UNIVERSITY OF LIVERPOOL

Biomarkers for treatment outcome in newly diagnosed epilepsy

Thesis submitted in accordance with the requirements of the University of Liverpool for the degree of Doctor in Philosophy by

Pauls Auce

September 2016

Table of contents

Abbreviations

ABCB1	ATP-Binding Cassette Subfamily B Member 1
ABCC1	ATP Binding Cassette Subfamily C Member 1
ABCC2	ATP Binding Cassette Subfamily C Member 2
AED	Antiepileptic Drug
Australia	TheRoyal Melbourne Hospital / The University of Melbourne cohort
CAROLE	A multicenter, prospective, and observational study (CAROLE; i.e., Coordination Active du Réseau Observatoire Longitudinal de l'Epilepsie) of patients with newly diagnosed unprovoked seizures
CBZ	Carbamazepine
CI	Confidence Interval
CLCN-2	Chloride Voltage-Gated Channel 2
CPS	Complex Partial Seizure
CT	Computerised Tomography
CYP2C9	Cytochrome P450 Family 2 Subfamily C Member 9
EEG	Electroencephalography
EKUT	Eberhard-Karls-Universität Tübingen
EPHX1	Epoxide Hydrolase 1
EpiPGX	Epilepsy Pharmacogenomics: delivering biomarkers for clinical use
GABA	Gamma-Aminobutyric Acid
Glasgow	Epilepsy Unit, Western Infirmary / University of Glasgow cohort
GWAS	Genome-Wide Association Study
ILAE	International league against epilepsy
IQR	Interquartile Range
JME	Juvenile Myoclonic Epilepsy
LCSB	Luxembourg Centre for Systems Biomedicine, University of Luxembourg
MAF	Minor allele frequency
MESS	Multicentre Study of Early Epilepsy and Single Seizures
MRI	Magnetic Resonance Imaging
NGPSE	National General Practice Survey of Epilepsy
NICE	National Institute for Health and Care Excellence
P-gp	P-glycoprotein
RCSI	Royal College of Surgeons in Ireland

RCT	Randomized Clinical Trial
rsid	Reference SNP cluster ID
SANAD	Standard Versus New Antiepileptic Drug trial
SCN1A	Sodium Voltage-Gated Channel Alpha Subunit 1
SCN2A	Sodium Voltage-Gated Channel Alpha Subunit 2
SCN3A	Sodium Voltage-Gated Channel Alpha Subunit 3
SCN8A	Sodium Voltage-Gated Channel Alpha Subunit 8
SGTC	Secondarily Generalized Tonic-Clonic seizure
SMEI	Severe Myoclonic Epilepsy of Infancy
SNP	Single Nucleotide Polymorphism
UCL	University College London cohort
ULB	Hôpital Erasme, Université Libre de Bruxelles cohort
ULIV	University of Liverpool cohort
VA	Veterans Affairs Healthcare System
VPA	Valproate
WP2	EpiPGX Work package 2

Abstract

Introduction and aims

Epilepsy is a common neurological condition and around 25% of patients are resistant to treatment with currently available drugs (Brodie et al., 2012). Currently there is only a limited ability to predict treatment outcome and no genome based biomarkers for treatment efficacy. The main aim for this thesis was to investigate clinical and genome based biomarkers for treatment response in newly diagnosed epilepsy as well as explore methodological aspects related to the assembly of a large scale international research cohort.

Methods

An EU-funded project entitled “Epilepsy Pharmacogenomics: delivering biomarkers for clinical use (EpiPGX)” was undertaken by a pan-European research consortium to explore genome-based biomarkers that could be used to individualize treatment of epilepsy. University of Liverpool led work on newly diagnosed epilepsy. Work presented in this thesis is solely based on this project. Cases with newly diagnosed epilepsy were either de-novo phenotyped or data was transferred from existing clinical databases. Analysis of clinical covariates using logistic and Cox regression, and a subsequent GWAS were performed. Methodological and data transfer quality aspects were assessed separately using descriptive statistics and Cohen's kappa and Lin’s coefficients.

Results and Conclusion

The following clinical factors were significantly associated with twelve month remission after application of first well tolerated antiepileptic drug: age at diagnosis, abnormal neurological examination, GTCs-only, epilepsy type, number of seizures before the treatment, MRI and EEG results. Heterogeneity of outcomes between cohorts, effect of mode of cases ascertainment was also demonstrated. Data quality assessment showed that simple variables can be robustly transferred between data bases whereas more complicated variables have a potential for introduction of bias. A GWAS was carried out on newly diagnosed cases with focal epilepsy and failed to identify any SNPs significantly associated with treatment outcome.

Clinical factors associated with treatment outcome potentially can be useful in daily clinical practice when assessing patients with newly diagnosed epilepsy. Large scale multi-centre studies utilizing historical retrospective data are possible but prospective recruitment should be preferred. Sound methodology and quality assurance methods should be applied in future epilepsy pharmacogenetic research particularly involving large multi-centre cohorts.

Declaration

I, Pauls Auce, confirm that the work presented in this thesis is my own. It is based on research cohort assembled by EpiPGX consortium and members of staff employed by it. Dr Sarah Ray Langley carried out phenotype derivation as a part of the EpiPGX project. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Acknowledgments

I would like to express my heartfelt gratitude to my academic supervisor Dr Graeme Sills as well as Dr Andrea Jorgensen for help with statistical aspects of my work.

I would also like to thank all my clinical supervisors professors Tony Marson, Martin Brodie and John Paul Leach for their support in my development both in Liverpool and Glasgow.

I would also like to acknowledge and express my gratitude to all members of EpiPGX consortium and in particular work package two. Particularly I would like to thank Drs Ben Francis and Sarah Ray Langley for their work in project which I relied on.

Oral and poster presentationss

Auce P., Francis B., Langley SR., Jorgensen A., Marson AG., Sills GJ., for the EpiPGX consortium. (May 2017) GWAS for early remission in newly diagnosed focal epilepsy. The Association of British Neurologists Annual Conference, Liverpool, UK. (accepted for poster presentation)

Auce P., Marson AG., Sills GJ., on behalf of the EpiPGX consortium. (September 2015) Biomarkers of treatment response in newly diagnosed epilepsy - the EpiPGX experience. 8th Baltic Congress of Neurology 2015 oral presentation and abstract, Riga, Latvia.

Auce P. (May 2015) Clinical and genomic biomarkers of treatment response in newly diagnosed epilepsy. University of Liverpool poster day.

Auce P., Francis B., Jorgensen A., Brodie MJ., Marson AG., G.Sills GJ., (October 2015) Genome-based biomarkers of early treatment response in newly-diagnosed epilepsy. 6th Eilat International Educational Course: Pharmacological Treatment of Epilepsy, Jerusalem, Israel.

1.0. Introduction

1.1.0. Epilepsy overview

Epilepsy is characterized by the recurrence of unprovoked seizures caused by excessive electrical discharges generated by hyper – synchronous and hyperexcitable neuronal aggregates or networks. Current clinical definition requires at least two unprovoked seizures 24 hours apart or a single seizure and risk of recurrence at least 60%, or diagnosis of a specific epilepsy syndrome (Fisher et al., 2014).

Prevalence of active epilepsy in Europe has been estimated to range between 3.3 – 7.8 per 1000 inhabitants (Forsgren et al., 2005). It not only has a significant social and health effect on the affected individual, but also has a significant economic burden to society costing around 13.8 billion euros per year for the whole Europe (Olesen et al., 2012). Seizure control is by far the main aim of epilepsy treatment. At the moment, modern medical science has no means of achieving full seizure remission in all patients with epilepsy. Already back in 1881 in his book "*Epilepsy and other chronic convulsive diseases: their causes, symptoms & treatment*" Gowers described a relatively good prognosis for epilepsy with around 70% of patients experiencing "arrest" of their seizures when treated with bromides (Gowers, 1885). In a later hospital-based study comprised from mixed patient population reported in 1965 by *Rodin* and co-authors two year seizure freedom were observed only in 32.2 % (n=29) and a five year seizure freedom in 16.6 % (n=15) of their patients. When authors compared their results with earlier studies they concluded that available newer anticonvulsants, have failed to improve the long term seizure control (Rodin et al., 1965). This difference might be explained by differences in the case ascertainment. A later study from United Kingdom recruiting patients with a definite epilepsy from general practice between 1984 and 1987 showed a five year terminal remission on 9 years' follow up in the 54 % of subjects (Cockerell et al., 1995). A more recent study of newly diagnosed cases demonstrated that 68 % patients were seizure free at their last visit and 52 % had ten year continuous

remission by the end of the study (Brodie et al., 2012). In conclusion treatment results for newly diagnosed epilepsy have not changed much with the time despite increasing numbers of AED (Antiepileptic Drugs) available over the last 20 years (Schmidt and Sillanpää, 2012, Shorvon and Luciano, 2007).

Importance of achieving full seizure freedom should not be underestimated. Seizure freedom not only reduces social stigma and lifts specific restrictions on daily activities, but also lowers mortality (Ridsdale et al., 2011, Salanova et al., 2002). Efficacy of treatment is just one side of the coin; there is also a tolerability aspect to the treatment. A survey looking at the impact of the disease from the patient's perspective, ranked by patients themselves, demonstrated that fewer side effects was one of their top priorities (Fisher et al., 2000). Currently ability to reliably predict both the efficacy and occurrence of most side effects is limited. There are great hopes that better understanding of genetics of epilepsy and drug response will provide us with such tools in future as well as expand our understanding of disease process itself thus potentially aiding in the quest for new treatment options.

Epilepsy is a highly heterogeneous disorder; with a complex syndrome classification relying on variety of factors like seizure type, age at onset, imaging and EEG results. With the evolution of genome based technology, genetic factors are due to play a more prominent role in the classification (Thomas and Berkovic, 2014). It has already been shown that common variants collectively might have a significant effect on phenotypic variation in epilepsy (Speed et al., 2014b).

AEDs at the moment are the mainstay of treatment, hence it is important to develop better understanding of factors related to treatment efficacy. Treatment outcome modelling in newly diagnosed epilepsy based on the Standard Versus New Antiepileptic Drug study (SANAD), utilising routine and recognised clinical factors, only have limited ability to predict outcomes (Bonnett et al., 2012, Bonnett et al., 2014a, Bonnett et al., 2014b). This fact opens the question if there are any additional factors determining treatment results? In newly diagnosed epilepsy, response to the initial treatment is a

predicting factor for development of further drug resistance and only a limited number of patients later develop drug resistance after initially responding (Kwan and Brodie, 2000, Brodie et al., 2012). This pattern could potentially indicate that treatment response in newly diagnosed epilepsy is an important part of the general epilepsy phenotype. More recently, data from a prospective genome – wide association study (GWAS) has shown that no common variant explains more than 4.4% variation of outcome suggesting that it is influenced by multiple and complex genetic factors and that biological pathways should be investigated (Speed et al., 2014a). Genome based biomarkers in the future will likely play a role in prognosticating treatment outcome in epilepsy, but it is unlikely it will be a single test, but rather integration of large individual genomic data encompassing relevant pathways with clinical predictors.

Pharmacogenetics as a term was first time coined by *Vogel* back in 1959. It *"is defined as the study of variability in drug responses attributed to hereditary factors in different populations*" (Roses, 2001). Response is not only efficacy of applied treatment, but also occurrence of any adverse effects.

At the moment, as pharmacogenetics in epilepsy is largely limited to preventative testing for a few serious cutaneous adverse reactions to carbamazepine (CBZ), there are no genetic markers predicting AED efficacy It has been reliably demonstrated that pre – treatment testing for HLA-B*1502 allele in an appropriate ethnic group is associated with significant reduction of Stevens–Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) (Chen et al., 2011). In the meantime it has also been shown that the introduction of genetic screening alone is insufficient to reduce rates of SJS and TEN as prescribers would substitute carbamazepine with other aromatic AEDs which do not require genetic testing. This emphasizes the important role of education and dissemination of information. (Chen et al., 2014b).

Utilization of genetic markers to predict treatment results is part of a wider concept of precision medicine. Precision medicine aims to tailor targeted treatment based on all factors that distinguish a given patient from others with a similar presentation (Jameson

and Longo, 2015). A first example of this approach applied to epilepsy is Migrating Partial Seizures of Infancy (MPSI) due to *KCNT1* mutation and its treatment with quinidine. MPSI has a heterogeneous genetic background, but mutations in *KCNT1* are the most common underlying known genetic cause; a case report has shown that quinidine, a partial agonist of KCNT1 but not an established anticonvulsant, could potentially be a valuable treatment option in those cases (Milligan et al., 2014, Bearden et al., 2014). After an initial case report and basic science exploration of concepts, further larger studies are required to substantiate initial reports.

1.2.0. Clinical features of newly diagnosed epilepsy

Epilepsy is a clinical diagnosis and relies less on para–clinical disciplines like imaging and neurophysiology. National Institute for Health and Care Excellence (NICE) has advised that patients should be seen by a specialist and diagnosis is based on a constellation consisting of a description of attacks and other symptoms present, while EEG and neuroimaging serve as supplementary methods (NICE, 2012). However, both neuroimaging and EEG play an important role in stratifying the risk of recurrence after the first seizure and in the case of relevant detected abnormalities might add enough supportive weight to allow a diagnosis of epilepsy after only one seizure (Fisher et al., 2014, Krumholz et al., 2015). Probably the most important role of neuroimaging (CT or MRI) is to elucidate underlying structural lesions. From the pharmacological treatment selection point of view results of neuroimaging have no direct effect on the selection of the anticonvulsant, but if there is an identified structural abnormality it might have to be approached separately, depending on the type. EEG has a role in defining epileptic syndrome and hence might influence selection of an anticonvulsant. Figure 1 summarizes clinical approach to patient presenting with first seizure.

Initial treatment is a monotherapy with a single AED, in case of focal onset seizures it is either carbamazepine, lamotrigine or levetiracetam, whereas in generalised epilepsy valproate or lamotrigine (NICE, 2012, Network, 2015).

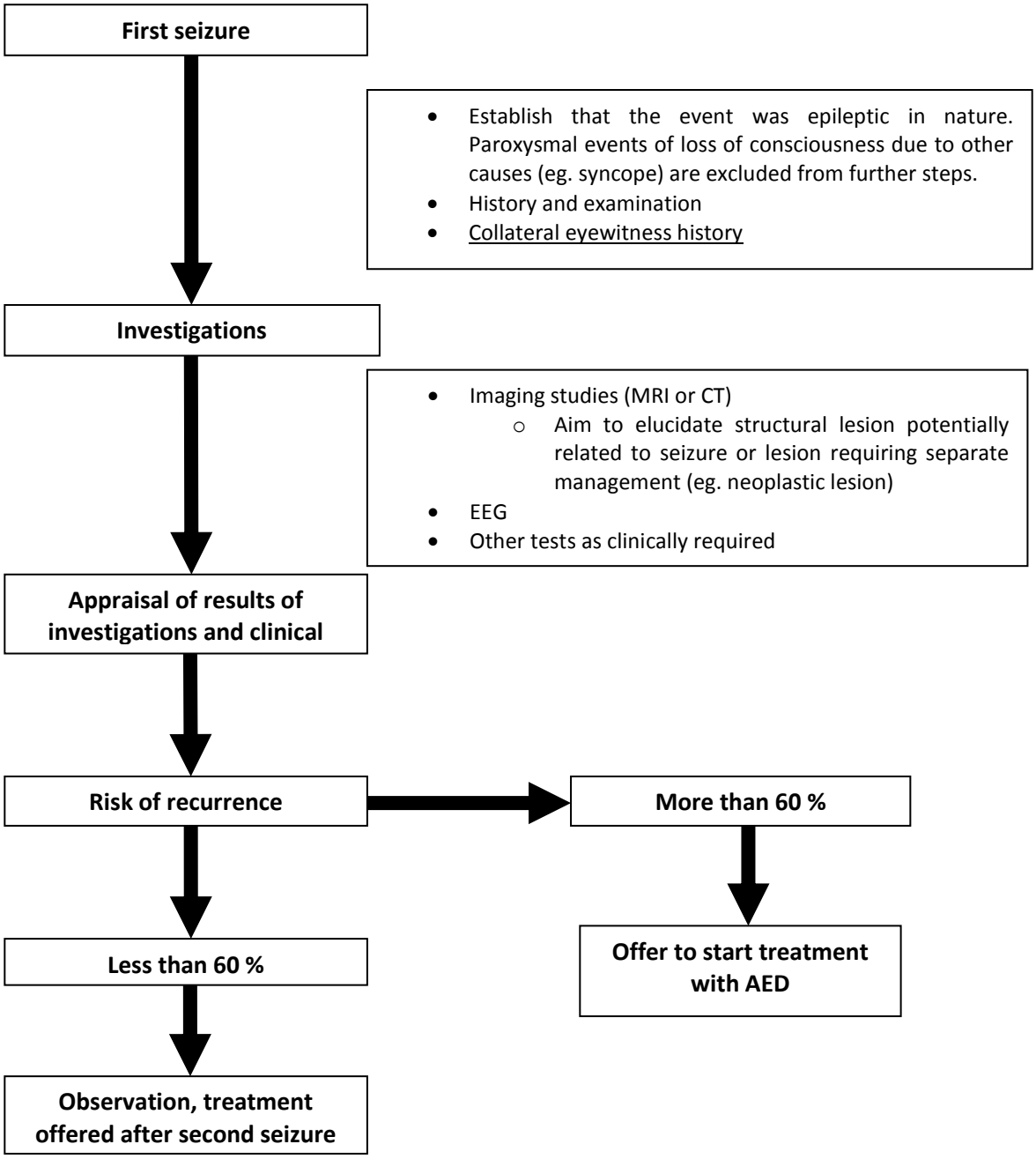


Figure 1. Flowchart illustrating clinical approach in case of first seizure.

1.3.0. Prognostic factors related to newly diagnosed epilepsy

Historically several clinical factors have been investigated for their links with treatment outcome in newly diagnosed epilepsy. The need to adjust analysis for clinical covariates in GWAS pharmacogenetic studies is neither clear nor well assessed. A recent large pharmacogenetic study elected to carry out two versions of analysis; in one the authors adjusted for clinical covariates whereas in the other, they did not (Speed et al., 2014a). The authors reasoned that if genetic factors influenced prognosis via clinical factors then adjustment might lead to failure to detect association, whereas lack of adjustment might reduce power. This study did not identify any associated (Single Nucleotide Polymorphisms) SNPs with GWAS significance, but all three loci with suggestive evidence had a similar effect size in both adjusted and unadjusted analysis. They suggested that association was therefore not mediated via clinical factors.

Based on data from previous analyses carried by *Bonnett et al* (Bonnett et al., 2012, Bonnett et al., 2014b) on the SANAD trial dataset, the following covariates have been selected for more in-depth description:

- Gender
- Febrile seizures
- Family history of epilepsy
- Imaging
- Treatment history
- Age at onset
- Time from first seizure to diagnosis
- Neurological examination
- Total number of pre – treatment seizures
- EEG
- Seizure type
- Epilepsy syndrome
- Aetiology
- Learning disabilities

1.3.1. Gender

Several epidemiological studies have shown a gender difference in the incidence of epilepsy; with males being affected more than females (Benn et al., 2008, Christensen et al., 2007, Lavados et al., 1992, Hauser et al., 1993). However, a systematic review and meta-analysis carried out by *Kotsopoulos et al* demonstrated showed no statistically significant difference in the incidence of epilepsy between genders, albeit with males having a slightly higher incidence (Kotsopoulos et al., 2002). Differences between gender in relation to epilepsy syndromes have been also reported. A study from Denmark based on the International League Against Epilepsy (ILAE) 1989 classification showed no overall difference in localization related epilepsies, but generalized epilepsy was observed more frequently in females (Christensen et al., 2005). The most frequent subtype of the localization related epilepsy was cryptogenic temporal lobe epilepsy which was more often observed in females, but males suffered more from the symptomatic localization related epilepsy. Juvenile myoclonic and juvenile absence epilepsy were more common in females.

Newly diagnosed epilepsy in general has a good prognosis with majority of patients experiencing prolonged periods of remission (Cockerell et al., 1995, Collaborative Group for the Study of, 1992). A landmark longitudinal study from Rochester, Minnesota utilizing a record linkage system from Mayo Clinic included 618 local residents of the county diagnosed with epilepsy between 1935 and 1974. *Annegers et al* observed that newly diagnosed epilepsy had a better prognosis for long-term remission than expected. One of the covariates assessed by the study was gender; females had slightly better remission rates than males initially, but in a long term differences were minimal (Annegers et al., 1979). Several later studies have failed to show any significant role of gender in relation to remission in newly diagnosed epilepsy (Beghi and Tognoni, 1988, Shorvon and Reynolds, 1982, MacDonald et al., 2000). Similarly, a prospective hospital based cohort of newly-diagnosed patients from Glasgow demonstrated no association of gender with the development of treatment resistant epilepsy (Hitiris et al., 2007). A study specifically investigating treatment outcomes for the newly-diagnosed patients in

the geriatric population also showed no significant difference in outcomes in relation to gender (Besocke et al., 2013). In the case of generalized epilepsies, a study from Glasgow showed that gender had no association with treatment outcome in non-paediatric newly-diagnosed idiopathic epilepsy (Mohanraj and Brodie, 2007).

Treatment outcome in newly-diagnosed epilepsy does not have a uniform and binary pattern. It has several different patterns: early or delayed remission, intermittent relapses and remissions, or continued persistence of seizures (Brodie et al., 2012). In a predominantly adult retrospective study from two tertiary epilepsy centres in Italy, specifically investigating prognostic factors for early versus late remission in patients with newly-diagnosed epilepsy, about 10 % of patients had a late remission and the only predictive factor for it in a multivariate analysis was more than 6 partial seizures before starting treatment; gender did not have any significance (Del Felice et al., 2010).

The SANAD trial is by far the largest multicentre randomised open-label controlled study in epilepsy. Arm A recruited mostly newly–diagnosed patients for whom the clinician deemed carbamazepine to be the first-line treatment. Thus an overall majority of patients had focal epilepsy (Marson et al., 2007a). A post hoc subgroup analysis based on the focal epilepsy arm (arm A) utilizing multivariable regression modelling identified several significant risk factors for time to treatment failure and 12 month remission. *Bonnett et al* demonstrated that females were more likely to fail due to unacceptable adverse effects, but no significant gender difference was seen for failure due to inadequate seizure control; men had a longer time to 12 month remission, but were marginally more likely to achieve a remission. (Bonnett et al., 2012). SANAD arm B was a separate un-blinded randomised multicentre study, which investigated effectiveness of valproate, lamotrigine and topiramate for generalised and unclassifiable epilepsy (Marson et al., 2007b). Contrary to similar analysis carried out in arm A, prognostic modelling showed that gender had no significant association with time to 12 and 24 month remission in generalised and unclassifiable epilepsy (Bonnett et al., 2014b). According to a further prognostic model involving the whole SANAD dataset, and looking

at the time to 12 month remission after failing the first anticonvulsant, men were again more likely to achieve 12 month remission (Bonnett et al., 2014a).

In summary, data from the literature are not straight-forward and that may be due to variety of reasons, particularly reporting bias and heterogeneity of disease and population studied. Gender likely has a small role, at best, in determining treatment outcome in newly diagnosed epilepsy.

1.3.2. Age of onset

In the developed world incidence of epilepsy has two peaks, one in early life and the other one in the elderly population; there are also some suggestions of a change in temporal trends over the time with reduction of the incidence in childhood and a rise of epilepsy numbers in the older age group (Neligan et al., 2012). Age groups differ by more characteristic epileptic syndromes and prevalent aetiology. Elderly patients with late onset epilepsy tend to have more symptomatic localization related and less idiopathic generalized epilepsy. Tumours and cerebrovascular disease are more common aetiologies of elderly onset epilepsy, whereas hippocampal sclerosis is observed less commonly than in both younger patients and elderly patients with early onset (Stefan et al., 2014). Hippocampal sclerosis (HS) has been shown to have worse treatment outcome compared to other causes of symptomatic focal epilepsy including post stroke epilepsy (Stephen et al., 2001).

Current literature regarding age of onset and treatment outcome in newly diagnosed epilepsy is contradictory. Some studies have observed no effect or a weak effect of age on treatment outcome (Shafer et al., 1988, Collaborative Group for the Study of, 1992, Cockerell et al., 1997, MacDonald et al., 2000). As already described earlier those studies have a variety of limitations, for example NGPSE included both treated and untreated cases whereas some patients in the major multicentre collaborative study from Italy (Collaborative Group for the Study of, 1992) were started on treatment after a first seizure if clinician felt treatment was required. Several hospital based cohorts have failed to demonstrate associations between age and treatment outcome in newly

diagnosed epilepsy (Elwes et al., 1984, Gasparini et al., 2013, Zhang et al., 2013, Shorvon and Reynolds, 1982, Su et al., 2013). These studies are heterogeneous by both population and their objectives. Furthermore, they have rather small sample size and they are not all prospective. Additionally, these studies have a wide time frame when they were conducted, thus anticonvulsant selection and treatment strategies might differ.

It has been suggested by Glasgow group that both elderly and young people have a more favourable response (Mohanraj and Brodie, 2006). Moreover in the case of newly diagnosed idiopathic generalized epilepsy in a non-paediatric setting, age has no influence on treatment outcome (Mohanraj and Brodie, 2007). Post hoc modelling based on SANAD study has suggested that older age has in general better prognosis in focal epilepsy (Bonnett et al., 2012) although this relationship is rather a complex. For time to 12 month remission in focal epilepsy, age showed a U-shaped relationship, with younger and older patients tending to respond more favourable. For time to failure, older patients had a higher chance to fail due to unacceptable adverse events and less due to inadequate seizure control. Younger patients had higher chance for treatment failure due to lack of effect and less due to poor tolerability (Bonnett et al., 2012). In a similar analysis based on SANAD arm B, where valproate was the standard of treatment, age had no predictive role either for time to remission or treatment failure (Bonnett et al., 2014b). The elderly also had better prospects for remission on a second AED after failure of the first drug (Bonnett et al., 2014a).

Observations from studies conducted specifically in elderly population also have shown in general good prognosis (Stephen et al., 2006, Besocke et al., 2013). Stephen et al demonstrated in general favourable outcome in a hospital based study with 62% of elderly patients becoming seizure free on a first drug and complete seizure freedom obtained in 79% of patients (Stephen et al., 2006). Similarly favourable outcome results were reported in a different study from Argentina, although additionally they suggested

that elderly patients might be more prone to medication adverse effects (Besocke et al., 2013).

A complex picture can be observed in the case of paediatric epilepsy where studies have shown different effects of age. These range from a significant association of younger age with a worse course of epilepsy (Geerts et al., 2010), to no difference if age of onset is above or below 6 years in regards of terminal remission of epilepsy (Sillanpää et al., 2014) or onset age 5 to 9 being significantly associated with 2 year remission (Berg et al., 2001).

Studies in newly diagnosed epilepsy are different from ones looking at established epilepsy. One of the differences is duration of disease at the moment of inclusion in the study and history of exposure to anticonvulsants. Age at onset of the condition is one of the factors from natural history that influences length of the treatment time with younger patients potentially having longer duration of epilepsy.

A study consisting of a mixture of newly diagnosed and established epilepsy cases showed that duration of epilepsy had a negative effect on response to newly administered AEDs (Schiller and Najjar, 2008). As already shown earlier, epilepsy has a complex course and often it involves mixtures of remissions and relapses even in the case of intractable disease. Duration of intractability has been shown to be negatively linked to chances of remission in this group, but had no association with potential for relapse in previously drug resistant patients that had entered remission (Callaghan et al., 2011). A similar hospital-based retrospective study involving patients with treatment resistant epilepsy showed that age of onset is not a significant predictor for treatment outcome (Selwa et al., 2003). In this study patients were young and from a very specific patient population who had been evaluated for epilepsy surgery, but found out not to be candidates for it. Thus, they were, by definition, treatment resistant but otherwise relatively young and healthy. In another hospital based surgical cohort study that assessed time to failure of the second anticonvulsant, younger age was shown to be strongly associated with longer latency period and past history of remission; age at

onset was also an independent factor associated with hippocampal atrophy, thus demonstrating the complex interaction between covariates (Berg et al., 2003). Research based on different cohorts of patients with mesial temporal lobe sclerosis also has controversial results; several studies have suggested early disease onset is linked to worse outcome (Kim et al., 1999, Varoglu et al., 2009, Sànchez et al., 2014) but at the same time, it has also been reported as associated with good seizure outcome (Kumlien et al., 2002).

As a covariate, age at onset has broad interactions with other factors such as epilepsy syndrome, seizure type, treatment duration, tolerability of drugs (comorbid conditions and their treatments), as well as specific disease aetiologies. Furthermore, age at onset can either be measured at the onset of first or second seizure or at the formal diagnosis. For example, a higher pre –treatment seizure frequency could potentially lead to earlier diagnosis and treatment, whereas some patients with rare seizures and low pre – treatment density might be diagnosed later in life. In summary, older age at onset of epilepsy confers better prognosis for treatment responsiveness.

1.3.3. Time from first seizure to diagnosis

Time from first seizure to diagnosis is influenced by several factors, some of which are related to epilepsy while others to health and social care as well as to the individual. For example, diagnostic delay might be due to over-reliance on investigations rather than accepting that epilepsy is a clinical diagnosis. A study exploring diagnostic delay for early onset childhood epilepsy showed that some cases were delayed due to over-reliance on the EEG (Berg et al., 2014). Thus, one can speculate that an abnormal EEG in childhood might lead to earlier treatment and shorter time between first seizure and treatment, while a normal EEG could potentially lead to a more 'watching and waiting' approach. There is also some overlap of clinical factors predicting earlier second seizure thus potentially a shorter time to diagnosis. Some of them are also potentially associated with treatment outcome once treatment has been started. The Multicentre Study of Early Epilepsy and Single Seizures (MESS) demonstrated that abnormal EEG,

neurological disorder and a higher number of seizures before randomisation were all significant factors predicting recurrent seizures (Kim et al., 2006a). In general, time from first seizure till diagnosis of epilepsy is very variable; it can be a single day in some cases, up to several decades. CAROLE study showed that roughly 59 % of patients will be diagnosed within the first year, 12 % in the second year and as many as 15 % more than five years after the initial seizure (Jallon et al., 2001).

Several studies have found no association between the time from the first seizure to diagnosis and either remission or treatment outcome in newly diagnosed epilepsy (MacDonald et al., 2000, Collaborative Group for the Study of, 1992, Park et al., 2014, Sillanpää, 1993). In a hospital based cohort of newly diagnosed epilepsy, duration of pre – treatment epilepsy was associated with a development of treatment resistance in the univariate analysis, but lost its significance in multivariate analysis (Hitiris et al., 2007). Time from the first seizure to diagnosis is closely related to time from first seizure to randomization, which has been assessed in post hoc analysis based on SANAD trial data. Results of this analysis showed that it has a role in the case of focal epilepsy. A longer duration between the first seizure and randomisation increases chances of remission but more seizures before treatment reduces it (Bonnett et al., 2012).

1.3.4. Neurological sequelae / abnormal exam

Neurological examination is performed in epilepsy with a goal to identify the sequeleae of previous neurological insults, or an on-going CNS disorder, for example brain tumour. However, damage of the central nervous system does not always produce focal neurological signs, and focal neurological signs do not always have a clear explanation. In epilepsy, a significant proportion of patients (10 – 15 %) has abnormal neurological examination (Okuma and Kumashiro, 1981, Heller et al., 1995, Collaborative Group for the Study of, 1992). It has also been demonstrated that neurological examination does not always pick up serious causes of seizures. A study from Finland has shown positive correlation between examination and changes on the CT, but not all patients with tumours had focal signs (Reinikainen et al., 1987). In

epilepsy research, there is a risk that hospital-based cohorts over-represent the proportion of patients with focal neurological signs or learning disabilities, particularly in historical studies where a significant proportion of patients was institutionalised. Furthermore, in cases of abnormal neurological examination one should bear in mind the link between it and underlying CNS disorder. Several studies have reported neurological symptoms as a negative predictor for remission in epilepsy (Okuma and Kumashiro, 1981, Sillanpää, 1993, Elwes et al., 1984) or in some cases as a combination of neurological and psychiatric handicap (Shorvon and Reynolds, 1982). In contrast, there are also several studies failing to demonstrate a significant association in this respect (Turnbull et al., 1985, Collaborative Group for the Study of, 1992, Lossius et al., 1999). It has been shown that neurological examination is not a prognostic factor for early- versus delayed-onset remission (Del Felice et al., 2010) or the development of treatment resistant epilepsy (Hitiris et al., 2007). Nevertheless, a covariate combining both learning disability and neurological deficit has been shown to be a significant predictive factor for time to 12 month remission for both focal and generalised/unclassified epilepsy (Bonnett et al., 2012, Bonnett et al., 2014b). In summary, neurological insult as a predictive factor probably reflects underlying brain pathology and singles out symptomatic from cryptogenic epilepsy.

1.3.5. Seizure type

Pre-treatment seizure type as a clinical predictor for treatment outcome has been assessed by several studies. Seizure type can be broadly classified into generalised and partial (focal). Both generalised and partial seizures are then further sub classified into smaller categories. Seizure classification according to ILAE 1981 Classification has been summarised in table 1 (1981).

Table 1. Summary of ILAE 1981 classification of seizures (Bancaud et al., 1981).

Seizure type	Notes
Partial seizures	
Simple partial seizure	Consciousness is not impaired
Complex partial seizure	Consciousness is impaired, sometimes might start as simple partial seizure
Partial seizure evolving into secondary generalised seizure	Can evolve from both simple and complex partial seizure
Generalised (convulsive and non-convulsive)	
Absence seizures	Might only have impairment of consciousness, or with additional mild tonic, clonic or autonomic components. Ictal EEG usually regular bilateral, symmetrical 3 Hz spike and wave complexes.
Atypical absence	Changes in tone are more pronounced, onset might not be as abrupt as in typical absence seizure. Ictal EEG is more heterogeneous.
Myoclonic seizures	Single or multiple myoclonic jerks.
Clonic seizures	Clonic jerks without tonic component
Tonic seizures	Rigid violent muscular contractions involving torso or limbs
Tonic-clonic seizures	Grand mal seizures. Tonic phase followed by clonic. Can be primarily or secondarily generalised.
Atonic (astatic) seizures	Sudden diminution of muscle tone, might be fragmentary
Unclassified epileptic seizures	All seizures that cannot be classified

Partial seizures have been shown by several studies to be a negative predictor for remission (Annegers et al., 1979, Shorvon and Reynolds, 1982, Elwes et al., 1984, Turnbull et al., 1985). A combined large VA study including patients with symptomatic localization-related epilepsy demonstrated that in this patient group, those with

generalized tonic-clonic seizures alone have the best prognosis for remission, complex partial seizures alone the worst and with all others somewhere in-between (Mattson et al., 1996). Contrary to their earlier report, a later and more extended study from Rochester County, Minnesota showed that never having had a generalized tonic-clonic seizure is a positive predictor (Shafer et al., 1988). A mixed seizure type also has been associated with poorer treatment responsiveness (Zhang et al., 2013, Su et al., 2013, Beghi and Tognoni, 1988).

Post-hoc analysis done on the SANAD trial data showed that in the case of focal epilepsy overall treatment failure is more likely if the subject experiences simple or complex partial seizures alone compared to those with additional secondary generalised tonic-clonic seizures. This link is significant for failure due to adverse events but not due to inadequate seizure control (Bonnett et al., 2012). Similar analysis done on patients with generalised or unclassified epilepsy showed that the presence of myoclonic or absence seizures is a negative predictor for 12 or 24 month remission, and that patients with absence seizures or a mixed seizure type (including absence or myoclonic seizures with generalized tonic-clonic seizures) are more likely to have treatment failure due to inadequate seizure control (Bonnett et al., 2014b). A further analysis of the SANAD data from patients who failed their first AED demonstrated that occurrence of tonic-clonic seizures while on treatment was a negative predictive factor for a subsequent remission. Likewise, patients with focal seizures before starting treatment had less chance of remission than those with secondary generalized seizures (Bonnett et al., 2014a).

On the other hand, there are several studies covering different patient groups that have failed to demonstrate any significant association of treatment outcome with pre – treatment seizure type (Mohanraj and Brodie, 2007, Park et al., 2014, Gasparini et al., 2013, Berg et al., 2001, Besocke et al., 2013, MacDonald et al., 2000).

Seizure type as a covariate can influence and is influenced by other factors like epilepsy syndrome and aetiology. Some seizure types are more intrusive and could prompt patients to seek medical attention earlier hence leading to earlier diagnosis and

lower pre – treatment seizure frequency. In short, expressing generalised tonic-clonic seizures only or as the predominant seizure type probably has a positive association with treatment outcome.

1.3.6. Epilepsy syndrome

Certain epilepsy syndromes are more prominent or occur almost exclusively during certain periods of life. Some childhood epilepsy syndromes like benign rolandic or occipital epilepsy has invariably very favourable prognosis with almost all patients entering into remission, whereas Lennox – Gastaut and West syndromes have a poor prognosis (Berg et al., 2001, Sillanpää et al., 2014). Age has an effect on the incidence of the epilepsy syndrome, with generalised epilepsies being more common than partial during the first five years of life, a roughly equivalent incidence during the adolescent years, and partial epilepsy taking over after age of 24. Interestingly there are also differences by gender (Hauser et al., 1993).

Several studies have demonstrated a lack of significant association between epilepsy syndrome and treatment outcome (Sillanpää and Schmidt, 2009, Park et al., 2014, Collaborative Group for the Study of, 1992). In newly diagnosed focal epilepsy, the lobe of origin also has been shown to have no association with response to the first anticonvulsant (Bonnett et al., 2012). Furthermore, there are no significant differences between cryptogenic and symptomatic focal epilepsies in terms of treatment outcome, although there are differences based on aetiology (Mohanraj and Brodie, 2005a).

On the other hand, there are also several studies demonstrating a link between syndrome and treatment outcome, as outlined below. In a paediatric population, idiopathic epilepsy syndromes have higher remission rates (Berg et al., 2001, Sillanpää et al., 2014). Similarly, in adults, idiopathic generalised epilepsy has been reported to have higher response rates than cryptogenic or symptomatic generalised epilepsies (Mohanraj and Brodie, 2006). Juvenile myoclonic epilepsy makes up around 5 – 10 % of all epilepsy and is the most common form of idiopathic generalised epilepsy (Camfield et al., 2013). Reported remission rates vary between studies, ranging from 17 % to 88 %

(Kleveland and Engelsen, 1998, Panayiotopoulos et al., 1994, Camfield and Camfield, 2009, Penry et al., 1989, Geithner et al., 2012, Siren et al., 2002). This wide range is probably related to study design, the definition of remission and a relatively small sample size in many studies. It has also been demonstrated that different seizure types have different remission rates (Kleveland and Engelsen, 1998).

Classification of epilepsy syndromes has changed several times hence it might be hard to compare older studies with newer ones. Epilepsy syndrome itself is intrinsically linked to the seizure type, aetiology, age at onset, EEG findings and genetic factors. All of these covariates are probably associated with the treatment outcome in their own right. In general, certain distinctive childhood syndromes have a very strong predictive effect, whereas in adults this effect might be less pronounced.

1.3.7. Aetiology

Aetiology can be linked to other co – variates, such as age, results of investigations, neurological deficit and epilepsy syndrome. For example, in the case of age at onset, cerebrovascular and neurodegenerative disorders are more often observed in the elderly, whereas congenital causes are more often observed in the younger patient group (Hauser et al., 1993). Due to the evolution of diagnostic capabilities, knowledge about genetic causes of epilepsy, and improved classification of aetiology, comparison between older and newer studies is becoming complicated. Furthermore, aetiological categories are often lumped together and different studies also employ different diagnostic criteria and investigation protocols making further comparison difficult.

Neurological dysfunction identified at birth has been associated with poorer prognosis (Annegers et al., 1979), whereas absence of early brain damage is associated with more favourable outcomes (Shafer et al., 1988). Likewise, remote symptomatic or merged similar categories as a non-idiopathic aetiology in paediatric population has been shown to have a worse treatment prognosis (Berg et al., 2001, Geerts et al., 2010). In adults with newly diagnosed focal epilepsy, cerebral atrophy and cerebrovascular disease are associated with more remissions than post-traumatic epilepsy. Interestingly,

primary brain tumours, cortical malformations and vascular lesions do not have a particularly worse treatment outcome than other aetiologies (Mohanraj and Brodie, 2005a). However, studies in patients with established epilepsy due to hippocampal sclerosis have shown a poor treatment response (Semah et al., 1998, Stephen et al., 2001, Hui et al., 2007). Then again, several studies covering different populations with regard to age and status of diagnosis at the time of inclusion, and using mostly condensed and simplified aetiological classifications, have shown no association with treatment outcome (Besocke et al., 2013, Su et al., 2013, Sillanpää and Schmidt, 2009, Collaborative Group for the Study of, 1992, Zhang et al., 2013, MacDonald et al., 2000).

In summary, the relationship between aetiology and treatment outcome is complex and will change in future as our knowledge about causes of epilepsy expands. With the development of genetic and imaging technologies, aetiology might increasingly guide us in the selection of the correct treatment approach.

1.3.8. Total number of pre-treatment seizures

Research based on patients undergoing ambulatory and video-EEG monitoring has shown that significant numbers of patients are not aware of their seizures and under-report seizure numbers. This applies to both adults (Blum et al., 1996, Hoppe et al., 2007, Kerling et al., 2006, Tatum et al., 2001) and children, where parents have been shown to under-report seizure numbers (Akman et al., 2009).

Several studies have demonstrated that total number of pre – treatment seizures is associated with treatment outcome. The National General Practice Study of Epilepsy (NGPSE) showed that the strongest predictive factor for all types of remission is the number of seizures in the first 6 months after the index seizure (MacDonald et al., 2000). A similar association was also demonstrated for the number of seizures prior to the index seizure. This effect is not linear and those with ten or more seizures before the index seizure have a better chance of remission. A further aspect arising from the NGPSE was the concept of “seizure density”, which is closely related to seizure frequency.

Similar results have been demonstrated by several other observational studies of newly diagnosed epilepsy, showing that a higher number of seizures before treatment is associated with worse treatment outcome (Shorvon and Reynolds, 1982, Elwes et al., 1984, Beghi and Tognoni, 1988, Collaborative Group for the Study of, 1992, Di Mascio et al., 1986, Stephen et al., 2006, Kwan and Brodie, 2000). Higher pre-treatment seizure count has also been associated with relapse of seizures in newly diagnosed epilepsy (Su et al., 2013).

Prognostic modelling based on the SANAD data has shown that in focal epilepsy, a higher total number of seizures before treatment is associated with treatment failure and also is a negative prognostic factor for time to 12 month remission (Bonnett et al., 2012). Similarly, in generalized epilepsy, the total number of tonic-clonic seizures was negatively associated with time to 12 and 24 month remission, but not with treatment failure (Bonnett et al., 2014b). Despite these observations, several studies have suggested that there is no association between treatment outcomes and number of pre-treatment seizures (Gasparini et al., 2013, Zhang et al., 2013, Park et al., 2014, Turnbull et al., 1985, Hitiris et al., 2007).

Pre – treatment seizure count is influenced by several factors. In some idiopathic generalized epilepsy syndromes, for example juvenile myoclonic epilepsy, total seizure number can be very high due to either absence or myoclonic seizures or both. This could explain the absence of a linear relationship in the NGPSE study. This covariate is related to epilepsy syndrome, as well as to localization of the epileptogenic zone and dominant seizure type. In conclusion, total pre-treatment seizure number probably is a clinical covariate associated with a treatment outcome.

1.3.9. Imaging CT/MRI

Neuroimaging has an important role in the assessment of patients with seizures and also in the management of epilepsy. Magnetic resonance imaging (MRI) is the preferred method to elucidate structural causes of the disease (NICE, 2004). Computed tomography (CT) is inferior to cranial MRI. In a study from Australia involving 300

consecutive patients from the first seizure clinic, computerised tomography (CT) failed to identify roughly one half of tumours detected by MRI (7 out of 15) as well as detecting lesions in less than one half of the cases (12 out of 28) (King et al., 1998). This fact has implications when assembling retrospective patient cohorts where some patients might have had CT and others MRI. A further large study involving adults referred to the first seizure clinic for an assessment, detected potentially epileptogenic lesions in 23% of MRI scanned individuals (Hakami et al., 2013). Authors observed statistically significant differences in the frequency of lesions between patients who were diagnosed as having an epileptic seizure and those with non-epileptic events (28% vs. 8% $p < 0.001$). This emphasises the fact that newly diagnosed seizures, although closely related, are not the same as newly diagnosed epilepsy as some patients with non-epileptic events still had abnormal imaging.

In a paediatric study of newly diagnosed epilepsy, close to 13% of imaged cases had a significant abnormality (Berg et al., 2000). Due to differences in the aetiology at various ages those results are not directly comparable, but nevertheless they demonstrate the importance of imaging in newly diagnosed epilepsy.

As discussed previously, MRI produces better diagnostic yield than CT, but imaging protocols and reporting is of importance as well. MRI carried out according to a specific epilepsy protocol and reported by radiologists experienced in epilepsy field has a better diagnostic output (McBride et al., 1998, von Oertzen et al., 2002). This potentially adds an extra layer of variability in cohorts where imaging studies are assessed locally by a mixture of general and neuroradiologists utilizing a variety of imaging protocols.

The number of studies assessing imaging in relation to treatment outcomes in newly diagnosed epilepsy is limited. The first study including imaging was from King's College hospital, conducted by *Elwes et al*, which recruited patients in the period between 1974 and 1979. It combined learning disabilities, focal neurological signs and abnormal CT scan under a variable termed 'neurological handicap'. This was shown to be one of the factors linked to worse treatment prognosis (Elwes et al., 1984). A study assessing only

initial response after the first 6 months of treatment in newly diagnosed epilepsy demonstrated no differences between responders and non – responders for the presence of structural lesions on MRI (Park et al., 2014). Lack of association of imaging with treatment outcome in newly diagnosed epilepsy has been shown both in adults (Zhang et al., 2013, Su et al., 2013) and in the elderly (Stephen et al., 2006). It also had no influence on prediction of early or late remission (Del Felice et al., 2010).

In the SANAD focal epilepsy arm, abnormal imaging studies were reported in 26% of patients and in a further subgroup analysis, imaging results were included in a multivariable prognostic model for time to 12 month remission with a hazard ratio of 0.88 but a wide 95% confidence interval of 0.76 – 1.03 (Bonnett et al., 2012). This could probably be explained by a complex interaction with other covariates. In the unclassified and generalised epilepsy arm of SANAD 5% of patients had an abnormal scan, with imaging results considered but not included in final prognostic models for time to 12 month remission and time to treatment failure (Bonnett et al., 2014b). An older hospital based cohort consisting of established epilepsy showed no link between CT imaging and recurrence of seizures (Lossius et al., 1999).

The nature of brain lesions is usually assessed by one of the imaging methods, most commonly MRI. In newly diagnosed, localization-related symptomatic epilepsy, it has been shown that cerebral atrophy and cerebrovascular disease had a better treatment outcome, while traumatic brain injury had worse results than patients with other symptomatic epilepsies (Mohanraj and Brodie, 2005a).

The Glasgow group has demonstrated earlier that mesial temporal sclerosis has a worse treatment outcome prognosis in newly diagnosed epilepsy than other common causes of symptomatic focal epilepsy (Stephen et al., 2001). In contrast, a different study based in a tertiary hospital setting, with the majority of cases being established epilepsy, showed an influence of MRI detected lesions on the intractability of partial epilepsy. The authors observed that patients with MRI detected lesions had a significantly less frequent remission (25% versus 42% seizure free, $p < 0.001$). Furthermore, hippocampal

sclerosis, dual pathology and cortical dysgenesis had the poorest outcomes, while tumours, vascular malformations and old strokes fared best (Semah et al., 1998). These studies are clearly different, one utilized newly diagnosed cases while the other one focused on established epilepsy and thus they cannot be directly compared. Interestingly the relationship between hippocampal sclerosis (HS) on MRI and treatment outcomes in patients with already established epilepsy is also not simple or straightforward. Some studies have shown that HS is a predictor of poor prognosis for treatment outcome (Semah et al., 1998, Van Paesschen et al., 1997, Pittau et al., 2009). Poor outcomes nevertheless have not been confirmed by all studies, for example *Kim et al* showed that one quarter of patients with a mesial temporal sclerosis (MTS) achieved remission and 37% had an improvement in seizure control with a medical treatment (Kim et al., 1999). MTS has also been observed in patients with sporadic benign temporal lobe epilepsy illustrating the complexity of this relationship (Kobayashi et al., 2001, Labate et al., 2006). Hippocampal atrophy is a part of mesial temporal lobe sclerosis but, so far, not all studies have demonstrated its clear association with treatment outcome (Andrade-Valença et al., 2003, Mohanraj and Brodie, 2005a). Interestingly, in a volumetric MRI study the pattern for grey matter atrophy for treatment resistant and relapsing-remitting epilepsy was similar, but different from responders (Bilevicius et al., 2010). The pharmacoresistant group had a lower age of onset and higher initial seizure frequency. Age of onset had previously been demonstrated to be strongly associated with time to intractability, and together with a history of febrile seizures, it was also shown to be associated with hippocampal atrophy (Berg et al., 2003).

In summary, results of imaging studies are not clear-cut predictors of treatment outcome. There are varieties of potential explanations like heterogeneous patient populations plus the majority of studies are hospital-based and retrospective and there are differences in acquisition of imaging and in its interpretation. Furthermore, imaging as a predictor might not be fully independent, but might bear some relationship to aetiology of epilepsy. There is a need for a prospective study of new and sophisticated

methods of neuroimaging and treatment response in newly diagnosed epilepsy. Such a study has been proposed by a research group in Halifax (Pohlmann-Eden et al., 2013).

1.3.10. Electroencephalography

Electroencephalography (EEG) is a valuable tool for diagnosis and classification of epileptic syndrome and localization of the ictal onset zone in the case of surgery. Epileptiform changes on the EEG performed after new onset seizures are observed in 12 to 50 % of adults (Wirrell, 2010). There is a correlation between focal abnormalities on the EEG and neuroimaging though it is not absolute (Reinikainen et al., 1987). Some of the idiopathic generalised epilepsy syndromes have specific EEG changes, which influences treatment choices (Panayiotopoulos, 2005). In addition some seizure types are more likely to have abnormal EEG, like absences and myoclonic seizures (Carpay et al., 1997). Timing of EEG can also influence the chance of detecting epileptiform abnormalities, with earlier EEG after seizure having a higher likelihood of abnormal findings (King et al., 1998). A higher number of seizures during the preceding 12 months also increased the probability of capturing an abnormal EEG. (Sundaram et al., 1990) Those factors and their relationship with epileptic syndrome itself, where EEG often is an integral part, should be taken into account when assessing EEG as a treatment outcome predictor.

EEG as a treatment outcome predictor has been assessed by multiple studies; results so far have been controversial. Several studies covering a variety of different populations have shown no association (Elwes et al., 1984, Beghi and Tognoni, 1988, Stephen et al., 2006, Turnbull et al., 1985, Zhang et al., 2013, Mohanraj and Brodie, 2007). In contrast, several others have shown a link between EEG and treatment outcome (Su et al., 2013, Bonnett et al., 2014b, Bonnett et al., 2012, Shafer et al., 1988). A study done by Shafer et al based on Rochester project showed that absence of generalised epileptiform activity was associated with the higher probability to achieve five year remission (Shafer et al., 1988). Similarly in a prognostic modelling based on SANAD presence of epileptiform abnormality was a significant factor for time to

treatment failure but not for remission (Bonnett et al., 2012, Bonnett et al., 2014b). Su et al demonstrated that epileptiform EEG before onset of treatment is associated with relapse of seizures in patients with newly diagnosed epilepsy (Su et al., 2013). Interestingly epileptiform EEG during the treatment did not have such association.

In summary, it is not clear if epileptiform abnormalities on EEG are predictive of treatment outcome. It may well be that specific seizure type and epileptic syndrome are associated with certain EEG changes, which themselves have prognostic value rather than the EEG itself.

1.3.11. Febrile seizures

Febrile seizures are defined as seizures occurring in a febrile child aged 6 to 60 months without a previous history of afebrile seizures and in the absence of intracranial infection or metabolic disturbances (Steering Committee on Quality Improvement Management, 2008). The risk of development of afebrile seizures in later life is slightly increased and has been estimated to be between 2-7% (Chungath and Shorvon, 2008). Interestingly, a large population based cohort from Denmark showed that this risk is higher in those with a family history of epilepsy, cerebral palsy and low Apgar scores at 5 minutes (Vestergaard et al., 2007).

Association of febrile seizures with treatment outcomes in newly diagnosed epilepsy has not been investigated as extensively as other covariates. A study from Glasgow involving 780 patients with newly diagnosed epilepsy treated in the same hospital-based outpatient clinic demonstrated an association of febrile seizures with worse treatment outcome (Hitiris et al., 2007). Not surprisingly, similar results were obtained from another study which was based on the same clinical cohort of newly diagnosed adults from Glasgow, but specifically looking at treatment outcomes for idiopathic generalized epilepsy syndromes in a non-paediatric setting (Mohanraj and Brodie, 2007). A Dutch study assessing the course and outcome of childhood epilepsy showed similar results (Geerts et al., 2010). This study included 413 subjects with the mean age of onset of 5.5 years (median 5.1; range 1 month – 15.5 years) and in whom history of febrile seizures

was one of the predictors of intractability. The role of febrile seizures as a prognostic factor was also assessed in the post – hoc subgroup analysis based on the data from SANAD trial. Although it failed to reach statistical significance, it was included in the final prognostic model for time to 24 month remission in subjects with generalised and unclassified epilepsy (Bonnett et al., 2014b). Febrile seizures have also been shown to be associated with drug resistant mesial temporal lobe epilepsy, while MTS in the same study was more likely to be intractable (Pittau et al., 2009). This study was based in a hospital tertiary care setting, case recruitment used both prospective and retrospective methods, and patients were not newly diagnosed. *Berg et al,* in a large, multicentre, pre–surgical, hospital-based study, demonstrated that history of febrile seizures and age at onset had independent effects on finding hippocampal atrophy on MRI, and that all three factors were associated with time to intractability. Multivariate analysis showed that age at onset provided the largest explanatory value and the authors suggested that all three variables are inter–related (Berg et al., 2003). In the meantime, a hospital-based retrospective study from Italy investigating factors related to long-term remission in newly diagnosed, cryptogenic focal epilepsy showed that individual history of febrile seizure had no association with the outcome (Gasparini et al., 2013).

1.3.12. Family history of epilepsy

First degree relatives of a proband with epilepsy have an increased incidence of epilepsy; the risk is higher for idiopathic generalized epilepsy, but is also present in the case of focal epilepsy (Peljto et al., 2014). Inclusion of data on family history in epilepsy studies largely depends on self – reporting or occasionally from interviewing relatives. It has been shown that this information has a relatively good precision for siblings and offspring, but not for parents (Ottman et al., 2011).

Several studies have included family history in their assessment, but not all have reported results. A heterogeneous group of studies involving newly diagnosed epilepsy from both community and hospital-based clinics have shown no association with treatment outcome (Shafer et al., 1988, Su et al., 2013, Beghi and Tognoni, 1988,

Shorvon and Reynolds, 1982, Zhang et al., 2013), while a few studies have demonstrated a link with poorer prognosis (Elwes et al., 1984, Hitiris et al., 2007). Interestingly, family history showed no association with treatment outcome in newly diagnosed generalized epilepsy syndromes in an adult population (Mohanraj and Brodie, 2007). Similar results have been demonstrated in juvenile myoclonic epilepsy, the commonest syndrome of idiopathic generalized epilepsy (Gelisse et al., 2001). It also appears that family history has no significant influence on the occurrence of either early or late remission (Del Felice et al., 2010). In the case of newly diagnosed cryptogenic epilepsy, results of a retrospective hospital-based cohort have suggested association between positive family history and long term remission (Gasparini et al., 2013). In contrast, mesial temporal lobe epilepsy studies have failed to show any significant role of family history (Pittau et al., 2009, Sànchez et al., 2014, Varoglu et al., 2009). The presence or absence of a first degree relative with epilepsy was also considered in a post – hoc subgroup analysis of the SANAD study data and it was included in a prognostic model for time to 12 month remission in generalised and unclassified epilepsy (Bonnett et al., 2012, Bonnett et al., 2014b).It was demonstrated that presence of the first degree relative with epilepsy is negatively associated with the time to 12 month remission (HR = 0.69; 95% CI 0.54 to 0.89).

The majority of participants in studies exploring family history as a predictor of prognosis are adults with adult-onset epilepsies. Epilepsy syndromes characteristic for childhood are often different in their natural history from typical syndromes encountered in adults. However, a more extensive review about paediatric epilepsy is outside of the scope of this chapter. In general, the relationship between family history and paediatric epilepsy is complex. Two studies assessing young children showed no association of family history with treatment intractability (Wirrell et al., 2012, Wirrell et al., 2013), but a large cohort of prospectively followed children have shown a negative association between family history and 2 year remission (Berg et al., 2001).

1.3.13. Treatment history

History of previous treatment is not a relevant predicting factor for treatment outcome in the case of newly diagnosed epilepsy as there is no previous treatment. However, it is valuable to explore it in more detail as it has a role in the case of established epilepsy. Furthermore, all patients with newly diagnosed epilepsy will develop a unique treatment history in due course, which will thus play a role in predicting outcome for more distant future.

It has been shown that, in newly diagnosed epilepsy, in the case of failure due to lack of efficacy of the first anticonvulsant, response to subsequent treatment decreases with every subsequent failure (Mohanraj and Brodie, 2006, Kwan and Brodie, 2000, Zhang et al., 2013). Studies have demonstrated several patterns of response. It has been proposed that some patients will go into remission, with some in indefinite or terminal remission and others relapsing later, some patients experience a relapsing and remitting course, and the remainder are treatment resistant from the outset (Brodie et al., 2012). Remission can be either early or delayed and in both cases patients who enter remission can later relapse and develop intractable epilepsy (Del Felice et al., 2010, Zhang et al., 2013, Brodie et al., 2012). It is the generally accepted view that intractable epilepsy harbours a worse long-term prognosis. Even in this group, a number of patients will enter into seizure remission, although a substantial number of them will subsequently relapse (Callaghan et al., 2011, Schiller, 2009). In both of these studies, previous treatment history was highlighted as a negative predictive factor. *Schiller* demonstrated that it is associated with seizure relapse and development of drug resistance (Schiller, 2009), while *Callaghan et al* additionally looked for predicting factors of relapse and showed that focal epilepsy is associated with reduced likelihood of relapse (Callaghan et al., 2011).

A study investigating treatment outcomes in established mesial temporal sclerosis showed that a noticeable proportion of patients still will experience some form of remission on pharmacological treatment (72 % in surgical group versus 23 % in non-surgical); responders had a lower number of previously used treatment schedules and

an earlier age at onset of the disease (Kumlien et al., 2002). Interestingly, a small community-based paediatric study involving only 77 patients with newly diagnosed temporal lobe epilepsy did not observe an effect of remission within the first two years on longer term outcome, although they did report a strong association between MRI lesions and the persistence of seizures (Spooner et al., 2006). These two studies are not comparable as one is a hospital-based adult pre-surgical series, while the other is a community-based paediatric study.

In the case of newly diagnosed epilepsy, the situation is more complicated as there is no previous treatment history, but some potentially useful prognostic information can be obtained from the initial months of the first adequate and appropriate treatment schedule. A recent post-hoc analysis of patient data from five double-blind, active control drug studies (three in newly diagnosed epilepsy) that assessed efficacy of oxcarbazepine against several mainstay AEDs showed that being seizure free after the first six months of treatment is strongly associated with being seizure free for further six months (Schmidt, 2007).

1.3.14. Learning Disabilities

Epilepsy is a commonly seen comorbid condition in children with learning disabilities; prevalence rate varies from 6 % in children with mild to 50 % in case of profound learning disability (Lhatoo and Sander, 2001). Only a few studies have reported on learning disabilities in adult newly diagnosed epilepsies. A study from China which included a mixed adult and paediatric population has shown that learning disability is not associated with treatment response in newly diagnosed epilepsy (Zhang et al., 2013). A further study from Glasgow, also including a mixed population but predominantly adults, similarly showed that presence of learning disability is not associated with development of pharmacoresistant epilepsy (Hitiris et al., 2007). However, in contrast, a study involving patients with both newly diagnosed and established epilepsy from China showed a significant association of learning disabilities with refractory epilepsy (Hui et al., 2007). Interestingly, earlier studies in newly

diagnosed epilepsy that explored the broad concept of 'neuropsychiatric handicap' also showed association with treatment failure (Shorvon and Reynolds, 1982, Elwes et al., 1984). It has similarly been demonstrated that prognosis for remission is poorer for patients with associated neurological dysfunction from birth (Annegers et al., 1979).

In summary, learning disability as a prognostic factor has been studied relatively infrequently. This clinical group is very heterogeneous both in the way that the LD diagnosis is made and also in aetiological aspects influencing the response to treatment (Lhatoo and Sander, 2001). Both learning disabilities and epilepsy might be a non-specific symptom for a specific genetic syndrome and in future with evolution of understanding of background genetics; precision based medicine could aid in a better treatment selection.

1.3.2 Prognostic research

Prognostic factor is present in the population with the given condition and is associated with the clinical outcome which can also be a response to a treatment (Riley et al., 2013). Prognostic factors are not always causally related to the condition (Moons et al., 2009). Stratified medicine on the other hand is targeting the intervention according to a shared risk common to the patient subpopulation (Hingorani et al., 2013).

Prognostic research has been carried out in the epilepsy field. For example, prognostic modelling has been applied to ascertain the risk of relapse after withdrawal from treatment in the case of remission, the recurrence of further seizures after the first seizure or treatment outcome after the application of the first AED in the case of newly diagnosed epilepsy (Group, 1993, Kim et al., 2006a, Bonnett et al., 2012, Bonnett et al., 2014b).

The general quality of prognostic research in medicine has been described as insufficient, which often limits their clinical application or reproducibility (Riley et al., 2003, Bouwmeester et al., 2012, Hemingway et al., 2009). Several important limitations

of current research have been suggested by *Riley et al*, like publication and reporting biases, poor study design and statistical analysis and lack of external validation (Riley et al., 2013). Prognostic research often relies on a retrospective design, which has several attractive features like a long observation time to collect a substantial number of events of interest but at the expense of data quality (Altman and Riley, 2005, Moons et al., 2009). On the other hand, it has significant limitations like, the lack of a pre-specified design, unclear inclusion criteria, lack of standardization and incompleteness of data. A prospective approach helps in reducing data dredging and type 1 errors, hence it is a preferred design (Riley et al., 2013).

A further important issue is data quality, and it has been suggested that factors with known inter-observer variability might be less suitable (Moons et al., 2012). Other potentially significant problems are the dichotomising of continuous co-variates and the assumption of their linearity as well as assumption that the data missingness is a random event (Moons et al., 2012, Riley et al., 2013). This could potentially be an issue in the case of newly diagnosed epilepsy with pre-treatment seizure frequency.

Sample size can also be an issue, with some studies being inadequately powered or having no reported sample size calculation (Hemingway et al., 2009, Bouwmeester et al., 2012). Small sample size can lead to over-fitted models particularly in cases where predictor selection is based on a relatively small significance level (Moons et al., 2012).

The establishment of new prospective clinical cohorts with an adequate sample size has been suggested as a way of improving the quality of prognostic research (Riley et al., 2013). Furthermore it is important that prognostic models are externally validated and updated and that the performance of new factors are assessed (Steyerberg et al., 2013).

1.4.0. Description of important cohorts

In this section, a brief description of the most important research cohorts of newly-diagnosed epilepsy will be provided; these are often cited in other relevant chapters. Furthermore, results described in this thesis are compared with them. It should also be stated that a significant proportion of the EpiPGX research cohort was derived from two of these pre-existing cohorts – those from SANAD and Glasgow.

Several studies involving large size cohorts with newly diagnosed epilepsy have assessed the effect of multiple covariates on treatment outcome or natural history. Some of them like SANAD and the Veterans Affairs multicentre studies were initially set up as clinical trials, but data were later used for analysis of association or prognostic modelling. All but two cohorts (NGPSE and Rochester study) were hospital based and all had a prospective follow up. Twelve month remission rates in all cohorts were reported to be between 50% to 80%, depending on study methodology, concordantly indicating a favourable prognosis for newly diagnosed epilepsy (Annegers et al., 1979, Collaborative Group for the Study of, 1992, MacDonald et al., 2000, Marson et al., 2007a, Marson et al., 2007b, Brodie et al., 2012, Speed et al., 2014a).

1.4.1. Department of Veterans Affairs multicentre studies (VA – 118 and VA – 264)

Both VA – 118 and VA – 264 studies were conducted using a very similar methodology, by the same research group, and mostly utilizing the VA healthcare setting. Inclusion criteria for both studies were newly diagnosed epilepsy in adults with partial onset seizures and/or secondarily generalised tonic-clonic seizures. Patients with previously documented treatment (if appropriate and adequate), severe learning disabilities, progressive neurological conditions, substance abuse and poor compliance were excluded. V-118 was conducted from 1978 to 1984, compared carbamazepine, phenobarbital, phenytoin and primidone, and recruited 622 adults in total. Overall, it showed superiority of carbamazepine and phenytoin over the other two drugs (Mattson

et al., 1985). V-264 utilized the same criteria, with 480 patients recruited between 1985 and 1991. This study compared valproate and carbamazepine and showed better results for carbamazepine in controlling complex partial seizures (CPS), but similar efficacy with regards to secondarily generalized tonic-clonic seizures (SGTC) (Mattson et al., 1992). A further post-hoc analysis of both studies showed that patients with SGTC had better prognosis than those with CPS alone (Mattson et al., 1996). Both studies had shortcomings, particularly due to the environment in which they were conducted, which led to significant selection bias. The proportion of enrolled women in both studies was around only 10%. Similarly, only around 10% of patients were not from VA Hospitals and up to 40% of patients had previous exposure to AEDs (Mattson et al., 1992, Mattson et al., 1985). Due to the specific population that VA healthcare serves, there is a potential for further selection bias by over-representation of focal post traumatic epilepsy (34% and 31% respectively) among veterans. On the other hand, both VA studies were multicentre, randomized and double-blind, with a prospective follow up and had a significant sample size.

1.4.2. Glasgow

The epilepsy research group in Glasgow has, over a long time, contributed significantly to knowledge about the pharmacological treatment of epilepsy. Its newly diagnosed patient cohort is drawn from a single hospital-based epilepsy unit and includes both adults and adolescents. Care for patients with epilepsy in the region is provided either by the Institute of Neurological Sciences at the Southern General Hospital, with expertise in diagnostics and epilepsy surgery, or the Epilepsy Research Unit at the Western Infirmary, which specialises in drug therapy. Hence, the catchment area of the epilepsy unit includes all of the Greater Glasgow area and the west of Scotland. The unit serves not only newly diagnosed patients, but also patients with an established diagnosis of epilepsy.

Although the Epilepsy Research Unit itself was based at the Western Infirmary, it was close to a community-based epilepsy practice. At certain time periods, the unit has provided additional specialised services like adolescent, pre-conception and learning disability & epilepsy clinics. The main source of referrals to this research cohort was from general practitioners, with less than 10% of patients referred from hospital A&E departments. Patients were followed up prospectively and reviewed regularly until seizure free for one year (Mohanraj and Brodie, 2005b). The unit also actively participated in regulatory trials and that was often reflected in recruitment of patients and in their treatment histories.

At the last analysis, the newly diagnosed cohort from Glasgow consisted of 1098 treatment naïve patients and was well balanced from a gender perspective with 52% (n = 575) males. Seizure freedom on the first AED in this cohort was observed in 50% of patients (Brodie et al., 2012). In addition to pharmacological and epidemiological studies, this cohort has participated in several genetic studies of newly diagnosed epilepsy (Sills et al., 2005, Makmor-Bakry et al., 2009, Szoeke et al., 2009). The sub-cohort for genetic studies is a mixture of regular patients from the unit with established epilepsy and patients enrolled into regulatory trials for newly diagnosed epilepsy conducted by the unit (Shazadi et al., 2014). Follow up for this cohort is mixed prospective and retrospective. This cohort contributes significantly to the research described in this thesis and the entire genetic sub-cohort has been phenotyped as part of this project.

1.4.3. The Italian Collaborative Group for the Study of Epilepsy

This was a multicentre, prospective observational study of patients with newly diagnosed epilepsy seen in 14 Italian hospitals and university centres. The majority of hospitals involved in the study were from northern Italy (n = 7), followed by central Italy (n=5) and the south of Italy (n =2). It started recruiting patients in 1982 and finished in 1985. Treatment was started at the time of recruitment or no earlier than three months before it. The study had a prospective design with regular follow up visits. Unselected

drugs at the standard dose for that time as a monotherapy were used. Cases where AEDs were prescribed as polytherapy or as prophylaxis were excluded (Beghi and Tognoni, 1988). Life table methods and Cox proportional hazards modelling were used for statistical processing of data. There were 280 patients included the study; 18% of them had experienced only a single seizure but the physician deemed it necessary to start treatment. The majority of patients had localisation related epilepsy (52.5%), followed by generalised (38.5%) and undetermined (6.8%) epilepsy. Phenobarbital was prescribed most often, followed by carbamazepine, valproate and phenytoin. The authors reported 1 year remission at 1 year follow-up in 62% of their patients and also demonstrated a negative association of pre-treatment seizure number with remission rates as well as a relationship between seizure relapses during the first year of follow-up and the subsequent risk of developing intractable epilepsy (Collaborative Group for the Study of, 1992).

1.4.4. The National General Practice Study of Epilepsy (NGPSE)

NGPSE study was a community-based prospective cohort study in general practice. It utilized the unique primary care system of the UK National Health Service where general practitioners (GPs) serve as gate-keepers to specialist medicine. Case identification ran from 1984 to 1987 and included 275 general practices. GPs were asked to include all cases with suspected or definite new diagnoses of epileptic seizures. The index seizure leading to inclusion in the study was not always was the first seizure. The study also included cases with symptomatic and febrile seizures. Exclusion criteria were a previous diagnosis of epilepsy and onset within the neonatal period. Data was collected from both GP practices and hospitals. Patients were not reviewed directly by the authors of the study but cases were classified by a panel based on information collected. In total, there were 1195 patients recruited, of whom 104 (9 %) were excluded (Sander et al., 1990).

From this group, 792 patients with definite or probable epileptic seizures were identified. Patients were followed up prospectively by collecting information from their

GPs, with a median follow up period of 7.2 years (MacDonald et al., 2000). The study demonstrated a relatively good prognosis for epilepsy with 68% achieving 5 year remission and pre-treatment seizure frequency was shown to be the most important factor associated with treatment response (MacDonald et al., 2000, Cockerell et al., 1997). This study was not designed to assess effectiveness of individual treatments.

1.4.5. SANAD study

The SANAD study is the largest ever randomised clinical trial in newly diagnosed epilepsy. Patients were stratified based on an initial clinical decision; if they were considered to have focal epilepsy, in which case carbamazepine was the preferred drug (arm A), or if they were suspected to have generalised or unclassified epilepsy, in which case valproate was the preferred drug (arm B). Patient recruitment lasted from December 1999 till August 2004 (Marson et al., 2007a, Marson et al., 2007b). Study design was pragmatic and resembled regular clinical practice. Hence, if the clinician felt that treatment was required it was started before investigations were carried out. The study was un-blinded. Two or more clinically definite unprovoked seizures were required before recruitment and the study also accommodated patients who were previously treated with inadequate AEDs or who had relapsed after earlier drug withdrawal. Patients with provoked seizures and known progressive neurological disorders were excluded. A further post-hoc analysis using predictive modelling was carried out by *Bonnett et al* (Bonnett et al., 2012, Bonnett et al., 2014a, Bonnett et al., 2014b).

In total, there were 1721 patients included in SANAD arm A; 378 randomised to carbamazepine; 377 to gabapentin; 378 to lamotrigine; 210 to oxcarbazepine; and 378 to topiramate. Of these, 82% were previously untreated and 10% of cases were later considered to have generalised or unclassified epilepsy (Marson et al., 2007a, Marson et al., 2007). SANAD arm B included 716 patients who were randomised either to lamotrigine (n = 239), valproate (n = 238) or topiramate (n = 239). Of these, 87.7% were

previously untreated (Marson et al., 2007, Marson et al., 2007b). Arm A showed that Lamotrigine is clinically superior to Carbamazepine for treatment of focal epilepsy, whereas arm B confirmed the superiority of valproate for the treatment of generalised and unclassified epilepsy (Marson et al., 2007b, Marson et al., 2007a). Furthermore extensive prognostic modelling for newly diagnosed epilepsy based on the SANAD cohort was carried out by *Bonnett et al* which have been reviewed in relevant sections of this thesis (Bonnett et al., 2014b, Bonnett et al., 2012).

DNA from ~40% of patients participating in SANAD was collected and has been used in previous genomic studies (Leschziner et al., 2006, Speed et al., 2014a). Like the Glasgow cohort, the SANAD study was a major contributor to the EpiPGX cohort of newly diagnosed epilepsy described in this thesis.

1.4.6. Rochester, Minnesota study

This was a longitudinal, observational study utilizing the record linkage system of the Mayo Clinic. Residents of Rochester, Minnesota with a diagnosis of epilepsy were identified from hospital records and their cases screened by single neurologist. Cases were identified between 1935 and 1984. Afterwards, patients were followed up using their medical records or via follow up enquiries (Hauser et al., 1993). This study was observational and patients were not always on treatment. Furthermore, the inclusion period covered a very long time span. That allowed the observation of temporal trends, but together with a lack of information on individual treatment schedules, this would potentially limit the ability to draw conclusions about individual drug efficacy.

1.5.0. Pharmacology of epilepsy and impact of genetic variation

Selection of the first anticonvulsant is a complex process and involves clinical factors as well as balancing patient's perspectives. Clinical factors like epilepsy syndrome or dominant seizure type, comorbid conditions, age, gender, family planning and potential drug interactions as well as costs and patient preferences all play an important role in this process. Future development of genetic knowledge will hopefully aid the process of accurate classification of epilepsy syndrome and help tailor treatment more effectively.

Drug pharmacology can be separated into two main groups, pharmacokinetics (dealing with absorption and distribution of the drug) and pharmacodynamics (targets of the drug). Anti-epileptic drugs have a complex pharmacology as their targets are CNS-based and therefore require the drugs to penetrate the blood-brain barrier. Furthermore, their targets are not necessarily specific to the epileptic process, hence there is a risk for adverse reactions.

There are several main theories trying to explain drug resistance in epilepsy. One of them – the drug transporter hypothesis - postulates that up-regulation of active efflux transporters is induced by a progressive increase in seizures, which in turn leads to a lower localised anticonvulsant concentration in the brain, rendering the drugs ineffective (Löscher and Potschka, 2002, Kwan and Brodie, 2005). It has also been proposed that epilepsy could be treatment resistant due to an intrinsic severity of the disease itself (Rogawski and Johnson, 2008, Rogawski, 2013). This approach has been substantiated by some epidemiological observations, particularly that a higher number of seizures (hence severity) before treatment is associated with a worse treatment outcome. Treatment resistance then would be an intrinsic property of the epilepsy and would be related to the same factors that are important in epilepsy severity (Rogawski and Johnson, 2008, Rogawski, 2013). Lastly, the drug target theory postulates that drug resistance is due to either acquired or genetic alteration of drug targets (Remy and Beck,

2006). All of these theories have some flaws and most likely treatment failure is due to multiple pathological mechanisms (Regesta and Tanganelli, 1999).

1.5.1. Drug efflux transporters

This section will focus on drug transporters located at the blood-brain barrier and predominantly on key members of the ATP binding cassette (ABC) family, such as P-glycoprotein (P-gp) and multidrug resistance associated proteins one (MRP1) and two (MRP2). The RLIP76 transporter also will be considered although this is not part of the ABC family.

P-gp was first described in 1976 by *Juliano and Ling* and as it altered drug permeability it was named permeability glycoprotein, or P-glycoprotein (Juliano and Ling, 1976). One of the sites where it is expressed is human capillary blood vessels at the blood-brain barrier (Cordon-Cardo et al., 1989). Increased expression of P-gp has been demonstrated in brain specimens removed from patients during surgery for intractable epilepsy (Tishler et al., 1995, Marchi et al., 2004, Dombrowski et al., 2001). Increased expression of MRP1 and MRP2 has also been associated with epileptic seizures (Miller, 2010). These proteins function as active efflux pumps, transporting a variety of substrates across cell membranes against their concentration gradient (Kerb et al., 2001). It has been proposed that over-expression of drug efflux transporters in the epileptogenic zone could lead to lower local concentrations of AEDs and thus failure to respond to treatment (Tishler et al., 1995, Sisodiya et al., 2002). For P-gp to effectively explain drug resistance, the majority of available AEDs should be substrates for transport. Research so far has produced conflicting results. There are suggestions that many commonly used AEDs are either definite or probable P-gp substrates (Zhang et al., 2012) but other research has demonstrated only limited affinity of AEDs (Löscher and Sills, 2007, Dickens et al., 2013, Anderson and Shen, 2007).

Polymorphisms in the *ABCB1* (ATP-Binding Cassette Subfamily B Member 1) gene that encodes P-gp have been extensively studied in relation to treatment resistance in epilepsy. One of the most studied single-nucleotide polymorphisms (SNPs) in *ABCB1* is

the synonymous 3435C>T SNP (rs1045642) located in exon 26 (Hoffmeyer et al., 2000). It was first described by *Hoffmeyer* et al in 2000, who demonstrated that volunteers expressing the T-allele have higher digoxin plasma level compared with those carrying the C-allele.

The first study exploring this variant in pharmacoresistant epilepsy showed an association with the CC genotype (Siddiqui et al., 2003), although later studies have reported controversial results covering multiple possible outcomes. A string of studies covering diverse ethnic backgrounds failed to show an association between CC genotype and intractable epilepsy in adults (Sills et al., 2005) (Kim et al., 2006b), as well as in paediatric cases (Vahab et al., 2009) (Dong et al., 2011). A Malaysian study investigating response to monotherapy with either CBZ or valproate (VPA) also did not show any association (Haerian et al., 2011a), nor did an additional study comparing CBZ blood levels or dose/concentration ratios (Ozgon et al., 2008). Similarly, there was no association between CBZ and PHT dosing and the *ABCB1* 3435C>T genetic polymorphism (Tate et al., 2005).

Two prospective studies exploring the link between this genetic polymorphism and treatment response in newly diagnosed epilepsy also did not yield any significant association (Szoeke et al., 2009, Leschziner et al., 2006). On the other hand, several studies have shown an association between CC genotype and drug resistant epilepsy (Siddiqui et al., 2003, Soranzo et al., 2004), although the patient cohort partially overlapped in these studies. Similar results have been reported from a study examining drug resistant and responsive patients with temporal lobe epilepsy (Zimprich et al., 2004). This study exclusively enrolled treatment resistant patients undergoing pre-surgical evaluation, with healthy volunteers used as controls. A study from Taiwan also demonstrated a positive association of the 3435C>T polymorphism with drug intractable epilepsy (Hung et al., 2007). This study defined drug resistance as the occurrence of 10 or more seizures per year in patients who had tried two or more appropriate AEDs at maximal tolerated dose, while drug responsiveness was defined as the absence of

seizures for two years. Such a definition excludes a number of subjects in-between those categories, thus it is not fully representative of the epilepsy population. A study from China has interestingly reported an association between drug intractable epilepsy and the TT genotype of 3435C>T (Kwan et al., 2007), likewise *Seo et al* have reported similar results for CBZ monotherapy in Japanese patients (Seo et al., 2006). Thus, one can conclude that conflicting results are not confined to one ethnic group.

The 3435C>T polymorphism is not the only SNP studied; other *ABCB1* polymorphisms also have been investigated, most notably 2677T/A>G and 1236C>T. The number of these studies is smaller, but nevertheless results show similar controversies. In the case of 2677T/A>G, several studies have failed to find an association with drug resistance (Vahab et al., 2009, Kim et al., 2006b, Kim et al., 2009a, Shahwan et al., 2007, Von Stülpnagel et al., 2009), although a handful of reports are positive (Kumari et al., 2011a, Kwan et al., 2009). The situation is similar with 1236C>T and most studies do not report an association (Grover et al., 2010, Haerian et al., 2011b, Kumari et al., 2011a, Kim et al., 2009b, Qu et al., 2012). Several meta-analyses have also been carried out, although, again, the results are not uniform. Some showed no association between 3435C>T polymorphism and drug resistance (Haerian et al., 2010, Haerian et al., 2011c, Bournissen et al., 2009), whereas others showed a weak association in Caucasians only but no association overall (Li et al., 2015b, Li et al., 2015a, Lv et al., 2014). When assessed separately, the *ABCB1* C1236T / G2677T / C3435T haplotype showed no association with drug resistance even when stratified by ethnicity and the same study also showed a lack of association for *ABCB1* C1236T and G2677T with drug resistance when analysed separately (Li et al., 2015a). One systematic review reported that heterogeneity is mainly derived from six positive studies (Haerian et al., 2010).

All of these studies investigating the drug transporter hypothesis report somehow conflicting results. There is more than one explanation for this situation. The main problems are related to either study design, phenotype definitions, or sample size. For example, none of the above studies reported initial sample size calculations or an

assessment to determine whether there was sufficient power to detect genetic association. Sample size in these studies was very variable, ranging from 60 patients to 685 patients, and most are significantly underpowered. Lack of unified and clear phenotype definitions is also an issue; a patient classified as drug resistant in one study could be classified as drug responsive in another (Kasperaviciute and Sisodiya, 2009). Definitions for drug resistance and response across different studies often identified very different patient groups. In some cases, generalized epilepsy was more prevalent in the drug responsive group and focal epilepsy, or those with co-morbid learning disabilities, was overrepresented in the intractable group. Several studies have reported demographic data only partially or not at all. Other potential bias is the fact that cohorts are often assembled from patients under the care of the tertiary level hospitals, that routinely deal with multi-drug resistant patients. Added to that, follow-up time is very variable, ranging from 3 months to 2 years, which might also contribute to heterogeneity (Li et al., 2015a).

