In Silico Drug Repurposing in Parkinson's Disease

Patrice Godard, Matthew Page and Jonathan van Eyll

UCB, Chemin du Foriest, Braine-l'Alleud, Belgium.

UCB, 216 Bath Road, Slough, Berkshire, UK.

Introduction

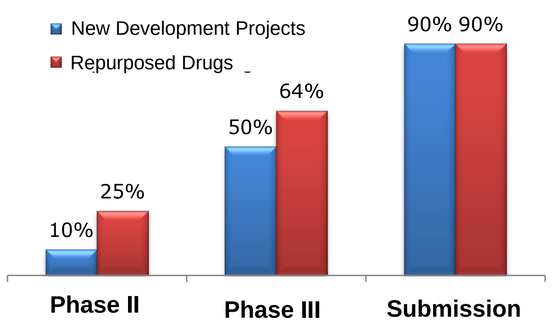


Fig. 1. Drug approval rates in clinical development. Repurposed drugs have a greater probability of success than new development projects. Data from the Centre for Medicines Research (CMR) International Pharmaceutical R&D Factbook.

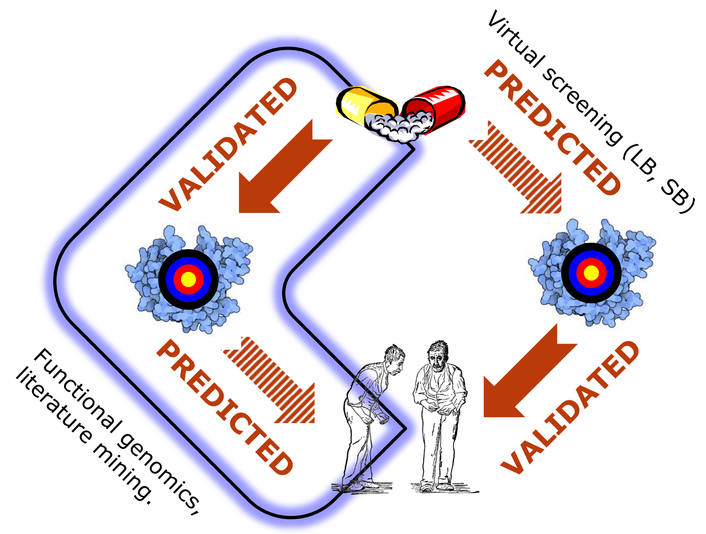
Costly, protracted drug development, coupled with increasingly high clinical attrition rates, have fuelled the pharmaceutical industry's interest in drug repurposing strategies. Drug repurposing is the rational application of a known drug to new indications and can lead to shorter, less costly drug development cycles with increased probability of success*.

Parkinson's Disease (PD) is the 2nd most common neurological disorder, involving progressive disruption of motor function with possible psychiatric complications, caused by depletion of dopaminergic neurons in the nigrostriatal system. Drug repurposing represents a potential strategy for cost saving and risk mitigation in the treatment of neurodegenerative diseases, including PD, where drug approval rates are low (~8%†).

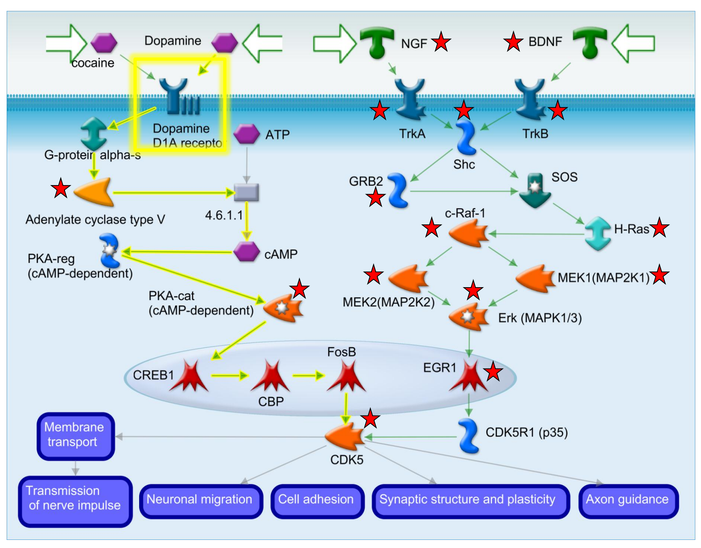
A data-driven strategy was adopted to identify drugs in clinical development that target molecules relevant to the patho-physiology of PD but that are not indicated for PD. Systems-level genomics data sets were combined with commercial knowledge management resources to enable a rational prioritisation of repurposing candidates.

