Identification of antiepilepitic drug-target interactions in public databases

Master Thesis

Master in Integrated Systems Biology

Author: Milena Zizovic

December 8, 2020

Preamble

The Master thesis was performed in the Bioinformatics Core Group of the Luxembourg Center for Systems Biomedicine, Luxembourg.
Principle Investigator: Prof. Dr. Reinhard Scheider
Supervisors: Dr. Roland Krause and Nikola de Lange
Reviewers: Assoc. Prof. Dr. Emma Schymanski and Dr. Aurélien Ginolhac

The part of the introduction addressing epilepsy, antiepileptics and partially therapy design was previously written for the research practical in the third semester and reused for this thesis. The research practical served as the starting point for the thesis project.

Contents

Abstract

Epilepsy is affecting people of all age and gender. The disease is traditionally treated with application of antiepileptic drugs. The therapy choice mostly relies on the differential diagnosis which is not always easy to be deducted. The treatment guidelines and antiepileptics are diverged according to major epilepsy types – generalized and focal epilepsy. However, retrospective studies of antiepileptic drug effectiveness on the European cohort have shown that pharmacoresponse is patient dependent. In this thesis the antiepileptic drug prescription trend in this cohort in generalized and focal epilepsy patients was investigated. Moreover, the use of antiepileptics in the clinics from the EpiPGX database was compared to the findings of their use in general practices in the UK. To explain the difference in patient response to therapy AED-target interactions were investigated on the level of databases. In addition, with the discovery of new genes implicated in epilepsy and success of drugs of other groups such as quinidine and fampridine in treating the symptoms, the drug-repurposing found its application in epilepsy. In this thesis, quinidine-KCNT1 and fampridine-KCNA2 interactions were investigated in order to estimate the feasibility of using public databases to select drug-target interactions for clinical application. The investigation relied mainly on the ChEMBL database. However, these genes were not found among antiepileptic drug targets in the database. Quinidine and fampridine were assay associated with other AED targets. The results suggest that the therapy choice for treatment of rare forms of epilepsy underlined by channelopathies could be significantly expanded, but that database approach requires high level of drug-target selection criteria and text mining.

Abbreviations

Term	Abbreviation
Absorption Distribution Metabolism Excretion	ADME
Adverse drug reaction	ADR
Aldehyde Dehydrogenase 5 Family Member A1	ALDH5A1
Antiepileptic drug	AED
Calcium channel	CACN
Calcium channel blocker	CCB
Central nervous system	CNS
Electroencephalogram	EEG
GABA analog	GA
International League against Epilepsy	ILAE
Mode of action	MOA
Multiple mechanism	MM
Potassium channel	KCN
Sodium channel	SCN
Sodium channel blocker	SCB
Synaptic vesicle A2 binder	SV

1 Introduction

Epilepsy is a disorder of brain function characterized by recurrent epileptic seizures (“Epilepsy” 2020). Affecting over 50 million people worldwide, epilepsy has been considered one of the most common neurological disorders and a great concern to the global population (“Epilepsy” 2020). It has been estimated to be present in the lives of 4 to 10 per 1000 people, and these figures are particularly elevated in impoverished countries populations (“Epilepsy” 2020) Since people are still stigmatized and discriminated in certain aspects of everyday life and have difficulties, for instance, to be licensed to manipulate a motor vehicle in some countries, people tend to hide the fact to be diagnosed with epilepsy (Driving Regulations Task Force | IBE Epilepsy). Antiepileptic drugs have been in use since the middle of the last century. However, epileptogenesis and pathophysiology remain unclarified. Together with the consequential difficulties of the adequate control of symptoms and a lack of a permanent cure, the unknown imposes a great responsibility on the medical and research community. Efforts are made to define proper disease classifications as it is a significant guideline for treatment protocol to manage patients.

1.1 Epilepsy classification

The International League Against Epilepsy (ILAE) has established many concepts of epilepsy classifications, each regularly revised and updated due to the progress of the research and gain of clinical experience, with the purpose of setting a world standard in the disease diagnosis process. A clinically used base definition for having epilepsy suggests a person should fulfill any of the following criteria: endure at least two unprovoked seizures occurring >24 h apart, or have one unprovoked seizure with a high probability of reccurrence after two unprovoked seizures over the next 10 years period, or have an epilepsy syndrome diagnosis (Fisher et al. 2014). After establishing epilepsy as a diagnosis, the next step is its classification. The latest ILAE Classification of Epilepsies from 2017 provides guidelines in three levels for epilepsy differentiation which would result in having a treatment choice to start with. The first differentiating feature of the disease is the seizure type (Fisher et al. 2017). According to the Operational classification of seizure types from 2014, a seizure is “a transient occurrence of signs and/or symptoms due to abnormal excessive or synchronous neuronal activity in

the brain" (Fisher et al. 2014). They can occur with focal (used to be known as partial), generalized or unknown onset depending on the starting point and the spreading manner through the brain (Fisher et al. 2014). Each group can be further divided into those with motor and nonmotor onset, meaning a patient can experience motoric impairment from mild to very extreme (Fisher et al. 2017). Furthermore, the important disease feature is also awareness or its absence during the seizure. The second level of diagnosis is to determine the epilepsy type which can be: generalized, focal, generalized and focal combined or unknown (Fisher et al. 2017). Generalized epilepsies involve seizures with a generalized onset, starting at one point and affecting both brain hemispheres from the start simultaneously, while focal epilepsies involve unifocal, multifocal and seizures originating and staying within one hemisphere (Fisher et al. 2017). Moreover, combinatorial epilepsies are reserved for patients experiencing both types of seizures. Unknown epilepsies are those that clinicians are unable to categorize as any of the three previously mentioned due to a lack of information. Finally, the third level of differential diagnosis is epilepsy syndrome diagnosis and has a purpose to help disease management (Fisher et al. 2017). The epilepsy syndrome combines a group of co-occurring features seizure types, electroencephalogram (EEG) and imaging features (Fisher et al. 2017).

Generalized epilepsy includes a major common subgroup of Idiopathic Generalized Epilepsy (IGE), also referred to as Genetic Generalized Epilepsy (GGE) in certain situations (Fisher et al. 2017). The important feature in idiopathic epilepsy is the genetic background and disease causative gene mutations. However, the genetic driving mechanism is not always identified and, therefore, only when identification is done and put in the disease context, the suitable term would be GGE for disease classification (Fisher et al. 2017). Epilepsies with genetic etiology are various and a majority of implicated genes and gene mutations have not yet been described (Fisher et al. 2017). Moreover, most of the noted mutations are found as de novo. Revealing and understanding the resulting phenotype is considered as an important puzzle piece of managing patients as it would explain disease with higher precision in individuals. Consequently, it would guide clinicians in making more effective and adequate treatment choices. However, as this is mostly in the research phase, medicine is still relying on disease pathophysiology and more robust disease classification.

Epileptogenesis and pathophysiology are two important aspects of the disease. While epileptogenesis as a latent period of chronic processes, triggered by external or internal factors (gene mutations), leads the brain into increased seizure susceptibility (Pitkänen et al. 2015), the pathophysiology is explaining consequentially occurred molecular and mechanical changes that characterize epilepsy (Engelborghs, D'Hooge, and Paul 2001). Common sense would suggest acting on the level of epileptogenesis to prevent the disease as the best solution. However, even with many existing definitions and studies, epileptogenesis is still not well defined which makes its modifications and disease prevention more challenging (Sloviter and Bumanglag 2013). On the other hand, the pathophysiology is

better understood. The basis for epileptic seizures is a disbalanced transmission in neuronal circuits. Its predominant factors are channelopathies and shifts in neuropeptide balance toward excitatory neurotransmission (Berkovic et al. 2006; Mazarati 2009). The epileptogenesis affected channels are voltage-gated (Na+, K+, Ca+) or ligand-gated channels. The hyperactivation of voltage-gated channels and subsequentially effect of excitatory neurotransmitters or lack of ligand-gated channels activation by inhibitory neurotransmitters (gamma-aminobutyric acid, GABA) results in excessive depolarization of neuronal membrane altering brain excitability and resulting in a seizure attack (Lerche, Jurkat-Rott, and Lehmann-Horn 2001) The same voltage-gated channels are widely distributed in different tissues, such as the heart muscle tissue. The ion channels are responsible for maintenance of the heartbeat by their selective permeability on the cell membrane which results in generation of the action potential (Grant Augustus O. 2009). Moreover, the best studied acetylcholine (Ach)-activated K+ channel ligand-gated channels are also found in the cardiac tissue. The excitatory neurotransmitter Ach binds to the G-protein coupled muscarinic receptor which induces the opening of the potassium channel. The muscarinic acetylcholine receptors have been found to play a significant role in the central nervous system (CNS) and generation of epilepsic seizures (Turski et al. 1989). The wide distribution of the ion channels in different tissues makes them the most common drug targets.

The pharmacological approach orients toward points that underwent the change and are active in the acute attack. The treatment's goal is to antagonize the effects of channels and reduce brain hyperexcitability. Although current studies intensively investigate genetic mutations as key epileptogenesis factors for purpose of disease development prevention, the antiepileptic drug treatment is the only available choice at the moment.

1.2 Antiepileptic drugs

The first use of a drug with anticonvulsant effect was mentioned in 1857 (Pearce 2002). Since then the number of antiepileptics increased significantly. They are roughly divided into well established antiepileptic drugs and the newer generation. The more precise drug classification by primary mode of action separates them into five groups: Sodium channel blockers (SCB), Calcium channel blockers (CCB), GABA analogs (GA), Synaptic vesicle protein 2A binders (SV) and drugs with multiple mechanisms of action (MM) (Brodie 2010).

Sodium channel blockers bind to the sodium channels in different conditions. While majority of these drugs bind to the inactivated channels to prolong their steady state, the newer antiepileptic drug lacosamide favors slow channel inactivation binding in a different manner according to its different molecular structure than other classical SCB, and, therefore, limiting rapid sequential action potentials generation of neuronal membrane (Chong and Bazil 2010) Ethosuximide with its specific binding to the

Table 1.1: ChEMBL approved drugs with epilepsy as drug indication and their properties. ChEMBL shows eslicarbazepine, eslicarbazepine acetate, gabapentin and gabapentin encarbil separately.

Drug	Abbrev	CHEMBL ID	MOA	ATC level 3	1st known approval year	aLogP	MW	Molecular species
Barbexaclone	BBX	CHEMBL3833301	GA	Antiepileptics			388	
Beclamide	BCM	CHEMBL64195	MM	Antiepileptics		1.93	198	Neutral
Brivaracetam	BRV	CHEMBL607400	SV2	Antiepileptics	2016	0.90	212	Neutral
Cannabidiol	CBD	CHEMBL190461	MM	Antiepileptics	2018	5.85	314	Neutral
Carbamazepine	CBZ	CHEMBL108	SCB	Antiepileptics	1968	3.39	236	Neutral
Clonazepam	CLZ	CHEMBL452	GA	Antiepileptics	1975	3.04	316	Neutral
Diazepam	DZP	CHEMBL12	GA	Anxiolytics	1963	3.15	285	Neutral
Eslicarbazepine	ESL	CHEMBL315985	SCB	Antiepileptics		2.49	254	Neutral
Eslicarbazepine Acetate	ESL	CHEMBL87992	SCB	Antiepileptics	2009	3.06	296	Neutral
Ethosuximide	ESM	CHEMBL696	CCB	Antiepileptics	1960	0.45	141	Neutral
Ethotoin	ETH	CHEMBL1095	SCB	Antiepileptics	1957	1.30	204	Neutral
Felbamate	FBM	CHEMBL1094	MM	Antiepileptics	1993	0.96	238	Neutral
Fosphenytoin	FPHT	CHEMBL1201336	SCB	Antiepileptics	1996	1.55	362	Acid
Gabapentin	GBP	CHEMBL940	GA	Antiepileptics	1993	1.37	171	Zwitterion
Gabapentin Enacarbil	GBP ENC	CHEMBL1628502	GA	Antiepileptics	2011	2.68	329	Acid
Lacosamide	LCM	CHEMBL58323	SCB	Antiepileptics	2008	0.45	250	Neutral
Lamotrigine	LTG	CHEMBL741	SCB	Antiepileptics	1994	2.01	256	Neutral
Levetiracetam	LEV	CHEMBL1286	SV2	Antiepileptics	1999	-0.13	170	Neutral
Mephenytoin	MPHT	CHEMBL861	SCB	Antiepileptics	1946	1.47	218	Neutral
Mesuximide	MSM	CHEMBL697	CCB	Antiepileptics	1957	1.33	203	Neutral
Metharbital	MTB	CHEMBL450	GA	Antiepileptics	1982	0.50	198	Neutral
Methylphenobarbital	MPB	CHEMBL45029	GA	Antiepileptics		1.04	246	Neutral
Midazolam	MDZ	CHEMBL655	GA	Hypnotics And Sedatives	1985	4.32	326	Neutral
Oxcarbazepine	OXC	CHEMBL1068	SCB	Antiepileptics	2000	2.64	252	Neutral
Paramethadione	PMT	CHEMBL1100	SCB	Antiepileptics	1949	0.76	157	
Perampanel	PER	CHEMBL1214124	MM	Antiepileptics	2012	4.44	349	Neutral
Phenacemide	PA	CHEMBL918	SCB	Antiepileptics	1951	0.42	178	Neutral
Pheneturide	PU	CHEMBL2107062	SCB	Antiepileptics		1.38	206	Neutral
Phenobarbital	PB	CHEMBL40	GA	Antiepileptics		0.70	232	Neutral
Phensuximide	PSM	CHEMBL797	CCB	Antiepileptics	1982	1.16	189	Neutral
Phenytoin	PHT	CHEMBL16	SCB	Antiepileptics	1953	1.97	252	Acid
Pregabalin	PGB	CHEMBL1059	GA	Antiepileptics	2004	1.08	159	Zwitterion
Primidone	PRM	CHEMBL856	GA	Antiepileptics	1954	0.54	218	Neutral
Retigabine	RTG	CHEMBL41355	GA	Antiepileptics	2011	3.59	303	Neutral
Rufinamide	RFM	CHEMBL1201754	SCB	Antiepileptics	2008	0.70	238	Neutral
Stiripentol	STP	CHEMBL1983350	GA	Antiepileptics	2018	2.84	234	Neutral
Sultiame	SUL	CHEMBL328560	MM	Antiepileptics		0.26	290	Neutral
Tiagabine	TGB	CHEMBL1027	GA	Antiepileptics	1997	5.04	376	Zwitterion
Topiramate	TPM	CHEMBL220492	MM	Antiepileptics	1996	-0.40	339	Neutral
Trimethadione	TMO	CHEMBL695	SCB	Antiepileptics	1946	0.37	143	
Valproic Acid	VPA	CHEMBL109	MM	Antiepileptics	1978	2.29	144	Acid
Vigabatrin	VGB	CHEMBL89598	GA	Antiepileptics	2009	0.36	129	Zwitterion
Zonisamide	ZNS	CHEMBL750	MM	Antiepileptics	2000	1.85	212	Acid

T-type calcium channels has been proven to be the important calcium channel blocker among others, used in absence epilepsy (generalized epilepsy) in younger patients, particularly in combination with valproate (Glauser et al. 2010).

GABA analogs exercise their pharmacological effect on different levels. Firstly, barbiturates and benzodiazepines allosterically bind to the GABA(A) receptor and potentiate the response to GABA. Secondly, valproate increases the neurotransmitter synthesis. Thirdly, vigabatrine acts on the level of GABA metabolism by reducing its clearance. Lastly, tiagabine temporarily prolongs the effect of the neurotransmitter by binding to GAT-1 GABA transporter responsible for glial re-uptake (Brodie 2010; Sills and Brodie 2002). As an element in synaptic vesicles membrane in inhibitory neurons, SV2A is found to be important in neuronal communication via neurotransmitters (Janz et al. 1999). Furthermore, it has been proven to be altered in epileptogenesis and its levels reduced in pharmacoresistant epilepsy (Löscher et al. 2016). To target and interfere with SV2A using levetiracetam has shown to result in antiepileptogenic and disease modifying effect possibly by stabilizing the vesicle structure (Daniels et al. 2013) to deliver GABA into the synaptic cleft. The antiepileptics with wider range of mechanism act on sodium channels, glutamate receptors and modulate GABA(A) responses. These include felbamate, topiramate, valproate, zonisamide, levetiracetam and rufinamide (Brodie 2010). AEDs such as gabapentin and levetiracetam still do not have clarified mode of action and seem to have multiple mechanisms.

AEDs are the drugs aiming the brain tissue. As such they are meant to cross the blood-brain barrier. The blood-brain barrier represents the contact point between the content of the blood vessels and the CNS tissue. This structure is particularly restrictive compared to other blood-tissue contacts. Among other regulating factors its tight cell junctions and astrocytes support protect a normal brain function from different harmful influences and molecules such as xenobiotics, pathogens, or injury (Daneman and Prat 2015). However, it is an obstacle in the case of drug delivery and distribution to the CNS. AEDs are mostly small molecules below approximately 300 Da and mostly lipophilic (Table 1.1).

1.3 Therapy design

To choose a treatment a clinician considers the diagnosed epilepsy type and individual patient characteristics, as well as all the aspects of the antiepileptics. The goal is to control seizures in the most efficient manner with least as possible side-effects and undesired drug interactions (Brodie and French 2000). A monotherapy as a first choice would be effective in approximately 60% of the patients, gaining full control of seizures (Kwan and Brodie 2000). The treatment option for patients with the refractory epilepsy where a monotherapy could not control the disease is a polytherapy. Moreover, a rational polytherapy suggests use of different mode of action-based combinations of antiepileptics

(St. Louis 2009). This approach would cover both pathophysiological underlying mechanisms of the seizure generation (increased excitatory and reduced inhibitory neurotransmission) and with careful therapy plan following the principles of a good practice (i.e. holding the current AED while titrating dose of the add-on until reaching the seizure control according to the drug pharmacokinetics), additive and supra-additive drug effect would be achieved with minimized side-effects (Czuczwar and Borowicz 2002; Panayiotopoulos 2005; St. Louis 2009). However, a polytherapy should not become an over-treatment and risk-to-benefit balance should be kept in mind. The monotherapy would be a treatment of choice when possible and the polytherapy should be well planned.

There are studies that have investigated the effectiveness of different AEDs in mono or polytherapy. Moreover, studies have been performed on the European cohort retrospective data ("EpiPGX – Epilepsy Biomarkers for Clinical Use" 2020) within our research group. The study of Androsova and colleagues in 2017 provided useful information on adverse drug reactions and AED retention in mesial temporal lobe epilepsy with hippocampal sclerosis (MTLE-HS) patients. The study suggested equality between the newer and older AEDs with carbamazepine as the most retained drug and the drug with the highest rate of seizure freedom. On the other hand, its analogue oxycarbamazepine was found with the highest rate of ADRs incidence (Androsova et al. 2017). Another study on this dataset performed by Silvennoinen et al. (2019) compared the effectiveness of five commonly used AEDs in patients with juvenile myoclonic epilepsy (JME). Depending on the gender, the most effective was valproate which was demonstrated as the AED to avoid in pregnancy or in women planning it. Lamotrigine or levetiracetam appeared as an alternative for valproate (Silvennoinen et al. 2019). Beside the studies helping the choice in monotherapy, there are investigations of the AED combinations based on their mode of action. As previously suggested, the combinations of drugs with different mode of action should be more adequate and rational. A study from 2014 on the American cohort data showed that different mode of action (MOA) based AED combinations gave better retention results in the focal epilepsy patients (Margolis et al. 2014). This was the model for the study in our group on the EpiPGX cohort performed by Hassanin and Krause (2018). The study was further elaborated into examination of MOA based AED combinations in focal and generalized epilepsy patients by me.

1.4 Drug-target interactions

The studies of effectiveness of MOA-based AED combinations suggested that overall different MOA combinations are longer retained in patients. The EpiPGX study raised the question of drug-target interactions and if they could explain the difference in AED-combination success in treatment. Different MOA stands for different group of AEDs, meaning that AEDs act on different set of targets or epilepsy genes. Polytherapy of right drug combination could be explained by the coverage of the

genes involved in the disease for more complete synergistical effect. Moreover, with the precision therapy and knowledge in AED-target interactions, specific gene variants could be targeted with a suitable drug. The application of the drug synergism has been already tried in the cancer treatment where therapies rely on it to compensate the individual drug imperfection (Narayan et al. 2020).