Available data about *ABCC1* (ATP Binding Cassette Subfamily C Member 1; multidrug resistance-associated protein; MRP1) and *ABCC2* (ATP Binding Cassette Subfamily C Member 2; MRP2) is not as extensive as for P-gp. Literature about AEDs transported by both MRP transporters is limited. *Luna-Tortós et al* reported that neither carbamazepine, valproate, levetiracetam, phenytoin, lamotrigine nor phenobarbital is transported by MRP1 or MRP2 (Luna-Tortós et al., 2010). Compared to P-gp, the link between genetic polymorphism in MRP genes and drug response has been investigated in a very limited fashion. Only a few studies have investigated *ABCC1* polymorphisms with regards to epilepsy treatment response. A study from India investigated ABC transporter genetic polymorphisms in relation to treatment response and included *ABCB1* and both *ABCC1* and *ABCC2* (Grover et al., 2012). It reported a statistically significant association between *ABCC2* promoter polymorphisms 1549G>A and 1019A>G and recurrent seizures in females despite using a highly conservative approach (i.e. Bonferroni method) to correct for multiple testing. Follow up period in this study was just 12 months, of which two months were assigned to achieve steady drug

concentration and 10 months to assess response. The authors initially enrolled 400 patients, but by the end of the study the dropout rate was 46% (184). This represents a significant drop out rate for a short study and undermines its power, as well as raising questions about potential attrition bias and the validity of results. Furthermore, it is doubtful if 10 months is sufficient follow up length to evaluate treatment response. In a Caucasian paediatric and adolescent cohort, the *ABCC2* 24T variant was reported to be significantly over-represented among non-responders (Ufer et al., 2009a). Symptomatic epilepsy was also over-represented in the non-responder group compared to responders. Meanwhile, a later study on childhood epilepsy from the same authors failed to replicate their previous finding (Ufer et al., 2011). Any link between the *ABCC2* 24T variant and drug resistance has not been confirmed in several adult studies (Seo et al., 2008, Zimprich et al., 2012), although some positive results also have been reported (Qu et al., 2012). Meta-analysis of available studies has shown conflicting results. One showed no statistically significant association for the most common SNPs whereas two others reported that *ABCC2* G1249A might have a decreased risk of treatment resistance (Wang et al., 2015, Chen et al., 2014a, Grover and Kukreti, 2013). With regard to newly diagnosed epilepsy, it has been shown that *ABCC2* polymorphisms have no association with treatment response to carbamazepine and furthermore, the authors showed that the drug is not a substrate of this transporter (Radisch et al., 2014).

RLIP76 is a non ABC family multi-specific transporter that has been found at the blood-brain barrier in surgically removed epileptic foci (Awasthi et al., 2005). In 2007, a study carried out by *Soranzo et al* showed that in both epileptic and control brain tissue RLIP76 was not endothelial. Furthermore, they genotyped six tagging SNPs in drug resistant and drug responsive groups and showed no association with phenotype (Soranzo et al., 2007). Later studies investigating common RLIP76 polymorphisms in epilepsy, including in newly diagnosed cases, failed to identify any influence on drug response (Leschziner et al., 2007, Manguoğlu et al., 2011).

In conclusion, it is unlikely that genetic polymorphisms in drug efflux transporters play a major role determining treatment outcomes.

1.5.2. Sodium channels

Voltage gated sodium channels have been extensively studied for their role in neurology, and particularly in epilepsy. In an adult human brain there are four main sodium channel subtypes, $Na_V1.1$, $Na_V1.2$, $Na_V1.3$, and $Na_V1.6$, encoded by the Sodium Voltage-Gated Channel Alpha Subunit 1 (*SCN1A)*, Sodium Voltage-Gated Channel Alpha Subunit 2 (*SCN2A)*, Sodium Voltage-Gated Channel Alpha Subunit 3 (*SCN3A)*, and Sodium Voltage-Gated Channel Alpha Subunit 8 (*SCN8A)* genes, respectively (Noebels et al., 2012). Mutations in these genes create a wide array of epileptic syndromes ranging from benign familial neonatal seizures to catastrophic epileptic encephalopathies, such as severe myoclonic epilepsy of infancy (SMEI). However, only 1 – 2 % of idiopathic epilepsies seems to be monogenic (Weber and Lerche, 2008). In terms of pharmacogenetics, SMEI is particularly important. There have been observations that AEDs with sodium channel properties, such as lamotrigine and carbamazepine (Rogawski and Loscher, 2004), can exacerbate seizures in patients with SMEI (Horn et al., 1986, Guerrini et al., 1998). It has now been postulated that selective impairment of sodium channels in GABAergic inhibitory neurons might be the underlying mechanism for SMEI (Noebels et al., 2012).

Genome based biomarkers have been extensively investigated as predictive factors for treatment response and dosage parameters of AEDs with a main action on sodium channels. An association between *SCNA1* IVS5–91 G>A and response to sodium channel blocking AEDs has been shown in Japanese and Han Chinese populations (Abe et al., 2008, Ma et al., 2014). Further studies have demonstrated an association between this polymorphism and clinically effective and/or maximal-tolerated doses of CBZ (Zhou et al., 2012, Hung et al., 2012, Tate et al., 2005), phenytoin (Tate et al., 2006, Tate et al., 2005) and oxcarbazepine (Ma et al., 2015). The methodology in some of these studies is

questionable, for example Zhou *et al* reported a significant association but apparently failed to correct their analysis for multiple testing, which would have resulted in the loss of statistical significance. As with the efflux transporter studies reviewed above, there are at least as many studies showing no link between the IVS5–91 G>A polymorphism and drug resistance (Sánchez et al., 2010, Kumari et al., 2013) or CBZ dosage (Zimprich et al., 2008).

Other *SCN1A* polymorphisms that have been investigated include 3166 A>G and 603-91G>A, both without association (Sanchez et al., 2010) (Manna et al., 2011). The situation with A3184G is interesting as there are two studies from the same institution with opposing results (Kumari et al., 2011a, Lakhan et al., 2009). The *SCN2A* 56G>A polymorphism has been associated with drug resistant epilepsy in a northern Indian population (Lakhan et al., 2009) and this is one of the few studies that has performed a sample size calculation. A study from Taiwan investigated multiple SNPs and found association between *SCN2A* IVS7-32A>G and resistance to AED treatment in a Han Chinese population. The authors of this study created a subgroup in their cohort of patients who were treated exclusively with sodium channel blocking AEDs but the association in this subgroup was no stronger than that seen in general. (Kwan et al., 2008). Two multi-centre cohort studies involving patients from Malaysia and Hong Kong have failed to demonstrate any association between treatment outcome and common *SCN1A, SCN2A, SCN3A* polymorphisms, and two further meta-analyses have been similarly unsuccessful (Haerian et al., 2013, Haerian et al., 2012).

Polymorphisms in sodium channel genes are unlikely to be a general biomarker for treatment response. Nevertheless, as sodium channel defects are causal in some epileptic syndromes, genetic testing and diagnosis might have implications for treatment selection. For example, in Dravet syndrome, the majority of patients will have a mutation in *SCN1A* and their treatment strategy might be different from other forms of early onset epilepsy (Dravet and Oguni, 2013). In conclusion, in selected patient populations there might be a role for testing mutations in sodium channel genes as a

part of the clinical work-up but otherwise these genes do not appear to harbour valuable biomarkers of treatment outcome in general.

1.5.3. Gamma-aminobutyric acid system

The GABAergic system is one of the main inhibitory systems in the brain and has been extensively studied in relation to almost all aspects of epilepsy. Mutations in genes encoding GABAergic system have been associated with generalised epilepsy, SMEI or febrile seizures (Baulac et al., 2001, Wallace et al., 2001, Cossette et al., 2002, Harkin et al., 2002, Audenaert et al., 2006, Johnston et al., 2014). Furthermore mechanism of action of some AED are via potentiation of GABAergic system (Kwan et al., 2001). This makes the GABAergic system an attractive target for pharmacogenomics research. Polymorphisms in Gamma-Aminobutyric Acid (GABA) related genes and treatment response in epilepsy has also been investigated, although not as extensively as P-gp or other potential drug targets. Several studies have investigated the potential link between genetic polymorphism in GABA receptor subunits and drug resistance. One such study evaluated *GABRA1* IVS11 + 15 A > G and *GABRG2* 588C > T and reported an association of the G-allele of the *GABRA1* with both susceptibility for epilepsy and drug resistance, but no association was observed for the *GABRG2* 588C>T variant (Kumari et al., 2010). A later study based on the same cohort investigated polymorphism in different GABA receptor subtypes. It showed no association for *GABRA6* c. 1512 T>C, *GABRB2* c. 1412 C>T, and *GABRR2* c. IVS2C>G polymorphisms and drug resistance (Kumari et al., 2011b). An further study utilising mesial temporal lobe epilepsy due to hippocampal sclerosis as a prototype resistant epilepsy and Juvenile Myoclonic Epilepsy (JME) as responsive epilepsy showed no difference in $GABA_A$ receptor SNP distributions between the groups (Balan et al., 2013). Likewise, a study from Taiwan investigating SNPs in the α-subunit of the $GABA_A$ receptor showed no significant difference in individual genotype distributions between responsive and treatment resistant patients but interestingly, a logistic regression model adjusted for age, aetiology and epileptic syndrome suggested that the combination of *GABRA1* (rs6883877 C>T), *GABRA2*

(rs511310 A>G) and *GABRA3* (rs4828696 C>T) SNPs was associated with treatment outcome (Hung et al., 2013).

In conclusion, as in the case of sodium channels, testing for genomic biomarkers related to the GABAergic system might ultimately have a role in a clinical practice, but as an aspect of diagnostic work-up rather than in the prediction of treatment outcome.

1.5.4. Hepatic metabolism

Some of the most commonly-used AED, like phenytoin, valproic acid, carbamazepine and lamotrigine, are entirely or at least partially metabolized in the liver. The most widely studied drug metabolising system is the cytochrome P450 (CYP) system. Several AEDs are substrates for CYP enzymes, including phenytoin which is metabolized extensively by Cytochrome P450 Family 2 Subfamily C Member 9 *(CYP2C9)* and *CYP2C19* (Klotz, 2007). Numerous studies have investigated efficacy, adverse effects and/or blood levels of this drug in relation to variability in the genes encoding these two enzymes. Studies examining PHT pharmacokinetics in the Japanese population have demonstrated a role of genetics in the metabolism of the drug (Odani et al., 1997, Hashimoto et al., 1996). A study from the Netherlands involving 60 individuals with learning difficulties, covering ages between 16 to 74 years, showed an association between dose requirements of PHT and *CYP2C19* genetic variants. However, only 13 of patients were on PHT monotherapy (van der Weide et al., 2001). A later investigation looking at dosing of CBZ and PHT reported an association between the *CYP2C9*3* polymorphism and maximum dose of PHT. The PHT group had 281 patients and carbamazepine had 425, thus 185 patients had been on dual treatment with potential interaction between drugs (Tate et al., 2005). Furthermore, there are individual case reports of significant phenytoin toxicity occurring after emergency usage in *CYP2C9* poor metabolizers (Dorado et al., 2013).

The Pharmacogenetics Working Group of the Royal Dutch Association for the Advancement of Pharmacy conducted a systematic review on a list of common

medicines and the effect of genetic factors on their dosing requirements and afterwards issued nationwide recommendations. The advice for phenytoin is a 25% dose reduction in carriers of *CYP2C9* common variants (*1/*2, *1/*3) and 50% reduction in carriers of less common but more dysfunctional alleles (*2/*2, *2/*3, *3/*3) (Swen et al., 2011). Phenobarbital also has been investigated in this respect and an association between dosage and *CYP2C19* polymorphisms seen (Mamiya et al., 2000).

Several studies have analysed CYP polymorphisms and response to treatment without regard to specific AEDs. A mixed paediatric and adolescent cohort from Germany was used to explore the link between several *CYP2C8*, *CYP2C9*, and *CYP2C19* polymorphisms and treatment outcome but arrived at inconsistent results and concluded that these SNPs have only minor impact on drug response (Ufer et al., 2009b). Another mixed adult and paediatric cohort from Spain showed no association for drug responsiveness and genetic polymorphisms in CYP enzymes (Sanchez et al., 2010). A study from India demonstrated that the *CYP2C9* 1075A>C (*3 variant) polymorphism shows significantly different genotype distributions in multi-drug resistant and drug-responsive patients (Kumari et al., 2011a). However, despite testing multiple polymorphisms, the authors failed to correct for multiple comparisons and once this is done, the results lose their statistical significance.

Another enzyme involved in drug metabolism that has been investigated in pharmacogenetic studies is epoxide hydrolase (encoded by *EPHX1*). A Japanese study showed that several *EPHX1* haplotypes are associated with altered CBZ metabolism (Nakajima et al., 2005). Furthermore, there are two studies that have evaluated CBZ dosage and *EPHX1* genetic polymorphisms, both of them reporting association of genetic polymorphisms with CBZ maintenance dose (Hung et al., 2011, Makmor-Bakry et al., 2009).

Polymorphisms in genes encoding enzymes belonging to UDP-glucuronosyltransferase family have also been investigated, although far less often than others. There is a reported effect of the *UGT1A4* L48V variant on LTG serum

concentration when used as monotherapy (Gulcebi et al., 2011). A further study used multivariate modelling and showed an association of *UGT2B7* c.802T>C, together with *SCN1A* IVS5–91 G>A and *EPHX1* c.337T>C, with concentration–dose ratios of CBZ (Hung et al., 2011). A similar investigation looking at VPA dose optimization indicated that SNPs in *UGT1A6* and *GRIN2B* might be associated with concentration–dose ratios of VPA (Hung et al., 2011).

In conclusion, the metabolism of phenytoin is most likely influenced by the genetic status of *CYP2C9*. Because of its complex pharmacokinetics, it probably plays a role in the development of neurotoxicity with this drug. However, generally speaking, genetic polymorphisms in metabolising enzymes have only a modest effect of AED pharmacokinetics. In the absence of a correlation between AED concentration and treatment response (Tomson et al., 2007), which explains why regular measurement of serum levels is not a widespread clinical practice, SNP genotyping is of limited value even where there is evidence of a link between genotype and dosing.

1.5.5. Adverse events

Five commonly used drugs for the treatment of epilepsy have labels that have been updated by the U.S. Food and Drug Administration (FDA) to include pharmacogenetic information. These are carbamazepine, valproic acid, clobazam, phenytoin and diazepam, although in the case of diazepam the FDA does not advise about genetic testing or genetic factors but simply informs about the role of *CYP3A4* and *CYP2C19* in its metabolism and in potential drug interactions. A study from Japan investigating recovery time after general anaesthesia when diazepam had been used as pre-medication showed an influence of *CYP2C19* genotype (Inomata et al., 2005). In the case of clobazam, the FDA advise changes in the starting dose and pace of titration for patients who are known to be *CYP2C19* poor metabolizers .

Carbamazepine and phenytoin are two commonly prescribed AEDs with known genome based markers associated with cutaneous hypersensitivity reactions. HLA-B*15:02 has a strong association with CBZ-induced Stevens-Johnson syndrome in the

Asian population (Yip et al., 2012) and a moderate association with similar reactions induced by phenytoin (Cheung et al., 2013). Additionally, HLA-A*31:01 is associated with a broader phenotype of CBZ cutaneous hypersensitivity across multiple ethnicities (Yip et al., 2012, Amstutz et al., 2014). Recent pharmacogenetics research for both drugs has resulted in changes in labelling (U.S. Food and Drug Administration, 2011a),(2011b). Interestingly, lamotrigine, which is another widely used anticonvulsant associated with cutaneous hypersensitivity reactions, so far has no robust HLA association identified (Bloch et al., 2014).

Valproic acid is contraindicated in persons with urea cycle disorders and there are recommendations that it is investigated further in patients who develop unexplained hyperammonemic encephalopathy (U.S. Food and Drug Adminstration, 2016). It also carries a risk of potential liver toxicity, particularly in patients with Alpers-Huttenlocher syndrome caused by a mutation in mtDNA replicase, polymerase gamma 1 (POLG) (Stewart et al., 2010, Saneto et al., 2010), although it has no formal warning in this regard.

1.6. Newly diagnosed epilepsy and GWAS

At the moment there is only one prospective pharmacogenomics study utilising GWAS for newly diagnosed epilepsy (Speed et al., 2014a). This study and the work described in this thesis shared common patient cohorts and methodological aspects. *Speed et al* analysed 1296 cases of newly diagnosed epilepsy, with the study cohort compromising SANAD patients and a prospectively followed cohort from Melbourne, Australia. Measured outcome was 12 months or longer seizure freedom and patients with incomplete follow-up were excluded from the study. Univariate analysis of clinical prognostic factors was performed on both sub-cohorts separately. Factors from the SANAD sub-cohort that were statistically significantly associated with the presence of 12 months remission were later used in the GWAS analysis. GWAS was performed both with and without adjusting for relevant clinical covariates. The following factors were used for adjustment; age at starting treatment, number of seizures before treatment,

EEG result, epilepsy type, presence or absence of neurological impairment, and treatment with gabapentin. After QC procedures, the GWAS was performed on 889 newly treated epilepsy patients who had complete follow-up data. This was just 68.6% of the total available cases and although the Melbourne sub-cohort lost more cases (61.8% included) than the SANAD sub-cohort (71.4% included), there was no difference in the proportions of responders (experiencing 12 month seizure freedom) and non-responders between excluded and included patient groups. The GWAS was performed separately on each sub-cohort and a fixed-effect meta-analysis subsequently carried out.

To summarise the findings, no single SNP reached genome-wide significance, but three loci had suggestive evidence of association; two from the unadjusted analysis (rs492146 and rs72700966) and one from the adjusted analysis (rs143536437). The study was sufficiently powered to detect variants with the strong effect size, but was not intended to assess treatment response as a polygenetic trait.

1.7. Methodological aspects

As a rapidly evolving field, pharmacogenomics has not been spared from controversies and problems. As expected, there is significant room for growth and improvement. The situation when significant number of molecular genetic studies fail to fully follow the basic principles of clinical epidemiology is not new (Bogardus, 1999). A study evaluating methodological quality of pharmacogenetic studies using binary responses showed that significant number of studies have problems with design and methods. From 65 reviewed studies, only one reported a planned sample size and had performed a prior sample size calculation, only one third provided detailed information on design, and a small minority undertook any correction for multiple testing (Cobos et al., 2011).

Epilepsy pharmacogenomics is not excluded from the methodological problems that are seen in the rest of the field. In 2004, *Neurology* had two articles in the same edition on ABCB1 polymorphisms and treatment response that reported different results (Zimprich et al., 2004, Tan et al., 2004). In the accompanying editorial, *Ott* drew the conclusion that a positive result is likely due to a combination of low pre-test probability and insufficient sample size (Ott, 2004). A further problem is with heterogeneity in the definitions of treatment outcome. A paediatric epilepsy study applying six different definitions of intractability that had been employed in previous investigations found that the proportion of children counted drug resistant ranged between 9% - 24%, depending on the definition used (Berg and Kelly, 2006). This demonstrates that differences in outcome can be created simply by applying various different definitions. This could be a potentially important issue in the case of small and underpowered pharmacogenomics studies in epilepsy. It has already been demonstrated that remission rates vary significantly between studies. These differences have been explained by several factors, including case ascertainment (prospective studies having more recurrent seizures than retrospective), the proportion of patients included with a single seizure (higher proportion of remission in cohorts with more single seizure cases), age of the patients included, and duration of follow-up (Abimbola et al., 2011). Epidemiological studies provide further corroborative evidence of the importance of case definition. It has been shown that the incidence of epilepsy is affected depending on how epilepsy has been defined, with higher incidence reported when single unprovoked and acute symptomatic seizures are included (Banerjee et al., 2009).

Case ascertainment is an important aspect of research design. Prospective study design is the gold standard in observational studies. If a case-control design is used, groups should only differ in respect of the outcome studied and controls should have a chance to develop the outcome of interest (Jorgensen and Williamson, 2008). Hence, using healthy controls is inappropriate as they have never been exposed either to treatment or to the disease itself.

A systematic review of prognostic factors for medically intractable epilepsy has shown that there is significant heterogeneity and bias towards hospital based studies (Wassenaar et al., 2013). On the other hand, randomized clinical trials are often designed for regulatory purposes, enrol highly selected patients and employ short follow-up periods that are consistent with short-term treatment exposure (Kwan and Sander, 2004).

Most research in epilepsy pharmacogenetics thus far has concentrated on a simple monogenic explanation, hypothesizing that one mechanism explains drug intractability. This approach is questionable from a biological perspective. It is not clear how it fits with currently available epidemiological data showing that some patients experience a remitting-relapsing pattern to their disease course (Brodie et al., 2012, Sillanpaa and Schmidt, 2006). It is unlikely that such a pattern could be explained by simple monogenic effect – it is more appropriate to suggest that the situation is considerably more complicated, and perhaps reflects the complexity of the epileptogenesis process itself.

2.0. Aims and objectives of this work

The introductory chapter described current knowledge about newly diagnosed epilepsy and more extensively elaborated on clinical factors associated with treatment outcome, together with methodological and pharmacogenomic aspects. As outlined earlier, the majority of patients with newly diagnosed epilepsy will have a favourable prognosis while around 30 – 40 % will fail their first anticonvulsant. Early identification of patients who are likely fail treatment would potentially pave the way to early intervention and improvement of treatment outcomes as well as improvement of quality of life.

To date, there is only one large scale GWAS study carried out exclusively on patients with newly diagnosed epilepsy (Speed et al., 2014a). It did not identify any SNPs with genome-wide significance. Data on associated clinical factors are more widely reported, but are often limited by sample size and the results are still controversial. As discussed in the introductory chapter, it is advisable to perform GWAS as a univariate analysis and also adjusted for clinical covariates. Significant numbers of pharmacogenomic studies have investigated specific aspects of epilepsy therapy (mostly drug resistance) but with the exception of one or two ADR studies these have been controversial and not consistently replicated. This problem could potentially be due to methodological aspects of those studies, such as the procedures for cohort assembly and follow up.

In an effort to improve current knowledge about prognosis of newly diagnosed epilepsy, the work described in this thesis had the following aims and objectives:

1. To assess clinical factors and their relationship with 12 month remission from seizures after application of the first well tolerated anticonvulsant.

a. Data was collected by manual phenotyping of cases, by transcription of existing databases into the EpiPGX eCRF, and collected from partners of the EpiPGx research consortium.

b. Data was then downloaded from the central repository and interrogated by using routine descriptive statistical methods.

c. A logistic regression model assessing association of clinical factors with 12 month remission of newly diagnosed epilepsy was applied.

d. A Cox regression model assessing association of clinical factors with time to onset of 12 month remission in newly diagnostic epilepsy was applied.

2. To investigate the impact of methodological heterogeneity in studies of treatment outcome in epilepsy.

a. Data was collected and processed in the same way as described for aim one above.

b. Different definitions of remission were applied and the proportions of responders compared in each case to assess the impact of terminology on treatment outcome.

c. The impact of ascertainment methods (prospective vs. retrospective) and duration of observation were investigated by routine statistical methods.

d. The effect of transcription and electronic upload on data quality and reliability were assessed by comparing the original SANAD genetic cohort with the newly-uploaded version of SANAD using routine statistical methods.

3. To explore genome based biomarkers of treatment response in newly diagnosed epilepsy.

a. A GWAS for treatment response to first well tolerated anticonvulsant was performed, with expert input from colleagues and collaborators in Work Package 2 of the EpiPGX consortium.

b. A separate GWAS of 12-month remission on first well tolerated AED in patients with newly diagnosed focal epilepsy was also performed using univariate analysis and also adjusting for important clinical covariates.

3.0. Methods

My thesis is based on the EpiPGX project; hence this chapter will briefly describe the overall project, followed by a more detailed description of EpiPGX Work Package 2 (Genomic Biomarkers of Early Treatment Outcome in Newly Diagnosed Epilepsy). Lastly, methods used for analysis of (a) association of clinical covariates, (b) methodological aspects of cohort construction, and (c) GWAS will be described.

3.1. General overview of EpiPGX project

EpiPGX is the acronym for the project whose full title is 'Epilepsy Pharmacogenomics: delivering biomarkers for clinical use'. The project was started on 1st November 2011 and finished on 31st October 2015. The main project coordinator was Professor Sanjay Sisodiya from University College London. It was funded by the European Union Seventh Framework Programme (FP7/2007-2013) under the grant agreement n° 279062, as a small to medium scale focused research project, involving collaboration between academic and public bodies with small to medium sized enterprise.

The project had the following participants:

1. University College London (UCL) United Kingdom.
2. Université Libre de Bruxelles (ULB) Belgium.
3. Istituto Giannina Gaslini (IGG) Italy.
4. Eberhard-Karls-Universität Tübingen (EKUT) Germany.
5. Stichting Epilepsie Instellingen Nederland (SEIN) The Netherlands.
6. Universitaetsklinikum Bonn (UKB) Germany.
7. Royal College of Surgeons in Ireland (RCSI) Ireland.
8. Belfast Health and Social Care Trust (BHSCT) Northern Ireland.
9. Islensk erfðagreining ehf (deCODE) Iceland.

10. Luxembourg Centre for Systems Biomedicine University of Luxembourg (LCSB) Luxembourg.
11. University Medical Center Utrecht (UMCU) The Netherlands.
12. University of Liverpool (ULIV) United Kingdom.
13. Imperial College London (IMP) United Kingdom.
14. University of Glasgow (UGLA) Scotland.
15. GABO:mi Gesellschaft für Ablauforganisation: Milliarium mbH & Co. KG Germany.

The project consisted of several work packages, which either provided the practical support required for implementation of the project or led a specific research direction. The overall main goal of EpiPGX was to investigate and identify genome-based predictive biomarkers for clinical application in patients with epilepsy. It covered all aspects of epilepsy care, including newly diagnosed and treatment resistant patients as well as selective drug response, adverse reactions, and valproate teratogenicity. Supportive work packages were involved in the development of unified phenotype definitions (in close collaboration with scientific work packages), the support and development of a centralised data repository, as well as genotyping and administrative support. The work packages, their goals, and leading institutions have been summarised in Table 2. The University of Liverpool led Work Package Two (WP2), investigating genome-based biomarkers related to treatment outcomes in newly diagnosed epilepsy.

Table 2. EpiPGX work package list and their main goals.

WP No	Title of the work package	Lead	Main goals
1	Characterisation of pharmacogenomics phenotypes	ULB	Generation of robust phenotype definitions for cross-consortium application Implementation, updating and phenotyping quality control Interaction with other pharmacogenomics networks and organizations
2	Genome-based biomarkers of early treatment response in newly diagnosed epilepsy	ULIV	Identification of biomarkers for treatment response with first well-tolerated AED General and selective drug responsiveness in newly diagnosed epilepsy Biomarkers of treatment failure with first AED
3	Genome-based biomarker discovery for broad AED resistance	UCL	Identification of common genetic variants (SNP/CNV) for broad drug resistance Genome-based biomarkers of shared rare variants and genes with increased burden of individual variants for individuals with extreme phenotypes
4	Genome-based biomarker discovery for late response to specific AEDs	EKUT	Genome-based biomarkers for late response/non-response to specific AEDs in focal and generalised epilepsies Common genome-based biomarkers for late drug response
5	Genome-based biomarker discovery for specific ADRs	RCSI	Assessment of common variants as genome-based biomarkers for each AED–associated ADRs via genetic mapping across cases and controls Rare genetic variants as biomarkers for a specific ADR via high-throughput sequencing in selected cases
6	Genome-based biomarker discovery for valproate teratogenesis	BHSCT	To establish genetic and clinical database for babies born with AED-induced major congenital malformations and their parents Genetic variants associated (including CNV) with major congenital malformations associated with in utero exposure to VPA Epigenetic changes induced by VPA Identification of rare variants associated with AED-induced major congenital malformations
7	Core analytic and bioinformatic processing	LCSB	To provide bioinformatics support To carry out multivariate analysis for genetic variants associated with treatment response Assessment of identified variants using computational-based functional and knowledge-driven analysis
8	Development of diagnostic tests and *in silico* database	deCODE	Genotyping Development of genomic assays for identified variants and clinical prediction tools Generation in-silico virtual test bed for drug development
9	Project Management	GABO:mi	Administrative support
10	Dissemination and training	UCL	Dissemination and coordination

Prior to the start of data collection and cohort assembly, unified phenotype definitions were agreed between the centres with subsequent testing and implementation (Appendix 2, page 222). As a part of the testing process for how phenotypes should be applied, each clinical work package provided a selection of anonymised cases. Those cases were then phenotyped and outcomes were compared between centres. This effort was coordinated by work-package 1. A similar exercise was repeated twice during the project and the results were analysed and reported during the annual general assembly of EpiPGX.

For the data collection, a central repository (electronic case record form; eCRF) was used. A paper version of the CRF was produced (Appendix 1, page 201) but the electronic version was used exclusively. The eCRF was based on the FileMaker platform and hosted on a server at deCODE (Iceland). Each clinical centre accessed the eCRF via remote access software. Access was controlled and password protected. Maintenance and updating of eCRF was a collaborative effort led by deCODE, LCSB, and assisted by the relevant WPs.

Cases were either manually phenotyped or data were transcribed and uploaded from existing clinical databases. Manual phenotyping was done locally utilising clinical notes. Policies on data transfer differed from centre to centre. The approach used in Liverpool is described in more detail later in this chapter.

Data relevant to each individual centre were accessible to the respective centre in browsing mode using distant access software. Prior to analysis, selected (in case of EpiPGX Work package 2, all newly diagnosed) cases were downloaded and transferred to Liverpool. Permission to access the data was obtained from individual centres ahead of the download. Further work with the data was carried out locally. Genomic data was stored on a cluster at the LCSB. The main GWAS analyses were performed on the same LCSB cluster, using remote access software.

3.2. Ethical aspects

The EpiPGX project complied with the highest ethical standards, national and local regulations. All clinical centres had received prior approval from either their local or national ethics committees for pharmacogenomic analysis of contributing samples and clinical data. All research activities were undertaken by EU member countries and one associated country (Iceland). Strict data confidentiality standards were followed; genotypic and phenotypic information was anonymized. In addition, the project had a separate ethical advisory board. The project office collected and archived all required ethical approvals at the outset of the project.

3.3. EpiPGX Work Package 2: Detailed description

3.3.1. Sample size

One of the main aims of the EpiPGX project was to collect and transfer to a single repository all available historical data. There was little active on-going patient recruitment during the project, hence no prior sample size calculation was carried out. A retrospective power calculation using CaTS software was performed afterwards (Skol et al., 2006). This software was designed based on work carried out by *Skol et al* which have shown that a joint analysis is more efficient that two stage replication approach (Skol et al., 2006).

3.3.2. Cohort assembly

Assembly of a newly diagnosed epilepsy cohort was based on patient status at the first visit recorded in their clinical records. Cases with previously existing epilepsy were excluded from EpiPGX WP2. Each centre collected data and assembled their cohort separately. The eCRF served as a unified and centralised data capture vehicle. Collection of data was carried out either manually, utilising contemporary clinical records, or by transferring pre-existing clinical trial data or clinical databases. Methods of assembly of each separate cohort used in EpiPGX WP2 are summarised in table 3. Once the data were collected, they were downloaded from the eCRF and phenotypes were derived.

Strict adherence to phenotype definitions was enforced, and only newly diagnosed cases were downloaded and utilised.

Table 3. Methods of assembly and sources of each individual cohort used to form the newly diagnosed epilepsy cohort in EpiPGX WP2.

Cohort	Origin	Method of data collection
UCL	Institute of Neurology, UCL	Manual phenotyping
EKUT	Eberhard-Karls-Universität Tübingen, Tuebingen, Germany	Manual phenotyping
ULB	Université Libre de Bruxelles	Data transfer from existing clinical database
RCSI	Royal College of Surgeons in Ireland, Beaumont Hospital	Manual phenotyping
ULIV	University of Liverpool and Walton Centre NHS Foundation Trust	Manual phenotyping
Glasgow	Epilepsy Unit, Western Infirmary, Greater Glasgow & Clyde NHS trust	Manual phenotyping
Australia	The Royal Melbourne Hospital and the Austin Hospital in Victoria	Data transfer from existing clinical trial database (coordinated at University of Liverpool)
SANAD	University of Liverpool	Data transfer from existing clinical trial database

3.3.3. Data collection and phenotyping

Phenotyping was carried out separately at each centre. It was done by experienced phenotypers (clinical fellows or research nurses) supported as required by clinical supervisors. In rare cases when specific questions regarding an individual case arose, it was discussed within the relevant work package.

Identification of cases for inclusion in EpiPGX was in the hands of each centre. There was very little ongoing recruitment of patients during the project; cohorts mostly

consisted of historical data. Assembly of historical cohorts was influenced by the respective purpose and aims of the original studies.

3.3.3.1. Author's contribution to phenotyping.

Manual phenotyping of all cases included in the EpiPGx by the WP2 was undertaken by the author of this thesis.

It included both patients with newly diagnosed and established epilepsy. As a part of the project, cases with established epilepsy were phenotyped and the data were later shared with partners. Cases were phenotyped either in the Walton Centre in Liverpool or Epilepsy Research Unit based at the Western Infirmary in Glasgow.

Manual phenotyping involved all aspects of data capture and extraction. A variety of old research and clinical databases were used to trace and cross-link historical cases with current clinical records. To provide a maximal capture where required, old scanned records were used and notes were retrieved from the off-site storage. All available sources including clinical notes, letters and investigation results were reviewed and data were extracted. Following extraction, information was entered into an electronic database (eCRF).

In total, 1387 cases with newly diagnosed and established epilepsy were manually phenotyped (977 in Glasgow and 410 in Liverpool).

Due to the patchy information available for several cases from Australia, the proportion of cases required a significant manual input by cross checking and adjusting the transferred information. All data uploads and phenotype derivation were also manually cross-checked.

3.3.4. Phenotype definitions and their application

Phenotype definitions were created prior to my involvement in the project; hence I claim no authorship of them. Full consortium phenotype definitions (EpiPGX, 2014) are available in Appendix 2, page 222.

Newly diagnosed epilepsy was defined as:

"Occurrence of ≥2 clinically definite unprovoked epileptic seizures in the previous year *or* the occurrence of one seizure and the clinician decides to start AEDs. Prospective data are preferred, but retrospective data are allowed if based on contemporary evidence (i.e. continuous records from initiation of the first AED onwards at a specialist epilepsy centre). Patients with known progressive neurological disorders at the time of first AED initiation are excluded; patients with prior AED exposure or rescue treatment should be noted and the indication recorded."

Epilepsy syndromes were defined according to ILAE 1989 Classification (Epilepsy, 1989).

For treatment outcome with the first well tolerated AED the following definitions were used (EpiPGX, 2014):

"**Remission** was defined as any continuous period of ≥12 months complete seizure freedom." It was separated into two subcategories:

- **Immediate** – occurring within 14 months of starting the first well-tolerated AED.
- **Deferred** – occurring later than 14 months after starting the first well-tolerated AED (and prior to initiation of another AED).

No remission was defined "as continuing seizures after starting the first *well-tolerated*, *adequately applied* and *appropriate* AED".

An AED was considered adequate if subjects were exposed to a minimum therapeutic dose of that AED for a sufficiently long time period. A six month period was considered a sufficiently long trial period to establish a lack of response. An AED was considered appropriate if it had been demonstrated to be effective by past research, preferably by randomised controlled trials (e.g. ethosuximide for focal seizures would be considered inappropriate). Table 4 summarises the most commonly used AEDs, appropriate seizure types, and the minimal therapeutic and World Health Organisation defined daily dose

for monotherapy in adult patients according to the EpiPGX phenotyping definitions (EpiPGX, 2014). The decision regarding whether or not the AED trial was adequate and appropriate was made by a phenotyper and recorded in eCRF.

Table 4. AEDs, appropriate seizure types, minimum and defined daily doses for AED in adults according to EpiPGX phenotyping definitions (EpiPGX, 2014).

Antiepileptic drug	Focal seizures	Primary generalised tonic-clonic seizures	Absence seizures	Other primary generalised / unclassified seizures	Minimum therapeutic dose (mg)	Defined daily dose (mg)
Carbamazepine	✓	✓	×	×	600	1000
Clobazam	✓	✓	✓	✓	10	20
Clonazepam	✓	✓	✓	✓	4	8
Eslicarbazepine	✓	✓	×	×	800	800
Ethosuximide	×	×	✓	×	1000	1250
Felbamate	✓	✓	✓	✓	1200	2400
Gabapentin	✓	×	×	×	1200	1800
Lacosamide	✓	×	×	×	200	300
Lamotrigine	✓	✓	✓	✓	150	300
Levetiracetam	✓	✓	✓	✓	1000	1500
Oxcarbazepine	✓	✓	×	×	900	1050
Phenobarbital	✓	✓	×	✓	60	100
Phenytoin	✓	✓	×	×	200	300
Pregabalin	✓	×	×	×	300	300
Primidone	✓	✓	×	✓	750	1250
Tiagabine	✓	×	×	×	30	30
Topiramate	✓	✓	✓	✓	100	300
Valproate	✓	✓	✓	✓	1000	1500
Vigabatrin	✓	×	×	×	1000	2000
Zonisamide	✓	✓	✓	✓	150	200

EpiPGX WP2 was separated into three tasks:

- The main aim for task one was to investigate the genome based biomarkers associated with 12 month remission of newly diagnosed epilepsy after application of the first well-tolerated anticonvulsant. The treatment outcome was split into three categories: immediate remission, deferred remission, and no remission. Case control and an innovative three-way competing risk survival model was used for analysis.

- Task two, on the other hand, investigated genome-based biomarkers for general versus selective drug responsiveness (i.e. patients who respond to any AED versus patients who require a specific AED to achieve seizure control). To achieve this aim, WP2 collaborated with WP4 which investigated genome-based biomarkers for late response.

- Task three investigated genome-based biomarkers associated with the first treatment failure. Treatment failure is categorised as either being due to lack of efficacy or adverse drug reactions. Treatment failure due to lack of efficacy requires appropriate and adequate exposure to the first AED. Failure due to adverse drug reactions were further split into either ‘on-target’ and ‘off-target’, depending on whether they were considered to be directly caused by the drug’s principal pharmacological mechanism of action.

To standardise clinical data collection, a special data entry manual was created (EpiPGX, 2013). It contained a section on general data collection and separate sections relevant for each work package. Creation of the manual was led by work-package 1 and three, but each subsection was written by the relevant work package. It was made available to all members of the consortium. Inclusion of general terms and definitions in addition to specific phenotypes ensured unified understanding and application of phenotypes and outcome definitions. Flowcharts related to newly diagnosed epilepsy and WP2 tasks are shown in Figures 2, 3 and 4. Figure 2 illustrates the general path for assessment of inclusion of the subject into WP2, requiring confirmation that the case is newly diagnosed and ensuring that data regarding the first anticonvulsant exposure is completely captured. Figure 3 is a flowchart illustrating the decision tree specifically for task one, evaluating the treatment outcome for the first well-tolerated anticonvulsant. This flowchart clearly underlines that the first well-tolerated anticonvulsant is not always the first applied drug. Figure 4 is a flowchart illustrating the decision tree for task 3, which was specifically designed to separate causes of failure of first applied anticonvulsant.

Flow charts were also later utilised in the process of development of code for phenotype derivation which was based on raw data.

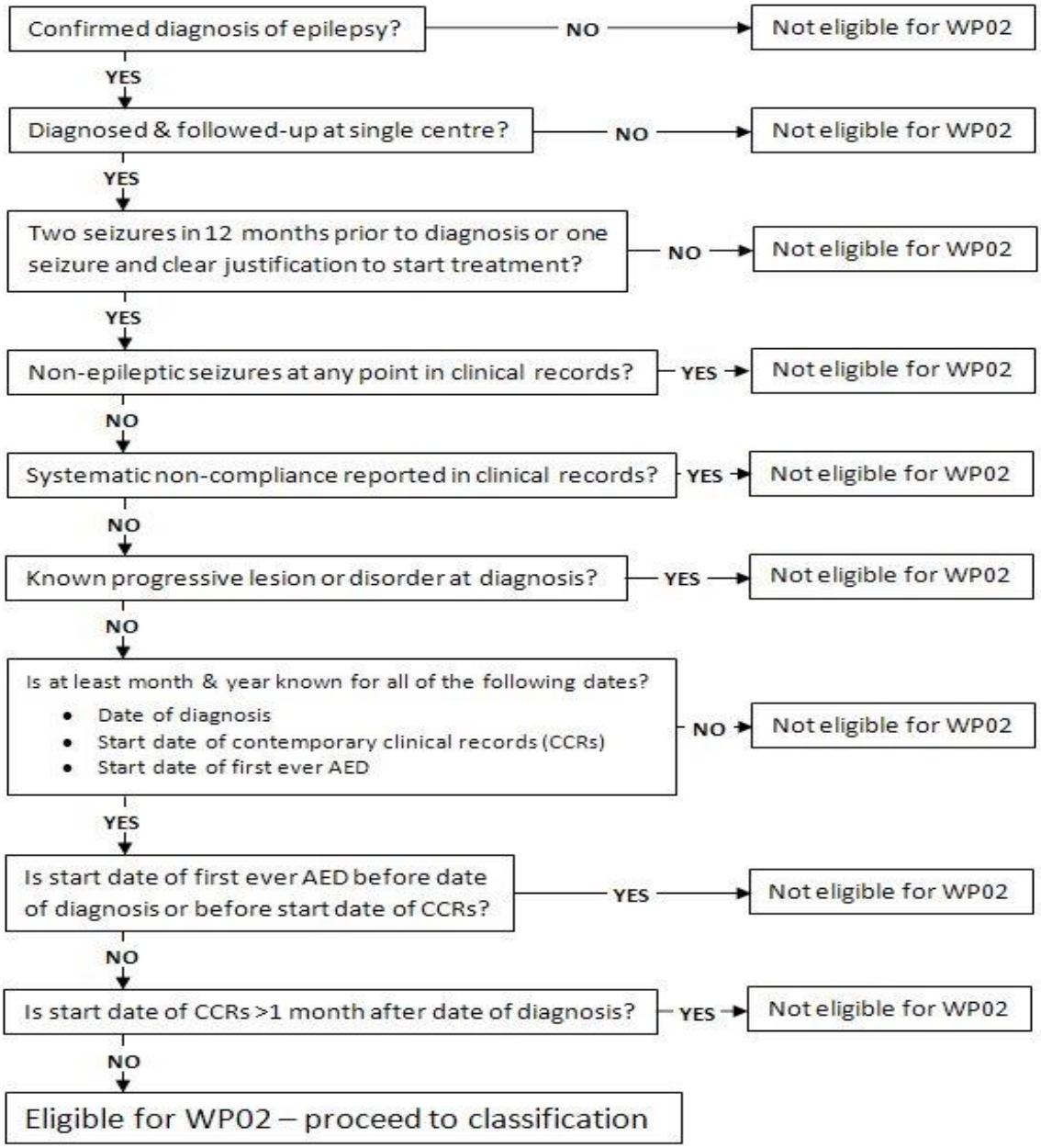


Figure 2. Flowchart to assist eligibility decisions for WP2; according to Data Entry Manual (EpiPGX, 2013)

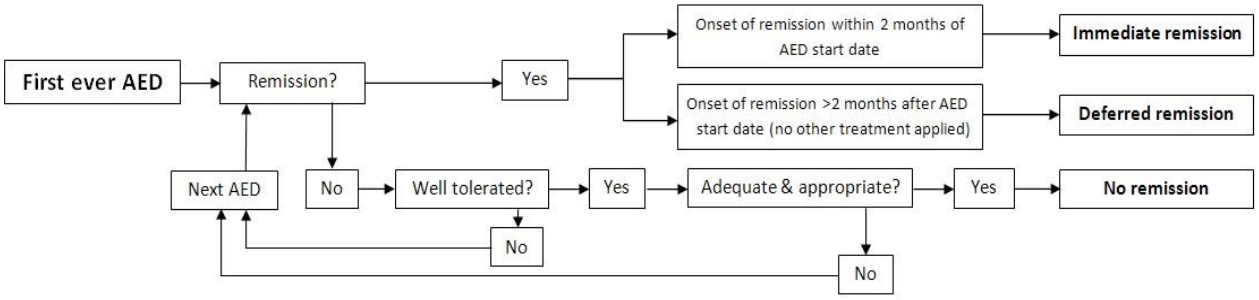


Figure 3. Flowchart to assess efficacy outcome with first well-tolerated anticonvulsant in newly diagnosed epilepsy used in WP2, Task 1; according to Data Entry Manual (EpiPGX, 2013)

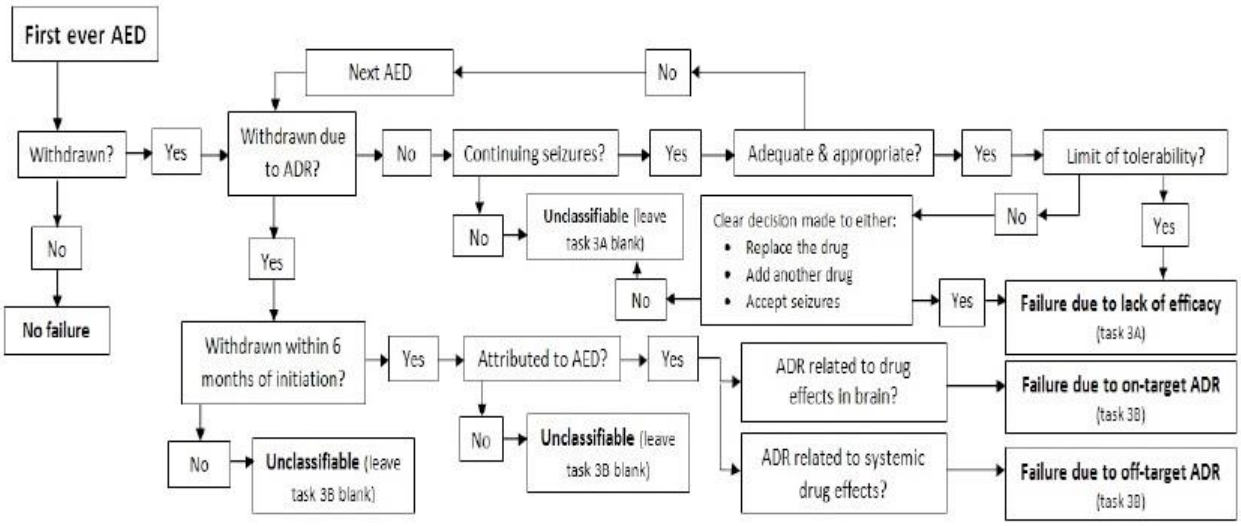


Figure 4. Flowchart illustrating decision tree to assess outcome after application of first ever anticonvulsant in newly diagnosed epilepsy used in WP2, Task 3; according to Data Entry Manual (EpiPGX, 2013)

3.3.4. Data transfer

Data for the following research cohorts were transferred from their respective original databases to the EpiPGX eCRF:

- SANAD
- Melbourne
- ULB

The SANAD and Melbourne transfers were carried out in Liverpool, whereas ULB was carried out locally in Brussels. Hence, a transfer code for the ULB cohort is not available; it was carried out as a collaboration between the local team in Brussels and Roland Krause from LCSB. Dr Ben Francis (Department of Biostatistics) was instrumental for all data uploads carried out in Liverpool.

For the SANAD and Melbourne transfers, separate dictionaries linking the original CRF with the EpiPGX paper CRF were created. All free text entries were manually reviewed and matched accordingly. Afterwards, the data were split into small clusters and batch transferred to the eCRF. Initially, the transferred data were uploaded to a test site, reviewed, and extensively tested for errors and discrepancies. After review, the data were moved to the permanent eCRF, where all cases were manually reviewed and treatment outcomes were classified manually according to EpiPGX definitions. During this process, treatment outcomes were cross checked against the original database.

The data transfer dictionary for the SANAD study has been provided in Appendix 3, page 227. SANAD patients with previous treatment were highlighted, which allowed their exclusion from analyses of newly diagnosed epilepsy. Pre-treatment seizure count in the SANAD study was provided in absolute numbers and was transferred accordingly to EpiPGX. Tables from 5 to 7 illustrate the approach used to match seizure types, epileptic syndromes and EEG results between cohorts. The SANAD study obtained additional, more detailed, neuroradiological data as well as follow up data from general practitioners after finishing the study. This data was incorporated into the data transfer. With regard to treatment outcome, the SANAD study collected data on the number of seizures between follow up visits and individual outcomes for each AED, which allowed a relatively straightforward calculation of onset of 12 month remission. Remission was assumed if there was either at least one 12-month period between recorded seizures while on treatment, or only an index seizure and no further seizures recorded. Onset of 12 month remission was set for the day after the most recent seizure for remissions on treatment, or the first day of treatment if there was immediate remission following the

onset of treatment. Reasons of treatment failure were recorded in the SANAD study and were also transferred to EpiPGX.

Table 5. Matching of seizure types between SANAD study and EpiPGX.

SANAD Seizure type	EpiPGX seizure type
Typical Absence	Absence
Atypical Absence	Absence
Febrile convulsions	Febrile
Simple Partial	Simple Partial
Complex Partial	Complex Partial
With Generalised Tonic-Clonic (partial)	Secondary GTCs
Myoclonic	Myoclonic
Tonic-Clonic (generalised)	Primary GTCs
Atonic	Atonic
Tonic-Clonic Uncertain	Unclassified GTCs
Other (free text entry, single isolated case)	Manual review of free text entry

The Melbourne cohort was based on a prospective observational study. Data organisation for it was completely different from the SANAD study, hence it was processed separately. In this cohort there were no cases with progressive neurological conditions other than brain tumours and these were manually filtered off. Information was also available regarding previous exposure to AEDs before recruitment and when those AEDs were started. Cases in which an AED was started more than sixty days before enrolment were excluded, as they were not considered newly diagnosed. Epilepsy syndrome classification was significantly different from that employed in EpiPGX, hence all cases were manually matched to EpiPGX definitions using all available covariates. Where matching was not possible, cases were recorded as unclassified. Follow up information was collected at three months, one year, and two years. The pre-treatment seizure count in the Australian cohort contained the following categories: 1, 2, 3, 4, 5 and more than 5 seizures before starting the treatment (Speed et al., 2014a). Categories

were matched accordingly where possible during data transfer. As there was no relevant category for 5 and more, it was matched with the unknown category in EpiPGX. In SANAD, the pre-treatment seizure count was collected as an absolute number, hence when the data were categorised, there were no cases in the unknown category. Subjects were considered in remission if there were no recorded seizures for a period covering at least 12 months.

Table 6. Matching of epilepsy syndromes between SANAD study and EpiPGX. EpiPGX had a three level classification of epileptic syndrome, with the first level indicating if the syndrome was generalised, localisation related or unclassified; whereas the third level described the specific syndrome. Some syndromes are omitted from this table as they were not represented in the SANAD study.

SANAD	**EpiPGX**		
Benign Childhood Epilepsy With Centrotemporal Spikes	Localisation Related	Idiopathic	Benign Childhood Epilepsy With Centrotemporal Spikes
Childhood Epilepsy With Centrotemporal Spikes	Localisation Related	Idiopathic	Other
Childhood Absence	Generalised	Idiopathic Age Related	Childhood Absence Epilepsy
Juvenile Absence	Generalised	Idiopathic Age Related	Juvenile Absence Epilepsy
Juvenile Myoclonic	Generalised	Idiopathic Age Related	Juvenile Myoclonic Epilepsy
Epilepsy With Tonic-Clonic Seizures on Awakening	Generalised	Idiopathic Age Related	Epilepsy with GTCs on Awakening
Generalised Not Specified	Generalised	Unclassified	
Unclassified	Unclassified		

Table 7. Matching of EEG between SANAD study and EpiPGX

SANAD EEG		EpiPGX EEG
Abnormal NO		Normal
Abnormal YES	Non Specific	Abnormal non specific
Abnormal YES	Generalised Activity Slow Wave Spiking	Abnormal Epileptiform
Abnormal YES	Generalised Activity Slow without spiking	Abnormal Non-specific
Abnormal YES	Focal Activity Paroxysmal with Spiking	Abnormal Epileptiform
Abnormal YES	Focal Activity Paroxysmal without Spiking	Abnormal Non Specific

The data for both cohorts were initially uploaded into a trial mode and were extensively checked for any inconsistencies and systematic errors, and afterwards were moved to a permanent database. Those checks were carried out by myself.

3.4. Statistical and genetic analysis

3.4.1. Phenotype derivation

Prior to analysis, all newly diagnosed cases were filtered and then downloaded from the EpiPGX eCRF. Raw data were processed and a phenotype derivation produced. This was carried out by Dr Sarah Langley (Imperial College London), with clinical input and quality control undertaken by myself.

Seizure types in EpiPGX were classified according to the ILAE 1981 classification (Bancaud et al., 1981).

During the phenotype derivation to ensure consistency with previous prognostic work carried out on the SANAD study (Bonnett et al., 2014b, Bonnett et al., 2012), seizure types were transformed accordingly. Matching of the seizure types is summarised in table 8. In the EpiPGx eCRF, the pre-treatment seizure count was collected separately for GTC, non-GTC and combined (unknown) seizure types. It was entered separately in

each of these categories either as an absolute or categorical value (including unknown), but not as both. This created a mixture of categorical and absolute values and rendered the creation of continuous values impossible. Hence, in order to avoid losing cases with categorical values, the data were categorised.

Table 8. Transformation of seizure type to match earlier prognostic work on the SANAD study.

Seizure type according earlier SANAD prognostic work (Bonnett et al., 2014b, Bonnett et al., 2012)	**Matched seizure types according to EpiPGX**
Simple or complex partial only	Simple partial or complex partial seizures without any others
Secondary generalised tonic-clonic	Secondarily generalised tonic-clonic seizures
Generalised tonic-clonic seizures only	Secondarily generalised tonic-clonic, primary generalised tonic-clonic or unclassified generalised tonic-clonic seizures without any other
Absence seizures	Absence seizures
Myoclonic or absence seizures with tonic-clonic seizures	Absence or myoclonic with primary generalised tonic-clonic or unclassified generalised tonic-clonic seizures
Unclassified tonic-clonic seizures	Unclassified generalised tonic-clonic seizures
Other/Uncertain	Other/Uncertain.

Epilepsy types were divided into three categories, either focal, generalized or unclassified.

The results of the imaging studies were coded either as normal, abnormal focal, abnormal nonspecific or not done/not known. Similar coding was used for EEG results where abnormal was replaced with either epileptiform or non-specific abnormalities. Results of the first investigation were used.

A positive family history was defined as the documented presence of any relative with epilepsy. The EpiPGX eCRF coded family history as a variable with three potential values – present, absent or unknown. In the phenotype derivation variables were dichotomised by collapsing together unknown or absent categories.

The outcome of the neurological examination in eCRF was coded as either normal, abnormal, not applicable or unknown. Similar to the other co-variates in the phenotype derivation, it was dichotomised. Two categories were created: abnormal and normal/unknown/not applicable.

Data collection on the presence of learning disabilities (LDs) began mid-way through the project; hence some of the cases phenotyped during the first half of the project lacked this information. To create a binary variable category for phenotype derivation, “yes” and “not known” were collapsed together.

The value of the variable time from first seizure to first AED was calculated in months by subtracting the date of the first seizure from the date of the first AED. Hence when the date of the first seizure was unknown, the value of the variable was coded as missing.

Follow up length was calculated in years by subtracting the date of the first visit from the date of the last visit.

3.4.2. Statistical analysis of clinical factors associated with treatment outcome

To ensure consistency between all analyses conducted during the EpiPGX project, a single phenotype derivation was used for statistical analysis of associations with treatment outcomes. Descriptive statistical methods were initially applied to explore the data. For analysis of association, logistic regression and Cox regression were used. A prior univariate analysis was performed with each factor included in the initial logistic and Cox regression models. The Bonferroni method was applied for corrections of multiple testing where there were multiple tests for the same hypothesis.

The logistic regression used backward selection using Wald statistics. The probability for stepwise entry was set at 0.05 and for removal 0.10. Results were plotted using forest plot as log transformation of Exp (B) value.

For the Cox regression, time to 12 month remission was calculated by subtracting the start date of the first well-tolerated AED from the first day of onset of the corresponding 12 month remission. This approach was different from the SANAD study where time to remission was calculated by subtracting the date of randomisation from the first date after continuous seizure freedom. Time to 12 month remission calculation is illustrated in figure 5.

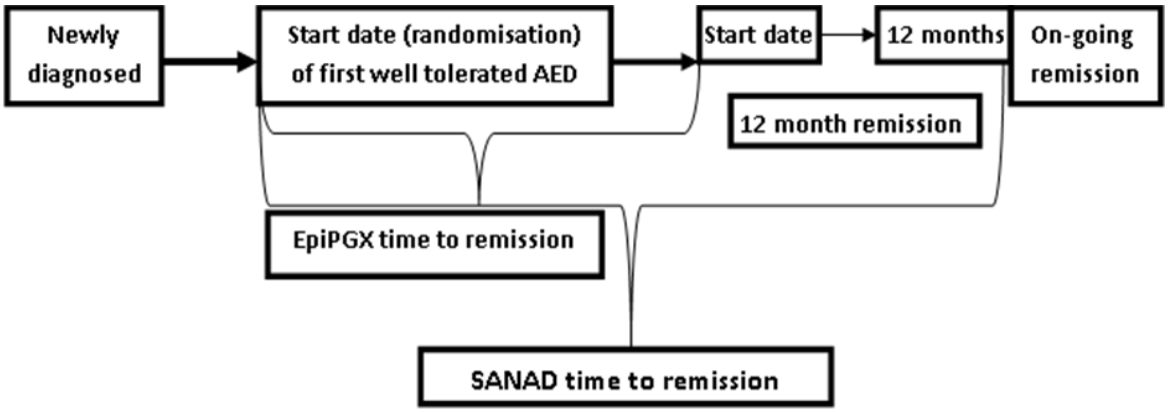


Figure 5. Flowchart illustrating the calculation of time to 12 month remission in EpiPGX and for comparison in the SANAD study.

Patients who did not experience remission with the first well-tolerated anticonvulsant, or had it withdrawn due to ADRs, were right-censored. For the survival analysis stepwise variable selection, Wald statistics were used; probability for stepwise entry was set at 0.05 and for removal at 0.10.

EpiPGX GWAS analyses were later adjusted for factors significantly associated with 12 months' remission.

3.4.3. Assessment of impact of study methodology

The impact of study methodology and methods of cohort assembly were assessed separately. The EpiPGX project is a multinational research consortium, hence effect of origin of cohort was explored in some detail. Additionally, as the EpiPGX research cohort was assembled from a mixture of old historical datasets, the effect of original mode of case ascertainment was explored. Impacts of various study parameters (duration of observation, definition of remission) were also separately assessed.

The definition of remission for our study required at least a continuous 12-month period of freedom from seizures. To assess how changing the length of the defined seizure freedom period affects the proportion of patients labelled as being in remission, the proportions of patients counted as being in remission after applying different length criteria of seizure freedom were compared. The one, two, five or 10-year periods of seizure freedom were separately applied to the patient cohort created from patients with at least equal follow up length to applied remission period. A separate group of remission lasting for a whole period of observation was also created. It was defined as seizure freedom up to the last visit, irrespective of how long the follow-up period was.

Association of the follow up length, origin of cohort, and mode of case ascertainment with the binary treatment outcome was assessed with a logistic regression. The same parameters were used as for the analysis of clinical factors. The digital data transfer process from existing clinical databases has the potential to introduce systematic error and distortion of data at various levels, hence it was also separately assessed. The genetic and statistical analysis performed by the SANAD study and EpiPGX utilized data derivations rather than the original raw data. Hence, to assess how the transformation and uploading of data affected their quality, SANAD cases from the EpiPGX phenotype derivation were compared with SANAD cases from the phenotype derivation created for an earlier genetic study (Speed et al., 2014a). As the SANAD study used different variables and definitions, the co-variates underwent some degree of transformation and were then uploaded to the central EpiPGx repository. To assess how this process affected them, the concordance of the basic demographic and most important co-variates was investigated.

Phenotype derivation for the EpiPGx study was created by selecting newly diagnosed cases from the eCRF and then extracting information in accordance with the task 1 definition. Similarly, the SANAD genetic dataset was based on the SANAD study and was created for an earlier GWAS on newly diagnosed epilepsy (Speed et al., 2014a). The inclusion criteria in SANAD differed slightly from the EpiPGx inclusion criteria. The SANAD study permitted some cases with previous inadequate and inappropriate exposure to AED or relapses after long-term remission. On the other hand, EpiPGX WP2 recruited only cases of newly diagnosed epilepsy. Cases with previous AED exposure in the SANAD genetic data set were filtered off based on whether there existed any history of previous monotherapy.

For exploration of the impact of study methodology and data transfer, a variety of descriptive statistics were used including contingency tables, Cohen's kappa coefficients for categorical variables, and Lin's concordance correlation coefficients for continuous variables.

For statistical data processing, MS Excel 2010 and IBM SPSS version 22 were used. Lin's concordance correlation coefficient was calculated using an internet calculator (http://services.niwa.co.nz/services/statistical/concordance).

3.4.4. Genotyping

There were two types of genetic data; either historical GWAS data which were genotyped as a part of previous studies, or *de novo* GWAS data genotyped at deCODE as part of EpiPGX.

In the majority of cases, DNA samples had been collected previously and extracted from either from blood or saliva.

Genotyping of samples sent to deCODE was performed on Illumina OmniExpress-12 v1.1 and -24 v1.1 single nucleotide polymorphism (SNP) arrays. BeadStudio (Illumina;

version 2.0) was used to call genotypes, normalise signal intensity data and establish the log R ratio and B allele frequency at every SNP.

The historical UK samples (SANAD and ULIV cohorts) were genotyped at the Wellcome Trust Sanger Institute on the Illumina 660 chip. The Melbourne samples were also genotyped using the same platform at the same institution, but at a different time. Sample QC criteria were the same for both the SANAD, ULIV and Melbourne cohorts and have been described previously (Speed et al., 2014a).

3.4.5. GWAS QC and imputation

All GWAS QC and imputation for EpiPGX was performed by deCODE (courtesy of Dr Andrés Ingason). QC involved several steps. Initially all markers with a very high (> 0.9) rate of missing genotypes were removed. As there were samples from multiple European countries, the cohorts were then split according to origin. A Hardy-Weinberg equilibrium (HWE) was calculated for each group separately. All samples with < 0.98 genotype rate and all markers with < 0.95 genotype rate, < 0.01 minor allele frequency (MAF), or $P < 1 \times 10^{-6}$ for HWE in any of the sample subgroups were removed. In the next QC step using a window size of 100 markers shifting by 25 markers at a time and removing one half of every SNP pair with genotypic $r^2 > 0.1$ a subset of markers independent of each other with respect to the linkage disequilibrium (LD) was created. Then by using this subset of markers heterozygosity (HET), identity by state (IBS) and sex was calculated. Samples with outlying HET values (> 5 standard deviations (SD) from the median of the whole sample); one half of all sample pairs with pihat > 0.9 IBS, and c) and all samples where sex determined from genotype did not match reported gender were removed. At the next step array-specific maps were retrieved from the website of Will Rayner at the Wellcome Trust (http://www.well.ox.ac.uk/~wrayner/strand/) and were used to update all marker positions and chromosome numbers to the Genome Reference Consortium Human Build 37 (GRCh37). Then to avoid strand issues all A/T and C/G markers were removed.In the next step extraction of genotypes for 2,766 ethnicity-sensitive SNPs common to all Illumina SNP arrays (Supplement 1) were done and STRUCTURE 2.2 (Pritchard et al., 2000) were used to derive European, Asian and African ancestry probabilities,respective

samples of Hapmap Yoruba in Ibadan, Nigeria Japanese in Tokyo, Japan and Han Chinese in Beijing, China and Utah residents with ancestry from northern and western Europe were used as reference populations, and samples with less than 90% European ancestry were removed. Quality control of genotypes was done using PLINK 1.9 (Purcell et al., 2007).

Genotypes were split up according to chromosome arms (the X chromosome was additionally split into pseudo-autosomal regions (PAR) and non-PAR) further step involved creation of phased haplotypes using SHAPEIT v2 (O'Connell et al., 2014, Delaneau et al., 2014, Delaneau et al., 2013b, Delaneau et al., 2013a, Delaneau et al., 2012) The imputation of genotypes in our dataset were done by using IMPUTE version 2.3.0 (Howie et al., 2009, Howie et al., 2012).

Further processing was conducted by Dr Ben Francis; this involved the merging of all genetic datasets and retaining only common SNPs. The merged dataset was pruned down to low linkage disequilibrium SNPs and principal component analysis was carried out.

All GWAS were performed by EpiPGX statisticians according to a defined analysis plan. Analysis was carried out on a server based at LCSB. GWAS was performed both as univariate and after adjusting for clinically significant factors as determined by logistic regression.

As a part of my PhD project, I undertook a separate GWAS using all newly diagnosed cases with focal epilepsy for which genetic data was available locally. It was done locally on the University of Liverpool High Performance Computing Cluster with SNPtest version 2.4.1 (Marchini et al., 2007).

In a similar way to the original EpiPGX GWAS, it was performed both as a univariate analysis and after adjusting for clinically important factors. To determine the significant clinical factors associated with 12 month remission for the further adjustment of GWAS, a binary logistic regression using the same parameters as described earlier but only including those with focal epilepsy and locally stored DNA (patients from SANAD,

Glasgow, ULIV and the Australian cohort) were carried out. The following clinical covariates were included in the logistic regression model:

- Age at onset in years
- Neurological examination previously undertaken
- Febrile seizures previously experienced
- Positive family history of epilepsy
- EEG
- CT
- MRI
- Pre-treatment seizure count
- Seizure type classified using the same approach as for a full logistic regression model
- Origin of cohort (separate co-variate included to represent each different cohort: ULIV, GLA, SANAD, and AUS)
- Whether study was prospective or retrospective

For the purposes of the logistic regression, the following variables were set as a baseline: 1 – 2 seizures before treatment, and normal for EEG, MRI and CT. Pre-treatment seizure frequency in the Australian research cohort was coded as a categorical variable with values from one to five and more than five. As described earlier, during the transfer and uploading of the Australian data, the values were categorised according to the EpiPGX eCRF, hence there were no subjects in the categories with 6 – 10, 11 – 20 or 21 or more seizures before starting the treatment. This lead to p-values of -1 when GWAS was adjusted simultaneous for both pre-treatment seizure count and origin of cohort, rendering adjustment for both fators impossible. As the logistic regression model based on the full cohort had shown that the mode of case ascertainment (prospective vs. retrospective) was statistically significantly associated with the treatment outcome, it was decided to replace the origin of the cohort with the variable representing whether the follow-up was prospective or retrospective. Cases from the SANAD study and Australia were considered to be prospective, whereas the rest of the cohorts were assembled retrospectively.

The GWAS analysis was adjusted for the first five principal components (population stratification) and co-variates statistically significantly associated with treatment outcome from the multivariable logistic regression model.

Initial GWAS was carried out using an additive model, and significant SNPs were later retested using the dominant model of inheritance. The genomic inflation factor was calculated for GWAS with additive mode of inheritance for both the univariate model and after adjustment for significant clinical factors. Results were plotted as Manhattan and QQ plots. For plotting results of individual SNPs, LocusZoom software were used (Pruim et al., 2010).

4.0. Results

4.1. Descriptive statistics

For the purpose of describing the research cohort, data on all of the patients included in work package 2 task 1 were used (n= 1906). Due to missing data, fewer cases (n=1723) were included in the logistic and Cox regression models. A full flowchart detailing the missing information is presented in Figure 6.

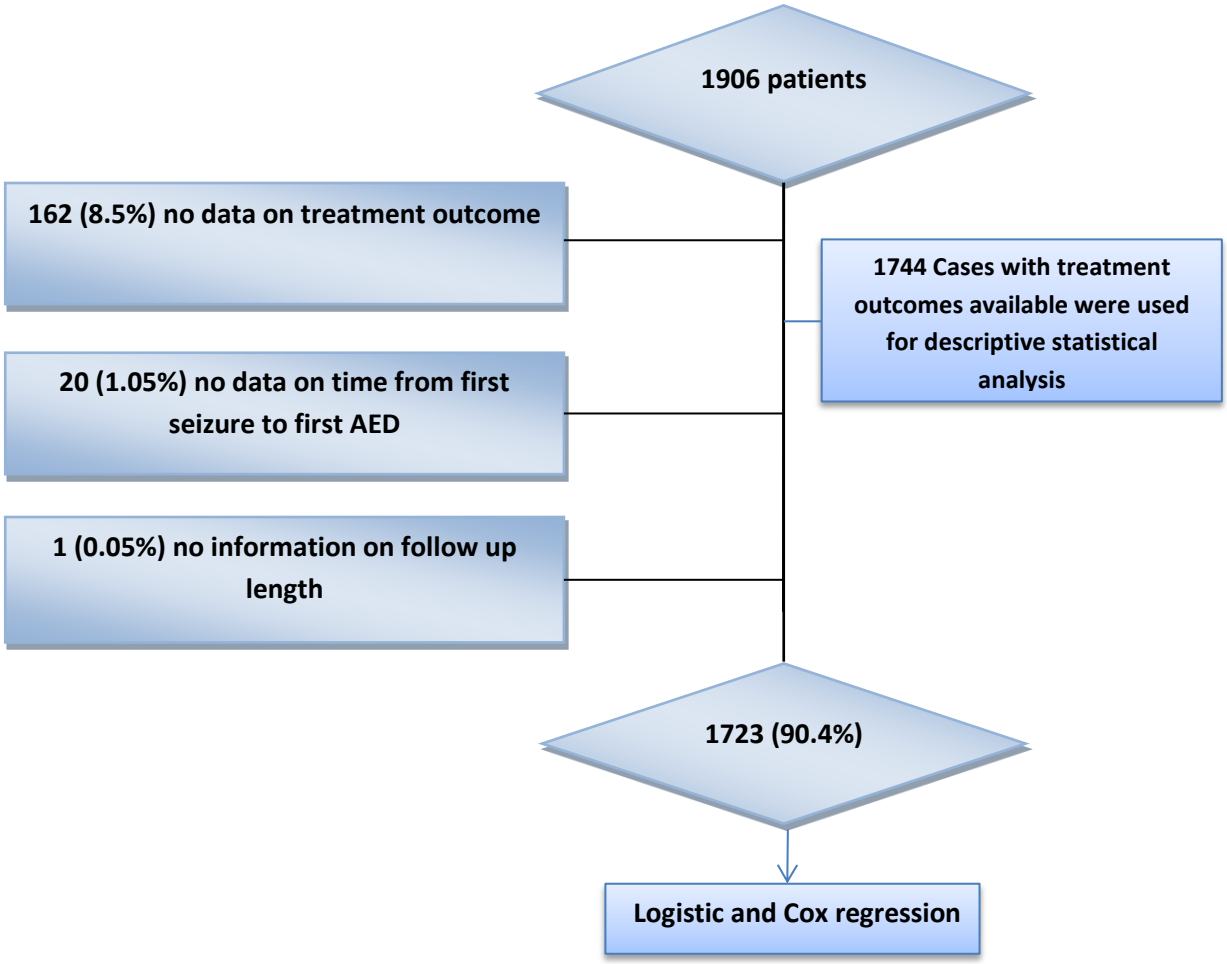


Figure 6. Flowchart detailing case attrition due to missing data. 1906 patients were included in the study and 1723 (90.4%) were analysed for a logistic and Cox regression, whereas 1744 were used for the descriptive statistical analysis.

For consistency with the GWAS and the rest of the analysis carried out as a part of the EpiPGX project, the phenotype derivation carried out by Dr Sarah R Langley were

used for all analysis unless stated otherwise. The phenotype derivation was created using a full data set; it captured phenotypes which were constructed before the onset of data collection using agreed definitions.

4.1.1. Remission

Remission was defined as any continuous period completely free of seizures, equal to 12 months or longer. No remission was defined as on-going seizures after starting the first well-tolerated, adequately applied and appropriate AED. Binary outcomes (remission vs. no remission) in newly diagnosed epilepsy cases were available on 1744 subjects (91.5%).The 12 month remission was observed in 982 of these (56.3%).

The proportion of subjects experiencing 12 month remission was different between cohorts. The Australian cohort had the highest proportion of remissions (n = 124; 72.5%), followed by SANAD (n = 439; 58.7%), Glasgow (n = 238; 57.3%) and EKUT (n = 133; 56.6%). The no remission rate was higher in the ULIV (n = 45; 84.9%), UCL (n = 27; 81.8%) and ULB (n = 13; 76.5%) cohorts. A Pearson's chi-square test showed statistically significant differences ($p < 0.001$) between cohorts in terms of the proportion of patients achieving 12 months' remission with the first well-tolerated AED. Details of patient outcome, stratified in detail by cohort, are summarised in Table 9.

In addition to the occurrence of remission, the time taken to achieve remission (time to start of 12 month remission) was separately assessed. From the 982 patients experiencing remission after the application of the first anticonvulsant, 699 (71.2%) had immediate remission, while the rest (n = 283, 28.8%) experienced deferred remission (onset of remission more than two months after the application of the first well-tolerated anticonvulsant). The data did not follow a normal distribution, but was skewed towards the left. A zero value for time to remission indicates an immediate onset of remission. The median time to a 12 months' period of remission was 0.0 days, the mean = 126, SD = 389 and IQR = 0 – 90. The relatively large difference (126 days) between the observed mean and median indicates the presence of outliers with a late onset of remission. On comparing time to 12 month remission between cohorts using a Kaplan-

Meier approach there was a statistically significant difference (p<0.001). In the majority of cohorts (ULB, Glasgow, Australia, SANAD) median time for the onset of 12 month remission was zero, whereas RCSI had the longest median time of 152 days. Further details of time to onset of 12 month remission stratified by cohort are presented in Table 10 and in Figures 7 and 8.

Table 9. Distribution of binary treatment outcomes (remission vs. no remission) in newly diagnosed epilepsy cases stratified by the cohort (n = 1744)

Cohort	12 month remission with the first AED					
	No remission		Remission		Total	
	n	%	n	%	n	%
UCL	27	81.8%	6	18.2%	33	1.9%
ULB	13	76.5%	4	23.5%	17	1.0%
EKUT	102	43.4%	133	56.6%	235	13.5%
RCSI	42	58.3%	30	41.7%	72	4.1%
Glasgow	177	42.7%	238	57.3%	415	23.8%
ULIV	45	84.9%	8	15.1%	53	3.0%
Australia	47	27.5%	124	72.5%	171	9.8%
SANAD	309	41.3%	439	58.7%	748	42.9%
Entire Cohort	762	43.7%	982	56.3%	1744	100.0%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

Table 10. Time to onset of 12 month remission in days (mean, median, SD and IQR) after application of first well tolerated anticonvulsant stratified by cohort (n = 982)

Cohort	Time to onset of 12 month remission (days)			
	Mean	Standard deviation	Median	IQR
UCL	24	42	8	0 – 19
ULB	5	10	0	0 – 10
EKUT	64	142	3	0 – 35
RCSI	1074	1599	152	0 – 1564
Glasgow	111	268	0	0 – 107
ULIV	68	128	1	0 – 100
Australia	29	79	0	0 – 29
SANAD	120	239	0	0 – 139
Entire Cohort	126	389	0	0 – 90

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

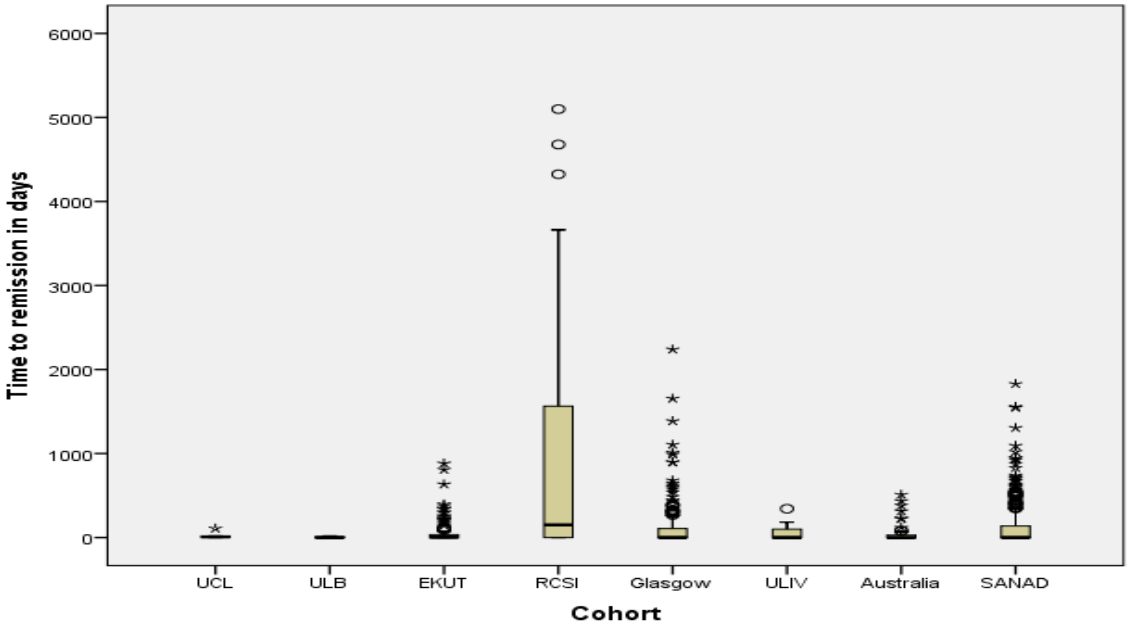


Figure 7. Box plot for time to onset of 12 month remission after application of first well tolerated anticonvulsant (n = 982). Distribution of time is shown in days separately for each cohort. The middle line indicates median, whereas boundaries indicates 75th and 25th percentile. The whiskers represent the lowest and highest non-outlier value. *UCL* – University College London; *ULB* - Hôpital Erasme, Université Libre de Bruxelles; *EKUT* - Eberhard-Karls-Universität Tübingen; *RCSI* - Royal College of Surgeons in Ireland; University of Liverpool; *SANAD* - Standard and New Antiepileptic Drugs Study.

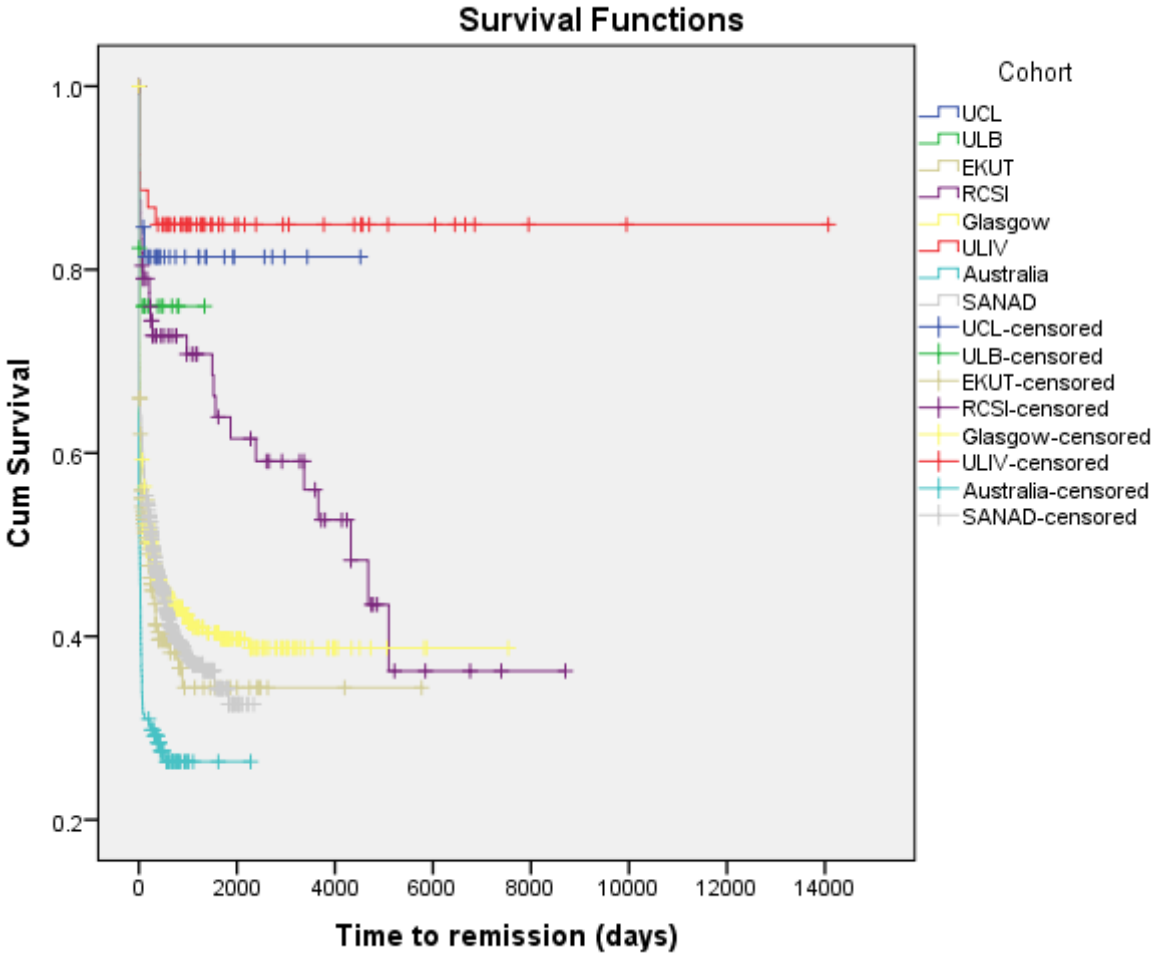


Figure 8. Kaplan-Meier plot for time to 12 month remission (days) in newly diagnosed epilepsy cases after application of the first well-tolerated anticonvulsant stratified by cohort (n = 1744). *UCL* – University College London; *ULB* - Hôpital Erasme, Université Libre de Bruxelles; *EKUT* - Eberhard-Karls-Universität Tübingen; *RCSI* - Royal College of Surgeons in Ireland; University of Liverpool; *SANAD* - Standard and New Antiepileptic Drugs Study.

4.1.2. Age at diagnosis

Age at diagnosis was available for all patients (n = 1906). The median age at diagnosis was 33.57 (IQR 19.92 - 50.27), and the data was positively skewed, with a single peak around adolescence (Figure 3). The cohort from Australia had the highest median age at diagnosis (38.5 years), whereas the EKUT cohort had the lowest (31.6 years). The data are further summarised in Table 11 and Figure 9. Only 13.3% (n = 252) of the subjects included in the study were aged 16 or younger, hence our study is more representative of the adult population. The majority (72.5%) of the patients included were diagnosed

between 2000 and 2010, with only 23.4% being diagnosed earlier than 2000. The proportion of the cases diagnosed with epilepsy per decade, is presented in Table 12.

