Haplotype mapping in epilepsy genetics and pharmacogenetics

Gianpiero L. Cavalleri

A thesis for submission to University College London for the degree of Doctor of Philosophy (Human Genetics)

University College London, Department of Biology

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

Abstract

Despite the success achieved in identifying mutations causing Mendelian forms of epilepsy, little progress has been made in illuminating variation contributing to the development of more common forms of the condition. The aim of this thesis was to improve methodology that would aid the detection of variation functional in the development and treatment of common forms of epilepsy.

Attempted replication of previous claims of association with common forms of epilepsy allowed resolution of true effects from false positive results and illustrated shortcomings in the application of association based genetic mapping. Where effects looked real, haplotype-based fine mapping techniques were applied to identify candidate causal variation. The tagging SNP method was applied to HapMap data in an attempt to identify variation that might guide the safe prescription of Vigabatrin, an effective antiepileptic drug limited by a serious adverse drug reaction.

Results outlined here show all previous claims of association with temporal lobe epilepsy and febrile seizures are likely to be false positives. However, supportive evidence is presented that common genetic variation predisposes to juvenile myoclonic epilepsy in a population specific manner. Evidence is also presented that variation that could be considered clinically relevant for the safe administration of Vigabatrin does not exist in the candidate genes examined here. Finally, a study design is proposed that seeks to significantly increase genetic power of detection through the incorporation of lessons learned from previous studies.

In conclusion, epilepsy appears to be as genetically complex as the phenotypic spectrum of the condition would suggest. This work illustrates key areas in study design that require improvement and presents methodological developments in genetic mapping techniques which together provide a solid platform for future success.

Table of contents

Acknowledgements

Thank you to David Goldstein, Nick Wood and Sanjay Sisodiya for opening up the world of epilepsy genetics to me and supervising this work with unrelenting drive, support and enthusiasm.

To all the members of the Goldstein lab for their help.- Sarah, Koursh, Anna,, Richard, Julia, Alice and Helen. To Mari-Wyn in particular for her efforts in pushing the TLE work through to completion, Mike Weale (co-author on Illumina chapter) who contributed enormously to all the statistics and tagging work, Nicole Soranzo who carried out all the mRNA work and Nicole Walley for her help with BRD2.
To Chantal Depondt for her massive effort in phenotyping and databasing, and Rinki Singh for Vigabatrin work.

To the Irish connection - Norman Delanty and Colin Doherty for helping create a great collaboration, may it lead to many discoveries. Peter Kinrons is a co-author on the Vigabatrin chapter. He makes a good drinking buddy and he'd be a decent scientist if he wasn't a clinician. John Lynch for his contribution to BRD2 and TLE. John let's not go near that robot ever again.

To my friends and everyone at the UCL postgrad society – Sarah P, Danny, Dhaval, Luke. Let no one say the Gaffa didn't spend the postgrad society money wisely. Dr. Wilson for whisky and moving to Scotland else this work might never have been completed. Sean Barry that Clapton house was good value for money. To Fiona for magicking me away to Sardinia and generally being the best buddy one could have. Finally to my family and especially mum and dad – I hope this a little something back!

List of tables

List of figures

Abbreviations

ADNFLE - autosomal dominant nocturnal frontal lobe epilepsy

ADPEAF - autosomal dominant partial epilepsy with auditory features.

ADR – Adverse drug reaction

AED – Anti epileptic drug

aoIGE - adolescent onset IGE

BFIC - benign familial infantile convulsions

BFNC - benign familial neonatal convulsions

BFNIS - benign familial neonatal and infantile seizures

CAE - childhood absence epilepsy

CEPH - Centre d'Etudes du Polymorphisme Humain

CI - Confidence interval

EA - episodic ataxia

EEG - Electroencephalograph

EGMA - epilepsy with grand mal upon awakening

EM – Expectation maximisation

FHM - familial hemiplegic migraine

FS - Febrile seizures

GEFS+ - generalised epilepsy with febrile seizures plus

HP – Haplotype

Hr2 – Haplotype r2 tagging

HS – Hippocampal sclerosis

htSNP – Haplotype tagging SNP

ICEGTC - Intractable childhood epilepsy with generalised tonic-clonic seizures

IGE - Idiopathic generalised epilepsy

ILAE - International League Against Epilepsy

JME - Juvenile myoclonic epilepsy

KB – Kilobase

KW – Kruskal Wallis

LD – Linkage disequilibrium

LOD – logarithm of the odds

MAF – Minor allele frequency

MRD - mean radial degrees

MRI - Magnetic resonance imaging

NCBI – National centre for biotechnology information

OR – Odds ratio

PCR – polymerase chain reaction

PPR - Photoparoxysmal response

SCA6 - spinocerebellar ataxia 6

SMEI - severe myoclonic epilepsy of infancy

SNP – Single nucleotide polymorphism

TDT – transmission disequilibrium test

TLE - Temporal lobe epilepsy

tSNP – Tagging SNP

UTR – Untranslated region

UV – Ultra violet

Chapter 1 Introduction

1.1 An introduction to the background and classification of epilepsy

1.1.1 What is an epileptic seizure? What is epilepsy?

An epileptic *seizure* is a transient experience commonly defined as the unprovoked, coordinated excessive firing of neurons in the brain. A seizure is an isolated event. Epilepsy itself is a *syndrome* defined as a disorder of the brain characterised by an enduring predisposition to generate epileptic seizures [2]. In the context of epilepsy two or more constitute 'enduring', thus an individual who has experienced two or more epileptic seizures is diagnosed as having epilepsy

Epilepsy is the most common serious neurological condition and occurs worldwide, across all age groups. In Europe, the syndrome is observed with a prevalence of 3-8 cases per 1,000 individuals, and an incidence of 43-82 per 100,000 per year [3, 4]. It is estimated that 3.1 million people in Europe today suffer from active epilepsy. Rates of epilepsy are elevated in young children and the elderly resulting in a characteristic U-shaped age specific incidence curve [3]. Societal costs for epilepsy are considerable as severe cases make up a substantial proportion of the epilepsy population.

1.1.2 Introduction to epilepsy classification

Aetiologically epilepsy is heterogeneous and therefore difficult to classify. The world's preeminent association of epilepsy physicians, the International League Against Epilepsy (ILAE) has in its mandate the task of classifying the epilepsies. The ILAE based its most recent classification of epilepsies on two major divisions – seizure *localisation* and *aetiology* [5]. At the most basal level, epilepsy was classified according to seizure *localisation* - the origin of electrical seizure activity in the brain. This can be determined using an electroencephalograph (EEG) - a device that monitors electrical brain activity through a series of electrodes placed on the scalp. Seizures that arise from large cortical areas in both hemispheres of the brain were termed "generalised". On the other hand, seizures that arise from a specific cortical area were termed "focal" in nature.

1.1.3 Classifying epilepsy according to seizure localisation

Generalised and focal epilepsies present with different seizure characteristics. Generalised seizures produce synchronous, high amplitude generalised spike-wave discharges on EEGs. These originate in both hemispheres of the brain and result, typically, but not always, in a tonic-clonic[i] seizure that produces the whole-body convulsions from which the historical term for epilepsy, "falling sickness", originated. Patients suffering from generalised epilepsies may also present with typical absence seizures (characterised by a momentary loss of contact with the surroundings), atypical absences (more prolonged

[i] Tonic-clonic are the most common generalised seizure. The patient loses consciousness, the body stiffens and the patient falls to the floor. The patient then suffers jerking movements. In all the seizure lasts around two minutes.

absences accompanied by myoclonic activity), or myoclonic seizures (characterised by localised twitching). Thirty percent of patients suffer from generalised epilepsy.

Focal epilepsy (also termed localisation-related or partial epilepsy) arises from a specific cortical area (often the hippocampus) and produces a characteristic 'inter-ictal' spike pattern on EEG. Focal seizures involve a discrete part of the body (e.g. a sudden speech arrest, a memory issue, jerking of the head or of a limb) and consciousness might be either preserved or impaired (in contrast to absence seizures where consciousness is always impaired). Seventy percent of patients suffer from focal epilepsy.

A second seizure closely following a first (*status epilipticus*) is potentially life threatening. This occurs in approximately 3% of all epilepsy patients, independently of whether the patient usually suffers from generalised or focal seizures. Patients can experience a focal epileptic seizure followed almost immediately by a generalised seizure, but it does not typically occur the other way around.

1.1.4 Classifying epilepsy according to seizure aetiology

The second major categorisation distinguishes epilepsies according to the underlying cause of the seizures. Epilepsy resulting from a known disorder of the central nervous system is termed 'symptomatic' epilepsy (or secondary epilepsy). 'Idiopathic' epilepsy (or primary epilepsy) has no apparent underlying cause but tends to have a strong family history.

Not all epilepsies can be classified as either 'symptomatic' or 'idiopathic'. A third category, 'cryptogenic' epilepsy, was created to account for epilepsies presumed to be symptomatic, but without known aetiology. This category of cryptogenic epilepsies suffers from imprecision in the definition as it accounts for essentially all epilepsies that cannot be classified either as symptomatic or idiopathic.

The division of epilepsies according to aetiology best reflects the presumed degree of genetic contribution to the different types. For example, the idiopathic epilepsies are thought to be largely or exclusively caused by genetic factors while the symptomatic epilepsies are thought to be mostly non-genetic in origin. The degree to which this presumption has been tested however, is limited.

As our knowledge of the biology of epilepsy develops, it has become increasingly apparent that the current ILAE classification scheme falls well short of representing the underlying pathophysiology of the condition. However the ILAE scheme is informative about prognosis and treatment and is therefore clinically relevant. For example, although we know little about the underlying biology or genetics of the ILAE syndrome 'juvenile myoclonic epilepsy', we do know that patients diagnosed with this syndrome tend to respond well to treatment with Sodium Valproate. Although far from ideal, the ILAE classification scheme remains for the most part the standard used in clinical practice. The ILAE task force is constantly working to address shortcomings in the classification scheme and publishes modifications as consensus is reached.

1.2 Control and treatment of epilepsy

1.2.1 Epilepsy in the community

Morbidity and increased mortality rates aside, epilepsy can be a major burden on the daily lives of sufferers. Epilepsy itself can result in serious social disadvantage and exclusion, for example through educational underachievement, unemployment or an inability to drive. Ignorance of the condition within a community can lead to stigmatisation and discrimination.

Although treatments for epilepsy are available, access is dependent on the localised perception of the condition. In some Third World regions for example epilepsy is thought of as a 'supernatural ailment' rather than a physical condition and is treated by traditional healers. The condition can be exacerbated if sufferers are denied access to medical help. Indeed, due to a combination of social and economic reasons, it has been estimated that over three-quarters of the 40 million sufferers worldwide are not medically treated. It is clear that for epilepsy to be addressed as effectively as possible the challenge is not only in the development of novel treatments but also in overcoming the common misconceptions surrounding the condition. It is only when epilepsy is recognised in a community as a treatable brain disorder that patients gain access to targeted care.

1.2.2 The history of epilepsy treatment

One of the forefathers of epilepsy treatment, Sir Edward Henry Sieveking, commented in his 1861 book that 'there is scarcely a substance in the world capable of passing through the gullet of man that has not at one time or other enjoyed the reputation of being an antiepileptic' [6]. Although our knowledge of antiepileptic compounds has developed since Sieveking's comment we still have a lot to learn about the treatment of epilepsy. The following is a brief description of the history of epilepsy care and management.

Perhaps the first true drug used to treat epilepsy was bromide, which was recognised to have antiepileptic properties in 1857. It was widely used across Europe and the United States through the late nineteenth and early twentieth centuries. With the birth of modern pharmaceuticals two drugs were quickly produced and became the principal treatments for epilepsy up to the 1950s: phenobarbital and diphenyhydantoin (marketed as Phenytoin). In the 1920s the development of EEG by Hans Berger represented a major advance in understanding the aetiology of epilepsy. This novel tool not only revealed the presence of electrical discharges in the brain but, critically, also showed that different seizure types resulted in distinct patterns of brain-related electrical activity. Better diagnosis improved treatment as it became apparent that certain seizure types responded better to certain drugs. For example phenobarbital was highly effective against generalised tonic-clonic seizures, whilst patients suffering from partial or secondarily generalised seizures (a focal seizure that expands during the seizure to become generalised) responded well to diphenylhydantoin.

The FDA in the United States tightened drug licensing laws during the 1960s in an effort to ensure that drugs were not only safe but also effective, inadvertently stagnating epilepsy drug development: since antiepileptic drugs had a comparatively small sales volume few pharmaceutical companies invested in epilepsy as a treatable condition. Between 1960 and 1974 only one new drug, diazepam (Valium), became available in the fight to treat epilepsy. The situation improved little through the late 1970s, although carbamazapine (Tegretol) and valproic acid, both effective antiepileptic drugs, were granted licences.

A major breakthrough came in the 1990s when the National Institute of Neurological Disorders and Stroke in the US announced its antiepileptic drug development program. This unique partnership between government, academia and industry has resulted in the appearance of a new crop of antiepileptic drugs.

Although antiepileptic drugs have immeasurably improved the lives of many epilepsy sufferers, around 30% of patients are resistant to all currently available medication. Clearly one of the principal goals of epilepsy research must be to greatly reduce, or negate altogether, the size of this group.

Resective surgery, the removal of the epileptogenic region of the brain, is one option for refractory patients. Viewed as a 'last resort' treatment, success rates are reasonable with 30-50% of patients achieving seizure freedom, and 50-90% reporting significant improvement in their condition after surgery [7, 8]. However, only 2-5% of individuals with medically refractory epilepsy are suitable for surgery as a number of criteria have to be considered – for example the prognosis is improved if the patient is young, there must be

minimal risk to memory or speech and, critically, there must be concordance between the epileptic lesion (the structural lesion associated with the development of seizures, measured by imaging such MRI) and the epileptic zone (the region of seizure onset and propagation, measured by EEG).

An alternative to drug treatment and resective surgery is the 'ketogenic diet'. This diet, developed in the 1920s yet still poorly understood, is high in fat and low in carbohydrate and appears to force the body to undergo ketosis (the conversion of fats to ketones). Ketones are then used as an energy source in place of glucose. Although the diet has proven effective in some cases (indeed some clinicians regard it as the single most effective treatment of epilepsy) and has recently enjoyed an increase in popularity, it is unpleasant for the patient and can result in serious complications [9].

Although some degree of seizure control is achievable for the majority of epilepsy sufferers receiving medical treatment, there is still much room for improvement. With current practice it often takes long periods of time to find controlling doses and/or drug combinations. Adverse reactions to antiepileptic drugs are commonplace, making drug regimes uncomfortable for the patient. As with many other diseases, drugs for epilepsy have been designed without great knowledge of the underlying biology involved, often by simply testing compounds on animal models of the condition. Such 'traditional' drug design tends to result in treatments with unpleasant side effects that are exacerbated when drugs are used in combination, as is frequently the case. It is not acceptable that a patient be required to weigh up the benefits of seizure control against the drawbacks of drug-related side effects. The hope is that genetics can play a crucial role in improving patient care by

not only helping guide clinicians in dosing but also in highlighting novel biological pathways which might in turn lead to targeted drug design.

1.3 A genetic component to epilepsy? Evidence from Mendelian, monogenic cases of epilepsy

1.3.1 Mendelian forms of epilepsy

That genetics plays a role in at least a subset of epilepsies is beyond doubt given the existence of numerous families who exhibit a Mendelian-like inheritance of the condition. As with other Mendelian diseases, the advent of genetic linkage mapping techniques led to the discovery of a number of genes and mutations causal in these familial forms of epilepsy. It was six years after the discovery of CFTR and cystic fibrosis in 1989 that Steinlen and colleagues reported the first gene that harboured mutations causing epilepsy [10, 11]. Since then, 16 additional Mendelian epilepsy genes have been mapped and shown to harbour hundreds of different disease causing mutations (see Table 1.1 below).

Table 1.1 Genes known to harbour mutations causing epilepsy

Gene name and class	Syndrome	Location	Year	References
Voltage gated ion channels				
Na channel				
SCN2A	BFNIS (MIM 607745), GEFS+ (MIM 604233)	2q23–q24.3	2001/(020	[12] [13, 14]
SCN1B	GEFS+ (MIM 604233)	19q13.1	1998	[15-17]
SCN1A	GEFS+ (MIM 604233), SMEI (607208) ICEGTC	2q24	2000/(01)	[17-29] for review see [30]
Ca channel				
CACNA1A	FHM(MIM 141500), SCA6 (MIM 183086), IGE (MIM 600669)	19q	2001	[31-33]
CACNA1H	CAE (MIM 607682)	16p13.3	2003	[34]
K channel				
KCNQ2	BFNC1 (MIM 125370), BFNC/myokymia (MIM 606437)	20q13.3	1998	[35-40]
KCNQ3	BFNC2 (MIM 121201)	8q24	1998	[37, 41, 42]
Cl channel				
CLCN2	CAE (MIM 607682), EGMA (MIM 607628)	3q26	2003	[43]
Ligand gated ion channels				
ACh receptor				
CHRNA4	ADNFLE1 (MIM 6000513)	20q13.2–q13.3	1995	[10]
CHRNB2	ADENFL3 (MIM 605375)	1p21	2000	[44, 45]
GABAA receptor				
GABRA1	JME (MIM 606904)	5q34	2002	[46]
GABRG2	GEFS+ (MIM 604233), CAE (MIM 607681)	5q34	2001	[47, 48] [49, 50]
GABRD	GEFS+ (MIM 604233)	1p36.3	2004	[51]
Non-ion channel				
LGI1	ADPEAF (MIM 600512)	10q24	2002	[52-54]
EFHC1	JME (MIM 254770)	6p12–p11	2004	[55]
ATP1A2	BFIC(MIM601764)	19p13	2003	[56]
MASS1	Febrile and afebrile seizures	5q14	2002	[57]

BFNIS, benign familial neonatal and infantile seizures; GEFS+, generalised epilepsy with febrile seizures plus; SMEI, severe myoclonic epilepsy of infancy; SCA6, spinocerebellar ataxia 6; IGE, idiopathic generalized epilepsy; CAE, childhood absence epilepsy; EA, episodic ataxia; FHM, familial hemiplegic migraine; BFNC, benign familial neonatal convulsions; EGMA, epilepsy with grand mal upon awakening; ADPEAF, autosomal dominant partial epilepsy with auditory features. ICEGTC, Intractable childhood epilepsy with generalised tonic-clonic seizures; ADNFLE, autosomal dominant nocturnal frontal lobe epilepsy; JME, juvenile myoclonic epilepsy; BFIC, benign familial infantile convulsions

The focus of this thesis is complex forms of epilepsy and so discussion of Mendelian forms will be limited to salient points of relevance, of which there are several. The Mendelian genes (detailed in Table 1.1), for the most part (13 of 17) fall in to the category of either voltage, or ligand gated ion channels, an observation that has led to the epilepsies being proposed as a group of channelopathies. The ongoing discovery of Mendelian genes has led to numerous studies on the function and role of these genes in an attempt to further uncover the biology of the condition. Indeed the pathways illuminated by the Mendelian genes overlapped with known targets for many of the commonly prescribed anticonvulsants. But these discoveries have also shed light on other important traits of epilepsy, namely genetic and phenotypic heterogeneity. For example, we know of five different genes that harbour mutations causing GEFS+ whilst to date mutations in *SCN1A* are known to cause three clinically distinct types of epilepsy. The genetic architecture underlying even the hypothetically simply Mendelian epilepsies appears complex; the task of deciphering the more common forms is clearly testing.

While these discoveries have undoubtedly provided important new insights in to the pathogenesis of epilepsy, it should be noted that Mendelian epilepsies account for only a small minority (1-2%) of all human epilepsies [4, 58] with the vast majority being complex, multifactorial disorders. However it is reasonable to hypothesise that the biological pathways uncovered by the Mendelian successes will also play a role in more complex forms of the disease. As such the Mendelian list detailed in Table 1.1 is a good starting point from which to embark on the study of complex epilepsy. Finally, it should also be noted that several of the apparently Mendelian epilepsies exhibit incomplete penetrance

and variable expressivity within families, suggesting involvement of other genetic and environmental factors.

1.3.2 Evidence from twin and familial aggregation studies

Evidence of a significant genetic contribution to more common forms of epilepsy has been uncovered from familial aggregation and twin studies.

Twin studies provide an invaluable source of information on the relative importance of the contributions of genetic and environmental factors in complex disease [59]. The twin design can disentangle and quantify the contribution of genetics, shared environment and individual-specific environment to observed variation in complex traits. Under the assumption that there are limited differences in the environments experienced by members of monozygotic and dizygotic twin pairs (they both share the same uterus and developmental environment), higher concordance among monozygotic pairs would indicate that genetic factors play an important role in the disease or trait under study. Concordance is typically calculated as: $2C/(2C+D)$, where C is the number of concordant pairs and D is the number of discordant pairs.

Three major twin studies have been carried out to date on epilepsy [60-62]. All three studies were carried out in cohorts of European Caucasian origin and reached similar conclusions. Firstly, the concordance rate for epilepsy was higher in monozygotic (56 – 62%), compared to dizygotic (14 – 21%) twins. This result provides strong evidence for a significant genetic component to complex forms of epilepsy. Secondly, concordance rates

were higher in generalised epilepsies (monozygotes 65 – 82%, dizygotes 12-27%) compared to focal epilepsies (monozygotes 9– 36%, dizygotes 5–10%). Focal epilepsies of a cryptogenic nature showed a higher genetic component than those of a symptomatic nature. Thirdly, up to 94% of concordant monozygotic twins and up to 71% of concordant dizygotic twins presented with the same epileptic syndrome, suggesting the influence of syndrome specific alleles. Intermediate values were seen for febrile seizures (monozygotic 21-58%, dizygotic 14-40%). While these results certainly provide not only evidence for a genetic contribution to epilepsy, but also a guide as to which forms of epilepsy are more 'genetic' than others, they must be qualified by the likelihood that clinicians will be biased when diagnosing siblings known to be twins.

The proportion of the risk associated with developing a disease that can be explained by genetic factors, termed 'heritability', is often used as a guide in the study of complex disease. Kjeldsen et al [62] reported that 80% (CI = 70-88%) of the liability to both seizures and epilepsy was found to be heritable with the remaining 20% attributable to environment-specific factors. However, this result must be qualified by the fact that it was carried out on a relatively small sample (214 twin pairs with seizures and/or epilepsy). It should also be noted that these values were calculated using "standard biometric assumptions" which included no epistasis and no gene-environment interaction. This is likely to be an oversimplified model, and as such the result should be regarded at best as a rough guide.

Studies examining the pattern of inheritance of epilepsy are an additional resource for questioning the genetic component of complex forms of epilepsy. Bianchi et al reported a comprehensive analysis on the pattern of inheritance of epilepsy in a large cohort of 11,000

epilepsy patients, which, when considering first degree relatives included over 50,000 individuals. The authors found 9.1% of epilepsy patients had a family history (affected first degree relative) of epilepsy.

Idiopathic epilepsies had the highest frequency of family history (15.3%) when compared against focal epilepsies of a cryptogenic (6.5%) or symptomatic (6.2%) nature. The prevalence of family history, defined as the ratio between the number of relatives with epilepsy and the total number of first degree relatives, is a useful measure of inheritance. Bianchi et al reported prevalence of family history values of 2.6% for all epilepsy cases, 5.3% for IGE, 1.3% for focal symptomatic and1.8% for focal cryptogenic. Consistent with twin study data, probands with IGE were highly concordant with respect to their relatives' epilepsy types.

Estimating concordance within families is an effective approach in separating shared from specific genetic effects. The procedure involves identifying families with two or more affected individuals and assessing whether the number of families concordant for seizure type exceeds that expected by chance. Through the application of this methodology, Winawer and colleagues showed evidence for distinct genetic effects on absence and myoclonic seizures [63]. The same authors replicated these results in an independent cohort of Australian families [64].

Together the twin and familial concordance as well as the familial aggregation study results support the notion that genetics plays a significant role in the development of common forms of epilepsy. Significantly, the results consistently order the principal epilepsy types in

terms of a genetic contribution as: idiopathic generalised, focal cryptogenic, focal symptomatic and finally other focal epilepsies of unknown aetiology. In addition, the studies were generally in agreement that febrile seizures[ii] carried a significant genetic component that appeared to lie somewhere between that for idiopathic and for focal cryptogenic epilepsies. Finally, the familial concordance studies in particular illustrate a distinct genetic effect on seizure type.

However, informative as these studies are in characterizing a genetic component, they do not explain the underlying architecture of that same component. Is the architecture complex involving many interacting factors of moderate effect or rather simple with a limited number of strong effect loci? As shall be discussed later, these issues are crucial in the design of studies that seek to elucidate the genetics of complex disease.

[ii] A fever accompanied by a seizure observed in the first three months to three years of life. A history of febrile seizures increases the chances of developing epilepsy in later life

1.4 Tools available for mapping complex traits – Availing of both linkage and association

1.4.1 The task in hand

The field of epilepsy genetics is, in parallel with other disease fields, undergoing a transition that could not be more pronounced. Over the past 20 years or so human geneticists have identified more than 1200 genes causing Mendelian diseases, just over one percent of which cause epilepsy. With few exceptions the identifications have been unambiguous and uncontroversial. On the other hand we know next to nothing about the genetics of common forms of disease, including epilepsy, which are influenced by multiple genes in a complicated interaction with the environment. In fact, there are only about 20 different polymorphisms that are generally accepted as risk factors for any form of so-called "complex" disease, none of which influence epilepsy [65-68].

Common disease genetics is clearly a different sort of research enterprise, yet the community of human geneticists has been moulded by its successes in studying Mendelian diseases. It is not clear that the prevailing perspectives and expectations resulting from this experience are relevant and useful in the study of complex traits such as predisposition to developing, and response to treatment for, common forms of epilepsy.

What little we do know about the genetics of common diseases, and of responses to their treatment, is that derived variants at a given gene can lead to increased or decreased

susceptibility, or result in good or bad responses to a given medicine (e.g. *SCN1a* and required dose of carbamazapine and phenytoin [69]). More fundamentally, it seems likely that there is a continuum of effect sizes for variants that influence common diseases, ranging from moderate effects such as that of ApoE4 for Alzheimer's, to apparently more marginal effects such as that of PPARγ and Calpain-10 on Type II diabetes (odds ratios estimated at 3.3, 1.23, and 1.19 respectively) [70-72]. Looked at this way, it is hard to see where one would draw the line to declare a particular gene a "susceptibility" gene for the condition; *ApoE4* is certainly a risk factor for Alzheimer's, though interestingly there are suggestions that it is deleterious only in certain environments [73]. But there is no reason to believe there is a small set of polymorphisms with clear and strong effects, like apoE4, and that the remainder of the polymorphisms have no effect. Rather it would appear more likely that there is a full continuum of effect sizes (and degrees of environmental dependency), and that whether a gene registers as a "susceptibility" gene will depend on the power of detection. Considering current sample sizes and study designs, it would seem that few genes carry polymorphisms of sufficiently strong effect to be consistently identified as risk factors [74]. For larger sample sizes, or designs that consider appropriate environmental interactions, it is probable that a much more sizable minority of the genes in the genome will have polymorphisms in or near them that have some effect on risk in some environments [72]. Currently, there is no way to guess how large this proportion might be. It seems premature to cast the search as one for a small number of susceptibility genes for each condition. It may be more appropriately considered not a search for "disease" genes, but rather an assessment of how polymorphisms in the human genome influence disease risk and drug responses in specific environments and genetic backgrounds.

Around 10 million of the approximately 3 billion bases in our genome are polymorphic in typical human populations (where polymorphic is defined as having a minor allele frequency greater than 1% [75]). Rather than focusing on identifying the few disease polymorphisms out of these 10 million, the challenge is to understand how these 10 million polymorphisms collectively moderate disease risk, and how they influence response to treatment, in their appropriate genetic and environmental backgrounds.

A focus on even this large set of sites assumes that the most important genetic contributions are due to sites that carry classical polymorphisms (i.e. minor allele frequency greater than 1%). Clearly these polymorphic sites contribute primarily to genetic differences among people on average, but there is considerable debate about their importance relative to less common variation in the genetics of common disease predisposition [76].

When the problem is cast this way it is apparent that very large scale and high quality genetic association studies, with all their problems [72], are indispensable in the effort to understand the genetics of common diseases and variable drug reactions.

1.4.2 Linkage as a tool for mapping complex disease

The best approach to decipher complex genetic disorders remains the subject of debate and is a rapidly evolving area in the field of molecular genetics. The traditional and highly successful approach for mapping monogenic traits is linkage analysis. The method focuses on the level of co-segregation between a disease locus and a battery of genetic markers, usually microsatellites. The level of co-segregation is measured using a LOD score, a

representation of the logarithm of the odds of the phenotype in question and a given genetic marker being linked, versus the odds they are unlinked. Parametric linkage analysis requires a precise genetic model, specifying the mode of inheritance, gene frequency and penetrance of the genotype under study. Such parameters are typically calculable for Mendelian diseases, a key factor in the dramatic success of parametric linkage in mapping these disorders. However for common forms of epilepsy and indeed for complex disease in general, knowledge of these parameters is typically the exception rather than the rule.

Non-parametric linkage analysis, being model free, is theoretically more amenable to the scenario presented by epilepsy. Here the focus is on alleles or chromosomal segments that are shared by affected relatives (e.g. sibling pairs) more often than expected. Easing parametric requirements is obviously advantageous, but the method has some important limitations. The statistical power of this approach is reduced compared to parametric linkage and defined candidate regions are usually large. Furthermore, the method assumes that the susceptibility factor is necessary and sufficient to cause disease, rarely the case in complex diseases.

An additional complicating factor that compromises both parametric and non-parametric linkage is that common forms of epilepsy are thought to be caused by the additive effect of numerous alleles at several interacting loci. Any individual variant in itself would typically be insufficient to cause the phenotype in the absence of specific variations at other genetic loci and/or additional environmental factors. Whilst linkage is robust against allelic heterogeneity, it suffers against locus heterogeneity. Considering that the genetic

component to epilepsy is likely to be complex with variants interacting at multiple loci, it is paramount that any mapping method is powered to detect weak to modest effects at any given locus.

Finally, use of linkage analysis for common forms of epilepsy is limited by the availability of large multiplex families. Although such families are more readily available for idiopathic generalised forms of epilepsy, they are very rare for the more difficult to treat focal epilepsies. Linkage strategies are therefore not ideally suited to common epilepsy syndromes; either the optimization of current linkage techniques or the development of alternative strategies are required.

1.4.3 Linkage results to date

Given the availability of suitable families, idiopathic generalised epilepsy has been the focus of linkage studies in epilepsy to date (see Table 1.2 for summary). Several risk loci have been proposed including the EJM2 region on chromosome 15q [77], 8q24 [78] 3q26, 14q23 and 2q36 [79]. However these peaks were not always reproducible in follow up studies [80-83]. The most robust linkage peak, termed EJM1, was proposed by Greenberg and colleagues in 1988 [84] to contain a risk factor for JME. This finding was successfully replicated in several studies [80, 85-87], although many also failed to detect this peak [88-91]. The recently proposed causative variant at this locus is discussed in Chapter 4 [1].

In summary, linkage studies have identified over 10 potential susceptibility loci for IGE, although few of these have stood up to replication. The genes responsible for the reliable

peaks remain to be characterised. An obvious strategy to identify the underlying genes is fine-mapping of the linked interval by means of association analysis. This approach has proven successful for identification of the *BRD2* gene in JME [1] and has now become significantly simplified thanks to the tagging approach.

Table 1.2 Summary of published linkage studies in complex forms of epilepsy

Study name	Year	Type	# families	LOD	Phenotype	Location	Note
Tauer[92]	2005	Genome wide	60	4.3[iii]	PPR	6p21	PPR at 6p21
Nakayama[93]	2004	Genome wide	59	3.68	FS	18p11.2	Proposed IMPA2 gene as causal
Pinto[94]	2005	Genome wide	16	3.47/2.44	PPR	7q32,16p13	
Pinto[90]	2004	6p11/21replication	18	3.17	JME	6p11	Partial replication of [95]
Sander[82]	2003	8p12 replication	176	Negative	IGE	No evidence	Failed replication of [95]
Robinson[96]	2002	scattered	33	5-15.1	CAE	16q12,15q11	Tested candidate genes
Sander[83]	2002	18q21 replication	130	Negative	IGE	No evidence	Failed replication of [95]
Windemuth [81]	2002	5p14 replication	99	Negative	IGE	No evidence	Failed replication of [95]
Bai[97]	2002	6p12-11	31	4.21	JME	Refined peak	Mexican
Durner[95]	2001	Genome wide	91	1.9-5.2	IGE	6p?, 8p?, 18q21	Propose oligogenic model
Obach[89]	2000	6	7	Negative	JME	No evidence	Spanish. Failed replication of [84]
Sander[79]	2000	Genome wide	130	4.19/3.28	IGE	3q26, 14q23	German No subtype analysis
Greenberg[87]	2000	6	85	4.2	JME	6p?	US replication of [84]
Nakayama[98]	2000	Genome wide	40	4.59	FS	5q14	Japanese pop.
Durner[99]	1999	8	88	3.24	aoIGE	8pcentromeric	US. aoIGE-JME
Sander[100]	1999	15q14	57	Negative	JME/CAE	No evidence	German. Replication of [77]
Sander[101]	1998	8q24	38	Negative	IGE/JME	No evidence	German. Failed replication of [14]
Elmslie[77]	1997	Various local	34	4.4	IGE/JME	15q14	UK
Sander[80]	1997	6p25	29	3.27	IGE/JME	6p21	Partial replication of [84]
Sander[102]	1997	15q11-15q13	94	1.4	Mixed IGE	JME 15q11-15q13	German
Liu[91]	1996	Genome wide	22	4.4	JME	6p12-p11	Belize Failed replication of [84]
Sander[103]	1996	5q32-5q35	63	Negative	IGE	No evidence	German
Elmslie[104]	1996	6p	19	Negative	JME	No evidence	UK
Sander[105]	1995	6p21	44	1.92	Mixed IGE	JME 6p21	German. Partial replication of [84]
Zara[78]	1995	Genome wide	10	10.61[iv]	IGE	8q24	US, failed to replicate [84]
Whitehouse[88]	1993	6p	25	Negative	JME	No evidence	UK, failed rep of [84]
Durner [85]	1991	6p	21	4.1	JME	6p21	Refined EJM1 peak
Weissbecker[86]	1991	HLA region	23	3.1	JME	6p21	Replicated[84]
Greenberg[84]	1988	HLA region	68	3.11	JME	6p21	First report of EJM1 peak

PPR=Photoparoxysmal response, FS=febrile seizures, JME=juvenile myoclonic epilepsy, CAE=childhood absence epilepsy, IGE=Idiopathic generalised epilepsy, aoIGE=adolescent onset IGE

[iii] nonparametric MOD linkage score. In a MOD score analysis, the LOD score is maximized not only over the genetic position of the disease locus, but also with respect to the penetrances of the disease gene genotypes and the disease allele frequency.
[iv] This figure represents the chi2 value (p=0.0006) from an extended sib pair analysis in the 10 families studied

1.4.4 Association as a tool for mapping complex disease

An alternative to linkage is the association study design. Association studies compare the frequency of specific alleles, hypothesised to be disease causing, in affected cases and unaffected controls. An allele is claimed to be 'associated' with the disease if its frequency differs between cases and controls by a margin greater than would be predicted by chance. Association studies assess allele frequencies in populations whilst linkage studies track the co-inheritance or otherwise of marker and disease alleles within families.

Association studies have several advantageous traits and has been viewed as strong where linkage is weak (note that the reverse is also the case). As association is population based, large multiplex families, rarely seen in most common forms of epilepsy, are not required and so sample collection is less restricted. Power of detection in association is directly related to the size of the populations under study. Generally speaking it is easier to identify and collect unrelated individuals for use in association studies rather than the deep multiplex pedigrees required for linkage. Cases and controls can be selected to match for environmental factors, such as age or lifestyle, which may also be important in the disease. The candidate region surrounding an associated variant is likely to be significantly smaller than a candidate gene region identified by linkage, as linkage looks at a few generations only, while association studies are based on historic recombination [106]. Finally, in an often cited paper, Rish and Merikangas argued association studies are likely to be more powerful than linkage for detection of weak susceptibility alleles [107]. They showed that

for modest effect sizes affected sibling pair analyses, for example, would require unrealistically large sample sizes.

On the other hand association has some distinct disadvantages when compared to linkage. Population stratification can be a confounding factor in association studies. Stratification occurs when multiple genetically distinct subgroups exist in different proportions in cases and controls. The case and control populations are thus genetically mismatched. Similar to any other confounder, stratification may cause spurious associations (type-1 error) or may mask real associations (type-2 error). The problem increases with population sample size in that slight stratification can become statistically significant (thus driving a spurious association), a concern for the many large scale association studies currently under way. Various methods have been developed to correct for stratification although conservative correction results in reduced power of detection [108-111].

An alternative to having to correct for stratification within a population is to use non-transmitted parental alleles as controls, which by definition cannot be stratified. The haplotype relative risk method [112] and transmission disequilibrium test (TDT) [113] are the commonly used examples of family based association that circumvent the stratification issue. Although these methods are valid and reliable, they have some practical drawbacks. Parents of affected individuals must be available, which may be a problem and as only heterozygous parents can be included for the TDT, some genotype information becomes redundant, resulting in a loss of statistical power to detect genuine allelic association.

The association study design performs poorly in detecting rare disease causing alleles - a problem that is exacerbated when the causal variant is of modest or weak effect or under conditions of locus or allelic heterogeneity. Whilst, as mentioned previously, linkage is somewhat more robust to allelic heterogeneity, both approaches suffer in the face of locus heterogeneity.

Although all these potentially confounding issues must be considered in the design of any association study, the most problematic confounder is type-one error due to multiple testing. If multiple independent tests are assessed within the same experiment, a researcher is required to correct for the total number of tests undertaken. The temptation is for a researcher to test multiple hypotheses without correction but report only those tests that proved 'significant'. Unfortunately this practice has led to a flood of false positive reports in the literature, an issue that has severely hindered progress not only the field of epilepsy genetics, but also in almost every other complex disease area.

The simplest and most conservative approach to correction for multiple testing is the Bonferroni method: $P=\leq\alpha/n$ where P is the experiment wide significance threshold, α is the significance value for each comparison (typically 0.05) and n equals the total number of tests undertaken in the experiment.

The conservative nature of the Bonferroni correction is such that the incidence of type-2 error can increase, in particular when a large number of tests are undertaken. Furthermore, Bonferroni assumes that tests are independent of each other and does not factor in the relationship between genetic variants being tested (i.e. levels of linkage disequilibrium).

Permutation testing considers the relationship between variants within a test (useful for haplotype analysis within a gene) but still does not address the problem of haplotype testing across a panel of genes that are essentially independent (i.e. there is little or no LD between the genes). The best method for correction is currently a matter of debate, but it is unlikely that any single statistical model can account for the variability in LD across the genome [106].