In addition to AED-target interactions, interactions of different drugs with the AED-targets would be useful for expansion of epilepsy treatment options. Many diseases share the targets and disease underlying defected genes. For instance, neurodegenerative diseases such as Parkinson's and Alzheimer's disease share over 40 genes (Dovrolis et al. 2017). Ion channels were already mentioned as commonly present in different tissues and, therefore, the channelopathies present in epilepsy are, for example , the cause of different heart conditions as well ("CV Pharmacology Antiarrhythmic Drugs" 2020). Common genes for different diseases indicate the possibility that there are also common drugs. The drug-repurposing approach has already been tried. This approach uses the pleiotropic effects of different drugs that act on many different proteins in the organism. One of the interesting examples is hydroxichloroquine, approved for the first time in 1949 according to the ChEMBL database. This drug has been indicated in treatment and profilaxys of malaria, and later in the treatment of autoimmune diseases rheumatoid arthritis and lupus erythematosus ("Chloroquine DrugBank Online" 2020). Moreover, cloroquine, different form of this compound, has been used as antiprotozoal medication. Recently hydroxichloroquine has been tried in patients with an infectuose disease (Lagier et al. 2020). The example of potential drug-repurposing for epilepsy is metformin and its effect on epileptic seizures. Metformin is an antidiabetic drug that has experimentally shown anticonvulsant and antiepileptic properties (Mehrabi et al. 2018; Yimer et al. 2019).

The concept of drug repurposing or repositioning means to use already approved drug without any modification and assign a new indication for its use in clinical treatment (Jourdan et al. 2020). In this approach, already developed drugs are considered by their chemical and biological properties to select the suitable for a specific disease. The drug application could be adapted in the new indication. For example, it could be administered via different route and at a new dose. The important elements in drug-repurposing before conducting the trials in patients are in silico research and analysis such as data mining, machine learning, approaches that are structure and ligand- based. In this process crucial segments are to describe the players in the disease, to establish drug-target interactions and to distinguish necessary properties for a drug to be adequate for achieving the therapeutic effect. With development of pharmacogenomics it could be possible to match the gene variant to a drug and to treat it. The drug-repurposing is much less costly and faster than development of a new drug, and thus, given all the factors it has become very popular (Jourdan et al. 2020).

1.5 Gene variants in epilepsy

Gene variants and their effect to the protein phenotype dictates the effect of drugs. Apart from variants that are not causing any functional change, typical consequence of those that are functional modifiers for the protein phenotype is loss- or gain-of-function (Vihinen 2020). The loss-of-function mutation reduces the function of the protein while the gain-of-function does the opposite. Moreover, the mutation interferes with the pharmacokinetics and pharmacodynamics of the treatment. Depending on the protein function modification, groups of AEDs that would have been expected to have a positive effect on a patient have resulted in symptoms aggravation. Such example is a mutation in the sodium channel alpha1 subunit encoding gene (SCN1A). Given the channel type, in some patients it was expected to control the symptoms with an SCB antiepileptic. However, biology turned out to be more complex than that. Since the variant was a loss-of-function type, the additional blockage of these channels has only made the disease worse. This distinction in the phenotype caused by the difference in the variant of the same gene was discovered in the clinics with individual patients (Balestrini and Sisodiya 2018; Hedrich et al. 2019). The evidence-based therapy recommendations would represent more solid guidelines. With this need developments have been done in the field of pharmacogenomics with the goal to achieve the precision medicine and application of personalized treatments in epilepsy patients. Since ion channelopathies are very common epileptogenic cause there have been ideas for a new project within our group to effectively target this group of gene mutations by precise AED-target interactions but also by all other drugs that could efficiently target the same targets.

1.6 Databases

A therapy guide that uses AED-target interactions and suggests drugs with different indications for treatment of epilepsy patients involves quite a few elements to be considered. Public databases that gather bioactivity data provide the access to centralized and comprehensively structured data for data mining and modeling. They represent efforts to collect all the scientific data in a meaningful manner. ChEMBL database gathers many different types of information with focus on both drugs and targets (Gaulton et al. 2017). Particularly, the database collects a large amount of information regarding compound chemical properties and activity data for over 13000 drugs, but also accumulates the information about as many drug targets in detail. The drugs could be distinguished for example by the development phase and Anatomical Therapeutic Chemical (ATC) drug classification and indications. Moreover, they have the assigned mechanistic targets which have been well recognized, but also the targets associated with them through the experimental data. The data is collected and verified with autocuration, intermediate and expert levels of curation. Moreover, a part of the data curation is man-

ual processing of assay-to-target relationships. Assigning a target to a drug in an assay is not always easy. When a single protein is assigned as a target in an assay a certain confidence score of 9 is attributed, while for yet uncurated data entries this score is 0 ("ChEMBL Data Questions" 2020). For cell-lines and tissues the level of certainty is described by 1, while for protein complex targets such as GABA receptors is 7. Assays and bioactivity data offer different type of experimental measurements such as drug potency, toxicity and effectiveness. Additionally, targets have been characterized by the type, organism of origin and gene component. The targets related data also assembles known gene variants. The ChEMBL database is constantly updated and currently in its 27th version released in May of 2020. It has been already used in different studies to model structure-activity based drug-target relationships (Bosc et al. 2019).

While the ChEMBL focuses on drug-target relations, the DISEASES database is more concentrated on disease and the disease related genes (Pletscher-Frankild et al. 2015). The database links diseases to the genes that appear to be changed using the information gathered by automatic text mining and manual curation of those results. In order to describe the disease causing potential, a confidence score that ranges from 1 to 5 is assigned to genes. The database gathers all genes related or causative to a disease. Contrary to other disease-gene associations databases the DISEASES is a public resource.
However, relying on the software that looks for disease and human genes associations in text mining of abstracts the DISEASES does not distinguish the ADME genes. The ADME genes are genes encoding proteins involved in the absorption, distribution, metabolism and excretion processes (Doogue and Polasek 2013). These genes are responsible for the difference in pharmacokinetics and the consequential pharmacoresponse in patients. The paper of Jing et al. (Li et al. 2011) uses the PharmaADME database ("Www.pharmaadme.org - Home" 2020) and lists all the ADME genes.

Another interesting database is PharmGKB that collects knowledge about the effect of genetic variations on drug response (Whirl-Carrillo et al. 2012). This pharmacogenomic database focuses on gene-drug associations supported by genotype-phenotype relationships. Furthermore, drugs are associated with reported variants of the gene with different levels of clinical annotation. The levels go from 1 for the highest to 4 for preliminary annotations based only on a case report, non-significant study or experimental assay evidence. PharmGKB is a useful database of variant-drug association information for personalized medicine.

There have been reports of different drugs with effect on epileptic seizures. Application of quinidine in a child with migrating partial seizures of infancy (MMPSI) was correlated with reduced frequency of epileptic seizures (Bearden et al. 2014). Moreover, this severe form of epilepsy has been most commonly caused by the gain-of-function mutation in KCNT1 gene. This change in potassium channels would be antagonized with potassium channel blockers. Quinidine is a class 1 antiarrhytmic drug

with partial antagonist activity on KCNT1 channels and, therefore, could be a good drug candidate for therapy of MMPSI. Another drug that has been tried in epilepsy patients is 4-aminopyridine, known under names of fampridine and dalfampridine. This drug is a potassium chanel blocker indicated in the treatment of multiple sclerosis (Judge and Bever 2006). It has been currently experimentally used in patients to reduce seizures and other epilepsy syndroms ("Experimental Epileptology : Hertie-Institut Für Klinische Hirnforschung" 2020). In these cases drugs approved for other indications were found effective to control epilepsy symptoms acting on the level of proteins affected by the gene mutation.

Although ChEMBL database is a very large structure, in order to consider the main important elements for establishing AED-target interactions for therapy guidelines and drug-AED target interactions for drug-repurposing, the DISEASES would have to be consulted. Although automatized text mining itself has its downsides, the DISEASES provide the descriptor of the epilepsy-gene associations. In addition, summarized ADME genes list from the study of Jing et al. (Li et al. 2011) would give an insight about the role of targets in ChEMBL database since drugs undergo the ADME processes. PharmGKB as the database that collects clinically actionable gene-drug associations would give another perspective to the drug-target interactions in ChEMBL database.

1.7 The context of the master thesis

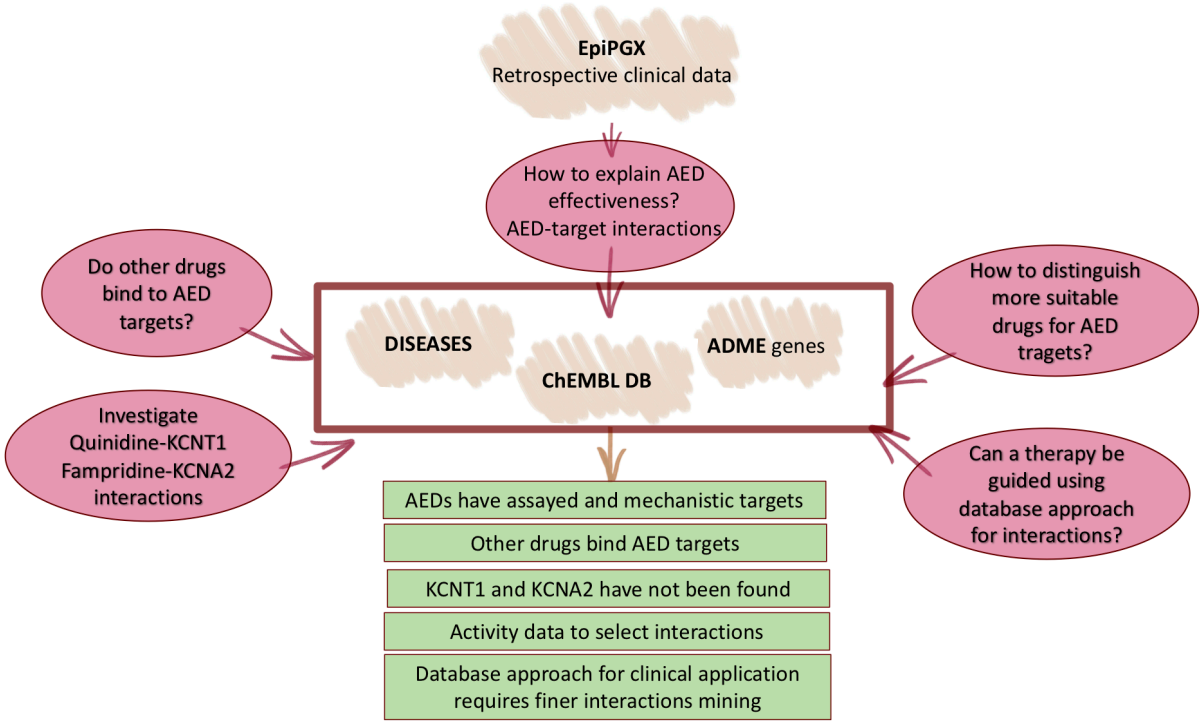

Figure 1.1: Master thesis overview

This master thesis represents an integration of two different scientific investigations on the same topic- AEDs effectiveness in different epilepsies using clinical data, and drug-repurposing based on drug-AED target interactions (Figure 1.1). The research practical in third semester was performed on EpiPGX retrospective clinical data to investigate AED combinations effectiveness in focal and generalized epilepsy patients based on their mode of action, following the work of Margolis et al. (2014). This retrospective study of the effectiveness of MOA-based AED combinations has been in preparation for the publication. The EpiPGX data has been previously used by our group in the examination of AEDs and their combinations by their retention in patients (Silvennoinen et al. 2019; Hassanin and Krause 2018; Androsova et al. 2017) and my work was an extension of those studies. The investigation of AEDs effectiveness further steered the question of "what can guide and help the rational therapy choice" to drug-target interactions in epilepsy as an important feature to describe the effectiveness of AEDs in the disease treatment. The work on drug-target interactions was proof of concept of the proposition that a therapy could be guided by association of AEDs and key epilepsy genes on individual patient basis using public databases. Moreover, use of databases that collect most of known drug-target interactions could lead to drug-repurposing and expansion of the list of possibly effective drugs in epilepsy treatment. As every new investigation motivated by new questions, it demands exploration and understanding of the sources before reaching a translational project end-product. The drug-target interactions approach was a first investigation of that kind in our group to use the ChEMBL database to explain drug effectiveness in a disease.

1.8 Aims

At the beginning, the aims of my thesis were to conclude the AED effectiveness using EpiPGX clinical data and then to investigate drug-target interactions. The drug-target interactions investigation meant to identify known AED targets and drug-AED target interactions, to rank drug-target interactions, and to rank AED combinations based on AED-target interactions to provide another dimension to the findings of the EpiPGX studies.

However, due to the time frame and with the exploration of the data the final and completed aims are:

- Conclude on EpiPGX data investigations of Silvennoinen et al. (2019); Hassanin and Krause (2018); Androsova et al. (2017),
- Collect known AED targets,
- Discover if there are other approved drugs that bind the AED targets,

- Investigate what are the features to distinct the drug-interaction targets that are available in the ChEMBL database,
- Elucidate if meaningful drug-target interactions information can be found and retrieved using public databases, and
- Try to reproduce the quinidine-KCNT1 and fampridine-KCNA2 interactions to answer the question if a therapy could be guided with the drug-target interactions using public databases.

The final aims were aligned with the first exploration of the resources to set up grounds for a new project for treating ion channels in the epilepsy.

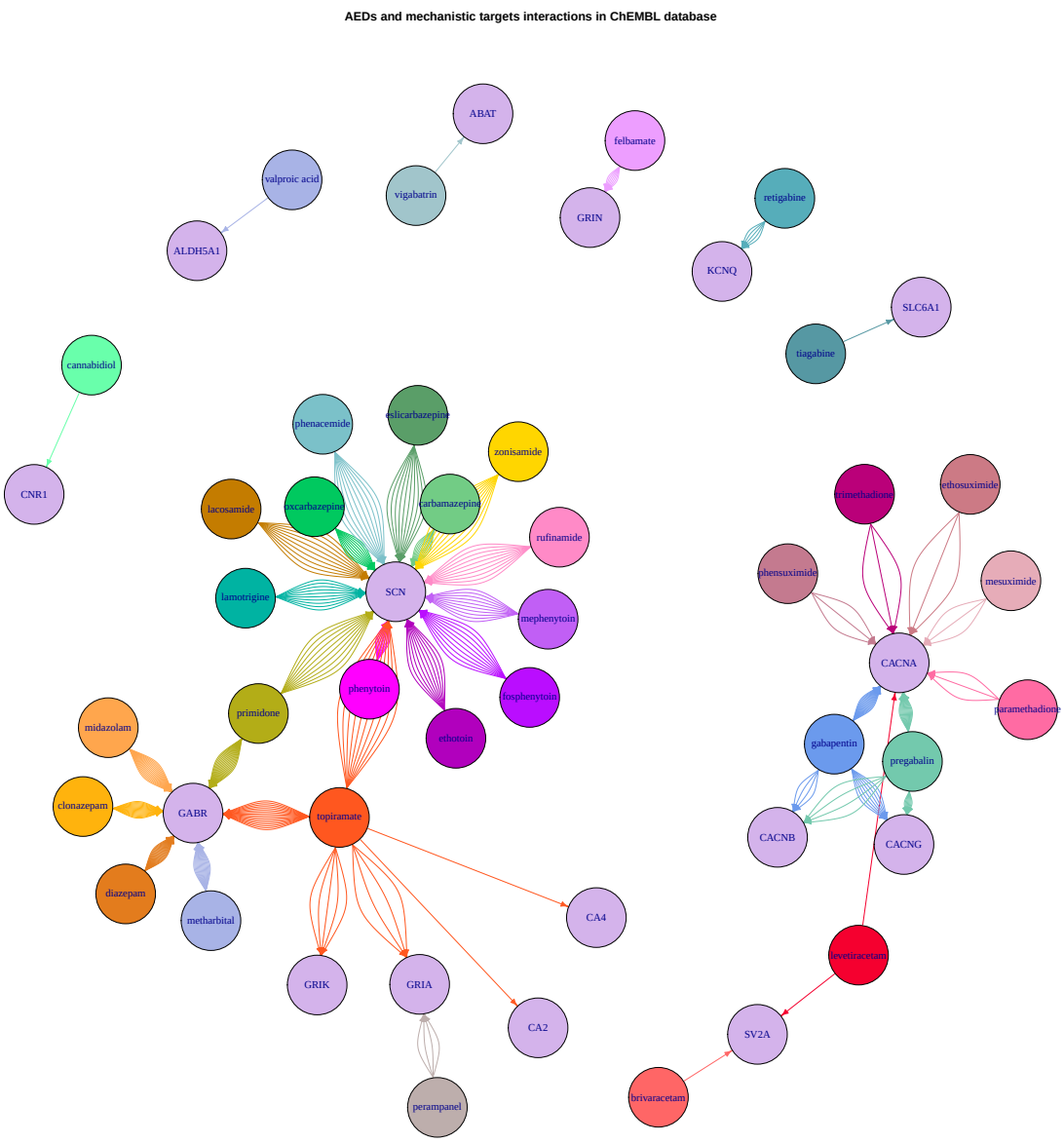

Figure 1.2: AED-mechanistic target interactions network. Number of arrows correspond to the number of collapsed genes within a target node.

2 Methods

2.1 Data sources

2.1.1 EpiPGX

As continuance of the MOA-based AED-combinations effectiveness in the research practical project the investigation of AEDs application in Europe was conducted on EpiPGX retrospective clinical epilepsy patient data ("EpiPGX – Epilepsy Biomarkers for Clinical Use" 2020). The data was collected in different European clinical centers for purposes of a pharmacogenomic project. The goal of the EpiPGX project is to make individualized pharmacological treatments for epilepsy patients based on epilepsy genome-biomarkers. The data covers records for 12829 epilepsy patients with over 40 000 AED trials from 1933 until 2015. The number of patients with AED trials gathers 6109 patients with focal epilepsy, 2097 with generalized epilepsy, 1756 patients diagnosed with some other form of epilepsy and 184 with missing diagnosis. AEDs included in the analyses were preselected after the study of Margolis et al. (2014) with later addition of ethosuximide to the list.

2.1.2 Utilization of AEDs data from General Practice Research Database

The study of Nicholas et al. (2012) investigates the changes in utilization of AEDs by epilepsy patients in the United Kingdom during 1993–2008 using General Practice Research database ("General Practice Research Database" 2020). The database provided them with data from 434 UK family practices for 63586 epilepsy patients and for the AEDs they were treated with after their registration to the GPRD. The used family practice data is up to standard for research by the policy of GPRD. This study provides the person-years for used AEDs in this cohort which was used to compare the primary care therapy trend with the protocols in epilepsy specialized secondary health care units.

2.1.3 ChEMBL database

The main data source for drug-target interactions study was ChEMBL database (Gaulton et al. 2017) version 27 published in May 2020. This database as part of EMBL-EBI ("About Us European Bioinformatics Institute" 2020) and a wide range of bioinformatics resources is built on variety of data sources - publicly available chemical structure and bioactivity databases such as PubChem BioAssay, BindingDB, together with toxicology data sets, drug/clinical candidate resources, deposited data sets, scientific literature and patents (Gaulton et al. 2017). Moreover, it covers over 65000 journal articles (Gaulton et al. 2017).

2.1.4 UniProt KB database

UniProtKB and *Retrieve/ID mapping* tool was used to retrieve gene names based on target component's accessions from ChEMBL database (Nicholas et al. 2012).

2.1.5 DISEASES database

DISEASES resource was used to describe disease-gene relations in epilepsy (Pletscher-Frankild et al. 2015). The source is a freely available text mining type of the database with manually curated disease-gene associations. Based on the source of information, disease-gene data is gathered in text mining, knowledge and experiments. The knowledge collection relies on Genetics Home Reference (GHR) and UniProtKB while the experiments collection relies on COSMIC and DistiLD. The text mining collection can overlap with the knowledge and experiments. The groups have a scoring system for confidence in disease-gene associations which is also the basis for text mining in STRING v9.1 and COMPARTMENTS. For purposes of large-scale data analysis with integration of all three collections, the associations are mapped to a common quality score for inter comparison. This confidence score scale ranges from 1 to 5 and has been used in this study to describe and order the gene-epilepsy associations. The cut off of confidence score > 2 for gene-epilepsy relations was used in the later ordering of AED-targets and drug-AED targets interactions. Moreover, in case of multiple confidence scores for an association, the higher value was used in rankings. "Epilepsy genes" have been interchangeably used with "AED targets" through the text.

2.1.6 ADME genes

The list of ADME genes from the study of Li et al. (2011) was used to consider assayed AEDs-target and drug-AED target interactions devoted to ADME processes. The researchers established a list of

31 core ADME genes and 252 extended ADME genes using PharmaADME database for the purpose of their study. The list excludes gender related ADME genes.

2.2 EpiPGX data description

2.2.1 Prescription AED trend

The investigation of the prescription AED trend was performed on the EpiPGX data ("EpiPGX – Epilepsy Biomarkers for Clinical Use" 2020) by consideration of the known prescription time points for each patient. Some AEDs are more adequate for one epilepsy type than the other, and therefore, the cohort was divided into focal and generalized epilepsy patient data to potentially give another dimension to the prescription distribution. Nitrazepam was found only in one of the diagnosis and, thus, it was excluded from the analysis. Due to low counts for some AEDs, plots were not scaled for the benefit of readability. The influence of start years between the two epilepsy types was tested using a nonparametric two-samples Wilcoxon rank-sum test for comparison of two independent groups of samples ("Unpaired Two-Samples Wilcoxon Test in R - Easy Guides - Wiki - STHDA" 2020). This statistical test is used as an alternative for an unpaired two-sample t-test when data does not follow normal distribution.

