Linkage analysis in a familial case of idiopathic epilepsy

and its implications in drug development

Maria Louisa Cassar

Thesis submitted to the University of Malta for the degree of Doctor of Philosophy

September 2008
To my husband Brian, my children Mikael, Gabriel, Samuel and my family
'Success is the ability to go from failure to failure without losing enthusiasm.'

Winston Churchill
Acknowledgements

I thank my supervisor, Dr. Janet Mifsud, for her support and guidance throughout these years and all the academic and support staff at the Department of Clinical Pharmacology and Therapeutics. My sincere thanks to the Head of the Department Prof Roger Ellul Micallef.

A sincere gratitude to my husband Brian who supported me during the difficult moments and who helped me around the house while I spent days in the laboratory or at the computer. To my older boys, Mikael and Gabriel, who helped a lot by babysitting the young one Samuel, to my mum and my aunt Helen who were always there to help me and to my father from whom I inherited the ambition to always strive to achieve the best. I would like to thank my brother Stephen for computer technical support.

This study would not have been possible without Dr. Doriette Soler who identified the family. I thank the family members who accepted to participate in the study. I thank Dr. Franz Rüschendorf who offered his full support in the use of the Alohomora software package and my work colleague Dr. Christopher Vidal for the long discussions we had on linkage analysis.

Finally yet importantly I would like to thank all those who supported and encouraged me throughout these years from when I first registered as an MPhil student at the University of Malta in 2002.
# Table of Contents

## Acknowledgements

## Table of Contents

## List of Tables

## List of Figures

## List of Software

## Abstract

## List of Abbreviations

## Chapter 1 - Introduction

1.1 Introduction

1.2 Epilepsy – historical perspective

1.3 Definition of epilepsy

1.4 Pathophysiology of epilepsy

1.5 Aetiology of epilepsy

1.6 Epidemiology of epilepsy

1.7 Classification of seizure types and syndromes

1.8 Inheritance factors associated with epilepsy

1.9 Genetics of the idiopathic epilepsies

1.10 Genetics of epilepsy and drug therapy
1.11 Methodological approaches for identification of genes causing epilepsy 31
1.12 Linkage analysis in genetic studies 32
1.13 Family recruiting for genetic studies 34
1.14 Genotyping in genetic studies 35
  1.14.1 SNP genotyping in genetic studies 35
  1.14.2 Hybridization of DNA strands to the GeneChip® 38
  1.14.3 Data analysis of SNP genotyping 39
1.15 STRs genotyping in linkage studies 39
1.16 Identifying the causative gene in genetic disorders 41
1.17 Aims of the study 43

Chapter 2 - Methodology

2.1 Introduction to methodology 47
2.2 Recruitment of the family with epilepsy 47
2.3 Syndrome characterization 49
2.4 DNA Extraction 50
  2.4.1 Sample collection from affected and non-affected members of the family 50
  2.4.2 DNA extraction from blood samples 50
  2.4.3 DNA Extraction from buccal rinse 51
  2.4.4 Determination of the quality of DNA 53
    2.4.4.1 Estimation of the concentration of DNA extracted from buccal rinse 53
    2.4.4.2 DNA purity 54
2.5 SNP genotyping 55
  2.5.1 Purification of DNA for SNP genotyping 55
  2.5.2 SNP genotyping using GeneChip® Mapping 250K (NspI) Assay Kit 56
2.6 Linkage analysis of SNP genotypes 58
2.7 Short tandem repeat (STR) genotyping for the confirmation of an identified linkage interval 59
  2.7.1 Amplification of STRs 59
  2.7.2 Electrophoresis of the fluorescently labelled PCR products 61
  2.7.3 Linkage analysis of the STR genotyping data 62
2.8 Sequencing of candidate genes in the linkage interval on Chromosome 20 62
  2.8.1 PCR amplification of exons of genes in the linkage interval 63
2.8.2 PCR product purification 67
2.8.3 DNA sequencing 68
2.8.4 Purification of the cycle sequencing products 69
2.8.5 Electrophoresis and analysis of the cleaned cycle sequencing products 70
2.8.6 Alignment of DNA Sequences 71
2.8.7 Identification of the affected allele 71
2.9 Population study to determine the frequency of the affected allele 72
2.10 Prediction of potential functional effects of identified variation 73
2.11 Investigating alternative splicing 74
  2.11.1 The pSPL3 plasmid vector 74
  2.11.2 Designing restriction sites in the PCR product 76
  2.11.3 Restriction digests of PCR product and plasmid 78
2.12 Transformation of competent cells 79
  2.12.1 Preparation of competent E.Coli cells 79
  2.12.2 Ligation reaction of digested insert and digested pSPL3 80
  2.12.3 Transformation reaction 81
  2.12.4 Extraction of plasmid DNA from bacterial cell cultures using 83
  2.12.5 Determination of the presence of insert in pSPL3 84
2.13 Sequencing of the mRNA transcript of the CHRNA4 84
  2.13.1 RNA extraction from leukocytes 84
  2.13.2 RNA Transcription 86
  2.13.3 cDNA amplification of trascript of CHRNA4 of patient and control 87
2.14 Copy Number Variation (CNV) analysis 89
2.15 Conclusion to methodology 89

Chapter 3 - Results

3.1 Introduction to results 92
3.2 Syndrome characterization in the family 93
3.3 Analysis of the genome-wide SNP data 98
  3.3.1 SNP genotyping 99
  3.3.2 Quality control of SNP genotypes 101
3.4 Linkage analysis of SNP genotyping data 102
3.4.1 Phenotype A – Family members who have had a witnessed seizure were considered as affected 102
3.4.2 Phenotype B – Family members who had had febrile seizures before the age of six were considered as affected 114
3.5 The best candidate region 117
3.6 Gene sequence analysis 132
3.7 Analysis of variations observed in Exon 5 of the CHRNA4 gene that differ between the affected and non-affected individuals 136
3.7.1 The synonymous SNPs in Exon 5 of CHRNA4 136
3.7.2 The variants of intron 4 of CHRNA4 138
3.7.3 Pre-mRNA folding prediction 141
3.7.4 Alternative splicing prediction 144
3.8 The minigene vector experiment 147
3.9 Sequencing of the CHRNA4 transcript 148
3.10 Copy Number Variation analysis 148
3.11 Conclusion 151

Chapter 4 - Discussion

4.1 Introduction 154
4.2 Syndrome in the recruited family 155
4.3 Linkage analysis of SNP genotyping data 157
4.4 The linkage interval on chromosome 20 161
4.5 CHRNA4 haplotype 163
4.6 Synonymous SNPs in Exon 5 of CHRNA4 165
4.7 Variations observed in intron 4 of CHRNA4 166
4.7.1 Prediction of secondary structure of pre-mRNA 167
4.7.2 Alternative splicing 170
4.7.3 Analysis of the cDNA transcript of CHRNA4 172
4.8 Copy Number Variation analysis 174
4.9 Genetic studies in epilepsy and the development of novel drug therapies 175
4.9.1 Drug targeting 176
4.9.2 Drug response in epilepsy 177
4.9.3 Gene therapy in epilepsy 178
4.10 Future work

4.11 Conclusion

References

Appendix 1 – Classification of epilepsy syndromes
Appendix 2 – Patient information leaflet (English and Maltese versions)
Appendix 3 – Patient Consent Form - Adult (over 18 years)
Appendix 4 – Patient Consent Form - Child (under 18 years)
List of Tables

Chapter 1 - Introduction

Table 1.1. Inherited diseases that have seizures as one of their symptoms 18
Table 1.2. Idiopathic epilepsy caused by mutations in voltage-gated or ligand gated ion channels 23
Table 1.3. Genetic disorders that cause brain abnormalities and have epilepsy among their manifestations 25
Table 1.4. Loci that have been linked to febrile seizures 27
Table 1.5. Modes of action of some antiepileptic drugs 30

Chapter 2 - Methodology

Table 2.1. Reaction mix for the amplification of the mitochondrial hypervariable region 1 (HVI) 54
Table 2.2. Thermal profile for the amplification of HV1 55
Table 2.3. STR markers on chromosome 20. Genetic position is in cM according to the deCode map 60
Table 2.4. Reagent Mix for the amplification of the STRs 61
Table 2.5. Thermal profile for the amplification of the STRs 61
Table 2.6. Reaction Mix for the amplification of the exons of CHRNA4, KCNQ2 and STMN3 63
Table 2.7. Reaction Mix for the amplification of the exons of CHRNA4, KCNQ2 and STMN3 that were not successfully amplified using the reaction mix in Table 2.6 64
Table 2.8. Thermal profile used for the amplification of the exons of CHRNA4, KCNQ2 and STMN3 64
Table 2.9. Sequences of primers for the amplification of exons of KCNQ2, size of PCR product and annealing temperature 65
Table 2.10. Sequences of primers for the amplification of exons of CHRNA4, size of PCR product and annealing temperature 66
Table 2.11. Sequences of primers for the amplification of exons of STMN3, size of PCR product and annealing temperature 67
Table 2.12. Cycle sequencing reaction mix 69
Table 2.13. Thermal cycling for cycle sequencing 69
Table 2.14. Reagents and volumes used to prepare 50ml of 8% polyacrylamide gel 72
Table 2.15. Primers used for amplifying the fragment having 400bp of intron 4, Exon 5 and 400bp of intron 5 77
Table 2.16. Reaction mix for the amplification of the insert using TLA DNA polymerase, a proof reading enzyme 77
Table 2.17. Thermal profile for the amplification of the insert 78
Table 2.18. Ligation reaction 80
Table 2.19. Samples and controls used in transformation 82
Table 2.20. Mix for reverse transcription 87
Table 2.21. Thermal profile used for reverse transcription 87
Table 2.22. Reaction mix for the cDNA amplification 88
Table 2.23. Thermal profile for the cDNA amplification 88

Chapter 3 - Results

Table 3.1. Disease status of family with age of onset and current health status of family members 96
Table 3.2. Chromosomal regions where the highest parametric LOD score was observed 107
Table 3.3. Multipoint linkage analysis of STRs on chromosome 20 showing parametric LOD and NPL scores 123
Table 3.4. Parametric LOD scores at different recombination fractions on chromosome 20 123
Table 3.5. The linkage analysis data where the NPL peaks showing npLOD, NPL, Zlr scores, markers and physical position of markers 124
Table 3.6. Genes (26) in the region of markers SNP_A-4214711 (rs1884328) and SNP_A-1933057 (rs1151621) 126
Table 3.7. Variations observed in the DNA sequence of CHRNA4, KCNQ2 and STMN3 in an affected and non-affected family member 133
Table 3.8. NetGene2 analysis of affected allele and Reference sequence
Table 3.9. Loss and gain at several chromosomal regions where individuals 2 and 6 showed variations when compared with the affected individuals as determined by the Genotyping Console™ software package

Chapter 4 - Discussion

Table 4.1. Cytogenetic band where the maximum parametric LOD was 2.67
Table 4.2. Haplotypes observed in CHRNA4
Table 4.3. Confidence level in the prediction of alternative donor and acceptor splice sites
Table of Figures

**Chapter 1 - Introduction**

- Figure 1.1. Preparation of genomic DNA for hybridization to the GeneChip®
- Figure 1.2. Principles involved in the fabrication of the GeneChip®
- Figure 1.3. Experiment design for identifying the causative gene

**Chapter 2 - Methodology**

- Figure 2.1. Pedigree of the family. Affected individuals are shown in black
- Figure 2.2. Transcript structure of KCNQ2 (Ensembl ENSG00000075043) showing 16 exons
- Figure 2.3. Transcript structure of CHRNA4 (Ensembl ENSG00000101204) showing 6 exons
- Figure 2.4. Transcript structure of STMN3 (Ensembl ENSG00000213603) showing 5 exons
- Figure 2.5. Splicing in the pSPL3 vector
- Figure 2.6. Plasmid pSPL3 showing restriction sites
- Figure 2.7. Graphical representation of Primer used for the amplification of the insert
- Figure 2.8. Primer pairs used for amplification of the transcript

**Chapter 3 - Results**

- Figure 3.1. Pedigree of the family studied
- Figure 3.2. A *.DAT file which shows the actual scan of the chip of Sample 8.
- Figure 3.3. A *.CEL file which shows the actual scan of the chip of Sample 8.
- Figure 3.4. Graphical representation of ‘Allele Call’ and confidence level as given by GTYPE software
- Figure 3.5. GRR of familial relationship between samples
- Figure 3.6. Parametric LOD scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%.
Figure 3.7. Zlr scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%.

Figure 3.8. NPL scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%.

Figure 3.9. Change in parametric LOD with change in penetrance.

Figure 3.10. HapMap of chromosome 7 between SNP_A-2137248 (rs4722286) and SNP_A-4226100 (rs2813838) (a region of 235kB) were parametric LOD was 2.67.

Figure 3.11. Variation of parametric LOD with variation in penetrance.

Figure 3.12. NPL scores on Chromosome 13 when penetrance was lowered to 0.7. In the region between 35cM and 50cM the NPL score is above 5 and this can be considered as significant evidence for linkage.

Figure 3.13. Parametric LOD scores on Chromosome 13 when penetrance was lowered to 0.7.

Figure 3.14. Parametric LOD scores on Chromosome 13 when penetrance was lowered to 0.5.

Figure 3.15. Chromosome 13 haplotype in chromosomal region where parametric LOD peaked.

Figure 3.16. Pedigree of the family studied. Affected individuals (family members that had febrile seizures before age 6) are shown in black.

Figure 3.17. Parametric LOD scores on Chromosome 3.

Figure 3.18. The haplotype for part of chromosomal region (3q21.3) where parametric LOD was 1.71.

Figure 3.19. Parametric LOD on chromosome 20 when inheritance pattern is taken as autosomal dominant inheritance with incomplete penetrance.

Figure 3.20. ZLR score on chromosome 20 when inheritance pattern is taken as autosomal inheritance with incomplete penetrance.

Figure 3.21. NPL scores on chromosome 20 when inheritance pattern is taken as autosomal inheritance with incomplete penetrance.
Figure 3.22. Haplotype observed in the family (the last five genotyped STRs on chromosome 20) 120
Figure 3.23. NPL scores on Chromosome 20 when using STRs markers 121
Figure 3.24. Parametric LOD scores on Chromosome 20 when using STRs markers 122
Figure 3.25. Chromosome 20 linkage interval 124
Figure 3.26. Haplotype of family showing the region where there was peaking of the ZLR score on chromosome 20 125
Figure 3.27. Expression of CHRNA4 in humans 129
Figure 3.28. Expression of KCNQ2 in humans 130
Figure 3.29. Expression of STMN3 in humans 131
Figure 3.30. Alignment of Affected allele with the Reference sequence ENST00000370263, region 61,451,218 to 61,451,339 135
Figure 3.31. DNA sequence of intron 4 of CHRNA4 in an affected individual 135
Figure 3.32. DNA sequencing of the region of Exon 5 of the CHRNA4 gene where there are the SNPs rs1044396 and rs1044397 137
Figure 3.33. Analysis of a region of Exon 5 of the CHRNA4 (coding sequence) using RESCUE-ESE 138
Figure 3.34. Alignment of part of the intron 4 of the CHRNA4 gene of a non-affected individual and the alternative assembly of chromosome 20, NW_001838671.1 139
Figure 3.35. Linkage disequilibrium in CHRNA4 gene (chromosome 20) in individuals of European origin (CEU) 141
Figure 3.36. Reference sequence (ENST00000370263) pre-mRNA folding (free energy of -δG of 217.23kcal/mol as given by mfold 142
Figure 3.37. Affected allele pre-mRNA folding (free energy of-δG of 253.15kcal/mol as given by mfold 142
Figure 3.38. Reference sequence (ENST00000370263) pre-mRNA folding (free energy of -δG of 83.4kcal/mol as given by RNA2 143
Figure 3.39. Affected allele pre-mRNA folding (free energy of-δG of 112.10kcal/mol as given by RNA2 143
Figure 3.40. Alternative donor and acceptor sites as given by SpliceView 145
Figure 3.41. Results of ASSP show the presence of a strong alternative isoform/cryptic donor at position 177 of the submitted affected allele sequence with a score of 8.778

Figure 3.42. Copy Number Variation analysis where affected family members are compared with non-affected family members

Chapter 4 - Discussion

Figure 4.1. Pre-mRNA folding of the affected allele and Reference sequence where a sequence in a loop region end up in a stem region due to the DNA variation in the affected allele

Figure 4.2. Sites involved in the splicing of pre-mRNA
<table>
<thead>
<tr>
<th>Software</th>
<th>URL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Transfer Tool (DTT)</td>
<td><a href="http://www.affymetrix.com">http://www.affymetrix.com</a></td>
</tr>
<tr>
<td>GeneChip® Operating Software (GCOS)</td>
<td><a href="http://www.affymetrix.com">http://www.affymetrix.com</a></td>
</tr>
<tr>
<td>GeneChip® Genotyping Analysis Software (GTYPE)</td>
<td><a href="http://www.affymetrix.com">http://www.affymetrix.com</a></td>
</tr>
<tr>
<td>Genotyping® Console</td>
<td><a href="http://www.affymetrix.com">http://www.affymetrix.com</a></td>
</tr>
<tr>
<td>Alohomora v0.29</td>
<td><a href="http://gmc.mdc-berlin.de/alohomora/">http://gmc.mdc-berlin.de/alohomora/</a></td>
</tr>
<tr>
<td>Graphical Relationship Representation (GRR)</td>
<td><a href="http://www.sph.umich.edu/csg/abecasis/index.html">http://www.sph.umich.edu/csg/abecasis/index.html</a></td>
</tr>
<tr>
<td>Pedcheck</td>
<td><a href="http://watson.hgen.pitt.edu/register">http://watson.hgen.pitt.edu/register</a></td>
</tr>
<tr>
<td>Allegro v 2.0</td>
<td><a href="http://www.decode.com/">http://www.decode.com/</a></td>
</tr>
<tr>
<td>Merlin</td>
<td><a href="http://www.sph.umich.edu/csg/abecasis/index.html">http://www.sph.umich.edu/csg/abecasis/index.html</a></td>
</tr>
<tr>
<td>Cyrillic 2.1</td>
<td><a href="http://www.cyrillicsoftware.com/">http://www.cyrillicsoftware.com/</a></td>
</tr>
<tr>
<td>Oligo Explorer 1.2</td>
<td><a href="http://www.genelink.com/tools/gl-oe.asp">http://www.genelink.com/tools/gl-oe.asp</a></td>
</tr>
<tr>
<td>Chromas with ClustalW</td>
<td><a href="http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm">http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm</a></td>
</tr>
<tr>
<td>NetGene2</td>
<td><a href="http://www.cbs.dtu.dk/services/NetGene2/">http://www.cbs.dtu.dk/services/NetGene2/</a></td>
</tr>
<tr>
<td>SpliceView</td>
<td><a href="http://bioinfo.itb.cnr.it/oriel/splice-view.html">http://bioinfo.itb.cnr.it/oriel/splice-view.html</a></td>
</tr>
<tr>
<td>ASSP</td>
<td><a href="http://www.es.embnet.org/~mwang/assp.html">http://www.es.embnet.org/~mwang/assp.html</a></td>
</tr>
<tr>
<td>FAS-ESS</td>
<td><a href="http://genes.mit.edu/fas-ess/">http://genes.mit.edu/fas-ess/</a></td>
</tr>
<tr>
<td>RESCUE-ESE</td>
<td><a href="http://genes.mit.edu/burgelab/rescue-ese/">http://genes.mit.edu/burgelab/rescue-ese/</a></td>
</tr>
<tr>
<td>Mfold</td>
<td><a href="http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi">http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi</a></td>
</tr>
<tr>
<td>RNA2</td>
<td><a href="http://www.genebee.msu.su/services/rna2_reduced.html">http://www.genebee.msu.su/services/rna2_reduced.html</a></td>
</tr>
</tbody>
</table>
HaploPainter v027  http://haplopainter.sourceforge.net/
HapMap Genome Browser (B36)  http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/
easyLINKAGE  http://genetik.charite.de/hoffmann/easyLINKAGE/index.html
Abstract

Epilepsy, which affects about 1% of the population worldwide, is a term that covers a range of heterogeneous disorders of brain function. The only common denominator of this condition is the susceptibility to repeated, unprovoked seizures of various phenotypes. While some epilepsies may be caused by a number of known neurological and metabolic conditions, others have unknown aetiology and are idiopathic. In recent years, significant research has been carried out on the genetics of epilepsy, and a number of mutations, mostly in genes that code for ion channels, have been linked to the epilepsy phenotype. Linkage analysis has played a major role in the discovery of these genes, using large families with several affected individuals.

In this study, linkage analysis was used to identify possible novel causative loci in a chromosomal interval that is linked to the epilepsy phenotype in a Maltese family, which has seven affected individuals with epilepsy or febrile seizures, over three generations. The inheritance pattern in the family was found to be autosomal dominant with incomplete penetrance. The syndrome in the family was characteristic of generalized epilepsy with febrile seizures plus (GEFS+) in being idiopathic, and with the affected family members having febrile seizures together with other types of epilepsy. The predominance of febrile seizures that remit before age six could also characterize the phenotype as familial febrile seizures. This study identified other loci beside the main linkage interval on chromosome 20, which may harbour genes that may explain phenotype heterogeneity and reduced penetrances in various family members.
Linkage analysis was performed using high-density single nucleotide polymorphisms (SNP) genechips. Analysis of the DNA from affected family members identified a linkage interval of about 20cM on chromosome 20 (20q13.32-33) which gave a parametric LOD score of 2.67. This score is suggestive evidence for linkage. By varying the penetrance and disease inheritance models, two other linkage intervals were observed on cytogenetic bands 3q21.3 and 13q13-14.

The linkage interval on chromosome 20 was confirmed using markers for Short Tandem Repeats (STRs). This region harbours several genes. The Zlr score peaked in a region of about 1.2cM, which has 18 genes, two of which have been already linked to other epilepsy syndromes. The best candidate genes in this region were CHRNA4 and KCNQ2 (which both code for ion channels) and STMN3, which is involved in neurite growth. DNA resequencing of the exons and intron/exon junctions of these genes, identified variations between the affected family members and the non-affected members. The affected individuals were found to have a different haplotype from the non-affected individuals. A novel 24bp deletion was, in fact, identified in all the affected individuals. This allele was found to normally occur only at a frequency of 0.04 in the general Maltese population.

Analysis of the affected allele, using bioinformatics software predicted an alternative splice site and a different folding pattern of the pre-mRNA, when compared to the reference sequence. cDNA sequencing of the CHRNA4 transcript of an affected individual, did not show any variation from the reference sequence. Copy Number Variation analysis identified regions where the two family members who have the
disease haplotype, but do not have epilepsy seizures, differ from the other affected individuals with epilepsy.

This project has identified novel genetic loci, which contribute to the phenotype of epilepsy. This could contribute to a better understanding of the cause of this condition. When considering that 30% of people who have epilepsy are pharmacoresistant, and other affected individuals still have a low quality of life in spite of treatment with antiepileptic drugs, it is imperative that the knowledge gained though genetic studies is used in the identification of novel therapeutic targets and the development of new drugs. This study also gives insights on the possible use of such genetic data for innovative drug development.
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3'UTR</td>
<td>Untranslated region at the 3' prime end of mRNA</td>
</tr>
<tr>
<td>ACh</td>
<td>Acetylcholine</td>
</tr>
<tr>
<td>ADNFLE</td>
<td>Autosomal dominant nocturnal frontal lobe epilepsy</td>
</tr>
<tr>
<td>AED</td>
<td>Antiepileptic drugs</td>
</tr>
<tr>
<td>C16orf45</td>
<td>Gene coding for the chromosome 16 open reading frame</td>
</tr>
<tr>
<td>CACNA2D2</td>
<td>Gene coding for the calcium channel, voltage-dependent, alpha-2/delta subunit 2</td>
</tr>
<tr>
<td>CAE</td>
<td>Childhood absence seizures</td>
</tr>
<tr>
<td>CGH</td>
<td>comparative genomic hybridization</td>
</tr>
<tr>
<td>CHRNA4</td>
<td>Gene coding for the neuronal acetylcholine receptor subunit alpha-4 precursor</td>
</tr>
<tr>
<td>CICN2</td>
<td>Gene coding for chloride channel 2</td>
</tr>
<tr>
<td>CLTC</td>
<td>Gene coding for the clathrin, heavy chain (Hc)</td>
</tr>
<tr>
<td>cM</td>
<td>centimorgan</td>
</tr>
<tr>
<td>CNTNAP2</td>
<td>Gene coding for contactin associated protein-like 2</td>
</tr>
<tr>
<td>CNV</td>
<td>Copy number variation</td>
</tr>
<tr>
<td>CT Scan</td>
<td>Computed tomography</td>
</tr>
<tr>
<td>CTGLF4</td>
<td>Gene coding for the centaurin, gamma-like family, member 4</td>
</tr>
<tr>
<td>DHX40</td>
<td>Gene coding for the DEAH (Asp-Glu-Ala-His) box polypeptide 40</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EDTA</td>
<td>Ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>EEG</td>
<td>Electroencephalogram</td>
</tr>
<tr>
<td>EMX2</td>
<td>empty spiracles homeobox 2</td>
</tr>
<tr>
<td>EPM2A</td>
<td>epilepsy, progressive myoclonus type 2A, Lafora disease (laforin)</td>
</tr>
<tr>
<td>EPM2B</td>
<td>epilepsy, progressive myoclonus type 2A, Lafora disease (laforin)</td>
</tr>
<tr>
<td>ESE</td>
<td>Exonic splicing enhancer</td>
</tr>
<tr>
<td>ESS</td>
<td>Exonic splicer inhibitor</td>
</tr>
<tr>
<td>Gene Symbol</td>
<td>Description</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>GABRD</td>
<td>Gene coding for the delta subunit of the gamma amino butyric acid (GABA) A receptor</td>
</tr>
<tr>
<td>GABRG2</td>
<td>Gene coding for the gamma 2 subunit of the gamma amino butyric acid (GABA) A receptor</td>
</tr>
<tr>
<td>GCOS</td>
<td>GeneChip® Operating Software</td>
</tr>
<tr>
<td>GEFS+</td>
<td>Generalized epilepsy with febrile seizures plus</td>
</tr>
<tr>
<td>GPR98</td>
<td>Gene coding for G coupled protein-receptor 98</td>
</tr>
<tr>
<td>GRR</td>
<td>Graphical relationship representation</td>
</tr>
<tr>
<td>HV1</td>
<td>Hypervariable region 1</td>
</tr>
<tr>
<td>IGE</td>
<td>Idiopathic generalized epilepsy</td>
</tr>
<tr>
<td>ILAE</td>
<td>International League Against Epilepsy</td>
</tr>
<tr>
<td>IMPA2</td>
<td>Gene coding for inositol(myo)-1(or 4)-monophosphatase 2</td>
</tr>
<tr>
<td>JAE</td>
<td>Juvenile absence seizures</td>
</tr>
<tr>
<td>JME</td>
<td>Juvenile myoclonic seizures</td>
</tr>
<tr>
<td>Kb</td>
<td>kilobases</td>
</tr>
<tr>
<td>KCNQ2</td>
<td>Gene coding for the voltage gated potassium channel, subfamily Q, member 2</td>
</tr>
<tr>
<td>KCNQ3</td>
<td>Gene coding for the voltage gated potassium channel, subfamily Q, member 3</td>
</tr>
<tr>
<td>KRIT1</td>
<td>Gene coding for ankyrin repeat containing</td>
</tr>
<tr>
<td>LD</td>
<td>Linkage disequilibrium</td>
</tr>
<tr>
<td>LOD</td>
<td>Log_{10} of odds</td>
</tr>
<tr>
<td>LYK5</td>
<td>Gene coding for protein kinase LYK5</td>
</tr>
<tr>
<td>MPBV17L</td>
<td>Gene coding for the mitochondrial membrane protein-like</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imagery</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger ribonucleic acid sequence</td>
</tr>
<tr>
<td>NMDA</td>
<td>N-methyl-D-aspartic acid</td>
</tr>
<tr>
<td>NPL</td>
<td>Non-parametric linkage</td>
</tr>
<tr>
<td>PAFAH1B1</td>
<td>Gene coding for platelet-activating factor acetylhydrolase, isoform 1b, alpha subunit</td>
</tr>
<tr>
<td>PAH</td>
<td>Phenylalanine hydroxylase</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>PARG</td>
<td>Gene coding for the poly (ADP-ribose) glycohydrolase</td>
</tr>
<tr>
<td>QC</td>
<td>Quality Control</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>RRN3</td>
<td>Gene coding for the RNA polymerase I transcription factor homolog</td>
</tr>
<tr>
<td>SCN1A</td>
<td>Gene coding for the sodium channel, neuronal type 1, alpha subunit</td>
</tr>
<tr>
<td>SCN1B</td>
<td>Gene coding for beta subunit of the sodium channels type 1</td>
</tr>
<tr>
<td>SCN2A</td>
<td>Gene coding for alpha subunit of the sodium channels type 2</td>
</tr>
<tr>
<td>SMEI</td>
<td>Myoclonic epilepsy in infancy</td>
</tr>
<tr>
<td>SNP</td>
<td>Single nucleotide polymorphism</td>
</tr>
<tr>
<td>STMN3</td>
<td>Gene coding for stathmin-like 3</td>
</tr>
<tr>
<td>STR</td>
<td>Short tandem repeat</td>
</tr>
<tr>
<td>TAE</td>
<td>Tris-acetate-EDTA buffer</td>
</tr>
<tr>
<td>TEMED</td>
<td>Tetramethylethylenediamine</td>
</tr>
<tr>
<td>TIMM23</td>
<td>Gene coding for the translocase of inner mitochondrial membrane 23 homolog</td>
</tr>
<tr>
<td>TLE</td>
<td>Temporal lobe epilepsy</td>
</tr>
<tr>
<td>TLE\</td>
<td>Temporal lobe epilepsy</td>
</tr>
<tr>
<td>TMEM49</td>
<td>Gene coding for the transmembrane protein 49</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>Zlr</td>
<td>Maximum likelihood ratio Z score</td>
</tr>
<tr>
<td>ZNF718</td>
<td>Gene coding for the Zinc finger protein 718</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction
1.1 Introduction

Epilepsy, which affects about 1% of the world population, is a term that covers a range of heterogeneous disorders of brain function (Shridharan, 2002). The only common denominator of these conditions is the susceptibility to repeated, unprovoked seizures of various phenotypes (Fisher et al., 2005). While some epilepsies may be caused by a number of known neurological and metabolic conditions, others have an unknown aetiology and are idiopathic. In recent years, a great deal of research has been carried out on the genetics of epilepsy, and, in fact, a number of gene mutations, such as those that code for various ion channels, have been linked to some epilepsy phenotypes. Such research is key to understanding the mechanisms underlying seizure disorders and the differences in drug response observed between patients (Scheffer et al., 2005). This research has great potential for the management of epilepsy in the future (Trent et al., 2003) since although a number of drugs have been developed to treat epilepsy, about 30% of epileptic patients remain refractory.

In this chapter an overview of the history of epilepsy is given; its incidence and epidemiology, classification, different aetiologies, pathophysiology, treatment presently available and a review of the research carried out in the last few years on the genetics of epilepsy and the implications known to date on the latter's effect on drug therapy. Different approaches to genetic studies that have been published will also be discussed with reference to linkage analysis, genotyping and data analysis and the software packages available for the statistical analysis of the data thus obtained.
1.2 Epilepsy – historical perspective

The oldest detailed account of epilepsy is found on a Babylonian tablet dating back to 1000BC held in the British Museum. The text on the tablet emphasizes the supernatural nature of epilepsy, and each seizure type is described as being associated with the name of a spirit or god and thus the only treatment given which was deemed suitable at the time was a spiritual one (SoRelle, 2005; Wilson and Reynolds, 1990). Sushruta, in the Indian system of medicine, Ayurveda (year 1000-800BC), noted for the first time the possible role of hereditary influences in epilepsy (Jain and Tandon, 2004; Tripathi, 2002).

The Greeks also considered epilepsy to be a disease caused by the invasion of the body by a god and in fact it was known as the “Sacred Disease.” In the “On the Sacred Disease” written by Hippocrates in 400BC he writes the following: “men being in want of the means of life, invent many and various things and devise many contrivances for all other things, and for this disease, in every phase of the disease, assigning the cause to a god........if they speak in a sharper and more intense tone, they resemble this state to a horse and say that Poseidon is the cause..... but if foam be emitted by the mouth, and the patient kick with his feet, Ares than gets the blame.”

Hippocrates, the Father of Medicine, believed that the disease was not sacred at all but has a natural cause like other diseases (Jain and Tandon, 2004). He also identified epilepsy as a brain disorder and even deducted that epilepsy has a hereditary component. “Its origin is hereditary, like that of other diseases. For if a phlegmatic person be born of a phlegmatic.......what is to hinder it from happening that where the father and mother were subject to this disease, certain of their
offspring should be so affected also? Hippocrates’ view of epilepsy as a brain disorder with a hereditary component did not begin to take root until the 19th century. The intervening 2000 years have in fact been indeed dominated by the earlier supernatural views.

Indeed, in Medieval Iran, some medical practitioners, such as Al-Tabari (838-870 AD), Abu Bakr Muhammed ibn Zakariya Al-Razi (860-940 AD) and Abu Ali Al-Hussain ibn Abdullah Ebn-e Sina (980-1037 AD), were of the opinion that, seizures result from the dynamic interplay of endogenous factors (e.g. fever); epileptogenic factors (e.g. hypoxia during delivery; head trauma; infectious diseases that lower seizure threshold) and precipitating factors (psychological stress; sleep deprivation; visual stimuli) (Gorji and Khaleghi, 2001). Al-Tabari, in his book ‘Paradise of Wisdom’, describes the occurrence of familial epilepsy (Gorji and Khaleghi, 2001).

