

## SPECIAL REPORT

# Epi4K: Gene discovery in 4,000 genomes

The Epi4K Consortium

Center for Human Genome Variation, Duke University, Durham, North Carolina, U.S.A.

### SUMMARY

**A major challenge in epilepsy research is to unravel the complex genetic mechanisms underlying both common and rare forms of epilepsy, as well as the genetic determinants of response to treatment. To accelerate progress in this area, the National Institute of Neurological Disorders and Stroke (NINDS) recently offered funding for the crea-**

**tion of a “Center without Walls” to focus on the genetics of human epilepsy. This article describes Epi4K, the collaborative study supported through this grant mechanism and having the aim of analyzing the genomes of a minimum 4,000 subjects with highly selected and well-characterized epilepsy.**

**KEY WORDS:** Epilepsy, Epileptic encephalopathies, Genetics, Phenotyping, Prognosis, Sequencing, Copy number variants.

Great progress has been made in the past two decades in unraveling the genetic bases of epilepsy. Specific gene mutations, mostly involving ion channels and neurotransmitter receptors, are now recognized as the cause of rare inherited epilepsy syndromes. De novo mutations, including copy number variants (CNVs), are associated with some of the epileptic encephalopathies and a small percentage of patients with common epilepsy syndromes (Poduri & Lowenstein, 2011; Sisodiya & Mefford, 2011). However, the genetic mechanisms that influence the development of epilepsy in the vast majority of patients remain unknown. Recognizing the importance of accelerating progress in this area of epilepsy research, the National Institute of Neurological Disorders and Stroke (NINDS) issued a Funding Opportunity Announcement in 2010 for the creation of a “Center without Walls” to focus on the genetics of human epilepsy. This article describes Epi4K, the collaborative study supported through this mechanism and having the aim of sequencing (at least) 4,000 subjects with epilepsy. We begin with a general overview of epilepsy genetics, followed by a more detailed discussion of approaches and progress in gene identification in the complex epilepsies. We then describe the genesis of the Epi4K Consortium, including its structure and the development of policies meant to uphold key philosophical underpinnings of the nature of the collaboration. Finally, we provide details on the scientific

rationale and proposed methods for the first four projects to be implemented by Epi4K.

### OVERVIEW OF EPILEPSY GENETICS

The risk for epilepsy is clearly genetically influenced: risk is increased two- to fourfold in the first-degree relatives of people with epilepsy of unknown cause {either genetic (previously called idiopathic generalized epilepsies [IGEs]) or nonlesional focal epilepsies [NFEs]} (Annegers et al., 1982; Ottman et al., 1996a,b); twin studies consistently show higher concordance in monozygotic than dizygotic pairs (Corey et al., 1991; Berkovic et al., 1998; Kjeldsen et al., 2003; Vadlamudi et al., 2004); and a large number of genes with a major effect on susceptibility has already been identified (Helbig et al., 2008; Reid et al., 2009; Ottman et al., 2010). The study of rare families with Mendelian inheritance has provided extremely important information about epileptogenic mechanisms. The availability of large autosomal dominant pedigrees with epilepsy combined with the application of linkage analysis and positional cloning or positional candidate approaches has led to the identification of epilepsy genes that, to date, mostly encode subunits of voltage-gated or ligand-gated ion channels (Mulley et al., 2003; Prasad & Prasad, 2008; Poduri & Lowenstein, 2011). However, despite the clear importance of genetics in the epilepsies, most patients have no affected relatives, and progress in identifying and characterizing the underlying genetic defects in these “genetically complex” epilepsies has so far been slow (Mulley et al., 2005; Ottman, 2005).

Some of the genetic epilepsies are not inherited, but instead result from de novo mutations occurring in the germ cell line of the parents. Such de novo mutations are extremely important in severe myoclonic epilepsy of infancy (SMEI), now known as Dravet syndrome, in which >70% of

Accepted March 22, 2012; Early View publication May 29, 2012.

Address correspondence to Darcy McMullin, Ph.D., Center for Human Genome Variation, Duke University, 308 Research Dr., 342A LSRC B-Wing, Durham, NC 27708, U.S.A. E-mail: darcy.mcmullin@duke.edu

Re-use of this article is permitted in accordance with the Terms and Conditions set out at [http://wileyonlinelibrary.com/onlineopen#OnlineOpen\\_Terms](http://wileyonlinelibrary.com/onlineopen#OnlineOpen_Terms).

Wiley Periodicals, Inc.

© 2012 International League Against Epilepsy

patients have *SCN1A* mutations of which more than 90% arise de novo (Claes et al., 2001; Depienne et al., 2009). Cases with de novo mutations have also been found in genes initially identified in large families (Phillips et al., 2000; Bissulli et al., 2004; Claes et al., 2004; Bertrand et al., 2005; Michelucci et al., 2007; Ishii et al., 2009).

Recently, copy number variants (CNVs) have been found to play an important role in the genetically complex epilepsies, and many of these occur de novo (Scheffer & Berkovic, 2010). CNVs are deletions, duplications, or insertions of DNA that range in size from ~1 kb to several megabases and may affect one or more genes. Three large, recurrent microdeletions at 15q13.3, 16p13.11, and 15q11.2 are each present in 0.5–1% of patients with epilepsy (Dibbens et al., 2009; Helbig et al., 2009; de Kovel et al., 2010; Mefford et al., 2010). Microdeletion of 15q13.3 appears to confer risk specifically for IGEs with an odds ratio of 68 (Dibbens et al., 2009). It is important to note that all three of these deletions are also established risk factors for related disorders, including intellectual disability, autism, and schizophrenia (Mefford & Mulley, 2010). Rare, nonrecurrent CNVs are also important. Large (>2 Mb), rare CNVs, including the deletion at 16p13.11, were found to be enriched in patients with diverse epilepsy syndromes (Heinzen et al., 2010). Rare CNVs were also found in ~10% of patients with various types of epilepsy (Mefford et al., 2010) and ~8% of patients with epileptic encephalopathies (Mefford et al., 2011).

## GENE IDENTIFICATION IN THE COMPLEX EPILEPSIES

Historically, studies aimed at gene identification in complex disorders have used linkage and association approaches. Linkage studies rely on genetic and phenotypic information from multiple generations to map regions of the genome that are co-inherited with disease in families. Regions of interest from linkage studies tend to be relatively large due to the few opportunities for recombination within the families investigated. Association studies, on the other hand, look for variants at individual loci that occur more commonly in unrelated patients than controls. Since association studies draw from many different unrelated individuals who each have a unique recombination history, the size of the associated region should theoretically be much smaller, and the identification of specific susceptibility genes may be more easily resolved than with a linkage study (Cardon & Bell, 2001). To identify the specific susceptibility genes generating a linkage signal, linkage studies are often followed by association mapping. This approach successfully identified the well-known apolipoprotein E (*APOE*) association with Alzheimer's disease (Pericak-Vance et al., 1991; Corder et al., 1993; Saunders et al., 1993), and the association of human leukocyte antigen (HLA) with many immune system disorders (Tomlinson & Bodmer, 1995).

Candidate gene association studies focus on genes thought likely to be involved in the disease pathophysiology. In epilepsy, the largest such study looked for associations with common variants in 229 candidate genes in 2,717 patients with epilepsy and 1,118 controls, and failed to identify any clear associations (Cavalleri et al., 2007b). Numerous small scale association studies in epilepsy have been published without yielding reproducible findings (Tan et al., 2010). Moreover, in a search for rare variants in idiopathic epilepsies, Klassen et al. (2011) resequenced protein coding regions of ion channel genes and did not find an excess of variation compared to controls. However, the small samples sizes may have precluded the identification of genetic risk factors of considerable effect sizes.

The lack of success of candidate gene studies in epilepsy and other complex diseases has been thought to result in part from inadequate knowledge of disease pathophysiology needed to select candidate genes. Consequently, the use of genome-wide association studies (GWAS), which allow for unbiased genome-wide association analyses, marked an important turning point in the field. GWAS employ genotyping chips to genotype “tagging” variants that capture almost all common variants (population frequency >5%) in the European and Asian populations, with slightly lower coverage for subjects of African ancestry (Barrett & Cardon, 2006). Prior to GWAS, only a handful of secure associations had been identified with linkage and association-based approaches in complex diseases (e.g., *APOE4*: Corder et al., 1993; Pericak-Vance et al., 1991; Saunders et al., 1993; HLA: Tomlinson & Bodmer, 1995); however, from 2005, when the first GWAS was published, to the middle of 2011, more than 1,449 associations have been published in more than 237 different conditions (Hindorff et al., n.d.). While some of these may be false positive, the majority of them satisfy significance thresholds sufficient to confirm the presence of real effects in the relevant genomic regions.