*Ashburn, T.T. & Thor, K.B. (2004). Drug repositioning: identifying and developing new uses for existing drugs. *Nat Rev Drug Discov* **3**, 673–683.
†Kola, I. & Landis, J. (2004). Can the pharmaceutical industry reduce attrition rates? *Nat Rev Drug Discov* **3**, 711–715.

In Silico Drug Repurposing: Target Re-indicating Vs Drug Re-targeting



Feature Generation



Drug target

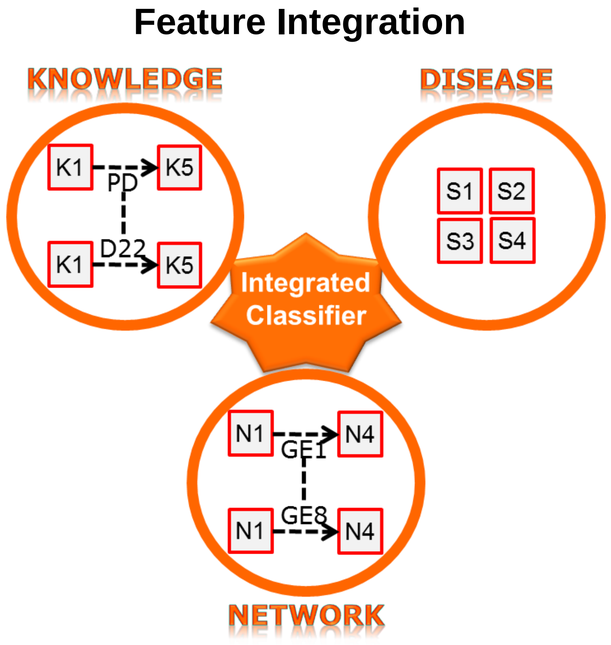
Drug target for PD

Linear pathway involving a PD drug target

Knowledge Based Features

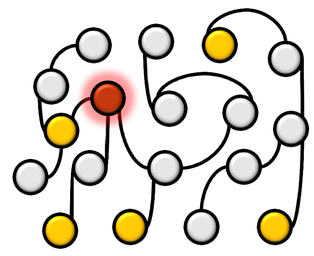
Capture *a priori* knowledge of the molecular aetiology of PD using curated disease pathways, therapeutically precedented mechanisms and molecular interrelations to known disease biomarkers. 5 knowledge-based features were captured for 22 diseases.

Feature Integration



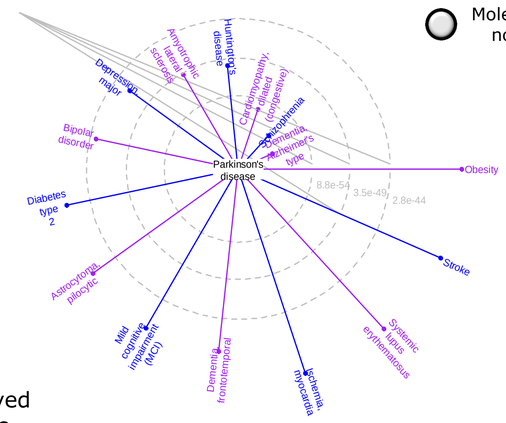
Drug, biomarker and pathway information were retrieved from Thomson Reuters' Integrity and MetaBase databases.

THOMSON REUTERS



Disease Based Features

Grossly approximate the similarity between PD and other diseases based on overlap between reported biomarkers and biomarker-enriched pathways. 4 Different similarity methods were applied.

