

Early Detection of Temporal Lobe Epilepsy: Identification of Novel Candidate Genes and Potential Biomarkers Using Integrative Genomics Analysis

Vishruth Nagam

BioScience Project, Wakefield, USA

Email: vnagam@outlook.com

How to cite this paper: Nagam, V. (2020) Early Detection of Temporal Lobe Epilepsy: Identification of Novel Candidate Genes and Potential Biomarkers Using Integrative Genomics Analysis. *Open Journal of Genetics*, 10, 65-81.
<https://doi.org/10.4236/ojgen.2020.104006>

Received: October 30, 2020

Accepted: December 27, 2020

Published: December 30, 2020

Copyright © 2020 by author(s) and Scientific Research Publishing Inc. This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

Currently afflicting more than 50 million people worldwide, epilepsy is the spectrum disorder characterizing seizures that occur without other plausible medical explanations. Temporal lobe epilepsy (TLE) is one of the most common forms of epilepsy. Current clinical methods; including MRI scans, EEG tests, and doctor visits; can take upwards of several months to confirm a TLE diagnosis; during this time, patients may experience additional seizures and are at an increased risk for other psychiatric disorders. The purpose of this study is to identify candidate genetic biomarkers to facilitate the earlier detection and diagnosis of TLE through gene-based testing (e.g., genomic heatmap analysis or genetic and/or microarray testing). It was hypothesized that potential biomarkers could be identified by analyzing genes that are normally significantly overexpressed in the temporal lobe relative to the gray matter. Statistical and functional analysis was performed on significantly overexpressed genes (≥ 3.000 fold change) in the gene expression profiles of four donors without epilepsy. The experimental-evidence-based STRING protein interactions analysis showed associations between genes found in DAVID keyword search and other genes facilitating network interconnectivity. After evaluation of the genes' STRING enriched functions, changes in the expression of the genes *CAMK2A*, *NPY*, *DLG4*, *MEF2C*, and *MAPK7* were concluded to be potential biomarkers for TLE, confirming the original hypothesis. Specifically, the identification of *MEF2C* and *MAPK7* for this purpose is relatively novel in the fields of bioinformatics and neurogenetics. Future work includes investigating the utility of the candidate genes in real-world gene-based diagnostic methods.

Keywords

Temporal Lobe Epilepsy (TLE), Gene Expression, Fold Change Values, Gene Ontology, Candidate Genes, Keyword Search, Four Individual Donors, STRING Enriched Functions, Protein, Genes Highest Fold

1. Introduction

When electrical activity in the brain is abnormal or irregular for a brief period of time, events called seizures can result [1]. Seizures take place suddenly and last from seconds to over five minutes [1] [2]. A person may experience changes in their behaviors, body movements, and level of awareness (consciousness) [1] [2]. Seizures can also come with other psychiatric or intellectual problems, including suppression of dendritic growth, which can impair learning and certain aspects of emotion [2].

There are several types of seizures, classified based on the affected areas of the brain. Occurring on both sides (hemispheres) of the brain, generalized seizures include *petit mal* or absence seizures (e.g., cause few seconds of staring or rapid blinking) and *grand mal* or tonic-clonic seizures (e.g., make the person cry out loudly, collapse, shake, or even lose consciousness) [2]. Focal seizures are the most common form of seizures (most common form in adults and one of the most common forms in children) and are localized to one area of the brain [2]. Secondary generalized seizures begin in a single brain region, but spread to both hemispheres of the brain to become generalized seizures [2].

Epilepsy is a condition that someone is diagnosed with when more than one such seizure occurs without any possible explanation by an underlying, non-permanent medical condition [2]. More than 50 million people have epilepsy worldwide [3]. Having many possible causes, epilepsy is thus considered, not a single disorder, but a spectrum [1] [2]. In addition, epilepsy is more common in infants ages 1 and below and in seniors ages 65 and above, which could possibly be attributed to comorbid neural phenomena (tumors, strokes, etc.) [1].

TLE is the most common form of focal epilepsy, accounting for focal seizures in the temporal lobe [1]. The temporal lobes lie by the ears, on the sides of the brain, and are involved in speech and language interpretation and production, visual and auditory processing, and involuntary or unconscious responses (e.g., hunger and sexual arousal) [1] [2]. In most cases, TLE can even further be localized to the mesial (*i.e.*, medial or inner) or lateral (*i.e.*, outer) regions of the temporal lobe. Nearly 80 percent of temporal lobe seizures can be attributed to mesial TLE, which often originate in the hippocampus or nearby areas [3]. About 75 percent of TLE cases can be traced to specific causes, including traumatic brain injuries, brain inflammation, and genetic mutations or factors [1].

Utilizing gene expression as an indicator of the presence of neurological disorders and conditions is already an established practice [4]. However, in clinical settings, TLE is currently diagnosed with a combination of magnetic resonance

imaging (MRI) scans (to detect changes in the temporal lobe), electroencephalogram (EEG) tests (to reveal abnormal neuronal electrical activity), and patients' descriptions of symptoms [3]. This process can take upwards of several months after symptoms first appear. During this time, patients may experience additional seizures and are at an increased risk for anxiety, depression, and other psychiatric disorders [5]. It is thus important to diagnose TLE before patients' symptoms worsen and susceptibility to other conditions increases.

In this study, earlier diagnosis of TLE is proposed using genomic heatmap analysis or genetic and/or microarray testing. Changes in gene expression could suggest potential genetic causes of TLE, indicate the presence of TLE in an individual before symptoms manifest, and allow for antiepileptic treatment to ease the onset of TLE. A previous study reported that a lack of the *GABA(B[1])* gene; which encodes the subunit receptor 1 of the gamma-aminobutyric acid (GABA) receptor subtype B and facilitates the inhibitory postsynaptic effects of GABA, the most prominent inhibitory neurotransmitter in the central nervous system (*i.e.*, brain and spinal cord); was associated with TLE in mice [6]. Other susceptibility genes and potential biomarkers for TLE include *PDYN* (prodynorphin) and the $\epsilon 4$ variant of *APOE4* (apolipoprotein E) [7].

The purpose of this project is to observe and examine gene expression patterns in the temporal lobe in order to identify candidate genes for TLE. It is hypothesized that certain genes normally significantly overexpressed in the temporal lobe compared to in the gray matter—as revealed through genomic heatmaps—could potentially be genetic biomarkers for TLE, while reflecting biological functions related to TLE.

2. Methods

The Allen Human Brain Atlas by the Allen Institute for Brain Science is a database that was used as the primary source for the gene expression data concerning the temporal lobe [8]. Heatmaps generated from microarray surveys of neural tissue that was donated to the Allen Institute for Brain Science were analyzed for the gene expression data of interest. A differential search was conducted comparing the gene expression in the temporal lobe to that in the gray matter. The data, released in 2010 and last updated in 2013, was obtained from four different donors (H0351.2001, H0351.1009, H0351.1012, and H0351.1015) between 18 and 68 years of age and with no known history of neuropsychiatric or neurological conditions, as gene expression data from individuals with epilepsy was not available in this database (see “Limitations and Future Directions”) [8]. Genes with a fold change of 3.000 or greater were collected for further statistical analysis.

The Venny 2.1 software was used to identify genes common to the four individual donors' gene lists [9]. The data was visually depicted with a Venn diagram.

The Google Sheets software was used to identify the top twenty genes with the highest fold change values in the gene lists of the four individual donors. The

data from the four gene lists was visually displayed in four bar graphs, with each bar graph corresponding to a gene list from a single donor. The software was also used to conduct further statistical analysis (*i.e.*, variance and other common statistical values for the fold change values of the gene lists).