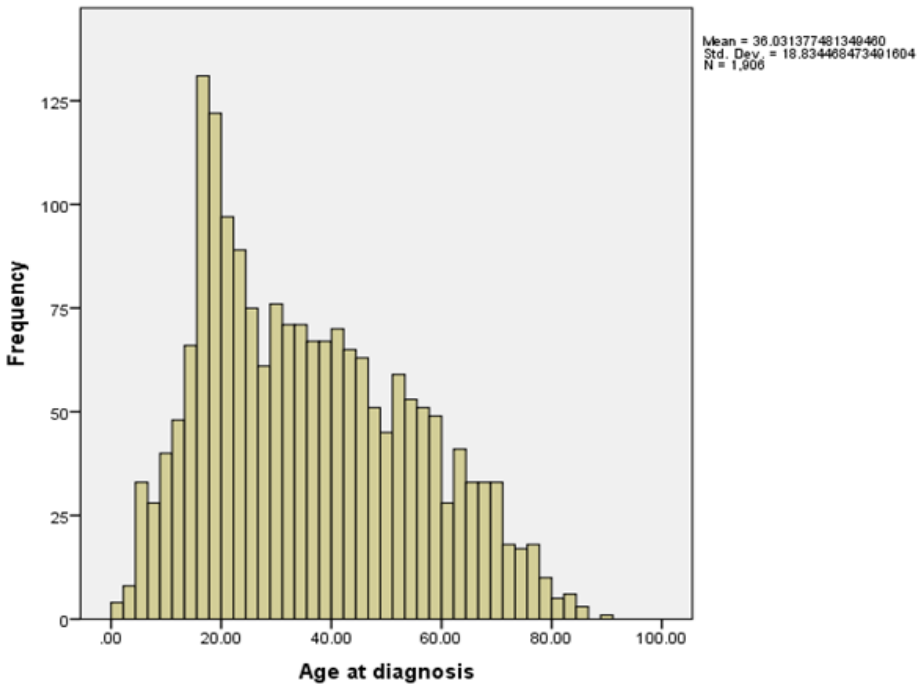


Figure 9. Histogram of distribution of age at diagnosis for patients with newly diagnosed epilepsy (n = 1906)

Table 11. Age at diagnosis of patients with newly diagnosed epilepsy (n = 1906) stratified by cohort

Cohort	Age at diagnosis (years)				
	Number	Mean	Standard deviation	Median	IQR
UCL	36	30.3	15.5	24.0	19.1 – 36.2
ULB	29	40.2	22.0	39.1	23.2 – 53.1
EKUT	247	23.4	17.0	16.8	11.9 – 31.1
RCSI	73	31.2	15.6	25.2	19.4 – 39.6
Glasgow	462	38.0	17.4	36.0	22.3 – 51.6
ULIV	53	32.8	13.6	31.6	20.2 – 45.1
Australia	228	40.6	18.9	38.5	23.6 – 53.9
SANAD	778	38.3	19.0	36.6	21.7 – 52.9
Entire Cohort	1906	36.0	18.8	33.6	19.9 – 50.3

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

Table 12. Period of diagnosis stratified in steps of 10 years (n = 1906)

Period of diagnosis (10 years)		
Period	**Frequency**	**%**
1960 – 1969	1	0.1
1970 – 1979	3	0.2
1980 – 1989	24	1.3
1990 – 1999	416	21.8
2000 – 2009	1382	72.5
2010 – current	80	4.2
Total	1906	100.0

4.1.3. Gender

Data on gender were available for all subjects (n=1906, 100%). There were marginally more males than females, with 989 males (51.9%) and 917 females (48.1%). The data are further summarised by cohort in Table 13 and Figure 10.

Cohorts from UCL (n = 14; 38.9%) and ULIV (n = 22; 41.5%) had the lowest proportion of males whereas Glasgow (n = 264; 57.1%) and ULB (n = 16; 55.2%) had the highest proportion of males. Gender distribution between the cohorts was uneven and a further Chi Square test confirmed a statistically significant difference in the distribution of gender between cohorts (p = 0. 008).

Additionally, stratification according to epilepsy type showed that focal epilepsy (males = 52.1%; females = 47.9%) and unclassified epilepsy (males = 60.7%; females = 39.3%) had a higher proportion of males than females, whilst the opposite was observed for generalised epilepsy (males = 45.5%; females = 54.5%).

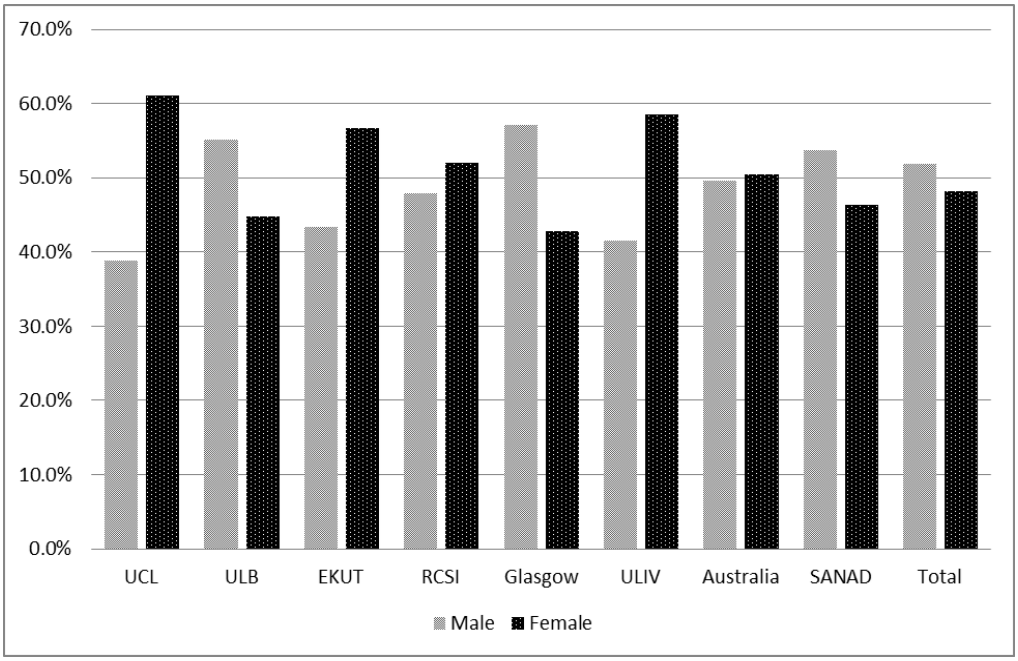


Figure 10. Gender of patients with newly diagnosed epilepsy stratified by cohort (n = 1906). *UCL* – University College London; *ULB* - Hôpital Erasme, Université Libre de Bruxelles; *EKUT* - Eberhard-Karls-Universität Tübingen; *RCSI* - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

Table 13. Gender of patients with newly diagnosed epilepsy stratified by cohort

Cohort	Gender			
	Male		Female	
	n	%	n	%
UCL	14	38.9%	22	61.1%
ULB	16	55.2%	13	44.8%
EKUT	107	43.3%	140	56.7%
RCSI	35	47.9%	38	52.1%
Glasgow	264	57.1%	198	42.9%
ULIV	22	41.5%	31	58.5%
Australia	113	49.6%	115	50.4%
SANAD	418	53.7%	360	46.3%
Entire Cohort	989	51.9%	917	48.1%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

4.1.4. Epilepsy type

Epilepsy type was categorised into three groups: focal (n = 1346; 70.6%), generalised (n = 341; 17.9%) or unclassified (n = 219; 11.5%). Due to the small numbers, undetermined syndromes with focal and generalised features (n = 3) were collapsed together with unclassified epilepsy. Table 14 summarises the distribution of epilepsy type according to the cohort. For the purpose of analysis, the SANAD arms A and B were collapsed together, however in order to better illustrate the distribution of epilepsy types between cohorts they are separated in Table 14. The SANAD arm A recruited only focal epilepsy, whereas arm B recruited generalised and unclassified cases. In the SANAD study, the initial assignment to arm A or B was based on initial syndromic diagnosis, but the epilepsy classifications were reviewed during the study as required. Hence there are 10 cases of focal epilepsy in arm B and 53 unclassified and 5 generalised epilepsy cases in arm A.

Table 14. Distribution of epilepsy type (focal, generalised and unclassified) stratified by cohort (n= 1906).

Cohort	Epilepsy classification						
	Focal		Generalised		Unclassified		Total
	n	%	n	%	n	%	n
UCL	25	69.4%	6	16.7%	5	13.9%	36
ULB	20	69.0%	3	10.3%	6	20.7%	29
EKUT	128	51.8%	116	47.0%	3	1.2%	247
RCSI	59	80.8%	13	17.8%	1	1.4%	73
Glasgow	362	78.4%	54	11.7%	46	10.0%	462
ULIV	43	81.1%	6	11.3%	4	7.5%	53
Australia	160	70.2%	26	11.4%	42	18.4%	228
SANAD B	10	5.5%	112	61.9%	59	32.6%	181
SANAD A	539	90.3%	5	0.8%	53	8.9%	597
Entire Cohort	1346	70.6%	341	17.9%	219	11.5%	1906

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

The distribution of epilepsy syndromes was different between cohorts; as expected, SANAD arm B and EKUT had the highest proportion of patients with generalised epilepsy. The EKUT cohort, similarly to the SANAD arm B, was originally assembled for

studies of generalised epilepsy plus unclassified epilepsy in case of SANAD arm B. The highest proportion of focal epilepsy was observed in ULIV (81.8%), RCSI (80.8%) and Glasgow (78.4%). The cohorts which were electronically transferred contained a higher number of unclassified cases (ULB had 20.7% whereas Australia had 18.4%).

In the phenotype derivation created for the purpose of data analysis, epilepsy types were divided into three broad categories (generalised, focal, unclassified) whereas the original eCRF contained the more elaborate ILAE 1989 classification (Epilepsy, 1989), hence for a more detailed description of epilepsy type the original data files were used. Figures 11 and 12 summarise the distribution of aetiological classifications for both focal and generalised epilepsy. The majority of patients with focal epilepsy had a cryptogenic epilepsy (n = 879; 65%), followed by symptomatic (n = 449; 33%).

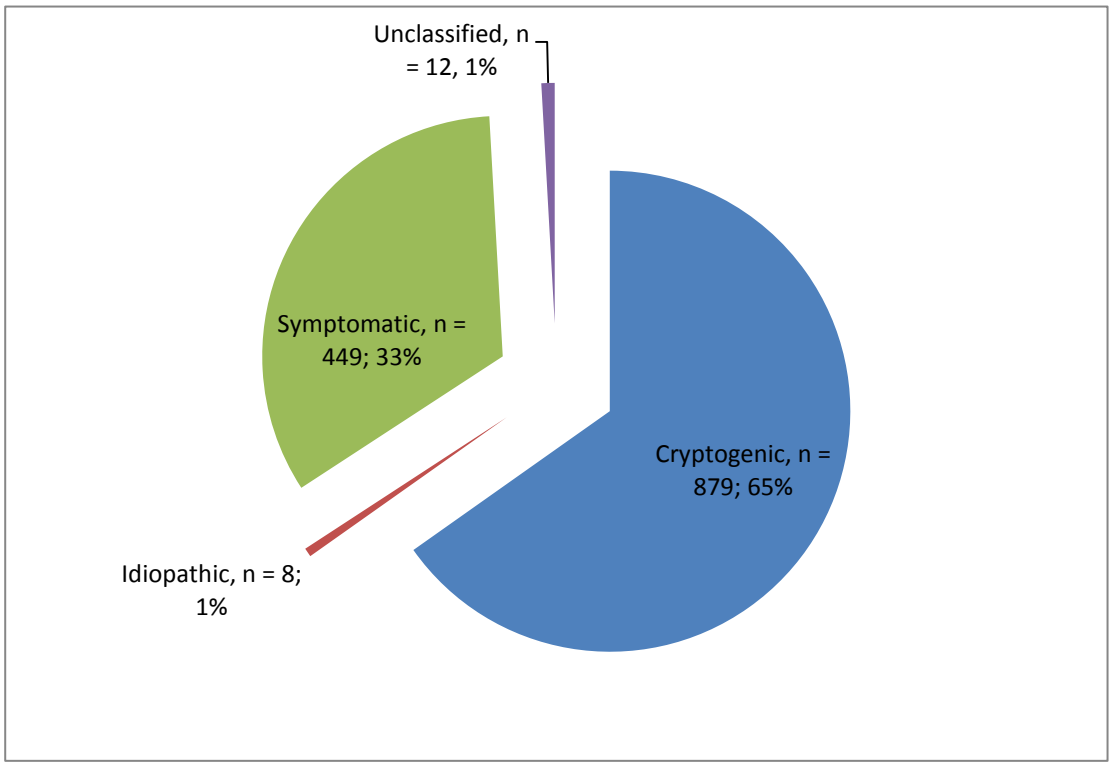


Figure 11. A pie chart based on raw data illustrating aetiological classification (symptomatic, cryptogenic, idiopathic or unclassified) for newly diagnosed focal epilepsy (n = 1348).

In most cases, the localisation of the onset lobe was unknown (n = 911; 67.58%). The most common lobe of onset documented was temporal (n = 332; 24.63%), followed by

frontal (n = 69; 5.12%), parietal (n = 22; 1.63%), and occipital (n = 14; 1.04%). There were 12 cases of mesial temporal lobe epilepsy in the cohort.

In cases of generalised epilepsy, the two most common syndromes were unclassified epileptic syndrome in 34% of cases, and JME (also 34%); childhood absence accounted for 12% and juvenile absence epilepsy for 9%. Rare epileptic syndromes were collapsed into the category "other", and were observed in 11% of cases with generalised epilepsy.

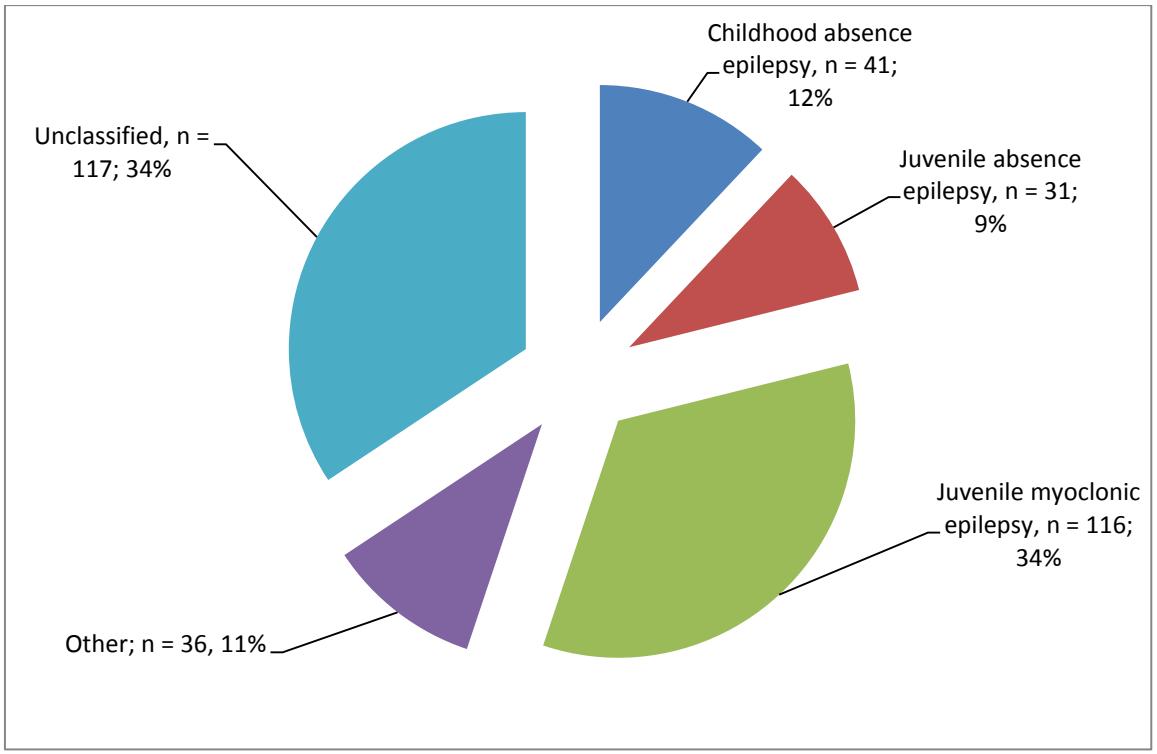


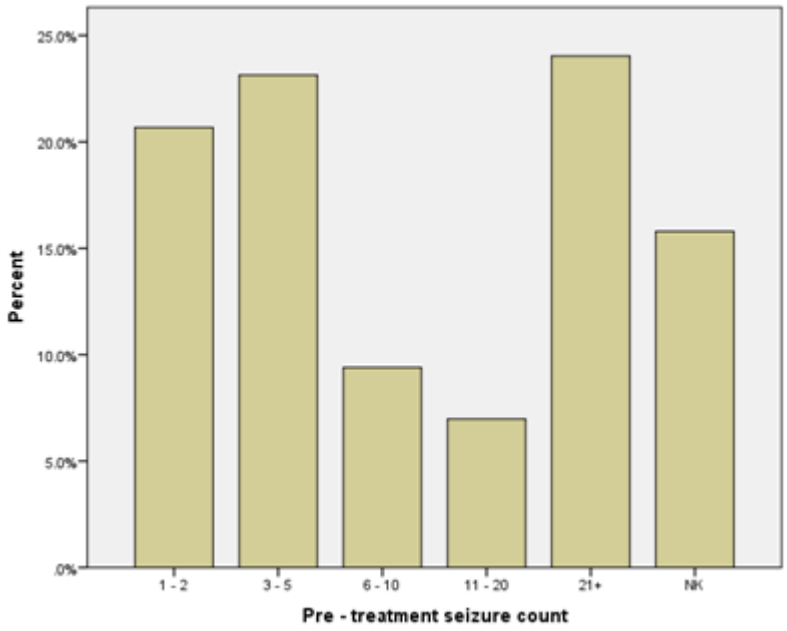
Figure 12. A pie chart based on a raw data illustrating aetiological classification (Juvenile myoclonic epilepsy, Juvenile absence epilepsy, Childhood absence epilepsy and unclassified or other) of generalised epilepsy (n = 308).

4.1.5. Pre-treatment seizure count

The reasons behind using categorical values have been described in detail in the methods section. The majority of cases (n = 1033; 59.23%) had categorical values for pre-treatment seizure count.

As can been seen from Figure 13, the seizure frequency distribution was uneven with two peaks. The majority of patients had either 1 or 2 seizures (n = 394; 20.7%), 3 – 5 seizures (n = 441 23.1%) or 21 or more (n = 458; 24.0%) before starting treatment. Furthermore, a significant proportion (n =301; 15.8%) of patients had an unknown pre-treatment seizure count.

Figure 13. Pre-treatment seizure count (%) stratified in categories (1-2; 3-5; 6-10; 11-20; 21+ and unknown).



The number of pre-treatment seizures differed between cohorts. Noticeably, the Australian cohort had a very small number (or no cases at all) in the 11 – 20 and 21+ categories. This observation is due to differences between the categories used for the data collection which did not exactly match the categories used in EpiPGX (Speed et al., 2014a). In SANAD, the pre-treatment seizure count was collected as an absolute number, hence when the data was categorised there were no cases in the unknown

category. More detailed data on pre-treatment seizure frequency, stratified by cohort, is presented in Table 15.

The effect of seizure type on the number of seizures reported was also assessed. Of all seizure types, as expected, myoclonic and absence seizures had the highest proportion of unknown pre-treatment seizure counts (26.8% and 28.5%) or 21 or more seizures (54.9% and 47.6%) before starting the treatment. This observation probably reflects the underlying high frequency and minimal intrusiveness of those seizure types. Interestingly, most (42.6%) patients with simple and complex partial seizures also reported more than 21 seizures before starting treatment. Epilepsies expressing generalised tonic-clonic seizures (either secondarily, primarily or unknown) had a lower number of unknown seizure counts before treatment, as well as a lower total number of seizures before the initiation of the treatment. Further data describing the pre-treatment seizure count, stratified by seizure type, are presented in Table 16.

Table 15. Categorised pre-treatment seizure count stratified by cohort (n = 1906)

Cohort	Pre-treatment seizure frequency											
	1 – 2		3 - 5		6 - 10		11 – 20		21+		not known	
	n	% of cohort	n	% of cohort	n	% of cohort	n	% of cohort	n	% of cohort	n	% of cohort
UCL	1	2.8%	6	16.7%	0	0.0%	3	8.3%	3	8.3%	23	63.9%
ULB	7	24.1%	0	0.0%	0	0.0%	3	10.3%	1	3.4%	18	62.1%
EKUT	57	23.1%	36	14.6%	11	4.5%	9	3.6%	79	32.0%	55	22.3%
RCSI	21	28.8%	7	9.6%	2	2.7%	1	1.4%	6	8.2%	36	49.3%
Glasgow	115	24.9%	137	29.7%	57	12.3%	25	5.4%	57	12.3%	71	15.4%
ULIV	8	15.1%	15	28.3%	4	7.5%	2	3.8%	3	5.7%	21	39.6%
Australia	78	34.2%	61	26.8%	8	3.5%	0	0.0%	4	1.8%	77	33.8%
SANAD	107	13.8%	179	23.0%	97	12.5%	90	11.6%	305	39.2%	0	0.0%
Entire cohort	**394**	**20.7%**	**441**	**23.1%**	**179**	**9.4%**	**133**	**7.0%**	**458**	**24.0%**	**301**	**15.8%**

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

Table 16. Distribution of pre-treatment seizure count stratified by seizure type (n = 1906)

			Simple or complex partial only	Secondary generalised tonic- clonic	Absence seizures	Generalised tonic-clonic seizures only	Myoclonic or absence seizures with tonic-clonic seizures	Unclassified tonic-clonic seizures	Other seizures
Pre-treatment seizure count	**1 - 2**	**n**	27	183	8	253	21	103	14
		%	6.9%	23.0%	5.6%	40.3%	8.5%	33.0%	20.6%
	3 - 5	**n**	42	214	13	239	22	110	28
		%	10.8%	26.9%	9.2%	38.1%	8.9%	35.3%	41.2%
	6 - 10	**n**	43	92	4	47	11	22	6
		%	11.0%	11.6%	2.8%	7.5%	4.5%	7.1%	8.8%
	11 - 20	**n**	55	54	1	22	5	13	3
		%	14.1%	6.8%	0.7%	3.5%	2.0%	4.2%	4.4%
	21+	**n**	166	146	78	19	117	18	12
		%	42.6%	18.4%	54.9%	3.0%	47.6%	5.8%	17.6%
	Not known	**n**	57	106	38	48	70	46	5
		%	14.6%	13.3%	26.8%	7.6%	28.5%	14.7%	7.4%
	Entire Cohort	**n**	390	795	142	628	246	312	68
		%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%	100.0%

4.1.6. Seizure type

The most commonly observed seizure type was secondarily generalised tonic-clonic (n= 795; 41.7%), followed by complex partial (n = 707; 37.1%) and simple partial (n = 426; 22.4%). On the other hand, tonic (n = 2; 0.1%), atonic (n = 3; 0.2%) or clonic (n = 72; 3.8%) seizure types were observed relatively rarely. It was impossible to classify partial seizures type in 2.7% (n = 52), and much more often it was impossible to classify generalised tonic-clonic seizures (n = 312; 16.4). Further detailed data on seizure types according to the original categories are presented in Table 17.

To ensure consistency with previous work based on the SANAD analysis of association seizure types were transformed according to the previous statistical modelling (Bonnett et al., 2014b, Bonnett et al., 2012), and those categories are described in more detail in the methods section..

The most frequently observed seizure types were those that included generalised tonic-clonic seizures. Secondarily generalised tonic-clonic seizures were observed in 41.7% of patients, and 32.9% of all patients experienced only generalised tonic-clonic seizures. Isolated simple or complex partial seizures were observed in 20.5% of the patients, and absence seizures in 7.5%. The frequency of the transformed seizure categories separated by cohort is further summarised in Table 18.

Seizure type data contained information on all of the seizure types ever experienced by the subject. It also includes seizure types appearing later during the course of the illness. EpiPGx did not collect data on the type of index seizure leading to diagnosis of epilepsy; hence it is impossible to derive the type of the index seizure. To assess the number of seizure types per subject, the original seizure categories were used. More than half of the subjects (n = 1050; 55.1%) had only one seizure type; 36.9% (n = 703) had two; 7.5% (n = 142) had three; 0.2% (n = 3) had four and a single patient (0.1%) had five seizure types. There were 7 (0.4%) cases without any seizure type entered.

Table 17. Distribution of seizure types according to ILAE 1981 seizure classification (Bancaud et al., 1981)

Seizure type	n	%
Primary generalised tonic-clonic	236	12.4
Absence	142	7.5
Clonic	72	3.8
Tonic	2	0.1
Atonic	3	0.2
Myoclonic	139	7.3
Simple partial	426	22.4
Complex partial	707	37.1
Secondarily GTC	795	41.7
Unclassified partial	52	2.7
Unclassified GTC	312	16.4

Table 18. Distribution (%) of seizure types categorised according to SANAD seizure classification and stratified by cohort

Cohort	Simple or complex partial only	Secondary generalised tonic-clonic	Generalised tonic-clonic seizures only	Absence seizures	Myoclonic or absence seizures with tonic-clonic seizures	Unclassified tonic-clonic seizures	Other seizures
UCL	22.2%	47.2%	25.0%	11.1%	13.9%	8.3%	5.6%
ULB	20.7%	41.4%	34.5%	0.0%	3.4%	13.8%	6.9%
EKUT	14.2%	38.1%	23.9%	22.7%	33.2%	0.4%	0.0%
RCSI	15.1%	56.2%	21.9%	6.8%	9.6%	15.1%	2.7%
Glasgow	12.3%	44.4%	37.2%	3.2%	9.7%	30.5%	1.7%
ULIV	7.5%	45.3%	11.3%	7.5%	11.3%	34.0%	1.9%
Australia	20.6%	35.1%	48.2%	5.7%	10.5%	20.6%	2.2%
SANAD	28.5%	41.4%	31.6%	5.8%	9.8%	11.2%	7.5%
Entire cohort	20.5%	41.7%	32.9%	7.5%	12.9%	16.4%	4.1%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study

4.1.7. Investigations

Where the available results of the EEG, MRI and CT were collected, they were categorised as either normal, abnormal-specific (epileptiform for the EEG and focal and attributable to epilepsy for imaging studies), abnormal non-specific, or not done. EEG results were available in 1575 (82.6%) cases. The majority of patients had normal EEG results (n = 650; 34.1%) followed by abnormal epileptiform EEG (n = 568; 29.8%), abnormal non-specific (n = 357; 18.7%) and not done (n = 331; 17.4%).

Data on CT head scans were available for 860 (45.4%) subjects. One half (n = 1046; 54.9%) of patients did not have a CT head scan performed, probably due to the general trend in neurology to utilise more MRI. It has already been pointed out that the majority of patients were diagnosed after the year 2000, when MRI was the method of choice for imaging in epilepsy. Diagnostic yield for the CT head scan was relatively low. From all the cases that included it, only 18.37% (n = 158) had relevant focal abnormalities.

MRI imaging was performed in 1412 (74.5%) cases. The majority of patients had normal imaging results (958 out of 1412; 67%); relevant focal abnormalities were detected in 311 out of 1412 (22.02%) patients. To replicate the imaging co-variate used in the SANAD study, all of the neuroimaging categories were collapsed together and a new mixed CT/MRI variable was created. It had three categories: normal, abnormal or not done, and MRI result would have precedence over CT. Some form of neuroimaging was performed on 1656 (86.9%) subjects and no imaging was done, or the results were unknown, in 250 (13.1%) cases. The majority of patients, 1140 out of 1656 (68.84%), had normal neuroimaging, whereas 516 (31.16%) had some abnormalities. More detailed information on the investigation results is presented in Table 19.

There were 616 individuals who had both a CT and an MRI examination. From these, 138 had a focal lesion demonstrated on an MRI, 120 (87.0%) had reported focal abnormalities on a CT scan, 12.3% (n = 17) were reported as normal and one (0.7%) had non-specific abnormalities on a CT scan. The highest concordance was observed between normal results of both modalities – 99.3% had concordant scans. Non-specific abnormalities on the other hand had the worst concordance; out of 43 patients who had

them reported on an MRI, only 37.2% had the same category based on the results of a CT scan, with 55.8% being reported as normal using the CT scan.

4.1.8. Family history

In total, 365 subjects (19.2%) had a positive family history, while the rest had an absent or unknown family history. A positive family history was unevenly distributed between cohorts. There were no cases with family history recorded as positive in the UCL cohort, whereas ULB (n = 11; 37.9%) and Glasgow (n = 136; 29.4%) had the highest proportion of patients with a positive family history. Further data on family history stratified by cohort has been summarised in Table 20.

Table 19. Positive family history stratified by cohort

		Positive family history			
		None /Unknown		Yes	
		n	%	n	%
Cohort	UCL	36	100.0%	0	0.0%
	ULB	18	62.1%	11	37.9%
	EKUT	193	78.1%	54	21.9%
	RCSI	55	75.3%	18	24.7%
	Glasgow	326	70.6%	136	29.4%
	ULIV	40	75.5%	13	24.5%
	Australia	167	73.2%	61	26.8%
	SANAD	706	90.7%	72	9.3%
	Entire cohort	1541	80.8%	365	19.2%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

As described earlier this variable was dichotomized hence to assess the number of cases with unknown family history, the original dataset were used. There were 201 (10.55%) cases with unknown status and 1340 (70.30%) with a negative family history out of 1906.

Table 20. Investigation results (EEG, CT, MRI) stratified by cohort

		Cohort																	
		UCL		ULB		EKUT		RCSI		Glasgow		ULIV		Australia		SANAD		Entire Cohort	
		n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
EEG	Normal	3	8.3%	5	17.2%	30	12.1%	15	20.5%	139	30.1%	11	20.8%	150	65.8%	297	38.2%	**650**	**34.1%**
	Epileptiform	16	44.4%	12	41.4%	126	51.0%	15	20.5%	112	24.2%	11	20.8%	76	33.3%	200	25.7%	**568**	**29.8%**
	Non-specific	8	22.2%	0	0.0%	28	11.3%	10	13.7%	127	27.5%	13	24.5%	0	0.0%	171	22.0%	**357**	**18.7%**
	Not done / NK*	9	25.0%	12	41.4%	63	25.5%	33	45.2%	84	18.2%	18	34.0%	2	0.9%	110	14.1%	**331**	**17.4%**
CT	Normal	3	8.3%	0	0.0%	24	9.7%	24	32.9%	149	32.3%	18	34.0%	0	0.0%	437	56.2%	**655**	**34.4%**
	Focal	2	5.6%	0	0.0%	4	1.6%	17	23.3%	22	4.8%	10	18.9%	0	0.0%	103	13.2%	**158**	**8.3%**
	Non-specific	0	0.0%	0	0.0%	7	2.8%	2	2.7%	22	4.8%	4	7.5%	0	0.0%	12	1.5%	**47**	**2.5%**
	Not done / NK*	31	86.1%	29	100.0%	212	85.8%	30	41.1%	269	58.2%	21	39.6%	228	100.0%	226	29.0%	**1046**	**54.9%**
MRI	Normal	18	50.0%	5	17.2%	109	44.1%	34	46.6%	229	49.6%	19	35.8%	195	85.5%	349	44.9%	**958**	**50.3%**
	Focal	9	25.0%	20	69.0%	42	17.0%	19	26.0%	57	12.3%	10	18.9%	33	14.5%	121	15.6%	**311**	**16.3%**
	Non-specific	6	16.7%	0	0.0%	25	10.1%	6	8.2%	88	19.0%	9	17.0%	0	0.0%	9	1.2%	**143**	**7.5%**
	Not done / NK*	3	8.3%	4	13.8%	71	28.7%	14	19.2%	88	19.0%	15	28.3%	0	0.0%	299	38.4%	**494**	**25.9%**

*NK = not known

4.1.9. Neurological examination / sequelae

The majority of patients (n=1754; 92.0%) had a normal/unknown/not applicable neurological examination and 152 (8.0%) had a documented abnormal examination. The ULB cohort (n = 7; 24.1%), ULIV (n = 12; 22.6%) and RCSI (n = 15; 20.5%) had the highest proportion of subjects with an abnormal neurological examination, whereas the Australian cohort (n = 10; 4.4%), Glasgow (n = 20; 4.3%) and SANAD (n = 46; 5.9%) had the least number of documented cases with an abnormal examination. More detailed results stratified by cohort are presented in Table 21.

Table 21. Results of neurological examination stratified by cohort

Cohort	Neurological examination			
	Normal / Not available		Abnormal	
	n	%	n	%
UCL	30	83.3%	6	16.7%
ULB	22	75.9%	7	24.1%
EKUT	211	85.4%	36	14.6%
RCSI	58	79.5%	15	20.5%
Glasgow	442	95.7%	20	4.3%
ULIV	41	77.4%	12	22.6%
Australia	218	95.6%	10	4.4%
SANAD	732	94.1%	46	5.9%
Entire Cohort	1754	92.0%	152	8.0%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

To assess the number of cases with unknown or not applicable outcomes, the original dataset was explored (n = 1906). There were 1658 (86.99%) cases with normal neurological examination, 151 with abnormal (7.92%) and 97 (5.09%) with either unknown or not applicable status.

4.1.10. Febrile convulsions

A history of febrile convulsions was documented in 75 (3.9%) cases. There were differences observed between cohorts, with ULB having no history and 6.5% of the EKUT

subjects and 5% of SANAD subjects having a history of febrile convulsions. The data on febrile convulsions is summarised by cohort in Table 22.

Table 22. Presence or absence of past medical history of febrile convulsion stratified by cohort (n = 1906)

Cohort	Febrile convulsions			
	No		Yes	
	n	%	n	%
UCL	35	97.2%	1	2.8%
ULB	29	100.0%	0	0.0%
EKUT	231	93.5%	16	6.5%
RCSI	71	97.3%	2	2.7%
Glasgow	448	97.0%	14	3.0%
ULIV	52	98.1%	1	1.9%
Australia	226	99.1%	2	0.9%
SANAD	739	95.0%	39	5.0%
Entire Cohort	1831	96.1%	75	3.9%

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

4.1.11. Learning disabilities

There were 1369 (71.8%) cases with unknown status or present known learning disabilities and 537 (28.2%) with no history of LDs. Because data was recoded in the phenotype derivation, the original data file was used to describe in detail the distribution of learning disabilities. Wide differences in the presence of LDs between cohorts were observed. The UCL and ULB cohorts had no cases with documented learning disabilities, and the ULB cohort had no collected data on LDs. UCL is a relatively small cohort with a high proportion of unknown (66.67%) LD status. The highest proportion of unknown data was in the ULB (100%), Australia (99.13%) and SANAD (93.44%) cohorts. More detailed results are summarised in Table 23.

In summary, due to the fact that the collection of the data on LDs was started in the middle of the project, the results of this co-variate are very unreliable.

Table 23. Distribution of learning disabilities stratified by cohort based on a raw data file.

Cohort	Learning disabilities						
	Yes	%	No	%	Not known	%	Total
Australia	1	0.44%	1	0.44%	227	99.13%	229
EKUT	4	1.63%	13	5.28%	229	93.09%	246
Glasgow	5	1.08%	434	93.94%	23	4.98%	462
RCSI	3	4.11%	38	52.05%	32	43.84%	73
SANAD	29	3.73%	22	2.83%	727	93.44%	778
UCL	0	0.00%	12	33.33%	24	66.67%	36
ULB	0	0.00%	0	0.00%	29	100.00%	29
ULIV	4	7.55%	17	32.08%	32	60.38%	53
Entire cohort	46	2.41%	537	28.17%	1323	69.41%	1906

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

4.1.12. Time from first seizure to first AED

Data for this co-variate were available for 1873 subjects; in 25 cases it was missing and in 8 cases values were negative. Negative values probably were due to errors in phenotyping. The distribution of this co-variate was skewed towards the left (see Figure 14.) The distribution of the data showed a single early peak with a relatively steep fall off. The median time from the first seizure to the first AED was 11.53 months, the mean time was 42.62 months and IQR was 4.01 - 37.42.

There were differences observed between cohorts, with RCSI (median = 4.57 months), EKUT (median = 4.99 months) and ULB (median = 6.14 months) having the shortest median time, whereas SANAD (median = 14.64 months) and UCL (median = 14.92 months) had the longest time from first seizure to first AED. The rest of the data, separated by cohort, are presented in Table 24. For the analysis of association, a logarithmic transformation of the data was used.

Table 24. Time (in months) from first seizure to first AED stratified by cohort. IQR – interquartile range.

Cohort	Time from first seizure to first AED			
	Mean	Standard deviation	Median	IQR
UCL	33.57	48.00	14.92	2.92 - 34.63
ULB	17.60	25.59	6.14	1.02 - 18.56
EKUT	18.26	40.74	4.99	1.12 - 15.97
RCSI	17.22	30.13	4.57	0.26 - 17.51
Glasgow	39.10	72.69	11.47	4.25 - 37.40
ULIV	34.41	62.75	10.78	3.06 - 32.95
Australia	44.34	83.90	13.04	3.84 - 36.93
SANAD	55.82	100.78	14.64	5.98 - 50.00
Entire Cohort	42.62	83.06	11.53	4.01 - 37.42

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

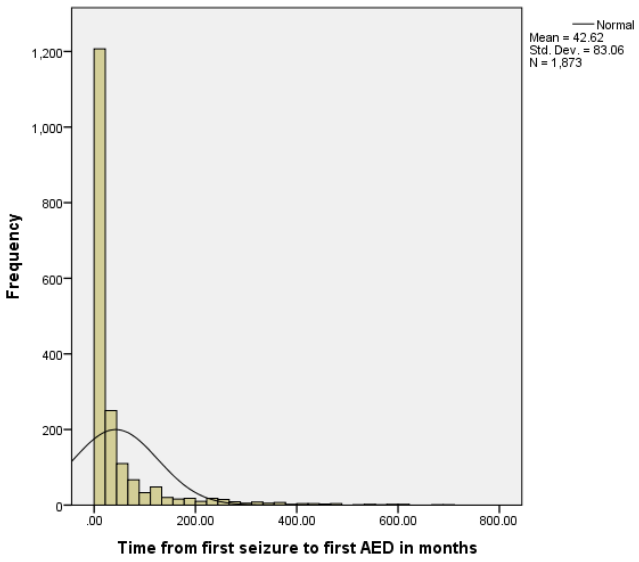


Figure 14. Histogram of frequency distribution for time from first seizure to first AED in months (n = 1873).

4.1.13. First well-tolerated anti-epileptic drug

Data on the application of the first applied anticonvulsant was available in 1874 cases (98.3%) and was missing in 32 cases (1.7%). This information is required for assessment of treatment outcome; as such, these 32 cases were excluded from further analysis.

The data on the first applied anticonvulsant stratified by cohort is presented in Table 25 and Figure 15 summarises the frequency of individual AEDs. Three subjects in Glasgow had participated in a randomised clinical trial for newly diagnosed epilepsy. They were exposed to one of the currently approved AEDs. The results of un-blinding after the end of the trial were not stored in clinical notes for those cases. As tasks one and three of work package two do not investigate selective drug response, they were included in the study. Several cohorts were initially assembled as clinical trials, for example, lamotrigine and levetricetam were extensively researched in the Glasgow and SANAD studies. Hence, our data are not representative of a baseline population of newly diagnosed epilepsy cases, nor do they characterise the routine non-clinical trial prescribing patterns of the recruiting centre. Overall, the most commonly prescribed anticonvulsant was lamotrigine (n= 478; 25.1%) followed by valproate (n = 415; 21.8%) and carbamazepine (n = 338; 17.7%).

Table 25. First applied anticonvulsant in newly diagnosed epilepsy according to the research cohort (n = 1874)

Generic name of AED	Cohort																	
	UCL		ULB		EKUT		RCSI		Glasgow		ULIV		Australia		SANAD		Entire Cohort	
	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%	n	%
Missing	0	0.0	1	3.4	0	0.0	0	0.0	10	2.2	0	0.0	8	3.5	13	1.7	32	1.7
Carbamazepine	6	16.7	3	10.3	20	8.1	17	23.3	34	7.4	20	37.7	87	38.2	151	19.4	338	17.7
Clobazam	1	2.8	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1
Ethosuximide	0	0.0	0	0.0	8	3.2	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1	9	0.5
Felbamate	0	0.0	0	0.0	0	0.0	0	0.0	3	0.6	0	0.0	0	0.0	0	0.0	3	0.2
Gabapentin	0	0.0	0	0.0	0	0.0	2	2.7	6	1.3	1	1.9	1	0.4	103	13.2	113	5.9
Lacosamide	0	0.0	0	0.0	0	0.0	0	0.0	1	0.2	0	0.0	0	0.0	0	0.0	1	0.1
Levetiracetam	3	8.3	7	24.1	36	14.6	12	16.4	76	16.5	0	0.0	18	7.9	7	0.9	159	8.3
Levetiracetam x Carbamazepine	0	0.0	0	0.0	0	0.0	0	0.0	2	0.4	0	0.0	0	0.0	0	0.0	2	0.1
Lamotrigine	14	38.9	3	10.3	61	24.7	11	15.1	151	32.7	9	17.0	25	11.0	204	26.2	478	25.1
Oxcarbazepine	0	0.0	0	0.0	15	6.1	6	8.2	17	3.7	0	0.0	0	0.0	59	7.6	97	5.1
Phenobarbital	0	0.0	0	0.0	2	0.8	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	2	0.1
Pregabalin	0	0.0	0	0.0	1	0.4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1
Phenytoin	3	8.3	2	6.9	4	1.6	14	19.2	2	0.4	6	11.3	12	5.3	1	0.1	44	2.3
Sulthiame	0	0.0	0	0.0	3	1.2	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	3	0.2
Tiagabine	0	0.0	0	0.0	0	0.0	0	0.0	2	0.4	0	0.0	0	0.0	0	0.0	2	0.1
Topiramate	2	5.6	1	3.4	8	3.2	0	0.0	30	6.5	1	1.9	3	1.3	158	20.3	203	10.7
Topiramate x Valproate	0	0.0	0	0.0	0	0.0	0	0.0	1	0.2	0	0.0	0	0.0	0	0.0	1	0.1
Vigabatrin	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1	1	0.1
Valproate	7	19.4	12	41.4	88	35.6	11	15.1	127	27.5	16	30.2	74	32.5	80	10.3	415	21.8
Zonisamide	0	0.0	0	0.0	1	0.4	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0	1	0.1

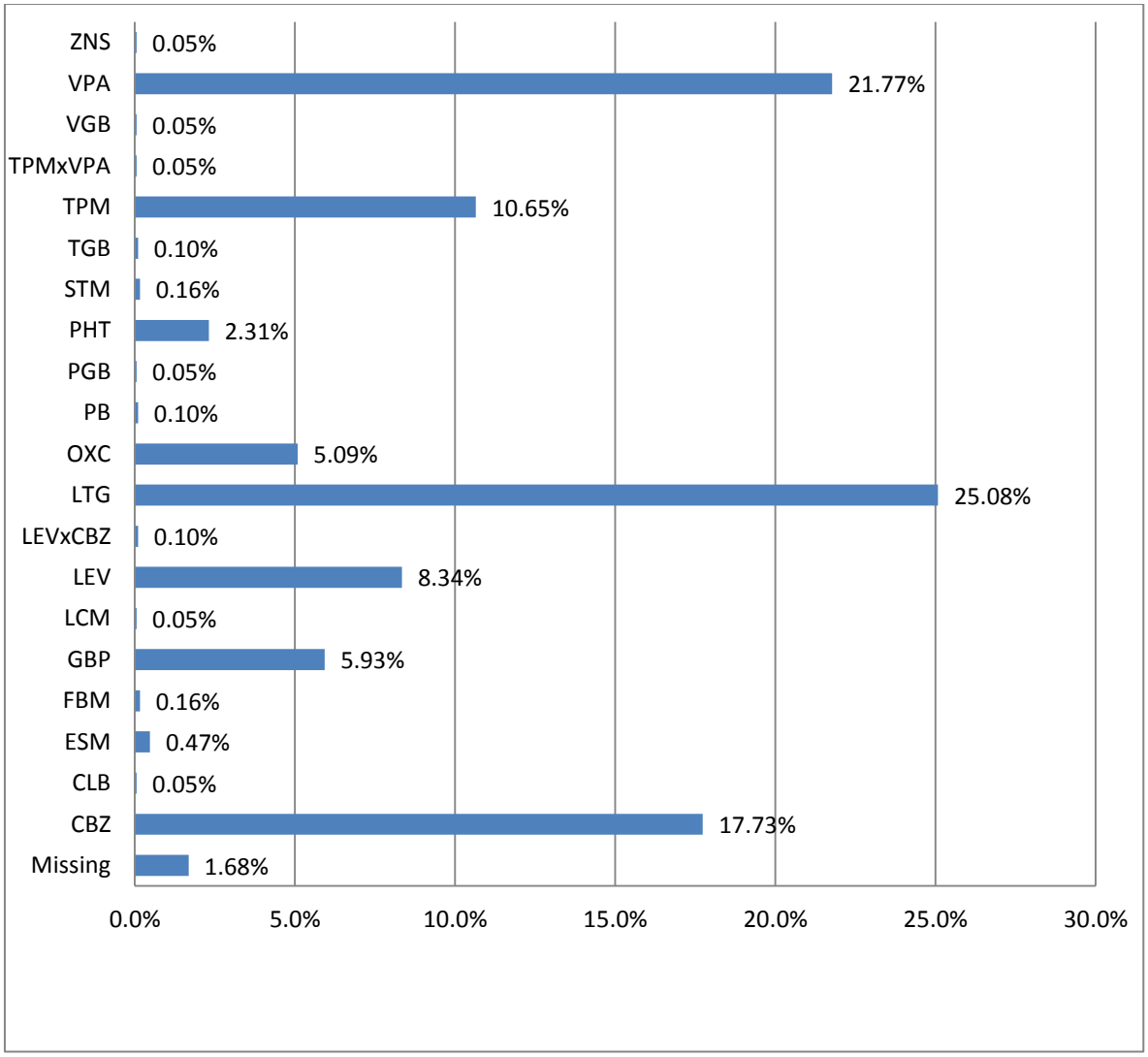


Figure 15. The distribution of the first well-tolerated anticonvulsant in cases of newly diagnosed epilepsy (n = 1874)

4.2. Logistic regression

A logistic regression analysis was used to assess the predictive role of clinical co-variates on the occurrence of seizure remission for at least 12 months in patients with newly diagnosed epilepsy after application of the first anticonvulsant. A full list of clinical covariates tested for association is presented in Table 26 a and b. In total, 1906 cases were selected for the analysis; 1723 (90.4%) were included and 183 (9.6%) were excluded due to missing information. Reasons for exclusion are the same as presented at the beginning of this chapter.

Univariate analysis showed that age at diagnosis, follow-up length, gender, epilepsy type, neurological examination results, generalised tonic-clonic seizures only, pre-treatment seizure count, cohort, and EEG and MRI results are statistically significantly associated at level of $p < 0.05$ with the presence of 12-month remission in newly diagnosed epilepsy after application of the first well-tolerated anticonvulsant. Full results of the univariate analysis are presented in Table 26 a and b.

Table 26a. Univariate analysis of all continuous clinical factors included in logistic regression model (IQR – inter quartile range)

Continuous co – variates	Total	Median; IQR	p-value	p-value (Bonferonni correction)
Age at diagnosis (years)	1744	Median = 33.21; IQR 19.85 - 49.63	0.001	0.021
Time from first seizure to AED in years (log)	1724	Median = 2.54; IQR 2.08 - 3.04	0.465	1.0
Follow-up length (years)	1743	Median = 3.62; IQR 2.22 - 5.70	0.008	0.155

Table 26b. Univariate analysis of all categorical clinical factors included in logistic regression model.

Categorical co-variates (n=1744)		Total	%	p-value	p-value (Bonferonni correction)
Gender	Males	898	51.5%	0.008	0.155
	Females	846	48.5%		
Positive family history	Yes	331	19.0%	0.963	1.0
	No / NK	1413	81.0%		
Neurological examination	Abnormal / NA	139	8.0%	0.000	0.000
	Normal	1605	92.0%		
Febrile seizures	Yes	73	4.2%	0.322	0.999
	No / NA	1671	95.8%		
Learning disabilities	Present / NK	489	28.0%	0.430	1.0
	Absent	1255	72.0%		
Seizure type					
Simple or complex partial only	Yes	357	20.5%	0.072	0.792
	No	1387	79.5%		
Secondary generalised tonic-clonic	Yes	722	41.4%	0.128	0.944
	No	1022	58.6%		
Generalised tonic-clonic seizures only	Yes	564	32.3%	0.000	0.000
	No	1180	67.7%		
Absence seizures	Yes	131	7.5%	0.613	1.0
	No	1613	92.5%		
Myoclonic or absence seizures with tonic-clonic seizures	Yes	226	13.0%	0.915	1.0
	No	1518	87.0%		
Unclassified tonic-clonic seizures	Yes	289	16.6%	0.064	0.751
	No	1455	83.4%		

Other seizures	Yes	75	4.3%	0.956	1.0
	No	1669	95.7%		
Epilepsy type					
Focal epilepsy - baseline		1235	70.8%	0.004	0.081
Generalised epilepsy		316	18.1%		
Unclassified epilepsy		193	11.1%		
Pre-treatment seizure count					
1 – 2 (baseline)		364	20.9%	0.000	0.000
3 - 5		395	22.6%		
6 – 10		165	9.5%		
11 - 20		127	7.3%		
21 +		435	24.9%		
Pre-treatment seizure count unknown		258	14.8%		
Cohort					
SANAD (baseline)		748	42.9%	0.000	0.000
UCL		33	1.9%		
ULB		17	1.0%		
EKUT		235	13.5%		
RCSI		72	4.1%		
Glasgow		415	23.8%		
ULIV		53	3.0%		
Australia		171	9.8%		
EEG					
EEG Normal (baseline)		586	33.6%	0.000	0.002
EEG Epileptiform		511	29.3%		
EEG Abnormal non-specific		342	19.6%		
EEG Not done/not known		305	17.5%		
MRI					
MRI Normal (baseline)		866	49.7%	0.000	0.007
MRI Focal		281	16.1%		
MRI Abnormal non-specific		130	7.5%		
MRI Not done/not known		467	26.8%		
CT					
CT Normal (baseline)		623	35.7%	0.239	0.996
CT Focal		155	8.9%		
CT Abnormal non-specific		42	2.4%		
CT Not done/not known		924	53.0%		

NK – Not known; NA – Not available

Fitting a multivariable binary logistic regression model showed that epilepsy type, pre-treatment seizure count, age at diagnosis, origin of cohort, EEG results, generalised tonic-clonic seizures only, CT and MRI results, and follow-up length were all significantly

associated with outcome. Older age at diagnosis and the presence of generalised tonic-clonic seizures increased the probability of attaining a 12-month remission. On the other hand, an abnormal neurological examination was associated with a reduced probability of experiencing a 12-month freedom from seizures. The direction of effect for epilepsy type pointed towards a higher probability of experiencing remission in the case of generalised epilepsy. A negative direction of effect was observed in cases of focal abnormalities on MRI image, and not done or unknown EEG. Direction of effect in the case of focal abnormal CT pointed towards a higher probability of experiencing remission. A more complex picture was observed in the case of pre-treatment seizure frequency, with 6 – 10, more than 21, or an unknown number of seizures before starting treatment having a negative direction of effect on the probability of a 12-month remission. The full results of the logistic regression are summarised in Figure 16 and Table 27. The Cox and Snell R Square for the final model was 0.135 and the Nagelkerke R Square was 0.181.

Table 27. Multivariable logistic regression model for 12-month seizure remission in newly diagnosed epilepsy (n = 1723)

Co-variate	**p-value**	**Odds Ratio**	**95% C.I. for Odds Ratio**	
Male	0.066	1.218	0.987	1.502
Age at diagnosis	0.000	1.019	1.012	1.026
Abnormal neurological examination	0.000	0.438	0.289	0.666
Generalised tonic-clonic seizures only	0.000	1.734	1.320	2.278
Follow-up length	0.000	1.054	1.023	1.085
Epilepsy type				
Focal (baseline)				
Generalised	0.000	1.944	1.385	2.728
Unclassified		0.954	0.658	1.381
Pre-treatment seizure count				
1 – 2 (baseline)				
3 - 5	0.001	1.033	0.743	1.437
6 - 10		0.600	0.393	0.914
11 - 20		0.742	0462	1.191

21 +		0.564	0.379	0.838
Not known pre-treatment seizure count		0.564	0.379	0.838
Cohort				
SANAD (baseline)				
UCL	0.000	0.135	0.047	0.388
ULB		0.320	0.092	1.119
EKUT		0.953	0.646	1.408
RCSI		0.430	0.225	0.823
Glasgow		0.740	0.541	1.011
ULIV		0.091	0.037	0.220
Australia		1.392	0.847	2.288
EEG				
EEG Normal (baseline)				
EEG Epileptiform	0.028	0.956	0.713	1.282
EEG Abnormal non-specific		0.915	0.676	1.239
EEG Not done/not known		0.624	0.451	0.864
CT				
CT Normal (baseline)				
CT Abnormal focal	0.030	1.633	1.009	2.645
CT Abnormal non-specific		0.576	0.275	1.207
CT Not done/not known		1.265	0.962	1.662
MRI				
MRI Normal (baseline)				
MRI Abnormal focal	0.009	0.560	0.385	0.814
MRI Abnormal non-specific		1.093	0.701	1.705
MRI Not done/not known		1.036	0.790	1.359

SANAD is the only included cohort that was a randomised clinical trial and it provides 43% of all cases. To assess the impact of this cohort on the results, a sensitivity analysis was carried out by repeating the above analyses but this time excluding all of the SANAD cases. Gender, epilepsy type, age at onset, generalised tonic-clonic seizures only, EEG, MRI, follow-up length and origin of cohort remained significant in this sensitivity analysis, whilst the variables pre-treatment seizure number and CT imaging results lost their significance.

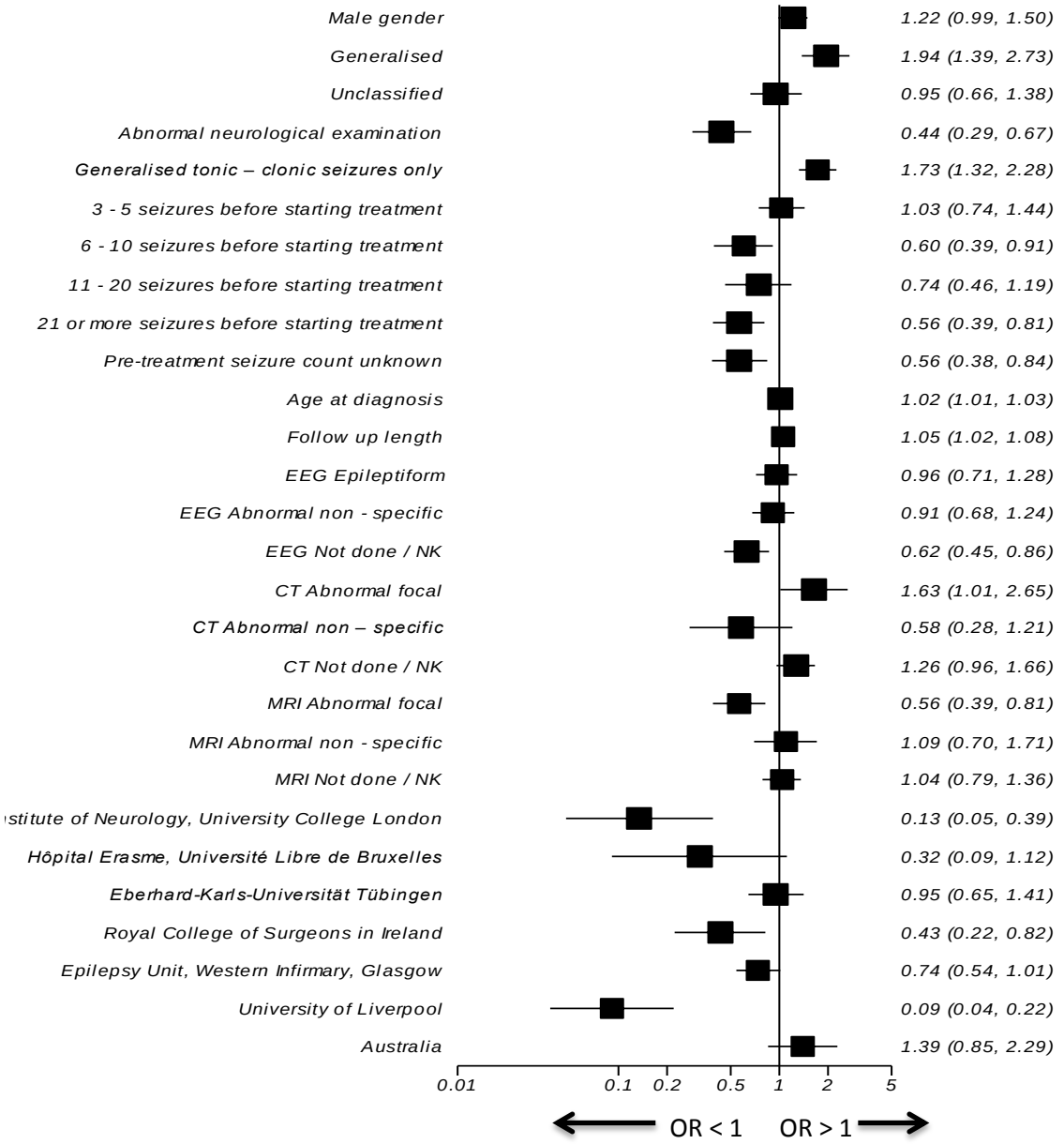


Figure 16. Forest plot (log scale) of results of logistic regression for treatment outcome (12 month remission vs. no remission) in newly diagnosed epilepsy after application of first anticonvulsant. The following categories are assumed baseline for the categorical covariates: female sex, focal epilepsy, 1-2 seizures before starting treatment, normal EEG, normal CT, normal MRI and SANAD. Right lateral column demonstrates 95% C.I. for Odds Ratio. OR > 1 indicates a higher probability of 12 months remission.

4.3. Cox regression

After the logistic regression analysis, a Cox regression proportional hazard analysis was carried out. This was used to assess the association between the predictive role of the same clinical covariates and the time to a 12-month seizure remission in patients with newly diagnosed epilepsy after application of the first anticonvulsant.

In total, 1906 cases were selected for the analysis, and 1723 (90.4%) were included; an event occurred in 50.8% (n = 968) cases and 39.6% (n = 755) were censored. The reasons for the missing data were identical to those described in the logistic regression paragraph. Before Cox regression univariate analysis was performed, for categorical variables log-rank test was used but for continuous variables Cox regression with single co-variate were used. Results of the univariate analysis are presented in table 28 a and b.

Table 28a. Univariate analysis of all continuous clinical factors included in Cox regression model. As distribution of data was identical with logistic regression summary statistics have been omitted.

Continuous co – variates	**Total**	**p-value**	**p-value (Bonferonni correction)**
Age at diagnosis (years)	1744	0.000	0.002
Time from first seizure to AED in years (log)	1724	0.185	0.986
Follow-up length (years)	1743	0.000	0.000

Table 28b. Univariate analysis of all categorical clinical factors included in Cox regression model. As distribution of data was identical with logistic regression summary statistics have been omitted.

Categorical co-variates (n=1744)		**p-value**	**p-value (Bonferonni correction)**
Gender	Males	0.004	0.074
	Females		
Positive family history	Yes	0.934	1.0
	No / NK		
Neurological examination	Abnormal / NA	0.000	0.000
	Normal		
Febrile seizures	Yes	0.256	0.998
	No / NA		
Learning disabilities	Present / NK	0.133	0.950
	Absent		

Seizure type			
Simple or complex partial only	Yes	0.072	0.790
	No		
Secondary generalised tonic-clonic	Yes	0.0622	0.740
	No		
Generalised tonic-clonic seizures only	Yes	0.000	0.000
	No		
Absence seizures	Yes	0.562	1.0
	No		
Myoclonic or absence seizures with tonic-clonic seizures	Yes	0.869	1.0
	No		
Unclassified tonic-clonic seizures	Yes	0.065	0.756
	No		
Other seizures	Yes	0.888381	1.0
	No		
Epilepsy type			
Focal epilepsy - baseline		0.000	0.007
Generalised epilepsy			
Unclassified epilepsy			
Pre-treatment seizure count			
1 – 2 (baseline)		0.000	0.000
3 - 5			
6 – 10			
11 - 20			
21 +			
Pre-treatment seizure count unknown			
Cohort			
SANAD (baseline)		0.000	0.000
UCL			
ULB			
EKUT			
RCSI			
Glasgow			
ULIV			
Australia			
EEG			
EEG Normal (baseline)		0.000	0.001
EEG Epileptiform			
EEG Abnormal non-specific			
EEG Not done/not known			
MRI			
MRI Normal (baseline)		0.000	0.003
MRI Focal			
MRI Abnormal non-specific			
MRI Not done/not known			
CT			
CT Normal (baseline)		0.0938	0.874
CT Focal			
CT Abnormal non-specific			
CT Not done/not known			

The Cox regression model showed that gender, epilepsy type, pre-treatment seizure count, age at diagnosis, origin of cohort, EEG results, generalised tonic-clonic seizures only and MRI results were all significantly associated with time to onset of 12 months remission after application of the first well tolerated AED. An increase in age at diagnosis and the presence of generalised tonic-clonic seizures were associated with shorter time to onset of 12 months remission. On the other hand, an abnormal neurological examination was associated with a longer time to experience a 12-month freedom from seizures. The direction of effect for epilepsy type pointed towards a shorter time to onset for remission in the case of generalised epilepsy. A negative direction of effect was observed in the case of focal abnormalities on the MRI image, and not done or unknown EEG. In the case of the pre-treatment seizure count, a more complex picture was observed, with 6 – 10, more than 21, or an unknown number of seizures before starting treatment having a negative direction on time to onset of 12 month-remission. The full results of the Cox regression are presented in Table 29 and in figure 17.

The assumption of proportionality of hazards over time was checked with Schoenfield residuals (data not shown) and this was not upheld for all covariates The internal validity of the model was assessed by Harrell's C and Somers' D index. The value for Harrell's C was 0.6853360185 and for Somers' D it was 0.3706720371 (Harrell et al., 1996).

Table 29. Cox regression model for 12 month seizure remission in newly diagnosed epilepsy (n = 1723)

Covariate	p-value	Hazard Ratio	95.0% CI for Hazard Ratio	
			Lower	Upper
Male	0.058	1.135	0.996	1.294
Age at diagnosis	0.000	1.011	1.007	1.015
Generalised tonic-clonic seizures only	0.000	1.405	1.201	1.644
Abnormal neurological examination	0.001	0.595	0.438	0.807
Epilepsy type				
Focal (background)	0.002			
Generalised		1.436	1.169	1.763
Unclassified		0.998	0.806	1.237
Pre-treatment seizure count				
1 – 2 (background)	0.001			

3 – 5		1.022	0.853	1.225
6 – 10		0.754	0.578	0.983
11 – 20		0.785	0.588	1.047
21 or more		0.678	0.541	0.850
Pre-treatment seizure count unknown		0.765	0.598	0.978
EEG				
EEG Normal	0.036			
EEG Epileptiform		0.994	0.833	1.188
EEG Abnormal non-specific		0.966	0.801	1.164
EEG Not done/not known		0.746	0.604	0.922
CT				
CT Normal (background)	0.088			
CT Abnormal focal		1.324	0.977	1.795
CT Abnormal non – specific		0.716	0.432	1.188
CT Not done/not known		1.116	0.942	1.321
MRI				
MRI Normal (background)	0.017			
MRI Abnormal focal		0.709	0.551	0.912
MRI Abnormal non-specific		1.082	0.824	1.423
MRI Not done/not known		1.057	0.893	1.250
Cohort				
SANAD (background)	0.000			
UCL		0.269	0.109	0.666
ULB		0.570	0.208	1.560
EKUT		1.169	0.917	1.490
RCSI		0.647	0.430	0.974
Glasgow		0.871	0.722	1.052
ULIV		0.192	0.094	0.394
Australia		1.185	0.901	1.557

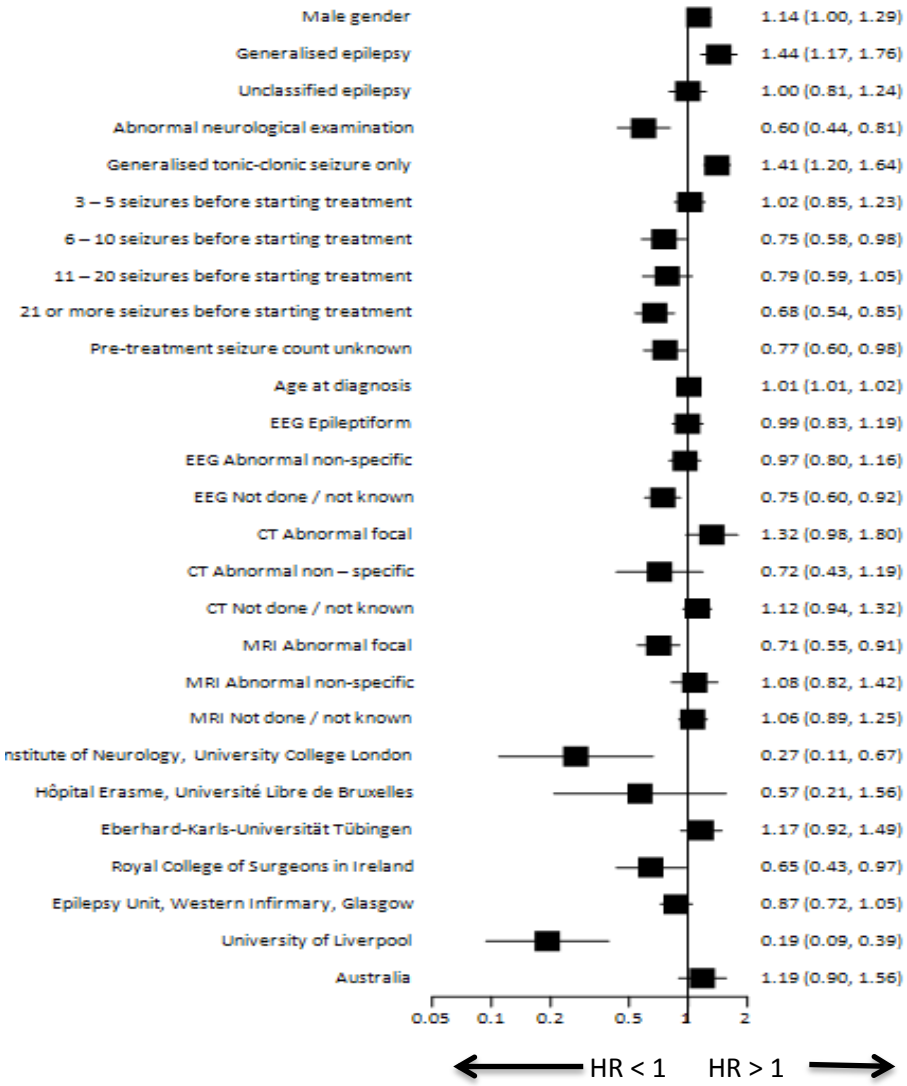


Figure 17. Forest plot (log scale) of results of Cox regression for time to onset of 12 months remission in newly diagnosed epilepsy after application of the first anticonvulsant. The following categories are assumed baseline for the categorical covariates: female sex, focal epilepsy, 1-2 seizures before starting treatment, normal EEG, normal CT, normal MRI and SANAD. Right lateral column demonstrates 95.0% CI for Hazard Ratio. HR > 1 indicates a shorter time to 12 months remission.

Due to the heterogeneity of the remission rates observed across the cohorts, a further Cox regression was performed and stratified by cohort. The parameters of the model were identical to the previously described model. The results of this model are presented in the table 30.

Table 30. Cox regression model for time to 12 month seizure remission in newly diagnosed epilepsy after the application of first well tolerated AED, stratified by cohort (n=1723)

Covariate	p-value	Hazard Ratio	95.0% CI for Hazard ratio	
			Lower	Upper
Male	0.049	1.140	1.000	1.299
Age at diagnosis	0.000	1.010	1.006	1.014
Abnormal neurological examination	0.001	0.598	0.441	0.812
Generalised tonic-clonic seizures only	0.000	1.402	1.198	1.640
Epilepsy type				
Focal (baseline)	0.000			
Generalised		1.487	1.213	1.822
Unclassified		0.996	0.804	1.235
Pre - treatment seizure count				
1 – 2 (baseline)	0.002			
3 – 5		1.013	0.845	1.214
6 – 10		0.750	0.575	0.978
11 – 20		0.787	0.590	1.050
21 or more		0.675	0.539	0.846
Pre - treatment seizure count unknown		0.768	0.600	0.983
EEG				
EEG Normal(baseline)	0.051			
EEG Epileptiform		1.001	0.838	1.196
EEG Abnormal non-specific		0.975	0.808	1.176
EEG Not done/not known		0.762	0.618	0.939
MRI				
MRI Normal (baseline)	0.070			
MRI Abnormal focal		0.808	0.657	0.994
MRI Abnormal non-specific		1.119	0.854	1.467
MRI Not done/not known		1.069	0.908	1.257

The stratified Cox regression model showed that age at onset, male sex, and generalised tonic-clonic seizures only are statistically significantly associated with a shorter time to 12 remission, whereas abnormal neurological examination was associated with a longer time to 12 month remission. Epilepsy type was also significantly associated with the outcome; generalised epilepsy had a positive direction of effect. Both MRI and EEG were not significantly associated with the outcome but remained in a final model.

4.4. Impact of methodology

4.4.1. Cohort

Both the logistic and Cox regression analyses described earlier in this chapter demonstrated that non-clinical factors are associated with 12 month remission in newly diagnosed epilepsy. The role of non-clinical variables such as the origin of the cohort, the length of the observation period, and the mode of case ascertainment will be assessed in this section.

The origin of the cohort does not have a clinical meaning, but it encompasses the way that the study was set up and conducted. The biggest contributor to our study was the SANAD study with 778 cases – 96.14% (n = 748) were included in the analysis. The next largest cohort was from Glasgow (n = 462), of which 43 cases (10.61%) were excluded. This indicates that cases can be lost both in the case of manual phenotyping and data transfer. Interestingly, a manually phenotyped cohort like ULIV had no losses at all, indicating a potential underlying selection bias, by selecting for initial genotyping cases with full clinical information available. The most significant proportional reduction of cases occurred in the ULB (34.48%), Australia (29.39%) and UCL (13.89%) cohorts. Due to the complicated and protracted data transfer from the clinical database, the ULB cohort lost proportionally the highest amount of cases. Data transfer was complicated by the complexity of linking values between the database used for clinical purposes and the EpiPGX eCRF. In the case of the cohort from Glasgow, losses were mostly due to incomplete (less than 12 months) follow-up due to a strain on services at certain periods of time. The Australian cohort, similarly to ULB and SANAD, was transferred electronically from the original research database. Cases from this cohort were lost due to data being incompatible with EpiPGX definitions. Table 31 summarises the total number of cases by cohort, as well as the number of cases each cohort contributed to the logistic and Cox regression analyses.

Table 30. Number of subjects included and excluded from the logistic and Cox regressions

Cohort	Total		Included in analysis	
	n	%	n	% out of total
UCL	36	1.9	31	86.11
ULB	29	1.5	16	65.52
EKUT	247	13.0	232	93.93
RCSI	73	3.8	69	94.52
Glasgow	462	24.2	413	89.39
ULIV	53	2.8	53	100.00
Australia	228	12.0	161	70.61
SANAD	778	9.5	748	96.14
Entire Cohort	1906	100.0	1723	100.00

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

4.4.2. Effect of method of case ascertainment

The complete research cohort contained a mixture of prospective and retrospectively ascertained cases. Out of 1906 cases, 1006 (52.8%) were prospectively ascertained. For an assessment of the effect of case ascertainment, only cases with complete information were used (n = 1723). The reasons for excluding cases were exactly the same as those described in this chapter for the logistic regression. Out of these 1723 cases, 909 (52.8%) were prospective. Only the SANAD and Australian cohorts were prospective. To assess the effect of the method of case ascertainment ,the final model of previously described logistic regression was refitted with an additional co-variate, “prospective v. retrospective”. Identical selection parameters as described in the logistic regression chapter were utilised.

The method of case ascertainment was statistically significantly ($p < 0.000$) associated with the occurrence of a 12-month remission. Similarly to the previously described logistic regression model, epilepsy type, abnormal neurological examination, presence of generalised tonic-clonic seizure only, pre-treatment seizure count, age at diagnosis, EEG, CT, MRI, follow-up length and origin of cohort were significantly associated with the treatment outcome. Prospectively recruited cases were more likely to experience remission (odds ratio 1.054, 95% CI 1.023 – 1.085).

4.4.3. Period of observation

Period of observation is an important methodological aspect as too short observation period would risk of failure to capture events of interest, on the other hand too long is associated with higher cost and logistic difficulties. Hence it is important to establish the minimal period required to capture most of the events of interest. *Speed et al* has assessed in their epilepsy pharmacogentic study and reported that 90% of patients with newly diagnosed epilepsy will experience remission within first 2 years after starting treatment (Speed et al., 2014a). This chapter will explore not only time required to a observe remission but a broader role of duration of whole follow up period.

To assess the effect of length of the period of observation on the treatment outcome three variables – time to onset of 12 month remission, follow-up length and length of remission were used.

Data on follow-up length were available in 1905 cases; it was only unavailable in one case. Its distribution is negatively skewed with an early peak and relatively steep fall off. The histogram is presented in Figure 18. Median follow-up length was 3.39 years (SD = 4.87; IQR =2.06 - 5.57), highlighting relatively good data capture and the probability of observing events of interest.

There were noticeable differences between cohorts in the period of observation, with the Australian cohort having the shortest median follow-up length (1.94 years) and ULIV (14.42 years) and RCSI (13.76 years) the longest. This was probably due to the fact that the Australian cohort is relatively recent compared to ULIV. ULIV and RCSI utilised direct retrospective phenotyping based on clinical notes and were able to trace back to the first anticonvulsant, hence they potentially had longer overall follow-up information. The SANAD study, was a pragmatic clinical trial and involved systematic data collection for a fixed time period but still individual follow up periods varied in-between patients.. The follow-up length stratified by the cohort are further summarised in Table 32.

The length of the follow-up was also assessed in the logistic regression model for a relationship with the 12-month period of remission. It was statistically significantly ($p <$

0.000, Exp(B) = 1.054, 95% CI 1.023 – 1.085) associated with remission. A longer length of follow-up was associated with a higher probability of remission.

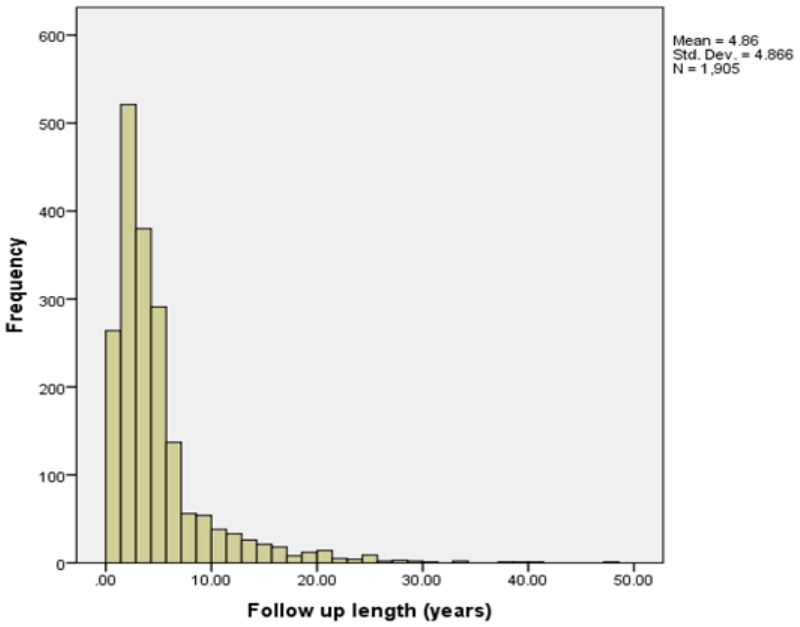


Figure 18. Histogram of distribution (frequency of cases) for follow-up length in years (n = 1905)

Table 31. Length of follow-up period in years stratified by cohort. IQR – interquartile range.

Cohort	Follow-up length (years)			
	Mean	Standard deviation	Median	IQR
UCL	8.25	7.06	7.36	1.68 - 12.92
ULB	3.87	3.50	3.08	1.28 - 5.30
EKUT	5.86	5.37	4.44	2.36 - 7.41
RCSI	13.76	8.15	13.62	7.68 - 18.57
Glasgow	5.63	4.62	4.37	2.41 - 7.62
ULIV	14.42	8.64	13.18	8.23 - 18.76
Australia	1.98	1.14	1.94	1.06 - 2.46
SANAD	3.32	1.52	3.06	2.09 - 4.48
Entire Cohort	4.86	4.87	3.39	2.06 - 5.57

UCL – University College London; ULB - Hôpital Erasme, Université Libre de Bruxelles; EKUT - Eberhard-Karls-Universität Tübingen; RCSI - Royal College of Surgeons in Ireland; University of Liverpool; SANAD - Standard and New Antiepileptic Drugs Study.

Data on time to onset of 12 month remission was available for all of the 982 patients experiencing remission; the median time to onset of remission was 0.0 years (mean = 0.35; SD = 1.07; IQR = 0 – 0.25, range 0 – 13.97), indicating immediate onset of remission.

A histogram (figure 19) illustrating frequency percentage distribution for time to onset of 12 month remission in years after application of the first well tolerated anticonvulsant shows a very steep drop after the first year followed by a tail lasting for several years. Ninety-five per cent of all remissions occurred within the 1.58 years from the onset of the treatment. The full range of time to remission in percentiles is presented in Table 33.

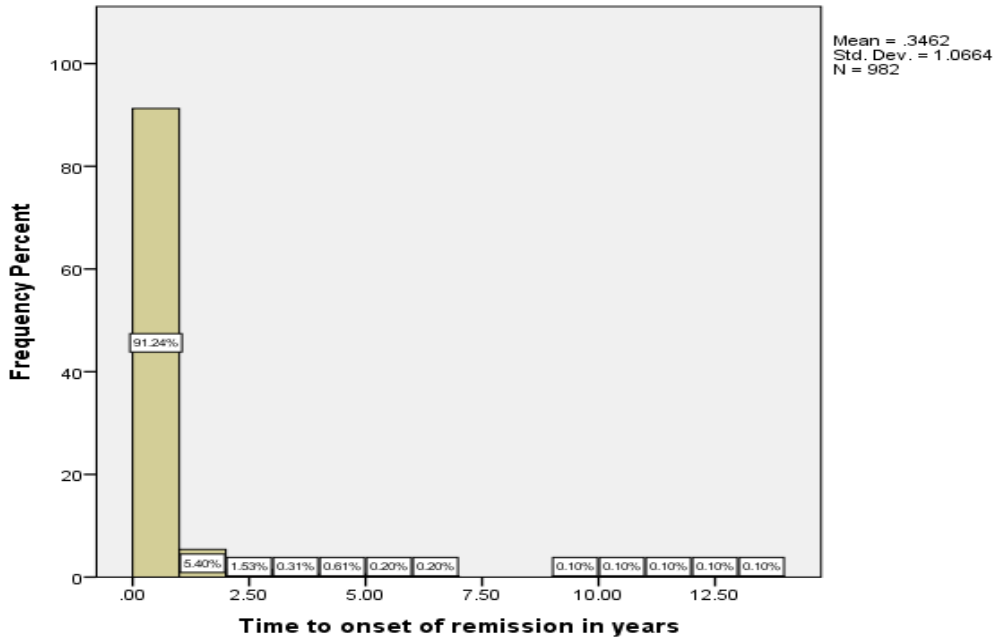


Figure 19. Histogram (n = 982) demonstrates frequency percentage distribution of patients over time who achieved 12 month remission after application of the first well tolerated anticonvulsant. Bin size is equal to one year. For majority of patients onset of remission was within a first year after application of anticonvulsant.

Table 32. Time to 12 month remission (years) separated into percentiles

	Percentile 05	Percentile 25	Percentile 50	Percentile 75	Percentile 90	Percentile 95	Percentile 99
Time to remission (days)	0	0	0	0.25	0.92	1.58	4.53

4.4.4. Duration of remission

It has already been shown that around 16% of patients with newly diagnosed epilepsy will have relapsing remitting course of disease with most of them experiencing 30 months long remission before relapse (Brodie et al., 2012). Those patients potentially represent a different phenotype and potentially different genetic association, particularly in pharmacogenomics research as those cases might not represent the effect of drug. Hence it is important to assess duration of remission as a separate methodological aspect to establish how long patients in remission should be observed for.

In total, 982 (56.3%) subjects experienced a 12-month remission after starting their first well-tolerated anticonvulsant. The median duration of remission on the first well-tolerated AED was 28.04 months (mean = 37.47; SD = 28.19; IQR = 18.92 - 46.52). There were differences between cohorts in the duration of remission, with UCL (mean = 84.06; median 68.60 months) and EKUT (mean = 53.91; median 42.48 months) having the longest, while Australian (mean = 27.71; median 24.90 months), ULIV (mean = 47.62; median = 24.97 months) and ULB (mean = 34.55, median = 25.94 months) had the shortest duration of remission. For 670 (68.2%) of the 982 patients with remission, the remission lasted for the entire duration of the follow-up period of observation, while 312 (31.8%) had a relapse. Subjects with continuous remission had the shortest period of follow up, whereas patients with initial remission but with subsequent relapse had the longest follow up period (median follow-up length in years for no remission = 3.66 years, remission with relapse = 5.20 and remission without recorded relapse 2.97). Duration of remission demonstrated a statistically significant difference between the

remission subgroups. Patients with initial remission but a relapse later, had a significantly shorter period of remission than those with sustained remission (independent sample median test $p < 0.001$, median length of remission in months for remission without a recorded relapse = 29.55 months and for 12 months remission with relapse = 24.80 months).

4.4.5. Assessment of effect of definition of remission

The application of an increased length of time of seizure freedom required for a case to be counted as being in remission resulted in a progressively reduced proportion of patients experiencing remission (1YR = 56.3%; 2YR = 43.5%; 5YR = 25.2%; 10YR = 12.2% and remission lasting for a whole period of observation = 38.4%). There were 197 patients who had follow-up for at least 10 years; 47.2% (n = 93) experienced seizure freedom for at least of 12 months. Only 12.2% (n = 24) experienced a 10-year seizure freedom. Data on the effect of changes in the length of seizure freedom on the proportion of patients classified as in remission are summarised in Table 34.

Table 34. Proportion of subjects with newly diagnosed epilepsy classified as experiencing remission after application of different periods of seizure freedom adjusted by the length of the follow up.

Length of follow up	Length of remission	n	%	95% CI
Whole cohort	No remission	762	43.7%	41.4% - 46.0%
	1 year remission	982	56.3%	54.0% - 58.6%
At least 2 year follow up	No 2 year remission	792	56.5%	53.9% - 59.1%
	2 year remission	610	43.5%	40.9% - 46.1%
At least 5 year follow up	No 5 year remission	419	74.8%	71.1% - 78.3%
	5 year remission	141	25.2%	21.7% - 28.9%
At least 10 year follow up	No 10 year remission	173	87.8%	82.7% - 91.8%
	10 year remission	24	12.2%	8.2% - 17.3%
Whole cohort	No remission	762	43.7%	41.4% - 46.0%
	Remission till last follow up date	670	38.4%	36.2% - 40.7%
	Remission but not for whole observation period	312	17.9%	16.1% - 19.7%

4.4.6. Impact of data transformation and upload

Data transfer between various digital databases is a potential source of systematic errors. Recording of data in particular is a potentially source of systematic errors especially if the approach in collecting parameters differs between datasets. This is a particularly important issue when multiple different digital sources are used. Often initial data after collection is further processed to create a specific derivation reflecting phenotype of interest. In some cases genetic studies are a continuation of previous older research hence differences in data handling at all levels (including phenotype derivation) can add some further complexity when interpreting results.