1.3 Definition of epilepsy

The word epilepsy is derived from the Greek word “epilepsia” which means “to take hold of” or “to seize” (WHOa). Epilepsy is defined as ‘a disorder of the brain characterized by an enduring predisposition to generate epileptic seizures’ and by the neurological, cognitive, psychological and social consequences of this condition. The definition of epilepsy requires the occurrence of at least one epileptic seizure (Fisher et al., 2005). Indeed ‘epilepsy’ covers a range of heterogeneous disorders of the brain function (Gillies et al., 1991) and the only common denominator of these diseases is the susceptibility to repeated, unprovoked seizures (Mak, 2001). Epilepsy is one of the commonest neurological disorders regardless of age, racial or
geographical boundaries (Sander and Kwan, 2004). The causes of epilepsy are numerous and include acquired brain damage, congenital and genetic factors, however some causes are still unknown (Parton and Cockerell, 2003; Steinlein, 2004).

There are several medical conditions which may be mistaken for epilepsy, such as syncope and psychogenic attacks. These have to be ruled out before a diagnosis of epilepsy is made (Dhillon and Sander, 1999). A clinical examination is also required, and this includes various blood tests in order to eliminate the possibility of any metabolic abnormality as a cause of the seizure. Supportive investigations, such as electroencephalogram (EEG), and magnetic resonance imagery (MRI), should also be used (Guberman and Bruni, 1999). An abnormal EEG is present in about 60% of patients having an epileptic seizure, while an MRI is very helpful in determining any underlying brain abnormality that could be causing the seizures such as stroke or brain tumours (Shorvon, 2000).

1.4 Pathophysiology of epilepsy

The brain is a series of interconnected networks where neurons are connected in a complex network in which each individual neurone is linked through synapses with hundreds of others (Dhillon and Sander, 1999; Blumenfeld, 2003). Electrical discharges by neurones cause the release of neurotransmitters that can either be excitatory or inhibitory. An excited neurone will activate an interconnected neurone while an inhibited one will not. In this way, information is conveyed, transmitted and processed in the central nervous system (Dhillon and Sander, 1999).
A normal neurone discharges repetitively at a low baseline frequency while in the case of seizures the low frequency discharges are replaced by hyperexcitability (Steinlein, 2004). If a single neurone discharges in an abnormal way, this does not usually have any clinical significance. However, when a whole population of neurones discharge synchronously in an abnormal way a seizure may be triggered (Dhillon and Sander, 1999). Since the brain is an interconnected neuronal network, a localized hyperexcitability spreads to other networks and, if not counterbalanced by inhibitory mechanisms, more neurones will be recruited and this leads to a clinically visible seizure (Blumenfeld, 2003; Chang and Lowenstein, 2003; Steinlein, 2004).

The underlying causes of seizures are thus abnormally discharging neurons. Indeed epileptic discharges are defined as "a pathological extreme of normal neuronal excitability and synchrony that are modulated by ion channel function" (Doman and Pellagra, 2004). Neuronal injury, damage or a metabolic insult can all lead to abnormal discharging of neurones. At the molecular level, perturbations in neurotransmitters, receptors and various ion channels contribute to the seizure. In the case of epilepsies where no underlying cause has been identified (idiopathic epilepsy), mutations in various ion-channels have been implicated as the possible cause of neuronal discharge abnormality (Escagy et al., 2000; Hirose et al., 2002; Steinlein, 2004; Wallace et al., 2001).

Nerve cell excitability is also associated with the flow of ions through specialized protein aggregates within neuronal membranes, ion-channels, and a mutation can thus interfere with the function of the channel leading to a difference in electrical
signals which are critical for the normal functioning of neurones (Doman and Pellagra, 2004; Felix, 2000). Mutations in neuronal acetylcholine receptor subunit alpha-4 precursor (CHRNA4) have been found to cause autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) (Phillips et al., 1998). CHRNA4 is one of the subunits of the nicotinic acetylcholine receptor, a ligand gated ion channel. The acetylcholine binds to the α subunit causing a conformational change in this subunit which causes the opening of the ion channel for the entry of cations Na⁺, K⁺ and Ca²⁺ leading to depolarization. Ca²⁺ entry also causes the release of neurotransmitters. Functional studies of these mutations have shown that the mutation increases sensitivity to acetylcholine thus increasing the probability of pore opening leading to increases in excitation (Bertrand et al., 2002).

1.5 Aetiology of epilepsy

As already discussed, epilepsy is a heterogeneous term, incorporating numerous epilepsy phenotypes and syndromes with different aetiologies (Fisher et al., 2005) and the causes are numerous and include acquired structural brain damage, inherited genetic and congenital causes (inherited or acquired) (Steinlein, 2004). The range of aetiologies also varies in different age groups and geographical locations. In early childhood, congenital and genetic conditions are the most common causes of epilepsy (Camfield, 2002; Kwong et al., 2007), while in older children and young adults, genetic predisposition, alcohol and drug abuse and trauma are important causes. In the elderly, cerebrovascular disease and Alzheimer’s disease are the main causes of epilepsy (Sridharan, 2002). Brain tumours and infections, e.g. meningitis, occurs at all ages, but certain endemic infections (malaria) are common to particular
geographical regions such as South America (Mexico), Africa and southern Asian countries such as India. Epilepsy can also be multifactorial, that is, a combination of acquired and genetic factors contribute to the development of the epilepsy syndrome (Shorvon, 2000).

Epilepsy can also occur as a complication of cerebrovascular disease. It has been estimated that the computed tomography (CT) scan of over 50% of those who present with epilepsy after the age of 50, reveals cerebrovascular disease (Paradowski, 2005). Degenerative disorders can also cause epilepsy. About 30% of patients who have Alzheimer’s disease, develop epilepsy in the late stages of the disease, and tonic-clonic seizures are a presenting feature on 10% of cases of Creutzfeldt-Jakob disease (Sridharan, 2002; Vercueil, 2004). About 40% of adults presenting with newly diagnosed epilepsies have an underlying tumour of the cerebral cortex. Tumours in other regions of the brain do not usually have seizures as one of the symptoms (Stefan, et al., 2005).

Head injuries resulting from trauma, such as from falls account for about 2-12% of epilepsies. The severity of the head injury is the main factor that determines the risk of epilepsy (Pagni and Zenga, 2005). Neurosurgery can also precipitate seizures. It has been found that about 17% of patients who undergo neurosurgery for non-traumatic conditions develop seizures in the first 2 years after the operation (Chung, 2005).

Epilepsy can occur in many infective diseases where there is cerebral involvement. Both bacterial and viral meningoencephalitis can precipitate seizures. Malaria, tuberculosis, cysticercosis and others can also cause epilepsy (Lavados et al., 1992).
The HIV virus can also cause seizures and it can predispose the brain to other infections (eg. toxoplasmosis) and brain neoplasms which in themselves can cause epilepsy (Kellinghaus et al., 2008). There is also evidence that certain viral infections example latent herpes infections can cause epilepsy (Parton, 2003).

Drugs, alcohol, toxins and metabolic disturbances can also provoke seizures. Metabolic disturbances, such as those that occur in late-stage liver or renal disease, can result in seizures (Boggs, 2007).

1.6 Epidemiology of epilepsy

Epilepsy is one of the commonest neurological disorders and has no age, racial or geographical boundaries (Sander and Kwan, 2004). There are a number of shortcomings in current knowledge regarding the epidemiology of epilepsy, which are partly due to methodological problems that include selection bias, case ascertainment and difficulty in making accurate diagnosis (Sander, 2003; Shridharan, 2002). For accurate diagnosis and precise identification of seizure type or syndrome a full assessment of the patient has to be made as described in Section 1.3.

Epilepsy is such a heterogeneous disease, that it makes it very difficult to establish a precise epidemiological statistics for the disease (Dhillon and Sander, 1999; Shridharan, 2002). In recent years, progress has been made in determining the epidemiology of epilepsy but there are still some areas that require further research. There are cases where patients are wrongly misdiagnosed. The fact that there is also
stigma attached to epilepsy, may cause patients to be reluctant to admit their condition (Fernandez et al., 2004).

A study carried out by Callenbach et al. in 2001, reported an increase in mortality in children who have been diagnosed as having epilepsy, but also found a clear distinction between children having symptomatic (epilepsy caused by an underlying neurological disorder) and nonsymptomatic (idiopathic) epilepsy. They concluded that in the children with nonsymptomatic epilepsy the death rate was the same as that of the normal population while in the cases of symptomatic epilepsy there was a 20-fold increase in the death rate (Callenbach et al., 2001).

1.6.1 Incidence of epilepsy

According to data collected by WHO, it has been estimated that there are about 50 million people world wide with active epilepsy and that the lifetime prevalence of epilepsy, (i.e. the number of people presently in the world who either have epilepsy now, have had it in the past, or will experience it in the future), is approximately 100 million people (Sridharan, 2002).

The overall (age adjusted) incidence of epilepsy in developed countries is estimated to be approximately 25-50/100,000 person years, while the cumulative lifetime incidence of epilepsy is 3-4% (Hauser et al., 1997). The incidence of epilepsy shows a characteristic age-dependence with the highest incidences at the extremes of life. Incidence is very high in neonates (100-200/100,000) and infancy, but then it falls dramatically after the first year of life and more gradually throughout childhood.
Incidence reaches a minimum in early adulthood (25/100,000), but then increases during late adulthood to a second peak in the elderly (Hauser et al., 1993).

In some developing countries, the age specific incidence is quite different from that in developed countries, with the peak incidence of epilepsy occurring in early adulthood and there is no increase in the elderly (Lavados et al., 1992). The incidence is also higher than that in developing countries because of the higher incidence of certain infections such as neurocysticercosis cysts (tapeworm infection), malaria, meningitis and tuberculosis which often lead to epilepsy (Mac et al., 2007; WHOb).

1.6.2 Prognosis of epilepsy

Up to 5% of the population will have at least one seizure in their lifetime, but the prevalence of active epilepsy is much lower, and most patients who develop seizures have a good prognosis (Dhillon and Sander, 1999). Over 70% of people who have epilepsy enter remission, and about half of these patients successfully withdraw their medication without relapsing (Sander et al., 2003). A minority of patients, about 20-30%, develop chronic epilepsy and treatment is more difficult. Studies have shown that patients with symptomatic epilepsy having more than one seizure type, associated mental retardation, neurological or psychiatric disorders are more likely to have a poorer outcome (Berg et al., 2001; Dhillon and Sander, 1999).
1.6.3 Mortality associated with epilepsy

People with epilepsy have a greater risk of premature death that other people who do not have epilepsy (Sander et al., 2003). Most studies have given overall standardized mortality ratios between two and three times higher than in the general population (Dhillon and Sander, 1999). This increased in mortality occurs especially in the younger patients and those with severe epilepsy (Callenbach et al., 2001, Dhillon and Sander, 1999). Common causes of death in people with epilepsy include accidents (e.g. head injury), status epilepticus, pneumonia, cerebrovascular disease, suicide and SUDEP (sudden unexpected death in epilepsy) (Callenbach et al., 2001; Dhillon and Sander, 1999).

1.6.4 Refractory epilepsy

Refractory epilepsy can be also termed as pharmacoresistent epilepsy (Brodie, 2004). When it comes to refractory epilepsy, even with the development of novel antiepileptic drugs in the last decade, the percentage of the patients who have uncontrolled seizures despite drug therapy has remained practically the same that is 30% (Dhillon, 1999; French, 2007). Pharmacoresistance in epilepsy can be due to a number of factors including the type and severity of epilepsy, environmental factors, such as trauma and prior drug exposure, and genetic factors that predetermine the rate of absorption, metabolism and uptake of a drug by target (Loscher and Potschka, 2002; French, 2007). Pharmacoresistance can also develop after a period of seizure control (Schmidt and Loscer, 2005).
Two main hypotheses have been developed to explain refractory epilepsy. The target hypothesis, postulates that intrinsic or acquired changes in AED targets in the brain form the basis of the development of drug-resistant epilepsy, whereas the multidrug-transporter hypothesis claims that the drug target is never reached because intrinsic or acquired overexpression of multidrug transporters at the blood–brain barrier restricts brain uptake of AEDs (Schimidt and Loscer, 2005).

1.7 Classification of seizure types and syndromes

Epilepsy, according to International League Against Epilepsy (ILAE) are classified into three main groups: partial, generalized and the unclassified epilepsies (ILAE, 1989). This classification is very limiting in that it does not include cause of the epilepsy, nor does it assist in the prognosis of the specific epilepsy nor in the choice of treatment and in fact is presently being reviewed by ILAE (Iinuma et al., 2006).

1.7.1 Partial Seizures

Partial seizures are those seizures that arise in specific, often small loci, of the cortex in one hemisphere (Mosewich and So, 1996) In simple partial seizures the discharge remains localized and consciousness is preserved (Shorvon, 2000). Simple partial attacks rarely occur on their own and usually progress to complex partial seizures where consciousness is impaired or lost. Complex partial seizures usually have three components:

- Aura - equivalent to simple partial seizures usually lasting few seconds
• Altered consciousness - altered consciousness takes the form of an absence and motor arrest. The patient is unable to respond to or carry out simple commands or to execute willed movement, and lacks awareness of surroundings and events.

• Automatisms - involuntary movements that occur during an epileptic seizure. Automatisms are most common in temporal and frontal lobe seizures and usually manifest themselves as orofacial movements such as chewing, gestural such as fiddling movements with hands, (Dhillon et al., 1999). Complex partial seizures can sometimes progress to secondarily generalized seizures. The spread can occur so quickly that no feature of the localized onset is apparent to an observer or patient. The involvement of the whole brain leads to a convulsive attack like a generalized tonic-clonic convulsion (Shorvon, 2000).

1.7.2 Generalised seizures

Generalised seizures result in impairment consciousness from the onset owing to the extensive cortical and subcortical involvement (Guberman and Bruni, 1999). There are various types of generalised seizures.

Typical absence seizures compromise an abrupt loss of consciousness and cessation of all motor activity. There is no fall but the patient is unaware, inaccessible and often appears blank. The attack last only few seconds and can occur hundreds of times a day (Shorvan, 2000). Typical absence seizures develop in childhood or adolescence and are encountered almost exclusively in idiopathic generalised epilepsy (Fong et al., 1998). Atypical absence seizures differ from typical absences in clinical form, EEG, aetiology and clinical context (Shorvon, 2000). Their duration
is longer, loss of consciousness is not complete and onset and cessation of attacks are not so abrupt. Atypical absences are associated with learning disability, other neurological abnormalities and multiple seizure types (Shorvon, 2000).

A myoclonic seizure is another type of generalised seizure is a brief contraction of muscles caused by a cortical discharge. It may involve the whole body or the arms or the head. Recovery from the seizure is immediate (Guberman and Bruni, 1999). Myoclonos can develop at any age and is one of the three seizure types in idiopathic generalised epilepsy. They usually occur on waking or when dropping off to sleep (Shorvon 2000). Clonic seizures consist of cloning jerking. These seizures are frequent in neonates and young children and are almost always symptomatic (Camfield and Camfield, 2002). The tonic seizure results in extension of the neck, contraction of the facial muscles, with eyes opening widely; up-turning of the eyeballs, contraction of the muscles of respiration and spasm of the proximal upper limbs muscle (Shorvon, 2000). These seizures usually last less than 60 seconds. Tonic seizures occur at all ages and are usually associated with diffuse cerebral damage and learning disabilities.

A tonic-clonic seizure is initiated by loss of consciousness, where the patient falls if standing; there is a brief period of tonic flexion and then a longer phase of rigidity. Respiration ceases and cyanosis occurs (Guberman and Bruni, 1999). The tonic phase takes about 10-30 seconds after which it is followed by the clonic phase during which convulsive movements occur. This phase lasts about 30-60 seconds. The final stage is characterised by flaccidity and can last from 2-30 minutes. Tonic-
clonic seizures occur at any age and are encountered in many epilepsy syndromes including the idiopathic generalised epilepsies (Mattson, 2003).

Atonic seizures comprise a sudden loss of muscle tone causing the patient to fall. These seizures are short followed by immediate recovery. These seizures occur at all ages and are always associated with severe cerebral damage, learning disability and severe symptomatic epilepsy (Shorvan, 2000).

Unclassifiable seizures includes about one-third of seizures, which cannot be classified using the current ILAE classification scheme (Michael, 1999). They have forms that do not conform with the typical clinical and EEG patterns of the above mentioned seizure types (Shorvon, 2000).

1.7.3 Epilepsy syndromes

According to ILAE, an epileptic syndrome is a ‘condition that has clinically recognizable pattern; the clustering of signs and symptoms that form this pattern suggests shared mechanisms. Those who have a particular syndrome differ in some fundamental way(s) from those who have other forms of epilepsy’ (Engel, 2006).

When epilepsy syndromes are classified, a number of factors are taken into account, including seizure type, typical EEG pattern, clinical features, the expected course of the disorder, precipitating features, expected response to treatment, and genetic factors (Berg et al., 2005). A full assessment of the patient is also made such as a description of the seizure from the patient and from eyewitnesses, together with a
family history, a history of recent diseases and any medications used and of life-style should also be collated (Manford, 2001).

Since 1981, ILAE has been working to establish a standardized classification and terminology for epileptic seizures and syndromes (Seino, 2006). The classification has had several modifications through the years and Appendix 1 lists the current classification of epilepsy syndromes.

1.8 Inheritance factors associated with epilepsy

About 1% of the population develop recurrent unprovoked seizures for no obvious reason and without any neurological abnormalities. It has estimated that 40% of the epilepsies are idiopathic generalised epilepsy and these are assumed to be mainly of genetic origin (Sander et al., 2000). These epilepsies display a complex pattern of inheritance and this suggests that several genetic factors contribute to the seizure phenotype (Sander et al., 2003). There are also some cases of epilepsy where the causative gene is known (monogenic) and the epilepsy is transmitted in a mendelian manner, however these account for only 1-2% of all epilepsies (Tripathi, 2002). The genes responsible usually code for ion channels e.g. autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE) where the gene responsible is the nicotinic acetylcholine (Ach) receptor subunit, CHRNA4. (Kullman, 2002).

Seizures may also occur in other inherited conditions where the seizure phenotype presents with other neurological features. These include DNA expansion syndromes such as Fragile X syndrome, where there is learning disability in conjunction with
seizures (Incorpora, 2002). Genetic syndromes such as Wilson’s disease, Niemann-Pick disease and Down’s Syndrome are also associated with seizures besides other neurological manifestations (Tripathi et al., 2003). Inborn errors of metabolism can also present with seizures. These include diseases such as phenylketonuria and porphyria (Martynyuk, 2007; Solinas, 2004).

Table 1.1. Inherited diseases that have seizures as one of their symptoms.

<table>
<thead>
<tr>
<th>Inherited factors</th>
<th>Example of a disease</th>
<th>Mode of inheritance</th>
<th>Implicated genetic abnormality</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizencephaly (genetic malformation of cerebral cortex)</td>
<td>Not clear but familial occurrence.</td>
<td>EMX2 (empty spiracles homeobox 2)</td>
<td>Brunelli et al., 1996</td>
<td></td>
</tr>
<tr>
<td>Single-gene disorders that cause brain abnormalities and have epilepsy among their manifestations</td>
<td>Lafora Disease</td>
<td>Autosomal recessive</td>
<td>EPM2A or EPM2B (epilepsy, progressive myoclonus type 2A, Lafora disease (laforin) or 2B)</td>
<td>Serratosa et al., 1999a</td>
</tr>
<tr>
<td>Down’s Syndrome</td>
<td>Trisomy 21</td>
<td>Corrado et al., 2005</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Myoclonus epilepsy ragged-red fibre syndrome</td>
<td>Maternal inheritance</td>
<td>Mitochondrial DNA mutation</td>
<td>Schoffner et al., 1990</td>
<td></td>
</tr>
<tr>
<td>Phenyketonuria</td>
<td>Autosomal recessive</td>
<td>PAH (phenylalanine hydroxylase)</td>
<td>Paine, 1957</td>
<td></td>
</tr>
<tr>
<td>Tuberosis sclerosis</td>
<td>Autosomal dominant</td>
<td>Tuberin on chromosome 16 and another locus on chromosome 9</td>
<td>Humfrey et al., 2004</td>
<td></td>
</tr>
<tr>
<td>Monogenic idiopathic idiopathic neonatal convulsions</td>
<td>Autosomal dominant</td>
<td>Potassium channels</td>
<td>Singh et al., 1998</td>
<td></td>
</tr>
</tbody>
</table>
1.9 Genetics of the idiopathic epilepsies

Genetic factors are suspected to have greater influence in idiopathic rather than symptomatic epilepsies (Bianchi, 2003; Tripathi, 2002), although genes causing abnormal neuronal development and migration leading to symptomatic epilepsy have been discovered (Puffenberger, 2007; Strauss, 2006). Idiopathic epilepsies are defined as the type of epilepsy ‘with a hereditary predisposition, whose aetiology cannot be attributed to an observable structural lesion, but that seems to result from dysfunction of the inhibiting mechanisms necessary at all times in cerebral auto-regulation’ (Berube, 1991).

In recent years it has been elucidated that idiopathic epilepsies are caused by genetic abnormalities, although the precise contribution is still not always clear (Escagy et al., 2001; Crunelli et al., 2002; Harkin et al., 2002; Hirose et al., 2002; Steinlein, 2004). Evidence in favour of the existence of mutant ‘epilepsy genes’ have been derived from studies of families that have a specific form of epilepsy (Tripathi, 2002). Large family studies have indicated a higher incidence of epilepsy in the affected family member’s family when compared to the incidence in the general population (Bianchi et al., 2003). Twin studies have demonstrated that there was a higher concordance rate of epilepsy in monozygotic twins (94%) as compared to the dizygotic twin group (71%) (Berkovic et al. 1998).

The contribution of genetics to the clinical manifestation of epilepsy was initially shown when a number of idiopathic epilepsies were described by simple Mendelian inheritance (Berkovic, 2000). Several families, where family members have epilepsy
have been found to have mutations in single genes that encode molecules involved in signal transmission between neurons: ion channels, molecules involved in neurotransmitter release and neurotransmitter receptors (Mak, 2001; Kullmann, 2002). Other proteins with less obvious roles in signal transmission in the brain such as transcription factors, intracellular signalling molecules and other proteins involved in the regulation of gene expression, cell proliferation, migration or cell death, have also been linked to epilepsy (Noebels, 2003).

Thus research to date indicates that the genes underlying epilepsy which have been identified so far, seem to belong to different functional categories which are involved in different molecular pathways that control the physiology of the nervous system and brain development (Noebels, 2003).

With the help of these family studies further discoveries have been made. It was elucidated that single gene mutations can cause both generalised and partial epilepsies (Ottoman, 2001). In monogenic epilepsy, penetrance is sometimes incomplete, whereas carriers of a mutation do not always have seizures. The phenotypes associated with some genes are variable, even within a family where there is the same point mutation (variable expressivity). This remarkable variability is consistent with a strong effect of modifying factors even when a single epilepsy gene is transmitted with high penetrance (Baulac et al, 1999).

A change on the molecular level may manifest differently in individuals as a result of modifying genetic or environmental factors (Mak, 2001). Conversely mutations in different genes can cause the same epilepsy phenotype (Kullmann, 2002; Steinlein,
2004). This is known as locus heterogeneity where a single syndrome is caused by different genes in different families (Aylsworth, 1998; Winawer et al, 2003). This reinforces the emerging evidence that idiopathic epilepsies result from individually rare, gene defects (Noebels, 2003).

Multifactorial (complex inheritance) epilepsies are more common than those caused by mutations in single genes. These conditions are more complicated to study since the seizure phenotype is influenced by the effect of variation at several or multiple genes (Tan et al. 2004). Each gene contributes a small or modest effect to seizure susceptibility and by itself is insufficient to cause epilepsy. In addition, environmental factors may also play a part in the development of the seizure (Briellmann et al, 2001; Kaneko et al., 2002). Idiopathic generalised epilepsies (IGE) are the best example of such complex epilepsies. Linkage studies in these epilepsies have shown that although the IGE can be classified in four main IGE subsyndromes, they are all linked to a common locus on chromosome 18 but then each sub-syndrome links also to different chromosomal regions. For example in cases of juvenile myoclonic epilepsy (subsyndrome of IGE) linkage to chromosome 6 was also found together with linkage to chromosome 18 (Tripathi, 2002). When such linkage studies are repeated in different populations or a different group of patients or families, sometimes these results cannot be replicated which further complicates the issue (Kaneco, 2002; Kullmann, 2002; Sander et al., 2002; Steinlein, 2004).

Studies have also shown that the risk to certain types of epilepsy has been found to be more than twice as high in the offspring of women then men in epilepsy (Ottman et al., 1988) Possible mechanisms for this maternal effect in the inheritance of
epilepsy include the possible involvement of mitochondrial genes, imprinting or selective expansion of repeat sequences (Max, 2001). All these mechanisms are a form of non-mendelian inheritance (Van Heyningen and Yeyati, 2004).

1.9.1 Monogenetic idiopathic epilepsy

Although complex genetic factors are thought to contribute to the pathogenesis of idiopathic epilepsy, several linkage studies have shown that a number of epilepsy syndromes are caused by mutations in single genes (George, 2004). Most genes that have been discovered to date in hereditary idiopathic epilepsies, were found to have mutations in the code for subunits of ion channels (Lerche et al., 2005). The LGI1 gene, which is mutated in autosomal dominant partial epilepsy with auditory features (Kalachicov, 2002), was initially thought to be unrelated to ion channels, but later it was shown to be a subunit of the K\(^+\) channel Kv1.1 associated protein complex (Berkovic et al, 2006; Schulte et al., 2006).
Table 1.2. Idiopathic epilepsy caused by mutations in voltage-gated or ligand-gated ion channels.

<table>
<thead>
<tr>
<th>Syndrome</th>
<th>Cytoband</th>
<th>Gene</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Generalized epilepsy with febrile seizures</td>
<td>19q13.1</td>
<td>SCN1B (gene coding for beta subunits of the sodium channels type 1)</td>
<td>Wallace, 1998</td>
</tr>
<tr>
<td>plus -Type 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GEFS+ Type 2</td>
<td>2q21-q33</td>
<td>SCN1A (gene coding for alpha subunit of the sodium channels type 1)</td>
<td>Bauluc et al., 1999; Escagy et al., 2000.</td>
</tr>
<tr>
<td>GEFS+ Type 3</td>
<td>5q31.1-q33.1</td>
<td>GABRG2 (gene coding for the gamma 2 subunit of the gamma amino butyric acid (GABA) A receptor)</td>
<td>Bauluc et al., 2001.</td>
</tr>
<tr>
<td>GEFS+ Type 4</td>
<td>2q22</td>
<td>SCN2A (gene coding for alpha subunit of the sodium channels type 2)</td>
<td>Audenaert et al., 2005</td>
</tr>
<tr>
<td>GEFS+ Type 5</td>
<td>1p36.3</td>
<td>GABRD (gene coding for the delta subunit of the gamma amino butyric acid (GABA) A receptor)</td>
<td>Dibbins et al., 2004</td>
</tr>
<tr>
<td>Benign familial neonatal-infantile seizures</td>
<td>2q23-q24.3</td>
<td>SCN2A (gene coding for alpha subunit of the sodium channel type 2)</td>
<td>Heron et al., 2002</td>
</tr>
<tr>
<td>Benign neonatal familial convulsions (EBN1)</td>
<td>20q13.3</td>
<td>KCNQ2 (voltage gated potassium channel, subfamily Q, member 2)</td>
<td>Singh et al., 1998</td>
</tr>
<tr>
<td>EBN2</td>
<td>8q24</td>
<td>KCNQ3 (voltage gated potassium channel, subfamily Q, member 3)</td>
<td>Charlier et al., 1998; Lewis et al., 1993</td>
</tr>
<tr>
<td>Autosomal dominant nocturnal frontal lobe epilepsy (ADNFLE)</td>
<td>20q13.2-q13.3</td>
<td>CHRNA4 (alpha subunit of acetylcholine receptor)</td>
<td>Phillips et al., 1998</td>
</tr>
</tbody>
</table>
1.9.2 Complex inheritance in common idiopathic epilepsies

Idiopathic generalised epilepsy account for over 40% of epilepsies. The most common types are childhood absence seizures (CAE), juvenile absence epilepsy (JAE) and juvenile myoclonic epilepsy (JME) (Sander, 2000). Inheritance in these syndromes is said to be complex (Serratosa, 1999b). Complex inheritance refers to non-Mendelian inheritance where more than a single gene mutation is required for the manifestation of the disease phenotype (van Heyningen and Yeyati, 2004). A study in a multigenerational family has shown that inheritance in CAE involves at least two genes (Marini, 2003).

In a study of 139 families with JME, maternal inheritance (parent of origin effect) was observed and there was also evidence found that independent inheritance of absence and myoclonic seizures in JME families indicates that combinations of loci confer susceptibility to the component seizure types of IGE (Pal, 2006). In another study of 130 IGE-multiplex families, the results provided suggestive evidence that at least three genetic loci, 3q26, 14q23 and 2q36, predispose to IGE syndromes, but due to the possibility of false-positive linkage findings, the loci were regarded as requiring further confirmation (Sander et al., 2000). In a study by Haug et al., a heterozygous mutation in CICN2 (found in 3q26 locus) was found in three IGE families (Haug et al., 2003). Klein et al. also confirmed linkage of IGE to 2q36-37 (Klein et al., 2008).
1.9.3 Symptomatic epilepsy genetics

ILAE defines symptomatic epilepsy as a syndrome in which the epileptic seizures are the result of one or more identifiable structural lesions of the brain. (Fisher et al., 2005). A number of studies have shown that genetics also play a central role in symptomatic epilepsy where gene mutations cause brain lesions which predispose to seizures (Puffenberger et al, 2007). Refer to Table 1.3.

Table 1.3. Genetic disorders that cause brain abnormalities and have epilepsy among their manifestations.

<table>
<thead>
<tr>
<th>Example of a disease</th>
<th>Mode of inheritance</th>
<th>Implicated genetic abnormality</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schizencephaly (genetic malformation of cerebral cortex)</td>
<td>Not clear but familial occurrence.</td>
<td>EMX2</td>
<td>Brunelli et al., 1996</td>
</tr>
<tr>
<td>Lafora Disease</td>
<td>Autosomal recessive</td>
<td>EPM2A or EPM2B</td>
<td>Gómez-Abad et al, 2005; Serratosa et al, 1995</td>
</tr>
<tr>
<td>Cortical dysplasia-focal epilepsy</td>
<td>Autosomal recessive</td>
<td>CNTNAP2 (contactin associated protein-like 2)</td>
<td>Strauss et al, 2006</td>
</tr>
<tr>
<td>Infantile-onset symptomatic epilepsy syndrome</td>
<td>Autosomal recessive</td>
<td>GM3 syntase</td>
<td>Simpson et al, 2004</td>
</tr>
<tr>
<td>Cerebral cavernous malformations</td>
<td>Sporadic, autosomal dominant or variable penetrance</td>
<td>KRIT1 (ankyrin repeat containing)</td>
<td>Denier et al, 2004; Gianfrancesco et al, 2007</td>
</tr>
<tr>
<td>Polyhydramnios, megalencephaly and symptomatic epilepsy</td>
<td>Autosomal recessive</td>
<td>7-kilobase deletion in LYK5 (protein kinase)</td>
<td>Puffenberger et al, 2007</td>
</tr>
<tr>
<td>Lissencephaly with posteriorly predominant gyral abnormality</td>
<td>Autosomal dominant</td>
<td>PAFAH1B1 (platelet-activating factor acetylhydrolase, isoform 1b, alpha subunit)</td>
<td>Guerrini et al, 2002</td>
</tr>
</tbody>
</table>
1.9.4 Febrile seizures and (generalized) epilepsy with febrile seizures plus (GEFS+)

In this present research, the genetics in a family diagnosed with a specific epilepsy syndrome, (generalized) epilepsy with febrile seizures plus (GEFS+) is studied. ILAE classifies febrile seizures (FS) as a condition with epileptic seizures which does not require a diagnosis of epilepsy. While febrile seizures are thought of as separate clinical entities, a genetic susceptibility to febrile seizures may cause epilepsy in other family members or later on in the same family member (Iwasaki, 2002). Simple febrile seizures consist of a brief (lasting less than 10 minutes) tonic-clonic convulsion which occurs only once within a 24-hour period. There are no focal features and it resolves spontaneously (Jones and Jacobson, 2007). Age of cessation in usually before age 6 years. Febrile seizures have been linked to a number of genetic loci, however the causative genes have not been identified in most patients (Nakayama and Arinami, 2006) (Table 1.4).
Table 1.4. Loci that have been linked to febrile seizures.

<table>
<thead>
<tr>
<th>Febrile Seizure</th>
<th>Locus and gene if identified</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Febrile seizures, familial 1 (FEB1)</td>
<td>8q13-q21</td>
<td>Wallace et al., 1996</td>
</tr>
<tr>
<td>Febrile seizures, familial 3 (FEB3)</td>
<td>2q24 (SCN1A-sodium channel, neuronal type 1, alpha subunit)</td>
<td>Peiffer et al., 1999</td>
</tr>
<tr>
<td>Febrile seizures, familial 4 (FEB4)</td>
<td>5q14-q15 (GPR98-G protein-coupled receptor 98)</td>
<td>Nakayama et al., 2000</td>
</tr>
<tr>
<td>Febrile seizures, familial 5 (FEB5)</td>
<td>6q22-q24</td>
<td>Nabbout et al. 2002</td>
</tr>
<tr>
<td>Febrile seizures, familial 6 (FEB6)</td>
<td>18p11.2 (IMPA2-inositol(myo)-1(or 4)-monophosphatase 2)</td>
<td>Nakayama et al., 2004</td>
</tr>
<tr>
<td>Febrile seizures, familial 7 (FEB7)</td>
<td>21q22</td>
<td>Hedera et al., 2006</td>
</tr>
<tr>
<td>Febrile seizures, familial 8 (FEB8)</td>
<td>5q31.1-q33.1 (GABRG2-gamma-aminobutyric acid (GABA)A receptor, γ2)</td>
<td>Audenaert et al, 2006</td>
</tr>
<tr>
<td>Febrile seizures, familial 9 (FEB9)</td>
<td>3p24.2-p23</td>
<td>Nabbout et al, 2007</td>
</tr>
</tbody>
</table>

*(Generalized) epilepsy with febrile seizures plus* (GEFS+) is a clinical subset of familial febrile convulsions in which affected individuals later develop afebrile seizures (Scheffer and Berkovic, 1997). GEFS+ is genetically heterogeneous and has been linked to a number of loci where mutations where found in genes coding for ion channels (Brice, 2004). To date, GEFS+ has been linked to 19q13.1 (SCN1B), 2q21-q33 (SCN1A), 5q31.1-q33.1(GABRG2), 2q24 (SCN2A) and 1p36.3 (GABRD) and recently also to 8p23-p21. (Wallace, 1998; Bauluc, 1999; Escagy, 2000; Lopes-Cendes et al, 2000; Bauluc, 2001; Audenaert, 2005; Dibbins, 2004; Baulac et al., 2008).
Still there are a number of families having GEFS+ phenotype that is not linked to these identified loci (Brice, 2004; Scheffer, 2005; Selmer et al., 2008). GEFS+ is characterized by phenotypic heterogeneity where family members having the same mutation exhibit a different phenotype including febrile seizures, and afebrile seizures; absence seizures, partial seizures (temporal or frontal), myoclonic and tonic seizures (Lopes-Cendes et al., 2000; Graves, 2006; Scheffer et al., 2007). ILAE have found difficulty in defining GEFS+ as a specific syndrome as there is no phenotypic homogeneity at an individual level.

