

**Characterisation of Novel Transcripts in the 3' Region of  
L-type Calcium Channel Genes in Human Brain**



**Hami Lee**

*Lincoln College, University of Oxford*

**Supervisor:** Dr. Elizabeth M. Tunbridge, Professor Paul Harrison

*Department of Psychiatry*

A thesis submitted for the degree of

*Doctor of Philosophy*

Hilary 2023

## Abstract

The  $\alpha 1$  subunit (*CACNA1* family) of L-type voltage-gated calcium channels (LTCCs) play significant roles in brain function and neuropsychiatric disorders. Transcription of LTCCs is complex, with each  $\alpha 1$  gene producing multiple isoforms. *CACNA1S* and *CACNA1F* were reported to have short-length transcripts from their 3' region in the human brain which were not well characterized. Moreover, whether other *CACNA1* genes also produce the 3' transcripts have not been studied.

This thesis aims to identify and understand the short-length transcripts arising from the 3' region of LTCC genes in human brain, especially *CACNA1S* and *CACNA1F*, using a series of technical approaches and analyses. Firstly, using PCR amplification and long-read nanopore sequencing, multiple 3' transcripts with exon-skipping possibilities were identified. Secondly, 5' Rapid Amplification of cDNA Ends (5' RACE) identified the 3' transcripts as being complete transcripts, arising from alternative transcription start sites (TSS). These findings were corroborated by exon expression level findings from Genotype-Tissue Expression (GTEx) data. Analyses using PhyloP showed that the 5'-untranslated region (5'-UTR) of the 3' transcripts are more conserved than the intronic regions. Finally, the translation and localization of the protein encoded by each 3' transcript was investigated in transfected HEK-293T and SH-SY5Y cell lines. All proteins were detectable by western blot. C-Cav1.3 (encoded by *CACNA1D*) localized to the nucleus in both cell types; C-Cav1.2 (encoded by *CACNA1C*) translocated to the nucleus in the excitable (SH-SY5Y) cells but not the non-excitable (HEK-293T) cells. C-Cav1.1 and C-Cav1.4 (encoded by *CACNA1S* and *CACNA1F* respectively) were not localized to the nucleus. This thesis shows that LTCC  $\alpha 1$  subunit genes produce 3' transcripts using alternative TSS in human brain. Their potential functionality is supported by the evolutionary conservation of their 5'-UTRs and by their

translation *in vitro*. However, further studies are required to identify the significance of these transcripts for brain function.

Key words: L-type calcium channel genes, Alternative TSS, 5'RACE, Long-read nanopore sequencing

## **Declaration**

This thesis is the product of my own work and includes nothing that is an outcome of collaborative research except when specified in the text. I state that no substantial part of my dissertation has already been submitted, or is being concurrently submitted for any such degree, diploma or other qualification at the University of Oxford or any other University or similar institution. In accordance with guidelines, this thesis does not exceed 50,000 words (40,083 words) and it contains less than 150 figures (55 figures).

DPhil Candidate in Psychiatry

17 April 2023

## ACKNOWLEDGEMENTS

Many nouns describe who I am. I am a person, a daughter, a friend, a citizen, a national, an international, an East Asian, a Korean, and a woman, but most of all, I am myself. During my DPhil years, I faced many incidents where my identity was not accepted or attacked for example, no matter how many times I corrected, my name was written or pronounced wrongly by everyone. However, due to the many people around me, I could finish my DPhil thesis by securing my own identity. I want to acknowledge the people who helped me to keep my identity as myself, Hami Lee, throughout my DPhil journey at the University of Oxford.

First of all, many people in academia helped me to finish my thesis. I want to thank my lab. My supervisors, Paul Harrison, and Elizabeth Tunbridge, Postdocs in my lab, Arne Mould, and Nicola Hall, and the collaborators – Wilfried Haerty at the Earlham Institute, Edward Drydale, Edward Drydale, and James Bancroft at the Wellcome Center for Human Genetics. Thank you for all your intellectual help to finish my experiments.

Some people provided unconditional support mentally and emotionally throughout the journey. If I were not a daughter of my parents nor a granddaughter of my grandparents, I would not be able to finish my journey. They supported me physically and spiritually. I think I am so lucky to have my family who gave me unconditional and unlimited support for me that I could not receive elsewhere. They always accepted me with warmth no matter who I am, what I do, and why I go back home. Finally, I want to devote my thesis to my cousin, and my grandpa who left the world too early. My cousin taught me how to laugh and how to stay positive. My grandpa taught me that I am smart enough to do anything. Nothing stops me.

I would love to acknowledge the support and love from my friends. There is a person who helped me a lot physically, mentally, emotionally, and financially. His name is Munim Husain not Syed Husain. He was the only one in Oxford who provided me with unlimited support, warmth, and love. We shared so many precious memories academically, and non-academically. He helped me to keep my identity as myself. This thesis is based on all the love I received from him. Kalina Naidoo also supported me without judging me. She listened to me very well so whenever I had serious issues or experienced racism, I could talk to her freely. Being there as her helped me a lot. Finally, Li Chen is my mother in Oxford. She was always there so that I can ask for some help. She treated me like her daughter by sharing precious wisdom based on her experiences. I won't forget all the food she cooked for me. She always cared about me and asked me whether I need anything. Without her, I would not be able to survive in Oxford and successfully work in the lab.

Moreover, I want to thank friends from OKAS, or my housemates who lived with me for many years on Cowley Road. We shared so many memories with warmth. They taught me how to be myself and helped me to keep my identity.

## Table of Contents

Abstract .....	i
Declaration .....	iii
ACKNOWLEDGEMENTS .....	iv
List of Figures .....	xii
List of tables .....	xiv
Abbreviations .....	xv
Introduction.....	1
1.1 Voltage-gated calcium channels (VGCCs) .....	1
1.1.1 Structure of VGCCs .....	1
1.1.2 Electrophysiologically-defined subcategories of VGCCs .....	2
1.1.3. Biophysiological properties of VGCCs .....	3
1.1.4. LTCCs in excitable tissues .....	4
1.2. Producing diverse mRNA transcripts.....	7
1.2.1. Using alternative transcription start site to produce multiple mRNA transcripts .	7
1.2.2. Alternative splicing to generate diverse mRNA transcripts .....	10
1.3. Producing diverse LTCCs at mRNA level .....	12
1.3.1. Usage of alternative TSSs to generate diverse LTCCs .....	12
1.3.2. Alternative splicing mechanisms to generate diverse LTCCs.....	15
1.4. Producing truncated LTCCs due to proteolytic cleavage.....	19
1.5. Observations of expression of the 3' region of <i>CACNAIS</i> and <i>CACNAIF</i> in the human brain .....	20
1.6. Aim of the thesis .....	27
1.7. Summary of the thesis.....	27
Chapter 2: Characterization of the short-length transcripts from the 3' region of <i>CACNAIS</i> and <i>CACNAIF</i> in the human brain .....	30
2.1. Introduction .....	30
2.1.1. Evidence for the expression of the short-length transcripts from the 3' region of <i>CACNAIS</i> and <i>CACNAIF</i> in the human brain .....	30
2.1.2. Long-read nanopore sequencing.....	31
2.1.2.1. What is long-read nanopore sequencing?.....	31
2.1.2.2. Advantages and disadvantages of long-read nanopore sequencing .....	32
2.2. The aim of the chapter .....	33
2.3. Methods .....	34
2.3.1. Human Brain Samples .....	34

2.3.2.	RNA extraction from the postmortem human brain regions .....	34
2.3.3.	Assessment of RNA quality and concentration .....	35
2.3.4.	Reverse Transcription of RNA to produce cDNA .....	37
2.3.5.	Designing primers.....	38
2.3.6.	PCR amplification of each target transcript from human brain cDNA .....	39
2.3.7.	Gel electrophoresis .....	40
2.3.8.	Gel extraction .....	40
2.3.9.	Barcoding PCR to add the barcode sequences .....	40
2.3.10.	PCR purification.....	41
2.3.11.	Measuring the concentration and quality of purified, barcoded PCR products .....	41
2.3.12.	Sequencing library preparation .....	42
2.3.13.	Analyzing long-read nanopore sequencing data using TAQLoRe program ..	44
2.3.14.	Examining the expression variations of the short-length transcripts from the 3' regions of <i>CACNA1S</i> and <i>CACNA1F</i> between brain regions and individuals .....	46
2.4.	Results .....	46
2.4.1.	The short-length transcript from the 3' region of <i>CACNA1S</i> was detected in cDNA from human brain, whereas the full-length <i>CACNA1S</i> transcript was not detected .....	46
2.4.2.	Long-read nanopore sequencing of the short-length transcript from the 3' region of <i>CACNA1S</i> in the human brain .....	48
2.4.2.1.	Long-read nanopore sequencing .....	48
2.4.2.2.	Highly expressed in-frame coding short-length transcripts from the 3' region of <i>CACNA1S</i> in the human brain were identified .....	51
2.4.2.3.	Examining regional and individual variation in the transcript profiles of short-length transcript from the 3' region of <i>CACNA1S</i> .....	57
2.4.3.	The short-length transcript from the 3' region of <i>CACNA1F</i> was detected in cDNA from human brain, whereas the full-length <i>CACNA1F</i> transcript was not detected .....	58
2.4.4.	Long-read nanopore sequencing of short-length transcript in the 3' region of <i>CACNA1F</i> in the human brain .....	59
2.4.4.1.	Long-read nanopore sequencing .....	59
2.4.4.2.	Highly expressed in-frame coding short-length transcripts from the 3' region of <i>CACNA1F</i> in the human brain were identified.....	62
2.4.4.3.	Examining regional and individual variations in the transcript profiles of the short-length transcript from the 3' region of <i>CACNA1F</i> .....	68
2.5.	Discussion.....	68



2.6.	Methodological limitations.....	72
Chapter 3 – Identification of the transcription start site of the 3’ transcripts of LTCC genes ..		74
3.1.	Introduction .....	74
3.1.1.	Rationale of the chapter.....	74
3.1.2.	Transcription start sites.....	74
3.1.3.	Cap Analysis of Gene Expression .....	75
3.1.4.	Previous studies examining the alternative transcription start site of VGCC genes .....	76
3.1.5.	The aim of this chapter .....	77
3.2.	Methods .....	78
3.2.1.	Failed attempts at 5’ RACE.....	78
3.2.2.	Successful attempts of 5’ RACE .....	79
3.2.2.1.	Using poly A <sup>+</sup> RNAs.....	80
3.2.2.2.	Reverse transcription.....	82
3.2.2.3.	Poly(A) tailing.....	83
3.2.2.4.	Outer PCR amplification.....	83
3.2.2.5.	Inner PCR amplification.....	84
3.2.2.6.	Gel electrophoresis.....	85
3.2.2.7.	Gel extraction of potential RACE products .....	86
3.2.2.8.	TA TOPO cloning with One shot TOP10 chemically competent cells .....	86
3.2.2.9.	Plasmid purification .....	87
3.2.2.10.	Restriction Enzyme Digestion.....	88
3.2.2.11.	Multiple Sequence Alignment.....	89
3.2.3.	CAGE data analysis using CAGEr.....	89
3.3.	Results of 5’ RACE experiment and CAGE analysis.....	91
3.3.1.	Result of 5’ RACE experiments .....	92
3.3.1.1.	The possible TSSs of the short-length transcripts from the 3’ region of LTCC genes – positive finding for the short-length transcripts from the 3’ region of <i>CACNAIS</i> .....	92
3.3.1.2.	The possible TSSs of the short-length transcripts from the 3’ region of LTCC genes – positive findings for the short-length transcripts from the 3’ region of <i>CACNAIC</i> , <i>CACNAID</i> , and <i>CACNAIF</i> .....	94
3.3.2.	5’ RACE and CAGE analysis of <i>CACNAIS</i> 3’ transcript.....	97
3.3.3.	5’RACE and CAGE analysis of <i>CACNAIF</i> 3’ transcript.....	101
3.3.4.	5’RACE and CAGE analysis of <i>CACNAIC</i> 3’ transcript.....	104
3.3.5.	5’RACE and CAGE analysis of <i>CACNAID</i> 3’ transcript .....	107

3.3.6.	Predicted similarities and differences across 3' transcripts of LTCC genes ...	110
3.4.	The existence of alternative TSS in the C-terminal region of VGCC $\alpha$ 1 subunits other than LTCCs.....	111
3.5.	Discussion.....	116
3.6.	Limitations of the experiments .....	118
Chapter 4: Investigating possible functionality of 5'UTR and expression of the 3' transcripts .....		120
4.1.	Rationale for studying conservation status and tissue specific expression ...	120
4.2.	Introduction .....	120
4.3.	Aim of chapter 4 .....	122
4.4.	Methods .....	123
4.4.1.	Genomic location data .....	123
4.4.4.	Expression level of exons in LTCC genes from GTEx data .....	126
4.5.	Results .....	128
4.5.1.	Conservation status of the 3' transcript of LTCC genes using PhyloP score .....	128
4.5.2.	Comparing the conservation status of the 5'UTRs with the exonic and intronic regions of the 3' transcripts .....	134
4.5.3.	Expression levels of exons before and after initiation of LTCC 3' transcript.....	137
4.5.3.1.	Expression level analyses after the transcription of 3' transcript of CACNA1S is initiated.....	137
4.5.3.2.	Expression level analyses after transcription of CACNA1F 3' transcript is initiated .....	140
4.5.3.3.	Expression level analyses after transcription of CACNA1C 3' transcript is initiated .....	143
4.5.3.4.	Expression level analysis after transcription of CACNA1D 3' transcript is initiated .....	146
4.6.	Discussion.....	149
4.6.1.	Findings from the conservation study .....	149
4.6.2.	Findings from the expression study.....	150
4.6.3.	Limitations and further studies .....	152
Chapter 5 – Localization of proteins encoded by 3' transcripts of LTCC genes .....		155
5.1.	Introduction .....	155
5.1.1.	Rationale for studying the localization of proteins predicted to be encoded by 3' LTCC transcripts .....	155

5.1.2.	What is the NLS?.....	155
5.1.3.	Aim of the chapter .....	156
5.1.4.	Overview of the experiments.....	156
5.2.	Methods .....	157
5.2.1.	Prediction of NLS in the C-LTCC .....	157
5.2.2.	C-flag pcDNA3 expression vector .....	157
5.2.3.	Primer design.....	159
5.2.4.	Plasmid design for 3' CACNA1F transcript.....	160
5.2.5.	Subcloning of vectors into DH5 $\alpha$ competent cells .....	161
5.2.6.	Plasmid purification.....	162
5.2.7.	Restriction Enzyme Digestion .....	164
5.2.8.	Dephosphorylation of digested expression vector.....	165
5.2.9.	Ligation of inserts into c-flag pcDNA3 expression vector.....	166
5.2.10.	Optimizing ligation of the 3' transcript of <i>CACNAIS</i> , <i>CACNAIC</i> , and <i>CACNAID</i> .....	168
5.2.11.	Plasmid precipitation.....	170
5.2.12.	Seeding the HEK293T cell line and the SH-SY5Y cell line.....	170
5.2.13.	Transfection of HEK293T cell line and SH-SY5Y cell line .....	171
5.2.14.	Harvesting the HEK293T cell line and the SH-SY5Y cell line .....	172
5.2.15.	Measuring protein concentrations .....	172
5.2.16.	Western blotting .....	173
5.2.17.	Immunofluorescent staining and imaging.....	174
5.2.18.	Co-Immunoprecipitation of the C-LTCC in HEK293T cells.....	175
5.2.19.	Coomassie blue Stain .....	176
5.3.	Results .....	177
5.3.1.	The proteins encoded by the 3' LTCC transcripts are predicted to localize to the nucleus .....	177
5.3.2.	Confirmation of C-LTCC expression in cell culture .....	179
5.3.2.1.	Protein lysates of transfected C-LTCC in HEK293T cells were separated and identified via western blotting .....	179
5.3.2.2.	Protein lysates of transfected C-LTCC in SH-SY5Y cells were separated and identified via western blotting .....	185
5.3.3.	Localization of C-LTCC .....	186
5.3.3.1.	Localization of C-LTCCs in HEK293T cells .....	186
5.3.3.2.	Localization of C-LTCCs in SH-SY5Y cells .....	189

5.3.4. Co-Immunoprecipitation of proteins encoded by the 3' transcripts of LTCC genes .....	192
5.3.4.1. Co-Immunoprecipitation of C-LTCCs in HEK293T cells .....	192
5.4. Discussion.....	196
Chapter 6 - General Discussion .....	200
6.1. Rationale of studying the 3' transcripts of LTCC genes .....	200
6.2. The findings regarding the 3' transcripts of LTCC genes .....	201
6.2.1. The 3' transcripts of <i>CACNA1S</i> and <i>CACNA1F</i> were found in the human brain .....	201
6.2.2. Could the 5' UTR contain functional motifs? .....	203
6.2.3. The expression of the 3' transcript of LTCC genes in human tissues .....	203
6.2.4. Where do the proteins encoded by the 3' transcripts of LTCC genes translocate in the human cell? .....	205
6.2.4.1. Possible functions .....	206
6.3. Future studies regarding to the findings in the thesis .....	208
6.3.1. Will the exon-skipping events in the 3' transcript of <i>CACNA1S</i> and <i>CACNA1F</i> have an influence on their function of translated protein in the human brain?.....	208
6.3.2. What are the non-coding short-length transcripts identified from the long-read nanopore sequencing?.....	211
6.3.3. The transcription mechanism of the LTCC 3' transcripts .....	213
6.3.4. Further research of examining transcription termination sites of 3' transcripts of LTCC genes via 3' RACE.....	215
6.4. Limitation in the experiment design .....	216
6.4.1. Experimental design could have been planned in a different way.....	216
6.4.2. Accessibility to other tissue types than the post-mortem human brain for positive control .....	217
6.4.3. The identification of the 3' transcripts of <i>CACNA1S</i> and <i>CACNA1F</i> could be validated in the human brain via Northern blot analysis. ....	218
6.5. Concluding remarks .....	219
REFERENCES .....	222
Appendix.....	256

## List of Figures

Figure 1.1. Schematic of the VGCC $\alpha 1$ subunit comprised of four different domains, each with six segments and a P loop (marked with orange arrow) .....	2
Figure 1.2. Schematic cartoon of alternative splicing patterns .....	12
Figure 1.3. Schematic figure of alternative splicing and alternative TSS location of LTCCs .....	15
Figure 1.4 Schematic figure of alternative splicing in the C-terminal region of CACNA1D .....	17
Figure 1.5. Schematic figure of alternative splicing in the C-terminal region of CACNA1F .....	19
Figure 1.6 LIBD Human DLPFC Development track data hub in UCSC genome browser (hg19) showing the result of differential expression analysis .....	24
Figure 1.7 Expression analysis from the GTEx portal .....	26
Figure 2.1 Overview of the PCR-targeted long-read nanopore sequencing method .....	33
Figure 2.2. An example of the integrity of RNA extracted from each sample .....	36
Figure 2.3. Gel electrophoresis results of PCR amplification of full-length CACNA1S and the predicted short-length transcript from the 3' region of CACNA1S .....	47
Figure 2.4. Read length histogram after sequencing .....	48
Figure 2.5. A pie chart showing the proportion of expression of short-length transcript from the 3' region of CACNA1S in the human brain .....	50
Figure 2.6. Proportion of high abundant (>1% of total expression of coding transcripts) coding short-length transcript from the 3' region of CACNA1S in each human brain region .....	51
Figure 2.7. The top 2 most highly expressed in-frame short-length coding transcripts from the 3' region of CACNA1S in the human brain .....	54
Figure 2.8. Principal Component Analysis (PCA) based on transcript expression .....	57
Figure 2.9. Gel electrophoresis result of PCR amplification of both full-length CACNA1F and predicted short-length transcript from the 3' region of CACNA1F from the postmortem human brain .....	59
Figure 2.10. Read length histogram after sequencing short-length transcript from the 3' region of CACNA1F .....	60
Figure 2.11. A pie chart showing the proportion of expression of short-length transcripts from the 3' region of CACNA1F in the human brain .....	61
Figure 2.12. Proportion of highly abundant (>1% of total expression level of coding transcripts) coding short-length transcripts from the 3' region of CACNA1F reads in each human brain region .....	62
Figure 2.13. The top 4 most highly expressed in-frame short-length coding transcripts from the 3' region of CACNA1F in the human brain .....	66
Figure 2.14. Principal Component Analysis (PCA) based on transcripts expression .....	68
Figure 3.1. Schematic representation of CAGE. This is retrieved from (Shiraki et al., 2003) .....	76
Figure 3.2. Overview protocol of 5'/3' RACE Kit, 2nd Generation. The 5'/3' RACE Kit uses a gene specific primer for RT and nested PCR .....	79
Figure 3.3. Flow chart of experiments done for this chapter to identify TSS of the short-length transcript of LTCC genes .....	80
Figure 3.4. Vector map of TOPO™ TA Cloning™ Kit with One Shot™ TOP 10 Chemically competent E. coli (Thermofisher, UK) indicating where the PCR products are inserted (Thermofisher.com) .....	87
Figure 3.5. The 5' RACE products of short-length transcript from the 3' region of LTCC genes .....	93
Figure 3.6. The 5' RACE products of short-length transcript from the 3' region of CACNA1C, CACNA1D, and CACNA1F transcript from the 3' region of CACNA1D. The lane names are described in Table 3.9. ....	96
Figure 3.7. TSS of the 3' transcript of CACNA1S with CAGE analysis and its location in the channel structure .....	100

Figure 3.8. TSS of the 3' transcript of CACNA1F with CAGE analysis and its location in the channel structure .....	103
Figure 3.9. TSS of the 3' transcript of CACNA1C with CAGE analysis and its location in the channel structure .....	106
Figure 3.10. TSS of the 3' transcript of CACNA1D with CAGE analysis and its location in the channel structure .....	109
Figure 3.11 Multiple Sequence Alignment analysis of 3' transcripts of LTCC genes .....	111
Figure 3.12 UCSC genome browser of CAGE analysis was aligned to the C-terminal region of each VGCC $\alpha$ 1 subunits genes .....	115
Figure 4.1. This schematic illustrates the process of the randomized resampling. ....	125
Figure 4.2. The PhyloP score around the TSS of 3' transcripts of LTCC genes .....	133
Figure 4.3 . Comparison of median PhyloP scores among the 5'UTR, exonic regions and intronic regions of the 3' transcripts of LTCC genes.....	136
Figure 4.4. Changes in expression level among exon 23, exon 24 and exon 25 of CACNA1S .....	140
Figure 4.5. Changes in expression level among exon 40, 41 and 42 of the 3' transcript of CACNA1F. .	143
Figure 4.6. Changes in expression level among exon 46, 47 and 48 of the 3' transcript of CACNA1C..	146
Figure 4.7. Expression levels among exon 46, 47 and 48 of the 3' transcript of CACNA1D .....	149
Figure 5.1 A vector map of c-Flag pcDNA3 from addgene (Addgene plasmid # 20011; <a href="http://n2t.net/addgene:20011">http://n2t.net/addgene:20011</a> ; RRID: Addgene_20011) .....	158
Figure 5.2. Explanation of Nested PCR experiments .....	160
Figure 5.3. Synthesized plasmid indicating locations of Buffer sequence, HindIII, coding sequence, XbaI and buffer sequences .....	161
Figure 5.4. Graphic description of restriction enzyme digestion which creates sticky ends .....	165
Figure 5.5. The ligation process of the 3' transcript of LTCC genes to the c-flag pcDNA3 .....	167
Figure 5.6. Flow chart of optimization for dephosphorylation and ligation step.....	168
Figure 5.7. Western blot result of protein lysate from the HEK293T cells transfected with the 3' transcript of LTCC genes.....	181
Figure 5.8. All the possible C-LTCC with different translation start sites .....	184
Figure 5.9. Western blot of the 3' LTCC in the SH-SY5Y cells .....	185
Figure 5.10. Immunofluorescence confocal imaging of transfected HEK293T cells which were transfected with the 3' LTCC .....	189
Figure 5.11. Immunofluorescence imaging of transfected SHSY5Y cells .....	191
Figure 5.12. A graph comparing number of flag tags in nuclei .....	191
Figure 5.13 Coomassie blue staining of co-immunoprecipitation with C-LTCC in HEK293T cells .....	194
Figure 5.14. Western blotting of immunoprecipitation of the C-LTCC in HEK293T cells .....	195

## List of tables

Table 1.1 Categorization of VGCC $\alpha 1$ subunits according to their physiological properties and voltage activation threshold	6
Table 2.2. Table of the reverse transcriptase (RT) mix to obtain cDNA from RNA sample extracted from the postmortem human brain	37
Table 2.3. The primer sequences used for PCR amplification of the target transcripts. The barcode adaptor sequences are in lowercase and the gene-specific sequences are in uppercase.	38
Table 2.4. Table of PCR mix to amplify gene of interest from the human brain	39
Table 2.5. Table of PCR amplification thermal cycle, annealing temperature and extension time	39
Table 2.6. The barcoding PCR reagent mix to attach a barcode. The barcode is unique to each sample, so they can be pooled for a single sequencing run.	41
Table 2.7 Genomic location of each exonic region (hg19) from the long-read sequencing of short-length transcript in the 3' region of CACNA1S from the human brain.	55
Table 2.8. Protein predicted to be encoded by each transcript	56
Table 2.9. Genomic location of each exonic region (hg19) from the long-read sequencing of short-length transcript in the 3' region of CACNA1F from the human brain.	64
Table 2.10. Protein predicted to be encoded by each transcript	67
Table 3.1. The poly A <sup>+</sup> RNA samples were used for the 5' RACE experiment. All poly A <sup>+</sup> RNA samples were pooled together.	81
Table 3.2. List of primers used for 5' RACE experiment	81
Table 3.3. Mixture of reagent for RT for 5'/3' RACE Kit, 2nd Generation	83
Table 3.4. Mixture of reagents for Poly(A) tailing for 5'/3' RACE Kit, 2nd Generation	83
Table 3.5. Mixture of reagents for Outer PCR amplification	84
Table 3.6. Mixture of reagents for Inner PCR amplification.	85
Table 3.7 Mixture of reagents for TOPO™ TA Cloning™ Kitwith One Shot™ TOP 10 Chemically competent E. coli (Thermofisher, UK).	87
Table 3.8 Mixture of reagents for Restriction Enzyme digestion with EcoRI (Promega, UK)	89
Table 3.9. List of CAGE data downloaded from the FANTOM5 project for the analysis.	90
Table 3.10. Name of each lane for figure 3.4	93
Table 3.11. Name of each lane for Figure 5	97
Table 4.22 Tissues in GTEx data	127
Table 4.23 Exon number where TSS is located	128
Table 5.24. Primer sequences used for the first round of PCR	160
Table 5.25. Primer sequences used for the second round of PCR	160
Table 5.26. Sequential restriction enzyme digestion methods	164
Table 5.27. Dephosphorylation methods using CIAP	166
Table 5.28. Ligation method of expression vector to determine if restriction enzyme digestion and dephosphorylation was successful	169
Table 5.29. Table of nuclear localization signal prediction.	177
Table 3.30. Total RNA samples used for the 5' RACE experiment. All total RNA samples were pooled together. Each number corresponds to a different individual (See Methods section 3.1 in chapter 2).	276

## Abbreviations

VGCCs: Voltage gated calcium channels  
Ca<sup>2+</sup>: calcium  
P loop: pore loop  
CaM: Calmodulin  
PCRD: Proximal C-terminal regulatory domain  
DCRD: Distal C-terminal regulatory domain  
CTM: C-terminal Modulator  
HVA: High voltage-activated  
LVA: Low-voltage-activated  
LTCCs: L-type Calcium channels  
CREB: cAMP response element-binding protein  
CaMK: Ca<sup>2+</sup>/Calmodulin-dependent protein kinase  
MAPK: Mitogen-activated protein kinase  
DREAM: Downstream regulatory element antagonist modulator  
DRE: Downstream regulatory element  
TSS: transcription start site  
5'UTRs: 5' untranslated regions  
CAGE: cap analysis of gene expression  
BDNF: Brain-derived neurotrophic factor  
CDS: coding sequence  
NRG1: neuregulin 1 gene  
NDF: neu differentiation factor  
GGF: glial growth factor  
snRNP: small nuclear ribonucleoproteins  
CCAT: Calcium Channel Associated Transcript  
NMDA: N-methyl-D-aspartate  
CDI: Calcium-dependent inactivation  
SAN: Sinoatrial node  
LIBD: Lieber Institute of Brain Development  
5' RACE: 5' Rapid Amplification of cDNA ends  
GTEX: Genotype-Tissue Expression  
DL-PFC: Dorsolateral prefrontal cortex  
CBLM: Cerebellum  
RIN: RNA Integrity Number  
RT: Reverse Transcription  
ApE: A Plasmid Editor  
ONT: Oxford Nanopore Technologies  
TMM: Trimmed mean of M-values normalization method  
PCA: Principal Component Analysis  
CIP: Calf Intestine Alkaline Phosphatase  
TAP: Tobacco Acid Pyrophosphatase  
PB: Phosphate Buffer



mRNA: messenger RNA  
3' UTR: 3' untranslated sequence  
NLSs: Nuclear localization signals  
NPC: Nuclear pore complex  
CIAP: Alkaline Phosphatase, Calf Intestinal  
DMEM: Dulbecco's Modified Eagle's Media  
FBS: Fetal Bovine Serum  
DPBS: Dulbecco Phosphate Buffered Saline  
CTF: C-terminal fragment  
ICD: Intracellular domain  
ChIP-seq: chromatin immunoprecipitation sequencing  
ncRNAs: non-coding RNAs  
lncRNAs: long non-coding RNAs  
eRNAs: enhancer RNAs  
PTBP1: polypyrimidine tract binding protein 1  
EZH2: enhancer of zeste homolog 2  
SNP: single nucleotide polymorphism  
TTS: transcription termination site  
GABA: gamma-aminobutyric acid  
SCA6: spinocerebellar ataxia type 6

## Introduction

### 1.1 Voltage-gated calcium channels (VGCCs)

VGCCs are essential for the function of excitable cells, mediating processes such as the release of neurotransmitters from neurons, the release of hormones by endocrine cells (Nanou & Catterall, 2018), as well as excitation-transcription coupling (Ma et al., 2011) and excitation-contraction coupling (Zamponi et al., 2015). These are transmembrane proteins that control the influx of calcium ( $\text{Ca}^{2+}$ ) by coupling depolarization of the plasma membrane to intracellular signaling pathways (Gonzalez-Ramirez & Felix, 2018). VGCCs mediate  $\text{Ca}^{2+}$  influx into excitable cells through a voltage sensor motif located in the transmembrane domain. This allows the channels to open in response to membrane depolarization (Catterall, 2010).

#### 1.1.1 Structure of VGCCs

VGCCs are comprised of an  $\alpha 1$  subunit and two auxiliary subunits: ( $\beta$ , and an  $\alpha 2\delta$ ). These subunits are encoded by separate genes (Zamponi et al., 2015). The  $\alpha 1$  subunits (encoded by ten separate *CACNA1* genes) are the pore-forming channel units that determine channel identity (Figure 1.1). Each  $\alpha 1$  subunit is a large protein (~200kDa) with four homologous domains, each with six transmembrane helices (S1-S6). Among six transmembrane helices, the S4 helix is positively charged and regulates voltage-dependent activation. A pore loop (P loop) motif between S5 and S6 of each domain forms the permeation pathway of the channel and ensures calcium selectivity (Figure 1.1). Each P loop contains an amino acid residue which is highly conserved and negatively charged in order to form a highly selective pore for calcium permeation (Simms & Zamponi, 2014). The transmembrane domains are connected by long cytoplasmic regions that are important for regulating channel function through protein-protein interactions (Hall et al., 2013; Simms & Zamponi, 2014). In particular, the cytoplasmic C-terminal region of VGCCs have

protein binding motifs for proteins such as Calmodulin (CaM) (Ma et al., 2011; Striessnig et al., 2014). In the C-terminal region, there are two putative alpha helices denoted the proximal C-terminal regulatory domain (PCRD) and the distal C-terminal regulatory domain (DCRD). The PCRD is located immediately after the IQ-domain, which is a CaM binding site, and the PCRD and DCRD together form a C-terminal Modulator (CTM) (Figure 1.1) (Singh et al., 2006, 2008).

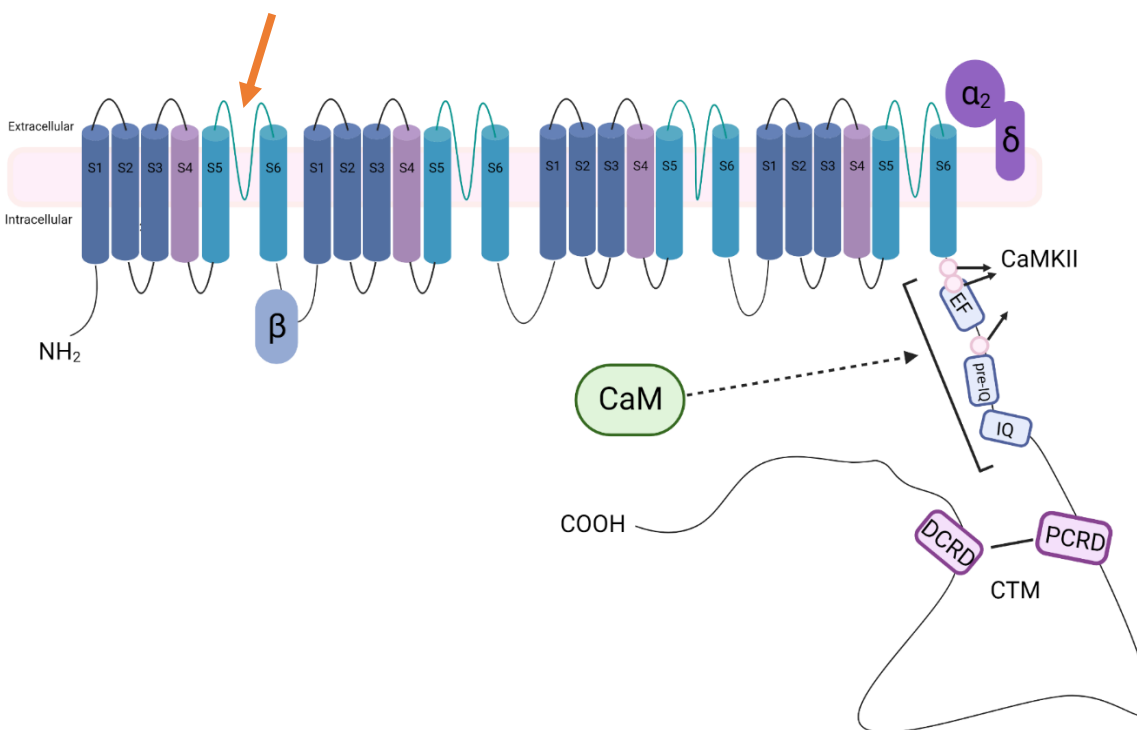


Figure 1.1. Schematic of the VGCC  $\alpha_1$  subunit comprised of four different domains, each with six segments and a P loop (marked with orange arrow). Created with BioRender.com

### 1.1.2 Electrophysiologically-defined subcategories of VGCCs

The VGCC  $\alpha_1$  subunits are divided into two types, high voltage-activated (HVA) channels and low voltage-activated (LVA) channels, based on the voltage threshold at which the channel is activated (Table 1.1). Within the HVA channel subclasses, there are L-type, P/Q-, N-, and R-type channels. L-type channels include a Cav1 type  $\alpha_1$  subunit, whilst the  $\alpha_1$  subunit for P/Q-, N-, and

R-type channels comes from the Cav2 subclass. In the LVA channels, there are T-type channels, which are from a Cav3 type  $\alpha 1$  subunit (Table 1.1). The LVA channels show slow activation, rapid deactivation, and require smaller membrane depolarization than HVA channels to open (Gonzalez-Ramirez & Felix, 2018; Zamponi, 2016)

Among the ten  $\alpha 1$  subunits, the Cav1 family (Cav1.1 – Cav1.4) are L-type Calcium channels (LTCCs) which require strong depolarization and maintain long-lasting currents during membrane depolarization. LTCCs are primarily recorded in cells that are electrically excitable such as smooth muscle, endocrine cells, and neurons. A strong depolarization is also required to activate N-type, P/Q-type and R-type calcium currents, which are primarily expressed in neurons where they stimulate neurotransmitter secretion and mediate calcium entry into cell bodies and dendrites. A weak depolarization is required to activate T-type calcium channels, which are expressed in various cell types and are responsible for shaping action potentials and regulating repetitive firing patterns. LTCCs are more sensitive to calcium blockers (e.g., dihydropyridines, phenylalkylamines, and benzothiazepines) and activators compared to other VGCCs. The function of  $\alpha 1$  subunits is modulated by auxiliary subunits -  $\beta$  (encoded by *CACNB*-genes), and  $\alpha 2\delta$  subunits (encoded by *CACNA2D*-genes). The function of auxiliary subunits is to refine channel characteristics and intracellular trafficking (Zamponi et al., 2015).

### **1.1.3. Biophysiological properties of VGCCs**

VGCCs are categorized into subunits based on their biophysical and pharmacological properties (Catterall et al., 2005) (Table 1.1). While the common function of VGCCs is to control calcium influx into a cell, each channel type has cell-specific functions (Lipscombe et al., 2013). In neurons, VGCCs regulate many key intracellular processes, including gene expression and neuronal transmission (Catterall, 2011; Ma et al., 2011). Specifically, calcium signaling plays an

important role in regulating neuronal excitability and synaptic plasticity (Berridge, 2014). In the brain, intracellular calcium levels increase through the opening of VGCCs, which in turn leads to the activation and nuclear localization of transcription factors followed by increases in gene expression (Simms & Zamponi, 2014). Several pathways activate transcription factors such as cAMP response element-binding protein (CREB) via VGCCs (Barbado et al., 2009). For example,  $\text{Ca}^{2+}$  through the LTCCs will activate nearby CaM that will trigger  $\text{Ca}^{2+}$ /Calmodulin-dependent protein kinase (CaMK) activation (Figure 1.1). This, in turn, activates CREB so that the  $\text{Ca}^{2+}$  activated Calmodulin can bind to the IQ domain in the C-terminal region of LTCCs, thereby activating the mitogen-activated protein kinase (MAPK) pathway. On the other hand, CaMKII can bind directly to the C-terminal domain of LTCCs and activate CREB pathway. Finally, transcription factors can directly bind to VGCCs genes and are activated by  $\text{Ca}^{2+}$  entering via VGCCs.  $\text{Ca}^{2+}$  binds to downstream regulatory element antagonist modulator (DREAM) so that transcriptional repression of promoters that contain downstream regulatory element (DRE) are deactivated and transcription is promoted (Barbado et al., 2009). As a result, alterations in calcium signaling, mediated by the action of VGCCs, influence complex neural functions such as learning and memory ability, and cognition (Berridge, 2014). Thus, VGCCs in the human brain play a significant role in regulating gene transcription and gene expression, which in turn affect neuronal activity and plasticity.

#### **1.1.4. LTCCs in excitable tissues**

Inside the cells, LTCCs engage in many biological processes such as contraction, transcription regulation and hormone secretion (Davenport et al., 2015). The LTCC show different expression profiles in different excitable cell types and have distinct gating characteristics and functions (Table 1.1) (Striessnig et al., 2014).

Cav1.1 channels, encoded by *CACNA1S*, are mainly expressed in the skeletal muscle where they function as voltage sensors (Catterall et al., 2005). The channels cause depolarization to release calcium from ryanodine receptors of the sarcoplasmic reticulum (Tuluc et al., 2009). These channels release calcium which triggers muscle contraction (Zamponi et al., 2015). Cav1.2 channels, encoded by *CACNA1C*, are expressed in several tissues such as the heart, smooth muscle, the brain, and endocrine cells (Striessnig et al., 2014). Cav1.3 channels, encoded by *CACNA1D*, are also expressed in the heart, smooth muscle, and the brain. Cav1.2 and Cav1.3 are expressed, and have crucial functions, in the brain. Both channels are postsynaptic channels mainly located in the spines and shafts of dendrites in neurons. They are responsible for neuronal connectivity, controlling neuronal firing and play a significant role in membrane excitability (Striessnig et al., 2014; Zamponi et al., 2015). Cav1.4 channels, encoded by *CACNA1F*, are predominantly expressed in retina (Zamponi et al., 2015); more specifically, rod photoreceptors and at retinal bipolar synapses (G. M. Y. Tan et al., 2012). *CACNA1F* plays a predominant role in calcium entry so that ribbon synapses can release glutamate to activate sensitive retinal photoreceptors (Catterall et al., 2005).

*Table 1.1 Categorization of VGCC  $\alpha 1$  subunits according to their physiological properties and voltage activation threshold*

Gene Name	Protein Name	Physiological Group	Voltage Type	Tissue Expression
<i>CACNAIS</i>	$Ca_v1.1$	L-type	HVA	Skeletal muscle
<i>CACNAIC</i>	$Ca_v1.2$			Brain, Heart
<i>CACNAID</i>	$Ca_v1.3$			Brain Heart
<i>CACNAIF</i>	$Ca_v1.4$			Retina
<i>CACNAIA</i>	$Ca_v2.1$	P/Q-type		Nervous system, Secretary tissue, Internal tissue
<i>CACNAIB</i>	$Ca_v2.2$	N-type		Nervous system, Secretary tissue, Internal tissue
<i>CACNAIE</i>	$Ca_v2.3$	R-type		Nervous system, Secretary tissue, Internal tissue
<i>CACNAIG</i>	$Ca_v3.1$	T-type	LVA	Nervous system
<i>CACNAIH</i>	$Ca_v3.2$			Nervous system, Secretary tissue, Internal tissue
<i>CACNAII</i>	$Ca_v3.3$			Nervous system, Secretary tissue, Internal tissue

Among the many genes encoding VGCCs, this thesis will focus on genes encoding L-type calcium channels (LTCCs), especially *CACNAIS* and *CACNAIF* which encode  $Ca_v1.1$  and  $Ca_v1.4$  respectively.  $Ca_v1.1$  is encoded by *CACNAIS* which is expressed predominately in skeletal muscle.  $Ca_v1.2$  is encoded by *CACNAIC* which is mostly expressed in the brain and heart.  $Ca_v1.3$  is encoded by *CACNAID* and is again primarily expressed in the brain and heart.  $Ca_v1.4$  is encoded by *CACNAIF* and is mostly expressed in retinal tissue.  $Ca_v2$  channels are divided into three distinctive types: P/Q type, N-type, and R-type.  $Ca_v2.1$  is of type P/Q, which is encoded by *CACNAIA*,  $Ca_v2.2$  is an N-type channel encoded by *CACNAIB*, and  $Ca_v2.3$  is an R-type channel encoded by *CACNAIE*. These genes are primarily expressed in the nervous system, secretary tissue and internal tissue.  $Ca_v3$  is a T-type calcium channel which is divided into 3 channels;  $Ca_v3.1$ ,  $Ca_v3.2$  and  $Ca_v3.3$ .  $Ca_v3.1$  is encoded by *CACNAIG* which is primarily expressed in the nervous system.  $Ca_v3.2$  is encoded by *CACNAIH* and  $Ca_v3.3$  is encoded by *CACNAII* which are mostly expressed in the nervous system and internal tissue.

## **1.2. Producing diverse mRNA transcripts**

Most protein-coding genes produce multiple alternative transcripts. These can arise from transcriptional mechanisms (using alternative transcription start site) and/or post-transcriptional mechanisms (alternative splicing). One gene can generate multiple pre-mRNAs by using alternative transcription initiation and those pre-mRNAs can experience alternative splicing, which in turn generates multiple transcript variants from the gene (Pal et al., 2011). A recent analysis claimed that alternative transcription initiation has a more considerable effect on transcriptome and proteome diversity, compared to alternative splicing (Reyes & Huber, 2018).

### **1.2.1. Using alternative transcription start site to produce multiple mRNA transcripts**

The majority of human protein-coding genes possess alternative transcription start sites (TSSs) (Carninci et al., 2006; Kimura et al., 2006). Largely, one gene has four TSSs (Xu et al., 2019). The 5' untranslated regions (5'UTRs) are regulatory regions which are transcribed to mRNA but do not encode protein. The highly conserved 5' UTRs contain regulatory motifs (transcription factor binding motifs) that regulate transcription by interacting with transcription factors (Birney et al., 2007; Byeon et al., 2021; Carninci et al., 2006). Alternative 5'UTRs, which may be constructed by alternative TSSs or alternative splicing, impact the expression of downstream coding regions (Smith, 2008). Genes with multiple alternative 5'UTRs result in having differential expression of each transcript (Resch et al., 2009). Any changes in the TSSs can change transcription expression, promoter usage, and translation initiation (Noderer et al., 2014), which in turn widen the possible range of tissue-specific or cell-specific expression patterns (Barrett et al., 2012).

Genes encoding ion channels do produce alternative transcripts using alternative TSSs (Okagaki et al., 2001; Drews et al., 2005; Shang & Dudley Jr., 2005; Gomez-Ospina et al., 2013).



These transcripts also show distinctive expression patterns in different tissues (Dai et al., 2002; Pang et al., 2003; Saada et al., 2003; Martin et al., 2007). However, other genes such as *Otx2*, Brain-derived neurotrophic factor (*BDNF*) gene, or neuregulin 1 gene (*NRG1*) also use alternative TSSs to produce multiple transcripts with either different functions or expression patterns. For example, *Otx2*, a gene involved in brain development, produces three different coding transcripts from different TSSs. All three transcripts (Type A, B, C) are expressed in the mouse brain but at different expression intensities between brain ventricles and medial habenula (Courtois et al., 2003). These transcripts are differentially expressed over the mouse developmental stages. In particular, a Type C is primarily expressed in the early development stage whereas Type A and B are mostly expressed in the late development stage. The inactivation of promoter for Type C overlaps the activation of a promoter for Type A (Fossat et al., 2005). Later, another alternative TSS of *Otx2* was identified, which is called Type C and the Type C transcript from the earlier studies became Type D transcript. Type A and Type B transcripts are required for anterior neuroectoderm, forebrain and mid brain development. Thus, the lack of these transcripts affects head morphogenesis whereas Type C and Type D transcripts are required to promote visceral endoderm anteriorization by activating *DKK1* and *Lefty1*. Also, protein encoded by Type C transcript is silenced at the end of gastrulation. These show that *Otx2* produces tissue-specific isoforms to guarantee normal gastrulation and head morphogenesis (Acampora et al., 2009).

Furthermore, *BDNF* also produces multiple transcripts from alternative TSSs and related promoters, which affect expression in tissue-specific manner. Human *BDNF* has a total of 8 alternative TSSs (Pruunsild et al., 2007). Even though all alternative transcripts encode the identical protein isoforms, they have cell-specific expression (Cattaneo et al., 2016). Among multiple transcript variants, transcript generated from promoter I and promoter II are associated

with regulating aggressive behavior in male. Transcripts with impaired promoter I and II reduced *BDNF* protein expression in hypothalamus and deficiency of 5-HT (Maynard et al., 2016). Moreover, transcript variants produced from promoter IV was studied to be linked with regulating sensory processing, fear regulation, sleep and modulating visual plasticity (Maynard et al., 2016) because mice with promoter IV silenced *BDNF* transcripts displayed elevated startle response, sleep disturbance and impaired sensory processing (Hill et al., 2016). These show that *BDNF* transcript variants possess independent functions *in vivo*, affected by selective usage of the promoters (Maynard et al., 2016).

On the other hand, alternative TSSs can encode alternative protein isoforms. TSSs affect translation initiation as loci near the TSS help to identify ribosome binding sites (Ingolia, 2014; Noderer et al., 2014). For example, a purine at position -3 from the TSS and guanine at position +4 from the TSS play significant roles in translation initiation efficiency (Kozak, 1984, 1986). In this manner, alternative TSSs can give rise to RNA transcripts with different coding sequences (CDSs) (Makhnovskii et al., 2022). For example, neuregulin 1 gene (*NRG1*) encodes multiple proteins with alternative TSSs affecting cell growth and differentiation such as neu differentiation factor (NDF), and glial growth factor (GGF). There are three known isoforms of *NRG1* with distinctive TSSs: *NRG1*-Type I, Type II and Type III. Each transcript encodes different protein isoforms. For example, Type II encodes GGF2 which is related to the increased susceptibility to schizophrenia (Stefansson et al., 2002). Later, 6 additional novel TSSs of *NRG1* were subsequently found, of which three produce novel protein coding transcripts: Type IV, Type V and Type VI (Steinthorsdottir et al., 2004a). The 5'UTR of Type I, Type II and Type III is conserved across rat, mouse and human. However, the Kozak sequence for Type IV is not conserved implying that the protein coding potential of this transcript is unique to human (Steinthorsdottir et al., 2004a).

Mutations in transcription regulatory sequences are known to affect various diseases including cancer, developmental disorders, and psychiatric disorders. Hence, understanding transcriptional regulation mechanisms is crucial to unravel complex disease aetiology (Casamassimi & Ciccodicola, 2019). In particular, understanding transcript regulation of LTCC genes such as *CACNA1C* may facilitate the development of new therapies for neuropsychiatric disorders such as Bipolar disorder, Schizophrenia, and Parkinson's disease (Striessnig et al., 2015; Zamponi et al., 2015). Therefore, a key step in understanding transcription regulation would be to study gene TSSs, since loci with TSSs contain transcription binding sites that interact with cis-regulatory elements (Goszczyński et al., 2021).

### **1.2.2. Alternative splicing to generate diverse mRNA transcripts**

The information in DNA is transferred to a mRNA molecule via transcription (Clancy & Brown, 2008). Pre-mRNA initially contains both exons and introns. Introns are removed from the mRNA via RNA splicing, leaving the chain of exons in the protein-coding mature transcripts (Abbasi et al., 2011). Splicing is a prevalent mechanism : more than 95% of human genes undergoing splicing events (Wahl et al., 2009; Y. Wang et al., 2015). In human cells, splicing is primarily mediated by rearrangements of the spliceosome, which is formed of various components including small nuclear ribonucleoproteins (snRNPs; U1, U2, U4, U5 and U6) (Y. Wang et al., 2015). The spliceosome play significant roles in mediating the splicing mechanisms (Baumgartner et al., 2015; Otake et al., 2002).

Prior to splicing events, exon-intron boundaries should be recognized as they are splice sites for major spliceosome binding. There are two models of the recognition which are exon definition and intron definition (Dvinge, 2018). Exon definition should have occurred correctly for conventional splicing. Exon definition occurs when snRNP binds to the exonic units to recognize

the exons. However, alternative splicing, such as exon skipping, may occur if the recognition is obstructed (M. Chen & Manley, 2009). The splicing mechanism may be regulated by cis-acting elements which act either as splicing enhancers or splicing silencers. These elements, along with protein regulators, determine whether the region is constitutively or alternatively spliced such as which exons to be included or skipped during splicing (M. Chen & Manley, 2009). Based on the location and function of the elements, they are categorized into four groups which are exonic enhancers, exonic silencers, intronic enhancers, and intronic silencers (Furlanis et al., 2019).

Alternative splicing is common in mammalian brain (Lipscombe et al., 2013), which occurs in a tissue-specific manner across different brain regions and developmental stages (Dillman et al., 2013; Licatalosi & Darnell, 2010), including for VGCC genes (Gray et al., 2007). For example, *BDNF*, which produces multiple transcripts using alternative TSSs, also undergoes extensive alternative splicing, which in turn affects its gene expression in tissue-specific or cell-specific manner. *BDNF* transcripts with exons II, III, IV, V and VII show brain-specific expression, whereas transcripts with exons VI and IX have the most abundant expression in non-neural tissues such as heart, lung, and skeletal muscle. Thus, the expression of *BDNF* is regulated at transcript-level to result in tissue- and cell-specific expression patterns (Cattaneo et al., 2016; Pruunsild et al., 2007). Not only does alternative splicing mechanism regulates the gene expression but also, the mechanism generates protein diversity (M. Chen & Manley, 2009; Wahl et al., 2009). Several different types of alternative splicing patterns (Figure 1.2) can result in various forms of mRNA with distinct functions (Y. Wang et al., 2015). Among many different patterns of alternative splicing, the most common pattern is cassette exon alternative splicing where exons are skipped (E. Kim et al., 2007).

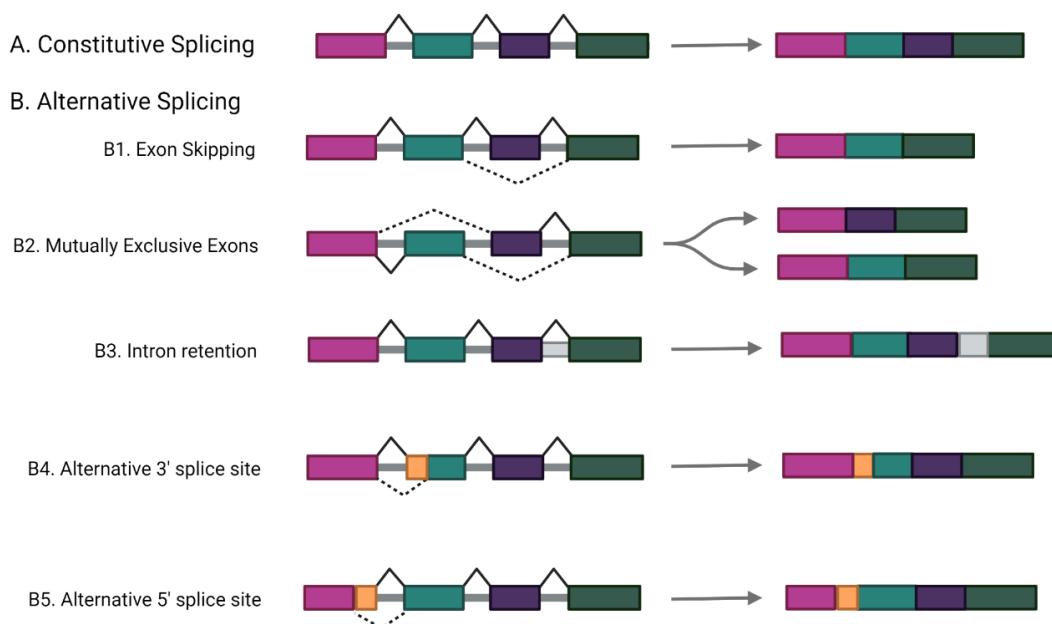


Figure 1.2. Schematic cartoon of alternative splicing patterns **A.** Constitutive splicing pattern **B.** Alternative splicing patterns Created with BioRender.com

### 1.3. Producing diverse LTCCs at mRNA level

#### 1.3.1. Usage of alternative TSSs to generate diverse LTCCs

Numerous numbers of novel coding transcripts were identified using long-read sequencing for example, among 983 novel transcripts, 333 transcripts have alternative TSSs (Wright et al., 2022). It was reported that the Cav1.2 channel, encoded by *CACNA1C*, has alternative first exon usage in a tissue-specific manner (Blumenstein et al., 2002). *CACNA1C* transcripts expressed in the heart include exon 1a, which encodes 46 amino acid N terminus, whereas exon 1b, expressed in the brain, encodes an alternate 16 amino acid sequence (Figure 1.3) (Blumenstein et al., 2002). The Cav1.2 with exon 1a shows functional differences to the channel containing exon 1b, including reduced channel activity (Shistik et al., 1998; Wei et al., 1996). It is also more likely to be modulated by protein kinase C (McHugh et al., 2000). Later, many researchers demonstrated that the tissue-specific N-terminal diversity of the human Cav1.2 channel is regulated by alternative

promoter regions. They identified two human Cav1.2 mRNA isoforms, where one had a longer N-terminus region with exon 1a (only expressed in the human heart) and the other with a shorter N terminus region with exon 1b (expressed ubiquitously in other human tissue) (Dai et al., 2002; Pang et al., 2003; N. I. Saada et al., 2005). The TSSs of two isoforms were detected by 5' RACE analysis. The promoter for transcripts with exon 1a functions to upregulate the expression of the exon 1a-containing transcripts without affecting the expression of exon 1b-containing transcripts. This suggests that the existence of alternative promoters allow the control of channel expression in one tissue type without affecting the channel expression in other tissue types (Dai et al., 2002; Pang et al., 2003).

Furthermore, the three-domain channel is produced through alternative transcription start site usage. Okagaki et al (2001) found that *TuCa1*, the protein of which is homologous to that of vertebrate LTCCs, produces three domain channels (Okagaki et al., 2001). The existence of the three-domain channel of *TuCa1* was first observed from the northern hybridization assay, which demonstrated that the shorter isoform was more intensely detected from the egg cells and embryos compared to the longer isoform. As the nucleotide sequences of cDNA of both three-domain channel and full-length channel are the same, it was assumed that two transcripts originate from the same gene. The three-domain channel encodes a truncated type of calcium channel without all the segments of the first domain, and S1-S2 of the second domain (Figure 1.3). 5'RACE experiment confirmed that the transcript encoding three-domain channel uses alternative TSS and alternative promoter site, locating between existing exons, compared to the transcript encoding the full-length channel. When the three-domain channel was co-expressed with the full-length channel in *Xenopus* oocyte, the three-domain channel did not produce a functional calcium current but had a significant inhibitory effect on the expression of the full-length channel in the ascidian muscle

blastomeres. These suggest that the three-domain channel regulates the LTCCs current amplitude during early developmental stages and cell differentiation (Okagaki et al., 2001). As *TuCa1* produces non-mammalian LTCCs, researchers created the similar three-domain channel of Cav1.2 by using the start codon at the same location as that creates mammalian three-domain channel of *TuCa1* and co-expressed to BHK-6 cells with the full-length Cav1.2. It was found that three-domain channel of Cav1.2 reduces the currents and expression of full-length Cav1.2. This shows that three-domain channel of Cav1.2 has a strong inhibitory effect on the full-length Cav1.2 (Ebihara et al., 2002), which may be due to the intact C-terminal region (Liao et al., 2015).

Moreover, Gomez-Ospina *et al* (2006) found that in neurons, Cav1.2 channels produce C-terminal fragments that function as transcription factors, named the Calcium Channel Associated Transcript (CCAT). CCAT is located in the nucleus of neurons and exported from the nucleus via  $\text{Ca}^{2+}$  influx through LTCCs and N-methyl-D-aspartate (NMDA) receptors. In the nucleus, the CCAT interacts with the transcriptional regulator p54/(nrb)/NonO. This interaction activates the transcription of neuronal genes, such as NMDA-receptor subunits, and the sodium-calcium exchanger. CCAT binds to the *Connexin 31.1* gene enhancer and regulates its expression. Moreover, CCAT expression can increase neurite extension in the primary neurons (Gomez-Ospina et al., 2006). Later, Gomez-Ospina et al. (2013) found that CCAT is produced by an internal promoter in exon 46 of *CACNA1C*, rather than as the result of protein cleavage (Figure 1.3) (Gomez-Ospina et al., 2013). As the expression of CCAT is independent from the expression of the full-length *CACNA1C*, CCAT is not cleaved from the full-length *CACNA1C*. They identified three novel TSSs in the 3' region of *CACNA1C*, of which TSS 4 is predicted to encode a 15kD protein which might represent a full-length CCAT. TSS 4 was confirmed using CAGE data, (Gomez-Ospina et al., 2013). The expression of CCAT reduces as the brain develops and is

speculated to control Cav1.2 channel expression, as it suppresses Cav1.2 expression during brain development (Gomez-Ospina et al., 2013).

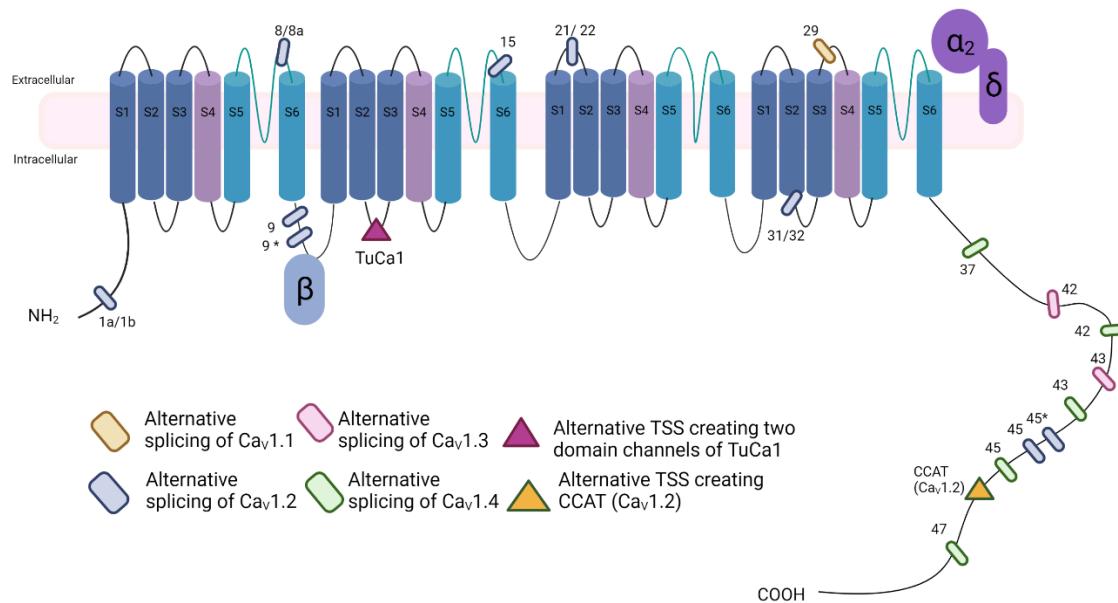


Figure 1.3. Schematic figure of alternative splicing and alternative TSS location of LTCCs Created with BioRender.com

### 1.3.2. Alternative splicing mechanisms to generate diverse LTCCs

LTCC genes undergo extensive alternative splicing, allowing for increased channel functional diversity with the capacity to modulate the biophysical and pharmacological properties of each channel depending on cell or tissue conditions (Liao et al., 2009). For example, *CACNA1S*, which encodes Cav1.1 channels, undergoes alternative splicing at exon 29, which is located on the loop between S3 and S4 helices in the last domain (Figure 1.3). Two transcripts were identified: one transcript has exon 29 and the other transcript lacks this exon. The major form of *CACNA1S* in early muscle development lacks exon 29; in contrast, exon 29 is increasingly included as



development progresses. The alternatively spliced exon 29 affects the electrophysiological property of Cav1.1, increasing its voltage sensitivity (Tuluc et al., 2009, 2016).

*CACNA1C* undergoes extensive alternative splicing such that 251 transcripts have been identified using long-read nanopore sequencing (Clark et al., 2020). There are several reported alternatively spliced exons, either novel exons or mutually exclusive exons; exon 8/8a, exon 9/9\*, exon 15, exon 21/22, exon 31/32, and exon 45/45\* (Figure 1.3) (Liao et al., 2009; Tang et al., 2004). Alternative splicing of exon 15 deletes 73 nucleotides, introducing a premature stop codon but encoding a transcript that is visible in both brain and heart tissue of humans and rats (Tang et al., 2004). This alternatively spliced transcript produces truncated two-domain isoform of Cav1.2 that reduces the current of full-length Cav1.2 in smooth muscle cells (Cox & Fromme, 2013). Alternative splicing provides a means to encode proteins with different amino acid sequences and functions in different tissues (Lipscombe & Andrade, 2015). For example, exon 9\*, which includes an additional 75nt between exon 9 and exon 10, is expressed in the adult heart but not the brain (Tang et al., 2004). Exon 45 is expressed in human fibroblast but not heart or brain, whereas the alternative Exon 45\* is expressed in the human heart and brain but not in fibroblasts (Tang et al., 2004).

Alternative splicing in the C-terminal region of Cav1.3, encoded by *CACNA1D*, changes the gating properties of the resulting channel (Bock et al., 2011). There are two different subtypes of *CACNA1D* resulting from alternative splicing in the C-terminal region (Figure 1.3; Figure 1.4). The *CACNA1D* transcript with exon 42 contains the CTM domain whereas the *CACNA1D* transcript with alternatively spliced exon 42A has a stop codon right after the IQ domain, which is a binding site for CaM, and does not have CTM. CTM reduces the calcium-dependent inactivation (CDI) therefore, spliced variant with exon 42A would show enhanced CDI, being activated at more

negative voltages and showing faster inactivation (B. Z. Tan et al., 2011). The alternatively spliced variant with exon 42A may affect CDI by hindering CaM binding (Zuccotti et al., 2011). Exon 43 of *CACNA1D* is also alternatively spliced using alternative 3' splice acceptor sites in exon 43. The alternative splicing of exon 43 introduces an early stop codon immediately after the PCRD. The alternatively spliced transcript (Cav1.3 43s) produces the C-terminal truncated Cav1.3 channel with an early stop codon. This transcript also shows tissue-specific expression because it can be found in various brain regions but has no expression in the sinoatrial node (SAN) (B. Z. Tan et al., 2011).

A switch from exon 42 to 42A could result in a neuroprotective effect, because the 42A splice variant would accelerate the CDI and therefore limit Ca<sup>2+</sup> influx, with a consequent positive impact on Ca<sup>2+</sup> neurotoxicity. However, a drug able to switch the splice variant from 42 to 42A might cause slower heart rate, because the 42A isoform is likely to initiate rather than sustain the pacemaker activity (B. Z. Tan et al., 2011).

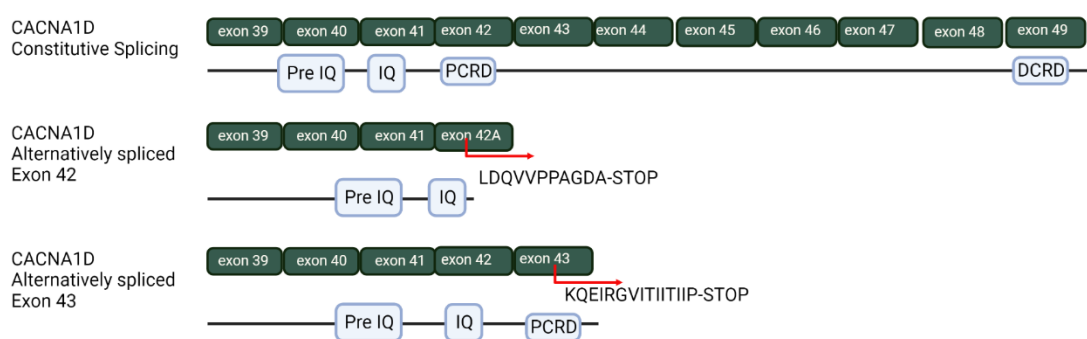
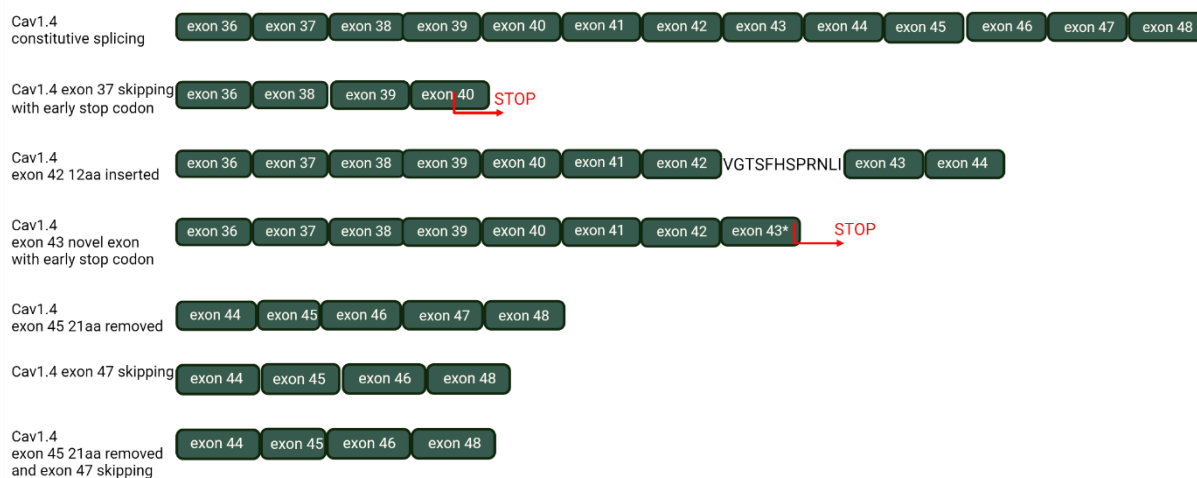


Figure 1.4 Schematic figure of alternative splicing in the C-terminal region of *CACNA1D*. PCRD is proximal carboxy-terminal regulatory domain, DCRD is distal carboxy-terminal regulatory domain.

Created with BioRender.com

There are many alternative splicing events in the C-terminal region of Cav1.4, encoded by *CACNA1F*, that affect channel properties (Figure 1.3; Figure 1.5). Alternative splicing in the C-terminal region of *CACNA1F* was reported to affect calcium inhibition in the human retina; more specifically, affecting CDI to stabilize the Ca<sup>2+</sup> currents for photoreceptor signaling (Singh et al., 2006; G. M. Y. Tan et al., 2012; Wahl-Schott et al., 2006). Firstly, exon 37 is excluded which results in truncation of the entire C-terminal region so that this variant does not show any functional channel activity. There are two variants with additional amino acids which are a transcript with 13 amino acids added to exon 42, and 19 amino acids added to exon 43 with an early stop codon so that the last 257 amino acids are missing. Other in-frame variants are one with 21 amino acids removed from exon 45, and a variant with exon 47 skipped. Another in-frame variant has a shorter exon 45 and 47 skipping and expresses in the primate retina but not in the mouse retina (Figure 1.5) (G. M. Y. Tan et al., 2012).

Among these alternatively spliced variants, those with an alternatively spliced exon 43, and an alternatively spliced exon 45 affect the channel property (Haeseleer et al., 2016; G. M. Y. Tan et al., 2012). The added amino acids in exon 43 activate CDI, increase current density, and trigger a hyperpolarized shift of voltage (G. M. Y. Tan et al., 2012). In addition, the transcript without exon 47 alters the function of the channel because the transcript does not have a part of CTM. As the exon 47 is removed, the channel shows more robust CDI than the full length of the channel (Singh et al., 2006; G. M. Y. Tan et al., 2012) and causes voltage activation to be positively shifted (Haeseleer et al., 2016). The exon 47 appears to be important for inhibiting CDI but does not affect the CaM or CaBP4 binding. However, exon 47 is needed for the functional modulation of channel activity by CaBP4 (Haeseleer et al., 2016).



*Figure 1.5. Schematic figure of alternative splicing in the C-terminal region of CACNA1F. Created with BioRender.com*

#### 1.4. Producing truncated LTCCs due to proteolytic cleavage

The preceding section summarized findings showing generation of diverse transcript variants or channels of LTCCs due to transcriptional mechanisms. Complementing these findings, there is also evidence that proteolytic cleavage can occur in the C-terminal region of LTCCs and produce shorter protein fragments with distinct functions. The proteolytic cleavage between PCRD and the DCRD of Cav1.1 and Cav1.2 creates a short C-terminal fragment which affect the calcium channel activity. For example, Cav1.1 undergoes proteolytic cleavage in the C-terminal region (A1664) (De Jongh et al., 1991; Hulme et al., 2005). The cleaved fragment contains amino acids 1802-1841 which interacts with the region (amino acid 1556-1612) in the non-cleaved C-terminus and increases calcium entry into the skeletal muscle. The cleaved cytoplasmic C-terminus fragment have the autoinhibitory function of the channel (Hulme et al., 2005). The C-terminus of Cav1.2 also experiences proteolytical cleavage between PCRD and DCRD (amino acid residue 1821), which interacts with the region (arginines 1696 and 1697) in the PCRD to inhibit Cav1.2

channel function. The cleaved fragment of Cav1.2 becomes an autoinhibitory domain that reduces channel opening probability in cardiac myocytes (Fuller et al., 2010; Hulme et al., 2006).

Furthermore, one study suggested that the C-terminal region of Cav1.3 is localized to the nucleus in atrial myocytes and functions as a transcriptional regulator for expression of *MyI2*, which encodes MLC2 protein (L. Lu et al., 2015). However, it is unclear how the C-terminus fragment of Cav1.3 is produced. It could be produced by proteolytic cleavage because the amino acid sequences in PCRD and DCRD are highly conserved among Cav1.1, Cav1.2 and Cav1.3, suggesting that those channels may use proteolytic cleavage in the C-terminal region as a channel regulatory mechanism to produce autoinhibited channel (Hulme et al., 2005; Y. Yang et al., 2022). On the other hand, the c-terminal fragment of Cav1.3 can be generated by introducing a stop codon by alternative splicing (Figure 1.4) (Scharinger et al., 2015; Y. Yang et al., 2022).

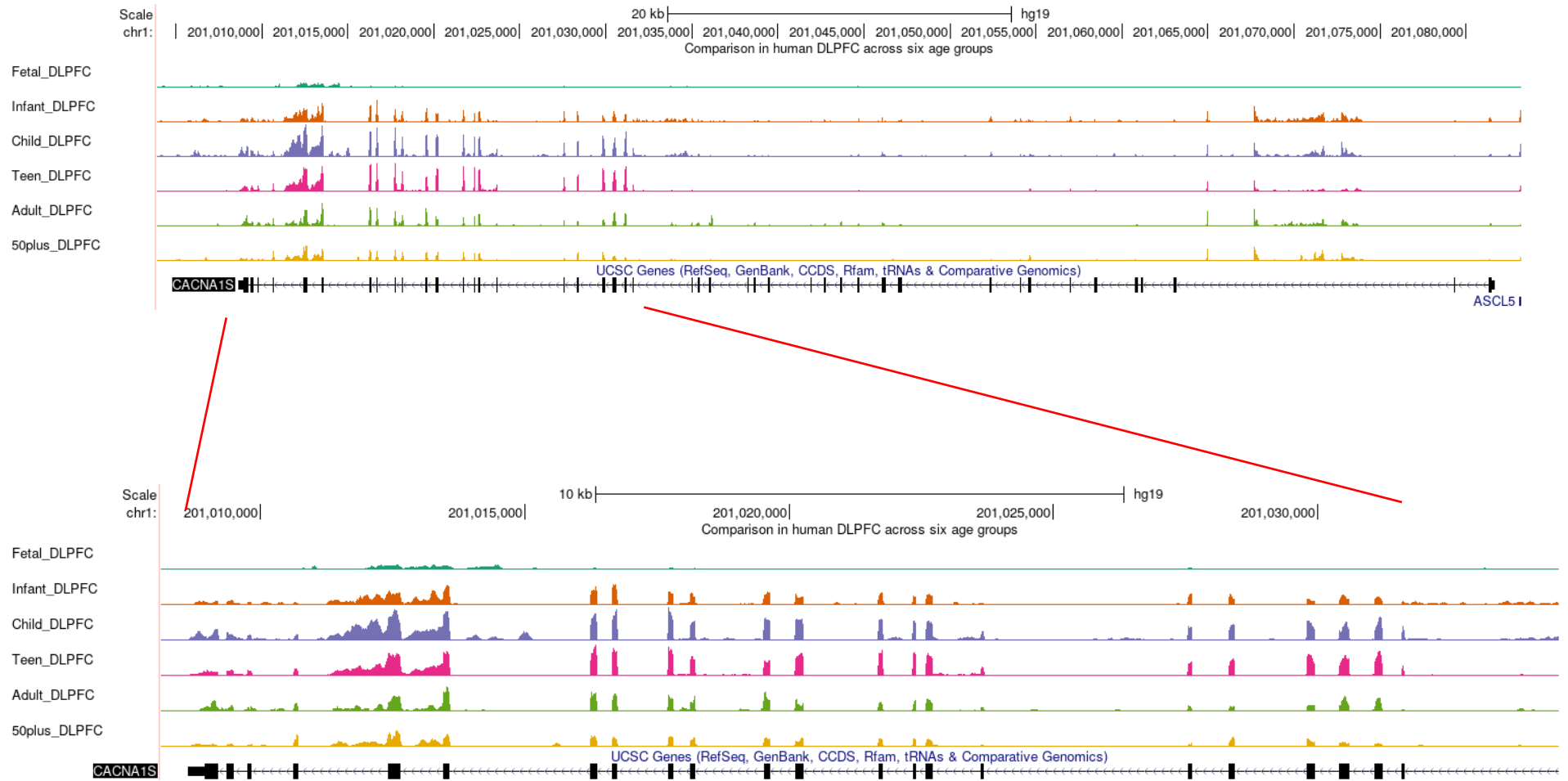
### **1.5. Observations of expression of the 3' region of *CACNAIS* and *CACNAIF* in the human brain**

The majority of VGCC genes are expressed in the brain, but *CACNAIS* and *CACNAIF* are reported to show restricted expression in the skeletal muscle and the retina, respectively. The expression of these genes in the human brain has not been reported (McRory et al., 2004; Tuluc et al., 2009; Zamponi et al., 2015). However the short-read RNA sequencing data in the human brain from Lieber Institute of Brain Development (LIBD) (Jaffe et al., 2015) suggests that a segment of the 3' region of both genes shows expression even though the full-length transcripts are not detected (Figure 1.6 A for *CACNAIS* and Figure 1.6 B for *CACNAIF*). Figure 1.6 C shows the predicted genomic location of short-length transcripts from the 3' region of *CACNAIS* (Blue arrow) and *CACNAIF* (Red arrow). Moreover, exon expression analysis and exon junction expression analysis from Genotype-Tissue Expression (GTEx) project (GTEx portal) suggests that in both

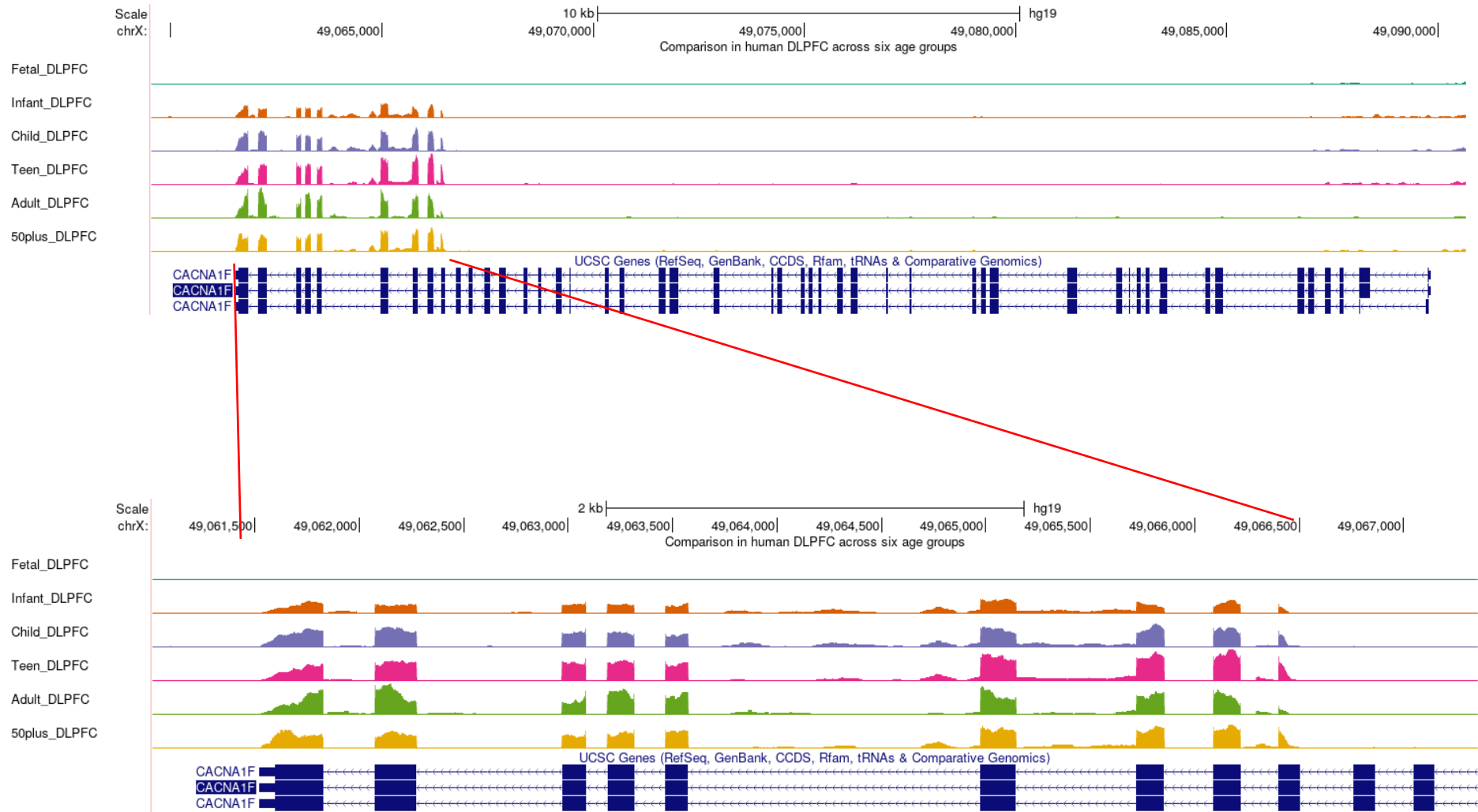
*CACNAIS* (Figure 1.7A) and *CACNAIF* (Figure 1.7B), exons toward the 3' region have higher expression level compared to the exons upstream towards the 5' region of each gene in the human brain. Despite exon expression data of *CACNAIS* (Figure 1.7A) shows that there are no exon expressions of *CACNAIS* in human tissues except skeletal muscle (blue color in Figure 1.7A), there is evidence of expression in exon junction expression data in human brain (red color in Figure 1.7A). Exon junctions toward the 3' region of the *CACNAIS* have higher expression, suggesting that *CACNAIS* transcript with those exon junctions have an expression in the human brain.

These observations suggest that *CACNAIS* and *CACNAIF* may produce short transcripts from their 3' region in human brain (Barrett et al., 2012; Dai et al., 2002; Pang et al., 2003; N. Saada et al., 2003). Therefore, this thesis will use a series of experimental methods to examine the short-length transcript from the 3' region of LTCC genes in the human brain.

A.



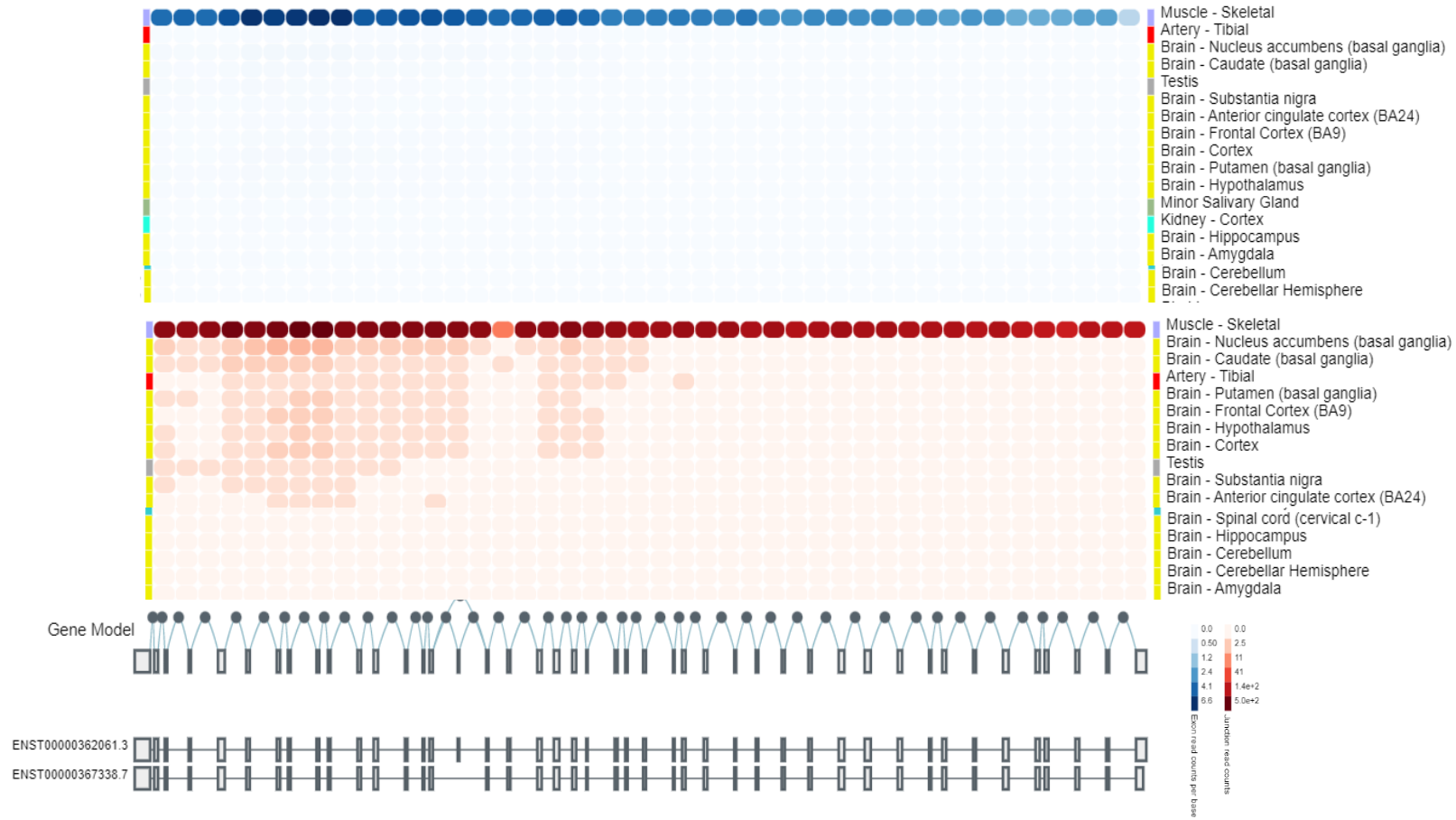
B.







A.



B.

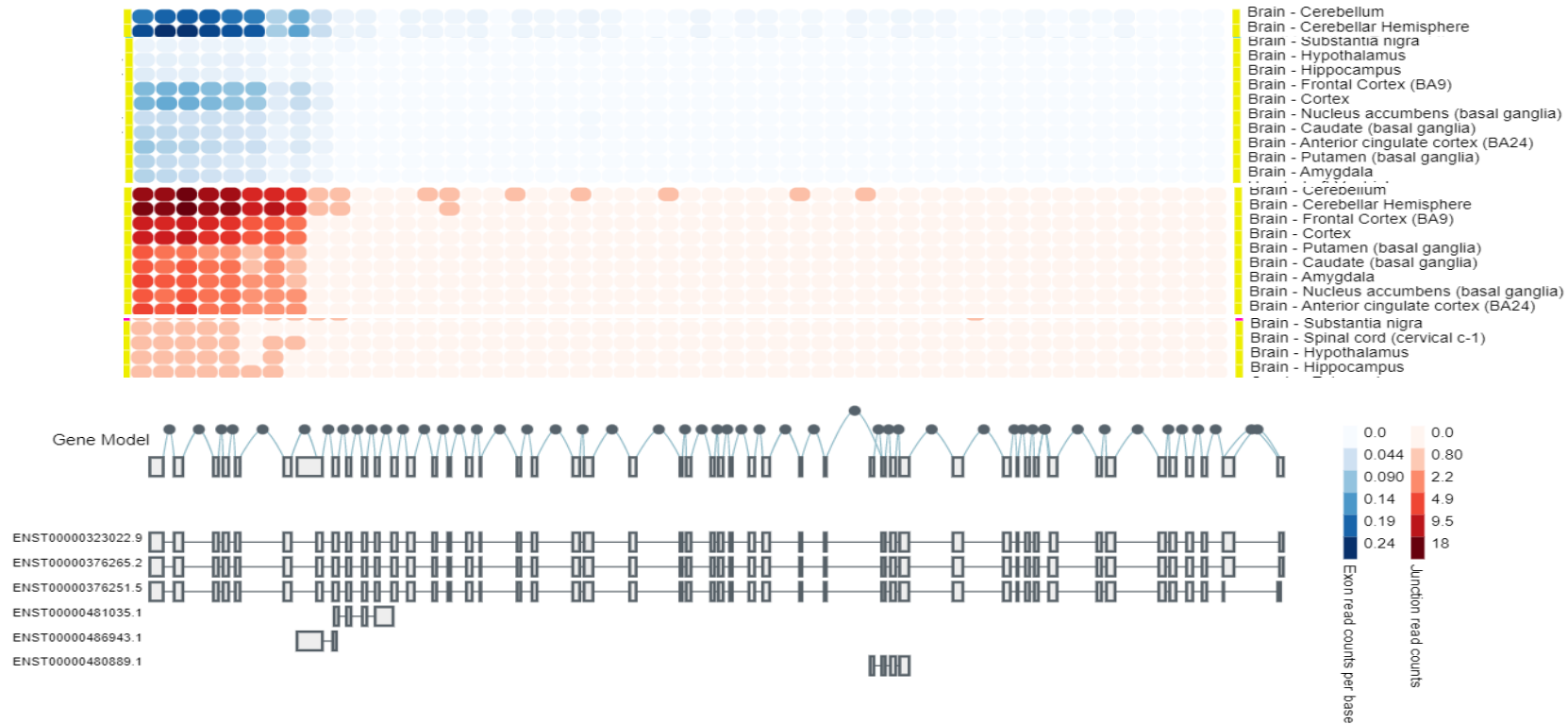


Figure 1.7 Expression analysis from the GTEx portal. Blue colored circles show the expression level of each exon in each tissue on the right. Red colored circles show expression level of exon junctions in each tissue on the right **A**. The exon and exon junction expression analysis of CACNA1S in different human brain. These show that there is evidence of expression of the short-length transcript from the 3' region of CACNA1S. **B**. The exon and exon junction expression analysis of CACNA1F in different human brain. These imply that short-length transcript from the 3' region of CACNA1F would express in the human brain.

## 1.6. Aim of the thesis

This thesis aims to characterize the short-length transcripts in the 3' region of *CACNAIS* and *CACNAIF* examined from the short-read RNA sequencing data using various methods and analyses. Firstly, I will aim to detect the existence of short-length transcripts in the human brain via PCR amplification, which will be sequenced using long-read nanopore sequencing. Then, the transcription start site of the short-length transcripts in the human brain will be identified using 5'RACE. If they are found to be complete transcripts with their own transcription start sites, I will examine the possible functionality of 5'UTR using randomized resampling of the PhyloP score from the Zoonomia project. In addition to that, the expression pattern of the short-length transcripts of LTCC genes will be examined using exon expression data from GTEx. These analyses will elucidate whether the short-length transcripts have tissue-specific expression patterns. Finally, using immunofluorescence, the localization of the protein encoded by the 3' transcripts of LTCC genes in both HEK293T cells and SHSY-5Y cells will be examined.

## 1.7. Summary of the thesis

In summary, LTCCs play a key role in normal brain functions, as well as being implicated in the aetiology and pathophysiology of psychiatric and neurological disorders (Harrison et al., 2020). Genes encoding LTCCs are known to undergo a variety of modifications at the transcriptional level, as well as at the post-translational level, which increases their functional diversity.