PythonAnywhere is an online integrated development environment utilizing the Python programming language [10]. It was used as a tool for the keyword search in DAVID, which is an online bioinformatics resource developed by the Laboratory of Immunopathogenesis and Bioinformatics [11]. Genes involved in functions associated with epilepsy were identified with the DAVID keyword search.

The Gene Ontology is a bioinformatics resource used to classify the genes common to the four donors (as determined through Venny 2.1) by the three domains of gene ontology: Molecular Function, Cellular Component, and Biological Process [12]. Molecular Function refers to the molecular activities that the gene product performs. Cellular Component concerns where in the cell or in what cellular structures the gene product functions. Biological Process refers to the broader, overarching processes that the gene product helps to accomplish.

The STRING Database was used to construct protein interaction networks based on genetic data to find candidate genes for TLE. It was also used to identify other genes, interactions, and pathways of interest. The first shell and, if needed, the second shell were selected to have a maximum of fifty and ten interactors, respectively. The confidence level was set at medium (0.400) to ensure that enough interactions were generated for analysis. Of the different levels of evidence, ranging from theoretical (least valid) to experimental (most valid), only experimental active interaction sources were used to generate the interactions.

3. Results and Discussion

3.1. Heatmap Analysis

The heatmap is a technique that was used to visualize the microarray values for the gene expression profiles of the four donors mentioned (refer to **Figure 1**). The bar at the top of the heatmap is of a unique color specific to each donor, indicating the donor that has been selected. The bar below represents the different brain regions in the heatmap. Each row and column of the heat map represent a particular gene and a tissue sample, respectively. The color of each box is indicative of the z-score over a probe ranging from red (z-score of +3 and above) to black (z-score of 0) to green (z-score of -3 and below).

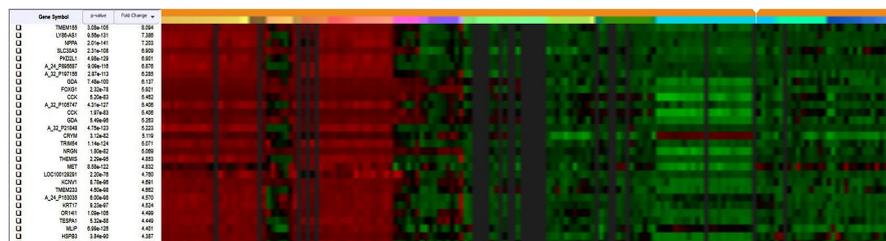


Figure 1. Allen Human Brain Atlas heatmap of Donor H0351.2001 generated by a differential search comparing gene expression in the temporal lobe to that in the gray matter.

The gene count varied noticeably among the donors. Donor H0351.2001 had 145 genes with a fold change greater than 3.000, while Donors H0351.1009, H0351.1012, and H0351.1015 had 184, 246, and 330 such genes, respectively.

3.2. Venny 2.1 Analysis

The venn diagram (refer to **Figure 2**) shows the common genes between, as well as the genes unique to, the gene lists of the four individual donors. 76 (29.3%) genes were identified to be commonly expressed in the temporal lobe amongst all four donors.

3.3. Bar Graphs

The bar graphs (refer to **Figure 3**) show the twenty genes with the highest fold change values from each of the four individual donors. The x-axis displays the symbols of the genes, and the y-axis displays the fold change values. Ten genes (*TMEM155*, *LY86-AS1*, *NPPA*, *PKD2LI*, *A_24_P895687*, *GDA*, *FOXL1*, *CCKA_32_P105747*, and *THEMIS*) appeared in all four lists of the twenty genes with the highest fold change values. Genes that appeared in exactly three of the four lists include *NRGN*, *HSPB3*, *SLC30A3*, and *A_32_P21848*.

3.4. Statistical Analysis

The statistical analysis (refer to **Table 1** for a full list of statistical values for each donor) for the 300 genes with the highest fold change values shows that the mean fold change ranges from 3.26 to 4.47 across the gene lists. Lower mean fold changes tend to be associated with lower minimum and maximum fold changes; however, from a qualitative standpoint, the minimum fold changes, spanning within a range of 2.43 to 3.11, did not differ greatly from donor to donor. The variance values amongst the donors are also relatively similar to each other, ranging from 0.85 to 2.35. Standard deviation appears to be quite similar amongst the donors, between 0.92 and 1.53.

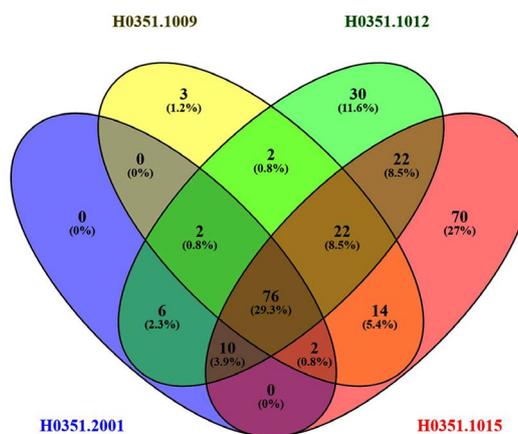


Figure 2. Venn diagram generated through Venny 2.1 depicting the common genes expressed in the temporal lobe between the four individual donors.

