are enriched in each region, there is no unified gene set that was used to generate all these correlations.

Supplemental References

Anders, S., and Huber, W. (2010). Differential expression analysis for sequence count data. Genome biology *11*, R106.
Anders, S., Reyes, A., and Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. Genome research *22*, 2008-2017.
Carlson, M., Falcon, S., Pages, H., and Li, N. org.Hs.eg.db: Genome wide annotation for Human (Bioconductor).
Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nature biotechnology *27*, 275-280.
Csardi, G., and Nepusz, T. (2006). The igraph software package for complex network research. InterJournal Complex Systems *1695*.
Fasano, C.A., Chambers, S.M., Lee, G., Tomishima, M.J., and Studer, L. (2010). Efficient derivation of functional floor plate tissue from human embryonic stem cells. Cell stem cell *6*, 336-347.
Gaspard, N., Bouschet, T., Hourez, R., Dimidschstein, J., Naeije, G., van den Ameele, J., Espuny-Camacho, I., Herpoel, A., Passante, L., Schiffmann, S.N.*, et al.* (2008). An intrinsic mechanism of corticogenesis from embryonic stem cells. Nature *455*, 351-357.
Homer, N., Merriman, B., and Nelson, S.F. (2009). BFAST: an alignment tool for large scale genome resequencing. PloS one *4*, e7767.
Kumar, L., and M, E.F. (2007). Mfuzz: a software package for soft clustering of microarray data. Bioinformation *2*, 5-7.
Lawrence, M., Gentleman, R., and Carey, V. (2009). rtracklayer: an R package for interfacing with genome browsers. Bioinformatics *25*, 1841-1842.
Lawrence, M., Huber, W., Pages, H., Aboyoun, P., Carlson, M., Gentleman, R., Morgan, M.T., and Carey, V.J. (2013). Software for computing and annotating genomic ranges. PLoS computational biology *9*, e1003118.
Qian, X., Davis, A.A., Goderie, S.K., and Temple, S. (1997). FGF2 concentration regulates the generation of neurons and glia from multipotent cortical stem cells. Neuron *18*, 81-93.
Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139-140.
Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B., and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome research *13*, 2498-2504.
Tarca, A.L., Draghici, S., Khatri, P., Hassan, S.S., Mittal, P., Kim, J.S., Kim, C.J., Kusanovic, J.P., and Romero, R. (2009). A novel signaling pathway impact analysis. Bioinformatics *25*, 75-82.
Team, R.C. (2012). R: A Language and Environment for Statistical Computing.
Wickham, H. (2009). Ggplot2 : elegant graphics for data analysis (New York: Springer).
Young, M.D., Wakefield, M.J., Smyth, G.K., and Oshlack, A. (2010). Gene ontology analysis for RNA-seq: accounting for selection bias. Genome biology *11*, R14.