1.10 Genetics of epilepsy and drug therapy

Most patients with epilepsy require long-term pharmacological therapy. The main therapeutic aim of the drug treatment available to date is to control the seizures, not to cure them, using just one drug with the lowest dose and with the least side-effects (Dhillon, 1999). In recent years, the armamentarium of drugs to control the symptoms of epilepsy has increased and over 70% of patients have their epilepsy controlled with one or a combination of drugs. Still the ‘wonder’ drug has not been found to date. Antiepileptic drugs have a number of severe side-effects and there are also about 30% of patients who have refractory (intractable) epilepsy, despite the use of different antiepileptic drugs in various doses and combinations (Berg, 2001).

Drug therapy for epilepsy is still in its infancy, since the mechanisms of what triggers and what stops the seizures are still not well understood and thus pharmacological treatment is not targeting the cause of the epilepsy (Doman, 2004).
Drug therapy is, in fact, directed to the symptoms of epilepsy, that is seizures, and not to the actual underlying cause.

The diagnosis of a particular type of epilepsy does not guide the physician to any specific drug, except in the case of absence seizures where ethosuximide or valproate is the drug of choice (Shorvon, 2000). In addition, except in the classification of absence seizures, the current classification of epilepsy does not assist in choosing treatment and there are still no official guidelines on drug treatment (Glauser, 2006) although some consensus has been agreed on the treatment of childhood epilepsies (Wheless, 2007). There are however several national and institutional general guidelines on treatment options (Feely, 1999).

The only way forward in order to seek a definite ‘cure’ for epilepsy is to understand it at a molecular level. The better a disease is understood the easier it would be to develop drugs that would target a specific biological mechanism, thus tailoring drug therapy and reducing side-effects (Tripathi, 2000; Johnson et al., 2001). As already described, in the last 25 years, a number of genetic studies have identified a number of genes as predisposing the brain to hyperexcitability (George, 2004) (Section 1.9). The more genes that are identified, the higher the understanding of this complex disease (Ottoman, 2001).

Various studies have shown that genetic variation can have an impact on drug response in epilepsy (Sisodiya, 2007). Genetic variations that influence the pharmacokinetics and pharmacodynamics of drugs can result in variation in drug response. Variation in drug metabolizing enzymes and multi-drug transporters have
been associated with variation in drug response in epilepsy (Spurr, 2006).

Antiepileptic drugs are known to act by different mechanisms that ultimately result in the modulation of the excitability of neurons (Ramachandran and Shorvan, 2003). Table 1.5 summarises and simplifies the mode of action of some of the currently used antiepileptic drugs.

Table 1.5. Modes of action of some antiepileptic drugs (Shovron, 2000).

<table>
<thead>
<tr>
<th>Drug</th>
<th>Modes of Action</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbamazepine</td>
<td>Acts of neuronal sodium channel conductance. Also action on monamine, acetylcholine and NMDA receptors</td>
</tr>
<tr>
<td>Clobazam</td>
<td>GABA_A receptor agonist. Also action on ion-channel conductance</td>
</tr>
<tr>
<td>Lamotrigine</td>
<td>Blockage of voltage-dependent sodium conductance</td>
</tr>
<tr>
<td>Phenytoin</td>
<td>Blockade of sodium channels and action on calcium and chloride conductance and voltage-dependent neurotransmission</td>
</tr>
<tr>
<td>Tiagabine</td>
<td>Inhibits GABA reuptake</td>
</tr>
<tr>
<td>Vigabatrin</td>
<td>Inhibition of GABA transaminase activity</td>
</tr>
</tbody>
</table>

Genetic variations in drug targets can also influence drug response. Mutations that alter the tertiary protein structure of an ion channel can result in altered drug responsiveness. This could be the reason why patients with severe myoclonic epilepsy in infancy (SMEI), that usually have missense mutations in the sodium channel protein that shorten this protein, show resistance to drugs acting on the sodium channel (Ramachandran and Shorvan, 2003). A study by Heinzen et al. also found that an intronic polymorphism in SCNA1 also influences drug response (Heinzen et al., 2007).
Chapter 1 - Introduction

The application of the gathered genetic information for drug development will thus greatly help in the future management of epilepsy (Scheffer, 2005).

1.11 Methodological approaches for identification of genes causing epilepsy

Single gene idiopathic epilepsies offer the best hope for the identification of epilepsy genes (Berkovic, 1997; Delgada-Escueta, 2004; George, 2004). Linkage analysis has been the traditional tool used in single gene diseases in order to narrow the chromosomal interval where the causative gene could be found. Linkage analysis requires large families with many affected individuals, or else a number of families that have the same disease and where the same gene is responsible for the phenotype (Tan, 2004).

The other approach used in order to identify epilepsy genes is through the study of candidate genes using animal models. A selected candidate gene, for example, an ion channel, is silenced or altered in a rodent, using knockout animal models. The observed neurological phenotype of the rodent is then evaluated. A number of novel genes linked to epileptogenesis have been detected by this method (Noebels, 2003). Subsequently a related human gene, calcium-channel beta4-subunit gene (CACNB4) has been linked to epilepsy, indicating that this method can be used as an approach in order to identify human epilepsy causing genes (Moore et al., 1998; Escagy et al., 2000; Noebels, 2003). Another candidate gene approach involves the identification of genetic variants in a gene or genes involved in the disease pathway that are present only in affected individuals. This approach does not require large families (Kwon et al., 2000). In these studies, the candidate genes (generally only the exons)
are sequenced in a number of affected individuals. Any identified variations from the reference sequence are studied, their frequency in population is determined and possible function of variant is experimentally elucidated (Greenberg, 2007). This type of study can however miss susceptibility mutations present in non-coding regions.

In the case of epilepsies linked with complex inheritance, linkage analysis can have less power when compared with linkage analysis carried out for monogenic disorders and in fact success with these approaches has been limited (Cardon and Bell, 2001; Mulley, 2005). The pathogenic effect of each susceptible gene may be small and non-mendelian inheritance may also be present which make linkage detection more difficult. In the case of complex epilepsies, association studies have been tried but for now they have given poor results (Tan, 2004). Association studies involve the comparison of the frequency of specific alleles in affected cases against that in unaffected controls. An allele is associated with the disease if its frequency differs between cases and controls more than would be predicted by chance.

1.12 Linkage analysis in genetic studies

One of the most commonly used methods in gene searches is genetic linkage analysis. In fact genome-wide linkage analysis is an established method of locating disease genes (Bevan, 2001). Genetic linkage refers to the tendency of alleles from two loci to segregate together in a family if they are located physically close to each other on a chromosome (de la Chapelle, 1998). The extent of linkage is a function of the distance between the two loci, which can be measured by the number of
crossovers between the two loci among the observed meiosis (recombination fraction - $\theta$). There is complete linkage of the two loci if there is no recombination ($\theta = 0$) and no linkage if the recombination fraction is 50% ($\theta = 0.5$) (Xu J, 1998). The LOD (log$_{10}$ of odds) score analysis is a likelihood-based parametric linkage approach to estimate the recombination fraction and the significant of the evidence of linkage (Kruglyak, 1996; Vogel, 1997; Xu J, 1998). The LOD score is the log$_{10}$ of the likelihood ratio of two hypothesis: $H_1$-The marker locus and the genetic trait (disease locus) are linked within the family ($\theta \leq 0.5$) and $H_0$ - the marker locus and the genetic trait are not linked ($\theta = 0.5$) (Kruglyak, 1996).

A LOD score of 3.0 or more is considered strong evidence for linkage. An LOD of 3 means that it is 1000 times more likely that the marker is linked to the disease genotype (over a range of recombinant fractions) than if it is not linked (the recombination fraction is 0.5) (Vogel, 1997). The ideal family for linkage is a pedigree with at least three generations, many matings and a large number of offspring since the number of informative meiosis in such a pedigree will be high (Vogel, 1997; Serratosa, 1999).

Parametric linkage analysis (LOD) can be highly sensitive to misspecification of the linkage model and in cases where the inheritance model is not at all clear the best solution is to use nonparametric linkage analysis (NPL). The linkage model defines mode of inheritance, penetrance, phenocopies and disease allele frequency. NPL provides a nonparametric pedigree-analysis method but loses some power when compared with the parametric method (Abreu et al., 1999). The NPL method does not require consideration of multiple models of inheritance nor advance knowledge.
of the correct model of inheritance (Kruglyak, 1996). NPL is also a likelihood ratio and measures whether affected individuals share alleles IBD (identical by descent) more often than expected under random segregation.

1.13 Family recruiting for genetic studies

The collection of data from families containing multiple affected individuals is essential for genetic research on the epilepsies (Ottman et al., 2005). It is possible to perform linkage analysis, both on a large family containing a number of affected individuals or else on a number of small families thought to have the same syndrome. In epilepsy, this is very difficult since even well defined syndromes have been linked to different genes (locus heterogeneity) such as benign neonatal familial convulsions which has been linked to both potassium voltage-gated channel, KQT-like subfamily, member 2 (KCNQ2) (20q13.3) and potassium voltage-gated channel, KQT-like subfamily, member 3 (KCNQ3) (8q24) (Singh et al., 2003). In such a case, if families of both types are studied together, a positive LOD score can be cancelled by a negative LOD score of the other type giving a non-statistically significant LOD score.

Recruiting families for genetic studies has its difficulty especially in diseases such as epilepsy that still carry a stigma. Families may not always be eligible or although the proband may be willing to participate, family members may refuse to co-operate with the researcher. Clinical information is at times very limited and the researcher has to rely on family anecdotes (Ottman et al., 2005).
1.14 Genotyping in genetic studies

A critical component of a gene mapping study is obtaining genotypes for the genetic markers used for linkage analysis (Vance, 1998). There are various methods and techniques used in genotyping, but presently the most commonly used genotyping markers are the Short Tandem Repeats (STRs) - Microsatellites. STRs are a class of polymorphisms that occurs when a pattern of two or more nucleotides are repeated and the repeated sequences are directly adjacent to each other. There is also another emerging technique which surpasses STRs genotyping. This new method is SNP genotyping using chip technology. Linkage analysis using high density SNP GeneChips® (SNP microarrays) can have higher power than STR studies and can narrow the linkage interval (Gonzalez Neira, 2007; Middleton, 2006). In this present study SNP genotyping was performed and STR genotyping was performed to confirm the result.

1.14.1 SNP genotyping in genetic studies

SNP (single nucleotide polymorphism) occurs when there is a single nucleotide variation in the DNA sequence between paired chromosomes of an organism (heterozygous) or between members of a species giving rise to the presence of different alleles (Jaing et al., 2003). For such a variation to be considered as an SNP it must be present in 1% of the population. It has been estimated that there are about 10 million SNPs in the human genome. SNPs can be found both in the coding and non-coding region of the genome. Recent studies have shown that even innocuous looking SNPs in intronic regions could cause disease (Krawczak et al., 2007). Silent mutations in the 5'-splice site of exon 10 of the tau gene lead to an increased
inclusion of exon 10 which affects the 4R to 3R tau splice isoforms ratio which is the root of the pathogenesis of frontotemporal dementia. The intronic variant c.306+2dupT in the MutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli) gene (MLH1) results in the loss of exon 3 and a frameshift mutation due to a new splice donor site 5 bp upstream leading to hereditary nonpolyposis colorectal cancer (Pagenstecher et al., 2006).

Various methods have been developed to interrogate SNPs and the method chosen depends on the number of SNPs that have to be genotyped (Dove, 2005). For a genome-wide scan, high density oligonucleotide SNP arrays are used. Chip technology allows a large number of SNPs to be interrogated simultaneously and is a hybridization based method. For the genome-wide scan in this study the GeneChip® Human Mapping Nsp 250K array (Affymetrix®) was used. In this technology, the DNA is fragmented using the restriction enzyme NspI. The fragmented DNA is then ligated to an adapter. One primer (complementary to adapter) amplification is performed followed by fragmentation of the PCR products to remove the adaptors. End-labelling (fluorescent dye) of the product is carried out followed by hybridization to the GeneChip® (Figure. 1.1).
The GeneChip® is produced by photolithographic in situ synthesis. A 5" x 5" quartz wafer is coated with a light-sensitive chemical (Kaminski, 2002). Lithographic masks are used to either block or transmit light onto specific locations of the wafer surface. The surface is then flooded with a solution containing either adenine, thymine, cytosine, or guanine, and coupling occurs only in those regions on the glass that have been deprotected through illumination. The dNTPs have also photoprotecting groups (Beier and Hoheisel, 2000). Unless the photoprotectant is removed, the nucleotide cannot bind to another nucleotide. By using the lithographic masks it is possible to direct light only on specific regions, removing the photoprotection and allowing oligomer extension (Beier and Hoheisel, 2000).
Chapter 1 - Introduction

Figure 1.2. Principles involved in the fabrication of the GeneChip®. A. Quartz having first dNTPs fixed to it. The red line indicates a layer of photoprotectant. B. A lithographic mask is placed on the quartz chip. C. The light removes the photoprotectant from the dNTPs that are not covered by the mask. D. Oligomer extension is only possible where the dNTPs have been deprotected.

1.14.2 Hybridization of DNA strands to the GeneChip®

Hybridization is the term used to describe the annealing of single stranded nucleic acid strands by hydrogen bonding (Rybicky, 1998). In the case of the GeneChip® or microarray, DNA strands (of the sample) anneal to their complementary oligomer on the chip, that is, the DNA on the chip acts as a probe in a hybridisation reaction. Each SNP is interrogated by a minimum of 24 oligonucleotides, such that a reliable SNP calling is obtained (e.g. in Affymetrix®). Following hybridization and washing, the gene chip is irradiated and a microarray reader reads the fluorescence emitted. The washing removes the unbound DNA strands (of sample) and during irradiation only bound (hybridized) DNA strands fluoresce.
1.14.3 Data analysis of SNP genotyping

Data analysis software selected for analyses of SNP microarrays depends on the type of information one needs to get from the data. In this case software for linkage analysis was used besides other software used for the quality control of data obtained. To date the number of studies where SNP gene chips are used for linkage analysis is increasing. Novel loci which had never before been linked to a condition when microsatellites were used are being detected when the genome-wide linkage analysis using SNP genotyping was performed (Afshari et al., 2008; Shugart et al., 2008; Stamm et al., 2006).

1.15 STRs genotyping in linkage studies

Short Tandem Repeats (STRs) are base pair repeats present throughout the genome. These repeats are usually present in non-coding regions of DNA. These repeats are highly polymorphic and therefore they are powerful for linkage analysis (Agrawal et al., 2004). When compared with SNPs, an STR is more powerful as heterozygosity for the marker is >0.6 while for SNPs the maximum is 0.5 although usually it is rarely more than 0.3. Average heterozygosity of the GeneChip® Human Mapping Nsp 250K array (Affymetrix®) is 0.29. Thus for a linkage study, more SNPs need to be genotyped in order to obtain the same power as with STRs (Sellick et al., 2004).

STRs can be dinucleotide, trinucleotide and tetranucleotide repeats. A dinucleotide repeat is a repeat of two bases pairs, a trinucleotide repeat is a repeat of three bases while a tetranucleotide repeat is a repeat of four bases. Dinucleotide repeats are the
most commonly found repeats in the genome, but the use of dinucleotide repeats should be kept to a minimum in genotyping since there use presents with several technical problems: the fact that alleles may be just two nucleotide apart can present difficulty when trying to distinguish between the two allele. In addition, the fact that many dinucleotides have “stutter bands” could make scoring the genotypes difficult especially when differentiating homozygous from heterozygous individuals (Vance, 1998). Trinucleotide and tetranucleotide are present in fewer numbers in the genome than the dinucleotide repeats but they have the advantage that they do not present with the same technical difficulties that the dinucleotide present (Smith et al., 1995). Although some trinucleotide repeats are unstable and have been associated with diseases such as Fragile X syndrome where an expansion of a CAG repeat in the FRM1 gene leads to the phenotype, most trinucleotide repeats are stable and can be used for genotyping (Hangerman, 2004).

A genome-wide linkage search is usually carried out with about 400 markers (STRs) which give a density of about 10cM separation across the genome (Bevan, 2001). STRs are amplified using Polymerase Chain Reaction (PCR) and fluorescently labelled primers. The PCR products are detected using Capillary electrophoresis, which offers high resolution for allele calling.

Reading of genotypes following electropheritic separation using capillary electrophoresis is known as “scoring”. To read genotypes, a software program is used called a GeneMapper Analysis Software (Applied Biosystems®). An internal lane standard (sizes of products of this standard are known) is run together with the sample. The software uses the sizing data of the internal lane standard in order to
size the PCR products. The allele size (base pairs) is used to call the allele (different size = different allele).

Genotyping data is then converted into a format that can be used by a linkage analysis software. For linkage analysis there are a number of software applications that can be used such as, GeneHunter v2.1v5 (Krugyak et al., 1996), or Allegro (Gudbjartsson et al., 2005). These can be used to calculate both LOD scores and NPL scores (Kruglyak, 1996).

1.16 Identifying the causative gene in genetic disorders

The parametric LOD score and NPL scores are used in these types of genetic studies in order to identify genomic regions of interest which link to the disease phenotype (Haines, 1998). The region of interest can be over 20cM interval on a particular chromosome. This 20cM region is still a very large region of the genome and usually a large number of genes can be found in this region (De la Chapelle, 1998). In order to possibly identify the gene, the procedure described in Figure 1.3 could be followed.
Linkage analysis identified a region of about 20cM where the LOD/NPL suggest linkage. A gene that links to this region is causing the epilepsy syndrome.

- **Literature search to identify gene function and region of expression**
  - Widely expressed genes or genes with unknown function. Try to narrow linkage by adding markers or by genotyping more family members.
- **High expression in brain**
  - No information of knockout gene experiments or studies. Try to narrow linkage by adding markers or by genotyping more family members.
- **Knockout gene experiments indicate an epilepsy phenotype**
  - More than 3 candidate genes. Narrow the linkage interval by using in between markers.
- **Only 1 to 3 candidate genes in the 20cM interval**
  - No mutation found in candidate genes. Consider less probable candidate genes.
- **Mutation found in a gene in the affected. DNA sequencing in a normal and an affected individual (starting with the smallest gene if more than one candidate gene)**
  - Mutation found not only in affected individuals. Probably this is a polymorphism. Possibility of incomplete penetrance? Consider sequencing other genes or sequencing other regions of gene e.g. promoter region.
- **Test all family members to determine if the mutation is present in all affected.**
  - Mutation found not only in affected individuals. Probably this is a polymorphism. Possibility of incomplete penetrance? Consider sequencing other genes or sequencing other regions of gene e.g. promoter region.
- **Variant present in all affected and not in normals. Perform literature search on particular variant.**
  - Perform functional studies to determine how the mutation affects function of gene (if this not done before literature) and how this could be correlated to the disease phenotype.

Figure 1.3. Approach to follow when trying to identify the causative gene in a wide linkage interval. The yellow colour indicates the way forward given the best scenario. The blue colour indicates other possible alternatives from the best scenario and how to proceed given that situation.
The experimental procedure described in Figure 1.3 is a time-consuming experimental design and in fact in a number of linkage studies that have been carried out the causative gene was not identified (Haines, 1998).

1.17 Aims of the study

This research thus focused on the use of gene-mapping studies using a high density SNP array for the identification of genes causing the epilepsy phenotype in a familial case of febrile seizures and epilepsy.

For best results with linkage analysis, a multigenerational family with a number of affected individuals should be used (Ottman et al., 2005). In such studies carried out, todate usually only one main locus that predisposed to epilepsy was identified (Table 1.2), however in another recent study, a genetic locus which could be harbouring a modifier gene was also observed (Nabbout et al., 2007). Such studies could be of great benefit in understanding phenotypic heterogeneity in the family. A number of other published studies have in addition demonstrated that high density SNPs genotyping can have higher power than STRs and can detect linked loci that are missed by 10cM spaced STRs (Middleton et al., 2004).

Copy Number variation (CNV analysis) can be carried out using high density SNPs data. Various studies have identified that CNVs could predispose to disease (McCarrol et al., 2007). It has been postulated that CNVs could also be responsible for an epilepsy phenotype (Gurnett and Hedera, 2007). A recent study has in fact identified that infantile spasms is caused by a hemizygous deletion (a loss) in chromosome 17 and it has been shown that this deletion results in the deletion of the
gene MAGI2 (Marshall et al., 2008). CNV can thus be used in order to identify deletions/insertions.

In the last decade great progress has been made in the genetics of epilepsy but there is still the need for more studies in order to understand the causes behind the phenotypic heterogeneity and incomplete penetrance of epilepsy.

Given the new developments in genotyping methods and the promising results coming out of these studies the main aims of this present study were:

1. The identification and recruitment of a multigenerational Maltese family with a number of affected individuals having febrile seizures or epilepsy that would give enough power to detect linkage.

2. To date no published study has yet identified how a single mutation in one gene could predispose to different phenotypes of epilepsy in a single family. Although one published study has demonstrated that more than one locus could predispose to epilepsy in a single family (Nabbout et al., 2007), penetrance or the relationships between loci and different phenotypes have not as yet been elucidated. Thus, one of the aims of this research was to perform a linkage study using a high density SNP chip in order to possibly identify other regions, beside a main locus, that could be predisposing to the different epilepsy phenotypes in the family.
3. Copy Number Variations have also been found to predispose to disease. Identical twins have been found to have difference in their copy number variation and this could explain why a twin may develop a disease and the other not (and not just environmental factors) (Bruder et al., 2008). It was therefore intended to use SNP data to determine the possibility of identifying any CNV between individuals in the family, having the mutation and the disease, and others having the mutation but no disease.

4. The identification of new mutations that could predispose to an epilepsy phenotype and the determination of how this could predispose to the seizure phenotype.

5. The study also reviewed ways by which genetic studies and pharmacogenomics could help in drug development and drug therapy in epilepsy. This type of research is extremely useful since, with the increase in knowledge of epilepsy as caused by gene mutations and variations, it would be possible to develop better drugs, which target these genes, and thus assist in improving the quality of life of persons with epilepsy.
Chapter 2

Methodology
2.1 Introduction to methodology

In this chapter details are given on the methodology used in this research. This research was carried out on a three generation family, having seven members affected by epilepsy or febrile seizures, which was recruited for the study, following their written informed consent. Samples were collected from both the affected and non-affected family members. In this chapter, details will be given of methodology used such as how DNA was extracted and purified and how whole genome scans were performed using GeneChip® Mapping 250K (NspI) Assay (Affymetrix®). A description will also be given of the procedures undertaken in order to ensure quality control of genotyping data using various software packages and how Allegro v2.0 was used to perform the linkage analysis. STR markers were also used to confirm the linkage interval. DNA sequencing of candidate genes was performed and bioinformatic tools were used to predict the structure of pre-mRNA and alternative splicing of the allele present in the affected individuals.

2.2 Recruitment of the family with epilepsy

The study was approved by the Ethics Research Committee of the Faculty of Medicine and Surgery, University of Malta. The family was identified through a proband who had attended the Paediatric Out-Patients clinic at St. Luke’s Hospital, the former largest general hospital in Malta, and who was referred to the researcher by the attending neurologist. The family of the proband was in fact found to be a three generation family with a number of family members having febrile seizures or epilepsy. The ideal family for linkage is indeed a kindred with at least three
generations, many matings and a large number of offspring (Vogel, 1997). This type of pedigree is indeed the best pedigree for detecting linkage since the number of informative meiosis is high. Thus by carrying out a complete genome scan there is a very high probability of identifying the loci where the disease co-segregates.

Figure 2.1. Pedigree of the family. Affected individuals are shown in black.

Following written informed consent, the family members, both affected and non-affected, were recruited for the study. The family had a number of affected individuals (refer to Fig. 2.1) which had either febrile seizures or else epilepsy. Thirteen family members were recruited. In linkage analysis it is very important that non-affected individuals are also included in the study. In order to determine linkage to a genomic interval, haplotypes of affected individuals have to be compared to that of the non-affected. Non-affected individuals should have a different haplotype from that of the affected individuals in the region that shows linkage.
2.3 Syndrome characterization

As described in Chapter 1, the ILAE Classification Commission classify febrile seizures as a condition with seizures that do not require a diagnosis of epilepsy (Section 1.9.4). Although it is known that febrile seizures are thought of as separate clinical entities, a genetic susceptibility to febrile seizures may cause epilepsy in other family members or later on in the same family member (Iwasaki, 2002).

The family under investigation was indeed a three generation family which has seven affected individuals (Fig. 2.1). Five of the family members had had febrile seizures from about nine months to about four years of age. Another member (family member 8 of the third generation, female, whose parents did not have seizures) had a febrile seizure at age six and since fourteen years of age has nearly weekly, panic attacks while another member (family member 14, third generation, male, whose parents did not have seizures) never had febrile seizures but developed focal onset seizures at age eight and also during the same period started having severe tantrums (now both well controlled with sodium valproate).

Following discussions with the attending neurologists, this family was characterised as having GEFS+ syndrome or Febrile Seizures (Gourfinkel-An I, 2004; Serratosa J, 1999). The type of febrile seizures in this syndrome are generalized tonic-clonic convulsions. Intelligence and neurological development were found to be normal except in family member 10 (third generation, male, whose father had had febrile seizures) who had moderate mental retardation.
2.4 DNA Extraction

2.4.1 Sample collection from affected and non-affected members of the family

A 6ml blood sample in an EDTA tube, or a buccal rinse was collected from affected and non-affected family members that consented to participate. The blood samples were stored at 2-8°C before they were processed. In order to avoid venipuncture in children and in adults who were reluctant to undergo the procedure, a buccal rinse was used as a source of DNA. To collect a buccal rinse, the family members were instructed to take about 20-50ml of sterilized saline swirl it in the mouth for a minute and then dispense it in a sterile centrifuge tube. The buccal rinses were processed within 2 hrs of collection.

2.4.2 DNA extraction from blood samples

The DNA extraction was carried out using AccuPrep® Genomic DNA Extraction Kit (Bioneer Corp.). The AccuPrep® Genomic DNA Extraction Kit (Cat. No: K-3032) employs glass fibres, fixed in a column, that specifically binds DNA in the presence of a chaotrophic salt. Proteins and other contaminants are eliminated through a series of short wash and spin steps. The genomic DNA is eluted by a low salt solution. The DNA extraction from blood was carried out according to the manufacturer’s instructions.

Using a filter tipped pipette, 200µl of blood were added to a labelled 1.5ml tube to which 20µl of Proteinase K and 200µl of Binding Buffer had been added. The tube
was then closed, vortexed to obtain a homogenous mix and incubated at 60°C for 10 minutes. Following incubation, 100μl of isopropanol were added to the tube and, the tube was vortexed. The mix was then transferred to an extraction column which was placed in a receiving tube. The assembly was then centrifuged at 8,000rpm for 1 minute. The receiving tube with the filtrate was discarded and the extraction column was placed in a new receiver tube. Wash Buffer 1 (500μl) were added to the extraction column and centrifugation at 8,000rpm for 1 minute was repeated. The receiving tube with the filtrate was discarded and the extraction column was placed in a new receiver tube. Wash Buffer 2 (500μl) were added to the extraction column and centrifugation at 13,000rpm for 1 minute was repeated. The receiving tube with the filtrate was discarded and the extraction column was placed in a new receiver tube. The assembly was centrifuged at 13,000rpm for 1 minute to remove all traces of Wash Buffer which could act as inhibitors in downstream reactions. The extraction column was placed in a labelled 1.5ml tube and 200μl of Elution Buffer was added to the column. The assembly was incubated at room temperature for 5 minutes after which it was centrifuged at 8,000rpm for 1 minute. The Extraction column was discarded and the 1.5ml tube containing the DNA was stored at 2-8°C.

2.4.3 DNA Extraction from buccal rinse

For the DNA extraction from the buccal rinse, the AccuPrep® Genomic DNA Extraction Kit (Cat. No: K-3032) was used. This kit does not have a protocol on how to extract DNA from a buccal rinse. A cell pelleting step was included before the actual manufacturer’s instruction could be followed as described below.
In order to extract the DNA from a buccal rinse, the centrifuge tubes containing the samples were centrifuged for 15 minutes at 3000g. The supernatant was discarded and 20μl of Proteinase K and 200μl of Lysis Buffer were added to the cell pellet and the cell pellet was resuspended by pipetting up and down. The suspension was transferred to a labelled 1.5 ml tube and the tube was incubated at 60°C for 3 hours. Then 200μl of Binding Buffer were added to the 1.5ml and the tube was vortexed to obtain a homogenous mix. The tube was incubated at 60°C for 10 minutes. Following incubation, 100μl of isopropanol were added to the tube and the tube was vortexed. The mix was then transferred to an extraction column which was placed in a receiving tube. The assembly was then centrifuged at 8,000rpm for 1 minute. The receiving tube with the filtrate was discarded and the extraction column was placed in a new receiver tube. 500μl of Wash Buffer 1 were added to the extraction column and centrifugation at 8,000rpm for 1 minute was repeated. The receiving tube with the filtrate was discarded and the extraction column was placed in a new receiver tube. 500μl of Wash Buffer 2 were added to the Extraction column and centrifugation at 13,000rpm for 1 minute was repeated. The receiving tube with the filtrate was discarded and the Extraction column was placed in a new receiver tube. The assembly was centrifuged at 13,000rpm for 1 minute to remove all traces of Wash Buffer which could act as inhibitors in downstream reactions. The extraction column was placed in a labelled 1.5ml tube and 200μl of Elution Buffer was added to the column. The assembly was incubated at room temperature for 5 minutes after which it was centrifuged at 8,000rpm for 1 minute. The extraction column was discarded and the 1.5ml tube containing the DNA was stored at 2-8°C.
2.4.4 Determination of the quality of DNA

Since there was no literature on the use of the AccuPrep® Genomic DNA Kit for the extraction of DNA from a buccal rinse, it was important to assess the quality of the extracted DNA (non-degraded DNA at a concentration similar to that obtained from blood and without PCR inhibitors present).

2.4.4.1 Estimation of the concentration of DNA extracted from buccal rinse

In order to determine the approximate yield of DNA, as compared to DNA extracted from blood, agarose gel electrophoresis was carried out. A 1% agarose gel was prepared by adding 0.4mg of agarose to 40ml of 1X TAE and then boiling the mixture until all the agarose melted. Ethidium bromide was added to the mixture to a final concentration of 0.5µg/ml. The mixture was poured in a gel tray to set. Combs of the required size were placed in the gel before it set in order to form wells for loading of the samples. The gel was left to set. Following setting of the gel, the gel was placed in 1X TAE buffer in an electrophoresis chamber. In a 96 well plate, 2µl of DNA loading buffer were placed in every well. DNA (5µl) samples were loaded in the wells containing the loading buffer. The sample was mixed with the loading buffer by pipetting and then loaded in the well of the gel. When all DNA samples were loaded (DNA extracted from blood and DNA extracted from buccal rinse), the gel was run for 15 minutes at 5V/cm. The gel was removed from the tray and visualized under UV light. The DNA bands of the DNA extracted from the buccal rinse had the same intensity as those of the DNA extracted from blood and no DNA degradation was observed (one sharp band and no smears).
2.4.4.2 DNA purity

In some instances, an extraction procedure does not produce DNA of acceptable purity, due to the presence of PCR inhibitors which might interfere with the successful amplification of DNA (Bassetti, 2007). In order to determine if the AccuPrep® Genomic DNA Extraction Kit removed all inhibitors when it was used to extract DNA from buccal rinses, a PCR was carried out to determine if inhibitors were present in the sample extracted from the buccal rinses.

The following PCR reaction mix that amplifies a fragment of 440bp, the Hypervariable region 1 of the D-loop of the mitochondrial DNA, was prepared according to Table 2.1.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Manufacturer</th>
<th>Volume per reaction</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X PCR buffer</td>
<td>Solis Biodyne</td>
<td>5.0 µl</td>
<td>2.5 mM MgCl₂</td>
</tr>
<tr>
<td>Hot FirePol (Taq DNA Polymerase)</td>
<td>Solis Biodyne</td>
<td>0.25</td>
<td>0.05 U/µl</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Solis Biodyne</td>
<td>2.0 µl</td>
<td>0.2 nmol/µl</td>
</tr>
<tr>
<td>HV1-Forward Primer</td>
<td>Bioneer Corp.</td>
<td>0.25 µl</td>
<td>0.5 pmol/µl</td>
</tr>
<tr>
<td>HV1-Reverse Primer</td>
<td>Bioneer Corp.</td>
<td>0.25 µl</td>
<td>0.5 pmol/µl</td>
</tr>
<tr>
<td>DNA</td>
<td>1 µl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Up to 25 µl</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

For amplification the following thermal profile on the ABI9700 thermal cycler (Applied Biosystems) was used as shown in Table 2.2.
Table 2.2. Thermal profile for the amplification of HV1.