One interesting observation regarding the modification of genes encoding LTCCs is the apparent expression of the short-length transcripts from the 3' region of LTCC genes. However, this possibility has not been explored for *CACNAIS* and *CACNAIF* in the human brain. The potential importance of the short-length transcripts is supported by evidence suggesting that

protein fragments from the C-terminal region of other LTCCs may act as transcription factors, or may modulate the properties of the full-length channels (Gomez-Ospina et al., 2013; Y. Yang et al., 2022).

In this thesis, the possible short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* will be examined using a series of experiments:

Chapter 2: The short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* were identified in the human brain using a combination of PCR amplification and long-read nanopore sequencing. In-frame coding transcripts were identified that are comprised of known exons with possibility of exon skipping. No novel exons were identified.

Chapter 3: In this chapter, it was hypothesized that the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* would be originated from the de novo TSS. The TSSs of the transcripts were identified using 5' Rapid amplification of cDNA ends (RACE), supported by genome-wide analysis using CAGE data from FANTOM5 project. The short-length transcripts from the 3' region of LTCC genes are derived from de novo TSSs within the exon of the full-length gene. From this chapter, '3' transcripts of LTCC genes (*CACNAIS*, *CACNAIC*, *CACNAID*, *CACNAIF*)' will be used throughout this thesis to describe the short-length transcript from the 3' region of LTCC genes.

Chapter 4: This chapter aimed to study conservation status of the 5' UTR of the 3' transcripts and tissue-specific expression of the 3' transcripts. It was hypothesized that the 5'UTR, which may contain functional regions (Byeon et al., 2021; Siepel et al., 2005), would be significantly more conserved than the exonic regions and/or intronic regions. The conservation status of the 5' UTR of the 3' transcripts of LTCC genes across 240 mammals was examined by analyzing PhyloP score from the Zoonomia project (Zoonomia Consortium, 2020). Also, the expression pattern of the 3'

transcripts of LTCC genes were examined by analyzing exon expression data from GTEx project. This chapter concluded that the 5'UTRs of 3' transcripts of LTCC genes are significantly more conserved than intronic regions across 240 mammals suggesting that 5'UTR may contain regulatory motifs. The 3' transcripts of *CACNAIF* and *CACNAIC* showed brain tissue specific expression pattern. The 3' transcript of *CACNAIS* showed higher expression once the transcription initiation in all tissue types. However, the 3' transcript of *CACNAID* did not show significantly increased expression after the transcription initiation in human tissue.

Chapter 5: The localization of the proteins encoded by the short-length transcripts of LTCC genes in HEK293T and SHSY5Y cells were examined. In this chapter, 'C-LTCCs, C-Cav1.1, C-Cav1.2, C-Cav1.3, C-Cav1.4' were used to describe the proteins encoded by the 3' transcripts of LTCC genes or each gene respectively. It was hypothesized that the C-LTCC will be localized to the nucleus in the cells based on the previous studies (Gomez-Ospina et al., 2013; L. Lu et al., 2015). C-Cav1.1 and C-Cav1.4 were not localized to the nucleus in both cell lines whereas C-Cav1.3 was localized to the nucleus in both cell lines. C-Cav1.2 was not localized to the nucleus in HEK293T cells but was localized to the SH-SY5Y cells. Co-immunoprecipitation to examine protein-protein interaction was attempted but was not successful.

Chapter 6: General discussion

## **Chapter 2: Characterization of the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* in the human brain**

### **2.1. Introduction**

#### **2.1.1. Evidence for the expression of the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* in the human brain**

As discussed in section 1.5 of chapter 1, accruing evidence suggests that exons in the 3' region of *CACNAIS* and *CACNAIF* are expressed in the human brain. The evidence for expression can be seen in the short-read RNA sequencing (RNA-seq) data of the human brain from Lieber Institute of Brain Development (LIBD; Jaffe et al., 2014), displayed in the UCSC genome browser (hg19) (<http://genome.ucsc.edu>), and short-read RNA-seq data from the Genotype-Tissue Expression (GTEx) project, displayed as exon expression levels and exon junction expression levels across multiple human tissues. According to the short-read RNA-seq data from both LIBD (see Figure 1.6 in chapter 1) and the GTEx project (see Figure 1.7 in chapter 1), exons toward the 3' of *CACNAIS* and *CACNAIF* have a much higher expression levels compared to exons toward the 5' end of the gene. Expression data shows that exon junctions towards the 3' regions of *CACNAIS* and *CACNAIF* have a higher expression than exon junctions towards the respective 5' regions. As expression data from GTEx project examined expression levels across multiple human tissues, this means that the transcripts containing the exon junctions have expression in the human tissues (Figure 1.7 in chapter 1). However, all existing evidence is from short-read RNA-seq, which has an intrinsic bias towards the 3' end of the transcripts. This is because the sequencing employs the poly(A) enrichment using oligo(dT) primers to enrich mRNA from the samples (Shi et al., 2021). Nevertheless, the sudden increase in reads at the exon towards 3' region of each gene

supports the hypothesis that these transcripts might express in the human brain, and not due to 3' bias. Therefore, the aim of this chapter is to detect the presence of the short-length 3' transcripts of *CACNAIS* and *CACNAIF* in the human brain via PCR amplification and characterise them using long-read nanopore sequencing.

## **2.1.2. Long-read nanopore sequencing**

### **2.1.2.1. What is long-read nanopore sequencing?**

Long-read nanopore sequencing uses flow cells with an array of nanopores which are embedded in an electrically resistant polymer membrane (Deamer et al., 2016; van Dijk et al., 2018). The electrodes measure the electrical currents that flows through the nanopore. Negatively charged single-stranded oligonucleotide molecules pass through the nanopore and electrodes measure the electrical currents flowing through the pore. Real-time sequencing is possible as the disruptions in ionic current (called “squiggle”) are decoded as each nucleotide passes through the pore (<https://nanoporetech.com/how-it-works>). Multiple methods have been developed to improve accuracy by reading the same single-stranded DNA molecule twice (2D method) or pairing reads from each complementary strand (1D<sup>2</sup>). However, the long-read nanopore sequencing method used in this chapter only supports the 1D method, which reads a single strand only once. In the 1D method, the motor protein unwinds double-stranded DNA molecules to single-stranded molecules to pass through the nanopore. Both ends of each strand are ligated with an adapter to get sequenced independently. Then the enzyme is ligated to the 5' end of the strand to translocate each nucleotide correctly. Once molecules have passed through the flow cell, data about number of active nanopores, speed of molecule translocation, and running time are measured (Y. Wang et al., 2021). In this chapter, MinION, which is a small portable sequencing device with powerful computing



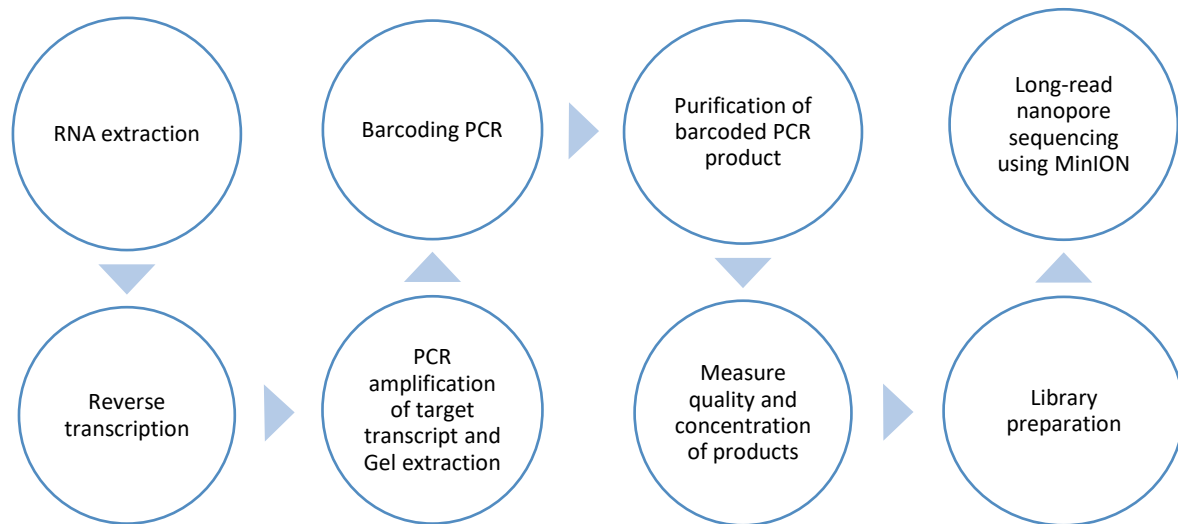
power, was used to sequence the short-length transcript from the 3' region of *CACNAIS* and *CACNAIF* in the human brain.

### **2.1.2.2. Advantages and disadvantages of long-read nanopore sequencing**

Previously, transcriptome profiling relied on the short-read RNA-seq technology (Stark et al., 2019; Z. Wang et al., 2009). Short-read RNA-seq provides accurate high coverage reads at low cost (Z. Wang et al., 2009), but produces reads under 200bp in length. As a result, data from short-read RNA-seq must be computationally mapped to a gene or transcript. Because the fragments produced by this method are short, sequencing and mapping can be ambiguous and transcript characterization can therefore be inaccurate (Dong et al., 2021; Kukurba & Montgomery, 2015). In contrast, the long-read nanopore sequencing enables long-read RNA sequencing up to 100kbp (Miga & Koren, 2020). The long-read nanopore sequencing is a cost-effective high throughput sequencing method that can sequence full-length transcripts (Byrne et al., 2017), thereby minimizing ambiguity (Stark et al., 2019). This feature makes it suitable for the study of splicing patterns such as alternative splicing or co-splicing of transcripts, including the identification of novel and deleted exons. Reads from the long-read nanopore sequencing can be aligned to a reference genome and used to reconstruct gene isoforms. Therefore, the long-read sequencing is used to elucidate genomic complexity by using unambiguous alignments. In contrast, the short-read RNA-seq is suited to characterize details of specific locations, such as splice site analysis, as it has higher accuracy and throughput. The main disadvantage of long-read nanopore sequencing is its comparably lower accuracy rate, although this can be overcome during the analysis stage (Y. Wang et al., 2021).

## 2.2. The aim of the chapter

Based on the short-read RNA-seq data indicating the possible existence of the short-length transcripts arising from the 3' region of *CACNAIS* and *CACNAIF* in the human brain, this chapter aims to detect and characterize these transcripts. It is hypothesized that the short-length transcripts would be present in the human brain and may have multiple transcript isoforms. This was achieved by combining PCR amplification and long-read nanopore sequencing (Figure 2.1) (N. A. L. Hall et al., 2021). The transcript isoform profiles were compared across brain regions.



*Figure 2.1 Overview of the PCR-targeted long-read nanopore sequencing method. The method was used to detect and characterise the short-length transcripts arising from the 3' regions of *CACNAIS* and *CACNAIF* in human brain.*

## 2.3. Methods

### 2.3.1. Human Brain Samples

High-quality human post-mortem brain samples were obtained via a collaboration with Professor Daniel Weinberger and colleagues at LIBD. Samples from eight brain regions, each from six healthy control individuals (Table 2.1), were utilized. The eight brain regions were: striatum, thalamus, parietal cortex, cingulate, dorsolateral prefrontal cortex (DL-PFC), superior temporal cortex, cerebellum (CBLM), and occipital cortex.

*Table 2.1. Demographics of the six healthy control individuals*

(F: female; M: male; AS: Asian; AA: African American; CAUC: Caucasian)

Individual ID	Sex	Race	Age of death (years)	Post-mortem intervals (hours)
5238	M	AA	37.16	10.5
5244	M	AA	21.2	28
5298	M	AA	50.34	12.5
5346	F	AS	25.11	24.5
5579	F	CAUC	41.85	13
5717	M	CAUC	51.88	35.5

### 2.3.2. RNA extraction from the postmortem human brain regions

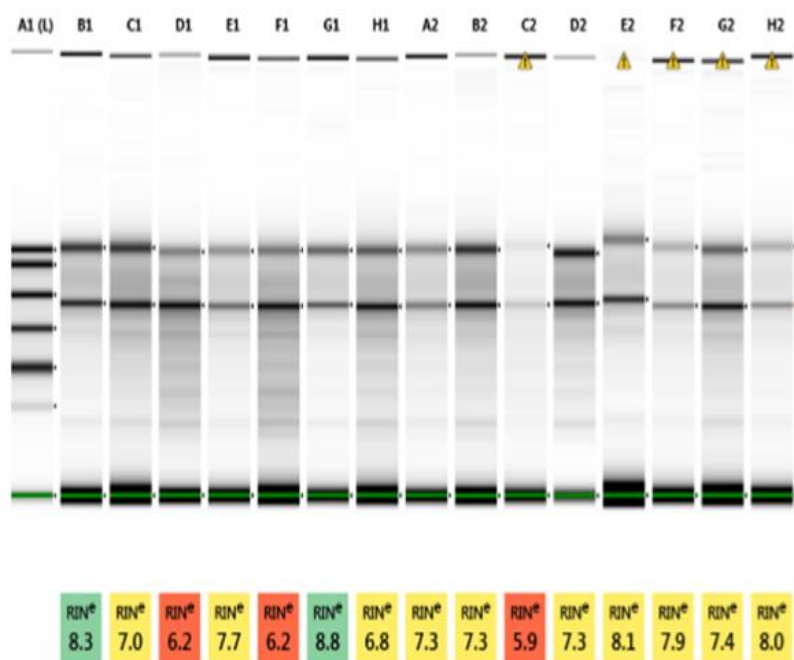
RNA was extracted from the postmortem human brain samples using the RNeasy Lipid Tissue Mini Kit (Qiagen, UK; 70804). Brain tissues were disrupted with the TissueLyserLT (Qiagen, UK; 85600) using RNase-free 5 mm stainless steel beads (Qiagen, UK; 69989) in QIAzol Lysis Reagent (Qiagen, UK; 79306). 200  $\mu$ l of chloroform was then added to each sample and shaken vigorously. The samples were incubated at room temperature for 3 minutes before being centrifuged at 12,000 x g for 15 minutes at 4°C. After centrifugation, the supernatants were carefully transferred to new 1.5 ml Eppendorf tubes. 500  $\mu$ l of 70% ethanol was added and

vortexed. 500  $\mu$ l of the samples were loaded onto RNeasy Mini spin columns and centrifuged for 15 seconds at 10,000 x g. This step was repeated until all of each sample had been loaded onto the columns. Then, 700  $\mu$ l of buffer RW1 was added to the RNeasy columns and centrifuged at 10,000 x g for 15 seconds to wash the column. After discarding the flow through, 500  $\mu$ l of buffer RPE was added to the column and centrifuged at 10,000 x g for 15 seconds. Once again, the flow through was discarded and 500  $\mu$ l of buffer RPE was added to the column and centrifuged for 2 minutes at 10,000 x g to wash the membrane. The columns were transferred to new 2 ml tubes and centrifuged at 13.3 x g for 1 minute to ensure no buffer or flow-through were carried over. The columns were transferred to new 1.5 ml Eppendorf tubes and 40  $\mu$ l of RNase-free water was added directly to the membrane. The columns were centrifuged at 10,000 x g for 1 minute and the eluate collected.

### **2.3.3. Assessment of RNA quality and concentration**

The quality of the extracted RNA was examined using a TapeStation 4200 (Agilent, UK) and quantified as RNA Integrity Number (RIN). RNA ScreenTape sample buffer (Agilent, UK; 0006407218) and RNA ScreenTape (Agilent, UK; 0201299-47) were incubated at room temperature for 30 minutes. 5  $\mu$ l of buffer and 1  $\mu$ l of ladder were mixed to 1  $\mu$ l of sample. The samples were incubated at 72°C for 3 minutes and put on ice for 2 minutes. Then, the samples were loaded onto the TapeStation (Figure 2.2; see appendix 1 for quality measurements of each RNA sample). The quality of RNA samples was indicated by RIN scores, which reflect the ratio of ribosomal RNAs (18s and 28s) determined by the result of microcapillary electrophoretic RNA separation (Schroeder et al., 2006). RNA samples with RIN scores above 6.5 indicate good quality RNA. RNA samples with RIN scores less than 6 are typically defined as low-quality, which are omitted for the sequencing as low-quality RNAs contain fragmented transcripts rather than

full-length transcripts (Kvastad et al., 2021). This can affect the sequencing results due to issues such as uneven gene coverage (Kukurba & Montgomery, 2015; Romero et al., 2014). Nevertheless, for the experiments in this chapter, RNA samples with low RIN scores were not discarded as one study showed that they could identify gene expression from RNA samples with an RIN score of around 4 (Romero et al., 2014). Moreover, library preparation methods for long-read nanopore sequencing were developed to reduce the effect of RNA degradation (Stark et al., 2019).



Default image (Contrast 50%), Image is Scaled to Simple, Image is Scaled to view larger Molecular Weight range

Figure 2.2. An example of the integrity of RNA extracted from each sample. The higher the RIN, the better the RNA integrity. The quality of extracted RNA is assessed using TapeStation 4200 (Agilent, UK). A1 is ladder, and from B1 to H2 are different samples. For the full data, see Appendix 1. RIN scores are colour coded: green is for RIN scores above 8.0; yellow is for RIN scores above 6.5; and red is for RIN scores below 6.5.

The concentration of extracted RNA was measured using the Qubit RNA Broad Range assay (ThermoFisher, UK; Q10210). A 1:200 working solution of reagent and buffer was mixed, to a volume sufficient for each sample and the two standards. Then, for each standard, 10  $\mu\text{l}$  was added to 190  $\mu\text{l}$  of working solution, and, for each sample, 1  $\mu\text{l}$  was added to 199  $\mu\text{l}$  of working solution. The mixtures were vortexed and incubated at room temperature for 2 minutes, then the concentration was measured using the Qubit (see appendix 2 for the concentration of each RNA sample). After the measurement, each RNA sample was diluted to 100 ng/ $\mu\text{l}$ .

#### 2.3.4. Reverse Transcription of RNA to produce cDNA

Reverse transcription (RT) was performed with GoScript™ Reverse Transcriptase (Promega, UK), using oligo dT primers. For each RNA sample, 1  $\mu\text{l}$  of oligo dT primer was added to 4  $\mu\text{l}$  of RNA. The samples were incubated at 70°C for 5 minutes, and chilled in ice water for 5 minutes. Then, RT mix (Table 2.2) was prepared. For each sample, 15  $\mu\text{l}$  of RT mix was added to 5  $\mu\text{l}$  of RNA-oligo dT primer. The samples were incubated at 25°C for 5 minutes, 42°C for 1 hour, and 70°C for 15 minutes.

*Table 2.2. Table of the reverse transcriptase (RT) mix to obtain cDNA from RNA sample extracted from the postmortem human brain*

	Volume per RNA sample ( $\mu\text{l}$ )
Nuclease Free (NF) Water	7.55
GoScript 5X Reaction Buffer	4
MgCl <sub>2</sub>	1.2
PCR Nucleotide Mix	0.75
Recombinant RNasin	0.50
GoScript Reverse Transcriptase	1
<i>Total volume</i>	<i>15</i>

### 2.3.5. Designing primers

Primers were designed for both the full-length *CACNA1S* and *CACNA1F* transcripts, and the predicted short-length transcripts from the 3' regions of *CACNA1S* and *CACNA1F*, with primers for the short-length sequences targeting sequences in the 5'-most regions to which short-read fragments were mapped in the LIBD and GTEx datasets (Table 2.3). PCR primers were located just after the predicted start sites and in the 3' untranslated regions after the stop codons. Barcode adaptor sequences were added to the 5' end of each PCR primer to allow for the later addition of unique barcode sequences to each sample. This allows for multiplexing and pooling of samples into a single nanopore sequencing run. Primer design was performed using Primer3 plus (Source code is available at <http://sourceforge.net/projects/primer3/>) (Untergasser et al., 2007) and A Plasmid Editor (ApE; RRID:SCR\_014266) (Davis & Jorgensen, 2022), and primers were checked using “In silico PCR” from the UCSC Genome Browser (<http://genome.cse.ucsc.edu/>) (Kent et al., 2002).

*Table 2.3. The primer sequences used for PCR amplification of the target transcripts. The barcode adaptor sequences are in lowercase and the gene-specific sequences are in uppercase.*

Target transcript	Sequence (5' - 3')
Full-length <i>CACNA1S</i>	F: tttctgttggtgctgatattgcAGAAAGCCAGATTCCCAGGG R:acttgctgctgctctatcttcCATCTAGCTGCTGAGAGGGAG
Short-length transcript from the 3' region of <i>CACNA1S</i>	F: tttctgttggtgctgatattgcGTACACAGCGACTTCCACTTC R:acttgctgctgctctatcttcCATCTAGCTGCTGAGAGGGAG
Full-length <i>CACNA1F</i>	F: tttctgttggtgctgatattgcCAGATGGCCCTTCAATCTCG R: acttgctgctgctctatcttcGGGCAGGAGGTTTATTGAGC
Short-length transcript from the 3' region of <i>CACNA1F</i>	F: tttctgttggtgctgatattgcGCGGATGAAACAGAAGCTGC R: acttgctgctgctctatcttcGGGCAGGAGGTTTATTGAGC

### 2.3.6. PCR amplification of each target transcript from human brain cDNA

Full-length *CACNA1S* and *CACNA1F*, and the short-length transcripts from the 3' region of *CACNA1S* and *CACNA1F* were amplified using PrimeSTAR GXL DNA polymerase (Takara, UK; R050B). Specifically, the transcripts were amplified from cDNA equating to 100 ng RNA template from the 48 post-mortem human brain samples. The reagents were mixed as specified in Table 2.4. Amplification was performed with the following thermal cycle, for either 30 or 35 cycles: 98°C for 10 seconds, annealing temperature (specific to each primer pair) for 15 seconds, and 68°C for 1 minute per 1 kb of target product length (Table 2.5).

*Table 2.4. Table of PCR mix to amplify gene of interest from the human brain*

	Per sample (µl)
Nuclease Free water	28
5x PrimerSTAR GXL Buffer	10
dNTP Mix (2.5mM each)	4
PrimeSTAR GXL DNA Polymerase	1
Forward Primer	1
Reverse Primer	1
cDNA (nuclease-free [NF] water for negative control reaction)	5
<i>Total</i>	<i>50</i>

*Table 2.5. Table of PCR amplification thermal cycle, annealing temperature and extension time*

Transcript to amplify	Thermal cycle	Annealing temperature	Extension time
Full-length <i>CACNA1S</i>	35	53°C	6 minutes
Full-length <i>CACNA1F</i>	30	57°C	6 minutes
Short-length transcript from the 3' region of <i>CACNA1S</i>	35	61°C	3 minutes
Short-length transcript from the 3' region of <i>CACNA1F</i>	30	57°C	1.5 minutes



### **2.3.7. Gel electrophoresis**

To optimally visualize the products of PCR amplification, a 1% of agarose gel was prepared in 100 ml 1 x TAE with 10  $\mu$ l of ethidium bromide (5 mg/ml) and visualized under UV light. However, when samples were later to be extracted from the gel, the ethidium bromide was replaced with 10  $\mu$ l of GelGreen (Biotium, UK; SCT125) and samples were visualized using standard LED light, to minimize DNA damage. For the ladder, 0.5  $\mu$ l of 1 kb DNA ladder (New England Biolabs, UK; N3232S) and 1  $\mu$ l of gel loading dye, purple (6x) (New England Biolabs, UK; B7025S) and 4.5  $\mu$ l of NF water were mixed and loaded onto the gel. 5  $\mu$ l of dye was mixed with 25  $\mu$ l of PCR product before loading onto the gel. All gels were run at 140 V for 40 minutes.

### **2.3.8. Gel extraction**

For the gel extractions, the Monarch DNA gel extraction kit (New England BioLabs, UK; T1020L) was used. The DNA bands were excised and transferred to 1.5 mL Eppendorf tubes, then weighed. 4 volumes of Gel Dissolving Buffer were added to the gel slice, followed by incubation at 55°C. Each dissolved sample was loaded onto a column and centrifuged for 1 minute at 16,000 x g (~13,000 RPM) then the flow-through was discarded. 200  $\mu$ l DNA wash buffer was added and centrifuged for 1 minute at 16,000 x g (~13,000 RPM). This wash step was repeated. Finally, the sample was eluted by adding 11  $\mu$ l of DNA elution buffer to the column membrane, incubated for 1 minute, and centrifuged for 1 minute at 16,000 x g.

### **2.3.9. Barcoding PCR to add the barcode sequences**

Each sample was barcoded via PCR amplification before being sequenced using barcodes from the Oxford Nanopore Technologies (ONT) PCR Barcoding Kit (EXP-PBC001) and PrimeSTAR GXL DNA polymerase (Takara, UK; R050B). PCR reagent mixes (Table 2.6) were set up in a total reaction volume of 25  $\mu$ l. PCR thermal cycle was, for 15 cycles: 98°C for

10 seconds, 62°C for 15 seconds, 68°C for 1 minute per 1 kb of target transcript length. The extension time for the short-length transcript from the 3' region of *CACNA1S* was 3 minutes and for the short-length transcript from the 3' region of *CACNA1F* was 1.5 minutes.

*Table 2.6. The barcoding PCR reagent mix to attach a barcode. The barcode is unique to each sample, so they can be pooled for a single sequencing run.*

	Per sample (µl)
Nuclease Free water	7
5x PrimeSTAR GXL buffer	5
dNTP mix (2.5 mM each)	2
PrimeSTAR GXL DNA polymerase	1
Nanopore barcode	1
Gel extracted cDNA	9
<i>Total volume</i>	<i>25</i>

### **2.3.10. PCR purification**

After the barcoding PCR, the Monarch PCR/DNA clean up kit (New England BioLabs, UK; T1030L) was used to purify the products. First, the samples were diluted with DNA Cleanup Binding Buffer at 5:1 ratio for the short-length *CACNA1F* transcript and at 2:1 ratio for the short-length *CACNA1S* transcript. The mixed samples were loaded onto columns and centrifuged at 16,000 x g for 1 minute and the flow-through was discarded. Then the samples were washed with 200 µl of DNA Wash Buffer and centrifuged at 16,000 x g for 1 minute. This wash step was repeated. The sample was eluted by adding 20 µl of DNA elution buffer to the column membrane, incubated for 1 minute, and centrifuged for 1 minute at 16,000 x g.

### **2.3.11. Measuring the concentration and quality of purified, barcoded PCR products**

The concentration of purified, barcoded PCR products was measured using the Qubit dsDNA Broad Range assay kit (Invitrogen, UK; 2295040). A 1:200 working solution of reagent and buffer were mixed to a volume sufficient for all samples and the two standards. Then, for each

standard, 10  $\mu$ l was added to 190  $\mu$ l of working solution, and, for each sample, 1  $\mu$ l was added to 199  $\mu$ l of working solution. The mixtures were vortexed and incubated at room temperature for 2 minutes, then the concentration was measured using the Qubit. The concentration of each sample is displayed in the Appendix 3 and Appendix 4.

The quality of PCR amplified samples was assessed by examining the size of the products using the Agilent Genomic DNA ScreenTape Assay for 4200 TapeStation System (Agilent, UK; 0006558133). Genomic DNA Reagents (Agilent, UK; 5067-5366) were equilibrated at room temperature for 30 minutes. Genomic DNA ScreenTape (Agilent, UK; 0201596-39) was flicked and loaded into the TapeStation machine. 10  $\mu$ l of Genomic DNA Sample Buffer was added to each of 1  $\mu$ l of Genomic DNA Ladder and 1  $\mu$ l of each sample. The samples were vortexed for 1 minute at 2000 rpm. It was used to examine the quality of purified, barcoded PCR products before the sequencing.

### **2.3.12. Sequencing library preparation**

After checking the quality of the purified barcoded PCR products, all the samples were pooled to a total 1000ng. Based on the concentration of each barcoded sample, the amount of sample input was calculated for each sample individually. This allows for the same amount of PCR product to be sequenced across samples. Then, the pooled samples were purified with Mag-Bind® TotalPure NGS (Omega Bio-Tek, USA) DNA-binding magnetic beads to remove low molecular weight DNA (e.g, primers, sheared amplicons) so that only molecules of the desired size are obtained and sequenced. The retained DNA was resuspended in 47  $\mu$ l NF water. For the long-read nanopore sequencing library preparation, the SQK-LSK109 (ONT, UK) ligation sequencing kit was used.

First, 1  $\mu$ l of DNA CS (a 3.6 kb DNA spike-in used for quality control), 1000 ng of DNA in 47  $\mu$ l NF water, 3.5  $\mu$ l of NEBNext FFPE DNA Repair Buffer, 2  $\mu$ l of NEBNext FFPE DNA Repair Mix, 3.5  $\mu$ l of Ultra II End-prep reaction buffer, and 3  $\mu$ l of Ultra II End-prep enzyme mix were mixed and incubated at 20°C for 5 minutes, then 65°C for 5 minutes. The solution was then purified with 1.8X Mag-Bind® TotalPure NGS (Omega Bio-Tek, USA) to remove the enzymes and buffers from the end repair step. The beads were added to the end-prepped reaction and incubated on a HulaMixer for 5 minutes at room temperature. Then, the sample was pelleted on a magnetic rack until the supernatant was clear. While the tube was on the magnetic rack, the supernatant was pipetted off. Whilst the tube was still on the magnetic rack, 200  $\mu$ l of 70% ethanol in nuclease-free water was added and then pipetted off. The 70% ethanol wash was repeated. The pellet was allowed to air dry, until dry but not cracked (~30 seconds). The tube was removed from the magnetic rack and the pellet was resuspended in 61  $\mu$ l nuclease-free water and incubated for 2 minutes at room temperature. The tube was put on the magnetic rack again to pellet the beads. The eluate was removed into a clean Lo-Bind 1.5 ml tube. Then, the Qubit Broad Range dsDNA concentration assay (Qiagen, UK) was used to check for the presence of sufficient product.

60  $\mu$ l of product, 25  $\mu$ l of Ligation Buffer, 10  $\mu$ l of NEBNext Quick T4 DNA Ligase, and 5  $\mu$ l of Adapter Mix were mixed and incubated for 10 minutes at room temperature. 1.8X Mag-Bind® TotalPure NGS (Omega Bio-Tek, USA) magnetic beads were used to purify the adapter-ligated samples. The beads were added to the reaction product and incubated on a HulaMixer for 5 minutes at room temperature. The beads were pelleted on a magnetic rack and the supernatant was discarded. The bead pellet was washed by resuspending beads in 250  $\mu$ l of Short Fragment Buffer and then pelleted on the magnet and the supernatant discarded. This wash step was repeated. Once the bead pellet was washed twice and then air dried, the pellet was

resuspended in 15  $\mu$ l of Elution Buffer and incubated for 10 minutes at room temperature. The tube was put on the magnetic rack until the eluate become clear and then 15  $\mu$ l of eluate, containing the DNA library, was retained into a new Lo-Bind 1.5 ml tube. Finally, the Qubit Broad Range dsDNA assay (Qubit, UK) was used to quantify the concentration of eluate sample. 50 fmol of DNA in a volume of 12  $\mu$ l was required for sequencing.

The quality of the flow cell was checked to ensure that the flow cell had more than 800 active pores. To prime the flow cell before sequencing, priming mix was made by adding 30  $\mu$ l of Flush Tether to a tube of Flush Buffer. 30  $\mu$ l of buffer inside the flow cell was withdrawn from the priming port of the flow cell. Then 800  $\mu$ l of the priming mix was loaded onto the flow cell through the priming port without introducing any air bubbles. The library was prepared for loading. 37.5  $\mu$ l of sequencing buffer, 25.5  $\mu$ l of Loading Beads, and 12  $\mu$ l of 50 fmol DNA library were mixed. 5 minutes after the priming mix was loaded onto the flow cell, another 200  $\mu$ l of priming mix was loaded via the priming port. Then, 75  $\mu$ l of the prepared DNA library was added to the flow cell via SpotON sample port in a dropwise manner. The SpotON sample port was covered, and the priming port was also closed.

Once the prepared library was loaded onto the flow cell, the sequencing program was set up and run via the MinKNOW software (ONT, UK) for 72 hours. After the sequencing run finished, Guppy software v6.4.2 (ONT, UK), which has Oxford Nanopore's basecalling algorithm and other bioinformatic features, was used to base-call the sequenced reads and de-multiplex the different barcoded samples.

### **2.3.13. Analyzing long-read nanopore sequencing data using TAQLoRe program**

The custom-designed TAQLoRe program (developed by WH and TW, Earlham Institute, available from <https://github.com/twrzes/TAQLoRe> with full documentation at

<https://taqlore.readthedocs.io/en/latest/>) was used to identify and quantify the transcript isoforms. TAQLoRe aims to identify novel exons and full-length isoforms, thereby improving the existing transcriptome annotation data. It also quantifies transcripts, outputting normalized read counts per transcript isoform. The program uses the trimmed mean of M-values normalization method (TMM) to normalize the read counts across the samples (Robinson & Oshlack, 2010).

The sequencing reads with an identified barcode were mapped to the transcriptome (GENCODE V35) and to the relevant *CACNAIS* or *CACNAIF* metagene with annotated exonic sequences using GMAP (Wu & Watanabe, 2005). From the reads aligned to the transcriptome, novel exons were identified by examining inserts of at least nine nucleotides at the exonic junctions between annotated exons. The novel exons should be located at least 200 nucleotides away from the existing exon junctions. Once potential novel exons were identified, the corresponding reads were mapped to the genome using LAST. In order for the novel exons to be included in the final annotation, at least 80% of nucleotides were required to have read coverage. Then, the novel exons were introduced to the *CACNAIS* or *CACNAIF* metagene, and reads were mapped again to the updated metagene for transcript annotation. Incomplete transcripts were collapsed into longer transcripts based on splicing pattern. Transcript isoforms with at least 100 reads in at least 2 libraries were retained for quantification.

Once the sequencing data were analyzed using TAQLoRe program, the highly expressed transcripts were examined using Excel 2008 and the protein that may be encoded by the sequenced coding transcripts were predicted by Uniprot (The UniProt Consortium, 2021).

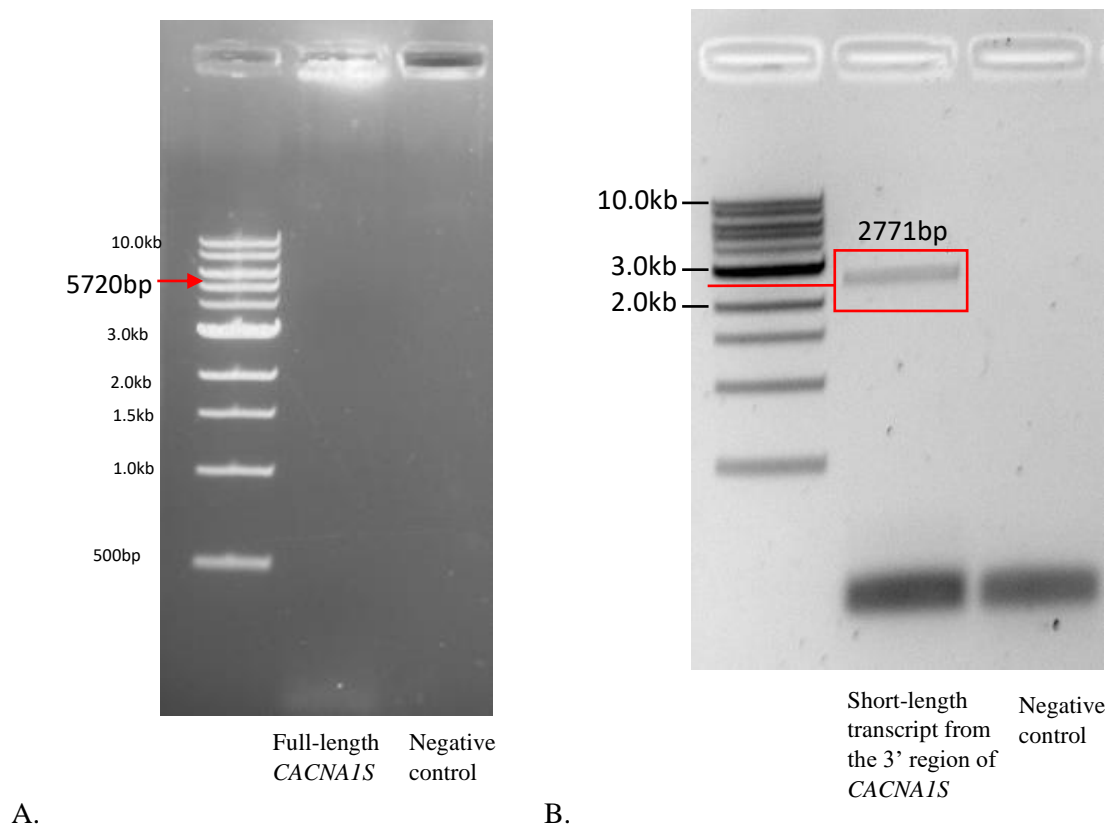
### **2.3.14. Examining the expression variations of the short-length transcripts from the 3' regions of *CACNA1S* and *CACNA1F* between brain regions and individuals**

Principal Component Analysis (PCA) was used to examine how variation in transcription isoform expression, using total number of reads per samples, varied between brain regions and individuals. The analysis was conducted in Python 3.7.6 via Visual Studio Code 1.74.1 using code developed by SMH, which is included in the appendix 5.

## **2.4. Results**

### **2.4.1. The short-length transcript from the 3' region of *CACNA1S* was detected in cDNA from human brain, whereas the full-length *CACNA1S* transcript was not detected**

Based on the short-read RNA-seq data analysis (see Figure 1.6A, 1.7A in chapter 1), I attempted to amplify both the full-length *CACNA1S* and the predicted short-length transcript from the 3' region of *CACNA1S* from the pooled cDNA of 48 postmortem human brain samples (see 2.3.1, 2.3.2, 2.3.3, 2.3.4, 2.3.5, 2.3.6 in Methods section). It was predicted that the short-length transcript of *CACNA1S*, but not the full-length *CACNA1S* transcript, would be detected in the human brain. A product with the same size of the predicted short-length transcript from the 3' region of *CACNA1S* was detected (2771bp) (Figure 2.3B; red box) and was sequenced using the long-read nanopore sequencing. However, the full-length *CACNA1S* transcript (5720bp) was not detected (Figure 2.3A; red arrow).



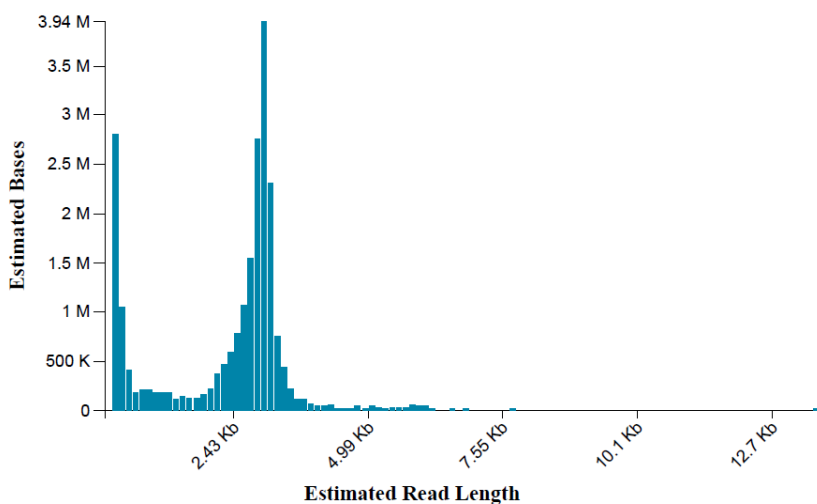
*Figure 2.3. Gel electrophoresis results of PCR amplification of full-length CACNA1S and the predicted short-length transcript from the 3' region of CACNA1S. A. PCR amplification of full-length CACNA1S (5720bp) from the human brain. The well labelled 'full length CACNA1S' contained the product of the PCR targeting the full-length CACNA1S with human brain cDNA template (5720 bp, red arrow) and the well labelled 'Negative control' contained the no template control reaction. There was no detectable amplification of full-length CACNA1S in human brain. B. Gel electrophoresis of the products of PCR amplification of short-length transcript from the 3' region of CACNA1S from human brain. Wells marked as 'Short-length transcript from the 3' region of CACNA1S' contains the products of PCRs targeting the predicted short-length transcript from the 3' region of CACNA1S with human brain cDNA template, whereas 'Negative control' wells contain the no template control reactions. There is an amplification product matching the predicted length of the short-length transcript from the 3' region of CACNA1S (2771 bp, red box) from the human brain cDNA.*



## 2.4.2. Long-read nanopore sequencing of the short-length transcript from the 3' region of *CACNAIS* in the human brain

### 2.4.2.1. Long-read nanopore sequencing

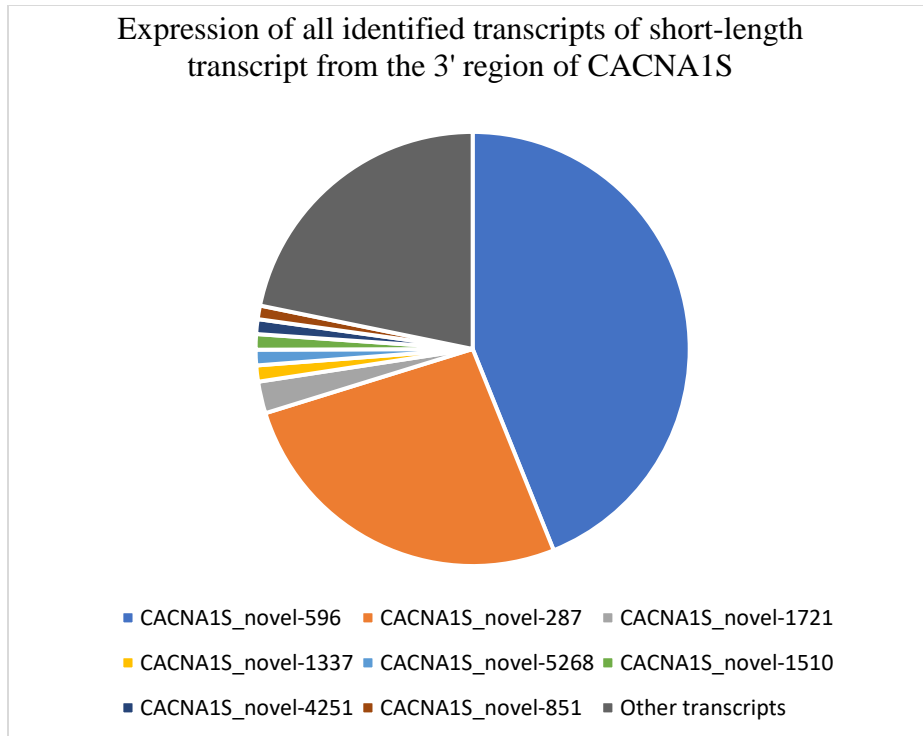
The short-length transcript predicted to arise from the 3' region of *CACNAIS* (Figure 2.3B) was sequenced with the long-read nanopore sequencing to verify the identity of amplified products. When the short-length transcript in the 3' region of *CACNAIS* was sequenced, 8 samples (5717 CBLM, 5238 CBLM, 5579 CBLM, 5298 CBLM, 5244 occipital lobe, 5717 occipital lobe, 5298 thalamus, 5298 parietal lobe) were excluded as the transcript was not amplified from the samples (Table 2.1; see section 2.3.1), which was not associated with the RNA quality measured in Methods section 2.3.3. In the end, 40 out of 48 samples were used for the sequencing. A total of 4.3 million reads was obtained from the sequencing (Figure 2.4). The output read length matched the input RNA length. Transcripts were aligned to the short-length transcript from the 3' region of *CACNAIS*.



*Figure 2.4. Read length histogram after sequencing. The short-length transcript from the 3' region of CACNAIS, a 2.771kb target product, was sequenced for 72 hours. The majority of reads cluster around the predicted size.*

The long-read nanopore sequencing result confirmed the existence of the short-length transcript from the 3' region of *CACNA1S* in the human brain, suggested by the LIBD and GTEx datasets. The expression proportion of each sequenced transcript in the human brain were identified (Figure 2.5). Transcripts with less than 1% expression of the total expression level were grouped as 'Other transcripts'. Sequencing identified a total of 996 transcripts, which were sub-grouped into coding transcripts and non-coding transcripts. If the open reading frame of the transcripts (from the first start codon to the stop codon of the sequenced transcripts) is divisible by three, they were categorized as coding-transcripts whereas if the open reading frame of transcripts is not divisible by three, they were categorized as non-coding transcripts.

Among 996 transcripts, 338 transcripts were identified as coding transcripts. For the analyses in this thesis, only coding transcripts were used, therefore non-coding transcripts were excluded. From the 338 coding transcripts, transcripts with greater than 1% of the total expression level in the human brain were used for downstream analyses. These were *CACNA1S\_novel-596* and *CACNA1S\_novel-287* (Figure 2.6).



*Figure 2.5. A pie chart showing the proportion of expression of short-length transcript from the 3' region of CACNA1S in the human brain. The transcripts that constituted less than 1% expression of the total transcript expression in the human brain were grouped as "other transcripts". CACNA1S\_novel-596 (43.9%) was the most highly expressed transcript in the human brain, followed by CACNA1S\_novel-287 (26.3%), and CACNA1S\_novel-1721 (2.37%), CACNA1S\_novel-1337 (1.2%), CACNA1S\_novel-5268 (1.16%), CACNA1S\_novel-1510 (1.15%), CACNA1S\_novel-4251 (1.1%), and CACNA1S\_novel-851 (1.01%). Other transcripts contribute 21.79% of the total expression. Among the highly expressed transcripts, CACNA1S\_novel-1721, CACNA1S\_novel-5268, CACNA1S\_novel-1510, and CACNA1S\_novel-851 are non-coding transcripts.*

### 2.4.2.2. Highly expressed in-frame coding short-length transcripts from the 3' region of *CACNA1S* in the human brain were identified

The highly expressed in-frame coding transcripts in each brain region were identified. In all eight brain regions, *CACNA1S\_novel-596* (55.3% of the total expression of coding transcripts) and *CACNA1S\_novel-287* (33.1% of the total expression of coding transcripts) are the most highly expressed in all brain regions. Transcripts that contribute less than 1% of the total expression of coding transcripts are grouped as ‘other coding transcripts’ (Figure 2.6). As other transcripts contribute the same or less than 1% of the total expression in the human brain, only two transcripts, *CACNA1S\_novel-596* and *CACNA1S\_novel-287*, were focused for the analyses in this chapter and further analyses in chapter 4.

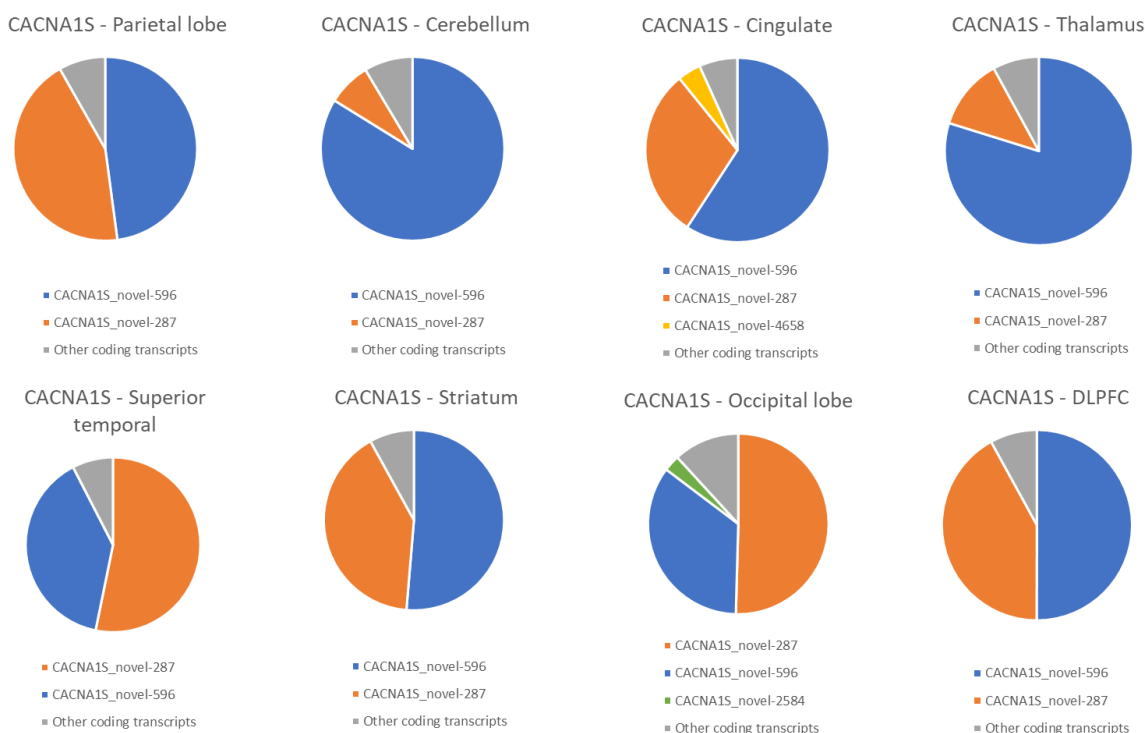
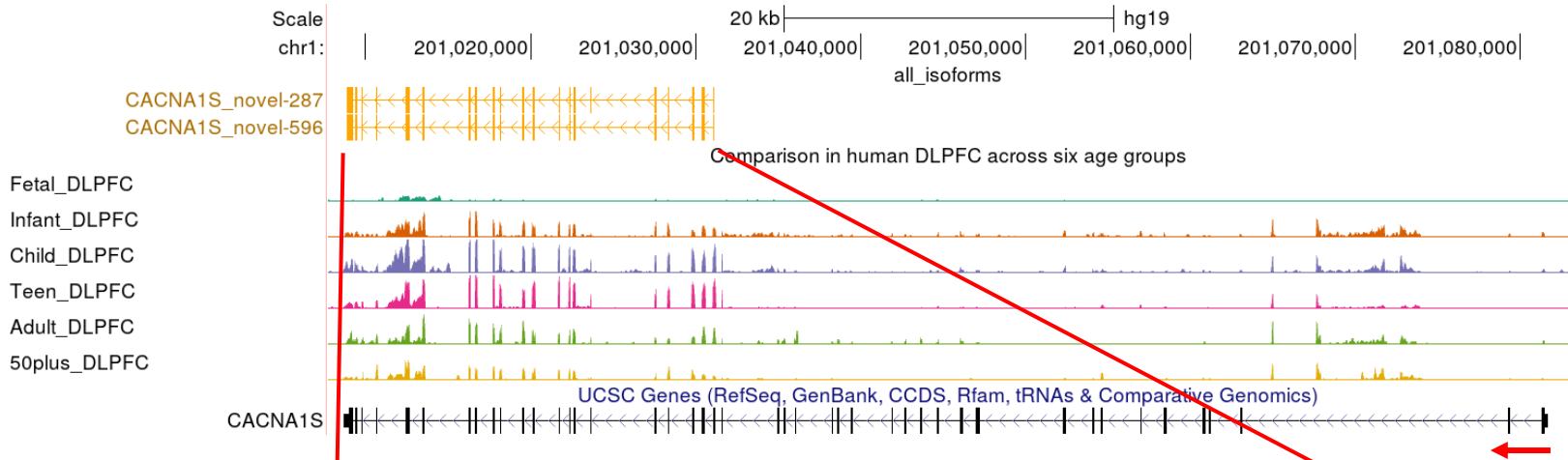


Figure 2.6. Proportion of high abundant (>1% of total expression of coding transcripts) coding short-length transcript from the 3' region of *CACNA1S* in each human brain region. Each transcript was

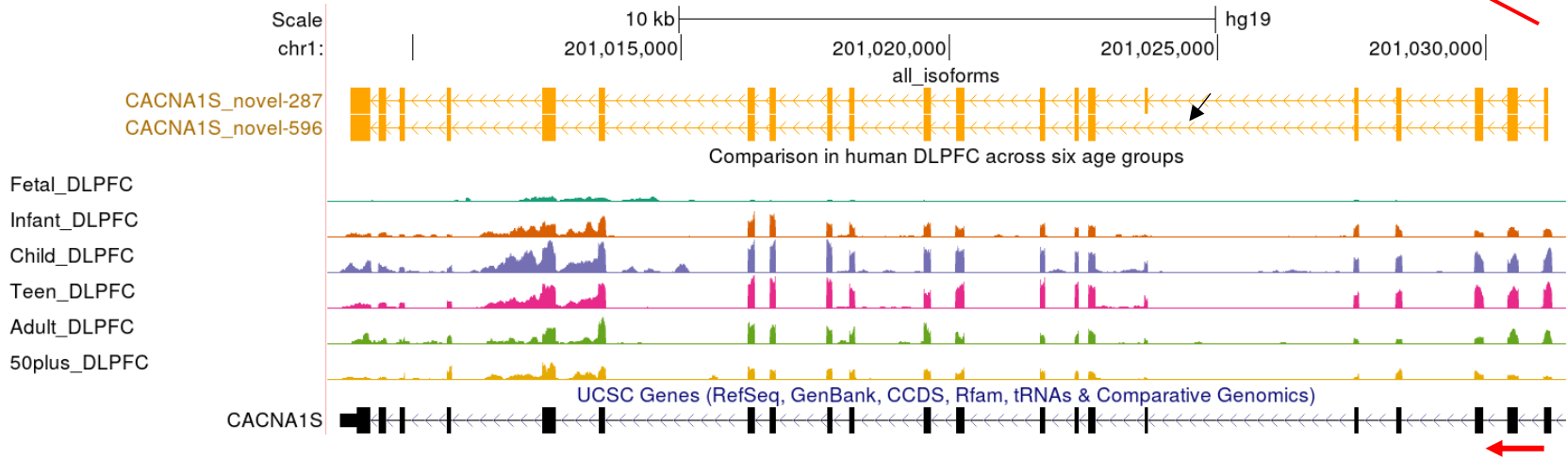
*color coded. Even though each transcript has a different proportion of expression in the brain region, CACNAIS\_novel-596 and CACNAIS\_novel-287 are the most expressed coding transcripts in all 8 different brain regions.*

The top 2 coding isoforms in the human brain are presented in the genome browser (hg19; <http://genome.ucsc.edu>; Figure 2.7). The short-length transcript from the 3' region of *CACNAIS* is comprised of 21 exons. These exons are known exons, meaning that they were already annotated in the *CACNAIS* metagene (Table 2.7.). No novel exons were identified, but exon skipping event was observed. *CACNAIS\_novel-596* transcript does not have exon 6 whereas *CACNAIS\_novel-287* has all the known exons matching to the sequences of the *CACNAIS* metagene. This suggests that the newly identified short-length transcript from the 3' region of *CACNAIS* in the human brain comprises of a pool of related but unique transcripts with exon skipping. Two transcripts encode different protein isoforms within voltage-dependent L-type calcium channel subunit alpha. *CACNAIS\_novel-287*, which has all the known exons, encode the part of the full-length Cav1.1 whereas *CACNAIS\_novel-596*, which has exon 6 skipped, encodes a protein with the entry name B1ALM2\_Human (Table 2.8.). The full protein sequences are available in the Appendix 6. Therefore, each transcript could possibly have distinct functions in the human brain.

A.



B.



*Figure 2.7. The top 2 most highly expressed in-frame short-length coding transcripts from the 3' region of CACNA1S in the human brain are shown in blue, mapped to the CACNA1S metagene (in black) displayed on the UCSC genome browser (hg19; <http://genome.ucsc.edu>). Colored tracks show LIBD short-read RNA Seq data. The orientation of the gene is from the right to the left (indicated with a red arrow). **A** shows the full CACNA1S locus. **B** shows a 'zoomed in' version focusing on the short-length coding transcript. Each transcript is an in-frame coding transcript. There were no identified novel exons from the long-read nanopore sequencing. However, one possible exon skipping event was observed, which is marked with black arrows.*

Table 2.7 Genomic location of each exonic region (hg19) from the long-read sequencing of short-length transcript in the 3' region of *CACNA1S* from the human brain.

Genomic location	Known exon or novel exon	Exon number in the full-length <i>CACNA1S</i>	Exon number in the short-length transcript in the 3' region of <i>CACNA1S</i>
Chr1:201,031,072–201,031,155	Known exon	Exon 24	Exon 1
Chr1:201,030,394–201,030,597	Known exon	Exon 25	Exon 2
Chr1:201,029,786–201,029,944	Known exon	Exon 26	Exon 3
Chr1:201,028,317–201,028,427	Known exon	Exon 27	Exon 4
Chr1:201,027,536–201,027,619	Known exon	Exon 28	Exon 5
Chr1:201,023,633–201,023,689	Known exon with possibility of exon skipping	Exon 29	Exon 6
Chr1:201,022,586–201,022,715	Known exon	Exon 30	Exon 7
Chr1:201,022,337–201,022,402	Known exon	Exon 31	Exon 8
Chr1:201,021,685–201,021,776	Known exon	Exon 32	Exon 9
Chr1:201,020,112–201,020,271	Known exon	Exon 33	Exon 10
Chr1:201,019,517–201,019,644	Known exon	Exon 34	Exon 11
Chr1:201,018,131–201,018,227	Known exon	Exon 35	Exon 12
Chr1:201,017,710–201,017,812	Known exon	Exon 36	Exon 13
Chr1:201,016,653–201,016,754	Known exon	Exon 37	Exon 14
Chr1:201,016,243–201,016,367	Known exon	Exon 38	Exon 15
Chr1:201,013,456–201,013,584	Known exon	Exon 39	Exon 16
Chr1:201,012,409–201,012,659	Known exon	Exon 40	Exon 17
Chr1:201,010,631–201,010,717	Known exon	Exon 41	Exon 18
Chr1:201,009,750–201,009,841	Known exon	Exon 42	Exon 19
Chr1:201,009,359–201,009,502	Known exon	Exon 43	Exon 20
Chr1:201,008,958–201,009,211	Known exon	Exon 44	Exon 21



‘Known exon’ means an exon that is aligned to metagene on GENCODE v35 of *CACNA1S* whereas ‘novel exon’ is an exon that was not previously annotated in the metagene of *CACNA1S* but newly identified from the long-read nanopore sequencing here. No novel exons were identified.

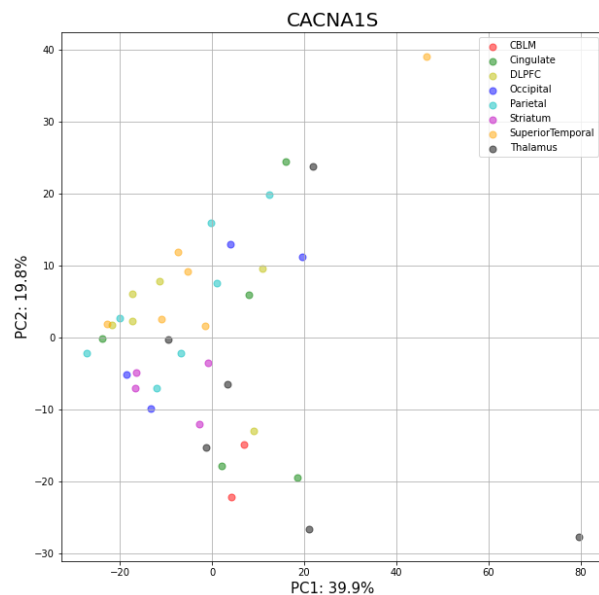
*Table 2.8. Protein predicted to be encoded by each transcript*

Transcript name	Entry	Entry Name	Protein names	Gene Names	Organism	Length
CACNA1S_novel-287	Q13698	CAC1S_HUMAN	Voltage-dependent L-type calcium channel subunit alpha-1S (Calcium channel, L type, alpha-1 polypeptide, isoform 3, skeletal muscle) (Voltage-gated calcium channel subunit alpha Cav1.1)	CACNA1S CACH1 CACN1 CACNL1A3	Homo sapiens (Human)	1004-1873
CACNA1S_novel-596	B1ALM3	B1ALM3_HUMAN	Voltage-dependent L-type calcium channel subunit alpha	CACNA1S	Homo sapiens (Human)	1004-1854

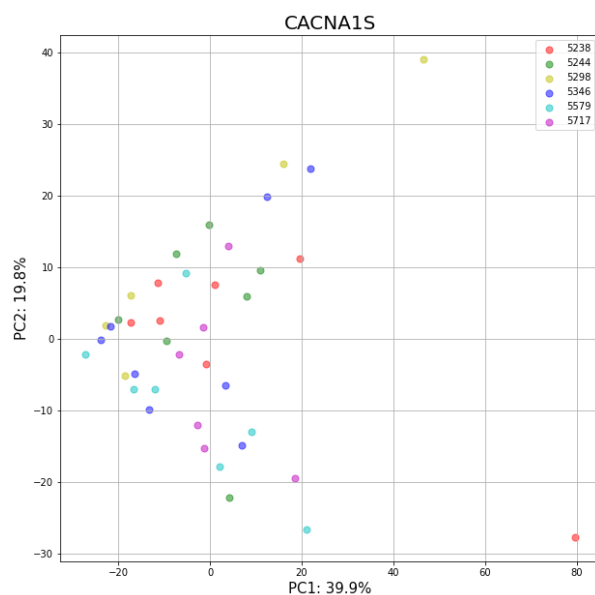
### 2.4.2.3. Examining regional and individual variation in the transcript profiles of short-length transcript from the 3' region of *CACNA1S*

PCA was used to examine how transcription isoform expression of short-length transcript from the 3' region of *CACNA1S* varied between brain regions and individuals, using total number of reads per samples, which each input of sample was standardized. PCA demonstrated no clear clustering of samples by either brain region (Figure 2.8A) or individual (Figure 2.8B).

**A.**



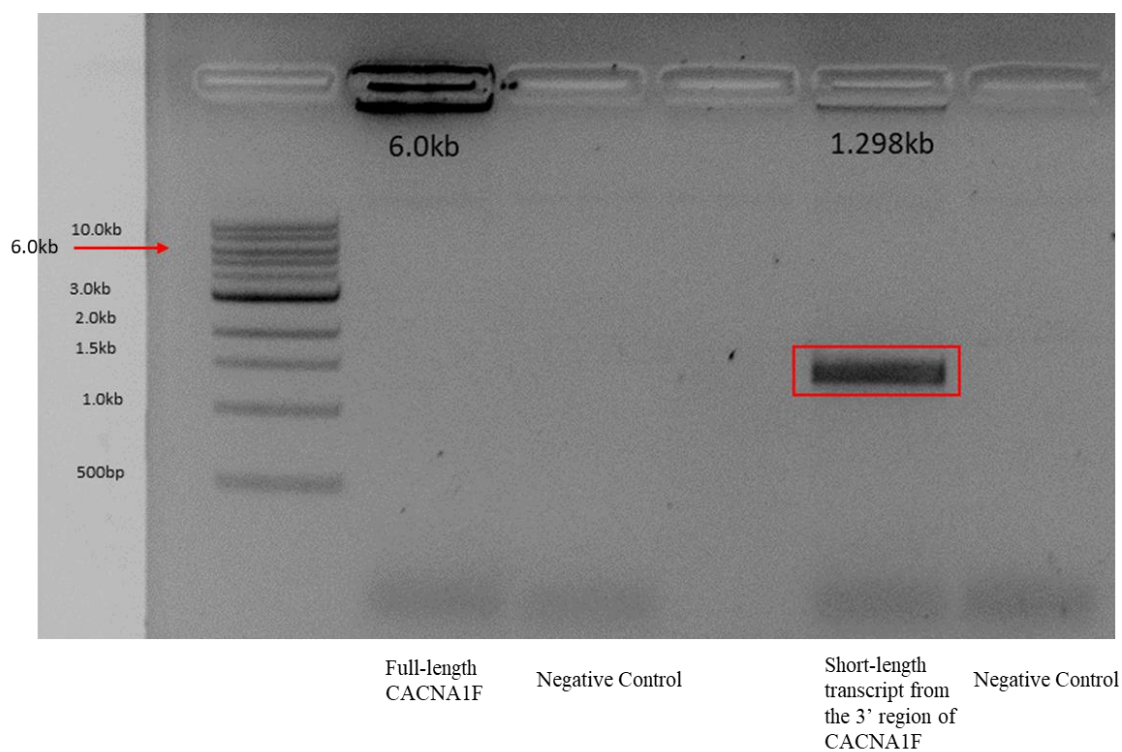
**B.**



*Figure 2.8. Principal Component Analysis (PCA) based on transcript expression. A PCA plot showing that expression of short-length transcript from the 3' region of *CACNA1S* does not evidence of clustering across the 8 human brain regions B PCA plot showing that transcript expression of short-length transcript in the 3' region does not show evidence of clustering across the 6 individuals.*

**2.4.3. The short-length transcript from the 3' region of *CACNA1F* was detected in cDNA from human brain, whereas the full-length *CACNA1F* transcript was not detected**

Based on the short-read RNA-seq data analysis (See Figure 1.6B, 1.7B in chapter 1), I attempted to amplify both the full-length *CACNA1F* and the predicted short-length transcript from the 3' region of *CACNA1F* from the pooled cDNA of 48 post-mortem human brain samples was attempted. A product with the same size of predicted short-length transcript from the 3' region of *CACNA1F* was detected (1298bp) (Figure 2.9; fifth well squared with red lines and labelled as 'Short-length transcript from the 3' region of *CACNA1F*') and was taken forward to be sequenced using the long-read nanopore sequencing. However, the full-length *CACNA1F* transcript (6 kb) was not detected (Figure 2.9; second well labelled as 'Full length *CACNA1F*').



*Figure 2.9. Gel electrophoresis result of PCR amplification of both full-length CACNA1F and predicted short-length transcript from the 3' region of CACNA1F from the postmortem human brain. A well, named 'Full length CACNA1F', has the PCR amplified product of the full-length CACNA1F with human brain cDNA template (6.0kb, red arrow) and the well labelled 'Negative control' contained the no template control reaction. There was no detectable amplification of full-length CACNA1F in the human brain. A well, named 'Short-length transcript from the 3' region of CACNA1F' contains PCR amplified product of short-length transcript from the 3' region of CACNA1F with human brain cDNA template (1.298kb, red box) whereas 'Negative Control' well has no template control reaction. There is an amplification product matching the predicted length of the short-length transcript from the 3' region of CACNA1F from the human brain cDNA.*

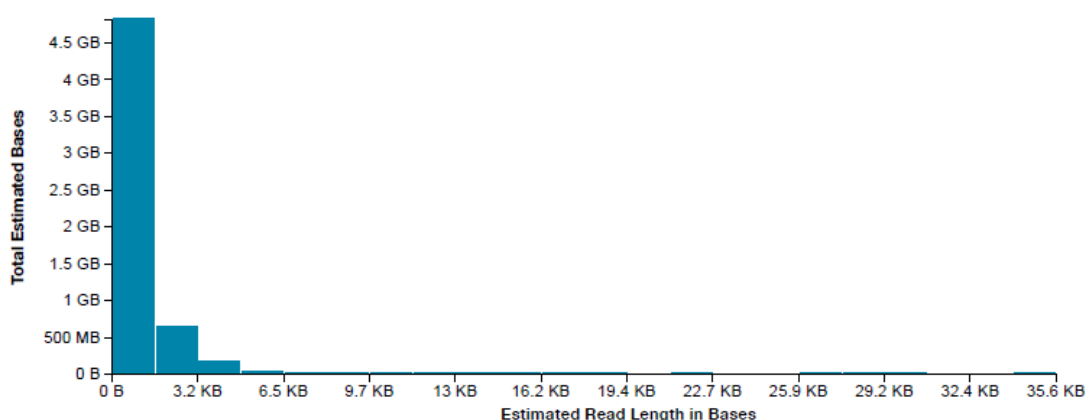
#### **2.4.4. Long-read nanopore sequencing of short-length transcript in the 3' region of CACNA1F in the human brain**

##### **2.4.4.1. Long-read nanopore sequencing**

The predicted presence of the short-length transcript from the 3' region of *CACNA1F*, identified from the PCR amplification (Figure 2.9), was sequenced with the long-read nanopore sequencing (Figure 2.1) to verify whether the identified PCR products are positively mapped to *CACNA1F* or not. When the short-length transcript from the 3' region of *CACNA1F* was sequenced, two samples (5244 thalamus, 5244 Striatum) were excluded as the transcript was not amplified (Table 2.1; see section 2.3.1). In the end, 46 out of 48 samples were used for the sequencing. A total 8.9 million reads were obtained from the sequencing (Figure 2.10). The output read lengths match the input RNA length which is 1.298kb.

## Read Length Histogram

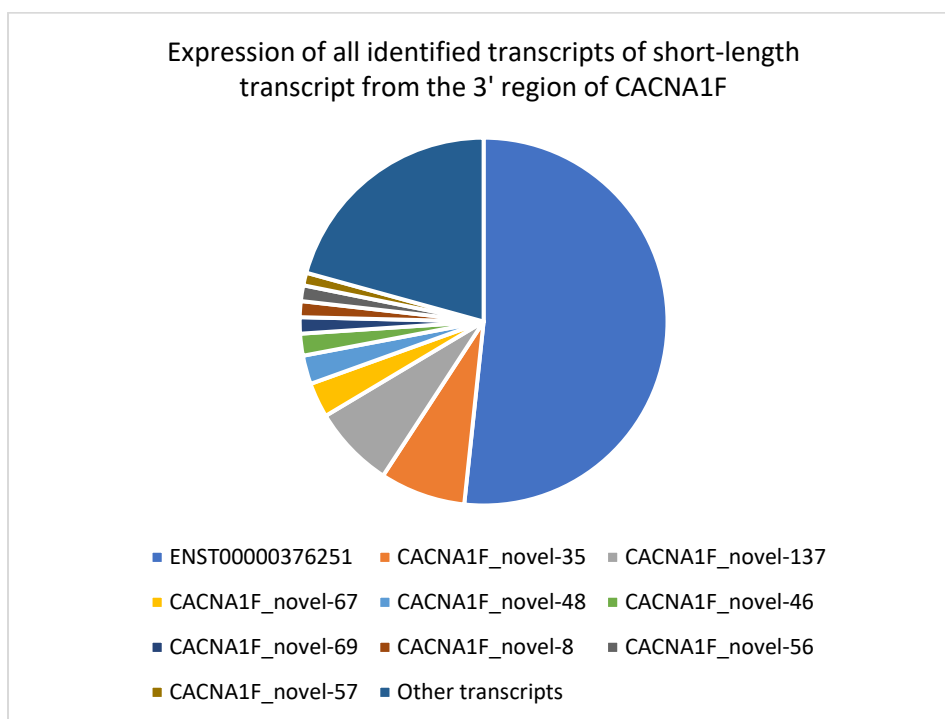
Summary read length distribution



*Figure 2.10. Read length histogram after sequencing short-length transcript from the 3' region of CACNA1F. The transcript is a 1.298kb and sequenced for 72 hours. There is a highest peak of the read length which matches the size of sequenced product. It was unfortunate that the scale of x-axis was not changeable, but it shows that the sequenced products is the same size as what was prepared to be sequenced.*

Long-read nanopore sequencing confirmed the existence of the short-length transcript from the 3' region of *CACNA1F* in the human brain, suggested by the LIBD and GTEx datasets. The expression proportion of each transcript in the human brain were identified (Figure 2.11). Transcripts with less than 1% expression of the total expression level were grouped as 'Other transcripts'. The long-read nanopore sequencing identified a total of 518 transcripts, which were sub-grouped into coding transcripts and non-coding transcripts. If the open reading frame of the transcripts (from the first start codon to the stop codon of the sequenced transcripts) is divisible by three, they were categorized as coding-transcripts whereas if the open reading frame of transcripts is not divisible by three, they were categorized as non-coding transcripts. Among the 518 transcripts, 170 transcripts were identified as coding transcripts. For the analyses in this thesis, only coding transcripts were used. Therefore, the non-coding transcripts were excluded from

further analyses. From the total 170 coding transcripts, the transcripts with more than 1% of the total expression level of the short-length transcript from the 3' region of *CACNA1F* in the human brain were used for the further analyses. These were ENST00000376251, *CACNA1F\_novel-55*, *CACNA1F\_novel-57*, and *CACNA1F\_novel-69*.



*Figure 2.11. A pie chart showing the proportion of expression of short-length transcripts from the 3' region of *CACNA1F* in the human brain. The transcripts that accounted for less than 1% of total expression in the human brain were grouped as "other transcripts". The transcript called ENST00000376251 was most highly expressed (51%) among the short-length transcripts in the 3' region of *CACNA1F* in the human brain, followed by *CACNA1F\_novel-35* (7.5%) and *CACNA1F\_novel-137* (7.2%). *CACNA1F\_novel-67* (3.07%), *CACNA1F\_novel-48* (2.52%), *CACNA1F\_novel-46* (1.93%), *CACNA1F\_novel-69* (1.43%), *CACNA1F\_novel-8* (1.40%), *CACNA1F\_novel-56* (1.39%), and *CACNA1F\_novel-57* (1.11%) also contributed more than 1% of the total expression of transcripts in the human brain. However, *CACNA1F\_novel-35*, *CACNA1F\_novel-137*, *CACNA1F\_novel-67*, *CACNA1F\_novel-48*,*

*CACNA1F\_novel-46*, *CACNA1F\_novel-56* and *CACNA1F\_novel-8* were identified as non-coding transcripts.

#### 2.4.4.2. Highly expressed in-frame coding short-length transcripts from the 3' region of *CACNA1F* in the human brain were identified

The long-read nanopore sequencing confirmed the existence of short-length transcripts from the 3' region of *CACNA1F* in the human brain as suggested by the LIBD and GTEx datasets. Among the sequenced in-frame coding transcripts of short-length transcript from the 3' region of *CACNA1F*, highly expressed coding transcripts (>1% of the total expression of coding transcripts) in each brain region were identified (Figure 2.12). In all eight brain regions, ENST00000376251 (83.32% of expression of total expression of coding transcripts) was the most expressed which was followed by *CACNA1F\_novel-69* (2.31%) and *CACNA1F\_novel-57* (1.79%) (Figure 2.12).

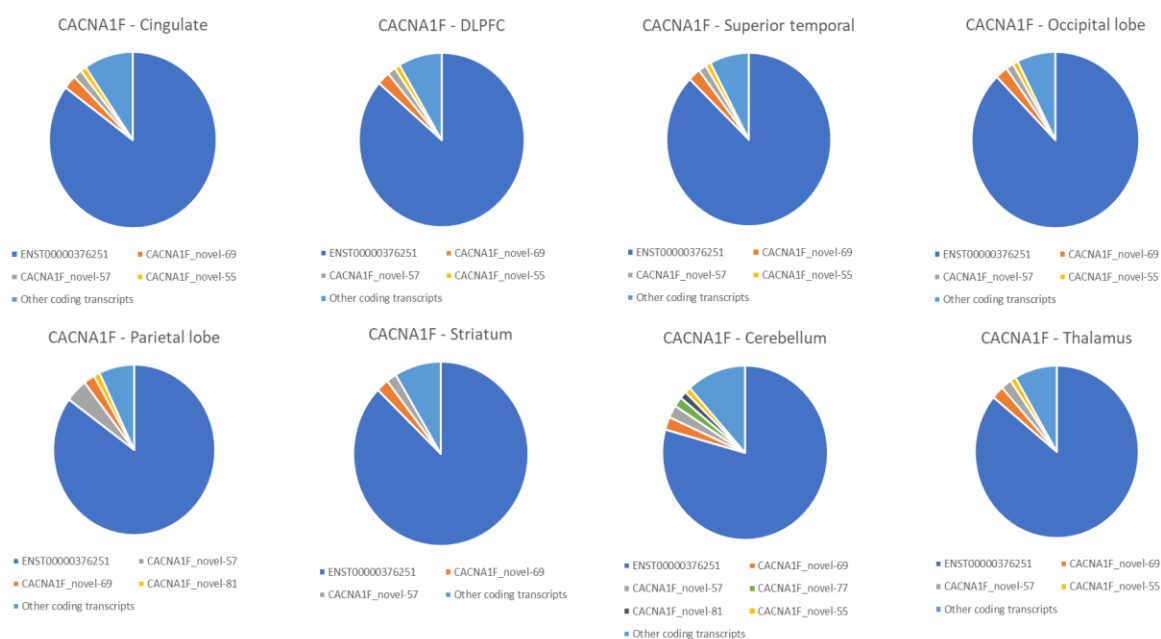


Figure 2.12. Proportion of highly abundant (>1% of total expression level of coding transcripts) coding short-length transcripts from the 3' region of *CACNA1F* reads in each human brain region. Each transcript

was color coded. About 90% of coding transcript expressions in the human brain was identified as the transcript ENST000003762.

ENST00000376251 is an isoform of the short-length transcript in the 3' region of *CACNA1F* that 100% aligned to the already annotated full-length *CACNA1F* transcript. As other transcripts are less than 1% read coverage, the four transcripts ENST00000376251, *CACNA1F\_novel-69*, *CACNA1F\_novel-57*, and *CACNA1F\_novel-55*, were the focus of this chapter, and were further analysed in Chapter 4.

The top coding isoforms in the human brain are presented in genome browser (hg19; <http://genome.ucsc.edu>; Figure 2.13). The short-length transcript from the 3' region of *CACNA1F* is comprised of 9 exons. These exons are known exons, meaning that they exist in *CACNA1F* metagene (Table 2.9). No novel exons were identified, but one possible exon skipping event was observed. *CACNA1F\_novel-57* transcript does not have exon 8 while ENST00000376251 has all the known exons matching to the sequences of *CACNA1F* metagene (Figure 2.13). This suggests that the newly identified short-length transcript from the 3' region of *CACNA1F* in the human brain comprises of diverse transcripts with exon skipping. Among identified transcripts, *CACNA1F\_novel-57* encodes different protein isoforms than other transcripts (Table 2.10). ENST00000376251, *CACNA1F\_novel-55*, and *CACNA1F\_novel-69* encode part of the protein isoform 2 or 3 of *CACNA1F*, while *CACNA1F\_novel-57* encodes part of protein isoform 6 of *CACNA1F* (Table 2.10). The protein sequences can be found in the Appendix 6. Therefore, each transcript, especially *CACNA1F\_novel-57*, could possibly have distinctive functions in the human brain.

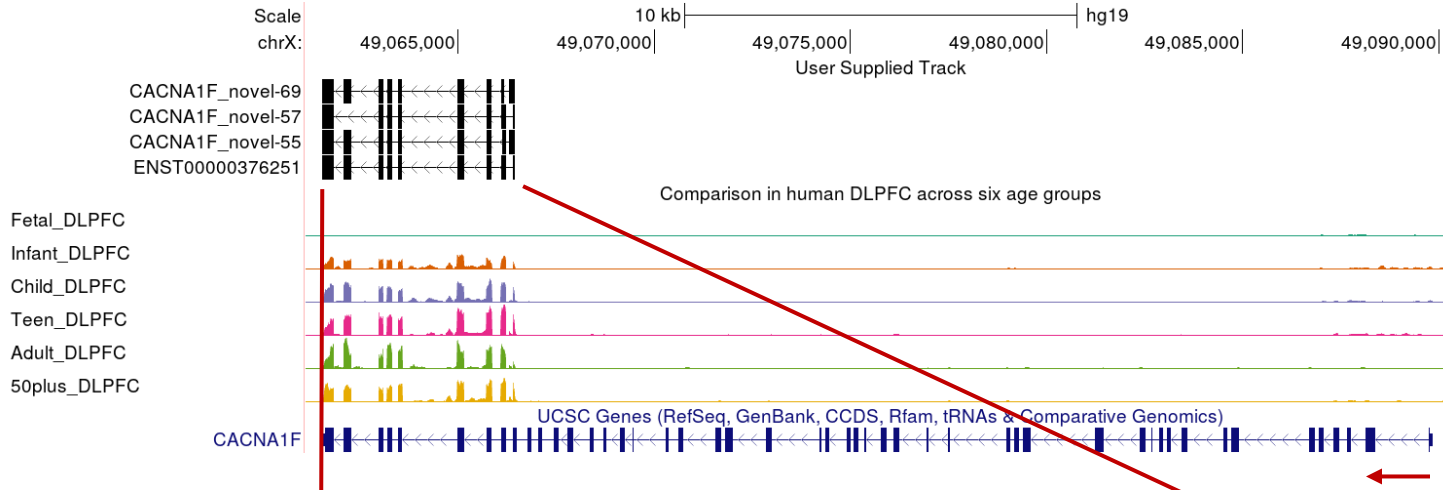


Table 2.9. Genomic location of each exonic region (hg19) from the long-read sequencing of short-length transcript in the 3' region of *CACNA1F* from the human brain.

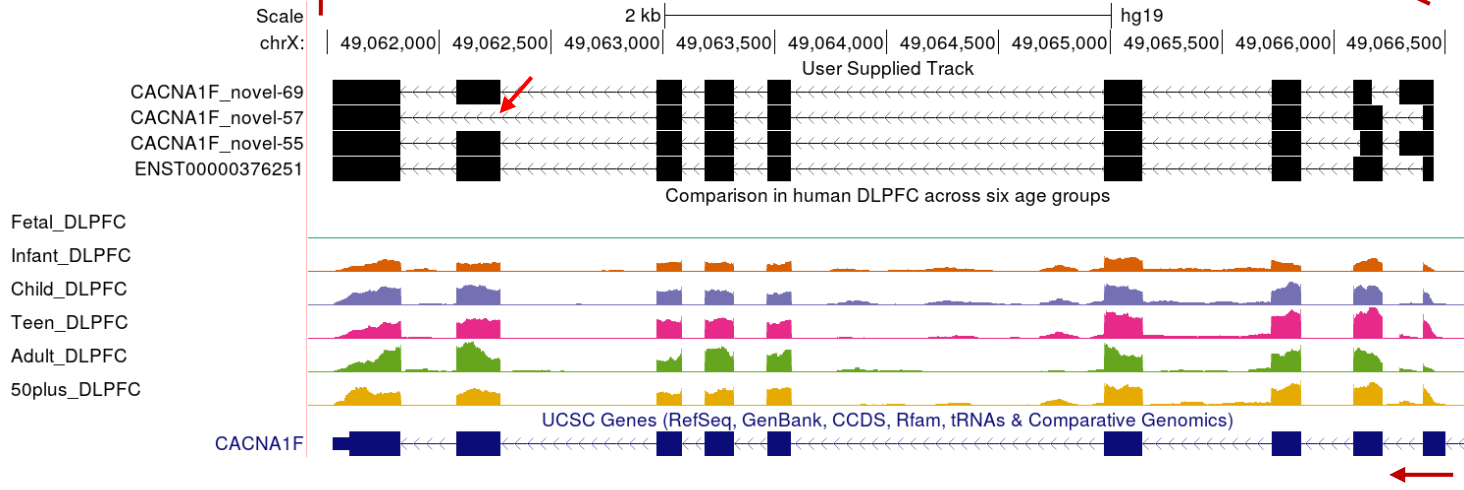
Genomic location	Known exons or novel exons	Exon number in the full-length <i>CACNA1F</i>	Exon number in the short-length transcript in the 3' region of <i>CACNA1F</i>
ChrX:49,066,401-49,066,447	Known exon	Exon 40	Exon 1
ChrX:49,066,089-49,066,219	Known exon	Exon 41	Exon 2
ChrX:49,065,721-49,065,853	Known exon	Exon 42	Exon 3
ChrX:49,064,974-49,065,144	Known exon	Exon 43	Exon 4
ChrX:49,063,463-49,063,573	Known exon	Exon 44	Exon 5
ChrX:49,063,189-49,063,316	Known exon	Exon 45	Exon 6
ChrX:49,062,972-49,063,084	Known exon	Exon 46	Exon 7
ChrX:49,062,076-49,062,273	Known exon with possibility of exon skipping	Exon 47	Exon 8
ChrX:49,061,597-49,061,828	Known exon	Exon 48	Exon 9

'Known exon' means an exon that is aligned to GENCODE V35 metagene of *CACNA1F* whereas 'novel exon' is an exon that was not previously annotated in the metagene of *CACNA1F* but newly identified from the long-read nanopore sequencing here. No novel exon was identified.

A.



B.



*Figure 2.13. The top 4 most highly expressed in-frame short-length coding transcripts from the 3' region of CACNA1F in the human brain. The transcripts are shown in black, mapped to the CACNA1F meta gene (in blue) displayed on the UCSC genome browser (hg19; <http://genome.ucsc.edu>). Colored tracks show LIBD short-read RNA-seq data. The orientation of this gene is from the right to the left (indicated with a red arrow). **A** shows the full CACNA1F locus, and **B** shows a 'zoomed in' version focusing on the short-length coding transcripts. Each transcript is an in-frame coding transcript. There are no identified novel exons from the long-read nanopore sequencing. However, one possible exon-skipping event was observed, which is marked with a red arrow.*

Table 2.10. Protein predicted to be encoded by each transcript

Transcript name	Entry	Entry name	Protein names	Gene names	Organism	Length
ENST00000376251, CACNA1F_novel-69, CACNA1F_novel-55,	O60840	CAC1F_HUMAN	Voltage-dependent L-type calcium channel subunit alpha-1F (Voltage- gated calcium channel subunit alpha Cav1.4)	CACNA1F CACNAF1	Homo sapiens (Human)	ENST00000376251: 1561-1977 CACNA1F_novel-69, CACNA1F_novel-55:1631 - 1977
	O60840-2		Isoform 2 of Voltage-dependent L-type calcium channel subunit alpha-1F (Voltage- gated calcium channel subunit alpha Cav1.4)			ENST00000376251: 1550 – 1966 CACNA1F_novel-69, CACNA1F_novel-55:1620 - 1966
	O60840-4		Isoform 3 of Voltage-dependent L-type calcium channel subunit alpha-1F (Voltage- gated calcium channel subunit alpha Cav1.4)			ENST00000376251:1496 – 1912 CACNA1F_novel-69, CACNA1F_novel-55:1566 - 1912
CACNA1F_novel-57	O60840-7	CAC1F_HUMAN	Isoform 6 of Voltage-dependent L-type calcium channel subunit alpha-1F (Voltage- gated calcium channel subunit alpha Cav1.4)	CACNA1F CACNAF1	Homo sapiens (Human)	1561 - 1911

### 2.4.4.3. Examining regional and individual variations in the transcript profiles of the short-length transcript from the 3' region of *CACNA1F*

PCA was used to examine how transcription isoform expression of the short-length transcript from the 3' region of *CACNA1F* varied between brain regions and individuals, using total number of reads per samples, which each input of sample was standardized. PCA shows no clear clustering of samples by either brain region (Figure 2.14A) or individual (Figure 2.14B).

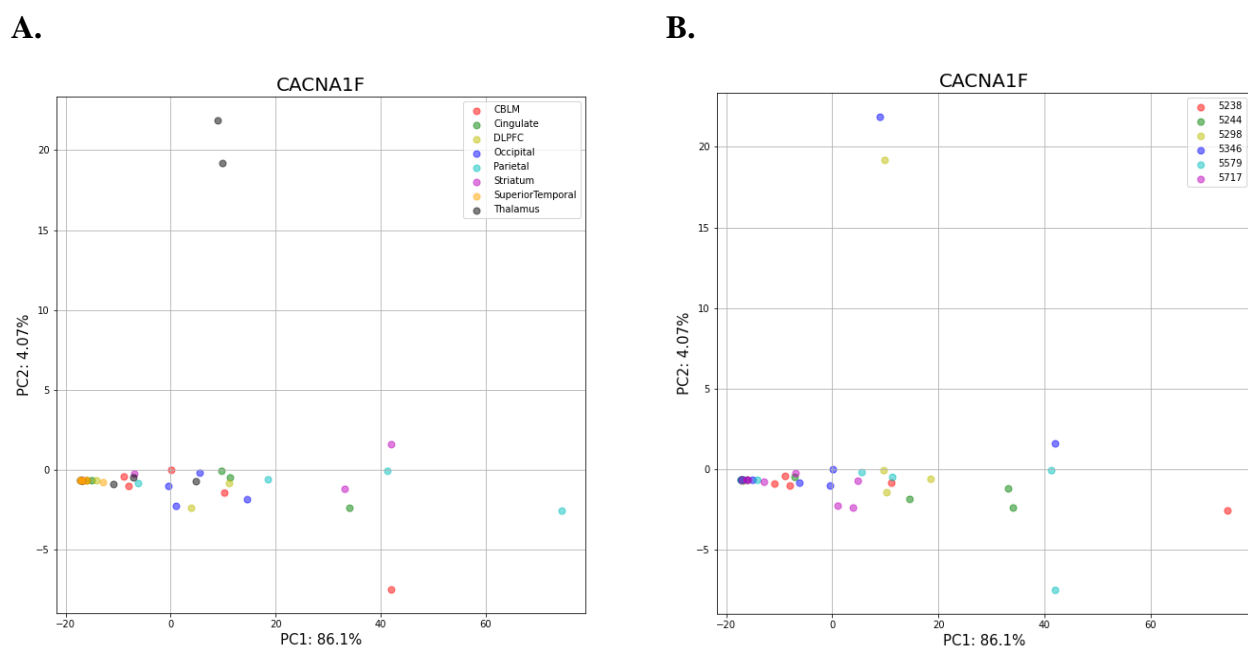


Figure 2.14. Principal Component Analysis (PCA) based on transcripts expression. **A** PCA plot showing that the expression of short-length transcript from the 3' region of *CACNA1F* does not have evidence of clustering across the 8 human brain regions. **B** PCA plot showing that transcript expression of short-length transcript in the 3' region does not show evidence of clustering across the 6 individuals.

## 2.5. Discussion

Previously, *CACNA1S*, a gene encoding Cav1.1 channels, and *CACNA1F*, a gene encoding Cav1.4 channels, were not thought to be expressed in the human brain. Instead, expression was thought to be limited to skeletal muscle and retina, respectively (McRory et al., 2004; Zamponi et al., 2015). However, short-read RNA seq data from LIBD and the GTEx Project (Figure 1.6 and 1.7 in chapter 1) suggest that the exons and exon junctions toward the 3' region of both *CACNA1S*

and *CACNAIF* are expressed in the human brain, whilst those toward the 5' region of the genes are not. Specifically, the exon junction expression data showed that transcripts with exon junctions in the 3' region of *CACNAIS* and *CACNAIF* are expressed in the human brain. Therefore, this chapter aimed to investigate the presence of the short-length transcript arising from the 3' region of *CACNAIS* and *CACNAIF* in the human brain using PCR amplification and long-read nanopore sequencing. In this chapter, I confirmed the presence of short-length transcripts of *CACNAIS* and *CACNAIF* but not full-length *CACNAIS* and *CACNAIF* transcripts in human brain. The collective evidence of both short-read RNA-seq data and long-read nanopore sequencing imply that the unknown transcript would arise from the 3' region of *CACNAIS* and *CACNAIF* in the human brain.