2.2.2 Person-years comparison

To investigate the difference in AED application in primary health care units and in the secondary epilepsy more specialized, EpiPGX data was compared to the UK cohort data (Nicholas et al. 2012). For the comparison person-years were used, which represent total years of a patient life under a certain treatment ("Concept: Person Years - Calculating in a Cohort Study" 2020). The time period of an AED treatment is calculated using the start date and the last visit date for an ongoing trial or the end date for terminated trials. This period is then transformed into weeks and divided by 52.25 to obtain the values in the "year" unit. Finally, person-years were accumulated for each AED. The UK cohort data was plotted by the provided AED person-years in the study of Nicholas et al. (2012). Since the UK cohort study covered the 1993-2008 time period, the EpiPGX data was restricted to the same period. Due to this time range, the only one barbexaclone trial and three mephobarbital trials were lost since their start dates were earlier than 1993. Moreover, phenobarbital, phenobarbitone and primidione were gathered into barbiturates to be compared to the barbiturates group in the findings of Nicholas.

2.3 ChEMBL data collection

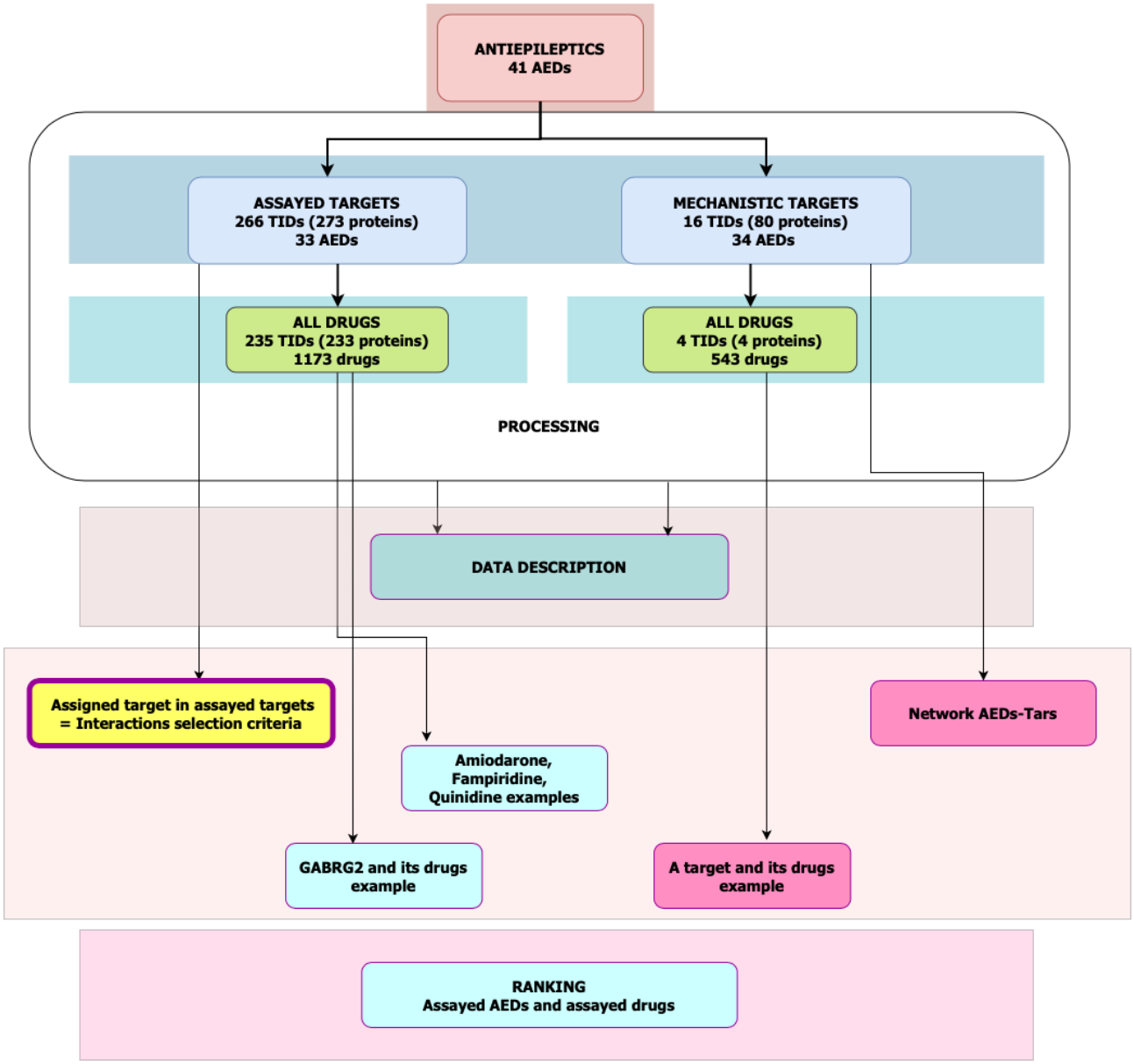

Figure 2.1: ChEMBL data exploration workflow with AED, drug and target numbers for corresponding data collections.

The ChEMBL Database version 27 was used to retrieve interactions data for:

1. AEDs and their mechanistic targets,
2. AEDs and their assayed targets,
3. AED-mechanistic targets and all other approved drugs,
4. AED-assayed targets and all other approved drugs.

ChEMBL Database gathers enormous amount of data which can be handled and searched using sev-

eral programs. In this study SQLite was used since it exists as a package for R program. The database has a specific scheme available on ftp://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/latest. The scheme points out several sections: source information, compound information, target information, experimental data, approved drugs data, drug mechanisms/indications, drug metabolism data and binding-site information. Each section contains separated tables with corresponding information.

Firstly, all drugs with drug indications containing “epilepsy” were retrieved. At the therapeutic/pharmacological ATC classification level these were drugs with N03 (antiepileptics) and N05 (pshycholeptics) classification. Furthermore, drugs as compounds and molecules have their *molregno* - internal and *CHEMBL ID* - external IDs, where the latter are created to be used by users on the web interface. Internally used by the database, molregnos can be parental or familiar depending on the form of the compound, meaning one compound can have neutral form as a parental molecule and many alternative forms such as salts, hydrates, isotopes etc. gathered as a family. More information on these compounds was found in compound related, MOLECULE_DICTIONARY, MOLECULE_HIERARCHY and ATC_CLASSIFICATION tables.

Secondly, the corresponding mechanistic and assayed targets for previously gathered AEDs were collected. The mechanistic targets were retrieved from the DRUG_MECHANISM table which stores the data only for approved compound forms. However, the approved form of a drug can be a parent or some of the family members which adds the complexity to the data collection process. The assayed targets were collected via ACTIVITIES and ASSAYS tables where target IDs were retrieved along with the assay and activity information for interactions with corresponding AEDs. In addition, details for all targets were completed with the information contained in TARGET_DICTIONARY, TARGET_COMPONENTS and TARGET_SEQUENCES tables.

Lastly, after the mechanistic and assayed targets were known and marked as AED targets, the data for their interactions with other approved drugs was assembled via ASSAYS and ACTIVITIES tables using the target ID -> assay ID –> drug ID relation. As these drug-target interactions have not been yet recognized as mechanistic, they belong to the experimental data in ChEMBL.

One of the available characteristics of targets is target type which can be a single protein, a protein complex, protein complex group, a protein family, a cell line, a tissue and other. In this analysis of interest were certainly targets of a single protein as the most reliably described interactions, targets of protein complex and protein complex group, for instance, in case of the important epilepsy targets like GABA-A receptors as the only assigned type in some conditions, and protein family. However, the target count and interpretation of drug-target interactions are difficult task given the complexity of the last three target types (Table 2.1). Moreover, these targets in ChEMBL have components which then have the corresponding gene names and UniProt accessions when they are described

Table 2.1: Example of AEDs with different target types

AED	Target ID	Target type	Uniprot gene name
Levetiracetam	31000	single protein	SV2A
Levetiracetam	10511	single protein	CACNA1B
Ethosuximide	105707	protein family	CACNA1H
Ethosuximide	105707	protein family	CACNA1G
Ethosuximide	105707	protein family	CACNA1I
Perampanel	104825	protein complex group	GRIA4
Perampanel	104825	protein complex group	GRIA3
Perampanel	104825	protein complex group	GRIA2
Perampanel	104825	protein complex group	GRIA1

and known. Since target names in ChEMBL are heterogeneous, based on the AED-target accessions, gene names were retrieved from the UniProt database using *Retrieval/Mapping ID tool* on the web interface. The gene names were referred to as target names in this study.

Approved drugs with “epilepsy” as drug indication count 43 drugs (Table 1.1). ChEMBL assembles different drug names - WHO recognized names and commonly preferred names. These names do not always overlap. WHO in the ATC drug classification employs INN names, except when they are not assigned in which case either USAN or BAN are usually used (“WHOCC - Structure and Principles” 2020). For example, ATC name “Lamotrigine” instead of preferred name “Erlosamide” was more used by clinicians. Another example is the case of “Phenytoin” which in ATC classification exists as “Phenytoin” and “Phenytoin combinations”, differing at the level 5 in the ATC classification, while in ChEMBL they have the same molregno and the same preferred name “Phenytoin”. On the other hand, some AEDs, like eslicarbazepine and gabapentin, in the ATC naming system have only one name while as preferred names have explicit names of two different drug forms (eslicarbazepine and eslicarbazepine acetate; gabapentin and gabapentin encarbil) and each form has a parent molregno. Moreover, one would expect one parent molregno for one form and family molregno for the other form as it is for majority of AEDs. Similar is seen in the other approved drugs data where some drugs have multiple parent molregno IDs for different forms of the same compound. In the analysis AED names were chosen according to the rest of the analysis elements. When the drug has many names but same molregno/CHEMBL ID, the ATC name was used. In the case of many names and different molregnos for a drug, the ATC name was used to gather and analyze the information for the same substance in different forms as for the rest of drugs. In addition, few of the AEDs were excluded from specific parts of the analysis due to the lack of data in experimental section for the mechanistic targets or in DRUG_MECHANISM table for assayed targets when they did not follow the setup criteria: AEDs-mechanistic targets and AEDs-assayed targets (Table 2.2).

Table 2.2: AEDs excluded from the corresponding analyses due to missing data.

No mechanistic targets	No assays for mechanistic targets/protein targets
Barbexaclone	Barbexaclone
Beclamide	Beclamide
Methylphenobarbital	Brivaracetam
Pheneturide	Fosphenytoin
Phenobarbital	Mesuximide
Stiripentol	Metharbital
Sultiame	Paramethadione
	Pheneturide

2.4 Epilepsy genes

Epilepsy targets gathered through assays in ChEMBL represent a collection of all tried out or observed AED interactions with protein targets. An observation of assayed AED targets for the share of ADME genes and epilepsy genes annotated in DISEASES was done. Moreover, a visualization of the epilepsy genes by the collections in DISEASES was done as well. In addition, the distribution of disease-gene confidence scores and their relation between text mining, knowledge, and experiments collections was done.

2.5 ChEMBL data description

2.5.1 Target and AED/drug counts

The retrieved mechanistic and assayed AED target data was observed for the number of AEDs and approved drugs associated with the target, and for the share of each target type in these interactions. The mechanistic targets data considers gabapentin and gabapentin encarbil (prodrug) as different compounds, as well as eslicarbazepine and eslicarbazepin acetate (prodrug). Similar trend is seen in the interactions data of all approved drugs with AED targets. Different forms of the same compound identified with multiple parent molregno IDs were considered as the same compound in the observations and counted once. All approved drugs were counted for AED target gene components taking into account the target type.

2.5.2 Assays and Activities

ASSAYS and ACTIVITIES tables in the experimental data contain different parameters to describe AED-assayed targets interactions and interactions between drugs and AED-assayed or AED-assigned targets. An assay captures key characteristics of a performed experiment for the assigned

target, roughly presented in the ASSAY_DESCRIPTION table, while related activity types and values are stored in the ACTIVITIES table. The same assay can have multiple IDs if it comes from different publications. Moreover, one assay can have multiple drugs investigated for one target such as assays collected from DrugMatrix in vitro pharmacology data ("DrugMatrix Database," n.d.) or The NCATS Chemical Genomics Center (NCGC) (Kim et al. 2019) database do.

To retrieve assayed targets for approved AEDs and assayed drugs for all AED-targets the starting criteria were:

- "Homo sapiens" as target organism,
- confidence score above 6,
- assay type Binding (B) and Functional (F),
- maximal phase equal to 4 for drugs for the indication they are approved.

Confidence score is part of manual data curation which reflects the target type assigned to an assay and the certainty that the target suits the assay. It ranges from 0 (uncurated) to 9 (direct single protein target assigned with high confidence) ("ChEMBL Data Questions" 2020). A confidence score of 7 means a direct protein complex subunit assigned, while 8 means homologous protein assigned to the assay ("ChEMBL Data Questions" 2020). The maximal phase of 4 for the drug means the drug has been approved and prescribable for an indication.

The represented experimental data in ChEMBL was observed for the influence of the first known approval year for a molecule in any state on the number of performed protein target assays. This was done for all AEDs with known approval year and an UniProt accession for the target's gene components.

2.6 AEDs-target network

The interactions network describes AEDs and their mechanistic assigned targets interactions (Figure 1.2). In this case, to each AED all associated genes were assigned as targets in the network. Interactions of an AED with a target node with multiple arrows indicate a protein family or protein complex group target type. Arrows represent connections to different genes within such a target. Moreover, such target nodes represent collapsed a protein family or a protein complex group for the benefit of better visualization. Single protein targets are connected to an AED with a single arrow.

2.7 Drugs-targets selection criteria

To describe drug-assayed AED targets and narrow down all found interactions the following criteria were applied:

- disease-gene confidence score above 2,
- confidence score above 6,
- data validity comment confirming good quality of the activity,
- standard flag set to 1 for curated data,
- pChembl value above 5,
- assay organism to be “Homo sapiens” (e.g. human cell lines).

The confidence score for a disease-gene association was selected to be above 2 to limit the number of genes associated with epilepsy by the automatic text mining and, therefore, to increase the implication of assayed AED targets in epilepsy (Pletscher-Frankild et al. 2015). In addition to the standardization of activity types/values/units, the ChEMBL database provides pChembl value for comparison on a logarithmic scale of different roughly comparable measurements such as potency, affinity, IC50, ED50, Ki, Kd (Papadatos et al. 2015). pChembl value is calculated as -Log(molar IC50, XC50, EC50, AC50, Ki, Kd or Potency). Assays with known pChembl values were chosen with perspective to use it as a feature in the later ranking of drug-target interactions. Furthermore, the arbitrary cut off > 5 was used to reduce the data for easier interpretation in the case of all approved drugs and AED targets interactions.

Data validity comment enables selection of the assays for example with values within the typical range for the particular drug and activity type (Papadatos et al. 2015). It allows selection of the good quality data. Moreover, the standard flag feature allows controlling of the selection by the curation status of the standardized data. The standardized units, relation, value or activity type can be curated or just imported by default from the published data.

The criteria were selected by investigating AED assays for SLC6A1 (GABA transporter 1), the mechanistic target for tiagabine. In addition, the investigation was also performed for CA4 (Carbonic anhydrase IV), the mechanistic target for topiramate, as a verification of the criteria selection. The aim was to obtain selection criteria to scale down the number of interactions drug-AED target interactions which is able to describe interactions of assayed AEDs for a mechanistic AED target along with preserving the AED from the mechanistic data.

2.8 Drug-targets interaction ranking

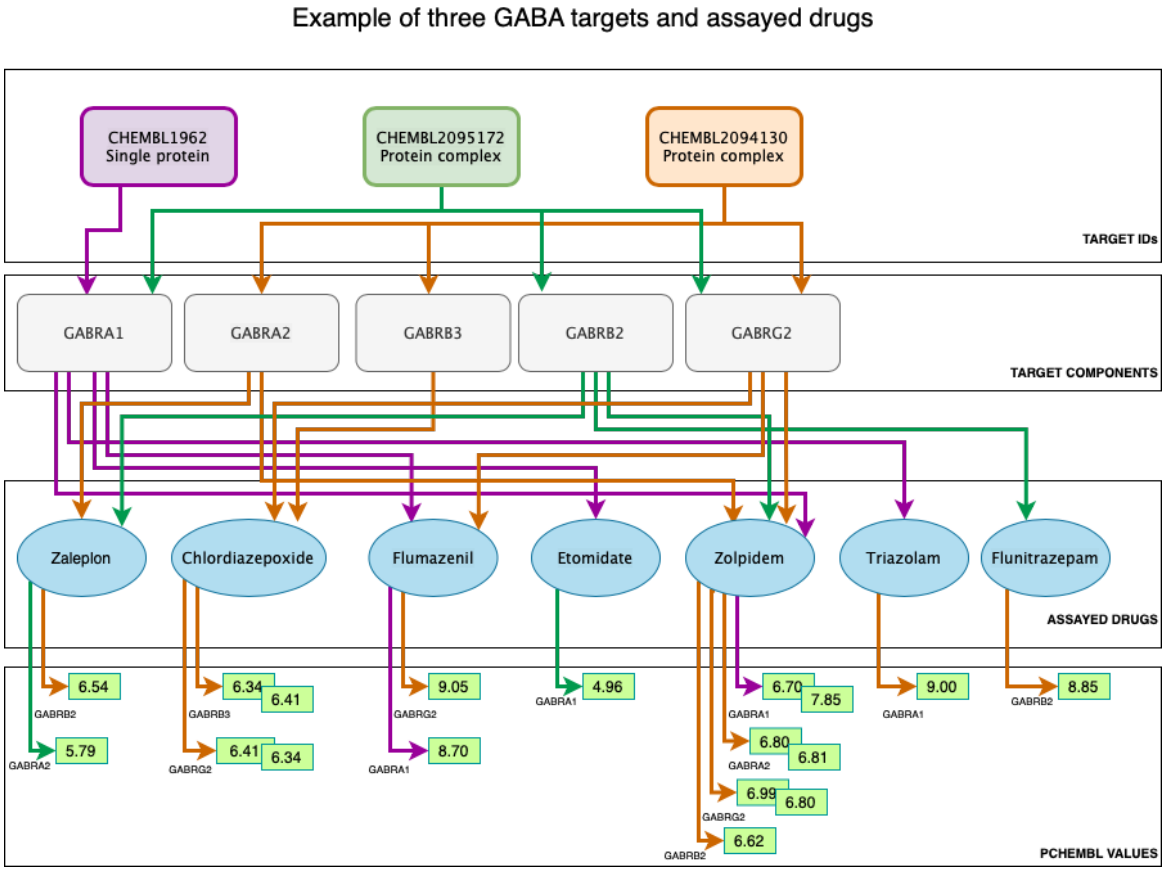

Figure 2.2: Targets with GABA component with assayed drugs and measured pChembl values for the interactions. Colours correspond to the target. GABA genes can have multiple pChembl values recorded from different assays.

The drug-AED target interactions were first ordered by the disease-gene confidence score collected for epilepsy in DISEASES and afterward by the pChembl values from the ACTIVITIES table in the ChEMBL database. For drugs with multiple assays and therefore multiple pChembl values an average of the value was done. For drugs interacting with targets annotated by different IDs but containing the same gene component, an average pChembl value was done across all targets with the same gene. (Figure 2.2) The average pChembl value would finally describe a drug-gene interaction.

2.9 Tools used in the analysis

The important tools that were used in this work were R program (R Core Team 2020) with many different packages such as Drake (Landau 2018), Renv (Ushey 2020), Ggplot2 (Wickham 2016), Tidyverse (Wickham et al. 2019), RSQLite, Igraph (Csardi and Nepusz 2006), and others. The use of these was supported and assisted by my supervisors Dr. Roland Krause and Nikola de Lange.

The analysis was done in RStudio using R program version 4.0.2. The project was developed using version control by Git and shared GitLab repository. In order to ease the reproducibility of the project environment, the Renv dependency manager was used as well. This package allows a user to create a private project library. The packages used in the project are recorded in a lockfile. The project environment is first initialized with `renv::init()`, then the renv.lock file is created using `renv::snapshot()` and can be updated as the work evolves with the same function. The environment is set via `renv::restore()`. This is convenient for collaborations to ensure that everyone involved in the project is using the same environment. Moreover, it is useful in the case of multiple projects to individualize and control the environments.