<table>
<thead>
<tr>
<th></th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Denaturation</td>
<td>95°C</td>
<td>for 5 minutes</td>
<td>1 cycle</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>for 1 minute</td>
<td>30 cycles</td>
</tr>
<tr>
<td>Annealing</td>
<td>55°C</td>
<td>for 1 minute</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>for 1 minute</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>for 10 minutes</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>

Following amplification, the PCR products were run on a 2% agarose gel and visualized under UV transillumination. The yield of the PCR products obtained from the DNA extracted from the buccal rinse was the same as that of the PCR products obtained from the DNA extracted from blood. This confirmed that the DNA extracted from buccal rinses could be used for further downstream reactions.

2.5 SNP genotyping

2.5.1 Purification of DNA for SNP genotyping

In order to further purify the DNA such that the DNA could be used for SNP genotyping using GeneChip® technology, the Microcon YM-100 Centrifugal Filter Units were used. The Microcon centrifugal YM-100 Centrifugal Filter (Millipore) provide a fast, efficient means for concentrating RNA/DNA samples, either single- or double stranded. These filters can be used to remove any salts or macromolecules from the DNA solution that could inhibit further downstream reaction. The manufacturers’s protocol for the use of Microcon® centrifugal YM-100 Centrifugal Filter was followed.
A sample reservoir was placed in a filtrate vial and 200μl of DNA solution were pipetted in the sample reservoir without touching the membrane. The lid of the filtrate vial was used to close the sample reservoir. The assembly was centrifuged at 500g for 12 minutes. The sample reservoir was removed from the filtrate vial, inverted and placed into a new filtrate buffer. About 100μl of water were added to the inverted sample reservoir and after incubation at room temperature for 5 minutes, the assembly was centrifuged at 1000g for 3 minutes. The filtrate was stored in fridge at 2-8°C until pending genotyping.

2.5.2 SNP genotyping using GeneChip® Mapping 250K (NspI) Assay Kit (Affymetrix®, Santa Clara, CA).

The GeneChip® Mapping 250K (NspI) Assay Kit forms part of the GeneChip® Mapping 500K Assay Set which has been widely used in genetic studies. The GeneChip® Mapping 250K (NspI) Assay Kit genotypes an average of 262,000 SNPs and only 250ng of sample DNA is required for this assay.

SNP genotyping of 12 family members was subcontracted to Case Western Reserve University (Wolstein Research Building, Cleveland, OH) since the necessary technology is not locally available.

The GeneChip® data of the 12 samples was sent electronically and the data was opened using Data Transfer Tool (DTT) and GeneChip® Operating Software (GCOS). This software automates the control of GeneChip® Fluidics Stations and Scanners but also acquires data, manages sample and experimental information, and supports GeneChip® Genotyping Analysis Software (GTYPE) and Genotyping
Console (Affymetrix®) which can be used both for performing genotyping and Copy Number Variation analysis. In GCOS it is possible to see the actual scans (*.DAT) of the GeneChips analysed and also the signal intensity (*.CEL). The GTYPE software was used to convert the scanned chip image in the genotypes of the family members. The genotype data was exported as a textfile (*.txt). ALOHOMORA software (Rüschendorf and Nürnberg, 2005)(http://gmc.mdc-berlin.de/alohomora/) was used to subject the genotype data to standard quality control routines before linkage analysis and to convert the genotype file into an appropriate linkage format file.

For quality control of data, ALOHOMORA v0.29 makes use of other freely available software: Graphical Relationship Representation (GRR) (http://www.sph.umich.edu/csg/abecasis/index.html) (Abecasis et al., 2001) in order to check for the correct relationship between family members, PedCheck (http://watson.hgen.pitt.edu/register)(O'Connell and Weeks 1997) in order to check for Mendelian errors and Merlin (Abecasis et al., 2001) to delete unlikely genotypes (http://www.sph.umich.edu/csg/abecasis/index.html). Uninformative markers were also removed from the genotype file. The quality controlled genotype data was then converted into linkage format for analysis by Allegro v2.0 (http://www.decode.com) (Gudbjartsson, 2005). HaploPainter v027 (http://haplopainter.sourceforge.net/) (Thiele and Nurnberg, 2005) was used to generate haplotype figures using data generated by Allegro v2.0. For Copy Number Variation Analysis, the software Genotyping Console™ (Affymetrix®) was used.
2.6 Linkage analysis of SNP genotypes

In the case of dominant inheritance a single allele is necessary for the expression of the phenotype while in the case of an autosomal recessive trait, two copies of the trait allele must be present for the expression of the phenotype (Speer, 1998).

In the family under investigation, although the pattern of inheritance appears to conform to Mendel’s laws, the fact that family members 2 and 6 (Figure 2.1) had never experienced any seizures, although their offspring are affected, could indicate a more complex mode of inheritance. In this family the condition was thus transmitted as an autosomal trait with incomplete penetrance and is phenotypically heterogeneous (Gourfinkel-An, 2004). Complex inheritance refers to traits that result from the influence of variation at more than one locus (polygenic) or where interaction with the environment might result in the expression of the phenotype (multifactorial) (Serratosa, 1999b).

Linkage analysis was carried out using Allegro v2.0, the deCode map and Caucasian SNP frequency data as provided by Affymetrix®. For the analysis, two phenotypes for affected individuals were analysed:

Phenotype A - An affected individual was defined as a family member that had had a witnessed seizure during his/her lifetime.

Phenotype B - An affected individual was defined as a family member that had started having febrile seizures in his/her first years of life which remitted before age six years.
Penetrance was varied from 1 to 0.5 to determine at which penetrance the highest parametric LOD is observed.

The HapMap Genome Browser B36 (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/) was used to analyse the degree of linkage disequilibrium in the regions where the highest LOD scores were obtained.

2.7 Short tandem repeat (STR) genotyping for the confirmation of an identified linkage interval

2.7.1 Amplification of STRs

Linkage analysis of SNP genotyping data identified loci that linked to the epilepsy/febrile seizures phenotype. One of the main loci was a region of 20cM on chromosome 20. To confirm the linkage of epilepsy/febrile seizures to Chromosome 20 that was determined by analysis of the NspI GeneChip®, the family was genotyped for a number of STR markers spaced at about 10cM throughout Chromosome 20 (CHLC markers version 10 screening set). Most markers having tetranucleotide repeats were chosen as this facilitated the analysis of the alleles after capillary electrophoresis. The primers were fluorescently labelled such that they can be analysed using capillary electrophoresis (Genetic Analyzer 3130 (Applied Biosystems). Table 2.3 is a list of the STR markers together with the primer sequences used for the amplification of the DNA loci. The genetic position of the markers is given in cM (deCode map).
Table 2.3. STR markers on chromosome 20. Genetic position is in cM according to the deCode map.

<table>
<thead>
<tr>
<th>Marker</th>
<th>Genetic position (cM)</th>
<th>cM between markers</th>
<th>Repeat and size (bp)</th>
<th>Fluorescently labelled primers</th>
</tr>
</thead>
<tbody>
<tr>
<td>D20S103</td>
<td>2</td>
<td>2</td>
<td>Di</td>
<td>5'-FAMGTTCATAGGGGACACACAGT-3' 5'-CCATGATTTGGTAATCACA-3'</td>
</tr>
<tr>
<td>D20S482</td>
<td>12</td>
<td>10</td>
<td>Tetra</td>
<td>5'-FAMCGCTCCATAACCACATGA-3' 5'-GAACCTAAACTCTAGGAAACG-3'</td>
</tr>
<tr>
<td>D20S851</td>
<td>25</td>
<td>13</td>
<td>Di</td>
<td>5'-FAMCTTCAAGTTATGTGTCACAC-3' 5'-GCCGAGACTCTGACATCA-3'</td>
</tr>
<tr>
<td>D20S1143</td>
<td>36</td>
<td>11</td>
<td>Tetra</td>
<td>5'-HEXGTTCTCCAGTGACAGACG-3' 5'-GGATCTTTTCAGCCTCATT-3'</td>
</tr>
<tr>
<td>D20S477</td>
<td>48</td>
<td>12</td>
<td>Tetra</td>
<td>5'-FAMGGATAACTCAGGGGTCAGT-3' 5'-TAGACCGTCCCTTTAACA-3'</td>
</tr>
<tr>
<td>D20S478</td>
<td>54</td>
<td>6</td>
<td>Tetra</td>
<td>5'-FAMCCAAGCAAGTTGTGTTATCTG-3' 5'-TGTTATAGCCAATAGAAAATGG-3'</td>
</tr>
<tr>
<td>D20S481</td>
<td>62</td>
<td>8</td>
<td>Tetra</td>
<td>5'-FAMCTGTATAGGTACAGACAC-3' 5'-AACGCAGAAAGCATACACAC-3'</td>
</tr>
<tr>
<td>D20S159</td>
<td>70</td>
<td>8</td>
<td>Tetra</td>
<td>5'-HEXAGTGGAGACACACATCCA-3' 5'-CCTCTCCCATGACGTTTCC-3'</td>
</tr>
<tr>
<td>D20S480</td>
<td>80</td>
<td>10</td>
<td>Tetra</td>
<td>5'-FAMGCTGTGACACACTATGG-3' 5'-AACGAATACACACTATTTC-3'</td>
</tr>
<tr>
<td>D20S451</td>
<td>90</td>
<td>10</td>
<td>Tetra</td>
<td>5'-FAMTCGTGAGACACACTATGG-3' 5'-TCCAGTGAGCAGATGG-3'</td>
</tr>
<tr>
<td>D20S164</td>
<td>96</td>
<td>6</td>
<td>Tetra</td>
<td>5'-FAMAGCGCTGGGACACACTCG-3' 5'-GGGTCTAGGTGCTGCTCA-3'</td>
</tr>
<tr>
<td>D20S173</td>
<td>98</td>
<td>2</td>
<td>Di</td>
<td>5'-FAMATCCACCTGCCACTTA-3' 5'-CCAGAGACTCGTGGACTCAT-3'</td>
</tr>
</tbody>
</table>

The Reaction Mix shown in Table 2.4 (except the DNA) was prepared and dispensed in labelled 200µl PCR tubes. The DNA was then added to the respective tube. The samples were subsequently placed in the ABI9700 Thermal cycler and the thermal profile shown in Table 2.5 was used.
Table 2.4. Reagent Mix for the amplification of the STRs.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Manufacturer</th>
<th>Volume per reaction</th>
<th>Final Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X PCR buffer</td>
<td>Solis Biodyne</td>
<td>2 µl</td>
<td>2.5mM MgCl₂</td>
</tr>
<tr>
<td>Hot FirePol (Taq DNA Polymerase)</td>
<td>Solis Biodyne</td>
<td>0.1 µl</td>
<td>0.01U/µl</td>
</tr>
<tr>
<td>dNTPs</td>
<td>Solis Biodyne</td>
<td>0.8 µl</td>
<td>0.2nmol/µl</td>
</tr>
<tr>
<td>Forward Primer</td>
<td>Bioneer Corp.</td>
<td>0.1 µl</td>
<td>0.05pmol/µl</td>
</tr>
<tr>
<td>Reverse Primer</td>
<td>Bioneer Corp.</td>
<td>0.1 µl</td>
<td>0.05pmol/µl</td>
</tr>
<tr>
<td>DNA</td>
<td>1 µl</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td></td>
<td>Up to 10 µl</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.5. Thermal profile for the amplification of the STRs.

<table>
<thead>
<tr>
<th>Stage</th>
<th>Temperature</th>
<th>Time</th>
<th>Repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial Denaturation</td>
<td>95°C</td>
<td>2 minutes</td>
<td>1 cycle</td>
</tr>
<tr>
<td>Denaturation</td>
<td>95°C</td>
<td>45 seconds</td>
<td>30 cycles</td>
</tr>
<tr>
<td>Annealing</td>
<td>55°C</td>
<td>45 seconds</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>1 minute</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>6 minutes</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>

2.7.2 Electrophoresis of the fluorescently labelled PCR products

A mix of formamide and LIZ500 (Applied Biosystems) (an internal lane standard 50-500bp, fluorescently labelled with the dye LIZ) was prepared (9 µl water and 0.25 µl LIZ500 per sample). The mix was dispensed in the sequencing plate. Then 1 µl of every fluorescently labelled PCR product was added to its respective well. The Sequencing plate was covered with a septum and the samples were denatured for 3 minutes at 95°C using the ABI9700 thermal cycler (Applied Biosystems), after which the plate was placed on ice for 3 minutes. The samples were then loaded on the ABI3130 Genetic Analyzer (Applied Biosystems). The Data Collection software v3.0 was used to set up the sample run while GeneMapper™ software was used in order to analyse and size the PCR products.
2.7.3 Linkage analysis of the STR genotyping data

Cyrillic 2.1 (http://www.cyrillicsoftware.com/) was used to draw the pedigree and to input the STR marker data. MLINK was used to convert the genotyping data of Cyrillic 2.1 to a linkage data file that can be used by Genehunter (gh2.1) (http://helix.nih.gov/apps/bioinfo/genehunter.html) (Kruglyak et al., 1996). Parametric linkage analysis (LOD) and non-parametric linkage analysis (NPL) were performed using Genehunter (penetrance 1.0, phenecopy rate 0.01, equal allele frequencies, genetic distance - deCode map). Two-point linkage analysis at different recombination fractions was carried out using MLINK (easyLINKAGE Plus).

2.8 Sequencing of candidate genes in the linkage interval on Chromosome 20

The chromosomal interval that was found to be linked to the phenotype A (Section 2.6) consisted of a region of about 6Mbp having 132 genes. NPL peaked in a region of about 364Kb and thus it was decided to start DNA sequencing of genes in this interval. In this region there are 18 genes. The best candidate genes in the region were KCNQ2 and CHRNA4 as they both code for ion channels and had already been associated with seizures (Table 1.2). The next gene just outside the linkage interval is stathmin-like 3 (STMN3) which is involved in neurite development.
2.8.1 PCR amplification of exons of genes in the linkage interval

Primer design was carried out using Oligo Explorer 1.2 (GeneLink™) (http://www.genelink.com/tools/gl-oe.asp). Tables 2.9, 2.10 and 2.11 describe the primers designed for DNA sequencing of the exons and intron/exon junctions of KCNQ2, CHRNA4 and STMN3 respectively. PCRs were optimised using the thermal gradient cycler, Eppendorf Master Cycler. All PCR reactions were carried out using the following Master Mix (Table 2.6) except the ones marked with an asterix (*) which were carried out according to Master Mix (Table 2.7).

Table 2.6. Reaction Mix for the amplification of the exons of CHRNA4, KCNQ2 and STMN3.

<table>
<thead>
<tr>
<th></th>
<th>Volume used per sample</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Buffer</td>
<td>5µl</td>
<td>2.5mM MgCl₂</td>
</tr>
<tr>
<td>dNTPs (Solis Biodyne)</td>
<td>2µl</td>
<td>2nmol/µl</td>
</tr>
<tr>
<td>Primer F (Bioneer Corp.)</td>
<td>0.25µl</td>
<td>0.5pmol/µl</td>
</tr>
<tr>
<td>Primer R (Bioneer Corp.)</td>
<td>0.25µl</td>
<td>0.5pmol/µl</td>
</tr>
<tr>
<td>Taq Polymerase (Solis Biodyne)</td>
<td>0.5µl</td>
<td>0.1U/µl</td>
</tr>
<tr>
<td>DNA</td>
<td>1µl</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Up to 25µl</td>
<td></td>
</tr>
</tbody>
</table>
Table 2.7. Reaction Mix for the amplification of the exons of CHRNA4, KCNQ2 and STMN3 that were not successfully amplified using the reaction mix in Table 2.6.

<table>
<thead>
<tr>
<th></th>
<th>Volume used per sample</th>
<th>Final concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Buffer</td>
<td>5µl</td>
<td>2.5mM MgCl₂</td>
</tr>
<tr>
<td>dNTPs (Solis Biodyne)</td>
<td>2µl</td>
<td>2nmol/µl</td>
</tr>
<tr>
<td>Primer F (Bioneer Corp.)</td>
<td>0.25µl</td>
<td>0.5pmol/µl</td>
</tr>
<tr>
<td>Primer R (Bioneer Corp.)</td>
<td>0.25µl</td>
<td>0.5pmol/µl</td>
</tr>
<tr>
<td>Taq Polymerase (Solis Biodyne)</td>
<td>0.5µl</td>
<td>0.1U/µl</td>
</tr>
<tr>
<td>Betaine (Sigma)</td>
<td>5.00µl</td>
<td>0.2M</td>
</tr>
<tr>
<td>DNA</td>
<td>1µl</td>
<td></td>
</tr>
<tr>
<td>Water</td>
<td>Up to 25µl</td>
<td></td>
</tr>
</tbody>
</table>

Table 2.8. Thermal profile used for the amplification of the exons of CHRNA4, KCNQ2 and STMN3.

<table>
<thead>
<tr>
<th></th>
<th>97°C</th>
<th>15 minute</th>
<th>1 cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>See Tables 2.9, 2.10 and 2.11.</td>
<td>1 minute</td>
<td>35 cycles</td>
</tr>
<tr>
<td>Denaturation</td>
<td></td>
<td>1 minute</td>
<td></td>
</tr>
<tr>
<td>Annealing</td>
<td></td>
<td>1 minute</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72°C</td>
<td>1 minute</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72°C</td>
<td>10 minutes</td>
<td>1 cycle</td>
</tr>
</tbody>
</table>

Figures 2.2, 2.3 and 2.4 are a graphical representation of the exons in the reference sequence of KCNQ2, CHRNA4 and STMN3 respectively.
Gene KCNQ2

Figure 2.2. Transcript structure of KCNQ2 (Ensembl ENSG0000075043) showing 16 exons. The filled in boxes represent exonic coding regions (cds).

Table 2.9. Sequences of primers for the amplification of exons of KCNQ2, size of PCR product and annealing temperature. The amplicon included the relevant exon-intron boundaries.

<table>
<thead>
<tr>
<th>Exon</th>
<th><strong>Forward and reverse primers</strong></th>
<th><strong>PCR Product Size</strong></th>
<th><strong>Annealing Temperature</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>5' - CCTTCCGCCTCTCCCTCTTTCT - 3'</td>
<td>590bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5' - CGTAGGCGGGTGATAGAAGAC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5' - GTGTTTGAGGATGATGAGG - 3'</td>
<td>637bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>5' - AGGATGGGGCTGATTACAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>5' - GTCAGCTGCCAGATCTCTTCCTTG - 3'</td>
<td>458bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>5' - CTTCCCTCTGAGATTAGAGGC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>5' - AACAGCCGGGGTGAGAG - 3'</td>
<td>541bp</td>
<td>63°C</td>
</tr>
<tr>
<td></td>
<td>5' - AGAAAGCCGGAGAGAAGAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>5' - ATGGTTGAGAGTTGATGAG - 3'</td>
<td>574bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - CTCGGCACTCTCAGACTCAGAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5' - GTCCTAGCTGGATGACTAG - 3'</td>
<td>417bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GGCACGGAGAGAAGAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>5' - GGCTGACCCCGAGATGAG - 3'</td>
<td>478bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GGGAGCCCTTCGTGAGACCG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>5' - TGCTCTACGAGGTTAAAATGAG - 3'</td>
<td>464bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - TACCCAGAGGAGAGACAC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>5' - ACAGCGAGAGATGGAGAG - 3'</td>
<td>640bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - TACGGGACCTGAGGATAC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>5' - AGACCCCCAGAGTGACAGAG - 3'</td>
<td>415bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - CACACCGGGACGACGACCA - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>5' - TCTCCTCCCTGTTTCTGTC - 3'</td>
<td>541bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GTTTCCTCTCGTGGTGATGAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>5' - GTCCTTTAGGCSSCGATATC - 3'</td>
<td>484bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GTCCTAGCCCTCCTTCTTCCC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>5' - ACTGGGAAAGAAGAGGCTAGAGAC - 3'</td>
<td>857bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GCATACAACCACCAACATAC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>5' - CTTGTAGCTACGAGAGAATGTGC - 3'</td>
<td>411bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - ACCTGAGGAGGAAATCCTCAC - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>5' - TTCTACTCTCCGTATAGTGTC - 3'</td>
<td>500bp</td>
<td>62°C</td>
</tr>
<tr>
<td></td>
<td>5' - GCCATGTAACCTACCTAGGCTG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16-fragment A</td>
<td>5' - CTCCTTTTTGCGAGGTCTTGTC - 3'</td>
<td>591bp</td>
<td>62°C</td>
</tr>
<tr>
<td>16-fragment B</td>
<td>5' - CGGGTACGCTTGGTGTAGAG - 3'</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>5' - TTCCGGCCACACTACAGTTAC - 3'</td>
<td>686bp</td>
<td>62°C</td>
</tr>
</tbody>
</table>
Chapter 2 - Methodology

CHRNA4

Figure 2.3. Transcript structure of CHRNA4 (Ensembl ENSG00000370263) showing 6 exons. The filled in boxes represent exonic coding regions (cds).

Table 2.10. Sequences of primers for the amplification of exons of CHRNA4, size of PCR product and annealing temperature. The amplicon included the relevant exon-intron boundaries.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Forward and reverse primers</th>
<th>PCR Product Size</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>5’-TCCCCCTCCCTCCCTTTTCAGTC-3’</td>
<td>590bp</td>
<td>54°C</td>
</tr>
<tr>
<td></td>
<td>5’-CCCCACCTCTCTCTCAAGTTC-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>5’-TGTGAGGGAGCTGAAAGGAAG-3’</td>
<td>546bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-CTGGGCTCTGTACTGAGAAC-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 and 4</td>
<td>5’-TGCCAGTGGCTGTAGATGAG-3’</td>
<td>695bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-TGCTTCGTCTTCTCTTTTC-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-fragment A</td>
<td>5’-CTCCCTGTGAGTGGAATGC-3’</td>
<td>551bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-CTTTTCTGTTGTGGGTAG-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-fragment B</td>
<td>5’-GACCCTACAAACACCAGGAGATAC-3’</td>
<td>845bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-ATCGTCTCGGGGAACACAGTAC-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5-fragment C</td>
<td>5’-CGCCCTCACCAGCTTCTGT-3’</td>
<td>813bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-ATCGTCTCRTCTCAAAAGCCA-3’</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>5’-CCATACATGGTGATGAGGC-3’</td>
<td>686bp</td>
<td>65°C</td>
</tr>
<tr>
<td></td>
<td>5’-GAGACTGGGGAGTGGTATTC-3’</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 2.11. Sequences of primers for the amplification of exons of STMN3, size of PCR product and annealing temperature. The amplicon included the relevant exon-intron boundaries.

<table>
<thead>
<tr>
<th>Exon</th>
<th>Forward and reverse primers</th>
<th>PCR Product Size</th>
<th>Annealing temperature</th>
</tr>
</thead>
<tbody>
<tr>
<td>1*</td>
<td>5’-CCTCACAAAGCCAAGTTTGAG-3’ 5’-GGGAAATTGGCCGTGCCGTGC-3’</td>
<td>562bp</td>
<td>65°C</td>
</tr>
<tr>
<td>2</td>
<td>5’-GAGGCGCAGAGAGACAGAAG-3’ 5’-CAGGGCAAGCCAGACAGTAGAG-3’</td>
<td>606bp</td>
<td>65°C</td>
</tr>
<tr>
<td>3</td>
<td>5’-GTGGAGAGGGAGGTCTAGG-3’ 5’-GTCTTTGAAAACCTCTCGACTCC-3’</td>
<td>584bp</td>
<td>65°C</td>
</tr>
<tr>
<td>4*</td>
<td>5’-GGGACGCCCAGTAACACAG-3’ 5’-GTCCTGAGAAGCCAGCAAAAGC-3’</td>
<td>951bp</td>
<td>65°C</td>
</tr>
<tr>
<td>5-fragment A</td>
<td>5’-TGAAAGTGCCCTGAAGGAAG-3’ 5’-AGACACGCCGCAGTGAAAGAC-3’</td>
<td>664bp</td>
<td>65°C</td>
</tr>
<tr>
<td>5-fragment B</td>
<td>5’-ATGCTTCTCTTCCCGAAGCTCAACGC-3’ 5’-GTCTCAAGAGATCTCTCGACTCC-3’</td>
<td>690bp</td>
<td>62°C</td>
</tr>
<tr>
<td>5-fragment C</td>
<td>5’-TGACAGATGCTCTTGAAAGAC-3’ 5’-CCACCATGAGGGATTATCCC-3’</td>
<td>727bp</td>
<td>65°C</td>
</tr>
<tr>
<td>5-fragment D</td>
<td>5’-TTTGTGGGAGTCAAGAAGAAG-3’ 5’-AGGAAAAGTGGCTGTGTTG-3’</td>
<td>376bp</td>
<td>65°C</td>
</tr>
</tbody>
</table>

2.8.2 PCR product purification

For further downstream reactions, the PCR product must be cleaned. The kit genPURE for PCR Cleaning (Genetix Ltd. Cat No. K3506) was used. The genPURE-PCR Purification Systems, have been specially developed for the fast, effective purification of PCR products. For the cleaning of the PCR products prior to cycle sequencing genPURE columns were used. The manufacturer’s protocol was followed.
Chapter 2 - Methodology

Twenty microlitres of PCR product were transferred into a 1.5ml tube and 500μl of Binding Buffer were added. The solution was pipetted up and down to get a uniform mix and then transferred into a pre-labelled genPURE spin column that has been placed into a collection tube. The assembly was centrifuged at 14,000g for 1 minute. The filtrate in the collection tube was discarded and the collection tube was replaced under the genPURE spin column. The spin column was opened and 700μl of Wash Buffer (to which ethanol has been already added according to manufacturer’s instructions) were added. The column was centrifuged for 2 minutes at 14,000g. The filtrate in the collection tube was discarded and the collection tube was replaced under the Spin column. The assembly was centrifuged for 3 minutes at 14,000g to remove any residual Wash Buffer. The genPURE spin column was then transferred to a pre-labelled 1.5ml tube and about 20μl of water were added to the centre of the membrane of the spin column. The assembly was incubated at room temperature for 1 minute and then centrifuged for 1 minute at 14,000g. The spin column was discarded and the cleaned PCR product was stored at -20°C.

2.8.3 DNA sequencing

DNA sequencing of amplified products, was carried out using the BigDye® Terminator Kit 3.1 cycle sequencing Kit (Applied Biosystems Cat. No. 4337455). This mix contains DNTPs (deoxynucleotides), fluorescently labelled ddNTPs (dideoxy nucleotides) and Taq polymerase in specific proportions suitable for Sanger sequencing which can be defined as the resolution of dideoxy terminated fragments using capillary electrophoresis (Luckey et al., 1990). The reaction mix (excluding the cleaned PCR product) was prepared according to Table 2.12 and dispensed in
labelled 200µl tubes. To every tube 2µl of the respective PCR product was added. The products were placed in the ABI9700 thermal cycler (Applied Biosystems) and submitted to the thermal profile shown in Table 2.13.

Table 2.12. Cycle sequencing reaction mix.

<table>
<thead>
<tr>
<th>Ingredient</th>
<th>Volume for 1 sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>BigDye® Terminator Kit 3.1 cycle sequencing Kit Reagent</td>
<td>2µl</td>
</tr>
<tr>
<td>Primer (Forward or reverse) at a concentration of 5pmol/µl</td>
<td>1µl</td>
</tr>
<tr>
<td>Cleaned PCR product</td>
<td>2µl</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 10µl</td>
</tr>
</tbody>
</table>

Table 2.13. Thermal cycling for cycle sequencing.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature</th>
<th>Time</th>
<th>Cycle</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>96°C</td>
<td>1 minute</td>
<td>1 cycle</td>
</tr>
<tr>
<td>Denaturation</td>
<td>96°C</td>
<td>10 seconds</td>
<td>25 cycles</td>
</tr>
<tr>
<td>Annealing</td>
<td>50°C</td>
<td>5 seconds</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>60°C</td>
<td>4.00 minutes</td>
<td></td>
</tr>
</tbody>
</table>

2.8.4 Purification of the cycle sequencing products

The genCLEAN column (Genetix Ltd., Cat No. K1005) allows the quick and easy removal of excess dye terminator prior to running samples on automated DNA sequencers. This cleaning procedure results in superior sequencing electropherograms through background reduction. Cleaning of cycle sequencing products eliminate “dye blob” problems and this results in an increase in read-length and better accuracy of the DNA sequence.

The manufacturer’s protocol was followed for sample processing. The genCLEAN column was opened from both ends and placed in a collection tube. The assembly
was centrifuged at 1,000g for 4 minutes. The filtrate was discarded from the collection tube and the collection tube was replaced under the column. About 200μl of water were added to the column and the assembly was centrifuged again at 1,000g for 4 minutes. The Collection tube was discarded and the geneCLEAN column was placed in a labelled 1.5ml tube. The cycle sequencing product was added to the centre of the column without the pipette tip touching the column. About 10μl of water were also added to the column. The assembly was centrifuged at 1,000g for 3 minutes. The column was discarded and the filtrate was transferred to an Applied Biosystems sequencing plate, covered with a septum and denatured at 94°C for 3 minutes. Following denaturation, the plate was placed on ice for 3 minutes.

2.8.5 Electrophoresis and analysis of the cleaned cycle sequencing products

The cleaned and denatured cycle sequencing products were run on the ABI3130 (Applied Biosystems) a four capillary Genetic analyzer. The Data Collection software 3.0 was used to run the sample, defining names and position of samples in plate, run modules and data storage files, while the Sequence Analysis software v5.3.1 with KB™ Basecaller v1.4 (Applied Biosystems) was used to analyse the raw sequence by defining baseline, and removing unclear sequences at the beginning and the end of the sequence.
2.8.6 Alignment of DNA Sequences

Alignment of the DNA sequences of the PCR products with the reference sequences ENSG00000075043 (for KCNQ2), ENST00000370263 (for CHRNA4) and ENSG00000213603 (for STMN3) was carried out using Chromas® (http://www.mb.mahidol.ac.th/pub/chromas/chromas.htm) with ClustalW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Following alignment, any changes from the reference sequences were annotated and reference was made to the NCBI SNP database.

2.8.7 Identification of the affected allele

DNA sequencing identified the superimposition of two sequences in the affected individual in the PCR product of CHRNA4. It was not possible to separate the two fragments on an agarose gel so a 8% polyacrylamide gel was used to determine if all the affected members had also two different alleles.

For the preparation of the polyacrylamide gel a 40% acrylamide solution was prepared (acrylamide: bisacrylamide, 40:1). The gel was prepared using reagents and volumes as indicated in Table 2.14. All the reagents (except TEMED) were mixed together in a 10ml beaker. TEMED (20ul) was then added and the contents of the beaker were mixed again. The mixture was poured in a gel tray and the gel was left to set for 30 minutes.
Table 2.14. Reagents and volumes used to prepare 50ml of 8% polyacrylamide gel.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volumes used to prepare a volume of 50ml</th>
</tr>
</thead>
<tbody>
<tr>
<td>40% acrylamide solution</td>
<td>10ml</td>
</tr>
<tr>
<td>50X TAE Buffer</td>
<td>1ml</td>
</tr>
<tr>
<td>10% ammonium persulphate solution</td>
<td>500μl</td>
</tr>
<tr>
<td>TEMED</td>
<td>20μl</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 50ml</td>
</tr>
</tbody>
</table>

Electrophoresis of the PCR products of Exon 5 (fragment A) of CHRNA4 of both affected and non-affected family members was carried out. The gel was stained using ethidium bromide. Affected family members and family members 2 and 6 had two bands (affected allele and normal allele) while non-affected family member had only one band (normal allele). DNA sequencing of the PCR products of the family members was performed.

2.9 Population study to determine the frequency of the affected allele

On a 8% polyacrylamide gel, it was possible to differentiate between the normal allele and ‘affected allele.’ PCR of Exon 5 (fragment A) was performed in a Maltese control population of 100 healthy individuals. DNA sequencing of those PCR products that were similar in size to the allele present only in the affected individuals. The frequency of the affected allele was found by counting the number of affected alleles found and divide this number by the total number of chromosome tested:

\[
\text{Affected allele frequency} = \frac{\text{Number of affected alleles}}{\text{Total number of alleles tested}}
\]
HapMap Genome Browser B36 (http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B36/) was used to study the degree of linkage disequilibrium between variations in the region of intron 4 and Exon 5 of CHRNA4.

2.10 Prediction of potential functional effects of identified variation

NetGene2 (http://www.cbs.dtu.dk/services/NetGene2/) (Brunak et al., 2001), SpliceView (http://bioinfo.itb.cnr.it/oriel/splice-view.html) (Rogozin and Milanesi, 1997) and Alternative Splicing Site Predictor (www.es.embnet.org/~mwang/assp.html) (Wang and Marin, 2006) were used to predict putative alternative acceptor or donor sites due to the variation in the ‘affected allele’. Mfold (http://mfold.bioinfo.rpi.edu/cgi-bin/dna-form1.cgi) (Zucker, 2003) and RNA2 (http://www.genebee.msu.su/services/rna2_reduced.html) (Brodsky et al., 1992) were used to predict the RNA structure of the pre-mRNA region harbouring the variant and compare it to reference sequence. RESCUE-ESE (http://genes.mit.edu/burgelab/rescue-ese/) (Fairbrother et al., 2002) and FAS-ESS (http://genes.mit.edu/fas-ess/) (Wang et al., 2004) were used to determine if synonymous SNPs could disrupt any exon splicing enhancer (ESE) or any exon splicing silencer (ESS).
2.11 Investigating alternative splicing

2.11.1 The pSPL3 plasmid vector

pSPL3 (Invitrogen®) is an exon-trapping vector and can be used for exon-skipping analysis (Stella et al., 2001). The plasmid has multiple cloning sites, an origin of replication and ampicillin resistance gene for growth in E.Coli and an SV40 segment that provides for transcription and replication in mammalian cells. The pSPL3 vector contains a minigene transcription occurs from the SV40 promoter and the RNA undergoes splicing under control of the host cell's RNA splicing machinery, resulting in fusion of the two vector exon sequences (Fig.2.7). The normal splicing pattern which is observed when only the vector exons are present, results in the fusion of exon 1 and 2 (Figure 2.5A). If a genomic DNA fragment cloned into pSPL3 contains an exon with functional splice donor (SD) and splice acceptor (SA) sequences, a different splicing pattern may be observed where this exon ends up between exon 1 and 2 (Figure 2.5B) The two splicing patterns can be distinguished at the cDNA level by using various vector-specific PCR primers and size-fractionation on gels can lead to recovery of the amplified exon from genomic DNA. This experiment was designed to determine if the sequence of the affected allele could result in a different splicing pattern of Exon 5 of CHRNA4.
Figure 2.5. Splicing in the pSPL3 vector. If there is no cloned insert or the cloned insert does not have a functional donor and acceptor site transcript A will result. If the cloned insert has a functional donor and acceptor site (shown in blue), the exon will be included in the transcript (B).

