UNIVERSITY OF LIVERPOOL

Predicting drugs for epilepsy using genetic and genomic data

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

By

Basel Taweel

August 2021

Table of Contents

Table of Contents ..... ***ii***

Acknowledgments ..... ***v***

Abstract ..... ***vi***

List of Tables ..... ***vii***

List of Figures ..... ***viii***

Abbreviations ..... ***x***

Chapter 1 Introduction to Epilepsy, Genomics and Drug repurposing ***1***

1.1 Introduction to Epilepsy **1**

1.1.1 Epidemiology, prevalence and distribution 1

1.1.2 Seizures and seizure types 1

1.1.3 Common and rarer epilepsies 2

1.1.4 Current Management modalities of Epilepsy 3

1.2 Introduction to novel drug development and drug repurposing **5**

1.2.1 Target identification 5

1.2.2 Target validation 6

1.2.3 Hit discovery 8

1.2.4 Lead generation and Lead optimisation 8

1.2.5 Candidate identification and preclinical development 11

1.2.6 Clinical testing 13

1.3 About drug repurposing **14**

1.3.1. Challenges of drug repurposing 15

1.3.2 Drug repurposing approaches 16

1.4 About GWAS **22**

1.4.1 GWAS study design 23

1.5 Summary **24**

1.6 Aims and structure of this thesis **24**

Chapter 2 : Creating and utilising the Seizure Associated Genes Across Species (SAGAS) database to predict drugs for epilepsy ***26***

2.1 Introduction 26

2.2 The SAGAS database 27

2.2.1 Methods: Electronic searches 27

2.2.2 Incorporating data from existing databases 28

2.2.3 Results: The SAGAS Database 29

2.3 Utilising SAGAS for drug repurposing 33

2.3.1 Methods: Utilising SAGAS for Drug repurposing 33

2.3.2 Results: Drug repurposing 34

2.4 Discussion 37

Chapter 3 : Genomics-based drug repurposing for Dravet Syndrome 42

3.1 Introduction 42

3.2 Methods 43

3.2.1 Data acquisition 43

3.2.2 Cosine distance 44

3.2.3 In Silico Validation strategy 46

3.2.4 Further validation 46

3.2.5 Improving drug predictions by creating and integrating the 'protein targets score' 48

3.2.6 Drug selection strategy 50

3.3 Results 52

3.3.1 Drug predictions for Dravet syndrome are concordant with the clinical effects of AEDs in Dravet syndrome 52

3.3.3 Drug predictions are improved by incorporating the protein target scores 56

3.3.4 Final drug selection 58

3.4 Discussion 58

Chapter 4 : Outlook, and future work 62

4.1 Future research 62

4.2 Conclusion 63

References 64

Appendices 74

Appendix 1: CACHEP search terms..... 74

A 1.1: Search for genes/proteins that cause seizures/epilepsy in animal models..... 74

A 1.2: Search for human epilepsy genes..... 75

Appendix 2: Entry proforma for part 1 of the CACHEP initiative..... 76

Appendix 3: GUILDIFY analysis selected genes..... 77

A 3.1: Genes used for Generalised epilepsy predictions..... 77

A 3.2: Genes used for Focal epilepsy predictions 78

Acknowledgments

My sincere thanks go out to my supervisor, Doctor Nasir Mirza, for his unwavering support and kindness throughout this degree, for the patience to help me over every hurdle, and for every opportunity that you have entrusted me with ever since the work began. I cannot thank you enough. My thanks also go out to Professor Tony Marson, who had the faith to set me on this path from the very beginning.

I would also like to thank the Wolfson Foundation and the Royal College of Physicians for supporting this work.

My warmest salutations go out to my friend, mentor and brother, Abdurrahman Islim, for continuing to be a source of inspiration, support, and wisdom. Thank you for guiding me into the world of academia and neurosurgery. Yours is a debt I cannot ever repay, and I can only wish you the very best for the days ahead.

My thanks go out to Conor Gillespie, George Richardson, Arish Mustafa, Roshan Babar, and Sumirat Keshwara, for the day to day help and support, for every smile you've put on my face, for every favour both big and small and most of all, for tolerating me throughout this year. I will hold dear the memories of our days together in the office, and I look forward to continuing to work with you in the coming years.

Thank you to Lara Gracie, and to Danial Rostami-Hochaghan, for their tireless efforts with the SAGAS database, and to the 60+ students for their crucial and diligent contributions to the CACHEP initiative. Without you, this could have never been possible.

My thanks go out to my mentor Mr. Emmanuel Chavredakis, for introducing me to the Walton centre, and having the faith to entrust me with my first ever research project, and to Professor Michael Jenkinson, Cathal Hanan, Ali Baksh, Chris Millward and Nisaharan Srikandarajah, for being friendly faces at the Walton centre, and for being shining sources of inspiration.

My thanks to Dia. For your contributions to my sanity and for your endless wisdom.
Finally, my dearest thanks go out to my family for their continual love, kindness, help and support. I cannot imagine where I would be without you. Your support and belief in me are what grant me the motivation to keep going, and I dedicate this work to you.

Truly, thank you all.

Abstract

65 million people have epilepsy. Current antiepileptic drugs produce adverse effects in 88% of users and fail to prevent seizures in 30% of people with epilepsy. New drugs for epilepsy are therefore required.

Traditional drug development methods are arduous and expensive, taking on average 10-15 years and $2.6 billion per drug. It is estimated that over 90% of drugs have a viable second indication and thus may be used for other purposes, making drug repurposing an attractive alternative.

This thesis aims to create drug repurposing resources for epilepsy and generate drug predictions for both monogenic and polygenic epilepsies.
We create and present the Seizure Associated Genes Across Species (SAGAS) database, the largest and most comprehensive existing database of epilepsy genes, containing over 9700 pieces of published evidence for the involvement of 3879 genes in the generation and potentiation of seizures across 6 species. We use genetic data from the SAGAS, alongside a publicly available network-based method of drug prediction, to generate drug prediction lists for polygenic focal and generalised epilepsies.

A monogenic epileptic syndrome is caused by a single mutant gene. However, knowing the identity of the mutant gene underlying a monogenic epileptic syndrome is not sufficient for predicting the effect of antiseizure medications on the syndrome. Dravet syndrome (DS), the archetypal monogenic epileptic encephalopathy, is typically caused by mutations in *SCN1A*. Some antiseizure medications that alleviate seizures in Dravet syndrome do not affect *SCN1A*, whilst some antiseizure medications that affect *SCN1A* aggravate seizures in Dravet syndrome. We are not aware of any genomics-based methods that can correctly predict the varying effects of different antiseizure medications on Dravet syndrome (or any other monogenic epileptic syndrome). We create a novel method to predict drugs for Dravet syndrome that takes into account not only the gene that causes Dravet syndrome but also other genes that can influence the expression of its phenotype and show that our predictions correctly identify the antiseizure drugs that are effective, aggravating and equivocal for Dravet syndrome.

List of Tables

Table 1.1: Properties of a drug that are considered prior to progression to the optimization phase..... 11

Table 2.1: Drug prediction table for Focal epilepsy 36

**: These drugs are tied in ranking as per their GUILDify score*..... 36

Table 2.2: Drug prediction table for generalised epilepsy 37

Table 3.1 AEDs classified as effective or aggravating..... 45

Table 3.2 List of all experimental drugs used in model validation 47

Table 3.3: The scoring system used in the enhanced model. Also shown are the sources from which this gene data was extracted..... 51

Table 3.4: Table of genes included in category II, which includes gene mutations that cause phenotypes similar to those observed in Dravet syndrome 52

Table 3.5: Average cosine distance values for different phenotypes. Please note that cosine distance values are not comparable across different phenotypes. 55

Table 3.6 Median percentile ranks of AED sets in drug predictions based on different scores 57

Table 3.7: Final drug prediction list after manual curation based on drug safety and appropriateness for epilepsy. 58

List of Figures

Figure 1.1: Simplified diagram outlining seizure classification, redrawn from the ILAE Classification of Epilepsies (9) 2

Figure 1.2: Traditional drug development pathway. Original figure..... 7

Figure 1.3: Traditional pre-clinical drug development. (ADMET = Absorption, Distribution, Metabolism, Elimination, Toxicity) Original figure. 7

Figure 1.4: Bar chart of success rate of compounds for each stage of clinical testing to approval. Reproduced from (89). Numbers are for compounds between the years 1991 and 2010. LOA = Likelihood of Approvals NDA/BLA = New drug application, Biologic license application 13

Figure 1.5: Graph showing cost of drug development vs. number of approved drugs. Reproduced from Xue H, Li J, Xie H, Wang Y. Review of Drug Repositioning Approaches and Resources..PhMRA = Pharmaceutical Research and Manufacturers of America, FDA = Food and Drug administration. 14

Figure 1.6: Pathways of drug development/approval. Reproduced, with permission, from Parvathaneni V, Kulkarni NS, Muth A, Gupta V. “Drug repurposing: a promising tool to accelerate the drug discovery process” 15

Figure 1.7: Illustrating the different magnitudes of effect genetic variation can phenotypically exert. So-called "Mendelian disorders” occupy the top left, whereas lower effect size variations typically identified by GWAs studies occupy the bottom right. Most disease-related genes lie on or around either of the two diagonals of the diagram. Reproduced from (127). 23

Figure 2.1: Screenshot of SAGAS website homepage 30

Figure 2.2: Screenshot of online SAGAS database at Sagas.ac . The “Citations” tab exhibits an outline of the genes included within the database, including the quantity and type of evidence for each species. The summary tab has one row per included gene. 31

Figure 2.3: Screenshot of the Citations tab, with one row per article of evidence. 32

Figure 2.4: Visualisation of the subnetworks produced. The subnetwork generated for focal epilepsy is generated on the left, whereas the subnetwork for generalised epilepsy is shown on the right. Green node = Seed, yellow node = top 1% non-seed node. Note that drugs are not visualised in the above figure.. 35

Figure 3.1: Heatmap illustrating average cosine distance values for effective, equivocal and aggravating AEDs, within the cosine distance values for all analysed drugs. a. Cosine distance ranges between +1 and −1, where +1 means that the disease-causing and drug-induced changes in the transcriptome are exactly the same, and −1 means that they are diametrically opposite. 53

Figure 3.4: Figure 1.8: Illustrating the prediction model after application of expression based-scoring system. Effective drugs rank highly, whereas aggravating drugs rank much lower. Note that these values utilise medians, hence the position of the aggravating average. 57

Abbreviations

ILAE	International League against Epilepsy
IGE	Idiopathic generalised epilepsy
JME	Juvenile Myoclonic epilepsy
CAE	Childhood Absence epilepsy
HS	Hippocampal sclerosis
AED	Anti-epileptic drugs
VNS	Vagal nerve stimulation
DBS	Deep brain stimulation
FDA	Food and Drug Administration
WHO	World Health Organisation
MHRA	Medicines and Health products Regulatory Agency
NIHR	National institute for Health Research
TANGO	Targeted Augmentation of Nuclear Gene Output


بِسْمِ ٱللَّهِ ٱلرَّحْمَٰنِ ٱلرَّحِيمِ

Chapter 1 Introduction to Epilepsy, Genomics and Drug repurposing

1.1 Introduction to Epilepsy

Epilepsy, or clinical features depicting epilepsy, have been described as far back as 2000 B.C. This disease has had, and continues to have, great social and cultural significance, famously being named the "*Sacred disease*" by Hippocrates.(1)

The term "Epilepsy" encompasses a large group of syndromes, the primary unifying feature of which is the predisposition to the occurrence seizures, defined as: “a paroxysmal alteration of neurologic function caused by the excessive, hypersynchronous discharge of neurons in the brain.”(2)

1.1.1 Epidemiology, prevalence and distribution

Epilepsy is one of the most common neurological disorders worldwide with an average worldwide point (or active) prevalence of 6.38-6.68 per 1000 of the population, and a lifetime prevalence of 7.6 per 1000.(3, 4)

The incidence of epilepsy follows a bimodal distribution in which the most affected are the young and the elderly.(4) The incidence of epilepsy in the young is highest in the first year of life and declines to adult rates by 10 years of life.(3) This is likely due to the high incidence of childhood epileptic disorders with genetic and developmental aetiologies.(5) Similarly, the elderly are affected more due to the higher incidence of other disorders that increase the risk of epilepsy, such as, strokes and brain tumours.(6)

1.1.2 Seizures and seizure types

Seizures can be broadly divided into focal onset and generalised onset. In a generalised onset seizure, electrical activity originates in and affects both hemispheres of the brain simultaneously, whereas in a focal onset seizure, electrical activity originates in a specific part of the brain, and then may either remain localised, or may progress to affect other parts of the brain.

Generalised seizures typically (but not always) cause impairment of consciousness, whereas focal seizures may or may not, depending on the subtype of seizure.(7)

1.1.2.2 Subtypes and subdivisions of seizures

The 2017 International League against Epilepsy (ILAE) classification of seizure types (8) subdivides seizures further based on clinical features/ semiology. This classification distinguishes between the seizure semiology at onset, and the semiology observed throughout the course of the seizure, as shown in Figure 1.1

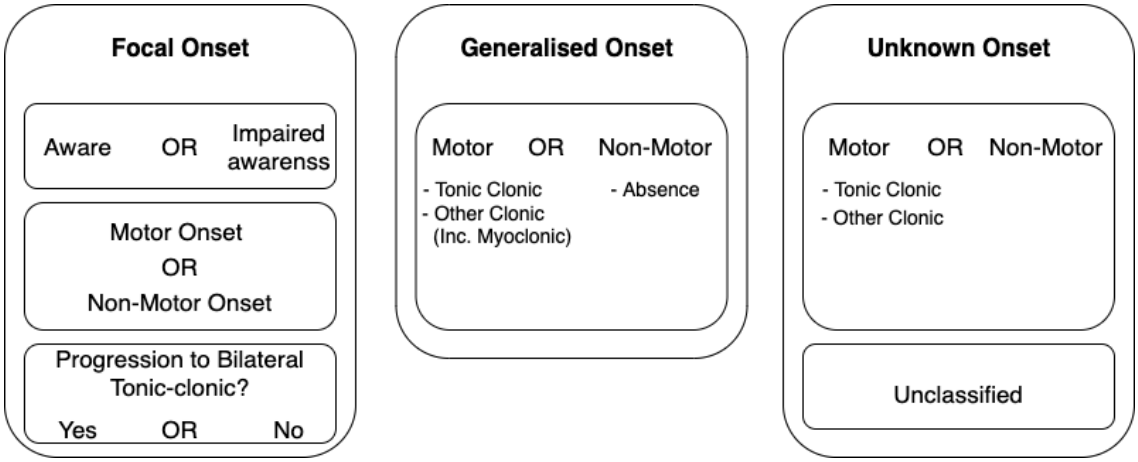


Figure 1.1: Simplified diagram outlining seizure classification, redrawn from the ILAE Classification of Epilepsies (9)

1.1.3 Common and rarer epilepsies

Common epilepsies have complex inheritance and genetic architecture. The two most common subtypes of Idiopathic epilepsies, Idiopathic generalised epilepsy (IGE) and Self-limited partial epilepsy, account for 15-20% and 15%-25% epilepsies respectively.(9-12) Other examples of common epilepsies include juvenile myoclonic epilepsy (JME), childhood absence epilepsy (CAE) and hippocampal sclerosis (HS).

Monogenetic epilepsies are much rarer, however. For comparison, West syndrome, tuberous sclerosis, and Angelmann syndrome, some of the most well-recognised single-gene disorders with seizures as a feature, occur in approximately 1 per 2500 births, 1 per 6000 births and 1 per 15,000 births, respectively.(13-15)

1.1.4 Current Management modalities of Epilepsy

The mainstay of epilepsy management is pharmacological, relying on the utilisation of anti-epileptic drugs (AEDs).(16) The aim of AED is to reduce the frequency and severity of seizures and, ideally, achieve seizure-freedom. However, 30% of patients will continue to suffer from seizures despite AED therapy.(17)

AEDs are typically used as long-term prophylaxis against seizures, but also may be taken acutely to abort an ongoing seizure (as is seen in status epilepticus).(16) AEDs may also be taken acutely by patients who are experiencing a known prodrome to a seizure, with the aim of preventing the seizure from occurring.(18)

AEDs vary in mechanism of action and in indication, and the minutiae of AED indications are beyond the scope of this thesis. AED selection for an individual, however, is guided by epilepsy syndrome and seizure types, as certain epilepsy syndromes are more responsive to certain AEDs, and certain seizure types are exacerbated by certain AEDs. The clinical issues surrounding AED selection in epilepsy are further discussed below (Section 1.1.4.1).

Following the first AED prescription approximately (53%) of patients will continue to have seizures despite taking the prescribed AED.(19) For these patients, a second AED will usually be trialled, and should that fail to control seizures, then other AEDs may continue to be trialled. For each AED that is trialled, after the first, the patient is 1.7 times less likely to respond to the subsequent drug.(20) Once a patient has failed to achieve seizure control following two appropriate AEDs at appropriate doses, the patient is considered to have drug-resistant epilepsy (DRE). Roughly 30-40% (20) of epilepsy patients will be drug resistant.

The surgical management options available are largely dependent on the anatomical location of the epileptogenic focus. If the seizures arise from a single non-eloquent locus, surgery aiming to resect the epileptogenic region may be an option. Epilepsy may also be surgically managed via neurostimulatory methods, namely vagal nerve stimulation (VNS), and deep brain stimulation (DBS).(21, 22)

For patients in whom AEDs are ineffective, diet-based treatment as well as surgical options might be available.(23) The ketogenic diet for example, leads to a significant improvement in seizure control for some patients.(24, 25)

Patients with epilepsy, in particular those with generalised epilepsy, are also advised to avoid seizure triggers, such as lack of sleep,(26, 27) excess alcohol consumption (28) and flashing lights.(29, 30)

1.1.4.1 Current Issues in the Medical Management of Epilepsy

Current AED selection tends to be centred around finding a regimen that reduces the occurrence rate of seizures, minimises side effect profiles and is compatible with the patient's past medical and drug history. Despite this, AED prescribing in epilepsy is far from simple.

Certain epileptic drugs, such as sodium valproate are highly teratogenic and are either relatively or completely contraindicated in pregnancy.(31, 32) Others tend to have cognitive and behavioural side effects and tend to be avoided in children.(33-35) Different AEDs also have varied side effect profiles, and the tolerance of the patient in clinic will change depending on individual baseline patient characteristics such as age and social history.(36) AEDs such as Valproate also affect liver enzyme metabolism of other drugs (37, 38) and Valproate is also strongly protein binding,(38, 39) which may displace other protein binding drugs causing acute toxicity.

The efficacy of an individual AED will vary from individual to individual,(40, 41) making it common for patients to trial more than one drug before finally settling on the most favourable drug. In addition, treatment regimens may feature an adjuvant drug,(42) with its own pharmacodynamic and pharmacokinetic issues, as well as its own side effect profile and efficacy variability. Long term AED therapy may also precipitate osteoporosis,(43-45) leading to increased risk of fractures in a population that is already at high risk due to seizures. These prescribing difficulties are compounded by patient co-morbidities, especially in the elderly population, such as kidney or liver disease, that may further influence drug metabolism and clearance.(46, 47)

Finally, there remains the population of patients with intractable epilepsy, who typically will trial a number of AEDs, but experience no significant improvement in seizure occurrence or severity.(48, 49) Other drug-specific idiosyncratic issues also exist that further increase the clinical difficulty in treating patients with epilepsy.(50)

These complex pharmacological characteristics mean that AED selection in epilepsy can prove to be clinically challenging. Better epilepsy drugs are needed to tackle these issues and provide medical treatment options that are less cumbersome. Drugs with more tolerable side effect profiles are needed, as well drugs with easier to manage drug-drug interactions and metabolism profiles. A marked need is present for high-efficacy drugs that are safe in pregnancy, and more therapy options are needed that reduce the need for adjuvant therapy.

1.2 Introduction to novel drug development and drug repurposing

Developing new drugs is a lengthy and costly process: it takes approximately £2.4 billion and 15 years to develop a new drug.(39, 51) The processes utilised in traditional novel drug development have varied with time, as improvements in available technology continue to provide more options to drug developers. It should be noted however that many drugs that are currently in use did not undergo this process, as many drugs were as a result of a false hypothesis or by chance observation. These observations are often made in empirical screens, where vast compound libraries are tested against predictive animal models. For example, target identification (discussed further in section 1.2.1) is often the first step taken towards novel drug discovery. However, for certain drugs, such as levetiracetam, the molecular targets of the drug were discovered years after the drug had reached the market.(52, 53)

Thus, the following section only aims to describe the common systematic processes often pursued for modern drug development, rather than give a comprehensive review of every possible avenue of drug discovery. Modern drug discovery processes typically follow the path illustrated in Figures 1.2 and 1.3.

1.2.1 Target identification

The initial step in drug design is the identification of a biological target.(54, 55) Targets that are considered are often ones that have been identified by the scientific literature as having a role in the pathophysiology of the disease.(56, 57) Pharmacological interactions with the target must be deemed theoretically safe and efficacious before development aiming to exploit this target progresses.(58)

This target's function must be potentially modifiable by pharmacological agents, i.e., the target must be "druggable".(59, 60) A target's druggability is typically estimated by grouping the target with known gene families that have been previously successfully targeted.(59) Other methods exist to determine druggability which may involve screening for an identifiable biological response, either *in vivo* or *in vitro*.(61) Targets may be identified through the exploration of published or publicly-available scientific literature.(56) The target identification process may also include novel research, whereby the roles of proteins and transcripts that may be culpable in the generation of the disease process are studied.

mRNA expression profiles can be studied in order to determine how the disease reacts to alterations within the transcriptome.(62) Target identification studies may instead focus on studying genetic polymorphisms theorised to be of importance in the disease process, and .(63) Genetic targets may also be explored via knockout animal models, to further study how the genome affects the disease and disease progression.(63)

Targets may also be chosen based on a process called "target deconvolution", whereby an effective drug that is currently in use is studied in order to identify its target.(64) The target of the existing efficacious drug is then considered for use as the target for new drugs to be developed.

1.2.2 Target validation

Once targets have been identified, the target must be validated for its ability to modulate disease pathophysiology.(55) Multiple methods of target validation exist, and it is preferable that multiple methods of target validation are used, to increase confidence in the validation process. Processes involved in target validation are not dissimilar to those involved in target identification.