The EpiPGX analysis follows a project structure where data import, processing and modeling are separated into different R scripts (Grolemund and Wickham 2017). Moreover, R scripts and the manuscript related Rmd files are in different directories. To construct paths to the files when necessary Here package was useful. It builds a path to the file using only a filename and the directory where it has been saved. The work on the drug-target interactions follows a similar pattern but within the Drake setup. Drake allows reproducibility but also to faster handle the analysis of bigger datasets such as the ChEMBL data collection. This package analyze the workflow and runs only the parts that are not up-to-date. Drake setup includes the functionalization of the analyses, creation of a Drake plane and a make file. Data import, wrangling and modeling can be in separate scripts and then invoiced into a Drake plan. The main used functions are `drake_plan()` to create a workflow, `make()` to run the plan and build the project, and `vis_drake_graph()` for visualization of the target state. Another useful function when it is necessary to move from one setup to Drake is code_to_plan() creates a relationship of the plan to the R script. Instead of objects, with Drake the work is done with targets which can be loaded into a manuscript with `loadd()` or inspected using `readd()` function. In the transition of the project setup to the use of Drake environment I was kindly supported by my supervisor Dr. Roland Krause. In addition, data transformation and analysis was relying on the use of the Tidyverse package.

ChEMBL database has over 25GB of the data and can be programatically accessed using database instances. In this project, it was communicated with SQLite as it has a package called RSQLite to connect to the locally stored SQLite database instance in R to retrieve the necessary data with SQL queries. To visualize the AED-target interactions ggraph (Pedersen 2020), igraph (Csardi and Nepusz 2006) and qgraph (Epskamp et al. 2012) packages were explored. These visualization packages offer numerous options and freedom in the network graph creation.

The thesis manuscript is done in R Markdown using the Bookdown package (Xie, Allaire, and Grolemund 2018). This setup contributes to the reproducibility of the thesis. The project for drug-target interactions can be found on https://git-r3lab.uni.lu/EpiPGX/drug-target-interactions. The EpiPGX part

of the thesis can be found on https://git-r3lab.uni.lu/roland.krause/mtle-hs/-/blob/master/multi-drug-trials/Exploratory.Rmd.

3 Results

There are generally two parts of the results of this master thesis project. The first part represents the findings of the investigation of the EpiPGX data. The EpiPGX data was examined to see the AEDs application in different types of epilepsy. Moreover, I wanted to compare the use of AEDs between different levels of the European healthcare system. The second part of the results is produced in the exploratory data analysis for the description of drug-target interactions. Particularly, this EDA shows the possible interaction selection criteria, challenges presented by the data structure and the interactions of quinidine and fampridine with potassium channels recognized as AED targets in the ChEMBL database. Furthermore, these results show the complexity of distinction and interpretation of meaningful drug-target interactions.

3.1 EpiPGX data findings

The EpiPGX data investigation shows results that were additions to Hassanin and Krause (2018) and my research practical work. The purpose of my current work was to investigate the treatment preference for focal and generalized epilepsy and compare to another treatment databases (“General Practice Research Database” 2020).

3.1.1 AED trials distribution

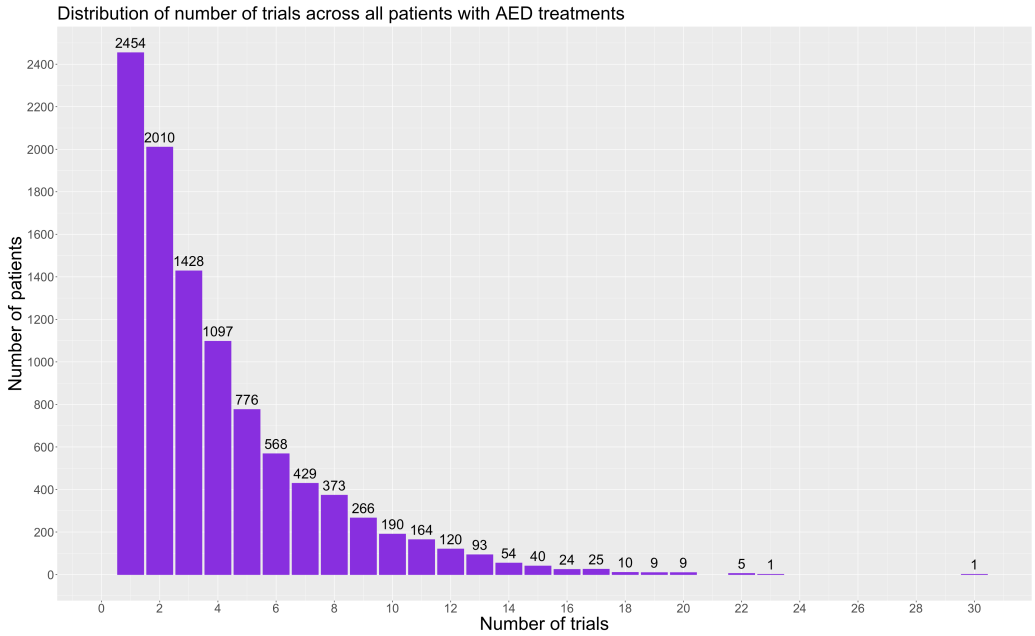

Figure 3.1: Number of trials over total cohort of epilepsy patients with recorded AED trials.

From a total of 10146 epilepsy patients with AED trials the majority of patients has tried 1-2 AEDs in their treatment of the disease. The next big group of patients has tried three or four, after which the number of patients who has tried more than 4 drugs lowers compared to the majority of the cohort. However, 32.07% of the cohort has been prescribed with more than three AEDs until they potentially achieved the control of the symptoms. Moreover, the number of trials for some of them reaches 22, 23, even 30 different AEDs. The trend of multiple treatments per patient could be a reflection of the specificity of epilepsy diagnosis (Figure 3.1).

3.1.2 AED prescription trend

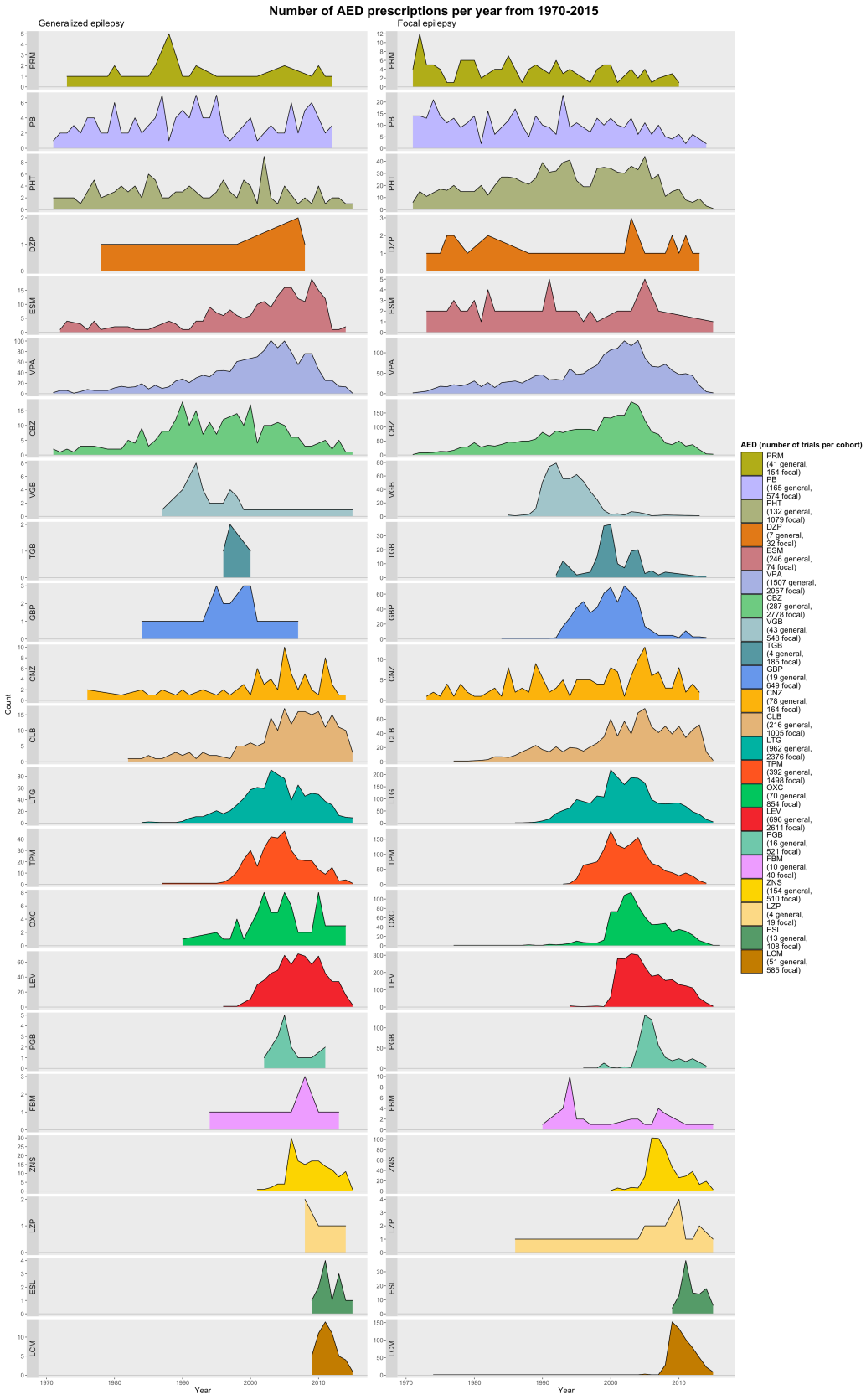

Figure 3.2: The prevalance of AED prescriptions by count. For readability reasons plots are not scaled between generalized and focal epilepsy cohorts.

Table 3.1: Wilcoxon rank-sum test comparing AED start years between generalized and focal epilepsy patients

AEDs	Statistic	p value
PRM	3.03e+03	7.04e-01
PB	3.89e+04	4.78e-04
PHT	7.64e+04	1.70e-01
DZP	1.22e+02	7.28e-01
ESM	3.89e+03	7.87e-14
VPA	1.37e+06	1.77e-09
CBZ	4.43e+05	1.75e-03
VGB	1.26e+04	4.49e-01
TGB	5.86e+02	4.39e-02
GBP	7.89e+03	3.75e-02
CNZ	4.60e+03	3.99e-04
CLB	8.10e+04	4.46e-09
LTG	9.39e+05	5.67e-16
TPM	2.18e+05	4.49e-15
OXC	2.88e+04	6.04e-01
LEV	6.65e+05	9.44e-28
PGB	4.75e+03	3.36e-01
FBM	9.25e+01	8.97e-03
ZNS	3.32e+04	3.50e-03
LZP	3.40e+01	7.75e-01
ESL	7.46e+02	7.12e-01
LCM	1.06e+04	5.13e-04

The prevalence of AED prescriptions reflect realistic events where older generations of drugs are superseded by the newer (Figure 3.2). The order of the AEDs follows the mean start year in the cohort. The so-called old-timers, phenobarbital, phenytoin and primidone have lower mean trials start year compared to the other AEDs. Lacosamide has been presented with the highest mean trials start year. Valproate with its side effects has been pushed aside by its newer safer alternative lamotrigine (Silvennoinen et al. 2019). Newer generation AED Levetiracetam has recorded over 300 prescriptions in a year in the treatment of focal epilepsy, more than the older carbamazepine in the same year, which was typically used in this epilepsy.

In addition, besides the time frame an important weight in the prescription choice has the epilepsy type. Vigabatrin and tiagabine with 548 and 185 recorded prescriptions have been prevalently given to patients with focal epilepsy compared to 43 and 1 recorded trials in the generalized epilepsy cohort, respectively. Ethosuximide has been tried 3 times more in the generalized epilepsy cohort than in the focal where the number of prescriptions lowers over time. As expected, there has been a statistically significant difference in ethosuximide prescriptions over time between generalized and focal epilepsy (p-value of 7.87e-14). Similar significance in time difference can be seen in the prevalence between the two diagnoses of other AEDs specific to the diagnosis, meaning that with time prescription trend shifts (Table 3.1).

3.1.3 Difference in general and clinical practice

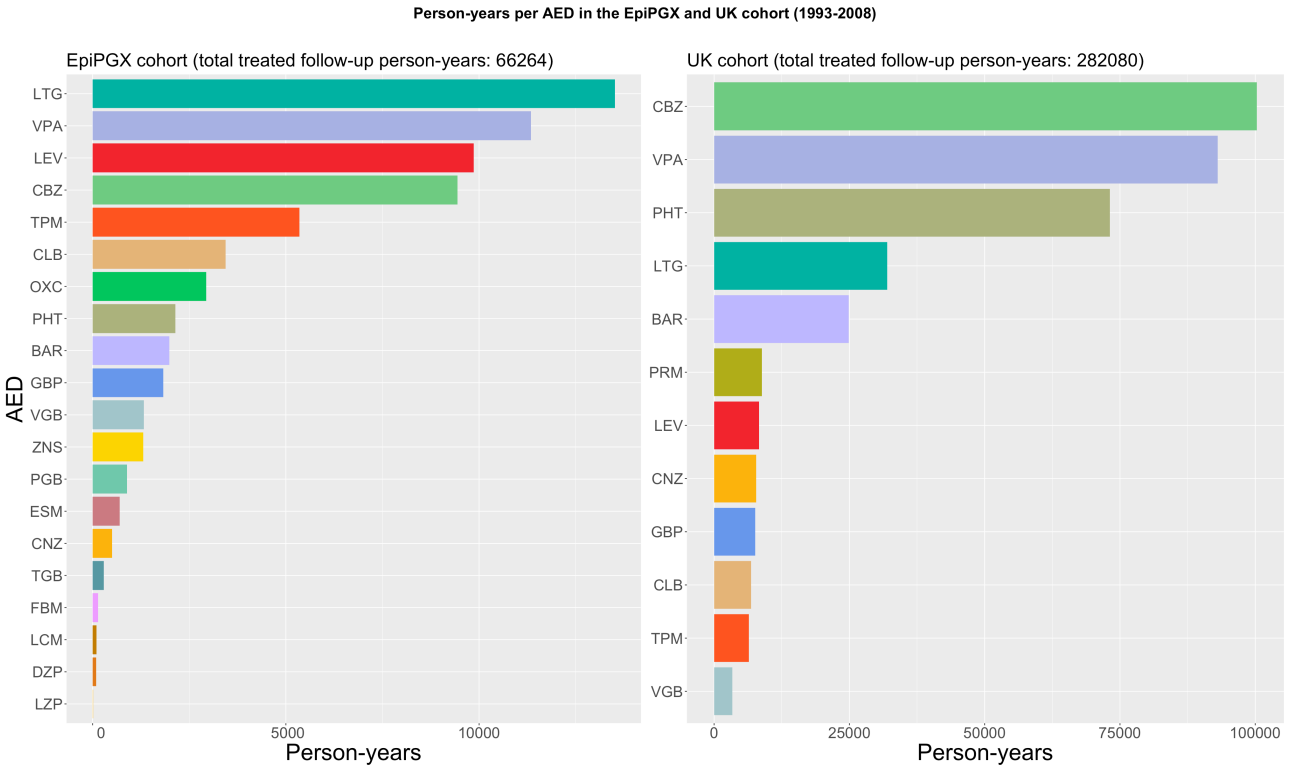

Figure 3.3: In the UK cohort ESM, OXC, PGB, TGB, ZNS, SUL, RFM, FPHT, STP, MSM, BCM, LCM were prescribed < 1% and not presented.

A different trend of treating protocols can be noticed between specialized clinics and general family practices. Moreover, although the EpiPGX cohort covers a larger region of the continent had 66264 person-years, where territorially smaller the UK cohort had 282080 person-years in total (Figure 3.3). The most prescribed and used AED in the primary health care units in the UK was carbamazepine (100296 person-years, 27.77%), followed by valproate (93051 person-years, 25.76) and phenytoin (73154 person-years, 20.25) (Figure 3.3). The relatively newer AEDs (year of UK license in the parantheses) oxcarbazepine (2000), pregabalin (2004), tiagabine (1998), zonisamide (2005), rufinamide (2007), lacosamide (2008) were used less than 1% of treated follow-up in the UK cohort according to the study of Nicholas et al. In the EpiPGX cohort the most used was lamotrigine (13512.1 person-years, 20.01%), followed by valproate (11343.6 person-years, 16.88%), levetiracetam (9861.4 person-years 14.67%) and carbamazepine (9440.8 person-years, 14.05%). Interestingly, levetiracetam was used for 8313 person-years in the primary health care units, 2.30% of total AEDs person-years, while in the secondary health care units it was used for 9861.4 person-years, covering 14.67% of the totality.

The presented AEDs are drugs present on the market. Some of the patients do not respond well to the treatment and have tried many AEDs until reaching the control of the symptoms. The AEDs and their mechanism of action were then explored in the ChEMBL database.

3.2 ChEMBL exploratory data analysis

This section shows results that are proof of concept of whether a database approach can help therapy protocols using AED-target interactions, but also to find drug candidates through drug-AED target interactions for drug-repurposing. Moreover, this new project is done to see if the discovery of quinidine or fampridine as AEDs could be reproduced from databases alone.

3.2.1 AEDs and assayed targets

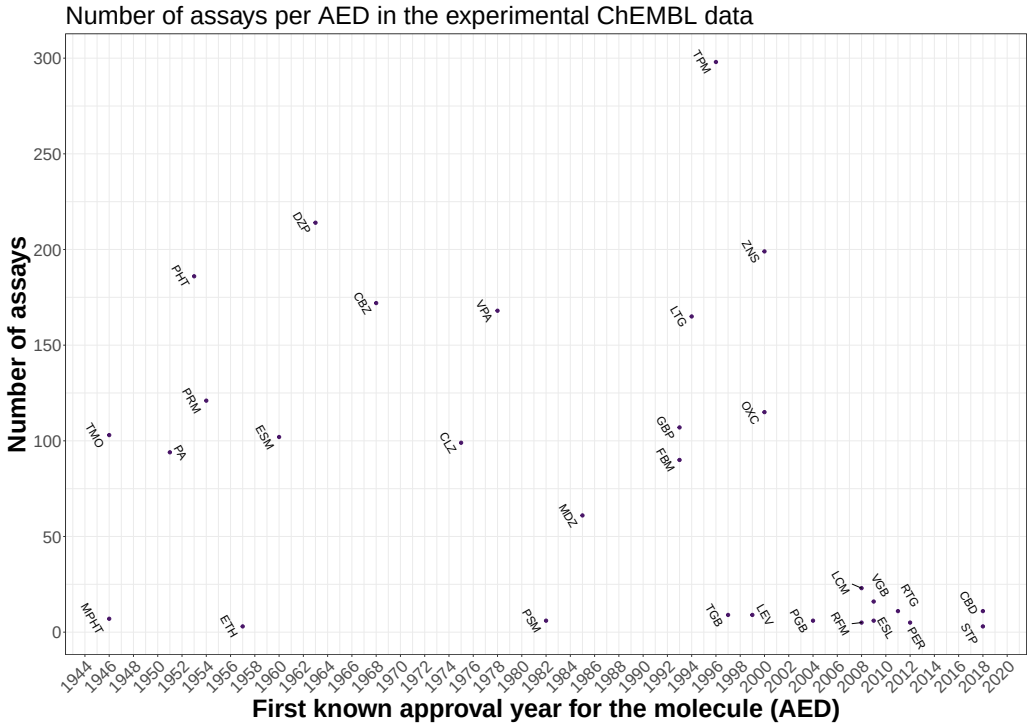

Figure 3.4: The number of assays for AEDs with known first approval year for any indication and with known gene components of assayed targets.

Among AEDs with known first approval year in at least one state for any indication the earliest released AED was mephenytoin (1946), while the latest approved were cannabidiol and stiripentol (2018) (Figure 3.4). Mephenobarbital, phenobarbital and sultiame, although studied and found in the experimental data, are not presented due to unknown first approval year. In addition, eslicarbazepine acetate and gabapentin are the forms of these compounds assayed with AED targets. The number of assays

performed for each AED differs. Typically, older AEDs such as those before the year of 2000 have more experimental data than the newer. The AED with the most performed assays is topiramate, followed by diazepam and zonisamide. However, AEDs with highest number of recorded activities for these assays are phenytoin with 1051267 times measured assay endpoints, followed by oxcarbazepine with 886012 and carbamazepine with 831541 activities. ChEMBL database gathers over 40 different standard activity types that have been measured multiple times for AEDs and targets in different assays. AEDs approved after the year of 2000 have less than 50 assays.

