

Work report on RNA-seq gene expression of hippocampal principal neurons

qin.he

October 2019

1 Introduction

The two typical excitatory cell types granule cell(GCs) and mossy cells of DG in chronically epileptic mice identified by optogenetic tagging are studied through the electrographic data of the seizure. To investigate the consequences of MC and GC loss on cognition, the activities of the cell neurons are recored by MEA. Further, gene expression profiling is applied to understand the functionality of them in addition to the intrinsic level. The major excitatorial neuronal classes of the hippocampus, GC, and MC of the dentate gyrus reveals some unknown properties to the epilepsy. The mouse model with hippocampal sclerosis, uses a combintion of optogenetic, electrophysiological, and behavioral approaches connecting the light-response features to gene expression in mammalian cells.

2 MEA analysis

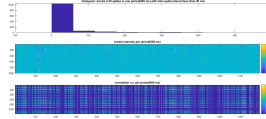
There are 5 ictal periods recorded (10 bins of interested separately.). For the whole dataset, the auto-correlation is calculated to detect the onset of seizure and preprocess including baseline correlation and normalization and filtering are applied. As the seizure is usually recorded on high frequency band compared to other signals, for instance the signal for learning tasks, the MEA signal is studied comparing different bands as well as on the whole 180 s time period. According to the 1, the signal is oscilated evenly on the gamma band (25 Hz to 100 Hz) and not detected as spikes at delta, theta, alpha, mu, SMR and beta bands according to the fpkm values (on log level) of the genes in the range -3 to 4. Specifically, bursts(defined as more than 30 spikes detected in one period: 500 ms with inter-spike interval less than 20 ms) are detected mainly in the time period: 0-80ms and is sparsely detected at 80ms to 410 ms(in the first 500 ms around ictal 1). The intensity is increased at 100-110 ms, 460-480 ms, 790-810 ms and 1100-1200 ms significantly without significant correlation. (See 2) As the time window shows negative correlated (Pearson) round 100 ms, it is not significantly correlated on any band with only a little bit synchronization on delta and gamma band (3). And finally, the 10 bins are classified into narrow

and broad spikes according to its asymmetry, duration, quantile and deviation distribution. (4) As the majority of the spikes are biased, they are classified as broad type.



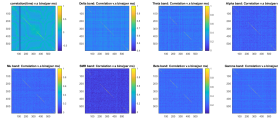
and f).png

Figure 1: STPS



and corr).png

Figure 2: bursts



and f).png

Figure 3: correlation

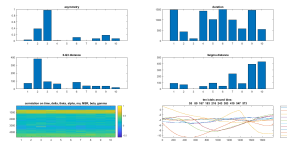


Figure 4: ictals

3 genetic analysis

As study of cellular processes existing inside intact organisms requires methods to visualize cellular functions such as gene expression in deep tissues, the 75 expression dataset (dg-d, dg-v and MC cell separately according to ??) is studied along with the gene regulation as the substrate of the observable trait at which the genotype gives rise to the phenotype (further cellular differentiation

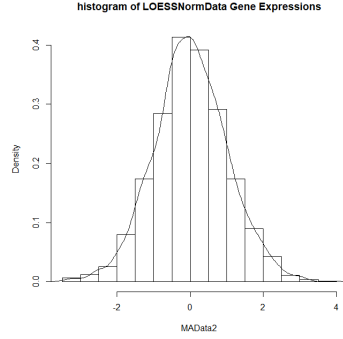
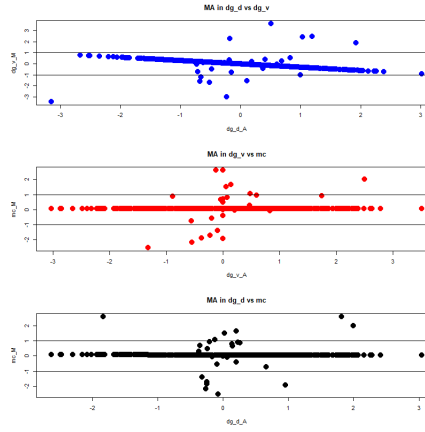


