The Hebrew University in Jerusalem

Faculty of Science – The Alexander Silberman Institute for Life Science

School of Computer Science and Engineering

Neuronal inducible transcription is comprised of scaled modules

By: Lior Ziv

Mentors: Prof.Ami Citri & Prof.Tommy Kaplan

Thesis Submitted for the degree: “Master of Science”

Submitted on: 8/1/2020

1. Abstract

Learning, memory and long-term plasticity in neurons depend on the induction of transcription programs. The function applied by a neuron in translating a defined pattern of stimulation into a transcription program is not known. My work is focused on computational analysis, comparing and contrasting these transcription programs with the aim of identifying the organizational principles defining their temporal structure. The common opinion today is that different stimuli or cell age induce different gene programs[4]. In our study we looked at it from a different approach, dividing the programs to modules. We found different programs induced by different stimuli or different cell culture age are in a large part scaled versions of the same program. This imply that the same genes are used repeatedly and may be regulated as groups. This step brings us closer to being able to describe the precise mechanisms driving the induction of individual genes.

Table of content

2. Motivation

Learning, memory and long-term plasticity all depend on the induction of gene expression in neurons[1-3,18]. Our lab has recently identified a transcriptional code of memory, whereby the recent experiences of individual mice can be decoded from a minimal fingerprint of inducible gene transcription programs across brain structures[6]. Studying inducible transcription in cultured neurons, it has been recently claimed that different patterns of neuronal stimulation drive different transcription programs[4]. Motivated by this work, we studied several transcriptional programs obtained in our lab following stimulation of neuronal cultures. These programs differ in the identity of the inducing stimulus, as well as in the maturity of the culture. The expressed gene transcription programs of those different experiments had shared and unshared genes. We hypothesized that transcription programs diverge in their recruitment of modules of gene expression, which can be scaled in their level of induction, defining ‘scaled modules’. This hypothesis implies that the common genes are used repeatedly across programs and may be regulated as groups. Thus, we first want to define a scaled module. A group of common genes which are induced as a response to different stimuli or cell age, the group can be characterized by unique scaling factor which is a scalar determined by the ratio between the peak fold change of the same gene in two different experiences.

3. Introduction

3.1 From experience to adaptive behavior

Our brain go through different experiences every day, each experience leads to neuronal activity which causes the neuron to react via a cascade of gene expression, starting with immediate early genes [1,2,5]. This chain of events is eventually captured in the cell memory and determines its adaptive behavior[1-3]. Memory can be divided based on its duration, into short and long-term[37]. A short-term memory is the retention of information for a brief time with no neuronal change. In contrast, a long-term memory occurs following learning, it changes neural pathways in order to store information, therefore months, or even years later recall of that information is possible. An important biological feature distinguishing short from long-term memory is phase of RNA and protein synthesis. If either RNA or protein synthesis are blocked before or immediately after experience, long-term memory formation is disrupted[39]. Experiences could be aversive – LiCl (lithium chloride), foot shock or rewarding such as sugary water and cocaine. Our lab showed[6] experiences can be distinguished using only five IEG as markers. Do different experiences mean the underlying gene expression will be necessarily different? Our goal is to characterize the relationship between stimulus and its translation to transcriptional programs.

3.2 Calcium ion signal transduction

A typical neuron consists of a cell body also known as the soma, dendrites and a single axon which are the input and output units of that neuron with correspondence. Neurons signal transduction starts with calcium influx into the neuron soma. Calcium is second messenger and it enters via two main routes, L-type, voltage-gated calcium channels (VSCCs), and glutamate receptors, N-methyl-d-aspartate (NMDA) or a-amino-3-hydroxy-5-methyl-4-iso-xazole -propionate, AMPA[30]. Here we will focus on the glutamate receptors NMDA and AMPA.

N-methyl- D-aspartate receptor, NMDAR, ionotropic glutamate receptor built as heteromeric complex constitute of R1 and R2 subunits. NMDA shown to be important for controlling synaptic plasticity and memory[9], its activation depends on calcium signaling cascade including Ca2+/CaM-dependent protein kinase (CaMKII) activation. NR1 subunit has a few variants which their functional significance remains undetermined. NR2 subunit, in contrast to the NR1 splice variants, are characterized by large differences in their electrophysiological profiles and very distinct expression patterns during development, therefore it has been speculated that they play a role in brain development[8]. Both subunit expression begins as early as 14 days[7].

α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid glutamate receptor, AMPAR, is an ionotropic transmembrane receptor for glutamate which mediates fast synaptic transmission. AMPARs are composed of four types of subunits encoded by different genes, designated as GluR1, GluR2, GluR3 and GRIA4 which combine to form tetramer. Phosphorylation of AMPARs can regulate channel localization, conductance, and open probability. AMPARs is responsible for the fast-excitatory synaptic transmission since it opens only for a brief time[10]. The AMPAR's permeability to calcium and other cations, such as sodium and potassium, is governed by the GluA2 subunit which its presence determined by post-transcriptional modification. During early development, neurons usually express calcium permeable AMPARs containing the GluR2 subunit. If an AMPAR lacks a GluA2 subunit, then it will be impermeable to the mentioned ions. Calcium-permeable AMPARs can drive gene expression directly[43], calcium impermeable AMPAR can indirectly activate calcium signaling cascade via glutamate binding which depolarizes the postsynaptic membrane and trigger opening of other calcium permeable channels, such as VSCCs[44].

3.3 Calcium regulation on gene transcription

The calcium ion (Ca2+) is a key intracellular signaling molecule which effects on many cellular processes. One of the most important such Ca2+ regulated processes is transcription[17,31]. Cells typically maintain an intracellular Ca2+ level

of 10^{-7}mM, which is 10^4 times lower from the level outside the cells which enable calcium influx via NMDAR or AMPA receptors and VSCCs. The route through which calcium enters the cell and the distinct signaling cascades recruited during synaptic transmission, are critical for determining the resulting pattern of induced transcription[12-18] . Hilmar Bading et al. showed that calcium influx via NMDARs or L-type VSCCs induce *Fos* transcription via independent signaling cascades[39].

Ca2+-bound calmodulin (Cam) binds and activates many downstream effectors, such as Ca2+/calmodulin-dependent protein kinases (CaMKs). CaMKs, most importantly CaMKIV, in turn phosphorylates and activates the transcription factor cyclic AMP (cAMP) response element-binding protein (CREB), which binds to cAMP response elements (CREs) in DNA to activate transcription[16]. Bading also showed that calcium influx will activate CamkII, but only calcium influx via L-type VSCCs induced transcription depends on the activation of CAMKII[39].

Additional path includes activation of PKC(Prcka) via Ca2+ binding. Ca2+ is release into the cytoplasm and binds PKC, which is then activated. Once activated, PKC phosphorylates and regulates many substrates, including CREB[17] and activator protein 1 (AP1)[40], transcription factors.

Moreover Ca2+ signaling can activate mitogen-activated protein kinase (MAPK) pathway via activation of small guanine nucleotide-binding protein ,Ras18. The regulation of transcription by MAPKs is highly complex with numerous phosphorylation targets transcription factors such as CREB, AP-1 and ATF-2[41].Binding of epidermal growth factor (EGF) to the EGFR activates the tyrosine kinase activity of the cytoplasmic domain of the receptor which promotes the removal of GDP from a member of the Ras family. Ras is activated through binding the GTP and activates protein RAF kinase. RAF kinase phosphorylates and activates MEK1 or MEK2 which then phosphorylates and activates MAPK.

3.4 CREB phosphorylation promotes gene expression

The CREB family of transcription factors is comprised of cAMP response element-binding protein (CREB), cAMP Response Element Modulatory Protein (Crem) and activating transcription factor 1 (Atf1). Activation of CREB in response to increased

levels of cAMP or Ca2+ appears to be regulated by the inducible phosphorylation of a specific amino acid, Ser133, which acts as a positive regulator of transcription[19]. Two CREB-binding proteins, CBP and p300, have been characterized to mediate the interaction of Ser133-phosphorylated CREB with the polymerase II transcription[20]. Phosphorylation of CREB at Ser133 by CaM kinases may involve the direct entry of Ca2+ into the nucleus. The ca2+ binds to calmodulin and activate the CaM kinases. The duration of CREB Ser-133 phosphorylation induced by NMDA receptor activation depends on the developmental maturity of cultured neurons. After 14 days, strong NMDA stimulation causes only a transient increase in phospho-CREB, contrasting with the prolonged increase in phospho-CREB seen at day 7[14]

3.5 Artificial stimuli patterns in cell culture

Our goal was to characterize the relationship between specific modes of neuronal activation as response to a stimulation or changes in cell age, and the generated transcriptional output. To address this, we exposed cortical neuronal cultures to KCl or Picrotoxin stimuli which induced artificial experience in those cells. **Potassium chloride (KCl)**, increase the ion influx into the cell. Since KCl have electrochemical influence a 40-55+ mM, used in most studies, leads to the opening of L-type VSCCs, resulting in the stimulation of multiple calcium-dependent signaling pathways. Our lab showed that at the use of more than 5Mm, KCl electrochemical influence will cause L-type voltage-gated calcium channels to open but not under 5Mm. Therefore, we used 5Mm to enable synaptic transmission with no L-type voltage-gated involved. **Picrotoxin,** $GABA_a$ antagonist, increases synaptic excitation indirectly by blocking inhibitory synaptic currents mediated by the $GABA_a$ receptors[21,22].

Inconclusion, when our brain goes through an experience the neuronal activity response is an outcome resulted by the expressed set of genes, were immediate early genes will impact first. Since it is not easy to measure this kind of activity in a living brain the Citri lab designed an experiment based on cell culture and artificial

stimuli. The artificial stimuli of KCl and picrotoxin lead to calcium signal transduction which result differential gene expression. Our hypothesis stats different stimuli or cell age will cause a different gene transcription programs, with shared and distinct genes. The shared genes might be used repeatedly as modules, while the difference between transcription programs is the result of different scaling factor used in the shared gene modules. Therefore, my aim is laying the foundation to support our hypothesis by using the designed experiment results, which portray how different experience our brain undergoes result different transcription programs.

4. Materials and methods

4.1 Experiment

This part summarizes the acquisition of the data by Doron Haritan a graduate M.Sc student in the Citri lab. Primary cortical cultures were prepared from newborn C57bl/6 mice on postnatal days 0-1. Experiments were performed on Day 10 and 14 *in vitro* (Div10/14). The neurons were stimulated by adding either KCl (to a final concentration of 5 µM unless defined otherwise) or Picrotoxin (final 10 µM) and cultures collected 1, 2, 4 or 8 hours later. RNA-seq was performed, libraries were sequenced using Illumina NextSeq-500. Read counts were processed using STAR alignment method[26] and HTSeq[27]

4.2 Normalization

Each experiment was normalized separately using DEseq2[25]. The DESeq2 method first performs a geometric mean for each gene using all its repeats, the resulted mean is then used to divide the read counts for that gene across those repeats. The next step is taking the median per repeat, out of the resulted read count divided by geometric mean. Finally, each median divides the read count of all genes in the associated repeat. The normalization is needed for a few reasons, for example a common problem is library depth differences which mean

experiments have different total read count between repeats. Moreover DESeq2 handle other complex problems, each individual gene normalization depends on the read count of other genes in the same sample (repeat), for example genes with zero counts or outliers which shift the entire mean[25]. To solve that DESeq2 uses log scale on the data (when performing the geometric mean) and divides the product by the median of each sample. Since each one of the experiments was conducted and normalized separately in order to compare them in the least biased way as possible, each gene was measured by the relative change in time after the exposure to before the exposure. This was done by dividing each gene time points with the read mean at the control (before exposure) which is called fold change (FC).

4.3 The GS model

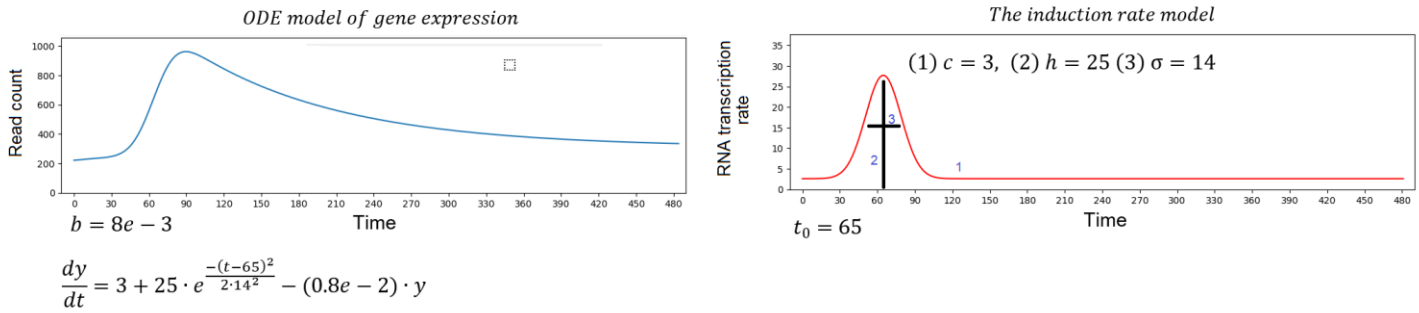
4.3.1 GS model equations and parameters

We developed a model which we called gene's shape (GS) model, it relies on two equations, first the Induction rate equation (Equation 1) which formulate the rate of RNA production in time t, Equation 2 is a differential equation which modules the change in the total number of RNA molecules in time t. The model is comprised of five parameters depicted in figure 1.A: c – basal level of a gene, h – height of the gaussian amplitude which describes the induction rate, b – degradation level, t_0 – gene's peak time and σ – the induction rate peak width. Figure 1 contain examples of equation 1 and 2 for two genes, arc and homer1 from picrotoxin 10 days experiment.

(1) $$Ir(t) = c + h \cdot e^{\frac{-(t-t_0)^2}{2\sigma^2}}$$

(2) $$GS := \frac{dy}{dt} = Ir(t) - b \cdot y$$

A



B

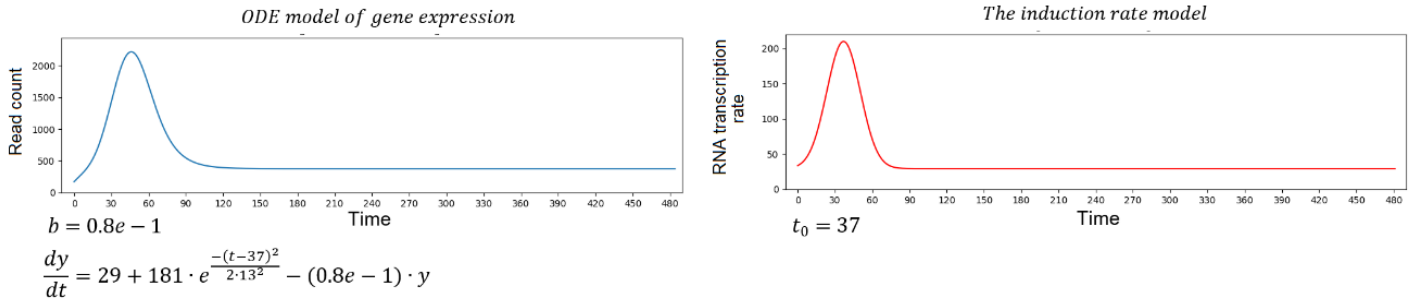


Figure1. The GS model equations 1 and 2 fitted to (A) Homer1 (B) Arc from picrotoxin 10 days experiment. Left the ordinary differential equation model of GS (equation 2). Right the induction rate (equation 1).