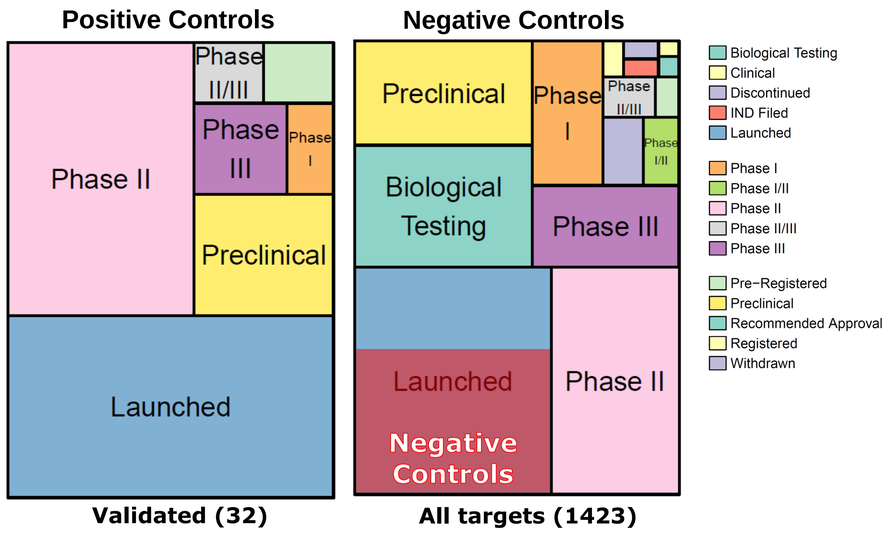


Molecular node | Differentially expressed gene | Candidate

Network Based Features

Prioritise topologically significant actors in the transcriptional regulatory network underlying PD using 4 different network metrics (e.g. visitation frequency from a Random Walk). Input gene signatures were derived from 8 public PD case-control genomic studies identified using NextBio.

Model Training and Selection

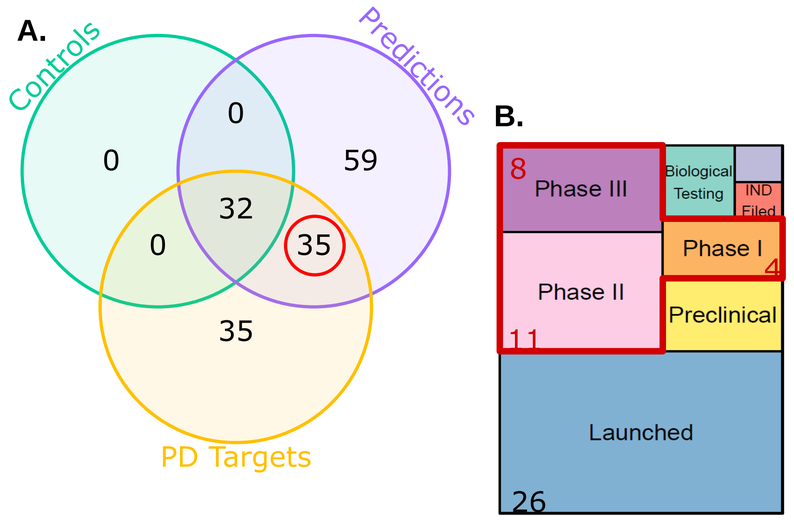


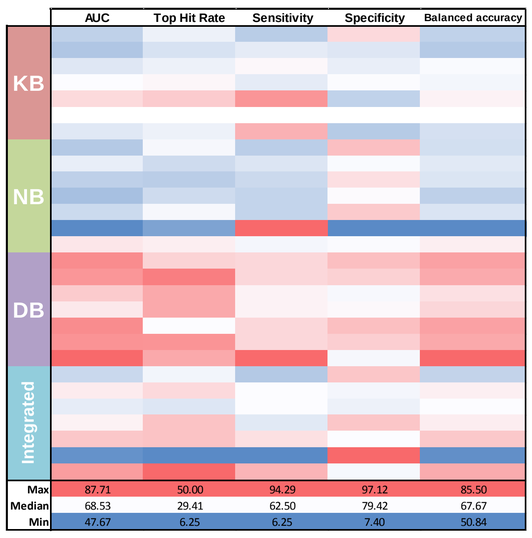
The 3 feature classes, totalling 146 features and encoding knowledge-, disease- and network-based considerations were used to enumerate a set of positive and negative PD drug controls.