Figure 2.6. Plasmid pSPL3 showing restriction sites.
2.11.2 Designing restriction sites in the PCR product

The Oligo Explorer Software (GeneLink™) was used to design primers for CHRNA4 Exon 5 and about 400bp of intronic flanking regions. The PCR product sequence of the CHRNA4 Exon 5 gene together with about 400bp of intron region at both the 5' and 3' end of the exon (i.e flanking regions) of an affected member of the family (having the intronic mutation), an unaffected member (not having the mutation) and the reference sequence were analysed using software NEB Cutter V2.0 (http://tools.neb.com/NEBcutter2/index.php) to determine which restriction enzymes can be used (restriction enzymes that do not cut PCR product and which have cloning sites in the pSPL3 plasmid). *NotI* (Bioneer Corp.), *EcoRV* (Promega Corp.), *EcoRI* (Promega), *XhoI* (Bioneer Corp.) and *BamHI* (Bioneer Corp.) were found to satisfy both criteria. A combination of primer pairs (with different restriction sites) shown in Table 2.15 were used for the amplification. PCR was carried out using the reagent mix in Table 2.16 and the thermal profile in Table 2.17.
Table 2.15. Primers used for amplifying the fragment having 400bp of intron 4, Exon 5 and 400bp of intron 5. Every primer had a restriction site and tail at the 5' (Figure 2.7). All possible combinations of primers and restriction enzymes were tried.

<table>
<thead>
<tr>
<th>Primers with restriction sites (bold) and tails</th>
<th>Restriction sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>P11F 5'-tgaacttgccagccgcagcgctccagcctctggagaaag-3'</td>
<td>NotI</td>
</tr>
<tr>
<td>P11R 5'-tgaacttgccagccgcagcgctccagcctctggagaaag-3'</td>
<td>BamHI</td>
</tr>
</tbody>
</table>
| P1F 5'-tgacaaatgggacctctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctgtgctggactctg

Figure 2.7. Primer used for the amplification of the insert, showing primer sequence (light blue), restriction site (yellow) and tail (red).

Table 2.16. Reaction mix for the amplification of the insert using TLA DNA polymerase, a proof reading enzyme.

<table>
<thead>
<tr>
<th>Volume used per sample</th>
<th>Concentration used</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Buffer (Bioneer Corp.)</td>
<td>2.5μl</td>
</tr>
<tr>
<td>dNTPs (Bioneer Corp.)</td>
<td>2.5μl</td>
</tr>
<tr>
<td>Primer F (Bioneer Corp.)</td>
<td>0.25μl</td>
</tr>
<tr>
<td>Primer R (Bioneer Corp.)</td>
<td>0.25μl</td>
</tr>
<tr>
<td>TLA DNA Polymerase (Bioneer Corp.)</td>
<td>0.5μl</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 25μl</td>
</tr>
<tr>
<td>DNA</td>
<td>2μl</td>
</tr>
</tbody>
</table>
Table 2.17. Thermal profile for the amplification of the insert.

<table>
<thead>
<tr>
<th>Thermal Step</th>
<th>Temperature (°C)</th>
<th>Time (min)</th>
<th>Cycles</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>96</td>
<td>5</td>
<td>1</td>
</tr>
<tr>
<td>Denaturation</td>
<td>96</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>Annealing</td>
<td>58</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Extension</td>
<td>72</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Final extension</td>
<td>72</td>
<td>10</td>
<td>1</td>
</tr>
</tbody>
</table>

2.11.3 Restriction digests of PCR product and plasmid

The PCR products were digested by two restriction enzymes. Sequential digestion was carried out for all pairs of restriction enzymes. All reactions were carried out in 20µl using the restriction enzyme specific 10X buffer. The tubes were incubated at 37°C for 8 hrs (except for XhoI where incubation was at 50°C for 8hrs) followed by incubation at 65°C for 20 minutes for heat inactivation of the enzyme. The digested products were cleaned using genPure columns (Genetix Ltd, Cat No. K3506). Elution of the cleaned digested product was in 20µl of water. To every tube, the respective 15µl cleaned digested PCR products (or 10µl of pSPL3) were added. The tubes were incubated at 37°C for 8 hrs followed by incubation at 65°C for 20 minutes for heat inactivation of the enzyme. Lambda DNA was also digested as a control of the restriction digest reaction (for all enzymes except NotI). The digested products were run on a 2% agarose gel. The products were cut from the gel and purified using genPure columns (Section 2.8.2) (Genetix Ltd., Cat No. K3506).
2.12 Transformation of competent cells

2.12.1 Preparation of competent *E.Coli* cells

Bacteria are said to be competent when they can take up extracellular naked DNA from the environment through their cell wall. In the laboratory it is possible to prepare chemically competent or elecrocompetent cells that can be used for transformation (Samrook and Russell, 2001). The chemical method was used to prepare competent cells.

An aliquot of DH5α™ *E.Coli* (Clontech) cells was thawed on ice and 20μl of the aliquot were added to 10ml of LB broth (Luria-Bertani) prepared according to manufacturer’s instructions) in a 50ml sterile centrifuge tube. The tube was incubated overnight (~16hrs) at 37°C and 250rpm. After the incubation 2ml of the bacterial suspension were added to a flask containing 200ml of LB Broth. The flask was incubated at 37°C for 4 hrs. About 50 mls of the bacterial suspension was transferred to a 50ml sterile centrifuge tube and the tube was incubated on crushed ice for 10 minutes. Then the tube was centrifuged for 20 minutes at 3000g (using a refrigerated centrifuge) after which the supernatant was discarded and the tubes turned upside down on a tissue paper to dry the pellet from as much medium as possible. To the cell pellet remaining in the tube, 10ml of sterile 100mM MgCl₂ were added and the pellet was resuspended gently (by flicking the side of the tube) to avoid damaging the cells. The suspension was incubated on crushed ice for 10 minutes after which it was centrifuged at 3000g for 10 minutes. The supernatant was then discarded and 10ml of 100mM CaCl₂ were added to the cell pellet. Again the
cell pellet was resuspended gently. This was followed by incubation on crushed ice for 30 minutes and centrifugation at 3000g for 10 minutes. The supernatant was then discarded and 1ml of glycerol/100mM CaCl$_2$ solution (15%v/v) was added to the dry cell pellet. Gentle resuspension of cell pellet was carried out and the suspension was aliquoted in 1.5ml tubes (100μl of suspension in each tube) and placed in the freezer at -20°C (Samrook and Russell, 2001).

2.12.2 Ligation reaction of digested insert and digested pSPL3

The T4 DNA Ligase (Promega Corp.) catalyzes the joining of two strands of DNA between the 5'-phosphate and the 3'-hydroxyl groups of adjacent nucleotides in either a cohesive-ended or blunt ended configuration. The ligation was carried out according to the manufacturer’s protocol.

The concentration of the digested plasmid and digested insert was estimated by gel electrophoresis. Ratios of insert: plasmid of 8:1, 5:1, 3:1, 1:1, 1:3 and 1:6 were used. The ligation reaction was carried using the volumes outlined in Table 2.18.

<table>
<thead>
<tr>
<th></th>
<th>X 1 Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>10X Ligation buffer (Promega Corp.)</td>
<td>1μl</td>
</tr>
<tr>
<td>T4 Ligase(Promega Corp.)</td>
<td>1μl</td>
</tr>
<tr>
<td>Plasmid (pSPL3)</td>
<td>1μl</td>
</tr>
<tr>
<td>Digested PCR product</td>
<td>2μl</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 10μl</td>
</tr>
</tbody>
</table>

The above mix was prepared (without the digested PCR product) and was then dispensed in 200μl tubes (8μl per tube). To the labelled tubes the respective digested
PCR product was added and the tubes were vortexed gently and centrifuged for a few seconds. The tubes were incubated overnight at 2-8°C.

2.12.3 Transformation reaction

The required aliquots of competent bacteria were removed from the -20°C storage and placed in crushed ice to thaw. The ligation reactions were taken from the 2-8°C fridge and 2µl of every reaction were placed in the bottom of a labelled 1.5ml tube respectively. To each tube, 25µl of competent cells were added (after thawing the tube containing the cell suspension was flicked to ensure a uniform suspension) and the mix was vortexed lightly. The tubes were incubated on crushed ice for 20 minutes after which they were placed at 42°C for 45 seconds and then returned on ice for 2 minutes. To every tube, 975µl of LB broth (BD Difco™ Cat no. 244620 - prepared according to the manufacturer’s instructions) were added and then every mix was transferred to respectively labelled 15ml sterile centrifuge tube. The centrifuge tubes were incubated in a shaking incubator at 37°C for 90 minutes at a shaking speed of 150rpm. After incubation, 500µl of the suspension in the tubes was plated on the respective labelled plates as indicated below.
Table 2.19. Samples and controls used in transformation.

<table>
<thead>
<tr>
<th>Sample Name</th>
<th>Sample Description</th>
<th>Media used for plating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control A</td>
<td>Unrestricted pSPL3</td>
<td>LB amp agar and Standard Method Agar</td>
</tr>
<tr>
<td>Control B</td>
<td>Plasmid restricted with one restriction enzyme only</td>
<td>LB amp agar and Standard Method Agar</td>
</tr>
<tr>
<td>Control C</td>
<td>Plasmid restricted with two restriction enzyme</td>
<td>LB amp agar and Standard Method Agar</td>
</tr>
<tr>
<td>Sample 1</td>
<td>Ligation reaction of affected individual 2nd generation having mutation</td>
<td>LB amp agar</td>
</tr>
<tr>
<td>Sample 2</td>
<td>Ligation reaction of affected individual 3rd generation having mutation</td>
<td>LB amp agar</td>
</tr>
<tr>
<td>Sample 3</td>
<td>Ligation reaction of non-affected individual 2nd generation (no mutation)</td>
<td>LB amp agar</td>
</tr>
<tr>
<td>Negative control</td>
<td>Plates with no bacteria added to them</td>
<td>LB amp agar and Standard Method Agar</td>
</tr>
</tbody>
</table>

LB agar (Conda Prorandiso™ Cat no. 1552.00) and Standard Method Agar (BD BBL™ Cat no. 211638) were prepared according to manufacturer’s instructions. When the LB agar had cooled to 50°C, ampicillin to a final concentration of 100μg/ml was added (LB amp agar).

The plates were incubated for 24hrs at 37°C. Controls were satisfactory indicating that the enzymes were added in enough quantity and incubated for a long enough time to cut the plasmid and the PCR product.

From every sample plate (LB plate), about 4 distinct colonies were picked. Using a sterile loop the colony was picked up and placed into 10ml of LB broth (BD Difco™ Cat no. 244620) to which 10μl of 1%w/v ampicillin solution were added. The broths
were incubated at 37°C for about 6 hours at 150rpm until turbidity in the broth (indicating growth) was observed.

2.12.4 Extraction of plasmid DNA from bacterial cell cultures using

AccuPrep® plasmid Mini Extraction Kit was used for the extraction of plasmid DNA from bacterial cells. The AccuPrep® plasmid Mini Extraction Kit was developed for the extraction of highly purified plasmid DNA from cultured bacterial cells. The overall principle is based on alkaline lysis method (Birnboim et al., 1979) the procedure was carried out according to the manufacturer’s protocol.

About 10 ml of bacterial suspension was transferred to a 15ml sterile centrifuge tube. The tubes were centrifuged at a 3,000g for 5 minutes. The supernatant was discarded and the tubes were placed upside down on a tissue to remove any remaining supernatant. The cell pellet was resuspended in 250μl of Buffer 1 (to which RNAse A Powder have been added) by pipetting. To the cell suspension, 250μl of Buffer 2 were added and the suspension was mixed by inverting the tube 2-4 times. The suspension was incubated at room temperature for 5 minutes until it turned clear. To the clear solution, 350μl of Buffer 3 were added and the solutions were mixed by inverting the tube 3-4 times. The tube was centrifuged at 16,600g for 10 minutes. The clear lysate was transferred to the DNA Binding column tube and this was centrifuged at 16,600g for 1 minute. The flow-trough was discarded and the collection tube was replaced under the DNA Binding column to which 700μl of Buffer 4 (to which absolute ethanol has been added) were added. The assembly was centrifuged at 16,600g for 1 minute and the flow-trough was discarded and the
collection tube was replaced under the DNA Binding column. The assembly was centrifuged at 16,600g for 1 minute to remove residual ethanol. The DNA Binding column was transferred to a labelled 1.5ml tube and 50µl of Buffer 5 (heated to 60°C) were added. The assembly was incubated at room temperature for about 5 minutes and then centrifuged at 16,600g for 1 minute.

2.12.5 Determination of the presence of insert in pSPL3

The plasmid pSPL3 has a size of about 6031bp while the insert is 2300bp (depending on which primer pair was used). Thus the plasmid with insert should have been about 8300bp. To determine the presence or absence of an insert, the cleaned plasmid DNA from clones (restricted with one enzyme) and restricted pSPL3 were run on a 0.4% agarose gel for 20 minutes. The presence of a band about 2300bp longer than the restricted plasmid, would have indicated the presence of the insertion in the plasmid.

2.13 Sequencing of the mRNA transcript of the CHRNA4

2.13.1 RNA extraction from leukocytes

The purity and integrity of RNA isolated from tissue or cultured cells are critical for its effective use in applications such as reverse transcription PCR (RT-PCR). RNA extraction from leukocytes was carried out using using SV Total RNA Isolation System (Promega Corp.). The SV Total RNA Isolation System provides a fast and simple technique for preparing purified and intact total RNA from tissues, cultured
cells and white blood cells. For the RNA extraction the manufacturer’s protocol was followed.

About 1ml of freshly collected EDTA anticoagulated blood was placed in a 2ml tube. The tube was centrifuged for 5 minutes at 4000g. The supernatant was removed and Red Blood Cell Lysis Solution was added. The cell pellet was resuspended by pipetting 4-5 times. The tube was centrifuged at 3000rpm for 5 minutes. The supernatant was removed and Red Blood Cell Lysis Solution was added. The cell pellet was resuspended by pipetting 4-5 times. The tube was centrifuged at 3000rpm for 5 minutes. The supernatant was removed and Red Blood Cell Lysis Solution was added. The cell pellet was resuspended by pipetting 4-5 times. The tube was centrifuged at 3000rpm for 5 minutes. All supernatant except 100μl were removed and 175ml of RNA Lysis Buffer (containing B mercaptoethanol) were added to the cell pellet. The cell pellet was resuspended and 350μl of RNA Dilution Buffer were added and the tubes were inverted 3-4 times to mix the contents. The tubes were incubated at 70°C for 3 minutes in a heating block. The tubes were centrifuged at 12,000g for 10 minutes. The clear lysate was transferred to the respectively labelled 1.5ml tube. A volume of 200μl of 95% ethanol were added to the lysate and mixed by pipetting 3-4 times. This mixture was transferred to a spin column assembly and the assembly was centrifuged at 12,000g for 1 minute. The flow-through in the collection tube was discarded and the collection tube was replaced under the spin column. To the spin column, 600μl of RNA Wash Solution were added and the assembly was centrifuged for 1 minute at 12,000g. The flow-through in the collection tube was discarded and the collection tube was replaced under the spin column. A mixture of 40μl Yellow Core Buffer, 5μl of 0.09M MnCl₂ and 5μl of
DNaseI enzyme was prepared for each tube. The solution was mixed by pipetting. To every spin column, 50µl of this mix was added directly on the filter in the column. The spin columns were incubated for 15 minutes at room temperature. After this incubation, 200µl of DNase Stop Solution were added to the spin column and the assembly was centrifuged at 12,000g for 1 minute. 600µl of RNA Wash Solution were added to the spin column and the assembly was centrifuged at 12,000g for 1 minute. The flow-through in the collection tube was discarded and the collection tube was replaced under the spin column. 250µl of RNA Wash Solution were added to the spin column and the assembly was centrifuged at 12,000g for 2 minute. The spin column was removed from the collection tube and placed into an RNase free 1.5ml Elution tube. About 100µl of Nuclease Free water was added to the spin column and the assembly was centrifuged at 12,000g for 1 minute. The spin column was discarded and the elution tube was stored at -80°C.

2.13.2 RNA Transcription

For the reverse transcription, Reverse Transcriptase ImPromII System (Promega Corp.) was used. The Reverse Transcriptase ImPromII System (Promega Corp.) is a convenient kit that includes a reverse transcriptase and an optimized set of reagents designed for efficient synthesis of first-strand cDNA. For the synthesis of cDNA, the kit was used according to the manufacturer's instructions.

Random hexamers were used for the synthesis of cDNA. The RNA was taken out from the -80°C storage and placed on crushed ice. Five microlitres of RNA were transferred to a pre-labelled 200ul tube and 2ul of 100pmol/ul random hexamers
were added and mixed by pipetting up and down. The tubes were placed in a thermal cycler at 70°C for 5 minutes and then replaced on ice. The mix shown in Table 2.20 was prepared and added to the tube.

Table 2.20. Mix for reverse transcription.

<table>
<thead>
<tr>
<th></th>
<th>X 1 Sample</th>
</tr>
</thead>
<tbody>
<tr>
<td>5X Reaction Buffer (Promega)</td>
<td>4μl</td>
</tr>
<tr>
<td>ImPromII (Promega)</td>
<td>1μl</td>
</tr>
<tr>
<td>Magnesium Chloride (25mM)</td>
<td>3.75μl</td>
</tr>
<tr>
<td>dNTPs (2.5mM of each dNPT/μl)</td>
<td>3μl</td>
</tr>
<tr>
<td>Water</td>
<td>Up to 13μl</td>
</tr>
</tbody>
</table>

The 200μl tubes were placed in the thermal cycler and submitted to the thermal profile in Table 2.21.

Table 2.21. Thermal profile used for reverse transcription.

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Annealing</td>
<td>25°C</td>
<td>5 minute</td>
</tr>
<tr>
<td>Extension</td>
<td>42°C</td>
<td>60 minutes</td>
</tr>
<tr>
<td>Enzyme Deactivation</td>
<td>70°C</td>
<td>15 minutes</td>
</tr>
</tbody>
</table>

2.13.3 cDNA amplification of transcript of CHRNA4 of patient and control

The PCR reaction was performed using the following primer pair, reaction mix and thermal profile in tables. The forward primer is in Exon 4 of the CHRNA4 gene while the reverse primer is in Exon 5 and they amplify a region of 568bp.
Primer pairs

T3-F 5' -CAGGAGTGGCAGACTACAAG-3'
T3-R 5' -AGATGCACAGCTATCTCTC-3'

caggagttg cagcctgacg aacctgagct gactatgaga atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc

cgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatct

ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac

ggggaccaag ccagagatcg aacctgagct gactatgaga atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctc
ttgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtcaccc tctttcccct ccgaggatcg caggacccac gccgggtgca ctgctgctca tctcctgcct caccgtgctg gtcttctacc tgtgcatctgacctgctgactcctcctggtgaac aagtgcacagcc gcggagtgggt cagcctgacg atgtcacttc catcgcgtaccc cccccgcagct cgcctccgacg tcatctggcg gcgggcatc ccaagctgccct cagaagctgcc catcgtggat ggcgggtgca ctgcgtcattctt gccgggtgcctg ccctccgagc tcatctggcg gccggacatc gtcctctaca acctgaccaa ggcccacctg ttccatgagc gcgggtgca ctggactccc cccggccttt acaagagctcc tgcacgatct gacgtc...
2.14 Copy Number Variation (CNV) analysis

The SNP genotyping data was also used for CNV analysis. Genotyping Console™ software package (Affymetrix®) was used for the analysis. SNP genotyping data was read using GCOS (Affymetrix®). The SNP data was then transferred to Genotyping Console™ software package (Affymetrix®) together with Sample Attributes files. Quality control of data was performed to determine which data can be used for the analysis.

Two analysis were made:-

1. To determine if affected individuals have any losses or gains that differ from those of the non-affected.

2. To determine if the two non-affected individuals who also have the disease haplotype have any losses or gains that differ from that in the affected individuals.

2.15 Conclusion to methodology

A three generation family having seven members affected by epilepsy or febrile seizures was identified as being suitable for the study. DNA samples were collected from both the affected and non-affected family members from blood or buccal rinse samples. The DNA was purified and a whole genome scan using GeneChip® Mapping 250K (NspI) Assay (Affymetrix®) was performed.
Quality control of genotyping data was performed using various software packages and Allegro v2.0 was used to perform the linkage analysis. STR markers were used to confirm the linkage interval. Genehunter gh2.1, MLINK and Cyrillic v2.1 were used for the linkage analysis.

PCR fragments of the exons (plus intro/exon junctions) of three candidate genes in the chromosomal interval, CHRNA4, KCNQ2 and STMN3 were sequenced. The DNA sequence obtained was aligned with the reference sequence. A deletion was found in intron 4 of CHRNA4. A population study was then carried out in order to determine frequency of the variation in the population. Bioinformatic tools were used in order to try to predict if the variation observed could have a function that could effect gene function.

A minigene plasmid vector experiment was attempted in order to try to determine if the intronic variation affected splicing, but it was not successful since the insert never entered the plasmid. Another experiment was performed to determine whether the intronic variation results in alternative splicing. The cDNA transcript from RNA extracted from leukocytes was sequenced across the Exon 4/Exon 5 region.

CNV analysis of the SNP genotyping data was performed using the Genotyping Console™ software package.
Chapter 3

Results
3.1 Introduction to results

In this chapter a description will be given of how the characterization of the syndrome in this family has been carried out using the relevant family history, medical history and literature search. SNP genotyping data of 12 family members was carried using the GeneChip® Human Mapping 250K Nsp Array. In order to generate genotypes from the GeneChip® scans, software supplied by Affymetrix® were used. Quality control of the genotyped data was carried out using various software packages while multipoint linkage analysis of SNP genotyping was carried using Allegro v2.0. By varying the disease status of family members in the ‘pedfile’ (pedigree file used for linkage analysis) and the penetrance, a number of loci with suggestive evidence for linkage were identified. The results as detailed later on indicate the maximum parametric LOD was 2.67 on chromosome 20 when disease inheritance pattern was taken as autosomal dominant with incomplete penetrance. These results are novel in that GEFS+ or familial febrile seizures have never been linked to chromosome 20 but this locus 20q13.32-33 had already been linked to other epilepsies (Table 4.1). Two other loci on chromosome 3 and 13 were also identified with a lower parametric LOD but still giving suggestive evidence for linkage.

This chapter will describe how the chromosomal interval on Chromosome 20 was confirmed using STRs. The best candidate genes in this region were CHRNA4, KCNQ2 and STMN3 (Section 3.5). Sequencing analysis of the exonic regions did not identify any changes in exonic sequence besides synonymous SNPs. In intron 4 of CHRNA4 a haplotype of variants segregated with the disease phenotype. A study
for the frequency of this haplotype in the Maltese population was carried out and its frequency was found to be 0.04.

This chapter also describes the bioinformatic-based analysis of the affected allele. In addition, this analysis also determined that the affected allele has a different pre-mRNA folding than that of the reference sequence (Section 3.7.3). The affected allele also has a possible alternative donor splice site with a higher score than the reference sequence (Section 3.7.4). Analysis of the cDNA sequence of the CHRNA4 transcript did not identify any variation in the affected individual when compared with the reference sequence.

Copy Number variation analysis was also performed in order to determine if there were any differences in CNV between affected and non-affected individuals (who do not have the disease associated haplotype) and between affected individuals and non-affected individuals who do have the disease associated haplotype (family members 2 and 6 - see Figure 3.1). CNV analysis identified CNV difference between family members 2 and 6 and the affected individuals (Figure 3.1) (Section 3.10).

3.2 Syndrome characterization in the family

A three generational Maltese family was recruited for the study following their informed consent. Their pedigree is described in Figure 3.1. Analysis of the family history both from information given by the family and also through hospital medical history files and attending neurologist, indicated that several individuals in the family had had febrile seizures, while in the third generation a seizure phenotype
was associated with learning difficulties and/or behavioural problems in all the affected individuals. Clinical data in their history files was very limited and no CT scans, MRIs or EEGs were available except for member 14 whose EEG had shown that epileptic discharges started in the temporal lobe. In this family five members had had febrile seizures in the first years of life. Of the other two affected individuals, one developed focal onset epilepsy at age eight years (individual 14) and the other (individual 8) developed febrile seizures at age six years. Individual 5 never suffered from any seizures through his life.

Individual 9 had had febrile seizures until about four years of age. She had her first fit when she was about one and a half years old. Individual 7 had had from febrile seizures till about four years of age. He had had his first fit when he was about nine months old. Individual 10 had had febrile seizures in the first years of life. Relatives were quoted as saying that his mental retardation was due to the severe seizure he had had when he was about one year nine months old. He was diagnosed as having moderate mental retardation (IQ68). He had developed behavioural problems at about nine years of age. Since eleven years old, he experiences sudden jerks while at times he rolls up his eyes and makes guttural sounds and is aggressive when scolded.

Individual 14 never had any febrile seizures. At age eight he had the first seizure and also started having exhibiting behavioural problems when scolded. Now both well controlled with sodium valporate. He has learning difficulties at school and needs a facilitator and also could be suffering from depression. The EEG showed that during a seizure, frequent short runs of epileptiform discharges are present over the temporo-occipital area which then spread to the right parietal and to the left temporo-
occipital areas. He was diagnosed as having focal onset seizure disorder. An MRI showed no cerebral pathology.

Individual 8 suffered her first febrile seizure when six years old while she had an upper respiratory tract infection. She had never had any other seizure but since puberty she suffers from panic attacks whenever someone scolds her. She is aware that her reaction to these situations is not normal but she has no control over this behaviour. Individual 18 suffered from febrile seizures until about four years of age. She had her first fit when she was about one and a half years old. Individual 17 suffered from febrile seizures until about four years of age. She had her first fit when she was about one and a half years old. At age two and a half she fell and hit her head and suffered an afebrile seizure. She has learning difficulties at school and needs a facilitator. The febrile seizures experience by the affected family members were always tonic clonic seizures. Individuals 11, 15, 2 and 6 have never had any seizures. The family was never recruited in any genetic study.

![Figure 3.1. Pedigree of the family studied. Family members who have never had any seizures but whose children are affected are indicated by a question mark (?). Affected individuals (family members that have a seizure phenotype) are shown in black. Individual 14 who suffered from focal onset seizures is highlighted with a border. Individuals shown in grey are family members who were unavailable to participate in the study.]
Table 3.1. Disease status of family with age of onset and current health status of family members.

<table>
<thead>
<tr>
<th>Family member</th>
<th>Affected/non-affected</th>
<th>Age of onset</th>
<th>Current health status</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>5</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>6</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>7</td>
<td>Febrile seizures</td>
<td>9 months</td>
<td>Cessation of febrile seizures at 4 years of age</td>
</tr>
<tr>
<td>8</td>
<td>Febrile seizures</td>
<td>6 years</td>
<td>Since puberty she has frequent panic attacks</td>
</tr>
<tr>
<td>9</td>
<td>Febrile seizures</td>
<td>18 months</td>
<td>Cessation of febrile seizures at 4 years of age</td>
</tr>
<tr>
<td>10</td>
<td>Febrile seizures</td>
<td>1 year of life</td>
<td>Cessation of febrile seizures at 4 years of age. Experiences jerks and uprolling of eyes since 11 years of age</td>
</tr>
<tr>
<td>11</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>14</td>
<td>Focal onset seizure</td>
<td>8 years</td>
<td>Controlled with sodium valproate</td>
</tr>
<tr>
<td>15</td>
<td>Non-affected</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>17</td>
<td>Febrile seizures</td>
<td>18 months</td>
<td>Cessation of febrile seizures at 4 years of age</td>
</tr>
<tr>
<td>18</td>
<td>Febrile seizures</td>
<td>18 months</td>
<td>Cessation of febrile seizures at 4 years of age</td>
</tr>
</tbody>
</table>
If only family members who had had febrile seizures in their first years of life (individuals 9, 10, 7, 18 and 17) are considered as affected, the inheritance pattern of the disease appears to be autosomal dominant. On the other hand if the phenotype in this family is taken as that of having had a witnessed seizure, family members 8 and 14 are also considered as affected and since their mothers have not had any seizures, the pattern of inheritance appears to be autosomal dominant with incomplete penetrance.

The affected family members have had febrile seizures but only one member had had a focal onset seizure although the EEG showed that the discharge starts from temporal and occipital lobe but moves on to other regions of the brain. This indicates that the syndrome in this family can be listed as either *Generalized epilepsy with febrile seizure plus* (GEFS+) or (due to the fact that no affected family member suffers from actual generalized seizures) *epilepsy with febrile seizure plus* (Ito, et al., 2002; Hiroto and Sunao, 2003). The description of GEFS+ syndrome is still being defined by the ILAE Classification Commission and although the abbreviation for this ‘syndrome’ is still GEFS+, ILAE is now referring to the syndrome as Epilepsy with febrile seizure plus (Engel, 2006). Based on the review of a number of published papers and with the gathered clinical history, the family syndrome fits with GEFS+ because in the family the condition appears to be idiopathic and febrile seizures are present together with other epileptic seizures (Baulat et al., 1999; Audenaert et al., 2005). The family phenotype is in accordance with the phenotype of GEFS+ (Scheffer and Berkovic, 1997; Singh et al. 1999). The presence of panic disorder, behavioural problems and intellectual impairment supports GEFS+ since families with GEFS+ have also been found to have these conditions (Baulac et al,
1999; Osaka et al., 2007). The presence of febrile seizures without epilepsy in a number of family members could also indicate a family with a phenotype of febrile seizures.

3.3 Analysis of the genome-wide SNP data

SNP genotyping was carried out on 12 family members. The SNP chip data was read using the Data Transfer Tool (DTT) (Affymetrix®) and Analysis of the 250K Nsp GeneChip® of all family members genotyped was performed using GeneChip® Operating Software (GCOS) (Affymetrix®) as described in Section 2.5.2. For every sample, a *.DAT file and a .CEL were generated. A .DAT file is the actual scan of the 250K Nsp GeneChip® while the *.CEL file stores the results of the intensity calculations on the pixel values of the DAT file.

Figure 3.2. A *.DAT file which shows the actual scan of the chip of Sample 8. Analysis of the 250K Nsp GeneChip® of Sample 8 using GeneChip® Operating Software (GCOS) (Affymetrix®) generates a *.DAT file.
Figure 3.3. A *.CEL file which shows the actual scan of the chip of Sample 8. Analysis of the 250K Nsp GeneChip® of Sample 8 using GeneChip® Operating Software (GCOS) (Affymetrix®) generates a *.CEL file.

3.3.1 SNP genotyping

Genotyping was performed using GeneChip® Genotyping Analysis Software (GTYPE) (Affymetrix®). Genotypes were generated for the 12 samples from the GeneChip® Human Mapping 250K Nsp Array (Section 2.5.2). Partial genotyping data of the 262,000 SNPs for some of the samples is shown in Figure 3.4. The software performs quality control of the Allele Call and gives a confidence level for the genotype of every SNP. The genotyping data (in text format) was transferred to another software package, Alohomora (Ruschendorf and Nurnberg, 2005) such that further quality control of data could be performed.
Figure 3.4. Graphical representation of 'Allele Call' and confidence level as given by GTYPE software. The table in the picture shows the SNP ID, the NCBI dbSNP RS ID, the chromosome and physical position of SNP. For every SNP for the family members there is the allele call together with the confidence level for the call. The triangle at the top of the picture represents confidence levels. The further to the corners of the triangle, the highest the level of confidence in the allele call.
3.3.2 Quality control of SNP genotypes

Graphical relationship representation (GRR) was used to check the correct relationship between individuals (Section 2.5.2) (http://www.sph.umich.edu/csg/abecasis/index.html). This software is very important in confirming accurate sample labelling and pedigree information since misspecification of relationship of DNA samples could be a problem in genetic studies leading to inaccurate results (Abecasis et al., 2001). Figure 3.5 shows the analysis of the genotyping data. Single coloured clusters indicate that there was no relationship misspecification.

![Figure 3.5. GRR of familial relationship between samples. The yellow cluster indicates a characteristic allele sharing between parent and offspring where it is expected that parent and offspring share at least one allele. The clustering of all the yellow squares indicate that the expected allele sharing between parent and offspring could be confirmed for the parents and offspring in the family, indicating that no misspecification of data labelling. The same applies for the other relatives (green) and sib-pairs (red).](image-url)
PedCheck (http://watson.hgen.pitt.edu/register/) was used to identify SNPs with Mendelian inconsistencies in the genotypes of the family members (Connell and Weeks, 1998). Analysis of the genotyping data using PedCheck identified 16842 SNPs with Mendelian inconsistencies. Merlin (http://www.sph.umich.edu/csg/abecasis/index.html) was used to identify unlikely genotypes. Out of a total of 262,000 SNPs, 72654 were non informative, 16842 had Mendelian inconsistencies and 853 had unlikely genotypes. In all 172768 remaining markers were used for the linkage analysis.