A comparison was made between 711 identical cases present in the EpiPGX and SANAD genetic study datasets. The results of this assessment have been summarised in Table 35. Absolute concordance was observed between the data sets for the basic demographic (age and gender) data and for epilepsy type, and for most of the binary co-variates a very good concordance were observed. Two variables describing seizure types (GTC only and other seizures) showed poor concordance between the data sets. Time to remission had a Lin's concordance coefficient of 0.6409, indicating poor concordance; this was due to a difference in how time to remission was calculated. In EpiPGX it was calculated as a time span from the first date of starting AED to the onset of the 12-month remission. In the SANAD study, the time to remission was calculated as a time span from the date of randomization to the last day of the 12-month period of seizure freedom.

Table 33. Concordance between compared data sets (SANAD original genetic derivation and EpiPGX) WP2 task 1 phenotype derivation

Binary co-variates		
Covariate	**Kappa value**	**Significance**
Abnormal neurological examination	1.000	0.000
Generalised epilepsy	0.962	0.000
Unclassified epilepsy	0.943	0.000
Gender	1.000	0.000
Generalised tonic-clonic seizures only	-0.136	0.000
Partial seizures only	0.986	0.000
Other seizures	-0.008	0.619
Absences or myoclonic seizures with GTC	0.721	0.000
CT abnormal	0.721	0.000
EEG abnormal	0.930	0.000
Continuous co-variates		
	Lin's Concordance Coefficient	**95% CI**
Time to remission	0.6409	0.6037 – 0.6753
Age at the onset	0.9996	0.9995 – 0.9997

A good concordance (kappa = 0.738, p = 0.000) for the treatment outcome (status of 12-month remission) was observed, with 86.8% (n = 617) of cases matching each other. Fifty (7.03%) had no remission in EpiPGx, but were marked as a remission in the SANAD original genetic derivation. In 27 (3.8%) cases there were no treatment outcomes in EpiPGx, whilst there were recorded outcomes in the SANAD genetic derivation; the opposite was observed in 17 (2.4%) patients. There were no discordant cases with patients recorded as no remission in SANAD but as a remission in EpiPGX. As some of the statistical methods rely on proportions, a comparison between the proportions of patients experiencing remission was also performed. Remission was observed in 56.8% of EpiPGx cases and 63.9% of the SANAD genetic dataset, whereas there was no

remission in 39.2% versus 33.6% subjects respectively, and no outcome available in 3.9% and 2.5% of cases respectively.

4.5. Genome-wide association analysis

4.5.1. Description of GWAS

Localisation-related or focal epilepsy is the commonest form of epilepsy. The aim of this genome-wide association study was to investigate genomic biomarkers associated with treatment outcomes after the application of the first appropriate and adequate anticonvulsant in those with focal epilepsy. The description of methods related to genotyping, imputation, QC and research cohort assembly have been provided separately in the methods section.

This GWAS was not in the EpiPGX initial analysis plan, hence only cases with locally held genotyping results were included. There were 935 cases included in this analysis. Remission was defined as at least 12 months of continuous seizure freedom after exposure to the first anticonvulsant.

4.5.2. Sample size calculation

To assess the power of GWAS, a retrospective sample size calculation was performed using several different scenarios (Skol et al., 2006).

The following parameters were used for the power calculation:

- Number of subjects experiencing 12 months remission = 524
- Number of subjects with no remission = 411
- Significance level = 0.0000001
- β = 0.8
- Prevalence = 0.44
- MAF = 0.05 or 0.1
- Genotype relative risk = 1.5; 2.0; 3.0; 4.0

As can be seen from table 36 the GWAS was underpowered to detect association in the case of low genetic relative risk for both additive and dominant model. For relative

risk = 2, 3 and 4 (MAF = 0.05) in the additive model and 3 and 4 (MAF = 0.05) for the dominant model combination of prevalence, RR and MAF failed to correspond to a possible genetic model. The same result was observed for RR = 2, 3, and 4 (MFA =0.1) under additive model and 4 (MAF= 0.1) under dominant model.

The GWAS was adequately powered to detect association under dominant model with RR 2 (MAF = 0.05 and 0.1) and RR 3 (MAF = 0.1).

Table 34. Results of power calculation for GWAS for focal newly diagnosed epilepsy applying both additive and dominant models of inheritance

MAF	RR	Power	Model
0.05	1.5	5.0%	Additive
0.05	2	NA	Additive
0.05	3	NA	Additive
0.05	4	NA	Additive
0.05	1.5	3.0%	Dominant
0.05	2	96%	Dominant
0.05	3	NA	Dominant
0.05	4	NA	Dominant
0.1	1.5	39%	Additive
0.1	2	NA	Additive
0.1	3	NA	Additive
0.1	4	NA	Additive
0.1	1.5	22%	Dominant
0.1	2	100%	Dominant
0.1	3	100%	Dominant
0.1	4	NA	Dominant

A further retrospective power calculation for was performed for EpiPGX WP2 task 2 GWAS, using the same scenarios, except the number of subjects experiencing 12 months remission was 850 and no remission 664. EpiPGX WP2 task 1 GWAS for a binary treatment outcome (remission v. failure) had 31% power (MAF = 0.05; RR 1.5;) to detect association with an additive one-stage model, and 23% power for a dominant inheritance model. The model assuming dominant inheritance mode with RR 2.0 (MAF= 0.05) had 100% power to detect association in a one-stage design. The additive model

with MAF = 0.1 and with a genetic relative risk of 1.5 had an 88% power to detect association. The same calculation was performed for the model with dominant mode of inheritance. For RR of 1.5 there was 71% power to detect association; for the RR = 2 and 3 model the power was 100%.

4.5.3. Results of GWAS

4.5.3.1. Clinical covariate analysis

There were 524 patients (56.0%) with at least a 12-month remission and 411 patients (44.0%) who had not experienced a remission. 928 cases were included in the logistic regression model; 6 cases were excluded due to lack of information in terms of time from first seizure to first AED.

The demographic data for cases included in the logistic regression model and subsequent GWAS are presented in Table 37. The following clinical factors had a significant association ($p<0.05$) with treatment outcome: neurological examination, generalised tonic-clonic seizures only, EEG, MRI, CT and age at onset.

The final model after backward selection using Wald statistics was applied included the following co-variates: neurological examination results, generalised tonic-clonic seizures only, EEG, MRI, CT results and pre-treatment seizure count. Results from the logistic regression are presented in Table 38. The variable 'prospective v. retrospective' was also included in the GWAS model, although it was not statistically significant in the final logistic regression model. It did however demonstrate significance in the full model and was retained in this model.

Table 35. The demographic data for cases included in the logistic regression model for 12 month remission after application of first AED in newly diagnosed focal epilepsy (n = 928) and subsequently GWAS.

Categorical variables		**n**	**%**
Gender	Males	482	52.0
	Females	446	48.0
Positive family history	Present	154	16.6
Abnormal neurological examination	Abnormal	67	7.2
Febrile seizures	Present	32	3.5
Simple or complex partial only	Yes	275	29.6
Generalised tonic-clonic seizures only	Yes	279	30.1
Unclassified tonic-clonic seizures	Yes	131	14.1
Other seizures	Yes	2	0.2
EEG Normal (baseline)		388	41.8
EEG Epileptiform abnormalities		159	17.1
EEG Abnormal non-specific		220	23.7
EEG Not known / Not done		161	17.3
MRI Normal (baseline)		443	47.7
MRI Focal abnormalities		169	18.2
MRI Abnormal non-specific		84	9.1
MRI Not known / not done		232	25.0
CT Normal (baseline)		393	42.3
CT Focal abnormalities		104	11.2
CT Abnormal non-specific		28	3.0
CT Not known / not done		403	43.4
1 – 2 Seizures before treatment (baseline)		166	17.9
3 – 5 seizures before treatment		210	22.6
6 - 10 seizures before treatment		124	13.4
11 - 20 seizures before treatment		87	9.4
21 + seizures before treatment		258	27.8
Pre – treatment seizure count unknown		83	8.9
Prospective v. retrospective	Prospective	584	62.9
Continuous variables		**Median**	**IQR**
Age at diagnosis	Years	41.72	29.01 - 56.16
Time from first seizure to first AED	Years(log)	2.61	2.20 - 3.11

Table 36. Multivariable logistic regression model for 12-month seizures remission in newly diagnosed focal epilepsy (n = 928). Normal EEG, CT, MRI and 1 – 2 seizure before the treatment were used as baseline variables (n = 928)

Co – variates	p value	Odds ratio	95% C.I.for Odds Ratio	
			Lower	Upper
Abnormal neurological examination	0.005	0.426	0.235	0.774
Generalised tonic-clonic seizures only	0.068	1.421	0.974	2.073
EEG				
EEG Epileptiform	0.00	1.355	0.878	2.092
EEG Abnormal non-specific		0.778	0.531	1.141
EEG Not done/not known		0.422	0.275	0.648
MRI				
MRI Abnormal focal	0.008	0.448	0.278	0.723
MRI Abnormal non-specific		1.005	0.570	1.773
MRI Not done/not known		0.812	0.551	1.195
CT				
CT Abnormal focal	0.004	1.759	0.968	3.197
CT Abnormal non-specific		0.512	0.208	1.263
CT Not done/not known		1.646	1.179	2.297
Pre-treatment seizure frequency				
3 – 5 seizures before treatment	0.000	0.747	0.459	1.217
6 – 10 seizures before treatment		0.374	0.211	0.666
11 – 20 seizures before treatment		0.507	0.265	0.971
21 + seizures before treatment		0.272	0.153	0.484
Pre-treatment seizure count unknown		0.284	0.145	0.554
Age at diagnosis	0.000	1.025	1.016	1.034
Prospective v. retrospective	0.056	1.398	0.991	1.973

4.5.3.2. Genetic analysis

Figures 20 and 21 show the Manhattan plot and qq plot for the GWAS (not adjusted for clinical factors). The genomic inflation factor (λgc) was 1.019. There were no SNPs with genome-wide significant association ($p < 5x10^{-8}$). There were 45 SNPs that had a p-value of less than $5x10^{-5}$. Genes within 50kb of SNPs with genome-wide p-values less than 5×10^{-5} are presented in Table 39. There were no SNPs with significant association in the exon regions of genes, and none with clear biological significance.

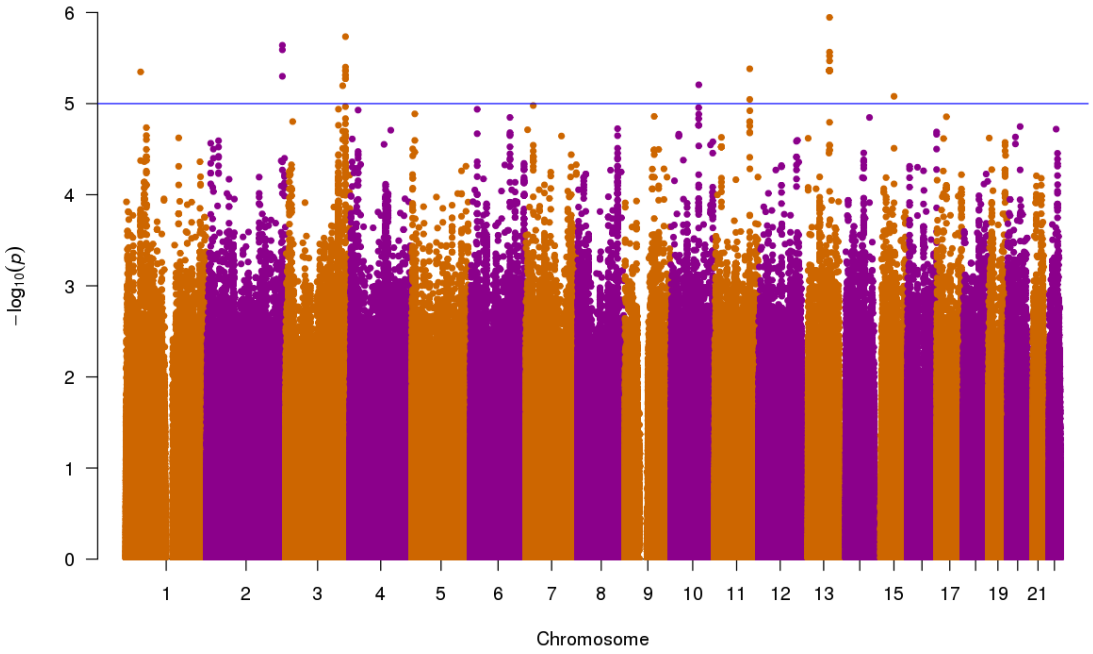


Figure 20. Manhattan plot for binary GWAS analysis for remission on the first well-tolerated anticonvulsant in newly diagnosed focal epilepsy, without adjustment for clinical factors

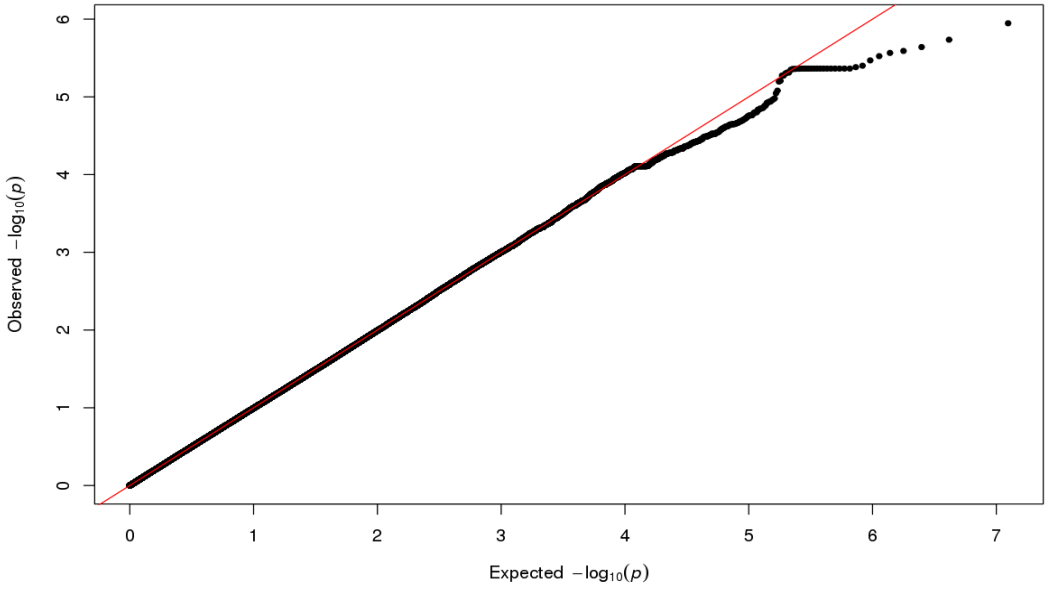


Figure 21. A qq plot for binary GWAS, not adjusted for clinical factors, for the 12 month remission on the first well-tolerated anticonvulsant in newly diagnosed focal epilepsy. The red line represents the null hypothesis of no association, whereas observed p-values (negative logarithm) are plotted on the y axis and expected p-values (negative logarithm) are plotted on the x axis. Deviation is only observed in the upper tail area indicating sufficient control of population stratification (genomic inflation factor = 1.019). The GWAS was adjusted for the first five principal components.

Table 37. SNPs with p-value less than $5x10^{-5}$ from binary unadjusted GWAS for 12 month remission on the first well-tolerated anticonvulsant in newly diagnosed focal epilepsy together with corresponding genes within 50 kb

SNP	Chromosome	MAF	P- value	Genes within 50 kb from SNP
rs76100028	3	0.077	1.47E-06	C3orf70, VPS8
rs10166451	2	0.371	1.83E-06	EFHD1, EIF4E2
rs114442102	2	0.372	2.06E-06	EFHD1, EIF4E2
rs59264451	3	0.074	3.22E-06	C3orf70, VPS8
chr11:107982503:D	11	0.193	3.36E-06	ACAT1, CUL5, NPAT
chr3:184791599:D	3	0.078	3.53E-06	C3orf70, VPS8
rs4687016	3	0.078	3.53E-06	C3orf70, VPS8
rs55882333	1	0.469	3.64E-06	HPDL, MUTYH, TOE1, ZSWIM5
rs744306	3	0.078	3.96E-06	C3orf70,VPS8
rs747507	3	0.078	3.96E-06	C3orf70, VPS8
rs7607256	2	0.371	4.08E-06	EFHD1, EIF4E2
rs2004208	3	0.078	4.33E-06	C3orf70, VPS8
rs2004207	3	0.078	4.33E-06	C3orf70, VPS8
rs17204959	15	0.414	6.83E-06	RORA
rs35219733	11	0.209	7.39E-06	ACAT1, CUL5, NPAT
rs2390716	7	0.166	8.66E-06	IL6, TOMM7
rs7624642	3	0.082	8.90E-06	VPS8
rs753796	10	0.360	9.13E-06	CDHR1, LRIT1, LRIT2, RGR
rs9460589	6	0.273	9.53E-06	CDKAL1
rs11727253	4	0.203	9.73E-06	SEL1L3
rs144267602	11	0.306	9.89E-06	C11orf65, KDELC2

Figures 22 and 23 show the Manhattan plot and qq plot for the GWAS adjusted for clinical factors. The genomic inflation factor was 1.061. Similar to the unadjusted analysis, there were no genome- wide significant hits ($p < 5x10^{-8}$). There were 37 SNPs that gave a p-value of less than 5×10^{-5}. Genes within 50kb of SNPs with genome-wide suggestive p-values from the GWAS (adjusted by clinical co-variates) have been presented in Table 40.

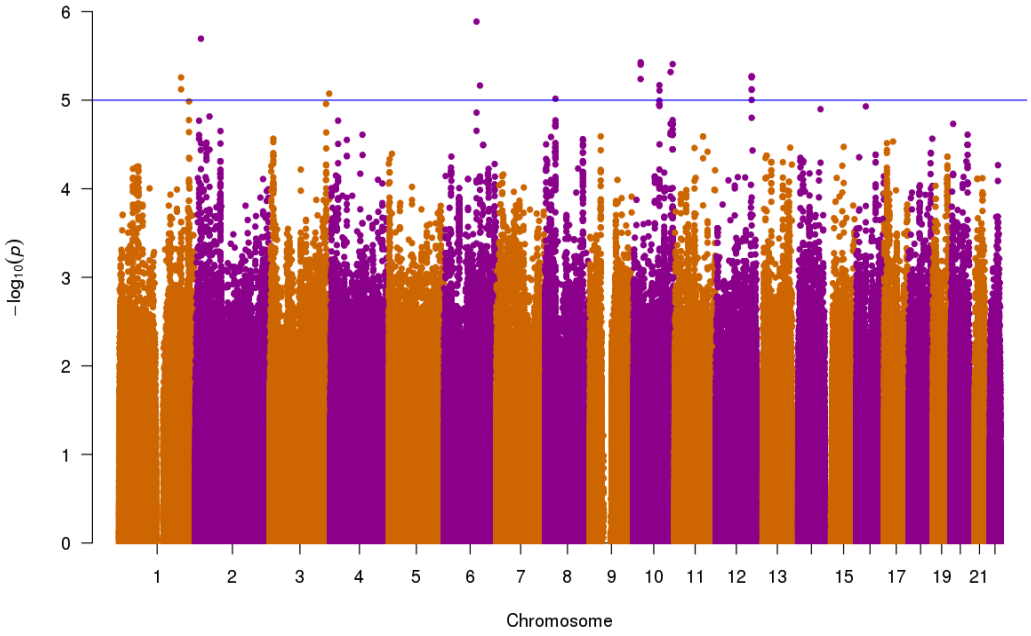


Figure 22. Manhattan plot for binary GWAS analysis for 12 month remission on the first well-tolerated anticonvulsant in newly diagnosed focal epilepsy, after adjustment for significant clinical factors

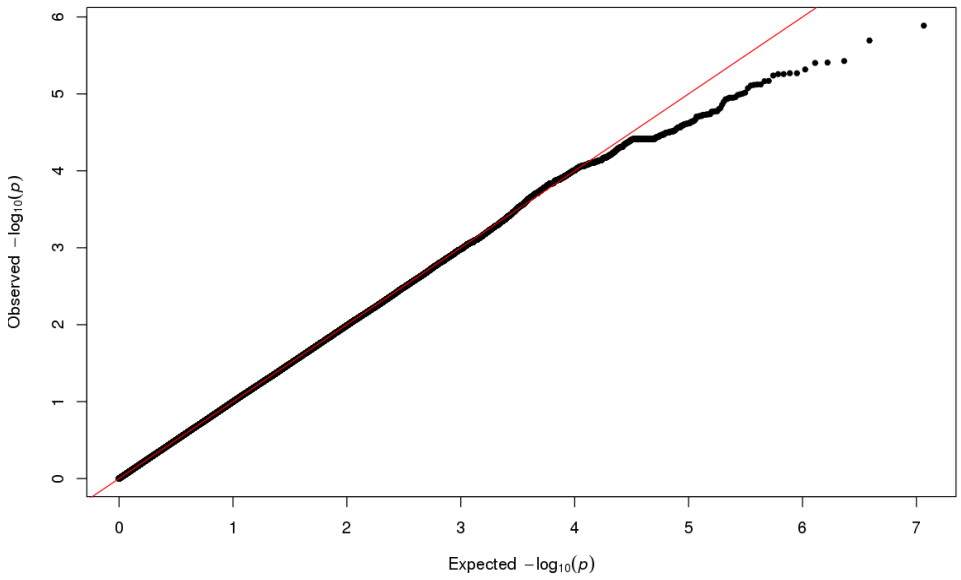


Figure 23. A qq plot for binary GWAS adjusted for significant clinical factors for 12 month remission on the first well-tolerated anticonvulsant in newly diagnosed focal epilepsy. The red line represents the null hypothesis of no association, whereas observed p-values (negative logarithm) are plotted on the y axis and expected p-values (negative logarithm) are plotted on the x axis. Deviation is only observed in the upper tail area indicating sufficient control of population stratification (genomic inflation factor = 1.061). The GWAS was adjusted for the first five principal components.

Table 40. SNPs with p-value less than $5x10^{-5}$ from binary GWAS for 12 month remission on first well-tolerated anticonvulsant in newly diagnosed focal epilepsy after adjustment for significant clinical factors.

SNP	Chromosome	MAF	P-value	Genes within 50 kb from SNP
rs11096544	2	0.334	9.89E-07	NT5C1B, RDH14
rs10828664	10	0.444	1.89E-06	KIAA1217
rs72835240	10	0.121	1.99E-06	NPS
rs12268215	10	0.450	2.01E-06	KIAA1217
rs1874438	12	0.489	2.79E-06	MAP1LC3B2
rs7968805	12	0.489	2.79E-06	MAP1LC3B2
rs11068100	12	0.495	2.86E-06	MAP1LC3B2
rs12725494	1	0.134	2.87E-06	PPFIA4, TMEM183A, TMEM183B
rs11014141	10	0.440	3.00E-06	KIAA1217
rs753796	10	0.360	3.56E-06	CDHR1, LRIT1, LRIT2, RGR
rs9400994	6	0.341	3.59E-06	ROS1, VGLL2
rs11068099	12	0.490	3.98E-06	MAP1LC3B2
rs11584217	1	0.136	3.98E-06	CYB5R1, PPFIA4, TMEM183A, TMEM183B
rs11068101	12	0.487	4.04E-06	MAP1LC3B2
rs713601	3	0.144	4.48E-06	FAM43A
rs7137033	12	0.488	5.33E-06	MAP1LC3B2
rs4508149	10	0.359	5.45E-06	LRIT1, LRIT2, RGR
rs12747999	1	0.088	5.56E-06	DUSP5P, RN5S1, RN5S2, RN5S3, RN5S4, RN5S5, RN5S6, RN5S7, RN5S8, RN5S9, RN5S10, RN5S11, RN5S12, RN5S13, RN5S14, RN5S15, RN5S16, RN5S17
rs73053778	3	0.095	5.94E-06	CHRD, CLCN2, EIF2B5, EIF4G1, FAM131A, POLR2H, PSMD2, SNORD66, THPO
rs753795	10	0.360	6.06E-06	CDHR1, LRIT1, LRIT2, RGR
rs7899757	10	0.360	6.06E-06	CDHR1, LRIT1, LRIT2, RGR
rs7904309	10	0.360	6.06E-06	CDHR1, LRIT1, LRIT2, RGR
rs7911510	10	0.476	6.29E-06	RPS3AP5
rs140775682	16	0.096	6.37E-06	ITGAM
rs11160087	14	0.122	6.88E-06	LGMN, RIN3
rs1530632	2	0.281	8.40E-06	EPAS1
rs7314617	12	0.486	8.71E-06	MAP1LC3B2
rs4412590	1	0.178	9.28E-06	RN5S1, RN5S2, RN5S3, RN5S4, RN5S5, RN5S6, RN5S7, RN5S8, RN5S9, RN5S10, RN5S11, RN5S12, RN5S13, RN5S14, RN5S15, RN5S16, RN5S17, RNF187
rs72835241	10	0.118	9.28E-06	NPS
rs10033588	4	0.200	9.43E-06	SEL1L3

SNPs with a p-value of less than 5×10^{-5} were re-tested for association with the 12 months remission under a dominant inheritance model, both with and without adjustment for significant clinical factors. Similarly to the initial analysis, both univariate and adjusted GWAS was adjusted for the first five principal components. Full results for the SNPs included in the GWAS with the dominant inheritance model have been summarised in table available in appendix 4 on page 238. SNPs with a p-value of less than 5×10^{-5} in both the adjusted and unadjusted analysis using dominant inheritance are presented in Table 41.

Table 38. Results of re-testing SNPs with a p-value of less than 5 x 10^{-5} from the additive model for association with outcome under a dominant inheritance model. On a left side of the table values from univariate analysis are presented whereas those adjusted for significant clinical factors are shown on the right.

CHR	rsid	position	info	p-value univariate		CHR	rsid	Position	info	p-value adjusted
13	chr13:85111634:D	85111634	0.993656	5.66E-07		13	chr13:85111634:D	85111634	0.993656	1.63E-06
14	rs11160087	93129982	0.964623	6.92E-06		14	rs11160087	93129982	0.964623	6.76E-06
13	rs12855432	85111586	0.998517	7.33E-07		13	rs12855432	85111586	0.998517	2.18E-06
11	rs144267602	108319016	0.991432	5.10E-07		11	rs144267602	108000000	0.991432	8.15E-06
13	rs1446766	85112660	0.992194	8.18E-07		13	rs1446766	85112660	0.992194	2.46E-06
13	rs1446767	85110202	0.999908	7.33E-07		13	rs1446767	85110202	0.999908	2.18E-06
13	rs1446768	85110189	0.999913	7.33E-07		13	rs1446768	85110189	0.999913	2.18E-06
13	rs1446770	85108206	1.000000	7.33E-07		13	rs1446770	85108206	1.000000	2.18E-06
13	rs1446782	85075320	0.999171	7.08E-07		13	rs1446782	85075320	0.999171	3.07E-06
10	rs4077084	86420972	0.989580	8.34E-06		10	rs4077084	86420972	0.98958	6.65E-06
3	rs59264451	184787463	0.997319	9.64E-07		3	rs59264451	185000000	0.997319	6.23E-06
13	rs7337244	85108915	0.999995	7.33E-07		13	rs7337244	85108915	0.999995	2.18E-06
13	rs7337326	85109087	0.999991	7.33E-07		13	rs7337326	85109087	0.999991	2.18E-06
13	rs7337490	85109106	0.999991	7.33E-07		13	rs7337490	85109106	0.999991	2.18E-06
13	rs7338700	85109002	0.999995	7.33E-07		13	rs7338700	85109002	0.999995	2.18E-06
13	rs7339022	85109110	0.999991	7.33E-07		13	rs7339022	85109110	0.999991	2.18E-06
3	rs76100028	184789338	0.997140	4.21E-07		3	rs76100028	185000000	0.997140	3.92E-06
13	rs7991045	85080910	0.998700	7.40E-07		13	rs7991045	85080910	0.998700	2.83E-06
13	rs7996369	85111713	0.997633	7.33E-07		13	rs7996369	85111713	0.997633	2.18E-06
13	rs7997475	85112098	0.995518	8.18E-07		13	rs7997475	85112098	0.995518	2.46E-06

13	rs9319058	85108979	0.999995	7.33E-07		13	rs9319058	85108979	0.999995	2.18E-06
13	rs9531612	85109868	0.999940	7.33E-07		13	rs9531612	85109868	0.999940	2.18E-06
13	rs9531613	85110793	0.999867	7.33E-07		13	rs9531613	85110793	0.999867	2.18E-06
13	rs9531615	85111407	0.999803	7.33E-07		13	rs9531615	85111407	0.999803	2.18E-06
13	rs9546751	85109272	0.999500	6.06E-07		13	rs9546751	85109272	0.999500	1.81E-06
13	rs9602496	85108738	1.000000	7.33E-07		13	rs9602496	85108738	1.000000	2.18E-06

The rs4077084 SNP retained significance in all of the analyses performed (both using the additive and dominant inheritance model, with and without adjustment for significant clinical factors). It is located in the intergenic region of chromosome 10 (BP = 86420972) with a minor allele frequency of 0.061. In our analysis it was not genotyped directly, but rather imputed and had an imputation INFO score of 0.989 suggesting reliable imputation accuracy. A plot of the region surrounding this SNP, using the software package 'Locuszoom' (Pruim et al., 2010), plot showed that other SNPs with a reasonable degree of linkage disequilibrium with it did not have similarly low p-values. This, together with its relatively low minor allele frequency, may suggest a spurious result. The Locuszoom plot is presented as Figure 24.

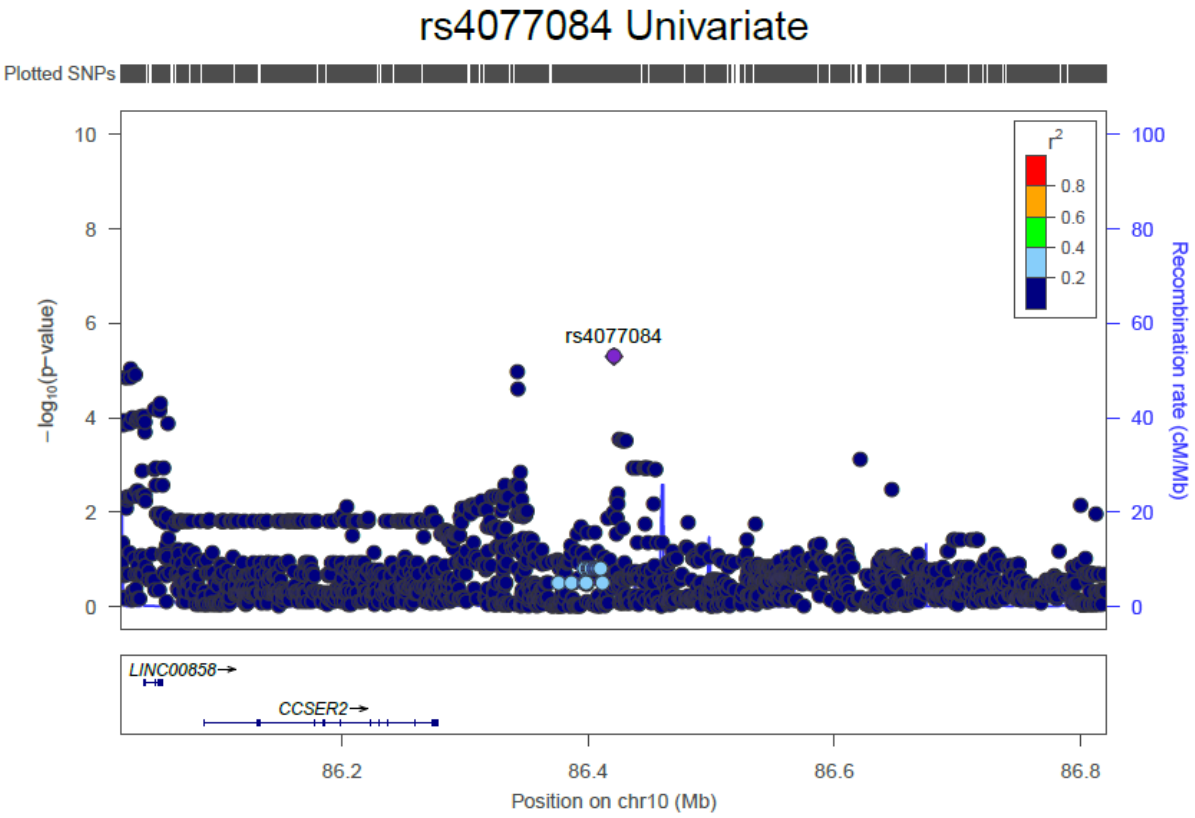


Figure 24. "Locuszoom" plot for the rs4077084 based on results of univariate GWAS analysis

The rs753796 SNP had a GWAS suggestive p-value in both univariate and adjusted GWAS with an additive inheritance model, but no significance in the dominant model. It is located in chromosome 10 (BP = 86028463), MAF = 0.360277 and its INFO score =

0.997041. It is an intergenic SNP located within 50kb from the following genes – CDHR1, LRIT1, LRIT2, and RGR. The Locuszoom plot is presented in Figure 25. There is no clear biological significance for this SNP.

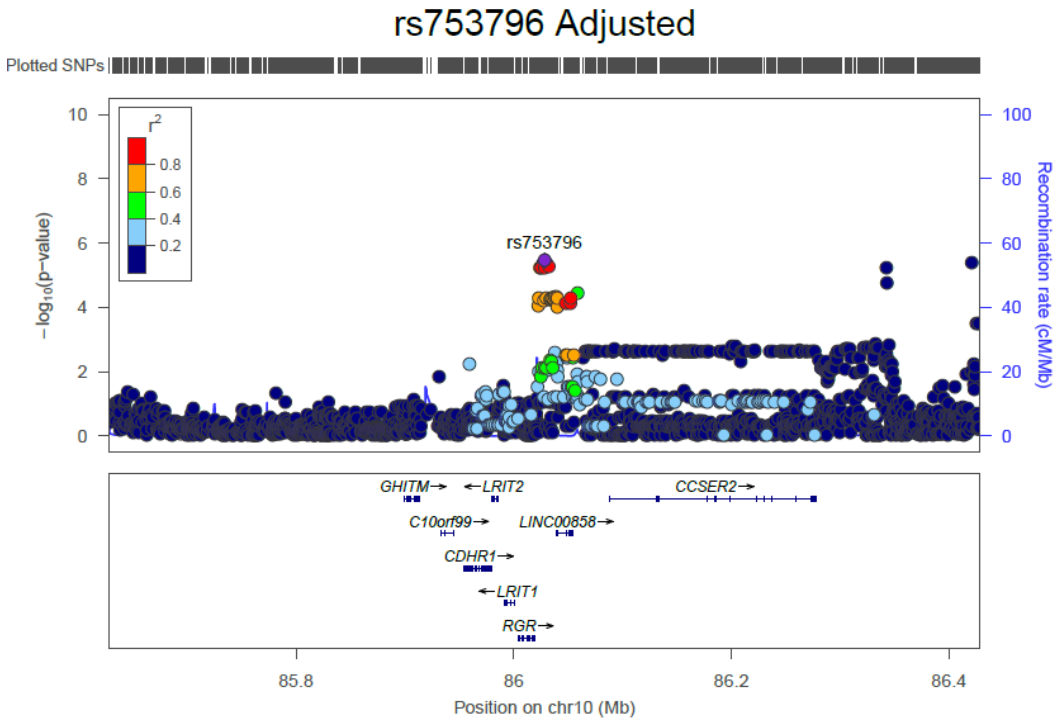


Figure 25. "Locuszoom" plot for the rs753796 based on results of GWAS analysis adjusted for significant clinical covariates

The rs73053778 SNP (BP = 184068305; MAF = 0.0951807; INFO score = 0.982405) had a low p-value in adjusted GWAS with the additive (p = 5.94E-06) and dominant (p = 7.51E-06) inheritance model, but not in either unadjusted analyses. This SNP is 500B downstream intron variant for a chloride voltage-gated channel 2 (*CLCN2*) gene. The Locuszoom plot for this SNP is presented in Figure 26.

In the additive model, unadjusted GWAS there were several SNPs located in close proximity to each other on chromosome 3, with a p-value of less than 5×10^{-5}. Their data are summarised in Table 42. The rs76100028 SNP had also a p-value of 3.92E-06 in the additive model with adjustment for significant clinical factors.

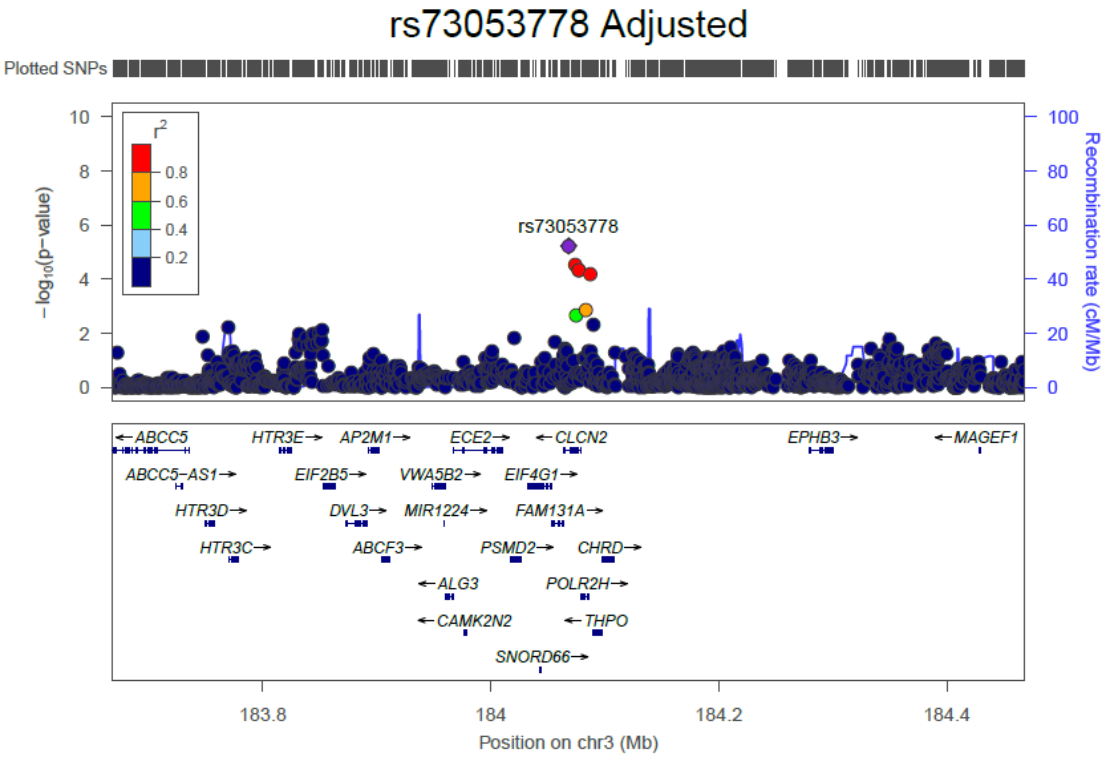


Figure 26. "Locuszoom" plot for the rs73053776 based on results of GWAS analysis adjusted for significant clinical covariates

Table 39. List of SNPs with a p-value of less than 5 x 10^{-5} clustering on chromosome 3 in close proximity to each other

CHR	SNP	BP	P	INFO	MAF
3	rs76100028	184789338	1.47E-06	0.997140	0.077367
3	rs4687016	184792444	3.53E-06	1.000000	0.077803
3	rs744306	184789748	3.96E-06	0.998469	0.078341
3	rs747507	184790127	3.96E-06	0.998684	0.078341
3	rs2004207	184790414	4.33E-06	0.997468	0.078161
3	rs2004208	184790407	4.33E-06	0.997456	0.078161
3	rs7624642	184561541	8.9E-06	0.994769	0.081797

All SNPs retained suggestive p-values in the dominant unadjusted analysis, but only rs76100028 retained a suggestive p-value in the dominant adjusted analysis. The Locuszoom plot for the SNP (rs76100028) with the smallest p-value is presented in Figure 27. All those SNPs are either in proximity or located in VPS8 (vacuolar protein sorting-associated protein 8 homolog) gene. In mammals there is in limited information

about biological function of this protein, current evidence is suggestive that it might be related to CORVET complex have a role in endosomal trafficking (Perini et al., 2014).

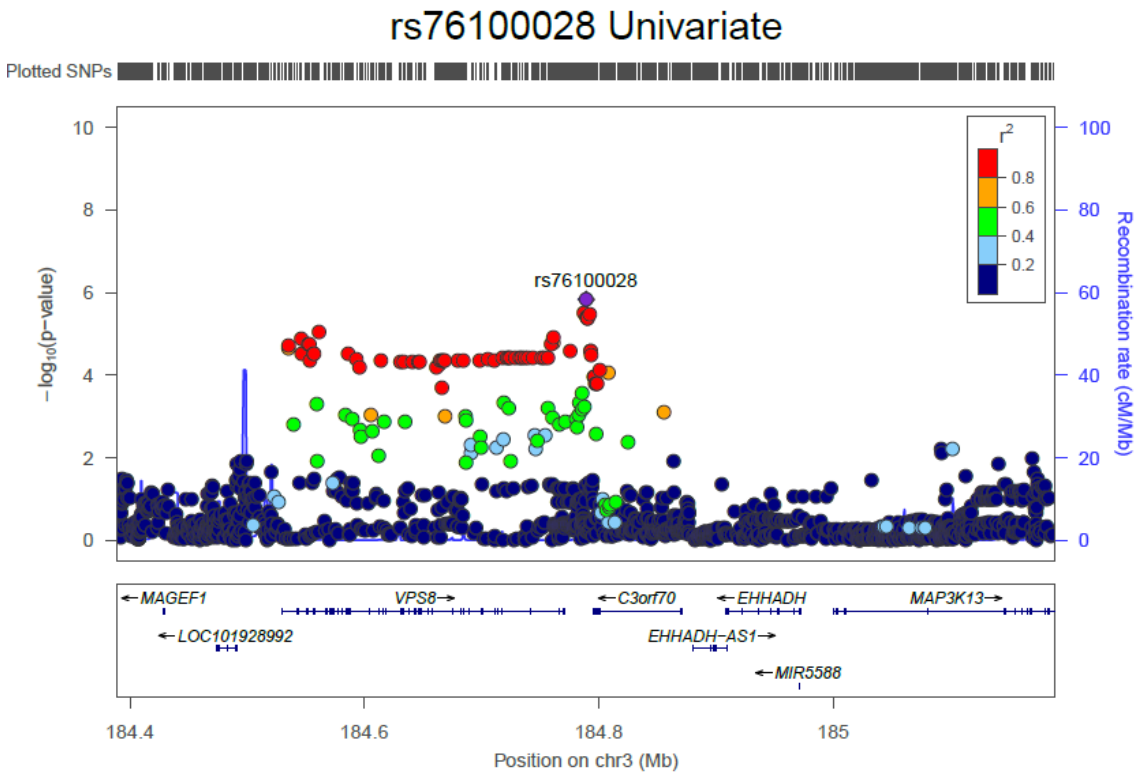


Figure 27. "Locuszoom" plot for the rs76100028 based on results of GWAS analysis without an adjustment for significant clinical covariates

A SEL1L family member 3 (SEL1L3) gene had intronic hits in both unadjusted and adjusted GWAS. Suggestive SNPs were not overlapping between analyses. The SNP rs11727253 was flagged up in an unadjusted analysis (p = 9.72784E-06; BP = 25770524; INFO score = 0.987954; MAF = 0.203088); it failed to retain significance in analysis with the dominant model. In the adjusted additive model, rs10033588 (p = 9.43009E-06; BP = 25754026; INFO score = 0.989671; MAF = 0.199519) retained a suggestive p-value in the dominant adjusted model but not in the dominant univariate model. Both SNPs are presented as locuszoom plots in figures 28 and 29.

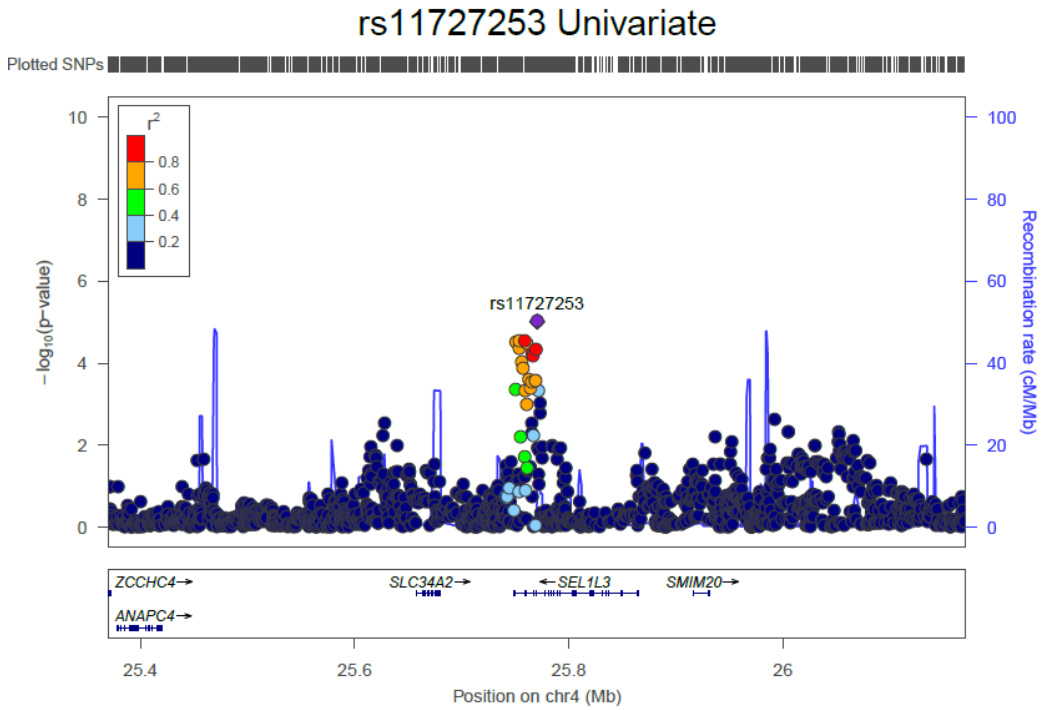


Figure 28. "Locuszoom" plot for the rs11727253 based on results of GWAS analysis without an adjustment for significant clinical covariates.

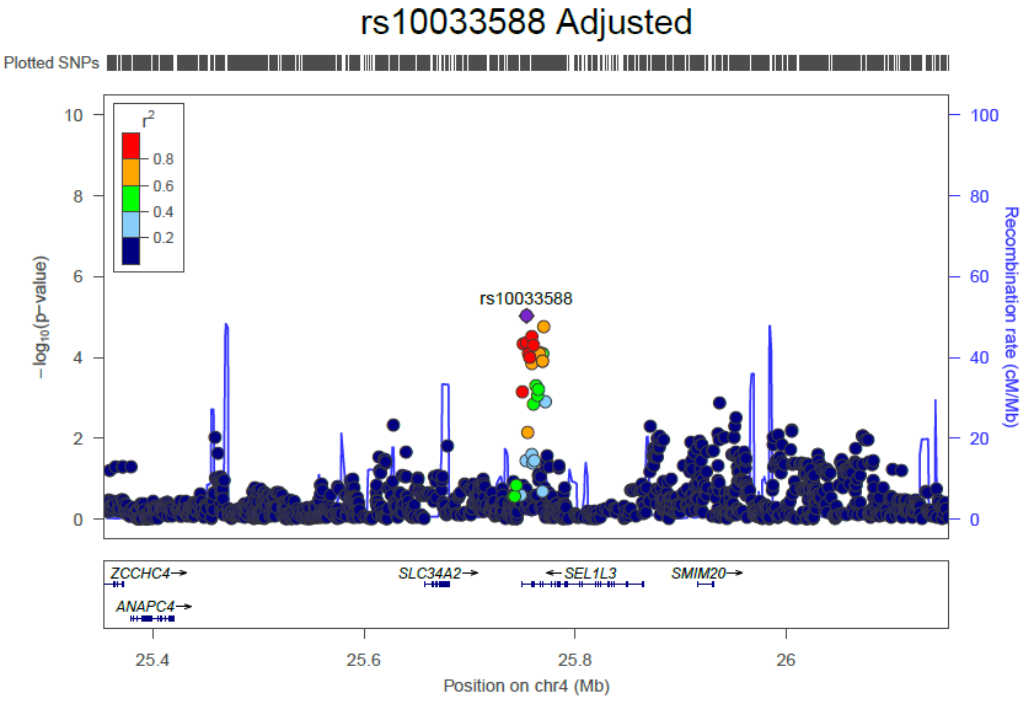


Figure 29. "Locuszoom" plot for the rs10033588 based on results of GWAS analysis adjustment for significant clinical covariates.

Locuszoom plots for all SNPs are available in the appendix 5 (univariate) and 6 (after adjustment for significant clinical factors) on pages 253 and 264 respectively.

5.0. Discussion

The work described in this thesis comprised three principal analyses, investigation of clinical factors associated with outcome in newly-diagnosed epilepsy, issues of methodology and their impact on results of studies that amalgamate existing data from multiple sites, and a genome wide association study that provisionally explored genetic contribution to the likelihood of 12-month remission in people with focal epilepsy. The general discussion will tackle each of these areas in turn.

5.1. Study of clinical factors

This work assessed the association between 12-month remission after application of the first well-tolerated anticonvulsant in newly diagnosed epilepsy, and multiple routine clinical factors. A significant association was demonstrated between remission and several clinically important factors, which are discussed more in detail below.

5.1.1. Remission

Our study reported a remission rate for the first well-tolerated AED in patients with newly diagnosed epilepsy to be 56.3%. This result is in line with previous work on newly diagnosed epilepsy which has reported the remission rate for the first AED to be slightly above 50% (Cockerell et al., 1995, Brodie et al., 2012, Collaborative Group for the Study of, 1992). Lack of deviation from previously published research adds validity to our results and suggests that our data are representative of newly diagnosed epilepsy; from a clinical perspective our data replicate previous knowledge of relatively good prognosis in regards of seizure control.

From the point of view of reliability, a significant proportion of cases (54.3%) in our cohort was transcribed directly from existing databases rather than be manually re-phenotyped. Thus, there would be a reasonable question as to whether there was any impact of data transfer and whether it introduced any significant bias during this

process. This was assessed using the SANAD dataset as it was the largest of our cohorts and we had access to the original study data. Reassuringly, remission in SANAD patients showed only minimal and insignificant discordance between original and transcribed data and no cases of no remission were turned into remission. The effect of data transfer is discussed in more detail in further sections.

Our overall research cohort was created by amalgamation of several different existing cohorts, and further analysis showed that there were statistically significant differences in proportions of patients experiencing remission.

Cohorts from UCL (81.8%), ULB (76.5%) and ULIV (84.9%) showed relatively high no remission rates; in contrast, the cohort from Australia (72.5%) had a noticeably higher rate of remission. A higher rate of no remission is likely related to the method of case ascertainment and the nature of the cohort.

The difference in proportion of remissions between some of the cohorts (i.e. SANAD, Glasgow, Australia) comprising the EpiPGX research cohort has previously been described (Shazadi et al., 2014). Several potential causes of this difference were already highlighted by the authors of that study, including factors such as a difference in recruitment and environment of the study (randomised trials vs. observational studies) or differences in local policies, drug selection and diagnostic approach. The same conclusions can be applied to our research cohort, plus differences in background of the population sampled can also be added. These differences illustrate the importance and the impact of case ascertainment and local clinical factors, and justify the adjustment for different origins of cohorts in the analysis of association.

Treatment response were assessed after application of the first well tolerated AED hence there were no cases in a cohort which would have been censored due to failure due to adverse reactions. Furthermore cases where the minimal effective dose were achieved but no further up titration had been undertaken due to ADRs but patients still had seizures were classified as no remission. This was due to fact that according to the EpiPGX definitions ADR had to occur within the first 6 months and lead to a dose reduction or drug withdrawal. For this analysis if patient experienced adverse effects

this AED were not classified as well tolerated. This created a situation where cases for survival analysis were only censored due to failure to achieve remission.

Both the analysis of association (logistic and Cox regression) as well as the genome wide association study relied on a 12-month remission as an outcome, hence they were specifically adjusted for origin of cohort. Furthermore, separate versions of the Cox regression stratified by origin of cohort were performed.

In summary, we demonstrated heterogeneity in distribution of the primary outcome (12-month remission) between research cohorts, lending support to the need to account for differences between centres in large multinational pharmacogenetic research initiatives.

5.1.2. Age at diagnosis

Previously published epidemiological observations have indicated two peaks of onset of epilepsy; one occurring in the early years of life and the other in elderly patients (Kotsopoulos et al., 2002, Neligan et al., 2012). In our cohort, we observed only a single relatively early peak around adolescence which is probably due to fact that most cohorts primarily recruited adults, but not children or geriatric population. Median age at diagnosis observed in our cohort was 33.57 years (mean 36.03), which is not too different to the age observed in other clinical trials for newly diagnosed epilepsy (Steinhoff et al., 2005, Bill et al., 1997, Christe et al., 1997, Rosenow et al., 2012, Kwan et al., 2011).

The cohort from Australia had the highest age at diagnosis, whereas the EKUT cohort had the lowest. The lower age for diagnosis at EKUT cohort probably reflects a research interest in genetic background of generalised epilepsy. The biggest single contributor to our research cohort was the SANAD study, which similarly observed a single peak of age distribution (Marson et al., 2007), probably reflecting that it was a clinical trial.

Age at diagnosis was significantly associated with likelihood of remission and time to 12-month remission in logistic and Cox regression models respectively. Furthermore, it

retained significance in sensitivity analysis when SANAD cases were excluded. It was also significantly associated with 12-month remission in the logistic regression model performed on focal epilepsy cases before GWAS. Direction of effect in all analyses of association was the same, indicating that older age at diagnosis was associated with higher probability of 12-month remission. Our observation would be in line with previous studies that have already shown that elderly patients have a more favourable prognosis (Stephen et al., 2006). From a data quality point of view, age at diagnosis showed high concordance when transcribed cases from SANAD were compared to the original genetic dataset.

In summary, age at diagnosis was robustly associated with a higher likelihood of 12-month remission in newly diagnosed epilepsy.

5.1.3. Gender

Previous epidemiological observations have shown a trend towards a higher incidence of epilepsy in males than females (Kotsopoulos et al., 2002, Hauser et al., 1993). In our study, there were marginally more males than females, which is consistent with previous reports. In addition, epidemiological studies have observed gender differences in some epilepsy syndromes, with generalised epilepsy being more commonly observed in females and symptomatic localisation-related epilepsy more common in males (Christensen et al., 2005). In our cohort, we also observed a higher proportion of females than males in the generalised epilepsy group. In both the logistic and Cox regression models for 12-month remission and time to remission after application of the first AED, gender approached the threshold for statistical significance, but failed to reach it.

5.1.4. Epilepsy type

The majority of patients (70.6%) in our cohort had focal epilepsy, followed by 17.9% with generalised epilepsy and 11.5% with unclassfied epilepsy. Those observations are consistent with already published epiedemiological data from Europe, where the majority of patients experience focal seizures/epilepsy (33–65%), followed by generalised (17-60%) and unclassifiable (2-8%) seizures (Forsgren et al., 2005). The combined research cohort is comprised of data taken from several studies conducted by various centres. The majority of the large cohorts, like SANAD (both arms), Glasgow, and Australia did not have specific exclusion criteria based on epilepsy type, but the EKUT cohort and a portion of the ULIV cohort, specifically those from the ReJuMEC study (Thomas et al., 2014), were originally enrolled to investigate specific epilepsy types.

Assessment of the reliability of data transfer showed almost perfect (the kappa value for generalised epilepsy = 0.962 and for unclassified epilepsy = 0.943) concordance for epilepsy type (McHugh, 2012). A limitation of this assessment is that in both the SANAD genetic derivation and EpiPGX phenotype derivation, epileptic syndromes were split into just three basic categories. Thus, potentially, data transfer for more elaborate syndromic categories might not have achieved such good concordance.

Epilepsy type was significantly associated with both likelihood of remission and time to 12-month remission in logistic and Cox regression, with a direction of effect favouring a more positive association for generalised epilepsy. This observation contradicts previously published studies (Collaborative Group for the Study of, 1992, Park et al., 2014) which reported no association of epilepsy type with treatment outcome. Arm B from the SANAD study provided 32.8% of all cases with generalised epilepsy, which was surpassed only by the EKUT cohort (34%). In a sensitivity analysis after removal of all SANAD cases, this association remained significant, indicating that it is not observed due to the SANAD study.

5.1.5. Neurological examination / sequelae

Abnormal neurological examination in cases of newly diagnosed epilepsy has been reported in around 15% of cases (Del Felice et al., 2010, Heller et al., 1995, Collaborative

Group for the Study of, 1992). It was observed in 8% of cases in our study, which is slightly lower than in previously reported data. There is a wide range between cohorts going from 4.4% in the Australian cohort to 24.1% in the ULB cohort. This probably represents differences in sampling population and case ascertainment. Of note, cases with progressive neurological conditions were excluded from EpiPGX. Abnormal neurological examination was statistically significantly associated with a lower probability of attaining 12-month remission in newly diagnosed epilepsy in both logistic and Cox regression.

5.1.6. Time from first seizure to first AED

Median time from the first seizure to first AED was 11.5 months. This observation is similar to previously published data from the CAROLE study, which showed that the majority of patients were diagnosed with epilepsy within seven months from the first seizure (range 1.5-30 months) (Jallon et al., 2001). This clinical factor failed to reach statistical significance in logistic and Cox regression for 12-month remission.

5.1.7. Investigations

There was noticeable variability between centres with respect to investigations, for example the proportion of patients with an abnormal focal MRI would range from as high as 69.0% to only 12.3%. Such a variability likely reflects diversity in the populations sampled, routine clinical practice, and data collection. For example, the Australian data was imported from a research database, and their CRF collected data from MRI but not CT results. Similarly, the higher proportion of abnormal epileptiform EEG results in the EKUT data could simply reflect a higher proportion of generalised epilepsy in this cohort. Those differences demonstrate the extent of variability between cohorts.

From a clinical perspective, results showed that CT and MRI had reasonably good concordance for focal abnormalities but were poor for non-specific abnormalities. However, it should be noted that this study was not designed to evaluate this particular issue. Furthermore, as we utilised only the initial imaging studies, the result is

potentially biased toward significant acute changes rather than specific epileptiform abnormalities that are often quite subtle. Previous modelling work carried out by colleagues in Liverpool (Bonnett et al., 2012) on arm A of the SANAD study indicated a potential association between abnormal imaging studies and time to 12-month remission. We were able not only to replicate this, but also to demonstrate the direction of effect for a lower likelihood of 12-month remission in cases with an abnormal focal MRI. This result was confirmed in the sensitivity analysis after removal of the SANAD cases. In a Cox regression stratified by cohort, MRI lost statistical significance but was still retained in a final model. Our assessment of imaging studies for newly diagnosed epilepsy was limited by the fact that we excluded all cases with progressive neurological conditions. A significant proportion of those cases likely would have abnormal imaging studies (e.g. brain tumours).

As for EEG, we showed that it is significant and that the direction of effect would be towards a lower probability of 12-month remission in cases where EEG was not done. This observation could be potentially explained by the fact that patients with higher seizure frequency and worse prognosis are not referred for EEG as there are fewer diagnostic doubts. Alternatively patients with poorer prognosis in general or those with focal epilepsy are not referred routinely for EEG by their doctors.

In summary, we showed that MRI results are a prognostic factor for 12 month remission for patients with newly diagnosed epilepsy after the application of a first well tolerated AED. As MRI is now considered as the routine imaging modality this result has the potential to influence daily clinical care. Observed differences between cohorts would support the introduction of a more standardised approach to the collection and interpretation of imaging studies in large pharmacogenetic cohorts, particularly as it might be a routinely used clinical marker that is significantly associated with treatment outcomes or in future employed in creation of endophenotypes.

5.1.8. Pre – treatment seizure count

The majority of patients (43.8%) experienced between one and five seizures before starting their treatment; this is consistent with previously published research (Del Felice et al., 2010, Collaborative Group for the Study of, 1992, MacDonald et al., 2000).

We observed a relatively high proportion of cases with unknown seizure number before starting treatment (15.8%). This could be explained by several factors related to clinical presentation, data collection, or transfer. As EpiPGX relied on the retrospective collection of data based on clinical records, it was impossible to resolve the problem of unknown data thus illustrating some of the limitations of this approach.

From a clinical point of view, certain seizure types have a higher propensity to be underreported. Myoclonic and absence seizures had a higher proportion of unknown pre-treatment seizure count. This probably reflects the higher frequency and minimal intrusiveness of those seizure types. Conversely, complex partial and simple partial seizures have a high proportion of 21 or more definite seizures before treatment, but a relatively low unknown seizure count. As expected, primarily generalised tonic-clonic and secondarily generalised tonic-clonic seizures have a lower unknown seizure count than any other seizure type.

For logistic and Cox regression analyses assessing associations with 12-month remission, pre-treatment seizure count is statistically significantly associated with the likelihood of remission. This observation is in line with previous work based both on the SANAD study and NGPSE (MacDonald et al., 2000, Bonnett et al., 2014b, Bonnett et al., 2012). In a sensitivity analysis after exclusion of the SANAD cases, significance was lost; this implies a degree of uncertainty about this association. NGPSE suggested seizure density (number of seizures within six months of the index seizure leading to inclusion in study) as a significant factor associated with treatment response in newly diagnosed epilepsy (MacDonald et al., 2000), although this factor was not collected in EpiPGX or assessed in this analysis.

5.1.9. Seizure type

Both logistic and Cox regression showed that presence of generalised tonic-clonic seizures only is positively associated with both likelihood of remission and time to remission. This seizure type was derived from cases with either secondarily generalised tonic-clonic seizures, primarily generalised tonic-clonic seizures, or unclassified generalised tonic-clonic seizures, without any other seizure types reported. There is an obvious question of whether the effect observed was potentially due to an underlying better prognosis of generalised epilepsy. In the final model of logistic and Cox regression, generalised tonic-clonic seizures only was significant along with epilepsy type indicating that the generalised tonic-clonic only seizure type has additional explanatory value itself.

Generalised tonic-clonic seizures only were one of the variables that were assessed for concordance between the original SANAD genetic derivation and the EpiPGX derivation. It produced a kappa value of -0.136 indicating more disagreement than agreement (McHugh, 2012). This is probably due to a difference in how seizure types were derived in SANAD and EpiPGX, as we included not only primary generalised tonic-clonic seizures, but also secondarily generalised tonic-clonic seizures in this category. Reassuringly, the generalised tonic-clonic only seizure type retained its significance after removal of the SANAD cases in the sensitivity analysis, indicating that the association is not affected by data transfer.

5.1.10. First well-tolerated antiepileptic drug

The most commonly prescribed anticonvulsants were lamotrigine, valproate and carbamazepine. This reflects both the setting up of the study (involving multiple historical studies) as well as the fact that those anticonvulsants are widely available throughout Europe (Baftiu et al., 2015). Furthermore all of them are established first-line treatments for newly diagnosed epilepsy and both valproate and lamotrigine can be used for both focal and generalised epilepsy.

Our study was not expressly designed to investigate the specific efficacy of any particular anticonvulsant, nor at this moment in time has EpiPGx interrogated data

regarding genetic biomarkers for treatment response to a specific drug. Nevertheless, EpiPGx as a consortium has aimed to assess certain aspects related to individual drugs, but that is outside of the scope of this thesis.

Treatment efficacy of anticonvulsants has not improved since the introduction of new drugs over last two decades (Schmidt and Sillanpää, 2012, Shorvon and Luciano, 2007). It is arguable that treatment response is rather a broad phenomenon which is either partially or fully unrelated to drug selection. Furthermore, it has been shown that the strongest predictor for a treatment outcome is failure of the first anticonvulsant itself (Kwan and Brodie, 2000). Hence it made sense to look into a broad treatment response rather than to narrow it down to an individual drug.

5.2. Logistic and Cox regression

Logistic regression was applied as a model with a dichotomous outcome for 12-month remission after the application of the first well-tolerated AED in newly diagnosed epilepsy. The Cox regression model was applied to time to onset of remission after application of the first well-tolerated anticonvulsant.

Reassuringly, most of the significant covariates overlapped between logistic and Cox regression, providing a degree of confidence that the observed associations are true. Only follow up length and results of CT head were significant in logistic but not in Cox regression. When the Cox regression was stratified by cohort, MRI lost its significance but was still retained in the model and gender became statistically significant. The direction of effects for all covariates was the same, providing an additional degree of reassurance.

Our study represents one of the largest research cohorts of newly diagnosed epilepsy. Only analysis based on the SANAD trial has had more combined cases. However, our study included only treatment naïve patients, whereas between 10% and 18% of patients recruited in SANAD had previous exposure to AEDs (Bonnett et al., 2014b, Bonnett et al., 2012). Furthermore, we were able to add an extra layer of

information about the impact of imaging studies by extracting additional information from the original trial data.

From a practical point of view, we were able to confirm previously observed associations with important clinical factors. As can be seen from table 43 below, most of the significant clinical factors have already been separately described by various previous research studies.

Table 40. Covariates significantly associated with 12-month remission after application of the first well-tolerated anticonvulsant in newly diagnosed epilepsy and their direction of effect based on logistic and Cox regression results.

Covariate	Logistic regression	Cox regression	Previously reported in literature	Reference
Male	Not significant	Not significant	Yes	(Bonnett et al., 2012, Annegers et al., 1979)
Age at diagnosis	Positive association	Shorter time to onset of remission	Yes	(Mohanraj and Brodie, 2006, Bonnett et al., 2012)
Abnormal neurological examination	Negative association	Longer time to onset of remission	Yes	(Okuma and Kumashiro, 1981, Elwes et al., 1984, Sillanpää, 1993, Bonnett et al., 2014b, Bonnett et al., 2012)
Generalised tonic-clonic seizures only	Positive association	Shorter time to onset of remission	Yes	(Mattson et al., 1996)
Epilepsy type	Generalised epilepsy – positive direction	Generalised epilepsy - positive direction	Yes	(Mohanraj and Brodie, 2006)
Pre-treatment seizure count	1 -2 seizures before treatment - positive direction	1 -2 seizures before treatment – positive direction	Yes	MacDonald et al., 2000
Cohort	The same direction of effect in both Cox and logistic regression			
EEG	EEG not done - negative direction	EEG not done - negative direction	Yes	(Shafer et al., 1988, Bonnett et al., 2014b, Bonnett et al., 2012)
CT	CT abnormal focal positive direction	Not significant in Cox regression	Yes	(Elwes et al., 1984, Bonnett et al., 2012)
MRI	Focal abnormality - negative direction	Focal abnormality - negative direction	Yes	(Bonnett et al., 2012, Spooner et al., 2006)
Follow up length	Positive association	Not significant in Cox regression		

Sophisticated statistical predictive modelling based on SANAD has previously demonstrated that maximal discriminatory capacity for models is 70% (Bonnett et al., 2014b, Bonnett et al., 2012), indicating that there are other factors predictive of the outcome which at the moment are not known.

Our sample size (n = 1723) of logistic and Cox regressions was relatively large and provides a degree of confidence that they are not underpowered (Hosmer et al., 2013, Peduzzi et al., 1996, Peduzzi et al., 1995). One specific limitation of the Cox regression was that it failed to uphold the assumption of proportionality of hazards over time. Another limitation of our study is that our research cohort was largely comprised of cases from Glasgow and SANAD; hence potentially observed associations would be due to the effect of initial cohort. To assess this, we did a sensitivity analysis by excluding SANAD (the largest contributor) in which all clinical factors except pre-treatment seizure number and CT retained significance. Regression as a method also allowed adjustment for observed significant differences between cohorts.

In summary, logistic and Cox regression analysis confirmed previously described clinical factors associated with treatment outcome in newly diagnosed epilepsy after the start of the first well-tolerated anticonvulsant. It also showed the direction of effect for abnormal focal MRI towards worse treatment prognosis in newly diagnosed epilepsy.

5.3. Methodology

Case ascertainment and the length of the period of observation are influenced by factors related to the initial aims and methodology applied to the original cohort. In our research, we had a limited ability to influence these factors as we relied on historically assembled cohorts.

5.3.1. Origin of cohort and method of case ascertainment

Our research showed that the origin of the cohort was highly statistically significantly associated with 12-month remission. This variable reflects local factors, methods, and aims of the original study. Often, genetic cohorts are assembled to answer an initial

question and then later re-used for a variety of different purposes. Our study demonstrated that it is feasible to create a large multi-national research cohort based on old and very different historical studies. Nevertheless, this is potentially at the expense of introducing of bias and researchers should account for this. We have demonstrated that the method of case ascertainment is associated with the measured outcome and, as expected, the prospective method of follow up was linked to a higher chance of observing a 12 month remission. This effect probably arises from the SANAD study (a randomised clinical trial) and the Australian cohort. All cohorts with a higher failure rate are based in tertiary care centres with an epilepsy surgery programme. Institutions like UCL, the Walton Centre (Liverpool), and ULB serve as tertiary referral centres thus creating a recruitment bias. In addition, in all of those cohorts, a retrospective design was used. Furthermore, they were initially developed for the research of treatment-resistant epilepsy. For example, the cohort from the University of Liverpool was composed of the Epilepsy Biobank (which is a random sample of patients attending out-patient clinics), ReJuMEC (Refractory Juvenile Myoclonic Epilepsy Cohort), and the so-called "Department of Health" study exploring the pharmacogenetics of clobazam and vigabatrin. ReJuMEC and Department of Health studies primarily targeted drug-resistant epilepsy; hence using retrospective phenotyping allowed the tracing back of cases to the time of diagnosis and first AED (Thomas et al., 2014, EU Clinical trial register accessed 24/04/2016 2004). Such an approach invariably introduced a selection bias.

Future genetic studies of newly diagnosed epilepsy should aim to avoid selection bias and employ prospective recruitment methods only. The effect of case ascertainment in epilepsy research has previously been demonstrated. Abimbola and colleagues assessed the effect of methodological aspects on remission and intractability in cohorts of newly diagnosed epilepsy (Abimbola et al., 2011). The authors showed that a convenience sample can introduce bias into the study and preferably a prospective population-based incident cohort should be used. They also suggested that the exclusion of patients with progressive neurological conditions leads to an increased proportion of immediate remission. Our research excluded cases with progressive neurological conditions and

had a lower proportion (8.0%) of patients with neurological deficit than in other studies (13.0% - 15.7%) of newly diagnosed epilepsy (Heller et al., 1995, Collaborative Group for the Study of, 1992). On the other hand but not unexpectedly our cohort had the same proportion as SANAD study arm A (Marson et al., 2007).

Our research cohort was a mixture of patients from observational studies and clinical trials. In cases of established epilepsy, it has already been demonstrated that observational studies are prone to various biases as well as possessing a bigger and more varied treatment effect size compared to randomised controlled trials (Maguire et al., 2008).

In summary, our research confirmed that the origin of the cohort and the method of case ascertainment are significantly associated with the probability to observe measured outcomes, and hence should be adjusted for.

5.3.2. Length of observation

Length of observation is also an important methodological aspect. It has been shown in the past that it takes time for those patients included in research to enter into sustained terminal remission (Abimbola et al., 2011). Our study also showed that a longer follow up period is associated with a higher probability of observing a 12-month remission. To capture more than 95% of all first 12-month remissions in newly diagnosed epilepsy cases, the observation period should be a minimum of two years. Out of 982 patients experiencing 12-month remission, 312 of them later relapsed; this raises questions around the significance of non-sustained remission. To further strengthen the argument about a minimal period of observation, the median time to relapse was 29.3 months; thus indicating that a very short period of observation could potentially miss it.

Most of the patients in our cohort started their treatment after the year 2000, but almost one quarter (23.3%) of the patients started their treatment earlier than that, with the earliest commencing their AED in the 1960s. During this period, treatment

strategies have changed significantly and a lot of new AEDs have entered into daily practice. However, there has been no major improvement in treatment efficacy (Shorvon and Luciano, 2007, Schmidt and Sillanpää, 2012). This issue could be a potential problem when adjusting for relevant clinical factors (particularly imaging studies) which tend to see technical advancement over the time. Furthermore, a period of recruitment can become an issue for studies assessing selective drug response as anticonvulsant usage changes over time, although this is not applicable to our work.

5.3.3. Duration of remission

Our definition of remission required at least 12 months of seizure freedom. However, there is no clear consensus in the case of newly diagnosed epilepsy about the duration of seizure freedom that is considered to constitute 'remission'.

Drug-responsive epilepsy has been defined by the ILAE task force as seizure freedom of 12 months, or three times the longest pre-treatment inter-seizure interval, whichever is longer (Kwan et al., 2010). The requirement for exactly 12 months of seizure freedom is rather arbitrary from a biological point of view, and based on evidence related to patients' quality of life (Kwan et al., 2010) and statutory limits on driving. On the other hand, using the longest inter-seizure interval might not be practical for retrospective studies as data is often limited. We assessed the effect on a proportion of subjects experiencing remission after applying different definitions of remission. Applying longer durations of seizure freedom resulted in a lower proportion of patients being counted as in remission (1yr = 56.3%; 2yr = 43.5%). The limitation of this approach is that not all patients were followed up for the same period of time. Arguably, some patients who entered into sustained remission were not followed up in the long term, hence skewing longitudinal data towards treatment-resistant cases. The median follow up period in our cohort was 3.39 years (IQR = 2.06 - 5.57) meaning that most of the subjects were followed up for at least two years.

In summary, our study demonstrates that methodological aspects can have a statistically significant association with the measured outcome. Future studies of newly

diagnosed epilepsy should account for those aspects and preferably utilise, where possible, a prospective community-based design and use a minimum period of observation. Whereas cautious assembly of large research consortiums utilising historical cohorts is feasible, it requires a careful approach. Results of short retrospective hospital-based genetic studies in newly diagnosed epilepsy should be interpreted with caution due to the propensity of introducing a whole spectrum of biases.

5.3.4. Impact of data transfer

Large scale genetic studies frequently rely on historical research cohorts. They have often been assembled for a different purpose and apply different phenotypes and measurement methods. During the process of assembly of a new research cohort, old cohorts are often transformed and fused together to form a new and much bigger one. To my knowledge, no previous large scale genetic study has assessed how this process affects the data. This step has the potential to introduce bias during the process of assembly, transformation, and fusion of the new cohort. Hence, it was decided to assess concordance between the original derivation of SANAD for a different genetic study and our phenotype derivation of SANAD cases in EpiPGX.

Our assessment was limited by several factors, such as the original genetic dataset used for comparison containing only a limited number of variables that could be compared. Both datasets represent phenotype derivation rather than original raw data. On the other hand, the fact that those phenotype derivations were used for both GWAS and analysis of association means that it would make sense to compare them.

The observed concordance for the binary treatment outcome of 12-month remission was very good, but not perfect (kappa = 0.738, p = 0.000). This is probably due to slight differences between definitions applied to different cohorts. Our research measured treatment response to the first well-tolerated anticonvulsant, whereas Speed et al. (2014) utilised the SANAD genetic dataset definition of remission which was not necessarily observed with the first well tolerated AED.

Another approach that could be used to assess how representative our sample is of the original cohort would be to assess proportions of patients being in remission to check if any significant deviation from the original cohort had occurred. Original publicly available data (Marson et al., 2007) from the SANAD study reported the proportion of treatment failures (SANAD arm A – 49%, arm B – 42%). As this is a binary outcome with minimal variability it can be compared between studies and datasets. The SANAD study did not collect DNA from all patients at the beginning, but from a selected population later during the study, potentially introducing a degree of selection bias. In the EpiPGx SANAD cohort,12 month remission was observed in 56.4%, failures were 39.7%, and no data was available in 3.9%. The SANAD study formed the UK cohort of the first large prospective GWAS in epilepsy (Speed et al., 2014a). The authors of the study reported the following treatment outcome proportions for their full UK cohort; remission in 66.7% of subjects and failure in 33.3%. Those differences raise the question of whether selective recruitment (i.e. collection of DNA) of patients from old studies will not introduce a degree of bias, adding a further argument in favour of prospective genetic studies.

Poor concordance (Lin's concordance coefficient = 0.6409) was observed for the time to remission data (McBride, 2005). Discordance between cohorts was due to the difference in how time to remission was calculated in both studies. In the SANAD study it was calculated from the first date of AED randomisation and last day of the 12-month period of seizure freedom. In our study, the time span was taken from the first date of AED application to the first day of 12-month remission. Nevertheless, this situation demonstrates that there is potential scope for the introduction of systematic error if calculated data rather than raw data are imported without quality assurance. It also strengthens the case for not using multiple old historical cohorts as there is a risk of introducing systematic error if cohorts used different methods for calculation of continuous variables.

Assessment of how data transfer affected seizure type was limited due to a small number of matched variables. It showed that there is a potential source of introduced

variability as two out of three assessed seizure types had a poor concordance between cohorts. Generalised tonic-clonic only and other seizures had poor concordance, whereas partial seizures only showed very good concordance. This was due to a difference in how seizure types were derived. This assessment demonstrates that even applying the same seizure types without directly applying the same code to derive them has the potential for introduction of bias, and should be monitored in large scale projects – particularly if they utilise a multi-stage, multi-centre design.

In the SANAD study, investigation outcomes were coded as normal, abnormal, and not available; however in EpiPGX, abnormal was further separated into abnormal non-specific and specific. Fortunately, the SANAD study collected imaging results separately hence during the data upload it was possible to classify the degree of abnormality. For EEG results, SANAD had a more elaborate classification separating various degrees of focal and generalised abnormalities, and for data transformation they were categorised and lumped together. In the original SANAD genetic dataset, investigation categories were also lumped together creating binary variables of abnormal EEG or CT. Abnormal CT had a kappa value of 0.721 ($p < 0.00$) and abnormal EEG was 0.930 ($p < 0.00$), indicating moderate and almost perfect concordance respectively between datasets for abnormal investigation results (McHugh, 2012).

Basic demographic variables (gender, age, epilepsy type, and abnormal neurological examination) had almost perfect concordance, indicating safe data transfer.

This analysis, in essence, shows that basic and robust demographic variables can be transformed and transferred between research studies. In contrast, more complex variables which have different definitions or classifications are at risk of introducing artificial variability during the process of transformation of data or derivation of new phenotypes.

This assessment also underscores limitations direct comparison of our results with previous research. For example SANAD cases form a significant proportion of our cohort and co-variates were partially modelled according previous research (Bonnett et al.,

2014b, Bonnett et al., 2012). Both impact of data transfer and fact that our cohort included only SANAD cases which contributed DNA (degree of selection bias) would raise a question about potential limitations for direct comparison of results.

If a DNA sample is obtained only from a proportion of the original research cohort there is a potential sub-selection of the population; hence, DNA sampling should be representative of the baseline population. It can be concluded that there is a requirement for a quality and concordance assurance process in large scale studies, particularly utilising multiple historical cohorts and clinical databases.

5.4. Genome based biomarkers for treatment response in newly diagnosed epilepsy

5.4.1. Phenotypes

In pharmacogenomic research, the phenotype requires not only presence of disease, but also exposure to the drug. The reaction of the subject is then observed and a potential genetic base for it is elucidated. Our phenotype definition was centred on the presence of a sustained 12-month period of seizure freedom after application of the first well-tolerated AED. It has been shown by previous research carried out in rural areas of developing countries that around 20-30% of patients with epilepsy experience spontaneous remission (Kwan and Sander, 2004, Placencia et al., 1994). This observation would imply that remission can occur in a significant proportion of patients irrespective of drug application, but as a part of the natural history of the illness. Hence, observed remission might not be a pharmacological effect. It is important to try to separate out patients who are truly responsive to treatment and without the benign natural course of the disease. One possible approach would be selecting patients who respond to the treatment, but have clinical factors associated with poorer prognosis.

Several patterns of drug response in newly diagnosed epilepsy have been described which could potentially be turned into separate phenotypes, like delayed remission and

a remitting-relapsing course (Brodie et al., 2012). Hence, another option would be to define patients with delayed remission as cases on the assumption that if seizures are on-going until the right drug dose has been reached, seizure freedom would represent a true effect of the drug.

Arguably, patients who relapse after an initial remission are not responsive to treatment or simply have more severe epilepsy. It is also possible that they were initially classified as in remission because they had infrequent seizures which were otherwise unaffected by AEDs. Alternatively, they might represent the development of progressive treatment intractability which would be in line with the "drug transporter" hypothesis of drug resistant epilepsy (Sisodiya et al., 2002). This subgroup also could potentially serve as a separate phenotype hence should be analysed separately.

An alternative approach would be to actually define patients who potentially have a benign course of illness and interrogate their genome for biomarkers, as this group might be exposed unnecessarily to AEDs.

In summary, future research is required to develop more nuanced phenotypes to effectively capture the underlying heterogeneity of epilepsy and treatment response. It has already been agreed within the pharmacogenomics field that with the development of genomic technologies and analytical methods, phenotypes have become the most complex part of the equation in the field as a whole (Relling and Evans, 2015).

5.4.2. Role of adjustment for clinical and non-clinical factors in genomic studies

It has already been argued by Speed et al. (2014) that epilepsy pharmacogenetic studies should be carried out as both unadjusted and adjusted for significant clinical factors (Speed et al., 2014a). We followed this trend and first performed a logistic regression of clinical factors associated with outcome in all cases included in the subsequent GWAS. Genetic analysis was then performed with and without adjustment for significant clinical factors. The prospective study design of pharmacogenomics is superior to retrospective design (Jorgensen and Williamson, 2008) and analyses of

association have demonstrated association of non-clinical factors with 12-month remission. Therefore, GWAS was also adjusted for non-clinical factors like mode of case ascertainment.

5.4.3. Sample size

The retrospective sample size calculation showed that the GWAS analysis on both EpiPGX WP2 task 1 and on newly diagnosed focal epilepsy was adequately powered to detect an association, assuming a dominant model of inheritance and relative risk of two; the additive model with similar parameters was incompatible with the statistical model. GWAS were adequately powered to detect SNPs with a high relative genotype risk, but studies were not designed to detect complex polygenetic traits. The GWAS of focal epilepsy was smaller than the original EpiPGX analysis on the newly diagnosed epilepsy cohort. The reason for a smaller sample size was that it was not planned at the beginning of the study, and hence we only included cases where genetic information was stored locally. Plus, it excluded all cases with generalised epilepsy (SANAD arm B). There was little active patient recruitments in the EpiPGX consortium; it relied on cases already collected, hence little possibility to increase the sample size.

Nevertheless, the sample size of the GWAS was bigger than those in previously published work, but Speed et al. (2014) also included generalised epilepsy. Hence, the GWAS on focal epilepsy represents the biggest reported cohort in newly diagnosed focal epilepsy so far.