1.4.5 Association results in epilepsy to date

A summary of all association studies published to date for common forms of epilepsy is shown in Addendum Table 8.1. It is striking that although over sixty studies have been published, none have been clearly replicated. For example, take studies examining *Il1-b*, a proinflamatory cytokine hypothesised to play a role in the development of febrile seizures[v]. Kanemoto et al [114] first reported an association with hippocampal sclerosis. Virta et al [115] then reported supportive evidence, but in patients with a history of febrile seizures rather than hippocampal sclerosis. A number of follow up studies failed to replicate either the original association with hippocampal sclerosis, nor that with febrile seizures although one report showed association with refractory partial epilepsy (a loosely related phenotype) [116-120]. The example of *Il1b* also illustrates several other weaknesses that are relevant to other association studies listed in the table. Firstly, studies are all small in size (maximum 112 cases) which effectively restricted power of detection to only strong effects. Secondly, only one variant was considered despite the fact that numerous others (of

[v] See section 3.1.2 for discussion of the hypothesised link between febrile seizures and hippocampal sclerosis

functional and unknown significance) were known. Finally, the phenotypes used were often poorly defined (refractory partial epilepsy is a very loose diagnosis).

The consequence of this poor application of the association study design is that despite the large number of publications, the method has contributed little if anything to our understanding of epilepsy genetics.

Clearly neither the linkage nor association approach is ideally suited to the situation presented by common syndromes of epilepsy. The challenge presented to the field is to refine both methods bearing in mind what we understand of complex disease genetic architecture. In this way our knowledge of epilepsy genetics will expand. This thesis focuses on developing and applying association based methodology for use in epilepsy genetic studies.

1.5 The evolution of map based association

1.5.1 Sequence versus map based association

Traditionally association based approaches in complex disease genetics (and in particular in epilepsy) have focused on one or a small number of putatively functional variants, located in the coding or promoter region of a gene. Whilst this approach is perfectly valid in the micro sense of those variants being tested, very little if anything is learnt about other variants in the gene that may be contributing to the phenotype in question. With this in mind, the method has evolved along two basic routes: termed **sequence based** and **map based**.

Botstein and Risch [66] have recently argued for a sequence based approach which focuses on the identification of all variants in and near exons, and in core promoter regions of genes. They argue that the experience of Mendelian disease dictates these regions to be most important in terms of disease causing variation. All identified variants in these putatively functional genomic regions would then be genotyped in individuals of known phenotype, and correlations assessed in a case-control or related design.

The map based approach on the other hand, refrains from making any assumptions about the precise genomic location of causal variants. Instead, the aim is to identify a set of markers sufficient to represent all other variation in the gene region of interest.

Both approaches have strengths and weaknesses. The principal advantages of the map based approach are economy and the fact that it does not require strong assumptions about where the important variants lie. The principal weaknesses of the map based approach are reduced power of association for any one variant examined and difficulty in accurately representing variants with low minor allele frequency. For these reasons, both approaches could be viewed as more complementary than competing. What follows is a description of the development of map based approaches.

1.5.2 LD mapping

Linkage disequilibrium gene mapping was first used successfully to fine localize Mendelian mutations within genomic regions implicated through linkage analyses, as for example in the cloning of the gene responsible for cystic fibrosis [11].

Two factors greatly increased interest in the application of LD mapping to common diseases. One was the previously mentioned demonstration of Risch and Merikangas [107] of the greater power of association mapping over linkage studies. The other was the rapidly developing understanding of the relationship between polymorphic markers spanning the human genome.

The potential application of LD mapping for common disease generated intense interest in deciphering actual patterns of LD in human populations. Until 2001 this work was predicated in many ways on expectations of gradual, though noisy, decay of LD with physical distance in the genome. It has long been appreciated that certain genes, such as the beta globin gene cluster [121, 122] harboured hotspots of recombination – regions of intense homologous recombination – and also that LD decay was often poorly correlated with distance, either because of uneven recombination in some regions or because of the stochastic effects.

Nevertheless, many empirical studies looked at the average decay in LD with distance over different genomic regions, effectively ignoring the precise pattern of decay within a given region. In addition, models used to generate expectation of the pattern of LD assumed

uniform recombination. In his influential report Leonid Kruglyak established a sort of null model for the expected extent of LD by assuming an idealized human demographic history and assuming homogeneous recombination rates [123]. Under these assumptions Kruglyak [123] showed that usable levels of LD (at that time set at $d^2>0.1$, which corresponds to a $r^2<0.2$, see below) would not extend further than about 3 kb in the human genome, implying that a genome wide map would require more than 1 million markers.

This conclusion was at odds however with empirical descriptions of LD, for example Reich et al [124], looking at multiple genomic regions, published conclusive evidence of much higher levels of LD extending over relatively long sequence tracts. The reasons for the discrepancy between early modelling and empirical patterns depend on both the ways in which human demographics differ from the models assumptions, and also the fact that recombination in the human genome is not uniform. The relative importance of these two factors in shaping patterns of human LD remains unclear.

The tendency to view LD decay with distance as gradual (although noisy) was dramatically changed by a set of papers published in the October 2001 issue of Nature Genetics. The challenge to the "gradual-decay" view had two components. Analysing 103 common SNPs (MAF>5%) across 500kb of chromosome 5q31, Daly et al [125] argued that the pattern was better viewed as discrete, with stretches of sequence showing little or no LD breakdown interspersed by regions of sharper LD breakdown. The stretches of limited haplotype diversity were called blocks, and within them up to 95% of the observed chromosomes were accounted for by 3 or 4 haplotypes [125]. Similar patterns were reported by Johnson et al, who observed a similar block-like pattern of LD among 122

SNPs across 135kb of 9 genes in European populations. Haplotype diversity within blocks was low with a maximum of 6 common haplotypes (frequency >5%) observed within any one block [126].

These publications led directly to the idea that blocks of LD, within which haplotype diversity is limited, is a prevailing characteristic of the genome. This is turn led to the concept that a set of SNPs could represent, or tag, each of the common haplotypes in a given region. These SNPs were first referred to as haplotype tagging SNPs (htSNPs). In an accompanying supplement to the Johnson et al study, David Clayton introduced an approach for selecting htSNPs that focused on the proportion of the haplotype diversity that could be explained by a set of tags [126].

1.5.3 The application of HapMap

The realisation that the tagging concept was indeed valid and applicable led to the development of the HapMap project. This multi-centre collaborative effort, funded by the NIH, was officially started in October 2002 with the aim of cataloguing human variation across global populations of European, Asian and African origin. Within three years the consortium managed to identify and genotype over four million SNPs (>1SNP per KB) in a total of 270 individuals representing the three populations. The completion of the HapMap project offers the opportunity to freely tag any gene or region across the genome in minutes, a task that was previously not only costly but also time consuming.

The HapMap is by no means perfect. The project set out to represent common variation across the genome, a goal it has achieved in spectacular fashion. However, this has resulted in a heavy bias in SNP ascertainment towards common variation. If indeed much of the genetic component to complex traits is explained by the interaction of rare variants, the utility of HapMap will be severely limited. Similarly, it is not known how well other kinds of variants (for example, repetitive elements or insertions and deletions) might be represented by SNP tags. Another complexity not yet adequately addressed concerns how well the four population groups studied in the HapMap project represent variation in other human populations. Nonetheless, the availability of a near complete catalogue of human genetic differences among populations will be invaluable in addressing one of the key questions facing human geneticists today – what role does common variation play in the development and treatment of common disease.

1.6 Conclusions and aims of this thesis

1.6.1 Conclusions

It is apparent that in many ways the field of epilepsy genetics is at a similar stage to other disease fields with the success of Mendelian work contrasting against the difficulties experienced in mapping more complex questions. A genetic component has clearly been illustrated by twin and family studies but deciphering this component is obviously more difficult than originally hoped. This difficulty is further complicated by the heterogeneity of epilepsy – obvious at the phenotypic level and extremely likely at the genotypic (already partially illustrated through Mendelian experience). In addition, epilepsy suffers a distinct

lack of intermediate phenotypes[vi] so abundantly available in for example cardiovascular or cancer related traits. The task ahead is clearly difficult. Pharmacogentic questions however tend to be much more straightforward in terms of biology and phenotype. Pathways of action and metabolism are known for the vast majority of epilepsy drugs. Intermediate phenotypes, for example drug plasma levels, are readily available.

Recent progress in both genetic mapping techniques and genotyping technology hold huge promise for epilepsy genetics. The challenge is to learn from past experiences and apply this knowledge to future project design.

1.6.2 Aims of this thesis

At the initiation of this thesis work in September 2003, the field of medical genetics was still high on the wave generated by Mendelian mapping successes. Epilepsy research groups seeking to translate this success to more common forms of the condition published numerous reports of association with hypothesised risk variation. At the same time, population geneticists were, through the illumination of LD patterning, developing novel methods of disease mapping. It was thus both an exciting and challenging time to begin this work. Exciting in that huge leaps were being made in mapping methods yet also challenging in how to interpret the published reports of association and apply the new mapping methods.

[vi] In intermediate phenotype is one closely related to the gene (i.e protein level) as opposed to an end-point phenotype (e.g. the disease).

In the first experimental chapter of this thesis seven previous reports of association between genetic variation and common forms of focal epilepsy are examined. The goal here was to solidify our knowledge of epilepsy genetics by critically examining for robustness, positive reports in the literature.

The second experimental chapter extends the theme of replication from focal to generalised epilepsy. This chapter examines the previously reported effect of variation in *BRD2* on the development of juvenile myoclonic epilepsy [1]. Again the goal here was to critically assess this reported association with a view to making a statement on the robustness of the association.

In the third experimental chapter, focus shifts from predisposition to treatment. Here an assessment is made of the potential use of genetic variation in guiding the prescribing of an effective antiepileptic drug whose use is severely limited by a serious adverse drug reaction. In this chapter focus also shifts from sequence, to map based methods of association. Data from the HapMap project is used to select tagging SNPs for a panel of candidate genes.

Finally, in the fourth experimental chapter, a study design is presented that attempts to address many of the difficulties illustrated and discussed in the first three experimental chapters. This study design represents the first truly large scale study of epilepsy genetics both in terms of the number of genes and variants considered and the size of the study population.

Chapter 2 Methods

2.1 Sample collection and storage

2.1.1 Patient recruitment

Patients were recruited from two independent epilepsy referral centers: the National Hospital for Neurology and Neurosurgery, Queen's Square, London, United Kingdom (incorporating the National Society for Epilepsy at Chalfont, London) and the epilepsy clinic at Beaumont Hospital, Dublin, Ireland.

All patients with a diagnosis of a common, non-Mendelian form of epilepsy were eligible for inclusion in the study. Patients were provided with information on the project and those interested provided a blood sample for DNA extraction after signing the informed consent form.

The study protocol was approved by Joint Research Ethics Committee of the National Hospital for Neurology and Neurosurgery and Institute of Neurology (JREC 00/N081 and 01/N088) and by the ethics panels of University College London and Dublin. As patient recruitment is ongoing, the number of DNA samples is continuously increasing. As of December 2005, the collection had passed 1800 samples. 90% of samples collected are of North Western European origin, the remaining 10% are from a variety of non-European ethnic groups.

During the course of the study, a patient panel of North Western European origin on which clinical data had been entered into a phenotype database at that time was assembled. The total number of samples in this panel was 673.

2.1.2 Patient phenotyping and databasing

Clinical details of participating patients were stored in two separate databases, stored at the Institute for Neurology, Queen's Square, London, and at Beaumont Hospital, Dublin. These databases cover in detail a wide variety of clinical phenotypes including sydromic diagnosis (according to ILAE classification), seizure diagnosis, seizure frequency and antiepileptic drug history.

2.1.3 Control cohort

Controls for this study were assembled from the National Twin Research Unit at Guy's and St Thomas' Hospitals, London, United Kingdom. To be considered for the control cohort, individuals must have been of British ancestry and unrelated to any other member in the selected cohort. Only one member of each sibship was included.

The use of these samples as controls is covered by ethics approval held at the Twin Research Unit, Guy's and St Thomas' Hospitals, London, United Kingdom.

2.1.4 DNA extraction from whole blood

DNA extraction from whole blood was carried using an Autogen extractor (AutoGenPrep® NA-2000).

Step 1: Obtain nuclei pellet:

Each blood sample (20ml) was split in to 2x50ml Falcon tubes (i.e. 10ml in each tube). To these tubes was added 30ml of a red blood cell lysis solution[vii]. The samples were inverted several times and left at room temperature for 5 minutes. In order to obtain a nuclei pellet, the samples were then spun at 2,000 x g for 20 minutes. The supernatant was discarded.

Step 2: Wash nuclei pellet:

The pellet was washed with 10ml PBS[vii] and centrifuged again at 2,000 x g for 10 minutes. The supernatant was discarded and the pellet resuspended in 5ml PBS[vii]. The samples were again spun at 2,000 x g for 10 minutes and the supernatant discarded.

Step 3: Digest/lyse pellet:

To digest protein and lyse the white blood cell nuclei 100ul (100ug/ml) proteinase K (Sigma-Aldrich P2308) and 1ml 10% SDS was added to each sample. The samples were then incubated overnight at 55ºC.

Step 4: Precipitate DNA:

On completion of digestion the samples were transferred to the Autogen machine which performed an automated phenol/choloroform extraction.

[vii] See Addenda for details of solutions

Step 5: Wash DNA

The autogen 2000 uses 70% ethanol to wash the DNA pellet resulting from the phenol/choloroform extraction.

Step 6: Resuspend DNA

Extracted samples were resuspended in 500ul of TBE (Sigma-Alrich 93306) and stored at -70ºC.

2.1.5 DNA quantitation and standardisation

Traditionally DNA is quantified by measuring absorbance at 260nm using a spectrophotometer. Although this method is highly accurate, it is not particularly sensitive and requires relatively large amounts of DNA for a read. Instead we availed of the Picogreen® method. Briefly, picogreen® is a fluorescent dye that undergoes a dramatic fluorescence enhancement upon binding with DNA. This fluorescence can be measure using a microplate fluorometer. The principal advantages of the picogreen® method are increased sensitivity (10,000 times more sensitive than ultra violet absorbance methods), the linear nature of the fluorescence over three orders of magnitude and robustness to protein contamination.

The "Quant-iT™ PicoGreen® dsDNA Assay Kit" from Molecular Probes (part # P7589) was used for DNA quantitation. Manufacturer's instructions were followed.

To 19.9ml TE[viii] solution was added 100 µl of Picogreen®. A standard curve was set up as follows. To 45 µl of the TE/Picogreen® solution was added 5 µl of standard concentration DNA (100 µg/ µl, delivered with Quant-iT™ PicoGreen® kit). A standard curve was constructed by filling the first five wells of a 96 well plate with the following volumes of the TE/Picogreen mix (solution A in Table 2.1) and the TE/Picogreen/DNA mix (solution B in Table 2.1).

Table 2.1 Dilution factors for picogreen quantitation

Position	*Solution A (µl)*	*Solution B (µl)*	*Concentration (ng/ul)*
A1	76	24	2.4
B1	88	12	1.2
C1	94	6	0.6
D1	97	3	0.3
E1	98.5	1.5	0.15

The remaining 91 wells on the plate were filled with 99ul of solution A. To each well was added 1ul of DNA of unknown concentration. Each well was thoroughly mixed and allowed to stand for 5 minutes. The fluorescence of each well was measure using a fluorometer (FluoroCount™, Packard) and the concentration of each unknown sample calculated from the standard curve.

Once the concentration of each sample was known, it was possible to create 96 well "working" plates at a standardised concentration much lower than that of the stock tubes. In general we aimed to standardise working plates at a concentration of 10ng/ul.

[viii] See Addenda for details of solutions

2.2 DNA sequencing

2.2.1 PCR amplification

Amplification of genomic DNA by PCR was performed using Applied Biosystems 9700 thermal cyclers. The standard reaction protocol is detailed below. Any changes from this protocol are detailed in the relevant chapters.

Standard PCR reaction mixture:

10ng DNA, 2.5 mM $MgCl_2$, 0.5 uM of each primer, 0.2mM of each dNTP, 1.6ul 10X Qiagen PCR buffer and 0.25 units of Qiagen HotStartTaq polymerase. The mixture is made up to 10µl using H_2O.

Standard PCR cycling conditions:

Initial denaturation at 95ºC for 15minutes.

35 cycles of:

Denaturation at 94ºC for 30 seconds,

Primer annealing at 60ºC for 30 seconds,

Primer extension at 72ºC for 30 seconds,

Final extension at 72ºC for 10 minutes

A negative control (H_2O instead of DNA) was included in all experiments.

2.2.2 Agarose gel electrophoresis

Before proceeding to the PCR cleanup and sequencing stages, the outcome of the PCR reaction was determined by running products on agarose gels. One µl Orange G® loading buffer (Trevigen) was added to 2 µl of each PCR product. DNA fragments were loaded on a 1% agarose gel containing 2.5 µg/25 µl ethidium bromide. A 100bp Hyperladder® (Midwest Scientific) size marker was added alongside the samples to allow identification of DNA fragments size. The gels were cast in 1x TBE buffer and run at a constant voltage of 100V. Gels were viewed over an UV light box.

2.2.3 PCR clean up for sequencing reaction

In order to remove single stranded DNA (primers) and unincorporated dinulceotides from each reacted PCR making them suitable for a sequencing reaction, the following protocol was applied.

An equal volume of MicroClean® (Microzone Ltd.) was added to each cycled PCR reaction. The sample was left at room temperature for 15 minutes then spun at 3750 RPM for 60 minutes. The supernatant was discarded. 150µl of ethanol was added to each sample and the plate spun at 3750 RPM for 10 minutes and the supernatant discarded. The samples were allowed to air dry for 30 minutes and room temperature. Once drying was complete, 5µl of water was added to each well.

2.2.4 Sequencing Reaction

Sequencing was performed using the dye terminator cycle sequencing technology as incorporated in the Applied Biosystems BigDye^{TM} reaction. Cycle sequencing utilizes successive rounds of denaturation, annealing and extension in a thermocycler to create a linear amplification of extension products. With dye terminator labelling, each of the four dideoxy terminators is tagged with a different fluorescent dye. The growing chain is simultaneously terminated and labelled with the dideoxy terminator dye that corresponds to the terminal base. All reactions are carried out in the same tube, all four colours are

assessed within the same capillary and false stops go undetected because no dye is attached. The ABI Prism Terminator Reaction Kit® contains the following reagents: didexoynucleotide terminators, each based type labelled with different fluorescent dyes, deoxynucleotides, AmpliTaq DNA polymerase FS (with thermally stable pyrophosphatase), MgCl2 and Tris-HCl buffer.

Standard sequencing reaction mix:

1 µl cleaned PCR product, 1 µl Ready Reaction mix, 2 µl 5x buffer, 3.2 µM primer. The mix was made up to a volume of 10µl with distilled H_2O.

Standard sequencing cycling conditions:

Initial denaturation at 96ºC for 10 seconds
25 cycles of:
Primer annealing at 50ºC for 5 seconds
Primer extension at 60ºC for 4 minutes

2.2.5 Sequencing reaction clean up

40µl of 75% ethanol was added to each sequencing reaction. The sample was mixed thoroughly and left at room temperature for 15 minutes. The plate was then centrifuged at 3750 RPM for 60 minutes and the supernatant discarded. Each pellet was then washed

with 150µl of 70% ethanol. The pellets were then allowed to air dry for 30 minutes at room temperature.

10µl of hi-dye formamide was added to each cleaned sequence reaction before loading on either an ABI Prism® 3100 or 3700 automated capillary electrophoresis DNA sequencer. The output traces were analyzed using the Sequencher® software.

2.3 Genotyping by TaqMan

Whilst genotyping by sequencing is reliable and accurate, it is labor intensive and costly, in particular when multiple SNPs are being genotyped in large cohorts. ABI TaqMan technology is an alternative to sequencing and a big improvement in terms of cost and labour. TaqMan is a PCR based system on quenching one of two allele specific fluorescent dyes (FAM and VIC). A TaqMan assays contains locus specific primers used to target a typical PCR reaction to the region containing the polymorphism of interest. However, the assay also contains two allele specific probes, each composed of a short stretch (ca. 20-25 bases) of oligonucleotides labelled with a specific fluorescent dye (FAM for one allele, VIC for the other). On the 5' terminus of each probe is a reporter dye and on the 3' terminus is a quenching dye. When the probe is intact, energy transfer occurs between the two dyes and emission from the reporter is quenched by the quencher. During the extension phase of PCR, the probe is cleaved by 5' nuclease activity of Taq polymerase thereby releasing the reporter from the oligonucleotide-quencher and producing an increase in reporter emission intensity. Fluorescence is detected using a CCD camera. If only one allele is present (i.e.

the individual is homozygous for the SNP in question) then only that allele specific probe fluoresces. If on the other hand the individual is heterozygous (i.e. a copy of each allele present) both probes will fluoresce. Fluorescence can be plotted in two dimensions, each axis representing a fluorescent dye (or allele). Measurement takes place directly in the well without post-PCR processing. This reduces the time of analyses, minimizes the risk of error, reduces the risk of cross-contamination and eliminates the labor and supply costs of post-PCR steps. Taqman assays were ordered directly from Applied Biosystems as either 'Assay-on-Demand' or 'Assay-by-Design' assays depending on availability. All genotyping runs were checked for Hardy-Weinberg equilibrium (the author notes that violation of Hardy-Weinberg can sometimes be observed in cases). What follows is a description of the standard TaqMan sequencing protocol used for the work described in this thesis.

Standard TaqMan reaction mix:

1 µl (10ng) DNA, 0.125 µl TaqMan assay, 2.5 µl 2X TaqMan Universal Master Mix. The mix was made up to a volume of 5µl with distilled H_2O.

Standard cycling conditions:

Initial denaturation at 95ºC for 10 seconds

60 cycles of:

Denaturation at 92ºC for 15 seconds

Primer extension at 60ºC for 1 minute

The assays were run on an ABI Prism® 7900HT machine and scored with SDS sequence Detection System) software.

2.4 Genotyping by digestion using restriction enzymes

The restriction digest method was used to generate genotype data for rs1044396 a SNP discussed in section 3.2.7. The protocol is described below.

The following primers were used in conjunction with the standard PCR protocol described in section 2.2.1 to amplify the region containing rs1044396.

F' 5'-3' CCTGGCCTCTCGCAACAC

R' 5'-3' TTGGTGCTGCGGGTCTTG

The resulting PCR product was digested with *Hha1* according to manufacturers directions (New England Biolabs, product number R0139S). Digested PCR products were analyzed on a 2% agarose gel by running at 100 volts for 45 minutes along size a 100bp size standard. Gels were scored on a UV light box. Genotype was scored according to product size (enzyme cuts with 'C' allele, does not with 'T'). Suspected heterozygotes of the PCR products were sequenced, for avoiding partial digestion mistaken for a heterozygote.

2.5 Genotyping of a variable number tandem repeat

The 68bp variable number tandem repeat in the promoter of the *PDYN* gene is discussed in section 3.2.3. The protocol used to genotype this locus is described below.

10ng DNA was added to a PCR reaction mix consisting of 20 mM Tris-HCl (pH 8.8), 50 mM KCl, 1.5 mM MgCl2, deoxynucleotide triphosphates each at 0.4 mM, 50 pmol of each primer (sequence described below), and 1.5 U of Taq polymerase. The total reaction volume was 30 µl.

Cycling conditions were as follows:

Initial denaturation at 94°C for 10 minutes,

30 cycles of
Denaturation at 94°C for 30 seconds,
Annealing at 62°C for 30 seconds
Extension at 72°C for 45 seconds

Final extension at 72°C for 10 minutes

F' 5'-3' AGCAATCAGAGGTTGAAGTTGGCAGC

R' 5'-3' GCACCAGGCGGTTAGGTAGAGTTGTC

The amplification products were resolved by running (alongside a 100bp size standard) for 45 minutes at 100 volts on a 2.5% agarose gel stained with ethidium bromide. Products were scored by viewing under a light box. Four common alleles, representing one (386bp), two (454bp), three (522bp) or four (590bp) copies of the variable number tandem repeat exist in the general population. To allow comparison with previous studies we grouped

alleles as low expression "L" (i.e., those with 1 or 2 repeats) and high expression "H" (i.e., those with 3 or 4 repeats).

2.6 Screening for variation across BRD2 *promoter and UTR regions*

The following primers pairs were reacted using standard protocol (described in section 2.2.1) to produce template for sequencing. To increase the sensitivity of SNP discovery, forward and reverse primers were reacted using the standard sequencing protocol (described in section 2.2.4). The label of each primer pair corresponds to that used in figure 4.1)

Promo S1 F'	5'-3'	TGCTACTAATGTAGGTATGGGTTCC
Promo S1 R'	5'-3'	AGAATAGAGGCCAGAATGTACCC
Promo S2 F'	5'-3'	AGGAACCTTTATGCTGTTTTCC
Promo S2 R'	5'-3'	CACAAATGGGATTACATCAAGC
UTR1 F'	5'-3'	CCATGCTGAACTCGTATGGA
UTR1 R'	5'-3'	TTGGTTAAGCACTTGACTGACA
UTR2 S1 F'	5'-3'	GGGCATGGGAACGTTAGAT
UTR2 S1 R'	5'-3'	AATTGGCTGAGCTGTGGTC
UTR2 S2 F'	5'-3'	CTTCCGCACCTCTTCCAAC
UTR2 S2 R'	5'-3'	CAGAATCCTCCAGCTCGTTC
UTR3 F'	5'-3'	CTTCCGCACCTCTTCCAAC
UTR3 R'	5'-3'	GATCTCAGACACCGTCGTCA

2.7 Quantitative mRNA work

Quantitative mRNA analysis was carried out as part of *BRD2* work described in section 4.2.7.

2.7.1 Samples used in quantitative mRNA work

The experiment was carried out in two separate, fully anonymised cohorts. The first cohort consisted of samples of temporal neocortical brain tissue from anterior temporal lobectomies obtained at therapeutic surgery from 40 patients with drug-resistant epilepsy and hippocampal sclerosis. The brain tissue was frozen in liquid nitrogen soon after resection. The second cohort also consisted of temporal neocortical tissue, but collected from post mortem brain samples of 23 patients who had had Parkinson's disease selected from a wider sample donated to the National Hospital for Neurology and Neurosurgery Parkinson's disease brain bank, based on the candidate causal variant rs3918149 genotype.

2.7.2 Preparation of mRNA

All tissue was stored at -80°C until extraction. Genomic DNA for samples used in the mRNA analysis was extracted from approximately 25 mg of the same brain tissue used for mRNA analysis using the Wizard Genomic DNA Purification Kit (supplied by Promega, catalogue number A1120) according to the manufacturer's conditions.

Total RNA was isolated from approximately 30 mg tissue using the RNeasy Lipid Tissue Purification Kit (supplied by Qiagen, catalogue number 74804), according to manufacturer's conditions. The RNA was quantified spectrophotometrically at 260 nm, and 1 μg RNA of each sample was reverse-transcribed to cDNA using the High Capacity cDNA Synthesis Kit (supplied by Applied Biosystems, catalogue number 4322171), according to manufacturer's conditions.

2.7.3 Quantitation of mRNA

Quantitative PCR was carried out using an Applied Biosystems 7300 real-time PCR system on a volume of cDNA corresponding to 10 ng starting RNA, using Assay on Demand assays for gene expression (supplied by Applied Biosystems). Quantitation of mRNA expression was obtained using the standard curve method. Each sample was run in triplicate, and the signal for each gene was normalised over the endogenous control *ACTB*, as outlined in User Bulletin 2 (Applied Biosystems, Catalogue No. P/N4303859). *ACTB*

was chosen as the endogenous control as it has been previously shown not to be upregulated in epileptogenic tissue [127]. The data for each sample were averaged over two independent replicates of the same experiment (from mRNA extraction onwards). The fit of the data to a normal distribution was assessed using a Kolmogorov-Smirnov (K-S) test for the composite hypothesis of normality. Correlation between different experiments was assessed using regression analysis. Association between mRNA expression and genotypes at the *BRD2* candidate causal variant rs3918149 was evaluated using a Kruskal-Wallis (K-W) rank test. All analyses were implemented in the software package R [128].

2.8 Selection and assessment of tagging SNPs

2.8.1 Tagging methodology and related issues

The term "tagging" was introduced by Johnson and colleagues who suggested the method to capture the variance in commonly observed haplotypes across a gene or region [126]. But tagging common haplotypes is only one of many possible ways to select a subset of SNPs that retain as much information as possible about the other SNPs. Broadly speaking the approaches that have been evaluated can be divided into 2 groups [129], those based on maximising the haplotype diversity present in the tagging set compared to the tagged set (*diversity based*) and those based on establishing as high an association as possible between the "tagging" and "tagged" set (*association based*). To avoid the close

identification with haplotype diversity in the selection of tags, some have suggested that tags be referred to as tSNPs as opposed to htSNPs (e.g. [129]).

The primary motivation for tSNP selection is their application in LD based gene mapping. For this reason, a tSNP selection criterion focused on the r^2 measure of LD seems logical because it allows quantification of the loss of power in typing the tSNPs instead of all the SNPs. Pritchard and Prezeworski showed that for two biallelic loci, power scales with r^2, such that typing the associated marker with n/r^2 individuals would have approximately the same power as n individuals in which the causative variant itself was typed, where r^2 is the association between the two variants [130]. This finding has been extended by Chapman et al [131] to include generalised r^2, including haplotype r^2 (see section 2.8.2 below).

2.8.2 Multi marker vs. pairwise approaches

There still remains the question of how to define the r^2 value. Relying on pairwise measures is straightforward, but may be inefficient. This is because pairs of SNPs will only have high pairwise association when their minor allele frequencies are very closely matched, thus meaning that SNPs that exhibit a full range of frequencies will need to be selected as tags. This can be overcome if combinations of the tSNPs are used to predict the other SNPs. Two principal approaches have been discussed in the literature; the haplotype r^2 method (Weale et al 2003; Goldstein et al 2003; Chapman et al 2003 and Clayton website: http://www-gene.cimr.cam. ac.uk/clayton/software) and the allelic r^2

method (Chapman 2003 and Clayton website: http://www-gene.cimr.cam.ac.uk/clayton/software).

Haplotype r^2 is defined as the proportion of variance in a "tagged" SNP of interest that is explained by an Analysis of Variance based on the *G* haplotypes formed by the set of tSNPs. :

$Yi = x_{i1}b_1 + x_{i2}b_2 + ... + x_{iG}b_G$

Where Yi is the predicted state of the tagged SNP of interest on the ith chromosome, $x_{i1}...x_{iG}$ are indicator variables for the *G* haplotypes and $b_1...b_G$ are coefficients estimated by standard least squares from the observed data.

Allelic r^2 only differs in that the tSNP alleles, as opposed to the haplotypes defined by the tSNPs as the indicator variables in the above regression equation. Although less efficient than haplotype r^2, allelic r^2 is attractive in that haplotype inference, a complicating factor when dealing with large genomic regions, is not required.

The combinatory approaches (haplotype and allelic r^2) are clearly more efficient in the sense of requiring fewer tags because they rely on combinations of haplotypes or alleles generated by tagging SNPs to predict the state of tagged SNPs. These combinations are identified by selecting the appropriate coefficients in a linear regression. At the present time, it is hard to predict which approaches to tSNP selection will prove the most useful in practise although it is clear that the LD properties of the region in question will influence any decision.

2.8.3 Block-based and block-free selection of tags

While the discovery of the block-like nature of LD and its effect on haplotype distribution inspired the idea of tags for given haplotypes (Johnson et al 2001), the use of tSNPs in no way depends on blocks of LD. Indeed, even if there is such a block structure, it is not apparent that tag selection should make reference to blocks. As noted in section 2.8.1 above, the early suggestions for the definition of htSNPs did not address this issue directly. More recently, however, it has been argued that block based identification of tags will always be less efficient (sometimes considerably) than methods that select across block boundaries. This is because tagging within blocks limits the effective range of a set of tags, and means that cross-block associations cannot be exploited (Goldstein et al 2003). For this reason, we advocate the selection of tSNPs across large contiguous sequence stretches, independently of any underlying block structure in the region. Even so, some questions remain in this approach. For example, computational issues make it difficult to select across very large regions without some sort of subdivision. In addition, selecting across large regions may result in a set of tSNPs that are not optimised for specific subregions, as for example a candidate gene (Goldstein et al 2003). These are just some of the issues that will need to be addressed in order to develop appropriate, efficient strategies for genome-wide tSNP selection.

2.8.4 Selecting tagging SNPs

All tags discussed in this thesis were selected using the haplotype r^2 method (as described in section 2.8.2 above) applied to data from the HapMap project. The HapMap project website (www.hapmap.org) offers a user-friendly interface allowing selection of any region of the genome of interest in manner similar to previously established genome browsers (UCSC, NCBI etc). Genotypic data is generated from 30 trios of a specific ethnic origin and can downloaded either as a large batch (i.e. all HapMap data or a chromosome), or for any defined region of the genome.

Haplotype inference is an essential prerequisite to tag selection using the haplotype r^2 method. The EM-related inference package PLEM (available from: http://www.people.fas.harvard.edu/~junliu/plem/click.html) is well equipped for this task. PLEM implements a partition-ligation approach (i.e. partitions the data in to subsets, infers haplotypes then ligates to neighbouring region etc) and is thus favoured over standard EM packages as it greatly helps in the analysis of larger datasets (i.e. haplotype inference across many SNPs), critically important as we move towards genome wide tags.

The TagIT package (available from http://www.genome.duke.edu/pressevents/lectures/pg2_050607), implemented in the Matlab programming environment was used for tag selection. TagIT allows tag selection using a wide range of variations of either diversity or association based methods (discussed briefly in section 2.8.2 and summarised in TagIT user guide). TagIT also offers

functions allowing the calculation, and display, of various measures of LD that aid in judging the suitability of the technique for a give gene or region.

The haplotype r^2 method focuses on the coefficient of determination (measured as haplotype r^2 or Hr^2) from a regression model involving the tagging and tagged SNPs. This provides a formal measure as to how well each individual SNP is being captured by the set of tSNPs. Increasing the number of tags improves performance. However, whilst improvement in tagging performance asymptotes, the cost of genotyping an increasing number of tags does not. A compromise is thus required and a consensus seems to be emerging that a Hr^2 value above 0.8 is acceptable [129, 132]. This implies that increasing the sample size to $n/0.8$, would be comparable to exhaustive typing of the hypothesised causal variant captured by a tag with a Hr^2 of 0.8. With this in mind, a Hr^2 value of 0.8 for any single observed SNP was set as the lower threshold for tSNP design in this thesis.

2.8.5 Evaluating the ability of tagging SNPs to detect unseen variation

The tagging method, when applied to incomplete data (such as that generated by the HapMap project), raises the inherent question of how well unknown variation is being captured. In order to address this crucial point the SNP dropping procedure was developed [133]. The approach entails taking the global set of known SNPs and for each SNP *'i'*, dropping it from the analysis in turn. For each reduced set of *N-1* SNPs new tags are selected, and their ability to represent the dropped SNP *'i'* is assessed (by Hr^2). In this way, a table of Hr^2 values for dropped SNPs is built, providing a statistical estimate of how well

the tSNPs can represent SNPs that are not observed (for example SNPs which are not yet discovered) in the region.

The trade off between tagging performance and genotyping cost necessitates the definition of a threshold for SNP dropping performance. At present no consensus exists as to what this value should be. Indeed it is difficult to define as any threshold would depend on the context of the region. However, a value somewhere between 0.5-0.7 as the lower threshold would seem acceptable.

When dealing with incomplete data such as that provided by HapMap, dips in SNP dropping performance can indicate regions of insufficient SNP coverage. If however, the region indicated by the dip in SNP dropping performance does not correspond to a drop in LD, it then becomes very difficult to nominate a region in which to focus resequencing efforts as genealogically informative loci could be located anywhere across the LD block.

2.8.6 Example of tSNP design using *SCN8a*

Following is a detailed illustration of the tagging SNP methodology employed in this thesis using the *SCN8A* gene as an example[ix]. The first step is to download and format the genotype data from the HapMap project website. An excel spreadsheet provides an efficient platform for data formatting, the principal steps being ordering samples as trios, and coding genotypes according to haplotype inference program used. Once haplotypes have been inferred, the researcher can proceed to tSNP selection.

[ix] See sections 2.8.4 and 2.8.5 for description of methodology

As described in section 2.8.4, tagging a region requires the examination of Hr^2 values for tSNP sets of increasing size in order that size that satisfies a Hr^2 threshold of 0.8. The blue trace in Figures 2.1 and 2.2 illustrate the effect of increasing tSNP set size on *SCN8a.* Whilst three tags provide reasonable performance, an increase in the number of tags from three to six brings all loci above the threshold Hr^2 of 0.8.

Figure 2.1 *SCN8a* tagged with three tSNPs

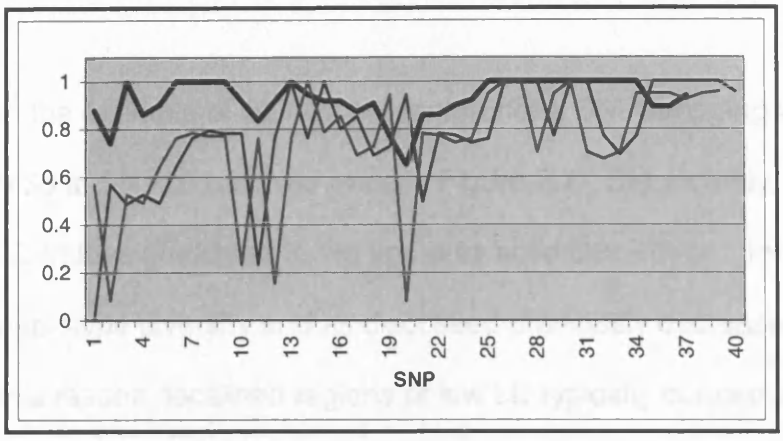
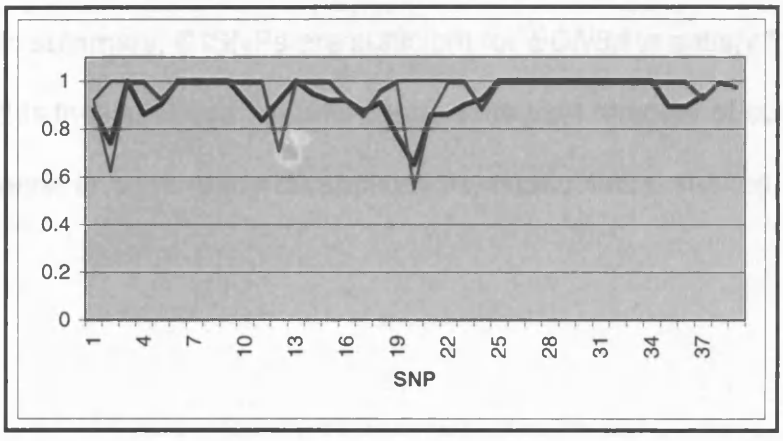


Figure 2.2 *SCN8a* tagged with six tSNPs



Legend for Figures 2.1. and 2.2
The X axis is Hr^2 values. The Y axis is SNP position. Solid black traces represent LD values calculated as a sliding window of D' values from pairwise comparisons between 3 consecutive SNPs. Blue traces are Hr^2 values. Hr^2 is the coefficient of determination value for that SNP in a linear regression between haplotypes generated by tSNPs and the global SNP set (see section 2.8.4).
Red traces are SNP dropping Hr^2 values. Each value represents the coefficient of determination value for SNP *n* in a linear regression between haplotypes generated by tSNPs (designed on the global SNP set minus SNP *n*) and the global SNP set (see section 2.8.5).