The highly expressed coding transcripts of short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* were identified (Figure 2.6; 2.12). No novel exons were identified, but a possible exon skipping event was identified (Figure 2.7; 2.13). There were two highly expressed short-length transcripts from the 3' region of *CACNAIS* (Figure 2.7); *CACNAIS\_novel-287* had all known exons, and *CACNAIS\_novel-596* had exon 6 skipped. These two transcripts encode slightly different proteins. *CACNAIS\_novel-287* encodes part of the protein which has an entry name of *CAC1S\_HUMAN*, whereas *CACNAIS\_novel-596* encodes part of a protein which has an entry name of *B1ALM3\_HUMAN* (The UniProt Consortium, 2021). Nevertheless, both proteins are encoded by short-length transcripts from the 3' region of *CACNAIS*. This shows that exon 6 skipping may influence the protein that each transcript encodes, which in turn may alter the function of each transcript. Exon 6 of the short-length transcript from the 3' region of *CACNAIS* is exon 29 of the full-length *CACNAIS* transcript which is studied to be skipped and

affects voltage sensitivity and gating properties of the full-length channel in muscle (Tuluc et al., 2016).

Moreover, there are four highly expressed short-length transcripts from the 3' region of *CACNA1F* (Figure 2.13). Among the transcripts, ENST00000376251, *CACNA1F\_novel-69*, *CACNA1F\_novel-55* contain all known exons, and are predicted to encode a part of three protein isoforms of Cav1.4, while *CACNA1F\_novel 57*, which has exon 8 skipped, is predicted to encode a part of protein isoform 6 of Cav1.4. This implies that the exon 8 skipping may affect the function of the transcript. Exon 8 of the short-length transcript from 3' region of *CACNA1F* is exon 47 of the full-length *CACNA1F* transcript. In the retina, exon 47 plays a crucial role in Ca<sup>2+</sup>-dependent inactivation (Haeseleer et al., 2016; The UniProt Consortium, 2021). From the long-read nanopore sequencing of the short-length transcripts from the 3' region of *CACNAIS* and *CACNA1F*, many different coding transcripts with the possibility of exon skipping were identified. Due to the exon skipping, each transcript may encode different proteins, which in turn may have different functions in the human brain.

These short-length transcripts from the 3' region of *CACNAIS* and *CACNA1F* have not been reported previously. However, there are multiple studies showing the presence of truncated calcium channels (Okagaki et al., 2001) or transcripts of VGCC genes, especially *CACNA1C*, as the result of an alternative transcription start site (TSS) (Gomez-Ospina et al., 2013; Pang et al., 2003). For example, *CACNA1C*, which encodes Cav1.2, produces two different families of transcripts in different tissue types via the use of an alternative TSS. One family of transcripts has exon 1a and is expressed in the heart, whereas the other has exon 1b and is expressed in the brain. These two transcript families use different promoters and TSSs (Dai et al., 2002; Pang et al., 2003; N. Saada et al., 2003). Moreover, *CACNA1C* also encodes the short-length Calcium Channel

Associated Transcript (CCAT), which corresponds to the C-terminal region of Cav1.2. Within the nucleus, CCAT binds to p54(nrb)/ NonO, a transcriptional regulator, and activates transcription of the regulator and other neuronal genes. For example, it binds to the enhancer of Connexin 31.1 gene (Cx31.1) and regulates the expression of both the enhancer and Cx31.1. The generation of CCAT was hypothesized to be via proteolytic cleavage (Gomez-Ospina et al., 2006) which has been examined from other VGCCs (Hell et al., 1993; Kubodera et al., 2003). However, later the team found that CCAT is produced from an exonic promoter in exon 46 of the full-length *CACNA1C* gene and localizes to the nucleus in neurons and functions as a transcription factor (Gomez-Ospina et al., 2013).

Although short *CACNA1C* transcripts have been reported, it is unknown how many isoforms are produced due to novel TSSs or alternative splicing. Long-read sequencing can reveal previously unknown transcripts (Wright et al., 2022). For example, Wright et al., (2022) identified 2567 novel transcripts via long-read sequencing of which 983 novel coding transcripts possess novel TSS, which are not known start sites. Among 983 novel coding transcripts with alternative TSS, the transcript start site of 333 transcripts were identified from Fantom5 CAGEseq data, which is genome-wide transcription start site data. This means that novel transcripts identified by long read sequencing techniques might be a novel complete transcript produced by a novel alternative transcription start site (Wright et al., 2022). Given their localization at the 3' end of the full-length gene, it is likely that the short-length transcripts arise from a novel TSS. I therefore characterized the 5' ends of the short transcripts, which will be described in chapter 3.



## 2.6. Methodological limitations

Given the highly restricted expression patterns of *CACNAIS* and *CACNAIF*, I did not have a positive control for the PCR amplification experiments i.e., a tissue known to express the full-length channel. *CACNAIS* is predominantly expressed in the skeletal muscle, and *CACNAIF* is highly expressed in the retina (Catterall et al., 2005; McRory et al., 2004; Zamponi et al., 2015). Without amplifying the full-length *CACNAIS* and *CACNAIF* from the positive control samples, it is not clear whether the full-length *CACNAIS* and *CACNAIF* were not amplified in the human brain because they do not exist or due to primers or settings used for PCR amplification. Therefore, it would make the chapter more concrete if I could examine the full-length genes in positive control samples (skeletal muscle and retina).

In order to address the limitation of not having a positive control, I optimised multiple PCR conditions, including number of cycles, and annealing temperatures. Nevertheless, it is possible that my failure to detect the full-length channels reflects experimental failure. However, arguing against this possibility is short-read RNA seq data from LIBD (Figure 1.6 in chapter 1) and GTEx (Figure 1.7 in chapter 1). Both data show that the full-length *CACNAIS* and *CACNAIF* are not expressed in the human brain, whereas the exons toward 3' region of *CACNAIS* and *CACNAIF* may have an expression in the human brain.

The protocol used in this chapter relies on the use of primers to perform targeted sequencing. As a result, PCR amplification-related factors could introduce bias into the data. For example, PCR can amplify contaminants or insignificant products that happen to have high affinity with the primers used (Garibyan & Avashia, 2013). In addition, by using specific primers to amplify regions of interest, I may have missed other short-length transcript in the 3' regions with

different start sites. Therefore, all the possible TSSs of short-length transcript in the 3' region of *CACNAIS* and *CACNAIF* will be examined in Chapter 3.

## **Chapter 3 – Identification of the transcription start site of the 3' transcripts of LTCC genes**

### **3.1. Introduction**

#### **3.1.1. Rationale of the chapter**

In chapter 2, even though the full-length *CACNAIS* and *CACNAIF* transcripts have not been reported to be expressed in the human brain (Zamponi et al., 2015), the short-length transcripts from their 3' regions were identified in the human brain via PCR amplification and long-read nanopore sequencing. However, it is still unclear how the short-length transcripts were produced. In this chapter, it is hypothesized that each short-length transcript represents a complete transcript with its own transcription start site (TSS). 5' Rapid amplification of cDNA ends (5' RACE) was used to characterize the 5' ends of the short-length transcripts. Then, Cap Analysis Gene Expression (CAGE) data were analyzed to confirm whether TSSs that were found using 5' RACE could be observed in publicly available genome-wide sequencing data of TSSs. This provides an orthogonal source of support for my observations and allows me to examine the possibility of de novo TSS in other genes encoding VGCC  $\alpha 1$  subunits.

#### **3.1.2. Transcription start sites**

A TSS is the first nucleotide where transcription of a gene starts (Sandelin et al., 2007). The genomic region around the TSS is a core promoter site, where proteins such as RNA polymerase and transcription factors bind and form a complex to initiate transcription (Smale & Kadonaga, 2003). Knowing the exact position of a TSS is crucial not only to know where the transcription of a transcript starts but also to identify the regulatory regions which might control gene expression (Xu et al., 2019). Many genes have more than one TSS, which are further downstream or upstream from the canonical TSS (The ENCODE Project Consortium, 2007).

Alternative TSSs can be differentially used across tissue or cell types (Shephard et al., 2007), and developmental stages (Davis Jr & Schultz, 2000; Steinhorsdottir et al., 2004b).

Having alternative TSSs means that a gene would have alternative 5' untranslated regions (5' UTRs) or different protein-coding regions, which in turn may affect translation efficiency and potentially the sequence and function of the encoded protein (Xu et al., 2019). Even when transcripts produce the same protein, if they have different 5' UTRs, each transcript might have different expression and localization patterns (Pruunsild et al., 2007; W. Tan et al., 2007). Therefore, examining where the transcriptions of the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* start adds to our understanding of the transcriptional regulation of calcium channel genes in the human brain.

### **3.1.3. Cap Analysis of Gene Expression**

As studying TSSs is crucial to understand the possible regulatory mechanism of genes and functions of the proteins that the gene produces, it is desirable to examine TSSs on a transcriptome-wide scale. CAGE is a high-throughput method that allows genome-wide analysis using the cap trapping method (Carninci et al., 2006) which isolates cDNA/mRNA hybrids; RNA is then chemically biotinylated, and magnetic beads capture the hybrids (Kodzius et al., 2006). CAGE allows for the mapping of most TSSs and promoter sites by creating CAGE tags. Total RNA is transcribed to cDNA to create CAGE tags, which are sequence of 20 nucleotides attached to the 5' end of the transcript. A biotinylated linker sequence is ligated to the 5' end of the cDNA so that the restriction endonuclease site, MmeI, can be introduced next to the 5' ends of transcripts. After the second strand is synthesized, MmeI cleaves the region about 20 to 18 nucleotides away (3') from the restriction site, which leaves two-base overhangs. The 20 to 18 nucleotide sequences become the CAGE tag. Next, the second linker, with a different restriction site, is ligated to the 3'

end of the CAGE tag to amplify the tag. The CAGE tag is subsequently released with restriction enzyme digestion. By mapping identified CAGE tags to the genome, multiple tags in the region can be counted (Figure 3.1).

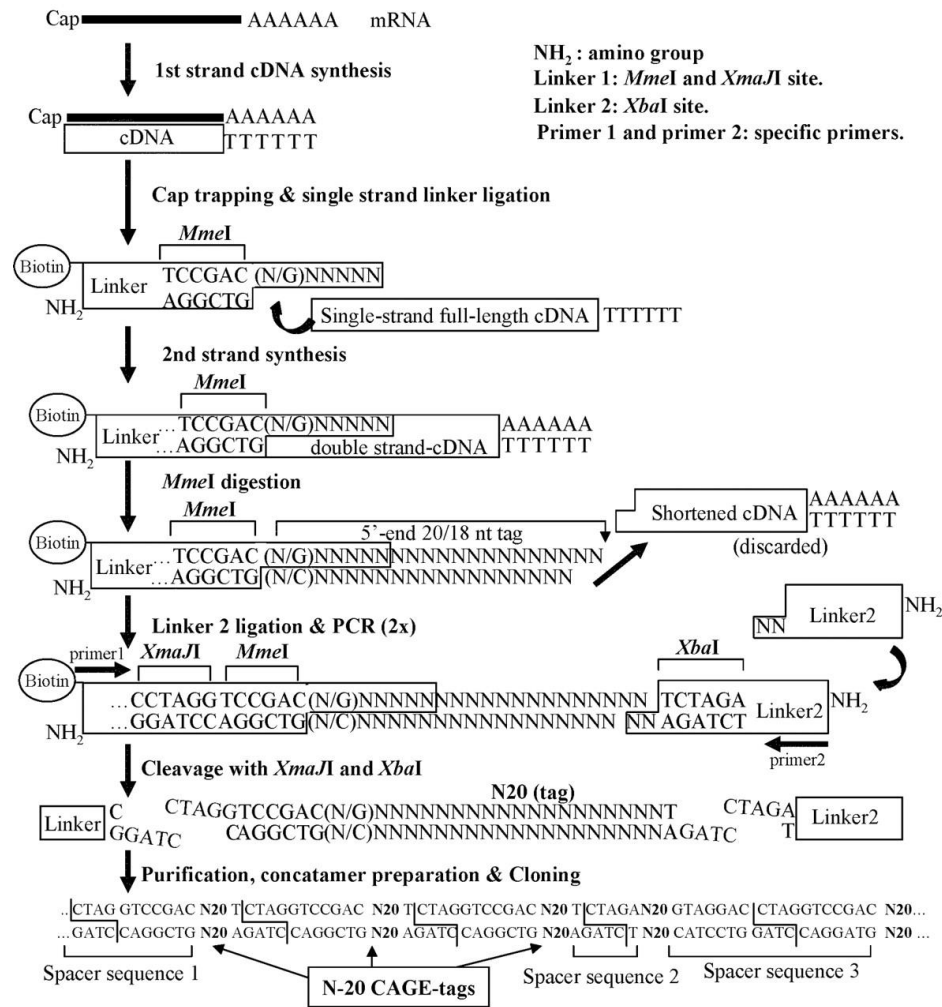


Figure 3.1. Schematic representation of CAGE. This is retrieved from (Shiraki et al., 2003)

### 3.1.4. Previous studies examining the alternative transcription start site of VGCC genes

As discussed in section 1.3 in chapter 1, prior studies examined diverse forms of channels by using alternative TSSs or by undergoing alternative splicing. One report demonstrated that three-domain calcium channels are produced not because of alternative splicing but due to alternative TSS usage. *TuCa1*, a non-mammalian VGCC which is homologous to the vertebrate

LTCCs, uses an alternative TSS to produce a three-domain channel. The three-domain channel of *TuCa1* lacks domain I and S1, S2 of domain II. It does not show functional calcium currents but inhibits the expression of the full-length channel (Okagaki et al., 2001).

Furthermore, several studies reported that *CACNA1C* produces multiple transcripts using alternative TSSs. *CACNA1C* uses alternatively spliced first exons in a tissue-specific manner: exon 1a, which encodes 46 amino acids, in heart and exon 1b, which translates to 16 amino acids, in the brain (Blumenstein et al., 2002). However, later, it was discovered that transcripts with exon 1a and exon 1b are produced from different TSSs, which were characterised using 5'RACE. The promoter region of exon 1a transcripts only upregulates expression in the heart, showing that a gene can use alternative TSSs to induce tissue-specific expression (Dai et al., 2002; Pang et al., 2003; N. I. Saada et al., 2005). Moreover, *CACNA1C* was known to encode a CCAT via an alternative promoter region in exon 46 of the full-length *CACNA1C* (Gomez-Ospina et al., 2013). These previous experiments show that it is possible that the LTCC genes use alternative TSSs to produce diverse transcripts or diverse channel structure.

### **3.1.5. The aim of this chapter**

It remains unclear whether the short-length transcripts identified in chapter 2 are produced as intact but shorter transcripts arising from alternative TSSs in the 3' region, or whether they are generated from mRNA cleavage, or whether they result from mRNA degradation or another molecular artefact. Therefore, the aim of this chapter is to examine how the short-length transcripts are produced from the full-length LTCC genes by characterising their TSSs in the human brain. Two complementary methods were used: 5' RACE and analysis of CAGE data from FANTOM 5 using the CAGEr package in R. Results from these analyses were aligned to see if the data were convergent. Even though only *CACNA1S* and *CACNA1F* were sequenced in chapter 2, the TSSs

in the 3' regions of the other LTCC genes (*CACNA1C* and *CACNA1D*) were examined as well. Moreover, using CAGE data, the possible existence of the short-length transcripts in the 3' region of other genes encoding VGCC  $\alpha$ 1 subunits were also examined. Overall, this chapter tested the hypothesis that the short-length transcript in the 3' region of these VGCC genes arises from de novo TSSs. Moreover, CAGE analysis will assess whether other genes encoding VGCC  $\alpha$ 1 subunits might also have novel transcripts with de novo transcription in their 3' regions.

## **3.2. Methods**

### **3.2.1. Failed attempts at 5' RACE**

To examine the TSSs of the short-length transcript from the 3' region of LTCC genes in the human brain, 5' RACE was used. 5' RACE uses full-length capped mRNAs and RT-PCR to detect TSS of the gene of interest. Two different 5' RACE experimental kits were used, including RLM firstchoice RACE (ThermoFisher, UK; AM1700) and 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche, UK; 03353621001; Figure 3.2). The technique proved to be problematic, and a range of conditions were tested before experiments were successful. The failed methodologies can be found in appendix 7 and only successful attempts of 5'RACE will be presented in this chapter.

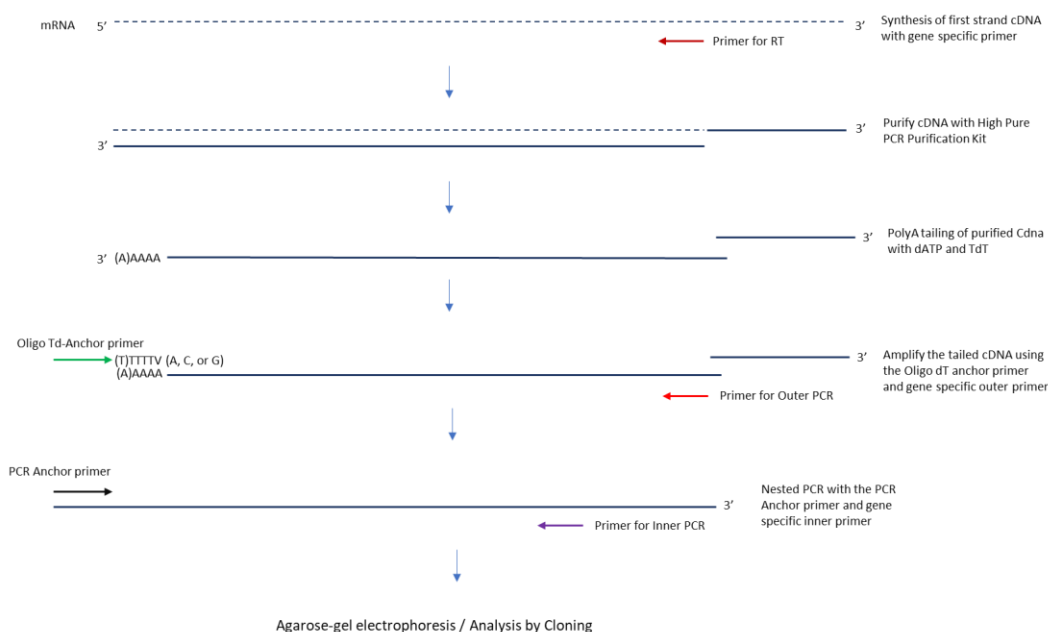


Figure 3.2. Overview protocol of 5'/3' RACE Kit, 2nd Generation. The 5'/3' RACE Kit uses a gene specific primer for RT and nested PCR. This gives more sensitive and gene-specific results.

### 3.2.2. Successful attempts of 5' RACE

After several optimizations, a TSS of each short-length transcript from the 3' region of LTCC genes were successfully identified with 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche, UK; 03353621001). This section will describe each step of 5' RACE experiment using 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche, UK; 03353621001). Moreover, this section includes gel electrophoresis pictures visualizing all possible TSSs of each short-length transcript from the 3' region of LTCC genes identified from 5' RACE which were excised and sequenced via Sanger sequencing. Even though these gel-electrophoresis figures show the results of 5' RACE, they were included into Methods section as they are also the result of the optimized protocol. They also aid in describing the flow of experiments completed for this chapter (Figure 3.3). First of all, instead of total RNAs, poly A<sup>+</sup> RNAs were used, then nested PCR was conducted which were checked via gel electrophoresis. The PCR amplified products were sequenced using Sanger sequencing to confirm that the TSS matches to the gene of interest.



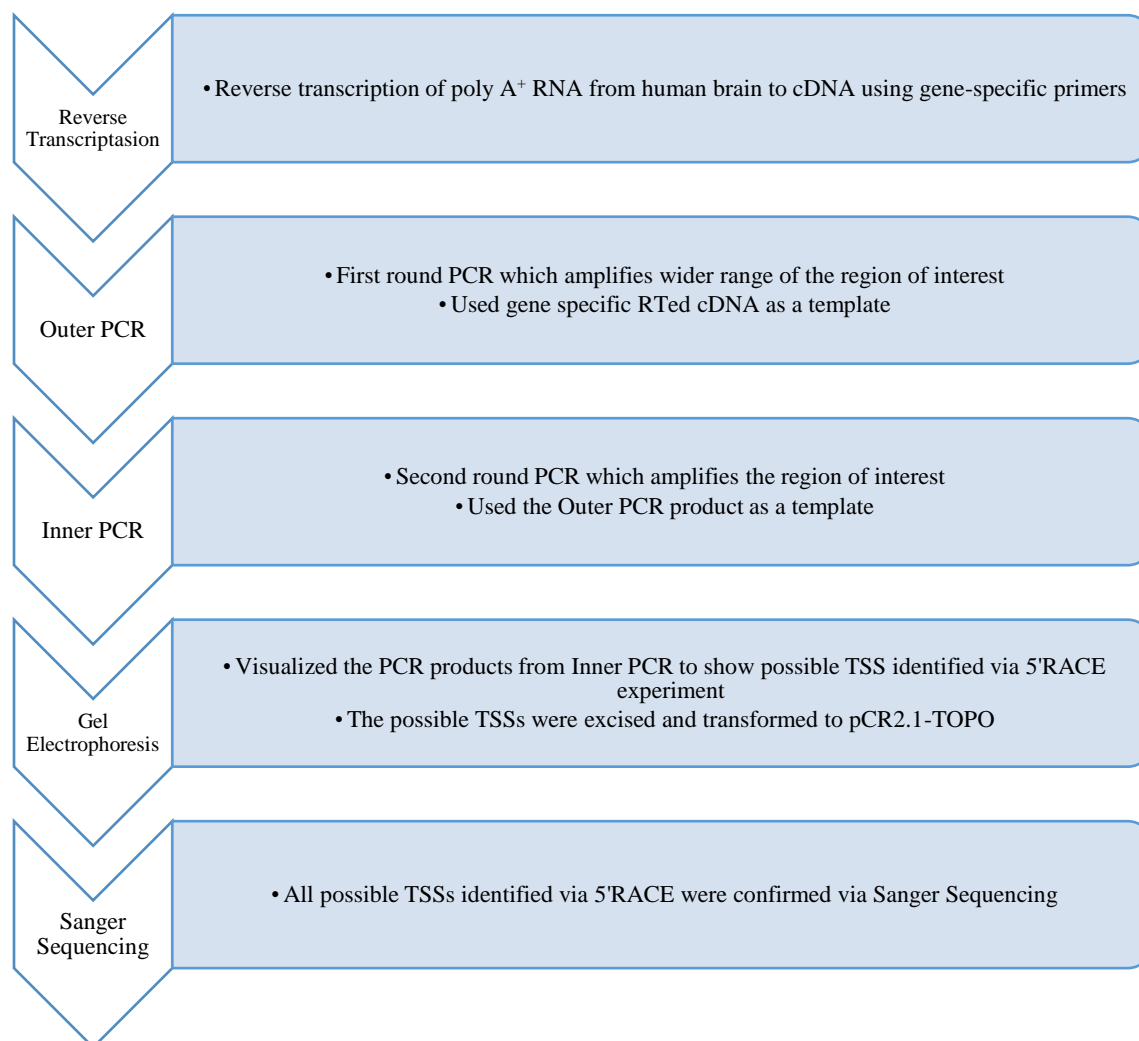


Figure 3.3. Flow chart of experiments done for this chapter to identify TSS of the short-length transcript of LTCC genes.

### 3.2.2.1. Using poly A<sup>+</sup> RNAs

Poly A<sup>+</sup> RNA, which contains coding sequences, was used instead of total RNA, which contains both coding and non-coding sequences (Kukurba & Montgomery, 2015). This increases the specificity of PCR amplification as only polyadenylated mRNAs would be amplified. Two commercial poly A<sup>+</sup> RNA samples were obtained to examine TSSs of short-length transcripts of L-type VGCC genes (Table 3.1).

Table 3.1. The poly A<sup>+</sup> RNA samples were used for the 5' RACE experiment. All poly A<sup>+</sup> RNA samples were pooled together.

Brain region	Source
Frontal Cortex	Takara
Cerebellum	Clontech

As the kit uses specific primers, three different primers for RT, outer PCR, and Inner PCR were designed (Table 3.2). Primers were designed using Primer3 plus (Source code is available at <http://sourceforge.net/projects/primer3/>) (Untergasser et al., 2007) and a plasmid Editor (ApE; RRID:SCR\_014266) (Davis & Jorgensen, 2022), and primers were checked using “In silico PCR” from UCSC (University of California, Santa Cruz) Genome Browser (<http://genome.cse.ucsc.edu/>) (Kent et al., 2002). The 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche, UK; 03353621001) provides oligo-DT anchor primers and PCR-anchor primers; therefore, the compatibility between provided primers and designed primers such as creating hairpins or primer dimers were checked during the design of the 5'RACE primers.

Table 3.11. List of primers used for 5' RACE experiment

Transcript of interest	Primer Name	Primer sequence (5'-3')	Primer usage	Coordinate
Short-length transcript from the 3' region of <i>CACNA1S</i>	SR3	GAAGAAGAGCATGAC GATGAGCAG	RT	chr1:201,022,361- 201,022,384
	SI2	GGTGACAATGTACCA CACCTG	Outer	chr1:201,029,852- 201,029,872
	SI1	CACACGGTTGTTGTA GATGGGAC	Inner	chr1:201,030,533- 201,030,555
Short-length transcript from the 3' region of <i>CACNA1C</i>	CI2	GTAGCTGACGTGGTT GCCGAACAG	RT	chr12:2,788,622- 2,788,645
	CI1	CTTGTCCAGCTCCTCC TCAG	Outer	chr12:2,786,966- 2,786,985

	CO5	CCTGCTCTTTGCGCTT CTTGAAC	Inner	chr12:2,786,321- 2,786,343
Short-length transcript from the 3' region of <i>CACNAID</i>	DO2	GGTGGTATTGGTCTG CTGAAGGG	RT	chr3:53,835,143- 53,835,165
	DO1	GCAGTGTCTTAATC CCGCC	Outer	chr3:53,834,275- 53,834,294
	DI3	CAGGGTACTTTCCCA CCAGTCC	Inner	chr3:53,820,932- 53,820,953
Short-length transcript from the 3' region of <i>CACNAIF</i>	FR2	GTAATCCTCACAAC TGCCCTG	RT	chrX:49,063,025- 49,063,046
	FI3	GGTGCCCATCCATGT ATCTCTG	Outer	chrX:49,063,228- 49,063,249
	FI2	GACAGGCAGAGACAC AGAAATC	Inner	chrX:49,065,082- 49,065,103

RT= primer used for RT, Outer = primer used as reverse primer for outer PCR amplification, Inner = primer used as reverse primer for inner PCR amplification.

### 3.2.2.2.Reverse transcription

Reagents (Table 3.3) were mixed to a total volume of 20µl. The mix was incubated at 55°C for 60 minutes and then at 85°C for 5 minutes. After the incubation, High Pure PCR Product Purification Kit (Roche, UK; 11796828001) was used to purify cDNA. 100µl of binding buffer was added to the 20µl of the first-strand cDNA reaction. The mixture was put into the collection tube and centrifuged for 30 seconds at 8000 x g. The flow-through was discarded and 200µl of wash buffer was added to the filter tube before being centrifuged for 30 seconds at 8000 x g. Once the flow-through was discarded, 200µl of wash buffer was added to the filter tube and centrifuged for at least 2 minutes at 8000 x g. The flow-through was again discarded. The filter tube was inserted into a sterile 1.5ml Eppendorf tube and 50µl of elution buffer was added. After 1 minute of incubation, the 1.5ml tube was centrifuged for 30 seconds at 8000 x g.

Table 3.12. Mixture of reagent for RT for 5'/3' RACE Kit, 2nd Generation

Reagent	Amount of volume ( $\mu$ l)
CDNA synthesis buffer	4.00
Deoxynucleotide Mixture	2.00
CDNA synthesis primer (primer for RT) (12.5 $\mu$ M)	1.25 (As the primer is 10 $\mu$ M, 1.25 $\mu$ l of the primer was added)
Poly(A)RNA (0.2-2 $\mu$ g)	2.00 (each sample is 100ng/ $\mu$ l therefore, 1 $\mu$ l of each sample was pooled to make 0.2 $\mu$ g)
Transcriptor Reverse Transcriptase	1.00
Nuclease Free water	9.25
Rnase inhibitor	0.50
<i>Total volume</i>	<i>20</i>

### 3.2.2.3. Poly(A) tailing

Once the RNA was reverse transcribed and purified, adenine nucleotides were left out from the poly A<sup>+</sup> RNA. Therefore, the poly(A) tail had to be ligated to the 3' end of the first-strand cDNA using recombinant terminal transferase and dATP. Reagents (Table 3.4) were mixed in the 1.5ml Eppendorf tube. The mixture was incubated at 94°C for 3 minutes. 1 $\mu$ l of Terminal Transferase Recombination (80U/ $\mu$ l) was added to the mixture. Then, it was incubated at 37°C for 20 minutes and then at 70°C for 10 minutes to heat inactivate the Terminal Transferase.

Table 3.13. Mixture of reagents for Poly(A) tailing for 5'/3' RACE Kit, 2nd Generation

Reagents	Amount ( $\mu$ l)
Purified cDNA sample	19.00
Reaction Buffer, 10x concentration	2.50
2 $\mu$ M dATP	2.50

### 3.2.2.4. Outer PCR amplification

After Poly(A) tailing, the first PCR amplification with gene specific primers were processed. For the outer PCR, illustra PuReTaq Ready-To-Go PCR Beads (Cytiva, UK; GE27-9559-01) were used. Reagents (Table 3.5) were mixed to the PCR tubes with bead taqs. The PCR

program for the outer PCR was programmed as follows: 5 minutes at 95°C for the first stage, 35 cycles of 15 seconds at 95°C, 15 seconds at 61°C, 1 minute at 72°C, which cycle was followed by 4 minutes at 72°C.

*Table 3.14. Mixture of reagents for Outer PCR amplification*

Reagents	Amount ( $\mu$ l)
Oligo-dT anchor primer (5'-GACCACGCGTATCGATGTCGAC TTTTTTTTTTTTTTTTV-3'; V= A, C, or G)	0.50 (18.75 $\mu$ M)
Gene specific reverse primer (10 $\mu$ M)	18.75 (187.5 $\mu$ M)
cDNA	1.00
Nuclease Free water	4.75
<i>Total</i>	<i>25.00</i>

Skewed PCR is performed for the first round of PCR following RT as it has been shown that increasing the concentration of the gene specific primer compared to the general primer (primer from the kit) in the outer PCR amplification significantly improves the amplification of rare transcripts (Bespalova et al., 1998). As *CACNAIS* and *CACNAIF* are known to have a low abundance in the human brain, the specific primer to general primer ratio was 10 to 1 (0.5/0.05mM). Reduced concentration of the general primer compared to the gene specific primer helps to reduce the amplification of non-specific PCR products which might be because of the attenuation of non-specific binding of the general primers. Therefore, 18.75 $\mu$ l of specific primers (187.5mM) and 0.5 $\mu$ l of general primer (18.75mM) were used for the outer PCR amplification.

### **3.2.2.5. Inner PCR amplification**

After the outer PCR amplification, inner PCR amplification was performed using the products from the outer PCR as templates. Inner PCR amplification was also performed with illustra PuReTaq Ready-To-Go PCR Beads (Cytiva, UK; GE27-9559-01). As inner PCR uses PCR amplified products as templates, 5 $\mu$ l of cDNA in the PCR reaction would be too concentrated and

would hinder specific primers to amplify the region of interest. To solve this problem, the outer PCR products were diluted to 1 in 20 and 1 in 400 to determine the most efficient concentration. Negative control samples lacking the outer PCR product template were also diluted accordingly. Reagents (Table 3.6) were mixed into the PCR tube with bead taq. The PCR program was programmed as follows: 5 minutes at 95°C, 35 cycles of 15 seconds at 95°C, 15 seconds at 61°C, 1 minute at 72°C, followed by 1 minute at 72°C.

*Table 3.15. Mixture of reagents for Inner PCR amplification.*

Reagents	Amount (µl)
PCR anchor primer (5'-GACCACGCGTATCGATGTTCGAC-3'; 12.5µM)	0.50
Gene specific reverse primer (10µM)	0.50
Outer PCR-ed template	1.00
Nuclease Free water	23.00
<i>Total</i>	<i>25.00</i>

### **3.2.2.6. Gel electrophoresis**

In order to visualize the final RACE products, the products were isolated on a 2% agarose gel prepared in 100ml 1x TAE with 10 µl of ethidium bromide (5mg/ml). 1 µl of loading dye purple (6x) (New England BioLabs, UK; B7025S) was mixed with 5 µl of RACE products. For the size detection, 0.5 µl of 50bp NEB ladder (New England BioLabs, UK; N3236L), 1 µl of gel loading dye, Purple (6x) (New England BioLabs, UK; B7025S) and 4.5 µl of NF water were mixed and loaded onto the gel. The gel was run at 140V for 40 minutes.

To excise the possible TSSs of the short-length transcripts from the 3' region of the LTCC genes, the final RACE products were isolated from a 2% agarose gel prepared in 100ml 1x TAE with 10µl of gel green (Biotium, UK) and extracted using the Monarch DNA gel extraction kit (New England BioLabs, UK; See Methods 2.3.8 in chapter 2).

### 3.2.2.7. Gel extraction of potential RACE products

The selected PCR amplifications were extracted from the 2% agarose gel prepared in 100ml of 1/10 TAE buffer with 10 $\mu$ l gel green (Biotium, UK) using the Monarch DNA gel extraction kit (New England BioLabs, UK). For the specific protocols, see methods section 2.3.8 in chapter 2. The extracted products were then transformed into pCR 2.1 TOPO (see section 3.2.2.8)

### 3.2.2.8. TA TOPO cloning with One shot TOP10 chemically competent cells

The final RACE products of short-length transcript from the 3' region of LTCC genes were isolated from 2% agarose gels prepared in 100ml of 1/10 TAE buffer with 10 $\mu$ l gel green (Biotium, UK) extracted from the gel, and subcloned using the TOPO™ TA Cloning™ Kit with One Shot™ TOP 10 Chemically competent E. coli (Thermofisher, UK; Figure 3.4). The kit was used because the gel purified products have A overhangs, and the kit allows for the insertion of Taq polymerase amplified PCR products into a plasmid vector (Thermofisher, UK).

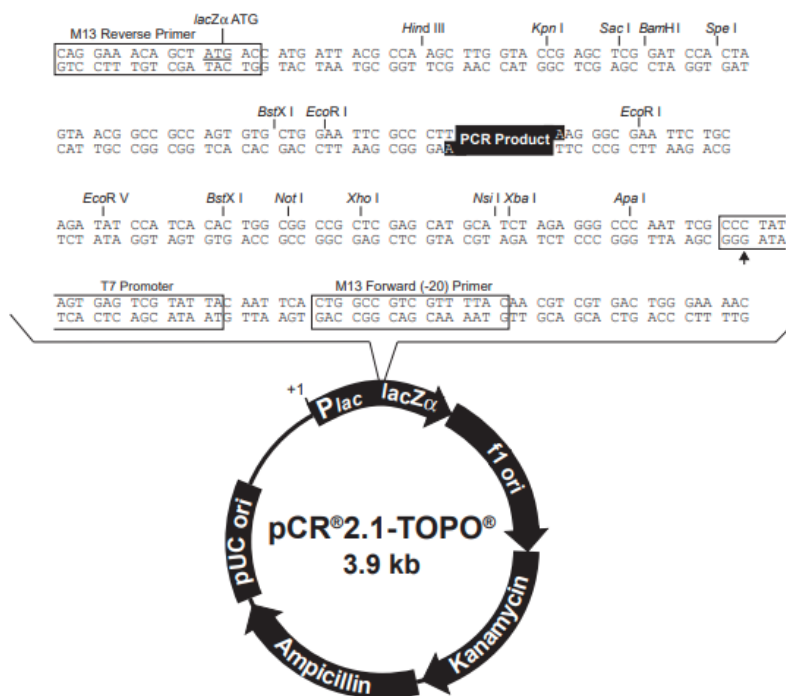


Figure 3.4. Vector map of TOPO™ TA Cloning™ Kit with One Shot™ TOP 10 Chemically competent *E. coli* (ThermoFisher, UK) indicating where the PCR products are inserted (ThermoFisher.com).

The reagents (Table 3.7) were mixed and incubated for 5 minutes at room temperature (22-23°C). 25µl of one shot TOP10 chemically competent *E. coli* was used per reaction. 2µl of TOPO cloning reaction was added to the cells and gently mixed. The mixture was incubated on ice for 30 minutes. The mixed cells were then heat-shocked at 42°C, using a water bath, for 30 seconds and then put directly on ice. Following which, 125µl of S.O.C. medium, at room temperature, was then added to the cell mixture. The mixture was shaken horizontally at 37°C for 1 hour, at a speed of 200rpm. The full volume (152µl) was plated onto the pre-warmed LB-agar plate (500mL of LB-agar with 500µl of ampicillin (100mg/mL)) and subsequently incubated at 37°C overnight (~17 hours).

Table 3.16 Mixture of reagents for TOPO™ TA Cloning™ Kit with One Shot™ TOP 10 Chemically competent *E. coli* (ThermoFisher, UK).

Reagents	Amount (µl)
Fresh PCR product (Elution buffer for Negative Control)	2.0
Salt solution	0.50
Nuclease Free water	Add to a total volume of 2.5
TOPO vector	0.50
<i>Total volume</i>	<i>3.00</i>

### 3.2.2.9. Plasmid purification

After cloning the PCR products into the Oneshot top10 chemically competent *E. coli*, several transformants from each plate were grown in 5ml LB with 5 µl of ampicillin mix (100mg/ml) and incubated overnight at 37°C, shaken at a speed of 150 rpm. The overnight cultured transformants were purified using the QIAprep Spin Miniprep Kit (Qiagen, UK; 27106). The next



day, 2ml of culture was aliquoted and centrifuged at 13.3 x g for 1 minute, and the supernatant was discarded. This step was repeated once more. The pellet was then resuspended in 250 µl of buffer P1 from the Qiagen Kit (supplemented with 200 µl of RNase A and 20 µl of Lyse Blue to buffer P1 first). Then, 250 µl of buffer P2 was added and the tube inverted 6 times. Within 5 minutes, 350 µl of buffer N3 was also added and the tube inverted 6 times. The mixture was then centrifuged at 17,000 x g for 10 minutes. Once the samples were centrifuged, 800 µl of supernatant was added to the column, centrifuged at 17,000 x g for 1 minute, and the flow-through was discarded. 500 µl of Phosphate Buffer (PB) was added and centrifuged at 17,000 x g for 1 minute, and the flow-through was again discarded. 750 µl of buffer PE was then added and centrifuged at 17,000 x g for 1 minute. The column was then placed in a new tube and centrifuged at 17,000 x g for 1 minute to ensure that the buffers were removed from the filter. The column was then inserted into a new collection tube, and 50 µl of elution buffer was added to the center of the membrane, left for 1 minute, and then centrifuged at 17,000 x g for 1 minute.

#### **3.2.2.10. Restriction Enzyme Digestion**

Finally, the minipreped transformants were digested with EcoRI restriction enzyme (Promega, UK), to examine whether the insert had been properly ligated to the vector. EcoRI cuts the vector backbone (pCR2.1 TOPO) on either side of insert site, thereby releasing the insert. The reagents (Table 3.8) were mixed. The mixture was incubated at 37°C for 3 hours and 10 minutes. The digested products were then analyzed by gel electrophoresis, using a 2% agarose gel prepared in 100 ml of 1/10 TAE buffer with 10µl of Ethidium Bromide (5mg/ml). Once the inserts were detected, the plasmids were sent for Sanger sequencing. The sequenced products were then aligned to each gene using the blat function on the UCSC genome browser (<http://genome.ucsc.edu>) to

confirm that the RACE products were specific to the gene of interest and in the anticipated regions (This will be shown in section 3.4.1).

*Table 3.17 Mixture of reagents for Restriction Enzyme digestion with EcoRI (Promega, UK)*

Reagents	Amount ( $\mu$ l)
Nuclease Free water	15.50
Restriction Enzyme 10x Buffer	2.0
Acetylated BSA (10ng/ $\mu$ l)	1.0
DNA (1ug/ $\mu$ l)	1.0
Restriction Enzyme (EcoRI, 10u/ $\mu$ l)	0.5
<i>Final volume</i>	<i>20.0</i>

### 3.2.2.11. Multiple Sequence Alignment

As the TSS of 3' transcript of LTCC genes were identified, the sequence of 3' transcript of LTCC genes were aligned to each other using clustal omega (<https://www.ebi.ac.uk/Tools/msa/clustalo/>) (McWilliam et al., 2013; Sievers et al., 2011). This is to examine the similarities among 3' transcripts of LTCC genes and identify regions that may be evolutionarily conserved.

### 3.2.3. CAGE data analysis using CAGEr

After examining the TSSs of transcripts via 5' RACE, identified TSSs were confirmed by analyzing CAGE data, in a publicly accessible database from the FANTOM5 project (Lizio et al., 2019). From the CAGE data, the exact position of the TSSs of captured RNAs were sequenced and mapped to the reference genome. The number of CAGE tags at each genomic location represents an index of the expression level of the TSS (de Hoon et al., 2015). From the CAGE data, we can infer a single-base resolution map of each TSS, alongside its expression level (de Hoon et al., 2015). For relatively highly expressed genes, 1/5 of TSSs are mapped within exons and known to generate alternative transcripts with different protein products (Carninci et al., 2006). Among

the data set, CAGE data of four brain regions (cerebellum, parietal cortex, occipital cortex, and thalamus) of two adults and one newborn were analyzed (Table 3.9).

*Table 3.18. List of CAGE data downloaded from the FANTOM5 project for the analysis.*

Brain Region	Individual Donors
Cerebellum	Adult10196
	Adult10252
	Newborn 10223
Occipital Cortex	Adult10196
	Adult10252
	Newborn 10223
Parietal Lobe	Adult10196
	Adult10252
	Newborn 10223
Thalamus	Adult10196
	Adult10252
	Newborn 10223

CAGEr, a software package in Bioconductor in R (Gentleman et al., 2004; Haberle et al., 2015) (detailed script in Appendix 9) was used to analyze CAGE data. The data was saved in a CTSS format file, which is a tab separated file with names of chromosomes, 1 based coordinate of the TSSs, genomic strands and number of CAGE tags. Chromosomes where LTCC genes are located were extracted (Chr1 for *CACNAIS*, Chr3 for *CACNAID*, Chr12 for *CACNAIC*, and ChrX for *CACANAIF*). There are two quality control steps in CAGEr which involve removing TSSs with tags below a threshold and removing G bias. Once CAGE tags were mapped to the genome, supporting tag thresholds could be set so that regions with a low number of tags are removed. In the CAGE experiment protocol, an extra G nucleotide is added to the 5' end of the CAGE tag during cDNA preparation (Harbers & Carninci, 2005). The G bias can be corrected (Haberle et al., 2015) by removing the first nucleotide of the 5' tag if it is G or by using systematic probability after mapping the BLANK to the genome (Carninci et al., 2006). Removing the 5' end G nucleotide addition bias allows for more exact TSS position prediction (Haberle et al., 2015).

Once the dataset has a more precise definition of the 5' ends of the CAGE tags, TSSs and supporting tag counts can be identified. After the identification, raw tag counts are normalized to quantify and to enable a comparison of the expression level of TSSs between multiple samples (Haberle et al., 2015). In this analysis, a power-law distribution has been used to normalize the raw CAGE tag counts (Balwierz et al., 2009). After the normalization of raw tag counts, overlapping tags are clustered by grouping nearby tags on the same strand.

For the analysis in this chapter, CAGEr was used only to examine the location of all possible TSSs near the short transcript in the 3' region of LTCC genes. The pipeline defines TSSs as regions with more than 5 tags. However, since *CACNAIS* and *CACNAIF* have extremely low expression levels in the human brain, it was set to define all the sites with equal to or more than 1 CAGE tag as a TSS. CAGE experiments have confirmed that 5' ends with a single CAGE tag can be genuine (Carninci et al., 2006). After analysis of the CAGE data set using CAGEr, the TSSs of each LTCC gene from the RACE experiment and the CAGEr analysis results were aligned to provide orthogonal sources of support for TSS identification.

### **3.3. Results of 5' RACE experiment and CAGE analysis**

The aim of this chapter is to understand the characteristics of the short-length transcript from the 3' region of LTCC genes, i.e., whether they are complete transcripts with their own TSS, or fragments cleaved from the full-length transcripts. To achieve this aim, 5' RACE and CAGE data analyses were conducted. 5' RACE allowed us not only to examine whether the short-length transcripts from the 3' region of LTCC genes are complete transcripts or not, but also to examine how many TSSs exist where different transcripts might have arisen and the location of TSSs where the transcription starts. 5' RACE experiments successfully identified TSSs for each short-length LTCC transcript (section 3.3.1). Therefore, the short-length transcripts from the 3' region of LTCC

genes will be referred to as the 3' transcripts of LTCC genes from now on in this thesis. In this results section, TSSs identified (visualized at Figure 3.4 and 3.5. in section 3.3.1 and sequenced via sanger sequencing) were aligned to their respective genes in the UCSC genome browser (<http://genome.ucsc.edu>) and overlaid with CAGE results to look for convergence (section 3.3.2 – 3.3.5).

The structure of the 3' transcripts of LTCC genes were speculated based on the 5' RACE results. Each TSS of the 3' transcripts was aligned to CAGE analysis (black bars in Figure 3.6A, 3.7A, 3.8A, 3.9A). In each case, as detailed further below, the alignment confirms that the existence of a complete transcript with a novel TSS in the 3' region of LTCC genes in the human brain can also be seen from the genome-wide CAGE data.

### **3.3.1. Result of 5' RACE experiments**

#### **3.3.1.1. The possible TSSs of the short-length transcripts from the 3' region of LTCC genes**

##### **– positive finding for the short-length transcripts from the 3' region of *CACNA1S*.**

The possible TSSs from the 5' RACE experiments (Figure 3.3; See section 3.2.2.1 – 3.2.2.6) were visualized in 2% agarose gel (See section 3.2.2.6). Among many bands, the bands marked with squares were excised to be sequenced (Figure 3.5, Table 3.10; See section 3.2.2.7). These squares were selected to be sequenced because the bands were not detected in the negative control lanes (Figure 3.5, Table 3.10), whereby nuclease-free water was used as a template. However, among those sequenced products, only the band with a red square was the positive TSS which means they had the correct gene-specific location (See Appendix 8). The reason why there were some bands in the negative control lanes might be that the primers used for outer PCR amplification bound to each other, since as they were diluted, the bands faded away (Figure 3.5, Table 3.10).

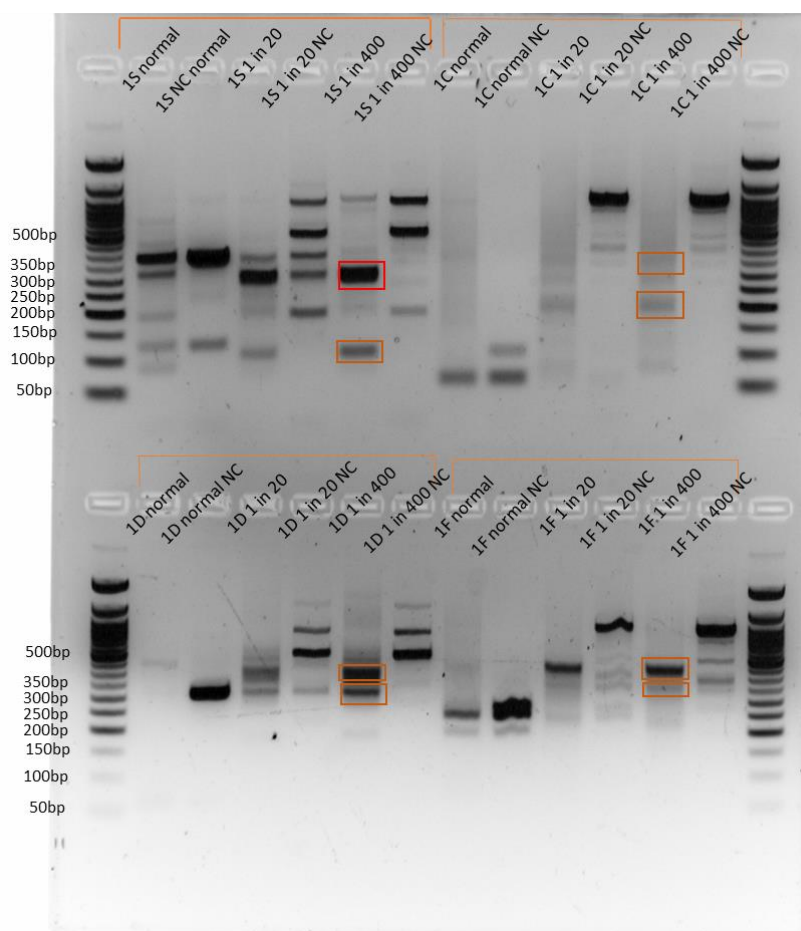


Figure 3.5. The 5' RACE products of short-length transcript from the 3' region of LTCC genes. The name of each lane is shown in Table 3.10. The band with the red square is the right TSS and other bands with orange colors were sequenced and turned out not to be specific to the gene. Those bands were chosen to be sequenced because they did not appear in the negative control lanes which means they are not a PCR artifact, but rather products amplified from the second round of PCR.

Table 3.19. Name of each lane for figure 3.4

Name on the gel	The gene which the target short-transcript is from	Template	Dilution
1S normal	<i>CACNA1S</i>	outer PCR product with cDNA	No dilution
1S NC normal	<i>CACNA1S</i>	outer PCR product with NFH <sub>2</sub> O	No dilution

1S 1 in 20	<i>CACNAIS</i>	outer PCR product with cDNA	1 in 20
1S 1 in 20 NC	<i>CACNAIS</i>	outer PCR product with NFH <sub>2</sub> O	1 in 20
1S 1 in 400	<i>CACNAIS</i>	outer PCR product with cDNA	1 in 400
1S 1 in 400 NC	<i>CACNAIS</i>	outer PCR product with NFH <sub>2</sub> O	1 in 400
1C normal	<i>CACNAIC</i>	outer PCR product with cDNA	No dilution
1C normal NC	<i>CACNAIC</i>	outer PCR product with NFH <sub>2</sub> O	No dilution
1C 1 in 20	<i>CACNAIC</i>	outer PCR product with cDNA	1 in 20
1C 1 in 20 NC	<i>CACNAIC</i>	outer PCR product with NFH <sub>2</sub> O	1 in 20
1C 1 in 400	<i>CACNAIC</i>	outer PCR product with cDNA	1 in 400
1C 1 in 400 NC	<i>CACNAIC</i>	outer PCR product with NFH <sub>2</sub> O	1 in 400
1D normal	<i>CACNAID</i>	outer PCR product with cDNA	No dilution
1D normal NC	<i>CACNAID</i>	outer PCR product with NFH <sub>2</sub> O	No dilution
1D 1 in 20	<i>CACNAID</i>	outer PCR product with cDNA	1 in 20
1D 1 in 20 NC	<i>CACNAID</i>	outer PCR product with NFH <sub>2</sub> O	1 in 20
1D 1 in 400	<i>CACNAID</i>	outer PCR product with cDNA	1 in 400
1D 1 in 400 NC	<i>CACNAID</i>	outer PCR product with NFH <sub>2</sub> O	1 in 400
1F normal	<i>CACNAIF</i>	outer PCR product with cDNA	No dilution
1F normal NC	<i>CACNAIF</i>	outer PCR product with NFH <sub>2</sub> O	No dilution
1F 1 in 20	<i>CACNAIF</i>	outer PCR product with cDNA	1 in 20
1F 1 in 20 NC	<i>CACNAIF</i>	outer PCR product with NFH <sub>2</sub> O	1 in 20
1F 1 in 400	<i>CACNAIF</i>	outer PCR product with cDNA	1 in 400
1F 1 in 400 NC	<i>CACNAIF</i>	outer PCR product with NFH <sub>2</sub> O	1 in 400

### 3.3.1.2. The possible TSSs of the short-length transcripts from the 3' region of LTCC genes

#### – positive findings for the short-length transcripts from the 3' region of *CACNAIC*, *CACNAID*, and *CACNAIF*

Since the TSSs of short-length transcripts from the 3' region of *CACNAIC*, *CACNAID*, and *CACNAIF* were not identified from the previous attempt (Figure 3.5), another attempt was performed with more stringent dilution of the outer PCR products. Before performing the outer PCR, gene-specific cDNAs were diluted. We have found that diluting cDNA increases the specificity of outer PCR products because there are less cDNA products for primers to bind to so that primers won't bind to random sequences that have higher expression in the human brain than the sequence of interest. If the specificity of the outer PCR is increased, the inner PCR results should also show higher specificity to the gene of interest. In some cases, outer PCR products were diluted even more than 1 in 400 to increase specificity.

Figure 3.6 shows the results of the inner PCR to examine TSSs of short-length transcripts from the 3' region of *CACNA1C*, *CACNA1D*, and *CACNA1F* which were run on a 2% agarose gel prepared in 100ml of 1x TAE with 10 $\mu$ l ethidium bromide (5mg/ml). Like the previous experiment, bands marked with squares were excised to be sequenced (Figure 3.6, Table 3.11; See section 3.2.2.7) because the bands were detected only in the lanes with template and not in the lanes without template. It is assumed that there are some bands in lane C, which is the negative control, due to outer PCR primers binding to each other, resulting in several nucleotide aggregates. Once diluted, they tended to fade away. Red square highlighted bands were verified as the correct target using Sanger sequencing, whereas orange squared band were confirmed as non-specific amplification. Therefore, red squared bands are the TSSs of short-length transcripts from the 3' region of *CACNA1C*, *CACNA1D*, and *CACNA1F* (see Appendix 8).



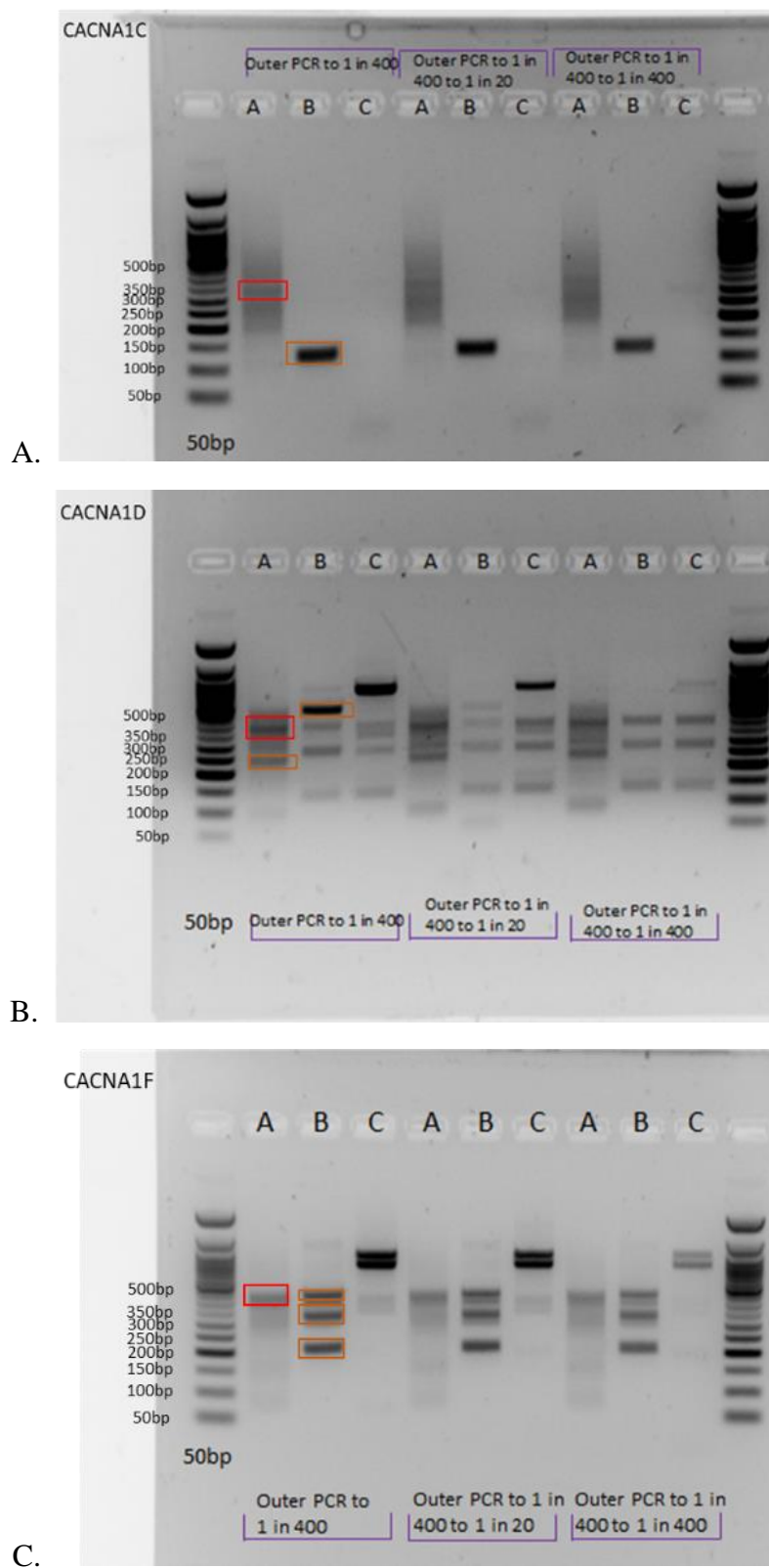


Figure 3.6. The 5' RACE products of short-length transcript from the 3' region of *CACNA1C*, *CACNA1D*, and *CACNA1F* A. The possible TSSs for the short-length transcript from the 3' region of *CACNA1C*. The

squared bands were sequenced, and the red squared band was the positive finding, whilst the orange squared bands were not specific to *CACNA1C*. **B.** The possible TSSs for the short-length transcript from the 3' region of *CACNA1D*. The squared bands were sequenced, and the red squared band was the positive finding and the orange squared bands were not specific to *CACNA1D*. **C.** The possible TSSs for the short-length transcript from the 3' region of *CACNA1F*. The squared bands were sequenced, and the red squared band represented the positive finding, whilst the orange squared bands were not specific to *CACNA1F*. The lane names are described in Table 3.9.

Table 3.20. Name of each lane for Figure 3.5

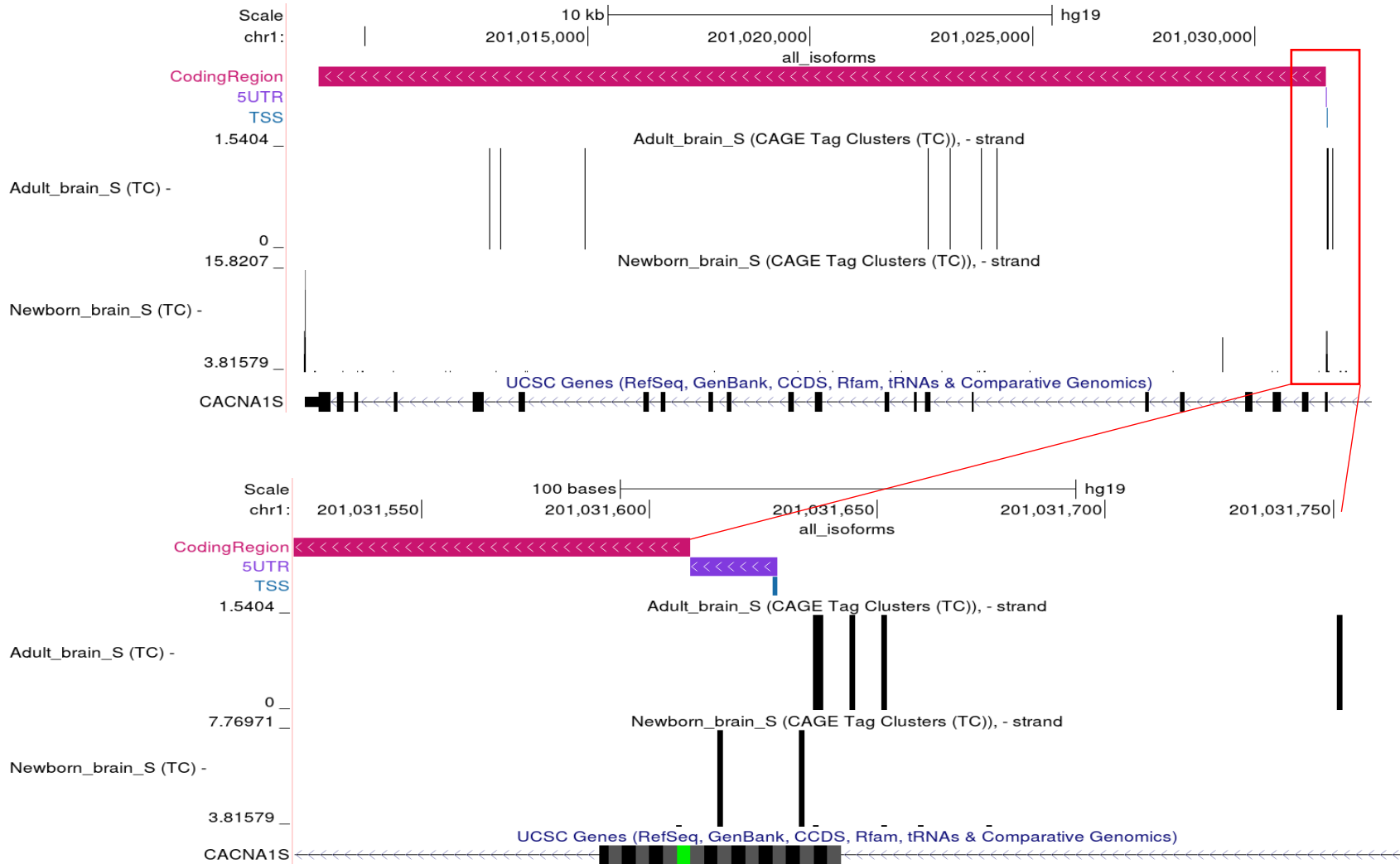
Name on the gel	cDNA for outer PCR
A	Non-diluted cDNA
B	cDNA diluted to 1 in 20
C	NF H <sub>2</sub> O
Name on the gel	Template to Inner PCR
Outer PCR to 1 in 400	Outer PCR product is diluted to 1 in 400
Outer PCR to 1 in 400 to 1 in 20	Outer PCR product is diluted to 1 in 8000
Outer PCR to 1 in 400 to 1 in 400	Outer PCR product is diluted to 1 in 160000

### 3.3.2. 5' RACE and CAGE analysis of *CACNA1S* 3' transcript

The *CACNA1S* 3' transcript alignment is shown in Figure 3.6 (Top bars in Figure 3.7A; the transcript runs from the right to the left). The blue bar labelled "TSS" represents the first nucleotide identified by the 5' RACE experiment, which is C, where transcription of 3' transcript of *CACNA1S* starts (genomic coordinates: chr1:201031628, hg19). The purple bar labeled "5UTR" is the untranslated region of the 3' transcript of *CACNA1S* (genomic coordinates: chr1:201031610-2010316278, hg19). The pink bar labeled "CodingRegion" is the coding region of the 3' transcript of *CACNA1S* starting where the first start codon is (genomic location: chr1:201008960-201031609, hg19). The 3' transcript of *CACNA1S* is expected to encode from the loop between S5 and S6 of repeat III to the cytoplasmic C-terminal region of Cav1.1 (Figure 3.7B). The 3' transcript of *CACNA1S* might not encode the proper functional calcium channel

however, as the  $\alpha 2\delta$  subunit binds to the region which could affect the function of the 3' transcript of *CACNA1S*. The alignment of the TSS from 5' RACE and CAGE analysis (Black bars in Figure 3.7A) shows that there is a single well-defined TSS in the C-terminus transcript of *CACNA1S*. The black bars represent normalized counts of clustered tags.

A.



B.

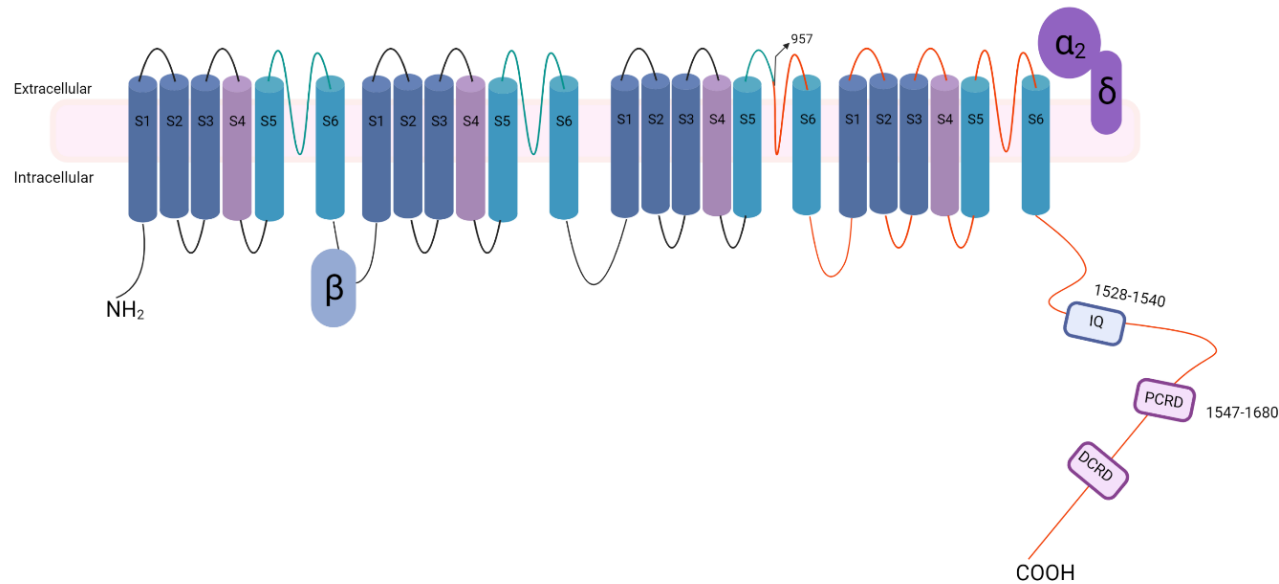
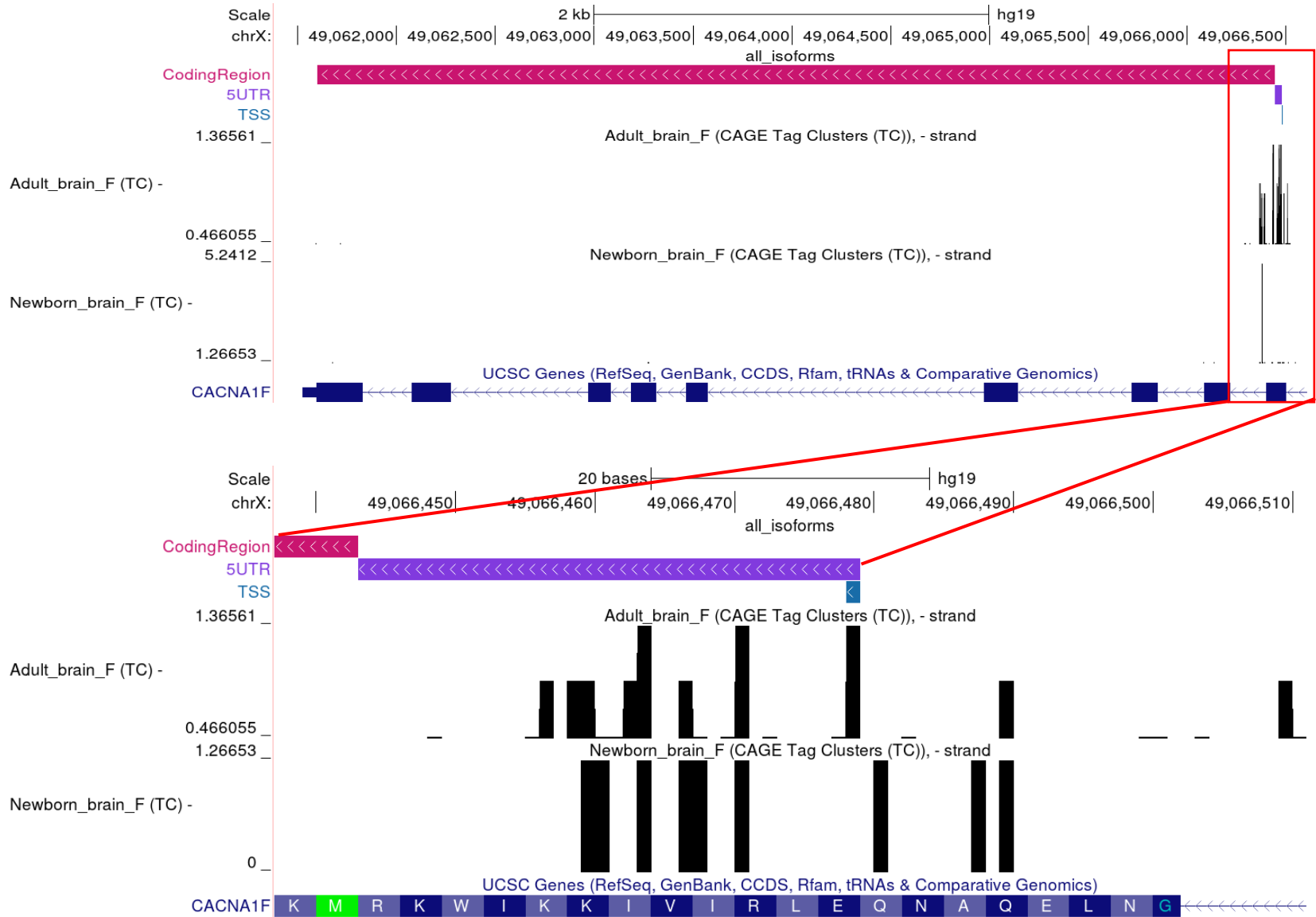


Figure 3.7. TSS of the 3' transcript of *CACNAIS* with CAGE analysis and its location in the channel structure. **A** UCSC genome browser, where the 5' RACE result of 3' transcript of *CACNAIS* and the CAGE analysis were aligned. The blue bar labelled "TSS" represents the first nucleotide, which is C, where transcription of 3' transcript in *CACNAIS* starts (genomic coordinates: chr1:201031628, hg19). The purple bar labeled "5UTR" is the untranslated region of the 3' transcript of *CACNAIS* (genomic coordinates: chr1:201031610-2010316278, hg19). The pink bar labeled "CodingRegion" is the coding region of the 3' transcript of *CACNAIS* (genomic location: chr1:201008960-201031609, hg19). The black bars represent normalized counts of clustered tags. **B** A schematic representation of what the 3' transcript of *CACNAIS* might encode. The 3' transcript of *CACNAIS* encodes from amino acid residue 957 in the full-length channel. EF is EF hands binding motifs, IQ is IQ domain where  $\text{Ca}^{2+}$ /Calmodulin binds. PCRD is proximal c-terminal regulatory domain and DCRD is distal c-terminal regulatory domain.

### 3.3.3. 5'RACE and CAGE analysis of *CACNA1F* 3' transcript

The 3' transcript of *CACNA1F* with the identified TSS from the 5' RACE is aligned to the UCSC genome browser (<http://genome.ucsc.com>; top bar in Figure 3.8A; the transcript runs from the right to the left). The blue bar labelled "TSS" represents the first nucleotide, C, where transcription of the 3' transcript in *CACNA1F* starts (the genomic coordinates: chrX:49066479; hg19). The purple bar labelled "5UTR" is the 5' untranslated region of the 3' transcript of *CACNA1F* (genomic coordinates: chrX:49066444-49066479, hg19). The pink bar labelled "CodingRegion" represents the coding region of the 3' transcript of *CACNA1F* starting where the first start codon is (genomic coordinates: chrX:49061598-49066443, hg19). The 3' transcript of *CACNA1F* is thought to produce a complete protein in the C-terminal region of *Cav1.4* (Figure 3.8B). The alignment of the TSS of the 3' transcript of *CACNA1F* between 5' RACE and CAGE analysis shows that there is a weakly defined yet dominant TSS in the 3' transcript of *CACNA1F* (Black bars in the Figure 3.8A). The black bars represent normalized counts of clustered tags.

A.



B.

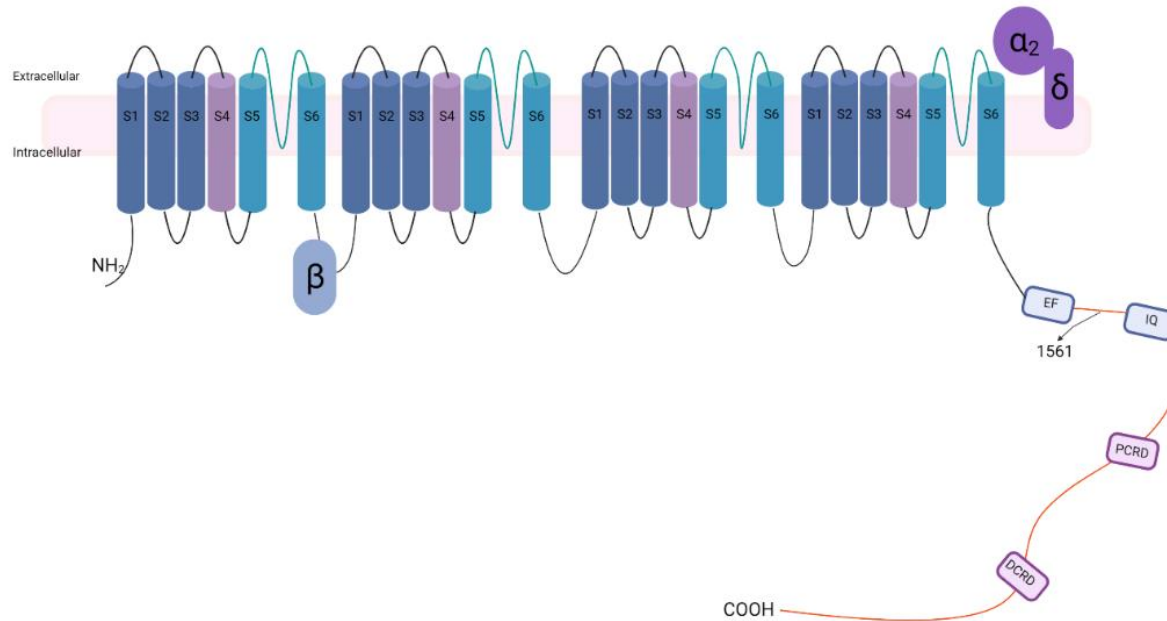


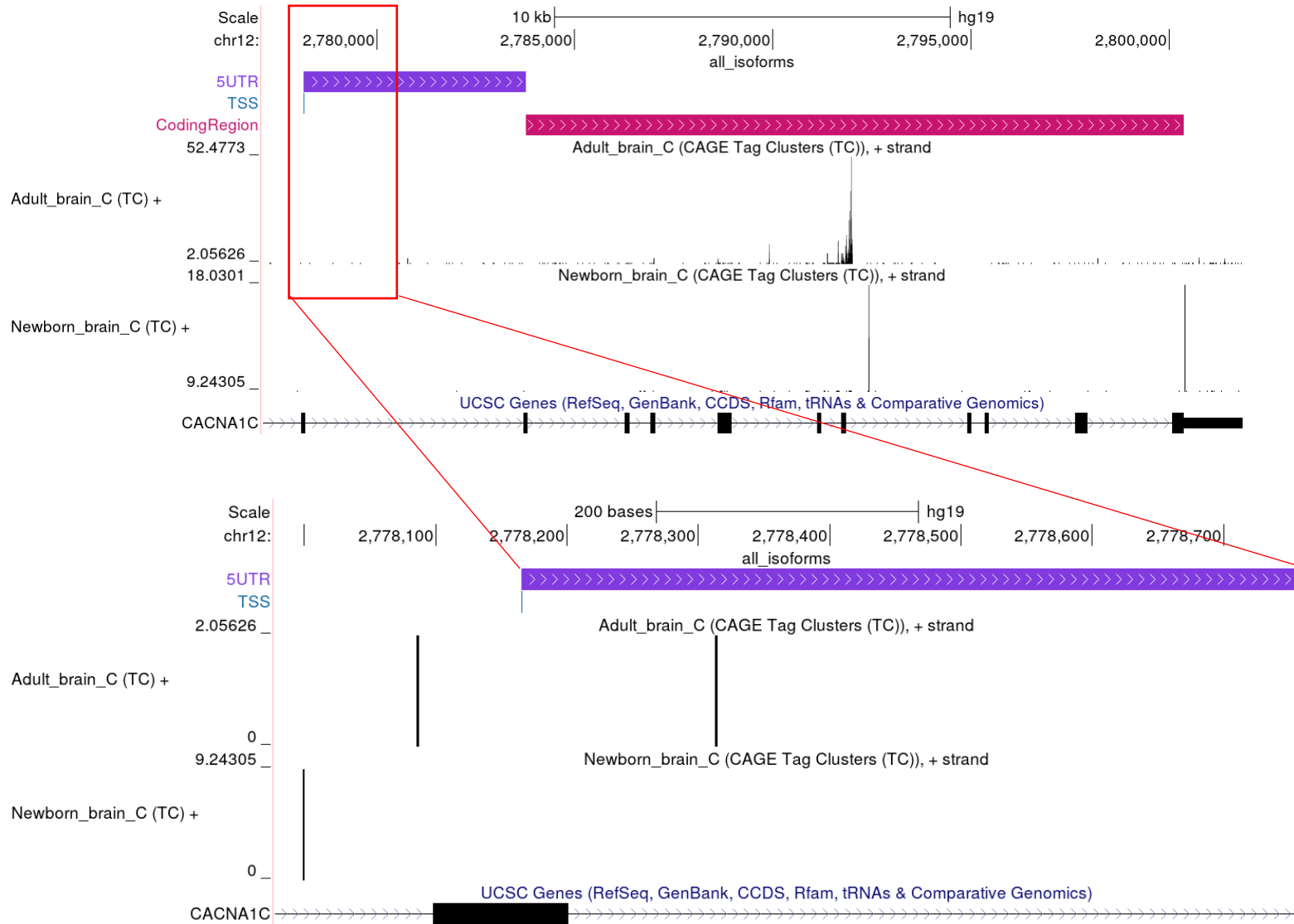
Figure 3.8. TSS of the 3' transcript of *CACNA1F* with CAGE analysis and its location in the channel structure. **A** UCSC genome browser which the 5' RACE result of 3' transcript of *CACNA1F* and CAGE analysis were aligned. The blue bar labelled "TSS" represents the first nucleotide, C, where transcription of the 3' transcript in *CACNA1F* starts (the genomic coordinates for the TSS are chrX:49066479; hg19). The purple bar labelled "5UTR" is the 5' untranslated region of the 3' transcript of *CACNA1F* (genomic coordinates: chrX:49066444-49066479, hg19). The pink bar labelled "CodingRegion" represents the coding region of the 3' transcript of *CACNA1F* (genomic coordinates: chrX:49061598-49066443, hg19; The black bars represent normalized counts of clustered tags). **B** A schematic representation of the 3' transcript of *CACNA1F*. The 3' transcript of *CACNA1F* encodes from amino acid residue 1561 in the full-length channel. EF is EF hands binding motifs, IQ is IQ domain where Ca<sup>2+</sup>/Calmodulin binds. PCRD is proximal c-terminal regulatory domain and DCRD is distal c-terminal regulatory domain.



### 3.3.4. 5'RACE and CAGE analysis of *CACNA1C* 3' transcript

The 3' transcript of *CACNA1C* with the identified TSS from 5' RACE was aligned to the UCSC genome browser (top bar in Figure 3.9A, the transcript runs from the left to the right). The blue bar labelled as TSS is the first nucleotide, which is A, where transcription of the 3' transcript in *CACNA1C* starts (genomic coordinates chr12:2778166, hg19). The purple bar labelled as 5UTR is the untranslated region of the 3' transcript of *CACNA1C* (genomic coordinates: chr12:2778166-2783771, hg19). Lastly, the pink bar labelled as CodingRegion represents the coding region of the 3' transcript of *CACNA1C* starting from the first start codon (genomic coordinates: chr12:2783773-2800365, hg19). The coding region is thought to produce protein in the C-terminal region of *Cav1.2*, which is encoded by the full-length *CACNA1C* (Figure 3.9B). The alignment of the TSS of *CACNA1C* between the 5' RACE and CAGE analysis shows that there is a very well-defined TSS in the 3' transcript of *CACNA1C*. The black bars represent normalized counts of clustered tags (Black bars in the Figure 3.9A).

A.



B.

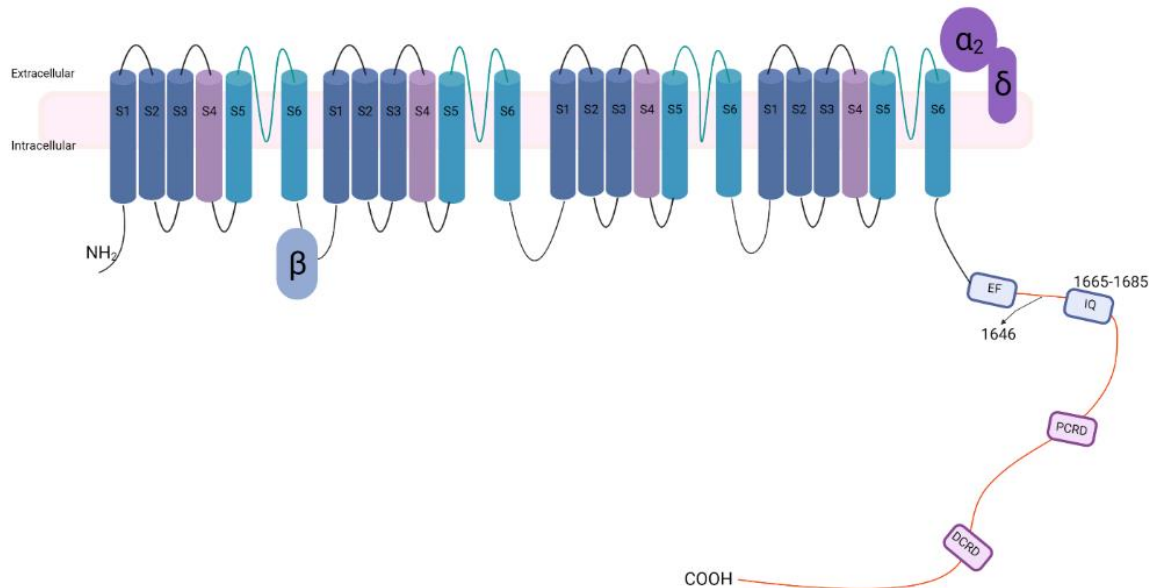
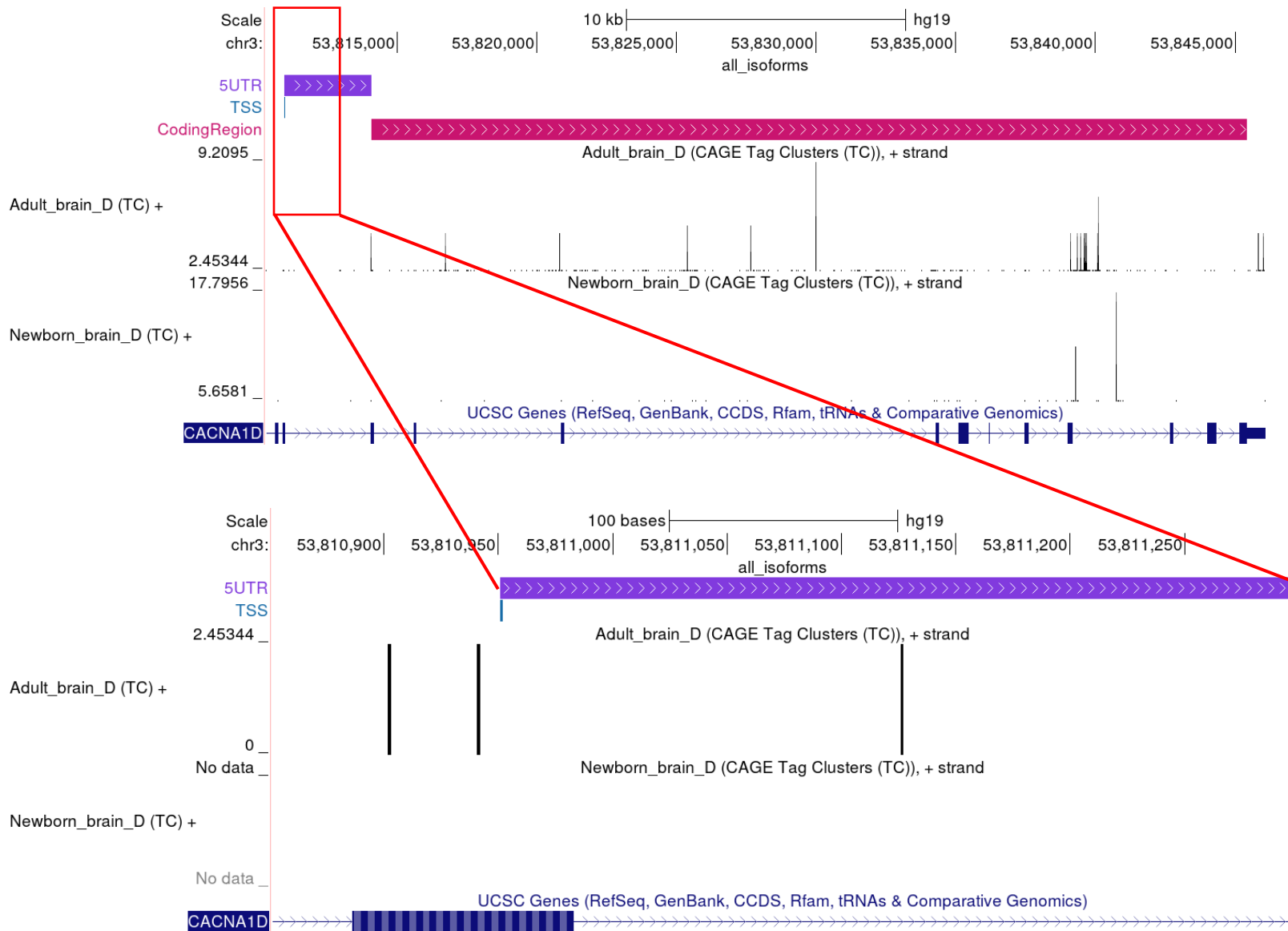


Figure 3.9. TSS of the 3' transcript of CACNA1C with CAGE analysis and its location in the channel structure. **A** UCSC genome browser showing the aligned 5' RACE result of 3' transcript of CACNA1C and CAGE analysis. The blue bar labelled as TSS is the first nucleotide, which is A, where transcription of the 3' transcript in CACNA1C starts (genomic coordinates chr12:2778166, hg19). The purple bar labelled as 5UTR is the untranslated region of the 3' transcript of CACNA1C (genomic coordinates: chr12:2778166-2783771, hg19). Lastly, the pink bar labelled as CodingRegion represents the coding region of the 3' transcript of CACNA1C (genomic coordinates: chr12:2783773-2800365, hg19; the black bars represent normalized counts of clustered tags). **B** A schematic representation of the C-terminus transcript of CACNA1C. The 3' transcript of CACNA1C encodes from amino acid residue 1646 in the full-length channel. EF is EF hands binding motifs, IQ is IQ domain where Ca<sup>2+</sup>/Calmodulin binds. PCRD is proximal c-terminal regulatory domain and DCRD is distal c-terminal regulatory domain.

### 3.3.5. 5'RACE and CAGE analysis of *CACNAID* 3' transcript

The 3' transcript of *CACNAID* with identified TSS from 5' RACE is aligned to the UCSC genome browser (top bar in Figure 3.10A, the transcript runs from the left to the right). The blue bar labelled "TSS" represents the first nucleotide, G, where transcription of the 3' transcript in *CACNAID* starts (genomic coordinate of the TSS is chr3:53810952, hg19). The purple bar labelled "5UTR" represents the 5' untranslated region of the 3' transcript of *CACNAID* (genomic coordinates: chr3:53810952-53814064, hg19). The pink bar labelled "CodingRegion" represents the coding region of the 3' transcript of *CACNAID* (genomic coordinates: chr3:53814065-53845433, hg19). The region is expected to produce a complete protein in the C-terminal region of the *Cav1.3*, which is encoded by the full-length *CACNAID* (Figure 3.10B). The alignment of TSS of *CACNAID* between 5' RACE and CAGE analysis shows that the TSS in the 3' transcript of *CACNAID* is well-defined (Black bars in the Figure 3.10A). The black bars represent normalized counts of clustered tags.

A.



B.

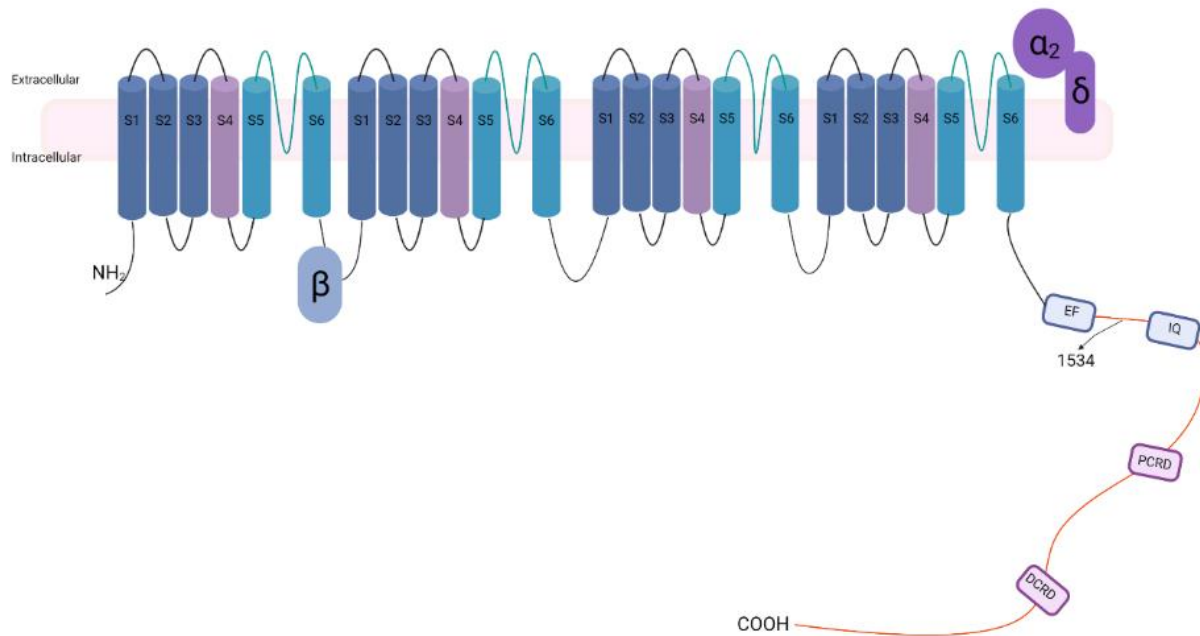


Figure 3.10. TSS of the 3' transcript of *CACNA1D* with CAGE analysis and its location in the channel structure. **A** UCSC genome browser depicting the aligned 5' RACE result of 3' transcript of *CACNA1D* and CAGE analysis. The blue bar labelled "TSS" represents the first nucleotide, G, where transcription of the 3' transcript in *CACNA1D* starts (genomic coordinate of the TSS is chr3:53810952, hg19). The purple bar labelled "5UTR" represents the 5' untranslated region of the 3' transcript of *CACNA1D* (genomic coordinates: chr3:53810952-53814064, hg19). The pink bar labelled "CodingRegion" represents the coding region of the C-terminus transcript of *CACNA1D* starting from the first start codon (genomic coordinates: chr3:53814065-53845433, hg19; the black bars represent normalized counts of clustered tags). **B** A schematic representation of the 3' of *CACNA1D*. The 3' transcript of *CACNA1D* encodes from amino acid residue 1534 in the full-length channel. EF is EF hands binding motifs, IQ is IQ domain where  $\text{Ca}^{2+}$ /Calmodulin binds. PCRD is proximal c-terminal regulatory domain and DCRD is distal c-terminal regulatory domain.