4.3.2 The fitting process

Each experiment we worked on contained a total of about 29,000 genes, were our goal was to identify the genes which induction rate is affected by our treatment. We wanted do perform the most sensitive identification process in order to maximize our gene set, this leaded to developing the GS model. Our algorithm finds the five GS model parameters individually per gene, allowing us to identify genes with a relatively small change. The process is described in the pseudo-code in figure2, first for each gene the FC values from the normalized read count matrix are taken, those values are used in the calculation of seven initial value arrays which contain the 5 GS parameters, the initial values represent a variety of genes behavior, for example – low transient induction, high transient induction and high long-lasting induction, the full details are shown in appendix S6. Each array

together with the GS model function and the gene FC values are given as input to optimize.fmin function, the function start with our initial value array and minimize the RMSE between the fitted values resulted by our model to the real gene values. The minimization is done using downhill simplex optimization algorithm[28] which iteratively pick five new values and assign them to the GS model, it stops when reaching convergence, which occur when the difference between the previous step RMSE and the current RMSE is sufficiently small. The minimal RMSE out the 7 results is chosen along with its 5 GS model values. The last step takes our optimal 5 GS model value array and search a grid around it to try and minimize our solution even further. The final GS model values array is appended to GS results table and used in the next step to calculate the p-value of our fit.

```
For all the genes with FC >=1.2 in experiment A:
   a. y_real = Extract gene data
   b. initial_guess= Prepare 7 different guesses based on the gene data (c,h,b,t0,σ)
   c. For each guess:
        i. RMSE[i], params[i] = find optimal parameters using optimize.fmin(find_GS_model_fit,
           initial_guess[i], y_real)
   d. minimal_rmse_guess = params[argmin(RMSE)]
   e. optimal_guess = minimiztion_optimiztion(find_GS_model_fit, minimal_rmse_guess, y_real)
   f. append optimal guess into the GS_results_table
```

Figure 2: The GS fitting algorithm pseudo-code.

4.3.3 Testing the model fit

To test the fit of our model to a gene shape over time we perform a variation of F test[45]. This F test compares between two different models, the first is a constant model, we use the maximum likelihood estimator (MLE) which is the expression of a gene remaining constant around the mean value of its reads . The second model will be our GS model. Accordingly, H_0: the gene is constant in time H_1: the gene is affected as a result of the treatment. The test is performed using equation 3 which calculates the F statistic value, where RSS_1 – is the root sum of squares, performed on the difference between our original data and the constant model (same value at all points), RSS_2 – the root sum of squares, performed on the difference between our original data and the GS model values. p_1 – number of freedom degrees in the constant model which is one, p_2 – number of freedom degrees in the GS model which is four. n is the total number of observations, in

our case the number of time points including repeats. From the F statistic we can derive a p-value which determines how much do we need those four extra parameters in the GS model (only one parameter at the constant model) in order to express that gene shape. Appendix S7 explain why we can use the F test, which assumes the data is normally distributed.

$$(3) \qquad F = \frac{\frac{RSS_1 - RSS_2}{p_2 - p_1}}{\frac{RSS_2}{n - p_2}}$$

4.4 Gene ontology analysis

Gene ontology annotations were derived from go-basic.obo[46] file which contain the entire gene ontology tree of GO. The file was processed into term tree keeping only its three top ancestors for each term. Using this tree, we manually grouped the terms under seven groups: receptor signaling, cell structure and adhesion, transporters and channels, calcium activity, transcription, ubiquitin like protein transferase and structural constituent of ribosome. The terms choice was focused towards relevant calcium induction and transcription related functions as well as plasticity in those cells resulted by our treatment. Group names were given by us, appendix S5 show the final table with group names and associated origin GO terms.

4.5 Transcription factor motif analysis

The binding motifs were derived from GSEA[47] database which take as an input a list of genes and retrieve a matrix containing column with one for each motif a gene is connected to, or empty otherwise. We aimed to early genes related binding motifs therefore only AP, CREB, ATF, E4F1, NFY and NFAT binding motifs were taken and since each binding motif can appear few times all associated definitions of the same motif were collapsed. Our result is matrix of the above binding motifs on the different gene groups from our experiments, containing one where a motif is connected to a gene or nothing otherwise.

5. Results

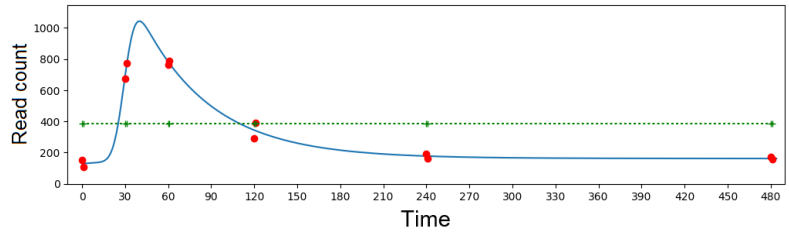
An experience causes neuronal activity which is translated into a defined transcription program. We hypothesize that inducible transcription programs are comprised of scaled modules. By this, we mean that a common set of genes is induced in different gene expression programs, albeit to a different extent. Furthermore, we hypothesize that these genes are regulated as groups (modules), subject to scaling by a common factor. To check our hypothesis, we used a dataset acquired in the simplified and accessible system of primary neuronal cultures, which were grown to 10 vs 14 days in vitro (DIV). The cultures were stimulated with either KCl (5 mM) or Picrotoxin (10 µM), cultures collected 1, 2, 4 or 8 hours later and RNA-seq was performed. In this study we used data from 3 different experiments: A) DIV10 picrotoxin; B) DIV14 picrotoxin and C) DIV10 KCl. Defining the transcriptional program induced by each treatment and comparing between the programs supported investigation of our hypothesis regarding the existence of scaled modules.

5.1 The GS model

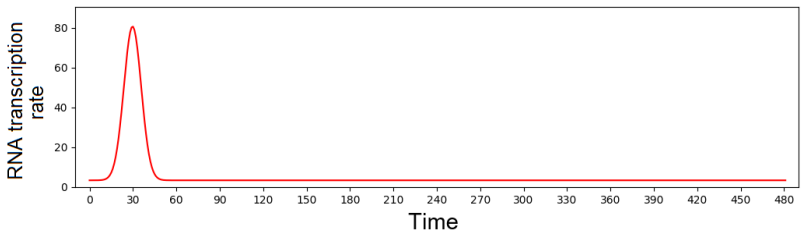
In order to identify induced genes developed a model that mathematically defines the induction profile of each gene. Those differently expressed genes will build the different gene transcription programs, this will enable us to compare between experiments and explore our hypothesis. Our model is based on a differential equation structure in order to capture changes in time. This 'Gene Shape' (GS) relies on two equations (1) Number of produced RNA molecules at time t, this equation contains a gaussian function which express the change in production rate over time in response to our treatment. The second equation (2) is the total change in RNA transcripts. The model is comprised of five parameters: c – basal level, h – height of amplitude, b – degradation level, t_0 – peak time and σ – the peak width. Applying the GS model provides with an objective and sensitive metric supporting the identification of genes that we define as induced by the treatment. With the GS model each gene was fitted with the described five paraments, using F test to evaluate the statistical significance of the improvement of the GS model over the

constant model describing a gene behavior (section 4.3 of materials and methods). Figure 3 show 3 different genes, for each gene the fit of both GS model and constant model are plotted (A1-C1), while A2-C2 show only the induction rate model (equation 2). At the top we see fos which is highly induced by our treatment, resulting a good fit to the GS model with a low p-value. At the middle Brinp2 which has a moderate induction first and steep induction by the end. The bottom figure show gapdh which is a housekeeping gene therefore it is not affected by our treatment resulting a poor fit to our model with high p-value.

A1



A2



Fos (KCl 10)

p-value: 4.16e-6

Adjusted p-value: 1e-3

Basal (c): 3.3

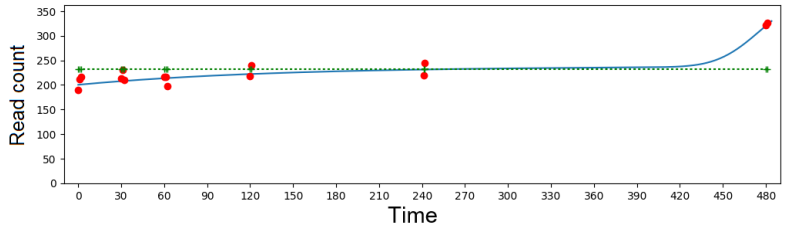
Induction rate (h): 77

Both c and h units - number of RNA transcripts

Degradation rate (b): 0.2e-1

t_0: 30, σ : 6

B1



Brinp2 (Picrotoxin 14)

p-value: 7.6e-5

Adjusted p-value: 1.6e-3

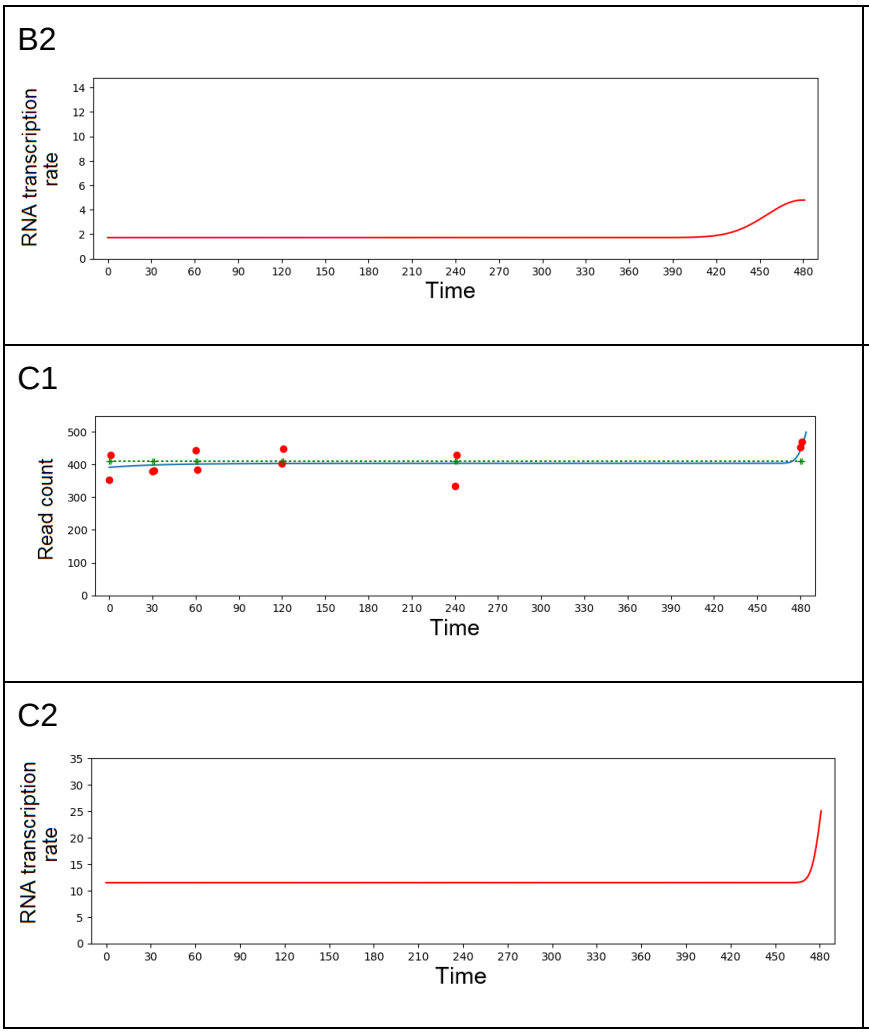
Basal (c): 1.7

Induction rate (h): 3

Both c and h units - number of RNA transcripts

Degradation rate (b): 0.7e-2

t_0: 478 σ : 24



Gapdh (KCl 10)

p-value: 0.55

Basal (c): 11.5

Induction rate (h): 23.4

Both c and h units - number of RNA transcripts

Degradation rate (b): 0.3e-1

t_0: 486, σ : 6

Figure 3: Example of three genes modeled using the GS model.

5.2 Identify genes which were affected by our treatment

Now we will perform a series of steps (Figure 4) using the GS model in order to identify the affected gene set in each one of our experiments. Since the experiments were performed separately, for a better comparison we decided to use the relative change of each gene and not its absolute values. The relative change is measured compared to the control and called fold change (FC). FC is the number of reads in a specific time frame divided by the control reads number. First, only genes with peak FC≥1.2 (average of ~5700 genes per experiment) are fitted with the GS model (mean of the peak time FC values) resulting a table with the five GS parameters and a p-value for each gene fit. FC of 1.2 is set as the

lower bound because we believe that at least 20% change in induction is needed in order to identify a gene as affected rather than noisy, this low bound supply us with reliable genes only due to the GS model. Since we do ~5700* F tests per experiment we use FDR correction using Benjamini Hochberg method[29]. Now we want to identify the genes which interest us, first only genes with adjusted p-value ≤0.1 will be included in the final set, containing genes from all three experiments (A, B and C in figure 4), leaving us with 932 genes for picrotoxin 10 days, 1481 at picrotoxin 14 days and 15 at KCl 10 days. Next using the set, we can relax a bit our constraints on the p-value and try to include more genes which are a bit below our detection level. For each one of the experiments we look on the data of the entire unified set (D in figure 4) and take new genes from it with unadjusted p-value ≤0.05. Therefore, at the end of the process each experiment will have genes with adjusted p-value ≤0.1 or genes with unadjusted p-value ≤0.05 that appear in another experiment with adjusted p-value ≤0.1 (E, F and G in figure 4). This step created the final set with 776 common genes, which are genes common to at least two out of our three experiments. 612 common genes were identified at picrotoxin 10 days , 648 at picrotoxin 14 and 504 at KCl 10.

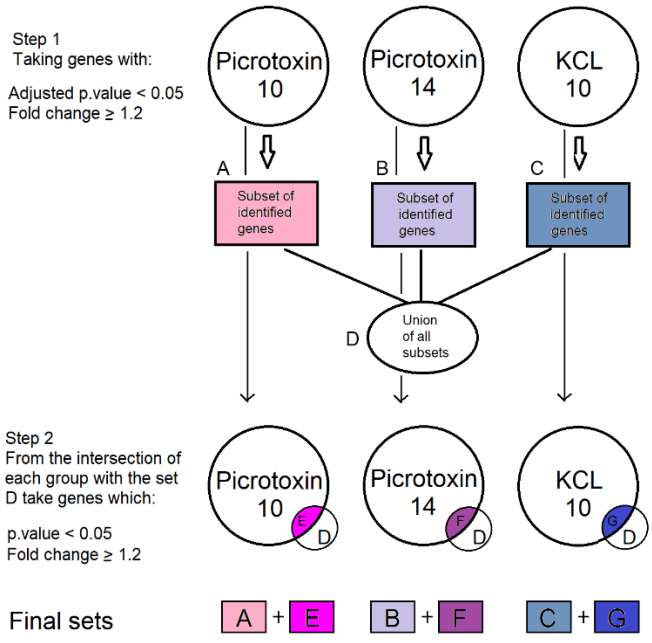


Figure 4: The entire pipeline to identify genes which were affected by our treatment.