The result of SNP dropping procedures using three and six tags is illustrated by the red trace in figures 2.1 and 2.2. Dips in Hr^2 values returned from SNP dropping id indicative of poor tag performance in predicting unknown variation. Given the poor performance of tagging with three tSNPs, it can be expected that SNP dropping would, in turn, perform poorly. Indeed this can be seen by comparing the red traces in figures 2.1 and 2.2. Increasing to six tSNPs returned an average Hr^2 value of 0.95 from SNP dropping indicating a significant improvement in detection of unseen variation compared to three tSNPs.

In the example of *SCN8a* presented here, SNP dropping with six tSNPs returns a value of 0.56 at SNP20 (see red trace in Figure 2.2). Significantly, this dip corresponds to a drop in LD values (indicated in the figure as solid black trace). Indeed, decay of LD generates haplotype diversity and as discussed previously decreases the performance of tagging. For this reason, localised regions of low LD typically correspond with dip in SNP dropping performance.

In summary, 6 tSNPs are sufficient for *SCN8a* to satisfy the selection criteria defined for this thesis. These variants capture the vast majority of common variation in the *SCN8a* gene and are ready for application in association studies.

2.9 Statistical methods used in association analysis

2.9.1 Allelic and genotypic association

For single SNP analysis we assessed the significance of genotypic and allelic contingency tables using Pearson's χ^2 distribution. For tables with insufficient cell counts (i.e. <5) we used an exact probability test as implemented in the program, RxC (available at http://bioweb.usu.edu/mpmbio/rxc.asp).

2.9.2 Haplotypic association

We assessed the distribution of haplotypes across case and control groups by calculating score statistics using the package haplo.score [18] in R [19]. Briefly, this method uses generalised linear models to generate score statistics which can be used to examine the correlation between expectation-maximisation generated haplotypes and the trait of interest. We coded cases and controls as a binomial trait. When generating score statistics, we only considered haplotypes observed at a frequency of 1% or greater. The likelihood of spurious inference by the EM algorithm increases with lower frequency haplotypes.

2.9.3 Power calculations for association studies

What follows is a description of power calculations as applied in Chapters 3 and 4.

In order to illustrate the expected power of replication in our cohort, we estimated power of detection using software available at http://statgen.iop.kcl.ac.uk/gpc/. To calculate power of detection at a 0.05 type-1 error rate level we used as parameter values: relative risk values, a disease prevalence of 0.1%, risk allele frequencies observed in our control population and the relevant case to control ratio. Direct positive replications are better estimators of genetic effect then original reports [Lohmueller *et al.* 2003]. However, as direct positive replications were not available, we used data from original reports to estimate relative risk and odds ratio values. We defined the risk allele as that conferring risk in the original report. We emphasise that we used the estimated relative risks, as opposed to the lower bounds on these relative risks. This means of course that we cannot rule out smaller effects for the variants even when our estimated power is high. Although TLE is the most common partial epilepsy, precise prevalence figures are not known. We view the use of a disease prevalence of 0.1% as a conservative estimate for all types of TLE tested here.

2.10 Visual assessment for quantification of adverse reaction to Vigabatrin

This section refers to visual field measurements discussed in Chapter 5.

2.10.1 Calculation of Goldmann perimetry values

Visual fields were recorded by an ophthalmologist using standard kinetic Goldmann perimetry. We calculated the mean radial degrees (MRD) for each eye by measuring the radial distance of the intact field in degrees from fixation at 12 points, 30 degrees apart and then taking the average of these values (as previously described [134]). We view this value as a quantitative measurement of vigabatrin-induced visual field constriction. The III4e and I4e isopters were used to calculate MRD values in cohorts *"A"* and *"B"* respectively. The reason for this discrepancy was that the isopter routinely used in at one referral center was different to that used at the other. We felt justified in doing this because a) the two cohorts were analyzed separately and b) the aim of the study was to associate the development and severity of constriction with particular genotypes and not to compare values for constricted fields.

Catch trials were performed and fixation was carefully monitored during testing by direct visualization of the fixing eye. We excluded patients in whom fields were unreliable due to poor compliance, i.e. they did not maintain fixation, they were slow to react, or gave inconsistent responses to the light stimulus producing a wide variation in radii (so called 'spiraling'). We also excluded patients with pre-existing eye disease, such as glaucoma, retinitis pigmentosa, optic nerve or macular degeneration or bilateral cataracts. We did not exclude patients with field defects due to previous intracranial surgery; however in such cases we did exclude the affected quadrant or hemifield from the calculation of the MRD. We used readings from the right eye in calculating MRD values unless local pathology,

such as cataract, prohibited this. Where serial fields were available we selected the measurement that corresponded to the greatest drug exposure time.

2.10.2 Estimation of Goldmann perimetry values in a control populations

MRD values from epilepsy patients previously unexposed to vigabatrin were used as estimates of a normal, unaffected population. MRD values of 49.3 and 47.6 degrees were used as cut-off for normal visual fields in *Cohorts A* and *B* respectively [135, 136].

Chapter 3 Attempted replication of previously reported genetic association studies with sporadic temporal lobe epilepsy. Lessons learnt for complex trait genetics

3.1 Introduction

Although major advances have been made in the field of Mendelian genetics, with the discovery of numerous genes underlying monogenic diseases, only a limited number of genetic susceptibility factors underlying common diseases have been convincingly identified [68]. This is mainly because common diseases arise from interaction of several genes, with additional environmental influences, and are therefore much more complex. Moreover, the best strategy to study such complex traits is a matter of ongoing debate [66].

3.1.1 The role of genetics in temporal lobe epilepsy

The study of genetic contributions to the aetiology of epilepsy exemplifies these issues. Temporal lobe epilepsy (TLE) is the most common form of partial epilepsy. According to the focus of seizure origin, TLE can be further subclassified into mesial TLE and lateral or neocortical TLE. Therefore, TLE is itself a heterogeneous condition with a wide range of possible underlying aetiologies. Although traditionally it was considered an acquired

disorder, it is now becoming clear that genes play a role in at least a subset of cases. Evidence for this comes from both human and animal studies.

TLE rarely occurs in families as a monogenic trait. So far, the only known gene associated with a monogenic type of TLE is *LGI1*, mutations in which are responsible for some cases of familial lateral TLE manifesting as autosomal dominant partial epilepsy with auditory features [52]. Several families with autosomal dominant mesial TLE have also been reported, and linkage has been established in some of them [137-139]. In addition, individuals with a TLE phenotype have been reported in other multiplex familial epilepsies, such as generalized epilepsy with febrile seizure [17] familial partial epilepsy with variable foci [140, 141] and partial epilepsy with pericentral spikes [142]. Finally, genes have been identified in mice, mutations in which cause a phenotype reminiscent of TLE. Examples include *Scn2a* in the Q54 mouse [143], *Jh8* in the jerky mouse[144] and *Plc-β1*[145]. These observations illustrate that there is good evidence for a genetic contribution to sporadic TLE.

3.1.2 A hypothesised link between febrile seizures and hippocampal sclerosis

Hippocampal sclerosis (HS) is the most common histopathology underlying sporadic forms of mesial TLE. In most cases, the cause of HS is not known. A long standing but controversial hypothesis links childhood febrile seizures (FS) with subsequent HS [146]; patients with TLE are more likely than patients with other focal epilepsies to have a history

of FS [147]. The heritability of FS appears to be high [148, 149]. Taken together, these data imply a genetic component to HS and thus to many cases of mesial TLE.

3.1.3 Previous genetic associations studies involving temporal lobe epilepsy

Over the last four years several studies have reported associations between common variants in specific genes and sporadic TLE including mesial TLE. Significant association with TLE has been reported in four genes: *Il-1b* [114], *PDYN* [150][Stogmann *et al.* 2002], *GABBR1* [151] and *PRNP* [152]. One study has reported an association of *APOE* with age of onset of TLE [153]. In addition, variants in *CHRNA4 [154]* and *GABRG2* [155] have been reported as risk factors for childhood FS. These associations, rightly or wrongly, have contributed to the growing view that TLE is at least partially genetic [156].

The well-accepted prevalence of false positive results in genetic association studies stresses the importance of replication to confirm or reject novel association results [106]. In fact, where follow-up studies have been reported (*Il-1b* [117] [120, 157] [116, 118, 119]; *PDYN* [158, 159] *APOE[160-162], CHRNA4 [163-167]*) results have been conflicting, making it currently unclear whether any of the genes have important effects on the development of TLE or FS.

3.1.4 Aims

In this chapter all seven claims of positive association (for TLE and FS) in the literature are examined by attempting to replicate them in an independent cohort of adult patients, which included subcohorts of 339 patients with a syndromic diagnosis of TLE and a partially overlapping subcohort of 107 patients with epilepsy of any type who had a definite previous history of FS.

Further, to make in total a cohort of 752 epilepsy patients, an additional 371 patients with forms of epilepsy other than TLE were added. This cohort was used to examine if the seven variants contribute to the development of any common type of epilepsy.

3.2 Results

3.2.1 Breakdown of patient cohort

The primary aim of this study was to re-test previous association studies involving various TLE phenotypes. We thus classified the 339 patients diagnosed with TLE in to relevant sub-categories as defined by the original studies reporting each association (see table 3.1). In this way we were able to re-test directly each of the claims for association. We note that these categories are not mutually exclusive, which is why the sum of the sub-categories is

greater than the total number of patients with TLE examined. For each gene, we make clear which group of patients was used for the association.

Table 3.1 Breakdown of TLE patient cohort

TLE Classification*	Number of patients
Hippocampal Sclerosis[1]	141
Familial non-lesional[2]	50
Non-lesional[3]	245
Refractory with Hippocampal Sclerosis[4]	121
Symptomatic Refractory[5]	181

* The categories of TLE defined here overlap. Thus their sum is greater then the total number of TLE patients. [1] As used by Kanemoto et al [Kanemoto *et al.* 2000]. HS was diagnosed using high-resolution magnetic resonance imaging. [2] As used by Stogmann et al [Stogmann *et al.* 2002]. Non-lesional TLE patients (HS allowed) with first or second degree relatives having a history of seizures. [3] As used by Gambardella et al [Gambardella *et al.* 2003a]. Non-lesional TLE patients (HS allowed). [4] As used by Walz et al [Walz *et al.* 2003]. We used criteria outlined in Siddiqui et al [Siddiqui *et al.* 2003] to define drug resistance. [5] As used by Briellmann et al [Briellmann *et al.* 2000]. Patients with symptomatic, drug resistant (as defined in [Siddiqui *et al.* 2003]) TLE. Includes patients with HS.

In order to examine any effect of these variants on other classifications of epilepsy we have divided our overall cohort of 752 epilepsy patients as having idiopathic generalised epilepsy (IGE, n=96), cryptogenic (n=233) or symptomatic epilepsy (n=330). This scheme follows presumed gradations in the level of genetic contribution to aetiology and, for our patients, corresponds approximately, though not exactly, to the to the International League Against Epilepsy classification of epilepsies and epileptic syndromes [5]. Ninety-three individuals were unclassifiable according to these criteria. These patients therefore were only included in the association tests involving "all epilepsy".

All patients and controls were of self-identified European ancestry. We used un-related individuals from a twin registry as controls [168].

3.2.2 Results for *Il-1b*

The *Il-1b* –511 variant has been previously reported to associate with TLE and HS [169], FS [115, 169], and refractory partial epilepsy [120]. The first of these reports [114] showed an association between *Il-1b* –511 genotype and development of TLE accompanied by HS in Japanese patients (p=0.0085; OR 3.29 (CI 1.28-8.47)). Our results failed to replicate, or show a trend, in support of this association in our European cohort (see table 3.2). We estimated our cohort to have 68% power of detection for this association[x].

We also examined the reported association with FS in Finnish children [115] and in Japanese patients with HS [169]. We were unable to replicate the association with childhood FS in our cohort of 107 patients with all forms of epilepsy and antecedent FS (p=0.557 and 0.408 for genotype and allele)

Restricting our analysis to HS patients with a history of FS (n=50) similarly failed to show significance (p=0.886 and 0.865 for genotype and allele). We estimated our cohort to have 90% power of detection for the febrile seizure association. We were unable to calculate power of detection for FS with HS as genotypic counts for these subgroups were not detailed in the original report.

Similarly, we found no association (p=0.780 and p=0.672 for genotype and allele) with refractory partial epilepsy in the 372 patients in our overall cohort who matched the

[x] *Il-1b* -511 was genotyped by Taqman (see section 2.3). The method used to calculate power of detection is detailed in section 2.9.3.

phenotype reported in the original positive association study [120]. We estimated our cohort to have almost complete power of detection for this association.

Examination in other forms of epilepsy failed to show any contribution of this allele to development of other types of epilepsy. Results are shown in Table 3.2

Table 3.2 Il-1b –511 genotype counts and analysis results

	TLE+HS[i] *n=141*	All Patients *n*=752	Symptomatic *n*=330	Cryptogenic *n*=233	IGE *n*=96	Controls *n*=384
G/G	59(0.45)	309(0.45)	133(0.44)	98(0.46)	37(0.42)	161(0.44)
G/A	57(0.44)	306(0.44)	137(0.45)	94(0.44)	41(0.47)	162(0.45)
A/A	15(0.11)	74(0.11)	34(0.11)	21(0.10)	10(0.11)	41(0.11)
MAF[†]	0.33	0.33	0.34	0.32	0.35	0.34
Genotype *p* *	0.9808	ns	ns	ns	ns	
Allele *p**	0.9272	ns	ns	ns	ns	

[i]TLE with hippocampal sclerosis as used by Kanemoto et al [Kanemoto *et al.* 2000].
**p* values are uncorrected and calculated from χ^2 distribution generated from contingency tables.
"ns" = non significant. [†]Minor allele frequency. Genotype counts are those for successful typings only

3.2.3 Results for *PDYN*

Stogmann et al. reported an effect of a functional *PDYN* promoter variation on development of non-lesional TLE in patients of middle-European descent whose first and/or second degree relatives also had a history of seizures [150]. The authors reported an over representation of the low expression allele ("L") in the patient cohort (p=0.0025; OR= 5.33(CI 1.94-14.65)). We failed to replicate this association in the 50 patients in our

overall cohort matching this phenotype (see Table 3.3). Our cohort is predicted have 89% power of detection for this association[xi].

The same authors also reported that patients with TLE carrying the low expression *PDYN* allele had a significantly higher risk of developing frequent secondarily generalised tonic clonic seizures and status epilepticus. We were unable to replicate these associations as our patient numbers were too low in the case with frequent secondarily generalised seizures and we lacked reliable information that could confirm the absence of a history of status. However, of the 14 patients we identified with frequent secondarily generalised seizures (as defined in the original association), only one carried the low expression allele.

Analysis of the broader cohort suggests the possibility that *PDYN* may act as a general risk factor for epilepsy (p=0.035) and for idiopathic generalised epilepsy (p=0.038). Following the previous observation [150] that *PDYN* genotype effect was limited to familial cases of epilepsy we examined the variant in familial cases of IGE and found the association to strengthen (p=0.002). In light of the many tests we have conducted here, these results can be considered only as a modest trend. To clarify the role of *PDYN* in epilepsy further analyses is required in larger patient cohorts.

[xi] The *PDYN* promoter repeat was genotyped as detailed in section 2.5. The method used to calculate power of detection is detailed in section 2.9.3.

Table 3.3 *PDYN* genotype counts and analysis results

	TLE+FH[i] *n*=50	All Patients *n*=752	Symptomatic *n*=330	Cryptogenic *n*=233	IGE *n*=96	IGE+FH[ii] *n*=32	Controls *n*=384
H/H	17(0.36)	336(0.48)	152(0.49)	106(0.50)	33(0.36)	8(0.25)	175(0.48)
L/H	22(0.47)	270(0.39)	120(0.39)	80(0.38)	46(0.49)	16(0.50)	160(0.44)
L/L	8(0.17)	92(0.13)	37(0.12)	27(0.13)	14(0.15)	8(0.25)	30(0.08)
MAF[†]	0.4	0.33	0.31	0.31	0.40	0.5	0.3
Genotype *p* *	0.0898	0.0348	ns	ns	0.038	0.0021	
Allele *p**	0.0429	ns	ns	ns	0.012	0.0011	

[i]Familial non-lesional TLE (history of seizures in first and/or second degree relatives) as used by Stogmann et al [Stogmann *et al.* 2002].
[ii]Familial IGE (history of seizures in first and/or second degree relatives). **p* values are uncorrected and calculated from χ^2 distribution generated from contingency tables. "ns" = non significant.
[†]Minor allele frequency. Genotype counts are those for successful typings only

3.2.4 Results for *GABBR1*

Gambardella et al. [151] reported the *GABBR1* G1465A variant to have a major effect on the development of non-lesional forms of TLE in Italian patients (p=<0.0001 OR= 37.95 (CI 8.84 - 162.98)). In our cohort of patients with non-lesional TLE, however, we see no association (see table 3.4). We have 81% power of detection for this association[xii]. However, this is a conservative estimate as we used the same relative risk value calculated for the heterozygote as an estimate for the homozygote risk genotype. This was required as it is impossible to calculate a relative risk value for the homozygote risk genotype as that genotype was not observed in the original control cohort.

We saw no effect of this variant on the development of other forms of epilepsy as classified here (see Table 3.4).

[xii] *GABBR1* G1465A was genotyped using Taqman as detailed in section 2.3. The method used to calculate power of detection is detailed in section 2.9.3.

Table 3.4 *GABBR1* G1465A genotype counts and analysis results

	nl-TLE *n*=245	All patients *n*=752	Symptomatic *n*=330	Cryptogenic *n*=233	IGE *n*=96	Controls *n*=1089
G/G	218(0.99)	683(0.99)	302(0.99)	210	89(0.99)	1062(0.99)
G/A	2(0.01)	6(0.01)	4(0.01)	0	1(0.01)	8(0.01)
A/A	0	0	0	0	0	0(0.0)
MAF†	0.002	0.004	0.006	0	0.006	0.004
Genotype*	1	ns	ns	ns	ns	
Allele*	0.6823	ns	ns	ns	ns	

‡Non-lesional TLE patients. Includes patients HS. As used by Gambardella et al [Gambardella *et al.* 2003a]
**p* values are uncorrected and generated using an exact probability test. "ns" = non significant.
†Minor allele frequency. Genotype counts are those for successful typings only

3.2.5 Results for *PRNP*

The non-synonymous *PRNP* variant Asn171Ser was reported by Walz et al [152] to contribute to the development of refractory TLE with HS in a Brazilian patient cohort which included patients of both European and African descent. The authors observed the variant in 23% of patients but were unable to detect it in controls suggesting the variant contributes greatly to disease development (p=<0.0001). We failed to replicate this association in our cohort of 121 patients with refractory TLE and HS[xiii] (see Table 3.5). In fact, we were unable to detect this variant in any of our refractory TLE patients. We observed the variant only twice, once in a patient with symptomatic epilepsy secondary to cerebrovascular disease and once in a control individual. As we could not estimate the genotypic relative risk of the variant in the original study (the variant was not observed in the control population), it was not possible to calculate power for this association.

[xiii] *PRNP* ASN171Ser was genotyped using Taqman as detailed in section 2.3.

Table 3.5 *PRNP* Asn171Ser genotype counts and analysis results

	Refrac. HS[i] *n*=121	All patients *n*=752	Symptomatic *n*=330	Cryptogenic *n*=233	IGE *n*=96	Controls *n*=384
Asn/Asn	109	718(0.99)	320(0.99)	219	93(0.99)	360(0.99)
Asn/Ser	0	1(0.01)	1(0.01)	0	0(0.01)	1(0.002)
Ser/Ser	0	0	0	0	0	0
MAF[†]	0	<0.001	<0.001	0	0	0.001
Genotype*	1	ns	ns	ns	ns	
Allele*	1	ns	ns	ns	ns	

[i]Refractory TLE with hippocampal sclerosis. As used by Walz et al [Walz *et al.* 2003]
**p* values are uncorrected and generated using an exact probability test. "ns" = non significant.
[†]Minor allele frequency. Genotype counts are those for successful typings only

3.2.6 Results for *APOE*

Briellmann et al. [153] reported a significant association in Australian patients between the *APOE* ε4 allele and age of onset of refractory TLE. Results indicated that the presence of the ε4 allele correlated with a mean age of onset of 5 (+/- 5) years, whilst in the absence of the allele, mean age of onset was 15 (+/- 10) years (Mann Whitley U test p=0.004). Our results, from 181 refractory temporal lobe epilepsy patients, do not support this association (Mann Whitley U test p=0.1808)[xiv]. Mean age of seizure onset in patients carrying at least one copy of the ε4 allele (30 of 181 patients) was 13.7 (+/- 10) years. Patients not carrying a copy of the ε4 allele had a mean age of seizure onset of 16.7 (+/- 11) years.

We further examined the distribution of the *APOE* haplotypes across refractory TLE and other sub groups of epilepsy (see table 3.6). We note a modest trend in haplotype distribution for the IGE (p=0.031) group and ε4 allele frequency for refractory TLE. In both cases the trend seems to be generated by an under representation of the hypothesised risk

[xiv] *APOE* ε4 was genotyped using Taqman as detailed in section 2.3.

allele (ε4). We feel this association is likely the result of multiple testing and thus would view it as tentative at best. As such, confirmatory testing in an independent cohort is required.

Table 3.6 *APOE* genotype counts and analysis results

	Refrac. TLE[i] *n*=181	All patients *n*=752	Symptomatic *n*=330	Cryptogenic *n*=233	IGE *n*=96	Controls *n*=384
ε3[†]	230(0.80)	926(0.79)	406(0.78)	285(0.8)	125(0.83)	469(0.74)
ε4[†]	36(0.13)	156(0.13)	72(0.14)	46(0.13)	14(0.09)	108(0.17)
ε2[†]	20(0.07)	92(0.08)	44(0.08)	24(0.07)	11(0.07)	57(0.09)
ε4 freq[‡]	0.13	0.13	0.14	0.13	0.09	0.17
Hap *p*[ii]	0.205	Ns	ns	ns	0.031	
ε4 *p*[iii]	0.078	Ns	ns	ns	ns	

[i]Symptomatic refractory TLE patients. As used by Briellman et al [Briellmann *et al.* 2000]
[ii]Uncorrected significance value for haplotype distribution. Detailed in methods section.
[iii]Uncorrected significance value for 2x2 χ^2 distribution between presence and absence of ε4 allele.
"ns" = non significant. [†]Haplotype counts are those inferred from successful genotyping at both loci only.
[‡]Frequency of ε4 allele

3.2.7 Results for *CHRNA4*

Chou et al reported, in a cohort of Taiwanese children, association of a synonymous SNP variant Ser543Ser with childhood FS (*p*=0.001; OR= 2.84 (CI 0.87-9.28)) [154]. Our results, from the 107 epilepsy patients with a history of FS in our cohort, failed to replicate this association. However, we do note our study is estimated to have only 17% power of detection for this association[xv].

Analysis across different subgroups of epilepsy failed to support a role for this variant in disease predisposition for the subtypes of epilepsy tested here (see Table 3.7).

[xv] *CHRNA4* Ser543Ser was genotyped using Taqman as detailed in section 2.3. The method used to calculate power of detection is detailed in section 2.9.3.

Table 3.7 *CHRNA4* Ser543Ser genotype counts and analysis results

	Feb. Con.[i] n=107	All patients n=752	Symptomatic n=330	Cryptogenic n=233	IGE n=96	Controls n=384
T/T	36(0.36)	241(0.36)	119(0.40)	70(0.33)	26(0.29)	120(0.32)
C/T	48(0.48)	301(0.45)	122(0.41)	97(0.46)	45(0.51)	200(0.52)
C/C	15(0.15)	131(0.19)	53(0.18)	44(0.21)	18(0.2)	61(0.16)
MAF	0.39	0.42	0.39	0.44	0.46	0.42
Genotype*	0.6525	ns	ns	ns	ns	
Allele*	0.4665	ns	ns	ns	ns	

[i]Epilepsy patients with a history of childhood febrile seizures. **p* values are uncorrected and generated using an exact probability test. "ns" = non significant. [†]Minor allele frequency
Genotype counts are those for successful typings only

3.2.8 Results for *GABRG2*

Chou et al reported a synonymous variant in the *GABRG2* gene (rs211037) to associate with FS in a cohort of 104 Taiwanese children (p=0.017, OR= 2.56 (CI 1.01-6.50)[155]. Our results in a cohort of 107 patients with a history of FS failed to support this association. Again, we note a lack of power, with our cohort estimated to have only 8% power of detection for this association[xvi].

Analysis of the *GABRG2* variant in the broader epilepsy cohort fails to support a role for this variant in the development of other forms of epilepsy tested here (see Table 3.8).

[xvi] *GABRG2* rs211037 was genotyped using restriction digestion as detailed in section 2.4. The method used to calculate power of detection is detailed in section 2.9.3.

Table 3.8 *GABRG2* rs211037 genotype counts and analysis results

	Feb. Con.[i]	All Patients	Symptomatic	Cryptogenic	IGE	Controls
	n=107	*n*=752	*n*=330	*n*=233	*n*=96	*n*=384
C/C	54(0.53)	391(0.60)	174(0.58)	126(0.64)	48(0.61)	203(0.61)
C/T	43(0.43)	213(0.33)	96(0.32)	62(0.31)	24(0.31)	114(0.35)
T/T	4(0.04)	45(0.07)	28(0.09)	10(0.05)	6(0.08)	13(0.04)
MAF[†]	0.25	0.23	0.26	0.21	0.23	0.21
Genotype *p* *	0.3188	ns	ns	ns	ns	
Allele *p**	0.2269	ns	ns	ns	ns	

[i]Epilepsy patients with a history of childhood febrile seizures. **p* values are uncorrected and generated using an exact probability test. "ns" = non significant. [†]Minor allele frequency. Genotype counts are those for successful typings only

3.3 Discussion

This study has examined all positive genetic associations reported thus far on sporadic forms of TLE, and two associations on FS, in a single patient cohort of greater size than used in any of the original reports. We also tested whether the variants might predispose to the development of other forms of epilepsy.

Replication studies have previously been published on five of the seven variants examined here[xvii]. In the case of *APOE*, data from other European populations published prior to the positive association [160, 161], strongly suggest the reported effect on age of onset [153, 162] to be a false positive association. The *IL1β*–511 association with HS was similarly not replicated in European populations, though samples sizes were small [117, 118, 157]. It is

[xvii] Note that at the time this work was published, replication had only been attempted on four of seven original reports. In fact numerous additional replication attempts have been published since our paper was published – all were negative, in agreement with our conclusions

possible that this association is due to a population-specific effect in terms of functionality or patterning of linkage disequilibrium (LD), but this would appear unlikely since there was also a failed replication in a Chinese population [119]. Previous re-testing of the *PDYN* polymorphism in a population of European origin failed to clearly replicate the original result [158, 159]. Similarly, replication attempts of *GABBR1* have also proved negative [156, 163, 165, 166]

Examination of the *CHRNA4* association in an Australian population failed to replicate the original association [164]. However, both the Australian and our study lack power of detection for the *CHRNA4* association. Similarly, we lacked power to replicate the *GABRG2* association. Further research for both of these associations is required, particularly in Asian populations.

Our own results are consistent with these failed replication efforts. In particular, *APOE* results fail to show an effect in the same direction as the original report either in our study or in the other replication efforts. The results for *PDYN* are more ambiguous. We do not see association with "familial" non-lesional TLE, as originally reported. However, our results suggest a possible role for *PDYN* as a risk factor for epilepsy. In particular we note a trend of association with IGE, which is enhanced in "familial" IGE patients (following the definition of familial as used in original report). We encourage examination of this variant across large, well-phenotyped cohorts in order to further examine this putative role.

For *PRNP* Asn171Ser, where there were no previous attempts at replication, our results suggest the original association to be likely false positive. In contrast to the original report,

we see no evidence for even a marginal effect of this variant in the European population tested here. We cannot, however, rule out the possibility that this variant may be disease-causing or in LD with disease-causing variants that are specific to one or more population groups. We thus encourage replication attempts in the respective population. However, a more likely explanation is that the original association is due, at least in part, to population stratification; the cases and controls were drawn from a genetically heterogeneous base population and no formal test for stratification was undertaken.

The results presented here suggest that most or all of the reported associations for TLE are not reproducible in the European population. When these results are considered with previous replication attempts the fact remains that not one of these associations has been accurately replicated in a population of similar or other ancestry. On the weight of current evidence it appears likely that most or all of the reported associations for TLE were false positives. Our data certainly do not rule out the presence of risk factors in any of the genes studied. But given that our sample sizes were always at least as big as the previous reports and usually bigger, it seems fair to conclude that the genes studied here are at best neutral in not having any more supporting evidence than any other part of the genome. It would therefore appear that at present we know little about the genetic basis of sporadic TLE. Even worse than our ignorance, however, is the fact that false associations can lead to a waste of effort in trying to understand the biological bases of the involvement of genes which are not in fact involved in the disease, in this case sporadic TLE. How can this situation be improved?

In agreement with recent recommendations [156], we believe methodological weaknesses in study design contribute to the variable results. For example, studies that fail to consider stratification in the context of an ethnically-mixed patient and control population (see [152]) should no longer be published. All seven of the variants tested here have been selected using a sequence-based approach, focusing on functional, or putatively functional, variation. Whilst this approach has obvious advantages, future studies should seek to merge a sequence- based approach with methods that systematically represent variation in candidate genes. Map-based approaches such as the tagging SNP method address this issue by taking advantage of LD to allow examination of most variation by typing only a small number of tagging SNPs [66, 170]. The publicly available HapMap resource, in conjunction with rapidly-developing tagging SNP selection techniques [132], makes this approach feasible even with only modest laboratory resource. In addition, for an association to be considered replicated, it should be replicated exactly. A follow up study that shows association with another polymorphism in a gene, or a somewhat different phenotype may be of interest, but should not be considered a replication. These sorts of "replications" must be treated with particular care because of the possibility of explored populations, polymorphisms, and phenotypes until something is found in a given gene that can be called a replication. Neale and Sham [170] have proposed a gene-based definition of replication that overcomes some of these issues.

These methodological issues deserve more serious attention than they have sometimes received, but we believe the central difficulty is more innate, and less easy to rectify.

The most fundamental problem with current efforts is the relatively low sample size in most studies, resulting in limited power to either detect, or to definitively rule out, association. Few investigators working on epilepsy genetics have large cohorts for any particular type of epilepsy, including our own current and previous efforts. It does not appear to us that in the near future individual investigators will be able to increase their own sample sizes to levels sufficient to carry out highly powered association studies within the various forms of epilepsy. For example, our own collection comes from one of the biggest referral centres in Europe and although our own total phenotyped collection is sizeable, the numbers of patients with any given subtype of epilepsy is still small. Our previous study of refractory epilepsy, for example, included only 200 refractory patients, reporting a modest association with a putatively functional variant in the *ABCB1 (MDR1)* gene [171]. This association was not replicated in an exact replication attempt [172], but was reported as "replicated" for a similar but not identical definition of pharmacoresistance [173].

We therefore see two central lessons from our failure to replicate the TLE associations. First, the field must concentrate more seriously on efforts to determine which polymorphisms have real effects, as opposed to always racing to publish a new association. False positives are clearly exacerbated by the publication of multiple small studies and the practice of data exploration to identify subgroups that show associations. This fact must be recognised and addressed. Second, and most fundamentally, we believe it is critical for different research groups to increase substantially the size of their patient cohorts. Given that this will take time, we feel that in the near term, groups should combine their epilepsy samples, and attempt to replicate one another's results in a population of similar ethnicity before publication. Although collaboration to such a level could be viewed

as controversial from a scientific point of view, we feel that the trade off here is justified. It is only with such steps that the epilepsy community, like other disease areas, will be able to arrive at a reasonable false to true discovery ratio in reported associations.

Chapter 4 Confirmation of *BRD2* as a risk factor for juvenile myoclonic epilepsy in populations of European, but not Indian origin.

4.1 Introduction

4.1.1 Juvenile myoclonic epilepsy

Juvenile myoclonic epilepsy accounts for 26% of all IGE and is characterised by adolescent-onset myoclonic jerks (especially matinal), generalized tonic-clonic and/or absence seizures and, a characteristic interictal electroencephalogram in untreated patients and occasionally, in asymptomatic relatives [174]. Although single gene mutations that cause Mendelian forms of idiopathic epilepsy, including familial forms of JME, have been identified in a number of families (*EFHC1*[55], *GABRA1[46]*, *CACNB4*[175] and *CLCN2* [43]), the vast majority of idiopathic epilepsies follow a more complex mode of inheritance.

4.1.2 The genetics of complex forms of juvenile myoclonic epilepsy

Progress in understanding the genetic bases of common, more genetically-complicated forms of IGE has been much slower. Greenberg et al.[87] proposed an oligogenic model in which no single locus is either sufficient or necessary to cause IGE but rather a number of loci interact to produce specific subtypes of IGE. In the case of JME, disease manifestation

may be due to interaction between variation at a locus linked specifically with JME termed EJM1, and susceptibility loci common to all IGE. Support for a role of the EJM1 peak comes from independent linkage studies [80, 85, 86, 95].

Recently it has been suggested that variation in *BRD2* is responsible for the EJM1 linkage peak. Association analysis in 20 probands implicated two promoter SNPs (rs3918149 and rs206787) located in the *BRD2 (RING3)* as risk factors for JME. Both SNPs associated individually, although rs3918149 was slightly more significant than rs206787 (odds ratio, OR, 2.80 (CI 1.19-6.64) and 2.21 (CI 1.08-4.52) respectively). When considered together, the two SNPs formed an at-risk haplotype that confers an odds ratio of 4.37 (CI 1.67 – 11.45) for developing JME [1].

4.1.3 Aims

We sought to replicate the role of *BRD2* in JME for two principal reasons. Firstly, both linkage and especially association studies have a high incidence of false positive results. We wished to determine whether putative risk allele did indeed confer increased risk. Secondly, estimation of the odds ratio in samples in which linkage has been identified is known to systematically overestimate effects. We wished to calculate a more accurate estimate of the true effect. We sought to replicate the role of *BRD2* in JME by analyses of three independent cohorts of differing ethnicity to determine if the effect was consistent across populations.

Our results support a role for *BRD2* (or a nearby gene) in the development of JME in European population samples. The variant does not, however, appear to confer a risk in a patient population of Southern Indian ancestry suggesting heterogeneity of disease causation, differing patterns of linkage disequilibrium (LD), or that the Europeans associations are false positives (though this seems unlikely). We were also unable to show any effect of the promoter variant on *BRD2* mRNA levels and so are unable to identify the causal variant.

4.2 Results

4.2.1 Replication in London cohort

The first stage of this study was to attempt replication of the association between *BRD2* promoter variation and JME as reported by Pal et al. [1]. We thus genotyped the two candidate causal promoter variants proposed by Pal et al. (rs3918149 and rs206787) in the 34 JME patients from our London cohort.

Rs3918149 showed a significant level of association at both an allelic (p=0.001) and genotypic (p=0.001) level, with the derived allele (A) conferring increased risk. The derived homozygous genotype of rs3918149 ("AA") conferred an OR of 29.1 (2.89-293.03). We consider this a clear replication of the Pal et al. association. The second candidate causal variant, rs206787 did not associate significantly with JME. Results are shown in Table 4.1.

Pal et al. also reported the haplotype defined by the derived state of the two candidate causal promoter variants ("AA") to confer an increased risk of JME (OR=4.37 (1.67-11.45)). We assessed the frequency of this haplotype across cases and controls (haplotype 3 in Table 4.2) and found it conferred an OR of 2.49 (1.29-4.81), comparable with the original report. We also assessed the distribution of all two-locus haplotypes defined by the two promoter SNPs across cases and controls using the log-likelihood ratio test and found it to be significant (p=0.012). Results are shown in Table 4.2.

Pal et al. reported a five-locus risk haplotype consisting of four SNPs and one microsatellite spanning a 2.3kb region of *BRD2* to confer an OR of 6.45 (2.36-17.58) (see orange coloured SNPs in Figure 4.1 for location in gene). We were unable to identify this haplotype in our patient population[xviii].

Given that rs206787 was the less significant of the two candidate causal variants proposed by Pal et al., and that it does not associate with JME here, we no longer considered this SNP as candidate causal. The fact that rs3918149 and rs206787 are in high LD (D'=1, r^2=0.117) would suggest the association of rs206787 with JME reported by Pal et al. was likely driven by LD with rs3918149.

[xviii] The original authors (i.e. Pal et al) did not provide sufficient information in their paper to identify the five locus haplotype. Despite acknowledging this via email, they have not provided the required information.

Table 4.1 Association analysis of BRD2 promoter variation in London JME patients and controls.

rs3918149	Patients n=34	Controls n=256	OR[a] (95% CI)	*p* Value	Rs206787	Patients n=34	Controls n=256	OR[a] (95% CI)	*p* Value
Genotype					**Genotype**				
GG	20(0.59)	194(0.78)	1		**TT**	8(0.24)	86(0.36)	1	
GA	11(0.32)	54(0.22)	1.98 (0.89-4.38)	0.001[ii]	**AT**	20(0.59)	120(0.50)	1.79(0.75-4.26)	0.375[i]
AA	3(0.09)	1(0.00)	29.1 (2.89-293.03)		**AA**	6(0.18)	35(0.14)	1.84(0.60-5.70)	
Allele					**Allele**				
G	51(0.75)	442(0.89)	1	0.001[iii]	**T**	36(0.53)	292(0.61)	1	0.229[i]
A	17(0.25)	56(0.11)	2.63 (1.42 to 4.87)		**A**	32(0.47)	190(0.39)	1.36(0.82-2.27)	

Numbers are those in which successful genotypes were obtained [a]Odds ratio [i] *p* value from χ^2 test using 2 degrees of freedom.[ii] Empirical *p* value based on 25,000 replicates.[iii] *p* value from χ^2 test using 1 degree of freedom.[na] Non applicable.

Table 4.2 Haplotype association of promoter variants in London JME patients

ID	Rs206787	Rs3918149	Freq Cases	Freq Controls
1	T	G	0.510	0.623
2	A	G	0.240	0.264
3	A	A	0.231	0.113
4	T	A	0.019	0.00

Likelihood ratio test: $\chi^2 = 10.96$, df=3, p=0.0119.
Odds ratio for haplotype "AA" against haplotype "TG" = 2.493 (CI 1.292 to 4.813)

4.2.2 Screen of *BRD2* to identify novel candidate causal variants.

Although rs3918149 associates strongly with JME both here and in the Pal et al. study, it is possible that the variant is not itself causal but associating through LD with an unidentified causal variant lying elsewhere in the gene. Pal et al., before proposing the promoter SNPs as candidate causal, screened all coding exons of *BRD2* and surrounding genes [1]. However, they did not screen the 2.1kb of untranslated exonic regions, located in three segments upstream and downstream of the *BRD2* coding region, nor a section of the promoter (see Figure 4.1).