5.4.4. Results of GWAS

The GWAS analysis for remission of newly diagnosed focal epilepsy failed to detect any significant signal, but several SNPs demonstrated potential suggestive associations. Most of them did not possess any potential biological significance, except rs73053778 which is an intronic variant 500bp downstream from chloride voltage-gated channel 2 (*CLCN-2*). Animal studies have shown that *CLCN-2* is widespread in the CNS and is involved in glial function, playing a role in ionic homeostasis (Blanz et al., 2007). Interestingly the authors of this study also been reported that *CLCN-2* knockout mice

have a normal seizure threshold. In humans, mutations in this gene have been linked to autosomal recessive leukoencephalopathy, but interestingly the phenotype does not include seizures (van der Knaap et al., 2015, Depienne et al., Di Bella et al., 2014). It has also been controversially linked to idiopathic generalised epilepsy, although that association has not been confirmed and the original article was partially retracted (Niemeyer et al., 2010, Saint-Martin et al., 2009, Kleefusz-Lie et al., 2009).

The other gene which had some signal is SEL1L3 which had intronic SNPs in the GWAS univariate and adjusted analysis, but they were not overlapping between analyses. This gene has not been implicated in epilepsy previously, although it has been shown to be up regulated in response to HIV-1 tat protein (Woollard et al., 2014). This association likely represents a spurious finding.

Another important aspect is that our GWAS failed to replicate the results of any previous epilepsy pharmacogenomics study. It can be concluded that there is no single SNP with a strong effect size under the dominant inheritance model determining treatment outcomes in newly diagnosed focal epilepsy.

5.5. Limitations of the study and suggestions regarding future research

One of the limitations of this study was that the research cohort was comprised of a mixture of studies some with a prospective and some with retrospective design as well as some studies were created as genetic studies whereas some as randomized clinical trials (RCT). Significant heterogeneity was observed for treatment outcomes, implicating significant differences in the aims and ways in which the original historical studies were set up. To counteract this problem, cases were re-phenotyped where possible and a statistical analysis performed, were either adjusted or stratified to the origin of the cohort. Re-phenotyping of the cohort itself would not resolve issues related to selection bias as it merely dealt with individual data but had no effect on how the cases were

selected. Selection bias is more pronounced in observational studies (Reeves et al., 2005). In our cohort, RCTs constituted more than half of it. On the other hand, RCTs are prone to selective and narrow inclusion criteria which can raise problems related to the ability to generalise the results to a wider and less homogeneous population (Britton et al., 1999).

Prognostic research is different from regulatory trials and epidemiological research. Prognostic factors can be either causally related to outcome or be only predictive in nature (Kamper et al., 2011). The best design for prognostic research is prospective cohort studies. RCTs can be used but, as they have fixed strict inclusion criteria, the results might not be generalizable (Moons et al., 2009). Furthermore, the usage of RCTs for the investigation of causational predictive factors would be unethical as it would require assigning people to variable degrees of causative factors (Kamper et al., 2011). The usage of a prospective design is preferable to a retrospective one (Riley et al., 2013). Our cohort was set up for the purpose of pharmacogenomics research and later adapted for the analysis of clinical markers for treatment outcome. Given the limited resources available, it would impossible to create a prospective cohort of the same size within the same time span; hence, we had to rely on historical data. Nevertheless, a significant proportion (52.7%) of patients were originally recruited into prospective studies.

A further weakness of our research was the categorisation of the continuous variable of the pre-treatment seizure count. Such an approach has been discouraged in prognostic research; furthermore, it has been shown in newly diagnosed epilepsy that there is a non-linear relationship for this variable (MacDonald et al., 2000, Riley et al., 2013). More than half (59.23%) of cases had categorical values; hence we would lose significant data by excluding them. To capture better the non-linear relationships, categories were created with relatively small steps.

An additional limitation was that we only analysed predictors for binary outcomes that might not fully reflect the real-life clinical situations. If more than one mutually exclusive event can occur, methods using competing risk analysis might be preferred over Kaplan-Meier estimates (Koller et al., 2012). Competing risk analysis has already

been applied in the past on SANAD data (Williamson et al., 2008). Two mutually exclusive events can occur in the case of epilepsy. Patients can either withdraw from treatment due to adverse reactions or a lack of efficacy. An analysis of WP2 Task 1 specifically interrogated the response to the first well tolerated AED; hence, competing risk analysis was not used. Such an approach makes more sense from the pharmacogenetic point of view rather than the clinical, as the genes that influence the probability of achieving remission may differ from those associated with the occurrence of ADR. Furthermore, WP2 Task 3 is planning to use a competing risk methodology for the analysis of treatment failure and the genetics of ADR were covered by a different work package.

As discussed in the results section, the proportionality of hazards over time was not upheld for all covariates included in the Cox model. Previous work carried out by Dr Laura Bonnett on SANAD has shown that, when an accelerated failure time approach is added to the analysis, which does not rely on the assumption of proportionality, the results remain the same (personal communication Bonnett, 2016). It is not fully possible to apply that to this cohort as it did not solely comprise cases from SANAD. The non-proportionality of hazards was also accounted for by stratifying the Cox regression by origin of the cohort. The stratified model that eliminated the CT Head and MRI results lost statistical significance but these were still retained in the model. In the meantime, gender gained statistical significance.

As a part of the assessment of the methodological aspects, the length of remission was assessed. A limitation of this assessment is that it is only relevant to a subset of patients experiencing it, plus some of the patients have been discharged at the onset of the 12 month remission period. A more robust approach would be to analyse the time to a first seizure following the treatment. Unfortunately, these data were not collected, representing a limitation of this analysis. Future research should collect more time to event data.

The effect of different definitions of remission was also assessed by applying different length criteria for remission. This approach has several limitations. There exists

selection bias, as not all patients in a cohort were followed for 5 - 10 years. There is a chance that patients who do not respond to treatment may be over represented when the sample is restricted to cases with long follow up periods. Patients in stable remission would be more likely to be discharged early compared to more complicated patients.

The assessment of the quality of the SANAD data transfer demonstrated that certain co-variates have poor concordance when compared to an earlier genetic study carried out based on the same cohort. Reassuringly critically important co-variates like remission and basic demographic data had very high concordance. These results show some limitations of the approach when data are transferred from the existing database. A sensitivity analysis, which was performed by excluding SANAD cases returned similar results, with the exception that CT and the pre-treatment seizure count lost statistical significance. This has implications for future research, particularly the underlying importance of data quality checks immediately after transfer.

The following practical conclusion can be drawn based on the EpiPGX experience which might be valuable for future epilepsy pharmacogenomics research:

- Preferably, a prospective design should be utilized. Our research showed that the mode of case ascertainment is significantly associated with treatment outcome.

- If a retrospective multi-centre design is used, researchers should account for and control for heterogeneity between centres or historical studies. Clinical practice varies between centres and countries and this includes the co-variates collected. This might affect the development of endophenotype later.

- Continuous data should preferably be collected in an attempt to avoid unnecessary categorization. If data are collected as categorical, they should not mutually exclude the collection of continuous data.

- The electronic transfer of existing clinical databases or historical studies should be planned well in advance and rigorous quality control procedures should be implemented before finalizing the transfer.

- If survival analysis is used time to event data should be collected.

Results of GWAS performed as part of the EpiPGX study and this thesis did not produce any genome wide significant SNPs, furthermore they failed to replicate results of previous studies. This raises a question whether there is a requirement to change the approach. It has been previously suggested that complex diseases phenotypes results from the interaction of many alleles and environment, furthermore lot of identified genetic loci in common disease have only a modest effect (Civelek and Lusis, 2014). Systems genetics has been suggested as a potential approach to explore complex traits and has already been successfully applied to epilepsy (Civelek and Lusis, 2014, Johnson et al., 2015, Delahaye-Duriez et al., 2016). In epilepsy pharmacogenomics pathway analysis also has already been used and have demonstrated potentially utility (Speed et al., 2014a). Future epilepsy pharmacogenomics research should explore role of systems genetics and best methods of its application.

Epilepsy pharmacogenomics future research lies with the large scale prospective multi-national consortia. It is unlikely that any single isolated centre will have the capacity to recruit sufficient patients on its own for adequately powered studies. With the digitalization of medicine, it might become feasible in future to interrogate routine clinical databases and electronic records in the case of common epilepsies. This would require quality control systems to be in place beforehand. Nevertheless, nuanced manual phenotyping likely may be required for phenotyping cases with suspected rare syndromes. Furthermore, the development and integration of electronic records have already paved the way for phenome-wide association studies which represent a reversion of the GWAS approach (Hebbring, 2014), although not yet in the epilepsy field.

The reduction of costs and evolution of genomic technologies like next generation sequencing will provide more tools for pharmacogenomics research and translation into daily clinical practice (Goodwin et al., 2016). This will probably facilitate the further development of pharmacogenomics and prognostic research. With new genomic technologies, the most complex aspect of research will become the development of the phenotype of interest (Relling and Evans, 2015).

6.0. Conclusions

This project explored three important aspects related to newly diagnosed epilepsy research. Routine clinical and genetic factors were assessed for an association with the remission after application of the first well tolerated anticonvulsant. Importantly methodological aspects related to creation large scale multinational epilepsy research project and data quality were also explored.

As discussed in the introduction several studies including few very large already in past have investigated association and predictive role of routine clinical factors in the newly diagnosed epilepsy. Assessment of clinical factors and their relationship with remission from seizures after the application of the first well-tolerated anticonvulsant demonstrated that age at diagnosis, generalised tonic-clonic seizures only, epilepsy type, and the results of the first MRI and EEG were predictive factors and have the same direction of effect in both logistic and Cox regression. Those results in large part replicated previous observations hence confirming current knowledge in the field.

The MRI is becoming a more important tool in modern epileptology, for stratification of results of imaging studies we used a relatively simple classification, a future research should explore it in more details. Particularly as in past it has been demonstrated already that not all abnormalities have the same prognostic implications (Stephen et al., 2001, Mohanraj and Brodie, 2005a).

Better knowledge about clinical factors and their association with treatment outcome might play an important in future epilepsy genetic research. It can be argued that some of clinical factors have a genetic background. Hence their in-depth exploration is not only beneficial for patient care, but in the process of creation of more nuanced phenotypes.

In future epilepsy research collaboration between research groups and countries will take even more prominent role. It was an important part of this research to assess what is the effect for pulling together resources and patient cohorts and how does that affect quality of data.

Our research showed that there is significant heterogeneity between centres on the various parameters assessed. We demonstrated that the origin of cohort, method of case ascertainment, and follow up length are statistically significantly associated with the measured outcome. The change of duration for seizure freedom required for remission resulted in a progressive reduction in the proportion of subjects experiencing it. This provides support for the introduction of quality standards in epilepsy pharmacogenetics research.

For the first time, we assessed the feasibility of using historical data cohorts, clinical databases, and data transfers. We demonstrated that robust demographic parameters are safe for data transfer, but complex variables have the potential to introduce systematic error during the data transfer process. Those results underline importance of internal quality assurance procedures during data transfers as well as during the process of derivation of a new phenotypes based on old historical data. Those aspects should be taken account during process assembly of future cohorts or perhaps giving a consideration of prospective multicentre recruitment.

A GWAS performed on newly diagnosed focal epilepsy for a 12-month remission after application of the first well-tolerated anticonvulsant demonstrated that there is no single SNP with a strong effect size. This finding does not conclude research in this field, but merely indicate requirement for a further large scale prospectively recruited genetic studies utilizing a careful selection of phenotype of interest. It has already been argued that GWAS-type approach might have a limited role in general in discovery of clinically useful predictors in pharmacogenetic field (Nelson et al., 2016). Furthermore in epilepsy pharmacogenetics there are already some arguments for application of pathway based analysis (Speed et al., 2014a).

Future epilepsy pharmacogenomics research should continue developing our knowledge about relationship of clinical factors with disease and treatment results. There is a need for further large scale genetic studies both for epilepsy itself and phamracogenetics with integration of new genomic technologies while maintaining a sound study methodology and quality.

References

ABE, T., SEO, T., ISHITSU, T., NAKAGAWA, T., HORI, M. & NAKAGAWA, K. 2008. Association between SCN1A polymorphism and carbamazepine-resistant epilepsy. *Br J Clin Pharmacol,* 66**,** 304-7.

ABIMBOLA, S., MARTINIUK, A. L. C., HACKETT, M. L. & ANDERSON, C. S. 2011. The Influence of Design and Definition on the Proportion of General Epilepsy Cohorts with Remission and Intractability. *Neuroepidemiology,* 36**,** 204-212.

AKMAN, C. I., MONTENEGRO, M. A., JACOB, S., ECK, K., CHIRIBOGA, C. & GILLIAM, F. 2009. Seizure frequency in children with epilepsy: factors influencing accuracy and parental awareness. *Seizure,* 18**,** 524-9.

ALTMAN, D. G. & RILEY, R. D. 2005. Primer: an evidence-based approach to prognostic markers. *Nat Clin Prac Oncol,* 2**,** 466-472.

AMSTUTZ, U., SHEAR, N. H., RIEDER, M. J., HWANG, S., FUNG, V., NAKAMURA, H., CONNOLLY, M. B., ITO, S., CARLETON, B. C. & THE, C. C. R. G. 2014. Recommendations for HLA-B*15:02 and HLA-A*31:01 genetic testing to reduce the risk of carbamazepine-induced hypersensitivity reactions. *Epilepsia,* 55**,** 496-506.

ANDERSON, G. D. & SHEN, D. D. 2007. Where is the evidence that p-glycoprotein limits brain uptake of antiepileptic drug and contributes to drug resistance in epilepsy? *Epilepsia,* 48**,** 2372-2374.

ANDRADE-VALENÇA, L. P., VALENÇA, M. M., RIBEIRO, L. T., MATOS, A. L. M., SALES, L. V., VELASCO, T. R., SANTOS, A. C. & LEITE, J. P. 2003. Clinical and Neuroimaging Features of Good and Poor Seizure Control Patients with Mesial Temporal Lobe Epilepsy and Hippocampal Atrophy. *Epilepsia,* 44**,** 807-814.

ANNEGERS, J. F., HAUSER, W. A. & ELVEBACK, L. R. 1979. Remission of Seizures and Relapse in Patients with Epilepsy. *Epilepsia,* 20**,** 729-737.

AUDENAERT, D., SCHWARTZ, E., CLAEYS, K. G., CLAES, L., DEPREZ, L., SULS, A., VAN DYCK, T., LAGAE, L., VAN BROECKHOVEN, C., MACDONALD, R. L. & DE JONGHE, P. 2006. A novel GABRG2 mutation associated with febrile seizures. *Neurology,* 67**,** 687-690.

AWASTHI, S., HALLENE, K., FAZIO, V., SINGHAL, S., CUCULLO, L., AWASTHI, Y., DINI, G. & JANIGRO, D. 2005. RLIP76, a non-ABC transporter, and drug resistance in epilepsy. *BMC Neuroscience,* 6**,** 61.

BAFTIU, A., JOHANNESSEN LANDMARK, C., NIKAJ, V., NESLEIN, I.-L., JOHANNESSEN, S. I. & PERUCCA, E. 2015. Availability of antiepileptic drugs across Europe. *Epilepsia,* 56**,** e191-e197.

BALAN, S., SATHYAN, S., RADHA, S. K., JOSEPH, V., RADHAKRISHNAN, K. & BANERJEE, M. 2013. GABRG2, rs211037 is associated with epilepsy susceptibility, but not with antiepileptic drug resistance and febrile seizures. *Pharmacogenet Genomics,* 23**,** 605-10.

BANCAUD, J., HENRIKSEN, O., RUBIO-DONNADIEU, F., SEINO, M., DREIFUSS FRITZ, E. & PENRY, J. K. 1981. Proposal for Revised Clinical and Electroencephalographic Classification of Epileptic Seizures. *Epilepsia,* 22**,** 489-501.

BANERJEE, P. N., FILIPPI, D. & ALLEN HAUSER, W. 2009. The descriptive epidemiology of epilepsy—A review. *Epilepsy Research,* 85**,** 31-45.

BAULAC, S., HUBERFELD, G., GOURFINKEL-AN, I., MITROPOULOU, G., BERANGER, A., PRUD'HOMME, J.-F., BAULAC, M., BRICE, A., BRUZZONE, R. & LEGUERN, E. 2001. First genetic evidence of GABAA receptor dysfunction in epilepsy: a mutation in the [gamma]2-subunit gene. *Nat Genet,* 28**,** 46-48.

BEARDEN, D., STRONG, A., EHNOT, J., DIGIOVINE, M., DLUGOS, D. & GOLDBERG, E. M. 2014. Targeted treatment of migrating partial seizures of infancy with quinidine. *Annals of Neurology,* 76**,** 457-461.

BEGHI, E. & TOGNONI, G. 1988. Prognosis of Epilepsy in Newly Referred Patients: A Multicenter Prospective Study. *Epilepsia,* 29**,** 236-243.
BENN, E. K. T., HAUSER, W. A., SHIH, T., LEARY, L., BAGIELLA, E., DAYAN, P., GREEN, R., ANDREWS, H., THURMAN, D. J. & HESDORFFER, D. C. 2008. Estimating the incidence of first unprovoked seizure and newly diagnosed epilepsy in the low-income urban community of Northern Manhattan, New York City. *Epilepsia,* 49**,** 1431-1439.
BERG, A. T. & KELLY, M. M. 2006. Defining Intractability: Comparisons among Published Definitions. *Epilepsia,* 47**,** 431-436.
BERG, A. T., LANGFITT, J., SHINNAR, S., VICKREY, B. G., SPERLING, M. R., WALCZAK, T., BAZIL, C., PACIA, S. V., SPENCER, S. S. & SURGERY, F. T. M. S. O. E. 2003. How long does it take for partial epilepsy to become intractable? *Neurology,* 60**,** 186-190.
BERG, A. T., LODDENKEMPER, T. & BACA, C. B. 2014. Diagnostic delays in children with early onset epilepsy: Impact, reasons, and opportunities to improve care. *Epilepsia,* 55**,** 123-132.
BERG, A. T., SHINNAR, S., LEVY, S. R., TESTA, F. M., SMITH-RAPAPORT, S., BECKERMAN, B. & EBRAHIMI, N. 2001. Two-Year Remission and Subsequent Relapse in Children with Newly Diagnosed Epilepsy. *Epilepsia,* 42**,** 1553-1562.
BERG, A. T., TESTA, F. M., LEVY, S. R. & SHINNAR, S. 2000. Neuroimaging in Children With Newly Diagnosed Epilepsy: A Community-Based Study. *Pediatrics,* 106**,** 527-532.
BESOCKE, A. G., ROSSO, B., CRISTIANO, E., VALIENSI, S. M., GARCÍA, M. D. C., GONORAZKY, S. E. & ROMANO, L. M. 2013. Outcome of newly-diagnosed epilepsy in older patients. *Epilepsy & Behavior,* 27**,** 29-35.
BILEVICIUS, E., YASUDA, C. L., SILVA, M. S., GUERREIRO, C. A. M., LOPES-CENDES, I. & CENDES, F. 2010. Antiepileptic drug response in temporal lobe epilepsy: A clinical and MRI morphometry study. *Neurology,* 75**,** 1695-1701.
BILL, P. A., VIGONIUS, U., POHLMANN, H., GUERREIRO, C. A. M., KOCHEN, S., SAFFER, D. & MOORE, A. 1997. A double-blind controlled clinical trial of oxcarbazepine versus phenytoin in adults with previously untreated epilepsy1. *Epilepsy Research,* 27**,** 195-204.
BLANZ, J., SCHWEIZER, M., AUBERSON, M., MAIER, H., MUENSCHER, A., HÜBNER, C. A. & JENTSCH, T. J. 2007. Leukoencephalopathy upon Disruption of the Chloride Channel ClC-2. *The Journal of Neuroscience,* 27**,** 6581-6589.
BLOCH, K. M., SILLS, G. J., PIRMOHAMED, M. & ALFIREVIC, A. 2014. Pharmacogenetics of antiepileptic drug-induced hypersensitivity. *Pharmacogenomics,* 15**,** 857-868.
BLUM, D. E., ESKOLA, J., BORTZ, J. J. & FISHER, R. S. 1996. Patient awareness of seizures. *Neurology,* 47**,** 260-4.
BOGARDUS, J. S. T. C. J. F. A. R. 1999. Clinical epidemiological quality in molecular genetic research: The need for methodological standards. *JAMA,* 281**,** 1919-1926.
BONNETT, L., SMITH, C. T., SMITH, D., WILLIAMSON, P., CHADWICK, D. & MARSON, A. G. 2012. Prognostic factors for time to treatment failure and time to 12 months of remission for patients with focal epilepsy: post-hoc, subgroup analyses of data from the SANAD trial. *The Lancet Neurology,* 11**,** 331-340.
BONNETT, L. J. 22/05/2016 2016. *RE: Prognostic Factors in SANAD Accelerated Failure time.* Type to AUCE, P.
BONNETT, L. J., TUDUR SMITH, C., DONEGAN, S. & MARSON, A. G. 2014a. Treatment outcome after failure of a first antiepileptic drug. *Neurology,* 83**,** 552-560.
BONNETT, L. J., TUDUR SMITH, C., SMITH, D., WILLIAMSON, P. R., CHADWICK, D. & MARSON, A. G. 2014b. Time to 12-month remission and treatment failure for generalised and unclassified epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry,* 85**,** 603-610.

BOURNISSEN, F. G., MORETTI, M. E., JUURLINK, D. N., KOREN, G., WALKER, M. & FINKELSTEIN, Y. 2009. Polymorphism of the MDR1/ABCB1 C3435T drug-transporter and resistance to anticonvulsant drugs: a meta-analysis. *Epilepsia,* 50**,** 898-903.

BOUWMEESTER, W., ZUITHOFF, N. P. A., MALLETT, S., GEERLINGS, M. I., VERGOUWE, Y., STEYERBERG, E. W., ALTMAN, D. G. & MOONS, K. G. M. 2012. Reporting and Methods in Clinical Prediction Research: A Systematic Review. *PLOS Medicine,* 9**,** e1001221.

BRITTON, A., MCKEE, M., BLACK, N., MCPHERSON, K., SANDERSON, C. & BAIN, C. 1999. Threats to Applicability of Randomised Trials: Exclusions and Selective Participation. *Journal of Health Services Research & Policy,* 4**,** 112-121.

BRODIE, M. J., BARRY, S. J. E., BAMAGOUS, G. A., NORRIE, J. D. & KWAN, P. 2012. Patterns of treatment response in newly diagnosed epilepsy. *Neurology,* 78**,** 1548-1554.

CALLAGHAN, B., SCHLESINGER, M., RODEMER, W., POLLARD, J., HESDORFFER, D., ALLEN HAUSER, W. & FRENCH, J. 2011. Remission and relapse in a drug-resistant epilepsy population followed prospectively. *Epilepsia,* 52**,** 619-626.

CAMFIELD, C. S. & CAMFIELD, P. R. 2009. Juvenile myoclonic epilepsy 25 years after seizure onset: a population-based study. *Neurology,* 73**,** 1041-5.

CAMFIELD, C. S., STRIANO, P. & CAMFIELD, P. R. 2013. Epidemiology of juvenile myoclonic epilepsy. *Epilepsy & Behavior,* 28**,** S15-S17.

CARPAY, J. A., DE WEERD, A. W., SCHIMSHEIMER, R. J., STROINK, H., BROUWER, O. F., PETERS, A. C. B., VAN DONSELAAR, C. A., GEERTS, A. T. & ARTS, W. F. M. 1997. The Diagnostic Yield of a Second EEG After Partial Sleep Deprivation: A Prospective Study in Children with Newly Diagnosed Seizures. *Epilepsia,* 38**,** 595-599.

CHEN, P., LIN, J.-J., LU, C.-S., ONG, C.-T., HSIEH, P. F., YANG, C.-C., TAI, C.-T., WU, S.-L., LU, C.-H., HSU, Y.-C., YU, H.-Y., RO, L.-S., LU, C.-T., CHU, C.-C., TSAI, J.-J., SU, Y.-H., LAN, S.-H., SUNG, S.-F., LIN, S.-Y., CHUANG, H.-P., HUANG, L.-C., CHEN, Y.-J., TSAI, P.-J., LIAO, H.-T., LIN, Y.-H., CHEN, C.-H., CHUNG, W.-H., HUNG, S.-I., WU, J.-Y., CHANG, C.-F., CHEN, L., CHEN, Y.-T. & SHEN, C.-Y. 2011. Carbamazepine-Induced Toxic Effects and HLA-B*1502 Screening in Taiwan. *New England Journal of Medicine,* 364**,** 1126-1133.

CHEN, P., YAN, Q., XU, H., LU, A. & ZHAO, P. 2014a. The Effects of ABCC2 G1249A Polymorphism on the Risk of Resistance to Antiepileptic Drugs: A Meta-Analysis of the Literature. *Genetic Testing and Molecular Biomarkers,* 18**,** 106-111.

CHEN, Z., LIEW, D. & KWAN, P. 2014b. Effects of a HLA-B*15:02 screening policy on antiepileptic drug use and severe skin reactions. *Neurology,* 83**,** 2077-2084.

CHEUNG, Y.-K., CHENG, S.-H., CHAN, E. J. M., LO, S. V., NG, M. H. L. & KWAN, P. 2013. HLA-B alleles associated with severe cutaneous reactions to antiepileptic drugs in Han Chinese. *Epilepsia,* 54**,** 1307-1314.

CHRISTE, W., KRÄMER, G., VIGONIUS, U., POHLMANN, H., STEINHOFF, B. J., BRODIE, M. J. & MOORE, A. 1997. A double-blind controlled clinical trial: oxcarbazepine versus sodium valproate in adults with newly diagnosed epilepsy1. *Epilepsy Research,* 26**,** 451-460.

CHRISTENSEN, J., KJELDSEN, M. J., ANDERSEN, H., FRIIS, M. L. & SIDENIUS, P. 2005. Gender Differences in Epilepsy. *Epilepsia,* 46**,** 956-960.

CHRISTENSEN, J., VESTERGAARD, M., PEDERSEN, M. G., PEDERSEN, C. B., OLSEN, J. & SIDENIUS, P. 2007. Incidence and prevalence of epilepsy in Denmark. *Epilepsy Research,* 76**,** 60-65.

CHUNGATH, M. & SHORVON, S. 2008. The mortality and morbidity of febrile seizures. *Nat Clin Pract Neuro,* 4**,** 610-621.

CIVELEK, M. & LUSIS, A. J. 2014. Systems genetics approaches to understand complex traits. *Nature reviews. Genetics,* 15**,** 34-48.

COBOS, A., SANCHEZ, P., AGUADO, J. & CARRASCO, J. L. 2011. Methodological quality in pharmacogenetic studies with binary assessment of treatment response: a review. *Pharmacogenet Genomics,* 21**,** 243-50.

COCKERELL, O. C., JOHNSON, A. L., SANDER, J. W. A. S. & SHORVON, S. D. 1997. Prognosis of Epilepsy: A Review and Further Analysis of the First Nine Years of the British National General Practice Study of Epilepsy, a Prospective Population-Based Study. *Epilepsia,* 38**,** 31-46.

COCKERELL, O. C., SANDER, J. W. A. S., HART, Y. M., SHORVON, S. D. & JOHNSON, A. L. 1995. Remission of epilepsy: results from the National General Practice Study of Epilepsy. *The Lancet,* 346**,** 140-144.

COLLABORATIVE GROUP FOR THE STUDY OF, E. 1992. Prognosis of Epilepsy in Newly Referred Patients: A Multicenter Prospective Study of the Effects of Monotherapy on the Long-Term Course of Epilepsy. *Epilepsia,* 33**,** 45-51.

CORDON-CARDO, C., O'BRIEN, J. P., CASALS, D., RITTMAN-GRAUER, L., BIEDLER, J. L., MELAMED, M. R. & BERTINO, J. R. 1989. Multidrug-resistance gene (P-glycoprotein) is expressed by endothelial cells at blood-brain barrier sites. *Proceedings of the National Academy of Sciences,* 86**,** 695-698.

COSSETTE, P., LIU, L., BRISEBOIS, K., DONG, H., LORTIE, A., VANASSE, M., SAINT-HILAIRE, J.-M., CARMANT, L., VERNER, A., LU, W.-Y., TIAN WANG, Y. & ROULEAU, G. A. 2002. Mutation of GABRA1 in an autosomal dominant form of juvenile myoclonic epilepsy. *Nat Genet,* 31**,** 184-189.

DEL FELICE, A., BEGHI, E., BOERO, G., LA NEVE, A., BOGLIUN, G., DE PALO, A. & SPECCHIO, L. M. 2010. Early versus late remission in a cohort of patients with newly diagnosed epilepsy. *Epilepsia,* 51**,** 37-42.

DELAHAYE-DURIEZ, A., SRIVASTAVA, P., SHKURA, K., LANGLEY, S. R., LAANISTE, L., MORENO-MORAL, A., DANIS, B., MAZZUFERI, M., FOERCH, P., GAZINA, E. V., RICHARDS, K., PETROU, S., KAMINSKI, R. M., PETRETTO, E. & JOHNSON, M. R. 2016. Rare and common epilepsies converge on a shared gene regulatory network providing opportunities for novel antiepileptic drug discovery. *Genome Biology,* 17**,** 245.

DELANEAU, O., HOWIE, B., COX, ANTHONY J., ZAGURY, J.-F. & MARCHINI, J. 2013a. Haplotype Estimation Using Sequencing Reads. *The American Journal of Human Genetics,* 93**,** 687-696.

DELANEAU, O., MARCHINI, J. & THE GENOMES PROJECT, C. 2014. Integrating sequence and array data to create an improved 1000 Genomes Project haplotype reference panel. *Nat Commun,* 5.

DELANEAU, O., MARCHINI, J. & ZAGURY, J.-F. 2012. A linear complexity phasing method for thousands of genomes. *Nat Meth,* 9**,** 179-181.

DELANEAU, O., ZAGURY, J.-F. & MARCHINI, J. 2013b. Improved whole-chromosome phasing for disease and population genetic studies. *Nat Meth,* 10**,** 5-6.

DEPIENNE, C., BUGIANI, M., DUPUITS, C., GALANAUD, D., TOUITOU, V., POSTMA, N., VAN BERKEL, C., POLDER, E., TOLLARD, E., DARIOS, F., BRICE, A., DE DIE-SMULDERS, C. E., VLES, J. S., VANDERVER, A., UZIEL, G., YALCINKAYA, C., FRINTS, S. G., KALSCHEUER, V. M., KLOOSTER, J., KAMERMANS, M., ABBINK, T. E. M., WOLF, N. I., SEDEL, F. & VAN DER KNAAP, M. S. 2013. Brain white matter oedema due to ClC-2 chloride channel deficiency: an observational analytical study. *The Lancet Neurology,* 12**,** 659-668.

DI BELLA, D., PAREYSON, D., SAVOIARDO, M., FARINA, L., CIANO, C., CALDARAZZO, S., SAGNELLI, A., BONATO, S., NAVA, S., BRESOLIN, N., TEDESCHI, G., TARONI, F. & SALSANO, E. 2014. Subclinical leukodystrophy and infertility in a man with a novel homozygous CLCN2 mutation. *Neurology,* 83**,** 1217-1218.

DI MASCIO, R., BEGHI, E., SASANELLI, F. & TOGNONI, G. 1986. Early prognosis of epilepsy. Effects of treatment in the first follow-up year. *The Italian Journal of Neurological Sciences,* 7**,** 421-429.

DICKENS, D., YUSOF, S. R., ABBOTT, N. J., WEKSLER, B., ROMERO, I. A., COURAUD, P.-O., ALFIREVIC, A., PIRMOHAMED, M. & OWEN, A. 2013. A Multi-System Approach Assessing the Interaction of Anticonvulsants with P-gp. *PLoS ONE,* 8**,** e64854.

DOMBROWSKI, S. M., DESAI, S. Y., MARRONI, M., CUCULLO, L., GOODRICH, K., BINGAMAN, W., MAYBERG, M. R., BENGEZ, L. & JANIGRO, D. 2001. Overexpression of multiple drug resistance genes in endothelial cells from patients with refractory epilepsy. *Epilepsia,* 42**,** 1501-6.

DONG, L., LUO, R., TONG, Y., CAI, X., MAO, M. & YU, D. 2011. Lack of association between ABCB1 gene polymorphisms and pharmacoresistant epilepsy: an analysis in a western Chinese pediatric population. *Brain Res,* 1391**,** 114-24.

DORADO, P., LOPEZ-TORRES, E., PENAS-LLEDO, E. M., MARTINEZ-ANTON, J. & LLERENA, A. 2013. Neurological toxicity after phenytoin infusion in a pediatric patient with epilepsy: influence of CYP2C9, CYP2C19 and ABCB1 genetic polymorphisms. *Pharmacogenomics J,* 13**,** 359-361.

DRAVET, C. & OGUNI, H. 2013. Chapter 65 - Dravet syndrome (severe myoclonic epilepsy in infancy). *In:* OLIVIER DULAC, M. L. & HARVEY, B. S. (eds.) *Handbook of Clinical Neurology.* Elsevier.

ELWES, R. D. C., JOHNSON, A. L., SHORVON, S. D. & REYNOLDS, E. H. 1984. The Prognosis for Seizure Control in Newly Diagnosed Epilepsy. *New England Journal of Medicine,* 311**,** 944-947.

EPILEPSY, C. O. C. A. T. O. T. I. L. A. 1989. Proposal for Revised Classification of Epilepsies and Epileptic Syndromes. *Epilepsia,* 30**,** 389-399.

EPIPGX 2013. *eCRF Data Entry Manual Version 0.9*.

EPIPGX 2014. EpiPGX Phenotype Definitions.

FISHER, R. S., ACEVEDO, C., ARZIMANOGLOU, A., BOGACZ, A., CROSS, J. H., ELGER, C. E., ENGEL, J., FORSGREN, L., FRENCH, J. A., GLYNN, M., HESDORFFER, D. C., LEE, B. I., MATHERN, G. W., MOSHÉ, S. L., PERUCCA, E., SCHEFFER, I. E., TOMSON, T., WATANABE, M. & WIEBE, S. 2014. ILAE Official Report: A practical clinical definition of epilepsy. *Epilepsia,* 55**,** 475-482.

FISHER, R. S., VICKREY, B. G., GIBSON, P., HERMANN, B., PENOVICH, P., SCHERER, A. & WALKER, S. 2000. The impact of epilepsy from the patient's perspective II: views about therapy and health care. *Epilepsy Research,* 41**,** 53-62.

FORSGREN, L., BEGHI, E., ÕUN, A. & SILLANPÄÄ, M. 2005. The epidemiology of epilepsy in Europe – a systematic review. *European Journal of Neurology,* 12**,** 245-253.

GASPARINI, S., FERLAZZO, E., BEGHI, E., TRIPEPI, G., LABATE, A., MUMOLI, L., LEONARDI, C. G., CIANCI, V., LATELLA, M. A., GAMBARDELLA, A. & AGUGLIA, U. 2013. Family history and frontal lobe seizures predict long-term remission in newly diagnosed cryptogenic focal epilepsy. *Epilepsy Research,* 107**,** 101-108.

GEERTS, A., ARTS, W. F., STROINK, H., PEETERS, E., BROUWER, O., PETERS, B., LAAN, L. & VAN DONSELAAR, C. 2010. Course and outcome of childhood epilepsy: A 15-year follow-up of the Dutch Study of Epilepsy in Childhood. *Epilepsia,* 51**,** 1189-1197.

GEITHNER, J., SCHNEIDER, F., WANG, Z., BERNEISER, J., HERZER, R., KESSLER, C. & RUNGE, U. 2012. Predictors for long-term seizure outcome in juvenile myoclonic epilepsy: 25-63 years of follow-up. *Epilepsia,* 53**,** 1379-86.

GELISSE, P., GENTON, P., THOMAS, P., REY, M., SAMUELIAN, J. C. & DRAVET, C. 2001. Clinical factors of drug resistance in juvenile myoclonic epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry,* 70**,** 240-243.

GOODWIN, S., MCPHERSON, J. D. & MCCOMBIE, W. R. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nat Rev Genet,* 17**,** 333-351.
GOWERS, W. R. 1885. *Epilepsy and other chronic convulsive diseases: their causes, symptoms & treatment,* N.Y., William Wood & Co.
GROUP, M. R. C. A. D. W. S. 1993. Prognostic index for recurrence of seizures after remission of epilepsy. Medical Research Council Antiepileptic Drug Withdrawal Study Group. *British Medical Journal,* 306**,** 1374-1378.
GROVER, S., BALA, K., SHARMA, S., GOURIE-DEVI, M., BAGHEL, R., KAUR, H., GUPTA, M., TALWAR, P. & KUKRETI, R. 2010. Absence of a general association between ABCB1 genetic variants and response to antiepileptic drugs in epilepsy patients. *Biochimie,* 92**,** 1207-12.
GROVER, S., GOURIE-DEVI, M., BALA, K., SHARMA, S. & KUKRETI, R. 2012. Genetic association analysis of transporters identifies ABCC2 loci for seizure control in women with epilepsy on first-line antiepileptic drugs. *Pharmacogenet Genomics,* 22**,** 447-65.
GROVER, S. & KUKRETI, R. 2013. A systematic review and meta-analysis of the role of ABCC2 variants on drug response in patients with epilepsy. *Epilepsia,* 54**,** 936-945.
GUERRINI, R., DRAVET, C., GENTON, P., BELMONTE, A., KAMINSKA, A. & DULAC†, O. 1998. Lamotrigine and Seizure Aggravation in Severe Myoclonic Epilepsy. *Epilepsia,* 39**,** 508-512.
GULCEBI, M. I., OZKAYNAKCı, A., GOREN, M. Z., AKER, R. G., OZKARA, C. & ONAT, F. Y. 2011. The relationship between UGT1A4 polymorphism and serum concentration of lamotrigine in patients with epilepsy. *Epilepsy Research,* 95**,** 1-8.
HAERIAN, B. S., BAUM, L., KWAN, P., TAN, H. J., RAYMOND, A. A. & MOHAMED, Z. 2013. SCN1A, SCN2A and SCN3A gene polymorphisms and responsiveness to antiepileptic drugs: a multicenter cohort study and meta-analysis. *Pharmacogenomics,* 14**,** 1153-1166.
HAERIAN, B. S., BAUM, L., TAN, H. J., KWAN, P., RAYMOND, A. A., SARUWATARI, J., NAKAGAWA, K. & MOHAMED, Z. 2012. SCN1A IVS5N+5 polymorphism and response to sodium valproate: a multicenter study. *Pharmacogenomics,* 13**,** 1477-1485.
HAERIAN, B. S., LIM, K. S., MOHAMED, E. H., TAN, H. J., TAN, C. T., RAYMOND, A. A., WONG, C. P., WONG, S. W. & MOHAMED, Z. 2011a. Lack of association of ABCB1 and PXR polymorphisms with response to treatment in epilepsy. *Seizure,* 20**,** 387-94.
HAERIAN, B. S., LIM, K. S., MOHAMED, E. H., TAN, H. J., TAN, C. T., RAYMOND, A. A., WONG, C. P., WONG, S. W. & MOHAMED, Z. 2011b. Lack of association of ABCB1 haplotypes on five loci with response to treatment in epilepsy. *Seizure,* 20**,** 546-53.
HAERIAN, B. S., LIM, K. S., TAN, C. T., RAYMOND, A. A. & MOHAMED, Z. 2011c. Association of ABCB1 gene polymorphisms and their haplotypes with response to antiepileptic drugs: a systematic review and meta-analysis. *Pharmacogenomics* 12**,** 713-725.
HAERIAN, B. S., ROSLAN, H., RAYMOND, A. A., TAN, C. T., LIM, K. S., ZULKIFLI, S. Z., MOHAMED, E. H., TAN, H. J. & MOHAMED, Z. 2010. ABCB1 C3435T polymorphism and the risk of resistance to antiepileptic drugs in epilepsy: a systematic review and meta-analysis. *Seizure,* 19**,** 339-46.
HAKAMI, T., MCINTOSH, A., TODARO, M., LUI, E., YERRA, R., TAN, K. M., FRENCH, C., LI, S., DESMOND, P., MATKOVIC, Z. & O'BRIEN, T. J. 2013. MRI-identified pathology in adults with new-onset seizures. *Neurology,* 81**,** 920-927.
HARKIN, L. A., BOWSER, D. N., DIBBENS, L. M., SINGH, R., PHILLIPS, F., WALLACE, R. H., RICHARDS, M. C., WILLIAMS, D. A., MULLEY, J. C., BERKOVIC, S. F., SCHEFFER, I. E. & PETROU, S. 2002. Truncation of the GABAA-Receptor γ2 Subunit in a Family with Generalized Epilepsy with Febrile Seizures Plus. *The American Journal of Human Genetics,* 70**,** 530-536.
HARRELL, F. E., LEE, K. L. & MARK, D. B. 1996. MULTIVARIABLE PROGNOSTIC MODELS: ISSUES IN DEVELOPING MODELS, EVALUATING ASSUMPTIONS AND ADEQUACY, AND MEASURING AND REDUCING ERRORS. *Statistics in Medicine,* 15**,** 361-387.

HASHIMOTO, Y., OTSUKI, Y., ODANI, A., TAKANO, M., HATTORI, H., FURUSHO, K. & INUI, K.-I. 1996. Effect of CYP2C polymorphysms on the pharmacokinetics of phentoin in Japanese patients with epilepsy *Biological & Pharmaceutical Bulletin,* 19**,** 1103-1105.

HAUSER, W. A., ANNEGERS, J. F. & KURLAND, L. T. 1993. Incidence of epilepsy and unprovoked seizures in Rochester, Minnesota: 1935-1984. *Epilepsia,* 34**,** 453-468.

HEBBRING, S. J. 2014. The challenges, advantages and future of phenome-wide association studies. *Immunology,* 141**,** 157-165.

HELLER, A. J., CHESTERMAN, P., ELWES, R. D., CRAWFORD, P., CHADWICK, D., JOHNSON, A. L. & REYNOLDS, E. H. 1995. Phenobarbitone, phenytoin, carbamazepine, or sodium valproate for newly diagnosed adult epilepsy: a randomised comparative monotherapy trial. *Journal of Neurology, Neurosurgery & Psychiatry,* 58**,** 44-50.

HEMINGWAY, H., RILEY, R. D. & ALTMAN, D. G. 2009. Ten steps towards improving prognosis research. *BMJ,* 339.

HINGORANI, A. D., WINDT, D. A. V. D., RILEY, R. D., ABRAMS, K., MOONS, K. G. M., STEYERBERG, E. W., SCHROTER, S., SAUERBREI, W., ALTMAN, D. G. & HEMINGWAY, H. 2013. Prognosis research strategy (PROGRESS) 4: Stratified medicine research. *BMJ : British Medical Journal,* 346.

HITIRIS, N., MOHANRAJ, R., NORRIE, J., SILLS, G. J. & BRODIE, M. J. 2007. Predictors of pharmacoresistant epilepsy. *Epilepsy Research,* 75**,** 192-196.

HOFFMEYER, S., BURK, O., VON RICHTER, O., ARNOLD, H. P., BROCKMĆ¶LLER, J., JOHNE, A., CASCORBI, I., GERLOFF, T., ROOTS, I., EICHELBAUM, M. & BRINKMANN, U. 2000. Functional polymorphisms of the human multidrug-resistance gene: Multiple sequence variations and correlation of one allele with P-glycoprotein expression and activity in vivo. *Proceedings of the National Academy of Sciences,* 97**,** 3473-3478.

HOPPE, C., POEPEL, A. & ELGER, C. E. 2007. Epilepsy: Accuracy of patient seizure counts. *Archives of Neurology,* 64**,** 1595-1599.

HORN, C. S., ATER, S. B. & HURST, D. L. 1986. Carbamazepine-exacerbated epilepsy in children and adolescents. *Pediatric Neurology,* 2**,** 340-345.

HOSMER, D. W., LEMESHOW, S. & STURDIVANT, R. X. 2013. *Wiley Series in Probability and Statistics : Applied Logistic Regression (3),* New York, US, Wiley.

HOWIE, B., FUCHSBERGER, C., STEPHENS, M., MARCHINI, J. & ABECASIS, G. R. 2012. Fast and accurate genotype imputation in genome-wide association studies through pre-phasing. *Nat Genet,* 44**,** 955-9.

HOWIE, B. N., DONNELLY, P. & MARCHINI, J. 2009. A Flexible and Accurate Genotype Imputation Method for the Next Generation of Genome-Wide Association Studies. *PLoS Genet,* 5**,** e1000529.

HUI, A. C. F., WONG, A., WONG, H. C., MAN, B. L., AU-YEUNG, K. M. & WONG, K. S. 2007. Refractory epilepsy in a Chinese population. *Clinical Neurology and Neurosurgery,* 109**,** 672-675.

HUNG, C.-C., CHANG, W.-L., HO, J.-L., TAI, J. J., HSIEH, T.-J., HUANG, H.-C., HSIEH, Y.-W. & LIOU, H.-H. 2011. Association of polymorphisms in EPHX1, UGT2B7, ABCB1, ABCC2, SCN1A and SCN2A genes with carbamazepine therapy optimization. *Pharmacogenomics,* 13**,** 159-169.

HUNG, C.-C., CHANG, W.-L., HO, J.-L., TAI, J. J., HSIEH, T.-J., HUANG, H.-C., HSIEH, Y.-W. & LIOU, H.-H. 2012. Association of polymorphisms in EPHX1, UGT2B7, ABCB1, ABCC2, SCN1A and SCN2A genes with carbamazepine therapy optimization. *Pharmacogenomics,* 13**,** 159-169.

HUNG, C.-C., CHEN, P.-L., HUANG, W.-M., TAI, J. J., HSIEH, T.-J., DING, S.-T., HSIEH, Y.-W. & LIOU, H.-H. 2013. Gene-wide tagging study of the effects of common genetic polymorphisms in the α subunits of the GABAA receptor on epilepsy treatment response. *Pharmacogenomics,* 14**,** 1849-1856.

HUNG, C.-C., JEN TAI, J., KAO, P.-J., LIN, M.-S. & LIOU, H.-H. 2007. Association of polymorphisms in NR1I2 and ABCB1 genes with epilepsy treatment responses. *Pharmacogenomics,* 8**,** 1151-1158.

INOMATA, S., NAGASHIMA, A., ITAGAKI, F., HOMMA, M., NISHIMURA, M., OSAKA, Y., OKUYAMA, K., TANAKA, E., NAKAMURA, T., KOHDA, Y., NAITO, S., MIYABE, M. & TOYOOKA, H. 2005. CYP2C19 genotype affects diazepam pharmacokinetics and emergence from general anesthesia. *Clin Pharmacol Ther,* 78**,** 647-655.

JALLON, P., LOISEAU, P., LOISEAU, J. & ON BEHALF OF GROUPE, C. 2001. Newly Diagnosed Unprovoked Epileptic Seizures: Presentation at Diagnosis in CAROLE Study. *Epilepsia,* 42**,** 464-475.

JAMESON, J. L. & LONGO, D. L. 2015. Precision Medicine — Personalized, Problematic, and Promising. *New England Journal of Medicine,* 0**,** null.

JOHNSON, M. R., BEHMOARAS, J., BOTTOLO, L., KRISHNAN, M. L., PERNHORST, K., SANTOSCOY, P. L. M., ROSSETTI, T., SPEED, D., SRIVASTAVA, P. K., CHADEAU-HYAM, M., HAJJI, N., DABROWSKA, A., ROTIVAL, M., RAZZAGHI, B., KOVAC, S., WANISCH, K., GRILLO, F. W., SLAVIERO, A., LANGLEY, S. R., SHKURA, K., RONCON, P., DE, T., MATTHEISEN, M., NIEHUSMANN, P., O'BRIEN, T. J., PETROVSKI, S., VON LEHE, M., HOFFMANN, P., ERIKSSON, J., COFFEY, A. J., CICHON, S., WALKER, M., SIMONATO, M., DANIS, B., MAZZUFERI, M., FOERCH, P., SCHOCH, S., DE PAOLA, V., KAMINSKI, R. M., CUNLIFFE, V. T., BECKER, A. J. & PETRETTO, E. 2015. Systems genetics identifies Sestrin 3 as a regulator of a proconvulsant gene network in human epileptic hippocampus. *Nature communications,* 6**,** 6031-6031.

JOHNSTON, A. J., KANG, J.-Q., SHEN, W., PICKRELL, W. O., CUSHION, T. D., DAVIES, J. S., BAER, K., MULLINS, J. G. L., HAMMOND, C. L., CHUNG, S.-K., THOMAS, R. H., WHITE, C., SMITH, P. E. M., MACDONALD, R. L. & REES, M. I. 2014. A Novel GABRG2 mutation, p.R136*, in a family with GEFS + and extended phenotypes. *Neurobiology of Disease,* 64**,** 131-141.

JORGENSEN, A. L. & WILLIAMSON, P. R. 2008. Methodological quality of pharmacogenetic studies: Issues of concern. *Statistics in Medicine,* 27**,** 6547-6569.

JULIANO, R. L. & LING, V. 1976. A surface glycoprotein modulating drug permeability in Chinese hamster ovary cell mutants. *Biochimica et Biophysica Acta (BBA) - Biomembranes,* 455**,** 152-162.

KAMPER, S. J., HANCOCK, M. J. & MAHER, C. G. 2011. Optimal Designs for Prediction Studies of Whiplash. *Spine,* 36**,** S268-S274.

KASPERAVICIUTE, D. & SISODIYA, S. M. 2009. Epilepsy pharmacogenetics. *Pharmacogenomics,* 10**,** 817-36.

KERB, R., HOFFMEYER, S. & BRINKMANN, U. 2001. ABC drug transporters: hereditary polymorphisms and pharmacological impact in MDR1, MRP1 and MRP2. *Pharmacogenomics,* 2**,** 51-64.

KERLING, F., MUELLER, S., PAULI, E. & STEFAN, H. 2006. When do patients forget their seizures? An electroclinical study. *Epilepsy Behav,* 9**,** 281-5.

KIM, D. W., LEE, S. K., CHU, K., JANG, I.-J., YU, K.-S., CHO, J.-Y. & KIM, S.-J. 2009a. Lack of association between ABCB1, ABCG2, and ABCC2 genetic polymorphisms and multidrug resistance in partial epilepsy. *Epilepsy Research,* 84**,** 86-90.

KIM, D. W., LEE, S. K., CHU, K., JANG, I. J., YU, K. S., CHO, J. Y. & KIM, S. J. 2009b. Lack of association between ABCB1, ABCG2, and ABCC2 genetic polymorphisms and multidrug resistance in partial epilepsy. *Epilepsy Res,* 84**,** 86-90.

KIM, L. G., JOHNSON, T. L., MARSON, A. G. & CHADWICK, D. W. 2006a. Prediction of risk of seizure recurrence after a single seizure and early epilepsy: further results from the MESS trial. *The Lancet Neurology,* 5**,** 317-322.

KIM, W.-J., PARK, S.-C., LEE, S.-J., LEE, J.-H., KIM, J.-Y., LEE, B.-I. & KIM, D.-I. 1999. The Prognosis for Control of Seizures with Medications in Patients with MRI Evidence for Mesial Temporal Sclerosis. *Epilepsia,* 40**,** 290-293.

KIM, Y. O., KIM, M. K., WOO, Y. J., LEE, M. C., KIM, J. H., PARK, K. W., KIM, E. Y., ROH, Y. I. & KIM, C. J. 2006b. Single nucleotide polymorphisms in the multidrug resistance 1 gene in Korean epileptics. *Seizure,* 15**,** 67-72.

KING, M. A., NEWTON, M. R., JACKSON, G. D., FITT, G. J., MITCHELL, L. A., SILVAPULLE, M. J. & BERKOVIC, S. F. 1998. Epileptology of the first-seizure presentation: a clinical, electroencephalographic, and magnetic resonance imaging study of 300 consecutive patients. *The Lancet,* 352**,** 1007-1011.

KLEEFUSZ-LIE, A., FRIEDL, W., CICHON, S., HAUG, K., WARNSTEDT, M., ALEKOV, A., SANDER, T., RAMIREZ, A., POSER, B., MALJEVIC, S., HEBEISEN, S., KUBISCH, C., REBSTOCK, J., HORVATH, S., HALLMANN, K., DULLINGER, J. S., RAU, B., HAVERKAMP, F., BEYENBURG, S., SCHULZ, H., JANZ, D., GIESE, B., MULLER-NEWEN, G., PROPPING, P., ELGER, C. E., FAHLKE, C. & LERCHE, H. 2009. CLCN2 variants in idiopathic generalized epilepsy. *Nat Genet,* 41**,** 954-955.

KLEVELAND, G. & ENGELSEN, B. A. 1998. Juvenile myoclonic epilepsy: clinical characteristics, treatment and prognosis in a Norwegian population of patients. *Seizure - European Journal of Epilepsy,* 7**,** 31-38.

KLOTZ, U. 2007. The Role of Pharmacogenetics in the Metabolism of Antiepileptic Drugs: Pharmacokinetic and Therapeutic Implications. *Clinical Pharmacokinetics,* 46**,** 271-279.

KOBAYASHI, E., LOPES–CENDES, I., GUERREIRO, C. A. M., SOUSA, S. C., GUERREIRO, M. M. & CENDES, F. 2001. Seizure outcome and hippocampal atrophy in familial mesial temporal lobe epilepsy. *Neurology,* 56**,** 166-172.

KOLLER, M. T., RAATZ, H., STEYERBERG, E. W. & WOLBERS, M. 2012. Competing risks and the clinical community: irrelevance or ignorance? *Statistics in Medicine,* 31**,** 1089-1097.

KOTSOPOULOS, IRENE A. W., VAN MERODE, T., KESSELS, FONS G. H., DE KROM, MARC C. T. F. M. & KNOTTNERUS, J. A. 2002. Systematic Review and Meta-analysis of Incidence Studies of Epilepsy and Unprovoked Seizures. *Epilepsia,* 43**,** 1402-1409.

KRUMHOLZ, A., WIEBE, S., GRONSETH, G. S., GLOSS, D. S., SANCHEZ, A. M., KABIR, A. A., LIFERIDGE, A. T., MARTELLO, J. P., KANNER, A. M., SHINNAR, S., HOPP, J. L. & FRENCH, J. A. 2015. Evidence-based guideline: Management of an unprovoked first seizure in adults: Report of the Guideline Development Subcommittee of the American Academy of Neurology and the American Epilepsy Society. *Neurology,* 84**,** 1705-1713.

KUMARI, R., LAKHAN, R., GARG, R. K., KALITA, J., MISRA, U. K. & MITTAL, B. 2011a. Pharmacogenomic association study on the role of drug metabolizing, drug transporters and drug target gene polymorphisms in drug-resistant epilepsy in a north Indian population. *Indian J Hum Genet,* 17 Suppl 1**,** S32-40.

KUMARI, R., LAKHAN, R., KALITA, J., GARG, R. K., MISRA, U. K. & MITTAL, B. 2011b. Potential role of GABAA receptor subunit; GABRA6, GABRB2 and GABRR2 gene polymorphisms in epilepsy susceptibility and pharmacotherapy in North Indian population. *Clin Chim Acta,* 412**,** 1244-8.

KUMARI, R., LAKHAN, R., KALITA, J., MISRA, U. K. & MITTAL, B. 2010. Association of alpha subunit of GABAA receptor subtype gene polymorphisms with epilepsy susceptibility and drug resistance in north Indian population. *Seizure,* 19**,** 237-41.

KUMARI, R., LAKHAN, R., KUMAR, S., GARG, R. K., MISRA, U. K., KALITA, J. & MITTAL, B. 2013. SCN1AIVS5-91G>A polymorphism is associated with susceptibility to epilepsy but not with drug responsiveness. *Biochimie,* 95**,** 1350-1353.

KUMLIEN, E., DOSS, R. C. & GATES, J. R. 2002. Treatment outcome in patients with mesial temporal sclerosis. *Seizure,* 11**,** 413-417.

KWAN, P., ARZIMANOGLOU, A., BERG, A. T., BRODIE, M. J., ALLEN HAUSER, W., MATHERN, G., MOSHÉ, S. L., PERUCCA, E., WIEBE, S. & FRENCH, J. 2010. Definition of drug resistant epilepsy: Consensus proposal by the ad hoc Task Force of the ILAE Commission on Therapeutic Strategies. *Epilepsia,* 51**,** 1069-1077.

KWAN, P., BAUM, L., WONG, V., NG, P. W., LUI, C. H., SIN, N. C., HUI, A. C., YU, E. & WONG, L. K. 2007. Association between ABCB1 C3435T polymorphism and drug-resistant epilepsy in Han Chinese. *Epilepsy Behav,* 11**,** 112-7.

KWAN, P. & BRODIE, M. J. 2000. Early Identification of Refractory Epilepsy. *New England Journal of Medicine,* 342**,** 314-319.

KWAN, P. & BRODIE, M. J. 2005. Potential role of drug transporters in the pathogenesis of medically intractable epilepsy. *Epilepsia,* 46**,** 224-35.

KWAN, P., BRODIE, M. J., KÄLVIÄINEN, R., YURKEWICZ, L., WEAVER, J. & KNAPP, L. E. 2011. Efficacy and safety of pregabalin versus lamotrigine in patients with newly diagnosed partial seizures: a phase 3, double-blind, randomised, parallel-group trial. *The Lancet Neurology,* 10**,** 881-890.

KWAN, P., POON, W. S., NG, H. K., KANG, D. E., WONG, V., NG, P. W., LUI, C. H., SIN, N. C., WONG, K. S. & BAUM, L. 2008. Multidrug resistance in epilepsy and polymorphisms in the voltage-gated sodium channel genes SCN1A, SCN2A, and SCN3A: correlation among phenotype, genotype, and mRNA expression. *Pharmacogenet Genomics,* 18**,** 989-98.

KWAN, P. & SANDER, J. W. 2004. The natural history of epilepsy: an epidemiological view. *Journal of Neurology, Neurosurgery & Psychiatry,* 75**,** 1376-1381.

KWAN, P., SILLS, G. J. & BRODIE, M. J. 2001. The mechanisms of action of commonly used antiepileptic drugs. *Pharmacology & Therapeutics,* 90**,** 21-34.

KWAN, P., WONG, V., NG, P. W., LUI, C. H. T., SIN, N. C., POON, W. S., NG, H. K., WONG, K. S. & BAUM, L. 2009. Gene-wide tagging study of association between ABCB1 polymorphisms and multidrug resistance in epilepsy in Han Chinese. *Pharmacogenomics* 10**,** 723-732.

LABATE, A., VENTURA, P., GAMBARDELLA, A., LE PIANE, E., COLOSIMO, E., LEGGIO, U., AMBROSIO, R., CONDINO, F., MESSINA, D., LANZA, P., AGUGLIA, U. & QUATTRONE, A. 2006. MRI evidence of mesial temporal sclerosis in sporadic “benign” temporal lobe epilepsy. *Neurology,* 66**,** 562-565.

LAKHAN, R., KUMARI, R., MISRA, U. K., KALITA, J., PRADHAN, S. & MITTAL, B. 2009. Differential role of sodium channels SCN1A and SCN2A gene polymorphisms with epilepsy and multiple drug resistance in the north Indian population. *Br J Clin Pharmacol,* 68**,** 214-220.

LAVADOS, J., GERMAIN, L., MORALES, A., CAMPERO, M. & LAVADOS, P. 1992. A descriptive study of epilepsy in the district of El Salvador, Chile, 1984–1988. *Acta Neurologica Scandinavica,* 85**,** 249-256.

LESCHZINER, G., JORGENSEN, A. L., PIRMOHAMED, M., WILLIAMSON, P. R., MARSON, A. G., COFFEY, A. J., MIDDLEDITCH, C., ROGERS, J., BENTLEY, D. R., CHADWICK, D. W., BALDING, D. J., JOHNSON, M. R. & ANDREW, T. 2006. Clinical factors and ABCB1 polymorphisms in prediction of antiepileptic drug response: a prospective cohort study. *The Lancet Neurology,* 5**,** 668-676.

LESCHZINER, G. D., JORGENSEN, A. L., ANDREW, T., WILLIAMSON, P. R., MARSON, A. G., COFFEY, A. J., MIDDLEDITCH, C., BALDING, D. J., ROGERS, J., BENTLEY, D. R., CHADWICK, D., JOHNSON, M. R. & PIRMOHAMED, M. 2007. The association between polymorphisms in RLIP76 and drug response in epilepsy. *Pharmacogenomics,* 8**,** 1715-1722.

LHATOO, S. D. & SANDER, J. W. A. S. 2001. The Epidemiology of Epilepsy and Learning Disability. *Epilepsia,* 42**,** 6-9.

LI, H., WANG, B., CHANG, C., WU, M., XU, Y. & JIANG, Y. 2015a. The Roles of Variants in Human Multidrug Resistance (MDR1) Gene and Their Haplotypes on Antiepileptic Drugs Response: A Meta-Analysis of 57 Studies. *PLoS ONE,* 10**,** e0122043.

LI, S.-X., LIU, Y.-Y. & WANG, Q.-B. 2015b. ABCB1 gene C3435T polymorphism and drug resistance in epilepsy: evidence based on 8,604 subjects. *Med Sci Monit,* 21**,** 861-8.

LÖSCHER, W. & POTSCHKA, H. 2002. Role of Multidrug Transporters in Pharmacoresistance to Antiepileptic Drugs. *Journal of Pharmacology and Experimental Therapeutics,* 301**,** 7-14.

LÖSCHER, W. & SILLS, G. J. 2007. Drug resistance in epilepsy: Why is a simple explanation not enough? *Epilepsia,* 48**,** 2370-2372.

LOSSIUS, M. I., STAVEM, K. & GJERSTAD, L. 1999. Predictors for recurrence of epileptic seizures in a general epilepsy population. *Seizure,* 8**,** 476-479.

LUNA-TORTÓS, C., FEDROWITZ, M. & LÖSCHER, W. 2010. Evaluation of transport of common antiepileptic drugs by human multidrug resistance-associated proteins (MRP1, 2 and 5) that are overexpressed in pharmacoresistant epilepsy. *Neuropharmacology,* 58**,** 1019-1032.

LV, W.-P., HAN, R.-F. & SHU, Z.-R. 2014. Associations between the C3435T polymorphism of the ABCB1 gene and drug resistance in epilepsy: a meta-analysis. *International Journal of Clinical and Experimental Medicine,* 7**,** 3924-3932.

MA, C.-L., WU, X.-Y., JIAO, Z., HONG, Z., WU, Z.-Y. & ZHONG, M.-K. 2015. SCN1A, ABCC2 and UGT2B7 gene polymorphisms in association with individualized oxcarbazepine therapy. *Pharmacogenomics,* 16**,** 347-360.

MA, C.-L., WU, X.-Y., ZHENG, J., WU, Z.-Y., HONG, Z. & ZHONG, M.-K. 2014. Association of SCN1A, SCN2A and ABCC2 gene polymorphisms with the response to antiepileptic drugs in Chinese Han patients with epilepsy. *Pharmacogenomics,* 15**,** 1323-1336.

MACDONALD, B. K., JOHNSON, A. L., GOODRIDGE, D. M., COCKERELL, O. C., SANDER, J. W. A. S. & SHORVON, S. D. 2000. Factors predicting prognosis of epilepsy after presentation with seizures. *Annals of Neurology,* 48**,** 833-841.

MAGUIRE, M. J., HEMMING, K., HUTTON, J. L. & MARSON, A. G. 2008. Overwhelming heterogeneity in systematic reviews of observational anti-epileptic studies. *Epilepsy Research,* 80**,** 201-212.

MAKMOR-BAKRY, M., SILLS, G. J., HITIRIS, N., BUTLER, E., WILSON, E. A. & BRODIE, M. J. 2009. Genetic variants in microsomal epoxide hydrolase influence carbamazepine dosing. *Clinical Neuropharmacology,* 32**,** 205-12.

MAMIYA, K., HADAMA, A., YUKAWA, E., IEIRI, I., OTSUBO, K., NINOMIYA, H., TASHIRO, N. & HIGUCHI, S. 2000. CYP2C19 polymorphism effect on phenobarbitone. *Eur J Clin Pharmacol,* 55**,** 821-825.

MANGUOĞLU, E., AKDENIZ, S., DÜNDAR, N. O., DUMAN, Ö., AKTEKIN, B., HASPOLAT, Ş., BILGE, U., ÖZEL, D. & LÜLECI, G. 2011. RLIP76 Gene Variants are not Associated with Drug Response in Turkish Epilepsy Patients. *Balkan Journal of Medical Genetics : BJMG,* 14**,** 25-30.

MANNA, I., GAMBARDELLA, A., BIANCHI, A., STRIANO, P., TOZZI, R., AGUGLIA, U., BECCARIA, F., BENNA, P., CAMPOSTRINI, R., CANEVINI, M. P., CONDINO, F., DURISOTTI, C., ELIA, M., GIALLONARDO, A. T., IUDICE, A., LABATE, A., LA NEVE, A., MICHELUCCI, R., MUSCAS, G. C., PARAVIDINO, R., ZACCARA, G., ZUCCA, C., ZARA, F. & PERUCCA, E. 2011. A functional polymorphism in the SCN1A gene does not influence antiepileptic drug responsiveness in Italian patients with focal epilepsy. *Epilepsia,* 52**,** e40-4.

MARCHI, N., HALLENE, K. L., KIGHT, K. M., CUCULLO, L., MODDEL, G., BINGAMAN, W., DINI, G., VEZZANI, A. & JANIGRO, D. 2004. Significance of MDR1 and multiple drug resistance in refractory human epileptic brain. *BMC Med,* 2**,** 37.

MARCHINI, J., HOWIE, B., MYERS, S., MCVEAN, G. & DONNELLY, P. 2007. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat Genet,* 39**,** 906-913.

MARSON, A., APPLETON, R., BAKER, G., CHADWICK, D. & DOUGHTY, J. 2007. A randomised controlled trial examining the longer-term outcomes of standard versus new antiepileptic drugs The SANAD trial. *Health Technology Assessment,* 11**,** 145.

MARSON, A. G., AL-KHARUSI, A. M., ALWAIDH, M., APPLETON, R., BAKER, G. A., CHADWICK, D. W., CRAMP, C., COCKERELL, O. C., COOPER, P. N., DOUGHTY, J., EATON, B., GAMBLE, C., GOULDING, P. J., HOWELL, S. J. L., HUGHES, A., JACKSON, M., JACOBY, A., KELLETT, M., LAWSON, G. R., LEACH, J. P., NICOLAIDES, P., ROBERTS, R., SHACKLEY, P., SHEN, J., SMITH, D. F., SMITH, P. E. M., SMITH, C. T., VANOLI, A. & WILLIAMSON, P. R. 2007a. The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. *The Lancet,* 369**,** 1000-1015.

MARSON, A. G., AL-KHARUSI, A. M., ALWAIDH, M., APPLETON, R., BAKER, G. A., CHADWICK, D. W., CRAMP, C., COCKERELL, O. C., COOPER, P. N., DOUGHTY, J., EATON, B., GAMBLE, C., GOULDING, P. J., HOWELL, S. J. L., HUGHES, A., JACKSON, M., JACOBY, A., KELLETT, M., LAWSON, G. R., LEACH, J. P., NICOLAIDES, P., ROBERTS, R., SHACKLEY, P., SHEN, J., SMITH, D. F., SMITH, P. E. M., SMITH, C. T., VANOLI, A. & WILLIAMSON, P. R. 2007b. The SANAD study of effectiveness of valproate, lamotrigine, or topiramate for generalised and unclassifiable epilepsy: an unblinded randomised controlled trial. *The Lancet,* 369**,** 1016-1026.

MATTSON, R. H., CRAMER, J. A. & COLLINS, J. F. 1992. A Comparison of Valproate with Carbamazepine for the Treatment of Complex Partial Seizures and Secondarily Generalized Tonic–Clonic Seizures in Adults. *New England Journal of Medicine,* 327**,** 765-771.

MATTSON, R. H., CRAMER, J. A. & COLLINS, J. F. 1996. Prognosis for total control of complex partial and secondarily generalized tonic clonic seizures. *Neurology,* 47**,** 68-76.

MATTSON, R. H., CRAMER, J. A., COLLINS, J. F., SMITH, D. B., DELGADO-ESCUETA, A. V., BROWNE, T. R., WILLIAMSON, P. D., TREIMAN, D. M., MCNAMARA, J. O., MCCUTCHEN, C. B., HOMAN, R. W., CRILL, W. E., LUBOZYNSKI, M. F., ROSENTHAL, N. P. & MAYERSDORF, A. 1985. Comparison of Carbamazepine, Phenobarbital, Phenytoin, and Primidone in Partial and Secondarily Generalized Tonic–Clonic Seizures. *New England Journal of Medicine,* 313**,** 145-151.

MCBRIDE, G. B. 2005. A proposal for strength-of-agreement criteria for Lin's Concordance Correlation Coefficient National Institute of Water & Atmospheric Research Ltd.

MCBRIDE, M. C., BRONSTEIN, K. S., BENNETT, B., ERBA, G., PILCHER, W. & BERG, M. J. 1998. Failure of standard magnetic resonance imaging in patients with refractory temporal lobe epilepsy. *Archives of Neurology,* 55**,** 346-348.

MCHUGH, M. L. 2012. Interrater reliability: the kappa statistic. *Biochemia Medica,* 22**,** 276-282.

MILLER, D. S. 2010. Regulation of P-glycoprotein and other ABC drug transporters at the blood–brain barrier. *Trends in Pharmacological Sciences,* 31**,** 246-254.

MILLIGAN, C. J., LI, M., GAZINA, E. V., HERON, S. E., NAIR, U., TRAGER, C., REID, C. A., VENKAT, A., YOUNKIN, D. P., DLUGOS, D. J., PETROVSKI, S., GOLDSTEIN, D. B., DIBBENS, L. M., SCHEFFER, I. E., BERKOVIC, S. F. & PETROU, S. 2014. KCNT1 gain of function in 2 epilepsy phenotypes is reversed by quinidine. *Annals of Neurology,* 75**,** 581-590.

MOHANRAJ, R. & BRODIE, M. J. 2005a. Outcomes in newly diagnosed localization-related epilepsies. *Seizure,* 14**,** 318-323.

MOHANRAJ, R. & BRODIE, M. J. 2005b. Pharmacological outcomes in newly diagnosed epilepsy. *Epilepsy & Behavior,* 6**,** 382-387.

MOHANRAJ, R. & BRODIE, M. J. 2006. Diagnosing refractory epilepsy: response to sequential treatment schedules. *European Journal of Neurology,* 13**,** 277-282.
MOHANRAJ, R. & BRODIE, M. J. 2007. Outcomes of newly diagnosed idiopathic generalized epilepsy syndromes in a non-pediatric setting. *Acta Neurologica Scandinavica,* 115**,** 204-208.
MOONS, K. G. M., KENGNE, A. P., WOODWARD, M., ROYSTON, P., VERGOUWE, Y., ALTMAN, D. G. & GROBBEE, D. E. 2012. Risk prediction models: I. Development, internal validation, and assessing the incremental value of a new (bio)marker. *Heart,* 98**,** 683-690.
MOONS, K. G. M., ROYSTON, P., VERGOUWE, Y., GROBBEE, D. E. & ALTMAN, D. G. 2009. Prognosis and prognostic research: what, why, and how? *BMJ,* 338.
NAKAJIMA, Y., SAITO, Y., SHISEKI, K., FUKUSHIMA-UESAKA, H., HASEGAWA, R., OZAWA, S., SUGAI, K., KATOH, M., SAITOH, O., OHNUMA, T., KAWAI, M., OHTSUKI, T., SUZUKI, C., MINAMI, N., KIMURA, H., GOTO, Y.-I., KAMATANI, N., KANIWA, N. & SAWADA, J.-I. 2005. Haplotype structures of EPHX1 and their effects on the metabolism of carbamazepine-10,11-epoxide in Japanese epileptic patients. *Eur J Clin Pharmacol,* 61**,** 25-34.
NELIGAN, A., HAUSER, W. A. & SANDER, J. W. 2012. Chapter 6 - The epidemiology of the epilepsies. *In:* HERMANN, S. & WILLIAM, H. T. (eds.) *Handbook of Clinical Neurology.* Elsevier.
NELSON, M. R., JOHNSON, T., WARREN, L., HUGHES, A. R., CHISSOE, S. L., XU, C.-F. & WATERWORTH, D. M. 2016. The genetics of drug efficacy: opportunities and challenges. *Nat Rev Genet,* 17**,** 197-206.
NETWORK, S. I. G. 2015. SIGN 143 Diagnosis and management of epilepsy in adults. *In:* SCOTLAND, H. I. (ed.).
NICE 2004. Clinical guideline 20 the epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care. October 2004. National Institute of Clinical Excelence.
NICE 2012. The epilepsies: the diagnosis and management of the epilepsies in adults and children in primary and secondary care (update). (Clinical guideline 137.). *In:* HEALTH, D. O. (ed.). National Institute of Clinical Excelence.
NIEMEYER, M. I., CID, L. P., SEPULVEDA, F. V., BLANZ, J., AUBERSON, M. & JENTSCH, T. J. 2010. No evidence for a role of CLCN2 variants in idiopathic generalized epilepsy. *Nat Genet,* 42**,** 3-3.
NOEBELS, J., AVOLI, M., ROGAWSKI, M., OLSEN, R. & DELGADO-ESCUETA, A. 2012. Jasper's Basic Mechanisms of the Epilepsies 4th edition [Internet] 4th edition ed. Bethesda (MD): National Center for Biotechnology Information (US).
O'CONNELL, J., GURDASANI, D., DELANEAU, O., PIRASTU, N., ULIVI, S., COCCA, M., TRAGLIA, M., HUANG, J., HUFFMAN, J. E., RUDAN, I., MCQUILLAN, R., FRASER, R. M., CAMPBELL, H., POLASEK, O., ASIKI, G., EKORU, K., HAYWARD, C., WRIGHT, A. F., VITART, V., NAVARRO, P., ZAGURY, J.-F., WILSON, J. F., TONIOLO, D., GASPARINI, P., SORANZO, N., SANDHU, M. S. & MARCHINI, J. 2014. A General Approach for Haplotype Phasing across the Full Spectrum of Relatedness. *PLoS Genet,* 10**,** e1004234.
ODANI, A., HASHIMOTO, Y., OTSUKI, Y., UWAI, Y., HATTORI, H., FURUSHO, K. & INUI, K.-I. 1997. Genetic polymorphism of the CYP2C subfamily and its effect on the pharmacokinetics of phenytoin in Japanese patients with epilepsy. *Clin Pharmacol Ther,* 62**,** 287-292.
OKUMA, T. & KUMASHIRO, H. 1981. Natural History and Prognosis of Epilepsy: Report of a Multi-institutional Study in Japan. *Epilepsia,* 22**,** 35-53.
OLESEN, J., GUSTAVSSON, A., SVENSSON, M., WITTCHEN, H. U., JÖNSSON, B., ON BEHALF OF THE, C. S. G. & THE EUROPEAN BRAIN, C. 2012. The economic cost of brain disorders in Europe. *European Journal of Neurology,* 19**,** 155-162.
OTT, J. 2004. Association of genetic loci: Replication or not, that is the question. *Neurology,* 63**,** 955-958.

OTTMAN, R., BARKER-CUMMINGS, C., LEIBSON, C. L., VASOLI, V. M., HAUSER, W. A. & BUCHHALTER, J. R. 2011. Accuracy of family history information on epilepsy and other seizure disorders. *Neurology,* 76**,** 390-396.

OZGON, G. O., BEBEK, N., GUL, G. & CINE, N. 2008. Association of MDR1 (C3435T) polymorphism and resistance to carbamazepine in epileptic patients from Turkey. *Eur Neurol,* 59**,** 67-70.

PANAYIOTOPOULOS, C. 2005. *The Epilepsies: Seizures, Syndromes and Management.,* Oxfordshire (UK), Bladon Medical Publishing.

PANAYIOTOPOULOS, C. P., OBEID, T. & TAHAN, A. R. 1994. Juvenile myoclonic epilepsy: a 5-year prospective study. *Epilepsia,* 35**,** 285-96.

PARK, K. M., HUR, Y., KIM, H. Y., JI, K.-H., HWANG, T. G., SHIN, K. J., HA, S. Y., PARK, J. & KIM, S. E. 2014. Initial response to antiepileptic drugs in patients with newly diagnosed epilepsy. *Journal of Clinical Neuroscience,* 21**,** 923-926.

PEDUZZI, P., CONCATO, J., FEINSTEIN, A. R. & HOLFORD, T. R. 1995. Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology,* 48**,** 1503-1510.

PEDUZZI, P., CONCATO, J., KEMPER, E., HOLFORD, T. R. & FEINSTEIN, A. R. 1996. A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology,* 49**,** 1373-1379.

PELJTO, A. L., BARKER-CUMMINGS, C., VASOLI, V. M., LEIBSON, C. L., HAUSER, W. A., BUCHHALTER, J. R. & OTTMAN, R. 2014. Familial risk of epilepsy: a population-based study. *Brain,* 137**,** 795-805.

PENRY, J. K., DEAN, J. C. & RIELA, A. R. 1989. Juvenile myoclonic epilepsy: long-term response to therapy. *Epilepsia,* 30 Suppl 4**,** S19-23; discussion S24-7.

PERINI, E. D., SCHAEFER, R., STÖTER, M., KALAIDZIDIS, Y. & ZERIAL, M. 2014. Mammalian CORVET Is Required for Fusion and Conversion of Distinct Early Endosome Subpopulations. *Traffic,* 15**,** 1366-1389.

PITTAU, F., BISULLI, F., MAI, R., FARES, J. E., VIGNATELLI, L., LABATE, A., NALDI, I., AVONI, P., PARMEGGIANI, A., SANTUCCI, M., CAPANNELLI, D., DI VITO, L., GAMBARDELLA, A., BARUZZI, A. & TINUPER, P. 2009. Prognostic factors in patients with mesial temporal lobe epilepsy. *Epilepsia,* 50**,** 41-44.

PLACENCIA, M., SANDER, J. W., ROMAN, M., MADERA, A., CRESPO, F., CASCANTE, S. & SHORVON, S. D. 1994. The characteristics of epilepsy in a largely untreated population in rural Ecuador. *Journal of Neurology, Neurosurgery & Psychiatry,* 57**,** 320-325.

POHLMANN-EDEN, B., CROCKER, C. E. & SCHMIDT, M. H. 2013. A conceptual framework for the use of neuroimaging to study and predict pharmacoresistance in epilepsy. *Epilepsia,* 54**,** 75-79.

PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.

PRUIM, R. J., WELCH, R. P., SANNA, S., TESLOVICH, T. M., CHINES, P. S., GLIEDT, T. P., BOEHNKE, M., ABECASIS, G. R. & WILLER, C. J. 2010. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics,* 26**,** 2336-2337.

PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, MANUEL A R., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, PAUL I W., DALY, MARK J. & SHAM, PAK C. 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. *American Journal of Human Genetics,* 81**,** 559-575.

QU, J., ZHOU, B.-T., YIN, J.-Y., XU, X.-J., ZHAO, Y.-C., LEI, G.-H., TANG, Q., ZHOU, H.-H. & LIU, Z.-Q. 2012. ABCC2 Polymorphisms and Haplotype are Associated with Drug Resistance in Chinese Epileptic Patients. *CNS Neuroscience & Therapeutics,* 18**,** 647-651.

RADISCH, S., DICKENS, D., LANG, T., BONNETT, L., ARLANOV, R., JOHNSON, M. R., SCHWAB, M., MARSON, A. G. & PIRMOHAMED, M. 2014. A comprehensive functional and clinical analysis of ABCC2 and its impact on treatment response to carbamazepine. *Pharmacogenomics J,* 14**,** 481-487.

REEVES, B. C., VAN BINSBERGEN, J. & VAN WEEL, C. 2005. Systematic reviews incorporating evidence from nonrandomized study designs: reasons for caution when estimating health effects. *Eur J Clin Nutr,* 59**,** S155-S161.

REGESTA, G. & TANGANELLI, P. 1999. Clinical aspects and biological bases of drug-resistant epilepsies. *Epilepsy Research,* 34**,** 109-122.

REGISTER, E. C. T. 2004. *Pharmacogenetics of GABAergic mechanisms of benefit and harm in epilepsy: : A prospective cohort study to determine the enviromental and genetic factors associated with response to clobazam.* [Online]. Available: https://www.clinicaltrialsregister.eu/ctr-search/trial/2004-003945-41/GB#D [Accessed 24/04/2016 2016].

REINIKAINEN, K. J., KERÄNEN, T., LEHTINEN, J. M., KÄLVIÄINEN, R., SAARI, T. & RIEKKINEN, P. J. 1987. CT brain scan and EEG in the diagnosis of adult onset seizures. *Epilepsy Research,* 1**,** 178-184.

RELLING, M. V. & EVANS, W. E. 2015. Pharmacogenomics in the clinic. *Nature,* 526**,** 343-350.

REMY, S. & BECK, H. 2006. Molecular and cellular mechanisms of pharmacoresistance in epilepsy. *Brain,* 129**,** 18-35.

RIDSDALE, L., CHARLTON, J., ASHWORTH, M., RICHARDSON, M. P. & GULLIFORD, M. C. 2011. Epilepsy mortality and risk factors for death in epilepsy: a population-based study. *Br J Gen Pract,* 61**,** e271-8.

RILEY, R. D., ABRAMS, K. R., SUTTON, A. J., LAMBERT, P. C., JONES, D. R., HENEY, D. & BURCHILL, S. A. 2003. Reporting of prognostic markers: current problems and development of guidelines for evidence-based practice in the future. *British Journal of Cancer,* 88**,** 1191-1198.

RILEY, R. D., HAYDEN, J. A., STEYERBERG, E. W., MOONS, K. G. M., ABRAMS, K., KYZAS, P. A., MALATS, N., BRIGGS, A., SCHROTER, S., ALTMAN, D. G., HEMINGWAY, H. & FOR THE, P. G. 2013. Prognosis Research Strategy (PROGRESS) 2: Prognostic Factor Research. *PLOS Medicine,* 10**,** e1001380.

RODIN, E. A., RHODES, R. J. & VELARDE, N. N. 1965. The Prognosis for Patients with Epilepsy. *Journal of Occupational and Environmental Medicine,* 7**,** 560-563.

ROGAWSKI, M. A. 2013. The intrinsic severity hypothesis of pharmacoresistance to antiepileptic drugs. *Epilepsia,* 54**,** 33-40.

ROGAWSKI, M. A. & JOHNSON, M. R. 2008. Intrinsic Severity as a Determinant of Antiepileptic Drug Refractoriness. *Epilepsy Currents,* 8**,** 127-130.

ROGAWSKI, M. A. & LOSCHER, W. 2004. The neurobiology of antiepileptic drugs. *Nat Rev Neurosci,* 5**,** 553-564.