3.4 Linkage analysis of SNP genotyping data

3.4.1 Phenotype A – Family members who have had a witnessed seizure were considered as affected

Multipoint linkage analysis was performed using Allegro (Gudbjartsson, 2005). In this family the inheritance pattern appeared to be autosomal dominant with incomplete penetrance when individuals who had had a witnessed seizure where considered as affected. Various analysis were carried out varying both penetrance and disease status of the mothers (individual 2 and 6 - Figure 3.1) of individuals 8 and 14 (phenocopy 0.01, deCode map and Caucasian SNP allele frequency). Highest parametric LOD, non-parametric Zlr scores and NPL were observed when the disease status of individual 2 and 6 was taken as ‘unknown’ taking in consideration that in GEFS+ there are cases where family members develop seizures later in life even if they do not have a history of febrile seizures (Baulac et al., 1999; Gourfinkel-An, 2004). The parametric LOD scores, Zlr scores and NPL scores obtained for all chromosomes are shown in Figure 3.6, 3.7 and 3.8.
Figure 3.6. Parametric LOD scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%. The top X axis shows the chromosomes while the bottom X axis shows genetic distance (cM). The Y axis is the parametric LOD score.
Figure 3.7. Zlr scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%. The top X axis shows the chromosomes while the bottom X axis shows genetic distance (cM). The Y axis is the parametric LOD score.
Figure 3.8. NPL scores across the genome following linkage analysis when phenotype of affected individuals was taken as that of having suffered a witnessed seizure, when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%. The top X axis shows the chromosomes while the bottom X axis shows genetic distance (cM). The Y axis is the NPL score.
When the disease status of individuals 2 and 6 was taken as non-affected, the highest parametric LOD score was obtained when penetrance was 90% (or 0.9). Fig 3.9 shows how the parametric LOD score changed when penetrance was varied from 1 to 0.5 (100% to 50%).

The highest parametric LOD 2.6747 (p<0.001), NPL of 8.4940 and Zlr of 3.3279 were observed for a number of chromosomes (Figure 3.6, 3.7 and 3.8). A parametric LOD score of 2.6747 is suggestive evidence of linkage. While the non-parametric scores (NPL and Zlr) have values that indicate significant linkage between the chromosomal region and the disease. Table 3.2 is a list of all the chromosomal regions where a parametric LOD of 2.6747 was observed.
Table 3.2. Chromosomal regions where the highest parametric LOD score was observed.

<table>
<thead>
<tr>
<th>Chromosomal region</th>
<th>SNP ID as given by Affymetrix</th>
<th>NCBI db SNP</th>
<th>deCode genetic position (cM)</th>
<th>Physical position</th>
</tr>
</thead>
<tbody>
<tr>
<td>7p15.3-15.2</td>
<td>SNP_A-2137248</td>
<td>rs4722286</td>
<td>39.87646354</td>
<td>23686,864</td>
</tr>
<tr>
<td></td>
<td>SNP_A-4226100</td>
<td>rs2813838</td>
<td>40.22425220</td>
<td>23922,532</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1787476</td>
<td>rs2158342</td>
<td>40.36704643</td>
<td>24019,292</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1842049</td>
<td>rs17149360</td>
<td>40.60438502</td>
<td>24180,117</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1989225</td>
<td>rs1476520</td>
<td>40.81867599</td>
<td>24569,469</td>
</tr>
<tr>
<td></td>
<td>SNP_A-4236596</td>
<td>rs10247954</td>
<td>41.12158772</td>
<td>25017,216</td>
</tr>
<tr>
<td>8q21.2</td>
<td>SNP_A-1950957</td>
<td>rs1529987</td>
<td>93.89291146</td>
<td>85151,359</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2056107</td>
<td>rs4740024</td>
<td>94.08221654</td>
<td>85614,478</td>
</tr>
<tr>
<td>8q22.1</td>
<td>SNP_A-2294853</td>
<td>rs4236836</td>
<td>101.36883055</td>
<td>96795,074</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1994334</td>
<td>rs496228</td>
<td>101.45726511</td>
<td>96918,711</td>
</tr>
<tr>
<td>8q22.2-22.3</td>
<td>SNP_A-1819121</td>
<td>rs2132349</td>
<td>105.17327046</td>
<td>101249,424</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2282103</td>
<td>rs17428846</td>
<td>105.73867864</td>
<td>101623,657</td>
</tr>
<tr>
<td>12p13.2</td>
<td>SNP_A-1836705</td>
<td>rs11053689</td>
<td>24.85533587</td>
<td>10260,274</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2175576</td>
<td>rs9739469</td>
<td>25.12397761</td>
<td>10629,565</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2263992</td>
<td>rs4763223</td>
<td>25.35223772</td>
<td>10943,345</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2032709</td>
<td>rs1816223</td>
<td>25.67373036</td>
<td>11385,288</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2118518</td>
<td>rs10743973</td>
<td>27.55014120</td>
<td>12026,051</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2313651</td>
<td>rs11054874</td>
<td>28.19352973</td>
<td>12434,500</td>
</tr>
<tr>
<td>13q26.3</td>
<td>SNP_A-2083555</td>
<td>rs2602036</td>
<td>127.70965031</td>
<td>97577,754</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2298375</td>
<td>rs8032018</td>
<td>128.27140400</td>
<td>97800,698</td>
</tr>
<tr>
<td>20q13.32-33</td>
<td>SNP_A-2240422</td>
<td>rs157093</td>
<td>94.03351563</td>
<td>55686,362</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2265391</td>
<td>rs6015139</td>
<td>94.73688179</td>
<td>55941,552</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1835033</td>
<td>rs841729</td>
<td>95.50550756</td>
<td>56220,419</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1936767</td>
<td>rs6015552</td>
<td>100.10203055</td>
<td>57706,242</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2174716</td>
<td>rs242821</td>
<td>101.30899824</td>
<td>58087,312</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1956938</td>
<td>rs1557184</td>
<td>106.97194355</td>
<td>59875,246</td>
</tr>
<tr>
<td></td>
<td>SNP_A-2131976</td>
<td>rs6089556</td>
<td>107.16907073</td>
<td>59937,484</td>
</tr>
<tr>
<td></td>
<td>SNP_A-1866638</td>
<td>rs6011286</td>
<td>114.28507122</td>
<td>62184,184</td>
</tr>
</tbody>
</table>
On chromosomes 7, 8, 12 and 15 the chromosomal region showing linkage was always less than 500kb and the parametric LOD observed could be due to random association (linkage disequilibrium) in the family, although scientifically no region should be excluded until it was analysed, since these regions are far larger than the average of 10-30kb of LD observed throughout the genome in the general population (Clarke, 2003; Moulard et al., 1999). Analysis of the HapMap for these regions identified a minimum of 3 main haplotype blocks for every chromosomal region.

![Image of HapMap of chromosome 7 between SNP_A-2137248 (rs4722286) and SNP_A-4226100 (rs2813838) (a region of 235kB) were parametric LOD was 2.67. CEU refers to the HapMap of a Population of European ancestry. To determine the degree of linkage disequilibrium between these two SNPs an imaginary line is drawn, from each of the SNPs, parallel to the triangle. The colour of map where the two lines intersect indicates the degree of linkage disequilibrium between the markers where red indicates strong LD and white no LD.](image)

On chromosome 13, (13q13-14) a NPL score of 8.4940 and Zlr of 3.3279 were observed in a region of about 16cM (Figure 3.11) Maximum parametric LOD score of 2.0352 was observed between markers SNP_A-1828416 (rs2843389) and SNP_A-2276734 (rs2246478) when penetrance was lowered to 0.7 (Figure 3.12 and 3.14). When penetrance was lowered to 0.5, the maximum LOD score decreased to
2.0132 but the LOD in the neighbouring region increased between SNP_A-1896289 (rs17081386) (physical position 33,985,131) and SNP_A-1799678 (rs6561300) (physical position 45,637,626) (Figure 3.13). The haplotype of the SNPs in this area (where parametric LOD peak was observed) indicated recombination in two non-affected individuals (Figure 3.15) (Greenberg and Berger, 1994).

![Figure 3.11. Variation of parametric LOD with variation in penetrance. The maximum multipoint parametric LOD was obtained when penetrance was 0.7.](image)
Figure 3.12. NPL scores on Chromosome 13 when penetrance was lowered to 0.7. In the region between 35cM and 50cM the NPL score is above 5 and this can be considered as significant evidence for linkage.
Figure 3.13. Parametric LOD scores on Chromosome 13 when penetrance was lowered to 0.7. A parametric LOD score above 3 is significant evidence for linkage while the value above 2 observed is suggestive evidence for linkage.
Figure 3.14. Parametric LOD scores on Chromosome 13 when penetrance was lowered to 0.5. A parametric LOD score above 3 is significant evidence for linkage while the value above 2 observed is suggestive evidence for linkage.
Figure 3.15. Chromosome 13 haplotype in chromosomal region where parametric LOD peaked. Recombination can be observed in individuals 11 and 15. In this region all affected individuals and family members 2 and 6 have the same ‘light blue’ chromosomal region inherited from the mother. In individuals 11 and 15, there is crossing over between chromosomes inherited from the mother (two coloured bar).
3.4.2 Phenotype B - Family members who had had febrile seizures before the age of six were considered as affected

Another linkage analysis was carried out where family members were considered as being affected if they had had febrile seizures before the age of six. The only family members that could be considered as affected were individuals 9, 7, 10, 18 and 17 as indicated in Figure 3.16. The mode of inheritance was taken as autosomal dominant.

![Pedigree of the family studied. Affected individuals (family members that had febrile seizures before age 6) are shown in black.](image)

The highest parametric LOD of 2.0983 (NPL 1.7635, ZLR 2.1928) was observed on a number of chromosomes but the chromosomal regions were always small (less than 500kb). A peak parametric LOD of 1.71 (NPL 1.3225, ZLR 1.9335) was observed on chromosome 3 in a region of about 5cM between markers SNP_A-2290852 (rs7612097) and SNP_A-2030317 (rs17277717).
Figure 3.17. Parametric LOD scores on Chromosome 3. Blue arrow indicates a region of about 5cM (3q21.3) where peak parametric LOD reaches 1.71 between markers SNP_A-2290852 (rs7612097) and SNP_A-2030317 (rs17277717).
Figure 3.18. The haplotype for part of chromosomal region (3q21.3) where parametric LOD was 1.71. In the case of individual 10 it is not possible to determine which allele of the father (7) he inherited (non-informative meiosis) since the father is homozygous. The pedigree shows also the SNP ID and the genetic distance in cM.
Other inheritance patterns were also considered. When mode of inheritance was taken as recessive, parametric LOD scores where less than zero on all chromosomes. When all non-affected family members were given the disease status as Unknown (individual 5 and spouses of second generation individuals where still considered as non-affected), the maximum parametric LOD was 2.07 when penetrance was 100%. The maximum scores were obtained on the same chromosomes where the highest parametric LOD was observed when disease status of individuals 2 and 6 was taken as unknown and penetrance 100%.

3.5 The best candidate region

The widest linkage interval (>500kb) with the highest parametric LOD score (when inheritance pattern was taken as autosomal dominant with incomplete penetrance) was observed on chromosome 20 (Figure 3.19, 3.20 and 3.21). The linked region was 20cM.

![Figure 3.19. Parametric LOD on chromosome 20 when inheritance pattern is taken as autosomal dominant inheritance with incomplete penetrance. The X axis is genetic distance in cM along chromosome 20 while the Y axis is the LOD score.](image-url)
Figure 3.20. Zlr score on chromosome 20 when inheritance pattern is taken as autosomal inheritance with incomplete penetrance. The X axis is in cM along chromosome 20 while the Y axis is the Zlr score. Peaking of Zlr score marked with a yellow arrow.

Figure 3.21. NPL scores on chromosome 20 when inheritance pattern is taken as autosomal inheritance with incomplete penetrance. The X axis is in cM along chromosome 20 while the Y axis is the NPL score. The peaking of NPL score is marked with a yellow arrow.

The linkage interval on chromosome 20 was confirmed using short tandem repeats (STRs). Eleven STRs spaced at an average of about 10cM were used for the genotyping. Linkage analysis was carried out using Genehunter v2.0 (Kruglyak, 1996) (phenocopy rate 0.01; penetrance 1.0; equal allele frequencies). The multipoint linkage analysis carried out confirmed the linkage interval on
chromosome 20. Figure 3.22 shows a region of chromosome 20 where the affected
individuals and individuals 2 and 6 have the same haplotype as indicated by bars of
the same colour. Figure 3.23 and 3.24 show the NPL scores and parametric LOD
scores for chromosome 20 while Table 3.3 shows the output of Genehunter (gh2.1)
for chromosome 20. Figure 3.25 compares the linkage interval on chromosome 20
when linkage analysis was carried out using SNP genotyping data and when linkage
analysis was carried out using STR genotyping data.
Figure 3.22. Haplotype observed in the family (the last five genotyped STRs on chromosome 20). All affected individuals and family members 2 and 6 have the same alleles as that on one of the chromosome of the mother at loci D20S451, D20S164 and D20S173.
Figure 3.23. NPL scores on Chromosome 20 when using STRs markers. A score above 5 is significant evidence for linkage.
Figure 3.24. Parametric LOD scores on Chromosome 20 when using STRs markers. A LOD score above 3 is significant evidence for linkage while the value of 2.7 is suggestive evidence for linkage.
Table 3.3. Multipoint linkage analysis of STRs on chromosome 20 showing parametric LOD and NPL scores. Position refers to the genetic position in cM (deCode) on chromosome 20, p-value is the confidence level and information is how much the marker is informative. The LOD scores observed in the highlighted region is suggestive evidence for linkage.

<table>
<thead>
<tr>
<th>Position</th>
<th>LOD score</th>
<th>NPL score</th>
<th>p-value</th>
<th>Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.00</td>
<td>0.053466</td>
<td>0.300781</td>
<td>0.669066</td>
<td></td>
</tr>
<tr>
<td>2.39</td>
<td>-0.005504</td>
<td>0.375000</td>
<td>0.542243</td>
<td></td>
</tr>
<tr>
<td>4.78</td>
<td>-0.070372</td>
<td>0.390625</td>
<td>0.465946</td>
<td></td>
</tr>
<tr>
<td>7.16</td>
<td>-0.142445</td>
<td>0.414063</td>
<td>0.425655</td>
<td></td>
</tr>
<tr>
<td>9.55</td>
<td>-0.223258</td>
<td>0.417969</td>
<td>0.418747</td>
<td></td>
</tr>
<tr>
<td>11.94</td>
<td>-0.314585</td>
<td>0.539063</td>
<td>0.462822</td>
<td></td>
</tr>
<tr>
<td>15.31</td>
<td>-0.202038</td>
<td>0.417969</td>
<td>0.339392</td>
<td></td>
</tr>
<tr>
<td>18.68</td>
<td>-0.146994</td>
<td>0.414063</td>
<td>0.310875</td>
<td></td>
</tr>
<tr>
<td>22.04</td>
<td>-0.149276</td>
<td>0.414063</td>
<td>0.338572</td>
<td></td>
</tr>
<tr>
<td>25.41</td>
<td>-0.213443</td>
<td>0.417969</td>
<td>0.424039</td>
<td></td>
</tr>
<tr>
<td>28.78</td>
<td>-0.339868</td>
<td>0.539063</td>
<td>0.610526</td>
<td></td>
</tr>
<tr>
<td>30.53</td>
<td>0.37497</td>
<td>0.539063</td>
<td>0.507059</td>
<td></td>
</tr>
<tr>
<td>32.27</td>
<td>-0.325252</td>
<td>0.539063</td>
<td>0.479526</td>
<td></td>
</tr>
<tr>
<td>34.01</td>
<td>-0.318977</td>
<td>0.539063</td>
<td>0.509736</td>
<td></td>
</tr>
<tr>
<td>35.76</td>
<td>-0.314483</td>
<td>0.539063</td>
<td>0.595230</td>
<td></td>
</tr>
<tr>
<td>37.50</td>
<td>-0.312636</td>
<td>0.539063</td>
<td>0.192037</td>
<td></td>
</tr>
<tr>
<td>41.47</td>
<td>-0.232079</td>
<td>0.507813</td>
<td>0.511041</td>
<td></td>
</tr>
<tr>
<td>45.44</td>
<td>-0.138743</td>
<td>0.414063</td>
<td>0.381210</td>
<td></td>
</tr>
<tr>
<td>49.42</td>
<td>-0.042800</td>
<td>0.302813</td>
<td>0.319667</td>
<td></td>
</tr>
<tr>
<td>53.39</td>
<td>0.045852</td>
<td>0.307078</td>
<td>0.339392</td>
<td></td>
</tr>
<tr>
<td>57.36</td>
<td>0.116560</td>
<td>0.234375</td>
<td>0.360499</td>
<td></td>
</tr>
<tr>
<td>59.30</td>
<td>0.140956</td>
<td>0.226663</td>
<td>0.405457</td>
<td></td>
</tr>
<tr>
<td>61.24</td>
<td>0.156540</td>
<td>0.226663</td>
<td>0.464692</td>
<td></td>
</tr>
<tr>
<td>63.10</td>
<td>0.161549</td>
<td>0.226663</td>
<td>0.544359</td>
<td></td>
</tr>
<tr>
<td>65.12</td>
<td>0.154071</td>
<td>0.226663</td>
<td>0.651811</td>
<td></td>
</tr>
<tr>
<td>67.06</td>
<td>0.132072</td>
<td>0.226663</td>
<td>0.651811</td>
<td></td>
</tr>
<tr>
<td>68.99</td>
<td>0.070444</td>
<td>0.300701</td>
<td>0.606050</td>
<td></td>
</tr>
<tr>
<td>70.93</td>
<td>-0.045735</td>
<td>0.382813</td>
<td>0.520334</td>
<td></td>
</tr>
<tr>
<td>72.86</td>
<td>-0.126085</td>
<td>0.414063</td>
<td>0.503673</td>
<td></td>
</tr>
<tr>
<td>74.79</td>
<td>-0.207810</td>
<td>0.417969</td>
<td>0.503673</td>
<td></td>
</tr>
<tr>
<td>76.72</td>
<td>-0.298620</td>
<td>0.515625</td>
<td>0.723921</td>
<td></td>
</tr>
<tr>
<td>79.92</td>
<td>-0.175632</td>
<td>0.417969</td>
<td>0.507535</td>
<td></td>
</tr>
<tr>
<td>83.12</td>
<td>-0.014257</td>
<td>0.382813</td>
<td>0.423332</td>
<td></td>
</tr>
<tr>
<td>86.32</td>
<td>0.196035</td>
<td>0.226663</td>
<td>0.429696</td>
<td></td>
</tr>
<tr>
<td>89.52</td>
<td>-0.561017</td>
<td>0.466524</td>
<td>0.518637</td>
<td></td>
</tr>
<tr>
<td>92.71</td>
<td>0.890290</td>
<td>0.708125</td>
<td>0.747363</td>
<td></td>
</tr>
</tbody>
</table>

Table 3.4. Parametric LOD scores at different recombination fractions on chromosome 20 from marker D20S480 to q telomere.

<table>
<thead>
<tr>
<th>Markers</th>
<th>θ =0</th>
<th>θ =0.1</th>
<th>θ =0.2</th>
<th>θ =0.3</th>
<th>θ =0.4</th>
<th>θ =0.45</th>
</tr>
</thead>
<tbody>
<tr>
<td>D20S480</td>
<td>-0.862120</td>
<td>-0.169265</td>
<td>0.001391</td>
<td>0.058415</td>
<td>0.053570</td>
<td>0.032685</td>
</tr>
<tr>
<td>D20S451</td>
<td>0.796469</td>
<td>0.680028</td>
<td>0.547243</td>
<td>0.394092</td>
<td>0.214694</td>
<td>0.112508</td>
</tr>
<tr>
<td>D20S451</td>
<td>2.418677</td>
<td>2.001569</td>
<td>1.536804</td>
<td>1.014398</td>
<td>0.445678</td>
<td>0.184441</td>
</tr>
<tr>
<td>D20S164</td>
<td>2.001569</td>
<td>1.536804</td>
<td>1.014398</td>
<td>0.445678</td>
<td>0.184441</td>
<td></td>
</tr>
<tr>
<td>D20S173</td>
<td>1.780330</td>
<td>1.462507</td>
<td>1.108138</td>
<td>0.712749</td>
<td>0.298995</td>
<td>0.120857</td>
</tr>
</tbody>
</table>
Figure 3.25. Chromosome 20. The disease linked chromosomal interval when SNPs were used is shown in blue. The disease linked chromosomal interval when STRs were used is shown in red.

The linkage interval on chromosome 20 has 132 known genes. It was observed that the Zlr peaked between markers SNP_A-4214711 (rs1884328 at 111.56cM) and SNP_A-1933057 (rs1151621 at 112.72cM) (Table 3.5) when using SNPs data. Between these 2 SNPs there is an interval of 1.3cM and 360kb (Physical Position 61.32Mb – 61.69Mb). This region has a relatively high gene density (Table 3.6), 18 falling exactly within this interval.

Table 3.5. The linkage analysis data where the NPL peaks showing npLOD, NPL, Zlr scores, markers and physical position of markers.

<table>
<thead>
<tr>
<th>npLOD</th>
<th>dhat</th>
<th>NPL</th>
<th>ZLR</th>
<th>info</th>
<th>marker</th>
<th>PhysPos</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.7894</td>
<td>2.5070</td>
<td>3.5585</td>
<td>2.8706</td>
<td>0.9995</td>
<td>SNP_A-4214711</td>
<td>61325513</td>
</tr>
<tr>
<td>2.3303</td>
<td>80.0132</td>
<td>3.8717</td>
<td>3.2759</td>
<td>0.9676</td>
<td>SNP_A-2118501</td>
<td>61628090</td>
</tr>
<tr>
<td>2.4049</td>
<td>80.0132</td>
<td>3.9153</td>
<td>3.3279</td>
<td>0.9986</td>
<td>SNP_A-2182944</td>
<td>61688033</td>
</tr>
<tr>
<td>1.7851</td>
<td>2.5032</td>
<td>3.5399</td>
<td>2.8671</td>
<td>0.9982</td>
<td>SNP_A-1933057</td>
<td>61690281</td>
</tr>
</tbody>
</table>
Figure 3.26. Haplotype of family showing the region where there was peaking of the Zlr score on chromosome 20.
Table 3.6. Genes (26) in the region of markers SNP_A-4214711 (rs1884328) and SNP_A-1933057 (rs1151621). Genes in bold (18) fall exactly between the two markers. O stands for orientation.

<table>
<thead>
<tr>
<th>start</th>
<th>stop</th>
<th>Gene Symbol</th>
<th>O</th>
<th>Cytogenetic band</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>61280296</td>
<td>61283382</td>
<td>MIRN124-3</td>
<td>+</td>
<td>20q13.33</td>
<td>microRNA 124-3</td>
</tr>
<tr>
<td>61290950</td>
<td>61291574</td>
<td>LOC100131876</td>
<td>+</td>
<td>20q13.33</td>
<td>hypothetical LOC100131876</td>
</tr>
<tr>
<td>61297226</td>
<td>61317983</td>
<td>YTHDF1</td>
<td>-</td>
<td>20q13.33</td>
<td>YTH domain family, member 1</td>
</tr>
<tr>
<td>61337721</td>
<td>61342299</td>
<td>BIRC7</td>
<td>+</td>
<td>20q13.33</td>
<td>baculoviral IAP repeat-containing 7 (livin)</td>
</tr>
<tr>
<td>61342584</td>
<td>61356335</td>
<td>NKAIN4</td>
<td>-</td>
<td>20q13.33</td>
<td>Na+/K+ transporting ATPase:interacting 4</td>
</tr>
<tr>
<td>61374610</td>
<td>61391587</td>
<td>ARFGAP1</td>
<td>+</td>
<td>20q13.33</td>
<td>ADP-ribosylation factor GTPase activating protein 1</td>
</tr>
<tr>
<td>61394983</td>
<td>61432729</td>
<td>COL20A1</td>
<td>+</td>
<td>20q13.33</td>
<td>collagen, type XX, alpha 1</td>
</tr>
<tr>
<td>61429238</td>
<td>61429657</td>
<td>LOC100131010</td>
<td>+</td>
<td>20q13.33</td>
<td>hypothetical protein LOC100131010</td>
</tr>
<tr>
<td>61445109</td>
<td>61463192</td>
<td>CHRNA4</td>
<td>-</td>
<td>20q13.2-q13.3</td>
<td>cholinergic receptor, nicotinic, alpha 4</td>
</tr>
<tr>
<td>61461784</td>
<td>61468463</td>
<td>LOC100130587</td>
<td>+</td>
<td>20q13.33</td>
<td>hypothetical protein LOC100130587</td>
</tr>
<tr>
<td>61469686</td>
<td>61470090</td>
<td>LOC100133187</td>
<td>+</td>
<td>20q13.33</td>
<td>hypothetical protein LOC100133187</td>
</tr>
<tr>
<td>61494367</td>
<td>61497805</td>
<td>LOC100130152</td>
<td>-</td>
<td>20q13.33</td>
<td>hypothetical protein LOC100130152</td>
</tr>
<tr>
<td>61507986</td>
<td>61574437</td>
<td>KCNQ2</td>
<td>-</td>
<td>20q13.3</td>
<td>potassium voltage-gated channel, KQT-like subfamily, member 2</td>
</tr>
<tr>
<td>61589810</td>
<td>61600949</td>
<td>EEF1A2</td>
<td>-</td>
<td>20q13.3</td>
<td>eukaryotic translation elongation factor 1 alpha 2</td>
</tr>
<tr>
<td>61622577</td>
<td>61623968</td>
<td>C20orf149</td>
<td>+</td>
<td>20q13.33</td>
<td>chromosome 20 open reading frame 149</td>
</tr>
<tr>
<td>61630210</td>
<td>61639151</td>
<td>PTK6</td>
<td>-</td>
<td>20q13.3</td>
<td>PTK6 protein tyrosine kinase 6</td>
</tr>
<tr>
<td>61642607</td>
<td>61649301</td>
<td>SRMS</td>
<td>-</td>
<td>20q13.33</td>
<td>src-related kinase lacking C-terminal regulatory tyrosine and N-terminal mYritylation sites</td>
</tr>
<tr>
<td>Gene Location</td>
<td>Gene/Protein Description</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>------------------------------------------------------------------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.33</td>
<td>C20orf195 chromosome 20 open reading frame 195</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.33</td>
<td>PRIC285 peroxisomal proliferator-activated receptor A interacting complex 285</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.33</td>
<td>LOCl00130226 hypothetical protein LOC100130226</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.33</td>
<td>GMEB2 glucocorticoid modulatory element binding protein 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.3</td>
<td>STMN3 stathmin-like 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.3</td>
<td>RTEL1 regulator of telomere elongation helicase 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.3</td>
<td>TNFRSF6B tumor necrosis factor receptor superfamily, member 6b, decoy</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.3</td>
<td>ARFRP1 ADP-ribosylation factor related protein 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20q13.3</td>
<td>ZGPAT zinc finger, CCCH-type with G patch domain</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The best candidate genes in this chromosomal region were CHRNA4 (cholinergic receptor, nicotinic, alpha polypeptide 4) and KCNQ2 (potassium voltage-gated channel, KQT-like subfamily, member 2) since both have already been associated with epilepsy and both are expressed in the brain (Figure 3.27, 3.28). Another gene which also appeared to be a good candidate was STMN3 (stathmin-like 3) since this gene is the next gene just outside this interval, it is small, is highly expressed in brain (Figure 3.29) and is involved in the pathway that inhibits neurite outgrowth (Avoli, 2007; Poulain et al., 2007). Figures 3.27, 3.28 and 3.29 show the normalized expression of CHRNA4, KCNQ2 and STMN3 in various human tissues.
Figure 3.27. Expression of CHRNA4 in humans. The light blue colour and the lemon green colour indicate expression in the brain (accessed by searching for CHRNA4 at URL: http://symatlas.gnf.org/SymAtlas/).
Figure 3.28. Expression of KCNQ2 in humans. The light blue colour and the lemon green colour indicate expression in the brain (accessed by searching for KCNQ2 at URL: http://symatlas.gnf.org/SymAtlas/).
Figure 3.29. Expression of STMN3 in humans. The light blue colour ■ and the lemon green colour □ indicate expression in the brain (accessed by searching for STMN3 at URL: http://symatlas.gnf.org/SymAtlas/).
3.6 Gene sequence analysis

DNA sequencing of the coding regions of the CHRNA4, KCNQ2 and STMN3 was carried out on the DNA of an affected person (18) and that of a non-affected person (11) (Figure 3.1). No nonsense SNPs were detected in these 3 genes. In CHRNA4 four synonymous SNPs were detected in Exon 5 while in KCNQ2 only 1 synonymous SNP was detected in Exon 15. In STMN3 a number of known SNPs were detected in the 3' UTR region. Also in this region a number of SNPs that were not found in the NCBI database were also detected. SNPs and variations were also detected in the partially sequenced introns (mainly intron/exon junctions) in both CHRNA4 and KCNQ2 (Table 3.7).

In KCNQ2 both affected and non-affected individuals had the same genotype while in CHRNA4 the affected individuals had a different genotype from the affected ones. Affected individuals and family members 2 and 6 (Figure 3.1) had a 24bp deletion at position 61,451,224. A region of about 100 bp also showed a number of variations from the reference sequence both in the affected and the non-affected family members (Figure 3.30). A recent addition of an alternative assembly of chromosome 20 to the NCBI DNA sequence database showed the same variants in the 100bp region that were observed in the affected allele and non-affected allele.
### Table 3.7. Variations observed in the DNA sequence of CHRNA4, KCNQ2 and STMN3 in an affected and non-affected family member.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Exon/Intron</th>
<th>NCBI dbSNP</th>
<th>Position</th>
<th>Affected</th>
<th>Non-affected</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHRNA4</td>
<td>Intron 4</td>
<td>Not listed as a polymorphism in NCBI SNP database</td>
<td>61,451,224</td>
<td>A deletion of 24bp / No deletion</td>
<td>No deletion / No deletion</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>61,451,290</td>
<td>No deletion / deletion of 6bp</td>
<td>deletion of 6bp / deletion of 6bp /</td>
</tr>
<tr>
<td></td>
<td></td>
<td>A region with a number of variants from the reference sequence</td>
<td></td>
<td>As per alignment in Fig 3.25.</td>
<td>As per alignment in Figure 3.25</td>
</tr>
<tr>
<td></td>
<td>rs13041103</td>
<td>C/C</td>
<td></td>
<td>Large Ins with a C at position 25 of insertion / Large Ins with a T at position 25 of insertion</td>
<td>Large Ins with a T at position 25 of insertion / Large Ins with a T at position 25 of insertion</td>
</tr>
<tr>
<td></td>
<td>rs35836786</td>
<td>Large Ins with a C at position 25 of insertion / Large Ins with a T at position 25 of insertion</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>rs55781913</td>
<td>A/G</td>
<td></td>
<td>G/G</td>
<td>G/G</td>
</tr>
<tr>
<td>Exon 5</td>
<td>rs2229960</td>
<td>G/G</td>
<td></td>
<td>A/A</td>
<td>A/A</td>
</tr>
<tr>
<td>Exon 5</td>
<td>rs2229959</td>
<td>A/A</td>
<td></td>
<td>C/T or C/C</td>
<td>T/T</td>
</tr>
<tr>
<td>Exon 5</td>
<td>rs1044396</td>
<td>A/G or G/G</td>
<td></td>
<td>A/A</td>
<td>A/A</td>
</tr>
<tr>
<td>Exon 5</td>
<td>rs1044397</td>
<td>G/G</td>
<td></td>
<td>G/G</td>
<td>G/G</td>
</tr>
<tr>
<td>KCNQ2</td>
<td>Intron 2</td>
<td>T/T</td>
<td></td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>GENE - Reference sequence NT_011333.5</td>
<td>rs6062939</td>
<td>T/T</td>
<td></td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>Intron 9</td>
<td>rs6062931</td>
<td>T/T</td>
<td></td>
<td>T/T</td>
<td>T/T</td>
</tr>
<tr>
<td>Intron 10</td>
<td>rs5842421</td>
<td>Ins</td>
<td></td>
<td>Ins</td>
<td>Ins</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACAGACA G/</td>
<td></td>
<td>ACAGACA G/</td>
<td>ACAGACA G/</td>
</tr>
<tr>
<td></td>
<td></td>
<td>ACAGACA G</td>
<td></td>
<td>ACAGACA G</td>
<td>ACAGACA G</td>
</tr>
<tr>
<td>Intron 12</td>
<td>rs3215651</td>
<td>+/- (del CA)</td>
<td></td>
<td>+/- (del CA)</td>
<td>+/- (del CA)</td>
</tr>
<tr>
<td>Intron 13</td>
<td>rs33838007</td>
<td>+/- (del GA)</td>
<td></td>
<td>+/- (del GA)</td>
<td>+/- (del GA)</td>
</tr>
<tr>
<td>Exon 15</td>
<td>rs1801475</td>
<td>G/G</td>
<td></td>
<td>G/G</td>
<td>G/G</td>
</tr>
<tr>
<td>STMN3</td>
<td>3' UTR</td>
<td>rs2150910</td>
<td>61743101</td>
<td>C/C</td>
<td>C/T</td>
</tr>
<tr>
<td>Reference sequence ENSG00000213603</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3' UTR</td>
<td>rs2738776</td>
<td>C/T</td>
<td>61742816</td>
<td>C/T</td>
<td>C/T</td>
</tr>
<tr>
<td>3' UTR</td>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61742706</td>
<td>C/T</td>
<td>C/T</td>
</tr>
<tr>
<td>3' UTR</td>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61742678</td>
<td>C/T</td>
<td>C/T</td>
</tr>
<tr>
<td>3' UTR</td>
<td>SNP</td>
<td>Reference</td>
<td>Allele</td>
<td>Allele</td>
<td></td>
</tr>
<tr>
<td>--------</td>
<td>----------</td>
<td>-------------</td>
<td>--------</td>
<td>--------</td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/G</td>
<td>61742655</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/A</td>
<td>61742630</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/C</td>
<td>61742611</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/G</td>
<td>61742550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>T/T</td>
<td>61742502</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/A</td>
<td>61742289</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61742270</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/C</td>
<td>61742245</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/C</td>
<td>61742187</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/T</td>
<td>61742161</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/G</td>
<td>61742122</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>T/C</td>
<td>61742060</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>G/A</td>
<td>61742058</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/G</td>
<td>61742035</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/G</td>
<td>61742020</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61741967</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/G</td>
<td>61741940</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61741938</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>C/T</td>
<td>61741923</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Not found in NCBI SNP database</td>
<td>A/G</td>
<td>61741911</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Chapter 3 - Results

Figure 3.30. Alignment of Affected allele with the Reference sequence ENST00000370263, region 61,451,218 to 61,451,339. The bases highlighted in yellow indicate variations from the reference sequence. The underlined text is the 6bp that are deleted in the non-affected allele.