Figure 4.1 Graphical representation of variation across *BRD2*

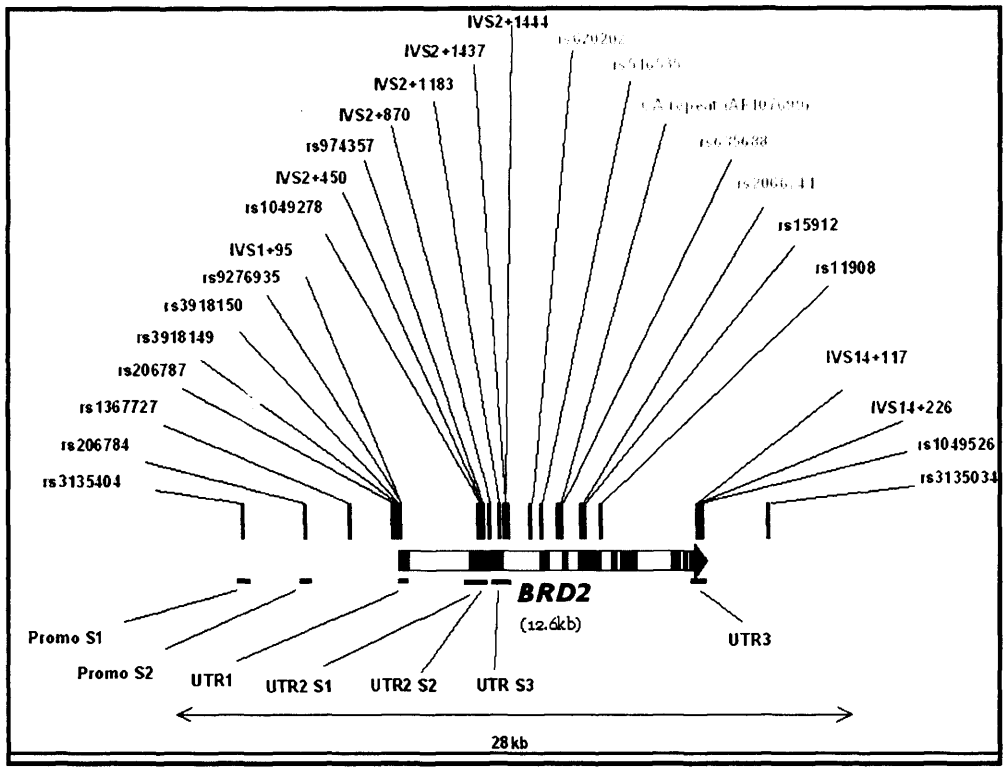


Figure 4.1 legend: A graphical representation of variation across a 28kb region containing *BRD2* considered in this study. The *BRD2* gene is indicated by the arrow. Green fill represents coding exons, orange fill UTR. Solid black horizontal lines represent regions screened for variation in this study. These regions were not screened by Pal et al. [1]. Blue vertical lines indicate variation. Coloured variation labels represent the following: Blue = Hapmap SNP, Red = Candidate causal from Pal et al.[1], Green = Detected here as a result of UTR and promoter screen, Orange = Variants that make up Pal et al. 5 locus risk haplotype [1]

In an attempt to identify novel candidate causal variants, we screened these regions in 16 unrelated CEPH individuals[xix]. The screen resulted in the detection of 14 SNPs, eight of which were previously unreported (see Table 4.3 for SNP details).

Inference of haplotypes generated by variation across *BRD2* allows characterisation of the relationship between SNPs in the region. This information can be used as a guide to select novel candidate causal variation.

4.2.3 Using the haplotypic structure of *BRD2* to identify novel candidate causal variants.

We genotyped the 14 SNPs discovered in the screen, the two candidate causal promoter SNPs and four SNPs from the Pal et al. five-locus risk haplotype in the same European American samples used in the HapMap project. This allowed genotypes from these 20 SNPs to be merged with genotypes from HapMap (five SNPs in *BRD2*), thus creating a dataset of 25 SNPs spanning the *BRD2* gene (i.e. all SNPs named in Figure 4.1).

The haplotypes inferred from these data, and the resulting network constructed from these haplotypes, are shown in Table 4.4 and Figure 4.2 respectively.

[xix] See section 2.6 for details of protocol used in this screen.

Table 4.3 SNPs discovered in *BRD2* screen

SNP name	Location	Sequence context	*f*[CEPH EU]
IVS1+95	5' UTR	ggccccccttg[I/D]ccccttggcgc	0.04
IVS2+450	5' UTR	ccgcggagag[G/A]tgttccttccc	0.02
IVS2+870	5' UTR	ctcggagtct[G/A]tccacccccctc	0.04
IVS2+1183	5' UTR	tttgtcgcct[G/C]gaggcccaaga	0.06
IVS2+1437	Int 2	gcgcgctgct[G/C]ttgcggcgcag	0.04
IVS2+1444	Int 2	gctgttgcgg[C/T]gcagctgtcg	0.01
IVS14+117	3' UTR	gacaaaacaa[C/T]attgaattccc	0.04
IVS14+226	3' UTR	ggagtgatct[C/G]ttggacacaga	0.05

f[CEPH EU] – Minor allele frequency in CEPH Europeans, UTR=untranslated region. Int=Intron

Table 4.4 *BRD2* haplotypes inferred in CEPH American Europeans

ID	*F*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
HP1	0.27	T	T	T	T	G	G	T	N	T	G	T	G	G	G	C	T	T	C	T	G	T	C	C	T	T
HP2	0.11	T	C	T	A	G	G	T	N	T	G	T	G	G	G	C	G	C	T	del	G	C	C	C	T	T
HP3	0.11	T	C	T	T	G	G	T	N	T	G	T	G	G	G	C	G	T	C	T	G	T	C	C	T	T
HP4	0.09	T	C	C	A	G	G	C	N	T	G	T	G	G	G	C	G	C	T	del	C	C	C	C	T	C
HP5	0.07	T	C	T	T	G	G	T	N	T	G	T	G	G	G	C	G	T	C	T	G	C	C	C	T	T
HP6	0.06	A	C	T	A	G	G	T	N	T	G	G	G	G	G	C	G	C	T	del	G	C	C	C	C	T
HP7	0.06	T	T	T	T	G	G	T	N	T	G	T	G	G	G	C	G	T	C	T	G	T	C	C	T	T
HP8	0.05	T	C	T	A	A	G	T	N	T	G	T	G	G	G	C	G	C	T	del	G	C	C	G	T	T
HP9	0.02	T	T	T	T	G	G	T	N	T	G	T	G	G	G	C	G	T	C	T	G	C	C	C	T	T
HP10	0.02	T	C	T	T	G	C	T	I/D	T	G	T	A	G	C	C	G	T	C	del	G	C	T	C	T	T
HP11	0.02	T	T	T	T	G	G	T	N	T	A	T	G	C	G	C	T	T	C	T	G	T	C	C	T	T
HP12	0.02	T	C	T	A	A	G	T	N	T	G	T	G	G	G	C	G	C	T	del	G	C	C	C	T	T
HP13	0.02	T	C	T	A	G	G	T	N	T	G	T	G	C	G	C	G	C	T	del	G	C	C	C	T	T
HP14	0.01	T	T	T	T	G	G	T	N	T	G	T	G	C	G	C	G	T	C	T	G	T	C	C	T	T
HP15	0.01	A	C	T	A	G	G	T	N	T	G	G	G	G	G	C	T	C	T	del	G	C	C	C	C	T
HP16	0.01	A	C	T	A	G	G	T	N	T	G	G	G	G	G	C	G	C	T	del	G	C	C	C	T	T
HP17	0.01	A	C	T	T	G	G	T	N	C	G	G	G	G	G	C	G	T	C	T	G	C	C	C	C	T
HP18	0.01	T	C	T	T	G	C	T	I/D	T	G	T	A	C	C	C	T	T	C	del	G	C	T	C	T	T
HP19	0.01	T	T	T	T	G	C	T	I/D	T	G	T	A	G	C	C	G	T	C	del	G	C	T	C	T	T
HP20	0.01	T	T	T	T	G	G	T	N	T	G	T	G	C	G	T	G	T	C	del	G	C	C	C	T	T
HP21	0.01	T	C	T	T	G	G	T	N	T	G	T	G	C	G	T	G	T	C	del	G	C	C	C	T	T

f = frequency HP=HaplotypeSNP coding: 1=rs3135404 ,2=rs206784, 3=rs1367727, 4=rs206787, 5=rs3918149, 6=rs3918150, 7=rs9276935, 8=IVS1+95, 9=rs1049278, 10=IVS2+450, 11=rs974357, 12=IVS2+870, 13=IVS2+1183, 14=IVS2+1437, 15=IVS2+1444, 16=rs620202, 17=rs516535, 18=rs635688, 19= rs2066741, 20=rs15912, 21=rs11908, 22=IVS14+117, 23=IVS14+226, 24=rs1049526, 25=rs3135034

Figure 4.2 Network of BRD2 haplotypes

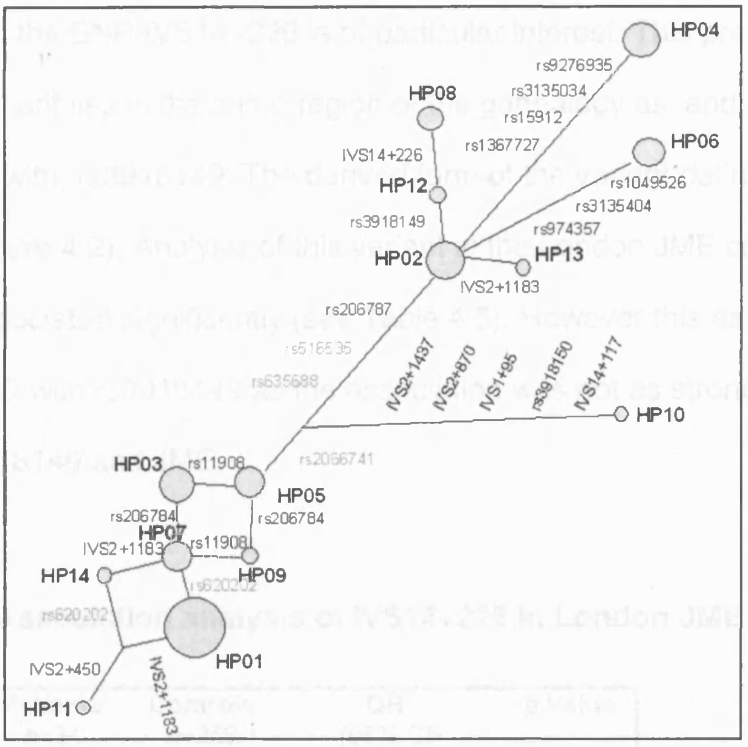


Figure 4.2 legend: A reduced median joining network illustrating haplotypes of *BRD2*. Coloured variation labels represent the following: Blue = HapMap SNP, Red = Candidate causal from Pal et al., Green = Detected here as a result of UTR and promoter screen, Orange = Variants that make up Pal et al. 5 locus risk haplotype. Filled yellow circles represent haplotypes. The size of the circle is directly proportional to the frequency of the haplotype as observed in CEPH American Europeans.

The network depicts the genealogical relationship between haplotypes of *BRD2*. SNPs found on the same branch of the genealogy are in complete LD (r^2=1). Haplotype HP12 represents the risk haplotype defined by the derived state of the candidate causal variant rs3918149. We calculated r^2 values between each SNP and rs3918149 in order to quantify the relationship between the candidate causal variant and other SNPs. Any variants in high r^2 values with rs3918149 can be considered as novel candidate causal variants and thus warrant examination for association in the patients and controls.

In this context, the SNP IVS14+226 is of particular interest. This previously unreported variant lies in the same region of the genealogy as, and is in high LD (D'=1, r^2=0.7) with, rs3918149. The derived form of the variant defines haplotype HP08 (see Figure 4.2). Analysis of this variant in the London JME cohort showed the variant associated significantly (see Table 4.5). However this association was likely due to LD with rs3918149 as the association was not as strong as that between rs3918149 and JME.

Table 4.5 Association analysis of IVS14+226 in London JME patients and controls.

IVS14+226	Patients *n*=34	Controls *n*=256	OR[a] (95% CI)	*p* Value
Genotype				
CC	26(0.76)	217(0.90)	1	
CG	8(0.24)	23(0.10)	2.9 (1.78-7.15)	0.062[i]
GG	0(0.0)	1(0.0)	*na*	
Allele				
C	60(0.88)	457(0.95)	1	0.032[ii]
G	8(0.12)	25(0.05)	2.44(1.05-5.65)	

Numbers are those in which successful genotypes were obtained.
[a]Odds ratio [i] *p* value from χ^2 test using 2 degrees of freedom.
[ii] Empirical *p* value based on 25,000 replicates.
[iii] *p* value from χ^2 test using 1 degree of freedom[na] Non applicable.

SNPs defining the branches representing haplotypes HP04, 06 and 13 are all in low r^2 with rs3918149 (r^2<0.007). It is thus highly unlikely that the causal variant lies on any of these haplotypes. Haplotype HP02, defined by rs206787, shows borderline association in the report of Pal et al., and no association here, and thus is not considered as candidate causal. Variants lying on branches of the

genealogy defined by the ancestral allele of rs206787 (i.e. haplotypes HP1,3,5,7,9-11 and 14) are all in low r^2 (<0.096) with rs3918149 and thus similarly highly unlikely to harbour the causal variant.

4.2.4 Replication of rs3918149 in a population of Irish ancestry.

To further rule out the possibility of a false positive, we examined the role of rs3918149 in an independent cohort of Irish ancestry. The association was borderline significant (genotype p=0.077, allele p=0.043). Unlike the London cohort it did not show a clear replication. Results are shown in Table 4.6.

Table 4.6 Association analysis of rs3918149 in Irish JME patients and controls

rs3918149	Patients *n*=57	Controls *n*=227	OR[a] (95% CI)	*p* Value
Genotype				
GG	43(0.75)	196(0.86)	1	
GA	13(0.23)	29(0.13)	2.04 (0.98-4.25)	0.077[i]
AA	1(0.02)	2(0.01)	2.28 (0.20-25.71)	
Allele				
G	64(0.87)	421(0.93)	1	0.043[ii]
A	8(0.13)	33(0.07)	1.93 (1.01-3.70)	

Numbers are those in which successful genotypes were obtained.
[a]Odds ratio [i] *p* value from χ^2 test using 2 degrees of freedom.
[ii] Empirical *p* value based on 25,000 replicates.[iii] *p* value from χ^2 test using 1 degree of freedom
[na] Non applicable.

4.2.5 The role of rs3918149 in the development of other common forms of epilepsy.

We next wished to examine if the rs3918149 plays a role in the development of other forms of epilepsy. We therefore assessed this variant in our full panel of 678 patients with epilepsy from the London cohort. We tested for association with a) all forms of epilepsy, b) IGE excluding JME, c) focal cryptogenic epilepsy, and d) focal symptomatic epilepsy. Considering that we are testing one polymorphism across four forms of epilepsy, we applied a significance threshold of 0.0125[xx]. Using this threshold there are no significant associations. We do note however, a trend with cryptogenic epilepsy (p=0.068). Results are shown in Table 4.7.

Table 4.7 Association analysis of rs3918149 with other forms of epilepsy in London cohort.

Rs3918149	Controls *n*=256	All Epi *n*=678	IGE - JME *n*=61	FC *n*=200	FS *n*=269
Genotype					
GG	194(0.78)	457(0.76)	43(0.75)	134(0.76)	184(0.78)
GA	54(0.21)	131(0.22)	13(0.23)	36(0.21)	50(0.21)
AA	1(0.004)	12(0.02)	1(0.02)	6(0.03)	1(0.004)
P		0.239[i]	0.395[ii]	0.068[ii]	1[ii]
Allele					
G	442(0.89)	1045(0.87)	99(0.87)	304(0.86)	418(0.89)
A	56(0.11)	155(0.13)	15(0.13)	48(0.14)	52(0.11)
P		0.341[i]	0.56[i]	0.29[i]	0.928[i]

Numbers are those in which successful genotypes were obtained
[i] *p* value from χ^2 test between case and control group using 2 degrees of freedom.
[ii] Empirical *p* value between case and control group based on 25,000 replicates

[xx] Bonferroni correction used here. The method is described in section 1.4.4.

4.2.6 Is rs3918149 a risk factor JME in India?

We next sought to identify a third population in which to test the association with JME detected in the London and Irish cohorts. In doing so, we wished to identify a population that would maximise the likelihood of replication, should the association indeed be real.

Knowledge of risk allele frequency helps identify populations with substantially greater power of detection. We thus assessed the frequency of the risk variant (rs3918149) in healthy individuals across a global panel of populations. Results illustrated in Figure 4.3 indicate that populations of Asian ancestry appear to carry the risk allele at a higher frequency and therefore may be better powered to replicate the association[xxi].

[xxi] Power calculations are detailed in section 2.9.3. Note that as direct positive replications are better estimators of genetic effect than original reports, we used, to calculate power of detection, relative risk values as determined from the London cohort presented here.

Figure 4.3 The frequency of rs3918149 across 8 global populations

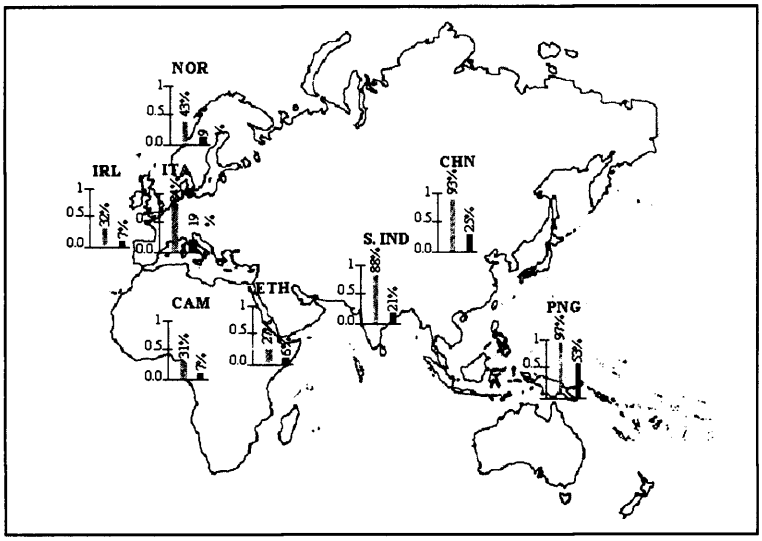


Figure 4.3 legend: A diagram illustrating the frequency of the risk variant rs3918149 across 8 global populations. Each population is represented by a bar chart indicating the frequency of the variant in black and the power of detection in grey. Parameters used in power of detection calculation are detailed in materials and methods section 2.9.3. IRL = Ireland, NOR = Norway, ITA = Italy, CAM = Cameroon, ETH = Ethiopia, S. IND = South India, CHN = China, PNG = Papua New Guinea

Given these results, we tested the association in a population of JME patients collected in Southern India. The variants failed to show significant association with JME, or even any trend in that direction. Results are shown in Table 4.8.

Table 4.8 Association analysis of rs3918149 in Southern Indian JME patients and controls.

rs3918149	Patients *n*=48	Controls *n*=144	OR[a] (95% CI)	*p* Value
Genotype				
GG	27(0.56)	82(0.57)	1	
GA	16(0.34)	46(0.32)	1.05 (0.52-2.16)	0.980[i]
AA	5(0.10)	16(0.11)	0.95 (0.32-2.83)	
Allele				
G	70(0.73)	210(0.73)	1	1[i]
A	26(0.27)	78(0.27)	1 (0.59-1.68)	

Numbers are those in which successful genotypes were obtained. [a]Odds ratio [i]*p* value from χ^2 test using 2 degrees of freedom. [ii]Empirical *p* value based on 25,000 replicates. [iii]*p* value from χ^2 test using 1 degree of freedom. [na] Non applicable.

4.2.7 A functional assessment of rs3819149

Finally, we wished to assess the functionality of rs3918149. Given the location of the variant in the promoter of the gene, we hypothesised the variant plays a role in influencing mRNA expression. To examine this, we carried out quantitative RT-PCR analysis on 40 samples of temporal neocortex from patients who had undergone resective surgery for refractory temporal lobe epilepsy[xxii]. Genotyping of these samples at the candidate causal variant, rs3918149, revealed 29 (72.5%) of the patients to be "GG" and 11 (27.5%) to be "GA". No homozygote patients for the risk allele "AA" genotype were observed.

The mRNA values were normalised over *ACTB* and shown to be normally distributed (K-S test P-value >0.05). The median expression levels for "GG" and "AG" genotypes were 1.39 ± 0.57 and 1.21 ± 0.32 respectively. The resulting ratio of mRNA expression for "AG" relative to "GG" genotype was 0.87. This result is non significant at the 5% level (K-W test, χ^2=0.64, df=1, p=0.42). Results are shown in panel "A" of Figure 4.4.

[xxii] See section 2.7 for methodology

Figure 4.4 A plot of the relative BRD2 mRNA levels in neocortical temporal lobe brain tissue categorised according to rs3918149 genotype

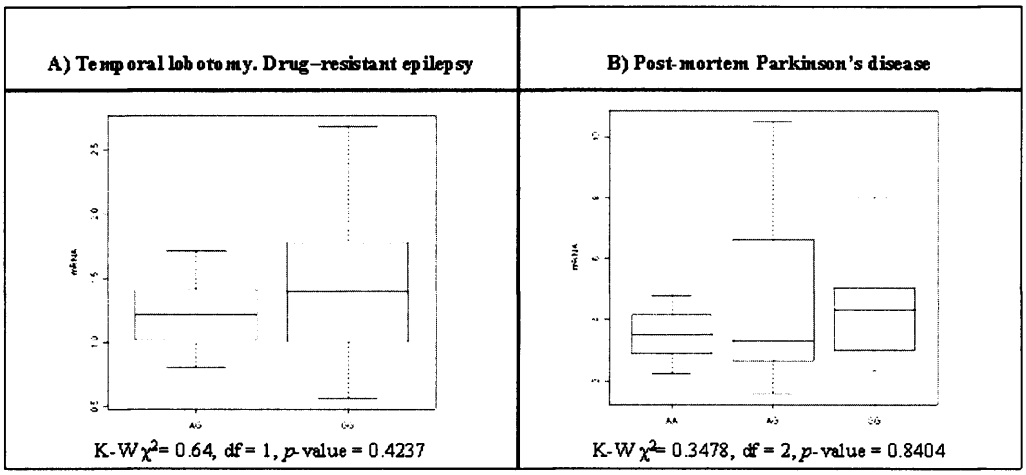


Figure 4.4 legend: A diagram indicating expression levels of *BRD2* mRNA categorised according to rs3918149 genotype. In the case of both plots, the X axis represents rs3918149 genotype, the Y axis mRNA expression levels. The mean, range and 95% confidence intervals of mRNA expression for each genotype category are indicated. Results in plot "A" are from temporal neocortical brain tissue from temporal lobe epilepsy patients having gone anterior temporal lobectomies. Results in plot "B" are from temporal neocortex brain tissue samples from a Parkinson's disease brain bank.

Given that we were limited to brain tissue from 40 from epilepsy patients, none of which carried the risk "AA" genotype at rs3918149, we assessed mRNA levels in a separate cohort of temporal neocortex brain tissue samples from a Parkinson's disease brain bank selected based on rs3918149 genotype (10 x "GG", 10 x "GA", 3 x "AA"). The median expression levels for "GG", "AG" and "AA" genotypes were 4.16 ± 1.98, 3.42 ± 2.6 and 3.52 ± 1.28 (see Figure 5). The resulting ratios of mRNA expression relative to "GG" genotype were 0.82 for "AG" and 0.85 for "AA". These results are not significant at the 5% level (K-W test, χ^2=0.35, df=2, p=0.84). Results are shown in panel "B" of Figure 4.

Our conclusion from mRNA analysis was that the rs3918149 variant does not appear to play a role in influencing expression levels of *BRD2* mRNA, at least not in the developed adult brain.

4.3 Discussion

When assessing the contribution of genetic variation to complex disease follow-up studies are essential in order to distinguish false positives from true associations. Whilst variation in several genes has been linked to common forms of IGE (*OPRM1*[176], *SLC4A3*[177], *ACP1[178]*, *LGI4* [53], *CACNA1A* [32], *GABRB3*[179], *GRIK1*[180] and *KCNJ3*[181]), few follow-up replication studies have been reported. In cases where they have, results have been conflicting (i.e. *CACNA1A* [182] and *GRIK1*[183], for review see [156]).

Taking this into account, the set of variants that have been reliably associated with complex, common diseases is still small, although growing [68]. Currently there are no examples for epilepsy among the set that appears to have broad acceptance.

We have shown a clear replication of the *BRD2* association in the London cohort and a borderline replication in the Irish population. We would consider the Irish result a very encouraging trend supporting the notion that the association is likely real. When considered in context with previous reports, these results support the

notion that variation in *BRD2* acts as a marker of increased risk for JME. Further, these results propose this association to be the most robust example to date of common variation predisposing to a form of epilepsy.

Power calculations showed a strong sensitivity to minor allele frequency. The weakened effect observed in the Irish sample may be a consequence of the slightly lower risk allele frequency compared to the London sample. Given the observation of increased risk allele frequency in Asian populations, we sought to replicate the association in a population of Southern Indian origin. Surprisingly, although we had considerably higher power of detection in the Southern Indian population (88% compared to 32% in Ireland), we found no association. Our results therefore showed the effect to be robust in populations of European origin, but apparently absent in the one Asian population studied here.

One explanation for this discrepancy is that the European results may be false positives. Isolated cases of replication do not always indicate a real effect [74], and it must be kept in mind that our sample sizes have been low. However, since three associations in populations of European ancestry have been significant, the weight of evidence would seem to favor a role for *BRD2* as a marker for JME. Should this indeed be the case, we see three possible explanations for the lack of association in the Indian cohort.

Firstly, it is possible that rs3918149 is not itself causal but rather a marker, associating through LD with the causal variant. The lack of association in the Southern Indian population would be explained by different patterns of LD between the marker and causal variant in European and Asian populations. If so, the extensive efforts in screening *BRD2* to date would suggest it more likely that the causal variant lies in a region outside of *BRD2*. Indeed, LD levels are high across the 150kb region containing *BRD2, HLA-DMA, HLA-DMB* and *HLA-DOA*. Although Pal et al. screened coding regions of these genes, it is not impossible the causal variant was missed. For example, we calculated that a SNP genotyped as part of the HapMap project, the non-synonymous variant rs1063478 lying in exon 2 of HLA-DMA, is in near complete LD with rs3918149 (D'=1, r^2=0.86) in Europeans.

The second explanation is genetic heterogeneity with different genes and/or alleles contributing to disease in different populations. Indeed, evidence for genetic heterogeneity exists from a previous linkage study where it was noted that linkage was based almost entirely on families of European origin [80, 87]. Families of non-European origin did not show evidence for linkage to EJM1 markers. Related to this, the functionality of *BRD2* variation in JME may be dependent on interactions with population-specific environmental factors. Promoter variation is arguably more likely to respond to environmental changes than other, typically functional variation (e.g. coding). As such, rs3918149 as a causal variant would fit with this theory. Recent research on the population

specificity of genetic variation functional in diseases would suggest population-specific functionality unlikely [68]. It must be noted, however, that this conclusion emerges from a lack of evidence for differences among populations rather than positive evidence that effects were the same [184].

The third explanation is that variation interacts in a trans-like manner to cause JME. That is, a general IGE risk locus interacts with a JME specific locus leading to disease development. Should the suspected general IGE risk locus be at a low frequency in a given population, then the apparent contribution of *BRD2* to the development of JME in that population would be negligible. Indeed there is support in the literature for such a scenario. Comparison of LOD peaks generated from linkage studies involving heterogeneous cohorts of IGE led to the proposal of an oligogenic model whereby variation at general risk loci interacts with variation across specific loci to produce characteristic forms of IGE such as JME, juvenile absence epilepsy and epilepsy with generalised tonic clonic seizures only ([1, 87, 95]). The EJM1 locus has been shown independently to be the site of a strong LOD peak for linkage analysis in JME samples, but the peak disappears when non-JME IGE samples are examined [85, 95, 105, 185]. The fact that we observed the association only for JME, and not other forms of IGE, is consistent with the hypothesis that EJM1/*BRD2* represents a risk locus specific to JME. Should the oligogenic model hold, it is possible that the variant with which *BRD2* interacts to confer risk in Europe is absent or rare in India. The nature of the data presented here, however, is such that we cannot confirm both

aspects of the oligogenic model, in particular the presence of general risk loci as suggested by linkage mapping [95] and follow-up association [186]. We do note, however, that we were unable to replicate the Malic enzyme 2 result (data not shown). It is now important to try and find such a variant to further understand the role of *BRD2*. In the absence of such identification or other explanation of the discrepancy, we must still consider the heterogeneity scenario as hypothetical.

The results of mRNA analysis presented here show no evidence that rs3918149 influences *BRD2* expression. However, our analysis was limited to post-mortem tissue from fully developed temporal lobe of adult brain. The effect of the agonal period on mRNA expression levels is unclear but likely substantial. In addition, when considered in the context of JME, neither refractory focal epilepsy or Parkinson's disease patients are ideal for studying the role of mRNA expression in JME. As such it is possible the variant is in fact functional, but that our assay was unable to detect an effect.

Although the results shed some light on the issue, the role of *BRD2* in JME remains unresolved. Further replication in populations of European origin would provide a solid foundation for a wider mapping project that seeks to examine, in detail, variation in surrounding genes. Indeed, follow-up mapping studies would need to be extensive as the low frequency of the risk variant would allow LD to extend for long distances both upstream and downstream of *BRD2*.

These findings suggest that association analysis in a genomic region identified (and preferably replicated) in linkage studies may be a successful approach to identify susceptibility genes for common forms of epilepsy. These results motivate the use of this strategy in other known candidate regions for IGE in which no gene has been identified yet [78, 79, 95, 187].

Chapter 5 A pharmacogenetic exploration of vigabatrin-induced Visual Field Constriction

5.1 Introduction

5.1.1 Vigabatrin

Vigabatrin is an effective anti epileptic drug (AED) licensed for use as an add-on therapy for partial epilepsy. Although its efficacy is similar to that other AEDs, it is particularly useful for treating infantile spasms where its efficacy appears to outweigh that of any other AED [188, 189]. Indeed Vigabatrin has been reported to be as effective as hormonal treatment for both controlling spasms and preventing neurodevelopmental decline in this otherwise very disabling condition [190]. In fact, it is widely known that in regions where the drug is unlicensed due to the side-effect profile, physicians regularly import Vigabatrin on a named-patient basis.

5.1.2 Vigabatrin induced irreversible visual field constriction

The use of Vigabatrin is severely limited by the development of retinopathy causing irreversible visual field constriction. This adverse reaction to the drug was first reported in 1997 [191] and is estimated to occur in approximately 40% of patients, although estimates of prevalence vary widely [135, 136, 192-197].

The natural history of the adverse drug reaction (ADR) is poorly understood, but there are a number of features that suggest the response to be idiosyncratic. Firstly, clinical experience shows that some patients, despite taking large doses of Vigabatrin over prolonged periods, fail to develop the ADR, while in others quite severe constriction appears to develop relatively rapidly. The extent of constriction can also vary markedly between affected individuals despite comparable dosing. Secondly, although most of the available data are retrospective, the findings with regard to dose-relationship are inconsistent, with a number of studies reporting the development of visual constriction to be unrelated to the daily or cumulative dose of the drug or the duration of therapy [135, 198, 199]. We wished to consider whether genetic variation in genes thought to play a role in the development and severity of Vigabatrin retinopathy could be used as a guide for predicting the severity of the ADR. Discovery of such variants may allow for safe individualized prescribing of Vigabatrin, similar to the use of mercaptopurine in chemotherapy [200], and thus the resurrection of an essentially orphaned drug.

5.1.3 The pharmacology of Vigabatrin

The pharmacology of Vigabatrin is relatively well understood, making it a good candidate drug for pharmacogenetic research. The drug works by irreversibly inhibiting GABA transaminase, the enzyme that metabolizes gamma-aminobutyric acid (GABA) the principal inhibitory neurotransmitter in the brain.

After oral intake of Vigabatrin, about 10% of the drug crosses the blood-brain barrier [188]. However, in animal models the drug has been shown to accumulate in the retina at concentrations up to 18.5 times that of brain [201, 202]. The precise mode of transport into the brain and retina is unclear but data suggest it is via one of the high affinity GABA transporters (GAT 1-4) [203-205]. In animal models, the drug appears to be taken up in the retina principally by the amacrine cells, expressing GAT-1 [205]. Vigabatrin-induced increased intracellular concentration of GABA reverses plasma membrane-expressed GABA transporter function causing GABA to be transported out of the cell resulting in a large rise in ambient GABA concentration [206]. This rise in ambient GABA increases tonic inhibition via its action on post-synaptic GABA receptors, including $GABA_A$ and $GABA_C$ receptors. $GABA_C$ receptors are a specialized type of GABA receptor located almost exclusively in the retina [207]. The highest concentration of these receptors is on axon terminals of rod bipolar cells, the predominant neuron in the peripheral retina, the region principally affected by Vigabatrin toxicity, as evidenced by the loss of peripheral vision. Elevated concentration of GABA can reverse chloride currents at GABA receptors, causing depolarization of the cell leading to a potentially cytotoxic calcium influx [208]. Although speculative, the drug may cause retinal toxicity in this manner.

5.1.4 Selection of candidate genes

Selection of our candidate genes focused on presumed transporters of Vigabatrin into the retina, its target enzyme GABA transaminase, and the $GABA_C$ receptor. We selected the following candidate genes: GABA transporters *GAT1-3* (*SLC6A1, SLC6A13, SCL6A11)*, GABA transaminase (*ABAT),* and the rho subunits of the $GABA_C$ receptor (*GABRR1* and *GABRR2*). An overview of the selection of candidate genes is shown in figure 5.1

Figure 5.1 Diagram of the retina showing sites relevant to pharmacogenetic study of vigabatrin toxicity

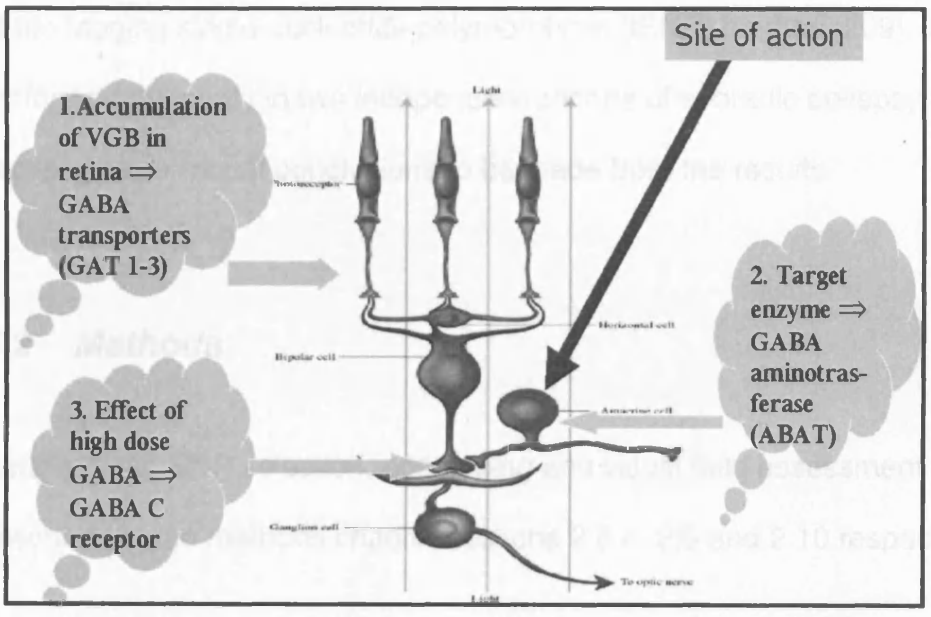


Figure 5.1 legend: (1) Vigabatrin enters the retina via GABA transporters. (2) Within amacrine cells, Vigabatrin targets the enzyme GABA transaminase. (3) The resultant high ambient GA BA concentrations may result in toxicity through its effect on $GABA_{C\ receptors}$ located at the terminal of bipolar cells.

5.1.5 Aims

We wished to test the hypothesis that common, functional variation in these candidate genes increases the risk of developing this ADR. Specifically, we are interested in detecting variants of strong, clinically relevant effect; variation that would be beneficial in guiding clinicians in prescribing Vigabatrin. We thus examined the correlation between variation in selected candidate genes and the amount of Vigabatrin-induced visual constriction. To ensure a thorough analysis, we have availed of recently made available HapMap data allowing the application of the tagging single nucleotide polymorphism (tSNP) method [209]. We have performed the study in two independent cohorts of sporadic epilepsy patients, allowing more robust conclusions to be made from the results.

5.2 Methods

Details of tag SNP selection, genotyping and visual field assessment are described in the methods chapter sections 2.8.4, 2.3 and 2.10 respectively

5.2.1 Subjects

The first cohort, termed the *Dublin* cohort, consisted of 73 patients recruited at the epilepsy clinic at Beaumont Hospital, Dublin, Ireland. The second cohort, termed the *London* cohort, consisted of 58 cases recruited at the epilepsy referral centre at the National Hospital for Neurology and Neurosurgery, London, United Kingdom. Patients were selected based on a history of current or prior exposure to Vigabatrin without prior knowledge of the results of visual field testing. To be included in the study patients must have been exposed to Vigabatrin for at least one year. For each patient we recorded age, sex, epilepsy diagnosis, duration of Vigabatrin therapy, cumulative dose and exposure to other medications. We included patients who were not currently taking Vigabatrin as the general consensus in the literature is that, with the exception of isolated case reports, the development of visual constriction is irreversible [210, 211]. Patient characteristics for the two groups are shown in Table 5.1.

Table 5.1 Summary of patient dosing data for *Dublin* and *London* cohorts

	Dublin (n=73)	*London (n=58)*
#Male/#Female	38/35	33/25
Mean Age (years)	36.7	38.1
Mean VGB Daily Dose(mg)	2060.4	2663.8
Mean VGB Duration (years)	7.2	6.1
Mean VGB Cumulative Dose (g)	5380.8	6502.2

VGB - Vigabatrin

5.2.2 Study design

We applied, using HapMap data, a tagging SNP strategy that seeks to represent most, or all, common variation across the candidate genes. Data generated by the HapMap project allow the quantification of linkage disequilibrium (LD) between variants and hence the selection of tSNPs for any gene or region across the genome.

In the *Dublin* cohort we examined the correlation between tSNPs and Vigabatrin-induced visual constriction. In the *London* cohort we attempted replication of any significant associations found in the *Dublin* cohort. We considered for replication any variant or haplotype associating significantly with the ADR in the *Dublin* before correction for multiple testing[xxiii]. We then corrected in the *London* cohort, but only for the number of tests undertaken in that same cohort. This methodology simultaneously reduces the incidence of type-1 and type-2 errors.