Figure 5: normalized dataset

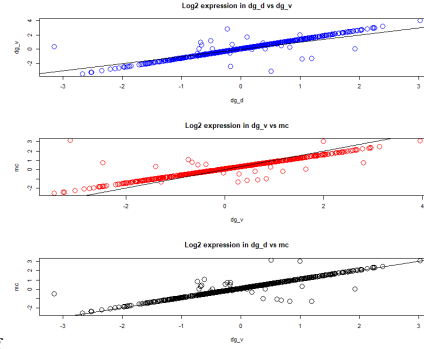


dg mc extended.png

Figure 6: MA plot

adaptability of organism in the versatility and adaptability). After the LOESS normalization, it is normalized according to QC test(5) and most genes of mc and dg-d are expressed stationarily with MA value mainly distributed around zero (since the expression of genes are assumed to be unchanged at certain tissue region.) while has decreased trend for the dg-v expression.(6). Detailedly, the dg-v to dg-d ratio is shown with higher skewness in the top figure while with skewness around 0 of mc to dg-v and mc to dg-d in the two bottom figures. The 13, 16 and 18 numbered genes are relatively densely expressed. As the profound effect on the functions (actions) of the gene in a cell is usually studied as a multicellular organism, the transcription, RN splicing, translation and etc. of a protein is usually modulated.

The data is 25 dimensioned with some specific biomarkers (For instances, the GTPases related Cdc45). Thus, the MDS plots for the first, second and third can be seen here: 8. The cell with different markers can be seen with their



expression dg mc extended.png

Figure 7: log2 after fdr correction

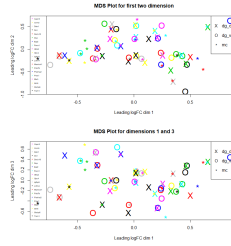


Figure 8:

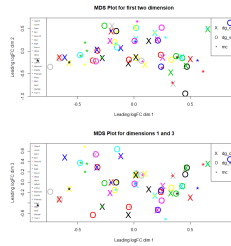


Figure 9: MDS on the first two dimensions

different expression levels. As the 13 and 18 marked gene, Prox1 and St18 (red and green) are expressed with larger density in dimension 2 than dimension 1, where the marks1 (yellow) are contrarily distributed evenly of different levels on both dimension 1 and dimension 2(along with extended dataset 58 53 74 26 18 34 66 and 42, 50 19 11 67 and 43) expressed more in cell 1 and cell 3, namely dg-d and mc. To be more specific, as shown in 9 marks1 is accumulated more centrally on dimension 2 than dimension 1(along with the extended synthetic dataset 31 and 47) expressed higher on dimension 1 only in cell 2. Similarly, for dimension 1 and 3, the Prox1 and St18 are expressed evenly on both dim 1 and

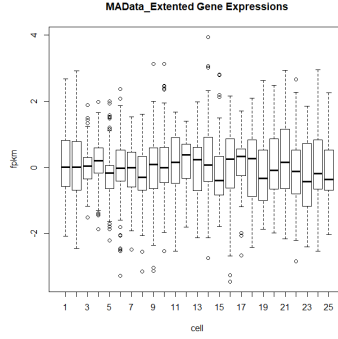
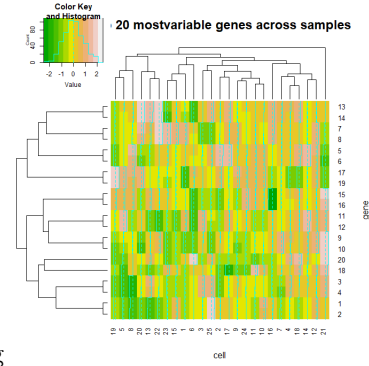


Figure 10: fpkm expression (on log level)

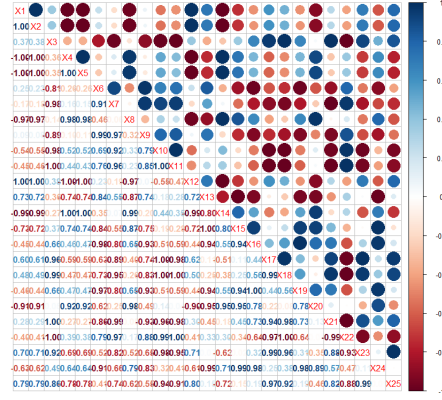
dim 3 while Marckl1 distributed only higher in cell 2.