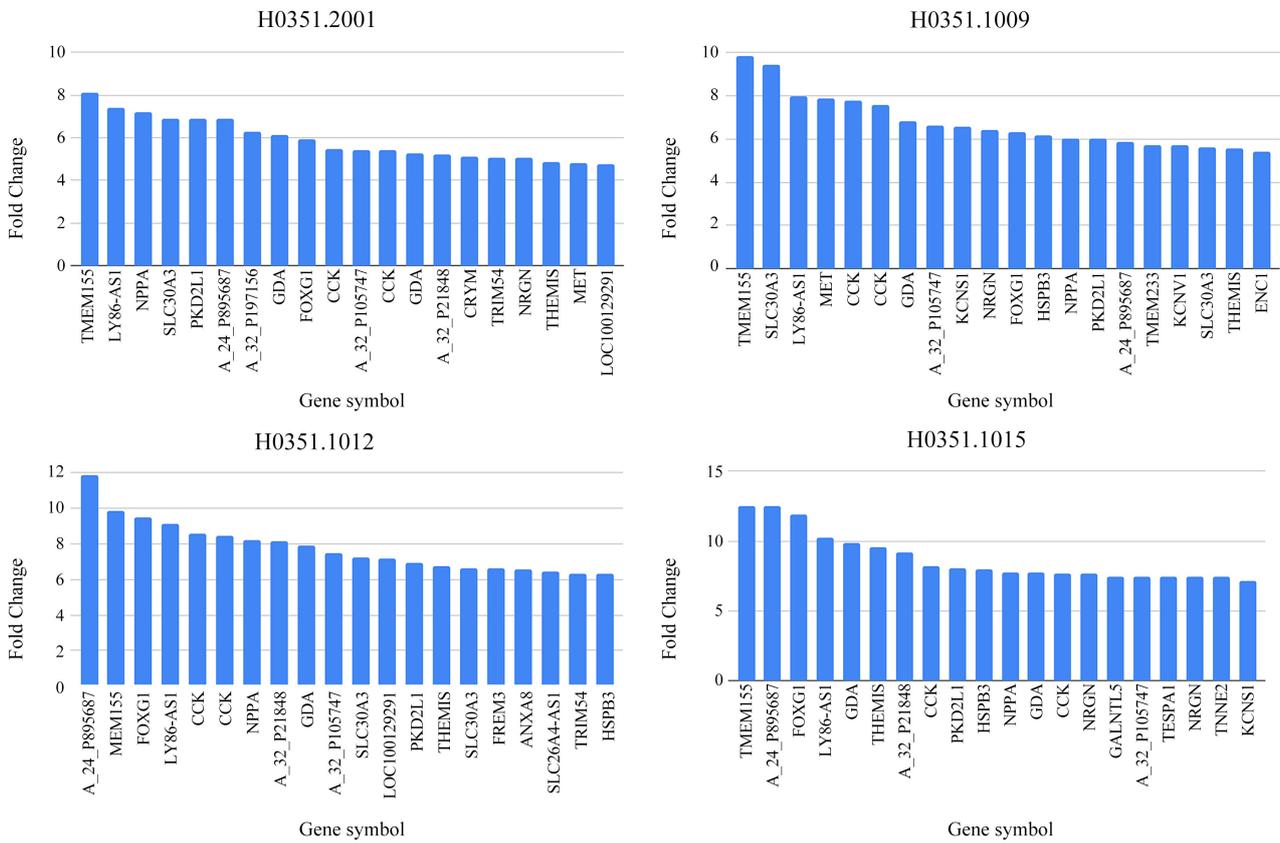


Figure 3. Bar graphs generated through Google Sheets depicting the fold change values of the top twenty genes in the gene lists of each of the four individual donors.

Table 1. Statistical analysis table for the gene lists of the four individual donors. The 300 genes with the highest fold change values for each donor were used.

	H0351.2001	H0351.1009	H0351.1015	H0351.1012
COUNT	300	300	300	300
MEAN	3.25939	3.506773333	4.472383333	3.921836667
STD	0.9228062649	1.102981348	1.532871011	1.321390844
MIN	2.431	2.578	3.11	2.803
Q1 (25%)	2.623	2.8135	3.455	3.073
Q2 (50%)	2.9725	3.1585	3.9995	3.4545
Q3 (75%)	3.58025	3.6805	4.83625	4.18825
MAX	8.094	9.869	12.481	11.902
VARIANCE	0.8515714026	1.216567855	2.349693535	1.746073763

3.5. DAVID Keyword Search

In the keyword analysis, the genes common to the four individual donors' gene lists were used to determine possible candidate genes for TLE. Within the DAVID bioinformatics database, the "GOTERM_BP_DIRECT," or biological processes, section of the Functional Annotation Table for the *Homo sapiens* species was used to find genes (refer to Figure 4 and Figure 5 for examples of the search methods used) associated with functions related to TLE.

The genes of interest (*MEF2C*, *HTR2A*, *SERPINF1*, etc.) tend to be related to excitatory postsynaptic potentials (*i.e.*, the increase in a postsynaptic cell's membrane potential resulting from the diffusion of sodium ions into the cell through ligand-gated sodium ion channels), inflammatory responses, and the neurotransmitters glutamate (excitatory) and GABA (inhibitory) [13]. In light of the known positive effects of a high-fat, low-carbohydrate (ketogenic) diet on remediating epilepsy symptoms, some genes were associated with saccharides and lipids as well (refer to **Table 2** for a full list of keywords used and the corresponding genes of interest) [2].

```

1 #!/usr/bin/python3.5
2 import pandas as pd
3 pd.set_option('display.max_rows', 500)
4 searchWord = "glutamate"
5 df = pd.read_csv('daviddemolists2+1.csv')
6 df2 = df[df['GOTERM_BP_DIRECT'].str.contains(searchWord) == True]
7 print (df2['Gene Name'])

```

Figure 4. Python script used for the keyword search in the DAVID bioinformatics database, with an example containing the search word “glutamate” shown above.

37711_at	myocyte enhancer factor 2C(MEF2C)	Related Genes	Homo sapiens
BIOCARTA	ALK in cardiac myocytes, Angiotensin II mediated activation of JNK Pathway via P42 dependent signaling, GABRI1 and Regulation of the Estrogen Receptor Role of ERK5 in Neuronal Survival, Control of skeletal myogenesis by HDAC 3 calcium/calmodulin-dependent kinase (CaMK), MAPKase Signaling Pathway, Signal Dependent Regulation of Myogenesis by Corepressor MIB3, NFAT and Hypertrophy of the heart (Transcription in the broken heart), p38 MAPK Signaling Pathway, Regulation of PGC-1 α .		
COG_ONTOLOGY	Transcription,		
GOTERM_BP_DIRECT	negative regulation of transcription from RNA polymerase II promoter, MAPK cascade, blood vessel development, osteoblast differentiation, neuron migration, B cell homeostasis, heart looping, mitochondrial ossification, blood vessel remodeling, chondrocyte differentiation, germinal center formation, regulation of germinal center formation, response to ischemia, primary heart field specification, secondary heart field specification, outflow tract morphogenesis, sinoatrial valve morphogenesis, cardiac ventricle formation, transcription, DNA-templated, regulation of transcription, DNA-templated, transcription from RNA polymerase II promoter, apoptotic process, humoral immune response, nervous system development, heart development, muscle organ development, skeletal muscle tissue development, muscle cell fate determination, learning or memory, response to virus, positive regulation of gene expression, negative regulation of gene expression, positive regulation of alkaline phosphatase activity, neural crest cell differentiation, cardiac muscle hypertrophy in response to stress, myoblast differentiation, dentate gyrus development, neuron differentiation, platelet formation, monocyte differentiation, negative regulation of ossification, melanocyte differentiation, positive regulation of bone mineralization, positive regulation of B cell proliferation, response to vitamin E, cellular response to drug, skeletal muscle cell differentiation, cellular response to trichostatin A, B cell proliferation, positive regulation of MAP kinase activity, regulation of neuron apoptotic process, negative regulation of neuron apoptotic process, regulation of megarakocyte differentiation, positive regulation of myoblast differentiation, positive regulation of neuron differentiation, positive regulation of osteoblast differentiation, positive regulation of transcription, DNA-templated, positive regulation of transcription from RNA polymerase II promoter, regulation of neurotransmitter secretion, regulation of synaptic plasticity, positive regulation of skeletal muscle tissue development, neuron development, cell morphogenesis involved in neuron differentiation, embryonic viscerocranium morphogenesis, negative regulation of epithelial cell proliferation, B cell receptor signaling pathway, smooth muscle cell differentiation, positive regulation of muscle cell differentiation, regulation of synapse assembly, regulation of synaptic transmission, glutamatergic ventricular cardiac muscle cell differentiation, palate development, regulation of synaptic activity, positive regulation of cardiac muscle cell proliferation, excitatory postsynaptic potential, transdifferentiation, regulation of sarcomere organization, cartilage morphogenesis, regulation of dendrite spine development, renal tubule morphogenesis, cellular response to lipopolysaccharide, cellular response to calcium ion, cellular response to retinoic acid, cellular response to glucose stimulus, cellular response to barbiturate hormone stimulus, cellular response to fluid shear stress, cellular response to transforming growth factor beta stimulus, positive regulation of cell proliferation in bone marrow, osteogenic morphogenesis, nephron tubule epithelial cell differentiation, positive regulation of protein homodimerization activity, positive regulation of macrophage apoptotic process, regulation of N-methyl-D-aspartate selective glutamate receptor activity, regulation of alpha-amino-3-hydroxy-5-methyl-4-isoxazole propionate selective glutamate receptor activity, positive regulation of cardiac muscle cell differentiation, positive regulation of behavioral fear response, epithelial cell proliferation involved in renal tubule morphogenesis, positive regulation of skeletal muscle cell differentiation.		