5.3 Gene transcription programs

Using the GS model identification process described in section 5.2 we created a gene set for each one of our experiments, including a total of 776 genes that are common to at least two experiments. Each experiment is characterized via its gene's expression pattern in time and their biological function. Genes peak time was depicted in an expression heatmap which show the induction waves for that experiment, each gene in the heatmap was normalized to the peak time to create a uniform range for all the depicted genes. We defined genes with peak at 30-60 minutes as immediate, with peak at 120-240 minutes as early and peak at 480 minutes as late. To define functionality each experiment gene set was checked for go-terms and binding motifs enrichment (section 4.4 in materials and methods), the results were plotted as heatmaps.

5.3.1 The transcription programs waves of induction

Cell culture of 10 days old cortical neurons, exposed to picrotoxin (figure 5.A), containing 932 induced genes, 612 (67%) genes overlap with at least one of the other experiments. Cortical neurons cell culture 14 days old (figure 5.C), exposed to Picrotoxin, containing 1481 genes, 648 (44%) genes are common to at least one of the other experiments. Cortical neurons cell culture 10 days old exposed to KCl (figure 5.B), contains 504 genes, while 15 are unique only to KCl 10 (Appendix figure S1.C). All together there is a large fraction of overlap between the different programs. When looking at the induction waves in figure 5, we see a difference in the peak time. The peak time of picrotoxin 14 genes is delayed compared to the peak time of picrotoxin and KCl 10, the effect is stronger between picrotoxin 10 and 14. In picrotoxin 10 Immediate genes are the majority with 38% while picrotoxin 14 and KCl 10 majority is at the late genes with 62% and 52% in correspondence.

5.3.2 The GO terms enrichment of transcription programs

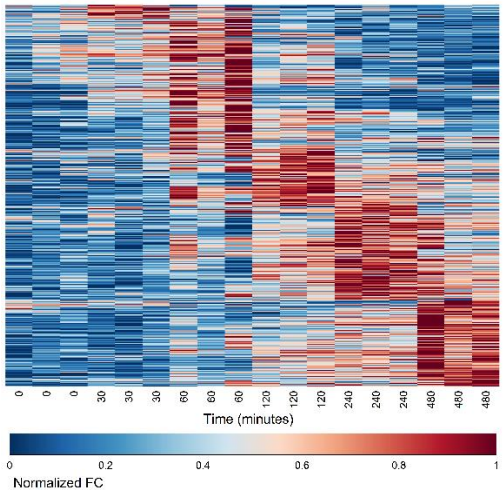
The final terms were defined as a summation of GO terms which are related to: receptor signaling, cell structure and cell adhesion, transports and channels, calcium activity, transcription, ubiquitin-like protein transferase activity and structural constituent of ribosome. Picrotoxin 10 is enriched for receptor signaling transcription but mostly to transcription, especially in the immediate genes (Figure 6.A). KCl 10 is mostly enriched for transports and channels also here especially at the immediate genes (Figure 6.B). Picrotoxin 14 is highly enriched with calcium activity but also with cell structure and cell adhesion, ubiquitin-like protein transferase activity and transcription (Figure 6.C).

5.3.3 Enrichment of transcription factor binding motifs at different transcription programs

Transcription factors binding motifs of each transcription program were derived summing up only specific transcription factors: CREB, ATF, AP, E4F1, NFY and NFAT and depicted as fractions in figure 7. AP-1 (Activator protein 1) controls a number of cellular processes including differentiation, proliferation, and apoptosis AP-1 show high enrichment across all time points in all the programs (Figure 7). CREB-TF (cAMP response element-binding protein) is a cellular transcription factor has a role in neuronal plasticity and long-term memory formation in the brain. CREB is closely related to ATF-1 (activating transcription factor-1) proteins which take part in cell growth and survival. The enrichment to CREB shows a high enrichment in immediate and early genes at picrotoxin and KCl 10 (Figure 7.A-B) while in picrotoxin 14 the effect is delayed relatively low in the immediate genes oppose to early and late genes (Figure 7.C). Moreover, the enrichment we see in CREB is correlated with the ATF enrichment.

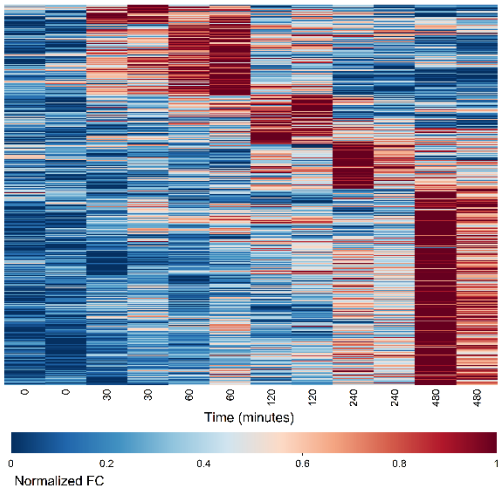
A

Picrotoxin 10



B

KCl 10



C

Picrotoxin 14

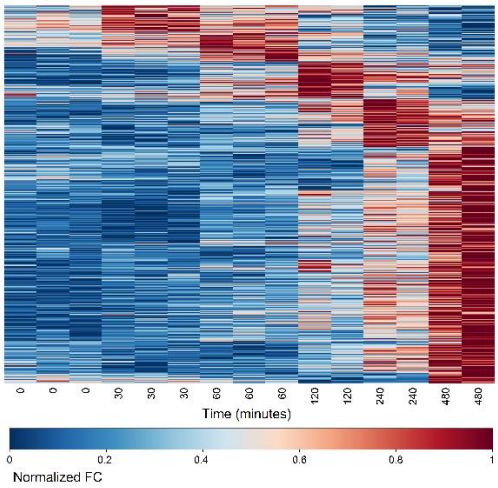
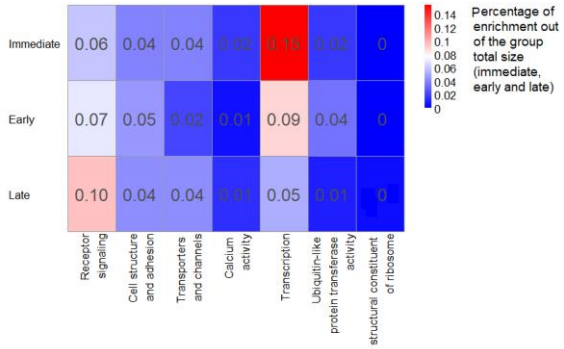


Figure 5: Waves of induction heatmaps. Each heatmap containing fold change expression for each gene normalized to its peak value. (A) Picrotoxin 10 (B) KCl 10 (C) Picrotoxin 14.

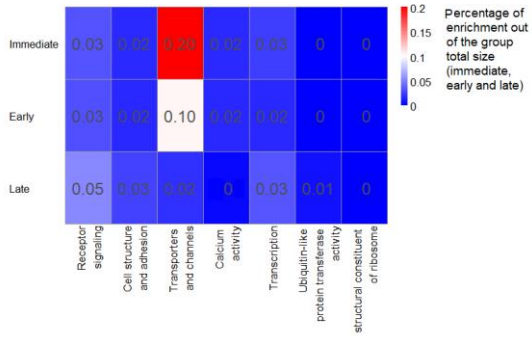
A

Picrotoxin 10



B

KCl 10



C

Picrotoxin 14

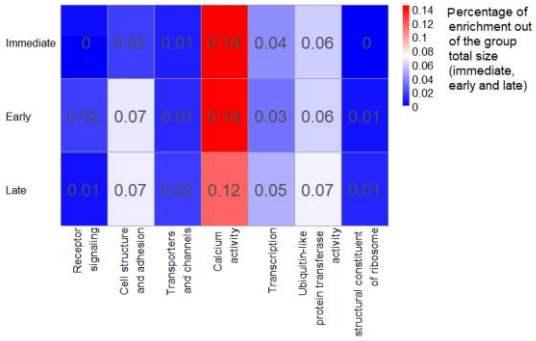


Figure 6: Heatmap with enriched GO terms. Each heatmap contains the fraction of each go term from immediate, early and late gene groups. (A) picrotoxin 10 (B) KCl 10 (C) picrotoxin 14.

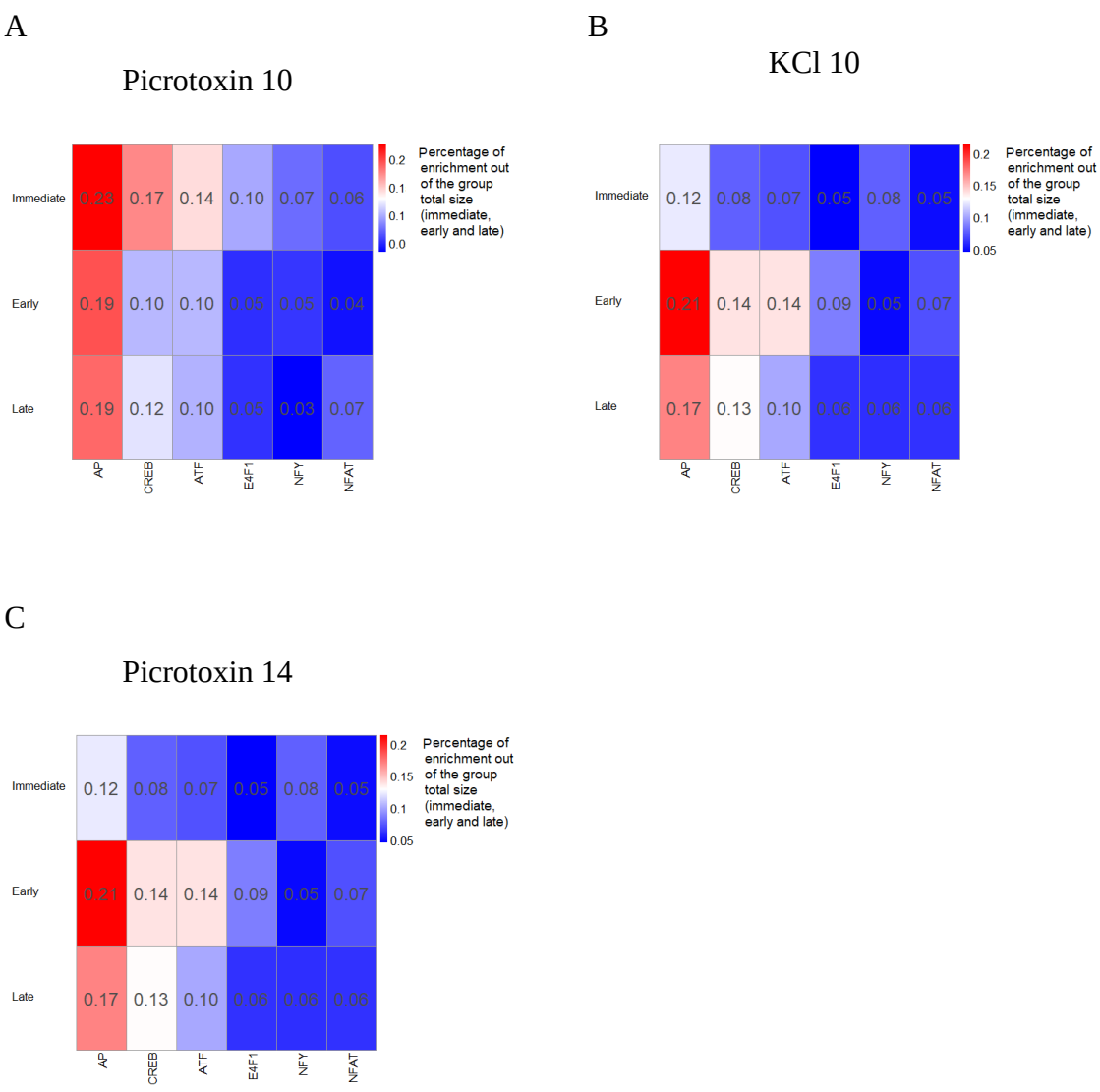


Figure 7: Heatmap with enriched binding motifs. Each heatmap contains the fraction of each motif from immediate, early and late gene groups. (A) Picrotoxin 10 (B) KCl 10 (C) Picrotoxin 14.

5.4 Genes are used repeatedly across programs

We have 776 genes in the affected common group, viz they are common to at least two experiments. If the same genes appear in multiple experiments the overlap will be larger than what you expect by chance, therefore we need to show our set size is significant and won't occur by random. Thus, if we take a random subset of genes using the same process described in section 5.2, the final set size should be significantly smaller. In order to address the probability that this overlap reflects enrichment of a set of common genes, we performed a permutation test, in which we shuffled genes with above zero FC in all time point are, as a result each gene get a new FC and parameters of the GS model. Next the identification step and modules separation are done the same as described before (section 5.2) only with the new shuffled data. The process described here was repeated for 100,000 iterations, figure 8 show the histogram of the results. None of the results reached a set size of 776 genes, hence the p-value is 1e-5 with 21 standard deviation from the shuffled data mean: 412. This result demonstrates a highly significant enrichment of the common gene set over chance, providing support for the notion that different gene expression programs share a common core. We are therefore interested in pursuing the question of whether this common gene set is comprised of modules of genes which are recruited to different degrees by each program.

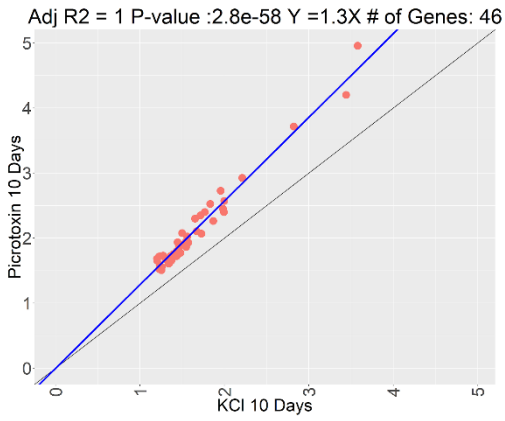
Figure 8: Histogram of the identified genes group size using permutation test.