*Figure 3.11 Multiple Sequence Alignment analysis of 3' transcripts of LTCC genes. A shows the guidetree of sequence alignment analysis of 3' transcripts of LTCC genes. B shows the nucleotide alignment among 3' transcripts of LTCC genes. The transcription start site of 3' transcript of CACNA1F is marked with a red arrow.*

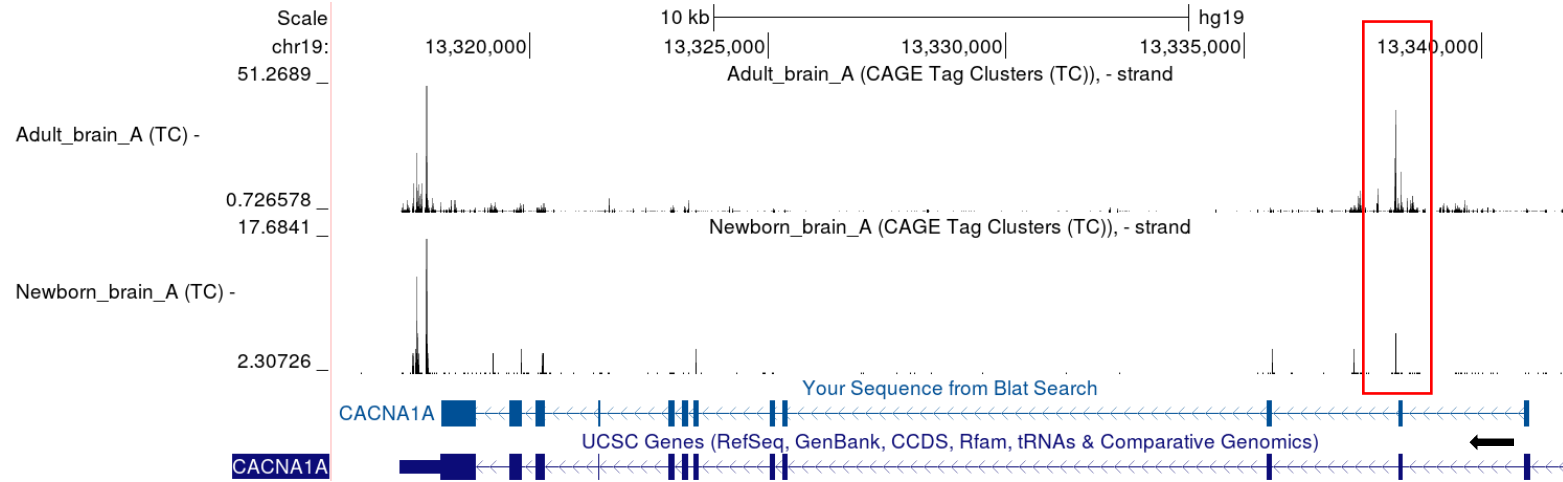
### **3.4. The existence of alternative TSS in the C-terminal region of VGCC $\alpha$ 1 subunits other than LTCCs.**

The data above show that alternative TSS usage to produce short transcripts is observed across LTCC genes. Therefore, I investigated whether evidence for the same pattern for the other VGCC  $\alpha$ 1 subunit genes could be seen in the CAGE data (Figure 3.12).

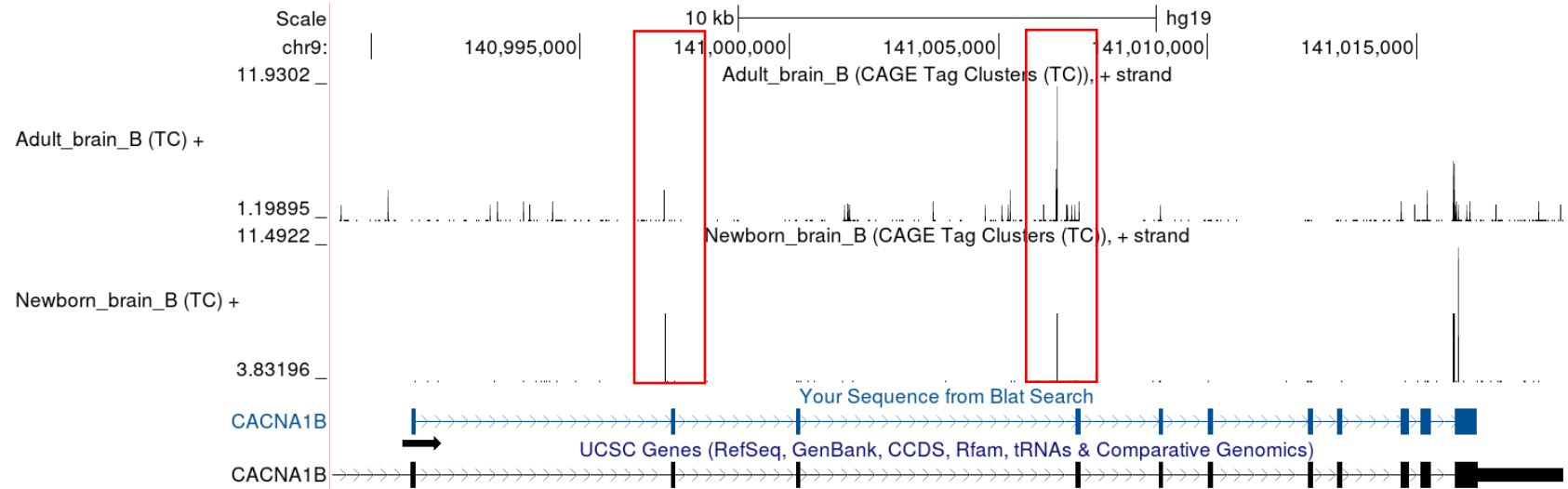
All the other VGCC  $\alpha$ 1 genes showed some evidence for the presence of 3' TSSs, indicated with red boxes (Figure 3.12). This highlights that the presence of short transcripts encoded by 3' TSSs may be a feature shared across VGCC genes.



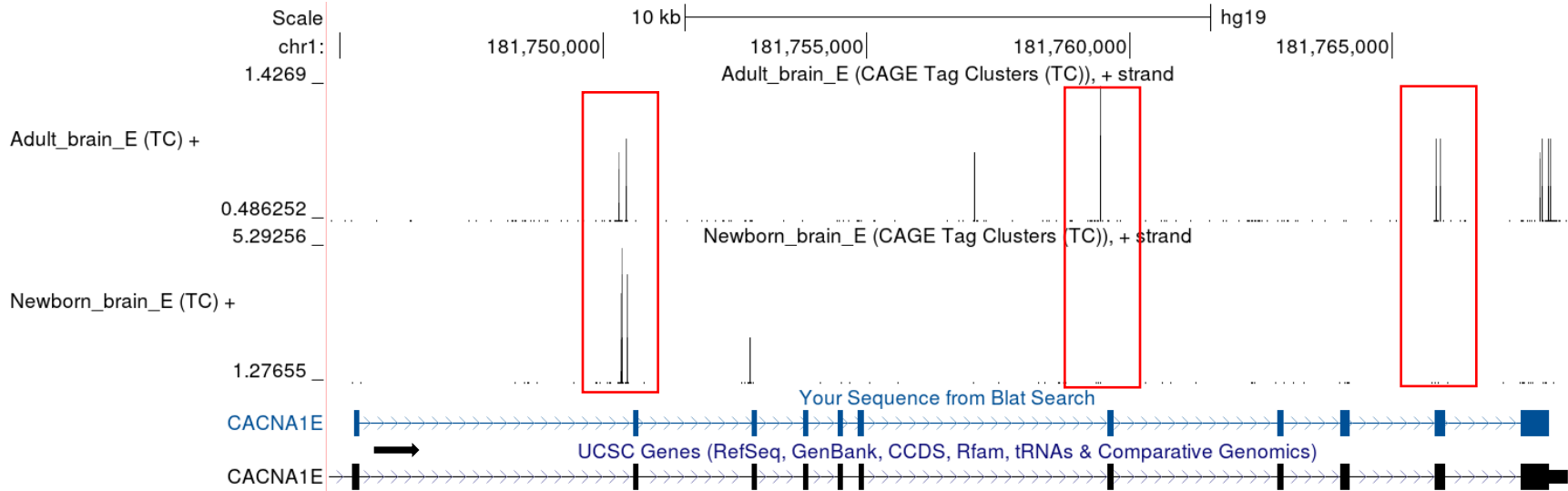
A.



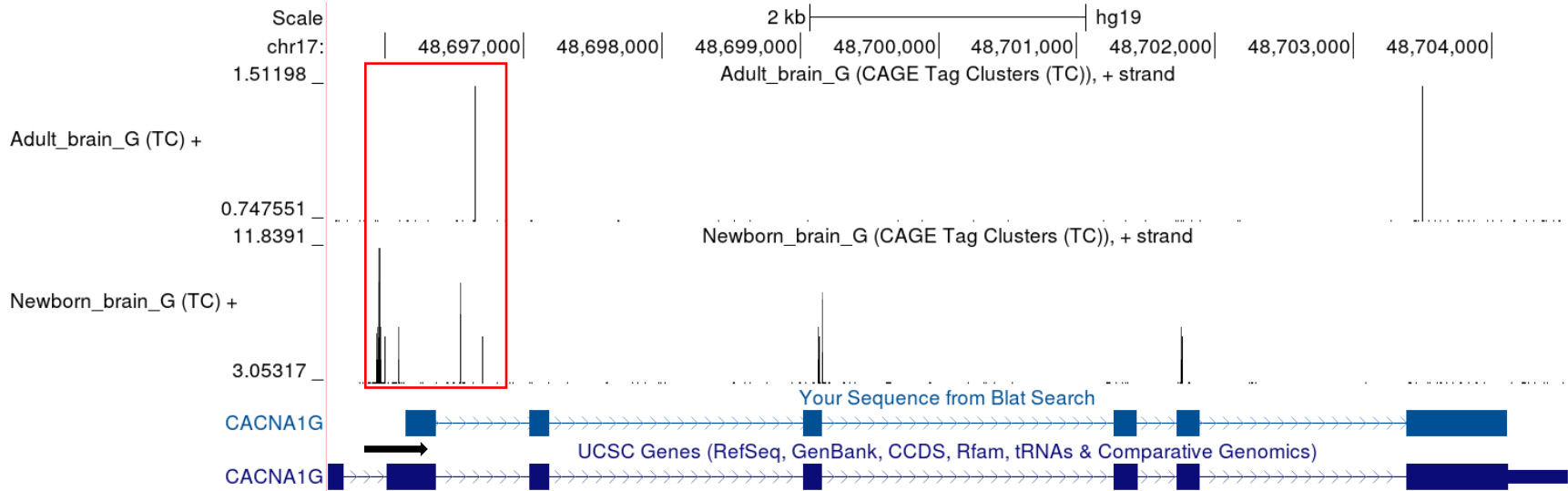
B.



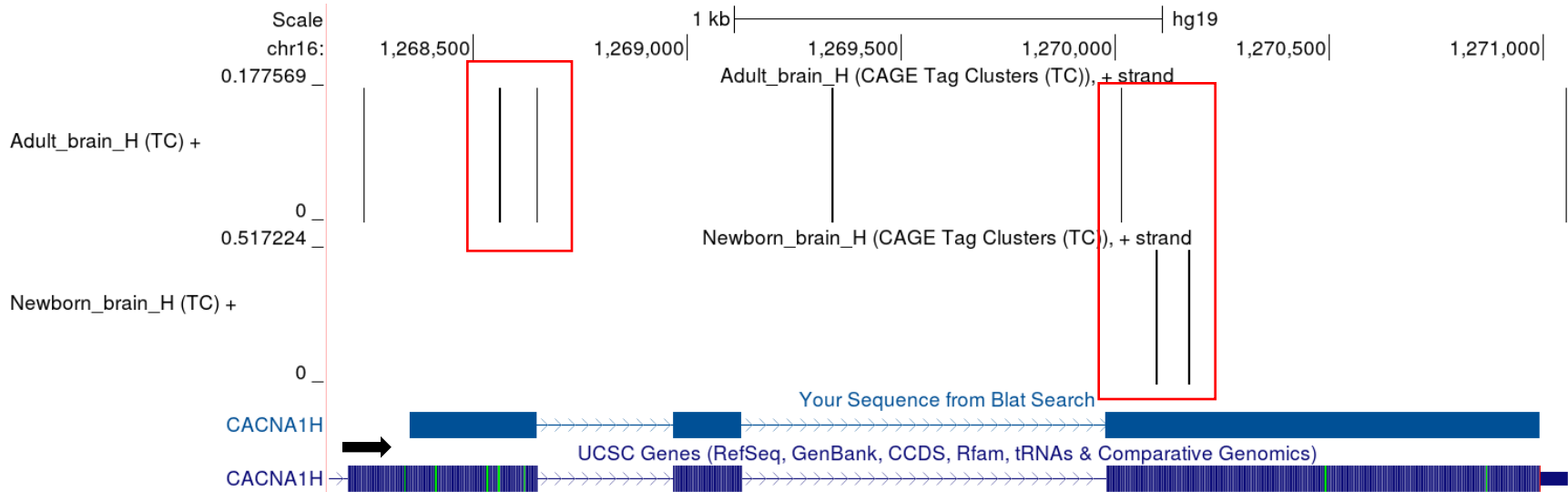
C.



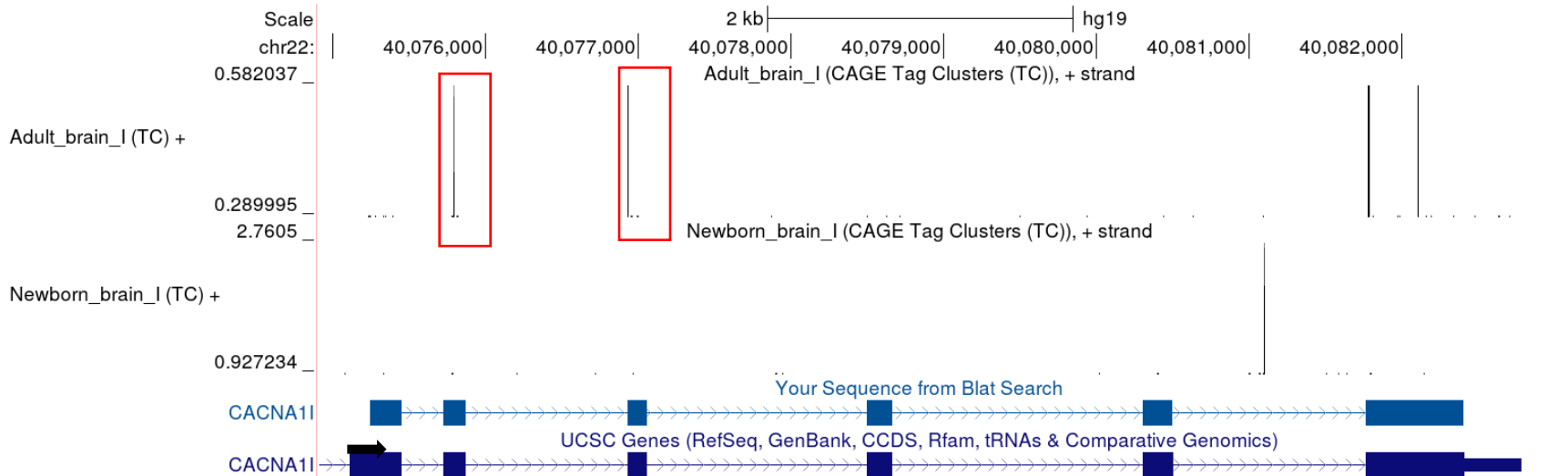
D.



E.



F.



*Figure 3.12 UCSC genome browser of CAGE analysis was aligned to the C-terminal region of each VGCC  $\alpha 1$  subunits genes. **A** Possible alternative TSSs in the C-terminal region of CACNA1A from the CAGE data. **B** Possible alternative TSSs in the C-terminal region of CACNA1B from the CAGE data. **C** Possible alternative TSSs in the C-terminal region of CACNA1E from the CAGE data. **D** Possible alternative TSSs in the C-terminal region of CACNA1G from the CAGE data. **E** Possible alternative TSSs in the C-terminal region of CACNA1H from the CAGE data. **F** Possible alternative TSSs in the C-terminal region of CACNA1I from the CAGE data.*

### 3.5. Discussion

It is common for one gene to have several different TSSs, allowing transcripts with different TSSs to produce proteins with distinctive structure and functions (Landry et al., 2003). Even among VGCCs, there are prior studies showing that one gene produces several transcripts or truncated channels which are originated from alternative TSSs (Dai et al., 2002; Gomez-Ospina et al., 2013; Okagaki et al., 2001; Pang et al., 2003). These studies mainly focused on *CACNA1C*; however, this chapter shows that other LTCC genes, such as *CACNA1S* and *CACNA1F*, produce short-length transcripts using an alternative TSS in the 3' region of each gene. Even though the full-length sequences of *CACNA1S* and *CACNA1F* are not expressed in the human brain (McRory et al., 2004; Zamponi et al., 2015), my 5'RACE and CAGE analysis validated the existence of short-length transcripts from the 3' region of each gene in the human brain as intact transcripts arising from de novo TSSs. The experimental results in this chapter suggest that 3' transcripts of LTCC genes are derived from novel TSSs within the gene: these were identified using 5'RACE, confirming the presence of TSSs in the region, and supported by genome-wide analysis using CAGE data.

Multiple sequence alignment analysis revealed the similarity among the 3' transcripts of LTCC genes. The 3' transcripts of *CACNA1C* and *CACNA1D* showed the most similarity (Figure 3.11A), which is concordance with previous studies identifying similarities among full-length LTCCs (Dolphin, 2006; Feng et al., 2018). Where transcription of the 3' transcript of *CACNA1F* starts showed high similarity among 3' transcripts of LTCC genes. This could be because the region may have a distinctive function such as binding motifs (Sofi et al., 2022).

Even though these 3' transcripts have low expression levels, their expression was still detectable across the whole transcriptome in the human brain. Furthermore, examining the TSSs

of short transcripts of *CACNA1C* and *CACNA1D* that were not sequenced in chapter 2 suggests that such transcripts occur within all LTCC genes. Finally, the analysis of the CAGE data suggests that having alternative TSSs in the 3' region is a phenomenon observed across all VGCC  $\alpha 1$  subunit genes.

Similar findings are observed from other ion channel families. Multiple sodium channels contain alternative TSSs including *SCN1A*, *SCN2A*, *SCN3A* (Martin et al., 2007), *SCN5A* (Shang & Dudley Jr, 2005), and *SCN8A* (Drews et al., 2005). *SCN8A*, which encodes the Nav1.6 channel, was known to produce four different transcripts with mutually exclusive 5' untranslated exons (exon 1a, exon 1b, exon 1c, and exon 1d) transcribed from different TSSs (Drews et al., 2005, 2007) of which untranslated exon 1c is conserved across human, mouse, chicken and zebrafish. This implies that the untranslated exon 1c might contain functional regions such as transcription binding sites (Drews et al., 2007). Conservation is frequently correlated with functional genetic elements (Pennacchio & Visel, 2010), so similar to sodium channels, a high conservation score across mammalian species of 5' UTRs for the 3' transcripts of each LTCC gene could also hint at functional regions. Therefore, in chapter 4, to what extent the 5'UTR of LTCC 3' transcripts are conserved across 240 mammals will be examined to understand the possibility of 5'UTR having functional regions.

Multiple transcripts produced from one gene using alternative TSSs may be expressed in a brain-region specific manner (Martin et al., 2007). Compared with the expression of full-length *SCN1A* in 9 mouse brain region, the variants with exon 1a, exon 1b and exon 1c showed different expression levels in a brain-region specific manner (Martin et al., 2007). Similar to sodium channels, which use alternative TSSs to regulate gene expression in brain tissues, the LTCCs may also use alternative TSSs to produce 3' transcripts that have a tissue-specific expression pattern.

Therefore, in chapter 4, the expression patterns of 3' transcripts of LTCC genes across human tissues will be examined.

Examining TSSs of 3' transcripts of LTCC genes helps to elucidate the origin of transcripts sequenced in chapter 2. Sequencing 3' transcripts of *CACNAIS* and *CACNAIF* didn't elucidate whether they are cleaved from the full-length transcripts, or whether they are complete transcripts derived from the novel TSSs. This chapter demonstrated that the latter situation is the case. It is evident that the 3' transcript of LTCC genes are complete transcripts, which may produce intact proteins of potential functional significance even though they may not produce functional channels. Previous studies reported that the CCAT (Gomez-Ospina et al., 2013) translocates to the nucleus in neurons and functions as a transcription factor. Additionally, non-functional three domain channels reduce the expression of full-length channel in *Xenopus* oocytes (Okagaki et al., 2001). Therefore, in chapter 5, the localization of the 3' transcripts of LTCC genes in both non-excitabile and excitabile cells will be investigated.

### **3.6. Limitations of the experiments**

The experiments in this chapter have a limited usage of samples, i.e., not having positive controls such as skeletal muscle and retina. *CACNAIS* and *CACNAIF* were studied to be predominantly expressed in the skeletal muscle and retina, respectively (Catterall et al., 2005; McRory et al., 2004; Zamponi et al., 2015). Therefore, the experiments in this chapter did not elucidate whether the identified TSSs of 3' transcripts of *CACNAIS* and *CACNAIF* are brain-specific TSSs or can be found in different tissues too. For example, *CACNAIC* produces tissue-specific transcripts from the alternative TSSs (Dai et al., 2002; Pang et al., 2003; Saada et al., 2005). Likewise, the full-length *CACNAIS* and the full-length *CACNAIF* could also produce 3' transcripts only in specific tissues. Moreover, like *SCN1A*, which transcripts with alternative

TSSs show brain-region-specific expressions (Martin et al., 2007), the 3' transcripts of the LTCC gene could also show brain-region-specific expression patterns. Therefore, having more than two brain regions will provide in-depth information on whether the 3' transcripts of LTCC genes express in all brain regions.

In order to overcome the limitation of not having a positive control, I will examine the tissue-specific expression pattern of the 3' transcripts of LTCC will be examined using exon expression level data from GTEx in Chapter 4. However, GTEx does not have exon expression level data in a retina, meaning that the existence of 3' transcripts of *CACNA1F* cannot be examined in the retina. Therefore, having additional tissues (skeletal muscle and retina) or additional brain regions will allow the experiments in this chapter stronger.



## **Chapter 4: Investigating possible functionality of 5'UTR and expression of the 3' transcripts**

### **4.1. Rationale for studying conservation status and tissue specific expression**

The previous chapters investigated the existence of novel transcripts arising from the 3' region of LTCC genes in the human brain. Specifically, chapter 2 identified the diversity of short-length transcripts produced from the 3' region of *CACNAIS* and *CACNAIF* in the human brain. Chapter 3 supported the hypothesis that short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* are complete transcripts originating from the novel TSSs within the *CACNAIS* and *CACNAIF* genes. Moreover, I confirmed the presence of truncated isoforms of *CACNAIC* and *CACNAID* arising from TSSs in their 3' regions. These were also supported by the analysis of publicly available genome-wide CAGE data. This chapter aims to investigate to examine the possible functionality of the 5'UTR of the 3' transcripts of LTCC genes and assess evidence of the 3' transcripts' expressions in human tissues.

### **4.2. Introduction**

For a gene to encode a protein in a cell, two major steps are needed: transcription and translation. Transcription involves the transfer of information contained within DNA to a messenger RNA (mRNA) molecule, whilst translation describes the production of proteins from the RNA template (Clancy & Brown, 2008). Initially, pre-mRNA is created, which contains both intronic and exonic sequences. Pre-mRNA undergoes splicing, during which the introns, non-coding regions, are removed, and exons, coding regions, are arranged into mature transcripts. Subsequently, the mature mRNA is translated into a protein (Abbasi et al., 2011). However, not all of the mature mRNA encodes the protein sequence: a 5' UTR lies upstream of the methionine

start site, and a 3' untranslated sequence (the 3' UTR) lies downstream of the stop codon (Clancy & Brown, 2008). The 5'UTR is a non-coding genomic region which stretches from the upstream of the beginning of coding sequences (CDS) until the TSS (Goss & Domashevskiy, 2016). The 5' UTR is a regulatory region which has a significant impact on transcription regulation by interacting with cis-regulatory elements (Goszczynski et al., 2021).

The majority of protein coding genes produce multiple transcript isoforms with different 5'UTRs originating from alternative TSSs (Carninci et al., 2006; Kimura et al., 2006; Reyes & Huber, 2018). The alternative 5'UTRs can affect the expression of downstream coding regions (Smith, 2008) and result in differential transcript expression between tissue and/or cell-types (Resch et al., 2009). Some specific regulatory motifs in the alternative TSSs can regulate gene expression in a tissue-specific manner by interacting with transcription factors (Makhnovskii et al., 2022). These motifs can be highly conserved across species (Siepel et al., 2005). This transcriptional regulation with alternative TSSs serves to widen the gene expression in a tissue-specific and/or cell-specific manner (Barrett et al., 2012). *BDNF* produces 11 transcripts from 8 different TSSs which all encode the identical protein isoforms. However, these transcripts show a cell-specific expression pattern (Cattaneo et al., 2016; Pruunsild et al., 2007). Moreover, the tissue specific usage of alternative TSS was observed from *CACNA1C* (Dai et al., 2002; Pang et al., 2003; N. I. Saada et al., 2005). *CACNA1C*, which encodes Cav1.2, produces two transcripts, which use different promoter regions for their transcription. One transcript has a longer N-terminal region with exon 1a which is expressed only in the human heart, whereas the other transcript has a shorter N-terminal region with exon 1b which is a dominant transcript in the human brain. Thus, the promoter of the transcript with exon 1a upregulates only the expression of the transcript in the heart but does

not affect the expression of the transcript with exon 1b. This means that the channel controls its expression via an alternative TSS of the gene (Pang et al., 2003; N. I. Saada et al., 2005).

Sequences that are highly conserved across species inform functional genomic sequences that exist in the mammalian genome (Siepel et al., 2005). For example, the 5'UTR is known to be highly conserved because it contains regulatory loci such as transcription factor binding sites (Algama et al., 2017; Byeon et al., 2021). However, a gap exists in our understanding of the behaviour of the 5'UTR across gene evolution (Dvorak et al., 2019). Recently, the Zoonomia project identified evolutionarily conserved genomic regions across 240 mammals using comparative genomic data (Arneson & Ernst, 2019). The data used PhyloP scores to detect the conservation status of each base which showed that the regions were not conserved in a cell-type dependent nor a tissue-type dependent manner (Arneson & Ernst, 2019). Therefore, in this chapter, I will also investigate whether the 3' transcripts of LTCC genes also exhibit tissue-specific expression.

### **4.3. Aim of chapter 4**

This chapter describes several analyses undertaken to examine (a) the conservation status of 5'UTR and (b) the expression pattern of the 3' transcripts of LTCC genes. Firstly, this chapter aims to examine to what extent the 5'UTR of the 3' transcripts of LTCC genes is conserved compared with the exonic and intronic regions of the transcripts across 240 mammals. It was hypothesized that if the 5' UTR contains functional regions, it will be more conserved compared with the exonic regions and/or the intronic regions. This will unravel whether the 5'UTR of the 3' transcripts could have functional regions such as the transcription binding site. Secondly, the chapter aims to investigate the exon expression pattern of the 3' transcript of LTCC genes in the human tissue using GTEx data. It was hypothesized that the amount of reads would be greater for

exons downstream of the TSS of the 3' transcript compared to the exons upstream. This will both provide additional evidence for the presence of 3' transcripts, and elucidate whether this occurs in a tissue-specific manner; in particular, whether the profile is similar in brain compared to other tissues. Together, the studies in this chapter will investigate the species and tissue specificity of 3' transcripts of LTCC genes.

#### **4.4. Methods**

Several analytical approaches were applied to explore the conservation status of 5' UTRs and the expression pattern of 3' transcripts of LTCC genes in the human tissues. Conservation level of 5'UTR, exonic regions and intronic regions of 3' transcripts were compared using randomized resampling analysis of PhyloP scores (Zoonomia Consortium, 2020) (see Methods 4.4.3). Then, exon expression level data from human-specific GTEx V8 were explored to examine the evidence of expression from the novel TSS of LTCC 3' transcripts across various human tissues or in a brain tissue-specific manner. All analyses were performed using Python 3.7.6 and Visual Studio Code 1.74.1.

##### **4.4.1. Genomic location data**

The identified TSSs of the 3' transcripts of LTCC genes from 5' RACE experiments in chapter 3 were used to determine the genomic locations of the 5'UTR, exonic regions, and intronic regions of the 3' transcripts of LTCC genes. For the 3' transcript of *CACNAIS* and *CACNAIF*, exonic regions in this chapter contain all the exons identified from the long-read nanopore sequencing, and intronic regions in this chapter contain all introns when there is no exon-skipping. All genomic locations were derived from GRCh38 Genome Reference Consortium Human Reference 38. Moreover, the Gencode V26 human reference genome

([genecodegenes.org/human/release-26.html](http://genecodegenes.org/human/release-26.html)) was used to retrieve nucleotide sequences of the 3' transcript of LTCC genes.

#### **4.4.2. PhyloP score from the Zoonomia project**

The Zoonomia Project shared phylogenetic genome assembly for 240 mammals, providing an insight on the genetic diversity among mammals (Zoonomia Consortium, 2020). The data from the Zoonomia Project showed a PhyloP score for each base, retrievable from the UCSC Genome Browser (<http://genome.ucsc.edu/cgi-bin/hgGateway>). The PhyloP score quantifies an evolutionary conservation for each base, from which a positive score means the site would be conserved; by contrast, a negative score means the site would evolve faster. Most pathogenic variants are placed where is highly conserved (Sun & Yu, 2019). Randomized resampling was exploited to examine whether the 5'UTR is more conserved than that of exonic regions or intronic regions of 3' transcripts of LTCC genes. Python 3.7.6 was used to analyze the data and generate appropriate figures (See methods 4.4.3). More specifically, Panda (McKinney, 2010) was used for data frame generation, Numpy (Harris et al., 2020) for calculations, and Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021) for visualization.

#### **4.4.3. Randomised resampling**

To examine the conservation of 5'UTRs of the 3' transcript of LTCC genes across 240 mammals, random resampling in Python 3.7.6 was exploited. This analysis led to comparing the conservation status of 5'UTR, exonic regions, and intronic regions under the same condition. Therefore, the same number of nucleotides in the 5'UTR of each 3' transcript was randomly resampled from the exonic regions and intronic regions 1,000 times. For example, 5'UTR of the 3' transcript of *CACNA1F* is comprised of 11 Ts, 7 As, 12 Cs, and 6 Gs (Figure 4.1). Therefore, a

set of 11 Ts, 7As, 12Cs and 3Gs were randomly resampled from each exonic regions and intronic regions, which was repeated 1000 times. The same process was applied to other 3' transcripts of LTCC genes. 5'UTR of the 3' transcript of *CACNA1S* has 4 Ts, 5 Cs, 6 Gs, and 4 As; the 5'UTR of the 3' transcript of *CACNA1D* has 10Ts, 12 As, 8 Cs, and 14 Gs; and the 5'UTR of the 3' transcript of *CACNA1C* has 11 Ts, 33 As, 25 Cs, and 32 Gs.

During randomized resampling, the estimated median PhyloP score from exonic regions and intronic regions were calculated with a 95% confidence interval. If the median PhyloP score of the 5' UTR of each 3' transcript of LTCC genes falls within the 95% confidential interval of the estimated PhyloP score from the exonic regions and intronic regions, then the 5'UTR is not significantly more (or less) conserved than exonic and/or intronic regions. However, if the median PhyloP score of 5' UTR falls outside of the 95% confidential interval, it will be either significantly less or more conserved. For the specific code, see the code in the Appendix 11.

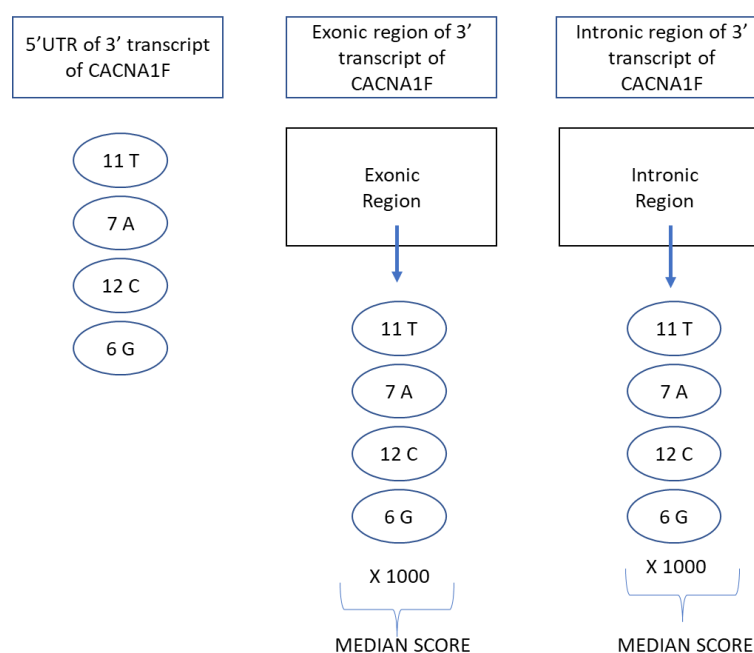


Figure 4.1. This schematic illustrates the process of the randomized resampling.

#### 4.4.4. Expression level of exons in LTCC genes from GTEx data

Data from the GTEx project allow for the study of mRNA expression in human tissue, and GTEx Analysis Release V8 data were explored to investigate whether the expression of LTCC 3' transcripts are from the novel TSS and tissue specificity of the expression. From the GTEx data, exon expression level data (read counts of each exon from the short-length RNA-sequencing) was used. The tissues (Table 4.1) from GTEx V8 data were categorized into two: brain (including all brain regions) and non-brain (all other tissues except brain tissues; <https://gtexportal.org/home/samplingSitePage>).

The statistically significance in changes of exon expression levels between exons was investigated using Mann-Whitney U test with Bonferroni correction for multiple testing. For analysis of exon expression level data, Excel (2018) and Python 3.7.6 were used. The exon expression level (Log10) of the exon containing the TSS of each 3' transcript was measured, along with that of the adjacent upstream and downstream exons.

A fold change calculation was used to examine to what extent the exon expression level changes as transcription of the 3' transcript of LTCC genes starts. With B as the changed expression level and A as the original expression level, the fold change calculation is  $(B-A)/A$ . The fold change was calculated using Excel (2018). Analyses were done by using Excel (2018), Panda (McKinney, 2010) for the data frame generation, Numpy (Harris et al., 2020) for calculations, and Matplotlib (Hunter, 2007) and Seaborn (Waskom, 2021) for visualisation and statistical analysis in Python 3.7.6. via Visual Studio Code 1.74.1.

Table 4.21 Tissues in GTEX data

Tissue name	Tissue name
Adipose - Subcutaneous	Esophagus - Gastroesophageal Junction
Adipose - Visceral (Omentum)	Esophagus - Mucosa
Adrenal Gland	Esophagus - Muscularis
Artery - Aorta	Fallopian Tube
Artery - Coronary	Heart - Atrial Appendage
Artery - Tibial	Heart - Left Ventricle
Bladder	Kidney - Cortex
Brain - Amygdala	Kidney - Medulla
Brain - Anterior cingulate cortex (BA24)	Liver
Brain - Caudate (basal ganglia)	Lung
Brain - Cerebellar Hemisphere	Minor Salivary Gland
Brain - Cerebellum	Muscle - Skeletal
Brain - Cortex	Nerve - Tibial
Brain - Frontal Cortex (BA9)	Ovary
Brain - Hippocampus	Pancreas
Brain - Hypothalamus	Pituitary
Brain - Nucleus accumbens (basal ganglia)	Prostate
Brain - Putamen (basal ganglia)	Skin - Not Sun Exposed (Suprapubic)
Brain - Spinal cord (cervical c-1)	Skin - Sun Exposed (Lower leg)
Brain - Substantia nigra	Small Intestine - Terminal Ileum
Breast - Mammary Tissue	Spleen
Cells - Cultured fibroblasts	Stomach
Cells - EBV-transformed lymphocytes	Testis
Cervix - Ectocervix	Thyroid
Cervix - Endocervix	Uterus
Colon - Sigmoid	Vagina
Colon - Transverse	Whole Blood



## 4.5. Results

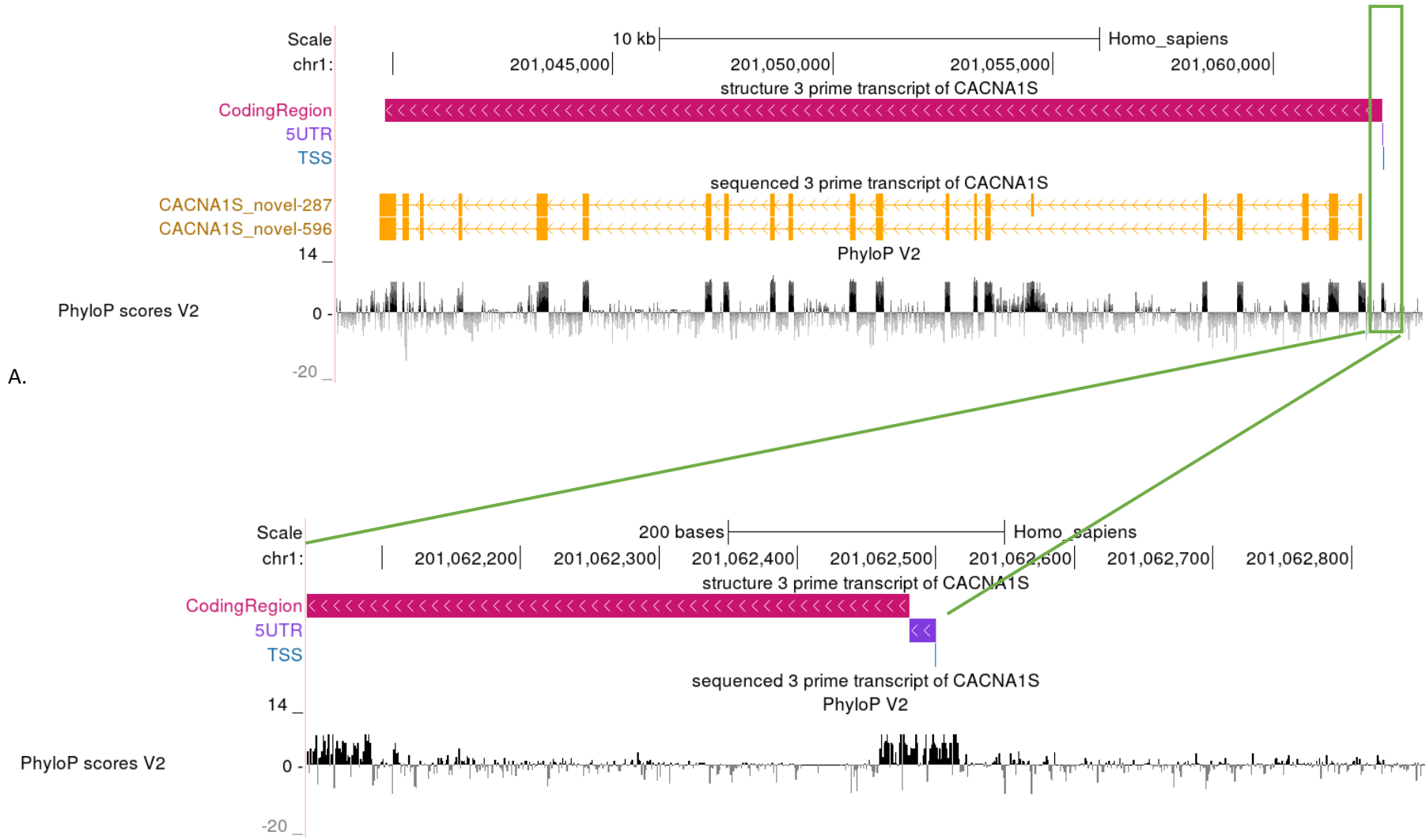
### 4.5.1. Conservation status of the 3' transcript of LTCC genes using PhyloP score

Based on the TSS of the 3' transcript of LTCC genes identified from chapter 3, the conservation status of the 3' transcript of LTCC genes were examined. For the 3' transcripts of *CACNAIS* and *CACNAIF*, not only the structures, but also sequenced in-frame coding transcripts, which were identified as highly expressed in the human brain in chapter 2, were mapped to the PhyloP scores from the Zoonomia Project via the UCSC Genome Browser (<http://genome.ucsc.edu>). The mapping provided insight into the evolutionary conservation pattern of 3' transcripts of LTCC genes (Figure 4.2).

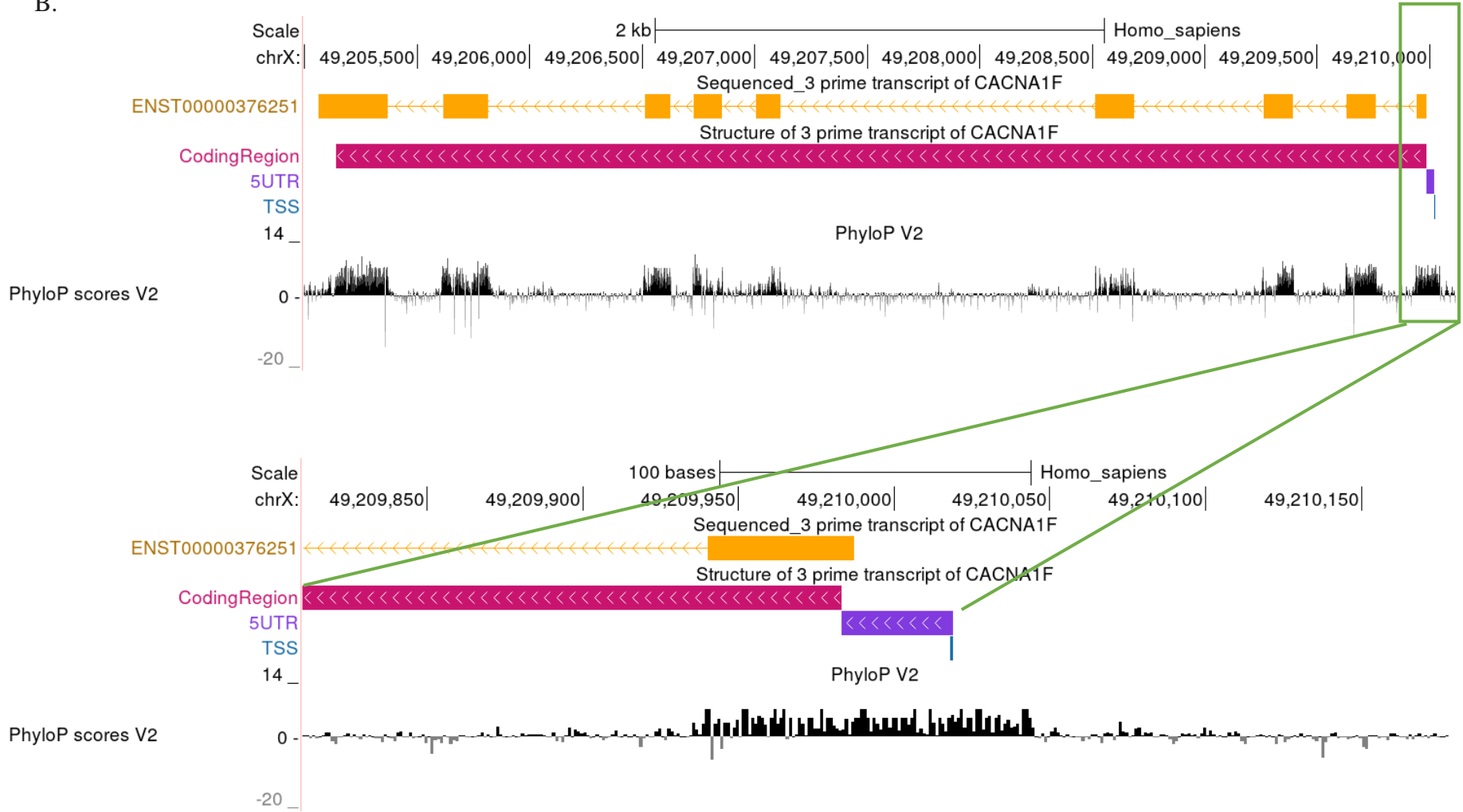
Moreover, PhyloP scores around the TSS were examined as shown in Figure 4.2. This shows that nucleotides upstream from the TSS of the 3' transcript of LTCC genes also have a high conservation rate, which might be because the TSS resides within the exon of the full-length LTCC genes. As it can imply that the 5'UTR of 3' transcripts may show a high conservation level due to being located on the existing exon (Table 4.2), further analysis was conducted to examine whether randomly selected intronic region sequences or exonic region sequences are also less conserved compared with the PhyloP score of the 5'UTR (See 4.6.3. for method).

*Table 4.22 Exon number where TSS is located*

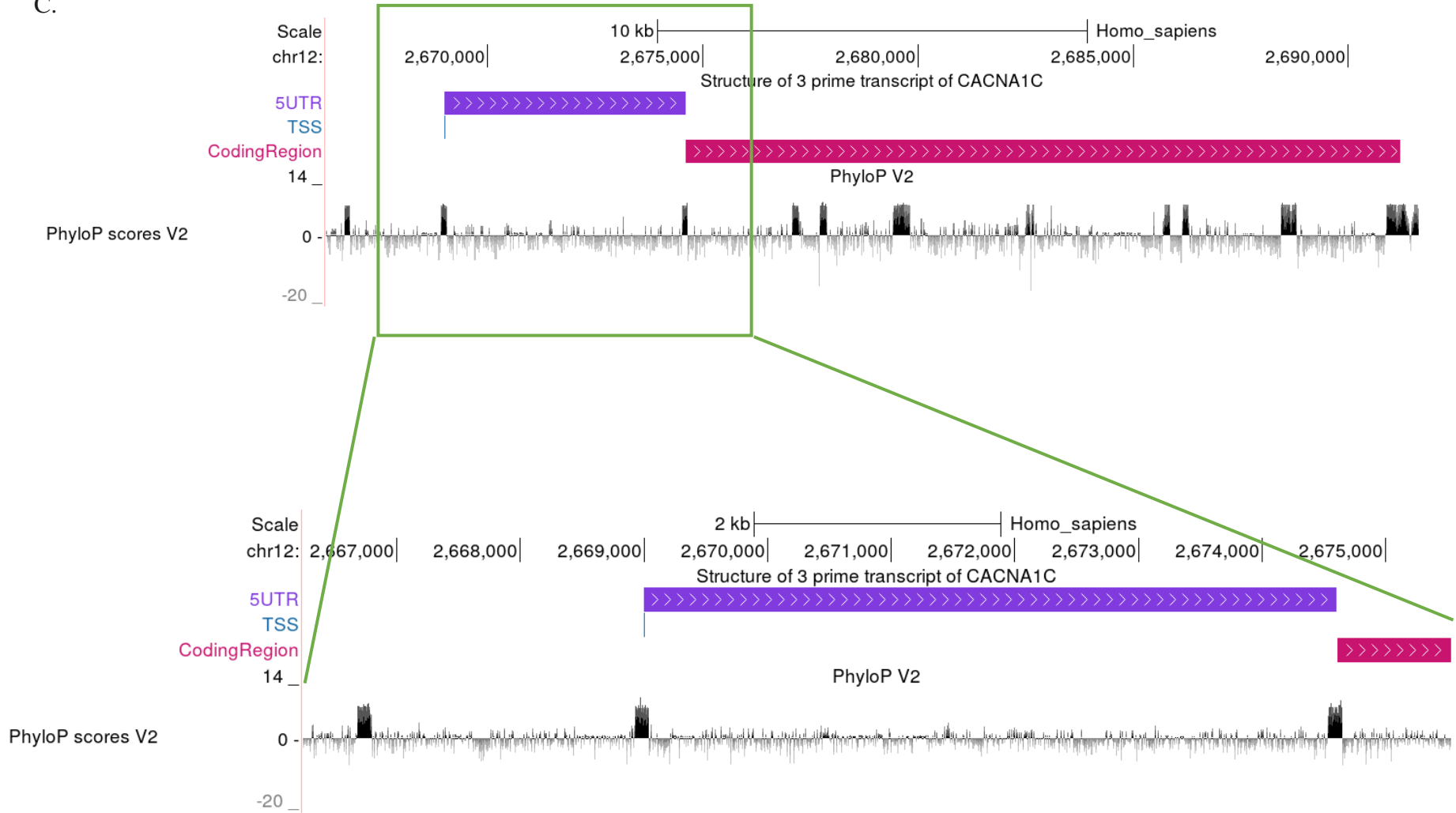
Name of the transcript	Where the TSS is located on the full-length gene (Exon number)
3' transcript of <i>CACNAIS</i>	Exon 24
3' transcript of <i>CACNAIF</i>	Exon 41
3' transcript of <i>CACNAIC</i>	Exon 47
3' transcript of <i>CACNAID</i>	Exon 47



B.



C.



D.

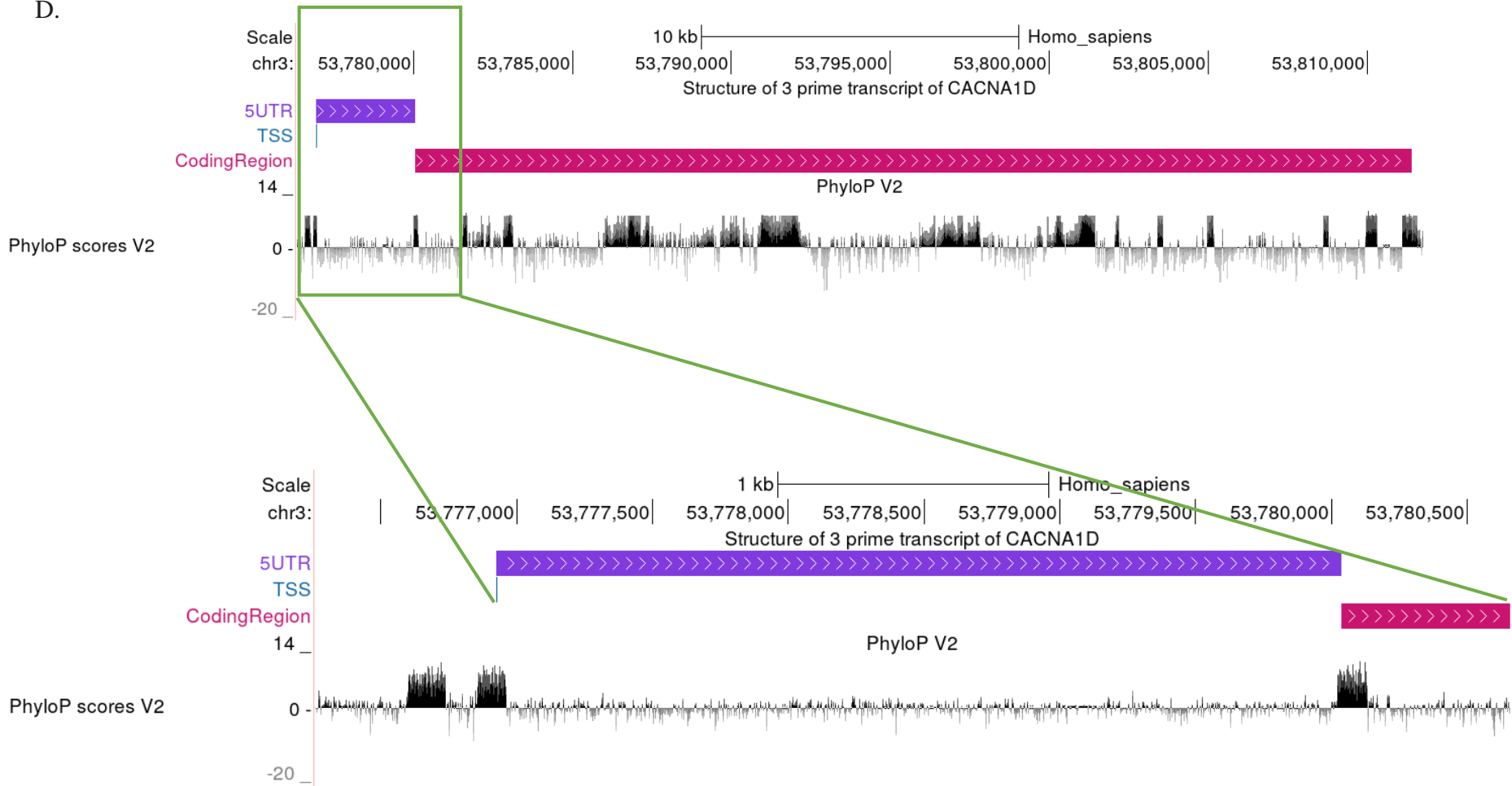


Figure 4.2. The PhyloP score around the TSS of 3' transcripts of LTCC genes. The UCSC Genome Browser track shows a PhyloP score for the 3' transcript of LTCC genes. A track named as Structure of 3 prime transcript of a gene is showing the structure of each 3' transcript (TSS, 5'UTR and coding region) identified from chapter 3. Black bars appearing as PhyloP scores V2 represent the PhyloP score of each nucleotide at the location. If the bar is positive, the nucleotide should be conserved, whereas if the bar is negative, the nucleotide may not be conserved. For **A** and **B**, sequenced coding transcripts of the 3' transcript of *CACNAIS* and *CACNAIF* from chapter 2 using long-read nanopore sequencing were also mapped to the PhyloP score (yellow track). **A**. Both the structure of the 3' transcript of *CACNAIS* and the sequenced coding 3' transcript of *CACNAIS* were mapped to a PhyloP score (yellow track). The below panel shows a magnified view to show the conservation status around the TSS of the 3' transcript of *CACNAIS*. **B**. Both the structure of the 3' transcript of *CACNAIF* and the sequenced coding 3' transcript of *CACNAIF* were mapped to a PhyloP score from the Zoonomia Project (yellow track). The below panel shows a magnified view to examine conservation status around the TSS of the 3' transcript of *CACNAIF*. **C**. Structure of the 3' transcript of *CACNAIC* was mapped to the PhyloP score. The below panel is a magnified view to examine conservation status around the TSS of the 3' transcript of *CACNAIC*. **D**. Structure of the 3' transcript of *CACNAID* was mapped to the PhyloP score. Below panel is a magnified view to examine conservation status around the TSS of the 3' transcript of *CACNAID*.

#### 4.5.2. Comparing the conservation status of the 5'UTRs with the exonic and intronic regions of the 3' transcripts

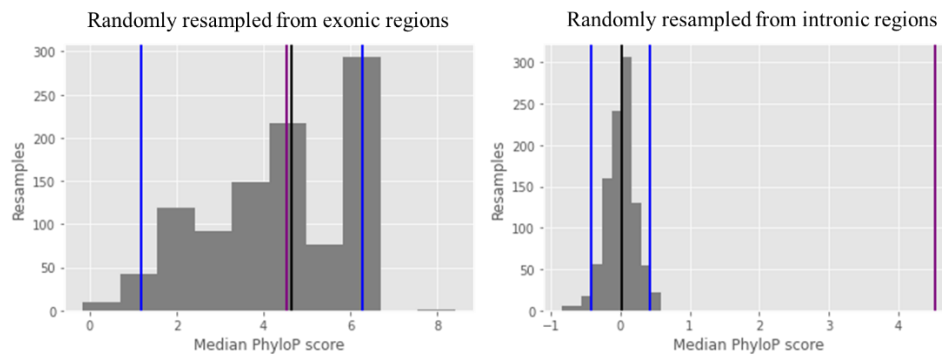
The possibility that whether the 5'UTR is more conserved compared to other regions in the 3' transcript of LTCC genes was investigated (Figure 4.3; see 4.4.3. for methods). The Median PhyloP score of the 5'UTR of each 3' transcript of the LTCC genes (purple line in Figure 4.3) was compared with the estimated median PhyloP score calculated by randomized resampling from either exonic regions or intronic regions of the 3' transcripts (estimated median PhyloP score is the black line and 95% confidence interval is represented by blue lines in Figure 4.3). If the median PhyloP score of the 5'UTR (purple line in Figure 4.3) lies above the confidence interval (blue lines in Figure 4.3), the 5'UTR is significantly more conserved than the exonic region or the intronic region. If the median PhyloP score of 5'UTR (purple line in Figure 4.3) lies between the confidence interval (blue lines in Figure 4.3), the conservation status of 5'UTR is not significantly different from that of other regions.

Figure 4.3A compares the PhyloP score of the 5'UTR of the *CACNAIS* 3' transcript with the distribution of a randomly resampled PhyloP score from the exonic and intronic regions of the transcript. The 5'UTR is seen to be significantly more conserved than the intronic, but not exonic, regions. In the case of the other LTCC genes examined, the 5'UTR was significantly more conserved than both the intronic and exonic regions (*CACNAIF* [Figure 4.3B]; *CACNAIC* [Figure 4.3C]; *CACNAID* [Figure 4.3D]).

Overall, the 5'UTR of each LTCC 3' transcript is significantly more conserved than the intronic regions across 240 mammals; the 5'UTR is also more conserved than the exonic regions for each of the 3' transcripts except for the 3' transcript of *CACNAIS*.

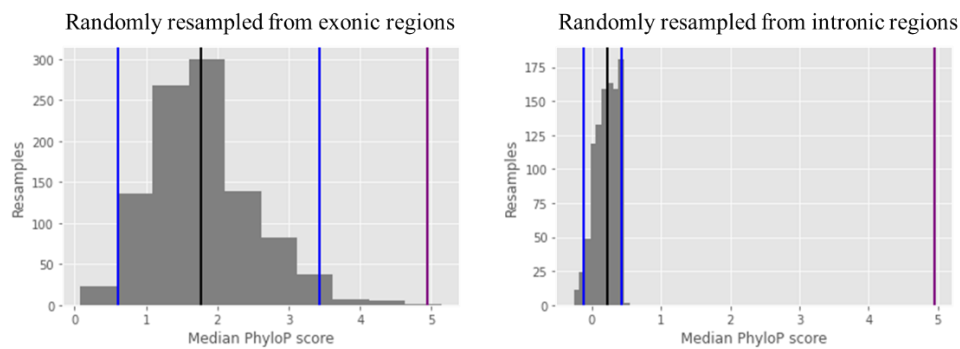
A.

## 3' Transcript of CACNA1S



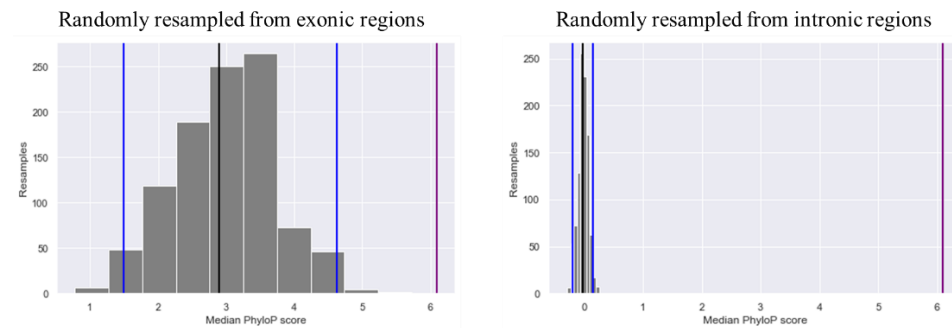
B.

## 3' Transcript of CACNA1F



C.

## 3' Transcript of CACNA1C



D.

## 3' Transcript of CACNA1D

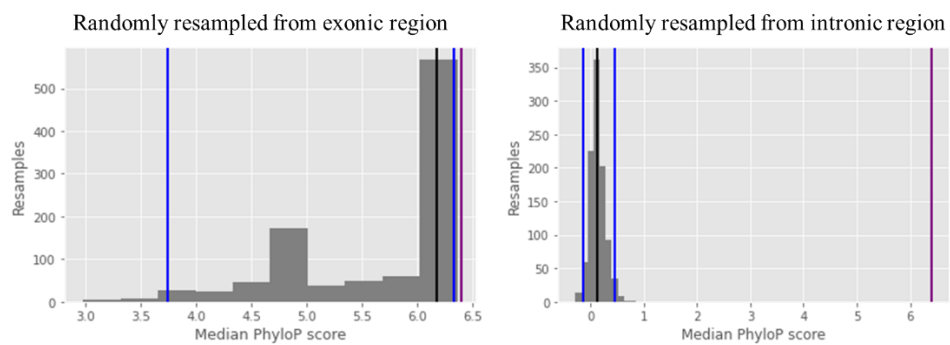




Figure 4.3 . Comparison of median PhyloP scores among the 5'UTR, exonic regions and intronic regions of the 3' transcripts of LTCC genes. The graphs named as 'Randomly resampled from exonic regions' shows the distribution of estimated median PhyloP scores which were randomly resampled from the exonic regions. The graphs named as 'Randomly resampled from intronic regions' shows the distribution of estimated median PhyloP scores which are randomly resampled from the intronic regions. The **Black line** represents an estimated median PhyloP score of a randomly resampled PhyloP score. **Blue lines** represent a 95% confidence interval, and the **purple line** represents the median PhyloP score of the 5'UTR of each 3' transcript of LTCC genes. **A.** The median PhyloP score comparison among 5'UTR and that of randomly resampled exonic regions and intronic regions of the CACNAIS 3' transcript. The median PhyloP score of 5'UTR is 4.53. Intronic regions have an estimated median PhyloP score of 0.023 with a 95% confidence interval [-0.43 0.43], whereas exonic regions have an estimated median PhyloP score of 4.63 with a 95% confidence interval [1.19 6.28]. **B.** The median PhyloP score comparison among 5'UTR and that of randomly resampled exonic regions and intronic regions of the CACNAIF 3' transcript. The median PhyloP score of the 5'UTR is 4.95. The intronic regions have an estimated median PhyloP score of 0.22 with a 95% confidence interval [-0.12 0.43], whereas the exonic regions have an estimated median PhyloP score of 1.77 with a 95% confidence interval [0.61 3.44]. **C.** The median PhyloP score comparison among the 5'UTR and that of randomly resampled exonic regions and intronic regions of the CACNAIC 3' transcript. The median PhyloP score of 5'UTR in the 3' transcript of CACNAIC is 6.14. The intronic regions have an estimated median PhyloP score of -0.02 with a 95% confidence interval [-0.2 0.14], whereas the exonic regions have an estimated median PhyloP score of 2.91 with a 95% confidence interval [1.49 4.63]. **D.** The median PhyloP score comparison among 5'UTR and that of randomly resampled exonic regions and intronic regions of the CACNAID 3' transcript. The median PhyloP score of the 5'UTR is 6.4. The intronic regions have an estimated median PhyloP score of 0.13 with a 95% confidence interval [-0.13 0.45], whereas the exonic regions have an estimated median PhyloP score of 6.17 with a 95% confidence interval [3.74 6.34].

### 4.5.3. Expression levels of exons before and after initiation of LTCC 3' transcript

This analysis is to assess the evidence of expression from the novel TSS of LTCC 3' transcripts using the exon expression level (read counts per exon) from the GTEx. It was hypothesized that there will be greater read counts from the novel TSS onwards compared to the exons upstream of the novel TSS. The transcription of the 3' transcripts of LTCC genes starts at a certain exon of the full-length LTCC genes. Therefore, if the 3' transcripts do exist in the human tissues, the downstream (-1) exon will show higher read counts compared to the upstream (+1) exon. The changes in the exon expression levels are identified using two methods: (a) comparing expression levels among exons and (b) comparing to what extent the expression level changes by calculating fold changes (see 4.4.4 for method). Changes in expression levels were compared among exons either side of where the TSS of 3' transcripts are located, *i.e.*, one exon upstream and one exon downstream. This allowed examination of the extent of expression changes as transcription initiates.

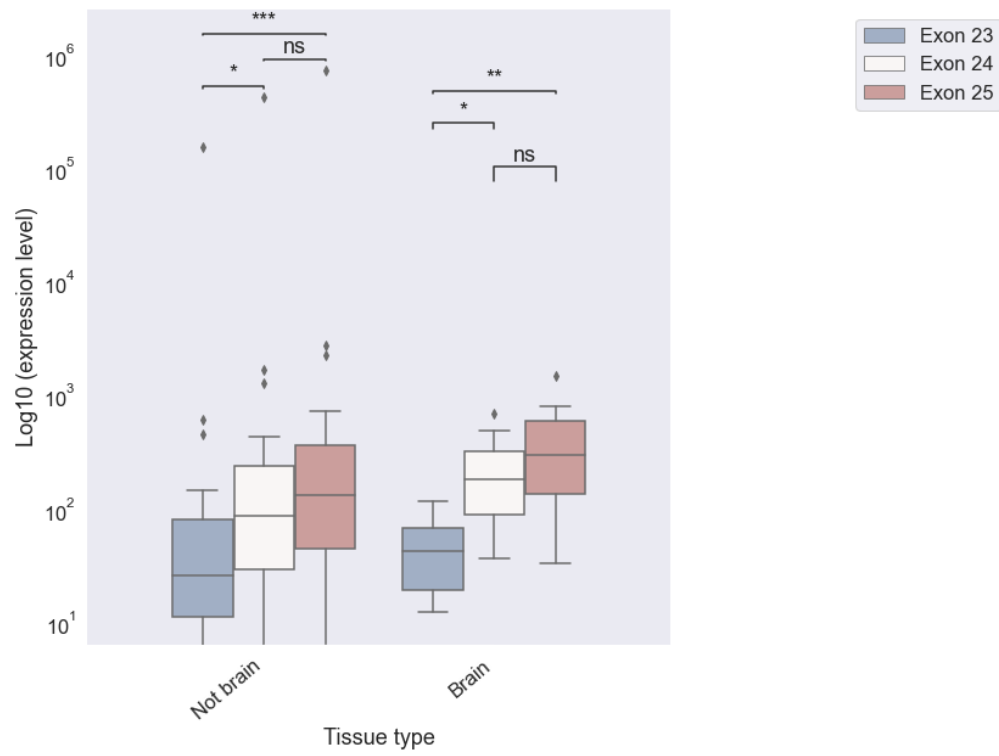
#### 4.5.3.1. Expression level analyses after the transcription of 3' transcript of *CACNA1S* is initiated

Transcription of the 3' transcript of *CACNA1S* initiates at Chr1: 201062500 (hg38), identified as exon 24 in the data from GTEx V8. The analysis presented here examines whether expression levels increase after the transcription of the 3' transcript of *CACNA1S* is initiated. Expression levels of exons were examined based on tissue types: whether the tissue is a brain tissue or a non-brain tissue. The analysis shows that regardless of the tissue origin, exon expression increases from exon 23 to exon 25 (Figure 4.4A). From exon 23 to 24, the exon expression level increases significantly in both non-brain tissues ( $p = 3.15e-02$ ; Mann-Whitney U test) and in brain tissues ( $p = 1.49e-02$ ; Mann-Whitney U test). Moreover, exon 25 has significantly increased

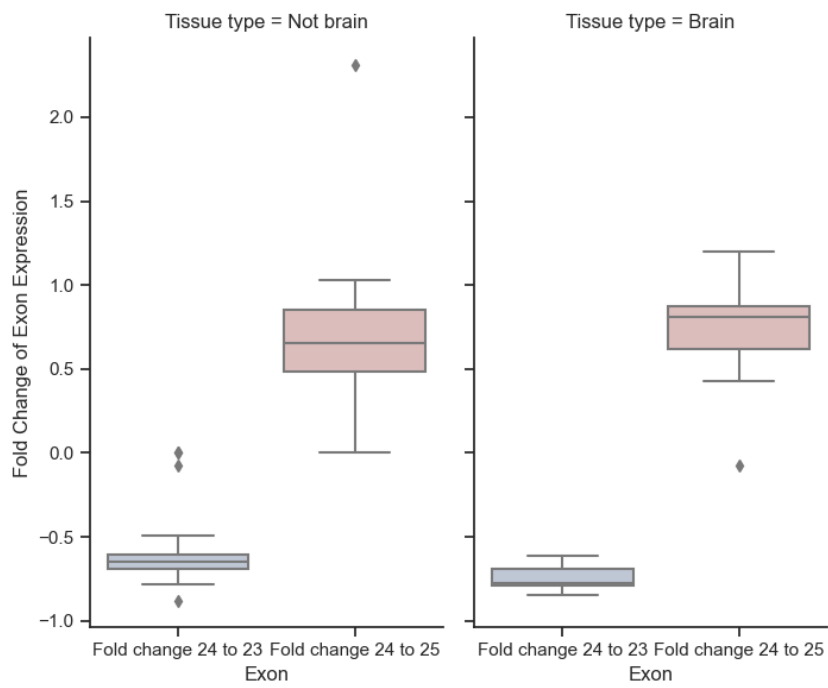
expression level compared to the exon 23 in both non brain tissues ( $p = 5.67e-04$ ; Mann-Whitney U test) and brain tissues ( $p = 1.63e-03$ ; Mann-Whitney U test). However, from exon 24 to exon 25, the exon expression level is not significantly increased in both non-brain tissues ( $p = 7.02e-01$ ; Mann-Whitney U test) and brain tissues ( $p = 1.00e+00$ ; Mann-Whitney U test).

The fold change analysis confirms that exon expression is higher after transcription of the 3' transcript of *CACNAIS* is initiated (Figure 4.4B). In non-brain tissues, exon 23 is 65% less expressed than exon 24 while exon 25 is 65% more expressed than exon 24. On the other hand, in brain tissues, exon 23 is 77% less expressed compared with the exon 24 whereas exon 25 is 81% more expressed than exon 24. These fold changes show that the exon expression level increases once the transcription of the 3' transcript of *CACNAIS* is initiated.

A.



B.



*Figure 4.4. Changes in expression level among exon 23, exon 24 and exon 25 of CACNA1S. A.*

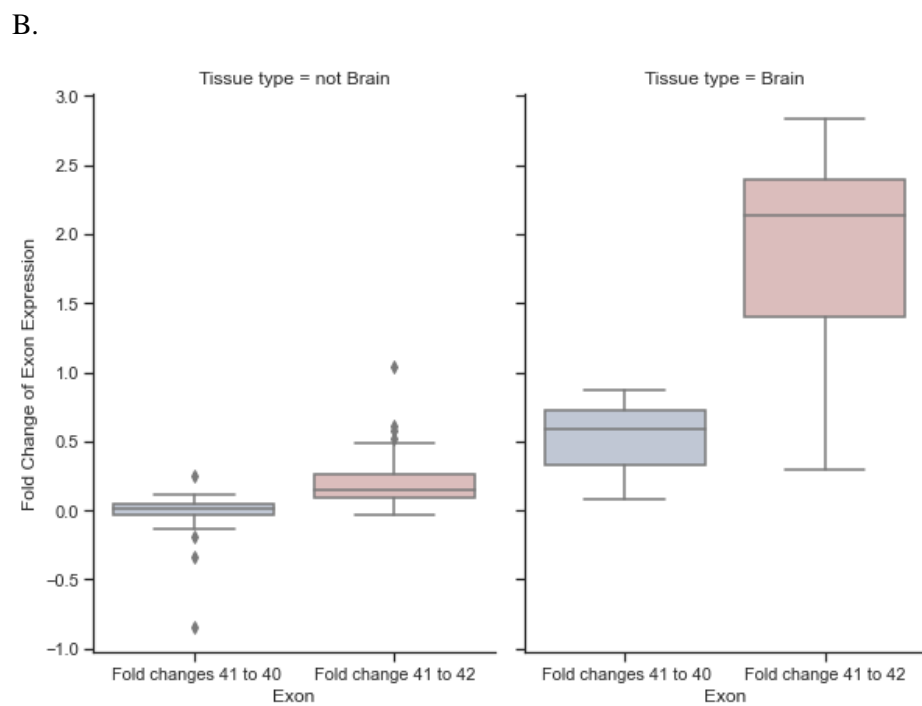
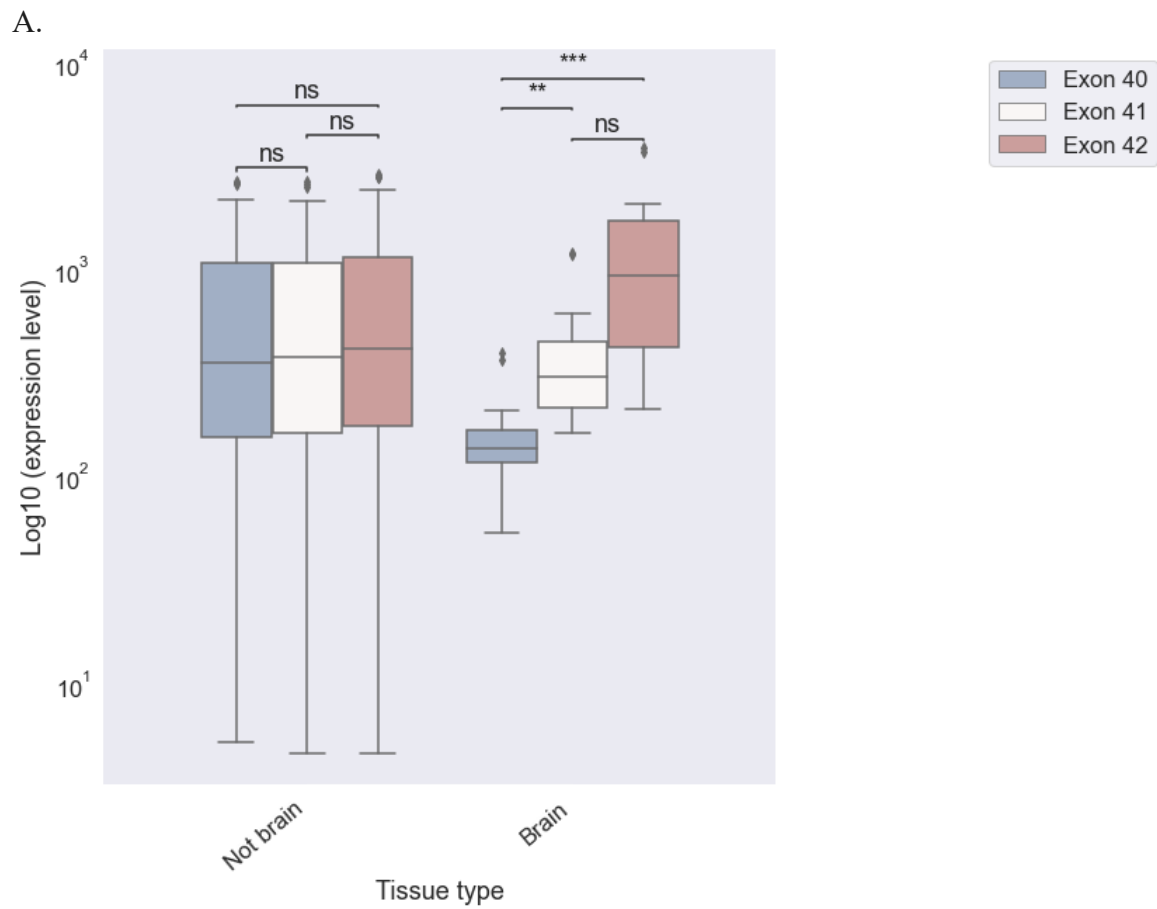
*Distribution of expression levels of exons 23, 24, and 25 based on tissue types. These levels reveal that the expression of exons increases as the transcription of the 3' transcript of CACNA1S initiates (exon 24). p-value annotation legend: ns:  $5.00e-02 < p \leq 1.00e+00$ , \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$ . In non-brain tissues, the median expression level of exon 23 is 1.40, of exon 24 is 1.93, and of exon 25 is 2.15. In brain tissues, the median expression level of exon 23 is 1.62, of exon 24 is 2.25, and of exon 25 is 2.47. B. Changes in the expression level of exons based on tissue types. In the not-brain tissues, the median fold change of exon expression level from exon 24 to exon 23 is -0.65, and from exon 24 to exon 25 is 0.65. In the brain tissues, the median fold change of exon expression level from exon 24 to exon 23 is -0.77, and from exon 24 to exon 25 is 0.81.*

#### **4.5.3.2. Expression level analyses after transcription of CACNA1F 3' transcript is initiated**

Transcription of the 3' transcript of *CACNA1F* initiates at ChrX: 49210019 (hg38), identified as exon 41 in GTEx V8 exon expression levels. The analysis examined whether expression levels increase after initiating transcription of the 3' transcript of *CACNA1F*. Expression levels were examined based on tissue type: brain tissues and non-brain tissues (Figure 4.5A). The expression level does not significantly change after the transcription of the transcript is initiated in non-brain tissues. The exon expression level changes from exon 40 to exon 41 ( $p = 1.00e+00$ ; Mann-Whitney U test), exon 41 to exon 42 ( $p = 1.00e+00$ ; Mann-Whitney U test), and exon 40 to exon 42 ( $p = 1.00e+00$ ; Mann-Whitney U test) are not significant in non-brain. On the contrary, in the brain tissues, exon expression level is noticeably increased as the transcription of *CACNA1F* 3' transcript was initiated. Expression level of exon 41 is significantly increased compared with that of exon 40 ( $p = 8.85e-03$ ; Mann-Whitney U test). Exon 42 shows significantly

increased expression level compared to exon 40 ( $p = 3.80e-04$ ; Mann-Whitney U test). However, exon 42 does not show significant increase in exon expression level compared to exon 41 ( $p = 5.35e-02$ ; Mann-Whitney U test).

In addition, fold changes between expression levels of exon 41 and exon 40, and between expression levels of exon 41 and exon 42 were examined to understand to what extent the expression level would change once the transcription is initiated (see Figure 4.5B). In non-brain tissues, around 1% of expression level changed between exon 41 and exon 40, and between exon 41 and exon 42. However, in brain tissues, exon 40 showed a 59% higher expression level than exon 41, whereas exon 42 showed an expression level change of more than 200% compared to exon 41.



*Figure 4.5. Changes in expression level among exon 40, 41 and 42 of the 3' transcript of CACNA1F. A. The distribution of expression levels of exons 40, 41, and 42 in different tissues (brain tissues and non-brain tissues). p-value annotation legend: ns:  $5.00e-02 < p \leq 1.00e+00$ , \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$ . In the non-brain tissues, the median expression level of exon 40 is 2.55, of exon 41 is 2.58, and of exon 42 is 2.61. In the brain tissues, the median expression level of exon 40 is 2.13, of exon 41 is 2.48, and of exon 42 is 2.97. B. Fold change of exon expression levels from exon 41 to exon 40 and exon 41 to exon 42 in different tissues (brain tissues or non-brain tissues). In the non-brain tissues, the median fold change of exon expression level from exon 41 to exon 40 is 0.01, and from exon 41 to exon 42 is 0.16. In the brain tissues, the median fold change of exon expression level from exon 41 to exon 40 is 0.59, and from exon 41 to exon 42 is 2.13.*

#### **4.5.3.3. Expression level analyses after transcription of CACNA1C 3' transcript is initiated**

Transcription of the 3' transcript of *CACNA1C* initiates at Chr12: 2669000 (hg38), identified as exon 47 in exon expression level data from GTEx V8. The analysis examined whether expression levels increase after the transcript initiation of the 3' transcript of *CACNA1C*. To examine whether expression level increases after transcription of the 3' transcript of *CACNA1C* is initiated, the expression levels were compared based on tissue type: brain tissues and non-brain tissues (Figure 4.6).

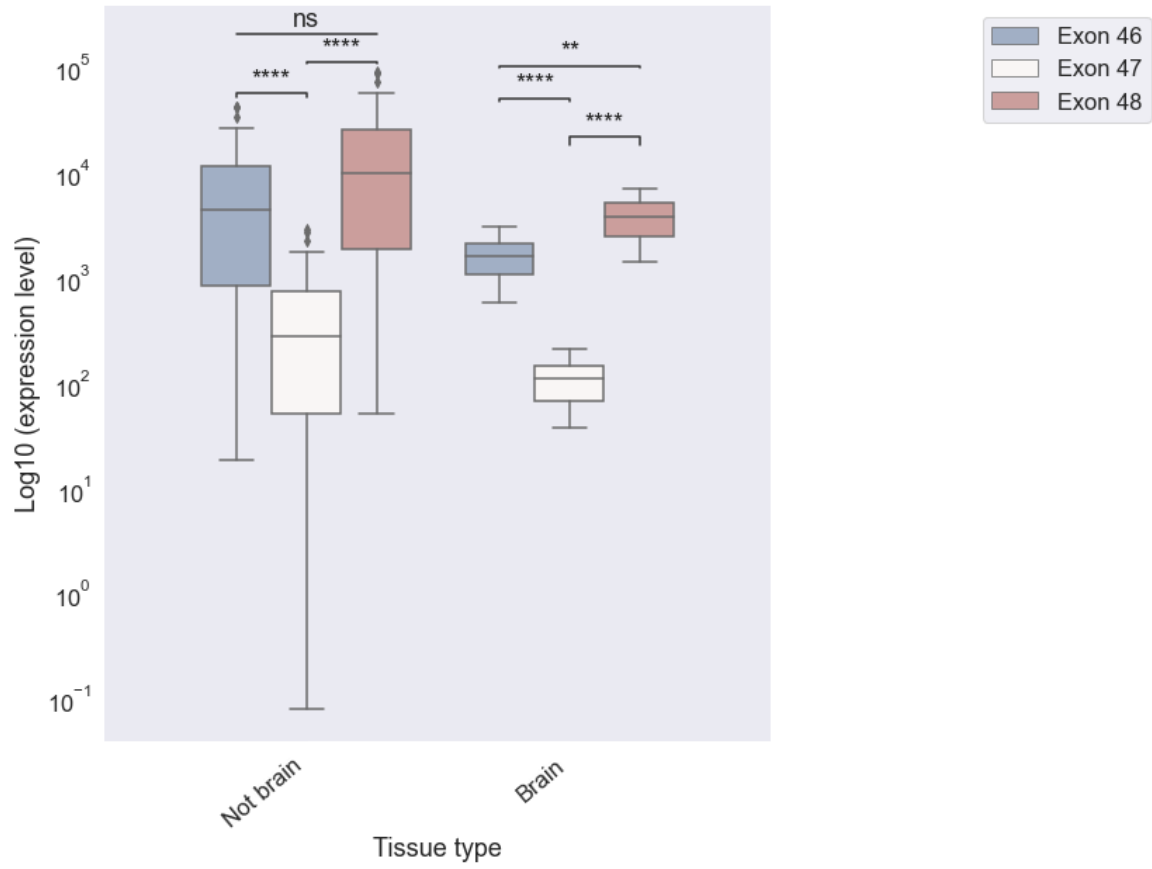
In both tissue types, the lowest expression level occurred in exon 47, where the TSS of *CACNA1C* 3' transcript is located compared to the adjacent exons, exons 46 and 48 (Figure 4.6A). In non-brain tissues, the exon expression level changes from exon 46 to exon 47 is significant ( $p = 2.82e-07$ ; Mann-Whitney U test), and the expression level changes from exon 47 to exon 48 is also significant ( $p = 5.09e-10$ ; Mann-Whitney U test). However, exon level changes from exon 46 to exon 48 are not significant ( $p = 2.17e-01$ ; Mann-Whitney U test). In the brain tissues, the exon



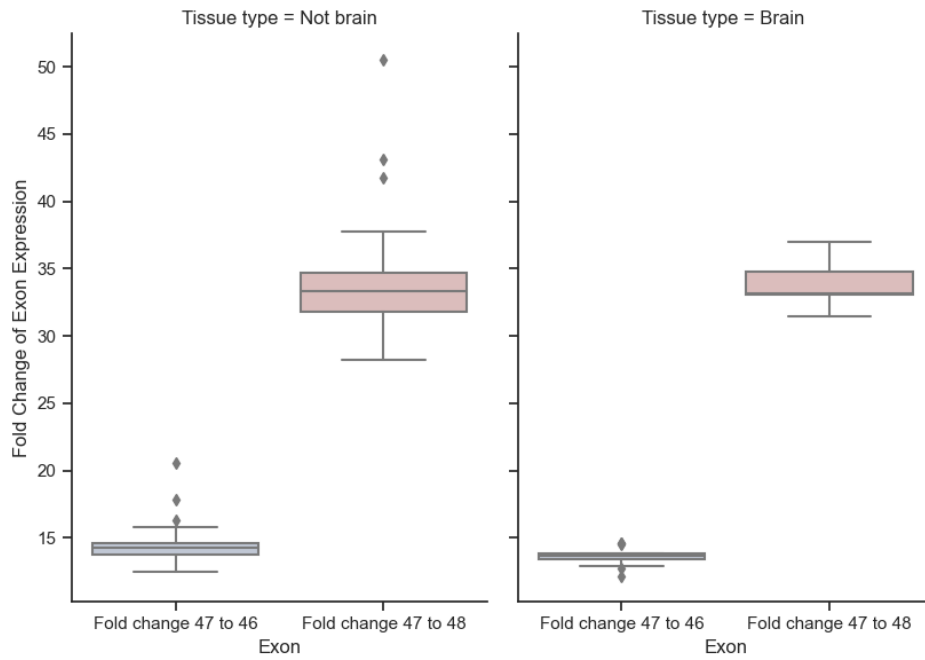
47 shows significantly lower expression level compared with the exon 46 ( $p = 9.90e-05$ ; Mann-Whitney U test), and exon 48 shows significantly higher expression level compared with the exon 47 ( $p = 9.90e-05$ ; Mann-Whitney U test). Also, exon 48 had significantly higher expression level compared with the exon 46 in the human brain tissues ( $p = 8.85e-03$ ; Mann-Whitney U test).

Additionally, fold change was examined to understand to what extent the expression level changes after the transcription of the 3' transcript of *CACNA1C* is initiated (Figure 4.6B). The analysis showed that expression level changed from exon 47 to exon 48 at a higher level than from exon 47 to exon 46 in both tissue types. The expression level increased from exon 47 to exon 46 by 14-fold whereas that increase from exon 47 to exon 48 by 33-fold. As the expression level increased once the transcription of *CACNA1C* 3' transcript is initiated, it shows that the 3' transcript of *CACNA1C* was initiated at exon 47 of the full-length *CACNA1C*.

A.



B.



*Figure 4.6. Changes in expression level among exon 46, 47 and 48 of the 3' transcript of CACNA1C. A. The distribution of expression levels of exons 46, 47, and 48 based on tissue types. p-value annotation legend: ns:  $5.00e-02 < p \leq 1.00e+00$ , \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$ . In the not-brain tissues, the median expression level of exon 46 is 3.64, of exon 47 is 2.44, and of exon 48 is 3.99. In the brain tissues, the median expression level of exon 46 is 3.19, of exon 47 is 2.03, and of exon 48 is 3.57. B. Fold change of expression levels from exon 47 to exon 46 and exon 47 to exon 48 based on tissue types. In the not brain tissues, the median fold change from exon 47 to exon 46 is 14.28, and from the exon 47 to exon 48 is 33.31. In the brain tissues, the median fold change from exon 47 to exon 46 is 13.67, and from exon 47 to exon 48 is 33.16.*

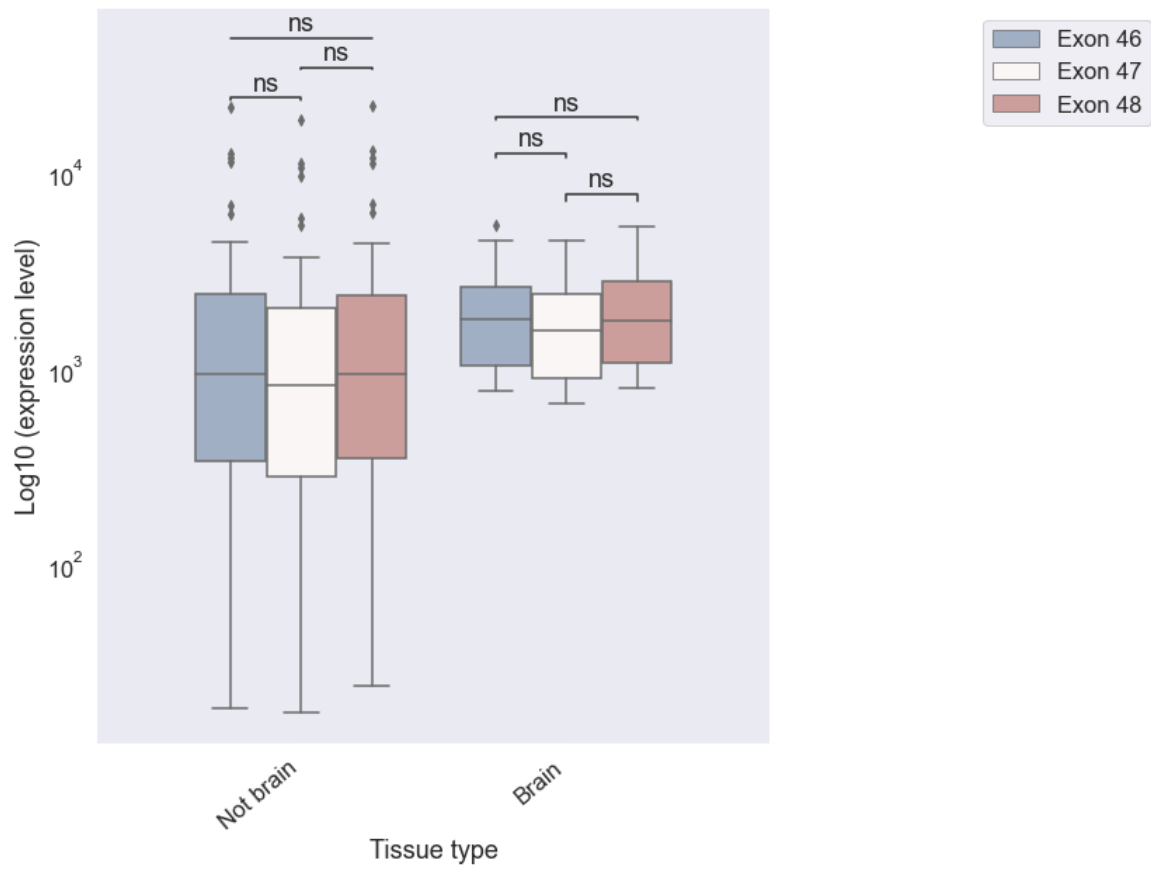
#### **4.5.3.4. Expression level analyses after transcription of CACNA1D 3' transcript is initiated**

Transcription of the 3' transcript of *CACNA1D* initiates at Chr3: 53776925 (hg38), identified as exon 47 in exon expression level data from GTEx V8. The expression levels of exons 46, 47, and 48 were compared to determine whether expression level increased after the initiation of the 3' transcript of *CACNA1D* (Figure 4.7A). It was identified that the exon expression level is slightly increased as the transcription is initiated. The changes in expression levels after the transcription of *CACNA1D* 3' transcript is initiated was examined via fold change calculation (Figure 4.7B).