The positive control set consisted of 32 targets of drugs with a "validated" status for PD in Integrity. The negative control set comprised 320 launched, non-PD drug targets.

Fig. 2. Development status of the positive and negative PD drug target controls. Positive controls: a *validated* status means the target is associated with the mechanism of action of a drug under active preclinical/clinical development or launched in PD. Negative controls: launched drug targets not associated with PD, either directly or indirectly, through related compounds or genes.

Prediction Results





6 different classification algorithms (including SVM, Random Forest and LDA) were trained using the 3 feature groups as input vectors both individually or in combination. A 10-fold cross validation procedure, together with ROC curve analysis was used to select discrimination thresholds and assess model performance.

A Random Forrest classifier, incorporating all 146 input features, was selected as the optimal model using the top hit rate metric.

Fig. 3. Global comparison of all classification models. A number of metrics of model performance were considered. Red and blue cell intensities correlate with distance above and below the median value for each performance metric.

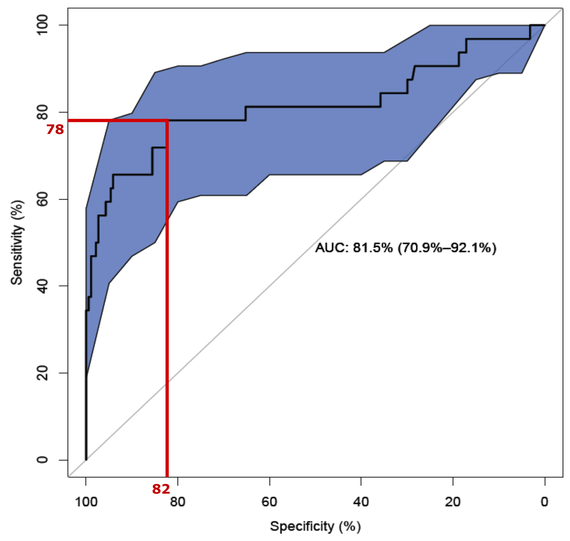


Fig. 4. Receiver operating characteristic curve analysis. Compares how sensitivity scales with specificity for different model discrimination thresholds. Thresholds were selected to maximise the balanced accuracy (red intersect), which avoids inflated performance estimates on imbalanced datasets. The area under the ROC curve (AUC) correlates with model performance and can be used as a standardised metric for method comparison.

The selected Random Forest classifier was run on all the available 1423 drug targets in Integrity and predicted a total of 126 repurposing candidates for PD (Fig. 5A). 35 of the 70 targets associated to PD but without a validated status, were also identified in the predictions. This overlap is highly significant (p-value<10^{-18}) and provides confidence for the remaining 59 predictions.

Of the 59 remaining predictions, 23 are products undergoing active clinical development with no existing consideration for PD (Fig. 5B) This category of repurposing candidates is of interest because clinical safety and tolerance data is available but the potential remains to be first in class in PD.

Conclusions

- This project represents an integrative, data-driven approach to enable unbiased, rational identification of drug repurposing opportunities in PD.
- Numerous, disparate biological information resources and systems-level genomic data sets were integrated with the aid of commercial knowledge management platforms (e.g. NextBio, Integrity, MetaBase).
- The 23 repurposing candidates were reviewed in detail and a final subset selected for follow-up *in vivo*. A similar initiative is now underway in Epilepsy.

Acknowledgments

UCB: Benoît Kenda and Dieter Scheller.

Thomson Reuters: Marina Bessarabova, Dorothea Emig, Alexander Ivliev, Lee Lancashire and Eugene Mishkin.