BRAIN
ORIGINAL ARTICLE

Towards realizing the vision of precision medicine: AI based prediction of clinical drug response

Johann de Jong,[1] Ioana Cutcutache,[2] Matthew Page,[2] Sami Elmoufti,[3] Cynthia Dilley,[4] Holger Fröhlich[1,5,6,†,‡] and Martin Armstrong[7,†]

[†]These authors contributed equally to this work.

See Chen *et al.* (doi:10.1093/brain/awab199) for a scientific commentary on this article.

Accurate and individualized prediction of response to therapies is central to precision medicine. However, because of the generally complex and multifaceted nature of clinical drug response, realizing this vision is highly challenging, requiring integrating different data types from the same individual into one prediction model.

We used the anti-epileptic drug brivaracetam as a case study and combine a hybrid data/knowledge-driven feature extraction with machine learning to systematically integrate clinical and genetic data from a clinical discovery dataset (*n* = 235 patients).

We constructed a model that successfully predicts clinical drug response [area under the curve (AUC) = 0.76] and show that even with limited sample size, integrating high-dimensional genetics data with clinical data can inform drug response prediction. After further validation on data collected from an independently conducted clinical study (AUC = 0.75), we extensively explore our model to gain insights into the determinants of drug response, and identify various clinical and genetic characteristics predisposing to poor response. Finally, we assess the potential impact of our model on clinical trial design and demonstrate that, by enriching for probable responders, significant reductions in clinical study sizes may be achieved.

To our knowledge, our model represents the first retrospectively validated machine learning model linking drug mechanism of action and the genetic, clinical and demographic background in epilepsy patients to clinical drug response. Hence, it provides a blueprint for how machine learning-based multimodal data integration can act as a driver in achieving the goals of precision medicine in fields such as neurology.

1 Data and Translational Sciences, UCB Biosciences GmbH, 40789 Monheim am Rhein, Germany
2 Data and Translational Sciences, UCB Pharma, Slough SL1 3WE, UK
3 Late Development Statistics, UCB Biosciences Inc., Raleigh, NC 27617, USA
4 Head of Asset Strategy, UCB Inc., Smyrna, GA 30080, USA
5 Fraunhofer Institute for Scientific Computing and Algorithms (SCAI), Business Area Bioinformatics, 53757 Sankt Augustin, Germany
6 Bonn-Aachen International Center for IT, University of Bonn, 53115 Bonn, Germany
7 Data and Translational Sciences, UCB Pharma, 1420 Braine l'Alleud, Belgium

[‡]The author performed most of this work while previously working at UCB Biosciences GmbH, Monheim, Germany

Correspondence to: Johann de Jong
UCB Biosciences GmbH, Alfred-Nobel-Straße 10, 40789 Monheim am Rhein, Germany
E-mail: johann.dejong@ucb.com

Correspondence may also be addressed to: Holger Fröhlich
University of Bonn, Bonn-Aachen International Center for Information Technology (B-IT)
Friedrich Hirzebruch-Allee 5, 53115 Bonn, Germany
E-mail: holger.froehlich@scai.fraunhofer.de

Martin Armstrong
UCB Pharma, Chemin du Foriest 1, 1420 Braine l'Alleud, Belgium
E-mail: martin.armstrong@ucb.com

Introduction

In recent years, precision medicine has emerged as a new paradigm for improved and more individualized patient care. Its key objective is to provide the right treatment, to the right patient at the right time, by basing medical decisions on individual patient characteristics (including biomarkers), rather than on averages of those characteristics across a whole patient population. Precision medicine has the potential to widely impact both patient care and drug development, via[1,2] (i) early disease diagnosis and prevention using individual patient characteristics; (ii) better disease management by more efficient and effective prescribing practices at the level of the individual patient, including reducing the risk of side effects and adverse events; and (iii) more efficient design of clinical trials by enriching for likely responders at baseline.

Precision medicine is closely connected to and partially dependent on pharmacogenetics, a field that has developed since the 1980s resulting from advances in genetics and molecular biology. In pharmacogenetics, statistical associations between genetic variants and drug response are studied within a defined target population. According to the US Food and Drug Administration (FDA) more than 400 pharmacogenetic biomarkers and biomarker signatures have appeared on ~160 drug labels,[3] almost half of which are within the field of oncology. However, in many cases, and specifically outside oncology, these tests have had a limited impact on prescribing practices. One of the reasons for this is that the nature of drug response is highly complex, resulting from the interaction of a multitude of factors including environmental, anthropometric and genetic factors, as well as biological subsystems affected by the disease.[4] In this respect, it is unlikely that any single biomarker or other single stratifying factor will fully capture this complexity, and consequently, the field of precision medicine is still in its infancy.[5]

Recent advances in the ability to generate molecular data, as well as parallel advances in the fields of artificial intelligence, specifically machine learning (ML), and high-performance computing, now allow for integrating different types of multivariate data from the same individual into one multimodal prediction model.[2] Such ML models can be used to make predictions at the individual patient level, and their interpretation can drive the development of a more holistic understanding of drug response. ML models can thus be instrumental in translating precision medicine into practice, and in optimizing clinical trial design by predictive enrichment for drug responders and the testing of therapies in selected populations.[6,7] However, before ML models can be moved into clinical routine, a rigorous validation pipeline is required that consists of (i) validation on an independent test set; (ii) further validation on an independent study (owing to the fact that all clinical studies are unavoidably biased by patient selection); (iii) a prospective validation via a clinical study to show a benefit compared to standard of care; and (iv) following a regulatory pathway to get approval as a diagnostic tool.[2] Because of this complexity, currently there are still only a few examples where ML models have reached clinical routine, and all of them are from the oncology field.[8,9] On the other hand, neurology is widely seen as a field in which there is a high potential for precision medicine, as well as high demand due to the increasing prevalence of neurological disorders.[10,11]

Epilepsy is one such neurological disorder that has a pressing need for the development of more individualized patient care paths in the context of precision medicine. The treatment of epilepsy and the development of new anti-epileptic drugs (AEDs) are hampered by high rates of patient non-response.[12] Despite almost 30 AEDs being licensed for use in the USA,[13] over 30% of patients remain inadequately controlled on their medication.[14–16] The reasons for this remain largely unknown, and iterative prescribing practices remain the norm.

Patients with drug-resistant epilepsy suffer more frequently from comorbid illnesses, psychological dysfunction, social stigmatization and reduced quality of life. Moreover, they are at increased risk of mortality and have an overall decreased life expectancy. As such, these patients represent most of the disease burden in the epilepsy patient population.[17] Additionally, in relation to the need to drive sustainable health care systems, patients with refractory/uncontrolled epilepsy contribute disproportionately to the cost of treatment. Studies of epilepsy subpopulations consistently found substantially higher per-person costs for patients with poor seizure control, ranging from ~2 to ~10 times higher compared to patients with unrefractory/controlled epilepsy.[18] With a high patient and socio-economic burden, there is therefore a pressing need to understand and predict the AED response better, to support the development of new AEDs and ensure that patients are optimally treated earlier in their disease course.

While previous work has attempted to improve our understanding of the AED response,[19–31] the focus has mostly been on specific pharmacokinetics related drug metabolizing enzymes and transporters, neglecting other genetic factors that could be important in determining response, such as a drug's mechanism of action or underlying disease mechanisms. However, as outlined above, focusing on single stratifying factors or biomarkers has severe limitations in its potential for achieving the goals of precision medicine. As such, it has so far been difficult for the—sometimes even conflicting—results to impact clinical practice.[32] We are aware of only one study that jointly analysed clinical and genetic data to predict AED resistance in patients with mesial temporal lobe

epilepsy.[33] While showing promise, the analysis was, again, limited to pharmacokinetic-related genes. Moreover, the authors did not validate their model in an independent patient dataset, so the extent to which their results generalize outside their discovery dataset is unclear. Finally, the study addressed the relatively general question of resistance against any AED. However, to achieve the objectives of precision medicine and support optimized prescription in clinical practice, specific drug response prediction models are needed, directly linking drug mechanism of action to patient-specific response.

In further related work, no attempts were made to link drug mechanism of action to patient response,[34–38] response was not predicted directly,[36,38] only limited numbers of single nucleotide polymorphisms (SNPs) were used,[35] no genetics data were used at all,[34–38] and/or the results were not validated on external data.[34,37,38]

For this purpose, we set out to use ML approaches to integrate a broader set of genetic and clinical data, to predict response to one specific AED. Specifically, we used the AED brivaracetam as a case study. Brivaracetam (trade name: Briviact®) is an AED whose mechanism of action is via blockade of the vesicle membrane protein SV2A and is licensed for the treatment of partial-onset (focal) seizures in patients 4 years of age and older. During phase III clinical studies, ~30–40% of subjects reached the primary end point of >50% reduction in seizure frequency.[39] By integrating clinical and whole-genome sequencing (WGS) data from 235 participants in these phase III clinical studies, we built and compared several ML models predictive of the brivaracetam response. Our results show how patient and clinical data, as well as genetic makeup (comprising epilepsy-specific and drug mechanism-specific genetic features) each uniquely contribute to predicting drug response. In addition, we demonstrate that our best ML model achieves a good prediction performance in a retrospective validation using an additional independent phase III clinical study. Finally, we quantify the impact that our best model could have on clinical trial design by its ability to enrich for responders in future clinical studies. We thus realize an important step towards precision medicine in neurological disorders.