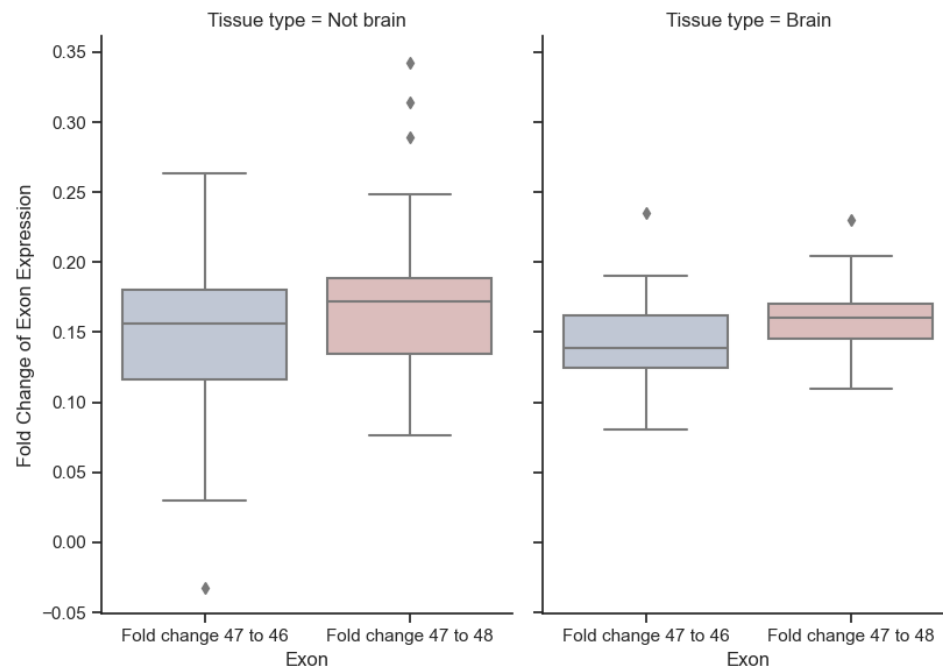
The expression level does not change significantly after the transcription is initiated. In non-brain tissues, the exon expression level changes from exon 46 to exon 47 ( $p = 1.00e+00$ ; Mann-Whitney U test), exon 47 to exon 48 ( $p = 1.00e+00$ ; Mann-Whitney U test) and exon 46 to exon 48 ( $p = 1.00e+00$ ; Mann-Whitney U test) were not significant. Similarly, in the brain tissues, exon 47 did not show significantly lower expression level than exon 46 ( $p = 1.00e+00$ ; Mann-Whitney U test), exon 48 also was not significantly more expressed than exon 46 ( $p = 1.00e+00$ ; Mann-Whitney U test) and exon 47 ( $p = 1.00e+00$ ; Mann-Whitney U test).

Fold change calculation revealed that in non-brain tissues, exon 46 is 16% higher expressed than exon 47 whereas exon 48 has a 17% higher expression level than exon 47. In the brain tissue, the expression level of exon 46 was 14% higher than exon 47, and exon 48 had a 16% higher expression level than exon 47. Therefore, the expression level change from exon 47 to exon 48 is slightly higher than that from exon 47 to exon 46 in both tissue types. These supports the existence of the 3' transcript of *CACNAID*, which is initiated at exon 47 of the full-length *CACNAID*.

A.



B.



*Figure 4.7. Expression levels among exon 46, 47 and 48 of the 3' transcript of CACNA1D. A. The distribution of expression levels of exons 46, 47, and 48. There is a slight change in expression levels as transcription starts (exon 47). p-value annotation legend: ns:  $5.00e-02 < p \leq 1.00e+00$ , \*:  $1.00e-02 < p \leq 5.00e-02$ , \*\*:  $1.00e-03 < p \leq 1.00e-02$ , \*\*\*:  $1.00e-04 < p \leq 1.00e-03$ , \*\*\*\*:  $p \leq 1.00e-04$ . In the non-brain tissues, the median expression level of exon 46 is 2.97, of exon 47 is 2.91, and of exon 48 is 2.97. In the brain tissues, the median expression level of exon 46 is 3.25, of exon 47 is 3.19, and of exon 48 is 3.24. B. Fold changes of expression level from exon 47 to exon 46 and from exon 47 to exon 48. In the not brain tissues, the median fold change from exon 47 to exon 46 is 0.16, and from exon 47 to exon 48 is 0.17. In the brain tissues, the median fold change from exon 47 to exon 46 is 0.14, and from exon 47 to exon 48 is 0.16.*

## **4.6. Discussion**

This chapter had two aims. Firstly, to examine possible functionality of the 5'UTR in the 3' transcript of LTCC genes across 240 mammals. Secondly, to assess the expression of the 3' transcripts in human brain and other tissues. To achieve these aims, the PhyloP score from the Zoonomia project and exon expression level from GTEx V8 data were investigated. The analyses were based on the data from chapter 3, which found the 3' transcripts of LTCC genes to be complete transcripts arising from de novo TSSs in the 3' region of the full-length LTCC genes.

### **4.6.1. Findings from the conservation study**

The 3' transcripts of LTCC genes were mapped to the PhyloP score V2 via the UCSC genome browser, which showed that 5'UTR seemed to be more conserved than introns (Figure 4.2). Additional analyses in this chapter showed that the 5'UTR is significantly more conserved compared with the exonic regions and intronic regions of the 3' transcript of LTCC genes except for the 3' transcript of *CACNA1S* that the 5' UTR is not significantly more conserved than its

exonic regions (Figure 4.3). These show that the 5'UTR of the 3' transcript of LTCC genes may contain functional regions like the alternative 5'UTR of *SCN8A* (Drews et al., 2007). *SCN8A*, which encodes the Nav1.6 channel, produces an alternative transcript arising from an alternative transcription start site containing untranslated exon 1c (Drews et al., 2005, 2007). The untranslated exon 1c was known to be conserved across humans, mice, chicken, and zebrafish and contains a transcription factor binding site (Drews et al., 2007). Therefore, 5'UTR of the 3' transcript of LTCC genes may also have a regulatory function (Byeon et al., 2021).

#### **4.6.2. Findings from the expression study**

To verify whether the expression of the LTCC 3' transcripts start from the novel TSS or not, and the expression pattern of the 3' transcript of LTCC genes in the human tissues, exon expression level (read counts) data set from GTEx was explored. In previous chapters, 3' transcripts of LTCC genes with their own TSS were identified in the human brain using 5' RACE. The exon expression analyses corroborated the findings and showed that the expression of LTCC 3' transcripts initiate from the identified novel TSS.

The analyses showed that the exon expression level increases once transcription is initiated (Figure 4.4 – 4.7). However, each 3' transcript of LTCC genes showed different patterns in changes of exon expression level. Firstly, the 3' transcript of *CACNAIS* showed significant increases in exon expression level after the transcription initiation in both tissue types (non-brain and brain). Additionally, once the transcription was initiated, expression of exon was not significantly increased. Moreover, fold change analysis showed that the exon expression level increases after the transcription is initiated in both tissue types. These suggest that the 3' transcripts of *CACNAIS* are not exclusively expressed in the brain. Secondly, exon expression level changes of the 3' transcript of *CACNAIF* were not significant in non-brain tissues whereas the expression

level changes significantly after the transcription initiation in the brain tissues. The expression level did not show significant increase when comparing expression level of exon where the transcription was initiated to that of the one exon downstream. Fold change analyses showed that expression level increases by 1% in non-brain tissues whereas expression level increases by 200% after transcription is initiated in the human brain tissues. These imply that *CACNAIF* produces brain tissue specific 3' transcripts from the de novo TSS.

The 3' transcript of *CACNAIC* did not show a significant increase in exon expression level when comparing exon 46 (1 downstream of the exon where TSS is located) to exon 48 (1 upstream of the exon where TSS is located) in non-brain tissues whereas the same comparison showed the significant increase in brain tissues. This shows that the 3' transcript of *CACNAIC* would have brain tissue specific expression. Moreover, fold change analysis also showed that exon 46 has a higher expression level than exon 47, and exon 48 has higher expression level than exon 47 in both tissue types. Interestingly, exon 47 in *CACNAIC* has significantly lower expression level compared with the expressions of exons downstream and upstream in both non-brain tissues and brain tissues (Figure 4.6). This may be because the exon is alternatively spliced. Alternative splicing of the C-terminal region of cardiac Cav1.2 in both human and rat was reported previously (Klößner et al., 1997; Wei et al., 1994). However, in rat myocytes, exon 47 was not found to be alternatively spliced (Fan et al., 2005) meaning that exon 47 may be alternatively spliced in human tissues. Finally, for the 3' transcript of *CACNAID*, the exon expression level did not significantly change after the transcription initiation in both tissue types. Also, fold change analysis showed that exon 46 has higher expression than exon 47, and exon 48 has higher expression than exon 47.

The fold change analyses imply the existence of the 3' transcript of LTCC genes, especially *CACNAIC* and *CACNAID*, in the human brain. In chapter 2, the 3' transcripts



of *CACNA1C* and *CACNA1D* were not sequenced due to their high expression in the human brain. As the method was PCR-amplicon sequencing, it would be unclear to distinguish whether the sequenced results are the actual 3' transcript itself or just amplified from the full-length transcript.

The production of alternative transcripts with tissue-specific expression can also be observed from the other genes (Martin et al., 2007; W. Tan et al., 2007). Neuregulin 1 gene (*NRG1*) is also known to use multiple promoters to generate different isoforms with tissue-specific expression. *NRG1* has 9 alternative promoters producing distinctive transcripts showing different expression patterns (Steinhorsdottir et al., 2004a). Type I and Type II mRNAs were detected in other human tissues except lymphocytes and skeletal muscle, and Type III mRNA was detected in both human heart, lung and adult brain whereas Type IV mRNA showed exclusive expression in the human brain, specifically in the fetal brain (W. Tan et al., 2007). Moreover, LTCC 3' transcripts may show expression not only in tissue specific manner but also in brain region specific manner. For example, *SCN1A* produces multiple transcripts using alternative TSSs, of which, the variant with untranslated exon 1a has higher expression in the thalamus and brain stem, and the variant with untranslated exon 1b has a higher expression in the hippocampus and striatum. Similarly, in comparison to the expression of full-length *SCN2A*, which has high expression in the cerebellum, the variant of *SCN2A* with untranslated exon 1b does not express in the cerebellum but is expressed in the hypothalamus and thalamus (Martin et al., 2007).

#### **4.6.3. Limitations and further studies**

Data in this chapter provide a support for the existence of the 3' transcripts of LTCC genes in the human brain (based on the exon-level analyses of GTEx data) and for the possible functionality of their 5'UTRs (based on their evolutionary conservation as shown by PhyloP scores). However, the transcription mechanisms and protein functions remain to be elucidated. In

chapter 5, localization of proteins predicted to be encoded by the 3' transcripts of LTCC genes will be examined and protein-protein interaction study will be attempted.

Some factors might have affected the exon expression analyses, such as RNA degradation or 3' bias. As this chapter did not focus on functional analysis or complex traits of transcripts, RNA degradation was not controlled during the exon expression analysis. Moreover, 3' bias might affect the trend of higher expression levels of the 3' transcript of LTCC genes. The exon expression data from GTEx were generated using short-read RNA sequencing which sequences poly(A) RNA from the 3' end, meaning that exons near the 3' end would have higher expression levels (read counts) compared to the exons toward the 5' end of the gene. To check the possibility of a 3' bias, exon expression levels of the full-length LTCC genes were examined but they did not show a 3' bias effect (see the Appendix 12).

In the GTEx analysis for this chapter, for convenience, the tissue-type examination included two groups: brain tissues and non-brain tissues. However, more specific categorization could provide a more in-depth understanding. The gene regulatory system applies to not only specific tissue types but also specific cell-type composition in the tissue (The GTEx Consortium, 2020), meaning that the 3' transcripts of *CACNAIF* and *CACNAIC*, which showed brain-specific expression, may display cell-type specific expression pattern within the brain tissues. Moreover, using GTEx data from the portal contains a limitation that gene or variant expression cannot be directly compared using more than two tissues concurrently (Stanfill & Cao, 2021). In order to do so, access to the controlled data set would be needed. Therefore, in this chapter, tissue variability of exon expression could not be investigated. However, the exon expression level of each 3' transcript was examined in all possible tissues from the portal. Moreover, GTEx data samples were

collected from deceased individuals which may affect the expression level data (Stanfill & Cao, 2021).

Finally, analysis of eQTL data from GTEx V8 would provide insight into what might affect the expression of the 3' transcript of LTCC genes in human tissues. The cis-eQTLs could play crucial roles in transcriptional regulation and the possible association with traits or disease states. Not only eQTL but also sQTL data would be useful to investigate because sQTL analysis shows the effect of splicing mechanisms on gene expression. Pairwise tissue analysis of sQTL indicates that splicing mechanisms are more tissue-specific compared to gene expression. Therefore, understanding the effects of splicing variants on disease across tissues and validating splicing effects from the cells in vivo would be a useful further investigation (The GTEx Consortium, 2020).

## **Chapter 5 – Localization of proteins encoded by 3' transcripts of LTCC genes**

### **5.1. Introduction**

#### **5.1.1. Rationale for studying the localization of proteins predicted to be encoded by 3'**

##### **LTCC transcripts**

It is controversial how the C-terminal fragments of LTCC genes are produced, with different studies claiming that they may arise from proteolytic cleavage, alternative splicing or alternative transcription start site usage (Heck et al., 2021; Scharinger et al., 2015). However, previous reports have claimed that the C-terminal region of VGCC genes is localized to the nucleus in the cell and functions as transcription factors (Gomez-Ospina et al., 2006, 2013; Kordasiewicz et al., 2006; L. Lu et al., 2015). CCAT, which is reported to be produced from the alternative TSS in the C-terminal region of *CACNA1C*, translocates to the nucleus in neuron and act as transcription factor (Gomez-Ospina et al., 2013). Therefore, the hypothesis tested in this chapter is that the proteins predicted by the 3' transcripts of LTCC genes would also localize to the nucleus, suggesting that they may also function as transcription factors. I first investigated whether the proteins predicted by the 3' LTCC transcripts contain putative nuclear localization signals (NLSs) and then examined their localization in both human excitable and non-excitable cells.

#### **5.1.2. What is the NLS?**

NLS is short peptide motifs that facilitate the transport of proteins from the cytoplasm to the nucleus (Fu et al., 2018). The NLS is important for mediating the communication between the nuclear and cytoplasmic compartments. Various proteins, including transcription factors, contain a NLS that is detected by nuclear transporters and guides them to be transported to the nucleus via the nuclear pore complex (NPC) (Bernhofer et al., 2018; Lowe et al., 2015). There are two groups of classical NLS: monopartite and bipartite (Bradley et al., 2007). The monopartite NLS is a single

cluster with 4-8 amino acids with four or more arginine or lysine residues, whereas the bipartite NLS is comprised of two clusters of 2-3 amino acids separated by a 9-12 amino acid linker region. These two clusters together determine the localization of protein in the cell. In addition, there are also non-classical NLSs which comprise of 20-30 amino acids with a disordered structure (J. Lu et al., 2021).

### **5.1.3. Aim of the chapter**

The aim of this chapter is to examine the localization of the protein encoded by the 3' LTCC transcripts in human non-excitabile (HEK293T) and excitable (SH-SY5Y) cells. Moreover, co-immunoprecipitation and mass spectrometry were aimed to be conducted so that protein-protein interactions in the cells could be examined. For clarity of nomenclature, proteins encoded by the 3' transcript of *CACNAIS* will be called C-Cav1.1, protein encoded by the 3' transcript of *CACNAIC* as C-Cav1.2, protein encoded by the 3' transcript of *CACNAID* as C-Cav1.3 and the protein encoded by 3' transcript of *CACNAIF* as C-Cav1.4. C-LTCC represent the proteins encoded by the 3' transcripts of LTCC genes.

### **5.1.4. Overview of the experiments**

SeqNLS (Lin & Hu, 2013) was used to investigate whether the peptides encoded by 3' LTCC transcripts contain a putative NLS. The proteins predicted to be encoded by the 3' LTCC transcripts were either synthesized with a C-terminal flag-tag (DYKDDDDK) or cloned into a pcDNA3 expression vector containing an in-frame C-terminal flag-tag (c-Flag pcDNA3), (Sanjabi et al., 2005). The flag-tagged 3' LTCC transcript expression vectors were transfected into the HEK293T and SHSY5Y cells. Once cells were lysed, the size of the proteins was determined through Western blotting. Localizations of the C-LTCC in the cells were examined using confocal immunofluorescence microscopy. As it was examined that the C-LTCC especially the C-Cav1.1

and C-Cav1.4 do not localize to the nucleus, co-immunoprecipitation was used to examine protein-protein interaction in the cells.

## **5.2. Methods**

### **5.2.1. Prediction of NLS in the C-LTCC**

The SeqNLS algorithm (<http://mleg.cse.sc.edu/seqNLS>) (Lin & Hu, 2013) uses two methods to predict the presence of NLSs: sequential pattern mining and linear motif scoring scheme. Most known NLS sequences are composed of well-conserved sequences with variable gaps. Therefore, this algorithm mines sequential patterns of NLS from the given query then, predicts the probability that an NLS is present using linear motif-scoring. Linear motif-scoring scheme is used because NLSs are linear motifs and mostly found in disordered region; therefore, the algorithm predicts disorder scores for all amino acid residues and calculates how well the residues interact with other proteins (Lin & Hu, 2013). Most transcription factors contain one or more NLS located close to the DNA binding domain (Shields & Yang, 1997). SeqNLS predicted over 50% of known NLSs and overall the algorithm shows a high prediction and precision score (Lin & Hu, 2013).

### **5.2.2. C-flag pcDNA3 expression vector**

The C-terminal flag tag (DYKDDDDK) was used because it is small and therefore less likely to interfere with the folding and trafficking of expressed proteins, compared with larger tags. The use of a C-terminal tag maximises the chance that the full protein sequence is inserted into the cell membrane without disturbance to protein folding. The c-Flag pcDNA3 was a gift from Stephen Smale (Figure 5.1; Addgene plasmid # 20011; <http://n2t.net/addgene:20011>; RRID: Addgene\_20011) and used as an expression vector. The expression vector was transformed into DH5 $\alpha$  Chemically Competent Cells (Thermofisher, UK, see methods section 5.2.5) and the

bacterial grow with ampicillin sodium salt (BCBZ9179, Sigma-Aldrich, UK) at a final concentration of 100ug/mL at 37°C. Plasmid was purified using Miniprep kit (Qiagen, UK, see method 5.2.6). In order to ligate the 3' transcript of LTCC genes with the c-Flag pcDNA3 (Addgene plasmid # 20011; <http://n2t.net/addgene:20011>; RRID: Addgene\_20011), HindIII enzyme (Promega, UK) was used as a 5' restriction site and XbaI enzyme (Promega, UK) was used as a 3' restriction site (Figure 5.1). By using two different restriction sites, the fragment could be directionally inserted into the expression vector and transformed to DH5 $\alpha$  Chemically Competent Cells (Thermofisher, UK) in the correct orientation.



Figure 5.1 A vector map of c-Flag pcDNA3 from addgene (Addgene plasmid # 20011; <http://n2t.net/addgene:20011>; RRID: Addgene\_20011). The 3' transcript of LTCC genes were inserted between HindIII and XbaI, marked with a red arrow, transformed to DH5 $\alpha$  Chemically Competent Cells and colonies grown under ampicillin selection.

### 5.2.3. Primer design

Primers were designed to amplify the full coding sequences of 3' LTCC transcripts and to allow them to be inserted in-frame into c-Flag-pcDNA3. Primers were designed using Primer3 plus (Source code is available at <http://sourceforge.net/projects/primer3/>) (Untergasser et al., 2007), A plasmid Editor (ApE; RRID:SCR\_014266) (Davis & Jorgensen, 2022), and checked for specificity using “In silico PCR” from UCSC (University of California, Santa Cruz) Genome Browser (<http://genome.cse.ucsc.edu/>) (Kent et al., 2002).

Nested PCR was performed to increase the specificity of PCR amplifications (Figure 5.2). Therefore, two sets of primers were designed for the 3' transcript of LTCC gene. First round amplification primers were designed to amplify a slightly longer region surrounding the 3' LTCC transcript coding sequences (Table 5.1). The nested PCR amplification used cDNA as a template (see Methods section 2.3.4). First round PCR products were purified using the Monarch<sup>®</sup> PCR & DNA Cleanup Kit (NEB, UK, see methods 2.3.10 in chapter 2) and used as a template for the nested PCR. To facilitate downstream cloning of PCR amplicons, a 6-nucleotide buffer sequence (GCATCT) followed by a HindIII (AAGCTT), and XbaI (TCTAGA) restriction enzyme site was added to the 5' ends of the forward and reverse primers, respectively. For the reverse primer, the endogenous stop codon was mutated by changing one base pair (i.e., TGC to GTA) (Table 5.2). Following this, the whole reverse primer sequence was reverse complemented. Nested PCR products were purified with a Monarch<sup>®</sup> DNA Gel Extraction Kit (NEB, UK; see methods 2.3.8 in chapter 2).



Table 5.23. Primer sequences used for the first round of PCR

Name	Sequences (5' - 3')
<i>CACNAIS</i>	F:GTCCAAGATGACAGAGGAGGAGTG R: CATCTAGCTGCTGAGAGGGAG
<i>CACNAIC</i>	F:CATGAACATGCCTCTGAACAGC R:CTACAAGCAGAGTCGAAAGACACC
<i>CACNAID</i>	F:GGAGGAGTATACATGTGGGAGC R:GAAGCAGACTCCACAGCAGTG
<i>CACNAIF</i>	F: GCGGATGAAACAGAAGCTGC R: GGCAGGAGGTTTATTGAGC

Table 5.24. Primer sequences used for the second round of PCR

Name	Sequences (5' - 3')
<i>CACNAIS</i>	F:GCATCTAAGCTTGTCCAAGATGACAGAGGAGGAG R:AGATGCTCTAGATGCTGTGTGGGCATACCAG
<i>CACNAIC</i>	F:GCATCTAAGCTTGAAGCGGACCAGCATGAAGC R:AGATGCTCTAGACCCACTACAGGCTGCTGACG
<i>CACNAID</i>	F:GCATCTAAGCTTAGTTGCCATGAACATGCCTCTC R:AGATGCTCTAGAGGGGATAACAAGGTGGTGATG
<i>CACNAIF</i>	F:GCATCTAAGCTTGC GGATGAAACAGAAGCTGC R:AGATGCTCTAGAAGGGGTGGGAATACAGAGGG

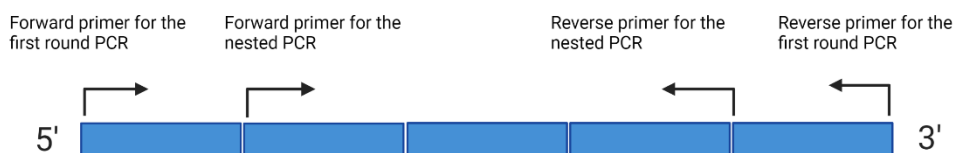


Figure 5.2. Explanation of Nested PCR experiments. The nested PCR technique was used to increase specificity of PCR amplification of the 3' transcript of *LTCC* genes which were inserted to *c-Flag pcDNA3* expression vector. To do so, two sets of primers were designed. The figure was created with BioRender.com

#### 5.2.4. Plasmid design for 3' *CACNA1F* transcript

The 3' transcript of *CACNA1F* could not be amplified, even with the nested PCR method. Therefore, the coding region of *CACNA1F* was synthesized by GeneArt Gene Synthesis (ThermoFisher, UK). The synthesized vector contained the 3' transcript of *CACNA1F*, as well as the buffer sequences, restriction enzyme site sequences of HindIII (AAGCTT), coding sequences (with stop codon sequence silenced), and restriction enzyme site sequences of XbaI (TCTAGA)

and the buffer sequences (Figure 5.3). The synthesised plasmid arrived as 5µg lyophilised product, which was resuspended in 50µl of elution buffer (New England Biolabs, UK) to make 100 ng/µl final concentration. The synthesized 3' transcript of *CACNA1F* vector was diluted to 10ng in 5µl and transformed into DH5α Competent Cells (Thermofisher, UK; see methods 5.2.5). Bacteria were grown under ampicillin selection (see section 5.2.5) and the plasmids purified via QIAprep Spin Miniprep Kit (Qiagen, UK; see methods 5.2.6).

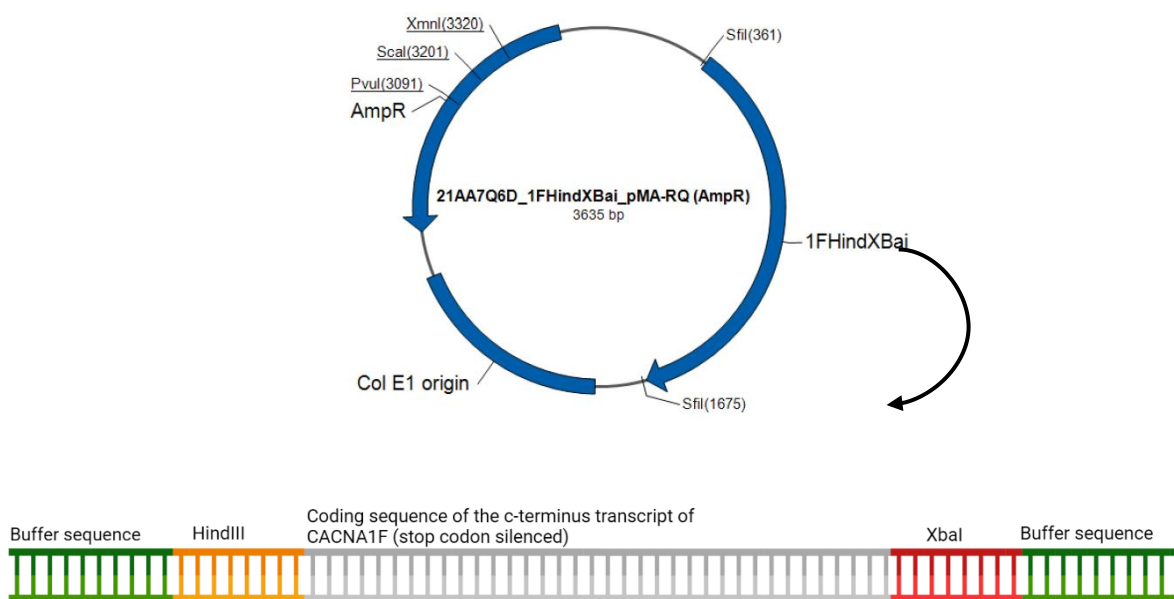


Figure 5.3. Synthesized plasmid indicating locations of Buffer sequence, HindIII, coding sequence, XbaI and buffer sequences. The vector map was provided by GeneArt, Thermofisher. The figure was created with BioRender.com

### 5.2.5. Subcloning of vectors into DH5α competent cells

Vectors (5 µl of 2ng of vectors) were added to 50 µl of DH5α competent cells (Thermofisher, UK) and mixed gently (without pipetting up and down). The mixture was incubated on ice for 30 minutes, followed by 20 seconds of heat shock at 42°C. The samples were incubated on the ice for 2 minutes. After addition of 950 µl of S.O.C. medium (Thermofisher, UK) samples

were shaken at at 37°C for 1 hour at 225 rpm. 200 µl of the transformation was spread on pre-warmed LB agar plates with ampicillin (final concentration of 100 ug/mL) (Sigma-Aldrich, UK). The plates were incubated overnight at 37°C.

### **5.2.6. Plasmid purification**

QIAprep Spin Miniprep Kit (Qiagen, UK) was used to purify plasmids from the cultured transformants. Cultures were aliquoted into 2 ml Eppendorf tube and centrifuged at 16,000 x g for 1 minute to pellet the bacteria. The supernatant was discarded, and this step was repeated. The bacterial pellet was resuspended in 250 µl of buffer P1 (a resuspension buffer containing RNase A and lyse blue dye). Then, 250 µl of buffer P2 (lysis buffer) was added and the solution was inverted 6 times. Within 5 minutes, 350 µl of buffer N3 (a neutralization buffer) was added and the resultant solution was inverted 6 times. The mixture was centrifuged at 17,000 x g for 10 minutes. The 800 µl of supernatant was added to a QIAprep Spin column, centrifuged at 17,000 x g for 1 minute and the flow through discarded. 500 µl of buffer PB was added, centrifuged at 17,000 x g for 1 minute, and the flow-through was discarded. Finally, 750 µl of buffer PE was added, centrifuged at 17,000 x g for 1 minute, and the flow-through was discarded. The column was added to the new tube and centrifuged at 17,000 x g for 1 minute. The column was added to the new tube and eluted in 50 µl of elution buffer (10 mM Tris·Cl, pH 8.5) for 1.5 minutes and centrifuged at 17,000 x g for 1 minute. The concentration of miniprep vectors were measured via nanodrop ND-1000 spectrophotometer.

For co-immunoprecipitation assay, plasmids were purified using HiSpeed Plasmid Maxi Kit (12662, Qiagen, UK). Bacteria were centrifuged at 6000 x g for 15 minutes at 4°C. The pellets were resuspended in 10mL of buffer P1 (a resuspension buffer containing RNase A and lyse blue). Then, 10 mL of buffer P2 (a lysis buffer) was added, mixed by inverting 5 times and incubated at

room temperature for 5 minutes. 10 mL of prechilled buffer P3 (a neutralization buffer) was added and mixed thoroughly by inverting 5 times. The lysate was poured into the QIAfilter Cartridge and incubated at room temperature for 10 minutes. HiSpeed Tip was equilibrated with 10 mL of buffer QBT until the buffer enters the resin. Gently insert the plunger into the QIAfilter Cartridge and filter the plasmid lysate into the equilibrated HiSpeed Tip. Once the lysate passed through, HiSpeed Tip was washed with 60 mL buffer QC. The DNA was eluted with 15 mL of buffer QF. The eluted plasmids were precipitated with 10.5ml (0.7 volumes) of room temperature isopropanol. The mix was incubated at room temperature for 5 minutes. During the incubation, the plunger was removed from 30 ml syringe and attached the QIAprecipitator Maxi Module onto the outlet nozzle. QIAprecipitator was placed over a waste bottle, the eluate/isopropanol mixture was transferred into 30 ml syringe, and the plunger was inserted. The eluate/isopropanol mixture was filtered through the QIAprecipitator using slow and constant pressure. The QIAprecipitator was removed from the 30 ml syringe and the plunger was pulled out. QIAprecipitator was re-attached and 2 ml of 70% ethanol was added to the syringe. The plasmid was washed with 70% ethanol by inserting the plunger and pressing through the QIAprecipitator with constant and slow pressure. QIAprecipitator was removed from 30 ml syringe and the plunger was pulled out. QIAprecipitator was attached to 30 ml syringe again and the plunger was inserted to dry the membrane by pressing air through the QIAprecipitator quickly and forcefully. This step was repeated. The outlet nozzle of the QIAprecipitator was dried out with absorbent paper to prevent ethanol carry-over. The plunger was removed from a new 5 ml syringe and the QIAprecipitator was attached onto the outlet nozzle. The outlet of the QIAprecipitator was hold over a 1.5 ml collection tube and 1 ml of Buffer TE was added to the 5 ml syringe. The plunger was inserted and the plasmid was eluted into the collection tube using constant and slow pressure. It is was important to make sure the outlet of the

QIAprecipitator was held over the collection tube properly as eluate can drip through the QIAprecipitator before the syringe barrel was inserted. The QIAprecipitator was removed from the 5 ml syringe, the plunger was pulled out and the QIAprecipitator was reattached to the 5 ml syringe. The eluate was transferred to the 5 ml syringe and elute for a second time into the same 1.5 ml tube. This re-elution step was conducted to ensure that the maximum amount of plasmids is solubilized and recovered. The concentration of purified plasmids was measured using a Nanodrop ND-1000 spectrophotometer.

### 5.2.7. Restriction Enzyme Digestion

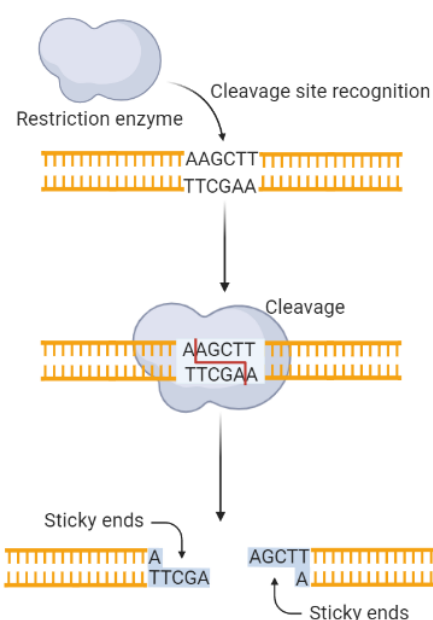
Restriction enzyme digestion was performed on the PCR amplified 3' transcripts of *CACNAIS*, *CACNAIC*, and *CACNAID*, on the synthesized vector of 3' *CACNA1F* transcript, and c-Flag pcDNA3, in the sequential digestion manner. Restriction enzyme digestion creates sticky ends so that the inserts and expression vectors can ligate together efficiently and in the correct 5'→3' orientation (Figure 5.4). The digestion protocol is indicated in Table 5.3. The digested PCR amplified 3' transcript of *CACNAIS*, *CACNAIC*, *CACNAID*, and synthesized vector of 3' transcript of *CACNA1F* were purified using Monarch<sup>®</sup> DNA Gel Extraction Kit (NEB, UK; see methods 2.3.8 in chapter 2). The concentrations of purified samples were measured via Nanodrop ND-1000 spectrophotometer.

Table 5.25. Sequential restriction enzyme digestion methods

PCR products and synthesized <i>CACNA1F</i> vector		c-Flag pcDNA3 expression vector	
Name	Amount (μl)	Name	Amount (μl)
DNA	16.8	Vector	10.0
RE buffer E (Promega, UK; R005A)	2.0	RE buffer E (Promega, UK; R005A)	2.0
Bovine Serum Albumin (BSA, 10mg/ml; Promega, UK; R396E)	0.2	Bovine Serum Albumin (BSA, 10mg/ml; Promega, UK; R396E)	0.2

HindIII (Promega, UK; R604A)	0.5	HindIII (Promega, UK; R604A)	0.5
NF water	0.00	NF water	6.8
Total	19.5	Total	19.5
Incubated at 37°C for 1 hour and 65°C for 15 minutes			
XbaI	0.5	XbaI	0.5
RE buffer D	2.0	RE buffer D	2.0
Incubated at 37°C for 1 hour and 65°C for 15 minutes			

## HindIII



## XbaI

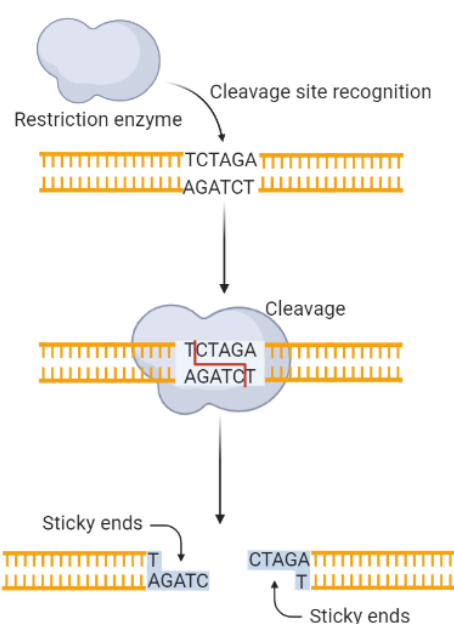


Figure 5.4. Graphic description of restriction enzyme digestion which creates sticky ends. The figure was created with BioRender.com

### 5.2.8. Dephosphorylation of digested expression vector

Following c-Flag pcDNA3 expression vector digestion, the vector was dephosphorylated using Calf Intestinal Alkaline Phosphatase (CIAP) (Promega, UK) to inhibit self-ligation. CIAP was diluted in CIAP 1X reaction buffer to a final concentration of  $0.01\mu\text{g}/\mu\text{l}$ . Each pmol of DNA ends require 0.01u CIAP. The mix (Table 5. 4) was then incubated at 37°C for 30 minutes. 0.6  $\mu\text{l}$

of diluted CIAP was then added and incubated firstly at 37°C for 30 minutes, and then 65°C for 15 minutes. The dephosphorylated vectors were purified using Monarch<sup>®</sup> PCR & DNA Cleanup Kit (NEB, UK; see methods 2.3.10 in chapter 2). The concentration of purified expression vector was measured using a Nanodrop ND-1000 spectrophotometer.

Table 5.26. Dephosphorylation methods using CIAP

Name	Amount (µl)
DNA	1µg (20 µl)
CIAP 10x buffer	5.0
Diluted CIAP (0.01u/ µl; up to 5 µl)	0.6
NF water	24.4
Total	50
Incubated at 37°C for 30 minutes	
Diluted CIAP	0.6
Incubated at 37°C for 30 minutes and 65°C for 15 minutes	

### 5.2.9. Ligation of inserts into c-flag pcDNA3 expression vector

Purified PCR amplified 3' transcript of *CACNAIS*, *CACNAIC*, and *CACNAID*, and synthesized *CACNAIF*, were ligated using three different molar ratios (1:1, 2:1, 3:1; See Appendix 13) into the digested and purified c-Flag pcDNA3 expression vector using T4 DNA ligase (Promega, UK; Figure 5.5). As ligation of 3' transcripts of LTCC genes with the expression vector was failing, the effect of dephosphorylation was examined. Therefore, the expression vector was used in two different conditions, dephosphorylated expression vector and non-dephosphorylated expression vector. Prior to the ligation, the concentration of each insert and expression vector was measured with a Nanodrop ND-1000 spectrophotometer. Based on the concentrations, the ligation method was optimized (See Appendix 13). All samples were ligated at 22°C for 3 hours, transformed into DH5α competent cells (Thermofisher, UK; see 5.2.5) spread onto LB-agar plates containing ampicillin (100 µg/ml) and grown over-night at 37°C Individual colonies were picked

grown overnight at 37°C in 5 ml of LB containing ampicillin (100 µg/ml), and miniprep using QIAprep spin Miniprep kit (Qiagen, UK; see methods section 5.2.6). In order to check which clones contained the proper insert, purified plasmids were sequentially restriction enzyme digested with HindIII (Promega, UK) and XbaI (Promega, UK) (see Methods 5.2.7). *CACNAIF* transformants were verified to have the proper insert by sanger sequencing using CMV-cDNA F primer (Source Bioscience, UK). The result of sanger sequencing is in Appendix 14. On the other hand, the 3' transcript of *CACNAIS*, *CACNAIC*, and *CACNAID* were not successfully ligated to the expression vector as evidenced by the lack of colonies on the selection plates.

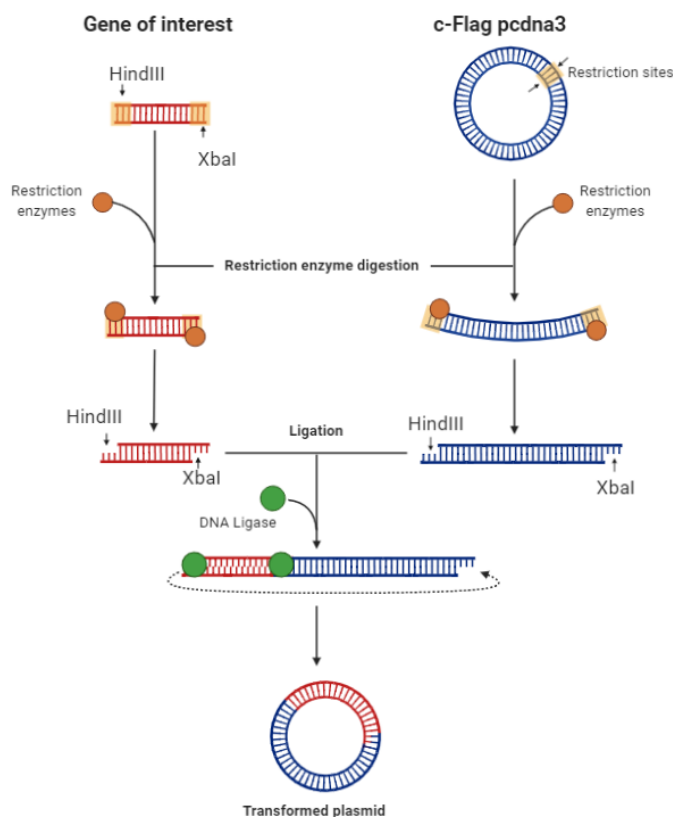


Figure 5.5. The ligation process of the 3' transcript of *LTCC* genes to the *c-flag pcDNA3*. The figure was created with BioRender.com



### 5.2.10. Optimizing ligation of the 3' transcript of *CACNAIS*, *CACNAIC*, and *CACNAID*

The synthesized 3' transcript of *CACNAIF* was the only insert to be successfully ligated with the expression vector and transformed to the DH5 $\alpha$  competent cells. As the ligation was not successful for the 3' transcript of *CACNAIS*, *CACNAIC*, and *CACNAID* following many attempts, each step of the process was analyzed to determine which step was the cause of the unsuccessful ligation (Figure 5.6).

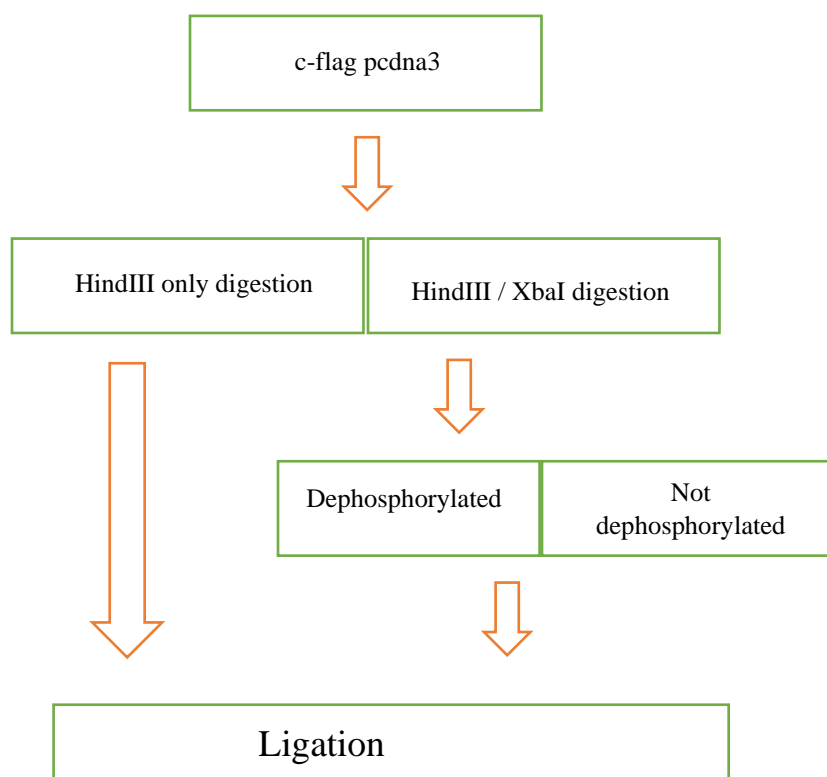


Figure 5.6. Flow chart of optimization for dephosphorylation and ligation step

Firstly, the expression vector was digested in two different ways (Figure 5.6): digestion only by HindIII enzyme (Promega, UK) and sequential digestion by HindIII (Promega, UK) then XbaI enzymes (Promega, UK). Secondly, the expression vector digested with both restriction enzymes were divided into either dephosphorylated, or not dephosphorylated, before ligation.

After the dephosphorylation step with purification, the expression vector was ligated without any inserts to examine the efficiency of ligation and dephosphorylation. The ligation mixes (Table 5.5) were incubated at 22°C for 3 hours. After the ligation step, the vectors were subcloned to the DH5 $\alpha$  competent cells (Thermofisher, UK). This was done to determine whether the transformation step or the ligation step was unsuccessful. The assumption is that if all the steps were successfully performed, after transformation, the dephosphorylated expression vector should have no transformants as there would be no self-ligation and no insert was ligated, whereas HindIII digested the expression vector without dephosphorylation it should have many transformants due to self-ligation.

*Table 5.27. Ligation method of expression vector to determine if restriction enzyme digestion and dephosphorylation was successful*

Name	Amount ( $\mu$ l)
Vector	4.03
10X buffer	1.00
Ligase	0.33
Nuclease Free water	4.64
<i>Total</i>	<i>10</i>

As a result, HindIII only digested expression vector without dephosphorylation, which was ligated with T4 DNA ligase (Promega, UK), had many transformants. Dephosphorylated expression vector and expression vector digested with both restriction enzymes did not have any transformants. This means that dephosphorylation and ligation steps were performed correctly therefore, the reason why the inserts were not ligated to the vector may be due to inefficient RE digestion of the PCR products.

As the 3' transcript of *CACNAIS*, *CACNAIC*, and *CACNAID* were not ligated successfully and due to time limitations, the coding regions of the 3' transcript of each gene were synthesized

with flag tag (DYKDDDDK) by GeneArt Gene Synthesis (ThermoFisher, UK). The synthesized sequence consisted in an order of the coding region sequence of the insert with stop codon silenced, buffer sequence (GCATCT), flag tag sequence (DYKDDDDK) and stop codon (TAA). Upon receipt, the vectors were transformed into DH5 $\alpha$  competent cells (see methods 5.2.5). The plasmids were grown up and after purification using a QIAprep spin Miniprep kit (Qiagen, UK; see methods section 5.2.6), the sequence was confirmed using Sanger sequencing (Source BioScience, UK). The results of Sanger sequencing are in the Appendix 14.

### **5.2.11. Plasmid precipitation**

The purified plasmids were precipitated with isopropanol (BCCC9126, Sigma-Aldrich, UK). For each sample, 10% volume of 3M Na Acetate was added. Then, 70% volume of room-temperature isopropanol (BCCC9126, Sigma-Aldrich, UK) was added then centrifuged at 15,000 x g for 30 minutes at 4°C. The pellets were then washed with 200  $\mu$ l of room temperature 70% ethanol (Sigma-Aldrich, UK) and centrifuged at 15,000 x g for 10 minutes. After the supernatant was discarded, the pellets were air-dried, and DNA was resuspended in 20  $\mu$ l of elution buffer. The concentration of plasmid was measured using a Nanodrop ND-1000 spectrophotometer.

### **5.2.12. Seeding the HEK293T cell line and the SH-SY5Y cell line**

HEK293T cells (85120602-1VL, Merck life science Ltd, UK) are non-excitable human cells. SH-SY5Y cells (94030304-1VL, Merck life science Ltd, UK) are excitable human neuroblastoma-derived cells. The HEK293T cell line was used as it has high transfection efficiency so that before attempting transfection into the SH-SY5Y cells, the size of the vector encoded protein can be examined. Cells were thawed and grown in initial plating media. Initial plating media consisted of Dulbecco's Modified Eagle's Media (DMEM high glucose, no glutamine) (11960044, ThermoFisher, UK), containing 10% Fetal Bovine Serum (FBS)

(65B001341, Lonza), 1X Glutamax (35050038, Thermofisher, UK) and 1X penicillin/streptomycin (15140122, Thermo Fisher Scientific, UK). The cells were grown in 10cm tissue culture dishes (Cell+, Sarstedt, Germany) in 10mL of the initial plating medium. The cells were washed with Dulbecco Phosphate Buffered Saline without  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  (DPBS without  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$ ; 2007927, Thermo Fisher Scientific, UK) and trypsinized with 0.25% Trypsin EDTA with phenol red (25200056, Thermo Fisher Scientific, UK). The trypsinized cells were pelleted by centrifugation (5 min at 400 x g at RT) and the cell pellet resuspended in 10mL of initial plating media minus antibiotics. Cells were counted using a hemocytometer and  $5 \times 10^5$  cells were seeded into 8cm<sup>2</sup> culture dishes (Cell+, Sarstedt, Germany) in 2 ml of media minus antibiotics.

### **5.2.13. Transfection of HEK293T cell line and SH-SY5Y cell line**

Once cells were 70-80% confluent, the plasmids of 3' transcript of LTCC genes and empty expression vector (c-Flag pcDNA3) were transfected to HEK293T cells and SH-SY5Y cells using Lipofectamine 3000 transfection reagent (L3000008, Thermofisher, UK). 5  $\mu\text{l}$  of Lipofectamine 3000 (Thermofisher, UK) was diluted into 125 $\mu\text{l}$  of Opti-MEM Reduced Serum Medium (2407807, Thermofisher, UK). 2.5 $\mu\text{g}$  of Plasmids were then diluted in 125  $\mu\text{l}$  of OptiMEM Reduced Serum Medium (2407807, Thermofisher, UK), and 5  $\mu\text{l}$  of P3000 Reagent (Thermofisher, UK) was added. The diluted plasmids with P3000 were mixed with diluted lipofectamine 3000 in 1:1 ratio and then incubated for 15 minutes at room temperature. The DNA-lipid complex was added dropwise to the HEK293T cells. Cells were then examined at 3 different timepoints -24-, 48- and 72-hours post transfection- to assess protein expression kinetics. SH-SY5Y cells were similarly transfected, but because transfected HEK293T cells showed a lot of cell detachment at 72 hours post transfection, SH-SY5Y transfection were only analysed at 24 and 48hr time points.

#### **5.2.14. Harvesting the HEK293T cell line and the SH-SY5Y cell line**

Whole cell lysates were prepared for analysis of protein expression. Media was removed and cells were washed with DPBS without  $\text{Ca}^{2+}$ ,  $\text{Mg}^{2+}$  (2007927, Thermofisher, UK). 120  $\mu\text{l}$  of RIPA buffer (RO278, Sigma-Aldrich, UK) containing 1x complete EDTA-free Protease Inhibitor Cocktail (11873580001, Roche, UK) was added to the plate. The cells were removed with a cell scraper (Sarstedt, Germany). The scraped cells were incubated on ice for 15 minutes and mixed by pipetting up and down. The lysates were centrifuged at 13,000 x g for 5 minutes at 4°C. The supernatant of each sample was aliquoted, snap-frozen on dry ice and stored at -80°C for downstream analysis.

#### **5.2.15. Measuring protein concentrations**

At first, Bradford assay was conducted to measure the protein concentration. For the standards, 10 mg/ml BSA and 1/25 dilution of RIPA buffer (SLBZ0792, Sigma-Aldrich, UK) in PBS was used. 5  $\mu\text{l}$  of both standards and samples were loaded on to 96-well Microplate which are round wells with flat bottom (E2996-1600; STARLAB, UK), and 200  $\mu\text{l}$  of Bradford reagent (B6916, Sigma-Aldrich, UK) was added to each well. The assay was measured using a SPECTRAMAX 190 with SOFTmax PRO 4.0. However, it was determined that Bradford assay is not compatible with the detergent present in the RIPA buffer. Therefore, DC protein assay (5000112, BIORAD, U.S.) was used instead. Proteins were diluted 1 in 10 in RIPA buffer containing 1x protease inhibitor cocktail. Working reagents were made according to the methods provided by supplier. The plate was incubated on room temperature for 15 minutes, and absorbance was read at 750nm using a SPECTRAMAX 190 with SOFTmax PRO 4.0.

### 5.2.16. Western blotting

10 µg of protein lysate in 20 µl total was used, if needed, samples were diluted in RIPA buffer (RO278, Sigma-Aldrich, UK). An equal volume of sample buffer, which consists of 95% of 2x Laemmli sample buffer (1610710, BioRad, USA) and 5% of 2-Mercaptoethanol (1610710, BioRad, USA), was added to each protein sample. Samples were heated at 90°C for 10 minutes. Running buffer was made with 100 mL of 10x Tris/glycerine SDS (cat number, BioRad, USA) and 900 mL of dH<sub>2</sub>O. Samples were run on 7.5% Mini-protean TGX Gels (4561023, BioRad, USA). 10µl of Precision plus protein dual colour standard (1610374, Bio-Rad, USA) was used as a molecular weight ladder. The gel was run at 100V until ladders were fully resolved. Protein was transferred to Immobilon-P Transfer Membranes (IPVH00010, Merck Millipore, Germany; pre-wet in Methanol before use). Transfer buffer was made with 100 ml of 10x Tris/Glycine buffer (BioRad, USA), 700 mL of dH<sub>2</sub>O, and 200 mL of Methanol. Transfer was carried out at 25V overnight. Membranes were incubated in Ponceau S solution BioReagent (p7170-1L, Merck life sciences, UK) to ensure successfully protein transfer. The membrane was then rinsed in PBS containing 0.05% Tween-20 (PBST) and blocked with 4% powdered skim milk (Marvel, Premier Foods, UK) in PBST for 1 hour. For the primary antibody, monoclonal anti-FLAG M2 antibody produced in mouse (F1804-50UG, Merck life science Ltd, UK) diluted 1:5000 in blocking buffer and incubated for 1 hour at room temperature. Then, the membrane was washed 3 times with PBST for 10 minutes. After the wash, the membrane was incubated in secondary antibody, Amersham ECL Mouse IgG HRP-linked whole Ab (Na931-100UL, VWR international Ltd, UK), diluted 1:5000 in blocking buffer, for 1 hour at room temperature. The membrane was washed with PBST 3 times for 10 minutes. Bound antibody was visualized by incubating the membranes in ECL Prime

Western Blotting Detection Reagents (17495087, Cytiva, USA) and exposing them to Amersham Hyperfilm ECL (18X24cm, 28906836, Cytiva, USA). Membranes were then stripped using Restore Western Blot stripping buffer (XA337524, Thermofisher, UK) and reprobred with anti-GAPDH antibody as a protein loading control. The primary antibody (rabbit anti-GAPDH mAb, clone, 2118s, cell signaling technologies Europe BV) was used at 1:5000 dilution in 4% of powdered skim milk (Marvel, Premier Foods, UK) in PBST. Secondary antibody for GAPDH (Amersham ECL Rabbit IgG, HRP-linked whole Ab (from donkey) (NA934-100UL, Cytiva, USA) was diluted in 1:5000 in 4% of powdered skim milk (Marvel, Premier Foods, UK) in PBST.

### **5.2.17. Immunofluorescent staining and imaging**

HEK293T and SH-SY5Y cells were seeded at  $2.5 \times 10^5$  cells per well on 8 well removable chamber slides (16260661, Thermofisher, UK). Transfections were performed as described above (see methods 5.2.13) and the cells fixed 24 hours after transfection for immunofluorescent staining. After washing the cells with 400 $\mu$ l of HBSS (14025092, Thermofisher, UK) to remove cell culture medium cells were fixed in 300  $\mu$ l of 4% PFA/PBS at room temperature for 10 minutes. Cells were then rinsed with 500  $\mu$ l of DPBS without  $\text{Ca}^{2+}$  or  $\text{Mg}^{2+}$  (2007927, Thermofisher, UK) twice. Cells were permeabilized by incubating in 300 $\mu$ l of 0.5% TritonX-100 in PBS for 15 minutes at room temperature. The permeabilized cells were rinsed with 500  $\mu$ l of 0.1% Triton X-100 in PBS. The cells were blocked with 200  $\mu$ l of 3% BSA in 0.1% Triton X-100 in PBS. Mouse monoclonal ANTI-FLAG M2 Antibody (F1804-50UG, Merck life sciences, UK) was used as a primary antibody which was diluted to 1 in 1000 in 3% BSA in 0.1% Triton X-100 in PBS and then applied to the cells, which were incubated overnight at 4°C. The cells were then washed with 300 $\mu$ l of 0.1% TritonX-100 in PBS for three times and incubated for 1 hour at room temperature in Alexa Fluor 488 donkey, anti-mouse IgG, diluted 1:400 in 3% BSA in 0.1% tritonx-100 in PBS containing

1/1000 dilution of phalloidin CruzFluor 647 Conjugate (sc363747, Insight Biotechnology Ltd, UK). Cells were washed twice with with 300  $\mu$ l PBS containing 0.1% Triton X-100 followed by a final wash in PBS to remove excess detergent. After removal of the chamber, 30  $\mu$ l of Vectashield Antifade Mounting medium with DAPI (ZG1202, 2BScientific, UK) was added to each well position and a glass coverslip was added. The slides were imaged using a Zeiss LSM900 with Airyscan 2 detection and zen blue 3.3 in collaboration with Edward Drydale and James Bancroft at the Wellcome Center for Human Genetics. For imaging processing, FIJI (Schindelin et al., 2012), Airyscan 2 detection and Photoshop CS6 Color Replace was used.

#### **5.2.18. Co-Immunoprecipitation of the C-LTCC in HEK293T cells**

HEK293T cells were seeded ( $6.0 \times 10^6$  per 10cm culture dish in 10 ml of media) and transfected with the plasmids of 3' transcript of LTCC genes (For method, see section 5.2.13 in chapter 5). They were harvested at 24 hours (Method 5.2.14). Prior to preparation of cell lysates, the TissueLyserLT (Qiagen, UK; 85600) was pre-chilled at  $-20^{\circ}\text{C}$ , and 200  $\mu$ l of glass beads, acid washed (G8772; Sigma-Aldrich, UK) were prepared in screwcap Eppendorf tubes and chilled on ice. For making cells lysates, culture media was removed, and cells were washed with 10 ml PBS. 1 ml of PBS containing 1x cOmplete Protein Inhibitor Cocktail (990  $\mu$ l of PBS with 10  $\mu$ l of 100x Protease Inhibitor Cocktail (Roche, UK; P8340-1ML)) was then added to the culture dishes and the cells detached using a cell scraper. The scrapped cells were transferred to a screw cap tube on ice. The scraped cells were transferred into the pre-chilled tubes containing glass beads. Cells were lysed using TissueLyserLT (Qiagen, UK; 85600) for 1 minute at 30Hz then put on the ice for 2 minutes. The disruption step was repeated an additional 3 times. The cell lysates were transferred to protein low bind tubes (Sarstedt, Germany) and centrifuged at  $13\,000 \times g$  for 10 minutes at  $4^{\circ}\text{C}$  to pellet any insoluble material. 10  $\mu$ l aliquots were made as 1% input of samples being co-



immunoprecipitated and the rest were used for co-immunoprecipitation. Tubes containing 200  $\mu$ l of anti-flag (DYKDDDK) magnetic beads (Life technologies, UK; A36797) were placed on the magnetic rack on ice. The supernatant was removed, and the beads washed twice with 500  $\mu$ l of PBS buffer. Tubes were mixed by inversion during each wash. 980  $\mu$ l of cell lysate was added to the anti-flag (DYKDDDK) magnetic beads (Life technologies, UK; A36797) and mixed on a HulaMixer at 4°C for overnight. The samples were then put on the magnetic rack on ice and the supernatant collected to checking whether there are proteins unbound to anti-flag magnetic beads (Life technologies, UK; A36797). The beads were washed twice with 500  $\mu$ l of PBS containing 1x cOmplete Protease Inhibitor Cocktail (Roche, UK; P8340-1ML). During each wash, samples were mixed on a HulaMixer at 4°C for 5 minutes, and then put on the magnetic rack on ice to allow removal of the wash (supernatant). 500  $\mu$ l of Milli-Q water was used to rinse the beads and the supernatant was removed on the magnetic rack on ice. The beads were eluted for 10 minutes with 100  $\mu$ l of Glycine elution buffer (90 $\mu$ l of Glycine, pH 2.0 (Merck, UK) plus 10 $\mu$ l of 1.5M NaCl).

#### **5.2.19. Coomassie blue Stain**

Coomassie blue stain was used to examine the outputs of co-immunoprecipitation. 10 $\mu$ l of Precision plus protein dual colour standards (1610374, Bio-Rad, USA) were run on the gel to estimate protein molecular weights. 20  $\mu$ l of each sample in 1 x sample buffer was prepared by dilution with water and 4 x Laemmli (BioRad, UK, 1610747) containing 10% 2-Mercaptoethanol (1610710, BioRad, USA) . The samples in were incubated at 90°C for 10 minutes and run on 7.5% Mini-protean TGX Gels (4561023, BioRad, USA) at 100V for 1.5 hours. The gel was washed thrice in 100mL dH<sub>2</sub>O for 5 minutes and incubated with the 20ml of Simply blue safe stain (Life technologies, UK; LC6060) for 1 hour. The gel was then washed with 100ml dH<sub>2</sub>O for 1 hour, after which 20ml of 20% NaCl was added, and the gel left overnight to destain.

### 5.3. Results

#### 5.3.1. The proteins encoded by the 3' LTCC transcripts are predicted to localize to the nucleus

SeqNLS was used to explore the sequences of 3'LTCC NLSs. The algorithm was set to define the sequence highly likely to be NLS if the sequence scores higher than 0.8. All 3' LTCCs were found to contain a nuclear localization signal that score higher than 0.8, which means that they are highly likely to be translocated to the nucleus (Table 5.6).

*Table 5.28. Table of nuclear localization signal prediction. The prediction score is color coded: navy is 0.1-0.3, blue is 0.3-0.5, cyan is 0.5-0.7, green is 0.7-0.8, yellow is 0.8-0.86, orange is 0.86-0.89 and red is higher than 0.89. It is difficult to distinguish navy from blue therefore, I wrote the sequence color coded as navy. There is no sequences in 3' Cav1.1 that are color coded in navy. However, for 3' Cav1.2, the sequence **PEEDKRD** is coded in navy, for 3' Cav1.3, **LLP-RRSSFNF**ECLR, and for 3' Cav1.4, **KGLL**.*

3' LTCC	Prediction result
3' Cav1.1	CTDLSKMTEEECRGYVVYKDGDPMQIELRHREWHVHSDF HFDNVL SAMMSLFTVSTFEGWPQLLYKAIDSNAEDVGPI YNNRVEMAIFFIIYIILIAFFMMNIFVGFVIVTFQEQQE TEYKNCELDKNQRQCVQYALKARPLRCYIPKNPYQYQVW YIVTSSYFEYLMFALIMLNTICLGMQHYNQSEQMNHISD ILNVAFTIIFTEMLKLMFAFKARGYFGDPWNVDFLIV IGSIIDVILSEIDTF LASSGGLYCLGGGCGNVDPDESAR ISSAFFRLFRVMRLIKLLSRAEGVRTLLWTFIKSFQALP YVALLIVMLFFIYAVIGMQMFGKIALVDGTQINRNNNFQ TFPQAVLLLFRCATGEAWQEILLACSYGKLCDPESDYAP GEEYTCGTNFAYYYFISFYMLCAFLVINL FVAVIMDNFD YLTRDWSILGPHHLDEFKAIWAEYDPEAKGRIKHLDVVT LLRRIQPPLGFGKFCPHRVACKRLVGMNMLNSDGTVTF NATLFALVRTALKIKTEGNFEQANEELRAIIKKIWKRTS MKLLDQVIPPIGDDEVTVGKFYATFLIQEHF <b>RKFMKRQE</b> <b>EYYGYRPKK</b> DIVQIQAGLRTIEEEAAPEICRTVSGDLAA EEELERAMVEAAMEEGIFRRTGGLFGQVDNFLERTNSLP PVMANQRPLQFAEIEEMEEMESPVLEDFPQDPRTPNPLAR ANTNNANANVAYGNSNHSNSHVFSVHYEREFPEETETP ATRGRALGQPCRVLGPHSKPCVEMLKLLTQRAMPGRQA PPAPCQCPRVESH <b>MPEDRKSST</b> PGSLHEETPHSRSTREN TSRCAPATALLIQKALVRGGLGTLAADANFIMATGQAL ADACQMEPEEVEIMATELLKGREAPEGMASLGLCLNLGS SLGSLDQHQSQETLIPRL

3' Cav1.2	<p>LVRTALRIKTEGNLEQANEELRAIKKIWKRTSMKLLDQ  VPPAGDDEVTVGKIFYATFLIQEYFRKFKKRKEQGLVGG  PSQRNALSQAGLRTLHDIGPEIRRAISGDLTAEELDK  AMKEAVSAASEDDIFRRAGGLFGNHVSYQSDGRSAFPQ  TFTTQRPLHINKAGSSQGDTEPSHEKLV DSTFTPSSYS  STGSNANINNANTALGRLPRPAGYPSTVSTVEGHGPPPL  SPAIRVQEVAWKLSNRERHVPMCEDELELRDSSGSAGTQ  AHCLLLRKANPSRCHSRESQAAMAGQEETSQDETYEVKM  NHDTEACSEPSLLSTEMLSYQDDENRQLTLPEEDKRDIR  QSPKRGFLRSASLGRRASFHLECLKRQKDRGGDISQKTV  LPLHLVHHQALAVAGLSPLLQRSHSPASFPRPFATPPAT  PGSRGWPPQPVPTLRLEGVESSEKLNSSFPSIHCGSWAE  TTPGGGSSAARRVRPVS LMVPSQAGAPGRQFHGSASSL  VEAVLISEGLGQFAQDPKFIETTQELADACDMTIEEME  SAADNILSGGAPQSPNGALLPFVNCRDAGQDRAGGEEDA  GCVRARGRPSEEEELQDSRVYVSSL</p>
3' Cav1.3	<p>KLCPHRVACKRLVAMNMLNSDGTVMFNATL FALVRTAL  KIKTEGNLEQANEELRAVIKKIWKKTSMKLLDQVPPAG  DDEVTVGKIFYATFLIQDYFRKFKKRKEQGLVGGYPAKNT  TIALQAGLRTLHDIGPEIRRAISCDLQDDEPEETKREEE  DDVFKRNGALLGNHVNHNVDSDRRDSLQQTNTTHRPLHVQ  RPSIPPASDTEKPLFPPAGNSVCHNHNHNSIGKQVPTS  TNANLNNANMSKAAHGKRPSIGNLEHVS ENGHSHKHD  REPQRSSVKRTRYETIYIRSDSGDEQLPTICREDPEIH  GYFRDPHCLGEQEFYSSEECYEDDSSPTWSRQNYGYYSR  YPGRNIDSERPRGYHHPQGFLEDDSPVCYDSRRSPRRR  LLPPTPASHRRSSFNFECLELRQSSQEVPSSPIFPHRTA  LPLHLMQQQIMAVAGLDSSKAQKYS PSHSTRSWATPPAT  PPYRDWTPCYTPLIQVEQSEALDQVNGSLPSLHRSSWYT  DEPDISYRTFTPASLTVPSSFRNKNSDKQRSADSLVEAV  LISEGLGRYARDPKFVSATKHEIADACDLTIDEMESAAS  TLLNGNVRPRANGDVGPLSHRQDYELQDFGPGYSDEEPD  PGRDEEDLADEMICITTL</p>
3' Cav1.4	<p>MKQKLLDEVI PPPDEEEVTVGKIFYATFLIQDYFRKFRRR  KEKGLLGNDAA PSTSSALQAGLRS LQDLGPEMRQALTCD  TEEEEEEGQEGVEEEDKDL ETNKATMVSQPSARRGSGI  SVSLPVGDRLPDSL SFGPSDDDRGTPTSSQPSVPQAGSN  THRRGSGALIFTIPEEGNSQPKGTKGONKQDEDEEVPDR  LSYLDEQAGTPPCSVLLPPHRAQRYMDGHLVPRRLLPP  TPAGRKPSFTIQCLQRQGSCELDLIPGTYHRGRNSGPNR  AQGSWATPPQRGRLLYAPLLLVEEGAAGEGYLGRSSGPL  RTFTCLHVPGTHSDPSHGKRG SADS LVEAVLISEGLGLF  ARDPRFVALAKQEIADACRLTLEMDNAASDLLAQTSS  LYSDEESILSRFDEEDLGDEMACVHAL</p>

### **5.3.2. Confirmation of C-LTCC expression in cell culture**

#### **5.3.2.1. Protein lysates of transfected C-LTCC in HEK293T cells were separated and identified via western blotting**

The protein encoded by the 3' transcripts of LTCC genes were tagged with a flag tag (DYKDDDDK). Expression vectors were transfected into the human HEK293T cells, and the relative molecular weight of the encoded proteins assessed by Western blotting (see methods section 5.2.16). Firstly, protease inhibitors from Sigma-Aldrich (Figure 5.7A; a well called 1S48H) and Roche (Figure 5.7A; a well called 1S48A) were compared for their ability to inhibit proteolysis in lysates. The potential cleavage product (below 37kDa) was observed in the presence of the protease inhibitor from Sigma-Aldrich indicative of incomplete protease inhibition (Figure 5.7A; a well called 1S48H) compared to protease inhibitor from Roche (Figure 5.7A; a well called 1S48A). Therefore, the protease inhibitor from Roche was used to lyse cells in future experiments (Figure 5.7A).

Among various culture durations, in HEK293T cells, cultured for 24 hours post-transfection showed the highest level of protein expression. C-Cav1.1 was not as highly expressed as the other C-LTCC transcripts; therefore, a higher quantity (10 $\mu$ g) of protein lysate was loaded, compared to 5 $\mu$ g of protein lysates of the C-Cav1.3 and empty vector were used, and 3 $\mu$ g of protein lysates of the C-Cav1.2 and C-Cav1.4 were used.

All C-LTCC, except C-Cav1.1, expressed proteins of the predicted sizes in HEK293T cells. The C-Cav1.2 translates the protein with the weight of 63.87 kDa (Figure 5.7b; a well named as C-Cav1.2 3 $\mu$ g), the C-Cav1.3 translates the protein with the weight of 72.59 kDa (Figure 5.7b; a

well named as C-Cav1.3 5ug), and the C-Cav1.4 would translate the protein with the weight of 46.38 kDa (Figure 5.7b; a well named as C-Cav1.4 3ug). All the possible alternative translation start sites were examined in Figure 5.8. The C-Cav1.1 band is predicted to be 104.7 kDa in size but it is smaller in the western blot (Figure 5.7b; a well named as C-Cav1.1 10ug). This may be due to an alternative translation start site (Figure 5.8A). There are many possible alternative translation start sites which would translate protein with sizes between 75 kDa and 100 kDa. All the possible proteins contain the same predicted NLS; therefore, it was decided to transfect the same construct to the human SH-SY5Y cells.

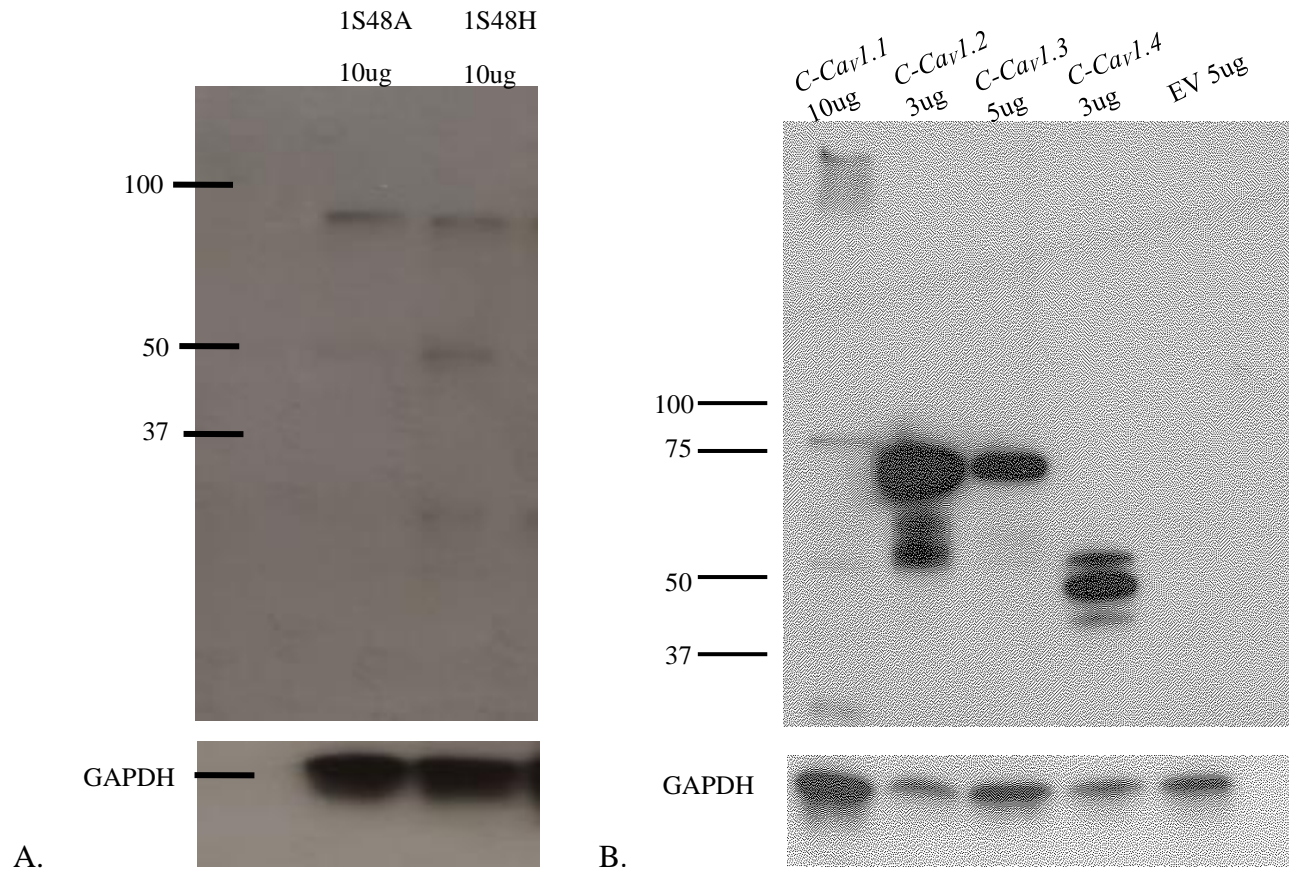
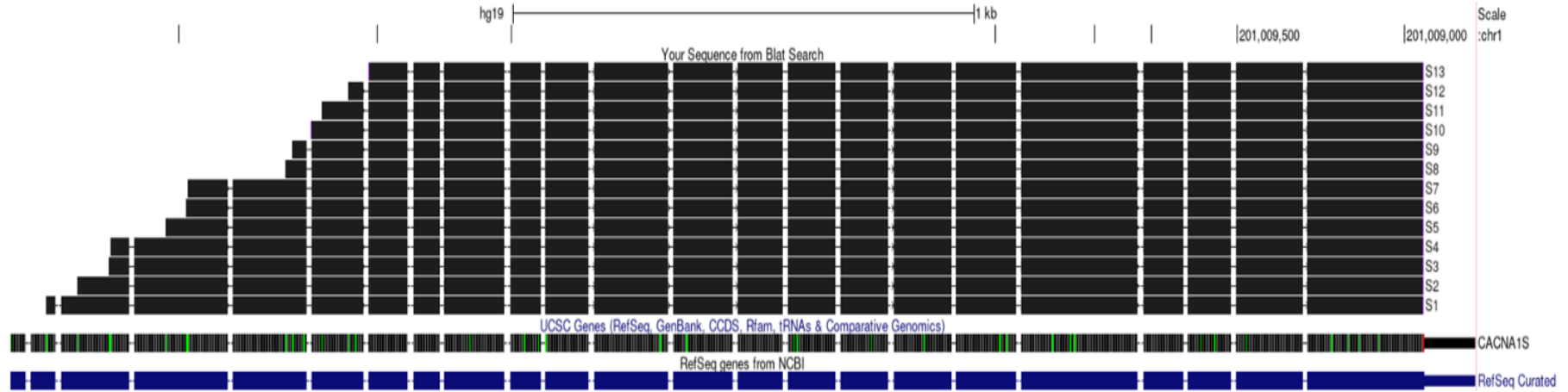


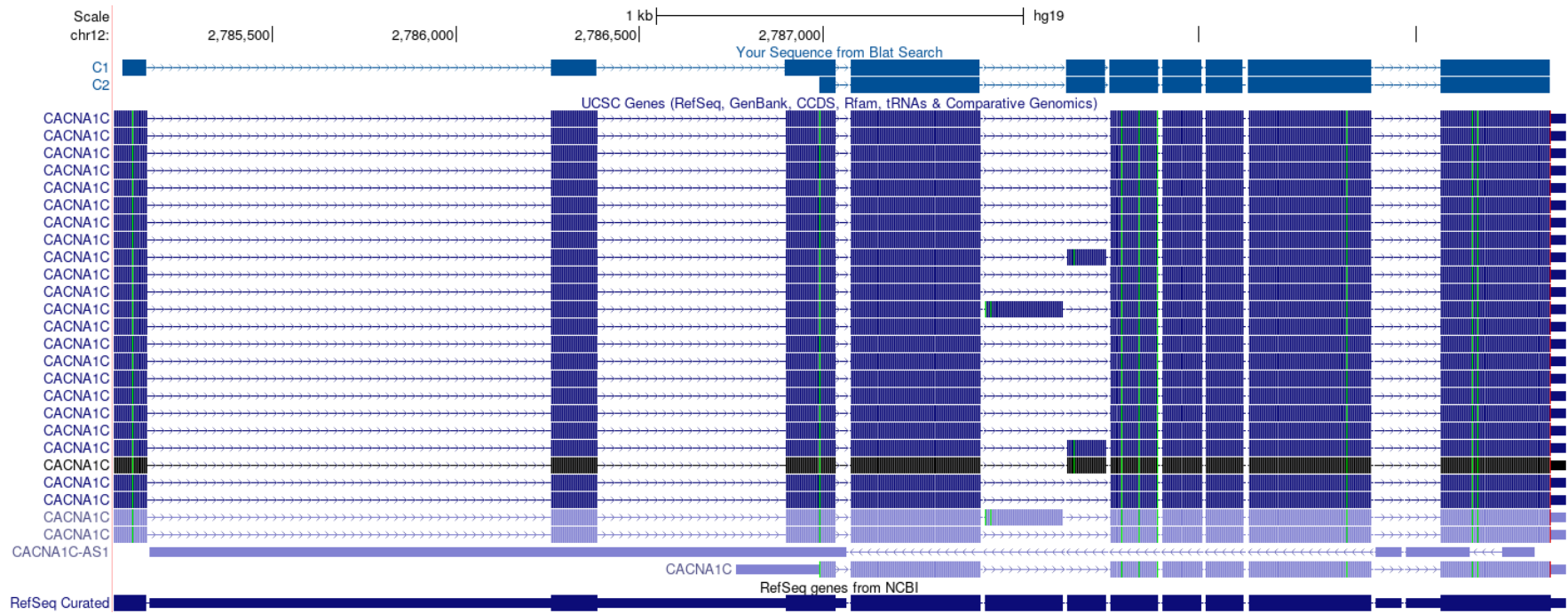
Figure 5.7. Western blot result of protein lysate from the HEK293T cells transfected with the 3' transcript of LTCC genes **A** Western blot result to test the best protease inhibitor. These western blots were developed with touch down exposure time. 1S48A is 10 $\mu$ g protein lysates from the HEK293T cells transfected with the C-Cav1.1 cultured for 48 hours and in the presence of protease inhibitor from Roche, whereas 1S48H is 10 $\mu$ g protein lysates from the HEK293T cells transfected with the C-Cav1.1 cultured for 48 hours and in the presence of protease inhibitor from Sigma-Aldrich. **B** Western blot of protein lysate from the

*HEK293T cells transfected with the C-LTCC cultured for 24 hours. The western blot was developed with touch-down exposure time. A well named as C-Cav1.1 10 $\mu$ g is 10 $\mu$ g of the protein lysate from the HEK293T cells transfected with the C-Cav1.1 (expected size 104.7kDa). A well named as C-Cav1.2 3 $\mu$ g is 3 $\mu$ g of the protein lysate from the HEK293T cells transfected with the C-Cav1.2 (expected size 63.87kDa). A well named as C-Cav1.3 5 $\mu$ g is 5 $\mu$ g of the protein lysate from the HEK293T cells transfected with the C-Cav1.3 (expected size 72.59 kDa). A well named as C-Cav1.4 3 $\mu$ g is 3 $\mu$ g of protein lysate from the HEK293T cells transfected with the C-Cav1.4 (expected size 46.38 kDa). A well named as EV24 5 $\mu$ g is 5 $\mu$ g of negative control (empty c-Flag pcDNA3 vector) from the HEK293T cells. GAPDH was used as a housekeeping protein to show whether there are proteins in the protein lysates harvested from each transfection.*

A

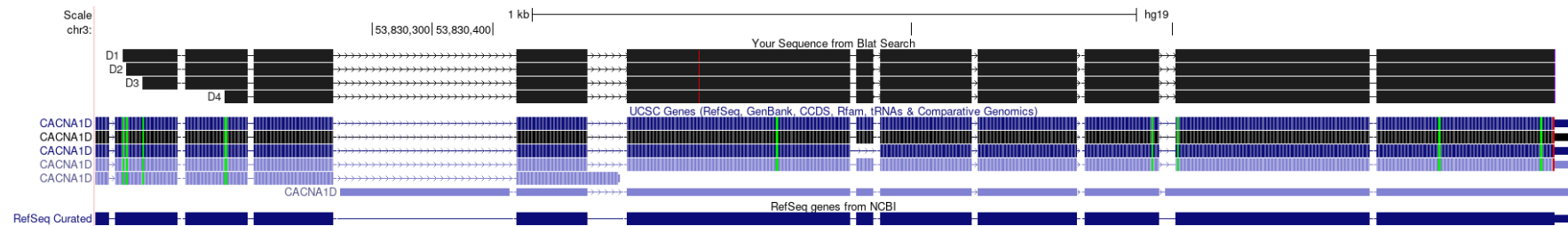


B





C



D

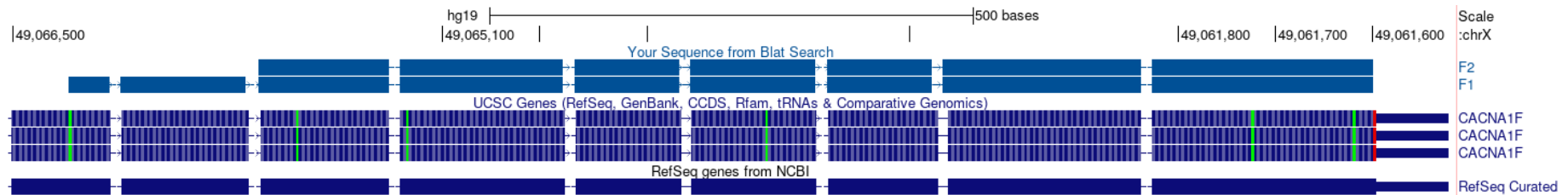


Figure 5.8. All the possible C-LTCC with different translation start sites **A**. All the possible C-Cav1.1 with different translation start sites. S1 is a protein which was originally aimed to be transfected to the HEK293T cells with protein size of 104.7kDa. S2 is a protein with size of 102.4kDa. S3 is a protein with size of 99.6kDa. S4 is a protein with size of 99.5kDa. S5 is a protein with size of 95.4kDa. S6 is a protein with size of 93.6kDa. S7 is a protein with size of 93.4kDa. S8 is a protein with size of 85.32kDa. S9 is a protein with size of 84.75kDa. S10 is a protein with size of 83.9kDa. S11 is a protein with size of 82.7kDa. S12 is a protein with size of 80.5kDa. S13 is a protein with size of 79.9kDa. **B**. All the possible C-Cav1.2 with different translation start sites. C1 is a protein which was originally aimed to be transfected to the HEK293T cells with protein size of 63.87kDa. C2 is a protein with size of 53.89kDa. **C**. All the possible C-Cav1.3 with different translation start sites. D1 is a protein which was originally aimed to be transfected to the HEK293T cells with protein size of 72.59kDa. D2 is a protein with size of 71.9kDa. D3 is a protein with size of 70.98kDa. D4 is a protein with size of 66.29kDa. **D**. All the possible C-Cav1.4 with different translation start sites. F1 is a protein which was originally aimed to be transfected to the HEK293T cells with protein size of 46.38kDa. F2 is a protein with size of 38.48kDa

### 5.3.2.2. Protein lysates of transfected C-LTCC in SH-SY5Y cells were separated and identified via western blotting

The C-LTCC were tagged with flag tag (DYKDDDDK) and transfected to the SH-SY5Y cells. The cells were cultured for 24 hours, and 48 hours. Western blotting (see methods section 5.2.16) was used to examine the transfection efficiency and measure the size of protein. As the SH-SY5Y cells have a lower transfection efficiency compared to the HEK293T cells, 10 $\mu$ g of protein lysates were loaded to the 7.5% gel (4561023, BioRad, USA). The size of each protein lysates should be the same as those of the HEK293T cells. However, in the SH-SY5Y cells, except the C-Cav1.2, the other C-LTCC might have too low transfection efficiency to be blotted visible. In SH-SY5Y cells, only the C-Cav1.2 showed evidence of expression (Figure 5.9).

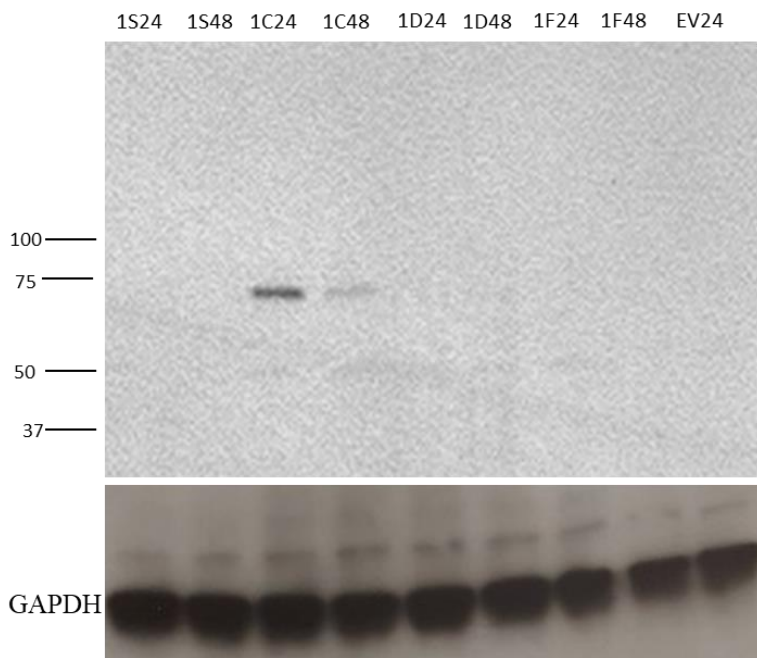


Figure 5.9. Western blot of the 3' LTCC in the SH-SY5Y cells. 1:5000 dilution of both primary antibody and the secondary antibody was used. 10 $\mu$ g of protein lysates were loaded to the 7.5% gel and developed on the film with touch-down exposure time. The wells were named with the gene name and time of being cultured. A well named as 1S24 is 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with the C-

*Cav1.1* cultured for 24 hours whereas 1S48 is 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with the C-*Cav1.1* cultured for 48 hours. The size of the protein should be 104.7kDa. A well named as 1C24 and 1C48 is 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with C-*Cav1.2* for 24 hours and 48 hours respectively, which should be 63.87kDa, 1D24 and 1D48 is 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with C-*Cav1.3* for 24 hours and 48 hours respectively, which should be 72.59kDa, and 1F24 and 1F48 are 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with C-*Cav1.4* for 24 hours and 48 hours respectively, which should be 46.38kDa. A well named with EV24 and EV48 are 10 $\mu$ g of protein lysates from the SH-SY5Y cells transfected with the empty c-Flag pcDNA3 for 24 hours and 48 hours respectively. The empty c-Flag pcDNA3 was transfected for negative control. GAPDH was used as a housekeeping protein to show whether there are proteins in the protein lysates harvested from each transfection.