Figure 5. Example of a keyword search in the DAVID Functional Annotation Table for the word “excitatory” (highlighted in orange).

Table 2. Keywords used in the search, along with the corresponding genes of interest. Some keyword searches, including “electric” and “spasm,” did not yield in any genes that were common amongst the four individual donors.

<u>Glutamate</u>	<u>Spasm</u>	<u>Electric</u>	<u>Tumor</u>	<u>Inflammatory Response</u>	<u>Movement</u>
ADCY2				SERPINF1	NPY
HTR2A				ADCYAP2	
MEF2C					
CAMK2A					
<u>Dendrite</u>	<u>GABA (gamma aminobutyric acid)</u>	<u>Neuron Development</u>	<u>Carbohydrate</u>	<u>Saccharide</u>	<u>Lipid</u>
MEF2C	HTR4	MEF2C		MEF2C	HTR2A
					SERPINF1
<u>Synapse</u>	<u>Excit</u>	<u>Inhibit</u>	<u>Sense</u>	<u>Calcium</u>	<u>Migration</u>
	MEF2C	HTR2A	SLC22A5	HTR2A	PRKCI
			ADCYAP2	ADCYAP2	SERPINF1
				MEF2C	
				NPY	

3.6. Gene Ontology Analysis

Of the 76 genes that were commonly expressed amongst the four donors, 16 were unmapped in the Gene Ontology database (refer to **Figure 6**), one was attributed to multiple PANTHER ID's, and the other 60 did have uniquely mapped ID's in the database. Hence, the 61 genes with ID's in the database were used in the Gene Ontology classification.

Concerning Molecular Function (refer to **Figure 7** for the complete pie charts and relative proportions of genes placed in each category), 21 genes (34.4%) were involved in "binding," while 9 genes (14.8%) were placed in the "catalytic activity" category. Several genes that were in the "binding" category (e.g., *MEF2C* and *HTR2A*), in the "catalytic activity" category (e.g., *ADCY2*), and in both of these categories (e.g., *CAMK2A* and *SERPINF1*), were also found to be associated with TLE through the DAVID keyword search.

The Cellular Component revealed that the "cell part" and "cell" categories each consisted of 27 genes (44.3%), the "organelle" category comprised 16 genes (26.2%), and the "membrane" category consisted of 15 genes (24.6%). The "cell part" and "cell" categories shared all 27 of their genes with each other. Though the "organelle" and "membrane" categories had a majority of their genes in common, the genes of interest found through the DAVID search were localized to either one of the categories. For example, *MEF2C* was only classified in the "organelle" category, and *HTR2A* and *CAMK2A* were only classified in the "membrane" categories.

In Biological Processes, the most common categories were "cellular process" and "biological regulation," with 22 (36.1%) and 19 (31.1%) genes, respectively. Within the "cellular process" category, 10 (45.5%) and 9 (40.9%) genes were associated with the "cellular metabolic process" and "cell communication" categories, respectively. These two findings suggest a possible connection to the abnormal neural excitation that is characteristic of TLE [1] [2] [3].

3.7. STRING Gene-to-Gene and Protein Interactions

Of the 10 unique genes that were commonly expressed amongst the four donors and appeared in the keyword search, *ADCYAP2* did not have data relevant to *Homo sapiens* in the STRING Database. Hence, the other nine genes were used to conduct the protein interaction analysis.

A protein interaction network (refer to **Figure 7**) was generated with the input genes *ADCY2*, *HTR2A*, *MEF2C*, *CAMK2A*, *HTR4*, *SLC22A5*, *SERPINF1*, *NPY*, and *PRKCI*. The network contained 50 and 10 interactors in the first and second shells, respectively, and its clusters centered on the following genes (in decreasing order of cluster size, measured by number of direct interactions): *PRKCI* (21 first shell interactions), *CAMK2A* (15 first shell interactions), *PSMD1* (2 first shell interactions and 8 second shell interactions), *MEF2C* (8 first shell interactions), *NPY* and *ADCY2* (each with 2 first shell interactions), *SERPINF1* and *HTR2A* (each with 1 first shell interaction), and *HTR4* and *SLC22A5* (each with no interactions).

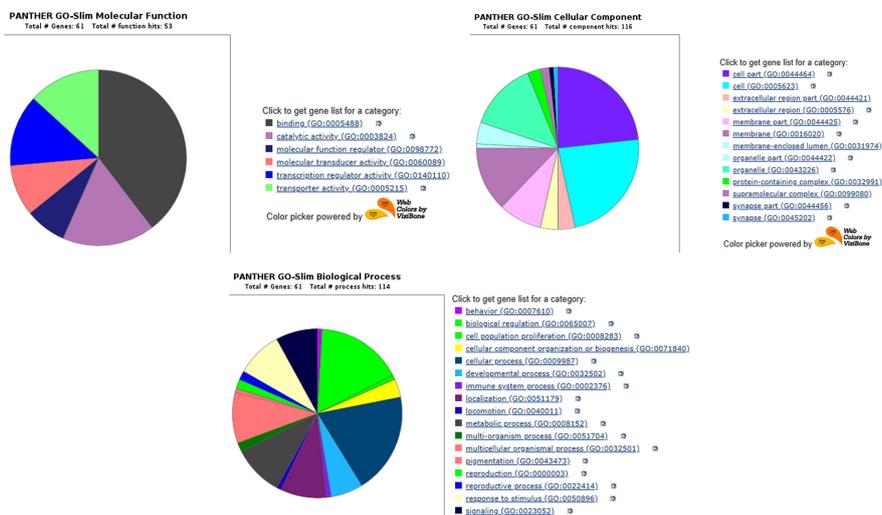


Figure 6. Molecular Function, Cellular Component, and Biological Process pie charts, respectively, generated through The Gene Ontology and the PANTHER classification system.

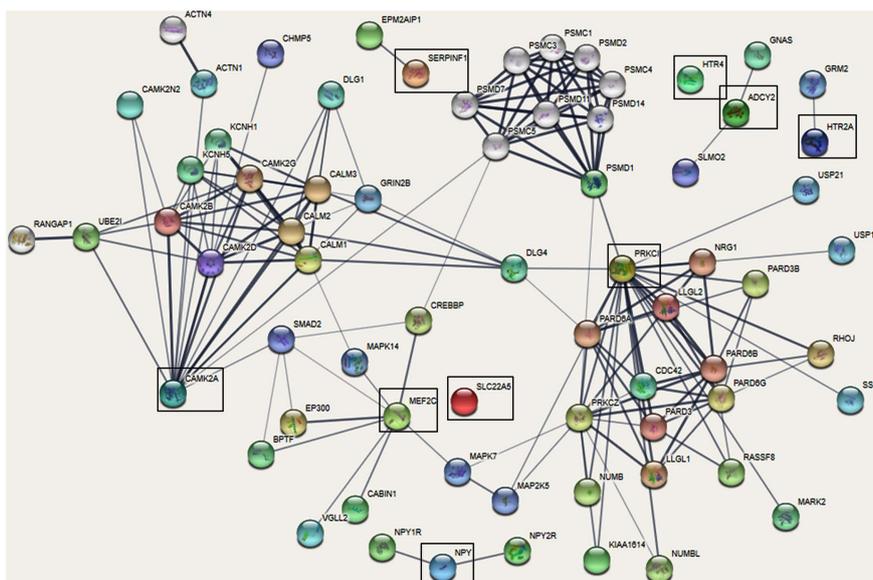


Figure 7. Protein interactions network generated through the STRING Database. The genes found through the keyword search (ADCY2, HTR2A, MEF2C, CAMK2A, HTR4, SLC22A5, SERPINF1, NPY, and PRKCI) were inputted and are boxed in black. The thickness of lines represents the relative strength of experimental support and validation. First shell interactors and second shell interactors are shown in colored nodes and white nodes, respectively. Empty nodes and filled nodes represent that the 3-dimensional structure is unknown and known or predicted, respectively.