Materials and methods

Processing the clinical data

Efficacy data were pooled from two phase III clinical trials [N01252 (NCT00490035), N01358 (NCT01261325)] for a therapeutic dose of >50 mg/day, enrolling adults (= 16 years) with focal seizures. Both initial clinical trials received approval by appropriate institutional review committees and were conducted in accordance with all regulatory, ethical, and good clinical practice requirements. After an 8-week prospective baseline period, patients were randomized to brivaracetam or placebo without titration and entered a 12-week treatment period. Patients randomized to brivaracetam >50 mg/day (100 and 200 mg/day) or placebo were included in the analysis. Although trial N01252 also included patients treated with 50 mg/day, these were excluded from the analysis to maximize correspondence between the validation dataset (N01252) and the training dataset (N01358), which did not include any patients treated with 50 mg/day. Data were used for subjects who consented to give a DNA sample and were treated with either 100 mg or 200 mg of brivaracetam per day, the approved brivaracetam doses.

Sample collection, WGS and processing

Total genomic DNA >850 ng (>10 ng/µl minimum concentration) was extracted from blood of 235 patients (NCT01261325) and 47 patients (NCT00490035) treated with 100 mg or 200 mg of brivaracetam per day. DNA samples were sequenced to 30× mean coverage on Illumina HiSeq X instruments with 150 bp, paired-end sequencing. Reads were aligned to the GRCh38/hg38 reference using BWA-mem 7.15.r1140.[40,41] Picard[42] was used to mark duplicates and GATK4[43] for base quality score recalibration, which was applied using a 4-bin base quality scheme. Picard was used to aggregate all data for each sample into individual CRAM files using lossless compression. SNPs and indels were called, per sample, using HaplotypeCaller (GATK 3.5)[43] and individual gVCFs consolidated into a single gVCF. Pre-computed CADD scores[44] were used for single nucleotide variants (SNVs) and CADD was run separately to generate scores for indels. The scores were converted to GRCh38 coordinates using CrossMap.[45] The gVCF was normalized and split to generate per-sample VCFs using bcftools.[46] We used snpEff[47] to annotate each sample VCF with CADD scores. MANTA v1.3.1[48] was used on all samples for CNV/SV calling.

Constructing the gene sets

Gene sets (Supplementary Table 1) were identified by enrichment of Gene Ontology Molecular Function terms; MetaBase canonical pathways; or phenotype categories, with genes that have an aetiological association to epilepsy or that are putatively linked to the mechanism of action of brivaracetam. Epilepsy genes were combined from different molecular evidence sources: rare genetic epilepsies or rare disorders that display phenotype traits ontologically descendant from 'seizure' (excluding symptomatic seizures and focal sensory seizures) and 'epileptic encephalopathy' in the Human Phenotype Ontology[49]; nearest mapped genes for significant SNPs from epilepsy genome-wide association studies (GWAS) in GWAS Catalog[50]; direct genetic associations from gene-disease databases DisGeNET[51] and OpenTargets[52]; experimentally validated target genes for miRNAs dysregulated in epilepsy from miRmap[53] and EpimiRBase[54]; and epigenetic mechanisms linked to epilepsy. In turn, drug mechanism of action genes were drawn from: canonical pathways related to synaptic vesicle biology; direct neighbours of SV2A, SV2B and SV2C in a protein-protein interaction network contextualized with brain gene expression data from GTEx[55]; genes encoding drug-metabolizing enzymes for brivaracetam; genes associated with comorbidities that influence drug response; and the targets of co-administered AEDs (carbamazepine, phenytoin, phenobarbital, levetiracetam). Epilepsy genes and drug mechanism of action genes were linked to gene sets from Gene Ontology Molecular Function terms using (i) TopGO[56]; (ii) Metabase[57] neurological canonical signalling pathways using a hypergeometric test; and (iii) knowledge-based epilepsy phenotype categories[58] based on gene membership.

Variant filtering and gene mapping

After preprocessing as outlined above, the WGS data represented ~20 million variants. From these, only variants that satisfied at least one of the following conditions were retained in subsequent gene set-based analyses: (i) a significant association to the disease phenotype in GWAS, as listed in the GWAS Catalog[50]; (ii) strong linkage disequilibrium (LD) to any of the above GWAS SNPs (r^2 > 0.8); (iii) proximity of at most 10 kb to SV2A and a CADD[44] score > 15; and (iv) expression quantitative trait loci (eQTL) evidence for SV2A and a CADD[44] score > 15.

This resulted in ~14 000 unique SNVs. These filtered variants were mapped to genes based on proximity in the genome (in ±10 kb region), brain *cis*-eQTLs from GTEx and significant SNPs from epilepsy related GWAS studies (Supplementary Table 2). Hg38 genomic intervals for genes were extracted from the SnpEff database[47]

using the 'dump' argument. Genomic coordinates for each gene were extended ±10 kb and converted to BED format. Significant brain *cis*-eQTLs were downloaded from https://gtexportal.org/home/datasets (v7; accessed 1 October 2020) and CrossMap[45] applied to convert to coordinates on the GRCh38 assembly before converting to BED format. GWAS Catalog[50] v1.0.1 was used to select all the SNPs significantly associated with specific epilepsy/seizure traits. For each such SNP the position, risk allele and mapped gene information were extracted from the catalogue. A BED file containing these positions and their corresponding LD blocks was prepared. Each BED formatted annotation file was used to annotate sample VCF files and a custom Python script used to extract a simplified annotation of each variant, which collapses together multiple gene mapping sources keeping: chromosome, position, reference allele, alternate allele, genotype, CADD scores, Entrez ids of mapped genes, and snpEff annotation. Gene mapping based on GWAS and eQTL information was done only if the expected risk allele was observed in the sample.

Gene set-wise mutational burden scores

Gene set-wise mutational burden scores were calculated by counting the number of unique mutations in the genes corresponding to each gene set.

Polygenic risk score

We retrieved log2 odds ratios for all SNPs with an epilepsy phenotype from the PheWAS Catalog,[59] and for each subject computed the dot product of the SNP log2 odds ratios with the patient-specific SNP genotypes. More specifically, let N be the total number of epilepsy related SNPs, i be a patient indicator, j be a SNP indicator, x_j be the log2 odds ratio for SNP j, g_{ij} be the genotype of patient i for SNP j. Then the polygenic risk score y_i for patient i is computed as follows:

$$y_i = \sum_{j=1}^{N} x_j g_{ij} \qquad (1)$$

Model training

All models were trained using nested cross-validation, with 10 inner folds for hyperparameter optimization, and 10 outer folds for performance estimation. To assess the dependence of performance estimation on the chosen folds, 20 repeats were done using different folds. For all models except the multimodal neural network, hyperparameter settings were selected using a grid search. Because of the large number of hyperparameters to be optimized for the multimodal neural network, as well as the computationally expensive model fitting procedure, hyperparameter settings were selected using Bayesian optimization instead.[60] Moreover, for the multimodal neural network, each hyperparameter optimization in the inner loop of the nested cross-validation was repeated three times for stability purposes. The models were implemented using the following software packages:

(i) Sparse multi-block PLS-DA: mixOmics.[61]
(ii) Multimodal neural network: Tensorflow 2.[62] A multimodal architecture was set up by limiting connections in the first layer(s) of the neural network to connections only between features from the same data modality. Given the limited number of samples, this model is highly complex, and needs to be strongly regularized to avoid overfitting. For this purpose, we implemented (a) a group lasso penalty[63] on the input layers for the different data modalities; (b) overall l1[64] and l2[65] penalties on the subsequent layers; (c) dropout[66] and batch normalization.[67] The neural networks were trained using an Adam optimizer,[68] with beta1 = 0.9 and beta2 = 0.999. The learning rate, together with the regularization penalties, dropout rate, batch size and architecture, was optimized using Bayesian optimization.[69]

(iii) Elastic net: glmnet.[70]
(iv) Gradient-boosted decision trees classifier: xgboost.[71]
(v) Stacked gradient-boosted decision trees classifier: xgboost.[71] Individual xgboost classifiers were trained for each data modality. Predictions of these individual classifiers were then stacked[72] using logistic linear regression.

Power calculations

For a given classifier threshold h, the positive predictive value for drug-treated subjects (ppv_{trt}) was determined using the classifier trained on both clinical and genetics data, as described above and in the main text. Since no WGS data were available for placebo-treated patients, we constructed an additional model using clinical data alone, and applied this model (with the same threshold h) to the placebo-treated patients to determine the positive predictive value ppv_{pcb}. These two proportions were interpreted as responder rates in a hypothetical confirmatory trial, and power calculations were performed for a one-sided two-sample test for proportions (using the R stats package; at significance level 0.01 and power 90%), to determine the minimum sample size for which ppv_{trt} is significantly larger than ppv_{pcb}. Note that because WGS data are not available for placebo-treated patients, ppv_{pcb} is necessarily based on clinical data only, whereas ppv_{trt} is based on clinical and genetics data, and hence the calculations are dependent on the assumption that the inclusion of epilepsy- and brivaracetam-specific genetics variants do not inflate placebo response too strongly.

Learning curve

We retrained our classifier, as described above, but with increasing sample sizes in the inner cross-validation loop (ranging from $n = 45$ to $n = 211$), while recording the area under the ROC curve (AUC) on a left-out set of fixed size ($n = 24$) in the outer cross-validation loop. This was repeated 20 times.