5.5 Defining modules using pairwise comparisons

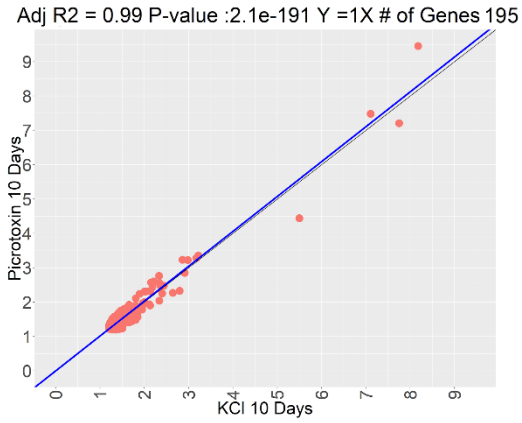
Our hypothesis implies that the common genes across programs are regulated as scaled modules. In order to test this hypothesis, I wanted to maximize the size of our shared gene pool containing genes which we believe were affected by the exposure to picrotoxin or KCl or by cell culture age. By maximizing the shared gene pool, we can have a bigger overlap between experiments allowing us to work with more data to support our hypothesis. Identifying genes affected by our treatment is a hard task especially with genes close to 1.2 FC which have a relatively small change. The GS model we developed allow us to examine each gene and calculate a p-value reflecting how much it was affected by our treatment, therefore using the GS model enabled us to fine tune the final gene pool, taking genes with a minimal FC of 1.2. Thus, our final set size is maximized and contains diverse gene scales, which are then divided into richer modules with a variety of scaling factors. Next for each common gene in all pairwise comparisons its scaling factor is calculated, this is done using the peak FC of that gene in one experiment divided by the peak FC of that gene in the second experiment (of course keeping it consistence across all genes in that comparison). For example, the peak value of fos is 6 (fold change) in KCl 10 days and 12 in picrotoxin 10 days therefore their peak FC ratio will be two (Figure 9.D). This division result approximates how much do we need to multiply the divisor gene expression level in order to get the same level as in the divided gene expression level also can be viewed as the slope of the line we can pass through a scatter plot of those gene expression levels. Using the scaling factor values, we cluster the data into nine different modules, their ranges are: [0.2-0.65], (0.65,0.8], (0.8,1.19], (1.19,1.4], (1.4,1.6], (1.6,1.8], (1.8,2.19], (2.19,3], (3,4], (4,5]. The clusters are divided in this manner up to 2.2 ration since we believe approximately 0.2 FC above and beneath each scaling factor are gene which behave the same, this takes under account the noise biological experiments tend to have. Above 2.2 the groups are sparse a therefore are more roughly sliced. Figure 9.E show an example for how subgroups association to a specific module is done, we see 3 different gene groups each group is characterized by different scaling factor resulted by its gene FC ratio, each

group is assigned to the module which contain this scaling factor. For each one of the modules a scatter plot is plotted showing all the FC ratios of genes in that module, using those a new linear equation is fitted, hence for each cluster of genes we get a single scaling factor with the optimal fit (minimal R2). This new FC ratio fit describes the overall ratio between the two compared gene groups in that module, figure 9.A-C show some fit examples.

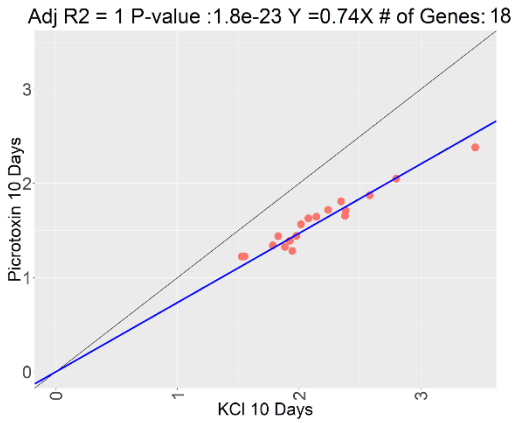
A



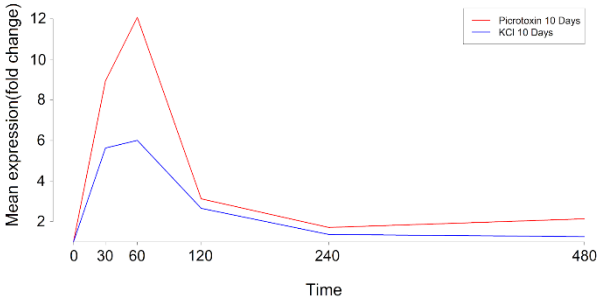
B



C



D



E

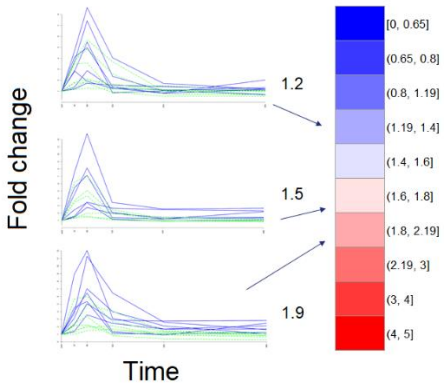


Figure 9: A-C: Example for clusters. Each pink dot represents the ratio between the peak fold change of the same gene in two compared experiments. The black line is linear line y= x, the blue line is the fit for that cluster. D: Fos number of reads over time (both picrotoxin and KCl 10). E: Cluster division, each graph are gene from two different experiments with an FC ratio of 1.2, 1.5 or 1.9, using the ratio they are divided into nine different clusters.

5.6 Characteristic of the defined modules in different programs

We clustered each pairwise comparison into nine modules by their scaling factor (described in section 5.5), the different comparisons results are shown in the next section. For each module GO-terms enrichment and binding motifs analysis is presented to illustrate the biological function of each one of the modules. This can support our hypothesis and show that different modules are regulated together as part of their biological functionality.

5.6.1 Scaled modules composition and overlap between experiments

Comparing cell culture maturity (picrotoxin 10 and 14) led to 515 common genes, which are 27% out of their total union, while comparing stimulation (picrotoxin and KCl 10) led to 355 common genes which are 33% out of their total union. Comparing KCl 10 with picrotoxin 14 showed 385 common genes which are 24%

out of their total union. The interesting result is the biological repeats comparison overlap, with 496 common genes which are 26% out of their total union. These are two replicates of the same experiment (stimulation of DIV14 cultures with picrotoxin) which were performed at least a year apart, on different cultures, and therefore reflect the variance of the experimental paradigm. We would anticipate that these cultures would have a larger shared component than cultures treated with different compounds or at different stages of maturity. Overall the percentage of common genes in pairwise experiments comparison is around ~28% (average) leaving us with ~72% unique genes. Figures 10-13.B show the modules composition, across all comparisons the largest module is with a slope of ~1, which imply those genes behave the same across experiments. Moreover, the biological repeats comparison shows low module sizes in the large scaling factors, namely most of the genes are centered in the 0.7,1 and 1.3 scaling factors which show a smaller difference in the expression of this comparison genes.

5.6.2 Scaled modules GO terms enrichment

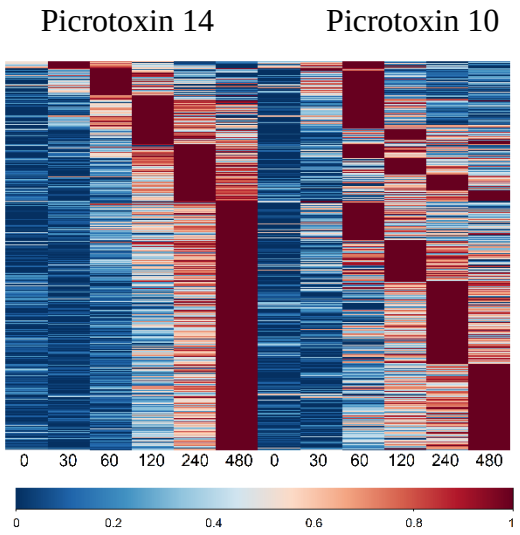
Figures 10-13.C show the GO-terms distribution across modules, although we don't see any specific function for each individual module at a specific comparison, the total enrichment in each comparison seems to be shifted by the compared experiments. The cell maturity comparison shows high enrichment in the transcription term (Figure 10.C). Stimulation comparison is enriched with calcium activity and transporters and channels terms (Figure 11.C). When altering both the stimulation and cell maturity we see enrichment at transcription and receptor signaling terms (Figure 12.C). Those can suggest the cell maturity shift the transcription program towards transcription enrichment while stimulation will affect more on the calcium signaling cascade. The biological repeat comparison most enriched terms (Figure 13.C) are receptor signaling and structural constituent of ribosome but examining closer the modules composition the highest enrichment is at module six which contain only nine genes. Appendix S2 show the fraction of GO-term enrichment, biological repeats comparison oppose to other comparisons show a low GO term enrichment and that is despite this comparison gene overlap

(section 5.6.1). This might indicate that indeed different comparisons common genes, which can be viewed as different modules, might have a more significant biological role which determine the rest of the transcription program composition.

5.6.3 Scaled modules binding motif enrichment

All comparisons show high enrichment for AP, CREB and ATF (Figures 10-13.D). Generally looking across the tables, binding motif distributed with higher enrichment across higher modules (5-9) for cell maturity (Figure 10.D) and stimulation (Figure 11.D) comparisons, while biological repeats comparison fractions (Figure 13.D) have higher enrichment in the former modules (1-4). Comparing picrotoxin 14 and KCl 10 (Figure 12.D) show a bit more random fractions, still the higher enrichment is for AP, CREB and ATF but here the distribution across smaller or larger scaling factor models is less coherent. Appendix S2 show a low fraction of binding motif enrichment for picrotoxin 14 in contrast to the other programs. This all together might suggest that there is a biological meaning for higher scaling factors modules, which are characterized by a larger gap between genes.

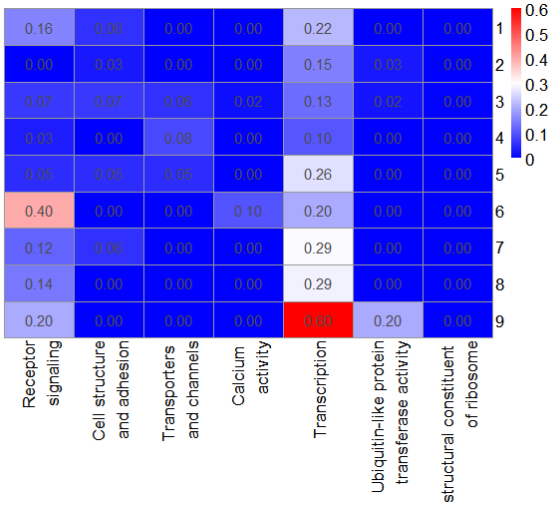
A



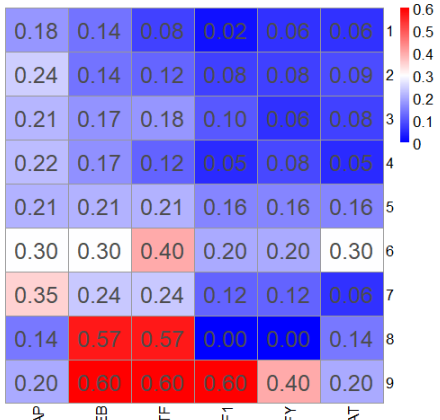
B

Module	slope	R2	p.value	size
1	0.44	0.93	6.2e-31	51
2	0.74	1.00	2.5e-83	66
3	0.99	0.99	4.0e-271	281
4	1.30	1.00	7.6e-79	59
5	1.52	1.00	5.7e-28	19
6	1.70	1.00	9.2e-19	10
7	2.01	1.00	7.6e-22	17
8	2.49	0.99	1.5e-07	7
9	3.68	0.97	2.9e-04	5

C



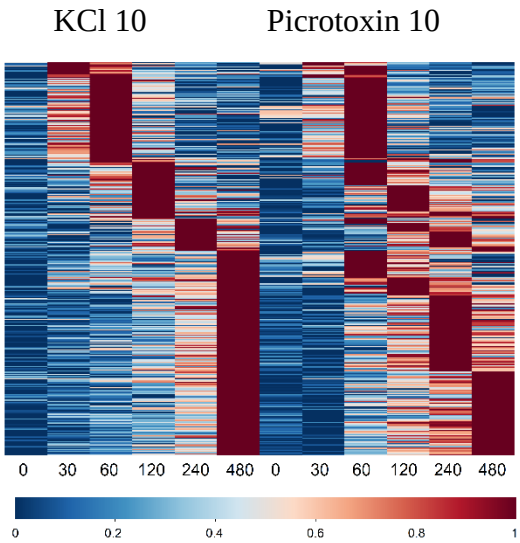
D



	AP	CREB	ATF	E4F1	NFY	NFAT
All modules	2.06	2.53	2.51	1.32	1.15	1.14
1-4 Modules	0.85	0.62	0.5	0.25	0.28	0.28
5-9 Modules	1.21	1.92	2.02	1.08	0.88	0.86

Figure 10: Comparing cell maturity, Picrotoxin 10 vs 14 days. (A) Heatmap containing fold change expression, each gene normalized to its peak value (B) The data clustered into modules. (C-D) Modules enrichment heatmap, for each module the fraction of enrichment is depicted (C) Go terms (D) binding motif. The smaller table contains summed values of D.

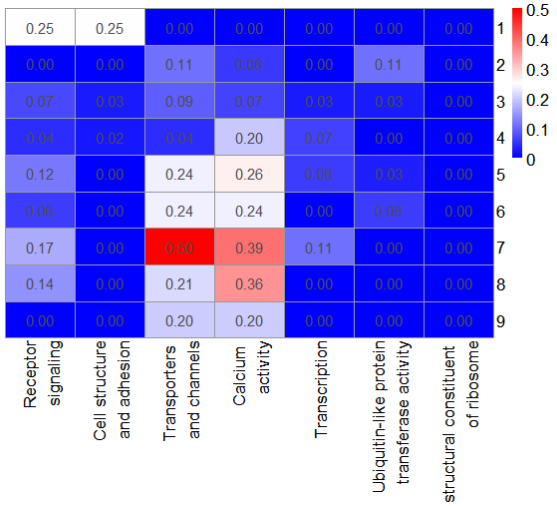
A



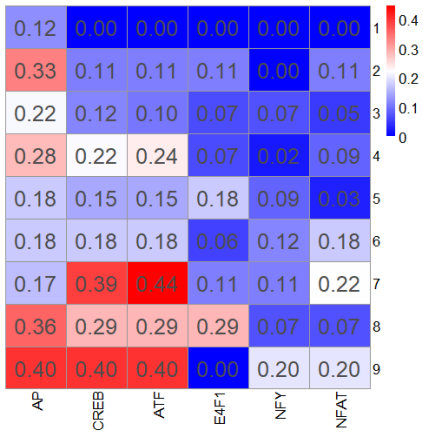
B

Module	slope	R2	p.value	size
1	0.55	0.99	1.4e-08	8
2	0.74	1.00	1.8e-23	18
3	1.02	0.99	2.1e-191	195
4	1.29	1.00	2.8e-58	46
5	1.50	1.00	5.0e-49	34
6	1.70	1.00	2.9e-24	17
7	1.95	1.00	2.1e-22	18
8	2.40	1.00	1.0e-17	14
9	3.57	0.99	4.1e-05	5

C



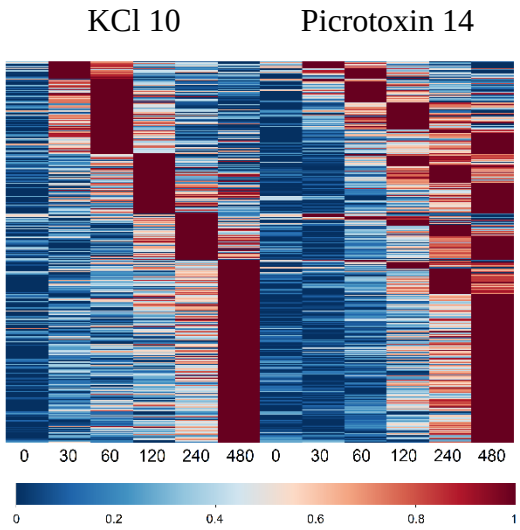
D



	AP	CREB	ATF	E4F1	NFY	NFAT
All modules	2.23	1.84	1.91	0.88	0.68	0.95
1-4 Modules	0.96	0.45	0.45	0.24	0.09	0.25
5-9 Modules	1.28	1.4	1.45	0.63	0.59	0.7

Figure 11: Comparing stimulation, Picrotoxin vs KCl 10 days. (A) Heatmap containing fold change expression for each gene normalized to the peak value. (B) The data clustered into. (C-D) Modules enrichment heatmap, for each module the fraction of enrichment is depicted (C) Go terms (D) binding motif. The smaller table contains summed values of D.