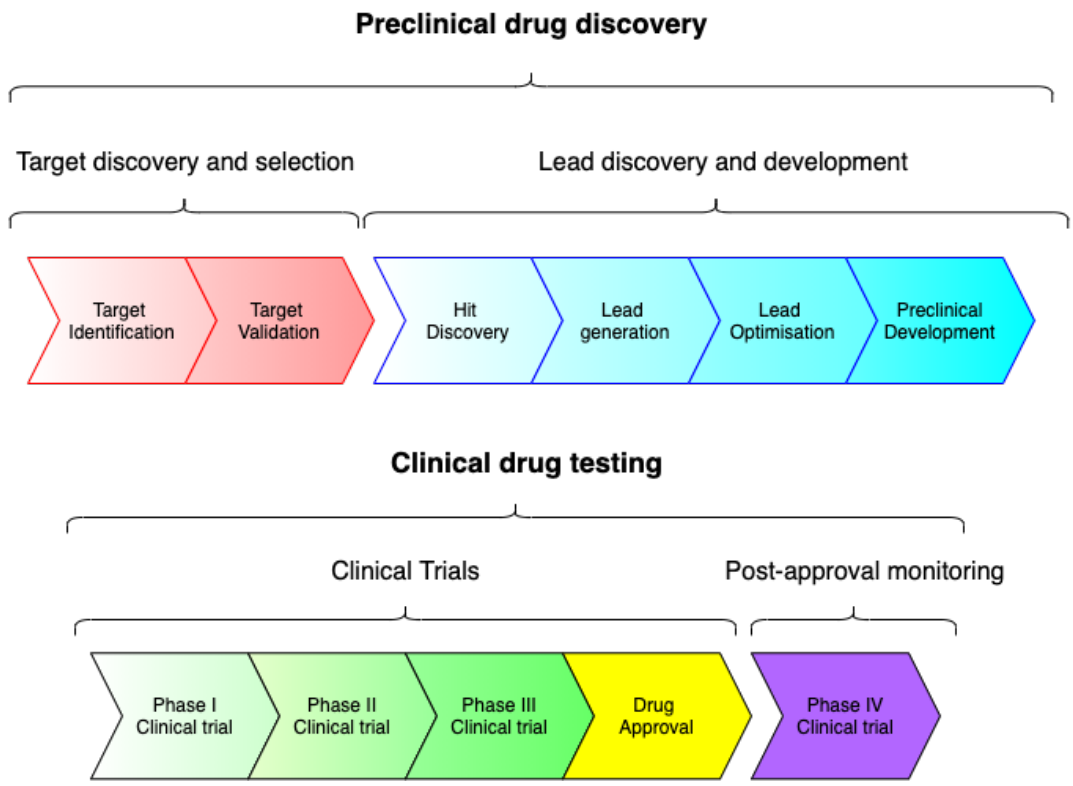


Figure 1.2: Traditional drug development pathway. Original figure.

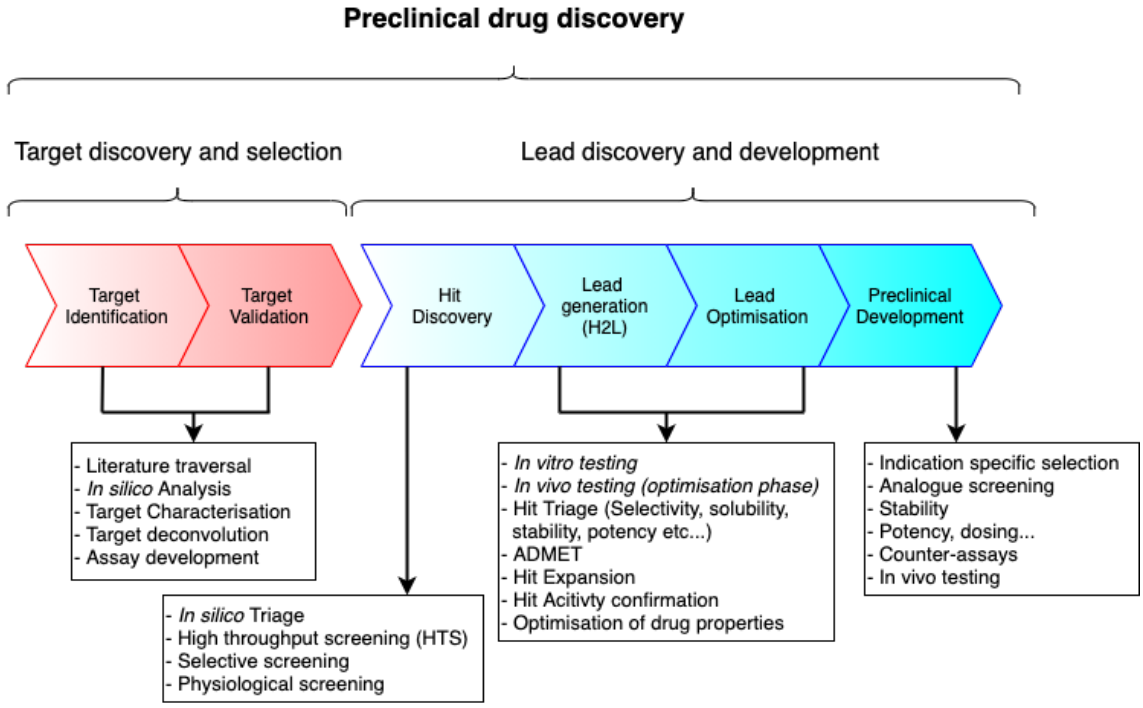


Figure 1.3: Traditional pre-clinical drug development. (ADMET = Absorption, Distribution, Metabolism, Elimination, Toxicity) Original figure.

1.2.3 Hit discovery

Once the target has been identified and validated, the goal becomes to identify a compound that has the desired activity on the target molecule, for example, activation or inhibition .(65) The compound found to fulfil this criterion is known at this stage of drug development as a "hit".

One common method of hit discovery is high throughput screening (HTS) of compounds, with compound libraries being screened against the target.(65, 66) HTS typically utilises laboratory automation in order to process the compounds at a high rate, and assumes no prior chemical knowledge regarding any compound's individual likelihood of being effective. An alternative is to use selective screening in which compounds that are thought to have a higher chance of being a hit are utilised,(67) thereby reducing the processing burden of screening. Knowledge of specific compound activity in this case is derived from the literature, and compound classes that are known to be effective may lead to the inclusion of similar compounds into the selective screen.

Increasingly, *in silico* methods are being utilised in order to identify the most promising hit compounds. Structure-aided drug design is one such method.(68) Structure-aided drug design aims to use knowledge of molecular structures to help design optimised molecules. This works often as an adjunct to HTS, and other screening methods, to provide information on how the hit compound could be modified chemically in order to optimise drug potency and target selectivity.

Physiological screening is a technique that tests the compound on a cell line or tissue of interest to investigate the effects of candidate compounds,(69) unlike HTS which tests compound affinity for the target molecule. Being reliant on tissue, however, this is lower throughout model, and may be used in retesting of hit compounds or as part of a focussed screen.

1.2.4 Lead generation and Lead optimisation

Lead generation, or the Hit-to-Lead process (H2L) is a process to progress the development of hits to make them more suitable for clinical use.(65) H2L aims to consider not only the efficacy of the compound, but to consider certain pharmacokinetic properties, optimising the usability of the drug, and capitalising on the most desirable properties of the hit

compound whilst reducing the less desirable properties.(65) This involves optimising hit compounds for potency, selectivity, solubility, permeability, metabolic stability, low cytochrome P450 activity and favourable absorption, distribution, metabolism, excretion and toxicity (ADMET) properties.(70) This also allows the quick exclusion of "dead-end" hits, which are compounds with unacceptable pharmacokinetic properties or potentially dangerous toxicity profiles.(71, 72)

The first step of the H2L process is called hit triage. In hit triage, the resulting hits from a HTS are assessed by medicinal chemists and computer aided drug design (CADD) scientists in order to identify the basic chemical scaffolding underlying each of the hit compounds.(72-74) The Hits are then grouped, with each group containing compounds that share a common chemical scaffold.

After grouping, the hits are then ranked based on desirability. The metrics by which the desirability is measured will vary from drug to drug, depending on the intended goals of the drug, indeed, it is common to see parameters added throughout the H2L process, as the need to optimise compounds for that parameter arises. Despite the variability in parameters across drug development processes, common scoring and ranking methodologies can be used once the parameters of a drug have been identified. One of the more prominent of these is the traffic light scoring system described by Lobell et al.,(75) whereby for each parameter, each drug is given one of three scores, good (0), warning (+1) and bad (+2), resulting in less desirable compounds having higher scores. The final scores can then be used alongside the grouping performed in the Hit triage stage in order to calculate aggregate scores such as means, medians and standard deviations for each scaffold group, thereby identifying the most desirable chemical scaffold for a given target. Once scaffold groups have been ranked, the more desirable scaffold groups are then selected for the purposes of resource allocation, with more resources typically put towards the development of more desirable hits and scaffolds.

At this stage several steps are taken towards optimising the lead simultaneously. One of these steps, the process of hit expansion,(65, 76) aims to identify more compounds that share a desirable scaffold but were not identified in the HTS. This may be done via purchase of select compound libraries for exploration of the desirability of individual compounds that share this scaffold.

This may utilise Quantitative Structural Analysis Relationship (QSAR) technology.(77) Previously, the only way to understand the properties of a compound was to synthesise every possible compound that could be formed from the parent scaffold. QSAR allows the prediction of compound properties by synthesising a small number of compounds to be tested within a lab, and then utilising the data from those tests to predict the chemical properties of every other possible compound.

Another step involved is confirmation of hit activity. This features rigorous study and testing of hit compounds for desirable pharmacokinetic properties including ADMET properties, compound potency and selectivity, and other pharmacokinetic properties such as lipophilicity. Earlier stages of these studies will resemble screening processes, however as the number of potential leads is narrowed, this process becomes more elucidative, with further and further properties of individual compounds being studied in-depth.

Hits and scaffold groups are then ranked based on these studies, with higher ranked compound and series more likely to progress to the lead optimisation phase. Prior to proceeding into the lead optimisation phase, a hit series must demonstrate the properties listed in table 1.1. It is very rare to find a compound that perfectly satisfies every property here, but a hit series is typically chosen for lead optimisation if it can be demonstrated that a number of different hits within the series can exhibit a substantial number of these desirable characteristics.

Lead optimisation is a much more costly process than lead generation (78) and is therefore conducted very selectively on very promising hit series. The process of lead optimisation itself utilises many of the methods used in lead generation, and primarily features optimisation of promising hit series via medicinal chemistry analyses,(79) wherein the properties of hits are further modified towards optimising favourable pharmacological properties.

Table 1.1: Properties of a drug that are considered prior to progression to the optimization phase.

Target Engagement	ADME	Safety/Toxicity	Chemical optimisation potential
Potency	Cell permeability	Low CYP* inhibition	Synthetic accessibility
Cellular target activity	Metabolic stability	No/low QT prolongation (tested via hERG gene patch-clamp essay)	Favourable QSAR predictions, indicating strong affinity for target and potency
In vivo proof of concept	Oral bioavailability (not all drugs, necessary in epilepsy)	Selectivity against related targets	Multiple modifiable sites available on scaffold
Biomarker evidence of target engagement		Selectivity against broad selectivity panel	Clear patent strategy

**CYP: Cytochrome P450*

1.2.5 Candidate identification and preclinical development

Once pharmacokinetic properties have been optimised, notable leads are then assessed for their usefulness as candidate drugs. Potential candidates must fulfil the basic needs of activity, potency and safety, but must also fulfil more practical needs.

The aims of candidate identification, much like lead generation, is highly variable, and depends on the characteristics of the drug itself, as well as the drug's intended target. Blood-brain permeability may be desirable in an Epilepsy drug, for example, but likely is not a desirable trait if the drug was intended to act primarily on hepatocytes. Compound analogues will be synthesised and tested against models to select an analogue of the lead that targets the least undesirable off target effects, while maintaining adequate target activity. Leads must be molecules that are stable, allowing for practical storage. The molecules must also be easy to be manufacture, for mass production.(80)

Clinical patient focussed questions are also addressed at this stage. For example, frequency and quantity of dosing as shown by animal models must be commensurate with a practical regimen for humans,(81) and the safety of the optimised lead will be reassessed to ensure that the molecule is not overtly toxic.(82) Counter assays may be used to test binding to

select receptors that are particularly undesirable for binding, such as those that are known to induce toxicity or those that may produce unpleasant adverse effects. If the compound is found to bind to undesirable receptors, further chemical medicinal modification of the compounds must take place to work these binding properties out of the final compound. Inevitably, every medicinal chemical modification will lead to sacrifices in other areas, such as absorption or target receptor selectivity. Nevertheless, in compound modification programs, lead selectivity is typically prioritised and optimised first.

Certain receptors are notorious for impeding drug progression at this stage and are almost certainly included in counter-assays. These are typically cardiac receptors and namely the I_{kr} receptor that is responsible for long-QT syndrome,(83, 84) Prolonged PQT will increase the risk of cardiac arrest, and therefore will work against the drug when seeking clinical approval. Cytochrome P450 (CYP450) activity (inhibition/stimulation) is another feature that may be problematic for the lead, and early assays against CYP450 are typically performed at this stage.(85)

Once medicinal chemical compound modification has concluded compound optimisation, the compound must then be tested both *in vitro* and *in vivo* to verify that the compound produces the desired response at both the cellular, tissue and organism level. *In vivo* assays aim to ascertain and verify the binding potential of leads to the target receptor. Additionally, assays will aim to establish the effects of the compound on the receptor, whether they are agonistic or antagonistic.

Animal models are utilised at this stage to verify the adequacy of compounds.(86) Different models can be used to assess different compound parameters, such as target bioavailability, drug response, and adequacy of the intended route of administration and other parameters. In addition, the testing period will vary based on the disease. Some drugs will require months of administration before efficacy can be measured, whereas other compounds will be expected to demonstrated efficacy within minutes.

Researchers will aim to identify a favoured compound and begin collecting data on the compound based on the *in vivo* and *in vitro* work. This is necessary in order to present these findings to national organisations (such as the United States of America's food and Drug administration (FDA) or the United Kingdom's Medicines and Health products Regulatory

Agency (MHRA) for the purposes of trial approval. Drug developers will need to show adequate safety and efficacy profiles, as well as a suitable ADMET profile within animal models, and suitable receptor selectivity in *in vitro* models before clinical testing on humans can be considered.

1.2.6 Clinical testing

Clinical trials are performed in three phases, with safety being the primary point of assessment in Phase I trials, efficacy in Phase II trials, and comparison vs. current standard of care in Phase III trials.(87) For epilepsy specifically, the epilepsy drug is sometimes trialled against a placebo, rather than the standard of care.(88) Many drugs will fail at different stages of the process with 68.2% of drugs failing at the preclinical stage, 24.9% at phase I trials, 50.0% at phase II trials, and 41.4% of drugs failing at phase III trials. Finally, a further 12.5% of drugs will fail at acquiring a license for the drug, despite successful Phase III testing (89)(Figure1.4).

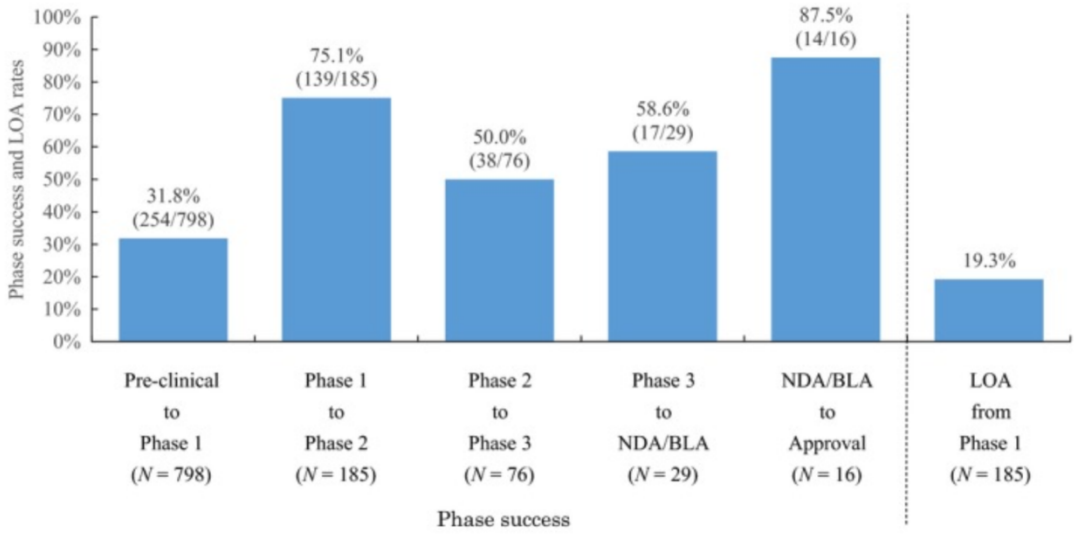


Figure 1.4: Bar chart of success rate of compounds for each stage of clinical testing to approval. Reproduced from (89). Numbers are for compounds between the years 1991 and 2010. LOA = Likelihood of Approvals NDA/BLA = New drug application, Biologic license application

The long and complex process of novel drug discovery will therefore inevitably leave many compounds by the wayside that will have failed at different stages of the process. These compounds represent an important source of potential therapeutic agents for future drug repurposing/repositioning initiatives.

1.3 About drug repurposing

Drug repurposing (or drug repositioning) is the use of compounds for indications other than the indications the compounds were originally designed for. The compounds used in drug repurposing may be drugs that are currently licensed for diseases, old drugs that are no longer used in clinical practice, or candidate compounds that were discarded during the traditional novel drug discovery process.

Modern traditional novel drug discovery approaches demand high time and resource costs. These burdens are reflected in current trends in drug development, with the costs of pharmaceutical development steadily increasing since 1995,(90) with no commensurate increase in number of newly licensed FDA drugs.

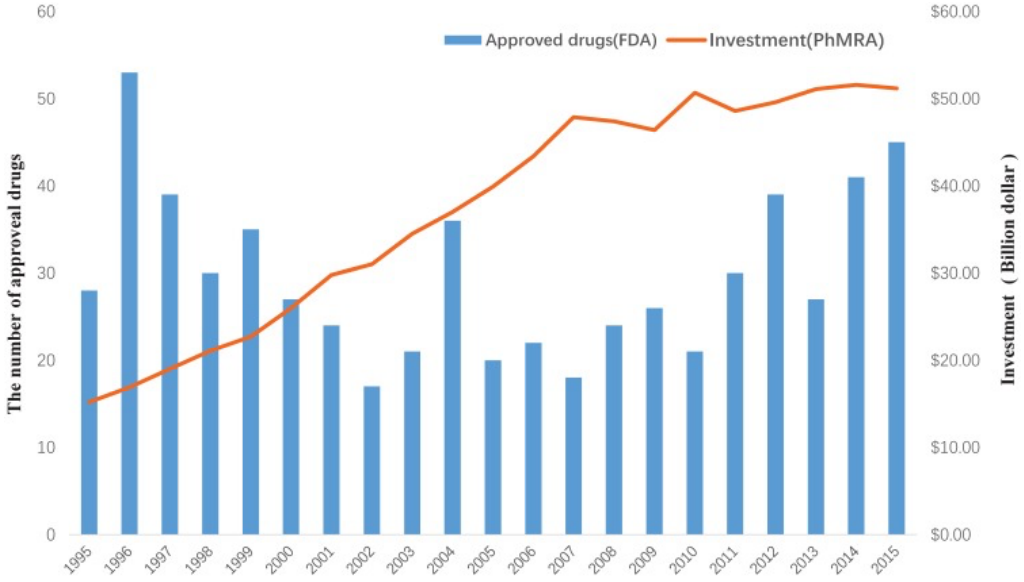


Figure 1.5: Graph showing cost of drug development vs. number of approved drugs. Reproduced from Xue H, Li J, Xie H, Wang Y. Review of Drug Repositioning Approaches and Resources..PhMRA = Pharmaceutical Research and Manufacturers of America, FDA = Food and Drug administration.

Drug repurposing therefore is an attractive prospect for several reasons, chief of which is the reduction in monetary and time investment required. A repurposed drug can bypass Phase I clinical testing, as drugs that have already been utilised in the treatment of other diseases will have a well-documented safety profile. Drug repurposing also has the potential of providing a new indication to already utilised low-cost generics. This allows the therapeutic utilisation of generics in place of much higher cost patented drugs, thereby reducing the cost of treatment to public health services. This is especially true in the case of orphan diseases, where the untreated disease can lead to costs associated with long-term supportive therapy for patients which may not be required, or may be required less, in the instance of

therapeutic availability. Orphan diseases are less likely to attract research investment from pharmaceutical companies, as there is a lower potential for financial return on the investment. Hence, drug repurposing, because of its lower costs, has a particular role in addressing the unmet need of finding treatments for orphan diseases.

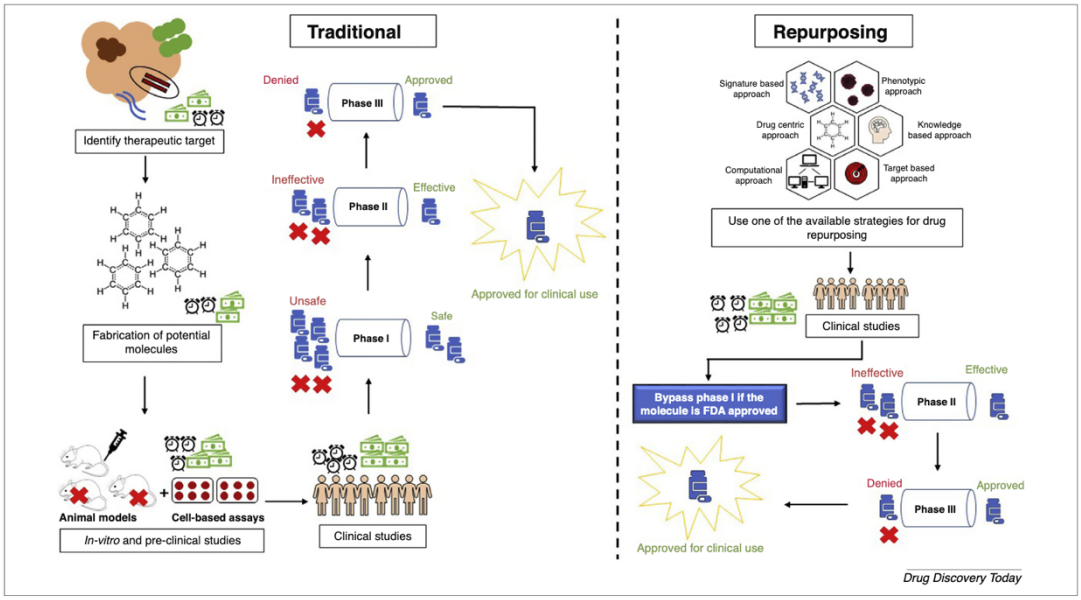


Figure 1.6: Pathways of drug development/approval. Reproduced, with permission, from Parvathaneni V, Kulkarni NS, Muth A, Gupta V. "Drug repurposing: a promising tool to accelerate the drug discovery process"

1.3.1. Challenges of drug repurposing

While pharmaceutical companies will show great interest in methods that may increase efficiency, and decrease costs, successfully repurposed drugs may not necessarily be a source of substantial profit, as profits from drug discovery largely come as a direct result of exclusive patents. Patenting in this way will typically not be available as an option for pharmaceutical companies that engage in drug repurposing as many of the candidate compounds will be available as generics. This is compounded by the fact that clinicians could prescribe off-label a generic cheaper version of an expensive patented drug.

An example of this is demonstrated by Avastin (bevicizumab), and its role in treating macular degeneration.(91-93) Bevicizumab is a biologic drug used in the therapy of colorectal and renal cancers, however, it was found to be effective in the treatment of wet age-related macular degeneration (AMD). With the emergence of this evidence, in 2017, twelve clinical commissioning groups in the north of England attempted to use bevacizumab in the treatment of AMD, however, it was challenged by pharmaceutical companies Novartis and Bayer(94), who are responsible for the marketing of ranibizumab and aflibercept.