While GWAS have been successful in identifying secure risk factors for common diseases; the impact of these findings has been somewhat constrained by two principal factors: (1) the effects of the variants have been generally modest and, even cumulatively, often explain relatively small proportions of the presumed heritability; and (2) causal variants that underlie the association signals have proven difficult to identify, limiting the biological inferences that can be drawn (Goldstein, 2009; Manolio et al., 2009). In epilepsy only two GWAS have been published so far. The first, which included nearly 4,000 European patients with a wide array of nonlesional and lesional focal epilepsies, identified no secure associations (Kasperaviciute et al., 2010). Power was sufficient to exclude variants with an odds ratio of 1.3 or greater, and thus the results argue against common genetic effects of this magnitude that are shared across different (and likely heterogeneous) forms of focal epilepsy in European populations. In a GWAS of 1,087 focal epilepsy

patients of Han Chinese descent, two suggestive but not definitive associations were reported, including one within a gene encoding a cytoskeletal protein (Guo et al., 2012). Analyses of smaller, more clinically homogenous subgroups (e.g., nonlesional focal epilepsies), in which family studies show a stronger genetic influence (Ottman et al., 1996a) may reveal associations masked by the clinical heterogeneity in both of these studies. GWAS are underway in other forms of epilepsy, including IGE, and meta-analyses using the data from a number of consortia (e.g., EPICURE, 2007; EPIGEN, 2008; Guo, 2012; Lohoff et al., 2005) are being undertaken through the efforts of a recently formed International League Against Epilepsy (ILAE) Consortium on Genetics of Complex Epilepsies. The success in identifying common variants that influence epilepsy susceptibility will rely on these collaborative efforts.

Given that GWAS have failed to identify many variants of large effect in complex disease, there is growing interest in the possibility that variants too rare to be represented well on GWAS chips may be important contributors to common diseases, including the epilepsies. This possibility has led to considerable interest in the use of next-generation sequencing (NGS), which allows for the near-comprehensive characterization of genetic variants across the genome, to identify rare risk factors for disease (Cirulli & Goldstein, 2010). The rare variant hypothesis is supported by the recent observation that rare CNVs can have large effects on neuropsychiatric disease risk (Need & Goldstein, 2010). This observation, in conjunction with the fact that only 9% of variants reported as disease-associated in the Human Gene Mutation Database are copy number variants (HGMD Professional version 2011.3) (Stenson et al., 2009), strongly suggests that other types of rare variants, such as single nucleotide and indel (insertion and/or deletions) variants, may be important contributors to complex disease. Early successes of next-generation sequencing in identifying the cause of several Mendelian diseases makes clear that pathogenic mutations can be found using this approach (Ng et al., 2010a,b; Sobreira et al., 2010), including recent studies in epilepsies (Corbett et al., 2010, 2011; Veeramah et al., 2012).

There is great potential for NGS to identify susceptibility genes for complex disease; however, two lines of evidence indicate that large sample sizes will be required. First, based on previous studies of epilepsy genetics, both locus and allelic heterogeneity are predicted to be high. Second, all genomes carry a large number of rare functional variants, which means that appropriately correcting for all the tested hypotheses requires very low p-values (Pelak et al., 2010). As discussed previously, direct evidence for the complexity of this variation in epilepsy was provided in a recent study that characterized the gene sequences of 237 ion channel genes in a cohort of individuals with IGE (Klassen et al., 2011).

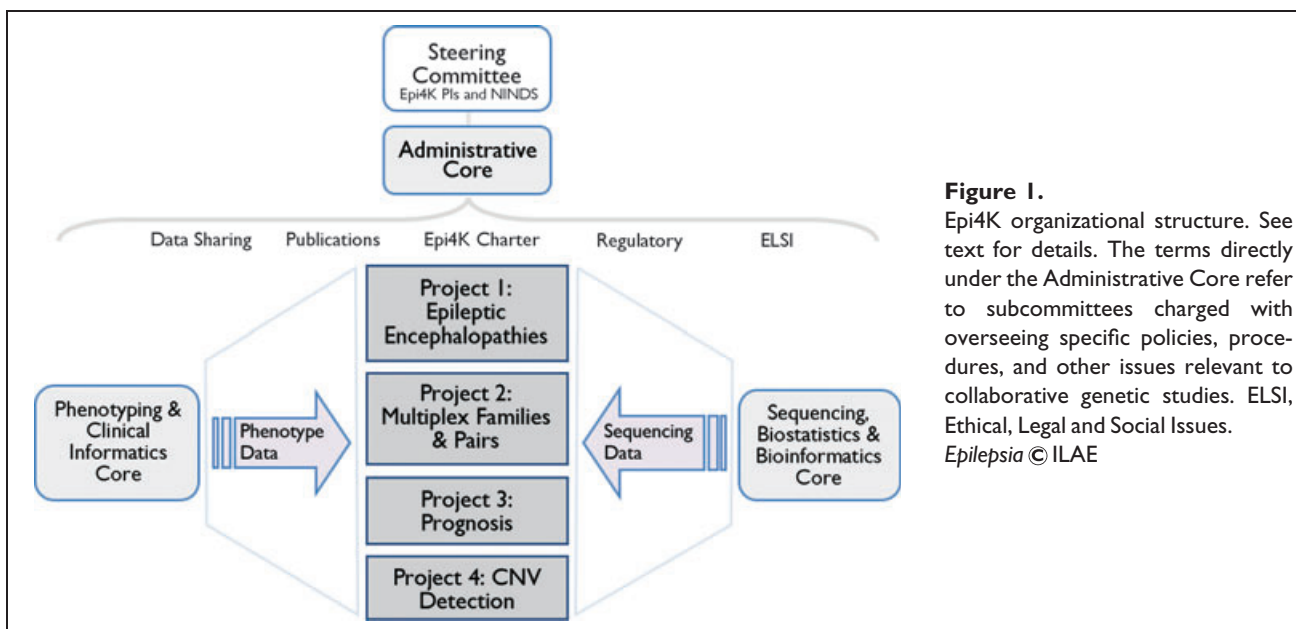
Given the complexity of epilepsy genetics and the difficulty for any single group to have sufficient sample sizes

for further gene discovery, the critical need for large-scale collaborative efforts has become increasingly clear.

## THE GENESIS OF EPI4K

In late summer 2010, NINDS sponsored an international workshop on the genetics of human epilepsy in San Diego, attended by epilepsy genetics researchers from throughout the world, as well as leaders in genome sequencing, statistical genetics, and functional genomics. The goals of the meeting were to advance epilepsy genetics research by reviewing the state of current investigations, promoting collaboration, and brainstorming about new strategies likely to be successful in identifying and characterizing genetic effects on complex epilepsies.

Following the workshop and internal discussions, NINDS issued a Funding Opportunity Announcement (FOA) in November 2010 to solicit applications for new “Centers without Walls for Collaborative Research in the Epilepsies: Genetics and Genomics of Human Epilepsies” (DHHS, 2010). The FOA encouraged applications from large, multidisciplinary groups and called for linked applications that could include an administrative core, a genomics core, and specific scientific projects. Catalyzed by this grant opportunity, we began meeting to identify the scientific questions of highest priority that could take advantage of the valuable epilepsy patient cohorts that have been assembled and the increased ability to apply NGS to the analysis of thousands of genomes. Our discussions also focused on the optimal structure for a Center without Walls, as well as our shared philosophy about key issues relating to collaboration and publication in multidisciplinary research. As a result, we developed a structure of three cores and four scientific projects, as well as a steering committee comprising the study primary investigators and representatives from NINDS (Fig. 1). The three cores include the Administrative Core (responsible for overall coordination of the Center’s activities), the Sequencing, Biostatistics and Bioinformatics Core (responsible for all NGS and initial identification of sequence variants), and the Phenotyping and Clinical Informatics Core (responsible for verifying and archiving the phenotypic data associated with every DNA sample that will be sequenced). The projects, described in more detail in the following sections, include three that analyze specific sets of epilepsy cohorts (epileptic encephalopathies, multiplex families and pairs, and prognosis) drawn from seven large-scale genetic studies from throughout the world (Table 1), and one that will apply cutting-edge analytic techniques related to the detection of CNVs. The proposed number of patients to be analyzed in the scientific projects is a minimum of 4,000; thus we named the Center “Epi4K: Gene Discovery in 4,000 Genomes.” The planned timeline for this effort spans 5 years (beginning October 2011), and includes milestones for the completion of project 1 (epileptic encephalopa-



**Figure 1.** Epi4K organizational structure. See text for details. The terms directly under the Administrative Core refer to subcommittees charged with overseeing specific policies, procedures, and other issues relevant to collaborative genetic studies. ELSI, Ethical, Legal and Social Issues. *Epilepsia* © ILAE