5.2.3 Statistical analysis

Association was examined at both the single SNP and haplotype level. We assessed the correlation between single SNPs and MRD using multiple linear

[xxiii] Using Bonferroni method

regression. We modeled considering both additive and dominant effects for each tSNP. We assessed the correlation of haplotypes generated by tSNPs with MRD values by calculating score statistics using the package haplo.score in 'R' [128, 212]. Briefly, this method uses generalized linear models to generate score statistics that can be used to examine the correlation between expectation-maximization (EM) generated haplotypes and the trait of interest. When generating score statistics, we only considered haplotypes observed at a frequency of 1% or greater. The likelihood of spurious inference by the EM algorithm increases with lower frequency haplotypes. MRD values were normalized where required.

Given that we intended to replicate any association detected in the *Dublin* cohort, we did not correct for multiple testing in that cohort and applied a significance threshold of 0.05. We considered as significant in the *London* cohort, only those associations that remained significant after Bonferroni correction for the relevant number of tests in the *London* cohort.

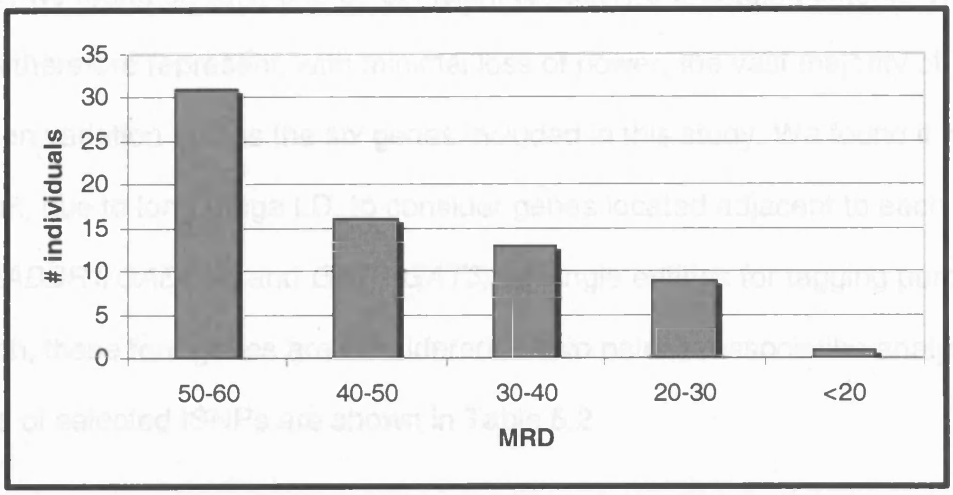
5.3 Results

5.3.1 Summary of clinical data

Demographic and drug-exposure data for both cohorts are shown in Table 5.1. In *The Dublin cohort*, 64 of the 73 patients were taking Vigabatrin at the time of their initial assessment. Using the estimated unaffected population cut-off of 49.3 degrees, 39 of the 73 patients in the *Dublin* cohort (53%) were considered as having constricted fields. In the *London* cohort, 45 of the 58 patients were taking Vigabatrin at the time of their initial assessment. Using the estimated unaffected population cut-off of 47.6 degrees, 45 of the patients in the *London* cohort (76%) could be considered as having constricted fields. The distribution of MRD values in both populations is shown in Figures 5.2 and 5.3[xxiv].

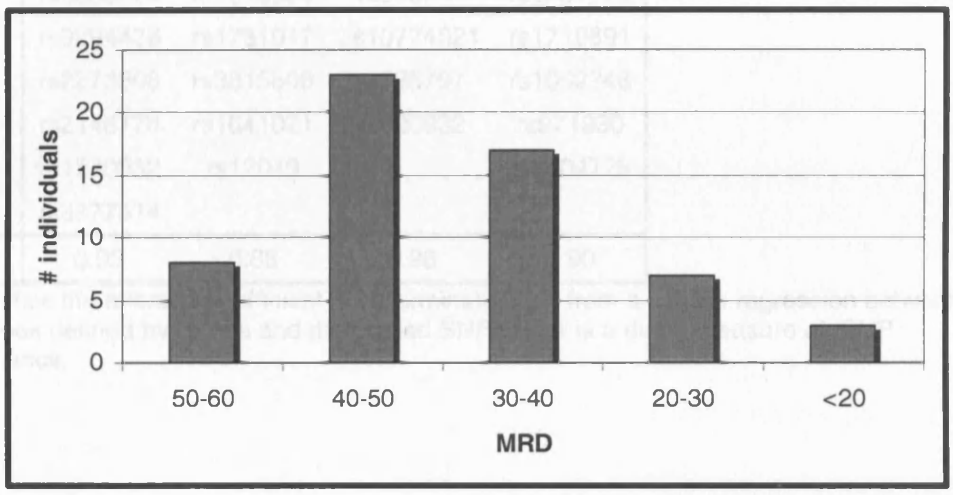
[xxiv] Note that as a different isopter was used in each cohort, direct comparison of figures 5.2 and 5.3 is not possible

Figure 5.2 Distribution of mean radial degrees as measured by III4e isopter from Goldman field tests in the *Dublin* cohort



MRD – Mean radial degrees. The MRD value represents the level of visual field constriction. The lower the MRD value, the more constricted the visual field. (The technique for calculating MRD is described in section 2.10.1).

Figure 5.3 Distribution of mean radial degrees as measured by I4e isopter in the *London* cohort



MRD – Mean radial degrees. The MRD value represents the level of visual field constriction. The lower the MRD value, the more constricted the visual field. (The technique for calculating MRD is described in section 2.10.1).

5.3.2 tSNP selection

In total, we found 32 tSNPs to be sufficient to satisfy our tagging criteria. These tSNPs therefore represent, with minimal loss of power, the vast majority of common variation across the six genes included in this study. We found it more efficient, due to long range LD, to consider genes located adjacent to each other (i.e. *GABBR1/GABBR2* and *GAT1/GAT3*) as single entities for tagging purposes. As such, these four genes are considered as two pairs in association analysis. Details of selected tSNPs are shown in Table 5.2

Table 5.2 Details of tSNP selection

tSNP#	*GABRR1/2*	*ABAT*	*GAT2*	*GAT1/3*
1	rs2297309	rs1345300	rs495360	rs1485142
2	rs1796740	rs3095512	rs2289954	rs1881354
3	rs282123	rs4984996	rs124440	rs746278
4	rs6902106	rs1640989	rs555044	rs2601126
5	rs9294426	rs1731017	rs10774021	rs1710891
6	rs2273508	rs3815508	rs525797	rs1062246
7	rs2148174	rs1641021	rs520932	rs971930
8	rs1570932	rs12049		rs2304725
9	rs3777514			
~r^2	0.92	0.88	0.98	0.90

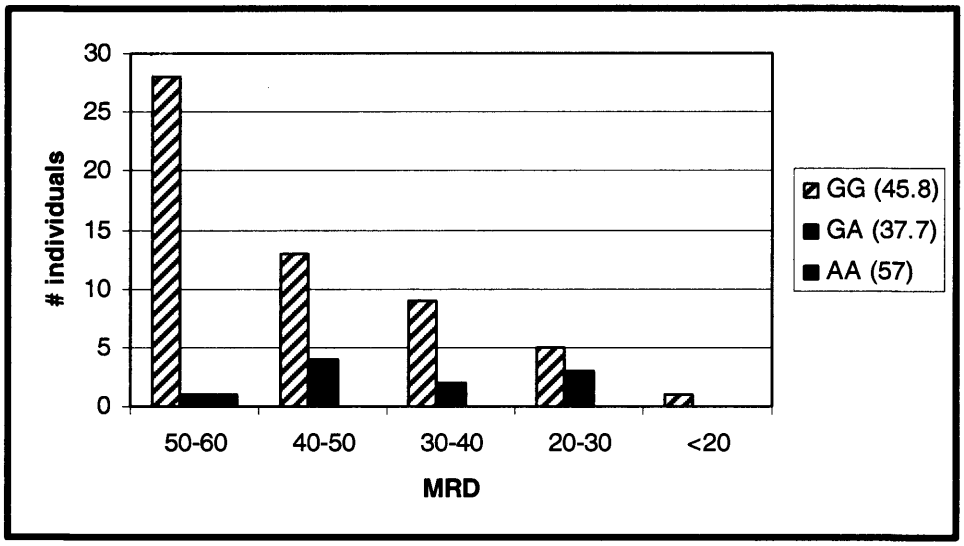
~r2 signifies the average coefficient of determination (r^2) from a logistic regression between the haplotypes defined by tSNPs and the tagged SNPs. This is a direct measure of tSNP performance.

5.3.3 Single SNP association analysis

Single SNP regression analysis in *The Dublin cohort* resulted in the detection of three significant associations between single tSNPs and MRD values (*GABRR1/2* tSNP6, *GAT1/3* tSNP7 and *GAT2* tSNP6).

In the case of *GABRR1/2* tSNP6, an intronic variant, the minor allele ("A") was associated with an increased risk of developing visual field constriction ($p=0.027$). The average MRD value for the ten individuals carrying the heterozygote genotype was 37.7 degrees compared to 45.8 degrees for the major allele homozygote. Although the sole individual carrying the homozygote risk genotype ("AA") did not show constriction (MRD = 57 degrees), it is impossible to make conclusions based on this one observation. A plot of MRD values against *GABRR1/2* tSNP 6 genotype is shown in Figure 5.4

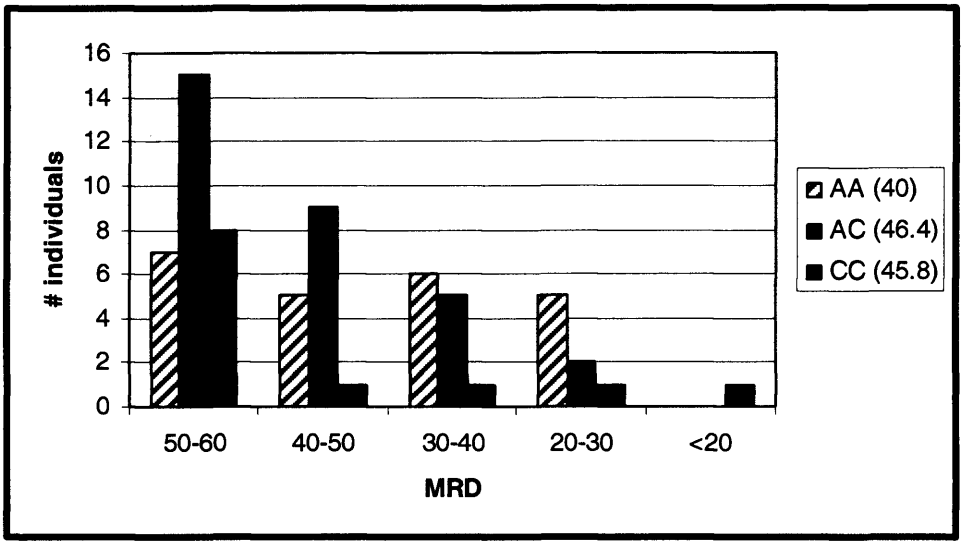
Figure 5.4 Distribution of mean radial degree values with GABRR1/2 tSNP 6 genotype in the *Dublin* cohort.



MRD – Mean radial degrees. Values in brackets represent mean MRD value for that genotype.

In the case of *GAT1/3* tSNP7 (intronic), the major allele ("A") was associated with increased risk of developing the ADR (p=0.047). Homozygote carriers of the risk genotype ("AA") had an average MRD of 40 degrees compared to 45.8 degrees for carriers of the homozygote protective "CC". A plot of MRD values against GAT1/3 tSNP7 genotype is shown in Figure 5.5.

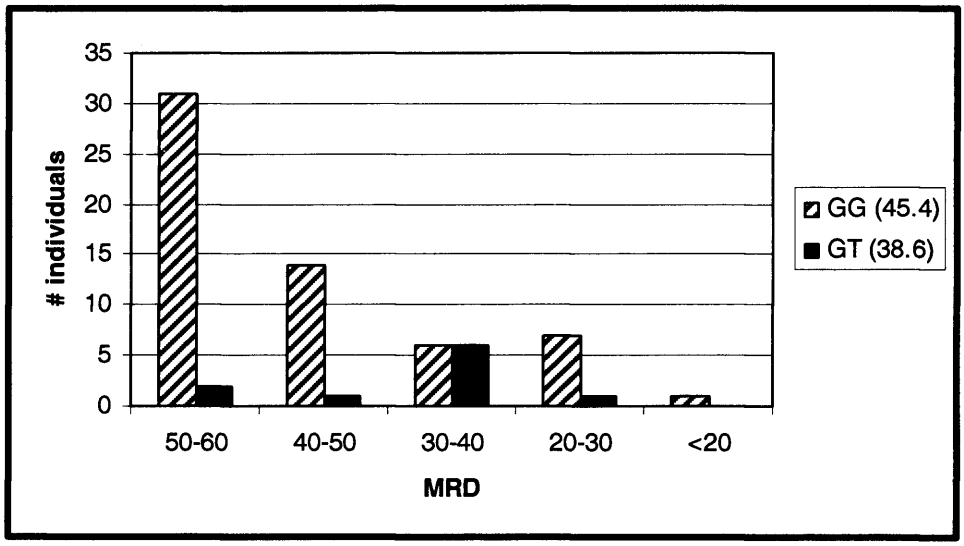
Figure 5.5 Distribution of MRD values with GAT1/3 tSNP 7 genotype in the *Dublin* cohort



MRD – Mean radial degrees. Values in brackets represent mean MRD value for that genotype.

Finally, the minor allele ("T") of *GAT2* tSNP6 (intronic) was associated increased risk with heterozygote carriers having an average MRD value of 38.6 degrees (p=0.05). No individuals homozygote for the risk allele were observed in *The Dublin cohort*. A plot of MRD values against *GAT2* tSNP6 genotype is shown in Figure 5.6.

Figure 5.6 Distribution of MRD values with GAT2 tSNP 6 genotype in the *Dublin* cohort.



MRD – Mean radial degrees. Values in brackets represent mean MRD value for that genotype.

5.3.4 Haplotype association analysis

At the haplotype level, the global statistic generated by *GAT2* haplotypes showed a significant correlation with MRD values (p=0.012). This result would suggest that either a single haplotype or a combination of haplotypes is correlating with MRD values. Indeed, one haplotype of *GAT2* showed a significant correlation, with carriers of this haplotype having a lower MRD reading than the Vigabatrin-exposed population average (p=0.039). Haplotypes generated by other tSNPs failed to show significant correlation with MRD values.

5.3.5 Attempted replication in the *London* cohort

We carried forward these five significant results for replication in the *London* cohort. However, we failed to replicate or show any trends towards significance for any of the single SNP or haplotype associations detected in the *Dublin* cohort.

5.4 Discussion

We have tested the hypothesis that common variation of strong, clinically relevant effect, exists in the selected candidate genes. Detection of such clinically relevant variation might allow the safe prescribing of this drug to a subset of patients. In the *Dublin* cohort, we detected three single SNP variants, and two haplotypes that showed significant association with the extent of Vigabatrin-induced visual field constriction. However, we were unable to reproduce these results in the *London* cohort suggesting that the initial findings are likely to represent false positive results. As such, our results do not the support the notion of clinically relevant variation residing in these genes.

Although we feel that the marginally significant associations detected in the *Dublin* cohort are likely to be false positives, we have not definitively ruled out an effect for these variants. It is entirely possible that these variants are indeed functional, or perhaps more likely tag functional variants of small to medium effect that influence the development of Vigabatrin-induced visual field constriction. Should this be the case, our study is clearly underpowered. Similarly, as the tagging method is known to perform poorly in capturing low

frequency variation, we have not ruled out the presence of rare functional variation in these six genes. A large patient population would help address the issue of power but such a collection is difficult to collect, given that few individuals are currently exposed to, or initiated on Vigabatrin.

Natural selection dictates that it is extremely rare for disease causing variation of strong effect to be observed at a high frequency in a population. Variants associated with altered drug pharmacology, on the other hand, may be evolutionarily silent in the absence of a substrate (in this case the drug Vigabatrin) and thus variants of strong effect may drift to high frequency in the population, as is evident from pharmacogenetic studies of chemotherapeutic agents [213-215].

Other confounding factors must also be considered as a potential explanation for our failure to detect any clear associations. Although every effort was made to exclude unreliable Goldmann fields, the subjective nature of the test introduces unavoidable variation to the measurement. For example, the rate at which the test object is moved, and the promptness with which the subject reports its appearance are not always consistent. This is especially true for patients where only one test was recorded. Even in subjects without retinal defects, repeat testing has shown that visual fields may vary by up to 14% in individuals [216].

Further, although prospective data on the natural history of the retinopathy are lacking, case reports do exist that suggest, at least in some cases, substantial

recovery of vision after discontinuation of the drug [217, 218]. If this were the case, any genetic signal would be obscured under our current study design. These isolated cases may have resulted from erroneous recording of either the initial or follow-up result, but one cannot out rule the possibility that a small subset of patients exist in whom real improvement does occur.

We felt that our gene selection represents the most appropriate front-line candidates based on what is understood of Vigabatrin pharmacology, but it is entirely possible that variation in other genes is responsible for the ADR. There is some evidence from mice models to suggest that the GAT-4 (BGT-1) transporter may be responsible for GABA (and thus possibly Vigabatrin) transport across the blood-brain barrier [203]. $GABA_A$ receptors are also present in the retina and variation within one of the many subunits of this receptor may influence toxicity. Mutations in the ornithine decarboxylase gene are known to cause rare cases of retinal degeneration and Vigabatrin is known to weakly inhibit this enzyme [188]. Indeed a previous study examined, in a small number of patients, the relationship between a limited number of variants in this gene and Vigabatrin-induced visual field defects [219] but no effect was reported.

Finally, its possible that, despite that apparent idiosyncrasy of the reaction, genetic factors do not play a major role and that some other unidentified factor such as a drug interaction or environmental agent, or perhaps a combination of factors predispose to the development of toxicity. Identification of either genetic

or environmental predisposing factors may allow for safe prescribing of an effective antiepileptic medication in selected individuals.

Chapter 6 Project design of a large scale map based association study

6.1 Introduction

In previous chapters I have highlighted some of the critical issues facing the field of epilepsy genetics such the need to increase sample sizes and incorporate map based association in to study design. In this chapter I will describe the initial stages of a project that seeks to integrate these lessons in to a large scale association study. The result should be a massive improvement over previous efforts in terms of overall study power and scope of variation examined.

6.1.1 Concept and overall study design

Our goal was to assemble the largest collection of epilepsy samples to date in an effort to identify variation across a panel of candidate genes that might predict the development or guide in the treatment of common forms of epilepsy. Coinciding with efforts to assemble this cohort came the release of phase one of the HapMap data which effectively opened the door to tagging any gene or region in the human genome. Similarly, recent technological advances have resulted in great reductions in the cost and time required for genotyping.

We wished to tag all common variation across a list of candidate genes but also directly examine known functional variation. Our approach was to a) assemble the list of candidate genes, b) use bioinformatics techniques to identify functional variation in each of the candidate genes, c) use the HapMap data to select tSNPs for each candidate genes but where possible, forcing functional variation identified in step 'b' as tags and finally d) genotype and analyse the data.

Since the concept of tagging SNPs was first proposed in late 2001, a diverse range of tagging SNP selection strategies have been developed [126, 129, 220]. Much work has been carried out in an effort to quantify the different methods in terms of efficiency to predict tagged variants. However, the critical issue for association work is not how well variants are predicted but rather what is the overall power of the selected tSNP set to detect association driven by an unknown (or invisible) functional variant. Given that little work had been carried out to evaluate the question of power of a given tag set, a number of experiments were run with the aim of selecting the optimum tagging method for the situation presented by this study.

There are a range of questions we would ideally like to address using genotype data generated from this study. The first and perhaps most straightforward in terms of phenotype collection is the question of predisposition to seizure and syndrome type. Extracting high quality phenotypes for pharmacogenetic questions tends to be more difficult as the phenotype must be defined in a

manner that maximises any potential genetic effect. Oftentimes the definition that is settled on for a given pharmacogenetic question is not in line with that stored in current phenotype databases. To further complicate matters, even if the phenotype were available at one collection centre, it is often not from another. These issues are certainly surmountable; it just takes a longer period of time to assemble the phenotype than for the predisposition question.

6.2 Methods

6.2.1 Selection of candidate genes

Candidate gene selection was based on current biological knowledge of epilepsy. This largely stems from Mendelian genes identified by positional cloning. As discussed in the introduction to this thesis (see section 1.3.1 and Table 1.1), positional cloning has identified a number of genes that have in turn highlighted key biological pathways involved in epilepsy. For example, many epilepsy causing mutations are located in genes that form subsets of the voltage gates ion channels, critical in the generation and propagation of action potentials. Included in the candidate gene list therefore, are all the voltage gated sodium (11), calcium (26) and chloride (11) channels as well as a subset of the large potassium channel family (22). Due to restrictions in space, the potassium channel genes were limited to those coding for subunits of types A,C,D,Q,AB and BK. Also included were key members of the calcium regulated potassium

channel, the inwardly rectifying channel and the hyperpolarization activated-cyclic nucleotide gated potassium channel.

The Mendelian genes listed in Table 1.1 also confirmed the importance of several key neurotransmitters in epilepsy such as acetylcholine, GABA and glutamate. Included in the candidate gene list are the principal metabolisers and transporters as well as key receptors for each of these neurotransmitters.

Apoptosis is receiving increasing attention in epilepsy and there is a growing suspicion that it plays an important role in epileptogenesis [221]. Key genes from the rapamycin signalling pathway, the related PI3K/PTEN/AKT pathway, the JAKs and STATS pathways were included in the candidate gene list. These pathways have been shown, through animal models to play a role in epileptogenesis [222].

All the principal metaboliser and transporters of the principal anti-epileptic drugs were included with a view to pharmacogenetic studies. We paid particular attention to leviteracetam (commonly known as Keppra).

6.2.2 Population sample

To date, over 3400 patient samples have been contributed from 7 collection centres across the globe. In order to reduce stratification, we have targeted samples of Caucasian origin. However, we have recruited additional

collaborators who have assembled cohorts of non-Caucasian origin that would allow testing of suspected functional (in terms of predisposition and treatment) variation in different ethnic groups.

Table 6.1 Summary of DNA sample contributions

Name	Ethnicity	# Patients	# Controls
UC London, UK	UK	833	
British 1958 birth cohort, UK	UK		2112
Beaumont Hospital, Dublin, Ireland	IRL	835	719
University of Anderlecht	BEL	234	
University of Antwerp	BEL	91	
University of Melbourne	AUS	220	
University of Melbourne	AUS	423	
GSK (Glaxo S K)	FIN	490	490
Total		3126	3321

UK – United Kingdom, IRL – Ireland, BEL – Belgium, AUS – Australia, FIN – Finland
Note that non-Caucasian populations have been omitted from this table

6.2.3 Retrieving the genomic location of each gene

In order to access the relevant data for functional and tagging SNP selection, it is first necessary to define the genomic region of each gene. In defining the genomic region, we wished to include rare isoforms of each gene. The 'ref seq' track from UCSC defines known protein gene locations by aligning mRNA from reference sequence collections. As 'ref seq' defines genomic locations using mRNA as opposed to protein alignment, rare isoforms are also represented. Additionally, as there can be several different 'refseq' entries for a given gene, the minimum of the start column to the maximum of the end column was taken as the actual genomic location. This is illustrated in Table 6.2 using the example of *OPRM1* below which was assigned the location: chromosome 6: 154,391,433 -

154,598,991.

Table 6.2 Assigning genomic location from ref seq data

GENE	ACC#	CHR	Tx Len	Start	End
OPRM1	NM_0010085	chr6	52,275	154,391,433	154,443,708
OPRM1	NM_0010085	chr6	54,213	154,391,433	154,445,646
OPRM1	NM_000914	chr6	79,864	154,391,433	154,471,297
OPRM1	NM_0010085	chr6	202,358	154,396,633	154,598,991
OPRM1	na	ch6	207,558	154,391,433	154,598,991

Acc# - NCBI accession number. CHR – Chromosome. Tx Len = Transcript length. Start = transcript start point (NCBI build 34). END = transcript stop point (NCBI build 34)

Using this methodology, the genomic address of each gene was recorded and used for both functional analysis and tSNP selection (described below in sections 6.2.2 and 6.2.5)

6.2.4 Selection of functional SNPs

Functional SNPs were identified and assembled using the bioinformatics tool TAMAL, developed at University of Chapel Hill, North Carolina [223]. The tool collects all SNPs from a number of publicly available databases then reduces the list to a set of functional variants (sometimes putative, sometimes biologically proven) according to user defined criteria.

The tool first collects all SNPs from the following freely accessible SNP databases HapMap (www.hapmap.org), Perlegen (www.genome.perlegen.com). This inclusive list is reduced to a set of functional SNPs for genotyping if there is evidence that the SNP is a) polymorphic in Caucasians and b) if the SNP is

functional according to the criteria defined by the user. The latter criterion refers to SNPs that: i) lead to non-synonymous or synonymous amino acid changes augmented with in silico prediction of functionality [224] or alteration of a splice site; ii) are in a promoter region (in silico prediction, but with biological validation)[225]; iii) are in regions with conservation scores =99th percentile genome wide for human-chimp-rat-mouse-chicken alignment via a hidden Markov model [226], iv)predicted regulatory potential [227], or v) a transfactor binding site [228]. An explanation of each of these functional criteria follows.

Coding variation (i.e. synonymous and non-synonymous) was accessed from the LS-SNP database (www.alto.compbio.ucsf.edu/ls-snp/). This database consists of all known coding variation supplemented with a prediction of functionality. The curators of the database have implanted a three module computational pipeline. In the first module they independently create a database of coding variation by extracting the genomic locations of all dbSNP variation (validated and non-validated) and map this on to human protein sequences in the Swissprot and TrEMBL databases. In the second module a database of high resolution protein structural information is created. The difficulty here is that only around 8% of all known proteins have been extensively studied using methods that define high resolution protein structure such as x-ray crytallography or nuclear magnetic resonance spectroscopy [229]. The curators attempt to address this issue by inferring the structure of a poorly studied protein from the structure of a putatively homologous protein solved by the previously mentioned experimental methods.

In the third module the curators use the outputs of the first two modules to compute a variety of annotations for each SNP examined including whether the SNP is near a ligand or a domain-domain interface as well as whether the SNP is putatively destabilizing. The database is updated twice a year.

Variation in promoter regions was defined by mapping SNP target sequence to promoter sequence at the Stanford promoters website (http://www-shgc.stanford.edu/). This database contains bioinformatically inferred promoter sequence (likely containing basal and upstream regulatory elements) from over 10,000 genes [225]. The database has been shown, through in-vitro confirmation, to be 91% accurate in predicting core promoter elements.

Variation lying in conserved sequence regions was determined by aligning SNP target sequence against the 'Evolutionarily Conserved Elements' database at (http://www.soe.ucsc.edu/~acs/conservation/). This database consists of over 1.31 million elements conserved across human, mouse, rat, chicken and *fugu rubripes.*

Polymorphisms lying in the critical regions transcription factor binding sites were identified using the TRANSFAC database (http://www.gene-regulation.com/). This database contains sequence information on bioinformatically inferred *cis*-acting DNA elements and *trans*-acting factors.

Selected SNPs genotyped by the HapMap consortium and observed with a frequency greater than 5% were set aside to be forced as tagging SNPs (see section 6.2.5 below). All other variants (i.e. those not genotyped by HapMap and those with a minor allele frequency less than 5%) were to be genotyped 'on the side' of tagging SNPs.

6.2.5 Selection of tagging SNPs

Tagging SNPs were selected using the allelic r^2 (or locus scoring r^2) method as detailed by Chapman et al [131]. This method is similar to the haplotype r^2 method in that it employs logistic regression. However, the regression model is fitted not to haplotype groups but to each tSNP entered as separate additive terms. The principal advantages of allelic, over the haplotype r^2 method are i) the abolishment of haplotype inference, ii) fewer predictor variables in the regression modal (thus less chance of over-fitting) and iii) fewer degrees of freedom in the association test. As allelic r^2 is, similar to haplotype r^2, a multimarker test, it would perform better than single locus methods such as pairwise- r^2 in predicting unknown variation.

6.2.6 Selection and filtering of HapMap data

The relevant SNP data was first extracted from HapMap data using the gene coordinate details as defined in section 6.2.3 and illustrated in Table 6.2. From this dataset was removed the following: SNPs with a minor allele frequency less than 5%, SNPs with greater than 10% missing data, SNPs that violated Mendelian inheritance in one or more trio units and SNPs which violated Hardy-Weinberg equilibrium with a significance values greater than p=0.01. In order to reduce computational burden large genes were broken down in to sub-regions containing a maximum of 52 SNPs.

6.2.7 Criteria used for tSNP selection

The following criteria had to be satisfied for tSNP set selection: any variation defined as functional in our bioinformatics screen and genotyped by HapMap was to be forced as a tSNP, every tagged SNP to be captured with a minimum allelic r^2 of 0.8 (that is, the coefficient of determination returned by the regression model must be greater than 0.8 for all tagged variants), a maximum average allelic r^2 of 0.3 between the selected tSNP set and a panel of unlinked variants (this test controls against over-fitting of the model).

6.2.8 Selection of neutral SNPs

Populations for this study although of Caucasian ancestry, have been assembled from five different geographic locations (see Table 6.1). Cryptic stratification

within a population can adversely affect results of an association study driving either false positive or false negative results depending on the nature of the substructure. Clearly some degree of stratification will exist between the five populations. Indeed stratification within any one of the populations is also a possibility. Knowledge of underlying structure is best estimated using a panel of neutral markers. We have thus included a panel of 100 neutral markers chosen as follows:

a) A panel of 300 SNPs were selected independent of Fst values from the HapMap dataset. SNPs were required to be at least 50kb from any known genes and have a minor allele frequency greater than 5%.

b) Of this panel of 300 SNPs, 100 SNPs, spread evenly across chromosomes 1-22 were selected for genotyping.

Estimation of population substructure is possible using the program STRUCTURE [230]. This information could be used as a guide in grouping or dividing populations for association analysis. The degree of structure within any one population can be estimated and used to correct association statistics. A brief discussion of relevant methodology is covered in section 1.4.4.

6.3 Results

6.3.1 Candidate gene selection

Four hundred and fifty eight candidate genes were selected. For each gene, genomic coordinates on NCBI build 34 were extracted from the UCSC database

and ten kibobases of upstream and downstream sequence added in order to capture regulatory regions. In total, these genes represent 54.4mb of sequence, an average of 119kb per gene. A small number of genes were not suitable as they had not been annotated on NCBI build 54. This set represents the vast majority of the majority of biologically plausible candidate genes for predisposition to, or treatment of, common forms of epilepsy. Details of selected genes and coordinates are shown in Table 8.2.

6.3.2 Functional analysis

Bioinformatic screening for functional variation across the 458 candidate genes resulted in the identification of 1650 putatively functional SNPs. On average, four functional variants were selected for each gene. Eight hundered and seven of the 1650 functional variants were genotyped by HapMap and thus forced as tags in SNP selection. Details of selected variants are shown in Table 6.3.

6.3.3 Tagging analysis

A total of 5447 tags, an average of 12 SNPs per gene, were found to be sufficient to represent variation across the 458 candidate genes. Forced functional variants accounted for 15% of all tagging SNPs. This level of efficiency, when extrapolated to a genome wide level (estimated at 3.07×10^9 at NCBI build 35), would predict approximately 308,000 tags to cover the entire genome. Genome-

wide tagging, but of genic regions only is an alternative to exhaustively tagging the entire genome. The UCSC genome browser has annotated 39,368 genes accounting for $2.33x10^9$ (equivalent to 75% of the genome) base pairs of sequence. However, many genes overlap or are nested within each other which reduces the actual amount of genic sequence considerably. In fact, it has been estimated that genic regions cover only 25% of the genome [231]. By extrapolating using a figure of 25%, 77,000 tags would be sufficient to tag all known .genes in the human genome

Table 6.3 Summary of functional SNP analysis

SNP class	SNP count[1]	HM[2]
Protein	459	201
Coding	292	147
Promoter	319	190
TFBS	230	105
Regulatory	92	38
Conserved	258	126
Total	1650	807

[1]All functional variants detected. [2]Functional variants also genotyped by HapMap
Protein = alters amino acid or protein stability. Coding = Located in an exon
Promoter = Located in inferred promoter region
TFBS = Located in inferred transcription factor binding site
Regulatory = Located in inferred regulatory region.
Conserved = Located in a region conserved at the 95 percentile across human-chimp-rat-mouse chicken

6.4 Discussion

The set of 6289 tagging and functional variants identified here represent a comprehensive assessment of variation across 489 candidate genes for epilepsy. The method described merges a sequence based, direct assessment with an indirect map based tagging SNP approach thus examining directly

putatively functional variation whilst at the same time capturing variation of unknown significance.

The broad approach taken to selection of putatively functional variation covers the principal functional groups identified to date; coding, promoter, regulatory and conserved. However, as information on SNP frequency in a Caucasian population was an inclusion criterion, the screen was limited to variation deposited in the two suitably annotated SNP databases – HapMap and Perlegen. Although these databases represent a major component of human variation, they are not complete. The alternative of including all functional variation ran the risk of inflating the proportion of non-variable SNPs.

There are numerous methods of tagging and each method has advantages and disadvantages. We chose allelic r^2 for three principal reasons, all related to the multi-marker nature of the method. Firstly the method is efficient in that it allows tags to work in combination to predict tagged variation. Secondly, it is powerful in detecting association as the number of degrees of freedom in the model is reduced when compared to other multimarker methods (for example haplotype r^2). Thirdly the method tends to behave better in capturing unknown variation than single marker methods. This is particularly important as we used Phase 1 HapMap data to run these experiments. Capturing unknown variation is less of an issue as SNP density is increased (HapMap consider Phase 2 to cover all common coding variation and the vast majority of non-genic variation). In

addition, haplotype inference is not an essential component of the allelic r^2 method. This greatly simplifies issues as haplotype inference can become unreliable across regions of complex LD structure.

Association studies are becoming increasingly large scale. The additional power and scope of such studies should begin to identify numerous loci across the genome that are contributing to the development and treatment of common disease. There currently are two approaches to large scale studies – candidate gene or genome wide. Whilst genome wide studies obviously have a larger scope, it would seem logical to first run large scale candidate gene studies as power of detection for the specific regions examined would be greater than the same regions by genome wide study. The correct execution of a candidate gene study should solidify current knowledge thus paving the way for examination or previously 'unsuspected' regions by genome wide association.

Indeed, considering the current extent of our knowledge, a third approach might be considered, one that examines, using a framework similar to that described here, all known genes across the genome. The advantage of this method is that whilst all coding regions are examined, power of detection is focused on coding regions as opposed to diluting it by including the 75% of the genome that does not contain any known genes. Such an approach raises interesting questions. Although the amount of sequence covered by genes is equivalent to 75% of the genome ($2x10^9$bp), the fact that some genes overlap reduces the actual size of

the region covered to 25% ($7.5x10^8$bp). The region could be reduced further by restricting the amount of intronic sequence examined.

The integration of Bayes theory to genome wide studies might allow the researcher to factor in biological knowledge a priori. Such an approach would allow known functional variation as well as key regions of the genome to be weighted thus maximising power of detection.

In a sense the candidate gene approach described seems logical as a first large scale study not only in that it clarifies what we currently suspect of disease pathways but also because the performance of current genome wide genotyping assays is not well understood. For example, most groups currently running genome wide studies avail of the Affymetrix chip systems. Whilst the number of variants examined is obviously impressive, the performance of those variants against a carefully chose tagging set is unknown. In the near future more carefully designed chips will become available that address many or all of these concerns.

Chapter 7 Discussion and conclusions

A pubmed search using the keywords 'epilepsy', 'association study', 'genetic' returns 136 listed publications, yet the field is little better off now than when it first embarked on mapping genetic variation for common forms of epilepsy. We know from heritability studies that genetics explains a significant proportion of the risk of developing epilepsy, the question we are attempting to address is thus fundamentally valid. The challenge now is to implement lessons learned from past difficulties to future study designs.

In epilepsy, as with many other disease fields, the so called 'low hanging fruit' (common variants of large effect) have not been identified. The apparent lack of such variation does not necessarily contradict the observation from heritability studies that a large component of epilepsy is genetic, in that the heritability of a trait explains nothing of the underlying genetic architecture. Indeed it is becoming more apparent that for complex disease the underlying architecture is seemingly characterised to a large extent by interaction between numerous small and medium effect variations. The effect of any single variant is thus marginal, making detection difficult.

With this in mind it would seem reasonable to view past epilepsy studies as exploratory. Although we have advanced our biological knowledge little, we have learnt the key lesson that power of detection must be increased through the

refinement of study design. Deciphering this complicated genetic signal will require developments on three fronts, all focused on maximising power of detection: a) increasing study cohort sizes; b) refining phenotype definition and c) improving genetic mapping techniques.

7.1.1 Increasing study cohort sizes

The broad phenotypic heterogeneity of epilepsy is such that any one phenotypic group represents a small proportion of the overall cohort size from a typical referral centre. For example, the cohort at the Institute of Neurology in London contains over 2000 individuals, a figure that initially appears substantial. However this figure must be contextualised by the underlying characteristics of the genetic component to epilepsy. Broad genetic risk factors (i.e. something that predisposes to epilepsy in general) probably do not exist, rather it is thought that genes play a role in developing specific syndrome or seizure types. When viewed in this way a 'large' cohort quickly becomes subdivided in to smaller groups in an effort to homogenise the genetic signal. For example only around 15% of patients in the London cohort fall in to what is regarded as a clearly definable genetic group such as IGE. An expansion of collection efforts is clearly required if any one subgroup within a cohort is going to reach a reasonable size to detect moderate and rare effect variants. Within a study centre this translates to greater investment in patient collection, databasing (fast and easy access to data of a specific nature) and, crucially, phenotyping.

7.1.2 Refining phenotype definition

Genotype and phenotype data play equally important roles in disease mapping, yet often the focus of attention is biased towards the former. As mentioned previously, one of the major difficulties with epilepsy genetics is the lack of intermediate phenotypes. This emphasises more the end-point phenotype of seizure or syndrome type. A key challenge facing the field is to align the end-point phenotype with the underlying genetic component. Correct alignment will maximise power. Any discrepancy in this alignment reduces power of detection considerably. For example, around 40% of TLE are caused by sclerosis in the hippocampus. Should TLE be considered as two independent groups - those with and those without sclerosis of the hippocampus - or should they instead be kept as a single group as has traditionally been the case? Genetic effect may well cross currently defined ILAE syndrome and seizure types, possibly leading to a realignment of current categories.