Ranibizumab and aflibercept are two other biologic drugs used in AMD, however, regimens using these two drugs are estimated to be 30 times more expensive than a regimen utilising bevacizumab, saving the NHS an estimated £100 million per year(95). The High court ruled in favour of the clinical commissioning groups in question, and against Novartis and Bayer in 2018.(96) The companies took the case to the court of appeal, which then ruled unanimously again in the favour of the NHS.(97) Avastin however, which is prescribable for AMD, remains unlicensed for this indication. This highlights yet another challenge of drug repurposing, as clinicians prescribing Avastin in the UK for AMD will be required to prescribe it as an off-label drug, leaving the prescribing doctor in a much more exposed and vulnerable legal position. This leads to reluctance in off-label prescribing, which may impede the usage of the drug, despite its advantages.

Currently, drug repurposing initiatives are largely pioneered by smaller institutions and academics. This leads to practical challenges, such as securing funding for large scale clinical trials. Traditionally, large-scale clinical trials are funded by pharmaceutical companies. For drug repurposing clinical trials, the funding must be obtained, generally speaking, from public/government bodies, for example, the National Institute for Health Research (NIHR). In most cases, the funding available from public bodies is much more limited than that available from the commercial sector. Finally, under current legislation, for a drug to be licensed, it must be sponsored by a pharmaceutical company. This is also a significant hindrance to drug repurposing(98).

1.3.2 Drug repurposing approaches

The approach utilised in drug repurposing will vary across diseases, as the approach is typically selected in accordance with the biology of the disease.(99) One of the most common causes of drug repurposing in clinical medicine has been serendipity,(100) with chance observation leading to hypothesis, which finally leads to structured research.(101) The time between point of initial chance observation and clinical suspicion and the point of initiation of clinical trials can take many decades, due largely to the lack of a systematic repurposing approach/process.

One example is the drug fenfluramine. Fenfluramine, a modulator of the sodium dependent serotonin transporter, was launched in the early 1970s for the treatment of obesity and was withdrawn in 1997 due to concerns regarding structural cardiac adverse-effects, such as,

valve abnormalities.(102) Interestingly, clinical case reports and case series published between 1980 and 1997 found a demonstrable therapeutic effect for fenfluramine in the treatment of seizures.(103-105) Fenfluramine, was then investigated further in 33 patients by Gastaut et al, one of the clinicians who originally published one of the first case series demonstrating fenfluramine's anti-seizure effects.(106) Belgian clinicians investigating fenfluramine for its potential new indication new indication sought approval from Belgian regulatory authorities to continue investigation of the drug for seizures, and permission was indeed granted.(107) Therefore, in the years following these investigative studies, as the use of fenfluramine diminished as an obesity drug, it's use as an AED become more common. Further case series, studies, and later on clinical trials,(108, 109) finally saw fenfluramine licensed for Dravet syndrome in July of 2020,(110) almost 50 years after its therapeutic potential was first noticed.

Modern drug repurposing approaches attempt to utilise the vast wealth of disease and drug related data, to accelerate the process, shortening the previously described decades-long timescales for the serendipitous discovery of a new indication for existing drugs. These methods also work to identify drugs as potential therapeutic agents that may have never garnered any clinical suspicion if left without dedicated investigation.(99)

1.3.2.1 Computational approaches:

Signature matching: Signature matching is based on the comparison of certain identified unique characteristics of candidate compounds with another biological signature of interest .(111) Comparator signatures may be certain aspects of a disease, or clinical phenotype, or may be aspects of a drug or class of drugs which has already been found effective for the treatment of the disease.

The disease signature is typically derived from one of the following types of data: transcriptomic, proteomic data, or metabolomic.(99) Signatures may also be identified from the chemical structures of suspect proteins, or from effective drugs. Adverse event profiles may also be used.

Transcriptomic disease differential expression signatures can be acquired by comparing the gene expression profile of disease-affected tissue with the gene expression profile of healthy tissue. Comparison of the gene expression profiles then allows the identification of

transcriptomic changes associated with the disease.(112) A drug signature may be acquired by comparing drug-treated tissue with control tissue, thereby identifying the transcriptomic effects of drug administration.(113)

Signatures may be used in several ways. One of the most common involves reverse-gene expression profiling. RNA Sequencing (RNA-seq) studies have found that the pathophysiological processes implicated in disease lead to changes in the transcriptomic profile of an organism. These organisms may be human (post-mortem/post excision tissue) or tissue resulting from animal models. The disease-state transcriptomic profile is known as the disease-signature. Similarly, just as disease can alter transcriptomic profiles, so can drugs. therefore, it can be said that each drug has an expression signature. Reverse gene expression profiling attempts to characterise the diseases-expression signature and find drugs that have the inverse expression signature, with the aim of reversing the pathological expression profile.(114, 115) This is typically done pre-clinically, using animal models to provide tissue for transcriptomic profiling of the disease. Drug transcriptome profiles can be obtained from animal models also, or from cell lines.

The genome-wide transcriptome of drugs is publicly available in online databanks such as the connectivity map (Cmap, available at https://portals.broadinstitute.org/cmap/).(116) Cmap is a large-scale publicly available library and collaborative project that aims to build a centralised, comprehensive "connectivity map" that includes data from assays on drugs, genes, and diseases, and how these three aspects interact or "connect". This includes data on drugs, diseases, genomics and transcriptomics among others, all of which are available publicly for querying. Much work has been done to expand the data provided by Cmap. The Library of Integrated Network-based Cellular Signatures (LINCS) works to expand Cmap by providing the transcriptomic profiles of drugs, therefore providing the necessary data to conduct drug repurposing studies.(116) The public availability of this data allows current research in signature-based drug repositioning to focus on computational elements of research methods, further reducing the time and resource cost of drug repurposing projects. This method, while seeming perhaps somewhat simplistic, has been demonstrated to be effective in predicting effective drugs for many different diseases, for example, the metabolic syndrome by Wagner et al.(117) Signature matching has also been used by Wei et al. to identify drugs that could be used as chemosensitizers in acute lymphoblastic leukaemia.(118)

This method has several limitations. The drug expression signature is acquired via *in vitro* application of the drug onto cell lines and tissue samples and observing for transcriptomic changes. This will not reproduce exactly the effects of the drug *in vivo,* as ingested drugs may be subject to processes not replicated within this experiment, such as gastrointestinal absorption mechanics, and related intra-hepatic drug transformations by liver enzymes such as CYP450, by drug-protein interactions within the plasma. This will also not take into account limitations of specific tissue penetration. This is especially true for CNS disorders such as epilepsy where blood-brain barrier penetration is essential for the success of any candidate drug The effects of the described *in vivo* drug transformations may lead to an *in vivo* transcriptomic response that differs from the response *in vitro*.(119)

Drug-drug similarity approaches (Guilt by association): This approach capitalises on knowledge of drugs/compounds that are already known to exhibit efficacy in the disease.(99) This method utilises similarities between the molecular signatures of drugs. The signature once again could be transcriptomic, proteomic or metabolomic. Drugs that are found to have similar signatures to known therapeutic agents are identified as having potential therapeutic benefit.(120) Drugs with similar molecular signatures are thought to share common targets, including targets that may not have been thought to be of relevance in the studied disease process, but may still be of therapeutic value.(120)

Similarity ensemble approach (SEA): First the chemical structure of efficacious drugs is elucidated and then studied to identify favourable chemical characteristics of the effective drug set.(99, 121) Compound libraries are then computationally screened for drugs which share any of these characteristics, and are ranked based on the number of shared structural aspects.(122) The chemical structure of the base drug is studied, as opposed to the drug forms that have undergone hepatic metabolism, or other pharmacokinetic transformations etc. Therefore, drugs which may seem chemically similar *in silico* may in fact be chemically distant in *vivo*.

Computational molecular docking: This utilises computational systems and 3D molecular conformational fingerprint structure data to attempt to identify the most complementary ligand for any given target by screening a drug library containing drug structures against the 3D structure of the target.(99, 123) It may also be used inversely, screening a multitude of implicated targets against a single drug, and then repeating the process for different drugs,

in order to rank drugs on their ability to bind to different targets.(124) The primary limitation of docking methods is the lack of sources which contain 3D data. This is especially the case for compounds and drugs that are less well known, such as drugs that were used exclusively for very rare diseases, or drugs that were never licensed.

Drug repurposing based on Genome-wide association studies (GWAS) studies:
GWAS data can be utilised to identify genes which may not be previously known to be complicit in pathogenesis. This therefore provides additional targets for pharmacological agents to potentially act on.(125) Databases such as DrugBank (126) are used to identify drugs that affect the function of disease genes, and this data can then be utilised to rank drugs based on their ability to affect the function of culprit genes. Further *in vivo* testing however is required to establish efficacy.

Drug repurposing methods based on drugs' ability to affect the function of disease genes have a number of limitations, the first of which is that they are not able to distinguish between drugs that are able to influence the target in a desirable fashion (thus demonstrating drug efficacy), from drugs that would influence the target negatively, therefore potentially aggravating the disease.(122) These methods are therefore said to have lack of directionality when producing drug predictions. Secondly, knowledge of all of the proteins changed in function by any individual drug is still incomplete Finally, GWAS data is population specific, therefore target culprit loci found in European populations may not be commonly present in African populations, for example.(125, 127) Other details of GWAS-based study design are discussed further in section 1.4.

Network-based approaches: Network-based methodologies work to construct a network of potential drug targets around the known targets(128-130). Recent studies have demonstrated that the genes associated with a disease tend to cluster in the same network neighbourhood, called the disease module, representing a connected subnetwork within the interactome rich in disease proteins. Hypothesized that for a drug to be effective for a disease, it must target proteins within or in the immediate vicinity of the corresponding disease module.(131)

Analysis of routinely collected clinical data or of clinical trial data:

There are several examples of clinicians serendipitously discovering that a licensed drug that has been prescribed for its original indication is unexpectedly improving the symptoms of a different disease. A relatively recent example of this phenomenon is the observation by a physician that candesartan prescribed for hypertension lead to an improvement in migraine.(132) The vast amounts of systematically collected adverse effect data in clinical trials offers a valuable opportunity for discovering or confirming drug repurposing candidates. Retrospective analysis of adverse effect data from clinical trials of candesartan was used to confirm the aforementioned observation that candesartan is associated with a reduction in migraines.(132) A well-known example of drug repurposing opportunities being revealed by the collection of adverse effect data during a clinical trial is that of sildenafil.

Sildenafil, perhaps one of the more well-known products of drug repurposing, was originally designed for the relief of angina. It came to be used in its current indication, erectile dysfunction, through exploitation of its side-effect profile. Sildenafil was created to relieve symptoms via inhibition of phosphodiesterase-5 (PDE5) which is an enzyme involved in the relaxation of coronary arteries. Clinical trials of sildenafil for angina failed to demonstrate sufficient therapeutic impact., However, after conclusion of the trial, via the monitoring of patient reported side effects, sildenafil was found to cause prolonged erections in male participants. This led Pfizer, the drug's manufacturer, to investigate this property in a dedicated clinical trial. Sildenafil was licensed shortly after for erectile dysfunction, and being a first-in-class drug, achieved incredible success, with annual sales exceeding 1.5 billion USD.(133)

In order to exploit drug adverse effect data for drug repurposing, statistical frameworks have been designed. One framework (134) utilises routinely collected data to emulate RCTs by treating drug-treated patients as the "treatment cohort" and versus controls. It does this using causal inference and has been used to evaluate the effects of 259 drugs in Parkinson's Disease. Routinely collected clinical data is also being utilised to identify drug repurposing candidates. Other computational frameworks also exist that aim to predict drug repurposing candidates based on side effect data.(135, 136)

1.4 About GWAS

Since the completion of the Human Genome project in April of 2003, techniques used to sequence genomes have improved considerably, lowering the cost, and the time needed in order to sequence entire human genomes.(137)

It should be noted that the term SNP is used in reference to *common* variants in the population. Rarer variants, such as those implicated in the pathogenesis of cystic fibrosis, are typically called mutations. It should be noted however, that mutations are not restricted to rarer diseases, and in fact in some cases mutations can underlie common disease processes.(138) While both a SNP and a mutation may only affect a single base pair, and while both may be implicated in the pathogenesis of a disease, it is the frequency of observation of the genetic polymorphism within the population which determines the nomenclature used in reference to the variant, and to a lesser extent the effect size of the alteration.(139)

GWAS studies aim to determine genetic risk factors from the entire human genome towards susceptibility to a certain disease. GWAS operates on the "common disease, common variant" hypothesis, which theorises that if a disease is common throughout the population, then there are common genetic variants associated with the disease within the genome.(140) This assumes therefore, that the penetrance of any of these more frequent variants is likely to be considerably smaller, and that the effect of any one of these variants on the overall phenotype is also likely to be less substantial. This explains why if the minor allele frequency of a SNP was, for example 40%, a disease prevalence of 40% is not observed within the population. It therefore follows that in order to entirely explain the genetic heritability of common diseases as observed clinically, there must be multiple genetic polymorphisms contributing to the overall observed disease heritability, all of which contribute a relatively small amount to the total.

The National Human Genome Research Institute - European Bioinformatics Institute GWAS Catalogue (www.genome.gov/gwastudies)(141) now has over 3600 SNPs linked to common diseases, many of which increase the risk of susceptibility by approximately 1.2-2.0 times the population risk.

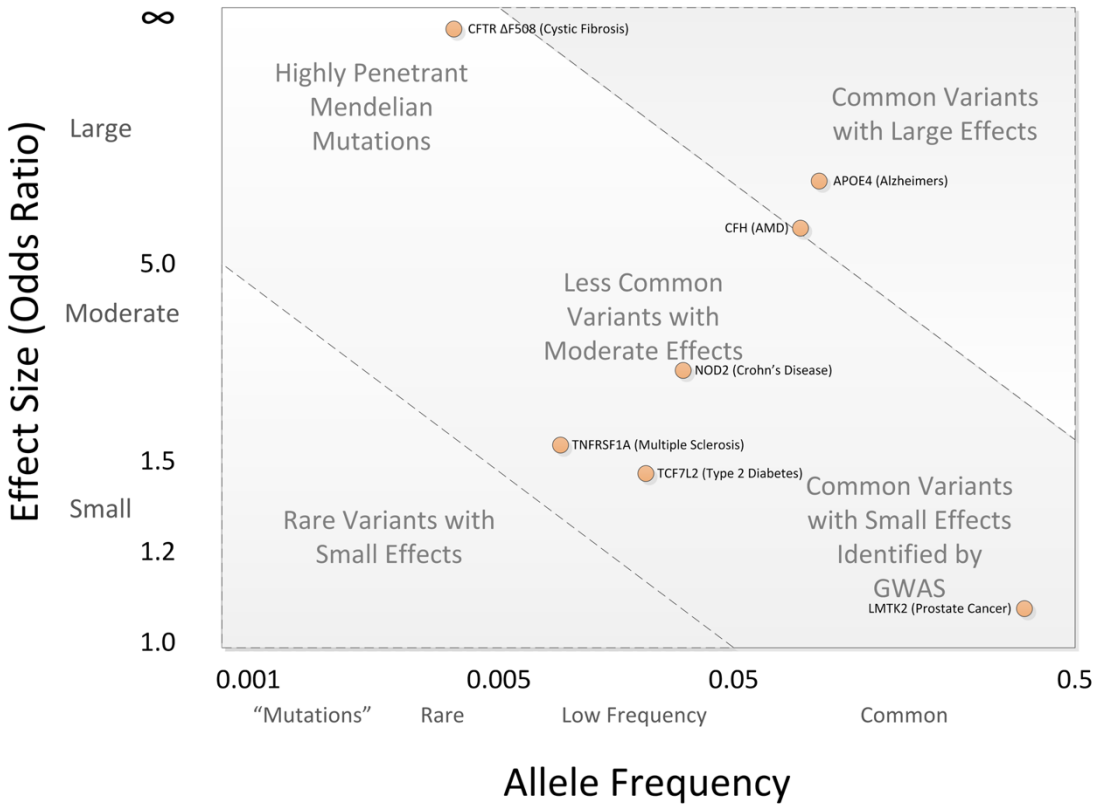


Figure 1.7: Illustrating the different magnitudes of effect genetic variation can phenotypically exert. So-called "Mendelian disorders" occupy the top left, whereas lower effect size variations typically identified by GWAs studies occupy the bottom right. Most disease-related genes lie on or around either of the two diagonals of the diagram. Reproduced from (127).

1.4.1 GWAS study design

Before delving into exploring the genotype, the phenotype of interest must first be defined. In genetic analyses, studied phenotypes can be quantitative or categorical. Studies examining categorical phenotypes requires phenotype standardisation, often in the form of strict clinical criteria that can be uniformly and consistently applied in order to diagnose or rule out a phenotype. Evidence demonstrating good sensitivity, specificity and intra-observer reliability for the set clinical criteria is essential before embarking on such studies.

Phenotype selection is essential for determining the possible statistical tests that can be conducted. The statistical tests chosen will be dependent on whether phenotype is categorical or quantitative. GWAS methodology will involve assessing for associations allele by allele, using appropriate statistical tests to measure the association between each locus and the trait.

In addition to the single locus associations, the wealth of genetic data produced by GWAS studies allows multi-locus analysis. However, as GWAs studies will often study 500,000 to 1 million SNPs, the vast quantity of genetic data prevents computational examination of all possible pair-wise combinations, even with the use of highly efficient computational algorithms. One method is to restrict study to SNPs that fall within a biological pathway or process.(142) Associations between SNPs within the selected subset can then be studied.
Following the identification of candidate SNPs, these SNPs are then mapped to genes. Identified genes may then be designated as candidate genes or loci, and may then become the subject of future study, and a target for drug repurposing approaches and initiatives.

1.5 Summary

Section 1.2 explored the traditional drug development process, its technical and scientific challenges, and the financial and chronological investment involved in the process. Sections 1.3 and 1.4 explored the current and developing drug-repurposing methods, alongside a number of practical examples of drugs that have been repurposed for new indications.

As of now, there is no one model or method that is superior to all others, leaving individual researchers to decide which drug repurposing method to use when attempting to produce drug prediction lists for an indication. Genetic, genomic, and transcriptomic drug repurposing methods have produced encouraging results for epilepsy.(143) Additionality, Monogenic and polygenic epilepsies can be viewed as separate entities requiring their own individual analyses, therefore, method selection must consider the suitability of the method for the type of epilepsy being studied.

1.6 Aims and structure of this thesis

New drugs are required for epilepsy. The aim of this thesis is to create drug repurposing resources and produce drug prediction lists for both monogenic and polygenic epilepsies.
Chapter 2 of this thesis describes a large-scale systematic review exploring genetic studies in epilepsy, with the goal of extracting the findings from the published work and collating it into a central database. Chapter 2 also describes one possible utilisation of the database, whereby the database is used alongside a public network-based drug repurposing platform to create drug prediction lists for common, polygenic generalised and focal epilepsies. Finally, Chapter 3 describes a computational transcriptome-based drug-repurposing strategy, suitable for monogenetic epilepsies, in this case applied to Dravet syndrome.

Chapter 2 : Creating and utilising the Seizure Associated Genes Across Species (SAGAS) database to predict drugs for epilepsy

2.1 Introduction

With new methods now available that allow for discovery of genes involved in many diseases, the ever-expanding wealth of literature in clinical genetics and bioinformatics leaves us with a vast body of knowledge regarding the genetic basis of disease. For many diseases, including epilepsy, the vast amounts of data that are now available to researchers make it difficult to pinpoint genes that may be of particular interest for future study. The diverse nature of scientific publication also creates difficulty in utilising the entirety of knowledge available to construct a set of genes that collectively have the highest degree of evidence for a particular disease, or to formulate a global view of the genomic networks that may underlie a disease process.

Thus, collecting, collating, and summarising available knowledge is paramount in maintaining the capacity to meaningfully utilise all available data for future work on a disease, which is especially key in diseases of primarily polygenetic origin, such as the common epilepsies. Creating a tool that aims to centralise all published information on epilepsy is therefore a key step towards accelerating mechanistic, diagnostic and therapeutic, genetic and genomic discoveries in epilepsy.

As of writing, we were not aware of an existing up-to-date database of epilepsy genes. A previously published database (available at: http://www.wzgenomics.cn/EpilepsyGene/download.php) was last updated seven years ago and is no longer fully functional. Previously published systematic collations of epilepsy genes have limited themselves to data from human studies, even though information from animal studies provides vital insights into the genetic basis of seizures/epilepsy. Previously published epilepsy gene datasets do not provide any objective numerical metrics that can be used to prioritize genes.

Our aim was to create a tool which provides a comprehensive list of all genes that may influence susceptibility to seizures, or that may modulate their frequency or severity. The tool would also provide a complete list of all the evidence for that gene, and summarise the

clinical phenotypes caused by mutations of that gene, for each instance of evidence. We also aimed to include data from GWAS studies in order to be able to give an indication as to the degree of associations of each gene with polygenic focal and generalised epilepsies. Finally, with large proportions of epilepsy research being conducted in animal models, we intend to provide the information from all genetic and genomic studies conducted in epilepsy, across all species, and to be able to display this information in a succinct and easily accessible fashion.

2.1.1 Structure of this chapter

This chapter is structured as two distinct works, each presented in its own subsection. Each subsection describes the methods pursued, and the reported results independently. Section 2.2 discusses the construction and creation of the SAGAS database, whereas section 2.3 discusses one possible utilisation of the SAGAS database for drug repurposing. Finally, section 2.4 summarises and discusses both works together.

2.2 The SAGAS database

2.2.1 Methods: Electronic searches

We searched SCOPUS for studies reporting one or more named genes whose mutation or manipulation led to seizures/epilepsy in people or animal models. Separate searches were conducted for human and for animal model studies. For animal studies, we included all years to 15 December 2020. For human studies, we included all years from 1 January 2016 to 15 December 2020. Earlier years were not included in the search for human studies as human studies up to and including 2015 had been collated by two previously published comprehensive human epilepsy gene databases/datasets, whose data was downloaded/extracted in bulk. Appendix 1 shows the search terms used in SCOPUS.

2.2.1.1 Inclusion criteria

For animal model studies, the article must demonstrate that direct manipulation of the gene/protein (such as CRISPR, RNA interference etc.) has caused seizures or epilepsy. Studies reporting clinical or electrical seizures were eligible. We have included gene manipulation studies that used the following techniques: gene knock-in/knock-out, gene knock down, utilisation of transgenic subjects, RNA interference, CRISPR, antisense oligonucleotides, CAS9, cre recombinase, optogenetic modulation, Targeted Augmentation of Nuclear Gene Output (TANGO), molecular switch and viral vectors. Administration of drugs or

pharmacological compounds that affect a protein/gene were not accepted, as the compound could also influence proteins/genes other than the target gene (as discussed in chapter 1) and, therefore, it cannot be definitively concluded that the effect of the compound on the intended target is the sole cause of the observed phenotype in studied subjects. We included human studies reporting an association between seizures/epilepsy and mutation(s), variation(s) or polymorphism(s) in one or more named genes.

2.2.1.2 Manual screening strategy

Given the large volume of articles that required screening, a number of students were invited to contribute to the data collection process. Of the invited students, 62 contributed to the screening and data extraction process. The students attended a seminar where they were instructed on how to perform the searches, alongside examples of studies to include and studies to exclude. Standardised instruction booklets were also distributed to the students. For screening studies and standardising data entry, we utilised an online standard data collection proforma using the online systematic review platform Sysrev (accessible at https://sysrev.com/). The data collection proforma is available in Appendix 2. All articles were screened by at least 2 authors to ensure accurate paper selection and data collection. All conflicts were resolved by a team of four senior reviewers. Finally, all collated information was downloaded, checked for errors and filtered by the author of this thesis.