**Table 1. Epi4K patient cohorts**

Cohort	Patient types (numbers)	References
The Epilepsy Phenome/Genome Project (EPGP)	First-degree relative pairs with IGE or NFE (n = 1,500) Trios with IS (n = 250), LGS (n = 250) or malformations of cortical development (n = 250)	EPGP (2007)
Epilepsy Family Study of Columbia University	Multiplex families with three or more affected patients with IGE or NFE (n = 32)	Ottman et al. (2005); Ottman and Susser (1992)
The Epilepsy Research Centre, University of Melbourne	Multiplex families with three or more affected patients with IGE or NFE (n = 250)	Marini et al. (2004)
Quebec Familial Epilepsy Study	Multiplex families with three or more affected patients with IGE or NFE (n = 47)	Kinirons et al. (2008)
Childhood Absence Epilepsy Study (CAE) SANAD and University of Melbourne	Singletons with CAE (n = 446) Newly diagnosed singletons with 716 IGE (n = 716) or focal epilepsy (n = 1,721) prospectively followed for seizure outcome	Glauser et al. (2010) Bleck (2007); Marson et al. (2007)
EPIGEN	Singletons with IGE (n = 700) or NFE (n > 3,000)	EPIGEN (2008)

IGE, idiopathic generalized epilepsy; NFE, nonlesional focal epilepsy; IS, infantile spasms; LGS, Lennox-Gastaut syndrome; CAE, childhood absence epilepsy.

thies) by the end of year 3 and project 2 (multiplex families and pairs) by the end of year 5. Completion of project 3 (prognosis) will depend on extra funds being obtained.

To establish a clear, upfront understanding among all Epi4K investigators regarding the nature of the collaborative effort, an “Epi4K Charter” was drafted and signed by all investigators. The Charter, which can be found at <http://www.epgp.org/epi4/charter>, specifies guiding principles (including the concept of “no surprises”; i.e., that all members of Epi4K are required to present and discuss any plans that utilize Epi4K data or resources with the Epi4K Steering Committee), the organizational structure, and policies for decision making and data sharing. We also wrote an explicit Publication Policy (see <http://www.epgp.org/epi4/PubPolicy>) that provides detailed guidelines for the planning and generation of all reports that rely on information that has

resulted from the collaborative nature of Epi4K. Among other things, the Publication Policy describes the primary papers expected to result from the Epi4K projects, and it stipulates that authorship for all primary papers will be under the title “The Epi4K Consortium.” The policy also emphasizes that a goal of Epi4K is to facilitate the career development of early-stage and midstage scientists and, to that end, priority will be given to junior investigators (defined as individuals who have not yet achieved the rank of full professor or its equivalent, including senior level postdoctoral fellows) when authors are named on secondary and other papers. Furthermore, senior investigators (those at the rank of full professor or equivalent) will, in general, not be named authors on Epi4K papers unless they have had a substantive role in being individually responsible for the conceptualization or analysis of the work.

## EPI4K PROJECT 1: EPILEPTIC ENCEPHALOPATHIES

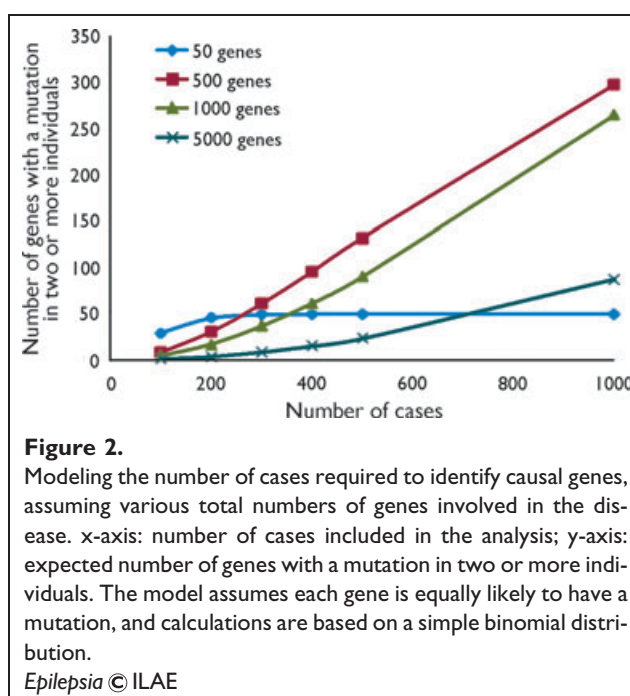
Project 1 of the Epi4K Center Without Walls will address the genetics of epileptic encephalopathies, conditions defined by severe epilepsy that are typically refractory to medication accompanied by comorbid cognitive and behavioral dysfunction. While epileptic encephalopathies are extremely diverse, the study focuses on two of the more common types: infantile spasms (IS) and Lennox-Gastaut syndrome (LGS).

Infantile spasms are among the most common of the epileptic encephalopathies with an incidence of 1 in 3,000 live births and an onset between 4 and 12 months of life (Trevathan et al., 1999; Frost & Hrachovy, 2003). IS are associated with the characteristic chaotic interictal EEG pattern of hypsarrhythmia, the sine qua non of the syndrome, and the spasms themselves are associated with an electrodecremental response. About 50–65% of IS cases have known causes such as developmental brain malformations, tuberous sclerosis complex, chromosomal syndromes (e.g., trisomy 21), and metabolic conditions (Osborne et al., 2010), reviewed in Paciorkowski et al. (2011a).

LGS, which typically has onset between 1 and 8 years, is characterized by mixed seizure types—including tonic, atonic, myoclonic, atypical absence, focal, and generalized tonic-clonic seizures—and intellectual disability (Camfield, 2011). The characteristic electroencephalography (EEG) patterns include a slow (<2.5 Hz) spike-wave pattern and generalized paroxysmal fast activity. Patients with IS and other early onset epilepsies may evolve into LGS, underscoring a likely shared etiology. As with IS, LGS may be symptomatic of underlying structural or metabolic abnormalities, but in about 25–35% of the cases, the cause remains unknown (Camfield, 2011).

In the past few years, *de novo* or inherited mutations in genes such as *ARX*, *STXBPI*, *CDKL5*, and *SCN1A* have been found in subjects with IS, although these genes seem to explain only a small percentage of the IS cases of hitherto unknown cause. Pathogenic CNVs have also been reported in children with IS (Mefford et al., 2011; Paciorkowski et al., 2011b). So far, there are far fewer established genetic etiologies for LGS. Since patients may evolve from one electroclinical syndrome to another over the course of childhood, and the genetic etiologies of the epileptic encephalopathies without known cause may be shared across syndromes, a study of both IS and LGS is likely to increase the chances for etiologic discovery. Given the relative severity of these disorders, the main hypothesis of project 1 is that a significant number of IS and LGS cases of unknown cause are due to dominant *de novo* mutations.

To undertake our gene discovery effort, we will conduct whole exome sequencing on 300+ family trios comprising patients with cryptogenic IS or LGS and their parents who do not have epilepsy (optimizing overlapping genetics by first selecting individuals who show evolution from IS to LGS).



**Figure 2.**

Modeling the number of cases required to identify causal genes, assuming various total numbers of genes involved in the disease. x-axis: number of cases included in the analysis; y-axis: expected number of genes with a mutation in two or more individuals. The model assumes each gene is equally likely to have a mutation, and calculations are based on a simple binomial distribution.

*Epilepsia* © ILAE

The probands and parents to be studied have been collected by the Epilepsy Phenome/Genome Project (EPGP) (see: <http://www.epgp.org>) and thus have detailed phenotypic data.

With the whole exome data we will screen for genes that have both *de novo* point mutations and *de novo* CNVs. We will also consider more complex genetics including genes with a combination of *de novo* and rare inherited mutations or genes with only rare inherited mutations. A recent study suggests that each new diploid genome has an estimated 70 *de novo* mutations and that only a small fraction of these are likely to fall in coding regions of the genome (~1.5%) (Roach et al., 2010). To obtain a rough idea of whether our sample size is sufficient for this approach, we can assume that IS and LGS cases are due to a mutation in any one of *N* genes, and that each of the *N* genes is equally likely to be responsible for any given case. We can then estimate the number of genes expected to carry a causal mutation in more than one case as a function of *N* (Fig. 2). This analysis suggests that if the number of genes is <500 we will be well powered to identify some of the responsible genes with our study design. Finally, although the primary discovery strategy for the IS/LGS cohort will focus on identifying *de novo* mutations affecting the same genes, we will evaluate rare inherited variants of interest in larger case control comparisons (described below), and we will sequence genes of interest in additional cases and controls.