Note that *PSMD1* and the input gene *PRKCI* were linked through a direct first shell interaction and that 8 of the 10 total second shell interactions were linked to *PSMD1*. Also, several members of the mitogen-activated protein kinase (MAPK) family, including *MAPK7* and *MAPK14*, play an integral role in linking different clusters to each other. In addition, *PSMC5* is the only one of the second shell interactors that has a direct link to an input gene (*CAMK2A*). *DLG4* and

members of the proteasome regulatory subunit group (e.g., *PSMD1* AND *PSMC5*) seem to link several clusters together as well.

Because the limits of STRING may have constrained the number of network edges shown, another network analysis was conducted with only the original input genes that were part of smaller clusters (*NPY*, *ADCY2*, *SERPINF1*, *HTR2A*, *HTR4*, and *SLC22A5*). This was done to reveal any network edges and interactions that may have not been displayed in **Figure 7**.

The second protein interaction network (refer to **Figure 8**) revealed that *HTR2A* was indirectly associated with *ADCY2* through several members of the guanine nucleotide-binding protein subunit group (e.g., *GNAI1*). However, *NPY*, *SERPINF1*, *HTR4*, and *SLC22A5* did not display any new or different interactions than in **Figure 7**. Given that the cluster centered around *HTR2A* contained members of the MAPK family (*MAPK1* and *MAPK3*), another protein network analysis was conducted to investigate any other possible proteins interactions between *HTR2A*, *ADCY2*, and the original input genes most directly associated with the MAPK family (*PRKCI*, *CAMK2A*, and *MEF2C*), as revealed in **Figure 7**.

Due to the display limits of STRING, two separate networks were generated, with one (refer to **Figure 9(a)**) having *HTR2A*, *ADCY2*, *CAMK2A*, and *MEF2C* as input genes and the other (refer to **Figure 9(b)**) replacing *CAMK2A* with *PRKCI*. **Figure 9(a)** suggests an indirect linkage between *HTR2A* and *CAMK2A* through *GNAO1* and *GNAI1*, members of the guanine nucleotide-binding protein subunit group. There are also indirect linkages between *HTR2A* and *MEF2C* through *MAPK14* (**Figure 9(a)**) and between *HTR2A* and *PRKCI* through *MAP2K1* (**Figure 9(b)**), confirming the earlier prediction that *HTR2A* may have additional interactions through MAPK's.

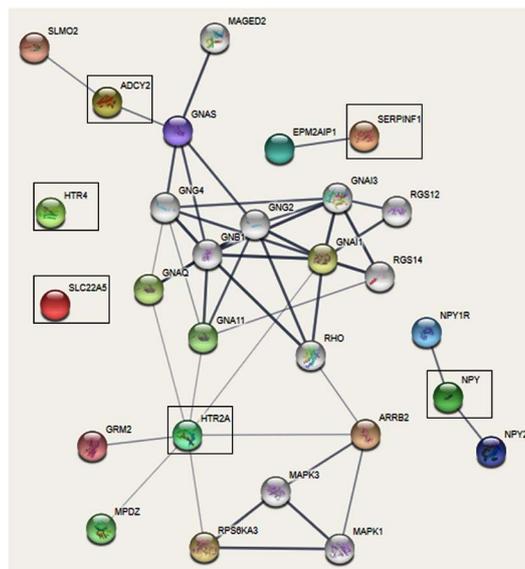
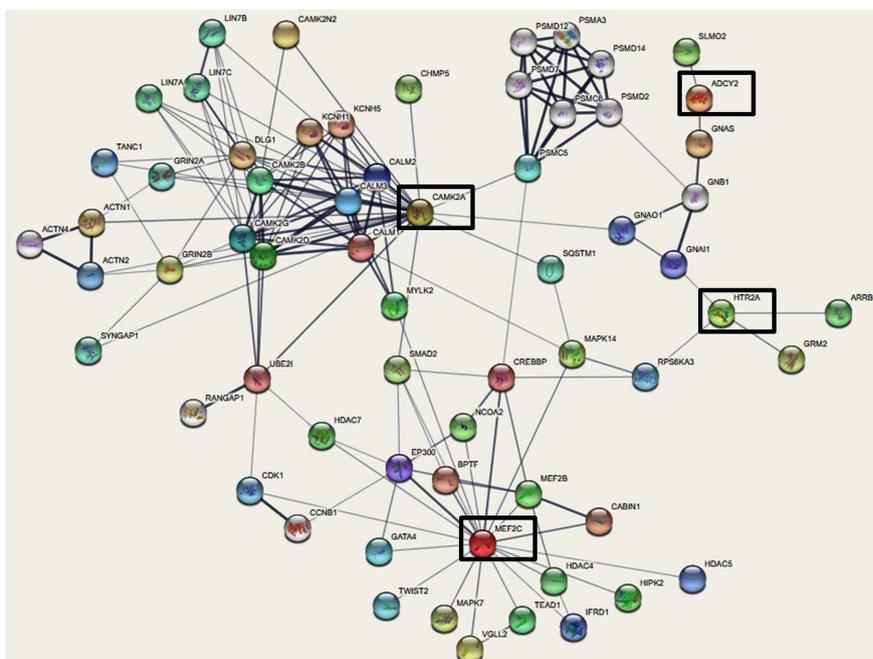
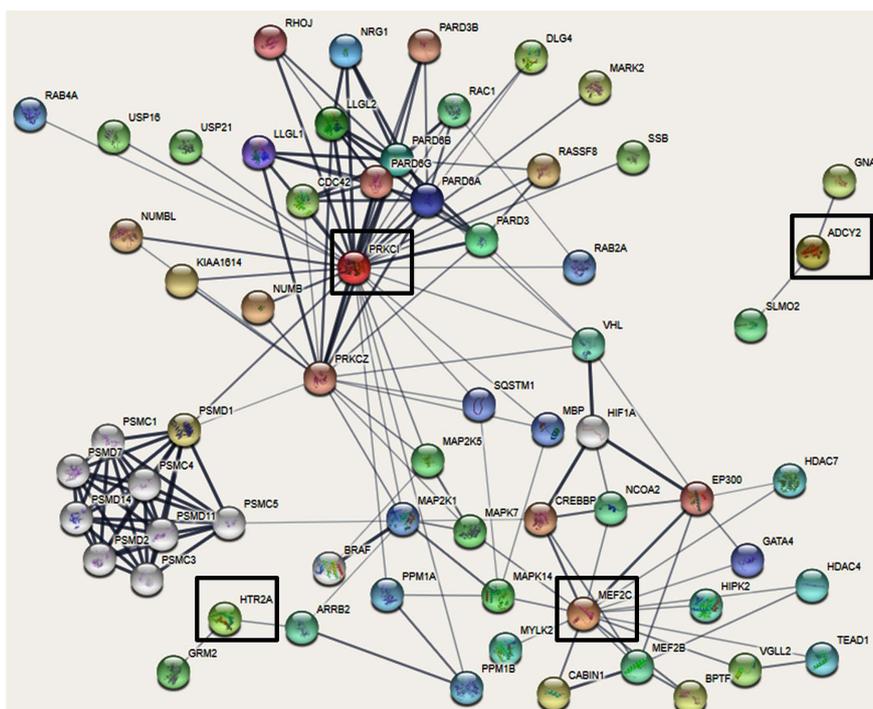


Figure 8. STRING protein interaction network with *NPY*, *ADCY2*, *SERPINF1*, *HTR2A*, *HTR4*, and *SLC22A5*. Refer to **Figure 7** for black boxes, line thickness, protein color, and node filling meanings.



