Genomics 108 (2016) 31–36

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

A genomic view on epilepsy and autism candidate genes

Kamel Jabbari *, Peter Nürnberg

Cologne Center for Genomics, University of Cologne, Cologne, Germany

ARTICLE INFO

Article history:
Received 27 August 2015
Received in revised form 15 December 2015
Accepted 1 January 2016
Available online 7 January 2016

Keywords:
Brain
Chromatin
Isochores
Replication timing
Lamina

ABSTRACT

Epilepsy is a common complex disorder most frequently associated with psychiatric and neurological diseases. Massive parallel sequencing of individual or cohort genomes and exomes led the identification of several disease associated genes. We review here the candidate genes in epilepsy genetics with focus on exome and gene panel data. Together with the examination of brain expressed genes and post synaptic proteome the results show that: (1) Non-metabolic epilepsies and autism candidate genes tend to be AT-rich and (2) large transcript size and local AT-richness are characteristic features of genes involved in developmental brain disorders and synaptic functions. These results point to the preferential location of core epilepsy and autism candidate genes in late replicating, GC-poor chromosomal regions (isochores). These results indicate that the genomic alterations leading to some brain disorders are confined to responsive chromatin areas harboring brain critical genes.

1. Introduction

Epilepsy and autism spectrum disorder (ASD) were early recognized as neurological diseases. They co-occur in approximately 30% of individuals with either ASD or epilepsy [25]. The consensus emerging from studies on ASD and epilepsy is that the same brain pathology accounts for the majority of children with co-occurring ASD and epilepsy.

Epilepsy is one of the most common complex neurological disorders in humans with up to 2% of people being diagnosed with epilepsy at some point in their life [5,49]. Worldwide, about 65 million people are estimated to have epilepsy [72]. Epilepsy is categorized according to what the primary cause is taught to be, although many disorders cannot fit straightforwardly to one given category. Seizures, the core symptom of epilepsy, can be the consequence of insults such as damage to the brain from injury or a microbial infection [28,50], otherwise a genetic cause or a separate disorder effect can be considered [8]. For recent reviews on epilepsy genetics, [see 48,67,80].

Depending on the mode of transmission, genetic epilepsies and ASD can be categorized into: (1) Mendelian, where a single gene is the major determinant and (2) non-Mendelian or complex, which involve many genes and/or epigenetic factors. Epilepsies are a heterogeneous group of disorders; specific genes have been linked to a few rare diffuse brain disorders accompanied with seizures, but the possible cause for Genetic Generalized Epilepsy (that includes entities such as juvenile myoclonic epilepsy and childhood absence epilepsy) was still missing till the identification of mutations in the gene of the acetylcholine receptor α4 subunit [69]. *CHRNA* was found to be associated with autosomal dominant nocturnal frontal lobe epilepsy. This discovery was achieved by linkage analysis sustained by large pedigree and Sanger sequencing. Other voltage- or ligand-gated ion channels, known to play roles in brain and especially in neuronal activity, led to the identification of epilepsy genes involved in voltage-gated potassium channels like *KCNQ2* [10,63], its paralog *KCNQ3* [13] and the voltage-gated sodium (Na+)-channel a1 subunit gene (*SCN1B*) [75]. The first mutations linked to Dravet syndrome, also known as Severe Myoclonic Epilepsy of Infancy, were detected in 2001 in a paralog of *SCN1B*, namely, *SCN1A* [14,81]. High-throughput sequencing and microarray technologies led later to the identification of further potential epilepsy genes. For instance, large consortia initiatives such as Epi4k [2], which enrolled 1500 families including 264 trios. In addition to the detection of known and unknown risk factors, Allen et al. [2] also found significant overlap between their protein–protein interactions network and the autism spectrum disorder (ASD) and intellectual disability networks, as one might expect from the observation that epilepsy is the medical condition most highly associated with autism [47].