## EPI4K PROJECT 2: MULTIPLEX FAMILIES AND PAIRS

Project 2 of the Epi4K Center without Walls will use whole genome sequencing (WGS) to identify genomic

variation that influences risk for common forms of IGE and NFE found in families with two or more affected members. The families to be studied include: (1) 1,500 pairs of first-degree relatives (sibling pairs or parent-offspring pairs) with either IGE or NFE collected in EPGP; and (2) approximately 300 families containing  $\geq 3$  individuals with IGE or NFE previously collected and phenotyped in detail by collaborating investigators. Our rationale for studying familial cases is that these cohorts are likely to be enriched for genetic influences on epilepsy, thereby greatly enhancing the likelihood of identifying the genes involved (Risch, 2001). This large, well-characterized dataset of relative pairs and multiplex epilepsy families provides an unparalleled opportunity for genetic analysis, both to detect clinically defined syndromes that have not previously been recognized, and to identify genomic variants that raise susceptibility to these specific syndromes.

One of the most challenging problems for genetic studies of the epilepsies is the extreme clinical and genetic heterogeneity of the disorder. Broadly defined by recurrent unprovoked seizures, the epilepsies encompass a wide range of different syndromes presumed to have different pathogenic mechanisms. Project 2 will therefore begin with an in-depth analysis of epilepsy phenotypes in the families to be studied. We will test the hypothesis that specific constellations of clinical epilepsy features aggregate within families and constitute “familial syndromes” that can be used to separate families into distinct subgroups likely to differ in their genetic influences.

We will then perform a comprehensive analysis of genomic variation in the families, including coding and noncoding single nucleotide variants, indels, and CNVs. In each family containing  $\geq 3$  affected individuals, we will sequence two affected individuals, selected to be as distantly related as possible. The rationale for this approach is that affected individuals within these multiplex families are likely to share the genetic variant(s) leading to their epilepsy, and distant relatives are unlikely to share as many variants by chance, so that examination of variants shared by the two distantly related affected individuals can be used as a filter for variants likely to have a pathogenic effect. To further enhance filtering, we also will genotype all affected family members for a panel of linkage markers and use the resulting information to examine whether the potentially pathogenic variants identified by sequencing are localized to genomic regions shared by all affected individuals in the family. In the EPGP first-degree relative pairs, we will perform a similar comprehensive analysis of genomic variation that influences risk for either IGE or NFE. WGS will be carried out in one individual from each of the 1,500 relative pairs (genome sequencing the second individual would provide little additional information, since the relatives are so closely related, but the second individual will be valuable for testing for the presence of proposed pathogenic variants). To improve filtering, we will also genotype the EPGP

full siblings for a linkage panel of markers and examine whether regions containing potentially pathogenic variants in the sequenced sibling are also inherited in the other sibling.

While control samples will not be sequenced directly through Epi4K, all of our association analyses will make use of the rapidly growing number of genomes that have been sequenced from a variety of sources. In particular, we anticipate having available several thousand genomes (a combination of exomes and genomes) sequenced at Duke University, and additional thousands available through public databases. This will allow us to implement both variant-based and gene-based association tests comparing the epilepsy genomes to population-control genomes.

We expect that numerous potentially disease-related variants will be identified; their likelihood of being pathogenic will be ranked by the extent of cosegregation with epilepsy within larger families, their recurrence in other families or relative pairs across the whole dataset and their predicted functional effects. Once our initial association analyses and filtering analyses have been performed, further validation of selected variants and genes of interest will be done by follow-up association studies in independent samples and functional genomic studies, both within Epi4K and with additional collaborators. We anticipate having available as many as 10,000 further samples for this effort, combining both further epilepsy samples and control samples. These samples will be drawn primarily from the EPIGEN consortium and from control samples available at Duke University and the University of Washington. Finally, we will assess the specificity of effect of identified variants for different clinically defined subsets of epilepsy and for families containing different numbers of affected individuals. In these analyses, we hypothesize that the variants that influence risk will differ between IGE and NFE, and among the syndromes identified in our analysis of familial phenotypes.

### EPI4K PROJECT 3: PROGNOSIS

The epilepsies have a highly variable prognosis, even among persons with seemingly the same type of epilepsy. Perhaps the most important aspect of epilepsy prognosis is how well a patient responds to medication, with the prevalence of pharmacoresistance in epilepsy largely unchanged by the new antiepileptic drugs (AEDs).

Project 3, which is planned as a future activity within Epi4K, will use established epilepsy cohorts with well-characterized data on seizure outcome to systematically explore the relationship between genetic variation and prognosis, focusing in particular on pharmacologic control of seizures. Our overarching hypothesis is that genetic variation influences the likelihood of response to specific AEDs. While the study design and discovery paradigms are agnostic about whether the variants that are most relevant to prognosis also influence disease predisposition, we postulate a genetically

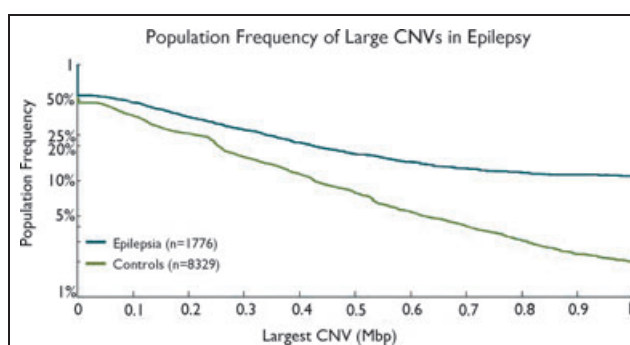
based interrelatedness of disease etiology and outcome (i.e., specific susceptibility variants increase the risk of subtypes of epilepsy that are inherently less responsive to currently available AEDs) (Johnson et al., 2003; Rogawski & Johnson, 2008).

The clinical data to be used in project 3 have been collected within established pharmacogenetic studies of epilepsy from the United Kingdom, United States, and Australia, providing a large combined cohort of newly treated epilepsy with prospectively measured seizure and drug response outcomes (see Table 1). The prospective epidemiologic design offers many advantages for the study of drug response outcomes including the facility to record uniform and unbiased patient data over time (Leschziner et al., 2006; Manolio et al., 2006). Prospective studies also mitigate many of the problematic aspects of the retrospective case-control design, such as choice of controls, recall bias, varying definitions of AED response, and potential biases in the allocation of outcome status. However, the inherent high-cost and long-term nature of prospective studies restricts the accrual of large numbers of cases at the extremes of phenotypic variation. Therefore, we will capitalize on the strengths of the collaboration to provide large numbers of appropriately phenotyped cases and include, in addition to the prospective study, a matched case-control (retrospective) design focusing on patients at the extremes of the drug-response phenotype (i.e., extreme responders and extreme treatment resistance).

Our approach to the genetic analysis of these cohorts is twofold. First, we will determine the prognostic implications of all variants identified as risk factors for epilepsy by genotyping susceptibility variants identified in Epi4K across the whole prospective epilepsy cohort ( $n = 2,000$ ). In this way, Epi4K will be able to leverage its established clinical resources not only for discovering susceptibility variants, but also for exploration of the prognostic implications of susceptibility variants. Second, we aim to exome-sequence patients at the extremes of drug-response phenotype, as well as across the prospective cohort, to identify genetic variants influencing response to medication independent of disease susceptibility.

## EPI4K PROJECT 4: COPY NUMBER VARIANTS

CNVs play an important role in the genetic risk and etiology of epilepsy. Large, recurrent CNVs as well as rare, non-recurrent CNVs have now been identified in a proportion of patients with generalized, focal, or encephalopathic epilepsies. The burden for large CNVs (5–10%) is significant when compared to controls (Fig. 3) and is comparable to those reported for children with autism (Cooper et al., 2011), albeit less so than for children with developmental delay and congenital abnormalities. Current platforms for detecting genome-wide CNV include array comparative



**Figure 3.** CNV size in patients with epilepsy (blue line) compared to population controls (green). Adapted from Cooper et al. (2011) *Epilepsia* © ILAE

genomic hybridization (CGH) and single nucleotide polymorphism (SNP) genotyping platforms. Although both have been used with great success to identify large CNVs, the ability to routinely and robustly detect CNVs <50 kb by either of these platforms is limited and largely restricted to unique regions of the genome (Alkan et al., 2011a). As a result, the role of small- to moderate-sized CNVs in disease is largely unknown and we predict will contribute significantly to the etiology of epilepsy. With the development of next generation sequencing, we now have a unique opportunity to detect the size spectrum of CNVs more comprehensively—from a single base pair to many megabases. Early studies confirm the ability to detect CNVs that are not detected by other platforms, increasing the rate of CNV discovery by more than fivefold (Hormozdiari et al., 2009; Conrad et al., 2010; Pelak et al., 2010; Mills et al., 2011). With the continued development and improvement of novel algorithms (Karakoc et al., 2011), the detection of CNVs from next-generation sequence data has improved significantly, but more work needs to be done to capture more complex variation as well as specific classes such as inversions and novel insertions. Data from Epi4K, which will include exome and genome sequences of 4,000 patients with various types of epilepsy, will be an incredibly rich resource for identifying disease-associated CNVs that are important within and across epilepsy subtypes.