heatmap of normed cells extended.png

Figure 11: top 20 expressed cell

And after studying the correlation between different genes(12), the cells and finally extended to 75 datasets for each cell and clustered(14) are classified based on the TF(13). The 8th, 9th, 20th, 21st, 22th, 23th and 25th are less correlated with 1st, 2nd, 7th, 11th, 15th, 20th and 24th individually. On average, the genes are more correlated positively rather than negatively. The extended TF is shown with the heat-map, giving rise to the density based cluster of top 20 genes across the 75 datasets (11). The highest expressed genes(negatively) are the first 4 genes in 19th, 5th, 8th and 20th dataset and the lowest for those genes are in the 6th, 3rd, 9th, 24th, 4th and 14th datasets are the first 4 genes in 19th, 5th, 8th and 20th dataset and the lowest for those genes are in the 6th, 3rd, 9th, 24th, 4th and 14th datasets while the positively highest expressed genes, 15th, 9th, 10th and 20th are all on the 21st dataset. The top 5 clusters(with lowest variant) in the dendrogram is marked in red, green yellow, blue and purple.(With dataset 1 47 72 22 31 56 6 , 39 14 64 28 3 53, 54 4 29, 27,2,52 and 30, 3, 55.) As cells can be classified by position of maximum changes, next



of TF.png

Figure 12: heatmap

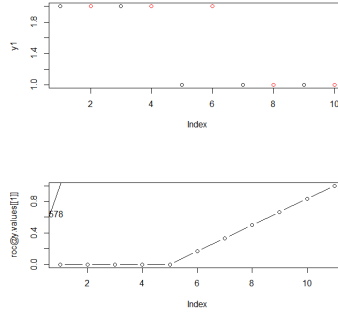
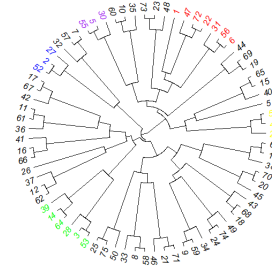


Figure 13: one classification example

study will be about the transcriptions of the cells giving out the relationships of the expression changes both temporarily and spatially.

4 Conclusion and discussion

In summary, according to 15 the correlation is largest within one window on delta band and gamma band (with the maximum around 90 ms. The histogram and the gap can be seen distributed with gamma and normal distribution. According to the filtered signal, the type of the spike is labeled with asymmetry (most weighted factor), quantile, duration of the spike which is related more to the time information. (The classification results is not with high enough accuracy mainly because the 10 ictal signal bins after the ANOVA is not distributed nor-



of normed genes(extended).png

Figure 14: dendrogram of the clusters

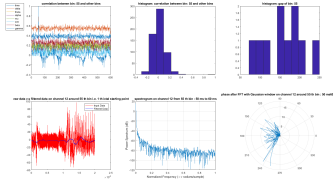
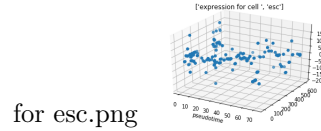


Figure 15:



for esc.png

Figure 16:

mally and will be modified in the next report.) However, the spectrogram and polar gram shows the information on phase domain which reveals the phase-lock property of the electrode signal also supporting the truth that the seizure component of the signal fires at the higher frequency. With the normalized expression data ($p = 0.0035$ with ks-test), Genes expressed in the 3-sigma distance, Cdc45, Prox1, St18, Marsckl1, SOX2, Cdk4 and etc. known to have distinct functions are shown in the heatmap and accordingly posses correlation with close to zero in the correlation figure. However, to explore the difference of the function on cellular scale, the welch two sampled t-test is applied. According to the result, the gene expression for dg-d, dg-v and mc($p = 0.009724$, $p = 0.008046$ and $p = 0.008584$) are not of the some mean on log2 level. Generally, dg-d and mc are expressed more evenly while dg-v is distributed on some specific dimeension (correction method will be applied in the next step exploration.) Further more, there can be further exploration on the cell difference and evolutaion with both

time and structure information.(TF will be analyzed with pseudo-time in the next step.¹⁶ There is the trial of the pseudo-time analysis shown in the first class of esc with first 70 dataset distributed with the coordinate(2D) recorded by the prone and expression in the microarraydata.) It might be still studied on log2 level in the further exploration.