ASD is a neurodevelopmental disease that comprises the severe end of the autism spectrum disorders, that include Asperger syndrome and Rett syndrome. Because over 70% of individuals with autism have intellectual disability, while in epilepsy it occurs in at least 25% [73,77], a parallel analysis of autism candidate genes may be useful as several biological pathways are expected to overlap [31,47]. More recently, insights from exome sequencing proved to be key in the understanding of the genetic etiology of ASD [20,82]. Many of the genes implicated encode proteins for synaptic formation, transcriptional regulation and chromatin-remodeling pathways. They also include voltage-gated ion channels and histone-modifying enzymes, as well as FMRP-associated

* Corresponding author at: Max Planck Institute For Biology Of Ageing, Cologne, Germany.
E-mail address: kjabbari@uni-koeln.de (K. Jabbari).

http://dx.doi.org/10.1016/j.ygeno.2016.01.001
0888-7543/

genes and embryonically expressed genes. Notice that fragile X syndrome is associated with epilepsy in 10%–20% of cases [36].

In addition to base-pair substitutions or single nucleotide variants and small frame shift mutations, double strand break repair can lead to chromosomal DNA loss/gain. Such copy number variations (CNVs) are major sources of genomic disorders. It is believed that CNVs account for a major proportion of human genetic variations and have an important role in genetic susceptibility to common disease and especially neuropsychiatric disorders [44]. Genome-wide surveys have demonstrated that rare CNVs altering genes in neuron-developmental pathways are implicated in epilepsy, autism spectrum disorder and schizophrenia [11]. From an evolutionary standpoint, genomic disorders associated with CNVs (dosage imbalances) appear to be highly penetrant and under strong negative selection [26,74].

In this study, we make use of the available data from different specialized resources and general databases with the aim of analyzing basic features such as gene size, base composition and pathway enrichment of candidate epilepsy and autism genes. We further compare the compositional properties of these genes to those expressed in the brain and with the postsynaptic proteome.

2. Material and methods

We first collected all genes associated with epilepsy that resulted from exome sequencing by mining the literature using PubMed with the key words: "epilepsy, gene and exome". Another list of epilepsy associated genes was obtained by searching ClinVar [38] using "epilepsy" as the phenotype and "pathogenic" as the feature. Autism candidate genes were obtained from the core autism candidate list (http://autismkb.cbi.pku.edu.cn/), which encompass high confident genes. For consolidation purposes we integrated other published datasets that will be mentioned in the next section. Brain expressed genes that are under negative selection are taken from [74]. Post-synaptic proteome was obtained from [6]. Ensembl annotation [23] was used to collect a non redundant set of human genes with HGCN (HUGO Gene Nomenclature Committee) protein coding entries (19,202 genes).

Gene GC percent and size are calculated for the total exon plus intron gene sequence (UTRs are not included). To test for differences in GC% distributions between the analyzed gene sets, we performed a Kruskal–Wallis rank sum test using the *kruskal.test* function in the R Project for Statistical Computing (http://www.r-project.org/). Isochores chromosomal coordinates are from [16].

Gene Ontology pathway enrichment was performed through http://cpdb.molgen.mpg.de/; for each of the gene sets examined, a *p*-value for enrichment was calculated according to the hypergeometric test based on the number of genes present in both the predefined set and our specified gene list. Only highly significant ($p < 4.55e\text{-}12$, the best 12 categories/nodes) enrichments are displayed (see results section) in Fig. 1a: GO:0097458 neuron part, GO:0043005 neuron projection, GO:0034702 ion channel complex, GO:0019226 transmission of nerve impulse, GO:1902495 transmembrane transporter complex, GO:0035637 multicellular organismal signaling, GO:0022838 substrate-specific channel activity, GO:0050877 neurological system process, GO:0003008 system process, GO:0007268 synaptic transmission, GO:0042995 cell projection, GO:0022803 passive transmembrane transporter activity. Similarly, for the GO enrichment of epilepsy gene panel from [40], only highly significant ($p < 3.24e\text{-}10$, the best 12 categories/nodes) enrichments are shown in Fig. 1b: GO:0044710 single-organism metabolic process, GO:0044444 cytoplasmic part, GO:0044281 small molecule metabolic process, GO:1901564 organonitrogen compound metabolic process, GO:0005737 cytoplasm, GO:0005775 vacuolar lumen, GO:0043436 oxoacid metabolic process, GO:0006082 organic acid metabolic process, GO:0005975 carbohydrate metabolic process, GO:1901135 carbohydrate derivative metabolic process, GO:0005739 mitochondrion, GO:0044763 single-organism cellular process.