(a)



(b)

Figure 9. (a). STRING protein interaction network with CAMK2A, HTR2A, ADCY2, and MEF2C. Refer to **Figure 7** for black boxes, line thickness, protein color, and node filling meanings. (b). STRING protein interaction network with PRKCI, HTR2A, ADCY2, and MEF2C. Refer to **Figure 7** for black boxes, line thickness, protein color, and node filling meanings.

In summary, the STRING Database protein interaction analysis revealed that potential candidate genes for TLE may be, not only the genes generated through

the DAVID keyword search, but also interconnecting or linking genes (*MAPK7*, *MAPK14*, *MAP2K1*, *PSMD1*, *PSMC5*, *DLG4*, *GNAO1*, and *GNAI1*). This calls for a reanalysis of each gene's specific functions as they relate to TLE in order to identify accurate candidate biomarkers (refer to **Table 3** for a full list of genes and their respective functions as shown in STRING). Only the genes with the most relevant functions will be discussed below (some genes have also been disregarded due to the lack of specificity in their enriched functions).

Table 3. Candidate genes found through the STRING protein interactions analysis listed with their corresponding STRING enriched functions.

Gene	STRING Enriched Functions
ADCY2	Adenylate cyclase type 2; Catalyzes the formation of the signaling molecule cAMP in response to G-protein signaling. Down-stream signaling cascades mediate changes in gene expression patterns and lead to increased IL6 production. Functions in signaling cascades downstream of the muscarinic acetylcholine receptors (By similarity); Belongs to the adenylate cyclase class-4/guanylate cyclase family.
HTR2A	5-hydroxytryptamine receptor 2A; G-protein coupled receptor for 5-hydroxytryptamine (serotonin). Also functions as a receptor for various drugs and psychoactive substances, including mescaline, psilocybin, 1-(2,5-dimethoxy-4-iodophenyl)-2-aminopropane (DOI) and lysergic acid diethylamide (LSD). Ligand binding causes a conformation change that triggers signaling via guanine nucleotide-binding proteins (G proteins) and modulates the activity of down-stream effectors. Beta-arrestin family members inhibit signaling via G proteins and mediate activation of alternative signaling pathways.
MEF2C	Myocyte-specific enhancer factor 2C; Transcription activator which binds specifically to the MEF2 element present in the regulatory regions of many muscle-specific genes. Controls cardiac morphogenesis and myogenesis, and is also involved in vascular development. Plays an essential role in hippocampal-dependent learning and memory by suppressing the number of excitatory synapses and thus regulating basal and evoked synaptic transmission. Crucial for normal neuronal development, distribution, and electrical activity in the neocortex. Necessary for proper development of megakaryocytes.
CAMK2A	Calcium/calmodulin-dependent protein kinase type II subunit alpha; CaM-kinase II (CAMK2) is a prominent kinase in the central nervous system that may function in long-term potentiation and neurotransmitter release. Member of the NMDAR signaling complex in excitatory synapses it may regulate NMDAR-dependent potentiation of the AMPAR and synaptic plasticity (By similarity). Phosphorylates transcription factor FOXO3 on "Ser-298". Activates FOXO3 transcriptional activity (By similarity); Belongs to the protein kinase superfamily. CAMK Ser/Thr protein kinase family. CaMK subfamily.
HTR4	5-hydroxytryptamine receptor 4; This is one of the several different receptors for 5-hydroxytryptamine (serotonin), a biogenic hormone that functions as a neurotransmitter, a hormone, and a mitogen. The activity of this receptor is mediated by G proteins that stimulate adenylate cyclase; 5-hydroxytryptamine receptors, G protein-coupled.
SLC22A5	Solute carrier family 22 member 5; Sodium-ion dependent, high affinity carnitine transporter. Involved in the active cellular uptake of carnitine. Transports one sodium ion with one molecule of carnitine. Also transports organic cations such as tetraethylammonium (TEA) without the involvement of sodium. Also relative uptake activity ratio of carnitine to TEA is 11.3; Solute carriers.
SERPINF1	Pigment epithelium-derived factor; Neurotrophic protein; induces extensive neuronal differentiation in retinoblastoma cells. Potent inhibitor of angiogenesis. As it does not undergo the S (stressed) to R (relaxed) conformational transition characteristic of active serpins, it exhibits no serine protease inhibitory activity; Serpin peptidase inhibitors.
NPY	Pro-neuropeptide Y; NPY is implicated in the control of feeding and in secretion of gonadotrophin-release hormone; Belongs to the NPY family.
PRKCI	Protein kinase C iota type; Calcium- and diacylglycerol-independent serine/ threonine-protein kinase that plays a general protective role against apoptotic stimuli, is involved in NF-kappa-B activation, cell survival, differentiation and polarity, and contributes to the regulation of microtubule dynamics in the early secretory pathway. Is necessary for BCR-ABL oncogene-mediated resistance to apoptotic drug in leukemia cells, protecting leukemia cells against drug-induced apoptosis. In cultured neurons, prevents amyloid beta protein-induced apoptosis by interrupting cell death process.
MAPK7	Mitogen-activated protein kinase 7; Plays a role in various cellular processes such as proliferation, differentiation and cell survival. The upstream activator of MAPK7 is the MAPK kinase MAP2K5. Upon activation, it translocates to the nucleus and phosphorylates various downstream targets including MEF2C. EGF activates MAPK7 through a Ras- independent and MAP2K5-dependent pathway. May have a role in muscle cell differentiation. May be important for endothelial function and maintenance of blood vessel integrity. MAP2K5 and MAPK7 interact specifically with one another and not with MEK1/E.