The primary goal of Epi4K project 4 will be to apply novel computational algorithms to whole-exome and whole-genome data for Epi4K patients to detect epilepsy-associated CNVs. Our approach will be to focus on structural variation in protein-coding sequence of all 4,000 individuals using a combination of algorithms based on read-depth and split-read analyses. Currently, there are three basic strategies that may be applied to the detection of structural variation from sequence data: read-pair, read-depth, and split-read methods (Medvedev et al., 2010; Alkan et al., 2011b; Mills et al., 2011). Read-pair methods detect structural variation based on discordances in length and/or orientation and can, in

principle, detect insertions, deletions, duplications, and inversions. Read-depth methods detect regions of significant excess or depletion of sequence reads as surrogates of duplications and deletions, respectively. Recently developed methods taking advantage of algorithms that map to multiple locations allow copy-number variation of genes within non-unique regions to be accurately assayed (Sudmant et al., 2010). Here, genome coverage is a key aspect in the power to detect events as a function of size (Bailey et al., 2002; Alkan et al., 2009; Yoon et al., 2009). Split-read methods detect breakpoints based on a read spanning a breakpoint and, as a result, length of the sequence read as well as the sequence complexity of the breakpoint determine sensitivity (Karakoc et al., 2011). In general, maximal discovery benefits from application of read-pair, read-depth, and split-read methods on the same datasets, and we will develop appropriate pipelines for the evaluation of exome data and whole genome data and for comparison of the set of CNVs identified in each set. Potentially pathogenic CNVs and structural variants will be subjected to experimental validation and familial segregation analysis, where applicable, and to further follow-up.

## CONCLUSIONS

Epi4K builds upon the substantial work that has already been focused on understanding the genetic basis of common forms of epilepsy and epileptic encephalopathies, including the prognostic determinants of these disorders. Our approach emphasizes careful phenotyping of subjects, the increased accessibility of next-generation sequencing, state-of-the-art methods for genomic analysis, and the unique opportunities that are created by large-scale collaborations. If successful, this work will lead to advances in our ability to identify the causes of many forms of so-called idiopathic and cryptogenic epilepsy, and it will drive the development of drugs that are antiseizure, and perhaps (by focusing on the interaction between etiology and outcome) even disease-modifying or antiepileptogenic. Of course, the identification of genetic risk factors for epilepsy is only a first step in such efforts, and new research programs building on a variety of platforms (e.g., knockin and knockout mouse models, high-throughput small molecule screens, in silico cellular and network analyses), will be required to study the biologic bases of those genetic risks and to assess their translational implications. To this end, the Epi4K investigators welcome the involvement of other researchers from throughout the world who see a potential benefit in collaborating with us. Those interested are encouraged to contact us via our website (<http://www.epgp.org/epi4K>).

## ACKNOWLEDGMENTS

Supported by grants: U01NS077274, U01NS077276, U01NS077303, U01NS077364, U01NS077367, and U01NS077275 from NINDS.

## ADDITIONAL CONTRIBUTORS

Sam Berkovic, M.D., FRS.: University of Melbourne – Principal Investigator, Epi4K Administrative Core, Phenotyping and Clinical Informatics Core, and Project 2

Patrick Cossette, M.D., Ph.D.: University of Montreal – Principal Investigator, Epi4K Phenotyping and Clinical Informatics Core and Project 2

Norman Delanty M.D. FRCPI: Beaumont Hospital and Royal College of Surgeons in Ireland, Dublin – Principal Investigator, Epi4K Project 3

Dennis Dlugos, M.D.: The Children's Hospital of Philadelphia – Principal Investigator, Epi4K Phenotyping and Clinical Informatics Core

Evan Eichler, Ph.D.: University of Washington – Principal Investigator, Epi4K Sequencing, Biostatistics and Bioinformatics Core and Project 4

Michael Epstein, Ph.D.: Emory University – Principal Investigator, Epi4K Project 2

Tracy Glauser, M.D.: Cincinnati Children's Hospital Medical Center – Principal Investigator, Epi4K Project 3

David Goldstein, Ph.D.: Duke University – Principal Investigator, Epi4K Administrative Core, Sequencing, Biostatistics and Bioinformatics Core, and Project 3

Erin Heinzen, Ph.D.: Duke University – Principal Investigator, Epi4K Sequencing, Biostatistics and Bioinformatics Core

Michael R. Johnson, M.D., Ph.D.: Imperial College, London – Principal Investigator, Epi4K Project 3

Ruben Kuzniecky, M.D.: New York University – Principal Investigator, Epi4K Phenotyping and Clinical Informatics Core

Daniel Lowenstein, M.D.: University of California, San Francisco – Principal Investigator, Epi4K Administrative Core and Phenotyping and Clinical Informatics Core

Tony Marson, M.D.: University of Liverpool – Principal Investigator, Epi4K Project 3

Heather Mefford, Ph.D.: University of Washington – Principal Investigator, Epi4K Sequencing, Biostatistics and Bioinformatics Core, Project 1, and Project 4

Terence O'Brien, M.D., FRACP, FRCPE: University of Melbourne – Principal Investigator, Epi4K Project 3

Ruth Ottman, Ph.D.: Columbia University – Principal Investigator, Epi4K Phenotyping and Clinical Informatics Core and Project 2

Ann Poduri, M.D.: Children's Hospital, Boston – Co-Investigator, Epi4K Project 1

Ingrid Scheffer M.B.B.S., Ph.D., FRACP: Florey Neuroscience Institutes, University of Melbourne – Principal Investigator, Epi4K Phenotyping and Clinical Informatics Core and Project 1

Elliott Sherr, M.D., Ph.D.: University of California, San Francisco – Principal Investigator, Epi4K Project 1

Kevin Shianna, Ph.D.: Duke University – Principal Investigator, Epi4K Sequencing, Biostatistics and Bioinformatics Core

## DISCLOSURE

Author Evan E. Eichler is a scientific advisory board member of Pacific Biosciences, Inc. and SynapDx Corp. The remaining authors have no conflicts of interest.

We confirm that we have read the Journal's position on issues involved in ethical publication and affirm that this report is consistent with those guidelines.