7.1.3 Improving genetic mapping techniques

The fruition of HapMap has made map based studies technically and financially feasible for a growing number of research groups. But the tagging methodology has yet to be fully optimised, a fact reflected by the numerous diverse methods

detailed in the literature. Further studies are required to compare and contrast the relative power of each method under different models (single functional locus, multiple interacting loci, different frequencies of functional loci etc.). Although this should be viewed as a natural maturation process, care must be taken in ensuring the optimal method is chosen for the study in question. Neither should the tagging methodology be used blindly. It should always be assumed that even under a best case scenario, the actual performance of tagging SNPs will not achieve the level predicted from the tag selection analysis due to differing patterns of LD in HapMap and test populations.

Regardless of the attention the HapMap project has generated, linkage mapping should continue to play a central in role epilepsy genetics. Indeed despite its limitations, linkage has been far more productive than association in terms of generating positive genetic knowledge[xxv]. However, for reasons described previously (see section 1.4.2), linkage in the form applied to epilepsy to date is poorly suited to the complex nature of the trait. Linkage needs to be refined and applied to carefully selected populations with deep pedigrees and homozygous environments. Homozygosity mapping is an alternative map based technique that has been under-examined in epilepsy. The application of the method in deep consanguineous pedigrees could yield candidate regions in the near term.

[xxv] Although linkage success stories have been largely restricted to cases of Medelian epilepsy.

7.1.4 Collaboration in epilepsy genetics

There are few if any research groups that can adequately tackle all three issues raised here, by increasing study cohort sizes, refining phenotype definition and improving genetic mapping techniques. It is clear that future epilepsy genetics teams will require the careful integration of clinicians, statistical geneticists and medical biologists. Collaboration is therefore key.

Despite the difficulties, epilepsy genetics is in the advantageous position of being, relative to other disease fields, in its infancy. In addition to having the luxury of learning from the experience of other more established disease fields an emerging field can develop large scale collaborative efforts at an early stage. Establishing an efficient system across groups greatly speeds research efforts. For example, streamlining phenotype databasing (a critical component of disease mapping) across collaborating centres allows an exact phenotype to be quickly and efficiently extracted at numerous centres simultaneously. Pooling of cohorts through collaborative efforts allows, in the near term, a significant increase in overall cohort size. Given the heterogeneous nature of the epilepsy phenotype pooling of cohorts appears to be the only way to achieve reasonable cohort sizes in the near term.

Collaboration must be built on a global scale. Large, ethnically homogenous, cohorts aid initial detection of functional variation, but access to patient cohorts of

different ethnicity is critical to determining the functionality of those variants across ethnic groups. Only through collaboration can this be carried out quickly and efficiently.

There are obvious disadvantages to large scale collaborative efforts, and care should be taken to avoid them. Diversity in how a question is tackled must be conserved. If the community is dominated by the thoughts of one or two key figures the development of alternative strategies might be hampered. However, for reasons discussed above, the system of research groups working in complete independence will compromise progress, at least in the short term.

The community must also think through the implications of the obviously complex nature of epilepsy genetics. The relevance of detecting variation of small to medium effects in known candidate genes is unclear. Deciphering interactions between such variants in order to explain a larger component of the risk would be beneficial in identifying high risk individuals who have not yet experienced a seizure. Perhaps of greater relevance to the management of epilepsy is the identification of novel, or confirmation of suspected, candidate genes, which would aid in the development of novel anti-epileptic drugs. Studies seeking to explain variability in dosing and efficacy of current anti-epileptic drugs are much required.

7.1.5 Pharmacogenetics in epilepsy

The community has recently focussed heavily on predisposition when the question of treatment might be more fruitful. The development of any new drug typically results in the identification of metabolisers, transporters and targets of the drug, thus immediately presenting a list of well motivated candidate genes for a pharmacogenetic study. Critically, these genes (in particular the drug metabolisers) have been extensively screened for variation, and furthermore, much of this variation has been tested for functionality. Intermediate phenotypes are often available (for example plasma levels of a drug). This allows known functional variants to be tested against drug related phenotypes, a strategy that has proven successful in epilepsy and other disease fields. For example variation in *CYP2C9* has been shown to affect plasma levels of Phenytoin [232]. In a more extensive study, part of the variability in dose of Phenytoin and Carbamazapine was shown to be explained by *SCN1a* genotype [69]. Although the proportion of the variation in dose explained is low (3-5%), it seems reasonable to predict that the addition of other weak to moderate effect variants would bring this above the threshold of clinical relevance.

The argument has been made that much of the genetic variation relevant to pharmacogenetics may be under positive and/or balancing selection to a much greater extent than disease causing variation. It is also possible that some variants may have been evolutionarily neutral until becoming functional on

exposure to previously unseen drug compounds. Under either scenario, 'functional' variants might drift to a relatively high frequency which would greatly increase not only the genetic power of detection but also the clinical relevance of the variant in question.

But there are still many challenges connected with pharmacogenetics in epilepsy. As most patient recruitment takes place at tertiary referral centres, patients tend to be on a cocktail of multiple drugs. The interaction of different drugs is poorly understood but likely to introduce a significant confounder effect. The type and severity of epilepsy also plays a role in drug response. Ideally study cohorts should be controlled with this in mind – for example a large cohort of JME patients might be better matched to a Valproate study than a cross-section of all epilepsy types. All these factors could and probably do play a role in determining the relative contribution of specific genetic variants to the efficacy of the drug.

7.1.6 Concluding thoughts

Epilepsy is clearly behind other common diseases fields such as diabetes and cardiovascular disease which have been subjected to large-scale studies comprising thousands of patients. As of 2005 the biggest association study in epilepsy considered one genetic variant in 563 cases and 660 controls. The execution of the study described in chapter six of this thesis looks set to change this picture and move epilepsy closer to the fore of complex disease genetics.

Epilepsy research could in principal emerge centre stage in genomics, alongside cancer research, because of the availability of affected tissue. This factor has allowed cancer studies to race ahead in a number of ways, yet the availability of affected tissue has been under-exploited in epilepsy. What is required is an integrated approach that correlates germline genetics with genomic patterns of protein related traits available from tissues (expression levels, alternative splicing, proteomics). This could not only further our understanding of epilepsy, but would also be a useful model for studying mechanisms such as apoptosis and the factors affecting long term potentiation.

Inherent potential and promise are insufficient alone if research funding bodies are to continue supporting a concept. Clear progress is required. Genomics research has been long promising to translate practically to the clinic; the time has come to deliver or the window of opportunity will be lost.

Addenda

Solutions:

-PBS (phosphate buffered saline)

For 1l: 2 PBS tablets (Sigma-Aldrich P08057) added to 1l of H_20

2 tablets in 1L of water yields in 0.01 M phosphate buffer, 0.0027 M KCl, 0.14 M NaCl, 0.05% Tween, pH 7.4

- Red blood cell lysis solution

320mM sucrose, 5mM $MgCl_2$, 10mM Tris/HCl, 1% Triton X-100.

For 1l

104.9g sucrose

1.02g $MgCl_2$.6H2O

10ml 1M Tris pH 8.0

10ml Triton-X-100

Make up to 1l with H_20

- TE

For 1l

10ml 1M Tris-HCl (pH 8.0)

400µl 0.25 M EDTA

Make up to 1l with H_20

Table 8.1 Summary of published association studies in complex forms of epilepsy

Study	Variant	#cases/#controls	Phenotype	Result	P value	Comment
Studies on focal epilepsies – initial positive report with or without replication attempts						
Il1-b						
Kanemoto[114]	-511	50/112	TLE+HS	Significant	0.0085	Japanese
Virta[115]	-511	35/400	FS	Significant	0.03	Finnish
Kanemoto[169]	-511	66/163	TLE+HS, FS	Significant	0.0022	Japanese
Tilgen	-511	99/126	FS	Negative	na	Failed replication of [115], Germans
Peltola[120]	-511	38/400	RPE	Significant	0.0023	Finnish
Heils[117]	-511	50/112	TLE+HS	Negative	na	Failed replication of [114],German
Buono[157]	-511	61/119	TLE+HS	Negative	na	Failed replication of [114], USA
Haspolat [118]	-511	73/152	FS	Negative	na	Failed replication of [114], Turkish
Jin [119]	-511	112/115	TLE+/-HS	Negative	Na	Failed replication of [114], Chinese
Il-1Ra						
Tsai [233]	I410	51/83	FS	Positive	0.03	Taiwan
Haspolat [118]	I410	73/152	FS	Negative	na	Failed replication of [233], Turkish
PDYN						
Stogmann[150]	promo	155/202	Fam nl-TLE	Significant	0.0025	German
Tilgen[158]	promo	182/205	nl-TLE	Negative	na	Failed replication of [150], German
Gambardella[159]	promo	60/259	Fam nl-TLE	Negative	na	Italian
GABBR1						
Gambardella[151]	G1465A	141/372	nl-TLE	Significant	<0.0001	Italian
Ma[165]	G1465A	120/118	nl-TLE	Negative	na	Failed replication of [151], USA
Ren [163]	G1465A	112/124	nl-TLE	Negative	na	Failed replication of [151], Chinese
Salzmann [166]	G1465A	134/145	nl/l TLE	Negative	na	Failed replication of [151], French
Tan [234]	G1465A	234/164	nl-TLE	Negative	na	Failed replication of [151], Australian
PRNP						
Walz[152]	Asn171Ser	100/180	rf TLE+HS	Significant	<0.0001	Brazilian

Table 8.1 continued... Summary of published association studies in complex forms of epilepsy

APOE						
Briellmann[153]	ε4	43	AoE TLE	Significant	0.004	Australian
Blumke[160]	ε4	125	AoE TLE	Negative	na	German
Gambardella[161]	ε4	63/220	AoE TLE	Negative	na	Italian
Yeni [162]	ε4	47/62	mTLE+HS	Negative	na	Turkish
CHRNA4						
Chou[154]	Ser543Ser	102/80	FS	Significant	0.001	Taiwanese
Mulley[164]	Ser543Ser	49/93	FS	Negative	na	Failed replication of [154], Australia
GABRG2						
Chou[155]	rs211037	104/83	FS	Significant	0.009	Taiwanese
Nakayama [167]	rs211037	94/106	FS	Negative	ns	Japanese (mutation screen)
Kananura [49]	rs211037	135/154	FS/CAE	Negative	ns	German (mutation screen)
CSNK1G2						
Yinan [235]	IVS2-33, A2241G	60/101	FS	Significant	0.009	Chinese
Studies on idiopathic generalised epilepsies – initial positive report with or without replication attempts						
OPRM1						
Sander [236]	Asn40Asp	72/340	IAE	Significant	0.019	German
Wilkie [176]	Asn40Asp	230/234	Mixed IGE	Significant	0.00026	UK (retracted)
CACNA1A						
Chioza [32]	'SNP8'	188/200	Mixed IGE	Significant	0.00033	UK
Sander [182]	'SNP8'	354/186	Mixed IGE	Negative	na	Failed replication of [32], German
CHRNA4						
Steinlein [237]	Cfolbp594	103/92	Mixed IGE	Significant	0.0394	German
Chioza [238]	Cfolbp594	182/178	Mixed IGE	Negative	na	Failed replication of [237] UK
GRIK1						
Sander [180]	AGAT rep	20 families[xxvi]	JAE	Significant	0.004	German
Izzi [183]	various	25[xxvii]	JAE	Negative	na	German (mutation screen)
KCNQ3						
Vijai [239]	D8S558	119probands[xxviii]	JME	Significant	0.008	South Indian
Haug [240]	D8S448, D8S1835	71 families[xxix]	Mixed IGE	Negative	na	German
KCNJ10						
Buono [241]	Arg271Cys	407/284	Mixed epilepsy	Significant		US/German
Lenzen [242]	Arg271Cys	563/660	Mixed IGE	Significant	0.03	Positive replication of [241], German

[xxvi] Family based analysis using HHRR statistic.
[xxvii] This study consisted of a screen in 25 patients. There was no case/control analysis.
[xxviii] Family based analysis using TDT statistic.
[xxix] Family based analysis using TDT statistic.

Table 8.1 continued... Summary of published association studies in complex forms of epilepsy

AE3						
Sander	Ala867Asp	366/183	Mixed IGE	Significant	0.021	German
KCNJ3						
Chioza [181]	T1038C	187/198	Mixed IGE	Significant	0.0097	UK, strongest in absence
DAT						
Sander [243]	3'UTR repeat	133/223	Mixed IGE	Significant	0.005	German, strongest IAE
CACNA1H						
Chen [244]	Mutation screen	118/230	CAE	na	na	Han Chinese
LGI4						
Gu [245]	1914GC	42/110	CAE	Significant	0.01	German
BRD2						
Pal [1]	Various	n/n	JME	Significant	nn	USA, mapping after linkage
ME2						
Greenberg [186]	7 locus haplotype	156/126	Mixed IGE	Significant	<0.0001	USA, mapping after linkage
Lenzen [246]	7 locus haplotype	660/666	Mixed IGE	Negative	na	Negative replication of [186], German
CX36						
Mas [247]	C588T	29/123	JME	Significant	0.017	European
Negative association studies involving genes for which no previous positive report existed in the literature.						
5HT2c						
Samochowiec [248]	Cys23Ser	119/242	Mixed IGE	Negative	na	German
ARC						
Haug [249]	C498	143/n	Mixed IGE	Negative	na	German, mutation screen
EAAT2/SERT						
Sander [243]	G603/promo	133/223	Mixed IGE	Negative	na	German
GABBR1						
Sander [250]	Various	118/130	JME/IAE	Negative	na	German
Lu [251]	Various	96/96	CAE	Negative	na	Han Chinese, mutation screen
GABRD						
Lenzen [252]	C659G	462/664	Mixed IGE	Negative	Na	German
GLRA3/GLRB						
Sobetzko [253]	Various	104/141	Mixed IGE	Negative	na	German, mutation screen
mGluR7/R8						
Goodwin [254]	Tyr433Phe/C2756T	n/n	Mixed IGE	Negative	na	UK

Table 8.1 continued... Summary of published association studies in complex forms of epilepsy

GRIM4						
Izzi [255]	C1455T	144/144	JME	Negative	na	German, mutation screen
KCNQ2						
Steinlein [256]	Thr752Asn	n/n?	Mixed IGE	Negative	na	German, mutation screen
Chou [257]	? paper not available	?	FS	Negative	na	Taiwan
hKCa3						
Sander [258]	CAG repeat	126/290	Mixed IGE	Negative	na	German
TASK-3						
Kananura [259]	C636T	65/	IAE	Negative	na	German, mutation screen
SCN2a						
Haug [260]	Various	46/	Mixed IGE	Negative	na	German, mutation screen
Nakayama	R19K	93/100	FS	Negative	na	Japanese
MAO-A						
Haug [261]	Promo variant	248/248	JME/IAE	Negative	na	German
PAX6						
Sander [262]	Promo variant	125/354	JME/IAE	Negative	na	German
Il-4						
Tsai [263]	Intron 3 repreat	94/83	FS/mixed IGE	Negative	na	Taiwan
Drug response studies						
MRD1						
Siddiqui [171]	C3435T	200/115	Any epilepsy	Significant	0.006	UK
Zimprich [173]	Various inc. C3435T	210/228	TLE	Significant	0.009	German
Tan [172]	C3435T	401/208*	Any epilepsy	Negative	na	UK
Sills [264]	C3435T	230/170*	Any epilepsy	Negative	na	UK
APOE						
Sporis [265]	ε4	30/30*	Refractory/non	Positive	0.002	Croation
DBH						
Depondt	C1021T	675/1087	Epi/controls	Negative	na	UK

*refractory/sensitive to drug treatment

Table 8.2 Summary of candidate genes and SNP selection for large scale epilepsy project

P1	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM[xxx]	#func HM	#tags
Voltage-Gated Sodium Channels							
1	SCN1A	chr2:167038217-167142693	104476	Na channel, voltage-gated, type I, alpha	1	2	4
2	SCN2A2	chr2:166343136-166459083	115947	Na channel, voltage-gated, type II, alpha 2	3	2	5
3	SCN3A	chr2:166136584-166273097	136513	Na channel, voltage-gated, type III, alpha	0	1	9
4	SCN4A	chr17:62479285-62533649	54364	Na channel, voltage-gated, type IV, alpha	0	1	9
5	SCN5A	chr3:38540211-38661744	121533	Na channel, voltage-gated, type V, alpha	3	2	20
6	SCN8A	chr12:50261287-50498365	237078	Na channel, voltage gated, type VIII, alpha	0	0	11
7	SCN9A	chr2:167247353-167380858	133505	Na channel, voltage-gated, type IX, alpha	2	4	11
8	SCN1B	chr19:40203565-40227014	23449	Na channel, voltage-gated, type I, beta	0	1	2
9	SCN2B	chr11:117563856-117594966	31110	Na channel, voltage-gated, type II, beta	1	1	6
10	SCN3B	chr11:123027549-123072607	45058	voltage-gated Na channel beta-3	0	2	13
11	SCN4B	chr11:117531744-117571187	39443	Na channel, voltage-gated, type IV, beta	4	1	6
Voltage-Gated Calcium Channels							
12	CACNA1A	chr19:13169088-13488274	319186	Ca channel, voltage-dependent, P/Q type, 1A	3	0	39
13	CACNA1B	chr9:135980855-136245245	264390	Ca channel, voltage-dependent, L type,	4	2	14
14	CACNA1C	chr12:2084680-2681584	596904	Ca channel, voltage-dependent, L type, α 1C	8	13	61
15	CACNA1D	chr3:53476031-53813420	337389	Ca channel, voltage-dependent, L type,	2	1	31
16	CACNA1S	chr1:198286899-198379951	93052	Ca channel, voltage-dependent, α 1S subunit	6	3	15
17	CACNA1F	chrX:48077757-48126067	48310	Ca channel, voltage-dependent, α 1F subunit	0	2	4
18	CACNA1E	chr1:178682262-179017473	335211	Ca channel, voltage-dependent, α 1E subunit	6	3	28
19	CACNA1G	chr17:49103459-49189264	85805	voltage-dependent Ca channel α 1G subunit	2	3	15
20	CACNA1H	chr16:1133242-1221583	88341	Ca channel, voltage-dependent, α 1H subunit	6	4	11
21	CACNA1I	chr22:38199797-38338774	138977	voltage-dependent T-type Ca channel	0	2	18
22	CACNA2D1	chr7:81181469-81694856	513387	Ca channel, voltage-dependent, α subunit	3	4	46
23	CACNA2D2	chr3:50351109-50509866	158757	Ca channel, voltage-dependent, α subunit	2	4	9
24	CACNA2D3	chr3:54104040-55075929	971889	Ca channel, voltage-dependent, α subunit	10	4	98
25	CACNA2D4	chr12:1761392-1908131	146739	voltage-gated Ca channel α(2)delta-4	4	3	20
26	CACNB1	chr17:37694669-37738864	44195	Ca channel, voltage-dependent, beta 1	0	5	12
27	CACNB2	chr10:18423611-18844043	420432	Ca channel, voltage-dependent, beta 2	6	4	52

[xxx] HM = HapMap

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P2	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Voltage-Gated Calcium Channels – continued.							
28	CACNB3	chr12:47488781-47518991	30210	Ca channel, voltage-dependent, beta 3	0	2	4
29	CACNB4	chr2:152887685-153168048	280363	Ca channel, voltage-dependent, beta 4	2	2	16
30	CACNG1	chr17:65580807-65613010	32203	voltage-dependent Ca channel gamma-1	2	0	5
31	CACNG2	chr22:35193437-35351660	158223	voltage-dependent Ca channel gamma-2	2	1	22
32	CACNG3	chr16:24224277-24350278	126001	voltage-dependent Ca channel gamma-3	2	1	19
33	CACNG4	chr17:65501127-65589611	88484	voltage-dependent Ca channel gamma-4	2	0	10
34	CACNG5	chr17:65413552-65441458	27906	voltage-dependent Ca channel gamma-5	0	1	8
35	CACNG6	chr19:59177354-59217730	40376	voltage-dependent Ca channel gamma-6	1	0	5
36	CACNG7	chr19:59097898-59147359	49461	voltage-dependent Ca channel gamma-7	1	0	6
37	CACNG8	chr19:59148102-59187915	39813	voltage-dependent Ca channel gamma-8	0	0	5
Voltage gates Chloride Channels							
38	CLCN1	chr7:142474149-142530027	55878	Cl channel 1, skeletal muscle (Thomsen	0	3	11
39	CLCN2	chr3:185375307-185410180	34873	Cl channel 2	0	0	5
40	CLCN3	chr4:171227689-171348047	120358	Cl channel 3 isoform e	0	0	8
41	CLCN4	chrX:9527833-9625025	97192	Cl channel 4	3	4	18
42	CLCN5	chrX:48741496-48784861	43365	Cl channel 5	0	0	1
43	CLCN6	chr1:11565814-11622731	56917	Cl channel 6 isoform ClC-6a	9	6	9
44	CLCN7	chr16:1425346-1475013	49667	Cl channel 7	1	2	12
45	CLCNKA	chr1:15701802-15732900	31098	Cl channel Ka	2	3	5
46	CLCNKB	chr1:15723056-15756160	33104	Cl channel Kb	1	3	7
47	SLC12A5	chr20:45333412-45384205	50793	solute carrier family 12 member 5	2	4	11
48	BSND	chr1:54824535-54854379	29844	barttin	2	0	4
Voltage-Gated Potassium Channels							
49	KCNQ1	chr11:2420530-2844530	424000	K voltage-gated channel, KQT-like	5	8	60
50	KCNQ2	chr20:62747915-62840206	92291	K voltage-gated channel KQT-like protein	4	5	15
51	KCNQ3	chr8:133087845-133459227	371382	K voltage-gated channel KQT-like protein	7	5	57
52	KCNQ4	chr1:40652978-40727654	74676	K voltage-gated channel KQT-like protein	2	2	15
53	KCNQ5	chr6:73317621-73911178	593557	K voltage-gated channel, KQT-like	9	1	45

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P3	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Voltage-Gated Potassium Channels – continued.							
54	KCNAB1	chr3:157149261-157587247	437986	K voltage-gated channel, shaker-related	5	3	32
55	KCNAB2	chr1:5800142-5880069	79927	K voltage-gated channel, shaker-related	0	5	13
56	KCNAB3	chr17:8016986-8043319	26333	K voltage-gated channel, shaker-related	0	2	4
57	KCNMB1	chr5:169776062-169807292	31230	K large conductance Ca-activated	2	2	9
58	KCNMB2	chr3:179587401-179893129	305728	Ca-activated K channel beta 2	3	1	26
59	KCNMB3	chr3:180271478-180300316	28838	Ca-activated K channel beta 3	0	4	7
60	KCNMB4	chr12:69036446-69121065	84619	Ca-activated K channel beta 4	1	3	12
61	KCNA1	chr12:4880806-4902293	21487	K voltage-gated channel, shaker-related	0	1	4
62	KCNA2	chr1:110434632-110456131	21499	K voltage-gated channel, shaker-related	1	0	2
63	KCNA3	chr1:110503807-110526105	22298	K voltage-gated channel, shaker-related, 3	0	0	2
64	KCNA4	chr11:29986307-30012797	26490	K voltage-gated channel, shaker-related	0	2	3
65	KCNA5	chr12:5013346-5036208	22862	K voltage-gated channel, shaker-related	2	2	7
66	KCNA6	chr12:4778603-4802839	24236	K voltage-gated channel, shaker-related, 6	1	2	7
67	KCNA7	chr19:54252487-54278010	25523	K voltage-gated channel, shaker-related	1	3	6
68	KCNA10	chr1:110348566-110370524	21958	K voltage-gated channel, shaker-related	2	0	5
69	KCNN1	chr19:17913111-17980800	67689	K intermediate/small conductance	1	3	12
70	KCNN2	chr5:113764232-113918405	154173	small conductance Ca-activated K	2	0	16
71	KCNN3	chr1:151886705-152069546	182841	small conductance Ca-activated K	3	2	42
72	KCNN4	chr19:48952527-48987249	34722	intermediate conductance Ca-activated	2	0	7
73	KCNJ6	chr21:37907185-38219058	311873	K inwardly-rectifying channel J6	1	1	37
74	KCNJ10	chr1:157217130-157266830	49700	K inwardly-rectifying channel J10	1	1	6
75	KCNC1	chr11:17711846-17768486	56640	Shaw-related voltage-gated K channel	1	3	11
76	KCND2	chr7:119458225-119954890	496665	K voltage-gated channel, Shal-related	3	2	22
77	HCN1	chr5:45297474-45751721	454247	hyperpolarization activated cyclic	1	1	14
78	HCN2	chr19:530893-578157	47264	hyperpolarization activated cyclic	1	2	10
79	HCN3	chr1:152454166-152485053	30887	hyperpolarization activated cyclic	0	3	4
80	HCN4	chr15:71319752-71386994	67242	hyperpolarization activated cyclic	0	2	11
Acetylcholine metabolism							
81	CHAT	chr10:50156686-50227753	71067	acetylcholine acetyltransferase isoform 2	4	3	14
82	ACHE	chr7:100089015-100113406	24391	acetylcholinesterase isoform E4-E5 precursor	5	1	2

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P4	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Acetylcholine receptors							
83	CHRNA1	chr2:175805151-175841714	36563	cholinergic receptor, nicotinic, α	1	2	5
84	CHRNA2	chr8:27330175-27368668	38493	cholinergic receptor, nicotinic, α 2 (neuronal)	4	3	8
85	CHRNA3	chr15:76593471-76639141	45670	cholinergic receptor, nicotinic, α	2	2	2
86	CHRNA4	chr20:62692799-62729096	36297	cholinergic receptor, nicotinic, α polypeptide 4	2	0	7
87	CHRNA5	chr15:76563725-76612277	48552	cholinergic receptor, nicotinic, α	2	2	3
88	CHRNA6	chr8:42615150-42650987	35837	cholinergic receptor, nicotinic, α	0	0	3
89	CHRNA7	chr15:30028757-30187178	158421	cholinergic receptor, nicotinic, α polypeptide 7	0	1	13
90	CHRNA9	chr4:40243011-40282515	39504	cholinergic receptor, nicotinic, α	0	1	3
91	CHRNA10	chr11:3641126-3666923	25797	cholinergic receptor, nicotinic, α	1	1	3
92	CHRNB1	chr17:7538983-7571455	32472	nicotinic acetylcholine receptor beta subunit	0	2	5
93	CHRNB2	chr1:151747049-151775875	28826	cholinergic receptor, nicotinic, beta	1	0	4
94	CHRNB3	chr8:42559964-42619577	59613	cholinergic receptor, nicotinic, beta	3	2	3
95	CHRNB4	chr15:76622455-76659406	36951	cholinergic receptor, nicotinic, beta	1	0	5
96	CHRM1	chr11:62441512-62474372	32860	cholinergic receptor, muscarinic 1	0	1	7
97	CHRM2	chr7:136105655-136127055	21400	cholinergic receptor, muscarinic 2 isoform a	3	3	9
98	CHRM3	chr1:237101901-237123589	21688	cholinergic receptor, muscarinic 3	2	1	4
99	CHRM4	chr11:46360977-46382416	21439	cholinergic receptor, muscarinic 4	1	1	1
100	CHRM5	chr15:32016484-32082987	66503	cholinergic receptor, muscarinic 5	0	1	4
GABA-B Receptors							
101	GABRA1	chr5:161246464-161317306	70842	GABA A receptor, α	0	0	5
102	GABRA2	chr4:46157248-46317487	160239	GABA A receptor, α 2	3	2	9
103	GABRA3	chrX:149944145-150247438	303293	GABA A receptor, α 3	0	0	14
104	GABRA4	chr4:46835784-46920997	85213	GABA A receptor, α 4	6	1	9
105	GABRA5	chr15:24712714-24767114	54400	GABA A receptor, α 5	0	0	2
106	GABRA6	chr5:161083865-161120014	36149	GABA A receptor, α 6	0	1	2
107	GABRG1	chr4:45947787-46051559	103772	GABA A receptor, gamma 1	1	1	4

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P5	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
GABA-B Receptors - continued							
108	GABRG2	chr5:161465724-161573421	107697	GABA A receptor, gamma 2	1	1	10
109	GABRG3	chr15:25318155-25390486	72331	GABA A receptor, gamma 3	2	1	9
110	GABRE	chrX:149729210-149770764	41554	GABA A receptor,	0	1	11
111	GABRP	chr5:170181660-170231943	50283	GABA A receptor, pi	3	2	6
112	GABRB1	chr4:46939133-47353987	414854	GABA A receptor, beta	4	4	28
113	GABRB2	chr5:160691766-160966025	274259	GABA A receptor, beta	2	2	17
114	GABRB3	chr15:24328824-24575550	246726	GABA A receptor, beta	5	2	26
115	GABRQ	chrX:150414250-150449438	35188	GABA receptor, theta	0	2	5
116	GABRR1	chr6:89873568-89932656	59088	GABA receptor, rho 1	4	0	10
117	GABRR2	chr6:89952835-90030550	77715	GABA receptor, rho 2	3	3	20
118	GABBR1	chr6:29677897-29708753	30856	GABA B receptor 1	10	0	3
119	GPR51	chr9:96420204-96861317	441113	G protein-coupled receptor 51	4	9	67
GABA transporters							
120	SLC32A1	chr20:38028533-38053443	24910	solute carrier family 32, member 1	0	0	5
121	SLC6A1	chr3:11022354-11063906	41552	solute carrier family 6 (neurotransmitter	1	1	11
122	SLC6A13	chr12:190052-252263	62211	solute carrier family 6 (neurotransmitter	2	1	9
123	SLC6A11	chr3:10822917-10965146	142229	solute carrier family 6 (neurotransmitter	3	2	18
124	SLC6A12	chr12:159513-202759	43246	solute carrier family 6 (neurotransmitter	0	0	6
GABA-Glut metabolism/synthesis							
125	GLS	chr2:191938140-192042808	104668	glutaminase C	1	1	4
126	GLS2	chr12:55141003-55178447	37444	glutaminase GA isoform a	1	3	4
127	GLUL	chr1:179581414-179610285	28871	glutamate-ammonia ligase (glutamine synthase)	4	3	6
128	GAD1	chr2:171865616-171930193	64577	glutamate decarboxylase 1 isoform GAD67	0	0	7
129	GAD2	chr10:26499600-26607493	107893	glutamate decarboxylase 2	2	1	5
130	ALPL	chr1:21295402-21384379	88977	tissue non-specific alkaline phosphatase	2	2	15
131	ABAT	chr16:8725743-8855729	129986	4-aminobutyrate aminotransferase precursor	7	2	25
132	ALDH5A1	chr6:24593204-24652837	59633	aldehyde dehydrogenase 5A1 precursor, isoform 1	2	4	10
GABA receptor associated proteins							
133	GPH	chr14:64955277-65648556	693279	gephyrin	1	2	21

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P6	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
GABA receptor associated proteins – continued.							
134	GABARAP	chr17:7334305-7356318	22013	GABA(A) receptor-associated protein	2	3	5
135	ARHGEF9	chrX:61711639-61851784	140145	Cdc42 guanine exchange factor 9	0	0	5
Glutamate receptors							
136	GRIA1	chr5:152888816-153229374	340558	glutamate receptor, ionotropic, AMPA 1	3	1	33
137	GRIA2	chr4:158709543-158872724	163181	glutamate receptor, ionotropic, AMPA 2	0	1	8
138	GRIA3	chrX:121005913-121330548	324635	glutamate receptor 3 isoform flip precursor	4	5	54
139	GRIA4	chr11:105009442-105398183	388741	glutamate receptor, ionotrophic	1	1	20
140	GRIK1	chr21:29836313-30242677	406364	glutamate receptor, ionotropic, kainate 1	4	5	41
141	GRIK2	chr6:101882752-102571956	689204	glutamate receptor 6 isoform 1 precursor	3	0	57
142	GRIK3	chr1:36684276-36937388	253112	glutamate receptor 7 precursor	2	0	21
143	GRIK4	chr11:120058680-120404621	345941	glutamate receptor KA1 precursor	3	5	37
144	GRIK5	chr19:47187783-47271797	84014	GRIK5	0	1	4
145	GRIN1	chr9:135381185-135429808	48623	glutamate receptor, N-methyl D-aspartate 1	0	0	4
146	GRIN2A	chr16:9812721-10253169	440448	glutamate receptor, N-methyl D-aspartate 2A	5	2	30
147	GRIN2B	chr12:13595411-14034289	438878	N-methyl-D-aspartate receptor subunit 2B	11	6	54
148	GRIN2C	chr17:73425403-73458354	32951	N-methyl-D-aspartate receptor subunit 2C	0	3	4
149	GRIN2D	chr19:53579944-53649205	69261	N-methyl-D-aspartate receptor subunit 2D	3	3	8
150	GRIN3A	chr9:99701473-99890700	189227	glutamate receptor, ionotropic,	3	7	19
151	GRIN3B	chr19:951436-960723	9287	glutamate receptor, ionotropic,	0	0	8
152	GRM1	chr6:146321224-146749304	428080	glutamate receptor, metabotropic 1	5	5	18
153	GRM2	chr3:51692046-51721176	29130	glutamate receptor, metabotropic 2 precursor	0	0	2
154	GRM3	chr7:85875259-86116220	240961	glutamate receptor, metabotropic 3 precursor	2	0	15
155	GRM4	chr6:34026484-34158298	131814	glutamate receptor, metabotropic 4	4	0	10
156	GRM5	chr11:87919131-88479343	560212	glutamate receptor, metabotropic 5	1	2	33
157	GRM6	chr5:178508933-178545725	36792	glutamate receptor, metabotropic 6 precursor	4	3	8
158	GRM7	chr3:6848405-7768404	919999	glutamate receptor, metabotropic 7 isoform b	9	5	93
159	GRM8	chr7:125623146-126447799	824653	glutamate receptor, metabotropic 8	5	15	134

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P7	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Glutamate transporters							
160	SLC1A3	chr5:36642256-36743935	101679	solute carrier family 1 (glial high affinity)	0	1	21
161	SLC1A2	chr11:35236697-35414899	178202	solute carrier family 1, member 2	0	2	18
162	SLC1A1	chr9:4470489-4587260	116771	solute carrier family 1, member 1	4	2	25
163	SLC1A6	chr19:14911991-14954730	42739	solute carrier family 1 (high affinity	2	0	5
164	SLC1A7	chr1:52913474-52988197	74723	solute carrier family 1 (glutamate transporter),	13	4	11
165	SLC17A6	chr11:22313976-22375352	61376	differentiation-associated Na-dependent	1	1	10
166	SLC17A7	chr19:54614470-54646596	32126	solute carrier family 17, member 7	3	4	6
167	SLC17A8	chr12:99243462-99328305	84843	Na-dependent inorg. phosphate cotransporter 8	0	1	9
Extra metabolism - over GABA							
168	GLUD1	chr10:88,454,821-88,539,200	84379	glutamate dehydrogenase 1	0	0	4
169	GLUD2	chrX:118,859,216-118,901,549	42333	glutamate dehydrogenase 2	0	0	3
adenosine							
170	ADORA1	chr1:200,328,038-200,444,789	116751	adenosine A1 receptor	3	3	11
171	ADORA2A	chr22:23,133,645-23,182,878	49233	adenosine A2a receptor	2	0	2
172	ADORA2B	chr17:16,028,796-16,099,776	70980	adenosine A2b receptor	0	1	6
173	ADORA3	chr1:111,320,781-111,365,470	44689	adenosine A3 receptor	1	2	5
174	ADK	chr10:75,235,567-75,833,663	598096	adenosine kinase isoform b	2	6	18
175	ADA	chr20:43,913,592-43,985,805	72213	adenosine deaminase	1	0	7
176	AHCY	chr20:33,563,751-33,626,799	63048	S-adenosylhomocysteine hydrolase	0	2	3
177	NT5E	chr6:86,135,404-86,221,092	85688	5' nucleotidase, ecto	1	1	5
178	NT5C1A	chr1:39,518,086-39,571,004	52918	5'-nucleotidase, cytosolic IA	1	2	6
179	NT5C1B	chr2:18,708,649-18,775,350	66701	5' nucleotidase, cytosolic IB isoform 1	0	1	7
180	CNT1	chr15:83,137,680-83,238,795	101115	solute carrier family 28 (Na-coupled	5	9	34
181	CNT2	chr15:43250489-43294189	43700	solute carrier family 28 (Na-coupled	1	1	3
182	CNT3	chr9:82,330,210-82,460,532	130322	concentrative Na+-nucleoside cotransporter	2	2	13
183	ENT1	chr6:44,214,247-44,268,733	54486	solute carrier family 29 , member 1	0	2	10
184	ENT2	chr11:65,885,352-65,934,651	49299	solute carrier family 29,member 2	0	0	1
185	ENT3	chr10:72,403,634-72,487,743	84109	solute carrier family 29, member 3	2	9	15
186	ENT4	chr7:5,046,912-5,108,028	61116	solute carrier family 29, member 4	1	3	10
Serotonin							

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P8	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
189	HTR1D	chr1:22,967,863-23,010,697	42834	Serotonin receptor 1D	0	1	2
190	HTR1E	chr6:87,623,005-87,741,992	118987	Serotonin receptor 1E	0	0	11
191	HTR1F	chr3:87,940,800-87,981,901	41101	Serotonin receptor 1F	0	0	1
192	HTR2A	chr13:45,185,513-45,288,176	102663	Serotonin receptor 2A	4	1	15
193	HTR2B	chr2:232,155,493-232,212,363	56870	Serotonin receptor 2B	2	0	5
194	HTR2C	chrX:112,562,803-112,928,877	366074	Serotonin receptor 2C	1	7	14
195	HTR3A	chr11:113,363,561-113,418,684	55123	Serotonin receptor 3A	0	2	10
196	HTR3B	chr11:113,293,240-113,374,934	81694	Serotonin receptor 3B	2	2	8
197	HTR3C	chr3:185,071,747-185,119,372	47625	Serotonin receptor 3 subunit C	1	2	4
198	HTR3D	chr3:185,050,244-185,098,069	47825	Serotonin receptor 3 family member D	0	1	4
199	HTR3E	chr3:185,118,879-185,165,695	46816	Serotonin receptor 3 subunit E	0	1	6
200	HTR4	chr5:147,839,242-148,082,226	242984	Serotonin receptor 4	1	0	16
201	HTR5A	chr7:154,234,349-154,287,937	53588	Serotonin receptor 5A	2	2	5
202	HTR6	chr1:19,441,254-19,495,530	54276	Serotonin receptor 6	0	1	4
203	HTR7	chr10:92,145,154-92,302,248	157094	Serotonin receptor 7 isoform d	0	0	7
204	TPH2	chr12:70,598,892-70,732,488	133596	neuronal tryptophan hydroxylase	0	2	8
205	MAOA	chrx:42626592-42662049	35457	monoamine oxidase A	0	0	5
206	SLC6A4	chr17:28,650,468-28,728,268	77800	solute carrier family 6 member 4	0	1	5
Dopamine							
207	DRD1	chr5:174,828,959-174,872,086	43127	dopamine receptor D1	1	0	6
208	DRD2	chr11:112,797,968-112,903,544	105576	dopamine receptor D2	9	2	16
209	DRD3	chr3:115,148,457-115,238,657	90200	dopamine receptor D3	0	1	6
210	DRD4	chr11:607,536-650,933	43397	dopamine receptor D4	1	4	5
211	DRD5	chr4:9,514,485-9,556,515	42030	dopamine receptor D5	0	0	2
212	SLC6A3	chr5:1,425,646-1,518,281	92635	solute carrier family 6 (neurotransmitter	0	3	17
Adrenoreceptors							
213	ADRA1A	chr8:26,607,576-26,764,832	157256	α-1A-adrenergic receptor isoform 1	6	1	18
214	ADRA1B	chr5:159,304,634-159,400,446	95812	α-1B-adrenergic receptor	0	1	8
215	ADRA1D	chr20:4,176,815-4,244,659	67844	α-1D-adrenergic receptor	0	0	10
216	ADRA2A	chr10:112,481,507-112,525,157	43650	α-2A-adrenergic receptor	0	1	3
217	ADRA2B	chr2:96,243,378-96,286,644	43266	α-2B-adrenergic receptor	0	1	2