Continued

MAPK14	Mitogen-activated protein kinase 14; Serine/threonine kinase which acts as an essential component of the MAP kinase signal transduction pathway. MAPK14 is one of the four p38 MAPKs which play an important role in the cascades of cellular responses evoked by extracellular stimuli such as proinflammatory cytokines or physical stress leading to direct activation of transcription factors. Accordingly, p38 MAPKs phosphorylate a broad range of proteins and it has been estimated that they may have approximately 200 to 300 substrates each. Some of the targets are downstream kinases.
MAP2K1	Dual specificity mitogen-activated protein kinase kinase 1; Dual specificity protein kinase which acts as an essential component of the MAP kinase signal transduction pathway. Binding of extracellular ligands such as growth factors, cytokines and hormones to their cell-surface receptors activates RAS and this initiates RAF1 activation. RAF1 then further activates the dual-specificity protein kinases MAP2K1/MEK1 and MAP2K2/MEK2. Both MAP2K1/MEK1 and MAP2K2/MEK2 function specifically in the MAPK/ERK cascade, and catalyze the concomitant phosphorylation of a threonine and a tyrosine resid.
PSMD1	26S proteasome non-ATPase regulatory subunit 1; Component of the 26S proteasome, a multiprotein complex involved in the ATP-dependent degradation of ubiquitinated proteins. This complex plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins, which could impair cellular functions, and by removing proteins whose functions are no longer required. Therefore, the proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage repair; Belongs to the proteasome subunit S1 family.
PSMC5	26S proteasome regulatory subunit 8; Component of the 26S proteasome, a multiprotein complex involved in the ATP-dependent degradation of ubiquitinated proteins. This complex plays a key role in the maintenance of protein homeostasis by removing misfolded or damaged proteins, which could impair cellular functions, and by removing proteins whose functions are no longer required. Therefore, the proteasome participates in numerous cellular processes, including cell cycle progression, apoptosis, or DNA damage repair. PSMC5 belongs to the heterohexameric ring of AAA.
DLG4	Disks large homolog 4; Interacts with the cytoplasmic tail of NMDA receptor subunits and shaker-type potassium channels. Required for synaptic plasticity associated with NMDA receptor signaling. Overexpression or depletion of DLG4 changes the ratio of excitatory to inhibitory synapses in hippocampal neurons. May reduce the amplitude of ASIC3 acid-evoked currents by retaining the channel intracellularly. May regulate the intracellular trafficking of ADR1B (By similarity); Belongs to the MAGUK family.
GNAO1	Guanine nucleotide-binding protein G(o) subunit alpha; Guanine nucleotide-binding proteins (G proteins) are involved as modulators or transducers in various transmembrane signaling systems. The G(o) protein function is not clear. Stimulated by RGS14; Belongs to the G-alpha family. <i>G(i/o/t/z)</i> subfamily.
GNAI1	Guanine nucleotide-binding protein G(i) subunit alpha-1; Guanine nucleotide-binding proteins (G proteins) function as transducers downstream of G protein-coupled receptors (GPCRs) in numerous signaling cascades. The alpha chain contains the guanine nucleotide binding site and alternates between an active, GTP-bound state and an inactive, GDP-bound state. Signaling by an activated GPCR promotes GDP release and GTP binding. The alpha subunit has a low GTPase activity that converts bound GTP to GDP, thereby terminating the signal.

According to the STRING enriched functions, CAMK2A is involved in the NMDAR signaling complex at excitatory synapses and induces increased activation of the FoxO3 transcription factor (promotes apoptosis of damaged neurons) in response to epileptic seizures; this corresponds to previous findings that *CAMK2A* overexpression is correlated with the onset of hippocampal seizures (deep in the temporal lobe), suggesting its candidacy as a TLE biomarker [14] [15]. MEF2C seems to play an integral role in regulating hippocampal excitatory synaptogenesis; furthermore, the fact that, in the donors (who had no history of epilepsy), *MEF2C* is significantly overexpressed (>3.000 fold change) in the temporal lobe compared to in the gray matter underscores MEF2C's important role in preventing abnormal neocortex activity [16]. The identification of *NPY* as a candidate TLE biomarker corresponds to previous work that shows the correlation between *NPY*'s neuroprotective effects and its overexpression in individuals with TLE, possibly as a defense mechanism [17]. Relatively few studies have reported *MAPK7* as a potential TLE biomarker; however, *MAPK7* contributes to the phosphorylation-dependent degradation of MEF2C, suggesting that

overexpression of MAPK7 may be linked with the onset of TLE and that underexpression of MAPK7 may be associated with cellular responses to TLE [18]. *DLG4*'s enriched function of regulating the ratio of hippocampal excitatory and inhibitory synapses suggests a connection to the symptoms of TLE (abnormal excitation); indeed, PSD-95, encoded by *DLG4*, is a scaffolding protein at excitatory synapses of the postsynaptic density, and downregulation of PSD-95 expression has been shown to be implicated with the onset of TLE [19] [20].

4. Conclusions

An individual is diagnosed with TLE, the most common form of focal epilepsy, when more than one temporal lobe seizure occurs and cannot be explained by another nonpermanent condition or disorder [2]. TLE can be localized to either the mesial or lateral temporal lobe; however, mesial TLE is more common (80%) compared to lateral TLE (20%). Although TLE as a whole has been proven to be influenced by genetic factors, it is currently diagnosed with EEG tests, MRI scans, and patient reports of symptoms, which can together take several months to interpret and confirm a diagnosis [3]. During this time, patients may experience additional seizures and have an increased risk for other anxiety or mood disorders, including depression [5].

This study proposes the earlier detection and diagnosis of TLE with quicker gene-based diagnostics, such as microarray and/or genetic testing or genomic heatmap analysis. It was hypothesized that candidate genes as biomarkers for the proposed diagnostic tests could be identified by analyzing gene expression patterns in the temporal lobe. Specifically, genes that were overexpressed in the temporal lobe compared to in the gray matter (≥ 3.000 fold change) were collected for further analysis from four donors through the Allen Human Brain Atlas. The number of genes with fold changes greater than 3.000 for the donors ranged from 145 to 330, and of these genes, 76 (29.3%) were commonly amongst the four individual donors, according to the Venny 2.1 analysis (refer to **Figure 2**).

When bar graphs (refer to **Figure 3**) were generated to depict the top twenty genes with the highest fold changes for each donor, ten genes were found to be common amongst all four bar graphs. Through further statistical analysis (refer to **Table 1**), it was determined that mean fold change values were directly correlated with the five number summary (minimum, quartile 1, quartile 2, quartile 3, and maximum) fold change values; the minimum fold changes, variances, and standard deviations only varied slightly from donor to donor.

The DAVID keyword analysis with the Python script (refer to **Figures 4-6**) revealed that the genes of interest were associated with many epilepsy-related phenomena, including inflammatory responses, excitatory postsynaptic potentials, the neurotransmitters glutamate (excitatory) and GABA (inhibitory), lipids, saccharides, neuron development and migration, and calcium concentrations. In the Gene Ontology analysis, the 76 commonly-expressed genes (from Venny 2.1) were revealed to be predominantly classified under “binding” and

“catalytic activity” (Molecular Function); “cell part,” “cell,” “organelle,” and “membrane” (Cellular Component); and “cellular process,” “biological regulation,” “cellular metabolic process,” and “cell communication” (Biological Processes). “Cellular metabolic process” and “cell communication” were subcategories of the “cellular process” category. Although the categories are relatively broad, these findings suggest a possible connection to the hallmark abnormal neural excitation of TLE.

The experimental-evidence-based STRING protein interactions analysis was first conducted with the 10 genes found through the keyword search (*ADCY2*, *HTR2A*, *MEF2C*, *CAMK2A*, *HTR4*, *SLC22A5*, *SERPINF1*, *NPY*, and *PRKCI*). As shown in **Figure 7**, the clusters were mainly centered around *PRKCI*, *MEF2C*, *PSMD1*, *CAMK2A*, *NPY*, *ADCY2*, *SERPINF1*, and *HTR2A*. However, *DLG4*, as well as members of the MAPK and proteasome regulatory subunit groups, did seem to have an integral role in linking different clusters together. Further investigation into interactions between the original input genes and MAPK’s (refer to **Figure 8** and **Figure 9**) not only revealed that *HTR2A* was indirectly associated with *ADCY2* via *GNAOI* and *GNAII*—members of the guanine nucleotide-binding protein subunit group—but also highlighted the importance of MAPK’s in facilitating interconnectivity between different clusters. With the analysis of each gene’s STRING enriched functions, it was concluded that changes in the expression of *CAMK2A*, *MEF2C*, *NPY*, *MAPK7*, or *DLG4* could potentially be biomarkers for TLE in gene-based diagnostics. These genes tended to have a high fold-change and were common to all four donors.