## REFERENCES

- Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE. (2009) Personalized copy number and segmental duplication maps using next-generation sequencing. *Nat Genet* 41:1061–1067.
- Alkan C, Coe BP, Eichler EE. (2011a) Genome structural variation discovery and genotyping. *Nat Rev Genet* 12:363–376.
- Alkan C, Sajjadian S, Eichler EE. (2011b) Limitations of next-generation genome sequence assembly. *Nat Methods* 8:61–65.
- Annegers JF, Hauser WA, Anderson VE, Kurland LT. (1982) The risks of seizure disorders among relatives of patients with childhood onset epilepsy. *Neurology* 32:174–179.
- Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE. (2002) Recent segmental duplications in the human genome. *Science* 297:1003–1007.
- Barrett JC, Cardon LR. (2006) Evaluating coverage of genome-wide association studies. *Nat Genet* 38:659–662.
- Berkovic SF, Howell RA, Hay DA, Hopper JL. (1998) Epilepsies in twins: genetics of the major epilepsy syndromes. *Ann Neurol* 43:435–445.
- Bertrand D, Elmslie F, Hughes E, Trounce J, Sander T, Bertrand S, Steinlein OK. (2005) The CHRN2 mutation I312M is associated with epilepsy and distinct memory deficits. *Neurobiol Dis* 20:799–804.
- Bisulli F, Tinuper P, Scudellaro E, Naldi I, Bagattin A, Avoni P, Michelucci R, Nobile C. (2004) A de novo LGII mutation in sporadic partial epilepsy with auditory features. *Ann Neurol* 56:455–456.
- Bleck TP. (2007) Valproate was more effective than lamotrigine and better tolerated than topiramate in generalized or unclassified epilepsy. *ACP J Club* 147:74–75.
- Camfield PR. (2011) Definition and natural history of Lennox–Gastaut syndrome. *Epilepsia* 52(Suppl. 5):3–9.
- Cardon LR, Bell JL. (2001) Association study designs for complex diseases. *Nat Rev Genet* 2:91–99.
- Cavalleri GL, Walley NM, Soranzo N, Mulley J, Doherty CP, Kapoor A, Depondt C, Lynch JM, Scheffer IE, Heils A, Gehrman A, Kinirons P, Gandhi S, Satishchandra P, Wood NW, Anand A, Sander T, Berkovic SF, Delanty N, Goldstein DB, Sisodiya SM. (2007a) A multicenter study of BRD2 as a risk factor for juvenile myoclonic epilepsy. *Epilepsia* 48:706–712.
- Cavalleri GL, Weale ME, Shianna KV, Singh R, Lynch JM, Grinton B, Szoek C, Murphy K, Kinirons P, O'Rourke D, Ge D, Depondt C, Claeys KG, Pandolfo M, Gumbs C, Walley N, McNamara J, Mulley JC, Linney KN, Sheffield LJ, Radtke RA, Tate SK, Chisoe SL, Gibson RA, Hosford D, Stanton A, Graves TD, Hanna MG, Eriksson K, Kantanen AM, Kalviainen R, O'Brien TJ, Sander JW, Duncan JS, Scheffer IE, Berkovic SF, Wood NW, Doherty CP, Delanty N, Sisodiya SM, Goldstein DB. (2007b) Multicentre search for genetic susceptibility loci in sporadic epilepsy syndrome and seizure types: a case–control study. *Lancet Neurol* 6:970–980.
- Cirulli ET, Goldstein DB. (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11:415–425.
- Claes L, Del-Favero J, Ceulemans B, Lagae L, Van Broeckhoven C, De Jonghe P. (2001) De Novo mutations in the sodium-channel gene SCN1A cause severe myoclonic epilepsy of infancy. *Am J Hum Genet* 68:1327–1332.
- Claes LR, Ceulemans B, Audenaert D, Deprez L, Jansen A, Hasaerts D, Weckx S, Claeys KG, Del-Favero J, Van Broeckhoven C, De Jonghe P. (2004) De novo KCNQ2 mutations in patients with benign neonatal seizures. *Neurology* 63:2155–2158.
- Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, MacArthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. (2010) Origins and functional impact of copy number variation in the human genome. *Nature* 464:704–712.
- Cooper GM, Coe BP, Girirajan S, Rosenfeld JA, Vu TH, Baker C, Williams C, Stalker H, Hamid R, Hannig V, Abdel-Hamid H, Bader P, McCracken E, Niyazov D, Leppig K, Thiess H, Hummel M, Alexander N, Gorski J, Kussmann J, Shashi V, Johnson K, Rehder C, Ballif BC, Shaffer LG, Eichler EE. (2011) A copy number variation morbidity map of developmental delay. *Nat Genet* 43:838–846.
- Corbett MA, Bahlo M, Jolly L, Afawi Z, Gardner AE, Oliver KL, Tan S, Coffey A, Mulley JC, Dibbens LM, Simri W, Shalata A, Kivity S, Jackson GD, Berkovic SF, Geck J. (2010) A focal epilepsy and intellectual disability syndrome is due to a mutation in TBC1D24. *Am J Hum Genet* 87:372–375.
- Corbett MA, Schwake M, Bahlo M, Dibbens LM, Lin M, Gandolfo LC, Vears DF, O'Sullivan JD, Robertson T, Bayly MA, Gardner AE, Vlaar AM, Korenke GC, Bloem BR, de Coo IF, Verhagen JMA, Lehesjoki AE, Geck J, Berkovic SF. (2011) A mutation in the golgi Qb-SNARE gene GOSR2 causes progressive myoclonus epilepsy with early ataxia. *Am J Hum Genet* 88:657–663.
- Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, Roses AD, Haines JL, Pericak-Vance MA. (1993) Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science* 261:921–923.
- Corey LA, Berg K, Pellock JM, Solaas MH, Nance WE, DeLorenzo RJ. (1991) The occurrence of epilepsy and febrile seizures in Virginian and Norwegian twins. *Neurology* 41:1433–1436.
- de Kovel CG, Pinto D, de Haan GJ, Kasteleijn-Nolst Trenite DG, Lindhout D, Koelman BP. (2007) Association analysis of BRD2 (RING3) and epilepsy in a Dutch population. *Epilepsia* 48:2191–2192.
- de Kovel CG, Trucks H, Helbig I, Mefford HC, Baker C, Leu C, Kluck C, Muhle H, von Spiczak S, Ostertag P, Obermeier T, Kleefuss-Lie AA, Hallmann K, Steffens M, Gaus V, Klein KM, Hamer HM, Rosenow F, Brilstra EH, Trenite DK, Swinkels ME, Weber YG, Unterberger I, Zimprich F, Urak L, Feucht M, Fuchs K, Moller RS, Hjalgrim H, De Jonghe P, Suls A, Ruckert IM, Wichmann HE, Franke A, Schreiber S, Nurnberg P, Elger CE, Lerche H, Stephani U, Koelman BP, Lindhout D, Eichler EE, Sander T. (2010) Recurrent microdeletions at 15q11.2 and 16p13.11 predispose to idiopathic generalized epilepsies. *Brain* 133:23–32.
- Depienne C, Trouillard O, Saint-Martin C, Gourfinkel-An I, Bouteiller D, Carpentier V, Keren B, Abert B, Gautier A, Baulac S, Arzimanoglou A, Cazeneuve C, Nabbout R, LeGuern E. (2009) Spectrum of SCN1A gene mutations associated with Dravet syndrome: analysis of 333 patients. *J Med Genet* 46:183–191.
- DHHS. (2010) Centers without Walls for Collaborative Research in the Epilepsies: Genetics and Genomics of Human Epilepsies (U01). Available at: <http://grants.nih.gov/grants/guide/rfa-files/RFA-NS-11-007.html> [Accessed March 1, 2012].
- Dibbens LM, Mullen S, Helbig I, Mefford HC, Bayly MA, Bellows S, Leu C, Trucks H, Obermeier T, Wittig M, Franke A, Caglayan H, Yapici Z, Sander T, Eichler EE, Scheffer IE, Mulley JC, Berkovic SF. (2009) Familial and sporadic 15q13.3 microdeletions in idiopathic generalized epilepsy: precedent for disorders with complex inheritance. *Hum Mol Genet* 18:3626–3631.
- EPGP. (2007) Epilepsy Phenome/Genome Project. Available at: <http://www.epgp.org> [Accessed March 1, 2012].
- EPICURE. (2007) Functional genomics and neurobiology of epilepsy: a basis for new therapeutic strategies. Available at: <http://www.epicure-project.eu> [Accessed March 1, 2012].
- EPIGEN. (2008) EPIGEN, An international consortium dedicated to tackling epilepsy genetics. Available at: <http://www.epilepsygenetics.eu/index.html> [Accessed March 1, 2012].
- Frost FD, Hrachovy RA. (2003) *Infantile spasms*. Kluwer Academic Publishers, Boston, MA.
- Glauser TA, Cnaan A, Shinnar S, Hirtz DG, Dlugos D, Masur D, Clark PO, Capparelli EV, Adamson PC. (2010) Ethosuximide, valproic acid, and lamotrigine in childhood absence epilepsy. *N Engl J Med* 362:790–799.
- Goldstein DB. (2009) Common genetic variation and human traits. *N Engl J Med* 360:1696–1698.
- Greenberg DA, Cayanis E, Strug L, Marathe S, Durner M, Pal DK, Alvin GB, Klotz I, Dicker E, Shinnar S, Bromfield EB, Resor S, Cohen J, Moshe SL, Harden C, Kang H. (2005) Malic enzyme 2 may underlie susceptibility to adolescent-onset idiopathic generalized epilepsy. *Am J Hum Genet* 76:139–146.
- Guo Y, Baum LW, Sham PC, Wong V, Ng PW, Lui CH, Sin NC, Tsoi TH, Tang CS, Kwan JS, Yip BH, Xiao SM, Thomas GN, Lau YL, Yang W, Cherny SS, Kwan P. (2012) Two-stage genome-wide association study