Data availability

Clinical studies (NCT01261325, NCT00490035) can be accessed via ClinicalTrials.gov.

Results

Defining discovery and validation datasets

To develop a predictive model for drug response, we used two datasets: (i) a discovery dataset; and (ii) a retrospective validation dataset.[2] We constructed a discovery dataset from phase III clinical trial data (NCT01261325) on drug response in 235 adult epilepsy patients, who consented to provide a DNA sample and were treated with either 100 mg or 200 mg of brivaracetam per day, the approved brivaracetam doses (Table 1).[73] The overall response rate in this dataset was 37%, with an average age of 40.7 years, 48% female, and focus localization mostly confirmed to be frontal or temporal. This discovery dataset was used for constructing, optimizing and internally validating the model. Additionally, we constructed a validation dataset by collecting data on 47 adult epilepsy patients from an independently conducted clinical trial (NCT00490035). These 47 patients consented to provide a DNA sample and were treated with 100 mg of brivaracetam per day (Table 1).[74] The overall response rate in this dataset was 36%, with an average age of 41.5 years, 43%

female, and focus localization mostly confirmed to be frontal or temporal. We used this dataset to determine the extent to which the internally validated model generalized to unseen patient data, with potentially different patient characteristics.

In addition to drug response (defined as >50% seizure frequency reduction 12 weeks after study baseline), available data modalities in both datasets included a comprehensive range of demographic and clinical characteristics, as well as WGS data for each patient. More details about the demographic and clinical data available for each patient can be found in [75,76(p15),77] and Supplementary Fig. 1 and Supplementary Table 3). The data were preprocessed as described in the 'Materials and methods' section.

Taking a knowledge-based approach to feature extraction from WGS data

WGS data contains potentially millions of variants. This is orders of magnitude larger than the typical number of patients enrolled in a clinical study, and therefore poses a high risk of overfitting, i.e. bad generalization performance of the ML model.[78] The sparsity, particularly of rare variants, further complicates this problem. Hence, we took a knowledge-based approach by extracting 40 disease-mechanism and drug mechanism of action associated features from the original ~20 million variants, to be used as input into our ML model (Fig. 1A). More specifically, we first defined literature-derived sets of genes that related to (i) epilepsy disease mechanisms; and (ii) brivaracetam's mechanism of action (Fig. 1B and 'Materials and methods' section). With respect to (i), we only considered SNVs that showed a significant association to the disease phenotype in GWAS, as listed in the GWAS Catalog,[50] or showed strong linkage disequilibrium (LD) to any of those SNVs ($r^2 > 0.8$). With respect to (ii), we only considered SNVs proximal to SV2A (±10 kb) and/or with eQTL evidence for SV2A, as documented by the Genotype-Tissue Expression (GTEx) project.[55] These SNVs were then mapped to genes by proximity of at most 10 kb upstream or downstream of transcription start site, or by significant influence on gene expression in brain tissues, as documented by GTEx.

To arrive at a single mutational load score per gene set, we counted the number of SNVs mapped to the genes in each gene set. This resulted in 38 features to be used as input into our ML models. In addition, we defined a polygenic epilepsy risk score using the PheWAS catalogue[59] (see 'Materials and methods' section), and a feature indicating the presence of at least one structural variant overlapping brivaracetam's receptor gene SV2A (±10 kb). This resulted in a total of 40 features derived from the genetics data, which by the approach taken above, we treated as three distinct genetic data modalities: (i) the gene set-wise mutational load

scores; (ii) the polygenic risk score; and (iii) the SV2A structural variance feature. Together with the clinical data, this resulted in four data modalities. For more details, see the 'Materials and methods' section.

Machine learning-based data integration successfully predicts anti-epileptic drug response

We applied and evaluated several ML approaches to systematically integrate our four data modalities for predicting response to brivaracetam:

(i)   Sparse multi-block partial least squares discriminant analysis (PLS-DA),[61,79] with blocks defined by the data modalities.
(ii) A multimodal neural network. Here, each data modality is represented by its own subnetwork at the input of the neural network, with initially no connections allowed between the subnetworks. Only at deeper network levels, the different subnetworks are connected to jointly compute a response probability. For more details regarding the architecture and hyperparameters, see the 'Materials and methods' section.
(iii) An elastic net classifier,[70,80] trained on all data modalities jointly.
(iv) A gradient-boosted decision trees classifier,[71] trained on all data modalities jointly.
(v)   Logistic regression-based stacking[72] of gradient-boosted trees classifiers.

These ML approaches spanned the whole range of data integration strategies,[81–83] from early (e.g. sparse multi-block PLS-DA) to intermediate (the multimodal neural network) to late (the stacked classifier). Our best model, a single gradient-boosted trees classifier trained jointly on all data modalities, significantly outperformed all other models and achieved an AUC of 0.76 ($P = 3.8 \times 10^{-6}$; Fig. 2A and B). Interestingly, the two linear methods (elastic net and sparse multi-block PLS-DA) showed substantially poorer performance than the two tree-based classifiers. This suggests that our problem of brivaracetam response prediction is, to an important extent, non-linear in nature. The multimodal neural network also performed relatively poorly. This is likely due to the large number of parameters specified by such a network, relative to the limited number of samples available for training the network.

Our model generalizes to independent clinical trial data

Each clinical study has unavoidable biases due to the selection and sampling process of patients from the overall population. To make sure that the internally cross-validated performance of our model

Table 1 Discovery, validation and placebo datasets

	Discovery dataset	Validation dataset	Placebo dataset
Number of patients	235	47	235
Brivaracetam dose, mg/day	100–200	100	0
Clinical trial	NCT01261325	NCT00490035	NCT01261325
Whole genome sequencing?	Yes	Yes	No
Responder rate	37%	36%	22%
Age, mean (SD)	40.7 (13.1)	41.5 (13.8)	39.6 (12.9)
Gender, % female	48	43	50
Focus localization			
Frontal	32%	17%	38%
Occipital	5%	6%	6%
Parietal	9%	4%	9%
Temporal	64%	60%	59%
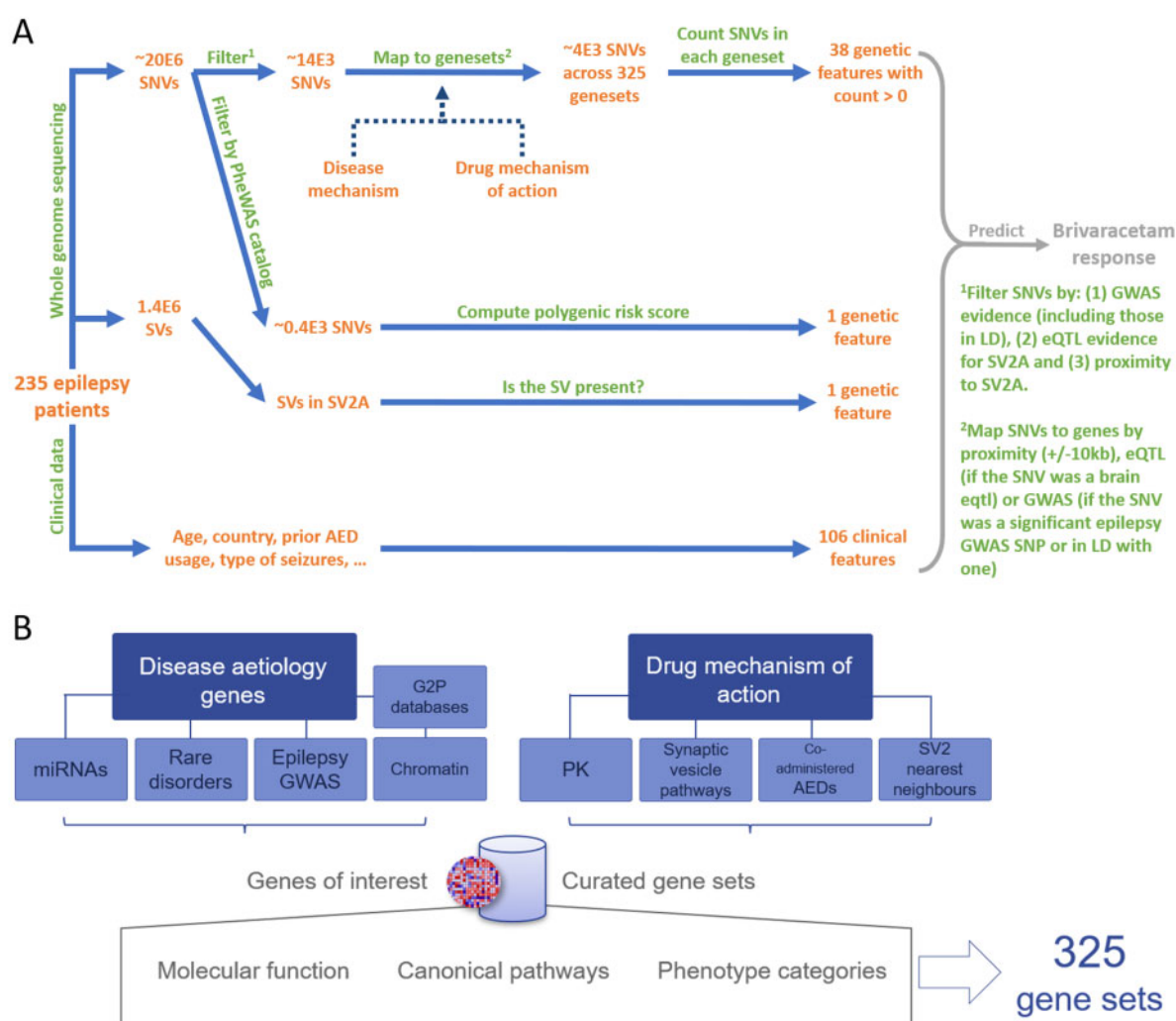
Figure 1 Analysis pipeline and gene set definition. (**A**) Combining hybrid data/knowledge-driven feature extraction with advanced ML to systematically integrate clinical and genetic data for predicting brivaracetam response. (**B**) Defining literature-derived gene sets that relate to (1) epilepsy disease aetiology and (2) brivaracetam's mechanism of action.

was not driven by certain characteristics unique to the discovery dataset, we collected data from an additional independently conducted clinical trial (NCT00490035, $n = 47$; Table 1), and used these data to externally validate our model. Despite significant overall differences in clinical and patient characteristics between the two datasets (Supplementary Fig. 1), our classifier generalized well: It achieved an AUC of 0.75 (Fig. 2C), very close to the cross-validated AUC of 0.76 on the discovery dataset, although due to the limited size of the validation dataset, the asymptotic 95% confidence interval (CI) was wide (0.6–0.9).