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P9	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Adrenoreceptors - continued							
218	ADRA2C	chr4:3,778,806-3,821,625	42819	α-2C-adrenergic receptor	0	1	4
219	ADRB1	chr10:115,448,401-115,490,115	41714	beta-1-adrenergic receptor	0	1	3
220	ADRB2	chr5:148,214,685-148,256,696	42011	adrenergic, beta-2-, receptor, surface	6	4	7
221	TH	chr11:2,129,468-2,177,344	47876	tyrosine hydroxylase isoform a	1	2	10
222	DDC	chr7:50,247,884-50,390,500	142616	dopa decarboxylase (aromatic L-amino acid	1	1	8
223	DBH	chr9:131,823,217-131,886,199	62982	dopamine beta-hydroxylase precursor	0	3	16
224	VMAT2	chr10:118,645,769-118,721,682	75913	solute carrier family 18 (vesicular monoamine),	2	2	9
225	SLC6A2	chr16:55,449,147-55,535,178	86031	solute carrier family 6 member 2	2	4	12
AED metabolisers - primary							
226	NR1I2	chr3:120,800,231-120,878,232	78001	pregnane X receptor isoform 2	1	0	7
227	CYP1A2	chr15:72,737,000-72,784,758	47758	cytochrome P450, family 1, subfamily A,	0	0	2
228	CYP2A6	chr19:46,021,283-46,068,180	46897	cytochrome P450, family 2, subfamily A,	5	0	2
229	CYP2B6	chr19:46,169,043-46,236,141	67098	cytochrome P450, family 2, subfamily B,	6	0	6
230	CYP2C9	chr10:96,343,026-96,433,734	90708	cytochrome P450, family 2, subfamily C,	0	7	9
231	CYP2C19	chr10:96,167,049-96,297,258	130209	cytochrome P450, family 2, subfamily C,	6	5	8
232	CYP2D6	chr22:40,747,095-40,791,440	44345	cytochrome P450, subfamily IID, polypeptide 6	9	1	3
233	CYP2E1	chr10:134,794,157-134,845,911	51754	cytochrome P450, family 2, subfamily E,	6	2	6
234	CYP3A4	chr7:98,946,000-99,013,205	67205	cytochrome P450, subfamily IIIA, polypeptide 4	0	0	2
235	CYP3A5	chr7:98,837,213-98,909,003	71790	cytochrome P450, subfamily IIIA, polypeptide 5	1	1	3
AED metabolisers - secondary							
236	UGT1A4	chr2:234935732-234967509	31777	UDP glycosyltransferase 1 family, polypeptide	3	9	27
237	UGT2B7	chr4:70,257,769-70,314,249	56480	UDP glycosyltransferase 2 family, polypeptide	0	4	6
238	EPHX1	chr1:222,967,606-223,027,892	60286	epoxide hydrolase 1, microsomal (xenobiotic)	8	5	10
239	GSS	chr20:34221912-34269277	47365	glutathione synthetase	3	0	0
240	GSTA1	chr6:52703224-52715493	12269	glutathione S-transferase A1	3	0	1
241	GSTA2	chr6:52662016-52675160	13144	glutathione S-transferase A2	3	0	4
242	GSTA4	chr6:52889586-52906976	17390	glutathione S-transferase A4	3	0	5
243	ODC1	chr2:10602246-10610192	7946	ornithine decarboxylase 1	6	0	5

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P10	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
AED metabolisers – secondary – continued.							
244	ABCB1	chr7:86,724,976-86,974,593	249617	ATP-binding cassette sub-family B member 1	2	1	11
245	ABCC1	chr16:15,989,872-16,222,627	232755	ATP-binding cassette, sub-family C, member 1	5	4	26
246	ABCC2	chr10:101,187,159-101,296,168	109009	ATP-binding cassette, sub-family C	7	4	7
247	ABCC3	chr17:49,166,865-49,263,700	96835	ATP-binding cassette, sub-family C, member 3	4	2	14
248	ABCC4	chr13:93,350,089-93,671,684	321595	ATP-binding cassette, sub-family C, member 4	7	1	31
249	ABCC5	chr3:184,938,638-185,076,588	137950	ATP-binding cassette, sub-family C, member 5	4	2	6
250	ABCC6	chr16:16,190,343-16,303,668	113325	ATP-binding cassette, sub-family C, member 6	5	3	7
251	ABCC10	chr6:43,426,343-43,485,018	58675	ATP-binding cassette, sub-family C, member 10	2	3	5
252	ABCA2	chr9:135,238,978-135,300,667	61689	ATP-binding cassette, sub-family A, member 2	1	3	6
253	RALBP1	chr18:9,445,529-9,548,105	102576	ralA binding protein 1	0	1	9
254	MVP	chr16:29,849,423-29,916,979	67556	major vault protein	0	1	3
255	LRPPRC	chr2:44,069,421-44,217,679	148258	leucine-rich PPR motif-containing protein	1	4	27
256	ABCG2	chr4:89,449,811-89,558,406	108595	ATP-binding cassette, sub-family G, member 2	2	1	7
AED transport - secondary							
257	SLCO2A1	chr3:134,952,450-135,089,637	137187	solute carrier organic anion transporter family,	0	2	22
258	SLCO1A2	chr12:21,293,093-21,459,638	166545	organic anion transporting polypeptide A isoform	0	4	13
259	SLCO2B1	chr11:74,568,470-74,663,605	95135	solute carrier organic anion transporter family,	3	1	18
260	SLCO3A1	chr15:90,106,713-90,456,547	349834	solute carrier organic anion transporter family,	7	4	64
261	SLCO4A1	chr20:61,980,145-62,049,996	69851	solute carrier organic anion transporter family	1	1	6
262	SLCO1C1	chr12:20,719,665-20,817,585	97920	solute carrier organic anion transporter family,	1	1	9
263	SLCO5A1	chr8:70,614,535-70,817,169	202634	organic anion transporter polypeptide-related	3	3	23
264	SLCO6A1	chr5:101,763,869-101,930,907	167038	solute carrier organic anion transporter family,	1	0	6
Keppra related							
265	SV2A	chr1:146687880-146704100	16220	synaptic vesicle glycoprotein 2	0	0	8
266	SV2B	chr15:89373312-89568278	194966	synaptic vesicle protein 2B homolog	1	0	25
267	SV2C	chr5:75405020-75667264	262244	synaptic vesicle protein 2C homolog	0	2	19
268	SYT1	chr12:78,093,750-78,367,311	273561	synaptotagmin I	2	0	10
269	SYT2	chr1:199,833,144-199,987,802	154658	synaptotagmin II	3	2	16
270	SYN1	chrX:46,458,244-46,545,239	86995	synapsin I isoform Ia	0	1	9

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P11	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Keppra related - continued							
271	STX1A	chr7:72,505,636-72,566,059	60423	syntaxin 1A (brain)	0	2	6
272	STX1B2	chr16:31,019,427-31,077,835	58408	syntaxin 1B2	0	1	2
273	SYP	chrX:48,050,498-48,102,895	52397	synaptophysin	0	0	3
274	SNAP25	chr20:10,174,476-10,303,065	128589	synaptosomal-associated protein 25 isoform	3	3	17
275	VAMP2	chr17:8,243,029-8,286,858	43829	VAMP2	0	2	3
276	VTI1A	chr10:113,851,602-114,182,119	330517	SNARE Vti1a-beta protein	6	0	28
277	VTI1B	chr14:66,087,909-66,151,589	63680	vesicle transport through interaction with	0	0	6
278	RAB3A	chr19:18,148,610-18,195,839	47229	RAB3A, member RAS oncogene family	1	4	7
279	RAB3B	chr1:51,734,754-51,846,266	111512	RAB3B, member RAS oncogene family	0	1	5
280	RAB3C	chr5:57,874,823-58,183,290	308467	RAB3C, member RAS oncogene family	5	3	27
281	RIMS1	chr6:72,572,324-73,126,014	553690	regulating synaptic membrane exocytosis 1	1	3	79
282	RNF40	chr16:30,789,689-30,842,596	52907	ring finger protein 40 isoform 1	0	1	2
283	NSF	chr17:45,123,036-45,329,632	206596	N-ethylmaleimide-sensitive factor	3	2	3
284	CPLX1	chr4:748,575-829,775	81200	complexin 1	2	5	8
285	CPLX2	chr5:175,184,532-175,311,946	127414	complexin 2	2	8	17
286	CSNK1A1	chr5:148,884,058-148,979,167	95109	casein kinase 1, α 1	1	2	5
287	CAMK2A	chr5:149,607,564-149,717,846	110282	CaM kinase II α	0	4	14
288	TRAPPC4	chr11:118,406,892-118,452,034	45142	trafficking protein particle complex 4	0	2	6
289	LRRC7	chr1:69,575,773-69,977,842	402069	leucine rich repeat containing 7	0	1	23
290	RPH3A	chr12:111,621,694-111,767,197	145503	rabphilin 3A homolog (mouse)	2	1	13
291	PNUTL1	chr22:18,056,540-18,105,396	48856	PNUTL1	0	0	3
292	STXBP1	chr9:125,730,612-125,851,040	120428	syntaxin binding protein 1	0	1	3
293	CAMK2B	chr7:43,979,676-44,126,004	146328	Ca/calmodulin-dependent protein kinase IIB	0	2	13
294	CAMK2D	chr4:114,812,392-115,161,548	349156	Ca/calmodulin-dependent protein kinase II	4	5	26
295	EPIM	chr12:129617247-129686907	69660	epimorphin isoform 2	0	2	4
296	PIK3CA	chr3:180,167,223-180,293,408	126185	phosphoinositide-3-kinase, catalytic, α	1	4	10
297	PIK3CG	chr7:106056455-106118121	61666	phosphoinositide-3-kinase, catalytic, gamma	2	1	5
298	RAB3GAP	chr2:136,000,621-136,158,313	157692	RAB3 GTPase-activating protein	0	1	6
299	RYR3	chr15:31309233-31894358	585125	ryanodine receptor 3	11	5	190
300	STX10	chr19:13105876-13132160	26284	syntaxin 10	0	2	4
301	STX11	chr6:144442239-144500071	57832	syntaxin 11	0	0	6

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P12	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Keppra related - continued							
303	stx3a	chr11:59288249-59355536	67287	syntaxin 3A	1	1	6
304	stx4a	chr16:31070963-31097544	26581	syntaxin 4A (placental)	0	0	2
305	stx5a	chr11:62339729-62384920	45191	syntaxin 5A	1	1	3
306	STX6	chr1:178174484-178241361	66877	syntaxin 6	1	1	6
307	STX7	chr6:132751725-132824736	73011	syntaxin 7	2	2	18
308	STX8	chr17:9344355-9689841	345486	syntaxin 8	1	2	42
309	SYN2	chr3:12010876-12217885	207009	synapsin II	4	5	8
310	SYN3	chr22:31223859-31737237	513378	synapsin III	3	10	83
311	SYNGR1	chr22:37979055-38034619	55564	synaptogyrin 1 isoform 1a	0	1	7
312	SYNGR3	chr16:1970027-1994276	24249	synaptogyrin 3	0	6	8
313	UNC13b	chr9:35142055-35405331	263276	UNC13 (C. elegans)-like	0	0	10
Apoptosis							
314	CSTB	chr21:44,030,033-44,072,461	42428	cystatin B	3	4	10
315	CASP2	chr7:142,436,331-142,495,714	59383	caspase 2 isoform 1 preproprotein	0	3	8
316	CASP3	chr4:186,224,740-186,286,491	61751	caspase 3 preproprotein	0	0	4
317	CASP6	chr4:111,048,604-111,103,449	54845	caspase 6 isoform α preproprotein	2	2	6
318	CASP8	chr2:202,280,755-202,374,974	94219	caspase 8, apoptosis-related cysteine protease	2	2	8
319	CASP10	chr2:202,230,397-202,316,653	86256	caspase 10 apoptosis-related cysteine protease	2	1	3
320	FADD	chr11:69,755,576-69,799,794	44218	Fas-associated via death domain	0	0	3
321	BAG4	chr8:38,031,473-38,105,905	74432	BCL2-associated athanogene 4	0	1	2
322	CFLAR	chr2:202,163,432-202,251,544	88112	CASP8 and FADD-like apoptosis regulator	1	1	3
323	BAX	chr19:54,129,928-54,176,866	46938	BCL2-associated X protein isoform beta	3	5	9
324	APAF1	chr12:97,521,545-97,651,678	130133	apoptotic protease activating factor isoform c	1	1	6
325	BCL10	chr1:85,134,524-85,186,176	51652	B-cell CLL/lymphoma 10	0	0	5
326	AKT1	chr14:103197384-103241239	43855	v-akt murine thymoma viral oncgene homolog 1	0	6	8
327	AKT2	chr19:45421635-45493034	71399	v-akt murine thymoma viral oncgene homolog 2	0	1	4
328	BAK1	chr6:33577188-33604836	27648	BCL2-antagonist/killer 1	2	1	11
329	BBC3	chr19:52405924-52436291	30367	BBC3	0	0	4
330	BCL2	chr18:58929548-59144899	215351	B-cell lymphoma protein 2 α isoform	1	4	29

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P13	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Apoptosis - continued							
331	BCL2L1	chr20:30987263-31065656	78393	BCL2-like 1 isoform 1	1	2	3
332	BCL2L11	chr2:111966469-112026954	60485	BCL2-like 11 isoform 1	0	1	7
333	BCL2L2	chr14:21756164-21781086	24922	BCL2-like 2 protein	1	0	4
334	BIRC4	chrX:121,681,802-121,755,576	73774	baculoviral IAP repeat-containing protein 4	2	2	6
335	CAB39	chr2:231,760,203-231,908,325	148122	Ca binding protein 39	0	0	9
336	casp4	chr11:104341321-104387006	45685	caspase 4 isoform α precursor	0	0	6
337	casp7	chr10:115094015-115165249	71234	caspase 7 isoform delta, large subunit	2	1	8
338	casp9	chr1:15171494-15223487	51993	caspase 9 isoform α preproprotein	0	4	5
339	DAPK3	chr19:3,889,451-3,940,826	51375	death-associated protein kinase 3	0	3	7
340	EIF4EBP1	chr8:68514-118365	49851	eukaryotic translation initiation factor 4E	0	0	0
341	FAS	chr10:90,394,864-90,460,118	65254	FAS	2	2	7
342	FASLG	chr1:169,847,560-169,895,386	47826	fas ligand	0	1	4
343	FRAP1	chr1:10,856,137-11,052,110	195973	FRAP1	1	5	6
344	HYOU1	chr11:118442552-118475477	32925	oxygen regulated protein precursor	0	3	4
345	IRS2	chr13:108094185-108146916	52731	insulin receptor substrate 2	1	1	6
346	LYK5	chr17:62243596-62302540	58944	protein kinase LYK5 isoform 1	2	3	4
347	MAP3K5	chr6:136848757-137104226	255469	mitogen-activated protein kinase 5	2	0	23
348	MDM2	chr12:67478238-67530481	52243	mouse double minute 2 homolog isoform MDM2	2	3	4
349	PDK1	chr2:173613380-173676069	62689	pyruvate dehydrogenase kinase, isoenzyme 1	1	0	5
350	PRKAA1	chr5:40798377-40853977	55600	PK, AMP-activated, a 1 catalytic subunit	0	1	3
351	PRKAA2	chr1:56470913-56553947	83034	AMP-activated protein kinase a 2 catalytic	0	0	9
352	PTEN	chr10:89277772-89400708	122936	PTEN	1	0	8
353	raptor	chr17:79209187-79650286	441099	raptor	13	9	40
354	RHEB	chr7:150545537-150618750	73213	Ras homolog enriched in brain	1	1	6
355	TNF	chr6:31637666-31660409	22743	tumor necrosis factor α	7	3	8
356	TNFRSF1A	chr12:6298185-6331472	33287	TNF receptor superfamily, member 1A	0	5	9
357	TSC1	chr9:131032782-131106053	73271	tuberous sclerosis 1 protein isoform 1	1	1	6
358	TSC2	chr16:2028516-2088636	60120	tuberous sclerosis 2 isoform 1	1	1	7
359	YWHAE	chr17:1,434,435-1,530,108	95673	YWHAE	0	1	8

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P14	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Apoptosis - continued							
360	YWHAG	chr7:75,548,205-75,620,405	72200	YWHAG	1	2	4
361	YWHAZ	chr8:101877497-101930260	52763	tyrosine 3/tryptophan 5 -monooxygenase	1	0	3
GSK							
362	PTGS2	chr1:183879960-183888548	8588	PTGS-endoperoxide synthase 2 precursor	1	0	3
363	SLC4A3	chr2:220694825-220709231	14406	solute carrier family 4, anion exchanger, memb	6	0	3
364	KCNV1	chr8:110935817-110943542	7725	K channel, subfamily V, member 1	3	0	5
365	NAT1	chr8:18077887-18091092	13205	N-acetyltransferase 1	1	0	5
366	GJA5	chr1:144735969-144752519	16550	gap junction protein, α 5	1	0	5
367	GJA10	chr1:38753032-38760583	7551	connexin 59	2	0	2
368	GJA3	chr13:18514119-18515427	1308	gap junction protein, α 3, 46kDa (connexin	1	0	6
369	CX62	chr6:90599785-90601417	1632	connexin 62	1	0	6
370	CX40.1	chr10:35898343-35901869	3526	connexin40.1	2	0	2
371	GJC1	chr17:38890071-38894110	4039	GJC1	0	0	1
372	GJB7	chr6:87988295-88034592	46297	hypothetical protein LOC375519	2	0	2
373	PANX3	chr11:124019104-124027901	8797	pannexin 3	2	0	3
374	BZRP	chr22:41778493-41790189	11696	peripheral benzodiazapine receptor isform PBR	3	0	4
375	CX3CL1	chr16:57181810-57194351	12541	chemokine (C-X3-C motif) ligand 1	1	0	4
376	CSEN	chr2:95447851-95536580	88729	Kv channel interacting protein 3	3	0	4
377	IL1B	chr2:113682482-113689502	7020	interleukin 1, beta proprotein	1	0	1
378	PLD1	chr3:172639557-172776785	137228	phospholipase D1, phophatidylcholine-specific	1	0	6
379	IL6	chr7:22509091-22513888	4797	interleukin 6 (interferon, beta 2)	1	0	5
380	KCNH2	chr7:150033789-150066755	32966	voltage-gated K channel, subfamily H,	1	0	5
381	RELN	chr7:102672767-103190493	517726	reelin	4	0	49
382	SGCE	chr7:93826570-93897507	70937	sarcoglycan, epsilon	1	0	4
383	PLAT	chr8:42050122-42082562	32440	plasminogen activator, tissue type isoform 1	2	0	4
384	DAPK1	chr9:85569914-85780662	210748	death-associated protein kinase 1	6	0	28
385	IFNA1	chr9:21430439-21431315	876	interferon, α 1	3	0	3
386	KCNJ11	chr11:17371106-17374515	3409	K inwardly-rectifying channel J11	3	0	2
387	ATXN2	chr12:110301970-110449433	147463	ataxin 2	4	0	3
388	SYT10	chr12:33419614-33484021	64407	synaptotagmin 10	2	0	2
389	ATP10A	chr15:23469780-23656231	186451	ATPase, Class V, type 10A	8	0	26

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P15	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
GSK - continued							
390	SLC12A6	chr15:32241721-32326986	85265	SLC12A6	4	0	2
391	A2BP1	chr16:6069424-7762298	1692874	ataxin 2-binding protein 1 isoform 1	12	0	0
392	MC3R	chr20:55509210-55510293	1083	melanocortin 3 receptor	2	0	5
Plasma proteins							
393	ALB	chr4:74736256-74753383	17127	albumin precursor	0	0	2
394	ORM1	chr9:112461380-112464802	3422	orosomucoid 1 precursor	0	0	6
395	ORM2	chr9:112468113-112471581	3468	orosomucoid 2	0	0	5
Neuromodulators of synaptic transmission							
396	NPY	chr7:24,046,080-24,093,749	47669	neuropeptide Y	1	3	5
397	NPY1R	chr4:164,802,743-164,851,375	48632	neuropeptide Y receptor Y1	0	0	3
398	NPY2R	chr4:156,687,407-156,735,854	48447	neuropeptide Y receptor Y2	0	1	4
399	NPY3R	chr2:137,062,688-137,106,494	43806	chemokine (C-X-C motif) receptor 4 isoform a	0	1	4
400	NPY4R	func only	0	pancreatic polypeptide receptor 1	0	0	5
401	NPY5R	chr4:164,822,717-164,870,711	47994	neuropeptide Y receptor Y5	0	0	4
402	NPY6R	chr5:137209677-137231630	21953	neuropeptide Y receptor Y6 (pseudogene)	0	0	2
403	BDNF	chr11:27,620,750-27,727,605	106855	brain-derived neurotrophic factor isoform c	0	2	7
404	NTRK2	chr9:82,721,744-83,114,528	392784	NTRK2	4	2	36
405	NGFR	chr17:48,027,304-48,086,999	59695	nerve growth factor receptor precursor	0	0	10
406	GAL	chr11:68,207,394-68,254,003	46609	galanin preproprotein	0	2	6
407	GALR1	chr18:73,068,709-73,129,070	60361	galanin receptor 1	0	1	6
408	GALR2	chr17:74,648,014-74,690,696	42682	galanin receptor 2	0	3	3
409	GALR3	chr22:36,442,427-36,484,541	42114	galanin receptor 3	0	1	4
Others							
410	NOVA1	chr14:24895216-25067088	171872	neuro-oncological ventral antigen 1 isoform 1	2	0	4
411	PAFAH1B1	chr17:2,683,526-2,815,479	131953	platelet-activating factor acetylhydrolase,	0	3	6
412	PDYN	chr20:1,934,402-1,989,702	55300	beta-neoendorphin-dynorphin preproprotein	0	0	4
413	CDK5	chr7:150,122,639-150,166,737	44098	cyclin-dependent kinase 5	2	2	4
414	EFHC1	chr6:52,311,947-52,424,054	112107	EF-hand domain (C-terminal) containing 1	6	4	23

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P16	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Mendelian/linkage - continued							
416	ME2	chr18:46,637,588-46,745,620	108032	malic enzyme 2, NAD(+)-dependent, mto	1	2	3
417	LGI1	chr10:95,162,264-95,242,503	80239	leucine-rich, glioma inactivated 1 precursor	1	0	10
418	LGI2	chr4:24,734,272-24,803,198	68926	leucine-rich repeat LGI family, member 2	2	1	11
419	LGI3	chr8:22,006,282-22,056,282	50000	leucine-rich repeat LGI family, member 3	0	2	5
420	LGI4	chr19:40,287,256-40,337,944	50688	leucine-rich repeat LGI family, member 4	2	4	6
421	MASS1	chr5:89,918,689-90,564,105	645416	very large G-protein coupled receptor 1	9	8	34
SOX related							
422	SOX2	chr3:182,730,634-182,773,134	42500	sex-determining region Y-box 2	0	0	3
Opioids/endocabinoids							
423	OPRM1	chr6:154,371,433-154,618,991	247558	opioid receptor, mu 1 isoform MOR-1X	1	3	59
424	OPRK1	chr8:54,171,732-54,233,932	62200	opioid receptor, kappa 1	2	2	6
425	OPRD1	chr1:28,806,165-28,897,717	91552	opioid receptor, delta 1	0	1	7
426	PENK	chr8:57,383,477-57,428,550	45073	proenkephalin	0	0	2
427	POMC	chr2:25,338,322-25,385,987	47665	proopiomelanocortin	1	1	5
428	PCSK1	chr5:95,780,191-95,863,025	82834	proprotein convertase subtilisin/kexin type 1	0	4	7
429	PCSK2	chr20:17,182,630-17,480,223	297593	proprotein convertase subtilisin/kexin type 2	9	4	29
430	CNR1	chr6:88,825,178-88,870,652	45474	central cannabinoid receptor isoform a	0	1	8
431	FAAH	chr1:46,209,908-46,269,433	59525	fatty acid amide hydrolase	0	2	4
Neurosteroids							
432	CYP11A1	chr15:72,325,920-72,395,784	69864	cytochrome P450, subfamily XIA precursor	3	2	5
433	SRD5A1	chr5:6,666,300-6,742,411	76111	steroid-5-α-reductase 1	4	1	5
434	AKR1C3	chr10:5,070,567-5,123,878	53311	aldo-keto reductase family 1, member C3	0	2	6
435	AKR1C2	chr10:4,965,965-5,034,207	68242	aldo-keto reductase family 1, member C2	2	3	9
Somatostatin							
436	SST	chr3:188,687,617-188,729,019	41402	somatostatin	1	0	3
437	SSTR1	chr14:36,647,242-36,692,307	45065	somatostatin receptor 1	0	0	7
438	SSTR2	chr17:71,738,282-71,785,183	46901	somatostatin receptor 2	0	3	6
439	SSTR3	chr22:35,825,283-35,871,392	46109	somatostatin receptor 3	1	2	8
440	SSTR4	chr20:22,991,120-23,032,287	41167	somatostatin receptor 4	0	0	4

Table 8.2 continued... Summary of candidate genes and SNP selection for large scale epilepsy project

P17	Gene	Gene coordinate	Size*	Gene Descriptions	#func non HM	#func HM	#tags
Somatostatin - continued							
441	SSTR5	chr16:1,048,869-1,089,964	41095	somatostatin receptor 5	0	2	6
cholecystokinin							
442	CCK	chr3:42,239,909-42,286,987	47078	cholecystokinin preproprotein	0	1	6
443	CCKAR	chr4:26,213,900-26,262,925	49025	cholecystokinin A receptor	0	0	5
444	CCKBR	chr11:6,225,274-6,277,665	52391	cholecystokinin B receptor	2	1	13
neurotachkynin							
445	TAC1	chr7:96,953,105-97,001,513	48408	tachykinin 1 isoform beta precursor	0	0	4
446	TACR1	chr2:75,230,766-75,420,810	190044	tachykinin receptor 1 isoform long	4	1	34
447	TACR2	chr10:70,489,184-70,540,682	51498	tachykinin receptor 2	0	3	7
448	TACR3	chr4:104,949,444-105,119,793	170349	tachykinin receptor 3	0	2	11
Zinc transporter protein							
449	GJB2	chr13:18,539,608-18,585,037	45429	gap junction protein, beta 2, 26kDa (connexin	0	0	3
450	GJE1	chr7:99,112,564-99,158,640	46076	gap junction protein, epsilon 1, 29kDa	1	2	6
451	GJB6	chr13:18,574,101-18,623,067	48966	gap junction protein, beta 6 (connexin 30)	0	0	7
452	GJB1	chrX:69,289,866-69,331,835	41969	gap junction protein, beta 1, 32kDa (connexin	0	2	7
453	CX36	chr15:32750734-32772745	22011	connexin-36	0	0	6
454	GJA1	chr6:121,717,363-121,771,448	54085	connexin 43	1	0	10
455	GJA7	chr17:43,337,159-43,378,350	41191	gap junction protein, α 7, 45kDa (connexin	0	0	3
456	GJA12	chr1:225,292,183-225,342,156	49973	connexin46.6	0	0	1
457	PANX1	chr11:93,530,246-93,623,288	93042	pannexin 1	3	2	7
458	PANX2	chr22:48,771,217-48,820,747	49530	pannexin 2	2	3	8

Bibliography

1. Pal, D.K., et al., *BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy.* Am J Hum Genet, 2003. **73**(2): p. 261-70.
2. Fisher, R.S., et al., *Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE).* Epilepsia, 2005. **46**(4): p. 470-2.
3. Forsgren, L., et al., *The epidemiology of epilepsy in Europe - a systematic review.* Eur J Neurol, 2005. **12**(4): p. 245-53.
4. Shorvon, S.D., *Handbook of Epilepsy Treatment.* 2000: Blackwell Science.
5. *Proposal for revised classification of epilepsies and epileptic syndromes. Commission on Classification and Terminology of the International League Against Epilepsy.* Epilepsia, 1989. **30**(4): p. 389-99.
6. Sieveking, S.E.H., *On Epilepsy and Epileptiform Seizures, their Causes, Pathology, and Treatment (2nd edition, 1861).* 1861, London: London.
7. Mittal, S., et al., *Long-term outcome after surgical treatment of temporal lobe epilepsy in children.* J Neurosurg, 2005. **103**(5 Suppl): p. 401-12.
8. Spencer, S.S., et al., *Predicting long-term seizure outcome after resective epilepsy surgery: the multicenter study.* Neurology, 2005. **65**(6): p. 912-8.
9. Sinha, S.R. and E.H. Kossoff, *The ketogenic diet.* Neurologist, 2005. **11**(3): p. 161-70.
10. Steinlein, O.K., et al., *A missense mutation in the neuronal nicotinic acetylcholine receptor alpha 4 subunit is associated with autosomal dominant nocturnal frontal lobe epilepsy.* Nat Genet, 1995. **11**(2): p. 201-3.
11. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.* Science, 1989. **245**(4922): p. 1066-73.
12. Sugawara, T., et al., *A missense mutation of the Na+ channel alpha II subunit gene Na(v)1.2 in a patient with febrile and afebrile seizures causes channel dysfunction.* Proc Natl Acad Sci U S A, 2001. **98**(11): p. 6384-9.
13. Heron, S.E., et al., *Sodium-channel defects in benign familial neonatal-infantile seizures.* Lancet, 2002. **360**(9336): p. 851-2.
14. Berkovic, S.F., et al., *Benign familial neonatal-infantile seizures: characterization of a new sodium channelopathy.* Ann Neurol, 2004. **55**(4): p. 550-7.
15. Wallace, R.H., et al., *Febrile seizures and generalized epilepsy associated with a mutation in the Na+-channel beta1 subunit gene SCN1B.* Nat Genet, 1998. **19**(4): p. 366-70.
16. Audenaert, D., et al., *A deletion in SCN1B is associated with febrile seizures and early-onset absence epilepsy.* Neurology, 2003. **61**(6): p. 854-6.
17. Abou-Khalil, B., et al., *Partial and generalized epilepsy with febrile seizures plus and a novel SCN1A mutation.* Neurology, 2001. **57**(12): p. 2265-72.
18. Escayg, A., et al., *Mutations of SCN1A, encoding a neuronal sodium channel, in two families with GEFS+2.* Nat Genet, 2000. **24**(4): p. 343-5.
19. Claes, L., et al., *De novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy.* Am J Hum Genet, 2001. **68**(6): p. 1327-32.

20. Escayg, A., et al., *A novel SCN1A mutation associated with generalized epilepsy with febrile seizures plus--and prevalence of variants in patients with epilepsy.* Am J Hum Genet, 2001. **68**(4): p. 866-73.
21. Sugawara, T., et al., *Nav1.1 mutations cause febrile seizures associated with afebrile partial seizures.* Neurology, 2001. **57**(4): p. 703-5.
22. Wallace, R.H., et al., *Neuronal sodium-channel alpha1-subunit mutations in generalized epilepsy with febrile seizures plus.* Am J Hum Genet, 2001. **68**(4): p. 859-65.
23. Annesi, G., et al., *Two novel SCN1A missense mutations in generalized epilepsy with febrile seizures plus.* Epilepsia, 2003. **44**(9): p. 1257-8.
24. Sugawara, T., et al., *Frequent mutations of SCN1A in severe myoclonic epilepsy in infancy.* Neurology, 2002. **58**(7): p. 1122-4.
25. Ohmori, I., et al., *Significant correlation of the SCN1A mutations and severe myoclonic epilepsy in infancy.* Biochem Biophys Res Commun, 2002. **295**(1): p. 17-23.
26. Fujiwara, T., et al., *Mutations of sodium channel alpha subunit type 1 (SCN1A) in intractable childhood epilepsies with frequent generalized tonic-clonic seizures.* Brain, 2003. **126**(Pt 3): p. 531-46.
27. Nabbout, R., et al., *Spectrum of SCN1A mutations in severe myoclonic epilepsy of infancy.* Neurology, 2003. **60**(12): p. 1961-7.
28. Wallace, R.H., et al., *Sodium channel alpha1-subunit mutations in severe myoclonic epilepsy of infancy and infantile spasms.* Neurology, 2003. **61**(6): p. 765-9.
29. Kimura, K., et al., *A missense mutation in SCN1A in brothers with severe myoclonic epilepsy in infancy (SMEI) inherited from a father with febrile seizures.* Brain Dev, 2005. **27**(6): p. 424-30.
30. Mulley, J.C., et al., *SCN1A mutations and epilepsy.* Hum Mutat, 2005. **25**(6): p. 535-42.
31. Jouvenceau, A., et al., *Human epilepsy associated with dysfunction of the brain P/Q-type calcium channel.* Lancet, 2001. **358**(9284): p. 801-7.
32. Chioza, B., et al., *Association between the alpha(1a) calcium channel gene CACNA1A and idiopathic generalized epilepsy.* Neurology, 2001. **56**(9): p. 1245-6.
33. Ophoff, R.A., et al., *Familial hemiplegic migraine and episodic ataxia type-2 are caused by mutations in the Ca2+ channel gene CACNL1A4.* Cell, 1996. **87**(3): p. 543-52.
34. Vitko, I., et al., *Functional characterization and neuronal modeling of the effects of childhood absence epilepsy variants of CACNA1H, a T-type calcium channel.* J Neurosci, 2005. **25**(19): p. 4844-55.
35. Charlier, C., et al., *A pore mutation in a novel KQT-like potassium channel gene in an idiopathic epilepsy family.* Nat Genet, 1998. **18**(1): p. 53-5.
36. Dedek, K., et al., *Myokymia and neonatal epilepsy caused by a mutation in the voltage sensor of the KCNQ2 K+ channel.* Proc Natl Acad Sci U S A, 2001. **98**(21): p. 12272-7.
37. Singh, N.A., et al., *KCNQ2 and KCNQ3 potassium channel genes in benign familial neonatal convulsions: expansion of the functional and mutation spectrum.* Brain, 2003. **126**(Pt 12): p. 2726-37.
38. Biervert, C., et al., *A potassium channel mutation in neonatal human epilepsy.* Science, 1998. **279**(5349): p. 403-6.

39. Pereira, S., et al., *Complete loss of the cytoplasmic carboxyl terminus of the KCNQ2 potassium channel: a novel mutation in a large Czech pedigree with benign neonatal convulsions or other epileptic phenotypes.* Epilepsia, 2004. **45**(4): p. 384-90.
40. Tang, B., et al., *A novel mutation in KCNQ2 gene causes benign familial neonatal convulsions in a Chinese family.* J Neurol Sci, 2004. **221**(1-2): p. 31-4.
41. Singh, N.A., et al., *A novel potassium channel gene, KCNQ2, is mutated in an inherited epilepsy of newborns.* Nat Genet, 1998. **18**(1): p. 25-9.
42. Zuberi, S.M., et al., *A novel mutation in the human voltage-gated potassium channel gene (Kv1.1) associates with episodic ataxia type 1 and sometimes with partial epilepsy.* Brain, 1999. **122 (Pt 5)**: p. 817-25.
43. Haug, K., et al., *Mutations in CLCN2 encoding a voltage-gated chloride channel are associated with idiopathic generalized epilepsies.* Nat Genet, 2003. **33**(4): p. 527-32.
44. De Fusco, M., et al., *The nicotinic receptor beta 2 subunit is mutant in nocturnal frontal lobe epilepsy.* Nat Genet, 2000. **26**(3): p. 275-6.
45. Phillips, H.A., et al., *CHRNB2 is the second acetylcholine receptor subunit associated with autosomal dominant nocturnal frontal lobe epilepsy.* Am J Hum Genet, 2001. **68**(1): p. 225-31.
46. Cossette, P., et al., *Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy.* Nat Genet, 2002. **31**(2): p. 184-9.
47. Baulac, S., et al., *First genetic evidence of GABA(A) receptor dysfunction in epilepsy: a mutation in the gamma2-subunit gene.* Nat Genet, 2001. **28**(1): p. 46-8.
48. Wallace, R.H., et al., *Mutant GABA(A) receptor gamma2-subunit in childhood absence epilepsy and febrile seizures.* Nat Genet, 2001. **28**(1): p. 49-52.
49. Kananura, C., et al., *A splice-site mutation in GABRG2 associated with childhood absence epilepsy and febrile convulsions.* Arch Neurol, 2002. **59**(7): p. 1137-41.
50. Harkin, L.A., et al., *Truncation of the GABA(A)-receptor gamma2 subunit in a family with generalized epilepsy with febrile seizures plus.* Am J Hum Genet, 2002. **70**(2): p. 530-6.
51. Dibbens, L.M., et al., *GABRD encoding a protein for extra- or peri-synaptic GABAA receptors is a susceptibility locus for generalized epilepsies.* Hum Mol Genet, 2004. **13**(13): p. 1315-9.
52. Kalachikov, S., et al., *Mutations in LGI1 cause autosomal-dominant partial epilepsy with auditory features.* Nat Genet, 2002. **30**(3): p. 335-41.
53. Gu, W., E. Brodtkorb, and O.K. Steinlein, *LGI1 is mutated in familial temporal lobe epilepsy characterized by aphasic seizures.* Ann Neurol, 2002. **52**(3): p. 364-7.
54. Bisulli, F., et al., *A de novo LGI1 mutation in sporadic partial epilepsy with auditory features.* Ann Neurol, 2004. **56**(3): p. 455-6.
55. Suzuki, T., et al., *Mutations in EFHC1 cause juvenile myoclonic epilepsy.* Nat Genet, 2004. **36**(8): p. 842-9.
56. Vanmolkot, K.R., et al., *Novel mutations in the Na+, K+-ATPase pump gene ATP1A2 associated with familial hemiplegic migraine and benign familial infantile convulsions.* Ann Neurol, 2003. **54**(3): p. 360-6.
57. Nakayama, J., et al., *A nonsense mutation of the MASS1 gene in a family with febrile and afebrile seizures.* Ann Neurol, 2002. **52**(5): p. 654-7.
58. Gardiner, R.M., *Impact of our understanding of the genetic aetiology of epilepsy.* J Neurol, 2000. **247**(5): p. 327-34.