- identifies variants in CAMSAP1L1 as susceptibility loci for epilepsy in Chinese. *Hum Mol Genet* 21:1184–1189.
- Heinzen EL, Radtke RA, Urban TJ, Cavalleri GL, Depondt C, Need AC, Walley NM, Nicoletti P, Ge D, Catarino CB, Duncan JS, Kasperaviciute D, Tate SK, Caboclo LO, Sander JW, Clayton L, Linney KN, Shianna KV, Gumbs CE, Smith J, Cronin KD, Maia JM, Doherty CP, Pandolfo M, Leppert D, Middleton LT, Gibson RA, Johnson MR, Matthews PM, Hosford D, Kalviainen R, Eriksson K, Kantanen AM, Dorn T, Hansen J, Kramer G, Steinhoff BJ, Wieser HG, Zumsteg D, Ortega M, Wood NW, Huxley-Jones J, Mikati M, Gallentine WB, Husain AM, Buckley PG, Stallings RL, Podgoreanu MV, Delanty N, Sisodiya SM, Goldstein DB. (2010) Rare deletions at 16p13.11 predispose to a diverse spectrum of sporadic epilepsy syndromes. *Am J Hum Genet* 86:707–718.
- Helbig I, Scheffer IE, Mulley JC, Berkovic SF. (2008) Navigating the channels and beyond: unravelling the genetics of the epilepsies. *Lancet Neurol* 7:231–245.
- Helbig I, Mefford HC, Sharp AJ, Guipponi M, Fichera M, Franke A, Muhle H, de Kovel C, Baker C, von Spiczak S, Kron KL, Steinich I, Kleefuss-Lie AA, Leu C, Gaus V, Schmitz B, Klein KM, Reif PS, Rosenow F, Weber Y, Lerche H, Zimprich F, Urak L, Fuchs K, Feucht M, Genton P, Thomas P, Visscher F, de Haan GJ, Moller RS, Hjalgrim H, Luciano D, Wittig M, Nothnagel M, Elger CE, Nurnberg P, Romano C, Malafosse A, Koelman BP, Lindhout D, Stephani U, Schreiber S, Eichler EE, Sander T. (2009) 15q13.3 microdeletions increase risk of idiopathic generalized epilepsy. *Nat Genet* 41:160–162.
- Hindorf LA, MacArthur J (European Bioinformatics Institute), Wise A, Junkins HA, Hall PN, Klemm AK, Manolio TA (n.d.). A catalog of published genome-wide association studies. Available at: <http://www.genome.gov/GWAStudies/> [Accessed January 15, 2012].
- Hormozdiari F, Alkan C, Eichler EE, Sahinalp SC. (2009) Combinatorial algorithms for structural variation detection in high-throughput sequenced genomes. *Genome Res* 19:1270–1278.
- Ishii A, Fukuma G, Uehara A, Miyajima T, Makita Y, Hamachi A, Yasukochi M, Inoue T, Yasumoto S, Okada M, Kaneko S, Mitsudome A, Hirose S. (2009) A de novo KCNQ2 mutation detected in non-familial benign neonatal convulsions. *Brain Dev* 31:27–33.
- Johnson MR, Milne RL, Torn-Broers Y, Hopper JL, Scheffer IE, Berkovic SF. (2003) A twin study of genetic influences on epilepsy outcome. *Twin Res* 6:140–146.
- Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE. (2011) Detection of structural variants and indels within exome data. *Nat Methods* 9:176–178.
- Kasperaviciute D, Catarino CB, Heinzen EL, Depondt C, Cavalleri GL, Caboclo LO, Tate SK, Jamnadas-Khoda J, Chinthapalli K, Clayton LM, Shianna KV, Radtke RA, Mikati MA, Gallentine WB, Husain AM, Alhusaini S, Leppert D, Middleton LT, Gibson RA, Johnson MR, Matthews PM, Hosford D, Heuser K, Amos L, Ortega M, Zumsteg D, Wieser HG, Steinhoff BJ, Kramer G, Hansen J, Dorn T, Kantanen AM, Gjerstad L, Peuralinna T, Hernandez DG, Eriksson KJ, Kalviainen RK, Doherty CP, Wood NW, Pandolfo M, Duncan JS, Sander JW, Delanty N, Goldstein DB, Sisodiya SM. (2010) Common genetic variation and susceptibility to partial epilepsies: a genome-wide association study. *Brain* 133:2136–2147.
- Kinirons P, Rabinowitz D, Gravel M, Long J, Winawer M, Senechal G, Ottman R, Cossette P. (2008) Phenotypic concordance in 70 families with IGE-implications for genetic studies of epilepsy. *Epilepsy Res* 82:21–28.
- Kjeldsen MJ, Corey LA, Christensen K, Friis ML. (2003) Epileptic seizures and syndromes in twins: the importance of genetic factors. *Epilepsy Res* 55:137–146.
- Klassen T, Davis C, Goldman A, Burgess D, Chen T, Wheeler D, McPherson J, Bourquin T, Lewis L, Villasana D, Morgan M, Muzny D, Gibbs R, Noebels J. (2011) Exome sequencing of ion channel genes reveals complex profiles confounding personal risk assessment in epilepsy. *Cell* 145:1036–1048.
- Layouni S, Buresi C, Thomas P, Malafosse A, Dogui M. (2010) BRD2 and TAP-1 genes and juvenile myoclonic epilepsy. *Neurol Sci* 31:53–56.
- Lenzen KP, Heils A, Lorenz S, Hempelmann A, Sander T. (2005) Association analysis of malic enzyme 2 gene polymorphisms with idiopathic generalized epilepsy. *Epilepsia* 46:1637–1641.
- Leschziner G, Jorgensen AL, Andrew T, Pirmohamed M, Williamson PR, Marson AG, Coffey AJ, Middleditch C, Rogers J, Bentley DR, Chadwick DW, Balding DJ, Johnson MR. (2006) Clinical factors and ABCB1 polymorphisms in prediction of antiepileptic drug response: a prospective cohort study. *Lancet Neurol* 5:668–676.
- Lohoff FW, Ferraro TN, Dahl JP, Hildebrandt MA, Scattergood TM, O'Connor MJ, Sperling MR, Dlugos DJ, Berrettini WH, Buono RJ. (2005) Lack of association between variations in the brain-derived neurotrophic factor (BDNF) gene and temporal lobe epilepsy. *Epilepsy Res* 66:59–62.
- Lorenz S, Taylor KP, Gehrmann A, Becker T, Muhle H, Gresch M, Tauer U, Sander T, Stephani U. (2006) Association of BRD2 polymorphisms with photoparoxysmal response. *Neurosci Lett* 400:135–139.
- Manolio TA, Bailey-Wilson JE, Collins FS. (2006) Genes, environment and the value of prospective cohort studies. *Nat Rev Genet* 7:812–820.
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A, Cho JH, Guttman AE, Kong A, Kruglyak L, Mardis E, Rotimi CN, Slatkin M, Valle D, Whittemore AS, Boehnke M, Clark AG, Eichler EE, Gibson G, Haines JL, Mackay TF, McCarroll SA, Visscher PM. (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753.
- Marini C, Scheffer IE, Crossland KM, Grinton BE, Phillips FL, McMahon JM, Turner SJ, Dean JT, Kivity S, Mazarib A, Neufeld MY, Korczyn AD, Harkin LA, Dibbens LM, Wallace RH, Mulley JC, Berkovic SF. (2004) Genetic architecture of idiopathic generalized epilepsy: clinical genetic analysis of 55 multiplex families. *Epilepsia* 45:467–478.
- Marson AG, Al-Kharusi AM, Alwaidh M, Appleton R, Baker GA, Chadwick DW, Cramp C, Cockerell OC, Cooper PN, Doughty J, Eaton B, Gamble C, Gouling PJ, Howell SJ, Hughes A, Jackson M, Jacoby A, Kellett M, Lawson GR, Leach JP, Nicolaides P, Roberts R, Shackley P, Shen J, Smith DF, Smith PE, Smith CT, Vanoli A, Williamson PR. (2007) The SANAD study of effectiveness of carbamazepine, gabapentin, lamotrigine, oxcarbazepine, or topiramate for treatment of partial epilepsy: an unblinded randomised controlled trial. *Lancet* 369:1000–1015.
- Medvedev P, Fiume M, Dzamba M, Smith T, Brudno M. (2010) Detecting copy number variation with mated short reads. *Genome Res* 20:1613–1622.
- Mefford HC, Mulley JC. (2010) Genetically complex epilepsies, copy number variants and syndrome constellations. *Genome Med* 2:71.
- Mefford HC, Muhle H, Ostertag P, von Spiczak S, Buysse K, Baker C, Franke A, Malafosse A, Genton P, Thomas P, Gurnett CA, Schreiber S, Bassuk AG, Guipponi M, Stephani U, Helbig I, Eichler EE. (2010) Genome-wide copy number variation in epilepsy: novel susceptibility loci in idiopathic generalized and focal epilepsies. *PLoS Genet* 6:e1000962.
- Mefford HC, Yendle SC, Hsu C, Cook J, Geraghty E, McMahon JM, Eeg-Olofsson O, Sadleir LG, Gill D, Ben-Zeev B, Lerman-Sagie T, Mackay M, Freeman JL, Andermann E, Pelakanos JT, Andrews I, Wallace G, Eichler EE, Berkovic SF, Scheffer IE. (2011) Rare copy number variants are an important cause of epileptic encephalopathies. *Ann Neurol* 70:974–985.
- Michelucci R, Mecarelli O, Bovo G, Bisulli F, Testoni S, Striano P, Striano S, Tinuper P, Nobile C. (2007) A de novo LGI1 mutation causing idiopathic partial epilepsy with telephone-induced seizures. *Neurology* 68:2150–2151.
- Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheatham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HY, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemes J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP, Stutz AM, Urban AE, Walker JA, Wu J, Zhang Y, Zhang ZD, Batzer MA, Ding L, Marth GT, McVean G, Sebat J, Snyder M, Wang J, Eichler EE, Gerstein MB, Hurler ME, Lee C, McCarroll SA, Korb JO. (2011) Mapping copy number variation by population-scale genome sequencing. *Nature* 470:59–65.
- Mulley JC, Scheffer IE, Petrou S, Berkovic SF. (2003) Channelopathies as a genetic cause of epilepsy. *Curr Opin Neurol* 16:171–176.
- Mulley JC, Scheffer IE, Harkin LA, Berkovic SF, Dibbens LM. (2005) Susceptibility genes for complex epilepsy. *Hum Mol Genet* 14 Spec No. 2:R243–249.
- Need AC, Goldstein DB. (2010) Whole genome association studies in complex diseases: where do we stand? *Dialogues Clin Neurosci* 12:37–46.
- Ng SB, Bigham AW, Buckingham KJ, Hannibal MC, McMillin MJ, Gildersleeve HI, Beck AE, Tabor HK, Cooper GM, Mefford HC, Lee C,