Figure 3.31. DNA sequence of intron 4 of CHRNA4 (reverse) in an affected individual showing superimposition of 2 sequences which starts at position 318.
3.7 Analysis of variations observed in Exon 5 of the CHRNA4 gene that differ between the affected and non-affected individuals

3.7.1 The synonymous SNPs in Exon 5 of CHRNA4

Exon 5 of the CHRNA4 gene was analysed in all family members to determine if the variations observed between the affected and non-affected family members were also present in the other family members. DNA sequencing determined that all affected family members (and individuals 2 and 6 - Figure 3.1) had the same haplotype at SNPs rs1044396 and rs1044397 which was different from that of the non-affected individuals (Figure 3.32). The genotype of the affected individuals (and individuals 2 and 6) at SNP rs1044396 is G/A (G/G as found in individuals 8 and 10) or C/T (C/C as found in individuals 8 and 10) in coding strand while non-affected individuals have A/A or T/T in coding strand. As regards SNP rs1044397 the genotype of the affected individuals (and individuals 2 and 6) is C/T (C/C as found in individuals 8 and 10) or G/A (G/G as found in individuals 8 and 10) in coding strand while non-affected individuals have T/T (or A/A) in coding strand. These SNPs are synonymous and do not cause any amino acid change in the polypeptide sequence.
Figure 3.32. DNA sequencing of the region of Exon 5 of the CHRNA4 gene where there are the SNPs rs1044396 and rs1044397 indicated with a red box (template strand). Individual 5 is non-affected and is homozygous T at rs1044397 and homozygous A at rs1044396 while individuals 18 and 8 are both affected. Individual 18 is heterozygous C/T at rs1044397 and heterozygous A/G at rs1044397. Individual 8 is homozygous C at rs1044397 and homozygous G at rs1044396.

For SNP rs1044396 the frequency of the C allele is reported to be 0.4 of the European population while in Sub-Saharan Africans, the frequency of the C allele is over 0.9. For SNP rs1044397 the frequency of the G allele is found in about 0.4 of European population while in Sub-Saharan Africans, the frequency of the G allele is over 0.9. Analysis of the sequence using RESCUE-ESE (Fairbrother et al., 2002) (http://genes.mit.edu/burgelab/rescue-ese/) which analyzes the region for exon splicing enhancer (ESE) elements showed that the presence of the G allele in rs1044397 disrupts an ESE (Figure 3.33).
Figure 3.33. Analysis of a region of Exon 5 of the CHRNA4 (coding sequence) using RESCUE-ESE to determine if the haplotype of the affected family members could interfere with any exonic splicing enhancer (ESE). The lower case (not upper case) indicate the SNPs rs1044396 and rs1044397.

A. DNA sequence of non-affected family members. B. DNA sequence of affected family members showing the disruption of an ESE, where ESES are shown in yellow.

Analysis of the same region using FAS-ESS (Wang et al., 2004) (http://genes.mit.edu/fas-ess/), a software tool for the detection of exonic splicing silencers (ESS) did not identify any changes between the different haplotypes of the affected and non-affected family members at SNPs rs1044396 and rs1044397.

3.7.2 The variants of intron 4 of CHRNA4

DNA Sequencing of part of intron 4 and the first part of Exon 5 of the CHRNA4 showed that there were two superimposed sequences in the affected individuals (and individuals 2 and 6 – Figure 3.1) but not in the non-affected individuals. The two PCR products could not be resolved by electrophoresis on a 2% agarose gel. It was however possible to distinguish between the two alleles using 8% polyacrylamide gel.
DNA sequencing of the allele in the affected individuals showed a number of variations from the reference sequence (ENST00000370263): a deletion of 24 bp at position 61,451,224, presence of a C at position 25 of the large insertion polymorphism (rs35836786) and a number of variations shown in Figure 3.30. These undefined variations which were also present in the allele of the non-affected family members matched an alternative assembly of chromosome 20 (NCBI NW_001838671.1). The alternative assembly sequence had the 6bp deletion like the non-affected allele.

Figure 3.34. Alignment of part of the intron 4 of the CHRNA4 gene of a non-affected individual and the alternative assembly of chromosome 20, NW_001838671.1 using BLAST (http://blast.ncbi.nlm.nih.gov/Blast.cgi). ‘Query’ reference to the sequence of the non-affected allele and ‘Sbjct’ refers to the sequence NW_001838671.1. Identities 355/355 (100%) and Gaps 0/355 (0%). The region highlighted shows the region of variants that were also found in the affected allele.
In all affected individuals the variants observed (deletion) in the intronic region were linked with rs35836786 having a C at position 25. A population study in healthy controls was carried out to determine the frequency of this affected allele haplotype in the Maltese population. Analysis of the PCR products of exon 5 of 100 controls showed the presence of the haplotype as observed on polyacrylamide gel, in 19% of the controls but no homozygosity for this allele was observed. DNA sequencing was carried out to determine the sequence of the alleles in the population. The affected allele haplotype was observed in eight of the alleles sequenced. The frequency of the allele in the population thus is 0.04. Due to heterozygosity it was impossible to determine how many of the 8 alleles that had the haplotype were linked to rs35836786 having a C at position 25. Linkage of the haplotype to rs35836786 having a T at position 25 or C at position 25 could be confirmed in some samples only.

DNA sequencing of the region of exon 5 of the CHRNA4 gene that has the SNPs rs1044396 and rs1044397, showed that all the 8 individuals that had the intron 4 haplotype were heterozygous at both SNPs, thus it was not possible to determine if the haplotype is linked with the C allele in rs1044396 and the G allele at rs1044397 as in the family members.

According to the HapMap, in individuals of European origin, there is strong linkage disequilibrium between SNPs in intron 4, exon 5 and part of intron 5 as observed in Figure 3.35.
3.7.3 Pre-mRNA folding prediction

In order to predict if the affected allele could have any effects on the expression of CHRNA4 gene, a number of bioinformatic software packages were used. Changes of pre mRNA folding and possible alternative splicing of Exon 5 due to the changes from the reference sequence were investigated.
Analysis of pre mRNA folding of about 300-350bp of intron 4 and the first 300bp of exon 5 of CHRNA4 was carried out using [mfold](http://frontend.bioinfo.rpi.edu/applications/mfold/cgi-bin/rna-form1.cgi) to determine if the variation observed in intron 4 could have an affect on the folding of this region. It was observed that the folding of the affected allele differs from the reference sequence (Figures 3.36, 3.37, 3.38 and 3.39).

![Reference sequence pre-mRNA folding](image1)

**Figure 3.36.** Reference sequence (ENST00000370263) pre-mRNA folding (free energy of $-\Delta G$ of 217.23 kcal/mol as given by mfold. Black arrow indicates 1st bp.

![Affected allele pre-mRNA folding](image2)

**Figure 3.37.** Affected allele pre-mRNA folding (free energy of $-\Delta G$ of 249.85 kcal/mol) as given by mfold. Black arrow indicates 1st bp.

Another software tool RNA2 (http://www.genebee.msu.su/services/rna2_reduced.html) was also used to predict pre-mRNA structure and the same observation was made as when mfold was used.
Figure 3.38. Reference sequence (ENST00000370263) pre-mRNA folding (free energy of $-\delta G$ of 83.4 kcal/mol as given by RNA2. Black arrow indicates $1^{\text{st}}$ bp.

Free Energy of Structure = $-83.4$ kcal/mol

Figure 3.39. Affected allele pre-mRNA folding (free energy of $-\delta G$ of 112.1 kcal/mol as given by RNA2. Black arrow indicates $1^{\text{st}}$ bp.

Free Energy of Structure = $-112.1$ kcal/mol
3.7.4 Alternative splicing prediction

When NetGene 2 (http://www.cbs.dtu.dk/services/NetGene2/) was used for the analysis of a DNA region having about 300-350bp of intron 4 where variation was observed and about 300bp of intron 5, the software did not predict any different possible alternative splice sites between the reference sequence and the variant.

Table 3.8. NetGene2 analysis of affected allele and Reference sequence resulted in identical output when it comes to the identification of alternative splice and donor sites. Donor/acceptor site predictions are made by TWO detection levels for true sites, one level where around 50% of the true sites are detected with very few false positive (marked with H), and another level where nearly all true sites are found, but with more false predictions as well. In this case the first level of detection missed the true acceptor splice site at position 355. Phase refers to where in the intron the splicing occurs.

<table>
<thead>
<tr>
<th>Donor splice sites, direct strand</th>
<th>Donor splice sites, complement strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos 5'→3' phase strand confidence 5' exon intron 3'</td>
<td>pos 5'→3' phase strand confidence 5' exon intron 3'</td>
</tr>
<tr>
<td>22 - + 0.00 CTCTGTGAC*GTGAGAATGC</td>
<td>469 187 0 - 0.55 GGGAAAGAA*GTGACGTCAA</td>
</tr>
<tr>
<td>398 258 0 - 0.60 CATGAAACAG*GTGGGCTTGG</td>
<td>353 273 0 - 0.71 CTTCTGACG*GTGGGTGAC</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Acceptor splice sites, direct strand</th>
<th>Acceptor splice sites, complement strand</th>
</tr>
</thead>
<tbody>
<tr>
<td>pos 3'→5' phase strand confidence 5' intron exon 3'</td>
<td>pos 3'→5' phase strand confidence 5' intron exon 3'</td>
</tr>
<tr>
<td>355 2 + 0.79 TCTCTGGCAG*TCTTGAAGGG</td>
<td>491 0 + 0.15 CATAAGCAG*AATCTGACCA</td>
</tr>
<tr>
<td>581 75 0 - 0.00 ACTCTTGCAG*AAGCTGACCT</td>
<td></td>
</tr>
</tbody>
</table>

CUTOFF values used for confidence:

- Highly confident donor sites (H): 95.0 %
- Nearly all true donor sites: 50.0 %
- Highly confident acceptor sites (H): 85.0 %
- Nearly all true acceptor sites: 20.0 %

Splice view (http://bioinfo.itb.cnr.it/oriel/splice-view.html) was also used to check for alternative splicing. The results indicate variation between the affected allele, and the reference control as indicated in Figure 3.40.
Chapter 3 - Results

A. Reference sequence

B. Affected allele

Figure 3.40. Alternative donor and acceptor sites as given by SpliceView. Donor sites are highlighted in yellow while acceptor sites are highlighted in pink. The region in black is part of the sequence of intron 4 of the CHRNA4 gene while the sequence in blue is part of the sequence of Exon 5. The G (reverse of C at position 25 of the large insertion polymorphism, rs35836786) is highlighted in green.

A. Reference sequence and B. Affected allele.

Alternative Splice Site predictor (www.es.embnet.org/~mwang/assp.html) was also used to identify possible alternative donor or acceptor sites. Donor site cut-off was set to 8.0 to increase stringency of the analysis. The software detected a strong
alternative isoform cryptic donor site in the affected allele (and alternative assembly) which was less strong in the Reference Sequence.

A. Score

<table>
<thead>
<tr>
<th>Position</th>
<th>Putative splice site</th>
<th>Sequence</th>
<th>Score</th>
<th>Intronic GC</th>
<th>Alt./Cryptic</th>
<th>Constitutive</th>
<th>Confidence**</th>
</tr>
</thead>
<tbody>
<tr>
<td>177</td>
<td>Alt. isoform/cryptic donor</td>
<td>GTGGGCATGGtgtgtgtgt</td>
<td>8.778</td>
<td>0.671</td>
<td>0.812</td>
<td>0.136</td>
<td>0.832</td>
</tr>
<tr>
<td>338</td>
<td>Constitutive acceptor</td>
<td>tctgccacTGCTGACGGG</td>
<td>8.369</td>
<td>0.643</td>
<td>0.431</td>
<td>0.552</td>
<td>0.219</td>
</tr>
<tr>
<td>471</td>
<td>Alt. isoform/cryptic acceptor</td>
<td>ctgaccacGAGAACTGCA</td>
<td>6.016</td>
<td>0.600</td>
<td>0.698</td>
<td>0.293</td>
<td>0.581</td>
</tr>
<tr>
<td>474</td>
<td>Alt. isoform/cryptic acceptor</td>
<td>cgaccagcAAGAATGCACCA</td>
<td>5.412</td>
<td>0.600</td>
<td>0.831</td>
<td>0.163</td>
<td>0.804</td>
</tr>
</tbody>
</table>

B. Diagrammatic representation of results

Figure 3.41. A. Results of ASSP show the presence of a strong alternative isoform/cryptic donor at position 177 of the submitted affected allele (see Figure 3.40 for sequence) sequence with a score of 8.778. B. The yellow box (in splice site box) indicates the alternative donor site while the green box indicates the constitutive acceptor site. In the 'codon usage' box the blue line indicates that the region before the alternative isoform/cryptic donor site can be coding and can be joined with the region after the constitutive acceptor site.
3.8 The minigene vector experiment

The PCR product of about 2300bp that spanned from intron 4 to intron 5 could not be inserted in the pSPL3 plasmid. A number of different primer sets with different restriction sites were tried and the amplification of the PCR product using the proof reading enzyme was always successful. A DNA control for the restriction enzymes confirmed that the restriction enzymes were digesting the DNAs. Also the control for digested plasmid did not show any growth on the LB/ampicillin plates. A range of PCR product: plasmid ratios were used. *E. coli* JM109 were also used for the transformation but still no insert was observed in the plasmid. Different batches of T4 ligase (Promega Corp) were also used. Different temperatures for the ligation reaction were also tried (2-8°C overnight, at 22°C for 3 hrs and at 15°C for 8hrs). The experiment to determine if the intron 4 haplotype in the affected individuals results in alternative splicing using the pSPL3 vector was not successful. Thus it was decided to extract RNA from the leucocytes of an affected person and sequence the RNA transcript of CHRNA4 to determine if there is any variation between the transcript sequence of the affected individual and the reference sequence.

The only plausible reasons for the unsuccessful insertion could be that the insert is quite big and big inserts are more difficult to insert in a plasmid than a smaller insert or that since the first part of the PCR product sequence (300bp) is very high in %G content, the PCR product may develop a conformation that hinders restriction digest or ligation.
3.9 Sequencing of the CHRNA4 transcript

RNA was extracted from an affected family member. Reverse transcription was carried out and sequencing of the cDNA was performed to determine if there is any variation in the transcription due to the intron 4 variations in the Affected allele. There was no change in the transcript sequence of exon 5 when compared with the reference sequence.

3.10 Copy Number Variation analysis

Copy number variation (CNV) could predispose to disease. In order to determine if CNV could contribute to the disease phenotype in this family, a copy number variation analysis between affected members of the family and non-affected members was performed. The Genotyping Console™ software (Affymetrix) was used to analyse the SNP data to determine and CNVs. The only chromosomal region where the affected family members (except individual 10) shared a loss or gain (individual 9) when compared with non-affected members was on Chromosome 12 region 12p11.1. No known genes are found in this region (Figure 3.42).

CNV analysis was also performed between affected members of the family having the disease haplotype and non-affected members having the disease haplotype to try to identify any possible genomic regions where a loss or gain could protect an individual that has the disease haplotype from developing the seizure phenotype. A number of chromosomal regions where the non-affected members having the disease haplotype had a loss/gain when compared with non-affected individuals were observed (Table 3.9).
Figure 3.42. Copy Number Variation analysis where affected family members are compared with non-affected family members. Region in red box indicates where all affected family members (except individual 10) have some variation from the non-affected individuals. A loss is indicated by a line lower that the main line (<2) while a gain is indicated by a line above the main line (>2).
Table 3.9. Loss and gain at several chromosomal regions where individuals 2 and 6 showed variations when compared with the affected individuals as determined by the Genotyping Console™ software package.

<table>
<thead>
<tr>
<th>Gain/Loss</th>
<th>Kb</th>
<th>Cytoband</th>
<th>Known genes in region</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain</td>
<td>193</td>
<td>4p16.3</td>
<td>ZNF718 Zinc finger protein 718</td>
</tr>
<tr>
<td>Gain</td>
<td>720</td>
<td>10q11.23</td>
<td>PARG poly (ADP-ribose) glycohydrolase</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CTGLF4 centrin, gamma-like family, member 4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TIMM23 translocase of inner mitochondrial membrane 23 homolog</td>
</tr>
<tr>
<td>Loss</td>
<td>492</td>
<td>15q11.2</td>
<td>LOC283755</td>
</tr>
<tr>
<td>Gain</td>
<td>75</td>
<td>16p13.11</td>
<td>RRN3 RNA polymerase I transcription factor homolog</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>MPBV17L mitochondrial membrane protein-like</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>C16orf45 chromosome 16 open reading frame</td>
</tr>
<tr>
<td>Gain</td>
<td>240</td>
<td>17q23.1</td>
<td>DHX40 DEAH (Asp-Glu-Ala-His) box polypeptide 40</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>CLTC clathrin, heavy chain (Hc)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>TMEM49 transmembrane protein 49</td>
</tr>
</tbody>
</table>
3.11 Conclusion

A family with GEFS+ or febrile seizures was recruited for the study. Multipoint linkage analysis showed a region of about 20cM with a maximum parametric LOD of 2.67 on chromosome 20 when penetrance was 0.9 and the disease status of family members 2 and 6 was taken as ‘Unknown’. Two other linkage intervals where identified, one on chromosome 13 when penetrance was lowered to 0.7 where a maximum parametric LOD of 2.0352 was obtained and the other on chromosome 3 when only family members having febrile seizures that stopped at 4 years of age where taken as affected where a maximum parametric LOD of 1.71 was obtained. These loci have never been linked to GEFS+ or to epilepsy or febrile seizures.

The linkage interval on chromosome 20 was confirmed using short tandem repeats (STRs). The ZlR peaked in a region having 18 genes. The best candidate genes in this region were CHRNA4, KCNQ2 and STMN3. CHRNA4 and KCNQ2 both code for sub-units of ion channels and have been associated with epilepsy (Lerche et al., 2001). STMN3 is highly expressed in brain and is involved in neurite growth (Kang et al., 2005). DNA sequencing of the exons and intron/exon junctions identified a number of SNPs and variations some of which not found in the NCBI database. In the CHRNA4 a haplotype was identified in the affected individuals (and family members 2 and 6 Figure 3.1). The frequency of this haplotype in the Maltese population is 0.04 and no homozygosity for the haplotype was observed.

Various bioinformatics software packages were used to predict possible variation in pre-RNA folding of the affected allele when compared with the reference sequence and also possible
cryptic donor/acceptor sites that differ from the reference sequence. Variation was observed in the putative folding of the pre-mRNA. With reference to prediction of alternative donor sites, Splice View and ASSP identified possible alternative donor sites that differed from the reference sequence.

cDNA analysis of RNA extracted from peripheral leucocytes did not identify any sequence variation from the reference sequence.

Copy Number Variation analysis identified a region on chromosome 12 where the affected individuals (except individual 10) show variation from the non-affected individuals. When the affected individuals genotypes were compared with that of family members 2 and 6, there were various DNA regions where variation in CNV was observed.
Chapter 4

Discussion
4.1 Introduction

As described in the Results chapter, this linkage analysis study of SNP genotyping data in a Maltese family, identified suggestive evidence of a novel locus for GEFS+ or febrile seizures on chromosome 20, together with two other potential loci on chromosomes 3 and 13 which could possibly explain the incomplete penetrance and the phenotypic heterogeneity of this disease in this family. These loci have not previously been linked to GEFS+ or familial febrile seizures (Section 4.2).

Variations in the haplotype of intron 4/ exon 5 of CHRNA4 were observed between the affected (and individuals 2 and 6 – Figure 3.1) and the non affected individuals. One of these variations was a 24bp deletion which was only present in the affected individuals (and individuals 2 and 6 – Figure 3.1). The frequency of the 24bp deletion variation in the population was 0.04. A number of bioinformatic tools were used to predict the possible genetic contribution that could predispose to the disease. DNA sequencing of the mRNA extracted from leukocytes did not indicate that the variation could affect RNA splicing.

CNV variations were observed between individuals 2 and 6 (Figure 3.1) (non-affected individuals that have the disease haplotype) and the affected individuals (section 4.8).

The usefulness of such genetic studies is being explored in great detail in order to assist in the development of more targeted drugs in these complex epilepsy syndromes (Evans and McLeod, 2003). Gene therapy in this area is also being
attempted (Section 4.9). These developments are only becoming possible with the
great advancement being made in the discovery of genetic information through
linkage studies and other genetic studies (Section 4.9).

4.2 Syndrome in the recruited family

A three-generation Maltese family who had seven affected members was recruited
for the study following their informed consent. Analysis of the family history both
from information given by the family and also through hospital medical history files,
indicated that febrile seizures were present in both the first and second generation
while in the third generation, other epilepsy types were present together with
learning difficulties and/or behavioural problems in all the affected individuals.

In the family medical history files there was very limited information on the actual
febrile seizures, especially for the first and second generation since the family
members had not been taken to hospital after having had the seizures. In the medical
files of all family members there is reference to febrile seizures as part of the general
family history. With respect to the third generation, some of the children were taken
to hospital when the seizures occurred but no EEG, CT Scan or MRI was performed
on these individuals except for family member 14. An EEG diagnosed the boy as
having a focal onset seizure disorder and also an MRI investigation showed no
cerebral pathology. Thus most of the information on seizure history was gathered
from the family members themselves. This could be a limitation of the study when
trying to characterize the syndrome of this family. The inheritance pattern of the
condition appears to be autosomal dominant with incomplete penetrance when the
pedigree is analysed in the light of the known clinical history. Studies have shown
that the inheritance pattern can change when the non-affected family members have an EEG which also exhibited epileptiform activity which did not manifest in an actual seizure (Jain et al., 1997). On the other hand there are many cases where individuals suffering from idiopathic epilepsy have a normal EEG (Betting et al., 2006). Thus an EEG may not always support the diagnosis of a syndrome or a linkage study. From the literature review undertaken and with the gathered clinical history it was possible to characterize the syndrome in this family as being GEFS+ or febrile seizures (Goulini-Kallael, 2004; Graves, 2006; Scheffer et al., 2007).

GEFS+ is genetically heterogeneous and has been linked to a number of loci where mutations have been found in genes coding for ion channels (Brice, 2004). To date, GEFS+ has been linked to 19q13.1 (SCN1B), 2q21-q33(SNC1A), 5q31.1-q33.1(GABRG2), 2q24 (SCN2A) and 1p36.3 (GABRD) and recently also to 8p23-p21. (Wallace, 1998; Bauluc, 1999; Escagy, 2000; Bauluc, 2001; Audenaert, 2005; Dibbins, 2004; Baulac et al., 2008). There are however a number of families having GEFS+ phenotype that is not linked to these identified loci (Brice, 2004; Selmer et al., 2008). GEFS+ is characterized by phenotypic heterogeneity where family members having the same mutation exhibit a different phenotype (Graves, 2006; Scheffer et al., 2007). Phenotypic heterogeneity could be caused by a number of factors including difference in the expression of the mutated gene, the mutation may lower the epileptic threshold and then other genetic factors or environmental factors may affect the phenotype of a family member (Brice 2004). The inheritance pattern of GEFS+ is autosomal dominant with incomplete penetrance. The incomplete penetrance indicates also the presence of other modifying factors besides the mutation that determine if a family member will
develop the condition. Both phenotypic heterogeneity and incomplete penetrance indicate a more complex inheritance pattern in GEFS+. In the study the recruited family had cases of classical febrile seizures that remit before age six, febrile seizures at six years of age and focal onset epilepsy together with behavioural difficulties which have been also noted in another case of GEFS+ (Grant and Vaquenz, 2005) and learning difficulties. The idiopathic nature, the autosomal inheritance with incomplete penetrance and the phenotypic heterogeneity are all characteristics of GEFS+ and also indicate the complex inheritance component where more than one genetic locus could be involved (Gourfinkel-An, 2004). The predominance of febrile seizures in the family could also characterize the phenotype in this family as being febrile seizures.

4.3 Linkage analysis of SNP genotyping data

Although it is well known that modifier genes could be responsible for the different seizure phenotypes in a family with GEFS+, only one linkage study has identified another chromosomal interval were a modifier gene might be present (Ito et al., 2006; Nabbout et al., 2007). In common idiopathic epilepsies where complex inheritance is present some loci have been identified (Mulley, 2005). In this study, the central aim was to try to identify any other possible loci besides the main locus that could be linked to the disease. A high density SNP chip was used for genotyping this family with 7 affected individuals. A single family with a number of affected individuals eliminates the problem of locus heterogeneity and in this study it was imperative that all linkage intervals observed belong to the same family. Also such a family is informative enough to give evidence of linkage. Studies have shown that
SNP genotyping can have higher power than 10cM spaced STR marker sets, can
narrow the linkage interval and are also able to detect other loci besides a major
locus (Schaid et al., 2004; Gonzalez Neira et al., 2007; Middleton et al., 2006). In
this study the GeneChip® Human Mapping 250K Nsp Array which queries 262,000
SNPs was used. The median physical distance between SNPs is 2.5kb while the
average physical distance between SNPs is 5.8kb. The median value indicates that
there are SNPs which are spaced by less than 2.5kb. The high density of SNPs poses
the problem of linkage disequilibrium (LD) between markers and could lead to
complications when reading results (Schaid et al., 2004). To reduce the incidence of
false positives due to LD, it was decided that STR genotyping of candidate regions
could be performed although there still exists the possibility of random association
between STRs (Moulard et al., 1999).

Linkage analysis using the GeneChip® Human Mapping 250K Nsp Array identified
a number of loci were maximum multipoint parametric LOD of 2.67 (p<0.0001),
NPL of 8.4940 and Zlr of 3.3279 were observed when the disease mode of
inheritance was taken as dominant and penetrance was 100% and phenotype of
individuals 2 and 6 (Fig 3.1) was taken as ‘unknown’. When individuals 2 and 6
were taken as non-affected maximum LOD of 2.4333 (a<0.0001), NPL of 8.4940
and Zlr of 3.3279 were obtained when mode of inheritance was taken as dominant
and penetrance was 90%. Both these LOD values are suggestive evidence of linkage
(parametric LOD score > -2 and < 3) while the NPL value which is > 5 provides
significant evidence for linkage. The main linkage interval observed was localized
within the chromosomal region 20q13.32-33, a region that has already been linked to
epilepsy (Berkovic et al., 1994; Dedek et al., 2001; Phillips et al., 1995). A
limitation of the study was that some family members refused to participate in this project. This missing data resulted in non-informative meiosis contributing to a potentially lower parametric LOD score. The main candidate chromosomal linkage interval is of about 20cM (94cM to 114cM - deCode Map, 55686Kb-62184Kb-Physical position). Linkage analysis of chromosome 20, using microsatellite markers, confirmed the linkage on the chromosome between 83cM to the telomere of chromosome 20 between markers D20S480 at position 83cM and telomere, 51M and 62M - Physical Position. The linked region using STRs is wider than that observed with SNP genotyping. The last 2 Megabases of chromosome 20 were only covered by SNPs (no STRs were genotyped in this region).

There were other chromosomal regions where the maximum parametric LOD of 2.67 was obtained when linkage analysis of the SNP data was carried out but these intervals were less than 500kb and it could be that the LOD observed is due to random association in the family and linkage disequilibrium between markers. In a study by Klein et al., it was noted that with increase in marker density there is an increase in false positive results when it comes to linkage. LD was mostly observed between markers < 200kb apart (Klein et al., 2005). Scientifically no region should be excluded until it is analysed since these regions are far larger than the average of 10-30kb of LD observed throughout the genome in the general population (Clarke, 2003; Moulard et al., 1999). STR genotyping using region specific high density STR panels can be performed in these regions to confirm or reject the linkage observed.
Chapter 4 - Discussion

Table 4.1. Cytogenetic band where the maximum parametric LOD was 2.67.

<table>
<thead>
<tr>
<th>Cytogenetic band</th>
<th>Linked to epilepsy in previous studies</th>
<th>Genes linked</th>
</tr>
</thead>
<tbody>
<tr>
<td>7p15.3-15.2</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>8q21.2</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>8q22.1</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>8q22.2-22.3</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>12p13.2</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>15q26.3</td>
<td>N/A</td>
<td></td>
</tr>
<tr>
<td>20q13.32-33</td>
<td>Myokymia with neonatal epilepsy</td>
<td>KCNQ2</td>
</tr>
<tr>
<td></td>
<td>Benign neonatal epilepsy, Type 1;</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(EBN1)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Nocturnal frontal lobe epilepsy, type</td>
<td></td>
</tr>
<tr>
<td></td>
<td>1 (ADNFLE)</td>
<td>CHRNA4</td>
</tr>
</tbody>
</table>

On chromosome 13, in a region of about 15cM (cytogenetic band 13q13-14), the NPL score was 8.4940 and the Zlr score was 3.3279. Maximum parametric LOD score of 2.0352 was observed when penetrance was lowered to 0.7. The lowered parametric LOD score is due to the recombination observed in two unaffected individuals. This region has been linked to a number of neurological disorders including autism, schizophrenia, major depressive disorder and anorexia nervosa (Smith et al., 2002).

Another interesting linked locus was that observed on chromosome 3 when the mode of inheritance was taken as autosomal dominant and only individuals that suffered from febrile seizures before six years of age were taken as being affected. A peak
parametric LOD score of 1.71 which is suggestive evidence for linkage was observed in a region of about 5cM at 3q21.3. All the affected family members had the same haplotype (Fig 3.18). Family member 10 (Fig 3.18) also shared the same haplotype but it was not possible to determine which allele he inherited from the father (family member 7) since the father was homozygous. This region had never before been linked to febrile seizures or epilepsy.

Both the regions on chromosome 13 and chromosome 3 are good candidates for further study. The parametric LOD score observed is suggestive evidence for linkage and the linked regions are large enough to possibly exclude linkage disequilibrium between markers. These regions may harbour a gene that could be interacting with the gene in the main locus on chromosome 20.

4.4 The linkage interval on chromosome 20

The widest linkage interval was that on chromosome 20, cytogenetic band 20q13.32-33. This region has been linked to epilepsy in other studies although in the studied families the phenotypes (ADNFLE, EBN1 and Myokymia with neonatal epilepsy) differ from those of the family recruited for this study (Table 1.2 and 4.1). The use of SNP genotyping was expected to give a narrower linkage interval when compared with STR genotyping using the 10cM spaced STR set but in this study the SNP genotyping still yielded a linkage interval of about 20cM (Yang et al., 2005) which is a very wide linkage interval. This region has about 132 genes. In such circumstances it would have been necessary to use in between STR markers or tagSNPs to try to narrow the linkage interval but the Zlr score peaked between
SNP_A-4214711 (rs1884328) and SNP_A-1933057 (rs1151621) (physical positions 61,325.513kb and 61,690.281kb respectively). Thus it was decided to investigate this region. This region has 18 genes, five of which are hypothetical proteins and two are chromosome 20 open reading frames (uncharacterized protein). The best candidate genes in the region included CHRNA4 and KCNQ2 as both genes have been already linked to an epilepsy phenotype, both code for ion channels and both are expressed in the brain. Most mutations causing epilepsy phenotypes have been found in ion channels (Lerche et al., 2001). Another potential candidate was STMN3 which is involved in the development of the nervous system and is highly expressed in the brain and nervous system. STMN3 controls the early phases of Purkinje cell dendritic differentiation during cerebellum development (Poulain et al., 2008). Purkinje cells are GABAergic neurones. STMN3 has not yet been linked to epilepsy.

DNA sequencing of the exons (plus intron/exon junctions) of CHRNA4, KCNQ2 and STMN3 was carried out in two affected individuals and a non-affected family member. Alignment of the sequences with the reference sequences identified a number of SNPs (Table 3.6). In CHRNA4 the affected individuals and family members 2 and 6 had a different haplotype than that observed in the non-affected individuals. A deletion of 24 bp at position 61,451,224 was observed only in affected individuals and in individuals 2 and 6 (Figure 3.1). A population study was carried out to determine the frequency of this 24bp deletion. Out of the 100 controls analysed, 8 had the deletion but no homozygosity for this deletion was observed. This could be due to the fact that since the frequency of the allele in the population is 0.04, the probability of observing homozygosity is 0.4% which means that in a population study of 100 controls it may not be detected (probability less than 1%).
Another reason for not finding any homozygosity could be that homozygosity could lead to a fatal phenotype. This study focused on trying to understand if the variations observed, in intron 4 of CHRNA4, in the affected individuals, could have an effect on the expression of CHRNA4. A point worth mentioning is that recently (February 2008) an alternate assembly of chromosome 20 was added to the NCBI database (NW 001838671.1) This DNA sequence is from the whole genome assembly released by the J Craig Venter Institute (Levy et al., 2007). The DNA sequence of this region of chromosome 20 aligns with a number of variations from the reference sequence that were present in both the affected and non-affected family members.

4.5 CHRNA4 haplotype

The affected individuals (and individuals 2 and 6 Figure 3.1) had a different haplotype on the CHRNA4 that the non-affected individuals and this also differed from the alternative assembly, NW 001838671.1.
Table 4.2. Haplotypes observed in CHRNA4.