3. Results

3.1. Exome and ClinVar data

The number of WES derived epilepsy candidate genes has dramatically increased during the last four years; more than 50% of the genes associated with epilepsy are the outcome of the last two years exome sequencing efforts, a great leap towards better comprehension of the disease and its co-morbidities. Remarkably, early 2011 reviewing of epilepsy genetics [54] listed only 24 genes associated with 18 epilepsy syndromes/conditions, whereas the total number of genes derived from WES and associated with epilepsy is by now more than 133, many of which were suspected with the array approach, sometimes aided with pedigree data. The set of genes associated with epilepsy and tagged as pathogenic in ClinVar [38] totals 54 genes, of which 25 (50%) are the result of WES.

We carried out an exhaustive review of the exome sequencing contribution to this challenge in an attempt to see common relevant themes emerging. Indeed, when considering the combined set of exome and ClinVar derived epilepsy risk genes, a clear GO term enrichment is observed (see Fig. 1a); of the top enriched Gene Ontology (GO) terms, the most significant ones are: 1) transmission of nerve impulse, 2) multicellular organismal signaling, 3) neuron part, 4) neurological system process, 5) neuron projection, 6) ion channel complex and 7) transmembrane transporter complex. These pathways are suggestive of neuronal/synaptic core genes as one would expect from a neurological disease. Other GO terms are also noticed, especially genes belonging to GO:0007612 for learning, which is enriched in genes such as *CHRNB2*, *CLN3*, *GRIN1*, *GRIN2A*, *SYNGAP1*, to name only some. To stress the heterogeneity of pathways associated with epilepsy phenotypes, one may mention GO:0043038 for amino acid activation which include several tRNA synthetase genes (*HARS*, *QARS*, *TARS2*, *VARS2* and *FARS2*); this is very likely to be specific to some syndromes accompanied with epilepsy, such as Alpers syndrome in the case of *FRAS2* or mitochondrial encephalomyopathies in the case of *TARS2* and *VARS2*.

3.2. Gene panel data

Gene panels aimed at diagnostics started already emerging. They provide higher depth (100× to 500×) than exome or genome sequencing (for a review, see [3]). This led to a paradigm change in clinical genomics, a "coming of age" field, where genetic diagnostics is of paramount importance. The first gene panel for epilepsies was designed three years ago [40]. The panel contained 265 genes (abbreviated here as EGP265) that, according to Online Mendelian Inheritance in Man (OMIM), are considered as involved in "*monogenic disorders including epilepsy as a phenotypic feature*", thus a search with this panel is confined to monogenic disorders where epilepsy may not be the core symptom. From this panel, presumed disease-causing mutations could be identified in 16 out of 33 well characterized epilepsy cases (sporadic and familial). Interestingly, Fig. 1b shows an extensive difference in GO terms featured by the exome plus ClinVar derived epilepsy risk genes on one hand and the genes belonging to the EGP265 panel on the other. Indeed, most of the GO terms associated with EGP265 are related to metabolic processes. This discrepancy may pertain to the phenotype range under consideration and stress the need of more focused gene panels for epilepsy. In fact, congenital disorders of glycosylation, lipofuscin accumulation and other inborn errors of metabolism have epilepsy as a feature [52]. The intersection between the exome plus ClinVar derived candidate genes and the EGP265 panel (22 genes) is significantly ($p < 1.81e\text{-}07$) enriched in GO terms like neurological system process, transmission of nerve impulse and multicellular organismal signaling; these GO terms are probably recapitulating salient core epilepsy gene pathways. We want finally point out that four other available diagnostic panels and one list of 44 known epilepsy genes [22] show pathway enrichments similar to the exome and ClinVar derived candidate genes (data not