Both clinical and genetic features uniquely impact drug response

A better understanding of the mechanisms driving drug (non-)response is instrumental in precision medicine. Hence, we next analysed the predictive strength of our model from the level of entire data modalities down to the level of individual clinical and genetic patient features. To this end, we trained gradient-boosted trees classifiers on each of the four individual data modalities alone and compared their performance with our best model. Our best model significantly outperformed all models trained on individual data modalities ($P = 1.9 \times 10^{-6}$; Fig. 3A), demonstrating the unique

contribution of both clinical and genetic factors to predicting drug response.

To investigate in more detail which individual patient features were most important in determining drug response, we computed feature-wise absolute SHAP values (SHapley Additive exPlanations)[84] (Fig. 3B), averaged across samples. SHAP value estimation decomposes each prediction into a sum of feature importance scores and a bias term, where each feature importance score represents the change in the expected prediction when conditioning on that feature.[84] Consistent with the results in Fig. 3A, the main determinants of drug response prediction represented a mix of clinical and genetic patient features. Strongly predictive clinical features included prior use of levetiracetam and epileptic focus localization. In the model, prior use of levetiracetam, an AED having the same primary receptor as brivaracetam, negatively impacted the probability of response to brivaracetam (Fig. 3C). Likewise, extra-temporal epileptic focus localization negatively impacted the probability of response (Fig. 3C). These two observations were also reflected in the patient population, where non-responders more often had used levetiracetam prior to study entry (Fig. 3D). This was consistent with a previous study suggesting that brivaracetam non-responders are more likely to suffer from drug-refractory epilepsy, and thus have generally used
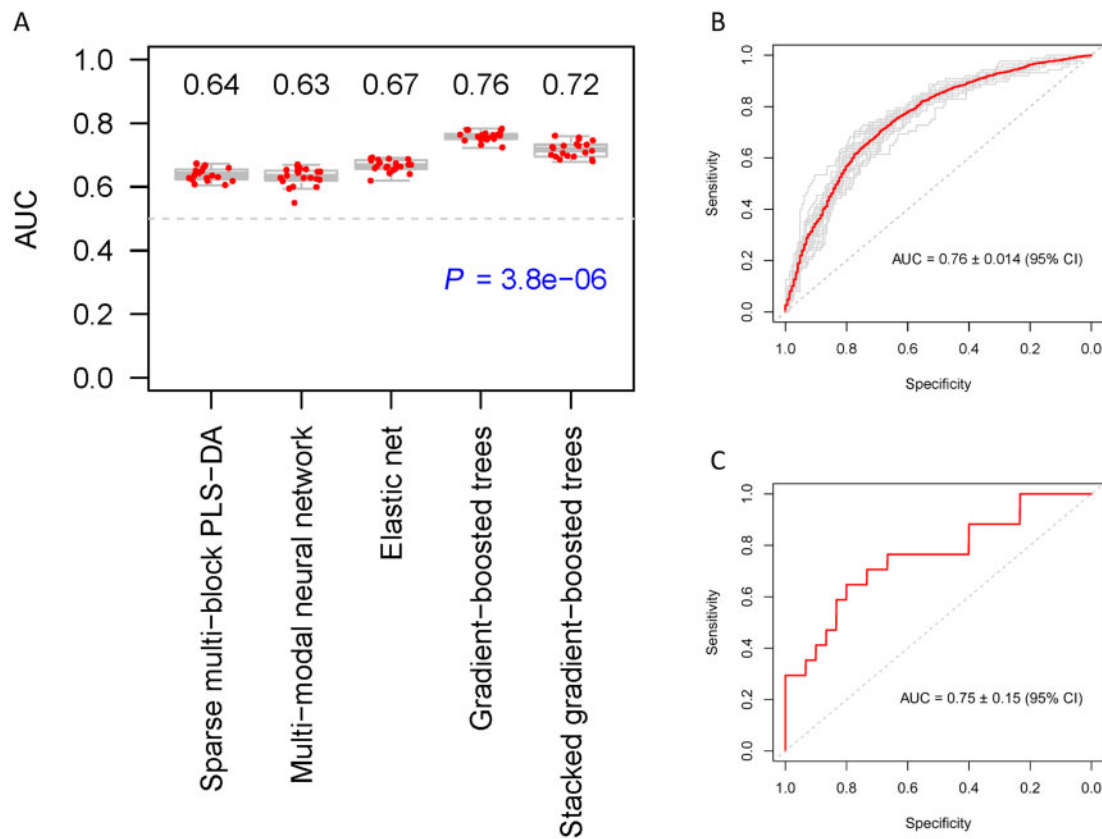
Figure 2 Model performance and external validation. (**A**) Performance of several ML approaches to systematically integrate the genetic and clinical data modalities for predicting response to brivaracetam, as estimated using repeated cross-validation with 20 repeats (each dot represents a repeat). The performance is shown in terms of area under ROC (AUC). An AUC of 0.5 corresponds to chance level, and an AUC of 1 is the best achievable performance. Statistical significance was assessed by a paired Wilcoxon-test between the best and the second-best model. (**B**) ROC of our best model, a gradient-boosted decision trees classifier trained jointly on all data modalities. In light grey, the individual ROCs for the 20 repeats, in red the average ROC across the 20 repeats and the associated empirical 95% CI based on the 20 repeats. The ROC depicts the trade-off between specificity and sensitivity of a classifier, while varying the diagnostic cut-off for the probability of response: Choosing to treat only patients with high predicted probability of response results in a high specificity (most treated patients will benefit from the treatment), but a low sensitivity (many patients that could potentially have benefited from treatment are not treated). Conversely, treating even patients with low predicted response probability results in a high sensitivity (most patients who could potentially benefit are treated) but low specificity (many treated patients do not benefit from the treatment). The AUC is a measure of overall prediction performance. An AUC of 0.5 corresponds to chance level, and an AUC of 1 is the best achievable performance. (**C**) ROC of validating of our best model on the independent validation dataset.

relatively more AEDs prior to study entry.[85] It should be noted that the brivaracetam responder rate in the levetiracetam-naïve group is roughly twice as high as the responder rate in the levetiracetam-exposed group (Supplementary Fig. 2A). This association is highly significant, even when correcting for prior exposure to any of the other AEDs in our data, or when correcting for confounders not restricted to AEDs (Supplementary Fig. 2B). This indicates that prior levetiracetam use is uniquely informative in predicting brivaracetam response, relative to the other AEDs and all other variables in the model, which is likely due to levetiracetam having the same primary receptor gene as brivaracetam.

Strongly predictive genetic features included the presence of structural variants overlapping brivaracetam's receptor gene SV2A, and the mutational load in the GO:0051011 gene set, a GO term representing microtubule minus end binding (Fig. 3C). Note that epilepsy can be caused by disorders of neurodevelopmental processes, and microtubule associated proteins are known to play a role in the regulation of such processes.[86,87] In the model, having more mutations in the GO:0051011 gene set positively impacted the probability of response to brivaracetam (Fig. 3D). On the other hand, having structural variants overlapping brivaracetam's

receptor gene SV2A (±10 kb) negatively impacted the probability of response (Fig. 3D). These two observations were also reflected in the patient population. Patients with a structural variant overlapping SV2A (±10 kb) were more likely not to respond to brivaracetam (Fig. 3D). Moreover, responders had more mutations in the GO:0051011 gene set (Fig. 3D).

SHAP dependence plots and brivaracetam response associations for all features can be found in Supplementary Figs 3 and 4.