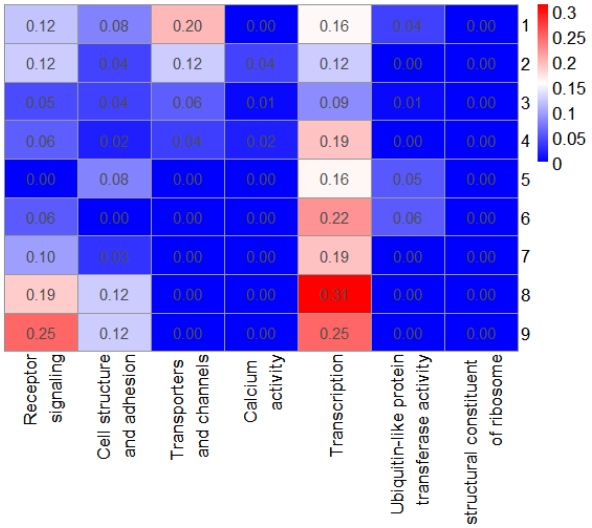
A



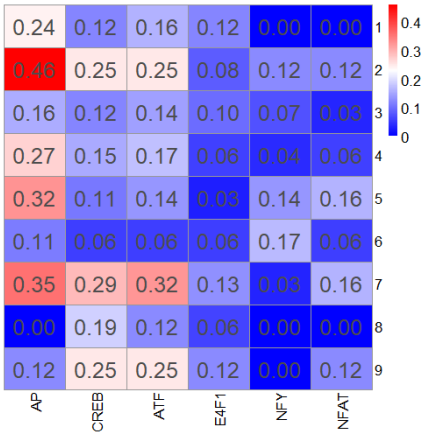
B

Module	slope	R2	p.value	size
1	0.50	0.96	3.1e-18	25
2	0.74	1.00	4.8e-32	24
3	0.98	0.99	8.4e-170	174
4	1.30	1.00	1.0e-74	52
5	1.48	1.00	1.4e-50	37
6	1.70	1.00	1.2e-26	18
7	2.00	1.00	1.8e-38	31
8	2.58	0.99	8.5e-18	16
9	4.44	0.86	2.2e-04	8

C



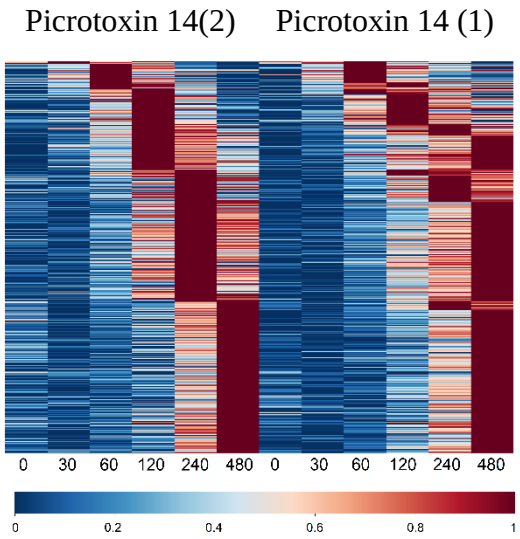
D



	AP	CREB	ATF	E4F1	NFY	NFAT
All modules	2.04	1.54	1.62	0.76	0.57	0.72
1-4 Modules	1.12	0.64	0.73	0.36	0.24	0.22
5-9 Modules	0.92	0.89	0.89	0.4	0.33	0.5

Figure 12: Comparing picrotoxin 14 days to KCl 10 days. (A) Heatmap containing fold change expression for each gene normalized to the peak value. (B) The data clustered into modules. (C-D) Modules enrichment heatmap, for each module the fraction of enrichment is depicted (C) Go terms (D) binding motif. The smaller table contains summed values of D.

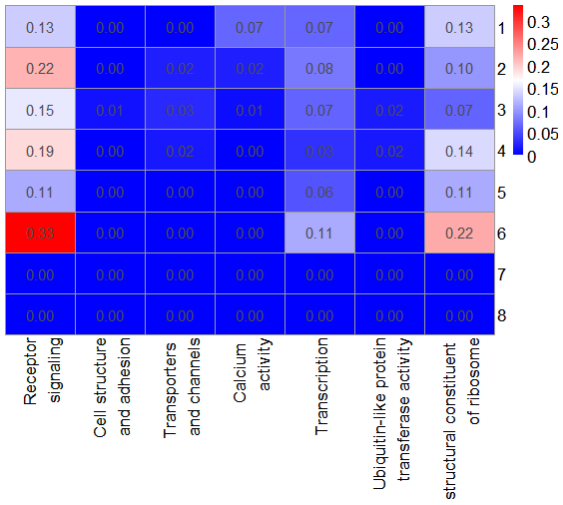
A



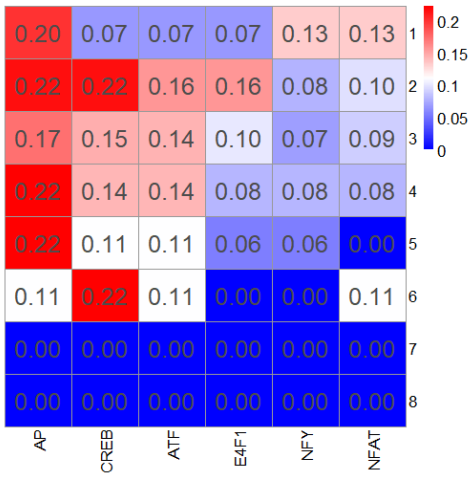
B

Module	slope	R2	p.value	size
1	0.62	1.00	1.5e-18	15
2	0.70	1.00	2.2e-62	51
3	0.95	0.99	2.2e-313	335
4	1.29	1.00	2.0e-83	63
5	1.50	1.00	2.2e-25	18
6	1.69	1.00	2.0e-13	9
7	2.00	0.99	2.8e-03	3
8	2.41	1.00	2.2e-02	2

C



D



	AP	CREB	ATF	E4F1	NFY	NFAT
All modules	1.14	0.9	0.73	0.46	0.42	0.51
1-4 Modules	0.81	0.57	0.51	0.4	0.36	0.4
5-8 Modules	0.33	0.33	0.22	0.06	0.06	0.11

Figure 13: Comparing biological repeats, Picrotoxin 14 1(original) vs picrotoxin 14 2. (A) Heatmap containing fold change expression for each gene normalized to the peak value. (B) The data clustered into modules. (C-D) Modules enrichment heatmap, for each module the fraction of enrichment is depicted (C) Go terms (D) binding motif. The smaller table contains summed values of D.

5.7 Higher enrichment for transcription factors in scaled common gene groups compared to non-scaled

In this section we take the overlap between our 3 experiments in order to show that the change in scale for those common genes group might be the cause for the difference between different experiments transcription program. Picrotoxin 10,14 and KCl 10 have 251 common genes. From this point we define, a linear group is a group of genes which appear in a module with scaling factor of one in both age and stimulation comparison, thus those genes behave the same across all three experiments. On the contrary a definition of a non-linear group is a group of genes which are part of non-linear module in at least one of the comparisons, thus their expression level is different across experiments. Appendix S3 table show 169 genes which appear in different modules (A1) and 82 which appear at the same module (A2). Deeper examination showed that 69 genes out of the 82 are part of a linear cluster in both pairwise comparisons, thus those 69 genes are a linear group and the rest 182 genes are a non-linear group. Figure 14.A show the motif enrichment, the non-linear group enrichment ratio is higher and almost doubles at CREB and ATF motifs compared to linear group. Figure 14.B is comparing the GO terms enrichment ratio; the trend is similar in most of the terms, but transcription which more than tripled itself in the non-linear group. Appendix S3.B-C show 21 IEGs contained in the common non-linear group, those genes start the entire gene expression cascade. The results might suggest that non-linear gene groups play a larger role in determine the entire transcription program in different experiments.

A B

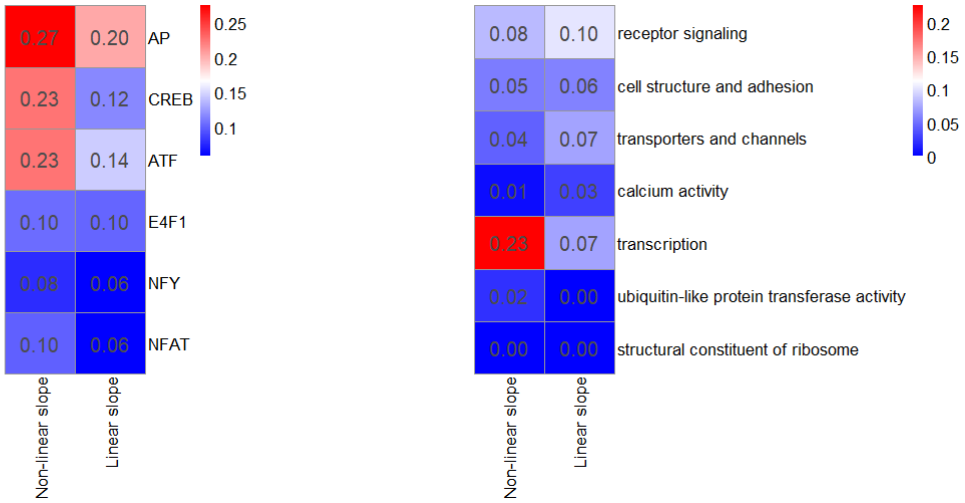


Figure 14: A. Maximal modules which intersect, comparing picrotoxin 14 and picrotoxin 10 fit picrotoxin 10. B-C enrichment for B – motifs, C- go terms of genes which belong to the same linear cluster across the maturity and stimuli or belong to a different cluster, were one must be non-linear.

To conclude the results, my goal was to show how different stimuli or cell age can cause a different transcription program while our hypothesis suggests the difference is caused by a core of shared genes that are expressed in different levels. The core of shared genes contains genes that expression intensity varies across different stimulation or cell age. I showed that the group size of shared genes identified by the GS model is significant compared to a randomly selected group, the significance was determined by applying a permutation test. The shared gene groups were divided into nine modules for each pairwise comparison, the later modules 5-9 in cell and age comparison show higher enrichment compare to the biological replicates comparison where modules 1-4 are more enriched. Moreover, the common genes of all three experiments which vary in their scale are more enriched with transcription factors. All those together could establish our hypothesis.

6. Discussion

Experience our brain undergo causes synaptic activity which result membrane depolarization of target neurons by an influx of calcium ions into the neurons soma. This depolarization triggers differential gene expression compared to control behavior in those neurons and alters their behavior, causing a refinement of the circuitry allowing adaptation to the environment[1,3]. Synaptic transmission which allow calcium influx to the neuron have a few alternative paths[32], NMDARs or AMPA, glutamate receptors, and VSCCs. Here we only address the glutamate receptors via exposure of cortical cell cultures to picrotoxin or KCl. The motivation behind this experiment is to identify the transfer function between different cell ages or different stimuli and the resulted gene expression program. We developed the GS model in order to identify those differentially expressed genes, using it we created groups of common genes and divided them into modules differ by their scaling factor. Our hypothesis states that the difference between experiments which vary in age or stimulation could be a consequence of those different modules which shift the gene expression cascade to different routes.

6.1 A Gene peak time is affected by the cell culture maturity

Figure 5.A show picrotoxin 14 days has a more delayed peak time compared to picrotoxin and KCl 10 days. The peak time of many genes is shifted towards 240 and 480 minutes. CREB transcription factor promotes transcription[19] when activated. Activation of CREB in response to increased levels of Ca2+ appears to be regulated by the inducible phosphorylation of a specific amino acid, Ser133. Previous study[14] has shown that cell culture of 14 days or older exposed to strong NMDA stimulation will cause a transient increase in phosphorylated CREB (phospho-CREB), contrasting with a prolonged increase in phospho-CREB seen at younger cell cultures. The transient effect is caused by the activation of a pathway that leads specifically to dephosphorylation of CREB and doesn't appear in low concentrations (20 Mm). In our study we use 5Mm KCl and 10Mm picrotoxin which are relatively low, therefore we expect a CREB motif associated gene

enrichment across all time points at both 10 and 14 days old cell cultures. Our data motif tables (Figures 7.A-B) show CREB motif connected gene enrichment at immediate, early and late time points with peak at the IEGs for 10 days cell culture of both picrotoxin and KCl. This continues enrichment is accompanied by ATF motif connected gene enrichment which appear nearly at the same levels. Picrotoxin 14 also show a sustained effect with a minimum at the IEGs which show only half of the enrichment compared to early and late (Figure 7.C). This finding of CREB motif enrichment in picrotoxin 14 might be the reason for the delayed gene peak we see compared to the other experiments. But still the question of why do we see such a low CREB enrichment at the IEGs remain unanswered.

Previous work[32] in neurons has indicated that synaptic stimulation recruits a fast calmodulin kinase IV (CaMK)-dependent pathway that dominates early signaling to CREB. Wu at el showed that phosphorylation of CREB is a convergence of fast and slow protein kinase pathways. CaMK-dependent pathway can be followed by a slower pathway that depends on Ras/mitogen-activated protein kinase (MAPK). The activation of the Ras-MAPK pathway was specifically effective in promoting a slow phase of CREB phosphorylation. Therefore, a possible explanation for the smaller enrichment at the immediate genes might be due to slow CREB phosphorylation caused by ras GTPase regulation on the MAPK pathway[42]. To support this claim Calcium/Calmodulin Dependent Protein Kinase Kinase 1(CamKK1, S4.A), part of the fast pathway, was identified as differentially expressed gene only at 10 days cell cultures experiments, which might point to the existence of the fast pathway. An indication for the slow pathway is Ribosomal Protein S6 Kinase A3 (RpS6ka3, S4.B) which is responsible for the phosphorylation in the slow pathway. RpS6ka3 appears in all programs and its highest peak is at picrotoxin 14.