3.2.2 Targets and AEDs distributions

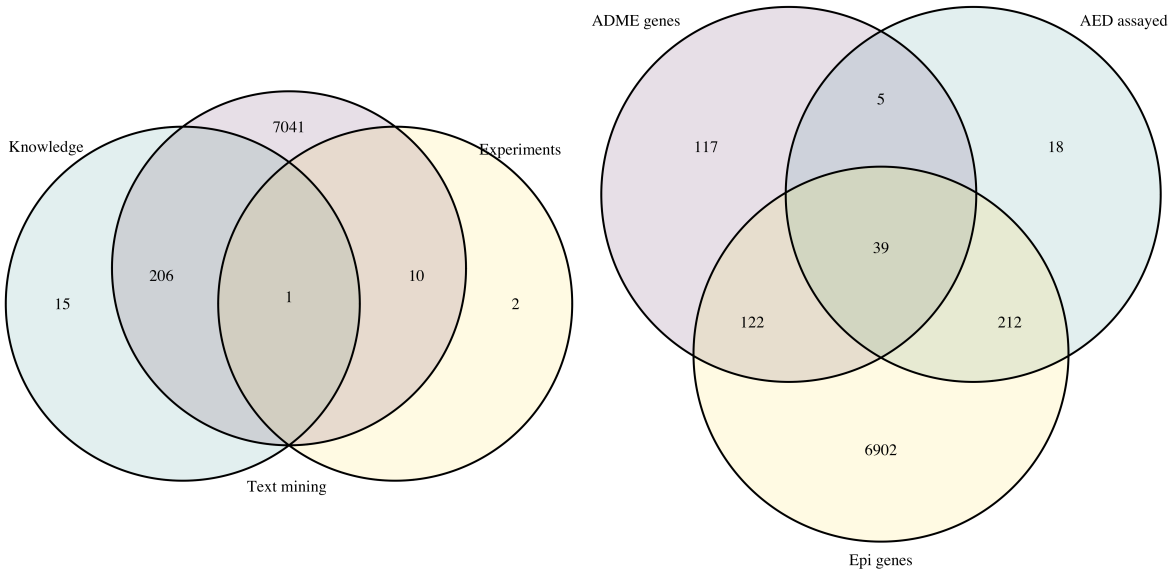

Figure 3.5: Epilepsy genes coverage: (left) Epilepsy genes source overlap in DISEASES database; (right) genes in ChEMBL, DISEASES and ADME

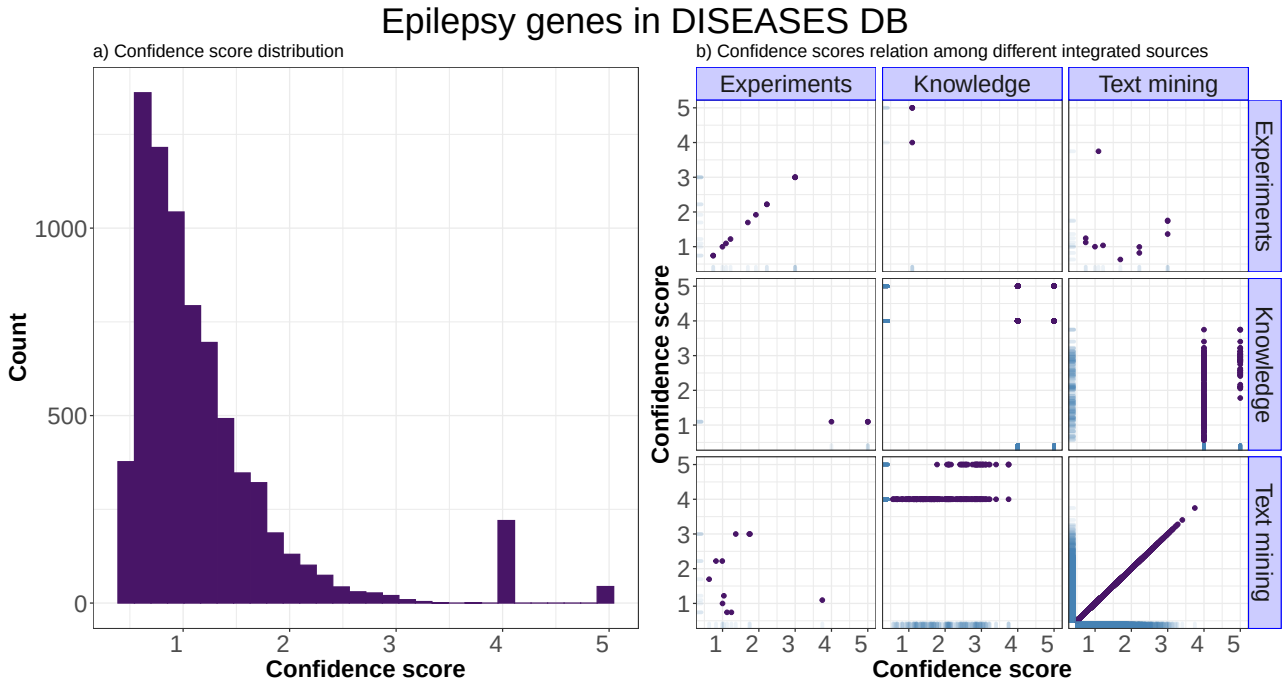

Figure 3.6: Distribution of disease-gene confidence scores presented in DISEASES database for epilepsy for total list of interactions and relations of confidence scores between three type of data collections. In the plot b) the rug shows density of confidence scores.

Epilepsy-gene interactions are one layer of information in understanding the disease and therapy choices. There are 7275 epilepsy genes in DISEASES database of which 13 are gathered in the experiments collection, 222 in the knowledge collection and 7258 in the text mining collection (Figure 3.5). Notably, the largest group is obtained with text mining and contains most of the genes from the other two collections as well. Moreover, this group has lower confidence scores ranging from 0.5 to 3.748 (Figure 3.6). Similarly, the genes gathered in experiments have been related to epilepsy with scores ranging from 0.743 to 2.999. On the other hand, the knowledge collection has confidence scores of 4 and 5. The confidence scores are mostly distributed around a value of 1.

From the assayed AED targets in ChEMBL database, 251 overlap with DISEASES gene list while 44 can be found in ADME gene list (Figure 3.5). From 44 18 are considered as core metabolic genes, while the rest is recognized in the extended list. As for the epilepsy-gene association, genes found as components of assayed AED targets in ChEMBL database are associated with epilepsy with minimum of 0.513 to 5 confidence score.

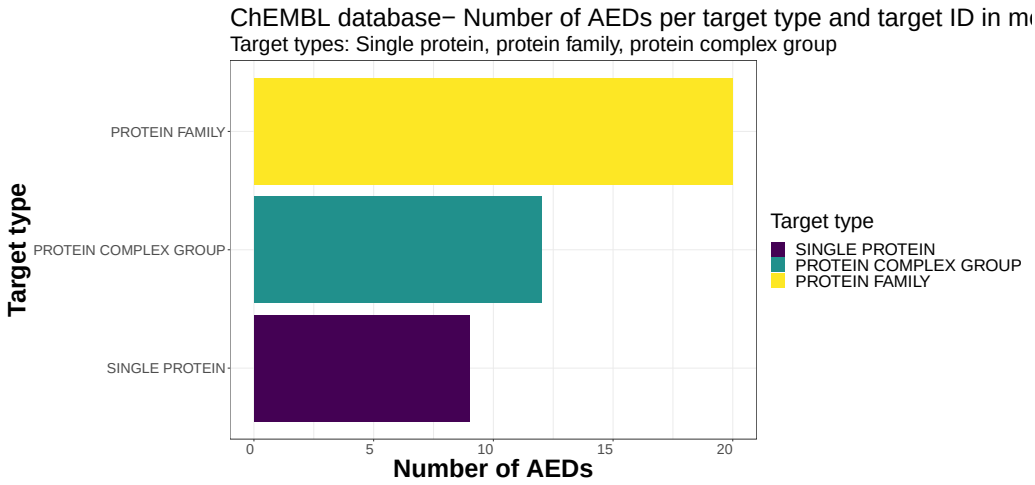

Figure 3.7: Mechanistic targets and the number of AEDs according to the target type.

There are 16 mechanistic AED targets counted by the ID. These have 80 genes. In particular, the much high number of considered epilepsy genes compared to the number of targets is contributed to by the protein complex group and protein family target types as more complex. Single protein type counts for 8 which are considered as mechanistic targets for 7 AEDs. Then, protein complex group type accounts for 5 targets involved in the mechanism of action of 10 AEDs. Finally, protein family target type appears in 3 targets and is connected to 20 AEDs. When the AED-target IDs interactions are as well considered in AED-target types relations the protein family target type counts the highest number of assigned AEDs, followed by the protein complex group and the single protein. In this case, some AEDs are assigned to multiple targets and counted repeatedly in different target types. For example, topiramate acts on three targets: GABA-A receptor, Glutamate receptor ionotropic AMPA and Glutamate receptor ionotropic kainate types. Each of these targets has many gene components. In this case topiramate has been counted three times (Figure 3.7). Moreover, topiramate has been assigned to a protein family target and two single protein targets and has been reported in those groups of target types as well. This shows the coverage of target types in the AED-mechanistic targets interactions by a single drug and contributes to the complexity of interpretation of interactions.

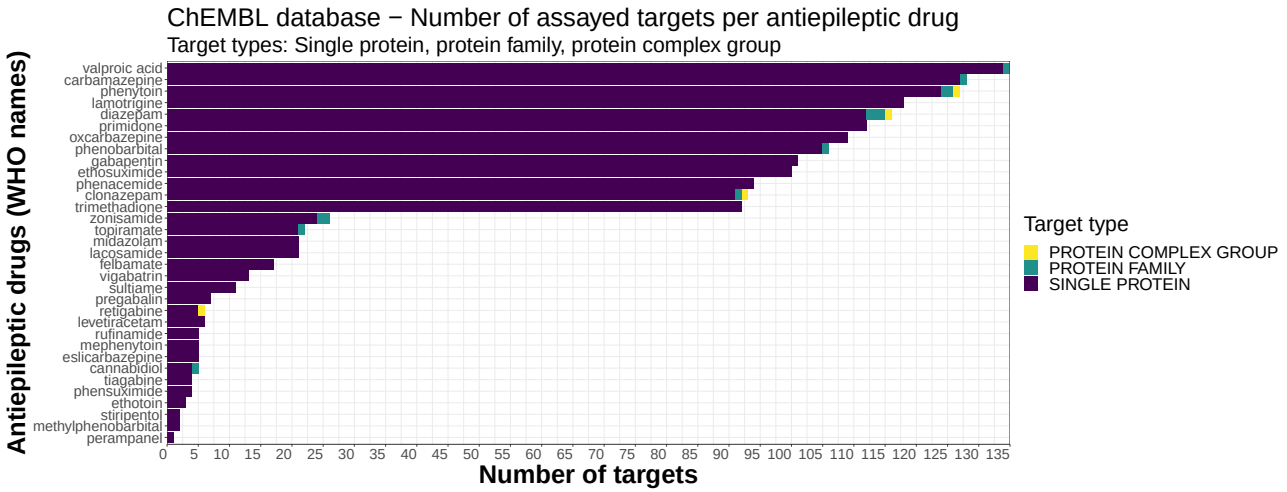

Figure 3.8: Distribution of assayed targets with the target type feature across AEDs with experimental data in ChEMBL.

Assays experimentally connect AEDs to 266 possible targets (Figure 3.8). Surprisingly, the predominant target type is the single protein, while protein family and protein complex group types cover 1-2 targets per AED. The confidence in assigned targets in these assays for AED-target interactions goes from 4 and 5 for protein family and protein complex group, to 9 and 8 for the single proteins. Valproic acid was displayed as the AED with the highest number of possible target interactions, followed by carbamazepine and phenytoin. One of the 137 assayed targets is ALDH5A1 that represents also the only target considered as mechanistic for the therapy effect of valproic acid. Perampanel was the drug with only one drug-target interaction in the experimental data. Moreover, this target was a single protein encoded by GRIA1 gene, which is also one of the gene components of the mechanistic target for this drug. Compared to the number of mechanistic targets, a much higher number of assayed targets shows more possibilities and additional space for drug repurposing.

Table 3.2: Top targets with highest number of assayed drugs. ADME genes are colored in grey

Uniprot gene names	Target-epilepsy confidence score	Total drugs per target
KCNH2	2.19	630
TDP1	1.08	627
CYP3A4	2.22	612
CYP2D6	1.96	609
CYP2C19	2.31	608
CYP2C9	2.28	608
CYP1A2	1.69	607
MAPK1	1.68	595
EGFR	1.71	554
DRD1	2.28	552
MAPK14	2.32	550
PRKCA	1.39	550
FYN	1.82	549
LCK	1.40	548
MAPK3	2.00	548

3.2.3 Drugs and AED assayed targets

Figure 3.9: Count of drug-assayed AED genes interactions. The graph presents targets with 10 and more drugs for purpose of readability. Colors identify mechanistic, ADME devoted targets as well as targets with GABA gene components due to their complexity. Although the readability is very low, the graph should convey the distribution trend.

From 266 AED assayed targets 235 were found experimentally associated with other approved drugs (Figure 3.9). The target types in these targets are single protein and protein complex, where protein complex describes targets with GABA gene components. When GABA targets are distinct, the rest of the genes represent single protein targets with only one target ID. The gene displayed as AED target with the highest number of drug-target interactions is KCNH2 with 630, followed by TDP1 with 627 (Table 3.2). There is a plateau-like part of the distribution with approximately 400-550 recorded drug-target interactions. The 84 targets are diverse, gathering different neurotransmitter and neu-

ropeptide receptors (Table 6.2 in the Appendix). This share of the AED assayed, autonomic nervous system (ANS) related targets consists of adrenergic, adenosine, muscarinic acetylcholine, histamine, dopamine, serotonin, opioid and neuropeptide Y receptors. From the 81 genes, two have not been found as related to epilepsy yet according to the DISEASES database. The confidence score for epilepsy-gene associations for the rest of them seems to largely sit below 2.5 and to belong to the text mining generated epilepsy gene collection.

3.3 Drug-target interactions ranking

The ranking of drug-AED target interactions was performed using assay pChembl values and confidence scores for gene-epilepsy relation. This section shows the complexity of the process to describe the drug-target interactions driven by the data structure. Furthermore, it presents the findings of quinidine, fampridine interactions with potassium channel AED assayed targets in the ChEMBL database. Additionally, few other interactions of the most druggable AED assayed target are shown.

3.3.1 Drug-assayed AED targets interactions

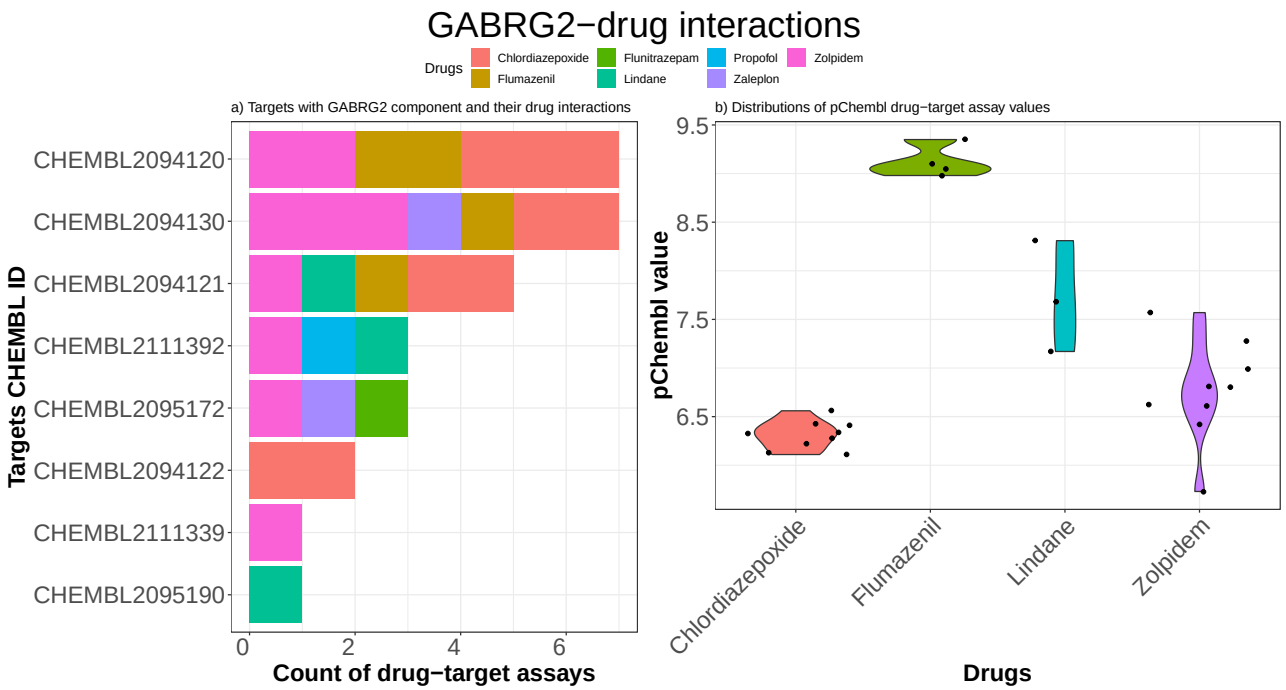

Figure 3.10: Example of drug-target interactions considering gene component to explain the necessity of data summarization. pChembl value distribution per drug explains the use of average pChembl value in drug-target interactions ranking. Drugs with less than 3 pChembl values were not presented in the distribution plot.

Table 3.3: GABRG2-drug interactions

Gene	Conf.score Epilepsy-gene	Drug	Conf. score drug-target	Avg. pChembl value
GABRG2	5	Flumazenil	7	9.12
GABRG2	5	Flunitrazepam	7	8.85
GABRG2	5	Lindane	7	7.72
GABRG2	5	Zolpidem	7	6.88
GABRG2	5	Chlordiazepoxide	7	6.32
GABRG2	5	Zaleplon	7	6.17
GABRG2	5	Propofol	7	5.10

Ranking of drug-target interactions was performed for all drugs and AED assayed targets. With the interactions selection criteria the number of 235 assayed targets was reduced to 50. Moreover, it means that 233 assayed genes was reduced to 49. To explain the complexity of the data in the ChEMBL database a GABA gene was taken as a representative since these genes belong to the protein complex group target type (Figure 3.10). GABRG2 gene is a component of eight targets which all have different drug interactions. While they essentially interact with the GABRG2 gene, some drugs establish the interaction through many of the targets and for each have multiple assays and activities recorded. For instance, zolpidem interacts with six of them and while for four of them has a single pChembl value, for two of them has three and two values recorded. The second example is chlordiazepoxide that has two pChembl values per interaction with three targets and three values for the fourth target. The last of them is flumazenil with four pChembl values for in total of three targets to characterize the flumazenil-GABRG2 interaction.

In order to summarize the description of GABRG2-drug interactions, when needed pChembl values were considered as the average of all existing values for the gene-drug interaction across all targets with the gene as a component. Particularly, the values seem to be grouped per drug and less different than between the drugs. The values were approximately around 6.30 in the case of chlordiazepoxide, 9.12 for flumazenil, 7.72 for lindane and 6.88 for zolpidem (Table 3.3). This distribution seemed to allow the use of average pChembl value for drug-gene interaction for purpose of ordering them. In the case of drug-gene interactions with a single value of pChembl, the averaging was not performed. According to the average pChembl value, drug-GABRG2 interactions displayed flumazenil as top positioned, followed by flunitrazepam. The lowest positioned was propofol with a pChembl value of 5.10. The data curation for the interaction of GABRG2 with propofol has been done by an expert. In the case of zolpidem, the lowest pChembl value was done by an expert while the rest of them were autocurated and data for the other five drugs was autocurated.

Table 3.4: Fampridine-KCNH2 target interaction

Drug CHEMBL ID	Conf.score Epilepsy-gene	Data validity comment	Curation	pChembl value	Confidence score
CHEMBL284348	2.19	Outside typical range	Expert		8
CHEMBL284348	2.19	Outside typical range	Autocuration		9

Table 3.5: Amiodarone-KCNH2 target interaction

Drug CHEMBL ID	Target CHEMBL ID	Assay type	Curation	pChembl value	Avg. pChembl	Confidence score
CHEMBL633	CHEMBL240	B	Autocuration	6.00	6.76	9
CHEMBL633	CHEMBL240	F	Autocuration	7.52	6.76	9

3.3.2 Specific drug-AED assayed targets examples

KCNH2 gene has been associated with epilepsy by text mining with a confidence score of 2.19. Since this gene has been seen as the target with the highest number of drug interactions it was investigated for certain drugs. The gene was assay associated with 15 AEDs and as previously displayed with 633 other approved drugs. After the interaction selection criteria were applied, KCNH2 was targeted by 70 approved drugs. According to clinical annotations about gene variant-drug interactions in the PharmGKB database KCNH2 was found associated with nitrendipine (level 3 evidence), amiodarone, dysopiramide, and quinidine (all with level 4 of evidence). Nitrendipine was not found in the drug-assayed AED targets data gathered in ChEMBL. Furthermore, disopyramide was excluded by the drug-interactions selection criteria due to pChembl value below 5. However, amiodarone and quinidine were present in the selected drug-target interactions. Dronedarone, structurally related to amiodarone, had also interaction with KCNH2 in this data. Additionally, 4-aminopyridine, also known as fampridine in the ATC classification and the ChEMBL database, was found associated with KCNH2 among other drugs.