Placebo response is not related to model determinants

Some of the features identified in Fig. 3B could potentially also be associated with placebo response. Therefore, we wanted to assess to what extent our model was also partly capturing placebo response. As WGS data were not available for placebo-treated patients, we focused on the model trained on only clinical data from brivaracetam-treated patients to address this question. We applied this model to 235 additional, placebo treated, patients collected from the same clinical study as our discovery dataset (Table 1, clinical study NCT01261325).[73] Placebo response could not be predicted significantly better than random (AUC = 0.56 ± 0.088),
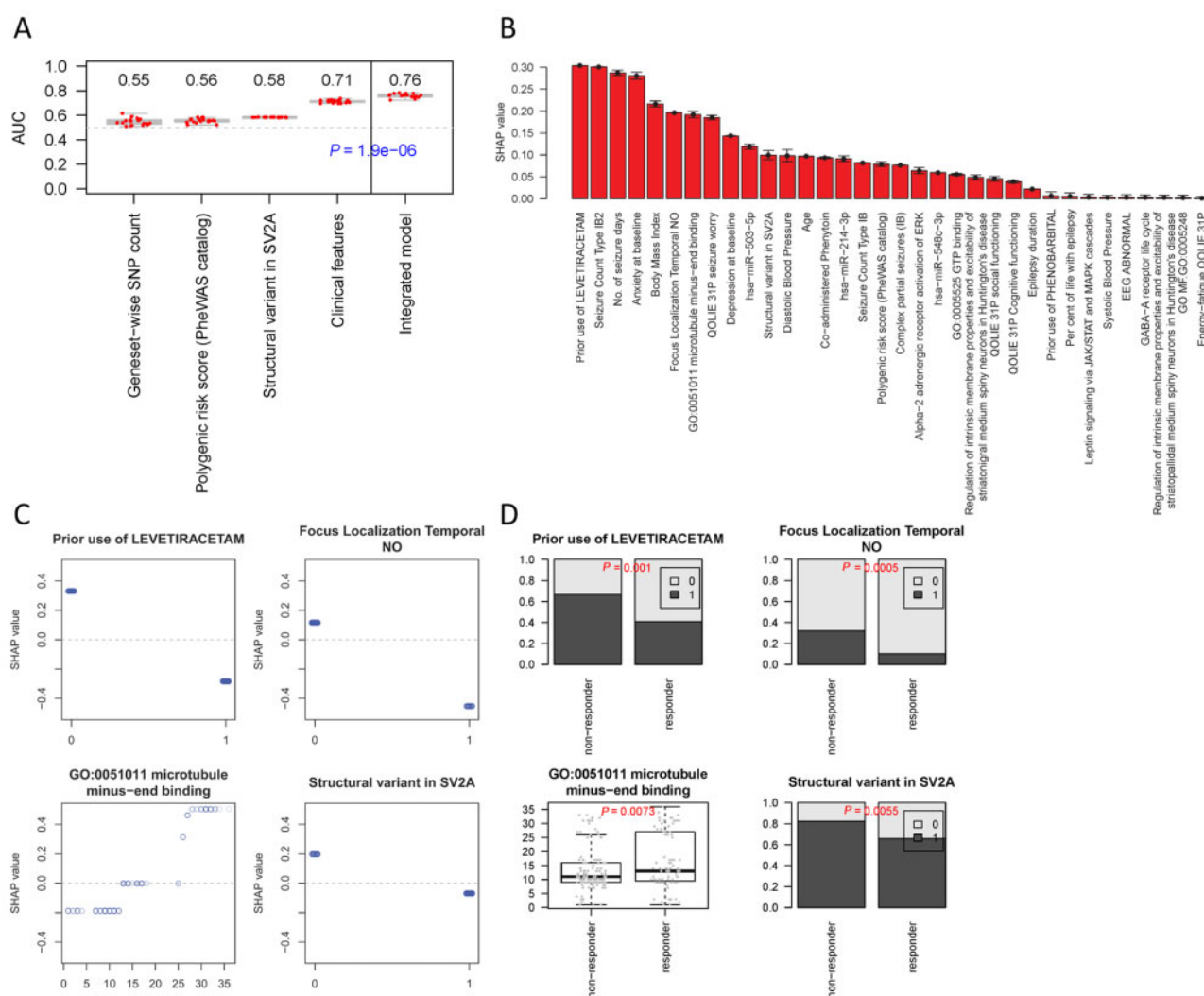
Figure 3 Determinants of drug response probability. (**A**) Performance of gradient-boosted trees classifiers on the individual data modalities, compared with the integrated model. Statistical significance was assessed by a paired Wilcoxon-test between the clinical-only model and the integrated model. (**B**) All patient features with non-zero average absolute SHAP values. Error bars represent the empirical 95% CIs based on 20 cross-validation repeats. (**C**) SHAP dependence plots for four selected variables. *Top* to *bottom*, *left* to *right*: Prior use of levetiracetam, extra-temporal focus localization, mutational load in gene set GO:0051011 and structural variants overlapping *SV2A*. Each circle represents a single patient (note that circles are often superimposed). (**D**) Univariate associations of selected variables with brivaracetam response. *Top* to *bottom*, *left* to *right*: Prior use of levetiracetam, extra-temporal focus localization, mutational load in gene set GO:0051011 and structural variants overlapping *SV2A*. Statistical significance was assessed by a Wilcoxon test (for GO:0051011) or a Fisher's exact test (for the other three features).

indicating that our models indeed did not capture placebo response to any significant degree (Supplementary Fig. 5A). In addition, we trained a new classifier on all clinical data available for the placebo-treated patients, for predicting placebo response directly. Supporting the analysis above, the model did not perform better than random (AUC = 0.5 ± 0.019; Supplementary Fig. 5C).

Finally, to determine whether placebo response was related to any of the individual main clinical determinants of the integrated model (Fig. 2A and B), we univariately tested the association of placebo response with the main clinical determinants of that model (Fig. 3B). No single significant feature could be identified at a false discovery rate of 10% (Supplementary Fig. 6).

Reducing sample size of confirmatory studies by enriching for drug responders

Clinical trials in epilepsy, but also in other disease areas, generally suffer from high non-responder rates.[12,88] A higher non-responder

rate means that more patients will need to be enrolled to establish a statistically significant difference in responder rates between drug- and placebo-treated subjects. However, models trained to predict drug response on data from an initial exploratory study, could be used to enrich for responders in subsequent confirmatory studies (so-called enrichment trial according to FDA terminology; https://www.fda.gov/media/121320/download; accessed 1 October 1 2020). As such, minimum required sample sizes for these confirmatory studies could be reduced.

More specifically, note that the output of our ML model is a probability for each individual patient to respond positively to brivaracetam after 12 weeks. In screening patients for a confirmatory study, a specific threshold could be applied to this probability to decide which patients would be accepted and which would not. Choosing to accept only patients with a very high probability of response results in a high specificity (many accepted patients are responders), but low sensitivity (many responders are not accepted). Conversely, accepting patients even at low predicted

response probabilities results in a higher sensitivity (many responders are accepted) but lower specificity (many accepted patients are non-responders). This trade-off between sensitivity and specificity is visualized by the receiver operating characteristic (ROC) in Fig. 2B.

As a concrete example, setting the classifier threshold at 0 would result in a specificity of 0% and a sensitivity of 100% (Fig. 2B), which would imply accepting all patients at screening, and lead to the original responder rate in our discovery dataset (38%; Fig. 4A). On the other hand, setting the classifier threshold at 0.45 would result in a specificity of 66% and a sensitivity of 74% (Fig. 2B), which could result in a 1.5 times higher responder rate (∼57%; Fig. 4A). This higher responder rate could subsequently reduce the minimum sample size required for a confirmatory study (at 90% statistical power and 0.01 significance threshold) from ∼230 subjects in each study treatment arm, to ∼50 subjects (classifier thresholds at 0 and 0.45, respectively; Fig. 4B), a reduction of ∼65%. This reduction would come at the expense of stricter confirmatory study inclusion and exclusion criteria, as determined by the classifier threshold: In the current example, setting the classifier threshold at 0.45 would imply that ∼50% of patients would not meet the criteria for study entry (Fig. 4A). A classifier with even better performance than the current model should be able to increase this acceptance rate, by showing a relative decrease in false negative calls at the same responder rate.

Note that the response rates for drug-treated subjects were estimated using both clinical and genetics data, whereas the response rates for placebo-treated subjects were necessarily estimated using clinical data alone, as no WGS data were available for placebo-treated subjects (for more details, see the 'Materials and methods' section).

Note also that, even when estimating both placebo and treatment response rates using only clinical data, substantial reductions in required sample size can still be observed (Supplementary Fig. 7).