Another possible explanation is dephosphorylation of CREB by Protein Phosphatase 1 (PP-1) in mature cell culture. Calcineurin (cabin1) is activated by Ca2 + /calmodulin (CaM) then calcineurin dephosphorylates PPP1r1a (I-1), which in its phosphorylated state is a potent inhibitor of PP-1. Thus, activation of calcineurin results in an increase in PP-1 activity by way of disinhibition[33]. From

the data we see that cabin1 was identified as differentially expressed only at picrotoxin 14 days (S4.E) with peak at 60 minutes, while I-1 appear in all three experiments (S4.C).

An additional explanation for the delay peak time in picrotoxin 14 regardless to CREB phosphorylation is inhibition of myocyte-enhancing factor 2 (MEF2), evolutionarily conserved family of transcription factors. Cabin1, present only in peak 14, can inhibit MEF2 transcription[34]. One strongly connected target gene of MEF2 is arc[35], our data show the lowest fold change for arc appear at picrotoxin (S4.D). Also, the peak time of arc in picrotoxin 14 is 60 minutes which is delayed compared to picrotoxin and KCl 10 days which get high expression values at 30 minutes. In order to track the origin of this delay and check some of the hypotheses we suggested here, new experiments should be conducted.

6.2 Do scaled gene modules determine the identity of different expressed gene subset in experiments?

Each pair of experiments common genes was clustered into nine groups. Clusters were defined with the slope fitted to them by the entire group (Figure 9.A-C). We wanted to show that the difference in those modules might be the key factor for the uncommon gene subset we see on each one of the transcription programs. But first we needed to support the significance of our results.

Initially, we showed the pairwise clusters size has significance, viz we won't get the same extent of clusters by picking random genes from our data. This was checked via permutation test, figure 8 show the results which confirm that the clusters we got are meaningful and far from exist in that extent at random, with p-value of 1e-5 and 21 standard deviation from the randomly permuted data mean.

Now that we have significance for the extent of our clusters, we wanted to show biological functionality which was checked by measuring the enrichment of motif and go terms in each cluster. We hoped clusters will varied in the enrichment term or motif they are related to, which will give us a way to define each cluster roll. Looking at tables (Figures 10-13.C) the go term enrichment looks more affected by the pair of experiments we compared than the cluster those genes are

in, therefore we can't connect a cluster to a specific biological go term. Generally looking at all the modules in a specific comparison, we saw transcription enrichment when comparing cell maturity and calcium signaling terms enrichment when comparing stimulation. Moreover, the number of genes in large scaling factor modules is low in picrotoxin 14 compared to other experiments and its highest scaling factor is 2.41 which is relatively low oppose the other comparisons with average of 3.9. This suggest that the subset of common genes in other comparisons is more go-term enriched plus the number of genes which behave differently is larger.

Binding motif enrichment showed a bit more interesting results (Figures 10-13.D). Across all comparisons we see that the highest enriched motifs are AP, CREB and ATF. Looking on it closer showed that the cell maturity and stimulation comparisons have a higher enrichment for all motifs at modules 5-9, with scaling factor of 1.5 and above, while the biological repeats comparison showed a more enriched gene population at 1-4 modules, which are centered around scaling factor of 1.

Additionally, appendix S2 show that all comparisons but biological repeats in total are more enriched both with go-term and binding motif. This occur even though the overlap extent of biological repeats comparison is about the same (26%) as other comparisons (average ~28%).

Is it possible that the binding motif and go terms enrichment we see suggest that the larger scaling factor modules which are also larger in their extent compared to the biological repeats, have more effect on determine the transcription program? Do they cause by their different scaling across experiments to a distinct gene cascade illustrated by the uncommon gene groups we see in each comparison?

6.3 Do common genes origin from high scaling factor modules determine the transcription program?

Intersecting three experiments together brings a total of 251 common genes, 82 appear at the same cluster and 169 appear in a different one (Appendix S3). The

groups which interest us were genes which act different across experiments. Examining closer our common gene group we saw 69 out of the 82 which appear in the same cluster are linear, thus this subset of genes behaves the same across all three experiments referred as the linear group. This leaves us with 13 genes which appear in the same nonlinear slope cluster plus 169 genes which appear in different clusters, all together we have a total of 182 genes referred as the non-linear group. To clarify, at the different cluster group with 169 genes one pairwise comparison can have linear slope while the other cant. Figure 14.A-B show a higher enrichment to AP, CREB, ATF and significantly higher enrichment (above three times) to transcription for the non-linear group. This significantly higher enrichment in the non-linear vs linear group might be the key to the diversity between the transcription programs. Staring with a core of 21 IEGs across all experiments (Appendix S3.B-C) each one of the genes appeared in a different module. The fact that common IEGs behave different may lead to the later non-linear shared modules which ultimately push the entire transcription program towards a specific direction, were the outcome is different transcription programs with a core of common scaled genes.

6.4 Conclusions

The GS model we developed identified differentially expressed genes across experiments. The comparison of stimulation and cell age showed many common (~28%) genes but also a large fraction (72%) of uncommon genes. We speculated here the scaling of those common genes might determine the identity of the entire transcription program. Dividing the pairwise comparison common genes into models showed the entire enrichment of binding motif and GO terms is higher across cell maturity and stimulation comparisons oppose to biological repeats comparison. Moreover, larger scale modules are more enriched when comparing cell maturity and stimulation. Another discovery is high enrichment for transcription factors in the common non-linear gene group. All of the above might suggest that indeed, a core of shared genes which are part of different scaled modules in

different experiments play a key role in the later differentially expressed gene waves and define the outcome of the entire transcription program.

7. References

1. Cohen, S. & Greenberg, M. E. Communication between the synapse and the nucleus in neuronal development, plasticity, and disease. *Annu. Rev. Cell Dev. Biol.* **24,** 183–209 (2008).
2. Alberini, C. M. Transcription Factors in Long-Term Memory and Synaptic Plasticity. *Physiol. Rev* 121–145 (2009). doi:10.1152/physrev.00017.2008.
3. Flavell, S. W. S. S. W. & Greenberg, M. E. M. E. M. E. Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.* **31,** 563–590 (2008).
4. Tyssowski, K.M & Gray, J.M. Different Neuronal Activity Patterns Induce Different Gene Expression Programs. Neuron 530-546 (2008) . doi: 10.1016/j.neuron.2018.04.001.2008.
5. Sheng, H. Z., Fields, R. D. & Nelson, P. G. Specific regulation of immediate early genes by patterned neuronal activity. *J. Neurosci. Res.* **35,** 459–467 (1993).
6. Diptendu, M & Ami, C. Salient experiences are represented by unique transcriptional signatures in the mouse brain. eLife. 7:e31220 (2008). doi: 10.7554/eLife.31220
7. Laurie, D.J & Seeburg, P.H. Regional and developmental heterogeneity in splicing of the rat brain NMDAR1 mRNA. J Neurosci. 14(5): 3180–3194(1994).
8. Cull-Candy, S.G & Leszkiewicz, D.N. Role of distinct NMDA receptor subtypes at central synapses. Sci STKE. 255, re16. DOI: 10.1126/stke.2552004re16(2004).
9. Fei Li, M.D., Ph.D. & Joe Z. Tsien, Ph.D. Memory and the NMDA Receptors. N Engl J Med. 361(3): 302–303(2009).
10. Simon R.Platt. The role of glutamate in central nervous system health and disease – A review. The Veterinary Journal. 173(2):278-86 (2007).
11. Mochly-Rosen, D. Localization of protein kinases by anchoring proteins: a theme in signal transduction. *Science* . 268(5208):247-51(1995).
12. Ghosh, A. & Greenberg, M. E. Calcium Signaling in Neurons: Molecular

Mechanisms and Cellular Consequences. *Science* (268-.). 5208:239-247(1995).

13. Buonanno, A. & Fields, R. D. Gene regulation by patterned electrical activity during neural and skeletal muscle development. *Current Opinion in Neurobiology* **9,** 110–120 (1999).
14. Sala, C., Rudolph-Correia, S. & Sheng, M. Developmentally regulated NMDA receptor-dependent dephosphorylation of cAMP response element-binding protein (CREB) in hippocampal neurons. *J Neurosci* **20,** 3529–36. (2000).
15. West, A. E. *et al.* Calcium regulation of neuronal gene expression. *Proc. Natl. Acad. Sci. U. S. A.* **98,** 11024–31 (2001)
16. Anna Vilborg, Maria C. Passarelli, and Joan A. Steitz. Calcium signaling and transcription: elongation, DoGs, and eRNAs. Receptors Clin Investig. 3(1): e1169(2016).
17. Clapham, D.E. Calcium signaling. Cell. **131**, 1047–1058(2007)
18. Bading, H. Nuclear calcium signalling in the regulation of brain function. *Nat. Rev. Neurosci.* **14,** 593–608 (2013).
19. Gonzalez, G. A and Montminy, M. R. Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell. 59(4), 675-80 (1989).
20. Kwok, R.P & Goodman, R.H. Nuclear protein CBP is a coactivator for the transcription factor CREB. Nature. **370**(6486), 223-6 (1994)
21. Maximov, A., Pang, Z. P., Tervo, D. G. R. & Südhof, T. C. Monitoring synaptic transmission in primary neuronal cultures using local extracellular stimulation. *J. Neurosci. Methods* **161,** 75–87 (2007).
22. Newland, C. F. & Cull-Candy, S. G. On the mechanism of action of picrotoxin on GABA receptor channels in dissociated sympathetic neurones of the rat. *J. Physiol.* **447,** 191–213 (1992).
23. Rubiolo, J. A. *et al.* Transcriptomic Profiling of Mice Primary Cortical Neurons in Response to Medium Change. *Transcr. Open Access* **4,** 4–11 (2016).

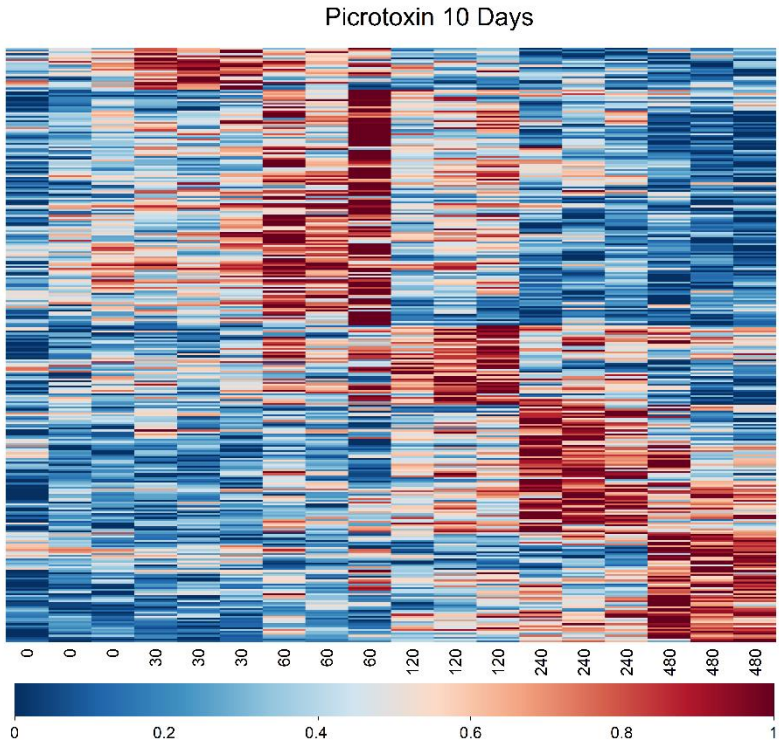
24. Levin, J. Z. *et al.* Comprehensive comparative analysis of strand-specific RNA sequencing methods. *Nat. Methods* **7,** 709–715 (2010).
25. Love, M.I, Huber, W and Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. **15**(12), 550(2014).
26. Alexander Dobin & Thomas R. Gingeras. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 29(1): 15–21(2013)
27. HTSeq manual-https://readthedocs.org/projects/htseq/downloads/pdf/latest/
28. Downhill simplex algorithm - https://academic.oup.com/comjnl/article-abstract/7/4/308/354237?redirectedFrom=fulltext
29. Benjamini, H and Hochberg, Y. Contorlling the False Discovery Rate: A Practical and Powerfull Approch to Mulitple Testing. Journal of the Royal Statistical Society. 57, 289-300(1995).
30. Flavell, S. W. S. S. W. & Greenberg, M. E. M. E. M. E. Signaling mechanisms linking neuronal activity to gene expression and plasticity of the nervous system. *Annu. Rev. Neurosci.* **31,** 563–590 (2008).
31. Lyons, M. R. & West, A. E. Mechanisms of specificity in neuronal activity-regulated gene transcription. *Prog. Neurobiol.* **94,** 259–295 (2011).
32. Wu, G.-Y., Deisseroth, K. & Tsien, R. W. Activity-dependent CREB phosphorylation: Convergence of a fast, sensitive calmodulin kinase pathway and a slow, less sensitive mitogen-activated protein kinase pathway. *Proc. Natl. Acad. Sci.* **98,** 2808–2813 (2001).
33. Mulkey, R.M & Malenka, R.C. Involvement of a calcineurin/ inhibitor-1 phosphatase cascade in hippocampal long-term depression. Nature. 369(6480), 486-8(1994).
34. Timothy, T.A & Olson E.N. MEF2: a calcium-dependent regulator of cell division, differentiation and death. Trends Biochem Sci. 27(1), 40-7(2002).
35. Rashid, A. J. & Josselyn, S. A. Emerging roles for MEF2 transcription factors in memory. Genes, Brain and Behavior. 13(1), 118-25 (2014).
36. Choi SH, Li Y, Parada LF, Sisodia SS. Regulation of hippocampal

progenitor cell survival, proliferation and dendritic development by BDNF. Mol Neurodegener. 2009;4:52. Published 2009 Dec 21.