59. Martin, N., D. Boomsma, and G. Machin, *A twin-pronged attack on complex traits.* Nat Genet, 1997. **17**(4): p. 387-92.
60. Berkovic, S.F., et al., *Epilepsies in twins: genetics of the major epilepsy syndromes.* Ann Neurol, 1998. **43**(4): p. 435-45.
61. Vadlamudi, L., et al., *Epilepsy in twins: insights from unique historical data of William Lennox.* Neurology, 2004. **62**(7): p. 1127-33.
62. Kjeldsen, M.J., et al., *Epileptic seizures and syndromes in twins: the importance of genetic factors.* Epilepsy Res, 2003. **55**(1-2): p. 137-46.
63. Winawer, M.R., et al., *Genetic influences on myoclonic and absence seizures.* Neurology, 2003. **61**(11): p. 1576-81.
64. Winawer, M.R., et al., *Familial clustering of seizure types within the idiopathic generalized epilepsies.* Neurology, 2005. **65**(4): p. 523-8.
65. Hirschhorn, J.N., et al., *A comprehensive review of genetic association studies.* Genet Med, 2002. **4**(2): p. 45-61.
66. Botstein, D. and N. Risch, *Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease.* Nat Genet, 2003. **33 Suppl**: p. 228-37.
67. Glazier, A.M., J.H. Nadeau, and T.J. Aitman, *Finding genes that underlie complex traits.* Science, 2002. **298**(5602): p. 2345-9.
68. Ioannidis, J.P., E.E. Ntzani, and T.A. Trikalinos, *'Racial' differences in genetic effects for complex diseases.* Nat Genet, 2004. **36**(12): p. 1312-8.
69. Tate, S.K., et al., *Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin.* Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5507-12.
70. Weedon, M.N., et al., *Meta-analysis and a large association study confirm a role for calpain-10 variation in type 2 diabetes susceptibility.* Am J Hum Genet, 2003. **73**(5): p. 1208-12.
71. McCarthy, M.I., *Progress in defining the molecular basis of type 2 diabetes mellitus through susceptibility-gene identification.* Hum Mol Genet, 2004. **13 Spec No 1**: p. R33-41.
72. Zondervan, K.T. and L.R. Cardon, *The complex interplay among factors that influence allelic association.* Nat Rev Genet, 2004. **5**(2): p. 89-100.
73. Pastor, P., et al., *Apolipoprotein Eepsilon4 modifies Alzheimer's disease onset in an E280A PS1 kindred.* Ann Neurol, 2003. **54**(2): p. 163-9.
74. Lohmueller, K.E., et al., *Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease.* Nat Genet, 2003. **33**(2): p. 177-82.
75. Kruglyak, L. and D.A. Nickerson, *Variation is the spice of life.* Nat Genet, 2001. **27**(3): p. 234-6.
76. Wright, A.F. and N.D. Hastie, *Complex genetic diseases: controversy over the Croesus code.* Genome Biol, 2001. **2**(8): p. COMMENT2007.
77. Elmslie, F.V., et al., *Genetic mapping of a major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q.* Hum Mol Genet, 1997. **6**(8): p. 1329-34.
78. Zara, F., et al.. *Mapping of genes predisposing to idiopathic generalized epilepsy.* Hum Mol Genet, 1995. **4**(7): p. 1201-7.
79. Sander, T., et al., *Genome search for susceptibility loci of common idiopathic generalised epilepsies.* Hum Mol Genet, 2000. **9**(10): p. 1465-72.
80. Sander, T., et al., *Refined mapping of the epilepsy susceptibility locus EJM1 on chromosome 6.* Neurology, 1997. **49**(3): p. 842-7.

81. Windemuth, C., et al., *No evidence for a susceptibility locus for idiopathic generalized epilepsy on chromosome 5 in families with typical absence seizures.* Epilepsy Res, 2002. **51**(1-2): p. 23-9.
82. Sander, T., et al., *Exploration of a putative susceptibility locus for idiopathic generalized epilepsy on chromosome 8p12.* Epilepsia, 2003. **44**(1): p. 32-9.
83. Sander, T., et al., *No evidence for a susceptibility locus for idiopathic generalized epilepsy on chromosome 18q21.1.* Am J Med Genet, 2002. **114**(6): p. 673-8.
84. Greenberg, D.A., et al., *Juvenile myoclonic epilepsy (JME) may be linked to the BF and HLA loci on human chromosome 6.* Am J Med Genet, 1988. **31**(1): p. 185-92.
85. Durner, M., et al., *Localization of idiopathic generalized epilepsy on chromosome 6p in families of juvenile myoclonic epilepsy patients.* Neurology, 1991. **41**(10): p. 1651-5.
86. Weissbecker, K.A., et al., *Confirmation of linkage between juvenile myoclonic epilepsy locus and the HLA region of chromosome 6.* Am J Med Genet, 1991. **38**(1): p. 32-6.
87. Greenberg, D.A., et al., *Reproducibility and complications in gene searches: linkage on chromosome 6, heterogeneity, association, and maternal inheritance in juvenile myoclonic epilepsy.* Am J Hum Genet, 2000. **66**(2): p. 508-16.
88. Whitehouse, W.P., et al., *Linkage analysis of idiopathic generalized epilepsy (IGE) and marker loci on chromosome 6p in families of patients with juvenile myoclonic epilepsy: no evidence for an epilepsy locus in the HLA region.* Am J Hum Genet, 1993. **53**(3): p. 652-62.
89. Obach, V., et al., *No evidence of linkage to 6p markers in spanish families with juvenile myoclonic epilepsy.* Neurosci Lett, 2000. **286**(3): p. 213-7.
90. Pinto, D., et al., *Evidence for linkage between juvenile myoclonic epilepsy-related idiopathic generalized epilepsy and 6p11-12 in Dutch families.* Epilepsia, 2004. **45**(3): p. 211-7.
91. Liu, A.W., et al., *Juvenile myoclonic epilepsy in chromosome 6p12-p11: locus heterogeneity and recombinations.* Am J Med Genet, 1996. **63**(3): p. 438-46.
92. Tauer, U., et al., *Genetic dissection of photosensitivity and its relation to idiopathic generalized epilepsy.* Ann Neurol, 2005. **57**(6): p. 866-73.
93. Nakayama, J., et al., *Linkage and association of febrile seizures to the IMPA2 gene on human chromosome 18.* Neurology, 2004. **63**(10): p. 1803-7.
94. Pinto, D., et al., *Genome-wide linkage scan of epilepsy-related photoparoxysmal electroencephalographic response: evidence for linkage on chromosomes 7q32 and 16p13.* Hum Mol Genet, 2005. **14**(1): p. 171-8.
95. Durner, M., et al., *Genome scan of idiopathic generalized epilepsy: evidence for major susceptibility gene and modifying genes influencing the seizure type.* Ann Neurol, 2001. **49**(3): p. 328-35.
96. Robinson, R., et al., *Linkage analysis between childhood absence epilepsy and genes encoding GABAA and GABAB receptors, voltage-dependent calcium channels, and the ECA1 region on chromosome 8q.* Epilepsy Res, 2002. **48**(3): p. 169-79.
97. Bai, D., et al., *Juvenile myoclonic epilepsy: linkage to chromosome 6p12 in Mexico families.* Am J Med Genet, 2002. **113**(3): p. 268-74.
98. Nakayama, J., et al., *Significant evidence for linkage of febrile seizures to chromosome 5q14-q15.* Hum. Mol. Genet., 2000. **9**(1): p. 87-91.

99. Durner, M., et al., *Evidence for linkage of adolescent-onset idiopathic generalized epilepsies to chromosome 8-and genetic heterogeneity.* Am J Hum Genet, 1999. **64**(5): p. 1411-9.
100. Sander, T., et al., *Evaluation of a putative major susceptibility locus for juvenile myoclonic epilepsy on chromosome 15q14.* Am J Med Genet, 1999. **88**(2): p. 182-7.
101. Sander, T., et al., *Replication analysis of a putative susceptibility locus (EGI) for idiopathic generalized epilepsy on chromosome 8q24.* Epilepsia, 1998. **39**(7): p. 715-20.
102. Sander, T., et al., *Linkage analysis between idiopathic generalized epilepsies and the GABA(A) receptor alpha5, beta3 and gamma3 subunit gene cluster on chromosome 15.* Acta Neurol Scand, 1997. **96**(1): p. 1-7.
103. Sander, T., et al., *Exclusion of linkage between idiopathic generalized epilepsies and the GABAA receptor alpha 1 and gamma 2 subunit gene cluster on chromosome 5.* Epilepsy Res, 1996. **23**(3): p. 235-44.
104. Elmslie, F.V., et al., *Linkage analysis of juvenile myoclonic epilepsy and microsatellite loci spanning 61 cM of human chromosome 6p in 19 nuclear pedigrees provides no evidence for a susceptibility locus in this region.* Am J Hum Genet, 1996. **59**(3): p. 653-63.
105. Sander, T., et al., *The phenotypic spectrum related to the human epilepsy susceptibility gene "EJM1".* Ann Neurol, 1995. **38**(2): p. 210-7.
106. Cardon, L.R. and J.I. Bell, *Association study designs for complex diseases.* Nat Rev Genet, 2001. **2**(2): p. 91-9.
107. Risch, N. and K. Merikangas, *The future of genetic studies of complex human diseases.* Science, 1996. **273**(5281): p. 1516-7.
108. Kohler, K. and H. Bickeboller, *Case-control association tests correcting for population stratification.* Ann Hum Genet, 2006. **70**(Pt 1): p. 98-115.
109. Devlin, B. and N. Risch, *A comparison of linkage disequilibrium measures for fine-scale mapping.* Genomics, 1995. **29**(2): p. 311-22.
110. Pritchard, J.K. and N.A. Rosenberg, *Use of unlinked genetic markers to detect population stratification in association studies.* Am J Hum Genet, 1999. **65**(1): p. 220-8.
111. Reich, D.E. and D.B. Goldstein, *Detecting association in a case-control study while correcting for population stratification.* Genet Epidemiol, 2001. **20**(1): p. 4-16.
112. Falk, C.T. and P. Rubinstein, *Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations.* Ann Hum Genet, 1987. **51 (Pt 3)**: p. 227-33.
113. Spielman, R.S., R.E. McGinnis, and W.J. Ewens, *Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM).* Am J Hum Genet, 1993. **52**(3): p. 506-16.
114. Kanemoto, K., et al., *Interleukin (IL)1beta, IL-1alpha, and IL-1 receptor antagonist gene polymorphisms in patients with temporal lobe epilepsy.* Ann Neurol, 2000. **47**(5): p. 571-4.
115. Virta, M., M. Hurme, and M. Helminen, *Increased frequency of interleukin-1beta (-511) allele 2 in febrile seizures.* Pediatr Neurol, 2002. **26**(3): p. 192-5.
116. Tilgen, N., et al., *Association analysis between the human interleukin 1beta (-511) gene polymorphism and susceptibility to febrile convulsions.* Neurosci Lett, 2002. **334**(1): p. 68-70.

117. Heils, A., et al., *Interleukin-1beta gene polymorphism and susceptibility to temporal lobe epilepsy with hippocampal sclerosis.* Ann Neurol, 2000. **48**(6): p. 948-50.
118. Haspolat, S., et al., *Interleukin-1alpha, interleukin-1beta, and interleukin-1Ra polymorphisms in febrile seizures.* J Child Neurol, 2005. **20**(7): p. 565-8.
119. Jin, L., et al., *Association analysis of a polymorphism of interleukin 1 beta (IL-1 beta) gene with temporal lobe epilepsy in a Chinese population.* Epilepsia, 2003. **44**(10): p. 1306-9.
120. Peltola, J., et al., *Polymorphism of the interleukin-1 gene complex in localization-related epilepsy.* Ann Neurol, 2001. **50**(2): p. 275-6.
121. Chakravarti, A., et al., *Nonuniform recombination within the human beta-globin gene cluster.* Am J Hum Genet, 1984. **36**(6): p. 1239-58.
122. Jeffreys, A.J., A. Ritchie, and R. Neumann, *High resolution analysis of haplotype diversity and meiotic crossover in the human TAP2 recombination hotspot.* Hum Mol Genet, 2000. **9**(5): p. 725-33.
123. Kruglyak, L., *Prospects for whole-genome linkage disequilibrium mapping of common disease genes.* Nat Genet, 1999. **22**(2): p. 139-44.
124. Reich, D.E., et al., *Linkage disequilibrium in the human genome.* Nature, 2001. **411**(6834): p. 199-204.
125. Daly, M.J., et al., *High-resolution haplotype structure in the human genome.* Nat Genet, 2001. **29**(2): p. 229-32.
126. Johnson, G.C., et al., *Haplotype tagging for the identification of common disease genes.* Nat Genet, 2001. **29**(2): p. 233-7.
127. Chen, J., et al., *Activity-induced expression of common reference genes in individual cns neurons.* Lab Invest, 2001. **81**(6): p. 913-6.
128. R foundation for Statisical Computing, V., Austria, *R Development Core Team. R: A language and environment for statistical computing.* 2004.
129. Weale, M.E., et al., *Selection and evaluation of tagging SNPs in the neuronal-sodium-channel gene SCN1A: implications for linkage-disequilibrium gene mapping.* Am J Hum Genet, 2003. **73**(3): p. 551-65.
130. Pritchard, J.K. and M. Przeworski, *Linkage disequilibrium in humans: models and data.* Am J Hum Genet, 2001. **69**(1): p. 1-14.
131. Chapman, J.M., et al., *Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power.* Hum Hered, 2003. **56**(1-3): p. 18-31.
132. Goldstein, D.B., et al., *Genome scans and candidate gene approaches in the study of common diseases and variable drug responses.* Trends Genet, 2003. **19**(11): p. 615-22.
133. Ahmadi, K.R., et al., *A single-nucleotide polymorphism tagging set for human drug metabolism and transport.* Nat Genet, 2005. **37**(1): p. 84-9.
134. Krauss, G.L., M.A. Johnson, and N.R. Miller, *Vigabatrin-associated retinal cone system dysfunction: electroretinogram and ophthalmologic findings.* Neurology, 1998. **50**(3): p. 614-8.
135. Newman, W.D., K. Tocher, and J.F. Acheson, *Vigabatrin associated visual field loss: a clinical audit to study prevalence, drug history and effects of drug withdrawal.* Eye, 2002. **16**(5): p. 567-71.
136. van der Torren, K., H.S. Graniewski-Wijnands, and B.C. Polak, *Visual field and electrophysiological abnormalities due to vigabatrin.* Doc Ophthalmol, 2002. **104**(2): p. 181-8.

137. Baulac, S., et al., *Evidence for digenic inheritance in a family with both febrile convulsions and temporal lobe epilepsy implicating chromosomes 18qter and 1q25-q31.* Ann Neurol, 2001. **49**(6): p. 786-92.
138. Berkovic, S.F., et al., *Familial temporal lobe epilepsy: a common disorder identified in twins.* Ann Neurol, 1996. **40**(2): p. 227-35.
139. Claes, L., et al., *Novel locus on chromosome 12q22-q23.3 responsible for familial temporal lobe epilepsy associated with febrile seizures.* J Med Genet, 2004. **41**(9): p. 710-4.
140. Scheffer, I.E., et al., *Familial partial epilepsy with variable foci: a new partial epilepsy syndrome with suggestion of linkage to chromosome 2.* Ann Neurol, 1998. **44**(6): p. 890-9.
141. Xiong, L., et al., *Mapping of a gene determining familial partial epilepsy with variable foci to chromosome 22q11-q12.* Am J Hum Genet, 1999. **65**(6): p. 1698-710.
142. Kinton, L., et al., *Partial epilepsy with pericentral spikes: a new familial epilepsy syndrome with evidence for linkage to chromosome 4p15.* Ann Neurol, 2002. **51**(6): p. 740-9.
143. Kearney, J.A., et al., *A gain-of-function mutation in the sodium channel gene Scn2a results in seizures and behavioral abnormalities.* Neuroscience, 2001. **102**(2): p. 307-17.
144. Liu, W., et al., *Jerky, a protein deficient in a mouse epilepsy model, is associated with translationally inactive mRNA in neurons.* J Neurosci, 2002. **22**(1): p. 176-82.
145. Bohm, D., et al., *Disruption of PLC-beta 1-mediated signal transduction in mutant mice causes age-dependent hippocampal mossy fiber sprouting and neurodegeneration.* Mol Cell Neurosci, 2002. **21**(4): p. 584-601.
146. Falconer, M.A., *Genetic and related aetiological factors in temporal lobe epilepsy. A review.* Epilepsia, 1971. **12**(1): p. 13-31.
147. Baulac, S., et al., *Fever, genes, and epilepsy.* Lancet Neurol, 2004. **3**(7): p. 421-30.
148. Tsuboi, T., *Genetic analysis of febrile convulsions: twin and family studies.* Hum Genet, 1987. **75**(1): p. 7-14.
149. Tsuboi, T. and S. Endo, *Genetic studies of febrile convulsions: analysis of twin and family data.* Epilepsy Res Suppl, 1991. **4**: p. 119-28.
150. Stogmann, E., et al., *A functional polymorphism in the prodynorphin gene promotor is associated with temporal lobe epilepsy.* Ann Neurol, 2002. **51**(2): p. 260-3.
151. Gambardella, A., et al., *GABA(B) receptor 1 polymorphism (G1465A) is associated with temporal lobe epilepsy.* Neurology, 2003. **60**(4): p. 560-3.
152. Walz, R., et al., *Surgical outcome in mesial temporal sclerosis correlates with prion protein gene variant.* Neurology, 2003. **61**(9): p. 1204-10.
153. Briellmann, R.S., et al., *APOE epsilon4 genotype is associated with an earlier onset of chronic temporal lobe epilepsy.* Neurology, 2000. **55**(3): p. 435-7.
154. Chou, I.C., et al., *Association of the neuronal nicotinic acetylcholine receptor subunit alpha4 polymorphisms with febrile convulsions.* Epilepsia, 2003. **44**(8): p. 1089-93.
155. Chou, I.C., et al., *Association analysis of gamma 2 subunit of gamma-aminobutyric acid type A receptor polymorphisms with febrile seizures.* Pediatr Res, 2003. **54**(1): p. 26-9.

156. Tan, N.C., J.C. Mulley, and S.F. Berkovic, *Genetic association studies in epilepsy: "the truth is out there".* Epilepsia, 2004. **45**(11): p. 1429-42.
157. Buono, R.J., et al., *Lack of association between an interleukin 1 beta (IL-1beta) gene variation and refractory temporal lobe epilepsy.* Epilepsia, 2001. **42**(6): p. 782-4.
158. Tilgen, N., et al., *Prodynorphin gene promoter polymorphism and temporal lobe epilepsy.* Ann Neurol, 2003. **53**(2): p. 280-1; author reply 281-2.
159. Gambardella, A., et al., *Prodynorphin gene promoter polymorphism and temporal lobe epilepsy.* Epilepsia, 2003. **44**(9): p. 1255-6.
160. Blumcke, I., et al., *The apolipoprotein E epsilon 4 allele is not associated with early onset temporal lobe epilepsy.* Neuroreport, 1997. **8**(5): p. 1235-7.
161. Gambardella, A., et al., *Apolipoprotein E polymorphisms and the risk of nonlesional temporal lobe epilepsy.* Epilepsia, 1999. **40**(12): p. 1804-7.
162. Yeni, S.N., et al., *Association between APOE polymorphisms and mesial temporal lobe epilepsy with hippocampal sclerosis.* Eur J Neurol, 2005. **12**(2): p. 103-7.
163. Ren, L., et al., *Lack of GABABR1 gene variation (G1465A) in a Chinese population with temporal lobe epilepsy.* Seizure, 2005. **14**(8): p. 611-3.
164. Mulley, J., et al., *Febrile convulsions and genetic susceptibility: role of the neuronal nicotinic acetylcholine receptor alpha 4 subunit.* Epilepsia, 2004. **45**(5): p. 561; author reply 561-2.
165. Ma, S., et al., *The GABBR1 locus and the G1465A variant is not associated with temporal lobe epilepsy preceded by febrile seizures.* BMC Med Genet, 2005. **6**: p. 13.
166. Salzmann, A., et al., *GABA receptor 1 polymorphism (G1465A) and temporal lobe epilepsy.* Epilepsia, 2005. **46**(6): p. 931-3.
167. Nakayama, J., et al., *Failure to find causal mutations in the GABA(A)-receptor gamma2 subunit (GABRG2) gene in Japanese febrile seizure patients.* Neurosci Lett, 2003. **343**(2): p. 117-20.
168. Andrew, T., et al., *Are twins and singletons comparable? A study of disease-related and lifestyle characteristics in adult women.* Twin Res, 2001. **4**(6): p. 464-77.
169. Kanemoto, K., et al., *Increased frequency of interleukin-1beta-511T allele in patients with temporal lobe epilepsy, hippocampal sclerosis, and prolonged febrile convulsion.* Epilepsia, 2003. **44**(6): p. 796-9.
170. Neale, B.M. and P.C. Sham, *The future of association studies: gene-based analysis and replication.* Am J Hum Genet, 2004. **75**(3): p. 353-62.
171. Siddiqui, A., et al., *Association of multidrug resistance in epilepsy with a polymorphism in the drug-transporter gene ABCB1.* N Engl J Med, 2003. **348**(15): p. 1442-8.
172. Tan, N.C., et al., *Failure to confirm association of a polymorphism in ABCB1 with multidrug-resistant epilepsy.* Neurology, 2004. **63**(6): p. 1090-2.
173. Zimprich, F., et al., *Association of an ABCB1 gene haplotype with pharmacoresistance in temporal lobe epilepsy.* Neurology, 2004. **63**(6): p. 1087-9.
174. Janz, D., *Juvenile Myoclonic Epilepsy in "Epilepsy A Comprehensive Textbook".* 1997. p. 2389-2400.
175. Escayg, A., et al., *Coding and noncoding variation of the human calcium-channel beta4-subunit gene CACNB4 in patients with idiopathic generalized epilepsy and episodic ataxia.* Am J Hum Genet, 2000. **66**(5): p. 1531-9.

176. Wilkie, H., et al., *Association of mu-opioid receptor subunit gene and idiopathic generalized epilepsy.* Neurology, 2002. **59**(5): p. 724-8.
177. Sander, T., et al., *Association of the 867Asp variant of the human anion exchanger 3 gene with common subtypes of idiopathic generalized epilepsy.* Epilepsy Res, 2002. **51**(3): p. 249-55.
178. Bottini, N., et al., *Convulsive disorder and the genetics of signal transduction; a study of a low molecular weight protein tyrosine phosphatase in a pediatric sample.* Neurosci Lett, 2002. **333**(3): p. 159-62.
179. Feucht, M., et al., *Possible association between childhood absence epilepsy and the gene encoding GABRB3.* Biol Psychiatry, 1999. **46**(7): p. 997-1002.
180. Sander, T., et al., *Allelic association of juvenile absence epilepsy with a GluR5 kainate receptor gene (GRIK1) polymorphism.* Am J Med Genet, 1997. **74**(4): p. 416-21.
181. Chioza, B., et al., *Suggestive evidence for association of two potassium channel genes with different idiopathic generalised epilepsy syndromes.* Epilepsy Res, 2002. **52**(2): p. 107-16.
182. Sander, T., et al., *Failure to replicate an allelic association between an exon 8 polymorphism of the human alpha(1A) calcium channel gene and common syndromes of idiopathic generalized epilepsy.* Epilepsy Res, 2002. **49**(2): p. 173-7.
183. Izzi, C., et al., *Sequencing of the GRIK1 gene in patients with juvenile absence epilepsy does not reveal mutations affecting receptor structure.* Am J Med Genet, 2002. **114**(3): p. 354-9.
184. Goldstein, D.B. and J.N. Hirschhorn, *In genetic control of disease, does 'race' matter?* Nat Genet, 2004. **36**(12): p. 1243-4.
185. Puranam, R.S., et al., *A locus for generalized tonic-clonic seizure susceptibility maps to chromosome 10q25-q26.* Ann Neurol, 2005. **58**(3): p. 449-58.
186. Greenberg, D.A., et al., *Malic enzyme 2 may underlie susceptibility to adolescent-onset idiopathic generalized epilepsy.* Am J Hum Genet, 2005. **76**(1): p. 139-46.
187. Fong, G.C., et al., *Childhood absence epilepsy with tonic-clonic seizures and electroencephalogram 3-4-Hz spike and multispike-slow wave complexes: linkage to chromosome 8q24.* Am J Hum Genet, 1998. **63**(4): p. 1117-29.
188. Ben-Menachem, E., *Vigabatrin.* Epilepsia, 1995. **36 Suppl 2**: p. S95-104.
189. Mackay, M., S. Weiss, and O.C. Snead, 3rd, *Treatment of infantile spasms: an evidence-based approach.* Int Rev Neurobiol, 2002. **49**: p. 157-84.
190. Lux, A.L., et al., *The United Kingdom Infantile Spasms Study (UKISS) comparing hormone treatment with vigabatrin on developmental and epilepsy outcomes to age 14 months: a multicentre randomised trial.* Lancet Neurol, 2005. **4**(11): p. 712-7.
191. Eke, T., J.F. Talbot, and M.C. Lawden, *Severe persistent visual field constriction associated with vigabatrin.* Bmj, 1997. **314**(7075): p. 180-1.
192. Malmgren, K., E. Ben-Menachem, and L. Frisen, *Vigabatrin visual toxicity: evolution and dose dependence.* Epilepsia, 2001. **42**(5): p. 609-15.
193. Schmitz, B., et al., *Visual field constriction in epilepsy patients treated with vigabatrin and other antiepileptic drugs: a prospective study.* J Neurol, 2002. **249**(4): p. 469-75.
194. Manuchehri, K., et al., *A controlled study of vigabatrin and visual abnormalities.* Br J Ophthalmol, 2000. **84**(5): p. 499-505.

195. Fledelius, H.C., *Vigabatrin-associated visual field constriction in a longitudinal series. Reversibility suggested after drug withdrawal.* Acta Ophthalmol Scand, 2003. **81**(1): p. 41-6.
196. McDonagh, J., et al., *Peripheral retinal dysfunction in patients taking vigabatrin.* Neurology, 2003. **61**(12): p. 1690-4.
197. Hardus, P., et al., *Vigabatrin: longterm follow-up of electrophysiology and visual field examinations.* Acta Ophthalmol Scand, 2003. **81**(5): p. 459-65.
198. P. Kinirons, G.L.C., D. O' Rourke, C.P. Doherty, I. Reid, P. Logan, B. Liggan, N. Delanty., *Vigabatrin retinopathy in an Irish cohort – lack of correlation with dose.* Epilepsia, 2006(in press).
199. Nicolson, A., et al., *The legacy of vigabatrin in a regional epilepsy clinic.* J Neurol Neurosurg Psychiatry, 2002. **73**(3): p. 327-9.
200. Weinshilboum, R.M. and S.L. Sladek, *Mercaptopurine pharmacogenetics: monogenic inheritance of erythrocyte thiopurine methyltransferase activity.* Am J Hum Genet, 1980. **32**(5): p. 651-62.
201. Sills, G.J., et al., *Visual field constriction: accumulation of vigabatrin but not tiagabine in the retina.* Neurology, 2001. **57**(2): p. 196-200.
202. Sills, G.J., et al., *Vigabatrin, but not gabapentin or topiramate, produces concentration-related effects on enzymes and intermediates of the GABA shunt in rat brain and retina.* Epilepsia, 2003. **44**(7): p. 886-92.
203. Takanaga, H., et al., *GAT2/BGT-1 as a system responsible for the transport of gamma-aminobutyric acid at the mouse blood-brain barrier.* J Cereb Blood Flow Metab, 2001. **21**(10): p. 1232-9.
204. Goodchild, M. and M.J. Neal, *The uptake of 3Hp -aminobutyric acid by the retina.* Br J Pharmacol, 1973. **47**(3): p. 529-42.
205. Pow, D.V., W. Baldridge, and D.K. Crook, *Activity-dependent transport of GABA analogues into specific cell types demonstrated at high resolution using a novel immunocytochemical strategy.* Neuroscience, 1996. **73**(4): p. 1129-43.
206. Wu, Y., W. Wang, and G.B. Richerson, *Vigabatrin induces tonic inhibition via GABA transporter reversal without increasing vesicular GABA release.* J Neurophysiol, 2003. **89**(4): p. 2021-34.
207. Chebib, M., *GABAC receptor ion channels.* Clin Exp Pharmacol Physiol, 2004. **31**(11): p. 800-4.
208. Staley, K.J., B.L. Soldo, and W.R. Proctor, *Ionic mechanisms of neuronal excitation by inhibitory GABAA receptors.* Science, 1995. **269**(5226): p. 977-81.
209. Altshuler, D., et al., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.
210. Nousiainen, I., M. Mantyjarvi, and R. Kalviainen, *No reversion in vigabatrin-associated visual field defects.* Neurology, 2001. **57**(10): p. 1916-7.
211. Johnson, M.A., et al., *Visual function loss from vigabatrin. effect Of stopping the drug.* Am J Ophthalmol, 2000. **130**(5): p. 691.
212. Schaid, D.J., et al., *Score tests for association between traits and haplotypes when linkage phase is ambiguous.* Am J Hum Genet, 2002. **70**(2): p. 425-34.
213. Relling, M.V., et al., *Mercaptopurine therapy intolerance and heterozygosity at the thiopurine S-methyltransferase gene locus.* J Natl Cancer Inst, 1999. **91**(23): p. 2001-8.
214. Milano, G. and H.L. McLeod, *Can dihydropyrimidine dehydrogenase impact 5-fluorouracil-based treatment?* Eur J Cancer, 2000. **36**(1): p. 37-42.

215. Iyer, L., et al., *Phenotype-genotype correlation of in vitro SN-38 (active metabolite of irinotecan) and bilirubin glucuronidation in human liver tissue with UGT1A1 promoter polymorphism.* Clin Pharmacol Ther, 1999. **65**(5): p. 576-82.
216. Ross, D.F., et al., *Variability of visual field measurements in normal subjects and patients with retinitis pigmentosa.* Arch Ophthalmol, 1984. **102**(7): p. 1004-10.
217. Krakow, K., et al., *Recovery of visual field constriction following discontinuation of vigabatrin.* Seizure, 2000. **9**(4): p. 287-90.
218. Versino, M. and P. Veggiotti, *Reversibility of vigabratin-induced visual-field defect.* Lancet, 1999. **354**(9177): p. 486.
219. Hisama, F.M., et al., *GABA and the ornithine delta-aminotransferase gene in vigabatrin-associated visual field defects.* Seizure, 2001. **10**(7): p. 505-7.
220. Halldorsson, B.V., S. Istrail, and F.M. De La Vega, *Optimal selection of SNP markers for disease association studies.* Hum Hered, 2004. **58**(3-4): p. 190-202.
221. Henshall, D.C. and R.P. Simon, *Epilepsy and apoptosis pathways.* J Cereb Blood Flow Metab, 2005. **25**(12): p. 1557-72.
222. Shinoda, S., et al., *Development of a model of seizure-induced hippocampal injury with features of programmed cell death in the BALB/c mouse.* J Neurosci Res, 2004. **76**(1): p. 121-8.
223. Hemminger, B.M., B. Saelim, and P.F. Sullivan, *TAMAL: an integrated approach to choosing SNPs for genetic studies of human complex traits.* Bioinformatics, 2006.
224. Karchin, R., et al., *LS-SNP: large-scale annotation of coding non-synonymous SNPs based on multiple information sources.* Bioinformatics, 2005. **21**(12): p. 2814-20.
225. Trinklein, N.D., et al., *Identification and functional analysis of human transcriptional promoters.* Genome Res, 2003. **13**(2): p. 308-12.
226. Siepel, A. and D. Haussler, *Combining phylogenetic and hidden Markov models in biosequence analysis.* J Comput Biol, 2004. **11**(2-3): p. 413-28.
227. Blanchette, M., et al., *Aligning multiple genomic sequences with the threaded blockset aligner.* Genome Res, 2004. **14**(4): p. 708-15.
228. TRANSFAC, *TRANSFAC v6.0.*
229. Boeckmann, B., et al., *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003.* Nucleic Acids Res, 2003. **31**(1): p. 365-70.
230. Pritchard, J.K., M. Stephens, and P. Donnelly, *Inference of population structure using multilocus genotype data.* Genetics, 2000. **155**(2): p. 945-59.
231. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.
232. Hung, C.C., et al., *Dosage recommendation of phenytoin for patients with epilepsy with different CYP2C9/CYP2C19 polymorphisms.* Ther Drug Monit, 2004. **26**(5): p. 534-40.
233. Tsai, F.J., et al., *Polymorphisms for interleukin 1 beta exon 5 and interleukin 1 receptor antagonist in Taiwanese children with febrile convulsions.* Arch Pediatr Adolesc Med, 2002. **156**(6): p. 545-8.
234. Tan, N.C., et al., *Is variation in the GABA(B) receptor 1 gene associated with temporal lobe epilepsy?* Epilepsia, 2005. **46**(5): p. 778-80.
235. Yinan, M., et al., *Polymorphisms of casein kinase I gamma 2 gene associated with simple febrile seizures in Chinese Han population.* Neurosci Lett, 2004. **368**(1): p. 2-6.
236. Sander, T., et al., *Genetic variation of the human mu-opioid receptor and susceptibility to idiopathic absence epilepsy.* Epilepsy Res, 2000. **39**(1): p. 57-61.

237. Steinlein, O., et al., *Possible association of a silent polymorphism in the neuronal nicotinic acetylcholine receptor subunit alpha4 with common idiopathic generalized epilepsies.* Am J Med Genet, 1997. **74**(4): p. 445-9.
238. Chioza, B., et al., *Failure to replicate association between the gene for the neuronal nicotinic acetylcholine receptor alpha 4 subunit (CHRNA4) and IGE.* Am J Med Genet, 2000. **96**(6): p. 814-6.
239. Vijai, J., et al., *Genetic association analysis of KCNQ3 and juvenile myoclonic epilepsy in a South Indian population.* Hum Genet, 2003. **113**(5): p. 461-3.
240. Haug, K., et al., *No evidence for association between the KCNQ3 gene and susceptibility to idiopathic generalized epilepsy.* Epilepsy Res, 2000. **42**(1): p. 57-62.
241. Buono, R.J., et al., *Association between variation in the human KCNJ10 potassium ion channel gene and seizure susceptibility.* Epilepsy Res, 2004. **58**(2-3): p. 175-83.
242. Lenzen, K.P., et al., *Supportive evidence for an allelic association of the human KCNJ10 potassium channel gene with idiopathic generalized epilepsy.* Epilepsy Res, 2005. **63**(2-3): p. 113-8.
243. Sander, T., et al., *Variation of the genes encoding the human glutamate EAAT2, serotonin and dopamine transporters and Susceptibility to idiopathic generalized epilepsy.* Epilepsy Res, 2000. **41**(1): p. 75-81.
244. Chen, Y., et al., *Association between genetic variation of CACNA1H and childhood absence epilepsy.* Ann Neurol, 2003. **54**(2): p. 239-43.
245. Gu, W., et al., *Genotypic association of exonic LGI4 polymorphisms and childhood absence epilepsy.* Neurogenetics, 2004. **5**(1): p. 41-4.
246. Lenzen, K.P., et al., *Association analysis of malic enzyme 2 gene polymorphisms with idiopathic generalized epilepsy.* Epilepsia, 2005. **46**(10): p. 1637-41.
247. Mas, C., et al., *Association of the connexin36 gene with juvenile myoclonic epilepsy.* J Med Genet, 2004. **41**(7): p. e93.
248. Samochowiec, J., et al., *Association analysis between a Cys23Ser substitution polymorphism of the human 5-HT2c receptor gene and neuronal hyperexcitability.* Am J Med Genet, 1999. **88**(2): p. 126-30.
249. Haug, K., et al., *Mutation screening of the chromosome 8q24.3-human activity-regulated cytoskeleton-associated gene (ARC) in idiopathic generalized epilepsy.* Mol Cell Probes, 2000. **14**(4): p. 255-60.
250. Sander, T., et al., *Association analysis of exonic variants of the gene encoding the GABAB receptor and idiopathic generalized epilepsy.* Am J Med Genet, 1999. **88**(4): p. 305-10.
251. Lu, J., et al., *The gene encoding GABBR1 is not associated with childhood absence epilepsy in the Chinese Han population.* Neurosci Lett, 2003. **343**(3): p. 151-4.
252. Lenzen, K.P., et al., *Association analysis of the Arg220His variation of the human gene encoding the GABA delta subunit with idiopathic generalized epilepsy.* Epilepsy Res, 2005. **65**(1-2): p. 53-7.
253. Sobetzko, D., T. Sander, and C.M. Becker, *Genetic variation of the human glycine receptor subunit genes GLRA3 and GLRB and susceptibility to idiopathic generalized epilepsies.* Am J Med Genet, 2001. **105**(6): p. 534-8.
254. Goodwin, H., et al., *No association found between polymorphisms in genes encoding mGluR7 and mGluR8 and idiopathic generalised epilepsy in a case control study.* Epilepsy Res, 2000. **39**(1): p. 27-31.

255. Izzi, C., et al., *Candidate gene analysis of the human metabotropic glutamate receptor type 4 (GRM4) in patients with juvenile myoclonic epilepsy.* Am J Med Genet B Neuropsychiatr Genet, 2003. **123**(1): p. 59-63.
256. Steinlein, O.K., et al., *The voltage gated potassium channel KCNQ2 and idiopathic generalized epilepsy.* Neuroreport, 1999. **10**(6): p. 1163-6.
257. Chou, I.C., et al., *The voltage-gated potassium channel KCNQ2 in Taiwanese children with febrile convulsions.* Neuroreport, 2002. **13**(15): p. 1971-3.
258. Sander, T., et al., *Length variation of a polyglutamine array in the gene encoding a small-conductance, calcium-activated potassium channel (hKCa3) and susceptibility to idiopathic generalized epilepsy.* Epilepsy Res, 1999. **33**(2-3): p. 227-33.
259. Kananura, C., et al., *Tandem pore domain K(+)-channel TASK-3 (KCNK9) and idiopathic absence epilepsies.* Am J Med Genet, 2002. **114**(2): p. 227-9.
260. Haug, K., et al., *The voltage-gated sodium channel gene SCN2A and idiopathic generalized epilepsy.* Epilepsy Res, 2001. **47**(3): p. 243-6.
261. Haug, K., et al., *Association analysis between a regulatory-promoter polymorphism of the human monoamine oxidase A gene and idiopathic generalized epilepsy.* Epilepsy Res, 2000. **39**(2): p. 127-32.
262. Sander, T., et al., *Association analysis of a regulatory promoter polymorphism of the PAX-6 gene with idiopathic generalized epilepsy.* Epilepsy Res, 1999. **36**(1): p. 61-7.
263. Tsai, F.J., et al., *Interleukin-4 intron 3 polymorphism is not related to susceptibility to febrile seizures.* Pediatr Neurol, 2002. **27**(4): p. 271-4.
264. Sills, G.J., et al., *Lack of association between the C3435T polymorphism in the human multidrug resistance (MDR1) gene and response to antiepileptic drug treatment.* Epilepsia, 2005. **46**(5): p. 643-7.
265. Sporis, D., et al., *Association of refractory complex partial seizures with a polymorphism of ApoE genotype.* J Cell Mol Med, 2005. **9**(3): p. 698-703.