<table>
<thead>
<tr>
<th>Affected individuals and family members 2 and 6</th>
<th>Non affected (0.04)</th>
<th>Population and family NCBI NW Ensemble</th>
<th>Alternative assembly NCBI NW 001838671.1</th>
<th>Reference sequence ENST0000370263</th>
</tr>
</thead>
<tbody>
<tr>
<td>24bp deletion</td>
<td>No deletion</td>
<td>24bp deletion</td>
<td>No deletion</td>
<td>No deletion</td>
</tr>
<tr>
<td>No deletion</td>
<td>6bp deletion</td>
<td>No deletion</td>
<td>6bp deletion</td>
<td>No deletion</td>
</tr>
<tr>
<td>Cluster of variations from reference sequence</td>
<td>Cluster of variations from reference sequence</td>
<td>Cluster of variations from reference sequence</td>
<td>Cluster of variations from reference sequence</td>
<td>Cluster of variations from reference sequence</td>
</tr>
<tr>
<td>SNP rs35836786 large insertion with C at position 25 of insertion</td>
<td>SNP rs35836786 large insertion with T at position 25 of insertion</td>
<td>SNP rs35836786 large insertion with C at position 25 of insertion</td>
<td>SNP rs35836786 large insertion with T at position 25 of insertion</td>
<td>Does not have the insertion</td>
</tr>
<tr>
<td>C allele at SNP rs1044396</td>
<td>T allele at SNP rs1044396</td>
<td>Not Determined as heterozygosity was observed in all samples.</td>
<td>T allele at SNP rs1044396</td>
<td></td>
</tr>
<tr>
<td>G allele at SNP rs1044397</td>
<td>A allele at SNP rs1044397</td>
<td>Not Determined as heterozygosity was observed in all samples.</td>
<td>A allele at SNP rs1044397</td>
<td></td>
</tr>
</tbody>
</table>

Although this region shows high linkage disequilibrium (Figure 3.35) between markers in HapMap, the haplotypes observed do not confirm this. The region analysed is less than 2000bp and the degree of variability in haplotypes (not all data...
of population study shown) confirms that this region has a lower LD than previously indicated in the European population or else LD in this region in the Maltese population is different from that of the European population studied. For the other populations in HapMap LD data in this region is scarce.

The different haplotype observed between the affected individuals (and family members 2 and 6 - Figure 3.1) and the non-affected individuals confirm that if the changes already observed in the region do not predispose to epilepsy another variant linked to this haplotype is potentially predisposing to the disease phenotype.

4.6 Synonymous SNPs in Exon 5 of CHRHA4

A SNP in which both alleles lead to the same polypeptide sequence is termed as synonymous. In this study both the C allele and the T allele at SNP rs1044396 form part of the codon that codes for serine (TCC or TCT) and both the A allele and the G allele at SNP rs1044397 form part of the codon that codes for alanine (GCA or GCG). Synonymous SNPs were thought to be innocuous and not to cause any disease but recent studies are demonstrating the even synonymous SNPs can cause or predispose to a disease or to variable response to a drug (Kimchi-Sarfaty et al., 2007; Komar, 2007; Sauna et al., 2007).

Exonic splicing enhancers (ESEs) are specific short oligonucleotide sequences that enhance pre-mRNA splicing when present in exons and play important roles in constitutive and alternative splicing. Studies have shown that the disruption of exonic splicing enhancers (ESEs) can lead to disease (Aretz et al., 2004; Foerster et al., 2004). Some mutations in cystic fibrosis are now known to disrupt an ESE
leading to exon skipping (Aznarez et al., 2003;). Thus even the most benign looking polymorphism in an exon cannot be ignored as it may affect the splicing process (Pagani et al., 2003).

It was also noted that the more proximal the SNP is to an intron/exon junction the greater the possibility of exerting an influence on the splicing pattern (Hull et al., 2007). The SNP rs1044397 is 100bp away from the exon/intron junction and is the last known SNP in Exon 5. When analysing the DNA sequence having the SNP rs1044396 and SNP rs1044397 using RESCUR-ESE (a software for predicting ESE motifs) the G allele at rs1044397 disrupts an ESE motif which could lead to a decrease in splicing of exon 5. This is just hypothetical and experimental evidence needs to be collected to test this hypothesis. A number of studies have been carried out to test for associations between the febrile seizures or generalized seizures or temporal lobe epilepsy (TLE) and polymorphisms in rs1044396 and SNP rs1044397. In a study conducted by Chou et al., individuals with the T allele at rs1044396 had a higher incidence of febrile convulsions (Chou et al., 2003). While in another study no significant association studies was found between having the T allele and a higher incidence of TLE (Cavalleri et al., 2005).

4.7 Variations observed in intron 4 of CHRNA4

The alignment of the partial sequence of intron 4 in the affected individuals with the Reference sequence showed a number of variations. The 24bp deletion was present only in the affected individuals and individuals 2 and 6 while the large insertion was present in both the affected and non-affected family members but in the affected
family members, there was a C at position 25 of insertion instead of a T. Although these variations are in the intron they can still exert an effect on the expression of the CHRNA4 (Heinzen et al., 2007).

4.7.1 Prediction of secondary structure of pre-mRNA

Pre-mRNA transcripts include both intronic and exonic sequences. Being single stranded the pre-mRNA folds onto itself to form the most stable structure. A study by Hiller et al., has shown that the secondary structure of pre-mRNA influence exon splicing (Hiller et al., 2007). The function of splicing enhancers and silencers depends on their localization within an RNA conformation. Stronger exon inclusion occurs when the enhancer is located in a loop compared with the location in the stem structure of the pre-mRNA secondary structure (Hiller et al., 2007). Thus base pair changes in introns will cause alteration of pre-mRNA structure favouring the minimum free energy structure and in so doing there could be changes which lead to an enhanced motif located in a loop region (single stranded RNA) ending up in stem region (double stranded) (Figure 4.1). This could result in a decrease in exon inclusion.
Figure 4.1. Pre-mRNA folding of the affected allele and Reference sequence where a sequence in a loop region end up in a stem region due to the DNA variation in the affected allele (shown in a blue box).

For example, mutations in the 5' splice site of exon 10 of the tau gene lead to increased inclusion of exon 10 leading to frontotemporal dementia and Parkinson (Donahue et al., 2006). These mutations are thought to destabilize a stem loop pre-mRNA structure at the 5' splice-site of exon 10 (Donahue et al., 2006). A study by Boom Wee et al., reported that it is possible to reduce the severance of Duchenne Muscular dystrophy by antisense nucleotide mediated exon skipping but the access of binding sites for antisense nucleotides for exon splicing are dependent on pre-mRNA folding (Boom Wee et al., 2008). Thus it is very evident that understanding pre-mRNA folding could also assist in the development of novel drug therapy.
Pre-mRNA is bound by numerous proteins, which influence its ability to fold freely. Recent studies have identified RNA chaperones that are involved in the proper folding of mRNA (Russell, 2008). Also, the formation of secondary structures occurs cotranscriptionally, which favours short-range over long-range base pairing where experiments have shown that it is limited to a region of about 50bp downstream of the transcribing polymerase (Hiller et al., 2007). The in vivo situation is complex and currently available software is unable to consider all these interactions that influence pre-mRNA folding. Thus although the software available (eg. mfold and RNA2) does predict the difference in the folding of the ‘affected allele’ when compared with the reference allele this is only a simulation and may not reflect what actually happens in vivo (Capriotti and Marti-Renom, 2008). The use of the two software packages varied considerably in the structures of the same alleles (both reference and affected allele) and the calculated free energy of the structure (although the ratios of the free energies remained the same) Mfold predicts RNA secondary structure (folding) by algorithms considering energy minimization using nearest neighbour energy parameters (Zucker, 2003). RNA2 predicts the secondary structure by finding a structure with the minimum free energy (Brodsky et al., 1992). Both software packages use the same parameters for predicting the secondary structure but still variation was observed. This shows the need of harmonising RNA prediction software if in the near future such software will be required to predict RNA structures for drug design.
4.7.2 Alternative splicing

Alternative splicing can be defined as RNA splicing variation mechanism in which the exons of the pre-mRNA, are separated and reconnected so as to produce alternative ribonucleotide arrangements. Alternative splicing occurs in nature and uses genetic expression to facilitate the synthesis of a greater variety of proteins from a one DNA sequence (Brett et al., 2001).

![Diagram of splicing](image)

**Figure 4.2. Sites involved in the splicing of pre-mRNA.**

Alternative splicing could also lead to disease (Garcia-Blanco et al., 2004; Nagau et al., 2005). This could be due to a change in the intronic sequence that activates a donor or acceptor site resulting in the insertion of a new exon in the mRNA sequence which can result in a different polypeptide or else the change in the intronic sequence may deactivate a donor or acceptor site resulting in exon skipping or the presence of an intronic region in the final mRNA (Baralle and Baralle, 2005). Also it is postulated that changes in the intronic sequence may activate pseudo splice sites. Changes at the branch site could also result in aberrant intron splicing (Janssen et al., 2000).
In order to postulate if the DNA sequence changes observed in the affected allele could result in alternative splicing of CHRNA4 when compared with the reference sequence, a number of predictive software were used. Analysis using NetGene 2 (Brunak et al., 2001) did not predict any potential donor/acceptor sites that were any different from those predicted for the Reference sequence. Analysis of the sequences using SpliceView (http://bioinfo.itb.cnr.it/oriel/splice-view.html) resulted in the detection of two possible donor sites that were different from the Reference sequence. When analysing the data using the software ASSP a putative donor site that in the terms of DNA sequence it was the same in both the reference sequence and the affected allele but the score for the affected allele was higher when compared with that of Reference sequence (8.778 vs 7.686) due to the neighbouring nucleotides (Wang and Marin, 2006). Analysis for alternative splicing sites in the affected allele as compared to the reference sequence mainly identified potential alternative donor sites. The software packages used differed in their predicted output.

There is substantial difference between the confidence levels given by software in the prediction of alternative donor and acceptor sites (Table 4.3) (Rogovin and Milanes, 1997; Thanaraj, 2000; Pertea, 2001; Wang and Marin, 2006).

Table 4.3. Confidence level in the prediction of alternative donor and acceptor splice sites.

<table>
<thead>
<tr>
<th>Software</th>
<th>Correct classification of donor sites</th>
<th>False negative donor sites</th>
<th>Correct classification of acceptor sites</th>
<th>False negative acceptor sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>NetGene2-Level 1</td>
<td>Nearly 100%</td>
<td>50%</td>
<td>Nearly 100%</td>
<td>50%</td>
</tr>
<tr>
<td>Level 2</td>
<td>33%</td>
<td>4.6%</td>
<td>33%</td>
<td>2.5%</td>
</tr>
<tr>
<td>Splice view</td>
<td>15%</td>
<td>11.9%</td>
<td>19%</td>
<td>16.3%</td>
</tr>
<tr>
<td>ASSP</td>
<td>71.23%</td>
<td>67.45%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The data in Table 4.3 indicates that ASSP is the most reliable software when it comes to detect donor and acceptor sites. NetGene2-Level 1 analysis failed to identify the constitutive acceptor site (3') of Exon 5.

4.7.3 Analysis of the cDNA transcript of CHRNA4

A study was carried out to determine if there will be any DNA sequence change in the transcription of Exon 5 of the CHRNA4 in an individual having the affected allele as compared with a normal control. No difference was observed in the transcript of the affected allele. The following hypothesis could be postulated from this evidence:

1. The changes observed in intron 4 of CHRNA4 in the affected individuals do not have any effect on the transcription of Exon 5 or else, the DNA sequence observed is only of that allele that does not have the polymorphisms.

2. The changes observed in intron 4 of CHRNA4 in the affected individuals could affect rate of transcription or any other parameter of the RNA transcription which however does not influence correct splicing of intron 4.

3. The RNA was extracted from leukocytes and it is possible that there could be variation in tissue expression where although in the transcript of RNA extracted from leukocytes, the transcript did not show any difference from the normal control, this cannot be extended for neuronal cells. In neuronal cells expression might be
different and the intronic variations observed in the ‘affected allele’ might have an effect in neuronal cells.

Hypothesis 1 implicates that the variation observed in intron 4 of the affected individuals might not be the mutation on chromosome 20 that is causing the disease phenotype and that another ‘mutation’ linked to the variations in intron 4 is causing the disease phenotype in this family. It could also implicate that the polymorphisms in the affected allele could have caused complete skipping of exon 5 and the cDNA sequence is only of the ‘normal’. Hypothesis 2 on the other hand implicates that the functional study performed could not access the function of the variation observed in intron 4 and other different functional studies need to be carried out in order to confirm this hypothesis. Any change in rate of transcription of a gene could effect further downstream processes such as translation leading to a decrease in protein production (Kimichi-Sarfati et al., 2007). Hypothesis 3 postulates that it is possible that the variation observed in the ‘affected allele’ might have an effect on the expression of CHRNA4 in neuronal cells but not in leucocytes. To test this hypothesis, RNA has to be extracted from brain neuronal cells. This can be carried out by collecting post-mortem brain specimens neuronal cells or from brain specimens removed from patients during resective surgery for intractable epilepsy. The study will involve RNA extraction from the cells to determine cDNA transcript sequence of CHRNA4 and DNA extraction to determine the sequence of intron 4.
4.8 Copy Number Variation analysis

Copy Number Variation (CNV) is defined as a DNA segment that is 1 kb or larger and present at variable copy number in comparison with a reference genome (Feuk et al., 2006). Identical twins have been found to have different CNVs and this could explain why in a pair of identical twin a member develops a disease while the other does not (phenotypically discordant twins) (Bruder et al., 2008) In the study for the construction of the first-generation CNV map of the human genome one of the platforms used for the detection of CNV was the Affymetrix GeneChip Human Mapping 500K array (Redon et al., 2006). In this study the SNP data used for genotyping was also used to determine CNVs in family members. The results obtained from CNV analysis depend a lot on the quality of the genotyping data. Data of poor quality (low call rates) give unsatisfactory results. The software used to analyse the SNP data (Genotyping Console™ - Affymetrix®) first performs a data QC which samples can be used for analysis. The hypothesis behind this analysis was the possibility the CNV might be responsible for the disease phenotype or that some CNVs present in individuals 2 and 6 might ‘protect’ these individuals from the disease even though they might have the disease causing genotype. The most interesting result was the presence of losses and gains of CNVs that were present in individuals 2 and 6 (Figure 3.1) but not in the affected individuals. These regions might harbour changes that could ‘protect’ individuals 2 and 6 from a seizure phenotype. If this is verified by other studies, CNV analysis would have helped in the identification of a possible scenario for the incomplete penetrance in the family.
Chapter 4 - Discussion

CNV analysis has only started in these recent years. Already some associations of CNV with disease have been made (Lee and Lupski, 2006). With the discovery of new CNV in the coming months, research will yield important insights, elucidating biological pathways and the mechanisms of human disease (McCarroll and Altshuler, 2007).

4.9 Genetic studies in epilepsy and the development of novel drug therapies

The increased understanding of how individual genetic variability may lead to a disease phenotype or a response to a specific drug, has great potential in the discovery and development of novel pharmacological targets and therapies (Holmes, 2002; Chancy and Kass, 2003).

The first step in a genetic study being a linkage study, an association study or a candidate gene approach study in order to discover the genetic variant that is predisposing to a disease/drug response. This is done by identifying variations between the DNA of affected individuals and that of either non-affected family members, as in linkage studies, or with normal controls as in association or candidate gene approach studies. A number of genetic studies in epilepsy have been carried out which have identified mutations and polymorphic variants in ion channels that predispose to a seizure phenotype (Kalachicov et al., 2002; Hirose et al., 2002; Tripathi and Jain, 2004; Muley et al., 2005). This has helped significantly in discovering genes involved in the generation of seizures (Graves, 2006).

Following the identification of a mutation or variant, functional studies need to be performed in order to understand how the variant affects function of the gene and
how this could predispose to an epilepsy phenotype (Hoda et al., 2008). Mutations in CHRNA4 have been found to cause increased sensitivity to acetylcholine, increasing the probability of the opening of the ion channel (Bertrand et al., 2002) leading to hyperexcitability. A mutations in KCNQ2 (5-bp insertion at the triplet encoding amino acid 534) has been found to cause protein truncation while another mutation (R207Q) caused hperexcitability (Bievert and Steilein, 1999; Wuttke et al., 2007). All these mutations led to a seizure phenotype.

Once the function of these mutations or variants have been understood, then one can start thinking of a possible way to counteract the effect of the mutation or variant and here is where drug development comes in (Evans and McLeod, 2003).

The study carried out promises to identify genetic variants predisposing to GEFS+ and also possible variants that will explain the phenotypic heterogencity and incomplete penetrance. The more the information gathered from genetic studies, the further the possibility of developing targeted drugs.

4.9.1 Drug targeting

CHRNA4 has been linked to ADNFLE (Table 4.1)(Phillips et al., 1995). A number of mutations have been identified in various families but in all of them the mutation seems to increase the sensitivity of the acetylcholine receptor (of which CHRNA4 is a subunit) that results in an increased predisposition to the opening of the receptor that is thought to cause the neuronal network dysfunction that results in epileptic seizures (Bertrand et al., 2002; Roda et al., 2008). Since the effect of this mutation
now seems clear, it is possible to develop novel drug therapy for this condition which targets this mutation.

This is still a daunting experience given the complexity of living organisms. There are various points where one can target. These include the receptor itself by the use of an anticholinergic agent, or else through gene expression by the development of oligonucleotides which target the 3’ UTR of CHRNA4 gene, thus down regulating the expression of the gene (Changxian et al., 2001).

4.9.2 Drug response in epilepsy

A number of pharmacogenomic studies have been carried out in order to understand what genetic variation or expression could be causing pharmacoresistance in epilepsy but progress to date has been minimal. Genetic variations in multidrug-transporters have been intensely studied but results have been unsatisfactory (Brodie, 2004, Leschzinget et al., 2007, Shahwan et al., 2007) When it comes to drug targets, several studies have also been carried out, but data at times is conflicting and no progress has been made to tailor any drugs depending on drug target changes observed (Remy and Beck, 2006; Shoeye et al., 2006).

Epilepsy can also be a progressive disorder (Tahvanainen et al., 1994). Some gene mutations that predispose to this phenotype have been discovered (Delgada-Esquito, et al., 2001). No drug is yet available that stops the progression of these types of epilepsies.
Although the number of available antiepileptic drugs have increased throughout the years, patients having epilepsy still have a decrease in quality of life when compared with normal individuals (Berto, 2002). With progress in the identification of mechanisms involved in seizure generation and the mechanisms involved in gene expression, the pharmaceutical industry starts developing target specific drugs (Evans and McLeod, 2003). The number of identified targets is already overwhelming and to develop a drug for each target is economically not feasible. But it might be possible that drugs that target gene expression at the molecular level might prove effective in a vast number of cases (Bhattacharyya et al., 2007).

4.9.3 Gene therapy in epilepsy

Genetic studies in recent years have elucidated the pathway for a seizure phenotype and this has led to a better understanding of the aetiology epilepsy. Armed with this information, gene therapy is being attempted experimentally (Li et al., 2003; Noe et al., 2007). In a recent study gene therapy strategy was found to decrease spontaneous seizures and suppress their progression in chronic epileptic rats, thus representing a promising new therapeutic strategy (Noe et al., 2008).

4.10 Future work

This study has shown that this family has a great deal of potential for other future work and there are a number of other studies that can be carried out in order to understand better the syndrome in this family.
1. The recruitment of other family members of the second generation, since there are four additional family members that were never contacted. Genotyping of these members will help in increasing the parametric LOD score and maybe also narrow the linkage interval.

2. Perform Quantitative Real Time PCR for mRNA of CHRNA4 to determine if there is variation in concentration between an affected and a normal individual. Develop a functional study that could be used to measure rate of pre-mRNA splicing to mRNA.

3. The variations observed in the affected allele are linked to the disease phenotype, and although it might in itself not be contributing to the disease phenotype, it could be linked to another nearby locus that is causing the disease phenotype. The 3' UTR and promoter region of CHRNA4 should be sequenced. SNP tagging of the region might also narrow the linked region.

4. If approach 3 above does not narrow the region, before moving on to sequencing new genes, the promoter region of KCNQ2 and STMN3 should be sequenced since previously published studies have indicated that SNPs or DNA variations in other regions beside the coding region can cause disease by a plethora of mechanisms; changing pre-mRNA secondary structure, alternative splice sites (Donahue et al., 2006).
5. The affected allele could also represent a common allele predisposing to febrile seizures. This will need to be proven using an association study in a large number of affected individuals and controls.

6. MicroRNAs are also involved in gene expression regulation by inducing cleavage of the mRNA or inhibit translation (John et al., 2004). Software that finds targets for microRNA can be used to analyse 3'UTR data for possible variants that could affect binding of microRNAs starting with the 3'UTR of STMN3 that has already been sequenced.

7. Future work can be carried out to confirm the linkage intervals on chromosome 13 and 3 using STRs and also narrowing the linkage interval by tagSNPs or high density STR markers. The linkage interval on chromosomes 7, 8, 12 and 15 will be confirmed by high density STR markers.

8. The confirmation of the CNV variation in family members 2 and 6 using another method such as comparative genomic hybridization (CGH) (Cho et al., 2006). It should also be possible to assess the possible interaction of these regions with any identified variants in the affected individuals which could result in protection against seizures.

4.11 Conclusion

This present research has indicated a parametric LOD of 2.67 at $p<0.0001$ giving suggestive evidence for linkage of GEFS+ or febrile seizures to a novel locus on
chromosome 20 in the recruited family. This locus was also confirmed by 10cM spaced STRs. Two other loci with a lower parametric LOD score were also identified. Disease modifier genes might be present at these loci which could help in understanding the incomplete penetrance and the phenotypic heterogeneity in the family. CNV analysis also identified chromosomal regions where the two non-affected family members having the disease haplotype differ from the affected individuals. The genes in these regions may provide a protective effect against seizure susceptibility or delay in onset of seizures.

DNA sequencing of three candidate genes in the chromosomal interval identified a haplotype in CHRNA4 that is only present in the affected individuals. The main difference was a 24bp deletion in intron 4 which was only present in affected individuals but also had a frequency of 0.04 in a random sample of the general Maltese population. The variation observed in intron 4 of CHRNA4 in the population shows that there is low LD in this region in the general Maltese population.

The use of bioinformatic software tools predicted that the DNA sequence of the 'affected allele' cause different folding of pre-mRNA when compared to the Reference sequence. An alternative donor splice site was also predicted which had a higher score than that predicted for the Reference sequence. Both these results indicated that the 'affected allele' could effect CHRNA4 splicing and expression. A study to determine the cDNA sequence of CHRNA4 mRNA in an individual having the affected allele did not show any variation in the DNA from the reference sequence. However this does not exclude the possibility that in neuronal cells
CHRNA4 might be regulated in a different way than in leukocytes (from which RNA was extracted) and that the variants in the ‘affected allele’ might still affect transcription regulation.

If this variant does not predispose to the seizure phenotype in this family, the presence of a disease linked haplotype indicates that the variant predisposing to the phenotype could be proximal since the region shows low LD in population. The affected allele could also represent a common allele predisposing to febrile seizures.

Such genetic studies have high potential in order to provide a better understanding of the cause of a complex disease as epilepsy, and to understanding a patient’s predisposition to drug response and side-effects. The use of this novel genetic information in novel drug design, will certainly improve epilepsy treatment in the future, improving the quality of life for persons with epilepsy (Depondt and Shorvan, 2006; Garcia-Blanco et al., 2004).
References
References


Aznarez I, Chan EM, Zielenski J, Blencowe BJ, Tsui LC. Characterization of disease-associated mutations affecting an exonic splicing enhancer and two cryptic


Boon Wee K, Pramono ZAD, Wang JL, MacDorman KF, Lai PS, Yee WC. Dynamics of Co-Transcriptional Pre-mRNA Folding Influences the Induction of


Engel J. ILAE classification of epilepsy syndromes. Epilepsy Res. 2006;70:5-10.


References


Gonzalez Neira A, Rosa-Rosa JM, Osario A, Gonzalez E, Southey M et al. Genomewide high-density SNP linkage analysis of non-BRCA1/2 breast cancer families identifies various candidate regions and has greater power than microsatellite studies. BMC Genomics 2007;8:299.


Ottman R. Progress in the genetics of the partial epilepsies. Epilepsia 2001;42:24-30


Tan NC, Mulley JC, Berkovic SF. Genetic association studies in epilepsy: “The truth is out there”. Epilepsia 2004;45:1429-42.


Winawer MR. Phenotype definition in epilepsy. Epilepsy and Behavior 2006;8:462-76.


Appendix
Appendix 1 - Classification of epilepsy syndromes


<table>
<thead>
<tr>
<th>Groups of Syndromes</th>
<th>Specific Syndromes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Idiopathic Focal Epilepsies of Infancy and Childhood</td>
<td>Benign Infantile Seizures (Non Familial)</td>
</tr>
<tr>
<td></td>
<td>Benign Childhood Epilepsy with Centrotemporal Spikes</td>
</tr>
<tr>
<td></td>
<td>Early Onset Benign Childhood Occipital Epilepsy (Panayiotopoulos type)</td>
</tr>
<tr>
<td></td>
<td>Late Onset Childhood Occipital Epilepsy (Gastaut type)</td>
</tr>
<tr>
<td>Familial (Autosomal Dominant) Focal Epilepsies</td>
<td>Benign Familial Neonatal Seizures</td>
</tr>
<tr>
<td></td>
<td>Benign Familial Infantile Seizures</td>
</tr>
<tr>
<td></td>
<td>Autosomal Dominant Nocturnal Frontal Lobe Epilepsy</td>
</tr>
<tr>
<td></td>
<td>Familial Temporal Lobe Epilepsy</td>
</tr>
<tr>
<td></td>
<td>Familial Focal Epilepsy with Variable Foci*</td>
</tr>
<tr>
<td>Symptomatic (or Probably Symptomatic) Focal Epilepsies</td>
<td>Limbic Epilepsies</td>
</tr>
<tr>
<td></td>
<td>• Mesial Temporal Lobe Epilepsy with Hippocampal Sclerosis</td>
</tr>
<tr>
<td></td>
<td>• Mesial Temporal Lobe Epilepsy Defined by Specific Etiologies</td>
</tr>
<tr>
<td></td>
<td>• Other Types Defined by Location and Etiology</td>
</tr>
<tr>
<td></td>
<td><strong>Neocortical Epilepsies</strong></td>
</tr>
<tr>
<td></td>
<td>• Rasmussen Syndrome</td>
</tr>
<tr>
<td></td>
<td>• Hemiconvulsion - Hemiplegia Syndrome</td>
</tr>
<tr>
<td></td>
<td>• Other Types Defined by Location and Etiology</td>
</tr>
<tr>
<td>Idiopathic Generalized Epilepsies</td>
<td>Benign Myoclonic Epilepsy in Infancy</td>
</tr>
<tr>
<td>----------------------------------</td>
<td>------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Epilepsy with Myoclonic Astatic Seizures</td>
</tr>
<tr>
<td></td>
<td>Childhood Absence Epilepsy</td>
</tr>
<tr>
<td></td>
<td>Epilepsy with Myoclonic Absences</td>
</tr>
<tr>
<td></td>
<td>Idiopathic Generalized Epilepsies with Variable Phenotypes</td>
</tr>
<tr>
<td></td>
<td>• Juvenile Absence Epilepsy</td>
</tr>
<tr>
<td></td>
<td>• Juvenile Myoclonic Epilepsy</td>
</tr>
<tr>
<td></td>
<td>• Epilepsy with Generalized Tonic-Clonic Seizures Only</td>
</tr>
<tr>
<td></td>
<td>Generalized Epilepsies with Febrile Seizures Plus*</td>
</tr>
<tr>
<td>Reflex Epilepsies</td>
<td>Idiopathic Photosensitive Occipital Lobe Epilepsy</td>
</tr>
<tr>
<td></td>
<td>Other Visual Sensitive Epilepsies</td>
</tr>
<tr>
<td></td>
<td>Primary Reading Epilepsy</td>
</tr>
<tr>
<td></td>
<td>Startle Epilepsy</td>
</tr>
<tr>
<td>Epileptic Encephalopathies (in which the epileptiform abnormalities may contribute to progressive dysfunction)</td>
<td>Early Myoclonic Encephalopathy</td>
</tr>
<tr>
<td></td>
<td>Ohtahara Syndrome</td>
</tr>
<tr>
<td></td>
<td>West Syndrome</td>
</tr>
<tr>
<td></td>
<td>Dravet Syndrome (Previously Known as Severe Myoclonic Epilepsy in Infancy)</td>
</tr>
<tr>
<td></td>
<td>Myoclonic Status in Non-Progressive Encephalopathies</td>
</tr>
<tr>
<td></td>
<td>Lennox-Gastaut Syndrome</td>
</tr>
<tr>
<td>Appendix</td>
<td></td>
</tr>
<tr>
<td>-----------------</td>
<td>--------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Progressive Myoclonus Epilepsies</td>
<td>See specific Myoclonic syndromes</td>
</tr>
<tr>
<td>Seizures Not Necessarily Requiring a Diagnosis of Epilepsy</td>
<td>Benign Neonatal Seizures</td>
</tr>
<tr>
<td></td>
<td>Febrile Seizures</td>
</tr>
<tr>
<td></td>
<td>Reflex Seizures Alcohol Withdrawal</td>
</tr>
<tr>
<td></td>
<td>Seizures Drug or Other Chemically-Induced Seizures Immediate and early Post Traumatic Seizures Single Seizures or Isolated Clusters of Seizures Rarely Repeated Seizures (Oligo-Epilepsy)</td>
</tr>
</tbody>
</table>
Appendix 2 - Patient information leaflet (English and Maltese versions)

Patient information leaflet

Epilepsy has an incidence of about 1% in the population. In most cases, the causative factor is unknown. In recent years, studies have shown that genetic variations in genes involved in neurotransmission could predispose to epilepsy.

Multigenerational families with a number of affected individuals can be very useful to detect these genetic variants.

All I would need from you is 6ml of blood that would be taken from the brachial vein by a qualified person. If you are reluctant to give blood you can give a buccal rinse which is collected by taking about 20-50ml of sterilized saline swirl it in the mouth for a minute and then dispense it in a sterile centrifuge tube. I would also be asking you a number of questions regarding the condition.

From the blood or buccal rinse I will extract the DNA. Tests will be carried out using the DNA to try to determine what variations could be causing the seizures.

This study could shed light on another possible cause of epilepsy. If the cause of a disease is known, it would be much easier to design new drugs for the treatment of epilepsy.

Thanks for your help.

Marisa Cassar B.Pharm (Hons)
Informazzjoni lill-Pazjent

L-epilessija ghandha incidenza ta’ 1% fil-popolazzjoni ingnerali. F’hafna mill-kazijiet, il-fatturi involuti fl-izvilupp ta’ din il-kondizzjoni mhumiex maghrufa. F’dawn l-ahhar snin saru ċerta studji li wrew li varjazzjoni ġenetika f’ġeni involuti fit-transmissjoni fin-nervituri jistghu jikkawżaw l-epilessija.

Familji kbar b’hafna membri effetwati bl-epilessija jistghu ikunu ta’ ghajnuna kbira f’dawn it-tip ta’ studji.

Kull ma jkolli bzonn minn ghandek hu ftit demm li jittiehed minn vina f’dirghajk li jittiehed minn persuna kwalifikata. Jekk tiddejjaq taghti d-demm, tista’ taghti ‘buccal rinse’ billi titfa ftit ilma f’halqek u tobużqu go flixkun li nkun ipprovdejtlek jien. Ser ikolli bzonn insaqskix mistoqsijiet dwar il-kondizzjoni tieghek.

Mid-demm naghmel estrazzjoni tad-DNA. Wara naghmel xi testijiet fuq id-DNA. Dawn it-testijiet jistghu jindikaw x’varjazzjoni ġenetika hemm li qed tikkawża l-aċċessjonijiet.


Grazzi tal-ghajnuna tieghek

Marisa Cassar B.Pharm (11ons)
Appendix 3 - Patient Consent Form - Adult

Jiena ghalaqt it-tmintax-il sena. Talbuni biex niehu sehem fl-istudju bl-isem ta’

“Linkage analysis in a familial case of idiopathic epilepsy and its implication in drug development”

Il-ghan u d-dettalji ta’ l-istudju spjegathomli Marisa Cassar, li wkoll iċċaratli xi mistoqsijiet li ghamilt.

Naghti l-kunsens tieghi lill-persuna responsabbli biex jtehdu l-kampjuni u dan nifhem li jista’ jikkawza xi skomdu.

Jien nifhem li r-rizultati ta’ dan l-istudju jistghu jintuzaw ghal skopijiet xjentiffiċi u jista’ jigi ppubblikat rapport bil-miktub: jekk isir hekk bl-ebda mod ma nista’ nkun identifikat/a, individualment jew bhala grupp minghajr il-kunsens tieghi.

Jien ma ghandi l-ebda dmir li niehu sehem

Jekk ikolli xi diffikulta waqt l istudju, nista’ nistaqsi ghal:

Dr. Janet Mifsud fuq in-numru tat-telefon 23402845
Marisa Cassar fuq in-numru tal-mobile 79270004

Isem il-partenċipant __________________________
Firma tal-partenċipant __________________________
Numru tal-identita __________________________

Dr. Janet Mifsud hi responsabbli ghal din ir-riċerka __________________________
Appendix 4 - Patient Consent Form – Child (under 18 years)

Jien ghandi il-kustodja ta’ ______________________

Talbuni biex __________________________ jiehu/tiehu sehem fl-istudju bl-isem ta’

“Linkage analysis in a familial case of idiopathic epilepsy and its implication in drug development”

Il-ghan u d-dettalji ta’ l-istudju spjegathomli Marisa Cassar, li wkoll iċċaratli xi mistoqsijiet li ghamilt.

Naghti l-kunsens tieghi lill-persuna responsabbli biex jiehdu l-kampjuni u dan nifhem li jista’ jikkawza xi skomdu.

Jien nifhem li r-rizultati ta’ dan l-istudju jistghu jintuzaw ghal skopijiet xjentifici u jista’ jiġi ppubblikat rapport bil-miktub: jekk isir hekk bl-ebda mod ma nista’ nkun identifikat/a, individualment jew bhala grupp minghajr il-kunsens tieghi.

Jien ma ghandi l-ebda dmir li nghati l-kunsens li __________________________ jiehu/tiehu sehem f’dan l-istudju u dan qed naghmlu minn raija.

Jekk ikolli xi diffikulta waqt l-istudju, nista’ nistaqsi ghal:

Dr. Janet Mifsud fuq in-numru tat-telephone 23402845
Marisa Cassar fuq in-numru tal-mobile 79270004

Isiem il-partecipant____________________________
Firma tal-partecipant____________________________
Numru tal-identita____________________________
Dr. Janet Mifsud hi responsabbli ghal din ir-riċerka __________________________