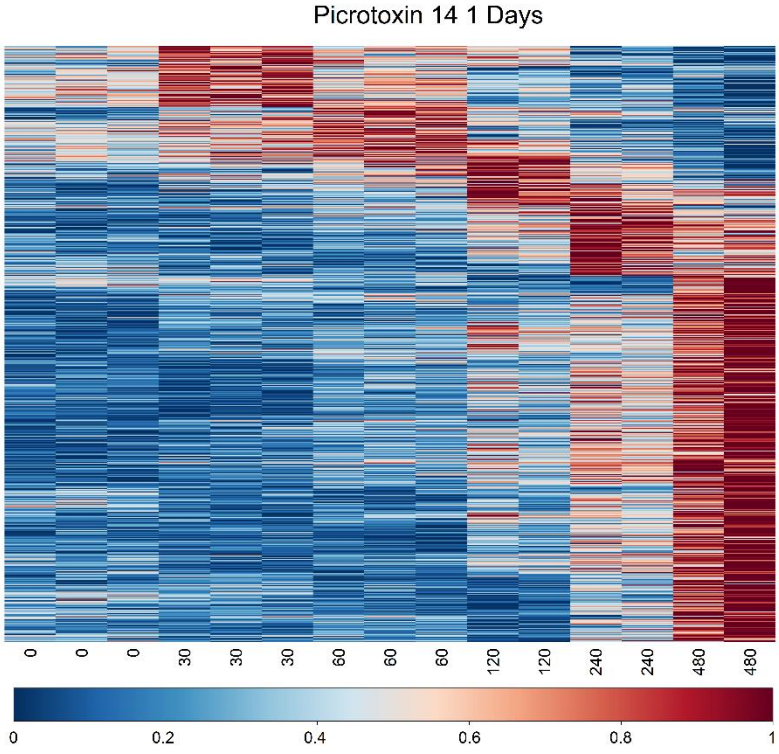
37. Davis, H.P and Squire, L.R. Protein synthesis and memory: a review. Psychol Bull. 96(3), 518-59 (1984)
38. Yamada, M and Hatanaka, H. Differences in survival-promoting effects and intracellular signaling properties of BDNF and IGF-1 in cultured cerebral cortical neurons. J Neurochem. 78(5), 940-51(2001).
39. Bading, H., Ginty, D. D. & Greenberg, M. E. Regulation of gene expression in hippocampal neurons by distinct calcium signaling pathways. Science (80-.). 260, 181–186 (1993).
40. Yokoyama, K & Mochida, J. C-Fos regulation by the MAPK and PKC pathways in intervertebral disc cells. PloS one. doi: 10.1371/journal.pone. 0073210. 2013.
41. Alan J.Whitmarsh. Regulation of gene transcription by mitogen-activated protein kinase signaling pathways. Biochimica et Biophysica Acta. 1773(8), 1285–1298(2007).
42. Sheng .M and Greenberg M.E. CREB: A Ca2+-Regulated Transcription Factor Phosphorylated by Calmodulin-Dependent Kinases. Science. 252, 1427-30(1991)
43. Greer, P. L. & Greenberg, M. E. From Synapse to Nucleus: Calcium-Dependent Gene Transcription in the Control of Synapse Development and Function. *Neuron* **59,** 846–860 (2008).
44. Higley, M. J. & Sabatini, B. L. Calcium Signaling in Dendritic Spines. *Cold Spring Harb. Perspect. Biol.* 4**,** a005686–a005686 (2012).
45. F-test: https://en.wikipedia.org/wiki/F-test
46. GO terms basic obo file: http://geneontology.org/docs/download-ontology/
47. gsea msigdb: https://www.gsea-msigdb.org/gsea/msigdb/annotate.jsp

8. Supplementary

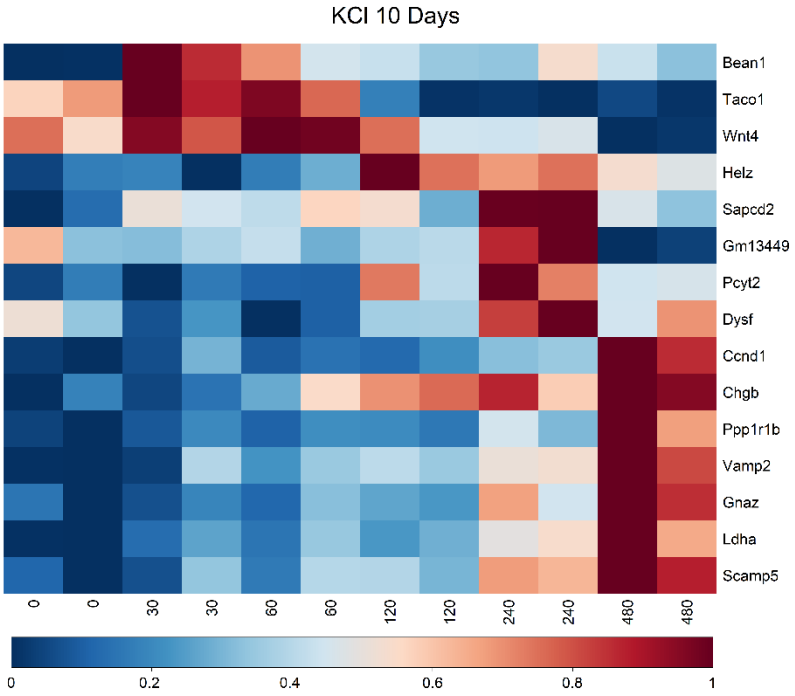
A



B



C



S1. Unique genes expression heatmaps, each gene is normalized to its peak. (A). Picrotoxin 10 days (B). Picrotoxin 14 days(1) (C). KCl 10 days.

	Go-terms	Motifs
Maturity	0.44	0.47
Stimulation	0.59	0.53
Picrotoxin 14 vs KCl 10	0.64	0.61
Biological repeats	0.34	0.32

S2: Fraction of common genes in each pairwise comparison that are connected to a term/motif

S3. 169 genes (A1) which appear in different clusters and 82 (A2) which appear at the same across comparisons. Slope 1/2 are the clusters slopes, Module 1/2

are the cluster number the gene is in. 1- comparison of picrotoxin 14 to 10, 2- comparison of KCl to picrotoxin 10.

A1				
Gene name	slope 1	slope 2	module 1 number	module 2 number
Egr2	3.68	1.95	9	7
Nr4a1	3.68	1.5	9	5
Grasp	2.49	1.95	8	7
Scg2	2.49	1.95	8	7
Fos	2.49	1.95	8	7
Egr1	2.49	1.29	8	4
Acan	2.49	1.02	8	3
Pim1	2.01	3.57	7	9
Fosl1	2.01	2.4	7	8
Pim3	2.01	2.4	7	8
Homer1	2.01	1.7	7	6
Ier2	2.01	1.7	7	6
Rasl11b	2.01	1.7	7	6
Vgf	2.01	1.7	7	6
Junb	2.01	1.5	7	5
Egr4	2.01	1.5	7	5
Rgs7	2.01	1.5	7	5
Mir670hg	2.01	1.02	7	3
Arc	2.01	1.02	7	3
1700016P03Rik	2.01	1.02	7	3
Fbxo33	1.7	3.57	6	9
Nr4a2	1.7	2.4	6	8
Tiparp	1.7	2.4	6	8
Rgs4	1.7	1.95	6	7
Gpr3	1.7	1.29	6	4
Dusp4	1.52	2.4	5	8
Per1	1.52	1.95	5	7
Nefm	1.52	1.7	5	6
Klf9	1.52	1.29	5	4
Ablim2	1.52	1.29	5	4
Dot1l	1.52	1.29	5	4
Nt5dc3	1.52	1.29	5	4
Bdnf	1.52	1.02	5	3
Slc6a17	1.52	1.02	5	3
Mir22hg	1.52	1.02	5	3
Fndc9	1.3	1.95	4	7
Arl4d	1.3	1.7	4	6

Fam150b	1.3	1.7	4	6
Arid5a	1.3	1.7	4	6
A830082K12Rik	1.3	1.5	4	5
Plk2	1.3	1.5	4	5
Gramd1b	1.3	1.5	4	5
Sik2	1.3	1.5	4	5
Sik3	1.3	1.5	4	5
Fosl2	1.3	1.5	4	5
Hcn1	1.3	1.02	4	3
Gad2	1.3	1.02	4	3
Ubash3b	1.3	1.02	4	3
Lgi2	1.3	1.02	4	3
Smap2	1.3	1.02	4	3
Ptprn	1.3	1.02	4	3
Mark1	1.3	1.02	4	3
Hdac9	1.3	1.02	4	3
Sik1	1.3	1.02	4	3
Dusp6	1.3	1.02	4	3
Lin7a	1.3	1.02	4	3
Dclk1	1.3	1.02	4	3
Plppr5	1.3	1.02	4	3
Rell2	1.3	0.74	4	2
Dlx6os1	1.3	0.74	4	2
Gadd45g	0.99	3.57	3	9
Atf3	0.99	3.57	3	9
Ppp1r15a	0.99	2.4	3	8
Spty2d1	0.99	2.4	3	8
Btg2	0.99	2.4	3	8
Dusp1	0.99	2.4	3	8
Prkar2a	0.99	2.4	3	8
Pcsk1	0.99	1.95	3	7
Arl5b	0.99	1.95	3	7
Gpr19	0.99	1.95	3	7
Siah2	0.99	1.95	3	7
Jun	0.99	1.95	3	7
Dnajb1	0.99	1.95	3	7
Gm13889	0.99	1.95	3	7
Nr4a3	0.99	1.95	3	7
Zdbf2	0.99	1.95	3	7
Ddx3y	0.99	1.7	3	6
Gla	0.99	1.7	3	6
Rgs2	0.99	1.7	3	6

Nfil3	0.99	1.7	3	6
Otud1	0.99	1.5	3	5
Nefl	0.99	1.5	3	5
Ets2	0.99	1.5	3	5
Mlip	0.99	1.5	3	5
Csrnp1	0.99	1.5	3	5
Irs2	0.99	1.5	3	5
Rnf217	0.99	1.5	3	5
Otud3	0.99	1.5	3	5
Nrd1	0.99	1.5	3	5
R3hdm1	0.99	1.5	3	5
Myc	0.99	1.5	3	5
Arpp21	0.99	1.29	3	4
Erf	0.99	1.29	3	4
Ing2	0.99	1.29	3	4
Nap1l5	0.99	1.29	3	4
Dennd5b	0.99	1.29	3	4
Phyhipl	0.99	1.29	3	4
Rab3ip	0.99	1.29	3	4
Ccnl1	0.99	1.29	3	4
Dusp14	0.99	1.29	3	4
Josd1	0.99	1.29	3	4
Slc38a2	0.99	1.29	3	4
Igf1	0.99	1.29	3	4
Amigo3	0.99	1.29	3	4
Ddx10	0.99	1.29	3	4
Lrrc4c	0.99	0.74	3	2
Rgs10	0.99	0.74	3	2
Psd3	0.99	0.74	3	2
Pgap1	0.99	0.74	3	2
Sv2c	0.99	0.74	3	2
Tmem35a	0.99	0.74	3	2
Lrrtm1	0.99	0.74	3	2
Sgip1	0.99	0.74	3	2
Kcnj3	0.99	0.55	3	1
Ntsr1	0.99	0.55	3	1
Slc16a7	0.99	0.55	3	1
Kcnh7	0.99	0.55	3	1
Ism1	0.74	1.7	2	6
Ifrd1	0.74	1.7	2	6
Maff	0.74	1.5	2	5
Lbh	0.74	1.5	2	5

Syt4	0.74	1.5	2	5
Pcdh8	0.74	1.5	2	5
Chka	0.74	1.29	2	4
Cldnd1	0.74	1.29	2	4
Zfp948	0.74	1.29	2	4
Tnks	0.74	1.29	2	4
Tbc1d1	0.74	1.29	2	4
Rabgef1	0.74	1.29	2	4
Cited2	0.74	1.29	2	4
Errfi1	0.74	1.29	2	4
Pde4d	0.74	1.02	2	3
Stmn4	0.74	1.02	2	3
Tsc22d1	0.74	1.02	2	3
Mbnl2	0.74	1.02	2	3
Atf4	0.74	1.02	2	3
Ciapin1	0.74	1.02	2	3
Slitrk4	0.74	1.02	2	3
Cyp51	0.74	1.02	2	3
Mon2	0.74	1.02	2	3
Camk2n2	0.74	1.02	2	3
Lrrn1	0.74	1.02	2	3
Pmvk	0.74	1.02	2	3
Eprs	0.74	1.02	2	3
Omg	0.74	0.55	2	1
Klf4	0.44	2.4	1	8
1190002N15Rik	0.44	1.95	1	7
Shisa2	0.44	1.5	1	5
Pvr	0.44	1.5	1	5
Emd	0.44	1.5	1	5
Kitl	0.44	1.5	1	5
Sap18	0.44	1.29	1	4
Gltscr2	0.44	1.29	1	4
Bmt2	0.44	1.29	1	4
Crem	0.44	1.29	1	4
Sgk1	0.44	1.29	1	4
Rassf1	0.44	1.29	1	4
Abhd17b	0.44	1.02	1	3
Dxo	0.44	1.02	1	3
Ier3	0.44	1.02	1	3
Gpr85	0.44	1.02	1	3
Cyr61	0.44	1.02	1	3
Nr2f1	0.44	1.02	1	3

Sap18b	0.44	1.02	1	3
Sst	0.44	1.02	1	3
Bhlhe22	0.44	1.02	1	3
Pcdh17	0.44	0.74	1	2
B3galt2	0.44	0.74	1	2
Klf6	0.44	0.74	1	2

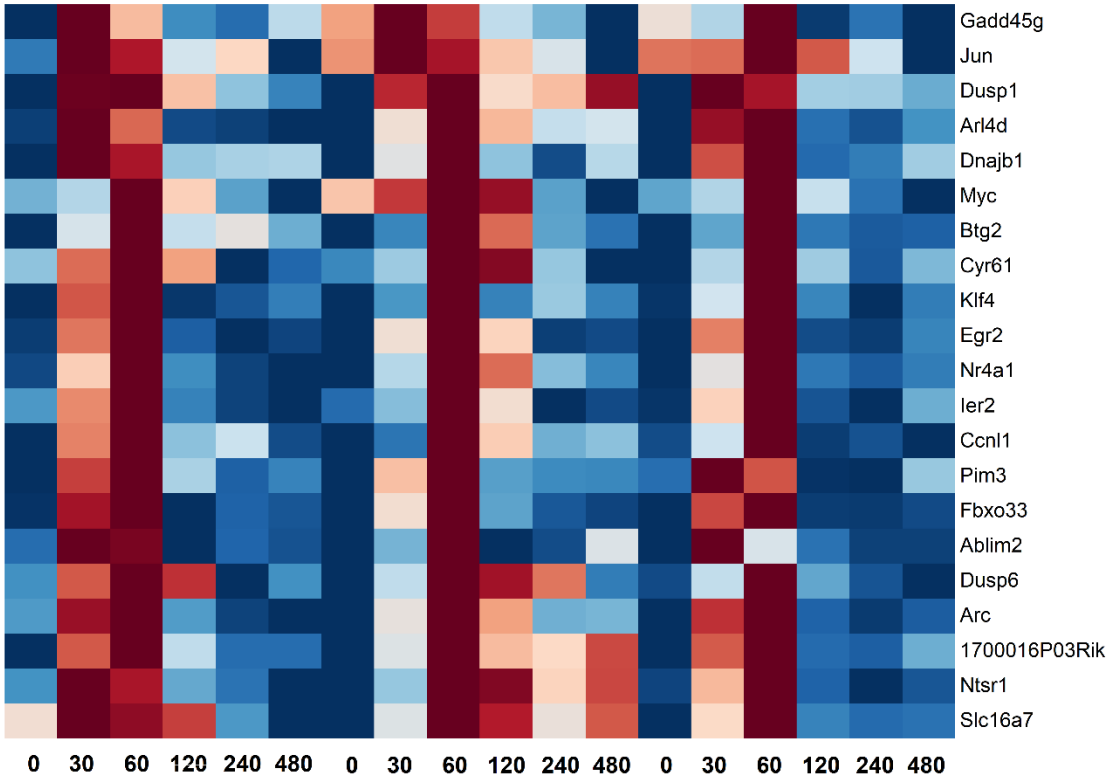
A2				
Gene name	slope 1	slope 2	module 1 number	module 2 number
Nptx2	3.68	3.57	9	9
C2cd4b	2.49	2.4	8	8
Gadd45b	2.49	2.4	8	8
Npas4	2.01	1.95	7	7
Gm11290	1.52	1.5	5	5
Fosb	1.52	1.5	5	5
Kcna4	1.3	1.29	4	4
Slco3a1	1.3	1.29	4	4
Cbx8	1.3	1.29	4	4
Nrsn1	1.3	1.29	4	4
A830010M20Rik	1.3	1.29	4	4
Cpeb4	1.3	1.29	4	4
Rprml	1.3	1.29	4	4
Elmo1	0.99	1.02	3	3
Sgtb	0.99	1.02	3	3
Hmgcs1	0.99	1.02	3	3
Gad1	0.99	1.02	3	3
Dmxl1	0.99	1.02	3	3
Nolc1	0.99	1.02	3	3
11-Mar	0.99	1.02	3	3
Trib1	0.99	1.02	3	3
Cxadr	0.99	1.02	3	3
Rcan2	0.99	1.02	3	3
Ppm1b	0.99	1.02	3	3
Adcyap1	0.99	1.02	3	3
Usp3	0.99	1.02	3	3
Rasa2	0.99	1.02	3	3
Ppp1r37	0.99	1.02	3	3
Rab6a	0.99	1.02	3	3
Ppme1	0.99	1.02	3	3
Lmo1	0.99	1.02	3	3