We then wanted to determine whether, given the current modelling approach, better models could likely be constructed. As outlined above, such models would be able to achieve higher study acceptance rates at similar responder rates, or higher responder rates at similar acceptance rates. For this purpose, we retrained our classifier with increasing sample sizes in the inner cross-validation loop (ranging from $n = 45$ to $n = 211$), while recording the AUC on a left-out set of fixed size ($n = 24$) in the outer cross-validation loop. The roughly linearly increasing AUC showed no signs of reaching a plateau for larger n, which indeed suggested that with more patients, even better performing models can be built (Fig. 4C).
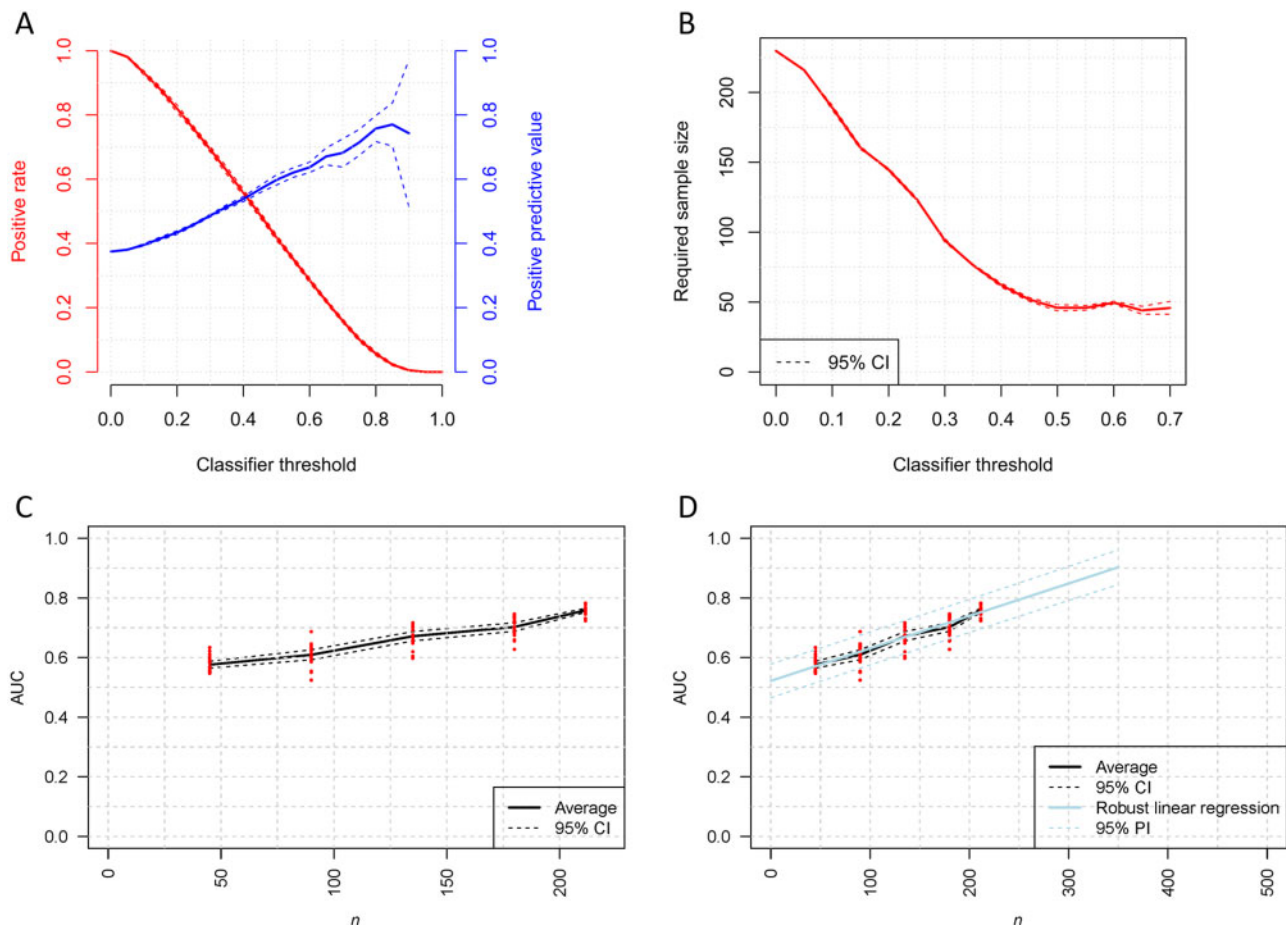
Figure 4 Application to clinical trial design. (**A**) The trade-off between positive rate (number of patients included in a trial) and positive predictive value (fraction of responders in a trial), as a function of the classifier threshold. 95% CIs were determined using the 20 cross-validation repeats. (**B**) Using the classifier to enrich for responders in confirmatory studies: Minimum required sample size in a confirmatory trial (at 90% statistical power and 0.05 significance level) as a function of classifier threshold. CIs were determined by computing sample sizes for the CIs in **A**. (**C**) Training the model on increasing sample sizes in the inner cross-validation loop, while recording the performance on a set of left-out samples of fixed size ($n = 24$) in the outer cross-validation loop. Each dot represents the performance of a model trained on a different randomly subsampled number of patients (n). To determine the 20 cross-validation repeats, 95% CIs were used. (**D**) Extrapolation of performance for larger n, using robust linear regression in blue, with a 95% prediction interval.

To get an indication of how many patients would be needed to achieve even better performance, we fitted a robust linear regression model to the data in Fig. 4C. Extrapolating the linear trend using this regression model suggested that it may be reasonable to expect that at least 350 patients would be needed to achieve a performance of AUC ∼0.9 (Fig. 4D).

Discussion

In this work, we presented a retrospectively validated ML model for predicting clinical drug response to the AED brivaracetam. We demonstrated that integrating WGS data with clinical data significantly improves predictive performance compared to a model based on clinical data or genetic data alone, i.e. that both clinical and genetic patient features uniquely contribute to model performance. Importantly, we also showed that it is unlikely that our model captured placebo response to any significant degree. This analysis was restricted to clinical determinants since, because of the cost considerations of generating WGS data and our focus on the determinants of clinical drug (non-)response, only genomes of brivaracetam-treated patients were sequenced.

In principle, WGS allows for a hypothesis-free way of analysing the influence of genomic variation on an outcome of interest. However, the dimensionality of WGS data is huge relative to the small number of patients available in our study, posing serious risks of overfitting. Hence, it is clear that prior knowledge needs to be used to maximally reduce the WGS feature space in the most informative way possible. For this reason, we initially analysed all variants in the WGS data at the level of sets of genes known to be related to epilepsy and/or brivaracetam's mechanism of action. Genes were organized into functional categories based on prior biological knowledge in the form of biological pathways, molecular functions, and human phenotypes. Such an approach allows us to simultaneously reduce the feature space based on those genes and linked genetic variants most likely to exert an effect on drug response, accommodate sparse rare variants, and generate features with a functional interpretation. In the end, gene set-wise mutational load scores were computed by simply counting the number of mutations mapped to the genes in a gene set. This can be thought of as a linear model that maps input variants, each with a coefficient equal to 1, to gene set-wise scores. One could think of more involved approaches, for example to estimate the coefficients of such a (possibly regularized) linear model using the class labels, as an integral part of the nested-cross validation. However, in our case this led to substantial overfitting.

Our results confirm that the mutational burden in our epilepsy- and mechanism of action-related gene sets are important factors in explaining and predicting response. Two of the most informative genetic features in the model were the presence of structural variants overlapping the *SV2A* gene, and the mutational load in a gene set representing microtubule minus end binding. Brivaracetam is a high affinity *SV2A* ligand whose purported mechanism of action involves blocking the *SV2A* membrane glycoprotein and inhibiting neurotransmitter release.[89] We showed that brivaracetam non-responders were significantly more likely to carry *SV2A* structural variants than responders (P = 0.0055; Fig. 3D). However, previous published work failed to find any impact of single *SV2A* variants on response to levetiracetam, an AED in the same drug class as brivaracetam, but with a less specific *SV2A* binding profile.[90] This again highlights the necessity for summarizing genetic variation into polygenic risk scores, gene-level and even gene set-level scores, to effectively address the challenges that are posed by the analysis of individual variants, which are often rare events.

It is worth noting that, except for prior use of levetiracetam, the clinical data alone contain no information that is specific to the action of, or response to brivaracetam. In this respect, a model built on clinical data alone would likely have some utility in predicting response to other drugs, i.e. would mostly have prognostic value. Including genetic features that are specific to the mechanism of action and disposition of brivaracetam means that the model is likely to have diagnostic value, i.e. be specific to predicting response to brivaracetam. This is supported by our observation that genetic features contributed significantly to the performance of the integrated model (Fig. 3A) and suggests that genetic variation is at least to some extent, complementary to clinical information in our setting. In other words, both genetic variation and clinical information each describe different aspects of brivaracetam response. Future work will entail pursuing this hypothesis by testing the model in equivalent datasets derived from other AEDs.

The above discussion demonstrates that our results can provide guidance as to the type of genetic features that are important in determining drug response and which should be considered when translating these approaches into practice, with a focus on genes related to the disease understudy and the drugs' mechanism of action, and pharmacokinetics.

Among the clinical features in the present study, we found that prior levetiracetam use was the most important predictor of brivaracetam response. Previous work has indeed shown that levetiracetam-naïve patients exhibited stronger response to brivaracetam.[76,85] This demonstrates that using ML, we are able to objectively identify variables for which clinical relevance has previously been established in the literature. In the context of this finding, it is important to note that despite showing a weaker response to brivaracetam, levetiracetam-exposed patients showed a better-than-placebo response.[76,85] Moreover, levetiracetam-exposed patients generally showed an attenuated response to a wide range of AEDs, leading to the suggestion that the observed brivaracetam non-response was due to a more general AED-refractoriness.[85] In our data however, we observed that prior use of levetiracetam was uniquely informative for predicting brivaracetam response relative to all other AEDs in the model, likely due to levetiracetam having the same receptor gene as brivaracetam.

In applying our model to responder enrichment for clinical trial design, we observed that it is likely that better models could be built when using more patients than available in our current discovery dataset. However, below about 100 patients, performance was little better than random, suggesting limited utility in responder enrichment for numbers lower than 100. Thus, there is a balance that needs to be considered when applying these approaches to inform clinical study design: Exploratory studies would need to be large enough to support the building of models that would allow for sufficiently strong predictive enrichment in, and reduction in size of, confirmatory studies.