### **5.3.3. Localization of C-LTCC**

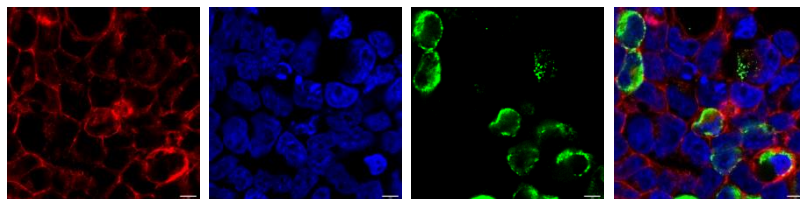
#### **5.3.3.1. Localization of C-LTCCs in HEK293T cells**

Plasmids containing C-terminal Flag-tagged 3' transcripts of LTCC genes were transfected into the HEK293T cells to examine the localization of the C-LTCCs using Zeiss confocal microscopy in collaboration with Edward Drydale and James Bancroft. Immunofluorescence labelling and confocal imaging shows that some, but not all, C-LTCC translocate to the nucleus. C-*Cav1.1* (Figure 5.10A), C-*Cav1.2* (Figure 5.10B), and C-*Cav1.4* (Figure 5.10D) were localised to the cytoplasm in HEK293T cells, whilst C-*Cav1.3* (Figure 5.10C) translocated to the nucleus. The empty expression vector, c-Flag pcDNA3, was examined as a negative control for FLAG immunolabeling (Figure 5.10E). The specificity of the secondary antibody was also checked by omitting the primary antibody (See Appendix 15). This showed that the secondary antibody was specific to the primary antibody. 3D confocal imaging of C-terminal flag tagged C-*Cav1.3* confirmed it localizes to the nucleus in the HEK293T cells (See Appendix 16). Finally, the relative

amount of FLAG-tagged protein inside of nuclei was counted and compared (Figure 5.12A). Only C-Cav1.3 located to the nucleus of HEK293T cells compared with other proteins (C-Cav1.1, C-Cav1.2, and C-Cav1.4)

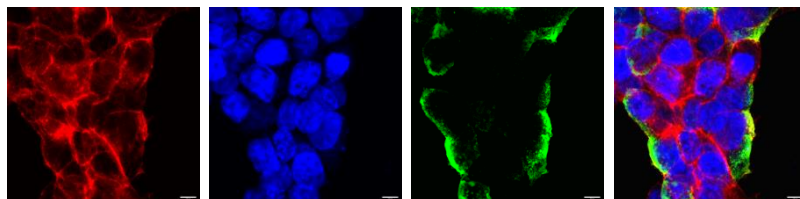
A. The HEK293T cells transfected with the C-Cav1.1

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged



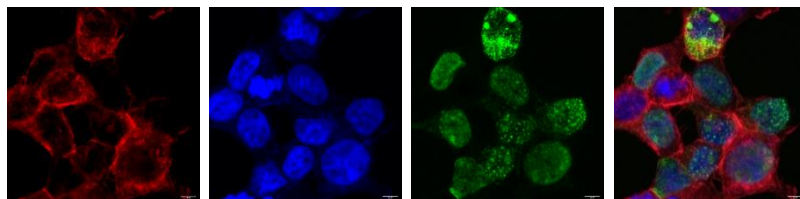
B. The HEK293T cells transfected with the C-Cav1.2

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged



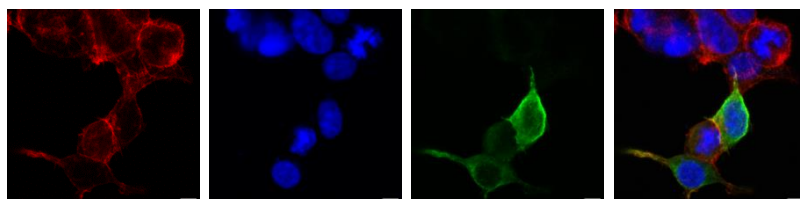
C. The HEK293T cells transfected with the C-Cav1.3

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged



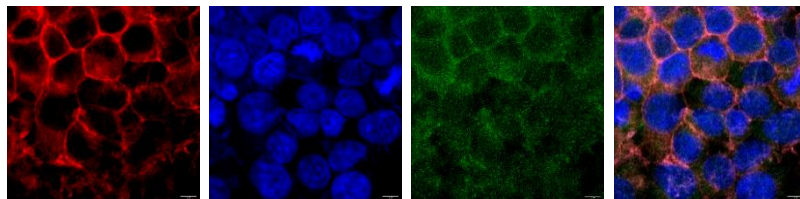
D. The HEK293T cells transfected with the C-Cav1.4

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged



E. The HEK293T cells transfected with the 3' transcript of empty c-Flag pcDNA3

Red: Phalloidin    Blue: DAPI    Green: non-specific    Merged



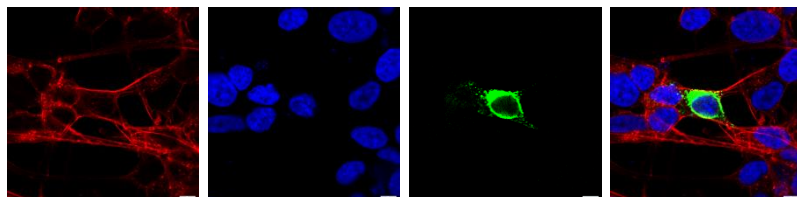
*Figure 5.10. Immunofluorescence confocal imaging of transfected HEK293T cells which were transfected with the 3' LTCC Cells were transfected for 24 hours and stained with Phalloidin and mounted with DAPI. The red channel is Phalloidin, the blue channel is DAPI, and the green channel is flag tag. The scale bar in the images is 100 $\mu$ M.*

### **5.3.3.2. Localization of C-LTCCs in SH-SY5Y cells**

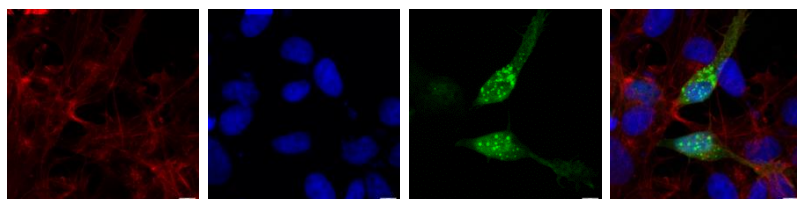
Plasmids containing the flag-tagged 3' transcript of LTCC genes were transfected into SH-SY5Y cells to examine the localization of the C-LTCCs in an excitable human cell type, by immunofluorescence and Zeiss confocal microscopy. Notably, not all C-LTCC translocated to the nucleus but stayed in the cytoplasm of SH-SY5Y cells. C-Cav1.1 (Figure 5.11A) and C-Cav1.4 (Figure 5.11D) localize to the cytoplasm, whilst C-Cav1.2 (Figure 5.11B) and C-Cav1.3 (Figure 5.11C) localised to the nucleus. Cells transfected with the empty expression vector, c-flag pcDNA3, served as a negative control for FLAG immunolabeling in SH-SY5Y cells (Figure 5.11E) as the vector does not express a FLAG protein in the absence of a coding insert. Lastly, the specificity of the secondary antibody was checked by omitting the primary antibody (See appendix 15). 3D confocal imaging of C-Cav1.2 and C-Cav1.3 confirmed nuclear localisation (See Appendix 16). Finally, the number of flag tags inside of nuclei of cells were counted and compared (Figure 5.12B), showing that only C-Cav1.2 and C-Cav1.3 localize to the nucleus compared with other proteins (C-Cav1.1 and C-Cav1.4).

A. The SH-SY5Y cells transfected with the C-Cav<sub>v</sub>1.1

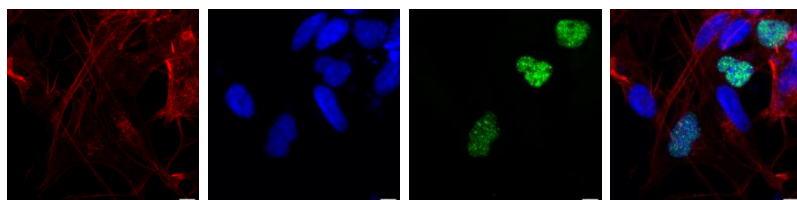
Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged

B. The SH-SY5Y cells transfected with the C-Cav<sub>v</sub>1.2

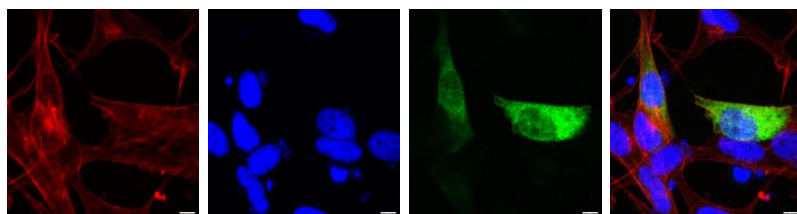
Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged

C. The SH-SY5Y cells transfected with the C-Cav<sub>v</sub>1.3

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged

D. The SH-SY5Y cells transfected with the C-Cav<sub>v</sub>1.4

Red: Phalloidin    Blue: DAPI    Green: Flag tag    Merged



## E. The SH-SY5Y cells transfected with the 3' transcript of empty c-Flag pcDNA3

Red: Phalloidin    Blue: DAPI    Green: non-specific    Merged

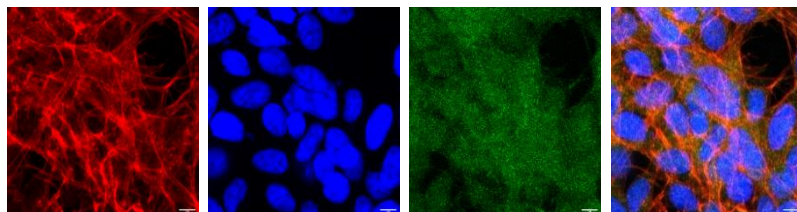


Figure 5.11. Immunofluorescence imaging of transfected SHSY5Y cells. These are immunofluorescence confocal imaging of SHSY5Y cells transfected with the C-LTCC. Cells were transfected for 24 hours and stained with Phalloidin and mounted with DAPI. The red channel is Phalloidin, the blue channel is DAPI, and the green channel is flag tag. The scale bar in the images is 100 $\mu$ M.

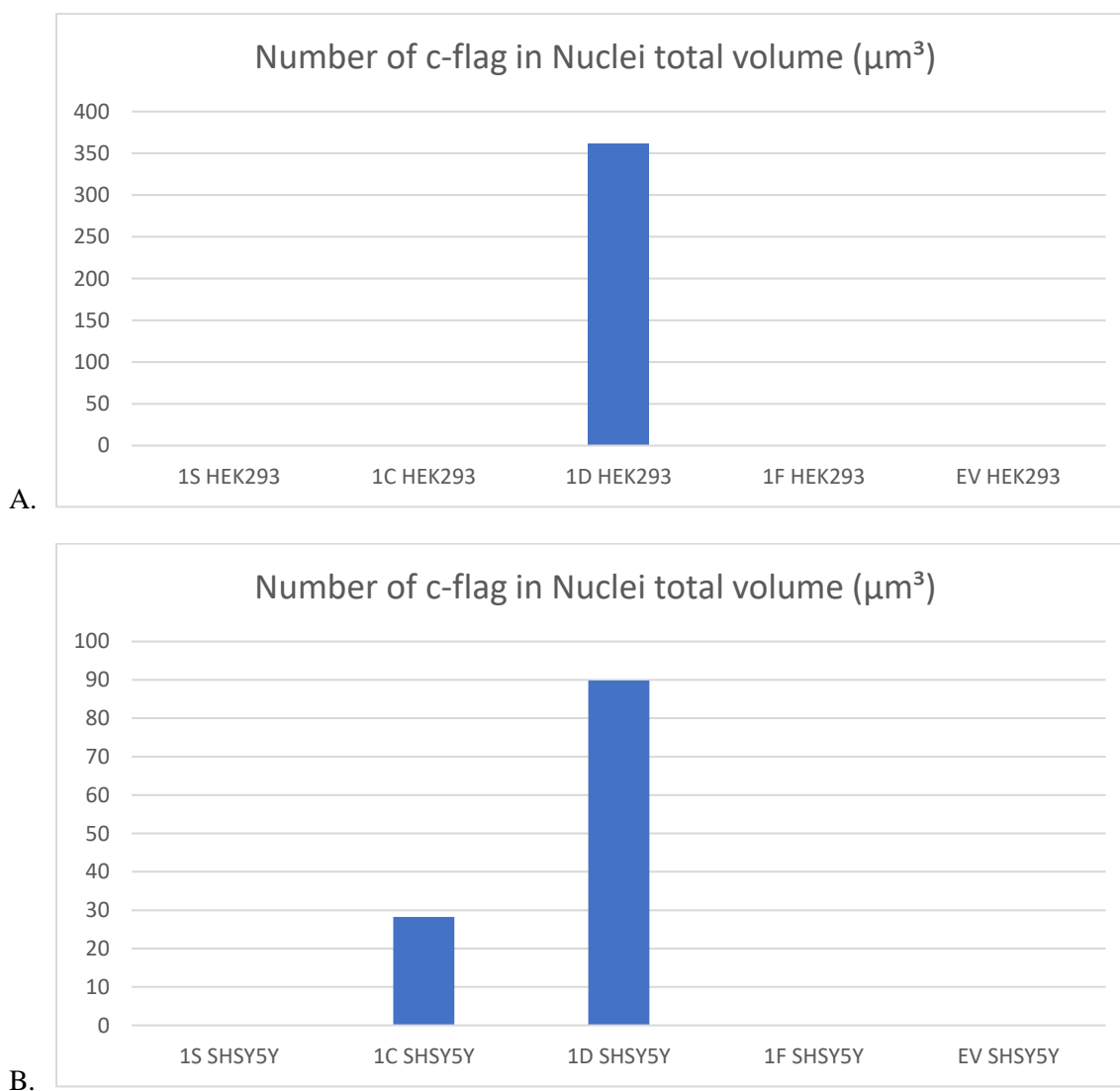


Figure 5.12. A graph comparing number of flag tags in nuclei. **A** This graph shows that the C-Cav1.3 (1D HEK293) located to the nucleus in HEK293T cells whereas other C-Cav1.1 (1S HEK293), C-Cav1.2 (1C HEK293), and C-Cav1.4 (1F HEK293) did not locate to the nucleus in HEK293T cells. EV HEK293 is an empty c-flag pcDNA3 transfected to HEK293T cells which is a negative control. **B** This graph shows that



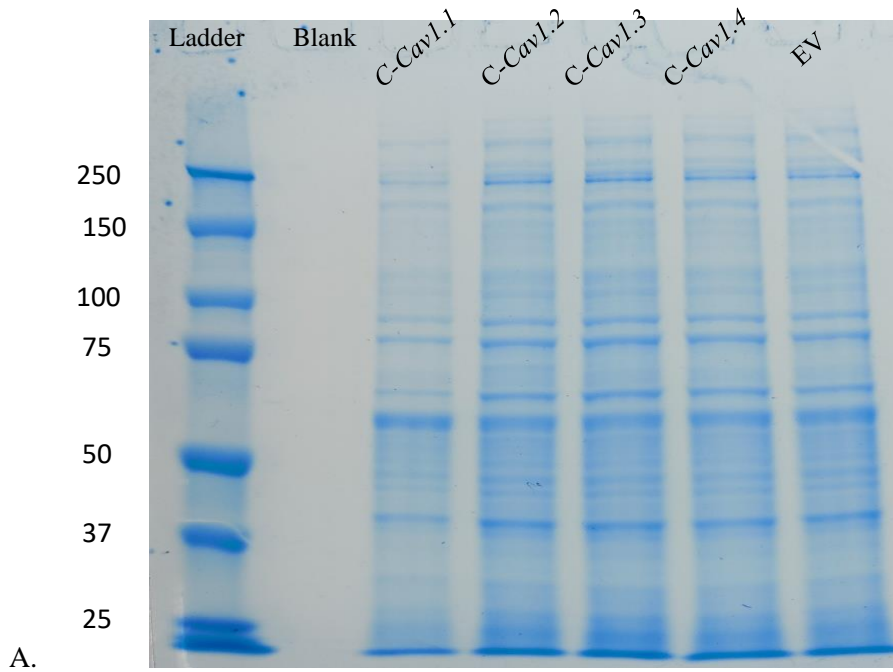
*the C-Cav1.2 (1C SHSY5Y) and C-Cav1.3 (1D SHSY5Y) located to the nucleus in SH-SHY5Y cells whereas other C-Cav1.1 (1S SHSY5Y), C-Cav1.4 (1F SHSY5Y) did not locate to the nucleus in SH-SHY5Y cells. EV SHSY5Y is an empty c-flag pcDNA3 transfected to SH-SY5Y cells which is a negative control.*

#### **5.3.4. Co-Immunoprecipitation of proteins encoded by the 3' transcripts of LTCC genes**

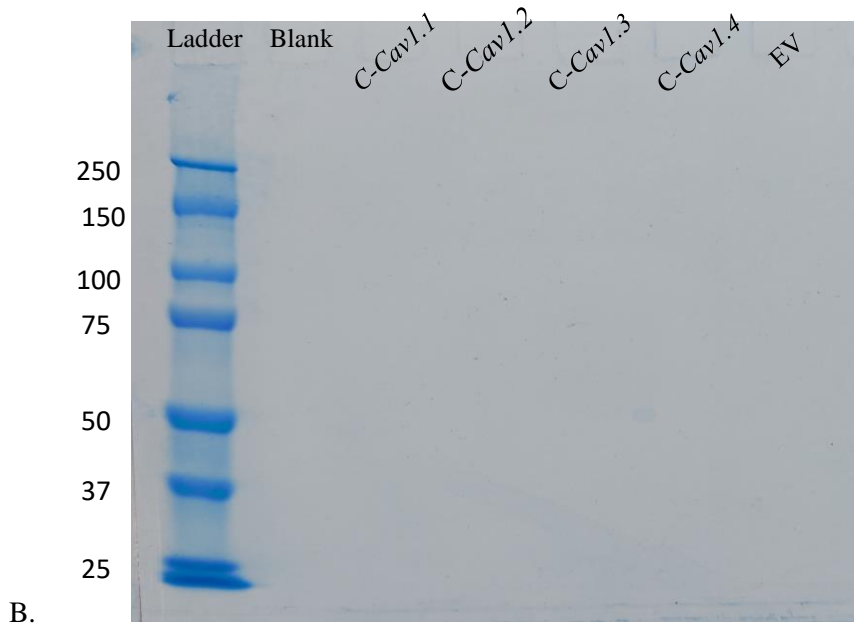
##### **5.3.4.1. Co-Immunoprecipitation of C-LTCCs in HEK293T cells**

In order to understand protein-protein interaction with C-LTCCs, co-immunoprecipitation was attempted with C-LTCC in HEK293T cells. The co-immunoprecipitation did not successfully identify proteins interacting with the C-LTCC in HEK293T cells (Figure 5.13B). It is assumed that anti-flag magnetic beads with glycine might be not strong enough to elute all proteins bound to the bead. As both the IPs of the target protein and co-IPs were not detected from the Coomassie stain, the mass spectrometry was not conducted. The Coomassie stain is less sensitive than western blotting therefore, western blotting assay was used to check whether the C-LTCC in HEK293T cells were correctly immunoprecipitated (Figure 5.14). The assay showed that except the C-Cav1.1 in HEK293T cells, other C-LTCCs were correctly immunoprecipitated (Figure 5.14).

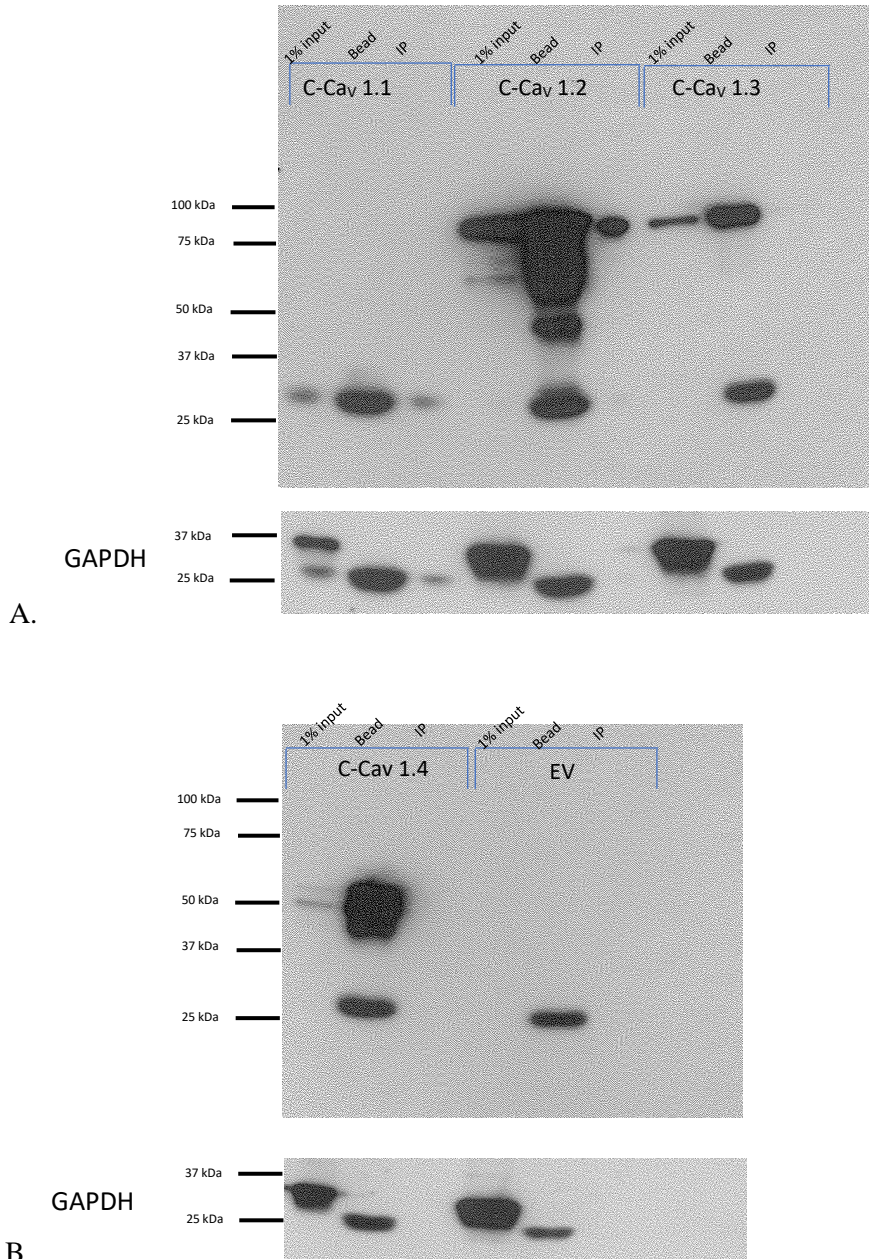
## 1% input to Co-Immunoprecipitation - Transfected to HEK293T



## Co-Immunoprecipitation - Transfected to HEK293T cells



*Figure 5.13 Coomassie blue staining of co-immunoprecipitation with C-LTCC in HEK293T cells. The well named as C-Cav1.1 is co-immunoprecipitation with the C-Cav1.1 in HEK293T cells, C-Cav1.2 is co-immunoprecipitation with the C-Cav1.2 in HEK293T cells, C-Cav1.3 is co-immunoprecipitation with the C-Cav1.3 in HEK293T cells, and C-Cav1.4 is co-immunoprecipitation with the C-Cav1.4 in HEK293T cells. **A** is Coomassie blue staining showing 1% input of the C-LTCC in HEK293T cells to co-immunoprecipitation. **B** is Coomassie blue staining showing the result of co-immunoprecipitation of the C-LTCC in HEK293T cells.*



*Figure 5.14. Western blotting of immunoprecipitation of the C-LTCC in HEK293T cells. A. Western blotting assay of immunoprecipitation of C-Cav1.1 (predicted to be 104.7 kDa), C-Cav1.2 (predicted to be 63.87 kDa), and C-Cav1.3 (predicted to be 72.59 kDa) in HEK293T cells show that the C-Cav1.1 was not immunoprecipitated while other C- LTCCs were immunoprecipitated correctly. GAPDH was used as protein loading control to show that equal amounts of protein were present in the input samples. Therefore, it should be absent from the IP. B. Western blotting assay of immunoprecipitation of the C-Cav1.4*

(46.38 kDa) and EV (empty c-flag pcdna3 vector) show that the C-Cav1.4 was immunoprecipitated correctly as FLAG antibody successfully detected the protein. GAPDH was used as a housekeeping protein to examine whether proper amount of proteins were immunoprecipitated.

#### 5.4. Discussion

In this chapter, it was hypothesized that the proteins predicted to be encoded by the 3' transcripts of LTCC genes will translocate to the nucleus in the human cells based on previous studies (Gomez-Ospina et al., 2013; L. Lu et al., 2015). Initial support for this hypothesis came from the results of the nucleus localization signal prediction (see 5.3.1 in Results). Then, to investigate their intracellular localization empirically, flag-tagged C-LTCCs were transfected to HEK293T cells (as an example of a non-excitabile cell type), and into SH-SY5Y cells (an excitable cell type). These experiments showed that only C-Cav1.3 localized to the nucleus in HEK293T cells (Figure 5.10D), whereas both C-Cav1.2 (Figure 5.11C) and C-Cav1.3 (Figure 5.11D) localized to nucleus in SH-SY5Y cells. However, C-Cav1.1 (Figure 5.10A; Figure 5.11A) and C-Cav1.4 (Figure 5.10B; Figure 5.11B) were not localized to nucleus in both cell types.

The localization results of C-Cav1.2 and C-Cav1.3 corroborate with the previous studies; CCAT from Cav1.2 which localizes to the nucleus of neurons and the C-terminal region of Cav1.3 (Gomez-Ospina et al., 2013; L. Lu et al., 2015). Despite several different mechanisms suggested to produce C-terminal fragment of VGCCs such as alternative splicing (Hulme et al., 2006; Scharinger et al., 2015), alternative promoter (Gomez-Ospina et al., 2013) and proteolytic cleavage (Kordasiewicz et al., 2006), it was suggested that the C-terminal fragments localize to the nucleus in neurons or cardiac cells and function as transcription factor (Gomez-Ospina et al., 2013; Kordasiewicz et al., 2006; L. Lu et al., 2015; Scharinger et al., 2015). The C-terminal fragment (CTF) that translocate to the nucleus can also be examined from the voltage-gated sodium channels.

The proteolytic cleavage by BACE1 and  $\gamma$ -secretase produce the intracellular domain (ICD) of Nav $\beta$  subunits, which regulate Nav  $\alpha$  subunits expression (D. Y. Kim et al., 2007; Wong et al., 2005). Especially, Nav $\beta$ 1-ICD enhances expression of *SCN4A* and *SCN5A* but reduces the expression of *SCN3A*. The Nav $\beta$ 2-ICD is 13kDa in size, which is similar to that of CCAT (Gomez-Ospina et al., 2006, 2013). Nav  $\beta$ 2-ICD along with Nav1.1 is highly expressed in cortical lysate from the Alzheimer's disease patients (Corbett et al., 2013). Even though the mechanism of Nav  $\beta$ -ICDs are different from that of C- LTCC in this project, C-Cav1.2 and C-Cav1.3, localized to the nucleus in SH-SY5Y cells, would also function to regulate other genes and possibly associated with disorders.

Although having a high prediction score to contain nuclear localization score in C-Cav1.1 and C-Cav1.4, they did not translocate to the nucleus in both cell types implying that they would not function as transcription factors. It is possible that they might interact with other proteins including full length channels to exert their functional effects. Therefore, to identify protein interaction with the C-LTCCs in each cell type, co-immunoprecipitation was conducted. Due to the lack of transfection sensitivity of C-LTCCs in both HEK293T cells and SH-SY5Y cells, insufficient levels of target protein immunoprecipitation were detected. As only a short period of time was available for the work, the co-immunoprecipitation procedure could not be optimized. As immunoprecipitation was detected by Western blot but not by Coomassie stain (Figure 5.13; 5.14), insufficient levels of immunoprecipitation of target protein to detect co-immunoprecipitation of binding partners were obtained, suggesting that glycine might not be strong enough to elute interact proteins. Glycine was used only for Mass spectrometry experiment (which was not conducted for this chapter) because mass spectrometry is not compatible with detergent. Therefore, it would be great to optimize co-immunoprecipitation procedure to yield

protein-protein interaction to examine what proteins bind to the C-LTCCs and the function proteins in the cells.

There are further experiments that could have done if there was more time. Firstly, the protein lysate of the C-Cav1.1 transfected in HEK293T cell showed different size than the predicted size of C-Cav1.1. There would be many reasons that transfected C-Cav1.1 show smaller than expected size such as protein degradation, protein being cleaved or alternative protein isoforms. While all processes of each sample were conducted simultaneously, I speculated that the antibody detected the smaller protein isoform in HEK293T cell. This means that the transcript might have different translation start site. If more time was available, more constructs of the C-Cav1.1 with different translation start sites could be transfected to the cells and examine the protein expression, and localization of the constructs. It is common that one transcript uses different translation initiation sites to produce different protein structure with different functions, which in turn diverse the genome (Touriol et al., 2003). If the translation employs a 5' cap-dependent mechanism, a translation start codon with a 'good' sequence (Kozak, 1987) would be predicted to be recognized, and used for translation initiation. However, if translation is dependent on a non 5' cap-dependent mechanism, translation can proceed from internal ribosome entry sites, which are located in the 5' UTR. Internal ribosome entry sites would help to recognize weak initiation sites of the transcript (Touriol et al., 2003). The effect of having alternative TSSs to the translation efficiency of the transcript can be seen from Runt-related transcription factor 1 (*RUNX1*) gene. The gene produced two different transcripts using different TSSs. The mRNA transcript produced from the distal TSS employs cap-dependent translation mechanism whereas the transcript originated from the proximal TSS has internal ribosome entry sites so that translation of the transcript depends on the 5' cap-independent translation (Pozner et al., 2000; Xu et al., 2019).

Therefore, predicting Kozak sequence and internal ribosome entry sites sequence of the C-LTCC would show how many protein structures can be produced from the transcript and how they localize in the cell.

Secondly, once co-immunoprecipitation is conducted successfully, protein-protein interactions can be examined utilizing mass spectrometry. Then, further functional studies can be conducted by using reporter gene analysis, mice knockout cell lines, or human knockout cell lines (Bunz, 2002; Kramer et al., 2018). These experiments would be able to compare existence of the C-LTCC to non-existence of C-LTCC in mice or human by co-expressing C-LTCC with other proteins. Lastly, C-Cav1.2 and C-Cav1.3 localized to the nucleus in the SH-SY5Y cells, they might function as transcription factor (Gomez-Ospina et al., 2013; L. Lu et al., 2015). Therefore, chromatin immunoprecipitation sequencing (ChIP-seq) (Park, 2009) can be utilized to investigate which DNA interact with C-Cav1.2 and C-Cav1.3.



## Chapter 6 - General Discussion

### 6.1. Rationale of studying the 3' transcripts of LTCC genes

The full-length *CACNAIS* and the full-length *CACNAIF* are known to be expressed primarily in the human skeletal muscle and the human retina respectively, but are not considered to be expressed in the human brain (Catterall et al., 2005; McRory et al., 2004; Zamponi et al., 2015). However, as demonstrated in section 1.5 in chapter 1, there was evidence of possible expression of previously unannotated transcripts arising from the 3' region of *CACNAIS* and *CACNAIF* in the human brain based on the short-read sequencing data from LIBD (Figure 1.6 in chapter 1), and exon expression level data and exon junction expression level data from GTEx (Figure 1.7 in chapter 1). Along with the short-read RNA-seq data results, several other relevant findings suggested the existence of 3' isoforms of VGCCs either at mRNA levels or protein levels (section 1.3 and 1.4 in chapter 1). Among those findings, some studies examined expression regulation of *CACNAIC* using an alternative TSS (Dai et al., 2002; Gomez-Ospina et al., 2013; Pang et al., 2003; N. I. Saada et al., 2005), or creating a three domain channel of *TuCa1*, which protein is homologous to that of vertebrate LTCCs, originated from an alternative TSS (Okagaki et al., 2001). Additionally, the CCAT was shown to arise from the exonic promoter in the C-terminal region of  $Ca_v1.2$ , and reported to be localized to the nucleus and to function as a transcription factor (Gomez-Ospina et al., 2013). In particular, CCAT serves as a transcription factor of Connexin 31.1 gene (*Cx31.1*). It regulates the expression of *Cx31.1* by binding to the final 148 base pairs of the *Cx31.1* promoter sequence (Gomez-Ospina et al., 2013).

Collectively, the short-read RNA sequencing data from LIBD and GTEx, and previous studies support the possible existence of the short-length transcript from the 3' region of *CACNAIS* and *CACNAIF* that may be arising from a de novo TSS rather than the canonical TSS of the full-

length *CACNAIS* and *CACNAIF*. This thesis aimed to investigate and characterize the possible short-length transcripts from the 3' region of LTCC genes in the human brain, especially focusing on *CACNAIS* and *CACNAIF*.

## **6.2. The findings regarding the 3' transcripts of LTCC genes**

The main finding of this thesis is that LTCC genes would produce 3' transcripts from the de novo TSSs in the human brain. The 3' transcripts may express in a tissue-specific manner. The conservation status analysis of 5'UTR of the 3' transcripts of LTCC genes implies that they may have functionality. Even though the potential functionality of each 3' transcript can only be speculated from the localization studies, the exact functions of the 3' transcripts remain to be studied. From now on, I will discuss the findings of each chapter with future research that will help to expand the findings from this thesis.

### **6.2.1. The 3' transcripts of *CACNAIS* and *CACNAIF* were found in the human brain**

The long-read nanopore sequencing identified the existence of the short-length transcripts arising from the 3' region of *CACNAIS* and *CACNAIF* in the human brain. After analyzing the long-read nanopore sequencing data with TAQLoRE (see 2.3.13 of methods section in chapter 2), the in-frame protein coding short-length transcripts from the 3' regions of *CACNAIS* and *CACNAIF* with high expression (> 1% of the total expression of coding transcripts) in the human brain were identified. They were found to be comprised of known exons, which are already annotated for the full-length *CACNAIS* and *CACNAIF*, with the possibility of exon skipping (Figure 2.7, 2.13 in chapter 2). This shows that similar to *CACNAIC* (Clark et al., 2020), the short-length transcript from the 3' region of *CACNAIS* and *CACNAIF* may undergo alternative splicing and produce multiple transcript isoforms in the human brain. Further, the exon skipping may affect the protein that each transcript isoform encodes (Table 2.8; 2.10 in chapter 2).

However, the origins of the sequenced transcripts were unclear. It has been shown that many novel transcripts, identified via long-read nanopore sequencing, are produced from novel TSSs other than canonical TSSs (Wright et al., 2022). Also, as *CACNA1C* produces multiple transcripts using alternative TSSs (Dai et al., 2002; Gomez-Ospina et al., 2013; Pang et al., 2003; N. I. Saada et al., 2005), it was hypothesized that the short-length transcripts from the 3' region of *CACNA1S* and *CACNA1F* would also originate from alternative TSSs. The results of the experiments described in this thesis support the hypothesis as the short-length transcripts from the 3' region of *CACNA1S* and *CACNA1F* were found to be complete transcripts originating from de novo TSSs. Based on the genomic location of the TSSs of the 3' transcripts of *CACNA1S* and *CACNA1F*, the genomic structure of these transcripts was speculated. The 3' transcript of *CACNA1S* encodes the channel from the loops between the S5 and S6 of domain III until the cytoplasmic C-terminal region (Figure 3.7 in chapter 3), whereas the 3' transcript of *CACNA1F* encodes only within the cytoplasmic C-terminal region (Figure 3.8 in chapter 3). In addition to that, alternative TSSs were found in the 3' region of *CACNA1C*, and *CACNA1D* in the human brain (Figure 3.9 in chapter 3 for the 3' transcript of *CACNA1C*; Figure 3.10 in chapter 3 for the 3' transcript of *CACNA1D*). These show that *CACNA1C* and *CACNA1D* also have a novel transcript that encode the C-terminal region of Cav1.2 and Cav1.3, respectively. The TSSs are located within an exon of the full-length LTCC genes and the short-length transcripts from the identified TSSs share the same nucleotide sequences as the sequences of the full-length genes. The results showed that there is a possible transcript regulation by having alternative TSSs in the 3' region of LTCC genes.

### 6.2.2. Could the 5' UTR contain functional motifs?

Functional regions in the transcripts are known to be more conserved than other regions such as intronic regions (Nemoto & Toh, 2012). As the 3' transcripts were found to be complete transcripts, the potential functionality of the 5' UTR of the 3' transcripts was examined using the PhyloP score across 240 mammals from the Zoonomia project (Zoonomia Consortium, 2020). It was hypothesized that if the 5'UTR contains functional regions, it will be significantly more conserved than other regions, especially intronic regions. The 5'UTR of the 3' transcript of LTCC genes showed significantly higher PhyloP scores than their intronic regions (Figure 4.3 in chapter 4). This suggests that the 5'UTR of the 3' transcript of LTCC genes may have functional roles affecting transcription regulation in the human brain as the binding motifs such as transcription factor binding sites in the 5'UTR show a high level of conservation (Byeon et al., 2021). The novel motifs show a similar conservation level to functional transcription regulators in both humans and mice (The ENCODE Project Consortium et al., 2020). Therefore, the novel 5'UTR of LTCC 3' transcripts in the human brain could contain transcription binding motifs that have not been studied yet.

### 6.2.3. The expression of the 3' transcript of LTCC genes in human tissues

The 5'UTR of the 3' transcript of LTCC genes was found to be well-conserved among 240 mammals in chapter 4. It has been suggested that a highly conserved 5'UTR may regulate post-transcriptional tissue-specific expression (Byeon et al., 2021). Also, other genes produce short-length transcripts from alternative TSSs with tissue-specific expressions. For example, *NRG1* produces diverse transcripts from different TSSs (Steinthorsdottir et al., 2004a). Among the multiple transcripts, *NRG1* Type I and II were not expressed in lymphocytes and skeletal muscle, and Type III mRNA expresses in both human heart, lung and adult brain whereas Type IV mRNA

showed exclusive expression in the human brain, specifically in the fetal brain (W. Tan et al., 2007). Therefore, not only to examine tissue-specificity but also to assess the evidence of expression of the LTCC 3' transcripts, exon expression data from GTEx was explored. It was hypothesized that once transcription is initiated, the expression level of the exons would be significantly increased compared to that of the exon where TSS or each 3' transcript is located.

To examine whether the exon expression increases after the transcription of the 3' transcripts of LTCC genes is initiated, the expression level of the exon where the TSS of the 3' transcript is located was compared with that of adjacent exons (1 upstream and 1 downstream). The exon expression was found to be significantly increased after the transcription initiation of the 3' transcript of *CACNAIS* (Figure 4.4 in chapter 4) in both tissue types (non-brain tissues and brain tissues). In the human brain tissues, the expression level of the one downstream exon from the exon where TSS of the 3' transcript of *CACNAIF* is located is significantly increased compared to the expression level of the exon where TSS is located. However, in non-brain tissues, the expression level of one downstream exon was not significantly increased compared to that of one upstream exon and that of exon where TSS is located. These suggest that the 3' transcript of *CACNAIF* may be expressed in a brain-tissue-specific manner (Figure 4.5 in Chapter 4). This result supports previous study which detected exon 28-29, exon 45-46 and exon 47-48 in human pineal gland tissue using RT-PCR analysis (Hemara-Wahanui et al., 2005) even though previously *CACNAIF* was not amplified via RT-PCR in the human brain (McRory et al., 2004). Exon expression level of the 3' transcript of *CACNAIC* is significantly increased after the transcription initiation in the brain tissue rather than non-brain tissues (Figure 4.6 in chapter 4). Interestingly, exon 47 where the TSS of the 3' transcript of *CACNAIC* is located, has significantly lower expression level compared to the adjacent exons in both tissue types, implying that the exon might

be alternatively spliced. Finally, the exon expression level of *CACNAID* did not show significant changes after the transcription initiation in both tissue types (Figure 4.7 in chapter 4).

Fold change analyses of expression level changes also shows that the expression level of one downstream exon increases compared to that of exon where TSS of each 3' transcript is located. Examining the expression level confirmed the possible existence of the 3' transcripts of LTCC genes in the human brain. These analyses support the evidence that the expression of the 3' transcripts arise from the novel TSS found in the thesis.

Moreover, LTCC genes may produce the 3' transcripts not only in tissue-specific manner but also in brain-region specific or cell-type specific manner. *SCN1A* and *SCN2A*, genes encoding voltage-gated sodium channels, produce multiple transcripts from alternative TSSs with brain-region specific expression (Martin et al., 2007). Furthermore, most genes show various preferred of TSS usage among cell types (Xu et al., 2019). For example, BDNF transcripts show not only tissue-specific but also cell-specific expression pattern when the expressions of each transcript were compared between hippocampus and Leukocyte (Cattaneo et al., 2016). Therefore, it would be interesting to examine brain-region specificity or cell-type specificity of the 3' transcript expression.

#### **6.2.4. Where do the proteins encoded by the 3' transcripts of LTCC genes translocate in the human cell?**

In line with a previous study that demonstrated CCAT localization to the nucleus in neurons to function as a transcription factor (Gomez-Ospina et al., 2013), the translated sequences of 3' transcripts of LTCC genes (C-LTCC) were predicted to contain a strong nuclear localization signal (Table 5.6 in chapter 5). Therefore, C-LTCCs were expected to localize to the nucleus in the cell. The translation and localization of the C-LTCCs were examined by transfecting them into

HEK293T cells (non-excitabile cells) and SH-SY5Y cells (excitable cells), respectively. Even though proteins were detected by western blot, C-Cav1.1 showed smaller than the predicted size (Figure 5.7bB; 5.8A in chapter 5) however, the detected protein was still expected to contain the strong nuclear localization signal therefore, the localization of the protein was examined in both cell types.

The C-Cav1.1 (Figure 5.10 A in chapter 5 for HEK293T cells; Figure 5.11A in chapter 5 for SH-SY5Y cells) and C-Cav1.4 (Figure 5.10D in chapter 5 for HEK293T cells; Figure 5.11D in chapter 5 for SH-SY5Y cells) did not localize to the nucleus in both cell types whereas C-Cav1.3 did localize to the nucleus in both cell types (Figure 5.10C for HEK293T cells; 5.11C for SH-SY5Y cells in chapter 5). C-Cav1.2 did not localize to the nucleus in HEK293T cells (Figure 5.10B in chapter 5) but localized to the nucleus in SH-SY5Y cells (Figure 5.11B in chapter 5). C-Cav1.2 might function as a transcription factor only in excitable cells which also supports the previous finding (Gomez-Ospina et al., 2013). These results did not support the hypothesis that the C-LTCCs would localize to the nucleus in the cells. However, the C-Cav1.3 localized to the nucleus which means that it might function as a transcription factor in both non-excitabile and excitable cells, which corroborates the previous finding (L. Lu et al., 2015). To examine protein-protein binding with the C-LTCCs, co-immunoprecipitation was attempted but was not successful.

#### **6.2.4.1. Possible functions**

In this thesis, the functions of LTCC 3' transcripts were not elaborated. However, based on previous studies, the functions can be speculated. Firstly, LTCC 3' transcripts are postulated to repress a target gene, especially the gene that encodes the full-length  $Ca^{2+}$  channel in the cell. Human *LEF1* that encodes lymphoid enhancer binding factor 1, produces two different transcripts from alternative TSS. The longer isoform interacts with  $\beta$ -catenin and Wnt target genes to control

Wnt signals however, the shorter isoform directly suppresses the target genes to regulate Wnt signals (Arce et al., 2006; Xu et al., 2019). In relation to LTCCs, *TuCa1* produces the three-domain channel from alternative TSS that does not encode functional calcium current but, it inhibits the expression of the full-length channel in the ascidian muscle blastomeres (Okagaki et al., 2001). Other researchers created a similar three-domain channel of Cav1.2 by introducing the start codon at the homologous region. The three-domain channel of Cav1.2 also repressed the expression of the full-length channel in BHK-6 cells (Ebihara et al., 2002). Additionally, the inhibition of the full-length channel may be associated with brain development because the expression of CCAT reduces as the brain develops and suppresses Cav1.2 expression during brain development (Gomez-Ospina et al., 2013).

Lastly, 3' transcripts of LTCC may regulate the development of neurons. Interneurons in the embryonic brain migrate to the target region once they were produced so that they can be active in cortical network (Warm et al., 2022). The multiple *NRG1* isoforms were studied to affect the migration of gamma-aminobutyric acid (GABA)ergic interneurons (Flames et al., 2004). The multiple isoforms of *NRG1*, which are produced from alternative promoters, were introduced earlier in this thesis. Type I, II, IV and V contain an immunoglobulin domain, EGF domain, transmembrane domain and C-terminal region. However, type III isoform has a different structure. Both N and C termini of type III are intracellular with additional transmembrane domain and cysteine rich domain but without immunoglobulin domain (Harrison & Law, 2006). *NRG1* isoforms with immunoglobulin directs the GABAergic interneurons to the cortex whereas the type III *NRG1* functions as a channel for the neurons from the lateral ganglionic eminence (Flames et al., 2004). Moreover, type III isoform provides an axonal signal that has a crucial role in Schwann cell development and myelination (Newbern & Birchmeier, 2010). Particularly, while type I and



IV translocate to neuronal soma and dendrites but not to axons, type III localizes to axons too, implying that the N-terminal sequences of type III may interact with proteins to transport the type III isoform to axons (Shamir & Buonanno, 2010).

Moreover, CCAT was the most enriched in GABAergic inhibitory neurons and had the highest expression in thalamus, inferior olive, and inferior colliculus, which implies that the CCAT would play a significant role in neuron development in cell-type specific manner (Gomez-Ospina et al., 2006). Like the distinctive function of Type III isoform from other types of isoforms or CCAT, the LTCC 3' transcripts can also have unique effect on neurons or cell development.

### **6.3. Future studies regarding to the findings in the thesis**

#### **6.3.1. Will the exon-skipping events in the 3' transcript of *CACNA1S* and *CACNA1F* have an influence on their function of translated protein in the human brain?**

The 3' transcripts are complete transcripts arising from the novel TSSs that would have their own functions. In chapter 2, exon skipping events were identified for each 3' transcript. Therefore, whether the exon skipping has an influence on translated protein of each 3' transcript will be useful to understand the 3' transcripts of *CACNA1S* and *CACNA1F* in the human brain.

The result of long-read nanopore sequencing of the 3' transcript *CACNA1S* is interesting as it demonstrates a possible exon 6 skipping event. Two highly expressed (>1% expression level of the total expression of coding transcripts) short-length transcripts from the 3' region of *CACNA1S* were identified. Among these two transcripts, *CACNA1S\_novel 596* had exon 6 skipped, which is equivalent to exon 29 of the full-length *CACNA1S*. Exon 29 of the full-length *CACNA1S* was found to be alternatively spliced in human muscle cells (Tuluc et al., 2009). Exon 29 is located on the extracellular loop between transmembrane segments 3 and 4 of the fourth domain (Tuluc et al., 2009). The exclusion and inclusion of exon 29 depends on the developmental

stage. The embryonic Cav1.1 channel variant does not contain exon 29 but the adult Cav1.1 channel variant does contain exon 29 (Sultana et al., 2016; Tuluc et al., 2016). Exon 29 is included in the predominant transcript variant in the human and mouse muscle cells (Tuluc et al., 2009). The CACNA1S transcript variant without exon 29 encodes functional channels in the human adult muscle and cultured muscle cells that show significantly increased voltage sensitivity and a longer open probability, which in turn causes eight-fold increased whole-cell calcium currents (Flucher, 2020; Sultana et al., 2016; Tuluc et al., 2009).

The 3' transcript of *CACNA1S* studied in this thesis would not encode a functional calcium channel; however, it does encode one domain with an  $\alpha 2\delta$  binding site and the C-terminal region. Similar to the three-domain channel of *TuCa1*, which originates from an alternative TSS without encoding functional channels but reduces the full-length channel expression in *Xenopus Oocyte* (Okagaki et al., 2001), the protein encoded by the short-length transcript from the 3' region of *CACNA1S* may have a similar function in the cell. Moreover, exon 6 skipping of the short-length transcripts from the 3' region of *CACNA1S* may affect its function. The short-length transcript from the 3' region of *CACNA1S* without exon skipping was predicted to encode a protein called CAC1S\_HUMAN whereas the transcript with exon 6 skipping is predicted to encode a protein called B1ALM3\_Human (Table 2.8 in chapter 2). This implies that exon 6 skipping could have influence on the function of the encoded protein.

Moreover, long-read nanopore sequencing also identified multiple transcript isoforms of the 3' transcript of *CACNA1F* in the human brain. Four highly expressed (>1% expression level of the total expression) short-length transcripts were identified. Among these transcripts, CACNA1F\_novel-57 had exon 8 skipped, which is equivalent to exon 47 of the full-length *CACNA1F*. In the retina, exon 47 of *CACNA1F* plays a crucial role in  $\text{Ca}^{2+}$ -dependent inactivation

(CDI) (Haeseleer et al., 2016; The UniProt Consortium, 2021). A splice variant of Cav1.4 without exon 47 removes a part of a C-terminal automodulatory domain (CTM) which is highly expressed in primate retina but not in mouse retina. Lacking exon 47 splice variant of Cav1.4 causes intense CDI, meaning that the channels should support the 3-fold higher  $\text{Ca}^{2+}$  influx into the channel compared with the full-length channel. However, CaBP4 binds to the C-terminal region of CACNA1F transcript without exon 47 and nullifies the intense CDI without affecting other activation mechanisms such as  $\text{Ca}^{2+}$ -bound calmodulin binding. Therefore, even though the 3' transcript of *CACNA1F* does not encode functional channels but just the C-terminal region of the full-length Cav1.4, the protein encoded by the transcript may have a distinct function in the cell because CCAT also encodes the C-terminal region of Cav1.2 and functions as transcription factor (Gomez-Ospina et al., 2013). Moreover, Exon 8 skipped transcript variant was predicted to encode part of protein isoform 6 of Cav1.4 whereas other highly expressed transcripts may encode part of different protein isoforms (Table 2.10 in chapter 2). This implies that the exon 8 lacking may influence the function of protein encoded by the splice variant.

Furthermore, these reported exon-skipping events in the full-length transcripts of *CACNAIS* and *CACNAIF* in muscle and retina tissues respectively were done primarily by using short-read sequencing or by amplifying those specific regions with PCR amplification. In such cases, it is impossible to distinguish between multiple different transcripts. For example, whether the exon skipping event is happening from a short-length transcript sharing the same genomic region as the full-length transcript or from the full-length transcript (Wright et al., 2022). By utilizing the long-read sequencing, multiple transcript variants of *CACNAIS* and *CACNAIF* can be identified. Therefore, studying whether previously investigated exon skipping events are predominately identified in the full-length *CACNAIS* and *CACNAIF* transcript originating from

the canonical TSS, or other transcript variants originating from alternative TSSs, would be interesting to investigate further.

### **6.3.2. What are the non-coding short-length transcripts identified from the long-read nanopore sequencing?**

From the long-read nanopore sequencing data of the short-length transcript from the 3' region of *CACNAIS* and *CACNAIF*, a large amount of the non-coding transcripts was identified (Figure 2.5 and 2.11 in chapter 2). There are several groups of non-coding RNAs (ncRNAs) which are infrastructural ncRNAs (ribosomal RNAs, transfer RNAs, small nuclear RNAs), regulatory ncRNAs (microRNAs, Piwi-interacting RNAs, small interfering RNAs, and long non-coding RNAs), and novel class ncRNAs (promoter-associated RNAs and enhancer RNAs) (Kaikkonen et al., 2011). Among these RNAs, the highly expressed ncRNAs identified in chapter 2 assumed to be long non-coding RNAs or enhancer RNAs due to their length (Kaikkonen et al., 2011).

The functions of the long non-coding RNAs (lncRNAs) has been widely studied recently (Statello et al., 2021). lncRNAs are transcripts that are longer than 200 nucleotides without being translated into functional proteins (Statello et al., 2021). Although one gene produces multiple splice variants, one transcript variant of the gene has the most expression (Gonzalez-Porta et al., 2013), which can be a non-coding transcript (Dhamija & Menon, 2018). lncRNAs have numerous functions such as regulating the expression of neighboring genes or distance genes. They affect the transcription of the genes or other structural and regulatory mechanisms of mRNA including splicing, translation, and signaling pathways. These effects in turn ultimately influence several cellular functions or the development of diseases. Therefore, the expression of lncRNAs could be used as disease biomarkers or therapeutic targets (Dhamija & Menon, 2018; Statello et al., 2021). Furthermore, the non-coding 3' transcripts of *CACNAIS* and *CACNAIF* could be enhancer RNAs

(eRNAs), which are 100bp to 9000bp in length. They are originated from enhancer regions where overlaps with H3K4me1 or PolIII (Kaikkonen et al., 2011). DNA sequences encoding eRNAs are evolutionarily conserved and studied to function as transcriptional activators (Kaikkonen et al., 2011). Thus, the un-analyzed non-coding transcripts from the long-read nanopore sequencing could have an important function distinct from the in-frame coding 3' transcripts of *CACNAIS* and *CACNAIF* in the biological regulation of mRNA transcription and translation.

Moreover, the non-coding transcripts of 3' transcripts of *CACNAIS* and *CACNAIF* can contain premature termination codon and undergo nonsense-mediated decay. Nonsense-mediated decay prevents the mRNAs with premature termination codon from being translated which in turn, may produce C-terminus truncated proteins with abnormal function (Nickless et al., 2017). Moreover, premature termination codon can be formed via alternative splicing and/or the use of alternative TSS. Alternative splicing can produce the premature termination codon that transcripts can undergo nonsense-mediated decay (Nickless et al., 2017). For example, exon 11 of polypyrimidine tract binding protein 1 (*PTBPI*), that regulates alternative splicing, mRNA stability, localization or IRES dependent translation, can be skipped and the alternatively spliced transcript experience nonsense-mediated decay (Wollerton et al., 2004). Moreover, having alternative TSSs can also produce transcripts with premature stop codon which results in nonsense-mediated decay (Arribere & Gilbert, 2013; Malabat et al., 2015) which may not be translated. It is not clear what the highly expressed non-coding 3' transcripts are, however, the functions of ncRNAs have been widely studied recently. Therefore, the examination of the non-coding 3' transcripts will be needed in the future.

### 6.3.3. The transcription mechanism of the LTCC 3' transcripts

The experiments in this thesis primarily aimed to characterize the 3' transcripts of LTCC genes in the human brain, but not to unravel the transcriptional regulation mechanism that produces the transcripts. Transcription begins once RNA polymerase and other transcription factors bind to the protein binding sites within the promoter region. A promoter is a sequence of bases that regulates the transcription of a gene that stretches from the upstream of the TSS of a gene and overlaps with the 5' end region of the gene. The regulatory elements are located near the promoter region (Mariño-Ramírez et al., 2004). In detail, transcription factors bind to the TATA box and start to form an RNA polymerase transcription complex which triggers transcription. The interaction between transcription factors and *cis*-acting elements in promoter regions controls gene expression (Haberle & Stark, 2018). The short-length transcripts may also be created using TATA box. For example, a promoter for *NRG1* type II mRNA contains TATA-like sequences (TAAATAAA) which transcription factors would bind to (W. Tan et al., 2007). However, the promoter region may not contain TATA-like sequences. *CCAT* uses the exonic promoter without TATA like sequences (Gomez-Ospina et al., 2013). Specifically, 238bp upstream from the M2011 in exon 46 is necessary for the transcription of *CCAT* (Gomez-Ospina et al., 2013). Likewise, the 3' transcripts of LTCC genes may also use the exonic promoter region to initiate the transcription in the human brain. Once the promoter region is identified, the transcription regulatory elements that bind to the promoter will be identified. In the human brain, a transcription factor, Enhancer of zeste homolog 2 (EZH2), binds to the promoter site of full-length *CACNA1C* and regulates the expression of *CACNA1C* in the developing brain (Billingsley et al., 2018). Moreover, it was predicted that other VGCC genes also contain EZH2 binding sites in their promoter region

(Billingsley et al., 2018). These show that transcription factors like EZH2 may bind to the promoter region of the 3' transcripts of LTCC genes to initiate the transcription in the human brain.

Identifying the promoter region of the 3' transcripts of LTCC genes will elucidate whether the production of the transcripts is affected by DNA variation at single nucleotide polymorphism (SNP) or epigenetic factors. For example, the activity of the promoter region for *NRG1* type IV is affected by schizophrenia risk-associated SNP (rs6994992). If there is T allele at rs6994992, the promoter activity is increased by 65% compared with C allele, implying that the association between a cis-acting regulatory element in the promoter region and DNA variation at SNP regulates the transcript expression (W. Tan et al., 2007).

Moreover, *BDNF* expression is affected by epigenetic factors including DNA methylation level and histone modifications (Cattaneo et al., 2016). CpG islands, which may undergo methylation, are more likely to be clustered near the promoter regions. Therefore, methylation can regulate the transcription of the 3' transcripts by affecting the promoter region. Increased methylation level at promoter IV of *BDNF* is associated with lower *BDNF* protein levels in the prefrontal cortex of the elders (Mellios et al., 2008). In post-mortem brain tissues, the increased histone methyltransferase H3K4 trimethylation, which is an active chromatin mark at promoters I and IV, is associated with increased expression of *BDNF* (Mellios et al., 2008). Moreover, a reduced H3K27 methylation state at promoter IV of *BDNF* was detected in post-mortem brain tissues of patients who used antidepressants compared to patients without antidepressant consumption (E. S. Chen et al., 2011). Similar to the promoters of *BDNF*, promoters of 3' transcripts of LTCCs may be affected by either SNP or an epigenetic mechanism, which would in turn regulate the transcription of these transcripts.

#### 6.3.4. Further research of examining transcription termination sites of 3' transcripts of LTCC genes via 3' RACE

As mentioned in chapter 5, the protein lysate of C-Cav1.1 transfected in HEK293T cells did not show the predicted size, meaning that it may use alternative translation start sites. However, transcription termination sites (TTSs) may also affect the translation or protein localization (Lekk et al., 2023). Over 70% of mammalian genes produce multiple transcript variants with different 3' UTRs by using alternative polyadenylation during transcription (Derti et al., 2012; Hoque et al., 2013). Therefore, a gene having alternative TTSs can be translated into different proteins with different functions (Reyes & Huber, 2018). Alternative 3'UTRs can affect protein localization, or produce non-coding RNA fragments (Lekk et al., 2023). If the alternative polyadenylation is located at the exonic regions, the transcript would have alternative last exons and be localized in neurons differently compared with the transcript with a canonical polyadenylation site (Pereira-Castro & Moreira, 2021).

Furthermore, in the transcript with different 3'UTR lengths, even though the protein-coding sequence is not affected, the gene expression is still influenced due to the cis-regulatory elements in the 3'UTR (Pereira-Castro & Moreira, 2021). The longer 3' UTRs are more likely to have more regulatory elements that function to repress translation by interacting with miRNAs and RNA binding proteins (Lekk et al., 2023). For example, *BDNF* produces different transcript variants with two different polyadenylation sites (Lau et al., 2010). One variant is with a shorter 3'UTR, which functions to maintain *BDNF* protein level at the basal level of neuronal activity. The other variant is with a longer 3' UTR, which was found to suppress *BDNF* translation at rest in neuronal dendrites to regulate the development of the normal dendritic spine and long-term potentiation (Lau et al., 2010). However, it was studied later that the longer 3'UTR does not repress



the translation of BDNF in cultured neurons (Lekk et al., 2023). Likewise, the 3' transcripts of LTCC genes could produce additional transcripts using alternative TSSs which may be translated differently. Therefore, the 3'RACE experiment could have given a more comprehensive structure of 3' transcripts of LTCC genes and the localization of translated protein in the human brain.

## **6.4. Limitation in the experiment design**

### **6.4.1. Experimental design could have been planned in a different way**

The experiments in this thesis could have been planned in a different way that would make the study more coherent. During the first year of my DPhil when I was specifying the project, I did PCR amplification of all VGCC genes that encode VGCC subunits: either  $\alpha 1$  subunits,  $\beta$  subunits, or  $\alpha 2\delta$  subunits in the human brain (see the table in Appendix 17). If the VGCC gene is assumed to have more than one start site according to the short-read RNA-seq data from LIBD (Jaffe et al., 2015), all the possible transcripts of the gene were amplified using PCR amplification. In the end, the 3' transcripts of *CACNAIS* and *CACNAIF* were chosen to be examined further for the DPhil project.

Therefore, based on the exon expression level data from short-read RNA-seq of each gene in the human brain, a combination of PCR and long-read nanopore sequencing was conducted to examine the 3' transcripts of *CACNAIS* and *CACNAIF* in the human brain. Subsequently, the 5' RACE was exploited to examine the TSS of each 3' transcript. Therefore, long-read nanopore sequencing didn't sequence from the TSS of each 3' transcript as they were identified after the sequencing. Thus, there is a limitation to claim that the splicing patterns of the 3' transcripts were examined via long-read nanopore sequencing because it did not sequence from the TSS of each 3' transcript. If the 5'RACE experiment was conducted once the possible existences of the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* were identified in the human

brain via PCR amplification, the complete 3' transcripts of *CACNAIS* and *CACNAIF* could have been sequenced via long-read nanopore sequencing. In this way, it would be possible to show that the short-length transcripts from the 3' region of *CACNAIS* and *CACNAIF* are complete transcripts originating from the alternative TSS. Then, the splicing patterns of each 3' transcript in the human brain could be examined via long-read nanopore sequencing.

The reason why an order of experiments might have a crucial impact on the results is that an amplicon long-read nanopore sequencing was used to sequence the 3' transcript of *CACNAIS* and *CACNAIF* using cDNA from the post-mortem human brain. As a combination of PCR amplification and long-read nanopore sequencing was used, primers were designed to target the region of interest. To sequence the full-length of the 3' transcript of *CACNAIS* and *CACNAIF* in the human brain, the primers had to be designed to target the TSS of the transcripts. Fortunately, the primer used to sequence the 3' transcript of *CACNAIF* covers the TSS of the 3' transcripts whereas the primers used for long-read nanopore sequencing did not cover the TSS of the 3' transcript of *CACNAIS*. Therefore, the splicing patterns of the full-length 3' transcript of *CACNAIS* were not perfectly elucidated.

#### **6.4.2. Accessibility to other tissue types than the post-mortem human brain for positive control**

Another limitation of the experiments in this project is the limited accessibility to other samples than the post-mortem human brain as positive controls for the experiments. Two genes, *CACNAIS* and *CACNAIF*, were reported to be highly expressed in the skeletal muscle and retina respectively, but not in the human brain (Catterall et al., 2005; McRory et al., 2004; Zamponi et al., 2015). However, one previous study detected *CACNAIS* (554bp fragment and 508bp fragment) in mouse retina and skeletal muscle via RT-PCR and northern blot analyses. Additionally, from

western blot analysis, Cav1.1 was strongly detected in the mouse skeletal muscle while weakly detected in the mouse heart, retina, and brain (Specht et al., 2009). Furthermore, a short cDNA fragment of mouse *CACNAIF* (759bp in length from exon 42 to exon 48) was detected in the mouse retina and pineal gland via RT-PCR and *in situ* hybridization analysis (Hemara-Wahanui et al., 2005). Even though *CACNAIF* could not be amplified in the human brain using RT-PCR in another study (McRory et al., 2004), there is the possibility of *CACNAIF* expression in the human brain but in a restricted manner (Hemara-Wahanui et al., 2005). As these studies used several different tissues than one tissue, the possible detection of each cDNA fragment of *CACNAIS* and *CACNAIF* can be compared to the well-known tissues. These studies showed the importance of positive control tissues in which *CACNAIS* and *CACNAIF* are strongly expressed.

Therefore, having those two additional tissues (Skeletal muscle, and/or Retina) included in the experiments would make the identification of the 3' transcripts of LTCC genes more concrete. This limitation is also related to the experimental method used in the project, PCR amplification. Not being able to amplify the full-length *CACNAIS* and *CACNAIF* in the human brain could depend on the primers or settings used for the PCR amplification. Therefore, it would make the project stronger if the identification of the 3' transcripts of LTCC genes in other tissues (skeletal muscle and retina) was also examined empirically.

#### **6.4.3. The identification of the 3' transcripts of *CACNAIS* and *CACNAIF* could be validated in the human brain via Northern blot analysis.**

The presence of the 3' transcripts of *CACNAIS* and *CACNAIF* in the human brain could be affected by PCR bias. PCR bias happens because the efficiency of amplification varies between samples (Polz & Cavanaugh, 1998), or it is forced to anneal at a certain temperature rather than anneal the most abundant sample by itself (Acinas et al., 2005). To reduce the effect of PCR

amplification bias, the number of PCR cycles was set to the minimum (30-35 cycles). However, the changes in PCR cycles do not have a significant effect on eliminating amplification bias (Krehenwinkel et al., 2017), meaning that even though there was an effort to reduce the amount of PCR amplification, there is always a chance that the results are affected by PCR amplification bias.

In order to neutralize the possible effects of PCR bias on the results, it would be useful to validate the existence of 3' transcripts of *CACNAIS* and *CACNAIF* in the human brain using Northern blot analysis, which was used to identify novel transcripts of LTCC genes previously (Gomez-Ospina et al., 2013; Okagaki et al., 2001). Northern blot analysis provides information on the expression level and size of transcripts. It is a useful approach to study transcript regulation as it isolates novel transcripts from mRNAs (T. Yang et al., 2022). The signal strength from the northern blot, which is not affected by PCR bias, directly shows the gene copy number which means that it allows direct quantitation and comparison of transcripts or RNA level of the samples (Streit et al., 2009). Therefore, this approach provides highly valid gene expression data and is widely used to validate high-throughput sequencing data (Schlamp et al., 2008). The limitation of using Northern blotting is that it has a low sensitivity because it is vulnerable to RNA degradation as it requires a longer time but, a protocol developed by Yang et al., (2022) has improved sensitivity and is less expensive (T. Yang et al., 2022).

## **6.5. Concluding remarks**

Previous reviews about VGCCs reported expression of *CACNAIS* and *CACNAIF* predominately in the skeletal muscle and the retina, respectively, though expression in the human brain was unclear (Hemara-Wahanui et al., 2005; Specht et al., 2009; Zamponi et al., 2015). This thesis contains a series of experiments and analyses to characterize and investigate the short-length transcripts expected to originate from the alternative TSS located at the 3' region of LTCC genes

in the human brain. Herein, it was studied that LTCC genes produce the short-length transcripts from alternative TSSs in their 3' regions. The 5'UTRs may contain functional motifs as they are more conserved than intronic regions across 240 mammals. Moreover, the expressions arising from the novel TSSs were supported by the exon expression dataset. The transcripts from alternative TSSs may show tissue-specific expression patterns. In particular, the 3' transcripts of *CACNAIF* and *CACNAIC* show a significantly higher expression in the human brain tissues compared with the non-brain human tissues. Though the protein-protein interaction was not elucidated, it was assumed that the protein encoded by the 3' transcripts of *CACNAIS* and *CACNAIF* may play a functional role in the cells by interacting with other proteins. The 3' transcripts of *CACNAIC* and *CACNAID* may function as a transcription factor in SH-SY5Y cells. In summary, LTCC genes use alternative TSSs to produce short-length transcripts in the human brain; however, the specific function in the cell remains to be elucidated.

As stated before, there are limitations regarding the experiments in this thesis. Nevertheless, this thesis claims the novel findings regarding the 3' transcripts of LTCC genes in the human brain that have not been investigated before. This thesis confirms that LTCC genes produce short-length transcripts from the novel TSSs in the 3' region of each gene.

To sum up, characterizing not only the 3' transcripts of LTCC genes but also the 3' transcripts of other VGCC genes in the human brain would provide an insightful understanding of genetic diseases and possible therapies. A study reported that the *CACNA1A* mRNA transcript produces the secondary protein of *CACNA1A*,  $\alpha 1$ ACT.  $\alpha 1$ ACT functions to regulate the expression of neuronal genes (Du et al., 2013) and is associated with many disorders such as spinocerebellar ataxia type 6 (SCA6), motor and developmental deficits in cerebellar atrophy (Du et al., 2019). However, inhibition of  $\alpha 1$ ACT in the cerebellum does not affect neurological or morphological

deterioration in the adult brain therefore,  $\alpha$ 1ACT would be a very effective potential therapeutic target (Du et al., 2019). Still, the mechanisms of producing the C-terminal proteins are unclear (Du et al., 2019; L. Lu et al., 2015). The 3' transcripts of LTCC genes would also be associated with neuronal disorders and can be therapeutic targets. Moreover, this thesis suggests that the C-terminal proteins could also be produced from the 3' transcripts of VGCC genes arising from the novel TSSs. This thesis brings an interesting yet important suggestion to the VGCC field, that there might be a more complex transcriptional regulation in the 3' region of VGCCs in the human brain.

## REFERENCES

- Abbasi, O., Rostami, A., & Karimian, G. (2011). Identification of exonic regions in DNA sequences using cross-correlation and noise suppression by discrete wavelet transform. *BMC Bioinformatics*, *12*, 430. <https://doi.org/10.1186/1471-2105-12-430>
- Acampora, D., Giovanni Di Giovannantonio, L., Di Salvio, M., Mancuso, P., & Simeone, A. (2009). Selective inactivation of Otx2 mRNA isoforms reveals isoform-specific requirement for visceral endoderm anteriorization and head morphogenesis and highlights cell diversity in the visceral endoderm. *Mechanisms of Development*, *126*(10), 882–897. <https://doi.org/10.1016/j.mod.2009.07.003>
- Acinas, S. G., Sarma-Rupavtarm, R., Klepac-Ceraj, V., & Polz, M. F. (2005). PCR-Induced Sequence Artifacts and Bias: Insights from Comparison of Two 16S rRNA Clone Libraries Constructed from the Same Sample. *Applied and Environmental Microbiology*, *71*(12), 8966–8969. <https://doi.org/10.1128/AEM.71.12.8966-8969.2005>
- Algama, M., Tasker, E., Williams, C., Parslow, A. C., Bryson-Richardson, R. J., & Keith, J. M. (2017). Genome-wide identification of conserved intronic non-coding sequences using a Bayesian segmentation approach. *BMC Genomics*, *18*, 259. <https://doi.org/10.1186/s12864-017-3645-2>
- Arce, L., Yokoyama, N. N., & Waterman, M. L. (2006). Diversity of LEF/TCF action in development and disease. *Oncogene*, *25*(57), 7492–7504. <https://doi.org/10.1038/sj.onc.1210056>
- Arneson, A., & Ernst, J. (2019). Systematic discovery of conservation states for single-nucleotide annotation of the human genome. *Communications Biology*, *2*(248). <https://doi.org/10.1038/s42003-019-0488-1>

- Arribere, J. A., & Gilbert, W. V. (2013). Roles for transcript leaders in translation and mRNA decay revealed by transcript leader sequencing. *Genome Research*, 23(6), 977–987.  
<https://doi.org/10.1101/gr.150342.112>
- Balwierz, P. J., Carninci, P., O Daub, C., Kawai, J., Hayashizaki, Y., Van Belle, W., Beisel, C., & van Nimwegen, E. (2009). Methods for analyzing deep sequencing expression data: Constructing the human and mouse promoterome with deepCAGE data. *Genome Biology*, 10(R79). <https://doi.org/10.1186/gb-2009-10-7-r79>
- Barbado, M., Fablet, K., Ronjat, M., & De Waard, M. (2009). Gene regulation by voltage-dependent calcium channels. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, 1793(6), 1096–1104. <https://doi.org/10.1016/j.bbamcr.2009.02.004>
- Barrett, L. W., Fletcher, S., & Wilton, S. D. (2012). Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cell and Molecular Life Sciences*, 69(21), 3613–3634. <https://doi.org/10.1007/s00018-012-0990-9>
- Baumgartner, M., Lemoine, C., Al Seesi, S., Karunakaran, D. K. P., Sturrock, N., Banday, A. R., Kilcollins, A. M., Mandoiu, I., & Kanadia, R. N. (2015). Minor splicing snRNAs are enriched in the developing mouse CNS and are crucial for survival of differentiating retinal neurons. *Developmental Neurobiology*, 75(9), 895–907.  
<https://doi.org/10.1002/dneu.22257>
- Bernhofer, M., Goldberg, T., Wolf, S., Ahmed, M., Zaugg, J., Boden, M., & Rost, B. (2018). NLSdb—Major update for database of nuclear localization signals and nuclear export signals. *Nucleic Acids Research*, 46(D1), D503–D508.  
<https://doi.org/10.1093/nar/gkx1021>



- Berridge, M. J. (2014). Calcium signalling and psychiatric disease: Bipolar disorder and schizophrenia. *Cell and Tissue Research*, 357(2), 477–492.  
<https://doi.org/10.1007/s00441-014-1806-z>
- Bespalova, I. N., Adkins, S., & Burmeister, M. (1998). 3' RACE: Skewed ratio of specific to general PCR primers improves yield and specificity. *BioTechniques*, 24(4), 575–577.  
<https://doi.org/10.2144/98244bm12>
- Billingsley, K. J., Manca, M., Gianfrancesco, O., Collier, D. A., Sharp, H., Bubb, V. J., & Quinn, J. P. (2018). Regulatory characterisation of the schizophrenia-associated CACNA1C proximal promoter and the potential role for the transcription factor EZH2 in schizophrenia aetiology. *Schizophrenia Research*, 199, 168–175.  
<https://doi.org/10.1016/j.schres.2018.02.036>
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigó, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Stamatoyannopoulos, J. A., Thurman, R. E., Kuehn, M. S., Taylor, C. M., Neph, S., Koch, C. M., Asthana, S., Malhotra, A., Adzhubei, I., Greenbaum, J. A., ... Transcriptional Regulatory Elements. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447(7146), Article 7146.  
<https://doi.org/10.1038/nature05874>
- Blumenstein, Y., Kanevsky, N., Sahar, G., Barzilai, R., Ivanina, T., & Dascal, N. (2002). A novel long N-terminal isoform of human L-type Ca<sup>2+</sup> channel is up-regulated by protein kinase C. *The Journal of Biological Chemistry*, 277(5), 3419–3423.  
<https://doi.org/10.1074/jbc.C100642200>

- Bock, G., Gebhart, M., Scharinger, A., Jangsangthong, W., Busquet, P., Poggiani, C., Sartori, S., Mangoni, M. E., Sinnegger-Brauns, M. J., Herzig, S., Striessnig, J., & Koschak, A. (2011). Functional properties of a newly identified C-terminal splice variant of Cav1.3 L-type Ca<sup>2+</sup> channels. *The Journal of Biological Chemistry*, *286*(49), 42736–42748. <https://doi.org/10.1074/jbc.M111.269951>
- Bradley, K. J., Bowl, M. R., Williams, S. E., Ahmad, B. N., Partridge, C. J., Patmanidi, A. L., Kennedy, A. M., Loh, N. Y., & Thakker, R. V. (2007). Parafibromin is a nuclear protein with a functional monopartite nuclear localization signal. *Oncogene*, *26*(8), Article 8. <https://doi.org/10.1038/sj.onc.1209893>
- Bunz, F. (2002). Human cell knockouts. *Current Opinion in Oncology*, *14*(1), 73–78.
- Byeon, G. W., Cenik, E. S., Jiang, L., Tang, H., Das, R., & Barna, M. (2021). Functional and structural basis of extreme conservation in vertebrate 5' untranslated regions. *Nature Genetics*, *53*(5), Article 5. <https://doi.org/10.1038/s41588-021-00830-1>
- Byrne, A., Beaudin, A. E., Olsen, H. E., Jain, M., Cole, C., Palmer, T., DuBois, R. M., Forsberg, E. C., Akeson, M., & Vollmers, C. (2017). Nanopore long-read RNAseq reveals widespread transcriptional variation among the surface receptors of individual B cells. *Nature Communications*, *8*, 16027. <https://doi.org/10.1038/ncomms16027>
- Carninci, P., Sandelin, A., Lenhard, B., Katayama, S., Shimokawa, K., Ponjavic, J., Semple, C. A. M., Taylor, M. S., Engström, P. G., Frith, M. C., Forrest, A. R. R., Alkema, W. B., Tan, S. L., Plessy, C., Kodzius, R., Ravasi, T., Kasukawa, T., Fukuda, S., Kanamori-Katayama, M., ... Hayashizaki, Y. (2006). Genome-wide analysis of mammalian promoter architecture and evolution. *Nature Genetics*, *38*(6), 626–635. <https://doi.org/10.1038/ng1789>

- Casamassimi, A., & Ciccodicola, A. (2019). Transcriptional Regulation: Molecules, Involved Mechanisms, and Misregulation. *International Journal of Molecular Sciences*, 20(6), 1281. <https://doi.org/10.3390/ijms20061281>
- Cattaneo, A., Cattane, N., Begni, V., Pariante, C. M., & Riva, M. A. (2016). The human BDNF gene: Peripheral gene expression and protein levels as biomarkers for psychiatric disorders. *Translational Psychiatry*, 6(11), e958. <https://doi.org/10.1038/tp.2016.214>
- Catterall, W. A. (2010). Ion channel voltage sensors: Structure, function, and pathophysiology. *Neuron*, 67(6), 915–928. <https://doi.org/10.1016/j.neuron.2010.08.021>
- Catterall, W. A. (2011). Voltage-gated calcium channels. *Cold Spring Harbor Perspectives in Biology*, 3(8), a003947. <https://doi.org/10.1101/cshperspect.a003947>
- Catterall, W. A., Perez-Reyes, E., Snutch, T. P., & Striessnig, J. (2005). International Union of Pharmacology. XLVIII. Nomenclature and Structure-Function Relationships of Voltage-Gated Calcium Channels. *Pharmacological Reviews*, 57(4), 411–425. <https://doi.org/10.1124/pr.57.4.5>
- Chen, E. S., Ernst, C., & Turecki, G. (2011). The epigenetic effects of antidepressant treatment on human prefrontal cortex BDNF expression. *International Journal of Neuropsychopharmacology*, 14(3), 427–429. <https://doi.org/10.1017/S1461145710001422>
- Chen, M., & Manley, J. L. (2009). Mechanisms of alternative splicing regulation: Insights from molecular and genomics approaches. *Nature Reviews. Molecular Cell Biology*, 10(11), 741–754. <https://doi.org/10.1038/nrm2777>
- Clancy, S., & Brown, W. (2008). Translation: DNA to mRNA to Protein. *Nature Education*, 1(1), 101.

- Clark, M. B., Wrzesinski, T., Garcia, A. B., Hall, N. A. L., Kleinman, J. E., Hyde, T., Weinberger, D. R., Harrison, P. J., Haerty, W., & Tunbridge, E. M. (2020). Long-read sequencing reveals the complex splicing profile of the psychiatric risk gene *CACNA1C* in human brain. *Molecular Psychiatry*, *25*(1), Article 1. <https://doi.org/10.1038/s41380-019-0583-1>
- Corbett, B. F., Leiser, S. C., Ling, H.-P., Nagy, R., Breysse, N., Zhang, X., Hazra, A., Brown, J. T., Randall, A. D., Wood, A., Pangalos, M. N., Reinhart, P. H., & Chin, J. (2013). Sodium Channel Cleavage Is Associated with Aberrant Neuronal Activity and Cognitive Deficits in a Mouse Model of Alzheimer's Disease. *Journal of Neuroscience*, *33*(16), 7020–7026. <https://doi.org/10.1523/JNEUROSCI.2325-12.2013>
- Courtois, V., Chatelain, G., Han, Z.-Y., Le Novere, N., Brun, G., & Lamonerie, T. (2003). New *Otx2* mRNA isoforms expressed in the mouse brain. *Journal of Neurochemistry*, *84*(4), 840–853.
- Cox, R. H., & Fromme, S. J. (2013). A naturally occurring truncated *Cav1.2*  $\alpha 1$ -subunit inhibits  $Ca^{2+}$  current in A7r5 cells. *American Journal of Physiology - Cell Physiology*, *305*(8), C896–C905. <https://doi.org/10.1152/ajpcell.00217.2013>
- Dai, B., Saada, N., Echeteu, C., Dettbarn, C., & Palade, P. (2002). A new promoter for  $\alpha 1C$  subunit of human L-type cardiac calcium channel *Ca(V)1.2*. *Biochemical and Biophysical Research Communications*, *296*(2), 429–433. [https://doi.org/10.1016/s0006-291x\(02\)00894-x](https://doi.org/10.1016/s0006-291x(02)00894-x)
- Davenport, B., Li, Y., Heizer, J. W., Schmitz, C., & Perraud, A.-L. (2015). Signature Channels of Excitability no More: L-Type Channels in Immune Cells. *Frontiers in Immunology*, *6*. <https://www.frontiersin.org/articles/10.3389/fimmu.2015.00375>

- Davis Jr, W., & Schultz, R. M. (2000). Developmental change in TATA-box utilization during preimplantation mouse development. *Developmental Biology*, 218(2), 275–283.  
<https://doi.org/10.1006/dbio.1999.9486>
- Davis, M. W., & Jorgensen, E. M. (2022). ApE, A Plasmid Editor: A Freely Available DNA Manipulation and Visualization Program. *Frontiers in Bioinformatics*.  
<https://doi.org/10.3389/fbinf.2022.818619>
- de Hoon, M., Shin, J. W., & Carninci, P. (2015). Paradigm shifts in genomics through the FANTOM projects. *Mammalian Genome*, 26(9), 391–402.  
<https://doi.org/10.1007/s00335-015-9593-8>
- De Jongh, K. S., Warner, C., Colvin, A. A., & Catterall, W. A. (1991). Characterization of the two size forms of the alpha 1 subunit of skeletal muscle L-type calcium channels. *Proceedings of the National Academy of Sciences of the United States of America*, 88(23), 10778–10782. <https://doi.org/10.1073/pnas.88.23.10778>
- Deamer, D., Akeson, M., & Branton, D. (2016). Three decades of nanopore sequencing. *Nature Biotechnology*, 34(5), 518–524. <https://doi.org/10.1038/nbt.3423>
- Derti, A., Garrett-Engle, P., MacIsaac, K. D., Stevens, R. C., Sriram, S., Chen, R., Rohl, C. A., Johnson, J. M., & Babak, T. (2012). A quantitative atlas of polyadenylation in five mammals. *Genome Research*, 22(6), 1173–1183. <https://doi.org/10.1101/gr.132563.111>
- Dhamija, S., & Menon, M. B. (2018). Non-coding transcript variants of protein-coding genes – what are they good for? *RNA Biol*, 15(8), 1025–1031.  
<https://doi.org/10.1080/15476286.2018.1511675>
- Dillman, A. A., Hauser, D. N., Gibbs, J. R., Nalls, M. A., McCoy, M. K., Rudenko, I. N., Galter, D., & Cookson, M. R. (2013). mRNA expression, splicing and editing in the embryonic

- and adult mouse cerebral cortex. *Nature Neuroscience*, *16*(4), 499–506.  
<https://doi.org/10.1038/nn.3332>
- Dolphin, A. C. (2006). A short history of voltage-gated calcium channels. *British Journal of Pharmacology*, *147*(Suppl 1), S56–S62. <https://doi.org/10.1038/sj.bjp.0706442>
- Dong, X., Tian, L., Gouil, Q., Kariyawasam, H., Su, S., De Paoli-Iseppi, R., Prawer, Y. D. J., Clark, M. B., Breslin, K., Iminoff, M., Blewitt, M. E., Law, C. W., & Ritchie, M. E. (2021). The long and the short of it: Unlocking nanopore long-read RNA sequencing data with short-read differential expression analysis tools. *NAR Genomics and Bioinformatics*, *3*(2). <https://doi.org/10.1093/nargab/lqab028>
- Drews, V. L., Lieberman, A. P., & Meisler, M. H. (2005). Multiple transcripts of sodium channel SCN8A (NaV1.6) with alternative 5'- and 3'-untranslated regions and initial characterization of the SCN8A promoter. *Genomics*, *85*(2), 245–257.  
<https://doi.org/10.1016/j.ygeno.2004.09.002>
- Drews, V. L., Shi, K., de Haan, G., & Meisler, M. H. (2007). Identification of evolutionarily conserved, functional noncoding elements in the promoter region of the sodium channel gene SCN8A. *Mammalian Genome*, *18*, 723–731. <https://doi.org/10.1007/s00335-007-9059-8>
- Du, X., Wang, J., Zhu, H., Rinaldo, L., Lamar, K.-M., Palmenberg, A. C., Hansel, C., & Gomez, C. M. (2013). Second cistron in CACNA1A gene encodes a transcription factor mediating cerebellar development and SCA6. *Cell*, *154*(1), 118–133.  
<https://doi.org/10.1016/j.cell.2013.05.059>
- Du, X., Wei, C., Pastor, D. P. H., Rao, E. R., Li, Y., Grasselli, G., Godfrey, J., Palmenberg, A. C., Andrade, J., Hansel, C., & Gomez, C. M. (2019). A1ACT is essential for survival and

- early cerebellar programming in a critical neonatal window. *Neuron*, 102(4), 770-785.e7.  
<https://doi.org/10.1016/j.neuron.2019.02.036>
- Dvinge, H. (2018). Regulation of alternative mRNA splicing: Old players and new perspectives. *FEBS Letters*, 592(17), 2987–3006. <https://doi.org/10.1002/1873-3468.13119>
- Dvorak, P., Leupen, S., & Soucek, P. (2019). Functionally Significant Features in the 5' Untranslated Region of the ABCA1 Gene and Their Comparison in Vertebrates. *Cells*, 8(6), 623. <https://doi.org/10.3390/cells8060623>
- Ebihara, T., Komiya, Y., Izumi-Nakaseko, H., Adachi-Akahane, S., Okabe, S., & Okamura, Y. (2002). Coexpression of a Cav1.2 protein lacking an N-terminus and the first domain specifically suppresses L-type calcium channel activity. *FEBS Letters*, 529(2–3), 203–207. [https://doi.org/10.1016/S0014-5793\(02\)03340-9](https://doi.org/10.1016/S0014-5793(02)03340-9)
- Fan, Q. I., Vanderpool, K. M., Chung, H.-S., & Marsh, J. D. (2005). The L-type calcium channel alpha 1C subunit gene undergoes extensive, uncoordinated alternative splicing. *Molecular and Cellular Biochemistry*, 269(1), 153–163. <https://doi.org/10.1007/s11010-005-3455-8>
- Feng, T., Kalyaanamoorthy, S., Barakat, K., Feng, T., Kalyaanamoorthy, S., & Barakat, K. (2018). L-Type Calcium Channels: Structure and Functions. In *Ion Channels in Health and Sickness*. IntechOpen. <https://doi.org/10.5772/intechopen.77305>
- Flames, N., Long, J. E., Garratt, A. N., Fischer, T. M., Gassmann, M., Birchmeier, C., Lai, C., Rubenstein, J. L. R., & Marín, O. (2004). Short- and long-range attraction of cortical GABAergic interneurons by neuregulin-1. *Neuron*, 44(2), 251–261.  
<https://doi.org/10.1016/j.neuron.2004.09.028>

- Flavell, S. W., & Greenberg, M. E. (2008). Signaling Mechanisms Linking Neuronal Activity to Gene Expression and Plasticity of the Nervous System. *Annual Review of Neuroscience*, *31*, 563–590. <https://doi.org/10.1146/annurev.neuro.31.060407.125631>
- Flucher, B. E. (2020). Skeletal muscle CaV1.1 channelopathies. *Pflügers Archiv - European Journal of Physiology*, *472*(7), 739–754. <https://doi.org/10.1007/s00424-020-02368-3>
- Fossat, N., Courtois, V., Chatelain, G., Brun, G., & Lamonerie, T. (2005). Alternative usage of Otx2 promoters during mouse development. *Developmental Dynamics*, *233*(1), 154–160.
- Fu, X., Liang, C., Li, F., Wang, L., Wu, X., Lu, A., Xiao, G., & Zhang, G. (2018). The Rules and Functions of Nucleocytoplasmic Shuttling Proteins. *International Journal of Molecular Sciences*, *19*(5), Article 5. <https://doi.org/10.3390/ijms19051445>
- Fuller, M. D., Emrick, M. A., Sadilek, M., Scheuer, T., & Catterall, W. A. (2010). Molecular mechanism of calcium channel regulation in the fight-or-flight response. *Science Signaling*, *3*(141), ra70. <https://doi.org/10.1126/scisignal.2001152>
- Furlanis, E., Traunmüller, L., Fucile, G., & Scheiffele, P. (2019). Landscape of ribosome-engaged transcript isoforms reveals extensive neuronal-cell-class-specific alternative splicing programs. *Nature Neuroscience*, *22*(10), 1709–1717. <https://doi.org/10.1038/s41593-019-0465-5>
- Garibyan, L., & Avashia, N. (2013). Research Techniques Made Simple: Polymerase Chain Reaction (PCR). *Journal of Investigative Dermatology*, *133*(3), 1–4. <https://doi.org/10.1038/jid.2013.1>
- Gentleman, R. C., Carey, V. J., Bates, D. M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A. J., ... Zhang, J. (2004). Bioconductor: Open



- software development for computational biology and bioinformatics. *Genome Biology*, 5(10), R80. <https://doi.org/10.1186/gb-2004-5-10-r80>
- Gomez-Ospina, N., Panagiotakos, G., Portmann, T., Pasca, S. P., Rabah, D., Budzillo, A., Kinet, J. P., & Dolmetsch, R. E. (2013). A Promoter in the Coding Region of the Calcium Channel Gene CACNA1C Generates the Transcription Factor CCAT. *PLOS ONE*, 8(4), e60526. <https://doi.org/10.1371/journal.pone.0060526>
- Gomez-Ospina, N., Tsuruta, F., Barreto-Chang, O., Hu, L., & Dolmetsch, R. (2006). The C Terminus of the L-Type Voltage-Gated Calcium Channel CaV1.2 Encodes a Transcription Factor. *Cell*, 127(3), 591–606. <https://doi.org/10.1016/j.cell.2006.10.017>
- Gonzalez-Porta, M., Frankish, A., Rung, J., Harrow, J., & Brazma, A. (2013). Transcriptome analysis of human tissues and cell lines reveals one dominant transcript per gene. *Genome Biology*, 14(7), R70. <https://doi.org/10.1186/gb-2013-14-7-r70>
- Gonzalez-Ramirez, R., & Felix, R. (2018). Transcriptional regulation of voltage-gated Ca<sup>2+</sup> channels. *Acta Physiologica*, 222(1). <https://doi.org/10.1111/apha.12883>
- Goss, D. J., & Domashevskiy, A. V. (2016). Messenger RNA (mRNA): The Link between DNA and Protein. *Encyclopedia of Cell Biology*, 1, 341–345. <https://doi.org/10.1016/B978-0-12-394447-4.10040-9>
- Goszczynski, D. E., Halstead, M. M., Islas-Trejo, A. D., Zhou, H., & Ross, P. J. (2021). Transcription initiation mapping in 31 bovine tissues reveals complex promoter activity, pervasive transcription, and tissue-specific promoter usage. *Genome Research*, 31(4), 732–744. <https://doi.org/10.1101/gr.267336.120>

- Gray, A. C., Raingo, J., & Lipscombe, D. (2007). Neuronal calcium channels: Splicing for optimal performance. *Cell Calcium*, 42(4–5), 409–417.  
<https://doi.org/10.1016/j.ceca.2007.04.003>
- Haberle, V., Forrest, A. R. R., Hayashizaki, Y., Carninci, P., & Lenhard, B. (2015). CAGER: Precise TSS data retrieval and high-resolution promoterome mining for integrative analyses. *Nucleic Acids Research*, 43(8), e51. <https://doi.org/10.1093/nar/gkv054>
- Haberle, V., & Stark, A. (2018). Eukaryotic core promoters and the functional basis of transcription initiation. *Nature Reviews Molecular Cell Biology*, 19, 621–637.  
<https://doi.org/10.1038/s41580-018-0028-8>
- Haeseleer, F., Williams, B., & Lee, A. (2016). Characterization of C-terminal Splice Variants of Cav1.4 Ca<sup>2+</sup> Channels in Human Retina \*. *Journal of Biological Chemistry*, 291(30), 15663–15673. <https://doi.org/10.1074/jbc.M116.731737>
- Hall, D. D., Dai, S., Tseng, P.-Y., Malik, Z., Nguyen, M., Matt, L., Schnizler, K., Shephard, A., Mohapatra, D. P., Tsuruta, F., Dolmetsch, R. E., Christel, C. J., Lee, A., Burette, A., Weinberg, R. J., & Hell, J. W. (2013). Competition between  $\alpha$ -actinin and Ca<sup>2+</sup>-calmodulin controls surface retention of the L-type Ca<sup>2+</sup> channel Ca(V)<sub>1.2</sub>. *Neuron*, 78(3), 483–497. <https://doi.org/10.1016/j.neuron.2013.02.032>
- Hall, N. A. L., Husain, S. M., Lee, H., & Tunbridge, E. M. (2021). Long read transcript profiling of ion channel splice isoforms. *Methods Enzymology*, 654, 345–364.  
<https://doi.org/10.1016/bs.mie.2021.02.015>
- Harbers, M., & Carninci, P. (2005). Tag-based approaches for transcriptome research and genome annotation. *Nature Methods*, 2(7), 495–502. <https://doi.org/10.1038/nmeth768>