2.2.2 Incorporating data from existing databases

Additionally, to ensure that all genes associated with epilepsy and seizures were identified by the CACHEP initiative, the results of the search was merged with data already present within existing databases.

The databases utilised were: Clinvar (https://www.ncbi.nlm.nih.gov/clinvar/),(144) OMIM (https://www.omim.org/),(145) HGMD(https://digitalinsights.qiagen.com/products-overview/clinical-insights-portfolio/human-gene-mutation-database/),(146) MGI (http://www.informatics.jax.org/),(147) IMPC (https://www.mousephenotype.org/), EpilepsyGene (http://www.wzgenomics.cn/EpilepsyGene/)(148) and epiGADS,(149) (http://www.epigad.org/).(150) Data from a review of large number of epilepsy-associated genes was also used.(148) Data from these databases was downloaded on (26/05/21).

Once the data from both the searches and external databases had been collated, data was extracted from each identified article of evidence, and then the articles were grouped by species. Any non-human genes were mapped to the orthologous human genes, using the *flybase* database (151) (available at: https://flybase.org/reports/FBgn0039647.html) for outlying Drosophila genes and via manual searching for all other genes. Additionally, from the studies found through literature search, the reported clinical syndromes were extracted by hand for all included records, and for the genes obtained from existing databases, the associated clinical syndromes were downloaded, if they were provided. A further final manual review of the database was then conducted to exclude any spurious records.

We also wished to include a measure of the degree of association between genes and common epilepsies, which polygenic and complex genetic conditions. GWAS gene-based p-values were calculated for the two main types of common epilepsy—focal and generalised—from their GWAS summary statistics (152) using FUMA (https://fuma.ctglab.nl/) with default settings. This GWAS, conducted in 2018, studies 15,212 patients with epilepsy, alongside 29,677 controls. The GWAS gene-based p-values were then added to the record for each gene. Finally, a parent summary table was then created that summarised the evidence for each gene by displaying the gene name alongside the GWAS p-values and the total number of pieces of evidence available within the database for that gene for each species.

2.2.3 Results: The SAGAS Database

The initial Medline search on Scopus yielded 9541 articles, which were then reviewed manually. After manual review, 3569 articles, which provided evidence of association with epilepsy fora total of 1402 genes, were included in the database. Additionally, a further 1628 genes, supported by a further 6334 pieces of evidence, were identified from external pre-existing databases, and included in our database. Our final database homepage is available at (https://www.liverpool.ac.uk/translational-medicine/research/groups/d3re/sagas/). The database is hosted online and is available for viewing, query and download at (Sagas.ac). The SAGAS database contains a total of 9752 published pieces of evidence for 2879 genes associated with epilepsy in 6 different animal species. Also provided is a summary of the phenotypes shown to be caused by perturbation of each gene in addition to a numerical value indicating the degree of genomic association between the gene and polygenic epilepsies.

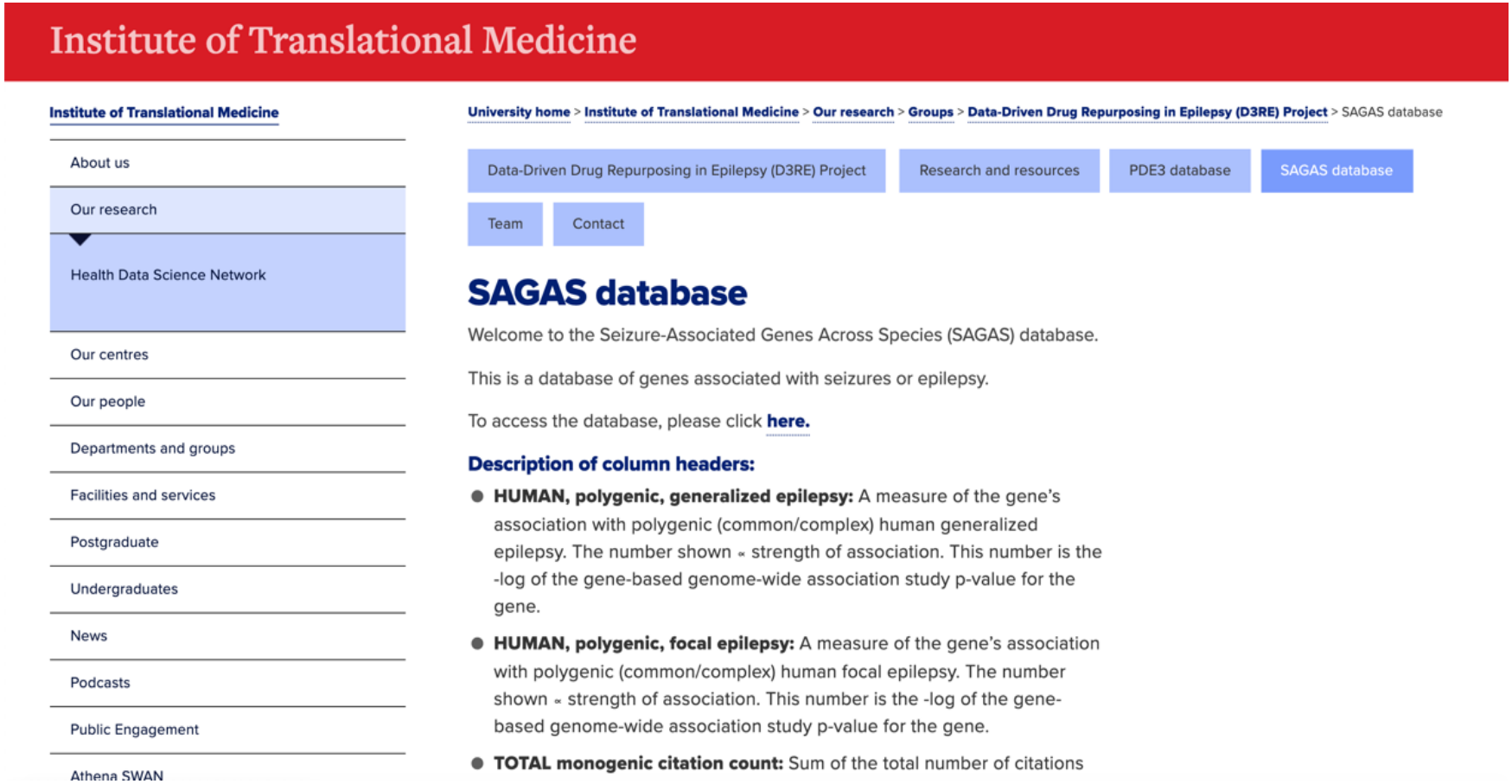
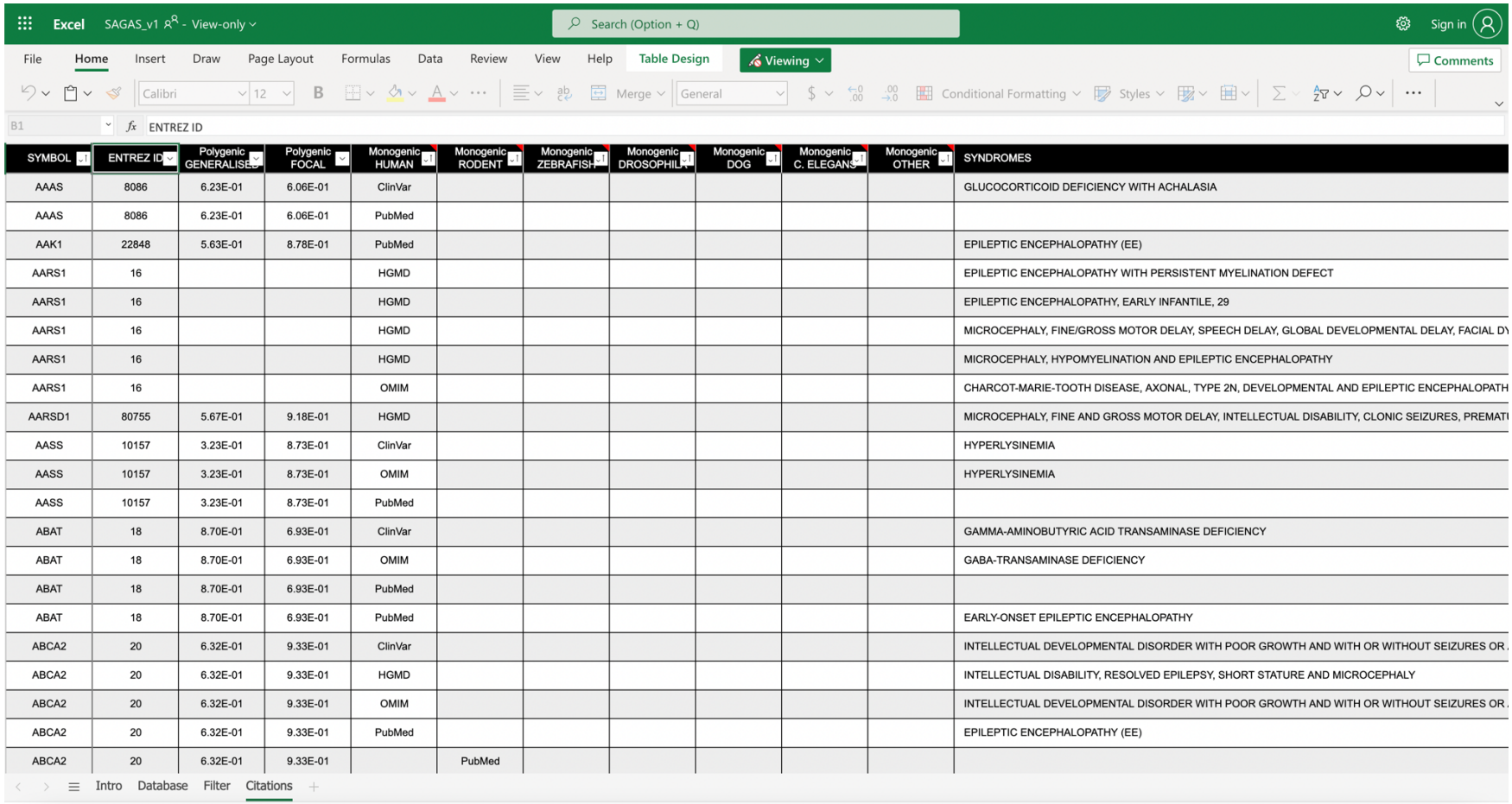


Figure 2.1: Screenshot of SAGAS website homepage

SYMBOL	ENTREZ ID	Polygenic GENERALISED	Polygenic FOCAL	Monogenic TOTAL	Monogenic HUMAN	Monogenic RODENT	Monogenic ZEBRAFISH	Monogenic DROSOPHIL	Monogenic DOG	Monogenic C. ELEGANS	Monogenic OTHER	SYNDROMES
AP3D1	8943	1.57E-03	8.63E-01	5	4	1	0	0	0	0	0	HERMANSKY-PUDLAK SYNDROME 10
CAMTA1	23261	2.13E-03	3.18E-01	2	2	0	0	0	0	0	0	CEREBELLAR ATAXIA, NONPROGRESSIVE, WITH MENTAL RETARDATION
SCN1A	6323	3.08E-03	8.57E-03	289	243	38	5	3	0	0	0	IGE, MTLE, ACUTE ENCEPHALITIS WITH REFRACTORY REPETITIVE PARTIAL SEIZURES, CRY
SETD1A	9739	1.14E-02	9.58E-01	3	3	0	0	0	0	0	0	DEVELOPMENTAL DISORDER, EPILEPSY AND PERSONALITY DISORDER, EPILEPSY, EARLY-O
STX1B	112755	2.16E-02	9.61E-01	8	6	2	0	0	0	0	0	FEBRILE SEIZURES & EPILEPSY, GENERALIZED EPILEPSY WITH FEBRILE SEIZURES PLUS, T
RAPGEF2	9693	2.85E-02	4.93E-01	4	3	1	0	0	0	0	0	EPILEPSY, FAMILIAL ADULT MYOCLONIC, 7
UBTF	7343	3.10E-02	9.11E-01	2	2	0	0	0	0	0	0	NEURODEGENERATION, CHILDHOOD-ONSET, WITH BRAIN ATROPHY, SEVERE EPILEPSY IN C
GABRA2	2555	3.23E-02	8.59E-01	9	5	4	0	0	0	0	0	ALCOHOL DEPENDENCE, SUSCEPTIBILITY TO, DEVELOPMENTAL AND EPILEPTIC ENCEPHAL
PCDH7	5099	3.31E-02	9.01E-01	1	1	0	0	0	0	0	0	
RIMS1	22999	3.31E-02	9.00E-01	1	0	1	0	0	0	0	0	
PHACTR1	221692	3.32E-02	9.01E-01	2	2	0	0	0	0	0	0	DEVELOPMENTAL AND EPILEPTIC ENCEPHALOPATHY 70, WEST SYNDROME
RBFOX1	54715	3.70E-02	8.52E-01	5	3	2	0	0	0	0	0	IDIOPATHIC GENERALIZED EPILEPSY (IGE), BENIGN EPILEPSY WITH CENTROTEMPORAL SPI
TTC21B	79809	3.84E-02	1.24E-01	2	2	0	0	0	0	0	0	FAILURE TO THRIVE, GLOBAL DEVELOPMENTAL DELAY, HYPOTONIA, SEIZURES, MUSCLE WI
MMP27	64066	4.38E-02	9.11E-01	1	1	0	0	0	0	0	0	EPILEPTIC ENCEPHALOPATHY (EE)
DOC2A	8448	4.79E-02	9.26E-01	1	0	0	1	0	0	0	0	
GRM4	2914	4.93E-02	8.76E-01	4	1	3	0	0	0	0	0	JUVENILE MYOCLONIC EPILEPSY
MYH14	79784	5.48E-02	7.19E-01	2	2	0	0	0	0	0	0	MALIGNANT MIGRATING PARTIAL SEIZURES OF INFANCY (MMPSI)
PFKM	5213	5.99E-02	9.00E-01	1	1	0	0	0	0	0	0	GLYCOGEN STORAGE DISEASE VII
CUX1	1523	6.80E-02	9.16E-01	2	2	0	0	0	0	0	0	MALIGNANT MIGRATING PARTIAL SEIZURES OF INFANCY (MMPSI)
MACF1	23499	6.90E-02	4.20E-01	2	2	0	0	0	0	0	0	LISSENCEPHALY 9 WITH COMPLEX BRAINSTEM MALFORMATION
ARID1B	57492	7.41E-02	8.86E-01	5	5	0	0	0	0	0	0	GLOBAL DEVELOPMENTAL DELAY AND SEIZURES, MODERATE INTELLECTUAL DISABILITY, SI

Figure 2.2: Screenshot of online SAGAS database at Sagas.ac . The "Citations" tab exhibits an outline of the genes included within the database, including the quantity and type of evidence for each species. The summary tab has one row per included gene.



SYMBOL	ENTREZ ID	Polygenic GENERALISED	Polygenic FOCAL	Monogenic HUMAN	Monogenic RODENT	Monogenic ZEBRAFISH	Monogenic DROSOPHILA	Monogenic DOG	Monogenic C. ELEGANS	Monogenic OTHER	SYNDROMES
AAAS	8086	6.23E-01	6.06E-01	ClinVar							GLUCOCORTICOID DEFICIENCY WITH ACHALASIA
AAAS	8086	6.23E-01	6.06E-01	PubMed							
AAK1	22848	5.63E-01	8.78E-01	PubMed							EPILEPTIC ENCEPHALOPATHY (EE)
AARS1	16			HGMD							EPILEPTIC ENCEPHALOPATHY WITH PERSISTENT MYELINATION DEFECT
AARS1	16			HGMD							EPILEPTIC ENCEPHALOPATHY, EARLY INFANTILE, 29
AARS1	16			HGMD							MICROCEPHALY, FINE/GROSS MOTOR DELAY, SPEECH DELAY, GLOBAL DEVELOPMENTAL DELAY, FACIAL DY
AARS1	16			HGMD							MICROCEPHALY, HYPOMYELINATION AND EPILEPTIC ENCEPHALOPATHY
AARS1	16			OMIM							CHARCOT-MARIE-TOOTH DISEASE, AXONAL, TYPE 2N, DEVELOPMENTAL AND EPILEPTIC ENCEPHALOPATH
AARSD1	80755	5.67E-01	9.18E-01	HGMD							MICROCEPHALY, FINE AND GROSS MOTOR DELAY, INTELLECTUAL DISABILITY, CLONIC SEIZURES, PREMATU
AASS	10157	3.23E-01	8.73E-01	ClinVar							HYPERLYSINEMIA
AASS	10157	3.23E-01	8.73E-01	OMIM							HYPERLYSINEMIA
AASS	10157	3.23E-01	8.73E-01	PubMed							
ABAT	18	8.70E-01	6.93E-01	ClinVar							GAMMA-AMINOBUTYRIC ACID TRANSAMINASE DEFICIENCY
ABAT	18	8.70E-01	6.93E-01	OMIM							GABA-TRANSAMINASE DEFICIENCY
ABAT	18	8.70E-01	6.93E-01	PubMed							
ABAT	18	8.70E-01	6.93E-01	PubMed							EARLY-ONSET EPILEPTIC ENCEPHALOPATHY
ABCA2	20	6.32E-01	9.33E-01	ClinVar							INTELLECTUAL DEVELOPMENTAL DISORDER WITH POOR GROWTH AND WITH OR WITHOUT SEIZURES OR
ABCA2	20	6.32E-01	9.33E-01	HGMD							INTELLECTUAL DISABILITY, RESOLVED EPILEPSY, SHORT STATURE AND MICROCEPHALY
ABCA2	20	6.32E-01	9.33E-01	OMIM							INTELLECTUAL DEVELOPMENTAL DISORDER WITH POOR GROWTH AND WITH OR WITHOUT SEIZURES OR
ABCA2	20	6.32E-01	9.33E-01	PubMed							EPILEPTIC ENCEPHALOPATHY (EE)
ABCA2	20	6.32E-01	9.33E-01		PubMed						

Figure 2.3: Screenshot of the Citations tab, with one row per article of evidence.

2.3 Utilising SAGAS for drug repurposing

2.3.1 Methods: Utilising SAGAS for Drug repurposing

We used the SAGAS database to predict drugs that could potentially be repurposed for epilepsy.

Different strategies can be envisaged for exploiting the SAGAS database in order to identify/predict potential drugs for epilepsy, to suit different research aims/priorities. Here, we use the SAGAS dataset to predict drugs for the two main types of common/complex/polygenic epilepsy: generalised and focal epilepsy. The Genome-Wide Association Study (GWAS) is becoming an increasingly powerful tool for revealing the distinct genetic determinants of different common epilepsies.(153-156) GWAS results are routinely used to predict new candidate drugs for complex diseases. In the standard approach, significant variants from the GWAS are mapped to genes; drugs that are known to affect the (protein products) of the genes, are predicted to affect the disease(157) This simplistic approach has a number of methodological deficiencies, one of which is that potential causal variants below the genome-wide disease significance threshold are ignored. Here, we exploit the comprehensive collation of data in the SAGAS database to identify the top GWAS 'hits' that are likely to be true associations, and potentially causal. Specifically, we select from the top 5% of genes with the most significant GWAS p-values those that also have evidence of association with epilepsy/seizures in monogenic epilepsies and/or animal models of epilepsy. To do this, we utilised an online drug repurposing platform, GUILDIFY (Available: http://aleph.upf.edu/guildify2/).(158) This constructs a network, by mapping the inputted genes ("seeds") onto a genome-wide protein-protein interaction network. GUILDIFY uses a series of graph theory algorithms to perform this. The result is a sub-network that includes both the seed genes and other genes that exist within that genetic neighbourhood.

It then uses drug target data extracted from DrugBank (available at: https://go.drugbank.com/),(126) DGIdb,(159) DrugCentral(160) and Chembl(161) to add drugs to the network. The top 1-5% of proteins are highlighted in a subnetwork (visually, these are represented as green nodes, see Figure 2.4 for results). Drugs are added to the network if their targets are available in the network. Each drug in the network is assigned a score ("GUILDIFY score", see table 2.1) based on the geometric distance to its target in the sub-network, and the number of its targets that are available in the generated network.

2.3.1.1 Algorithm selection

GUILDIFY offers five different algorithms for creating a protein-protein interaction network from seed genes: NetScore, NetZscore, NetShort, DIAMOnD and NetCombo. NetCombo is a consensus algorithm which combines the results of NetScore, NetZscore and NetShort, and outputs the mean of each the three scores. We used the NetCombo algorithm, as it has been shown to perform better than the other 4 algorithms available.(162)

2.3.1.2 Seed selection

We selected from the top 5% of genes with the most significant GWAS p-values those that also have evidence of association with epilepsy/seizures in monogenic epilepsies and/or animal models of epilepsy. This was done separately for both the GWAS for focal epilepsy, and the GWAS for generalised epilepsies. These "seed sets" were then inputted separately into GUILDify with the goal of generating a drug-prediction list for each of generalised and focal epilepsy. The decision to separate the seed-sets used in the analyses for focal/generalised epilepsy takes into consideration that different antiepileptic drugs are most effective for generalized and focal epilepsies, and some antiepileptic drugs can aggravate some generalised seizure types.

2.3.1.3 Drug selection

Once GUILDify had generated the sub-network of interest, and the accompanying drug prediction list, each of the resulting drug prediction lists was manually filtered to remove drugs that are unusable as long-term preventative antiseizure treatments in people with epilepsy. This included: antibiotics, antineoplastic drugs, drugs known to be toxic according to published literature, drugs known to aggravate seizures, drugs requiring parenteral administration, and drugs that with no published evidence of blood brain barrier permeability. The top 10 drugs from the resulting list were then selected.

2.3.2 Results: Drug repurposing

The genes selected for each GUILDify NetCombo analysis are listed in Appendix 3. A total of 2 subnetworks were generated, which successfully formed cohesive and interconnected gene networks, visualised in figure 2.4.

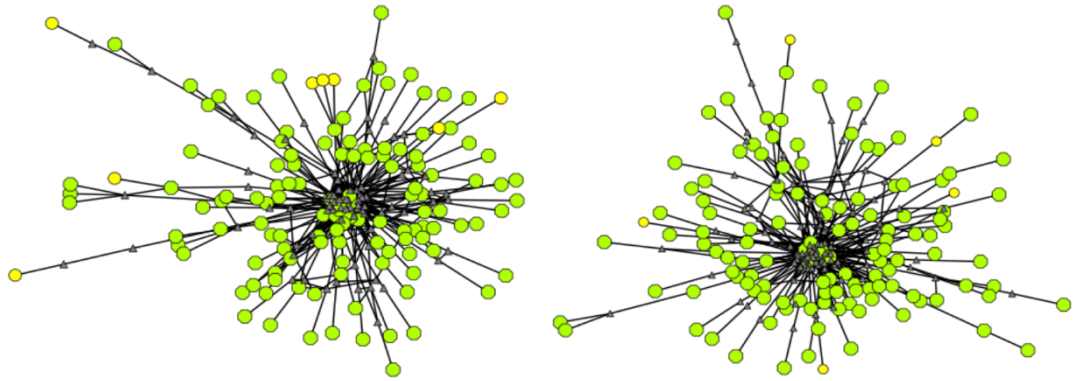


Figure 2.4: Visualisation of the subnetworks produced. The subnetwork generated for focal epilepsy is generated on the left, whereas the subnetwork for generalised epilepsy is shown on the right. Green node = Seed, yellow node = top 1% non-seed node. Note that drugs are not visualised in the above figure..