After the selection, there were two activities for Amiodarone-KCNH2 interactions in the experimental data (Table 3.5). These belong to assays with binding and functional assay type with maximum of confidence in target assignment. With the average pChembl value was 6.76 this interaction was positioned as 19th out of total of 70 drug-KCNH2 interactions. Dronedarone had only one binding interaction described with pChembl value of 6.5, with confidence score of 9 for the assigned target. Quinidine-KCNH2 interaction was displayed with four activities from assays investigating drug-target binding with 9 and 8 confidence score for the assay assigned target (Table 3.6). The average pChembl

Table 3.6: Quinidine-KCNH2 target interaction

Drug CHEMBL ID	Target CHEMBL ID	Assay type	Curation	pChembl value	Avg. pChembl	Confidence score
CHEMBL1294	CHEMBL240	B	Expert	6.49	6.1	9
CHEMBL1294	CHEMBL240	B	Autocuration	6.49	6.1	9
CHEMBL1294	CHEMBL240	B	Autocuration	5.57	6.1	9
CHEMBL1294	CHEMBL240	B	Autocuration	5.84	6.1	8

value for the interaction of 6.1 placed quinidine as 36th drug associated with KCNH2 gene.

Although present, all fampridine-KCNH2 interactions assay data was characterized by activity values annotated as outside of typical range by expert and autocuration data processing, and therefore, they were presented in the data with no pChembl value. The confidence score of target assignment in the assay was 8 and 9 (Table 3.4).

4 Discussion

Studies have been done on the matter of AED effectiveness in mono and polytherapy. The EpiPGX retrospective clinical data was utilized for exploration of the AED usage in several levels (Silvennoinen et al. 2019; Hassanin and Krause 2018; Androsova et al. 2017). In these studies, their effectiveness has been compared in certain epilepsy types. The difference in the retention of AEDs have been related to gender, their MOA and ADRs. However, they have not compared the AEDs usage between two major epilepsy types and two levels of the healthcare system. Moreover, these studies have not addressed the AED-target interactions behind the AED effectiveness as a feature to guide therapy with higher precision. In addition, the drug-repurposing for AED targets has not yet been explored on a larger scale.

Here I show the trend of AEDs usage in time and in generalized and focal epilepsy patients. I also show what AED and AED target-related data could be gathered using public databases and more importantly how easily it could be employed for clinical use.

4.1 Drug usage in EpiPGX data

AEDs are traditionally used drugs in the treatment of epilepsy. The EpiPGX powerfully gathers epilepsy patient data from multiple clinical centers in Europe over a period of around 80 years. Beside AED trials, the patient's data documents also all other known applied drugs in these patients and non-pharmacological treatments in treating epilepsy. Many patients with recorded AED trials were treated with over 10 drugs in the process of finding the best protocol that would obtain the best possible control of the disease symptoms without excessive manifestation of adverse drug reactions. This could be due to the epilepsy types found in these patients which were typically characteristic and demanding for therapy optimization. Although the patients were mostly diagnosed with two major epilepsies, focal and generalized, generalized epilepsy is known to be largely genetically underlined, caused by different gene variants. This contributes to the complexity of diagnose making and adds another layer in therapy choice. The AEDs are currently divided into drugs adequate for treatment of focal and those that are more suited for treatment of generalized epilepsy. Some AEDs have been

recognized as better or worse for epilepsy caused by a specific gene mutation. After the group of AEDs is chosen, elements for fine tuning of the treatment could lay in the match of the drug with the genetic feature of the disease.

The real-world EpiPGX data shows the change in therapy protocols over time. Ethosuximide is an AED used in treatment of childhood absence epilepsy (Glauser et al. 2010). The prescription of ethosuximide has been displayed with an increasing trend over time in the generalized epilepsy patients, meaning, the success in the first trials could have led to later increased application in younger patients with this epilepsy type. Moreover, better understanding of the epilepsy and effects of ethosuximide led to categorization of this drug. On the other hand, there were also trial records of this drug in the focal epilepsy patients. However, these patients were trying approximately 10 drugs in average. Ethosuximide was part of the process of therapy optimization by try-out. Vigabatrin and tiagabine are drugs dedicated for treatment of focal epilepsy and can even worsen the symptoms of the generalized type (Panayiotopoulos et al. 1997; Genton 2000). Nevertheless, there are cases of patients with generalized epilepsy who have tried these two AEDs. As in the case of Ethosuximide, these patients were part of those who have tried in average around 10 AEDs. In addition, according to patient records they have been most likely specific cases where the correct diagnosis was not made at the early stage of the therapy plan.

Another effect of drugs development and medical progress is the superseding of older AEDs characterized with plenty of ADRs by newer promising safer drugs. For instance, valproate was pushed aside by lamotrigine that has a better side effect profile in female patients planning or carrying pregnancy (Silvennoinen et al. 2019). Furthermore, therapy trend differs between different levels of the health care system from the perspective of older versus newer drug generations. Generally, clinical practice seems to outreach at a moment established protocols and try new available drugs in the secondary health care units such as those used in the data gathering of the EpiPGX project. Compared to it, general doctors in the primary health care units tend to prescribe commonly used drugs while the new AEDs are reaching them later. Moreover, different countries have different pace of approving medicines. For example, levetiracetam as an interesting new generation drug with not so well-established mechanism of action yet, has been readily applied in the specialized clinics. According to family practices in the UK, this AED has been reaching general practice in a slower dynamic. In addition to its yet unknown MOA, it was licensed in the UK from 2000 (Nicholas et al. 2012) while the first known approval year for this compound according to ChEMBL is 1999 which contributes to the latency in drug acceptance as common therapy choice. Another example is lamotrigine which despite the better ADR profile takes fourth place in the UK family practices data after valproate compared to the first place in the EpiPGX data. However, this analysis does not consider the epilepsy type or gender distribution. The difference in drug retention could be due to higher response rate in use of

valproate as reported in the study of Silvennoinen et al. (2019).

These events show the need to understand and establish the disease diagnosis on the genomic level in each patient rather than only applying AEDs according to the general epilepsy type in order to aim the key targets at the very beginning of the treatment. There are around 500 epilepsy-related genes that underline major epilepsy types (Wang et al. 2017). The individualized approach to every patient would shorten the period of searching for the optimal therapy and number of drug trials per patient. Although larger data collection because it covers family practices all over the UK compared to the EpiPGX data that covers only specialized clinics in the Europe, the UK data shows much lower use of newer AEDs. The patient-individualized approach would potentially ease the use of newer drugs in the general practices.

4.2 How hard is it to get meaningful data out of ChEMBL

Many different databases that gather drug and target interactions are publicly available for use. However, the ChEMBL database seem to be the one that covers most of them in addition to the enormous scientific literature coverage. Beside the information about the mechanism of action there are established and characterized drug-target interactions by the experimental data and activity information for different assays. ChEMBL shows for AEDs the targets that are considered to be the point of mechanism of their action, and all other targets that have been assigned as the point of interaction in different performed assays.

The assay's information seemed exploitable for the distinction of AEDs and drugs that interact better with the gene of interest from those interacting poorer. Yet, the assays data seemed to be different from drug to drug. Some AEDs such as brivaracetam were not found to have any for the protein-type of targets, although they had a known mechanistic target while others like barbexaclone or beclamide simply lacked both assay and mechanistic target data. Another example is levetiracetam for which the most interesting to investigate and the most reported mechanistic target is SV2a together with CACNA1B (Abou-Khalil 2008). Nevertheless, the experimental data in ChEMBL reports only activities for targets such as EHMT2, ABCB1, ABCB11, ABCC2, ABCC3 and ABCC4, which prevents the characterization of levetiracetam-SV2a interaction and its comparison to the brivaracetam-SV2a. The rest of the AEDs seemed to have different number of assays and recorded activities due to the year of their appearances on the market. The older AEDs have been much more investigated so far than the AEDs that have appeared after the year of 2000. ChEMBL gathers the first approval year for the compound known in any state for any indication. However, lamotrigine, that was on the market in the UK in 1991 ("Lamotrigine (CHEBI:6367)" 2020) while ChEMBL reports the first ever known approval year as 1994, represents one of the data inconsistencies.

The assays offer all possible targets for AEDs, associated in literature and in databases such as DrugMatrix in vitro pharmacology data ("DrugMatrix Database," n.d.). There were few mechanistic AED targets among the assayed. Although the ADME assay type was excluded, some ADME genes were still found in the drug-assayed AED targets interactions data. From the 39 overlapping genes between the ADME, DISEASES and ChEMBL experimental data, CYP2C9, UGT2B7, ALDH5A1, CYP3A4, CYP2C19, ABCB1 were found with epilepsy-gene association confidence score over 2 in the text mining collection. However, only ALDH5A1 was found as pathogenic and causative epilepsy gene ("ALDH5A1[gene] EPILEPSY - ClinVar - NCBI" 2020). The ALDH5A1 encodes succinic semialdehyde dehydrogenase (SSDH) which is involved in the degradation of GABA. The most common disorder of GABA metabolism is caused by SSDH deficiency and patients with mutation in this gene can manifest seizures (Lorenz et al. 2006). Moreover, this gene is considered as mechanistic target interaction for valproate. In addition, it was the only approved compound for this target. Polymorphism in CYP genes and UGT2B7 was found associated with the response-rate to valproate but not as causative epilepsy genes (Feng et al. 2018).

AEDs and their interactions with assayed targets indicate for some drugs a high off-target effects. On the other hand, this opens the possibility that other AED interactions within the organism contribute to the control of the disease. Moreover, this high number of experimentally associated targets enlarges the space for drug repurposing which becomes more and more necessary and popular as the development of new compounds stagnates. Nevertheless, the line is thin between the contribution to the desired and to the side effects. Valproate appeared as one with the highest number of possible target interactions and was recently reported as the drug with considerable ADRs incidence (Silvennoinen et al. 2019). However, lamotrigine, the safer alternative to valproate, shares 95 assayed genes with it. The two drugs could be further investigated for responsible drug-target interactions for the difference in their profiles. Generally, the high number of target interactions could contribute to the rational polytherapy guidelines in treating epilepsy where low dosage of few drugs lowers the manifestation of side effects.

The assays data indicated high druggability of AED assayed targets whether through binding or functionality assays. The odd distribution of drugs per target perhaps could be explain by the composition of this group. The big share of assayed AED targets with over 400 different drug interactions each was mostly composed of neurotransmitter receptors which are typically the G-protein coupled receptors (GPCR) (Levitan et al. 2002). GPCRs represent 50-60% of known drug targets which makes them the most common (Lundstrom 2009). Numerous drugs have been developed based on GPCRs. In addition, their bioassay activity data was assembled using databases that provide information from high-throughput drug screening such as PubChem (Y, T, and Sh 2017). The odd distribution of number of drugs per target would have to be considered and normalized in the statistical testing and

ranking of the drug combinations.

The other important portion of the assayed but also mechanistic AED targets represent the channel proteins. Calcium and sodium channels are also recognized in the drug-target interactions. Whether the channel endures loss or gain of function modification depends on the gene variant itself. The knowledge of the gene variant and its effect would be a significant additional layer of information in the drug-interactions mining. Moreover, it would be an inevitable parameter to make the choice of right drugs in their repurposing to epilepsy.

KCNH2, also known as hERG1, emerged in data as the AED assayed target with the highest number of associated drugs. The gene encodes 4 alpha protein subunits of the potassium voltage-gated channes which are considered as very complex. This gene is abundantly expressed in different tissues ("KCNH2 Gene - GeneCards KCNH2 Protein KCNH2 Antibody" 2020). In the heart muscle tissue, it is involved in the muscle recharging after the impulse with purpose to preserve the regular cardiac rhythm. The mutations in this gene that alter function of the potassium channels are suggested as cause of different arrythmia syndromes like short QT syndrome (SQTS) (Sun et al. 2011). Antiarrhythmics are the group of drugs usually used to prevent and treat abnormal cardiac rhythm. In the case of KCNH2 gain-of-function mutations, the caused channel hyperactivation and cell excitation are treated with drugs such as amiodarone and sotalol that slow down the conduction of impulse in all heart cells ("CV Pharmacology Antiarrhythmic Drugs" 2020). These antiarrhytmics belong to the class 3 and are known as potassium channel blockers. Because of the expression of KCNH2 in the brain and the fact that both arrhythmia and epilepsy belong to ion channelopathies there have been studies to find the connection of KCNH2 mutations to idiopathic epilepsy and some of them report it exists (Partemi et al. 2013). But so far, they only report the loss-of-function mutations in KCNH2 to be related to epilepsy. In this case, amiodarone or other potassium channel blockers would only aggravate the symptoms. However, other hERG type potassium channels such as KCNH1 and KCNH5 are connected to epilepsy when carrying a gain-of-function variants (Niday and Tzingounis 2018). Furthermore, dronedarone which was found to have an interaction with KCNH2, has been reported rather as a channel opener acting on KCNH1 in cancer tissue (Meléndez et al. 2020). This represents a difference compared to the action of it's analogue amiodarone. Aside the action type, although distributed in the brain, both amiodarone and dronedarone are not very lipophilic drugs, meaning the blood-brain barrier represents an issue for their distribution in therapeutic concentration (Brien et al. 2011; Iram et al. 2016).

Drugs with experimentally reported efficacy in treatin epilepsy seizures have been found among approved drugs that are associated with AED-targets. Interactions of fampridine with potassium channels was explored in the retrieved ChEMBL assay data because of his current trials in epilepsy patients ("Experimental Epileptology : Hertie-Institut Für Klinische Hirnforschung" 2020). This licensed

drug has been found effective in reduction of currents through potassium channels induced by gain-of-function variants. Fampridine is indicated in treatment of multiple sclerosis as a neurofunctional modifier to ease the mobility in these patients ("Dalfampridine" 2020). The group investigates its ability to modify epilepsy symptoms by acting on the mutation effect on the KCNA2 channel protein subunit ("Experimental Epileptology : Hertie-Institut Für Klinische Hirnforschung" 2020). However, this gene has not been found among the ChEMBL AED assayed or mechanistic targets. Additionally, the only interaction of this drug with a potassium channel protein was with KCNH2 and was not well described in the ChEMBL data due to estimation of measured activities as outside of typical range. Similarly, experimentaly reported interaction of quinidine with KCNT1 was not found among drug-AED target interactions. Moreover, the KCNT1 target was not among any AED targets. However, quinidine had described interactions with KCNH2 and one of them was curated by expert. Furthermore, the other channels involved in epilepsy channelopathies are also widely distributed in different tissues, equally as the potassium channels. In addition to the necessity to match the drug action type to the channel phenotype, the drug tissue distribution complicates the distinction of epilepsy relevant interactions as brain tissue and CNS differs from the rest. Although some drug-target interactions were found, it was not for the reported targets. Moreover, the available activity information related to these interactions was not sufficient to estimate all of them in detail. Fampridine-KCNH2 interaction did not have the data to enable its comparison with other drugs targeting this gene. As partial potassium channel blocker, quinidine would be inadequate for treatment of loss-of-function phenotype of KCNH2. According to this, it seems that the AED assayed targets data lacks the reported epilepsy related genes and their interactions. To use the drug-target interactions for drug-repurposing and later clinical application requires a lot of further discussion and additional information on the level of gene mutation phenotype and chemical properties of drugs.

Generally, blood barrier is specific by her low permeability achieved by her highly lipophilic nature, and by requirements to allow a drug crossover and drug distribution in the brain tissue. In order to transfer this border a drug molecule has to be small and lipophilic enough for the efflux process or to be a candidate for one of the available transporters which are GLUT1, MCT1, LAT1, CAT1, CNT2 for mediated transport (Pardridge 2012). In order to undergo the efflux a molecule needs to have a logP between 1.5-2.7 (Pajouhesh and Lenz 2005). The molecular weight should be significantly reduced compared to other group of drugs, within 400- 600 Da or even below that range. Most of the CNS licensed drugs have logP around 2.5. and mean weight around 310 MW (Pajouhesh and Lenz 2005). AEDs in the ChEMBL database have a logP from -0.4 to 5.85, and a molecular weight between 129.16 and 387.52 Da. Amiodarone although with a logP of 6.94 has a molecular weight of 645.32 Da. Dronedarone weights 556.77 Da and has logP of 7.05. Low distribution of these drugs in the brain tissue could be because they are not sufficiently small for the active efflux. On the other

hand, fampridine is a very small molecule with only 94.12 Da which helps its transfer even with a lower logP of 0.66. Finally, there is room for fine adjustments of different compound properties beside the solubility and molecular weight, to achieve a certain therapeutic concentration in the tissue.

GABA genes in the AED mechanistic data shows single protein target type, but also a protein complex group of targets that considers many GABA gene components in one target. The assays follow this pattern as well. The protein complex target type challenges the interpretation of the drug-target interactions. Moreover, it makes difficult to distinguish and know with certainty which of the gene components are actually interacting and with which strength. The drug-target interaction could be misinterpreted for the therapy guidelines.

ChEMBL data base enables large-scale data mining. The indicator is the number of drug-AED target interactions. Additionally, the database provides a lot of compound and target related properties. The choice of the parameters to be included in the selection criteria depends on what is considered to have the greatest impact on the drug-target interaction that would significantly contribute to achieve an optimal performance in therapy. Here only few of them were used as proof of concept. The number of binding drugs was reduced using available pChembl value, target type, target confidence scores in assays and disease associations, and data quality parameters in the databases. They could be also ranked by the values of the important features. However, remains the question if the result matches the reality. The outcome list of ranked interactions should be verified by, for instance, comparison to previous rankings of AEDs in the clinical data studies.

4.3 Master thesis work reflection

During the master thesis internship, I have experienced to work on two different projects. The first project has started during the research practical in the third semester. This project involved the investigation of the effectiveness of AED combinations based on their mode of action. The outcome in the AED combinations study raised a question if the retention of these combinations could be explained by the specific AED-target interactions and if the target interactions underline the distinction of better from less retained therapies. Moreover, this question was supported by the idea of treating epilepsy by treating distinct genes causing channelopathies. To transit from one project to the other meant adaptation and further starting research. It also meant to join the two for the thesis. The drug-target interactions and the approach exploiting different databases were the first of this kind in our team. The drug-interactions project would be a proof of concept for the later ion channelopathies project in which the sodium, calcium and potassium channel loss and gain of function mutations would be specifically targeted.

The reproducibility was very important and equally applied as much as possible in both projects. Therefore, both projects were done with the utilization of the GitLab repository. The emphasis in this type of work should be on the good project structure as this enables easier team collaborations and faster analyses. The Git version control allowed me to follow my progress and track changes through commits history. It also allowed discussing different issues with the team members using the Issues section. The opportunity to diverge the work into different branches to separate the work of different investigators was also useful. This was performed by my supervisor Dr. Roland Krause. After the progress was found as desired branches would be converged into the same. Moreover, the EpiPGX study involved dealing with personal patient data which meant that data confidentiality had to be respected. For this purposes, Boxcryptor and ownCloud were used. The encrypted EpiPGX data was stored in ownCloud and accessed using the Boxcryptor. The reproducibility benefits also from the writing of the manuscript in R Markdown (Xie, Allaire, and Grolemund 2018). It allows to visualize the included results while explaining them, it also enables to use the inline code to retrieve the important values directly from the dataframes within the text which in case of any changes in the analysis updates the used values in the comments. Furthermore, it eases the figure and table referencing, content table and reference list updating.

The analysis in R demanded different environment managers. Specifically, Renv, Drake and Here packages were the one I got familiar with. Renv enabled that each team member easily run the analysis regardless of the R version and package versions on their computers. Drake (Landau 2018) came at the time when the amount of ChEMBL data was extremely slowing down the analysis, bringing the machine to the memory vector exhaustion which was highly inconvenient in the search of proper SQL queries. Since the ChEMBL data segregates different information into tables, in order to retrieve the full information on interactions meant to use join statements. Also, in the starting exploration numerous columns had to be retrieved in order to decipher whether the assays matched the target and the compound information correctly. Drake enabled faster analysis run. However, it was not possible to employ `code_to_plan()` for some parts of the previously written code. Furthermore, here package that solves all path dependencies makes the sourcing process easier. For AED-target interactions network ggraph (Pedersen 2020), igraph (Csardi and Nepusz 2006) and qgraph (Epskamp et al. 2012) were the choice. To organize the nodes on the layout as necessary to achieve good readability was challenging. For example, the functions in igraph packages are numerous and while some allow high flexibility the rest have some options fixed. This represented an issue in the layout adjustment and was solved with the use of qgraph (Epskamp et al. 2012). The RSQLite was another important package since it allowed the search of the ChEMBL database using the SQL statements. Within all the data transformation and manipulations the Tidyverse (Wickham 2019) was largely used, while data visualization relied on ggplot2 (Wickham 2016).

As for the analysis itself, the interest in drug-target interactions was essentially the interest in drug-gene associations. The complexity of the data structure made the interpretation problematic. The pChembl value itself although created to ease the activities comparison does not account for the drug action type on the target. Much other available information in this database, its interrelation and meaning has not been clarified yet. Given the shift from one project to another and the complexity of this database and the amount of information it gathers, the internship time frame was sufficient to start the exploration analysis and to get familiar with the database. Nevertheless, it was not enough to do it in sufficient detail nor to reach a ready-to-use product for therapy choice and much less to suggest other drugs that could be efficient in epilepsy treatment.

4.4 Limitations

While this thesis shows the difference in AEDs usage between different epilepsies and healthcare system, and important aspects of the database approach in drug-target interaction mining, it also has limitations. Some limitations are data related.