ROSENOW, F., SCHADE-BRITTINGER, C., BURCHARDI, N., BAUER, S., KLEIN, K. M., WEBER, Y., LERCHE, H., EVERS, S., KOVAC, S., HALLMEYER-ELGNER, S., WINKLER, G., SPRINGUB, J., NIEDHAMMER, M., ROTH, E., EISENSEHR, I., BERROUSCHOT, J., ARNOLD, S., SCHRÖDER, M., BEIGE, A., OERTEL, W. H., STRZELCZYK, A., HAAG, A., REIF, P. S., HAMER, H. M. & GROUP, F. T. L. S. 2012. The LaLiMo Trial: lamotrigine compared with levetiracetam in the initial 26 weeks of monotherapy for focal and generalised epilepsy—an open-label, prospective, randomised controlled multicenter study. *Journal of Neurology, Neurosurgery & Psychiatry,* 83**,** 1093-1098.

ROSES, A. D. 2001. Pharmacogenetics. *Human Molecular Genetics,* 10**,** 2261-2267.

SAINT-MARTIN, C., GAUVAIN, G., TEODORESCU, G., GOURFINKEL-AN, I., FEDIRKO, E., WEBER, Y. G., MALJEVIC, S., ERNST, J.-P., GARCIA-OLIVARES, J., FAHLKE, C., NABBOUT, R., LEGUERN, E.,

LERCHE, H., PONCER, J. C. & DEPIENNE, C. 2009. Two novel CLCN2 mutations accelerating chloride channel deactivation are associated with idiopathic generalized epilepsy. *Human Mutation,* 30**,** 397-405.

SALANOVA, V., MARKAND, O. & WORTH, R. 2002. Temporal Lobe Epilepsy Surgery: Outcome, Complications, and Late Mortality Rate in 215 Patients. *Epilepsia,* 43**,** 170-174.

SÀNCHEZ, J., CENTANARO, M., SOLÍS, J., DELGADO, F. & YÉPEZ, L. 2014. Factors predicting the outcome following medical treatment of mesial temporal epilepsy with hippocampal sclerosis. *Seizure,* 23**,** 448-453.

SANCHEZ, M. B., HERRANZ, J. L., LENO, C., ARTEAGA, R., OTERINO, A., VALDIZAN, E. M., NICOLAS, J. M., ADIN, J. & ARMIJO, J. A. 2010. Genetic factors associated with drug-resistance of epilepsy: relevance of stratification by patient age and aetiology of epilepsy. *Seizure,* 19**,** 93-101.

SÁNCHEZ, M. B., HERRANZ, J. L., LENO, C., ARTEAGA, R., OTERINO, A., VALDIZÁN, E. M., NICOLÁS, J. M., ADÍNA, J. & ARMIJO, J. A. 2010. Genetic factors associated with drug-resistance of epilepsy: Relevance of stratification by patient age and aetiology of epilepsy. *Seizure,* 19**,** 93-101.

SANDER, J. W., HART, Y. M., JOHNSON, A. L. & SHORVON, S. D. 1990. National General Practice Study of Epilepsy: newly diagnosed epileptic seizures in a general population. *Lancet,* 336**,** 1267-71.

SANETO, R. P., LEE, I.-C., KOENIG, M. K., BAO, X., WENG, S.-W., NAVIAUX, R. K. & WONG, L.-J. C. 2010. POLG DNA testing as an emerging standard of care before instituting valproic acid therapy for pediatric seizure disorders. *Seizure,* 19**,** 140-146.

SCHILLER, Y. 2009. Seizure relapse and development of drug resistance following long-term seizure remission. *Archives of Neurology,* 66**,** 1233-1239.

SCHILLER, Y. & NAJJAR, Y. 2008. Quantifying the response to antiepileptic drugs: Effect of past treatment history. *Neurology,* 70**,** 54-65.

SCHMIDT, D. 2007. How reliable is early treatment response in predicting long-term seizure outcome? *Epilepsy & Behavior,* 10**,** 588-594.

SCHMIDT, D. & SILLANPÄÄ, M. 2012. Evidence-based review on the natural history of the epilepsies. *Current Opinion in Neurology,* 25**,** 159-163 10.1097/WCO.0b013e3283507e73.

SELWA, L. M., SCHMIDT, S. L., MALOW, B. A. & BEYDOUN, A. 2003. Long-term Outcome of Nonsurgical Candidates with Medically Refractory Localization-related Epilepsy. *Epilepsia,* 44**,** 1568-1572.

SEMAH, F., PICOT, M.-C., ADAM, C., BROGLIN, D., ARZIMANOGLOU, A., BAZIN, B., CAVALCANTI, D. & BAULAC, M. 1998. Is the underlying cause of epilepsy a major prognostic factor for recurrence? *Neurology,* 51**,** 1256-1262.

SEO, T., ISHITSU, T., ONIKI, K., ABE, T., SHUTO, T. & NAKAGAWA, K. 2008. ABCC2 haplotype is not associated with drug-resistant epilepsy. *J Pharm Pharmacol,* 60**,** 631-5.

SEO, T., ISHITSU, T., UEDA, N., NAKADA, N., YURUBE, K., UEDA, K. & NAKAGAWA, K. 2006. ABCB1 polymorphisms influence the response to antiepileptic drugs in Japanese epilepsy patients. *Pharmacogenomics,* 7**,** 551-61.

SHAFER, S. Q., HAUSER, W. A., ANNEGERS, J. F. & KLASS, D. W. 1988. EEG and Other Early Predictors of Epilepsy Remission: A Community Study. *Epilepsia,* 29**,** 590-600.

SHAHWAN, A., MURPHY, K., DOHERTY, C., CAVALLERI, G. L., MUCKIAN, C., DICKER, P., MCCARTHY, M., KINIRONS, P., GOLDSTEIN, D. & DELANTY, N. 2007. The controversial association of ABCB1 polymorphisms in refractory epilepsy: An analysis of multiple SNPs in an Irish population. *Epilepsy Research,* 73**,** 192-198.

SHAZADI, K., PETROVSKI, S., ROTEN, A., MILLER, H., HUGGINS, R. M., BRODIE, M. J., PIRMOHAMED, M., JOHNSON, M. R., MARSON, A. G., O'BRIEN, T. J. & SILLS, G. J. 2014.

Validation of a multigenic model to predict seizure control in newly treated epilepsy. *Epilepsy Research,* 108**,** 1797-1805.
SHORVON, S. & LUCIANO, A. L. 2007. Prognosis of chronic and newly diagnosed epilepsy: revisiting temporal aspects. *Current Opinion in Neurology,* 20**,** 208-212 10.1097/WCO.0b013e3280555175.
SHORVON, S. D. & REYNOLDS, E. H. 1982. *Early prognosis of epilepsy*.
SIDDIQUI, A., KERB, R., WEALE, M. E., BRINKMANN, U., SMITH, A., GOLDSTEIN, D. B., WOOD, N. W. & SISODIYA, S. M. 2003. Association of Multidrug Resistance in Epilepsy with a Polymorphism in the Drug-Transporter Gene ABCB1. *New England Journal of Medicine,* 348**,** 1442-1448.
SILLANPÄÄ, M. 1993. Remission of Seizures and Predictors of Intractability in Long-Term Follow-Up. *Epilepsia,* 34**,** 930-936.
SILLANPÄÄ, M., SAARINEN, M. & SCHMIDT, D. 2014. Clinical conditions of long-term cure in childhood-onset epilepsy: A 45-year follow-up study. *Epilepsy & Behavior,* 37**,** 49-53.
SILLANPAA, M. & SCHMIDT, D. 2006. Natural history of treated childhood-onset epilepsy: prospective, long-term population-based study. *Brain,* 129**,** 617-24.
SILLANPÄÄ, M. & SCHMIDT, D. 2009. Early seizure frequency and aetiology predict long-term medical outcome in childhood-onset epilepsy. *Brain,* 132**,** 989-998.
SILLS, G. J., MOHANRAJ, R., BUTLER, E., MCCRINDLE, S., COLLIER, L., WILSON, E. A. & BRODIE, M. J. 2005. Lack of association between the C3435T polymorphism in the human multidrug resistance (MDR1) gene and response to antiepileptic drug treatment. *Epilepsia,* 46**,** 643-7.
SIREN, A., ERIKSSON, K., JALAVA, H., KILPINEN-LOISA, P. & KOIVIKKO, M. 2002. Idiopathic generalised epilepsies with 3 Hz and faster spike wave discharges: A population-based study with evaluation and long-term follow-up in 71 patients. *Epileptic Disorders,* 4**,** 209-216.
SISODIYA, S. M., LIN, W.-R., HARDING, B. N., SQUIER, M. V. & THOM, M. 2002. Drug resistance in epilepsy: expression of drug resistance proteins in common causes of refractory epilepsy. *Brain,* 125**,** 22-31.
SKOL, A. D., SCOTT, L. J., ABECASIS, G. R. & BOEHNKE, M. 2006. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet,* 38**,** 209-213.
SORANZO, N., CAVALLERI, G. L., WEALE, M. E., WOOD, N. W., DEPONDT, C., MARGUERIE, R., SISODIYA, S. M. & GOLDSTEIN, D. B. 2004. Identifying Candidate Causal Variants Responsible for Altered Activity of the ABCB1 Multidrug Resistance Gene. *Genome Research,* 14**,** 1333-1344.
SORANZO, N., KELLY, L., MARTINIAN, L., BURLEY, M. W., THOM, M., SALI, A., KROETZ, D. L., GOLDSTEIN, D. B. & SISODIYA, S. M. 2007. Lack of support for a role for RLIP76 (RALBP1) in response to treatment or predisposition to epilepsy. *Epilepsia,* 48**,** 674-83.
SPEED, D., HOGGART, C., PETROVSKI, S., TACHMAZIDOU, I., COFFEY, A., JORGENSEN, A., ELEFTHEROHORINOU, H., DE IORIO, M., TODARO, M., DE, T., SMITH, D., SMITH, P. E., JACKSON, M., COOPER, P., KELLETT, M., HOWELL, S., NEWTON, M., YERRA, R., TAN, M., FRENCH, C., REUBER, M., SILLS, G. E., CHADWICK, D., PIRMOHAMED, M., BENTLEY, D., SCHEFFER, I., BERKOVIC, S., BALDING, D., PALOTIE, A., MARSON, A., O'BRIEN, T. J. & JOHNSON, M. R. 2014a. A genome-wide association study and biological pathway analysis of epilepsy prognosis in a prospective cohort of newly treated epilepsy. *Human Molecular Genetics,* 23**,** 247-258.
SPEED, D., O'BRIEN, T. J., PALOTIE, A., SHKURA, K., MARSON, A. G., BALDING, D. J. & JOHNSON, M. R. 2014b. Describing the genetic architecture of epilepsy through heritability analysis. *Brain,* 137**,** 2680-2689.

SPOONER, C. G., BERKOVIC, S. F., MITCHELL, L. A., WRENNALL, J. A. & HARVEY, A. S. 2006. New-onset temporal lobe epilepsy in children: Lesion on MRI predicts poor seizure outcome. *Neurology,* 67**,** 2147-2153.

STEERING COMMITTEE ON QUALITY IMPROVEMENT MANAGEMENT, S. O. F. S. 2008. Febrile Seizures: Clinical Practice Guideline for the Long-term Management of the Child With Simple Febrile Seizures. *Pediatrics,* 121**,** 1281-1286.

STEFAN, H., MAY, T. W., PFÄFFLIN, M., BRANDT, C., FÜRATSCH, N., SCHMITZ, B., WANDSCHNEIDER, B., KRETZ, R., RUNGE, U., GEITHNER, J., KARAKIZLIS, C., ROSENOW, F. & KERLING, F. 2014. Epilepsy in the elderly: comparing clinical characteristics with younger patients. *Acta Neurologica Scandinavica,* 129**,** 283-293.

STEINHOFF, B. J., UEBERALL, M. A., SIEMES, H., KURLEMANN, G., SCHMITZ, B. & BERGMANN, L. 2005. The LAM-SAFE Study: Lamotrigine versus carbamazepine or valproic acid in newly diagnosed focal and generalised epilepsies in adolescents and adults. *Seizure,* 14**,** 597-605.

STEPHEN, L. J., KELLY, K., MOHANRAJ, R. & BRODIE, M. J. 2006. Pharmacological outcomes in older people with newly diagnosed epilepsy. *Epilepsy & Behavior,* 8**,** 434-437.

STEPHEN, L. J., KWAN, P. & BRODIE, M. J. 2001. Does the Cause of Localisation-Related Epilepsy Influence the Response to Antiepileptic Drug Treatment? *Epilepsia,* 42**,** 357-362.

STEWART, J. D., HORVATH, R., BARUFFINI, E., FERRERO, I., BULST, S., WATKINS, P. B., FONTANA, R. J., DAY, C. P. & CHINNERY, P. F. 2010. Polymerase γ Gene POLG determines the risk of sodium valproate-induced liver toxicity. *Hepatology,* 52**,** 1791-1796.

STEYERBERG, E. W., MOONS, K. G. M., VAN DER WINDT, D. A., HAYDEN, J. A., PEREL, P., SCHROTER, S., RILEY, R. D., HEMINGWAY, H., ALTMAN, D. G. & FOR THE, P. G. 2013. Prognosis Research Strategy (PROGRESS) 3: Prognostic Model Research. *PLOS Medicine,* 10**,** e1001381.

SU, L., DI, Q., KWAN, P., YU, N., ZHANG, Y., HU, Y. & GAO, L. 2013. Prediction for relapse and prognosis of newly diagnosed epilepsy. *Acta Neurologica Scandinavica,* 127**,** 141-147.

SUNDARAM, M., HOGAN, T., HISCOCK, M. & PILLAY, N. 1990. Factors affecting interictal spike discharges in adults with epilepsy. *Electroencephalography and Clinical Neurophysiology,* 75**,** 358-360.

SWEN, J. J., NIJENHUIS, M., DE BOER, A., GRANDIA, L., MAITLAND-VAN DER ZEE, A. H., MULDER, H., RONGEN, G. A. P. J. M., VAN SCHAIK, R. H. N., SCHALEKAMP, T., TOUW, D. J., VAN DER WEIDE, J., WILFFERT, B., DENEER, V. H. M. & GUCHELAAR, H. J. 2011. Pharmacogenetics: From Bench to Byte[mdash] An Update of Guidelines. *Clin Pharmacol Ther,* 89**,** 662-673.

SZOEKE, C., SILLS, G. J., KWAN, P., PETROVSKI, S., NEWTON, M., HITIRIS, N., BAUM, L., BERKOVIC, S. F., BRODIE, M. J., SHEFFIELD, L. J. & O'BRIEN, T. J. 2009. Multidrug-resistant genotype (ABCB1) and seizure recurrence in newly treated epilepsy: data from international pharmacogenetic cohorts. *Epilepsia,* 50**,** 1689-96.

TAN, N. C. K., HERON, S. E., SCHEFFER, I. E., PELEKANOS, J. T., MCMAHON, J. M., VEARS, D. F., MULLEY, J. C. & BERKOVIC, S. F. 2004. Failure to confirm association of a polymorphism in ABCB1 with multidrug-resistant epilepsy. *Neurology,* 63**,** 1090-1092.

TATE, S. K., DEPONDT, C., SISODIYA, S. M., CAVALLERI, G. L., SCHORGE, S., SORANZO, N., THOM, M., SEN, A., SHORVON, S. D., SANDER, J. W., WOOD, N. W. & GOLDSTEIN, D. B. 2005. Genetic predictors of the maximum doses patients receive during clinical use of the anti-epileptic drugs carbamazepine and phenytoin. *Proceedings of the National Academy of Sciences of the United States of America,* 102**,** 5507-5512.

TATE, S. K., SINGH, R., HUNG, C.-C., TAI, J. J., DEPONDT, C., CAVALLERI, G. L., SISODIYA, S. M., GOLDSTEIN, D. B. & LIOU, H.-H. 2006. A common polymorphism in the SCN1A gene associates with phenytoin serum levels at maintenance dose. *Pharmacogenet Genomics,* 16**,** 721-726 10.1097/01.fpc.0000230114.41828.73.

TATUM, W. O. I., WINTERS, L., GIERON, M., PASSARO, E. A., BENBADIS, S., FERREIRA, J. & LIPORACE, J. 2001. Outpatient Seizure Identification: Results of 502 Patients Using Computer-Assisted Ambulatory EEG. *Journal of Clinical Neurophysiology,* 18**,** 14-19.

THOMAS, R. H. & BERKOVIC, S. F. 2014. The hidden genetics of epilepsy - a clinically important new paradigm. *Nat Rev Neurol,* 10**,** 283-292.

THOMAS, R. H., WALSH, J., CHURCH, C., SILLS, G. J., MARSON, A. G., BAKER, G. A. & REES, M. I. 2014. A comprehensive neuropsychological description of cognition in drug-refractory juvenile myoclonic epilepsy. *Epilepsy & Behavior,* 36**,** 124-129.

TISHLER, D. M., WEINBERG, K. I., HINTON, D. R., BARBARO, N., ANNETT, G. M. & RAFFEL, C. 1995. MDR1 Gene Expression in Brain of Patients with Medically Intractable Epilepsy. *Epilepsia,* 36**,** 1-6.

TOMSON, T., DAHL, M.-L. & KIMLAND, E. 2007. Therapeutic monitoring of antiepileptic drugs for epilepsy. *Cochrane Database of Systematic Reviews* [Online]. Available: http://onlinelibrary.wiley.com/doi/10.1002/14651858.CD002216.pub2/abstract.

TURNBULL, D. M., HOWEL, D., RAWLINS, M. D. & CHADWICK, D. W. 1985. Which drug for the adult epileptic patient: phenytoin or valproate? *Br Med J (Clin Res Ed),* 290**,** 815-9.

U.S. FOOD AND DRUG ADMINISTRATION 2011a. Dilantin (phenytoin sodium) Injection FDA Approved Labeling.

U.S. FOOD AND DRUG ADMINISTRATION 2011b. Tegretol, Prescribing Information.

U.S. FOOD AND DRUG ADMINSTRATION 2016. Depakene (valproic acid) capsules and oral solution FDA Approved Labeling.

UFER, M., MOSYAGIN, I., MUHLE, H., JACOBSEN, T., HAENISCH, S., HÄSLER, R., FALTRACO, F., REMMLER, C., VON SPICZAK, S., KROEMER, H. K., RUNGE, U., BOOR, R., STEPHANI, U. & CASCORBI, I. 2009a. Non-response to antiepileptic pharmacotherapy is associated with the ABCC2 -24C>T polymorphism in young and adult patients with epilepsy. *Pharmacogenetics and Genomics,* 19**,** 353-362 10.1097/FPC.0b013e328329940b.

UFER, M., MOSYAGIN, I., MUHLE, H., JACOBSEN, T., HAENISCH, S., HÄSLER, R., FALTRACO, F., REMMLER, C., VON SPICZAK, S., KROEMER, H. K., RUNGE, U., BOOR, R., STEPHANI, U. & CASCORBI, I. 2009b. Non-response to antiepileptic pharmacotherapy is associated with the ABCC2 −24C>T polymorphism in young and adult patients with epilepsy. *Pharmacogenet Genomics,* 19**,** 353-362.

UFER, M., VON STULPNAGEL, C., MUHLE, H., HAENISCH, S., REMMLER, C., MAJED, A., PLISCHKE, H., STEPHANI, U., KLUGER, G. & CASCORBI, I. 2011. Impact of ABCC2 genotype on antiepileptic drug response in Caucasian patients with childhood epilepsy. *Pharmacogenet Genomics,* 21**,** 624-30.

VAHAB, S. A., SEN, S., RAVINDRAN, N., MONY, S., MATHEW, A., VIJAYAN, N., NAYAK, G., BHASKARANAND, N., BANERJEE, M. & SATYAMOORTHY, K. 2009. Analysis of Genotype and Haplotype Effects of ABCB1 (MDR1) Polymorphisms in the Risk of Medically Refractory Epilepsy in an Indian Population. *Drug Metabolism and Pharmacokinetics,* 24**,** 255-260.

VAN DER KNAAP, M. S., DEPIENNE, C., SEDEL, F. & ABBINK, T. E. M. 2015. CLCN2-Related Leukoencephalopathy. *In:* PAGON, R. A., ADAM, M. P., ARDINGER, H. H., WALLACE, S. E., AMEMIYA, A., BEAN, L. J. H., BIRD, T. D., FONG, C. T., MEFFORD, H. C., SMITH, R. J. H. & STEPHENS, K. (eds.) *GeneReviews(R).* Seattle (WA): University of Washington, Seattle

VAN DER WEIDE, J., STEIJNS, L. S. W., VAN WEELDEN, M. J. M. & DE HAAN, K. 2001. The effect of genetic polymorphism of cytochrome P450 CYP2C9 on phenytoin dose requirement. *Pharmacogenet Genomics,* 11**,** 287-291.

VAN PAESSCHEN, W., DUNCAN, J. S., STEVENS, J. M. & CONNELLY, A. 1997. Etiology and early prognosis of newly diagnosed partial seizures in adults: A quantitative hippocampal MRI study. *Neurology,* 49**,** 753-757.

VAROGLU, A. O., SAYGI, S., ACEMOGLU, H. & CIGER, A. 2009. Prognosis of patients with mesial temporal lobe epilepsy due to hippocampal sclerosis. *Epilepsy Research,* 85**,** 206-211.

VESTERGAARD, M., PEDERSEN, C. B., SIDENIUS, P., OLSEN, J. & CHRISTENSEN, J. 2007. The Long-Term Risk of Epilepsy after Febrile Seizures in Susceptible Subgroups. *American Journal of Epidemiology,* 165**,** 911-918.

VON OERTZEN, J., URBACH, H., JUNGBLUTH, S., KURTHEN, M., REUBER, M., FERNÁNDEZ, G. & ELGER, C. E. 2002. Standard magnetic resonance imaging is inadequate for patients with refractory focal epilepsy. *Journal of Neurology, Neurosurgery & Psychiatry,* 73**,** 643-647.

VON STÜLPNAGEL, C., PLISCHKE, H., ZILL, P., BÄUMEL, C., SPIEGEL, R., GRUBER, R. & KLUGER, G. 2009. Letter: Lack of association between MDR1 polymorphisms and pharmacoresistance to anticonvulsive drugs in patients with childhood-onset epilepsy. *Epilepsia,* 50**,** 1835-1837.

WALLACE, R. H., MARINI, C., PETROU, S., HARKIN, L. A., BOWSER, D. N., PANCHAL, R. G., WILLIAMS, D. A., SUTHERLAND, G. R., MULLEY, J. C., SCHEFFER, I. E. & BERKOVIC, S. F. 2001. Mutant GABAA receptor [gamma]2-subunit in childhood absence epilepsy and febrile seizures. *Nat Genet,* 28**,** 49-52.

WANG, Y., TANG, L., PAN, J., LI, J., ZHANG, Q. & CHEN, B. 2015. The recessive model of MRP2 G1249A polymorphism decrease the risk of drug-resistant in Asian Epilepsy: A systematic review and meta-analysis. *Epilepsy Research,* 112**,** 56-63.

WASSENAAR, M., LEIJTEN, F. S. S., EGBERTS, T. C. G., MOONS, K. G. M. & UIJL, S. G. 2013. Prognostic factors for medically intractable epilepsy: A systematic review. *Epilepsy Research,* 106**,** 301-310.

WEBER, Y. G. & LERCHE, H. 2008. Genetic mechanisms in idiopathic epilepsies. *Developmental Medicine & Child Neurology,* 50**,** 648-654.

WILLIAMSON, P. R., KOLAMUNNAGE-DONA, R., PHILIPSON, P. & MARSON, A. G. 2008. Joint modelling of longitudinal and competing risks data. *Statistics in Medicine,* 27**,** 6426-6438.

WIRRELL, E., WONG-KISIEL, L., MANDREKAR, J. & NICKELS, K. 2012. Predictors and course of medically intractable epilepsy in young children presenting before 36 months of age: A retrospective, population-based study. *Epilepsia,* 53**,** 1563-1569.

WIRRELL, E. C. 2010. Prognostic Significance of Interictal Epileptiform Discharges in Newly Diagnosed Seizure Disorders. *Journal of Clinical Neurophysiology,* 27**,** 239-248 10.1097/WNP.0b013e3181ea4288.

WIRRELL, E. C., WONG-KISIEL, L. C. L., MANDREKAR, J. & NICKELS, K. C. 2013. What predicts enduring intractability in children who appear medically intractable in the first 2 years after diagnosis? *Epilepsia,* 54**,** 1056-1064.

WOOLLARD, S. M., BHARGAVAN, B., YU, F. & KANMOGNE, G. D. 2014. Differential Effects of Tat Proteins Derived from HIV-1 Subtypes B and Recombinant CRF02_AG on Human Brain Microvascular Endothelial Cells: Implications for Blood–Brain Barrier Dysfunction. *Journal of Cerebral Blood Flow & Metabolism,* 34**,** 1047-1059.

YIP, V. L., MARSON, A. G., JORGENSEN, A. L., PIRMOHAMED, M. & ALFIREVIC, A. 2012. HLA genotype and carbamazepine-induced cutaneous adverse drug reactions: a systematic review. *Clinical Pharmacology And Therapeutics,* 92**,** 757-765.

ZHANG, C., KWAN, P., ZUO, Z. & BAUM, L. 2012. The transport of antiepileptic drugs by P-glycoprotein. *Advanced Drug Delivery Reviews,* 64**,** 930-942.

ZHANG, Y., YU, N., SU, L. & DI, Q. 2013. A prospective cohort study of prognosis for newly diagnosed epilepsy in east China. *BMC Neurology,* 13**,** 116.

ZHOU, B.-T., ZHOU, Q.-H., YIN, J.-Y., LI, G.-L., XU, X.-J., QU, J., LIU, D., ZHOU, H.-H. & LIU, Z.-Q. 2012. Comprehensive analysis of the association of SCN1A gene polymorphisms with the retention rate of carbamazepine following monotherapy for new-onset focal seizures in the Chinese Han population. *Clinical and Experimental Pharmacology and Physiology* 39**,** 379-384.

ZIMPRICH, F., HILGER, E., REINTHALER, E. M., STOGMANN, E., HOTZY, C., PATARAIA, E., BAUMGARTNER, C. & ZIMPRICH, A. 2012. Lack of association between ABCC2 gene variants and treatment response in epilepsy. *Pharmacogenomics,* 13**,** 185-190.

ZIMPRICH, F., STOGMANN, E., BONELLI, S., BAUMGARTNER, C., MUELLER, J. C., MEITINGER, T., ZIMPRICH, A. & STROM, T. M. 2008. A functional polymorphism in the SCN1A gene is not associated with carbamazepine dosages in Austrian patients with epilepsy. *Epilepsia,* 49**,** 1108-1109.

ZIMPRICH, F., SUNDER-PLASSMANN, R., STOGMANN, E., GLEISS, A., DAL-BIANCO, A., ZIMPRICH, A., PLUMER, S., BAUMGARTNER, C. & MANNHALTER, C. 2004. Association of an ABCB1 gene haplotype with pharmacoresistance in temporal lobe epilepsy. *Neurology,* 63**,** 1087-1089.

Appendix 1

EpiPGX paper CRF final version as of 09/05/2012 (created by EpiPGX Consortium).

EpiPGX

Site Code:	☐ ☐ ☐ ☐
Date of CRF	DD/MM/YYYY
Data Entered By:	
Data Source	**Medical** ☐
	Database ☐ **Spe**
	Other ☐ **Spe**
DNA nr:	
DNA Source *(tick only one)*	**Blood** ☐ **Sali** ☐ **Brain** ☐
	Othe ☐ **Spe**
Genotyped	**Yes** ☐ **N** ☐
	If 'Yes' Platform
Imputed	**Yes** ☐ **N** ☐

C.I.
Professor Sanjay M Sisodiya
Department of Clinical & Experimental Epilepsy
UCL Institute of Neurology
Queen Square
WC1N 3BG
London
United Kingdom

General Guidelines Regarding Completing The CRFS

All forms should be completed in **black ink** in a clear manner. Any changes or corrections should be made by drawing a line through the data, making sure the amended script is still legible, entering the corrected information and initialling and dating the change.

Note: *Do not use Correction Fluid*

Following standard notation should be used in the event that values or answers cannot be provided:

- NA: Not applicable
- NK: Not known
- ND: Not done
- NR: Not retrievable/Not available

Patient Enrolment - Notes

Patient Enrolment

Patient Demographics

Sex	Male ☐ Female ☐
Date of Birth	DD/MM/YYYY

Ethnic origin

Please tick ethnic origin, as self reported by Patient	European ☐ African ☐ Chinese ☐ Japanese ☐ Indian ☐ Pakistani ☐ Middle Eastern ☐ Mixed race ☐ Specify ☐ Other ☐ Specify ☐
Date of Recruitment / DNA	DD/MM/YYYY
Date of epilepsy diagnosis	DD/MM/YYYY
Start date of contemporary	DD/MM/YYYY
Status at start of contemporary records *(please tick all that apply)*	New epilepsy ☐ Existing epilepsy on treatment ☐ Existing epilepsy off treatment ☐ Existing epilepsy, off treatment but previous AED <12 months ☐ >12 months ☐ Existing epilepsy on treatment ☐
Date of most recent update to Contemporary medical records	DD/MM/YYYY

Epilepsy Diagnosis Notes

Epilepsy Diagnosis

Epilepsy syndrome according to 1989 ILAE classification:	
Hippocampal sclerosis	Y ☐ N ☐ If 'Yes' please state below L ☐ R ☐
Confirmed by:	MRI ☐ Histology ☐ Unknown ☐

Known Progressive Neurological Disorder

	Y ☐ N ☐ Unkno ☐ If 'Yes' please tick one
Type	Neoplastic / ☐ Metabolic ☐ Infectious ☐ Inflammatory ☐ Degenerative ☐ Genetic ☐ O ☐ Specify ☐
Onset Date	DD/MM/YYYY

Neurological Examination - Notes

Neurological Examination

Nor ☐ Abnor ☐ N ☐

If abnormal *(please tick all that apply)*

- Higher cortical functions ☐
- Speech disturbance ☐
- Cranial nerve ☐
- Motor abnormalities ☐
- Sensor abnormalities ☐
- Co-ordination ☐

Other ☐ Specif

Seizures

Seizures *(please tick all that apply)*

- Primary generalised tonic ☐
- Absence ☐
- Clonic ☐
- Tonic ☐
- Atonic ☐
- Myoclonic ☐
- Simple partial ☐
- Secondarily GTCS ☐
- Unclassified partial ☐
- Unclassified GTCS ☐
- Uncertain epileptic ☐
- Non-epileptic ☐

Seizure Frequency - Notes

Seizure Frequency

Date / Year of first ever seizure?	DD/MM/YYYY
Total number of seizures prior to	Absolute number *(if known):*

Total number of seizures prior to first ever AED

GTC seizures	Absolute number *(if* 1 ☐ 3 ☐ 6 ☐ 1 ☐ 2 ☐ NK ☐
Non-GTC seizures *(any type):*	Absolute number *(if known)* ☐ 1 ☐ 3 ☐ 6 ☐ 1 ☐ 2 ☐ NK ☐
Combined *(if type unknown)*	Absolute number *(if known)* 1 ☐ 3 ☐ 6 ☐ 1 ☐ 2 ☐ NK ☐
Any periods of remission	Y ☐ N ☐ Unk ☐
Remission 1	Start date DD/MM/YYYY Stop Date DD/MM/YYYY
Remissions 2	Start date DD/MM/YYYY Stop Date DD/MM/YYYY
Remission 3	Start date DD/MM/YYYY Stop Date DD/MM/YYYY
Remission 4	Start date DD/MM/YYYY Stop Date DD/MM/YYYY
Total number of seizures in the	GTC seizures Non-GTC seizures *(any* Combined *(if type*
Total number of seizures in the prior to epilepsy surgery *(if*	GTC seizures Non-GTC seizures *(any* Combined *(if type*

Non Medical Epilepsy Treatment – Notes

Non Medical Epilepsy Treatment

Date of Procedure	DD/MM/YYYY
Type	Specify

AED History / General

Non-epileptic seizures:	Y ☐ N ☐ NK ☐
Total number of appropriate and AED trials	specify
Number of AEDS that failed due to efficacy at minimum therapeutic	specify
Responder to VPA + LTG in	Yes ☐ No ☐ NK ☐ NA ☐

Investigations (First Ever) - Notes

Investigations (First Ever)

EEG *(please tick all that apply)*	Normal ☐ Abnormal ☐ Abnormal *(non-specific)* ☐ Not done ☐ Not known ☐
Imaging - MRI/CT *(Please tick all that*	Normal ☐ Abnormal *(focal)* ☐ Abnormal *(non-specific)* ☐ Not ☐ Not known ☐

Investigations 2 (most relevant, if different from above)

EEG *(please tick all that apply)*	Normal ☐ Abnormal ☐ Abnormal *(non specific)* ☐ Not done ☐ Not known ☐
Imaging - MRI/CT *(Please tick all that*	Normal ☐ Abnormal *(focal)* ☐ Abnormal *(non-specific)* ☐ Not done ☐ Not known ☐

AED History per AED - Notes

[1] Reason for stopping:

1. inadequate seizure control
2. unacceptable adverse effects
3. both inadequate seizure control and unacceptable adverse effects
4. remission
5. unknown

[2] Outcome of this AED trial

1. response
2. failure
3. extreme late response
4. unclassified
5. unknown

AED History per AED

AED trial number	specify
AED generic name	specify
Start date	DD/MM/YYYY
Stop date (NA if patient still on AED)	DD/MM/YYYY N/A
Patient known to be non-adherent for	Yes No Unknown
Initiated as *(Tick all that apply)*	Monotherapy Add-on
Maximum dose reached	Dose
Average monthly seizure frequency 6 months before starting this AED	GTC seizures Non-GTC seizures *(any type):* Combined (if type unknown):
Average monthly seizure frequency on	GTC seizures Non-GTC seizures *(any type)*: Combined *(if type unknown):*
Reason for stopping [1]	1 2 3 4 5 Other Specify
Can this AED trial be considered And adequate	Yes No Unknown
Outcome of this AED trial [2]	1 2 3 4 5

Adverse Drug Reactions

Adverse drug reactions	Yes ☐ No ☐ Unknown ☐ If 'Yes' please tick all that apply and label ADR 1, 2...etc
Behavioural disorder **Ascertained through:**	Agitation ☐ Aggression ☐ Irritability ☐ Confusion ☐ Prospective neuropsychological ☐ Retrospective contemporary data ☐
Cognitive impairment **Ascertained through:**	Amnesia ☐ Forgetfulness ☐ Concentratio ☐ Slowed mentation ☐ Prospective neuropsychological ☐ Retrospective contemporary data ☐
Hepatic dysfunction	Highest GOT/AST *(IU/l + reference* ☐ Highest GPT/ALT *(IU/l + reference* ☐ Highest PT *(seconds):* ☐
Hyponatraemia / SIADH	Lowest plasma Na+ *(mEq/l):* ☐
Neutropenia / agranulocytosis	Lowest absolute neutrophil count *(/µl):* ☐
Psychosis	Psychosis according to ICD10 ☐ Confirmed by yes ☐ no ☐ NK ☐
Cutaneous adverse reactions	Confirmed by lymphocyte transformation test: Yes ☐ No ☐ Unknown ☐ Confirmed by dermatologist Yes ☐ No ☐ Unknown ☐

Type

Maculopapular exanthema	Hypersensitivity syndrome *(please tick all that apply)* Prolonged recovery phase *(despite AED* Fever Internal organ involvement *(tick & specify* Liver ☐ Specify ☐ Gastro-intestinal ☐ Specify Kidne ☐ Specify ☐ Lung ☐ Specify ☐ Central Nervous ☐ Specify Heart ☐ Specify ☐ Muscle ☐ Specify ☐ Thyroid ☐ Specify ☐ Haematologica ☐ Specify Lymphoid ☐ Spec Stevens-Johnson syndrome ☐ Toxic epidermal necrolysis ☐
Speech Disorder	Speech difficulties witnessed by physician Yes ☐ No ☐ Unknown ☐
Tremor	Mild ☐ Modera ☐ S ☐ Family History of tremor Yes ☐ No ☐ Unknown ☐
Visual field constriction	Confirmed Goldmann ☐ Humphrey ☐ OCT
Weight change	Weight before KG Weight after KG Weight change: KG

Type (Continued) - Notes

Type (Continued)

Miscellaneous *(please tick all that*	Arthralgia ☐ Cardiac conduction ☐ Syncope ☐ Depressed mood ☐ Depression ☐ Encephalopathy ☐ Erectile dysfunction ☐ Impotence ☐ Gastrointestinal symptoms ☐ Gum hypertrophy ☐ Gingivitis ☐ Hair loss ☐ Headache ☐ Hirsutism ☐ Insomnia ☐ Osteoporosis ☐ Paraesthesia ☐ Polycystic ovaries ☐ polycystic ovary syndrome ☐ Renal stones ☐ Sleepine ☐ Somnolence ☐ Sedation ☐ Fatigue ☐ Lethargy ☐ Thrombocytopenia ☐ Unsteadiness ☐ Dizziness ☐ Vertigo ☐ Ataxia ☐ Urinary retention ☐ Other ☐ Specify ☐

Summary - Notes

Summary

Patient can be included in WP02	Has newly diagnosed epilepsy ☐ Immediate remission ☐ Deferred remission ☐ No remission ☐ First AED failed due to ☐ First AED failed due to ADRs ☐
Patient can be included in WP03	Is drug resistant ☐ Is drug responsive ☐ Is extremely drug resistant ☐ Is control for extremely drug ☐
Patient can be included in WP04	
1.	Specify ☐ Response ☐ Failure ☐ Extreme late response
2.	Specify ☐ Response ☐ Failure ☐ Extreme late response
3..	Specify ☐ Respo ☐ Failure ☐ Extreme late response
4.	Specify ☐ Respo ☐ Failure ☐ Extreme late response
5.	Specify ☐ Respo ☐ Failure ☐ Extreme late response
6.	Specify ☐ Respo ☐ Failure ☐ Extreme late response
7.	Specify ☐ Respo ☐ Failure ☐ Extreme late response

Summary (Continued)

8.	Specify AED's ☐ Response ☐ Failure ☐ Extreme late response
9.	Specify AED's ☐ Response ☐ Failure ☐ Extreme late response
10.	Specify AED's ☐ Response ☐ Failure ☐ Extreme late response

Patient can be included in WP05				
1.	AED	☐	ADR	☐
2.	AED	☐	ADR	☐
3.	AED	☐	ADR	☐
4.	AED	☐	ADR	☐
5.	AED	☐	ADR	☐
6.	AED	☐	ADR	☐
7.	AED	☐	ADR	☐
8.	AED	☐	ADR	☐
9.	AED	☐	ADR	☐
10.	AED	☐	ADR	☐

Patient can be included in WP06						
1.	AED	☐	case	☐	control	☐
2.	AED	☐	case	☐	control	☐
3.	AED	☐	case	☐	control	☐
4.	AED	☐	case	☐	control	☐
5.	AED	☐	case	☐	control	☐
6.	AED	☐	case	☐	control	☐
7.	AED	☐	case	☐	control	☐
8.	AED	☐	case	☐	control	☐
9.	AED	☐	case	☐	control	☐
10.	AED	☐	case	☐	control	☐

Appendices

ADR's - Notes

ADR's

ADR __ __	ADR dosing at time of reaction ADR start date DD/MM/YYYY ADR stop date *(N/A if* DD/MM N ☐ Can ADR be reasonably attributed to AED Yes ☐ No ☐ Unknow ☐ Did ADR lead to dose reduction Yes ☐ No ☐ Unknow ☐ Did ADR lead to ADR withdrawal Yes ☐ No ☐ Unknow ☐
ADR __ __	ADR dosing at time of reaction ADR start date DD/MM/YYYY ADR stop date *(N/A if ongoing)* DD/M N ☐ Can ADR be reasonably attributed to AED Yes ☐ No ☐ Unknown ☐ Did ADR lead to dose reduction Yes ☐ No ☐ Unknown ☐ Did ADR lead to ADR withdrawal Yes ☐ No ☐ Unknown ☐
ADR __ __	ADR dosing at time of reaction ADR start date DD/MM/YYYY ADR stop date *(N/A if ongoing)* DD/MM N ☐ Can ADR be reasonably attributed to AED Yes ☐ No ☐ Unknown ☐ Did ADR lead to dose reduction Yes ☐ No ☐ Unknown ☐ Did ADR lead to ADR withdrawal Yes ☐ No ☐ Unknown ☐

Pregnancy - Notes

Pregnancy

Number of this Specify

Personal Malformation Yes No Unknown

If 'yes' please specify

Classification of

Family history Yes No Unknown

If 'yes' please specify

Classification of MCM

Folic acid Yes No Unknown

If 'yes' please specify below

Preconceptual folic Yes No

Dose

AEDs taken First 3 months

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Name Dose

Pregnancy (Continued) - Notes

[1] Outcome

1 = Completed
2 = Miscarriage
3 = Induced abortion
4 = Still birth
5 = Ongoing

Pregnancy (Continued)

Other	Yes ☐ No ☐ Unknown ☐ If 'yes' please specify drug (generic) names and doses below Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐
Other conception	Yes ☐ No ☐ Unknown ☐ If 'yes' please specify drug (generic) names and doses below Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐ Name ☐ Dose ☐
Alcohol use	Yes ☐ No ☐ Unknown ☐
Smoking in	Yes ☐ No ☐ Unknown ☐
Gestational	☐
Birth	☐
Sex	Male ☐ Female ☐ Unknown
Outcome	1 ☐ 2 ☐ 3 ☐ 4 ☐ 5

Pregnancy Continued - Notes

[1] Major congenital malformation

1 = Cardiac malfunction
2 = Cleft palate
3 = Facial dysmorphism (other than cleft palate)
4 = Gastro-intestinal tract defect
5 = Genito-urinary tract defect
6 = Neural tube defect
7 = Spina bifida
8 = Skeletal malfunction
9= unspecified

Pregnancy Continued

Major congenital malformation [1	Yes ☐ No ☐ Unknown ☐ If 'Yes' tick all that apply 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7 ☐ 8 ☐ 9 ☐ Other ☐ Specify ☐
Neurodevelopmental delay	Yes ☐ No ☐ Unknown ☐

Appendix 2

Full set of phenotype definitions developed and applied by the whole consortium as of 10/09/2014.

WP 02 – Genome-based biomarkers of early treatment response in newly-diagnosed epilepsy

Only patients with newly diagnosed epilepsy will be included for this purpose.
- **Newly diagnosed epilepsy** is defined as the occurrence of ≥2 clinically definite unprovoked epileptic seizures in the previous year, *or* the occurrence of one seizure and the clinician decides to start AEDs.
- Prospective data are preferred, but retrospective data are allowed if based on contemporary evidence (i.e. continuous records from initiation of the first AED onwards from a specialist epilepsy centre).
- Patients with known progressive neurological disorders at time of first AED initiation are excluded
- Make a note of those patients with prior AED exposure or rescue treatment + the indication.
- **Focal & generalized epilepsy** are defined as in the 1989 ILAE classification

Task 1: Identifying genome-based biomarkers of remission with *first well-tolerated* drug.
- **Remission** is defined as any continuous period of ≥12 months complete seizure freedom. A titration period of 2 months is taken into account.
- **Immediate remission** is defined as remission within 14 months of starting the first well-tolerated AED.
- **Deferred remission** is defined as remission that is first recorded later than 14 months after starting the first well-tolerated AED (and prior to initiation of another AED).
- **No remission** is defined as continuing seizures after starting the first *well-tolerated, adequately applied* and *appropriate* AED (see Appendix for guidance).

Task 2: Identifying genome-based biomarkers that distinguish general and selective drug responsiveness.
No additional definitions needed here.

Task 3: Identifying genomic biomarkers of first drug failure.
- **Treatment failure (withdrawal) due to lack of efficacy** is defined as continuing seizures after the first *appropriate* AED has been *adequately* applied. Where seizure frequency data is not available, there should be clear written evidence that the AED was withdrawn specifically because it failed to control seizures.
- **Treatment failure due to ADRs**: In order to be attributed to the AED in question, ADRs should (i) occur within 6 months of initiation of the AED (not applicable for visual field defects), (ii) lead to withdrawal of the AED, and (iii) not be attributed to another cause by the treating clinician or the phenotyping clinician.

ADRs are sub-classified into 'on-target' and 'off-target' reactions:
- **On-target** : neurological in origin, related to dose or concentration, associated with dose increase, resolve on dose reduction or drug withdrawal.
- **Off-target:** non-neurological, not necessarily related to dose or concentration, not necessarily associated with dose increase, do not necessarily resolve on dose reduction or drug withdrawal.

WP 03 – Genome-based biomarker discovery for broad AED resistance

Task 2: Undertaking GWAS for broad drug resistance
- **Broad AED resistance** is defined as seizures recurring at a frequency of ≥4/year over the last year till latest recorded visit, despite *adequate* trials of ≥2 *tolerated* and *appropriately chosen and used* AED schedules, whether as monotherapies or in combination (see Appendix for guidance).
- **Drug responsiveness** is defined as freedom from seizures for ≥12 months up to latest recorded visit.

- Patients who have had epilepsy surgery and fulfilled the above criteria for broad AED resistance before surgery can also be included.
- Patients who have had epilepsy surgery can never be classified as drug responsive thereafter
- Patients known to be systematically non-adherent should be excluded.
(1 seizure a year due to non-adherence may be disregarded)
- Make a note of patients with a history of alternating remissions (≥12 months) and relapses on any AED + specify the number of remissions.
- During its long and sometimes fluctuating course a person's epilepsy may not fulfil the definition criteria for either drug resistant or drug-responsive epilepsy at certain time points. In such circumstances, drug responsiveness should be temporarily classified as ''undefined.''

Task 3: Search for rare variants causing broad drug resistance
- **Extreme AED resistance** is defined as:
- Clearly identifiable (MRI- or histologically-confirmed), stable lesion
- Follow-up of ≥5 years
- Seizures recurring at a frequency of ≥4/year over the last year till latest data entry, despite *adequate* trials of ≥5 *tolerated* and *appropriately chosen and used* AED schedules (whether as monotherapies or in combination).
- Never been seizure-free for ≥12 months
- Patients who have had epilepsy surgery and fulfilled the above criteria before surgery can also be included
- **Drug responsiveness** is defined as:
- Clearly identifiable, stable lesion
- Free from seizures for ≥5 years up to latest data entry
- Patients cannot have had surgery for their epilepsy.

WP 04 – Genome-based biomarker discovery for late response to specific AEDs

Only patients who have failed at least one AED trial due to lack of efficacy will be included for this purpose (although this AED does not necessarily have to be withdrawn).

Task 1: Search for variants for late response to specific AEDs in focal epilepsies
- AEDs to be included: any.
- Response to specific AEDs is defined as freedom from seizures lasting for ≥12 months which according to the treating clinician and/or the phenotyper can be attributed to the AED, e.g. after an increase of dose (and prior to initiation of another treatment for epilepsy).
- **Failure of specific AEDs** is defined as seizures recurring at >50% of the pretreatment seizure frequency after the *appropriate* AED has been *adequately* applied (see Appendix for guidance).
- Patients known to be systematically non-adherent should be excluded.
(1 seizure a year due to non-adherence may be disregarded)

- **Extreme late AED response** is defined as:
    - Patients who failed *adequate* trials of ≥2 *tolerated* and *appropriately chosen and used* AED schedules (whether as monotherapies or in combination).
- Became seizure free for ≥12 months after reaching the minimum therapeutic dose of the AED 'X' (see Appendix 1)
- Where appropriate, concomitant AEDs have been withdrawn to leave the patient on AED 'X' in monotherapy.
- Patients cannot have had surgery for their epilepsy.

NB: cases in which some details are missing (e.g. exact seizure frequencies, exact AED exposure duration) but a response can be derived from the clinical context can also be included. These are coded as follows:
R1= response + all criteria fulfilled
R2= response with details missing
R3= extreme late response
F1= failure + all criteria fulfilled
F2= failure with details missing

Task 2: Search for variants for late response to specific AEDs in generalised epilepsies
- AEDs to be included: LTG & VPA, maybe others (e.g. LEV)
- Same definitions as above
- Make a note of patients responding to VPA+LTG combination therapy, when both AEDs alone have failed, and to those not responding to VPA+LTG

<u>WP 05 – Genome-based biomarker discovery for specific ADRs</u>

- In order to be attributed to the AED in question, **ADRs** should:
- Occur within 6 months of initiation of the AED (not applicable for visual field defects) *and*
- Where appropriate, lead to withdrawal or dose reduction of the AED *and*
- Where appropriate, reverse or improve after withdrawal or dose reduction (e.g. not for visual field defects) *and*

- Not be attributed to another cause by the treating clinician or the phenotyping clinician.

NB: fulfilment of all of the above criteria (as well as specific ADR definitions below) classifies a case as '**strict**'; fulfilment of only criterion 4 classifies a case as '**loose**'.

- Specific ADR definitions:
Rash/hypersensitivity:
Maculopapular exanthema (MPE) is defined as mild rash
Hypersensitivity syndrome (HSS) is defined as:
Rash
Involvement of ≥1 internal organs (see Appendix 3)
Prolonged recovery phase (despite AED withdrawal) and/or fever
Stevens-Johnson syndrome (SJS) and toxic epidermal necrolysis (TEN) are defined as skin detachment (1-30% for SJS and >30% for TEN)
Weight change is defined as a >10% weight gain or loss
The definitions of **behavioural disorder and cognitive impairment** are based on the prospective neuropsychological assessments developed by participant 07. Retrospective cases can be included on the basis of contemporary evidence (i.e. improvement on neuropsychological testing after coming off TPM).
Speech disorder on TPM at a dose of ≤100 mg/d and witnessed by a physician.
Neutropenia is defined as a documented neutrophil count of <1000/µl.
Hyponatremia is defined as a documented plasma sodium concentration of <125 mEq/l.
Visual field defect is measured by optical coherence tomography according to the protocol established by participant 01.
Tremor is defined as the appearance of a new-onset, severe tremor in patients <40 years old on VPA monotherapy at a dose of ≤1000 mg/d and without a family history of tremor.
Hepatotoxicity is defined as A) ≥5-fold elevation above the ULN of ALT or AST *or* B) ≥2-fold elevation above the ULN for ALP (when there is no bony cause for rise in ALP) *or* C) ≥3-fold elevation in ALT plus bilirubin elevation exceeding 2x ULN.
Thrombocytopaenia is defined as a documented platelet count of <100 $*10^9$/L (<100.000/µl).
Psychotic reaction is defined as vivid hallucinations, misidentifications, delusions and/or ideas of reference (often of a paranoid or persecutory nature), psychomotor disturbances (excitement or stupor), and an abnormal affect, which may range from intense fear to ecstasy. The sensorium is usually clear but some degree of clouding of consciousness, though not severe confusion, may be present.
The diagnosis is confirmed by a psychiatrist.
In addition, **any idiosyncratic reaction** on any AED will be recorded.

<u>WP 06 – Genome-based biomarker discovery for valproate teratogenesis</u>

- **Cases** are women with epilepsy who were taking any AED (either in monotherapy or polytherapy) during a period including the first trimester of pregnancy where the infant was identified to have any major congenital malformation (MCM).
- **Controls** are women with epilepsy who were taking any AED (either in monotherapy or polytherapy) during a period including the first trimester of pregnancy where the infant did not have a MCM.

- A **MCM** is defined as any structural abnormality with surgical, medical, functional or cosmetic importance
- MCMs resulting in spontaneous or induced abortion are also included.
- Make a note in case of siblings with MCM.

Appendix to phenotypes defintions.

Antiepileptic drug	Focal seizures	Primary generalised tonic-clonic seizures	Absence seizures	Other primary generalised / unclassified seizures	Minimum therapeutic dose (mg)	Defined daily dose (mg)
Carbamazepine	✓	✓	×	×	600	1000
Clobazam	✓	✓	✓	✓	10	20
Clonazepam	✓	✓	✓	✓	4	8
Eslicarbazepine	✓	✓	×	×	800	800
Ethosuximide	×	×	✓	×	1000	1250
Felbamate	✓	✓	✓	✓	1200	2400
Gabapentin	✓	×	×	×	1200	1800
Lacosamide	✓	×	×	×	200	300
Lamotrigine	✓	✓	✓	✓	150	300
Levetiracetam	✓	✓	✓	✓	1000	1500
Oxcarbazepine	✓	✓	×	×	900	1050
Phenobarbital	✓	✓	×	✓	60	100
Phenytoin	✓	✓	×	×	200	300
Pregabalin	✓	×	×	×	300	300
Primidone	✓	✓	×	✓	750	1250
Retigabine	✓	×	×	×	600	900
Rufinamide	×	×	×	✓	1200	1400
Tiagabine	✓	×	×	×	30	30
Topiramate	✓	✓	✓	✓	100	300
Valproate	✓	✓	✓	✓	1000	1500
Vigabatrin	✓	×	×	×	1000	2000
Zonisamide	✓	✓	✓	✓	150	200

- **Adequate:** the AED has been administered during an adequate time period and at an adequate dose (see also Tables 1 & 2 below).
- **Appropriate:** previously shown to be effective, preferably in randomized controlled studies (e.g. ethosuximide for focal seizures is considered inappropriate and therefore does not count). Please note that some patients may ‘‘fail’’ several AEDs before they fail one that is ‘‘appropriate’’ and in a way that is ‘‘informative.’’

Table 1. AEDs, appropriate seizure types, minimum and defined daily doses for AED monotherapy in adult patients

The above doses are given for guidance only. Final judgment of adequacy of any AED trial is left to the discretion of the treating clinician and/or phenotyper.

Table 2. Minimum dataset required to determine whether the trial of a therapeutic intervention is informative

Nature of the intervention (e.g., type of drug, in the case of antiepileptic drug treatment)
Mode of application (e.g., formulation, dose, dosing interval, and patient's compliance in case of an antiepileptic drug)
Duration of exposure
Occurrence of seizures and adverse effects during the trial period
Whether there was any effort to optimize dose
Reason(s) for discontinuation (if applicable)
Unsatisfactory seizure control
Adverse effects
Long-term seizure freedom
Psychosocial reasons, for example, planning for pregnancy
Administrative reasons, for example, lost to follow up
Financial issues, for example, cannot afford treatment
Patient/caretaker preference
Other reasons

Kwan P. et al, Epilepsia 2010 (http://www.ncbi.nlm.nih.gov/pubmed/19889013)

Appendix 3

SANAD to EpiPGX data transfer dictionary (created by Dr Ben Francis, University of Liverpool). Dictionary was based on EpiPGX CRF and corresponding numbers reflect item in SANAD CRF.

General note: use "NA" whenever data are not available

1. General data

- Site code (e.g. UCL, ULB…): ULIV
- Date of CRF completion:
- Person entering data: Ben Francis
- Data source (tick one):
  - o medical records
  - o database (specify): SANAD
  - o other (specify):
- DNA nr:
- DNA source (tick one):
  - o blood
  - o saliva
  - o brain tissue
  - o other (specify):
- Genotyped : yes

If yes: -Platform:

-Imputed: yes

- Gender: male / female 2A
- DOB: 2A
- Ethnicity: European, African, Asian, other (specify), mixed (specify), unclassified
- Date of recruitment / DNA collection: 2B
- Date of epilepsy diagnosis:
- Start date of continuous contemporary clinical records: 2B
- Status at start of continuous contemporary clinical records (tick one): 2A
  - o new epilepsy
  - o existing epilepsy, off treatment

- existing epilepsy, off treatment but previous AED treatment (<12 months / >12 months)
- existing epilepsy, on treatment

- Date of latest recorded visit: 4: Follow-up

2. <u>Epilepsy diagnosis</u>

- Epilepsy syndrome according to 1989 ILAE classification: 1.5
- Hippocampal sclerosis: yes / no
  - If yes: - left / right
    - confirmed by: MRI / histology / unknown

3. <u>Known progressive neurological disorder</u>

- Yes / no 1.2
- If yes: - Type (tick one):
  - Neoplastic/paraneoplastic
  - Metabolic
  - Infectious
  - Inflammatory
  - Degenerative
  - Genetic
  - Other
- Details:
- Onset date:

4. **<u>Neurological examination</u>**

- Normal / abnormal / NA 1.1
- If abnormal: tick one or more:
  - Higher cortical functions
  - Speech disturbance

- Cranial nerve abnormalities
- Motor abnormalities
- Sensory abnormalities
- Coordination
- Other

- Details: 1.1

5. Seizures

- Seizure types (tick any that apply): 1.4
  - primary generalized tonic clonic (GTC)
  - absence
  - clonic
  - tonic
  - atonic
  - myoclonic
  - simple partial
  - complex partial
  - secondarily GTC
  - unclassified partial
  - unclassified GTC
  - uncertain epileptic
  - non-epileptic

6. Seizure frequency

- Date/year of first ever seizure: 1.4
- Total number of seizures prior to first ever AED
  1) GTC seizures: 1.4
     - Absolute number (if known):
     - Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown
  2) Non-GTC seizures (any type):

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

3) Combined (if type unknown): 1.4

- Absolute number (if known):

- Categorical: 1-2, 3-5, 6-10, 11-20, 21+, unknown

- Did the patient experience at least one seizure in the 12 months prior to starting the first AED? Yes / No / Unknown 1.4

- Any periods of remission: unknown

- Total number of seizures in the last 12 months prior to latest recorded visit

1) GTC seizures: 1.4

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

2) Non-GTC seizures (any type): 1.4

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

3) Combined (if type unknown): 1.4

- Absolute number (if known):

- Categorical: 1-3, ≥4, unknown

8. Investigations 1 (first ever)

- EEG: 3B1

- normal
- abnormal (epileptiform)
- abnormal (non-specific)
- not done
- not known

- Imaging (MRI / CT – delete as appropriate): 3C

- normal
- abnormal (focal)
- abnormal (non-specific)
- not done

- not known

9. Investigations 2 (most relevant, if different from above)

- EEG: 3B1
  - normal
  - abnormal (epileptiform)
  - abnormal (non-specific)
  - not done
  - not known

- Imaging (MRI / CT – delete as appropriate): 3C
  - normal
  - abnormal (focal)
  - abnormal (non-specific)
  - not done
  - not known

10. AED history per AED (*Fill out 1 form per AED tried, in chronological order)*

- AED generic name: 2A
- Start date: 2B
- Stop date (NA if patient still on AED): 4.3
- Patient adherent for this AED: unknown
- Initiated as monotherapy / add-on / unknown (delete as appropriate) 4.2
- Maximum dose reached (mg/d): 2B
- Average monthly seizure frequency during ≥3 months before starting this AED

1) GTC seizures: 1.4

2) Non-GTC seizures (any type): 1.4

3) Combined (if type unknown): 1.4

- Average monthly seizure frequency while on this AED (and prior to any subsequent change in epilepsy treatment)

1) GTC seizures: 4.1

2) Non-GTC seizures (any type): 4.1

3) Combined (if type unknown): 4.1

- Reason for stopping: 4.6

- o inadequate seizure control
- o unacceptable adverse effects
- o both inadequate seizure control and unacceptable adverse effects
- o remission
- o other (specify):
- o unknown
- o NA

- Can this AED trial be considered appropriate and adequate: yes / no / unknown 4.3

- Outcome of this AED trial: response / failure / extreme late response / unclassified / unknown 4.3/4.6

- **Adverse drug reactions**: yes / no / unknown 4.6

If yes (tick any, label "ADR1, 2 …" and fill out details below for each ADR):

o Behavioural disorder (agitation / aggression / irritability / confusion). Ascertained through: 4.6

- o Prospective neuropsychological assessment
- o Retrospective contemporary data 4.6

o Cognitive impairment (amnesia / forgetfulness / concentration difficulties / slowed mentation). Ascertained through: 4.6

- o Prospective neuropsychological assessment
- o Retrospective contemporary data 4.6

o Hepatic dysfunction 4.6

- o Highest GOT/AST (IU/l + reference values):
- o Highest GPT/ALT (IU/l + reference values):
- o First elevated AP (IU/l + reference values):
- o Highest AP (IU/l + reference values):

- First elevated bilirubin (mg/dl + reference values):
- Highest bilirubin (mg/dl + reference values):
- Highest PT (seconds):

- Hyponatraemia / SIADH 4.6
  - Lowest plasma Na+ (mEq/l):
- Neutropenia / agranulocytosis 4.6
  - Lowest absolute neutrophil count (/µl):
- Psychosis 4.6
  - Psychosis according to ICD10 definition
  - Confirmed by psychiatrist: yes / no / unknown
- Cutaneous adverse reactions 4.6
  - Confirmed by lymphocyte transformation test: yes / no / unknown
  - Confirmed by dermatologist: yes / no / unknown

Type:

- Maculopapular exanthema
- Hypersensitivity syndrome (tick any that apply):
  - ☐ Prolonged recovery phase (despite AED withdrawal)
  - ☐ Fever
  - ☐ Internal organ involvement (tick + give details):
    - ☐ Liver
    - ☐ Gastro-intestinal
    - ☐ Kidney
    - ☐ Lung
    - ☐ Central nervous system
    - ☐ Heart
    - ☐ Muscle
    - ☐ Thyroid
    - ☐ Haematological
    - ☐ Lymphoid system
- Stevens-Johnson syndrome
- Toxic epidermal necrolysis

- Speech disorder

- Speech difficulties witnessed by physician
- Thrombocytopenia
  - Lowest thrombocyte count (/µl):
- Tremor
  - Mild / moderate / severe
  - Family history of tremor: yes / no / unknown
- Visual field constriction
  - Confirmed by : Goldmann / Humphrey / OCT
- Weight change
  - Weight before AED: kg
  - Weight after AED: kg
  - Weight change: kg
- Miscellaneous
  - Cardiac conduction abnormality / syncope
  - Depressed mood / depression
  - Encephalopathy
  - Other:

Fill out details for each of the above ADRs:

- **ADR1** : 4.6
- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- **ADR2** : 4.6
- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):

- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- **ADR3** : 4.6
- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

- **ADR4** : 4.6
- ADR dosing at time of reaction (mg/d):
- ADR start date:
- ADR stop date (NA if ongoing):
- Can ADR be reasonably attributed to AED: yes / no / unknown
- Did ADR lead to dose reduction: yes / no / unknown
- Did ADR lead to ADR withdrawal: yes / no / unknown

12. Summary *(tick any that apply)*

- ☐ Patient has generalized epilepsy 1.5
- ☐ Patient has focal epilepsy 1.5/1.6
- ☐ WP02: Patient has newly-diagnosed epilepsy: yes

If yes: 4.1

- ○ Immediate remission
- ○ Deferred remission

- o No remission
- o First AED failed due to inefficacy 4.3
- o First AED failed due to ADRs 4.3

 Patient can be included in WP03: unknown

- o Is drug resistant
  - Total number of adequate, appropriate and tolerated AEDs:
- o Is drug responsive 4.1
- o Is extremely drug resistant
- o Is a control for extremely drug resistant

 Patient can be included in WP04 (list AED names and outcome below): yes / no / unknown

Include only patients who have failed at least one AED trial due to lack of efficacy!

1) ……4.2…… : response / failure / extreme late response 4.1/4.3
2) …… : response / failure / extreme late response
3) …… : response / failure / extreme late response
4) …… : response / failure / extreme late response
5) …… : response / failure / extreme late response
6) …… : response / failure / extreme late response
7) …… : response / failure / extreme late response
8) …… : response / failure / extreme late response
9) …… : response / failure / extreme late response
10) …… : response / failure / extreme late response
11) …… : response / failure / extreme late response
12) …… : response / failure / extreme late response

Responder to VPA + LTG in combination only: yes / no / unknown / NA

 Patient can be included in WP05 (list AED names and ADRs below) 4.2

AED name	ADR

Appendix 4

Table with the full list and results of SNPs include in GWAS with dominant inheritance model. Left sided columns are univariate analysis whereas right sided are after adjustment for significant clinical factors.

CHR	rsid	Position	info	P value univariate analysis		CHR	rsid	position	info	P value adjusted analysis
11	chr11:107982503:D	107982503	0.993793	0.000187703		11	chr11:107982503:D	108000000	0.993793	0.00032
13	chr13:85111634:D	85111634	0.993656	5.65715E-07		13	chr13:85111634:D	85111634	0.993656	1.63E-06
2	chr2:12639256:D	12639256	0.998011	0.000063158		2	chr2:12639256:D	12639256	0.998011	0.0000411
3	chr3:184791599:D	184791599	0.995788	1.15897E-06		3	chr3:184791599:D	184791599	0.995788	1.097E-05
8	chr8:33952031:D	33952031	0.991777	0.000609636		8	chr8:33952031:D	33952031	0.991777	6.06E-06
4	rs10033588	25754026	0.989671	0.000040116		4	rs10033588	25754026	0.989671	6.77E-06
2	rs10166451	233479640	0.984334	0.000396878		2	rs10166451	233479640	0.984334	0.0028749
10	rs10828664	24814503	1	0.00153067		10	rs10828664	24814503	1	0.000984
10	rs11014141	24814223	0.991028	0.00227079		10	rs11014141	24814223	0.991028	0.001581
12	rs11068099	117034918	0.996878	7.47441E-05		12	rs11068099	117000000	0.996878	0.0000816
12	rs11068100	117035067	0.995765	0.000162272		12	rs11068100	117000000	0.995765	0.000189
12	rs11068101	117036472	0.997107	0.000105591		12	rs11068101	117000000	0.997107	0.000106
2	rs11096544	18767384	1	0.00451517		2	rs11096544	18767384	1	0.0001044
14	rs11160087	93129982	0.964623	6.92383E-06		14	rs11160087	93129982	0.964623	6.76E-06
2	rs114442102	233478687	0.978042	0.000448273		2	rs114442102	233478687	0.978042	0.0032818
1	rs11584217	202982240	0.995155	0.00248646		1	rs11584217	202982240	0.995155	8.071E-05
3	rs116415722	163130184	0.979521	1.52929E-05		3	rs116415722	163000000	0.979521	0.000486
4	rs11727253	25770524	0.987954	1.21643E-05		4	rs11727253	25770524	0.987954	0.0000208
10	rs12268215	24814142	0.988971	0.00169348		10	rs12268215	24814142	0.988971	0.001027
8	rs12676472	33972141	0.979782	0.000342764		8	rs12676472	33972141	0.979782	0.0000125

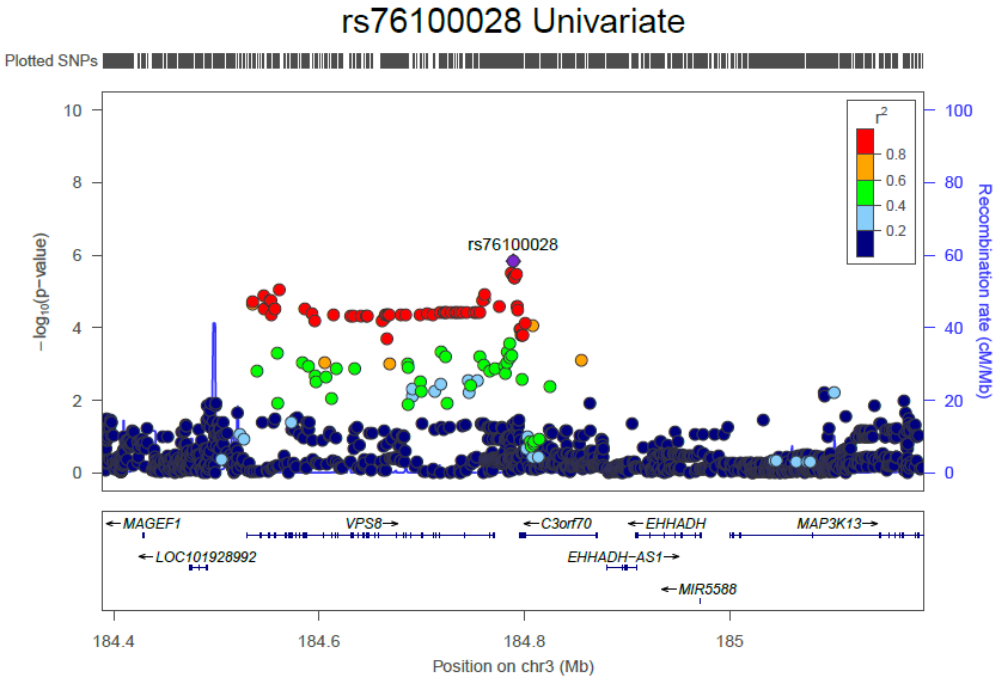
1	rs12725494	202996589	0.984188	0.00151839		1	rs12725494	202996589	0.984188	6.298E-05
1	rs12747999	228796705	0.991716	0.000316075		1	rs12747999	228796705	0.991716	1.389E-06
13	rs12855432	85111586	0.998517	7.33108E-07		13	rs12855432	85111586	0.998517	2.18E-06
16	rs140775682	31290557	0.983502	9.79997E-05		16	rs140775682	31290557	0.983502	0.0000148
11	rs144267602	108319016	0.991432	5.0952E-07		11	rs144267602	108000000	0.991432	8.15E-06
13	rs1446766	85112660	0.992194	8.1794E-07		13	rs1446766	85112660	0.992194	2.46E-06
13	rs1446767	85110202	0.999908	7.33108E-07		13	rs1446767	85110202	0.999908	2.18E-06
13	rs1446768	85110189	0.999913	7.33108E-07		13	rs1446768	85110189	0.999913	2.18E-06
13	rs1446770	85108206	1	7.33108E-07		13	rs1446770	85108206	1	2.18E-06
13	rs1446782	85075320	0.999171	7.08097E-07		13	rs1446782	85075320	0.999171	3.07E-06
2	rs1530632	46590384	0.992424	0.00229165		2	rs1530632	46590384	0.992424	0.0001708
15	rs17204959	61481102	0.998619	0.00045448		15	rs17204959	61481102	0.998619	0.002085
12	rs1874438	117034544	0.998148	7.79997E-05		12	rs1874438	117000000	0.998148	0.0000803
6	rs1876550	106213051	0.999807	0.00116238		6	rs1876550	106000000	0.999807	0.0000444
10	rs1937818	122511887	0.991968	0.000031822		10	rs1937818	123000000	0.991968	4.02E-06
3	rs2004207	184790414	0.997468	1.44048E-06		3	rs2004207	185000000	0.997468	0.0000143
3	rs2004208	184790407	0.997456	1.44048E-06		3	rs2004208	185000000	0.997456	0.0000143
7	rs2390716	22807922	0.997207	1.75423E-05		7	rs2390716	22807922	0.997207	0.0000628
11	rs35219733	107989244	0.997716	0.00022103		11	rs35219733	108000000	0.997716	0.000778
10	rs4077084	86420972	0.98958	8.33869E-06		10	rs4077084	86420972	0.98958	6.65E-06
1	rs4412590	228720037	0.987423	0.148345		1	rs4412590	228720037	0.987423	0.12195
10	rs4508149	86032370	0.99926	0.000575802		10	rs4508149	86032370	0.99926	0.000162
3	rs4687016	184792444	1	1.15897E-06		3	rs4687016	185000000	1	0.000011
3	rs526346	176210523	0.994518	0.000159524		3	rs526346	176000000	0.994518	0.001169
1	rs55882333	45757393	0.953498	6.29281E-06		1	rs55882333	45757393	0.953498	NA
3	rs59264451	184787463	0.997319	9.64284E-07		3	rs59264451	185000000	0.997319	6.23E-06
3	rs713601	194459554	0.977859	0.0106843		3	rs713601	194000000	0.977859	0.016326

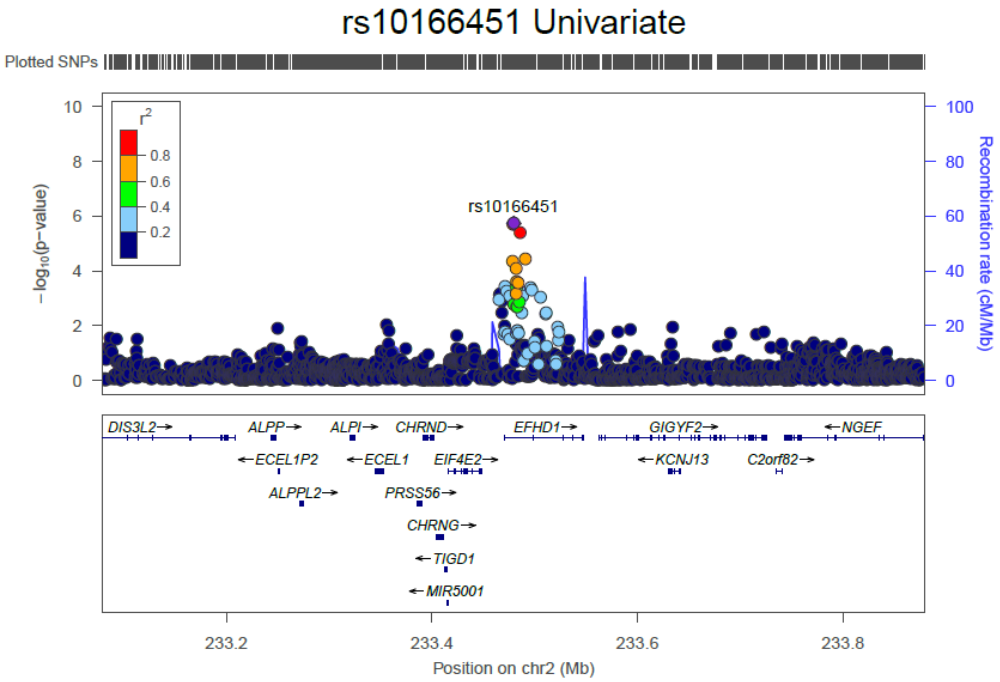
12	rs7137033	117035544	0.997265	0.000132653		12	rs7137033	117000000	0.997265	0.000132
10	rs72835240	129359897	0.993357	2.97218E-05		10	rs72835240	129000000	0.993357	3.28E-06
10	rs72835241	129360226	0.996921	9.55415E-05		10	rs72835241	129000000	0.996921	0.0000154
3	rs73053778	184068305	0.982405	6.71075E-05		3	rs73053778	184000000	0.982405	7.51E-06
12	rs7314617	117031946	1	0.000160505		12	rs7314617	117000000	1	0.000211
13	rs7337244	85108915	0.999995	7.33108E-07		13	rs7337244	85108915	0.999995	2.18E-06
13	rs7337326	85109087	0.999991	7.33108E-07		13	rs7337326	85109087	0.999991	2.18E-06
13	rs7337490	85109106	0.999991	7.33108E-07		13	rs7337490	85109106	0.999991	2.18E-06
13	rs7338700	85109002	0.999995	7.33108E-07		13	rs7338700	85109002	0.999995	2.18E-06
13	rs7339022	85109110	0.999991	7.33108E-07		13	rs7339022	85109110	0.999991	2.18E-06
3	rs744306	184789748	0.998469	1.30976E-06		3	rs744306	185000000	0.998469	0.0000138
3	rs747507	184790127	0.998684	1.30976E-06		3	rs747507	185000000	0.998684	0.0000138
10	rs753795	86028579	0.998618	0.0006652		10	rs753795	86028579	0.998618	0.000179
10	rs753796	86028463	0.997041	0.000428384		10	rs753796	86028463	0.997041	0.000102
2	rs7607256	233485696	1	0.000573714		2	rs7607256	233485696	1	0.0043705
3	rs76100028	184789338	0.99714	4.20872E-07		3	rs76100028	185000000	0.99714	3.92E-06
3	rs7624642	184561541	0.994769	3.48511E-06		3	rs7624642	185000000	0.994769	0.0000243
10	rs7899757	86024199	0.998313	0.0006652		10	rs7899757	86024199	0.998313	0.000179
10	rs7904309	86025310	0.998455	0.0006652		10	rs7904309	86025310	0.998455	0.000179
10	rs7911510	86342382	0.999702	0.000324525		10	rs7911510	86342382	0.999702	0.00019
12	rs7968805	117034246	0.998675	7.79997E-05		12	rs7968805	117000000	0.998675	0.0000803
13	rs7991045	85080910	0.9987	7.3992E-07		13	rs7991045	85080910	0.9987	2.83E-06
13	rs7996369	85111713	0.997633	7.33108E-07		13	rs7996369	85111713	0.997633	2.18E-06
13	rs7997475	85112098	0.995518	8.1794E-07		13	rs7997475	85112098	0.995518	2.46E-06
13	rs9319058	85108979	0.999995	7.33108E-07		13	rs9319058	85108979	0.999995	2.18E-06
6	rs9320132	106246658	0.998851	0.00269345		6	rs9320132	106000000	0.998851	0.000189
6	rs9400994	117565453	0.993985	0.00637079		6	rs9400994	118000000	0.993985	0.000121

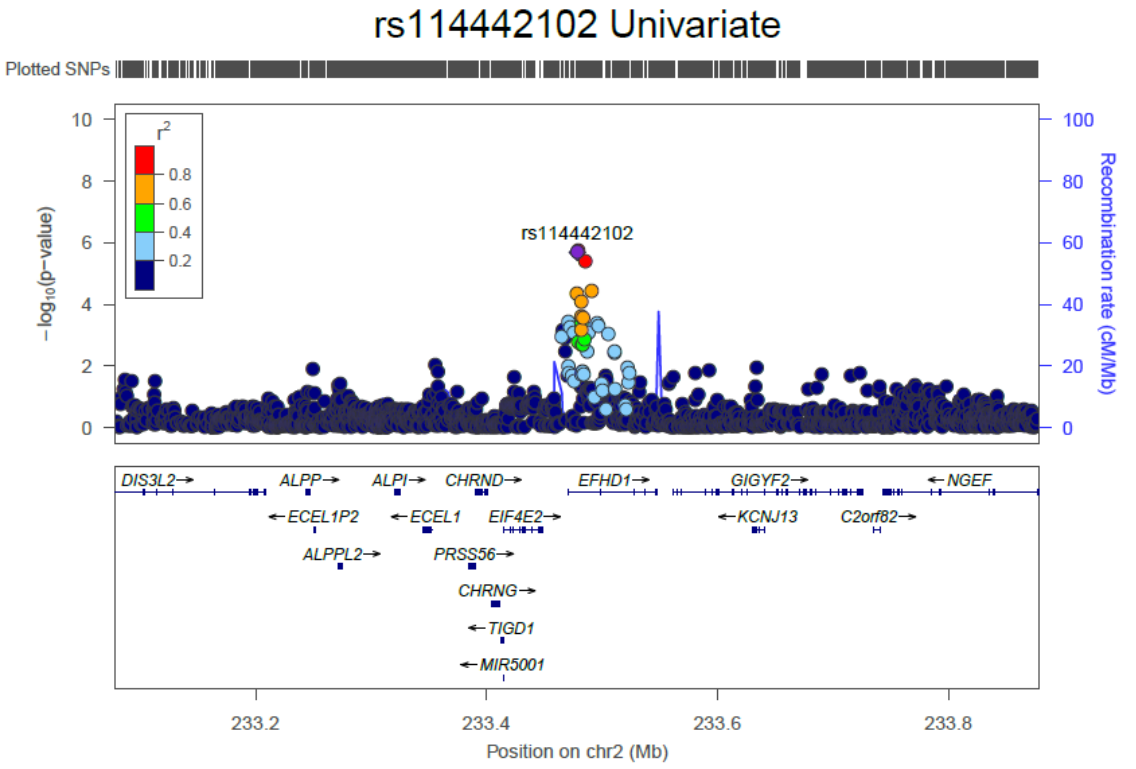
6	rs9460589	21044895	0.997508	4.52763E-07		6	rs9460589	21044895	0.997508	0.000056
13	rs9531612	85109868	0.99994	7.33108E-07		13	rs9531612	85109868	0.99994	2.18E-06
13	rs9531613	85110793	0.999867	7.33108E-07		13	rs9531613	85110793	0.999867	2.18E-06
13	rs9531615	85111407	0.999803	7.33108E-07		13	rs9531615	85111407	0.999803	2.18E-06
13	rs9546751	85109272	0.9995	6.05683E-07		13	rs9546751	85109272	0.9995	1.81E-06
13	rs9602496	85108738	1	7.33108E-07		13	rs9602496	85108738	1	2.18E-06

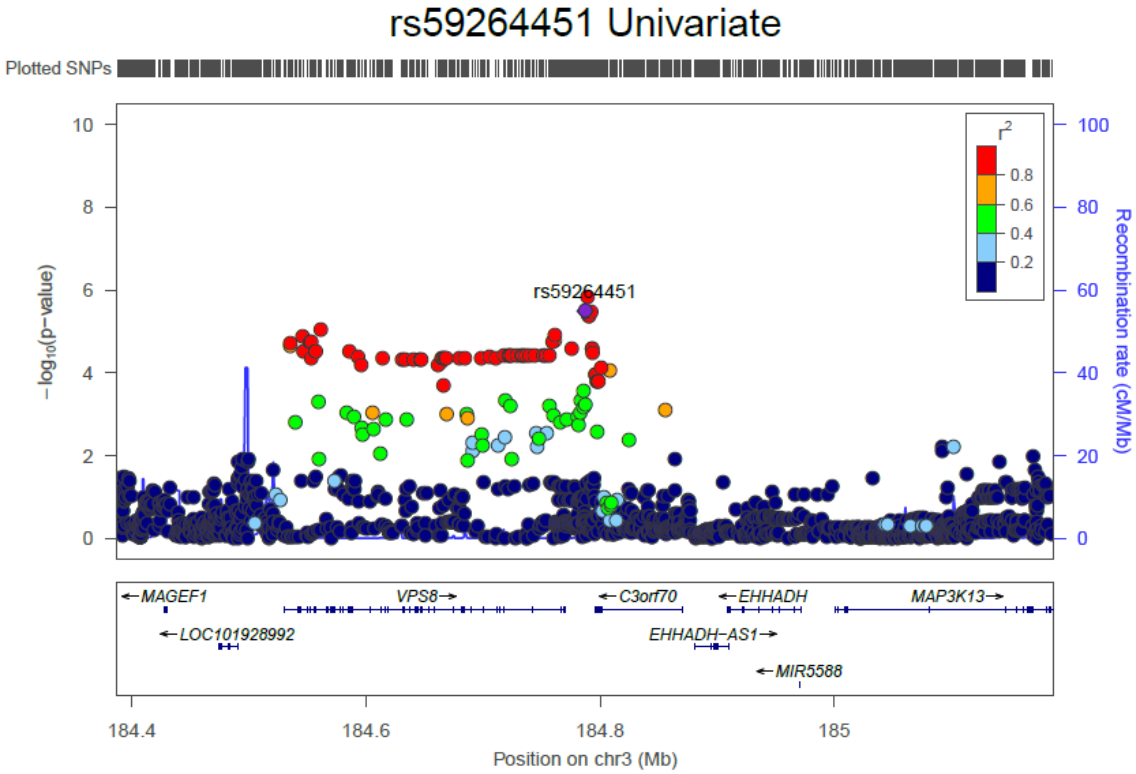
Appendix 5

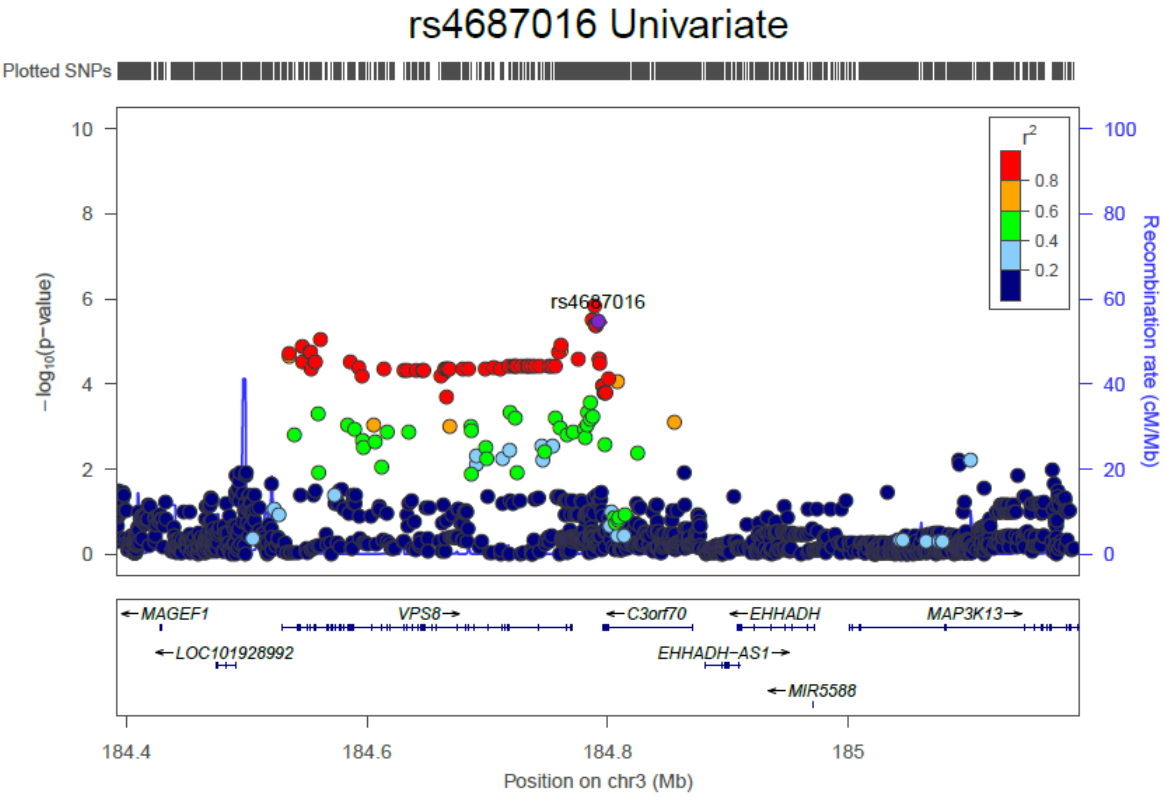
Full list of LocusZoom plots for all SNPs with p value less than $5x10^{-5}$ from univariate GWAS for newly diagnosed focal epilepsy after application of first well tolerated anticonvulsant.