The results of the GUILDify analyses are available online. The GWAS based results are accessible for focal epilepsy at: (http://aleph.upf.edu/guildify2/result/b215138c-62a0-48df-bece-cea0d3d04d1b/1/20/1) and for generalised epilepsy at: (http://aleph.upf.edu/guildify2/result/1f71fa0c-f769-48c2-97de-e3e201bbb4f5/1/20/1)

2.3.2.1 SAGAS and GUILDify drug predictions

Presented within tables 2.1 and 2.2 (overleaf) are the drug repurposing lists for focal and generalised epilepsy.

Table 2.1: Drug prediction table for Focal epilepsy

Rank	Drug	Investigational Status	GUILDify Score	Target
1	Metformin	Approved	0.83256	PRKAB1 ETFDH GPD1
2	Copper	Investigational, approved	0.83119	140+ known targets
3	Caffeine	Approved	0.83116	ADORA PDE PIK3
4	Riluzole	Investigational, approved	0.82134	SCN5A SLC7A11
5*	Bepridil	Approved, withdrawn	0.81679	CACNA1 CACNA2D2 KCNQ1
6*	Verapamil	Approved	0.81679	CACNA1 KCNJ11 ADRA1
7*	Dronedarone	Approved	0.81679	SCN5A KCNH2 CACNA1
8*	Safinamide	Investigational, approved	0.81679	MAOB
9	Primidone	Approved, vet_approved	0.80483	GABRA(1-6) CHRNA(4+7) GRIA2
10	Topiramate	Approved	0.80483	GABRA1 SCN(1-10)A GRIK(1-5)

**: These drugs are tied in ranking as per their GUILDify score*

Table 2.2: Drug prediction table for generalised epilepsy

Rank	Drug	Investigational Status	GUILDify Score
1*	Nifedipine	Approved	0.78470
2*	Lamotrigine	Investigational, approved	0.78470
3†	Imipramine	Approved	0.66985
4†	Quinidine	Investigational, approved	0.66985
5	Valproic Acid	Investigational, approved	0.66512
7‡	Progabide	Investigational, approved	0.66038
8‡	Nisoldipine	Approved	0.66008
8§	Topiramate	Approved	0.65081
9§	Carbamazepine	Investigational, approved	0.65081
10§	Primidone	Approved	0.65081

**†‡§: These drugs are tied in ranking as per their GUILDify score*

2.4 Discussion

We present the SAGAS database. SAGAS represents the largest and most comprehensive of genetic repositories for epilepsy, comprising 9752 articles of published evidence for 2879 genes associated with epilepsy, and containing evidence from 6 species, and presenting a summary of syndromes that be caused by a mutation/variation of each gene.

The SAGAS database possesses notable strengths. The following points make our database different from and substantially better than all similar resources:

- The latest published systematic collation of epilepsy genes includes only 977 genes, whereas our database includes 2879 genes. Our database demonstrates that the number of genes that can potentially contribute to seizures/epilepsy is much higher than previously envisaged.
- Previously published epilepsy databases/datasets are limited to genes that cause monogenic forms of human epilepsy. Our database also incorporates the strength of association of each gene with polygenic forms of human epilepsy.
- Previously published epilepsy databases/datasets present only data from gene mutation/variation analyses in people with epilepsy. The discovery of a gene mutation in an individual with epilepsy is not always sufficient to establish that the gene causes epilepsy. For all genes found to bear mutations/variations in people

with epilepsy, we also present any existing evidence of their association with seizures in animal models, which can be critical for establishing the genes' pathogenicity.

- We also include genes that have evidence of association with seizures in animals, but do not yet have such evidence in humans. These genes are potentially important leads for mechanistic, diagnostic and therapeutic discovery in human epilepsy.
- Previously published epilepsy databases/datasets are limited to studies of inherited or de novo gene mutations/variations. We have also included gene manipulation studies that used various techniques such as: gene knock-in/knock-out, RNA interference, CRISPR, antisense oligonucleotides, cre recombinase, optogenetic modulation, and viral vectors.
- For each gene, our database displays a numeric value for the strength of its association with polygenic (common/complex) forms of epilepsy, and the number of articles of evidence that show its association with monogenic seizures/epilepsy in each animal species. This allows users to rank genes according to the amount of evidence of their association with seizures/epilepsy, and to select the genes that meet any threshold chosen by the user.
- It is possible to identify genes that have evidence of association with seizures/epilepsy in one or more specific specie(s) of interest and that have the desired number of articles of evidence for the specie(s).
- The SAGAS database allows for flexible search options. It is possible to search for specific genes and/or epileptic syndromes. Finally, SAGAS is designed to be suitable for both manual browsing as well as bulk computational querying.

While a concerted effort has been made to include in the SAGAS database all genes with published evidence of association with epilepsy, some genes may have been mistakenly omitted, if our literature search drugs failed to identify the relevant studies. Conversely, in a database of this size, which also includes data collated from pre-existing databases, it is possible that a small number of genes are erroneously included. We have not assessed or evaluated the quality of the studies reporting associations of gene mutations/variations with epilepsy; some reports might be erroneous because of poor study design or execution. We provide links to each study/site/database reporting an association between a gene and epilepsy, so individual reports of interest can be scrutinized by readers. Additionally, readers

can select genes that have corroborated evidence of association with epilepsy from multiple studies/species, as these are more likely to be true associations.

We use a publicly available network-based drug repurposing method to predict drugs for both generalised and focal epilepsies, based on genes selected from the SAGAS database. Of the drug predictions generated, 2 of the 10 predicted drugs for focal epilepsy are already licensed AEDs, whereas 6 of the 10 drugs predicted for generalised epilepsy are AEDs, which is an encouraging indicator of our methods' ability to select drugs effective for the treatment of epilepsy. Additionally, of the drugs that are currently not classified as AEDs, some have already been found to exhibit antiepileptic effects in experimental models, and in small sample human studies. Imipramine for example, a tricyclic anti-depressant, ranked 3-5th for generalised epilepsy has been found to produce an anti-epileptic effect in small case series,(163-165) and has been recognised as having potential as a future AED by a more recent review.(166) Metformin, the top ranked drug for focal epilepsy, is a commonly used drug used to control blood glucose levels in type II diabetes mellitus (T2DM). However, despite the pathophysiological differences between diabetes mellitus and epilepsy, metformin was ranked first in our analysis and has also been found experimentally to be therapeutic in animal models of epilepsy.(167, 168) A recent review summarising the results of 11 published papers exploring the effect of metformin in animal models found that metformin has been shown to control seizures, attenuate seizure generation and delay the onset of long-term cognitive effects of epilepsy.(167) Riluzole, ranked 4th on the drug prediction list for focal epilepsy, is a drug used in the treatment of amyotrophic lateral sclerosis. Riluzole functions by blocking glutaminergic neurotransmission in the CNS,(169) and has been found to have anti-epileptic properties, both as stand-alone therapy (170) and as an adjunct.(171)

The drug prediction methodology used in this work lacks directionality. Directionality refers to a drug prediction method's ability to predict not only that a drug will affect a disease, but also whether that effect is to alleviate or aggravate the disease. This is a recognised limitation of methods that use data for the ability of drugs to alter the function of genetically-associated disease-proteins in order to predict drugs that can affect the disease (128, 172, 173) as the direction of change in protein activity occurring in the disease is unknown. The lack of directionality could lead to the inclusion, amongst our drug predictions, of some compounds that may provoke seizures. For example, carbamazepine, ranked 8-10th in our

list for generalised epilepsy, is known to aggravate some generalised seizure types,(174, 175) despite being an AED that is widely used in focal epilepsy.(176)

Our drug predictions and analysis can be used as starting points for new drug-repurposing initiatives in generalised and focal epilepsy, potentially reducing the time and cost of finding new AEDs for these types of epilepsy. Of course, as with all drugs, our predicted candidate drugs require further *in vivo* animal model and/or human clinical trial evidence before being considered for deployment in clinical practice.

In conclusion, we present the SAGAS database, the largest database to date of epilepsy genes, and we showcase one possible use of the SAGAS database in predicting drugs for generalised and focal epilepsy.

Chapter 3 : Genomics-based drug repurposing for Dravet Syndrome

3.1 Introduction

More than 30 drugs are licensed for the treatment of epilepsy.(177) At least 173 other drugs have evidence of antiseizure efficacy in animal models.(178) Yet, few licensed or experimental antiseizure drugs are efficacious in people with or animal models of Dravet syndrome.(179, 180) Indeed, many of the currently licensed antiseizure drugs aggravate seizures in Dravet syndrome.(181) Complete seizure control remains unattainable for most people with the condition, and the goal of current therapy is to reduce the frequency of seizures, while minimizing the adverse effects of drugs. As such, it is desirable to identify additional drugs that can be used to treat Dravet syndrome, so that any individual with the condition has more treatment options available and, hence, better chances of finding an efficacious and well-tolerated treatment.

A monogenic epileptic syndrome is caused by a single mutant gene. However, knowing the identity of the mutant gene underlying a monogenic epileptic syndrome is not sufficient for predicting the effect of antiseizure medications on the syndrome. Dravet syndrome, the archetypal monogenic epileptic encephalopathy, is caused by mutations in *SCN1A*. Some antiseizure medications that alleviate seizures in Dravet syndrome do not affect *SCN1A*, whilst some antiseizure medications that affect *SCN1A* aggravate seizures in Dravet syndrome. We are not aware of any genomics-based methods that can correctly predict the varying effects of different antiseizure medications on Dravet syndrome (or any other monogenic epileptic syndrome). We create a novel method to predict drugs for Dravet syndrome that considers not only the gene that causes Dravet syndrome but also other genes that can influence the expression of its phenotype and show that our predictions correctly identify the antiseizure drugs that are effective, aggravating and equivocal for Dravet syndrome.

In recent years, there has been an increasing interest in using transcriptomics to aid and accelerate the drug discovery process. The approach is based on the following precepts: Every biological state and, hence, every disease state can be described by a gene expression signature. Treatments that restore gene expression patterns to their norm are associated

with amelioration of the disease phenotype.(8) The methodology can be summarised as follows. The first step is to generate a signature of differential gene expression for the disease through a genome-wide gene expression analysis comparing normal tissue with the disease tissue of interest. Then, using the gene expression signature of disease, databases of drug-induced gene expression signatures, such as Connectivity Map (Cmap) and Library of Integrated Network-based Cellular Signatures (LINCS) are queried.(116, 182) These databases contain drug-induced signatures of differential gene expression for thousands of compounds. Each signature is generated through genome-wide gene expression analysis comparing cells before and after drug exposure. If disease- and drug-induced signatures are sufficiently opposed (i.e. the genes upregulated in the disease-induced signature are downregulated in the drug-induced signature and *vice versa*) then the effect of the drug on transcription is opposite to the effect of the disease. Hence, the drug might revert the disease signature of differential gene expression and the disease phenotype itself. This method has successively produced a number of therapeutic leads for different diseases.(183)

In Dravet syndrome, a monogenic disorder, the alteration in function of the causative gene, *SCN1A*, leads to wider alterations in gene expression, which may come as a direct result of the mutation, or may be compensatory.(184) These wider transcriptomic changes allow the application of the above methodology to predict drugs that can potentially reverse Dravet syndrome's transcriptome and improve its clinical phenotype.

In this chapter, we use the above transcriptomic method to predict the efficacy of drugs for Dravet syndrome. We then enhance the drug predictions by also considering the effect of drugs on the function of protein products of genes that cause epilepsies like Dravet syndrome.

3.2 Methods

3.2.1 Data acquisition

The transcriptome Dravet syndrome was obtained from a published RNA-Seq analysis of hippocampi from a mouse model of Dravet syndrome.(184) Specifically, we extracted genes differentially expressed between the hippocampi of wildtype and *Scn1a*$^{+/-}$ mice of an epilepsy-susceptible strain, after the age of seizure onset, but without recent seizures. The disease gene expression signature comprised all genes identified as significantly differentially expressed in the published results. For comparative analysis:

- We extracted, from the same study, the transcriptomes of three other types of *Scn1a*$^{+/-}$ mice:
    1. *Scn1a*$^{+/-}$ mice of an epilepsy-susceptible strain ([129xC57BL/6J] F1) after age of seizure onset (3rd postnatal week, with GTCS occurring at P16-P19), with recent seizures. This is the transcriptome of Dravet syndrome, contaminated with transcriptomic changes induced by seizures.
    2. *Scn1a*$^{+/-}$ mice of an epilepsy-susceptible strain ([129xC57BL/6J] F1), before age of seizure onset. This is the transcriptome of mutant mice that are susceptible to but have not yet developed Dravet syndrome.
    3. *Scn1a*$^{+/-}$ mice of an epilepsy-resistant strain (29S6/SvEvTac). This is the transcriptome of mutant mice that are resistant to Dravet syndrome.
- Gene-expression changes associated with human focal epilepsy were imputed from the results of the most recent genome-side association study (GWAS) for epilepsy,(153) as previously described. we utilised focal epilepsy over generalised epilepsy as it more suited the role of a "baseline" for our model. Our hypothesised outcome for focal epilepsy was that all three classes of drugs in Dravet syndrome (effective, equivocal and aggravating) would be predicted to be efficacious by our model in focal epilepsy (as all drugs included are AEDs). Generalised epilepsy is less suited to this as some drugs known to be aggravating in Dravet syndrome are also aggravating in generalised epilepsy,(181) which would disrupt testing of this hypothesis.

3.2.2 Cosine distance

Disease and drug-induced transcriptomes were compared in order to predict each drug's relative ability to rectify disease-associated transcriptomic changes. This analysis was performed using the Combination Connectivity Mapping Bioconductor package and the Library of Integrated Network-Based Cellular Signatures (LINCS) data,(182) as previously described. This package utilizes cosine distance as the (dis)similarity metric. A higher (more negative) cosine distance value indicates that the drug induces gene-expression changes more strongly opposed to those associated with the disease. A lower (more positive) cosine distance value indicates that the drug induces gene-expression changes more similar to those associated with the disease.

Specific AEDs are considered first-line or second-line treatments for Dravet syndrome, whilst a number of other AEDs exacerbate seizures in Dravet syndrome. We divided AEDs into effective (those classed as first-line or second-line treatments for Dravet syndrome), aggravating, and equivocally effective (neither first-line or second-line treatments, nor aggravating) (Table 3.1). The set of first and second-line AEDs was based on recommendations from a North American panel (185) and an independent international (European and South American) panel. The set of aggravating AEDs was compiled as follows: A Medline search was conducted on the 25th of September 2020 for review articles with the terms Dravet AND (treatment OR management) in the title. Single-author papers were excluded; only multi-author papers were retained. This was done as multi-author papers typically have the advantage of being peer-reviewed by contributing authors, thereby improving the reliability and relative quality of the work. We include in our list of aggravating AEDs those that were listed as aggravating in two or more review papers. It should be noted that not all drugs identified through literature search are present in the drug prediction list due the unavailability of the drug expression signature for these drugs from the Cmap/L1000 database.

Table 3.1 AEDs classified as effective or aggravating.

AEDs	**Classification**	**Targets**
Clobazam	Effective	*GABARs*
Cannabidiol	Effective	*CNRs* + 42 others
Fenfluramine	Effective	*SLC6A4, HTRs*
Stiripentol	Effective	*GABARs, LDH*
Topiramate	Effective	*SCNs, GRIK1, CACNAs*
Valproicacid	Effective	*SCNs, PPARs*
Carbamazepine	Aggravating	*SCNs, CHRNA4*
Gabapentin	Aggravating	*CACNAs, KCNs*
Lamotrigine	Aggravating	*CACNAs, ADORAs,*
Oxcarbazepine	Aggravating	*AKRs, CBRs*
Phenytoin	Aggravating	*SCNs, KCNs, CACNAs*
Pregabalin	Aggravating	*CACNA2D1*
Tiagabine	Aggravating	*SLC6A1*
Vigabatrin	Aggravating	*ABAT*

3.2.3 In Silico Validation strategy

We compare the average cosine distance values and ranks of effective, equivocal, and aggravating AEDs within our drug prediction results. To ease conceptualisation and interpretation of results, we convert ranks to percentile ranks. For example, a drug with a percentile rank of 90 is ranked higher/better than 90% of all drugs. We use the median in order to compute the average of ranks, as it is less liable to being skewed by outliers. We determined the statistical significance of the drug prioritisation results by comparing the results to those from a null distribution generated by performing 106 random permutations.

3.2.4 Further validation

3.2.4.1 Further validation based on experimental drugs

As further validation, a literature search (using Medline on the 1st of October 2020) was conducted to collate a list of *experimental* drugs reported to have efficacious in Dravet syndrome. The experimental drugs were further classified as those that progressed to clinical trials in humans, and those drugs that did not (Table 3.2). We hypothesized that the average cosine distanced for all experimental drugs included would be negative. We also hypothesized that the drugs that progressed to human trials would exhibit a more negative average cosine distanced than the drugs that did not.

3.2.4.2 Further validation based on large-scale medium-throughput zebrafish model screening results

To further validate our model, we investigated whether our prediction model is able to predict the efficacy of drugs found to be effective in a large-scale medium-throughput zebrafish model drug screen by Dinday and Baraban.(186, 187) Dinday et al. utilised an *Scn1a* knockout model of zebrafish for large-scale medium-throughput screening of drug libraries. We acquired the database containing this data from https://barabanlab.ucsf.edu/drug-discovery-database (accessed on date 15/12/2020) and identified the drugs that were present in both this database, and the L1000 database. Effective drugs reported by Dinday et al. from their large-scale drug screening were noted. Equivocal and aggravating drugs are not identified by this model.

Table 3.2 List of all experimental drugs used in model validation

Drug Name	**Status**	**Targets**
ataluren	Trialled	Unknown
clemizole	Trialled	5-HT receptors
lorcaserin	Trialled	*HTR2C*
tak935	Trialled	*CH24H*
verapamil	Trialled	*CACNA, KCN*
62aminopropylbenzofuran	Experimental	*SLC6A, 5-HT2*
aa43279	Experimental	*SCN1A*
allopregnanolone	Experimental	*GABRA*
azd7325	Experimental	*GABRA2+3*
detomidine	Experimental	*ADRA2A*
dimethadione	Experimental	*CACNA1G*
donepezil	Experimental	*ACHE, BCHE*
efavirenz	Experimental	*pol* (HIV gene)
fluoxetine	Experimental	*SLC6A4, HTR2C*
gr46611	Experimental	*HTR1D*
gs967	Experimental	*SCN8A*
hm1a	Experimental	*SCN10A*
huperzinea	Experimental	*ACHE*
liraglutide	Experimental	*GLP1R*
methylergonovine	Experimental	*DRD1, HTR2B*
mifepristone	Experimental	*PGR, NR3, NR1*
mv1312	Experimental	*SCN1A, SCN8A*
mv1369	Experimental	*SCN1A, SCN8A*
pargyline	Experimental	*MAOB, MAOA*
progesterone	Experimental	*PGR, ESR*
promethazine	Experimental	*HRH1, DRD1*
pyrilamine	Experimental	*HRH1*
Trazodone	Experimental	*HTR, ADRA*

3.2.5 Improving drug predictions by creating and integrating the 'protein targets score'

We postulated that drug predictions could be improved by also considering the relevance to Dravet syndrome of proteins whose function is affected by each drug. For brevity, we use the terms proteins and genes interchangeably in the remainder of this chapter.

We hypothesised that: A drug is more likely to affect Dravet syndrome if it affects the function of genes that cause diseases that are more like Dravet syndrome. Our hypothesis is based on the following reasoning:

- A drug is likely to affect a disease if it affects another similar disease, the more similar the disease, the higher the likelihood.
- A drug is likely to affect a disease if it affects the gene(s) underlying the disease.
- Therefore, a drug is likely to affect a disease if it affects the gene(s) underlying another similar disease, the more similar the disease, the higher the likelihood.

Data for drugs and the proteins they affect was downloaded from the following databases: Drug-Gene Interaction Database 3.0 (http://www.dgidb.org/downloads; accessed 01/09/2020). STITCH 5.0 (http://stitch.embl.de/; accessed 01/09/2020). DrugBank (https://www.drugbank.ca/unearth/advanced/drugs; accessed 01/09/2020). The ChEMBL database (https://www.ebi.ac.uk/chembl/g/#browse/mechanisms_of_action; accessed 12/08/2019)

Genes were divided into seven categories, ranging from the most likely (*SCN1A*) to the least likely to cause diseases like Dravet syndrome (Table 3.3). Three of the categories contain genes that are not known to cause monogenic epilepsies/encephalopathies: genes expressed, elevated and enriched in the human brain (and are not necessarily implicated in epilepsy). The biological justification of their inclusion and order in our scoring system is as follows. Genes in these categories are not known to cause epilepsies, but are likely to contain yet undiscovered epilepsy-causing genes, with undiscovered epilepsy-causing genes being more enriched in brain-enriched genes than in brain-elevated genes than in brain-expressed genes. Compared to the set of genes expressed in the brain, the set of genes elevated in the brain is >2.2-fold enriched with genes that are known to cause monogenic epilepsies (hypergeometric distribution p-value < 6.0e-11). Compared to the set of genes expressed in the brain, the set of genes enriched in the brain is >5.6-fold enriched with genes that are

known to cause monogenic epilepsies (hypergeometric distribution p-value < 9.2e-15). This suggests that brain-enriched genes, more than brain-elevated genes, more than brain-expressed genes are likely to be causally relevant to epilepsy.

Every gene was ascribed a score equal to the fold enrichment of AEDs that affect Dravet syndrome amongst the drugs that affect any gene. This was calculated individually for the drugs that effect each gene category (I-VII) (Table 3.3). As alternative scoring systems, score increments ranging from 10% to 100% were applied to each successive gene category, from least to most likely to cause epilepsies like Dravet syndrome. These alternative scoring systems yielded similar results and the same conclusion (data not shown). The sum of the scores of all the genes a drug changes in function is the drug's 'protein targets score'.
All genes (whose protein products are expressed in the brain and changed in function by drugs) were categorised and scored as shown (Table 3.3). Enrichment is the fold enrichment of AEDs that affect (alleviate or aggravate) Dravet syndrome amongst all the drugs that affect the function of any gene in each category. Enrichment is calculated relative to all the drugs that affect the function of any gene across the human genome. Except for category VII, all enrichments were statistically significant (FDR-corrected hypergeometric equation p-value < 0.05). All weblinks were accessed on 13/11/2020.

Genes included in category II (Table 3.3) are genes in which mutations were shown to cause a phenotype in humans similar to that of Dravet syndrome. This was determined via literature search. Category II genes, alongside the studies reporting the studied phenotypes, are shown in Table 3.4.

3.2.5.1 Combined score

The combined score was calculated as follows:

1. Each drug's protein targets and cosine distance scores were multiplied together and then:
2. The drugs' cosine distance scores were rescaled between 1 and -1

The above strategy assigns the cosine distance score twice the weight as the protein targets score. We opted to assign the cosine distance score twice the weight as the protein targets score because the cosine distance score has directionality (see below) whereas the protein targets score does not.

3.2.6 Drug selection strategy

To ensure clinical viability, and suitability for use in epilepsy, drugs were then manually screened against the following set of criteria:

- Drugs must be safe for long-term use
- Drugs must be available for oral dosing.
- Drugs must be able to penetrate the Blood Brain Barrier.
- Drugs must not be an antimicrobial (Suggesting long term use of drugs with antimicrobial properties may encounter obstacles with adoption and licensing by health regulatory agencies.)
- Drug must not be an anti-cancer agent (because their clinically recognised side effect profiles)
- Drug must not be known to be aggravate seizures in any human epilepsy

Screening was conducted using published literature, and by utilising public drug information repositories, such as DrugBank (126).