More specifically, underexpression of *MEF2C* or *DLG4* could be linked with the onset of TLE. On the other hand, changes in the expression of *CAMK2A*, *MAPK7*, or *NPY* could signal either the onset of TLE (overexpression of *MAPK7*, underexpression of *CAMK2A* or *NPY*) or a biological defense response to TLE (overexpression of *CAMK2A* or *NPY*, underexpression of *MAPK7*). Only *CAMK2A*, *NPY*, and *DLG4* have a known correlation with TLE. Though *MEF2C*’s and *MAPK7*’s correlations with TLE have yet to be proven with substantial evidence, they do have functions similar to symptoms of TLE.

Limitations and Future Directions

Limitations in this study include the list of keywords used in the DAVID keyword search. Keywords were selected based on their relevance to TLE, and due to time constraints, the number of keywords that could be used was limited, potentially decreasing the scope of the study. Additionally, the candidate genes identified could possibly overlap with genetic factors of other neurological conditions, potentially making it difficult for genomic heatmap analysis and microarray and/or genetic testing to narrow down a diagnosis to TLE. Increasing the number of donors may have also increased the accuracy of the results. The fact that data was collected from donors without a history of epilepsy is not considered a severe limitation, as the initial hypothesis was that certain genes *normally* overexpressed in the temporal lobe as a whole may be candidate biomark-

ers for TLE. Furthermore, the database of choice (Allen Human Brain Atlas) did not have gene expression profiles specific to donors only with epilepsy.

More research is required to confirm the utility of the candidate genes in genomic heatmap analysis and microarray and/or genetic testing. In addition, integrative genomic analysis studies could also be conducted to identify candidate genes common to TLE and other neurological conditions, as epilepsy is often accompanied with other neurological or psychological symptoms [1]. Another possibility may be to analyze gene expression profiles of donors of different ages, investigating whether different genes contribute to the onset of TLE at different stages in life.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

- [1] Cleveland Clinic (n.d.) Temporal Lobe Epilepsy. <https://my.clevelandclinic.org/health/diseases/17778-temporal-lobe-seizures>
- [2] Beverly, J.M., Blumenrath, S., Chiu, L., Davis, A., Fessenden, M., Galinato, M., Halber, D., Hopkin, K., Kelly, D., Parks, C., Richardson, M., Rojahn, S., Sheikh, K.S., Weintraub, K., Wessel, L., Wnuk, A. and Zyla, G. (2012) Brain Facts: A Primer on the Brain and Nervous System. The Society for Neuroscience, Washington DC, 75.
- [3] Klein, H. and Joshi, C. (2019) Temporal Lobe Epilepsy (TLE). <https://www.epilepsy.com/learn/types-epilepsy-syndromes/temporal-lobe-epilepsy-aka-tle>
- [4] Born, J., Matos, H.C., de Araujo, M.A., Castro, O.W., Duzzioni, M., Peixoto-Santos, J.E., Leite, J.P., Garcia-Cairasco, N., Paçó-Larson, M.L. and Gitaí, D. (2017) Using Postmortem hippocampi Tissue Can Interfere with Differential Gene Expression Analysis of the Epileptogenic Process. *PLoS ONE*, **12**, e0182765. <https://doi.org/10.1371/journal.pone.0182765>
- [5] Vosburgh, S. and Osborne Shafer, P. (2018) Overview of Depression. <https://www.epilepsy.com/living-epilepsy/healthy-living/emotional-health/overview-depression>
- [6] Gambardella, A., Manna, I., Labate, A., Chifari, R., La Russa, A., Serra, P., Cittadella, R., Bonavita, S., Andreoli, V., LePiane, E., Sasanelli, F., Di Costanzo, A., Zappia, M., Tedeschi, G., Aguglia, U., Quattrone, A. (2003) GABA(B) Receptor 1 Polymorphism (G1465A) Is Associated with Temporal Lobe Epilepsy. *Neurology*, **60**, 560-563. <https://doi.org/10.1212/01.WNL.0000046520.79877.D8>
- [7] Vadlamudi, L., Scheffer, I.E. and Berkovic, S.F. (2003) Genetics of Temporal Lobe Epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry*, **74**, 1359-1361. <http://dx.doi.org/10.1136/jnnp.74.10.1359>
- [8] Allen Human Brain Atlas (2010) Allen Institute for Brain Science. <http://human.brain-map.org/>
- [9] Oliveros, J.C. (2007-2015) Venny. An Interactive Tool for Comparing Lists with Venn's Diagrams. <https://bioinfogp.cnb.csic.es/tools/venny/index.html>
- [10] Smithson, R. and Thomas, G. (2012) Python Anywhere.

- <https://www.pythonanywhere.com/>
- [11] Database for Annotation, Visualization and Integrated Discovery (2003-2020) DAVID Bioinformatics Resources 6.8. <https://david.ncifcrf.gov/>
- [12] The Gene Ontology Consortium (2019) The Gene Ontology Resource: 20 Years and Still GOing Strong. *Nucleic Acids Research*, **47**, D330-D338. <https://doi.org/10.1093/nar/gky1055>
- [13] Harvey, V.L. and Dickenson, A.H. (2010) EPSPs and IPSPs. In: Stolerman, I.P., Ed., *Encyclopedia of Psychopharmacology*, Springer, Berlin, Heidelberg.
- [14] Lee, M.C., Ban, S.S., Woo, Y.J. and Kim, S.U. (2001) Calcium/Calmodulin Kinase II Activity of Hippocampus in Kainate-Induced Epilepsy. *Journal of Korean Medical Science*, **16**, 643-648. <https://doi.org/10.3346/jkms.2001.16.5.643>
- [15] Salih, D.A. and Brunet, A. (2008). FoxO Transcription Factors in the Maintenance of Cellular Homeostasis during Aging. *Current Opinion in Cell Biology*, **20**, 126-136. <https://doi.org/10.1016/j.ceb.2008.02.005>
- [16] Huang, Y.Y., Wu, X.L., Guo, J. and Yuan, J.X. (2016) Myocyte-Specific Enhancer Binding Factor 2A Expression is Downregulated during Temporal Lobe Epilepsy. *The International Journal of Neuroscience*, **126**, 786-796. <https://doi.org/10.3109/00207454.2015.1062003>
- [17] Noè, F., Pool, A.-H., Nissinen, J., Gobbi, M., Bland, R., Rizzi, M., *et al.* (2008) Neuropeptide Y Gene Therapy Decreases Chronic Spontaneous Seizures in a Rat Model of Temporal Lobe Epilepsy. *Brain*, **131**, 1506-1515. <https://doi.org/10.1093/brain/awn079>
- [18] Badodi, S., Baruffaldi, F., Ganassi, M., Battini, R. and Molinari, S. (2015) Phosphorylation-Dependent Degradation of MEF2C Contributes to Regulate G2/M Transition. *Cell Cycle*, **14**, 1517-1528. <https://doi.org/10.1080/15384101.2015.1026519>
- [19] Coley, A.A. and Gao, W.J. (2018) PSD95: A Synaptic Protein Implicated in Schizophrenia or Autism? *Progress in Neuro-Psychopharmacology & Biological Psychiatry*, **82**, 187-194. <https://doi.org/10.1016/j.pnpbp.2017.11.016>
- [20] Sun, Q.J., Duan, R.S., Wang, A.H., Shang, W., Zhang, T., Zhang, X.Q. and Chi, Z.F. (2009) Alterations of NR2B and PSD-95 Expression in Hippocampus of Kainic Acid-Exposed Rats with Behavioural Deficits. *Behavioural Brain Research*, **201**, 292-299. <https://doi.org/10.1016/j.bbr.2009.02.027>