- Harris, C. R., Millman, K. J., van der Walt, S. J., Gommers, R., Virtanen, P., Cournapeau, D., Wieser, E., Taylor, J., Berg, S., Smith, N. J., Kern, R., Picus, M., Hoyer, S., van Kerkwijk, M. H., Brett, M., Haldane, A., Fernandez del Rio, J., Wiebe, M., Peterson, P., ... Oliphant, T. E. (2020). Array programming with NumPy. *Nature*, *585*, 357–362.
- Harrison, P. J., & Law, A. J. (2006). Neuregulin 1 and schizophrenia: Genetics, gene expression, and neurobiology. *Biological Psychiatry*, *60*(2), 132–140.  
<https://doi.org/10.1016/j.biopsych.2005.11.002>
- Harrison, P. J., Tunbridge, E. M., Dolphin, A. C., & Hall, J. (2020). Voltage-gated calcium channel blockers for psychiatric disorders: Genomic reappraisal. *The British Journal of Psychiatry*, *216*(5), 250–253. <https://doi.org/10.1192/bjp.2019.157>
- Heck, J., Palmeira Do Amaral, A. C., Weißbach, S., El Khallouqi, A., Bikbaev, A., & Heine, M. (2021). More than a pore: How voltage-gated calcium channels act on different levels of neuronal communication regulation. *Channels*, *15*(1), 322–338.  
<https://doi.org/10.1080/19336950.2021.1900024>
- Hell, J. W., Westenbroek, R. E., Warner, C., Ahlijanian, M. K., Prystay, W., Gilbert, M. M., Snutch, T. P., & Catterall, W. A. (1993). Identification and differential subcellular localization of the neuronal class C and class D L-type calcium channel alpha 1 subunits. *Journal of Cell Biology*, *123*(4), 949–962. <https://doi.org/10.1083/jcb.123.4.949>
- Hemara-Wahanui, A., Berjukow, S., Hope, C. I., Dearden, P. K., Wu, S.-B., Wilson-Wheeler, J., Sharp, D. M., Lundon-Treweek, P., Clover, G. M., Hoda, J.-C., Striessnig, J., Marksteiner, R., Hering, S., & Maw, M. A. (2005). A CACNA1F mutation identified in an X-linked retinal disorder shifts the voltage dependence of Cav1.4 channel activation. *PNAS*, *102*(21), 7553–7558. <https://doi.org/10.1073/pnas.0501907102>

- Hill, J. L., Jimenez, D. V., Maynard, K. R., Kardian, A. S., Pollock, C. J., Schloesser, R. J., & Martinowich, K. (2016). Loss of promoter IV-driven BDNF expression impacts oscillatory activity during sleep, sensory information processing and fear regulation. *Translational Psychiatry*, 6, e873. <https://doi.org/10.1038/tp.2016.153>
- Hoque, M., Ji, Z., Zheng, D., Luo, W., Li, W., You, B., Park, J. Y., Yehia, G., & Tian, B. (2013). Analysis of alternative cleavage and polyadenylation by 3' region extraction and deep sequencing. *Nature Methods*, 10(2), Article 2. <https://doi.org/10.1038/nmeth.2288>
- Hulme, J. T., Konoki, K., Lin, T. W.-C., Gritsenko, M. A., Camp, D. G., Bigelow, D. J., & Catterall, W. A. (2005). Sites of proteolytic processing and noncovalent association of the distal C-terminal domain of CaV1.1 channels in skeletal muscle. *Proceedings of the National Academy of Sciences*, 102(14), 5274–5279. <https://doi.org/10.1073/pnas.0409885102>
- Hulme, J. T., Yarov-Yarovoy, V., Lin, T. W.-C., Scheuer, T., & Catterall, W. A. (2006). Autoinhibitory control of the CaV1.2 channel by its proteolytically processed distal C-terminal domain. *The Journal of Physiology*, 576(Pt 1), 87–102. <https://doi.org/10.1113/jphysiol.2006.111799>
- Hunter, J. D. (2007). Matplotlib: A 2D Graphics Environment. *Computing in Science & Engineering*, 9(3), 90–95. <https://doi.org/10.1109/MCSE.2007.55>
- Ingolia, N. T. (2014). Ribosome profiling: New views of translation, from single codons to genome scale. *Nature Reviews. Genetics*, 15(3), 205–213. <https://doi.org/10.1038/nrg3645>
- Jaffe, A. E., Shin, J., Collado-Torres, L., Leek, J. T., Tao, R., Li, C., Gao, Y., Jia, Y., Maher, B. J., Hyde, T. M., Kleinman, J. E., & Weinberger, D. R. (2015). Developmental regulation

- of human cortex transcription and its clinical relevance at single base resolution. *Nature Neuroscience*, *18*(1), 154–161. <https://doi.org/10.1038/nn.3898>
- Kaikkonen, M. U., Lam, M. T. Y., & Glass, C. K. (2011). Non-coding RNAs as regulators of gene expression and epigenetics. *Cardiovascular Research*, *90*(3), 430–440. <https://doi.org/10.1093/cvr/cvr097>
- Kent, W. J., Sugnet, C. W., Furey, T. S., Roskin, K. M., Pringle, T. H., Zahler, A. M., & Haussler, D. (2002). The Human Genome Browser at UCSC. *Genome Research*, *12*(6), 996–1006. <https://doi.org/10.1101/gr.229102>
- Kim, D. Y., Carey, B. W., Wang, H., Ingano, L. A. M., Binshtok, A. M., Wertz, M. H., Pettingell, W. H., He, P., Lee, V. M.-Y., Woolf, C. J., & Kovacs, D. M. (2007). BACE1 regulates voltage-gated sodium channels and neuronal activity. *Nature Cell Biology*, *9*(7), 755–764. <https://doi.org/10.1038/ncb1602>
- Kim, E., Magen, A., & Ast, G. (2007). Different levels of alternative splicing among eukaryotes. *Nucleic Acids Research*, *35*(1), 125–131. <https://doi.org/10.1093/nar/gkl924>
- Kimura, K., Wakamatsu, A., Suzuki, Y., Ota, T., Nishikawa, T., Yamashita, R., Yamamoto, J., Sekine, M., Tsuritani, K., Wakaguri, H., Ishii, S., Sugiyama, T., Saito, K., Isono, Y., Irie, R., Kushida, N., Yoneyama, T., Otsuka, R., Kanda, K., ... Sugano, S. (2006). Diversification of transcriptional modulation: Large-scale identification and characterization of putative alternative promoters of human genes. *Genome Research*, *16*(1), 55–65. <https://doi.org/10.1101/gr.4039406>
- Klößner, U., Mikala, G., Eisfeld, J., Iles, D. E., Strobeck, M., Mershon, J. L., Schwartz, A., & Varadi, G. (1997). Properties of three COOH-terminal splice variants of a human cardiac

- L-type Ca<sup>2+</sup>-channel alpha1-subunit. *The American Journal of Physiology*, 272(3 Pt 2), H1372-1381. <https://doi.org/10.1152/ajpheart.1997.272.3.H1372>
- Kodzius, R., Kojima, M., Nishiyori, H., Nakamura, M., Fukuda, S., Tagami, M., Sasaki, D., Imamura, K., Kai, C., Harbers, M., Hayashizaki, Y., & Carninci, P. (2006). CAGE: cap analysis of gene expression. *Nature Methods*, 3, 211–222. <https://doi.org/10.1038/nmeth0306-211>
- Kordasiewicz, H. B., Thompson, R. M., Clark, H. B., & Gomez, C. M. (2006). C-termini of P/Q-type Ca<sup>2+</sup> channel alpha1A subunits translocate to nuclei and promote polyglutamine-mediated toxicity. *Human Molecular Genetics*, 15(10), 1587–1599. <https://doi.org/10.1093/hmg/ddl080>
- Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. *Nucleic Acids Research*, 12(2), 857–872. <https://doi.org/10.1093/nar/12.2.857>
- Kozak, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell*, 44(2), 283–292. [https://doi.org/10.1016/0092-8674\(86\)90762-2](https://doi.org/10.1016/0092-8674(86)90762-2)
- Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Research*, 15(20), 8125–8148.
- Kramer, N. J., Haney, M. S., Morgens, D. W., Jovičić, A., Couthouis, J., Li, A., Ousey, J., Ma, R., Bieri, G., Tsui, C. K., Shi, Y., Hertz, N. T., Tessier-Lavigne, M., Ichida, J. K., Bassik, M. C., & Gitler, A. D. (2018). CRISPR-Cas9 screens in human cells and primary neurons identify modifiers of C9orf72 dipeptide repeat protein toxicity. *Nature Genetics*, 50(4), 603–612. <https://doi.org/10.1038/s41588-018-0070-7>

- Krehenwinkel, H., Wolf, M., Lim, J. Y., Rominger, A. J., Simison, W. B., & Gillespie, R. G. (2017). Estimating and mitigating amplification bias in qualitative and quantitative arthropod metabarcoding. *Scientific Reports*, 7(1), Article 1. <https://doi.org/10.1038/s41598-017-17333-x>
- Kubodera, T., Yokota, T., Ohwada, K., Ishikawa, K., Miura, H., Matsuoka, T., & Mizusawa, H. (2003). Proteolytic cleavage and cellular toxicity of the human  $\alpha 1A$  calcium channel in spinocerebellar ataxia type 6. *Neuroscience Letters*, 341(1), 74–78. [https://doi.org/10.1016/s0304-3940\(03\)00156-3](https://doi.org/10.1016/s0304-3940(03)00156-3)
- Kukurba, K. R., & Montgomery, S. B. (2015). RNA Sequencing and Analysis. *Cold Spring Harbor Protocols*, 11, 951–969. <https://doi.org/10.1101/pdb.top084970>
- Kvastad, L., Carlberg, K., Larsson, L., Villacampa, E. G., Stuckey, A., Stenbeck, L., Mollbrink, A., Zamboni, M., Magnusson, J. P., Basmaci, E., Shamikh, A., Prochazka, G., Schaupp, A.-L., Borg, Å., Fugger, L., Nistér, M., & Lundeberg, J. (2021). The spatial RNA integrity number assay for in situ evaluation of transcriptome quality. *Communications Biology*, 4(1), Article 1. <https://doi.org/10.1038/s42003-020-01573-1>
- Landry, J.-R., Mager, D. L., & Wilhelm, B. T. (2003). Complex controls: The role of alternative promoters in mammalian genomes. *Trends in Genetics: TIG*, 19(11), 640–648. <https://doi.org/10.1016/j.tig.2003.09.014>
- Lau, A. G., Irier, H. A., Gu, J., Tian, D., Ku, L., Liu, G., Xia, M., Fritsch, B., Zheng, J. Q., Dingleline, R., Xu, B., Lu, B., & Feng, Y. (2010). Distinct 3'UTRs differentially regulate activity-dependent translation of brain-derived neurotrophic factor (BDNF). *Proceedings of the National Academy of Science*, 107(36), 15945–15950. <https://doi.org/10.1073/pnas.1002929107>

- Lekk, I., Cabrera-Cabrera, F., Turconi, G., Tuvikene, J., Esvald, E.-E., Rähni, A., Casserly, L., Garton, D. R., Andressoo, J.-O., Timmusk, T., & Koppel, I. (2023). Untranslated regions of brain-derived neurotrophic factor mRNA control its translatability and subcellular localization. *Journal of Biological Chemistry*, *299*(2), 102897. <https://doi.org/10.1016/j.jbc.2023.102897>
- Liao, P., Yu, D., Hu, Z., Liang, M. C., Wang, J. J., Yu, C. Y., Ng, G., Yong, T. F., Soon, J. L., Chua, Y. L., & Soong, T. W. (2015). Alternative Splicing Generates a Novel Truncated Cav1.2 Channel in Neonatal Rat Heart\*. *Molecular Biophysics*, *290*(14), P9262-9272. <https://doi.org/10.1074/jbc.M114.594911>
- Liao, P., Zhang, H. Y., & Soong, T. W. (2009). Alternative splicing of voltage-gated calcium channels: From molecular biology to disease. *Pflügers Archiv - European Journal of Physiology*, *458*(3), 481–487. <https://doi.org/10.1007/s00424-009-0635-5>
- Licatalosi, D. D., & Darnell, R. B. (2010). RNA processing and its regulation: Global insights into biological networks. *Nature Reviews Genetics*, *11*(1), 75–87. <https://doi.org/10.1038/nrg2673>
- Lin, J., & Hu, J. (2013). SeqNLS: nuclear localization signal prediction based on frequent pattern mining and linear motif scoring. *PLOS ONE*, *8*(10), e76864. <https://doi.org/10.1371/journal.pone.0076864>
- Lipscombe, D., & Andrade, A. (2015). Calcium Channel CaV $\alpha_1$  Splice Isoforms—Tissue Specificity and Drug Action. *Current Molecular Pharmacology*, *8*(1), 22–31. <https://doi.org/10.2174/1874467208666150507103215>



- Lipscombe, D., Andrade, A., & Allen, S. E. (2013). Alternative splicing: Functional diversity among voltage-gated calcium channels and behavioral consequences. *Biochimica Et Biophysica Acta*, 1828(7), 1522–1529. <https://doi.org/10.1016/j.bbamem.2012.09.018>
- Lizio, M., Abugessaisa, I., Noguchi, S., Kondo, A., Hasegawa, A., Hon, C. C., de Hoon, M., Severin, J., Oki, S., Hayashizaki, Y., Carninci, P., Kasukawa, T., & Kawaji, H. (2019). Update of the FANTOM web resource: Expansion to provide additional transcriptome atlases. *Nucleic Acids Research*, 47(D1), D752–D758.
- Lowe, A. R., Tang, J. H., Yassif, J., Graf, M., Huang, W. Y., Groves, J. T., Weis, K., & Liphardt, J. T. (2015). Importin- $\beta$  modulates the permeability of the nuclear pore complex in a Ran-dependent manner. *ELife*, 4, e04052. <https://doi.org/10.7554/eLife.04052>
- Lu, J., Wu, T., Zhang, B., Liu, S., Song, W., Qiao, J., & Ruan, H. (2021). Types of nuclear localization signals and mechanisms of protein import into the nucleus. *Cell Communication and Signaling*, 19(1), 60. <https://doi.org/10.1186/s12964-021-00741-y>
- Lu, L., Sirish, P., Zhang, Z., Woltz, R. L., Li, N., Timofeyev, V., Knowlton, A. A., Zhang, X.-D., Yamoah, E. N., & Chiamvimonvat, N. (2015). Regulation of Gene Transcription by Voltage-gated L-type Calcium Channel, Cav1.3 \*. *Journal of Biological Chemistry*, 290(8), 4663–4676. <https://doi.org/10.1074/jbc.M114.586883>
- Ma, H., Groth, R. D., Wheeler, D. G., Barrett, C. F., & Tsien, R. W. (2011). Excitation-transcription coupling in sympathetic neurons and the molecular mechanism of its initiation. *Neuroscience Research*, 70(1), 2–8. <https://doi.org/10.1016/j.neures.2011.02.004>
- Makhnovskii, P. A., Gusev, O. A., Bokov, R. O., Gazizova, G. R., Vepkhvadze, T. F., Lysenko, E. A., Vinogradova, O. L., Kolpakov, F. A., & Popov, D. V. (2022). Alternative

- transcription start sites contribute to acute-stress-induced transcriptome response in human skeletal muscle. *Human Genomics*, *16*, 24. <https://doi.org/10.1186/s40246-022-00399-8>
- Malabat, C., Feuerbach, F., Ma, L., Saveanu, C., & Jacquier, A. (2015). Quality control of transcription start site selection by nonsense-mediated-mRNA decay. *ELife*, *4*, e06722. <https://doi.org/10.7554/eLife.06722>
- Mariño-Ramírez, L., Spouge, J. L., Kanga, G. C., & Landsman, D. (2004). Statistical analysis of over-represented words in human promoter sequences. *Nucleic Acids Research*, *32*(3), 949–958. <https://doi.org/10.1093/nar/gkh246>
- Martin, M. S., Tang, B., Ta, N., & Escayg, A. (2007). Characterization of 5' untranslated regions of the voltage-gated sodium channels SCN1A, SCN2A, and SCN3A and identification of cis-conserved noncoding sequences. *Genomics*, *90*(2), 225–235. <https://doi.org/10.1016/j.ygeno.2007.04.006>
- Maynard, K. R., Hill, J. L., Calcaterra, N. E., Palko, M. E., Kardian, A., Paredes, D., Sukumar, M., Adler, B. D., Jimenez, D. V., Schloesser, R. J., Tessarollo, L., Lu, B., & Martinowich, K. (2016). Functional Role of BDNF Production from Unique Promoters in Aggression and Serotonin Signaling. *Neuropsychopharmacology*, *41*, 1943–1955. <https://doi.org/10.1038/npp.2015.349>
- McHugh, D., Sharp, E. M., Scheuer, T., & Catterall, W. A. (2000). Inhibition of cardiac L-type calcium channels by protein kinase C phosphorylation of two sites in the N-terminal domain. *Proceedings of the National Academy of Sciences*, *97*(22), 12334–12338. <https://doi.org/10.1073/pnas.210384297>

- McKinney, W. (2010). DataStructuresforStatisticalComputinginPython. *Proceedings of the 9th Python in Science Conference*, 445. <https://doi.org/10.25080/Majora-92bf1922-00a>
- McRory, J. E., Hamid, J., Doering, C. J., Garcia, E., Parker, R., Hamming, K., Chen, L., Hildebrand, M., Beedle, A. M., Feldcamp, L., Zamponi, G. W., & Snutch, T. P. (2004). The CACNA1F gene encodes an L-type calcium channel with unique biophysical properties and tissue distribution. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 24(7), 1707–1718. <https://doi.org/10.1523/JNEUROSCI.4846-03.2004>
- McWilliam, H., Li, W., Uludag, M., Squizzato, S., Park, Y. M., Buso, N., Cowley, A. P., & Lopez, R. (2013). Analysis Tool Web Services from the EMBL-EBI. *Nucleic Acids Research*, 41(W1), W597–W600. <https://doi.org/10.1093/nar/gkt376>
- Mellios, N., Huang, H.-S., Grigorenko, A., Rogae, E., & Akbarian, S. (2008). A set of differentially expressed miRNAs, including miR-30a-5p, act as post-transcriptional inhibitors of BDNF in prefrontal cortex. *Human Molecular Genetics*, 17(19), 3030–3042. <https://doi.org/10.1093/hmg/ddn201>
- Miga, K. H., & Koren, S. (2020). Telomere-to-telomere assembly of a complete human X chromosome. *Nature*, 585, 79–84.
- Nanou, E., & Catterall, W. A. (2018). Calcium Channels, Synaptic Plasticity, and Neuropsychiatric Disease. *Neuron*, 98(3), 466–481. <https://doi.org/10.1016/j.neuron.2018.03.017>
- Nemoto, W., & Toh, H. (2012). Functional region prediction with a set of appropriate homologous sequences-an index for sequence selection by integrating structure and

- sequence information with spatial statistics. *BMC Structural Biology*, 12, 11.  
<https://doi.org/10.1186/1472-6807-12-11>
- Newbern, J., & Birchmeier, C. (2010). Nrg1/ErbB signaling networks in Schwann cell development and myelination. *Seminars in Cell & Developmental Biology*, 21(9), 922–928. <https://doi.org/10.1016/j.semcdb.2010.08.008>
- Nickless, A., Bailis, J. M., & You, Z. (2017). Control of gene expression through the nonsense-mediated RNA decay pathway. *Cell & Bioscience*, 7(1), 26.  
<https://doi.org/10.1186/s13578-017-0153-7>
- Noderer, W. L., Flockhart, R. J., Bhaduri, A., Diaz de Arce, A. J., Zhang, J., Khavari, P. A., & Wang, C. L. (2014). Quantitative analysis of mammalian translation initiation sites by FACS-seq. *Molecular Systems Biology*, 10(8), 748.  
<https://doi.org/10.15252/msb.20145136>
- Okagaki, R., Izumi, H., Okada, T., Nagahora, H., Nakajo, K., & Okamura, Y. (2001). The Maternal Transcript for Truncated Voltage-Dependent Ca<sup>2+</sup> Channels in the Ascidian Embryo: A Potential Suppressive Role in Ca<sup>2+</sup> Channel Expression. *Developmental Biology*, 230(2), 258–277. <https://doi.org/10.1006/dbio.2000.0119>
- Otake, L. R., Scamborova, P., Hashimoto, C., & Steitz, J. A. (2002). The Divergent U12-Type Spliceosome Is Required for Pre-mRNA Splicing and Is Essential for Development in *Drosophila*. *Molecular Cell*, 9(2), 439–446. [https://doi.org/10.1016/S1097-2765\(02\)00441-0](https://doi.org/10.1016/S1097-2765(02)00441-0)
- Pal, S., Gupta, R., Kim, H., Wickramasinghe, P., Baubet, V., Showe, L. C., Dahmane, N., & Davuluri, R. V. (2011). Alternative transcription exceeds alternative splicing in

- generating the transcriptome diversity of cerebellar development. *Genome Research*, 21(8), 1260–1272. <https://doi.org/10.1101/gr.120535.111>
- Pang, L., Koren, G., Wang, Z., & Nattel, S. (2003). Tissue-specific expression of two human Cav1.2 isoforms under the control of distinct 5' flanking regulatory elements. *FEBS Letters*, 546(2), 349–354. [https://doi.org/10.1016/S0014-5793\(03\)00629-X](https://doi.org/10.1016/S0014-5793(03)00629-X)
- Park, P. J. (2009). ChIP–seq: Advantages and challenges of a maturing technology. *Nature Reviews Genetics*, 10, 669–680. <https://doi.org/10.1038/nrg2641>
- Pennacchio, L. A., & Visel, A. (2010). Limits of sequence and functional conservation. *Nature Genetics*, 42(7), 557–558. <https://doi.org/10.1038/ng0710-557>
- Pereira-Castro, I., & Moreira, A. (2021). On the function and relevance of alternative 3'-UTRs in gene expression regulation. *WIREs RNA*, 12(5), e1653. <https://doi.org/10.1002/wrna.1653>
- Polz, M. F., & Cavanaugh, C. M. (1998). Bias in template-to-product ratios in multitemplate PCR. *Applied and Environmental Microbiology*, 64(10), 3724–3730. <https://doi.org/10.1128/AEM.64.10.3724-3730.1998>
- Pozner, A., Goldenberg, D., Negreanu, V., Le, S. Y., Elroy-Stein, O., Levanon, D., & Groner, Y. (2000). Transcription-coupled translation control of AML1/RUNX1 is mediated by cap- and internal ribosome entry site-dependent mechanisms. *Molecular and Cellular Biology*, 20(7), 2297–2307. <https://doi.org/10.1128/MCB.20.7.2297-2307.2000>
- Pruunsild, P., Kazantseva, A., Aid, T., Palm, K., & Timmusk, T. (2007). Dissecting the human BDNF locus: Bidirectional transcription, complex splicing, and multiple promoters. *Genomics*, 90(3), 397–406. <https://doi.org/10.1016/j.ygeno.2007.05.004>

- Resch, A. M., Ogurtsov, A. Y., Rogozin, I. B., Shabalina, S. A., & Koonin, E. (2009). Evolution of alternative and constitutive regions of mammalian 5'UTRs. *BMC Genomics*, *10*(162). <https://doi.org/10.1186/1471-2164-10-162>
- Reyes, A., & Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, *46*(2), 582–592. <https://doi.org/10.1093/nar/gkx1165>
- Romero, I. G., Pai, A. A., Tung, J., & Gilad, Y. (2014). RNA-seq: Impact of RNA degradation on transcript quantification. *BMC Biology*, *12*(42). <https://doi.org/10.1186/1741-7007-12-42>
- Saada, N., Dai, B., Echeteu, C., Sarna, S. K., & Palade, P. (2003). Smooth muscle uses another promoter to express primarily a form of human Cav1.2 L-type calcium channel different from the principal heart form. *Biochemical and Biophysical Research Communications*, *302*(1), 23–28. [https://doi.org/10.1016/s0006-291x\(03\)00097-4](https://doi.org/10.1016/s0006-291x(03)00097-4)
- Saada, N. I., Carrillo, E. D., Dai, B., Wang, W.-Z., Dettbarn, C., Sanchez, J., & Palade, P. (2005). Expression of multiple CaV1.2 transcripts in rat tissues mediated by different promoters. *Cell Calcium*, *37*(4), 301–309. <https://doi.org/10.1016/j.ceca.2004.11.003>
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., & Hume, D. A. (2007). Mammalian RNA polymerase II core promoters: Insights from genome-wide studies. *Nature Reviews Genetics*, *8*(6), 424–436. <https://doi.org/10.1038/nrg2026>
- Sanjabi, S., Williams, K., Sacconi, S., Zhou, L., Hoffmann, A., Ghosh, G., Gerondakis, S., Natoli, G., & Smale, S. T. (2005). A c-Rel subdomain responsible for enhanced DNA-binding affinity and selective gene activation. *Genes & Development*, *19*(18), 2138–2151. <https://doi.org/10.1101/gad.1329805>

- Scharinger, A., Eckrich, S., Vandael, D. H., Schönig, K., Koschak, A., Hecker, D., Kaur, G., Lee, A., Sah, A., Bartsch, D., Benedetti, B., Lieb, A., Schick, B., Singewald, N., Sinnegger-Brauns, M. J., Carbone, E., Engel, J., & Striessnig, J. (2015). Cell-type-specific tuning of Cav1.3 Ca<sup>2+</sup>-channels by a C-terminal automodulatory domain. *Frontiers in Cellular Neuroscience*, *9*. <https://doi.org/10.3389/fncel.2015.00309>
- Schlamp, K., Weinmann, A., Krupp, M., Maass, T., Galle, P., & Teufel, A. (2008). BlotBase: A northern blot database. *Gene*, *427*(1), 47–50. <https://doi.org/10.1016/j.gene.2008.08.026>
- Shamir, A., & Buonanno, A. (2010). Molecular and cellular characterization of Neuregulin-1 type IV isoforms. *Journal of Neurochemistry*, *113*(5), 1163–1176. <https://doi.org/10.1111/j.1471-4159.2010.06677.x>
- Shang, L. L., & Dudley Jr, S. C. (2005). Tandem promoters and developmentally regulated 5'- and 3'-mRNA untranslated regions of the mouse Scn5a cardiac sodium channel. *Journal of Biological Chemistry*, *280*(2), 933–940. <https://doi.org/10.1074/jbc.M409977200>
- Shephard, E. A., Chandan, P., Stevanovic-Walker, M., Edwards, M., & Phillips, I. R. (2007). Alternative promoters and repetitive DNA elements define the species-dependent tissue-specific expression of the FMO1 genes of human and mouse. *The Biochemistry Journal*, *406*(3), 491–499. <https://doi.org/10.1042/BJ20070523>
- Shi, H., Zhou, Y., Jia, E., Pan, M., Bai, Y., & Ge, Q. (2021). Bias in RNA-seq Library Preparation: Current Challenges and Solutions. *BioMed Research International*, *6647597*. <https://doi.org/10.1155/2021/6647597>
- Shields, J. M., & Yang, V. W. (1997). Two Potent Nuclear Localization Signals in the Gut-enriched Krüppel-like Factor Define a Subfamily of Closely Related Krüppel Proteins\*.

*Journal of Biological Chemistry*, 272(29), 18504–18507.

<https://doi.org/10.1074/jbc.272.29.18504>

Shiraki, T., Kondo, S., Katayama, S., Waki, K., Kasukawa, T., Kawaji, H., Kodzius, R., Watahiki, A., Nakamura, M., Arakawa, T., Fukuda, S., Sasaki, D., Podhajska, A., Harbers, M., Kawai, J., Carninci, P., & Hayashizaki, Y. (2003). Cap analysis gene expression for high-throughput analysis of transcriptional starting point and identification of promoter usage. *Proceedings of the National Academy of Sciences*, 100(26), 15776–15781. <https://doi.org/10.1073/pnas.2136655100>

Shistik, E., Ivanina, T., Blumenstein, Y., & Dascal, N. (1998). Crucial Role of N Terminus in Function of Cardiac L-type Ca<sup>2+</sup> Channel and Its Modulation by Protein Kinase C\*. *Journal of Biological Chemistry*, 273(28), 17901–17909.

<https://doi.org/10.1074/jbc.273.28.17901>

Siepel, A., Bejerano, G., Pedersen, J. S., Hinrichs, A. S., Hou, M., Rosenbloom, K., Clawson, H., Spieth, J., Hillier, L. W., Richards, S., Weinstock, G. M., Wilson, R. K., Gibbs, R. A., Kent, W. J., Miller, W., & Haussler, D. (2005). Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Research*, 15(8), 1034–1050. <https://doi.org/10.1101/gr.3715005>

Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., & Higgins, D. G. (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology*, 7(1), 539. <https://doi.org/10.1038/msb.2011.75>



- Simms, B. A., & Zamponi, G. W. (2014). Neuronal voltage-gated calcium channels: Structure, function, and dysfunction. *Neuron*, 82(1), 24–45.  
<https://doi.org/10.1016/j.neuron.2014.03.016>
- Singh, A., Gebhart, M., Fritsch, R., Sinnegger-Brauns, M. J., Poggiani, C., Hoda, J.-C., Engel, J., Romanin, C., Striessnig, J., & Koschak, A. (2008). Modulation of Voltage- and Ca<sup>2+</sup>-dependent Gating of CaV1.3 L-type Calcium Channels by Alternative Splicing of a C-terminal Regulatory Domain \*. *Journal of Biological Chemistry*, 283(30), 20733–20744.  
<https://doi.org/10.1074/jbc.M802254200>
- Singh, A., Hamedinger, D., Hoda, J.-C., Gebhart, M., Koschak, A., Romanin, C., & Striessnig, J. (2006). C-terminal modulator controls Ca<sup>2+</sup>-dependent gating of Cav1.4 L-type Ca<sup>2+</sup> channels. *Nature Neuroscience*, 9(9), Article 9. <https://doi.org/10.1038/nn1751>
- Smale, S. T., & Kadonaga, J. T. (2003). The RNA polymerase II core promoter. *Annual Review of Biochemistry*, 72, 449–479.  
<https://doi.org/10.1146/annurev.biochem.72.121801.161520>
- Smith, C. W. J. (2008). Alternative Splicing in the Postgenomic Era, edited by Benjamin J. Blencowe and Brenton R. Graveley. 2007, XXIV, Springer, New York. ISBN: 978-0-387-77373-5. *RNA*, 14(12), 2460–2461. <https://doi.org/10.1261/rna.1340908>
- Sofi, M. Y., Shafi, A., & Masoodi, K. Z. (2022). Chapter 6—Multiple sequence alignment. In M. Y. Sofi, A. Shafi, & K. Z. Masoodi (Eds.), *Bioinformatics for Everyone* (pp. 47–53). Academic Press. <https://doi.org/10.1016/B978-0-323-91128-3.00011-2>
- Specht, D., Wu, S.-B., Turner, P., Dearden, P., Koentgen, F., Wolfrum, U., Maw, M., Brandstatter, J. H., & tomDieck, S. (2009). Effects of Presynaptic Mutations on a Postsynaptic Calcium Channel Colocalized with mGluR6 at Mouse



- [rg/10.1016/j.gene.2004.07.029](https://doi.org/10.1016/j.gene.2004.07.029)<https://doi.org/10.1016/j.gene.2004.07.029><https://doi.org/10.1016/j.gene.2004.07.029>
- Steinthorsdottir, V., Stefansson, H., Ghosh, S., Birgisdottir, B., Bjornsdottir, S., Fasquel, A. C., Olafsson, O., Stefansson, K., & Gulcher, J. R. (2004b). MultiphenoveltranscriptioninitiationsitesforNRG1. *Gene*, *342*(1), 97–105. <https://doi.org/10.1016/j.gene.2004.07.029>
- Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., & Friess, H. (2009). Northern blot analysis for detection and quantification of RNA in pancreatic cancer cells and tissues. *Nature Protocols*, *4*(1), 37–43. <https://doi.org/10.1038/nprot.2008.216>
- Striessnig, J., Ortner, N. J., & Pinggera, A. (2015). Pharmacology of L-type Calcium Channels: Novel Drugs for Old Targets? *Current Molecular Pharmacology*, *8*(2), 110–122. <https://doi.org/10.2174/1874467208666150507105845>
- Striessnig, J., Pinggera, A., Kaur, G., Bock, G., & Tuluc, P. (2014). L-type Ca<sup>2+</sup> channels in heart and brain. *Wiley Interdisciplinary Reviews. Membrane Transport and Signaling*, *3*(2), 15–38. <https://doi.org/10.1002/wmts.102>
- Sultana, N., Dienes, B., Benedetti, A., Tuluc, P., Szentesi, P., Sztretye, M., Rainer, J., Hess, M. W., Schwarzer, C., Obermair, G. J., Csernoch, L., & Flucher, B. E. (2016). Restricting calcium currents is required for correct fiber type specification in skeletal muscle. *Development (Cambridge, England)*, *143*(9), 1547–1559. <https://doi.org/10.1242/dev.129676>
- Sun, H., & Yu, G. (2019). New insights into the pathogenicity of non-synonymous variants through multi-level analysis. *Scientific Reports*, *9*(1), Article 1. <https://doi.org/10.1038/s41598-018-38189-9>

- Tan, B. Z., Jiang, F., Tan, M. Y., Yu, D., Huang, H., Shen, Y., & Soong, T. W. (2011). Functional characterization of alternative splicing in the C terminus of L-type CaV1.3 channels. *The Journal of Biological Chemistry*, 286(49), 42725–42735. <https://doi.org/10.1074/jbc.M111.265207>
- Tan, G. M. Y., Yu, D., Wang, J., & Soong, T. W. (2012). Alternative Splicing at C Terminus of CaV1.4 Calcium Channel Modulates Calcium-dependent Inactivation, Activation Potential, and Current Density. *The Journal of Biological Chemistry*, 287(2), 832–847. <https://doi.org/10.1074/jbc.M111.268722>
- Tan, W., Wang, Y., Gold, B., Chen, J., Dean, M., Harrison, P. J., Weinberger, D. R., & Law, A. J. (2007). Molecular cloning of a brain-specific, developmentally regulated neuregulin 1 (NRG1) isoform and identification of a functional promoter variant associated with schizophrenia. *Journal of Biological Chemistry*, 282(33), 24343–24351. <https://doi.org/10.1074/jbc.M702953200>
- Tang, Z. Z., Liang, M. C., Lu, S., Yu, D., Yu, C. Y., Yue, D. T., & Soong, T. W. (2004). Transcript scanning reveals novel and extensive splice variations in human l-type voltage-gated calcium channel, Cav1.2 alpha1 subunit. *The Journal of Biological Chemistry*, 279(43), 44335–44343. <https://doi.org/10.1074/jbc.M407023200>
- The ENCODE Project Consortium. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature*, 447, 799–816. <https://doi.org/10.1038/nature05874>
- The ENCODE Project Consortium, Moore, J. E., Purcaro, M. J., Pratt, H. E., Epstein, C. B., Shores, N., Adrian, J., Kawli, T., Davis, C. A., Dobin, A., Kaul, R., Halow, J., Van Nostrand, E. L., Freese, P., Gorkin, D. U., Shen, Y., He, Y., Machkiewicz, M., Pauli-

- Behn, F., ... Weng, Z. (2020). Expanded encyclopaedias of DNA elements in the human and mouse genomes. *Nature*, *583*, 699–710. <https://doi.org/10.1038/s41586-020-2493-4>
- The GTEx Consortium. (2020). The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*, *369*(6509), 1318–1330. <https://doi.org/DOI:10.1126/science.aaz1776>
- The UniProt Consortium. (2021). UniProt: The universal protein knowledgebase in 2021. *Nucleic Acids Research*, *49*(D1), D480–D489. <https://doi.org/10.1093/nar/gkaa1100>
- Touriol, C., Bornes, S., Bonnal, S., Audigier, S., Prats, H., Prats, A.-C., & Vagner, S. (2003). Generation of protein isoform diversity by alternative initiation of translation at non-AUG codons. *Biology of the Cell*, *95*(3–4), 169–178. [https://doi.org/10.1016/S0248-4900\(03\)00033-9](https://doi.org/10.1016/S0248-4900(03)00033-9)
- Tuluc, P., Benedetti, B., Coste de Bagneaux, P., Grabner, M., & Flucher, B. E. (2016). Two distinct voltage-sensing domains control voltage sensitivity and kinetics of current activation in CaV1.1 calcium channels. *The Journal of General Physiology*, *147*(6), 437–449. <https://doi.org/10.1085/jgp.201611568>
- Tuluc, P., Molenda, N., Schlick, B., Obermair, G. J., Flucher, B. E., & Jurkat-Rott, K. (2009). A CaV1.1 Ca<sup>2+</sup> channel splice variant with high conductance and voltage-sensitivity alters EC coupling in developing skeletal muscle. *Biophysical Journal*, *96*(1), 35–44. <https://doi.org/10.1016/j.bpj.2008.09.027>
- Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R., & Leunissen, J. A. M. (2007). Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research*, *35*(1). <https://doi.org/10.1093/nar/gkm306>

- van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends in Genetics: TIG*, 34(9), 666–681.  
<https://doi.org/10.1016/j.tig.2018.05.008>
- Wahl, M. C., Will, C. L., & Lührmann, R. (2009). The spliceosome: Design principles of a dynamic RNP machine. *Cell*, 136(4), 701–718. <https://doi.org/10.1016/j.cell.2009.02.009>
- Wahl-Schott, C., Baumann, L., Cuny, H., Eckert, C., Griessmeier, K., & Biel, M. (2006). Switching off calcium-dependent inactivation in L-type calcium channels by an autoinhibitory domain. *Proceedings of the National Academy of Sciences of the United States of America*, 103(42), 15657–15662. <https://doi.org/10.1073/pnas.0604621103>
- Wang, Y., Liu, J., Huang, B., Xu, Y.-M., Li, J., Huang, L.-F., Lin, J., Zhang, J., Min, Q.-H., Yang, W.-M., & Wang, X.-Z. (2015). Mechanism of alternative splicing and its regulation (Review). *Biomedical Reports*, 3(2), 152–158.  
<https://doi.org/10.3892/br.2014.407>
- Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. *Nature Biotechnology*, 39, 1348–1365.
- Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: A revolutionary tool for transcriptomics. *Nature Reviews Genetics*, 10(1), 57–63. <https://doi.org/10.1038/nrg2484>
- Warm, D., Schroer, J., & Sinning, A. (2022). Gabaergic Interneurons in Early Brain Development: Conducting and Orchestrated by Cortical Network Activity. *Frontiers in Molecular Neuroscience*, 14, 807969. <https://doi.org/10.3389/fnmol.2021.807969>
- Waskom, M. L. (2021). seaborn: Statistical data visualization. *The Journal of Open Source Software*, 6(60). <https://doi.org/10.21105/joss.03021>

- Wei, X., Neely, A., Lacerda, A. E., Olcese, R., Stefani, E., Perez-Reyes, E., & Birnbaumer, L. (1994). Modification of Ca<sup>2+</sup> channel activity by deletions at the carboxyl terminus of the cardiac alpha 1 subunit. *The Journal of Biological Chemistry*, *269*(3), 1635–1640.
- Wei, X., Neely, A., Olcese, R., Lang, W., Stefani, E., & Birnbaumer, L. (1996). Increase in Ca<sup>2+</sup> channel expression by deletions at the amino terminus of the cardiac alpha 1C subunit. *Receptors & Channels*, *4*(4), 205–215.
- Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., & Smith, C. W. J. (2004). Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Molecular Cell*, *13*(1), 91–100. [https://doi.org/10.1016/s1097-2765\(03\)00502-1](https://doi.org/10.1016/s1097-2765(03)00502-1)
- Wong, H., Sakurai, T., Oyama, F., Kaneko, K., Wada, K., Miyazaki, H., Kurosawa, M., De Strooper, B., Saftig, P., & Nukina, N. (2005). Beta Subunits of voltage-gated sodium channels are novel substrates of beta-site amyloid precursor protein-cleaving enzyme (BACE1) and gamma-secretase. *Journal of Biological Chemistry*, *280*(24), 23009–23017. <https://doi.org/10.1074/jbc.M414648200>
- Wright, D. J., Hall, N. A. L., Irish, N., Man, A. L., Glynn, W., Mould, A., De Los Angeles, A., Angiolini, E., Swarbreck, D., Gharbi, K., Tunbridge, E. M., & Haerty, W. (2022). Long read sequencing reveals novel isoforms and insights into splicing regulation during cell state changes. *BMC Genomics*, *23*(42). <https://doi.org/10.1186/s12864-021-08261-2>
- Xu, C., Park, J.-K., & Zhang, J. (2019). Evidence that alternative transcriptional initiation is largely nonadaptive. *PLOS Biology*, *17*(3), e3000197. <https://doi.org/10.1371/journal.pbio.3000197>

- Yang, T., Zhang, M., & Zhang, N. (2022). Modified Northern blot protocol for easy detection of mRNAs in total RNA using radiolabeled probes. *BMC Genomics*, *23*, 66.  
<https://doi.org/10.1186/s12864-021-08275-w>
- Yang, Y., Yu, Z., Geng, J., Liu, M., Liu, N., Li, P., Hong, W., Yue, S., Jiang, H., Ge, H., Qian, F., Xiong, W., Wang, P., Song, S., Li, X., Fan, Y., & Liu, X. (2022). Cytosolic peptides encoding CaV1 C-termini downregulate the calcium channel activity-neuritogenesis coupling. *Communications Biology*, *5*(484). <https://doi.org/10.1038/s42003-022-03438-1>
- Zamponi, G. W. (2016). Targeting voltage-gated calcium channels in neurological and psychiatric diseases. *Nature Reviews Drug Discovery*, *15*(1), 19–34.  
<https://doi.org/10.1038/nrd.2015.5>
- Zamponi, G. W., Striessnig, J., Koschak, A., & Dolphin, A. C. (2015). The Physiology, Pathology, and Pharmacology of Voltage-Gated Calcium Channels and Their Future Therapeutic Potential. *Pharmacological Reviews*, *67*(4), 821–870.  
<https://doi.org/10.1124/pr.114.009654>
- Zoonomia Consortium. (2020). A comparative genomics multitool for scientific discovery and conservation. *Nature*, *587*, 240–245. <https://doi.org/10.1038/s41586-020-2876-6>
- Zuccotti, A., Clementi, S., Reinbothe, T., Torrente, A., Vandael, D. H., & Pirone, A. (2011). Structural and functional differences between L-type calcium channels: Crucial issues for future selective targeting. *Trends in Pharmacological Sciences*, *32*(6), 366–375.  
<https://doi.org/10.1016/j.tips.2011.02.012>

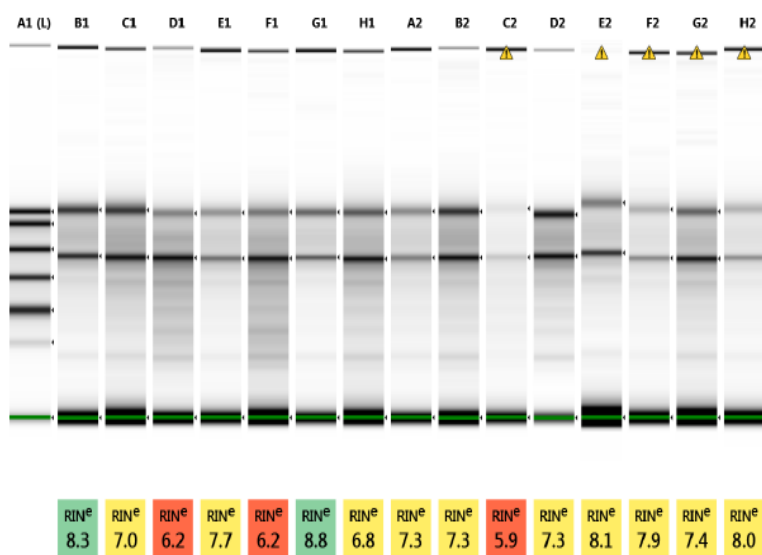


## Appendix

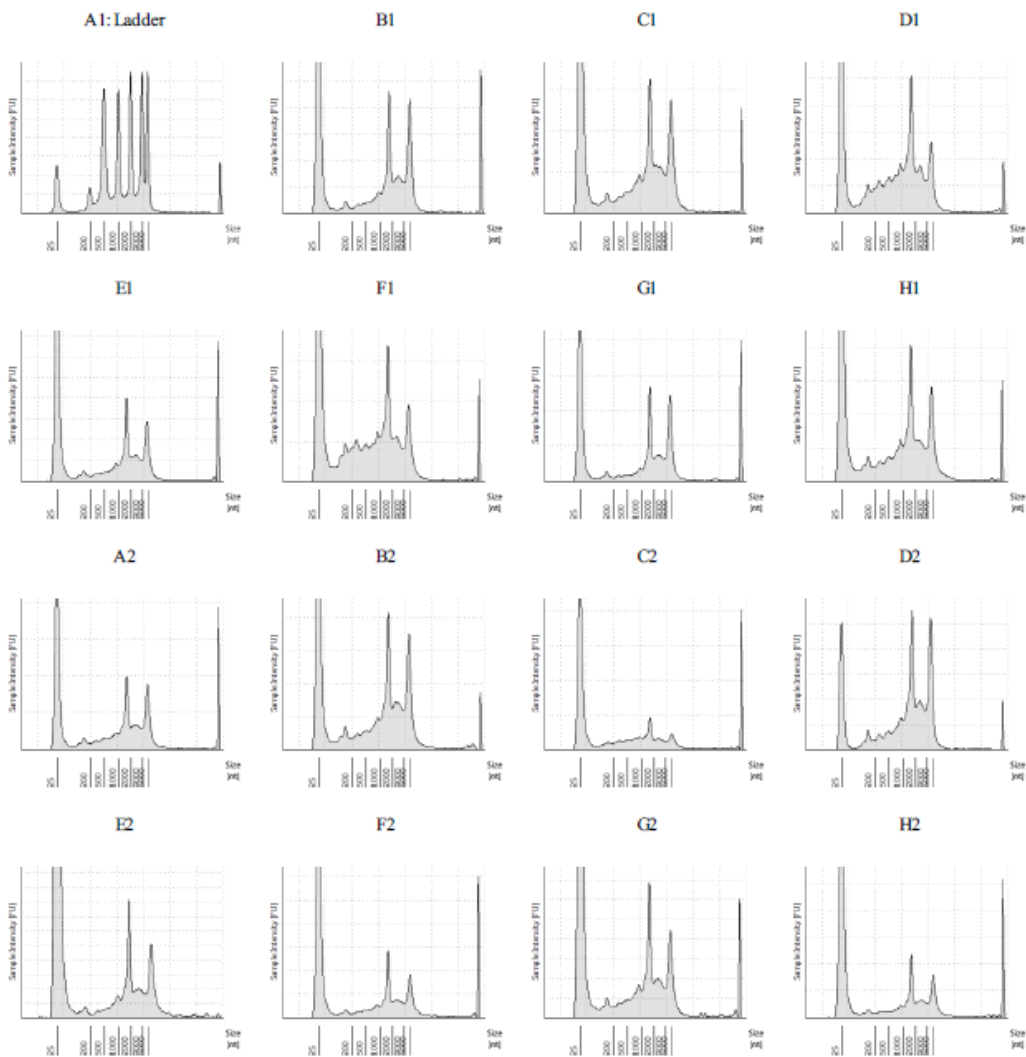
### Appendix 1. Tapestation of RNA samples that shows quality of RNA samples

Filename: 2018-03-21 - 17.00.10.RNA

#### Gel Images



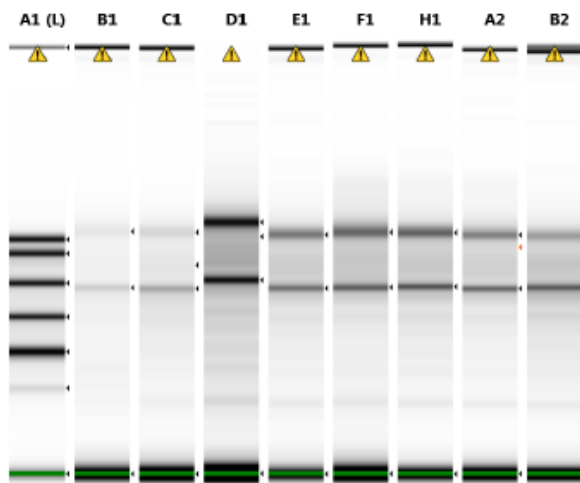
Default image (Contrast 50%), Image is Scaled to Sample, Image is Scaled to view larger Molecular Weight range



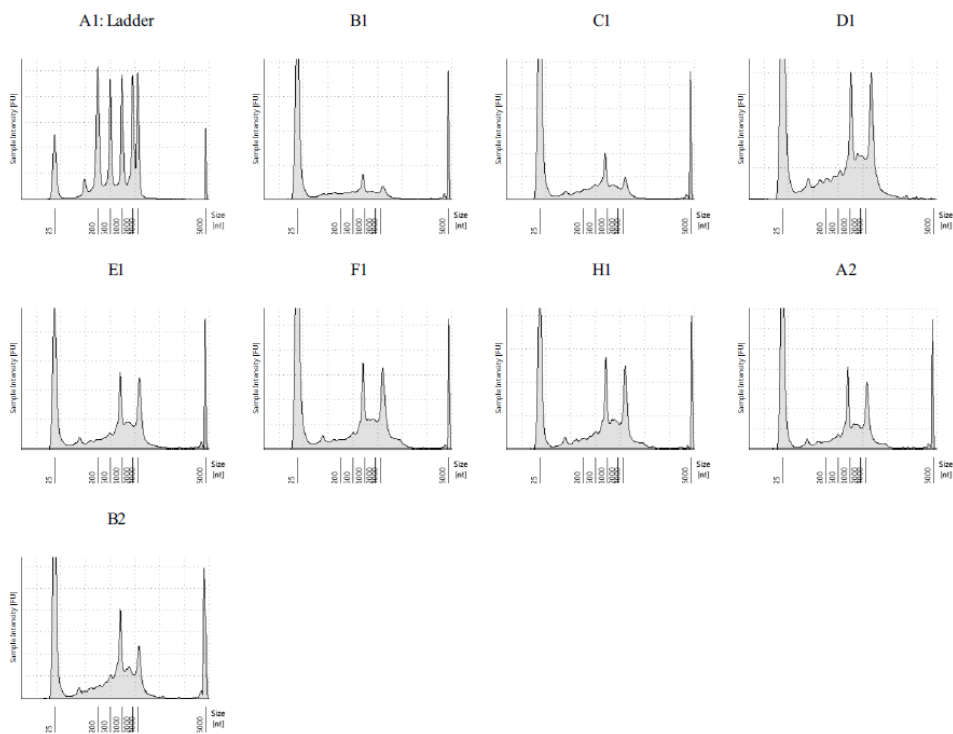
A1	Ladder	A2	5298 DL-PFC
B1	5244 striatum	B2	5298 cingulate
C1	5244 thalamus	C2	5346 parietal cortex
D1	5346 striatum	D2	5346 superior temporal cortex
E1	5298 parietal cortex	E2	5244 occipital cortex
F1	5346 cingulate	F2	5244 DL-PFC
G1	5298 striatum	G2	5298 thalamus
H1	5244 cingulate	H2	5298 superior temporal cortex

Filename: 2018-03-21 - 17.25.34.RNA

## Gel Images




RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>
6.6	6.5	7.5	7.8	8.0	8.0	8.2	7.2	

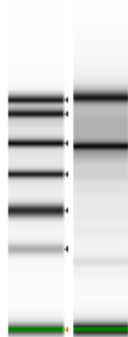


A1	Ladder	A2	5346 cerebellum
B1	5346 DL-PFC	B2	5346 thalamus
C1	5346 occipital cortex		
D1	5244 superior temporal cortex		
E1	5298 occipital cortex		
F1	5244 cerebellum		
H1	5298 cerebellum		

Filename: 2018-03-26 - 10.37.03.RNA

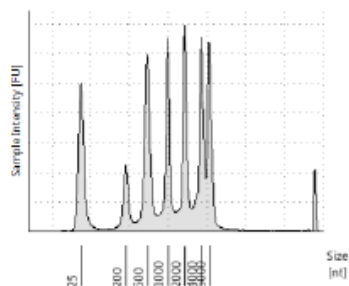
## Gel Images

A1 (L) B1  


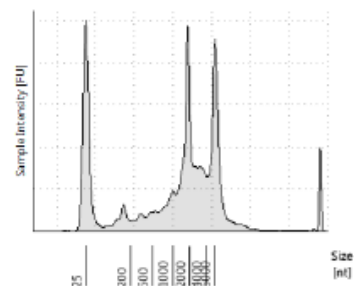


RIN<sup>e</sup>  
7.9

A1: Ladder



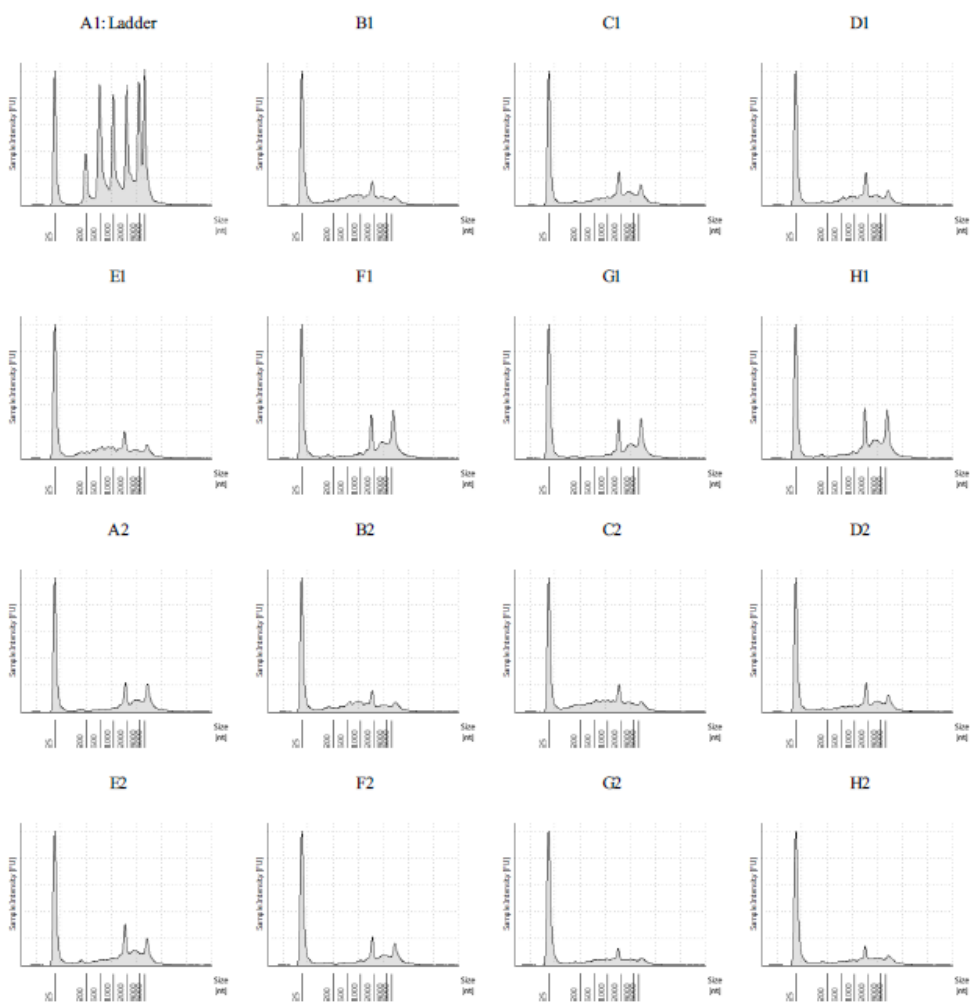
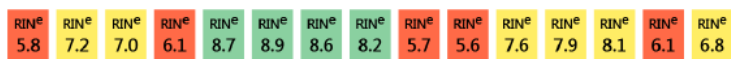
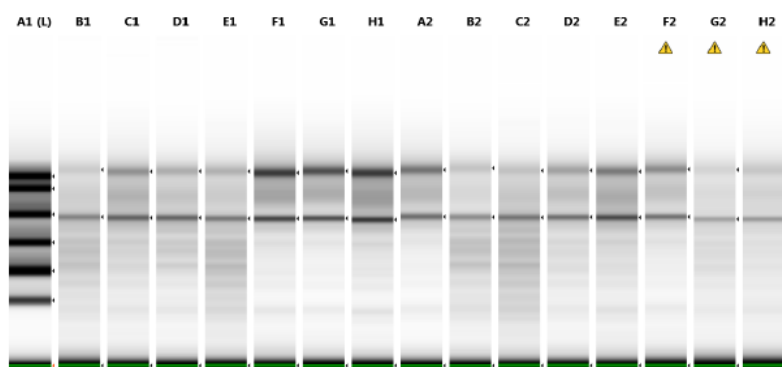
B1



A1	Ladder	B1	5244 parietal cortex
----	--------	----	----------------------

Filename: 2018-04-06 - 11.21.19.RNA

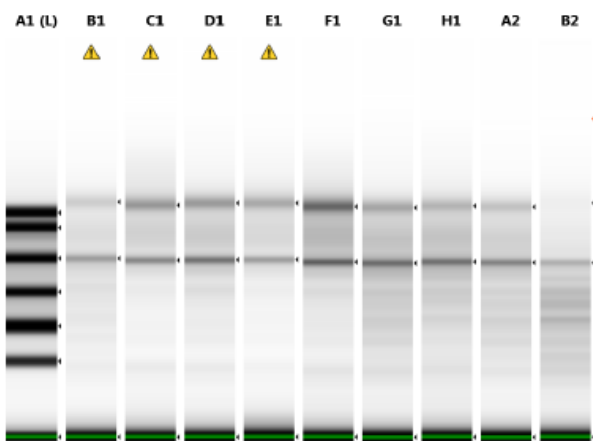
Gel Image



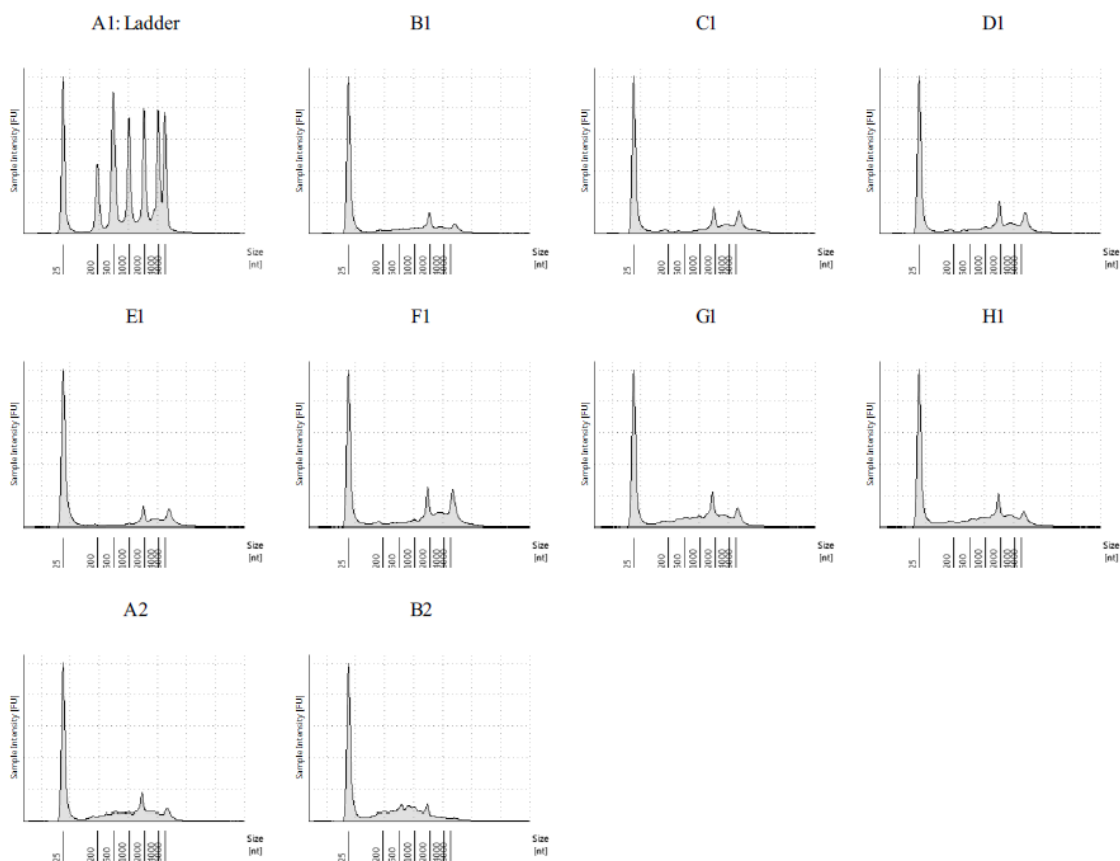
A1	Ladder	A2	5238 cingulate
B1	5717 thalamus	B2	5579 cingulate
C1	5717 parietal cortex	C2	5579 cerebellum
D1	5579 DL-PFC	D2	5717 superior temporal cortex
E1	5579 striatum	E2	5717 DL-PFC
F1	5238 superior temporal cortex	F2	5579 parietal cortex
G1	5238 DL-PFC	G2	5579 parietal cortex
H1	5238 parietal cortex	H2	5717 cerebellum

Filename: 2018-04-09 - 10.00.57.RNA

## Gel Image



RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>	RIN <sup>e</sup>
6.9	8.3	7.9	8.1	8.0	6.6	6.5	6.6	6.6	4.2



A1	Ladder	A2	5717 striatum
B1	5717 cingulate	B2	5717 occipital cortex
C1	5238 cerebellum		
D1	5238 thalamus		
E1	5238 striatum		
F1	5238 occipital cortex		
G1	5579 thalamus		
H1	5579 superior-temporal cortex		



**Appendix 2 RNA concentration**

Sample name (number)	RNA concentration	Sample name (number)	RNA concentration	Sample name (number)	RNA concentration
5244 striatum (1)	232 ng/ $\mu$ l	5244 thalamus (2)	310 ng/ $\mu$ l	5717 thalamus (3)	472 ng/ $\mu$ l
5346 striatum (4)	108 ng/ $\mu$ l	5298 parietal cortex (5)	270 ng/ $\mu$ l	5717 parietal cortex (6)	278 ng/ $\mu$ l
5346 cingulate (7)	186 ng/ $\mu$ l	5298 striatum (8)	832 ng/ $\mu$ l	5579 DL-PFC (9)	488 ng/ $\mu$ l
5579 striatum (10)	484 ng/ $\mu$ l	5238 superior temporal cortex (11)	174 ng/ $\mu$ l	5238 DL-PFC (12)	276 ng/ $\mu$ l
5244 parietal cortex (13)	136 ng/ $\mu$ l	5244 cingulate (14)	103 ng/ $\mu$ l	5238 parietal cortex (15)	368 ng/ $\mu$ l
5238 cingulate (16)	208 ng/ $\mu$ l	5298 DL-PFC (17)	640 ng/ $\mu$ l	5298 cingulate (18)	282 ng/ $\mu$ l
5579 cingulate (19)	188 ng/ $\mu$ l	5579 cerebellum (20)	228 ng/ $\mu$ l	5717 superior temporal cortex (21)	296 ng/ $\mu$ l
5717 DL-PFC (22)	510 ng/ $\mu$ l	5346 parietal cortex (23)	192 ng/ $\mu$ l	5346 superior temporal cortex (24)	430 ng/ $\mu$ l
5244 occipital cortex (25)	338 ng/ $\mu$ l	5244 DL-PFC (26)	340 ng/ $\mu$ l	5298 thalamus (27)	1180 ng/ $\mu$ l
5298 superior-temporal cortex (28)	574 ng/ $\mu$ l	5346 DL-PFC (29)	676 ng/ $\mu$ l	5346 occipital cortex (30)	332 ng/ $\mu$ l
5579 occipital cortex (31)	830 ng/ $\mu$ l	5579 parietal cortex (32)	600 ng/ $\mu$ l	5717 cerebellum (33)	686 ng/ $\mu$ l
5717 cingulate (34)	608 ng/ $\mu$ l	5238 cerebellum (35)	494 ng/ $\mu$ l	5238 thalamus (36)	604 ng/ $\mu$ l
5238 striatum (37)	260 ng/ $\mu$ l	5238 occipital cortex (38)	270 ng/ $\mu$ l	5244 superior temporal cortex (39)	372 ng/ $\mu$ l
5244 cerebellum (40)	314 ng/ $\mu$ l	5298 occipital cortex (41)	368 ng/ $\mu$ l	5298 cerebellum (42)	206 ng/ $\mu$ l
5346 thalamus (43)	534 ng/ $\mu$ l	5346 cerebellum (44)	286 ng/ $\mu$ l	5579 thalamus (45)	390 ng/ $\mu$ l
5579 superior-temporal cortex (46)	342 ng/ $\mu$ l	5717 striatum (47)	584 ng/ $\mu$ l	5717 occipital cortex (48)	314 ng/ $\mu$ l

**Appendix 3. The amount of input to the long-read nanopore sequencing based on the concentration of the short-length transcript from the 3' region of *CACNA1F* in each sample. Each sample was pooled to the final concentration of 1000ng / 47µl. Each number corresponds to the sample number in RNA extraction (Appendix 1).**

Sample name	Input (concentration)	Sample name	Input (concentration)	Sample name	Input (concentration)
TF1	25.19 ng (1.48 ng/µl)	TF2	19.21 ng (1.13 ng/µl)	TF3	18.36 ng (1.08 ng/µl)
TF4	25.33 ng (1.49 ng/µl)	TF5	56.52 ng (3.14 ng/µl)	TF6	24.14 ng (1.42 ng/µl)
TF7	40.28 ng (2.12 ng/µl)	TF8	56.52 ng (3.14 ng/µl)	TF9	38.38 ng (2.02 ng/µl)
TF10	17.73 ng (0.985 ng/µl)	TF11	132.24 ng (6.96 ng/µl)	TF12	121.98 ng (6.42 ng/µl)
TF13	113.63 ng (5.98 ng/µl)	TF14	75.24 ng (3.96 ng/µl)	TF15	119.32 ng (6.28 ng/µl)
TF16	80.94ng (4.26 ng/µl)	TF17	137.56 ng (7.24 ng/µl)	TF18	62.7 ng (3.3 ng/µl)
TF19	96.9ng (5.1 ng/µl)	TF20	22.32 ng (1.24 ng/µl)	TF21	96.9 ng (5.1 ng/µl)
TF22	5168ng (272 ng/µl)	TF23	3838ng (202 ng/µl)	TF24	245.1 ng (12.9 ng/µl)
TF25	4046ng (238 ng/µl)	TF26	4932ng (274 ng/µl)	TF27	3914ng (206 ng/µl)
TF28	4370ng (230 ng/µl)	TF29	3287ng (173 ng/µl)	TF30	4332ng (228 ng/µl)
TF31	4560ng (240 ng/µl)	TF32	2698ng (142 ng/µl)	TF33	3515ng (185 ng/µl)
TF34	3325ng (175 ng/µl)	TF35	4370ng (230 ng/µl)	TF36	2318ng (122 ng/µl)
TF37	3990ng (210 ng/µl)	TF38	3363ng (177 ng/µl)	TF39	3800ng (200 ng/µl)
TF40	3325ng (175 ng/µl)	TF41	4370ng (230 ng/µl)	TF42	2088ng (116 ng/µl)
TF43	2679ng (141 ng/µl)	TF44	2166ng (114 ng/µl)	TF45	3838ng (202 ng/µl)
TF46	3952ng (208 ng/µl)	TF47	2546ng (208 ng/µl)	TF48	1710ng (90 ng/µl)

**Appendix 4. Amount of input into the long-read nanopore sequencing based on concentration of the short-length transcripts from the 3' region of *CACNAIS* in each sample. Each sample was pooled to the final concentration of 1000ng / 47 $\mu$ l. Samples with too low concentration were not pooled. Each number corresponds to the sample number in RNA extraction (Appendix 1).**

Sample name	Input (concentration)	Sample name	Input (concentration)	Sample name	Input (concentration)
TS1	1472.4ng (81.8 ng/ $\mu$ l)	TS2	315ng (17.5 ng/ $\mu$ l)	TS3	354.6ng (19.7 ng/ $\mu$ l)
TS4	273.6ng (15.2 ng/ $\mu$ l)	TS5	Too low	TS6	325.8ng (18.1 ng/ $\mu$ l)
TS7	1101.6ng (61.2 ng/ $\mu$ l)	TS8	651.6ng (36.2 ng/ $\mu$ l)	TS9	237.6ng (13.2 ng/ $\mu$ l)
TS10	73.44ng (4.08 ng/ $\mu$ l)	TS11	774ng (43 ng/ $\mu$ l)	TS12	666ng (37 ng/ $\mu$ l)
TS13	579.6ng (32.2 ng/ $\mu$ l)	TS14	432ng (24 ng/ $\mu$ l)	TS15	180ng (10 ng/ $\mu$ l)
TS16	766.8ng (42.6 ng/ $\mu$ l)	TS17	198ng (11 ng/ $\mu$ l)	TS18	522ng (29 ng/ $\mu$ l)
TS19	453.6ng (25.2 ng/ $\mu$ l)	TS20	Too low	TS21	493.2ng (27.4 ng/ $\mu$ l)
TS22	604.8ng (33.6 ng/ $\mu$ l)	TS23	266.4ng (14.8 ng/ $\mu$ l)	TS24	716.4ng (39.8 ng/ $\mu$ l)
TS25	Too low	TS26	626.4ng (34.8 ng/ $\mu$ l)	TS27	Too low
TS28	45.36ng (2.52 ng/ $\mu$ l)	TS29	117.36ng (6.52 ng/ $\mu$ l)	TS30	120.24ng (6.68 ng/ $\mu$ l)
TS31	102.6ng (5.7 ng/ $\mu$ l)	TS32	217.8ng (12.1 ng/ $\mu$ l)	TS33	Too low
TS34	63.72ng (3.54 ng/ $\mu$ l)	TS35	Too low	TS36	79.56ng (4.42 ng/ $\mu$ l)
TS37	385.2ng (21.4 ng/ $\mu$ l)	TS38	471.6ng (26.2 ng/ $\mu$ l)	TS39	1472.4ng (81.8 ng/ $\mu$ l)
TS40	57.24ng (3.18 ng/ $\mu$ l)	TS41	52.56ng (2.92 ng/ $\mu$ l)	TS42	Too low
TS43	73.8ng (4.1 ng/ $\mu$ l)	TS44	248.4ng (13.8 ng/ $\mu$ l)	TS45	777.6ng (43.2 ng/ $\mu$ l)
TS46	468ng (26 ng/ $\mu$ l)	TS47	576ng (32 ng/ $\mu$ l)	TS48	Too low

**Appendix 5. PCA plot script**

```

####Import programs on Python##
import sklearn
import pandas as pd
import matplotlib.pyplot as plt
from sklearn.preprocessing import StandardScaler
from sklearn.decomposition import PCA
import numpy as np
from matplotlib.patches import Ellipse
import matplotlib.transforms as transforms

### PCA plot of long-read nanopore sequencing data of the short-length transcripts from the 3'
region of CACNA1S###
## read data file in csv ##
df = pd.read_csv("CACNA1S_tp.csv", sep=",", index_col=0)
df = df.reset_index(drop=True)

print(df.head(5))
loadfile = "PCA_loadings.csv"

x = pd.read_csv("CACNA1S_tp.csv", sep=",")
x = df.iloc[:, 3:]
print(x.head(5))
x = StandardScaler().fit_transform(x)

## Create data frame ##
dataset1 = df['Tissue']
dataset1 = pd.DataFrame(data=dataset1, columns = ['Tissue'])

pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)

transcripts = df.columns[3:]

loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=transcripts)
loadings.to_csv(loadfile)

principalDf = pd.DataFrame(data = principalComponents, columns = ['PC1', 'PC2'])
TissueDf = pd.concat([principalDf, dataset1['Tissue']], axis = 1)
TissueDf.to_csv("principalDf.csv")

## PCA ##

principalComponents = principalComponents.T
print(principalComponents[0])
print("The PCA explained variance ratio is:")

```

```

print(pca.explained_variance_ratio_)
fig = plt.figure(figsize = (10,10))
ax = fig.add_subplot(1,1,1)
PC1 = "PC1: " + str(pca.explained_variance_ratio_[0] * 100)[:4] + "% "
ax.set_xlabel(PC1, fontsize = 15)
PC2 = "PC2: " + str(pca.explained_variance_ratio_[1] * 100)[:4] + "% "
ax.set_ylabel(PC2, fontsize = 15)
ax.set_title("CACNA1S", fontsize = 20)
ax.scatter(principalComponents[0][0:7], principalComponents[1][0:7], c='r', alpha=0.5, s = 50)
ax.scatter(principalComponents[0][7:14], principalComponents[1][7:14], c='g', alpha=0.5, s =
50)
ax.scatter(principalComponents[0][14:19], principalComponents[1][14:19], c= 'y',alpha=0.5, s =
50)
ax.scatter(principalComponents[0][19:27], principalComponents[1][19:27], c='b',alpha=0.5, s =
50)
ax.scatter(principalComponents[0][17:25], principalComponents[1][17:25], c='c', alpha=0.5, s =
50)
ax.scatter(principalComponents[0][25:32], principalComponents[1][25:32], c='m', alpha=0.5, s =
50)

ax.legend(['5238', '5244', '5298', '5346', '5579', '5717'])
ax.grid()
plt.savefig("CACNA1S_in_")

#### PCA plot of long-read nanopore sequencing data of the short-length transcripts from the 3'
region of CACNA1S####

## read data file in csv ##
df = pd.read_csv("CACNA1F_individuals.csv", sep=",", index_col=0)
df = df.reset_index(drop=True)

print(df.head(5))
loadfile = "PCA_loadings.csv"

x = pd.read_csv("CACNA1F_individuals.csv", sep=",")
x = df.iloc[:, 3:]
print(x.head(5))

x = StandardScaler().fit_transform(x)

## Create data frame ##
dataset1 = df['Tissue']
dataset1 = pd.DataFrame(data=dataset1, columns = ['Tissue'])

pca = PCA(n_components=2)
principalComponents = pca.fit_transform(x)

```

```

transcripts = df.columns[3:]

loadings = pd.DataFrame(pca.components_.T, columns=['PC1', 'PC2'], index=transcripts)
loadings.to_csv(loadfile)

principalDf = pd.DataFrame(data = principalComponents, columns = ['PC1', 'PC2'])
TissueDf = pd.concat([principalDf, dataset1['Tissue']], axis = 1)
TissueDf.to_csv("principalDf.csv")

###PCA###

principalComponents = principalComponents.T
print(principalComponents[0])
print("The PCA explained variance ratio is:")
print(pca.explained_variance_ratio_)
fig = plt.figure(figsize = (10,10))
ax = fig.add_subplot(1,1,1)
PC1 = "PC1: " + str(pca.explained_variance_ratio_[0] * 100)[4] + "%"
ax.set_xlabel(PC1, fontsize = 15)
PC2 = "PC2: " + str(pca.explained_variance_ratio_[1] * 100)[4] + "%"
ax.set_ylabel(PC2, fontsize = 15)
ax.set_title("CACNA1F", fontsize = 20)
ax.scatter(principalComponents[0][0:8], principalComponents[1][0:8], alpha=0.5, c='r', s = 50)
ax.scatter(principalComponents[0][8:14], principalComponents[1][8:14], alpha=0.5, c='g', s = 50)
ax.scatter(principalComponents[0][14:22], principalComponents[1][14:22], alpha=0.5, c= 'y', s = 50)
ax.scatter(principalComponents[0][22:30], principalComponents[1][22:30], alpha=0.5, c='b', s = 50)
ax.scatter(principalComponents[0][30:38], principalComponents[1][30:38], alpha=0.5, c='c', s = 50)
ax.scatter(principalComponents[0][38:46], principalComponents[1][38:46], alpha=0.5, c='m', s = 50)

ax.legend(['5238', '5244', '5298', '5346', '5579', '5717'])
ax.grid()
plt.savefig("CACNA1F_individuals")

```

**Appendix 6. Protein sequence of highly expressed short-length transcripts from the 3' region of *CACNA1S* and *CACNA1F* in the human brain identified from long-read nanopore sequencing**

```
>sp|Q13698|CAC1S_HUMAN Voltage-dependent L-type calcium channel
subunit alpha-1S OS=Homo sapiens OX=9606 GN=CACNA1S PE=1 SV=4
MEPSSPQDEGLRKKQPKKPVEILPRPPRALFCLTLENPLRKACISIVEWKPFETIILLT
IFANCVALAVYLPMPEDDNNLNLGLEKLEYFFLIVFSIEAAMKIIAYGFLFHQDAYLRS
GWNVLDFTIVFLGVFTVILEQVNVIQSHTAPMSSKAGLDVKALRAFRVLRPLRLVSGVP
SLQVVLNSIFKAMPLPLFHIALLVLFMVIIYAIIGLELFGKGMHKTCYFIGTDIVATVENE
EPSPCARTGSGRRCTINGSECRGGWPGPNHGITHFDNFGFSMLTVYQCITMEGWTDVLYW
VND AIGNWPWIYFVTLILLGSFFILNLVVLGVLSGEFTKEREKAKSRGTFQKLREKQQLD
EDLRGYMSWITQGEVMDVEDFREGKLSLDEGGSDESLEYEIAGLNKIIQFIRHWRQWNRI
FRWKCHDIVKSKVFYWLVLIVALNTLSIASEHNNQPLWLTRLQDIANRVLLSLFTTEML
MKMYGLGLRQYFMSIFNRFDCFVVCSGILEILLVESGAMTPLGISVLR CIRLLRIFKITK
YWTSLSNLVASLLNSIRSIASLLLLLFLFIVIFALLGMQLFGGRYDFEDTEVRRSNFDFN
PQALISVFQVLTGEDWTSMMYNGIMAYGGPSYPGMLVCIYFIILFVCGNYILLNVFLAIA
VDNLAEAESL TSAQKAKAEKRRKMSKGLPKSEEEKSTMAKKLEQKPKGEGIPTAKL
KIDEFESNVNEVKDPYPSADFPGDDEEDEPEIPLSPRPRPLAELQLKEKAVPIPEASSFF
IFSPTNKIRVLCHRIVNATWFTNFILLFILLSSAALAAEDPIRADSMRNQILKHF DIGFT
SVFTVEIVLKMTTYGAFLHKGSFCRNYFNMLDLLVAVSLISMGLESSAISVVKILRVLR
VLRPLRAINRAKGLKHVVQCMFVAISTIGNIVLVTTLLQFMFACIGVQLFKGKFFRCTDL
SKMTEEECRGYYVYKDGDPMQIELRHREWHSDHFHFDNVLSAMMSLFTVSTFEGWPQLL
YKAIDSNAEDVGPIYNNRVEMAIFFIIYIILIAFFMMNIFVGFVIVTFQE QGETEYKNC
LDKNQRQCVQYALKARPLRCYIPKNPYQYQVWYIVTSSYFEYLMFALIMLNTICLGMQHY
NQSEQMNHISDILNVAFTIIFTLEMILKLMFAFKARGYFGDPWNVDFDLIVIGSIIDVILS
EIDTFLASSGGLYCLGGCGNVDPDESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTF
IKSFQALPYVALLIVMLFFIYAVIGMQMFGKIALVDGTQINRNNNFQTFPQAVLLLFRCA
TGEAWQEILLACSYGKLCDPESDYAPGEEYTCGTNFAYYYFISFYMLCAFLVINL FVAVI
MDFDYLTRDWSILGPHHLDEFKAIWAEYDPEAKGRIKHLDVVTLLRRIQPPLGFGKFCP
HRVACKRLVGMNMP LNSDGTVTFNATL FALVRTALKIKTEGNFEQANEELRAIIKKIWK
TSMKLLDQVIPIPGDDEVTVGKFYATFLIQEHFRKFMKRQEEYYGYRPKKDIVQIQAGLR
TIEEEAAPEICRTVSGDLAAEEELERAMVEAAMEEGIFRRTGGLFGQVDNFLERTNSLPP
VMANQRPLQFAE IEMEEMESPVFLEDFPQDPRTNPLARANTNNANANVAYGNSNHSNSHV
FSSVHYEREFPEETETPATRGRALGQPCRVLGPHSKPCVEMLKGLLTQRAMP RGQAPPAP
CQCPRV ESSMPEDRKSSTPGSLHEETPHSRSTRENTSRCSAPATALLIQKALVRGGLGTL
AADANFIMATGQALADACQMEPEEVEIMATELLKGREAPEGMASSLGCLNLGSSLGSLDQ
HQGSQETLIPPR L
```

>tr|B1ALM3|B1ALM3\_HUMAN Voltage-dependent L-type calcium channel subunit alpha OS=Homo sapiens OX=9606 GN=CACNA1S PE=1 SV=1  
MEPSSPQDEGLRKKQPKKPVEILPRPPRALFCLTLENPLRKACISIVEWKPFETIILLT  
IFANCVLAVYLPMPEDDNNLNGLGLEKLEYFFLIVFSIEAAMKIIAYGFLFHQDAYLRS  
GWNVLDFTIVFLGVFTVILEQVNVIQSHTAPMSSKGAGLDVKALRAFRVLRPLRLVSGVP  
SLQVVLNSIFKAMLPLFHIALLVLFMVIIYAIIGLELEFKGKMHKTCYFIGTDIVATVENE  
EPSPCARTGSGRRCTINGSECRGGWPGPNHGITHFDNFGFSMLTVYQCITMEGWTDVLYW  
VND AIGNEWPWIIYFVTLILLGSFFILNLVVLGVLSGEFTKEREKAKSRGTFQKLRKQQLD  
EDLRGYMSWITQGEVMDVEDDFREGKLSLDEGGSDTESLYEIAGLNKIIQFIRHWRQWNRI  
FRWKCHDIVKSKVFWLVILIVALNTLSIASEHHNQPLWLTRLQDIANRVLLSLFTTEML  
MKMYGLGLRQYFMSIFNRFDCFVVCSGILEILLVESGAMTPLGISVLR CIRLLRIFKIK  
YWTSLSNLVASLLNSIRSIASLLLLLFLFIVIFALLGMQLFGGRYDFEDTEVRRSNFDFN  
PQALISVFQVLTGEDWTSMMYNGIMAYGGPSYPGMLVCIYFIILFVCGNYILLNVFLAIA  
VDNLAEAEESL TSAQKAKAEKRRKMSKGLPKSSEEEKSTMAKKLEQKPKGEGIPTTAKL  
KIDEFESNVNEVKDPYPSADFPGDDEEDEPEIPLSPRPRPLAELQLKEKAVPIPEASSFF  
IFSPNTKIRVLRCHRVNATWFTNFILLFILLSSAALAAEDPIRADSMRNQILKHFDIGFT  
SVFTVEIVLKMTTYGAFLHKG SFCRNYFNMLDLLVAVSLISMGLESSAISVVKILRVLR  
VLRPLRAINRAKGLKHVVQCMFVAISTIGNIVLVTTLLQFMFACIGVQLFKGKFFRCTDL  
SKMTEEECRGYYYVYKDGDPMQIELRHREWHSDHFHFDNVLSAMMSLFTVSTFEGWPQLL  
YKAIDSNAEDVGPIYNNRVEMAIFFIIYIILIAFFMMNIFVGFVIVTFQE QGETEYKNCE  
LDKNQRQCVQYALKARPLRCYIPKNPYQYQVWYIVTSSYFEYLMFALIMLNTICLGMQHY  
NQSEQMNHISDILNVAFTIIFTLEMILKLMAFKARGYFGDPWNVDFDFLIVIGSII DVILS  
EIDDPDESARISSAFFRLFRVMRLIKLLSRAEGVRTLLWTFIKSFQALPYVALLIVMLFF  
IYAVIGMQMFGKIALVDGTQINRNNNFQTFPQAVLLLFRCATGEAWQEILLACSYGKLCD  
PESDYAPGEEYTCGTNFAYYYFISFYMLCAFLVINL FVAVIMDNFDYLRDWSILGPHHL  
DEFKAIWAEYDPEAKGR IKHLDVVTLLRRIQPPLGFGKFCPHRVACKRLVGMN MPLNSDG  
TVTFNATLFALVRTALKIKTEGNFEQANEELRAIIKKIWKRTSMKLLDQVIPIPIGDDEV  
VGKFYATFLIQEHFRKFMKRQEEYGYRPKKDIVQIQAGLRTIEEEAAPEICRTVSGDLA  
AEELERAMVEAAMEEGIFRRTGGLFGQVDNFLERTNSLPPV MANQRPLQFAEIEEMEEME  
SPVFLEDFPQDPRTNPLARANTNNANANVAYGNSNHSNSHVFSSVHYEREFPEETETPAT  
RGRALGQPCRVLGPHSKPCVEMLKGLLTQRAMPRGQAPPAPCQCPRV ESSMPEDRKSSTP  
GSLHEETPHSRSTRENTSRCSAPATALLIQKALVRGGLGTLAADANFIMATGQALADACQ  
MEPEEVEIMATELLKGREAPEGMASSLGCLNLGSSSLGSLDQHQGSQETLIPPRL



>sp|O60840|CAC1F\_HUMAN Voltage-dependent L-type calcium channel subunit alpha-1F OS=Homo sapiens OX=9606 GN=CACNA1F PE=1 SV=2  
MSESEGGKDTTPEPSPANGAGPGPEWGLCPGPPAVEGESSGASGLGTPKRRNQHSKHKT  
AVASAQRSPRALFCLTLANPLRRSCISIVEWKPFDILILLTIFANCVALGVYIPFPEDDS  
NTANHNLEQVEYVFLVIFTVETVLKIVAYGLVLHPSAYIRNGWNLLDFIIVVVGFLFSVLL  
EQGPGRPGDAPHTGGKPGGFVDKALRAFRVLRPLRLVSGVPSLHIVLNSIMKALVPLLHI  
ALLVLFVIIIIYAIIGLELFLGRMHKTCYFLGSDMEAEEDPSPCASSGSGRACTLNQTECR  
GRWPGPNGGITNFDNFFFAMLTVFQCVTMEGWTDVLYWMQDAMGYELPWVYFVSLVIFGS  
FFVLNLVLGVLSGEFSKEREKAKARGDFQKQREKQQMEEDLRGYLDWITQAEELDMEDPS  
ADDNLGSMAEGRAGHRPQLAELTNRRRGRRLRWFHSTRSTHSTSSHASLPASDTGSMTE  
TQGDEDEEEGALASCTRCLNKIMKTRVCRRLRRANRVLRARCRRAVKSNACYWAVLLLVF  
LNTLTIASEHHGQPVWLTQIQEYANKVLLCLFTVEMLLKLYGLGPSAYVSSFFNRFDCFV  
VCGGILETTLVEVGAMQPLGISVLRVLLRIFKVTRHWASLSNLVASLLNSMKSIASLL  
LLFLFIIFSLGMLQFVGGKFNFDQHTKRSTFDTFPQALLTVFQILTGEDWNVVMYDG  
IMAYGGPFFPGMLVCIYFIILFICGNYILLNVFLAIAVDNLSAGDAGTAKDKGGEKSNEK  
DLPQENEGLVPGVEKEEEEEGARREGADMEEEEEEEEEEEEEEEEEEGAGGVELLQEVVPKE  
KVVPIPEGSAFFCLSQTNPLRKGCHTLIHHVFTNLIILVFIILSSVSLAAEDPIRAHSFR  
NHILGYFDYAFTSIFTVEILLKMTVFGAFLHRGSFCRSWFNMLDLLVSVSLISFGIHSS  
AISVVKILRVLRLRPLRAINRAKGLKHVVQCVFAIRTIIGNIMIVTLLQFMFACIGVQ  
LFGKGFYCTDEAKHTPQECKGSFLVYPGDVSRPLVRERLWVNSDFNFDNVLSAMMALF  
TVSTFEGWPALLYKAIDAYAEDHGPIYNYRVEISVFFIVYIIIIAFFMMNIFVGFVIITF  
RAQGEQEYQNCELDKNQRQCVEYALKAQPLRRYIPKNPHQYRVWATVNSAAFEYLMFLLI  
LLNTVALAMQHYEQTAPFNYAMDILNMVFTGLFTIEMVLKIIAFKPKHYFTDAWNTFDAL  
IVVGSIVDIAVTEVNNGGHLGESSEDSRISITFFRLFRVMRLVKLLSKGEGIRTLLWTF  
IKSFQALPYVALLIAMIFFIYAVIGMQMFGKVALQDGTQINRNNNFQTFPQAVLLLFRCA  
TGEAWQEIMLASLPGNRCDPESDFGPGEEFTCGSNFAIAYFISFFMLCAFLIINLFAVI  
MDNFDYLTRDWSILGPHHLDEFKRIWSEYDPGAKGRIKHLDVVALLRRIQPPLGFGKLCF  
HRVACKRLVAMNMLNSDGTVTFNATLFAVVRTSLKIKTEGNLEQANQELRIVIKKIWKR  
MKQKLLDEVIPPPDEEEVTVGKFYATFLIQDYFRKFRRRKEKGLLGNDAA PSTSSALQAG  
LRSLQDLGPEMRQALTCDEEEEEEGQEGVEEEDKDLKATMVSQPSARRGSGISVS  
LPVGDRLPDSLSFGPSDDDRGTPTSSQPSVPQAGSNTHRRGSGALIFTIPEEGNSQPKGT  
KGQNKQDEDEEVPDRLSYLDEQAGTPPCSVLLPPHRAQRYMDGHLVPRRLLPPTPAGRK  
PSFTIQCLQRQGSCELDLPIPGTYHRGRNSGPNRAQGSWATPPQRGRLLYAPLLLVEEGAA  
GEGYLGRSSGPLRTFTCLHVPGTHSDPSHGKRGSA DSLVEAVLISEGLGLFARDPRFVAL  
AKQEIADACRLTLDEMNAASDLLAQGTSSLYSDEESILSRFDEEDLGDEMACHAL

>sp|O60840-2|CAC1F\_HUMAN Isoform 2 of Voltage-dependent L-type calcium channel subunit alpha-1F OS=Homo sapiens OX=9606