The above considerations represent the first step in the translational pipeline towards clinical practice, supporting the development of companion diagnostics. The actual translation of our present method into routine clinical practice would require several additional steps. First, further validation of our model is required; after internal and retrospective validation as performed here, the next step would be a prospective clinical validation study, in which the benefit compared to standard routes for patient selection has to be demonstrated.[2] After this prospective clinical validation, the test itself, genotyping and algorithm, would need to be made available in a format, and on a platform that could be employed in a routine clinical care environment, such as a hospital. In the current format our model uses 106 clinical and 40 diverse genetic features, comprising 4695 genetic variants, hence being technically challenging and relatively labour and cost

intensive. In this context, it should also be noted that clinical features captured in clinical trials are not necessarily in line with what is measured in routine clinical practice. Future work should thus focus on streamlining the model towards translation into clinical practice, including development of a cost-effective assay for the most informative genetic variants, routes required to make algorithms available and accessible and inclusion of clinical features that might be more routinely collected. A systematic way of assessing feature importance in ML models, such as the SHAP values used in our work, can support this translation by helping to deconvolute the model and identify the features that are most informative.

In case of success, the regulatory implications for an impact on the drug label and companion diagnostic co-approval need to be considered. Regulatory authorities are interested in the ability to target treatments to patients who would most benefit, needing to ensure that the label accurately reflects the enrichment strategies used to select the patients, and that there is an approved method available to identify these patients once any drug is approved. In general, the potential effect of enrichment strategies on labelling and the route of approval of any test as a companion diagnostic should be part of an ongoing dialogue with regulatory bodies during drug development (https://www.fda.gov/media/121320/download; accessed 1 October 2020).

Our work represents the first retrospectively validated ML model linking drug mechanism of action and the genetic, clinical and demographic background of epilepsy patients to clinical drug response. As such, it provides a blueprint for applying ML approaches to achieve the key ambition of precision medicine across a variety of disease areas, including neurological disorders. In particular, extracting informative features from very high dimensional and sparse genomic data is challenging but relevant to many disease areas. Our work demonstrates that even with limited sample size, integrating genetics data with clinical data for informing drug response prediction is possible, when using advanced ML methods combined with literature knowledge to manage the huge dimensionality of the genetics data. Furthermore, such models have substantial potential utility for clinical practice and in future clinical trial design by their ability to enrich for responders. Integration of additional data modalities, such as EEG or imaging, may further improve model performance. Additionally, this ability to identify and characterize subpopulations more likely to respond, shows that within the domain of neurological disorders and specifically epilepsy, we can begin to think more systematically about rational and personalized clinical decision making and reducing the disease burden for patients on an individual level.

Acknowledgements

Funding

Competing interests

All authors received salaries from UCB Pharma. UCB Pharma had no influence on the content of this work.

Supplementary material

Supplementary material is available at *Brain* online.

References

1. Vogenberg FR, Isaacson Barash C, Pursel M. Personalized medicine: Part 1: Evolution and development into theranostics. *P & T*. 2010;35(10):560–576.

2. Fröhlich H, Balling R, Beerenwinkel N, et al. From hype to reality: Data science enabling personalized medicine. *BMC Med*. 2018;16(1):150.

3. FDA. Table of pharmacogenomic biomarkers in drug labeling. 2020. Accessed 1 October 2020. https://www.fda.gov/drugs/science-and-research-drugs/table-pharmacogenomic-biomarkers-drug-labeling

4. Armstrong M. The Genetics of Adverse Drug Reactions. In: Cohen N, ed. *Pharmacogenomics and Personalized Medicine. Methods in Pharmacology and Toxicology*. Humana Press; 2008: 121–147.

5. Peck RW. Precision medicine is not just genomics: The right dose for every patient. *Ann Rev Pharmacol Toxicol*. 2018;58(1): 105–122.

6. Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. *Trends Pharmacol Sci*. 2019;40(8):577–591.

7. De Martini D. Empowering phase II clinical trials to reduce phase III failures. *Pharm Stat*. 2020;19(3):178–186.

8. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. *N Engl J Med*. 2004;351(27):2817–2826.

9. van 't Veer LJ, Dai H, van de Vijver MJ, et al. Gene expression profiling predicts clinical outcome of breast cancer. *Nature*. 2002;415(6871):530–536.

10. Strafella C, Caputo V, Galota MR, et al. Application of precision medicine in neurodegenerative diseases. *Front Neurol*. 2018;9: 701.

11. Tan L, Jiang T, Tan L, Yu J-T. Toward precision medicine in neurological diseases. *Ann Transl Med*. 2016;4(6):104.

12. Daniel F, French J. Designing better trials for epilepsy medications: The challenge of heterogeneity. *Clin Invest*. 2013;3(10): 927–934.

13. CenterWatch. *FDA approved drugs: Listings in epilepsy*. 2020. Accessed 1 October 2020. https://www.centerwatch.com/directories/1067-fda-approved-drugs/topic/184-epilepsy

14. Kwan P, Brodie MJ. Early identification of refractory epilepsy. *N Engl J Med*. 2000;342(5):314–319.

15. 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):279–286.

16. Perucca E, Brodie MJ, Kwan P, Tomson T. 30 years of second-generation antiseizure medications: Impact and future perspectives. *Lancet Neurol*. 2020;19(6):544–556.

17. Laxer KD, Trinka E, Hirsch LJ, et al. The consequences of refractory epilepsy and its treatment. *Epilepsy Behav*. 2014;37:59–70.

18. Begley CE, Durgin TL. The direct cost of epilepsy in the United States: A systematic review of estimates. *Epilepsia*. 2015;56(9): 1376–1387.

19. Aronica E, Gorter JA, Ramkema M, et al. Expression and cellular distribution of multidrug resistance–related proteins in the hippocampus of patients with mesial temporal lobe epilepsy. *Epilepsia*. 2004;45(5):441–451.

20. Feldmann M, Asselin M-C, Liu J, et al. P-glycoprotein expression and function in patients with temporal lobe epilepsy: A case-control study. *Lancet Neurol*. 2013;12(8):777–785.

21. Klotz U. The role of pharmacogenetics in the metabolism of antiepileptic drugs. *Clin Pharmacokinetics*. 2007;46(4):271–279.

22. Ghosh C, Marchi N, Desai NK, et al. Cellular localization and functional significance of CYP3A4 in the human epileptic brain. *Epilepsia*. 2011;52(3):562–571.

23. Emich-Widera E, Likus W, Kazek B, et al. CYP3A53 and C3435T MDR1 polymorphisms in prognostication of drug-resistant epilepsy in children and adolescents. *BioMed Res Int*. 2013;2013: 526837.

24. Boussadia B, Ghosh C, Plaud C, et al. Effect of status epilepticus and antiepileptic drugs on CYP2E1 brain expression. *Neurosci*. 2014;281:124–134.

25. Dauchy S, Dutheil F, Weaver RJ, et al. ABC transporters, cytochromes P450 and their main transcription factors: Expression at the human blood–brain barrier. *J Neurochem*. 2008;107(6): 1518–1528.

26. Nurmohamed L, Garcia-Bournissen F, Buono RJ, Shannon MW, Finkelstein Y. Predisposition to epilepsy—Does the ABCB1 gene play a role? *Epilepsia*. 2010;51(9):1882–1885.

27. Grover S, Kukreti R. A systematic review and meta-analysis of the role of ABCC2 variants on drug response in patients with epilepsy. *Epilepsia*. 2013;54(5):936–945.

28. Bournissen FG, Moretti ME, Juurlink DN, Koren G, Walker M, Finkelstein Y. Polymorphism of the MDR1/ABCB1 C3435T drug-transporter and resistance to anticonvulsant drugs: A meta-analysis. *Epilepsia*. 2009;50(4):898–903.

29. Balan S, Bharathan SP, Vellichiramal NN, et al. Genetic association analysis of ATP binding cassette protein family reveals a novel association of ABCB1 genetic variants with epilepsy risk, but not with drug-resistance. *PLoS One*. 2014;9(2): e89253–10.

30. van der Weide J, Steijns LSW, van Weelden MJM, de Haan K. The effect of genetic polymorphism of cytochrome P450 CYP2C9 on phenytoin dose requirement. *Pharmacogenet Genom*. 2001;11(4): 287–291.

31. Dombrowski SM, Desai SY, Marroni M, et al. Overexpression of multiple drug resistance genes in endothelial cells from patients with refractory epilepsy. *Epilepsia*. 2001;42(12): 1501–1506.

32. Cavalleri GL, McCormack M, Alhusaini S, Chaila E, Delanty N. Pharmacogenomics and epilepsy: The road ahead. *Pharmacogenomics*. 2011;12(10):1429–1447.

33. Silva-Alves MS, Secolin R, Carvalho BS, et al. A prediction algorithm for drug response in patients with mesial temporal lobe epilepsy based on clinical and genetic information. *PLoS One*. 2017;12(1):e0169214.

34. Petrovski S, Szoeke CE, Sheffield LJ, D'souza W, Huggins RM, O'brien TJ. Multi-SNP pharmacogenomic classifier is superior to single-SNP models for predicting drug outcome in complex diseases. *Pharmacogenet Genom*. 2009;19(2):147–152.

35. Shazadi K, Petrovski S, Roten A, et al. Validation of a multigenic model to predict seizure control in newly treated epilepsy. *Epilepsy Res*. 2014;108(10):1797–1805.

36. Devinsky O, Dilley C, Ozery-Flato M, et al. Changing the approach to treatment choice in epilepsy using big data. *Epilepsy Behav*. 2016;56:32–37.