The EpiPGX data is very valuable as it covers the main epilepsy clinical centers in Europe and over a large period of time. However, it is retrospective clinical data, meaning, it was not possible to account for epilepsy misdiagnosis. Moreover, the comparison of AED use in primary and secondary health care system does not consider the patient gender, age, epilepsy type or AEDs mode of action which were previously shown to play an important role in the AEDs retention (Silvennoinen et al. 2019; Hassanin and Krause 2018; Androsova et al. 2017) .

The ChEMBL database was chosen to investigate drug-target interactions because of its broadness and diversity. However, this results in complexity of the data and its structure in the database. From the beginning, it has been shown as rather difficult to establish drug-target interactions due to the complexity of the drugs and target data structure. Compounds as a group are convoluted due to different molecular forms and their properties. Their interrelations are captured in the MOLECULE_HIERARCHY table as parent molecule and its family members manner. Moreover, it can be noted that some compounds for their forms have many main IDs (parent molecule molregno) while for others the same forms are related as parent-family member molecules. Among AEDs this happens with eslicarbazepine and gabapentine molecules and their acetate salts. For both of these AEDs, one was a prodrug and the other was an active drug form. This raises the question of why acetate forms are in some cases, for example, for desmopressin, related to the neutral parental form, while in other cases are considered as equal to the main molecule form. In addition, beside the chemical nature these also have the bioactivity captured in the ASSAYS and ACTIVITIES tables. Bioactivity is very important feature as it is the connection between the drugs and the targets in the

database. On the other hand, targets are unique by molregno ID but their gene components happen to overlap between differently identified targets, which makes the interpretation of the interactions challenging when the goal is to describe a specific drug-gene interaction. GABA receptors, one of the main players in epilepsy, are particularly vague for interpretation. This complexity is a result of the assays and the uncertainty about which subunit of the protein complex interacts with the tried drug (Gaulton et al. 2012). The solution for that here was the retrieval of the drug-gene interactions across all target IDs containing the same gene and involved in assays with the same drug. However, to just average the pChembl values can oversee some aspects of the assay information, such as, the difference in confidence given by the data curation or the fact that the assay does not specifically suggest the specific and direct interaction with this protein.

The assays data gathers numerous activity types to describe drug binding or function. Although activity types are standardized, the entries are not very well curated yet as, for example, drug potency exists as “Potency” and “potency”. In this work I have not looked into individual activity types. Since they are so different, I rather chose the intercomparable parameter, which is the pChembl value. The pChembl value is calculated for few of the activity types measured in binding and functional assays. However, the variety in activity types might still contain some useful information for a finer distinction of the drug-target interactions. In addition, the confidence scores for the assigned targets in assays might change due to the manual curation process that is still ongoing.

The lack of certain interactions has been noted and it shows that the experimental data in ChEMBL does not contain all the reported scientific information. As it was previously discussed and shown, KCNA2 and KCNT1 targets have been reported as involved in epilepsy (Corbett et al. 2016; Bearden et al. 2014; Møller et al. 2015) but not found in the relation with AEDs in this database. However, the other databases such as PharmGKB do not suggest any drugs for these targets as well. Furthermore, drug-target interactions in ChEMBL are represented in different extent. While some are better studied others are supported with less data due to different factors, possibly related to a shortcoming of the literature but also to the autocuration data process. It seems that to capture all recently discovered epilepsy genes and their drug interactions a manual text mining would be required. The inequality of data quantity would need to be considered in the case of ranking the AED combinations. The inequality of available interactions related information would favor some interactions over others.

The obvious limitation of this work for drug-repurposing or therapy guidelines using AED-target interactions is that it has not looked into the gene variants and their phenotypes. The ChEMBL database contains information for annotated variants in assays. However, it does not put the type of protein modification into a perspective. The PharmGKB associates drugs with the gene variants and provides the variant type but not the phenotype. Moreover, it gathers less data compared to ChEMBL. A solution for this could be the use of another source that presents phenotype, and to bring this layer

of information to the drug-target interactions gathered in ChEMBL. This aspect of drug-target interactions has been shown as crucial when it comes to therapy of rare forms of epilepsies (Hedrich et al. 2019).

The text mining and data autocuration has been present in all the databases. As it exists in ChEMBL and PharmGKB, it exists also in the DISEASES database. The automatic text mining process can have a certain percent of error and confidence scores cannot be as reliable as with the manual abstract curation (Pletscher-Frankild et al. 2015). Although this could be solved by selection criteria of epilepsy-gene confidence score of 4 and 5, genes such as KCNA2 would be lost since they have been less studied. The strict exclusion of the epilepsy-gene associations would give more certain predictions but not the revealing ones. The same applies to the selection criteria for drug-target interactions. Although autocuration lowers the confidence in the data quality, it is necessary nowadays with the fast progress of research and growth of available information in science.

To sum this up, these data quantity and quality aspects are some examples that show the limitations of the database approach itself in the drug-target interactions mining. Additionally, due to the exploratory nature of the project, my work has limitations related to the time frame and experience. However, my thesis integrates different aspects involved in a project of this kind, such as bioinformatics and biology data interpretation. I have tried to contribute to both. The data interpretation emphasizes points to be considered in the new project of ion channel treatments or in any other drug-repurposing related project. The bioinformatics leaves the code to comprehensively and programmatically access the ChEMBL database, integrate it with the DISEASES, and sort the interactions by certain selection criteria. The fact that my thesis was written completely in R distinguishes it from some others. All the figures presenting the results are generated within the project setup and numerical values used in the text are mostly derived directly from the code. This makes the thesis reproducible.

4.5 Outlook

This work definitely shows that there is more to the effectiveness of drugs in any treatment and especially in brain disorders such as epilepsy. So far, AEDs have been used mostly according to the major epilepsy type and recognized pathophysiology. However, treating patients with precision using a pharmacogenomic approach needs finer investigation and distinction of the AED-target interactions. The drug-repurposing would need even higher level of data mining considering drug pharmacokinetics and pharmacodynamics. Therefore, the database approach could be refined with the inclusion of drug chemical properties and gene variants phenotype as features for drug-target interactions selection. In addition, this work shows drugs and AEDs on the level of involved genes, but it would be more precise to see their positioning on the level of gene variants for the purpose of treatment of rare

epilepsy forms and drug-repurposing. Moreover, since previous findings indicate that better disease control is achieved by different MOA-based AED combinations than with the same MOA-based, it would be interesting to see how this is achieved on the level of AED-target interactions. In this work, the AED-targets are collected but the AED-target interactions in different MOA-based combinations have not been investigated. Since the AED-gene variant requires more detailed time-consuming work, the contribution to the AED polytherapy therapy guideline might be done in the meantime. After the formation of different MOA-based AED combinations is done, they could be ranked according to the epilepsy-gene association strength, the number of covered genes, and AED-gene interaction characterization. The epilepsy-gene association strength could be described by the confidence scores from the DISEASES and AED-gene interactions would be characterized by the binding drug affinity captured in pChEMBL value in ChEMBL database. The ranking would be then validated by result comparison with the EpiPGX study. This could be done for epilepsy and epilepsy subtypes, the focal and generalized.

4.6 Conclusion

The application of AEDs has been the only option in treating patients with epilepsy. The traditional therapy cannot always achieve a response in patients with rare forms of the disease underlined by the gene variants. Only some of them have been discovered and annotated in different pharmacogenomic projects and clinics. The prescription of AEDs in general is different in patients with focal and patients with generalized epilepsy. The AED-target interactions could explain this difference on the mechanistic level. In addition, the difference is reflected in AEDs retention in family general practices and specialized clinics. With AED-target interactions included in the therapy protocols and guidelines, the control of symptoms would be achieved faster and in a less painful manner for patients.

My thesis shows the entanglement of the information retrieved from the ChEMBL database and the starting elements in drug-target interactions mining. It also points out some of the important features for the selection of meaningful interactions and lack of information reported in the literature. Different drugs associated with the same potassium channel protein subunit act with a different mechanism. In general, the gene variant phenotype must be matched to the drug mechanism of action for every type of target. Moreover, the structural drug analogs have been reported for exhibiting opposite effects. However, this approach with proper selection criteria and integration of databases could be appropriate to investigate and compare drug-AED target interactions for different drugs targeting the same gene. Nevertheless, to go from the database search to a clinical application in epilepsy patients, drug-target interactions mining requires a significant amount of discussion and work.

5 References

Abou-Khalil, Bassel. 2008. “Levetiracetam in the Treatment of Epilepsy.” *Neuropsychiatric Disease and Treatment* 4 (3): 507–23. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2526377/.

“About Us European Bioinformatics Institute.” 2020. Accessed October 30. https://www.ebi.ac.uk/about.

“ALDH5A1[gene] EPILEPSY - ClinVar - NCBI.” 2020. Accessed October 29. https://www.ncbi.nlm.nih.gov/clinvar/?term=ALDH5A1%5Bgene%5D++EPILEPSY.

Androsova, Ganna, Roland Krause, Mojgansadat Borghei, Merel Wassenaar, Pauls Auce, Andreja Avbersek, Felicitas Becker, et al. 2017. “Comparative Effectiveness of Antiepileptic Drugs in Patients with Mesial Temporal Lobe Epilepsy with Hippocampal Sclerosis.” *Epilepsia* 58 (10): 1734–41. doi:10.1111/epi.13871.

Balestrini, Simona, and Sanjay M. Sisodiya. 2018. “Pharmacogenomics in Epilepsy.” *Neuroscience Letters* 667 (February): 27–39. doi:10.1016/j.neulet.2017.01.014.

Bearden, David, Alanna Strong, Jessica Ehnot, Marissa DiGiovine, Dennis Dlugos, and Ethan M. Goldberg. 2014. “Targeted Treatment of Migrating Partial Seizures of Infancy with Quinidine: MPSI and Quinidine.” *Annals of Neurology* 76 (3): 457–61. doi:10.1002/ana.24229.

Berkovic, Samuel F., John C. Mulley, Ingrid E. Scheffer, and Steven Petrou. 2006. “Human Epilepsies: Interaction of Genetic and Acquired Factors.” *Trends in Neurosciences* 29 (7): 391–97. doi:10.1016/j.tins.2006.05.009.

Bosc, Nicolas, Francis Atkinson, Eloy Felix, Anna Gaulton, Anne Hersey, and Andrew R. Leach. 2019. “Large Scale Comparison of QSAR and Conformal Prediction Methods and Their Applications in Drug Discovery.” *Journal of Cheminformatics* 11 (1): 4. doi:10.1186/s13321-018-0325-4.

Brien, James F., Sandra Jimmo, F. James Brennan, Sally E. Ford, and Paul W. Armstrong. 2011. “Distribution of Amiodarone and Its Metabolite, Desethylamiodarone, in Human Tissues.” *Canadian Journal of Physiology and Pharmacology*, February. doi:10.1139/y87-062.

Brodie, Martin J. 2010. “Antiepileptic Drug Therapy the Story so Far.” *Seizure* 19 (10): 650–55. doi:10.1016/j.seizure.2010.10.027.

Brodie, M. J., and J. A. French. 2000. “Management of Epilepsy in Adolescents and Adults.” *Lancet (London, England)* 356 (9226): 323–29. doi:10.1016/S0140-6736(00)02515-0.

“ChEMBL Data Questions.” 2020. Accessed October 29. https://chembl.gitbook.io/chembl-interface-documentation/frequently-asked-questions/chembl-data-questions.

“Chloroquine DrugBank Online.” 2020. Accessed October 28. https://go.drugbank.com/categories/DBCAT001399.

Chong, Derek J., and Carl W. Bazil. 2010. “Update on Anticonvulsant Drugs.” *Current Neurology and Neuroscience Reports* 10 (4): 308–18. doi:10.1007/s11910-010-0120-8.

“Concept: Person Years - Calculating in a Cohort Study.” 2020. Accessed October 27. http://mchp-appserv.cpe.umanitoba.ca/viewConcept.php?printer=Y&conceptID=1196.

Corbett, Mark A., Susannah T. Bellows, Melody Li, Renée Carroll, Silvana Micallef, Gemma L. Carvill, Candace T. Myers, et al. 2016. “Dominant KCNA2 Mutation Causes Episodic Ataxia and Pharmacoresponsive Epilepsy.” *Neurology* 87 (19): 1975–84. doi:10.1212/WNL.0000000000003309.

Csardi, Gabor, and Tamas Nepusz. 2006. “The Igraph Software Package for Complex Network Research.” *InterJournal* Complex Systems: 1695. http://igraph.org.

“CV Pharmacology Antiarrhythmic Drugs.” 2020. Accessed October 27. https://www.cvpharmacology.com/antiarrhy/antiarrhythmic.

Czuczwar, Stanisław J, and Kinga K Borowicz. 2002. “Polytherapy in Epilepsy: The Experimental Evidence.” *Epilepsy Research* 52 (1): 15–23. doi:10.1016/S0920-1211(02)00181-X.

“Dalfampridine.” 2020. Accessed October 27. https://go.drugbank.com/drugs/DB06637.

Daneman, Richard, and Alexandre Prat. 2015. “The Blood–Brain Barrier.” *Cold Spring Harbor Perspectives in Biology* 7 (1): a020412. doi:10.1101/cshperspect.a020412.

Daniels, V, M Wood, K Leclercq, R M Kaminski, and M Gillard. 2013. “Modulation of the Conformational State of the SV2A Protein by an Allosteric Mechanism as Evidenced by Ligand Binding Assays: Allosteric Modulation of SV2A.” *British Journal of Pharmacology* 169 (5): 1091–1101. doi:10.1111/bph.12192.

Doogue, Matthew P., and Thomas M. Polasek. 2013. “The ABCD of Clinical Pharmacokinetics.” *Therapeutic Advances in Drug Safety* 4 (1): 5–7. doi:10.1177/2042098612469335.

Dovrolis, Nikolas, George Kolios, George Spyrou, and Ioanna Maroulakou. 2017. “Laying in Silico Pipelines for Drug Repositioning: A Paradigm in Ensemble Analysis for Neurodegenerative Diseases.” *Drug Discovery Today* 22 (5): 805–13. doi:10.1016/j.drudis.2017.03.009.

“DrugMatrix Database.” n.d., 40.

Engelborghs, Sebastiaan, Rudi D’Hooge, and De Paul. 2001. “Pathophysiology of Epilepsy.” *Acta Neurologica Belgica* 100 (January): 201–13.

“Epilepsy.” 2020. Accessed October 28. https://www.who.int/news-room/fact-sheets/detail/epilepsy.

“EpiPGX – Epilepsy Biomarkers for Clinical Use.” 2020. Accessed October 28. https://www.epipgx.eu/.

Epskamp, Sacha, Angélique O. J. Cramer, Lourens J. Waldorp, Verena D. Schmittmann, and Denny Borsboom. 2012. “qgraph: Network Visualizations of Relationships in Psychometric Data.” *Journal of Statistical Software* 48 (4): 1–18. http://www.jstatsoft.org/v48/i04/.

“Experimental Epileptology : Hertie-Institut Für Klinische Hirnforschung.” 2020. Accessed October 26. https://www.hih-tuebingen.de/en/research/neurology-and-epileptology/research-groups/experimental-epileptology/.

Feng, Weixing, Shenghui Mei, Leting Zhu, Yazhen Yu, Weili Yang, Baoqin Gao, Xiaojuan Wu, Zhigang Zhao, and Fang Fang. 2018. “Effects of UGT2B7, SCN1A and CYP3A4 on the Therapeutic Response of Sodium Valproate Treatment in Children with Generalized Seizures.” *Seizure* 58 (May): 96–100. doi:10.1016/j.seizure.2018.04.006.

Fisher, Robert S., Carlos Acevedo, Alexis Arzimanoglou, Alicia Bogacz, J. Helen Cross, Christian E. Elger, Jerome Engel, et al. 2014. “ILAE Official Report: A Practical Clinical Definition of Epilepsy.” *Epilepsia* 55 (4): 475–82. doi:10.1111/epi.12550.

Fisher, Robert S., J. Helen Cross, Jacqueline A. French, Norimichi Higurashi, Edouard Hirsch, Floor E. Jansen, Lieven Lagae, et al. 2017. “Operational Classification of Seizure Types by the International League Against Epilepsy: Position Paper of the ILAE Commission for Classification and Terminology.” *Epilepsia* 58 (4): 522–30. doi:10.1111/epi.13670.

Gaulton, Anna, Louisa J. Bellis, A. Patricia Bento, Jon Chambers, Mark Davies, Anne Hersey, Yvonne Light, et al. 2012. “ChEMBL: A Large-Scale Bioactivity Database for Drug Discovery.” *Nucleic Acids Research* 40 (Database issue): D1100–D1107. doi:10.1093/nar/gkr777.

Gaulton, Anna, Anne Hersey, Michał Nowotka, A. Patrícia Bento, Jon Chambers, David Mendez, Prudence Mutowo, et al. 2017. “The ChEMBL Database in 2017.” *Nucleic Acids Research* 45 (D1): D945–D954. doi:10.1093/nar/gkw1074.

“General Practice Research Database.” 2020. Accessed October 29. http://www.gprd.com/home/default.html.

Genton, Pierre. 2000. “When Antiepileptic Drugs Aggravate Epilepsy.” *Brain and Development* 22 (2): 75–80. doi:10.1016/S0387-7604(99)00113-8.

Glauser, Tracy A., Avital Cnaan, Shlomo Shinnar, Deborah G. Hirtz, Dennis Dlugos, David Masur, Peggy O. Clark, Edmund V. Capparelli, and Peter C. Adamson. 2010. “Ethosuximide, Valproic Acid, and Lamotrigine in Childhood Absence Epilepsy.” *New England Journal of Medicine* 362 (9): 790–99. doi:10.1056/NEJMoa0902014.

Grant Augustus O. 2009. “Cardiac Ion Channels.” *Circulation: Arrhythmia and Electrophysiology* 2 (2): 185–94. doi:10.1161/CIRCEP.108.789081.

Grolemund, Garrett, and Hadley Wickham. 2017. *R for Data Science*. https://r4ds.had.co.nz/.

Hassanin, Emadeldin, and Roland Krause. 2018. “Comparative Effectiveness of Antiepileptic Combination Therapy Based on Mode of Action.” PhD thesis.

Hedrich, Ulrike B. S., Henner Koch, Albert Becker, and Holger Lerche. 2019. “Epileptogenese Und Konsequenzen Für Die Therapie.” *Der Nervenarzt* 90 (8): 773–80. doi:10.1007/s00115-019-0749-8.

Iram, Farah, Sadaf Ali, Aftab Ahmad, Shah Alam Khan, and Asif Husain. 2016. “A Review on Dronedarone: Pharmacological, Pharmacodynamic and Pharmacokinetic Profile.” *Journal of Acute Disease* 5 (2): 102–8. doi:10.1016/j.joad.2015.10.002.

Janz, Roger, Yukiko Goda, Martin Geppert, Markus Missler, and Thomas C Südhof. 1999. “SV2A and SV2B Function as Redundant Ca2+ Regulators in Neurotransmitter Release.” *Neuron* 24 (4): 1003–16. doi:10.1016/S0896-6273(00)81046-6.

Jourdan, Jean□Pierre, Ronan Bureau, Christophe Rochais, and Patrick Dallemagne. 2020. “Drug Repositioning: A Brief Overview.” *Journal of Pharmacy and Pharmacology* 72 (9): 1145–51. doi:10.1111/jphp.13273.

Judge, Susan I. V., and Christopher T. Bever. 2006. “Potassium Channel Blockers in Multiple Sclerosis: Neuronal Kv Channels and Effects of Symptomatic Treatment.” *Pharmacology & Therapeutics* 111 (1): 224–59. doi:10.1016/j.pharmthera.2005.10.006.

"KCNH2 Gene - GeneCards KCNH2 Protein KCNH2 Antibody." 2020. Accessed October 27. https://www.genecards.org/cgi-bin/carddisp.pl?gene=KCNH2#expression.

Kim, Sunghwan, Jie Chen, Tiejun Cheng, Asta Gindulyte, Jia He, Siqian He, Qingliang Li, et al. 2019. "PubChem 2019 Update: Improved Access to Chemical Data." *Nucleic Acids Research* 47 (D1): D1102–D1109. doi:10.1093/nar/gky1033.

Kwan, Patrick, and Martin J. Brodie. 2000. "Early Identification of Refractory Epilepsy." *New England Journal of Medicine* 342 (5): 314–19. doi:10.1056/NEJM200002033420503.

Lagier, Jean-Christophe, Matthieu Million, Philippe Gautret, Philippe Colson, Sébastien Cortaredona, Audrey Giraud-Gatineau, Stéphane Honoré, et al. 2020. "Outcomes of 3,737 COVID-19 Patients Treated with Hydroxychloroquine/Azithromycin and Other Regimens in Marseille, France: A Retrospective Analysis." *Travel Medicine and Infectious Disease* 36 (July): 101791. doi:10.1016/j.tmaid.2020.101791.