GN=CACNA1F

MSESEGGKDTTPEPSPANGAGPGPEWGLCPGPPAVEGESSGASGLGTPKRRNQHSKHKT  
AVASAQRSPRALFCLTLANPLRRSCISIVEWKPFDILILLTIFANCVLGVYIPFPEDDS  
NTANHNLEQVEYVFLVIFTVETVLKIVAYGLVLHPSAYIRNGWNLLDFIIVVVGFLFSVLL  
EQGPGRPGDAPHTGGKPGGFDVKALRAFRVLRPLRLVSGVPSLHIVLNSIMKALVPLLHI  
ALLVLFVIIIIYAIIGLELFLGRMHKTCYFLGSDMEAEEDPSPCASSGSGRACTLNQTECR  
GRWPGPNGGITNFDNFFFAMLTVFQCVTMEGWTDVLYWMQDAMGYELPWVYFVSLVIFGS  
FFVLNLVLGVLSGEFSKEREKAKARGDFQKQREKQQMEEDLRGYLDWITQAEELDMEDPS  
ADDNLGPQLAELTNRRRGRRLRWFHSTRSTHSTSSHASLPASDTGSMETETQGEDEEEEGA  
LASCTRCLNKIMKTRVCRRLRRANRVLRARCRRAVKSNACYWAVLLLVLNLTITASEHH  
GQPVWLTQIQEYANKVLLCLFTVEMLLKLYGLGPSAYVSSFFNRFDCFVVCGGILETTLV  
EVGAMQPLGISVLRCVRLLRIFKVTRHWASLSNLVASLLNSMKSIASLLLLLFLFIIIFS  
LLGMQLFGGKFNFDQTHTKRSTFDTFPQALLTVFQILTGEDWNVVMYDGIMAYGGPFFPG  
MLVCIYFIIILFCGNYIILLNVFLAIAVDNLASGDAGTAKDKGGEKSNEKDLPQENEGLPV  
GVEKEEEEGARREGADMEEEEEEEEEEEEEEEEEGAGGVELLQEVVPKEKVVPIPEGSFA  
FCLSQTNPLRKGCHTLIHHHVFTNLIILVFIILSSVSLAAEDPIRAHSFRNHILGYFDYAF  
TSIFTVEILLKMTVFGAFLHRGSFCRSWFNMLDLLVVSLSLISFGIHSSAISVVKILRVL  
RVLRPLRAINRAKGLKHVVQCVFVAIRTIGNIMIVTLLQFMFACIGVQLFKGKFYTCTD  
EAKHTPQECKGSFLVYPDGDVSRPLVRERLWVNSDFNFDNVLSAMMALFTVSTFEGWPAL  
LYKAIDAYAEDHGPIYNYRVEISVFFIVYIIIIAFFFMMNIFVGFVITFRAQGEQEYQNC  
ELDKNQRCVEYALKAQPLRRYIPKNPHQYRVWATVNSAAFEYLMFLILLNTVALAMQH  
YEQTAPFNAMDI LNMVFTGLFTIEMVLKIIAFKPKHYFTDAWNTFDALIVVGSIVDIAV  
TEVNNGGHLGESSEDSSRISITFFRLFRVMRLVKLLSKGEGIRTLLWTFIKSFQALPYVA  
LLIAMIFFIYAVIGMQMFGKVALQDGTQINRNNNFQTFPQAVLLLFRCATGEAWQEIMLA  
SLPGNRCDPESDFGPGEFTCGSNFAIAYFISFFMLCAFLIINLFVAVIMDNFDYLTRDW  
SILGPHHLDEFKRIWSEYDPGAKGRIKHLDVALLRRIQPPLGFGKLCPHRVACKRLVAM  
NMPLNSDGTVTFNATLFAVVRTSLKIKTEGNLEQANQELRIVIKKIWKRMKQKLLDEVIP  
PPDEEEVTVGKFYATFLIQDYFRKFRRRKEKGLLGNDAA PSTSSALQAGLRS LQDLGPEM  
RQALTCDEEEEEEGQEGVEEEDKDL ETNKATMVSQPSARRGSGISVSLPVGDRLPDSL  
SFGPSDDDRGTPTSSQPSVPQAGSNTHRRGSGALIFTIPEEGNSQPKGTKQNKQDEDEE  
VPDRLSYLDEQAGTPPCSULLPPHRAQRYMDGHLVPRRLLPPTPAGRKPSFTIQCLQRQ  
GSCEDLPIPGTYHRGRNSGPNRAQGSWATPPQRGRLLYAPLLLVEEGAAGEGYLGRSSGP  
LRTFTCLHVPGTHSDPSHGKRG SADS LVEAVLISEGLGLFARDPRFVALAKQEIADACRL  
TLDEMDNAASDLLAQGTSSLYSDEESILSRFDEEDLGDEMACVHAL

>sp|O60840-4|CAC1F\_HUMAN Isoform 3 of Voltage-dependent L-type calcium channel subunit alpha-1F OS=Homo sapiens OX=9606

GN=CACNA1F

MSESEGGKGERILPSLQTLGASIVEWKPFDILILLTIFANCVLGVYIPFPEDDSNTANH  
NLEQVEYVFLVIFTVETVLKIVAYGLVLHPSAYIRNGWNLLDFIIVVVGFLFSVLLLEQGP  
RPGDAPHTGGKPGGFVDKALRAFRVLRPLRLVSGVPSLHIVLNSIMKALVPLLHIALLVL  
FVIIIIYAIIGLELFLGRMHKTCYFLGSDMEAEEDPSPCASSGSGRACTLNQTECRGRWPG  
PNGGITNFDNFFFAMLTVFQCVTMEGWTDVLYWMDAMGYELPWVYFVSLVIFGSFFVLN  
LVLGVLSGEFSKEREKAKARGDFQKQREKQQMEEDLRGYLDWITQAEELDMEDPSADDNL  
GSMAEEGRAGHRPQLAELTNRRRGRRLRWFHSTRSTHSTSSHASLPASDTGSMTEQTQDE  
DEEEGALASCTRCLNKIMKTRVCRRLRRANRVLRARCRRRAVKSNACYWAVLLLVLNLT  
IASEHHGQPVWLTQIQEYANKVLLCLFTVEMLLKLYGLGPSAYVSSFFNRFDCFVVCGGI  
LETTLVEVGAMQPLGISVLRVLLRIFKVTRHWASLSNLVASLLNSMKSIASLLLLLFL  
FIIIFSLGMQLFGGKFNFDQHTKRSTFDTFPQALLTVFQILTGEDWNVVMYDGIMAYG  
GPFPPGMLVCIYFIILFICGNYILLNVFLAIAVDNLAGDAGTAKDKGGEKSNEKDLPQE  
NEGLVPGVEKEEEEEGARREGADMEEEEEEEEEEEEEEEEEEGAGGVELLQEVVPKEKVVPI  
PEGSAFFCLSQTNPLRKGCHTLIHHVFTNLILVFIIILSSVSLAAEDPIRAHSFRNHILG  
YFDYAFTSIFTVEILLKMTVFGAFLHRGSFCRSWFNMLDLLVSVSLISFGIHSSAISV  
KILRVLRLRPLRAINRAKGLKHVVQCVFAIRTI GNIMIVTLLQFMFACIGVQLFKGK  
FYTCTDEAKHTPQECKGSFLVYPGDVSRPLVRERLWVNSDFNFDNVLSAMMALFTVSTF  
EGWPALLYKAIDAYAEDHGPIYNYRVEISVFFIVYIIIIAFFMMNIFVGFVIITFRAQGE  
QEYQNCELDKNQRQCVEYALKAQPLRRYIPKNPHQYRVWATVNSAAFEYLMFLLILLNTV  
ALAMQHYEQTAPFNYAMDILNMVFTGLFTIEMVLKIIAFKPKHYFTDAWNTFDALIVVGS  
IVDIAVTEVNNGGHLGESSEDSSRISITFFRLFRVMRLVKLLSKGEGIRTLTLLWTFIKSFQ  
ALPYVALLIAMIFFIYAVIGMQMFGKVALQDGTQINRNNNFQTFPQAVLLLFRCATGEAW  
QEIMLASLPGNRCDPESDFGPGEFTCGSNFAIAYFISFFMLCAFLIINLFVAVIMDNFD  
YLTRDWSILGPHHLDEFKRIWSEYDPGAKGRIKHLDVVALLRRIQPPLGFGKLCPHRVAC  
KRLVAMNMPNLSDGTVTFNATLFAVVRTSLKIKTEGNLEQANQELRIVIKKIWKRMKQKL  
LDEVI PPPDEEEVTVGKFYATFLIQDYFRKFRRRKEKGLLGNDAA PSTSSALQAGLRSLQ  
DLGPEMRQALTCDTEEEEEEGQEGVEEEDKDLTKATMVSQPSARRGSGISVSLPVG  
RLPDSLSFGPSDDDRGTPTSSQPSVPQAGSNTHRRGSGALIFTIPEEGNSQPKGTGKQNK  
QDEDEEVPDRLSYLDEQAGTPPCSVLLPPHRAQRYMDGHLVPRRLLPPTPAGRKPSFTI  
QCLQRQGSCEDLPIPGTYHRGRNSGPNRAQGSWATPPQRGRLLYAPLLLVEEGAAGEGYL  
GRSSGPLRTFTCLHVPGTHSDPSHGKRGSA DSLVEAVLISEGLGLFARDPRFVALAKQEI  
ADACRLTLDEMDNAASDLLAQGTSSLYSDEESILSRFDEEDLGDEMACVHAL

>sp|O60840-7|CAC1F\_HUMAN Isoform 6 of Voltage-dependent L-type calcium channel subunit alpha-1F OS=Homo sapiens OX=9606

GN=CACNA1F

MSESEGGKDTTPEPSPANGAGPGPEWGLCPGPPAVEGESSGASGLGTPKRRNQHSKHKT  
AVASAQRSPRALFCLTLANPLRRSCISIVEWKPFDILILLTIFANCVALGVYIPFPEDDS  
NTANHNLEQVEYVFLVIFTVETVLKIVAYGLVLHPSAYIRNGWNLLDFIIVVVGFLFSVLL  
EQGPGRPGDAPHTGGKPGGFVVKALRAFRVLRPLRLVSGVPSLHIVLNSIMKALVPLLHI  
ALLVLFVIIIIYAIIGLELFLGRMHKTCYFLGSDMEAEEDPSPCASSGSGRACTLNQTECR  
GRWPGPNGGITNFDNFFFAMLTVFQCVTMEGWTDVLYWMQDAMGYELPWVYFVSLVIFGS  
FFVLNLVLGVLSGEFSKEREKAKARGDFQKQREKQQMEEDLRGYLDWITQAEELDMEDPS  
ADDNLGSMAEGRAGHRPQLAELTNRRRGRLRWFHSHSTRSTHSTSSHASLPASDTGSMTE  
TQGDEDEEEGALASCTRCLNKIMKTRVCRRLRRANRVLRARCRAVKSNAFYWAVLLLVF  
LNTLTIASEHHGQPVWLTQIQEYANKVLLCLFTVEMLLKLYGLGPSAYVSSFFNRFDCFV  
VCGGILETTLVEVGAMQPLGISVLRVRLRLRIFKVTRHWASLSNLVAVSLLNSMKSIASLL  
LLLFLFIIIFSLGMLQFVGGKFNFDQTHTKRSTFDTFPQALLTVFQILTGEDWNVVMYDG  
IMAYGGPFFPGMLVCIYFIILFICGNYILLNVFLAIAVDNLAGDAGTAKDKGGEKSNEK  
DLPQENEGLVPGVEKEEEEEGARREGADMEEEEEEEEEEEEEEEEEEGAGGVELLQEVVPKE  
KVVPIPEGSAFFCLSQTNPLRKGCHTLIHHVFTNLIILVFIILSSVSLAAEDPIRAHSFR  
NHILGYFDYAFTSIFTVEILLKMTVFGAFLHRGSFCRSWFNMLDLLVSVSLISFGIHSS  
AISVVKILRVLRLRPLRAINRAKGLKHVVQCVFAIRTIGNIMIVTLLQFMFACIGVQ  
LFKGFYCTCTDEAKHTPQECKGSFLVYPDGDVSRPLVRERLWVNSDFNFDNVLSAMMALF  
TVSTFEGWPALLYKAIDAYAEDHGPIYNYRVEISVFFIVYIIIIIAFFMMNIFVGFVIITF  
RAQGEQEYQNCLELKNQRQCVEYALKAQPLRRYIPKNPHQYRVWATVNSAAFEYLMFLLI  
LLNTVALAMQHYEQTAPFNYAMDILNMVFTGLFTIEMVLKIIAFKPKHYFTDAWNTFDAL  
IVVGSIVDIAVTEVNNGGHLGESSEDSRISITFFRLFRVMRLVKLLSKGEGIRTLTLLWTF  
IKSFQALPYVALLIAMIFFIYAVIGMQMFGKVALQDGTQINRNNNFQTFPQAVLLLFRCA  
TGEAWQEIMLASLPGNRCDPESDFGPGEEFTCGSNFAIAYFISFFMLCAFLIINLFAVI  
MDNFDYLTRDWSILGPHHLDEFKRIWSEYDPGAKGRIKHLDVALLRRIQPPLGFGKLCF  
HRVACKRLVAMNMLNSDGTVTFNATLFAVVRTSLKIKTEGNLEQANQELRIVIKKIWK  
MKQKLLDEVIPPPDEEEVTVGKFYATFLIQDYFRKFRRRKEKGLLGNDAA PSTSSALQAG  
LRSLQDLGPEMRQALTCDEEEEEEGQEGVEEEDKDLKATMVSQPSARRGSGISVS  
LPVGDRLPDSLSFGPSDDDRGTPTSSQPSVPQAGSNTHRRGSGALIFTIPEEGNSQPKGT  
KGQNKQDEDEEVPDRLSYLDEQAGTPPCSVLLPPHRAQRYMDGHLVPRRRLLPPTPAGRK  
PSFTIQCLQRQGSCELDLPIPGTYHRGRNSGPNRAQVLISEGLGLFARDPRFVALAKQEIA  
DACRLTLDEMDNAASDLLAQGTSSLYSDEESILSRFDEEDLGDEMACVHAL

## Appendix 7. Failed 5' RACE experiments

### 1. RLM first choice RACE kit (Thermofisher, UK; AM1700)

To use RLM first choice RACE kit, RNA samples were reverse transcribed with a 5' adapter provided by the kit. The same RNA samples that were extracted for the long-read nanopore sequencing in chapter 2, which were extracted from Methods section 3.2 in chapter 2, were used. Total RNAs extracted from three different brain regions (cerebellum, cingulate, DL-PFC) from 6 different individuals were pooled (Table Appendix 7.1). The cerebellum is included because the cerebellum has distinctive cell types compared to the other brain regions included. This will allow to maximize the number of cell types included to identify TSS of the short-length transcripts in the 3' region of LTCC genes and to reduce the possibility of missing cell-type specific alternative TSSs.

*Table Appendix 7.1 Total RNA samples used for the 5' RACE experiment. All total RNA samples were pooled together. Each number corresponds to a different individual (See Methods section 3.1 in chapter 2).*

Brain region	Sample name
DLPFC	5346, 5238, 5298, 5717, 5579, 5244
Cingulate	5346, 5238, 5298, 5717, 5579, 5244
Cerebellum	5346, 5238, 5298, 5717, 5579, 5244

Prior to the reverse transcription (RT), pooled total RNAs was treated with Calf Intestine Alkaline Phosphatase (CIP) to remove 5'-phosphatases but leaving the cap structure on the 5' end of mRNA undamaged. The mRNA with the cap structure was treated with Tobacco Acid Pyrophosphatase (TAP) to remove the cap structure and leave only 5' monophosphate. Before ligation, RNA without TAP treatment was prepared as a negative control to verify whether the RACE product is the intended RACE product or a random fragment. RNA adapter oligonucleotides, which are 45 base pairs in length, are ligated to the RNA. As RNA is dephosphorylated, T4 RNA ligase is used to ligate the adapter to the RNA with 5' monophosphate.

After ligation, de-capped mRNAs acquire adapter sequences on their 5' ends. Then, reverse transcription with random primers is used to obtain cDNAs. Nested PCR is then used to amplify the 5' ends of a specific transcript. The kit supplies two nested forward primers corresponding to the 5' RACE adapter sequence so that the user can use reverse primers which are specific to the gene of interest (Figure 3.2). After the nested PCR, it is expected to obtain RACE products that are 100bp to 300bp in length.

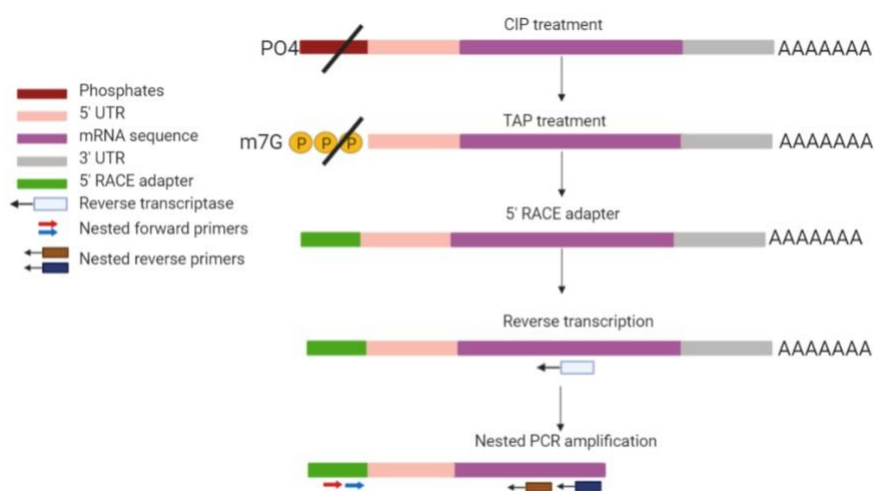


Figure 3.15. Flow chart of RLM firstchoice RACE

Started with 1 $\mu$ g or less total RNA.

Firstly, mix reagents (Table 1) and incubated at 37°C for 1 hour to treat RNA samples with CIP.

Table 1 Reagents to treat RNA samples with CIP

Component	Amount ( $\mu$ l)
1 $\mu$ g of total RNA	X
10X CIP buffer	2
Calf Intestine Alkaine Phosphatase (CIP)	2
Nuclease-free water	Up to total 20

Once RNA samples were treated with CIP, add 15  $\mu$ l of Ammonium Acetate Solution, 115  $\mu$ l of Nuclease-free water and 150  $\mu$ l of acid phenol:chloroform then centrifuge at 13,000 x g for 5 minutes at room temperature. The supernatants were transferred to a new tube. The samples were precipitated with 150  $\mu$ l of isopropanol. The samples with 150  $\mu$ l of isopropanol were vortexed thoroughly and chilled on ice for 10 minutes. They were centrifuged at 13,000 x g for 20 minutes and supernatant were disregarded. The pellet was rinsed with 500 $\mu$ l of 70% ethanol and centrifuged at 13,000 x g for 5 minutes. The ethanol was removed, and pellet were air dried. The samples were resuspended in 4  $\mu$ l of 1X TAP buffer and 1  $\mu$ l of CIP-treated RNA was reserved for minus TAP control reaction.

The CIP-treated RNA samples were treated with TAP at 37°C for 1 hour by adding the following (Table 2)

*Table 2 Reagents for TAP treatment*

Component	Amount ( $\mu$ l)
CIP'd RNA	4
Tobacco Acid Pyrophosphatase	1
Total	5

After TAP treatment, 5'RACE adapter was ligated to the samples by adding the following (Table 3). The mixed components were incubated at 37°C for one hour.

*Table 3 Ligation of 5'RACE adapter to the samples*

Component	Amount ( $\mu$ l)
CIP/TAP-treated RNA	5
5'RACE Adapter	1
10X RNA Buffer	1
T4 RNA Ligase (2.5U/ $\mu$ l)	2
Nuclease-free Water	1

Processed RNA samples are reverse transcriptased at 42°C for one hour. The following reagents (Table 4) were added to the ligated RNA samples / minus-TAP control.

*Table 4 Reagents for reverse transcription*

Component	Amount ( $\mu$ l)
Ligated RNA (or minus-TAP control)	2
dNTP Mix	4
Random Decamers	2
10X RT Buffer	2
RNase Inhibitor	1
M-MLV Reverse Transcriptase	1
Nuclease-free water	Up to the total of 20 $\mu$ l

With reverse transcriptased RNA samples, nested PCR was executed.

## 2. Failed attempts of 5' RACE with 5'/3' RACE Kit, 2<sup>nd</sup> Generation

Examining the TSS of the short-length transcript from the 3' region of *CACNA1S* and *CACNA1F* with the kit was unsuccessful. Given that the short-length transcript from the 3' region of *CACNA1S* and *CACNA1F* have extremely low expression levels in the human brain, the experimental strategy had to be highly tailored to each gene of interest. It would be beneficial if the mRNA was reverse transcribed with primers specific to the gene of interest. Therefore, a new experimental approach using a different kit was employed.

The 5'/3' RACE Kit, 2<sup>nd</sup> Generation (Roche, UK; 03353621001) was used as an alternative method (Figure 3.1). This kit uses a gene specific primer for the RT step, enhancing the selectivity of the gene of interest. For the first attempt with this kit, total RNAs of DLPFC, cingulate cortex, and cerebellum from six different individuals (the same samples as used previously for RLM first choice RACE kit; Table 3.1) were pooled. However, despite the gene-specific method, TSSs were not identified: all the RACE products mapped to non-targeted genes, indicating amplification of non-specific products. Various modifications were attempted, such as using diluted outer PCR products to perform inner PCR amplification or performing skewed PCR amplifications for the outer PCR amplification step. However, the 5'RACE experiment with total RNAs of DLPFC, cingulate, and cerebellum from six different individuals was not successfully done. Therefore, different samples were used which will be described in the successful attempts of the 5' RACE section.





## Appendix 9. R script used to analyze CAGE data using CAGEr package

```

##Using bash to grep my bed data (chrX)
cd /t1-data/user/hlee
grep "chrX" brain%2c%20adult%2c%20donor1.CNhs11796.10084-102B3.hg19.ctss.bed > Filename
grep -w "chr1" file name > file name

## Format the bed data file to ctss file
awk 'OFS="\t" {print $1,$3,$6,$5} ' Filename > New filename

##download and install package in Rstudio
if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("CAGEr")
library(CAGEr)

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("rtracklayer")
library(rtracklayer)

if (!requireNamespace("BiocManager", quietly = TRUE))
  install.packages("BiocManager")
BiocManager::install("BSgenome.Hsapiens.UCSC.hg19")
library(BSgenome.Hsapiens.UCSC.hg19)

## load my data file
setwd("E:/from cbrg folder/CACNA1B")
inputFiles = list.files("E:/from cbrg folder/CACNA1B", pattern="b_ctss.bed$", full.names=TRUE)
inputFiles

## creating the object called HumanB
HumanB <- CAGExp( genomeName = "BSgenome.Hsapiens.UCSC.hg19"
, inputFiles = inputFiles
, inputFileType = "ctss"
, sampleLabels = sub( ".chr9.ctss", "", basename(inputFiles)
)

HumanB

## Reading in the data -> reads the provided files in the order as in inputFiles. This command creates a table with
counts of CAGE tags of each transcription start sites in all dataset in inputFiles.
HumanB <- getCTSS(HumanB)
HumanB
CTSSStagCountSE(HumanB) ## Genomic coordinates of all transcription start sites and CAGE tag counts can be
retrieved
CTSScoordinatesGR(HumanB) ## To access the coordinates of CAGE tags
CTSSStagCountDF(HumanB) ## To access the expression values of CAGE tags
CTSSStagCountGR(HumanB, 12) ## Granges for a sample with 1 expression count
sampleLabels(HumanB) ## To check sample labels and their orders

## To merge several datasets to one group. mergeIndex argument specifies which datasets will be merged and the
order of the final dataset. Samples with the same number label will be merged by summing CAGE tags per TSS.
The final dataset will be ordered by the labels provided in mergedSampleLables argument.

```

```
HumanB <- mergeSamples(HumanB, mergeIndex = c(1,1,2,1,1,2,1,1,2,1,1,2),
mergedSampleLabels = c("Adult_brain_B", "Newborn_brain_B"))
```

## To annotate CAGE tags with the reference genome and categorize into regions (Promoter, Exon, Intron and Unknown). The reference genome from GENCODE was loaded using import.gff function of the rtracklayer package.

```
hg19annot <- import("gencode.v19.annotation.gff3")
HumanB <- annotateCTSS(HumanB, hg19annot)
colData(HumanB)[,c("librarySizes", "promoter", "exon", "intron", "unknown")]
CTSScoordinatesGR(HumanB)
plotAnnot(HumanB, "counts") ## To provide a bar plot of annotation result
```

## Normalization

As library sizes (number of total CAGE tags) are different to each dataset, normalization is needed to compare two dataset. The librarySizes function allows to see the total number of CAGE tags in each dataset.

Tags per million normalization was used to get normalized tags per million tags.

```
librarySizes(HumanB)
```

## To check whether the CAGE datasets follow power-law distribution with total distribution of million tags

```
plotReverseCumulatives(HumanB, fitInRange = c(2, 1000000), onePlot = TRUE)
```

## A reference power-law distribution is based on alpha (slope in the log-log graph) and T (total number or tags = million) parameters.

```
HumanB <- normalizeTagCount(HumanB, method = "powerLaw", fitInRange = c(2, 1000000), alpha = 1.27, T = 2*10^6)
```

```
HumanB[["tagCountMatrix"]]
```

## To cluster nearby transcription start sites together and create set of tag clusters. Tags within 20bp range were clustered together. The final set of tag clusters include singletons which are clusters with only one transcription start site. The clusterCTSS function creates a set of tag clusters for each data set separately. It returns genomic coordinates of each cluster, number of transcription start sites within the cluster, frequently used transcription start sites, calculate the total CAGE signal within cluster, and CAGE signal of frequently used transcription start site only. tagClusterGR function allows to extract tag clusters of each data set.

```
HumanB <- clusterCTSS( HumanB
```

```
, threshold = 0
```

```
, thresholdIsTpm = TRUE
```

```
, nrPassThreshold = 0
```

```
, method = "distclu"
```

```
, maxDist = 20
```

```
, removeSingletons = TRUE
```

```
, keepSingletonsAbove = 1)
```

```
tagClustersGR(HumanB, sample = "Adult_brain_B")
```

```
tagClustersGR(HumanB, sample = "Newborn_brain_B")
```

## Promoter Width – analyze promoter width across all dataset in CAGEexp object.

```
HumanB <- cumulativeCTSSdistribution(HumanB, clusters = "tagClusters", useMulticore = T)
```

```
HumanB <- quantilePositions(HumanB, clusters = "tagClusters", qLow = 0.1, qUp = 0.9)
```

```
tagClustersGR( HumanB, "Adult_brain_B"
```

```
, returnInterquantileWidth = TRUE, qLow = 0.1, qUp = 0.9)
```

```
tagClustersGR( HumanB, "Newborn_brain_B"
```

```
, returnInterquantileWidth = TRUE, qLow = 0.1, qUp = 0.9)
```

```
plotInterquantileWidth(HumanB, clusters = "tagClusters", tpmThreshold = 3, qLow = 0.1, qUp = 0.9)
```

```
plotInterquantileWidth(HumanB, clusters = "tagClusters", tpmThreshold = 10, qLow = 0.1, qUp = 0.9)
```

```

## Export – normalized CAGE tag counts were exported to a bedGraph file format to upload to a genome browser.
trk <- exportToTrack(CTSSnormalizedTpmGR(HumanB, "Adult_brain_B"))
HumanB |> CTSSnormalizedTpmGR("all") |> exportToTrack(HumanB, oneTrack = FALSE)
trkBKG <- split(trk, strand(trk), drop = TRUE)
trkBKG[['+']]@trackLine@name <- paste0(trkBKG[['+']]@trackLine@name, "+")
trkBKG[['-']]@trackLine@name <- paste0(trkBKG[['-']]@trackLine@name, "-")
trkBKG[['+']]@trackLine@description <- paste0(trkBKG[['+']]@trackLine@description, "+ strand")
trkBKG[['-']]@trackLine@description <- paste0(trkBKG[['-']]@trackLine@description, "- strand")
rtracklayer::export.bedGraph(trkBKG, "AdultBrainB.bedGraph")
trk <- exportToTrack(CTSSnormalizedTpmGR(HumanB, "Newborn_brain_B"))
HumanB |> CTSSnormalizedTpmGR("all") |> exportToTrack(HumanB, oneTrack = FALSE)
trkBKG <- split(trk, strand(trk), drop = TRUE)
trkBKG[['+']]@trackLine@name <- paste0(trkBKG[['+']]@trackLine@name, "+")
trkBKG[['-']]@trackLine@name <- paste0(trkBKG[['-']]@trackLine@name, "-")
trkBKG[['+']]@trackLine@description <- paste0(trkBKG[['+']]@trackLine@description, "+ strand")
trkBKG[['-']]@trackLine@description <- paste0(trkBKG[['-']]@trackLine@description, "- strand")
rtracklayer::export.bedGraph(trkBKG, "NewbornBrainB.bedGraph")

```

## Appendix 10. Multiple sequence alignment of 3' transcripts of LTCC genes

CLUSTAL O(1.2.4) multiple sequence alignment

1S	GTGCACCGACTTGTCCAAGATGACAGAGGAGGAGTGCAGGGGCTACTACTACGTGTACAA	60
1F	-----	0
1C	-----	0
1D	-----	0
1S	GGACGGGGACCCCATGCAGATAGAGCTGCGTCACCGCGAGTGGGTACACAGCGACTTCCA	120
1F	-----	0
1C	-----	0
1D	-----	0
1S	CTTCGACAAATGTGCTCTCAGCCATGATGTCCCTCTTCACGGTCTCCACCTTCGAGGGATG	180
1F	-----	0
1C	-----	0
1D	-----	0
1S	GCCTCAGCTGCTGTACAAGGCCATAGACTCCAATGCGGAGGACGTGGGTCCCATCTACAA	240
1F	-----	0
1C	-----	0
1D	-----	0
1S	CAACCGTGTGGAGATGGCCATCTTCTTCATCATCTACATCATCCTCATTGCCTTCTTCAT	300
1F	-----	0
1C	-----	0
1D	-----	0
1S	GATGAACATCTTTGTGGGCTTCGTTCATTGTCACCTTCCAGGAGCAGGGAGAGACTGAGTA	360
1F	-----	0
1C	-----	0
1D	-----	0
1S	CAAGAACTGTGAGCTGGACAAGAACCAGCGCCAATGTGTACAGTATGCCCTGAAGGCCCG	420
1F	-----	0
1C	-----	0
1D	-----	0
1S	CCCCTGAGGTGCTACATTCCCAAAAACCCATACCAGTACCAGGTGTGGTACATTGTCAC	480
1F	-----	0
1C	-----	0
1D	-----	0
1S	CTCCTCCTACTTTGAATACCTGATGTTTGCCCTCATCATGCTCAACACCATCTGCCTCGG	540
1F	-----	0
1C	-----	0
1D	-----	0
1S	CATGCAGCACTACAACCAGTCGGAGCAGATGAACCACATCTCAGACATCCTCAATGTGGC	600
1F	-----	0
1C	-----	0
1D	-----	0
1S	CTTCACTATCATCTTCAACCTGGAGATGATCCTCAAGCTCATGGCCTTCAAGGCCAGGGG	660
1F	-----	0
1C	-----	0
1D	-----	0

1S	CTACTTTGGAGACCCCTGGAATGTGTTTGACTTCCTGATTGTCATTGGCAGCATCATTGA	720
1F	-----	0
1C	-----	0
1D	-----	0
1S	TGTCATCCTCAGTGAGATCGACTTTCCTGGCCTCCAGCGGGGACTGTATTGCCTGGG	780
1F	-----	0
1C	-----	0
1D	-----	0
1S	TGGAGGCTGCGGGAACGTTGACCCAGATGAGAGTGCCCGCATCTCCAGCGCCTTCTCCG	840
1F	-----	0
1C	-----	0
1D	-----	0
1S	CCTGTTCCTGTCATGAGGCTGATCAAGCTGCTGAGCCGGGCAGAAGGAGTGCGAACCCCT	900
1F	-----	0
1C	-----	0
1D	-----	0
1S	CCTGTGGACGTTTCATCAAGTCCTTCCAGGCCCTACCTACGTGGCTCTGCTCATCGTCAT	960
1F	-----	0
1C	-----	0
1D	-----	0
1S	GCTCTTCTTCATCTACGCTGTCATCGGCATGCAGATGTTTGGGAAGATCGCCTTGGTGGA	1020
1F	-----	0
1C	-----	0
1D	-----	0
1S	TGGGACCCAAATAAACCGGAACAACAACCTCCAGACCTTCCCACAAGCTGTGCTACTGCT	1080
1F	-----	0
1C	-----	0
1D	-----	0
1S	CTTCAGGTGTGCAACAGGTGAGGCCTGGCAGGAGATCCTACTGGCCTGCAGCTATGGGAA	1140
1F	-----	0
1C	-----	0
1D	-----	0
1S	GCTGTGTGACCCAGAGTCGGACTATGCCCCAGGGAGGAGTACACATGTGGCACCAACTT	1200
1F	-----	0
1C	-----	0
1D	-----	0
1S	TGCATACTACTACTTCATCAGCTTCTACATGCTCTGTGCCTTCTGGTCATCAACCTCTT	1260
1F	-----	0
1C	-----	0
1D	-----	0
1S	TGTGGCTGTTCATCATGGACAATTTTGACTACCTCACCCGGGACTGGTCCATCCTGGGCC	1320
1F	-----	0
1C	-----	0
1D	-----	0
1S	TCATCACCTGGATGAGTTCAAGGCCATCTGGGCAGAGTATGACCCAGAGGCTAAGGGGAG	1380
1F	-----	0
1C	-----	0
1D	-----	0

1S	AATCAAACACCTGGACGTGGTGACCCTGCTGAGAAGGATTCAGCCCCCTCTGGGCTTTGG	1440
1F	-----	0
1C	-----	0
1D	-----G	1
1S	GAAGTTCTGCCACATCGGGTAGCTTGTAAGCGGCTGGTGGGCATGAACATGCCCTGAA	1500
1F	-----	0
1C	-----	0
1D	GAAGTTATGTCCACACAGGCTAGCGTGCAAGAGATTAGTTGCCATGAACATGCCTCTCAA	61
1S	CAGCGACGGCACAGTCACCTTCAATGCCACACTCTTTGCCCTGGTCCGCACGGCACTCAA	1560
1F	-----	0
1C	-----CCCTGGTCAGGACGGCCCTGAG	22
1D	CAGTGACGGGACAGTCATGTTTAAATGCAACCTGTTTGCCTTGGTTCGAACGGCTCTTAA	121
1S	GATCAAGACGGAAGGTAACCTTTGAGCAGGCCAACGAGGAGCTGAGGGCCATCATCAAGAA	1620
1F	-----GAGCTGCGGATTGTCATCAAAAA	23
1C	GATCAAAACAGAAGGGAACCTAGACAAGCCAATGAGGAGCTGCGGGCGATCATCAAGAA	82
1D	GATCAAGACCGAAGGGAACCTGGAGCAAGCTAATGAAGAACTTCGGGCTGTGATAAAGAA	181
	* * * * * * * * * *	
1S	GATCTGGAAGAGAACCAGCATGAAGCTCTTGGACCAGGTCATCCCTCCAATAGGAGATGA	1680
1F	GATCTGGAAGCGGATGAAACAGAAGCTGCTAGATGAGGTCATCCCCCACCAGACGAGGA	83
1C	GATCTGGAAGCGGACCAGCATGAAGCTGCTGGACCAGGTTGGTCCCTGCAGGTGATGA	142
1D	AATTTGGAAGAAAACCAGCATGAAATTACTTGACCAAGTTGTCCTCCAGCTGGTGTGATGA	241
	* * * * * * * * * * * * * * * * * * * *	
1S	TGAGGTGACAGTGGGGAAGTTCTACGCCACATTCCCTCATCCAGGAGCACTTCCGGAAGTT	1740
1F	GGAGGTACCCGTGGGCAAATTCTACGCCACATTTCTGATCCAGGACTATTTCCGCAAATT	143
1C	TGAGGTACCCGTGGCAAGTTCTACGCCACGTTCCCTGATCCAGGAGTACTTCCGGAAGTT	202
1D	TGAGGTAACCGTGGGGAAGTTCTATGCCACTTTCCCTGATACAGGACTACTTTAGGAAATT	301
	* * * * * * * * * * * * * * * * * * * *	
1S	CATGAAACGCCAAGAGGAGTATTATGG-----	1767
1F	CCGGCGGAGGAAAGAAAAGGGCTACTAGGCAACGACGCCGCCCTAGCACCTCTTCCGC	203
1C	CAAGAAGCGCAAAGAGCAGGGCCTTGTGGCAAGCCCTCCAGAGGA---ACGCGTGTGTC	259
1D	CAAGAAGCGGAAAGAACAGGACTGGTGGGAAAGTACCCTGCGAAGAACACCACAATTGC	361
	* * * * * * * * * *	
1S	-----	1767
1F	CCTTCAGGCTGGTCTGCGGAGCCTGCAGGACTTGGGTCTGAGATGCGGCAGGCCCTCAC	263
1C	TCTGCAGGCTGGCTTGCGCACACTGCATGACATCGGGCCTGAGATCCGACGGGCCATCTC	319
1D	CCTACAGGCGGGATTAAGGACACTGCATGACATTGGGCCAGAAATCCGGCGTGTATATC	421
1S	-----	1767
1F	CTGTGACACAGAGGAGGAGGAAGAAGAGGGGCAGGAGGGAGTGGAGGA-----GGA	314
1C	TGGAGATCTCACCGTGTAGGAGGAGCTGGACAAGGCCATGAAGGAGGCTGTGTCCGCTGC	379
1D	GTGTGATTTGCAAGATGACGAGCCTGAGGAAACAAAACGAGAA-----	464
1S	-----	1767
1F	AGATGAAAAGGACTTGAAACTAACA---AAGCCACGATGGTCTCCAGCCCTCAGCTCG	371
1C	TTCTGAAGATGACATCTTCAGGAGGCCGGTGGCCTGTTCCGCAACCACGTCAGCTACTA	439
1D	-GAAGAAGATGATGTGTTCAAAAGAAATGGTGCCTGCTTGAAACCATGTCAATCATGT	523
1S	-----	1767
1F	CCGGGGCTCCGGGATTTCTGTGTCTCTGCCTGTGCGGGACAGACTTCCAGATTCACTCTC	431
1C	CCAAAGCGACGGCC-----GGAGCGCTTCCCC-----AGACCTTCACCA	480
1D	TAATAGTGATAGGA-----GAGATTCCTTTCAGC-----AGACCAATACCA	564
1S	-----	1767
1F	CTTTTGGGCCAGTGTGATGAT---GACAGGGGGACTCCACCTCCAGTCAGCCAGTGTGCC	488
1C	CTCAGCGCCCTGCACATCAACAAGGCGGGC-----AGCAGCCAGGGCGACAC	529
1D	CCCACCGTCCCTGCATGTCCAAGGCCTTCAATTCCACCTGCAAGTGATACTGAGAAAC	624

1S	-----	1767
1F	CCAGGCTGGATCCAACACCCACAGGAGAGGCTCTGGGGCTCTCATTTTCACCATCCCAGA	548
1C	TGAGTCGCCATCCCACGAGAAGCTGGT--GGA CTC-CACCTTCACCCCGAGCAGCTAC--	584
1D	CGCTGTTTCCTCCAGCAGGAAATTCGGTGTGTCTAT-AACCATCATAACCCATAATTCATA	683
1S	-----	1767
1F	AGAAGGAAATTCAGCCC-AAGGGAACCAAAGGGCAAACAAGCA-----	593
1C	-----TCGTCCACCGGCTCCAACGCCAACATCAACAACGCCAACACACCGCCCTG	635
1D	GGAAAGCAAGTTCACCTCAACAAATGCCAATCTCAATAATGCCAATATGTCCAAAGCT	743
1S	-----	1767
1F	-----AGATGAGGATGAGGAAGTC	612
1C	GGTCGCCTCCCTCGCCCCGCGGCTACCCAGCACGGTCAGC-----ACTGTGGAGGGC	689
1D	G-----CCCATGGAAAGCGGCCAGCATTTGGGAACCTTGAGCATGTGTCTGAA	791
1S	-----	1767
1F	CCTGATCGGCTTTCCTACCTAGATGAGCAGGCAGGGACTCCCCGTGCTCAGTCCTT---	669
1C	CACGGGCCCCCTTGTCCCTGCCATCCGGGTGCAGGAGGTGGCGTGGAAAGCTCAGCTCC	749
1D	AATGGGCATCA-TTCTTCCAC--AAGCATGACCGGGAGCCT-CAGAGAAGGTCCAGTGT	847
1S	-----	1767
1F	-----TTGCCACCTCACAGAGCTCAGAGATACATGGATGGGCACCTGGT	713
1C	AACAGGGAAAGG-----CACGTTCCGATGTGTGAGGATCTGGAGCTCAGGAGGGATT	801
1D	GAAAAGAACCCGCTATTATGAAACTTACATTAGGTCCGACTCAGGAGATGAACAGCTCCC	907
1S	-----	1767
1F	ACCACGCGCGCTCTG-----CTGCCCCCACACCTGCAGGTCGGAA----GCCCTCCTT	764
1C	CAGGCTCAGCAGGGACTCAGGCTCACTGCCTTCTGCTCAGGAAAGCAAACCCCTTAGGT	861
1D	AACTATTTGCGGGGAAGACC---CAGAGATACATGGCTATTTCAGGGACCCCCACTGCTT	964
1S	-----	1787
1F	-----CTATCGGCCCAAGAAGGACA	820
1C	CACCATCC---AGTGTCTGCAGCGCCAGGGCAGTGTGAGGATTTACCCATCCCAGGCA	921
1D	GCCACTCCCGGGAGAGCCAGGCAGCCATGGCGGGTCCAGGAGGACGTCTCAGGATGAGA	1011
	GG-----GGGAGCAGGAGTA---TTTCAGTAGTGAGGAATGCTACGAGGATGACA	
	* * *	
1S	TT--GTACAGATC---CAGGCAGGGCTGCGGACCATTGAGGAAGAGGCA--GCCCCGAG	1840
1F	CCTATCATCTGTG---GGCGAAATTCAGGG-CCCAATAGGGCTCAGGGTTCCCTGGGCAAC	875
1C	CCTATGAAGTGAAGATGAACCATGACACGG-AGGCCTG---CAGTGAG---CCCAGCCTG	974
1D	GCTCGCCACCTGGAGCAGGCAAACTATG-GCTACTACAGCAGATACCCAGGCAGAAAC	1070
	* * *	
1S	ATCTGTGCGACGGTCTCAGGAGACCTGGCTGCTGAGGAGGAGCTGGAGAGAGCCATGGTG	1900
1F	ACCACCTCAGCGGGT--CGGCTC-----C---TGTATGCCCCGCTGTTGTGGTG	921
1C	CTCTCCACAGAGATGCTCTCTACCAGGATGACG---AAAATCGGCAACTGACGCTCCCA	1031
1D	ATCGACTCTGAGAGGCCCGGAGGTACC-----ATCATCCCAAGGATTCTTGAG	1121
	* * *	
1S	GAGGCTGCGA---TGGAGGAGGGAATATTCGGAGGACT--GGAGGCCTGTTGGCCAGG	1955
1F	GAAGAGGGCGCAGCGGGGAGGGGTACCTCGGCAGA--TCCAGTGGCCACTGCGCACCT	979
1C	GAGGAGGACA-----AGAGGGACATCCGGCAATCTCCGAAGAGGGGTTTCCTCCGC	1082
1D	GACGATGACTCGCCCGTTTGCTATGATTACGGAGATCTCCAAGGAGACGCCTACTACCT	1181
	** * * * * *	
1S	TGGACAACCTCCTGGAAAGGACCAAC-----	1981
1F	-----	979
1C	TCT---GCCTCACTAGGTGCAAGGGCTCCTTCCACCTGGAATGTCTGAAGCGACAGAAG	1139
1D	CCCACCCAGCATCCCACCGGAGATCCTCCTTCAACTTTGAGTGCCTGCGCCGGCAGAGC	1241
1S	-----	2026
1F	-----	979
1C	GACCGAGGGGA-----GACATCTCTCAGAAGACAGTCTGCCCTTGCA	1184



1D	AGCCAGGAAGAGGTCCCGTCGTCTCCCATCTTCCCCATCGCACGGCCCTGCCTCTGCAT	1301
1S	GAGATAGAGATGGAAGAGATGGAG-TCACCTGTCTTCTTGGAGGACTTCCCACAAGATCC	2085
1F	-----	979
1C	CTGGTTCATCATCAGGCATTGGCAGTGGCAGGCCCTGAGCCCCCTCCTCC--AGAGAAGCC	1242
1D	CTAATGCAGCAACAGATCATGGCAGTTGCCGGCCTAGATTCAAGTAAAG--CCCAGAAGT	1359
1S	ACGCACCAACCCCTG-----GCTCGTGCCAATACCAACAATGCCAACGCCAATGTGCGC	2140
1F	-----	979
1C	ATTCCCTGCCTCATTTCCCTAGGCCTTTTGGCACCCACCAGCCACACCTGGCAGCCGAG	1302
1D	ACTCACCGAGTCACTCGACCCGGTCTGGGCCACCCCTCCAGCAACCCCTCCCTACCGGG	1419
1S	TATGGCAACAGCAACCATAGCAACAGCCATGTGTTTTCCAGTGTCCACTATGAAAGGGAG	2200
1F	-----	979
1C	GCTGGCCCCACAGC-----CCGTCCCCACCCT-----GCGGCTT-----GAG	1340
1D	ACTGGACACCGTGT-----ACACCCCTGTAT-----CCAAGTG-----GAG	1457
1S	TTCCAGAAAGAGACAGAGACGCCTGCTACCAGAGGACGAGCCCTGGCCAACCCT----G	2256
1F	-----	979
1C	GGGGTCGAG-----TCCAGTGAGAACTCAACAGCAGCTTCCCATCCATCCACT	1389
1D	CAGTCAGAG-----GCCCTGGACCAGGTGAACGGCAGCCTGCCGTCCCTGCACC	1506
1S	CAGGGTCTGGGACCCACAGCAAACCCTGTGTGGAGATGCTGAAGGGACTGCTGACCCA	2316
1F	-----	979
1C	GCGGCTCCTGGGCTGAGACCACCCCT-----GGTGGCGGG	1424
1D	GCAGTCTTGGTACACAGACGAGCCC-----	1532
1S	GAGGGCAATGCCAGAGGCCAGGCACCTCCTGCCCTGCCAGTGCCCCAGGGTGGAGTC	2376
1F	-----TCACCTGTCTGCACGTGCCTGGAACCC-----	1006
1C	GGCAGCAGCGCCCGGAGAGTCCGGCCGTCTCCCTCATGGTGCCAGCCAGG-----	1479
1D	---GACATCTCTACCGACTTTCACACCAGCCAGCCTGACTGTCCCCAGCAGCT-----	1584
	*   * *   * * * *	
1S	CTCCATGCCTGAGGACAGAAAGAGCTCCACACCAGGGTCTTTCATGAGGAGACACCCCA	2436
1F	-----ACTCGGACCCAGCC-----	1021
1C	-----CTGGGGCCCCAGGGA-----	1494
1D	-----TCCGGAACAAAAACA-----	1599
	*   *	
1S	CAGCAGGAGCACCAGGGAGAATACTTCCAGGTGCTCAGCACAGCTACAGCCCTGCTGAT	2496
1F	-----ATGGGAAGAGGGGCAGTGCCGACAGCTTGGT	1052
1C	-----GGCAGTCCACGGCAGTGCCAGCAGCTGGT	1525
1D	-----GCGACAAGCAGAGGAGTGGGACAGCTTGGT	1630
	*   *   *   * * *	
1S	CAAAAGGCTCTGGTTCGAGGGGGCCTGGGCACCTTGGCAGCTGATGCAAACCTTCATCAT	2556
1F	GGAGGCTGTGCTTATCTCAGAGGGTCTGGGCCTCTTTGCTCGAGACCCACGTTTCTGTGGC	1112
1C	GGAAGCGTCTTGATTTCAAGAGGACTGGGGCAGTTTGTCAAGATCCCAAGTTTCATCGA	1585
1D	GGAGGCAGTCTGATATCCGAAGGCTTGGGACGCTATGCAAGGGACCCAAATTTGTGTCT	1690
	*   *   *   *   *   * *   * * *   *   * *   *   * *   *   * *   *	
1S	GGCAACAGGCCAGGCCCTGGCAGATGCCTGCCAAATGGAACCAGAGGAAGTGGAGATCAT	2616
1F	CCTGGCCAAGCAGGAGATTGCAGATGCGTGTGCGCTGACGCTGGATGAGATGGACAATGC	1172
1C	GGTCACCCACAGGAGCTGGCCGACGCCTGCGACATGACCATAGAGGAGATGGAGAGCGC	1645
1D	AGCAACAAAACGAAATCGCTGATGCCTGTGACCTACCATCGACGAGATGGAGAGTGC	1750
	*   * * *   * * * *   *   *   * * *   * * *   * * *   * * *   *	
1S	GGCAACAGAGCTACTGAAAGGACGAGAGGCCCCAGAGGGCATGGCCA-----	2663
1F	TGCCAGTGACCTGCTGGCACAGGGAACCAGCTCTCTCTATAGCGACGAGGAGTCCATCCT	1232
1C	GGCCGACAACATCCTCAGCGGGGGCGCCCCACAGAGCCCAATGGCGCCTCTTACCCTT	1705
1D	AGCCAGCACCTGCTAATGGGAACGTGCGTCCCCGAGCCAACGGGGATGTGGGCCCCCT	1810
	* *   *   * *   *   *   *	
1S	-----GCTCCCTGGGATG-----CCTGAACCT	2685
1F	CTCCCGCTTCGATGAGGAGGACTTGG-----GAGACGAGAT	1268



## Appendix 11. Python code to generate graphs for chapter 4

```

## Python code randomly resample 1000 times from introns/ exons
###import packages and programs to python###
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('ggplot')
import seaborn as sns

###open PhyloP score file of 5'UTR of 3' transcripts of CACNA1F###
dfu = pd.read_csv('1f_comb_5utr.csv')
### calculate median PhyloP score of 5'UTR###
median= np.median(dfu.Score)
print(median)

###open PhyloP score file of exons###
FET = pd.read_csv("1f_comb_exonT.csv")
FEC = pd.read_csv("1f_comb_exonC.csv")
FEG = pd.read_csv("1f_comb_exonG.csv")
FEA = pd.read_csv("1f_comb_exonA.csv")

###Random resampling###
boot_sample_mediansE = []
for i in range (1000):
    FETS = FET.groupby('Base').apply(lambda x:x.sample(11))
    FECS = FEC.groupby('Base').apply(lambda x:x.sample(12))
    FEGS = FEG.groupby('Base').apply(lambda x:x.sample(6))
    FEAS = FEA.groupby('Base').apply(lambda x:x.sample(7))
    frames= [FETS, FECS, FEGS, FEAS]
    fexonboot = pd.concat(frames, keys=["FETS", "FECS", "FEGS", "FEAS"], ignore_index=True, sort=False)
    boot_medianE = np.median(fexonboot.Score)
    boot_sample_mediansE.append(boot_medianE)
median_median = np.median(boot_sample_mediansE) ###calculate median
median_stdE = np.std(boot_sample_mediansE) ###calculate standard deviation
boot_median_CIE = np.percentile(boot_sample_mediansE, [2.5, 97.5]) #95% CI
print(boot_median_CIE)
print(median_stdE)
print(median_median)
###create figure###
plt.hist(boot_sample_mediansE, alpha = 1, color = 'grey')
plt.axvline(np.percentile(boot_sample_mediansE,2.5), color = 'blue',linewidth=2)
plt.axvline(np.percentile(boot_sample_mediansE,97.5), color = 'blue', linewidth = 2)
plt.axvline(median_median, color = 'black', linewidth = 2)
plt.axvline(4.95, color = 'purple', linewidth = 2)
plt.xlabel("Median PhyloP score")
plt.ylabel("Resamples")

###open PhyloP score file of introns###
FIT = pd.read_csv("1FintronT.csv")
FIC = pd.read_csv("1FintronC.csv")
FIG = pd.read_csv("1FintronG.csv")
FIA = pd.read_csv("1FintronA.csv")

###Random resampling###
boot_sample_medians = []

```

```

for i in range (1000):
    FITS = FIT.groupby('Base').apply(lambda x:x.sample(11))
    FICS = FIC.groupby('Base').apply(lambda x:x.sample(12))
    FIGS = FIG.groupby('Base').apply(lambda x:x.sample(6))
    FIAS = FIA.groupby('Base').apply(lambda x:x.sample(7))
    frames= [FITS, FICS, FIGS, FIAS]
    fintronboot = pd.concat(frames, keys=["FITS", "FICS", "FIGS", "FIAS"], ignore_index=True, sort=False)
    boot_median = np.median(fintronboot.Score)
    boot_sample_medians.append(boot_median)
print(fintronboot)
median_median = np.median(boot_sample_medians) ###calculate median
median_std = np.std(boot_sample_medians) ###calculate standard deviation
boot_median_CI = np.percentile(boot_sample_medians, [2.5, 97.5]) #95%CI
print(median_median)
print(boot_median_CI)
print(median_std)
###create figure###
plt.hist(boot_sample_medians, alpha = 1, color = 'grey')
plt.axvline(np.percentile(boot_sample_medians,2.5), color = 'blue',linewidth=2)
plt.axvline(np.percentile(boot_sample_medians,97.5), color = 'blue', linewidth = 2)
plt.axvline(median_median, color = 'black', linewidth = 2)
plt.axvline(4.95, color = 'purple', linewidth = 2)
plt.xlabel("Median PhyloP score")
plt.ylabel("Resamples")

###Import packages and programs to python###
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
plt.style.use('ggplot')
import seaborn as sns
import matplotlib.path_effects as path_effects
from utils import *
from scipy.stats import mannwhitneyu, normaltest
import statannot as statannot
from statannot import add_stat_annotation
from statannotations.Annotator import Annotator

###read exon expression file in csv###
F = pd.read_csv("If gtex 40 41 42 sum expression level.csv")
sns.set_theme(style="ticks")
f, ax = plt.subplots(figsize=(8,9))
ax.set_yscale("log")
sns.set_theme(style="ticks")
sns.set(font_scale = 1.5)

# Plot the orbital period with horizontal boxes
x = "Tissue type"
y = "Level"
hue = "Exon"
order = ['Not brain', 'Brain']
f2 = sns.boxplot(data=F, x=x, y=y, hue=hue, order=order, palette="vlag", width = 0.8)
f2.set_xticklabels(f2.get_xticklabels(), rotation=40, ha="right")
plt.subplots_adjust(wspace=2, hspace=6)
plt.tight_layout()

```

```

add_stat_annotation(f2, data=F, x=x, y=y, hue=hue, order=order, box_pairs=[(("Not brain", "Exon 40"), ("Not
brain", "Exon 41")), (("Not brain", "Exon 41"), ("Not brain", "Exon 42")), (("Not brain", "Exon 40"), ("Not brain",
"Exon 42")), (("Brain", "Exon 40"), ("Brain", "Exon 41")), (("Brain", "Exon 41"), ("Brain", "Exon 42")), (("Brain",
"Exon 40"), ("Brain", "Exon 42"))], test="Mann-Whitney", text_format="star", loc="inside", verbose = 2).
plt.legend(loc="upper left", bbox_to_anchor=(1.30, 1))
f2.yaxis.grid(False)
f2.xaxis.grid(False)
f2.set(yscale="log")
f2.set(ylabel="Log10 (expression level)")
sns.despine(trim=True, left=True)

```

```

###Plot the fold changes###

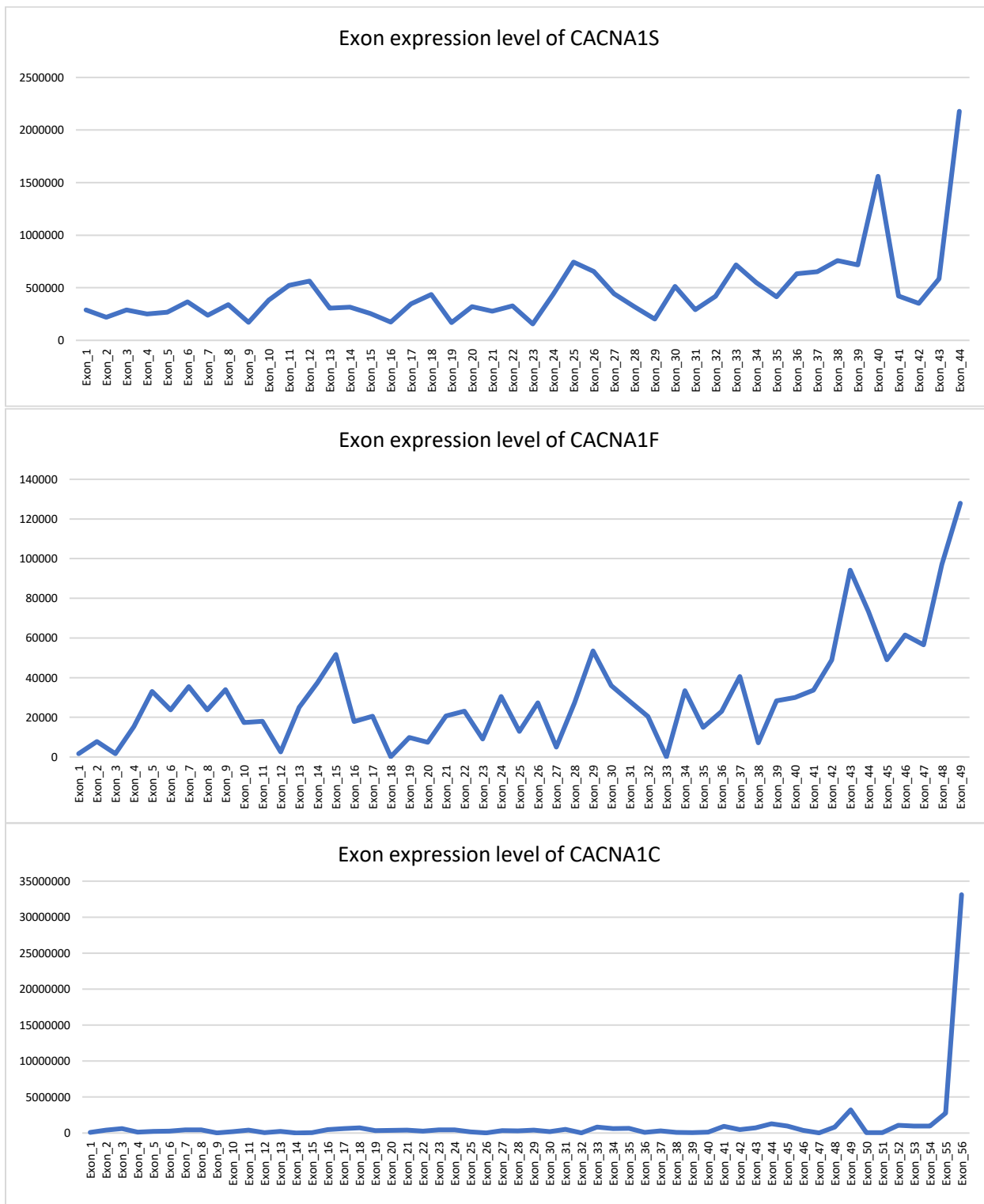
```

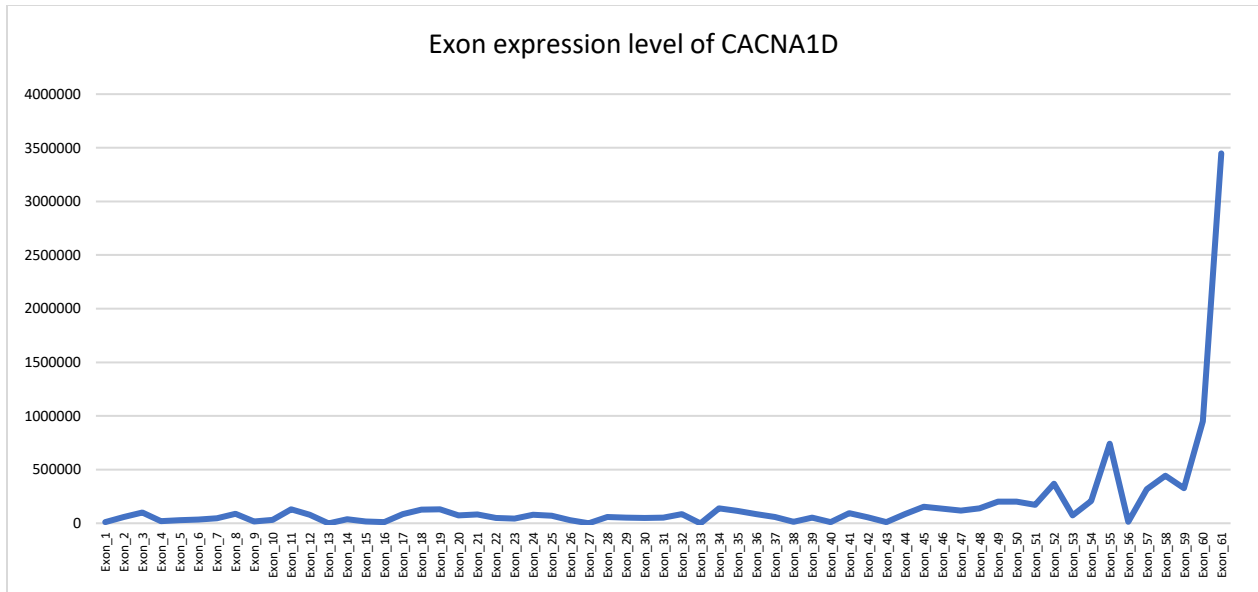
```

Ffold = pd.read_csv("1f fold change 40 41 42.csv") ###import foldchange data in csv file###
gfold1 = sns.catplot(x="Exon", y="Fold change", col="Tissue type",
                    data=Ffold, kind="box", palette="vlag",
                    height=6, aspect=.7)
gfold1.set(ylabel="Fold Change of Exon Expression")
gfold1.set(xlabel="Exon")

```

## Appendix 12. Examining the potential 3' bias from GTEx data





## Appendix 13

### 13.1 Ligation of inserts to the dephosphorylated expression vector for chapter 5

Name	Molar ratio (Insert: Vector)	Calculation of insert (ng)	How much to mix
1s	1:1	$(100 \times 2.75) / 5.4 \times 1/1 = 50.93$	Insert: 0.93 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 5.67 $\mu$ l
	2:1	$(100 \times 2.75) / 5.4 \times 2/1 = 101.86$	Insert: 1.87 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 4.73 $\mu$ l
	3:1	$(100 \times 2.75) / 5.4 \times 3/1 = 152.79$	Insert: 2.8 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 3.8 $\mu$ l
1c	1:1	$(100 \times 1.64) / 5.4 \times 1/1 = 30.37$	Insert: 0.43 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 6.17 $\mu$ l
	2:1	$(100 \times 1.64) / 5.4 \times 2/1 = 60.74$	Insert: 0.86 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 5.74 $\mu$ l
	3:1	$(100 \times 1.64) / 5.4 \times 3/1 = 91.11$	Insert: 1.30 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 5.3 $\mu$ l
1d	1:1	$(100 \times 1.89) / 5.4 \times 1/1 = 35$	Insert: 0.67 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 5.93 $\mu$ l
	2:1	$(100 \times 1.89) / 5.4 \times 2/1 = 70$	Insert: 1.34 $\mu$ l Vector: 2.07 $\mu$ l 10x buffer = 1 $\mu$ l Ligase = 0.33 $\mu$ l Water: 5.26 $\mu$ l
	3:1	$(100 \times 1.89) / 5.4 \times 3/1 = 105$	Insert: 2.02 $\mu$ l



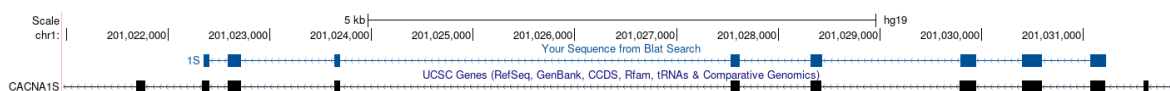
			Vector: 2.07µl 10x buffer = 1µl Ligase = 0.33µl Water: 4.58µl
1f	1:1	$(100*1.25) / 5.4 * 1/1 = 23.15$	Insert: 1.8µl Vector: 2.07µl 10x buffer = 1µl Ligase = 0.33µl Water: 4.8µl
	2:1	$(100*1.25) / 5.4 * 2/1 = 46.3$	Insert: 3.62µl Vector: 2.07µl 10x buffer = 1µl Ligase = 0.33µl Water: 2.98µl
	3:1	$(100*1.25) / 5.4 * 3/1 = 69.45$	Insert: 5.43µl Vector: 2.07µl 10x buffer = 1µl Ligase = 0.33µl Water: 1.7µl

### 13.2 Ligation method of inserts to the non-dephosphorylated expression vector

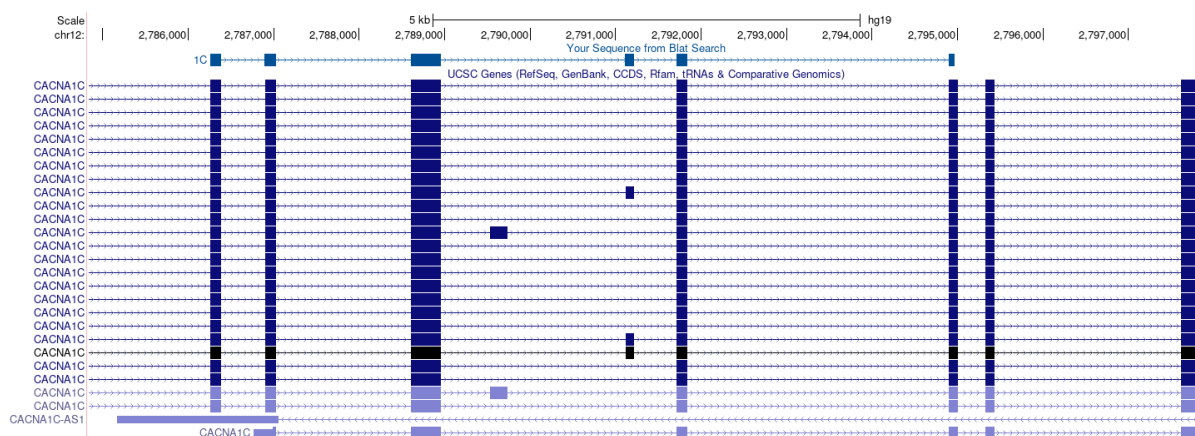
Name	Molar ratio (insert: vector)	Calculation of insert (ng) = same as above	How much to mix
1s	1:1	$(100*2.75) / 5.4 * 1/1 = 50.93$	Insert: 0.93µl Vector: 2.2µl 10x buffer = 1µl Ligase = 0.33µl Water: 5.54µl
	2:1	$(100*2.75) / 5.4 * 2/1 = 101.86$	Insert: 1.87µl Vector: 2.2µl 10x buffer = 1µl Ligase = 0.33µl Water: 4.6µl
	3:1	$(100*2.75) / 5.4 * 3/1 = 152.79$	Insert: 2.8µl Vector: 2.2µl 10x buffer = 1µl Ligase = 0.33µl Water: 3.67µl
1c	1:1	$(100*1.64) / 5.4 * 1/1 = 30.37$	Insert: 0.43µl Vector: 2.2µl 10x buffer = 1µl Ligase = 0.33µl Water: 6.04µl
	2:1	$(100*1.64) / 5.4 * 2/1 = 60.74$	Insert: 0.86µl Vector: 2.2µl 10x buffer = 1µl Ligase = 0.33µl

			Water:	5.61µl
	3:1	$(100*1.64) / 5.4 * 3/1 = 91.11$	Insert: Vector: 10x buffer = Ligase = Water:	1.30µl 2.2µl 1µl 0.33µl 5.17µl
1d	1:1	$(100*1.89) / 5.4 * 1/1 = 35$	Insert: Vector: 10x buffer = Ligase = Water:	0.67µl 2.2µl 1µl 0.33µl 5.8µl
	2:1	$(100*1.89) / 5.4 * 2/1 = 70$	Insert: Vector: 10x buffer = Ligase = Water:	1.34µl 2.2µl 1µl 0.33µl 5.3µl
	3:1	$(100*1.89) / 5.4 * 3/1 = 105$	Insert: Vector: 10x buffer = Ligase = Water:	2.02µl 2.2µl 1µl 0.33µl 4.45µl
1f	1:1	$(100*1.25) / 5.4 * 1/1 = 23.15$	Insert: Vector: 10x buffer = Ligase = Water:	1.8µl 2.2µl 1µl 0.33µl 4.67µl
	2:1	$(100*1.25) / 5.4 * 2/1 = 46.3$	Insert: Vector: 10x buffer = Ligase = Water:	3.62µl 2.2µl 1µl 0.33µl 2.85µl
	3:1	$(100*1.25) / 5.4 * 3/1 = 69.45$	Insert: Vector: 10x buffer = Ligase = Water:	5.43µl 2.2µl 1µl 0.33µl 1.04µl

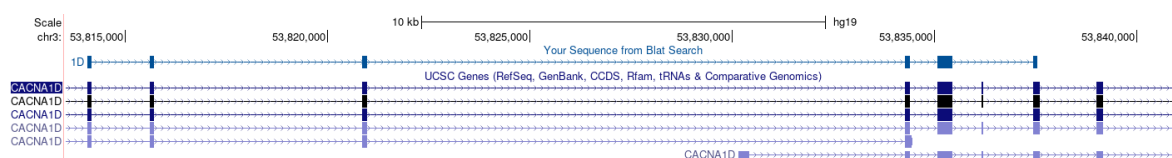


14.2. Sanger sequence result of the 3' transcript of *CACNA1S* was blatted to protein**voltage-dependent L-type calcium channel subunit alpha-1S [Homo sapiens]**Sequence ID: [NP\\_000060.2](#) Length: 1873 Number of Matches: 3[See 2 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)Range 1: 963 to 1282 [GenPept](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Matc](#)

Score	Expect	Method	Identities	Positives	Gaps
611 bits(1575)	0.0	Compositional matrix adjust.	320/320(100%)	320/320(100%)	0/320(0%)
Query 1		MTEEECRGYYYYYKDGDPMQIELRHREWYHSDHFHFDNVLSAMMSLFTYSTFEGWPQLLYK			60
Sbjct 963		MTEEECRGYYYYYKDGDPMQIELRHREWYHSDHFHFDNVLSAMMSLFTYSTFEGWPQLLYK			1022
Query 61		AIDSNAEDVGP IYNNRYEMAiffiiyiiil affMMNIFVGFYIVTFQEQGETEYKNCELD			120
Sbjct 1023		AIDSNAEDVGP IYNNRYEMAiffiiyiiil affMMNIFVGFYIVTFQEQGETEYKNCELD			1082
Query 121		KNQRQCYQYALKARPLRCYIPKNPYQYQYWIYVTSSYFEYLMFALIMLNTICLGMQHYNQ			180
Sbjct 1083		KNQRQCYQYALKARPLRCYIPKNPYQYQYWIYVTSSYFEYLMFALIMLNTICLGMQHYNQ			1142
Query 181		SEQMNHISDILNYAFTIIFTLEMILKLMAFKARGYFGDPWNYFDLIVIGSIDVILSEI			240
Sbjct 1143		SEQMNHISDILNYAFTIIFTLEMILKLMAFKARGYFGDPWNYFDLIVIGSIDVILSEI			1202
Query 241		DTFCLASSGglyclgggcnYDPDESARISSAFFRLFRYMRLIKLLSRAEGVRTLLWTFIK			300
Sbjct 1203		DTFCLASSGGLYCLGGGCGNYDPDESARISSAFFRLFRYMRLIKLLSRAEGVRTLLWTFIK			1262
Query 301		SFQALPYVALLIYMLFFIYA 320			
Sbjct 1263		SFQALPYVALLIYMLFFIYA 1282			

14.3. Sanger sequence result of the 3' transcript of *CACNA1C* was blotted to protein**voltage-dependent L-type calcium channel subunit alpha-1C isoform 1 [Homo sapiens]**Sequence ID: [NP\\_955630.3](#) Length: 2221 Number of Matches: 1Range 1: 1646 to 1964 [GenPept](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Match](#)

Score	Expect	Method	Identities	Positives	Gaps
639 bits(1647)	0.0	Compositional matrix adjust.	318/319(99%)	319/319(100%)	0/319(0%)
Query 1	MKLLDQVYPPAGDDEVTVGKIFYATFLIQEYFRKFKKRKEQGLVGKPSQRNALSLQAGLRT				60
Sbjct 1646	MKLLDQVYPPAGDDEVTVGKIFYATFLIQEYFRKFKKRKEQGLVGKPSQRNALSLQAGLRT				1705
Query 61	LHDIGPEIRRAISGDLTAEELDKAMKEAVSAASEDDIFRRAGGLFGNHYSYYQSDGRSA				120
Sbjct 1706	LHDIGPEIRRAISGDLTAEELDKAMKEAVSAASEDDIFRRAGGLFGNHYSYYQSDGRSA				1765
Query 121	FPQTFITQRPLHINKAGSSQGDTEPSHEKLYDSTFTPSSYSSTGSnaninnanntaLGR				180
Sbjct 1766	FPQTFITQRPLHINKAGSSQGDTEPSHEKLYDSTFTPSSYSSTGSNANINNANNTALGR				1825
Query 181	LPRPAGYPTVSTVEGHGPPLSPAIRVQEVAWKLSNRERHYPMCEDLELRDSSGSAGTQ				240
Sbjct 1826	LPRPAGYPTVSTVEGHGPPLSPAIRVQEVAWKLSNRERHYPMCEDLELRDSSGSAGTQ				1885
Query 241	AHCLLLRKANPSRCHSRESQAAMAGQEETSQDETYEYKMNHDTEACSEPSLLSTEMLSYQ				300
Sbjct 1886	AHCLLLRRANPSRCHSRESQAAMAGQEETSQDETYEYKMNHDTEACSEPSLLSTEMLSYQ				1945
Query 301	DDENRQLTLPEEDKRDIRQ				319
Sbjct 1946	DDENRQLTLPEEDKRDIRQ				1964

14.4. Sanger sequence result of the 3' transcript of *CACNA1D* was blatted to protein**calcium voltage-gated channel subunit alpha1 D [Homo sapiens]**Sequence ID: [KAI2530071.1](#) Length: 2170 Number of Matches: 1[See 1 more title\(s\)](#) [See all Identical Proteins\(IPG\)](#)Range 1: 1543 to 1855 [GenPept](#) [Graphics](#)[▼ Next Match](#) [▲ Previous Match](#)

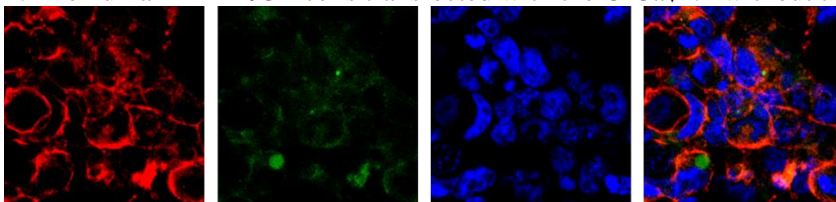
Score	Expect	Method	Identities	Positives	Gaps
608 bits(1567)	0.0	Compositional matrix adjust.	313/313(100%)	313/313(100%)	0/313(0%)
Query 1		MNMP LNSDGT YMFNATL FALV RTALK I KTEGNLEQ ANEELRAVIK KIWK KTSMKLLDQYV			60
Sbjct 1543		MNMP LNSDGT YMFNATL FALV RTALK I KTEGNLEQ ANEELRAVIK KIWK KTSMKLLDQYV			1602
Query 61		PPAGDDEVT YGK FYATFL IQDYFRKF KKRKEQGLV GKYP AKNTT IALQAGLR TLHDIGPE			120
Sbjct 1603		PPAGDDEVT YGK FYATFL IQDYFRKF KKRKEQGLV GKYP AKNTT IALQAGLR TLHDIGPE			1662
Query 121		IRRA I SCDLQddepeet kreeddYFKRNGALLGNHYNHYNSDRRDSLQQTNTTHRPLHY			180
Sbjct 1663		IRRA I SCDLQddepeet kreeddYFKRNGALLGNHYNHYNSDRRDSLQQTNTTHRPLHY			1722
Query 181		QRPS IPPASDTEKPLFPPAGNSYChnhhnhnS IGKQYPTSTNANLNNANMSKAAHGKRPS			240
Sbjct 1723		QRPS IPPASDTEKPLFPPAGNSYChnhhnhnS IGKQYPTSTNANLNNANMSKAAHGKRPS			1782
Query 241		IGNLEHYSENGHHSSHKH DREPQRRSSYKRTRY YET YIRSDSGDEQLPT ICREDPE IHGY			300
Sbjct 1783		IGNLEHYSENGHHSSHKH DREPQRRSSYKRTRY YET YIRSDSGDEQLPT ICREDPE IHGY			1842
Query 301		FRDPHCLGEQEYF 313			
Sbjct 1843		FRDPHCLGEQEYF 1855			

### Appendix 15. Immunofluorescence imaging of transfected HEK293T cells and SHSY5Y cells to test specificity of the secondary antibody.

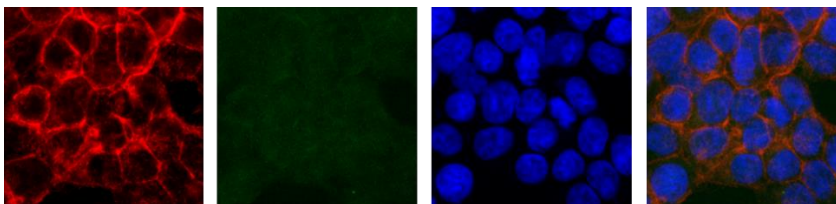
Cells were stained with Phalloidin and mounted with DAPI. The red channel is Phalloidin, the blue channel is DAPI, and the green channel is flag tag.

15.1. Immunofluorescence imaging of transfected HEK293T cells to test specificity of the secondary antibody

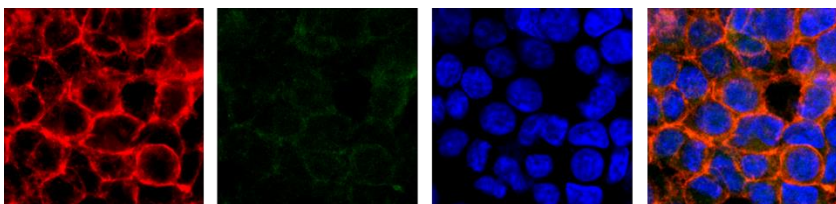
A. The human HEK293T cells transfected with the C-Cav1.1 without the primary antibody



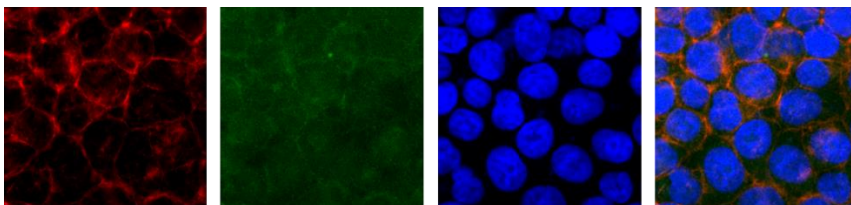
B. The human HEK293T cells transfected with the C-Cav1.2 without the primary antibody



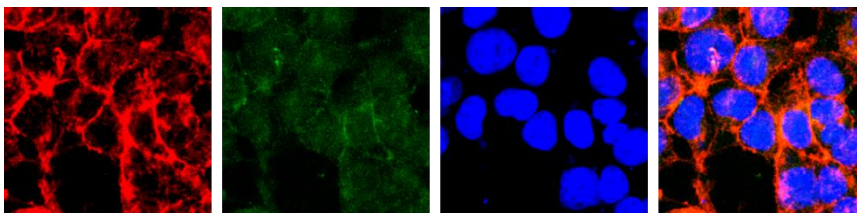
C. The human HEK293T cells transfected with the C-Cav1.3 without the primary antibody



D. The human HEK293T cells transfected with the C-Cav1.4 without the primary antibody

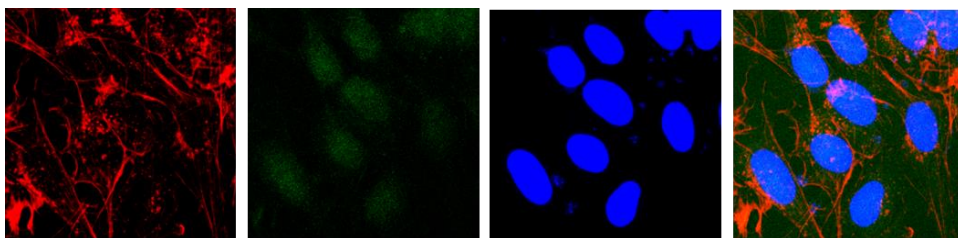


E. The human HEK293T cells transfected with the empty C-flag pcDNA3 the primary antibody

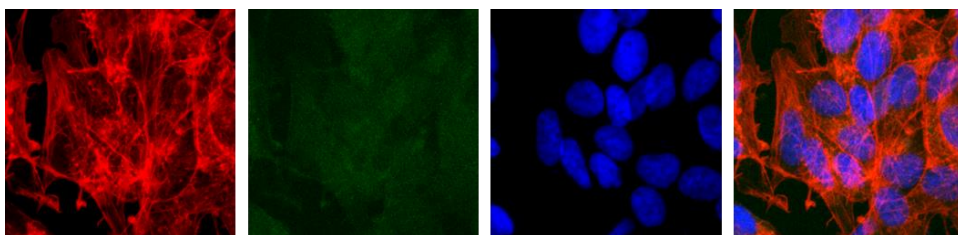


15.2. Immunofluorescence imaging of transfected SH-SY5Y cells to test specificity of the secondary antibody

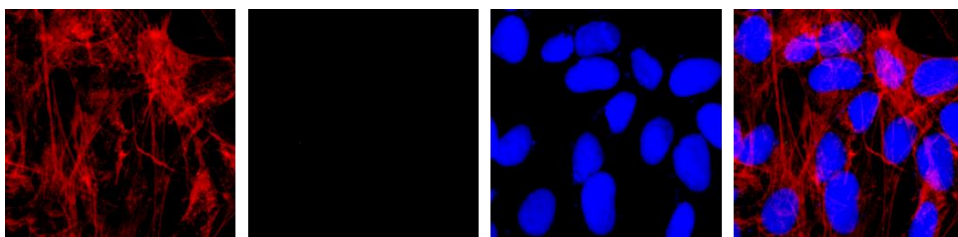
A. The human SH-SY5Y cells transfected with the C-Cav1.1 without the primary antibody



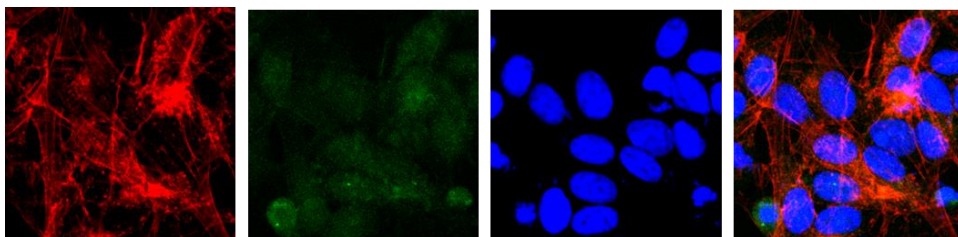
B. The human SH-SY5Y cells transfected with the C-Cav1.2 without the primary antibody



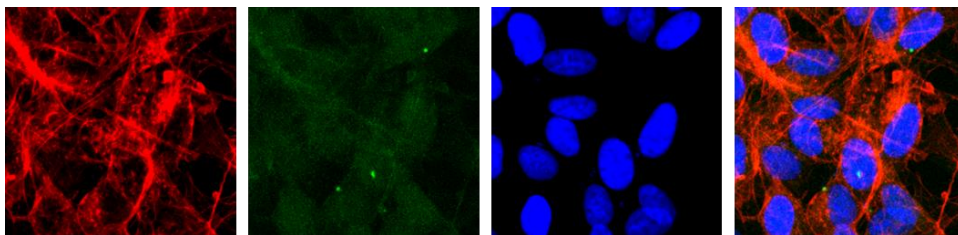
C. The human SH-SY5Y cells transfected with the C-Cav1.3 without the primary antibody



D. The human SH-SY5Y cells transfected with the C-Cav1.4 without the primary antibody



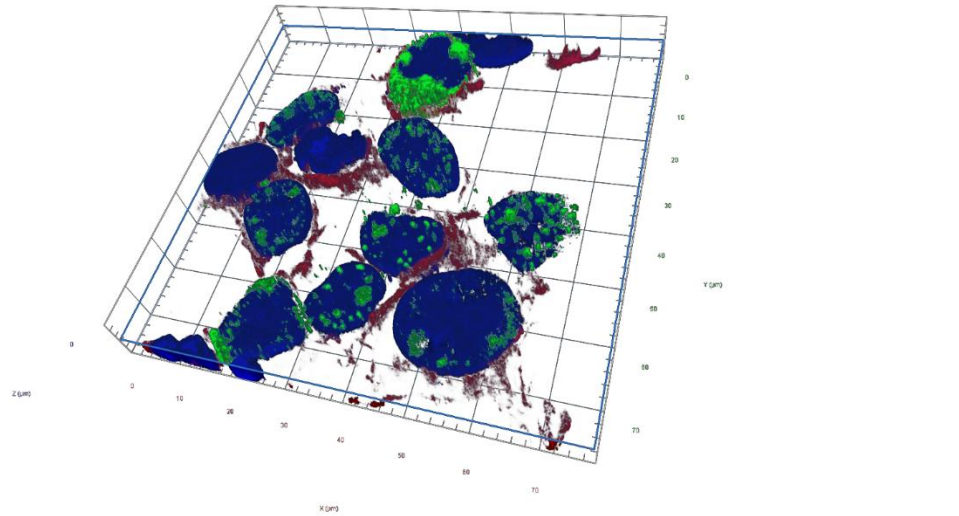
E. The human SH-SY5Y cells transfected with the empty c-flag pcDNA3 without the primary antibody



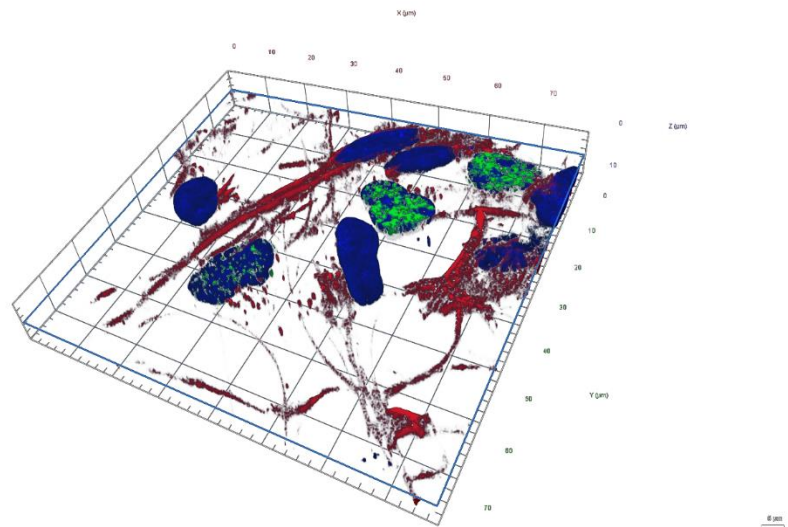


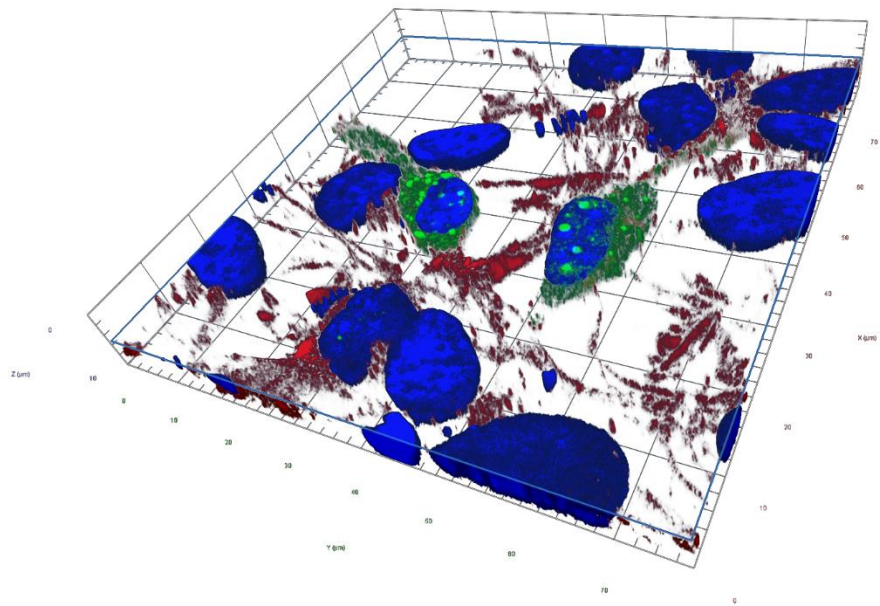
## Appendix 16. 3D image of confocal immunofluorescence images

### 16.1. 3D half cut image of C-Cav1.3 in HEK293T cells



### 16.2 3D half cut image of C-Cav1.3 in shsy5y cells



16.3 3D half cut image of C-Ca<sub>v</sub>1.2 in shsy5y cells

**Appendix 17. The results of both the short-length and the full-length PCR amplification of VGCC genes.**

Gene	Short fragment amplification	Full length CDS amplification	Expression in human brain based on LIBD data and GTEx data
<i>CACNA1A</i>	Yes	Yes	Yes
<i>CACNA1B</i>	Yes	No	Yes
<i>CACNA1E</i>	Yes	Yes	Yes
<i>CACNA1F</i>	Yes	No for full-length, but yes to truncated <i>CACNA1F</i> predicted based on LIBD data	Yes
<i>CACNA1G</i>	Yes	No	Yes
<i>CACNA1H</i>	No	No	No
<i>CACNA1I</i>	Yes	No	Yes
<i>CACNA1S</i>	Yes	No for full-length, but yes to truncated <i>CACNA1S</i> predicted based on LIBD data	No
<i>CACNB1</i>	Yes	Yes	Yes
<i>CACNB2</i>	Yes	Yes	Yes
<i>CACNB3</i>	Yes	Yes	Yes
<i>CACNB4</i>	Yes	Yes	Yes
<i>CACNA2D1</i>	Yes	Yes	Yes
<i>CACNA2D2</i>	No	No	No
<i>CACNA2D3</i>	Yes	No	Yes
<i>CACNA2D4</i>	No	No	No