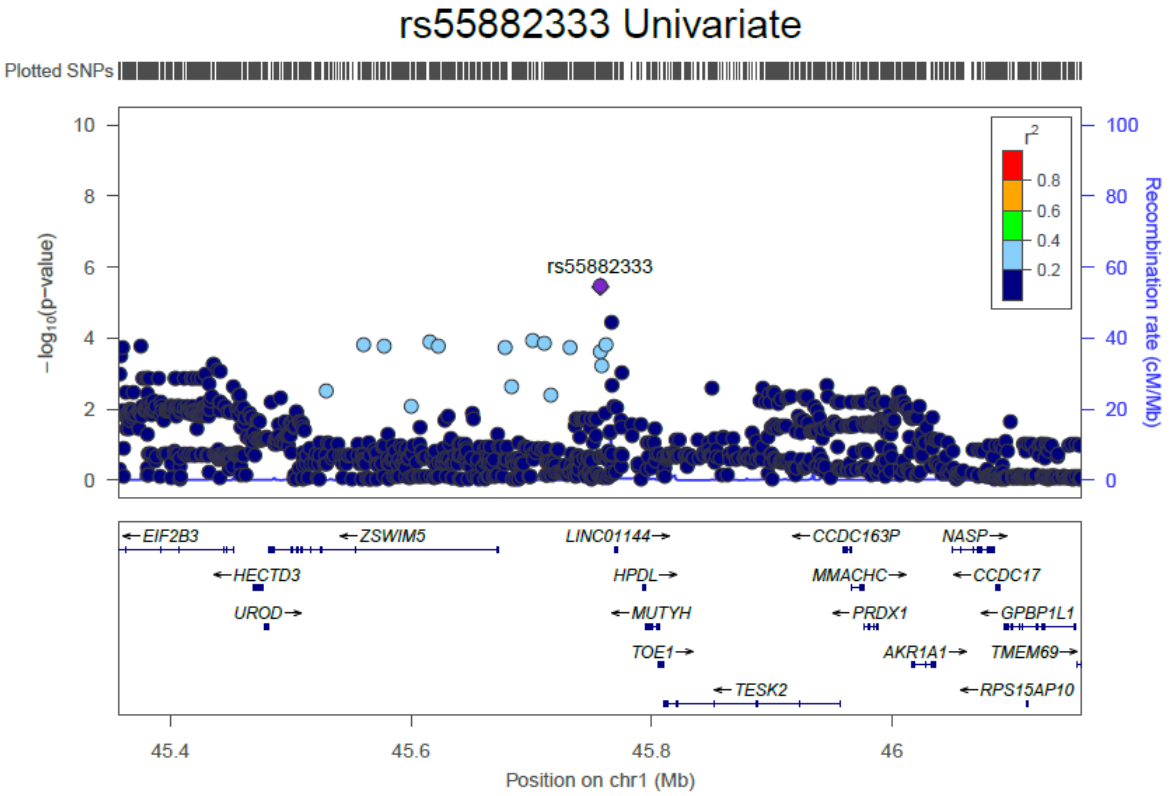


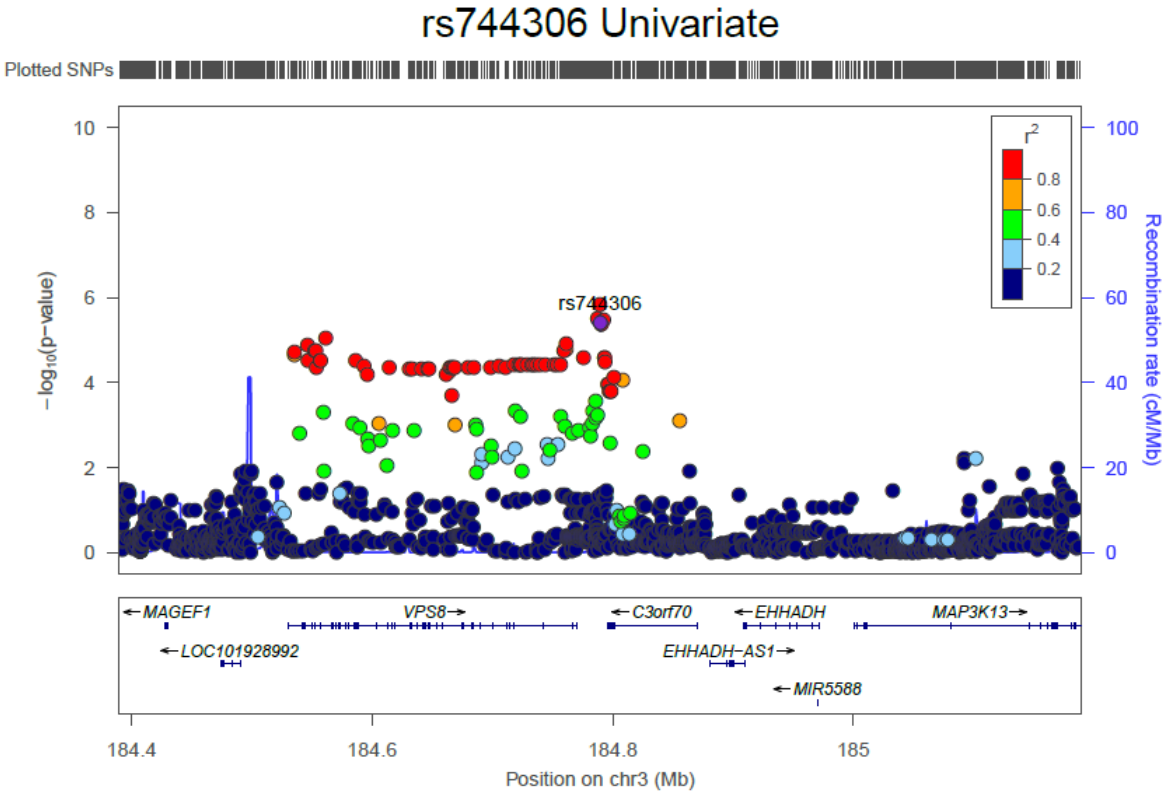


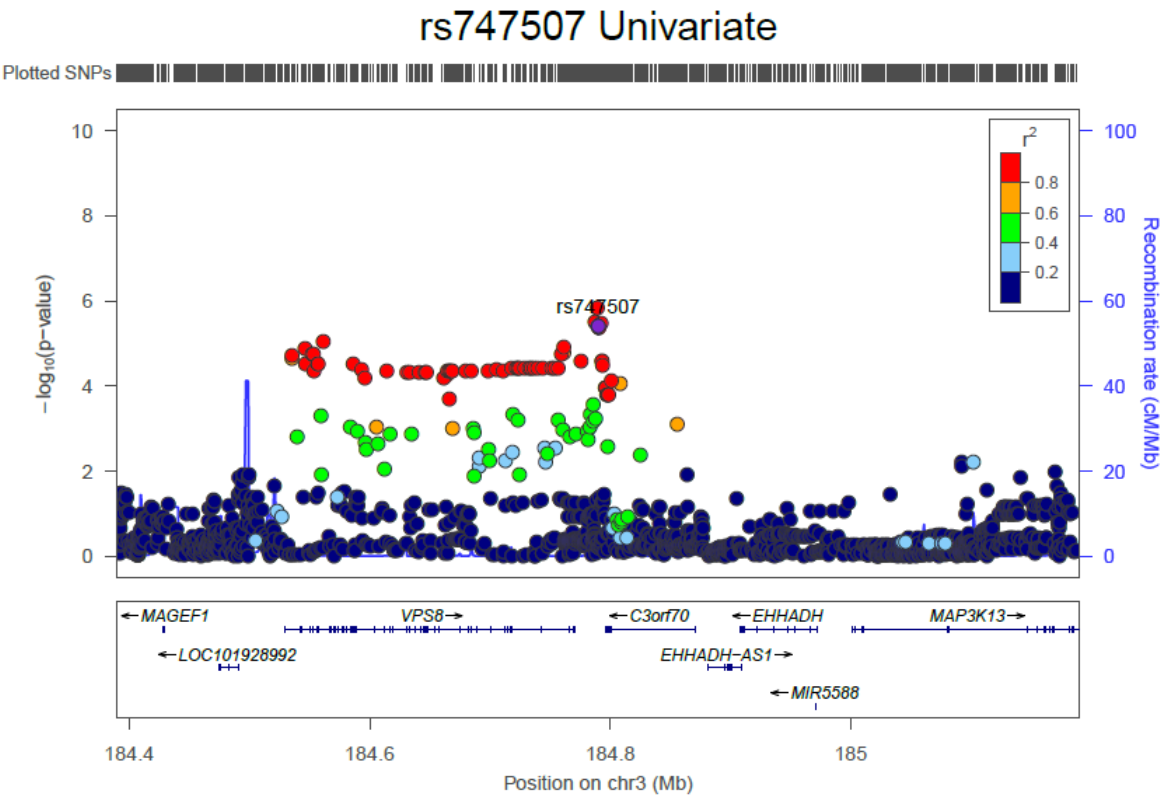

rs114442102 Univariate
Plotted SNPs
rs114442102
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
DIS3L2→
ALPP→
ALPI→
CHRND→
EFHD1→
GIGYF2→
←NGEF
←ECEL1P2
←ECEL1
EIF4E2→
←KCNJ13
C2orf82→
ALPPL2→
PRSS56→
CHRNG→
←TIGD1
←MIR5001
233.2
233.4
233.6
233.8
Position on chr2 (Mb)

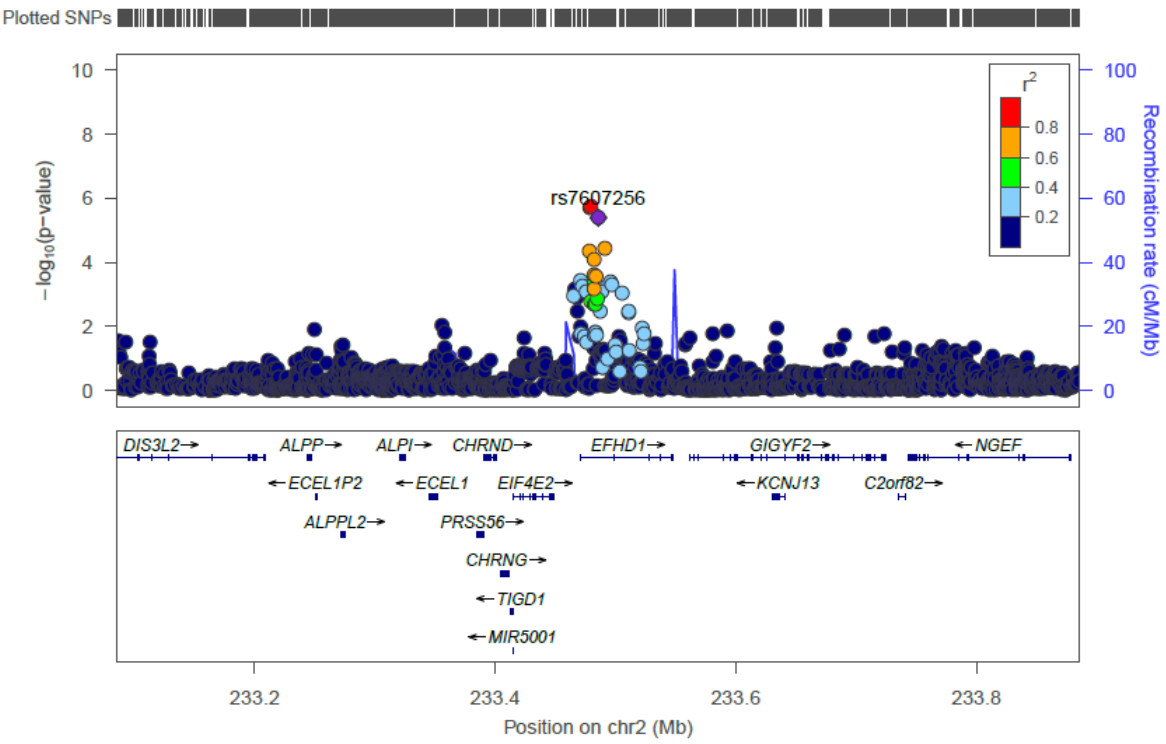

rs59264451 Univariate
Plotted SNPs
rs59264451
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
←MAGEF1
VPS8→
←C3orf70
←EHHADH
MAP3K13→
←LOC101928992
EHHADH−AS1→
←MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

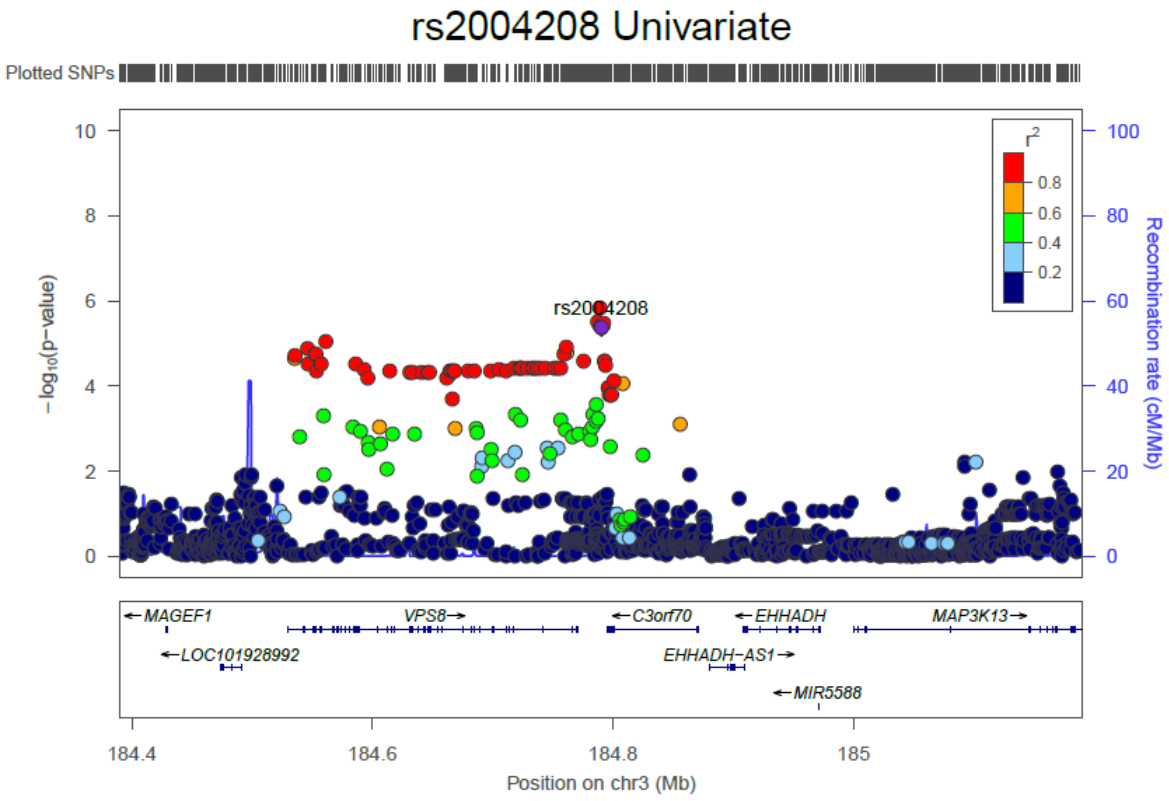

rs4687016 Univariate
Plotted SNPs
rs4687016
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MAGEF1
VPS8
C3orf70
EHHADH
MAP3K13
LOC101928992
EHHADH−AS1
MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

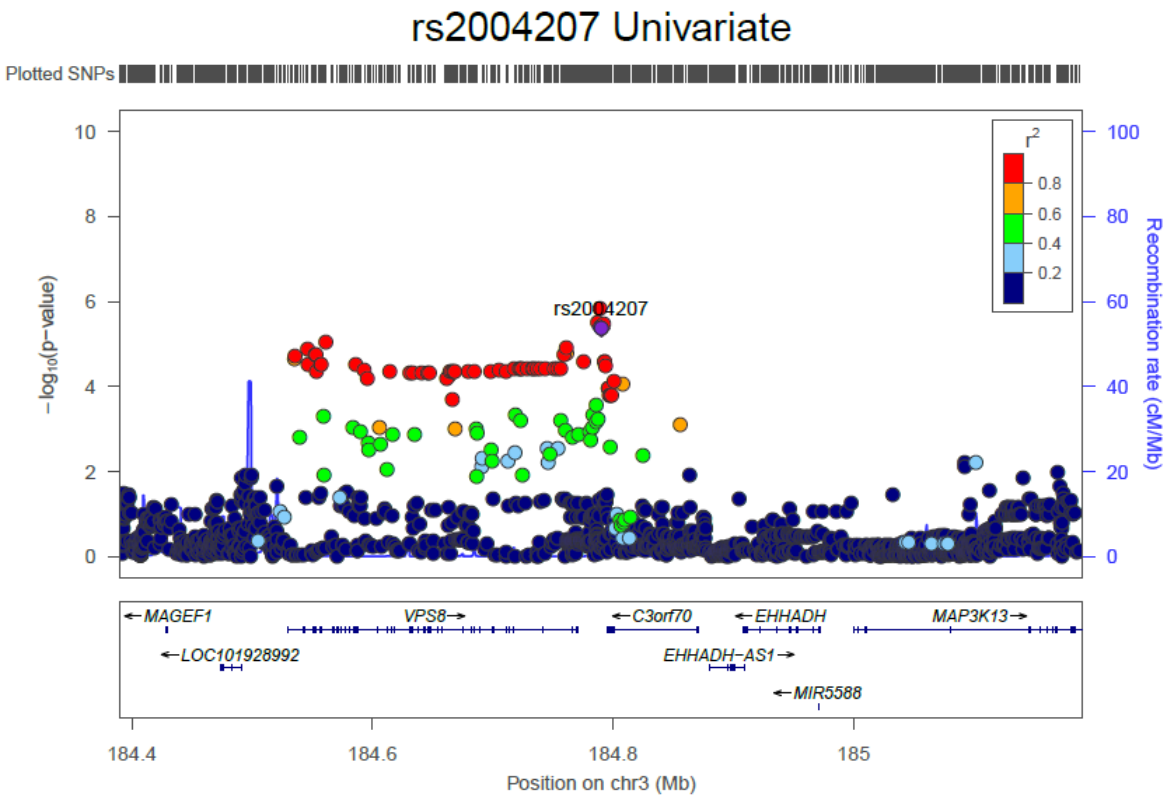

rs55882333 Univariate
Plotted SNPs
rs55882333
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
EIF2B3
ZSWIM5
LINC01144
CCDC163P
NASP
HECTD3
HPDL
MMACHC
CCDC17
UROD
MUTYH
PRDX1
GPBP1L1
TOE1
AKR1A1
TMEM69
TESK2
RPS15AP10
45.4
45.6
45.8
46
Position on chr1 (Mb)

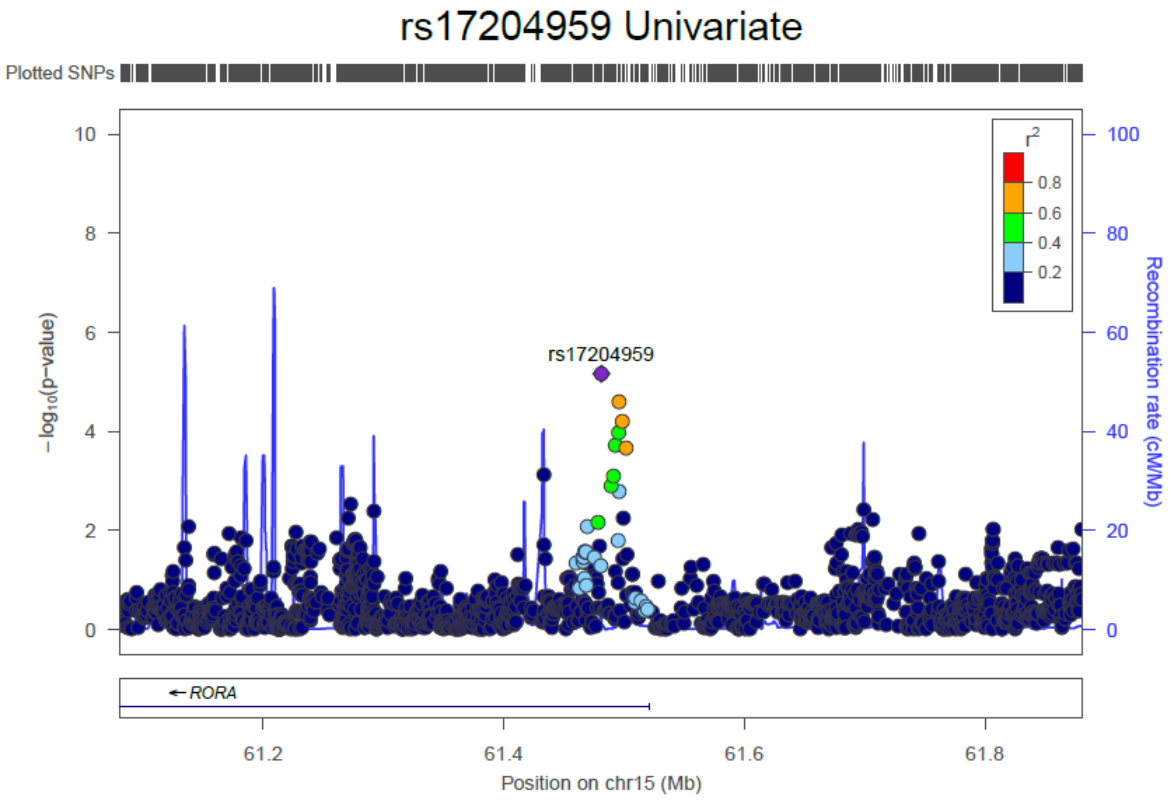

rs744306 Univariate
Plotted SNPs
rs744306
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MAGEF1
VPS8
C3orf70
EHHADH
MAP3K13
LOC101928992
EHHADH−AS1
MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

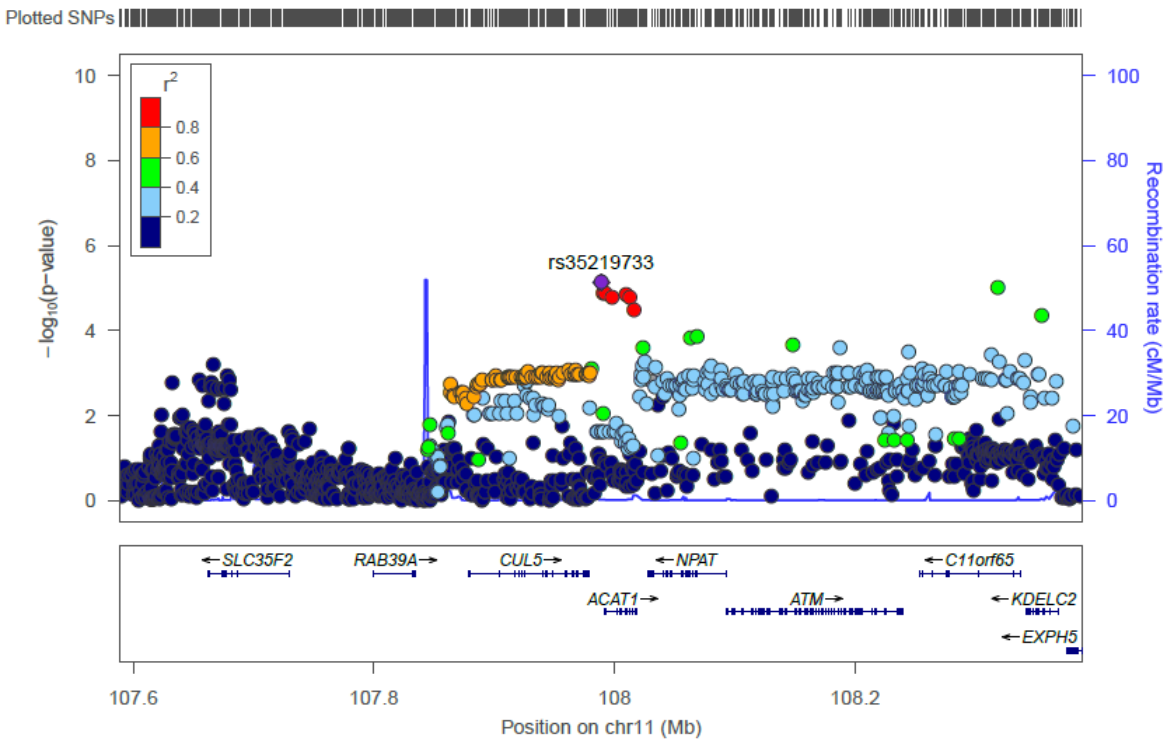

rs747507 Univariate
Plotted SNPs
rs747507
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MAGEF1
VPS8
C3orf70
EHHADH
MAP3K13
LOC101928992
EHHADH−AS1
MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

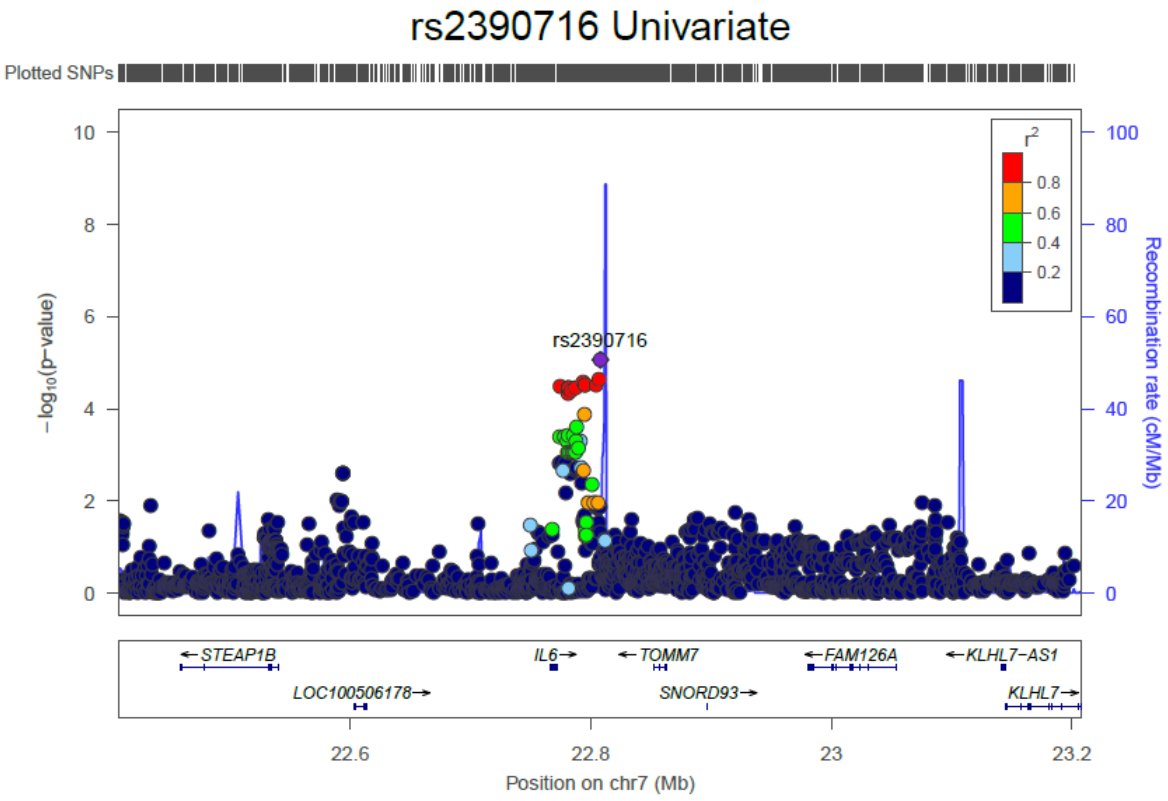

rs7607256 Univariate
Plotted SNPs
rs7607256
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
DIS3L2
ALPP
ALPI
CHRND
EFHD1
GIGYF2
NGEF
ECEL1P2
ECEL1
EIF4E2
KCNJ13
C2orf82
ALPPL2
PRSS56
CHRNG
TIGD1
MIR5001
233.2
233.4
233.6
233.8
Position on chr2 (Mb)

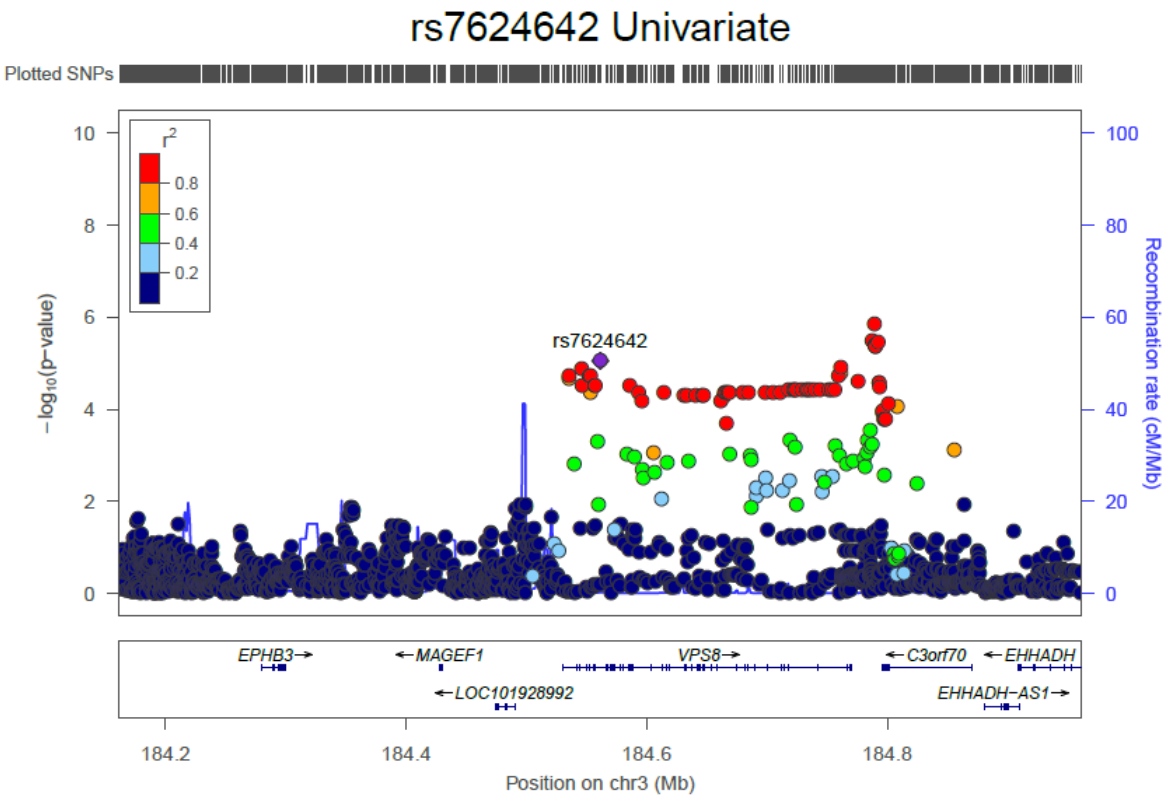

rs2004208 Univariate
Plotted SNPs
rs2004208
r²
0.8
0.6
0.4
0.2
−log₁₀(p−value)
Recombination rate (cM/Mb)
MAGEF1
VPS8
C3orf70
EHHADH
MAP3K13
LOC101928992
EHHADH−AS1
MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

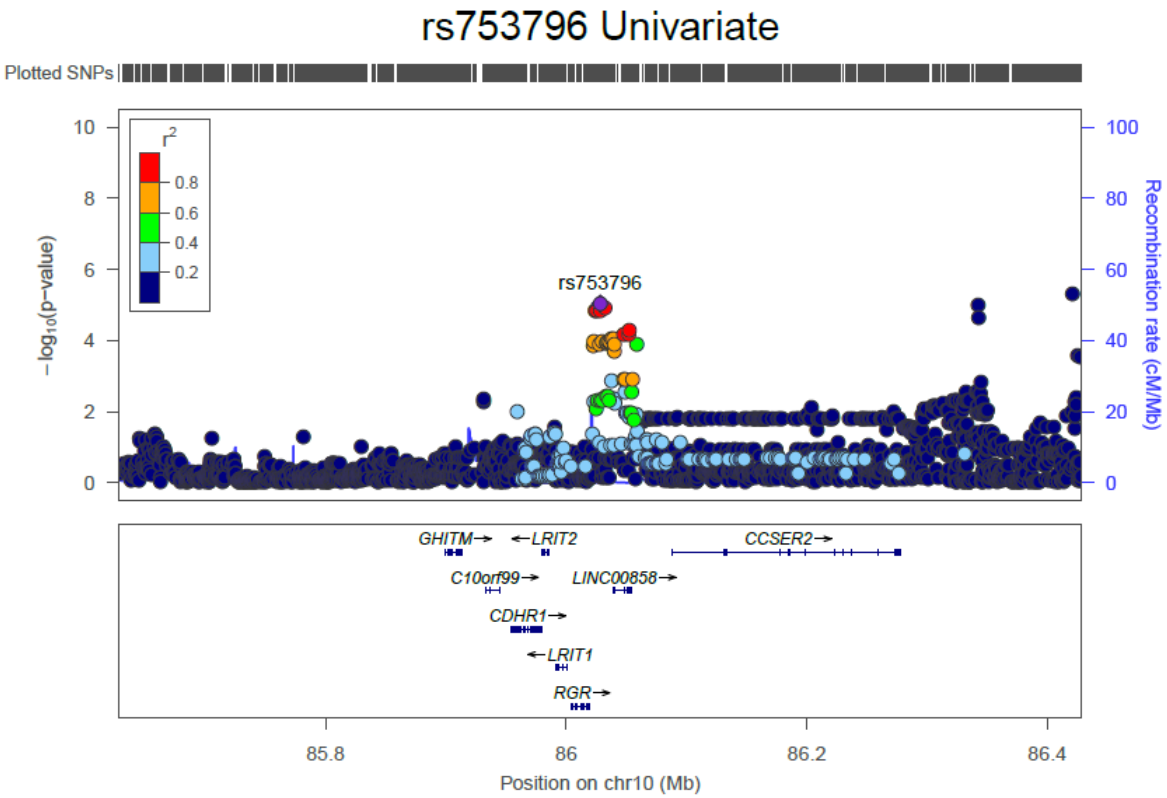

rs2004207 Univariate
Plotted SNPs
rs2004207
r²
0.8
0.6
0.4
0.2
−log₁₀(p−value)
Recombination rate (cM/Mb)
MAGEF1
VPS8
C3orf70
EHHADH
MAP3K13
LOC101928992
EHHADH−AS1
MIR5588
184.4
184.6
184.8
185
Position on chr3 (Mb)

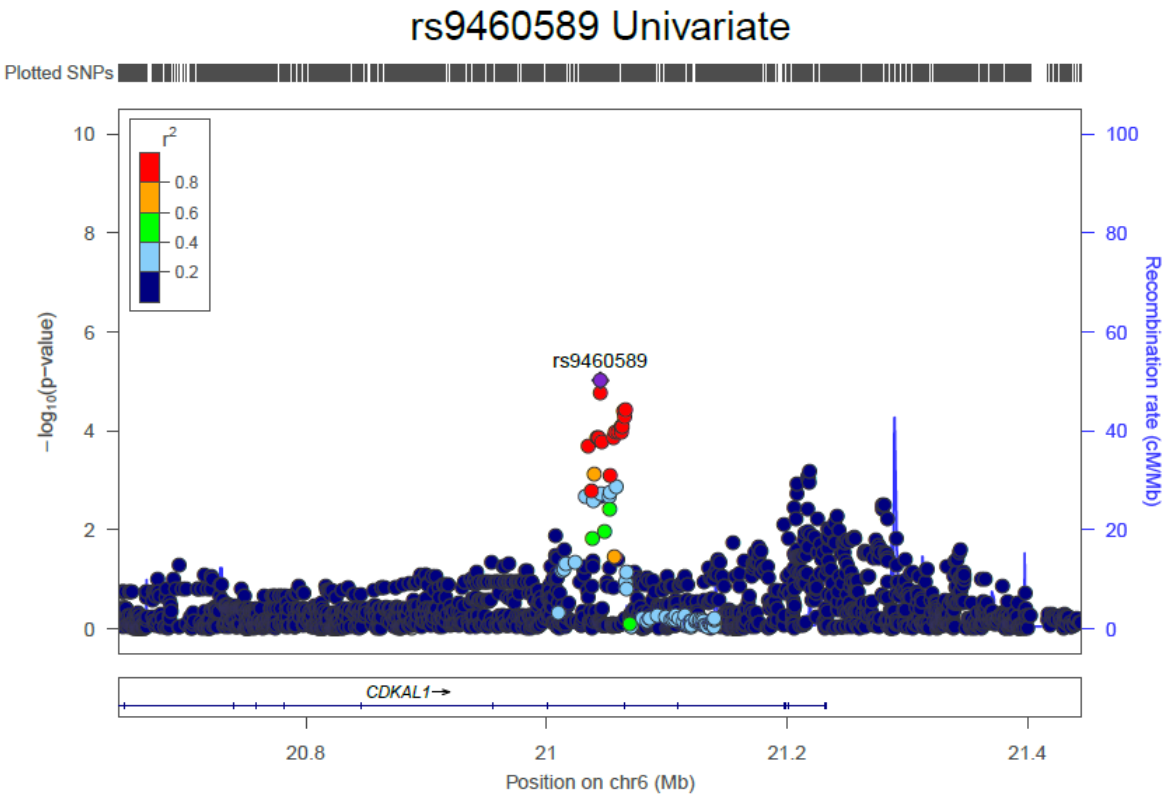

rs17204959 Univariate
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs17204959
−log10(p−value)
Recombination rate (cM/Mb)
RORA
61.2
61.4
61.6
61.8
Position on chr15 (Mb)

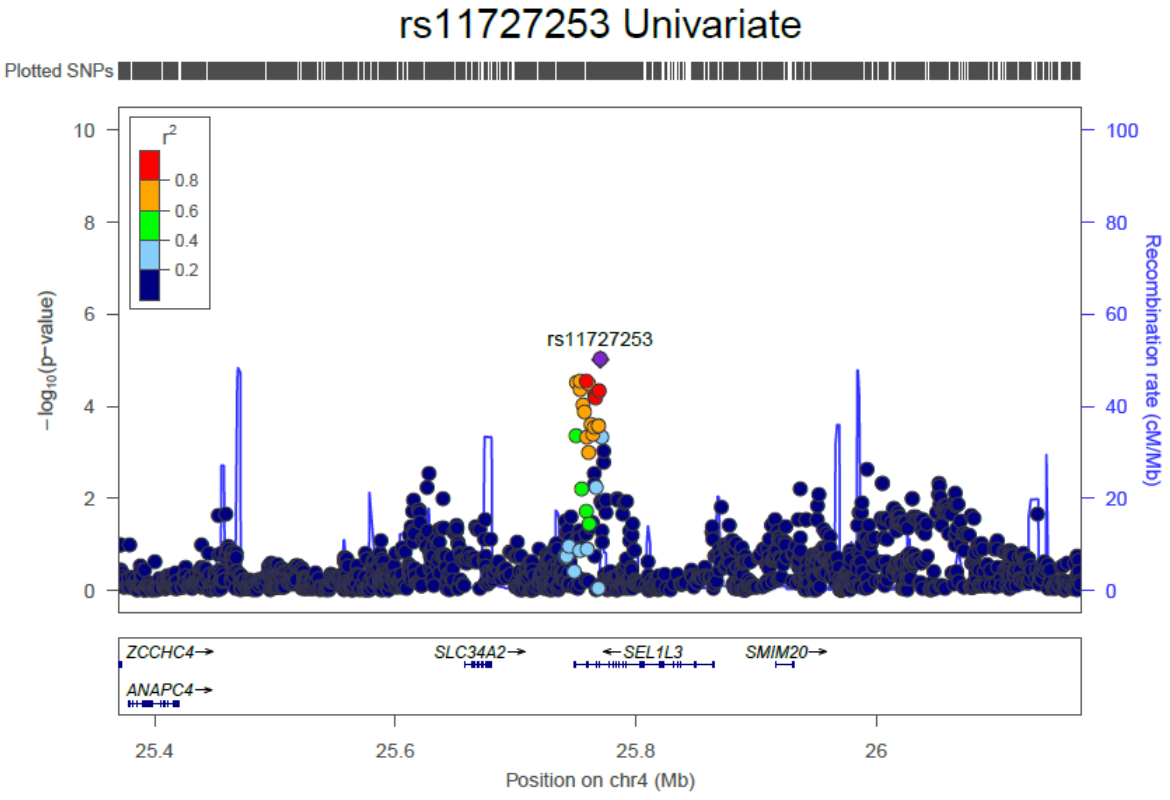

rs35219733 Univariate
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs35219733
−log10(p−value)
Recombination rate (cM/Mb)
SLC35F2
RAB39A
CUL5
NPAT
C11orf65
ACAT1
ATM
KDELC2
EXPH5
107.6
107.8
108
108.2
Position on chr11 (Mb)

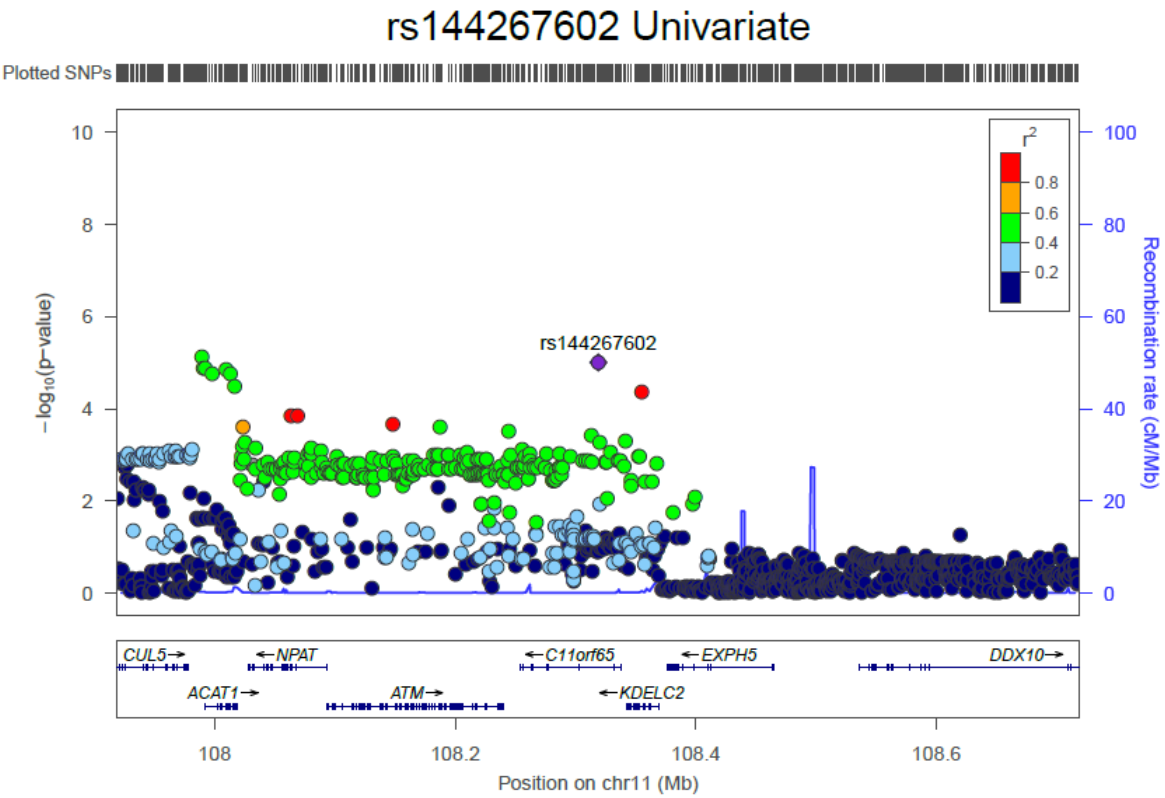

rs2390716 Univariate
Plotted SNPs
rs2390716
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
STEAP1B
IL6
TOMM7
FAM126A
KLHL7−AS1
LOC100506178
SNORD93
KLHL7
22.6
22.8
23
23.2
Position on chr7 (Mb)


rs7624642 Univariate
Plotted SNPs
rs7624642
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
EPHB3
MAGEF1
VPS8
C3orf70
EHHADH
LOC101928992
EHHADH−AS1
184.2
184.4
184.6
184.8
Position on chr3 (Mb)


rs753796 Univariate
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs753796
−log10(p−value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

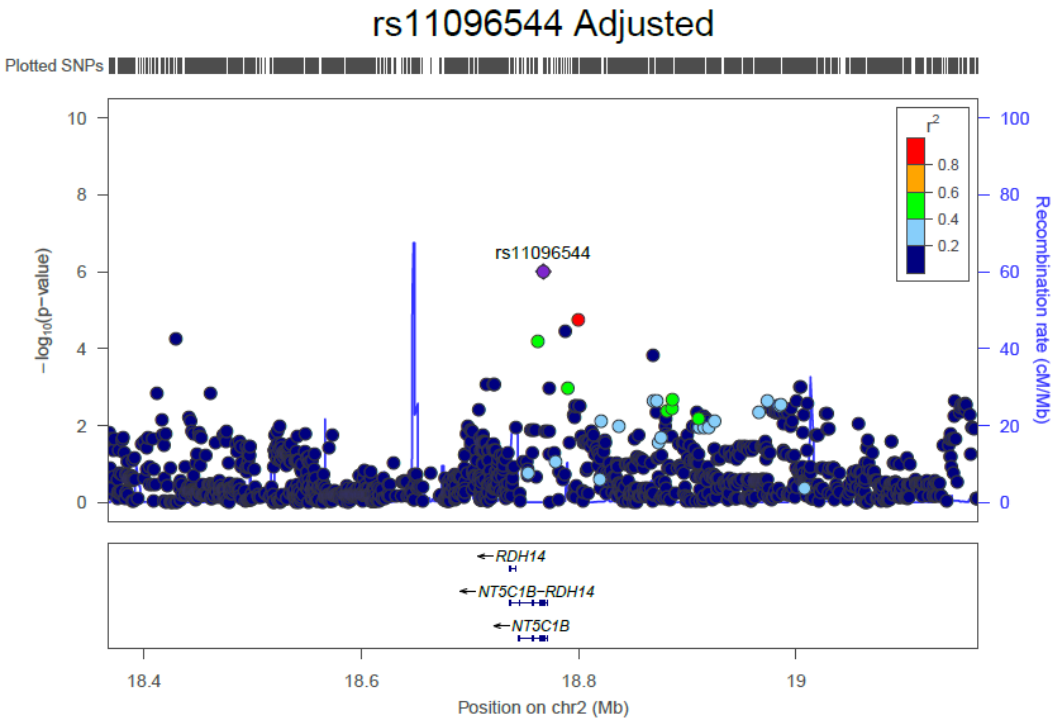

rs9460589 Univariate
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs9460589
−log10(p−value)
Recombination rate (cM/Mb)
CDKAL1
20.8
21
21.2
21.4
Position on chr6 (Mb)

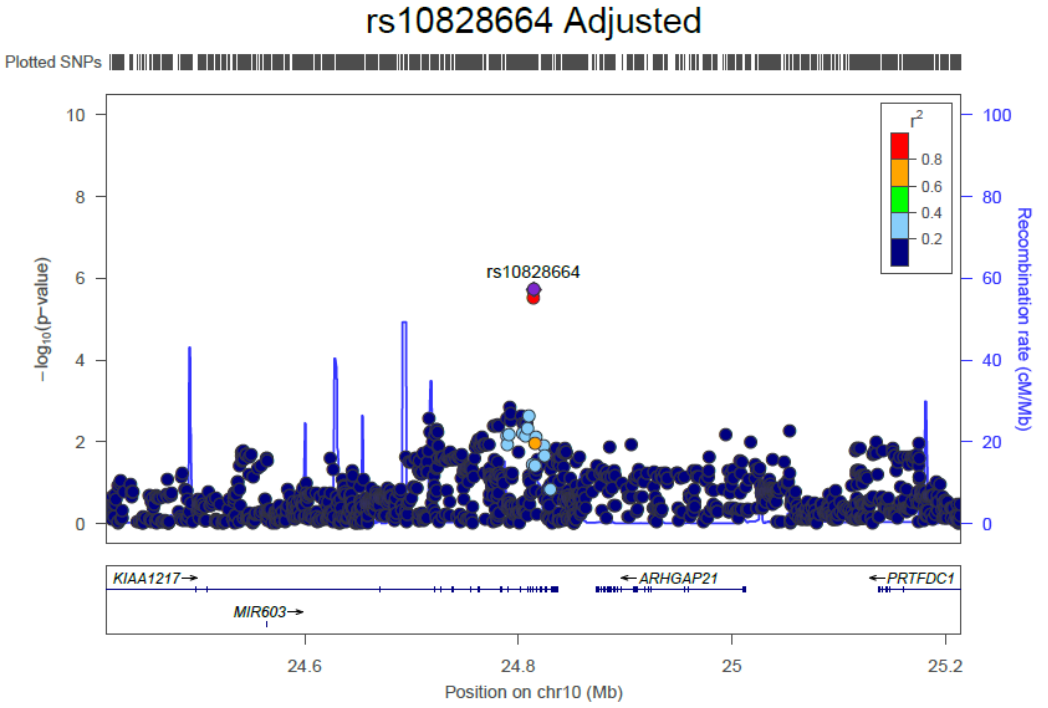

rs11727253 Univariate
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs11727253
−log10(p−value)
Recombination rate (cM/Mb)
ZCCHC4
SLC34A2
SEL1L3
SMIM20
ANAPC4
25.4
25.6
25.8
26
Position on chr4 (Mb)

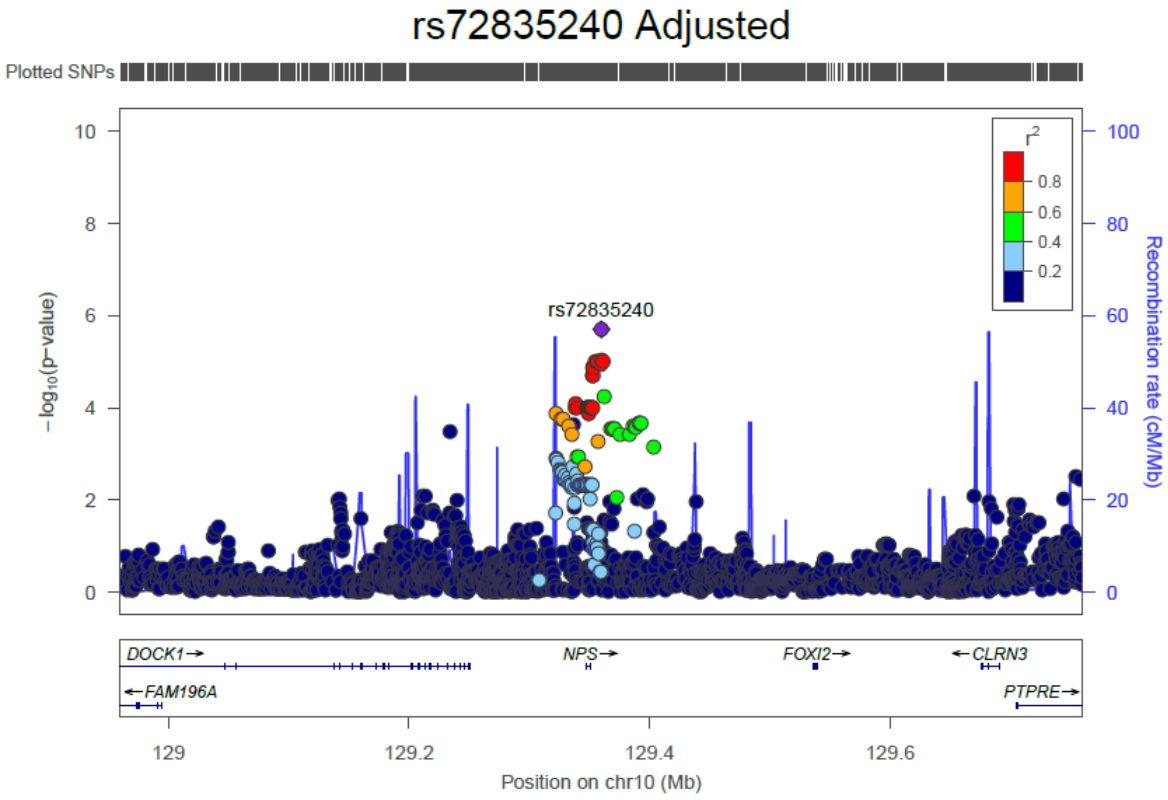

rs144267602 Univariate
Plotted SNPs
rs144267602
r2
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CUL5
NPAT
C11orf65
EXPH5
DDX10
ACAT1
ATM
KDELC2
Position on chr11 (Mb)

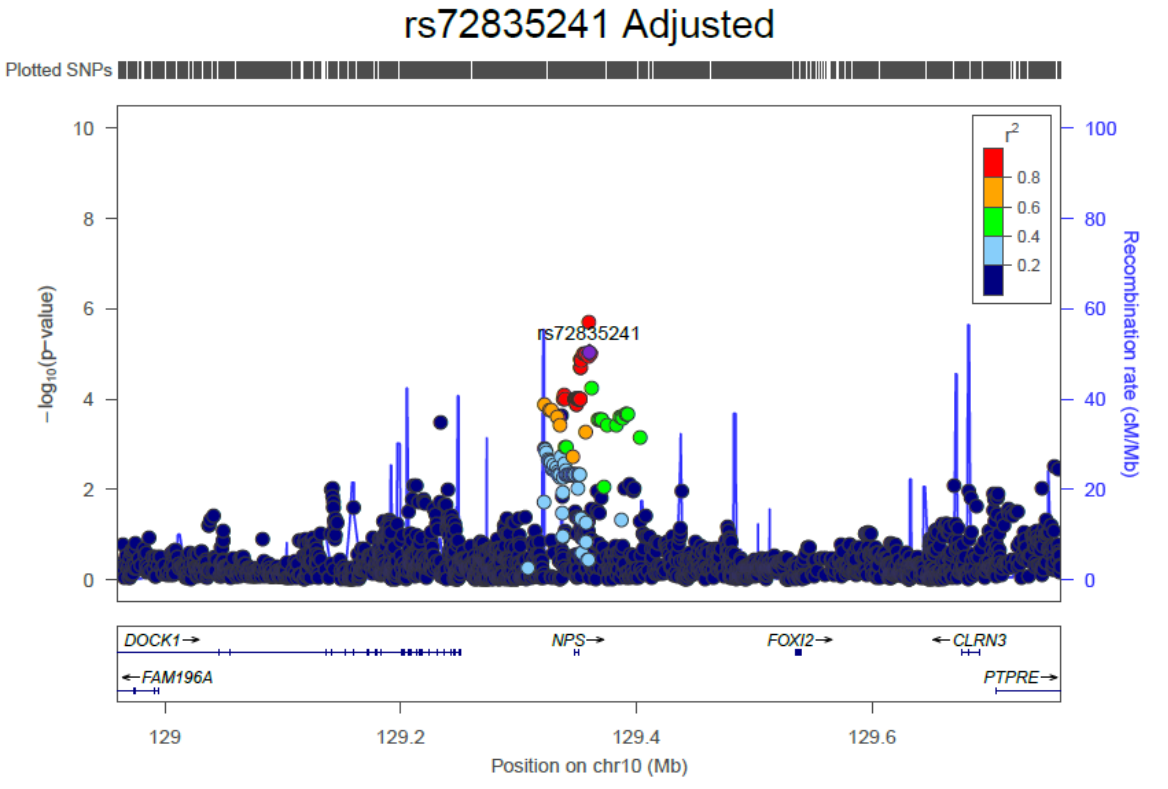
Appendix 6

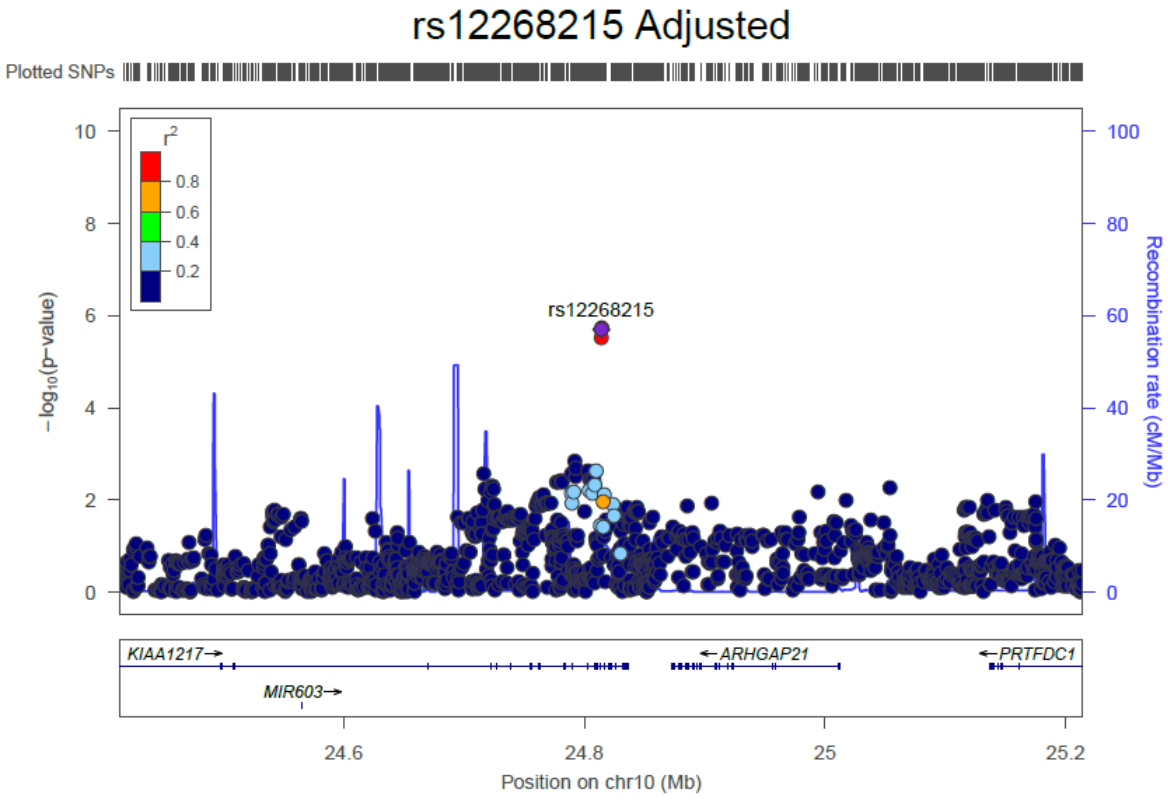
Full list of LocusZoom plots for all SNPs with p value less than 5x10-5 from GWAS for newly diagnosed focal epilepsy after application of first well tolerated anticonvulsant after adjustment for significant clinical factors.

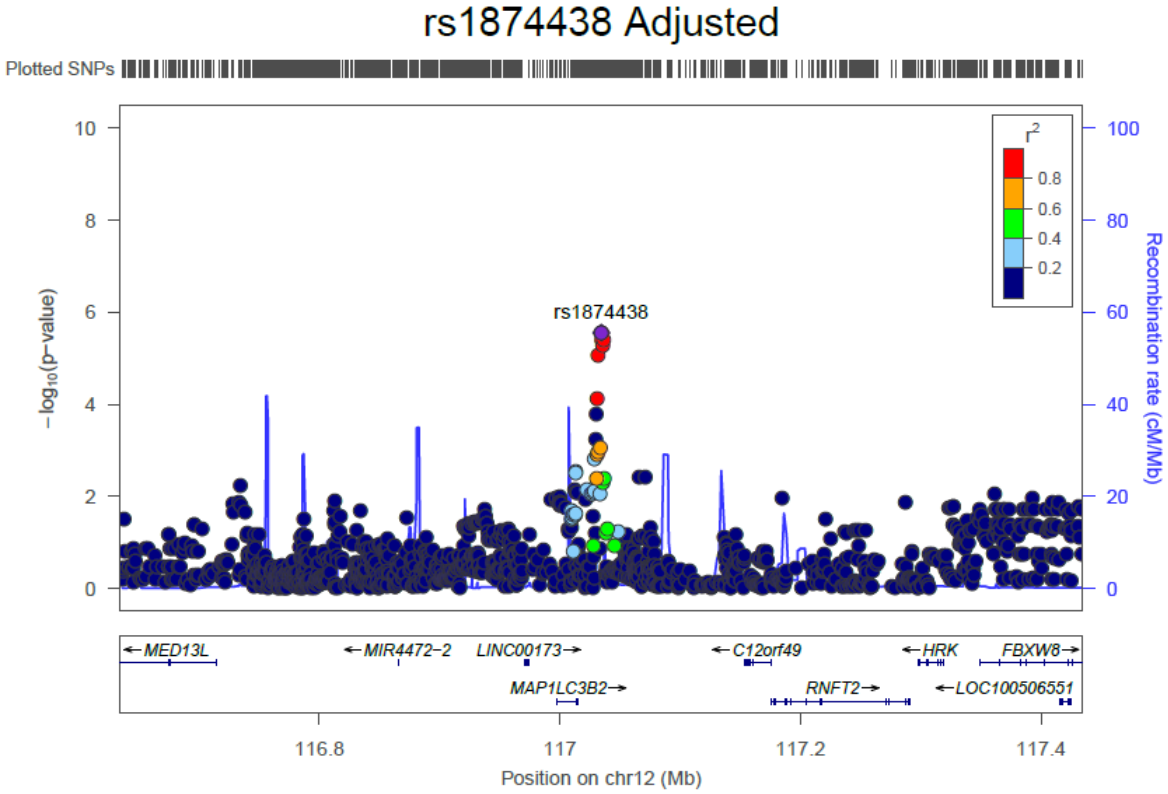


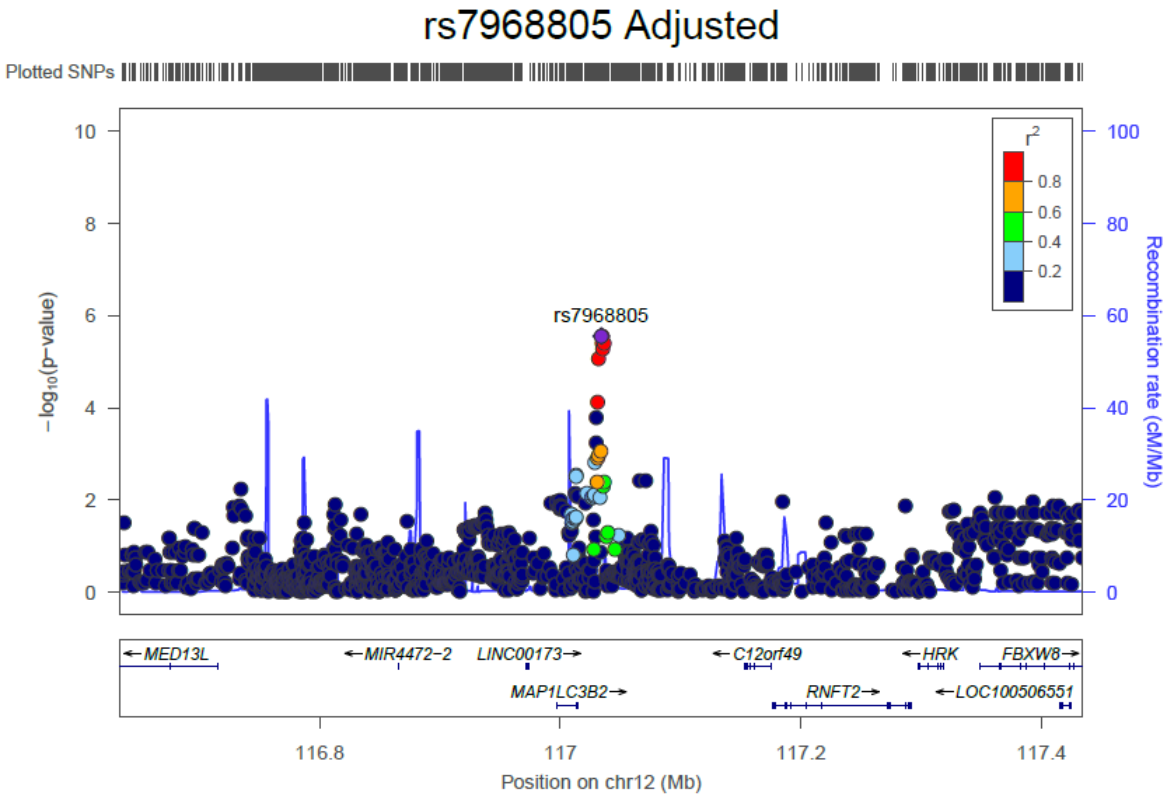


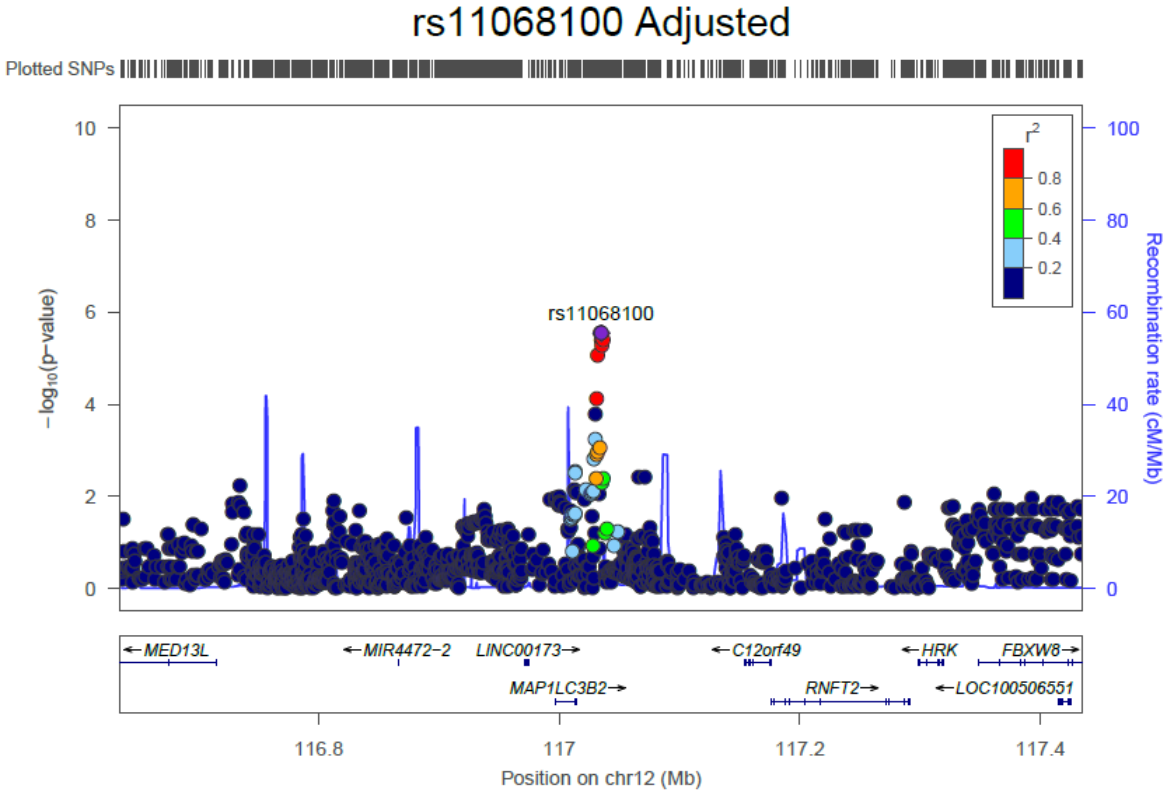

rs72835240 Adjusted
Plotted SNPs
rs72835240
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
DOCK1
NPS
FOXI2
CLRN3
FAM196A
PTPRE
129
129.2
129.4
129.6
Position on chr10 (Mb)

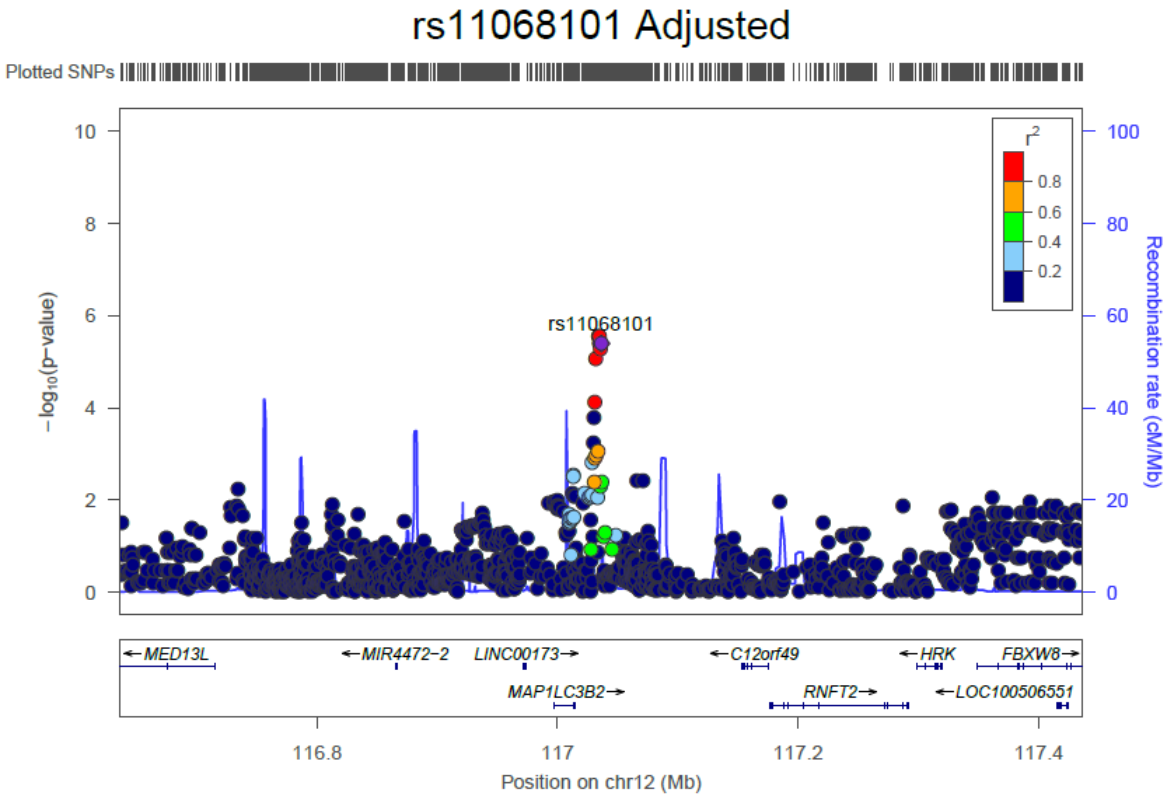

rs72835241 Adjusted
Plotted SNPs
rs72835241
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
DOCK1
NPS
FOXI2
CLRN3
FAM196A
PTPRE
129
129.2
129.4
129.6
Position on chr10 (Mb)

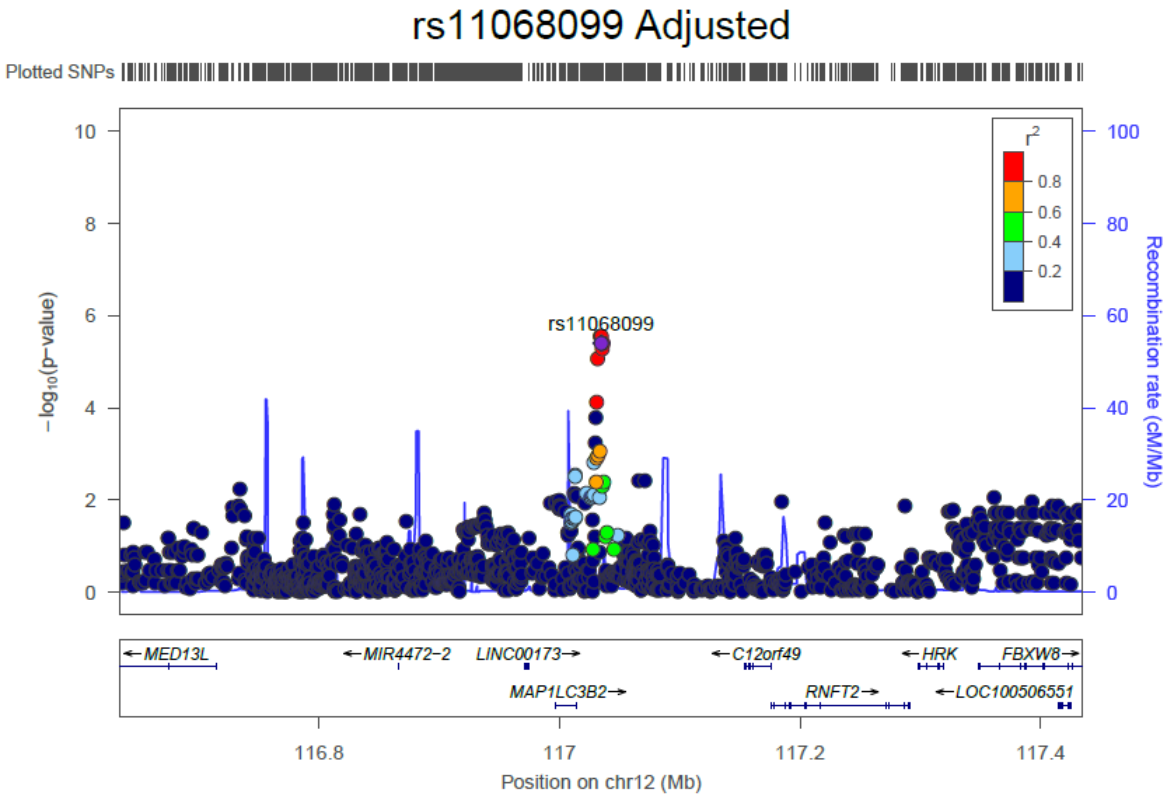

rs12268215 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs12268215
−log10(p−value)
Recombination rate (cM/Mb)
KIAA1217
ARHGAP21
PRTFDC1
MIR603
24.6
24.8
25
25.2
Position on chr10 (Mb)

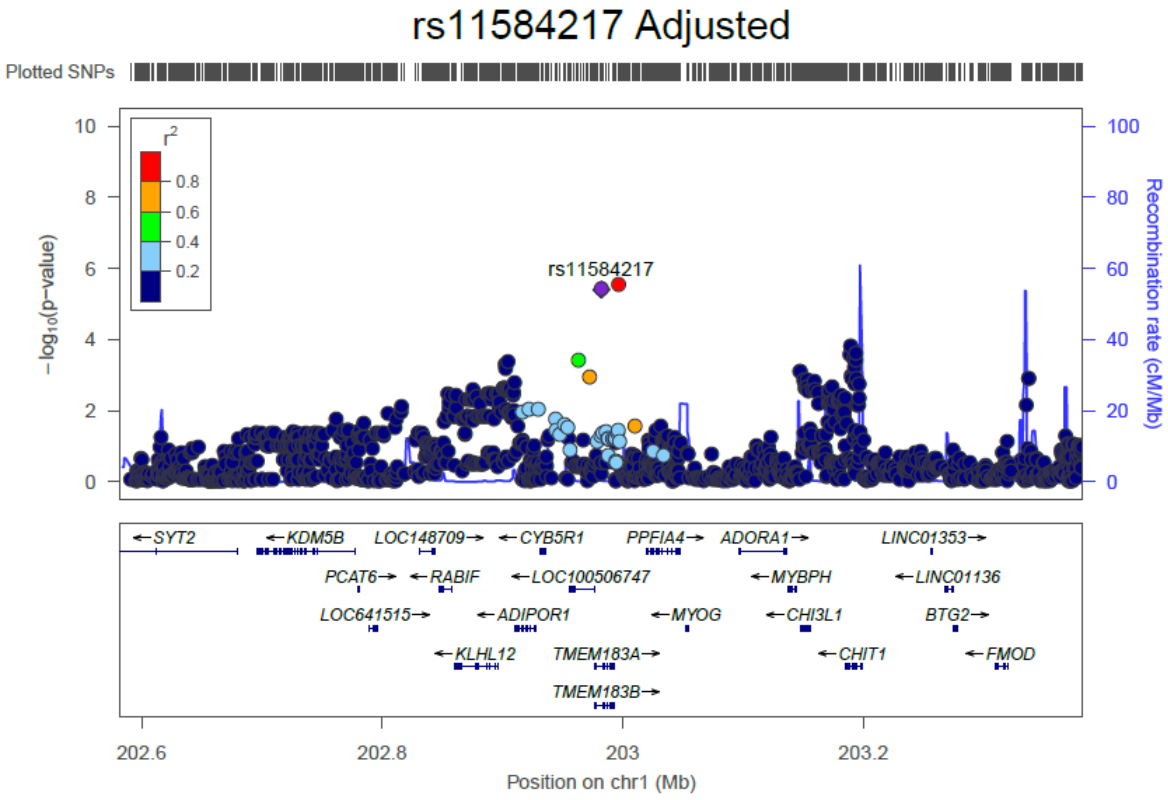

rs1874438 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs1874438
−log10(p−value)
Recombination rate (cM/Mb)
MED13L
MIR4472−2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

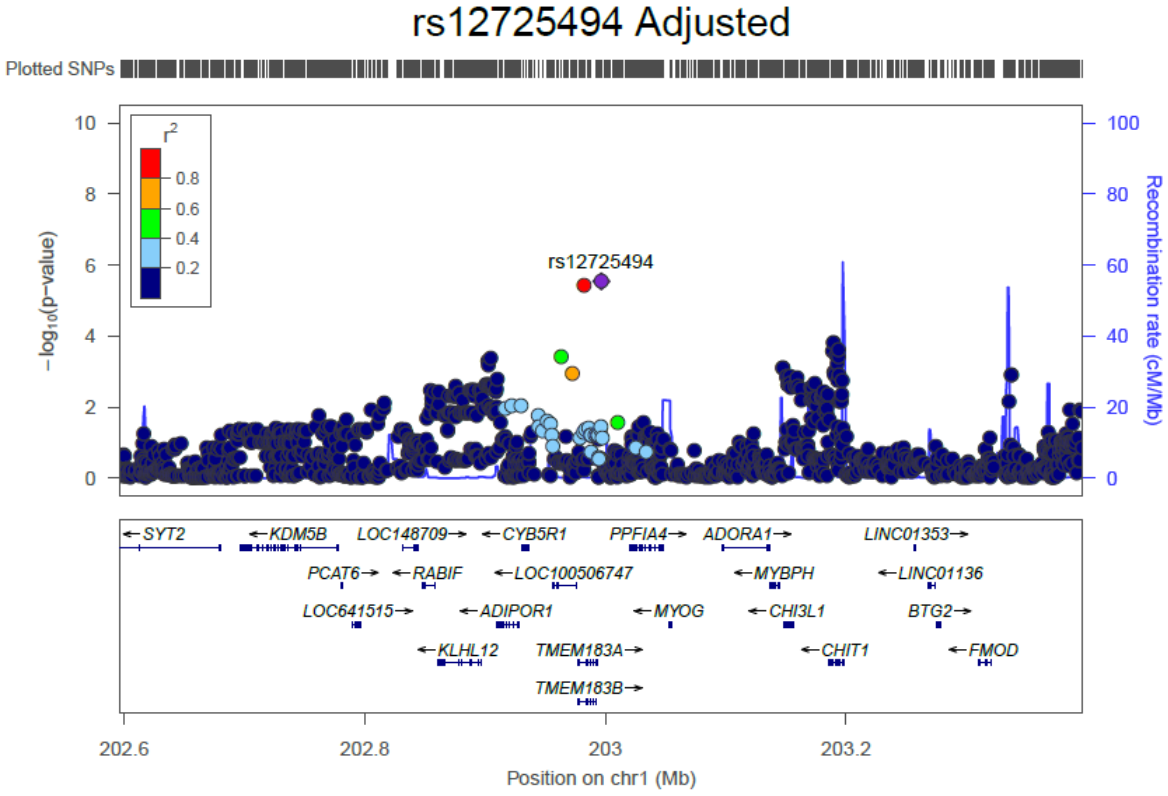

rs7968805 Adjusted
Plotted SNPs
rs7968805
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
MED13L
MIR4472-2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