37. Zhang J-H, Han X, Zhao H-W, et al. Personalized prediction model for seizure-free epilepsy with levetiracetam therapy: A retrospective data analysis using support vector machine. *Br J Clin Pharmacol*. 2018;84(11):2615–2624.

38. Yao L, Cai M, Chen Y, Shen C, Shi L, Guo Y. Prediction of antiepileptic drug treatment outcomes of patients with newly diagnosed epilepsy by machine learning. *Epilepsy Behav*. 2019;96: 92–97.

39. Steinhoff BJ, Staack AM. Levetiracetam and brivaracetam: A review of evidence from clinical trials and clinical experience. *Therap Adv Neurol Disord*. 2019;12:1756286419873518.

40. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2009;25(14): 1754–1760.

41. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. *Bioinformatics*. 2010;26(5):589–595.

42. Picard Toolkit. Broad Institute; 2019. Accessed 1 October 2020. http://broadinstitute.github.io/picard/

43. McKenna A, Hanna M, Banks E, et al. The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.

44. Rentzsch P, Witten D, Cooper G, Shendure J, Kircher M. CADD: Predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886–D894.

45. Zhao H, Sun Z, Wang J, Huang H, Kocher J-P, Wang L. CrossMap: A versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*. 2014;30(7):1006–1007.

46. Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21): 2987–2993.

47. Cingolani P, Platts A, Wang LL, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. *Fly (Austin)*. 2012;6(2):80–92.

48. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.

49. Köhler S, Carmody L, Vasilevsky N, et al. Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res*. 2019;47(D1):D1018–D1027.

50. Buniello A, MacArthur JAL, Cerezo M, 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–D1012.

51. Piñero J, Ramírez-Anguita JM, Saüch-Pitarch J, et al. The DisGeNET knowledge platform for disease genomics: 2019 update. *Nucleic Acids Res*. 2020;48(D1):D845–D855.

52. Carvalho-Silva D, Pierleoni A, Pignatelli M, et al. Open targets platform: New developments and updates two years on. *Nucleic Acids Res*. 2019;47(D1):D1056–D1065.

53. Vejnar CE, Blum M, Zdobnov EM. miRmap web: Comprehensive microRNA target prediction online. *Nucleic Acids Res*. 2013; 41(Web Server issue):W165–W168.

54. Mooney C, Becker BA, Raoof R, Henshall DC. EpimiRBase: A comprehensive database of microRNA-epilepsy associations. *Bioinformatics*. 2016;32(9):1436–1438.

55. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;454, (6):580–585.

56. Alexa A, Rahnenfuhrer J. TopGO: Enrichment analysis for gene ontology. Accessed 1 October 2020. http://www.bioconductor.org/packages/release/bioc/html/topGO.html

57. Nikolsky Y, Kirillov E, Zuev R, Rakhmatulin E, Nikolskaya T. Functional analysis of OMICs data and small molecule compounds in an integrated "knowledge-based" platform. *Methods Mol Biol*. 2009;563:177–196.

58. Wang J, Lin Z-J, Liu L, et al. Epilepsy-associated genes. *Seizure Eur J Epilepsy*. 2017;44:11–20.

59. Denny JC, Bastarache L, Ritchie MD, et al. Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data. *Nat Biotechnol*. 2013;31(12):1102–1111.

60. Bischl B, Richter J, Bossek J, Horn D, Thomas J, Lang M. MlrMBO: A modular framework for model-based optimization of expensive black-box functions. 2017. Accessed 1 October 2020. http://arxiv.org/abs/1703.03373

61. Rohart F, Gautier B, Singh A, Lê Cao K-A. mixOmics: An R package for 'omics feature selection and multiple data integration. *PLoS Comput Biol*. 2017;13(11):e1005752–19.

62. Abadi M, Agarwal A, Barham P, et al. TensorFlow: Large-scale machine learning on heterogeneous systems. 2015. Accessed 1 October 2020. https://www.tensorflow.org/

63. Yuan M, Lin Y. Model selection and estimation in regression with grouped variables. *J R Stat Soc*. 2006;68(1):49–67.

64. Tibshirani R. Regression shrinkage and selection via the Lasso. *J R Stat Soc*. 1996;58(1):267–288.

65. Tikhonov AN, Vy A. *Solution of ill-posed problems*. New York: John Wiley & Sons; 1977.

66. Hinton GE, Srivastava N, Krizhevsky A, Sutskever I, Salakhutdinov RR. Improving neural networks by preventing co-adaptation of feature detectors. *CoRR*. 2012;abs/1207.0580. Accessed 1 October 2020. http://arxiv.org/abs/1207.0580

67. Ioffe S, Szegedy C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In: *Proceedings of the 32nd International Conference on International Conference on Machine Learning - Volume 37*. JMLR; 2015:448–456.

68. Kingma DP, Ba J. Adam: A method for stochastic optimization. In: Bengio Y, LeCun Y, eds. *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*; 2015. Accessed 1 October 2020. http://arxiv.org/abs/1412.6980

69. Mockus JB, Mockus LJ. Approach to global optimization and application to multiobjective and constrained problems. *J Optim Theory Appl*. 1991;70(1):157–172.

70. Friedman J, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1–22.

71. Chen T, Guestrin C. XGBoost: A scalable tree boosting system. In: *KDD '16*. Association for computing machinery; 2016:785-794. doi:10.1145/2939672.2939785

72. Wolpert DH. Stacked generalization. *Neural Networks*. 1992;5(2):241–259.

73. ClinicalTrials.gov. Double-blind, randomized study evaluating the efficacy and safety of brivaracetam in adults with partial onset seizures; 2016. Accessed 1 October 2020. https://clinicaltrials.gov/ct2/show/NCT00464269

74. ClinicalTrials.gov. Brivaracetam efficacy and safety study in subjects with partial onset seizures; 2016. Accessed 1 October 2020. https://clinicaltrials.gov/ct2/show/NCT01261325

75. Biton V, Berkovic SF, Abou-Khalil B, Sperling MR, Johnson ME, Lu S. Brivaracetam as adjunctive treatment for uncontrolled partial epilepsy in adults: A phase III randomized, double-blind, placebo-controlled trial. *Epilepsia*. 2014;55(1):57–66.

76. Klein P, Schiemann J, Sperling MR, et al. A randomized, double-blind, placebo-controlled, multicenter, parallel-group study to evaluate the efficacy and safety of adjunctive brivaracetam in adult patients with uncontrolled partial-onset seizures. *Epilepsia*. 2015;56(12):1890–1898.

77. Rheims S, Ryvlin P. Pharmacotherapy for tonic–clonic seizures. *Expert Opin Pharmacother*. 2014;15(10):1417–1426.

78. Johnstone IM, Titterington DM. Statistical challenges of high-dimensional data. *Philos Trans R Soc A Math Phys Eng Sci*. 2009;367(1906):4237–4253.

79. Singh A, Shannon CP, Gautier B, et al. DIABLO: An integrative approach for identifying key molecular drivers from multi-omics assays. *Bioinformatics*. 2019;35(17):3055–3062.

80. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc B*. 2005;67(2):301–320.

81. Ahmad A, Fröhlich H. Integrating heterogeneous omics data via statistical inference and learning techniques. *Genom Comput Biol*. 2016;2(1):e32.

82. Pavlidis P, Weston J, Cai J, Grundy WN. Gene functional classification from heterogeneous data. In: RECOMB '01: *Proceedings of the Fifth Annual International Conference on Computational Biology*. Association for Computing Machinery; 2001:249–255.

83. Maragos P, Gros P, Katsamanis A, Papandreou G. Cross-modal integration for performance improving in multimedia: A review. In: Maragos P, Potamianos A, Gros P, eds. Multimedia processing and interaction, audio, video, text. *Multimedia Systems and Applications*. Vol 33. New York: Springer; 2008:3–45.

84. Lundberg SM, Lee S-I, et al. A unified approach to interpreting model predictions. In: Guyon I, Luxburg UV, Bengio S, eds. *Advances in neural information processing systems 30*. New York: Curran Associates, Inc.; 2017:4765–4774.

85. Asadi-Pooya AA, Sperling MR, Chung S, et al. Efficacy and tolerability of adjunctive brivaracetam in patients with prior antiepileptic drug exposure: a post-hoc study. *Epilepsy Res*. 2017;131:70–75.

86. Bonini SA, Mastinu A, Ferrari-Toninelli G, Memo M. Potential role of microtubule stabilizing agents in neurodevelopmental disorders. *Int J Mol Sci*. 2017;18(8):1627.

87. Lasser M, Tiber J, Lowery LA. The role of the microtubule cytoskeleton in neurodevelopmental disorders. *Front Cell Neurosc*. 2018;12:165.

88. Hwang TJ, Carpenter DP, Lauffenburger JC, Wang B, Franklin JM, Kesselheim AS. Failure of investigational drugs in late-stage clinical development and publication of trial results. *JAMA Intern Med*. 2016;176(12):1826–1833.

89. Klitgaard H, Matagne A, Nicolas J-M, et al. Brivaracetam: Rationale for discovery and preclinical profile of a selective SV2A ligand for epilepsy treatment. *Epilepsia*. 2016;57(4):538–548.

90. Lynch JM, Tate SK, Kinirons P, et al. No major role of common SV2A variation for predisposition or levetiracetam response in epilepsy. *Epilepsy Res*. 2009;83(1):44–51.