Table 3.3: The scoring system used in the enhanced model. Also shown are the sources from which this gene data was extracted

#	Name	Description	Enrichment	Score
I	Cause Dravet syndrome	*SCN1A*	53	53
II	Cause Dravet-like syndrome	Gene mutations that cause phenotypes identical/similar to Dravet syndrome; collated from studies identified through a Medline literature search	26	26
III	Cause epileptic encephalopathy	https://www.omim.org/phenotypicSeries/PS308350 and https://www.omim.org/phenotypicSeries/PS617711	19	19
IV	Cause monogenic epilepsy	https://www.omim.org/	14	14
V	Enriched in brain	At least Eight-fold higher mRNA level in brain compared to any other tissues. https://www.proteinatlas.org/humanproteome/brain/human+brain	8	8
VI	Elevated in brain	At least four-fold higher mRNA level in brain compared to any other tissues or at least four-fold higher average level of mRNA expression in a group of 2-5 tissues, including brain, compared to all other tissues or at least four-fold higher mRNA level in brain compared to the average level in all other tissues https://www.proteinatlas.org/humanproteome/brain/human+brain	3	3
VII	Expressed in brain	https://www.proteinatlas.org/humanproteome/brain/human+brain	1	1

Table 3.4: Table of genes included in category II, which includes gene mutations that cause phenotypes similar to those observed in Dravet syndrome

Gene	Paper
SCN1A	https://onlinelibrary.wiley.com/doi/full/10.1111/dmcn.12709
SCN2A	https://pubmed.ncbi.nlm.nih.gov/19783390/
SCN1B	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3681183/pdf/tacg-4-113.pdf
SCN8A	https://www.ncbi.nlm.nih.gov/sites/books/NBK379665/
SCN9A	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2730533/pdf/pgen.1000649.pdf
GABRA1	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001207/pdf/NEUROLOGY2013548636.pdf
GABRG2	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5426359/
CHD2	https://www.cell.com/action/showPdf?pii=S0002-9297%2813%2900458-8
HCN1	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4325594/pdf/i1535-7511-14-6-348.pdf
KCNA2	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4380508/pdf/emss-62156.pdf
STXBP1	https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4001207/pdf/NEUROLOGY2013548636.pdf

3.3 Results

3.3.1 Drug predictions for Dravet syndrome are concordant with the clinical effects of AEDs in Dravet syndrome

In our predictions for Dravet syndrome, 100% of the AEDs that are clinically effective in Dravet syndrome (and are also found in the LINCS database) were ascribed negative cosine distance values, indicating that they are all predicted to improve the Dravet syndrome's transcriptome and phenotype. Additionally, effective drugs are ranked higher than equivocally effective drugs, which are ranked higher than aggravating drugs, with median percentile ranks 96, 64 and 24, respectively. These findings are unlikely to occur by chance (permutation-based $p<0.001$). The median cosine values for the effective, equivocally effective and aggravating group of drugs (-0.70, -0.31, 0.54, respectively) indicate that the

effective group of AEDs is predicted to be twice as effective as the equivocally effective group of AEDs, while the aggravating groups of AEDs is correctly predicted to be aggravating.

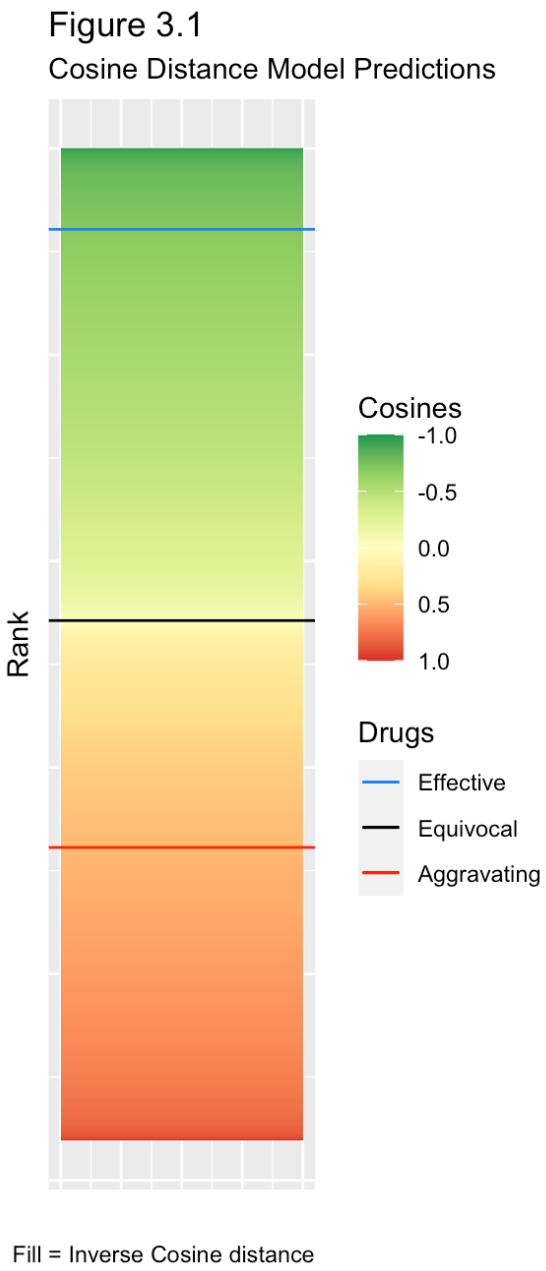


Figure 3.1: Heatmap illustrating average cosine distance values for effective, equivocal and aggravating AEDs, within the cosine distance values for all analysed drugs. a. Cosine distance ranges between +1 and −1, where +1 means that the disease-causing and drug-induced changes in the transcriptome are exactly the same, and −1 means that they are diametrically opposite.

The background colour gradient represents the model in use, with a higher intensity of green representing higher predicted efficacy, and more red intensity representing higher predicted aggravating potential (see legend). Each coloured row of pixels forming the gradient represents one drug, with the colour ascribed to that row representing the cosine difference (and hence predicted efficacy). Overall, 24,030 drugs make up the gradient for this figure.

3.3.1.1 Drug predictions for Dravet syndrome prioritise experimental drugs with evidence of efficacy, especially drugs that have progressed to clinical trials

In our drug predictions for Dravet syndrome, experimental drugs with evidence of efficacy in Dravet syndrome are prioritised (median percentile rank = 97), and experimental drugs that have progressed to clinical trials are prioritised even more (median percentile rank = 77) (Figure 3.3).

3.3.1.2 Drug predictions for Dravet syndrome prioritise drugs effective in high-throughput screening of compounds in the Zebrafish model of Dravet syndrome

At the time of analysis, Dinday et al. had screened 3333 drugs in the Dravet zebrafish model, of which 11 drugs were found to be effective (16/12/2020, last database update 17/2/2020). A total of 855 drugs were found both their database and our predictions dataset, 7 (out of a possible 11) of which were effective. In our drug predictions for Dravet syndrome, these drugs are prioritised (median percentile rank) (Figure 3.3).

3.3.1.3 Drug predictions for other transcriptomes are **not** concordant with the clinical effects of AEDs in Dravet syndrome

Drug predictions based on transcriptomes *not* of Dravet syndrome are *not* concordant with the clinically observed effects of drugs on Dravet syndrome (Table 3.5). This also shows that the drug predictions are consistent with clinical experience only when the correct diseased transcriptome is used.

For the data taken from human focal epilepsy, the prediction model produced negative cosine distances for each of the drug sets, which is consistent with expectations, as all three sets of drugs are clinically effective for focal epilepsy.

Table 3.5: Average cosine distance values for different phenotypes. Please note that cosine distance values are not comparable across different phenotypes.

Transcriptome Source	Description	Average cosine distance of antiseizure medications		
		Effective	Equivocal	Aggravating
Scn1a$^{+/-}$ mice of an epilepsy-susceptible strain, after age of seizure onset, without recent seizures	Transcriptome of Dravet syndrome, uncontaminated with transcriptomic changes induced by seizures	-0.70	-0.31	0.54
Scn1a$^{+/-}$ mice of an epilepsy-resistant strain	Transcriptome of mutant mice that are resistant to Dravet syndrome	0.99	0.99	-0.18
Scn1a$^{+/-}$ mice of an epilepsy-susceptible strain, before age of seizure onset	Transcriptome of mutant mice that are susceptible to but have not yet developed Dravet syndrome	0.97	-0.98	0.90
Scn1a$^{+/-}$ mice of an epilepsy-susceptible strain, after age of seizure onset, with recent seizures	Transcriptome of Dravet syndrome, contaminated with transcriptomic changes induced by seizures	0.31	0.16	0.14
Human focal epilepsy	Transcriptome of human focal epilepsy	-0.44	-0.36	-0.30

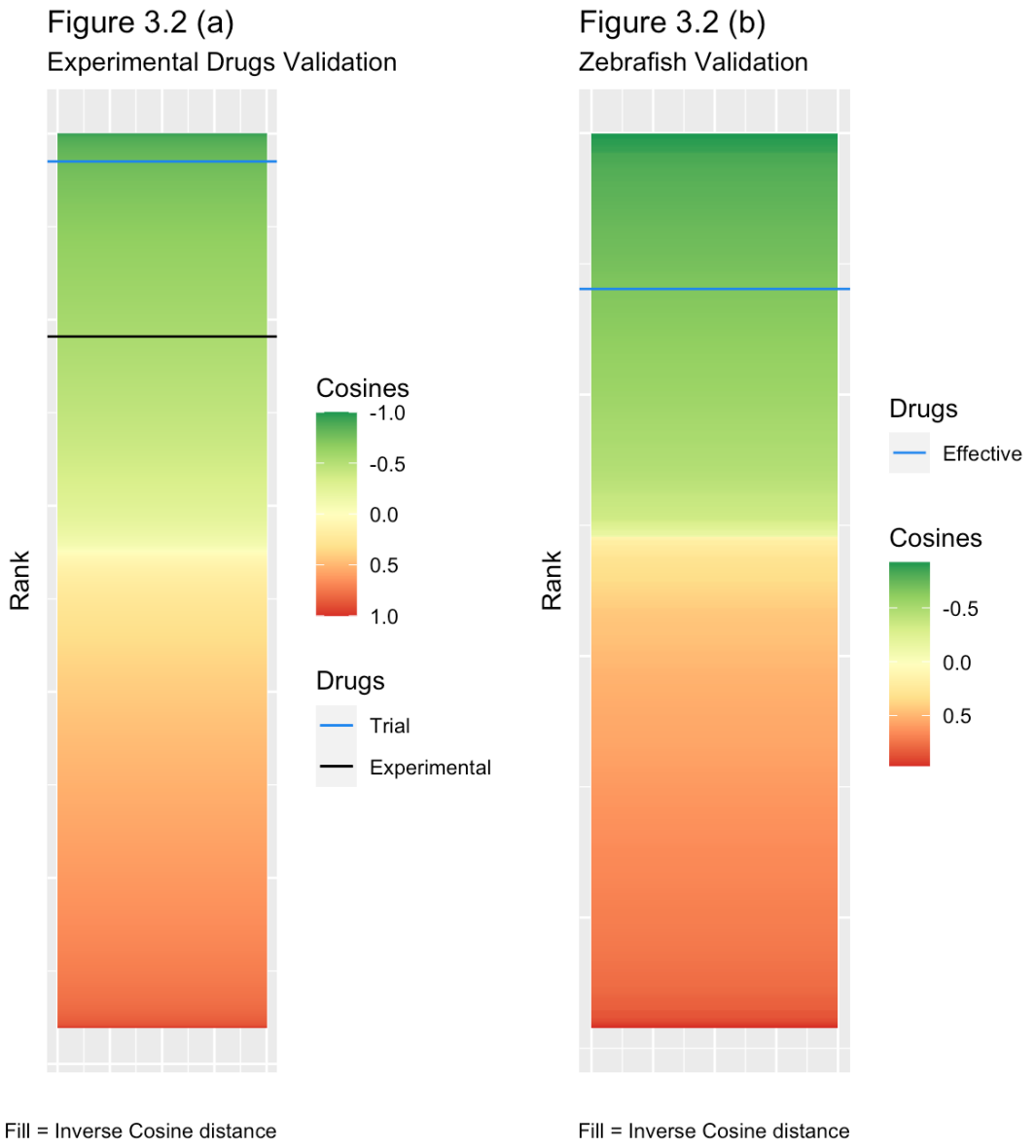


Figure 3.2: Further validation – 3.2(a): The prediction model applied to experimental drugs, both drug sets produce negative cosine distances, and drugs that progressed to trial having a higher predicted efficacy. 3.2(b): Drug predictions utilising a drug set that was found to be effective in a zebrafish model of Dravet syndrome.

The background colour gradient represents the model in use, with a higher intensity of green representing higher predicted efficacy, and more red intensity representing higher predicted aggravating potential (see legend). Each coloured row of pixels forming the gradient represents one drug, with the colour ascribed to that row representing the cosine difference (and hence predicted efficacy). Overall, 24,030 drugs make up this model.

3.3.3 Drug predictions are improved by incorporating the protein target scores

We created a combined score by integrating the cosine distance and protein target scores. This led to improved prioritisation of the AEDs that are clinically effective for Dravet syndrome and increased *deprioritisation* of AEDs that are clinically aggravating for Dravet syndrome (Table 3.5 and Figure 3.3).

Table 3.6 Median percentile ranks of AED sets in drug predictions based on different scores

	Using the cosine distance score only	Using the combined score
Effective AEDs	92	100
Equivocal AEDs	52	90
Aggravating AEDs	30	2

Figure 3.3

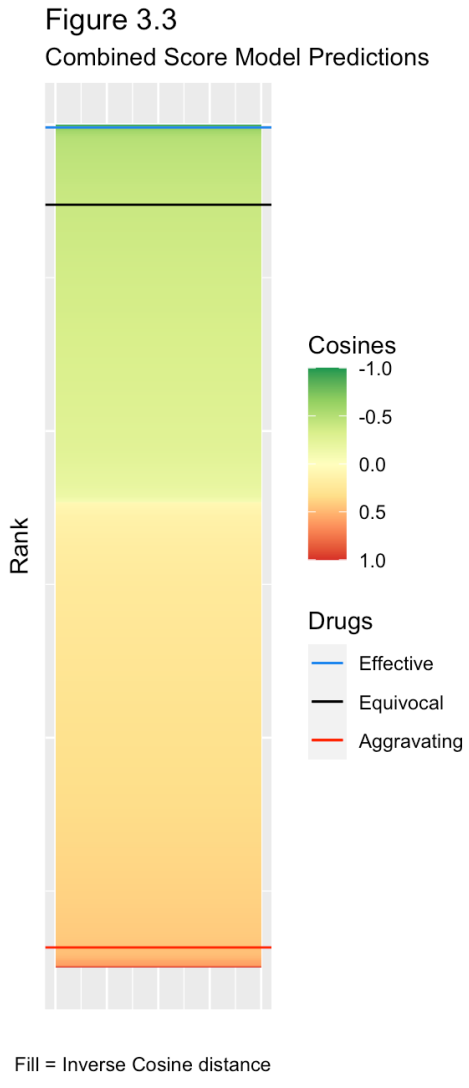


Figure 3.2: Figure 1.8: Illustrating the prediction model after application of expression based-scoring system. Effective drugs rank highly, whereas aggravating drugs rank much lower. Note that these values utilise medians, hence the position of the aggravating average.

The background colour gradient represents the model in use, with a higher intensity of green representing higher predicted efficacy, and more red intensity representing higher predicted aggravating potential (see legend). Each coloured row of pixels forming the gradient represents one drug, with the colour ascribed to that row representing the cosine difference (and hence predicted efficacy). Overall, 1,374 drugs are included in this model.

3.3.4 Final drug selection

The final list showing the top 10 drugs is available in Table 3.7.

Table 3.7: Final drug prediction list after manual curation based on drug safety and appropriateness for epilepsy.

Drug Name	Rank based on cosine distance score only	Rank based on combined score
topiramate	1	1
valproicacid	4	2
primidone	5	3
pentobarbital	9	4
verapamil	12	5
memantine	13	6
atomoxetine	17	7
stiripentol	18	8
pentoxifylline	21	9
carisoprodol	22	10

3.4 Discussion

We have used a computational genomics strategy to predict drugs' relative efficacy against seizures in Dravet syndrome. This is the first analysis that predicts not only effective drugs but also aggravating drugs for a monogenic epilepsy, and correctly ranks the sets of drugs that are effective, equivocally effective and aggravating for a monogenic epilepsy. Our dataset is a valuable resource for selecting candidate drugs that could potentially be repurposed for treating seizures in Dravet syndrome. The chosen candidate drugs will have to be validated in future animal model studies and/or human clinical trials before being deployed in clinical practice.

Our method has several limitations. *SCN1A* mutation-induced genome-wide transcriptomic dysregulation was assayed in tissue from an animal model rather than from people with Dravet syndrome. It is not known how closely the transcriptomic changes in this animal model of Dravet syndrome recapitulate the transcriptomic changes in people with Dravet syndrome. However, this animal model reproduces the clinical features (seizures, early

mortality etc.) and response to AEDs seen in people with Dravet syndrome, which suggests that its transcriptomic profile reproduces the transcriptomic profile of people with Dravet syndrome. This is reinforced by studies indicating that clinical diagnoses (of other disorders) and disease severity can be predicted by studying organism transcriptomics, which work to highlight the link between clinical manifestation and disease expression profiles.(188, 189) Importantly, we show that *SCN1A* mutation-induced genome-wide transcriptomic dysregulation must be assayed in brain tissue from Dravet syndrome model animals that have become susceptible to but have not yet developed seizures, as seizure-induced transcriptomic changes will likely confound *SCN1A* mutation-induced transcriptomic changes, once seizures have developed. Clearly, it is impossible to obtain human surgical/post-mortem brain tissue that meets these conditions. For some drugs, the cosine distance score is not currently available, as their transcriptomic effects are still to be assayed.

The protein targets score relies upon knowledge of the proteins changed in function by drugs. At present, knowledge of the proteins that are changed in function by each drug is incomplete, and it is more incomplete for some drugs than for others. The more incomplete the knowledge of the proteins changed in function by a drug, the more likely it is that the drug's protein targets score will be underestimated. By extension, the protein targets score is more likely to be underestimated for drugs that are less studied, as their modes of action are less analysed and, hence, knowledge of the proteins changed in function by them is less complete. Nevertheless, drug predictions can only be based upon available data, and the drugs with the higher protein targets scores are the ones that affect the higher number of causal genes, according to available data. Whilst the cosine distance score predicts directionality of drugs' effect (alleviating or aggravating), the protein targets score does not.

There is increasing interest in medium throughput animal model drug screening using lower animal species (C. elegans, and Drosophila etc.). Our findings suggest that using our method to create a shortlist of candidate drugs for medium throughput animal model screening can potentially increase the yield of the animal model screening many fold, decreasing the time and resource cost of the animal model screening. Furthermore, efficacy of a drug in lower animal species does not indicate that the drug is human blood-brain barrier permeable, but a drug preselected using our computational method is more likely to penetrate the brain and produce a clinical antiseizure effect in people. Additionally, cellular models of Dravet syndrome utilising induced pluripotent stem cells (IPSCs) have been developed.(190-192)

These models have elucidated the loss-of-function of the GABAergic system implicated in pathogenesis.(192) There may be potential to utilise these new cellular models, alongside knowledge of the pathobiology of Dravet syndrome, as a basis for a higher-throughput screen for candidate Dravet syndrome drugs.

Our method is potentially generalizable to other epileptic encephalopathies, other monogenic epilepsies and other monogenic diseases. Its applicability to other diseases will depend on the availability of appropriate animal models.

Chapter 4 : Outlook, and future work

65 million people have epilepsy worldwide,(193) and current AEDs have significant shortcomings, including an unfavourable side effect profile,(194) and an inability to produce seizure-freedom in 30% of patients.(195) Given the high time and resource costs of novel drug development, exploring alternative avenues of providing better drugs to epilepsy patients is desirable.

This thesis primarily aimed to create new drug repurposing resources, with the ultimate goal of creating a set of resources suitable for both polygenic and monogenic epilepsies. In the first chapter of this thesis, we discussed how drug repurposing provides an attractive alternative to traditional drug development methods. Practical examples of drugs being repurposed were discussed, including the fifty year-long repurposing journeys of fenfluramine. Emerging methods within drug repurposing were discussed that aim to systematically elucidate therapeutic opportunities for existing drugs, with the goal of reducing the decades needed for serendipitous discovery to a far more practicable timeframe.

The second chapter of this thesis described the creation of the SAGAS database, the largest the largest database of epilepsy genes across species. We used the SAGAS database, alongside the online GUILDify platform, to create drug prediction lists for both focal and generalised epilepsy.

Lastly, the third chapter of the thesis aimed to employ transcriptomics-based drug repurposing for Dravet syndrome, the archetypal monogenic epileptic encephalopathy. We enhanced our transcriptomic-based drug predictions by also considering the effect of drugs on the function of genes that cause Dravet syndrome or syndromes like it.

4.1 Future research

Following on from the work described within this thesis, animal model validation of some of the higher-ranking drugs would be the natural next step.

The drug repurposing techniques employed in this thesis are dependent on the availability of accurate drug-related data in order to produce accurate drug predictions. Therefore,

further studies are warranted expanding the publicly available data available to drug-repurposing teams. Specifically, more data on the protein targets and transcriptomic profiles of pharmacological compounds, especially unlicensed experimental compounds, would be valuable. The majority of data on available drugs is centred around licensed drugs which are currently in clinical use, but approximately 1,500 drugs only are currently licensed by the FDA for all human conditions.(196)

Further work can also be aimed at gathering further data on epilepsy syndromes, including disease transcriptomic profiles, and data on culprit genes associated with various epilepsy phenotypes. Recent studies, including the latest GWAS have revealed that the genes associated with different epilepsies are many and diverse. Gathering genetic, transcriptomic, and proteomic data on the different epilepsies will further enable drug repurposing analyses for these phenotypes.

In summary, the success of drug repurposing analyses is dependent on the data available for the analyses. The more comprehensive and accurate the data, the more promising the results we can hope to see from future drug repurposing initiatives.

4.2 Conclusion

In conclusion, this thesis has created new computational resources for drug repurposing and generated drug predictions for epilepsies of varying genetic aetiologies. We describe the development of the SAGAS database (available online at sagas.ac), the largest and most comprehensive database of genes implicated in epilepsy, covering all known and published genetic aetiologies in epilepsy, across all studied species. We also present our enhanced model for predicting drugs in epilepsies where data for both effective and aggravating drugs are available. And finally, we generate drug predictions for epilepsies of varying genetic aetiologies.