Cars	0.99	1.02	3	3
Msmo1	0.99	1.02	3	3
Kcnn1	0.99	1.02	3	3
Pcdh19	0.99	1.02	3	3
Tceal5	0.99	1.02	3	3
Acsl4	0.99	1.02	3	3
Rheb	0.99	1.02	3	3
Atp8a1	0.99	1.02	3	3
Inpp5f	0.99	1.02	3	3
Asns	0.99	1.02	3	3
Slc2a3	0.99	1.02	3	3
Clstn3	0.99	1.02	3	3
Mif	0.99	1.02	3	3
Ndel1	0.99	1.02	3	3
Ksr1	0.99	1.02	3	3
Hdac5	0.99	1.02	3	3
Penk	0.99	1.02	3	3
Brinp1	0.99	1.02	3	3
B230216N24Rik	0.99	1.02	3	3
Map3k9	0.99	1.02	3	3
Ube2ql1	0.99	1.02	3	3
Hmgcr	0.99	1.02	3	3
Grin1	0.99	1.02	3	3
Slc25a25	0.99	1.02	3	3
Hars	0.99	1.02	3	3
Dpysl3	0.99	1.02	3	3
Cep120	0.99	1.02	3	3
Malat1	0.99	1.02	3	3
Pard6a	0.99	1.02	3	3
Pnoc	0.99	1.02	3	3
Ly6h	0.99	1.02	3	3
Ubald1	0.99	1.02	3	3
Dgkk	0.99	1.02	3	3
Srf	0.99	1.02	3	3
Arxes1	0.99	1.02	3	3
Rps6ka3	0.99	1.02	3	3
Rasgrf1	0.99	1.02	3	3
Map9	0.99	1.02	3	3
Adar	0.99	1.02	3	3
Zrsr1	0.99	1.02	3	3
Pitpna	0.99	1.02	3	3
Tanc2	0.99	1.02	3	3

Coq10b	0.99	1.02	3	3
Tmem198	0.99	1.02	3	3
Nmnat2	0.99	1.02	3	3
Dbpht2	0.99	1.02	3	3
Flrt2	0.99	1.02	3	3
Mirg	0.99	1.02	3	3
Dip2c	0.99	1.02	3	3
Slco5a1	0.99	1.02	3	3
Ctage5	0.99	1.02	3	3

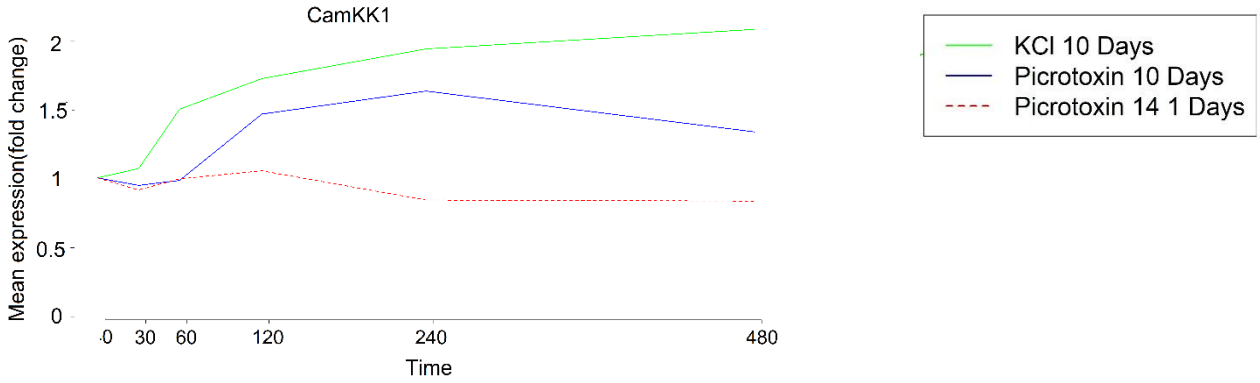
S3.B Only IEGs of the common set of 3 experiments. The fit columns are the slope of each gene in the pairwise comparison. Module 1/2 are the cluster number the gene is in.

Gene name	Fit picrotoxin 14 to picrotoxin 10	Fit KCl 10 to picrotoxin 10	Module number 1	Module number 2	cluster diff
Ccnl1	1.18	1.3	3	4	1
Dusp6	1.32	1.18	4	3	1
Ablim2	1.45	1.26	5	4	1
Ier2	2.01	1.63	7	6	1
Pim3	1.82	2.46	7	8	1
Cyr61	0.32	0.94	1	3	2
Ntsr1	0.96	0.52	3	1	2
Slc16a7	1.01	0.63	3	1	2
Myc	1.08	1.57	3	5	2
Arl4d	1.22	1.8	4	6	2
Egr2	3.21	1.93	9	7	2
Fbxo33	1.73	3.15	6	9	3
Jun	0.82	2.12	3	7	4
Dnajb1	1.19	2.14	3	7	4
Arc	2.19	1.16	7	3	4
1700016P03Rik	2.11	0.93	7	3	4
Nr4a1	3.28	1.46	9	5	4
Btg2	0.81	2.37	3	8	5
Dusp1	1.07	2.33	3	8	5
Gadd45g	1.04	3.79	3	9	6
Klf4	0.35	2.61	1	8	7

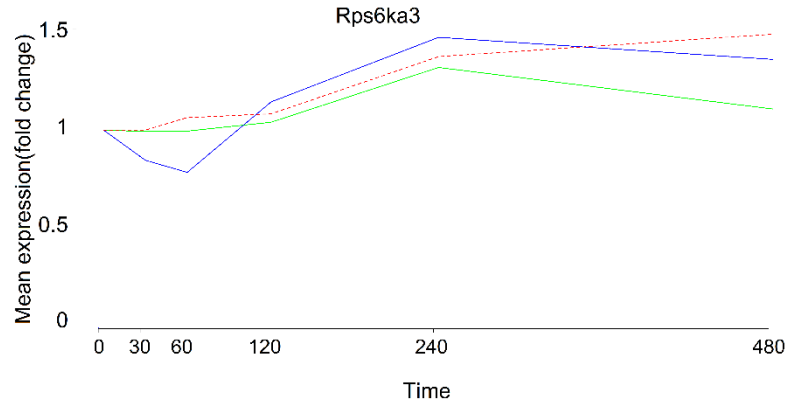


S3.C. IEG's gene expression heatmap.

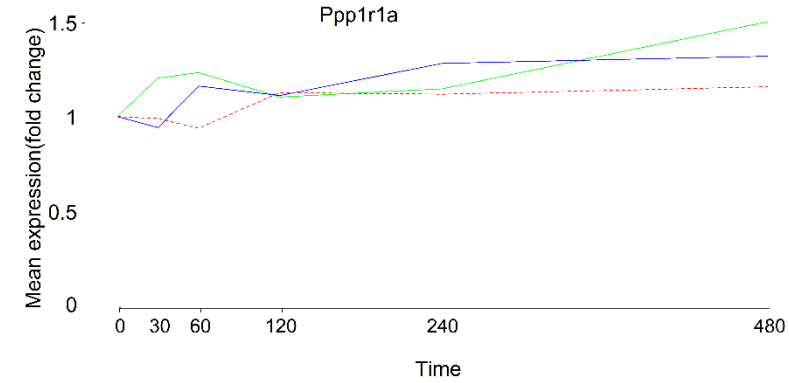
A


CamKK1
Mean expression(fold change)
2
1.5
1
0.5
0
0
30
60
120
240
480
Time
KCl 10 Days
Picrotoxin 10 Days
Picrotoxin 14 1 Days

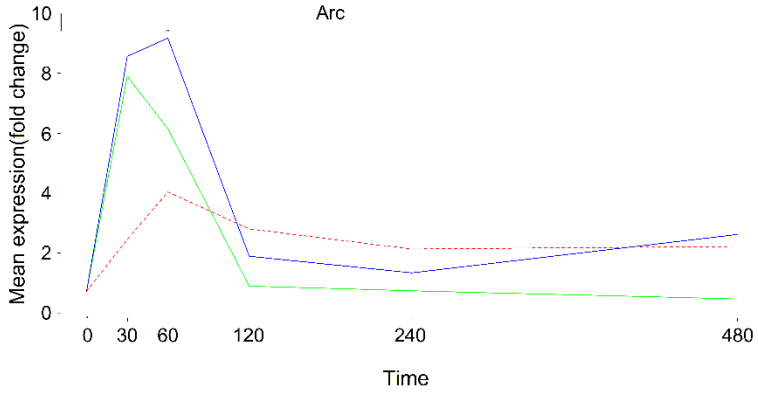
B


Rps6ka3
Mean expression(fold change)
1.5
1
0.5
0
0
30
60
120
240
480
Time

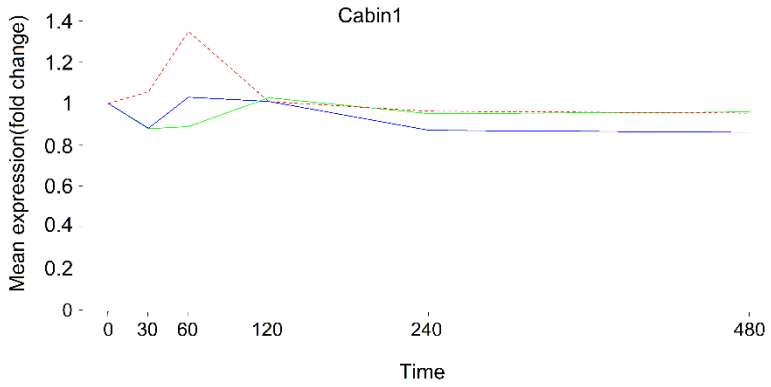
C


Ppp1r1a
Mean expression(fold change)
1.5
1
0.5
0
0
30
60
120
240
480
Time

D



E



S4: Line graph of individual genes. The values are in mean fold change.

S5 GO terms table

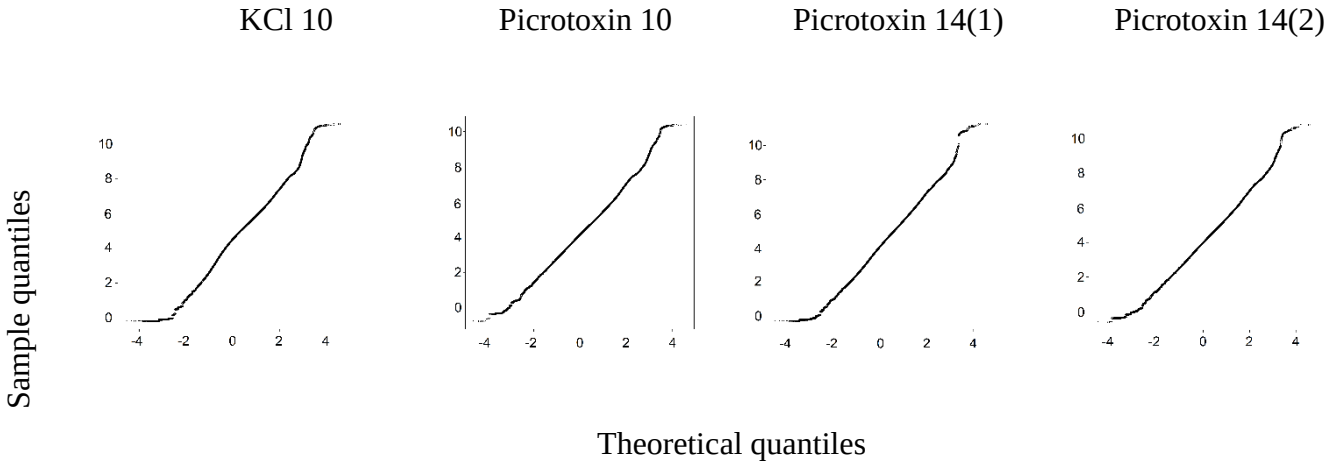
GUESSES	c	h	b	t_0	σ
1	Mean of control	0.01	0.001	Argmax(y_real)	10
2	Mean of control	2	0.2	Argmax(y_real)	10
3	Mean of control	10	0.2	Argmax(y_real)	10

4	Mean of control	10	0.001	Argmax(y_real)	200
5	Mean of control	10	0.001	Argmax(y_real)	8
6	Mean of control	0	0	480	10
7	100	$(max(real\_y) - min(real\_y\_i)) \cdot \frac{1}{2}$	0.399	Argmax(y_real)	10

S6: The 7 initial guesses the GS model uses for each gene, real_y is the gene data in FC, calculated from the normalized read count matrix.

S7 - Why can we use F test?

An F test is done only for normally distributed data. Gene expression or fold change data in our case is not distributed normally usually, in order to overcome this obstacle, I converted the values into log2 values which is supposed to spread the data more proportionally and thus create a normally distributed data. To show our data was indeed normally distributed a Q-Q plot was created. A Q-Q plot is a scatter plot created by two sets of quantiles plotted against one another. If both sets of quantiles came from the same distribution, we will see a roughly straight line.



Q-Q plots. Compare between our data distribution in each quantile and compare it to a normal distribution which is derived from the same mean and variance.