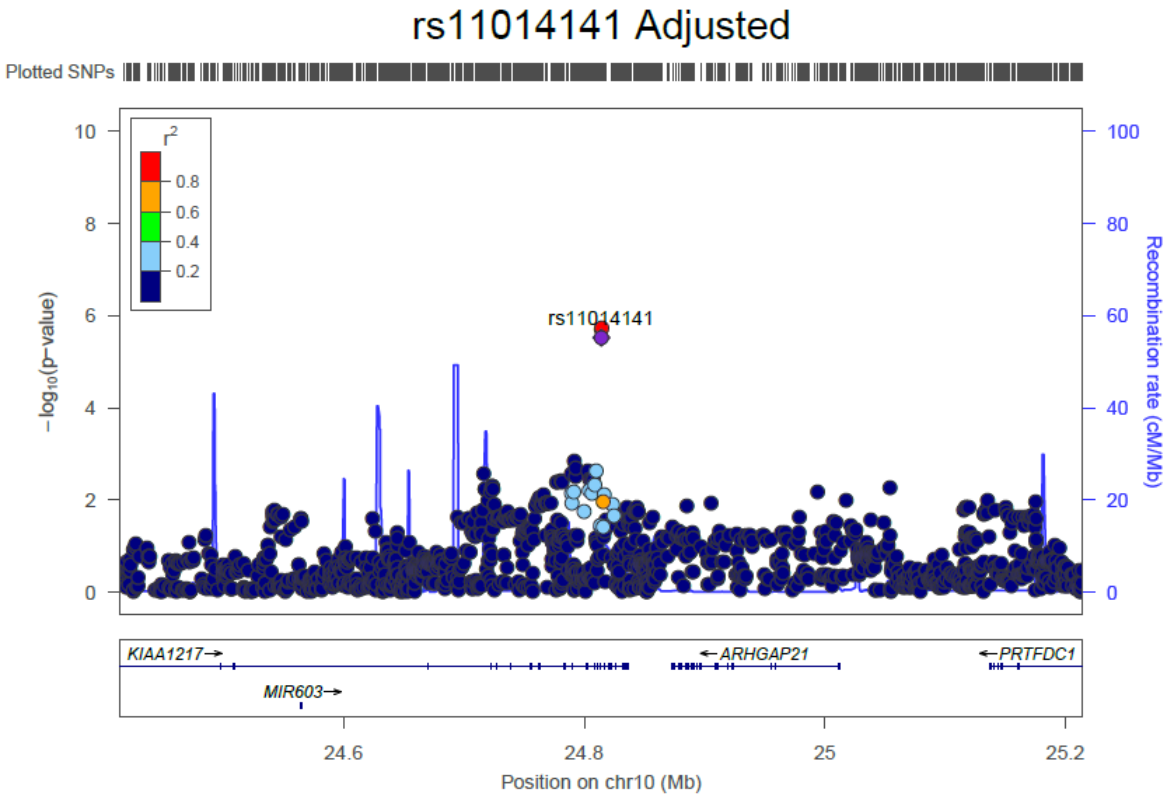

rs11068100 Adjusted
Plotted SNPs
rs11068100
r^2
0.8
0.6
0.4
0.2
$-\log_{10}$(p-value)
Recombination rate (cM/Mb)
MED13L
MIR4472-2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

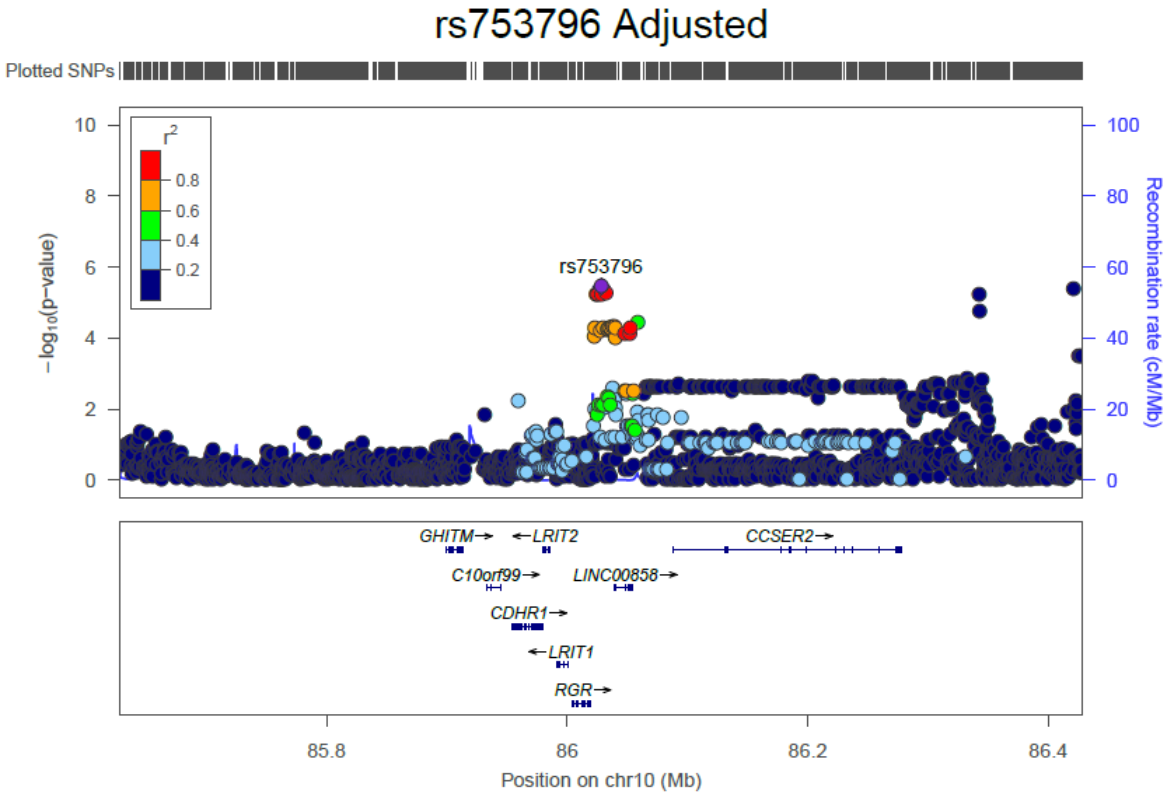

rs11068101 Adjusted
Plotted SNPs
rs11068101
−log10(p-value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MED13L
MIR4472-2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

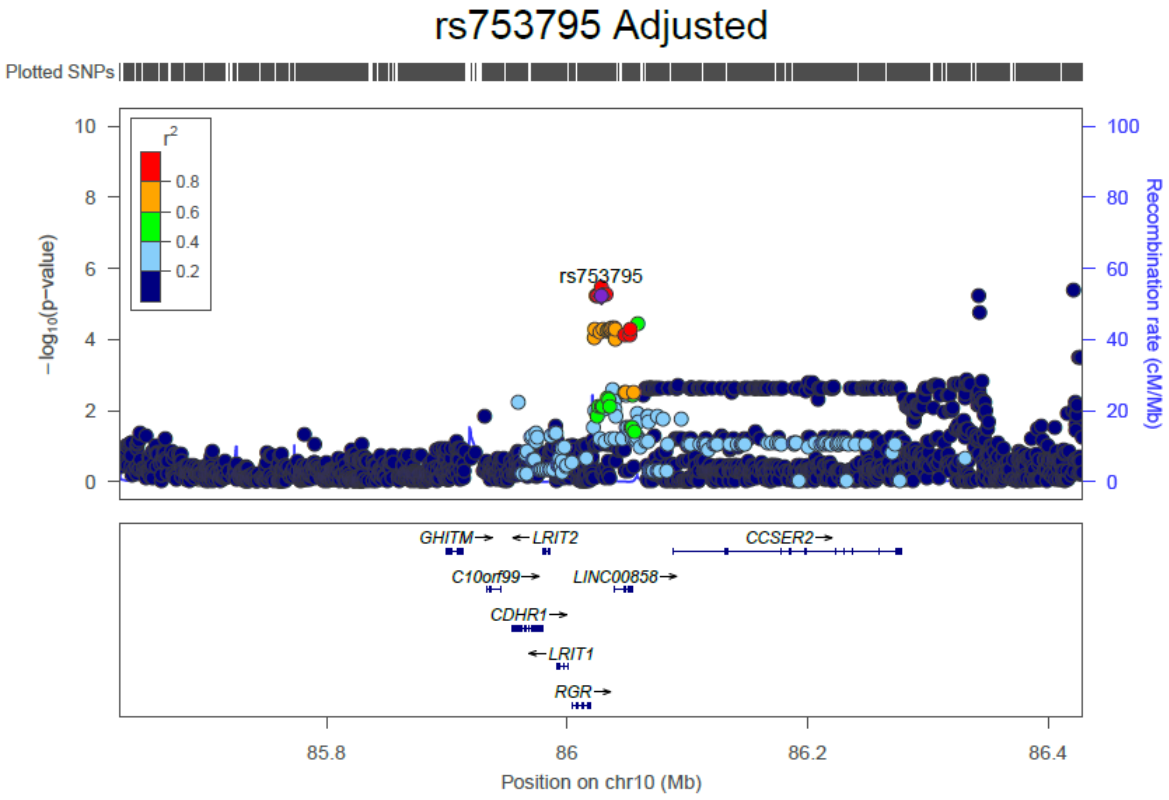

rs11068099 Adjusted
Plotted SNPs
rs11068099
−log10(p-value)
Recombination rate (cM/Mb)
r2
0.8
0.6
0.4
0.2
MED13L
MIR4472-2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

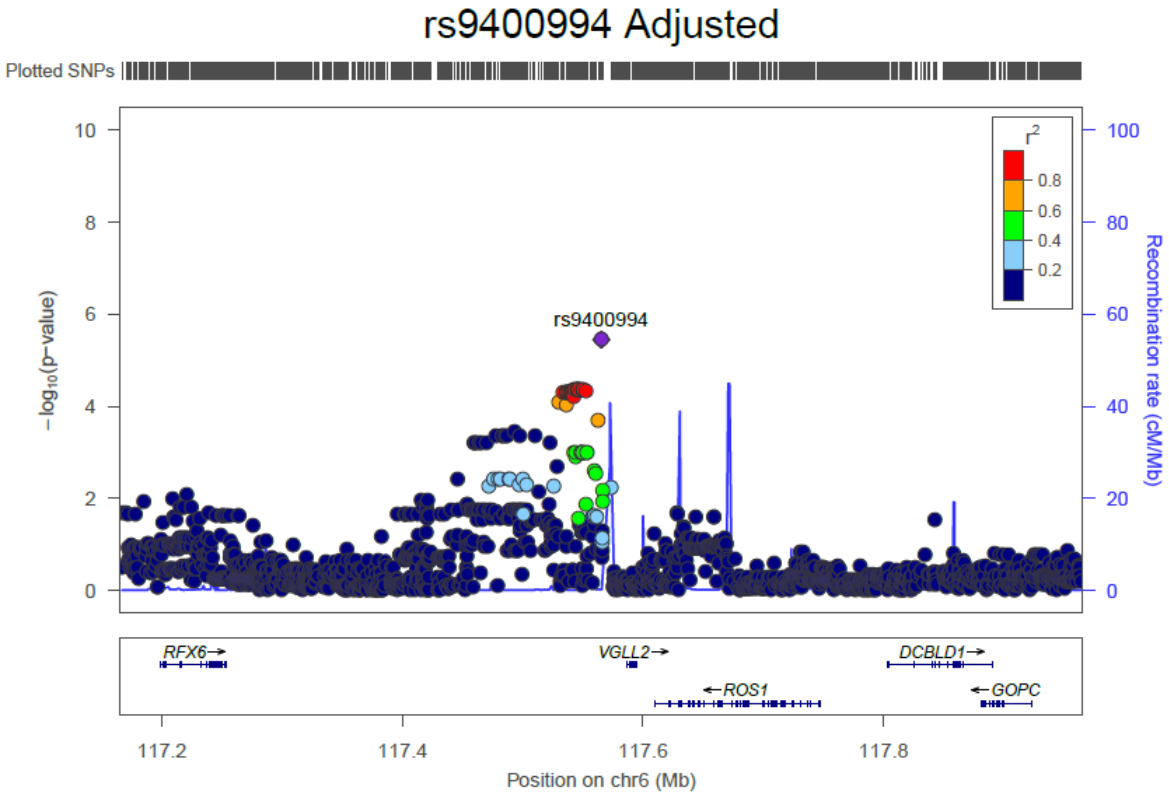

rs11584217 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs11584217
−log₁₀(p−value)
Recombination rate (cM/Mb)
SYT2
KDM5B
LOC148709
CYB5R1
PPFIA4
ADORA1
LINC01353
PCAT6
RABIF
LOC100506747
MYBPH
LINC01136
LOC641515
ADIPOR1
MYOG
CHI3L1
BTG2
KLHL12
TMEM183A
CHIT1
FMOD
TMEM183B
202.6
202.8
203
203.2
Position on chr1 (Mb)

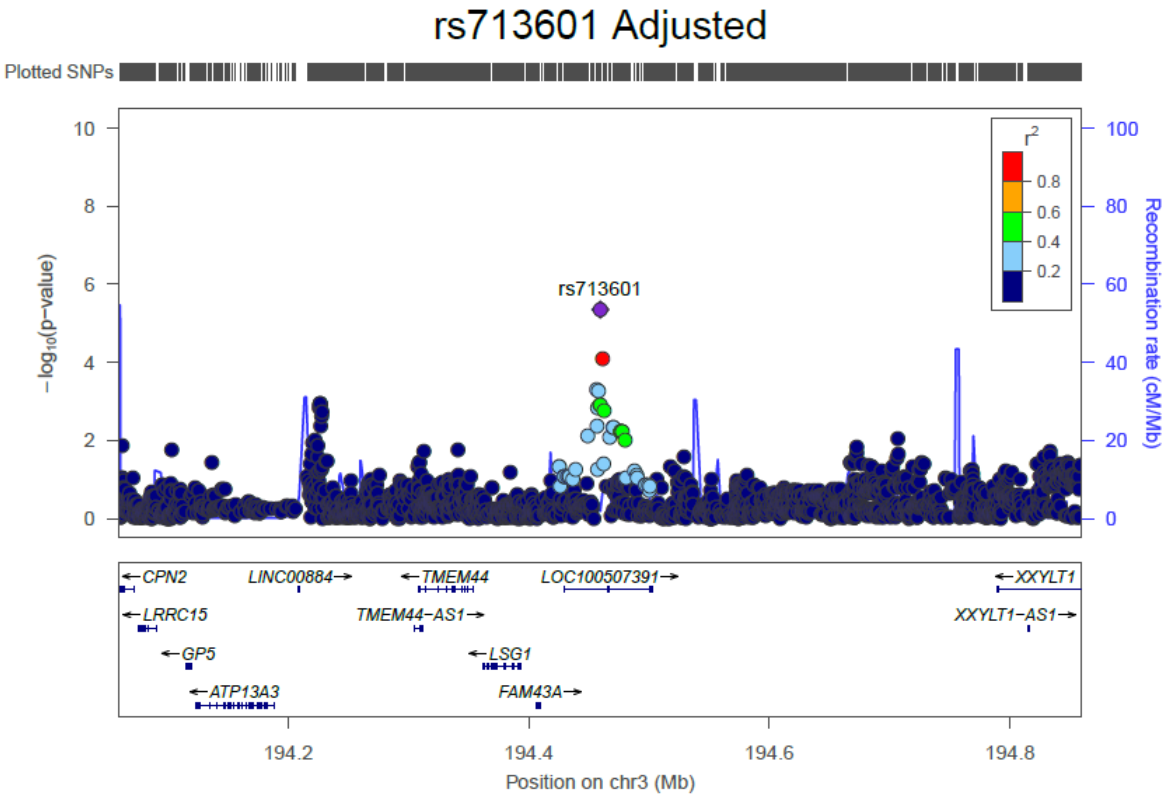

rs12725494 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs12725494
−log₁₀(p−value)
Recombination rate (cM/Mb)
SYT2
KDM5B
LOC148709
CYB5R1
PPFIA4
ADORA1
LINC01353
PCAT6
RABIF
LOC100506747
MYBPH
LINC01136
LOC641515
ADIPOR1
MYOG
CHI3L1
BTG2
KLHL12
TMEM183A
CHIT1
FMOD
TMEM183B
202.6
202.8
203
203.2
Position on chr1 (Mb)

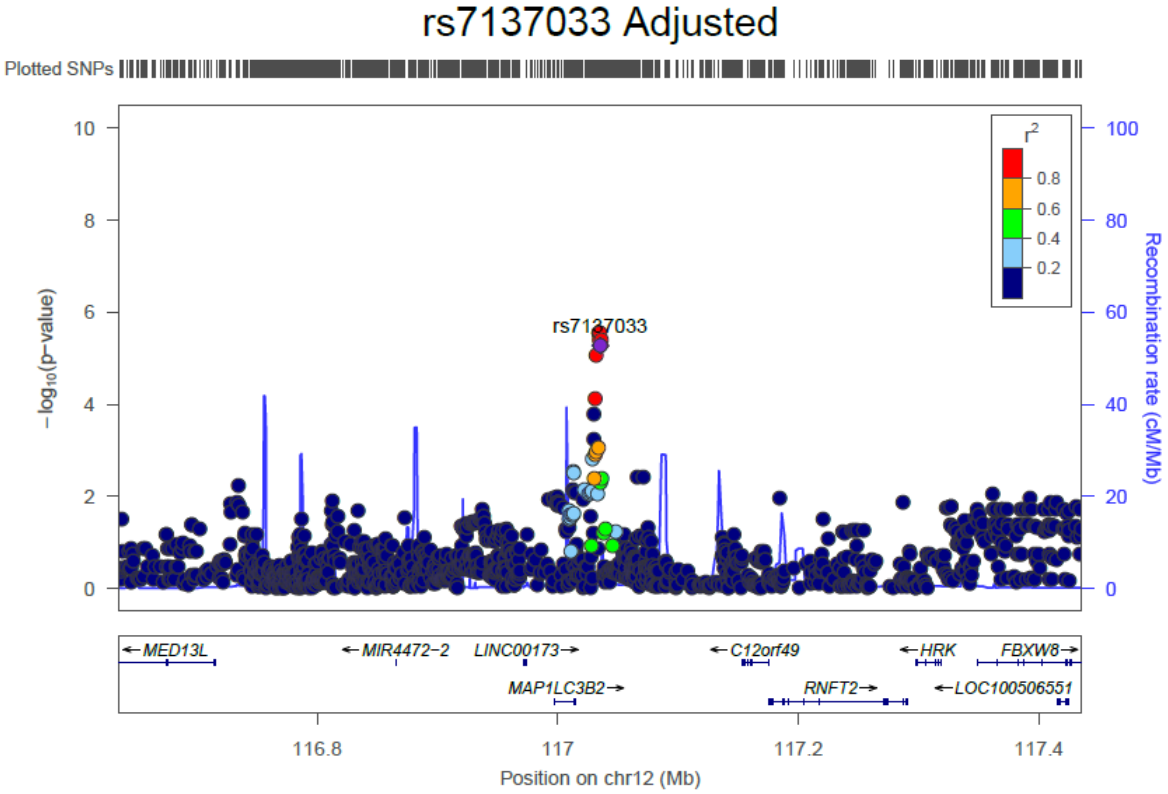

rs11014141 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs11014141
−log10(p−value)
Recombination rate (cM/Mb)
KIAA1217
ARHGAP21
PRTFDC1
MIR603
24.6
24.8
25
25.2
Position on chr10 (Mb)

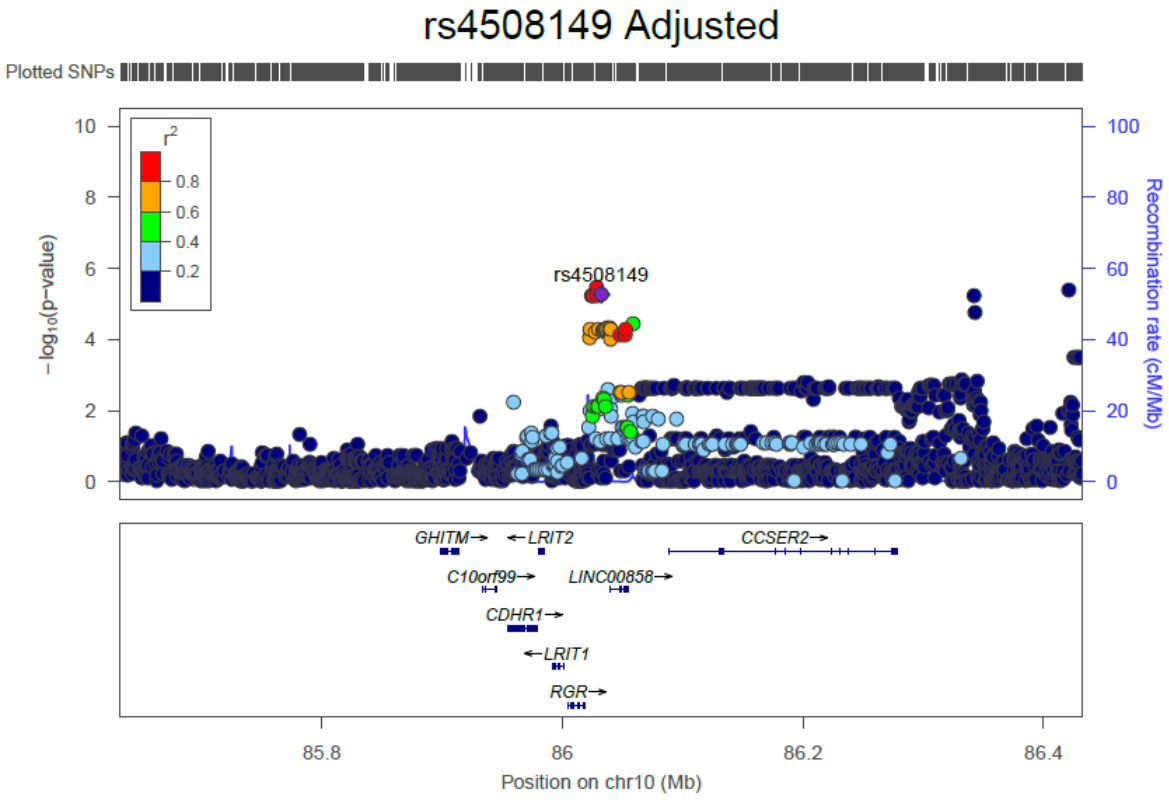

rs753796 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs753796
−log10(p−value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

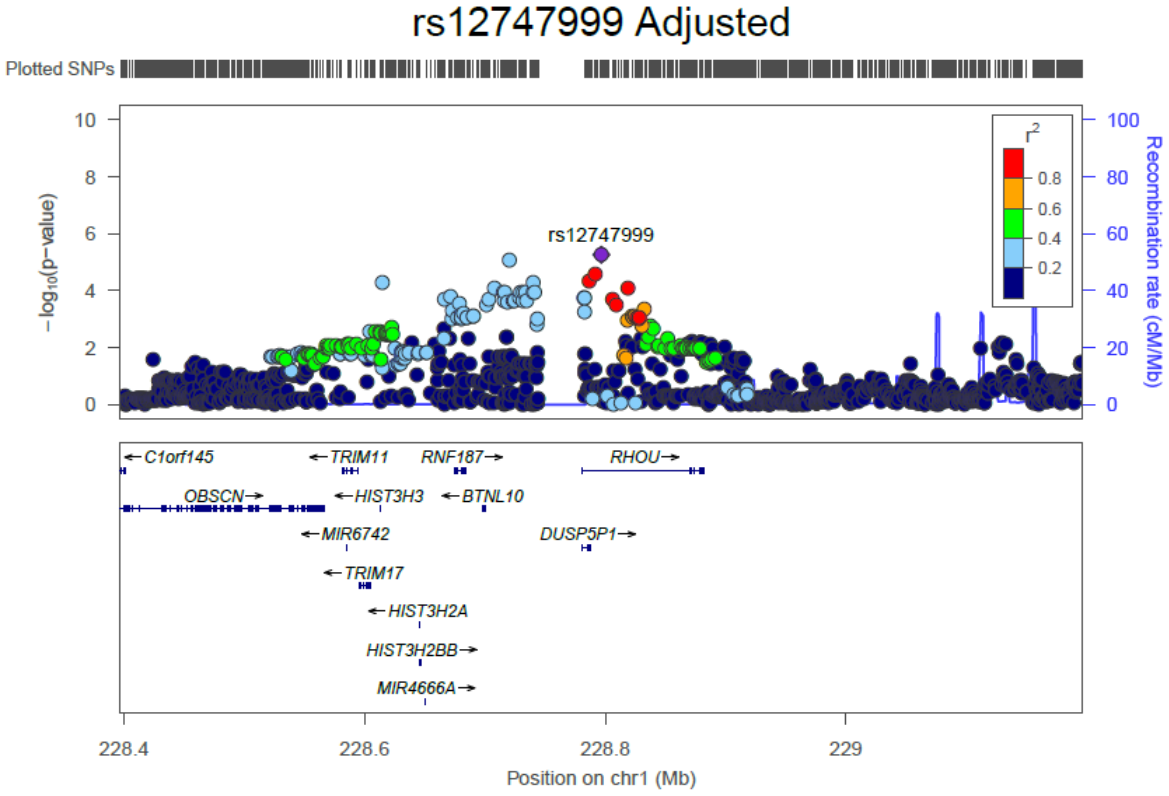

rs753795 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs753795
−log10(p−value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

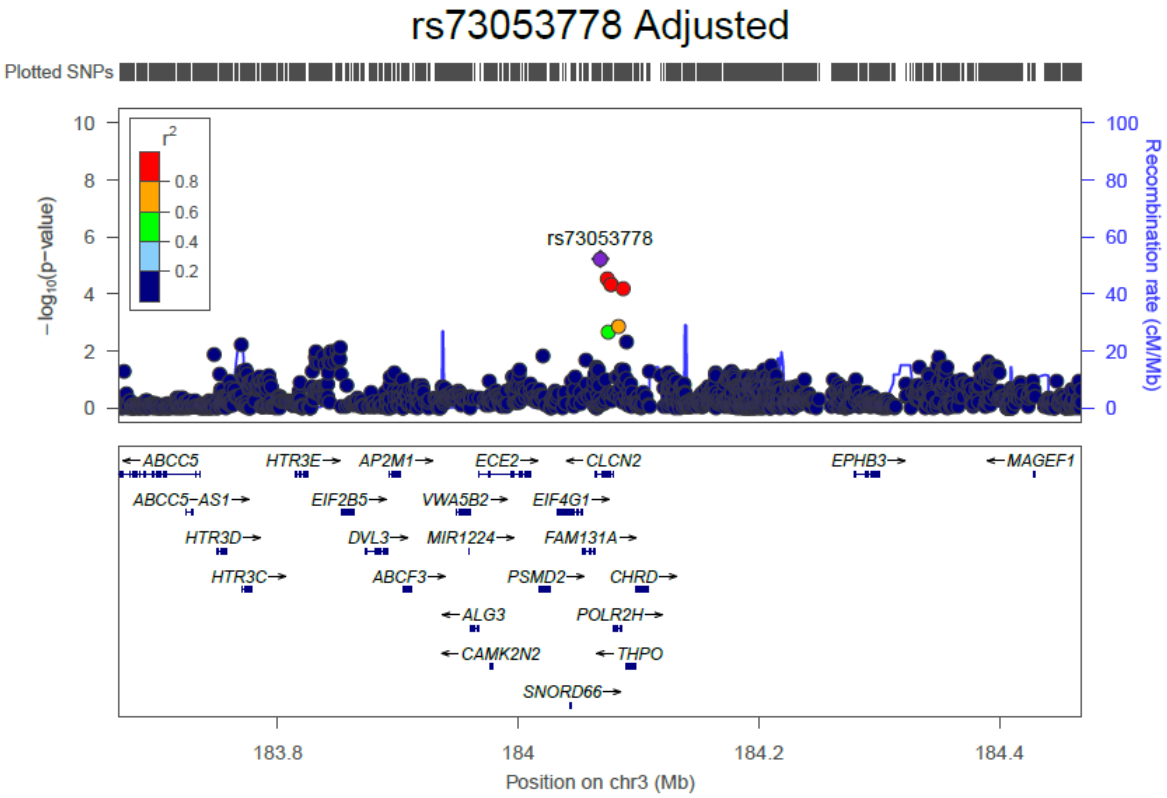

rs9400994 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs9400994
−log10(p−value)
Recombination rate (cM/Mb)
RFX6
VGLL2
DCBLD1
ROS1
GOPC
117.2
117.4
117.6
117.8
Position on chr6 (Mb)

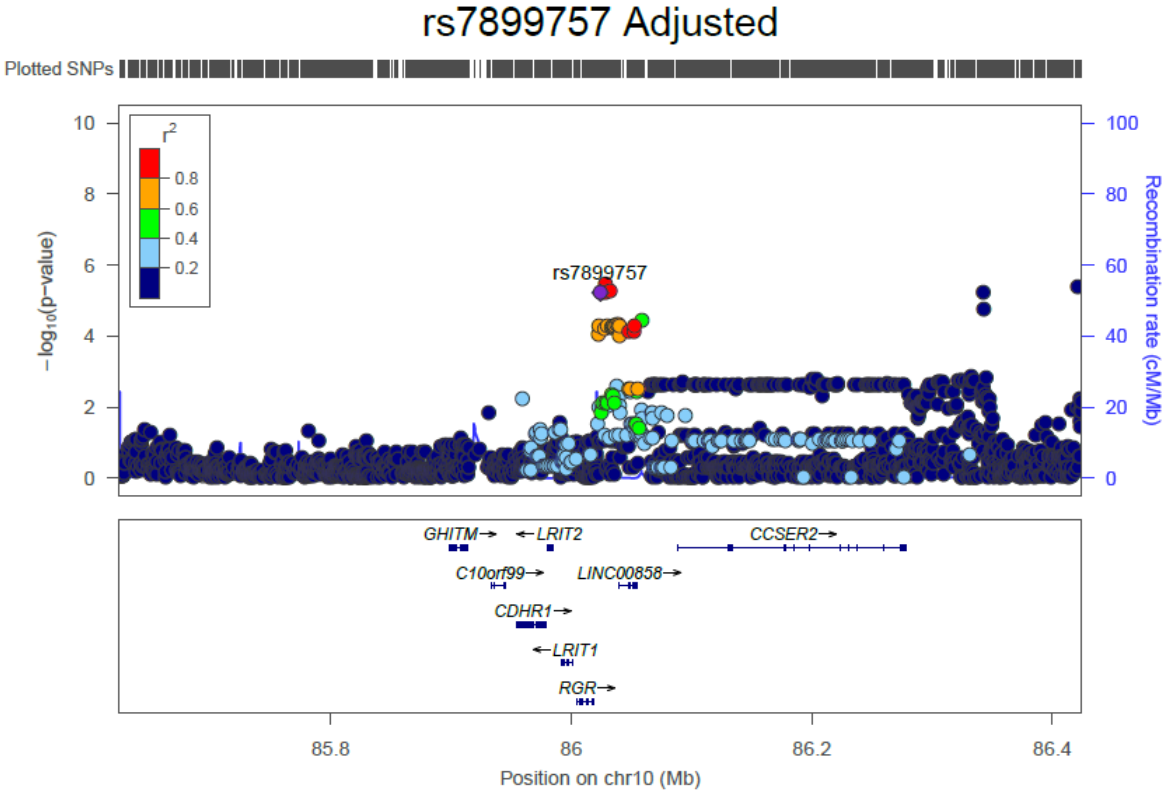

rs713601 Adjusted
Plotted SNPs
rs713601
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
CPN2
LINC00884
TMEM44
LOC100507391
XXYLT1
LRRC15
TMEM44−AS1
XXYLT1−AS1
GP5
LSG1
ATP13A3
FAM43A
194.2
194.4
194.6
194.8
Position on chr3 (Mb)

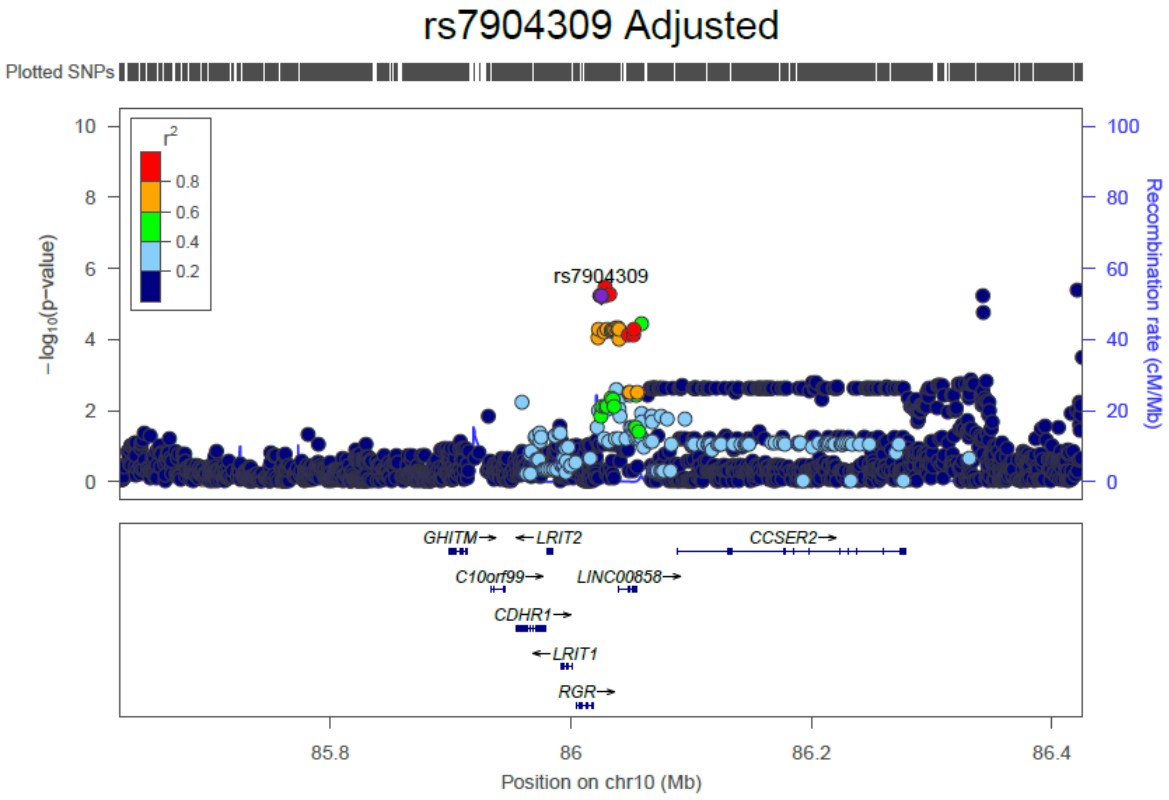

rs7137033 Adjusted
Plotted SNPs
rs7137033
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
MED13L
MIR4472−2
LINC00173
C12orf49
HRK
FBXW8
MAP1LC3B2
RNFT2
LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)

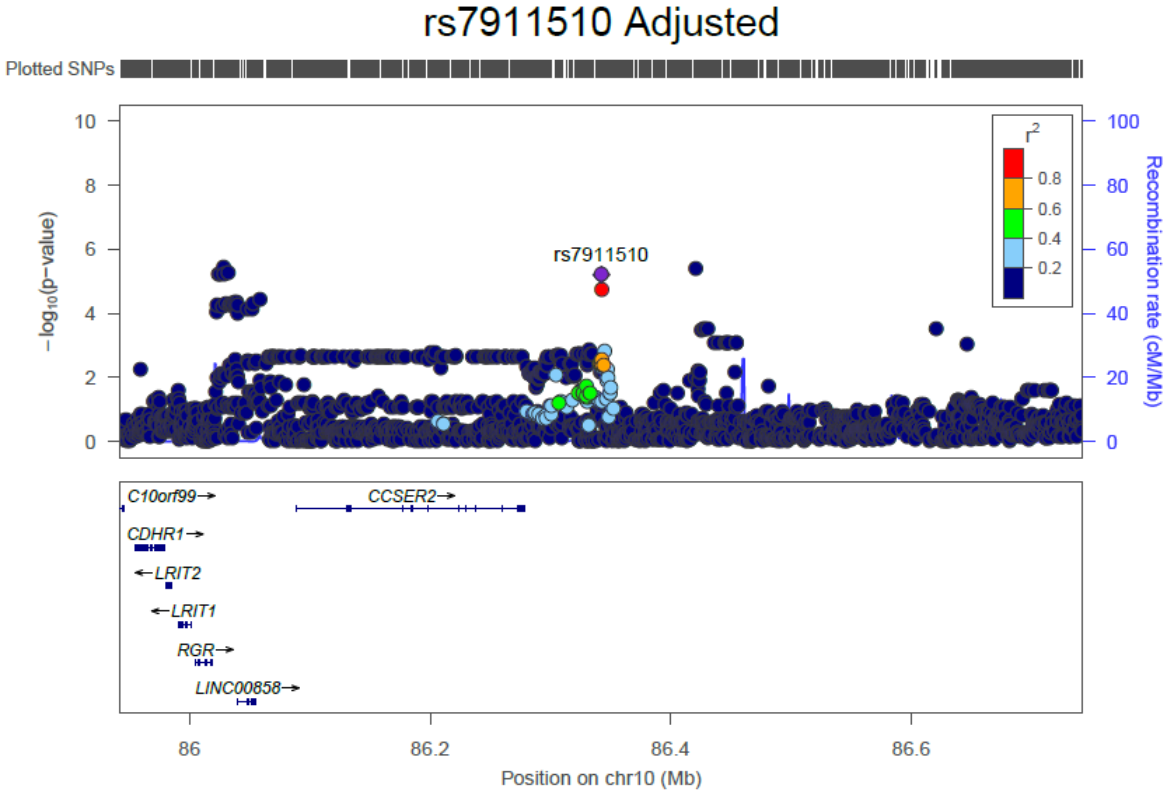

rs4508149 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs4508149
−log10(p-value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

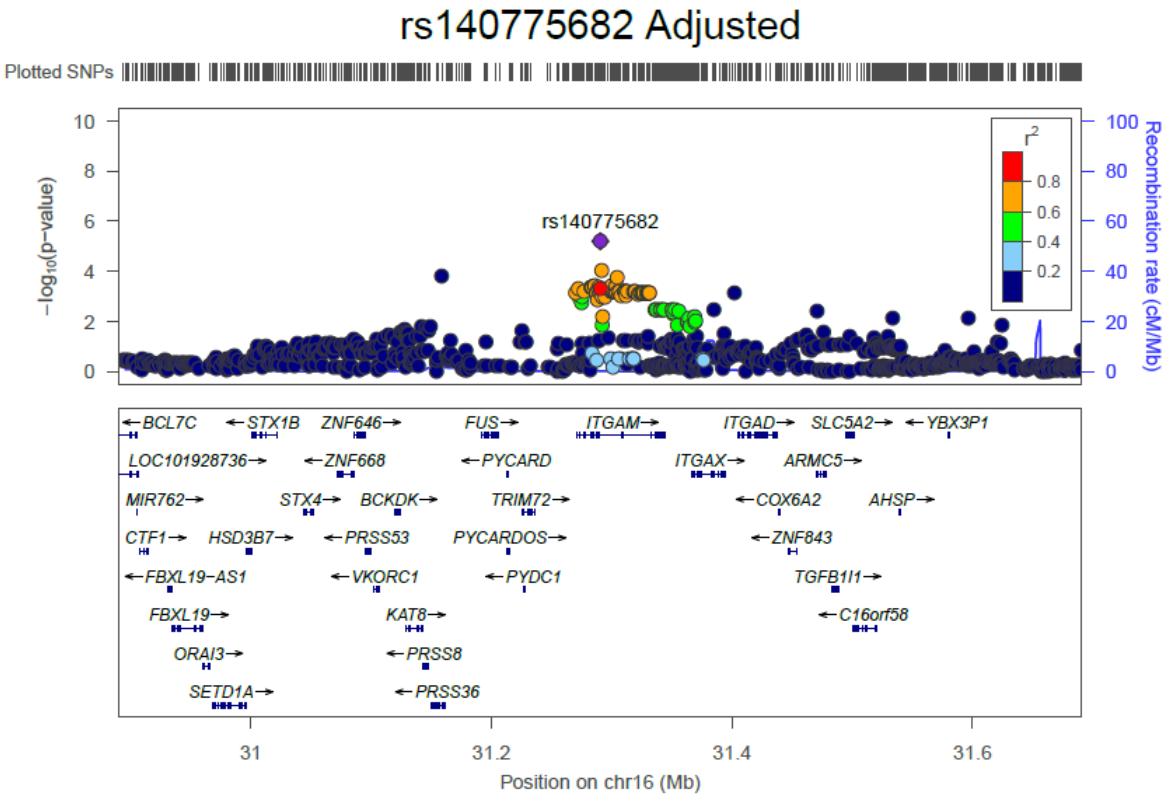

rs12747999 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs12747999
−log10(p-value)
Recombination rate (cM/Mb)
C1orf145
TRIM11
RNF187
RHOU
OBSCN
HIST3H3
BTNL10
MIR6742
DUSP5P1
TRIM17
HIST3H2A
HIST3H2BB
MIR4666A
228.4
228.6
228.8
229
Position on chr1 (Mb)

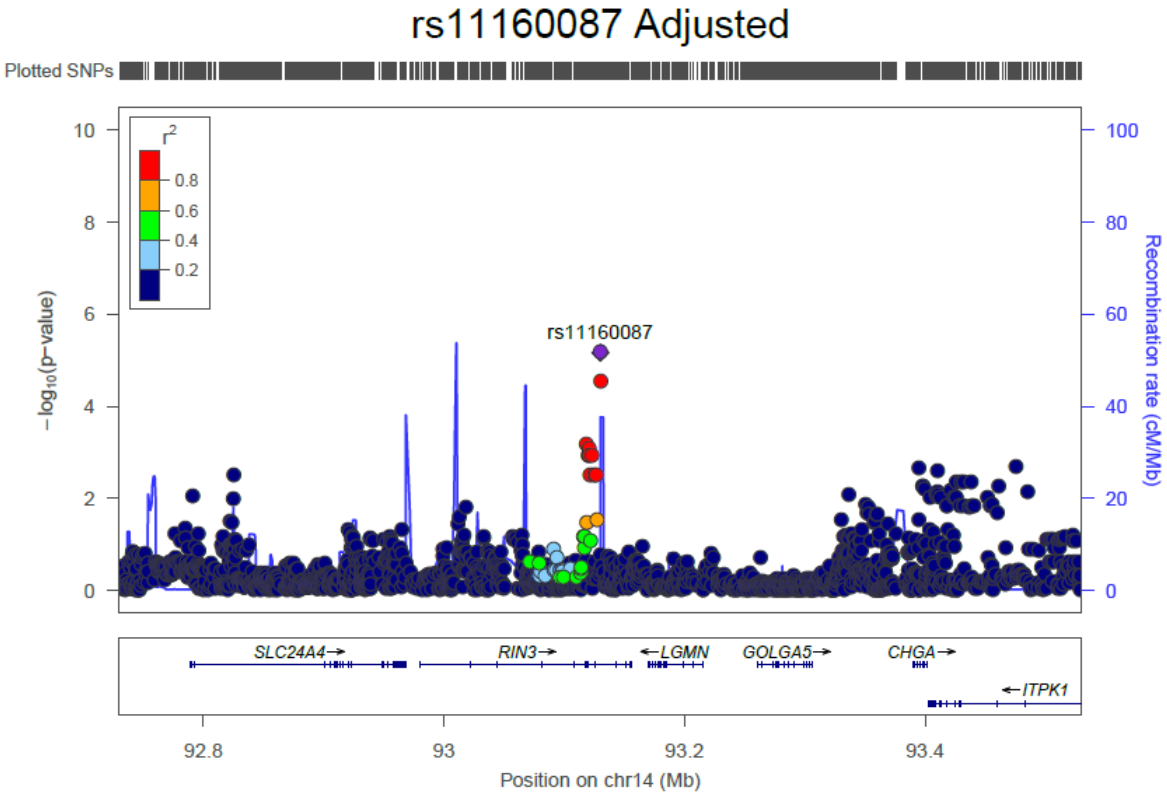

rs73053778 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs73053778
−log10(p−value)
Recombination rate (cM/Mb)
ABCC5
HTR3E
AP2M1
ECE2
CLCN2
EPHB3
MAGEF1
ABCC5−AS1
EIF2B5
VWA5B2
EIF4G1
HTR3D
DVL3
MIR1224
FAM131A
HTR3C
ABCF3
PSMD2
CHRD
ALG3
POLR2H
CAMK2N2
THPO
SNORD66
183.8
184
184.2
184.4
Position on chr3 (Mb)

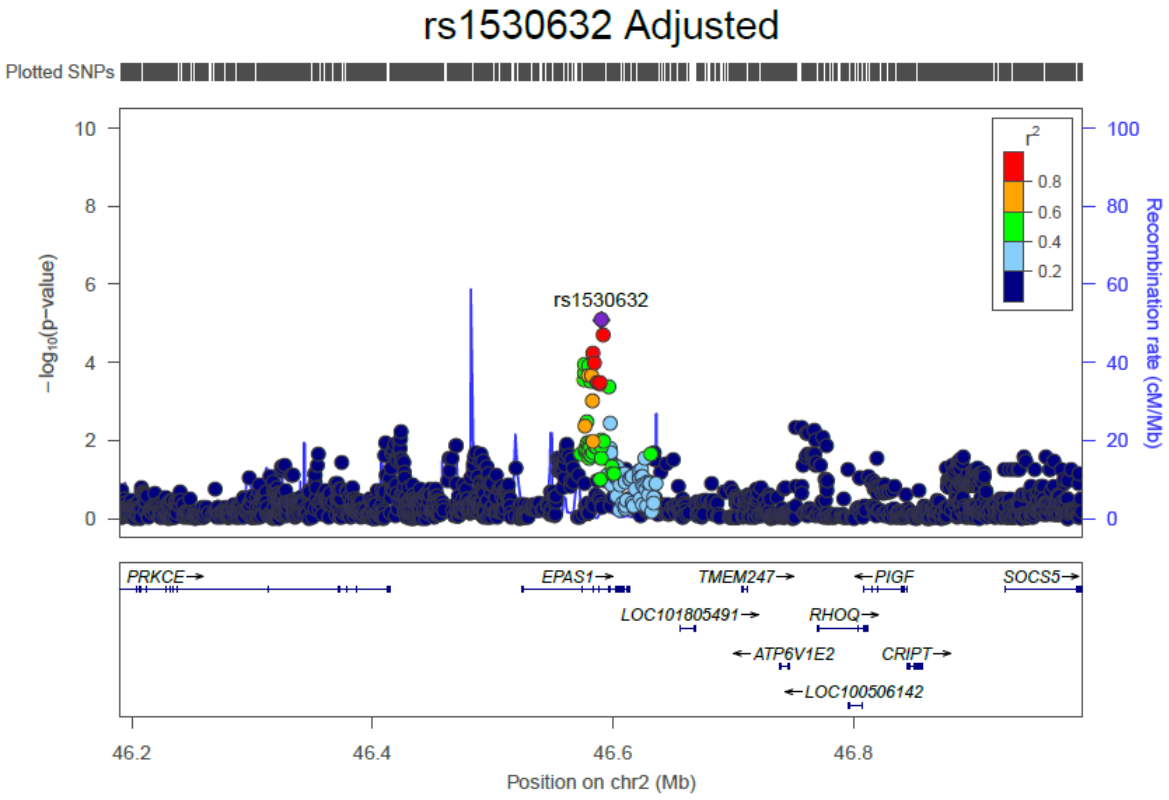

rs7899757 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs7899757
−log10(p−value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

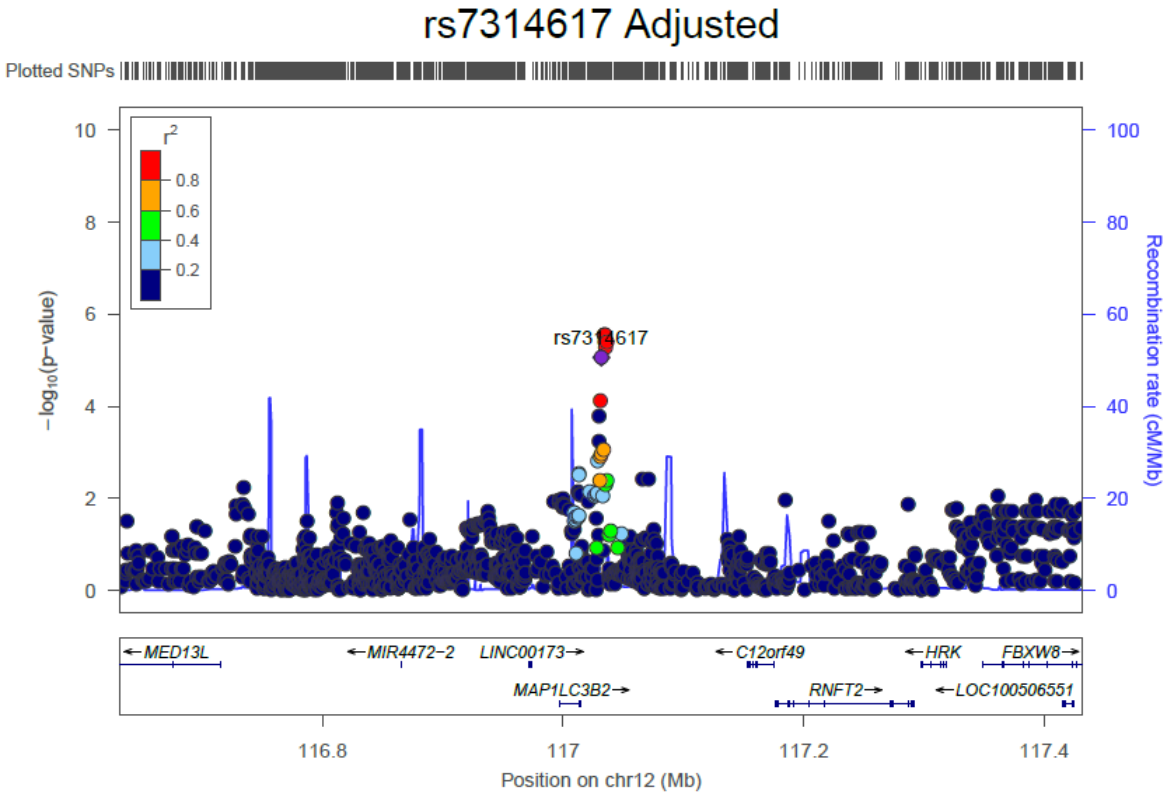

rs7904309 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs7904309
−log10(p−value)
Recombination rate (cM/Mb)
GHITM
LRIT2
CCSER2
C10orf99
LINC00858
CDHR1
LRIT1
RGR
85.8
86
86.2
86.4
Position on chr10 (Mb)

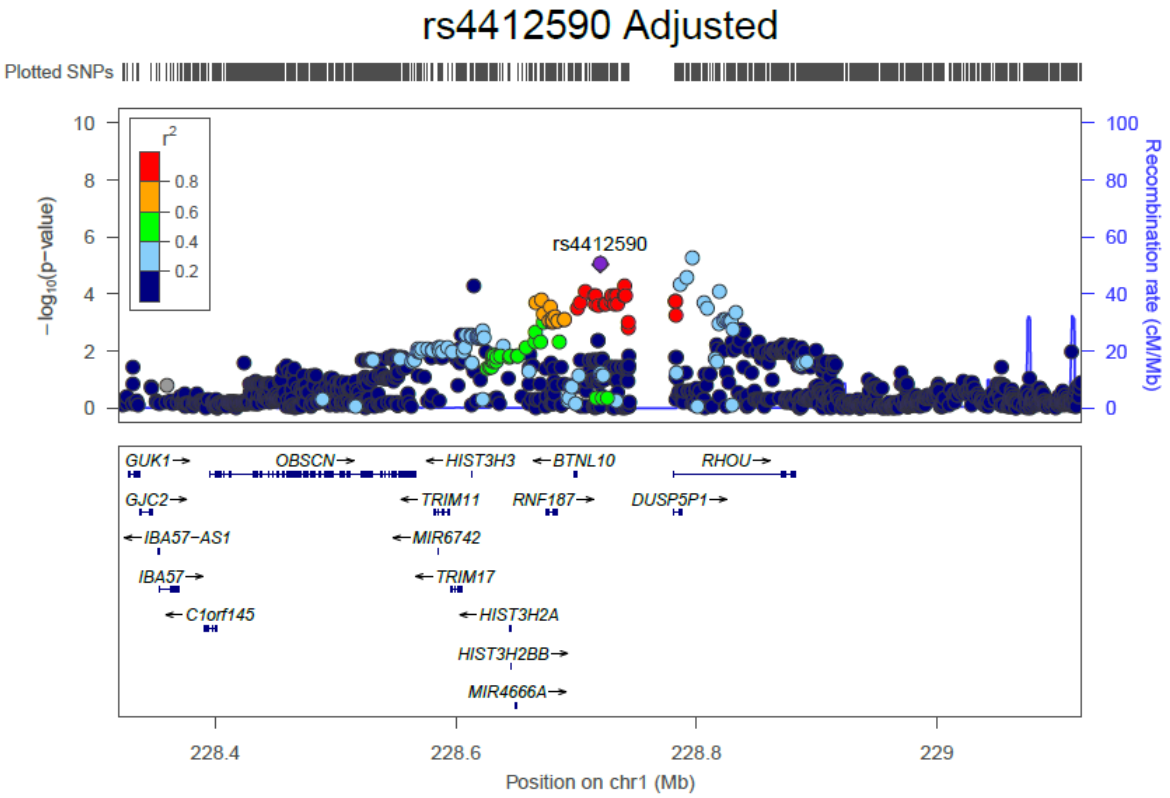

rs7911510 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs7911510
−log10(p−value)
Recombination rate (cM/Mb)
C10orf99
CCSER2
CDHR1
LRIT2
LRIT1
RGR
LINC00858
86
86.2
86.4
86.6
Position on chr10 (Mb)

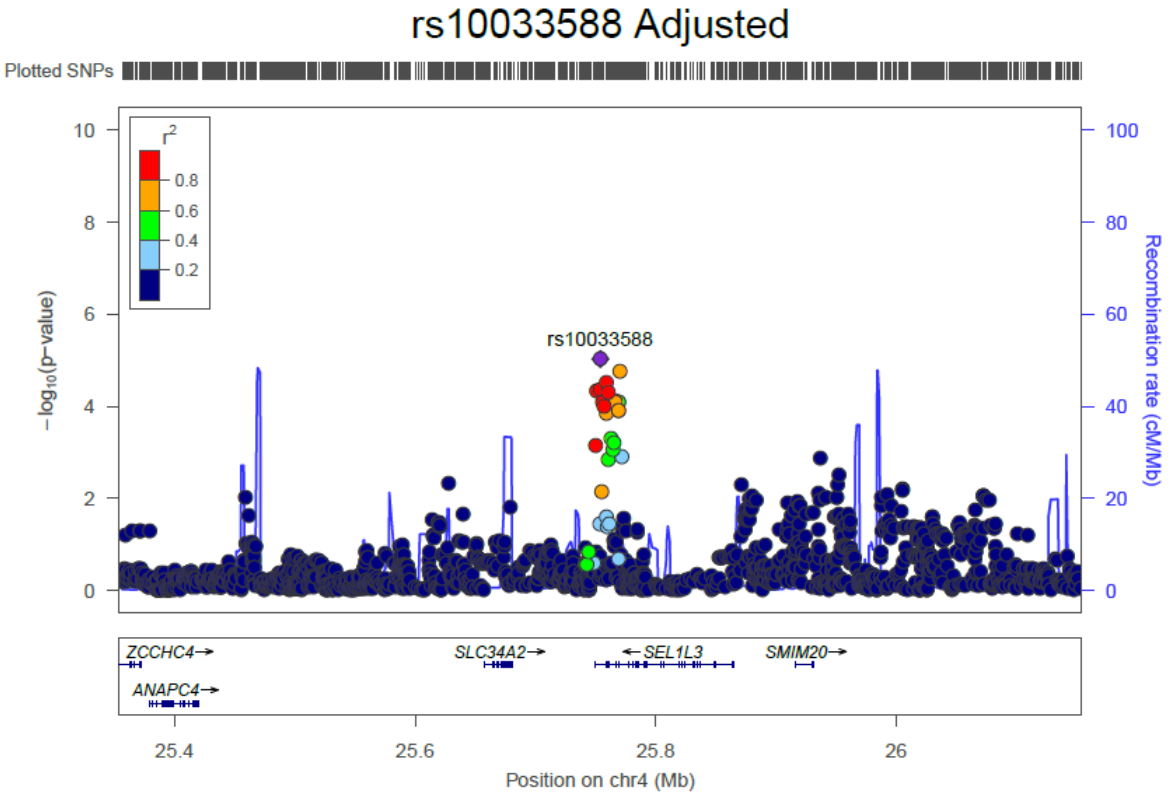

rs140775682 Adjusted
Plotted SNPs
rs140775682
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
BCL7C
STX1B
ZNF646
FUS
ITGAM
ITGAD
SLC5A2
YBX3P1
LOC101928736
ZNF668
PYCARD
ITGAX
ARMC5
MIR762
STX4
BCKDK
TRIM72
COX6A2
AHSP
CTF1
HSD3B7
PRSS53
PYCARDOS
ZNF843
FBXL19-AS1
VKORC1
PYDC1
TGFB1I1
FBXL19
KAT8
C16orf58
ORAI3
PRSS8
SETD1A
PRSS36
Position on chr16 (Mb)


rs11160087 Adjusted
Plotted SNPs
rs11160087
r²
0.8
0.6
0.4
0.2
−log10(p-value)
Recombination rate (cM/Mb)
SLC24A4
RIN3
LGMN
GOLGA5
CHGA
ITPK1
Position on chr14 (Mb)


rs1530632 Adjusted
Plotted SNPs
rs1530632
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
PRKCE→
EPAS1→
TMEM247→
←PIGF
SOCS5→
LOC101805491→
RHOQ→
←ATP6V1E2
CRIPT→
←LOC100506142
46.2
46.4
46.6
46.8
Position on chr2 (Mb)


rs7314617 Adjusted
Plotted SNPs
rs7314617
r²
0.8
0.6
0.4
0.2
−log10(p−value)
Recombination rate (cM/Mb)
←MED13L
←MIR4472−2
LINC00173→
←C12orf49
←HRK
FBXW8→
MAP1LC3B2→
RNFT2→
←LOC100506551
116.8
117
117.2
117.4
Position on chr12 (Mb)


rs4412590 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs4412590
−log10(p−value)
Recombination rate (cM/Mb)
GUK1→
OBSCN→
←HIST3H3
←BTNL10
RHOU→
GJC2→
←TRIM11
RNF187→
DUSP5P1→
←IBA57−AS1
←MIR6742
IBA57→
←TRIM17
←C1orf145
←HIST3H2A
HIST3H2BB→
MIR4666A→
228.4
228.6
228.8
229
Position on chr1 (Mb)


rs10033588 Adjusted
Plotted SNPs
r²
0.8
0.6
0.4
0.2
rs10033588
−log10(p−value)
Recombination rate (cM/Mb)
ZCCHC4→
SLC34A2→
←SEL1L3
SMIM20→
ANAPC4→
25.4
25.6
25.8
26
Position on chr4 (Mb)

Appendix 7

This appendix contain the final progress report created on behalf of EpiPGX WP2 by Dr Graeme Sills, it was submitted to relevant funding body. It was created on 01/12/2015.

WP02: Genomic biomarkers of early treatment response in newly-diagnosed epilepsy

Objectives

To identify biomarkers of remission with the first well-tolerated antiepileptic drug in newly-diagnosed epilepsy
To identify biomarkers that distinguish general and selective drug responsiveness in newly-diagnosed epilepsy
To identify biomarkers of first drug failure in newly-diagnosed epilepsy

Summary of progress towards objectives and details for each task

Summary & progress update

WP02 represents the combined interests and efforts of the Universities of Liverpool (ULIV) and Glasgow (UGLA) and Imperial College London (IMP), with work coordinated from ULIV and with contributions from all other EpiPGX partners. The principal objectives for WP02 are to identify genomic biomarkers of clinically-relevant treatment outcomes following initial antiepileptic drug exposure in patients with newly-diagnosed epilepsy. During this final reporting period, we have made significant progress towards achieving those objectives, as detailed below. This has been achieved thanks to the commitment of staff supported by EpiPGX – Dr Ben Francis and Dr Pauls Auce at ULIV and Dr Sarah Langley and Dr Prashant Srivastava at IMP – and senior investigators (Jorgensen, Marson, Johnson, Sills) at both sites. The WP02 team has continued to meet face-to-face on a six-monthly basis during the final reporting period and our regular teleconferences have increased in frequency to fortnightly. This coordinated effort, together with a clear analysis plan (as described in the previous periodic report in Dec 2014), means that we have now met the majority of our intended milestones and have fully completed two of our three main tasks (and the associated deliverables) and have partially completed the remaining task (with an interim deliverable available). It was hoped that a short no-cost extension to EpiPGX would allow full completion of all WP02 tasks and deliverables but this was not forthcoming. The remaining work – which will necessitate collaboration across investigators in WP02, WP03 and WP04 – will be completed in the first half of 2016.

Progress on specific tasks

A key component of the intended analysis in WP02 was to first establish the relative influence of demographic (i.e. age, sex) and clinical (i.e. epilepsy type) factors in the variability in treatment outcomes in newly diagnosed epilepsy. Understanding and adjusting for these factors (or covariates) should allow a more sensitive investigation of the relative contribution of genomic variants in the subsequent genome-wide association analyses. This essential work was undertaken by Dr Pauls Auce at ULIV using a logistic regression method to identify significant non-genetic predictors of treatment outcome and to quantify the extent of their influence. Although this work was not associated with any specific task or deliverable, it contributed to all of the genomic analyses undertaken in WP02 and the main findings are accordingly reported in the 'significant results' section below.

Task 1: Identifying genome-based biomarkers of remission with first well-tolerated drug

All phenotype and genotype data for this analysis was assembled by end of 2014. Subsequent quality control checks were implemented in early 2015. These resulted in a loss of cases due to either missing phenotype information essential to the determination of treatment outcome or missing genotype information above a pre-determined ceiling for inclusion. The final genome-wide association study (GWAS) was undertaken in a population of 1,514 individuals with newly-diagnosed epilepsy, who had been followed-up prospectively at a single epilepsy centre from initial diagnosis and treatment initiation and until such time that they reached an efficacy end-point associated with their first well-tolerated antiepileptic drug (AED). Patients were stratified into those experiencing an immediate remission from seizures (n=604), those experiencing a later (or deferred) remission (n=248), and those who did not experience remission on their first AED (n=664). This population was then interrogated using a variety of statistical approaches in an effort to identify genomic variants associated with treatment outcome, using binary, multinomial and survival GWAS methods. The latter approach (survival GWAS) necessitated the development of novel statistical methodology – three-way mixed modelling GWAS – which was outlined in the previous periodic report and has been the subject of dissemination activities. The key findings from the Task 1 analysis are reported under 'significant results' below and are described in more detail, together with further information on the statistical model, in the corresponding deliverable (D2.3). This task is now completed.

Task 2: Identifying genome-based biomarkers that distinguish general and selective drug responsiveness

Chronologically, this was always intended as the final task for WP02. It was originally designed as a collaborative effort between WP02 and WP04 and reliant on the identification of genome-wide significant biomarkers of early and late remissions in WP02 Task 1 and WP04 Tasks 1 & 2, respectively. To date, no such biomarkers have been reliably identified in either WP. In a revised approach, the intention is now to undertake collaboration across WP02, WP03 and WP04 and to explore response versus non-response to

specific medications or classes of medication, with analyses adjusted according to whether response is early (first drug) or late (second or subsequent drugs). This will essentially achieve the same result but using a different approach. To this end, we have undertaken an initial GWAS of the patient cohort described above for Task 1 analyses but restricted to those individuals whose first well-tolerated AED was a drug targeting voltage-gated sodium channels as its primary mechanism of action. This group includes several first-line AEDs, including carbamazepine, lamotrigine, phenytoin and oxcarbazepine. This sub-population was then interrogated using a binary GWAS method comparing genotypes between individuals who experienced a remission from seizures on a sodium channel blocker as their first well-tolerated AED compared to those who failed to experience a remission despite adequate drug exposure. The key findings from this initial Task 2 analysis are reported under 'significant results' below and are described in more detail in the corresponding deliverable (D2.6), which is submitted in (partial) fulfilment of this task. Additional analyses to further explore general versus selective drug responsiveness will continue in conjunction with WP03 and WP04 during the first half of 2016.

Task 3: Identifying genomic biomarkers of first drug failure

The patient cohort for this analysis is essentially identical to that for Task 1. As described above, all genotypes and phenotypes had been assembled by end of 2014 and subsequent quality control checks were implemented in early 2015. Cases were again lost due to missing phenotype and/or insufficient genotype data. For this analysis, the final GWAS drew on a population of individuals with newly-diagnosed epilepsy who had been followed-up prospectively at a single epilepsy centre from initial diagnosis and treatment initiation and until such time as their first ever AED failed. Failure was defined as withdrawal of the first drug and/or addition of a second drug. Patients who did not experience drug failure were censored at the time of last recorded clinical visit (n=437). Patients who experienced failure of the first drug were stratified into those failing due to unacceptable adverse events (UAE; n=340) and those failing due to inadequate seizure control (ISC; n=276). This population was then interrogated a novel two-way competing risks GWAS approach, which was developed specifically for EpiPGX. The original intention of using three-way statistical methodology was later revised for several reasons; insurmountable difficulties in data handling within the statistical model, artificial distinction in sub-groups of UAE (i.e. on- and off-target adverse events), and a lower than anticipated number of available cases (there was an unexpectedly large group of patients for whom the reason for first drug failure was recorded as "unknown"). The key findings from the Task 3 analysis are reported under 'significant results' below and are described in more detail, together with further information on the statistical model, in the corresponding deliverable (joint report for D2.4 and D2.5). This task is now completed.

<u>Milestones and deliverables for WP02</u>

The following milestones and deliverables were outstanding at the time of the previous periodic report (Dec 2014).

Type	No.	Title	Due date	Status
Milestone	M24	Association analysis for first well tolerated AED	Feb 2015*	Milestone reached
Milestone	M25	Development of three-way competing risks approach for analysis of drug failure	Jan 2015	Milestone reached
Milestone	M26	Association analysis for first drug failure	Apr 2015	Milestone reached
Milestone	M27	Development of prediction models for response to specific AEDs	Apr 2015	Milestone not reached (expected in first half of 2016)
Deliverable	D2.3	Validated genome-based biomarkers [of remission with first AED]	Feb 2015*	Delivered
Deliverable	D2.4	List of discovery genome-based biomarkers of first drug failure	Apr 2015*	Delivered (joint report for D2.4 and D2.5)
Deliverable	D2.5	List of validated genome-based biomarkers of first drug failure	Oct 2015	Delivered (joint report for D2.4 and D2.5)
Deliverable	D2.6	List of validated genome-based biomarkers of general responsiveness to AEDs	Oct 2015	Partially delivered (interim report available)

**Original date revised as described in second periodic report (Dec 2014)*

Significant results

Clinical and demographic predictors of treatment outcome in newly diagnosed epilepsy

A total of 1,906 newly-diagnosed epilepsy patients were identified in the EpiPGX eCRF as eligible for inclusion in WP02. Of those, key phenotype information was missing in 183 (9.6%), leaving 1,723 individuals in the logistic regression analysis of non-genetic influences on treatment outcome. Of those, 43.4% came from the SANAD trial (UK), 24.0% from the Western Infirmary, Glasgow, 13.5% from Tübingen, and 9.3% from Melbourne. Other centres contributed less than 5% of cases. Males comprised 51.5% of the cohort, mean age at diagnosis was 35.6 years and mean duration of follow-up was 5.1 years. Over 70% of the study population had focal epilepsy. Remission was achieved with the first well-tolerated AED in 56.2% of cases.

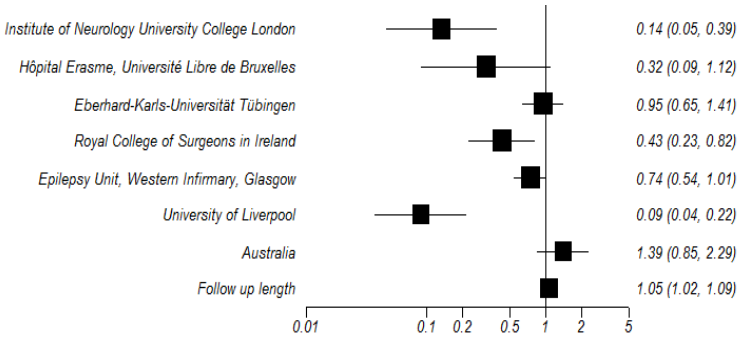


Figure 1: Forest plot (including odds ratios and 95% confidence intervals) for association between source cohort and duration of clinical follow-up and the likelihood of remission with first well-

Statistically significant positive predictors of remission included generalised epilepsy (OR=1.94, 95%CI=1.39-2.73) and older age at diagnosis (OR=1.02, 95%CI=1.01-1.03). Negative predictors of remission included abnormal neurological exam (OR=0.44, 95%CI=0.29-0.67), high pre-treatment seizure count (OR=0.56, 95%CI=0.38-0.84), and abnormal focal MRI (OR=0.56, 95%CI=0.39-0.81). Other predictors included duration of follow-up (OR=1.05, 95%CI=1.02-1.09) and source cohort (figure 1).

The data are entirely consistent with pre-existing knowledge on treatment outcomes in newly diagnosed epilepsy. This gives confidence that the EpiPGX cohort is representative of new-onset epilepsy. In addition, this analysis has identified several key covariates that need to be adjusted for in subsequent GWAS analyses.

Genomic biomarkers of remission with the first well-tolerated antiepileptic drug in newly-diagnosed epilepsy

The first ever AED to which a patient with newly-diagnosed epilepsy is exposed is often discontinued due to

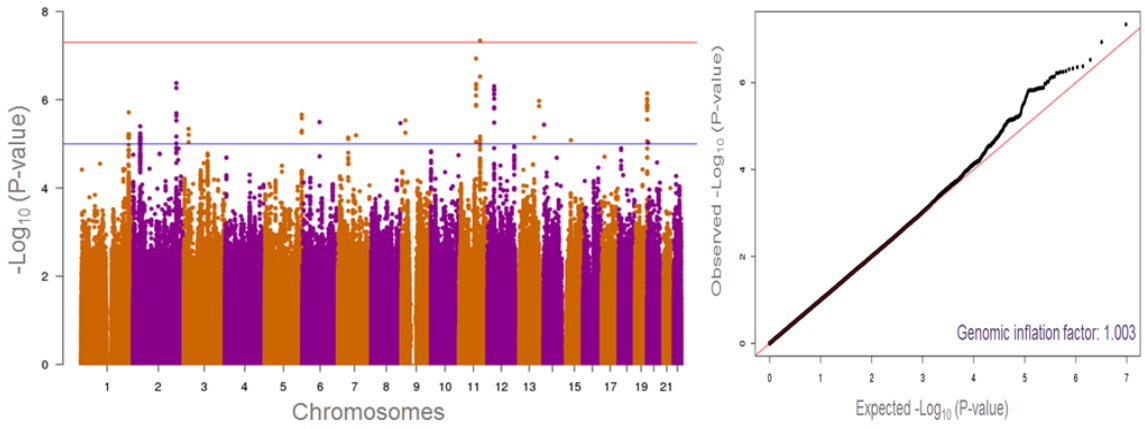


Figure 2: Manhattan plot and associated QQ plot for binary GWAS analysis of deferred vs no remission with the first well-tolerated antiepileptic drug in patients with newly-diagnosed epilepsy. This analysis was unadjusted for clinical and demographic variables known to influence treatment outcome. Genotypes were filtered for MAF<0.05 and INFO<0.5. There were 122 SNPs with $p<10^{-5}$ including one SNP (rs72996844) that reached genome-wide significance ($p=4.7x10^{-8}$).

adverse effects before a satisfactory evaluation of its efficacy can be made. As a result, we chose to explore the first well-tolerated AED which is the same as the first ever AED in most patients but can be the second or even third medication to which a patient is exposed in cases where tolerability is an initial problem. Treatment outcome on the first well-tolerated AED was stratified into immediate remission (12 months seizure freedom commencing within the first 2 months since diagnosis), deferred remission (12 months seizure freedom commencing at some point beyond 2 months from diagnosis) and no remission (no 12 month period of seizure period at any point during exposure to the first well tolerated drug). We used binary (comparison of remission vs no remission), multinomial (3-way comparison) and survival (time to achieve remission) analyses to interrogate the data. An example of the output from this analysis is shown below (figure 2 and table 1). Full results are provided in accompanying deliverable D2.3.

Table 1: Top 5 loci for binary GWAS of deferred vs no remission (unadjusted for clinical covariates)

SNP	Imp	Gene	Chr	Pos (bp)	Test allele	Ref allele	MAF	P-value	Adj P-value
rs72996844	yes	CNTN5	11	99916286	C	T	0.06	4.56E-08	4.78E-08
rs12791153	yes	-	11	80685181	T	A	0.08	1.16E-07	1.21E-07
rs770584	yes	CNTN5	11	99902569	G	A	0.06	2.98E-07	3.10E-07
rs6715132	yes	-	2	208141163	C	G	0.26	4.19E-07	4.36E-07
rs11603610	yes	-	11	80668994	G	T	0.08	4.40E-07	4.58E-07

In this analysis, a single SNP (rs72996844) on chr11 and located in the *CNTN5* gene, which encodes contactin-5 was marginally genome-wide significant in a binary analysis of patients with deferred vs no remission but only when clinical covariates were excluded from the analysis. This SNP is of potential biological significance, having been reported to harbour de novo epilepsy-causing mutations in a recent exome-sequencing analysis by the Epi4K group. Other than this single SNP in *CNTN5*, none of the other binary, multinomial or survival analyses undertaken as part of WP02 Task 1 revealed any association that achieved genome-wide significance. This would suggest that there are no strong genomic influences on response to treatment with the first well-tolerated AED in newly-diagnosed epilepsy.

<u>Genomic biomarkers distinguishing general and selective drug responsiveness in newly-diagnosed epilepsy</u>

This analysis represent the first step in an attempt to distinguish between patients are who are essentially responsive to whatever AED they first receive and those who require a very specific drug or drug from a specific class in order to control their seizures. In this case, we chose to focus on patients who received a sodium channel blocking AED (i.e. carbamazepine, lamotrigine, phenytoin, oxcarbazepine) as their first well-tolerated drug. Treatment outcome was stratified into remission (12 months seizure freedom occurring at any point during follow-up) and no remission (no 12 month period of seizure period at any point during

exposure to the first well tolerated drug) and we used a binary GWAS to interrogate the data. An example of the output from this analysis is shown below (figure 3 and table 2). Full results are provided in accompanying deliverable D2.6.

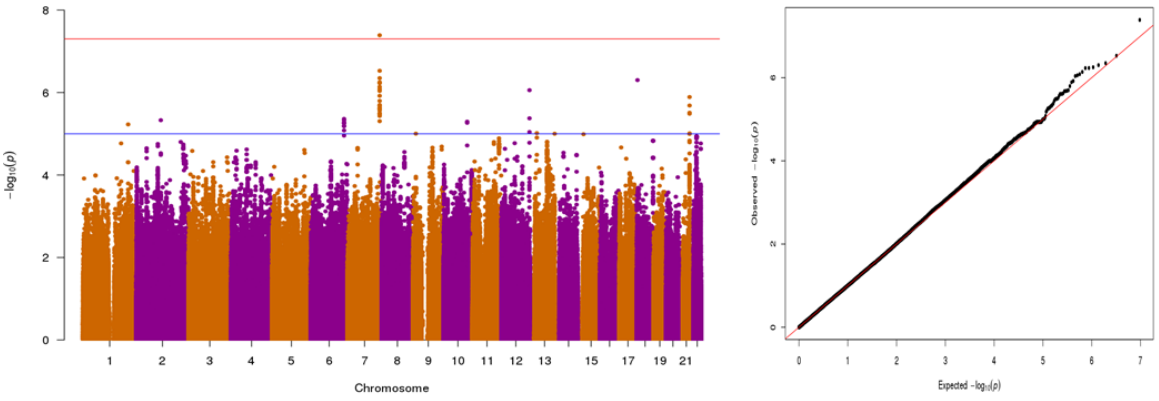


Figure 3: Manhattan plot and associated QQ plot for binary GWAS analysis of remission vs no remission in patients with newly-diagnosed epilepsy for whom the first well-tolerated antiepileptic drug was a sodium channel blocking agent. This analysis was unadjusted for clinical and demographic variables known to influence treatment outcome. Genotypes were filtered for MAF<0.05 and INFO<0.5. There were 49 SNPs with $p<10^{-5}$ including one SNP (rs1967394) that reached genome-wide significance ($p=4.08x10^{-8}$).

Table 2: Top 5 loci for binary GWAS of remission vs no remission in patients treated with sodium channel blocking AEDs

SNP	Imp	Gene	Chr	Pos (bp)	Test allele	Ref allele	MAF	P-value	Adj P-value
rs1967394	yes	-	7	149681561	G	C	0.06	4.08E-08	-
rs7235163	no	DLGAP1	18	3725553	G	A	0.46	4.99E-07	-
chr12:130642445:D	yes	FZD10-AS1	12	130642445	ATAC	-	0.13	8.76E-07	-
rs144402785	yes	-	21	46245856	T	A	0.10	1.28E-06	-
rs883987	yes	CNKSR3	6	154781155	G	C	0.35	4.35E-06	-

In this analysis, a single SNP (rs1967394) on chr7 was marginally genome-wide significant in a binary analysis of remission vs no remission in patients treated with sodium channel blocking AEDs. This SNP is located in an intergenic region, with no clear association with any particular gene. The analysis reported above was not adjusted for clinical and demographic factors known to influence treatment outcome, however this SNP remained significant (albeit not genome-wide) when the analysis was adjusted for such covariates. This lends weight to the role of rs1967394 in treatment response with sodium channel blocking agents. Further analysis is planned with these compounds, both individually and as a class of drugs, in conjunction with WP03 and WP04 as described above.

Genomic biomarkers of first drug failure in newly-diagnosed epilepsy

The first ever AED to which people with newly-diagnosed epilepsy are exposed typically fails in up to 70% of cases. As discussed above, there are two key reasons for that failure – inadequate seizure control (ISC) or unacceptable adverse events (UAE). This analysis was undertaken to seek genomic biomarkers that might predict treatment failure and the reason for it. It required the development of a novel statistical methodology – two-way competing risks GWAS – which was then applied to a subset (n=799) of our newly-diagnosed epilepsy cohort who had a clear failure of their first ever AED recorded in the eCRF. Reason for first drug failure (ISC or UAE) was identified for all cases and a two-way competing risks GWAS used to interrogate the data. In this methodology, the hazard of failing from either cause is continually being calculated, which gives rise to two association plots and two sets of loci associated with outcome. A typical output from this analysis is shown below (figure 4 and tables 3 & 4). Full results are provided in corresponding deliverable (joint report for D2.4 and D2.5).

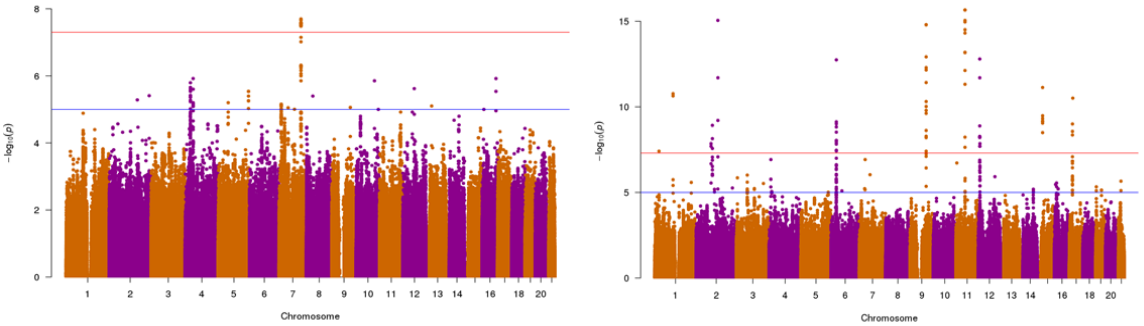


Figure 4: Comparative Manhattan plots for two-way competing risks GWAS of failure with first ever AED in newly-diagnosed epilepsy. Left-hand plot shows association analysis for failure due to inadequate seizure control (ISC), while right-hand plot shows association analysis for failure due to unacceptable adverse events (UAE). Both analyses were run without adjustment for clinical covariates. In both cases, genotypes were filtered for MAF<0.05 and INFO<0.5. In the ISC analysis, there were 68 SNPs with $p<10^{-5}$ including 12 SNPs that reached genome-wide significance ($p<5x10^{-8}$). In the UAE analysis, there were 190 SNPs with p<10-5 including 68 SNPs that reached genome-wide significance ($p<5x10^{-8}$).

Table 3: Top 5 loci for competing risks GWAS of first drug failure due to inadequate seizure control

SNP	Imp	Gene	Chr	Pos (bp)	Test allele	Ref allele	MAF	P-value	Adj P-value
rs6950012	yes	GRM8	7	126668420	C	T	0.23	2.20E-08	-
rs9319547	yes	-	16	80042084	T	G	0.11	1.20E-06	-
rs139820017	yes	-	4	45027060	A/T	-	0.19	1.20E-06	-
rs12250166	yes	SLK	10	105776055	T	C	0.07	1.40E-06	-
rs1821692	yes	-	4	29279092	A	C	0.20	1.60E-06	-

Table 4: Top 5 loci for competing risks GWAS of first drug failure due to unacceptable adverse events

SNP	Imp	Gene	Chr	Pos (bp)	Test allele	Ref allele	MAF	P-value	Adj P-value
rs139118386	yes	-	11	54767936	A	T	0.23	2.20E-16	-
rs138964936	yes	-	2	131208125	T	A	0.05	8.90E-16	-
rs140792738	yes	-	9	97079046	C	G	0.11	1.60E-15	-
rs6487118	yes	-	12	8594235	A	T	0.06	1.60E-13	-
rs9469157	yes	-	6	32545390	C	G	0.05	1.80E-13	-

In this analysis, several SNPs were seen to achieve genome-wide significance, most notably in the analysis of failure due to UAE. The reason for such a large number of statistically significant associations is unclear but may be related to the relatively modest numbers of cases in each failure group. In the analysis of failure due to ISC, the most significant locus corresponded to a region on chr7 within the GRM8 gene, which encodes a specific sub-type of metabotropic glutamate receptor with strong plausibility for involvement in epilepsy and response to AED therapy. In contrast, the most significant loci associated with failure due to UAE fell within intergenic regions on various chromosomes. There was no clear association with any gene and no functional conclusions can be drawn. Further exploration of the data is merited, particularly in light of the large number of genome-wide significant signals. Identification of further cases for this analysis, via re-evaluation of phenotypes and the method of phenotype derivation, would add to the robustness of the findings and their reliability.