References

1. Magiorkinis E, Diamantis A, Sidiropoulou K, Panteliadis C. Highights in the History of Epilepsy: The Last 200 Years. Epilepsy Research and Treatment. 2014;2014:582039.
2. Stafstrom CE, Carmant L. Seizures and epilepsy: an overview for neuroscientists. Cold Spring Harb Perspect Med. 2015;5(6):a022426.
3. Beghi E. The Epidemiology of Epilepsy. Neuroepidemiology. 2020;54(2):185-91.
4. Fiest KM, Sauro KM, Wiebe S, Patten SB, Kwon C-S, Dykeman J, et al. Prevalence and incidence of epilepsy: A systematic review and meta-analysis of international studies. Neurology. 2017;88(3):296-303.
5. Sokka A, Olsen P, Kirjavainen J, Harju M, Keski-Nisula L, Räisänen S, et al. Etiology, syndrome diagnosis, and cognition in childhood-onset epilepsy: A population-based study. Epilepsia open. 2017;2(1):76-83.
6. World Health O. Atlas : epilepsy care in the world. Geneva: World Health Organization; 2005.
7. Fisher R ILAEI. Instruction manual for the ILAE 2017 operational classification of seizure types 2017 [Available from: https://www.ilae.org/files/dmfile/Operational-Classification---Instruction-manual-Fisher_et_al-2017-Epilepsia-1.pdf.
8. Epilepsy ILA. ILAE Classification of the Epilepsies (2017) 2017 [Available from: https://www.ilae.org/guidelines/definition-and-classification/ilae-classification-of-the-epilepsies-2017.
9. Jallon P, Latour P. Epidemiology of Idiopathic Generalized Epilepsies. Epilepsia. 2005;46(s9):10-4.
10. Cavazzuti GB. Epidemiology of different types of epilepsy in school age children of Modena, Italy. Epilepsia. 1980;21(1):57-62.
11. Banerjee PN, Filippi D, Allen Hauser W. The descriptive epidemiology of epilepsy-a review. Epilepsy Res. 2009;85(1):31-45.
12. Shinnar S, O'Dell C, Berg AT. Distribution of epilepsy syndromes in a cohort of children prospectively monitored from the time of their first unprovoked seizure. Epilepsia. 1999;40(10):1378-83.
13. Wataya-Kaneda M, Tanaka M, Hamasaki T, Katayama I. Trends in the prevalence of tuberous sclerosis complex manifestations: an epidemiological study of 166 Japanese patients. PLoS One. 2013;8(5):e63910-e.
14. Fejerman N, Caraballo R, Tenembaum SN. Atypical evolutions of benign localization-related epilepsies in children: are they predictable? Epilepsia. 2000;41(4):380-90.
15. (NORD) NOfRD. Angelman Syndrome 2018 [Available from: https://rarediseases.org/rare-diseases/angelman-syndrome/#:~:text=Angelman%20syndrome%20affects%20males%20and,prevalence%20in%20the%20general%20population.
16. (NICE) NioCE. Epilepsies: diagnosis and management 2020 [Available from: https://www.nice.org.uk/guidance/cg137.
17. Schiller Y. Seizure Relapse and Development of Drug Resistance Following Long-term Seizure Remission. Archives of Neurology. 2009;66(10).
18. Wolf P. Acute Drug Administration in Epilepsy: A Review. 2011;17(5):442-8.
19. Kwan P, Brodie MJ. Effectiveness of first antiepileptic drug. Epilepsia. 2001;42(10):1255-60.
20. Chen Z, Brodie MJ, Liew D, Kwan P. Treatment Outcomes in Patients With Newly Diagnosed Epilepsy Treated With Established and New Antiepileptic Drugs: A 30-Year Longitudinal Cohort Study. JAMA Neurol. 2018;75(3):279-86.
21. Pycroft L, Stein J, Aziz T. Deep brain stimulation: An overview of history, methods, and future developments. Brain Neurosci Adv. 2018;2:2398212818816017.

22. Lozano AM, Lipsman N, Bergman H, Brown P, Chabardes S, Chang JW, et al. Deep brain stimulation: current challenges and future directions. Nat Rev Neurol. 2019;15(3):148-60.
23. (NICE) NioCE. Epilepsies: diagnosis and management; Clinical guideline [CG137] 2020 [Available from: https://www.nice.org.uk/guidance/cg137.
24. Rogovik AL, Goldman RD. Ketogenic diet for treatment of epilepsy. Can Fam Physician. 2010;56(6):540-2.
25. Williams TJ, Cervenka MC. The role for ketogenic diets in epilepsy and status epilepticus in adults. Clin Neurophysiol Pract. 2017;2:154-60.
26. Samsonsen C, Sand T, Bråthen G, Helde G, Brodtkorb E. The impact of sleep loss on the facilitation of seizures: A prospective case-crossover study. Epilepsy Res. 2016;127:260-6.
27. Malow BA. Sleep deprivation and epilepsy. Epilepsy Curr. 2004;4(5):193-5.
28. Hamerle M, Ghaeni L, Kowski A, Weissinger F, Holtkamp M. Alcohol Use and Alcohol-Related Seizures in Patients With Epilepsy. Front Neurol. 2018;9:401.
29. Wolfe CE, Wood DM, Dines A, Whatley BP, Yates C, Heyerdahl F, et al. Seizures as a complication of recreational drug use: Analysis of the Euro-DEN Plus data-set. Neurotoxicology. 2019;73:183-7.
30. Brust JC. Neurologic complications of illicit drug abuse. Continuum (Minneap Minn). 2014;20(3 Neurology of Systemic Disease):642-56.
31. Macfarlane A, Greenhalgh T. Sodium valproate in pregnancy: what are the risks and should we use a shared decision-making approach? BMC Pregnancy Childbirth. 2018;18(1):200.
32. Duncan S. Teratogenesis of sodium valproate. Curr Opin Neurol. 2007;20(2):175-80.
33. Guilfoyle SM, Follansbee-Junger K, Smith AW, Combs A, Ollier S, Hater B, et al. Antiepileptic drug behavioral side effects and baseline hyperactivity in children and adolescents with new onset epilepsy. Epilepsia. 2018;59(1):146-54.
34. Chen B, Choi H, Hirsch LJ, Katz A, Legge A, Buchsbaum R, et al. Psychiatric and behavioral side effects of antiepileptic drugs in adults with epilepsy. Epilepsy Behav. 2017;76:24-31.
35. Meador KJ, Loring DW. Developmental effects of antiepileptic drugs and the need for improved regulations. Neurology. 2016;86(3):297-306.
36. Elsharkawy AE, Thorbecke R, Ebner A, May TW. Determinants of quality of life in patients with refractory focal epilepsy who were not eligible for surgery or who rejected surgery. Epilepsy Behav. 2012;24(2):249-55.
37. Marvanova M. Pharmacokinetic characteristics of antiepileptic drugs (AEDs). Ment Health Clin. 2016;6(1):8-20.
38. French JA, Gidal BE. Antiepileptic drug interactions. Epilepsia. 2000;41 Suppl 8:S30-6.
39. Dutta S, Faught E, Limdi NA. Valproate protein binding following rapid intravenous administration of high doses of valproic acid in patients with epilepsy. J Clin Pharm Ther. 2007;32(4):365-71.
40. Sweeney GD. Variability in the human drug response. Thromb Res Suppl. 1983;4:3-15.
41. Dorne JL. Impact of inter-individual differences in drug metabolism and pharmacokinetics on safety evaluation. Fundam Clin Pharmacol. 2004;18(6):609-20.
42. Becerra JL, Ojeda J, Corredera E, Ruiz Giménez J. Review of therapeutic options for adjuvant treatment of focal seizures in epilepsy: focus on lacosamide. CNS Drugs. 2011;25 Suppl 1:3-16.

43. Miziak B, Chrościńska-Krawczyk M, Czuczwar SJ. An update on the problem of osteoporosis in people with epilepsy taking antiepileptic drugs. Expert Opin Drug Saf. 2019;18(8):679-89.
44. Nakken KO, Taubøll E. Bone loss associated with use of antiepileptic drugs. Expert Opin Drug Saf. 2010;9(4):561-71.
45. Ensrud KE, Walczak TS, Blackwell TL, Ensrud ER, Barrett-Connor E, Orwoll ES. Antiepileptic drug use and rates of hip bone loss in older men: a prospective study. Neurology. 2008;71(10):723-30.
46. Asconapé JJ. Use of antiepileptic drugs in hepatic and renal disease. Handb Clin Neurol. 2014;119:417-32.
47. Vidaurre J, Gedela S, Yarosz S. Antiepileptic Drugs and Liver Disease. Pediatr Neurol. 2017;77:23-36.
48. Kalilani L, Sun X, Pelgrims B, Noack-Rink M, Villanueva V. The epidemiology of drug-resistant epilepsy: A systematic review and meta-analysis. Epilepsia. 2018;59(12):2179-93.
49. French JA. Refractory epilepsy: clinical overview. Epilepsia. 2007;48 Suppl 1:3-7.
50. Zaccara G, Franciotta D, Perucca E. Idiosyncratic adverse reactions to antiepileptic drugs. Epilepsia. 2007;48(7):1223-44.
51. Workman P, Draetta GF, Schellens JHM, Bernards R. How Much Longer Will We Put Up With $100,000 Cancer Drugs? Cell. 2017;168(4):579-83.
52. Rogawski MA. Brivaracetam: a rational drug discovery success story. Br J Pharmacol. 2008;154(8):1555-7.
53. Gower AJ, Noyer M, Verloes R, Gobert J, Wülfert E. ucb L059, a novel anti-convulsant drug: pharmacological profile in animals. Eur J Pharmacol. 1992;222(2-3):193-203.
54. Schenone M, Dančík V, Wagner BK, Clemons PA. Target identification and mechanism of action in chemical biology and drug discovery. Nat Chem Biol. 2013;9(4):232-40.
55. Mohs RC, Greig NH. Drug discovery and development: Role of basic biological research. Alzheimers Dement (N Y). 2017;3(4):651-7.
56. Kafkas Ş, Dunham I, McEntyre J. Literature evidence in open targets - a target validation platform. Journal of Biomedical Semantics. 2017;8(1):20.
57. Hur J, Schuyler AD, States DJ, Feldman EL. SciMiner: web-based literature mining tool for target identification and functional enrichment analysis. Bioinformatics (Oxford, England). 2009;25(6):838-40.
58. Brennan RJ. Target Safety Assessment: Strategies and Resources. Methods Mol Biol. 2017;1641:213-28.
59. Owens J. Determining druggability. Nature Reviews Drug Discovery. 2007;6(3):187-.
60. Hopkins AL, Groom CR, editors. Target Analysis: A Priori Assessment of Druggability. Small Molecule — Protein Interactions; 2003 2003//; Berlin, Heidelberg: Springer Berlin Heidelberg.
61. Trosset JY, Cavé C. In Silico Target Druggability Assessment: From Structural to Systemic Approaches. Methods Mol Biol. 2019;1953:63-88.
62. Yang X, Kui L, Tang M, Li D, Wei K, Chen W, et al. High-Throughput Transcriptome Profiling in Drug and Biomarker Discovery. Frontiers in Genetics. 2020;11(19).
63. Floris M, Olla S, Schlessinger D, Cucca F. Genetic-Driven Druggable Target Identification and Validation. Trends Genet. 2018;34(7):558-70.
64. Terstappen GC, Schlüpen C, Raggiaschi R, Gaviraghi G. Target deconvolution strategies in drug discovery. Nature Reviews Drug Discovery. 2007;6(11):891-903.
65. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. Br J Pharmacol. 2011;162(6):1239-49.
66. Frearson JA, Collie IT. HTS and hit finding in academia--from chemical genomics to drug discovery. Drug Discov Today. 2009;14(23-24):1150-8.

67. Valler MJ, Green D. Diversity screening versus focussed screening in drug discovery. Drug Discov Today. 2000;5(7):286-93.
68. Kalyaanamoorthy S, Chen Y-PP. Structure-based drug design to augment hit discovery. Drug Discovery Today. 2011;16(17):831-9.
69. SigmaAldritch. Hit Discovery & Confirmation for Early Drug Discovery 2021 [Available from: https://www.sigmaaldrich.com/GB/en/technical-documents/technical-article/research-and-disease-areas/pharmacology-and-drug-discovery-research/hit-discovery-and-confirmation-for-early-drug-discovery.
70. Keseru GM, Makara GM. Hit discovery and hit-to-lead approaches. Drug Discov Today. 2006;11(15-16):741-8.
71. Reitz AB, Smith GR, Tounge BA, Reynolds CH. Hit triage using efficiency indices after screening of compound libraries in drug discovery. Curr Top Med Chem. 2009;9(18):1718-24.
72. Dahlin JL, Walters MA. The essential roles of chemistry in high-throughput screening triage. Future Med Chem. 2014;6(11):1265-90.
73. Zhao H, Akritopoulou-Zanze I. When analoging is not enough: scaffold discovery in medicinal chemistry. Expert Opin Drug Discov. 2010;5(2):123-34.
74. Tintori C, Selvaraj M, Badia R, Clotet B, Esté JA, Botta M. Computational studies identifying entry inhibitor scaffolds targeting the Phe43 cavity of HIV-1 gp120. ChemMedChem. 2013;8(3):475-83.
75. Lobell M, Hendrix M, Hinzen B, Keldenich J, Meier H, Schmeck C, et al. In silico ADMET traffic lights as a tool for the prioritization of HTS hits. ChemMedChem. 2006;1(11):1229-36.
76. Bergner A, Parel SP. Hit expansion approaches using multiple similarity methods and virtualized query structures. J Chem Inf Model. 2013;53(5):1057-66.
77. Neves BJ, Braga RC, Melo-Filho CC, Moreira-Filho JT, Muratov EN, Andrade CH. QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. Frontiers in Pharmacology. 2018;9(1275).
78. EvinceBiosciences. The Economics of Drug Discovery and Development are Not What You Think 2017 [Available from: https://medium.com/@EvinceBio/the-economics-of-drug-discovery-and-development-are-not-what-you-think-91cd4b1528b.
79. Zhu T, Cao S, Su P-C, Patel R, Shah D, Chokshi HB, et al. Hit Identification and Optimization in Virtual Screening: Practical Recommendations Based on a Critical Literature Analysis. Journal of Medicinal Chemistry. 2013;56(17):6560-72.
80. Baumann A. Preclinical development of therapeutic biologics. Expert Opin Drug Discov. 2008;3(3):289-97.
81. Musuamba FT, Manolis E, Holford N, Cheung S, Friberg LE, Ogungbenro K, et al. Advanced Methods for Dose and Regimen Finding During Drug Development: Summary of the EMA/EFPIA Workshop on Dose Finding (London 4-5 December 2014). CPT Pharmacometrics Syst Pharmacol. 2017;6(7):418-29.
82. Holt JD, Nuttall JP. Preclinical safety evaluation. Curr Top Microbiol Immunol. 2014;383:55-78.
83. Karlsson M, Danielsson BR, Nilsson MF, Danielsson C, Webster WS. New proposals for testing drugs with IKr-blocking activity to determine their teratogenic potential. Curr Pharm Des. 2007;13(29):2979-88.
84. Yang T, Chun YW, Stroud DM, Mosley JD, Knollmann BC, Hong C, et al. Screening for acute IKr block is insufficient to detect torsades de pointes liability: role of late sodium current. Circulation. 2014;130(3):224-34.
85. Foti RS, Wienkers LC, Wahlstrom JL. Application of cytochrome P450 drug interaction screening in drug discovery. Comb Chem High Throughput Screen. 2010;13(2):145-58.
86. Everitt JI. The future of preclinical animal models in pharmaceutical discovery and development: a need to bring in cerebro to the in vivo discussions. Toxicol Pathol. 2015;43(1):70-7.

87. National_Research_Council_(US)_Panel_on_Handling_Missing_Data_in_Clinical_Trials. The Prevention and Treatment of Missing Data in Clinical Trials.2010.
88. Goldenholz DM, Goldenholz SR. Response to placebo in clinical epilepsy trials--Old ideas and new insights. Epilepsy Res. 2016;122:15-25.
89. Takebe T, Imai R, Ono S. The Current Status of Drug Discovery and Development as Originated in United States Academia: The Influence of Industrial and Academic Collaboration on Drug Discovery and Development. Clin Transl Sci. 2018;11(6):597-606.
90. Xue H, Li J, Xie H, Wang Y. Review of Drug Repositioning Approaches and Resources. Int J Biol Sci. 2018;14(10):1232-44.
91. Moshfeghi AA, Rosenfeld PJ, Puliafito CA, Michels S, Marcus EN, Lenchus JD, et al. Systemic bevacizumab (Avastin) therapy for neovascular age-related macular degeneration: twenty-four-week results of an uncontrolled open-label clinical study. Ophthalmology. 2006;113(11):2002.e1-12.
92. Avery RL, Pieramici DJ, Rabena MD, Castellarin AA, Nasir MA, Giust MJ. Intravitreal bevacizumab (Avastin) for neovascular age-related macular degeneration. Ophthalmology. 2006;113(3):363-72.e5.
93. Jyothi S, Chowdhury H, Elagouz M, Sivaprasad S. Intravitreal bevacizumab (Avastin) for age-related macular degeneration: a critical analysis of literature. Eye. 2010;24(5):816-24.
94. Cohen D. CCGs face legal threat for offering off-label drug for wet AMD. Bmj. 2017;359:j5021.
95. NIHR_Southampton_Clinical_Research_Facility. Landmark study could save NHS £100 million
2019 [Available from: https://www.uhs.nhs.uk/ClinicalResearchinSouthampton/Research/Facilities/NIHR-Southampton-Clinical-Research-Facility/Ourroleandimpacts/Impactcasestudies/Landmark-study-could-save-NHS-100-million.aspx.
96. Cohen D. CCGs win right to offer patients Avastin for wet AMD. Bmj. 2018;362:k4035.
97. Bayer plc & Or v NHS Darlington CCG & Ors (2020) EWCA Civ 449.
98. Pantziarka P, Pirmohamed M, Mirza N. New uses for old drugs. Bmj. 2018;361:k2701.
99. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, et al. Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov. 2019;18(1):41-58.
100. Hargrave-Thomas E, Yu B, Reynisson J. Serendipity in anticancer drug discovery. World J Clin Oncol. 2012;3(1):1-6.
101. Parvathaneni V, Kulkarni NS, Muth A, Gupta V. Drug repurposing: a promising tool to accelerate the drug discovery process. Drug Discov Today. 2019;24(10):2076-85.
102. Gogou M, Cross JH. Fenfluramine as antiseizure medication for epilepsy. Developmental Medicine & Child Neurology. 2021.
103. Aldrich MS, Garofalo EA, Drury I. Epileptiform abnormalities during sleep in Rett syndrome. Electroencephalography and Clinical Neurophysiology. 1990;75(5):365-70.
104. Gastaut H, Zifkin B, Rufo M. Compulsive respiratory stereotypies in children with autistic features: polygraphic recording and treatment with fenfluramine. J Autism Dev Disord. 1987;17(3):391-406.
105. Aman MG, Kern RA. Review of Fenfluramine in the Treatment of the Developmental Disabilities. Journal of the American Academy of Child & Adolescent Psychiatry. 1989;28(4):549-65.
106. Gastaut H ZB. Antiepileptic effects of fenfluramine: pilot study. Ann Neurol. 1987.
107. Boel M, Casaer P. Add-on therapy of fenfluramine in intractable self-induced epilepsy. Neuropediatrics. 1996;27(4):171-3.

108. Lagae L, Sullivan J, Knupp K, Laux L, Polster T, Nikanorova M, et al. Fenfluramine hydrochloride for the treatment of seizures in Dravet syndrome: a randomised, double-blind, placebo-controlled trial. Lancet. 2019;394(10216):2243-54.
109. Boyd B, Smith S, Gammaitoni A, Galer BS, Farfel GM. A phase I, randomized, open-label, single-dose, 3-period crossover study to evaluate the drug-drug interaction between ZX008 (fenfluramine HCl oral solution) and a regimen of stiripentol, clobazam, and valproate in healthy subjects . Int J Clin Pharmacol Ther. 2019;57(1):11-9.
110. Youssef Bessada P, Kevin W. Chamberlin, PharmD. Fenfluramine for the Treatment of Seizures Associated With Dravet Syndrome 2020 [Available from: https://www.drugtopics.com/view/fenfluramine-for-the-treatment-of-seizures-associated-with-dravet-syndrome.
111. Keiser MJ, Setola V, Irwin JJ, Laggner C, Abbas AI, Hufeisen SJ, et al. Predicting new molecular targets for known drugs. Nature. 2009;462(7270):175-81.
112. Casamassimi A, Federico A, Rienzo M, Esposito S, Ciccodicola A. Transcriptome Profiling in Human Diseases: New Advances and Perspectives. Int J Mol Sci. 2017;18(8):1652.
113. Stathias V, Jermakowicz AM, Maloof ME, Forlin M, Walters W, Suter RK, et al. Drug and disease signature integration identifies synergistic combinations in glioblastoma. Nature Communications. 2018;9(1):5315.
114. Le BL, Andreoletti G, Oskotsky T, Vallejo-Gracia A, Rosales R, Yu K, et al. Transcriptomics-based drug repositioning pipeline identifies therapeutic candidates for COVID-19. Scientific Reports. 2021;11(1):12310.
115. Kim I-W, Jang H, Kim JH, Kim MG, Kim S, Oh JM. Computational Drug Repositioning for Gastric Cancer using Reversal Gene Expression Profiles. Scientific Reports. 2019;9(1):2660.
116. Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 2017;171(6):1437-52.e17.
117. Wagner A, Cohen N, Kelder T, Amit U, Liebman E, Steinberg DM, et al. Drugs that reverse disease transcriptomic signatures are more effective in a mouse model of dyslipidemia. Mol Syst Biol. 2015;11(3):791-.
118. Wei G, Twomey D, Lamb J, Schlis K, Agarwal J, Stam RW, et al. Gene expression-based chemical genomics identifies rapamycin as a modulator of MCL1 and glucocorticoid resistance. Cancer Cell. 2006;10(4):331-42.
119. Musa A, Ghoraie LS, Zhang S-D, Galzko G, Yli-Harja O, Dehmer M, et al. A review of connectivity map and computational approaches in pharmacogenomics. Briefings in Bioinformatics. 2017:bbw112.
120. Hodos RA, Kidd BA, Shameer K, Readhead BP, Dudley JT. In silico methods for drug repurposing and pharmacology. Wiley Interdiscip Rev Syst Biol Med. 2016;8(3):186-210.
121. Issa NT, Kruger J, Byers SW, Dakshanamurthy S. Drug repurposing a reality: from computers to the clinic. Expert Rev Clin Pharmacol. 2013;6(2):95-7.
122. Wang Z, Liang L, Yin Z, Lin J. Improving chemical similarity ensemble approach in target prediction. J Cheminform. 2016;8:20.
123. Tatar G, Ozyurt E, Turhan K. Computational drug repurposing study of the RNA binding domain of SARS-CoV-2 nucleocapsid protein with antiviral agents. Biotechnol Prog. 2021;37(2):e3110.
124. Sohraby F, Bagheri M, Aryapour H. Performing an In Silico Repurposing of Existing Drugs by Combining Virtual Screening and Molecular Dynamics Simulation. Methods Mol Biol. 2019;1903:23-43.
125. Tam V, Patel N, Turcotte M, Bossé Y, Paré G, Meyre D. Benefits and limitations of genome-wide association studies. Nat Rev Genet. 2019;20(8):467-84.

126. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. Nucleic Acids Res. 2018;46(D1):D1074-d82.
127. Bush WS, Moore JH. Chapter 11: Genome-wide association studies. PLoS Comput Biol. 2012;8(12):e1002822.
128. Guney E, Menche J, Vidal M, Barábasi AL. Network-based in silico drug efficacy screening. Nat Commun. 2016;7:10331.
129. Stolfi P, Manni L, Soligo M, Vergni D, Tieri P. Designing a Network Proximity-Based Drug Repurposing Strategy for COVID-19. Frontiers in Cell and Developmental Biology. 2020;8(1021).
130. Zeng X, Zhu S, Hou Y, Zhang P, Li L, Li J, et al. Network-based prediction of drug-target interactions using an arbitrary-order proximity embedded deep forest. Bioinformatics. 2020;36(9):2805-12.
131. Guney E, Menche J, Vidal M, Barábasi A-L. Network-based in silico drug efficacy screening. Nature Communications. 2016;7(1):10331.
132. Tronvik E, Stovner LJ, Helde G, Sand T, Bovim G. Prophylactic Treatment of Migraine With an Angiotensin II ReceptorBlocker: A Randomized Controlled Trial. JAMA. 2003;289(1):65-9.
133. Renaud RC, Xuereb H. Erectile-dysfunction therapies. Nat Rev Drug Discov. 2002;1(9):663-4.
134. Ozery-Flato M, Goldschmidt Y, Shaham O, Ravid S, Yanover C. Framework for identifying drug repurposing candidates from observational healthcare data. JAMIA Open. 2020;3(4):536-44.
135. Ye H, Liu Q, Wei J. Construction of Drug Network Based on Side Effects and Its Application for Drug Repositioning. PLoS One. 2014;9(2):e87864.
136. Yang L, Agarwal P. Systematic Drug Repositioning Based on Clinical Side-Effects. PLoS One. 2011;6(12):e28025.
137. Schwarze K, Buchanan J, Fermont JM, Dreau H, Tilley MW, Taylor JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet Med. 2020;22(1):85-94.
138. Bomba L, Walter K, Soranzo N. The impact of rare and low-frequency genetic variants in common disease. Genome Biology. 2017;18(1):77.
139. Karki R, Pandya D, Elston RC, Ferlini C. Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics. 2015;8:37-.
140. Hemminki K, Försti A, Bermejo JL. The 'common disease-common variant' hypothesis and familial risks. PLoS One. 2008;3(6):e2504-e.
141. Buniello A, MacArthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005-d12.
142. Ma L, Han S, Yang J, Da Y. Multi-locus test conditional on confirmed effects leads to increased power in genome-wide association studies. PLoS One. 2010;5(11):e15006.
143. Brueggeman L, Sturgeon ML, Martin RM, Grossbach AJ, Nagahama Y, Zhang A, et al. Drug repositioning in epilepsy reveals novel antiseizure candidates. Ann Clin Transl Neurol. 2018;6(2):295-309.
144. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, et al. ClinVar: improving access to variant interpretations and supporting evidence. Nucleic Acids Res. 2018;46(D1):D1062-d7.
145. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Res. 2005;33(Database issue):D514-7.

146. Stenson PD, Ball EV, Mort M, Phillips AD, Shaw K, Cooper DN. The Human Gene Mutation Database (HGMD) and its exploitation in the fields of personalized genomics and molecular evolution. Curr Protoc Bioinformatics. 2012;Chapter 1:Unit1.13.
147. Law M, Shaw DR. Mouse Genome Informatics (MGI) Is the International Resource for Information on the Laboratory Mouse. Methods Mol Biol. 2018;1757:141-61.
148. Wang J, Lin ZJ, Liu L, Xu HQ, Shi YW, Yi YH, et al. Epilepsy-associated genes. Seizure. 2017;44:11-20.
149. Muñoz-Fuentes V, Cacheiro P, Meehan TF, Aguilar-Pimentel JA, Brown SDM, Flenniken AM, et al. The International Mouse Phenotyping Consortium (IMPC): a functional catalogue of the mammalian genome that informs conservation. Conserv Genet. 2018;19(4):995-1005.
150. Tan NC, Berkovic SF. The Epilepsy Genetic Association Database (epiGAD): analysis of 165 genetic association studies, 1996-2008. Epilepsia. 2010;51(4):686-9.
151. Thurmond J, Goodman JL, Strelets VB, Attrill H, Gramates LS, Marygold SJ, et al. FlyBase 2.0: the next generation. Nucleic Acids Res. 2019;47(D1):D759-d65.
152. Abou-Khalil B, Auce P, Avbersek A, Bahlo M, Balding DJ, Bast T, et al. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nature Communications. 2018;9(1):5269.
153. International League Against Epilepsy Consortium on Complex E. Genome-wide mega-analysis identifies 16 loci and highlights diverse biological mechanisms in the common epilepsies. Nat Commun. 2018;9(1):5269.
154. Consortium E, Consortium EM, Steffens M, Leu C, Ruppert AK, Zara F, et al. Genome-wide association analysis of genetic generalized epilepsies implicates susceptibility loci at 1q43, 2p16.1, 2q22.3 and 17q21.32. Hum Mol Genet. 2012;21(24):5359-72.
155. International League Against Epilepsy Consortium on Complex Epilepsies. Electronic address e-auea. Genetic determinants of common epilepsies: a meta-analysis of genome-wide association studies. Lancet Neurol. 2014;13(9):893-903.
156. Kasperaviciute D, Catarino CB, Matarin M, Leu C, Novy J, Tostevin A, et al. Epilepsy, hippocampal sclerosis and febrile seizures linked by common genetic variation around SCN1A. Brain. 2013;136(Pt 10):3140-50.
157. Sanseau P, Agarwal P, Barnes MR, Pastinen T, Richards JB, Cardon LR, et al. Use of genome-wide association studies for drug repositioning. Nat Biotechnol. 2012;30(4):317-20.
158. Aguirre-Plans J, Piñero J, Sanz F, Furlong LI, Fernandez-Fuentes N, Oliva B, et al. GUILDify v2.0: A Tool to Identify Molecular Networks Underlying Human Diseases, Their Comorbidities and Their Druggable Targets. J Mol Biol. 2019;431(13):2477-84.
159. Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, et al. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res. 2018;46(D1):D1068-d73.
160. Ursu O, Holmes J, Knockel J, Bologa CG, Yang JJ, Mathias SL, et al. DrugCentral: online drug compendium. Nucleic Acids Res. 2017;45(D1):D932-d9.
161. Gaulton A, Hersey A, Nowotka M, Bento AP, Chambers J, Mendez D, et al. The ChEMBL database in 2017. Nucleic Acids Res. 2017;45(D1):D945-d54.
162. Guney E, Oliva B. Exploiting Protein-Protein Interaction Networks for Genome-Wide Disease-Gene Prioritization. PLoS One. 2012;7(9):e43557.
163. Fromm GH, Amores CY, Thies W. Imipramine in Epilepsy. Archives of Neurology. 1972;27(3):198-204.
164. Fromm GH, Wessel HB, Glass JD, Alvin JD, Van Horn G. Imipramine in absence and myoclonic-astatic seizures. Neurology. 1978;28(9 Pt 1):953-7.
165. Hurst DL. The use of imipramine in minor motor seizures. Pediatric Neurology. 1986;2(1):13-7.

166. Turner AL, Perry MS. Outside the box: Medications worth considering when traditional antiepileptic drugs have failed. Seizure. 2017;50:173-85.
167. Yimer EM, Surur A, Wondafrash DZ, Gebre AK. The Effect of Metformin in Experimentally Induced Animal Models of Epileptic Seizure. Behavioural Neurology. 2019;2019:6234758.
168. Sanz P, Serratosa JM, Sánchez MP. Beneficial Effects of Metformin on the Central Nervous System, with a Focus on Epilepsy and Lafora Disease. Int J Mol Sci. 2021;22(10).
169. Doble A. The pharmacology and mechanism of action of riluzole. Neurology. 1996;47(6 Suppl 4):S233-41.
170. Kim JE, Kim DS, Kwak SE, Choi HC, Song HK, Choi SY, et al. Anti-glutamatergic effect of riluzole: comparison with valproic acid. Neuroscience. 2007;147(1):136-45.
171. Borowicz KK, Sekowski A, Drelewska E, Czuczwar SJ. Riluzole enhances the anti-seizure action of conventional antiepileptic drugs against pentetrazole-induced convulsions in mice. Pol J Pharmacol. 2004;56(2):187-93.
172. Cheng F, Desai RJ, Handy DE, Wang R, Schneeweiss S, Barabási A-L, et al. Network-based approach to prediction and population-based validation of in silico drug repurposing. Nature Communications. 2018;9(1):2691.
173. Himelstein AL, Foster JC, Khatcheressian JL, Roberts JD, Seisler DK, Novotny PJ, et al. Effect of Longer-Interval vs Standard Dosing of Zoledronic Acid on Skeletal Events in Patients With Bone Metastases: A Randomized Clinical Trial. Jama. 2017;317(1):48-58.
174. Liu L, Zheng T, Morris MJ, Wallengren C, Clarke AL, Reid CA, et al. The mechanism of carbamazepine aggravation of absence seizures. J Pharmacol Exp Ther. 2006;319(2):790-8.
175. French JA. Seizure exacerbation by antiepileptic drugs. Epilepsy currents. 2005;5(5):192-3.
176. Suresh SH, Chakraborty A, Virupakshaiah A, Kumar N. Efficacy and Safety of Levetiracetam and Carbamazepine as Monotherapy in Partial Seizures. Epilepsy research and treatment. 2015;2015:415082-.
177. Goldenberg MM. Overview of drugs used for epilepsy and seizures: etiology, diagnosis, and treatment. P T. 2010;35(7):392-415.
178. Sivapalarajah S, Krishnakumar M, Bickerstaffe H, Chan Y, Clarkson J, Hampden-Martin A, et al. The prescribable drugs with efficacy in experimental epilepsies (PDE3) database for drug repurposing research in epilepsy. Epilepsia. 2018;59(2):492-501.
179. Carvill GL, Weckhuysen S, McMahon JM, Hartmann C, Møller RS, Hjalgrim H, et al. GABRA1 and STXBP1: novel genetic causes of Dravet syndrome. Neurology. 2014;82(14):1245-53.
180. (NORD) NOfRD. Dravet Syndrome
. 2020.
181. Shi X-Y, Tomonoh Y, Wang W-Z, Ishii A, Higurashi N, Kurahashi H, et al. Efficacy of antiepileptic drugs for the treatment of Dravet syndrome with different genotypes. Brain and Development. 2016;38(1):40-6.
182. Koleti A, Terryn R, Stathias V, Chung C, Cooper DJ, Turner JP, et al. Data Portal for the Library of Integrated Network-based Cellular Signatures (LINCS) program: integrated access to diverse large-scale cellular perturbation response data. Nucleic Acids Res. 2018;46(D1):D558-d66.
183. Mirza N, Sills GJ, Pirmohamed M, Marson AG. Identifying new antiepileptic drugs through genomics-based drug repurposing. Human Molecular Genetics. 2017;26(3):527-37.
184. Hawkins NA, Calhoun JD, Huffman AM, Kearney JA. Gene expression profiling in a mouse model of Dravet syndrome. Exp Neurol. 2019;311:247-56.
185. Wirrell EC, Laux L, Donner E, Jette N, Knupp K, Meskis MA, et al. Optimizing the Diagnosis and Management of Dravet Syndrome: Recommendations From a North American Consensus Panel. Pediatr Neurol. 2017;68:18-34.e3.

186. Dinday MT, Baraban SC. Large-Scale Phenotype-Based Antiepileptic Drug Screening in a Zebrafish Model of Dravet Syndrome. eneuro. 2015;2(4):ENEURO.0068-15.
187. Baraban SC. Baraban lab, zebrafish. 2020.
188. Byron SA, Van Keuren-Jensen KR, Engelthaler DM, Carpten JD, Craig DW. Translating RNA sequencing into clinical diagnostics: opportunities and challenges. Nat Rev Genet. 2016;17(5):257-71.
189. Jain R, Ramaswamy S, Harilal D, Uddin M, Loney T, Nowotny N, et al. Host transcriptomic profiling of COVID-19 patients with mild, moderate, and severe clinical outcomes. Comput Struct Biotechnol J. 2021;19:153-60.
190. Liu Y, Lopez-Santiago LF, Yuan Y, Jones JM, Zhang H, O'Malley HA, et al. Dravet syndrome patient-derived neurons suggest a novel epilepsy mechanism. Ann Neurol. 2013;74(1):128-39.
191. Jiao J, Yang Y, Shi Y, Chen J, Gao R, Fan Y, et al. Modeling Dravet syndrome using induced pluripotent stem cells (iPSCs) and directly converted neurons. Hum Mol Genet. 2013;22(21):4241-52.
192. Higurashi N, Uchida T, Lossin C, Misumi Y, Okada Y, Akamatsu W, et al. A human Dravet syndrome model from patient induced pluripotent stem cells. Mol Brain. 2013;6:19.
193. Ngugi AK, Bottomley C, Kleinschmidt I, Sander JW, Newton CR. Estimation of the burden of active and life-time epilepsy: a meta-analytic approach. Epilepsia. 2010;51(5):883-90.
194. Kowski AB, Weissinger F, Gaus V, Fidzinski P, Losch F, Holtkamp M. Specific adverse effects of antiepileptic drugs--A true-to-life monotherapy study. Epilepsy Behav. 2016;54:150-7.
195. Shorvon SD. The epidemiology and treatment of chronic and refractory epilepsy. Epilepsia. 1996;37 Suppl 2:S1-s3.
196. Erakovic Haber V, Spaventi R. Discovery and Development of Novel Drugs. Springer International Publishing; 2017. p. 91-104.

Appendices

Appendix 1: CACHEP search terms

A 1.1: Search for genes/proteins that cause seizures/epilepsy in animal models

(

((TITLE-ABS (((spAEDs) OR (absences) OR clonic OR myoclon* OR convuls* OR epilep* OR kindl* OR seizure* OR eeg OR electroencephalogra* OR polyspikes OR "Spike-and-wave" OR "Spike-wave" OR "sharp-wave" OR afterdischarge* OR "fast ripples" OR "delta power" OR "gamma power" OR "theta power" OR “early myoclonic encephalopathy” OR “glut1 deficiency” OR “landau kleffner” OR “lennox gastaut” OR “pyridoxamine 5'-phosphate oxidase deficiency” OR “unverricht lundborg” OR “west syndrome” OR aicardi OR angelman OR dravet OR ohtahara OR panayiotopoulos OR rAEDussen OR rett OR “ring chromosome 20”) AND (rat OR rats OR rattus OR mouse OR mice OR murine OR zebrafish OR danio OR rerio OR (fly) OR (flies) OR drosophila OR melanogaster OR (worm) OR (worms) OR caenorhabditis OR elegans) AND (allel* OR mutation* OR mutant* OR mutate* OR variation* OR variant* OR polymorphi* OR (gene* W/50 ("insertion" OR delet* OR duplicat* OR "substitution" OR "transloation" OR "inversion" OR "amplification")) OR "trinucleotide repeats" OR knock* OR transgen* OR crispr OR homozyg* OR heterozyg* OR (+/+) OR (-/-) OR (+/-) OR (-/+) OR rnai OR "RNA interference" OR (gene* W/50 silenc*) OR (gene W/50 ablat*) OR cas9 OR "Targeted Augmentation of Nuclear Gene Output" OR tango OR "antisense oligonucleotide" OR (gene* W/50 edit*) OR "Cre recombinase" OR haploinsufficien* OR optogenetic* OR photoinhibit* OR crispra OR overexpress* OR (gene* W/50 modulat*) OR "molecular switch" OR (gene* W/50 manipulat*) OR (gene* W/50 disrupt*) OR (vir* W/50 vector*) OR (gene* W/50 rescue) OR (gene* W/50 modifier) OR (KO) OR (cKO) OR (gene W/50 target*) OR (modifier) OR (dominant-negative) OR (dominant negative) OR (null)))))

)

AND LANGUAGE(english) AND (DOCTYPE(ar) OR DOCTYPE(cp) OR DOCTYPE(le)) AND SRCTYPE(j)

A 1.2: Search for human epilepsy genes

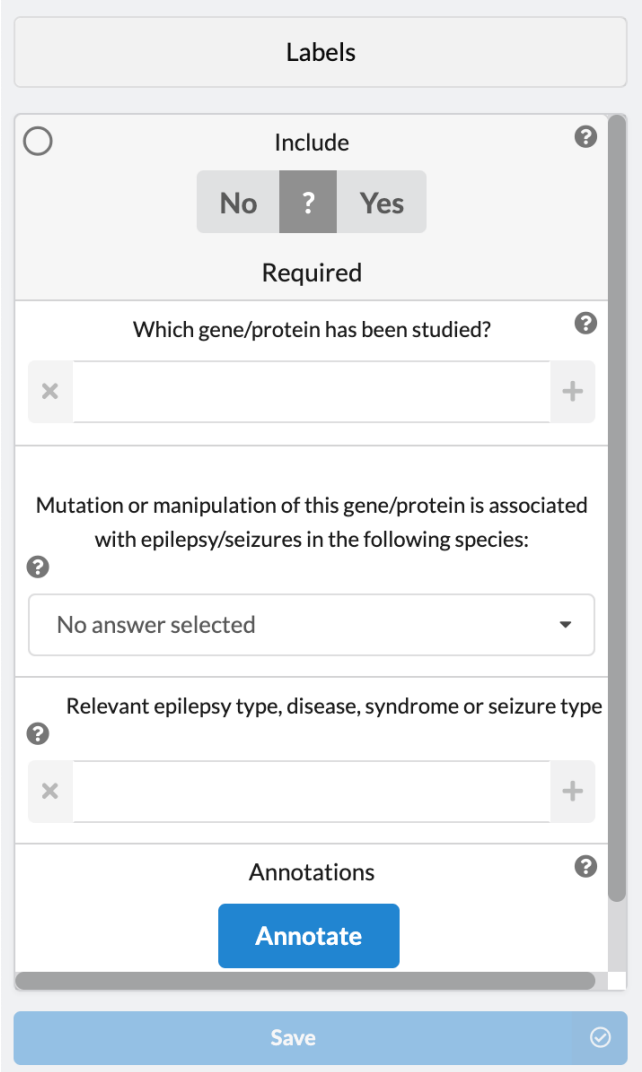
(

(((TITLE-ABS (((spAEDs) OR (absences) OR clonic OR (tonic) OR myoclon* OR convuls* OR epilep* OR seizure* OR “early myoclonic encephalopathy” OR “glut1 deficiency” OR “landau kleffner” OR “lennox gastaut” OR “pyridoxamine 5'-phosphate oxidase deficiency” OR “unverricht lundborg” OR “west syndrome” OR aicardi OR angelman OR dravet OR ohtahara OR panayiotopoulos OR rAEDussen OR rett OR “ring chromosome 20”) AND (allel* OR mutation* OR mutant* OR mutate* OR variation* OR variant* OR polymorphi* OR (gene* W/50 ("insertion" OR delet* OR duplicat* OR "substitution" OR "transloation" OR "inversion" OR "amplification")) OR "trinucleotide repeats" OR homozyg* OR heterozyg* OR (+/+) OR (-/-) OR (+/-) OR (-/+)) AND (human OR patient OR family OR families OR (generations) OR proband OR child* OR boy OR girl OR man OR woman OR men OR women OR male OR female OR people OR person OR individual))) AND (PUBYEAR > 2015)))

)

AND LANGUAGE(english) AND (DOCTYPE(ar) OR DOCTYPE(cp) OR DOCTYPE(le)) AND SRCTYPE(j)

Appendix 2: Entry proforma for part 1 of the CACHEP initiative



Appendix 3: GUILDIFY analysis selected genes

A 3.1: Genes used for Generalised epilepsy predictions

AP3D1
CAMTA1
SCN1A
SETD1A
STX1B
RAPGEF2
UBTF
GABRA2
RIMS1
PCDH7
PHACTR1
RBFOX1
TTC21B
MMP27
DOC2A
GRM4
MYH14
PFKM
CUX1
MACF1
ARID1B
PRKAB1
DEAF1
ITPR1
LBR
CTC1
KCNK2
HTT
ZNHIT3
PNPO
TNRC6A
KAT8
TMEM163
ACVRL1
KCNK4
SCYL1
FAM102A
DPM2
ALPL
NCK1
RORB
MYRF
KMT2D
PIGP
HSF2
CRADD
ARHGEF15
KCNJ4
VAMP2
PDE2A
DENND5A
MICAL1
NPAS4
SMARCA4
COL6A2
WWOX
HECW2
ADORA2B
PSEN2
CACNA2D2
BCL11A
SPEN
PIGL
STAG1
EXOC6B
RP1L1
CEP152
FIG4
TAT
CNOT1
SEMA3F
IFNG
FUS
PIGW
UBP1
KCNJ6
FBP1
SUOX
IQSEC1
ETV6
HNRNPH2
PANK2
MYO3A
ALG10B
STARD9
ATAD2B
TYRO3
IPO9
RYR2
PRKACA
SPTBN2
MPI
CENPW
MAGI2
NFASC
ADCY1
BCKDHA
TPPP
KDM2B
ALG2
PRF1
PEPD
CLPP
RAI1
STARD7
SCN9A
SCYL2
PC
PARP10
GNB5
AIMP2
KCNJ3
NDUFS8
YIF1B
DHX30
UBE4A
AUH
IL13
CACNB4
C11orf80
IFIH1
KCNMB4
SRCAP
MPDU1
HTR7
CNTN5
SBDS
SIN3A
AKAP6
ECE1
CFH
TNKS2
S100B
RRP1B
PEX3
NUS1
DLD
MAPK1
NOD2
ACOT4
DNASE1
ATP7B
AR
CTSF

A 3.2: Genes used for Focal epilepsy predictions

SCN1A
SCN9A
TTC21B
SERPINE2
ASPM
RD3
KCNMA1
LIAS
UPB1
FGF7
GCDH
EFHC1
TGFB1
BPTF
ADORA2A
PPARA
COQ3
TSFM
CR2
CYP27B1
HLF
IQSEC1
SPATA5
VPS13A
IL6
CAMTA1
FARSA
RBFOX3
ST8SIA2
GABRG2
COASY
FGFR1
MRS2
RET
SMAD4
TRMT61A
CADPS2
DDHD2
CRTAC1
FGF14
NEUROD1
ARC
EN2
TMED7
ZBTB38
HPGDS
CKB
POLA1
DGKE
JAM
GABBR1
MTHFD1
MACF1
CAMK2B
CYFIP1
SLC20A2
USP15
TRIM32
NLGN4X
CDH15
ATP8A2
LAMA1
UBE4A
CHD8
PTGR1
TOR1A
CLCN4
FGF2
KCNH1
THRB
ZBTB20
TRAPPC9
SPECC1L
GRM7
KCNJ3
HCK
SV2B
CELSR1
ABHD12
AKAP9
FGF1
GALR1
KCNN3
MICU1
PRMT7
RAP1B
RLIM
SH3GL2
SLC27A4
TLK2
VWA3B
YWHAE
CLN8
IER3IP1
LARS2
MAGI2
RAPGEF2
SENP2
CDKN2D
DNA2
PCDH19
POLR3B
PACS1
DNMT3A
SNX14
DNASE1
NDUFS3
ACADSB
AMPD2
BICD2
CNTNAP2
SGK1
NFAT5
GNAQ
B4GALNT1
SUMF1
AHI1
BBS4
FAT4
IDH3A
MED27
RAB11B
SCNN1D
TBX2
MFN1
RHEB
DBP
PTF1A
NEB
ADORA1
ACY3
ELP3
SLC44A1
GTPBP3
AP4E1
AGT
CHAT
IFIH1
STT3B
EPO
DDT
DDTL
ACTB
ADAM22