- Turner EH, Smith JD, Rieder MJ, Yoshiura K, Matsumoto N, Ohta T, Niikawa N, Nickerson DA, Bamshad MJ, Shendure J. (2010a) Exome sequencing identifies MLL2 mutations as a cause of Kabuki syndrome. *Nat Genet* 42:790–793.
- Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J, Bamshad MJ. (2010b) Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 42:30–35.
- Osborne JP, Lux AL, Edwards SW, Hancock E, Johnson AL, Kennedy CR, Newton RW, Verity CM, O'Callaghan FJ. (2010) The underlying etiology of infantile spasms (West syndrome): information from the United Kingdom Infantile Spasms Study (UKISS) on contemporary causes and their classification. *Epilepsia* 51:2168–2174.
- Ottman R. (2005) Analysis of genetically complex epilepsies. *Epilepsia* 46(Suppl. 10):7–14.
- Ottman R, Susser M. (1992) Data collection strategies in genetic epidemiology: The Epilepsy Family Study of Columbia University. *J Clin Epidemiol* 45:721–727.
- Ottman R, Annegers JF, Risch N, Hauser WA, Susser M. (1996a) Relations of genetic and environmental factors in the etiology of epilepsy. *Ann Neurol* 39:442–449.
- Ottman R, Lee JH, Risch N, Hauser WA, Susser M. (1996b) Clinical indicators of genetic susceptibility to epilepsy. *Epilepsia* 37:353–361.
- Ottman R, Berenson K, Barker-Cummings C. (2005) Recruitment of families for genetic studies of epilepsy. *Epilepsia* 46:290–297.
- Ottman R, Hirose S, Jain S, Lerche H, Lopes-Cendes I, Noebels JL, Serratos J, Zara F, Scheffer IE. (2010) Genetic testing in the epilepsies – report of the ILAE Genetics Commission. *Epilepsia* 51:655–670.
- Paciorkowski AR, Thio LL, Dobyns WB. (2011a) Genetic and biologic classification of infantile spasms. *Pediatr Neurol* 45:355–367.
- Paciorkowski AR, Thio LL, Rosenfeld JA, Gajecka M, Gurnett CA, Kulkarni S, Chung WK, Marsh ED, Gentile M, Reggin JD, Wheless JW, Balasubramanian S, Kumar R, Christian SL, Marini C, Guerrini R, Maltsev N, Shaffer LG, Dobyns WB. (2011b) Copy number variants and infantile spasms: evidence for abnormalities in ventral forebrain development and pathways of synaptic function. *Eur J Hum Genet* 19:1238–1245.
- Pal DK, Evgrafov OV, Tabares P, Zhang F, Durner M, Greenberg DA. (2003) BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. *Am J Hum Genet* 73:261–270.
- Pelak K, Shianna KV, Ge D, Maia JM, Zhu M, Smith JP, Cirulli ET, Fellay J, Dickson SP, Gumbs CE, Heinzen EL, Need AC, Ruzzo EK, Singh A, Campbell CR, Hong LK, Lornsen KA, McKenzie AM, Sobreira NL, Hoover-Fong JE, Milner JD, Ottman R, Haynes BF, Goedert JJ, Goldstein DB. (2010) The characterization of twenty sequenced human genomes. *PLoS Genet* 6:e1001111.
- Pericak-Vance MA, Bebout JL, Gaskell PC Jr, Yamaoka LH, Hung W-Y, Alberts MJ, Walker AP, Bartlett RJ, Haynes CA, Welsh KA, Earl NL, Heyman A, Clark CM, Roses AD. (1991) Linkage studies in familial Alzheimer disease: evidence for chromosome 19 linkage. *Am J Hum Genet* 48:1034–1050.
- Phillips HA, Marini C, Scheffer IE, Sutherland GR, Mulley JC, Berkovic SF. (2000) A de novo mutation in sporadic nocturnal frontal lobe epilepsy. *Ann Neurol* 48:264–267.
- Poduri A, Lowenstein D. (2011) Epilepsy genetics – past, present, and future. *Curr Opin Genet Dev* 21:325–332.
- Prasad AN, Prasad C. (2008) Genetic influences on the risk for epilepsy. In Pellock JM, Bourgeois BFD, Dodson WE et al. (eds) *Pediatric epilepsy*. Demos, New York, pp. 117–134.
- Reid CA, Berkovic SF, Petrou S. (2009) Mechanisms of human inherited epilepsies. *Prog Neurobiol* 87:41–57.
- Risch N. (2001) Implications of multilocus inheritance for gene–disease association studies. *Theor Popul Biol* 60:215–220.
- Roach JC, Glusman G, Smit AF, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M, Shendure J, Drmanac R, Jorde LB, Hood L, Galas DJ. (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–639.
- Rogawski MA, Johnson MR. (2008) Intrinsic severity as a determinant of antiepilepsy drug refractoriness. *Epilepsy Curr* 8:127–130.
- Saunders AM, Strittmatter WJ, Schmechel D, George-Hyslop PH, Pericak-Vance MA, Joo SH, Rosi BL, Gusella JF, Crapper-MacLachlan DR, Alberts MJ, Hulette C, Crain B, Goldgaber D, Roses AD. (1993) Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43:1467–1472.
- Scheffer IE, Berkovic SF. (2010) Copy number variants – an unexpected risk factor for the idiopathic generalized epilepsies. *Brain* 133:7–8.
- Sisodiya SM, Mefford HC. (2011) Genetic contribution to common epilepsies. *Curr Opin Neurol* 24:140–145.
- Sobreira NL, Cirulli ET, Avramopoulos D, Wohler E, Oswald GL, Stevens EL, Ge D, Shianna KV, Smith JP, Maia JM, Gumbs CE, Pevsner J, Thomas G, Valle D, Hoover-Fong JE, Goldstein DB. (2010) Whole-genome sequencing of a single proband together with linkage analysis identifies a Mendelian disease gene. *PLoS Genet* 6:e1000991.
- Stenson PD. (2008). Human Gene Mutation Database. Available at: <http://www.hgmd.org> [Accessed February 21, 2012].
- Stenson PD, Mort M, Ball EV, Howells K, Phillips AD, Thomas NS, Cooper DN. (2009) The Human Gene Mutation Database: 2008 update. *Genome Med* 1:13.
- Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, 1000 Genomes Project, Eichler EE. (2010) Diversity of human copy number variation and multicopy genes. *Science* 330:641–646.
- Tan NC, Berkovic SF. (2010) The Epilepsy Genetic Association Database (epiGAD): analysis of 165 genetic association studies, 1996–2008. *Epilepsia* 51:686–689.
- Tomlinson IP, Bodmer WF. (1995) The HLA system and the analysis of multifactorial genetic disease. *Trends Genet* 11:493–498.
- Trevathan E, Murphy CC, Yeargin-Allsopp M. (1999) The descriptive epidemiology of infantile spasms among Atlanta children. *Epilepsia* 40:748–751.
- Vadlamudi L, Andermann E, Lombroso CT, Schachter SC, Milne RL, Hopper JL, Andermann F, Berkovic SF. (2004) Epilepsy in twins: insights from unique historical data of William Lennox. *Neurology* 62:1127–1133.
- Veeramah KR, O'Brien JE, Meisler MH, Cheng X, Dib-Hajj SD, Waxman SG, Talwar D, Girirajan S, Eichler EE, Restifo LL, Erickson RP, Hammer MF. (2012) De novo pathogenic SCN8A mutation identified by whole-genome sequencing of a family quartet affected by infantile epileptic encephalopathy and SUDEP. *Am J Hum Genet* 90:502–510.
- Yoon S, Xuan Z, Makarov V, Ye K, Sebat J. (2009) Sensitive and accurate detection of copy number variants using read depth of coverage. *Genome Res* 19:1586–1592.