"Lamotrigine (CHEBI:6367)." 2020. Accessed October 26. https://www.ebi.ac.uk/chebi/searchId.do?chebiId=CHEBI:6367.

Landau, William Michael. 2018. "The Drake R Package: A Pipeline Toolkit for Reproducibility and High-Performance Computing." *Journal of Open Source Software* 3 (21). https://doi.org/10.21105/joss.00550.

Lerche, Holger, Karin Jurkat-Rott, and Frank Lehmann-Horn. 2001. "Ion Channels and Epilepsy." *American Journal of Medical Genetics* 106 (2): 146–59. doi:10.1002/ajmg.1582.

Levitan, Irwin B., Founding Chair of the Department of Neuroscience Irwin B. Levitan Ph.D, Irwin B. Levitan, and Leonard K. Kaczmarek. 2002. *The Neuron: Cell and Molecular Biology*. Oxford University Press.

Li, Jing, Luyong Zhang, Hang Zhou, Mark Stoneking, and Kun Tang. 2011. "Global Patterns of Genetic Diversity and Signals of Natural Selection for Human ADME Genes." *Human Molecular Genetics* 20 (3): 528–40. doi:10.1093/hmg/ddq498.

Lorenz, Susanne, Armin Heils, Kirsten P. Taylor, Anne Gehrmann, Hiltrud Muhle, Meike Gresch, Tim Becker, Ulrike Tauer, Ulrich Stephani, and Thomas Sander. 2006. "Candidate Gene Analysis of the Succinic Semialdehyde Dehydrogenase Gene (ALDH5A1) in Patients with Idiopathic Generalized Epilepsy and Photosensitivity." *Neuroscience Letters* 397 (3): 234–39. doi:10.1016/j.neulet.2005.12.030.

Löscher, Wolfgang, Michel Gillard, Zara A. Sands, Rafal M. Kaminski, and Henrik Klitgaard. 2016. “Synaptic Vesicle Glycoprotein 2A Ligands in the Treatment of Epilepsy and Beyond.” *CNS Drugs* 30 (11): 1055–77. doi:10.1007/s40263-016-0384-x.

Lundstrom, Kenneth. 2009. “An Overview on GPCRs and Drug Discovery: Structure-Based Drug Design and Structural Biology on GPCRs.” In *G Protein-Coupled Receptors in Drug Discovery*, edited by Wayne R. Leifert, 51–66. Methods in Molecular Biology. Totowa, NJ: Humana Press. doi:10.1007/978-1-60327-317-6_4.

Margolis, Jay M., Bong-Chul Chu, Zhixiao J. Wang, Ronda Copher, and Jose E. Cavazos. 2014. “Effectiveness of Antiepileptic Drug Combination Therapy for Partial-Onset Seizures Based on Mechanisms of Action.” *JAMA Neurology* 71 (8): 985. doi:10.1001/jamaneurol.2014.808.

Mazarati, A. 2009. “Neuropeptides: Epilepsy.” In *Encyclopedia of Neuroscience*, 907–14. Elsevier. doi:10.1016/B978-008045046-9.01464-9.

Mehrabi, Soraya, Nima Sanadgol, Mahmood Barati, Ali Shahbazi, Gelareh Vahabzadeh, Mitra Barzroudi, Morteza Seifi, Mazaher Gholipourmalekabadi, and Fereshteh Golab. 2018. “Evaluation of Metformin Effects in the Chronic Phase of Spontaneous Seizures in Pilocarpine Model of Temporal Lobe Epilepsy.” *Metabolic Brain Disease* 33 (1): 107–14. doi:10.1007/s11011-017-0132-z.

Meléndez, T. A., A. Huanosta-Gutiérrez, C. Barriga-Montoya, M. González-Andrade, and F. Gómez-Lagunas. 2020. “Dronedarone Blockage of the Tumor-Related Kv10.1 Channel: A Comparison with Amiodarone.” *Pflügers Archiv - European Journal of Physiology* 472 (1): 75–87. doi:10.1007/s00424-019-02342-8.

Møller, Rikke S., Sarah E. Heron, Line H. G. Larsen, Chiao Xin Lim, Michael G. Ricos, Marta A. Bayly, Marjan J. A. van Kempen, et al. 2015. “Mutations in KCNT1 Cause a Spectrum of Focal Epilepsies.” *Epilepsia* 56 (9): e114–e120. doi:10.1111/epi.13071.

Narayan, Ravi S., Piet Molenaar, Jian Teng, Fleur M. G. Cornelissen, Irene Roelofs, Renee Menezes, Rogier Dik, et al. 2020. “A Cancer Drug Atlas Enables Synergistic Targeting of Independent Drug Vulnerabilities.” *Nature Communications* 11 (1): 2935. doi:10.1038/s41467-020-16735-2.

Nicholas, Jennifer M., Leone Ridsdale, Mark P. Richardson, Mark Ashworth, and Martin C. Gulliford. 2012. “Trends in Antiepileptic Drug Utilisation in UK Primary Care 1993–2008: Cohort Study Using the General Practice Research Database.” *Seizure* 21 (6): 466–70. doi:10.1016/j.seizure.2012.04.014.

Niday, Zachary, and Anastasios V. Tzingounis. 2018. “Potassium Channel Gain of Function in Epilepsy: An Unresolved Paradox.” *The Neuroscientist: A Review Journal Bringing Neurobiology, Neurology and Psychiatry* 24 (4): 368–80. doi:10.1177/1073858418763752.

Pajouhesh, Hassan, and George R. Lenz. 2005. “Medicinal Chemical Properties of Successful Central Nervous System Drugs.” *NeuroRx* 2 (4): 541–53. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1201314/.

Panayiotopoulos, C. P. 2005. *Principles of Therapy in Epilepsies*. Bladon Medical Publishing. https://www.ncbi.nlm.nih.gov/books/NBK2607/.

Panayiotopoulos, C. P., A. Agathonikou, I. Ahmed Sharoqi, and A. P. J. Parker. 1997. “Vigabatrin Aggravates Absences and Absence Status.” *Neurology* 49 (5): 1467.1–1467. doi:10.1212/WNL.49.5.1467.

Papadatos, George, Anna Gaulton, Anne Hersey, and John P. Overington. 2015. “Activity, Assay and Target Data Curation and Quality in the ChEMBL Database.” *Journal of Computer-Aided Molecular Design* 29 (9): 885–96. doi:10.1007/s10822-015-9860-5.

Pardridge, William M. 2012. “Drug Transport Across the Blood–Brain Barrier.” *Journal of Cerebral Blood Flow & Metabolism* 32 (11): 1959–72. doi:10.1038/jcbfm.2012.126.

Partemi, Sara, Sandrine Cestèle, Marianna Pezzella, Oscar Campuzano, Roberta Paravidino, Vincenzo L. Pascali, Federico Zara, et al. 2013. “Loss-of-Function KCNH2 Mutation in a Family with Long QT Syndrome, Epilepsy, and Sudden Death.” *Epilepsia* 54 (8): e112–e116. doi:10.1111/epi.12259.

Pearce, J. M. S. 2002. “Bromide, the First Effective Antiepileptic Agent.” *Journal of Neurology, Neurosurgery & Psychiatry* 72 (3): 412–12. doi:10.1136/jnnp.72.3.412.

Pedersen, Thomas Lin. 2020. *Ggraph: An Implementation of Grammar of Graphics for Graphs and Networks*. https://CRAN.R-project.org/package=ggraph.

Pitkänen, Asla, Katarzyna Lukasiuk, F. Edward Dudek, and Kevin J. Staley. 2015. “Epileptogenesis.” *Cold Spring Harbor Perspectives in Medicine* 5 (10): a022822. doi:10.1101/cshperspect.a022822.

Pletscher-Frankild, Sune, Albert Pallejà, Kalliopi Tsafou, Janos X. Binder, and Lars Juhl Jensen. 2015. “DISEASES: Text Mining and Data Integration of Disease–Gene Associations.” *Methods*, Text mining of biomedical literature, 74 (March): 83–89. doi:10.1016/j.ymeth.2014.11.020.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Sills, Graeme J., and Martin J. Brodie. 2002. “Update on the Mechanisms of Action of Antiepileptic Drugs.” *Epileptic Disorders* 3 (4): 165–72. http://www.jle.com/fr/revues/epd/e-docs/update_on_the_mechanisms_of_action_of_antiepileptic_drugs_110045/article.phtml?tab=texte.

Silvennoinen, Katri, Nikola Lange, Sara Zagaglia, Simona Balestrini, Ganna Androsova, Merel Wassenaar, Pauls Auce, et al. 2019. "Comparative Effectiveness of Antiepileptic Drugs in Juvenile Myoclonic Epilepsy." *Epilepsia Open* 4 (3): 420–30. doi:10.1002/epi4.12349.

Sloviter, Robert S., and Argyle V. Bumanglag. 2013. "Defining 'Epileptogenesis' and Identifying 'Antiepileptogenic Targets' in Animal Models of Acquired Temporal Lobe Epilepsy Is Not as Simple as It Might Seem." *Neuropharmacology* 69 (June): 3–15. doi:10.1016/j.neuropharm.2012.01.022.

St. Louis, Erik K. 2009. "Truly 'Rational' Polytherapy: Maximizing Efficacy and Minimizing Drug Interactions, Drug Load, and Adverse Effects." *Current Neuropharmacology* 7 (2): 96–105. doi:10.2174/157015909788848929.

Sun, Yaxun, Xiao-Qing Quan, Samantha Fromme, Robert H. Cox, Ping Zhang, Li Zhang, Donglin Guo, et al. 2011. "A Novel Mutation in the KCNH2 Gene Associated with Short QT Syndrome." *Journal of Molecular and Cellular Cardiology* 50 (3): 433–41. doi:10.1016/j.yjmcc.2010.11.017.

Turski, L., C. Ikonomidou, W. A. Turski, Z. A. Bortolotto, and E. A. Cavalheiro. 1989. "Review: Cholinergic Mechanisms and Epileptogenesis. The Seizures Induced by Pilocarpine: A Novel Experimental Model of Intractable Epilepsy." *Synapse (New York, N.Y.)* 3 (2): 154–71. doi:10.1002/syn.890030207.

"Unpaired Two-Samples Wilcoxon Test in R - Easy Guides - Wiki - STHDA." 2020. Accessed October 30. http://www.sthda.com/english/wiki/unpaired-two-samples-wilcoxon-test-in-r.

Ushey, Kevin. 2020. *Renv: Project Environments*. https://CRAN.R-project.org/package=renv.

Vihinen, Mauno. 2020. "Functional Effects of Protein Variants." *Biochimie*, October. doi:10.1016/j.biochi.2020.10.

Wang, Jie, Zhi-Jian Lin, Liu Liu, Hai-Qing Xu, Yi-Wu Shi, Yong-Hong Yi, Na He, and Wei-Ping Liao. 2017. "Epilepsy-Associated Genes." *Seizure* 44 (January): 11–20. doi:10.1016/j.seizure.2016.11.030.

Whirl-Carrillo, M., E. M. McDonagh, J. M. Hebert, L. Gong, K. Sangkuhl, C. F. Thorn, R. B. Altman, and T. E. Klein. 2012. "Pharmacogenomics Knowledge for Personalized Medicine." *Clinical Pharmacology and Therapeutics* 92 (4): 414–17. doi:10.1038/clpt.2012.96.

"WHOCC - Structure and Principles." 2020. Accessed October 9. https://www.whocc.no/atc/structure_and_principles/.

Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. https://ggplot2.tidyverse.org.

———. 2019. *Tidyverse: Easily Install and Load the 'Tidyverse'*. https://CRAN.R-project.org/package=tidyverse.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. doi:10.21105/joss.01686.

"Www.pharmaadme.org - Home." 2020. Accessed October 29. http://www.pharmaadme.org/joomla/.

Xie, Yihui, J. J. Allaire, and Garrett Grolemund. 2018. *R Markdown: The Definitive Guide*. Boca Raton, Florida: Chapman; Hall/CRC. https://bookdown.org/yihui/rmarkdown.

Y, Wang, Cheng T, and Bryant Sh. 2017. "PubChem BioAssay: A Decade's Development Toward Open High-Throughput Screening Data Sharing." *SLAS Discovery : Advancing Life Sciences R & D* 22 (6): 655–66. doi:10.1177/2472555216685069.

Yimer, Ebrahim M., Awol Surur, Dawit Zewdu Wondafrash, and Abadi Kahsu Gebre. 2019. "The Effect of Metformin in Experimentally Induced Animal Models of Epileptic Seizure." Review Article. *Behavioural Neurology*. doi:https://doi.org/10.1155/2019/6234758.

6 Appendix

Table 6.1: AED assayed targets found in ADME genes list

Gene	Type
ABCB1	core
ABCB11	extended
ABCC1	extended
ABCC2	core
ABCC3	extended
ABCC4	extended
ALDH1A1	extended
ALDH5A1	extended
AOX1	extended
ARSA	extended
CES1	extended
CES2	extended
CYP1A1	core
CYP1A2	core
CYP2A6	core
CYP2B6	core
CYP2C19	core
CYP2C8	core
CYP2C9	core
CYP2D6	core
CYP2E1	core
CYP2J2	extended
CYP3A4	core
NR1I2	extended
NR1I3	extended
PPARG	extended
SLC22A1	core
SLC22A5	extended
SLC22A8	extended
SLCO1B1	core
SLCO1B3	core
SLCO2B1	extended
UGT1A1	core
UGT1A10	extended
UGT1A3	extended
UGT1A4	extended
UGT1A6	extended
UGT1A8	extended
UGT1A9	extended
UGT2A1	extended
UGT2B10	extended
UGT2B15	core
UGT2B4	extended
UGT2B7	core

Table 6.2: Assayed AED targets with 400-550 assayed drugs.

TID	Uniprot gene name	Target-epilepsy confidence score	DISEASE source collection	Target name in ChEMBL	Number of approved drugs per target	Number of assays per gene
11755	FYN	1.819	Text mining	Tyrosine-protein kinase FYN	549	
10140	LCK	1.400	Text mining	Tyrosine-protein kinase LCK	548	
11639	MAPK3	2.004	Text mining	MAP kinase ERK1	548	
72	DRD2	2.378	Text mining	Dopamine D2 receptor	545	1272
107	HTR2A	2.187	Text mining	Serotonin 2a (5-HT2a) receptor	533	713
15	CA2	2.266	Text mining	Carbonic anhydrase II	531	1708
108	HTR2C	2.099	Text mining	Serotonin 2c (5-HT2c) receptor	531	684
129	OPRM1	2.163	Text mining	Mu opioid receptor	531	842
19	ESR1	1.602	Text mining	Estrogen receptor alpha	530	1076
86	MAOA	1.952	Text mining	Monoamine oxidase A	530	726
130	DRD3	2.225	Text mining	Dopamine D3 receptor	529	788
188	ERBB2	1.273	Text mining	Receptor protein-tyrosine kinase erbB-2	529	685
121	SLC6A4	2.240	Text mining	Serotonin transporter	529	858
10979	FLT1	1.192	Text mining	Vascular endothelial growth factor receptor 1	528	602
43	ADRB2	1.656	Text mining	Beta-2 adrenergic receptor	527	785
127	HRH1	1.633	Text mining	Histamine H1 receptor	527	671
25	NR3C1	1.741	Text mining	Glucocorticoid receptor	527	1111
61	CHRM1	1.440	Text mining	Muscarinic acetylcholine receptor M1	525	636
155	SLC6A3	2.137	Text mining	Dopamine transporter	525	751
136	OPRD1	1.925	Text mining	Delta opioid receptor	524	735
100	SLC6A2	1.728	Text mining	Norepinephrine transporter	524	792
47	CHRM2	1.340	Text mining	Muscarinic acetylcholine receptor M2	523	598
90	DRD4	2.251	Text mining	Dopamine D4 receptor	523	728
93	ACHE	2.368	Text mining	Acetylcholinesterase	522	1076
137	OPRK1	2.011	Text mining	Kappa opioid receptor	521	734
52	ADRA2A	1.563	Text mining	Alpha-2a adrenergic receptor	520	614
219	CHRM3	1.364	Text mining	Muscarinic acetylcholine receptor M3	520	565
219	CHRM3	2.999	Experiments	Muscarinic acetylcholine receptor M3	520	565
214	CHRM4	1.208	Text mining	Muscarinic acetylcholine receptor M4	520	552
11272	SIGMAR1	1.250	Text mining	Sigma opioid receptor	520	614
10627	HTR6	1.253	Text mining	Serotonin 6 (5-HT6) receptor	519	588
126	PTGS2	2.136	Text mining	Cyclooxygenase-2	519	1134
227	HTR2B	1.722	Text mining	Serotonin 2b (5-HT2b) receptor	518	557
218	ADRA2C	1.533	Text mining	Alpha-2c adrenergic receptor	517	583
50	ADRB1	1.551	Text mining	Beta-1 adrenergic receptor	517	631
215	CHRM5	0.908	Text mining	Muscarinic acetylcholine receptor M5	516	550
96	PTGS1	1.249	Text mining	Cyclooxygenase-1	515	865
249	EDNRA	1.236	Text mining	Endothelin receptor ET-A	514	542
174	ESR2	1.346	Text mining	Estrogen receptor beta	514	790
103	ADRA1D	1.691	Text mining	Alpha-1d adrenergic receptor	513	569
216	ADRA2B	1.643	Text mining	Alpha-2b adrenergic receptor	512	546
216	ADRA2B	4.000	Knowledge	Alpha-2b adrenergic receptor	512	546
134	AVPR1A	1.148	Text mining	Vasopressin V1a receptor	512	537
10142	MC4R	1.430	Text mining	Melanocortin receptor 4	512	571
12592	MMP9	2.002	Text mining	Matrix metalloproteinase 9	512	520
87	CNR1	2.755	Text mining	Cannabinoid CB1 receptor	511	696
235	ELANE	0.889	Text mining	Leukocyte elastase	511	519
3	PDE5A	1.708	Text mining	Phosphodiesterase 5A	511	603
10692	AGTR2	1.295	Text mining	Angiotensin II type 2 (AT-2) receptor	510	529
11624	CASP1	1.934	Text mining	Caspase-1	510	526
10494	CTSG	0.731	Text mining	Cathepsin G	510	516
24	HMGCR	1.044	Text mining	HMG-CoA reductase	510	586
252	ADORA2A	2.388	Text mining	Adenosine A2a receptor	509	583
280	ADORA3	2.355	Text mining	Adenosine A3 receptor	509	554
226	ADRB3	1.384	Text mining	Beta-3 adrenergic receptor	509	569
102	HRH2	1.340	Text mining	Histamine H2 receptor	509	522
13000	MMP1	0.905	Text mining	Matrix metalloproteinase-1	509	513
251	PTAFR	0.807	Text mining	Platelet activating factor receptor	509	514
250	TACR1	1.566	Text mining	Neurokinin 1 receptor	509	528
114	ADORA1	2.497	Text mining	Adenosine A1 receptor	508	560
10034	BDKRB2	0.859	Text mining	Bradykinin B2 receptor	508	520
11575	CCR2	1.538	Text mining	C-C chemokine receptor type 2	508	515
10580	CCR5	1.397	Text mining	C-C chemokine receptor type 5	508	526
20073	CYP2A6	1.481	Text mining	Cytochrome P450 2A6	508	511
11003	MC3R	1.230	Text mining	Melanocortin receptor 3	508	532
11006	MC5R	1.151	Text mining	Melanocortin receptor 5	508	528
10475	NPY1R	1.550	Text mining	Neuropeptide Y receptor type 1	508	521
12088	PTPRC	1.595	Text mining	Leukocyte common antigen	508	512
248	TBXAS1	0.704	Text mining	Thromboxane-A synthase	508	524
119	CALCR	1.058	Text mining	Calcitonin receptor	507	510
10472	CCKAR	0.751	Text mining	Cholecystokinin A receptor	507	512
10579	CCR4			C-C chemokine receptor type 4	507	518
11574	CXCR1	0.703	Text mining	Interleukin-8 receptor A	507	518
10773	CXCR2	0.978	Text mining	Interleukin-8 receptor B	507	510
20056	CYP2E1	1.401	Text mining	Cytochrome P450 2E1	507	510
179	CYSLTR1			Cysteinyl leukotriene receptor 1	507	521
10477	NPY2R	1.627	Text mining	Neuropeptide Y receptor type 2	507	515
12579	PPP3CA	1.226	Text mining	Serine/threonine protein phosphatase 2B catalytic subunit, alpha isoform	507	513
12579	PPP3CA	4.000	Knowledge	Serine/threonine protein phosphatase 2B catalytic subunit, alpha isoform	507	513
10184	TACR2	0.911	Text mining	Neurokinin 2 receptor	507	510
100334	VIPR1	0.995	Text mining	Vasoactive intestinal polypeptide receptor 1	507	510
103711	GMNN	0.745	Text mining	Geminin	484	736
103668	LMNA	1.534	Text mining	Prelamin-A/C	420	462
102672	EHMT2	1.440	Text mining	Histone-lysine N-methyltransferase, H3 lysine-9 specific 3	418	441